Volume 16 Number 15 1988 Nucleic Acids Research

Insertion of a retrotransposon within the 3′ end of a mouse gene provides a new functional polyadenylation signal

Thomas Baumruker, Claudia Gehe and Ivan Horak*

Institut für Virologie und Immunbiologie der Universität Würzburg, Versbacher Str. 7, 8700 Würzburg, FRG

Received June 9, 1988; Accepted June 29, 1988

ABSTRACT

A site of genomic insertion of the mouse retrotransposon LTR-IS/MuRRS was analysed. The comparison of the genomic and the cDNA clones indicates the insertion of the LTR-IS element into the 3' untranslated region of a mouse gene. The fact that the isolated cDNA clone ends with a poly A tail 20 nucleotides downstream from the LTR-IS AATAAA box and the result of the S1-nuclease mapping provides evidence that the 3' end of the mouse gene transcript was generated under the control of the LTR-IS polyadenylation signal.

INTRODUCTION

Transposable elements are probably the major source of mutagenic activity in the genome of lower eukaryotes (for review see ref. 1). The site of the element insertion in the genome is a critical factor in determining its genetic consequences. The relative accumulation of mobile elements withing the repetitive DNA and the DNA of unknown function indicates a selection of non-lethal insertions into "allowed" sites in the genome. The analysis of the insertions into active genes shows examples of more tolerated regions, like introns, in contrast to exons and regulatory regions, where the insertions have more severe effects on the phenotypic expression (1, 2). However, also intron insertions associated with mutant phenotypes were described (3).

The evaluation of the impact of mobile genetic element insertion in higher eukaryotes is hampered by the lack of genetic analysis. Therefore, the molecular analysis of individual integration events is a source of valuable information towards our understanding of mechanisms involved in insertional mutagenesis.

© IRL Press Limited, Oxford, England.

Previously we have described mouse retrotransposons MuRRS (Murine retrovirus-related sequences), which are 5.7 kb long elements present in about 100 copies per mouse haploid genome (4). In addition to these proviruslike structures, there are about 1000 copies per haploid genome of solo long terminal repeats (LTR) of MuRRS elements, called LTR-IS (5).

Three possible activities of LTR-IS/MuRRS elements in the genome may be considered. Firstly, it is the generation of DNA re-arrangements via homologous or non-homologous recombination. Some evidence for high recombinational activity of LTR-IS elements was reported (6, 7). Secondly, it is expression of their genes. The nucleotide sequence analysis of the LTR-IS/MuRRS elements revealed putative transcription regulatory signals, including a TATA box and a polyadenylation signal AATAAA (5), as well as partial homologies to retroviral gag and pol genes (4). We have shown previously that the LTR-IS elements contain functional RNA polymerase II promoters, which require enhancement by cis- or trans-activating factors (8). However, their transcriptional or translational products, if any, in vivo, have not yet been identified. Finally, one possible activity might be related to the insertion of new regulatory signals into the genome. Experimental evidence for the lastly mentioned activity will be provided in this paper.

METHODS

cDNA library

Total RNA from A20/J cells was prepared by guanidinium isothiocyanate extraction and CsCl centrifugation (9). The cDNA library in lambda gt10 phages was prepared according to the procedure of Gubler and Hofmann (10).

Genomic library from BALB/c mice

The library was prepared from male liver DNA in lambda EMBL 3 as described (11) with minor modifications (7). Hybridization probes: the A1/Sau3a probe is the 300bp fragment of the LTR-IS 5' flanking sequence; the A1/EcoRI/BgIII is the 5' end of the A1 cDNA clone.

S1-nuclease mapping

The experiments were performed essentially as described by Berk and Sharp (13). The RNA from A20/J cells was annealed with the

labelled HincII/SphI fragment (Fig. 4) at 49° C for 12 hrs, incubated with 58u of S1 nuclease, and analysed on a sequencing gel.

DNA sequencing

DNA fragments were subcloned in M13mp18/19 vectors and sequenced by the dideoxymethod (14).

