Human Endogenous Retrovirus-like Sequences

R. Brack-Werner[1], C. Leib-Mösch[3], T. Werner[2], V. Erfle[1], and R. Hehlmann[4]

A. Introduction

One of the most salient features of the replication strategy used by retroviruses is the transcription of the retroviral (RNA) genome into DNA followed by integration of this DNA product into the host cell genome. The integrated viral DNA copy, termed "provirus", can then serve as a template for the synthesis of further infectious virus particles. Stably integrated proviruses have been found to also persist in the germ line of animal cells. In this case, they have become an endogenous constituent of their host cell's genome and are passed on as stable Mendelian genes from one generation to the next.

Endogenous retroviruses have been detected in a number of vertebrate species, including primates and birds. As a rule, they persist as silent retroviral copies in their host cell's genome since deletions and mutations in the provirus genome have often led to the loss of their pathogenic potential. There are exceptions, however, and activation of endogenous retroviruses has been found to occur spontaneously, as in the case of the leukemogenic ecotropic provirus of the 101 mouse [31]. Other factors, such as treatment with carcinogens [24] and chemicals such as IUdR (iododeoxyuridine) and BrdU (bromodeoxyuridine) and irradiation can also lead to the production of infectious viral particles from endogenous proviruses [32, 55, 25]. Furthermore, the synthesis of pathogenic retroviruses as a result of recombination events between different endogenous proviral sequences has been shown for the highly leukemogenic murine MCF (mink cell focus-forming) virus [7, 13].

Besides delivering the basis for the induction of potentially pathogenic viral particles, the biological potential of endogenous retroviruses can be found on at least two additional levels. First, even replication-defective proviruses can give rise to products such as the p15E envelope-related proteins, which have been shown to possess immunosuppressive and anti-inflammatory activity [59]. Second, insertion of a proviral sequence can take place within host cell genes, causing changes in expression of the latter (insertion mutagenesis). Furthermore, once the provirus is installed it can influence the expression of adjacent cellular sequences by virtue of its own transcription control signals [reviewed in 40]. Some examples illustrating the mutagenic potential of tumor-associated proviral insertion have been reported for intracisternal A-type particles (IAP) in mouse plasmacytoma [6], MoMuLV-induced tumors [56], and avian leukosis virus (ALV)-induced erythroblastosis [14, 16, 17].

The fact that almost all vertebrate species analyzed to date have been shown to contain endogenous retroviruses makes it highly conceivable that these are also an integral component of the human genome. The evidence pointing to the existence of human endogenous retroviruses runs in three lines. First, particles with

[1] GSF-Abt. für Molekulare Zellpathologie and [2] GSF-Institut für Säugetiergenetik, D-8042 Neuherberg, FRG
[3] Medizinische Poliklinik der Universität München, D-8000 Munich, FRG
[4] III. Medizinische Klinik Mannheim der Universität Heidelberg, Wiesbadenerstr. 7–11, D-6800 Mannheim 31, FRG

Table 1. Copy number and chromosomal localization of human endogenous retroviral sequences

Endogenous retroviral sequence	Length (kb)	Copy no. per haploid genome	Chromosomal localization	Reference
H51 related	4.4	35–50	dispersed to multiple human chromosomes	[61]
4-1 related	8.8	35–50		
ERV1	8.0	1	18q22–q23	[41, 52]
additional ERV1-related sequences	n.d.	11	n.d.	[2]
ERV3	9.9	1	7	[42]
S71	6	1	18q21–q22	[3]
S71-related	n.d.	35	n.d.	
HuRRS-P	8.1	20–40	n.d.	[29]
RTVL-H	5.8	800–1000	n.d.	[34]
HLM	9.7	50	chromosomes 7, 8, 11, 14, and 17	[23]
HM	6–8	30–40	n.d.	[10]
HERV-K	9.5	50	n.d.	[45]
THE1 repeats	2.3	10000	n.d.	[11]
THE solitary LTRs	0.35	10000	n.d.	

n.d., not determined.

retrovirus-like morphology have been visualized by electron microscopy of various human tissues and cell lines, many of which are of neoplastic origin [1, 28, 33, 38]. The second line of evidence is the detection of proteins related to exogenous animal retroviruses in human tissues or body fluids [18, 21]. We previously reported that antibodies against structural components of the simian sarcoma-associated virus (SSAV) recognize proteins in leukemic sera. Proteins immunologically related to the p30 constituent of the SSAV group-specific antigen were detected only in sera from patients with acute leukemia and CML blast crisis, but not in nonleukemic controls [19]. Furthermore, proteins related to the SSAV envelope gp70 protein seem to be of diagnostic value for the prognosis of patients with acute leukemias or CML blast crisis [20].

