Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 2072–2076, April 1985
Genetics

Transposon-like sequences in extrachromosomal circular DNA from mouse thymocytes

(small polydisperse circular DNA/cloning/R repetitive sequence/flanking direct repeat/immunoglobulin gene)

SHINJI FUJIMOTO, TADASHI TSUDA*, MASAAKI TODA, AND HIDEO YAMAGISHI†

Department of Biophysics, Faculty of Science, Kyoto University, Sakyo-ku, Kyoto 606, Japan

Communicated by Susumu Ohno, November 6, 1984

ABSTRACT **Small polydisperse circular (spc) DNA was isolated from mouse thymocytes and cloned into the *Hin*dIII site of λ vector Charon 7. Fifty-six recombinants from this spc DNA library were analyzed. R repeats, which were originally found near immunoglobulin genes, were enriched in spc DNA clones relative to their representation in the chromosome. In one clone, the R sequence was linked to *Bam* and MIF sequences and the contiguous arrangement was truncated from both ends. In another clone, composite *Bam*/R and R repeats existed as a pair in inverted repeat orientation. Truncation occurred from the 5′ side without affecting the 3′ ends. In both clones, short direct repeats flanked the repeated sequences. The possible role of R sequences in transposition and circular formation is discussed.**

Somatic rearrangement of the immunoglobulin genes (1–3) and T-cell receptor genes (4, 5) is believed to occur in the course of lymphocyte development. All the possible mechanisms of translocation, inversion, and deletion were proposed for variable region κ chain–constant region κ chain (V_κ–C_κ) gene joining (6) and the looping-out excision model was for variable region heavy chain–constant region heavy chain (V_H–C_H) recombination (7). An inverted repeat structure formed between conserved sequences of embryonic joining (J) and V region DNAs was shown as a possible looping-out structure (8, 9). Although the details of the mechanism of gene rearrangement are unknown, extrachromosomal circular DNAs may be produced as products of intramolecular recombination. In fact, we found large-sized extrachromosomal circular DNAs [3–150 kilobase pairs (kbp)], which predominated in primary lymphoid cells (10–12) but not in established cultured cell lines (13, 14). They appeared at the critical stage in the ontogeny of B lymphocytes in chicken bursa (15, 16).

To relate extrachromosomal circular DNAs with abortive recombination in primary lymphoid cells, we constructed a cloned library of extrachromosomal circular DNAs from mouse thymocytes and examined the nucleotide sequences of several DNA molecules. We found that a certain interspersed repetitive sequence was significantly enriched in these cloned DNAs. Sequencing and computer search for homology with known major repeat families showed that the repetitive sequence was a member of the R-repeat family (17, 18).

MATERIALS AND METHODS

Mitochondrial DNA-free small polydisperse circular (spc) DNA was prepared from thymocytes of 120 mice (4 weeks old) as described (12).

*Hin*dIII digests of the spc DNA were cloned into λ Charon 7 phage as described (19) and are referred to as MT001-122. The *in vitro*-packaged phages were plated to form individual plaques without amplifying the phage DNA library to avoid cloning of siblings. As a probe in plaque and blot hybridizations, the cloned *Hin*dIII fragments were recloned into the hybrid plasmid vector pKAT1 derived from pBR322 (14) and are referred to as p001-122. Plasmid pEH2 (a 2.0-kbp *Eco*RI/*Hin*dIII fragment of rDNA cloned into pBR322) was used as a probe of the B1 sequence (ref. 20; Ryo Kominami, personal communication) and p3.5 (a 3.5-kbp *Eco*RI fragment of rDNA cloned into pSV-2) as type 2 *Alu* or M2 sequence (21), respectively. pEH2 carried a 130-base-pair (bp) sequence with 88% homology with the consensus B1 sequence of Krayev *et al.* (22). DNA probes were labeled with ^{32}P by nick-translation (23) to specific activities of 10^7–10^8 cpm/μg. Plaque and Southern blot hybridizations were done as described by Maniatis *et al.* (19). Plaque sizes were similar in every spc DNA clone analyzed so far. Washing was done with 5 mM NaH_2PO_4/1 mM EDTA/0.2% $NaDodSO_4$, pH 7, at room temperature. The DNA sequence was determined by a combination of chemical degradation (24) and dideoxy-chain elongation (25, 26) methods.

