Volume 13 Number 14 1985 Nucleic Acids Research

Chromosomal rearrangements associated with LINE elements in the mouse genome

Susan Shyman and Steven Weaver

Laboratory of Cell, Molecular and Developmental Biology, Department of Biological Sciences, University of Illinois, Box 4348, Chicago, IL 60680, USA

Received 6 May 1985; Revised 10 June 1985; Accepted 19 June 1985

ABSTRACT

Two segments of DNA that have apparently inserted in the interval between the two adult β-globin genes in BALB/c (Hbb^d haplotype) but not in C57Bl/10 (Hbb^s haplotype) mouse strains have been described (1). These putative insertions, each about 1000 bp in length, mapped near a repetitive element. To determine the precise position of these alleged insertions, their target sites, and the nature of their boundaries, we cloned and sequenced the appropriate regions of both chromosomes. One of the two segments is not an insertion but rather a region between two independently integrated L1 repetitive elements (LINEs)(2), one in Hbb^d and the other in the Hbb^s chromosome. The other segment is an insertion of 940 bp which is located within the L1 element in the Hbb^d chromosome. This insert is unusual in that it exists in only one copy in the BALB/c genome.

INTRODUCTION

A major mechanism by which chromosomes diverge is through alterations in chromosome structure. Chromosomal rearrangements include inversions, translocations, duplications, and deletions. Rearrangements of a deletional as well as an insertional nature are often mediated by mobile elements. One example of this is the *his4* gene in yeast, where deletions are promoted by the TY1 transposable element (3).

The short direct repeats that border transposable elements are illustrative of the insertion process. The TY1 element of yeast and the copia element of *Drosophila* both produce 5 bp direct repeats (4,5). Integration of retroviral proviruses also results in short direct repeats, 4-6 bp long (6). This feature is also common to prokaryotic transposable elements (7).

Some elements contain terminal inverted repeats that constitute part of the element itself. The P element of *Drosophila* for example, is bounded by 31 bp inverted repeats; these are necessary for transposition (8). Foldback (FB) elements, also found in *Drosophila* contain terminal inverted repeats of variable length (9).

Dispersal of families of highly repetitive elements represent another

© IRL Press Limited, Oxford, England.

source of chromosomal variability. Interspersed throughout the mammalian genomes in 10^4 - 10^5 copies, these repeats have been divided into two classes: SINEs, or short interspersed repetitive elements, are typically less than 500 bp in length, whereas LINEs, or long interspersed repetitive elements, range in length from a few hundred base pairs up to 7 kb (2). These elements also create short direct repeats of about 7-20 bp upon insertion, but lack terminal inverted repeats (10). Characteristic of both classes is an A-rich region found at the end referred to as the 3' end. LINE family members are usually truncated at variable points, at the other "non-poly A" or 5' ends. SINEs, on the other hand, have a constant 3' and 5' end.

A clearer understanding of the evolution of the mammalian genome can be gained through knowledge of chromosomal differences within a species, and between closely related species or subspecies. We have investigated chromosomal rearrangements between the allelic β-globin gene clusters of BALB/c ($\underline{Hbb}^d$ haplotype) and C57BL/10 ($\underline{Hbb}^s$ haplotype) mouse strains. These rearrangements create length differences between the haplotypes in the interval between the β1 and β2 loci which encode adult β-globin. Inconsistencies in restriction maps and measurements from electron micrographs of snap-back and heteroduplexed molecules involving genomic clones indicated that two regions of DNA were present in the $\underline{Hbb}^d$ haplotype but not at the corresponding sites in the $\underline{Hbb}^s$ haplotype (1). These two regions, apparently insertions, were designated as the A and B segments. They are in the vicinity of a repetitive element now known to be a member of the L1 family of LINEs.

