JOURNAL OF VIROLOGY, July 1993, p. 3845–3853
0022-538X/93/073845-09$02.00/0

Vol. 67, No. 7

A Model System for Nonhomologous Recombination between Retroviral and Cellular RNA

ADELINE M. HAJJAR[1,2,3] AND MAXINE L. LINIAL[1,3]*

Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104,[1] and Departments of Comparative Medicine[2] and Microbiology,[3] University of Washington, Seattle, Washington 98195*

Received 19 January 1993/Accepted 25 March 1993

A current model for the generation of transforming retroviruses proposes that read-through RNAs, containing both viral and cellular sequences, are copackaged with viral genomic RNA. It is, however, possible that a cellular mRNA is occasionally encapsidated into a retroviral particle, even though viral packaging sequences are absent. We have generated recombinant proviruses following copackaging of an avian leukosis viral genomic RNA and a *neo*-containing RNA completely devoid of retroviral sequences. In these studies, we used the packaging cell line *SE*21Q1b, which has the unique ability to randomly package cellular mRNA into retroviral particles. We describe 10 recombinants obtained following copackaging of nonhomologous RNAs. Our data show that recombination is not occurring at the DNA level in the parental *SE*21Q1b cells but is occurring at the RNA level, during reverse transcription. These data further suggest that reverse transcriptase can preferentially jump between templates at short stretches of homology in otherwise unrelated RNAs. We conclude that retroviral sequences are not required for packaged mRNA to be reverse transcribed and to be included in integrated proviruses.

Acutely transforming retroviruses arise following recombination between retroviral sequences and cellular protooncogenes (c-*onc*). The mechanism of nonhomologous recombination that generates transducing retroviruses has been hypothesized to involve recombination during reverse transcription, following copackaging of a viral genomic RNA and a c-*onc* mRNA (7, 8, 12), generating a recombinant provirus. Viral oncogenes (v-*onc*), unlike their cellular counterparts, generally lack introns (12, 23, 26, 45), although 5′ recombination junctions occasionally occur within an intron (26, 43). Some v-*onc*s also contain short stretches of A's at their 3′ junctions (22–24). These features support the role of RNA as an intermediate during the acquisition of a c-*onc* by a retrovirus. Rous sarcoma virus (RSV) is the only example of an acutely transforming virus that is replication competent, since transduction of c-*src* sequences occurred without the deletion of viral sequences. Sequence analysis of RSV suggested a model for its generation in which both RNA recombination (during reverse transcription) and recombination at the DNA level were required (43, 44). It is proposed that retroviral sequences including a long terminal repeat (LTR) first integrated upstream of c-*src* sequences, which allowed transcription of a read-through *src* RNA containing viral packaging sequences. Such an RNA could then be copackaged with a viral genomic RNA, allowing recombination during reverse transcription (40). Since the models for reverse transcription propose that reverse transcriptase (RT) jumps twice during proviral DNA synthesis, it is hypothesized that RT can jump at other times between the RNA templates, for example, when the RNA strand is broken (forced copy-choice model of recombination [7, 8]). However, attempts to induce recombination by generation of RNA breaks with gamma radiation have not proven successful (21). Template switching, for whatever reason, could result in RT jumping to a copackaged *onc* RNA, giving rise to a transforming virus. Homologous recombination in retroviruses, also proposed to occur during reverse transcription (15, 19, 20, 40), occurs at a very high frequency, with markers as close as 1 kb apart segregating independently, suggesting that RT jumps frequently between copackaged genomic RNA (29). In the case of nonhomologous recombination, it is not known whether certain sequences or short stretches of homology promote RT jumps, since examples of recombination breakpoints at areas of either patchy homology (11, 23, 45) or no homology (1, 26, 38, 43) are found in naturally occurring transforming viruses.

Several systems have been described to study nonhomologous recombination during reverse transcription. Coffin and collaborators have used avian leukosis virus (ALV) with a mutation in the polyadenylation sequence and have demonstrated that read-through messages containing ALV and downstream neomycin phosphotransferase gene (*neo*) sequences can give rise to recombinant proviruses following reverse transcription (18, 42). In another system which has been used to study nonhomologous recombination, retroviral packaging sequences have been included at the 5′ end of the nonhomologous message and *env* sequences have been included at the 3′ end of the selectable marker (41). However, this introduces a short stretch of homology between the nonviral message and the retroviral genomic RNA. Zhang and Temin have described a system that generates recombinants after a single cycle of retroviral replication (46). Retroviral sequences are present on both of the packaged RNAs, providing short stretches of homology, although the recombination events scored are at nonhomologous regions. In other studies, the regeneration of Harvey sarcoma viruses following infection of *ras*-transformed cells containing truncated Harvey sarcoma virus with Moloney murine leukemia virus has been examined (13, 14, 16). Again, the constructs used contained retroviral sequences. More recently, studies of circular proviral DNA intermediates suggest that nonhomologous recombination might occur during reverse transcription. In these studies, cellular RNA

* Corresponding author.

found in the cytoplasm of the infected cell was implicated as a recombination substrate (10). However, copackaging was not rigorously ruled out.

In this study, we describe a system that allows copackaging of a viral genomic RNA and an RNA containing *neo* which is completely devoid of retroviral sequences. This system utilizes the quail packaging cell line *SE*21Q1b, which randomly packages cellular mRNAs into virions (3, 31). *SE*21Q1b virions have been shown to mediate retrofection, the transfer of a cellular gene by retroviral infection in the absence of retroviral genes (27, 28). In those studies, *SE*21Q1b virions were shown to be capable of reverse transcribing and integrating *neo* RNA transcribed from pRSVneo in the absence of helper viruses. In order for cells containing the *neo* gene to be G418 resistant (G418^r), the *neo* RNA had to contain sequences encoding a functional promoter (27). The *neo*-containing plasmid used in this work (pCMVneo) contains a single strong promoter upstream of the cap site. Therefore, the RNA transcripts do not contain promoter sequences. Thus, copackaging of a *neo* RNA and a genomic RNA is necessary to lead to the generation of a provirus which contains a viral LTR driving expression of *neo*. In the absence of a helper virus, transmission of the G418^r phenotype does not occur. We have been unable to detect preexisting recombinant proviruses in the DNA of the parental *SE*21Q1b cells. Thus, it is most likely that recombination occurred after infection of recipient QT35 cells, during reverse transcription. From the analysis of 10 5′ recombination junctions, we propose a model for the generation of recombinant proviruses following copackaging of a viral genomic RNA and a *neo* RNA.

