Sequence organization and genomic distribution of the major family of interspersed repeats of mouse DNA
(interspersed repeated sequences/genome organization/recombinant DNA)

Michele Meunier-Rotival, Philippe Soriano, Gérard Cuny, François Strauss, and Giorgio Bernardi

Laboratoire de Génétique Moléculaire, Institut de Recherche en Biologie Moléculaire, 2 Place Jussieu, 75005 Paris, France

Communicated by Gary Felsenfeld, September 30, 1991

ABSTRACT We have investigated the organization and the distribution of a family of interspersed DNA repeats in the mouse genome. The repeats are at least 5600 base pairs (bp) in size and contain two contiguous BamHI endonuclease fragments, 4000 and 540 bp in size, the larger of which includes a 1350-bp EcoRI fragment studied by previous authors. The repeats are polymorphic in their restriction maps, and represent the major family of interspersed repeats in the mouse genome. The repeats are present almost exclusively in the two light major components of mouse DNA, and the base composition of their large BamHI fragments matches that of those components. The genomic distribution of the repeats is different from that of structural genes, which are present not only in the two light components but also in the two heavy components of mouse DNA. This distribution indicates that the repeats are not involved, at least in any simple way, in the regulation of gene expression.

Degradation of mouse main-band DNA with restriction enzymes EcoRI, HindIII, HindII + III and Hae III yields discrete fragments that appear in gel electrophoresis as distinct bands over a continuous background smear (1, 2). In particular, EcoRI yields 1.3-kilobase (kb) fragments, representing 1.6% of main-band DNA; the degree of repetition of these fragments (30,000 copies per haploid genome) puts them in the intermediate repetitive kinetic class (1). Further studies on the 1.3-kb EcoRI fragments (3–6) have confirmed their interspersion in the main-band DNA and their lack of correlation with satellite DNA and have shown that these fragments are heterogeneous in some of their internal restriction sites and are included in larger repeats 3 kb in size.

In the present work, we have studied the sequence organization and the genome distribution of the interspersed repeats of mouse DNA. The main conclusion reached is that the major family of interspersed repeats of the mouse genome consists of sequences at least 5600 bp in size, polymorphic in their restriction maps, and present almost exclusively in the two major light components of mouse DNA. Such genomic distribution is different from that of structural genes, which are present on all four major components of mouse DNA (7–14); this indicates that the major family of interspersed repeats is not involved, at least in any simple way, in the regulation of gene expression.

MATERIALS AND METHODS

BALB/c mouse DNA was prepared from thymus or liver as described (11, 15). Its major components were isolated as reported (10, 11); the average molecular weights of the two preparations used in the experiments of Figs. 1–3 and 4 were 20 kb and 100 kb, respectively.

Restriction endonucleases were either purchased from commercial suppliers or prepared by L. Mallet and A. Meier of our Institute. Vertical gels of agarose (type II, Sigma) or polycrylamide (Bio-Rad) were used for electrophoresis.

Extraction of the large [4000-base pair (bp)] BamHI fragment and of the (1350-bp) EcoRI fragment from agarose were performed by dissolving the gel in sodium perchlorate at 65°C and by chromatographing the mixture at 65°C on hydroxyapatite (16). Extraction of the small (540-bp) BamHI fragment from polycrylamide was done by diffusion. EcoRI and BamHI fragments were inserted, using phage T4 ligase (Bethesda Research Laboratories) at the corresponding sites of plasmids pSF2124 and pBR322, respectively. These plasmids were used to transform competent Escherichia coli HVC 45 and HB101 cells, respectively. Recombinant clones were identified by the colicin test (17) in the first case or by plating the bacterial cells on L-agar supplemented with tetracycline at 10 μg/ml or ampicillin at 40 μg/ml in the second one.

