Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 6024–6028, August 1990
Microbiology

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*

(plus strand synthesis/template switching/RNase H)

VINAY K. PATHAK AND HOWARD M. TEMIN†

McArdle Laboratory for Cancer Research, 1400 University Avenue, University of Wisconsin, Madison, WI 53706

Contributed by Howard M. Temin, May 24, 1990

ABSTRACT **In the preceding paper we described an experiment that determined the *in vivo* forward mutation rate in a single replication cycle for spleen necrosis virus. In addition to substitutions, frameshifts, and hypermutations, the mutated proviruses contained two classes of deletions. One class of deletions contained short direct repeats at the deletion junctions. Another class of deletions had short stretches of sequences inserted at the deletion junctions. In this report, we describe the deletion mutations, and we present models for their generation. Detailed analysis of two deletions with insertions indicates that these mutations occurred as a result of template switching during plus-strand DNA synthesis. The analysis also indicates that fragments of viral RNA generated by the viral RNase H endonuclease are used as templates and contribute to the sequences inserted at the deletion junctions.**

It is commonly observed that when animal viruses are passaged at high multiplicities of infection, deletion mutants arise (1). These mutants are called defective interfering particles.

Deletions in retroviral genomes appear to occur frequently, and they may be an obligatory step in the transduction of oncogenes (2).

When Rous sarcoma virus was subjected to 50 undiluted passages, mutants deleted at two separate portions of the genome appeared (3). These two deletions did not arise through homologous recombination, and different deletion mutants had similar but not identical deletion end points.

In the course of sequencing of spleen necrosis virus (SNV) DNA and host cell DNA junctions, long terminal repeats (LTRs) of differing sizes were obtained at a high frequency (4).

Dougherty and Temin (5) constructed an SNV-based retroviral vector that contained two selectable marker genes, *hygro* and *neo*. In this vector the *neo* gene was expressed only in vectors that sustained mutations. The mutations that allowed *neo* expression occurred at a frequency of 0.5% per cycle of replication. These mutations were primarily deletions (G. A. Pulsinelli, personal communication).

In the preceding paper (6), we reported the development of an *in vivo* forward mutation assay for SNV. In addition to single nucleotide changes (substitutions, hypermutations, and frameshifts), we also observed many deletions. These deletions could be classified into two classes: deletions and deletions with insertions. In this report, we describe the nature of these deletions and suggest models for their generation.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. Rates of deletion

Mtn. type (mutant)	Nucleotide change(s), bp	Mtn. rate, mtn. per bp per cycle*
Deletion		2×10^{-6}
D1 (5, 13, 21)	−8	
D3 (93, 95, 96, 101, 103)	−23	
D4 (15, 16)	−330	
D2 (78)	−376	
D5 (6)†	−388	
Deletion with insertion		2×10^{-6}
DI1 (104)	−630 +14	
DI2 (41)	−8 +13	
DI3 (3, 7, 8, 10, 12)	−1006 +39	
DI4 (102)	−3 +1	

*Mutation rates of deletion and deletion with insertion were calculated as follows: 11 deletions per 16,867 total proviruses recovered per 280 target nucleotides (length of *lacZα* gene plus the *lacZα* promoter).

†D5 (mutant no. 6) was phenotypically blue; it was identified through restriction mapping and DNA sequencing. Therefore, this mutation was not used in the calculation of the deletion mutation rate.

MATERIALS AND METHODS

Plasmid Construction. Retroviral shuttle vector plasmids pVP212 and pVP232 were constructed with DNA fragments taken from pUC19, pMC, and JD215 (5, 7, 8). Standard DNA cloning procedures were used (9). A detailed description of the cloning steps is available on request.

Protocol for *lac* Repressor-Mediated Purification of Proviral DNA. A method was developed for obtaining a 100-fold enrichment of proviral DNAs from genomic DNA based on the tight and specific binding of the *lac* repressor to the *lac* operator. The details of the method were previously described (6).

DNA Sequencing. All plasmid sequencing was performed using the Sequenase kit and protocols (United States Biochemical), using the dGTP reactions.