RESULTS

LTR-IS-hybridizing RNA in mouse tissues

To search for LTR-IS-specific transcripts total and poly A-selected RNA from various mouse tissues and cell lines was prepared and hybridized to a radioactively labeled cloned BgIII/HindIII fragment of the LTR-IS clone B8 (5). We have shown previously that this part of the LTR-IS sequences is specific for the elements and does not contain sequences present in any other known elements (7). Northern blot analysis (Fig. 1) revealed the RNA from A20/J B-lymphoid cell line as the only one positively hybridizing with the LTR-IS probe. All the other

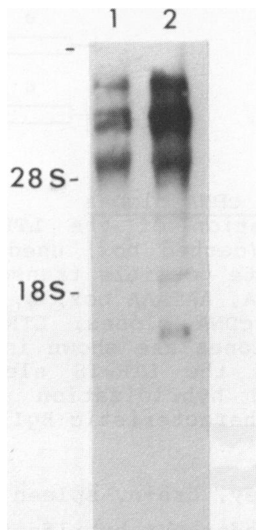

Fig. 1
Northern blot analysis of RNA from A20/J cells
4 µg of total (1) or poly A-selected (2) RNA from A20/J cells were analysed on 1.5 % agarose gel and hybridized with the Sp6-labeled LTR-IS BgIII/HindIII fragment (see Methods).

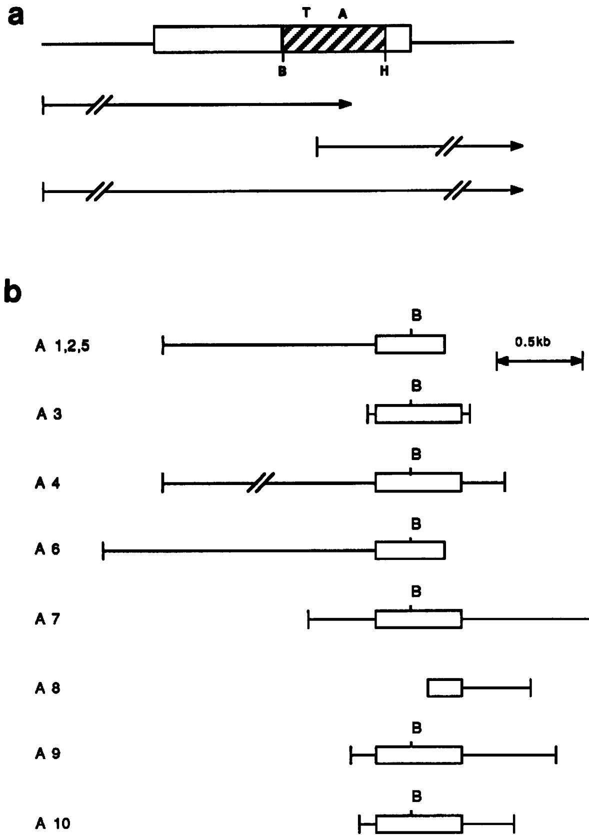

Fig. 2
Identification of LTR-IS cDNA clones
a) Schematic representation of the LTR-IS element and its BgIII/HindIII fragment (dashed box) used as the hybridization probe. The arrows indicate possible transcripts detectable with the probe. T, TATA box; A, AATAAA box; B, BgIII; H, HindIII. b) Summary of the LTR-IS cDNA clones. LTR-IS-hybridizing EcoRI fragments of ten cDNA clones are shown in the 5' to 3' orientation. The position of the LTR-IS element, as located by restriction mapping and hybridization is indicated by open boxes. Only the LTR-IS characteristic BgIII site (B) is shown.

tested RNAs (liver, kidney, brain, spleen, thymus, L-cells, EL4 thymoma) were negative under our hybridzation conditions.

Identification of LTR-IS cDNA clones from the A20/J cells

The only unequivocal identification of LTR-IS specific transcripts is by isolation and characterization of their corresponding cDNA clones. For this purpose a cDNA library from

```
B8   TATAAAAACT GAAAACTCTT TCCCCTCGAG GTGGACTCCT CTACCCCTGC  50
A1   .......... .GG.C.C... ..T.....G. ..CA...... ..........  50
A6   .......... ..G.C.C... ..T.....G. ..C....... ..........  50

B8   ATGGGATATG AGTCGTCCCC AGAGCTCTGG CTTTCCCCGA ATAAAGCCTC 100
A1   G......... ....A.T... .......... ........ . .......... 100
A6   G......... ....AT T.. ........A. .... ..... .......TC. 100

B8   ATGTGGTTTG CAACAAGCTC GGTCTATCGT GAGTTCTTGG            150
A1   G......... ..T...AAAA AAAAA.AAAA A.AAAAAAAA            150
A6   TA........ ..T...AAAA AAAAA.AAAA A.                    150
```

Fig. 3
3' terminal sequences of the two cDNA clones A1 and A6
The nucleotide sequences of the LTR'IS B8 element (5), and A1, and A6 cDNA clones are compared. Only the LTR-IS sequence is printed in full, dots indicate identical nucleotides. TATA and AATAAA signals are underlined.

the RNA of the A20/J cells was constructed as described in Methods. The library was screened with the cloned BgIII/HindIII LTR-IS fragment as a probe and the positively hybridizing clones were isolated.

