JOURNAL OF VIROLOGY, Feb. 1985, p. 447–455
0022-538X/85/020447-09$02.00/0

Vol. 53, No. 2

# Reverse Transcription of Retroviral Genomes: Mutations in the Terminal Repeat Sequences

LESLIE I. LOBEL AND STEPHEN P. GOFF*

*Department of Biochemistry and Institute for Cancer Research, Columbia University, College of Physicians and Surgeons, New York, New York 10032*

Received 2 August 1984/Accepted 22 October 1984

**The process of reverse transcription of retroviral genomes begins with the synthesis of a short DNA molecule near the 5′ end of the RNA template. This molecule, termed minus-strand strong-stop DNA, is then translocated to the 3′ end of the viral RNA by means of a repeated sequence, the R region, present at both ends of the template. The translocation should result in the transfer of genetic information from the 5′ R region to the 3′ R region. We have generated a series of mutants of Moloney murine leukemia virus with alterations in the R regions by in vitro mutagenesis of a cloned DNA copy of the viral genome. The altered DNAs were introduced into mouse cells by transfection, and the translocation of the mutations during viral replication was assessed. Some mutations were not transferred from the 5′ R region to the 3′ R region; these results were not in accord with current models for reverse transcription. The results can be explained if DNA molecules shorter than strong-stop DNA, formed by premature termination of synthesis, are sometimes translocated. A number of mutants with large deletions in the R region were tested and were able to replicate with normal strong-stop DNA translocation. Thus, short stretches of homology can be used by the virus to carry out strong-stop translocations.**

The retroviruses carry out an unusual reaction soon after infection: a double-stranded DNA copy of the viral genome is synthesized by reverse transcription of the virion RNA. The reaction is known to be complex (Fig. 1). Synthesis begins with the extension of a tRNA primer molecule that is base paired to the template RNA near its 5′ end (6) and quickly stops when the 5′ end of the template is copied. The product of this reaction, known as minus-strand strong-stop DNA, contains some sequences that are unique in the viral RNA, but also includes a repeated sequence, the R region, which is present at both the 5′ and 3′ termini of the viral RNA (3). Subsequent extension of this short DNA requires a extraordinary step: the translocation or "jump" of the strong-stop DNA to the 3′ end of the template. The mechanism of the translocation is not well understood, but presumably involves base pairing of the strong-stop DNA to the R region at the 3′ acceptor end of the template. After the jump has occurred, DNA synthesis resumes and results in the formation of a nearly full-length minus-strand DNA.

The completion of the process of reverse transcription requires several more steps. The viral RNA is degraded by the RNase H activity of the reverse transcriptase enzyme. Next, synthesis of the second DNA strand is begun with a plus-strand strong-stop DNA (15). This molecule is initiated at a polypurine stretch in unique viral sequences at the edge of the U3 region at the 3′ end of the genome; DNA synthesis stops after copying part of the tRNA at the 5′ end of the minus-strand strong-stop DNA. A second jump occurs, in which the plus-strand strong-stop DNA is paired with sequences at the 3′ end of the minus strand to allow its extension. After extension of both strands, the product is complete: a fully duplex linear DNA containing a duplication of sequences called long terminal repeats (LTRs). The linear DNA is ultimately transported to the nucleus and circularized, and one or more of the viral DNAs are integrated into the host chromosome.

* Corresponding author.

Analysis of the structures of the intermediate species has confirmed many features of this model (4). The reaction as a whole can be studied because it can be carried out in vitro with detergent-permeabilized virions; it is much more difficult to study the reaction in vivo, and very few of the steps have been examined in infected cells. Earlier work has shown (26) that during intergenic recombination, progeny virus always retain identical termini derived from one or the other of the parental genomes; this result suggests that termini are quickly made identical as would be predicted by the model. An important prediction of the current model for reverse transcription of retroviral genomes is the transfer of genetic information from one end of the genome to the other (Fig. 1). A mutation in the 5′ R region should be copied during synthesis of the minus-strand strong-stop DNA and carried to the 3′ end of the template. The mutation should be maintained at the 3′ end and transferred back to the 5′ end of the genome by the translocation of the plus-strand strong-stop DNA; the final product should carry the mutation at both ends. Thus, the mutation should replace the wild-type sequence at the 3′ end and should thereafter be retained by the virus in both R regions.

To test the genetic predictions of the model for reverse transcription, we have constructed a series of mutations in the R region of a cloned DNA copy of the genome of Moloney murine leukemia virus (M-MuLV). We have then introduced the altered DNAs into permissive cells in culture and allowed replication of the virus to occur. The presence of the mutation at the 5′ and 3′ ends of the viral genome was assessed by restriction enzyme digestion of the viral DNA. Surprisingly, mutations were not always transferred according to the predictions of the model. We suggest that DNA molecules shorter than the strong stop DNAs may be translocated during strong-stop jumps.

## MATERIALS AND METHODS

**Cells and viruses.** NIH/3T3 and XC cells were grown in Dulbecco modified Eagle medium containing 10% calf se-

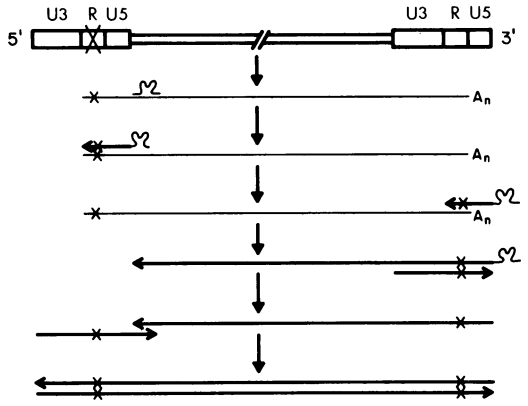



FIG. 1. Replication of the retroviral genome and predicted behavior of mutations in the 5' R region. RNA is indicated by thin lines, and DNA is indicated by thick lines. The top line depicts a proviral DNA bearing a mutation in the 5' R region (×). This mutation would be copied to RNA during forward transcription (line 2) and to DNA during formation of strong-stop DNA (line 3). It would be transferred to the 3' end of the template via a strong-stop DNA jump (line 4) and transferred back to the 5' end via the second strong-stop jump (lines 5 and 6). The final proviral DNA (line 7) would contain the mutation in both LTRs.

rum. Introduction of viral DNAs into cells was carried out as described previously (13). Briefly, DNA in phosphate-buffered saline was mixed with DEAE Dextran (Pharmacia Fine Chemicals) to a final concentration of 500 μg/ml, and the mixture was applied to cells previously washed twice with phosphate-buffered saline. After 10 to 60 min, the cells were washed twice again and fed with medium. When confluent (3 to 6 days), the cells were transferred at a 1:10 dilution and allowed to grow to confluence (3 to 4 days). The monolayers were overlaid with XC cells soon after reaching confluence (18), and plaques were scored after staining with hematoxylin. The assay effectively measures the number of cells producing virus at the time of the 1:10 dilution; this number is highly reproducible for a given DNA, being 500 to 1,000 PFU per 100 ng for wild-type DNA.