Cells, Transfection, and Infection. The handling of D17 cells and .2G helper cells, DNA transfections, virus harvesting, and virus infections were as previously described (6).

RESULTS

Rate of Deletion. Four mutant proviruses containing simple deletions (D1–D4) were identified by their white colony

Abbreviations: SNV, spleen necrosis virus; LTR, long terminal repeat; PPT, polypurine tract.

*This paper is no. 2 in a series. Paper no. 1 is ref. 6.

†To whom reprint requests should be addressed.

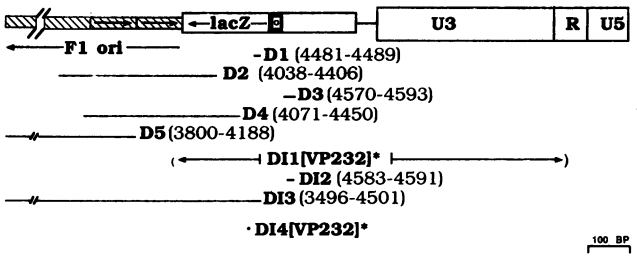

FIG. 1. Deletions and deletions with insertions in the VP212 and VP232 proviruses. An expanded view of a portion of the F1 origin of replication containing direct repeats of 110 nucleotides (hatched boxes with arrows), the *lacZα* gene (open box with arrow), the *lac* operator sequence (black box marked o), and the 3′ LTR are shown. The positions and lengths of deletions (D1–D5) and deletions with insertions (DI1–DI4) are shown. The numbers in parentheses refer to numbers beginning at the end of the *Sac* I restriction site near the beginning of the 5′ LTR. The asterisks indicate deletion with insertion clones that were derived from a VP232 provirus in which the *lacZα* gene is in the opposite orientation; consequently to avoid confusion sequence position numbers are not provided. U3, unique 3′ region in SNV LTR; U5, unique 5′ region in SNV LTR.

phenotype; mutant provirus D5 exhibited a blue colony phenotype and contained a deletion in the F1 origin of replication (Table 1). The rate of occurrence of simple deletions was 2×10^{-6} per base pair (bp) per cycle.

Rate of Deletion with Insertion. Proviruses containing deletions with insertions (DI1–DI4) had deletion of sequences and additional nucleotides inserted at the deletion junctions (Table 1). The rate of occurrence of DI mutations was 2×10^{-6} per bp per cycle.

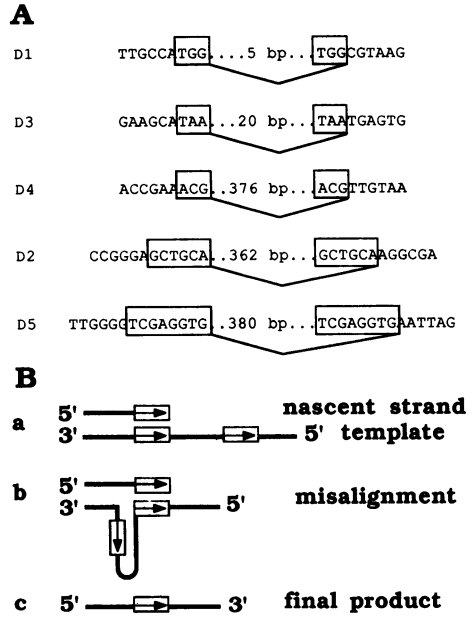

FIG. 2. Simple deletions and model for their generation. (*A*) Plus strand DNA sequence of the upstream and downstream deletion junction sequences for simple deletions D1–D5. Direct repeats of 3–8 nucleotides present at deletion junctions are shown in open boxes. The length of sequence present between the direct repeats is indicated. In each case, the sequence between the direct repeats as well as one copy of the direct repeat was deleted. (*B*) Model for generation of deletions. (*a*) The nascent strand growing point reaches the end of one copy of the direct repeat. (*b*) Template misalignment occurs, and the second copy of the direct repeat in the template strand is aligned with homologous sequences on the nascent strand. (*c*) Synthesis of the nascent strand is completed to generate the final product. The direction of DNA synthesis is arbitrary and does not imply plus-strand synthesis.