To determine the precise locations and the nature of the A and B segments, we cloned and sequenced their boundaries from the $\underline{Hbb}^d$ haplotype as well as the equivalent positions in the $\underline{Hbb}^s$ haplotype. We show that only one of these sequences, the A segment, is consistent with an insertion event. The B segment, on the other hand, represents the interval between the integration sites for two distinct LINE elements, one in the $\underline{Hbb}^s$ haplotype and one in the $\underline{Hbb}^d$ haplotype.

MATERIALS AND METHODS

Clones

To determine the location of the A and B segments and their putative target sites through restriction map analysis, the 510 bp (R) and 720 bp (S) BamHI fragments were transferred from the C57BL/10 genomic clone BA3 (1) into an M13mp8 vector. The 2.3 kb BamHI-EcoRI fragment (U) of BALB/c was isolated from a pBR322 subclone (obtained from M. Garrick) and transferred into M13mp8

and M13mp9 in opposite orientations.

Cloned fragments used in sequencing the ends of the A and B segments and their putative target sites were generated by further subcloning and deletions. Deletions in fragments R and U were made by sequential digestion with Exonuclease III and S1 nuclease (11,12). M13 phage clones were linearized with a variety of enzymes including HindIII, SacI, and EcoRI and then digested with Exonuclease III to create single stranded deletions. The single stranded ends were trimmed with S1 nuclease and religated. The extent of the deletions obtained were characterized by agarose gel electrophoresis. Subclones of fragment U were also generated by isolation of the 2.3 kb BamHI-EcoRI fragment (U) from pBR322 vector using low melting agarose (13). The insert was then digested with HincII or HinfI and ligated into the HincII site of an M13mp18 vector. HinfI fragments were made blunt by S1 nuclease before ligation.

Sequencing

All sequencing reactions were done using single stranded M13 recombinant phages and the dideoxy chain termination method of Sanger et al. (14). Sequences were determined on either both strands of the DNA or sequenced in multiple experiments.

Nomenclature

The adult globin genes are described using the locus designations $\beta 1$ or $\beta 2$ with superscripts s or d referring to the haplotype. Previously $\beta 1^{s}$ has been referred to as β^{s}, $\beta 1^{d}$ as β^{d} major, $\beta 2^{s}$ as β^{t}, and $\beta 2^{d}$ as β^{d} minor.

In this paper we refer to the LINE family as L1Md, following the nomenclature introduced by Voliva et al. (15). L1 refers to LINE family 1. Md identifies the species - Mus domesticus. Digits added to the designation distinguish between different family members. Referred to in this paper are L1Md-6 in BALB/c and L1Md-9 in C57Bl/10.

RESULTS

There are at least eight L1 elements interspersed throughout the β-globin cluster of the mouse Hbb^{d} haplotype (15,16). The A and B segments are in the vicinity of one of these elements, L1Md-6, (15) located 5' of the β2 gene (Fig. 1B). On the basis of electron micrographic measurements, Weaver et al. concluded that the A segment, measured as about 1100 bp, is situated within L1Md-6 while the B segment, roughly 1500 bp, lies between the L1 and the $\beta 2^{d}$ gene (1).

