JOURNAL OF VIROLOGY, June 1996, p. 4063–4070 Vol. 70, No. 6
0022-538X/96/$04.00+0


# Sequence Tags of Provirus Integration Sites in DNAs of Tumors Induced by the Murine Retrovirus SL3-3

ANNETTE B. SØRENSEN,[1] MOGENS DUCH,[1] HENRIK W. AMTOFT,[1]
POUL JØRGENSEN,[1] AND FINN SKOU PEDERSEN[1,2]*

*Department of Molecular and Structural Biology[1] and Department of Medical Microbiology and Immunology,[2] University of Aarhus, DK-8000 Aarhus, Denmark*

Received 4 January 1996/Accepted 19 March 1996

**The murine retrovirus SL3-3 is a potent inducer of T-cell lymphomas when inoculated into susceptible newborn mice. The proviral integration site sequences were surveyed in tumor DNAs by a simple two-step PCR method. From 20 SL3-3-induced tumors a total of 39 provirus-host junctions were amplified and sequenced. Seven showed homology to known sequences. These included the known common integration site c-*myc* as well as genes not previously identified as targets of provirus integration, namely N-*ras* and the genes coding for major histocompatibility complex class II E-β, protein kinase C-η, and T-cell receptor β-chain. Among these genes, the integrations in c-*myc* as well as the one in N-*ras* were found to be clonal. One of the remaining 32 proviral integration site sequences that show no similarities to known sequences may represent a common integration site, as 2 of the 20 tumors demonstrated clonal provirus insertion into this region.**

Tumor induction by the nonacute transforming retroviruses is a multistep process which has not yet been clearly defined. However, several studies have provided evidence that retroviral insertional mutagenesis plays a central role in viral oncogenesis (9, 24). The effects of provirus integration vary depending on the particular location of the insertion site relative to the targeted gene; that is, integrated proviruses may affect the neighboring genes in a variety of ways by promoter insertion, enhancer insertion, and/or truncation of a normal cellular gene (9).

The normal approach when investigating the relationship of retroviral integrations to disease development has been cloning and hybridization to define a common viral integration site (1–3, 9, 12). In this study, proviral flanking sequences have been isolated by a relatively simple and fast PCR-based method (22). This approach should facilitate the identification of common integration sites as well as genes more rarely affected by an integrated provirus but that nevertheless could play a role in tumor development. Besides, ongoing genome analyses with the concomitant and quickly growing sequence databases require strategies by which provirus integration sites can be identified directly by sequence comparisons.

SL3-3 is a highly leukemogenic retrovirus which induces malignant lymphomas in 100% of inoculated mice of several strains, with latency periods of 2 to 5 months (6, 10, 17). From 20 tumors induced by SL3-3 in NMRI mice we have isolated a total of 39 provirus integration site sequences. From each tumor, one to five provirus-flanking cellular fragments were amplified, and the sequences of these fragments were compared with sequences in the GenBank and EMBL databases. As the c-*myc* and *pim-1* genes have been identified as integration sites in SL3-3-induced tumors, with c-*myc* being the one most frequently affected (20 to 25% of the tumors showed rearrangements in the c-*myc* gene) (7, 14), we expected to see insertions into these genes. Six of 20 tumors indeed showed rearrangements in the c-*myc* gene. However, we did not find any integrations into the *pim-1* gene either by the PCR method or by Southern blot hybridizations. Four provirus-flanking sequences showed homology to genes not previously identified as provirus integration sites, including N-*ras* and the genes coding for major histocompatibility complex (MHC) class II E-β (MHCII E-β), protein kinase C-η (nPKC-η), and T-cell receptor β-chain (TCR-β). The fact that the integration into N-*ras* was found to be clonal suggests that it may have a role in the development of this particular tumor. The integrations into nPKC-η, TCR-β, and MHCII E-β were not clonal, so if they play a role in tumorigenesis, their influence must be of a secondary nature. Yet, for studies of retroviral integration targets in general, such nonclonal integration sites may be valuable.

The remaining flanking sequences showed no homology to any known sequences. Still, as they all represent proviral integration sites, they should, like the nonclonal integrations into known genes, be useful in studies considering provirus integration in general. Besides, some of the sequences may represent unidentified proto-oncogene loci. One such candidate was discovered when Southern blot hybridizations with a preintegration site probe revealed rearrangements in 2 of 20 tumors.

## MATERIALS AND METHODS

**PCRs, Dynabead-streptavidin purification, and sequencing of PCR fragments.** These procedures have been described previously (22).

**Sequence comparison.** Nucleotide sequences were compared with sequences in the GenBank (release 92.0 [12.95]) and EMBL (release 45.0 [12.95]) databases by using the Wisconsin Package EGCG (version 8.1-beta-0.2, November 1995) *fasta* program with the parameter "word size: 6."

**Preparation of oligonucleotides.** Oligonucleotides (DNA Technology ApS, Aarhus, Denmark) were synthesized with an Applied Biosystems 381B DNA synthesizer and purified with Oligo Purification Cartridges.

**Tumors.** Tumors originated from experiments described by Hallberg et al. (6). Newborn NMRI mice were injected with SL3-3 virus. Control mice were mock injected with complete medium.

**Southern blotting and hybridization.** DNA was extracted from frozen tumor tissues. A 10-μg sample of each tumor DNA was digested with *Hin*dIII, and the products were separated by electrophoresis on 0.8% agarose gels. The fragments were then transferred to nylon membranes (Zeta-Probe; Bio-Rad) by alkaline blotting. The membranes were incubated twice in prehybridization (hybridization) buffer (1 mM EDTA, 0.5 M $Na_2HPO_4$ [pH 7.2], 7% sodium dodecyl sulfate [SDS]) at 65°C for 1 h. $^{32}$P-labeled probe ($10^6$ to $3 \times 10^6$ cpm/ml; random primed DNA labeling) (Boehringer Mannheim) was added to the hybridization buffer, and the membranes were incubated in this mixture for 18 to 24 h at 65°C. The

* Corresponding author. Mailing address: Department of Molecular Biology, University of Aarhus, C. F. Møllers Allé, Building 130, DK-8000 Aarhus C, Denmark. Phone: 45 8942 3188. Fax: 45 8619 6500. Electronic mail address: fsp@mbio.aau.dk.

membranes were washed twice for 1 h at 65°C in 0.1× SSC containing 0.1% SDS (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). Hybridization patterns were analyzed and reproduced with the PhosphorImager SF (Molecular Dynamics, Inc.).

**DNA probes.** The N-*ras* probe employed was a 0.9-kb *Pvu*II fragment from a plasmid clone of human N-*ras* (5). The ecotropic-virus-specific probe was a *gp70 Sma*I fragment from the Akv retrovirus (positions 6240 to 6570). The remaining probes were all prepared by PCR with the primers described below. The PCR amplification products were electrophoresed in a 1.5% (wt/vol) low-melting agarose (SeaPlaque; FMC) gel; fragments were excised from the gel and purified by phenol-chloroform extractions.

**Primer sequences.** The provirus-specific primers for amplification from the 5′ long terminal repeat (LTR) (primers 1, 2, and 4; see Fig. 1) have been described previously (22); hence, only the primers used for amplification from the 3′ LTR will be described herein.

