Volume 10 Number 23 1982 Nucleic Acids Research

A short interspersed repetitive element found near some mouse structural genes

Kira K.Lueders* and Bruce M.Paterson

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205, USA

Received 27 August 1982; Revised and Accepted 2 November 1982

ABSTRACT

We have isolated and characterized a family of interspersed repetitive elements which make up about 1% of the mouse genome. The elements represent a group of homologous but non-identical units about 400 bp in length. Individual members of the family show considerable divergence from one another. The spacial relationships between members of the family and a number of other identified mouse sequences including structural genes have been determined; these elements are found on the 5' as well as 3' sides of various genes at distances ranging from less than 1 to 7.5 kilobases (Kb). The sequences are present in the DNA of all species of Mus. Related sequences are present in the rat genome at a repetition frequency similar to that in the mouse genome. A partial sequence of one member of the family is presented.

INTRODUCTION

A number of moderately repetitive interspersed sequence families have been identified in the mouse genome (for reviews, see 1,2). With copy numbers in the range of 10^4-10^5, each of the families may represent 1% or more of the genomic DNA. Members of the families range from quite small, such as the 130 bp B1 family unit (3), and 475 bp R family unit (4), to ones which are quite long, such as the recently described 5.6 Kb repeat family (5). In many cases, the unit size of the family members has not yet been determined, and it appears that repetitive sequences of more than one type can be associated with one another in different combinations (6). There is no known function for most of these sequences, and very few have been sufficiently studied to determine their relationship, if any, to structural genes.

In the course of our studies with a family of repetitive dispersed retroviral genes (7,8), we observed a much more reiterated sequence element close to the 5' end of one of these genes. This sequence was used as probe to isolate other members of this highly reiterated family and to determine their properties, and spatial relationship to other genes.

0305-1048/82/1023-7715 $2.00/0

METHODS

Sources of DNA

High molecular weight DNA was prepared from mouse myeloma, rat hepatoma, and Syrian hamster kidney as previously described (9). Monkey BSC-1 cell DNA was from Maxine Singer, NCI. Cloned IAP genes were isolated from a library of mouse DNA in λ phage Charon 4A (generously provided by J. Seidman, NIH) (7). The recombinant used in these experiments, λMIA14, contains the entire IAP gene sequence. Recombinant clones containing various other mouse genes isolated from gene libraries in λ phage were obtained from a number of individuals as indicated in Table 1. The rat gene library (10) was provided by T. Sargent, NCI.

Sub-cloning of DNA Fragments in pBR322

DNAs were digested with restriction endonucleases (N.E. Biolabs) and the digests were electrophoresed in agarose gels as previously described (7). DNA fragments were recovered from the gels by electroelution. Fragments were ligated to pBR322 and the mixtures used to transfect *E. coli*. Ampicillin resistant transformants were screened by hybridization as previously described (7).

Filter Hybridization

For homologous DNA reactions, hybridization of filters to nick-translated or end-labeled probes was as described (7) except all three washes were in 0.1 x SSC (1 x SSC = 0.15 M NaCl, 0.015 M sodium citrate), 0.25% SDS for 15 min. at 55°C. Heterologous reactions were done at 50°C, with washes in 6 x SSC at 55°C.

RESULTS

Identification of an Interspersed Repeat Family in Mouse Genomic DNA

In the course of our studies on a family of retrovirus genes related to the intracisternal A-particle (IAP), we prepared fragments representing different parts of the viral genome for use as hybridization probes. Figure 1 shows a partial restriction map for the particular cloned IAP gene, λMIA14, that was used in these experiments. When most of the fragments from this gene were labeled and hybridized to mouse genomic DNA which had been cut with restriction endonucleases, the blot hybridization patterns were those expected from the known restriction maps of the IAP genes, which are present in the mouse genome at 1,000 to 2,000 copies (9). However, a 2 Kb *EcoRI* fragment, containing the 5' viral long terminal repeat (LTR) and about 1.5 Kb of flanking sequence, gave a very heavy heterogeneous pattern of hybridization, which