**Enzyme reactions and assays.** Restriction enzymes were purchased from New England BioLabs and used according to their specifications. Digestion with BAL 31 and ligation of DNAs were carried out as described previously (19). Reverse transcriptase assays were carried out by the rapid-dot method (5); harvests of transfected cells were taken after the first passage, immediately before XC overlay. Nick translations were as described previously (16). We thank J. Van Oostrum for the generous gift of T4 DNA ligase.

**Construction of mutants.** The 12-base-pair (bp) insertion mutations in the 5' or 3' LTRs were generated as described previously (10); insertions were made into the *Sma*I sites of pT11, a plasmid containing a cloned provirus, to form mutants *in*479 and *in*8743. Viral DNA containing the mutations in both LTRs (*in*479/8743) was formed by transferring the 2.5-kilobase (kb) *Sac*I fragment from mutant *in*479 into the equivalent fragment of the permuted viral clone pa8.2 (10). The viral insert was then excised with *Hin*dIII and oligomerized by ligation at high concentration (19) to form a provirus with two mutated LTRs.

Mutant *dl*8743 was constructed by joining two fragments: a *Sal*I-to-*Eco*RI fragment of *in*8743 and an *Eco*RI-to-*Sal*I fragment of *in*9050 (a derivative of pT11 bearing a unique *Eco*RI site in the DNA flanking the proviral insert).

Mutations in the 5' half of the R region were generated by a complex procedure (Fig. 2). First, a small fragment of p8.2 containing the LTR was subcloned into pBR322 to form the plasmid pCLANDE. This plasmid will be described in detail elsewhere (J. Colicelli and S. P. Goff, manuscript in preparation). Two fragments were isolated from pCLANDE: a

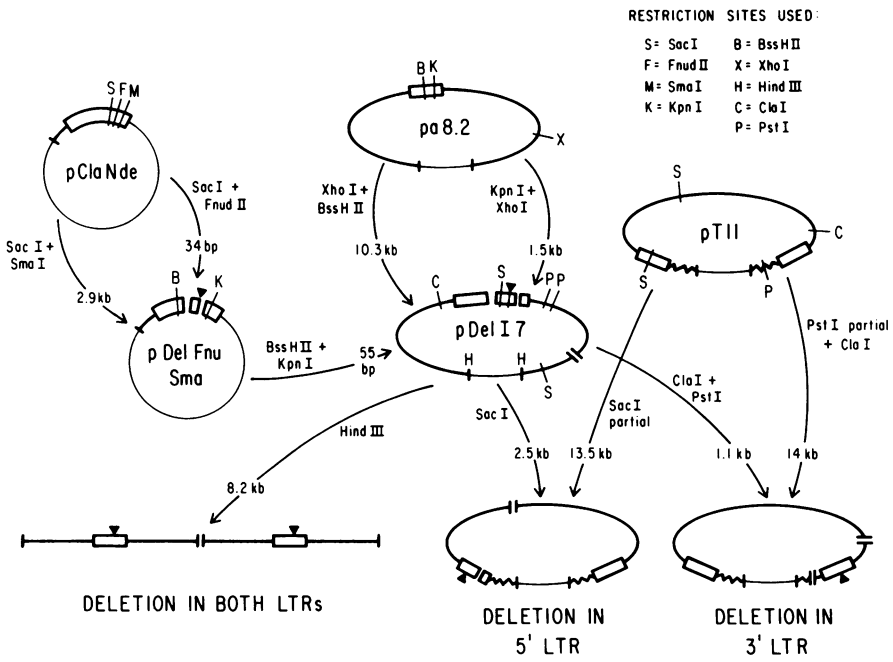



FIG. 2. Construction of a deletion mutation removing the 5' half of the R region and movement of the mutation into various LTRs of the provirus. See the text for a detailed description of the manipulations. The final constructs (bottom line) consist of proviruses bearing mutations in both LTRs (mutant *dl*450/8714), in the 5' LTR (mutant *dl*450), and in the 3' LTR (mutant *dl*8714).

2.9-kb fragment generated by cleavage with *Sac*I plus *Sma*I and a 34-bp fragment generated by cleavage with *Sac*I plus *Fnu*dII. These two were ligated, and plasmid pDelFnuSma, which contained the desired deletion of the 5′ half of the R region, was isolated after bacterial transformation. A complete viral genome containing the deletion in this plasmid was generated by the ligation of three DNAs: a 55-bp DNA produced by cleavage of pDelFnuSma with *Bss*HII plus *Kpn*I, a 10.3-kb fragment produced by cleavage of pa8.2 with *Bss*HII plus *Xho*I, and a 1.5-kb fragment produced by cleavage of pa8.2 with *Xho*I plus *Kpn*I. Oligomerization of the viral insert in this plasmid, pDelI7, yielded proviral DNA with the mutation in both LTRs. Proviral DNA carrying the mutation at the 5′ LTR only was generated by joining two DNAs: the 13.5-kb *Sac*I fragment of pT11 and the 2.5-kb *Sac*I fragment of pDelI7. Proviral DNA carrying the mutation at the 3′ LTR only was also generated by joining two DNAs: the 1.1-kb fragment produced by cleavage of pDelI7 with *Cla*I plus *Pst*I and the 14.0-kb fragment produced by treatment of pT11 DNA so as to give partial cleavage with *Pst*I and complete cleavage with *Cla*I.

Mutant *dl*439 was generated by cleavage of viral DNA with *Sma*I, digestion of the linear DNAs with BAL 31, cyclization with DNA ligase, and introduction into *Escherichia coli*. We are indebted to P. Schwartzberg and N. Tanese for preparing and characterizing this mutant DNA.

**Analysis of unintegrated viral DNAs.** NIH/3T3 cells were infected at high multiplicity (1 to 10 PFU per cell) in the presence of polybrene (8 μg/ml), and low-molecular-weight DNA was isolated 24 h later by the Hirt (7) procedure. The DNA was purified by repeated phenol and chloroform extractions, treated with RNase, and collected by ethanol precipitation. The DNAs were cleaved, fragments were separated by agarose gel electrophoresis and blotted to nitrocellulose (21), and the viral DNAs were detected by hybridization (24) to radioactive viral DNA (clone p8.2).

