Unusual Features of Integrated cDNAs Generated by Infection with Genome-Free Retroviruses

KATHRYN L. LEVINE,¹ BART STEINER,¹ KYLE JOHNSON,^{1†} RACHEL ARONOFF,^{1,2}
TINA J. QUINTON,^{1‡} AND MAXINE L. LINIAL^{1,2*}

Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104,^{1} and
Department of Microbiology, University of Washington, Seattle, Washington 98195²*

Received 21 November 1989/Accepted 18 January 1990

We previously demonstrated that when nonretroviral RNAs are encapsidated in retroviral particles they can be reverse transcribed into cDNAs, which are then integrated into the cellular genome. This transfer of genetic information via retroviral infection has been designated retrofection. Further analyses of three genes transferred in this manner (retrogenes) revealed that each was present in a single copy at a different site in the recipient quail cell genome and included a transcriptional promoter encoded by the encapsidated *neo* RNA. A unique feature of the retrogenes was a common 16-nucleotide sequence at or near a recombination border, which was not present in either recombination partner. The existence of this sequence suggests a common mechanism of retrogene formation and/or integration mediated by retrofection.

During recent years, the image of the eucaryotic genome as a static entity has undergone radical transformation. The molecular details of programmed or developmentally regulated DNA rearrangements (such as those that occur in B and T lymphocytes [19, 26], during antigenic or mating-type switching in African trypanosomes [2] or *Saccharomyces cerevisiae* [18], or in ciliated protozoans such as *Tetrahymena* sp. [51]) have been accessible to study because they occur in large populations of cells in response to cellular and organismal functional selections. Unscheduled rearrangements such as chromosomal translocation, inversion, or deletion have been studied post facto in tumor-derived cell lines (35). Gene rearrangements caused by retroviral integrations have also been analyzed (29, 47).

Several systems have been developed to study the recombination events that result in the acquisition of transferred genes by host cells (9, 12, 30, 41, 43); these systems primarily utilize transfection or microinjection to deliver the gene of interest to the host cell. Genes transferred in these ways, however, are frequently integrated as tandem arrays or in multiple sites. More, recently, we and others described systems that allow the study of the integration and expression of single-copy genes (10, 22). These studies entail infection of tissue culture cells with retroviral particles containing nonretroviral RNAs and authentic (virally encoded) reverse transcriptase, a process we have called retrofection. In our system (Fig. 1), we have used a quail cell line, SE21Q1b, that produces retroviral particles containing randomly packaged cellular mRNAs (14). Transfection of SE21Q1b cells with a plasmid containing the selectable marker *neo* under control of the Rous sarcoma virus (RSV) long terminal repeat (LTR) promoter produces retroviral particles that contain *neo* RNAs. Subsequent infection of the quail cell line QT35 (28) with the *neo* RNA containing viral particles leads to the establishment of G418-resistant clones

that harbor single-copy *neo* genes integrated into the chromosome. This process is RNA mediated and does not appear to require any retroviral sequences in *cis*, although inclusion of retroviral sequences increases retrofection efficiency.

DNA sequences occasionally appear in new genomic locations, both in germ line and somatic cells. In many instances, there is good evidence that these molecules (including certain retrotransposons such as long interspersed nuclear elements and intracisternal A particles) have transposed via RNA intermediates (1). Structural similarities between many of these genes and processed pseudogenes, including the lack of functional transcriptional control sequences and the presence of 3'-terminal poly(dA), have led to the suggestion that processed pseudogenes arose by the integration of cDNAs produced by reverse transcription (39, 49). To evaluate this suggestion and to study the processes involved in the retrovirally mediated transfer of nonretroviral genes, we cloned and sequenced *neo* retrogenes from three independently derived G418-resistant QT35 clones.

MATERIALS AND METHODS

Retrofection. SE21Q1b cells were transfected with plasmid DNAs by the method of Chen and Okayama (5), and individual G418-resistant clones were picked and grown into G418^r cell lines. The G418^r cell line SE21Q1b Cl.25 (22) was used in these experiments. Viral supernatants from individual cultures of SE21Q1b Cl.25 were filtered to remove cells and used to infect QT35 cells (retrofection assay as defined in Fig. 1). In the case of the mass G418^r cultures described in Table 1 and Fig. 7, all of the G418^r SE21Q1b colonies on a single plate (several hundred) were pooled.

Molecular cloning of the B2, G1, and J3 *neo* retrogenes and the B2 preintegration site. DNA was extracted from individual G418^r clones of QT35 cells that were obtained after infection by SE21Q1b particles containing mRNAs encoded by the plasmid pRSVneo. DNAs were digested with restriction endonucleases that do not cut within the pRSVneo plasmid. Using a 2.3-kilobase-pair (kb) *Bam*HI-*Hind*III *neo* probe and Southern blot analysis, we determined that suitably sized fragments for cloning could be obtained after digestion with *Sac*I or *Xba*I. Size-selected *Sac*I-digested DNA fragments from QT35 G418^r clones B2, G1, and J3

* Corresponding author.

† Present address: Department of Biochemistry, Biophysics and Genetics, University of Colorado, Health Sciences Center, Denver, CO 80262.

‡ Present address: Department of Pathology, University of Washington, Seattle, WA 98195.