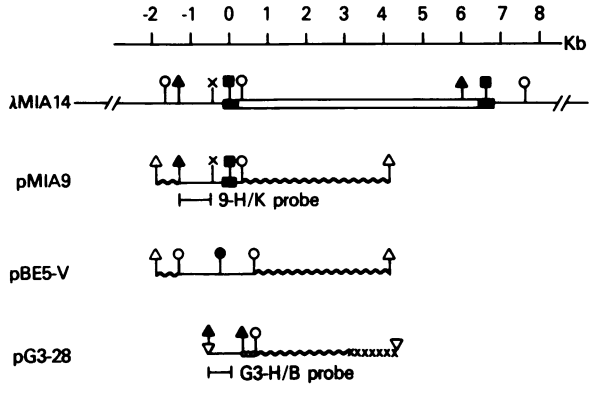

Figure 1. Physical maps of clones used for preparation of probes and for heteroduplex analysis. A partial restriction map of recombinant λMIA14 containing the 7 Kb IAP gene is shown at the top (double line); the ends of the IAP gene are delineated by the long terminal repeats (LTRs) ■. Mouse flanking sequences are shown as a single line. An EcoRI/HindIII fragment containing the IAP gene 5' LTR and mouse flanking sequences was subcloned in pBR322. The resulting recombinant, pMIA9, is shown; pBR322 sequences are shown as ~~~. pBE5-V contains a 2 Kb EcoRI fragment from λBE5, a genomic clone with a mouse β-globin gene in it (11). pG3-28 is a clone containing a 0.85 Kb mouse DNA fragment in the plasmid pK04, a derivative of pK01 (12). xxxx, E. coli galK gene sequence; 000, λ o gene sequence. The 850 bp HindIII/KpnI fragment from pMIA9 (designated 9-H/K probe) and the 650 bp HindIII/BglI fragment from pG3-28 (designated G3-H/B) were recovered from agarose gels by electroelution. Restriction sites are as follows: ♀, EcoRI; ♠, HindIII; ♥, PstI; ✱, KpnI; ♙, SphI; ♥, XbaI; ▽, BamHI.

could not be accounted for in terms of the IAP genes (data not shown). This suggested that the flanking region 5' to the LTR contained another more highly reiterated sequence element which was interspersed in the mouse genome.

This more abundant component was separated from the LTR sequences by cutting the plasmid pMIA9 (Fig. 1) with HindIII and KpnI. The resulting 850 bp HindIII/KpnI fragment (9-H/K probe) containing the highly repeated sequence, was hybridized to recombinants in a mouse genomic library consisting of 15 Kb inserts in λ phage. Twenty-seven percent of the clones reacted with the 9-H/K probe (not shown), indicating that the homologous sequences are interspersed in the mouse genome. The intensity of the reactions was relatively uniform. Assuming only one repeat per positive insert, we calculate there are ~ 10^5 copies per genome. With an average unit size of ~ 400 bp (see below), these sequences represent 1% of mouse DNA.

We then tested 114 recombinants selected in other laboratories because they contained identifiable mouse genes (unrelated to IAP genes). The major-

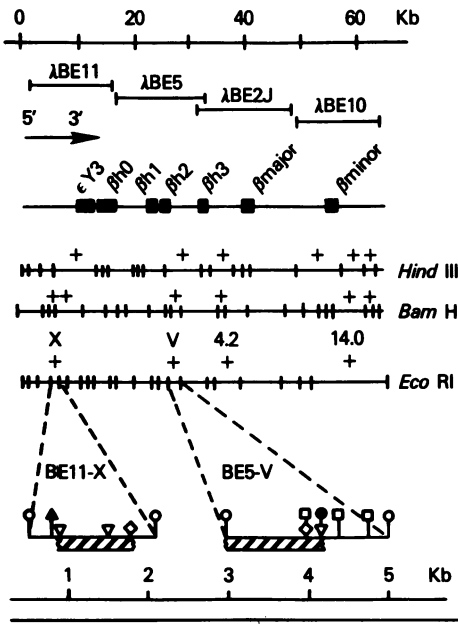

Figure 2. Localization of the repeated sequence element in the β-globin gene cluster. The β-globin gene containing phage were provided by Norman Hansen and Philip Leder: A. The location and designation (11) of these recombinants is shown at the top. The nomenclature for the genes, the restriction maps, and designation for the EcoRI fragments is according to Jahn et al. (13). Reactive restriction fragments were mapped by blot hybridization to recombinant phage DNAs and are indicated with a + on the maps. Two of the reactive EcoRI fragments, X and V, were subcloned into pBR322, and the resulting recombinants were designated pBE11-X andpBE5-V. The repetitive sequences in these subclones were mapped to within an approximately 1 Kb region indicated by ▨. Restriction sites are as follows: ⚲, EcoRI; ↑, HindIII; ▽, BamHI; ⚲, XbaI; ⚲, BglII; ◇, PvuII.

ity of these were also from genomic libraries that consisted of 15-18 Kb fragments of mouse DNA inserted into λ phage. Thirteen of these recombinants reacted on spot tests with the 9-H/K probe (see Table 1). Four of these were clones containing various β-globin genes, spanning about 60 Kb in the mouse genome (11). Since the restriction maps for these clones were known in some detail (13), they were used to isolate other examples of the reiterated sequences. Figure 2 shows the positions of the reactive fragments on these clones (λ BE11, λBE5, λBE2J, and λBE10) with reference to the restriction maps published by Jahn et al. (13).

EcoRI fragments V and X (2.0 Kb and 1.6 Kb, respectively) were subcloned into pBR322; the resultant recombinants were designated pBE5-V and pBE11-X. Hybridization of the 9-H/K probe to restriction fragments from pBE5-V and pBE11-X was used to localize the reactive sequences to within a 1 Kb region in each recombinant (Fig. 2).

