

Evidence that a major class of mouse endogenous long terminal repeats (LTRs) resulted from recombination between exogenous retroviral LTRs and similar LTR-like elements (LTR-IS)

(transposons/mouse species/retrovirus evolution)

MARTIN SCHMIDT, KLAUS GLÖGGLER, THOMAS WIRTH, AND IVAN HORAK*

Institut für Virologie und Immunbiologie der Universität Würzburg, Versbacher Strasse 7, D-8700 Würzburg, Federal Republic of Germany

Communicated by M. Lindauer, July 25, 1984

ABSTRACT Two endogenous retroviral long terminal repeats (LTRs) were sequenced and compared to LTR-IS (a family of insertion-element-like sequences with structural features of solitary retroviral LTRs) and to Moloney murine leukemia virus DNA. The sequence comparisons revealed that the major difference between these two endogenous LTRs is a 190-base-pair segment which is also present in LTR-IS elements. Hybridization analysis of DNAs from several mouse species using specific probes shows linkage of the 190-base-pair segment to a LTR-IS specific fragment. It is concluded that the major class of endogenous LTRs has been generated by recombination between exogenous retroviral LTRs and LTR-IS sequences.

The mouse genome harbors a substantial amount of retrovirus-like DNA sequences. These can be grouped into A, B, and C particle and VL30 sequences (1). DNA copies of retroviruses are vertically transmitted as stable integral components of chromosomal DNA and, in some cases, can be expressed as infectious viruses. However, only a minor fraction of these sequences represent complete copies equivalent to infectious viruses; a majority are subgenomic sequences of unknown function (2).

The long terminal repeats (LTRs) of proviruses have received particular attention because they represent regulatory units responsible for virus integration, replication, and transcription (3). It has been shown that LTRs might be responsible for the determination of viral tropisms (4, 5) and tumorigenicity (6-8).

Recently, a family of middle-repetitive DNA elements, LTR-IS, has been described (9). These sequences resemble insertion elements (IS) and have the structural features of solitary retroviral LTRs. The LTR-IS elements were isolated by screening a mouse genomic library with a probe from a cloned endogenous retroviral LTR, 36.1. These elements are about 500 base pairs (bp) long, start and end with 11-bp inverted repeats, and contain signals implicated in RNA polymerase II transcriptional regulation. The number of LTR-IS elements in the mouse haploid genome has been estimated to be ≈ 500 . Further studies (10) have shown that LTR-IS elements are present in the genomes of *Mus musculus* subspecies of all geographic locations examined and in *Mus cooki* and *Mus caroli*. Their arrangement is polymorphic in different mouse strains. LTR-IS sequences, therefore, appear to have arisen in early mouse ancestors and have been mobile at some point during their evolution.

Recently, several endogenous proviral sequences have been cloned and characterized (11, 12). It has been noted that the majority of these clones contain LTRs that are homologous to Moloney murine leukemia virus (MuLV) LTR except for a 190-bp central sequence of unknown origin. It

has been speculated (11) that this sequence might be a transposon inserted into a LTR. Here we present evidence that this 190-bp insert is an integral part of LTR-IS elements and that the major class of endogenous LTRs was generated by recombination between LTRs of exogenous retroviruses and LTR-IS sequences.

METHODS

Recombinant DNA Clones. Cloning and subcloning of LTR-IS elements were described previously (9). The clone 36.1 was a gift of R. Mural and J. Ihle (13).

DNA Sequence Analysis. DNA fragments for sequence analysis were labeled at either the 5' ends by polynucleotide 5'-hydroxyl-kinase or at the 3' ends by *Escherichia coli* DNA polymerase. The end-labeled fragments were analyzed according to the method of Maxam and Gilbert (14) with modifications (9). Sequence comparisons were performed using a computer program of the University of Wisconsin.