DNA fragments were transferred from agarose gels to nitrocellulose filters (BA 85, Schleicher & Schuell) by blotting. Radioactive labeling of plasmid DNA was done by nick-translation. Prehybridization and hybridization were done as described (18). Low stringency washing conditions (150 mM NaCl/15 mM sodium citrate, 65°C) were used with filters loaded with genomic DNA; more stringent conditions (15 mM NaCl/1.5 mM sodium citrate, 65°C) were used with filters loaded with recombinant plasmid DNA. Autoradiography was done at −80°C on preflashed Kodak X-Omat or Fuji Rx films, using Du Pont Cronex Lightning Plus intensifying screens.

RESULTS

Repeated Fragments in Digests of Mouse DNA and of Its Major Components. When digested with EcoRI, BamHI, Kpn I, HindII, and Hae III and subjected to gel electrophoresis, mouse DNA exhibits several bands over a smear (Figs. 1 A and C, 2 A and A', 3A, and 4A). The lengths, in bp, of the fragments corresponding to the bands are the following: EcoRI, 1350; BamHI, 4000, 1350, 970, 540, and 540; Kpn I, 2500 and 820; HincII, 2500; Hae III, 3200, 1960, 1860, 1180, 1120, 1000, 950, 850, 350, and 180.

When restriction digests of isolated major components of mouse DNA (11) were examined, it was observed (Figs. 1 A and C and 4A) that the bands mentioned above were exclusively (or almost so in the case of the 1350-bp EcoRI and 540-bp BamHI bands) present in the two light components (1.699 and 1.701 g/cm³); the buoyant density of the 4000-bp BamHI fragment, as isolated from plasmid pMB1.1, was found to be 1.698 g/cm³. In contrast, the 1350-bp BamHI band was present in all four major components.

Abbreviations: kb, kilobase(s); bp, base pair(s).
Restriction Mapping of Cloned Repeated Fragments. Fig. 5 shows the restriction maps of the repeated fragments carried by recombinant plasmids pMRE1, pMRB1.1–3, and pMRB5, namely one 1350-bp EcoRI fragment, three 4000-bp BamHI fragments, and one 540-bp BamHI fragment. The maps were established on the basis of single, double, and triple digests and cross-hybridization experiments which showed: (i) that the 4000-bp BamHI fragment of pMRB1.1 is homologous in sequence with the 4000-bp fragments of pMRB1.2 and pMRB1.3, (ii) that the 1350-bp EcoRI fragment of pMRE1 is homologous in sequence with the 1350-bp EcoRI fragment contained in the inserts of pMRB1.1, pMRB1.2, and pMRB1.3, (iii) that the 540-bp BamHI fragment does not exhibit any sequence homology with the other cloned fragments. It should be noted that the three cloned 4000-bp BamHI fragments exhibit very similar, yet not identical, maps and that the cloned 1350-bp EcoRI fragment has the same size and a map similar to, yet not identical with, the map of the EcoRI fragments contained in the 4000-bp BamHI fragments.

Restriction Mapping of Genomic Repeated Fragments. Cloned repeated EcoRI and BamHI fragments were found to hybridize with genomic fragments of identical size (Table 1 and Fig. 6), under conditions in which unique sequences would show no hybridization. In addition to such a band hybridization, a weak smear hybridization was also detected, indicating the presence of fragments showing sequence homology with the probes but having EcoRI and BamHI sites located differently.

Hybridization of recombinant plasmids pMRE1, pMRB1.1, and pMRB5 on other single and double restriction digests (Table 1) allowed the construction of a restriction map of genomic repeated fragments (Fig. 5). This map rests on the hybridization of the probes with genomic fragments obtained with EcoRI, BamHI, Kpn I, HincII, and combinations of these enzymes. Some fragments (2900-bp EcoRI, 2500-bp Kpn I, 2500-bp HincII) were absent from the probes but could easily be placed on the restriction map of genomic fragments. In contrast, another genomic fragment also absent from the probes (820-bp Kpn I) could not be put on the map of Fig. 5; such a fragment was probably due to the polymorphism of the restriction map of the genomic repeated fragments (see Discussion); in all likelihood, for the same reason, the 1720-bp HincII fragment of pMRB1.1 (Fig. 5) was not found in the genomic fragments. It should also be noted that at least one Hae III fragment (1960