For the first PCR, the primer sequence was 5′-TGCGGCCGCGATTCCCA GATGACCGGGGATC-3′ (located at positions 356 to 376 in the LTR; 1 = first base in U3, the underlined sequence; remainder of primer consists of linker sequences added for other purposes). For the second PCR, the primer sequence was 5′-TTAAACTAACCAATCAGCTCGCTTC-3′ (located at positions 391 to 415 in the LTR; 1 = first base in U3). The sequence of the sequencing primer was 5′-TCCGAATCGTGGTCTCGCTGATCCTTGG-3′ (located at positions 550 to 577 in the LTR; 1 = first base in U3). The sequences of c-*myc* primers *a* and *b* (see Fig. 2A) were 5′-TGTGTATGTATACGTTTGGGGATTGTAC-3′ and 5′-CACTCCAGCACCTCCGGTTCGGACT-3′, respectively. The sequences of the *pim-1* PCR probe primers were 5′-TACCTCTTACCTGCTGCT TCTCAAACT-3′ (positions 6091 to 6117) and 5′-TAAAGGGAATTGGTCAT AGTCAGGGAGT-3′ (positions 7190 to 7163). The positions indicated are from the sequence in the GenBank-EMBL databases (nucleotide sequence accession number M13945). (The PCR probe of 1.1 kb obtained by these primers extends from both sides of the *Hin*dIII restriction site located in the 3′-terminal *pim-1* exon. Hence, when used with *Hin*dIII-digested DNA, this probe will scan both the upstream and downstream regions in which proviral integrations have been found to cluster [21].) The sequences of PKC-η primers *a* and *b* (see Fig. 5) were 5′-CCCCGCGCGCTCTCAGAAGGACGAT-3′ and 5′-TGCGCAACAGCTC CTGGAACTGCAGCGT-3′, respectively. The sequences of the TCR-β (J2 region) primers (used for construction of the PCR probe and for verification of integration) were as follows: upstream primer, 5′-GCTTCTTGGCAACTG CAGCGGGGAGT-3′ (located at positions 1295 to 1320); downstream primer, 5′-CTGGGTCTCCAACACTGCTTCAAGTG-3′ (located at positions 2058 to 2033). (Positions in parentheses correspond to the sequence in the GenBank-EMBL databases [accession number K02802].) The sequence of the MHCII E-β primer (used for verification of the integration) was 5′-TTGACCCACCTGGGGGCCAA GAAAT-3′ (located at positions 1719 to 1695; nucleotide sequence accession number M37515). For the MHCII E-β PCR probe, the following primers were employed: 5′-CGGGCCGAGGTGGACACGGTGTGC-3′ (located at positions 204 to 227; nucleotide sequence accession number M28591) and 5′-TGCTCGGGCCAGT GAGGAGATCAGC-3′ (located at positions 45 to 21; nucleotide sequence accession number M37515).

**Nucleotide sequence accession numbers.** The nucleotide sequences of the provirus-flanking sequences have been assigned the EMBL data bank accession numbers Z69798 to Z69836 (see Table 1).

## RESULTS AND DISCUSSION

**Amplification and sequencing of DNA flanking integrated proviruses.** The overall strategy for amplifying the provirus-flanking cellular coherent fragments is illustrated in Fig. 1 (22). DNA containing integrated provirus(es) is amplified in two steps. In the first PCR step, one of the partly degenerate primers (FP, Fig. 1) is used, as well as a biotinylated provirus-specific primer. The amplified, biotinylated fragments are isolated with streptavidin-coated magnetic beads and used as templates in the second PCR. In the second PCR, a primer corresponding to the nondegenerate part of the flanking primer is biotinylated (primer 3, Fig. 1) and used with a nested provirus-specific primer. The isolated biotinylated fragments from this PCR are directly sequenced. To avoid amplification of internal proviral sequences from the other LTR, the DNA was digested before the first PCR (22). When amplification was from the 5′ LTR, the DNA was digested with *Afl*II, which cuts immediately upstream of the 3′ LTR, thereby preventing amplification of internal proviral fragments from the 3′ LTR. *Pvu*I was used to cut off the 5′ LTR when amplification was from the 3′ LTR.

The above method was used on DNAs from 20 SL3-3-induced tumors. SL3-3 murine leukemia virus (MuLV) was injected into newborn NMRI mice, causing malignant lymphomas (predominately thymic lymphomas) in all animals (6). The results of the PCR are summarized in Table 1. The amplified fragments varied in size between 400 and 2,000 bp, as determined by electrophoresis on agarose gels (data not shown). From each tumor, at least one of the flanking primers (FP, Fig. 1) used in the first PCR resulted in amplification of at least one provirus-cellular DNA junction fragment; a total of 39 different provirus-flanking DNA fragments from the 20 tumors were isolated, and their sequences were determined.

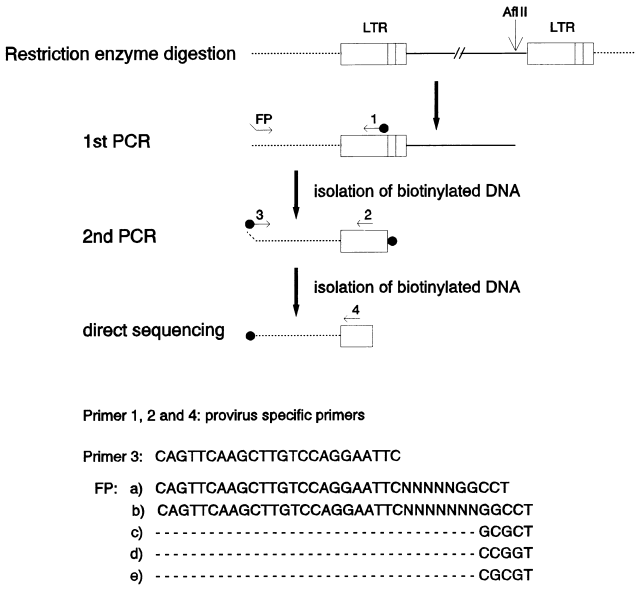


FIG. 1. Strategy for amplifying DNA flanking integrated proviruses. The figure shows amplification from the 5′ LTR. Amplification from the 3′ LTR follows the same principle. The integrated provirus is represented by the two boxes (the LTRs) and the solid line, while broken lines indicate unknown flanking DNA. The light vertical arrow marks the *Afl*II restriction site. The horizontal arrows indicate the oligonucleotide primers (FP [flanking primer] indicates that any of the FPs written below is used). A closed circle symbolizes biotin coupled to the 5′ end of the DNA.

The number of provirus-flanking cellular coherent fragments obtained in a given tumor does not necessarily reflect the number of integrated proviruses in that tumor. As seen in Table 1, the ratio of the number of PCR fragments to the number of ecotropic proviral integrations varied among the tumors, as deduced from Southern blot hybridizations (data not shown). The number of PCR fragments obtained depends on the accessibility of the PCR to the integrated proviruses. First, whether an integrated provirus will be amplified depends on the distance to the nearest cellular pentamer site corresponding to the 3′ end of the flanking primer; that is, for the PCR to take place, the flanking primer has to hybridize within a suitable range. We have overcome this problem somewhat by using flanking primers containing different 3′ ends (FP, Fig. 1), thereby increasing the probability that at least one of the flanking primers will hybridize within a suitable distance from the integrated provirus. By including more flanking primers, still more provirus-flanking cellular DNA junction fragments may be obtained from each tumor. Second, structural features of the flanking DNA may in some cases make it difficult to amplify the DNA by PCR, and ergo some of the proviral integration site sequences will not be amplified.

