Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 3966–3970, July 1983
Biochemistry

Kpn I family of long interspersed repeated DNA sequences in primates: Polymorphism of family members and evidence for transcription

(interspersed DNA/processed genes)

MICHAEL I. LERMAN, RONALD E. THAYER, AND MAXINE F. SINGER

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Building 37, Room 4E28, Bethesda, Maryland 20205

Contributed by Maxine F. Singer, April 6, 1983

ABSTRACT **An ≈2-kilobase-pair-long member (*Kpn* I-LS1) of the African green monkey *Kpn* I family of repeated sequences has been cloned, subjected to sequence analysis, and compared to other family members which are over 6 kilobase pairs (*Kpn* I-α7) and 829 base pairs (*Kpn* I-RET) long. Both *Kpn* I-LS1 and *Kpn* I-RET lack sequences found at the ends of the longer family member and their structures resemble those of processed genes. *Kpn* I-LS1 sequences are colinear with part of the long family member, *Kpn* I-α7. However, although all sequences in *Kpn* I-RET are represented in *Kpn* I-LS1, the two are not colinear; *Kpn* I-RET is missing 731 base pairs found in *Kpn* I-LS1 and one segment flanking the deletion is inverted. The results demonstrate that *Kpn* I family members are not only of different lengths but may also contain scrambled arrangements of common sequences. Sequences in *Kpn* I-LS1 hybridize to RNA from monkey and human cells, indicating that some family members are transcribed.**

Eukaryote genomes typically contain multiple families of repeated DNA sequences that are interspersed among other genomic segments. In primates, the *Kpn* I family members are polymorphic in length and sequence, occur on the order of 10^4 times, and have been observed flanking genes and within satellite DNA (1–8). Several long [>6 kilobase pairs (kbp)] *Kpn* I family members have been cloned from both human (1, 5) and African green monkey (AGM) (6) DNA and their restriction maps have been defined, including the ordering of the *Kpn* I, *Eco*RI, and *Hin*dIII endonuclease fragments that are produced from typical family members upon digestion of total DNA. Very similar but distinctive maps were found for human and AGM family members (1, 3, 5, 6). A very short [829 base pairs (bp)] AGM family member called *Kpn* I-RET interrupts a cloned stretch of AGM α-satellite DNA (8). Its structure resembles that of a processed gene and it is flanked by 14-bp direct repeats of the interrupted α-satellite sequence, suggesting that it was inserted as a moveable element.

In this paper we describe an AGM *Kpn* I family member of intermediate length called *Kpn* I-LS1. *Kpn* I-LS1 was discovered by chance within a 17.5-kbp cloned AGM DNA segment, λCaOri7 (9), that also contains a segment homologous to the region around the origin of replication of simian virus 40 (SV40) and several Alu family members (Fig. 1*A*) (10–12). The structure of *Kpn* I-LS1 is compared here with that of *Kpn* I-RET and *Kpn* I-α7, a typical long AGM family member (6). Also, we report that transcripts homologous to *Kpn* I family sequences have been detected in both AGM and human cells.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

METHODS

Materials. The recombinant phages λCaα7 (6) and λCaOri7 (9–12) were described previously; the *Kpn* I family members included in these clones are called *Kpn* I-α7 and *Kpn* I-LS1, respectively. For convenience we refer to them as α7 and LS1 in this paper and *Kpn* I-RET (8) is called RET. DNA was isolated (13) and purified (14) from monkey liver. Phage and plasmid DNAs were isolated as described (9). Total RNA from AGM liver and HeLa cells was isolated by the guanidinium isothiocyanate method (15) and the polyadenylylated RNA was selected as described (16). Total liver nuclear RNA was isolated from purified nuclei and was treated with DNase I (17). The primary human fibroblast cDNA library was kindly supplied by H. Okayama and P. Berg (18); two cDNA plasmids, pcDK1 and pcDK2, were selected by screening (19) the library with sequences from LS1 (see *Results*).