MATERIALS AND METHODS

Cell culture. *SE*21Q1b (31) and QT35 (34) cells were grown in GM+D+Ck (Ham's F10 medium containing 5% tryptose phosphate broth [Difco], 5% calf serum, 1% heat-inactivated chicken serum, and 1% dimethyl sulfoxide) as previously described (2). Twenty micrograms of plasmid DNA was transfected into appropriate cells by using the modified calcium phosphate procedure of Chen and Okayama (4) as described elsewhere (2). To select for pools and single cell clones, 0.15 mg of G418 (Sigma) per ml in CM (Ham's F10 medium containing 12% calf serum, 4% heat-inactivated chicken serum, 1% nonessential vitamins [GIBCO], and 10% tryptose phosphate broth) was used. Colonies were isolated by using cloning cylinders (Bellco). Viral supernatants were harvested from cell clones established from individual G418^r *SE*21Q1b cells every 24 h and filtered through 0.45-μm-pore-size Millipore filters. QT35 cells (5×10^5) were infected in triplicate with either 0.5 or 1.0 ml of viral supernatant in GM+D+Ck containing 2 μg of Polybrene (Sigma) per ml.

Plasmids. DNA manipulations were performed by standard techniques (33). pCMVneo (28) and p882-29 (39) have been previously described. pRCAS-BP is a derivative of pRCAS (25) and contains the Bryan high-titer virus polymerase gene in place of pRCAS *pol*. pRCAS-BP and p882-29 were obtained from S. Hughes (Frederick Cancer Center).

DNA and RNA preparation from cells and virions. Viral RNA was extracted from supernatants harvested every 24 h as previously described (2). Total cellular RNA was extracted by the method of Chomczynski and Sacchi (6) as previously described (2). Total cellular DNA was extracted following separation of nuclei by standard methods (33).

DNA hybridization and sequence analysis. For Southern blots, 10 μg of cell DNA was digested with appropriate restriction endonuclease and electrophoresed through a 0.8% agarose gel. Blotting and hybridizations were performed according to standard procedures (33). The *neo* probe consisted of a 900-bp pCMVneo *Hin*dIII-*Sph*I fragment, and the LTR probe was a 210-bp pRCAS-BP *Bst*EII-*Pvu*I fragment. These probes were nick translated according to standard procedures (33). The dideoxy-chain termination method of DNA sequencing (37) was performed on plasmid DNA obtained from the polymerase chain reaction (PCR)-generated clones, using the Sequenase kit (U.S. Biochemical Corp.).

PCR. Oligonucleotide primers were designed with added restriction sites for cloning. The RSV U3 sense primer (5′-GCGGTACCGTATCAGACGGGTCTAAC-3′) contains a *Kpn*I site (underlined), and the antisense *neo* primer (5′-GCTCTAGAACCTGCGTGCAATCCATC-3′) contains an *Xba*I site (underlined). One microgram of cell DNA and 0.5 μg of each oligonucleotide primer were used in 50-μl PCRs. Reaction conditions consisted of 100 μM each deoxynucleoside triphosphate, 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-Cl (pH 8.3), 0.01% (wt/vol) gelatin, 10% glycerol, and 2.5 U of AmpliTaq DNA polymerase (Perkin-Elmer/Cetus). The first denaturation cycle at 94°C for 5 min was followed by 30 cycles of a 1-min denaturation step at 94°C, a 2-min annealing step at either 60 or 65°C, and finally a 5-min extension step at 72°C. Three separate reactions from each DNA template were pooled, and products were isolated by gel electrophoresis. Amplified products were purified from the agarose gels by using the QIAEX kit (Qiagen) according to the manufacturer's protocol. The gel-purified DNA was then digested with *Xba*I and *Kpn*I and subjected to phenol-chloroform and chloroform extractions. The digested DNA was concentrated by precipitation and ligated to the phagemid vector pVZ1 (17) previously digested with *Xba*I and *Kpn*I.

Primer extension. An antisense *neo* 22-mer (5′-GCAGCCC TTGCGCCCTGAGTGC-3′) located 117 nucleotides downstream of the cap site of pCMVneo was used in primer extension experiments with cell RNA as a template as previously described (30).

RESULTS

Generation of recombinant proviruses. *SE*21Q1b cells were transfected with pCMVneo, and four G418^r clones derived from single cells (CN3, CN4, CN6, and CN8) were isolated. Virions produced by these cell clones were shown to contain packaged *neo* RNA by Northern blotting (data not shown). However, when these virions were used to infect QT35 cells, they did not transmit the G418^r phenotype (Table 1) because there was no promoter driving the *neo* gene. The individual CN cell clones were then transfected in parallel with either of two constructs encoding ALV subgroup A replication-competent (helper) viruses, p882-29 (39) and pRCAS-BP (25) (Fig. 1). The cells were passaged for 2 weeks to allow the helper viruses to spread within the cultures. Virions produced by the helper virus-containing CN3, CN4, CN6, and CN8 clones (designated CN RCAS-BP and CN 882-29; Fig. 1) packaged *neo* RNA and were now capable of transmitting the G418^r phenotype to QT35 cells (Table 1). CN3 was found, by Southern blotting, to contain a single copy of integrated pCMVneo and was used for subsequent experiments. The QT35 G418^r M series of clones was obtained following infection of QT35 cells with supernatant from the CN3 RCAS-BP cells, whereas the J series was derived from the CN3 882-29 cells. These single cell-derived secondary

TABLE 1. Transmission of G418 resistance by using filtered supernatants

Cell line	No. of G418-resistant QT35 colonies/ml of viral supernatant	Total ml of supernatant assayed
CN3	<0.1	7.5
CN3 882-29	0.9	9.0
CN3 RCAS-BP	0.4	9.0
CN4	<0.1	9.0
CN4 882-29	0.2	4.5
CN4 RCAS-BP	1.1	4.5
CN6	<0.1	13.5
CN6 882-29	2.2	4.5
CN6 RCAS-BP	1.3	4.5
CN8	<0.1	9.0
CN8 882-29	2.8	4.5
CN8 RCAS-BP	2.7	4.5

clones (M and J) were expanded and used in subsequent studies. Cellular DNAs from the secondary clones were digested with a restriction endonuclease which does not have a recognition site in pCMVneo but does cut in *pol* and *env*. A Southern blot of these DNAs probed with *neo* indicated a single site of *neo* integration in most of these clones (Fig. 2). We detected two bands in M1 cells. Upon continued passage of M1 G418^r cells, a second, faster-migrating band, seen in this Southern blot, increased in intensity (data not shown). We therefore think that this DNA species is derived from a provirus deleted downstream of *neo*. The bands that hybridize to a *neo* probe in the secondary clones also hybridize to an ALV LTR probe (data not shown). CN3 itself does not contain a DNA fragment with both LTR and *neo* sequences (data not shown). This result suggests that recombination had occurred to generate a structure in which an LTR is now driving expression of *neo*. We reasoned that the two most likely sites of recombination were either at the DNA level in the CN3 cells containing