RESULTS

Cloning of spc DNA. Mouse thymus lymphocytes contained ≈200 copies of spc DNA per cell (10). Mouse thymus covalently closed circular (ccc) DNAs were obtained at 96% purity (by weight) by use of ATP-dependent DNase (12). Mouse thymus spc DNA, after removal of mitochondrial DNA by *Xho* I digestion, ranged in size from 0.6 to 87 kbp, with a mean length of 16.7 ± 13.0 kbp and a median length of 14.6 kbp (12). Mouse spc DNA fragments obtained by *Hin*dIII digestion were ligated into λ Charon 7 phage vector. In all, 10,400 clear plaque-forming units were recovered, and 99 units were selected at random and further characterized by restriction analysis. Of these phage clones, 56 (MT1-34 and 101-122) were found to have DNA insertions of various lengths (1.0–7.7 kbp) with a mean and a median length of 3.9 kbp (Table 1). Therefore, the intact spc DNAs must be fragmented to an average of 1/5th their original length.

Homology to Chromosomal Sequences and Sequence Abundance on Chromosome. We carried out a plaque hybridization experiment on spc DNA clones with ^{32}P-labeled total mouse thymus genomic DNA. Since spc DNA copies are as low as ≈200 per cell, genomic DNA was not necessarily separated from spc DNA. Various degrees of autoradiographic intensity were observed and were classified arbitrarily into four classes (Table 1): class I of highly repetitive sequences (+++), 2 clones; class II of intermediate repetitive sequences (++), 5 clones; class III of low repetitive

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.

Abbreviations: spc, small polydisperse circular; ccc, covalently closed circular; bp, base pair(s); kbp, kilobase pair(s); C, V, and J, constant, variable, and joining regions, respectively, of immunoglobulin gene.

*Deceased, October 20, 1983.
†To whom reprint requests should be addressed.

Table 1. Properties of cloned spc DNA from mouse thymocyte

Clone	Size of insert, kbp*	Plaque hybridization probe												
		Thymus (genome)	pEH2 (B1)	p104 (++)	p104-C (++)	p014 (+++)	p010 (++)	p114 (+)	p004 (+)	p009 (−)	p011 (−)	p032 (−)	p012 (−)	p023 (−)
MT113	7.7	+	±	0	0	+	0	0	0	0	0	0	0	0
114	7.7	+	0	0	0	±	0	++	0	0	++	0	0	0
115	7.7	−	0	0	0	0	+	0	0	0	0	0	0	0
006	7.4	−	0	0	0	0	0	0	0	++	0	0	0	0
009	7.4	−	0	0	0	0	0	0	0	++	0	0	0	0
112	7.3	+	0	0	0	0	0	0	0	0	0	+	0	0
109	4.8 + 2.5	−	0	0	0	0	0	0	0	0	0	+	0	0
104	7.2	++	0	++	++	+	++	0	0	0	0	0	0	0
102	7.0	+	0	0	0	0	0	0	++	0	0	0	0	0
103	4.9 + 2.1	−	0	0	0	0	0	0	0	±	0	0	0	+
004	6.2	+	0	0	0	0	0	0	++	0	0	0	0	0
110	5.8	−	0	++	++	0	0	0	0	0	0	0	0	0
106	5.6	+	0	0	0	0	0	0	0	0	0	0	0	0
005	5.3	+	0	0	0	0	0	0	++	0	0	0	0	0
014	4.8	+++	±	+	++	++	+	0	0	0	0	0	0	0
034	4.7	+++	0	+	++	+	+	0	0	0	0	0	0	0
011	4.7	−	0	0	0	+	0	++	0	0	++	0	0	0
015	4.6	++	0	+	++	++	+	0	0	0	0	±	0	++
026	4.6	++	±	±	+	++	±	0	0	0	0	0	0	0
108	4.6	−	0	0	0	0	0	0	0	0	0	0	0	0
031	4.5	+	0	0	0	0	0	0	++	0	0	0	0	±
107	2.9 + 1.5	−	0	0	0	0	0	0	0	0	0	0	0	0
028	2.7 + 1.6	+	+	0	0	+	0	±	++	0	0	0	0	0
101	4.3	−	0	++	++	0	0	0	0	0	0	0	0	0
010	4.1	++	0	+	++	+	++	0	0	0	0	0	0	0
003	4.0	+	0	0	0	±	0	0	0	0	0	0	0	0
117	4.0	−	0	0	0	++	0	±	0	0	++	0	0	0
027	3.9	−	0	0	0	0	0	0	0	0	0	0	0	0
119	3.9	−	0	0	0	0	0	0	0	0	0	0	0	0
033	3.7	−	0	0	0	0	0	0	0	0	0	0	0	0
017	3.4	−	0	0	0	0	++	0	0	0	0	0	0	0
018	3.4	−	0	0	0	±	++	0	0	0	0	0	0	0
025	3.1	−	0	0	0	0	0	0	0	±	0	0	0	0
029	3.1	−	0	0	0	0	0	0	0	0	0	0	0	0
121	2.5	++	0	0	0	+	+	0	0	0	0	0	0	0
008	2.5	−	0	0	0	0	0	0	0	0	0	0	0	0
032	2.5	−	0	0	0	0	0	0	0	0	0	++	0	0
116	2.3	−	0	0	0	0	0	0	0	0	0	0	0	0
012	2.2	−	0	0	0	+	0	0	0	0	0	0	++	0
013	2.2	−	0	0	0	+	0	0	0	0	0	0	++	0
111	2.2	−	0	0	0	0	0	0	0	0	0	0	0	0
105	2.1	−	0	0	0	0	0	0	0	±	0	0	0	0
001	2.0	−	0	0	0	0	±	0	0	0	0	0	0	0
022	2.0	−	0	0	0	0	0	0	0	0	0	0	0	0
007	1.9	−	0	0	0	0	0	0	0	0	0	+	0	0
023	1.9	−	0	0	0	0	0	0	0	±	0	0	0	++
002	1.8	−	0	0	0	0	0	0	0	0	0	0	0	0
016	1.8	−	0	0	0	+	0	0	0	0	0	0	0	0
024	1.8	−	0	0	0	0	0	0	0	0	0	0	0	0
118	1.5	+	0	0	0	0	0	0	0	0	0	0	0	0
019	1.5	−	0	0	0	0	0	0	0	0	0	0	0	0
030	1.2	−	0	0	0	0	0	0	0	0	0	0	0	0
122	1.2	−	0	0	0	0	0	0	0	0	0	0	0	0
020	1.1	−	0	0	0	0	0	0	0	0	0	0	0	0
021	1.1	−	0	0	0	0	0	0	0	0	0	0	0	0
120	1.0	−	0	0	0	0	0	0	0	0	0	0	0	0