**Manipulation of DNAs.** Plasmid DNAs were prepared from small cultures by a rapid method (8) and from large cultures by cesium chloride-ethidium bromide centrifugation (9). DNAs were eluted from agarose gels by the glass powder method (23). The detailed structure of the mutations was confirmed by direct determination of DNA sequence by the procedure of Maxam and Gilbert (12). DNAs were introduced into bacteria by the calcium chloride method (11). Analysis of strong stop DNA was as described previously (6), with minor modifications.

## RESULTS

**Insertion mutations in the R region.** For our initial experiments, two mutant viral genomes were constructed, one with an alteration in the 5′ R region and one with an alteration in the 3′ R region. Plasmid pT11, which contains a cloned copy of a wild-type M-MuLV provirus, was mutagenized by the "suppressor-linker" method (10). The two mutations consisted of insertions of 12 bp at the *Sma*I site in the R region of either the 5′ or the 3′ LTR (Fig. 3). The added bases caused the loss of the *Sma*I site normally present in the wild-type virus and the insertion of a recognition site for cleavage by *Eco*RI (Fig. 5). The genome of wild-type M-MuLV contains no *Eco*RI sites; thus, the presence of these mutations could be readily detected by analysis of the products of *Eco*RI digestion of the viral DNAs.

**Analysis of viral structure after replication.** Cellular genes can be introduced into mammalian cells as a mixture of DNA and DEAE Dextran (13); the gene is usually expressed only transiently and is not maintained by the cell for many generations. The transfection of mammalian cells with the proviral DNA of a replication-competent retrovirus also allows the transient expression of the genome; in this case, however, the virus spreads by infection of neighboring cells and is thus maintained. The efficiency of the process can be estimated by measuring the release of virion-associated reverse transcriptase or by counting plaques after overlay of the cells with the XC plaque indicator cells. Eventually all of the cells in a recipient culture become stable producers of virus.

To analyze the virus produced by the mutated DNAs, each DNA was applied to NIH/3T3 cells with DEAE Dextran. Both of the mutants were replication competent, giving rise to 500 to 1,000 XC plaques per 100 ng of applied DNA (Fig. 3). The supernatant medium from the cultures revealed wild-type levels of reverse transcriptase activity. Thus, the presence of the alterations in the R region did not reduce the ability of the DNAs to initiate virus production.

To determine whether the mutants had retained the *Eco*RI site during replication, these supernatants were filtered and then used without dilution to infect fresh NIH/3T3 cells. The low-molecular-weight DNA was purified and cleaved with *Eco*RI, and the resulting fragments were separated by agarose gel electrophoresis. Viral DNA fragments were detected by transfer to nitrocellulose, hybridization with a radioactive viral probe, and autoradiography (Fig. 4). The results of the experiment were surprising. Virus recovered from the mutant altered in the 5′ LTR synthesized normal amounts of all viral DNA species (lane 3), but these DNAs did not contain an *Eco*RI site (lane 4). Thus, the mutation in the 5′ LTR was not successfully transferred to the 3′ end as expected. Instead, the sequences originally present at the 3′ end had been retained and were now present at both ends. This was demonstrated by testing for the reappearance of the *Sma*I site that was originally disrupted in the mutagenesis. Cleavage of these DNAs with *Sma*I (lanes 8 and 9) showed that the *Sma*I site was restored, and that the wild-type sequence from the 3′ end was the source of this region of the viral DNA. We estimate that at least 95% of the DNA made was wild type in structure. The virus recovered from the mutant altered in the 3′ LTR was also wild type in structure (lane 5); thus, as expected in this case, the wild-type sequences were transferred from the 5′ end to the 3′ end and were retained. In agreement with earlier studies (10), insertions in other regions of the genome were stably retained under these conditions (lanes 6 and 7).

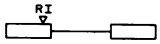
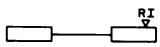
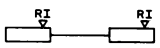
| MUTANT | STRUCTURE | FATE OF MUTATION | RT | XC |
|---|---|---|---|---|
| in 479 |  | Lost | + | ~$10^3$,$10^3$,$10^3$ |
| in 8743 |  | Lost | + | ~$10^3$,$10^3$,$10^3$ |
| in 479/8743 |  | Maintained | + | ~500,500,500 |

FIG. 3. Structure and properties of insertion mutants of M-MuLV. The positions of the linker insertion mutations are indicated by the open triangles. RT indicates the appearance of virion-associated reverse transcriptase in the medium one passage after transfection, as measured by enzymatic assay (5). XC indicates the number of XC syncytial plaques detected after DEAE dextran transfection of 100 ng of the indicated DNA into NIH/3T3 cells. Multiple entries indicate results of duplicate assays.