FIG. 1. Retrofection protocol. SE21Q1b is a quail cell line that contains a mutant RSV provirus. RSV particles produced by these cells lack specifically packaged RSV genomes but rather contain randomly packaged poly(A)⁺ cellular mRNAs (24). SE21Q1b cells were transfected with either pCMVneo or pRSVneo (relevant plasmid sequences are shown in greater detail in Fig. 4A and 7), and Neo^r cell clones were selected in G418-containing medium (4, 22). Cell-free viral supernatants from individual SE21Q1b Neo^r clones (hatched) were used to infect QT35 cells (which lack viral sequences closely related to RSV), and individual clones were selected in G418-containing medium. Resultant Neo^r QT35 clones (retrofectants: stipple) contain single-copy, integrated *neo* cDNAs (retrogenes). These clones resulted from infections by virus particles that contain *neo* mRNA (n). Clones B2, G1, and J3 (described in the text) were from different infections with SE21Q1b Neo^r Cl.25 cell-free virions.

were extracted from low-melting-point agarose gels by hexadecyltrimethylammonium bromide extraction (21) and treated with calf intestine alkaline phosphatase (CAP) (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) *SacI* fragments were cloned into *SacI*-digested arms of λ 2001 (Stratagene, La Jolla, Calif.). *Escherichia coli* LE392 (*hsdR hsdM514 supE44 supF58 LacZ⁺ galK2 GalT22 metB1 trpR55*) cells were infected with recombinant phage, and plate lysates were prepared. Eluted phage were then used to infect *E. coli* B9 (which is LE392 [λ O_{am29}]; obtained from T. St. John, Fred Hutchinson Cancer Research Center, Seattle, Wash.), which was plated on LB (10.0 g of Bacto-tryptone, 5.0 g of Bacto-yeast extract, 10.0 g of NaCl per liter of H₂O;

pH 7.5) agar plates containing 30 μ g of kanamycin (Sigma Chemical Co., St. Louis, Mo.) per ml. Kan^r colonies were picked and grown overnight in LB broth containing 30 μ g of kanamycin per ml. CHCl₃ was added to 0.1%, cultures were cleared of cells by low-speed centrifugation, and titers of supernatants were determined on *E. coli* LE392. Clear plaques that hybridized to a 2.3-kb *Bam*HI-*Hind*III *neo* probe were picked and used for phage DNA preparations. Lambda phage clones were obtained from all three independently derived QT35 G418^r clones: B2, J3, and G1. Restriction endonuclease-generated fragments containing the complete *neo* retrogene and cellular flanking sequences (*SacI* for G1 and J3 and *XbaI* for B2) were subcloned into the plasmid vector pVZ1 (pBluescribe [Stratagene, La Jolla, Calif.] containing additional restriction enzyme sites in the poly-linker [17]) for sequence analysis.

Nick-translated probes used to characterize the *neo* retrogenes were as follows: a 2.3-kb *XbaI* fragment from the preintegration site of the B2 retrogene, a 220-base-pair (bp) *Bgl*III-*SacI* fragment from the 3'-flanking sequence of the G1 retrogene, and a 700-bp *SmaI* fragment from the 3'-flanking sequence of the J3 retrogene. Methods of DNA preparation, transfer, and hybridization have been described previously (23).

The preintegration site of B2 was cloned from a quail genomic library (kindly provided by C. Emerson, Department of Biology, University of Virginia, Charlottesville) by using a 1.2-kb *XbaI*-*Eco*RI fragment prepared from the 5' region of the cloned B2 retrogene as a probe. The 2.3-kb *XbaI* preintegration site fragment was subcloned into pVZ1, and the DNA sequences around the integration breakpoints were determined.

Construction of the plasmids used in retrofection. (i) **pRSVneo16⁺.** pRSVneo was kindly provided by F. Yoshimura (Department of Biostructure, University of Washington, Seattle). Both strands of a 40-bp oligonucleotide were synthesized on an Applied Biosciences DNA synthesizer. This 40-mer has *AvaI* ends, contains *XhoI* and *Asu*II sites, and includes the 16-mer TGCAATTAAACCAGCT referred to below. After annealing of the complementary strands, the 40-mer was ligated into pRSVneo that had been previously cleaved with *AvaI* and treated with CAP (see Fig. 6). Ligated (T4 ligase; U.S. Biochemical Corp., Cleveland, Ohio) DNAs were transfected into competent *E. coli* SJ72 cells. *XhoI* digestion was used to screen plasmid miniprep DNAs, and positive clones were confirmed by DNA sequencing.

(ii) **pCMVneo.** pBC12/CMV/IL-2 was kindly provided by Brain Cullen (Hoffman-LaRoche, Inc., Nutley, N.J.). Details of the construction of this plasmid were reported previously (22).

(iii) **pCMVneo16⁺/–.** pCMVneo was digested with *AvaI*; the smaller fragment, which contained the rat preproinsulin gene and the simian virus 40 polyadenylation signal, was discarded. The larger fragment was ligated to form pCMV-neoAva^P. As described above for pRSVneo16⁺, the annealed 40-mer was ligated into pCMVneoAva^P. Positive clones were screened by *XhoI* digestion followed by confirmation of plus or minus orientation of the 40-mer by DNA sequencing.

(iv) **pCMVTKneo.** pTK102, kindly provided by H. Weintraub (Fred Hutchinson Cancer Research Center), was digested with *Pvu*II and *Bgl*III; a 250-bp fragment containing the herpes simplex virus thymidine kinase promoter was isolated and treated with Klenow enzyme (Boehringer) to blunt the ends. After phosphorylation of the fragment by treatment with T4 polynucleotide kinase (U.S. Biochemical

FIG. 2. (A) Southern blot analyses of DNAs prepared from parental QT35 cells and three clonal cell lines that each contain a single *neo* retrogene (B2, G1, and J3). DNAs (10 μ g of each) were digested with *XbaI* or *SacI*. Each panel was probed with DNA fragments derived from the flanking sequences of the cloned retrogene (described in the text). The *neo*⁺ retrogene fragment(s) are denoted by dots to the right of each panel. The two dots in panel A point out the B2 retrogene fragments produced by digestion of B2 DNA by *SacI* (left) or *XbaI* (right). The other bands in all panels are fragments that hybridized to preintegration probes; a single fragment in the cases of B2 and G1 and a gene family in the case of J3. Molecular weight markers were λ DNA digested with *HindIII*. (B) Structures of the B2 retrogene pre- and postintegration sites. Directly repeated 40-bp sequences (sequence data not shown) found at the breakpoints are shown by stippled boxes in lines 1 and 2. The vertical arrows below line 2 indicate the recombination points

Corp.), *HindIII* linkers (New England BioLabs, Beverly, Mass.) were added. This fragment was then ligated into pCMVneo that had been previously digested with *HindIII* and treated with CAP. Positive clones were detected by colony hybridization with a thymidine kinase promoter probe, and orientation of the insert was confirmed by DNA sequencing.

