JOURNAL OF VIROLOGY, Sept. 1991, p. 4786–4797
0022-538X/91/094786-12$02.00/0
Copyright © 1991, American Society for Microbiology
Vol. 65, No. 9

Characterization of Large Deletions Occurring during a Single Round of Retrovirus Vector Replication: Novel Deletion Mechanism Involving Errors in Strand Transfer

GARY A. PULSINELLI AND HOWARD M. TEMIN*

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

Received 25 March 1991/Accepted 4 June 1991

Retroviruses mutate at a high rate during replication. We used a spleen necrosis virus-based vector system and helper cell line to characterize mutations occurring during a single round of retrovirus replication. The vector used, JD216HyNeo, codes for two drug resistance genes, hygromycin resistance (*hygro*) and neomycin resistance (*neo*). The downstream *neo* gene is expressed only when a mutation alleviates a block to splicing which is located in the upstream *hygro* gene. The mutations allowing splicing were large deletions, ranging in size from about 500 to about 2,000 bp. Most of the mutant proviruses lacked the encapsidation sequence, as shown by our inability to rescue the mutant proviruses with wild-type reticuloendotheliosis virus strain A and confirmed by Southern blotting and direct DNA sequence analysis. We therefore concluded that most of the deletions arose during reverse transcription in the target cell, rather than during transcription in the host cell. The sequence data also indicated that the deletions occurred by at least three different mechanisms: (i) misalignment of the growing point; (ii) incorrect synthesis and termination in the primer-binding sequence during synthesis of the plus-strand strong-stop DNA; and (iii) incorrect synthesis and termination before the primer-binding sequence during synthesis of the plus-strand strong-stop DNA. The second mechanism also led to the incorporation of cellular sequences into the proviral genome, pointing to a potential novel mechanism by which retroviruses can acquire cellular genes.

Retroviruses are known to mutate at high rates (1, 24, 25). This high rate of mutation means that retrovirus species exist as heterogeneous populations, with the members representing a wide range of variants. These variants provide the genetic material on which selection can operate, allowing retroviruses to adapt quickly to changing environmental conditions. Thus, a high rate of mutation is an important component of the success of retroviruses.

An important source of these mutations is the reverse transcription process. Reverse transcriptase itself is an error-prone enzyme (22), and it makes many mistakes in the copying of the viral RNA genome into the viral DNA genome. These mistakes include base-pair substitutions, deletions, and insertions (20, 21), all of which increase the variation in a retrovirus population. The reverse transcription process also has other unusual features—in particular, the strand transfer steps (1, 25)—which may be important in the creation of mutant progeny.

To study the mutation process in more detail, we used a system previously developed in this laboratory to study mutations occurring during a single round of retrovirus vector replication (5). This system uses a series of helper cells and nonhelper target cells to generate a provirus that has been propagated through one round of replication and can proceed no further. This system enabled us to obtain directly a per-cycle mutation rate and also to isolate the events of a single round of replication from the cumulative events of multiple rounds (5).

The vector used in these studies (JD216HyNeo) codes for two resistance genes, *hygro* and *neo* (*hygro* encodes the hygromycin B gene, conferring resistance to hygromycin; *neo* encodes the bacterial aminogycoside 3′-phosphotransferase gene, conferring resistance to G418). Expression of the downstream *neo* gene is observed only if there is a mutation in the upstream *hygro* gene (5). We characterized mutations in this vector generated during a single round of retrovirus replication. The observed mutations were large deletions which we propose occurred by three mechanisms. The first of these mechanisms involved misalignment of the growing point, as previously described by Pathak and Temin (21). The second and third mechanisms involved errors during plus-strand strong-stop DNA synthesis and transfer, either within or just before the primer-binding site (PBS). These latter mechanisms represent a new class of errors, one that is unique to retroviral replication.

MATERIALS AND METHODS

Plasmids and vectors. The vector pJD216HyNeo has been described previously (5). The p designates the plasmid vector; names without the p (e.g., JD216HyNeo) denote viruses (or proviruses) derived from this vector. All sequence coordinates are given relative to the 5′ end of the 5′ long terminal repeat (LTR) in the plasmid.

Cells, transfection, and infection. D17 cells are derived from a dog osteosarcoma and are permissive for spleen necrosis virus (SNV) infection. C3A2 cells are a helper cell line created by transfection of DNA for a defective wild-type reticuloendotheliosis virus strain A (REV-A) provirus into D17 cells (27). Chicken embryo fibroblasts are primary chicken cells permissive for SNV infection. Cell culture and selection were done as previously described (5, 7, 26).

Transfection was performed by the dimethyl sulfoxide-Polybrene method (13). Infections were done with 0.2 ml of

* Corresponding author.

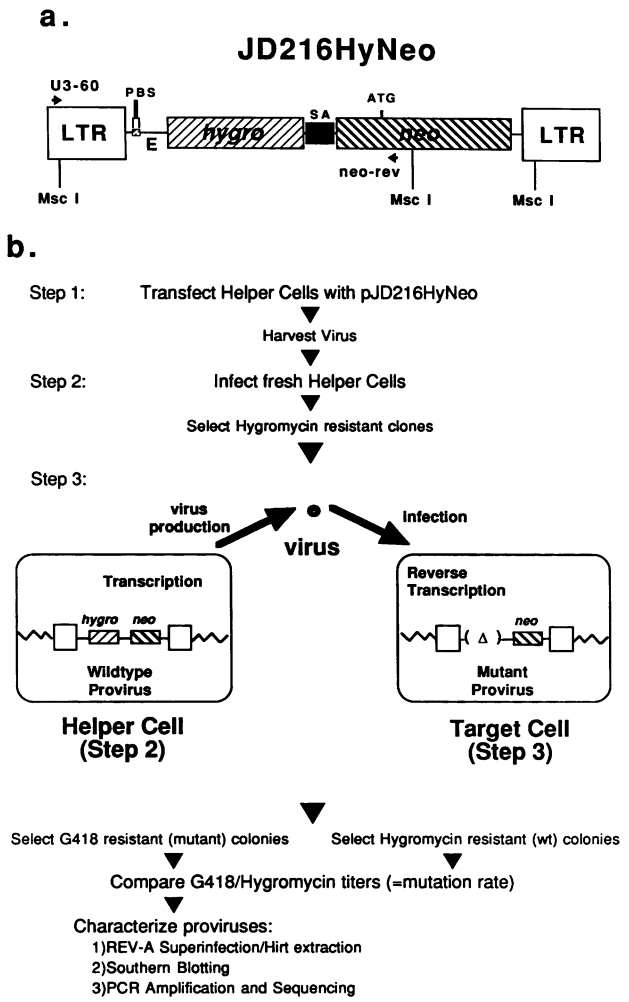

FIG. 1. (a) Vector JD216HyNeo. The positions of the *Msc*I sites and the locations of the PCR primers are indicated. *hygro*, hygromycin B resistance gene; *neo*, bacterial aminoglycoside 3′-phosphotransferase gene; LTR, SNV LTR; E, SNV encapsidation sequence; SA, REV-A splice acceptor fragment; ATG, initiation codon for aminoglycoside 3′-phosphotransferase; U3-60, U3-60 PCR primer; neo-rev, neo-rev PCR primer. (b) Experimental scheme. pJD216HyNeo DNA was transfected into C3A2 helper cells (step 1). After selection for hygromycin resistance, helper-free vector virus was harvested and was used to infect fresh C3A2 helper cells (step 2). Hygromycin-resistant colonies were selected, and helper-free vector virus was harvested. This vector virus was used to infect two sets of target (nonhelper) cells (step 3). The target cells were selected for hygromycin or G418 resistance, and the resulting colonies were counted to ascertain the viral titer. The hygromycin-resistant titer was the overall (wild-type) titer, and the G418-resistant titer was the mutant titer; the ratio of the mutant to overall titer was the mutation rate. The mutant proviruses in the G418-resistant colonies were further characterized by REV-A superinfection and then Hirt extraction of unintegrated viral DNA, Southern blotting of genomic DNA, and PCR amplification and sequencing of mutant proviruses. Open boxes, SNV LTRs; Δ, deletion in the *hygro* gene; hatched boxes, resistance genes; wavy lines, cell DNA; straight lines, vector provirus DNA.

virus plus 0.2 ml of Polybrene (100 μg/ml for D17 cells or 15 μg/ml for chicken embryo fibroblasts) for 40 min.

To obtain the hygromycin- and neomycin-resistant titers, we used 10-fold dilutions of virus to infect separate sets of D17 cells. One set was selected with hygromycin B, and the wild-type titer was determined at endpoint dilution; the other set was selected with G418, and the mutant titer was determined at endpoint dilution (5).