Schematic Map of Deletions. Fig. 1 shows the locations of deletions and deletions with insertions. As expected, most of the deletion mutations altered the open reading frame of the *lacZα* gene; two mutations (D3 and DI2) altered sequences close to the translational start, presumably reducing the expression of the *lacZα* peptide and thus generating the light-blue colony phenotype.

Deletions Have Small Direct Repeats at Deletion Junctions. Five deletions involved deletion of sequences of 8–388 bp, without addition of nucleotides at deletion junctions (see Fig. 2*A*). These deletions were characterized by the presence of small direct repeats of 3–8 nucleotides at the deletion junctions, suggesting that in the course of DNA synthesis template misalignments occurred within these blocks of homology as outlined in Fig. 2*B* (discussed later).

Deletion D5 was identified by restriction enzyme analysis of 800 proviral plasmid clones of blue colony phenotype (data not shown). DNA sequencing of the F1 origin of replication of this clone revealed the unselected deletion mutation.

Deletions with Insertions Involve Template Switching and May Involve RNA Fragments as Templates. A general scheme for synthesis of a DNA molecule from viral genomic RNA is given in Fig. 3. Four deletions with insertions at the deletion junctions were characterized (Figs. 4–6). Models to account for the nucleotides inserted at the deletion junctions of two mutant proviruses are presented in Figs. 4 and 5 and are

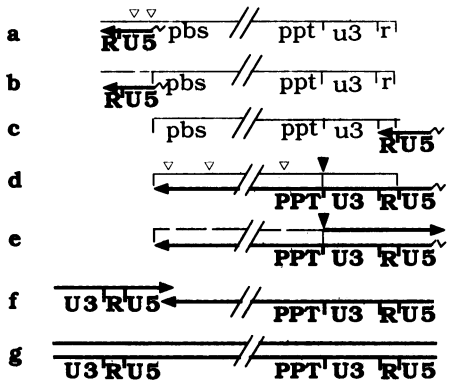

FIG. 3. Synthesis of a DNA molecule from viral genomic RNA. Retroviral DNA polymerization steps involved in the generation of a double-stranded DNA molecule from single-stranded genomic RNA. For convenience, events for only one of the two genomic RNA molecules and intrastrand minus strand strong stop DNA transfer are depicted (W.-S. Hu, personal communication). Thin lines and lowercase letters designate RNA, whereas thick lines and capital letters designate DNA. Abbreviations: r or R, short terminal repeats at the ends of viral RNA; u5 or U5, unique 5′ sequence of viral RNA that is duplicated in generation of the LTRs; u3 or U3, unique 3′ sequence of viral RNA that is duplicated in generation of the LTRs; pbs or PBS, tRNA primer binding site that primes the synthesis of minus-strand DNA; ppt or PPT, polypurine tract that primes the synthesis of plus-strand DNA. Vertical bars below lines are boundaries between regions of RNA or DNA; ▽, possible RNase H digestion sites; ▼, specific cuts by RNase H required to begin plus-strand DNA synthesis. Slashes indicate sequences encoding viral proteins that are not shown. Zig-zag lines designate primer tRNA. (*a*) Primer tRNA hybridized to pbs initiates minus-strand DNA synthesis. (*b*) RNase H digests viral r-u5 RNA in RNA·DNA hybrids. (*c*) Minus-strand DNA is hybridized to the 3′ end of viral RNA. (*d*) Minus-strand DNA synthesis progresses toward the 5′ end of pbs. RNase H further digests viral RNA in RNA·DNA hybrids. A specific RNase H cut is made at the end of ppt. (*e*) Plus-strand DNA synthesis is initiated at ppt and progresses past the 5′ end of minus-strand DNA to 18 nucleotides into the primer tRNA. RNase H digests primer tRNA. Minus-strand DNA synthesis may not be completed at this step. (*f*) Plus-strand DNA hybridizes to the PBS sequence of minus-strand DNA. (*g*) DNA synthesis on both the minus and the plus strand completes synthesis of a double-stranded DNA molecule with two LTRs.

discussed later (see *Discussion*). The models involve template switching and utilization as a template of viral RNA fragments generated by RNase H.