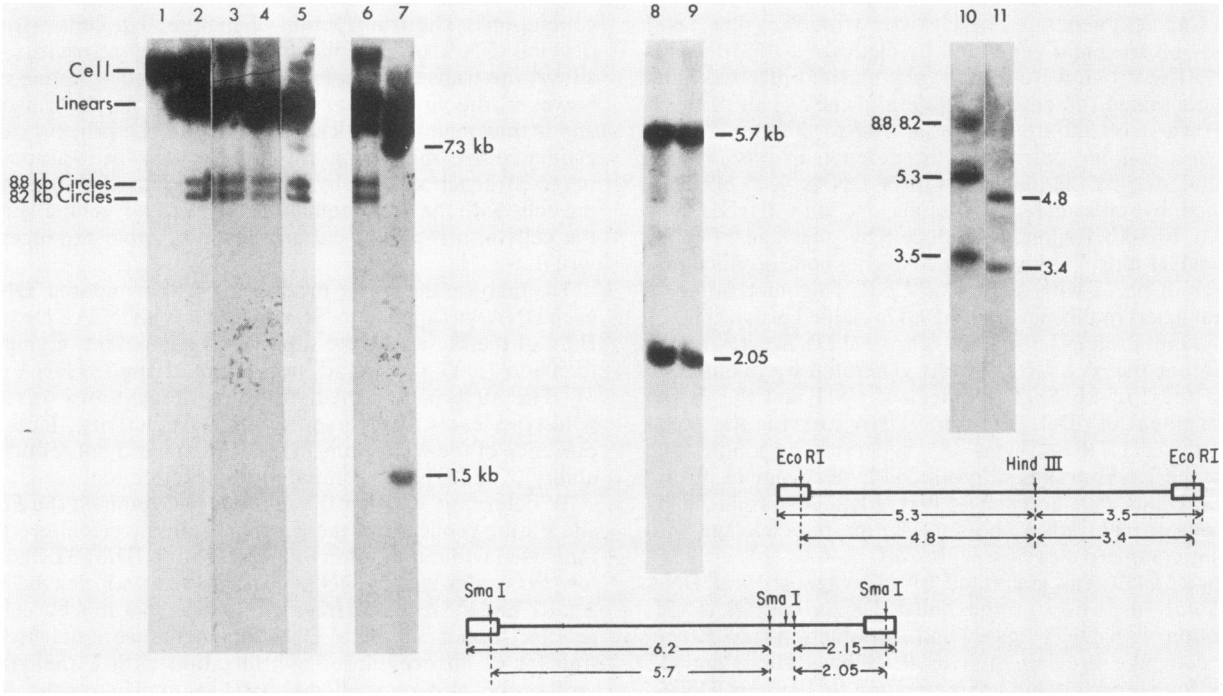



FIG. 4. Southern blot of low-molecular-weight DNAs synthesized by mutant viruses. NIH/3T3 cells were infected with various virus preparations, and the low-molecular-weight DNA was isolated by the Hirt (7) procedure 24 h later. The DNA was displayed by agarose gel electrophoresis, transferred to nitrocellulose, and hybridized to a radioactive viral DNA probe. Lanes 1 and 2 are controls. Lane 1 shows uninfected cell DNA. The host DNA, migrating near the top of the gel, is labeled because of the endogenous viral sequences with homology to the genome of M-MuLV. In lane 2 wild-type virus directs the synthesis of full-length linear and two circular DNAs. The positions of the viral species are indicated. Lanes 3, 4, and 5 show loss of the *Eco*RI site originally present at either the 5′ or 3′ LTR. In lane 3 the cells were infected with virus from mutant *in*479. The DNAs were not treated with a restriction enzyme. Lane 4 shows mutant *in*479 DNAs, after treatment with *Eco*RI. All of the DNA forms were resistant to cleavage. Lane 5 shows mutant *in*8743 DNAs, after treatment with *Eco*RI; these DNAs were also resistant. Lane 6 shows previously described (10) linker insertion mutant, *in*1525-12, bearing an insertion in the *gag* gene, used as a control. Lane 7 shows *In*1525-12 DNAs after cleavage with *Eco*RI. The two circles were completely cut, and the linear DNAs were cut to yield fragments 7.3 and 1.5 kb in size. Lanes 8 and 9 show restoration of the *Sma*I site occurring concomitantly with loss of the *Eco*RI site from mutant *in*479 DNAs. The appended restriction map shows the relative positions of the *Sma*I sites in wild-type and mutant DNAs. Lane 8 shows wild-type DNAs after *Sma*I cleavage. Lane 9 shows mutant *in*479 DNAs after *Sma*I cleavage. The fragments produced, 2.05 and 5.7 kb in length, were as predicted for wild-type virus. Loss of the terminal *Sma*I sites would have resulted in the appearance of a fragment 2.15 kb in length (see Fig. 7). Lanes 10 and 11 show retention of the *Eco*RI site in a mutant with inserts in both LTRs. The appended restriction map shows the relative positions of the *Eco*RI and *Hin*dIII sites in wild-type and mutant DNAs. In lane 10 wild-type virus was used and cleaved with *Eco*RI plus *Hin*dIII; the fragments, 5.3 and 3.5 kb in size, are as expected for virus without *Eco*RI sites. In lane 11 mutant *in*479/8743 was used and cleaved as in lane 10. The fragments produced, 4.8 and 3.4 kb in size, were as predicted for a viral genome retaining the *Eco*RI sites.

One possible explanation for the loss of the 5′ mutation was that any virus that carried the *Eco*RI site at both LTRs was defective and could not be propagated. Thus, there would be selection for an unusual event that would repair the mutation. To test this possibility, a viral DNA termed *in*479/8743 was constructed which carried the 12-bp insertion in both 5′ and 3′ R regions (Fig. 3 and 5). This DNA was applied to sensitive cells as before and analyzed by the same procedures. The DNA was infectious and gave rise to both XC plaques and high levels of reverse transcriptase. The supernatant virus was harvested and used to infect fresh cells, and the viral DNA was analyzed by blot hybridization (Fig. 4). This virus stably retained the *Eco*RI site; it could be shown that the site was retained in both LTRs by digestion of the DNA with *Hin*dIII plus *Eco*RI (lanes 10 and 11). Thus, the insertion itself was not lethal to the replication of the virus and could easily be detected by the procedures used. If viral DNA with the insertion in the 5′ LTR alone had efficiently transferred the mutation to the 3′ end, the resulting virus would have been able to propagate and would have been detected. We conclude that the 5′ mutation in mutant *in*479 was not faithfully carried to the 3′ end as predicted by the current model for reverse transcription. In contrast, a wild-type 5′ end was carried to the 3′ end in mutant *in*8743 as predicted by the model.

**Construction of deletion mutations in the R region.** One explanation for the failure of the mutation in the 5′ sequences to translocate is that incomplete DNAs might sometimes jump prematurely to the 3′ end, using only a portion of the homology in the R region. If the translocation occurred before reverse transcription of the mutation, no transfer of the mutation would be expected, and the resulting wild-type virus could spread throughout the culture. To test whether short regions of homology could be used for transfers, deletion mutations were constructed in the R region that provided a shorter region of homology for RNA-DNA annealing during the strong-stop translocation.

First, a proviral DNA was generated that lacked the 3′ half of the 3′ R region (Fig. 5). This mutant was constructed by the ligation of two DNA fragments from two insertion mutants (see above for details). The resulting proviral DNA was intact, but for the deletion of the sequences from the