Unit Size of the Reiterated Sequence Element

The size of the sequence element was determined by examination of heteroduplexes formed between different cloned isolates. Hybridization of λMIA14 and λBE5 revealed a single small homology region of 390±60 bps within the 16-18 Kb inserts (Fig. 3A). This same homology was also studied in heteroduplexes between the subclones pMIA9 and pBE5-V (see Fig. 1). Both plasmids

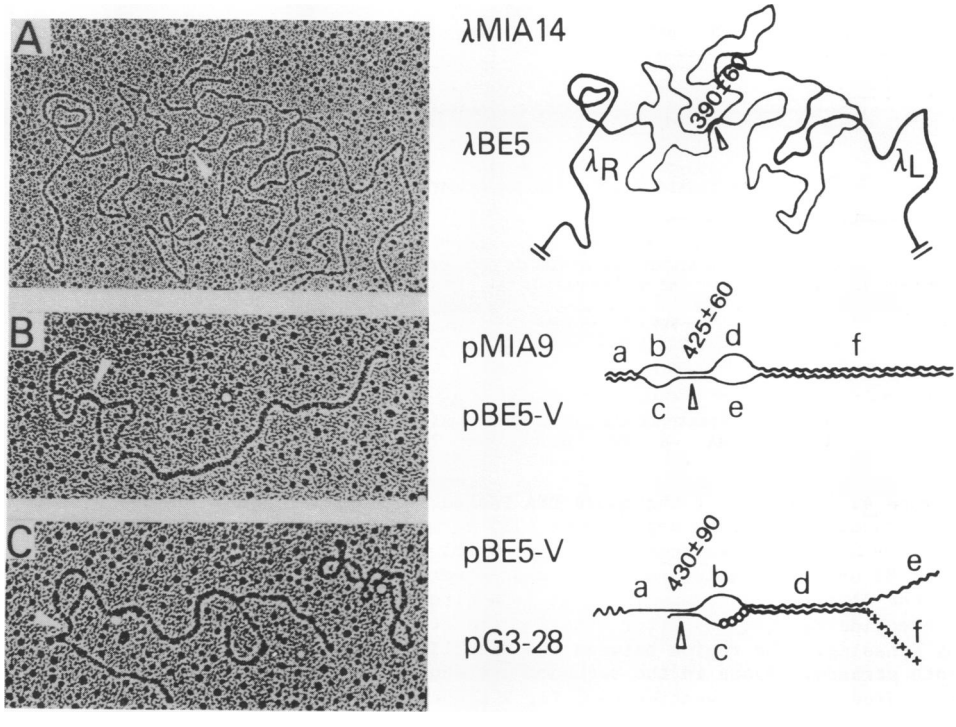

Figure 3. Determination of the size of the sequence element by heteroduplex analysis. The phage or DNAs were treated with EDTA and alkali, and the heteroduplexes prepared and mounted by the formamide technique of Davis et al. (14) as previously described (8). Regions of homology are indicated by white arrows. A. Heteroduplex formed between λMIA14 and λ BE5. SV40 DNA (5.24 Kb) and Øx 174 DNA (5.386 Kb) were used as double stranded and single stranded DNA markers, respectively. B. Heteroduplex formed between pMIA9 and pBE5-V. The molecules were linearized with SphI, which cuts pBR322 at map position 561. (See Fig. 1.) The known sizes of the pBR322 double stranded regions (a=532 bp, f=3802 bp) were used as size standards. The sizes of the other regions in bps were: b, 320±60; c, 825±100; d, 1150±90; e, 960±83. C. Heteroduplex formed between pBE5-V and pG3-28. pBE5-V DNA was linearized with SphI as above. pG3-28 DNA was linearized with BamHI, which cuts the plasmid pK04 16 bp from the HindIII site at one end of the mouse DNA insert (see Fig. 1). The homology region lies close to this end of the insert, leaving a short single strand region of pG3-28 opposite the long single strand region of pBE5-V. Part of the pBR322 sequences in pBE5-V remain single stranded because these sequences have been replaced with the E. coli galK gene in pG3-28 (shown as xxx here and in Fig. 1). Markers were those used in A. The sizes of the regions measured in bps were as follows (numbers in parentheses give expected lengths for pBR322 and pK04 sequences): a, 1400±200; b, 1080±70; c, 660±130; d, 2430±100 (2300); e, 1520±190 (1500); f, 1370±110 (1300).

```
          10          20          30          40          50          60
//GATGATGATT ACAGT AATC TAT GGATGG AACACAGGGC CCCCAGTGGA GGAGCTAGAG AACGTACCCA
  TG    C C      C        T                          A                 A

  70          80          90          100         110         120         130
A GAGCTGAA GGGGCCTGCA ACCCTATAGG TGGAACAACA AAATAAACTA ATCAGTAACC CCAGAGCTC
 G G           T                              T G       C     C           T

  140         150         160         170         180         190         200
 GTATCTCTT GCTGCATATG TAGCAGAAGA TGGACTAGTC AGCCATCATT GGAAAGAGAG GCCCCAAGGT
G  G     A              T           C       G           G              T   C

  210         220         230         240         250         260         270
CTAGCAAACT TTATATGTCT CAGTAC GGG GA CGCC GT TGCC AGAAG TGGGAGTGGG TGGGTAAGGG
     T           AC C       A     A    T G G  TT                      G  A

     280         290         300         310         320         330         340
AGCAGGAGAC AGGGAGGAAA TAGGGAACTT TCAGGATAGC ATTTGAAATG TAAATGAGGA AAAATATAAA
      GC G G      GT                                                 G    GT

     350         360         370         380
AAAAAATTTT TTTAAATTAA AAAAAAACAG ATACACTAAA
 T    AAGA AAA   GA   TGT   TA A TA AGAA  T
```