The sequence determination of the flanking DNA fragments

TABLE 1. Amplification of provirus integration sites in tumor DNAs employing the strategy shown in Fig. 1

| Tumor no. | Mouse age (days)[a] | No. of ecotropic proviral integrations[b] | Junction sequences[c] | | 4-bp repeat[d] | | Integration site[e] | Accession number[f] |
|---|---|---|---|---|---|---|---|---|
| | | | 5′ end | 3′ end | | | | |
| 1 | 57 | 2 | 1 | 0 | 5′ | TACA | n.h. | Z69830 |
| 2 | 71 | 2 | 1 | 1 | 5′ | ACTC | n.h. | Z69828 |
| | 71 | 2 | 1 | 1 | 3′ | GCTG | n.h. | Z69829 |
| 3 | 76 | 3 | 1 | 0 | 5′ | GTGG | c-*myc* promoter | Z69825 |
| 4 | 76 | 4 | 1 | 0 | 5′ | GTAC | n.h. | Z69826 |
| 5 | 82 | 1 | 2 | 0 | 5′ | CTTA | n.h. | Z69824 |
| | 82 | 1 | 2 | 0 | 5′ | ATAC | c-*myc* promoter | Z69827 |
| 6 | 90 | 2 | 1 | 1 | 5′ | CTTG | n.h. | Z69822 |
| | 90 | 2 | 1 | 1 | 3′ | CACT | n.h. | Z69821 |
| 7 | 91 | 4 | 1 | 1 | 5′ | NTAC | n.h. | Z69820 |
| | 91 | 4 | 1 | 1 | 3′ | TNTT | n.h. | Z69818 |
| 8 | 92 | 1 | 2 | 0 | 5′ | GTTG | n.h. | Z69816 |
| | 92 | 1 | 2 | 0 | 5′ | CAGC | n.h. | Z69823 |
| 9 | 95 | 1 | 1 | 0 | 5′ | ATTA | n.h. | Z69819 |
| 10 | 106 | 1 | 1 | 1 | 5′ | GCCT | n.h. | Z69817 |
| | 106 | 1 | 1 | 1 | 3′ | CTGT | n.h. | Z69814 |
| 11 | 110 | 1 | 0 | 1 | 3′ | CTCN | n.h. | Z69811 |
| 12 | 110 | 1 (2) | 3 | 2 | 5′ | CAAC | n.h. | Z69812 |
| | 110 | 1 (2) | 3 | 2 | 5′ | CTAT | TCR-β | Z69813 |
| | 110 | 1 (2) | 3 | 2 | 5′ | AAGG | MHCII E-β | Z69809 |
| | 110 | 1 (2) | 3 | 2 | 3′ | TNAC | n.h. | Z69806 |
| | 110 | 1 (2) | 3 | 2 | 3′ | TTCT | n.h. | Z69807 |
| 13 | 114 | 1 (2) | 0 | 2 | 3′ | AGTC | n.h. | Z69810 |
| | 114 | 1 (2) | 0 | 2 | 3′ | TTGG | c-*myc* promoter | Z69808 |
| 14 | 119 | 4 | 3 | 0 | 5′ | ATGG* | n.h. | Z69803 |
| | 119 | 4 | 3 | 0 | 5′ | ATGT | n.h. | Z69802 |
| | 119 | 4 | 3 | 0 | 5′ | AGCA | n.h. | Z69801 |
| 15 | 120 | 1 (1) | 3 | 1 | 5′ | GAGC* | n.h. | Z69804 |
| | 120 | 1 (1) | 3 | 1 | 5′ | CTAT | PKC-η | Z69805 |
| | 120 | 1 (1) | 3 | 1 | 5′ | CATC | n.h. | Z69798 |
| | 120 | 1 (1) | 3 | 1 | 3′ | GCTC* | n.h. | Z69799 |
| 16 | 124 | 5 | 1 | 0 | 5′ | AAAG | n.h. | Z69800 |
| 17 | 140 | 2 (1) | 2 | 1 | 5′ | CGGT | n.h. | Z69815 |
| | 140 | 2 (1) | 2 | 1 | 5′ | CCCT | n.h. | Z69831 |
| | 140 | 2 (1) | 2 | 1 | 3′ | TGTG | n.h. | Z69832 |
| 18 | 141 | 3 | 1 | 1 | 5′ | AACC | N-*ras* exon −1 | Z69833 |
| | 141 | 3 | 1 | 1 | 3′ | ACCA | n.h. | Z69834 |
| 19 | 249 | 3 | 1 | 0 | 5′ | NTAC | n.h. | Z69835 |
| 20 | 267 | 3 | 1 | 0 | 5′ | CTTC | n.h. | Z69830 |

[a] Age of mouse at death.
[b] The number of ecotropic proviral integrations as determined by Southern blot hybridizations. The numbers in parentheses indicate additional weakly hybridizing fragments.
[c] The number of flanking fragments amplified from the 5′ and 3′ LTRs.
[d] The 4-bp repeat at the site of integration, presented directly from the sequence of the second PCR product. An asterisk indicates that the same integration site was amplified with different flanking primers.
[e] All flanking sequences were compared with sequences in the GenBank and EMBL databases. n.h., no homology (no significant homology was found). However, a few flanking fragments contained stretches of repetitive sequences (tumors 7 [Z69820], 14 [Z69802], and 18 [Z69834]).
[f] The nucleotide sequences of the proviral flanking sequences have been assigned EMBL data bank accession numbers as indicated.

was done in one step with a sequencing primer located in the LTR, as shown in Fig. 1, and 450 bp was the maximum determination. However, for most of the fragments it would be possible to obtain longer sequences (up to about 1,800 bp) by including more sequencing steps. In a few cases, no more than about 50 bp could be determined because the flanking primer had hybridized very close to the proviral LTR. All the sequences were compared with sequences in the GenBank and EMBL databases as well as with each other.

**Clonal integrations into the c-*myc* gene.** Of the 39 provirus-flanking sequences, 3 revealed homology to the mouse c-*myc* promoter sequence (tumors 3, 5, and 13, Fig. 2A and Table 1). All three integrations were found within an 800-bp region of the c-*myc* promoter. Almost all proviral integrations that have been found within this region of the c-*myc* gene in transcriptional orientations opposite that of c-*myc* (2, 9, 11, 15). In our experiments, one of the integrations was found to be in the same transcriptional orientation as the c-*myc* gene while the other two were found to be in the opposite orientation (Fig. 2A).

