Volume 11 Number 24 1983 Nucleic Acids Research

The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end

Charles F.Voliva, Carolyn L.Jahn*, Mary B.Comer, Clyde A.Hutchison, III, and Marshall H.Edgell

Department of Microbiology and Immunology, Curriculum in Genetics, and Program in Molecular Biology and Biotechnology, University of North Carolina, Chapel Hill, NC 27514, USA

Received 6 September 1983; Revised and Accepted 8 November 1983

ABSTRACT

We have characterized a large repetitive element which has been found at seven different locations within the beta globin locus of the BALB/c mouse. This repeat has an unusual structure in that each of the different members has the same end of the element conserved while the other end terminates at a different point in each repeat member. The sequences within the repeats from the beta globin locus have homology with other repetitive families such as the MIF-1, Bam-5, R, and the BamH1 families. These were recently proposed (T. Fanning, (1983) Nucleic Acids Res. 11, 5Ø73-5Ø91) to be part of a structure with the same organization which we found in the globin locus. Probing plaques from a BALB/c genomic library with sequences derived from the repeats in the globin locus shows that virtually all of the repeats from this family are organized in a manner consistent with the proposed structure.

INTRODUCTION

A current basic problem in genetics is to explain how and why sequence homology is maintained between members of large interspersed repetitive DNA families. Because of the large copy number and the lack of identifiable (and assayable) functions, it is difficult to study these sequences using traditional genetic methods. Therefore, the usual approach has been to determine the structure and sequence of the repetitive element with the hope that these might reveal the function of the repeat or the mechanisms responsible for the sequence homology. Unfortunately, the functions of an element or the mechanisms maintaining its homology within a family are seldom immediately obvious even from sequence data. However, such data has been used to construct models. Some examples include the dispersal mechanisms of repetitive elements such as transposons, Alul repeats and small nuclear RNA pseudogenes (1,2,3). The L1Md repeat family is, as we will describe, a repetitive element with unusual structural features. We expect that these unusual features will assist in reducing the number of possible models to explain the sequence homology seen within this family.

The average organization of repetitive DNA in the genome of many organisms was first investigated by reassociation kinetic analysis of genomic DNA. By this method, repetitive sequences fall into two categories that differ in repeat length and dispersion frequency. One repeat type is very short (about 3ØØ bp) and instances are separated by single copy sequences that average 7ØØ to 11ØØ bp in length (4). The other type is much larger (greater than 1 to 2 kb) and instances are separated by very long single copy DNA segments (5).

The detailed organization of dispersed repeat families gathered from cloned members shows that these repetitive DNA sequences may be organized in a variety of ways. Repeat families may, for example, have the properties of a simple archetypic repeat, that is, each member being like the others in length and sequence. Examples of this type of repeat family have been found in every eucaryote examined. Elements like Alu (6) and repetitive sequences complimentary to small nuclear RNAs (3) have such properties. More complex families have been described where several small "discrete" repetitive elements are found mixed together into larger arrays but with no apparent conserved ordering of the smaller elements with respect to each other within the larger array. This organization has been called a "scrambled and clustered organization" (5) and examples have been described in Drosophila (5), chicken (7) and within the rabbit beta globin locus (8). Length variation exists in both of these two types of repeats. Small insertions and deletions within the "discrete" repetitive elements account for variations in the repeat length (9). The length of clustered and scrambled repetitive elements would depend upon the particular arrangement of each repeat (1Ø).

There is a set of repetitive sequences in the mouse which does not seem to fit either of these repeat types. Several groups (11, 12, 13, 14) have characterized the repetitive sequences MIF-1, Bam-5, and R which are each found in different abundances in the mouse genome. While this variation in abundance leads to the natural assumption that the repetitive elements are independent of each other, a recent proposal links these repeats into a single unit which was called the BamH1 family (24). The proposal suggests that this family has a structure where the same end of each element is conserved and the other end terminates at different points. We will show here that there is a repetitive element with sequence homology to the BamH1 repeat family in the murine beta globin locus that has exactly this structure.

METHODS

Clones

The Charon 4A clones, CE17 and CE18, that contain the embryonic gene region and the clone that contains the adult gene β2dmin, CE14, were described previously (15). The CA4 and CA11 clones were isolated from a BALB/c mouse adult liver library (a gift of Dr. Norman Arnheim, SUNY, Stony Brook) by probing with an adult beta-globin cDNA clone (15). The construction of the BALB/c liver DNA library has previously been described (15).

