JOURNAL OF VIROLOGY, June 1986, p. 937–944
0022-538X/86/060937-08$02.00/0

Vol. 58, No. 3

Molecular Cloning and Long Terminal Repeat Sequences of Human Endogenous Retrovirus Genes Related to Types A and B Retrovirus Genes

MASAO ONO

Department of Molecular Biology, School of Medicine, Kitasato University, Sagamihara-shi, Kanagawa 228, Japan

Received 18 December 1985/Accepted 12 March 1986

By using a DNA fragment primarily encoding the reverse transcriptase (*pol*) region of the Syrian hamster intracisternal A particle (IAP; type A retrovirus) gene as a probe, human endogenous retrovirus genes, tentatively termed HERV-K genes, were cloned from a fetal human liver gene library. Typical HERV-K genes were 9.1 or 9.4 kilobases in length, having long terminal repeats (LTRs) of ca. 970 base pairs. Many structural features commonly observed on the retrovirus LTRs, such as the TATAA box, polyadenylation signal, and terminal inverted repeats, were present on each LTR, and a lysine (K) tRNA having a CUU anticodon was identified as a presumed primer tRNA. The HERV-K LTR, however, had little sequence homology to either the IAP LTR or other typical oncovirus LTRs. By filter hybridization, the number of HERV-K genes was estimated to be ca. 50 copies per haploid human genome. The cloned mouse mammary tumor virus (type B) gene was found to hybridize with both the HERV-K and IAP genes to essentially the same extent.

All vertebrates, including human beings, are presumed to have endogenous retrovirus genes transmitted from parents to offspring (12). However, we have yet to fully understand the types, structures, functions, and causal relationships of these endogenous retrovirus genes to tumors. Intracisternal A particle (IAP; type A retovirus) genes are moderately repetitive endogenous retrovirus genes interspersedly present in rodents such as mice, rats, and Syrian hamsters (18). Since the integration of a murine IAP gene has been reported to cause either activation of a proto-oncogene, c-*mos* (9), or inactivation of an actively transcribed kappa immunoglobulin gene (16), some endogenous retrovirus genes which can transpose as insertional mutagens may thus cause genetic diseases such as cancer.

Although the sequence homologous to the mouse IAP gene has been detected in DNAs from mink, raccoon, and monkey cells, as well as in DNAs from rodents (12), the presence of a sequence homologous to the rodent IAP gene has yet to be detected in human cells. Our recent determination of the complete nucleotide sequence of the Syrian hamster IAP gene not only led to an understanding of a typical LTR-*gal-pol-env*-LTR structure of the IAP gene but also confirmed the presence of conserved *pol* regions among retrovirus genomes and revealed a close evolutionary relationship of the type A retrovirus gene to types B and D oncovirus genes (29).

By using the fragment primarily encoding the conserved *pol* region of the Syrian hamster IAP gene as a probe, I was able to detect and isolate the sequence hybridizable with this probe in normal human DNA. Typical human endogenous retrovirus (HERV) genes were 9.1 or 9.4 kilobases (kb) in length and contained typical long terminal repeat (LTR) sequences of ca. 1.0 kb at both ends. Approximately 50 copies of these HERV genes were present per haploid human genome and were similarly homologous to the mouse mammary tumor virus (MMTV; type B retrovirus) gene in addition to the type A retrovirus gene. Since the lysine (K) tRNA was identified as a presumed primer for reverse transcription, I tentatively classified them as an HERV-K family.

MATERIALS AND METHODS

Materials and clones. Restriction endonucleases, T4 ligase, T4 polynucleotide kinase, and terminal deoxynucleotidyl transferase were obtained from Takara Shuzo Co. Digestion of the endonucleases was carried out as recommended in instructions of the supplier. The cloned IAP gene, IAP-H18, from the Syrian hamster was described previously (29, 37). Cloned MMTV (5), Rous sarcoma virus (RSV; clone SRA-2) (13), and Friend murine leukemia virus (24) genes in pBR322 were kindly provided by H. Diggelmann (Swiss Institute of Experimental Cancer Research, Lausanne, Switzerland), J. M. Bishop (University of California, San Francisco), and Y. Ikawa (Institute of Physical and Chemical Research, Japan), respectively.

Preparation of DNAs and screening. Human DNAs from peripheral blood leukocytes and chicken liver DNA were prepared as described previously (2). Southern blot transfer, nick translation, and filter hybridization were carried out as described previously (26). In the present work, hybridization carried out at 65°C is referred to as the stringent condition, and that carried out at 50°C is referred to as the less stringent condition. After hybridization, the filter was washed five times (10 min per wash) in 0.3× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate at 65°C (stringent washing condition) or 1× SSC–0.1% sodium dodecyl sulfate at 50°C (less stringent condition). A human fetal liver DNA library described by Lawn et al. (17) was kindly provided by T. Maniatis (Harvard University, Cambridge, Mass.) and screened under the less stringent hybridization and washing conditions by using the IAP-H18 fragment containing mainly the *pol* region (probe a; see Fig. 2C) as a probe. Individual clones were plaque purified twice, and some lambda DNAs containing HERV-K genes were digested with *Eco*RI, *Hin*dIII, or *Bam*HI and then subcloned into pBR322.

