Characterization of Unintegrated Retroviral DNA with Long Terminal Repeat-Associated Cell-Derived Inserts

MARTHA M. DUNN,¹ JOHN C. OLSEN,² AND RONALD SWANSTROM^{1,2*}

Department of Biochemistry and Biophysics¹ and UNC Lineberger Comprehensive Cancer Center, CB7295,² University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7295

Received 31 January 1992/Accepted 22 June 1992

We have used a replication-competent shuttle vector based on the genome of Rous sarcoma virus to characterize genomic rearrangements that occur during retrovirus replication. The strategy involved cloning circular DNA that was generated during an acute infection. While analyzing a class of retroviral DNA clones that are greater than full length, we found several clones which had acquired nonviral inserts in positions adjacent to the long terminal repeats (LTRs). There appear to be two distinct mechanisms leading to the incorporation of cellular sequences into these clones. Three of the molecules contain a cell-derived insert at the circle junction site between two LTR units. Two of these molecules appear to be the results of abortive integration attempts, because of which, in each case, one of the LTRs is missing 2 bases at its junction with the cell-derived insert. In the third clone, pNO220, the cellular sequences are flanked by an inappropriately placed copy of the tRNA primer-binding site on one side and a partial copy of the U3 sequence as part of the LTR on the other side. A fourth molecule we characterized, pMD96, has a single LTR with a U5-bounded deletion of viral sequences spanning *gag* and *pol*, with cell-derived sequences inserted at the site of the deletion; its origin may be related mechanistically to pNO220. Sequence analysis indicates that all of the cellular inserts were derived from the cell line used for the acute infection rather than from sequences carried into the cell as part of the virus particle. Northern (RNA) analysis of cellular RNA demonstrated that the cell-derived sequences of two clones, pNO220 and pMD96, were expressed as polyadenylated RNA in uninfected cells. One mechanism for the joining of viral and cellular sequences suggested by the structures of pNO220 and pMD96 is recombination occurring during viral DNA synthesis, with cellular RNA serving as the template for the acquisition of cellular sequences.

The genetic instability of the retroviral genome is the result of processes which contribute to high error rates, frequent homologous recombination, numerous rearrangements, and the ability to recombine with host sequences. Insights into the process of illegitimate recombination between viral and cellular sequences can be inferred from the structures of the recombination products, a subset of which are represented by products of the transduction of cellular oncogenes in the form of acutely transforming retroviruses. For avian retroviruses, the presence of cellular intron sequences in the *v-src* (53, 54), *v-myc* (26), and *v-erbB* (43) oncogenes suggests that one of the recombination events in the transduction of these sequences took place at the DNA level. Experimental systems in which the left half of viral DNA is first fused to a heterologous sequence to mimic a DNA recombination event and then rescued by infection to create a molecule with viral sequences both upstream and now downstream of the heterologous sequence have been developed (13, 14, 16, 17, 48). In this model system, the second recombination event is thought to occur at the RNA level. The presence of a poly(A) tract in the genomes of certain *v-fps*- and *v-erbB*-containing viruses implies that a poly(A)-containing RNA participated in the recombination event generating the 3' ends of these transduced sequences (20, 43).

Other models of illegitimate recombination involve RNA in generating the upstream junction between viral and cellu-

lar sequences. These models are based on (i) transcriptional readthrough to give a greater-than-full-length RNA transcript (19, 49, 50) and (ii) readthrough with splicing to generate a smaller chimeric transcript (22, 34, 36, 44, 47). The final events in recombination are then envisioned as taking place during the next round of viral DNA synthesis.

One last model involves the copackaging of cellular RNA into the virion, thus making it accessible as an inappropriate template during viral DNA synthesis. In the avian sarcoma-leukosis virus system, a packaging mutant (SE21Q1b) which is deficient for viral RNA packaging but capable of packaging cellular RNAs has been identified (11, 32, 45), and these RNAs can serve as templates during reverse transcription (5, 30, 31, 55). Another mutant of the avian sarcoma-leukosis virus group, TK15, is less well characterized but appears to have similar properties (24, 27, 37, 38). Under certain circumstances, a specific cellular RNA is included in virions in the absence of any known viral defect (1, 23). All of these observations indicate that the packaging process has the potential to include cellular RNA in the virion, possibly juxtaposing this RNA with viral RNA and the viral DNA polymerase, reverse transcriptase. Strand jumping (6, 12) to the copackaged RNA during viral DNA synthesis could result in a recombinant viral DNA with a cellular insert, as has been seen with the TK15 mutant (38). While it is possible that more than one of these mechanisms contributed to the known recombinants, all of these models make a very specific prediction about the source of the cell-derived sequence: i.e., it must first be incorporated within the virion

* Corresponding author.

either as a separate RNA or as a chimeric RNA recombination intermediate.