The approximate location of one boundary of the A segment was deduced from

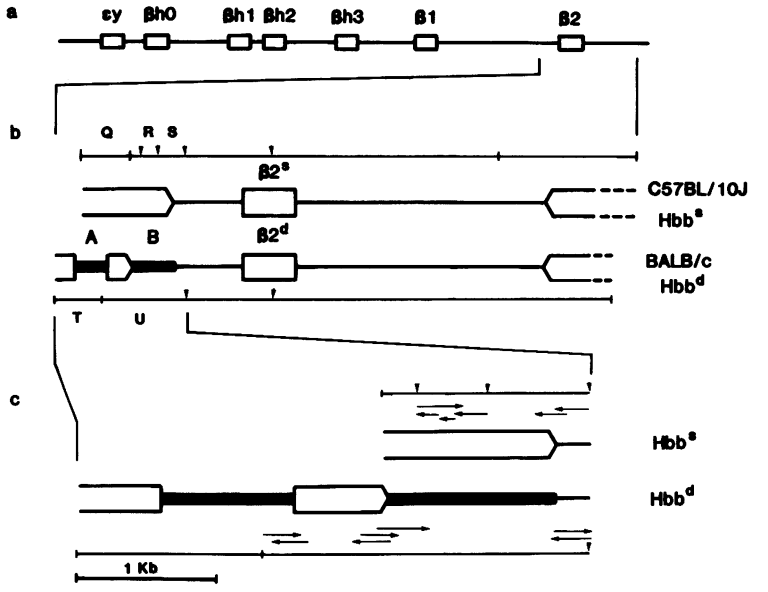

Figure 1. (a) The mouse β-globin gene cluster is shown. It includes the embryonic β-like genes, εy, βh0, and βh1, the pseudogenes, βh2 and βh3, and the adult β-globin genes, β1 and β2. (b) L1 repetitive elements indicated by large open arrows are located on either side of the β2 gene in both the Hbb^s and Hbb^d haplotypes. Dashed lines indicate uncertainty of length. Within and bordering L1Md-6 the A and B segments are represented as heavy black lines. EcoRI sites are shown by vertical ticks, BamHI as arrows. (c) The sequencing strategy to determine the location of the ends of the A and B segments as well as the corresponding positions in the Hbb^s haplotype. The BamHI sites at the right of each haplotype are homologous, as are the left EcoRI sites. Sequencing reactions were done according to the dideoxy chain termination method. Sequences not present in both orientations were determined in multiple experiments. Fragment T was previously sequenced by C. Voliva (personal communication).

electron micrographs of heteroduplexed molecules (1). Sequence at the EcoRI site on the right of fragment Q was homologous to sequence at the EcoRI site on the left of fragment T (Fig. 1B,1C). Measurements of these heteroduplexes placed the left boundary of the A segment within fragment T, about 600 bp from the left EcoRI site. These results led us to believe that the other end of the A segment would be located within fragment U and that the target site in the Hbb^s haplotype would lie within fragment R (Fig. 1B,1C).

We sequenced fragment R as well as the left end of fragment U (data not shown) (Fig. 1C). Alignment of the sequences of fragments R and U and that of fragment T which was previously sequenced, (C. Voliva, personal communication) (Fig. 2) defined the extent of the A segment. The A segment is 940 bp long,

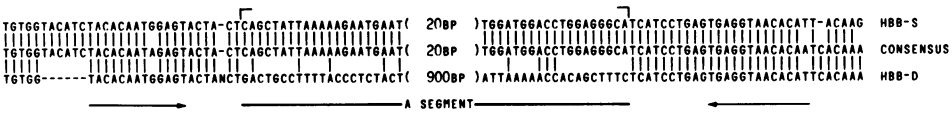

Figure 2. Part of the A segment and its surrounding sequence is aligned with the L1 consensus sequence as well as with its putative target site within the L1 element in the $\underline{Hbb}^{s}$ haplotype. Arrows underline 13 bp repeats forming mirror symmetry. Within the brackets are the 60 bases of the L1 sequence missing from the $\underline{Hbb}^{d}$ haplotype. Gaps are added to increase homology.

extending from a point 745 bp from the right EcoRI site within fragment T to a point 195 bp from the EcoRI site within fragment U (Fig. 1B). The point in fragment R corresponding to the position of the A segment is located 170 bp from the left BamHI site (Fig. 1B).

