A novel transposon-like repeat interrupted by an LTR element occurs in a cluster of B1 repeats in the mouse *c-mos* locus

Friedrich Propst and George F. Vande Woude

LBI-Basic Research Program, NCI-Frederick Cancer Research Facility, P.O. Box B, Frederick, MD 21701, USA

Received 2 August 1984; Revised and Accepted 22 October 1984

ABSTRACT

The mouse genomic locus containing the oncogene *c-mos* was analyzed for repetitive DNA sequences. We found a single B1 repeat 10 kb upstream and three B1 repeats 0.6 kb, 2.7 kb, and 5.4 kb, respectively, downstream from *c-mos*. The B1 repeat closest to *c-mos* contains an internal 7-bp duplication and a 18-bp insertion. Localized between the last two B1 repeats is a copy of a novel mouse repeat. Sequence comparison of three copies of this novel repeat family shows that they a) contain a conserved *Bgl*II site, b) are ~420 bp long, c) possess internal 50-bp polypurine tracts, and d) have structural characteristics of transposable elements. They are present in about 1500 copies per haploid genome in the mouse, but are not detectable in DNA of other mammals. The *Bgl*II repeat downstream from *c-mos* is interrupted by a single 632-bp LTR element. We estimate that ~1200 copies of this element are present per haploid genome in BALB/c mice. It shares sequence homology in the R-U5 region with an LTR element found in 129/J mice.

INTRODUCTION

The *Alu* repeat family is the most prominent family of short interspersed repetitive DNA sequences in the mammalian genome (1). The human *Alu* repeat, ~300 bp long, appears as a dimer of the related mouse B1 repeat. Within each species, individual repeats are usually more than 80% homologous in nucleotide sequence, but of the few that have been examined none are identical. There are 10^5 to 5×10^5 copies of mouse B1 repeats per haploid genome. While their average distribution would be expected to be from 6-10 kb between individual repeats, clusters of B1 repeats (2) and *Alu* repeats (3) separated only by a few hundred base pairs have been described. An additional group of interspersed repetitive DNA sequences in the mouse genome are sequences closely related to retroviral proviruses. Among these are the endogenous murine leukemia virus (MuLV)-related proviruses (50-100 copies per haploid genome) (4), genes for intracisternal A-particles (1000-2000 copies) (5), VL30-related sequences (50-100 copies) (6), and solitary LTR elements (~500 copies) (7,8,9).

We have mapped the highly repetitive DNA sequences in the mouse genomic locus containing the c-mos oncogene. We identified three B1 repeats clustered within a 5.4-kb segment of DNA downstream from c-mos. Prompted by reports that DNA sequences between closely spaced Alu repeats can be polymorphic and repetitive as well (3), we looked for the complexity of the sequences located between two of the B1 repeats and found a novel mouse repeat. By sequencing this region and homologous sequences in independently isolated genomic clones, we were able to define this novel repeat and show that it has structural features resembling a transposon. This repeat downstream from c-mos is itself interrupted by a single LTR element.

MATERIALS AND METHODS

DNA

Clone pMS2 containing the mouse c-mos gene from an EcoRI site 3 kb upstream of the gene to an EcoRI site ~10 kb downstream was a plasmid subclone from a phage previously described (10). Recombinant phage λ MSU was purified from a mouse genomic library provided by M. Oskarsson. This library was prepared from MboI partial digests of NIH3T3 DNA cloned into the BamHI sites of the lambda phage vector L47.1 (11) and was screened with a 3.3-kb SacI/HindIII fragment containing c-mos. The insert of λ MSU contains the c-mos gene including ~10 kb of upstream and 1.7 kb of downstream mouse sequences.

Clone MS5 is a 1.4-kb HindIII/BglII fragment 3 kb downstream from c-mos cloned into HindIII and BglII sites of pHC79 (12). It was subcloned from pMS2. Clone MS57 is a subclone of MS5 containing only the LTR element on a BglII/MboII fragment cloned into the BglII site and the blunt-ended HindIII site of pHC79 (12).