We have taken another approach to identifying mechanisms at work in generating recombinants between retroviral sequences and cellular sequences. We have used a retrovirus shuttle vector in which we clone circular unintegrated DNA to examine genetic instability associated with virus replication (39, 40). These circular DNA molecules are not competent to participate in integration and thus are dead-end by-products of DNA synthesis (4). However, they provide a convenient source of viral DNA that can be analyzed for patterns of genetic instability which may also be reflected in linear DNA, the precursor to both circular and integrated DNAs (46). There are many types of aberrant products seen in circular viral DNA, including molecules that are greater than full length, molecules with internal deletions, and molecules with deletions in which one boundary of the deletion is near the edge of a long terminal repeat (LTR). The greater-than-full-length class includes dimers and trimers that are likely the result of homologous recombination (18, 29). We have discovered other members of this class of molecules whose larger DNAs can be attributed to internal duplications and to the presence of linked cellular sequences. In this report, we describe the structures of several molecules with cellular inserts adjacent to the viral LTRs.

MATERIALS AND METHODS

Cell lines. Fertilized chicken eggs (C/O) were obtained from the Regional Poultry Research Laboratory, East Lansing, Mich. Chicken embryo fibroblasts (CEF) and QT6 cells, a continuous Japanese quail cell line (35), were maintained as previously described (40).

Plasmids and bacteria. pANV-A has been described previously (39). It was derived from the Schmidt-Ruppin A strain of Rous sarcoma virus (7) and modified by introducing a *Cla*I site in place of the *v-src* gene (21). Plasmids pNO220, pMD1, pMD37, pMD66, and pMD96 were obtained from the previously described collection of unintegrated avian *neo* virus subgroup A (ANV-A) circular DNA clones rescued in the DH5 α strain of *Escherichia coli* (39).

Analysis of DNA. Chromosomal DNA was isolated from chicken and quail cells by the method of Dillella and Woo (10). DNA was digested with restriction endonucleases obtained from Life Technologies, Inc. (Gaithersburg, Md.) or from New England Biolabs (Beverly, Mass.). Digested DNA was fractionated by electrophoresis in agarose gels (Seakem LE; FMC Corp., Marine Colloids Div., Rockland, Maine) and transferred from agarose to nitrocellulose (BA85; Schleicher & Schuell Inc., Keene, N.H.) as described previously (40). Multiprimer DNA probes (Multiprimer DNA labeling system; Amersham Corp., Arlington Heights, Ill.) labeled by using [α -³²P]dCTP (800 Ci/mmol; Dupont, NEN Research Products, Boston, Mass.) were hybridized as described previously (40). Nucleotide sequence analysis was performed by using the dideoxy chain termination method (Sequenase protocol; U.S. Biochemical Corp., Cleveland, Ohio).

Amplification and cloning of cell-derived sequences related to the inserted sequences in pNO220, pMD37, pMD66, and pMD96. CEF genomic DNA (2 μ g), QT6 genomic DNA (2 μ g), and plasmid DNA (60 pg) were subjected to polymerase chain reaction (PCR) amplification in 100- μ l reaction mixtures containing 10 mM Tris-HCl (pH 8.3); 50 mM KCl; 1.5 mM MgCl₂; 0.2 mM each dATP, dGTP, dCTP, and dTTP; 10

μ M each oligonucleotide primer; and 2.5 U of *Taq* DNA polymerase (Perkin-Elmer Cetus, Norwalk, Conn.). The primers used were 220L (5'-GCTAGTCTAGAAGAAGTAT TGCTTGGTTGCA-3') and 220R (5'-CGATGGAGCTCCT GTATTITGACAGCACTGC-3') for pNO220, 37L (5'-GCTCTAGAGTCGACCTTCCAGTGTCTGTGTGACA-3') and 37R (5'-CGTCTAGAGTCGACGAATGAGTATCAGAGC CTTG-3') for pMD37, 66L (5'-CGTCTAGAGTCGACGGA TAGAGTATAGCATGAGG-3') and 66R (5'-GCTCTAGA GTCGACGCTCCTGAGCCACTGCAGAT-3') for pMD66, and 96L (5'-GCTCTAGAGTCGACCATGATGCTAAGGT GATA-3') and 96R (5'-GCTCTAGAGTCGACTAAGTGGT TAATAGTACT-3') for pMD96. The reactions were subjected to 30 cycles of 30 s at 95°C and 30 s at 60°C. The PCR products were purified by gel electrophoresis (NuSieve GTG; FMC), and impurities were removed by using an NaI-glass bead procedure (GeneClean II kit; Bio 101, Inc., La Jolla, Calif.). PCR products were then cloned into the pIBI31 plasmid vector (International Biotechnologies Inc., New Haven, Conn.) and subjected to DNA sequence analysis.

Sequence organization and comparison was done by using the GCG Sequence Analysis package (9). Data base similarity searches were conducted by using the BLAST program (2).

Analysis of RNA. Polyadenylated cellular RNA was isolated by the method of Gonda et al. (15), fractionated by electrophoresis in agarose containing formaldehyde, and transferred to nitrocellulose as described elsewhere (33). Strand-specific DNA probes were generated by PCR by using one primer from the inserted cell sequence. Gel-purified PCR-amplified DNA fragments (500 ng) were amplified in 100- μ l reaction mixtures containing 10 mM Tris-HCl (pH 9.0); 50 mM KCl; 0.1% Triton X-100; 1.5 mM MgCl₂; 4 μ M each dATP, dGTP, and dTTP; 0.16 μ M dCTP; 10 μ Ci of [α -³²P]dCTP (800 Ci/mmol; Dupont, NEN); and 5 U of *Taq* DNA polymerase (Promega Corp., Madison, Wis.). The reaction conditions were 23 cycles of 30 s at 95°C, 30 s at 55°C, and 2 min at 72°C. Each strand of the insert sequence was individually amplified and then hybridized to immobilized QT6 RNA and immobilized *Eco*RI-digested QT6 DNA. The QT6 RNA was subsequently hybridized with a multiprimer actin DNA probe to ensure consistent loadings from lane to lane.

