Volume 13 Number 10 1985 Nucleic Acids Research

Structure and genomic organization of a new family of murine retrovirus-related DNA sequences (MuRRS)

Martin Schmidt, Thomas Wirth, Burkhard Kröger and Ivan Horak*

Institut für Virologie und Immunbiologie der Universität Würzburg, Versbacher Strasse 7, D-8700 Würzburg, FRG

Received 11 March 1985; Revised and Accepted 24 April 1985

ABSTRACT
A new class of murine retrovirus-related sequences (MuRRS) is described. These 5.7 kb long transposon-like DNA-elements start and end with ~ 600 bp long repeats identical to previously identified solitary LTR-like elements (LTR-IS). There are about 50 - 100 5.7 kb elements and about 500 - 1000 solo LTR-IS elements per mouse haploid genome. Sequence analysis of one cloned MuRRS element revealed several possible open reading frames with partial sequence homologies to retroviral gag, pol and env genes.

INTRODUCTION

Mobile elements have been described in the genomic DNA of prokaryotes, plants and invertebrates (for review see 1). So far no corresponding sequences capable of direct DNA-DNA transposition have been found in vertebrate DNA but several genetic elements of vertebrates are considered as candidates for mobile elements (2 - 7). The evidence for their mobility is, however, only circumstantial. It has been speculated, that some of these sequences could be mobile via an RNA intermediate (8, 9). Most striking similarity exists between prokaryotic transposons, copia-like elements of Drosophila, Ty-elements of yeast and retroviral proviruses of vertebrates. This similarity includes in some cases not only the overall structural homology, but also partial nucleotide or amino acids sequence homologies and functional analogy (10, 11, 12).

The chromosomal DNA of vertebrates contains multiple gene families that are related to the genomes of retroviruses. Several distinct retroviral gene families have been detected in mouse genomic DNA. Two of them represent sequences related to competent retroviruses (of type B and C), the others, IAPs,

VL 30, LTR-IS and ETn (4 - 7) do not code for infectious retroviruses. Previously described LTR-IS families of middle repetitive DNA elements closely resemble insertion elements and have the structural features of solitary retroviral LTRs. They are 500 - 600 bp long, start and end with 11 bp inverted repeats and are flanked by 4 bp duplicated sequences of host DNA at either end (6). We have shown that the LTR-IS elements contain weak RNA polymerase II promoters which require enhancement by cis- or transactivating factors (13). Evidence was obtained indicating that the LTR-IS elements have arisen in early mouse ancestors and were mobile at least at some point during their evolution (14, 15).

Due to the analogy of LTR-IS elements to proviral LTRs one could expect that longer structures having LTR-IS elements on their both ends, also exist in the mouse genome. Here we describe the isolation, nucleotide sequence and characterization of such elements.

MATERIAL AND METHODS

Material

The 129/J mouse DNA library in λ Charon 4 a was described previously (6). A 129/Sv mouse library in cosmids pcos 2 EMBL (16) was a generous gift of A. M. Frischauf and H. Lehrach (EMBL, Heidelberg).

Cloning and DNA sequence analysis

Basic recombinant DNA manipulations were carried out as outlined by Maniatis et al. (17). Overlapping sets of DNA fragments subcloned in M13 mp 18/19 vectors were analyzed by the dideoxy sequencing method (18) with modifications (19). The complete nucleotide sequence of both complementary strands was determined. Sequence comparisons were performed using a computer program of the University of Wisconsin on a VAX 11/750 computer.

Hybridization

All hybridizations were carried out in 50 % formamide/5 x standard saline citrate/5 x Denhardt's reagent at 42° for 20 - 36 hrs (17).

RESULTS

Isolation of the MuRRS elements

Fifteen LTR-IS clones characterized by restriction mapping did not reveal any common flanking sequences (6), which would indicate that they might be a part of a larger element. Moreover, all nine sequenced elements (6, unpublished) are flanked by 4 bp target site duplications, diagnostic for solo LTR elements. This suggested that if any longer elements do exist in the mouse genome they constitute only a minor fraction of the 500 - 1000 LTR-IS hybridizing sequences.