Recombinant phages λ MS51 and λ MS52 were obtained by screening a mouse genomic library (a partial MboI digest of BALB/c embryonic DNA cloned into BamHI sites of phage CH28; gift from P. Leder) with nick translated clone MS5 DNA as a probe. Clone 51B is a 727-bp BglII fragment obtained from λ MS51 which hybridizes to nick translated clone MS5 DNA. This fragment was cloned into the BglII sites of pHC79 (Boehringer Mannheim; 12). Clone 52H is a 2.2-kb HindIII fragment obtained from λ MS52, which hybridizes to labeled clone MS5 DNA. It was cloned into the HindIII site of pBR322. Clone 521 is a subclone of 52H containing a complete BglII repeat with 19 bp of upstream flanking sequence on a BanI/AccI fragment cloned into the BamHI site of pBR322 with the help of BamHI linkers.

Clone pMr118 contains a mouse B1 repeat cloned into the PstI site of pBR322. It was a gift from N. Hastie.

High-molecular-weight genomic DNA of various species was a gift of M. Tainsky.

Southern hybridization analysis

The standard laboratory techniques were used in digestion, agarose gel electrophoresis, and Southern blotting (13) of genomic and cloned DNA. Probes were prepared by nick translation (14) of gel purified fragments. High-molecular-weight mouse genomic DNA was sonified for 10-20 minutes to an average fragment length of 600 bp prior to labeling. Two different hybridization conditions were used: a) High-stringency hybridization was carried out in 50% formamide, 0.1% (w/v) Ficoll 400, 0.1% (w/v) polyvinylpyrrolidone, 0.1% (w/v) bovine serum albumine, 0.75 M NaCl, 0.075 M sodium citrate, 0.05 M sodium phosphate (pH 6.5), and 0.5 mg/ml sonicated salmon sperm DNA at 42°C for at least 24 hours. Filters were then rinsed three times for 5 minutes in 0.3 M NaCl, 0.03 M sodium citrate, and 0.1% (w/v) sodium dodecylsulfate at room temperature and three times for 20 minutes in 0.015 M NaCl, 0.0015 M sodium citrate, and 0.1% (w/v) sodium dodecylsulfate at 55°C. b) Low-stringency hybridization was carried out as under a), except that 30% formamide was used during hybridization and the final three washes were done in 0.15 M NaCl, 0.015 M sodium citrate, and 0.1% (w/v) sodium dodecylsulfate at 55°C. Filters were exposed to x-ray film (XAR, Kodak) for the times indicated in the figure legends.

DNA sequencing

Sequences were determined by the method of Maxam and Gilbert (15) from overlapping fragments of one or the other strand, except for the B1_{0.6} repeat and part 1 of the BglII repeat in MS5, where both complementary strands were sequenced.

Computer analyses of DNA sequences

Sequence analyses was performed using a DecSystem-10 computer. Sequences obtained in this study were compared to sequences in the Genetic Sequence Data Bank (GenBank) and selected other sequences using the programs SRCHN and NUCALN (16) and SEQ (17) available in the DecSystem-10.

RESULTS

B1 repeats

High-copy-number repeats were first mapped by the Southern hybridization technique using labeled mouse genomic DNA as a probe. Under low-stringency

Fig. 1

The location of identified repeats in the mouse *c-mos* locus

Small solid bars indicate the position of the B1 repeats B1_u, B1_{0.6}, B1_{2.7}, and B1_{5.4}. The subscript U refers to upstream and the subscript numbers refer to the distance in kb downstream from *c-mos*. Not all of the known restriction endonuclease sites present in this locus are indicated.

conditions, this probe detects B1 repeats as well as other highly repetitive DNA sequences (not shown). Four such repeats were found and carefully mapped within ~20 kb of cloned DNA of the mouse *c-mos* locus. Subsequent hybridization with a cloned B1-specific probe gave an identical pattern, indicating that all four repeats are B1 repeats (B1_u, B1_{0.6}, B1_{2.7}, and B1_{5.4}; Fig. 1). Using mouse genomic DNA as a probe no other repeats are detected