Amplification of cellular DNA flanking sequences related to the insert sequences in pNO220 (inverse PCR). *Eco*RI-digested QT6 cellular DNA was self-ligated (circularized) in a 50- μ l reaction mixture containing 50 mM Tris-HCl (pH 7.6), 10 mM MgCl₂, 1 mM dithiothreitol, 5% (wt/vol) polyethylene glycol-8000, 0.6 μ g of DNA, and 2 U of T4 DNA ligase (Life Technologies) for 16 h at room temperature. The ligase was then inactivated by heating at 68°C for 30 min, and the DNA was concentrated by ethanol precipitation. The self-ligated DNA product (0.2 μ g) was then subjected to PCR amplification as described above by using primer A (5'-CGATC GAATTCGCCAGGGAAGTTTCTACA-3') and primer B (5'-AGATCGAATTCTCATGTTGTTGCTCTGCA-3'). The amplification reaction consisted of 30 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 60 s. The product of the PCR reaction (1.2 kbp) was cleaved with *Eco*RI and gel purified. The purified products were cloned and subjected to nucleotide sequence analysis as described above.

Nucleotide sequence accession numbers. The GenBank accession numbers for pMD37, pMD66, pMD96, and pNO220 are M97290, M97291, M97292, and M97293, respectively.

FIG. 1. Structure of the ANV-A genome and four clones that contain cellular inserts. Thin line, sequence derived from the Schmidt-Ruppin strain of Rous sarcoma virus; large open box, LTR unit; stippled box, sequence from the pBR322 origin of DNA replication (*ori*); narrow open box, sequence from the simian virus 40 (SV40) early enhancer-promoter region; thick line, sequence from the Tn5 Neo^r gene, including the promoter region (*neo*). For pNO220, the arrow indicates the region of sequence duplication in the molecule. The dark boxes between the LTRs in pNO220, pMD37, and pMD66 and adjacent to the LTR in pMD96 indicate sequences of nonviral origin. The numbers of nucleotides missing from the LTRs bordering the inserts are denoted by Δ .

RESULTS

Analysis of greater-than-full-length genomes. We previously isolated a large number of clones of circular retroviral DNA by using a shuttle vector strategy for rapid cloning in bacteria (39). The shuttle vector is a replication-competent retrovirus genome derived from a DNA clone of the Schmidt-Ruppin A strain of Rous sarcoma virus (7) with the Neo^r gene and a bacterial plasmid origin of replication in place of the *v-src* gene (21, 39). A virus stock of this vector, referred to as pANV-A, grown in CEF cells was used to infect the continuous quail cell line QT6. Unintegrated circular viral DNA was isolated at 24 h postinfection and introduced into bacteria, which were then selected for kanamycin resistance. The plasmid DNA present in each resistant bacterial clone was mapped with several restriction enzymes and cataloged on the basis of various features of its structure relative to that of the parental viral genome (39). Approximately 95% of the circular viral DNAs isolated were either the same size as or smaller than the parental genome, while the remaining 5% were larger than the parental genome (39). A number of these greater-than-full-length clones contained duplications of viral sequences (data not shown).

Analysis of one greater-than-full-length clone, pNO220, revealed two alterations in the genome structure, both of which contributed to the increased size (Fig. 1). One change resulted in a direct repeat of approximately 1.2 kbp near the 3' terminus of the *pol* gene. The other change appeared in the LTR region. The most common structures in the circle

junction region for circular viral DNAs have one or two copies of the LTR sequence. Restriction enzymes that cleaved on either side of the LTRs in pNO220 gave DNA fragments that were larger than those expected for a normal circle junction region. However, cleavage with enzymes that recognized sites within the LTR suggested that the viral sequences flanking the LTRs and internal to the genome were intact.

pNO220, pMD37, and pMD66 contain inserts between two altered LTRs. Further mapping of pNO220 indicated that the molecule had two LTR sequences in their normal positions with respect to the flanking viral sequences. However, the distance between the two LTRs was greater than would be expected for LTRs that were in tandem at the circle junction site (Fig. 1). In pNO220, the fragment spanning the circle junction region of the two LTRs (generated with *EcoRI*, which cleaves within the LTR) was approximately 600 bp long rather than the 330-bp fragment that is generated from tandem LTRs (52). This *EcoRI* fragment was subcloned and sequenced through the boundaries of the LTRs by using oligonucleotide primers which anneal within the LTR for each of the two LTR sequences.