Figure 4. Sequence of the mouse DNA insert in pG3-28. Sequencing was done by the procedure of Maxam and Gilbert (15) as modified by Smith and Colvo (16). The A + G reaction was carried out with 88% formic acid. Products were run on 6 or 8% urea-acrylamide gels. Sequences were obtained from the HindIII site in the 3' flank (not shown) and in both directions from the DdeI site at nucleotide 224 (arrow) after labeling the fragments at the 5' ends with ^{32}P by kinasing. The region between the HindIII and DdeI sites was sequenced on both strands. Bases in the sequence are shown in the 5' → 3' direction, according to the convention used for R sequences by Gebhard et al (4). The sequence of pG3-28 is compared with that of R1 published by Gebhard et al.; bases different in R1 are shown on the lower line. R1 sequence extends another 123 nucleotides in the 5' direction (not shown) so that position 1 in the pG3-28 sequence corresponds to 124 in R1. The boxed nucleotides in positions 294-303 indicate a region of homology with the SV40 72 bp repeat in positions 163-172. The numbering for SV40 is that of Fiers et al. (17); the SV40 control region included positions 5207-272. Two other regions of homology with the SV40 72 bp repeat were at the following positions in pG3-28: 58-74 (76% homology with positions 172-158 of SV40) and 129-136 (81% homology with positions 169-157 of SV40).

were linearized with SphI which cuts pBR322 in such a way as to provide double-stranded regions of 532 and 3802 bp as convenient internal size markers. The heteroduplex between these molecules is shown in Fig. 3B. In these experiments, the homology region was measured as 425±60 bp.

Another example of the reiterated element was detected in a recombinant consisting of a 850 bp HindIII fragment of mouse genomic DNA cloned into pKO4 (12), which was provided by S. Segal, NCI. A heteroduplex between the BamHI linearized DNA from this clone, designated pG3-28 (see Fig. 1), and SphI linearized pBE5V had a homology region of 430±90 bp (Fig. 3C). Similarly, a heteroduplex between pG3-28 and pMIA9 had a size of 360±60 pb (not shown).

The evidence suggests that the repetitive sequence element has a size of approximately 400±50 bp.

Sequence Analysis of a Repeat Element

While this paper was in preparation, a paper by Gebhard et al. (4) appeared which described a family of repeats, the R sequences, in the mouse genome. Similarities between the properties of R sequences and the repeats we identified strongly suggested that they belong to the same family. Comparison of a partial sequence of pG3-28 with the R sequences published by Gebhard et al. confirmed that this was indeed the case. Figure 4 shows the partial sequence from the pG3-28 and indicates the bases which differ in the R1 sequence. Most positions which differ between the two sequences consist of one or two base changes or additions/deletions. Over the 358 bp compared, the two sequences share 300 bases or 84% homology. Other common features that we noted were the very AT rich flanking region, and the presence of many direct repeats throughout the sequence.

Relationships Between R Family Members and Some Identified Mouse Sequences

Six members of the R family have been identified by Gebhard et al. (4) in recombinant clones containing K variable region genes. Table 1 summarizes our data obtained by restriction enzyme analysis and/or heteroduplex analysis on the relationships between R sequences and other mouse sequences.

It is clear that R sequences occur on either the 5' or 3' sides of a number of genes, at distances ranging from less than 1 Kb up to 7.5 Kb. We could detect no obvious constant spacial relationship between R elements and the other sequences.

We found R sequences in a number of genomic clones containing IAP genes isolated in our laboratory. Three genomic clones selected for other genes contained both R sequences and IAP sequences (the β-globin $_{min}$ gene, K light chain variable region gene, and λ light chain variable region gene). Clones carrying IAP genes or partial AKV genes were the only clones with retroviral sequences that contained R sequences. Both the IAP and AKV sequences are present as endogenous virogenes in the mouse genome. Other genomic clones containing viral genes that did not contain R sequences included MMTV, MoMSV, VL30, and Abelson leukemia virus.