In order to search for these type of elements, the mouse genomic DNA was digested with BglII restriction nuclease, which cuts only once within a LTR-IS element and the DNA was blotted and hybridized to the LTR-IS probe. This analysis revealed a multicopy band of size 1.7 kb, indicating a common flanking sequence. The 1.7 kb DNA-band was cloned and several of the cloned fragments were characterized by restriction mapping. The majority of these clones also have the characteristic BalI site within the LTR-IS hybridizing segment, suggesting that the 1.7 kb fragment contains a 5' part of the LTR-IS element and 5' flanking sequences, which are identical for several LTR-IS elements. This means that either several LTR-IS elements are inserted into an identical middle repetitive DNA sequence or that the 1.7 kb fragments represent the 3' end of a provirus-like structures with LTR-IS elements on both ends.

In order to isolate such provirus-like structures, a mouse cosmid library was screened with the 5' segment from one of the 1.7 kb fragments described above. Four randomly chosen positively hybridizing cosmid clones were further characterized by restriction mapping and hybridization. This analysis (Fig. 1) revealed a set of 5.7 kb long structures with LTR-IS elements on its both ends. The majority of restriction sites within the 5.7 kb element are identical among the four clones, but their flanking sequences are different. Due to their resemblance to retroviral proviruses we named these 5.7 kb elements MuRRS (Murine Retrovirus-Related Sequences).

Nucleotide sequence of the MuRRS-5 clone

Overlapping sets of DNA fragments of the MuRRS-5 clone were

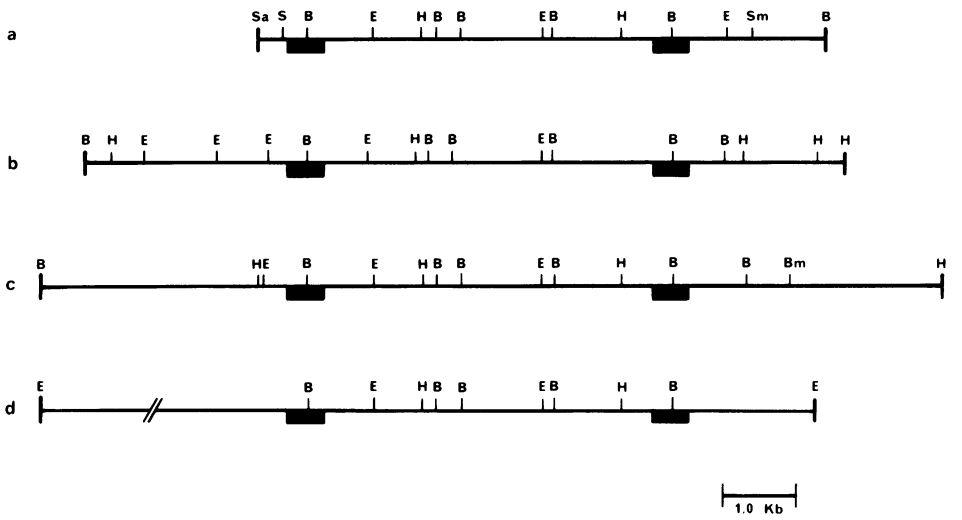

Fig. 1
Restriction maps of four MuRRS clones
Bars indicate the position of LTR-IS elements. B, Bgl II; Bm, Bam HI; E, Eco RI; H, Hind III; S, Sst I; Sa, Sal I; Sm, Sma I. a) MuRRS-1, b) MuRRS-2, c) MuRRS-5, d) MuRRS-6.

subcloned in M13 phage vectors and the nucleotide sequence was determined. The sequence in Fig. 2 starts and ends with the host 4 bp target site duplication ATAG. The LTR-IS elements of the MuRRS-5 clone belong to the S-subclass of LTR-IS (14) with which they are more than 90 % homologous. Both LTR-IS elements contain 11 bp inverted repeats at either end and the RNA polymerase II regulatory signals, CCAAT, TATAAA and ATAA. The 5' LTR-IS element (nt 5 - 513) and 3' LTR-IS (nt 5204 - 5685) are not absolutely identical in that, besides five point mutations there is a 27 bp duplication of the sequence (nt 417 - 444) in the 5' LTR-IS.