Hybridization. All hybridizations were carried out in 50% formamide/5 \times standard saline citrate (NaCl/Cit; 1 \times is 0.15 M NaCl/0.015 M sodium citrate, pH 7)/5 \times Denhardt's reagent (1 \times is 0.02% each in Ficoll, bovine serum albumin, and polyvinylpyrrolidone) containing denatured salmon sperm DNA at 100 μ g/ml. Stringent conditions were defined as follows: hybridize at 42°C for 36 hr, wash in 0.2 \times NaCl/Cit at 60°C; nonstringent conditions: hybridize at 37°C for 36 hr, wash in 2 \times NaCl/Cit at 42°C. Transfers from agarose gels were performed as outlined by Southern (15).

Hybridization Probes. The LTR-IS-specific probe was a subcloned *Bgl* II/*Hind* III fragment of the pB8 clone (9). The MuLV-specific probe was a 245-bp *Dde* I/*Dde* I fragment of the clone 36.1 subcloned in pUC8. The common probe for endogenous LTRs and LTR-IS elements was a 217-bp *Bal* I/*Bgl* II fragment of the clone pA6 (9) subcloned in pUC8 (see Fig. 2). The 4.3-LTR probe was a 385-bp *Pst* I/*Sma* I fragment subcloned in pUC8.

RESULTS AND DISCUSSION

Nucleotide Sequence Comparisons of LTR-IS Elements with Retroviral LTRs. The LTR-IS elements were isolated from a mouse genomic library by screening with a representative retroviral LTR probe from the clone 36.1 (9). It was, therefore, of interest to delineate the sequence homology between LTR-IS elements and the clone 36.1 that was responsible for the hybridization and to search for possible homologies to other endogenous LTR sequences. For this purpose, the 36.1 clone was sequenced.

The 36.1 clone itself was isolated from the C3H mouse

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: LTR, long terminal repeat; LTR-IS, insertion-element-like sequences with structural features of solitary retroviral LTRs; MuLV, murine leukemia virus; L, long; S, short; bp, base pair(s).

*To whom reprint requests should be addressed.

genomic library by screening with a Moloney MuLV probe (13). Sequencing data (Fig. 1, line b) and hybridizations (data not shown) confirmed that the 36.1 clone contains a retrovirus-like LTR sequence flanked with *env*-like sequences at its 5' end and with mouse sequences at its 3' end. The entire LTR is about 700 bp long and contains large areas of homology to Moloney MuLV LTR (Fig. 1, line a). The obvious exception is the 190-bp central segment, which is >90% homologous to the LTR-IS sequences. Comparison of the 36.1 LTR sequence to *Pst* I/*Sma* I fragments of several endogenous retroviral LTRs recently sequenced by Khan and Martin (11) indicates that these sequences belong to the same class of endogenous LTRs. Khan and Martin noted that these fragments also contain LTRs that are homologous to Moloney MuLV LTR except for a 190-bp central segment of unknown origin. They suggested that this sequence might be a transposon inserted into a LTR. This 190-bp segment is also present in LTR-IS elements (Fig. 1).

A consensus sequence compiled from nine randomly chosen LTR-IS elements was used for the comparisons shown in Fig. 1. Based on their length the members of the LTR-IS family can be divided into two subclasses, long (L) and short (S). The length differences are due to two deletions, one (12 bp) at positions 158–170 and another (93 bp) at positions 181–274. The homology (without the deletions) between L and S sequences is 95%.

The 36.1 LTR and the endogenous isolates of Khan and

Martin (11) share an almost identical 190-bp segment with the S subclass of LTR-IS elements. The corresponding segment in the L subclass of LTR-IS elements is not 190 but 305 bp long and the two deletions that distinguish the L and S subclasses are located within this 305-bp segment. Interestingly, the length and the boundaries of the two deletions are identical in LTR-IS S subclass, 36.1, and the isolates of Khan and Martin (11), indicating that they probably had a common ancestor.

A second endogenous retrovirus-related sequence, clone 4.3, was isolated from DNA of inbred mouse strain 129/J by screening with an *env* probe derived from the clone 36.1. The nucleotide sequence of this clone (data not shown) is >90% homologous to the NFS-xenotropic LTR (11) and has no homology to the LTR-IS elements.