Fig. 1. Repeated EcoRI and BamHI fragments in mouse DNA and its major components. Thirty micrograms of mouse DNA (T) and 25 µg each of its major components of buoyant density 1.699 (lanes 1), 1.701 (lanes 2), 1.704 (lanes 3), and 1.708 (lanes 4) g/cm³ were digested by EcoRI (A) or BamHI (C), electrophoresed on 1% agarose, and stained with ethidium bromide. (B) Autoradiogram obtained by hybridizing the pMRE1 probe with EcoRI fragments. Molecular weight markers (M) were a phage DNA digested by EcoRI and simian virus 40 DNA digested by Hae III.

Fig. 2. Repeated Kpn I fragments in mouse DNA. Mouse DNA major component 1.699 g/cm³ was digested by Kpn I (lanes 1) and Kpn I plus Bam HI (lanes 2), electrophoresed on 1% agarose, and stained with ethidium bromide (A and A'). (B, C, and D) Autoradiograms obtained after hybridizing labeled pMRB1.1, pMRE1, and pMRB5, respectively, with the transferred digests. Molecular weight markers (M) are as in Fig. 1.

Fig. 3. Repeated HincII fragments in mouse DNA. Mouse DNA major component 1.699 g/cm³ was digested by HincII (lanes 1) and HincII plus BamHI (lanes 2), electrophoresed on 1% agarose, and stained with ethidium bromide (A). (B, C, and D) Autoradiograms obtained after hybridizing labeled pMRE1, pMRB1.1, and pMRB5, respectively, with the transferred digests. Molecular weight markers (M) are as in Fig. 1.

Fig. 4. Repeated Hae III fragments in mouse DNA and its major components. Mouse total DNA (T) and its major components of buoyant density 1.699 (lanes 1), 1.701 (lanes 2), 1.704 (lanes 3), and 1.708 (lanes 4) g/cm³ (in quantities proportional to their amounts in total DNA; see ref. 13) were digested by Hae III, electrophoresed on 2.5% agarose, and stained with ethidium bromide (A). (B, C, and D) Autoradiograms obtained after hybridizing labeled pMRB1.1, pMRE1, and pMRB5, respectively, with the transferred digests. Molecular weight markers (M) were a phage DNA digested by EcoRI plus HincIII.
Fig. 5. Restriction maps of cloned and genomic repeated restriction fragments of mouse DNA. In addition to the sites indicated on the maps, the following information is available on the cloned fragments: pMRE1 contains one Hae III site (whose cleavage produces two fragments, 1000 and 350 bp) but no BamHI, HindII, or Kpn I sites; pMRB1.1 produces six fragments upon digestion with Hae III: 1230, 1120, 880, 350, 315, and 115 bp (the 1230- and 315-bp fragments are BamHI/Hae III fragments); pMRB5 contains two Alu I sites, one Hpa II site, and no Bgl II, EcoRI, Hae III, HindIII, Kpn I, or Sau3A sites. Concerning the genomic map, Bgl II and HindIII sites were not mapped, and only two Hae III sites were mapped in order to show that the 4000-bp and 540-bp BamHI fragments are linked.

bp) visible after ethidium bromide staining did not hybridize with any of the probes; on the other hand, one Hae III fragment (1500 bp) not visible after staining hybridized with pMRB1.1.

The common hybridization of the three probes on the 1860-bp Hae III genomic fragment (Fig. 7, lanes 2) revealed that the 540-bp BamHI fragment is contiguous, or nearly contiguous, to the 4000-bp BamHI fragment. Moreover, because a Kpn I site can also be mapped at about 1000 bp from the end of the 4000-bp BamHI fragment (see above and Fig. 5), the BamHI genomic fragments belong in segments at least 5600 bp long.