(v) **pCMVneoSPo.** pACYC184 (kindly provided by J. Cooper, Fred Hutchinson Cancer Research Center) was digested with *SspI* and *PvuII* to free a fragment that contained the procaryotic origin of replication. After gel purification, this fragment was ligated into pCMVneoAva^D that had been digested with *SmaI* and treated with CAP. Positive clones and insert orientation were determined by digestion of plasmid miniprep DNAs with *XmnI*.

(vi) **pCMVneoLTRjxn.** pCMVneoLTRjxn was constructed by Richard Katz (Fox Chase Cancer Center, Institute for Cancer Research, Philadelphia, Pa.). Two copies of an avian leukosis virus (ALV) LTR junction (75 bp from U5 and 100 bp from U3) were cloned in tandem (head to tail) into the single *SmaI* site in pCMVneoAva^D; the orientation of the insert (5' *neo* U5U3U5U3) was confirmed by diagnostic restriction endonuclease mapping and DNA sequencing.

DNA sequence analysis. All DNA sequencing was done using the Sanger et al. dideoxy method (37). *XbaI* or *SacI* fragments were subcloned into pVZ1. Detailed restriction enzyme analysis with sites known to be in the pRSVneo plasmid (15) indicated where flanking sequences would start. To sequence the junctions, we either subcloned smaller fragments and used the universal sequencing primer (for B2 3' and J3 3' and 5' breakpoints) or synthesized specific oligonucleotide primers with an Applied Biosystems DNA synthesizer. The sequence of each junction was confirmed by synthesizing an oligonucleotide primer for the complementary strand.

RESULTS

***neo* retrogene cDNAs integrate by illegitimate recombination into different quail DNA sequences.** We molecularly cloned three *neo* retrogenes—B2, G1, and J3—derived from three independent G418-resistant QT35 cell clones after retrofection. These *neo* retrogenes arose from infections with particles produced by SE21Q1b Cl.25 cells transfected with the plasmid pRSVneo (Fig. 1). DNA sequencing and Southern blot analyses revealed that each of these molecular clones contained *neo* sequences flanked by various amounts of quail sequences both 3' and 5' (*neo* hybridization data not shown). Cellular flanking sequences were used as hybridization probes on Southern blots of quail cellular DNAs and to screen a library of nonretrofed quail DNA. Each cloned retrogene probe detected a single *neo*-containing restriction enzyme fragment, which was missing in parental cell lanes

between quail and *neo* cDNA sequences. The structure of the B2 retrogene is shown in line 3. The black boxes show the remnant of the 40-bp direct repeat retained at the 3' junction of *neo* and quail DNA sequences. This sequence is also present at the 5' junction. The deletion of approximately 900 bp of quail DNA (indicated by hatching in line 2) is discussed in the text. (C) Structure of the G1 retrogene. The G1 *neo* cDNA (solid line) is flanked by 215-bp direct repeats (stippled rectangles) of quail DNA sequences (dotted lines). The 3' repeat is immediately adjacent to the 3' end of the retrogene. The 5' direct repeat is separated from the *neo* sequence by 33 bp (dashed line).

FIG. 3. Retrogene cDNAs integrate into quail DNA via illegitimate recombination. Sequences present at recombination junctions between quail DNA and pRSVneo cDNAs in three *neo* retrogenes, B2, G1, and J3, are shown colinearly. Retrogene sequences are shown in boldface type, and sequences shown in lightface type above and below the continuous sequence of the retrogenes are continuations of pRSVneo not present in the retrogenes. The consensus polyadenylation signals of *neo* mRNAs are indicated by pA. Numbers below the colinear sequences indicate 5' and 3' breakpoints in pRSVneo. Arrows and numbers show the directions and lengths of repeated sequences referred to in the text.

(Fig. 2A). The sizes of the cloned retrogene fragments were as follows: B2, 5.8 kb; G1, 5.0 kb; and J3, 5.1 kb. It appeared that the fragment containing the B2 retrogene was present in less than one copy per cell (Fig. 2A), which we believe to be a consequence of long-term passage of the cells in medium lacking G418. For G1, a 3'-flanking probe that hybridized only to unique cellular sequences was used, since the 5'-flanking sequences primarily consisted of repetitive sequences (discussed below). The insertion site of J3 appeared to be a family comprised of at least three members. It is likely, based on the relative hybridization intensities of the fragments containing the three potential preintegration sites with the 3'-flanking probe, that the largest *SacI* fragment represents the J3 *neo* cDNA insertion site. Use of a J3 5'-flanking probe indicated that the 5'- and 3'-flanking sequences were not on the same *SacI* fragment in normal quail DNA; this suggests that in J3 the integration of *neo* cDNA was accompanied by a large deletion of quail cellular DNA (data not shown).

Using a B2 flanking probe, we molecularly cloned the preintegration site of the B2 retrogene (Fig. 2B). Sequence analysis of the quail cell preintegration site indicated that integration of the B2 *neo* cDNA was not a simple event such as is seen in the integration of retroviral proviruses. The preintegration site contained two 40-bp direct repeats flanking approximately 900 bp. Most of one repeat and the central 900 bp were deleted upon cDNA integration (Fig. 2B). This rearrangement resulted in 4-bp direct repeats flanking the *neo* cDNA.

The sequence of the recombination sites is shown in greater detail in Fig. 3. Comparisons of the plasmid sequences expected to extend 3' or 5' of the recombination sites of the cDNAs with those of the cellular flanking sequences revealed no obvious complementarity. Lack of homology at recombination junctions between integrated plasmid or proviral and host sequences has been a common finding in transfected as well as retrovirally infected eucaryotic cells (32, 40). Short, duplicated sequences of characteristic lengths flank integrated proviruses. With RSV, the provirus is flanked by a 6-bp duplication of the host integration site (47). Such precise 6-bp duplications were not seen in any of the three clones.

neo retrogenes contain internal promoters, lack 3'-terminal poly(dA), and share common 3' and 5' breakpoints. We

previously showed that the transcription of pRSVneo sequences in SE21Q1b cells starts both at the expected LTR R-region cap site (position +1) as well as at upstream sites (22). This led to the hypothesis that the integrated *neo* cDNAs are copies of 5'-extended mRNAs such that the cDNAs contain internal LTR promoter-enhancer elements that allow efficient gene expression. Sequence analysis revealed that the 5' ends of the integrated cDNAs were located upstream of the LTR cap site (position +1) at positions -228, -230, and -1025 (denoted by closed circles in Fig. 4A). The region at about position -230 is where we had previously mapped a transcriptional start site in SE21Q1b (22).