Deletion with insertion DI1 had a deletion of 630 bp near the 3′ end of a VP232-derived provirus and an insertion of 14 nucleotides (Figs. 1 and 4). Sequences including the 3′ end of the *lacZα* gene, the SNV polypurine tract (PPT), the 3′ SNV LTR U3 sequence, and a portion of the R region of the 3′ SNV LTR were deleted. Sequences complementary to 14 nucleotides from the 5′ end of the vector, 427–413, were inserted at the deletion junction of DI1 (Fig. 4).

Deletion with insertion DI2 had a deletion of 8 nucleotides within the *lacZα* gene promoter sequences and an insertion of 13 nucleotides (Figs. 1 and 5). Sequences complementary to 10 nucleotides from the middle of the vector, 2503–2494, were inserted at the deletion junction of DI2 (Fig. 5). An A residue 5′ to the 10 nucleotides and AG residues 3′ to the 10 nucleotides were also inserted.

Deletion with insertion DI3 had a deletion of 1006 nucleotides of the F1 and *lacZα* sequences and an insertion of 39 nucleotides at the deletion junction (Figs. 1 and 6). Sequences from elsewhere in VP212 that were inserted at the deletion junction and their nucleotide positions are shown in Fig. 6. Three of the homologous sequences are located within a 69-bp stretch (573–642).

Deletion with insertion DI4 had a deletion of 3 nucleotides (ATT) in the *lacZα* gene and an insertion of one nucleotide (G) at the deletion junction (Figs. 1 and 6). A 2-nucleotide direct repeat (CA) was present at the deletion junction in the plus-strand DNA. The CA nucleotides at the upstream deletion junction were preceded by a G nucleotide.

DISCUSSION

Deletions and Lack of Duplications. The deletions found in this study were characterized by the presence of direct repeats at the deletion junctions. Direct repeats at deletion junctions in cellular DNAs have been found in spontaneous deletions and in ionizing radiation-induced deletions (refs. 10–12; D. Klinedinst and N. Drinkwater, personal communication).

It is noteworthy that we found no duplications in the set of mutations analyzed here. This result indicates that duplications are much less common than deletions. The model outlined in Fig. 2*B* for simple deletions indicates why duplications occur less frequently than deletions. In order for a duplication to occur, the growing point of the nascent strand must progress to the end of the second direct repeat. Next, a misalignment of the growing point and the first direct repeat of the template strand must occur. Since the DNA between the direct repeats is now double stranded, hydrogen bonding

