MOLECULAR AND CELLULAR BIOLOGY, Jan. 1990, p. 68–74
0270-7306/90/010068-07$02.00/0

Vol. 10, No. 1

cDNA Genes Formed after Infection with Retroviral Vector Particles Lack the Hallmarks of Natural Processed Pseudogenes

RALPH DORNBURG† AND HOWARD M. TEMIN*

McArdle Laboratory for Cancer Research, University of Wisconsin, 450 North Randall Avenue, Madison, Wisconsin 53706

Received 8 August 1989/Accepted 26 September 1989

Retroviral proteins can encapsidate RNAs without retroviral *cis*-acting sequences. Such RNAs are reverse transcribed and inserted into the genomes of infected target cells to form cDNA genes. Previous investigations by Southern blot analysis of such cDNA genes suggested that they were truncated at the 3′ and the 5′ ends (R. Dornburg and H. M. Temin, Mol. Cell. Biol. 8:2328–2334, 1988). To analyze such cDNA genes further, we cloned three cDNA genes (derived from a hygromycin B phosphotransferase gene) in lambda vectors and analyzed them by DNA sequencing. We found that they did not correspond to the full-length mRNA: they were truncated at both the 3′ and the 5′ ends, did not contain a poly(A) tract, and were not flanked by direct repeats. The 3′-end junctions to chromosomal DNA of five more cDNA genes were amplified by polymerase chain reaction, cloned in pUC vectors, and sequenced. All of these cDNA genes had 3′-end truncations, and no poly(A) tracts were found. Further polymerase chain reaction experiments were performed to detect hygromycin B phosphotransferase cDNA genes with a poly(A) tract in DNA extracted from a pool of about 500 colonies of cells containing cDNA genes. No hygromycin B phosphotransferase cDNA gene with a poly(A) tract was found. Investigation of two preintegration sites by Southern analysis revealed that deletions were present in chromosomal DNA at the site of the integration of the cDNA genes. Naturally occurring processed pseudogenes correspond to the full-length mRNA, contain a poly(A) sequence, and are flanked by direct repeats. Our data indicate that cDNA genes formed by infection with retrovirus particles lack the hallmarks of natural processed pseudogenes. Thus, it appears that natural processed pseudogenes were not generated by retrovirus proteins.

The transposition of genetic material from one locus to another through an RNA intermediate is a powerful evolutionary mechanism for sequence duplication, dispersion, and rearrangements which contribute to the fluidity of eucaryotic genomes. DNA sequence data obtained from many laboratories suggest that 5 to 10% of the mammalian genome consists of sequences that arose by retroposition of cellular RNAs. Besides retroviruses and retrotransposons, these sequences mainly comprise short interspersed repeated sequences, long interspersed repeated sequences, and processed pseudogenes of protein-coding genes (2, 18, 24, 25, 27, 32, 34).

Aside from genetic lesions (e.g., base substitutions and frameshift mutations), most of these processed pseudogenes have four characteristics that distinguish them from their parental functional genes. First, they do not contain introns. The sequences 5′ and 3′ to intervening sequences are joined as in the mature mRNA of their functional counterparts. Second, they usually represent a full-length copy of the processed transcript from the functional gene. Third, they contain a poly(A) tract at the 3′ end. Fourth, they are flanked by 7- to 17-base-pair (bp)-long direct repeats that probably result from the mechanism by which they are inserted into the genome (27, 34).

Reverse transcription was first discovered as a key step in the replication of RNA tumor viruses (1, 22). Retroviruses (and most probably retrotransposons) have all *trans*- and *cis*-acting sequences required for active retroposition. They code for a reverse transcriptase, an integrase, and other proteins needed for the replication of retroviral RNA (17, 24, 28, 30). However, the *trans*- or *cis*-acting sequences that are involved in the retroposition of other polymerase II or III transcripts are not known, nor is the mechanism of their formation.

Recently, it was shown that cDNA genes could be formed after infection with retrovirus particles (6, 14). We described a retroviral vector system to investigate whether retroviral proteins are involved in the formation of processed pseudogenes of protein-coding genes (6). We introduced a gene unit consisting of a double promoter, a hygromycin B phosphotransferase (*hygro*) gene, and a polyadenylation processing site into retrovirus helper cells by means of a self-inactivating retroviral vector (Fig. 1; 7). RNAs (without retroviral *cis*-acting sequences) were transcribed from the internal promoters only. We showed that retrovirus vector particles encapsidate such RNAs (6). They were reverse transcribed and integrated into the genomes of infected target cells to form functional cDNA genes. An intron inserted between the double promoter was lost, establishing that an RNA intermediate had existed (R. Dornburg and H. M. Temin, J. Virol., in press). Similar results were established recently in another system (14). Investigation of eight processed pseudogenes by Southern blots indicated that all of them were truncated at the 3′ and 5′ ends (6; Dornburg and Temin, in press). To further investigate cDNA genes formed after infection with retrovirus particles, we investigated cloned cDNA genes by DNA sequencing and polymerase chain reaction (PCR).

* Corresponding author.

† Present address: Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, NJ 08854-5635.