Subcloning. DNA fragments were isolated from phage or plasmid digests by preparative agarose gel electrophoresis and were ligated into pBR322 directly or after attaching *Hin*dIII linkers (Collaborative Research, Waltham, MA).

Hybridization Analyses. Procedures for labeling plasmid and genomic DNA and for restriction endonuclease digestion, gel electrophoresis, transfer to nitrocellulose filters, and hybridization were as described (6, 8, 9).

Nucleotide Sequence Analysis. The nucleotide sequence determinations were by the Maxam and Gilbert procedure (20) as modified (21). Computer analysis (22) and alignment (23) of sequences were by published methods.

RESULTS

The LS1 Segment. Hybridization of ^{32}P-labeled total monkey DNA and Alu sequences (24) to restriction endonuclease digests of λCaOri7 (Fig. 1*A*) identified a region containing highly repetitive, but non-Alu sequences, starting about 3 kbp away from the SV40-like segment (data not shown). The region also hybridized with subcloned probes prepared from previously characterized AGM *Kpn* I family members, including α7 and RET; the cross-hybridizing segments in λCaOri7 are contained within a maximum of 2.4 kbp, as summarized in Fig. 1 (data not shown). No hybridization to ^{32}P-labeled total monkey DNA was detected in restriction endonuclease fragments located immediately adjacent to the 2.4-kbp region. Thus LS1, the *Kpn* I family member in λCaOri7, is bounded by low copy number sequences. Present data do not permit the precise localization of the left end of LS1; the right end is very close to the right hand *Hin*dIII site (see below). The sequence of 1,784 bp within

Abbreviations: AGM, African green monkey; kbp, kilobase pair(s); bp, base pair(s); SV40, simian virus 40.