The third line of evidence is the existence of numerous retrovirus-like sequences which are indigenous to the human genome. These endogenous retroviral sequences constitute a complex variety of retroviral information in the human genome. A conservative estimate based on the copy number of endogenous retroviral sequences published to date (Table 1) shows that at least 0.6% of the human genome consists of retrovirus-like elements. The actual percentage is probably much higher, since new families of retrovirus-related sequences are being discovered continuously.

B. Identification and Isolation of Human Endogenous Retroviral Sequences

A number of different strategies have been employed to identify retrovirus-related sequences in the human genome (Table 2). Human C-type retrovirus-related sequences were initially discovered by utilizing probes from primate endoge-

Table 2. Identification and isolation of human endogenous retroviral sequences

Source of DNA for human library	Hybridization probe used for screening of human DNA library	Stringency	Identification of	Group of human endogenous retroviral sequences	Reference
Human fetal liver	gag-pol-related fragment from African green monkey endogenous retroviral sequence	low	λH51	C-type-related	[36]
	fragment from λH51-pol-related sequence	high/low	~30 additional retrovirus-related sequences		[53, 58]
	*pol-related fragment from chimpanzee endogenous retroviral sequence	low	ERV1		[2]
	and Baev LTR probe	low	ERV3		[42]
Burkitt's lymphoma	SSAV proviral DNA and various fragments from different regions of the SSAV genome	low	S71		[30]
	DNA fragment containing the retrovirus-related region in S71	high	clones only from S71 genomic locus		
Human male blood cells	Synthetic oligonucleotide complementary to murine $tRNA^{Pro}$	low and medium	Pλ1		[29]
	LTR probe from P1	high	HuRRS-P		
–	–	–	RTVL-H1		[34]
Human embryonic fibroblasts	Various RTVL-H1 fragments	stringent	RTVL-H2		[35]
Human fetal liver	MMTV provirus	low	HLM-2	A-, B-, and D-type-related sequences	[4]
	gag-pol of MMTV provirus	low	HM16		[10]
	pol region of Syrian hamster IAP	low	HERV-K		[45]
Human breast cancer cell line	MMTV provirus as well as gag-pol and LTR region of MMTV provirus	low	NMWV4		[37]
n.s.	Total human genomic DNA and cloned Alu-family member	n.s.	THE1 repeats	retroposons with LTRs	[62]

* ERV3 was isolated by employing the same chimpanzee endogenous retroviral fragment together with the BaEV LTR probe.

nous retroviral sequences for low-stringency hybridization of human genomic libraries. In 1981, Martin and co-workers used a cloned segment of African green monkey DNA which specifically hybridized with C-type murine and primate proviruses to identify related sequences in the human genome [36]. One of these sequences was isolated from a human DNA library (clone λ51-1). High-stringency hybridization of the same library with a retrovirus-related probe from 51-1 yielded over 30 additional type-C retrovirus-related sequences [53]. One of these (4-1) was also shown to contain a full-length provirus [50, 54]. An additional full-length provirus (NP-2) was cloned by low-stringency hybridization using a 51-1 pol probe [58]. Another human C-type retroviral sequence (ERV1) was isolated by Bonner et al. [2] with the help of a fragment from a cloned chimpanzee retrovirus-like sequence homologous to the polymerase genes of the baboon endogenous virus (BaEV) and the Moloney murine leukemia virus (MoMuLV). Low-stringency screening of a human genomic library with the same cloned chimpanzee fragment and a probe containing the BaEV LTR led to the isolation of a full-length human endogenous provirus termed ERV3 [42].