Southern Blot Analysis Using Specific LTR Segments as Probes of Various Mouse DNAs. Sequence comparisons revealed the existence of type-specific and common DNA regions within LTRs of endogenous (clone 36.1), exogenous (Moloney MuLV), and LTR-IS sequences. Accordingly, we have subcloned corresponding DNA fragments to obtain type-specific and common probes (see *Methods* and Fig. 2).

Using these probes, we have screened the DNA from a number of species and strains for corresponding sequences. Previously, we had shown that LTR-IS-specific sequences are present in all inbred strains and subspecies of *M. musculus* tested, as well as in *M. cooki* and *M. caroli* (10). Fig. 3A



FIG. 1. Comparison of LTR-IS and retroviral LTR nucleotide sequences. Line a, Moloney MuLV LTR (16); line b, clone 36.1 LTR; line c, LTR-IS-S, consensus sequence derived from sequences of five clones of the S subclass; line d, LTR-IS-L, consensus sequence derived from sequences of four clones of the L subclass. Dots indicate identical nucleotides and blank spaces, absent nucleotides. The U3, R, and U5 regions of Moloney MuLV LTR are marked with arrows. The beginning and the end of the 190-bp segment are indicated with triangles.

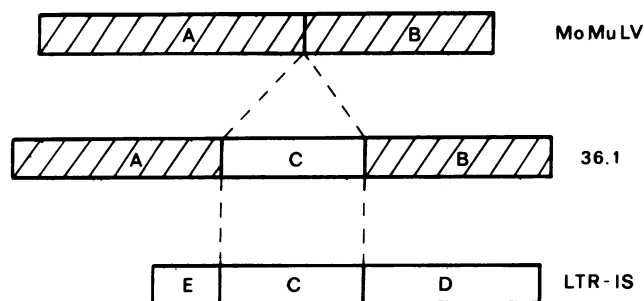


FIG. 2. Schematic illustration of sequence homologies between LTR-IS, Moloney (Mo) MuLV, and 36.1 LTR. Homologous sequences are identified with identical letters. Hybridization probes were derived from segment B of the clone 36.1 (specific probe for MuLV LTRs), C of the LTR-IS clone (common probe for both 36.1 and LTR-IS sequences), and D of LTR-IS (LTR-IS-specific probe).

shows that the common probe for LTR-IS and endogenous LTRs hybridizes to all DNAs previously found positive for LTR-IS-specific sequences. Note that DNA from all the *M. musculus* subspecies and inbred strains hybridizes to the MuLV-specific probe but that DNAs from the other *Mus* species, *caroli*, *cooki*, *platythrix*, and *pahari*, do not hybridize even under nonstringent conditions (Fig. 3 B and C). Results of the Southern blot analysis are summarized in Table 1. These data show that the common fragment and the LTR-IS-specific fragment are linked and segregate together in all mouse DNAs tested.

We have characterized two different types of endogenous LTR sequences, 36.1 and 4.3. The major difference between

Table 1. MuLV-specific, LTR-IS-specific, and MuLV/LTR-IS-common DNA sequences in various mouse species and subspecies

Species	Probe		
	MuLV-specific	LTR-IS-specific	Common
<i>M. musculus</i> inbred	+	+	+
<i>M. musculus musculus</i>	+	+	+
<i>M. musculus castaneus</i>	+	+	+
<i>M. musculus molossinus</i>	+	+	+
<i>M. musculus spicilegus</i>	±	+	+
<i>M. caroli</i>	-	+	+
<i>M. cooki</i>	-	+	+
<i>M. platythrix</i>	-	±	±
<i>M. pahari</i>	-	-	-

The presence of specific DNA fragments was estimated from hybridization experiments described in the legend to Fig. 3 and in ref. 10. +, Positive hybridization under stringent conditions. ±, Positive hybridization under nonstringent conditions only.

