The EMBO Journal vol.3 no.2 pp.467–472, 1984

Looking for relationships between the most repeated dispersed DNA sequences in the mouse: small R elements are found associated consistently with long MIF repeats

Karen L. Bennett[1] and Nicholas D. Hastie[2]*

Department of Cell and Tumor Biology, Roswell Park Memorial Institute, Buffalo, NY 14263, USA

[1]Present address: Carnegie Institution of Washington, Department of Embryology, 115 West University Parkway, Baltimore, MD 21210, USA
[2]Present address: MRC Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK
*To whom reprint requests should be sent
Communicated by J. Bishop

Four highly-repeated dispersed DNA sequence families have been described in the mouse genome. These are the three small elements B1, B2 and R and the large (6 kb) MIF element. Together these comprise ~10% of the mouse genome. Possible relationships between these families are pertinent to the genome as a whole. We report here that the B1s, B2s and Rs are all randomly organized in the genome with respect to each other. Surprisingly though, the R and MIF families are found together consistently in a set of random genomic clones and in selected clones. We find Rs often located on one end of the MIF at a consistent site and conclude that a minority of Rs are an integral part of MIF while the majority of Rs are not associated with MIFs. We propose that isolated R elements are truncated forms of MIF. Also we speculate on the mechanism of dispersal of these elements through the mouse genome.
Key words: repeated dispersed DNA/R elements/MIF elements

Introduction

In the mouse genome four different dispersed DNA families have been described which are present in 30 000–150 000 copies per genome. These include the B1, which is the mouse Alu equivalent (Krayev *et al.*, 1980); the B2 (Krayev *et al.*, 1982); the MIF, the mouse interspersed fragment, originally called the 1.3-kb *Eco*RI fragment now known to be greater than 5 kb in its entirety (Cheng and Schildkraut, 1980; Heller and Arnheim, 1980; Brown and Dover, 1981; Meunier-Rotival *et al.*, 1982); and the R sequence, a 400–500 bp size family first located in multiple copies amidst the immunoglobulin cluster (Gebhard *et al.*, 1982; Lueders and Paterson, 1982).

In a separate study, we have constructed a small repetitive DNA sequence library from *Mus musculus* genomic DNA and identified clones corresponding to these four families (Bennett *et al.*, 1984). These clones have been used to study the repetition frequency, transcription and heterogeneity of these elements in the mouse genome. We concluded that these are probably the only families repeated >30 000 times in the mouse genome apart from simple polypyrimidine stretches. In trying to understand the evolution and amplification of these sequences in the genome, it is necessary to know whether they always exist as discrete entities or whether they can be found in association with each other at high frequency. As these families together constitute ~10% of the mouse genome, the question of relationships between them is pertinent to the organization of the genome as a whole.

In *Drosophila*, short repetitive families (1 kb) are organized in densely spaced and scrambled clusters (Wensink *et al.*, 1979). In plants, similar large clusters of scrambled arrays of short repeats have been reported (Flavell, 1980). It has been proposed that the short repeats of mammalian DNA are also arranged in tandem clusters (Moyzis *et al.* (1981); however, none of the studies of repetitive sequences in selected genomic clones to date has revealed large blocks of intermingled short repeats in mammals. In the mouse, the 30 000 or more MIF sequences are heterogeneous. However, there is no indication that scrambled rearrangements are responsible for this heterogeneity. When a full repeat is present in the genome, it follows a particular order of conserved restriction sites. The heterogeneity seems to result from less than full length repeats and numerous polymorphic forms. The *Kpn*I family, the most abundant long repeat in humans and monkeys, appears to be similarly arranged (Manuelidis and Biro, 1982; Grimaldi and Singer, 1983).

In this study, we have used the Benton-Davis plaque assay to look for associations between these five families in randomly selected genomic clones. We report that the three small element families, the B1s, B2s and Rs, are all organized in a random fashion with respect to one another, with no evidence of clustering. Surprisingly though, MIF is always found in association with R but R also exists in more copies in isolation. In contrast, the B1 and B2 elements are found at far less than the expected frequency in MIF-containing genomic clones. A consistent structural arrangement of the R and MIF elements is found in individually chosen genomic clones and in mouse DNA. A subset of R elements is precisely located to one end of the MIF long repeat. We discuss why we believe that isolated R sequences are really truncated MIFs and we also discuss a possible mechanism for the transposition of these elements involving reverse transcripts, using RNA intermediates.

Results

Random relationships between small repetitive sequences

Benton-Davis replicate filters can be used to evaluate the relationships between families (Figure 1 and Bennett *et al.*, 1984). For example, the B1 family is found in 80% of any set of random mouse genomic clones (Bennett *et al.*, 1984). If the location of a B1 element is totally independent of any other repetitive element, the B1s should be found in 80% of the plaques containing any non-B1 family for which we have cloned probes.