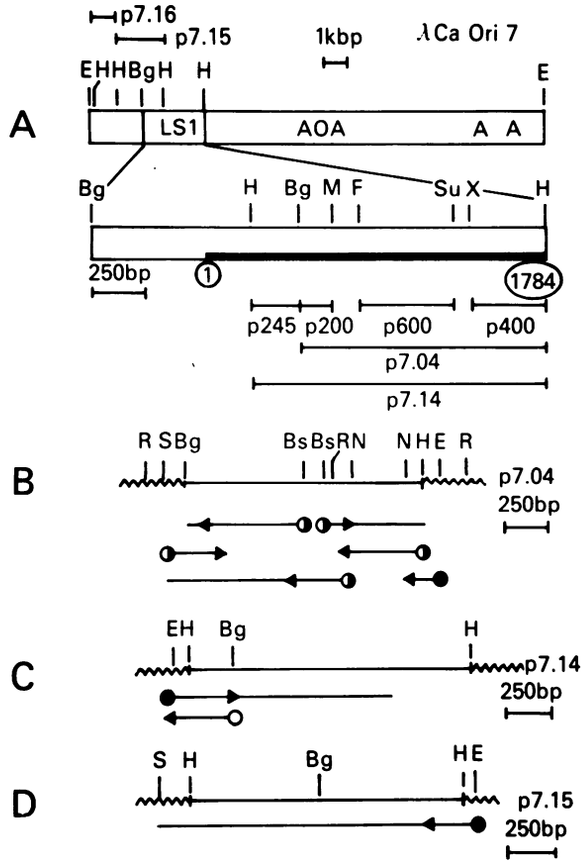

FIG. 1. Restriction endonuclease map, description of subclones, and sequence analysis strategy for LS1. Some of the data were described previously (9–12); however, the map here is given in the opposite direction. (*A*) Partial restriction endonuclease map of λCaOri7 showing the regions homologous to the SV40 origin of replication (O), the Alu sequences (A), and the *Kpn* I family probes (LS1) (the left boundary of the *Kpn* I family member is unknown). The various subclones are indicated as follows (residue numbers correspond to Fig. 2): p7.04 (469–1,783); p7.14 (218–1,783); p7.15 and p7.16 were subcloned from λCaOri7, whereas p200 (469–668), p600 (756–1,294), and p400 (1,413–1,783) were subcloned from p7.04 and p245 was subcloned from the homologous region of RET (its identity to residues 208–434 in LS1 was confirmed by sequence analysis; see Fig. 3*A*). The heavy bar shows the region that was subjected to sequence analysis; encircled numbers correspond to those in Fig. 2. (*B–D*) Sequence analysis strategy for p7.04, p7.14, and p7.15, respectively (wavy lines are pBR322); filled circles indicate 5′-end labeling and open circles indicate 3′-end labeling, with half-filled circles indicating sequence analysis of both strands. The arrowheads show the limit of sequence determination. Bg, *Bgl* II; Bs, *Bst*N1; E, *Eco*RI; F, *Fnu*4H1; H, *Hin*dIII; M, *Mbo* II; N, *Nco* I; R, *Rsa* I; S, *Sal* I; Su, *Sau*3A1; and X, *Xmn* I.

LS1 was determined according to the strategy outlined in Fig. 1 *B–D* and is shown in Fig. 2. It contains 61% A·T base pairs, little more than AGM DNA itself (25), but is asymmetric in that 40% of the bases in the strand shown in Fig. 2 are adenines. There are within the sequence several direct repeats, symmetries, and dyad symmetries that are long enough and good enough matches to be statistically significant but their biological significance, if any, is unknown. Most notably, there is a 61% homology between residues 323–394 and 1,443–1,506. Several sequences typical of transcriptional control elements are present on both strands but their arrangement relative to one another appears to be random. Analysis of the coding potential in all three reading frames on both strands shows numerous stop codons; the longest open reading frame is about 246 bp. One marked feature of the sequence is the 40-residue-long A-rich stretch at the 3′ end (as written).

Computer searches (22, 23) showed no statistically significant homologies between LS1 and Alu sequences (24, 26), 7SL RNA sequences (27, 28), or any sequence in the Los Alamos Data Bank for all vertebrates as of December 1982.

Comparison Between LS1 and Other Cloned *Kpn* I Family Members. α7 (Fig. 3*A*) is typical of several long (>6 kbp) human (5) and monkey (6) *Kpn* I family members. It includes the characteristic 1.5- and 1.2-kbp *Kpn* I fragments (in that order, left to right) and contains at least 700 bp homologous to the 1.8-kbp *Kpn* I fragment mapped to the far left end of many family members (5, 6). Also, α7 is interrupted by an Alu sequence. Restriction endonuclease digests of phage λCaα7 were electrophoresed, transferred to nitrocellulose, and hybridized to ^{32}P-labeled subclones p245, p200, p600, and p400, which represent sequences in LS1 (Fig. 1). The results (not shown) indicated that sequences in all four probes are present in α7 in an order consistent with the arrangement in LS1 (Fig. 3*A*). The experiments did not determine the relative positions of p400 and p600 sequences in the 2.6-kbp *Bam*HI/*Eco*RI fragment of α7. Placement of the p600 segment (Fig. 3*A*) is based on nucleotide sequence comparison (see below), whereas the precise location of p400 remains unknown. The sequences represented in p400 and p600 with the Alu sequence only account for 1.4 kbp of the 2.6 kbp in the *Bam*HI/*Eco*RI fragment of α7; thus, α7 must contain at least 1 kbp of sequence that is not in LS1. Because none of the LS1 probes hybridized to restriction endonuclease fragments from the left end of α7, these too are missing.

The complete sequence of the 829 bp in RET has been reported (8) and is compared schematically to LS1 and α7 in Fig. 3*A*. The entire sequence of RET is present in LS1 but 731 bp of LS1 (residues 651–1,381, Fig. 2) are missing from RET and the segment on one side of the deletion is inverted in RET. Thus, residues 208–650 of LS1 are 86% homologous to residues 446–1 of RET and homology is again detected from residue 1,382 through the end of LS1 where it matches (77%) residues 449–829 of RET.

A part of the nucleotide sequence of α7 was analyzed to determine if it is colinear with LS1 or with RET in the region of the deletion and inversion in RET. The α7 sequence is shown in Fig. 2 and the pertinent comparisons are outlined in Fig. 3*B*; the partial sequence of α7 is colinear with residues 313–807 of LS1 (87% homology). Because α7 is similar to other family members (6) we conclude that RET is the unusual member.