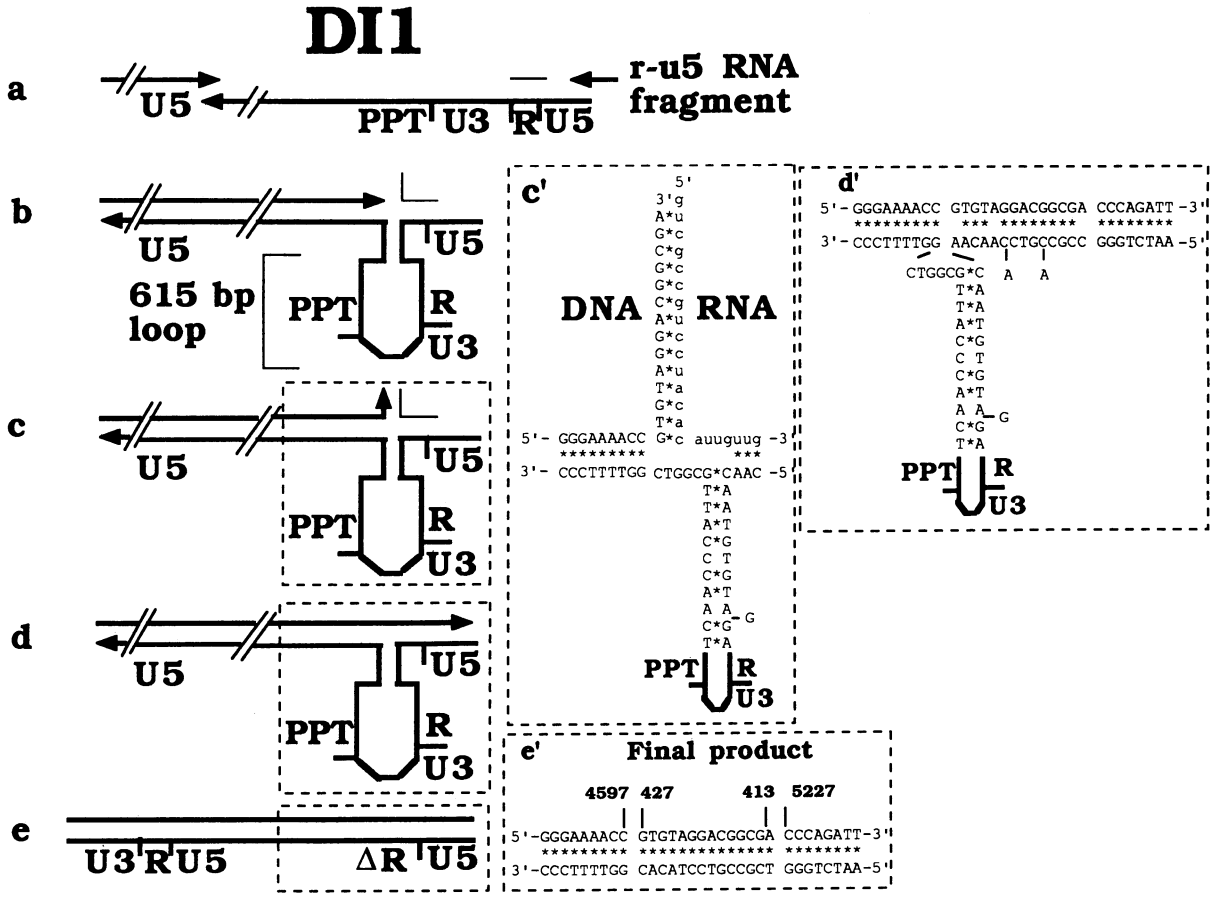

FIG. 4. Model for generation of deletion with insertion DI1. All abbreviations are the same as described in the legend of Fig. 3. Asterisks between nucleotides indicate complementary base pairs. (*a*) An r-u5 RNA fragment resulting from RNase H digestion of RNA·DNA hybrids (see Fig. 3*b*) hybridizes to the 3′ R-U5 sequence of minus-strand DNA. (*b*) Plus-strand DNA synthesis progresses to the upstream deletion junction next to the partially displaced r-u5 RNA. A 615-bp stem–loop forms in the minus-strand DNA. (*c*) The plus-strand DNA growing point switches templates to the r sequence of the partially displaced RNA and progresses toward the 5′ end of RNA, ceasing DNA synthesis 1 bp away from the 5′ end of RNA (see *c′* for details). (*d*) RNA is released from the DNA by RNase H digestion. The growing point of plus-strand DNA switches back and hybridizes to a 14-bp sequence (11/14 nt) on the minus-strand DNA just upstream of the 3′ deletion junction; thus the growing point of plus-strand DNA slips forward by four base pairs and is now juxtaposed to the downstream deletion junction (see *d′* for details). (*e*) Plus-strand DNA synthesis is completed to generate the final product (see *e′* for details).