Preparation of DNA and Southern blotting. Unintegrated viral DNA was prepared by the method of Hirt (10). Total genomic DNA was prepared by standard procedures (15).

Southern blotting was performed by standard protocols (15, 23). *Msc*I restriction enzyme was purchased from New England BioLabs. Nylon membranes (Pall; MSI New England Nuclear) were used, and blotting was performed with a vacuum apparatus (LKB 2016 VacuGene). Probes were labeled by nick translation with the Amersham labeling kit (kit N.5000). Hybridization was done overnight at 65°C.

PCR. The polymerase chain reaction (PCR) was done on genomic DNA isolated from the resistant cell clones of interest (8). Primers were purchased from Genetic Designs/Genosys (The Woodlands, Tex.). The following primers were used: U3-60, GCTTGCCCTGGCCACTAACCG, located in the U3 region of the 5′ LTR at position 60; and neo-rev, ACCCAAGCGGCCGGAGAACCT, located just past the initiation codon for the aminoglycoside 3′-phosphotransferase protein (*neo* gene) at position 2600. Amplifications were performed in a buffer consisting of 50 mM KCl, 10 mM Tris (pH 8.3), 0.01% gelatin, 1.8 mM $MgCl_2$, and 40 μM each deoxynucleoside triphosphate. Genomic DNA (1 μg) was added to this buffer with the two primers (5 pmol each for symmetric PCR or 0.5 pmol:50 pmol for asymmetric PCR). *Taq* polymerase was obtained from PE-Cetus. Amplifications were performed for 35 cycles in a Coy Laboratories model 60 Tempcycler. Products were analyzed by electrophoresis on 1% TBE agarose gels (15) (1/10th of the total PCR per lane).

PCR sequencing. The primers used in sequencing were, in addition to the amplification primers, U3-360, CACCCTG TAAGCTGTAAGCG, located in the U3 region of the 5′ LTR at position 360; and U5-550, TTGGCTCGCCTACT GGGTGGG, located in the U5 region of the 5′ LTR at position 550. Templates were prepared for sequencing by one of the following protocols.

For direct sequencing of symmetric PCR products, primers and excess nucleotides were removed with a Linkers 6 spin column (Boehringer Mannheim Biochemicals) according to the manufacturer's directions or by purification from agarose gels with NA-45 paper (Schleicher & Schuell). This product was prepared for sequencing by either NaOH denaturation, following the method in reference 6, or boiling, as described in reference 14.

For direct sequencing of asymmetric PCR products, the amplification was performed as above, but using a ratio of 50 pmol of neo-rev primer to 0.5 pmol of U3-60 primer, thus generating single strands of the minus strand of the provirus. The PCR product was precipitated away from the excess primers and nucleotides with ammonium acetate and isopropanol, as described in Chapter 5 of reference 8.

For some mutants, the PCR product was cloned to facilitate sequencing. Symmetric PCR product was passed through the Linkers 6 column as described above. The amplified DNA was then cut with appropriate enzymes (Boehringer Mannheim) and ligated into the vector pTZ19R (Pharmacia). The resulting double-stranded plasmid clones were denatured following the instructions in the Sequenase manual (United States Biochemical).

For all template preparation methods, sequencing was performed using the Sequenase Kit (United States Biochemical), essentially as described in the Sequenase manual, except that the annealing was done for 10 min at 37°C rather

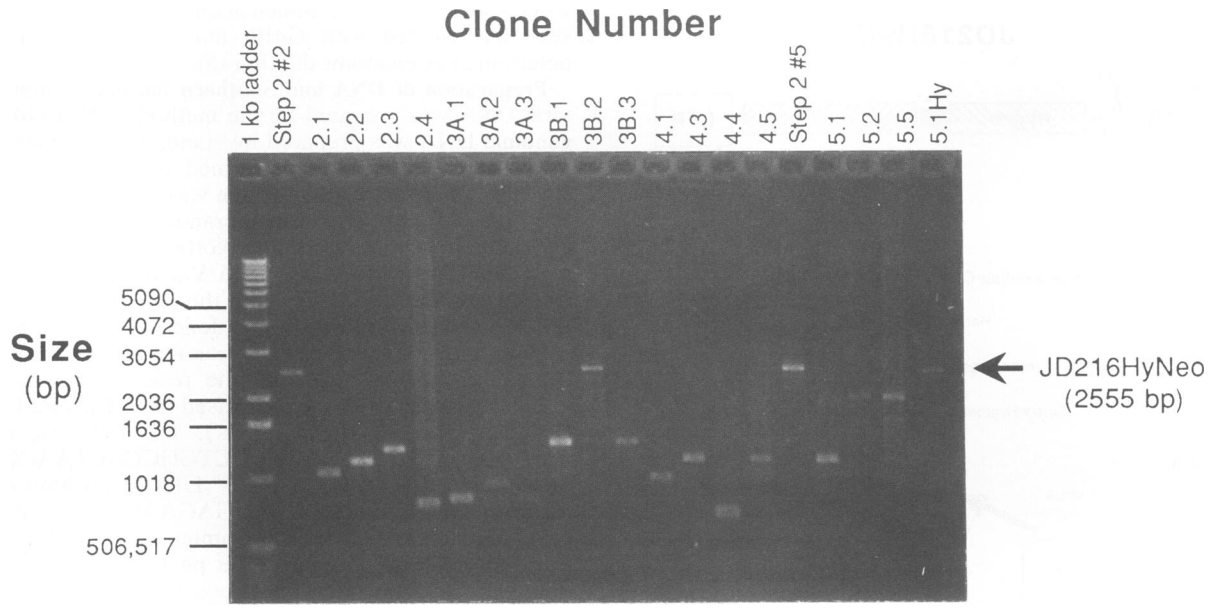

FIG. 2. Ethidium bromide-stained agarose gel of the PCR amplification products. The size markers are indicated on the left. Step 2 #2 and step 2 #5 are parental (Hygro^r) helper cell clones. 5.1Hy is a wild-type (Hygro^r) step 3 clone; the remaining clones are Neo^r step 3 clones. The expected size for a wild-type provirus (2,555 bp) is marked as JD216HyNeo.

than at 65°C. For direct sequencing of the PCR products, the supplied Mn^{2+} buffer was added; this addition was not necessary for plasmid sequencing. To obtain sequence far from the primers, we used the supplied sequence extension buffer.

The computer program DNA Strider (16) was used for sequence organization; the GCG Sequence Analysis package (3) was used to perform homology searches of the GenBank and EMBL sequence data bases.

Nucleotide sequence accession numbers. The inserted sequences in mutants 5.14 and 3A.1 have been submitted to GenBank (accession numbers M73045 and M73046, respectively).

RESULTS

Generation of mutant proviruses. The vector pJD216 HyNeo has been previously described, as has been the protocol for generating mutants in a single cycle of retroviral vector replication (5) (Fig. 1). Briefly, this vector was introduced into the C3A2 helper cell line by transfection. After selection for hygromycin resistance, helper-free JD216HyNeo virus was harvested and was used to infect fresh C3A2 helper cells. Individual hygromycin-resistant colonies were selected, and the resulting cell clones were expanded. Four of these clones (numbered 2 through 5) were selected for further study. Helper-free vector virus harvested from each of these clones was diluted serially and was used to infect two parallel sets of D17 cells. One set was selected for hygromycin resistance (Hygro^r), and the other was selected for G418 (neomycin) resistance (Neo^r). The titer of virus under each selection was determined, and the two titers were compared. Mutant (Neo^r) viruses appeared at a rate of about 1% of wild-type (Hygro^r) viruses, comparable to the rate reported previously (5). Individual Neo^r clones were picked and expanded. These expanded clones were the source of the mutant proviruses analyzed below. In the names of these clones, the first number designates the parental step 2 clone that was the source of the infecting virus and the second number designates the individual Neo^r step 3 clone.

Analysis of mutant proviruses. We first attempted to rescue the mutant proviruses by superinfection with REV-A helper virus. The resulting vector plus helper virus was used to infect chicken embryo fibroblasts, from which we prepared unintegrated viral DNA. This DNA was Southern blotted and hybridized to the *neo* gene probe. On these blots, 18 of 36 clones did not give detectable bands, suggesting that some necessary *cis* sequence was missing from the mutants (data not shown). For those 18 clones that did have detectable bands, 13 were wild-type JD216HyNeo proviruses which were present in addition to the mutant provirus that led to the Neo^r phenotype (see Discussion). Six cell clones did have rescuable mutant proviruses, and these were further mapped by restriction analysis and Southern blotting of the unintegrated DNA.