We also compared (22, 23) the sequence of LS1 with that of known portions of human *Kpn* I family members (refs. 2 and 7 and A. Nienhuis, personal communication). As expected from the positions of these segments on restriction endonuclease maps, we found no extensive long homologies to LS1. However, two previously unidentified human DNA segments that were cloned at random (29) from bulk repeated DNA are homologous to LS1. Residues 1,427–1,767 in LS1 are 73% homologous to the 325 bp of pPD16 and residues 389–650 are 73% homologous to the non-Alu portion of BLUR 14 (29). Thus, both of these cloned human segments represent portions of *Kpn* I family members.

Hybridization of LS1 Sequences to Digests of Total AGM DNA. Approximately 4 kbp of the sequences in α7 are missing from LS1, including segments at both the left and right side of the α7 map (see Fig. 3*A*). To confirm the conclusion that many family members are longer than LS1 on both sides, we hybridized subcloned portions of LS1 to endonuclease digests of total AGM DNA. The probe p245 (Fig. 4, lane D′) hybridizes to all of the typical genomic *Kpn* I fragments (1.2, 2.8, 3.4, and 4.6 kbp) except the 1.5-kbp fragment. In contrast, p7.04 (Fig. 4, lane A′) hybridizes to none of them. Because p7.15 (Fig. 1*A*,

FIG. 2. The nucleotide sequence of the portion of LS1 indicated by the black bar in Fig. 1*A*. The sequences read 5′ to 3′ left to right and top to bottom. The sequence of a part of $\alpha 7$ (see Fig. 3*A*) is shown on the line below (dots indicate identical bases); it was obtained by cleaving λCα7 (6) with *Bam*HI, labeling 3′ termini, and cleaving with *Eco*RI to obtain the desired fragments. P is a pyrimidine, X is undetermined, O is a deletion, and V is an insertion.

top), which contains the junction between LS1 and unrelated cellular sequences, also did not hybridize to the 1.5-kbp fragment in $\alpha 7$ (not shown), we conclude that the *Kpn* I family sequences in LS1 cannot extend more than about 50 bp into the 1.5-kbp *Kpn* I fragment.

Our attempt to demonstrate that many family members include sequences that are missing at the right side of LS1 was not satisfactory. In earlier work, probes from the right-hand portion of $\alpha 7$ gave only a smear of hybridization to AGM DNA digested with various endonucleases (6). Similarly, p7.04 gave only a smear with *Kpn* I (Fig. 4, lane A′) and *Hin*dIII (not shown) digests. In the course of studying additional enzymes we found that endonuclease *Msp* I produces discrete repeated genomic fragments, some of which include *Kpn* I family sequences; fragments that are abundant enough to be seen on stained gels (although they may not be easily visible in Fig. 4, lanes B and E) are 1.6, 1.9, 2.2, 3.6, and 4.2 kbp long. Of these, the 2.2- and 4.2-kbp fragments hybridize to p7.04 (Fig. 4, lane B′) and p245 (Fig. 4, lane E′) as well as to p200, p600, and p400 (not shown). The regions included in the 2.2- and 4.2-kbp fragments encompass most, if not all, of the sequence in LS1. All probes also hybridize to a smear of larger *Msp* I fragments. We do not know if the 1.6- and 3.6-kbp *Msp* I fragments are related to the *Kpn* I family but the 1.9-kbp fragment hybridizes to other *Kpn* I family probes (unpublished data). Notably, much less of the bands in question appears upon digestion of AGM DNA with *Hpa* II compared to *Msp* I (Fig. 4, lane C′), indicating that many of the corresponding C-C-G-G sequences are methylated (30).