Indeed B1s are found to be present in ~80% of clones containing B2s and Rs (Figure 1). We have shown elsewhere that R elements are located in 35–40% of a set of random mouse genomic clones (Bennett *et al.*, 1984). As expected, Rs are present in ~40% of the clones containing B1s and B2s

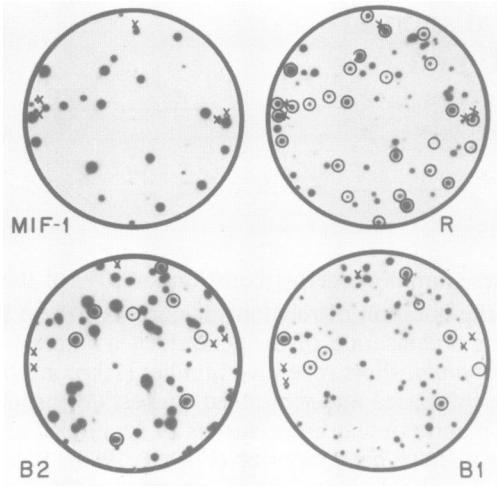

Fig. 1. Relationships of MIF elements to the other highly repeated families in the mouse: the non-random association of the R family with MIF elements. Replicate Benton-Davis filters were hybridized with cloned repetitive DNA from four different sequence families. Plasmids pMR134 (MIF), pMR290 (R), pMR142 (B2) and pMR225 (B1) were each nick-translated and hybridized to replicate filters. Those plaques which contain both the MIF family and other repetitive sequences are circled.

(Figure 1). Similarly, B2 sequences are found as predicted (Bennett *et al.*, 1984) in ~45% of clones containing B1 and R elements (Figure 1).

Also, >20 individual genomic clones have been mapped to varying extents in this study, and only in a very few cases are any two small elements found on the same restriction fragment. Thus, we conclude that three small repeat length families, the B1s the B2s and the Rs, exist in a random relationship with respect to one another. If large arrays of multiple short repeats exist, these are infrequent in comparison with the solitary locations of family members.

A surprising relationship between the MIF and R elements

In contrast to the randomness described above, Figure 1 shows a very different result when comparing plaques on replicate filters between the R and MIF families. We have shown that R is present in 35–40% of random genomic mouse clones whereas MIF is located in ~15% of the genomic clones (Bennett *et al.*, 1984). Though the R family is present in 2.5 times as many plaques as the MIF elements, whenever a plaque contains an MIF sequence it also has an R element. In the series shown in Figure 1, the MIF probe detects 27 plaques, some of them rather faintly. In every case, the R sequence is also present. This non-random relationship is consistent through many trials. Only very occasionally (in 5% out of several hundred MIF plaques) is an R element not detected. When MIF-containing plaques are isolated and re-screened, in all cases R sequences are located in the same plaques.

B1 and B2 sequences are found in MIF-containing genomic clones at far less than the expected frequency

As discussed above, Figure 1 shows that 27/27 MIF-containing genomic clones contain R sequences. We would predict that 80% of MIF-containing clones or 22/27 also contain B1 sequences based on their frequency in the whole genomic library (Figure 1, and Bennett *et al.*, 1984). The results in Figure 1 show that only 8/27 MIF-containing clones also have B1 elements. This result is confirmed by Southern blot analysis of 12 genomic clones selected because they contain MIF (see below). Only four of the 12 contain B1 sequences. Likewise, we would predict that 14/27 MIF-containing clones have B2 elements, when in fact only eight contain B2s. More strikingly, Southern analysis shows that only two out of 12 MIF-containing clones have B2 elements. The possible significance of these results is considered in the Discussion.

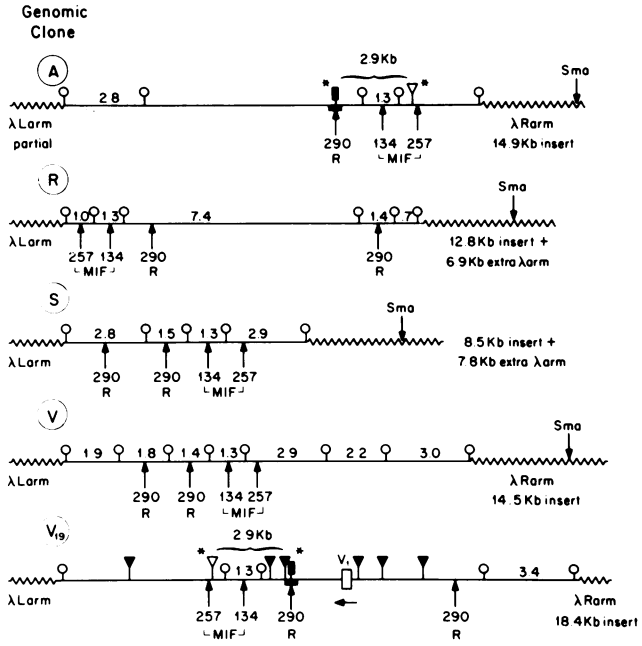

Fig. 2. Preliminary mapping of the R:MIF relationships in five selected genomic clones. The results of mapping with *Eco*RI digests (complete and partials) in addition to using the *Sma*I fragment technique (see description in text and Schibler *et al.*, 1982) are illustrated here. λV19 is a kind gift of Steve Crews. A map of V19 after an *Eco*RI digest is published in Crews *et al.* (1981); the location of *Bam*HI sites are from Steve Crews (personal communication). The other four clones were selected from the Davis mouse library based only on the presence of an MIF element in the insert. The probes pMR290, pMR134 and pMR257 are described in the text. *Only those *Sst*I and *Kpn*I sites which have been definitively mapped are shown. *Eco*RI ○; *Sst*I ■; *Bam*HI ▼; *Kpn*I ▽.