We then analyzed the mutant proviruses using Southern blotting of genomic DNA. *Msc*I was chosen because it cuts only in regions presumably intact in the mutant proviruses (i.e., the U3 region of the LTRs and within the selected *neo* gene; see map in Fig. 1a). After digesting with *Msc*I and hybridizing with the *neo* fragment, we determined that all mutants had large deletions, ranging in size from about 500 to about 2,000 bp (data not shown). We also determined, using the encapsidation sequence as a probe on Southern blots of genomic DNA, that the nonrescuable proviruses had deleted the encapsidation sequence region (data not shown).

We next analyzed the mutant proviruses by the PCR. Primers in the U3 region of the LTR and just past the initiation codon of the *neo* gene, regions which were expected to remain in the mutant proviruses, were used to amplify a 2,555-bp region in the wild-type provirus. These same primers were then used to amplify the corresponding region in the mutant proviruses. The difference in size between the wild-type and mutant amplification products

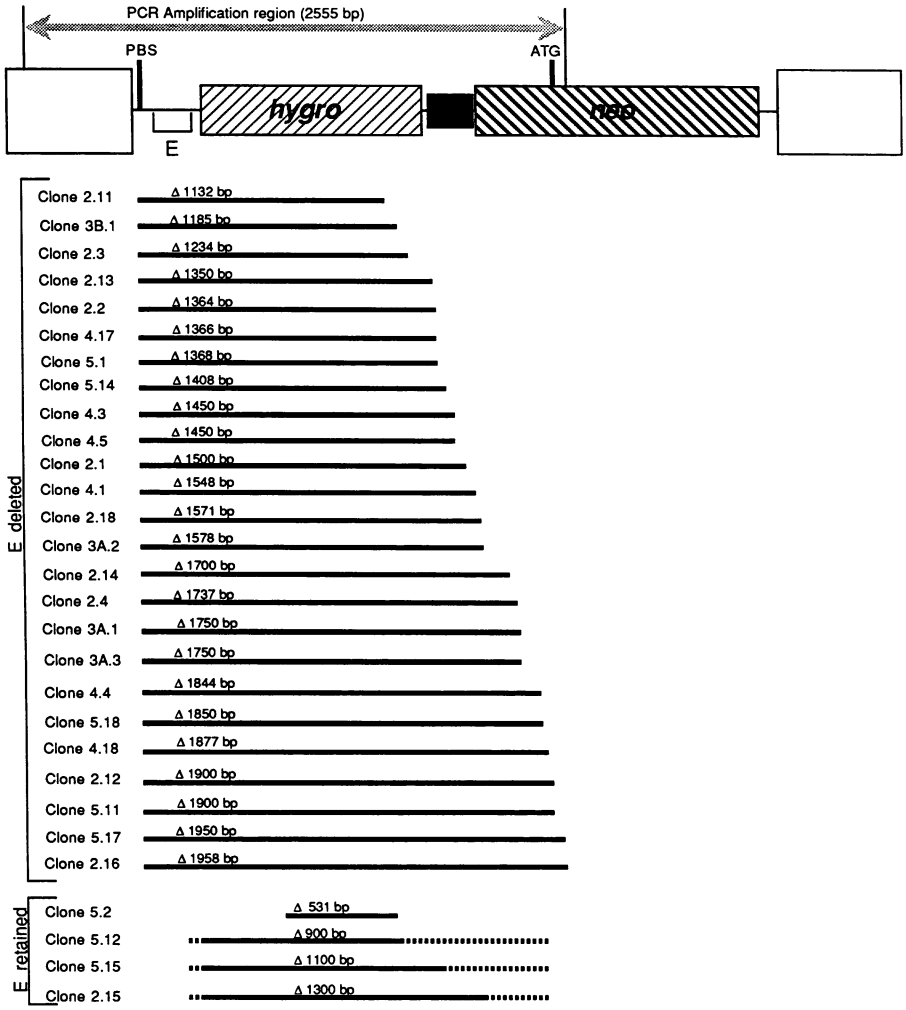

FIG. 3. Sizes of deletions in JD216HyNeo. Sizes were either estimated from the size of the PCR amplification product on agarose gels or obtained directly by sequencing. All the E-deleted deletions are shown with a common starting site for size comparison. The choice of the starting point is based on the observation that these deletions lack the E sequence, coupled with the sequence data, which shows that most deletions begin in or near the PBS; this narrows the starting range down to a small region, and so for simplicity we chose the PBS as the common start point. The dotted lines on the E-retained deletions indicate the range in which the alterations may lie; the exact endpoints have not been determined.

gave the size of the alteration. Using this approach, we confirmed and extended the results from the Southern blotting. Figure 2 shows an agarose gel with the PCR amplification products from a variety of step 3 clones. Again, the mutants were found to have large deletions, ranging in size from 500 to 2,000 bp (summarized in Fig. 3). In the cases in which both Southern blotting and PCR were used to analyze the same clone, the sizes of the alterations measured by the two methods were in good agreement.

Using PCR and several different direct sequencing protocols (see Materials and Methods), we determined the deletion endpoints of 16 of the mutants (Fig. 4 and 5). We divided these 16 alterations into three main classes, based on their position relative to the PBS region: the misalignment class, which occurred after the PBS (3 of 16); the PBS class, which occurred within the PBS region (10 of 16); and the pre-PBS class, which occurred just before the PBS region (3 of 16) (Fig. 4) (see Discussion).

Several of the mutants also contained insertions. One of the 16 clones, clone 5.14 (involving an error in plus-strand strong-stop DNA termination), contained a long sequence (489 bp) of nonviral origin, which was an open reading frame encoding 162 amino acids (Fig. 5a). A search of the GenBank and EMBL data bases revealed no sequences similar to this inserted sequence. Two other mutants (clones 2.2 and 3A.1) contained inserted tRNA sequences, arising from a failure of reverse transcriptase to terminate at the correct base in the tRNA primer. Clone 3A.1 contained, in addition to the 14 extra base pairs copied from the tRNA, 697 bp of inserted sequence, 505 of which arose from a partial duplication of LTR sequences and 192 of which arose from an unknown sequence of nonviral origin (Fig. 5b and see Fig. 8a); this mutant will be addressed more thoroughly in the Discussion. A search of the GenBank and EMBL data bases revealed that the inserted sequence in clone 3A.1 had a strong similarity to the canine blood-clotting factor IX mRNA (GenBank accession number M33826); the GCG program BESTFIT (3) revealed 81.6% identity between the two