The internal region between the two LTR-IS structures starts with a sequence (nt 516 - 534) homologous to the Moloney murine leukemia virus (MoMLV) (20) primer-binding site, and continues as long leader sequence (535 - 1062). By maximizing the nucleic acid sequence homology between MuRRS and MoMLV (Fig. 3) it is possible to recognize long stretches of amino acid homology between partial open reading frames in the MuRRS element and the MoMLV gag, pol and env proteins. This homology does not

Fig. 2
Complete nucleotide sequence of the clone MuRRS-5
The arrows mark 4 bp target site duplication, 11 bp inverted repeats and beginning of gag, pol and env genes.

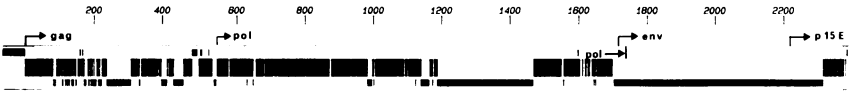

Fig. 3
Comparison of aminoacid sequences derived from nucleotide sequences of MoMLV and MuRRS-5
Computer-assisted comparison has been made using "gap show plot" program after aligning the two amino acid sequences with "gap out" program. "Gap show plot" displays two sequences by plotting one above the other and marking the points of similarity. Vertical lines indicate amino acid identity, deletions in MuRRS sequence appear as bars in the bottom line, insertions in the upper line. The numbers indicate the amino acids of MoMLV derived from the nucleotide sequence (20).

span the whole coding region of MoMLV. In addition to several small differences which necessitate frame shifts to maintain amino acid homology, there is an 840 nt deletion in the MuRRS element at a position corresponding to the pol region (Δ pol) and a 1840 deletion at a position corresponding to the env gene (Δ env). Another feature common to all retroviral proviruses that is conserved in the MuRRS clone is a polypurine track to the 5' side of the right hand LTR-IS (5190 - 5201).

Copy number of MuRRS in the mouse genome

The number of MuRRS elements in the mouse haploid genome was estimated by two methods. Firstly, from the frequencies of MuRRS$^+$ plaques in the mouse λ library and secondly by comparison of the signal intensities obtained from a dot-blot hybridization of a homologous MuRRS probe (2.2/kb Eco RI fragment, Fig. 1) to the mouse DNA (129/J), with hybridizations containing a known number of cloned MuRRS copies. The number of MuRRS in the 129/J mouse haploid genome was calculated, from both procedures, to be about 50 - 100.

Southern blot analysis of DNA from various species using MuRRS specific hybridization probes

Sequence analysis revealed two major deletions, Δ pol and Δ env, in the MuRRS-5 clone. All four randomly chosen MuRRS clones are of the same length and their restriction maps are also very similar, indicating that all four clones contain similar Δ pol and Δ env deletions (Fig. 1). It was also of interest to know whether in the 50 - 100 MuRRS copies in the