Mapping the R:MIF complex in selected genomic clones

To characterize the relationship of MIF and R families further, a mapping analysis of the two elements in selected genomic clones was carried out. The majority of MIF elements in the genome have 1.3-kb *Eco*RI fragments, but the neighboring MIF-containing *Eco*RI fragments may vary in length. However, a distinct subset has a 2.9-kb *Eco*RI fragment next to the 1.3-kb fragment. We had probes available for both parts of MIF; pMR134 for the 1.3-kb section, pMR257 for the 2.90-kb one. We chose to map the R:MIF relationship more precisely in five clones that hybridized to all three probes and had a 1.3-kb *Eco*RI fragment.

Figure 2 summarizes the results of the mapping studies in five 'typical' MIF clones. As further evidence of the heterogeneity of MIF members, none of these five shared a similar sized *Bam*HI or *Msp*I restriction fragment (not shown). In each of the five genomic clones, with an *Eco*RI digest, the pattern is an R element followed by the 1.3-kb fragment and then the 2.9-kb fragment (R:1.3:2.9). As there are no *Eco*RI sites in the known sequence of the R element, it is not surprising that if R were adjacent to the conserved *Eco*RI fragments of MIF rather than internal, the R-containing fragment would be of variable length. In these five selected clones there

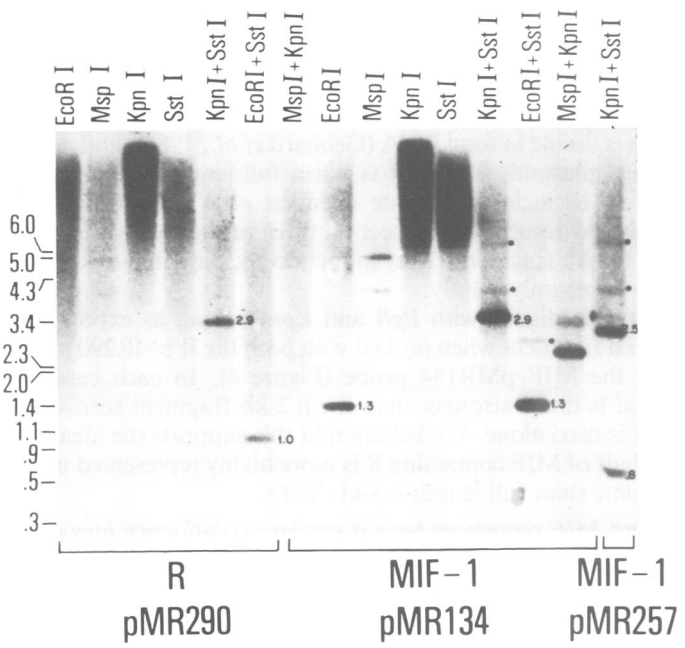

Fig. 3. Analysis of genomic DNA establishes the nature of the R:MIF repeat. 1 μg of genomic mouse DNA was digested with the appropriate enzyme or combination of enzymes, electrophoresed on an agarose gel, transferred to nitrocellulose and hybridized with the specific probes pMR290 (R), pMR134 (MIF). The exposure was 5 h. All digests are limit, with 10 units enzyme/μg DNA at 37°C overnight. MIF members in the genome which presumably have lost restriction sites are indicated with dots. These MIF members confirm the organization of the MIF repeat in the same manner as a partial digest.

is no second R on the other side of MIF. If the relationship of R to MIF were similar to that of long terminal repeats (LTRs) to retroviral genomes or insertion sequences (IS) to procaryotic transposons, we would expect a flanking repeat of the R sequences. We do not find such a structure. In the case of three of these clones, we would not have detected a second R as they join the λ arm immediately past the fragment which hybridizes with the pMR257 probe. However, at least two clones do continue with mouse DNA beyond this fragment and no R is detected in either clone up to an additional 5.2 kb of DNA (in genomic V). Interestingly in 4/5 of these genomic clones, there is a second R element adjoining the first such that the order is two Rs followed by the MIF sequences, R:R;1.3:2.9. However, the distance of the second R from the first is not the same in each case, and in one case is >7 kb from the first (in genomic clone R). Furthermore, for the immunoglobulin clone V_{19}, the coding sequences of the V_H gene lies between the two Rs. The size of the MIF element determined by S1 nuclease heteroduplex analysis of genomic DNA (Bennett *et al.*, 1984) was only slightly larger than 6 kb with no evidence for larger elements. As yet, we have not been able to determine the relevance, if any, of these additional R elements. Their locations in 4/5 clones beyond the full length R:MIF complex may be only a result of the high repetition frequency of R in the genome, and the tendency of Rs to exist in multiple nearby locations, as in the globins and the immunoglobulins, perhaps due to the high frequency of transposability of these elements.