We also tested R sequences for homologies to other short mouse repetitive sequence families by DNA dot hybridization as previously described (23). There was no homology to the repetitive sequences in the mouse ribosomal DNA non-transcribed spacer (24) or to the evolutionarily conserved TG repeats (25)

TABLE 1

R Sequence	Gene Designation	Clone	Distance Between Gene and R Sequence (Kb)	Side of Gene on Which R Sequence Occurs (Orientation)[i]
1	β-globin εY3	λβE11[a]	4	5'
2	β-globin βh2	λβE5[a]	<1	3' (S)
2	β-globin βh3	λβE5[a]	4	5' (S)
3	β-globin βh3	λβE5[a]	≤3	3' (S)
3	β-globin major	λβE2J[a]	≤5	5' (O)
4	β-globin minor	λβE10[a]	1-6	3' (O)
5	α-globin related	λα20[b]	ND	ND
6	K light chain variable	λMIA80[c]	≤5	3'
7	Partial AKV	λAK70[d]	3-4	3' (O)
8	Partial AKV	λAK68[d]	ND	ND
9	Partial AKV	λAK46[d]	ND	ND
10	H2	λ9[e]	ND	ND
11	H2	λ11[e]	ND	ND
12	α-amylase Amy-1[a]	IID[f]	≤1	3'
13	IAP	λMIA14[g]	0.5	5' (O)
14	IAP	λMIA21[g]	≤3	3'
15	IAP	λMIA58[g]	1-2	5'
16	IAP	λMIA63[g]	≤3	3'
17	λ light chain variable	Vλ1[h]	7.5	5' (S)
18	Partial IAP	Vλ1[h]	3	5' (O)

ND = not determined; clones positive on dot test. Sources of clones were as follows: a, Norman Hansen and Philip Leder (11); b, Aya Leder; c, Uli Siebenlist and Philip Leder; d, Arifa Khan and Malcolm Martin; e, Philippe Kourilsky; f, Ueli Schibler (18); g, clones from our Laboratory which have been previously described (7,8); h, Ursula Storb (19); i, orientation is indicated as S = same or O = opposite with respect to the structural gene; the 5' → 3' orientation of R sequence was designated according to the convention of Gebhard et al. (4). Additional clones which contained integrated retroviral sequences along with mouse flanking cellular sequences were obtained from the following: Clone GR40 (20) with mouse mammary tumor virus (MMTV) sequences from Nancy Hynes; c-mos sequences, from which the v-mos of Moloney murine sarcoma virus (MoMSV) is derived, from William McClements (21); Abelson murine leukemia virus sequences from Steven Goff; and the endogenous virus-like (VL-30) sequences from Eli Keshet (22).

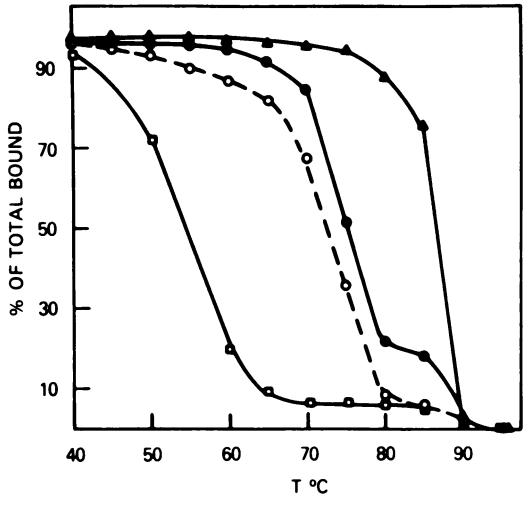

Figure 5. Thermal stability of R family members. The probe was the ^{32}P-end labeled G3-H/B fragment. For chromatography on hydroxylapatite, the samples were diluted and loaded on the column in 0.12 M sodium phosphate (1:1) at 40°C. Three 2 ml fractions were eluted at each temperature. The probe was hybridized in 50 µl reactions for 10 min with (▲) 4 µg (C_ot=0.8) of the 850 bp HindIII insert from pG3-28 or with 100 µg (C_ot=20) of genomic DNAs at 65°C for (●) M. musculus, and at 50°C for (○) M. cervicolor and (□) rat. Genomic DNAs were sonicated to an average size of 450 bp as previously described (9).

(N. Arnheim, personal communication). Gebhard et al. (4) have shown that the R family is not related to the mouse B1 family (3).

R Sequences in Mus musculus

It was apparent from hybridization of blots of cloned DNAs containing different members of the R family with 9-H/K probe that the reaction intensities for similar amounts of DNA varied a great deal. In addition, we could not find common patterns of restriction sites for different cloned R sequences. This suggested that although these sequences are related, individual members could be quite heterogeneous. To examine this question, we determined the thermal stability of hybrids formed between members of the family.