		40
B1 consensus	CCGGGCGTGGTGGCGCACGCCTTTAATCCAGCACTCGGGAG	
B1 _{0.6}	taaaggaacagttgttaa..A.CA...A...A...TA.....ATA...	
B1 _{2.7}	ttagaatatcttcttttg.T...T.....A..T.....A.....T...	
B1 _{5.4}	ttgagagaaaacagaaag.T...AG...ATA...A.....G.....TT...	
		95
B1 consensus	GCAGAGGCAGGCGGATTCT-----GAGTTCGAGGCCAGCCTGGTCTACAGAGTGAGT	
B1 _{0.6}	TT...C...GATCTCT..A..TA.....A.....A.....	
B1 _{2.7}	T.....A.....A.....	
B1 _{5.4}	AA.....T.....A.....A.....	
		134
B1 consensus	TCCAGGAC-----AGCCAGGGCTACACAGAGAAACCCTGTCTNN	
B1 _{0.6}	...A...TACAGGCTACTACTGTAT.....T..T....AGaac	
B1 _{2.7}	aaacaaacaaacaaacaaatca <u>tccttttc</u> gaaacc	
B1 _{5.4}	.T..... <u>gagagagaga</u> aaac	
		134
B1 _{0.6}	aagcagaacaaacaaacaaatgaaaaggttcctttcccttcagtt	
B1 _{2.7}	aaacaaacaaacaaacaaatca <u>tccttttc</u> gaaacc	
B1 _{5.4}	aaacaaacaaacaaacaaacaaac <u>gagagagaga</u> aaac	

Fig. 2

The sequence of the B1 repeats downstream from *c-mos*

The sequences of B1_{0.6}, B1_{2.7}, and B1_{5.4} are compared to the B1 consensus sequence (2). Only nucleotides that differ from the B1 consensus sequence are indicated. Dashes indicate gaps introduced to obtain an optimal alignment. The numbering refers to the B1 consensus sequence (2). Lower case letters indicate flanking sequences. Putative direct repeats flanking the B1 repeats are underlined.

Fig. 3

Hybridization of a MS5 probe to a Southern blot of genomic DNA

This probe contains a part of the *Bgl*II repeat plus the LTR element and flanking cellular sequences (Fig. 1). Hybridization was carried out at low stringency to *Eco*RI-digested genomic DNA of human (lane 1), monkey (lane 2), bovine (lane 3), dog (lane 4), cat (lane 5), mouse (BALB/c) (lanes 6 and 10), mink (lane 7), rat (lane 8), and hamster (lane 9). The x-ray film was exposed for 15 hours at -70°C .

in this DNA fragment.

B10.6, B12.7, and B15.4 were sequenced. They are respectively 84%, 93%, and 88% homologous to the B1 consensus sequence (2) (Fig. 2). B10.6 has a 7-bp duplication at position 62 and a 18-bp insertion at position 103. The repeats are flanked by imperfect direct repeats (underlined sequences in Fig. 2) of 12 bp (7 matches), 11 bp (9 matches), and 16 bp (12 matches), respectively. All three repeats have the typical A-rich sequence at their 3' end and are in the same orientation as the *c-mos* gene. The position of

The BglII repeat downstream from c-mos (MS5) is compared to the BglII repeats in clones 52H and 51B. For this alignment, the LTR element inserted in the BglII repeat downstream from c-mos and one of the 4-bp direct repeats flanking it (Fig. 6A) were omitted. Clone 51B contains only 326 bp of the repeat sequence, since the BglII site within the repeat was used as cloning site. The numbers refer to the BglII repeat in MS5. For the repeats in 52H and 51B, only nucleotides different from the sequence in MS5 are indicated. Gaps, introduced to produce an optimal alignment, are marked by dashes. Inverted repeats at the ends of the BglII repeats are indicated by horizontal arrows. Cellular sequences flanking the repeats are given in lower case letters. The direct repeats flanking the BglII repeat in MS5 and the BglII sites are boxed. The position of the interruption of the BglII repeat in MS5 by the LTR element at the BglII site is indicated.