All probes were ^{32}P-labeled by nick-translation except the p104-C fragment 5′-end-labeled with ^{32}P. +++, Very strong; ++, strong; +, moderate; ±, weak; 0, no detectable hybridization.
*MT109, 103, 107, and 028 contain two inserts.

sequences (+), 11 clones; class IV of very low repetitive sequences or unique sequences (−), 38 clones. Most class IV spc DNA clones showed very weak autoradiographic intensity but significant intensity upon prolonged exposure. Genomic abundance of spc DNA was estimated to be on the order of 10^5 copies for class I, 10^4 copies for class II, 10^3 copies for class III, and $<10^3$ copies for class IV in the genome by comparison with the autoradiographic intensity

of the human DNA clones containing *Alu*, *Kpn* I, and mitochondrial sequence (14).

Homology Among spc DNA Clones. We tested for the presence of a known major repeat family of the B1 sequence and the unidentified p104 repetitive sequence in 56 spc DNA clones (Table 1). The B1 probe (pEH2) showed moderate (+) or weak (±) homology to four clones carrying a repetitive sequence of class I to III (MT028, -113, -014, and -026). The low repetitive sequence of MT028 and -113 may be regarded as a minor member of the B1 sequence. The p104 probe showed strong homology (++) to two clones with a unique or very low repetitive sequence (MT110 and -101), moderate homology (+) to four clones with a repetitive sequence (MT014, -034, -015, and -010), and weak homology (±) to MT026. Hybridizations were stronger with the C fragment of p104 (Fig. 1) as a probe. Repetitive sequence of p104 was interspersed in the genome, as indicated by smeared hybridization in Southern blots (not shown).