DNA sequencing. DNA fragments were labeled either at 5′ ends with [γ-^{32}P]ATP and T4 polynucleotide kinase or at 3′ ends with [α-^{32}P]ddATP and terminal deoxynucleotidyl transferase. The nucleotide sequence of the fragments was

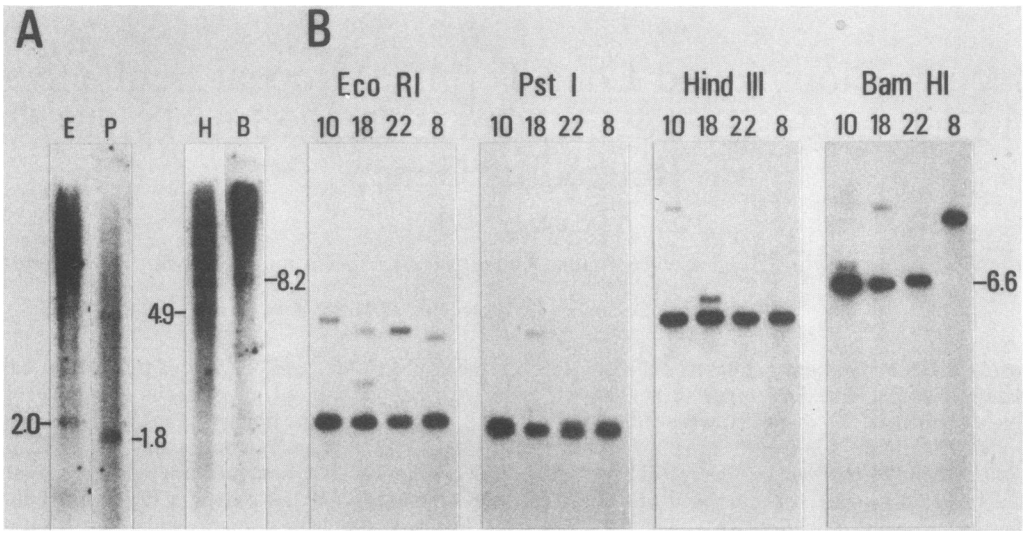

FIG. 1. Identification and cloning of IAP-gene-related sequences in a human genome. (A) Human DNAs (2.0 μg) prepared from peripheral blood leukocytes or (B) cloned lambda DNAs (0.25 μg) were digested with restriction enzymes and analyzed by the Southern hybridization method under the less stringent condition with probe a, generated from the Syrian hamster IAP gene (see legend to Fig. 2C). After hybridization, the filter was washed under the less stringent condition. (A) Lanes: E, *Eco*RI digest; P, *Pst*I digest; H, *Hin*dIII digest; B, *Bam*HI digest. (B) Hybridization with cloned lambda DNAs. Fragment size, indicated in kilobases, was estimated by *Hin*dIII digests of wild-type lambda DNA as standards.

determined by the method of Maxam and Gilbert (21). All sequences shown were determined in both DNA strands.

RESULTS

Cloning of HERV-K genes. Restriction enzyme-digested human DNAs from peripheral blood leukocytes were Southern blot transferred and hybridized with the ^{32}P-labeled *pol* fragment of IAP-H18 under the less stringent condition. After being washed under the less stringent condition, several discrete bands corresponding to 2.0 kb by *Eco*RI digestion, 1.8 kb by *Pst*I digestion, 4.9 kb by *Hin*dIII digestion, or 8.2 kb by *Bam*HI digestion were observed in addition to a smeared background (Fig. 1A). Since these discrete bands were detected in DNAs from all three males and three females, the general presence of the sequences corresponding to these bands in the human genome was expected, and the sequences from the human gene library were cloned.

From 9 × 10^4 plaques screened under the less stringent hybridization and washing conditions, 26 positives were

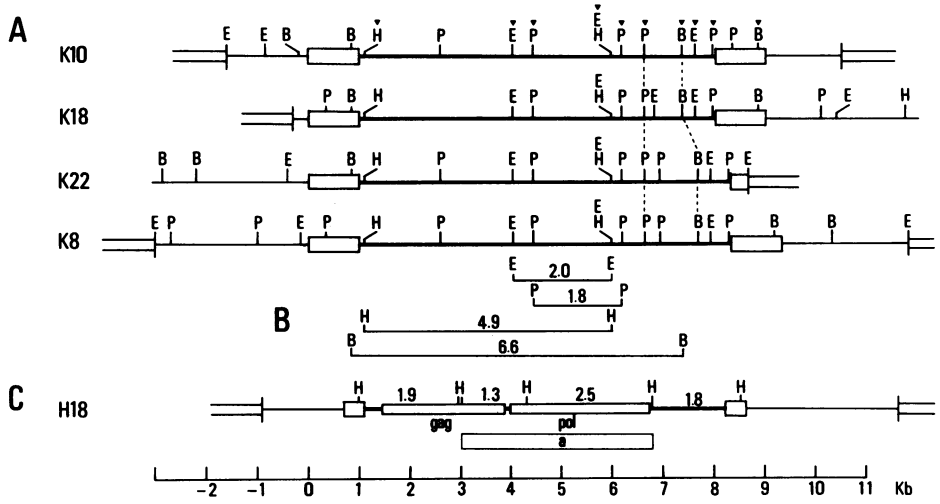

FIG. 2. Restriction maps of cloned HERV-K genes. (A) Relative positions of the restriction enzyme sites determined by digestion of subcloned DNAs with *Eco*RI (E), *Pst*I (P), *Hin*dIII (H), and *Bam*HI (B). These cloned sequences were aligned according to the common restriction enzyme sites. The orientation of the HERV-K genes was determined on the basis of the LTR sequences shown in Fig. 3. Restriction enzyme sites commonly present in all four clones are indicated by ▼. LTRs are boxed, and Charon 4A arms are indicated by open bars. The 5′ end of the gene is positioned at the left of the figure. (B) Positions of the *Eco*RI (2.0 kb), *Pst*I (1.8 kb), *Hin*dIII (4.9 kb), and *Bam*HI (6.6 kb) fragments hybridizable with probe a, as shown in Fig. 1B. (C) Structure of the Syrian hamster IAP gene, IAP-H18 (5, 6). LTR, *gag*, and *pol* regions are boxed. Probe a, used for the cloning of the HERV-K genes, is indicated.