Transcription of *Kpn* I Family Sequences. We have detected transcripts that hybridize with sequences in LS1 by several different methods. First, labeled total monkey liver nuclear RNA hybridizes to the monkey segment in digests of p7.04 (Fig. 5*A*). Similarly labeled poly(dT)-selected total liver RNA and poly(dT)-selected cytoplasmic RNA from CV-1 cells also hybridized (data not shown). Second, RNA sequences homologous to sequences from residues 469–1,104 and from residues 1,227–1,784 in LS1 were detected in equal abundance by the contact hybridization method (not shown) (31). Third, HeLa cell RNA between about 400 and 9,000 nucleotides long hybridizes to p7.14 (Fig. 5*B*, lane 1). The hybridizable material disappeared upon treatment with RNase (Fig. 5*B*, lane 2) but not with DNase (Fig. 5*B*, lane 3). Finally, two cloned plasmids, pcDK1 and pcDK2, that hybridized with the sequence cloned in p7.14 were isolated from a cDNA library prepared with total poly(dT)-selected RNA of primary human fibroblasts (18). The plasmids were cleaved with *Xho* I, which releases the insert along with about 210 bp of flanking vector sequence (18). Hybridization of the segment cloned in p7.14 to the inserts in pcDK1 and pcDK2 is shown in Fig. 5*C*; the calculated sizes of the inserts are approximately 800 and 250 bp, respectively.

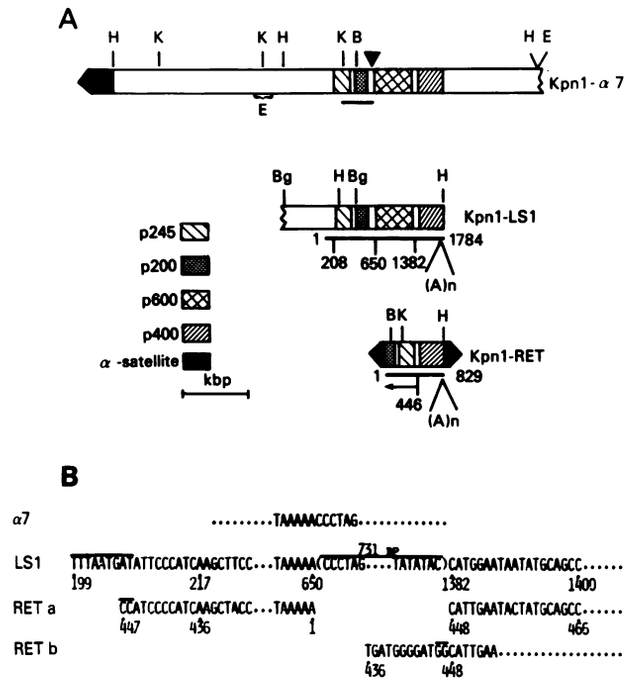

FIG. 3. Comparison of LS1, RET, and α7. (*A*) Schematic diagrams. Similar markings indicate homologous regions as determined by nucleotide sequence comparison or hybridization of cloned probes to restriction endonuclease digests. Underlined black bars are sequenced regions. The residue numbers correspond to the sequence of LS1 (Fig. 2) and RET (8), respectively. The arrow indicates the opposite direction of a portion of the RET compared to α7 and LS1 and the triangle (▼) indicates the position of the Alu sequence in α7 (unpublished data). The filled segments at the ends of α7 and RET indicate junctions with α-satellite. (*B*) Sequences close to the discontinuity in RET. All sequences are shown 5′ to 3′. The sequence of RET is given twice to show (a) homology with LS1 and (b) contiguous sequence of RET itself. Overlines mark sequences in LS1 or RET that do not appear in both family members. B is *Bam*HI; other nuclease symbols as in Fig. 1.

DISCUSSION

Only a few members of the *Kpn* I family have been characterized in any detail thus far. Random selection of family members from both human (1, 5) and AGM (6, 8) genomes has turned up family members that are more than 6 kbp long as well as shorter ones. Most or all segments within the long members are repeated about 10^4 times in both primate genomes but none appears to be repeated within the long members themselves (1, 3–6). We report here that transcripts homologous to *Kpn* I family sequences occur in human and monkey cells. A previous report described the inclusion of *Kpn* I family sequences in transcripts that were synthesized by RNA polymerase III *in vitro* upon initiation at an upstream Alu sequence (32). Altogether, too little is now known about transcription to warrant discussion at this time.