Figure 4B summarizes the structures of the three retrogenes. In each case, LTR sequences from pRSVneo were retained. LTR sequences were at 5' recombination junctions between plasmid-derived and cellular sequences in two clones, B2 and G1. In G1, they were truncated by two bases relative to the end of a bona fide LTR, and in B2, they were complete and identical. The 3' ends of the G1 and J3 cDNAs did not contain any LTR sequences, although inverted U3 sequences were present at the 3' end of the B2 cDNA. These results indicate that, although the LTR sequences are not critical to the integration of the *neo* cDNAs, their inclusion may promote a hybrid integration pathway, with both retroviral integrative and host illegitimate recombination mechanisms operating. G1 contained both a complete coding *neo* gene and a portion of a second *neo* gene. The 5' end of the truncated *neo* gene in G1 was joined to the 3' end of the complete gene just 3' of a consensus polyadenylation site, suggesting the possible involvement of two RNA molecules in the recombination event. None of the *neo* retrogenes contained 3'-terminal poly(dA), although in each case the plasmid-derived *neo* sequences terminated shortly after a consensus polyadenylation site (AATAAA or CATAAA) derived from the plasmid (50). The 3' ends of the complete *neo* sequences in B2 and in G1 occurred at the same site.

Different directly repeated quail sequences flank integrated B2 and G1 *neo* cDNAs. Comparison of the sequences of the B2 *neo* cDNA integration site to the B2 preintegration site revealed that B2 contained an inverted 25-bp repeat internal to and composed of the ends of the pRSVneo-derived sequence (Fig. 3). It is interesting that the first two bases (AA) at the B2 cDNA 5' end were not included as part of the inverted repeat; these are the same two bases that are

A.

B.

FIG. 4. Molecular structure of three *neo* retrogenes, B2, G1, and J3. (A) Relevant sequences derived from pRSVneo present in the *neo* retrogenes. +1, *neo* mRNA cap site; +1187, termination of *neo* translation. The numbers and closed circles above the diagram indicate the 5' ends of the *neo* cDNAs in each of the retrogenes. The RSV LTR sequences are indicated by U3, R, and U5. The numbers in parentheses refer to the number of bases in each region; the U5 region of wild-type RSV is 80 bp. (B) Long vertical lines show recombination junctions between quail (Q) and retrogene (neo) sequences. The location of the promoters (P) and consensus polyadenylation signals (pA) of the *neo* transcriptional unit are shown. The numbers refer to breakpoints in the pRSVneo sequence shown in panel A. The open boxes indicate the fragment of the RSV LTR as shown in detail in panel A, with +1 indicating the LTR cap site. The asterisk shows the location of a novel 16-bp sequence present in the three retrogenes (see Fig. 5).

deleted upon normal avian retroviral integration (6, 20). The presence of the two A's at the 5' end of the B2 cDNA suggests that integration was not via the normal retroviral mechanism. Clone B2 was flanked by a 4-bp direct repeat of quail DNA (Fig. 3). This is reminiscent of the repeats found flanking many processed pseudogenes. However, as described above, the B2 short repeats were actually the remnants of two copies of a preexisting 40-bp repeat, one copy of which was partially deleted during integration of the cDNA (Fig. 2B).

Sequence analysis of the G1 clone showed that the integrated cDNA sequences were flanked by a 215-bp perfect direct repeat of quail DNA (Fig. 2C). Whereas the 3' repeat was contiguous with the 3' end of the G1 *neo* retrogene, the 5' repeat was separated from pRSVneo sequences by a 33-bp unique sequence of unknown origin. We found that this 215-bp repeat occurred approximately every 1,000 kb in the Japanese quail genome (data not shown). Retroviruslike families such as QR1, distantly related to chicken exogenous retroviruses, have been previously described in the Japanese quail genome (3, 11). Although it is possible that the 215-nucleotide repeat is an LTR-like sequence that is part of an endogenous retroviral family, it is not highly related (by DNA sequence similarity) to the QR1 family (data not shown).

Although clone J3 did not contain obvious direct repeats, we were unable to clone the preintegration site from genomic libraries. Southern blot analysis and analysis of several genomic phage clones that hybridized to either 3'- or 5'-flanking probes suggested that the 5' and 3' sequences surrounding the J3 *neo* gene were not closely linked in the quail genome (data not shown). These results suggest that a deletion or some other rearrangement may have taken place during cDNA integration in this clone.

Each retrogene contains an identical sequence at a recombination site that is not present before integration. Comparisons of the sequences present at the junctions between cDNA and host DNA in the three clones led to a striking observation. We found that each clone contained, at or near one recombination site, an identical 16-bp sequence (referred to as the 16-mer). The sequence TGCAATTAAAC CAGCT was found in all three of the retrogene clones (Fig. 4B and 5A), and sequence analysis with a 16-mer oligonucleotide as a sequencing primer confirmed that each clone contained only a single copy of the 16-mer (data not shown). In addition, the 16-mer was preceded by the sequence AAAATNNN in each clone. The 16-bp sequence was present in different locations in each retrogene clone (denoted by asterisks in Fig. 4B and 5B). In clones B2 and J3, the sequence was found at or near the 3' or 5' cDNA-quail DNA

A. AAAATNNNTGCAATTAAACCAGCT

B.