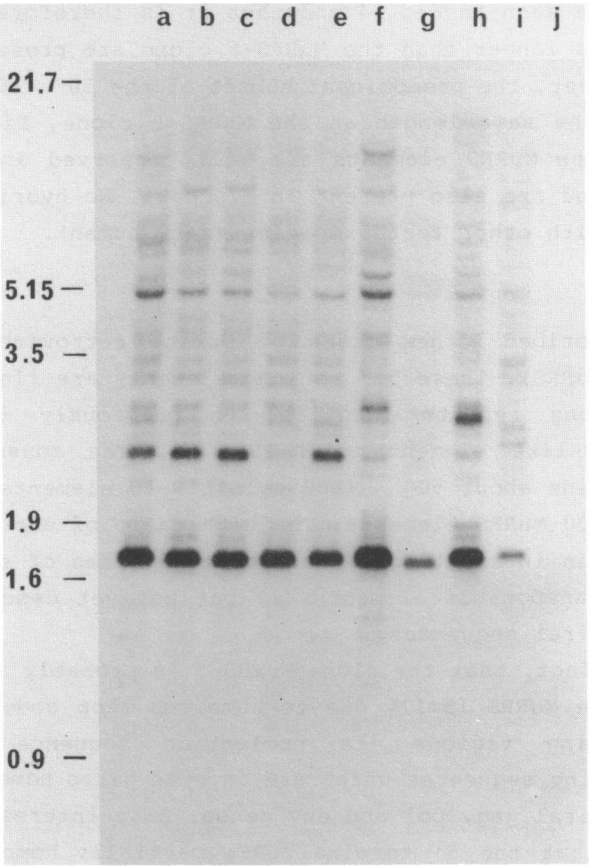

Fig. 4
Southern blot analysis of various DNAs using MuRRS-specific hybridization probe
Liver DNAs (10 ug) digested with Bgl II were analyzed by Southern blotting and hybridized to the subcloned ^{32}P-labeled MuRRS/Bgl II/Pst I 3' fragment. a) 129/J, b) C57Bl/6, c) BALB/c, d) C3H, e)DBA, f) M. m. molossinus, g) M. m. castaneus, h) M. m. spicelegus, i) M. cooki, j) R. rattus.

genome undeleted MuRRS sequences exist and how representative is the clone we have isolated and sequenced. Therefore we have analyzed the size of the MuRRS internal restriction fragments in genomic DNA. Fig. 4 shows that the majority of the BglII fragments which hybridize with the MuRRS 3' probe are of the same size (1.7 kb), as the BglII fragments of the four MuRRS clones we have analyzed (Fig. 1). Nevertheless, larger BglII

fragments are seen in Fig. 4 and thus it is therefore possible that elements longer than the MuRRS-5 clone are present in the genome. However, the predominant subset of the 50 - 100 genomic copies has the same length as the MuRRS-5 clone. Fig. 4 also shows that the MuRRS elements are well conserved among M. m. subspecies and are also present in M. cooki. No hybridizations were found with other tested species (rat, human).

DISCUSSION

We have described a new class of murine retrovirus-related sequences (MuRRS). These 5.7 kb long elements are flanked with ~ 600 bp long repeats identical to previously identified solitary LTR-like elements (LTR-IS) (6). The mouse haploid genome contains about 500 - 1000 solo LTR-IS elements and only about 50 - 100 MuRRS elements. The high ratio of short to long elements is an interesting feature of this class of sequences, common to transposable elements (1) but not yet described for other retroviral sequences.

Despite the fact, that the clone MuRRS-5 is probably a "silent" member of the MuRRS family due to numerous stop codons within protein coding regions its nucleotide sequence revealed possible coding sequences which are in some parts homologous to other retroviral gag, pol and env genes. Also interestingly we have noted that the 5' terminal ORF, partially homologous to MoMLV p15 might code for a protein which has a N-terminal extension of 62 amino acids preceding the beginning of p15. The initiating methionine codon of MoMLV p15 is replaced by threonine in the MuRRs element. In addition to small insertion/deletion differences there are two major deletions, Δ pol and Δ env, in the MuRRS sequence compared to MoMLV. It is noteworthy, that the Δ pol (840 bp long), is almost identical in size and position with the pol deletion of spleen focus forming virus (SFFV) (21). The Δ env encompasses almost the entire env gene, but the remaining 3' terminal 71 amino-acids are homologous to MoMLV p15E.

A Southern blot analysis indicate that the majority of MuRRS elements have the same deletion (Δ pol, Δ env), not only within one mouse inbred strain but also within different

M. species, This suggests that these elements were amplified and spred after the deletion event.

At the present time the evolutionary origin of MuRRS/LTR-IS sequences is not clear. Southern blot analysis (Fig. 4 and ref. 14, 15) revealed a high degree of conservation of these sequences among mouse species and suggested their involvement in the generation of C-type endogenous retroviruses. Another indication of a relationship of MuRRS to murine C-type viruses is the scattered nucleotide sequence homology to MoMLV, AKR and SFFV viruses. No apparent homology has been found to the IAP sequence (22).