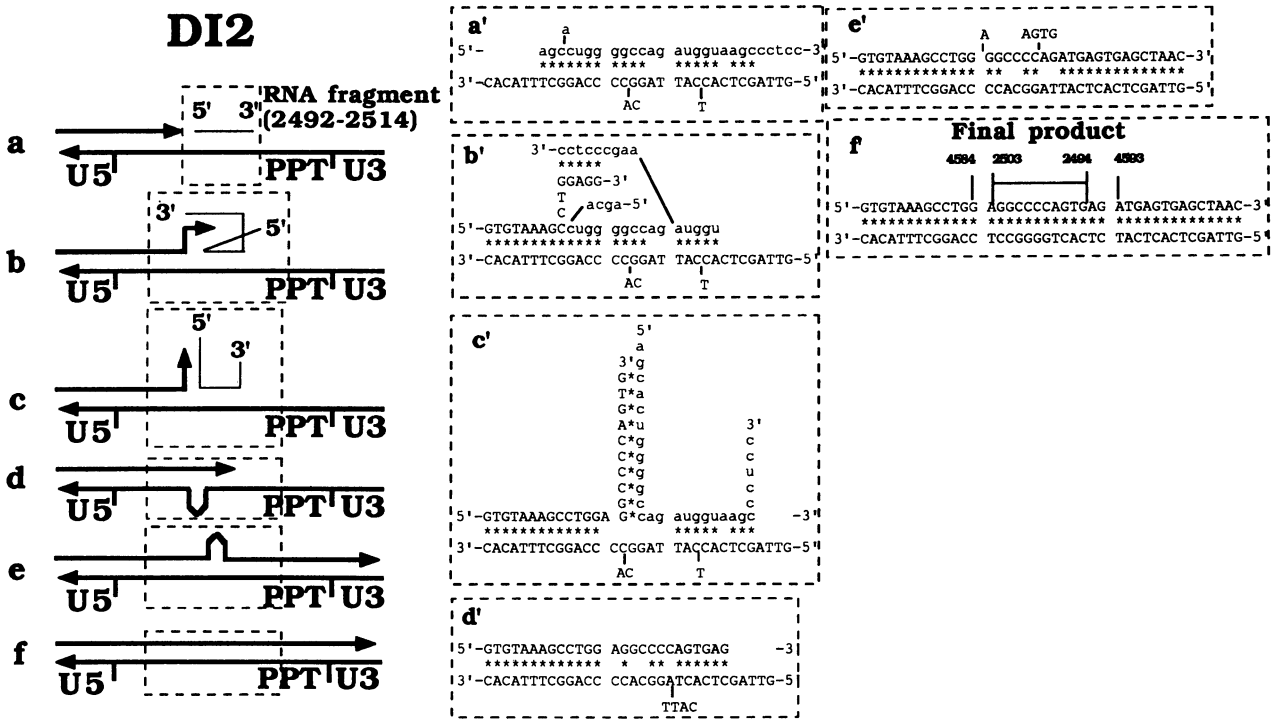

FIG. 5. Model for generation of deletion with insertion DI2. (*a*) An RNA fragment (2492–2514) resulting from RNase H digestion of RNA·DNA hybrids (see Fig. 3*b*) hybridizes to minus-strand DNA 5′ and 3′ of the deletion junctions (*a* and *a*′). Nineteen of 23 bases can form base pairs. (*b*) The plus-strand DNA growing point progresses to the first deletion junction. The growing point of the nascent DNA strand switches templates to the 3′ end of the RNA fragment, and plus-strand DNA is extended by 3 nucleotides (*b* and *b*′). (*c*) Misalignment of the nascent DNA strand growing point on the viral RNA template occurs, and plus-strand DNA is extended by an additional 8 nucleotides (*c* and *c*′). (*d*) The nascent DNA strand growing point switches templates to the minus-strand DNA, and plus-strand DNA is extended by an additional 2 nucleotides (*d* and *d*′). (*e*) Misalignment of the nascent DNA strand growing point on the minus-strand DNA template occurs; DNA synthesis continues past the second deletion junction (*e* and *e*′). (*f*) Plus-strand DNA synthesis is completed to generate the final product (*f* and *f*′).

must be disrupted for this misalignment to occur. No disruption of hydrogen bonds between the direct repeats is required for misalignments involving deletions. This model, and the lack of duplications identified in our collection of mutants, suggests that large duplications will rarely occur; however, short duplications would involve denaturation of short sequences and may be detectable in a larger data set. For example, denaturation of 8 bp is expected in the generation of D5.

