MOLECULAR AND CELLULAR BIOLOGY, Sept. 1985, p. 2197–2203
0270-7306/85/092197-07$02.00/0

Vol. 5, No. 9

Long Interspersed Repeated DNA (LINE) Causes Polymorphism at the Rat Insulin 1 Locus

MALATHI S. LAKSHMIKUMARAN,[1] ETTORE D'AMBROSIO,[1] LAIMONIS A. LAIMINS,[2] DAVID T. LIN,[1] AND ANTHONY V. FURANO[1]*

Section on Genomic Structure and Function, Laboratory of Biochemical Pharmacology, National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases,[1] and Section on Virus Tumor Biology, Laboratory of Molecular Virology, Division of Cancer Etiology, National Cancer Institute, Bethesda, Maryland 20205[2]

Received 22 April 1985/Accepted 3 June 1985

The insulin 1, but not the insulin 2, locus is polymorphic (i.e., exhibits allelic variation) in rats. Restriction enzyme analysis and hybridization studies showed that the polymorphic region is 2.2 kilobases upstream of the insulin 1 coding region and is due to the presence or absence of an approximately 2.7-kilobase repeated DNA element. DNA sequence determination showed that this DNA element is a member of a long interspersed repeated DNA family (LINE) that is highly repeated (>50,000 copies) and highly transcribed in the rat. Although the presence or absence of LINE sequences at the insulin 1 locus occurs in both the homozygous and heterozygous states, LINE-containing insulin 1 alleles are more prevalent in the rat population than are alleles without LINEs. Restriction enzyme analysis of the LINE-containing alleles indicated that at least two versions of the LINE sequence may be present at the insulin 1 locus in different rats. Either repeated transposition of LINE sequences or gene conversion between the resident insulin 1 LINE and other LINE sequences in the genome are possible explanations for this.

About one-third of various mammalian genomes is composed of families of repeated DNA sequences (4, 32). A major portion of this DNA is highly repeated (i.e., >20,000 copies per family) and is interspersed among nonrepeated DNA (7). Two classes of interspersed repeat DNA families have been identified, based on the length of the repeated element. Short interspersed repeated DNA families contain 100- to 300-base-pair (bp) repeat elements (12, 28, 32), whereas long interspersed repeated DNA families (LINE family) consist of elements that are quite complex and at least several kilobase pairs long (1, 9, 10, 16, 20, 21, 30, 32, 33).

About one-third of the rat genome also is repeated DNA (25). This fraction is dominated by several highly repeated short interspersed repeated DNA families and at least one LINE family, as is the case for other mammalian genomes (42). These families, which did not reanneal as highly repeated DNA (25), are highly transcribed in the rat (42).

Whatever their function, repeated DNA sequences may affect contiguous DNA sequences. These include effects on gene activity due to regulatory elements in the repeated DNA (28), as well as deletion, duplication, inversion, and translocation of the contiguous DNA sequences by homologous recombination between the resident repeated DNA sequence and other members of the repeat family in the genome. Such repeat DNA-mediated events are not just theoretical possibilities. One type of familial hypercholesteremia in humans, for example, is due to deletion of part of the gene for a lipoprotein receptor by recombination between two members of the human Alu short interspersed repeated DNA family (14).

It recently was shown (8) that two single-copy DNA loci of rats are polymorphic due to the presence or absence of a full-length (~6.5-kilobase [kb]) member of the rat LINE family: the *Igh* locus and the *Mlvi-2* locus, which is a site for the integration of Moloney leukemia virus (38). Here, we demonstrate that the earlier detected *Eco*RI restriction enzyme site polymorphism at the rat insulin 1 locus (6) also is due to the presence or absence of a highly repeated DNA element. DNA sequence determination of this region shows that the repeated DNA element is a portion of a typical rat LINE sequence and is one member of the inverted repeats that were found to flank the insulin 1 gene (18). LINE-containing insulin 1 alleles are present in the heterozygous or homozygous states in seven of eight individual rats, including a recently trapped wild rat and laboratory strains that were established from wild rats at widely different times and locations. Restriction enzyme analysis suggests that at least two different LINE sequences are present at LINE-containing alleles from different individuals.

LINE-related polymorphisms at three genetically independent loci indicate that LINE sequences have undergone (and perhaps are presently undergoing) transposition and that this type of polymorphism might be quite general in rats. Since LINE sequences might profoundly affect contiguous DNA sequences (see above), then a LINE-containing locus may be subjected to regulatory phenomena and evolutionary fates that are quite different from its counterpart without a LINE. The implications for genomic evolution are discussed.

MATERIALS AND METHODS

Animals. The DNAs from Sprague-Dawley (T) (SDT) and Sprague-Dawley (R) (SDR) rats were supplied by Linda Jagodzinski, Biotech Research Laboratories, Rockville, Md. The DNAs from the Osborne-Mendel, Sprague-Dawley (P) (SDP), Fischer, and Buffalo (B) rats were given to us by Philip N. Tsichlis, Fox Chase Cancer Center, Philadelphia, Pa. The sources of the other DNAs were as described by Witney and Furano (41).