It is likely that IAPs, VL 30, ETn and the MuRRS sequences described here originated from partially deleted proviruses. The functional significance of these sequences is still a matter for speculation. There is evidence that at least some members of these individual families are transcribed in vivo (4, 5, 7, 13), but their translational products, if any, have not yet been identified. Another possible function is in the generation of DNA rearrangements via homologous or non-homologous recombination. The most likely route whereby the solo LTR-IS elements originated was via the homologous recombination between the left and the right LTR-IS sequences of a MuRRS element, and precise excision of the intervening MuRRS genomic sequence. The existance of the high copy number of solo LTR-IS suggests that the homologous recombination between LTR-IS elements might be frequent. Another line of evidence for the high recombinational activity of LTR-IS elements is the previously described generation of endogenous retroviral LTRs by recombination between exogenous retroviral LTRs and LTR-IS elements (14) and the isolation of a new class of elements (23) which contain part of the LTR-IS sequence.

ACKNOWLEDGEMENTS

We thank Isabella Sauer-Clark and Margot Keidel for technical assistance and C. Stoppe for preparation of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 165) and from Fonds der Chemischen Industrie.

*To whom correspondence should be addressed

REFERENCES

1. Shapiro, J. A. (1983). Mobile genetic elements (Academic Press, New York, N. Y.) 1-688
2. Schmid, C. W. and Jelinek, W. R. (1982) Science 216, 1065-1070
3. Kramerov, D. A., Grigoryan, A. A., Ryskov, A. P. and Georgiev, G. P. (1979) Nucleic Acids Res. 6, 697-713
4. Lueders, K. K. and Kuff, E. L. (1977) Cell 12, 963-972
5. Itin, A. and Keshet, E. (1983) J. Virol. 47, 656-659
6. Wirth, T., Glöggler, K., Baumruker, T., Schmidt, M. and Horak, I. (1983) Proc. Natl. Acad. Sci. USA 80, 3327-3330
7. Brulet, P., Kaghad, M., Xu, Y.-S., Croissant, O. and Jacob, F. (1983) Proc. Natl. Acad. Sci. USA 80, 5641-5645
8. Van Arsdell, S. W., Denison, R. A., Bernstein, L. B., Weiner, A. M., Manser, T. and Gesteland, R. F. (1981) Cell 26, 11-17
9. Jagadeeswaran, P., Forget, B. G. and Weissman, S. M. (1981) Cell 26, 141-142
10. Varmus, H. E. (1983) in Mobile genetic elements, ed. Shapiro, J. A. (Academic Press, New York, N. Y.) pp. 411-503
11. Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K. and Yuki, S. (1984) Nature 312, 659-661
12. Roeder, G. S. and Fink, G. R. (1983) in Mobile Genetic elements, ed. Shapiro, J. A. (Academic Press, New York, N. Y.) pp. 229-328
13. Köhrer, K., Grummt, I. and Horak, I. (1985) Nucleic Acids Res., in press
14. Schmidt, M., Glöggler, K., Wirth, T. and Horak, I. (1984) Proc. Natl. Acad. Sci. USA 81, 6696-6700
15. Wirth, T., Schmidt, M., Baumruker, T. and Horak, I. (1984) Nucleic Acids Res. 12, 3603-3610
16. Poustka, A., Rackwitz, H. R., Frischauf, A. M., Hohn, B. and Lehrach, H. (1984) Proc. Natl. Acad. Sci. USA, 81, 4129-4133
17. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular cloning, Cold Spring Harbor Laboratory
18. Sanger, F. (1980) J. Mol. Biol. 143, 161-178
19. Messing, J. (1983) Method in Enzymology 101, 20-78
20. Shinnick, T. M., Lerner, R. A. and Sutcliffe, J. G. (1981) Nature 293, 543-548
21. Clark, S. P. and Mak, T. W. (1983) Proc. Natl. Acad. Sci. USA 80, 5037-5041
22. Burt, D. W., Reith, A. D. and Brammar, W. J. (1984) Nucleic Acids Res. 12, 8579-8593
23. Propst, F. and Vande Woude, G. F. (1984) Nucleic Acids Res. 12, 8381-8392