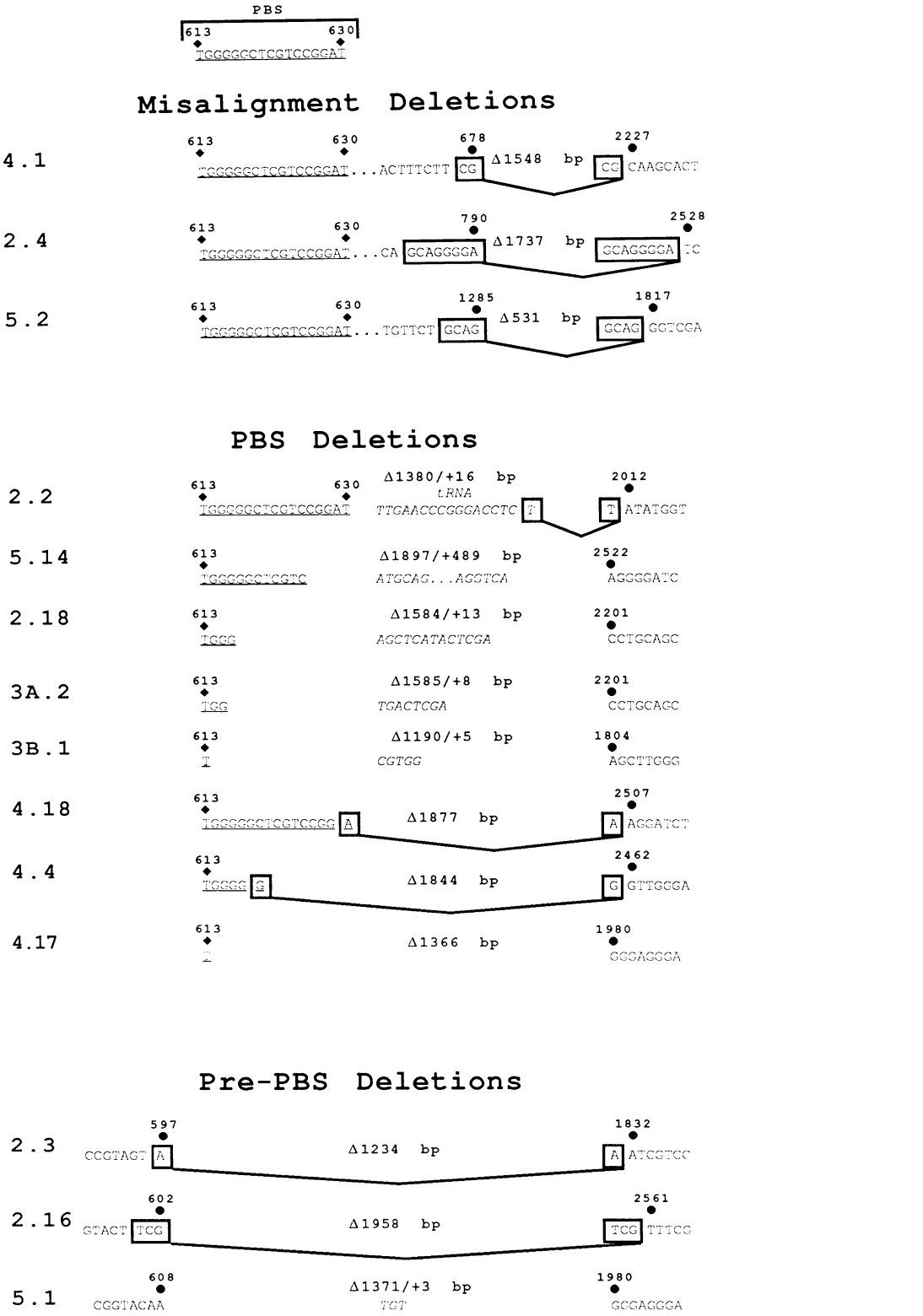

FIG. 4. Alteration sequences. Sequences of 14 of the deletion endpoints are shown. The numbers above the diamonds indicate positions in the PBS; the numbers above the large dots indicate the locations of the deletion endpoints relative to the 5′ end of the 5′ LTR. Underlining indicates the position of the PBS. Three small dots in the sequence indicate regions of sequence that are present but not shown. Italics indicate regions of sequence not found in the wild-type sequence. The bases in boxes indicate overlapping regions present in the normal sequence of both endpoints; only one copy remains in the mutant provirus.

sequences over a region of 161 bases. Clone 2.11 contained an unusual inserted sequence of 42 bp which consisted mainly of Cs and Ts and contained no Gs (Fig. 5c). This clone also contained at the deletion endpoint a 9-bp sequence from the *neo* gene and an overlapping 10-bp sequence from the *hygro* gene. A similar misincorporation of sequences from other parts of the viral genome was previously designated deletion with insertion by Pathak and Temin (21). Finally, four mutants (2.18, 3A.2, 3B.1, and 5.1) contained small inserts which were too short to allow speculation as to their origin; however, we believe that the mechanism which gave rise to these mutants may also bear some similarity to the deletions with insertions observed by Pathak and Temin (21).

DISCUSSION

We used a system developed by Dougherty and Temin (5) to study mutations generated during a single round of retrovirus replication. The vector virus construct used, JD216HyNeo, expressed the downstream *neo* gene only if a mutation occurred in the upstream *hygro* gene. Dougherty and Temin (5) determined that the wild-type construct did not splice correctly and that the mutations in some way alleviated this block to splicing. Our data from REV-A rescue and Southern blotting experiments indicated that these mutations were large deletions, ranging in size from about 500 to about 2,000 bp (Fig. 3), which presumably deleted the sequences that interfered with correct splicing in the wild-type vector.

Over one-third (13 of 36) of the G418-resistant clones contained two proviruses, one of which was generally wild-type and the other of which was the mutant. These doubly infected clones were observed because, in order to see the relatively rare mutants (which arose at a rate of 1%) while using a reasonably small number of petri dishes, we infected the cells with viral stocks which were fairly concentrated; that is, the infection was done at a high multiplicity of

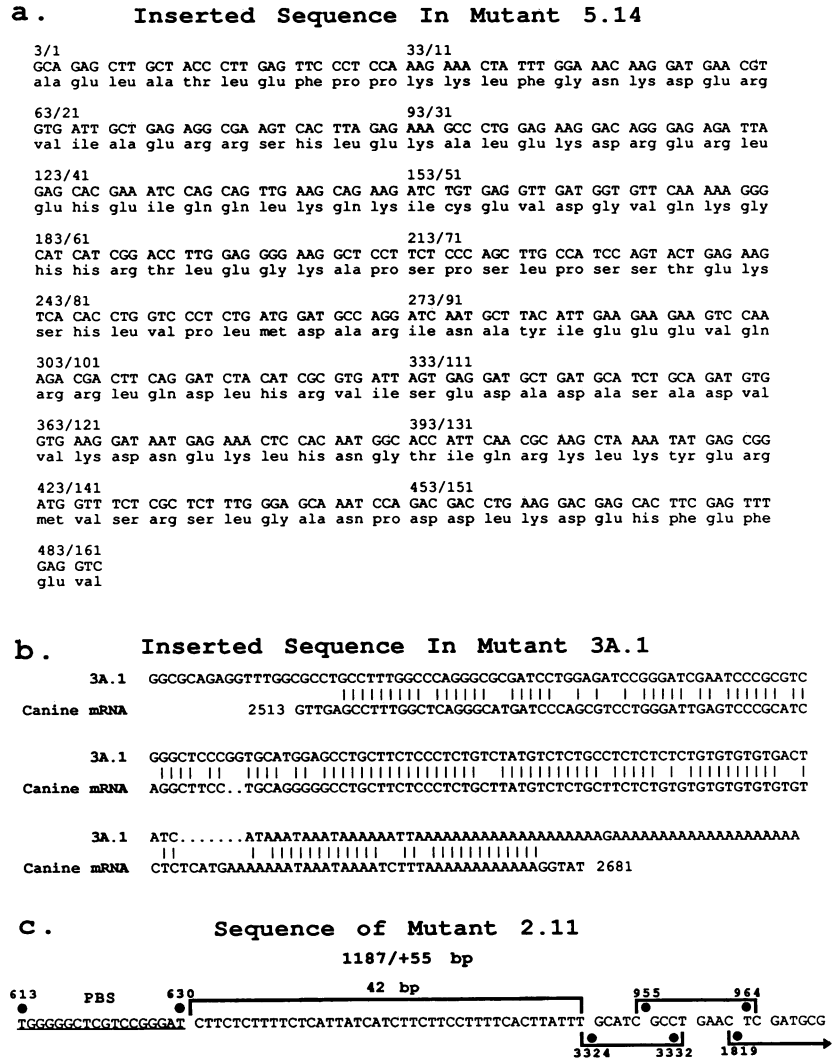

FIG. 5. (a) Sequence of 489-bp insert in mutant 5.14. A translation of the open reading frame contained in this sequence is also shown. (b) Sequence of 192-bp insert of nonviral origin in mutant 3A.1. The GCG BESTFIT alignment with the canine blood-clotting factor IX mRNA is also shown. (c) Sequence of mutant 2.11, showing the 42-bp inserted sequence (large upper bracket) and the overlapping regions from elsewhere in the vector sequence (smaller brackets, with the origins of the inserted base pairs in the proviral sequence indicated).