Two probes (p014 and p010) showed homologies not only with 6 clones with a p104-repetitive sequence family (MT104, -014, -034, -015, -026, and -010), but also with 5 other clones with repetitive sequences (MT113, -114, -028, -003, and -121), and 9 clones with very low repetitive sequences. Of these, 4 clones shared the B1-like sequence. These 20 clones hybridizable with p014 or p010 together with 2 clones with very low repetitive sequences hybridizable with p104 may have been derived from closely related chromosomal regions. The repetitive sequence in these clones is referred to as the p014 family. Five clones hybridizable with p004 (MT102, -004, -005, -031, and -028) may carry another repetitive sequence family (p004 family). The other three clones (MT112, -106, and -118) may carry another uncharacterized repetitive sequence. No clone showed homology to p3.5 carrying M2 sequence (not shown).

Extensive homologies were also found among the various clones with very low repetitive sequences. Six clones shared homology to the probe p009. Four other probes with very low repetitive sequence also showed homology to several spc DNA clones.

Localization of Repetitive DNA Sequences on Two spc Clones, MT104 and -026. Various restriction digests of MT104 and -026 were analyzed by Southern blotting with nick-translated mouse brain DNA as a probe. This probe detects repetitive DNA exclusively, because it lacks sufficient quantities of unique sequences to allow their detection. Fragments containing a repetitive sequence are shown by bidirectional arrows over the restriction maps in Fig. 1. From dot-blot experiments, we estimated the reiteration frequency of the repetitive sequence in the C fragment of MT104 as 2×10^4 per haploid genome (not shown).

Sequence of Repetitive Element of MT104. A portion of the sequence of the B and C fragments of MT104 was determined (Fig. 2). The sequence strategy is shown in Fig. 1. In performing the computer search for homology to the mouse major repetitive sequence families, we unexpectedly found an internal C region of MT104 (303 bp; position, R 170–473) with 92% homology to the 3′ region of the consensus R sequence of Gebhard *et al.* (17). Another repetitive sequence located close to the left end of the C fragment was found to be part of a long *Bam*/R composite sequence, which extends another 268 nucleotides in the 3′ direction into the B fragment. Over the 683-bp regions compared (position, *Bam* 298–507; R 1–473), the two sequences share 600 bases or 88% homology. Other common features that we noted were the presence of a poly(A) addition signal A-A-T-A-A-A (29), an adenine-rich region at the 3′ end, and direct repeats flanking the repetitive sequence. Since the MIF, *Bam*, and R sequences are linked together into a superfamily in this order in the mouse genome (18, 27, 30), the present repetitive elements are truncated forms from the 5′ side of the superfamily. These two truncated repeats are arranged as a pair in inverted repeat orientation (Fig. 1).

Sequence of the Repetitive Element of MT026. By sequencing of the HB, BE, and EB fragments of MT026, the contiguous sequence of the MIF/*Bam*/R superfamily was assigned to the region of the repetitive sequence (Fig. 2). Over the 907-bp region (position, MIF 27–1; *Bam* 1–507; R 1–373), the two sequences share 750 bases or 83% homology. Unexpectedly, the MIF/*Bam*/R superfamily sequence was truncated from both ends. Again flanking, short direct repeats were found in the 3′ flank of the truncated sequence. These are expected to be the target site of the transposition element (31, 32).