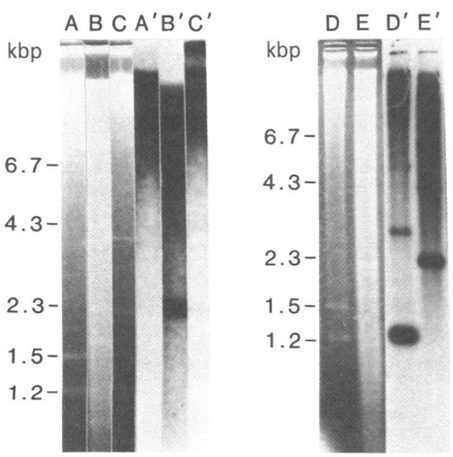

FIG. 4. Hybridization of subclones of LS1 to restriction endonuclease digests of AGM DNA. Each digest represents 10 μg of DNA electrophoresed on 1% agarose. Lanes A–E are ethidium-stained gels, whereas lanes A′–E′ are the corresponding autoradiograms (exposure times: lanes A–C, 9 hr with screen at −70°C; lanes D and E, 2 days at 23°C). Lanes: A, *Kpn* I digest with p7.04 probe; B, *Msp* I digest with p7.04 probe; C, *Hpa* II digest with p7.04 probe; D, *Kpn* I digest with p245 probe; and E, *Msp* I digest with p245 probe.

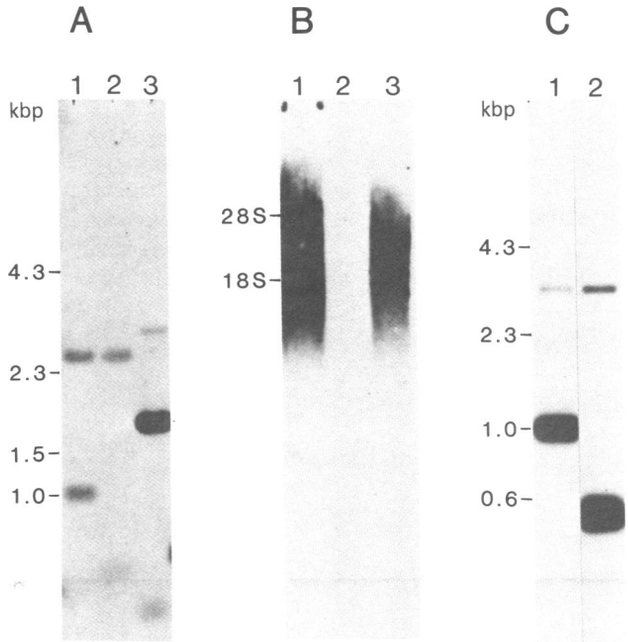

FIG. 5. Hybridization of subcloned regions of LS1 to monkey and human transcripts. (*A*) Total nuclear RNA from monkey liver was digested briefly with alkali, end-labeled with [γ-^{32}P]ATP and polynucleotide kinase, and used as a probe (2×10^7 cpm/μg) against: lane 1, p7.04 digested with *Rsa* I (see Fig. 1*B*); lane 2, p7.04 digested with *Rsa* I and *Hin*dIII (see Fig. 1*B*); and lane 3 (control), p7.06 (11), which contains an Alu sequence, digested with *Acc* I. (*B*) Total RNA from HeLa cells was electrophoresed on a 1.3% agarose/methylmercury gel, transferred to nitrocellulose, and hybridized with p7.14; the position of the rRNA on the ethidium bromide-stained gel is indicated: lane 1, untreated sample; lane 2, treated with RNase before electrophoresis; and lane 3, treated with RNase-free DNase I before electrophoresis. (*C*) DNA from pcDK1 (lane 1) and pcDK2 (lane 2) was digested with *Xho* I, electrophoresed on 1% agarose, transferred to nitrocellulose, and hybridized with the partially purified insert of p7.14. The insert was contaminated with pBR322 vector, thus accounting for the residual hybridization to the 3-kbp vector (pcD) fragment.