Southern blot preparation and hybridization. DNA was digested with restriction endonucleases and electrophoresed in 1% agarose gels. The DNA was transferred to nitrocellulose filters (35) which were prehybridized at 65°C for 14 to 20

* Corresponding author.

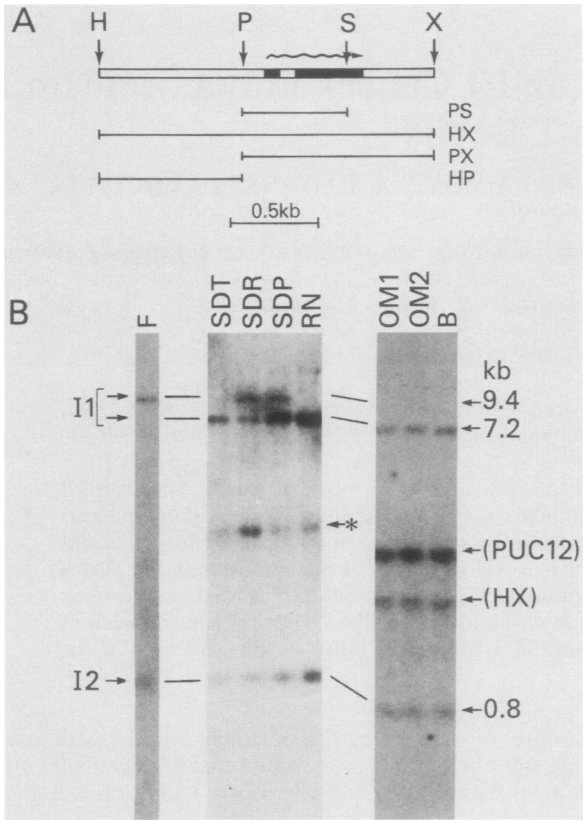

FIG. 1. *Eco*RI site polymorphism of the insulin 1 locus. (A) Partial restriction map of the rat insulin 1 locus. H, P, S, and X indicate the sites for *Hin*dIII, *Pst*I, *Sau*3A, and *Xba*I, respectively. The filled regions correspond to the mature insulin 1 transcript (6, 18). The lines indicated by PS, HX, PX, and HP represent the different subclones that we prepared. The arrow indicates the direction of transcription. (B) A blot of *Eco*RI-restricted total rat DNA from eight different rats was hybridized with the PS probe. Lanes: F, Fischer rat; SDT, SDT rat; SDR, SDR rat; SDP, SDP rat; RN, RN (wild rat); OM1, OM1 rat; OM2, OM2 rat; B, B rat. *Hin*dIII-digested λ DNA and internal autoradiographic standards (PUC12 linear and HX fragment) were used to determine the size of insulin 1 fragments. The band marked with an asterisk was detected only when the temperature of hybridization was lowered from 65 to 55°C. It was not seen when hybridization to these DNAs was carried out at 65°C (data not shown) or in lanes F, OM1, OM2, and B which were hybridized at 65°C.

h with a hybridization solution containing 0.4 M sodium phosphate buffer (pH 7.0), 0.25% (wt/vol) sodium dodecyl sulfate, 1 mg of polyvinylpyrrolidone per ml, 1 mg of Ficoll per ml, 100 μg of denatured, sonicated salmon sperm DNA per ml, and 50 μg of heparin per ml (34). They then were hybridized to the appropriate ^{32}P-labeled DNA probes. After hybridization, the filters were washed twice with 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and twice with 3 mM Tris base at room temperature, dried, and exposed to X-ray film at −70°C with an intensifier.

DNA sequencing. The DNA sequence was determined by the chain termination procedure (27). DNA fragments were cloned into M13 mp18 and M13 mp19 (24). *Exo*III deletion clones were prepared by the procedure of Henikoff (11). Nucleotide sequences were analyzed by previously published computer programs (5, 40), as well as by programs for detecting short regions of homology between long sequences (M. Kanehisa and D. Lipman, personal communication).

RESULTS

Two nonallelic genes for insulin are present and expressed in the rat, and the coding regions of both the insulin 1 and the insulin 2 genes are homologous (6, 18). Earlier studies showed that the location of an *Eco*RI site upstream of the insulin 1 locus is polymorphic, such that depending on the source of DNA, the insulin 1 coding sequences were located on either a 7.2- or a 9.4-kb *Eco*RI fragment or both (6). The insulin 2 locus did not exhibit this polymorphism (6; see below).

To investigate the DNA polymorphism at the insulin 1 locus in more detail, different portions of the insulin 1 region were cloned into PUC12 (Fig. 1A). All the clones except HP contain the insulin coding sequences and can detect insulin 1 and insulin 2 genes (see below).

