GENOMICS 4, 68–75 (1989)

Human SSAV-Related Endogenous Retroviral Element: LTR-like Sequence and Chromosomal Localization to 18q21

RUTH BRACK-WERNER,* DAVID E. BARTON,† THOMAS WERNER,‡ BRIGITTE E. FOELLMER,† CHRISTINE LEIB-MÖSCH,§ UTA FRANCKE,† VOLKER ERFLE,* AND RÜDIGER HEHLMANN§

*Abt. für Molekulare Zellpathologie and ‡Institut für Säugetiergenetik, Gesellschaft für Strahlen- und Umweltforschung (GSF), D-8042 Neuherberg, Federal Republic of Germany; †Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06510; and §Medizinische Poliklinik der Universität München, D-8000 Munich, Federal Republic of Germany

Received July 12, 1988; revised September 14, 1988

A new family of human endogenous retroviral sequences was recently discovered by way of its relationship to the simian sarcoma-associated virus (SSAV). One molecular clone, termed S71, contains sequences related to the genes coding for the group-specific antigens (gag) and polymerase (pol) proteins of SSAV. At the 3′ end of this human retroviral element we have now found a 535-bp region which shows features characteristic of a retroviral long terminal repeat, including potential signal sequences essential for transcriptional control. By means of Southern blotting and *in situ* hybridization, the sequence was mapped to chromosome 18 band q21. © 1989 Academic Press, Inc.

INTRODUCTION

A number of human endogenous retroviral elements have been isolated on the grounds of their homology either to mammalian C-type retroviruses (Bonner *et al.*, 1982; Leib-Mösch *et al.*, 1986; Martin *et al.*, 1981; O'Connell *et al.*, 1984; Repaske *et al.*, 1985) or to A- and B-type viruses (Callahan *et al.*, 1985; Deen and Sweet, 1986; May and Westley, 1986; Ono, 1986). Others have been identified by their retrovirus-like structure (Kröger and Horak, 1987; Mager and Freeman, 1987; Mager and Henthorn, 1984; Paulson *et al.*, 1985). Some retroviral elements which occur in only one copy per genome equivalent have been assigned to distinct chromosomes (O'Brien *et al.*, 1983; O'Connell *et al.*, 1984; Renan and Reeves, 1987); others are members of families with 50 to 10,000 copy numbers and are widely dispersed through the human genome (Horn *et al.*, 1986; Kröger and Horak, 1987; Mager and Freeman, 1987; Ono, 1986; Paulson *et al.*, 1985; Steele *et al.*, 1984, 1986). Despite the abundant occurrence of human endogenous retroviral elements, their functional significance is not yet known. Nevertheless, they are obviously not only silent components of the human genome since some have been shown to be transcriptionally active (Kato *et al.*, 1987; Rabson *et al.*, 1985). Expression of human endogenous retroviral sequences has also been associated with neoplastic disease (Gattoni-Celli *et al.*, 1986); however, conclusive evidence for a direct role of human retroviral elements in pathogenicity is still forthcoming. We have found human DNA to contain a number of sequences related to the simian sarcoma-associated virus/gibbon ape leukemia virus (SSAV/GALV) primate retrovirus group. Under relaxed hybridization conditions we isolated one of these sequences from a human genomic library (Leib-Mösch *et al.*, 1986). This clone, S71, contains an incomplete proviral element with sequences related to the gag and pol genes of SSAV. S71 occurs in only one copy per genome equivalent, but under less stringent hybridization conditions a number of additional S71-related sequences are observed in the human genome. Although SSAV/GALV-related proteins have been detected in human sera and tissues by immunological methods and have been associated with various types of leukemia (Derks *et al.*, 1982; Hehlmann *et al.*, 1983, 1984; Schetters *et al.*, 1985), experimental evidence for a potential role of SSAV/GALV-related sequences in human neoplasias has been lacking. We now present data that localize the S71 sequence to the same band of human chromosome 18 as that of the bcl-2 breakpoint cluster region of human follicular B-cell lymphomas (Bakhshi *et al.*, 1985, 1987; Pegoraro *et al.*, 1984; Tsujimoto *et al.*, 1984). Furthermore, sequencing of the 3′ end of S71 showed the region directly adjacent to the S71 pol sequence to contain a long terminal repeat (LTR)-like structure with several possible regulatory elements.

MATERIALS AND METHODS

Dideoxynucleotide DNA Sequencing

S71 subclones in plasmid vector pUC9 were sequenced as described (Leib-Mösch *et al.*, 1986) using

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. J03061.

0888-7543/89 $3.00

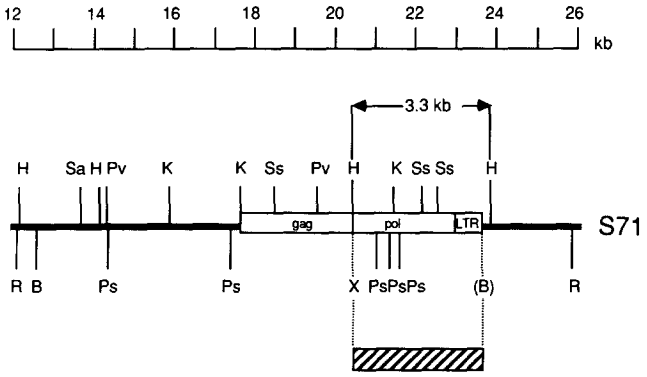


FIG. 1. Structural organization of the SSAV-related human endogenous retroviral element in the S71 locus. Open rectangles in the restriction map mark the extent of the regions containing gag, pol, and LTR-like sequences in S71. A hatched rectangle designates the S71 pol-LTR hybridization probe which detects a 3.3-kb *Hin*dIII fragment (horizontal arrow) in human genomic blots. Abbreviations: B, *Bam*HI; H, *Hin*dIII; K, *Kpn*I; Ps, *Pst*I; Pv, *Pvu*II; R, *Eco*RI; Ss, *Sac*I; X, *Xho*I. The *Bam*HI site in parentheses was generated by vector sequences in the original λS71 clone.

either Klenow fragment (Pharmacia, Sweden) or the modified T7 DNA polymerase (Tabor and Richardson, 1987) Sequenase (United States Biochemical Corp.). The sequences of both strands were determined.