FIG. 5. Locations of a conserved 16-bp sequence found at a recombination junction in each *neo* retrogene. (A) Sequence of the conserved 16-mer. The perfectly conserved nucleotides are shown in boldface type. Shown in lightface type is a second block of conserved nucleotides separated by three nucleotides that vary. (B) Proposed origin of the components of the 16-mer in each retrogene. Boundaries between quail DNA (Q, single line) and *neo* cDNAs (boxed) are indicated. In B2, the 16-mer is located internally to a 25-bp inverted repeat, as discussed in the text. Two diagrams (I and II) show alternative origins for the 16-mer in B2. X, 7 bp of unknown origin. In G1, both the 5' and 3' ends of the 16-mer are shown as originating from pRSVneo sequences. In J3, the 5' end of the 16-mer is tentatively indicated to originate from quail DNA sequences, although this has not been verified by sequence analysis of the preintegration site. The sequence 5' of the breakpoint is not present in the sequence of pRSVneo in either orientation.

junctions, respectively. In clone G1, the 16-bp sequence was located within cDNA-derived sequences at the junction between the complete and truncated *neo* genes. This 16-bp sequence was not present in its entirety anywhere in the original pRSVneo plasmid, nor was it present in RSV or other sequenced avian retroviruses. Additionally, we used the 16-mer to search the GenBank data base, which includes the sequences of many pseudogenes; the 16-mer did not occur in these sequences. However, a caveat of this analysis is that there are as yet no examples of complete avian pseudogenes including both flanking sequences in any data base. Part of the 16-mer in each clone was apparently derived from plasmid sequence, but the portion of the 16-mer that was plasmid derived differed in each clone (discussed below). Interestingly, the 16-mer was in the same orientation relative to *neo* coding sequences in each clone.

Efficiency of retrofection is increased by the inclusion of the 16-mer or other sequences in the transfected plasmids. Because we found the 16-mer sequence in all of the pRSVneo retrofectants we analyzed, we reasoned that it might be important in one or more of the processes involved in the establishment of these retrogenes. To test this, we cloned a 40-bp oligonucleotide containing the 16-mer in both orientations (+ or -) into pCMVneo between the terminators of *neo* transcription and translation (Fig. 6; the + sign indicates the same orientation relative to *neo* transcription as found in G1, J3, and B2). pCMVneo was used in these experiments because it contains no retroviral sequences. Table 1 summarizes the results of retrofection experiments with pCMVneo, pCMVneo16+, and pCMVneo16-. We did not detect any transfer of G418 resistance with SE21Q1b cells transfected with pCMVneo without the 16-mer. Inclusion of the 16-mer

FIG. 6. Construction of pRSVneo and pCMVneo plasmids, which contain copies of a 16-bp sequence found at or near a recombination junction in three *neo* retrogenes. Details of the construction of these plasmids are included in Materials and Methods.

in the minus orientation allowed detection of G418-resistant colonies, whereas inclusion of the 16-mer in the plus orientation resulted in retrofection efficiencies that resembled those found with pRSVneo. Retrofection experiments with particles containing *neo* RNAs produced after transfection of pCMVneo into SE21Q1b cells that contained helper virus (ALV) were as efficient as those with pCMVneo16+ or pRSVneo in helper-virus-free SE21Q1b cells (discussed below).

We were interested to see whether other sequences might affect retrofection efficiency. We chose enhancer-promoter sequences to test, as well as an *E. coli* plasmid sequence thought to be functionally inactive in eucaryotic cells as a control (Fig. 7). We cloned the thymidine kinase promoter

and associated transcriptional control sequences from herpes simplex virus type I (27) into pCMVneo between the cytomegalovirus immediate-early promoter and the *neo* transcriptional unit (pCMVTKneo). Additionally, we cloned a 900-bp fragment that included the *E. coli* plasmid pACYC origin of replication into the same site as the 16-mer in pCMVneo. Neither of these sequences increased the efficiency of retrofection when compared with retrofection with pCMVneo alone. We did find, however, that inclusion of sequences other than the 16-mer could positively affect retrofection efficiencies. A 434-bp fragment containing two copies of the junction of ALV U3-U5 LTR sequences from a closed circular provirus was cloned into pCMVneo at the 16-mer site in the same orientation as *neo* transcription. The major RSV (and, by analogy, ALV) enhancers reside in the U3 region of this fragment (8, 25, 36). We were able to select G418-resistant colonies with this plasmid, although the retrofection efficiencies were lower than those with pRSVneo (Fig. 7).

Interestingly, when the 16-mer was cloned into pRSVneo, no stable retrofectants could be isolated (Table 1). The initial efficiencies of retrofection were nearly as high with this plasmid as with pRSVneo, but G418-resistant QT35 clones were small and reverted to G418 sensitivity. This phenomenon might be explained by the de novo formation of an additional 16-mer in these transient retrofectants and subsequent loss, through homologous recombination, of functional *neo* sequences.

DISCUSSION

Our analysis of three *neo* cDNAs transferred by retroviral infection as *neo* RNAs reveals that the integrated structures of the *neo* retrogenes are more complex than processed pseudogenes. The retrogenes also lack many of the hallmarks of naturally acquired processed pseudogenes. Although most processed pseudogenes possess 3'-terminal poly(dA), these sequences were not present in the three *neo* retrogenes analyzed. Many processed pseudogenes are flanked by directly repeated sequences of various lengths thought to be produced by repair DNA synthesis (13, 16). Two of the three *neo* retrogenes analyzed are also flanked by directly repeated sequences. In the case of G1, one repeat was separated from *neo* sequences by 33 bp. Additionally, the G1 repeats were longer than the repeats found flanking many processed pseudogenes. In B2, the repeats were short (4 bp), but we know that they were already present at the preintegration site. The presence of such repeats in the immediate vicinities of the ends of the *neo* cDNAs suggests that pairing of these repeats was involved (perhaps peripherally) with *neo* cDNA integration.

The structure of the inverted repeats at the ends of the B2 cDNA (Fig. 3) could be explained by a model in which a circular intermediate was involved in integration. Hairpin cDNA synthesis at the 5' end (similar to the hairpin synthesis proposed to occur during reverse transcription of cauliflower mosaic virus [7]) could generate an inverted repeat. If such a molecule were ligated to form a circle and then cut between the inverted repeats before integration (ligation and subsequent cleavage occurring at nonidentical sites), the structure seen in B2 could be formed. The deletion of the 5'-terminal two nucleotides at the 3' end of the inverted repeat is consistent with the involvement of the retroviral integrase protein (4, 34).