In pNO220, the outer boundary of each LTR was altered relative to the sequence arrangement seen in the circle junction consisting of tandem LTRs. The LTR derived from the right end of the linear viral DNA (ending in U5) had an extra copy of the primer-binding sequence (PBS), which is normally present adjacent to the U5 domain of the other

FIG. 2. Nucleotide sequences from each of the nonviral inserts showing the junctions with the ANV sequence and comparisons with quail and chicken DNAs. The viral sequences flanking each of the inserts are in boldface type along with the ANV sequences given above the junction sequence of each of the clones. The sense strand (plus strand) of the viral sequence is shown in each case. The boundaries of the U5 and U3 regions of the LTRs in ANV and the boundaries of PBSs are indicated by bold vertical lines. Sequences related to each of the inserts were amplified from quail and chicken DNAs by PCR, and the gel-purified amplified fragments were cloned and subjected to nucleotide sequence analysis. A 50-bp segment of each insert is compared with quail and chicken DNAs. Positions of identity are denoted by dots, with percentages of sequence identity to the entire insert given in parentheses. The chicken DNA fragment amplified by using the pMD96 primers shows poor alignment to the plasmid insert sequence over its entire length, while no major chicken DNA fragment was amplified by using the pMD37 primers.

LTR and internal to the genome. This extra copy contained the entire PBS except for the 3' terminal nucleotide (Fig. 2). Sequence analysis showed that the region of pNO220 which should contain the normal PBS was intact (data not shown). The LTR derived from the left end of the linear viral DNA carried a 7-bp deletion at the outer U3 boundary. Between both of these aberrant LTRs was an insert of 268 bp. Computer searches of the inserted sequence showed no significant sequence similarity to any region of the viral genome nor to any sequence deposited in the GenBank data base.

Restriction enzyme mapping using *EcoRI* showed that pMD37 and pMD66 also generated restriction fragments spanning the circle junction site that were larger than expected for two tandem LTRs (Fig. 1). Subcloning of the *EcoRI* fragments and sequence analysis through both LTR boundaries showed that each molecule contained altered LTRs along with a nonviral insert between the LTRs. The outer U3 boundary in the left-end LTR of pMD37 had a

16-bp deletion, while the outer U5 boundary in the right-end LTR had 2 bp deleted (Fig. 2). The insert between these alterations was 813 bp. In pMD66, the LTR derived from the left end of the linear viral DNA had 2 bp deleted at the outer U3 boundary, while the LTR derived from the right-end LTR carried a 22-bp deletion at the outer U5 boundary (Fig. 2). The nonviral insert in this molecule was determined to be 657 bp. A computer search revealed no similarity between any of these insert sequences and any region of the viral genome or any sequence deposited in the GenBank data base. The absence of the terminal 2 bp from one of the LTRs of pMD66 and pMD37 suggests that these DNAs may have formed as the result of an abortive integration event (see Discussion).

pMD96 contains a U5-associated deletion and insertion. A separate screen of clones involving hybridization to cellular DNA showed that pMD96 contained a cell-derived insert. Subsequent restriction enzyme mapping revealed that this clone had only one LTR and a deletion extending from the

U5 boundary of the LTR removing *gag*, *pol*, and a portion of *env* (Fig. 1). Sequence analysis through the LTR boundary with an oligonucleotide primer that anneals within the U5 region revealed nonviral sequences adjacent to the LTR at the U5 boundary. There are also 2 bp deleted from the U5 boundary (Fig. 2). The insert between the LTR and the *env* region was found from sequence analysis to be 349 bp.

Of all circular viral DNAs isolated using the ANV-A shuttle vector, approximately 30% have deletions in the 5' portion of the viral genome. The deletions in these clones typically have one endpoint near the outside U5 boundary of the LTR (39). Because of the discovery of the insert in pMD96, we sequenced approximately 50 clones with the deletion site extending from the U5 region but did not find another example of a nonviral sequence insertion in this class of molecules.

Nonviral inserts are derived from quail cells. We next considered the possibility that the nonviral inserts were derived from a previously uncharacterized region of the host genome. Because of the design of the experiment, it was necessary to consider two different species types as potential donors of the insert in each case. Virus stock was generated in chicken cells, while acute infection prior to cloning of unintegrated viral DNA was carried out in the quail cell line QT6. In the case of pNO220, we made a radioactive probe consisting of the insert sequence and hybridized it to both chicken and quail DNAs. The probe showed good hybridization to quail DNA but poor hybridization to chicken DNA (data not shown).

To address more directly the question of the derivation of the inserted sequences, we used PCR to amplify a region of the insert sequence from both chicken and quail DNAs. For pNO220, a short region (insert nucleotides 189 to 238) from chicken and quail DNAs was amplified, subcloned, and subjected to DNA sequence analysis. Comparison of the sequence of the insert present in pNO220 with the sequences of the equivalent regions of chicken and quail DNAs showed that the insert was identical to the quail-derived sequence and divergent from the chicken sequence (Fig. 2).

For pMD37, pMD66, and pMD96, PCR primers were designed within the nonviral sequence adjacent to the boundary with the viral sequence, and amplification was carried out by using the parental plasmid, chicken DNA, and quail DNA. PCR using quail DNA and primers for the pMD37 insert yielded a fragment the same size as that amplified from the plasmid plus a larger fragment. Multiple minor fragments of various sizes were amplified in chicken DNA; none of these fragments was the same size as the plasmid-derived product. Amplification of quail DNA by using PCR primers designed from the pMD66 insert resulted in a single fragment the same size as the plasmid product, while chicken DNA yielded at least three fragments, one slightly larger than the plasmid product and two significantly smaller. The pMD96 PCR insert primers also amplified a single fragment in quail DNA of a size similar to that of the fragment in the plasmid; amplification of chicken DNA resulted in one slightly smaller product and at least four larger products. The prominent amplified DNA fragments in all cases were cloned and subjected to DNA sequence analysis. Figure 2 shows the sequence comparisons of regions within the inserts for these molecules. The sequence of the plasmid insert in all instances was most similar and usually identical to the quail-derived sequence and diverged from the chicken-derived sequence. The source of the cell-derived sequence in each of these four cases was the QT6 cell line.