Known transposable elements are bordered by characteristic features such as inverted repeats of roughly 20-40 bp or short direct repeats, 3-12 bp, generated by duplication of bases at the target site (7). Neither inverted repeats nor direct repeats are found at the borders of the A segment. Alignment of the ends of the A segment with the corresponding sequences in the $\underline{Hbb}^{s}$ haplotype reveals a 60 bp deletion within the $\underline{Hbb}^{d}$ haplotype at the site of insertion (Fig. 2). We are confident that the missing bases represent a deletion in the $\underline{Hbb}^{d}$ chromosome rather than an insertion into the $\underline{Hbb}^{s}$ chromosome since the 60 bases are part of the L1 consensus sequence (16). A remarkable sequence feature of uncertain significance is present 7-10 bp outside of either end of the A segment. Two 13 bp sequences, consisting of 5' TACACAATGGAGT 3' and 5' TGAGGTAACACAT 3', form perfect mirror images of one another (Fig. 2). A computer analysis (unpublished results) revealed no other such mirror images of this length anywhere in the L1 consensus sequence.

The boundaries of the B region and the corresponding areas in the $\underline{Hbb}^{s}$ haplotype were defined according to the sequencing strategy presented in Figure 1C. The BamHI sites on the right of fragment S and fragment U were found to be located in homologous sequence environments, demonstrating that the B region lay to the left, entirely within fragment U (Fig. 1B). The homology between the two haplotypes ended about 240 bp to the left of the BamHI sites, thereby defining one end of the B segment in the $\underline{Hbb}^{d}$ haplotype.

Since electron micrographic measurements indicated that the B segment was 1500 bp in length (1) we anticipated that its other end would be located about 1000 bp from the EcoRI site in fragment U (Fig. 1B). Deletions of fragment U were made (see Materials and Methods) and appropriate clones were sequenced (Fig. 1C) and compared to the corresponding region of the $\underline{Hbb}^{s}$ haplotype (Fig.

1B). The left boundary is located 935 bp from the EcoRI site, at a point immediately flanking the poly-A tract of L1Md-6 (Fig. 1B). The entire B segment is approximately 1200 bp long.

As was the case with the A segment, the boundaries of the B segment showed no direct repeats representative of the integration process, nor were there inverted repeats. However, alignment of the ends of the B segment with its corresponding region in the Hbb^s haplotype demonstrated that the location of both of its ends corresponded precisely to the 3' poly A tract of an L1 element. The left end of the B segment borders L1Md-6 while the right end is in the position corresponding to the 3' end of L1Md-9 (Fig. 4).

DISCUSSION

In general, L1 elements share a conserved 3' end that terminates in a poly A tract and characteristically are truncated at one end, the 5' end. They range in length from a few hundred base pairs to a maximum of about 6.5 kb. The L1Md-9 element in the Hbb^s haplotype discussed in this paper is roughly 6 kb. The 5' boundary of L1Md-9 was confirmed by the existence of 13 bp imperfect direct repeats at both the putative 5' boundary and adjacent to the 3' poly A tract (Fig. 3). Interestingly, the DNA sequence immediately adjacent to the 5' end of L1Md-9 is identical to the right end of the B segment (Fig. 3). That is, at least part of the region we have designated as B, originally

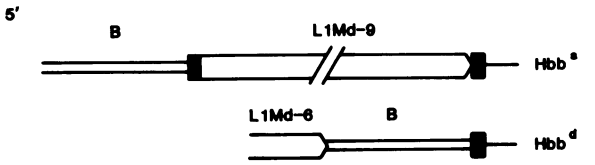

Figure 3. Alignment of the 5' and 3' ends of L1Md-9 in the Hbb^s haplotype with its target site in the Hbb^d haplotype. Within the open and black boxes, in the sequence and schematic diagram respectively, are the 13 bp imperfect direct repeats and the target site. The sequence 5' of the direct repeats in L1Md-9 is homologous to sequence 5' of the target site in the Hbb^d haplotype and is also part of the B segment, (open lines in the lower diagram). Roughly 4 kb has been omitted from L1Md-9 which is illustrated by slashed lines. Number designation for the L1 element in the Hbb^s haplotype as well as the sequence of the 5' end was obtained from Burton et al. (20).