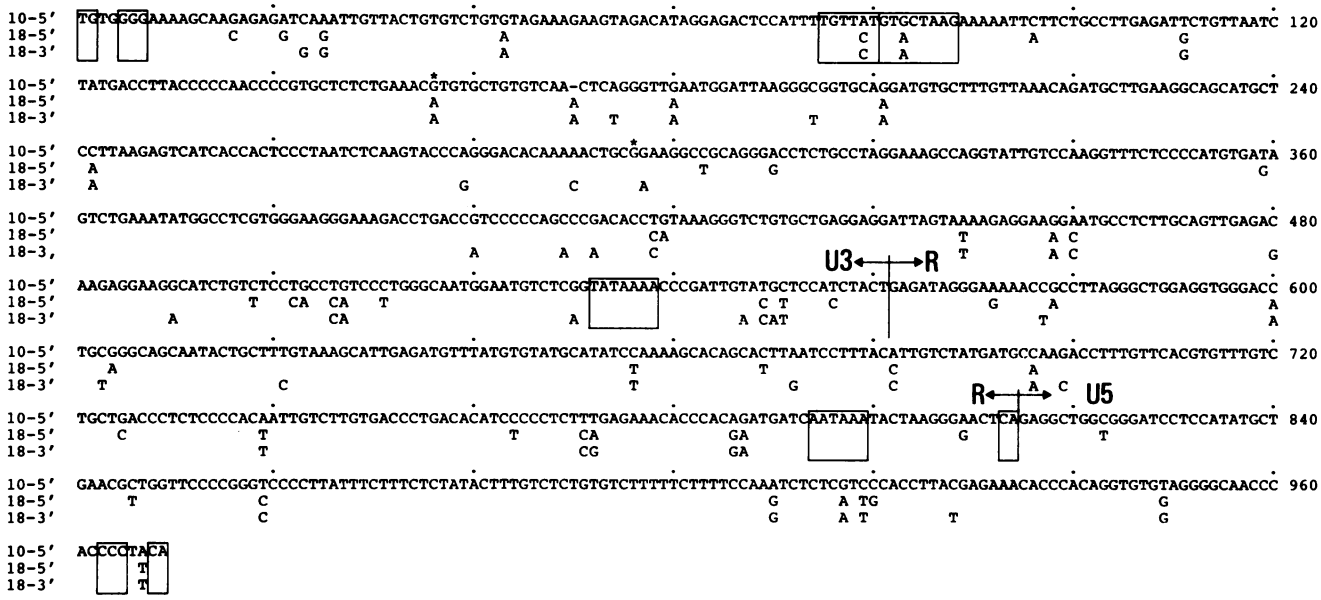

FIG. 3. LTR sequences of the HERV-K10 and HERV-K18 genes. The DNA sequences of the coding strand are given, and only the altered nucleotides are shown for the HERV-K18 LTRs. A dash at position 170 indicates the absence of a nucleotide. The HERV-K10 LTR situated at the 3′ end had two substituted adenines at positions 156 and 296, as indicated by asterisks. The inverted repeats, glucocorticoid responsive element (32), enhancer core sequence (15), TATAA box, polyadenylation signal, and polyadenylation site are boxed, and putative U3, R, and U5 regions are indicated.

obtained. Of 26 clones, 9 (35%) had hybridizable 2.0-kb *Eco*RI fragments (see examples in Fig. 1B) corresponding to the discrete band found in Fig. 1A. Similarly, 10 clones (38%) had hybridizable 1.8-kb *Pst*I fragments (Fig. 1B), and 9 clones (35%) had 4.9-kb *Hin*dIII fragments. But no clones having the 8.2-kb *Bam*HI fragment could be found from among these 26 clones. For a more detailed analysis, four clones (HERV-K8, HERV-K10, HERV-K18, and HERV-K22) having common restriction fragments hybridizable with probe a (see Fig. 2C) were selected. After mapping of these clones with restriction enzymes, the common restriction enzyme sites of each clone were aligned to give a preliminary organization map of the HERV-K genes (Fig. 2A).

LTR sequences and their flanking regions. From the organ-

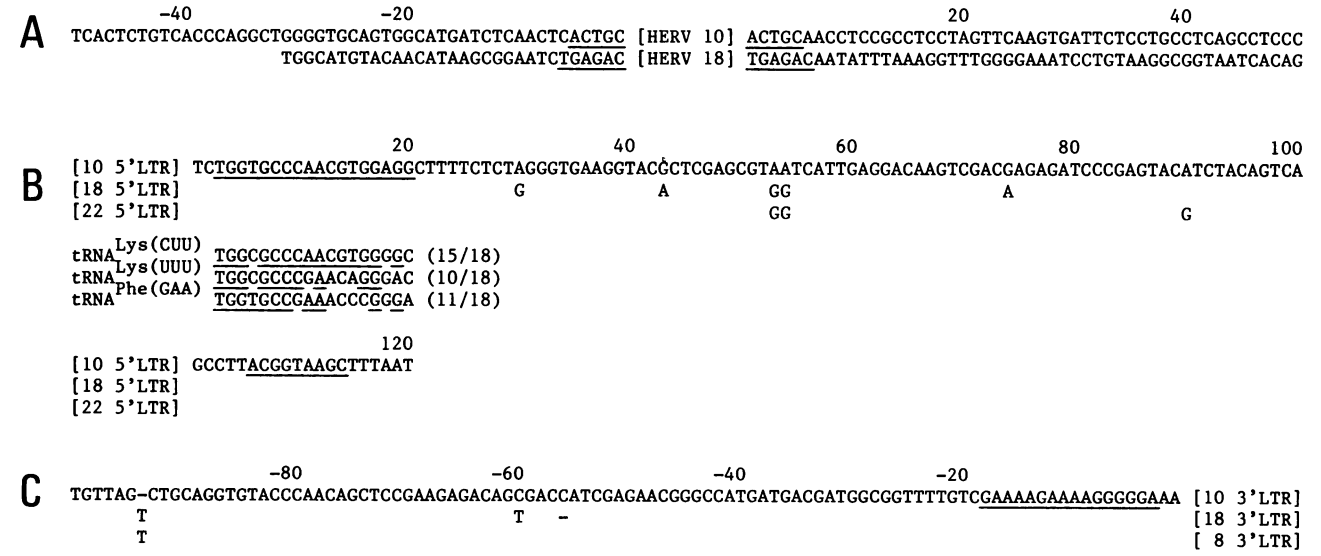

FIG. 4. Nucleotide sequence of the flanking regions of the HERV-K LTRs. (A) Sequence of cellular DNA adjacent to the HERV-K LTRs. The direct repeats are underlined. (B) Sequences downstream from the 5′ LTR of the HERV-K genes. Only the altered nucleotides are shown for HERV-K18 and HERV-K22. The sequence complementary to the last 18 nucleotides of three tRNAs (36) is also shown. The putative primer binding site at positions 3 to 20 and the splicing donor site (22) at positions 106 to 114 are underlined. (C) Sequence upstream from the 3′ LTR of the HERV-K genes. Only the altered nucleotides are shown for HERV-K18 and HERV-K8 LTRs, and the purine-rich region is underlined.