Insights into the mechanism of retrogene formation are supplied by retrogene G1, which was formed by two *neo*

TABLE 1. Retrofection efficiencies

Plasmid	Clone ^a	No. of colonies per ml ^b	Activity ^c
pCMVneo	Mass	<0.25	—
pCMVneo	T5	<0.1	—
pCMVneo16+	Mass	13.9	++
pCMVneo16+	T6	11.7	++
pCMVneo16-	Mass	1.3	+
pCMVneoLTRjxn	Mass	0.6	+
pCMVneoLTRjxn	3	1.7	+
pCMVTKneo	5	<0.4	—
pCMVTKneo	8	<0.4	—
pCMVneoSPo	A5	<0.4	—
pCMVneoSPo	B2	<0.4	—
pCMVneo	2C (ALV) ^d	25.5, 15.8, 7.0	++
pRSVneo	25	39.2, 19.0	++
pRSVneo16+	T3	<0.1	—
pRSVneo16+	T4	<0.12	—

^a In some instances, as denoted by letter and/or numeral, individual G418^r QT35 cell clones were isolated and analyzed; data from representative clones are presented. "Mass" indicates that pools of G418^r SE21Q1b cells from a single transfected dish were analyzed.

^b Figures of >1.0 indicate the numbers of G418^r QT35 colonies per ml of viral supernatant (collected from transfected SE21Q1b cells); multiple numbers indicate the results from independent retrofection experiments with virus from different cell passages. Figures of <1.0 indicate that no G418^r QT35 colonies were found in the amount of supernatant analyzed (e.g., <0.4 = <1 colony per 2.5 ml of supernatant in multiple infected plates).

^c Activity is empirically derived from the number of colonies obtained per milliliter.

^d SE21Q1b cells were infected with ALV when they were transfected with pCMVneo.

FIG. 7. Plasmids derived from pRSVneo and pCMVneo used in this study, and their activities in retrofection experiments. (A) Retroviral sequences important in replication and reverse transcription (PBS, primer binding site; P, promoter; pA, polyadenylation signal; cap, 5' terminus of retroviral RNA; E, enhancer) and viral RNA encapsidation (psi) are diagrammed. (B) Sequences important for *neo* transcription and translation in parental plasmids. Abbreviations are as above plus the following: TAG, terminator of translation. (C) Plasmid derivatives. Details of the construction of these plasmids are given in Materials and Methods. Retrofection: +, G418^r QT35 colonies could be detected after infection of QT35 cells with virus produced by SE21Q1b Cl.25 cells transfected with the indicated plasmid; -, G418^r was not transferred.

genes. The internal location of the 16-mer in G1 (and perhaps in B2) could indicate that recombination occurred during the integration of a complete retrogene into the site of a previously integrated *neo* fragment that included the 16-mer in SE21Q1b cells during G418 selection. This is unlikely, given the low frequency of homologous recombination in higher eucaryotes (48), unless the 16-mer is strongly recombinogenic. Alternatively, it is possible that the G1 recombinant was generated during reverse transcription in a particle that contained two different encapsidated *neo* mRNAs. Data suggest that RNAs are encapsidated as multimers, perhaps as dimers, in SE21Q1b cells (24). We are currently analyzing the structures of retrogenes generated from infections by particles that have encapsidated heterologous RNAs (i.e., a

neo mRNA plus a retroviral genomic mRNA). Preliminary data suggest that such recombinations do occur during reverse transcription and, further, that they may involve short regions of identity between the two heterologous RNAs. In fact, the high levels of retrofection previously reported to have occurred with SE21Q1b cells transfected with pCMVneo (22) occurred because of the presence of ALV (helper virus) in the SE21Q1b cells. All of the pCMV *neo* retrogenes analyzed from the previous study have acquired LTR and other sequences from ALV. We surmise that copackaging of *neo* RNA and a retroviral genome leads to recombination. This process is currently under investigation.

The most striking feature of the three retrogenes is the

presence of an identical 16-bp sequence found near a recombination junction in each case. We found that breakpoints in the *neo* cDNAs occurred in similar locations in each clone (Fig. 4). We suspect this conservation is significant because each of these breakpoints is able to donate a segment of the 16-mer. The origin of the G1 retrogene 16-mer sequence can be precisely determined, as a large portion of the sequence was derived from the two adjacent *neo* cDNA gene sequences (Fig. 5B). The central three deoxynucleotides (AAA) were not present in either *neo* sequence at the sites of recombination. Interestingly, the three A's occurred just downstream from a consensus polyadenylation site. These nucleotides may have arisen from a polyadenylated *neo* mRNA that recombined (during reverse transcription or as a cDNA copy) with a second truncated copy of *neo*. Such an event could have occurred in the parental SE21Q1b cells, since we have found that acquisition of new RNA-mediated cDNA integrations is an ongoing process in these cells (unpublished data).

Because we have cloned the preintegration site of retrogene B2, we can also deduce the origins of the 16-mer in this clone. The first 7 bp (TCGAATT) are derived from the pRSVneo sequence, and the last two (CT) could be from the 3' 25-bp inverted repeat sequence (Fig. 5B). The central sequence of 7 bp (denoted by $\times$ in Fig. 5B, B2, I) does not match any sequence in its vicinity in either pRSVneo or preintegrative quail DNA sequences. These 7 bp may represent filler DNA, not unlike that proposed to explain the short blocks of unassignable nucleotides at many illegitimate recombination borders (reviewed in reference 33). Alternatively, with analogy to G1, the central three A's may be remnants of terminal poly(dA). If this is so, the next four nucleotides may have been derived from the 5' terminus of a second, essentially completely deleted, *neo* RNA (Fig. 5B, B2, II). The mechanism by which the B2 16-mer was created cannot as yet be determined; however, it seems clear that the process was more complex than simple integrative recombination.

We have no information about the origin of the 16-mer in retrogene J3. The complete 16-bp sequence appears to bridge the host-*neo* junction. As we have not yet cloned and analyzed the preintegration site of J3, however, we cannot be sure what sequences or events contributed to the genesis of the 16-bp sequence.

