Retroviral Integration Sites in Transgenic Mov Mice Frequently Map in the Vicinity of Transcribed DNA Regions

KATRIN MOOSLEHNER, URSULA KARLS, AND KLAUS HARBERS*

Heinrich-Pette-Institut für Experimentelle Virologie und Immunologie an der Universität Hamburg, Martinistrasse 52, 2000 Hamburg 20, Federal Republic of Germany

Transcription of cellular sequences flanking proviral insertion sites was studied in several Mov mouse strains, each of which carried one copy of the Moloney murine leukemia virus in its germ line. In three out of five randomly chosen Mov strains, the provirus had integrated in the vicinity of DNA regions transcribed in the embryonal stem cell line CCE and the embryonal carcinoma cell line F9. Assuming that CCE and F9 cells are developmentally equivalent to the early embryonic cells that were infected to establish the Mov strains, our results suggest that retroviruses integrate preferentially into actively transcribed DNA regions.

An essential step in the life cycle of retroviruses is the integration of a double-stranded DNA copy of their RNA genome into the host cell chromosome (for a review, see reference 18). The proviral DNA is always colinear with the viral RNA, except for the presence of long terminal repeat sequences at either end of the provirus which are generated during reverse transcription. Insertion is precise to the nucleotide level with respect to the viral DNA. In contrast, integration into the host cell DNA occurs at many sites, perhaps randomly. Recently, however, the screening of a large number of integration events revealed the presence of some strongly preferred sites in addition to the random sites (14). Although the role of specific DNA sequences is not quite clear, evidence is accumulating that other levels of chromosomal organization might influence the site of provirus integration. For instance, it has been shown that proviral integrations occur preferentially within a few hundred base pairs of a DNase I-hypersensitive chromatin region (12, 19). DNase-hypersensitive sites in many cases correlate with gene expression and are located in open chromatin domains (4). Therefore these results suggest that retroviruses might integrate preferentially into actively transcribed chromatin domains. In this study, we analyzed transcription of cellular sequences flanking the Moloney murine leukemia virus (M-MuLV) provirus in five different Mov mouse strains. These mouse strains were derived by infection of preimplantation mouse embryos with M-MuLV, and each carries one copy of the provirus in its germ line (9, 16). Here we show that in three out of five randomly chosen Mov mouse strains, the M-MuLV provirus is integrated in the vicinity of DNA regions transcribed in embryonal stem cells.

The Mov strains used in the present study and some of the properties of their respective M-MuLV proviruses are listed in Table 1. Homozygosity at any of the five Mov loci did not interfere with normal development or postnatal life (8, 15). Previous results have shown a close association between DNase-hypersensitive sites and proviruses in all Mov mouse strains studied (12). Cloning of the proviruses and their flanking cellular sequences as well as isolation of Mov-specific cellular probes have been described previously for Mov-7, Mov-9, and Mov-10 (3, 7). The same approach was used in the case of Mov-12 and Mov-21 (data not shown). To obtain the cellular sequences representing the proviral pre-

integration sites, BALB/c genomic libraries were screened with the Mov-specific probes. For each locus one clone carrying the largest insert was purified and further characterized by restriction analysis and by mapping the site of provirus integration. Restriction maps of the Mov-10, Mov-12, and Mov-21 loci are shown in Fig. 1.

As a first step in analyzing transcription of the cellular sequences flanking the proviruses, subfragments from each clone were hybridized to Southern blots with DNAs from several species (mice, humans, rats, chickens) to select for sequences that are unique and have been conserved during evolution. Figure 2 shows representative Southern blots of fragments indicated as hatched bars in Fig. 1. Since these fragments seem to represent single-copy sequences, they were used in subsequent experiments as probes for analyzing transcription on Northern (RNA) blots. The fragments were subcloned into ribovector pSPT18 or pSPT19, and single-stranded RNA probes transcribed from each of the two strands (11) were hybridized to total RNA or poly(A)-enriched RNA of the embryonal stem cell line CCE (5) or the embryonal carcinoma cell line F9 (2). In some cases RNA isolated from F9 cells differentiated into parietal endoderm cells with retinoic acid and cyclic AMP (17) was included in the analysis. Fragments p10a, p12b, and p21a (hatched bars in Fig. 1) from the Mov-10, Mov-12, and Mov-21 loci, respectively, hybridized with defined transcripts when RNAs from CCE and F9 cells were analyzed by Northern blots (Fig. 3). In all three cases transcripts were less abundant or not detectable in RNA from differentiated F9 cells. A second unique fragment from each Mov locus (also indicated by hatched bars in Fig. 1) gave the same result (data not shown). In the case of Mov-9, only one fragment was shown to be unique. All other fragments tested gave a repetitive pattern on Southern blots. When used as a probe on Northern blots, the unique probe hybridized to a large number of RNA species, resulting in a smear (data not shown). The origin of the multiple bands is unclear, and therefore no convincing evidence was found for integration of the Mov-9 provirus in the vicinity of a transcribed gene. Although unique DNA fragments were isolated from the Mov-7 locus, none of the fragments tested hybridized with RNAs from CCE and F9 cells, indicating that about 14 kilobases 5′ and about 2.5 kilobases 3′ from the proviral integration site are not transcribed in these cells (data not shown).