Our initial interest in human endogenous retroviral sequences arose from the observation mentioned above that human sera contain proteins immunologically related to structural components of SSV/SSAV and the closely related gibbon ape leukemia virus (GALV) [19]. Low-stringency Southern blot hybridization of a number of human genomic DNAs with various probes derived from the SSAV genome showed multiple SSAV-related sequences in the human genome [30]. Therefore, we decided to use a direct approach and screen a human DNA library with a probe containing the complete SSAV provirus as well as probes derived from various regions of the SSAV genome under low-stringency conditions. The initial hybridization yielded quite a few positive plaques corresponding to at least 35 copies of SSAV-related sequences per haploid genome. Washing the filters under higher stringency conditions caused a number of the positive signals to grow more or less weaker or to disappear altogether, which indicates that the retrovirus-related sequences detected during initial screening were of varying homologies to SSAV. One clone which gave a particularly strong hybridization signal with an SSAV pol-env probe was termed S71 and chosen for further analysis. The region containing the retrovirus-related sequences in S71 was used for renewed screening of the human DNA library, this time under high-stringency conditions. All positive clones obtained overlapped with clone S71 to some extent, comprising about 36 kb of the S71 genomic locus. Contrary to Repaske et al. [53], we had not been able to isolate any additional retrovirus related sequences by high-stringency screening of a human DNA library with S71 probes. This suggests that the SSAV-related human endogenous retroviral sequences are less similar to each other than the members of the 51-1/4-1 family.

A further family of C-type retrovirus-related sequences was isolated by virtue of the fact that retroviruses contain short sequences complementary to tRNA molecules, which are used as primers for reverse transcription. Screening of a human DNA library with an oligonucleotide complementary to $tRNA^{Pro}$ (murine) yielded a human LTR-like sequence which could be utilized for renewed screening and isolation of a retrovirus-like sequence termed HuRRS-P [29]. Finally, one multicopy endogenous retrovirus-like element termed "RTVL-H" was discovered fortuitously during attempts to clone a region of the human β-globin gene cluster region [34]. Additional RTVL-H elements were isolated by screening a human DNA library with RTVL-H1 probes [35].

The strategy of direct screening of human DNA libraries with probes derived from recombinant rodent proviruses was

used to initially identify a second large group of human endogenous retroviral elements (Table 2). This group consists of sequences related to the B-type mouse mammary tumor virus (MMTV) as well as to the Syrian hamster IAP and to the D-type squirrel monkey retrovirus (SMRV). Members of this group were isolated by low-stringency hybridization with DNA probes encompassing various regions of the MMTV genome [5, 10, 37] or by employing a probe from the polymerase gene of the Syrian hamster IAP [45].

The final group of human endogenous retroviral sequences consists of elements flanked by two sequences with the hallmarks of retroviral long-terminal repeats (LTRs); [48]. This group of elements, designated THE 1 repeats by Sun et al. [62], was isolated from a human DNA library as clones hybridizing to human genomic DNA but not to an Alu family member. Like the other endogenous retroviral sequences discussed here, these elements possess features indicative of having been generated by the reverse flow of genetic information from RNA to DNA. Such elements are known collectively as retroposons [67].

C. Chromosomal Localization

Some human retroviral elements occur singly or in a few copies in the human genome enabling their assignment to distinct chromosomes (Table 1). Hybridization of DNA from rodent x human somatic cell hybrids revealed that the full-length retroviral sequence ERV3 resides at a single locus on human chromosome 7 [42]. The long arm of chromosome 18 carries two incomplete proviral sequences: S71 at band q21 [3] and ERV1 at bands q22–q23 [41, 52]. The chromosomal location of these retroviral elements was determined by Southern blot analysis of DNA from hybrid cell lines as well as by in situ hybridization. The members of the closely related 4-1 and 51-1 families were found to be widely dispersed over the human genome, indicating that the 50–100 copies of these sequences may have been generated by amplification processes [61]. Clone λNP-2, a full-length proviral sequence related to 4-1 and 51-1, was localized in two copies on the Y chromosome. Conservation of cellular flanking sequences suggests that the second copy results from gene duplication, rather than from provirus insertion [58]. Some members of the B-type-related multicopy HLM-family were mapped to chromosomes 1, 5, 7, 8, 11, 14, and 17 [23]. The RTLV-H elements and the THE 1 repeats occur in much higher copy numbers in the human genome than the other retroviral elements (Table 1).