In the present work we examined eight different individual rats that were not analyzed earlier. Total rat DNA was digested with *Eco*RI, and blots of these digests were hybridized with PS DNA. Figure 1B shows that the insulin 1 sequences are located on a 7.2-kb *Eco*RI fragment in a recently trapped wild specimen of *Rattus norvegicus* (RN) and in Osborne-Mendel 1 (OM1), Osborne-Mendel 2 (OM2), B, and SDT rats, whereas in the Fischer rat the insulin 1 gene is located on a 9.4-kb *Eco*RI fragment. The SDP and SDR rats have two alleles for insulin 1 located on a 7.2- and a 9.4-kb *Eco*RI fragment. Our data confirm the earlier studies (6) that the insulin 1 gene is polymorphic with respect to the *Eco*RI site. The earlier studies showed that insulin 2 sequences are located on a 4.0- and a 0.8-kb *Eco*RI fragment (6, 18). In all of the lanes, an 0.8-kb band was observed which corresponds to insulin 2 sequences (Fig. 1B). The PS DNA probe does not hybridize to the 4.0-kb insulin 2-containing *Eco*RI fragment, since the probe does not extend into this region of the insulin 2 gene. An extra band (labeled with an asterisk) is present in the SDT, SDR, SDP, and RN lanes. This band was observed only when the temperature of hybridization was lowered from 65 to 55°C and does not correspond to insulin 1 or 2 sequences (also see the legend to Figure 1).

To determine whether the polymorphism at the insulin 1 locus also involved other restriction enzyme sites, total rat DNA was digested with different restriction enzymes. The results (Table 1) show that the insulin 1 locus is also polymorphic with repect to *Bam*HI, *Bgl*II, and *Xba*I sites. In all cases, the restriction fragments containing the insulin 2 gene also were identified. The sizes of these (not shown in

TABLE 1. Insulin 1-containing restriction endonuclease fragments

Source of DNA	Size (kb) of insulin 1-containing fragment							
	*Eco*RI	*Bam*HI	*Bgl*II	*Pst*I	*Xba*I	*Hin*dIII	*Kpn*I	*Sac*I
RN (wild)	7.2	5.3		5.0	6.7	5.8	14	4.1
SDP	9.4	10.2	23	5.0	4.0	5.8	14	4.1
	7.2	5.3	10.0	5.0	6.7	5.8	14	4.1
SDR	9.4	10.2	23	5.0	4.0	5.8	14	
	7.2	5.3	10.0	5.0	6.7	5.8	14	
SDT	7.2	15	10.0	5.0	6.7	5.8	14	
OM1	7.2	5.3	10.0	5.0	6.7	5.8		
		15						
OM2	7.2	5.3	10.0	5.0	6.7	5.8		
		15						
B	7.2	5.3	10.0	5.0	6.7	5.8		
Fischer	9.4	10.2	23	5.0	4.0	5.8		

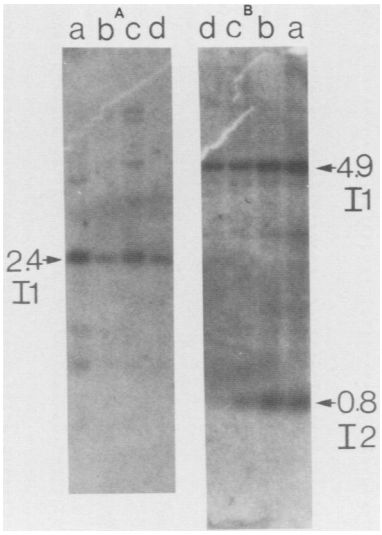

FIG. 2. Double restriction endonuclease digests of total rat DNA. Total rat DNA (5 μg per lane) that was digested with restriction endonucleases and electrophoresed on agarose gels was blotted and hybridized with the PS probe (Fig. 1) exactly as described in the text. A 2.4-kb insulin 1 band was obtained when rat DNA was digested with both *Hin*dIII and *Bam*HI (A), and a 4.9-kb band was observed after digestion with *Pst*I and *Eco*RI (B). The size of the bands was determined with marker DNA (λ DNA digested with *Hin*dIII). Lanes: a, RN (wild); b, SDT- rat; c, SDR- rat; d, Fischer- rat.

Table 1) agreed with those previously determined for the cloned insulin 2 gene (6, 18) and exhibited no polymorphism (data not shown).

The above results indicated that the polymorphism at insulin 1 is not due to a single base change, and in the case of the *Bam*HI restriction, three different sized fragments were identified in different rats. This suggests that there are at least three alleles of insulin 1 (see below).

We determined the location of the polymorphic restriction sites with respect to the insulin 1 gene by performing different sets of double restriction enzyme digestions. To determine whether the polymorphism was downstream of the insulin 1 gene, the DNA was first digested with *Pst*I, which digests the insulin 1 locus 341 bp upstream of the translation start site (6; see below), and then with other

TABLE 2. Size of insulin 1-containing fragments after sequential digestion with two restriction enzymes

First enzyme	Source of DNA	Size (kb) after digestion with following second enzyme		
		*Eco*RI	*BglII*	*Xba*I
Pst	RN (wild)	4.9	5.0	1.1
	SDP	4.9	5.0	1.1
	Fischer	4.9	5.0	1.1
	OM1	4.9	5.0	1.1
	SDT	4.9	5.0	1.1
	B	4.9	5.0	1.1
*Xba*I	RN (wild)	3.3	3.9	
	SDP	3.3, 4.0	3.9, 4.0	
	Fischer	4.0	4.0	
	OM1	3.3	3.9	
	SDT	3.3	3.9	
	B	3.3	3.9	