ization map and cross-hybridization experiment (data not shown), two *Bam*HI sites positioned at 0.8 and 8.9 kb on the restriction map of the HERV-K10 and HERV-K18 genes appeared to be situated in the putative LTR sequences of the HERV-K genes. To elucidate the fine structure of the putative LTR sequence, the nucleotide sequences of the LTR regions of the HERV-K10 and HERV-K18 genes were determined (Fig. 3 and 4).

The HERV-K10 LTRs were 968 base pairs (bp) in length and differed by only two bases, indicated by the asterisks in Fig. 3, at positions 156 and 296. The HERV-K18 LTRs were one base longer than HERV-K10 LTRs, containing an additional A residue at position 170, and differed from one another in 42 positions. The maximum number of base substitutions among HERV-K10 and HERV-K18 LTRs was 75 (7.7%).

At the ends of the LTR, an inverted repeat of 5′ TG---CA 3′, commonly present in the retrovirus LTRs (39, 40), was observed in addition to imperfect inverted repeats of 5 or 6 out of 7 bases found in HERV-K18 or HERV-K10 LTRs, respectively. Starting at position 532, HERV-K10 and HERV-K18 LTRs had the sequence (TATAAAA) corresponding to the consensus sequence of the TATAA box (TATA$^{A}_{T}$A$^{A}_{T}$) (4), a presumed promotor sequence for RNA polymerase II. A typical polyadenylation signal (4), AATAAA, starting from position 744, and a presumed polyadenylation site, CA, were found 13 bp downstream from the polyadenylation signal. On these HERV-K LTRs, there was no sequence corresponding to a CAT box (4), another presumed promotor, present 40 to 50 bp upstream from the TATAA box on the LTRs of the IAP and MuLV genes. Starting at position 81, a sequence (GT$^{G}_{A}$CTAAG) corresponding to an enhancer core (GTGG$^{AAA}_{TTT}$G) (15) was found, and immediately upstream from this sequence, a sequence (TGTT$^{A}_{C}$T) corresponding to a glucocorticoid responsive element (32) found in the MMTV LTR was present.

The retrovirus LTRs can be subdivided into 3 regions, U3, R, and U5 (10, 39, 40). The R region always starts with G positioned ca. 30-bp downstream from the TATAA box and ends with polyadenylation site CA. Assuming the G nucleotide located between the nucleotides 26 and 34 downstream from the initial T of the TATAA box to be the 5′ end of the R region (4), the R and U5 regions were calculated to be 253 and 155 bp, respectively, and the estimated U3 regions were calculated to be 560 bp in the HERV-K10 LTR and 561 bp in the HERV-K18 LTR. No open reading frame capable of encoding more than 60 amino acid residues was found on any reading frame of the coding strand of either the HERV-K10 or the HERV-K18 LTR. In addition, these LTRs had little sequence homology to IAP (27), MMTV (14), RSV (33), Moloney murine leukemia virus (35), or human T-cell leukemia virus (HTLV) type I LTR (34).

Cellular direct repeats of 6 bp were present immediately outside the HERV-K18 LTRs, whereas these LTRs in HERV-K10 were 5 bp (Fig. 4A). Fifteen out of eighteen bases of the primer binding site of the three HERV-K genes were complementary to the last 18 nucleotides of chicken and rat lysine (K) tRNA having a CUU anticodon (36), whereas the sequence of these HERV-K primer binding sites was less complementary to the corresponding region of either phenylalanine tRNA (a putative primer of the IAP gene) (28) or lysine tRNA having a UUU anticodon (a primer tRNA of the MMTV gene) (40). Hence, the lysine (CUU) tRNA was presumed to be the primer tRNA of the HERV-K genes. Starting at position 106 downstream from the 5′ LTRs, a sequence (ACG-GTAAGC) corresponding to the consensus-splicing donor ($^{C}_{A}$AG-GT$^{A}_{G}$AGT) (22) was found. Adjacent upstream from the 3′ LTRs, typical purine-rich regions (40), each 18 bp long, were observed.

The restriction maps between 5′ and 3′ LTRs were almost identical in four HERV-K clones, except that HERV-K22 and HERV-K8 were 0.3-kb longer than HERV-K10 and HERV-K18 between a *Pst*I site at 6.55 kb and a *Bam*HI site at 7.4 kb. Since the length of the LTR was 1.0 kb and the distance between the two *Bam*HI sites in the LTR was 1.0 kb and the distance between the two *Bam*HI sites in the LTR was 8.1 kb in HERV-10 and HERV-18 or 8.4 kb in HERV-K22 and HERV-K8, the size of a typical HERV-K gene, according to type, was either 9.1 or 9.4 kb.

Organization of HERV-K genes in the human genome. To elucidate the organization of HERV-K genes in the human genome, human peripheral leukocyte DNA was digested with several restriction enzymes and analyzed by the Southern hybridization method by using probe b containing the sequence cross-hybridizable with probe a or by using probe c having almost the entire region of the HERV-K gene (see Fig. 5B). With *Eco*RI digestion, probe b mainly hybridized with the 2.0-kb fragment, as expected from the restriction map. Minor 2.9- or 3.7-kb fragments appeared to be generated from the two *Eco*RI sites, the first and third or the first and fourth (numbered from the 5′ end), respectively, as observed within the HERV-K gene. In addition to the 2.0-kb fragment, probe c hybridized with the 1.8-kb fragment which appeared to correspond to the fragment generated from the second and fourth *Eco*RI sites. The 0.9-kb fragment detected by probe c was expected to be generated from the second and third or the third and fourth *Eco*RI sites.