The V, 1.35, and U fragments were prepared by digestion of the appropriate Charon 4A clone with EcoR1 followed by electrophoresis in agarose gels. These fragments were recovered by electroelution and were ligated into M13mp2 RF DNA (16) cleaved at the single EcoR1 site. The transfection of E. coli JM1Ø1 as well as the detection, isolation, and preparation of DNA from clones carrying the appropriate insert were done as described previously (15).

Restriction Digests, Southern Blots, and Hybridization

Digestion of clones and genomic DNA with restriction enzymes, their analysis on agarose gels, and their transfer to nitrocellulose paper were carried out essentially as described previously (15). All hybridizations were to blots of a standard set of digests that includes the entire beta-globin gene region (Fig. 1).

The hybridizations with genomic DNA were done in the presence of dextran sulfate and formamide using the procedure of Wahl (17) modified as previously described (15). All other hybridizations were done without dextran sulfate and formamide as described previously (15).

Probes

M13mp2 RF DNA carrying inserts of V, 1.35, or U fragments, and EcoR1 digested genomic DNA prepared from BALB/c mouse livers were labelled with ^{32}P by nick translation as described previously (15).

Electron Microscopy

The self-annealing and visualization by electron microscopy of DNA from the CE18 clone were done as previously described (18).

RESULTS

Location of Repetitive Sequences in the Globin Region

We have looked for sequences in the beta-globin gene complex of the BALB/c mouse which are repetitive in the genome by using radiolabelled total genomic DNA probes. Many locations in the beta globin complex locus hybridized to the genomic probe, some

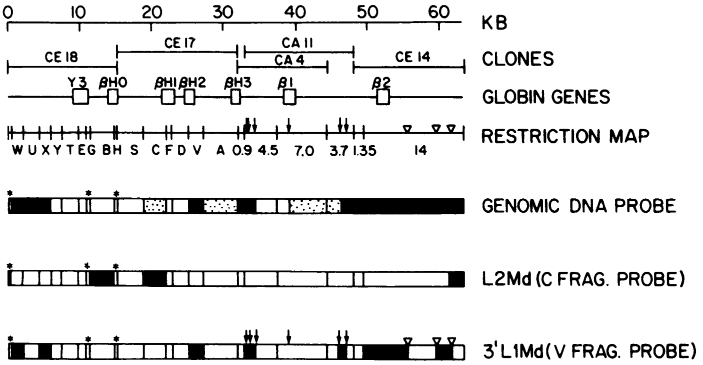

Figure 1. The location of repetitive sequences in the globin gene region. Probes made from total genomic DNA, EcoRl fragment C, and EcoRl fragment V were hybridized to blots of restriction digests of the globin gene region. The C fragment probe was two HindIII fragments (12ØØ and 9ØØ bp) from the interior of the EcoRl fragment and does not include coding sequences to the hl gene. The digests in the analysis include (clone/enzyme): CE18/EcoRl, CE17/EcoRl, CA4/EcoRl, CA4/BamHl, CA11/EcoRl, CA11/BamHl, CE14/EcoRl, CE14/HindIII, and CE14/EcoRl + HindIII. Only the restriction sites included in these digests are shown. Restriction sites are abbreviated as: + = EcoRl, ↓ = BamHl, and ▽ = HindIII. An intense hybridization signal is represented as a filled-in block and a weak hybridization signal as a dotted block. An asterisk (*) indicates which fragments were not included in the analysis.

with more intensity than others (Fig. 1). The variation in signal intensity that we see depends on homology and repeat size, as well as copy number in the probe. It has previously been estimated that this method will detect sequences that are repeated at least 5Ø times per genome (8). This was based on the size of the mammalian genome, the amount and specific activity of the probe, and the amount of DNA transferred in a Southern blot.

One of the many locations in the beta globin locus shown to be repetitive in the genome by this method had been shown to be a B1 sequence element (19). This B1 repetitive element is just 3' to the β1 gene (Fig. 1).