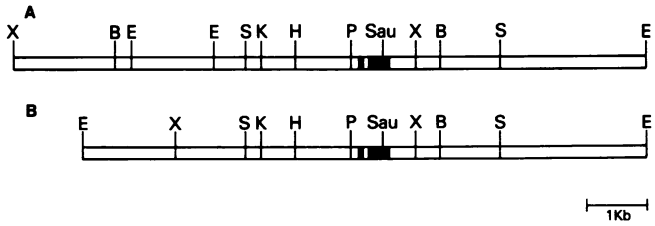

FIG. 3. Restriction endonuclease map of the rat insulin 1 locus representing the two alleles. The letters represent the various restriction enzyme sites. B, E, H, K, P, S, Sau, and X indicate the restriction sites for the enzymes *Bam*HI, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Sac*I, *Sau*3A, and *Xba*I, respectively. The coding region of insulin 1 is indicated by the filled boxes.

enzymes: *Bgl*II, *Eco*RI, and *Xba*I. An example of the blot hybridization of the *Pst*I-*Eco*RI digests is shown in Fig. 2B, and the results are summarized in Table 2. The insulin 1 gene was located in the same size fragments for all of the rats investigated for each set of double restrictions. Therefore, the *Eco*RI, *Xba*I, and *Bgl*II restriction sites downstream of insulin 1 are at the same position in the various alleles.

To determine whether the three types of *Bam*HI polymor-

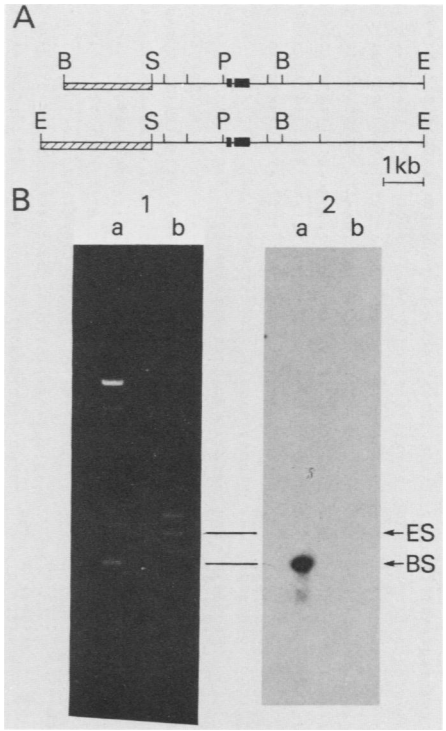

FIG. 4. Polymorphism at the insulin 1 locus is due to the presence or absence of repeated DNA sequences. (A) Partial restriction endonuclease map of the two alleles of insulin 1. B, S, P, and E indicate the sites for *Bam*HI, *Sac*I, *Pst*I, and *Eco*RI, respectively. (B) Panel 1 shows an ethidium bromide-stained gel. Lane a contains the *Bam*HI-*Sac*I digest of a pBR322 subclone of a LINE-containing insulin 1 allele (see the upper diagram of panel A and Fig. 3). The band labeled BS corresponds to the crosshatched box. Lane b contains an *Eco*RI-*Sac*I digest of a PUC12 clone that we prepared from the 9.4-kb *Eco*RI insulin 1 allele (see the lower diagram of panel A and Fig. 3). The band labeled ES corresponds to the crosshatched box. Panel 2 shows an autoradiogram of the hybridization of total rat DNA to a nitrocellulose blot of this gel.

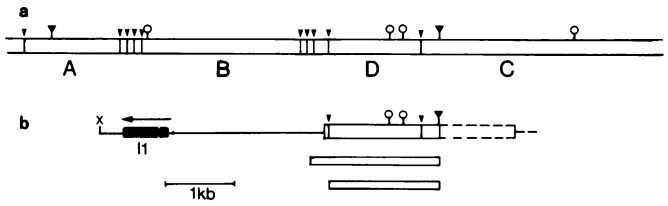

FIG. 6. Comparison of the LINE-containing insulin 1 allele with a typical rat LINE. (a) Structure of a member of the rat LINE that we determined from its DNA sequence. Only the recognition sites for *Eco*RI (▼), *Bam*HI (▼), and *Hin*dIII (▼) are shown. All of the DNA between the left-hand *Bam*HI site and 1.1 kb beyond the right-hand *Bam*HI site and is highly repeated and highly transcribed in rats (unpublished data). The depicted structure, in which we designate four segments (A, B, D, and C), seems to be quite typical of this family and can account for prominently stained electrophoretic bands seen in total rat DNA after digestion with *Eco*RI, *Bam*HI, or *Hin*dIII as well as several other restriction enzymes (unpublished data). Furthermore, it is generally colinear with the LINE sequences that make the *Igh* and *Mlvi-2* locus polymorphic (8) as well as with another randomly selected member of the rat LINE family. The region near the most-right-hand *Eco*RI site and *Bam*HI site, is very similar to the corresponding portion of the mouse LINE (33). (b) Structure of the LINE-containing insulin 1 allele that we determined from the DNA sequence (Fig. 5) and restriction enzyme (Fig. 3) analyses. The filled boxes and arrow indicate the insulin 1 coding sequence and direction of transcription, respectively. X depicts the *Xba*I site just downstream of the coding sequence (Fig. 3), and the other symbols are as described above for part a. The open box on the right indicates the LINE sequence, and the dashed lines indicate that portion of the LINE which we deduced to be present from restriction enzyme analysis (see the text and Fig. 3). The two thin open boxes under the insulin 1 LINE depict the corresponding portion of two other rat LINEs that we sequenced (Fig. 5). The first of these two boxes is from LINE 3 which is the LINE shown in part a.