Oligonucleotide Preparation

Specific sequencing primers were synthesized on an Applied Biosystems Model 381A DNA synthesizer, purified by polyacrylamide gel electrophoresis, and desalted by gel exclusion chromatography as specified in the Applied Biosystems User Bulletin No. 13 (1984).

Sequence Analysis

Sequence homology analysis for the TG motif was carried out with the Dayhoff program package (Protein Identification Resource, PIR) on a VAX computer and with the MacGene package on a Macintosh Plus PC. The LTR sequences shown in Fig. 3 were aligned by eye. All sequences for comparison were taken from the EMBL database release 10.

Hybrid Cell Lines

The construction and characterization of all somatic cell hybrid lines used have been described in detail elsewhere. Series XII, XVII, and XVIII hybrids were derived from fusions between Chinese hamster 380-6 (a HPRT-deficient derivative of lung fibroblast line V79) and MN-3 leukocytes (XII), CM-6 fibroblasts (XVII), or KC-8 fibroblasts (XVIII) (Francke, 1984; Francke *et al.*, 1976; Oliver *et al.*, 1978). Series XXI hybrids were derived from fusions between Chinese hamster Don/a3 TK$^-$cells and KG-7 fibroblasts (Francke and Francke, 1981).

Southern Blot Hybridization

Genomic DNA (10 μg) was digested with restriction enzymes, and the resulting fragments were separated by electrophoresis in 0.8% agarose gels and transferred to nitrocellulose filters by the methods of Southern (1979). Filters were pretreated and hybridized with gel-purified, ^{32}P-labeled probe as described previously (Barton *et al.*, 1986).

In Situ Hybridization

Preparation of human metaphase chromosome spreads, labeling of probe with ^{3}H by nick-translation, and *in situ* hybridization were carried out as described (Harper and Saunders, 1981), except that after hybridization, slides were washed to a final stringency of 50% formamide in 0.2 × SSC at 39°C to minimize cross-hybridization with related sequences.

RESULTS

S71 LTR-like Sequence

Figure 1 is a schematic of the structural organization of the SSAV-related incomplete proviral element in the S71 locus. The S71 retroviral element contains two regions shown by hybridization analysis to be related to the gag and pol genes of SSAV and other C-type retroviruses. Comparison of the deduced amino acid sequences of several sections of the S71 pol gene yielded 40–55% identity with the pol gene of the C-type murine leukemia virus AKV (Leib-Mösch *et al.*, 1986). The nucleotide sequence of a 535-bp stretch of the 3′ end of the S71 retroposon is depicted in Fig. 2. This sequence exhibits a number of features typical of retroviral LTRs (Chen and Barker, 1984). The boundaries of retroviral LTRs are formed by inverted repeats 2 to 15 nucleotides in length. In most integrated proviruses the inverted repeats begin with TG and end with the

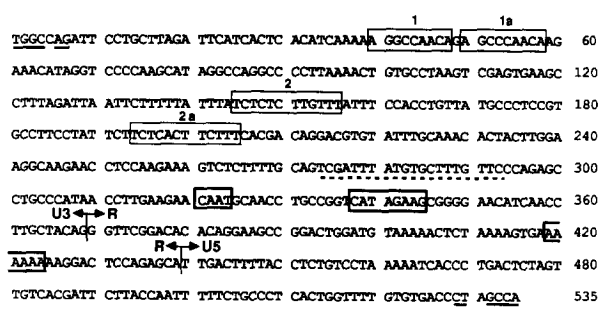


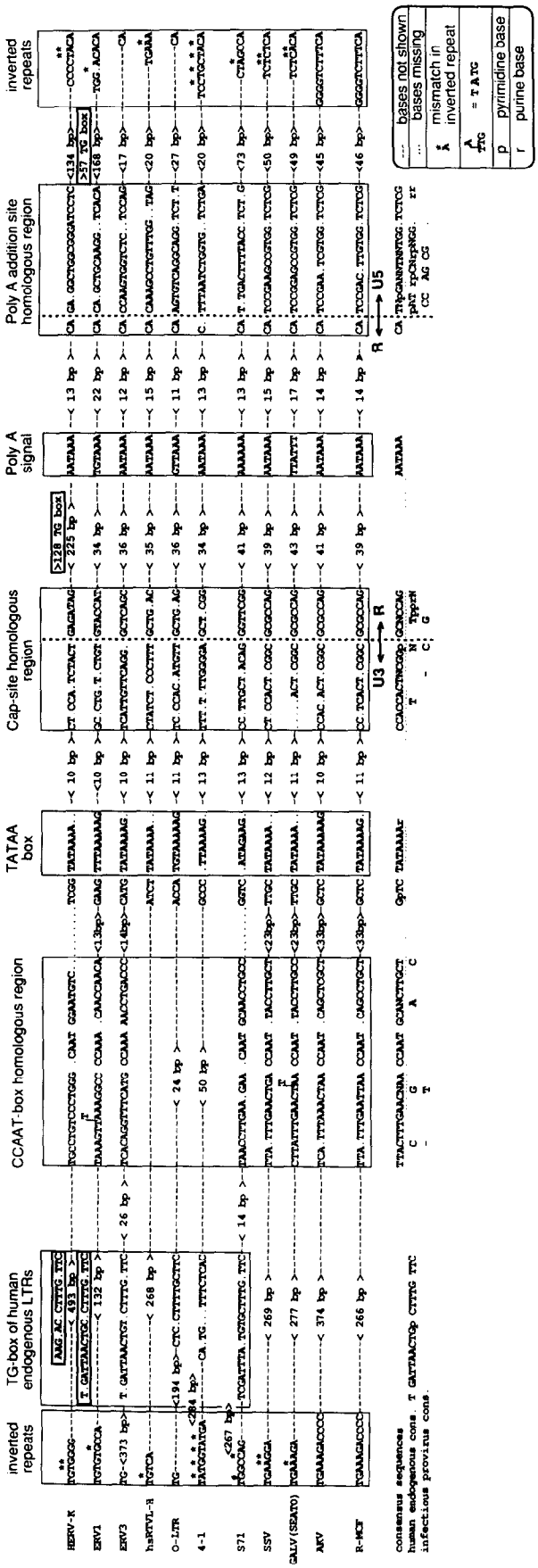
FIG. 2. Nucleotide sequence of the 3′ LTR-like region of S71. Inverted repeats are underlined, and direct repeats are boxed and designated 1, 1a, and 2, 2a. Transcription regulatory sequences are boxed with boldface lines. The dotted line shows the TG box related sequence. Vertical lines mark the borders of the U3/R and R/U5 regions.

