Volume 10 Number 16 1982 Nucleic Acids Research

Characterization of a highly repetitive family of DNA sequences in the mouse

Thomas G.Fanning

Department of Pathology, School of Medicine, University of California, Davis, CA 95616, USA

Received 30 March 1982; Revised 4 June 1982; Accepted 16 July 1982

ABSTRACT

A large proportion (0.5-1%) of total mouse DNA is cleaved by Bam HI into fragments whose size is about 500 base pairs. A cloned member of this repetitive family of DNA sequences (BAM5 family) was sequenced by the dideoxy chain termination procedure and shown to contain 507 base pairs. The sequence exhibited no unusual or remarkable features. Repetitive sequences complementary to the cloned BAM5 fragment were found in rat DNA, but not in feline or human DNA. Restriction mapping suggested that many BAM5 sequences were components of much larger repetitive DNAs which were scattered throughout the mouse genome. The BAM5 sequences within the larger repetitive DNAs did not appear to be arranged tandemly or as members of scrambled tandem repeats. RNA homologous to the cloned BAM5 sequence was detected in cultured mouse cells, but not in cultured rat cells.

INTRODUCTION

Highly reiterated DNA sequences are present in the genomes of all eukaryotes (1). The simplest of these DNAs, satellite DNAs, are usually composed of relatively short sequences repeated tandemly hundreds of times (2). The major satellites of Drosophila, for example, are composed of tandemly repeated 5-10 base pair units (3), while mouse satellite DNA consists of tandem repeats of a 234 base pair unit (4).

In contrast to the tandem arrangement of satellite DNAs, many reiterated sequences in eukaryotic genomes are interspersed with single copy DNA. Interspersed repetitive DNAs have been described in a number of organisms (1). Although little is known concerning the function of most of these DNA sequences, one class of highly reiterated sequences in humans, the Alu family, has been implicated as a possible origin of DNA replication and/or RNA polymerase III transcription (5).

0305-1048/82/1016-5003$2.00/0

Many highly reiterated, interspersed DNAs may have a regulatory function (6,7), although some have postulated that these sequences simply represent "junk" DNA (8,9). A distinction between these two extreme positions is hampered by the general lack of information concerning repetitive DNA function and genomic organization. Several models describing the genomic organization of repetitive DNAs exist (1) and recently a model has been presented which suggests that many highly repetitive eukaryotic DNAs are arranged in scrambled tandem clusters (10).

The present report describes the initial characterization of a family of highly reiterated, interspersed DNA sequences in the mouse genome (BAM5 family). Many, perhaps all, BAM5 sequences are components of much larger (greater than 6 kilobase pairs) repetitive DNAs. Each BAM5 sequence within the larger repeat appears to be present once, suggesting that the larger repeats are composed neither of tandemly repeated units nor organized as scrambled tandem clusters.

MATERIALS AND METHODS

GR/A mice, tissue culture lines, and the extraction of DNA have been described previously (11,12). RNA was isolated from tissue culture cells by the method of Pedersen *et al.* (13). The cells were disrupted with Triton X-100, the unbroken nuclei pelleted through sucrose and the RNA-containing supernatant phenol extracted and ethanol precipitated. RNA prepared by this procedure is highly enriched for cytoplasmic RNAs. Blot-transfer of DNA was by the method of Southern (14). RNA was electrophoresed and blot-transfered essentially as described by Thomas (15). Generally, $2x10^6$ cpm of cloned DNA, labeled by nick translation (16) was used for filter hybridization. Filter hybridization buffers and procedures have been described previously (11).

The reiterated 500 base pair mouse DNA fragments (BAM5 family) were partially purified by sucrose gradient centrifugation (17) of *Bam HI*-cleaved mouse DNA. The fragments were ligated into *Bam HI*-cleaved pBR322 and the sought after sequences identified by colony hybridization (18) of transformed *E. coli* JM103 cells. The probe used for the colony hybridizat-

ion was Bam HI-generated total mouse BAM5 fragments isolated from an agarose gel (19) and labeled by nick translation.