FIG. 3. RNA blot analysis of insert sequences in pNO220, pMD96, pMD37, and pMD66. Approximately 5 μ g of polyadenylated RNA from QT6 cells was fractionated by agarose-formaldehyde gel electrophoresis. Transcripts were detected by strand-specific probes which were generated by using insert sequences from each of the four clones as templates for asymmetric PCR. Probes used in the odd-numbered lanes detected (i.e., annealed to) sequences of the sense shown in Fig. 2. Probes used in the even-numbered lanes detected sequences complementary to the sense shown in Fig. 2. Probes were generated from the pNO220 insert (lanes 1 and 2), pMD96 insert (lanes 3 and 4), pMD37 insert (lanes 5 and 6), and pMD66 insert (lanes 7 and 8). The band below each of the lanes represents subsequent hybridization of the same filter strip when an actin probe was used to verify the RNA loaded in each lane. All of the probes showed strong hybridization to *Eco*RI-digested QT6 DNA (not shown).

Cell-derived sequences from pNO220 and pMD96 are expressed as RNA. Recombination between host and viral sequences could occur by using either host DNA or host RNA. To determine if host RNA could have participated in the recombination event, we asked if the sequences in the cell-derived insert present in each of the molecules were expressed as RNA in uninfected cells. In this experiment, we made use of strand-specific probes which would allow us to determine the polarity of the insert relative to the viral genome.

When probes generated by the pNO220 insert were used, the probe to one strand of the insert hybridized to multiple RNA species in a quail poly(A)⁺ RNA preparation (Fig. 3, lane 1). These RNAs ranged in size from 2.0 to 8.5 kb. The other strand-specific probe did not hybridize to RNA at detectable levels (Fig. 3, lane 2). Furthermore, the polarity of the probe that hybridized indicated that the plus strand of viral DNA was linked to the minus or cDNA strand of the cell-derived sequence.

One strand-specific probe generated by the pMD96 insert also hybridized to quail poly(A)⁺ RNA (Fig. 3, lane 4). The hybridization was to multiple RNA species ranging in size from 4.0 to 9.5 kb. Hybridization was not detected with the probe generated by the other strand (Fig. 3, lane 3). In this case, the polarity of the probe that hybridized indicated that the minus strand of viral DNA was linked to the minus or cDNA strand of the cell-derived sequence.

Strand-specific probes were also generated for pMD37 (Fig. 3, lanes 5 and 6) and pMD66 (Fig. 3, lanes 7 and 8) and subjected to Northern (RNA) analysis with quail poly(A)⁺ RNA. No hybridization with any of these probes was detected.

Inserts in pNO220, pMD66, and pMD96 are derived from single-copy genes. To determine whether the cellular inserts were present as a single copy or multicopies, we hybridized single-stranded DNA probes generated from each of the inserts to QT6 DNA and CEF DNA, each singly digested with three different restriction enzymes (data not shown). Probes derived from pNO220, pMD66, and pMD96 cellular

FIG. 4. Analysis of QT6 cell DNA adjacent to sequences related to the insert sequences in pNO220. Flanking-cell DNA sequences related to the insert sequences in pNO220 were amplified by using inverse PCR, cloned into a pBI31 vector, and subjected to sequence analysis. Underlined sequences are shared at the recombination sites.

sequences each showed a single band (the bands were of different sizes) in QT6 DNA digested with *Bam*HI, *Eco*RI, or *Xba*I and in CEF DNA digested with the same enzymes. The pMD37 probe hybridized to a number of fragments in all lanes of digested QT6 DNA and CEF DNA. Thus, for three of the inserts, we can identify with confidence the source of the sequence linked to viral DNA. The final example appears to be associated with a DNA sequence that is repeated within the avian genome.

Sites of recombination in the pNO220 insert show little homology. To determine whether sequence homology played a role in the recombination events that generated pNO220, the normal cellular sequence was determined and compared with viral sequences at the sites of joining (PBS and U3 boundary). Primers from within the cell-derived insert were designed to point outward from the insert, and these primers were used in an inverse-PCR strategy to clone the flanking sequence in cellular DNA. The amplified fragment was cloned and subjected to DNA sequence analysis at the sites of recombination with viral sequences (Fig. 4). At one end of the insert (near the PBS), there were 4 bp of homology between the quail sequence and the 3' end of the PBS at the site of recombination (GTGA). At the other end of the insert, the sequences had only a single nucleotide (T) in common at the site of recombination between the cellular sequence and the U3 region. These results suggest that extended homology was not utilized in the recombination event.

DISCUSSION

We have previously observed that approximately 5% of the cloned unintegrated circular DNAs derived from an acute infection of QT6 cells with the ANV-A retrovirus shuttle vector were greater than full length (39). A portion of this population consisted of oligomeric viral DNA and viral DNA containing other types of duplications (data not shown), similar to molecules that have been described previously (18, 29). In addition, we reasoned that this population might also contain molecules that had recombined with cellular sequences. While the frequency of this event has not been measured, examples of circular viral DNAs from this class with linked cellular sequences were found in the molecules pNO220, pMD37, and pMD66. In addition, we characterized one smaller molecule with a cell-derived insert, pMD96.