inverted complement CA. In some proviral LTRs (i.e., AKV, R-MCF, and SSV) the inverted repeat sequence can be extended to include two or three nucleotides upstream of the TG and downstream of the CA dinucleotide (AKV/R-MCF: 5'AATGAAGACCCC/GGGGTCTTTCATT; Van Beveren *et al.*, 1982; Vogt *et al.*, 1985; SSV 5'LTR: AATTGAAGGA/TCTCTCATTT; Devare *et al.*, 1983). The S71 LTR-like sequence contains a 7-bp inverted repeat sequence beginning with TG (TGAGCCAG) and ending with CA (CTAGCCA) at positions 1–7 and 529–535, respectively. Inclusion of the three bases AAG 5' of TG and CTT 3' of CA extends the inverted repeat in the S71 sequence to 10 bp. The S71 retroviral element lacks an envelope gene as well as the polypurine-rich region normally preceding retroviral 3' LTRs.

Retroviral LTRs consist of the three subregions U3, R, and U5 (5'-U3-R-U5-3'). The U3 region contains control elements for the initiation of transcription (promoter and enhancer). The R region begins with the cap site (G) and ends with the poly(A) addition site (CA). The remaining LTR sequence constitutes the U5 region. In order to define the U3, R, and U5 regions of the S71 LTR-like sequence and to locate putative transcription control elements, we compared the S71 sequence with the published nucleotide sequences of 10 retroviral LTRs. Four of the latter were derived from infectious murine or primate proviruses and the remaining six from different human endogenous retroviral elements. Initially, we used a computer program based on the Needleman–Wunsch algorithm (Needleman and Wunsch, 1970) for alignment of the 10 proviral LTR sequences. This method yielded extensive overall sequence homology (ca. 75% matches) for the LTR sequences of the four primate and murine infectious proviruses, with matching of the respective transcriptional signals. However, no such alignment was obtained for the six human endogenous retroviral LTR sequences with the same program, indicating that these show no extended overall homology to other retroviral LTRs. Like infectious proviruses, human endogenous LTRs also contain signal sequences for transcriptional control which, in at least one case, have been shown to be functionally active (ERV3; Kato *et al.*, 1987). Therefore we concluded that the transcriptional signal sequences within the LTRs should represent the most highly conserved sequences and consequently be appropriate for positioning the human endogenous LTR sequences for alignment. Figure 3 shows that homology between retroviral LTRs includes not only the control elements themselves but also the sequences in the vicinity of the CAAT box, the cap site, and the poly(A) addition site. The S71 U3 region contains a CAAT box at position 321 (Fig. 2). The sequence surrounding the CAAT box (CAAT box homologous region) matches at 17 of 29 positions (59%) to the simian sarcoma virus (SSV) LTR (Fig. 3). It is noteworthy that three of six human endogenous retroviral LTRs used for comparison with the S71 sequence lack a CAAT box. From the alignment in Fig. 3 it is evident that a TATAA box is located approximately 40 nucleotides downstream of the CAAT box in the infectious murine and primate LTRs. No typical TATAA box can be found at a corresponding position in the S71 sequence. However, there is a CATAGAAG sequence about 15 nucleotides downstream of the S71 CAAT box at position 338 (Fig. 3). This sequence is identical in 7 of 9 positions to the TATAA consensus sequence G/C T/A AT T/A T/A AAG (Bonner *et al.*, 1982). Furthermore, its spacing relative to the CAAT box is analogous to that of the TATAA box in the HERV-K LTR (Fig. 3).

In addition to the typical LTR elements described above, the U3 region of S71 contains a TG box at position 274 which matches similar motifs found in some human and primate LTRs. With the exception of hsRTVL-H (Mager and Henthorn, 1984), all human endogenous LTR sequences analyzed seem to contain at least part of this box (Fig. 3). Related sequences can also be found in the LTRs of ATLV (Seiki *et al.*, 1982), HIV-1 (Desai *et al.*, 1986), and HIV-2 (Guyader *et al.*, 1987). No function has yet been assigned to this motif, but in the case of the human retroviral sequence ERV3 and the baboon endogenous virus BaEV, homology extends for over 40 bases with only one mismatch (O'Connell and Cohen, 1984). The TG box is found preferentially in the U3 region but appears in R and U5 as well; therefore its function, if any, should be independent of its relative position. This property is a feature characteristic of enhancers. In the case of S71, the TG box contains one copy of the sequence TGTGCTTTG (Fig. 2, position 282). With the exception of a C at position 286, this sequence is a perfect fit to the enhancer core consensus sequence TGTGG T/A T/A T/A G which is present in a number of viral enhancers including Moloney sarcoma virus (Weiher *et al.*, 1983). A similar sequence has also been found in the U3 region of the GALV LTR (Trainor *et al.*, 1984). Furthermore, the U3 region of the S71 LTR contains two imperfect repeats (Fig. 2). Some repeated sequences have been shown to possess enhancer function in several mammalian retroviruses (Trainor *et al.*, 1984).