The following observations indicate that the transcripts
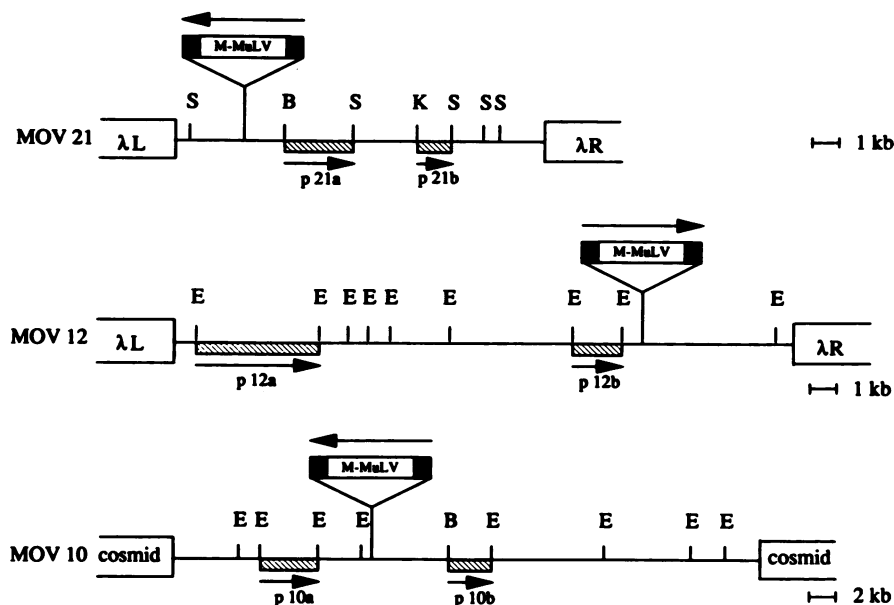
---

* Corresponding author.

FIG. 1. Schematic representation of cloned *Mov* loci. *Mov-21* and *Mov-12* sequences were cloned in lambda vector EMBL 3, and *Mov-10* sequences were cloned in cosmid vector pNNL. Arrows indicate the transcriptional orientation of the M-MuLV proviruses as well as of the fragments hybridizing with RNA (hatched bars). Complete restriction patterns are shown for *Eco*RI (E) in the case of *Mov-10* and *Mov-12* and for *Sst*I (S) in the case of *Mov-21*. Other restriction sites are incomplete and are only shown when necessary for understanding the described experiments. K, *Kpn*I; kb, kilobases.

from the *Mov-10*, *Mov-12*, and *Mov-21* loci were genuine and not artifacts as sometimes seen with riboprobes: (i) transcripts were seen with two independently derived probes; (ii) only single-stranded RNA probes from one strand hybridized; and (iii) the RNA-RNA hybrids on the Northern blots were resistant to RNase A digestion, indicating that they were not a result of unspecific hybridization (data not shown). The results of the transcription analysis are summarized in Fig. 1. The *Mov-10* provirus had obviously integrated within the transcribed region of a gene, since cellular fragments 5' and 3' from the integration site hybridized with RNA. This was confirmed by preliminary data, which indicated that integration had occurred into the first intron of the gene (L. Hamann and K. Harbers, unpublished observation). In the case of *Mov-12* and *Mov-21*, the position of the respective provirus relative to the transcribed gene is less clear. The results presented in Fig. 1 suggest that the proviruses might have integrated outside the transcribed region, either 5' (*Mov-21*) or 3' (*Mov-12*). However, larger DNA regions must be analyzed to confirm these results. With the exception of *Mov-12*, all proviruses integrated in the opposite transcriptional orientation relative to the gene.