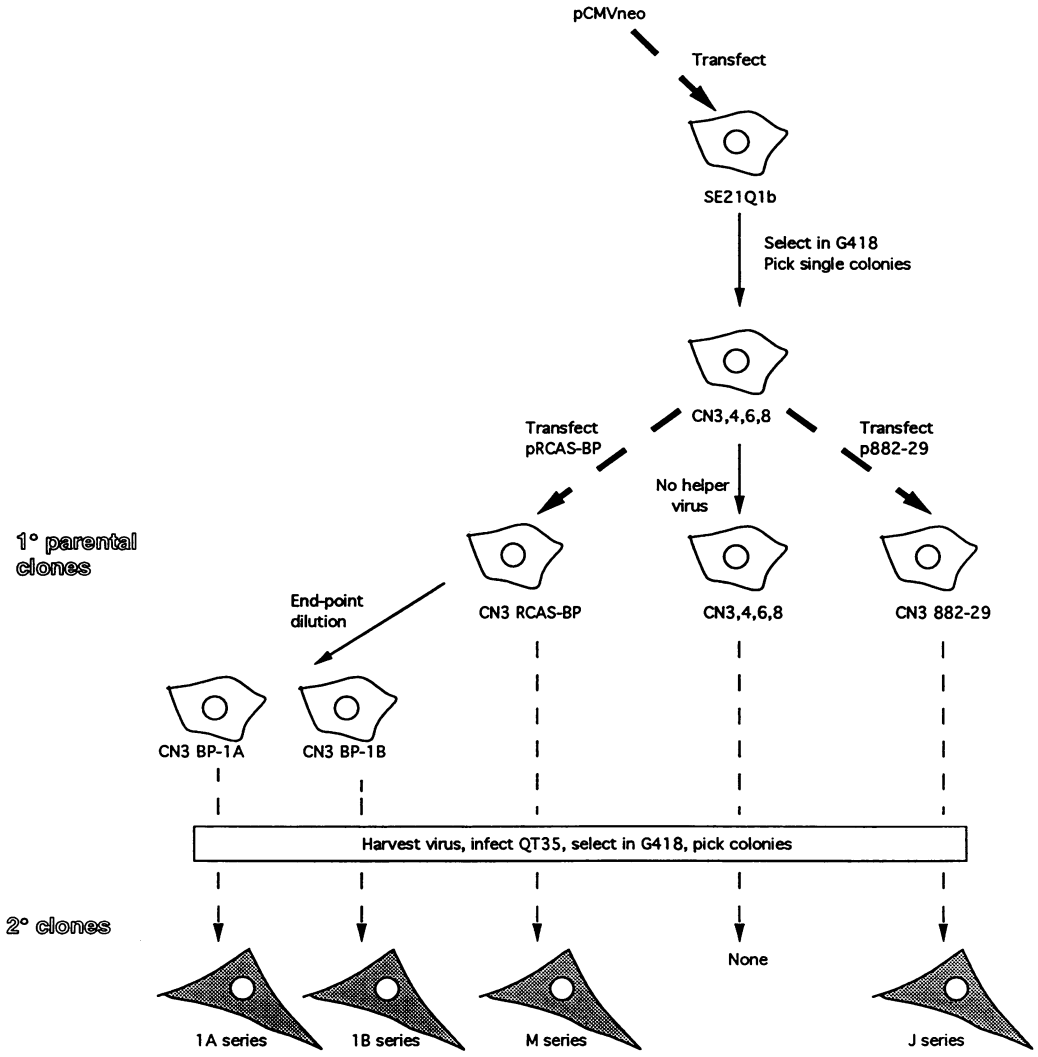


FIG. 1. System used to generate recombinant proviruses. The clear cells (CN, CN RCAS-BP, and CN 882-29) represent *SE*21Q1b clones. The stippled cells (1A, 1B, M, and J series) represent QT35 clones. Thick dashed lines indicate transfection of 20 μg of indicated plasmid; thin lines indicate passaging of cells in culture to obtain single cell clones; dashed lines indicate use of filtered supernatants to infect QT35 cells.

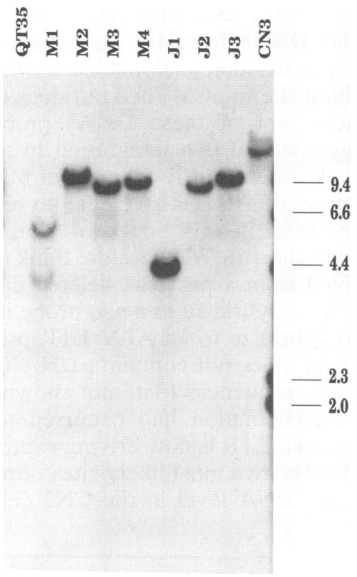


FIG. 2. Southern analysis of cell DNA digested with *Xba*I and probed with *neo*. Each indicated cell DNA (10 μg) was digested with *Xba*I according to the manufacturer's directions, electrophoresed through 0.8% agarose, and transferred to nitrocellulose. The *neo* probe was used at a specific activity of 1.5×10^6 cpm/ml of hybridization fluid. *Xba*I recognition sites in ALV are at nucleotides 3370 (*pol*) and 5670 (*env*). λ DNA digested with *Hin*dIII was used as a molecular weight marker, and the fragment sizes in kilobase pairs are indicated.

helper viruses or at the RNA level following infection of QT35 cells, during reverse transcription of copackaged *neo* and viral genomic RNA.