To test the clonality of the c-*myc* integrations, Southern blot analyses were performed on *Hin*dIII-digested tumor DNAs with the c-*myc* promoter-specific PCR probe shown in Fig. 2A. For tumors 3 and 13, the Southern blots showed rearranged fragments of the expected sizes (6.3 and 7.0 kb, respectively; Fig. 2C). For tumor 5, two fragments of 3.4 and 5.6 kb were expected. However, one rearranged fragment of about 9 to 9.5 kb was observed (Fig. 2C). This could be explained if this particular provirus did not contain an internal *Hin*dIII restriction site, and in order to examine this, a PCR was done on DNA from tumor 5 with a provirus primer located 3′ of the internal *Hin*dIII site (indicated in Fig. 2B) together with c-*myc*

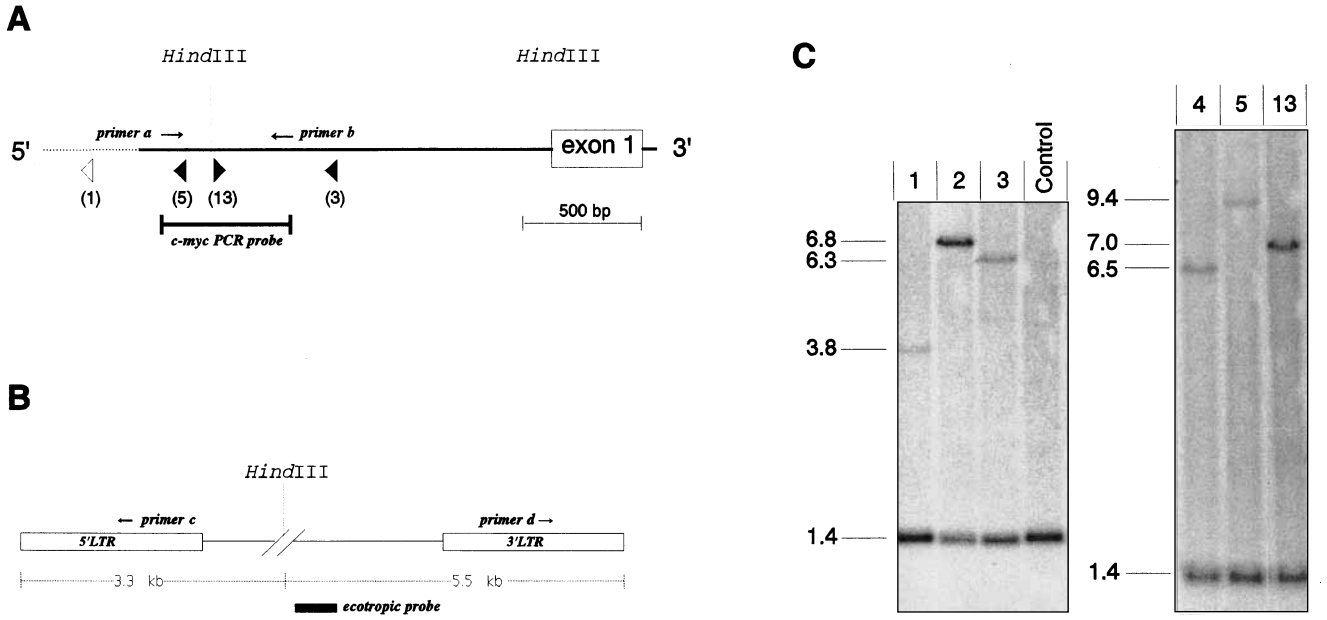


FIG. 2. Integrations into the c-*myc* gene. (A) The positions and orientations of the proviruses integrated into the c-*myc* promoter sequence are indicated by triangles. A closed triangle represents an integration determined by the two-step PCR method, while an open triangle represents an integration determined by Southern blot analyses (see text). The corresponding tumor numbers are given in parentheses. The broken line indicates the nonavailable sequence of the c-*myc* promoter. Primers *a* and *b* were used to verify the integrations. These primers were also employed to amplify the c-*myc* probe (indicated by the bar) which was used in Southern blot hybridizations. The *Hin*dIII restriction sites are shown. (B) The SL3-3 MuLV provirus. The LTR-specific primers (*c* and *d*) are indicated by the arrows. The *Hin*dIII cleavage site is shown. The location of the ecotropic-virus-specific probe used in Southern blot hybridizations is indicated. (C) Southern blot hybridizations with the c-*myc* PCR probe shown in panel A of *Hin*dIII-digested DNA from tumors (number indicated above each lane) and control DNA (liver DNA from mock-injected control mice). Fragment sizes (in kilobases) are indicated at the left.

primer *b* shown in Fig. 2A. The resulting fragment of about 4 kb could not be digested with *Hin*dIII (data not shown), thus supporting the lack of an internal *Hin*dIII site in this integrated provirus.

In addition, alterations in the c-*myc* region were seen in three cases in which no integrations into the c-*myc* gene were found by the two-step PCR method: tumors 1, 2, and 4 (Fig. 2C). This could be due to integrations in the c-*myc* gene outside the known sequences. Thus, homology searches in the databases would not detect these integration sites as c-*myc* sequences. Alternatively, the observed alterations do not result from proviral integrations but rather from nonviral chromosomal rearrangements in the c-*myc* locus. To examine this, the same Southern blots were hybridized with an ecotropic-virus-specific probe (the location of the probe is indicated in Fig. 2B). For tumors 1 and 4, the ecotropic-virus-specific probe did not hybridize to the rearranged c-*myc* fragments of 3.8 and 6.5 kb, respectively, while for tumor 2, this probe apparently did hybridize to the 7.3-kb rearranged c-*myc* fragment (data not shown). Additionally, for further examinations, combinatorial PCRs were done with c-*myc* primers *a* and *b* (Fig. 2A) and provirus-specific primers *c* and *d* (Fig. 2B). In the PCR amplifications of DNA from tumor 1, one combination, c-*myc* primer *b* and LTR primer *c*, provided a fragment of about 1,300 bp. This result, together with the hybridization data, confirms that in tumor 1 a provirus has integrated in the opposite transcriptional orientation about 1,800 bp upstream of c-*myc* exon 1, which is outside the known sequence of the c-*myc* promoter (Fig. 2A). In the PCRs of DNA from tumors 2 and 4 with the four pairwise combinations of primers, no amplification products were obtained, so whether the rearranged c-*myc* fragments observed in these tumors result from proviral integrations and/or nonviral chromosomal rearrangements cannot be ascertained.

**Clonal integration into the N-*ras* gene.** One of the 5′ LTR-flanking sequences from tumor 18 showed homology to the 5′ end of the mouse N-*ras* gene. To verify this integration, an N-*ras* primer located downstream of the integration site was constructed and used together with a provirus-specific primer in order to amplify the 3′ LTR-flanking sequence. The size and sequence of the resulting fragment (data not shown) confirmed that the provirus had integrated into the N-*ras* gene at the 3′ end of the first noncoding exon, exon −1 (Fig. 3A), in the same transcriptional orientation. To examine if this integration was clonal, Southern blotting was performed on *Hin*dIII-digested tumor DNAs with a human N-*ras* 5′ probe (5). As shown in Fig. 3B, the integration into the N-*ras* gene in tumor 18 was clonal, as a rearranged fragment of about 9.4 kb was seen. In the other tumor DNAs, no rearrangements of the N-*ras* gene were observed (data not shown). There have been no reports of proviral activation of N-*ras*. However, retroviral activation of the other *ras* genes, H-*ras* and K-*ras*, has been described (4, 8, 23). A proviral integration into the H-*ras* in a T-cell line from Moloney MuLV-induced tumors was shown to be associated with high-level expression of the H-*ras* locus (8). In this case, the provirus has integrated between the 5′ noncoding exon and the first coding exon in a cotranscriptional orientation. Overexpression of K-*ras*, due to a Friend MuLV provirus insertion, has been implicated in the transformation of a murine myeloid cell line (4, 23). Again, the provirus has integrated between the noncoding exon and the first coding exon in a cotranscriptional orientation. The sites of proviral integration found in H-*ras* and K-*ras* are thus similar to the integration site we have found in N-*ras*, indicating that in this case, overexpression of the N-*ras* gene may contribute to tumor development.