The results presented here extend previously described observations of the close association of proviruses and DNase I-hypersensitive sites (12, 19) and show for the first time that proviral genomes map frequently near transcribed DNA regions. Analysis of proviral integration sites in transgenic mice therefore allows the identification of genes that are expressed during early embryogenesis and regulated in a differentiation-specific manner, as shown here for *Mov-10*, *Mov-12*, and *Mov-21*. Embryonal stem cells and embryonal carcinoma cells were chosen for these transcription studies because they serve as a model system for early embryonic

TABLE 1. Mov strains and M-MuLV proviruses

Stage of exposure	Genetic locus	Virus expression	HS site near provirus[a]
4- to 16-cell pre-implantation stage	*Mov-7*	–	+
	Mov-9	+	+
	Mov-21	+	ND
Blastocyst	*Mov-10*	–	+
	Mov-12	–	+

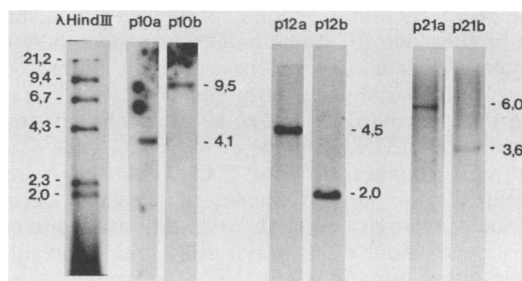[a] HS, DNase I-hypersensitive site (12). ND, Not determined.

FIG. 2. Southern blot analysis of unique probes from the *Mov-10*, *Mov-12*, and *Mov-21* loci used for transcription analysis. Probes p10a, p10b, p12a, and p12b (see Fig. 1) were hybridized to *Eco*RI digests, and probes p21a and p21b were hybridized to *Sst*I digests, of BALB/c mouse DNA. Lengths of λ *Hin*dIII marker and hybridizing fragments are indicated in kilobases. DNAs probed with p10a, p10b, p12a, and p12b and the radioactively labeled λ *Hin*dIII marker were run on same gel. Probes p21a and p21b were analyzed in a separate experiment. DNA (15 μg per lane) was separated on an 0.8% agarose gel and transferred to GeneScreen Plus membranes (Dupont, NEN Research Products). Hybridization of radioactively labeled probes (6) and washing were performed as recommended by the manufacturer; the final wash was with 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate at 50°C.
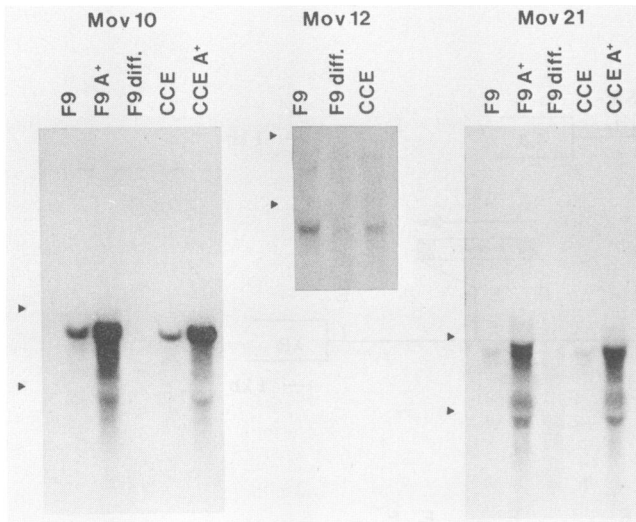
FIG. 3. Transcription analysis of cellular sequences flanking proviral integration sites in Mov mice. Single-stranded RNA probes from fragments p10a, p12b, and p21a (Fig. 1) were hybridized on Northern blots to total or poly(A)-enriched RNA from CCE and F9 cells and to total RNA from F9 cells differentiated with retinoic acid and cyclic AMP (F9 diff.) to parietal endoderm cells (17). Total RNA was isolated by the method of Auffray and Rougeon (1) and enriched for poly(A)$^+$ RNA by chromatography on oligo(dT)-cellulose. Arrows indicate positions of 18S and 28S rRNAs. Each lane contained 15 μg of total RNA or 2 μg of poly(A)-enriched RNA. RNA amounts on the gel were controlled by intensity of rRNA bands.

cells (5, 17). If one assumes that these cells are developmentally equivalent to cells of the preimplantation embryo that were infected with M-MuLV to establish the Mov mice, our results suggest that three out of five proviral integrations have occurred into actively transcribed DNA regions. The frequency at which randomly selected lambda or cosmid clones from a genomic library contain sequences that are transcribed in embryonal stem cells is not known. However, coding sequences are estimated to represent only a small percentage of the total genome of higher organisms (reviewed in reference 10), and it is therefore unlikely that 60% of randomly chosen clones from a genomic library contain DNA transcribed in a given cell type. Therefore, our results suggest that retroviruses integrate preferentially into actively transcribed DNA regions.