The R sequence is shown to map at a consistent site to one end of the MIF sequence in genome mouse DNA

We have mapped the position of R sequences relative to MIF sequences in five genomic clones. However, as there are

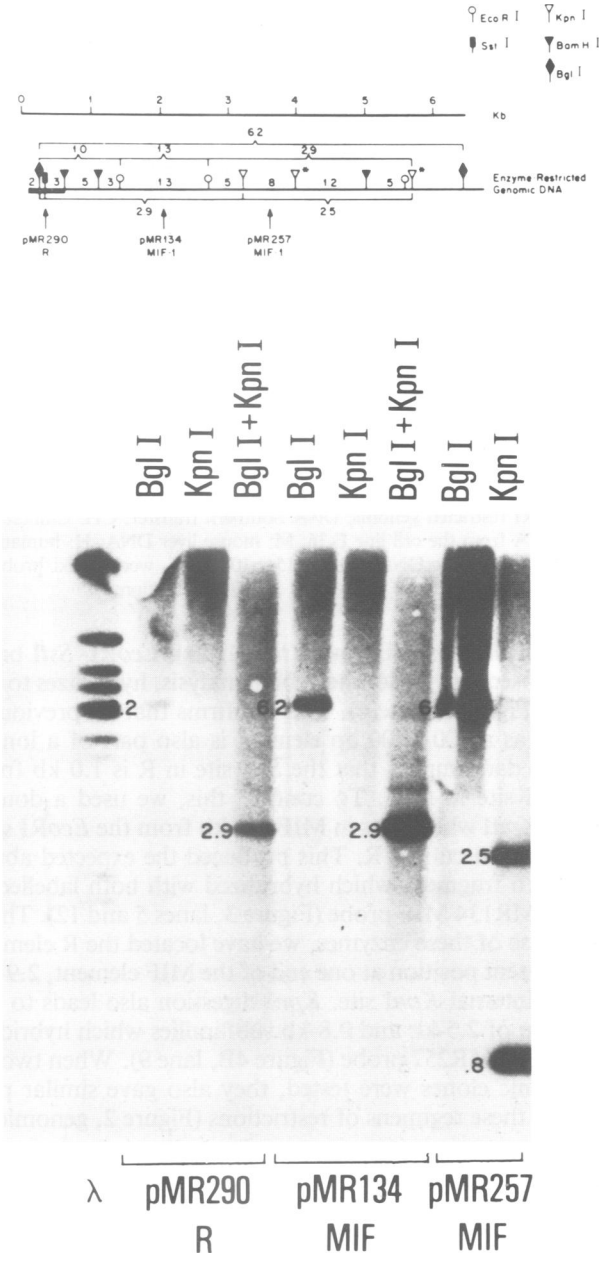

Fig. 4. A generalized map of the R:MIF structure in genomic DNA and the use of an enzyme which produces a large consistent-sized R:MIF element. **(A)** The *Bam*HI restriction sites are based on the map of Meunier-Rotival *et al.* (1982). **(B)** Hybridization with appropriate probes after Southern transfer. The exposure shown here is 14 h. *These appear to be two different *Kpn*I subgroups which are both present in the genome.

30 000 MIF elements in the genome we next thought it would be appropriate to map the relationship in genomic mouse DNA. The consensus sequence of R reveals a single *Sst*I site (and a single *Bgl*I site 30 bp from the *Sst*I site) (Gebhard *et al.*, 1982). We reasoned that if the R element is a precise distance from the 1.3-kb *Eco*RI site and from the known *Kpn*I sites in MIF (see map, Figure 4A), then double digestion of mouse DNA should produce abundant, discrete fragments, containing R and MIF sequences. Indeed ethidium bromide staining after gel electrophoresis shows the predicted bands (data not shown). Digestion with only *Eco*RI gives the prominent 1.3-kb band which was the first identifying characteristics of the family (Heller and Arnheim, 1980). A digestion with both *Eco*RI and *Sst*I gives the same 1.3-kb R1 band

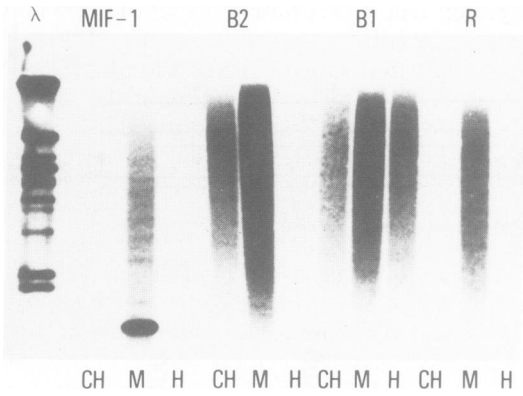

Fig. 5. The different patterns of evolutionary conservation of the four most repeated dispersed DNA sequences in the mouse. Probes were from pMR134 (MIF-1), 142 (B2), 225 (B1) and 290 (R), and were hybridized against *Eco*RI restricted genomic DNA Southern transfer. CH: Chinese hamster DNA from the cell line E-36; M: mouse liver DNA; H: human lymphocyte DNA. 1 μg DNA/lane. ≈15 x 10^6 c.p.m. were added/probe. The λ markers are a mix of *Eco*RI and *Hind*III restrictions.

which hybridises to MIF and a new 1.0-kb *Eco*RI- *Sst*I band which, subsequent to Southern blot analysis, hybridizes to the R probe (Figure 3, lane 6). This confirms that R, previously identified as a 400−500 bp element is also part of a longer unit. This data implies that the *Sst*I site in R is 1.0 kb from the *Eco*RI site in MIF. To confirm this, we used a double digest of *Kpn*I which cuts in MIF 1.9 kb from the *Eco*RI site, and *Sst*I which cuts in R. This produced the expected abundant 2.9-kb fragment which hybridized with both labelled R and the pMR134 MIF probe (Figure 3, lanes 5 and 12). Thus, with the use of these enzymes, we have located the R element to a consistent position at one end of the MIF element, 2.9 kb from the internal *Kpn*I site. *Kpn*I digestion also leads to the production of 2.5-kb and 0.8-kb subfamilies which hybridize to our MIF pMR257 probe (Figure 4B, lane 9). When two of our genomic clones were tested, they also gave similar patterns with these regimens of restrictions (Figure 2, genomic A and V_{19}).