A 650 bp end-labeled HindIII/BgI fragment from pG3-28 (G3-H/B probe, see Fig. 1), which contained an entire R sequence as defined by heteroduplex analysis, was used as probe. The homoduplex formed between this probe and pG3-28 DNA had a sharp melting curve with a T_m of 87°C (Fig. 5).

The melting curve for hybrids formed between the probe and genomic DNA from Mus musculus was relatively homogeneous and nearly as sharp, but the T_m was 12.5° lower than that of the homologous hybrids, indicating a divergence of

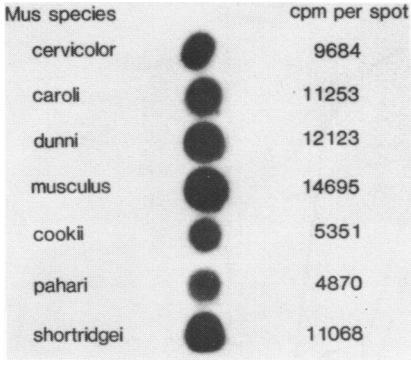

Figure 6. Hybridization of DNAs from Mus species to R sequences. Aliquots (5 μg) of DNAs from livers of the indicated Mus species were spotted onto a nitrocellulose filter (provided by R. Callahan, NCI) and the filter was hybridized with the 9-H/K probe for 40 hrs. at 50°C. The filter was washed in 0.1 x SSC at 50°C. The amount of label bound to each DNA was quantitated by cutting out the spots and counting them in a liquid scintillation counter.

about 8% among the family members. It should be noted that all the divergent sequences detected in the melting curve would be capable of forming heteroduplexes within the range of stringency used for electron microscopy.

R Sequences in Other Mus Species

Genomic DNAs from seven distantly related species of Mus all contained R sequences as shown in Figure 6. DNAs (5 μg) were bound to a nitrocellulose filter and hybridized to nick-translated 9-H/K probe under conditions to detect divergent sequences. The amount of label bound was determined by cutting and counting the spots. There was only a three-fold difference in the amount of probe bound by DNAs of Mus musculus and the most distantly related species (Mus pahari).

The melting curve of hybrids formed between the G3-H/B probe and genomic DNA from the Asian mouse species Mus cervicolor is shown in Figure 5. Probe self-annealing (15%) noted in the reaction with M. musculus DNA was not seen because the hybridization was done at a temperature below the optimum for homologous reaction. The T_m of the hybrids was only 2.5° lower than that for hybrids formed between the probe and M. musculus DNA. Since the ΔT_m for single-copy DNA of these mice is 5-6° (26), these experiments indicate that the R sequences are relatively conserved between the M. cervicolor and M. musculus genomes even though they are internally divergent within each species. These results suggest that R sequences were amplified before the Mus species diverged.

R Sequences in the DNAs of Other Species

Equivalent amounts of HindIII digested DNAs from mouse, rat, Syrian hamster and monkey were fractionated by electrophoresis and hybridized to the end-labeled G3-H/B fragment under conditions for detection of highly divergent sequences. Mouse and rat DNAs gave equally strong reactions over a wide size

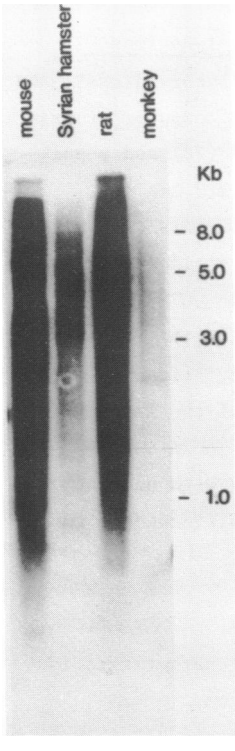

Figure 7. R-related sequences in the DNAs of other species. Aliquots of the indicated DNAs (5µg) were digested with HindIII, separated by electrophoresis in an agarose gel, transferred to a nitrocellulose filter, and hybridized at 50° for 18 hrs with the end labeled G3-H/B probe. The filter was washed in 6 x SSC at 50°C.

range (Fig. 7). Few discrete reactive fragments reflecting classes of sequences with similar HindIII sites could be distinguished. When the 9-H/K fragment from pMIA9 (the same probe used to scan the mouse genomic library) was hybridized to recombinants in a rat genomic library consisting of 10-20 Kb inserts in λ phage (10), the same percentage of clones reacted (27%) as had reacted with recombinants from the mouse gene library. Thus, the R sequences seem to be dispersed in the rat genome at a reiteration frequency similar to that seen in the mouse.

Figure 5 shows the melting curve for hybrids formed by rat genomic DNA and the G3-H/B probe. The shape of the curve was similar to that for the two mouse genomic DNAs but the T_m was 17.5° lower, consistent with the phylogenetic distance between these species (26).