More structural information is required before serious consideration can be given to questions concerning the genomic amplification, dispersal, and maintenance of long family members. However, the structures of the shorter members, RET (8) and LS1 (this paper), raise distinctive questions. Both lack sequences from each end of the long family members. At the 3′ end, as written, both terminate in homologous segments, including an A-rich stretch (see Fig. 3*A*); LS1 contains more of the 5′ side of long *Kpn* I family members than does RET. Also, a 731-bp long internal portion of LS1 is absent in RET and a segment on one side of the deletion is inverted. Comparison with α7, a typical long family member, suggests that it is RET that is unusual; α7 and LS1 are colinear in this region.

There are several ways to account for this situation. *Kpn* I

family members might be constructed from discrete elements that are joined in a scrambled fashion at different genomic localities, analogous to the scrambled clusters of short repetitive elements in chickens (33) and *Drosophila* (34); the mechanism by which such clusters form remains obscure. Another possibility is that all family members might have been typical long units at first, with some being shortened by deletions both at the ends and within the element itself. An inversion might arise if a deletion occurred between two neighboring family members with opposite orientations and other mechanisms are also possible; however, site-specific inversions of the sort associated with *Salmonella* flagellar phase variation (35) and the G segment of phage Mu (36) are unlikely, given the absence of inverted repeats at the two ends of the inverted segment in RET (8). Recently, unusual deletions and inversions were observed within DNA regions associated with immunoglobulin heavy chain genes (37, 38); multiple recombinational events were proposed to account for these rearrangements and these depend on specific short repeated sequences within joining and switch regions. There are, close to position 650 of LS1 [the inversion point in RET (Fig. 3*B*)], three repeats of sequences similar to one part of the V–J joining sequence (5′-A-C-A-A-A-A-A-C-C). However, because *Kpn* I family members are polymorphic, the precursor of RET may have had quite a different sequence in this region.

Other possibilities involve RNA intermediates or a combination of reactions at both the RNA and DNA level. A mechanism involving transcription followed by reverse transcription and insertion back into a new position in the genome has been proposed to explain the dispersal and abundance of Alu sequences (39, 40) and processed pseudogenes (40, 41). The detection of transcripts homologous to LS1, the presence of A-rich stretches at one 3′ end of both LS1 and RET, and the fact that RET is flanked by direct tandem repeats of the target site (8) make it possible that RET and LS1 are processed genes. The left-hand end of LS1 has not yet been defined so we do not know if it too is flanked by target-site duplications. This class of models, which we favor, implies that at least some *Kpn* I family members include active transcription units and raises the possibility that the large deletion in RET represents the loss of an intervening sequence. However, splicing of introns is not known to lead to inversions so we would need to postulate an independent inversion event, perhaps at the DNA level. Thus, the unusual structure of RET could be the end result of a complex series of recombinational events, whereas that of LS1 might represent a simpler processed gene.

We are grateful to G. Grimaldi, D. Lipman, and J. Saffer for assistance and critical discussions, to H. Okayama and P. Berg for the cDNA library, to M. Zasloff and R. Billiter for assistance, to Charles Mock for photography, and to May Liu, who prepared the manuscript.