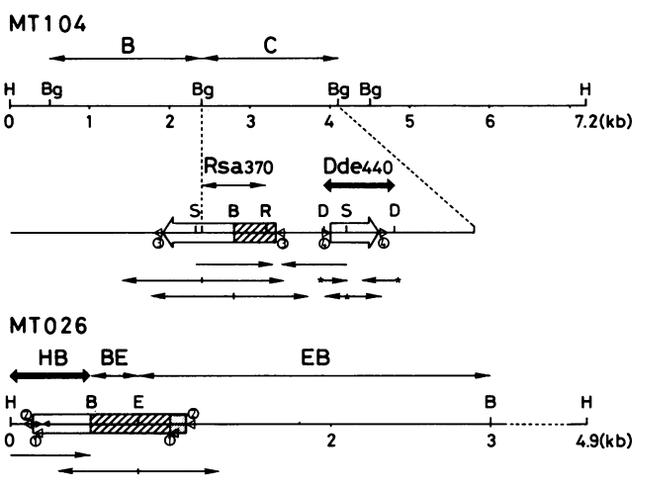

FIG. 1. Restriction maps of two spc DNA clones MT104 and MT026. Restriction fragments containing repetitive sequences are indicated by bidirectional arrows with designation of fragments. Fragment exhibiting homology to the Dde440 fragment is shown by a heavy bidirectional arrow. H, *Hin*dIII; Bg, *Bgl* II; D, *Dde* I; R, *Rsa* I; B, *Bam*HI; S, *Sac* I; E, *Eco*RI. Not all sites for a given enzyme are shown. Single-headed arrows denote region and direction sequenced (5′ to 3′). Most regions were sequenced by the dideoxy-chain elongation method (25) with M13 subclones, but Dde440 regions (asterisks) were sequenced by the chemical degradation method (24). Composite repetitive sequences are shown on a single line by thick arrows or a box, and R (open), *Bam* (hatched), and MIF (stippled) repeats are distinguished. Arrowhead parts represent adenine-rich sequences. Arrows showing the sequence are indicated in the 5′ to 3′ direction, according to the convention used for R sequences by Gebhard *et al.* (17). Pairs of open triangles designated ① to ④ show the occurrence of short direct repeats flanking the composite repetitive sequence. The pair of solid triangles shows the presence of a palindrome structure.

DISCUSSION

spc DNA has been isolated and cloned from the BSC-1 line of African green monkey kidney cells (33), the CHO-K1 line of Chinese hamster ovary cells (34), and HeLa cells (14). The best characterized sequence is the BSC-1 spc DNA containing the *Alu* repetitive sequence (35–37). In *Drosophila melanogaster*, most, if not all, of spc DNA is homologous to middle repetitive DNA (38). Flavell and Ish-Horiwicz (39) found extrachromosomal circular DNAs of the *copia* transposable element in cultured *Drosophila* cells.

Cloned spc DNA from mouse thymus is fragmented to an average of 1/5th the intact circular form. The spc DNA clone homologous to the repetitive sequence, referred to as the p014 family, was found in 11 of 56 clones, or 1/5th the total clones. This finding suggests that each circular molecule of 16.7-kbp in average size retains a partial sequence of the p014-family or R-linked superfamily. Thus, R family se-