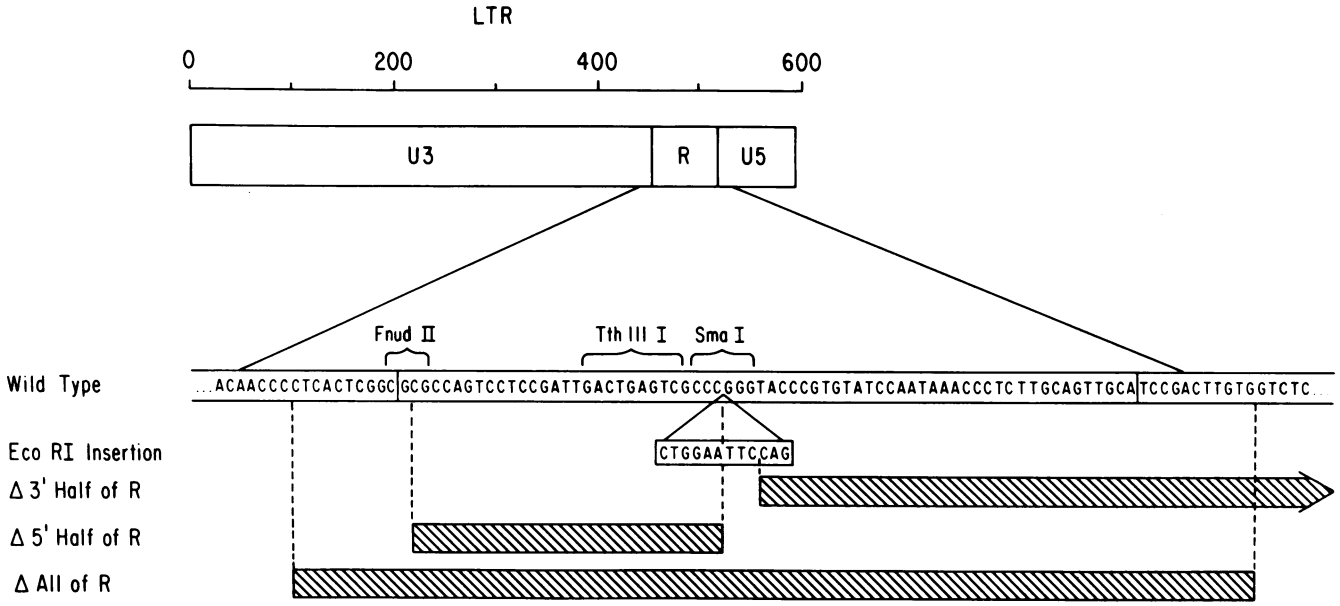



FIG. 5. Structure of the mutations generated in the R region of the LTR. The top shows the U3, R, and U5 regions of the viral LTR as boxed regions; the scale indicates length in base pairs. The bottom shows a blow-up of the R region. The DNA sequence of the wild-type virus is indicated (from reference 20, with one base change at position 449); the positions of the recognition sites for cleavage by *Fnu*DII, *Tth*111I, and *Sma*I are shown. The 12 bases inserted at the *Sma*I site in mutants *in*479, *in*8743, and *in*479/8743 are shown. The mutation removing the 3' half of the R region is shown next; this deletion extends from the inserted *Eco*RI site to a distant site downstream. The mutation removing the 5' half of the region is indicated; this deletion extends from the *Fnu*DII site to the *Sma*I site. Finally, the large deletion of the entire R region is shown.

*Sma*I site in the center of the R region to a site in the downstream flanking DNA.

Second, a clone of the viral LTR was constructed which lacked the 5' half of the R region, from an *Fnu*DII site at the point of initiation of transcription to the *Sma*I site in the center of the R region (see above and Fig. 2). A fragment containing this altered LTR was then introduced into the 5' LTR of a provirus, into the 3' LTR, or into both LTRs (see above for details of the construction). The resulting proviruses were intact except for the 29-bp deletion lying entirely within the R region. It should be noted that the deletion began exactly after the first base of the viral RNA and thus did not extend into the 5'-proximal sequences required for transcription (Fig. 5). The construction of the deletion depended on the appearance of an *Fnu*DII site at this position; the site is not present in the published sequence of M-MuLV (20), but is present in many clones of M-MuLV by virtue of a 1-bp difference at position 449 (data not shown).

**Analysis of deletion mutants.** Each of the DNAs described above was applied to NIH/3T3 cells as before and analyzed by XC overlay and reverse transcriptase assays (Fig. 6). Mutant *dl*8743, containing a 3' deletion in the 3' LTR, gave

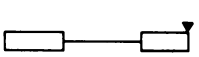
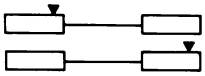
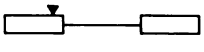
| MUTANT | STRUCTURE | POSITION OF DELETION | FATE OF MUTATION | RT | XC |
|---|---|---|---|---|---|
| dl 8743 |  | 3' Half of R (8744-8858) | Lost | + | ~$10^3$,$10^3$,$10^3$ |
| dl 450 |  | 5' Half of R (451-479) | Lost | + | ~$10^3$,$10^3$,$10^3$ |
| dl 8714 | | 5' Half of R (8715-8743) | Lost | + | ~$10^3$,$10^3$,$10^3$ |
| dl 450/8714 | | 5' Halves of Rs (451-479 and 8715-8743) | Maintained | + | 0,0,0 (~$10^3$,$10^3$,$10^3$) |
| dl 439 |  | All of R (440-528) | | +/− | 0,10,5,0,0,20 |

FIG. 6. Structure and properties of deletion mutants. The position of the deletions in each mutant is indicated by the closed triangles. RT indicates the appearance of virion-associated reverse transcriptase in the medium one passage after transfection: +, wild-type levels; +/−, very low levels (roughly 100-fold reduced), which increased with time (2 to 4 weeks) to finally reach wild-type levels. XC indicates the number of XC syncytial plaques detected after DEAE Dextran transfection of the indicated DNAs into NIH/3T3 cells. Multiple entries represent results of duplicate transfections. Entries within parentheses indicate results after one additional passage of the recipient NIH/3T3 cells.