One of the BAM5-containing pBR322 plasmids (designated pMBA14) was cleaved with Bam HI and the BAM5 fragment isolated by polyacrylamide gel electrophoresis. The fragment was then ligated directly into the cloning/sequencing vector M13mp9 (20) at the vector's Bam HI site. Alternatively, the BAM5 fragment was cleaved with either Alu I or Rsa I and the resultant fragments blunt-end ligated into the Hinc II site of the M13mp9 vector. Phage containing BAM5 inserts were isolated (21) and the inserts sequenced by the dideoxy chain termination procedure (22,23).

All experiments involving recombinant DNA were performed under guidelines set forth by the NIH and the IBC-UC Davis.

RESULTS

Identification of BAM5 sequences.

Several families of reiterated DNA sequences were observed after Bam HI digestion of total mouse liver DNA and agarose gel electrophoresis (Fig. 1). Densitometer tracings of the stained gel indicated that a major family of sequences (approximately 500 base pairs in size and designated the BAM5 family) accounted for about 0.5-1% of the total mouse DNA. This would suggest that as many as $0.5\text{-}1\times10^5$ copies of these sequences are present in each diploid mouse cell.

Cloning and sequence of a BAM5 fragment.

Mouse DNA was cleaved with Bam HI and the BAM5 fragments were partially purified on sucrose gradients and cloned in pBR322. One clone, designated pMBA14, was selected for further study.

Mouse liver DNA was cleaved with Bam HI and the BAM5 fragments isolated following preparative polyacrylamide gel electrophoresis. In a like manner, the cloned insert was isolated from pMBA14. Digestion of the two DNA samples with the enzymes Alu I or Rsa I demonstrated that both enzymes cleaved the DNAs several times giving rise to virtually identical patterns (Fig. 2). Digestion with other restriction enzymes produced similar results (not shown). Thus, on the

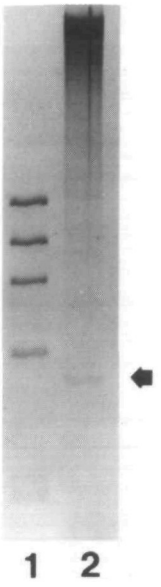

Fig. 1: Agarose gel electrophoresis of Bam HI-cleaved mouse liver DNA. 10 ug of DNA was digested to completion with Bam HI and electrophoresed into a 1.5% agarose gel (lane 2). The arrow marks the position of the BAM5 family of repetitive sequences. 1 ug of ØX 174 RF DNA was digested with Hae III and run as size markers (lane 1). The ØX 174 band in lane 1 just above the BAM5 band is 603 base pairs.

basis of conserved restriction enzyme sites, the insert in pMBA14 is representative of many of the highly repetitive BAM5 sequences in the mouse genome.

Having established that the insert in pMBA14 was representative of the bulk of the BAM5 family of sequences, the insert was recloned in a phage M13 derivative and sequenced by the dideoxy chain termination procedure. The BAM5 fragment was 507