The cap site signifies the beginning of the R region, and the corresponding G nucleotide is found at position 370 of the S71 LTR-like sequence (Fig. 2). The alignment agrees in sequence and position with the published cap sites of HERV-K (Ono, 1986), ERV3 (O'Connell and Cohen, 1984), the murine leukemia virus AKV (Etzerodt *et al.*, 1984), SSV (Devare *et al.*, 1983), and GALV (Trainor *et al.*, 1984) (Fig. 3). For AKV and ERV3, the cap site has also been confirmed by mapping the corresponding RNA (Etzerodt *et al.*,



1984; Kato *et al.*, 1987). Significantly, the S71 LTR shows 40–50% matches to the cap site homologous regions of ERV3, AKV, and R-MCF (Vogt *et al.*, 1985). Using our alignment we are able to suggest which of two possible cap sites in ERV1 (Bonner *et al.*, 1982) is more likely to correspond to an actual transcription initiation site. Assumption of the first guanosine (nucleotide 215 in the ERV1 sequence) as cap site reduces the match to the consensus cap homologous sequence from 55 to 35%. This points to the second G at position 217 as the actual cap site. The presumed polyadenylation site CA is located at position 438 in S71 (Fig. 2). The relative spacings of the putative TATAA box, the cap site, and the poly(A) addition site are nearly identical in S71 and SSV. S71 contains no typical poly(A) addition signal (AATAAA), but there is a poly(A) stretch in the corresponding position. The presence of an AATAAA consensus sequence does not seem to be essential for LTR function because the LTR of the GALV SEATO strain is missing this signal.

According to our alignment shown in Fig. 3, the U3 region of S71 LTR encompasses 369 nucleotides, the R region 70 nucleotides, and the U5 region 96 nucleotides. This is in good agreement with the other LTR sequences, with the exception of the HERV-K LTR. We found that the S71 control regions are generally more homologous to the LTRs of the infectious primate retroviruses SSV/SSAV and GALV (SEATO strain) and the murine leukemia viruses AKV and R-MCF (50%) than to the LTRs of the human endogenous retroviral elements (35%).

Chromosomal Localization

The chromosomal location of human S71 sequences was determined by Southern blot analysis of DNA from 13 Chinese hamster × human hybrid cell lines. The probe used was a 3.0-kb *Hin*dIII/*Bam*HI insert of a subclone containing S71 pol and LTR sequences (Fig. 1) and practically no flanking cellular sequences. The flanking cellular regions of S71 turned out to be unsuitable for hybridization experiments since they contain repetitive sequences such as *Alu* repeats (Leib-Mösch *et al.*, 1986). We tried several different fragments from the S71 genome to reduce background due to S71-related sequences and found the pol-LTR *Hin*dIII/*Bam*HI fragment best suited for this purpose.

FIG. 3. Alignment of LTR control elements. LTRs of human endogenous retroviral sequences and those of infectious primate and murine proviruses were aligned to obtain the best possible match in the sequences surrounding the control elements. Human endogenous retroviral sequences: HERV-K (37), ERV1 (4), ERV3 (34), hsRTVL-H (29), O-LTR (38), 4-1 (48). Primate proviruses: SSV (12) and GALV SEATO strain (50). Murine proviruses: Akv (13) and R-MCF (54).

TABLE 1

Human Chromosome Content and Presence of S71 Sequences in Hybrid Cell Lines[a]

Hybrid	Lane	S71 Signal	Human chromosome																					
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
XII-4A-la-aza	3	−	−	−	+	−	−	−	−	+	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−
XII-4A-ld-E HAT	4	+	−	−	+	−	−	−	−	−	−	−	+	−	−	P	−	+	−	+	−	−	−	−	P
XII-2D-ld aza	5	−	−	−	+	−	−	−	−	−	−	−	−	−	+	+	+	P	−	−	−	−	+	−	P
XII-2D-lf HAT	6	+	+	+	+	+	−	+	−	+	+	−	−	+	+	+	+	+	−	+	+	−	+	+	P
XII-4A-aza	7	+	−	−	+	−	−	−	L	+	−	−	+	L	+	−	−	−	−	L	−	L	−	−	−
XVIII-54A aza	8	+	−	+	+	+	+	+	−	−	+	−	+	+	+	+	−	+	−	+	+	+	−	−	L
XXI-22A-g-la	9	+	−	−	−	−	+	+	−	+	−	−	+	−	+	−	+	−	+	+	+	+	+	+	−
XVIII-23C-b aza	10	+	+	+	+	−	−	+	L	+	−	+	+	−	−	+	−	+	−	+	L	−	+	+	P
XVII-10A-12a	11	−	−	−	+	P	−	+	+	−	−	+	+	+	−	−	+	−	−	−	−	+	+	+	+
XXI-23A-2c	12	−	+	−	−	−	−	−	P	+	−	−	P	−	−	+	−	−	+	−	+	−	P	+	+
XVIII-23H-a aza	13	−	−	−	−	+	−	+	−	P	−	−	−	−	+	+	−	+	−	−	+	−	+	+	P
XVIII-54A-la HAT	14	+	−	−	L	−	−	−	−	−	−	−	+	+	L	+	−	−	−	L	+	+	L	−	P
XII-4A-ld-I-HAT	15	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	+

[a] Symbols: +, presence of chromosome or 3.3-kb *Hin*dIII S71 signal; −, absence of chromosome or signal; P, part of chromosome present; L, chromosome present at a frequency of 0.1 or less.