Deletions with Insertions. The deletions with insertions we encountered seemed to occur during both plus- and minus-strand DNA synthesis. In DI2, the final product contained 5 bp more than the wild type. This process may represent a mechanism by which the virus can generate additional sequences lost by the high rate of deletions. The additional sequences can become a source of genetic raw material upon which mutation and natural selection can act to generate variation in retroviral populations.

Model for Generation of DI1. The steps involved in the polymerization of a double-stranded DNA molecule from a viral genomic RNA are outlined in Fig. 3 (13).In the process of reverse transcription, the PPT is required for initiation of plus-strand DNA synthesis (Fig. 3 *d* and *e*). The absence of PPT in the recovered DI1 provirus indicated that the deletion could not have occurred before initiation of plus-strand DNA synthesis and established that at least some DIs were formed during plus-strand DNA synthesis (Fig. 4).

The plus strand of the 14-nucleotide inserted sequence found at the deletion junction was compared with the VP232 plus-strand and minus-strand nucleotide sequences. Nucleotides complementary to the 14 inserted nucleotides were identified in the plus-strand R region, just 1 bp downstream of the 5′ end of R. The presence of the complementary sequence in the DI1 plus-strand nucleotide sequence indicated that, during plus-strand DNA synthesis, misalignment and DNA synthesis templated by the viral RNA occurred

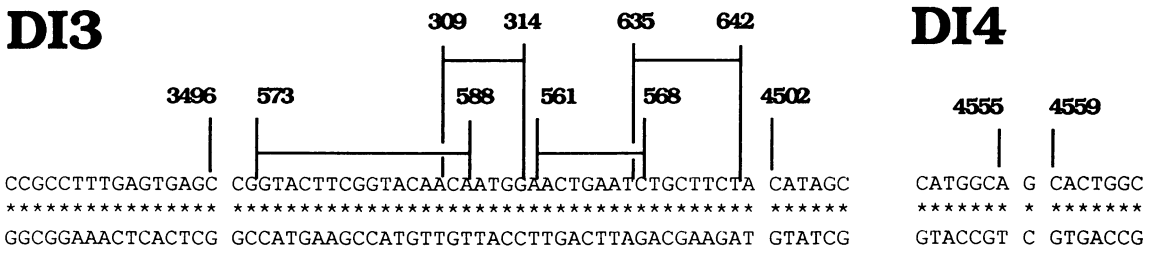

FIG. 6. Final products of deletions with insertions DI3 and DI4. The DNA sequence of the deletion junctions of DI3 and the inserted sequence are shown. Vertical lines with sequence numbers indicate positions of homologous sequences within the VP212 plus-strand nucleotide sequence. Sequence positions 573–588 are homologous to nucleotides on the plus strand adjacent to the U5–PBS junction. The DNA sequence of the deletion junctions of DI4 and the inserted nucleotide are shown.

(Fig. 4 *c* and *c'*). Since the DNA synthesis templated by the plus-strand sequences terminated just 1 nucleotide from the 5' end of the R region, which exists as a free 5' end only in the viral RNA, it is likely that viral RNA was utilized as a template.

Further analysis of the deletion junction sequence suggested that a second misalignment followed by DNA synthesis templated by the minus-strand DNA occurred (Fig. 4 *d* and *d'*). Hybridization of the newly synthesized DNA sequence complementary to the viral RNA to the DNA sequences of the minus strand just upstream of the second deletion junction is postulated to occur as shown in Fig. 4*d'*. The two sequences are highly homologous; 11 of the 14 nucleotides can form base pairs. DNA synthesis directed by the minus-strand DNA would then reinitiate and continue to complete plus-strand DNA synthesis. The 630-bp deleted sequence (shown as a stem–loop in Fig. 4*d*) may be removed by the host cell DNA repair mechanisms. Alternatively, integration of viral DNA followed by DNA replication and segregation could occur to generate the final proviral DNA sequence (Fig. 4 *e* and *e'*).