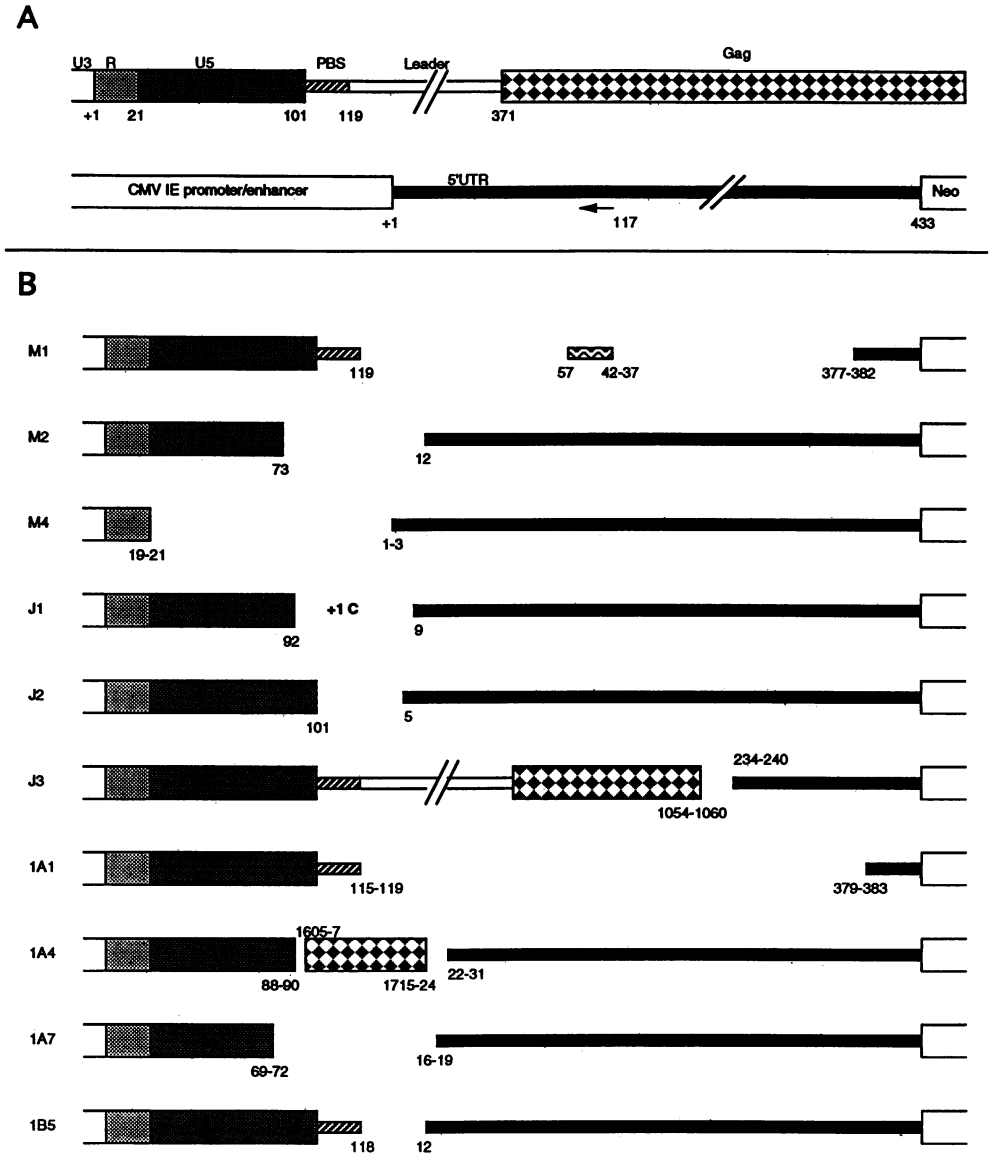
Recombination did not occur between DNA molecules in *SE*21Q1b parental cells. Supernatants from the ALV-infected CN3 cells produced only a low level of secondary G418^r colonies (<3 colonies per ml) on QT35 cells. Since we had transfected the CN3 cells with constructs encoding helper viruses, it was possible that in a small subset of the resulting pool of cells, the helper virus had integrated upstream of *neo*, creating a chimeric LTR-*neo* RNA. Such rare recombinants would not be detected by Southern blot hybridization, but such rare cells could be responsible for the transfer of the *neo* gene. To determine whether such DNA recombination had occurred in CN3 cells after transfection, we diluted the CN3 RCAS-BP cells and obtained five clones derived from single cells (designated CN3 BP-1A, CN3 BP-1B, CN3 BP-1C, CN3 BP-2, and CN3 BP-3; Table 2). Supernatants from these clones were used to test transmission of G418 resistance on QT35 cells. All five subclones were able to transmit the G418^r phenotype at equivalent frequencies (Table 2). This finding indicates that most cells within the CN3 RCAS-BP pool are capable of generating secondary G418^r QT35 clones. Furthermore, the restriction maps of secondary QT35 clones obtained from the same *SE*21Q1b CN3 BP subclone were distinct (data not shown), indicating that a separate recombination event generated each secondary clone. Southern blots of the CN3 BP subclones did not reveal a fragment that hybridized with both LTR and *neo* probes (data not shown). Therefore, the LTR-*neo* recombination event was unlikely to have occurred in the CN3 cells, following integration of an LTR just upstream of the *neo* gene. Instead, the most likely site of recombination between the ALV LTR and *neo* is during reverse transcription of copackaged RNAs.

TABLE 2. Analysis of single cell clones obtained from the CN3 RCAS-BP pool for the ability to transmit G418 resistance to QT35 cells

Cell line	No. of G418-resistant QT35 colonies/ml of supernatant	Total ml of supernatant assayed
CN3 RCAS-BP	0.4	9.0
CN3 BP-1A	2.4	4.5
CN3 BP-1B	1.3	6.0
CN3 BP-1C	2.7	4.5
CN3 BP-2	2.0	4.5
CN3 BP-3	1.6	4.5

Analysis of 5′ recombination junctions. To elucidate the mechanism of recombination in our system, we sequenced recombination junctions. Such information would allow us to determine whether there are certain preferred sites of recombination, such as short stretches of homology in otherwise unrelated sequences, or whether the process is random. We first examined 5′ recombination junctions. Since our data from Southern blots indicated that *neo* sequences were downstream of an LTR, we designed oligonucleotide primers that could amplify the 5′ junctions in PCRs. To clone the 5′ junction between *neo* and viral sequences, we used a sense primer in the U3 region of the LTR and an antisense primer at the 5′ end of the *neo* open reading frame in PCR experiments on cell DNA extracted from the secondary clones. Six of the seven M and J series clones (Fig. 1) yielded amplified products, and these were cloned and sequenced. Cell DNAs from QT35 G418^r cell clones obtained following infection with supernatant from CN3 RCAS-BP subclones 1A and 1B were also amplified in PCRs. We cloned and sequenced amplified products from five such cell clones. No products were amplified from the parental CN3 RCAS-BP cells or the CN3 BP-1A and -1B subclones (data not shown), confirming our conclusion that recombination was not occurring at the DNA level in the *SE*21Q1b cells. Figure 3A depicts the 5′ ends of the parental DNAs used initially in transfections, and Fig. 3B summarizes the structures of the 5′ junctions of 10 recombinant proviruses.