Two-dimensional electrophoresis blot hybridization (2Ø, C.A. Hutchison, III, personal communication) was used to indicate the number of different repetitive sequences in the globin gene region. This technique narrowed the possibilities to at most five sequences repetitive within the locus (data not shown). These results were refined by direct hybridization of purified fragments to our "standard" Southern blot. The results (Fig. 1) showed that there were in fact two distinct families repetitive

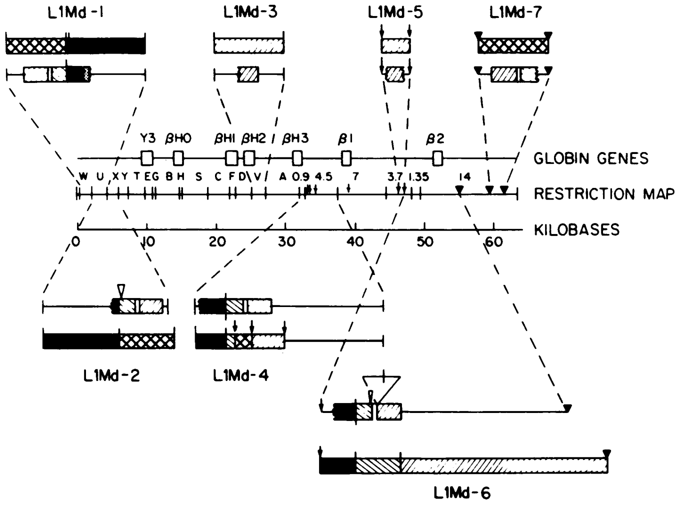

Figure 2. The location of L1Md repeats in the globin region. Probes made from the V, 1.35, and the U fragment were hybridized to blots of the standard digests of the globin gene region. The three probes do not hybridize to one another, but hybridize to adjacent positions at other locations in the globin region. This suggested that the three probes contained portions of a larger repeat structure. These seven positions are expanded and the hybridization results, as well as our interpretation of these results (based on sequence analysis of the three probes (see Results)) are presented. The actual location of L1Md-7 within the 2 kb HindIII fragment, and the location and orientation of L1Md-5 within the 900 bp BamHl fragment are not presently known. Therefore, the interpretation of the hybridization data for these two repeats is centered within the fragments. The location of homology to the V fragment probe is indicated by right slanted hatching, homology to the 1.35 probe by right slanted hatching, and homology to the U fragment probe by a filled-in area. Homology to both the V and 1.35 fragment probes is represented by cross hatching. The location of restriction sites are marked as follows: EcoRl = +, BamHl = ⊥, and HindIII = ▼.

within the murine beta-globin locus. One family, which we call L1Md, was found at seven different locations (Figs. 1 and 2) and another, L2Md, was present at three different locations (Fig. 1). Both of these families are repetitive outside of the beta-globin locus as well. Hence we find no repeat families, except for the globin genes, confined to the beta globin complex locus.

The L1Md Repeat

The L1Md repeat family was initially defined in the beta globin locus by finding sequences at seven locations which would hybridize to a probe prepared from the EcoRl V fragment. Five of these locations were found to have a common sequence adjacent to