The 16-mer, when included in a plasmid that contains no retroviral sequences, increases the efficiency of retrofection in an orientation-dependent manner. We do not know how this effect is mediated. Retroviral sequences known to have transcriptional enhancer activity (pCMVneoLTRjxn) (Fig. 7 and Table 1) can also affect retrofection efficiency. The inclusion of these transcriptional enhancer sequences increases the retrofection efficiency to the level seen with pCMVneo16-. This effect seems to be restricted to retroviral enhancer sequences, because the retrofection efficiency is not increased by either the cytomegalovirus enhancers or the herpes simplex virus thymidine kinase promoter-associated transcriptional control sequences. However, in these cases the enhancers are 5' rather than 3' to the *neo* gene. Whether nonretroviral enhancer sequences placed 3' to *neo* allow transfer of G418 resistance has not yet been tested. We are currently testing whether the 16-mer has transcriptional enhancer activity, although how such an activity might be acting to increase retrofection efficiency remains obscure. As expected, random sequences such as the pACYC *ori* region do not increase retrofection efficiency.

A second way in which the 16-mer might be acting is via an

involvement in cDNA integration. It is conceivable that the 16-mer functions like a bacteriophage lambda chi sequence to stimulate recombination in neighboring regions of the DNA (42). If this is so, interesting questions are raised by its location in retrogene G1, where it appears between an expressed *neo* coding sequence and a truncated *neo* fragment rather than at a boundary between flanking and *neo* sequences, as in J3.

A third possibility is that the 16-mer acts as a primer-binding site during reverse transcription. Normally, retroviral genomic mRNAs are reverse transcribed via characteristic tRNA primers. The reverse transcriptase encoded by the retroviral *pol* gene and the retroviral primer-binding site determine which tRNA primer will be used (31, 38). Previous experiments by Taylor and Cywinski (45) have characterized the primer for the reverse transcriptase in SE21Q1b virions as tRNA-like in size but not as extensively base paired as the primer (tRNA^{Trp}) for wild-type RSV reverse transcription. Interestingly, the extended 16-mer common to each *neo* retrogene shares sequence similarities to a large number of tRNAs from diverse organisms, most strikingly to two short blocks of highly conserved nucleotides within the 18 nucleotides at the 3' termini of the tRNAs. This is the region of the tRNA that binds the retroviral primer-binding site and allows the initiation of reverse transcription of the minus strand of the proviral cDNA (44, 46). We are replacing the wild-type primer-binding site in RSV with an oligonucleotide that includes the 16-mer to test whether this sequence can function as a primer-binding site *in vivo*. The 16-mer, if present in the encapsidated *neo* mRNA, may thus be able to function as a tRNA-like primer-binding site to allow the efficient production of *neo* retrogene cDNAs.

ACKNOWLEDGMENTS

We thank Mark Groudine, Harvey Eisen, Harold Weintraub, Meng-Chao Yao, Susan Amundsen, Nancy Hutchison, and members of our laboratory for thoughtful comments and advice.

This work was supported by Public Health Service grant CA 18232 from the National Cancer Institute. R.A. was supported by institutional training grant funds through the University of Washington (Public Health Service grant GM07270 from the National Institute of General Medical Sciences).

LITERATURE CITED

- Boeke, J. D., and V. G. Corces. 1989. Transcription and reverse transcription of retrotransposons. *Annu. Rev. Microbiol.* 43: 403-434.
- Borst, P., and G. Cross. 1982. Molecular basis for trypanosome antigenic variation. *Cell* 29:291-303.
- Boyce-Jacino, M. T., R. Resnick, and A. J. Faras. 1989. Structural and functional characterization of the unusually short long terminal repeats and their adjacent regions of a novel endogenous avian retrovirus. *Virology* 173:157-166.
- Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop. 1989. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc. Natl. Acad. Sci. USA* 86:2525-2529.
- Chen, C., and H. Okayama. 1987. High efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.* 7:2745-2752.
- Colicelli, J., and S. P. Goff. 1988. Sequence and spacing requirements of a retrovirus integration site. *J. Mol. Biol.* 199:47-59.
- Covey, S. N., and D. S. Turner. 1986. Hairpin DNAs of cauliflower mosaic virus generated by reverse transcription *in vivo*. *EMBO J.* 5:2763-2768.
- Cullen, B. R., K. Raymond, and G. Ju. 1985. Functional analysis of the transcriptional control region located within the avian retroviral long terminal repeat. *Mol. Cell. Biol.* 5:438-447.
- de Saint Vincent, B. R., and G. M. Wahl. 1983. Homologous