With *Pst*I or *Hin*dIII digestion, the size of the predominant fragments hybridized with either probe b or probe c was that expected from the restriction map. The major 8.2-kb fragment detected by probes b and c upon *Bam*HI digestion seemed to lack the third *Bam*HI site commonly observed in all four clones in Fig. 2A and corresponded to the fragment generated from the first and fourth *Bam*HI sites. The 6.6-kb fragment plausibly generated from the first and third *Bam*HI sites was found as a minor band in Fig. 5A (*Bam*HI). The minor 2.9-kb fragment shown in Fig. 5A (*Bam*HI) was assumed to be the fragment generated from the second and third *Bam*HI sites, and the 1.5-kb fragment in Fig. 5A (*Bam*HI-2) appeared to be from the third and fourth *Bam*HI sites. In other human peripheral leukocyte DNAs, obtained from five individuals, the position and intensity of the main bands detected by the Southern hybridization were virtually the same as those observed in Fig. 5A. Since the 8.2-kb band in Fig. 5A (*Bam*HI) corresponds to 8.1 or 8.4 kb in the cloned HERV-K genes but lacks the internal *Bam*HI site, a considerable number of HERV-K genes in the human genome may be integrated as 9.1- or 9.4-kb HERV-K genes, as was observed in the clones.

From the dot hybridization experiment, the number of genes hybridizable with either probe b or c was calculated to be 50 copies per haploid human genome (Fig. 6A, B, and D). This was ca. 20 times less than that of IAP genes present in the rodent genome (27). By Southern hybridization, 20 to 40 copies per haploid human genome of the HERV-K genes were estimated to each have the 2.0-kb *Eco*RI fragment corresponding to probe b (Fig. 6C).

Sequence homology of the HERV-K gene to other oncovirus genes. On the basis of the homology of the amino acid sequences deduced from the nucleotide sequences of the cloned retrovirus genes, the interrelationship of six representative oncovirus genes (types A, B, avian C, mammalian

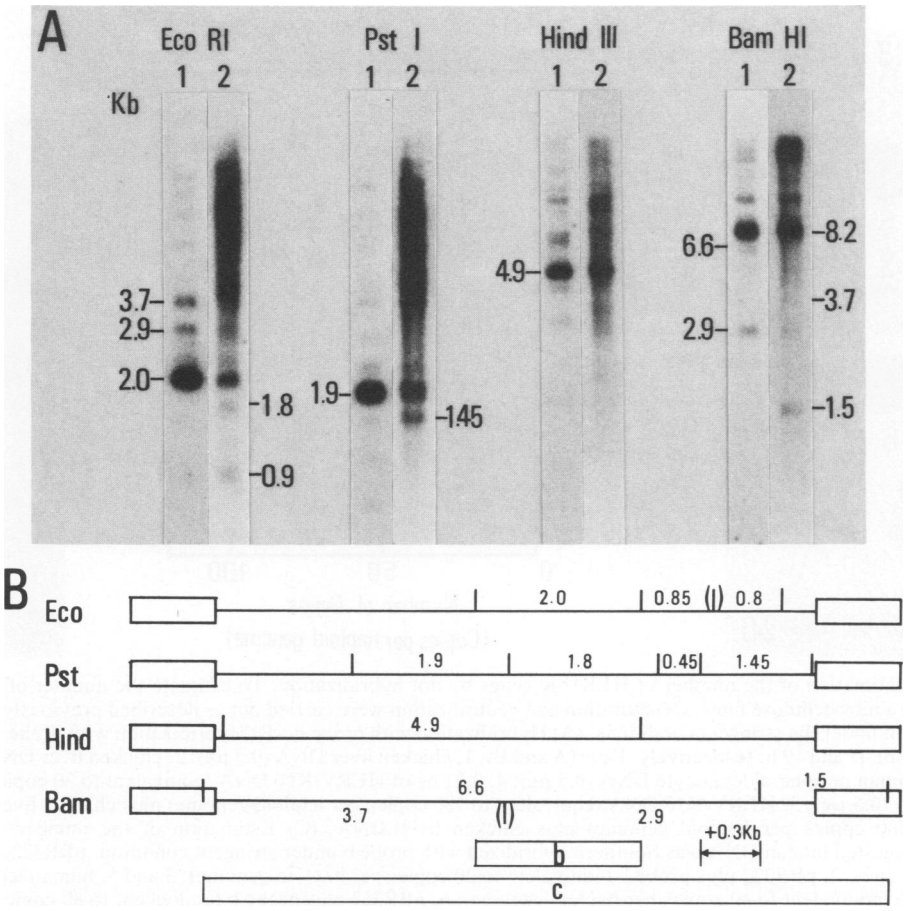

FIG. 5. Southern hybridization analysis of HERV-K genes in a human genome. (A) Human leukocyte DNA (2.0 μg) was digested with restriction enzymes and analyzed by the Southern hybridization method with probes prepared from the HERV-K10 gene. The hybridization and washing were carried out under the stringent conditions. Lanes marked 1, hybridization with probe b; lanes marked 2, hybridization with probe c. (B) 9.1-kb-type HERV-K genes, indicated with boxed LTRs at both ends. Restriction enzyme sites commonly present in the cloned HERV-K genes are indicated by vertical lines and numbered in order from the 5′ end. A less-common third *Eco*RI site and a presumed second *Bam*HI site appear in parentheses. A 0.3-kb region longer in 9.4-kb-type HERV-K genes is indicated beside the boxed probe b. Probe b (2.0-kb *Eco*RI fragment) and probe c (8.1-kb *Bam*HI fragment), indicated at the bottom of the figure, were prepared from the subcloned HERV-K10 genes.