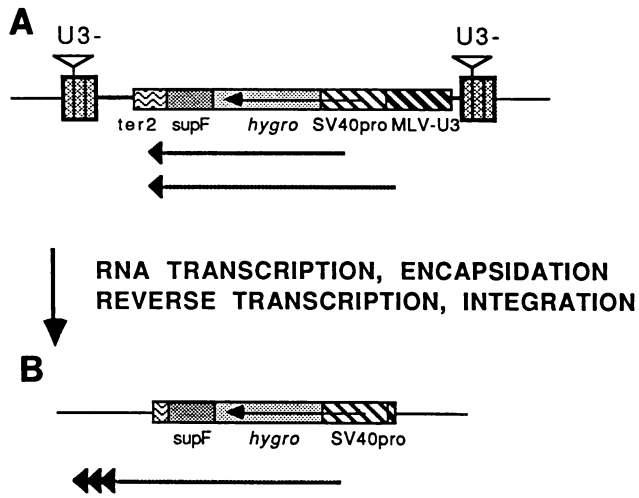

FIG. 1. Experimental protocol. A gene unit consisting of two promoters (SV40 and MLV U3), the *hygro* gene, and a polyadenylation signal (ter2) was introduced into a retrovirus helper cell by a retroviral vector that forms a provirus with U3-deleted long terminal repeats (A) (6, 7; Dornburg and Temin, in press). RNAs were transcribed from the internal promoters only. These RNAs lack all *cis*-acting sequences for retroviral replication. They were encapsidated into retrovirus particles supplied by the helper cell. Infection of fresh target cells resulted in the formation of cDNA genes (B). Functional cDNA genes (derived from the transcript starting from the MLV U3 promoter) were transcribed from the SV40 promoter, enabling the selection of hygromycin-resistant cell colonies (6). Polyadenylation is dependent on the presence of a polyadenylation site in the chromosomal DNA. ter2, Polyadenylation processing site of the herpes simplex virus thymidine kinase gene; supF, suppressor tRNA gene of *Escherichia coli*; *hygro*, hygromycin B phosphotransferase gene of *E. coli*; SV40pro, promoter of the SV40 early genes; MLV-U3, promoter of the MLV U3 region; U3-, deletion of the U3 region in the long terminal repeats. Arrows indicate RNA transcripts.

MATERIALS AND METHODS

Nomenclature and cells used. *hygro* is the hygromycin B phosphotransferase gene isolated from pLg89 (10). All vectors were derived from spleen necrosis virus, an avian reticuloendotheliosis virus. Experiments were performed on D17 cells (a dog cell line). Retrovirus helper cells (C3A2) were derived from D17 cells (33).

Plasmid constructions. All plasmids were described previously (6; Dornburg and Temin, in press) and were constructed by standard cloning procedures (15).

Southern blot analysis. Southern blots were performed by using standard procedures (15). DNA was transferred to nitrocellulose papers with a vacuum blotting system (Pharmacia-LKB).

Cloning of chromosomal DNA in lambda libraries. Chromosomal DNAs were partially digested with *Sau*3AI and were subjected to high-pressure liquid chromatography (HPLC) (LKB Instruments, Inc.) on a Genepack column (Waters Associates, Inc.). Fractions with fragments larger than 10 kbp were monitored by Southern analysis with a *hygro*-specific probe, and DNAs of positive fractions were ligated with EMBL4 lambda arms (digested with *Bam*HI purchased from Promega Biotec). A detailed protocol of the HPLC procedure will be published elsewhere (R. Dornburg, manuscript in preparation). The lambda libraries were screened by standard techniques (15). Fragments containing *hygro* cDNA genes flanked by chromosomal DNA were subcloned in pUC vectors for sequencing by standard techniques (15).

DNA sequencing. DNA sequences were determined in plasmid DNAs by using the dideoxy method (20, 31). Plasmids for sequencing were purified by HPLC (5) or with Geneclean (BIO 101).

PCR amplification. *Taq* polymerase was purchased from UBL or Beckman Instruments, Inc. Primers were purchased from Genetic Design. Reactions were carried out under salt conditions suggested by the supplier of the *Taq* polymerase and under the following temperatures: denaturation at 94°C for 2 min, annealing at 55°C for 2 min, and polymerization at 72°C for 5 min. The probes were cycled 50 times (9, 19).

Cloning of PCR-amplified DNA. The amplified DNA was treated with Klenow polymerase (Boehringer Mannheim Biochemicals) for 25 min at room temperature, extracted with phenol-chloroform, and precipitated with ethanol. The pellet was dissolved in water, adjusted to salt concentrations for restriction enzyme digestion as suggested by the supplier of the enzyme, and digested for 3 h at 37°C. The samples were extracted with phenol-chloroform, precipitated with ethanol, dissolved in water, and adjusted to T4 DNA ligase salt conditions. The DNA was treated with polynucleotide kinase (Boehringer) for 1 h. The samples were heated to 70°C for 10 min and cooled on ice, and alkaline phosphatase (Boehringer)-treated cloning vector (pUC) and DNA ligase were added. Ligation was done overnight at 14°C.

cDNA synthesis. In vitro cDNAs were prepared by standard techniques, using avian myeloblastosis virus reverse transcriptase (Boehringer) and oligo$(dT)_{12-18}$ (15).

RESULTS

Sequence analysis of three cDNA genes. To investigate in detail the structures of cDNA genes formed after infection with retrovirus particles, we made gene libraries of chromosomal DNAs isolated from cell clones containing a *hygro* cDNA gene. DNAs from *hygro*-positive bacteriophages were isolated and subcloned in pUC vectors, and the sequences of the junctions of the cDNA genes with the chromosomal DNA were determined (Fig. 2).