D. Organization of Human Endogenous Retrovirus-like Sequences

Hybridization studies and nucleotide sequence analysis showed that each group of human endogenous retroviral sequences has one or more members resembling full-length proviruses; i.e., their retroviral sequences are arranged 5′LTR-gag-pol-env-LTR3′ as in proviruses resulting from infection with exogenous viruses (Fig. 1, MoMuLV). In the group of C-type-related retroviral sequences, 4-1 and ERV3 show a proviral organization [54, 42], and in the group of B-type-related sequences this holds true for the HERV-K family [46] (Fig. 2). However, 4-1 and ERV3 both contain stop codons and frame shifts in their nucleotide sequence, precluding the synthesis of infectious virus particles. In 4-1, complete nucleotide sequence analysis revealed these mutations to be dispersed over the whole genome inactivating all three retroviral genes [54; see also 22]. It seems that these sequences are of sufficient danger to the human cell to warrant an efficient blockade of their expression.

A great proportion of human endogenous retroviral elements consist of retrovirus-related sequences organized in a manner suggestive of truncated provirus-

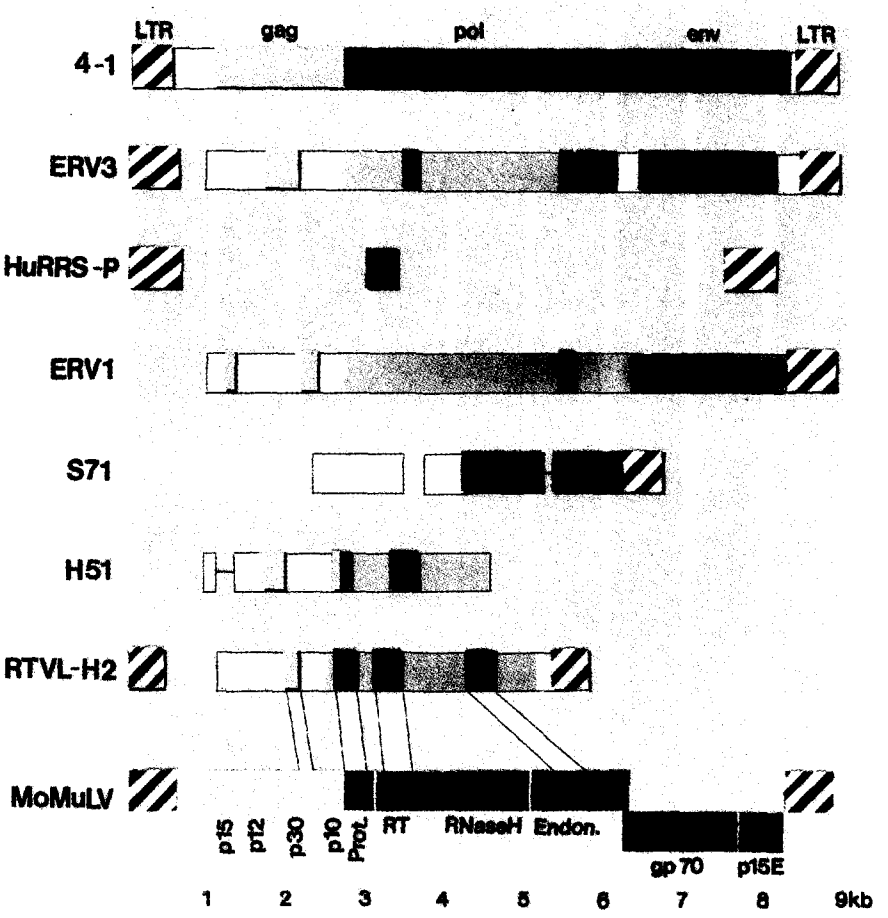

Fig. 1. Genomic organization of C-type-related human endogenous retroviral sequences. The genomic organization of C-type-related human retroviral elements was deduced primarily from sequence comparison with the MoMuLV genome (depicted at *bottom*). The gene assignment of *lightly shaded regions* was inferred on the basis of their location between sequenced regions or from hybridization data. LTR-like sequences are *hatched*. *Horizontal lines* in H51 and S71 delineate deleted sequences. The *left (gray) box* in the S71 element marks the gag region. The *right (white) box* immediately adjacent to the S71 pol sequence shows the minimal extent of nonretroviral sequences in S71. References: 4-1 [54, 60], ERV3 [42, 43], HuRRS-P [29], ERV1 [2], S71 [3, 30], H51 [53], RTVL-H2 [35], MoMuLV [57]