phism (Table 1) are upstream or downstream of insulin 1 coding sequences, DNA from the appropriate rats was restricted with *Hin*dIII, which digests the insulin 1 region 1 kb upstream of the insulin 1 coding region (6, 18; see below), and *Bam*HI. The insulin 1 gene was located in the 2.4-kb fragment in all of the rats studied (Fig. 2A). Therefore, all of the polymorphisms that we detected are upstream of the insulin 1 coding region.

From the results of the restriction enzyme analysis shown above (as well as other data not shown), we constructed restriction site maps (Fig. 3). The polymorphic region is about 2.2 kb upstream of insulin 1 coding sequences, and a comparison of the relative positions of the upstream *Xba*I sites suggests that the diagrammed insulin 1 alleles differ by the presence of a 2.7-kb DNA element.

Repetitive DNA element at the polymorphic region. To investigate the nature of the inserted DNA element, the polymorphic region from each allele was hybridized to total rat DNA. A pBR322 clone of the *Bam*HI-*Xba*I fragment of the insulin 1 allele associated with the 7.2-kb *Eco*RI fragment (Fig. 3) was digested with *Bam*HI and *Sac*I to produce the fragment corresponding to the crosshatched region of Fig. 4A (upper diagram). The *Eco*RI-*Sac*I fragment from the 9.4-kb insulin 1 *Eco*RI fragment (2) (Fig. 4A, lower diagram, crosshatched bar) was cloned into PUC12. The appropriate restriction enzyme digests of these plasmid clones were resolved by electrophoresis on 1% agarose gel (Fig. 4B, panel 1), and blots of these gels were hybridized to radioactive total rat DNA (Fig. 4B, panel 2). Total rat DNA hybridized to the *Bam*HI-*Sac*I region of the insulin 1 allele that is on the 7.2-kb *Eco*RI fragment, but not to the corresponding region of the insulin 1 allele that is present on the 9.4-kb fragment. These results indicated that the inserted DNA is a repetitive element. Therefore, the allelic forms of the insulin 1 locus are due to the presence or absence of repetitive DNA sequences.

Sequence of the insulin 1 region. Both the length of the repeated DNA segment at the insulin 1 locus and its pattern of hybridization to various restriction enzyme digests of total rat DNA suggested that the repeated DNA was a member of the rat LINE family (data not shown). To confirm this, we determined the DNA sequence of the portion of a LINE-containing insulin 1 allele that extends from the *Bam*HI site that is upstream of the insulin 1 coding sequence to the *Xba*I site downstream of it (Fig. 3A). The results are shown in Fig. 5, but here the sequence is written with the *Xba*I site on the left and the *Bam*HI site on the right.

The last 1,600 bp or so of the sequence shown in Fig. 5 is very homologous to a part of two other rat LINE members that we also sequenced (LINEs 3 and 4). (The entire sequence and other characteristics of the rat LINE family will be published elsewhere.) Comparison of the insulin 1 sequence and the rat LINE 3 sequence clearly shows the demarcation between the repeated and nonrepeated DNA at the insulin 1 allele. Figure 6a shows a diagram of an entire rat LINE member aligned with the insulin 1 sequence presented in Fig. 5 (see the legend for Fig. 6 for more details).

In addition to being almost identical in sequence to the two other rat LINE members, the insulin 1 LINE also contains an open reading frame that extends from the asterisk in Fig. 5 (nucleotide 4129) for 265 codons to the *Bam*HI site, which is as far as we sequenced. However, Soares et al. also have sequenced the same insulin 1 allele (personal communication) and showed that the open reading frame extends beyond the *Bam*HI site into what we have called the C segment (Fig. 6; 8). Furthermore, rat LINEs 3 and 4 also contain open reading frames that begin in segment D of the LINE and also extend into the C segment (data not shown).