There are several notable features of these clones. (i) Although the *neo* RNA transcribed from pCMVneo contains a 432-base untranslated region (black line in Fig. 3A), 7 of 10 of the secondary clones (M2, M4, J1, J2, 1A4, 1A7, and 1B5) have a recombination junction within the first 30 nucleotides of the RNA. (ii) Two clones contain insertions between the LTR and *neo*. M1 contains 15 nucleotides of antisense Trp tRNA sequences. 1A4 contains *gag* sequences inserted at the 5′ junction, indicating that two recombination events had occurred to generate this 5′ junction. (iii) In 4 of 11 recombination junctions (M4, 1A1, 1A4, and 1A7), there is a short stretch of homology at the breakpoint (two to four bases). 1A4 has two junctions, one between LTR and *gag* and one between *gag* and *neo*, both of which contain short homologies. (iv) Three of the eleven recombination junctions have longer sequence identities at the breakpoint. M1 has six identical nucleotides between the *neo* and Trp tRNA sequences. J3 has a stretch of 7 identical nucleotides at the breakpoint, and 1A4 has a stretch of 9 identical nucleotides at the breakpoint, with an additional 50% sequence identity shared between the parental sequences over the 12 nucleotides upstream of the breakpoint. The alignments are



C

```
Neo   TGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGG
        ** *   ***   *   **    ***************
M1    CCGACGTGATTCGAACACGCAGCCTTCTGATGGCGCAGGG
      ******************************   *

αtrp  CCGACGTGATTCGAACACGCAGCCTTCTGATCTGGAGTCA
tRNA

Neo   AGAGAAAGCAGGTAGCTTGCAGTGGGCTTACAT
        *      *** ********************
J3    GCAAGGGTAAGGGAGCTTGCAGTGGGCTTACAT
      ********************          *
ALV   GCAAGGGTAAGGGAGCTTGCGAGTACTGGTCCG
                      /\
                     AGG

Neo   CCTGGAGACGCCATCCACGCTGTTTTGACCTCCAT
        * **    * ** **********************
1A4   ACGGGCAGCACCCTCCACGCTGTTTTGACCTCCAT
      **********************      **    *
ALV   ACGGGCAGCACCCTCCACGCTGACCACCCCAGGAG
```

FIG. 3. 5′ recombination junctions derived from sequencing of PCR-generated fragments. (A) Parental DNAs used. At the top is ALV; at the bottom is pCMVneo. +1 indicates the start of transcription. Numbers correspond to the nucleotide numbers of the RNA. Symbols: ▩, R; ■, U5; ▨, PBS; ⧓, *gag* sequences; ■, 5′-untranslated region (5′UTR) of *neo*. The U3 sense oligonucleotide used in PCR is located at −90, and the *neo* antisense oligonucleotide is at +460, just downstream of the neo ATG (not shown). The arrow indicates the site of annealing of the *neo* primer used for primer extension analysis in Fig. 4. CMVIE, cytomegalovirus immediate-early. (B) Recombination junctions determined for recombinant proviruses. Numbers indicate the nucleotides at the recombination junction. Ranges of numbers indicate that the parental sequences are identical at the breakpoint. +1C indicates an extra C nucleotide present at the 5′ breakpoint in J1. ⧓, antisense Trp tRNA sequences. (C) Sequence alignment of M1, J3, and 1A4 virus-*neo* 5′ junctions. The middle lines show the sequences of the recombinant proviruses. The top lines are the parental *neo* sequences, and the bottom lines are the parental viral sequences for J3 and 1A4 or the antisense Trp tRNA (αtrp RNA) sequence for M1. Asterisks indicate identical nucleotides between recombinant and parental sequences. Underlined nucleotides in the Trp tRNA sequence hybridize to the PBS. The caret indicates three nucleotides deleted in J3 but present in ALV.

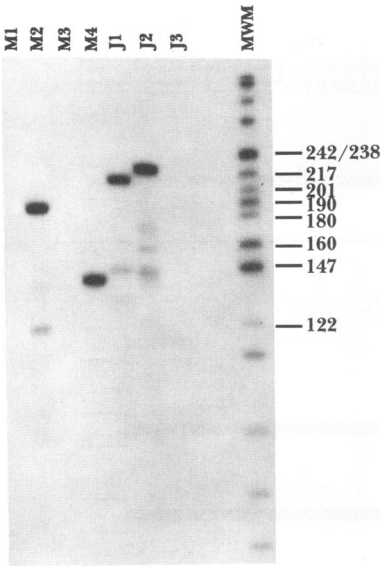


FIG. 4. Primer extension analysis of cell RNA. A *neo* antisense primer (see Fig. 3A, arrow) with the 3′-end 95 nucleotides downstream of the pCMVneo cap site, was used in this experiment. The shorter bands in the M2, J1, and J2 lanes are probably strong stops. MWM, *Msp*I-digested pBR322 molecular weight marker. Sizes are indicated in bases.

shown in Fig. 3C. (v) Six of the ten clones lack the primer binding site (PBS). The three clones (M1, 1A1, and 1B5) with breakpoints at the PBS contain virtually the entire PBS. (The 10th clone, J3, has the entire PBS, since the breakpoint occurred downstream of the PBS, in *gag*.) (vi) Two clones (M2 and 1B5) utilized the same breakpoint in *neo*. In addition, we characterized one other recombinant with unusual features (acquisition of 7SL sequences [5]), which will be reported separately. We have preliminary evidence from Northern (RNA) and Southern blots (data not shown) that 80% of the clones contain 3′ LTR sequences.

Primer extension analysis confirms the PCR data. To exclude the possibility that artifacts such as the adventitious joining of products from different molecules had occurred during PCR amplification, we used an independent approach to confirm the validity of the PCR-generated 5′ recombination junctions. We determined the 5′ end of the *neo* RNA transcribed in the secondary clones by performing primer extension analysis using a 5′ antisense *neo* primer on total cell RNA extracted from the secondary G418^r QT35 clones. The site of the primer is denoted by an arrow in Fig. 3A. This primer yields a 117-base extension product with use of pCMVneo-derived RNA (data not shown). If the first nucleotide in the retroviral R region denotes the start of transcription in the recombinant *neo*-containing proviruses sequenced, the predicted extension products, using this primer with M2, M4, J1, and J2 RNAs, are 180, 140, 200, and 210 nucleotides, respectively. Products of these sizes were obtained following primer extension analysis of cell RNAs (Fig. 4). In clones M1 and J3, the *neo* breakpoint occurred downstream of the primer, thus deleting the site for the primer to anneal, and no extension products were detected, as predicted. From these data, we concluded that the junctions obtained by PCR are not artifacts, that the cells are actively transcribing these recombinant proviruses, and that the first nucleotide of R is the start of transcription in many of the G418^r clones.