Naturally occurring processed pseudogenes of protein-coding genes usually correspond to full-length mature mRNAs, are flanked by direct repeats, and contain poly(A) sequences (27, 34). A hypothetical processed *hygro* pseudogene with these hallmarks and the structures of three *hygro* cDNA genes formed in our experiments are shown in Fig. 2a and b, respectively. The DNA sequences of the junctions of the cDNA genes and the chromosomal DNAs are shown in Fig. 2c. None of the *hygro* cDNA genes formed in our experiments corresponded to the full-length mRNA transcribed in the parental helper cells (Figure 1; 6; Dornburg and Temin, in press). They were truncated at the 5′ and 3′ ends. However, the truncations did not include the (second internal) simian virus 40 (SV40) promoter (Fig. 1). These data confirm our earlier finding obtained by Southern analysis (6; Dornburg and Temin, in press). Further, none of these cDNA genes was flanked by direct repeats, and none of them contained a poly(A) tract, as a result of the extensive 3′-end truncations. No consensus sequence between these three integrates could be found at the sites of integration. These data suggest that integration occurred randomly and was not mediated by the retroviral integrase, in agreement with our previous suggestion (6).

Southern blot analysis of approximately 500 cDNA genes. Formation of cDNA genes by retroviral infections is 7 to 8

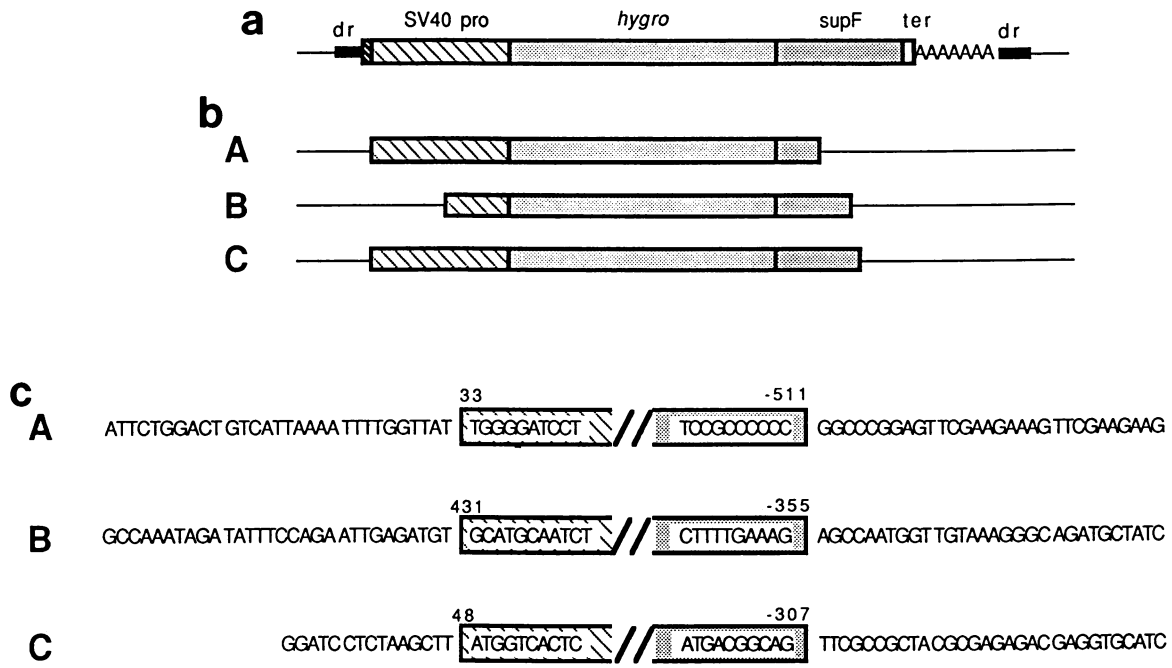

FIG. 2. Analysis of cDNA genes formed after infection with retrovirus vector particles. (a) Structure of a hypothetical cDNA derived from RD17 (see Fig. 1 and text). (b) Structures of three cDNA genes (A, B, and C) formed in our experiments. (c) Sequences of the junctions of these cDNA genes with the chromosomal DNA. Sequences of the cDNA genes are boxed. Each number indicates the number of the nucleotide in the parental RNA transcript from which the cDNA gene was derived. Negative numbers indicate the distance from the 3′ end [nucleotide position before the poly(A) addition site].

orders of magnitude less efficient than formation of a retroviral provirus (6). To increase the efficiency of cDNA gene formation, we constructed retroviral vectors with an additional packaging sequence between the two internal promoters. We found that the presence of the encapsidation sequence of murine leukemia virus (MLV) increased the formation of cDNA genes 100-fold (Dornburg and Temin, in press). Approximately 500 independent cell colonies containing *hygro* cDNA genes obtained in these experiments were pooled (from 10 different plates with about 50 colonies each), and chromosomal DNA was extracted and subjected to Southern analysis.

The DNA was digested with *Bam*HI plus *Hin*dIII, *Xba*I plus *Hin*dIII, and *Xba*I plus *Bgl*II. If the cDNA genes formed in these experiments corresponded to full-length mRNA, digestion with these enzymes and hybridization with a *hygro* probe should result in bands of 2.0, 1.1, and 1.5 kbp, respectively (the restriction enzyme map of a hypothetical full-length cDNA gene is shown in Fig. 3A). Analysis of chromosomal DNA isolated from the parental helper cell (from which the virus was harvested) and digested with these enzymes resulted in bands as expected (positive control; Fig. 3B, lanes a, c, and e, respectively). However, no bands could be detected when the chromosomal DNA of cells containing *hygro* cDNA genes was digested with *Bam*HI plus *Hin*dIII and with *Xba*I plus *Bgl*II (Fig. 3B, lanes b and f, respectively). These data indicate that the majority of the cDNA genes were truncated at the 5′ and 3′ ends, in agreement with our previous findings (6; Dornburg and Temin, in press). Further, these data establish that the pool of colonies was derived from independent clones. This conclusion is based on the finding that no clonal bands were observed with these enzymes. Since we selected for *hygro* expression in our experiments, an intact *hygro* gene (*Xba*I-to-*Hin*dIII fragment) was expected to be present in all cells. The *Xba*I-plus-*Hin*dIII-digested DNA gave a 1.1-kbp band, as expected (Fig. 3B, lane d).