1. Adams, J. W., Kaufman, R. E., Kretschmer, P. J., Harrison, M. & Nienhuis, A. W. (1980) *Nucleic Acids Res.* **8,** 6113–6128.
2. Manuelidis, L. (1982) *Nucleic Acids Res.* **10,** 3211–3219.
3. Manuelidis, L. & Biro, P. (1982) *Nucleic Acids Res.* **10,** 3221–3239.
4. Shafit-Zagardo, B., Maio, J. J. & Brown, F. L. (1982) *Nucleic Acids Res.* **10,** 3175–3193.
5. Shafit-Zagardo, B., Brown, F. L., Maio, J. J. & Adams, J. W. (1982) *Gene* **20,** 397–407.
6. Grimaldi, G. & Singer, M. F. (1983) *Nucleic Acids Res.* **11,** 321–338.
7. Forget, B. C., Tuan, D., Biro, P. A., Jagadeeswaran, P. & Weissman, S. M. (1981) *Trans. Assoc. Am. Physiol.* **25,** 204–210.
8. Thayer, R. E. & Singer, M. F. (1983) *Mol. Cell. Biol.*, in press.
9. McCutchan, T. F. & Singer, M. F. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 95–99.
10. Queen, C., Lord, S. T., McCutchan, T. F. & Singer, M. F. (1981) *Mol. Cell. Biol.* **1,** 1061–1068.
11. Saffer, J. D. & Lerman, M. I. (1983) *Mol. Cell. Biol.* **3,** 960–964.
12. Grimaldi, G., Queen, C. & Singer, M. F. (1981) *Nucleic Acids Res.* **9,** 5553–5568.
13. Britten, R. J., Graham, D. E. & Neufeld, B. R. (1974) *Methods Enzymol.* **38,** 281–298.
14. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, K. & Quon, D. (1978) *Cell* **15,** 687–701.
15. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
16. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
17. Glisin, V., Crkvenjakov, R. & Byus, C. (1974) *Biochemistry* **13,** 2633–2637.
18. Okayama, H. & Berg, P. (1983) *Mol. Cell. Biol.* **3,** 280–289.
19. Thayer, R. E. (1979) *Anal. Biochem.* **98,** 60–63.
20. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
21. Smith, D. R. & Calvo, J. M. (1980) *Nucleic Acids Res.* **8,** 2255–2274.
22. Queen, C. & Korn, L. (1980) *Methods Enzymol.* **65,** 595–609.
23. Wilbur, J. & Lipman, D. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726–730.
24. Jelinek, W. R. & Schmid, C. W. (1982) *Annu. Rev. Biochem.* **51,** 813–844.
25. Maio, J. J. (1971) *J. Mol. Biol.* **56,** 579–595.
26. Singer, M. F. (1982) *Int. Rev. Cytol.* **76,** 67–112.
27. Ullu, E. & Melli, M. (1982) *Cell* **29,** 195–202.
28. Li, W.-Y., Reddy, R., Henning, D., Epstein, P. & Busch, H. (1982) *J. Biol. Chem.* **257,** 5136–5142.
29. Deininger, P. L., Jolly, D. J., Rubin, C. M., Friedman, T. & Schmid, C. W. (1981) *J. Mol. Biol.* **151,** 17–33.
30. Waalwijk, C. & Flavell, R. A. (1978) *Nucleic Acids Res.* **5,** 3231–3236.
31. Zasloff, M. & Santos, T. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5668–5672.
32. Manley, J. L. & Colozzo, M. T. (1982) *Nature (London)* **300,** 376–379.
33. Musti, A. M., Sobieski, D. A., Chen, B. B. & Eden, F. C. (1981) *Biochemistry* **20,** 2989–2999.
34. Wensink, P. C., Tabata, S. & Pachi, C. (1979) *Cell* **18,** 1231–1246.
35. Simon, M., Zieg, J., Silverman, M., Mandel, G. & Doolittle, R. (1980) *Science* **209,** 1370–1374.
36. Kamp, D., Kahmann, R., Zipser, D., Broker, T. R. & Chow, L. T. (1978) *Nature (London)* **271,** 577–580.
37. Alt, F. W. & Baltimore, D. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4118–4122.
38. Greenberg, R., Lang, R. B., Diamond, M. S. & Marcu, K. B. (1982) *Nucleic Acids Res.* **10,** 7751–7761.
39. Jagadeeswaran, P., Forget, B. G. & Weissman, S. M. (1981) *Cell* **26,** 141–142.
40. Van Arsdell, S. W., Denison, R. A., Bernstein, L. B., Weiner, A. M., Mancer, T. & Gesteland, R. F. (1981) *Cell* **26,** 11–17.
41. Denison, R. A. & Weiner, A. M. (1982) *Mol. Cell. Biol.* **2,** 815–828.