these two LTR types is the presence in 36.1 of a 190-bp segment homologous to LTR-IS elements that is not present in the clone 4.3. The remaining parts of both LTRs are ≈90% homologous to Moloney MuLV LTR. LTR types 36.1 and 4.3 both possess single *Pst* I and *Kpn* I sites. Due to the 190-bp segment the *Pst* I/*Kpn* I fragments from the two LTR types are of different length. This was used to advantage in determining the relative amounts of the two LTR-types in the mouse genome. Digesting mouse DNA with *Pst* I and *Kpn* I and hybridizing it to a 4.3-type LTR DNA probe revealed (Fig. 4) an abundant 552-bp fragment corresponding

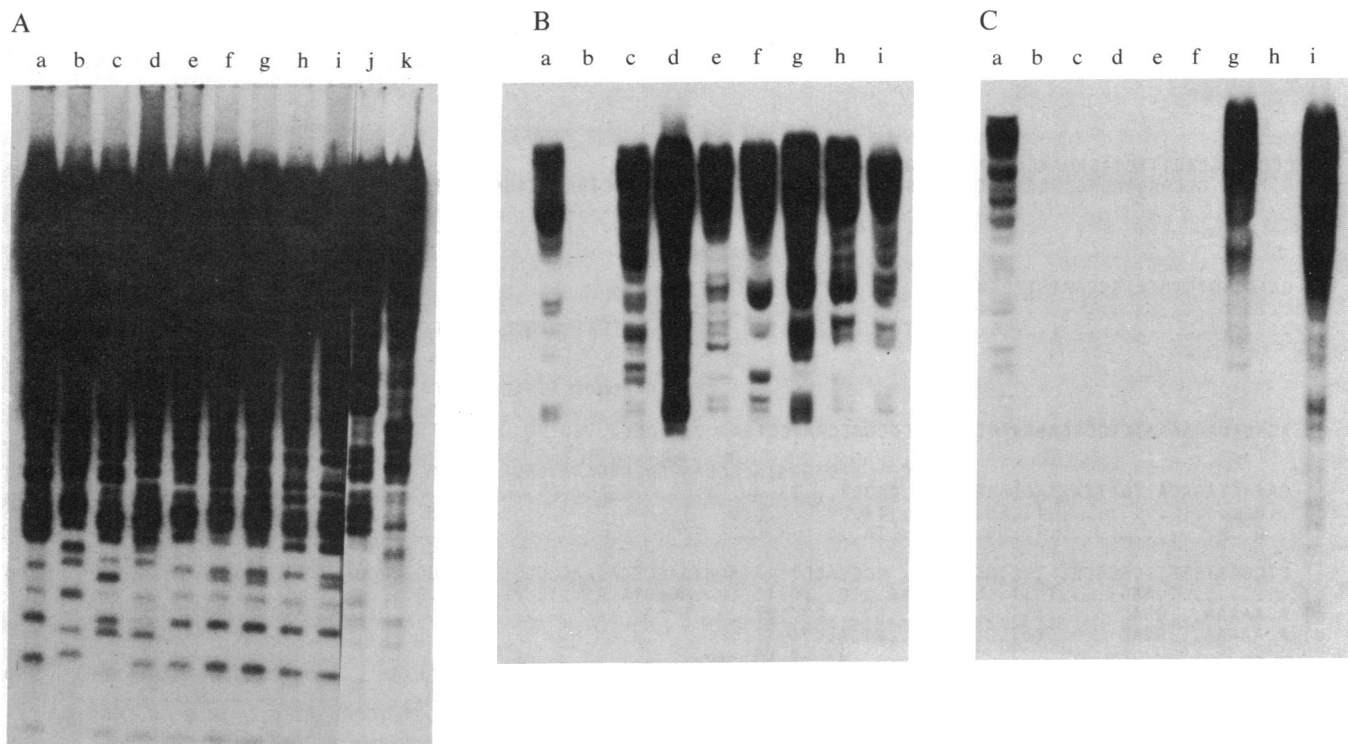


FIG. 3. Southern blot hybridizations of various animal DNAs using specific LTR segments as probes. Liver DNAs (10 μ g) digested with *Eco*RI were analyzed by Southern blotting and hybridized to 32 P-labeled probes. (A) The common probe for LTR-IS and for endogenous LTRs was hybridized with DNA restriction fragments from *M. musculus musculus* (lane a), *M. m. spicilegus* (lane b), *M. m. castaneus* (lane c), *M. m. molossinus* (lane d), C3H (lane e), BALB/c (lane f), C57BL/6 (lane g), DBA (lane h), 129/J (lane i), *M. cooki* (lane j), and *M. caroli* (lane k). (B) The MuLV-specific probe was hybridized with restriction digests. Lane designations are the same as in A. (C) The MuLV-specific probe was hybridized with DNA restriction fragments from *M. m. castaneus* (lane a), *M. platythrix* (lane b), *M. pahari* (lane c), *Leggata minutoides* (lane d), *M. cooki* (lane e), *M. caroli* (lane f), *Gerbillus gerbillus* (lane g), *Rattus rattus* (lane h), and *Mesocricetus auratus* (lane i). Conditions were stringent for hybridizations shown in A and B and nonstringent for those shown in C. For details see *Methods*.

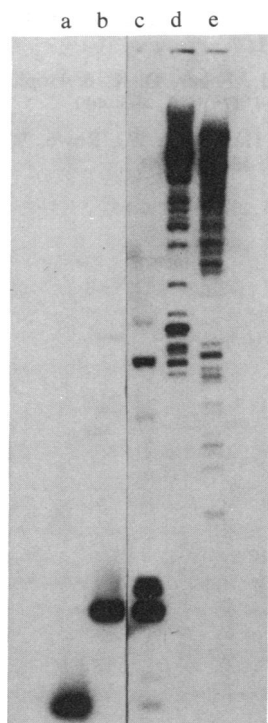


FIG. 4. Endogenous LTR types in inbred mouse strain 129/J as determined by hybridization of internal *Pst* I/*Kpn* I fragments to an LTR probe. Liver DNA (10 μ g) or plasmid DNA (1 μ g) was digested with restriction enzymes and analyzed by electrophoresis in 1.2% agarose gel. The gel was blotted and hybridized under stringent conditions to the LTR probe from the clone 4.3. Lane a, 4.3-LTR clone digested with *Kpn* I/*Pst* I; lane b, 36.1-LTR clone digested with *Kpn* I/*Pst* I; lane c, 129/J DNA digested with *Kpn* I/*Pst* I; lane d, 129/J DNA digested with *Kpn* I; lane e, 129/J DNA digested with *Pst* I.