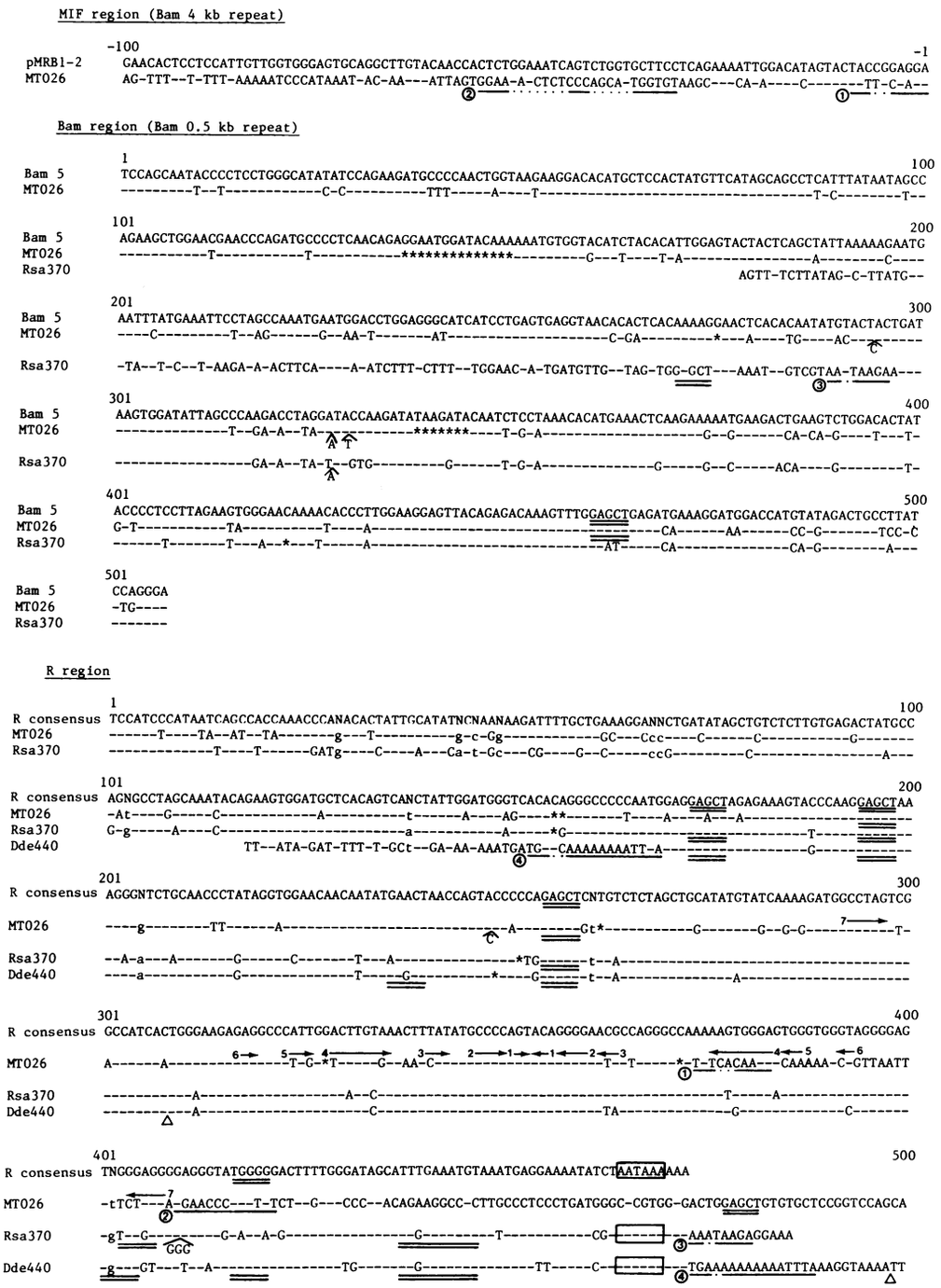

FIG. 2. Repetitive and flanking sequences of MT026 and MT104 clones from mouse thymus spc DNA. Our data on MT026 and the region including Rsa370 and Dde440 from MT104 are compared with those for pMRB1-2 (27), Bam5 (28), and R consensus (17). Sequences are aligned to give maximal homology. Dashes indicate that bases in given positions are unchanged; asterisk indicates that the base is missing; insertions are indicated by an inverted V. A variant nucleotide is named N in the consensus sequence and the corresponding position is shown by a lower case letter. Direct repeats flanking composite repeat are underlined with interruption at nonhomologous positions and designated ① to ④ in pairs. Inverted repeats of special interest are marked by numbered superscript arrows. Boxed nucleotides show a poly(A) addition signal (29). G-A-G-C-T, T-G-G-G-G, and T-T-G-G-A-A-A-T-G-T found in the IgH S region and J_T flank are shown by double underlines. Dde440-sequence homologous to the kallikrein intron is between the two open triangles. MIF/*Bam*/R sequence of MT121 was truncated to the position *Bam*214 from the 3′ side.

quences appeared in the cloned spc DNA at much higher frequency than 2×10^4 per haploid genome or 150-kbp interval on average. For confirmation of the presence of the p014-family sequence in covalently closed spc DNA, partially purified spc DNAs were fractionated by CsCl/EtBr density gradient centrifugation and hybridized to ^{32}P-labeled DNA probe. This gradient dot-blot experiment with p104 and pEH2 revealed two peaks of hybridization, which corresponded in position to covalently closed spc DNA and the linear chromosomal fragment (not shown). Hybridization intensity in ccc DNA was stronger with the probe of p104 than with pEH2. The R-linked superfamily sequence may play a specific role in formation of circular DNA from the thymocyte chromosome.

R-sequence families have homologous RNA polymerase II-dependent transcripts in mouse liver cells (40). At the 3′ end of the R sequence of spc DNA, a run of deoxyadenine residues was found just after the poly(A) addition site A-A-T-A-A-A. This is indicative of the transcription–retrotranscription mechanism proposed by Gebhard *et al.* (17). The retrotranscription may start from the poly(A) tract at the 3′ end and stop prematurely within the R or adjacent sequence. Evidence for R translocation via retrotranscription is the short direct repeats flanking the R sequence, which may be duplicated at a target site for insertion (28, 31). A rearranged sequence of the R element probably caused by transposition has been reported (17, 18). Truncation of the R sequence found in genomic DNA occurs from the 5′ side but never from the 3′ side. In the spc R sequence, however, we observed truncation from the 3′ side. Mouse interspersed repetitive sequences, including the R superfamily, can be isolated from DNA or heterogeneous RNA (double-stranded RNA-A and double-stranded RNA-B) as either intramolecular or intermolecular base-paired duplexes (41, 42). As

shown in the R1/R2-containing fragment 5′ to J_κ (17), the $C_\lambda A/C_\lambda B$-containing fragment 3′ to $C_\lambda 1$ (43) and the MT104 fragment, R sequences were arranged as a pair in inverted repeat orientation. The pair of inverted R sequences was convergent in the first case and divergent in the latter two cases. Truncation from the 3′ side of R repeats may arise by excision of a looped-out structure formed in a pair of divergent inverted repeats. This sort of recombination may generate a new inverted repeat structure. In fact, long stretches of inverted repeat were observed in the region (position, R 293–408) close to the 3′ truncated end. Permuted rearrangement of the *Kpn* I element (44), human long interspersed repeats closely related to the MIF/*Bam*/R superfamily, may also be explained by excision of the circular form.