This probe detected a single strongly hybridizing 3.3-kb fragment in Southern blots of *Hin*dIII-digested human DNA and in all hybrids which had retained human chromosome 18 (Table 1, Fig. 4A). This is the size of the fragment expected from the restriction map in Fig. 1. All other human chromosomes were ruled out as possible sites of this major hybridizing fragment by three or more discordant hybrids (Table 2). In addition, numerous weakly hybridizing fragments were detected in human DNA and in most hybrids, but none of these fragments showed a pattern consistent with that of a single locus. It is noteworthy that in hybrid XII-4A-1d-I-HAT, in which human chromosomes 18 and X are retained, only the major 3.3-kb fragment was detected (Fig. 4A, lane 15). We conclude that, although sequences possibly related to the S71 pol gene are widely dispersed in the human genome, there is a single S71-specific locus on chromosome 18 and none on the X chromosome. No cross-hybridization was seen with DNA from Chinese hamster (Fig. 4A, lane 1), mouse, or rat cell lines (data not shown). No attempts were made, as part of this project, to map the weakly hybridizing related sequences present on other autosomes. To define the regional location of the S71 sequence on chromosome 18, we carried out *in situ* hybridization of tritiated S71 probe to human metaphase chromosomes. After autoradiography and Giemsa-banding, a total of 162 metaphase spreads, from two unrelated individuals, was scored for silver grains associated with chromosome 18. Label was detected over chromosome 18 in 49 cells (30%). As illustrated in Fig. 4B, the majority of these grains (43/57 or 75%) were clustered in the center region of the long arm, bands 18q12–q22 with a peak at band 18q21 (26/57 grains or 46%).

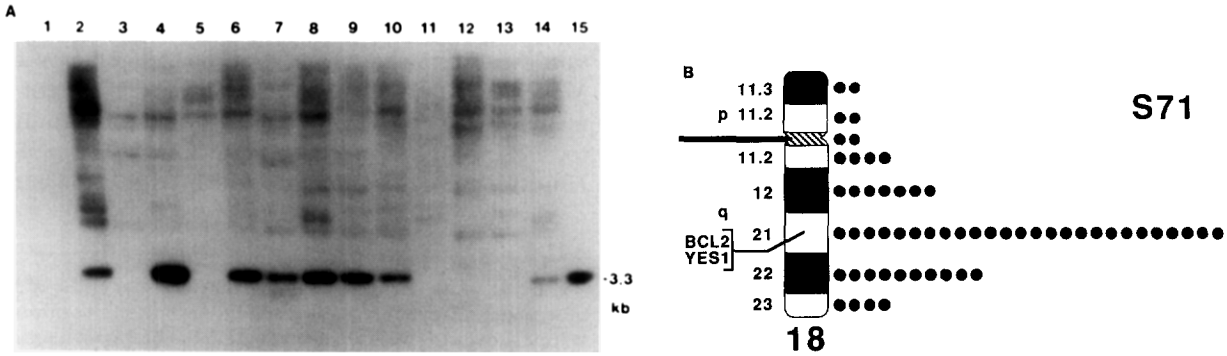


FIG. 4. Chromosomal mapping. **(A)** Hybridization of ^{32}P-labeled S71 DNA with a Southern blot of *Hin*dIII-digested DNA from human × Chinese hamster cells and controls. Lane **1,** Chinese hamster V79/380-6 DNA; lane **2,** human DNA, lanes **3** to **15,** human × Chinese hamster hybrid cell DNAs. Lane numbers correspond to those in Table 1. **(B)** Silver grain distribution on chromosome 18 after *in situ* hybridization with ^{3}H-labeled S71 probe. Data represent combined results from hybridizations to chromosomes from two individuals.

TABLE 2

Correlation of Human S71 Sequences with Human Chromosomes in Chinese Hamster × Human Somatic Cell Hybrids[a]

Hybridization/ chromosome	Human chromosome																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
+/+	2	3	5	2	2	4	0	4	2	1	6	3	4	4	2	4	1	7	5	3	3	3	1
−/−	4	5	2	3	5	3	3	2	5	4	3	4	2	2	3	3	4	4	3	4	1	2	1
+/−	1	0	3	1	0	2	1	2	0	1	1	1	3	3	2	1	1	0	2	1	4	3	2
−/+	6	5	2	6	6	4	7	4	6	7	2	4	3	3	6	4	7	0	3	4	3	5	2
Discordant hybrids	7	5	5	7	6	6	8	6	6	8	3	5	6	6	8	5	8	0	5	5	7	8	4
Informative hybrids	13	13	12	12	13	13	11	12	13	13	12	12	12	12	13	12	13	11	13	12	11	13	6

[a] The numbers of hybrids showing concordant (+/+ or −/−) and discordant (+/− or −/+) segregation with S71 are given for each chromosome. Data on rearranged chromosomes or chromosomes present in 10% or less of cells were excluded.

DISCUSSION

The S71 retroviral element contains a 535-bp region at its 3′ terminus which discloses structural features characteristic of retroviral LTRs. In agreement with previous reports, we found that the nucleotide sequences of human endogenous LTRs, including that of the S71 retroviral element, did not contain sufficient overall homology with other retroviral LTRs to permit identification of these sequences by computer searches of nucleotide sequence databases (Mager and Henthorn, 1984; Steele *et al.*, 1984; O'Connell and Cohen, 1984; Paulson *et al.*, 1985; Ono, 1986). Furthermore, dot matrix analysis of human endogenous LTRs with one another and with other proviral LTRs revealed no significant homologies. The only exceptions were ERV1, ERV3, and BaEV and were due mainly to a stretch of 33–41 nucleotides which has previously been reported to be nearly identical in all three LTRs (Bonner *et al.*, 1982; O'Connell and Cohen, 1984). We detected part of this conserved sequence of the ERV1, ERV3, and BaEV LTRs in various regions of most human endogenous LTRs analyzed (TG box). It is remarkable that even the ERV3 LTR, which has been shown to be functional (Kato *et al.*, 1987), does not exhibit an overall homology to other retroviral LTRs stronger than that of the other human endogenous LTRs. Therefore we conclude that identification of human endogenous retroviral LTR-like sequences should be carried out on the basis of conserved structural features, such as signal sequences for transcription control, rather than overall sequence homology. Interestingly, we found the sequences surrounding the control elements to be conserved and therefore useful for identification of human LTR-like sequences.