**Nonclonal integrations into the TCR-β- and MHCII E-β-encoding genes.** In tumor 12, two of the flanking sequences

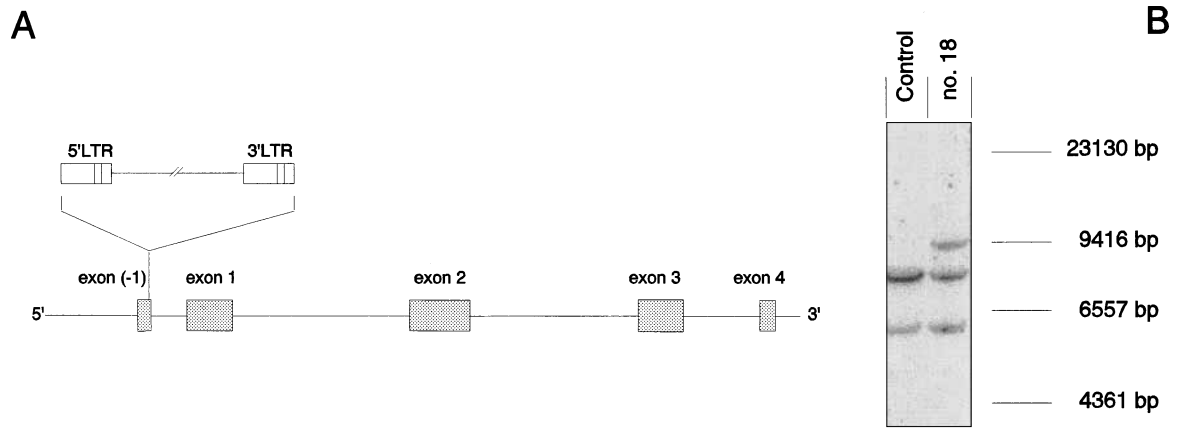


FIG. 3. Integration into the N-*ras* gene. (A) The mouse N-*ras* gene is shown with the location and orientation of the integrated provirus. (B) DNA from tumor 18 and control DNA (DNA from liver from mock-injected control mice) were digested with *Hin*dIII and hybridized with a human N-*ras* 5′ probe (5). Size markers are shown at the right.

showed homology to known sequences, namely, the genes encoding TCR-β and MHCII E-β. The locations and orientations of these integrations as well as the sequence alignments are shown in Fig. 4.

Alignment of the provirus-flanking sequence and the published MHCII E-β sequence reveals several gaps (Fig. 4B). To verify the integration, a primer from the published sequence downstream of the integration was constructed (positions are given in Materials and Methods). This primer was used together with a 5′ LTR-specific primer (primer 1, Fig. 1) in a

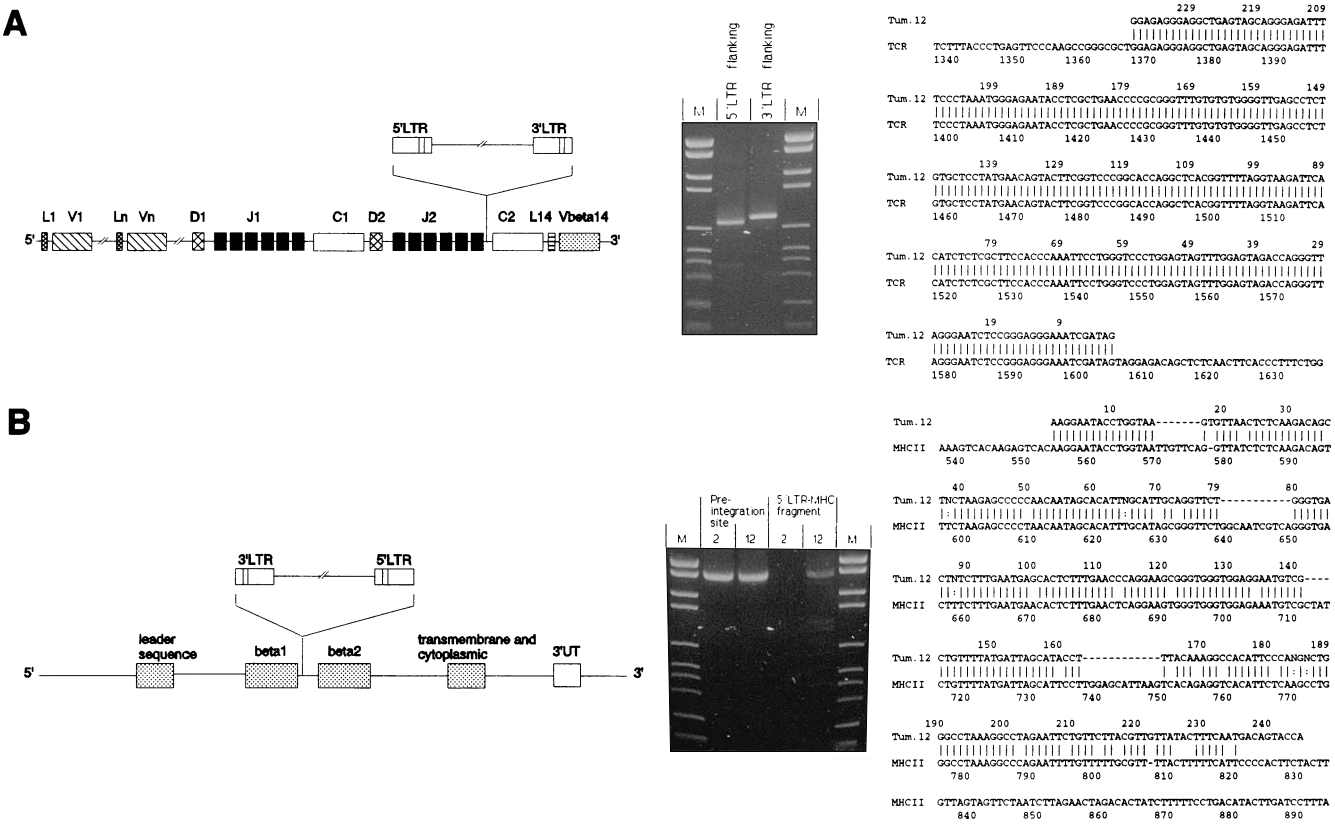


FIG. 4. Integrations into the TCR-β and MHCII E-β genes in tumor (Tum.) 12. (A) Left: orientation and integration site of the provirus in the mouse TCR-β gene, shown here in germline configuration. Center: PCR amplifications of the 5′ LTR– and 3′ LTR–TCR-β junction fragments (primer positions provided in Materials and Methods). M, molecular size markers (2,176, 1,766, 1,230, 1,033, 653, 517, 453, 394, 298, and 234/220 bp, respectively). Right: alignment of the provirus-flanking sequence (238 bp) and the TCR-β J2 sequence (numbers below refer to the GenBank-EMBL sequence, accession number K02802). (B) Left: orientation and integration site of the provirus in the mouse MHCII E-β gene. Center: PCR amplifications of the preintegration site fragment and the 5′ LTR-MHCII E-β junction fragment of DNA from tumors 2 and 12 (primer positions provided in Materials and Methods). M, molecular size markers as in panel A. Right: alignment between this provirus-flanking sequence (245 bp) and the MHCII E-β sequence (accession number M37515).