The probe should detect theoretically transcripts which a) initiate at the LTR-IS promoter, b) are processed/terminated at the LTR-IS AATAAA signal, but initiate at other cellular promoters, or c) are readthrough transcripts, which initiate and terminate at other cellular signals (Fig. 2 a).

To distinguish among these three possibilities the position of the LTR-IS segment within the cDNA clones was mapped. Figure 2 b is a summary of identified clones by Southern blot analysis. It shows that the majority of the cDNA clones belongs to the "read through" type, and contains the LTR-IS sequence surrounded by other cellular sequences. The clone A8 has the LTR-IS sequence located at its 5' end and thus may correspond to a transcript which initiates within the LTR-IS element. The clones A1 and A6 have the LTR-IS sequence at their 3' ends and thus may represent transcripts which terminate in the LTR-IS elements.

Nucleotide sequece analysis of the cDNA clones

The nucleotide sequence of the clone A8 revealed a truncated LTR-IS element in the 5' to 3' orientation. The first nucleotides of the sequence correspond to the LTR-IS TATA box, the

fact which rules out the initiation at the LTR-IS promoter (data not shown).

The sequences of the clones A1 and A6 (Fig. 3) show homology with the LTR-IS sequence at their 3' ends up to 20 nt behind the polyadenylation signal of the LTR-IS element and continue further with poly A tails. This result confirms the assumption that the clones A1 and A6 represent cellular transcripts with their 3' ends processed and polyadenylated under the control of the LTR-IS AATAAA box. The clones A2 and A5 are identical with the A1 clone, suggesting that this type of transcript is abundant in A20/J cells.

The structure of the genomic A1 locus in BALB/c mice

The 700 bp EcoRI/BglII fragment from the 5' end of the cDNA clone A1 was used as a hybridization probe to screen genomic library from BALB/c mouse liver. The restriction maps of the isolated genomic locus and the cDNA clone A1 are collinear at their 3' ends up to the DraI site (Fig. 4). An open reading frame terminated with a stop codon at its 3' end, is interrupted at its 5' end, 4 bp distal from the DraI site, with an intron (sequence not shown).

Since the A1 locus displays all characteristics of a genuine protein coding gene, we conclude that the LTR-IS insertion is within its 3' untranslated region, despite of the fact that the translation remains to be shown.

The 3' terminal sequences of the both clones are compared in the lower part of the Figure 4. 153 bp downstream from the LTR-IS AATAAA signal is the sequence AAGAAATAAG, which could be a former mutated polyadenylation signal.

The genomic organization of the A1 locus

Southern blot analyses were performed to find out whether the A1 locus of the A20/J tumor cells is identical with the A1 locus of normal mouse tissues (e. G. liver), and how many copies of the A1 locus are in the genome.

Figure 5 shows fragments obtained from the DNA of A20/J cells and BALB/c mouse liver, respectively, which hybridize to the A1 probe. The restriction patterns indicate that the A1 locus is identical in the both cell types, and therefore is not caused by a DNA rearrangement in the tumor cells.

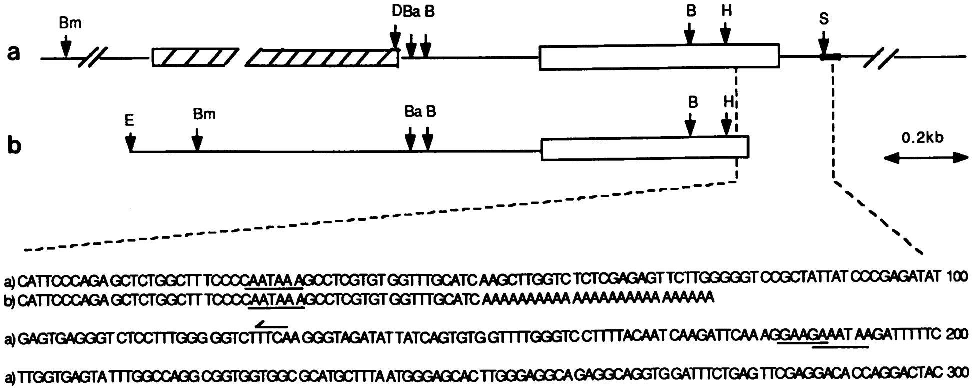

Fig. 4
The comparison of the A1 genomic and cDNA clones
The upper part shows the restriction maps, and the lower part the nucleotide sequences of the 3' ends of a) the A1 genomic clone, and b) the A1 cDNA clone. Only the relevant sites are indicated; B, BglII; Ba, BalI; Bm, BamHI; D, DraI; E, EcoRI; H, HincII; S, Sphl. The LTR-IS sequence is indicated as an open box, the B1 sequence as a black bar, the intron sequence is a dashed box; the polyadenylation signals are underlined; the 3' end of the LTR-IS sequence is marked with an arrow.