C, D, and HTLV) has recently been elucidated (11, 29, 31). Among the oncovirus subfamily, the IAP gene (type A) was closely related to both MMTV (type B) and squirrel monkey retrovirus (type D) genes and homologous to the RSV (avian type C) gene (29). Thus, the sequence homology of the HERV-K gene to the MMTV or RSV gene was investigated by the Southern hybridization method under the less stringent condition. In Fig. 6E, the cloned MMTV gene hybridized with the HERV-K and Syrian hamster IAP genes to essentially the same extent. Probe b slightly hybridized with cloned RSV DNA, but no hybridization was detected between probe b and the Friend murine leukemia virus (mammalian type C) gene (data not shown).

DISCUSSION

Recently, HERV genes homologous to mammalian type C or type B genes or to oncovirus genes of an unidentified type were isolated and characterized. Since the length of LTRs, the LTR sequences, the presumed primer tRNA, and the nucleotide number of the directly repeated cellular sequences of the HERV-K genes in the present study all differed from those of the murine leukemia virus (mammalian type C) gene-related HERV genes isolated by either Martin et al. (20) and Repaske et al. (30) or Bonner et al. (3) and O'Connell and Cohen (25), the HERV-K genes can be concluded to differ from their HERV genes. HERV genes isolated by Noda et al. (23) were homologous to the baboon endogenous virus (mammalian type C) LTR, and ca. 300 copies were present per haploid human genome. Significant differences in gene number and absence of any detectable sequence homology between baboon endogenous virus LTR (38) and HERV-K LTR indicate that their HERV genes are distinct from those in the present study. A HERV gene of unidentified type, present downstream from the human beta-globin gene cluster and cloned by Mager and Henthorn (19), could be distinguished from the HERV-K gene, since its LTR length, LTR sequences, putative primer tRNA, and copy number differed completely from those found in the HERV-K gene.

HERV genes homologous to the MMTV (type B) gene were isolated and characterized by Callahan et al. (7). In a recent article (6), Callahan et al. showed that one of their clones, HLM-2, had a sequence hybridizable with the M432