es. These elements may lack only a small part of the retroviral genome, such as one of the two LTRs at either end (ERV1; Fig. 1), or they may be completely devoid of sequences corresponding to one or more proviral genes. We have found the SSAV-related human retroviral element S71 to provide a good example for such a truncated endogenous provirus. By hybridization of molecular clone S71 with probes derived from various SSAV genes, the S71 retroviral element was delineated to a region of approximately 6 kb. Since a full-length C-type provirus ranges from 8.5 to 9.5 kb in length, the S71 retroviral element is obviously lacking part of the retroviral genome. Interestingly, the retroviral region in S71 is surrounded by Alu repeats, which, although nonviral, are also retroposons. Other human C-type-related retroviral elements have also been reported to be associated with retroposons, such as the Alu or the Kpn I family of reiterated sequences [53, 60].

The retroviral region in S71 contains sequences related to the gag and pol genes of SSAV. In addition, hybridization with an SSAV LTR probe suggested the presence of an LTR-like sequence. To obtain a better idea of the organization of the pol-related sequences in S71 we compared the sequence of the 3′ half of the S71 retroviral element with the pol-gene sequence of the Moloney murine leukemia virus [57]. The pol-gene sequence of retroviruses codes for three ac-

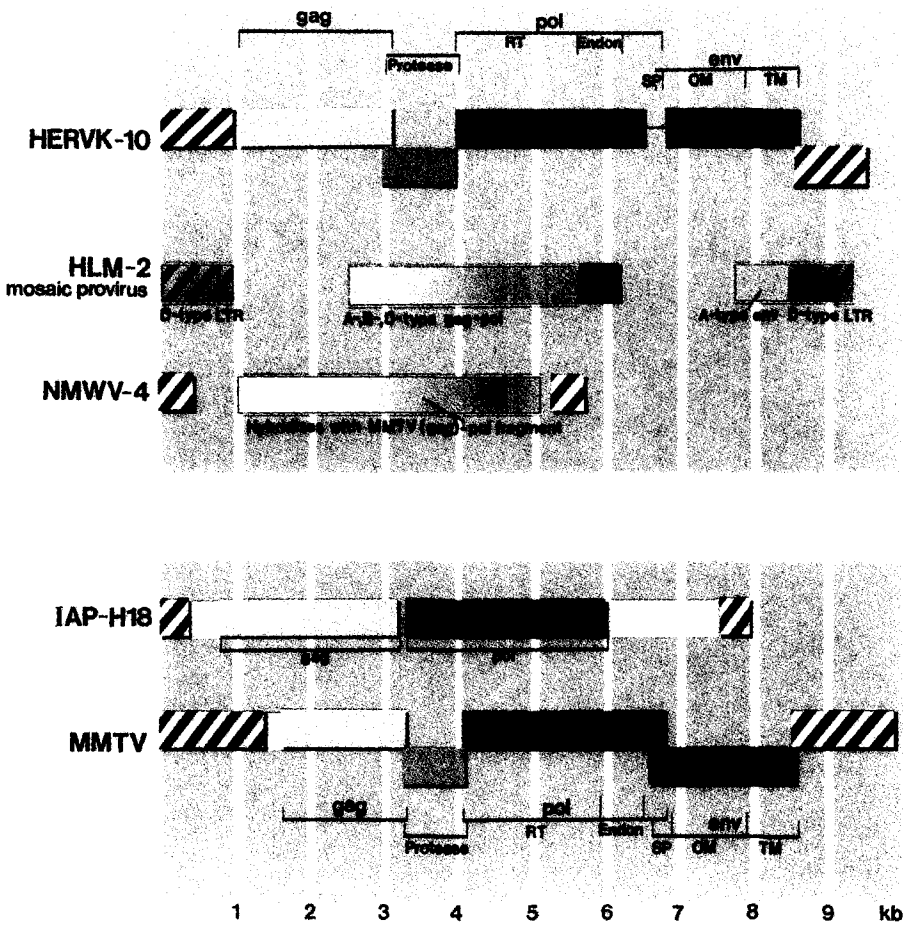

Fig. 2. Genomic organization of A-, B-, and D-type-related human endogenous retroviral sequences. The genomic organization of the human retroviral elements was deduced from sequence comparison with the genome of the Syrian hamster IAP H18 and/or the MMTV provirus (both shown at the *bottom*) or from hybridization data (*lightly shaded regions* of HLM-2 and NMWV-4). The B-type-related endogenous element HM16 [10] was omitted since, aside from the presence of a 2.1-kb pol sequence and restriction fragments containing repeated sequences, the data available did not allow further deduction of the genomic organization of HM16. References: HERVK-10 [46], HLM-2 [5], NMWV-4 [37], IAP-H18 [44], MMTV [39]

tivities: the RNA-directed DNA polymerase (reverse transcriptase), a ribonuclease H, responsible for degradation of viral RNA in RNA·DNA hybrids, and an endonuclease which is essential for integration of the viral information into the host cell genome. In the polymerase genes of C-type retroviruses these activities are arranged 5′ reverse transcriptase – RNAse H – endonuclease 3′ [26]. We found the polymerase-related sequences in S71 to correspond to a region of the MoMuLV pol gene beginning in the 3′ half of the reverse transcriptase domain and extending through the RNase H and most of the endonuclease domain (Fig. 1). With the exception of a small deletion at the 5′ terminus of the endonuclease domain (indicated by a