Analysis of 3′ ends of PCR-amplified cDNA genes. The detection of a single or a few full-length cDNA genes in a pool of about 500 by Southern blot analysis is below the sensitivity level of this technique. To investigate whether some of the cDNA genes formed by retroviral infections contained the complete 3′ end of the mRNA plus a poly(A) tract, we amplified the 3′-end junctions of *hygro* cDNA genes with the chromosomal DNA by PCR with two sets of primers that selected for cDNA genes extending beyond the *Bgl*II restriction enzyme site [located 145 bp upstream of the poly(A) addition site; Fig. 3A]. The experimental strategy for this DNA amplification (with one set of primers) is shown in Fig. 4. (The same experimental protocol was applied with a second set of primers that were specific for a region about 50 bp 5′ of primers 1 and 2 shown in Fig. 4. In this experiment, chromosomal DNA was digested with *Hha*I, ligated, and linearized with *Eco*RI before amplification by PCR.)

The application of this technique resulted in the isolation of five cDNA genes that contained the *Bgl*II site described above. However, none of these sequences contained a complete 3′ end of the corresponding mRNA, and none contained a poly(A) tract (Fig. 5). These data further indicate that cDNA genes formed after infection with retrovirus particles are truncated and do not contain a poly(A) tract.

To test this finding further, we performed additional PCR experiments with the chromosomal DNA isolated from the pool of about 500 cell colonies as follows. Two primers were synthesized. One contained an oligo(dT) tract followed by sequences specific for the 3′ end of the mRNA transcribed from the *hygro* gene (P1 in Fig. 6A); the other was specific for sequences located 96 to 118 bp upstream of the putative 3′ end of the putative mRNA (P2 in Fig. 6A). Under

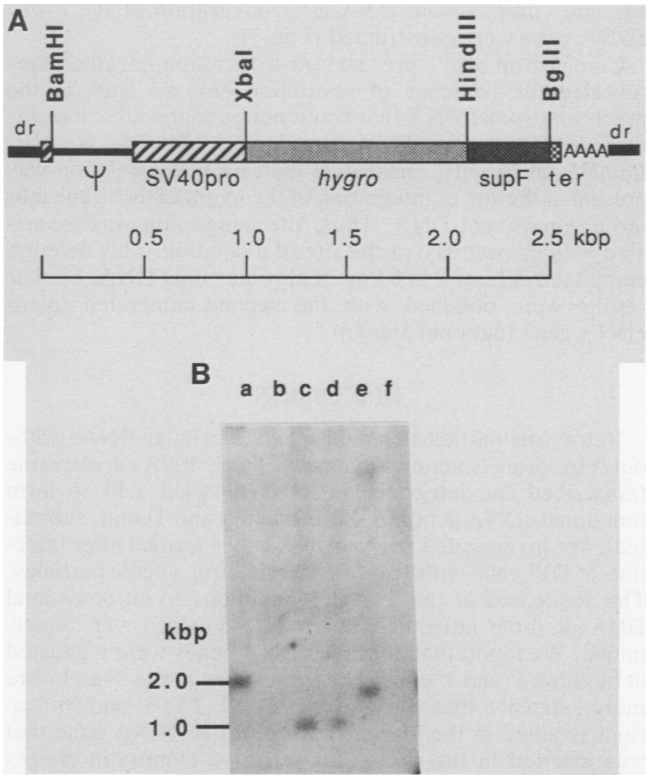

FIG. 3. Southern blot analysis of cDNA genes formed after infection with retroviral vector particles containing an additional encapsidation sequence of MLV (ψ). (A) Restriction enzyme map of a hypothetical full-length cDNA gene. (B) Analysis of DNA isolated from a pool of approximately 500 colonies containing *hygro* cDNA genes. Lanes: a and b, DNA digested with *Bam*HI plus *Hin*dIII isolated from the helper cell (lane a) and the pool of infected target cells (lane b); c and d, the same DNAs digested with *Xba*I plus *Hin*dIII; e and f, the same DNAs digested with *Xba*I plus *Bgl*II. Abbreviations are as in Fig. 1.