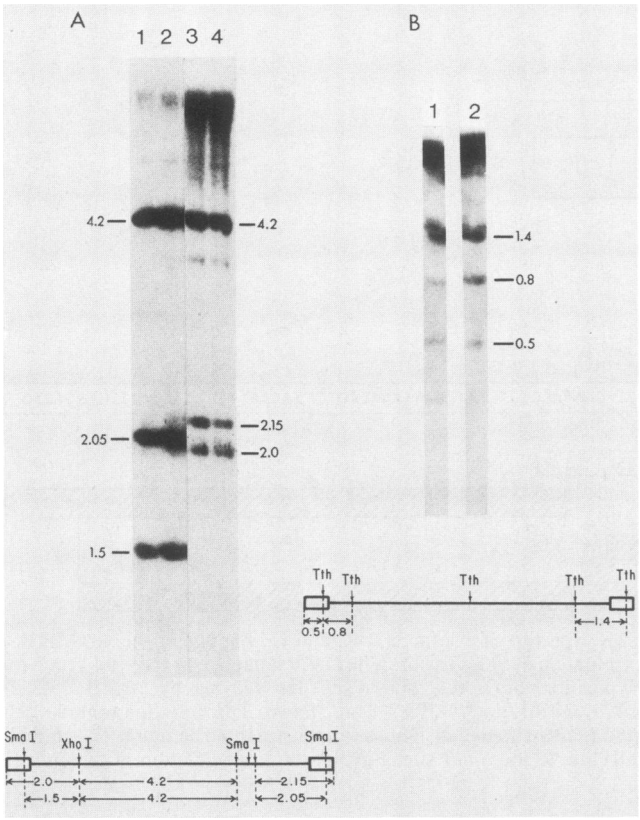



FIG. 7. Southern blot of low-molecular-weight DNAs synthesized by mutant viruses. (A) DNAs were digested with *Xho*I plus *Sma*I. The appended restriction map gives the expected fragment sizes for wild-type DNA (containing the *Sma*I sites in the LTRs) and for deletion mutant DNA (lacking these sites). Lane 1 shows wild-type virus. The three prominent bands indicated are of the sizes expected for viral DNA that contains the *Sma*I site. Lane 2 shows the *dl*450 virus. This DNA has reacquired the *Sma*I sites and is indistinguishable from wild-type DNA. Lanes 3 and 4 show two independent isolates of *dl*450/8714 virus. These DNAs show the fragments predicted for viral DNA lacking the terminal *Sma*I sites. (B) DNAs were digested with *Tth*111I. The map shows the small fragments predicted after cleavage with this enzyme. Lane 1 shows wild-type virus, and lane 2 shows *dl*450 virus. This DNA has also reacquired the *Tth*111I sites in the LTRs.

rise to wild-type numbers of plaques and normal levels of reverse transcriptase. Analysis of the viral DNA synthesized by the recovered virus showed a wild-type restriction pattern (data not shown); thus, the 5′ R region had been moved during the jump to the 3′ end and then was retained. These results show that limited homology (the 5′ half of R) is sufficient to allow strong stop transfers.

Mutant *dl*450, bearing a 5′ deletion in the 5′ LTR, also gave wild-type numbers of plaques and levels of reverse transcriptase activity. This mutant could only synthesize a prematurely terminated strong-stop DNA; the final half of the R region had been removed by the deletion. Thus, analysis of this mutant also suggests that shorter regions of homology can be used to jump and, in addition, shows that short DNAs can jump to full-size acceptors. DNA synthesized by this virus was isolated and cleaved with *Sma*I plus *Xho*I and with *Tth*111I. These digests showed that the virus had recovered a wild-type R region (Fig. 7). Thus, the wild type sequences present at the 3′ LTR were used as templates after the jump. These sequences would be in the correct position to act as templates after the transfer of the shortened strong-stop DNA.

The mutant with the same deletion at the 3′ LTR, *dl*8714, also yielded wild-type numbers of plaques and enzyme activity; digests showed that it also retained the *Sma*I site (data not shown). In this case, the region of homology is at the very 3′ end of the RNA; the wild-type strong-stop DNA was apparently extended across the deletion to restore at least part of the deleted region. To carry out this jump and remove the deletion, a full-length strong-stop DNA must be used.

Mutant *dl*450/8714, bearing the same deletion at both LTRs, did not give rise to XC plaques when tested shortly after transfection, but was able to replicate in cells as judged by the presence of high levels of reverse transcriptase (Fig. 6). After two passages, the cells did become XC positive. Analysis of the DNA synthesized by this virus showed that the mutation had been retained; cleavage with *Sma*I plus *Xho*I showed that there was no *Sma*I site in the LTRs (Fig. 7A, lanes 3 and 4). The deletion had some deleterious effect on the virus and slowed its replication rate.

**Construction and analysis of mutant lacking the entire R region.** When homologous DNAs are introduced into cells in a calcium phosphate precipitate, recombination can occur between them which will repair lesions (17, 22, 25) and restore wild-type function. The frequency of such events is reduced when the DEAE Dextran procedure is used (unpublished observations), but is still detectable. Thus, one explanation for the observed behavior of the mutant viral DNAs is that virus is formed only after a DNA-DNA recombination event between the LTRs. This explanation is made unlikely by the fact that the mutant DNAs give rise to wild-type numbers of plaques; when recombination is required to generate virus, we usually find a 100 to 1,000-fold reduction in the number of plaques (unpublished observations). To eliminate this possibility, however, another deletion mutant was constructed.

A subclone was prepared containing only the 5′ LTR and small amounts of flanking DNA. This DNA was cleaved with *Sma*I, digested with BAL 31, and ligated. The DNA was used to transform *E. coli* HB101 to ampicillin resistance, and deletion mutants were identified by screening DNAs from colonies. The DNA sequence of one mutant, *dl*439, was determined; 89 bp, including the entire R region, was removed (Fig. 5 and 6). This mutation was then transferred back to plasmid pT11 to reconstruct an intact provirus carrying the deletion in the 5′ LTR. This DNA could not lead to the formation of virus by strong-stop jumping, but direct recombination between the two LTRs would generate a wild-type viral DNA. The region of homology within which such recombination could occur is essentially identical to that available in our other constructs: namely, the entire U5 region. The results showed that, in contrast to the other constructs, this DNA had an enormously reduced biological activity. In repeated tests we found that the *dl*439 DNA was either totally XC negative or gave at most 2 to 20 plaques per 100 ng when applied to cells. The resulting virus, recovered only after extended incubation of the recipient cells, was wild type in structure (data not shown). Thus, recombination during transfection did occur, but the frequency of virus formation was 100 to 500 times lower than that observed when DNAs that can carry out strong-stop transfers were used. This recombination event cannot account for the virus seen in our earlier experiments. The efficient generation of virus from those DNAs must have resulted from the use of the viral replicative functions.

**Strong-stop DNAs synthesized in vitro by mutant viruses.** Based on the above results, the simplest explanation for the aberrant loss of mutations in the 5′ R region was that a prematurely terminated strong-stop DNA ("weak-stop DNA" [6]), which did not contain the mutation, was sometimes jumping to the 3′ end of the template and serving as a primer for further synthesis. This event could be specific to the mutants or could be occurring at low frequency even in replication of wild-type virus.