DISCUSSION

*SE*21Q1b cells were used in these studies because of their unique ability to package cellular messages completely devoid of retroviral sequences (3, 31). They also retain some ability to package retroviral RNA more efficiently than do other mRNAs (3), although the presence of viral RNA does not abrogate packaging of cellular RNAs. This enabled us to copackage a *neo* RNA lacking any retroviral sequences with a viral genomic RNA. In other systems that have been used to study nonviral gene transduction, vectors were designed such that the nonviral RNA copackaged with the viral RNA shared at least a short stretch of homology (14, 16, 41). In another system, virions copackaged vector RNAs, each containing LTR and retroviral packaging sequences (46). One system, which mimics the model for RSV generation, yields a read-through RNA that contains the entire ALV genome with 3′ *neo* sequences (42). The presence of such homologous sequences on both RNAs in all those studies could potentially complicate interpretation of the results, since the actual recombination breakpoints could not be determined if recombination occurred in the homologous regions. Our system allows us to examine how recombination occurs in the absence of any homology, as it is unnecessary to have a packaging sequence on the selectable RNA. While some models of nonhomologous recombination in retroviruses propose that DNA recombination generates transcripts containing viral and cellular sequences that are packaged because of the presence of packaging sequences, it is possible that an *onc* RNA could be packaged rarely even without a packaging sequence. It is such rare events that our system amplifies.

In our system, *neo* RNA is efficiently packaged into *SE*21Q1b virions. However, packaging of *neo* RNA is not sufficient to generate G418^r QT35 colonies, since supernatants from the CN cells did not allow the transfer of G418 resistance. The *neo* RNA transcribed from pCMVneo begins just downstream of the cytomegalovirus promoter, generating a *neo* RNA lacking sequences that could act as a promoter following reverse transcription and integration. Thus, even if the *neo* RNA is reverse transcribed and integrated in QT35 cells, it would have to be integrated just downstream of a cellular promoter in order to be expressed. We have no direct evidence for the rare integration events downstream of cellular promoters in our system; however, clone M3 could represent such an event. Addition of helper viruses to the CN cells provides an LTR promoter through recombination that can then drive expression of *neo*. Therefore, the 5′ recombination junctions are under strong selection for expression in that both the retroviral U3 promoter/enhancer and the coding region for *neo* must be retained.

Based on the 5′ recombination junctions that we have sequenced, one model that requires the fewest number of steps to generate these recombinants during reverse transcription is shown in Fig. 5. We hypothesize that reverse transcription initiates normally at the PBS on the viral genomic RNA (Fig. 5A). Minus-sense strong-stop DNA is made normally and jumps to the 3′ end of the genomic RNA, generating an LTR (Fig. 5B). Minus-sense DNA synthesis then proceeds on the genomic RNA. At some point, a jump occurs onto the *neo* RNA in G418-selected recombinants. Whether this is caused by a break in the RNA, a secondary structure, such as a hairpin where RT may pause, or a region of homology is unknown at this time. In a majority of cases, the *neo* RNA is then reverse transcribed to its 5′ end (7 of the 10 clones had breakpoints within the first 30 nucleotides of

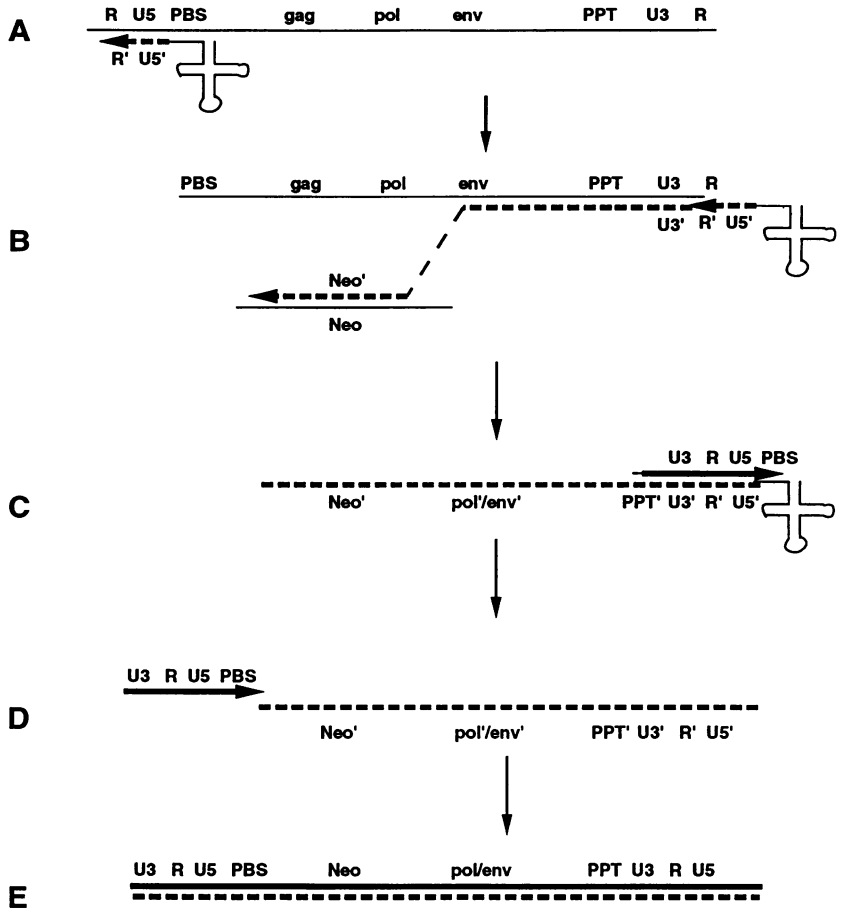


FIG. 5. Proposed model for the generation of recombinant proviruses. Symbols: thin line, RNA; dashed line, minus-sense DNA; thick line, plus-sense DNA. The cloverleaf structure represents the Trp tRNA primer. PPT, polypurine tract. See text for explanation of panels A to E.

the RNA). We propose that plus-sense DNA synthesis initiates normally at the polypurine tract (Fig. 5C), and in cases where the PBS is present without insertions in the secondary clones (1A1 and 1B5), plus-sense strong-stop DNA is made. A jump then occurs to the 3′ end of the minus-sense DNA without any homology (Fig. 5D), although in normal reverse transcription the PBS provides homology for this jump to occur. It is possible that the ends of the minus-sense DNA are held in close proximity, allowing a jump to occur even without any homology due to the juxtaposition of the ends. Darlix and coworkers (9a) have shown that minus-sense strong-stop DNA can still jump with 1% efficiency when there is no homology between the ends (R is deleted) and that this efficiency can be increased by adding one to four nucleotides of homology. The presence or absence of a PBS does not seem to have an effect on the amount of homology at the breakpoint. Once the jump has occurred, plus-sense DNA is synthesized, generating a *neo*-containing recombinant provirus (Fig. 5E). In the secondary clones in which the breakpoint occurs in R or U5 (M2, M4, J1, 1A1, 1A4, and 1A7), and which thus do not contain a PBS, we believe that prematurely terminated plus-sense strong-stop DNA is translocated to the 3′ end of the DNA, as has been previously hypothesized for spleen necrosis virus deletion mutants (36). Premature termination is thought to occur in the case of minus-sense strong-stop DNA in order to generate a primer that can then jump to the 3′ end of the viral RNA (32).