The BglII Repeat

To identify the repetitive sequences in MS5, two recombinant phages (λ MS51 and λ MS52) possessing sequences homologous to an MS5 probe were isolated from a mouse library. Restriction fragments containing the related

Fig. 5**Hybridization at high stringency of a 51B probe to mouse genomic DNA**

The probe contains 326 bp of BglII repeat sequence plus 400 bp of flanking cellular sequence. 20 μ g samples of genomic mouse DNA were digested with HindIII (lane 1), BglII (lane 2), SacI (lane 3), or XbaI (lane 4). The x-ray film was exposed for 17 hours at -70°C .

sequences in the DNA recombinant phage were subcloned into plasmid vectors (clones 51B and 52H, respectively), and, in turn, served as probes to map the exact position of the homologous sequences in MS5. By sequencing the related regions in 51B, 52H, and MS5, we identified a novel mouse repeat family (Fig. 4). Because the three sequences contain a conserved BglII site, we refer to them as BglII repeats. The repeats of MS5 and 52H are 418 bp and 417 bp, respectively. They end in inverted repeats, the first and last two nucleotides being TG and CA, respectively. The repeat in MS5 but not the one in clone 52H is flanked by 6-bp direct repeats. All three repeats have a polypurine tract of about 50 bp (approximately at position 190 to 240), which is the least conserved region between the individual members. The overall homology is 85% between the repeats in 51B and 52H, 78% between the repeats in MS5 and 52H, and 79% between the repeats in MS5 and 51B. There is only one copy of the BglII repeat in the 20-kb c-mos locus and in the two independent recombinant phages λ MS51 and λ MS52 (not shown).

When labeled 51B DNA was hybridized at high stringency to a Southern blot

(A) Sequence organization of

of mouse genomic DNA, intense regions of hybridization were observed (Fig. 5). The major hybridizing band in BglII digested DNA (Fig. 5, lane 2) is ~1.1 kb, and there are minor bands of ~0.9 kb and ~0.8 kb. The same result was obtained when clone 521 was used as probe. When DNA from λ MS51 and λ MS52 was digested with BglII and probed with MS5 DNA, a ~0.7-kb fragment was detected in each case. These two fragments differed in size by about 50 bp (not shown).

8388

the mammalian and viral sequences in the Genetic Sequence Data Bank (GenBank release 24, September 4, 1984) or to many recently published mouse repetitive sequences.

The LTR element

By sequencing the BglII repeat downstream from c-mos, we found that it was interrupted by an LTR element at the BglII site (Fig. 6A). The LTR element is flanked by 4-bp direct repeats which are part of the sequence of the BglII repeat probably duplicated upon integration of the LTR element or the putative retrovirus to which it belonged. The structure of the element (Fig. 6B) is similar to LTR elements of mammalian proviruses (19). A 7-bp inverted repeat and the putative U3, R, and U5 sequences are indicated (Fig. 6B). Under low-stringency hybridization conditions this element could not be detected in genomic DNA of other mammals (Fig. 3) or in the cloned human mos locus (18). To estimate the copy number of the LTR element we hybridized an LTR-specific probe (clone MS57) at low stringency to a lambda phage library of BALB/c genomic DNA (not shown). 1500 copies per haploid genome were estimated. Most, but not all, of the recombinant phages detected with the clone MS57 probe also hybridized to the U3-specific BglII/XhoI fragment of MS57 (yielding a copy number of ~1200 [not shown]). This indicates that R and U5 regions are shared among LTRs with different U3 sequences. We found 72% sequence homology between the LTR element in MS5 and a novel LTR element associated with a VL30 element (E. Keshet, personal communication); this homology was confined to R and U5 sequences. In addition, a solitary LTR element found in 129/J mice (9) shares 66% homology to the R and U5 regions of the LTR element in MS5 (Fig. 6B). A computer search did not detect homology to other published LTR sequences in the Genetic Sequence Data Bank (GenBank release 24, September 4, 1984).