The analysis of plaques (Figure 1) showed that only 40% or so of R sequences are associated with MIF elements. Thus, we would not expect all the genomic DNA which hybridises to the R probe to be in the discrete 1.0-kb *Sst*I and 2.9-kb *Kpn*I-*Sst*I fragments. This is the case; there is an obvious smear in addition to specific fragments. Also, the R probe reveals a greater smear relative to the specific bands in the *Kpn*I-*Sst*I double digest than does the MIF pMR134 probe.

If there were another R at the other end of the MIF elements in a significant proportion of the MIF population, the second R would have been detected as an additional hybridizing fragment beyond the second *Kpn*I site in the *Kpn*I-*Sst*I double digest. As the *Kpn*I internal sites are not symmetric the additional R fragment would not be hidden in identical sized fragments to those already hybridizing. Similarly, we do not find the second R element on the same side as the known R at a characteristic measurable distance from the first; the *Sst*I single digest would have revealed such an exact R:R relationship.

In terms of the possible significance of this R:MIF association, it is important to try to distinguish whether the R element is at the exact end of the MIF or has inserted itself into the MIF and subsequently amplified with the MIF in the mouse genome. It was expedient to find an enzyme which would cut full-length MIF elements very near their ends. Such an enzyme exists. We have made use of a *Bgl*I digest of genomic DNA reported by Fanning (1982) which cuts MIF into a large, 6.2-kb fragment (Figure 4B). This abundant size class is visible in total DNA (Gebhard *et al.*, 1982), and as our cloned plasmid, though it is not a full length R element, is known to include that site (Bennett *et al.*, 1984, and unpublished results), we expect all three probes to hybridize to this 6.2-kb fragment. They in fact do so, though our R probe hybridizes only faintly.

Double digests with *Bgl*I and *Kpn*I reveal, as expected, a 2.9-kb fragment when probed with both the R pMR290 probe and the MIF pMR134 probe (Figure 4). In each case the signal is much stronger than the 6.2-kb fragment seen when *Bgl*I is used alone. We believe that this supports the idea that the half of MIF containing R is more highly represented in the genome than full length 6.5-kb MIFs.

R and MIF sequences have a similar evolutionary history

We have shown that R can apparently exist as isolated discrete units and at one end of the large MIF elements. It is of interest to consider whether R has a separate origin from MIF and became associated with MIF recently in evolutionary time or whether the elements arose as a single entity. This would require a comprehensive knowledge of structure and number of MIF and R elements in a range of mammalian and perhaps other vertebrate species. This is outside the scope of this study but a preliminary evolutionary study does suggest that MIF and R have evolved at similar rates which are quite different from those seen for B1 and B2. Figure 5 shows the result of an experiment in which the probes for all the highly repeated families were hybridized to mouse, hamster and human DNA at the relatively low stringency of 60°C and 4 x SSC. Several patterns are seen. As expected, the B1 probes cross-hybridize with both hamster and human DNA. The B2 family hybridizes to the mouse DNA and as strongly to the hamster DNA, as expected from the 90% sequence homology (Haynes and Jelinek, 1981), but not at all to the human and is the only sequence with such a pattern. Both the MIF and R sequences are mouse-specific under these conditions, though with lower temperature (55°C), a Chinese hamster MIF counterpart is seen very faintly.

This inability to detect MIF and R hybridization to hamster and human DNA could be due to sequence divergence, a low concentration of these elements in the other mammals or a combination of both. Certainly sequence divergence is a major factor as we observe a low level of hybridization to hamster DNA at 55°C, and Lueders and Paterson (1982) have shown that mouse R sequences form a duplex with rat DNA which melts at 18°C below the *T*m of R duplexes formed with mouse DNA.

As a practical consideration, the MIF, R and, in some cases, B2 probes would all be useful in detecting the presence of mouse DNA in gene transfer experiments into cells of heterologous species.