to the *Pst* I/*Kpn* I fragment of 36.1-type LTR but only a weakly hybridizing 391-bp fragment corresponding to 4.3-type LTR. It seems, therefore, that the majority of the endogenous MuLV-derived LTR sequences belong to the class of sequences containing the LTR-IS-homologous 190-bp segment.

The length differences between L and S subclass LTR-IS elements are interpreted as being due to two deletions. Alternatively, insertions could be responsible, in which case the L elements would have been derived from the S elements. Frequent deletions or insertions in areas of short repeated sequences have been noted in prokaryotic as well as eukaryotic DNA (17, 18). It seems likely that the two deletions/insertions happened before the amplification of the LTR-IS elements in the mouse genome. There are two distinct patterns of microheterogeneity because of point mutations, one for L and one for S elements (data not shown), which suggest independent evolution and amplification of both classes after the deletion/insertion event.

The idea that the 190-bp segment could be an insertion element, as suggested by Khan and Martin (11), is based mainly on the 6-bp direct repeats (T-C-A-G-T-T) at its ends. Thus, the exact ends of the 190-bp element are known. If it had been inserted into another sequence, one should find the same ends of the 190-bp element flanked by different 6-bp repeats. However, the ends of the 190/305-bp segment in LTR-IS elements are not flanked by any direct repeats and the sequence T-C-A-G-T-T is located within the 190-bp segment. In addition, the clone 4.3, which does not contain the 190-bp LTR-IS-homologous segment, has T-C-A-G-T-T tandemly repeated (Fig. 5). For this reason and because of the data discussed above, it seems more likely that the 190-bp



FIG. 5. Comparison of the nucleotide sequences of three endogenous LTRs and two LTR-IS isolates in the region of putative recombination. Line a, clone 4.3; line b, clone B34 (11); line c, clone 36.1; line d, LTR-IS A6; line e, LTR-IS B8. Putative crossover points are indicated by triangles and vertical lines, and the repeats described for the clone B34 (11) are boxed.

segment of the endogenous LTRs was derived from recombination between exogenous retroviral LTRs and LTR-IS-like sequences.