FIG. 5. DNA sequence of a LINE-containing allele of insulin 1. The region sequenced extends from the *Bam*HI site (B) which is 4 kb upstream of the beginning of the insulin 1 gene to the *Xba*I site (X) that is just downstream of the gene (Fig. 3A). The sequence here is written with the *Xba*I site at the 5′ end and the *Bam*HI site at the 3′ end and thus is in the opposite orientation to that shown in Fig. 3. This is so the LINE portion of the sequence is in the same orientation as that generally presented for these structures (33; Fig. 6). Some of the restriction enzyme sites shown in Fig. 3 are indicated, as well as the initiation and termination codons of the insulin coding sequence of which both strands are shown. The LINE portion of the insulin 1 allele was compared with the relevant portions of two other LINE sequences (LINEs 3 and 4), randomly selected from a recombinant DNA library of rat DNA (42). Identities and differences among those sequences are indicated by a · and by a letter or –, respectively. The * indicates the ATG which is at the start of the open reading frame in the insulin 1 LINE and extends to the *Bam*HI site (see the text). The imperfect, inverted repeat and tandem array of four GAATTs are underlined (see the text). The region of the insulin 1 sequence enclosed in brackets was not determined by us but was taken from the Gen Bank data base (Genetic Sequence Data Bank, compiled and maintained by the Los Alamos National Laboratory, Los Alamos, NM 87545). About 95% of the sequence was determined in both directions or from a series of overlapping *Exo*III deletion clones (11).

Open reading frames have been identified in the corresponding region of mouse LINEs and primate LINEs (19) and thus appear to be a general feature of mammalian LINEs.

We analyzed all of the non-LINE insulin 1 sequence for homology to all segments of the rat LINE diagrammed in Fig. 6 by computer methods suitable for detecting either general or local regions of homology. No statistically significant homologies were detected. We did find an imperfect, inverted repeat of 33 bp which contains four copies of the sequence AGCT about 1,900 bp to the left of the LINE–non-LINE junction and a tandem array of four GAATT sequences about 500 bp to the left of this junction (Fig. 5). Neither of these was found on randomized versions of the DNA sequence between the insulin coding sequence and the LINE sequence.

DISCUSSION

Polymorphism at the rat insulin 1 locus is due to the presence or absence of an approximately 2.7-kb DNA element. Our DNA sequence determination, as well as that of Soares et al. (personal communication), shows that this DNA is a member of the highly repeated rat LINE family. The presence or absence of LINE sequences at a given locus indicates that LINE elements are (or were) mobile in the rat genome.

Seven of the eight rats examined by us have at least one LINE-containing insulin 1 allele, and only one rat was homozygous for the absence of LINE sequences. Since these rats included laboratory strains established as independent colonies from wild rat populations at widely different times and locations (17), as well as a recently trapped wild rat, LINE-containing insulin 1 alleles are quite common in the population of *R. norvegicus*. Therefore, either this allele arose early in the evolutionary history of rats or LINE sequences are now undergoing transposition at the insulin 1 locus. The *Bam*HI restriction enzyme analysis of LINE-containing alleles indicates that at least two different versions of the rat LINE sequences are present at the insulin 1 locus. Repeated transposition of LINE sequences or gene conversion between the resident insulin 1 LINE and other LINE members in the genome are possible explanations for this. Determination of the DNA sequence of individual alleles that lack the LINE as well as of the junction between LINE and non-LINE sequences from different LINE-containing alleles should reveal whether this locus has undergone repeated independent transpositional events.

It recently was shown that two other single-copy loci, the *Igh* locus and the *Mlvi-2* locus, are polymorphic due to the presence or absence of rat LINE sequences (8). The LINE sequences at these other two loci are more than 6 kb long and are colinear with the LINE depicted in Fig. 6. By contrast, the LINE sequences at the insulin 1 locus consist of just the D and C segments (Fig. 6). Other studies (unpublished data) showed that at least half of the B and D LINE segments in the genome are present in 5.5-kb *Bam*HI fragments. These results as well as others indicate that at least half of the LINE sequences are organized as the >6.0-kb structure shown in Fig. 6a.

The presence or absence of LINE sequences at three genetically independent loci does not necessarily indicate that LINEs have transposed randomly in the genome. Both the *Igh* and *Mlvi-2* loci normally undergo recombination events (8): the *Igh* locus during B-cell differentiation (26) and the *Mlvi-2* locus as a site of integration in Moloney leukemia virus (38). Recent evidence from Soares et al. (personal communication) indicates that the insulin 1 gene probably arose from the integration of a DNA copy of an aberrant transcript of the ancestral insulin 2 gene. This suggests that the insulin 1 locus also may be located in a naturally recombinogenic region of the genome. Perhaps LINE transposition involves only regions such as these, and LINE-related polymorphisms may in fact identify such regions. Whether the inverted repeat or the tandem array of four GAATTs present at the insulin 1 region (Fig. 5 and 6) is related to LINE transposition to this site might be clarified when the DNA sequence of the other LINE integration sites is known.

There is no homology detectable by hybridization (8) or by DNA sequence comparison (this work) between LINE sequences and the three sites involved in LINE transposition. This suggests that homologous recombination does not play a role in the transposition of LINEs at these loci. This is not surprising since illegitimate recombination (i.e., nonhomologous recombination) appears to have been involved in the transposition of a wide variety of DNA sequences that has occurred in mammalian genomes (10, 13, 15, 22, 23, 28, 31, 36, 37, 39). Many of these transposed sequences, which include highly repeated ones, are transcribed, and the transposition is thought to involve the integration of a DNA copy of an RNA intermediate (10, 13, 15, 22, 23, 28, 31, 36, 37, 39). Since all segments of the rat LINE are highly transcribed in various rat cells (42; unpublished data), this pathway also may be used for the transposition of rat LINE sequences. However, other mechanisms also are possible.