Our studies show that the human SSAV-related retroviral element in S71 maps to chromosome 18 band q21. It is of interest that another human endogenous retroviral sequence, ERV1, has also been assigned to chromosome 18 (O'Brien *et al.*, 1983), but *in situ* hybridization revealed more distal localization in bands 18q22–qter (Renan and Reeves, 1987). The full-length proviral element ERV3, parts of which have been shown to be expressed in various human tissues (Kato *et al.*, 1987), is located on chromosome 7. In addition, two endogenous B-type related retroviral elements have been mapped to human chromosomes 1 and 5 (Horn *et al.*, 1986).

The localization of S71 to chromosome 18q21 places it on the same chromosomal band as that of several loci implicated in neoplastic transformation. The yes-1 proto-oncogene, the human homolog of the Yamaguchi sarcoma viral oncogene v-yes-1 (Kitamura *et al.*, 1982; Yoshida *et al.*, 1985), and bcl-2, a breakpoint region involved in chromosome translocations in B-cell neoplasia (Bakhshi *et al.*, 1985, 1987; Pegoraro *et al.*, 1984; Tsujimoto *et al.*, 1984), have been mapped to chromosome 18, subband 18q21.3. An additional breakpoint cluster region also rearranged in many follicular lymphomas has recently been discovered at a distance of more than 20 kb from the first (Cleary *et al.*, 1986a). It has been postulated that the bcl-2 locus encodes a proto-oncogene which becomes activated when translocated to the immunoglobulin heavy chain region at 14q32.3 (Tsujimoto *et al.*, 1984). Similar rearrangements in other types of neoplasms have been shown to be due to aberrant V–J joining involving spurious consensus joining sequences close to oncogenes on another chromosome (Finger *et al.*, 1986). Such rearrangements could also be triggered by recombination between homologous or closely related sequences on different chromosomes. There is evidence suggesting that t(14;18) translocations lead to altered transcription through a *cis*-acting mechanism (Cleary *et al.*, 1986b). Although the IgH enhancer on chromosome 14 is a potential candidate for such a mechanism, it would be required to exert its effect over a distance of more than 370 kb (Seto *et al.*, 1988; Tsu-

jimoto *et al.*, 1987). In this context, a potential transcriptional activity of the S71 LTR-like sequence located on 18q21 would be of special interest. Elucidation of an actual role of S71 sequences in human neoplasia awaits further study, in particular molecular analysis of DNA from tumors bearing rearrangements involving 18q21. The finding of SSAV-related antigens in human leukemic sera and human leukemic cells (Derks *et al.*, 1982; Hehlmann *et al.*, 1983, 1984; Schetters *et al.*, 1985) offers a tantalizing clue that such a role may exist.

ACKNOWLEDGMENTS

We thank Richard Albang for expert technical assistance. We are grateful to Günter Herrmann for valuable advice on computer-assisted evaluation of sequence data. This work was supported by the Deutsche Forschungsgemeinschaft (Grant SFB 324), by the National Institutes of Health (Grant RO1-GM 26105), and by Contract B16-156D of the Commission of the European Communities.