Table 1. Hybridization of cloned repeated fragments to restriction digests of mouse DNA

<table>
<thead>
<tr>
<th>Digest</th>
<th>BamHI</th>
<th>EcoRI</th>
<th>BamHI + EcoRI</th>
<th>Kpn I</th>
<th>Kpn I + BamHI</th>
<th>Kpn I + EcoRI</th>
<th>HindII</th>
<th>HindII + BamHI</th>
<th>HindII + EcoRI</th>
<th>Hae III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pMRE1</td>
<td>4000</td>
<td>1350</td>
<td>1350</td>
<td>Smear</td>
<td>Smear</td>
<td>2050</td>
<td>2050</td>
<td>Smear</td>
<td>2050</td>
<td>1350</td>
</tr>
<tr>
<td>Figs.</td>
<td>6A1</td>
<td>1B</td>
<td>6A3</td>
<td>6A2</td>
<td>2C1</td>
<td>2C2</td>
<td>NS</td>
<td>3B1</td>
<td>3B2</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4B</td>
</tr>
<tr>
<td>pMRB1.1</td>
<td>4000</td>
<td>1350</td>
<td>2400</td>
<td>2500</td>
<td>2050</td>
<td>2500</td>
<td>2500</td>
<td>1850</td>
<td>1850</td>
<td>1350</td>
</tr>
<tr>
<td></td>
<td>2900</td>
<td>1350</td>
<td>820</td>
<td>1500</td>
<td>1350</td>
<td>Smear</td>
<td>540</td>
<td>Smear</td>
<td>540</td>
<td>1180</td>
</tr>
<tr>
<td></td>
<td>Smear</td>
<td>290 (NS)</td>
<td>820</td>
<td>Smear</td>
<td>820</td>
<td>Smear</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Figs.</td>
<td>6B1</td>
<td>6B3</td>
<td>6B2</td>
<td>2B1</td>
<td>2B2</td>
<td>NS</td>
<td>3C1</td>
<td>3C2</td>
<td>NS</td>
<td>4C</td>
</tr>
<tr>
<td>pMRB5</td>
<td>540</td>
<td>Smear</td>
<td>540</td>
<td>Smear</td>
<td>540</td>
<td>Smear</td>
<td></td>
<td>540</td>
<td>Smear</td>
<td>1860</td>
</tr>
<tr>
<td></td>
<td>6C1</td>
<td>6C3</td>
<td>6C2</td>
<td>2D1</td>
<td>2D2</td>
<td>NS</td>
<td>3D1</td>
<td>3D2</td>
<td>NS</td>
<td>4D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NS, not shown.

DISCUSSION

Distribution of Repeated Fragments in the Major Components of Mouse DNA. The results of Figs. 1 and 4 indicate that, except for the 1350-bp BamHI fragment that appears to be present in all four components (Fig. 1C), all the fragments are present, at about the same extent, in the two light components. Bands corresponding to the 1350-bp EcoRI fragment and to some Hae III fragments appear to be also present in the 1.704 g/cm³ component, but hybridization with cloned probes was barely detectable on those bands (Figs. 1B and 4 B–D). Because there is independent evidence for a small extent of contamination of the 1.704 g/cm³ component by the lighter ones (11), the conclusion should be drawn that the family of repeated sequences studied here is present almost exclusively in the light components of the mouse genome. Such presence is not due
to the fact that the 4000-bp BamHI fragments have a buoyant density, 1.698 g/cm³, close to those of the light components, because they are short relative to the average size of the DNA preparations used. Instead, it indicates that the flanking sequences of several thousand interspersed repeats scattered in the genome have about the same base composition, in keeping with the large compositional homogeneity of the major components of warm-blooded vertebrates (11). Finally, the different genomic distribution of the 1350-bp BamHI fragment compared to the other fragments points to the existence of different families of repeated fragments in the mouse genome. Such a conclusion cannot be reached with certainty merely on the basis of a lack of cross-hybridization, as shown by the fact that the 540-bp BamHI fragment does not cross-hybridize with the 4000-bp BamHI fragment and yet is contiguous to it.