PCR of DNA from tumor 12, resulting in a fragment of the expected size of about 1,700 bp (Fig. 4B, 5′ LTR-MHC fragment, tumor 12). As a control, DNA from tumor 2 was used in a similar PCR; no fragment amplification occurred (Fig. 4B, 5′ LTR-MHC fragment, tumor 2). For further verification, the preintegration site fragments of the same DNAs were amplified, using primers constructed from the published sequence (positions are given in Materials and Methods). As seen in Fig. 4B, the expected fragment of about 1,670 bp was obtained in both cases. The observed gaps in the alignment may thus reflect sequence differences in this intron among different mice strains. The preintegration site fragment was used as a probe in a Southern blot hybridization on *Hin*dIII-digested tumor DNA, revealing that the integration was not clonal (data not shown).

The integration in TCR-β was verified by PCR with primers located upstream and downstream of the integration site (constructed from the published sequence; positions are given in Materials and Methods) as well as a 5′ LTR- and a 3′ LTR-specific primer, respectively (the LTR-specific primers were identical to those used in the first PCR of the two-step PCR). As seen in Fig. 4A, the expected fragments of 720 and 830 bp were obtained, thus confirming this integration. Employing the same TCR-β-specific primers, the preintegration site fragment was amplified and used as a probe in a Southern blot hybridization of *Hin*dIII-digested DNA, which showed that the integration was not clonal (data not shown).

Since the two integrations in tumor 12 were found to be nonclonal, no immediate indications for a possible role in tumorigenesis were provided. There have been no reports of proviral integrations into these genes. However, with respect to the TCR-β gene, it is tempting to speculate if this would occur in tissues or cell lines in which this gene is active in recombination or transcription. It has been shown that actively transcribed regions are preferred sites for proviral integrations (13, 18, 19, 25). The PCR strategy presented here would be a useful implement in such studies.

**Nonclonal integration into an nPKC-η gene.** For tumor 15 an integration into a protein kinase C gene was observed. A 5′ LTR-flanking sequence of about 450 bp showed homology in the last 200 bp to an nPKC-η gene which is predominantly expressed in lung and skin (16). This indicates an overlap between the provirus-flanking sequence and the published nPKC-η sequence (Fig. 5). To verify this overlap and rule out PCR artifacts, primers were constructed from the published sequence (primer *b*, Fig. 5A, and second underlined sequence in Fig. 5B) and from the obtained provirus-flanking sequence (primer *a*, Fig. 5A, and first underlined sequence in Fig. 5B), both located outside the overlap. These primers were used in PCRs of control DNA (DNA of liver from mock-injected mice) showing the expected fragment of 615 bp (Fig. 5B).

Southern blot hybridizations with a PCR probe constructed by primers *a* and *b* (shown in Fig. 5A) showed no rearranged nPKC gene fragments; thus, this integration was not clonal.

**Lack of integrations into the *pim-1* gene.** *pim-1* has been identified as a common integration site in tumors induced by SL3-3 (7, 14) as well as in other MuLV-induced tumors (3, 20). Accordingly, we expected to see proviral integrations into this gene. However, none of the flanking sequences showed homology to *pim-1* (Table 1). Therefore, in order to exclude the possibility that this indicates a deficiency of the method, we performed Southern blot hybridizations with a *pim-1* probe on *Hin*dIII-digested DNA from all the tumors. (The *pim-1* probe scans both the upstream and downstream regions in which provirus insertions have been found to cluster [3, 21]). No rearranged *pim-1* fragments were detected (data not shown), thus supporting the results obtained by PCR.

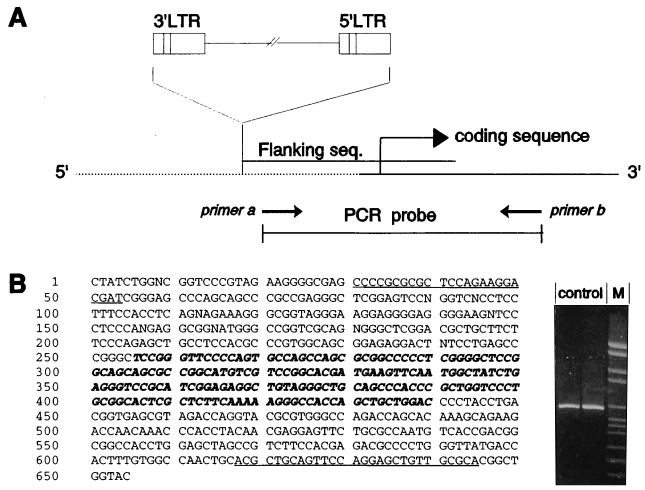


FIG. 5. Integration into the PKC-η gene in tumor 15. (A) The mouse nPKC-η gene is shown, along with the location and orientation of the integrated provirus. The dotted line represents the upstream nPKC-η sequence which is not available. The broken arrow shows the start point of the coding sequence (16). The extent of the determination of the provirus-flanking sequence is indicated. Primers *a* and *b* were used to verify the overlap. (B) Left: the sequence of the provirus-flanking DNA (1 to 440) overlaps with the published nPKC-η sequence (256 to 655+). The overlap is shown in boldface. The sequences of primers *a* and *b*, indicated in panel A, are underlined. Right: PCRs of liver DNA from two mock-injected control mice, employing primers *a* and *b*. M, molecular size markers (2,176, 1,766, 1,230, 1,033, 653, 517, 453, 394, 298, 234/220, and 154 bp, respectively).

**Integrations into unknown DNA sequences.** As seen in Table 1, most of the integration site sequences showed no similarities to any known sequences in the databases. Some of these sequences may represent integrations into already known genes or DNA regions; however, if they are located in introns or nontranscribed regions these integration sites may not be recognized, since such sequences most likely are underrepresented in the databases. Extension of available sequences, however, may reveal overlap between provirus-flanking sequences and known sequences, like in the case of the nPKC-η integration.

The sequences were compared with each other, but no overlaps were detected. Still, this does not rule out the possibility that some of the proviruses have integrated into the same region, as the sequence determinations of the flanking cellular DNA fragments are all less than 450 bp. Consequently, proviruses integrated by distances of 400 bp or more would not be detected in the comparison analyses.