The reaction with Syrian hamster DNA was much weaker than those with mouse and rat DNAs (Fig. 7), and discrete components over a wide size range were detectable. Thus, some of the related elements in this species may be in higher order repeat units not detected in mouse and rat. Although the reaction with monkey DNA was weak as expected for a very distantly related

species, several discrete fragments could be detected.

Identification of Sequences Related to Transcriptional Control Regions

The sequence of pG3-28 was searched for eucaryotic regulatory signals and compared with sequences known to have such signals using the computer program of Queen and Korn (27). A DNA polymerase II promoter TATAAA (28) and a polyadenylation signal AATAAA (29) were identified at positions 220 and 109, respectively (see Fig. 4). No DNA polymerase III promoter signals (30) were found. In comparing other sequences, the only matches considered were those with greater than 75% homology and a probability of less than 10^{-4} that they occurred randomly. When the pG3-28 sequence was compared with that of SV40 control region, numerous such regions of homology were found, primarily with the origin of replication or the 72 bp repeat (31). Three of the latter included the region of the SV40 72 bp repeat which has been shown to have "enhancer" activity (32). The best of these, a 90% homology with nucleotides 163-172, is indicated as a boxed region in Fig. 4. The significance of the presence of these signal sequences remains to be determined.

DISCUSSION

In this paper we have described the identification and characterization of a family of repetitive sequence elements in the mouse genome.

The R family was defined by Gebhard et al. (4) on the basis of six members, all of which were found in recombinants containing K light chain variable region genes (33). We have shown that members of the family are found near other genes as well and in several cases determined their location with respect to these genes (Table 1).

The approximately 60 Kb region of the mouse β-globin gene cluster has been studied in detail with respect to both coding sequences (11,13) and repeated elements (13,34). This locus in the mouse is similar to that in other mammals in having multiple repetitive elements interspersed among the β-globin genes (see 35). Although it is clear from previous work that the mouse locus contains at least five classes of repeated sequences (6), their nature and the relationships between them are not known, except for a B1 element which has been mapped to a position 3' to the β- major gene (34) and another element which has been identified as a satellite-like sequence (6). The 9-H/K probe gave positive reactions with four regions of the β-gene locus. In two of these instances we have localized the R sequence within an approximately 1 Kb region (see Fig. 2), thus determining the size and location of a major repetitive element in this locus.

In the case of the α-amylase gene (18), we assigned the R sequence to the 3' flank because in addition to the α-amylase gene itself, the clone which gave a positive reaction contains only a 1 Kb 3' flanking sequence. However since the intervening sequences in the α-amylase gene have been shown to contain repeated DNA (18), it is possible that one of these could represent R sequences.

Sequences homologous to, but quite different from, the mouse R family were also detected in the DNAs of rat, Syrian hamster, and monkey. Gebhard et al. (4) noted that the strength of R sequence reaction with rat DNA was much weaker than that with mouse DNA under their conditions. Using more relaxed hybridization conditions, we did not observe such a difference (see Fig. 6). Our hybridization of R sequences to rat library plaques and the melting curve (Fig. 4) show that the rat genome contains R related sequences whose interspersed nature and reiteration frequency are very similar to those of R sequences in the mouse genome, but that they are quite divergent from the mouse sequences.

In spite of the extensive literature on the properties and location of a number of repetitive sequence families, their functions are still largely unknown, although various suggestions for possible roles have been put forth (see 1 for review). Clone pG3-28 was originally selected on the basis of its ability to provide promoter activity for a bacterial gene (Segal, in preparation). Examination of the partial sequence indicates that there are at least 8 sets of sequences which correspond to the consensus sequences for this activity (36). This fragment also contains a TATAAA box (28) and poly(A) signal (29). Experiments are under way (Segal) to determine whether these signals can be recognized in eukaryotic cells and promote the transcription of a promoterless gene.

Another type of function which could be envisaged for R sequences is that of "enhancer". This class of regulatory sequence can stimulate transcription of genes from their own promoters with a great deal of latitude in position and orientation. Although enhancers were first discovered as 72 bp repeated sequences in SV40 (31), homologous repetitive sequences have now been found in genomic DNA (37). There is evidence that sequences essential for enhancing activity in the SV40 72 bp repeat are in a 17 bp EcoRII fragment (nucleotides 162-178) (32). We found several homologies to the SV40 72 bp repeat, including a 90% homology with nucleotides 163-172 (see Fig. 4). It is not known whether the presence of these sequences is invariably associated

with enhancer activity and whether R sequences have such actvity remains to be determined.

Acknowledgments

We are grateful to the many individuals cited in this paper who generously provided cloned genes. We thank Dr. Jeffrey Saffer for his invaluable instruction in carrying out the computer analysis and Dr. Maxine Singer for providing us with her modification of the Queen and Korn program. We also thank Dr. Michael Botchan for making his results available before publication, and Dr. Norman Arnheim for testing our clones. KKL is especially grateful to Dr. Edward Kuff for many invaluable discussions and suggestions.