The notion that the large family of LTR-IS elements can be involved in recombination with LTRs of exogenous viruses is of interest. It is known that the development of spontaneous thymic lymphoma in AKR mice is accompanied by production of new recombinant viruses (19). Several of these recombinant viruses have altered host-range properties and are highly leukemogenic (20). Recent evidence indicates that not only recombination within viral structural genes (21) but also subtle alteration in LTR sequence (22) could be involved in this process.

We are thankful to Dr. J. Klein (Max-Planck-Institute, Tübingen, F.R.G.) for generously providing *M. musculus* subspecies and *Mus* species and to Ms. I. Sauer-Clark for excellent technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft, SFB 165.

- Coffin, J. (1982) in *RNA Tumor Viruses*, eds. Weiss, R., Teich, N., Varmus, H. & Coffin, J. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 1109-1203.
- Rowe, W. P. (1978) *Harvey Lect.* 71, 173-192.
- Varmus, H. E. (1982) *Science* 216, 812-820.
- Laimins, L. A., Khoury, G., Gorman, C., Howard, B. & Gruss, P. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6453-6457.
- DesGroseillers, L., Rassart, E. & Jolicoeur, P. (1983) *Proc. Natl. Acad. Sci. USA* 80, 4203-4207.
- Neal, B. G., Hayward, W. S., Robinson, H. L., Fang, J. & Astrin, S. M. (1981) *Cell* 23, 323-334.
- Payne, G. S., Courtneidge, S. A., Crittenden, L. B., Fadley, A. M., Bishop, J. M. & Varmus, H. E. (1981) *Cell* 23, 311-322.
- Hayward, W. S., Neel, B. G. & Astrin, S. M. (1981) *Nature (London)* 290, 475-480.
- Wirth, T., Glöggler, K., Baumruker, T., Schmidt, M. & Horak, I. (1983) *Proc. Natl. Acad. Sci. USA* 80, 3327-3330.
- Wirth, T., Schmidt, M., Baumruker, T. & Horak, I. (1984) *Nucleic Acids Res.* 12, 3603-3610.
- Khan, A. S. & Martin, M. A. (1983) *Proc. Natl. Acad. Sci. USA* 80, 2699-2703.
- Ou, C. Y., Boone, L. R. & Yang, W. K. (1983) *Nucleic Acids Res.* 11, 5603-5619.
- Roblin, R., Young, J. M., Mural, R. J., Bell, T. E. & Ihle, J. N. (1982) *J. Virol.* 43, 113-126.
- Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
- Southern, E. M. (1975) *J. Mol. Biol.* 98, 503-517.
- Beveren, C. V., Rands, E., Chattopadhyay, S. K., Lowy, D. R. & Verma, I. M. (1982) *J. Virol.* 41, 542-556.

17. Albertini, A. M., Hofer, M., Calos, M. P. & Miller, J. H. (1982) *Cell* **29**, 319-328.
18. Tlsty, T. D., Albertini, A. M. & Miller, J. F. (1984) *Cell* **37**, 217-224.
19. Hartley, J. W., Wolford, N. K., Old, L. J. & Rowe, W. P. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 789-792.
20. Chattopadhyay, S. K., Cloyd, M. W., Linemeyer, D. L., Lander, M. R., Rands, E. & Lowy, D. R. (1982) *Nature (London)* **295**, 25-31.
21. Rommelaere, J., Faller, D. K. & Hopkins, N. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 495-499.
22. Lung, M. L., Hartley, J. W., Rowe, W. P. & Hopkins, N. (1983) *J. Virol.* **45**, 275-290.