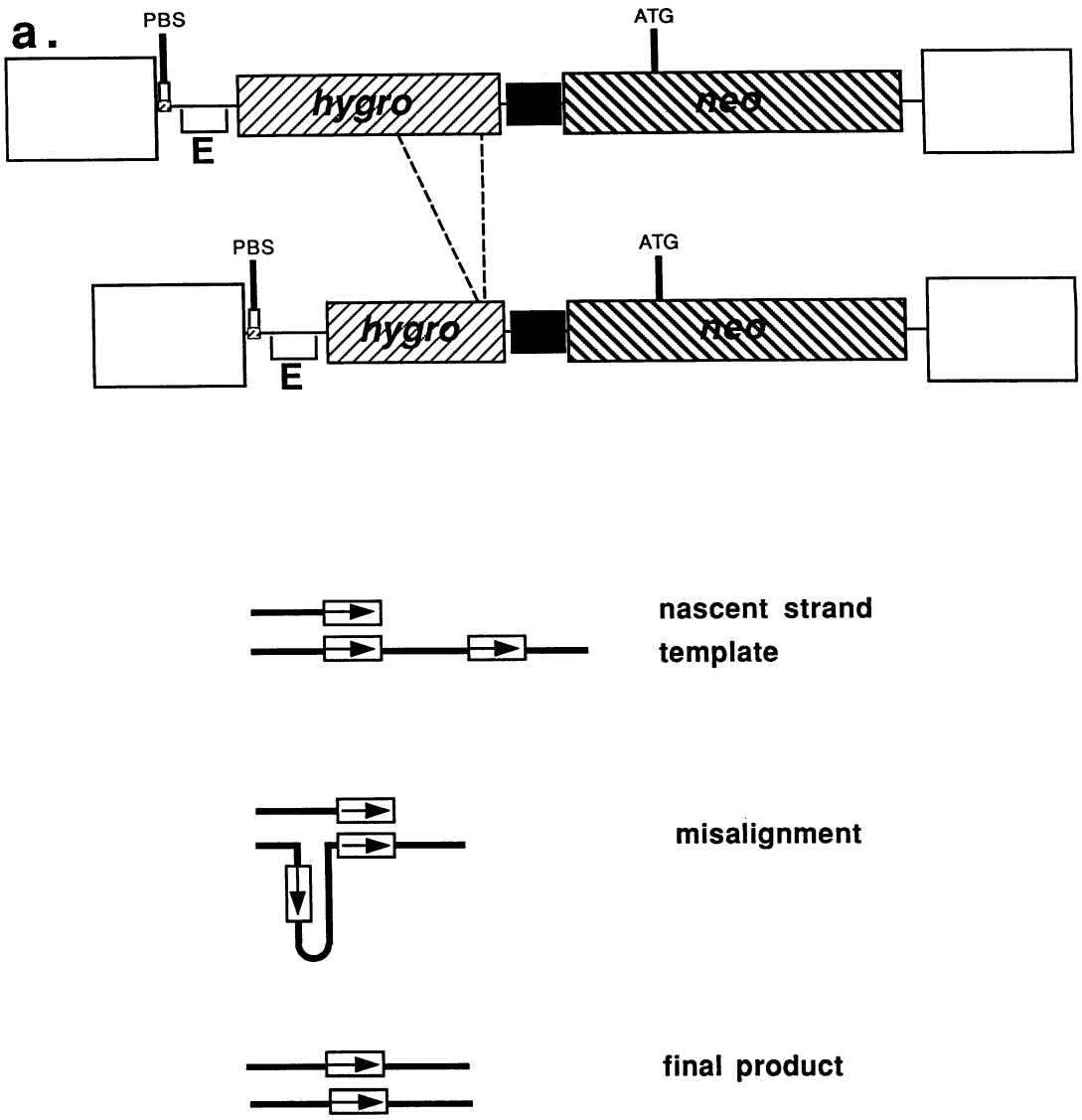

FIG. 6. Mechanisms of deletion. Symbols are as defined in the legend to Fig. 1. (a) Misalignment mechanism. The generation of mutant 5.2 is shown as an example. During reverse transcription, the growing point misaligned to a sequence elsewhere in the vector similar to the sequence being copied, deleting the intervening sequence. (b) PBS mechanism. The generation of mutant 2.2 is shown as an example. During plus-strand strong-stop DNA synthesis, reverse transcription failed to terminate at the correct place in the tRNA primer and instead copied 17 extra bases of the tRNA (hatched lines). The resulting plus-strand strong-stop DNA product contained PBS sequences, and so it could partially hybridize to the PBS, but it also contained extra 3′ bases that could not hybridize. These free bases annealed internally in the vector and the sequence was extended, resulting in the deletion of the intervening sequences. (c) Pre-PBS mechanism. The generation of mutant 2.16 is shown as an example. Since the minus-strand strong-stop DNA transfer was interstrand (see text) and reverse transcription occurred on only one of the two viral RNA molecules, the (full-length) minus-strand viral DNA contained R and U5 regions at its 3′ end (see also Fig. 7). These regions provided the homology by which the truncated (and incorrect) plus-strand strong-stop DNA product could anneal to the minus-strand DNA, with the exception of the incorrect 3′ bases. Plus-strand synthesis then proceeded as described in panel b above, resulting in the observed deletion.

infection of hygromycin resistance (but at an endpoint multiplicity for G418 resistance). Thus, cells which contained mutant proviruses also often contained additional wild-type proviruses.

A single cycle of retrovirus replication involves one round of transcription and one round of reverse transcription. The mutations could have arisen at either of these steps. However, since most mutant proviruses lacked the encapsidation sequence, we concluded that they arose during reverse transcription. If the deletions had occurred during transcription, the viral RNA from the step 2 cells would have lacked the encapsidation sequence and would not have been encapsidated into viral particles; therefore, it would not have been seen in the following round of infection of the step 3 cells.

Without an encapsidation sequence, the viral RNA from the step 3 cells cannot be encapsidated efficiently; in fact, we were unable to rescue the mutant proviruses using wild-type REV-A virus. The deleted proviruses are therefore unable to

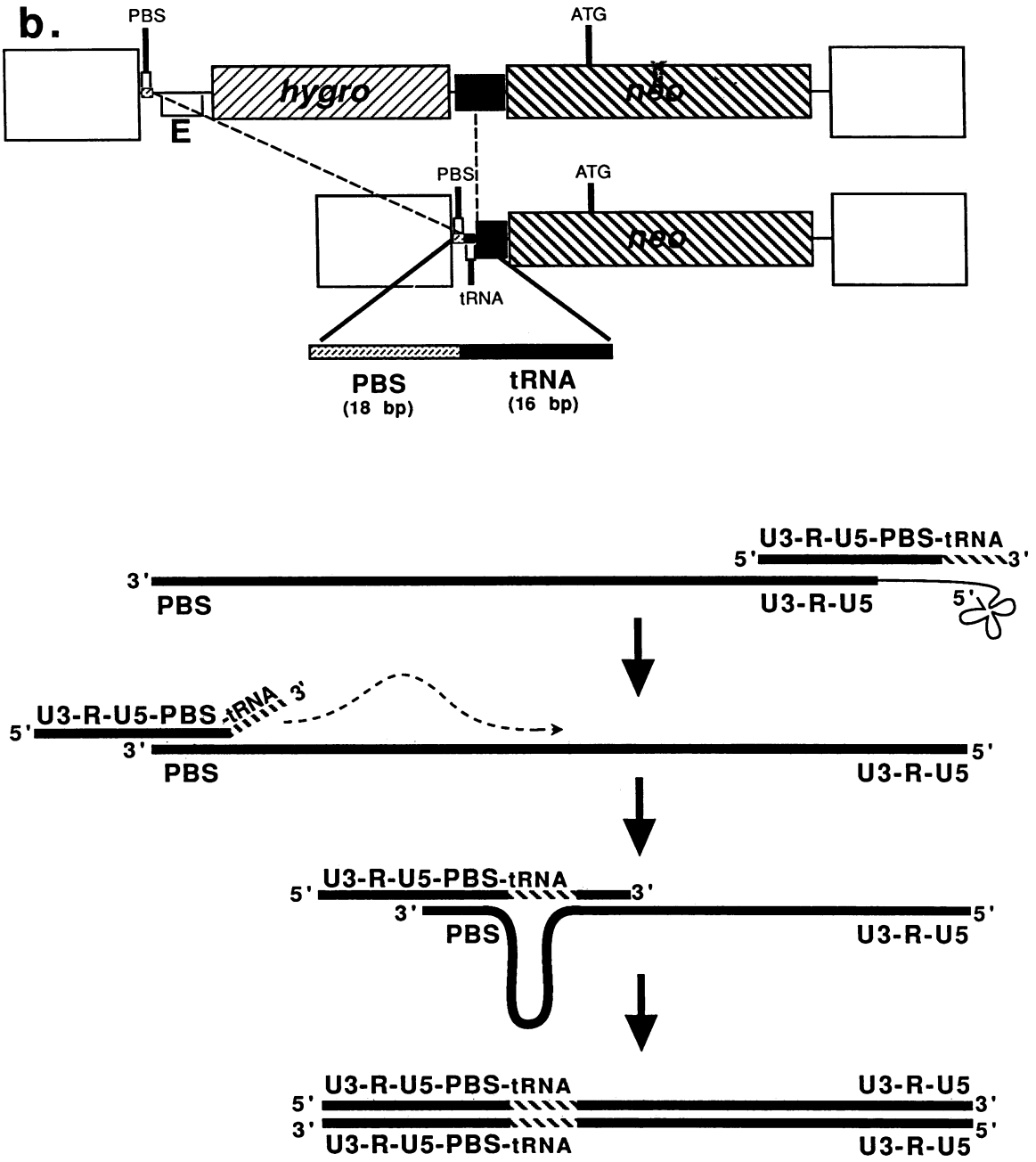

FIG. 6—*Continued.*

produce progeny virus that can infect other cells, even in the presence of a helper virus; these deletions are lethal in terms of further propagation.