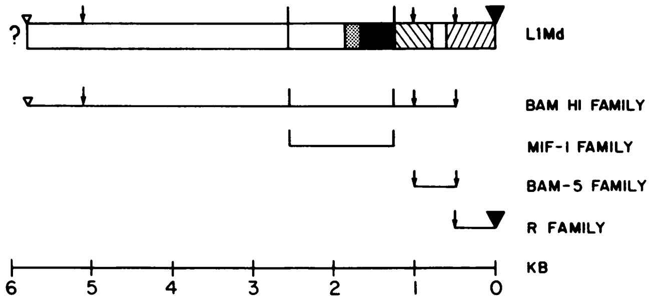

Figure 3. The relationship of probes to other mouse repeats. Probes made to the V fragment, the 1.35 fragment, and to the U fragment were hybridized to nitrocellulose blots of mouse genomic DNA that was digested with EcoRl or BamHl and electrophoresed in agarose gels. All three probes hybridize to a smear of fragments of all sizes, but each also hybridizes strongly to one or more bands that stain intensely with ethidium bromide. The V fragment hybridizes to the Ø.5 kb BamHl band. The 1.35 probe hybridizes to the Ø.5 kb BamHl and the 4.1 kb BamHl bands. The U probe hybridizes to the 4.1 kb BamHl band and the 1.3 kb EcoRl band. These bands are portions of previously described repeats including the BamHl family, the MIF-1 family, and the Bam-5 family. The positions and lengths of the L1Md repeat sequence in each of the three probes (determined by DNA sequencing, manuscript in preparation) are also shown. The L1Md repetitive sequence found in the V fragment is represented by right-slanted hatching, in the 1.35 fragment by left-slanted hatching, and in the U fragment by the filled-in area. The uncertainty of the endpoint of the repeat in the U fragment is indicated by the dotted area. Restriction sites are indicated as follows: EcoRl = ⊥, BamHl = ↓, and Kpnl = ∇. The large filled-in triangle indicates the position of the conserved 5'-TAATAAAAAA-3' sequence.

them, defined by using the 1.35 EcoRl fragment as a probe. Four of these five locations were also found to have another common sequence adjacent to them defined by using the EcoRl fragment U as a probe (Fig. 2). The order of these three sequences, defined by homology to EcoRl V, EcoRl 1.35, and EcoRl U fragment probes was the same in each of the four members of the L1Md repeat family which have homology to all three probes (Fig. 2).

This suggested that each of the probes contained non-overlapping portions of a larger repetitive sequence. Each of these probes used to define the distribution of repetitive sequences contained a portion of the canonical L1Md repeat (Fig. 3). We now know from sequence data (manuscript in preparation) that the V fragment contains 65Ø bp of the L1Md repeat. The 1.35 fragment contains 5ØØ bp of the L1Md repeat and 85Ø bases of a unique sequence inserted into L1Md-6 (see below). Finally, the U

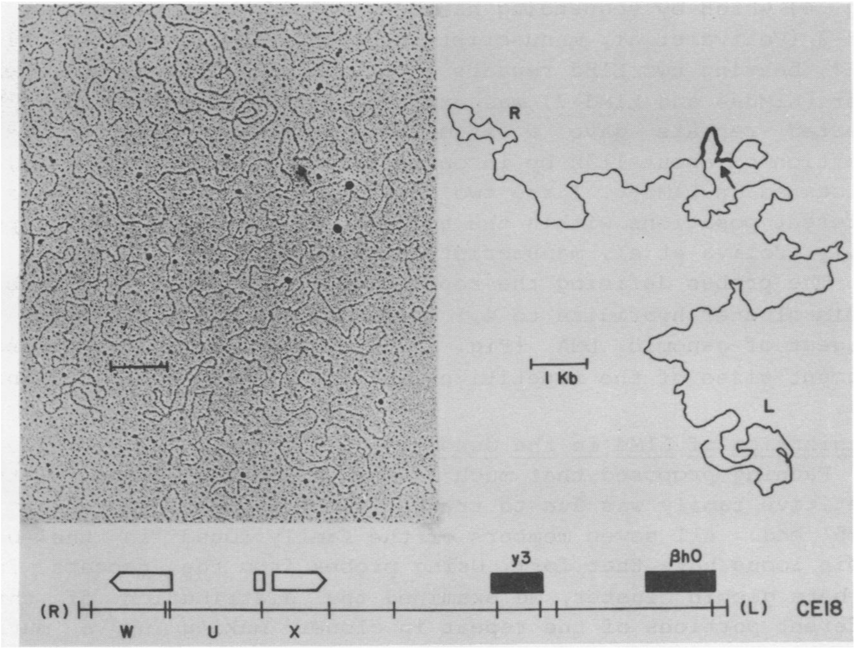

Figure 4. L1Md-1 and L1Md-2 form an inverted repeat in CE18. The electron micrograph and the interpretation of the molecule derived by melting and self-annealing the CE18 clone is shown. The length of the stem is 1500 +/- 200 bp, and the length of the loop is 2100 +/- 300 bp. The measurements were made on 17 molecules. The arrow points to a small knob which is interpreted as an single stranded region due to an insertion in one of the repeats. The knob is detectable at the same location in 9 of the 17 molecules measured. Since it is just detectable, we suspected that the insertion was about 200 bp in length. The DNA sequence of both of these MDR1 repeats is almost complete. That analysis confirms the EM measurements of the stem length and the location of a 186 bp insertion in L1Md-2.

fragment contains portions of two truncated L1Md repeats. It has 600-800 bp from the L1Md repeat and about 1400 bases of sequence unrelated to L1Md.