We know of only one other report of a single-copy locus that is polymorphic due to the presence or absence of a highly repeated DNA sequence (29). In this case, polymorphism at the rat prolactin locus is due to a member of a short interspersed repeated DNA family. As mentioned in the introduction, whatever its function, a transcriptionally active or highly repeated, or both, DNA element could dramatically affect contiguous DNA sequences. In addition to altering gene activity, the repeated DNA sequences could cause genetic rearrangement of the contiguous DNA. It has long been speculated (e.g., see references 3 and 31) that genomic rearrangements mediated by homologous recombination among repeated DNA sequences could expose functional genes to novel regulatory environments and thereby produce plasticity in the genome, which in the short term may have adaptive value and in the long term may be fundamental to genomic evolution and speciation. It is clear that transposition of repeated DNA to (or from) a given locus also could alter the regulatory environment of a functional gene. If the locus is polymorphic with respect to the repeated DNA sequence, then the original functioning locus would be maintained in the population, whereas the altered locus could be subjected to any of the above putative effects of the repeated DNA sequences.

ACKNOWLEDGMENTS

We thank William Rutter for providing the pBR322 clone of the 9.4-kb *Eco*RI insulin 1 allele (2) and Peter Lomedico for the clone of the *Bam*HI-*Xba*I fragment of the insulin 1 allele that is associated with the 7.4-kb *Eco*RI fragment. We are especially appreciative of Argiris Efstratiadis and his colleagues M. Soares and E. Schon for communicating to us their results and their DNA sequence of the LINE-containing insulin 1 allele before publication.

LITERATURE CITED

1. **Adams, J. W., R. E. Kaufman, P. J. Kretschmer, M. Harrison, and A. W. Nienhuis.** 1980. A family of long reiterated DNA sequences, one copy of which is next to the human beta globin