REFERENCES

1. BAKHSHI, A., JENSEN, J. P., GOLDMAN, P., WRIGHT, J. J., MCBRIDE, O. W., EPSTEIN, A. L., AND KORSMEYER, S. J. (1985). Cloning the chromosomal breakpoint of t(14;18) human lymphomas: Clustering around JH on chromosome 14 and near a transcriptional unit on 18. *Cell* **41:** 899–906.
2. BAKHSHI, A., WRIGHT, J. J., GRANINGER, W., SETO, M., OWENS, J., COSSMAN, J., JENSEN, J. P., GOLDMAN, P., AND KORSMEYER, S. J. (1987). Mechanism of the t(14;18) chromosomal translocation: Structural analysis of both derivative 14 and 18 reciprocal partners. *Proc. Natl. Acad. Sci. USA* **84:** 2396–2400.
3. BARTON, D. E., YANG-FENG, T. L., AND FRANCKE, U. (1986). The human tyrosine aminotransferase gene mapped to the long arm of chromosome 16 (region 16q22–q24) by somatic cell hybrid analysis and in situ hybridization. *Hum. Genet.* **72:** 221–224.
4. BONNER, T. I., O'CONNELL, C., AND COHEN, M. (1982). Cloned endogenous retroviral sequences from human DNA. *Proc. Natl. Acad. Sci. USA* **79:** 4709–4713.
5. CALLAHAN, R., CHIU, I.-M., WONG, J. F. H., TRONICK, S. R., ROE, B. A., AARONSON, S. A., AND SCHLOM, J. (1985). A new class of endogenous human retroviral genomes. *Science* **228:** 1208–1211.
6. CHEN, H. R., AND BARKER, W. C. (1984). Nucleotide sequences of the retroviral long terminal repeats and their adjacent regions. *Nucleic Acids Res.* **12:** 1767–1779.
7. CLEARY, M. L., GALILI, N., AND SKLAR, J. (1986a). Detection of a second t(14;18) breakpoint cluster region in human follicular lymphomas. *J. Exp. Med.* **164:** 315–320.
8. CLEARY, M. L., SMITH, S. D., AND SKLAR, J. (1986b). Cloning and structural analysis of cDNAs for bcl-2 and a hybrid bcl-2/immunoglobin transcript resulting from the t(14;18) translocation. *Cell* **47:** 19–28.
9. DEEN, K. C., AND SWEET, R. W. (1986). Murine mammary tumor virus pol-related sequences in human DNA: Characterization and sequence comparison with the complete murine mammary tumor virus pol gene. *J. Virol.* **57:** 422–432.
10. DERKS, J. P. A., HOFMANS, L., BRUNING, H. W., AND VAN ROOD, J. J. (1982). Synthesis of a viral protein with molecular weight of 30,000 (p30) by leukemic cells and antibodies cross-reacting with simian sarcoma virus p30 in serum of a chronic myeloid leukemia patient. *Cancer Res.* **42:** 681–686.
11. DESAI, S. M., KALYANARAMAN, V. S., CASEY, J. M., SRINIVASAN, A., ANDERSEN, P. R., AND DEVARE, S. G. (1986). Molecular cloning and primary nucleotide sequence analysis of a distinct human immunodeficiency virus isolate reveal significant divergence in its genomic sequences. *Proc. Natl. Acad. Sci. USA* **83:** 8380–8384.
12. DEVARE, S. G., REDDY, E. P., LAW, J. D., ROBBINS, K. C., AND AARONSON, S. A. (1983) Nucleotide sequence of the simian sarcoma virus genome: Demonstration that its acquired cellular sequences encode the transforming gene product p28sis. *Proc. Natl. Acad. Sci. USA* **80:** 731–735.
13. ETZERODT, M., MIKKELSEN, T., PEDERSEN, F. S., KJELDGAARD, N. O., AND JØRGENSEN, P. (1984). The nucleotide sequence of the Akv murine leukemia virus genome. *Virology* **134:** 196–207.
14. FINGER, L. R., HARVEY, R. C., MOORE, R. C. A., SHOWE, L. C., AND CROCE, C. M. (1986). A common mechanism of chromosomal translocation in T- and B-cell neoplasia. *Science* **234:** 982–985.
15. FRANCKE, U. (1984). Random X inactivation resulting in mosaic nullisomy of region Xp21.1–p21.3 associated with heterozygosity for ornithine transcarbamylase deficiency and for chronic granulomatous disease. *Cytogenet. Cell Genet.* **38:** 298–307.
16. FRANCKE, U., AND FRANCKE, B. (1981). Requirement of the human chromosome 11 long arm for replication of Herpes Simplex Virus Type 1 in nonpermissive Chinese hamster × human diploid fibroblast hybrids. *Somat. Cell Genet.* **7:** 171–191.
17. FRANCKE, U., BUSBY, N., SHAW, D., HANSEN, S., AND BROWN, M. G. (1976). Intrachromosomal gene mapping in man: Assignment of nucleoside phosphorylase to region 14cen–14q21 by interspecific hybridization of cells with a t(X;14)(p22;q21) translocation. *Somat. Cell Genet.* **2:** 27–40.
18. GATTONI-CELLI, S., KIRSCH, K., KALLED, S., AND ISSELSBACHER, K. J. (1986). Expression of type C-related endogenous retroviral sequences in human colon tumors and colon cancer cell lines. *Proc. Natl. Acad. Sci. USA* **83:** 6127–6131.
19. GUYADER, M., EMERMAN, M., SONIGO, P., CLAVEL, F., MONTAGNIER, L., AND ALIZON, M. (1987). Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature (London)* **326:** 662–669.
20. HARPER, M. E., AND SAUNDERS, G. F. (1981). Localization of single copy DNA sequences on G-banded human chromosomes by in situ hybridization. *Chromosoma* **83:** 431–439.
21. HEHLMANN, R., SCHETTERS, H., ERFLE, V., AND LEIB-MÖSCH, C. (1983). Detection and biochemical characterization of antigens in human leukemic sera that cross-react with primate C-type viral proteins (M_r 30,000). *Cancer Res.* **43:** 392–399.
22. HEHLMANN, R., ERFLE, V., SCHETTERS, H., LUZ, A., ROHMER, H., SCHREIBER, M. A., PRALLE, H., ESSERS, U., AND WEBER, W. (1984). Antigens and circulating immune complexes related to the primate retroviral glycoprotein SiSVgp70: Indicators of early mortality in human acute leukemias and chronic myelogenous leukemias in blast crisis. *Cancer* **54:** 2927–2935.
23. HORN, T. M., HUEBNER, K., CROCE, C., AND CALLAHAN, R. (1986). Chromosomal locations of members of a family of novel endogenous human retroviral genomes. *J. Virol.* **58:** 955–959.
24. KATO, N., PFEIFER-OHLSSON, S., KATO, M., LARSSON, E., RYDNERT, J., OHLSSON, R., AND COHEN, M. J. (1987). Tissue-specific expression of human provirus ERV3 mRNA in human placenta: Two of the three ERV3 mRNAs contain human cellular sequences. *J. Virol.* **61:** 2182–2191.
25. KITAMURA, N., KITAMURA, A., TOYOSHIMA, K., HIRAYAMA, Y., AND YOSHIDA, M. (1982). Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus. *Nature (London)* **297:** 205–208.