Although R sequences were originally found near immunoglobulin genes, they have been detected throughout the genome—for instance, near endogenous virogenes (IPA and AKV) and in the β-globin and H2 regions (45). In performing computer search for homology between the Dde440-R sequence and the vertebrate portion of the NIH nucleic acid sequence data bank at Los Alamos, we found a significant homology 190-bp long with an 84% base match between the Dde440-R sequence (position, R 309–498) and the third intron of a mouse kallikrein gene (position, 6458–6647) (46). The homologous region of the kallikrein gene was flanked by two pairs of short direct repeats, G-T-G-G and T-A-T-$(^{T}_{A})$-C-C. Thus, a putative transposable sequence flanked by short direct repeats was not confined to the consensus repetitive sequences.

R sequences were frequently found near immunoglobulin light chain genes, 5′ to $V_\lambda 1$, 5′ to $C_\lambda 3$, 3′ to $C_\lambda 1$, 5′ and 3′ to $V_\lambda 2$, 5′ to $C_\lambda 2$, 3′ to $C_\lambda 4$, 5′ to $V_\kappa n$, 5′ to J_κ and 3′ to C_κ (17, 43). Although R sequences were found at the flanking regions of the V_H gene (30) and the intervening segments between C_δ and $C_\gamma 1$ (unpublished results), their presence near immunoglobulin heavy-chain genes has not been extensively studied. However, R sequences share short common sequences, such as G-A-G-C-T and T-G-G-G-G, with the immunoglobulin class switch regions (47) and the J_T flanking region of the T-cell receptor gene (4). A prevalent sequence of $S_\gamma 2b$, T-T-G-G-A-A-A-T-G-T (47), was also found close to the 3′ end of the R sequence. According to Ohno (48), the 5′ noncoding sequence of each immunoglobulin C_H is made up of multiple copies at various stages of degeneration of one primordial 20-bp-long sequence: (A-G-C-T-G)-(A-G-C-T-G)-(A-G-C-T-G)-(G-G-G-T-G). We found 38 nonoverlapping repeats of these primordial pentamers and single or two-base derivatives in a 304-bp stretch of the Dde440-R sequence. These short homologous repeats may serve as preferred recombination sites with R sequence transposed near immunoglobulin (or T-cell receptor) gene and the unique sequence located between them may be mobilized and abortively rearranged.

We thank Dr. H. Ozeki for his continuous interest, S. Aota for performing the computer search for sequence similarities, and Dr. R. Kominami for providing plasmids pEH2 and p3.5 and for information on the B1 sequence of pEH2 before its publication. This work was supported by Grants-in-Aid for Scientific Research and for Special Project Research from the Ministry of Education, Science and Culture of Japan.