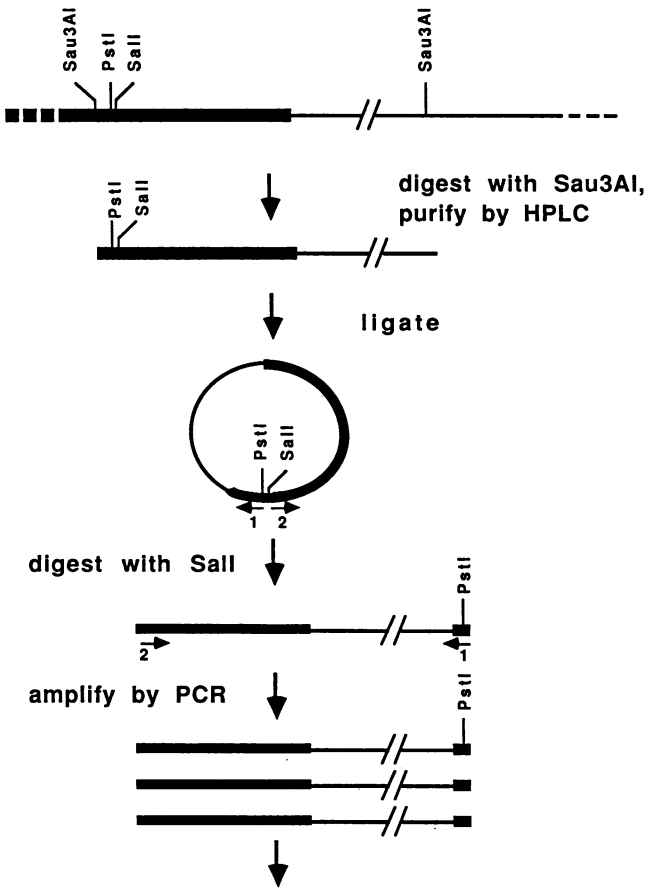

FIG. 4. Protocol for PCR amplification of integrated cDNA genes (21). Chromosomal DNA isolated from a pool of approximately 500 cell colonies carrying *hygro* cDNA genes was digested with *Sau*3AI and purified by HPLC from RNA and DNA fragments smaller than 50 bp. The DNA was circularized with T4 ligase and linearized with *Sal*I. The linear DNA was amplified by PCR with *hygro* cDNA gene-specific primers (arrows 1 and 2), digested with *Pst*I, and cloned into a *Sma*I-plus-*Pst*I-digested pUC vector for DNA sequencing. The *Sau*3AI (GATC) site in the *hygro* cDNA gene is part of the *Bgl*II (aGATCt) site shown in Fig. 3.

appropriate experimental conditions, polymerase-catalyzed reactions should have resulted in the amplification of a 127-bp fragment even if only a single colony contained a *hygro* cDNA gene with a poly(A) tract (Fig. 6B).

In our experiments, no amplification of a 127-bp fragment was observed (Fig. 6C, lane b). The bands longer than 200 bp are interpreted as unspecific amplifications of A+T-rich regions with primer 2. This interpretation is based on the finding that the same bands were also observed in a D17 control DNA without *hygro* cDNA genes (Fig. 6C, lane c). In further control experiments, RNA was isolated from virus particles that had packaged *hygro* mRNA. cDNA was made in vitro with avian mycloblastosis virus reverse transcriptase primed with oligo(dT). This minus-strand cDNA should contain all sequences corresponding to the 3′ end of the *hygro* mRNA plus an oligo(T) tract. Thus, hybridization of primer 1 (Fig. 6A) and DNA polymerase reaction should result in a double-stranded cDNA with a poly(A) tract. Therefore, PCR of this cDNA with primers 1 and 2 would result in the amplification of a 127-bp fragment. This band was observed in two different experiments with cDNAs made from RNA from virus particles isolated from two different helper cell lines (Fig. 6C, lanes a and d).

The results of these experiments show that (i) the *hygro* mRNA encapsidated in virus particles was processed as postulated, coinciding with our earlier observations obtained by Northern (RNA) blot analysis (6), and (ii) cDNA genes formed after infection with retrovirus particles do not contain a poly(A) tract.

Analysis of preintegration sites. The molecular cloning of three *hygro* cDNA genes in lambda vectors also enabled us to investigate the preintegration sites of the cDNA genes. First, unique sequences next to the integrated *hygro* cDNA gene were identified that were suitable as probes for the restriction enzyme mapping of chromosomal DNA by Southern blot analysis. Chromosomal DNA sequences next to the integrated *hygro* cDNA gene were subcloned in pUC plasmid vectors. The inserts were purified, labeled, and hybridized to genomic DNA. In the case of two lambda clones, unique sequences next to the integrated *hygro* cDNA gene were found. The third cDNA gene was integrated into repetitive DNA, and no unique sequences could be detected.

Chromosomal DNA of D17 cells, chromosomal DNA containing the integrated *hygro* cDNA gene, and DNA of the lambda clone were digested with several enzymes and

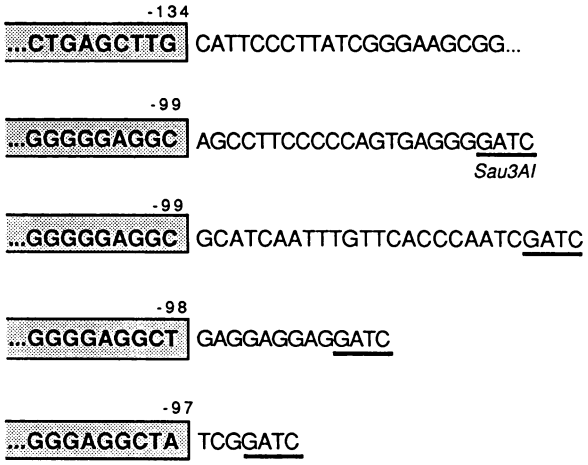

FIG. 5. Sequences of 3'-end junctions of *hygro* cDNA genes with chromosomal DNAs. PCR-amplified DNAs (see Fig. 4) were cloned into pUC vectors and sequenced. Sequences of the cDNA genes are boxed. Negative numbers indicate the nucleotide position in the parental RNA upstream of the poly(A) addition site. The *Sau*3AI site in the chromosomal DNA is underlined. For further details, see text.

enzyme combinations and subjected to Southern blot analysis. Hybridization of all filters was performed with the unique sequences described above. From the results of these experiments, restriction enzyme maps of the preintegration site and chromosomal DNA after integration of the *hygro* cDNA gene were constructed (Fig. 7).