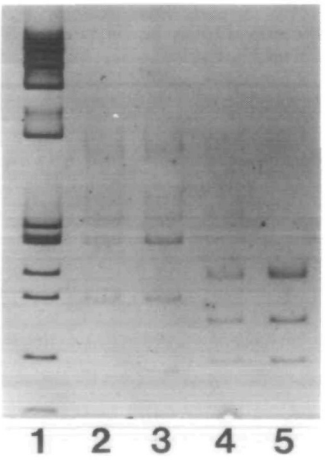

Fig. 2: Restriction mapping total mouse BAM5 reiterated sequences and a cloned BAM5 family member. Lane 1, ØX 174 RF DNA cleaved with Hae III; lanes 2 and 4, total BAM5 sequences cleaved with Alu I and Rsa I respectively; lanes 3 and 5, a cloned BAM5 family member (isolated from the plasmid pMBA14) cleaved with Alu I and Rsa I respectively. The Alu I digests shown in lanes 2 and 3 are partial digests. The gel is 4.5% polyacrylamide.

```
        10         20         30         40         50         60         70         80
TCCAGCAATACCCCTCCTGGGCATATATCCAGAAGATGCCCCAACTGGTAAGAAGGACACATGCTCCACTATGTTCATAG

        90        100        110        120        130        140        150        160
CAGCCTCATTTATAATAGCCAGAAGCTGGAACGAACCCAGATGCCCCTCAACAGAGGAATGGATACAAAAAATGTGGTAC

       170        180        190        200        210        220        230        240
ATCTACACATTGGAGTACTACTCAGCTATTAAAAAGAATGAATTTATGAAATTCCTAGCCAAATGAATGGACCTGGAGGG

       250        260        270        280        290        300        310        320
CATCATCCTGAGTGAGGTAACACACTCACAAAAGGAACTCACACAATATGTACTACTGATAAGTGGATATTAGCCCAAGA

       330        340        350        360        370        380        390        400
CCTAGGATACCAAGATATAAGATACAATCTCCTAAACACATGAAACTCAAGAAAAATGAAGACTGAAGTCTGGACACTAT

       410        420        430        440        450        460        470        480
ACCCCTCCTTAGAAGTGGGAACAAAACACCCTTGGAAGGAGTTACAGAGACAAAGTTTGGAGCTGAGATGAAAGGATGGA

       490        500
CCATGTATAGACTGCCTTATCCAGGGA
```

Fig. 3: Sequence of the BAM5 family member cloned in the plasmid pMBA14. The sequence was determined by the dideoxy chain termination method using the M13mp9 cloning/sequencing vector. The sequence shown is that corresponding to the phage + strand.

base pairs long, contained 57% A+T, and exhibited no remarkable features, e.g., blocks of short repeated sequences (Fig. 3). In addition, the entire sequence appeared to be unique and not the product of a duplication involving a shorter sequence.

Analysis of rat, feline and human DNAs.

Southern blot analysis of rat, feline and human DNAs was performed using labeled pMBA14 as probe. Rat DNA exhibited hybridization with the cloned mouse BAM5 fragment (Fig. 4b, lane 1) and, as in mouse (see below), the homologous sequences appeared to be scattered throughout the rat genome. However, only a minor fraction of the rat sequences migrated to a position corresponding to 500 base pairs in size after Bam HI cleavage (not shown).

No detectable hybridization of the BAM5 probe was apparent with either the feline or human DNAs (Fig. 4b, lanes 2 and 3) indicating that BAM5 homologous sequences are either lacking, or present in low amounts in these DNAs. Thus, BAM5 sequences may be unique to the order Rodentia.

Organization of BAM5 sequences in the mouse genome.

Mouse liver DNA was digested with Bam HI, Eco RI or Hind III and analyzed by the Southern blotting procedure using labeled pMBA14 as probe. Sequences capable of hybridizing to the probe were present in all size classes of DNA fragments

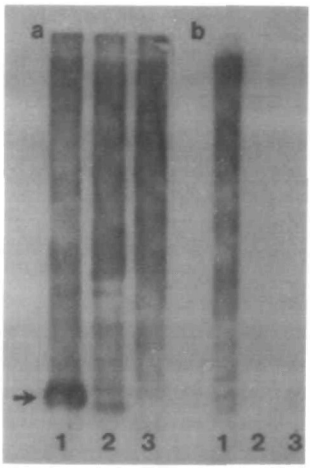

Fig. 4: Analysis of mouse, rat, feline and human DNAs for the presence of BAM5 sequences. (a) 10 ug of mouse liver DNA was digested with Bam HI (lane 1), Eco RI (lane 2) or Hind III (lane 3) and analyzed by Southern blotting and hybridization to nick translated pMBA14. BAM5 fragments in lane 1 are shown by the arrow. The autoradiogram is overexposed to bring out faint bands. (b) 5 ug of rat DNA (lane 1), 8 ug of feline DNA (lane 2) and 15 ug of human DNA (lane 3) were digested with Hind III and analyzed as in (a).