The Southern blot analysis indicates furthermore that the A1 locus is probably a single copy gene, since no other A1 homologous sequences were detected in the BALB/c mouse. This is in agreement with the observed frequency of A1 hybridizing phages from the genomic library.

Determination of the 3' ends of the A1 transcripts in BALB/c mice

Northern blot hybridization of the RNA from the A20/J cells with the A1 probe revealed a uniform transcript of about 1600 nt (Fig. 6 a). To determine the 3' end of this transcript, S1 nuclease protection assay was carried out. The size of the protected fragment (Fig. 6 b) maps the 3' end of the transcript to about 20 nt distal from the LTR-IS polyadenylation signal, which corresponds well to the 3' end of the A1 cDNA clone determined by sequencing (Fig. 3). It proves that the LTR-IS AATAAA box is responsible for generating the 3' end of the A1 transcript. The fact that no other protected fragments were detected indicate that the distal AATAAG signal is not used.

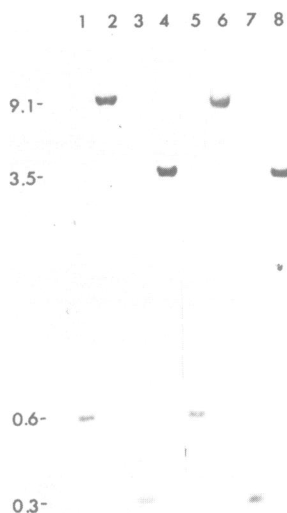

Fig. 5
Southern blot analysis of the Al locus in A20/J cells and in BALB/c mouse liver
10 ,ug aliquots of A20/J DNA (1-4) or BALB/c liver DNA (5-8) were digested with restriction enzyme BgIII (;? &), EcoRI (2, 6) Sau3a, (3, 7), Xbal (4, 8), analysed on 0.8 % agarose gel, and hybridized to Al/Sau3a probe (see Methods)

DISCUSSION

The generation of the 3' end of eukaryotic mRNA is an important mechanism to regulate gene expression (for review see ref. 15, 16). It is known that sequences downstream from the polyadenylation site are required for generating correct 3' ends of some mRNAs (17, 18). In some cells, a selection of alternative polyadenylation sites is used to produce different 3' ends of one transcript (19-22). The sequences within 3' untranslated region can affect the mRNA stability as it was shown for a population of eukaryotic mRNAs (23). Also, a transposon insertion into 3' untranslated region with a modulatory effect on gene expression was described in Drosophila (24).

Here we report that the insertion of a mobile element, LTR-IS/MuRRS, into the 3' untranslated region of a mouse gene (Al) provided a functional polyadenylation signal.

The finding of a AATAAG sequence 50 bp downstream of the LTR-IS insertion suggests that this could have been an original poly-

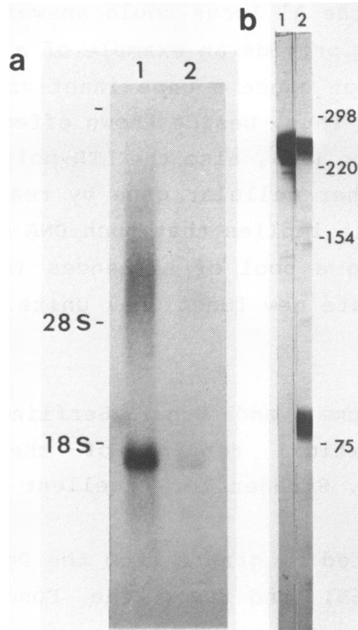

Fig 6
Analysis of the Al transcripts from A20/J cells
a) 5 μg of poly A-selected (1) or total (2) RNA from A20/J cells were analysed by Northern blotting and hybridized to the Al/EcoRI/BglII probe. b) Determination of the 3' end of the Al transcript by S1 nuclease protection. The Al/HincII/SphI fragment, 262 bp, was labeled by T4 polymerase, annealed with the RNA, and digested with S1 nuclease (see Methods). (1) labeled fragment, (2) 65 μg of A20/J RNA and the labeled fragment. The pBR322/HinfI marker is indicated.