From the data generated by direct PCR sequencing, we concluded that the deletions fall into three classes. The first and simpler class accounted for 3 of the 16 sequenced mutants. This class of mutants seemed to arise from a misalignment of the growing point to a similar sequence occurring elsewhere in the viral genome (Fig. 6a). These mutants were very much like those described previously by Pathak and Temin (21), who reported that they occur at a rate of 2×10^{-6} per nucleotide per single round of replication. From this reported rate, we estimated that our mutation target size was about 1,000 nucleotides (13 of 16 of this type X 1% per cycle = 2×10^{-3} per cycle [the observed overall rate]; 2×10^{-3}/1,000 nucleotides = 2×10^{-6} per nucleotide per cycle [the reported per-nucleotide rate]).

Mutant 5.2, which fell into this class, was the smallest deletion sequenced. It had deleted a region of 531 bp near the 3′ end of the *hygro* gene. Thus, this deletion helped define the minimal sequence that interferes with the expression of *neo* resistance. This region presumably contains sequences that in some way interfere with splicing, although we have not yet shown this property directly.

The other deletions were caused by mechanisms involving errors in the synthesis or termination of the plus-strand

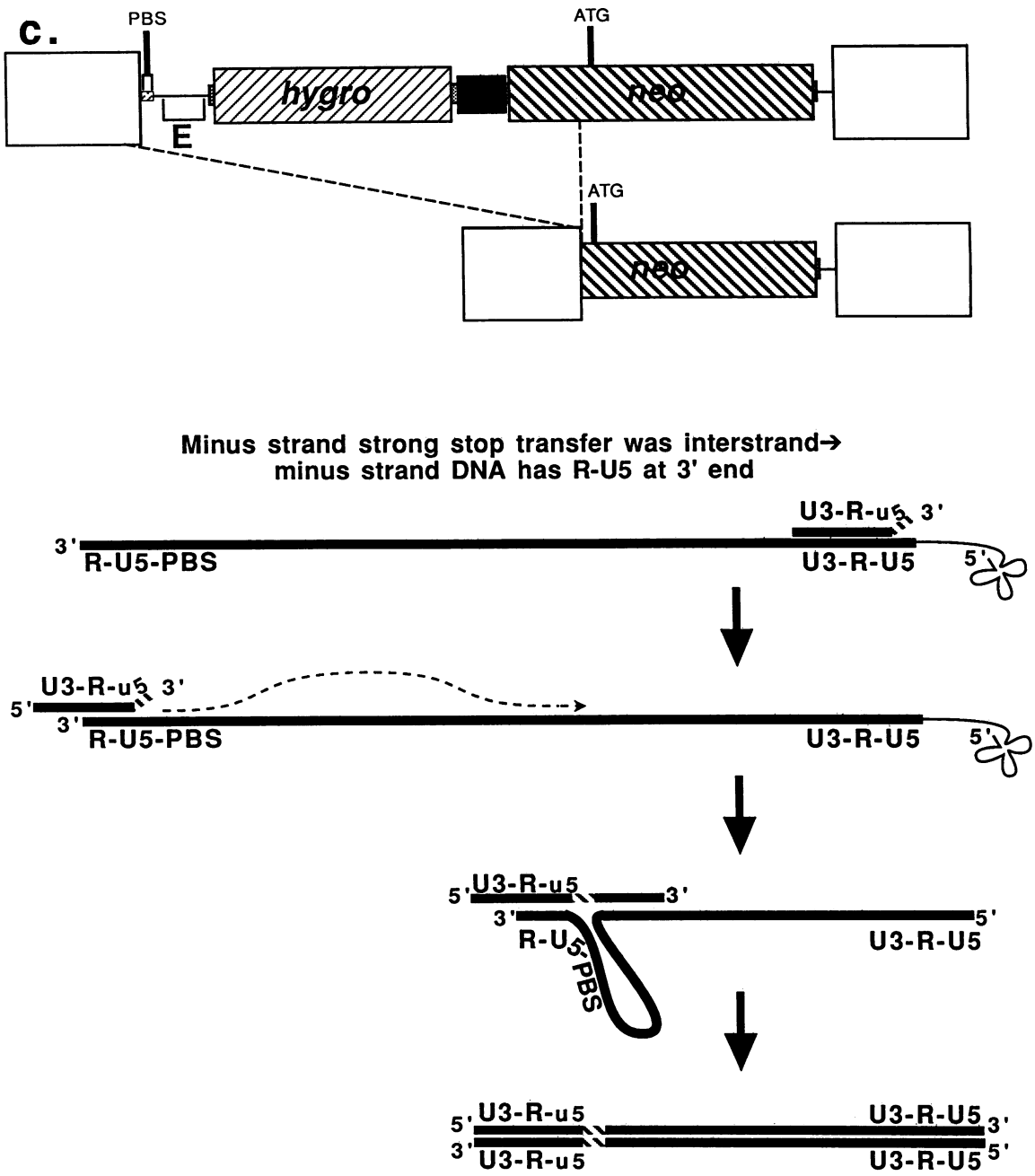

FIG. 6—*Continued.*

strong-stop DNA. These mutants are distinguished by the amount of the PBS copied: those that synthesized all or part of the PBS form the PBS termination class, and those that did not reach the PBS form the pre-PBS termination class.

The PBS termination class model is diagrammed in Fig. 6b. (For reference, the normal reverse transcription process is diagrammed in Fig. 7.) Misincorporation during synthesis, followed by incorrect termination of the plus-strand strong-stop DNA, resulted in a product with incorrect nucleotides at the 3′ end of the DNA. The sequence shown in the example, that of mutant 2.2, most clearly demonstrated this mechanism (Fig. 4 and 6b). Plus-strand strong-stop DNA synthesis failed to terminate at the correct base at the end of the PBS in the tRNA. Such a failure could have arisen either from a mistake during reverse transcription or from a tRNA that was not correctly methylated at the A that is the normal termination position. Synthesis continued into the tRNA, terminating at an unmodified base near the end of the codon loop. As a consequence, the plus-strand strong-stop DNA contained 16 extra base pairs of tRNA sequence at its 3′ end. (In the other mutants in this class, other mechanisms, such as simple misincorporation, could lead to incorrect terminal bases on the strong-stop DNA.) After strand transfer, these terminal bases could not pair with the minus-strand DNA and so could not be extended as the growing point at the correct position; they then annealed to a position farther 5′