generated by these enzymes (Fig. 4a). The restriction patterns of phage lambda DNA, which had been added to the cellular DNA prior to cleavage, demonstrated that digestion was complete in each case. Thus, only a portion of the sequences homologous to the cloned BAM5 fragment contained Bam HI sites 507 base pairs apart (Fig. 4a, lane 1). The fact that all size classes of DNA appeared to hybridize to the probe, plus the absence of detectable multimers of the BAM5 fragment, suggested that most BAM5 sequences were not arranged tandemly in the mouse genome.

To further test for tandemly linked BAM5 sequences, mouse DNA was partially digested with Bam HI and analyzed by Southern blotting. No evidence for BAM5 multimers was found (not shown).

Large repetitive DNA fragments contain BAM5 sequences.

Restriction mapping using a number of enzymes indicated that many BAM5 sequences were contained within much larger repeat structures in the mouse genome (Fig. 5). The largest BAM5 containing repetitive DNA that has been found thus far is approximately 6.2 kilobase pairs in size (Fig. 5, lane 5).

The 6.2 kilobase pair Bgl I fragment (Fig. 5, lane 5) was isolated from an agarose gel, digested with several restriction enzymes and Southern blotted and hybridized with labeled BAM5 probe (Fig. 6a). The blot was then placed in boiling water for several minutes to remove the hybridized BAM5 probe and rehybridized with nick translated 6.2 kilobase pair Bgl I

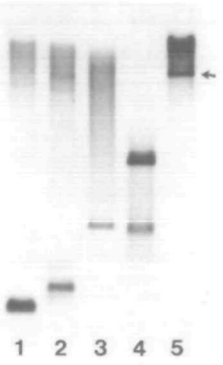

Fig. 5: Restriction mapping of total mouse DNA and detection of repetitive DNAs containing BAM5 sequences. 5 ug of mouse liver DNA was digested with Bam HI (lane 1), Msp I (lane 2), Taq I (lane 3), Hae III (lane 4) or Bgl I (lane 5) and analyzed by Southern blotting and hybridization to nick translated BAM5 DNA isolated from the plasmid pMBA14. The size of the fragment in lane 5 (arrow) was estimated to be about 6.2 kilobase pairs based upon the positions of phage lambda marker fragments (not shown).

fragments to identify all restriction fragments (Fig. 6b).

As shown in Fig. 6b, cleavage with either Bam HI or Msp I produced two restriction fragments. The size of the larger fragment in the Bam HI digest was approximately 5.5 kilobase pairs, while the larger fragment in the Msp I digest was about 5.4 kilobase pairs (Fig. 6b). In addition, each digest produced a smaller fragment which contained the total BAM5 homologous DNA present in the larger 6.2 kilobase pair Bgl I fragment (Fig. 6a). Both of the smaller fragments were identical in size to those produced by digestion of total mouse DNA with the same enzymes (Fig. 5).

Thus, the results presented in Fig. 6 indicate that the

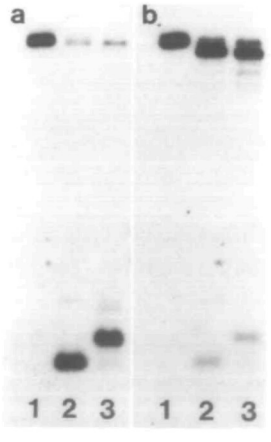

Fig. 6: Restriction mapping the 6.2 kilobase pair Bgl I repetitive fragment. The fragment (see Fig. 5, lane 5) was isolated from an agarose gel and aliquots were untreated (lane 1) or digested with Bam HI (lane 2) or Msp I (lane 3) and Southern blotted. The blot was probed with labeled BAM5 DNA isolated from the plasmid pMBA14 (a). Following autoradiography the blot was immersed in boiling water to remove the BAM5 probe and then probed with labeled 6.2 kilobase pair Bgl I fragments (b).