DISCUSSION

The novel interspersed BglII repeat described here has two features characteristic of transposable elements. First, the copy downstream from c-mos is flanked by 6-bp direct repeats of cellular sequence. The duplication of the target site is characteristically associated with transposition (20). Analogous direct repeats have been observed for interspersed repetitive sequences of other families as well (1). Moreover, the ends of the BglII repeat form 7-bp inverted repeats, and the first two nucleotide pairs of the inverted repeat TG/CA are found in most prokaryotic and eukaryotic transposable elements (20). Our observation of predominant hybrid-

izing bands in Southern analyses of genomic mouse DNA (Fig. 5) may indicate that a number of BglII repeats are part of a larger repeat. However, the three genomic loci analyzed in this study (MS51, MS52, and the MS5 c-mos locus) contain one copy of a BglII repeat with no homology in sequences flanking the repeats.

An interesting feature of the BglII repeat in the c-mos locus is its interruption by an LTR element at the BglII site. Comparison of the repeat in MS5 with the other two BglII repeats shows that the LTR element was inserted into this copy of the BglII repeat. This could mean that at a certain point in the evolution of the mouse, a retroviral provirus has integrated 4 kb downstream from c-mos. Excision of this provirus by homologous recombination would result in a single LTR flanked by direct repeats. Alternatively, the BglII repeat in which the putative provirus was already integrated (and excised) moved to its present location downstream from c-mos.

The two novel transposon-like elements described in this paper, the BglII repeat and the LTR element, were found by using methods that did not depend on homology to previously cloned mouse repetitive sequences or LTR elements. It is possible that with these methods other repetitive elements will be found and that the mouse genome harbors more of these elements than originally anticipated.

We found only one high-copy-number repeat (B1_g) upstream from the c-mos gene. Downstream from c-mos, however, we found three B1 repeats (B1_{0.6}, B1_{2.7}, and B1_{5.4}) within a distance of 5.4 kb from the c-mos gene. Between B1_{2.7} and B1_{5.4} we found the novel BglII repeat which is interrupted by a single LTR element. This raises the question whether this peculiar arrangement of repetitive sequences has any influence on the expression of the c-mos gene. Expression of this gene has not been detected in normal cells (21, 22), but activation has occurred in several plasmacytomas in vivo by integration of an intracisternal A-particle genome 5' to c-mos (23-25). Similar activations of other oncogenes by 5' or 3' insertion of a provirus have also been observed (26-28). In vitro, the c-mos gene can be activated by linking a Moloney murine sarcoma virus LTR 5' or 3' to the gene, but the removal of a cis-acting inhibitor sequence 1-1.5 kb upstream from the gene is required (29). In preliminary transfection experiments, the BglII repeat and the LTR element did not enhance the transforming activity of the v-mos or the c-mos gene in NIH3T3 cells. On the other hand, clusters of Alu repeats have been shown to confer genetic instability

to a genomic locus (3), and rearrangements of genetic loci containing proto-oncogenes like the *c-myc* oncogene (although at the 5' end of the gene) may contribute to development of certain tumors (30). As a first step towards understanding a possible biological role for this cluster of repetitive DNA sequences, we are looking for restriction fragment length polymorphism in this locus.

ACKNOWLEDGEMENT

We thank N. Hastie for the generous gift of cloned mouse repetitive sequences and E. Keshet for making his results available to us prior to publication.

Research sponsored by the National Cancer Institute, DHHS, in part through the Laboratory of Molecular Oncology, NCI-FCRF and in part under Contract No. NO1-CO-23909 with Litton Bionetics, Inc. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U. S. Government.