- recombination in mammalian cells mediates formation of a functional gene from two overlapping gene fragments. *Proc. Natl. Acad. Sci. USA* **80**:2002–2006.
10. Dornburg, R., and H. M. Temin. 1988. Retroviral vector system for the study of cDNA gene formation. *Mol. Cell. Biol.* **8**: 2329–2334.
 11. Dunwiddie, C., and A. J. Faras. 1985. Presence of retrovirus reverse transcriptase-related gene sequences in avian cells lacking endogenous avian leukosis viruses. *Proc. Natl. Acad. Sci. USA* **82**:5097–5101.
 12. Folger, K. R., E. A. Wong, G. M. Wahl, and M. R. Capecchi. 1982. Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. *Mol. Cell. Biol.* **2**: 1372–1387.
 13. Galas, D. J., and M. Chandler. 1981. On the molecular models of transposition. *Proc. Natl. Acad. Sci. USA* **78**:4858–4862.
 14. Gallis, B., M. Linial, and R. Eisenman. 1970. An avian oncovirus mutant deficient in genomic RNA: characterization of the packaged RNA as cellular messenger RNA. *Virology* **94**:146–161.
 15. Gorman, C., R. Padhamabhan, and B. H. Howard. 1983. High efficiency DNA-mediated transformation of primate cells. *Science* **221**:551–553.
 16. Harshey, R. M., and A. I. Bukhari. 1981. A mechanism of DNA transposition. *Proc. Natl. Acad. Sci. USA* **78**:1090–1094.
 17. Henikoff, S., and M. Eghtedarzadeh. 1987. Conserved arrangement of nested genes at the *Drosophila Gart* locus. *Genetics* **117**:711–725.
 18. Hicks, J. B., J. N. Strathern, and I. Herskowitz. 1977. The cassette model of mating type interconversion, p. 457–462. In A. Bukhari, S. Shapiro, and S. Adhya (ed.), *DNA insertion elements, plasmids and episomes*. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory, N.Y.
 19. Hood, L., M. Kronenberg, and T. Hunkapiller. 1985. T cell antigen receptor and the immunoglobulin supergene family. *Cell* **40**:225–231.
 20. Ju, G., and A. M. Skalka. 1980. Nucleotide sequence analysis of the long terminal repeat (LTR) of avian retroviruses: structural similarities with transposable elements. *Cell* **22**:379–386.
 21. Langridge, J., P. Langridge, and P. L. Berquist. 1980. Extraction of nucleic acids from agarose gels. *Anal. Biochem.* **103**: 264–271.
 22. Linial, M. 1987. Creation of a processed pseudogene by retroviral infection. *Cell* **49**:93–102.
 23. Linial, M., N. Gunderson, and M. Groudine. 1985. Enhanced transcription of *c-myc* in bursal lymphoma cells requires continuous protein synthesis. *Science* **230**:1126–1132.
 24. 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.
 25. Luciw, P. A., J. M. Bishop, H. E. Varmus, and M. R. Capecchi. 1983. Location and function of retroviral and SV-40 sequences that enhance biochemical transformation after microinjection of DNA. *Cell* **33**:705–716.
 26. Lutzker, S. G., and F. W. Alt. 1989. Immunoglobulin heavy chain class switching, p. 693–714. In D. E. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.
 27. McKnight, S. L., E. R. Gavis, R. Kingsbury, and R. Axel. 1981. Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. *Cell* **25**:385–398.
 28. 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.
 29. Moulding, C., A. Rapoport, P. Goldman, J. Battey, G. M. Lenoir, and P. Leder. 1985. Structural analysis of both products of a reciprocal translocation between *c-myc* and immunoglobulin loci in Burkitt's lymphoma. *Nucleic Acids Res.* **13**:2141–2152.
 30. Palmiter, R. D., and R. L. Brinster. 1985. Transgenic mice. *Cell* **41**:343–345.
 31. Peters, G. G., and J. Hu. 1980. Reverse transcriptase as the major determinant for selected packaging of tRNAs into avian sarcoma virus particles. *J. Virol.* **36**:692–700.
 32. Roth, D., and J. Wilson. 1988. Illegitimate recombination in mammalian cells, p. 621–653. In R. Kucherlapati and G. Smith (ed.), *Genetic recombination*. American Society for Microbiology, Washington, D.C.
 33. Roth, D. B., T. N. Porter, and J. H. Wilson. 1985. Mechanisms of nonhomologous recombination in mammalian cells. *Mol. Cell. Biol.* **5**:2599–2607.
 34. Roth, M. J., P. L. Schwartzberg, and S. P. Goff. 1989. Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. *Cell* **58**:47–54.
 35. Rowley, J. D. 1980. Chromosomal abnormalities in human leukemia. *Annu. Rev. Genet.* **14**:17–39.
 36. Ruddell, A., M. Linial, W. Schubach, and M. Groudine. 1988. Lability of leukosis virus enhancer-binding proteins in avian hematopoietic cells. *J. Virol.* **62**:2728–2735.
 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. Sawyer, R. C., and H. Hanafusa. 1979. Comparison of the small RNAs of polymerase-deficient and polymerase-positive Rous sarcoma virus and another species of avian retrovirus. *J. Virol.* **29**:863–871.
 39. Sharp, P. A. 1983. Conversion of RNA to DNA in mammals: *Alu*-like elements and pseudogenes. *Nature (London)* **301**:471–472.
 40. Shimoto, K., S. Mitzutani, and H. M. Temin. 1980. No apparent nucleotide sequence specificity in cellular DNA juxtaposed to retrovirus proviruses. *Proc. Natl. Acad. Sci. USA* **77**:7357–7361.
 41. Smith, A. J. H., and P. Berg. 1984. Homologous recombination between defective *neo* genes in mouse 3T6 cells. *Cold Spring Harbor Symp. Quant. Biol.* **49**:171–181.
 42. Smith, G. R. 1983. General recombination, p. 175–209. In R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg (ed.), *Lambda II*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 43. Smithies, O., R. G. Gregg, S. S. Boggs, M. A. Koralewski, and R. S. Kucherlapati. 1985. Insertion of DNA sequences into the human chromosomal B-globin locus by homologous recombination. *Nature (London)* **317**:230–234.
 44. Taylor, J. M. 1977. An analysis of the role of tRNA species as primers for the transcription into DNA of RNA tumor virus genomes. *Biochim. Biophys. Acta* **473**:57–71.
 45. 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.
 46. Varmus, H. E. 1983. Retroviruses, p. 411–503. In J. A. Shapiro (ed.), *Genetic elements*. Academic Press, Inc., New York.
 47. Varmus, H. E. 1987. Cellular and viral oncogenes, p. 271–346. In G. Stamatoyannopoulos, A. W. Nienhuis, P. Leder, and P. W. Majerus (ed.), *Molecular basis of blood diseases*. The W. B. Saunders Co., Philadelphia.
 48. Wasmuth, J. J., and L. V. Hall. 1984. Genetic demonstration of mitotic recombination in cultured Chinese hamster cell hybrids. *Cell* **36**:697–707.
 49. Weiner, A. M., P. L. Deininger, and A. Efstratiadis. 1986. Nonviral retrotransposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* **55**:631–661.
 50. Wickens, M., and P. Stephenson. 1984. Naturally occurring polyadenylation sequences in eukaryotic mRNAs. *Science* **226**: 1045–1051.
 51. Yao, M.-C. 1989. Site specific chromosome breakage and DNA deletion in ciliates, p. 715–734. In D. E. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.