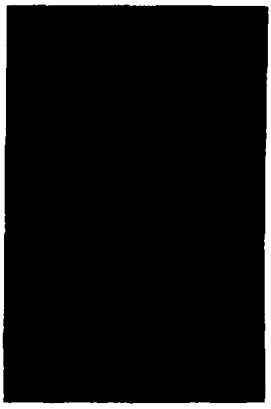

Fig. 7: Analysis of mouse and rat cell RNAs for the presence of BAM5 homologous sequences. The RNAs were isolated by a method designed to enrich for cytoplasmic RNA. Aliquots containing 10-12 ug of RNA from cultured mouse cells were untreated (lane 1) or treated with 1 ug of RNase A for 10 min. at 37 °C (lane 3). Similarly, 10-12 ug of RNA from cultured rat cells were untreated (lane 2) or RNase A treated (lane 4). The blot was probed with labeled BAM5 DNA isolated from the plasmid pMBA14. The positions of the 28S (upper arrow) and 18S (lower arrow) rRNAs in the gel are indicated.

vast majority of 6.2 kilobase pair Bgl I fragments contain BAM5 sequences located near one end of the fragment, but nowhere else.

Transcription of BAM5 sequences.

In order to investigate the possibility that BAM5 sequences are transcribed, RNA was extracted from two cell lines; GR 3A, a mouse mammary tumor cell line, and M1.19, a rat hepatoma cell line. An isolation procedure was chosen (see Materials and Methods) which was designed to remove contaminating nuclear material and reduce the possible interference posed by the enormous quantities of BAM5 homologous sequences present in the nuclear DNA.

Figure 7 demonstrates that RNA homologous to BAM5 sequences was present in the mouse cells, but not in the rat cells, BAM5 homologous DNA, undoubtedly from ruptured nuclei, was also present in each sample, but did not obscure the result. Thus, in certain mouse cells (long term culture cells) BAM5 sequences are transcribed.

DISCUSSION

The BAM5 family of repetitive DNA sequences in the mouse consist of 507 base pair stretches of DNA, bounded by Bam HI restriction sites, which account for approximately 0.5-1% of the total mouse genome. BAM5 sequences are dispersed throughout the mouse genome and appear to be unique; that is to say, the sequences do not appear to have arisen by the duplication of a shorter sequence(s) (Fig. 3).

RNA homologous to BAM5 sequences is present in cultured mouse cells, but not in cultured rat cells (Fig. 7), although both cell types contain reiterated BAM5 DNA sequences (Fig. 4). However, a much more thorough study of BAM5 transcription will be necessary to determine the biological significance of this finding. Studies designed to assess the tissue specificity and developmental timing of BAM5 DNA transcription in mice and rats are currently in progress. In any case, the BAM5 family member that has been sequenced (Fig. 3) contains numerous stop codons in all three reading frames. Thus, it is unlikely that BAM5 sequences code for cellular proteins.

Based upon restriction mapping (Fig. 2,4,5) and sequence analysis (Fig. 3), the BAM5 sequence reported here is not homologous to the repetitive interspersed mouse DNAs described previously (24,25,26,27,28,29). The 6.2 kilobase pair Bgl I fragment, however, does appear to contain subfragments described by others (24,25,29). Recently, Meunier-Rotival et al. (30) reported studies with a cloned BAM5 family member (designated by them as a 0.54 kilobase pair fragment). Their results suggested that BAM5 sequences were scattered throughout much, but not all, of the mouse genome and were parts of larger (greater than 5.6 kilobase pairs) repetitive sequences. Thus, the results presented here and those of Meunier-Rotival et al. are in agreement.

Restriction mapping of mouse DNA revealed that BAM5 sequences were components of larger repetitive DNAs (Fig. 5). The Bgl I fragment (Fig. 5, lane 5) was estimated to have a size of about 6.2 kilobase pairs and restriction mapping demonstrated that the BAM5 sequence was located exclusively near one end of this large fragment (Fig. 6). This result clearly shows that BAM5 sequences within the larger fragment are not arranged in tandem scrambled clusters, an organizational pattern that has been proposed for much of the repetitive DNA in an organism (Syrian hamster) closely related to the mouse (10).