horizontal line in Fig. 1), the S71 pol sequence aligns to the corresponding region of the MoMuLV pol gene in a colinear manner. Thus, the S71 pol sequences show the same structural organization as the corresponding sequences of infectious C-type retroviruses. Translation of the S71 pol nucleotide sequence yields an amino sequence which is 40%–60% identical with the MoMuLV pol sequence, depending on the region of the polymerase gene used for comparison. The S71 pol amino acid sequence contains three stop codons, one each in the deduced reverse transcriptase and RNAse H domains and one in the endonuclease sequence. Therefore, the situation in S71 is similar to the pol region in the C-type-related 4-1 [54] and the B-type-related HM16 element [10], in that numerous stop codons seem to serve the purpose of preventing synthesis of functional polymerase proteins from these endogenous retroviral se-

quences. Indeed, the polymerase sequence of only one human endogenous retroviral element, the B-type-related HERV-K [46], has yet been reported to constitute an open reading frame long enough to allow synthesis of full-length polymerase proteins.

In a biological sense, expression of sequences enabling random reverse flow of genetic information from RNA to DNA would pose a great threat for the evolutionary stability of the human genome. A prerequisite for the maintenance of such sequences in the human genome is therefore a very rigid control mechanism precluding their random expression. The numerous stop codons and frame shifts observed in the pol sequences of C-type-related human endogenous retroviral elements may be a significant factor contributing to this stringent control.

Replication of viral RNA in the host cell leads to the duplication of sequences specific for the 5′ and 3′ ends of the viral RNA. Therefore, the integrated provirus is flanked by long terminal repeats (LTRs) which in the case of mammalian C-type retroviruses are typically 500–600 bp in length [8]. It is still not clear whether endogenous retroviral sequences were generated via the same replication mechanism essential for the spread of infectious retroviruses. However, it is remarkable that quite a few human endogenous retroviral elements are also flanked by LTRs (Figs. 1 and 2). Like the LTRs of infectious proviruses, the endogenous LTRs contain signal sequences implicated in transcriptional control. Indeed, the LTR of the C-type-related endogenous retroviral sequence ERV3 was recently demonstrated to drive transcription of the retroviral and adjacent cellular sequences in a tissue-specific manner [43, 27]. In some human retroviral elements, e.g., ERV1 (Fig. 1), LTR-like sequences were not discovered as duplicated sequences at both ends of the retroviral element. Rather, they were identified as possessing the same sequence features and structural organization as the LTRs of infectious proviruses.