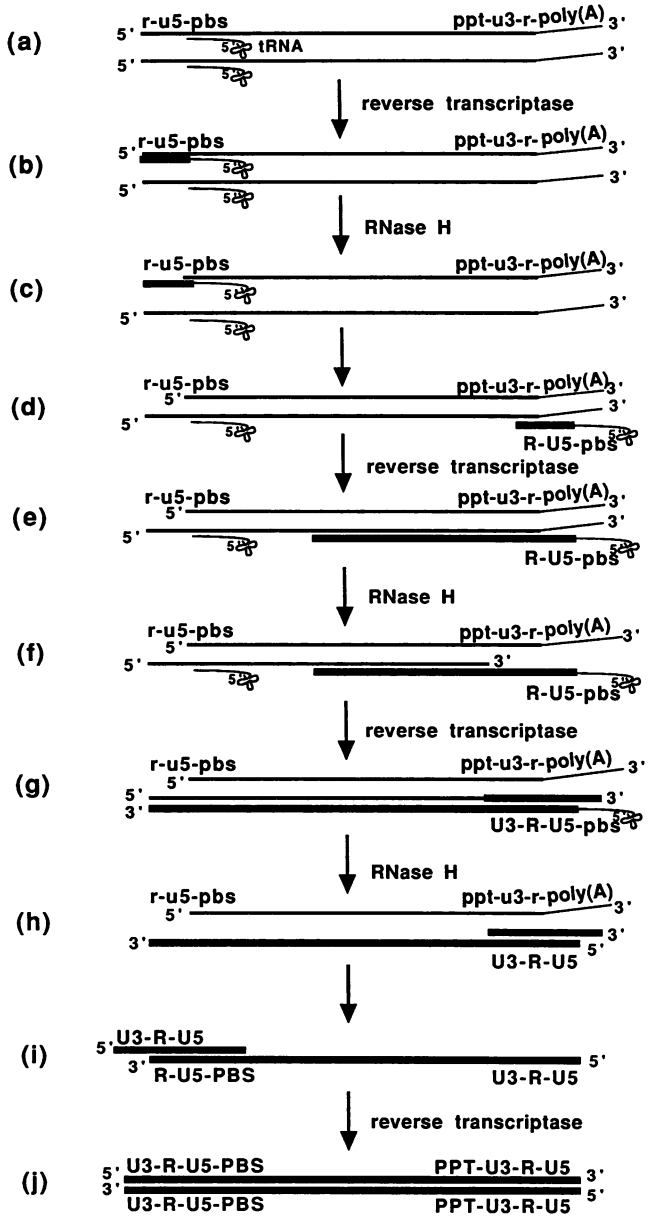

FIG. 7. Reverse transcription process. Thin lines represent RNA, thick lines represent DNA. (a) Starting products. (b) Minus-strand strong-stop DNA synthesis by reverse transcriptase. (c) Removal of the bound RNA from the minus-strand strong-stop DNA by RNase H. (d) Minus-strand strong-stop DNA transfer (intermolecular transfer is shown). (e) Minus-strand DNA synthesis by reverse transcriptase. (f) Nicking of RNA at the polypurine tract (ppt) and removal of 3′ sequence by RNase H. (g) Minus-strand DNA synthesis and plus-strand strong-stop synthesis by reverse transcriptase. (h) Removal of RNA by RNase H. (i) Plus-strand strong-stop DNA transfer. (j) Completion of minus-strand and plus-strand DNA syntheses by reverse transcriptase.

in the minus-strand DNA and continued synthesis at this point, deleting the nucleotides between and resulting in the observed sequence. Presumably, the incorrect nucleotides on the strong-stop DNA were complementary to the site to which they annealed in the minus strand and thus determined the 3′ deletion boundary.

Insertion of tRNA sequences into retroviral DNA has been observed in a slightly different context. Colicelli and Goff (2) reported sequences at the junction of the LTRs in two-LTR circles. They observed a much longer portion of the tRNA inserted into this region. They also observed that the reverse transcriptase was able to copy correctly certain unusual bases (Ψ, dihydro-U, methyl-G, and methyl-C) into their correct complementary bases (A, A, C, and G, respectively), which we observed as well. However, the two types of insertions actually have their origins in different steps of the reverse transcription process. The errors we observed arose from incorrect termination of synthesis of the plus-strand strong-stop DNA, while those reported by Colicelli and Goff (2) in two-LTR circle junctions arose from the failure of the RNase H to remove the unpaired tRNA sequences from the end of the DNA molecule. More recently, Olsen et al. (17) have reported the incorporation of tRNA sequences into circular viral DNA of Rous sarcoma virus. They also present other sequences from a class of mutants, deletions extending from the U5 region of tandem LTRs, which are very similar to the ones we observed in our PBS and pre-PBS classes. These could have arisen from the mechanisms we propose, with the deleted viral DNA undergoing blunt-end ligation to form the two-LTR circles studied by Olsen et al. (17) or integrating normally to form the proviruses studied by us.

The pre-PBS termination class arose from a mechanism similar to that of the PBS termination class (Fig. 6c), except that the misincorporation and premature termination occurred earlier, before any of the PBS sequences were copied from the tRNA primer. This mechanism is distinct from the PBS mechanism in that the misincorporation took place during the copying of a DNA template, rather than an RNA template. The result of this abnormal synthesis was a plus-strand strong-stop DNA product that had improper bases at the 3′ end but lacked any PBS sequences. For there to be accurate jumping of the plus-strand strong-stop DNA, homology must exist between the plus-strand strong-stop DNA and the 3′ end of the minus-strand DNA. Such homology exists when the minus-strand strong-stop DNA starts on only one strand of the RNA and its transfer is intermolecular; that is, when there is asynchronous synthesis on the two strands (11, 19). In such a situation, following interstrand transfer, the minus-strand viral DNA would have intact R and U3 regions at its 3′ end, and so the plus-strand strong-stop DNA molecule could have paired with these and then have proceeded as in the model presented in Fig. 6c.

An alternative model for the generation of the pre-PBS termination class deletions would not require homology between the ends of the plus-strand strong-stop DNA and the minus-strand DNA. Without homology, the strand transfer step would be unable to proceed correctly. The 3′ bases of the plus-strand strong-stop DNA instead would bind to a homologous position either in the vector or on another RNA molecule, and there extend the sequence. The result would be the deletion of that portion of the sequence between the position at which the plus-strand strong-stop DNA terminated and the point at which it randomly began synthesizing the plus-strand DNA. However, we consider this mechanism unlikely; since the ends of the two viral DNA strands would not be annealed, this mechanism would require the postulation of a rather unlikely series of steps to create a viral DNA capable of integration into the cell genome. O'Rear and Temin (18) previously suggested such an erroneous termination and transfer of the plus-strand strong-stop DNA; however, they observed a different outcome (inser-

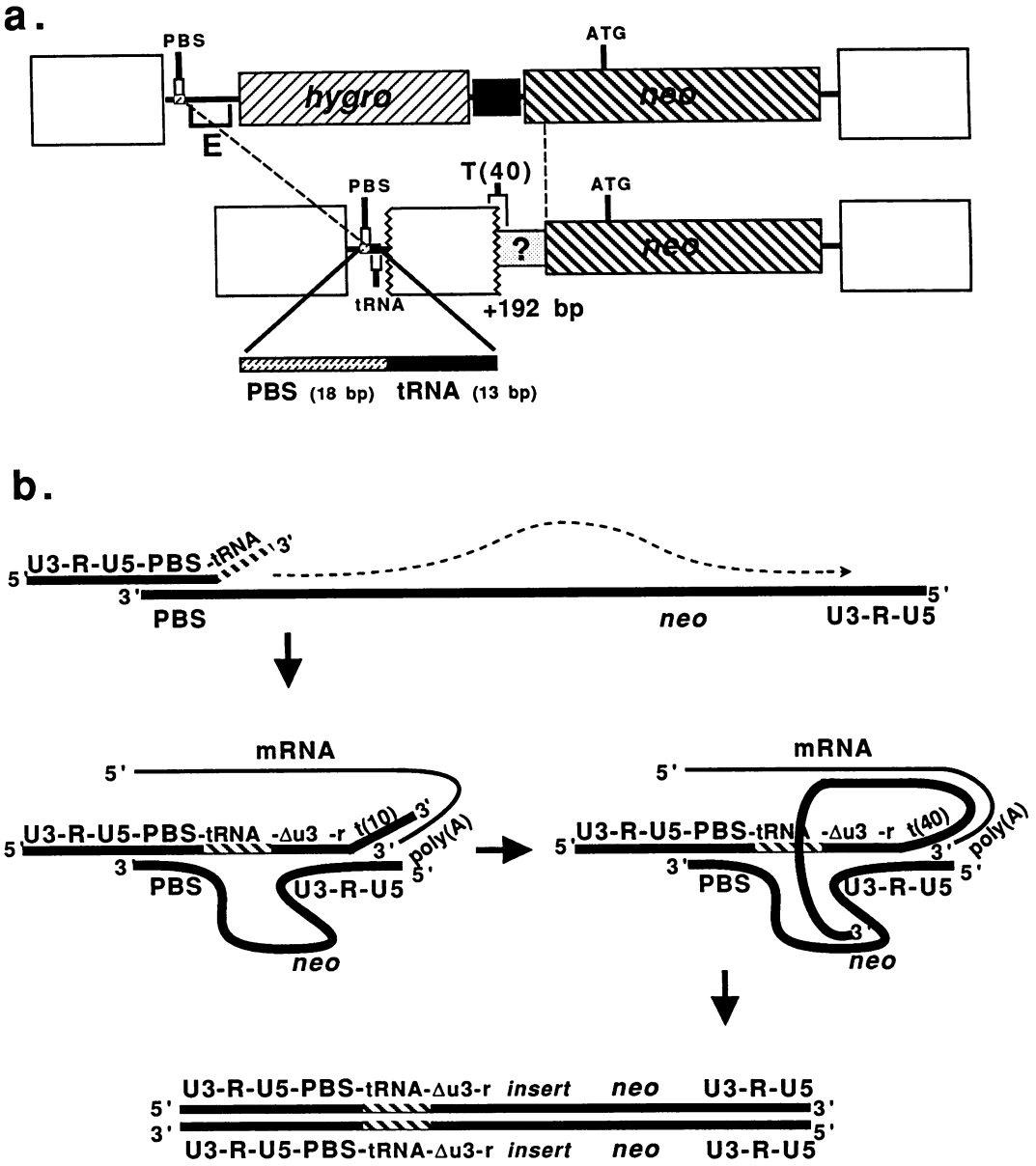

FIG. 8. (a) Structure of mutant 3A.1. Symbols are as defined in the legend to Fig. 1. The box with ragged edges represents a partial duplication of the LTR; the gray box with the question mark represents a 192-bp inserted cellular sequence. (b) Proposed mechanism for generating mutant 3A.1. Thick lines represent DNA; thin lines represent RNA. The plus-strand strong-stop DNA product, which contained 14 extra bases of tRNA sequence at the 3′ end (hatched lines), transferred correctly to the opposite end of the minus-strand DNA. The unpaired bases at the 3′ end then annealed near the beginning of the U3 region in the 3′ LTR, and synthesis continued at this point. After copying U3 and part of R, the reverse transcriptase growing point switched strands to a copackaged canine mRNA. This switch was presumably facilitated by the binding of a section of 10 T residues in the R region to a poly(A) region in the mRNA. The mRNA was copied for 192 bases, and then the growing point switched back to the minus-strand viral DNA, near the beginning of the *neo* gene. The rest of the viral DNA was copied normally, and after resolution, the final observed provirus was produced.

tion instead of deletion), and they postulated a second plus-strand strong-stop DNA transfer event that yielded correct, double-stranded DNA ends.