Discussion

The nature of the MIF:R association. Are R sequences really truncated MIFs

Our data suggests that most and probably all MIF elements contain an R sequence at one end. This was first revealed by Benton-Davis analysis which showed that R sequences are present in 2.5 times as many genomic clones as the part of

MIF which hybridizes to our clone pMR134 (Figure 1). However, at least 95% of the MIF-containing clones contain an R sequence. In total this relationship has been confirmed in several hundred clones by Benton-Davis analysis. Also 12/12 clones selected as MIF-containing were also shown to include Rs by Southern analysis. Finally, analysis of genomic DNA has shown that the *Sst*I site in R is found consistently 2.9 kb away from the *Kpn*I site in MIF. In fact this 2.9-kb fragment gives a slightly stronger signal when hybridized with our pMR134 probe than does the characteristic 1.3-kb fragment. This suggests that most and probably all MIFs include R and that R is an integral part of MIF. Of course we cannot conclude this for the MIFs which produce a smear on Southern analysis of genomic DNA which is presumably due to variation within MIFs. However, the constant association in genomic clones suggests that the relationship holds for all MIFs. The evidence suggests that the MIF-associated Rs all map very close to the end of MIF. Our data suggests that a full length MIF is between 6.2 and 6.5 kb in length. The S1 nuclease-resistant duplexes containing MIF are a continuum of sizes up to a very obvious maximum cut off at 6.5 kb (Bennett *et al.*, 1984). The largest consensus restriction fragment shown to include MIF is a 6.2-kb *Bgl*I fragment (Figure 4). As *Bgl*I cuts within R at least 200 nucleotides from the end, the terminus of R must be at least 6.4 kb from the other *Bgl*I site. Taken together, these data support the idea that R is at one end of MIF. It seems that there are two alternative explanations for the relationship between MIF and R. Either Rs are truly discrete small elements which have also become associated with long MIFs or, alternatively, Rs are really truncated versions of MIF. We favour the latter explanation for several reasons. Rs were originally described as a more or less discrete population of small elements with a consensus size of ~475 bp (Gebhard *et al.*, 1982). Also the heteroduplex analysis carried out by ourselves (Bennett *et al.*, 1984) and Lueders and Paterson (1982) supported this conclusion. However, although all the Rs sequenced were shown to have a common terminus (ascribed to be 3′), the 5′ terminus was never actually determined (Gebhard *et al.*, 1982). It was shown that one R stopped at ~200 nucleotides from the 3′ end but the remaining four clones were only sequenced to a common *Bam*HI site at 475 nucleotides from the 3′ end, thus the position of the consensus 5′ end was not determined. Hence, it was possible that the R elements extended into the 0.5-kb *Bam*HI region of MIF. These R elements would not have been much longer than 500 nucleotides as repetitive sequence hybridization did not extend beyond the *Bam*HI site. During the preparation of this manuscript, Wilson and Storb (1983) reported on the sequence of R-containing elements near immunoglobulin K chain genes and showed that several Rs extended into the 0.5-kb *Bam*HI sequence of MIF [which they call BAM5 and which had been previously sequenced by Fanning (1982), see Figure 4] such that the small *Bam*HI fragment and R together were flanked by direct repeats. Again, all these elements showed very similar A-rich 3′ ends which were nearly identical to those reported by Gebhard *et al.* (1982). There was, however, no indication of a shared 5′ end although two still finish within 10 nucleotides of each other. Wilson and Storb only found these *Bam* fragment sequences associated with Rs or Rs alone, never BAM5 fragment sequences alone. They found no cases where the next 1.3-kb *Eco*RI MIF sequences were also present. They concluded that BAM5 and R sequences are two small elements which have fused together. Our interpretation is that they are both truncated versions of MIF.

Our own data also support the conclusion that there is a range of overlapping MIFs all starting at the same 3′ end. Our S1 nuclease data (Bennett *et al.*, 1984) shows that although Rs appear to be mainly in the size range of 400–500 nucleotides there is a continuum of R-containing duplexes up to 6.5 kb. What is more there are a number of large discrete size classes of duplex DNAs (which hybridize both to R and to MIF probes suggesting that they all share the same 3′ end at R but probably have truncated 5′ ends). Also, hybridization of MIF and R probes to genomic DNA shows that the 2.9-kb *Kpn*I-*Bgl*I fragment is more highly represented in the genome than the 6.2-kb *Bgl*I fragment (Figure 4). Again this supports the idea that the R end of MIF is in higher concentration than full length MIFs. Finally, we have shown that, whereas 35–40% of genomic clones hybridize to pMR290, 15–20% hybridize to the pMR134 probe and only 10–12% to the pMR257 probe. The majority of clones which hybridize to pMR257 also hybridize to pMR134 and pMR290. The most likely explanation is that there are different length overlapping MIF members which all share the same 3′ end but have different 5′ ends with a distinct polarity to the elements. It is important to note that in our mapped genomic clones and total genomic DNA the R is in the same orientation on the abundant long MIF elements as found in the truncated versions detailed by Wilson and Storb (1983). Gebhard and Zachau (1983) have recently come to the same conclusion that Rs are truncated MIFs.

We have also shown that MIF-containing genomic clones have less than half the number of B1 and B2s expected from an analysis of random clones. Also we have not detected a B1 or B2 in the middle of an MIF element. As MIFs (those detected by our pMR134 probe) are on average ~5–6 kb long and genomic clones are 10–15 kb long, it seems likely that the presence of MIF excludes B1s and B2s. There are several possible explanations for this. Perhaps B1s and B2s inserted into some MIFs but were removed by gene conversion or selected against. Alternatively, the data could imply that MIF sequences have amplified and dispersed through the genome since the B1s and B2s did so.