thought of as an insertion into the Hbb^d chromosome just 3' to the L1 element, is also present at the 5' end of the L1 in the Hbb^s chromosome (Fig. 4). Additional information as to the presence of the B element 5' of L1Md-9 comes from hybridization experiments using a restriction fragment which maps entirely within the B segment (17). This probe is homologous to a single restriction fragment in the genomes of both C57BL/10 and BALB/c. With C57BL/10 genomic DNA, this probe hybridized to a 5.5 kb BamHI fragment, the same fragment within which the 5' end of L1Md-9 is located (Fig. 4). Finally, the 13 bp target site for this L1 was located within the Hbb^d chromosome and consisted of the point we had defined to be the right boundary of B.

These observations indicate that the B region, which had been characterized to be an insert based on heteroduplex analyses, is in fact a single copy DNA sequence common to both the Hbb^s and Hbb^d haplotypes. The B segment is simply the interval between two homologous but independently integrated members of the L1 family.

A tentative extension of this conclusion is that the B region is a segment of a somewhat larger chromosomal neighborhood that is a preferred site for the integration of L1 elements. This neighborhood has served as the integration site for at least two repetitive elements, L1Md-6 in the Hbb^d chromosome and L1Md-9 in the Hbb^s chromosome. Additional hybridization data supports this notion. A probe specific for the 3' end of the L1 family hybridizes to a cloned segment of the Hbb^s chromosome, between the 3' end of the β1 locus and the left end of L1Md-9 (data not shown). Although the exact location of this homology is not known, it is near the position of the B segment. Also in this region is a SINE, a member of the B1 (mouse Alu-equivalent) family, located 2.8kb downstream of the β1 locus in the Hbb^d haplotype, to the left of L1Md-6 (18).

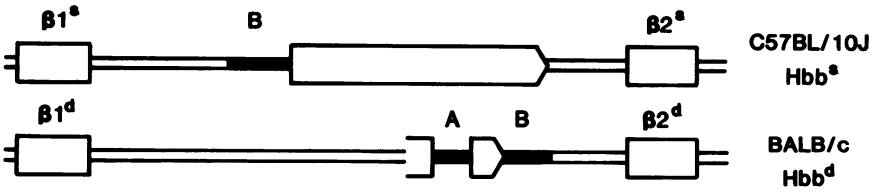

Figure 4. Location of the A and B segments relative to L1 elements. Large arrows pointing in the direction of the 3' A-rich regions are L1Md-9 in the Hbb^s haplotype and L1Md-6 in the Hbb^d haplotype. Both are bordered on opposite ends by the B segment, represented by heavy black lines. The A segment interrupts L1Md-6. The interval between the β1 genes and the L1 elements in both haplotypes contain other repetitive elements not shown.

Regions bordering the 3' poly A tracts of both L1 elements discussed in this paper are A-rich but not homologous. Sequences of five L1 elements from the embryonic region in the Hbb^{d} haplotype, ranging in length from 180-2095 bp, were recently published (16). Each is bordered by different direct repeats, none of which is homologous to the direct repeats presented in this paper. The insertions of L1Md-6 and L1Md-9 are consequently separate, non-sequence specific events that have occurred since the divergence of the Hbb^{d} and Hbb^{s} haplotypes.

The region of DNA designated as the A segment interrupts the L1Md-6 element at a point about 700 bp from its 3' A rich region and results in a net insertion of 940 nucleotides with respect to the L1 consensus sequence. Comparison of the A segment to the L1 consensus sequence, as well as to those of the B1 and B2 repetitive elements (18,19) (SINEs) shows no evidence of any sequence homology. Several features within and surrounding the A segment are worth noting: within the 940 bp there are numerous short perfect direct and imperfect inverted repeated sequences both ranging from 5-20 bp in length; adjacent to either side of the A segment there are a pair of 13 bp repeats that exhibit perfect mirror symmetry (Fig. 2). These repeats are not to be confused with the more commonly seen inverted repeats or palindromes forming a dyad symmetry. There are two deletions in the neighborhood of the A segment. 60 bp of the L1 consensus sequence are missing from L1Md-6 at the position of the A segment as well as 6 bp from the region adjacent to one of the 13 bp repeats (Fig. 2).