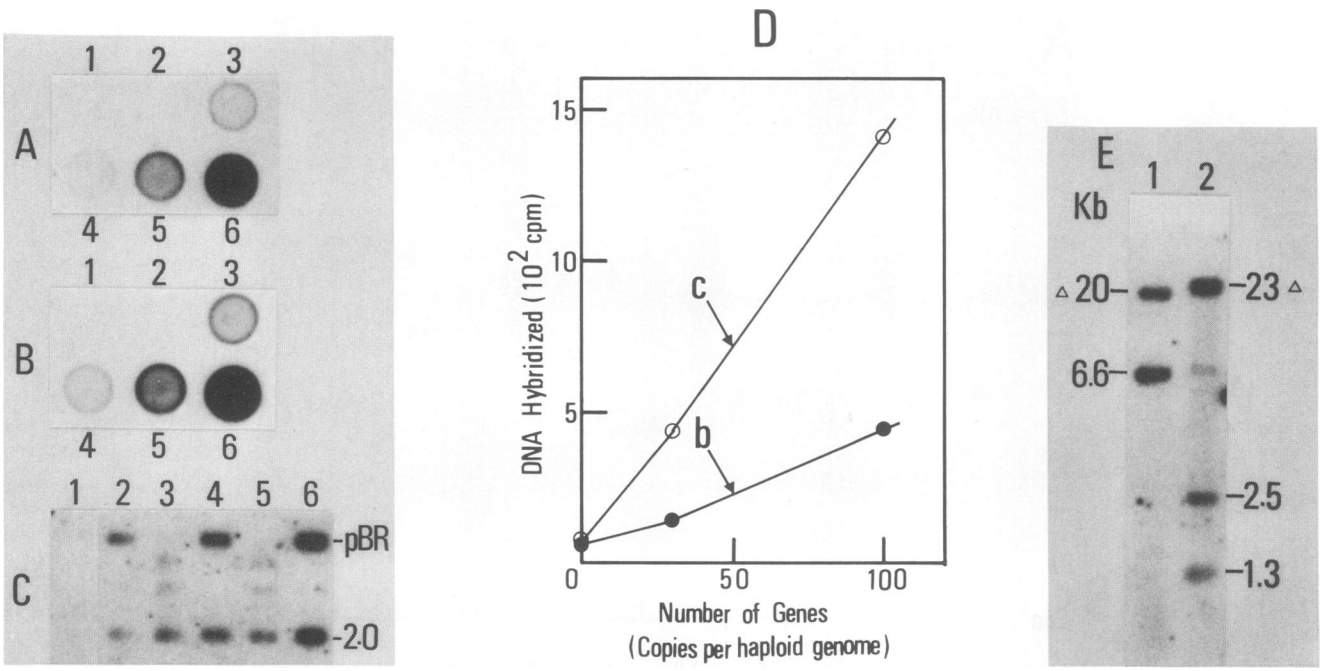

FIG. 6. (A and B) Estimation of the number of HERV-K genes by dot hybridization. To estimate the number of genes, *Eco*RI-digested DNA was spotted onto a nitrocellulose filter. Denaturation and neutralization were carried out as described previously (1). Hybridization and washing were carried out under the stringent conditions. (A) Hybridization with probe b. (B) Hybridization with probe c. Nitrocellulose filters A and B were exposed for 47 and 19 h, respectively. Dots (A and B): 1, chicken liver DNA (0.5 μg); 2, chicken liver DNA (0.5 μg) plus Charon 4A DNA (2.3 ng); 3, human peripheral leukocyte DNA (0.5 μg); 4, 0.21 ng of HERV-K10 DNA (equivalent to 30 copies per haploid genome) plus chicken liver DNA (0.5 μg); 5, HERV-K10 DNA (equivalent to 100 copies per haploid genome) plus chicken liver DNA; 6, HERV-K10 DNA (equivalent to 300 copies per haploid genome) plus chicken liver DNA. (C) Estimation of the number of genes by Southern hybridization. *Eco*RI-digested human DNA was Southern hybridized with probe b under stringent condition. pBR322 containing probe b was used as the standard. Lanes: 2, pBR322 plus probe b (equivalent to 20 copies per haploid genome); 3 and 5, human leukocyte DNA (1.0 μg); 4, pBR322 plus probe b (equivalent to 40 copies per haploid genome); 6, pBR322 plus probe b (equivalent to 80 copies per haploid genome). Lanes 1, 2, 4, and 6 also had 1.0 μg of chicken liver DNA (*Eco*RI digested). (D) Graphical presentation of the dot hybridization experiment. Radioactivity on the spots in panels A and B was counted, and the average of duplicate samples was plotted. Symbols: ●, Hybridized with probe b; ○, hybridized with probe c. Positions of the radioactive label hybridized with human leukocyte DNA are indicated by arrows. (E) Southern hybridization of HERV-K gene and IAP gene with MMTV gene. The MMTV gene, subcloned as four *Pst*I fragments into pBR322, was used as a probe under the less stringent condition. Lane 1, *Bam*HI-digested HERV-K10 DNA (0.25 μg); lane 2, *Hin*dIII-digested IAP-H18 DNA (0.25 μg). A restriction map of HERV-K10 and IAP-H18 genes is shown in Fig. 2. Positions of the 6.6-kb fragment on the HERV-K10 DNA and the 1.3- and 2.5-kb fragments on the HERV-H 18 DNA are shown in Fig. 2A and 2C, respectively. △, Indicates the fragments cross-hybridized with vector pBR322; pBR322 itself did not hybridize with either the 6.6-kb fragment in HERV-K 10 or the 1.3- or 2.5-kb fragment in HERV-H 18.

gene assumed to be a recombinant between an IAP gene in the *Mus* genus and an unknown retrovirus gene (8). Since the HERV-K gene had the sequence homologous to the MMTV gene as shown in Fig. 6E, the HLM-2 gene is likely to be a member of the HERV-K gene family. In fact, the estimated gene number and hybridization pattern shown in Fig. 5A agree well with their results (6, 7). For conclusive confirmation of this identity, we are now carrying out nucleotide sequencing of the putative endonuclease domain of the *pol* region whose partial sequence has already been reported (6).

Since both type A and B oncovirus genes are similarly homologous to the type D gene (11, 29, 31) in the most conserved region, *pol*, the HERV-K gene, among the oncovirus genes, must be similarly homologous to these three oncovirus genes in the *pol* region. In addition, the HERV-K gene is slightly homologous to the RSV (avian type C) gene in the *pol* region, so that the HERV-K gene family may be considered to have diverged from the common progenitor of the oncovirus genes almost at the same time as type A, B, and D genes after divergence of the avian type C gene.

Among the reported retrovirus LTRs, the HERV-K LTRs are the second longest next to the MMTV LTR (1,332 bp) (10). In contrast to the presence of an open reading frame capable of encoding ca. 320 amino acid residues on the MMTV LTR (14), no open reading frame encoding more than 60 amino acid residues was found on any reading frame of the coding strand of the HERV-K LTRs.

In the retrovirus LTRs, the R region is usually less than 100 bp in length (10), but in the HTLV- and bovine leukemia virus-type genes this region is more than 200 bp in length (31, 34). On the HTLV and bovine leukemia virus LTRs, the polyadenylation signal is positioned 260- to 270-bp upstream from the polyadenylation site, and the splicing donor site is present in the middle of the R region (31, 34). Thus, the secondary structure of the R region may be essential for controlling transcription and polyadenylation (31). In contrast, the significance of the comparably long R region (253 bp) in HERV-K LTR remains to be clarified, since the location of the putative polyadenylation signal and splicing donor site differ considerably from those in HTLV and bovine leukemia virus genes, and no thermodynamically

stable secondary structure could be found in the HERV-K R region.

In this work, a family of the HERV-K genes was cloned and characterized. As one step for elucidating the significance of the HERV-K gene family in the human genome, an investigation is currently being conducted on the transcription and translocation of the HERV-K gene in normal and malignant human cells.

ACKNOWLEDGMENTS

I am grateful to M. Kawakami for his encouragement throughout the course of this work and to H. Diggelmann, J. M. Bishop, Y. Ikawa, and T. Maniatis for kindly providing the clones and gene library.

This work was supported by a grant-in-aid for cancer research from the Ministry of Education, Science, and Culture of Japan.