The M1 clone contains Trp tRNA sequences, in the antisense orientation, inserted between the LTR and *neo*, just downstream of the PBS. These sequences are contiguous with the sequences that bind to the PBS. Such insertions have been previously described in cloned circular viral DNA (9, 35). The model proposed for their generation consists of synthesis of a longer than normal plus-sense strong-stop DNA, with the tRNA primer serving as template for the additional sequences (9). Translocation of such a primer to the 3′ end of the minus-strand DNA would result in the apparent insertion of sequences downstream of the PBS, as occurred in M1.

In our studies, several jumps occurred with two to four bases of homology at the breakpoint (M4, 1A4, and 1A7). Only one of the five clones with *neo* breakpoints within the first 12 nucleotides of the *neo* RNA (M4) has any homology at the breakpoint (two bases). The remaining four do not have any homology at the breakpoint, although two (M2 and 1B5) have the same *neo* breakpoint, and thus this site may be a preferred jump site for RT. All five of the remaining clones, with breakpoints downstream of nucleotide 12 (M1, J3, 1A1, 1A4, and 1A7), have three to nine bases of identity at the breakpoint. In clone 1A4, we found *gag* sequences inserted between the LTR and *neo*. There are 9 identical nucleotides

at the *gag-neo* breakpoint with 50% homology over the 12 nucleotides upstream of the breakpoint, perhaps facilitating an RT jump back to the viral RNA. J3 also has a short region of homology (seven identical nucleotides) at the 5′ breakpoint. RT may have jumped back to the viral genomic RNA at this region and reverse transcribed the viral RNA to its 5′ end (PBS, since R and U5 were presumably degraded by the RNase H activity of RT during minus-sense strong stop DNA synthesis). Plus-strand DNA translocation would then have occurred normally in J3, since the PBS sequences would have been present at the 3′ end of the minus-sense DNA. These results suggest that such short stretches of high homology in otherwise nonhomologous RNAs could be preferred sites for RT jumps. An alternative explanation is that recombination occurred following separately transcribed cDNAs from the viral and *neo* RNAs prior to integration. However, there is no evidence for the presence of enzymes in the RT complex which could carry out such reactions. Four of the ten clones (M2, J1, J2, and 1B5) have breakpoints with no homology. It must be emphasized, however, that the 5′ recombination junctions are under strong selection for *neo* expression. Thus, the frequency of 5′ breakpoints with no homology may not reflect recombination breakpoints in general.

The 3′ recombination junctions should not display such selection, other than retention of *neo* coding sequences, since RT could jump from the genomic RNA to the *neo* RNA (Fig. 5B) in a longer region than at the 5′ end. Comparison of the 3′ recombination junctions will allow us to speculate more fully on RT preferences for sites of jumps. In work reported by Zhang and Temin (46), 6 of 10 general-type recombination junctions (3′ junctions occurring upstream of the polypurine tract) were found to contain short stretches of homology. We are currently examining 3′ junctions, and preliminary results indicate that 3′ LTR sequences are present in 8 of the 10 recombinant proviruses as is predicted from our model. Only one of our clones (J3) has intact 5′ viral sequences, including packaging sequences and part of *gag*, similar to acutely transforming viruses. We have found that despite the absence of packaging sequences in the proviruses, several of the recombinants can give rise to tertiary G418^r clones (data not shown). Thus, such clones could represent intermediates in the generation of transmissible viruses that could acquire leader sequences in subsequent rounds of infection and recombination.

ACKNOWLEDGMENTS

We are grateful to Steve Hughes for the ALV plasmids. We thank Kate Levine, Paul Neiman, Julie Overbaugh, and Shuyuarn Yu for critical review of the manuscript.

This work was supported by NCI grant CA18282 to M.L.L. A.M.H. was supported by NIH training grant RR07019 and research scientist development award RR00079.

REFERENCES

1. **Alitalo, K., J. M. Bishop, D. H. Smith, E. Y. Chen, W. W. Colby, and A. D. Levinson.** 1983. Nucleotide sequence of the v-*myc* oncogene of avian retrovirus MC29. Proc. Natl. Acad. Sci. USA **80:**100–104.
2. **Aronoff, R., A. M. Hajjar, and M. L. Linial.** 1993. Avian retroviral RNA encapsidation: reexamination of functional 5′ RNA sequences and the role of nucleocapsid Cys-His motifs. J. Virol. **67:**178–188.
3. **Aronoff, R., and M. Linial.** 1991. Specificity of retroviral RNA packaging. J. Virol. **65:**71–80.
4. **Chen, C., and H. Okayama.** 1987. High-efficiency transformation of mammalian cells by plasmid DNA. Mol. Cell. Biol. **7:**2745–2752.
5. **Chen, P.-J., A. Cywinski, and J. M. Taylor.** 1985. Reverse transcription of 7S L RNA by an avian retrovirus. J. Virol. **54:**278–284.
6. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156–159.
7. **Coffin, J. M.** 1979. Structure, replication and recombination of retrovirus genomes: some unifying hypotheses. J. Gen. Virol. **42:**1–26.
8. **Coffin, J. M.** 1990. Retroviridae and their replication, p. 1437–1500. *In* B. N. Fields and D. M. Knipe (ed.), Fields virology, 2nd ed., vol. 2. Raven Press, New York.
9. **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.

9a. **Darlix, J.-L., et al.** Personal communication.