Although the A segment appears to be an insertion it does not contain features typical of insertion elements. Inverted repeats are not present nor are there target site derived direct repeats. The latter may be due to the deletion bordering the A segment, however, we have no access to a specific target site for the A segment. As a result of independent insertions of the two L1 elements, L1Md-6 and L1Md-9, it is now evident that the region of the Hbb^{s} chromosome originally thought to be a target site for the A segment is not strictly an allele of its Hbb^{d} counterpart.

Insertion elements are expected to be present in multiple copies in the genome. Interestingly, the A segment is present as single copy DNA. A Southern blot hybridization of an A segment specific probe to BamH1 digests of BALB/c genomic DNA resulted in hybridization to a single restriction fragment. In addition, homology to the A segment is not present in the genome of C57BL/10, the prototype Hbb^{s} strain (17).

It is not likely that the A segment is a characteristic specific to L1Md-6

or L1 elements in general. Sequences homologous to the A segment have not been found in other L1 elements previously reported (16). The A segment is therefore unique in that it is not bordered by direct repeats and is present in only one copy per genome.

ACKNOWLEDGEMENTS

The authors thank Drs. Carolyn L. Jahn and Brian P. Nichols for helpful discussions. This work was supported by Public Health Service Grant GM30140.

REFERENCES

1. Weaver, S., Comer, M.B., Jahn, C.L., Hutchison, C.A., III, Edgell. M.H. (1981) Cell 24, 403-411.
2. Singer, M.F. (1982) Cell 28, 433-434.
3. Roeder, G. S., and Fink, G. R. (1980) Cell 21, 239-249.
4. Farabaugh, P.J., and Fink, G.R. (1980) Nature 286, 352-356.
5. Dunsmuir, P. Brorein, W.J., Simon, M.A., Rubin, G.M. (1980) Cell 21, 575-579.
6. Temin, H.M. (1981) Cell 27, 1-3.
7. Calos, M.P., and Miller, J.H. (1980) Cell 20, 579-595.
8. Spradling, A.C., and Rubin, G.M. (1982) Science 218, 341-347.
9. Potter, S.S. (1982) Nature 297, 201-204
10. Rogers, J. (1983) Nature 306, 113-114.
11. Roberts, T.M. and Lauer, G.D. (1979) Methods in Enzymology 68, 479-480.
12. Guo, L. and Wu, R. (1982) Nucl. Acids Res. 10, 2065-2084.
13. Maniatis, T., Fritsch, E.F., Sambrook, J. (1982) Molecular Cloning 170.
14. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. 74, 5463-5467.
15. Voliva, C.F., Jahn, C.L., Comer, M.B., Hutchison, C.A., III, Edgell, M.H. (1983) Nucl. Acids Res. 11, 8847-8859.
16. Voliva, C.F., Martin, S.L., Hutchison, C.A., III, Edgell, M.H., (1984) J. Mol. Biol. 178, 795-813.
17. Burton, F.H., Voliva, C.F., Edgell, M.H., Hutchison, C.A., III (1983) DNA 2, 82 (abstr).
18. Coggins, L.W., Vass, J.K., Stinson, M.A., Lanyon, W.G., Paul, J. (1982) Gene 17, 113-116.
19. Krayev, A.S., Markusheva, T.V., Kramerov, D.A., Ryskov, A.P., Skryabin, K.G., Bayev, A.A., Georgiev, G.P. (1982) Nucl. Acids Res. 10, 7461-7475.
20. Burton, F.H., Loeb, D.D., Chao, S.F., Hutchison, C.A., III, Edgell, M.H. Nucl. Acids Res., accompanying manuscript.