gene. Nucleic Acids Res. **8**:6113–6128.
2. **Bell, G. I., R. Pictet, and W. J. Rutter.** 1980. Analysis of the regions flanking the human insulin gene and sequence of an Alu family member. Nucleic Acids Res. **8**:4091–4109.
3. **Britten, R. J., and E. H. Davidson.** 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q. Rev. Biol. **46**:111–138.
4. **Britten, R. J., and D. E. Kohne.** 1968. Repeated sequences in DNA. Science **161**:529–540.
5. **Brutlag, D. L., J. Clayton, P. Friedland, and L. H. Kedes.** 1982. SEQ: a nucleotide sequence analysis and recombination system. Nucleic Acids Res. **10**:279–294.
6. **Cordell, B., G. I. Bell, E. Tischer, F. M. DeNoto, A. Ullrich, R. Pictet, W. J. Rutter, and H. M. Goodman.** 1979. Isolation and characterization of a cloned rat insulin gene. Cell **18**:533–543.
7. **Davidson, E. H., B. R. Hough, C. S. Amenson, and R. J. Britten.** 1973. General interspersion of repetitive and non-repetitive sequence elements in the DNA of *Xenopus*. J. Mol. Biol. **77**:1–23.
8. **Economou-Pachnis, A., M. A. Lohse, A. V. Furano, and P. N. Tsichlis.** 1985. Insertion of long interspersed repeated elements (LINEs) at the *Igh* and *Mlvi-2* loci of rats. Proc. Natl. Acad. Sci. U.S.A. **82**:2857–2861.
9. **Fanning, T. G.** 1983. Size and structure of the highly repetitive *Bam*HI element in mice. Nucleic Acids Res. **11**:5073–5091.
10. **Grimaldi, G., and M. F. Singer.** 1983. Members of the *Kpn*I family of long interspersed repeated sequences join and interrupt α-satellite in the monkey genome. Nucleic Acids Res. **11**:321–338.
11. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted break points for DNA sequencing. Gene **28**:351–359.
12. **Kramerov, D. A., A. A. Grigoryan, A. P. Ryskov, and G. P. Georgiev.** 1979. Long double stranded sequences (dsRNA-B) of nuclear pre-mRNA consist of a few highly abundant classes of sequences: evidence from DNA cloning experiments. Nucleic Acids Res. **6**:697–713.
13. **Kuff, E. L., A. Feenstra, K. Lueders, L. Smith, R. Hawley, N. Hozumi, and M. Shulman.** 1983. Intracisternal A-particle genes as movable elements in the mouse genome. Proc. Natl. Acad. Sci. U.S.A. **80**:1992–1996.
14. **Lehrman, M. A., W. J. Schneider, T. C. Südhof, M. S. Brown, J. L. Goldstein, and D. W. Russell.** 1985. Mutations in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. Science **227**:140–146.
15. **Lemischka, I., and P. A. Sharp.** 1982. The sequences of an expressed rat α-tubulin gene and a pseudogene with an inserted repetitive element. Nature (London) **300**:330–335.
16. **Lerman, M. I., R. E. Thayer, and M. F. Singer.** 1983. *Kpn*I family of long interspersed repeated DNA sequences in primates: polymorphism of family members and evidence for transcription. Proc. Natl. Acad. Sci. U.S.A. **80**:3966–3970.
17. **Lindsey, J. R.** 1979. Historical foundations, p. 1–36. *In* H. J. Baker, J. R. Lindsey, and S. H. Weisbroth (ed.), The laboratory rat, vol. 1. Academic Press, Inc., New York.
18. **Lomedico, P., N. Rosenthal, A. Efstratiadis, W. Gilbert, R. Kolodner, and R. Tizard.** 1979. The structure and evolution of the two nonallelic rat preproinsulin genes. Cell **18**:545–558.
19. **Martin, S. L., C. F. Voliva, F. H. Burton, M. H. Edgell, and C. A. Hutchison.** 1984. A large interspersed repeat found in mouse DNA contains a long open reading frame that evolves as if it encodes a protein. Proc. Natl. Acad. Sci. U.S.A. **81**:2308–2312.
20. **Meunier-Rotival, M., and G. Bernardi.** 1984. The Bam repeats of the mouse genome belong in several super families the longest of which is over 9 kb in size. Nucleic Acids Res. **12**:1593–1608.
21. **Meunier-Rotival, M., P. Soriano, G. Cuny, F. Strauss, and G. Bernardi.** 1982. Sequence organization and genomic distribution of the major family of interspersed repeats of mouse DNA. Proc. Natl. Acad. Sci. U.S.A. **79**:355–359.
22. **Nishioka, Y., A. Leder, and P. Leder.** 1980. Unusual α-globin like gene that has cleanly lost both globin intervening sequences. Proc. Natl. Acad. Sci. U.S.A. **77**:2806–2809.
23. **Nomiyama, H., T. Tsuzuki, S. Wakasugi, H. Fukuda, and K. Shimada.** 1984. Interruption of a human nuclear sequence homologous to mitochondrial DNA by a member of the *Kpn*I 1.8 kb family. Nucleic Acids Res. **12**:5225–5234.
24. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26**:101–106.
25. **Pearson, W. R., J. Wu, and J. Bonner.** 1978. Analysis of rat repetitive DNA sequences. Biochemistry **17**:51–59.
26. **Sakano, H., R. Maki, W. Kurosawa, W. Roeder, and S. Tonegawa.** 1980. Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy chain genes. Nature (London) **286**:676–683.
27. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
28. **Schmid, C. W., and W. R. Jelinek.** 1982. The Alu family of dispersed repetitive sequences. Science **216**:1065–1070.
29. **Schuler, L. A., J. L. Weber, and J. Gorski.** 1983. Polymorphism near the rat prolactin gene caused by insertion of an Alu-like element. Nature (London) **305**:159–160.
30. **Shafit-Zagardo, B., F. L. Brown, J. J. Maio, and J. W. Adams.** 1982. *Kpn*I families of long, interspersed repetitive DNA's associated with the human β-globin gene cluster. Gene **20**:397–407.
31. **Sharp, P. A.** 1983. Conversion of RNA to DNA in mammals: Alu-like elements and pseudogenes. Nature (London) **301**: 471–472.
32. **Singer, M. F.** 1982. Highly repeated sequences in mammalian genomes. Int. Rev. Cytol. **76**:67–112.
33. **Singer, M. F., R. E. Thayer, G. Grimaldi, M. J. Lerman, and T. G. Fanning.** 1983. Homology between the *Kpn*I primate and *Bam*HI (MIF-1) rodent families of long interspersed, repeated sequences. Nucleic Acids Res. **11**:5739–5745.
34. **Singh, L., and K. W. Jones.** 1984. The use of heparin as a simple cost effective means of controlling background in nucleic acid hybridization procedures. Nucleic Acids Res. **12**:5627–5638.
35. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
36. **Streeck, R. E.** 1982. A multicopy insertion sequence in the bovine genome with structural homology to the long terminal repeats of retroviruses. Nature (London) **298**:767–769.
37. **Thayer, R. E., and M. F. Singer.** 1983. Interruption of an α-satellite array by a short member of the *Kpn*I family of interspersed, highly repeated monkey DNA sequences. Mol. Cell. Biol. **3**:967–973.
38. **Tsichlis, P. N., P. G. Strauss, and L. F. Hu.** 1983. A common region for proviral DNA integration in MoMuLV induced rat thymic lymphomas. Nature (London) **302**:445–449.
39. **Van-Arsdell, S. W., R. A. Denison, L. B. Bernstein, A. M. Weiner, T. Manser, and R. F. Gestland.** 1981. Direct repeats flank three small nuclear RNA pseudogenes in the human genome. Cell **26**:11–17.
40. **Wilbur, W., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acids and protein data banks. Proc. Natl. Acad. Sci. U.S.A. **80**:726–730.
41. **Witney, F. R., and A. V. Furano.** 1983. The independent evolution of two closely related satellite DNA elements in rats (Rattus). Nucleic Acids Res. **11**:291–304.
42. **Witney, F. R., and A. V. Furano.** 1984. Highly repeated DNA families in the rat. J. Biol. Chem. **259**:10481–10492.