Linker insertion mutants might synthesize high levels of such prematurely terminated DNAs, because the template would contain a long palindrome (18 bp) which might cause the reverse transcriptase to pause before copying the mutation (D. Mills, unpublished observations). To test this notion, virions were harvested from cell lines producing three viruses: mutant *in*479/8743, which stably carried the *Eco*RI insertion in both LTRs; *dl*450/8714, which stably carried the R region deletion in both LTRs; and wild-type virus. Endogenous reverse transcriptase reactions were carried out with these virions, and the products were analyzed by polyacrylamide gel electrophoresis (Fig. 8) after treatment with RNase. The major product synthesized by wild-type virus was the normal 145-nucleotide strong-stop DNA; abundant weak-stop species 105 and 75 nucleotides in size were also apparent (lane 1). Mutant *in*473/8743 yielded only two major species; a full-length strong-stop DNA, migrating more slowly than the wild type and corresponding to a size approximately 12 nucleotides longer, and a prominent species approximately 120 nucleotides in length, corresponding to a DNA terminating at the position of the insertion of the palindrome (lane 2). This species was much more abundant relative to the full-length molecule than the shorter weak-stop DNAs, which were present in both the wild type and mutants. No DNA of this size was formed by wild-type virus; thus, the presence of the insertion does cause reverse transcription to pause and does result in the accumulation of shorter DNAs. Mutant *dl*450/8714 showed only a single major species, a short strong-stop DNA of 116 nucleotides (lane 3). This corresponds closely to termination of synthesis at the end of the template RNA. As expected, this size was close to the size of the prominent stop in mutant *in*479/8743 because both terminations occur at the *Sma*I site in the R region.

## DISCUSSION

Our initial goal was to test whether the model of reverse transcription, as developed from analyses of in vitro reactions, was in fact correct in vivo. An important prediction of the model is that alterations in the 5′ R region of the LTR should be transferred to the 3′ LTR and then transmitted to progeny virus thereafter (Fig. 1). The results presented above suggest that DNA sequences in the R region of the retroviral LTR do not always behave in vivo as would be predicted by the current model of reverse transcription (4). Instead, mutations in this region are sometimes lost. This loss of the mutation does not occur when it is initially present in both LTRs.

One possible explanation for the loss of the mutation was that recombination occurred during the transfection procedure, either between the two LTRs of one mutant DNA molecule, or between LTRs of two different molecules. This laboratory (P. Schwartzberg and S. P. Goff, unpublished observations) and others (17, 22, 25) have shown that such homologous recombination events can occur, and that such events can reconstruct complete genes from two overlapping portions of a gene (17). In our hands, mixtures of two mutant viral DNAs give rise to wild-type recombinants after DEAE Dextran transfection at rather low efficiencies: usually 100- to 1000-fold fewer plaques per μg are observed than when intact DNA is used. Since the mutant DNAs with alterations in the 5′ LTR yielded as many plaques as wild-type DNA, this explanation seemed unlikely to account for the loss of 5′ mutations. To directly test whether this recombination could be the source of wild-type virus, we constructed a larger deletion (*dl*439) that removed all of the R region and therefore could not carry out a strong-stop jump; this mutant could give viable virus only by recombination during the transfection process. The region available for such recombination events was nearly identical to the region available in the earlier mutants (the U5) region). The results showed that this molecule was extremely poorly infectious, yielding 100- to 500-fold fewer plaques than wild-type DNA. Thus, DNA-DNA recombination does occur, but at such low frequencies that it cannot account for the high-efficiency appearance of wild-type virus from mutants like *in*479.

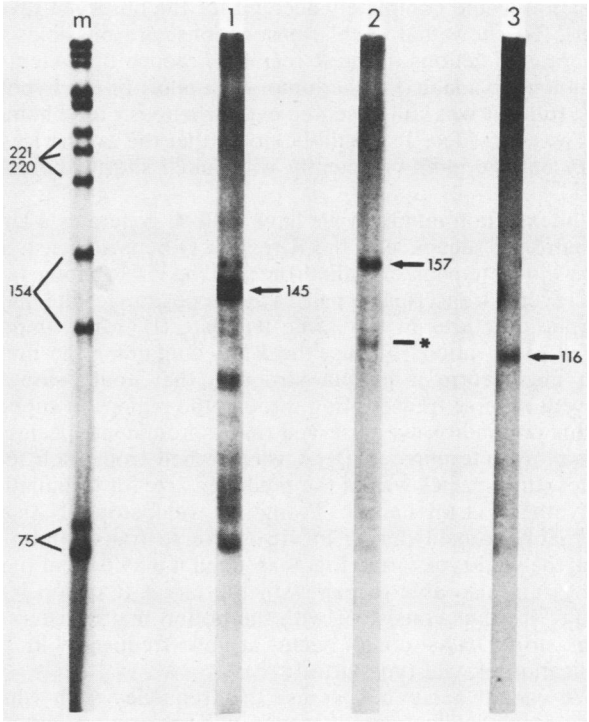



FIG. 8. Polyacrylamide gel electrophoresis of strong-stop DNAs synthesized during in vitro reverse transcription reactions by purified virions of mutant viruses. Lane m shows marker DNA fragments produced by cleavage of pBR322 DNA by *Hin*fI. The sizes of the fragments in bases are indicated; two bands are produced from each double-stranded fragment because of differences in mobility between the two separated strands. Lane 1 shows wild-type virus, lane 2 shows *in*479/8743 virus, and lane 3 shows *dl*450/8714 virus. The arrows indicate the largest prominent specie synthesized by each virus; sizes are as predicted from the known structure of the mutant DNA. Wild-type virus shows the prominent strong-stop DNA migrating at 145 bases and many shorter weak-stop DNAs (6). The insertion mutant makes a strong-stop DNA that is longer by approximately 12 bases and also shows a new major weak-stop DNA at approximately 120 bases (asterisk). The deletion mutant makes a shorter strong-stop DNA, migrating at approximately 116 bases.

A second possibility that we cannot rule out is that the mutation is successfully copied into strong-stop DNA and translocated, but that the mutation is subsequently repaired by using the R region at the 3′ end as a template. Reverse transcriptase has not been reported to be capable of "proofreading" mismatched DNAs. The rather low level of fidelity reported for the avian enzyme (1) would suggest that the enzyme does not, in fact, carry out such proofreading functions. Nevertheless, this or some other enzyme might be able to correct the mismatches between the translocated strong-stop DNA and the target RNA.

It is also possible that the insertion mutations are lost by some excision process. If this event can occur, however, it must occur precisely: the viruses that have lost the *Eco*RI site have been shown to carry the *Sma*I site that was originally interrupted by the insertion. In addition, such an excision event was not observed in viruses that carry the mutation in both LTRs. We therefore consider this possibility to be remote.

The explanation that we favor is that the entire R region is not always transferred from the 5′ end to the 3′ end during a strong-stop jump. The occasional transfer of prematurely terminated (weak-stop) DNAs that have not yet copied the mutation would completely account for the observed reversions. To show that such translocations. are possible, we generated deletions in the R region to reduce the extent of homology available for the jump. In support of the hypothesis, transfer was still observed even when very little homology was provided. It should be noted that the avian viruses carry out the jumps efficiently with much shorter homologies.