Origins of pNO220 and pMD96 may have involved illegitimate recombination involving cellular RNA. The occurrence of a cell-derived insert between LTRs in viral DNA could be the result of imprecise excision of integrated DNA (25, 57), insertion of cellular DNA during circularization of viral

DNA, or incorporation of cellular sequences during viral DNA synthesis. In the case of pNO220, the fact that the cell-derived insert is adjacent to a nearly complete and extra copy of the PBS suggests that the insert arose from an aberrant jump during reverse transcription. A likely point for this event is after completion of synthesis of plus-strand strong-stop DNA. The structure of plus-strand strong-stop DNA includes a PBS at its 3' end which has been shown to be a stable intermediate during reverse transcription (12, 28, 42, 51, 56). Normally, the plus-strand strong-stop DNA is elongated by using minus-strand DNA as template after a jump to the 3' end of the nearly full-length minus-strand DNA (12). The precise linkage between the PBS and the cellular sequence in pNO220 can be explained by an inappropriate jump of plus-strand strong-stop DNA to a cellular

FIG. 5. Possible mechanism for the generation of pNO220 (A) and pMD96 (B) clones containing cellular inserts. Thin line, viral RNA; medium line, minus-strand DNA; thick line, plus-strand DNA; thin line with poly(A) tail, cell RNA. The U3, R, and U5 regions of the LTR are indicated.

sequence (Fig. 5A). The observations that the insert is expressed as RNA in QT6 cells and is linked to the viral plus strand in a cDNA sense are consistent with cellular RNA as the target for the jump.

In the case of pMD96, the cell-derived insert occurred in conjunction with a deletion of *gag*, *pol*, and a portion of *env*. Hybridization to QT6 cell RNA revealed that the insert sequence in pMD96 is joined to the viral DNA minus strand in a cDNA sense. The structure of this molecule, which has only part of *env* and one LTR, together with the expression of the insert as RNA in QT6 cells supports the suggestion that an aberrant jump to the cellular RNA occurred during minus-strand synthesis (Fig. 5B). However, the U5 boundary of the LTR is missing 2 bases in pMD96, making it ambiguous whether these bases were lost in a recombination event associated with DNA synthesis or whether this molecule was the result of an aberrant integration event.

Cell-derived sequences from pMD37 and pMD66 appear to be the result of aborted integration. Both pMD37 and pMD66 have a 2-bp deletion from one LTR boundary along with two LTRs in a normal configuration with respect to the remainder of the viral genome, indicative of integration (reviewed in reference 4). Neither strand of the inserted sequence in these clones showed any detectable hybridization to QT6 cell RNA. The structures of pMD37 and pMD66 imply that the cellular inserts are the outcomes of a prematurely terminated integration event.

Implications. We interpret the structures of these molecules as arising by two mechanisms: (i) abortive integration and (ii) aberrant strand jumping during DNA synthesis. While it is possible that pNO220 and pMD96 are both the result of prematurely terminated integration events, the correlative evidence suggests that this is not the case. Although the exact nature of the cellular target in these cases is not known for certain, the data are consistent with cellular RNA serving as the template for recombination during viral DNA synthesis.

The structures of pNO220 and pMD96 have several implications for the reverse transcription process with respect to recombination with host sequences. The first is that recombination can occur with cellular RNA sequences serving as a template for reverse transcription. The inserted sequences in pNO220 and pMD96 are expressed as RNA, and the polarity of strand linkage suggests that either the plus or the minus strand can participate in aberrant jumping to a cellular mRNA target. Pulsinelli and Temin, studying mutations in spleen necrosis virus (42), have recently identified clones containing nonviral sequences which they suggest were derived from mRNA sequences. One such insert shows strong similarity to a known cellular mRNA. The acquisition of this cell sequence was thought to occur during reverse transcription by a switch to an mRNA template presumed to be copackaged from the parental helper cell line. A similar conclusion regarding RNA-mediated recombination in yeast cells was reached by Derr et al. (8). Using a reporter gene, they found that RNA can serve as an intermediate in the formation of the functional equivalent of a pseudogene, facilitated by the expression of the retrotransposon Ty.

A second implication of the structures of pNO220 and pMD96 is based on the observation that the nonviral sequences which participated in the recombination event were present in the acutely infected cell and therefore must have had some degree of accessibility to the subviral ribonucleoprotein complex (3) active in reverse transcription. We propose that a cellular mRNA not intimately associated with the replication complex was the direct target for the recom-

bination events that resulted in both pNO220 and pMD96. However, we have not completely ruled out a mechanism of copackaging viral and cellular sequences into a virus particle followed by reinfection and reverse transcription, since these events could have occurred within the time allowed for the acute infection. Nevertheless, we consider a copackaged RNA target an unlikely source of the recombining sequence in this case. First, the virus was exposed to the quail cells for a fairly short period (24 h), allowing little time for the initiation of a second round of infection. Second, we have found that virus vectors with oversized genomes (like the parental ANV-A with its Neo^r insert) are blocked at a late step in replication in QT6 cells (39a), making a second round of infection improbable.