ACKNOWLEDGEMENTS

I thank Gisela Heidecker, John Puma, Dave Morris, Rich Gardner and R. D. Cardiff for suggestions and discussion.

I especially thank R. D. Cardiff for continual support and encouragement and Gisela Heidecker for instruction in the use of the M13mp9 cloning/sequencing system. In addition, thanks to Jeff Vieira and Jo Messing for generously making the M13mp9 phage available. Supported by Public Health Service grant 5 R01 CA 21454 from the National Cancer Institute.

REFERENCES

1. Lewin, B. (1980) Gene Expression 2-Eucaryotic Chromosomes, John Wiley and Sons, New York.
2. Brutlag, D.L. (1980) Ann. Rev. Genet. 14, 121-144.
3. Fry, K. and Grutlag, D.L. (1979) J. Mol. Biol. 135, 581-593.
4. Horz, W. and Altenburger, W. (1980) Nuc. Acids Res. 9, 683-696.
5. Jelinek, W.R., Toomey, T.P., Leinwand, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L. and Schmid, C.W. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 1398-1402.
6. Britten, R.J. and Davidson, E.H. (1969) Science 165, 349-357.
7. Davidson, E.H. and Britten, R.J. (1979) Science 204, 1052-1059.
8. Doolittle, W.F. and Sapienza, C. (1980) Nature 284, 601-603.
9. Orgel, L.E. and Crick, F.H.C. (1980) Nature 284, 604-607.
10. Moyzis, R.K., Bonnet, J., Li, D.W. and Ts'o, P.O.P. (1981) J. Mol. Biol. 153, 870-896.
11. Fanning, T.G., Puma, J.P. and Cardiff, R.D. (1980) J. Virol. 36, 109-114.
12. Fanning, T.G., Puma, J.P. and Cardiff, R.D. (1980) Nuc. Acids Res. 8, 5715-5723.
13. Pedersen, P.L., Greenawalt, J.W., Reynafarje, B., Hullihen, J., Decker, G.L., Soper, J.W. and Bustamente, E. (1978) Methods Cell Biol. 20, 411-481.
14. Southern, E.M. (1975) J. Mol. Biol. 38, 503-517.
15. Thomas, P. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 5201-5205.
16. Rigby, P.W., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
17. Lane, D. and Gardner, R.C. (1979) J. Bact. 139, 141-151.
18. Hanahan, D. and Meselson, M. (1980) Gene 10, 63-67.
19. Tabak, H.F. and Flavell, R.A. (1978) Nuc. Acids Res. 5, 2321-2332.
20. Vieira, J. and Messing, J., in preparation. M13mp9 is a derivative of M13mp7 described in - Messing, J., Crea, R. and Seeberg, P.H. (1981) Nuc. Acids Res. 9, 309-321.
21. Gronenborn, B. and Messing, J. (1979) Nature 272, 375-377.
22. Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
23. Heidecker, G., Messing, J. and Gronenborn, B. (1980) Gene 10, 69-73.
24. Cheng, S.M. and Shildkraut, C.L. (1980) Nuc. Acids Res. 8, 4075-4090.
25. Heller, R. and Arnheim, N. (1980) Nuc. Acids Res. 8,

5031-5042.
26. Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskov, A.P., Bayer, A.A. and Georgiev, G.P. (1980) Nuc. Acids Res. 8, 1201-1215.
27. Manuelidis, L. (1980) Nuc. Acids Res. 8, 3247-3258.
28. Miesfeld, R., Krystal, M. and Arnheim, N. (1981) Nuc. Acids Res. 9, 5931-5947.
29. Brown, S.D.M. and Dover, G. (1981) J. Mol. Biol. 150, 441-466.
30. Meunier-Rotival, M., Soriano, P., Cuny, G., Strauss, F. and Bernardi, G. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 355-359.