Figure 3 shows the LTR structure of a typical mammalian C-type provirus. The boundaries of retroviral LTRs are formed by inverted repeats, beginning with TG and ending with CA. The LTRs consist of three entities: the U3, R, and the U5 region. The U3 region contains signal sequences necessary for transcription initiation, including the CCAAT and TATAA boxes, and an enhancer region, which often contains directly repeated sequences. However, it should be pointed out that at least three human endogenous LTRs lack a CCAAT box (hsRTVL-H [34]; O-LTR [48]; 4-1 [60].) The beginning of the R region is marked by the cap site, a G nucleotide. As a rule, the R region also contains a poly A signal, although this signal seems to be dispensable for LTR function in some cases [64], and a poly A addition site (CA) which marks the end of the R region. The remaining sequence, including the 3′ inverted repeat counterpart, makes up the U5 region.

We determined the nucleotide sequence of a 535 bp region located at the 3′ terminus of the S71 retroviral element directly adjacent to the pol-related sequences. By comparison of the S71 sequence with the aligned nucleotide sequences of 11 LTRs, six of which were derived from human endogenous retroviral elements and four from infectious proviruses [3], we were able to identify all salient features characteristic for mammalian C-type proviral LTRs. In addition, alignment of the human endogenous LTR sequences demonstrated a common sequence motif, all or part of which is reflected in five of the six human endogenous LTRs analyzed. In the S71 LTR-like sequence this motif contains a 9-bp region with eight matches to the enhancer core consensus sequence present in a number of viral enhancers [66]. Sequences with potential enhancer function, such as this common motif or direct repeats, two of which are also contained in the S71 LTR-like sequence, may enable human endogenous LTRs to influence the expression of adjacent cellular genes in *cis*.

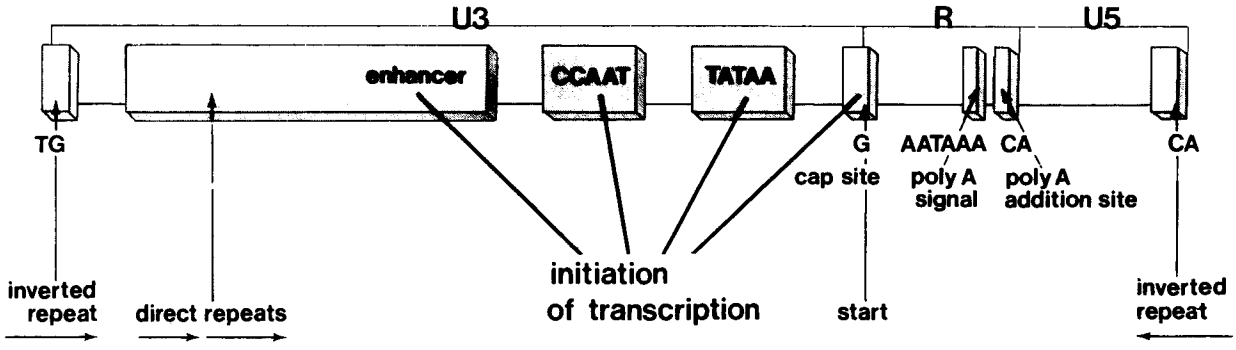

Fig. 3. Structure of mammalian C-type proviral LTRs

Hybridization analysis of a 3-kb restriction fragment directly bordering the 5′ terminus of the S71 pol sequences shows this region to contain sequences related to the gene coding for the group-specific antigen (gag) of SSAV. Preliminary sequence analysis disclosed the S71 gag-related region to encompass about 1 kb (H. Backhaus, personal communication; Fig. 1). Furthermore, the gag- and pol-related sequences in S71 are separated from each other by a sequence about 0.5–1 kb in length. This part of S71 discloses no similarity to any known retroviral genes or control elements (Fig. 1), and therefore is most likely of cellular origin.

Overall, the structure of the S71 retroviral element shows remarkable parallels to the genomic organization of the *sis* oncogene-transducing retrovirus SSV (simian sarcoma virus [12]). This acutely transforming virus is thought to have arisen by recombination of SSAV with the cellular homologue of the v-*sis* sequence which codes for a component of the platelet-derived growth factor. The SSV genome lacks a substantial portion of the pol gene, and most of the envelope gene has been replaced by cellular sequences *(sis)*. In analogy, the S71 retroviral element is likewise missing part of the pol gene – although the pol deletion is not as extensive as in SSV – as well as envelope sequences. In addition, the S71 element also contains nonretroviral sequences embedded in retroviral sequences. These analogies imply that the generation of the oncogene-transducing retrovirus SSV and the human endogenous SSV/SSAV-related sequence in S71 may have involved similar mechanisms.