A third implication of our results is that template switching by reverse transcriptase can occur with only limited homology; in the case of pNO220, only 4 bases of perfect homology appear to have mediated the jump from plus-strand strong-stop DNA to the cellular sequence (Fig. 4). Pathak and Temin (41) have found that small deletions and duplications of viral sequences apparently generated by template switching during reverse transcription can be mediated by only a few bases of homology. Pulsinelli and Temin (42), who analyzed large deletions of viral sequences, have noted similar results.

The system we have used includes cell lines from two different species for virus production (chicken) and acute infection (quail), allowing us to determine the origin of the recombined nonviral sequences. The model for recombination that we propose here, which involves extracapsid cellular RNA, is a mechanism that could also give rise to the formation of viral genomes with internal cellular sequences, i.e., transduction. We are currently exploring the implications of this new model for virus-host recombination.

ACKNOWLEDGMENTS

We thank Carol Bova-Hill and Pascale Nantermet for assistance with portions of this study.

This work was supported by Public Health Service grant RO1-CA33147 and by a grant from the North Carolina Biotechnology Center.

REFERENCES

- Adkins, B., and T. Hunter. 1981. Identification of a packaged cellular mRNA in virions of Rous sarcoma virus. *J. Virol.* 39:471-481.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Bowerman, B., P. O. Brown, H. E. Varmus, and J. M. Bishop. 1989. A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev.* 3:469-478.
- Brown, P. O. 1990. Integration of retroviral DNA. *Curr. Top. Microbiol. Immunol.* 157:19-48.
- 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.
- Coffin, J. M. 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J. Gen. Virol.* 42:1-26.
- DeLorbe, W. J., P. A. Luciw, H. M. Goodman, H. E. Varmus, and J. M. Bishop. 1980. Molecular cloning and characterization of avian sarcoma virus circular DNA molecules. *J. Virol.* 36:50-61.
- Derr, L. K., J. N. Strathern, and D. J. Garfinkel. 1991. RNA-mediated recombination in *S. cerevisiae*. *Cell* 67:355-364.
- Devereux, J., P. Haeblerli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic*