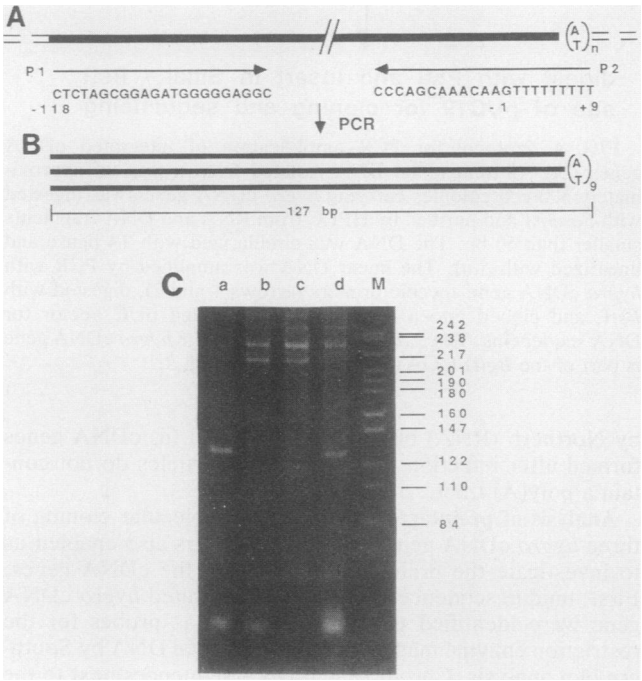

FIG. 6. PCR of *hygro* cDNA genes containing poly(A) sequences. (A) Sequence of a hypothetical *hygro* cDNA gene. (B) Sequences of primers 1 and 2 (P1 and P2). PCR of chromosomal DNA would result in the amplification of a 127-bp fragment if a full-length *hygro* cDNA gene with a poly(A) tract was present. (C) Result of the experiment outlined above. Lanes: a and d, amplification of in vitro-synthesized cDNA of RNA purified from virus particles; b, PCR of chromosomal DNA extracted from a pool of about 500 cell colonies with *hygro* cDNA genes; c, PCR of chromosomal DNA purified from D17 cells without *hygro* cDNA genes.

Comparison of the pre- and postrestriction enzyme maps revealed the presence of restriction enzyme sites in the prechromosomal DNA that could not be found after integration of the *hygro* cDNA gene (*Sac*I, *Eco*RI, *Hin*dIII, *Bam*HI, and *Kpn*I). These data indicate that a deletion was present at the site of integration of the *hygro* cDNA gene into the chromosomal DNA. Thus, the integration was associated with or occurred at the site of a deletion. This deletion comprised at least 7 to 8 kbp of chromosomal DNA. Similar results were obtained with the second integrated *hygro* cDNA gene (data not shown).

DISCUSSION

Retrovirus particles can encapsidate cellular RNAs without retroviral *cis*-acting sequences. These RNAs are reverse transcribed and integrated in infected target cells to form functional cDNA genes (6, 14; Dornburg and Temin, submitted). We investigated *hygro* cDNA genes formed after infection of D17 cells with spleen necrosis virus vector particles. The sequences of the 5' and 3' junctions to chromosomal DNA of three integrated *hygro* cDNA genes were determined. We found that all three cDNA genes were truncated at both the 5' and 3' ends. The truncations at the 3' end were more extensive than those at the 5' end. The 3'-end truncations resulted in the loss of the suppressor tRNA gene that was inserted in the vector for selective cloning in phages with amber mutations. Since we selected for functional *hygro* cDNA genes, we selected for an intact SV40 promoter. This most probably accounts for the finding that two *hygro* cDNA genes extended almost to the 5' end of the parental mRNA. In the case of one cDNA gene, the enhancer region of the SV40 promoter was lost (Fig. 2B). None of these three *hygro* cDNA genes contained a poly(A) tract, and none was flanked by direct repeats.

During integration of retroviral DNA into the host genome, direct 4- to 6-bp repeats that abut the integrated provirus are formed (30). The lack of direct repeats in cDNA genes formed in our experiments suggests that integration was not mediated by the retroviral integrase. This conclusion is further supported by the finding that deletions of chromosomal sequences are present at the site of integration of two *hygro* cDNA genes. Such deletions are not found after integration of retroviral DNA mediated by the viral integrase (30). This conclusion also agrees with our earlier proposal based on the comparison of growth curves of D17 cells infected with retrovirus particles containing a *hygro* mRNA and retrovirus vector particles containing an RNA with all retroviral *cis*-acting sequences (6). It might be postulated that a full-length cDNA gene was integrated and that deletions of chromosomal DNA after the integration event resulted in truncated cDNA genes. However, we consider it very unlikely that two independent deletions occurred in the cDNA genes investigated after the integration, since there is no special mechanism for such deletions.