Restriction Maps of Cloned Repeated Fragments. The data of Fig. 5 allow two main conclusions to be reached. The first is that the repeated fragments are polymorphic in their restriction maps, a finding in agreement with previous results (4–6). In spite of this restriction site polymorphism, both the conservation of a number of restriction sites and the hybridization results obtained among cloned fragments and with genomic fragments (see below) strongly indicate a basic homology in the family of repeated sequences as studied here. The second conclusion is that the cloning and the mapping of the 4000-bp BamHI fragments has allowed us to observe that the 1350-bp EcoRI fragment is a part of a larger repeat.

Restriction Maps of Genomic Repeated Fragments. The use of labeled cloned repeated fragments as probes allowed us to link together most fragments seen by gel electrophoresis of restriction digests of mouse DNA into repeats over 5600 bp long. In fact, the 540-bp BamHI fragment is contiguous to the 4000-bp BamHI fragment, and the latter appears to extend to a KpnI site located at about 1000 bp from its other end (Fig. 5). The only exceptions found were (i) the 1350-bp BamHI fragments; (ii) the 540- and 970-bp BamHI fragments (Fig. 1C) and the 1860-bp HaeIII fragments (Fig. 4). The former have a different genomic distribution and belong, therefore, in a different family of interspersed repeated sequences, and the latter have the same distribution as the 4000- and 540-bp BamHI fragments but do not hybridize with any of the probes and cannot be placed on the genomic map (Fig. 5). They might either represent segments adjacent to the main, 4000- and 540-bp BamHI fragments or correspond to one (or more) additional family (families) of repeated sequences. It appears, however, that some of the repeated sequences discussed here may not be included in larger repeats, because 1350-bp EcoRI fragments not included in 4000-bp BamHI fragments have been found close to the β-globin gene cluster (19); interestingly enough, the β-globin gene is also localized in one of the two light components (13) that contain the EcoRI fragments.

Estimating the stoichiometry of the repeated restriction fragments contained within the 5.6-kb units is not easy; at the present time it can only be said that the data obtained in the present work are compatible with the idea that the map of Fig. 5 depicts the situation of the majority of the fragments. Equally difficult is to estimate the amount of the fragments belonging in the 5.6-kb units relative to that of the fragments unrelated to them. Because of the fact that the background smear varies in intensity from the top to the bottom of the gel and in different major components of mouse DNA at corresponding molecular weight levels, the intensity of the bands stained by ethidium bromide provides indications that may be misleading as to the relative amounts of DNA fragments contained in the bands, as clearly demonstrated by hybridization experiments (Fig. 1A and B).

In spite of these problems, the gel electrophoresis results suggest that the family of repeated sequences studied here is the major one in the mouse genome. Two independent arguments are in agreement with this conclusion. The first argument is that the 1350-bp EcoRI fragments have been estimated to represent 0.5–3% of mouse DNA (1, 4, 5); if the majority of them belong in larger repeats 5.6 kb in size, then these may correspond to up to 2–12% of mouse DNA. The second argument is that the genomic distribution of the repeated sequences studied here parallels that of interspersed repeated sequences, as estimated by the reassociation kinetics of the major components, in that the latter are very abundant in the light components and very scarce in the heavy ones (12). If such a scarcity is accounted for, at least to a large extent, by the absence of the BamHI repeated sequences in the heavy components, then these should represent a sizable fraction of the mouse genome.

As far as the functional role of the repeated sequences studied here is concerned, an important point is that their genomic distribution does not parallel that of genes, because the latter are present, in roughly equal relative amounts, in all four of the components of the mouse genome (11) [in particular, the α-globin genes are on the 1.708 g/cm³ component, whereas the β-globin genes are on the 1.701 g/cm³ component (13)]. This rules out at least any simple correlation between the major family of interspersed repeated sequences and the regulation of gene expression in the mouse genome. In agreement with this conclusion is the finding (12) of very large differences in the distribution of interspersed repeated sequences in the major components of the mouse and human genomes, which contrasts with the conserved distribution of structural genes, because these differences are unlikely to concern sequences involved in the regulation of gene expression. On the other hand, the very existence of distinct families of interspersed repeated sequences, demonstrated here by their different genomic distribution, suggests by itself different functions for such families.