Model for Generation of DI2. The presence of the 10-nucleotide plus-strand sequence complementary to the sequence in the VP212 viral RNA indicated that DI2 was also formed during plus-strand DNA synthesis using a viral RNA template (Fig. 5). We postulate that hybridization to the minus-strand DNA of an RNase H-generated RNA fragment containing the sequence from 2492 to 2514 occurred (Fig. 5 *a* and *a'*). Nineteen of 23 nucleotides are identical. Next, misalignment of the growing point of plus-strand DNA and the viral RNA occurred, and the sequence near the 3' end of the viral RNA fragment was utilized as a template (Fig. 5 *b* and *b'*). A second misalignment of the growing point and sequences near the middle of the viral RNA occurred, and the viral RNA was further utilized as a template (Fig. 5 *c* and *c'*). A third misalignment of the growing point and minus-strand DNA occurred (Fig. 5 *d* and *d'*), and the plus-strand DNA was extended by two additional base pairs. Finally, a fourth misalignment of the growing point and the minus-strand DNA occurred (Fig. 5 *e* and *e'*), and DNA synthesis directed by the minus-strand DNA reinitiated to complete plus-strand DNA synthesis and generate the final proviral DNA sequence (Fig. 5 *f* and *f'*).

DI3 and DI4. The orientation of the inserted sequences in DI3 sequences indicated that this deletion with insertion occurred during minus-strand DNA synthesis (Fig. 6). The number of homologous sequences that may be involved in the generation of DI3 indicated that several misalignments and other potentially complex events were involved; therefore, a detailed model for the generation of DI3 is not presented.

We postulate that DI4 was formed as a result of a transient misalignment involving the CA direct repeats during minus-strand synthesis, followed by extension of the growing point by 1 nucleotide (Fig. 6). A realignment of the growing point followed by reinitiation of DNA synthesis directed by the plus-strand viral RNA then could have occurred to generate this deletion.

Conclusion. The results presented here indicate that deletions in retroviral genomes can occur by at least two mechanisms. Deletions that do not fit into the classes described here may also arise at a lower frequency. These results confirm our previous conclusion that reverse transcriptase is an enzyme of low processivity (6). The apparent low affinity of the enzyme for the template is perhaps an essential property for an enzyme that participates in the strand transfer events during viral replication.

We thank J. Couch, B. Fritz, K. Krebsbach, P. Olson, J. Rein, and R. Wisniewski for their technical assistance. We also thank Drs. N. Drinkwater, A. Panganiban, and B. Sudgen for critical reading of the manuscript and Drs. M. Hannink, W.-S. Hu, and G. A. Pulsinelli for helpful discussions and comments on the manuscript. This work is supported by Grants CA-22443 and CA-07175 from the National Cancer Institute. V.K.P. was supported by Training Grant CA-09075. H.M.T. is an American Cancer Society Research Professor.

1. Huang, A. S. & Baltimore, D. (1970) *Nature (London)* **226,** 325–327.
2. Swanstrom, R., Parker, R. C., Varmus, H. E. & Bishop, J. M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2519–2523.
3. Voynow, S. L. & Coffin, J. M. (1985) *J. Virol.* **55,** 67–78.
4. Shimotohno, K. & Temin, H. M. (1982) *J. Virol.* **41,** 163–171.
5. Dougherty, J. P. & Temin, H. M. (1986) *Mol. Cell. Biol.* **6,** 4387–4395.
6. Pathak, V. K. & Temin, H. M. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6019–6023.
7. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33,** 103–119.
8. Stanssens, P., Opsomer, C., McKeown, Y. M., Kramer, W., Zabeau, M. & Fritz, H.-J. (1989) *Nucleic Acids Res.* **17,** 4441–4454.
9. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
10. Grosovsky, A. J., De Boer, J. G., De Jong, P. J., Drobetsky, E. A. & Glickman, B. W. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 185–188.
11. De Jong, P. J., Grosovsky, A. J. & Glickman, B. J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 3499–3503.
12. Pastink, A., Vreeken, C., Schalet, A. P. & Eeken, J. C. J. (1988) *Mutat. Res.* **207,** 23–28.
13. Varmus, H. (1988) *Science* **240,** 1427–1435.