Naturally occurring processed pseudogenes of protein-coding genes usually correspond to full-length mature mRNAs of their functional counterparts, contain a poly(A) tract, and are flanked by direct repeats (27, 34). However, some processed pseudogenes of protein-coding genes (approximately 5%) are truncated and do not have a poly(A) tract. These truncations and the lack of the poly(A) tract in such processed pseudogenes are considered to result from deletions that occurred after their formation later in evolution

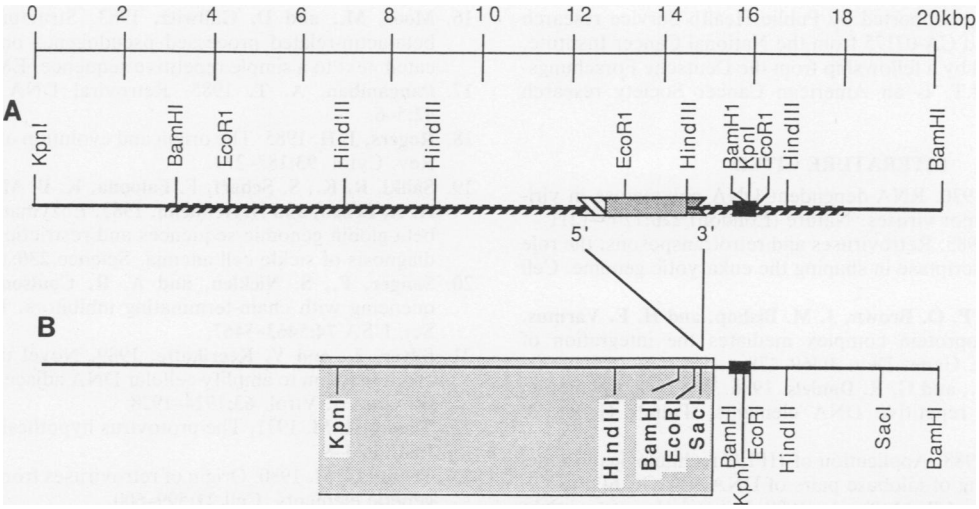

FIG. 7. Restriction enzyme maps of chromosomal DNA containing a cDNA gene and of the chromosomal DNA before integration of the *hygro* cDNA gene. The maps were obtained by digestion of chromosomal DNA and lambda clone DNA with several enzymes, followed by Southern analysis. (A) Restriction enzyme map of the chromosomal DNA with the integrated cDNA gene. The hatched area indicates the sequences that were cloned in lambda. The sequence used for hybridization is indicated by the solid box 3′ of the integrated *hygro* cDNA gene. (B) Restriction enzyme map of the chromosomal DNA before integration of the *hygro* cDNA gene. The restriction enzyme sites in the shaded box were not present after integration of the *hygro* cDNA gene. A scale in kilobase pairs is shown at the top.

(27, 34). We further tested whether 3′-end truncations and the lack of a poly(A) tract are general features of cDNA genes formed after retroviral infections.

Approximately 500 independent *hygro* cDNA genes in chromosomal DNA extracted from a pool of about 500 cell colonies with *hygro* cDNA genes were investigated by Southern blot analysis and PCR amplifications. Southern blot analysis indicated that at least 95% of the *hygro* cDNA genes did not contain a *Bgl*II site located 145 bp upstream of the putative poly(A) addition site (Fig. 3). This finding suggests that those cDNA genes had extensive 3′-end truncations, coinciding with our earlier observations. PCR with primers selecting for *hygro* cDNA genes that extended beyond the *Bgl*II site led to the isolation of the 3′ sequences of five more *hygro* cDNA genes. All of them revealed truncations at the 3′ end, and no poly(A) tracts were found. Further PCR experiments were performed to selectively amplify a *hygro* cDNA gene with a poly(A) tract in the DNA extracted from about 500 pooled cell colonies with *hygro* cDNA genes. No *hygro* cDNA gene with a poly(A) tract could be detected. (Similar PCR experiments proved to be sensitive enough to detect a single human immunodeficiency provirus in 1 of 10^5 infected cells [12].) Control experiments established that the *hygro* mRNA was correctly polyadenylated and that the failure of detection of a *hygro* cDNA gene with poly(A) was not due to inappropriate experimental conditions.

Our data indicate that cDNA genes formed after infection with retrovirus particles are truncated and do not contain a poly(A) tract. Thus, the mechanism of their formation appears different from that of formation of naturally occurring processed pseudogenes. A model for the formation of natural processed pseudogenes that accounts for the poly(A) tract has been proposed (27). This model suggests that natural processed pseudogenes are formed by the hybridization of the poly(A) tail of mRNA to a T-rich single-stranded region of chromosomal DNA after staggered chromosome breaks (27). The T-rich chromosomal DNA tail may serve as the primer, and cDNA synthesis of the RNA may take place at the site of the integration. Alternatively, the T tail of a minus-strand cDNA may hybridize to an A-rich region of chromosomal DNA (16). In contrast, recent data suggest that cDNA synthesis of retroviral RNA takes place in RNA-protein complexes (retroviral core structures) and that the double-stranded DNA copy of the retroviral RNA is integrated after the completion of DNA synthesis (3; J. Coffin, personal communication). Since the *hygro* mRNA was encapsidated into retroviral cores, the *hygro* cDNA was probably also synthesized in the core of the retroviral vector particle. However, as a result of the lack of a specific primer-binding site, the *hygro* cDNA might have been primed randomly. This hypothesis is supported by the finding that all *hygro* cDNA genes investigated were truncated, and no consensus sequence at the integration sites was found. Alternatively, the truncations may have occurred as a result of unspecific integration of the double-stranded *hygro* cDNA.