Another unlikely alternative model would involve a prematurely terminated minus-strand viral DNA product, created perhaps by a break in the viral RNA. The plus-strand strong-stop DNA transfer would be unable to transfer to the (nonexistent) PBS on the minus-strand DNA and would instead transfer to the random end of minus-strand DNA, potentially losing some bases at the 3′ end to an endonuclease. We consider this mechanism unlikely since the high rate of retrovirus recombination tends to repair RNA strand breaks (11), and random transfer to an incomplete minus-strand DNA is unlikely in a process as ordered as reverse transcription.

Both the misalignment model and the incorrect plus-strand strong-stop termination models have in common a requirement for sequence homology at the deletion junctions. All the mutants (with the exception of clone 4.18) either had a recognizable overlap of at least 1 bp at the junction or

contained inserted sequences of unknown origin which were presumably incorporated because they had homology to the deletion endpoints.

Another finding from this work was the incorporation of foreign DNA sequences into the proviruses. Clone 5.14 contained a long insert (489 bp) of a sequence of unknown origin (Fig. 5a). Because this sequence was a long open reading frame, we believe it arose from an mRNA sequence. Clone 3A.1 also contained an insert of 192 bp which was very similar to the mRNA for canine blood-clotting factor IX. Perhaps mRNAs from the parental canine helper cells were encapsidated at some low frequency (4, 9, 12), or one of the viral RNAs contained readthrough sequences from cell DNA (1). In either case, the unpaired bases at the end of the plus-strand strong-stop DNA could then have paired with one of these RNA sequences and extended it for some length before again transferring back to the original viral DNA molecule (see also Fig. 8).

In addition to this incorporation of foreign DNA, clone 3A.1 had other interesting features (Fig. 5b). Clone 3A.1 contained 14 bp of tRNA sequence at the end of the PBS sequence, similar to clone 2.2; it also contained a partial duplication of LTR sequences (505 bp) and 192 bp of inserted DNA of nonviral origin (Fig. 8a). We believe that this mutant arose by the mechanism shown in Fig. 8b. As in the creation of clone 2.2, plus-strand strong-stop DNA synthesis did not stop at the end of PBS, and 14 bases of tRNA were copied. After normal strand transfer, the unpaired tRNA sequences at the end of the plus-strand strong-stop DNA annealed near the 5′ end of U3 in the 3′ LTR. Synthesis continued at this point, causing the insertion of the partial LTR. Then, in a region of 10 consecutive T residues in the R region, the reverse transcriptase growing point switched templates to a copackaged mRNA; this switch was presumably facilitated by the binding of the poly(T) stretch to a poly(A) region on the mRNA. The reverse transcriptase continued copying this mRNA for 192 bases and then switched templates again, this time back to the original minus-strand viral DNA, just upstream of the *neo* gene. The synthesis was then finished to the end of the DNA in normal fashion. The resulting product has the final structure observed for clone 3A.1.

ACKNOWLEDGMENTS

We thank Rebecca Wisniewski, Bonnie Fritz, Tim Jacoby, Barbie Pietz, and Kevin Krebsbach for technical assistance; W. Sugden and A. Panganiban for critical reading of the manuscript; and W.-S. Hu, L. Mansky, V. Pathak, and S. Yang for discussion and helpful comments.

This work was supported by grants CA-07175 and CA-22443 from the U.S. Public Health Service. H.M.T. is an American Cancer Society research professor. G.P. was supported by training grant CA-09135 from the National Institutes of Health.

REFERENCES

1. **Coffin, J. M.** 1990. Retroviridae and their replication, p. 1437–1500. *In* B. N. Fields, D. M. Knipe, et al. (ed.), Virology, 2nd ed., vol. 1. Raven Press, New York.
2. **Colicelli, J., and S. P. Goff.** 1986. Structure of a cloned circular retroviral DNA containing a tRNA sequence between the terminal repeats. J. Virol. **57**:674–677.
3. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–396.
4. **Dornburg, R., and H. M. Temin.** 1988. Retroviral vector system for the study of cDNA gene formation. Mol. Cell. Biol. **8**:2328–2334.
5. **Dougherty, J. P., and H. M. Temin.** 1986. High mutation rate of a spleen necrosis virus-based retrovirus vector. Mol. Cell. Biol. **6**:4387–4395.
6. **DuBose, R. F., and D. L. Hartl.** 1990. Rapid purification of PCR products for DNA sequencing using Sepharose CL-6B spin columns. Biotechniques **8**:271–274.
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. **Erlich, H. A.** 1989. PCR technology. Stockton Press, New York.
9. **Gallis, B., M. Linial, and R. Eisenman.** 1979. An avian oncovirus mutant deficient in genomic RNA: characterization of the packaged RNA as a cellular messenger RNA. Virology **94**:146–161.
10. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–369.
11. **Hu, W.-S., and H. M. Temin.** 1990. Retrovirus recombination and reverse transcription. Science **250**:1227–1233.
12. **Ikawa, Y., J. Ross, and P. Leder.** 1974. An association between globin messenger RNA and 60S RNA derived from Friend leukemia virus. Proc. Natl. Acad. Sci. USA **71**:1154–1158.
13. **Kawai, S., and M. Nishizawa.** 1984. New procedure for DNA transfection with polycation and dimethyl sulfoxide. Mol. Cell. Biol. **4**:1172–1174.
14. **Kusukawa, N., T. Uemori, K. Asada, and I. Kato.** 1990. Rapid and reliable protocol for direct sequencing of material amplified by the polymerase chain reaction. Biotechniques **9**:66–72.
15. **Maniatis, T., E. F. Fritsch, and J. Sambrook (ed.).** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Marck, C.** 1988. 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16**:1829–1836.
17. **Olsen, J. C., C. Bova-Hill, D. P. Grandgenett, T. P. Quinn, J. P. Manfredi, and R. Swanstrom.** 1990. Rearrangements in unintegrated retroviral DNA are complex and are the result of multiple genetic determinants. J. Virol. **64**:5475–5484.
18. **O'Rear, J. J., and H. M. Temin.** 1982. Spontaneous changes in nucleotide sequence in proviruses of spleen necrosis virus, an avian retrovirus. Proc. Natl. Acad. Sci. USA **79**:1230–1234.
19. **Panganiban, A. T., and D. Fiore.** 1988. Ordered interstrand and intrastrand DNA transfer during reverse transcription. Science **241**:1064–1069.
20. **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.
21. **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.
22. **Roberts, J. D., B. D. Preston, L. A. Johnston, A. Soni, L. A. Loeb, and T. A. Kunkel.** 1989. Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis in vitro. Mol. Cell. Biol. **9**:469–476.
23. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
24. **Temin, H. M.** 1989. Evolution of retroviruses, p. 189–196. *In* A. L. Notkins and M. B. A. Oldstone (ed.), Concepts in viral pathogenesis III. Springer-Verlag, New York.
25. **Varmus, H., and P. Brown.** 1989. Retroviruses, p. 53–108. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
26. **Watanabe, S., and H. M. Temin.** 1982. Encapsidation sequences for spleen necrosis virus, an avian retrovirus, are between the 5′ long terminal repeat and the start of the *gag* gene. Proc. Natl. Acad. Sci. USA **79**:5986–5990.
27. **Watanabe, S., and H. M. Temin.** 1983. Construction of a helper cell line for avian reticuloendotheliosis virus cloning vectors. Mol. Cell. Biol. **3**:2241–2249.