1. Tonegawa, S. (1983) *Nature (London)* **302,** 575–581.
2. Honjo, T. (1983) *Annu. Rev. Immunol.* **1,** 499–528.
3. Hood, L., Davis, M., Early, P., Calame, K., Kim, S., Crews, S. & Huang, H. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45,** 913–923.
4. Chien, Y.-h., Gascoigne, N. R. J., Kavaler, J., Lee, N. E. & Davis, M. M. (1984) *Nature (London)* **309,** 322–326.
5. Saito, H., Kranz, D. M., Takagaki, Y., Hayday, A. C., Eisen, H. N. & Tonegawa, S. (1984) *Nature (London)* **309,** 757–762.
6. Hozumi, N. & Tonegawa, S. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 3628–3632.
7. Honjo, T. & Kataoka, T. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2140–2144.
8. Tonegawa, S., Sakano, H., Maki, R., Traunecker, A., Heinrich, G., Roeder, W. & Kurosawa, Y. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45,** 839–858.
9. Leder, P., Max, E. E., Seidman, J. G., Kwan, S. P., Scharff, M., Nau, M. & Norman, B. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **45,** 859–865.
10. Yamagishi, H., Kunisada, T. & Tsuda, T. (1982) *Plasmid* **8,** 299–306.
11. Tsuda, T., Yamagishi, H., Ohnishi, N., Yamada, Y., Izumi, H. & Mori, K. J. (1983) *Plasmid* **10,** 235–241.
12. Yamagishi, H., Tsuda, T., Fujimoto, S., Toda, M., Kato, K., Maekawa, Y., Umeno, M. & Anai, M. (1983) *Gene* **26,** 317–321.
13. Kunisada, T., Yamagishi, H. & Sekiguchi, T. (1983) *Plasmid* **10,** 242–250.
14. Kunisada, T. & Yamagishi, H. (1984) *Gene* **31,** 223–233.
15. DeLap, R. J. & Rush, M. G. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 5855–5859.
16. Toda, M. & Yamagishi, H. (1984) *Dev. Growth Differ.* **26,** 197–204.
17. Gebhard, W., Meitinger, T., Hochtl, J. & Zachau, H. G. (1982) *J. Mol. Biol.* **157,** 453–471.
18. Gebhard, W. & Zachau, H. G. (1983) *J. Mol. Biol.* **170,** 255–270.
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
20. Kominami, R., Urano, Y., Mishima, Y. & Muramatsu, M. (1981) *Nucleic Acids Res.* **9,** 3219–3233.
21. Kominami, R., Muramatsu, M. & Moriwaki, K. (1983) *Nature (London)* **301,** 87–89.
22. Krayev, A. S., Kramerov, D. A., Skryabin, K., Ryskov, A. P., Bayev, A. A. & Georgiev, G. P. (1980) *Nucleic Acids Res.* **8,** 1201–1215.
23. Rigby, P. W., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
24. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
25. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143,** 161–178.
26. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26,** 101–106.
27. Meunier-Rotival, M. & Bernardi, G. (1984) *Nucleic Acids Res.* **12,** 1593–1608.
28. Fanning, T. G. (1982) *Nucleic Acids Res.* **10,** 5003–5013.
29. Fitzgerald, M. & Shenk, T. (1981) *Cell* **24,** 251–260.
30. Benett, K. L. & Hastie, N. D. (1984) *EMBO J.* **3,** 467–472.
31. Calos, M. P. & Miller, J. H. (1980) *Cell* **20,** 579–595.
32. Temin, H. M. (1980) *Cell* **21,** 599–600.
33. Bertelsen, A. H., Humayun, M. Z., Karfopoulos, S. G. & Rush, M. G. (1982) *Biochemistry* **21,** 2076–2085.
34. Stanfield, S. W. & Helinski, D. R. (1984) *Mol. Cell. Biol.* **1,** 173–180.
35. Krolewski, J. J., Bertelsen, A. H., Humayun, M. Z. & Rush, M. G. (1982) *J. Mol. Biol.* **154,** 399–415.
36. Krolewski, J. J. & Rush, M. G. (1984) *J. Mol. Biol.* **174,** 31–40.
37. Krolewski, J. J., Schindler, C. W. & Rush, M. G. (1984) *J. Mol. Biol.* **174,** 41–54.
38. Stanfield, S. W. & Lengyel, J. A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 6142–6146.
39. Flavell, A. J. & Ish-Horowicz, D. (1981) *Nature (London)* **292,** 591–595.
40. Heller, D., Jackson, M. & Leinwand, L. (1984) *J. Mol. Biol.* **173,** 419–436.
41. Kramerov, D. A., Ryskov, A. P. & Georgiev, G. P. (1977) *Biochim. Biophys. Acta* **475,** 461–475.
42. Jelinek, W. R., Toomey, T., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L. & Schmid, C. W. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 1398–1402.
43. Wilson, R. & Storb, U. (1983) *Nucleic Acids Res.* **11,** 1803–1817.
44. Potter, S. S. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1012–1016.
45. Lueders, K. K. & Paterson, B. M. (1982) *Nucleic Acids Res.* **10,** 7715–7729.
46. Mason, A. J., Evans, B. A., Cox, D. R., Shine, J. & Richards, R. I. (1983) *Nature (London)* **303,** 300–307.
47. Kataoka, T., Miyata, T. & Honjo, T. (1981) *Cell* **23,** 357–368.
48. Ohno, S. (1981) *Differentiation* **18,** 65–74.