Preferential integration into transcribed DNA regions might explain the high frequency with which germ line integration of retroviruses leads to mutations. In the case of the Mov mice, 2 out of 48 proviral integrations resulted in recessive lethal mutations (13, 15). A compilation of data (both published and unpublished) shows that about 5% of all proviral insertions in transgenic mice result in insertional mutations. The molecular basis for preferential integration into actively transcribed DNA regions is not known. Presumably the altered chromatin structure associated with this region might facilitate recombination and serve as a preferential target for the integrase. Studies are under way to characterize the genes described here in more detail and to analyze the possible effects that the integrated proviruses might have on transcription.

LITERATURE CITED

1. **Auffray, C., and F. Rougeon.** 1980. Purification of mouse immunoglobulin heavy-chain messenger RNA's from total myeloma tumor RNA. Eur. J. Biochem. **107**:303–314.
2. **Bernstine, E. G., M. L. Hooper, S. Grandchamp, and B. Ephrussi.** 1973. Alkaline phosphatase activity in mouse teratoma. Proc. Natl. Acad. Sci. USA **70**:3899–3903.
3. **Chumakov, I., H. Stuhlmann, K. Harbers, and R. Jaenisch.** 1982. Cloning of two genetically transmitted Moloney leukemia proviral genomes: correlation between biological activity of the cloned DNA and viral genome activation in the animal. J. Virol. **42**:1088–1098.
4. **Conklin, C. F., and M. Groudine.** 1984. Chromatin structure and gene expression, p. 293–351. *In* A. Razin, H. Cedar, and A. D. Riggs (ed.), DNA methylation. Springer Verlag, New York.
5. **Evans, M. J., and M. H. Kaufman.** 1981. Establishment in culture of pluripotential cells from mouse embryos. Nature (London) **292**:154–156.
6. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**:266–270.
7. **Harbers, K., A. Schnieke, H. Stuhlmann, and R. Jaenisch.** 1982. Infectivity and structure of clones obtained from genetically transmitted Moloney leukemia proviral genomes. Nucleic Acids Res. **10**:2521–2537.
8. **Jaenisch, R., K. Harbers, A. Schnieke, J. Löhler, I. Chumakov, D. Jähner, D. Grotkopp, and E. Hoffmann.** 1983. Germline integration of Moloney murine leukemia virus at the Mov 13 locus leads to recessive lethal mutation and early embryonic death. Cell **32**:209–216.
9. **Jaenisch, R., D. Jähner, P. Nobis, I. Simon, J. Löhler, K. Harbers, and D. Grotkopp.** 1981. Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. Cell **24**:519–529.
10. **Lewin, B.** 1980. Gene expression 2: eukaryotic chromosomes, p. 488–502. John Wiley & Sons, Inc., New York.
11. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
12. **Rohdewohld, H., W. Reik, R. Jaenisch, and M. Breindl.** 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNaseI-hypersensitive sites. J. Virol. **61**:336–343.
13. **Schnieke, A., K. Harbers, and R. Jaenisch.** 1983. Embryonic lethal mutation in mice induced by retrovirus insertion into the α1(I) collagen gene. Nature (London) **304**:315–320.
14. **Shih, C., J. P. Stoye, and J. M. Coffin.** 1988. Highly preferred targets for retrovirus integration. Cell **53**:531–537.
15. **Soriano, P., T. Gridley, and R. Jaenisch.** 1987. Retroviruses and insertional mutagenesis in mice: proviral integration at the Mov 34 locus leads to early embryonic death. Genes Dev. **1**:366–375.
16. **Soriano, P., and R. Jaenisch.** 1986. Retroviruses as probes for mammalian development: allocation of cells to the somatic and germ line lineages. Cell **46**:19–29.
17. **Strickland, S., K. K. Smith, and K. R. Marotti.** 1980. Hormonal induction of differentiation in teratocarcinoma stem cells: generation of parietal endoderm by retinoic acid and dibutyryl cAMP. Cell **21**:347–355.
18. **Varmus, H. E., and R. Swanstrom.** 1984. Replication of retroviruses, p. 369–512. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Vijaya, S., D. L. Steffen, and H. L. Robinson.** 1986. Acceptor sites for retroviral integrations map near DNaseI-hypersensitive sites in chromatin. J. Virol. **60**:683–692.