*To whom reprint requests should be sent

References

1. Singer, M.F. (1982) Int. Rev. Cytol. 76, 67-112.
2. Singer, M.F. (1982) Cell 28, 433-434.
3. Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskov, A.P., Bayer, A.A., and Georgiev, G.P. (1980) Nucleic Acids Res. 8, 1201-1215.
4. Gebhard, W., Meitinger, T., Höchtl, J., and Zachau, H.G. (1982) J. Mol. Biol. 157, 453-471.
5. Meunier-Rotival, M., Soriano, P., Cuny, G., Strauss, F., and Bernardi, G. (1982) Proc. Natl. Acad. Sci. USA 79, 355-359.
6. Haigwood, N.L., Jahn, C.L., Hutchison, C.A., III, and Edgell, M.H. (1981) Nucleic Acids. Res. 9, 1133-1150.
7. Lueders, K.K. and Kuff, E.L., (1980) Proc. Natl. Acad. Sci. USA 77, 3571-3575.
8. Kuff, E.L., Smith, L.A., and Lueders, K.K. (1981) Mol. Cell. Biol. 1, 216-227.
9. Lueders, K.K. and Kuff, E.L. (1977) Cell 12, 963-972.
10. Sargent, T.D., Wu, J.-R., Sala-Trepat, J.M., Wallace, R.B., Reyes, A.A., and Bonner, J. (1979) Proc. Natl. Acad. Sci. USA 76, 3256-3260.
11. Leder, P., Hansen J.N., Konkel, D., Leder, A., Nishioka, Y., Talkington, C. (1980) Science 209, 1336-1342.
12. McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, C. and Rosenberg, M. (1981) Gene Amplification and Analysis 2, 383-415.
13. Jahn, C.L., Hutchison, C.A., III, Phillips, S.J., Weaver, S., Haigwood, N.L., Voliva, C.F., and Edgell, M.H. (1980) Cell 21, 159-168.
14. Davis, R.W., Simon, M., and Davidson, N. (1971) Methods Enzymol. 21, 413-428.
15. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
16. Smith, D.R. and Colvo, J.M. (1980) Nucleic Acids Res. 8, 2255-2274.
17. Fiers, W.R. et al. (1978) Nature 273, 113-120.
18. Schibler, U., Pittet, A.C., Young, R.A., Hagenbüchle, M.T., Gellman, S., and Wellauer, P.K. (1982) J. Mol. Biol. 155, 247-266.
19. Selsing, E., Miller, J., Wilson, R., and Storb, U. (1982) Proc. Natl. Acad. Sci. USA 79, 4681-4685.
20. Hynes, N.E., Kennedy, N., Rahmsdorf, U., and Groner, B. (1981) Proc. Natl. Acad. Sci. USA 78, 2038-2042.

21. Oskarsson, M., McClements, W.L., Blair, D.G., and Maizel, J.V. (1980) Science 297, 1222-1223.
22. Keshet, E., Shaul, Y., Kaminchik, J., and Aviv, H. (1980) Cell 20, 431-439.
23. Lueders, K.K. and Kuff, E.L. (1981) Nucleic Acids Res. 9, 5917-5930.
24. Arnheim, N., Seperack, P., Banerji, J., Lang, R.B., Miesfeld, R., and Marcu, K.B. (1980) Cell 22, 179-185.
25. Miesfeld, R., Krystal, M., and Arnheim, N. (1981) Nucleic Acids Res. 9, 5931-5947.
26. Rice, N.R. and Straus, N.A. (1973) Proc. Natl. Acad. Sci. USA 70, 3546-3550.
27. Queen, C. and Korn, L.J. (1980) Methods Enzymol. 65, 595-609.
28. Gannon, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Carni, B., and Chambon, P. (1979) Nature 278, 428-434.
29. Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251-260.
30. Shenk, T. (1981) Curr. Top. Microbiol. Immunol. 93, 25-40.
31. Gruss, P., Dhar, R., Khoury, G. (1981) Proc. Natl. Acad. Sci. USA 78, 943-947.
32. Banerji, J., Rusconi, S., and Schaffner, W. (1981) Cell 27, 299-308.
33. Steinmetz, M., Höchtl, J., Schnell, H., Gebhard, W., and Zachau, H. (1980) Nucleic Acids Res. 8, 1721-1729.
34. Coggins, L.W., Vass, J.K., Stinson, A.M., Lanyon, W.G., and Paul, J. (1982) Gene 17, 113-116.
35. Fritsch, E.F., Shen, C.K.J., Lawn, R.M., and Maniatis, T. (1981) Cold Spring Harbor Symp. Quant. Biol. 45, 761-775.
36. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
37. Conrad, S.E. and Botchan, M.R. (1982) Mol. Cell. Biol. 2, 949-965.