In summary, cDNA genes formed after infection with retrovirus particles lack the hallmarks of naturally occurring pseudogenes. Thus, we conclude that retroviral proteins were not involved in their formation. This conclusion is further supported by the recent discovery of a processed pseudogene in plants that revealed all of the hallmarks of pseudogenes found in animals (8). No genuine exogenous or endogenous retroviruses are known in plants, although some plant DNA viruses replicate through an RNA intermediate (30). The recent discovery of reverse transcriptase genes in bacteria (11, 13) suggests that this enzyme evolved very early in evolution (22, 23, 25). Thus, a cellular reverse transcriptase gene(s) may also be present in higher eucaryotes and may still be involved in shaping the genome by the retrotransposition of cellular RNAs. However, the 5% of natural processed pseudogenes without poly(A) may have a retroviral history and thus may represent a separate subclass of pseudogenes.

ACKNOWLEDGMENTS

We thank S. Hinz, J. Couch, and J. Rein for technical assistance. We thank M. Hannink, R. Risser, and G. Bhat for helpful comments on the manuscript.

This research was supported by Public Health Service research grants CA-22443 and CA-07175 from the National Cancer Institute. R.D. was supported by a fellowship from the Deutsche Forschungsgemeinschaft. H.M.T. is an American Cancer Society research professor.

LITERATURE CITED

1. **Baltimore, D.** 1970. RNA-dependent DNA polymerase in virions of RNA tumor viruses. Nature (London) **226**:1209–1211.
2. **Baltimore, D.** 1985. Retroviruses and retrotransposons: the role of reverse transcriptase in shaping the eukaryotic genome. Cell **40**:481–482.
3. **Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus.** 1989. A nucleoprotein complex mediates the integration of retroviral DNA. Genes Dev. **3**:469–478.
4. **Deininger, P. L., and G. R. Daniels.** 1986. The recent evolution of mammalian repetitive DNA elements. Trends Genet. **3**: 76–80.
5. **Dornburg, R.** 1988. Application of HPLC technologies in rapid DNA sequencing of kilobase pairs of DNA. LC.GC **6**:254–259.
6. **Dornburg, R., and H. M. Temin.** 1988. Retroviral vector system for the study of cDNA formation. Mol. Cell. Biol. **8**:2328–2334.
7. **Dougherty, J. P., and H. M. Temin.** 1987. A promoterless retroviral vector indicates that there are sequences in U3 required for 3′ RNA processing. Proc. Natl. Acad. Sci. USA **84**:1197–1201.
8. **Drouin, G., and G. A. Dover.** 1987. A plant processed pseudogene. Nature (London) **328**:557–558.
9. **Faloona, F., and K. B. Mullis.** 1987. Specific synthesis of DNA in vitro via polymerase catalyzed chain reaction. Methods Enzymol. **155**:335–350.
10. **Gritz, L., and J. Davies.** 1983. Plasmid encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. Gene **25**:179–188.
11. **Inouye, S., M.-Y. Hsu, S. Eagle, and M. Inouye.** 1989. Reverse transcriptase associated with the biosynthesis of branched RNA-linked msDNA in Myxococcus xanthus. Cell **56**:709–719.
12. **Kumar, R., J. J. Goedert, and S. H. Hughes.** 1989. A method for the rapid screening of human blood samples for the presence of HIV-1 sequences: the probe shift assay. AIDS Res. Hum. Retroviruses. **5**:345–354.
13. **Lim, D., and W. K. Maas.** 1989. Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in E. coli B. Cell **56**:891–904.
14. **Linial, M.** 1987. Creation of a processed pseudogene by retroviral infection. **49**:93–102.
15. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Moos, M., and D. Gallwitz.** 1983. Structure of two human beta-actin-related processed pseudogenes one of which is located next to a simple repetitive sequence. EMBO J. **2**:757–761.
17. **Panganiban, A. T.** 1985. Retroviral DNA integration. Cell **42**:5–6.
18. **Rogers, J. H.** 1985. The origin and evolution of retroposons. Int. Rev. Cytol. **93**:187–279.
19. **Saiiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim.** 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230**:1350–1354.
20. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA-sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
21. **Silver, J., and V. Keerikatte.** 1989. Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. J. Virol. **63**:1924–1928.
22. **Temin, H. M.** 1971. The protovirus hypothesis. J. Natl. Cancer Inst. **46**:3–8.
23. **Temin, H. M.** 1980. Origin of retroviruses from cellular movable genetic elements. Cell **21**:599–600.
24. **Temin, H. M.** 1985. Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. Mol. Biol. Evol. **2(6)**:455–468.
25. **Temin, H. M.** 1989. Retrons in bacteria. Nature (London) **339**:254–255.
26. **Temin, H. M., and S. Mitzutani.** 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature (London) **226**:1211–1213.
27. **Vanin, E. F.** 1985. Processed pseudogenes. Characteristics and evolution. Annu. Rev. Genet. **19**:253–272.
28. **Varmus, H. E.** 1982. Form and function of retroviral proviruses. Science **216**:812–820.
29. **Varmus, H. E.** 1985. Reverse transcriptase rides again. Nature (London) **314**:583–584.
30. **Varmus, H., and P. Brown.** 1989. Retroviruses, p. 53–108. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
31. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
32. **Wagner, M.** 1986. A consideration of the origin of processed pseudogenes. Trends Genet. **5**:134–137.
33. **Watanabe, S., and H. M. Temin.** 1983. Construction of a helper cell line for avian reticuloendotheliosis virus cloning vectors. Mol. Cell. Biol. **3**:2241–2249.
34. **Weiner, A. M., P. L. Deininger, and A. Efstratiadis.** 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. **55**:631–661.