The insertion mutation we have studied generates a large palindromic sequence in the R region. Q beta replicase has been shown to pause at palindromes in the viral genome (14), and reverse transcriptase pauses at the position of the major hairpin structures in this same template (D. Mills, unpublished observations). Thus, the RNA containing the mutation might form a hairpin structure that could slow or prevent reverse transcription through the region. In support of this possibility, we observed that a prominent species of prematurely terminated DNA was formed from such templates; this species was of the predicted size for termination of synthesis at the hairpin. Prominent weak-stop DNAs are invariably formed during in vitro reverse transcription of even the wild-type virus (Fig. 8) (6), and it may be that these DNAs are also able to pair with the target R region. Our results are thus consistent with the notion that transfers of weak-stop DNAs could occur at low frequency in the replication of wild-type virus.

We cannot accurately assess the frequency with which weak stops might jump, because if both weak stops and strong stops can jump, forming a mixture of wild-type and mutant viruses, the wild-type virus would probably outgrow the mutant. This would give a falsely high impression of the frequency of jumping of the shorter weak-stop DNAs. It might be, for example, that weak stops jump at a frequency of 10% and form wild-type virus at an abundance of 10%; this virus could then become the vast majority of the virus in the population after several rounds of replication. It should be noted, however, that the virus undergoes a minimum of replicative cycles in our experiments before analysis of the genome structure (probably less than 10), and mutant DNAs such as *in*479 yielded fully wild-type viral genomes with no detectable mutant genomes present (less than 5%). Before any wild-type virus could overgrow the population, this virus had to arise at some frequency, and it had to arise in very early events. It could not have arisen later, because we have shown that virus formed with the mutation in both LTRs stably retain the insertion.

There are several steps in the life cycle of the virus which may select for the outgrowth of wild-type virus over mutants with alterations in the R region. Even when the region is copied, the strong-stop DNA might form a hairpin and might be poorly able to pair with the 3′ end of the RNA. Other steps in the life cycle of the virus, including transcriptional termination and polyadenylylation, could also be affected quantitatively (2). Direct analysis of the steps affected by these mutations might help define further functions encoded in this region.

## LITERATURE CITED


1. **Battula, N., and L. A. Loeb.** 1974. The infidelity of avian myeloblastosis virus deoxyribonucleic acid polymerase in polynucleotide replication. J. Biol. Chem. **249**:4086–4093.
2. **Benz, E. W., R. M. Wydro, B. Nadal-Ginard, and D. Dina.** 1980. Moloney murine sarcoma proviral DNA is a transcriptional unit. Nature (London) **288**:665–669.
3. **Coffin, J. M., T. C. Hageman, A. M. Maxam, and W. A. Haseltine.** 1978. Structure of the genome of Moloney murine leukemia virus: a terminally redundant sequence. Cell **13**: 761–773.
4. **Gilboa, E., S. W. Mitra, S. P. Goff, and D. Baltimore.** 1979. A detailed model of reverse transcription and tests of crucial aspects. Cell **18**:93–100.
5. **Goff, S. P., P. Traktman, and D. Baltimore.** 1981. Isolation and properties of Moloney murine leukemia virus mutants: use of a rapid assay for release of reverse transcriptase. J. Virol. **38**:239–248.
6. **Haseltine, W. A., D. G. Klein, A. Panet, E. Rothenberg, and D. Baltimore.** 1976. Ordered transcription of RNA tumor virus genomes. J. Mol. Biol. **106**:109–131.
7. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–371.
8. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**:193–197.
9. **Katz, L., D. T. Kingsbury, and D. R. Helinski.** 1973. Stimulation by cyclic adenosine monophosphate of plasmid deoxyribonucleic acid replication and catabolite repression of the plasmid deoxyribonucleic acid-protein relaxation complex. J. Bacteriol. **114**:577–591.
10. **Lobel, L. I., and S. P. Goff.** 1984. Construction of Mutants of Moloney murine leukemia virus by "Suppressor linker" insertional mutagenesis: the positions of viable insertion mutations. Proc. Natl. Acad. Sci. U.S.A. **81**:4149–4153.
11. **Mandel, M., and A. Higa.** 1970. Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. **53**:159–162.
12. **Maxam, A., and W. Gilbert.** 1979. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–599.
13. **McCutchan, J. H., and J. S. Pagano.** 1968. Enhancement of the infectivity of SV40 deoxyribonucleic acid with diethyl-aminoethyl-dextran. J. Natl. Cancer Inst. **41**:351–357.
14. **Mills, D. R., C. Dobkin, and F. R. Kramer.** 1978. Template-determined, variable rate of RNA chain elongation. Cell **15**: 541–550.
15. **Mitra, S. W., S. P. Goff, E. Gilboa, and D. Baltimore.** 1979. Synthesis of a 600-nucleotide-long plus-strand DNA by virions of Moloney murine leukemia virus. Proc. Natl. Acad. Sci. U.S.A. **76**:4355–4359.
16. **Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling of DNA to high specific activity by nick translation. J. Mol. Biol. **113**:237–258.
17. **Robert de Saint Vincent, B., and G. M. Wahl.** 1983. Homologous recombination in mammalian cells mediates formation of a functional gene from two overlapping gene fragments. Proc. Natl. Acad. Sci. U.S.A. **80**:2002–2006.

18. **Rowe, W. P., W. E. Pugh, and J. Hartley.** 1970. Plaque assay techniques for murine leukemia viruses. Virology **42**:1136–1139.
19. **Schwartzberg, P., J. Colicelli, and S. P. Goff.** 1983. Deletion mutants of Moloney murine leukemia virus which lack glycosylated *gag* protein are replication competent. J. Virol. **46**:538–546.
20. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
21. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
22. **Subramani, S., and P. Berg.** 1983. Homologous and nonhomologous recombination in monkey cells. Mol. Cell. Biol. **3**: 1040–1052.
23. **Vogelstein, B., and D. Gillespie.** 1979. Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. U.S.A. **76**:615–619.
24. **Wahl, G. M., M. Stern, and G. R. Stark.** 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl paper and rapid hybridization using dextran sulphate. Proc. Natl. Acad. Sci. U.S.A. **76**:3683–3687.
25. **Wake, C. T., and J. H. Wilson.** 1979. Simian virus 40 recombinants are produced at high frequency during infection with genetically mixed oligomeric DNA. Proc. Natl. Acad. Sci. U.S.A. **76**:2876–2880.
26. **Wang, L.-H., P. H. Duesberg, T. Robins, H. Yokota, and P. K. Vogt.** 1977. The terminal oligonucleotides of avian tumor virus RNAs are genetically linked. Virology **82**:472–492.