- Acids Res. 12:387-396.
10. Dillela, A. G., and S. L. C. Woo. 1985. Cosmid cloning of genomic DNA. *Focus* 7:1-5.
 11. Gallis, B., M. Linial, and R. Eisenman. 1979. An avian oncovirus deficient in genomic RNA: characterization of the packaged RNA as cellular messenger RNA. *Virology* 94:146-161.
 12. Gilboa, E., S. W. Mitra, S. Goff, and D. Baltimore. 1979. A detailed model of reverse transcription and tests of crucial aspects. *Cell* 18:93-100.
 13. Goff, S. P., C. J. Tabin, J. Y. Wang, R. Weinberg, and D. Baltimore. 1982. Transfection of fibroblasts by cloned Abelson murine leukemia virus DNA and recovery of transmissible virus by recombination with helper virus. *J. Virol.* 41:271-285.
 14. 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.
 15. Gonda, T. J., D. K. Sheiness, and J. M. Bishop. 1982. Transcripts from the cellular homologs of retroviral oncogenes: distribution among chicken tissues. *Mol. Cell. Biol.* 2:617-624.
 16. 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.
 17. 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.
 18. Goubin, G., and M. Hill. 1979. Monomer and multimer covalently closed circular forms of Rous sarcoma virus. *J. Virol.* 29:799-804.
 19. Herman, S. A., and J. M. Coffin. 1987. Efficient packaging of readthrough RNA in ALV: implications for oncogene transduction. *Science* 236:845-848.
 20. 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.
 21. Hughes, S., and E. Kosik. 1984. Mutagenesis of the region between *env* and *src* of the SR-A strain of Rous sarcoma virus for the purpose of constructing helper independent vectors. *Virology* 136:89-99.
 22. Ikawa, S., K. Hagino-Yamagishi, S. Kawai, T. Yamamoto, and K. Toyoshima. 1986. Activation of the cellular *src* gene by transducing retrovirus. *Mol. Cell. Biol.* 6:2420-2428.
 23. 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.
 24. Kawai, S., and T. Koyama. 1984. Characterization of a Rous sarcoma virus mutant defective in packaging its own genomic RNA: biological properties of mutant TK15 and mutant-induced transformants. *J. Virol.* 51:147-153.
 25. Kleckner, N. 1989. Transposon *Tn10*, p. 227-268. In D. E. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.
 26. Klempnauer, K. H., T. J. Gonda, and J. M. Bishop. 1982. Nucleotide sequence of the retroviral gene *v-myb* and its cellular progenitor *c-myb*. The architecture of a transduced oncogene. *Cell* 31:453-463.
 27. Koyama, T., F. Harada, and S. Kawai. 1984. Characterization of a Rous sarcoma virus mutant defective in packaging its own genomic RNA: biochemical properties of mutant TK15 and mutant-induced transformants. *J. Virol.* 51:154-162.
 28. Kung, H. J., Y. K. Fung, J. E. Majors, J. M. Bishop, and H. E. Varmus. 1981. Synthesis of plus strands of retroviral DNA in cells infected with avian sarcoma virus and mouse mammary tumor virus. *J. Virol.* 37:127-138.
 29. Kung, H. J., P. R. Shank, J. M. Bishop, and H. E. Varmus. 1980. Identification and characterization of dimeric and trimeric circular forms of avian sarcoma virus-specific DNA. *Virology* 103:425-433.
 30. 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.
 31. Linial, M. 1987. Creation of a processed pseudogene by retroviral infection. *Cell* 49:93-102.
 32. Linial, M., E. Medeiros, and W. S. Hayward. 1978. An avian oncovirus mutant (SE21Q1b) deficient in genomic RNA: biological and biochemical characterization. *Cell* 15:1371-1381.
 33. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 34. Martin, P., C. Henry, F. Ferre, C. Bechade, A. Begue, C. Calothy, B. Deguire, D. Stehelin, and S. Saule. 1986. Characterization of a *myc*-containing retrovirus generated by propagation of an MH2 viral subgenomic RNA. *J. Virol.* 57:1191-1194.
 35. Moscovici, C., M. G. Moscovici, H. Jimenez, 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.
 36. Nilsen, T. W., P. A. Maroney, R. G. Goodwin, F. M. Rottman, L. B. Crittenden, M. A. Raines, and H.-J. Kung. 1985. *c-erbB* activation on ALV-induced erythroblastosis: novel RNA processing and promoter insertion result in expression of an amino-truncated EGF receptor. *Cell* 41:719-726.
 37. Nishizawa, M., T. Koyama, and S. Kawai. 1985. Unusual feature of the leader sequence of Rous sarcoma virus packaging mutant TK15. *J. Virol.* 55:881-885.
 38. Nishizawa, M., T. Koyama, and S. Kawai. 1987. Frequent segregation of more defective variants from a Rous sarcoma virus packaging mutant TK15. *J. Virol.* 61:3208-3213.
 39. 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.
 - 39a. Olsen, J. C., W. P. Osheroff, and R. Swanstrom. Unpublished data.
 40. Olsen, J. C., and R. Swanstrom. 1985. A new pathway in the generation of defective retrovirus DNA. *J. Virol.* 56:779-789.
 41. 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.
 42. 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.
 43. Raines, M. A., N. J. Mailhe, C. Moscovici, L. Crittenden, and H.-J. Kung. 1988. Mechanisms of *c-erbB* transduction: newly released transducing viruses retain poly(A) tracts of *erbB* transcripts and encode C-terminally intact *erbB* proteins. *J. Virol.* 62:2437-2443.
 44. Rosson, D., D. Dugan, and E. P. Reddy. 1987. Aberrant splicing events that are induced by proviral integration: implications for *myb* oncogene activation. *Proc. Natl. Acad. Sci. USA* 84:3171-3175.
 45. Shank, P. R., and M. Linial. 1980. Avian oncovirus mutant (SE21Q1b) deficient in genomic RNA: characterization of a deletion in the provirus. *J. Virol.* 36:450-456.
 46. Shank, P. R., and H. E. Varmus. 1978. Virus-specific DNA in the cytoplasm of avian sarcoma virus-infected cells is a precursor to covalently closed circular viral DNA in the nucleus. *J. Virol.* 25:104-114.
 47. Shen-Ong, G. L. C., H. C. Morse, M. Potter, and J. F. Mushinski. 1986. Two modes of *c-myb* activation in virus-induced mouse myeloid tumors. *Mol. Cell. Biol.* 6:380-392.
 48. Stuhlmann, H., M. Dieckmann, and P. Berg. 1990. Transduction of cellular *neo* mRNA by retrovirus-mediated recombination. *J. Virol.* 64:5783-5796.
 49. Swain, A., and J. M. Coffin. 1989. Polyadenylation at correct sites in genome RNA is not required for retrovirus replication or genome encapsidation. *J. Virol.* 63:3301-3306.
 50. Swain, A., and J. M. Coffin. 1992. Mechanism of transduction by retroviruses. *Science* 255:841-845.
 51. Swanstrom, R., J. M. Bishop, and H. E. Varmus. 1982. Structure of a replication intermediate in the synthesis of Rous sarcoma virus DNA in vivo. *J. Virol.* 42:337-341.
 52. Swanstrom, R., W. J. DeLorbe, J. M. Bishop, and H. E. Varmus.

1981. Nucleotide sequence of cloned unintegrated avian sarcoma virus DNA: viral DNA contains direct and inverted repeats similar to those in transposable elements. *Proc. Natl. Acad. Sci. USA* **78**:124–128.
53. **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.
54. **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.
55. **Taylor, J. M., and A. Cywinski.** 1984. A defective retrovirus particle (SE21Q1b) packages and reverse transcribes cellular RNA, utilizing tRNA-like primers. *J. Virol.* **51**:267–271.
56. **Varmus, H. E., S. Heasley, H. J. Kung, H. Oppermann, V. C. Smith, J. M. Bishop, and P. R. Shank.** 1978. Kinetics of synthesis, structure and purification of avian sarcoma virus-specific DNA made in the cytoplasm of actively infected cells. *J. Mol. Biol.* **120**:55–82.
57. **Varmus, H. E., N. Quintrell, and S. Ortiz.** 1981. Retroviruses as mutagens: insertion and excision of a nontransforming provirus alter expression of a resident transforming provirus. *Cell* **25**:23–36.