REFERENCES

1. Schmid, C. W. and Jelinek, W. R. (1982) *Science* 216, 1065-1070.
2. Kalb, V. F., Glasser, S., King, D. and Lingel, J. B. (1983) *Nucl. Acids Res.* 11, 2177-2184.
3. Calabretta, B., Robberson, D. L., Barrera-Saldana, H. A., Lambrou, T.P. and Saunders, G.F. (1982) *Nature* 296, 219-225.
4. Coffin, J. (1982) *Molecular Biology of Tumor Viruses*, 2nd edn. pp. 1109-1203, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
5. Kuff, E. L., Feenstra, A., Lueders, K., Smith, L., Hawley, R., Hozumi, N. and Shulman, M. (1983) *Proc. Natl. Acad. Sci. USA* 80, 1992-1996.
6. Itin, A. and Keshet, E. (1983) *J. Virol.* 47, 656-659.
7. Ou, C., Boone, L. R. and Yang, W. K. (1983) *Nucl. Acids Res.* 11, 5603-5620.
8. Rotman, G., Itin, A. and Keshet, E. (1984) *Nucl. Acids Res.* 12, 2273-2282.
9. Wirth, T., Glogler, K., Baumbaker, T., Schmidt, M. and Horak, I. (1983) *Proc. Natl. Acad. Sci. USA* 89, 3327-3330.
10. Oskarsson, M., McClements, W. L., Blair, D. G., Maizel, J. V. and Vande Woude, G. F. (1980) *Science* 207, 1222-1224.
11. Loenen, W. A. M. and Brammar, W. J. (1980) *Gene* 10, 249-259.
12. Hohn, B. and Collins, J. (1980) *Gene* 11, 291-298.
13. Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517.
14. Rigby, P. W. J., Deckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
15. Maxam, A. M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
16. Wilbur, W. J. and Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* 80, 726-730.
17. Brutlag, D. L., Clayton, J., Friedland, P. and Kedes, L. H. (1982) *Nucl. Acids Res.* 10, 279-294.

18. Watson, R., Oskarsson, M. and Vande Woude, G.F. (1982) *Proc. Natl. Acad. Sci. USA* 79, 4078-4082.
19. Chen, H. R. and Barker W. C. (1984) *Nucl. Acids Res.* 12, 1767-1778.
20. Temin, H. M. (1980) *Cell* 21, 599-600.
21. Gattoni, S., Kirschmeier, P., Weinstein, B., Escobedo, J. and Dina, D. (1982) *Mol. Cell. Biol.* 2, 42-51.
22. Müller, R., Slamon, D. J., Tremblay, J. M., Cline, M. J. and Verma, I. M. (1982) *Nature* 299, 640-641.
23. Canaani, E., Dreazen, O., Klar, A., Rechavi, G., Ram, D., Cohen, J.B. and Givol, D. (1983) *Proc. Natl. Acad. Sci. USA* 80, 7118-7122.
24. Cohen, J. B., Unger, T., Rechavi G., Canaani, E. and Givol, D. (1983) *Nature* 306, 797-799.
25. Gattoni-Celli, S., Hsiao, W. W. and Weinstein, I. B. (1983) *Nature* 306, 795-796.
26. Hayward, W. S., Neel, B. G. and Astrin, S. M. (1981) *Nature* 290, 475-480.
27. Nusse, R., van Ooyen, A., Cox, D., Fung, Y. K. and Varmus, H. (1984) *Nature* 307, 131-136.
28. Tschlis, P. N., Strauss, P. G. and Hu, L. F. (1983) *Nature* 302, 445-449.
29. Wood, T. G., McGeady, M.L., Baroudy, B., Blair, D.G. and Vande Woude, G.F. (1984) *Proc. Natl. Acad. Sci. USA*, in press.
30. Leder, P., Battey, J., Lenoir, G., Moulding, C., Murphy, W., Potter, H., Stewart, T. and Taub, R. (1983) *Science* 222, 765-771.