We did not detect sequences in humans which are homologous to MIFs by hybridization (Figure 5). It is not clear whether this is due to sequence divergence or lack of the sequences in humans. It is interesting to note that the *Kpn*I element in humans is also 6.4 kb long in its entirety (Grimaldi and Singer, 1983). Sequence data will reveal whether there is any homology of this element to the MIF.

Possible mechanism for transposition of Rs

Jagadeeswaran *et al.* (1981) proposed a model for transposition of human Alu sequences in which DNA intermediates were synthesized from RNA by a reverse transcriptase. Similarly Van Arsdell *et al.* (1981) proposed that reverse transcripts of U1, U2, and U3 snRNAs were intermediates in dispersal of the pseudogenes of these sequences throughout the genome. This model was constructed to explain the structure of the pseudogenes which all had 5′ ends which coincided with the 5′ of RNA transcripts but had truncated 3′ ends. The whole unit is flanked by short direct repeats of a sequence at the insertion site. Bernstein *et al.* (1983) have obtained more direct evidence supporting the idea by showing that U3 RNAs can be reverse transcribed *in vitro* to produce

the predicted intermediates. Similarly it has been suggested that pseudogenes for various structural genes including those coding for human and rat tubulin and human and rat metallothionin are derived from reverse transcripts of these mRNAs (as reviewed in Sharp, 1983). This was proposed to explain several observations. Firstly, the pseudogenes lack introns; secondly, a run of deoxyadenine residues is found at the 3′ end of the peudogene just after the poly(A) addition site AATAAA and finally, in some cases the pseudogene ends at the position of the mRNA 5′ terminus. One of the factors common to all R sequences is an A-rich end which includes the sequence AATAAAA just prior to the flanking direct repeat (Gebhard *et al.*, 1982; Wilson and Storb, 1983). In some cases this is followed by a stretch of A residues. Because of this, Gebhard *et al.* (1982) proposed that R elements dispersed by way of reverse transcripts of a RNA transcript.

If, as we try to argue above, R sequences are really truncated MIFs then it might be reasonable to propose that MIFs code for long RNA transcripts which have the R sequence at the 3′ end. Then reverse transcription of this RNA takes place but is often incomplete stopping within 400−500 nucleotides from the 3′ end of the RNA, thus generating transposition intermediates for R and the other truncated MIFs. Alternatively, perhaps some transcripts initiating towards the 3′ end of MIF could be reverse transcribed to produce truncated sequences. It is too early to speculate whether full length MIFs could have arisen by such a mechanism as the ends have not yet been characterised.

Materials and methods

Materials

Inbred mice of C57Bl/6J and BALB/cJ strains were purchased from West Seneca Laboratories. Restriction enzymes were purchased from Bethesda Research Laboratories and New England BioLabs and were used as per manufacturer's recommendations. The Charon 4A mouse library is from Dr. M. Davis.

Maintenance and analysis of recombinant DNA clones

The original library of genomic repetitive sequences was constructed at the *Pst*I site of pBR322. Repetitive hybrids were cloned with the use of GC tails (Pietras *et al.*, 1983). The insert containing plasmids have been maintained in the *Escherichia coli* strain X1776. Subsequently, plasmids representative of each repetitive family have been transformed into the *E. coli* strain HB101. For initial identifications of plasmids, small scale isolations of plasmid DNA were performed following the method of Birnboim and Doly (1979). Large scale isolation of both plasmid DNA and bacteriophage λ DNA follow standard procedures as outlined in Maniatis *et al.* (1982). Assays of recombinant λ plaques followed the protocols as outlined by Benton and Davis (1977).

DNA preparation

Mouse liver DNA was extracted as described by Piccini *et al.* (1982).

Southern blot analysis and hybridizations

After agarose gel electrophoresis, DNA was denatured and transferred to nitrocellulose according to Southern (1975). All hybridizations were done with 4 x SSC (0.6 M NaCl/0.06 M sodium citrate), 5 x Denhardt's (Denhardt, 1966), 0.1% SDS and 0.1% sodium pyrophosphate with the addition of 150 μg of denatured salmon sperm DNA/ml hybridization solution. In those hybridizations that involved genomic clones of mouse DNA in Charon 4a (some Southern blots and all Benton-Davis assays) denatured *E. coli* DNA at 100 μg/ml and denatured λ DNA at 2−3 μg/ml were also added to eliminate background hybridization. All hybridizations were done at 60°C.

Restriction mapping of lambda clones

This was carried out with the aid of the *Sma* mapping techniques described by Schibler *et al.* (1982).

Plasmids

The plasmid probes used in this study were all isolated from a small library of repeated sequences described in Pietras *et al.* (1983) and Bennett *et al.* (1984). The plasmids pMR225 (for the B1 family), pMR142 (for the B2 family), pMR290 (for the R family) and pMR134 and pMR257 (for the MIF-1 family) all contain inserts which are between 150 and 200 nucleotides in size. The identity of the inserts has been confirmed by DNA sequencing (pMR225, pMR142, and pMR290) or by use of the probes in comprehensive restriction mapping analysis (pMR34, and pMR257).