One individual integration into an unknown DNA was examined. From tumor 15 both 5′ LTR- and 3′ LTR-flanking sequences from the same proviral insertion were obtained, as suggested by the 4-bp repeat at the site of integration (Table 1). To verify that these flanking sequences originated from the same integrated provirus, the preintegration site was amplified by using primers constructed from the 5′ and 3′ LTR-flanking sequences (Fig. 6A). The expected preintegration site fragment of 390 bp was obtained (data not shown), and this fragment was used as a probe in Southern blot hybridizations. As seen in Fig. 6B, the integration into this unknown DNA sequence in tumor 15 was clonal, as two rearranged fragments of about 13 and 10 kb were obtained. Interestingly, hybridizations of the preintegration site probe to DNA from the remaining tumors revealed a rearranged fragment of about 10 kb in

A

```
  1 CAGCGGGTCC CCTGTGGAGG GAAGGACCCT GTTACACAGT ACAGCCCCCC TAGAGTCACA CCATGGTACA CCAGCCCACA AAAGCCAAAC
 91 CTGAGAGCTG TAGCCAAGTG AGAATGCACA GGTGACAGAT GACTGCTGTG ACCCAAACTC TCCCCACATG GAGTTAAGGA CAGCAAGATG
181 GCGTTGGAGT CCCACAGGGT ATACTCCAGG TCAGGGGTCT TCTGAACTGA AGGAAGGGCT TAATCAACAC CACATGGCAC CATTTCACAT
271 GGCAAGTGTG AGCATGTGCA GGACCGTTAG AGCAACACAG CAAGCCCCAT GGGCAACACG AAGGGCCAGT GCCTGGGCTG GGGACCCTGG
361 AAGGAGGAAA AGATGTCCTG TGGATGCAGG GAATCANCCG CCTCCAGCCC AGTGATTCTC TCCCTCTGCT CCGCTCATCT CCATGGGTGT
451 TGGCATCTCA GCCCACTGTG GGCAACTGGT CACACTTGAG TCTCAGCCCA ACCGCTCTCA GGCCCTAGAA AGCCGACGAG TCAGTCACAG
541 CGCAGGGCAC AGCTGCTGCT TCCGGGGAGG CCGTGTCAGG CTCGAGGGTG ATTGGGAAGC AGCTGGAGCC CTGTGCCTTG TGGCAGCTAA
631 ATGACACTCG GTCTTACTCT GGTCTGGCCA GAGCAAGGGC GGNCTTGTCC TGCTTGCTCC CCCGGTGGCG GGGTGACCAT CTGAGCAGAT
721 GCTGCCCCTG CCTCTCCTGC CCAGGCC
```

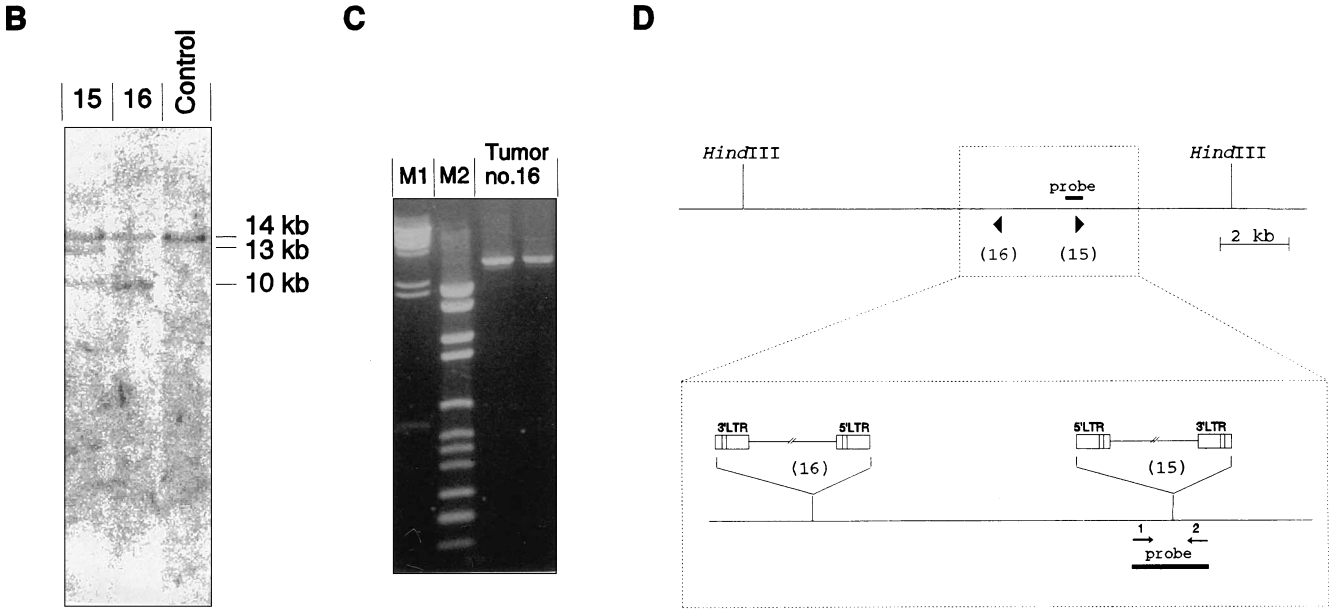


FIG. 6. Integration into unknown DNA. (A) Sequence of the preintegration site in tumor 15. Both 5′ LTR- and 3′ LTR-flanking sequences from the same site of integration were obtained by the strategy shown in Fig. 1. The 4-bp repeat at the integration site is shown in boldfaced italics. The sequences of the primers that were constructed to amplify the preintegration site fragment are underlined (primers 1 and 2, respectively, shown in panel D). (B) DNA from tumors 15 and 16 and control DNA (DNA from liver from mock-injected control mice) were digested with *Hin*dIII and hybridized with the preintegration site fragment. Size markers are shown at the right. (C) Two independent PCRs of DNA from tumor 16, with primer 2 (shown in panel D) and with the provirus-specific primer *c* (shown in Fig. 2B). M1 and M2, molecular size markers (M1: 27,000/23,130, 9,420, 6,560, 4,360, 2,320, 2,020, and 560 bp, respectively; M2: 2,176, 1,766, 1,230, 1,033, 653, 517, 453, 394, 298, 234/220, and 154 bp, respectively. (D) The relative location and orientation of the integrated proviruses in tumors 15 and 16, deduced from the Southern blot (panel B) and the PCRs (panel C). The preintegration site probe is indicated with a black bar. In the lower enlarged panel, the two primers, 1 and 2, that were used to amplify the preintegration site fragment are shown.

tumor 16 as well (Fig. 6B), indicating that a provirus in this tumor had integrated into the same region of unknown DNA. Combinatorial PCRs of DNA from tumor 16 were done with the primers shown in Fig. 6A and D and the provirus-specific primers shown in Fig. 2B. One combination, LTR primer *c* (Fig. 2B) and the primer constructed from the 3′ LTR-flanking sequence of tumor 15 (primer 2, Fig. 6D), provided a fragment of about 3 kb (Fig. 6C). This result, together with the rearranged fragment sizes observed in the Southern blot experiments, indicates that the proviruses in the two tumors have integrated in opposite orientations and that the two integration sites are separated by approximately 2.5 kb (Fig. 6D). These results strongly suggest that this unknown DNA sequence represents a common integration site.

**Sequence tag strategies.** Our results illustrate important aspects of a sequence tag analysis as the starting point for analysis of retroviral integration sites. Homology searches immediately identified integration sites in c-*myc*, N-*ras*, and the genes coding for MHCII E-β, TCR-β, and nPKC-η. While the role of clonal integrations in the upstream c-*myc* promoter region in tumor development has been extensively studied, the N-*ras* gene represents a previously unknown target for proviral insertion. However, analogy with earlier findings of corresponding integrations into H-*ras* and K-*ras* strongly suggests a role of this proviral integration in tumor development. The three remaining integrations into known sequences were all found to be nonclonal. Hence, they may represent purely random integrations without any role in tumor initiation, or they may contribute to secondary effects in tumor progression.