10. **Dunn, M. M., J. C. Olsen, and R. Swanstrom.** 1992. Characterization of unintegrated retroviral DNA with long terminal repeat-associated cell-derived inserts. J. Virol. **66:**5735–5743.
11. **Felder, M.-P., A. Eychene, J. V. Barnier, I. Calogeraki, G. Calothy, and M. Marx.** 1991. Common mechanism of retrovirus activation and transduction of c-*mil* and c-R*mil* in chicken neuroretina cells infected with Rous-associated virus type 1. J. Virol. **65:**3633–3640.
12. **Goff, S. P., E. Gilboa, O. N. Witte, and D. Baltimore.** 1980. Structure of the Abelson murine leukemia virus genome and the homologous cellular gene: studies with cloned viral DNA. Cell **22:**777–785.
13. **Goldfarb, M. P., and R. A. Weinberg.** 1981. Generation of novel, biologically active Harvey sarcoma viruses via apparent illegitimate recombination. J. Virol. **38:**136–150.
14. **Goodrich, D. W., and P. H. Duesberg.** 1988. Retroviral transduction of oncogenic sequences involves viral DNA instead of RNA. Proc. Natl. Acad. Sci. USA **85:**3733–3737.
15. **Goodrich, D. W., and P. H. Duesberg.** 1990. Retroviral recombination during reverse transcription. Proc. Natl. Acad. Sci. USA **87:**2052–2056.
16. **Goodrich, D. W., and P. H. Duesberg.** 1990. Evidence that retroviral transduction is mediated by DNA, not by RNA. Proc. Natl. Acad. Sci. USA **87:**3604–3608.
17. **Henikoff, S., and M. Eghtedarzadeh.** 1987. Conserved arrangement of nested genes at the Drosophila *Gart* locus. Genetics **117:**711–725.
18. **Herman, S. A., and J. M. Coffin.** 1987. Efficient packaging of readthrough RNA in ALV: implications for oncogene transduction. Science **236:**845–848.
19. **Hu, W.-S., and H. M. Temin.** 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proc. Natl. Acad. Sci. USA **87:**1556–1560.
20. **Hu, W.-S., and H. M. Temin.** 1990. Retroviral recombination and reverse transcription. Science **250:**1227–1233.
21. **Hu, W.-S., and H. M. Temin.** 1992. Effect of gamma radiation on retroviral recombination. J. Virol. **66:**4457–4463.
22. **Huang, C.-C., C. Hammond, and J. M. Bishop.** 1984. Nucleotide sequence of v-*fps* in the PRCII strain of avian sarcoma virus. J. Virol. **50:**125–131.
23. **Huang, C.-C., C. Hammond, and J. M. Bishop.** 1985. Nucleotide sequence and topography of chicken c-*fps*: genesis of a retroviral oncogene encoding a tyrosine-specific protein kinase. J. Mol. Biol. **181:**175–186.
24. **Huang, C.-C., N. Hay, and J. M. Bishop.** 1986. The role of RNA molecules in transduction of the proto-oncogene c-*fps*. Cell **44:**935–940.
25. **Hughes, S. H., J. J. Greenhouse, C. J. Petropoulos, and P. Sutrave.** 1987. Adaptor plasmids simplify the insertion of foreign DNA into helper-independent retroviral vectors. J. Virol. **61:** 3004–3012.
26. **Klempnauer, K.-H., T. J. Gonda, and J. M. Bishop.** 1982. Nucleotide sequence of the retroviral leukemia gene v-*myb* and its cellular progenitor c-*myb*: the architecture of a transduced

oncogene. Cell **31**:453–463.
27. **Levine, K. L., B. Steiner, K. Johnson, R. Aronoff, T. J. Quinton, and M. L. Linial.** 1990. Unusual features of integrated cDNAs generated by infection with genome-free retroviruses. Mol. Cell. Biol. **10**:1891–1900.
28. **Linial, M.** 1987. Creation of a processed pseudogene by retroviral infection. Cell **49**:93–102.
29. **Linial, M., and S. Brown.** 1979. High-frequency recombination within the *gag* gene of Rous sarcoma virus. J. Virol. **31**:257–260.
30. **Linial, M., and M. Groudine.** 1985. Transcription of three c-*myc* exons is enhanced in chicken bursal lymphoma cell lines. Proc. Natl. Acad. Sci. USA **82**:53–57.
31. **Linial, M., E. Medeiros, and W. S. Hayward.** 1978. An avian oncovirus mutant (*SE*21Q1b) deficient in genomic RNA: biological and biochemical characterization. Cell **15**:1371–1381.
32. **Lobel, L. I., and S. P. Goff.** 1985. Reverse transcription of retroviral genomes: mutations in the terminal repeat sequences. J. Virol. **53**:447–455.
33. **Maniatis, T., E. Fritsch, and J. Sambrook.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
34. **Moscovici, C., M. G. Moscovici, H. Jiminez, M. M. C. Lai, M. J. Hayman, and P. K. Vogt.** 1977. Continuous tissue culture cell lines derived from chemically induced tumors of Japanese quail. Cell **11**:95–103.
35. **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.
36. **Pulsinelli, G. A., and H. M. Temin.** 1991. Characterization of large deletions occurring during a single round of retrovirus vector replication: novel deletion mechanism involving errors in strand transfer. J. Virol. **65**:4786–4797.
37. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
38. **Shibuya, M., and H. Hanafusa.** 1982. Nucleotide sequence of Fujinami sarcoma virus: evolutionary relationship of its transforming gene with transforming genes of other sarcoma viruses. Cell **30**:787–795.
39. **Sorge, J., W. Ricci, and S. H. Hughes.** 1983. *cis*-acting RNA packaging locus in the 115-nucleotide direct repeat of Rous sarcoma virus. J. Virol. **48**:667–675.
40. **Stuhlmann, H., and P. Berg.** 1992. Homologous recombination of copackaged retrovirus RNAs during reverse transcription. J. Virol. **66**:2378–2388.
41. **Stuhlmann, H., M. Dieckmann, and P. Berg.** 1990. Transduction of cellular *neo* mRNA by retrovirus-mediated recombination. J. Virol. **64**:5783–5796.
42. **Swain, A., and J. M. Coffin.** 1992. Mechanism of transduction by retroviruses. Science **255**:841–845.
43. **Swanstrom, R., R. C. Parker, H. E. Varmus, and J. M. Bishop.** 1983. Transduction of a cellular oncogene: the genesis of Rous sarcoma virus. Proc. Natl. Acad. Sci. USA **80**:2519–2523.
44. **Takeya, T., and H. Hanafusa.** 1983. Structure and sequence of the cellular gene homologous to the RSV *src* gene and the mechanism for generating the transforming virus. Cell **32**:881–890.
45. **Van Beveren, C., F. van Straaten, T. Curran, R. Muller, and I. M. Verma.** 1983. Analysis of FBJ-MuSV provirus and c-*fos* (mouse) gene reveals that viral and cellular *fos* gene products have different carboxy termini. Cell **32**:1241–1255.
46. **Zhang, J., and H. M. Temin.** 1993. Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. Science **259**:234–238.