Acknowledgements

We acknowledge Dr. Robert E. Hill for continued helpful discussion. We appreciate the excellent technical assistance of Colleen Kane-Haas, Mary Bodis and Fred Gaskin. We thank Katie Rae for excellent typing and processing of the manuscript and Sandy Bruce for the very high quality of the photographic work produced under duress. We thank Steve Crews of Dr. Leroy Hood's laboratory for the three immunoglobulin genomic clones and additional restriction maps. We thank Drs. Sue Naylor and Alan Sakaguchi for the gift of human lymphocyte DNA and Dr. Peter Lalley for DNA for the Chinese hamster cell line E-36. We thank Drs. Gebhard and Zachau for communicating to us the results of their studies prior to publication. This work was supported by National Institutes of Health grant GM-24771-05 to NDH.

References

Bennett,K.L., Hill,R.E., Pietras,D.F., Woodworth-Guati,M., Kane-Haas, C., Houston,J.M., Heath,J.K. and Hastie,N.D. (1984) *Mol. Cell. Biol.*, in press.
Benton,W.D. and Davis,R.W. (1977) *Science (Wash.),* **196**, 180-182.
Bernstein,L.B., Mount,S.M. and Weiner,A.M. (1983) *Cell,* **32**, 461-472.
Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.,* **7**, 1513-1523.
Brown,S.D.M. and Dover,G. (1981) *J. Mol. Biol.,* **150**, 441-466.
Cheng,S.M. and Schildkraut,C.L. (1980) *Nucleic Acids Res.,* **8**, 4075-4091.
Crews,S., Griffin,J., Huang,H., Calame,K. and Hood,L. (1981) *Cell,* **25**, 59-66.
Denhardt,D. (1966) *Biochem. Biophys. Res. Commun.,* **23**, 641-646.
Fanning,T.G. (1982) *Nucleic Acids Res.,* **10**, 5003-5013.
Flavell,R.B. (1980) *Annu. Rev. Plant Physiol.,* **31**, 569-596.
Gebhard,W., Meitinger,T., Hochtl,J. and Zachau,H.G. (1982) *J. Mol Biol.,* **157**, 453-471.
Gebhard,W. and Zachau,H.G. (1983) *J. Mol. Biol.,* in press.
Grimaldi,A. and Singer,M.F. (1983) *Nucleic Acids Res.,* **11**, 321-338.
Haynes,S.R. and Jelinek,W.R. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 6130-6134.
Heller,R. and Arnheim,N. (1980) *Nucleic Acids Res.,* **8**, 5031-5042.
Jagadeeswaran,P., Forget,B. and Weissman,S.M. (1981) *Cell,* **26**, 141-142.
Krayev,A.S., Kramerov,D.A., Skryabin,K.G., Ryskov,A.P., Bayev,A.A., Georgiev,G.P. (1980) *Nucleic Acids Res.,* **8**, 1201-1215.
Krayev,A.S., Markusheva,T.V., Kramerov,D.A., Ryskov,A.P., Skryabin, K.G., Bayev,A.A. and Georgiev,G.P. (1982) *Nucleic Acids Res.,* **10**, 7461-7475.
Lueders,K.K. and Paterson,B.M. (1982) *Nucleic Acids Res.,* **10**, 7715-7729.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual,* published by Cold Spring Harbor Laboratory Press, NY.
Manuelidis,L. and Biro,P.A. (1982) *Nucleic Acids Res.,* **10**, 3221-3239.
Meunier-Rotival,M., Soriano,P., Cuny,G., Strauss,F. and Bernardi,G. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 355-359.
Moyzis,R.K., Bonnet,L., Li,D.W. and Ts'o,P.O.P. (1981) *J. Mol. Biol.,* **153**, 871-896.
Piccini,N., Knopf,J.L. and Gross,K.W. (1982) *Cell,* **30**, 205-213.
Pietras,D.F., Bennett,K.L., Siracussa,L.F., Woodworth-Gutai,M., Chapman,V.M., Gross,K.W., Kane-Haas,C. and Hastie,N.D. (1983) *Nucleic Acids Res.,* **11**, 6965-6983.
Schibler,U., Pittet,A.-C., Young,R.A., Hagenbuchle,O., Tosi,M., Gellman, S. and Wallauer,P.K. (1982) *J. Mol. Biol.,* **155**, 247-266.
Sharp,P.A. (1983) *Nature,* **301**, 471-472.
Southern,E.M. (1975) *J. Mol. Biol.,* **98**, 503-517.
Van Arsdell,S.W., Denison,R.A., Bernstein,L.B., Weiner,A.M., Manser,T. and Gesteland,R.F. (1981) *Cell,* **26**, 11-17.
Wensink,P.C., Tabata,S. and Pachl,C. (1979) *Cell,* **18**, 1231-1246.
Wilson,R. and Storb.U. (1983) *Nucleic Acids Res.,* **11**, 1803-1817.

Received on 21 November 1983

Note added in proof

The paper by Gebhard and Zachau has now appeared in *J. Mol. Biol.,* (1983) **170**, 255-270. Since submission of this paper it has been shown by DNA sequencing that R sequences are at the end of MIF-1 elements (Fanning,T.G. (1983) *Nucleic Acids Res.,* **11**, 5073-5091). Also it has been shown by sequencing that there is homology between the MIF-1 sequence of the mouse and the human *Kpn*I family (Singer,M.F., Thayer,R.E., Grimaldi,G., Lerman,M.I. and Fanning,T.G. (1983) *Nucleic Acids Res.,* **11**, 5739-5745).