Sequence tags without homology to known sequences will be available in databases for studies of proviral integration sites of possible relevance for understanding oncogenesis as well as the general target specificity of the retroviral integration machinery. Moreover, such sequences may provide an entry to analysis of surrounding DNA with specific PCR primers and derived hybridization probes. The finding reported here of proviral integration into a common region in two tumors presents an example of the value of this approach.

## ACKNOWLEDGMENTS

We thank Lone Højgaard for technical assistance.

This work was supported by the Danish Cancer Society, the Karen Elise Jensen Foundation, the Danish Biotechnology Programme, the Danish Natural Sciences Research Council, and the Leo Nielsen Foundation. H.W.A. was the recipient of a student fellowship from the Danish Cancer Society.

## REFERENCES


1. **Ben David, Y., E. B. Giddens, and A. Bernstein.** 1990. Identification and mapping of a common proviral integration site Fli-1 in erythroleukemia cells induced by Friend murine leukemia virus. Proc. Natl. Acad. Sci. USA **87:** 1332–1336.
2. **Corcoran, L. M., J. M. Adams, A. R. Dunn, and S. Cory.** 1984. Murine T lymphomas in which the cellular myc oncogene has been activated by retroviral insertion. Cell **37:**113–122.
3. **Cuypers, H. T., G. Selten, W. Quint, M. Zijlstra, E. R. Maandag, W. Boelens, P. van Wezenbeek, C. Melief, and A. Berns.** 1984. Murine leukemia virus-induced T-cell lymphomagenesis: integration of proviruses in a distinct chromosomal region. Cell **37:**141–150.
4. **George, D. L., B. Glick, S. Trusko, and N. Freeman.** 1986. Enhanced c-Ki-ras expression associated with Friend virus integration in a bone marrow-derived mouse cell line. Proc. Natl. Acad. Sci. USA **83:**1651–1655.
5. **Hall, A., C. J. Marshall, N. K. Spurr, and R. A. Weiss.** 1983. Identification of transforming gene in two human sarcoma cell lines as a new member of the ras gene family located on chromosome 1. Nature (London) **303:**396–400.
6. **Hallberg, B., J. Schmidt, A. Luz, F. S. Pedersen, and T. Grundström.** 1991. SL3-3 enhancer factor 1 transcriptional activators are required for tumor formation by SL3-3 murine leukemia virus. J. Virol. **65:**4177–4181.
7. **Hays, E. F., G. Bristol, and S. McDougall.** 1990. Mechanisms of thymic lymphomagenesis by the retrovirus SL3-3. Cancer Res. **50:**5631S–5635S.
8. **Ihle, J. N., B. Smith-White, B. Sisson, D. Parker, D. G. Blair, A. Schultz, C. Kozak, R. D. Lunsford, D. Askew, Y. Weinstein, and R. J. Isfort.** 1989. Activation of the c-H-*ras* proto-oncogene by retrovirus insertion and chromosomal rearrangement in a Moloney leukemia virus-induced T-cell leukemia. J. Virol. **63:**2959–2966.
9. **Kung, H. J., C. Boerkoel, and T. H. Carter.** 1991. Retroviral mutagenesis of cellular oncogenes: a review with insights into the mechanisms of insertional activation. Curr. Top. Microbiol. Immunol. **171:**1–25.
10. **Lenz, J., R. Crowther, S. Klimenko, and W. Haseltine.** 1982. Molecular cloning of a highly leukemogenic, ecotropic retrovirus from an AKR mouse. J. Virol. **43:**943–951.
11. **Li, Y., C. A. Holland, J. W. Hartley, and N. Hopkins.** 1984. Viral integration near c-myc in 10–20% of mcf 247-induced AKR lymphomas. Proc. Natl. Acad. Sci. USA **81:**6808–6811.
12. **Liao, X., A. M. Buchberg, N. A. Jenkins, and N. G. Copeland.** 1995. *Evi-5*, a common site of retroviral integration in AKXD T-cell lymphomas, maps near *Gfi-1* on mouse chromosome 5. J. Virol. **69:**7132–7137.
13. **Mooslehner, K., U. Karls, and K. Harbers.** 1990. Retroviral integration sites in transgenic Mov mice frequently map in the vicinity of transcribed DNA regions. J. Virol. **64:**3056–3058.
14. **Morrison, H. L., B. Soni, and J. Lenz.** 1995. Long terminal repeat enhancer core sequences in proviruses adjacent to c-*myc* in T-cell lymphomas induced by a murine retrovirus. J. Virol. **69:**446–455.
15. **O'Donnell, P. V., E. Fleissner, H. Lonial, C. F. Koehne, and A. Reicin.** 1985. Early clonality and high-frequency proviral integration into the c-*myc* locus in AKR leukemia. J. Virol. **55:**500–503.
16. **Osada, S. I., K. Mizuno, T. C. Saido, Y. Akita, K. Suzuki, T. Kuroki, and S. Ohno.** 1990. A phorbol ester receptor/protein kinase, nPKC-eta, a new member of the protein kinase C family predominantly expressed in lung and skin. J. Biol. Chem. **265:**22434–22440.
17. **Pedersen, F. S., R. L. Crowther, D. Y. Tenney, A. M. Reimhold, and W. A. Haseltine.** 1981. Novel leukaemogenic retroviruses isolated from cell line derived from spontaneous AKR tumor. Nature (London) **292:**167–170.
18. **Rohdewohld, H., H. Weiher, W. Reik, R. Jaenisch, and M. Breindl.** 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. J. Virol. **61:** 336–343.
19. **Scherdin, U., K. Rhodes, and M. Breindl.** 1990. Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. **64:** 907–912.
20. **Selten, G., H. T. Cuypers, and A. Berns.** 1985. Proviral activation of the putative oncogene Pim-1 in MuLV induced T-cell lymphomas. EMBO J. **4:** 1793–1798.
21. **Selten, G., H. T. Cuypers, W. Boelens, E. Robanus-Maandag, J. Verbeek, J. Domen, C. van Beveren, and A. Berns.** 1986. The primary structure of the putative oncogene pim-1 shows extensive homology with protein kinases. Cell **46:**603–611.
22. **Sørensen, A. B., M. Duch, P. Jørgensen, and F. S. Pedersen.** 1993. Amplification and sequence analysis of DNA flanking integrated proviruses by a simple two-step polymerase chain reaction method. J. Virol. **67:**7118–7124.
23. **Trusko, S. P., E. K. Hoffman, and D. L. George.** 1989. Transcriptional activation of cKi-ras proto-oncogene resulting from retroviral promoter insertion. Nucleic Acids Res. **17:**9259–9265.
24. **Tsichlis, P. N., and P. A. Lazo.** 1991. Virus-host interactions and the pathogenesis of murine and human oncogenic retroviruses. Curr. Top. Microbiol. Immunol. **171:**95–171.
25. **Vijaya, S., D. L. Steffen, and H. L. Robinson.** 1986. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. J. Virol. **60:**683–692.