LITERATURE CITED

1. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
2. **Blin, N., and D. W. Stafford.** 1976. A general method for isolation of high molecular weight DNA from eukaryotes. Nucleic Acids Res. **3**:2303–2308.
3. **Bonner, T. I., C. O'Connell, and M. Cohen.** 1982. Cloned endogenous retroviral sequences from human DNA. Proc. Natl. Acad. Sci. USA **79**:4709–4713.
4. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
5. **Buetti, E., and H. Diggelmann.** 1981. Cloned mouse mammary tumor virus DNA is biologically active in transfected mouse cells and its expression is stimulated by glucocorticoid hormones. Cell **23**:335–345.
6. **Callahan, R., I.-M. Chiu, J. F. H. Wong, S. R. Tronick, B. A. Roe, S. A. Aaronson, and J. Schlom.** 1985. A new class of endogenous human retroviral genomes. Science **228**:1208–1211.
7. **Callahan, R., W. Drohan, S. Tronick, and J. Schlom.** 1982. Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. Proc. Natl. Acad. Sci. USA **79**:5503–5507.
8. **Callahan, R., E. L. Kuff, K. K. Lueders, and E. Birkenmeier.** 1981. Genetic relationship between the *Mus cervicolor* M432 retrovirus and the *Mus musculus* intracisternal type A particle. J. Virol. **40**:901–911.
9. **Canaani, E., O. Dreazen, A. Klar, G. Rechavi, D. Ram, J. B. Cohen, and D. Givol.** 1983. Activation of the c-*mos* oncogene in a mouse plasmacytoma by insertion of an endogenous intracisternal A-particle genome. Proc. Natl. Acad. Sci. USA **80**:7118–7122.
10. **Chen, H. R., and W. C. Barker.** 1984. Nucleotide sequences of the retroviral long terminal repeats and their adjacent regions. Nucleic Acids Res. **12**:1767–1778.
11. **Chiu, I.-M., R. Callahan, S. R. Tronick, J. Schlom, and S. R. Aaronson.** 1984. Major *pol* gene progenitors in the evolution of oncoviruses. Science **223**:364–370.
12. **Coffin, J.** 1982. Endogenous viruses, p. 1109–1203. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Molecular biology of tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. **DeLorbe, W. J., P. R. Luciw, H. M. Goodman, H. E. Varmus, and J. M. Bishop.** 1980. Molecular cloning and characterization of avian sarcoma virus circular DNA molecules. J. Virol. **36**:50–61.
14. **Kennedy, N., G. Knedlitschek, B. Groner, N. E. Hynes, P. Herrlich, R. Michalides, and A. J. J. van Ooyen.** 1982. Long terminal repeats of endogenous mouse mammary tumor virus contain a long open reading frame which extends into adjacent sequences. Nature (London) **295**:622–624.
15. **Khoury, G., and P. Gruss.** 1983. Enhancer elements. Cell **33**:313–314.
16. **Kuff, E. L., A. Feenstra, K. Lueders, L. Smith, N. Hawley, N. Hozumi, and M. Shulman.** 1983. Intracisternal A-particle genes as movable elements in the mouse genome. Proc. Natl. Acad. Sci. USA **80**:1992–1996.
17. **Lawn, R. M., E. F. Fritsch, R. C. Parker, G. Blake, and T. Maniatis.** 1978. The isolation and characterization of linked δ- and β-globin genes from a cloned library of human DNA. Cell **15**:1157–1174.
18. **Lueders, K. K., and E. L. Kuff.** 1981. Sequence homologous to retrovirus-like genes of the mouse are present in multiple copies in the Syrian hamster genome. Nucleic Acids Res. **9**:5917–5930.
19. **Mager, D. L., and P. S. Henthorn.** 1984. Identification of a retrovirus-like repetitive element in human DNA. Proc. Natl. Acad. Sci. USA **81**:7510–7514.
20. **Martin, M. A., T. Bryan, S. Rasheed, and A. S. Khan.** 1981. Identification and cloning of endogenous retroviral sequences present in human DNA. Proc. Natl. Acad. Sci. USA **78**:4892–4896.
21. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
22. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
23. **Noda, M., M. Kurihara, and T. Takano.** 1982. Retrovirus-related sequences in human DNA: detection and cloning of sequences which hybridize with the long terminal repeat of baboon endogenous virus. Nucleic Acids Res. **10**:2865–2878.
24. **Obata, M., H. Amanuma, H. Harada, N. Sagata, and Y. Ikawa.** 1984. *env*-related leukemogenic genes (gp55 genes) of two closely related polycythemic strains of Friend spleen focus-forming virus possess different recombination points with an endogenous mink cell focus-forming virus *env* gene. Virology **136**:435–438.
25. **O'Connell, C. D., and M. Cohen.** 1984. The long terminal repeat sequences of a novel human endogenous retrovirus. Science **226**:1204–1206.
26. **Ono, M., M. D. Cole, A. T. White, and R. C. C. Huang.** 1980. Sequence organization of cloned intracisternal A particle genes. Cell **21**:465–473.
27. **Ono, M., H. Kitasato, H. Ohishi, and Y. Motobayashi-Nakajima.** 1984. Molecular cloning and long terminal repeat sequences of intracisternal A-particle genes in *Mus caroli*. J. Virol. **50**:352–358.
28. **Ono, M., and H. Ohishi.** 1983. Long terminal repeat sequences of intracisternal A particle genes in the Syrian hamster genome: identification of the $tRNA^{Phe}$ as a putative primer tRNA. Nucleic Acids Res. **11**:7169–7179.
29. **Ono, M., H. Toh, T. Miyata, and T. Awaya.** 1985. Nucleotide sequence of the Syrian hamster intracisternal A-particle gene: close evolutionary relationship of type A particle gene to types B and D oncovirus genes. J. Virol. **55**:387–394.
30. **Repaske, R., P. E. Steele, R. R. O'Neill, A. B. Rabson, and M. A. Martin.** 1985. Nucleotide sequence of a full-length human endogenous retroviral segment. J. Virol. **54**:764–772.
31. **Sagata, N., T. Yasunaga, J. Tsuzuku-Kawamura, K. Ohishi, Y. Ogawa, and Y. Ikawa.** 1985. Complete nucleotide sequence of the genome of bovine leukemia virus: its evolutionary relationship to other retroviruses. Proc. Natl. Acad. Sci. USA **82**:677–681.
32. **Scheidereit, C., and M. Beato.** 1984. Contacts between hormone receptor and DNA double helix within a glucocorticoid regulatory element of mouse mammary tumor virus. Proc. Natl. Acad. Sci. USA **81**:3029–3033.
33. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
34. **Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida.** 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA **80**:3618–3622.
35. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
36. **Sprinzl, M., J. Moll, F. Messner, and T. Hartmann.** 1985.

Compilation of tRNA sequences. Nucleic Acis Res. **13**:1–49.
37. **Suzuki, A., H. Kitasato, M. Kawakami, and M. Ono.** 1982. Molecular cloning of retrovirus-like genes present in multiple copies in the Syrian hamster genome. Nucleic Acids Res. **10**:5733–5746.
38. **Tamura, T.-A., M. Noda, and T. Takano.** 1981. Structure of the baboon endogenous virus genome: nucleotide sequences of the long terminal repeats. Nucleic Acids Res. **9**:6615–6626.
39. **Temin, H. M.** 1981. Structure, variation and synthesis of retrovirus long terminal repeat. Cell **27**:1–3.
40. **Varmus, H. E.** 1982. Form and function of retroviral proviruses. Science **216**:812–820.