Two of the L1Md repeats (L1Md-2 and L1Md-6) were found to contain insertions. A clone (CE18) containing the 5' end of the beta-globin locus and two members of the L1Md family (L1Md-1 and L1Md-2) inverted with respect to one another was melted and allowed to self-anneal. A stem and loop structure was seen with the electron microscope at a position coincident with the location of L1Md-1 and L1Md-2. One strand of that stem was consistently interrupted by a 200 bp insertion/deletion loop

(Fig. 4) which by sequencing has been mapped as an insertion in L1Md-2 (Voliva et al, manuscript in preparation). Another clone (CE14) bearing two L1Md repeats inverted with respect to each other (L1Md-6 and L1Md-7) was examined in the same fashion. The inverted repeats gave a stem-loop structure with a large insertion of about 1100 bp in one strand previously shown (18) to be located in L1Md-6. These two insertions were found to occur at different positions within the two members of the L1Md repeat family (Voliva et al, manuscript in preparation).

The probes defining the repetitive sequences in the beta globin cluster hybridize to 4.1 kb and 0.5 kb BamH1 fragments in a digest of genomic DNA (Fig. 3). These are characteristic fragment sizes of the repetitive family called BamH1 by Fanning (24).

Organization of L1Md in the Genome

Fanning proposed that much of the variation seen in this repetitive family was due to truncation of a canonical repeat at its 5' end. All seven members of the family found in the beta globin locus have that form. Using probes from the repeats from the beta globin cluster, we examined the distribution of three different portions of the repeat in clones making up a murine genomic library. Replica transfers of plaques from a BALB/c library of EcoR1 partial digest fragments were probed with V, 1.35, and U sequences and individual plaques scored for their sequence content.

A model of the repeat family, each member of which is truncated from only one end would predict that there would be few, if any, clones bearing the sequences from the conserved end of the repeat (V probe) and sequences further 5' (U probe) which do not contain the sequences in between (1.35 probe). This is what was found, as only 0.2% of the clones contained repeats that fall into this class (Fig. 5). A less stringent analysis would be the fraction of plaques where one found portions of the repeat in the absence of the sequences from the conserved end (V). The model predicts that there would be few if any such cases. Only 6.6% of the plaques contained repeats that fall into this category. The remainder of the plaques that hybridize to any of the probes met the expectation of a canonical repeat truncated from one end. Finally, the relative abundance of the three sequences defined by our probes (Fig. 5) was consistent with truncation from one end of the repeat. The sequences in the V probe (which contains the portion of the repeat nearest the conserved end) were most abundant, followed by 1.35 and then the U probe. All of these results are consistent with the model that this repeat family is organized as a canonical repeat each member

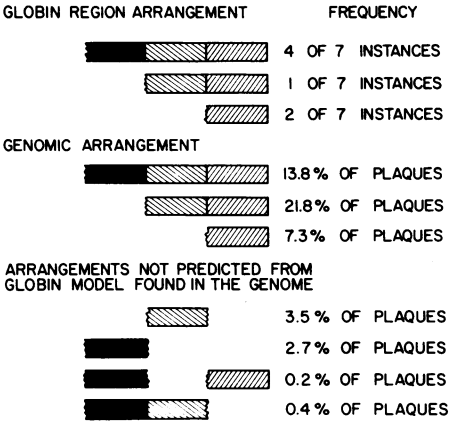

Figure 5. The conserved arrangement of L1Md. Probes made from the V fragment, the 1.35 fragment, and the 1.3Mml clone (22) (which is homologous to the U fragment) were hybridized to nitrocellulose filters of mouse DNA library plaques. Plaques were scored according to which probe they hybridized to, and the results are represented graphically below. These results are compared to the arrangement of the portions found at seven locations in the globin gene region. Four plates with about 400 plaques each were scored. The order of the three portions in both the genome and in the globin gene region is conserved as (from right to left): "V homology- 1.35 homology- U homology". The relative frequencies of the three probes in the genome suggest that many instances of the repeat are truncated from one end. The relative size of each portion of the L1Md repeat contained in each fragment is not implied by the size of the blocks in the diagram.

of which shares a common 3' end and is terminated at its 5' end at points random with respect to the other members.