26. KRÖGER, G., AND HORAK, I. (1987). Isolation of novel human retrovirus-related sequences by hybridization to synthetic oligonucleotides complementary to the tRNAPro primer-binding site. *J. Virol.* **61:** 2071–2075.
27. LEIB-MÖSCH, C., BRACK, R., WERNER, T., ERFLE, V., AND HEHLMANN, R. (1986). Isolation of an SSAV-related endogenous sequence from human DNA. *Virology* **155:** 666–677.
28. MAGER, D. L., AND FREEMAN, J. D. (1987). Human endogenous retroviruslike genome with type C pol sequences and gag sequences related to human T-cell lymphotropic viruses. *J. Virol.* **61:** 4060–4066.
29. MAGER, D. L., AND HENTHORN, P. S. (1984). Identification of a retrovirus-like repetitive element in human DNA. *Proc. Natl. Acad. Sci. USA* **81:** 7510–7514.
30. MARTIN, M. A., BRYAN, T., RASHEED, S., AND KHAN, A. S. (1981). Identification and cloning of endogenous retroviral sequences present in human DNA. *Proc. Natl. Acad. Sci. USA* **78:** 4892–4896.
31. MAY, F. E. B., AND WESTLEY, B. R. (1986). Structure of a human retroviral sequence related to mouse mammary tumor virus. *J. Virol.* **60:** 743–749.
32. NEEDLEMAN, S. B., AND WUNSCH, C. D. (1970). A general method applicable to the search for similarities in the acid sequence of two proteins. *J. Mol. Biol.* **48:** 443–453.
33. O'BRIEN, S. J., BONNER, T. I., COHEN, M., O'CONNELL, C., AND NASH, W. G. (1983). Mapping of an endogenous retroviral sequence to human chromosome 18. *Nature (London)* **303:** 74–77.
34. O'CONNELL, C. D., AND COHEN, M. (1984). The long terminal repeat sequences of a novel human endogenous retrovirus. *Science* **226:** 1204–1206.
35. O'CONNELL, C., O'BRIEN, S., NASH, W. G., AND COHEN, M. (1984). ERV3, a full-length human endogenous provirus: Chromosomal localization and evolutionary relationships. *Virology* **138:** 225–235.
36. OLIVER, N., FRANCKE, U., AND PELLEGRINO, M. A. (1978). Regional assignment of genes for mannose phosphate isomerase, pyruvate kinase-3, and β2-microglobulin expression on human chromosome 15 by hybridization of cells from a t(15;22)(q14;q13.3) translocation carrier. *Cytogenet. Cell Genet.* **22:** 506–510.
37. ONO, M. (1986). Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. *J. Virol.* **58:** 937–944.
38. PAULSON, K. E., DEKA, N., SCHMID, C. W., MISRA, R., SCHINDLER, C. W., RUSH, M. G., KADYL, L., AND LEINWAND, L. (1985). A transposon-like element in human DNA. *Nature (London)* **316:** 359–361.
39. PEGORARO, L., PALUMBO, A., ERIKSON, J., FALDA, M., GIOVANAZZO, B., EMANUEL, B. S., ROVERA, G., NOWELL, P. C., AND CROCE, C. M. (1984). A 14;18 and an 8;14 chromosome translocation in a cell line derived from an acute B-cell leukemia. *Proc. Natl. Acad. Sci. USA* **81:** 7166–7170.
40. RABSON, A. B., HAMAGISHI, Y., STEELE, P. E., TYKOCINSKI, M., AND MARTIN, M. A. (1985). Characterization of human endogenous retroviral envelope RNA transcripts. *J. Virol.* **56:** 176–182.
41. RENAN, M. J., AND REEVES, B. R. (1987). Chromosomal localization of human endogenous retroviral element ERV1 to 18q22–q23 by in situ hybridization. *Cytogenet. Cell Genet.* **44:** 167–170.
42. REPASKE, R., STEELE, P. E., O'NEILL, R. R., RABSON, A. B., AND MARTIN, M. A. (1985). Nucleotide sequence of a full-length human endogenous retroviral segment. *J. Virol.* **54:** 764–772.
43. SCHETTERS, H., HEHLMANN, R., ERFLE, V., KREEB, G., ROHMER, H., AND SCHMIDT, J. (1985). Antigens and circulating immune complexes related to the primate retroviral glycoprotein SiSV gp70: Prevalence and distribution in human sera. *Eur. J. Cancer Clin. Oncol.* **21:** 687–700.
44. SEIKI, M., HATTORI, S., AND YOSHIDA, M. (1982). Human adult T-cell leukemia virus: Molecular cloning of the provirus DNA and the unique terminal structure. *Proc. Natl. Acad. Sci. USA* **79:** 6899–6902.
45. SETO, M., JAEGER, U., HOCKETT, R. D., GRANINGER, W., BENNETT, S., GOLDMAN, P., AND KORSMEYER, S. J. (1988). Alternative promoters and exons, somatic mutation and deregulation of the Bcl-2-Ig fusion gene in lymphoma. *EMBO J.* **7:** 123–131.
46. SOUTHERN, E. (1979). Gel electrophoresis of restriction fragments. *In* "Methods in Enzymology" (R. Wu, Ed.), Vol. 68, pp. 152–164, Academic Press, New York.
47. STEELE, P. E., MARTIN, M. A., RABSON, A. B., BRYAN, T., AND O'BRIEN, S. J. (1986). Amplification and chromosomal dispersion of human endogenous retroviral sequences. *J. Virol.* **59:** 545–550.
48. STEELE, P. E., RABSON, A. B., BRYAN, T., AND MARTIN, M. A. (1984). Distinctive termini characterize two families of human endogenous retroviral sequences. *Science* **225:** 943–947.
49. TABOR, S., AND RICHARDSON, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84:** 4767–4771.
50. TRAINOR, C. D., SCOTT, M. L., JOSEPHS, S. F., FRY, K. E., AND REITZ, M. S., JR. (1984). Nucleotide sequence of the large terminal repeat of two different strains of gibbon ape leukemia virus. *Virology* **137:** 201–205.
51. TSUJIMOTO, Y., BASHIR, M. M., GIVOL, I., COSSMAN, J., JAFFE, E., AND CROCE, C. M. (1987). DNA rearrangements in human follicular lymphoma can involve the 5′ or 3′ region of the bcl-2 gene. *Proc. Natl. Acad. Sci. USA* **84:** 1329–1331.
52. TSUJIMOTO, Y., FINGER, L. R., YUNIS, J., NOWELL, P. C., AND CROCE, C. M. (1984). Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. *Science* **226:** 1097–1099.
53. VAN BEVEREN, C., RANDS, E., CHATTOPADHYAY, S. K., LOWY, D. R., AND VERMA, I. M. (1982). Long terminal repeat of murine retroviral DNAs: Sequence analysis, host–proviral junctions and preintegration site. *J. Virol.* **41:** 542–556.
54. VOGT, M., HAGGBLOM, C., SWIFT, S., AND HAAS, M. (1985). Envelope gene and long terminal repeat determine the different biological properties of Rauscher, Friend, and Moloney mink cell focus-inducing viruses. *J. Virol.* **55:** 184–192.
55. WEIHER, H., KÖNIG, M., AND GRUSS, P. (1983). Multiple point mutations affecting the simian virus 40 enhancer. *Science* **219:** 626–631.
56. YOSHIDA, M. C., SASAKI, M., SEMBA, K., YAMANASHI, Y., NISHIZAWA, M., SUKEGAWA, J., YAMAMOTO, T., AND TOYOSHIMA, K. (1985). Localization of the human cellular oncogene c-yes-1 to chromosome 18 band q21. *Cytogenet. Cell Genet.* **40:** 786.