adenylation signal of the Al gene. The AATAAG mutation was described for a thalassemic globin gene, where it retains only about 15 % of the activity (25). It is possible that the original polyadenylation signal of the Al gene was mutated before the insertion event, and the LTR-IS insertion was selected because it provided an advantage of a better polyadenylation signal. However, it cannot be excluded that the downstream signal could have been mutated after the insertion event, since it was no longer under selection pressure. We do not know, whether the alteration of the original 3' end of the Al transcript, due to the LTR-IS insertion, had an effect on the Al gene expression. A finding of a mouse strain without the

LTR-IS insertion in the A1 locus could answer this question. The described results provide an example of a mobile element insertion, which did not cause a gene inactivation, but probably led to a gene rescue. Thus, beside known effects of LTR enhancer and promoter insertion (26), also the LTR-polyadenylation signal can be fused to another cellular gene by rearrangements in the mammalian genome. This implies that such DNA elements may have a function by providing a pool of sequences in the genome which can be used to generate new functional units.

ACKNOWLEDGEMENTS

We thank Ingrid Grummt and Edgar Serfling for stimulating discussions and critical reading of the manuscript, and I. Sauer-Clark and C. Straßer for excellent technical assistance.

This work was supported by grants from the Deutsche Forschungsgemeinschaft (SFB 165) and from the Fonds der Chemischen Industrie.

*To whom correspondence should be addressed

REFERENCES

1. Shapiro, J. A. (1983) Mobile genetic elements. Academic Press, Inc. New York.
2. Levis, R., O'Hare, K. and Rubin, G. M. (1984) Cell 38, 471-481.
3. Kidd, S., Locket, T. J. and Young, M. W. (1983) Cell 34, 421-433.
4. Schmidt, M., Wirth, T., Kröger, B. and Horak, I. (1985) Nucl. Acids Res. 13, 3461-3470.
5. Wirth, T., Glöggler, K., Baumruker, T., Schmidt, M. and Horak, I. (1983) Proc. Natl. Acad. Sci. USA 80, 3327-3330.
6. Wirth, T., Schmidt, M., Baumruker, T. and Horak, I. (1984) Nucl. Acids Res. 12, 3603-3610.
7. Schmidt, M., Glöggler, K., Wirth, T. and Horak, I. (1984) Proc. Natl. Acad. Sci. USA 81, 6696-6700.
8. Köhrer, K., Grummt, I. and Horak, I. (1985) Nucl. Acids Res. 13, 2631-2645.
9. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979) Biochemistry 18, 5294-9299.
10. Gubler, U. and Hoffman, B. J. (1983) Gene 25, 263-269.
11. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.
12. Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., and Green, M. R. (1984) Nucleic Acids Res. 12, 7035-7056.
13. Berk, A. J. and Sharp, P. A. (1977) Cell 12, 721-732.

14. Sanger, F., Nicklen, S. and Coulsen, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467
15. Birnstiel, M. L., Busslinger, M. and Strub, K. (1985) Cell 41, 349-359.
16. Platt, T. (1986) Ann. Rev. Biochem. 55, 339-372.
17. McDevitt, M. A., Imperiale, M. J., Ali, H. and Nevins, J. R. (1984) Cell 37, 993-999.
18. Gil, A. and Proudfoot, N. J. (1984) Nature 312, 473-474.
19. Aho, S., Tate, V. and Boedtker, H. (1983) Nucleic Acids Res. 11, 5443-5450.
20. Henikoff, S., Sloan, J. S. and Kelly, J. D. (1983) Cell 34, 405-414.
21. Hagenbüchle, O., Wellauer, P. K., Cribbs, D. and Schibler, U. (1984) Cell 38, 737-744.
22. Amara, S. G., Evans, R. and Rosenfeld, M. G. (1984) Molec. Cell. Biol. 4, 2151-2160
23. Shaw, G. and Kamen, R. (1986) Cell 46, 659-667.
24. Campuzano, S., Balcells, L., Villres, R., Carramolino, L., Garcia-Alonso, L. and Modolell, J. (1986) Cell 44, 303-312.
25. Higgs, D. R., Goodbourn, S. E. Y., Lamb, J., Clegg, J. B., Weatherall, D. J., and Proudfoot, N. J. (1983) Nature 306, 398-400.
26. Hayward, W. S., Neal, B. G. and Astrin, S. M. (1981) Nature 290, 475-480.