The Conserved Endpoint of L1Md

The 3' end of this repeat, which was called the R family, has been defined by sequence analysis (11). Our sequence of L1Md-3 (found in the V fragment) shares extensive homology (88%, counting "N" as a mismatch) with the R repeat family consensus sequence (Fig. 6). Sequence homology between all members of this family ends at the 3' end with 5'-...TAATAAAAAA-3' which is followed by an "A-rich" region. Sequences in this 3' region from other genomic locations (21,24) also loose their homology to each other at this same point.

All instances of the repeat in the globin gene region and most instances in the genome terminate at or near this 5'-TAATAAAAAA-3' sequence. Each instance of the repeat in the beta globin gene region and almost all of the plaques from the

```
         .         .         .         .        50         .         .         .         .       100
GGATCCATCCCATAATTAGCCTCCAAACGATGACACCATTGCATACACTAGCAAGCGTTTGAAGCAAGGACCATGATATAGCTGTCTCTTGTGAGACTAG  L1MD-3
|||||||||||||||| |||| ||||||    |||| ||||||||  | |  |||  ||||  | |||||   |||||||||||||||||||||||||| 
GGATCCATCCCATAATCAGCCACCAAACCCANACACTATTGCATATNCNAANAAGATTTTGCTGAAAGGANNCTGATATAGCTGTCTCTTGTGAGACTAT  R FAMILY CONSENSUS

         .         .         .         .       150         .         .         .         .       200
GCCGGGGCCTAGCAAACACAGAAGTGGATGCTCACAGTCAACTATTAGATGGATCACA..GGGCTCC.AATGGAGGAGCTAGAGAAAGTATCCAAGGAGC  L1MD-3
||| | |||||||||| ||||||||||||||||||||||| ||||| ||||| |||||  |||| || |||||||||||||||||||||| |||||||||
GCCAGNGCCTAGCAAATACAGAAGTGGATGCTCACAGTCANCTATTGGATGGGTCACACAGGGCCCCCAATGGAGGAGCTAGAGAAAGTACCCAAGGAGC  R FAMILY CONSENSUS

         .         .         .         .       250         .         .         .         .       300
TAAAGAGATCTGCAACTCTGTAGGTGC...AACATTATGAACTAACCAGTACCCC.AGAGCTCTTGTCTCTAGCTGCATATGTATCAAAAGATGGCCTAG  L1MD-3
||||| | |||||||| || ||||||    |||| |||||||||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||||
TAAAGGGNTCTGCAACCCTATAGGTGGAACAACAATATGAACTAACCAGTACCCCCAGAGCTCNTGTCTCTAGCTGCATATGTATCAAAAGATGGCCTAG  R FAMILY CONSENSUS

         .         .         .         .       350         .         .         .         .       400
TCGGCCATCATTGGAAAGAGAGGCCCATTGGACAGGCAAACTTTATATGCCCCAGTACAGGGGAACGCCAGGGCCAAAAAATGGGAATGGGTGGGTAGGG  L1MD-3
|||||||||| ||| ||||||||||||||||||  | ||||||||||||||||||||||||||||||||||||||||||| ||||| |||||||||||||
TCGGCCATCACTGGGAAGAGAGGCCCATTGGACTTGTAAACTTTATATGCCCCAGTACAGGGGAACGCCAGGGCCAAAAAGTGGGAGTGGGTGGGTAGGG  R FAMILY CONSENSUS

         .         .         .         .       450         .         .         .         .       500
GAGTGTGGTGGGGGGAGAGTGTGGGAGACTTTTGGGATAGCATTGGAAATGTAATTGAGGAAAATACGTGATAAAAAA                        L1MD-3
||||  ||   |||||| || |||| |||||||||||||||||| ||||||||| |||||||||||  | ||||||||
GAGTNGGGA..GGGGAGGGTATGGGGGACTTTTGGGATAGCATTTGAAATGTAAATGAGGAAAATATCTAATAAAAAA                        R FAMILY CONSENSUS
```

Figure 6. Sequence homology between L1Md-3 and the R family. The sequence from a portion of the V fragment and the consensus sequence for the R repeat family (11) are shown aligned to display the homology betweeen them. Dots (.) are added to create gaps and "N" indicates a position for which no consensus sequence could be assigned. The complete nucleotide sequence of L1Md-3 will be published elsewhere (Voliva et al., manuscript in preparation).

genomic library that hybridized to our probes have homology to the V probe carrying these 3' sequences (Fig. 5). This is consistent with the model that each member of this family terminates at or near this same 3' sequence.

DISCUSSION

This repeat family is a dispersed highly repetitive sequence at least 7 kb in length which is conserved at the 3' end and terminated at many different points in the 5' portion of the sequence. Various portions of this repeat have been described and named as separate entities. These include the R family (11), the BamH1 family (12), the MIF-1 family (13), and the Bam-5 family (14). Fanning suggested that all of these sequences are part of a single repeat and called that repeat the BamH1 family. This family is present in many species each with its own distinctive restriction pattern (Frank Burton, manuscript in preparation). We propose here a name for the family, that is L1, which is independent of the predominant restriction fragments of each species. L1 can be used with each species to carry the connotation of shared sequence. The L is drawn from the designation of Singer (25) for long interspersed repetitive sequences as LINES. Species designators can be added as L1Md

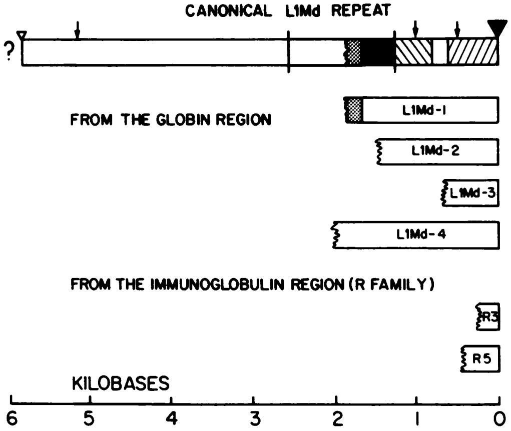

Figure 7. The random truncation endpoint of six L1Md repeats. Enough sequence has been determined for four instances of L1Md in the globin gene region (manuscript in preparation) and for two instances of R family repeats in the immunoglobulin gene region (11) that the location of the truncated endpoint can be accurately determined. Each of the six repeats ends at a different position relative to the other repeats. The shaded area at the truncation of L1Md-1 indicates the uncertainty of the location of that endpoint. It must lie in a 200 bp unsequenced region since no homology can be found to L1Md-4 when sequence resumes across the gap.

(for Mus domesticus) and trivial lab names can be appended as L1Md-4.

The L1Md structure is unusual amongst characterized repeats in that one end of the element is virtually always conserved while the other end is almost always truncated by various amounts. Each of the seven instances of L1Md within the beta-globin locus (Hbb-d) is truncated at a different point. Four of these 5' endpoints have been determined by DNA sequencing (manuscript in preparation) and none terminate at the same point (Fig. 7). The 5' endpoints of two other instances of this repeat family within the immunoglobulin gene region have been sequenced as R family repeats and those endpoints are not the same (11).

The frequency that various portions of L1Md are found in the genome suggests that the bulk of the repeats in the genome are also truncated at the 5' end. Portions of the repetitive element which are closest to the 3' end of L1Md are most abundant (Fig.5). Abundance estimates of the independently characterized portions of L1Md give the same picture. The R family sequences are present in about 100,000 copies per haploid genome (11), the Bam-5 sequences in 50,000 copies (13), and the MIF-1 family in

20,000 copies (13).

A small number of plaques in the BALB/c liver DNA library had hybridization patterns inconsistent with this canonical arrangement. The small number of plaques which had L1Md sequences, but did not include the 3' conserved end (6.6% of the plaques), are most reasonably explained by surmising that these repeats were severed from the conserved sequence when they were cloned. The plaques with sequences missing from within L1Md are also inconsistent with this model. This arrangment occurs in only 0.4% of the hybridizing plaques and probably represent deletions within an L1Md repeat.

The 3' endpoint is conserved at or near the same position in all the L1Md repeats. The sequence comparison between L1Md-3 and the R family consensus sequence shows that sequence homology ends at the sequence 5'-TAATAAAAAA-3', followed by a "A rich" region (Fig.6). This endpoint is also shared with an instance of the repeat that flanks sequences homologous to an Intercisternal A Particle (21) and with instances of the repeat randomly cloned from genomic DNA (24). In addition, all instances of the L1Md repeat within the globin gene region (Fig. 2) and almost all instances in the genome (Fig. 5) share homology to a probe that contains the 3' end of the repeat. This suggests that all the repeats end at or near the 5'-TAATAAAAAA-3'.

An "A rich" region like that found at the conserved endpoint of the L1Md repeat is found in other repetitive sequences, including the Alu-1 family of repeats (6), repetitive sequences homologous to some families of small nuclear RNAs (3), and retroviral proviral sequences (23). This "A rich" region is thought to reflect the participation of a poly-adenylated RNA in the dispersion of those sequences to new locations (2,3). Since L1Md sequences appear to be transcribed (14), transcripts of L1Md could be involved in their dispersal as well. A dispersal mechanism involving cDNA copies of these transcripts would explain the random 5' truncation of the repeat.

ACKNOWLEDGEMENTS-The author wishes to thank S.C. Hardies and S.L. Martin for helpful reading of the manuscript and discussion. This research was supported by Public Health Service Grants AI08998 and GM21313 from the National Institutes of Health. M.B.C. was supported by a grant from the Winchester Foundation. CFV was supported by a predoctoral training grant GM07092 to the University of North Carolina from the NIH.

*Present address: Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309, USA

REFERENCES

1. Galas, D.J., and Chandler, M. (1981) Proc. Natl. Acad. Sci. USA 78, 4858-4862.
2. Jagadeeswaran, P., Forget, B.G., and Weissman, S.M. (1981) Cell 26, 141-142.
3. Van Arsdell, S.W., Denison, R.A., Bernstein, L.B., Weiner, A.M., Manser, T., and Gesteland, R.F. (1981) Cell 26, 11-17.
4. Davidson, E.H., and Britten, R.J. (1973) Quart. Rev. Biol. 48, 565-613.
5. Wensink, P.C., Tabata, S., and Pachl, C. (1979) Cell 18, 1231-1246.
6. Jelinek, W.R. and Schmid, C.W. (1982) Ann. Rev. Biochem. 51, 813-44.
7. Musti, A.M., Sobieski, D.A., Chen, B.B., and Eden, F.C. (1981) Biochem. 20, 2989-2999.
8. Shen, C.-K. J., and Maniatis, T. (1980) Cell 19, 379-391.
9. Haynes, S.R., Toomey, T.P., Leinwand, L., Jelinek, W (1981) Mol. Cell. Biol. 1, 573-583.
10. Pearson, W.R., Mukai, T., and Morrow, J.F. (1981) J. Biol. Chem. 256, 4033-4041.
11. Gebhard, W., Meitlinger, T., Hochtl, J. and Zachau, H.G. (1982) J. Mol. Biol. 157, 453-471.
12. Meunier-Rotival, M., Soriano, P., Cuny, G., Strauss, F., and Bernardi, G. (1982) Proc. Natl. Acad. Sci. USA 79, 355-359.
13. Brown, S.D.M., and Dover, G. (1981) J. Mol. Biol. 1500, 441-446.
14. Fanning, T.G. (1982) Nucleic Acids Res. 10, 5003-5013.
15. Jahn, C.J., Hutchison, C.A., III, Phillips, S.J., Weaver, S., Haigwood, N.L., Voliva, C.F., and Edgell, M.H. (1980) Cell 21, 159-168.
16. Gronenborn, B., and Messing, J. (1978) Nature 272, 375-377.
17. Wahl, G.M., Stern, M., and Stark, G.R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683-3687.
18. Weaver, S., Comer, M.B., Jahn, C.L., Hutchison, C.A., III, and Edgell, M.H. (1981) Cell 24, 403-411.
19. Coggins, L.W., Vass, J.K., Stinson, M.A., Lanyon, W.G., and Paul, J. (1982) Gene 17, 113-116.
20. Sato, S., Hutchison, C.A.,III, and Harris, I. (1977) Proc. Nat. Acad. Sci. USA 74, 542-546.
21. Leuders, K., and Paterson, B.M. (1982) Nucleic Acids Res. 10. 7715-7729.
22. Heller, R., and Arnheim, N. (1980) Nucleic Acids Res. 8, 5031-5042.
23. Temin, H. (1980) Cell 21, 599-600.
24. Fanning, T.G. (1983) Nucleic Acids Res. 11, 5073-5091.
25. Singer, M.F. (1982) Cell 28, 433-434.