E. Expression of Human Endogenous Retroviral Sequences

Although all human endogenous retroviral elements examined so far are replication defective, some of them have been shown to be transcriptionally active in human tissues and cell lines. Several discrete mRNA species hybridizing to LTR and env DNA probes derived from the 4-1 element were detected in human placenta, spleen, normal colon mucosa, and primary colon cancers, as well as in colon cancer cell lines (SW1116, HCT, Caco2), in a breast carcinoma cell line (T47D), and in a T-cell acute lymphocytic cell line (8402) [15, 50, 51]. In colon tumors an increase of env-LTR-related 1.7- and 3.0-kb transcripts was observed compared with normal colon tissue, whereas a 3.6-kb transcript abundant in normal colon mucosa was decreased in tumor cells [15]. Partial cDNA clones of 4-1 env-related mRNA transcripts were isolated from human placenta. Sequence analysis of two placental cDNA clones, however, revealed in-frame termination codons, so that neither of them could encode full-length env proteins [51].

The env region of another C-type-related full-length retroviral element, ERV3, contains a long open reading frame corresponding to approximately

650 amino acids. This potential polypeptide was found to exhibit features characteristic of retroviral glycoproteins, including several potential glycosylation sites and sequences indicative of transmembrane proteins [9]. An ERV3 env-specific c-DNA of 2.85 kb was isolated from a human fetal cDNA library and found to be identical to ERV3 by DNA sequence analysis. Three polyadenylated RNAs of 9, 7.3, and 3.5 kb were identified in human placental chorion and characterized by Northern blotting and S1 nuclease mapping [27]. The RNAs were found to be spliced mRNAs lacking the gag and most of the pol gene. The two larger mRNAs extended through the polyadenylation site in the 3′ LTR and contained adjacent cellular sequences.

We have also identified S71-related cDNA clones in a human osteosarcoma and a placenta cDNA library, indicating that S71-related sequences are expressed in these tissues. Sequence analysis of the cDNA inserts in these clones is currently in progress (Leib-Mösch et al. manuscript in preparation).

The MMTV-related human proviral sequence HERV-K was found to be expressed as an 8.8-kb full-length mRNA transcript in cell lines from breast carcinoma (T47D), gastric carcinoma (Kato-III), malignant melanoma (HMT-2), and epidermoid carcinoma (HEp-2, Hela). Stimulation of HERV-K expression was observed in steroid-treated T47D cells [47].

In spite of abundant transcription of endogenous retroviral sequences in various cells, the corresponding proteins have not yet been identified. The only case in which there is at least some indirect evidence for expression at the protein level is the truncated retroviral element ERV1. Antibodies were raised against a synthetic undecapeptide, the sequence of which was derived from the gag-related region of ERV1. These antibodies identified a 75 kD protein in renal adenocarcinoma, placenta, and trophoblastic tumors [63, 65].

However, it should be pointed out that mammalian C-type retroviral gag proteins were recently shown to share an antigenic determinant with the snRNP-associated 70 kD protein [49]. This most likely represents an example of molecular mimicry resulting from convergent evolution of otherwise unrelated proteins. Although the sequence of the antigenic determinant common to retroviral $p30^{gag}$ and the 70 kD protein is not contained in the ERV1 gag synthetic peptide, involvement of a similar phenomenon cannot be ruled out at this stage.

F. Concluding Remarks

To summarize briefly, endogenous retroviral elements are a substantial component of the human genome. In structure, they resemble either full-length or truncated proviruses. Retrovirus-related sequences seem to be dispersed to all human chromosomes; however, single-copy retroviral elements could be assigned to distinct chromosomal loci. Although their function is still unknown, RNA expression has been detected in various human materials, including tumor-derived tissues and cell lines as well as placenta. Human retroviral elements exhibit a number of features giving them a potential for involvement in carcinogenesis. One of them is their likelihood of being transposed, thereby enabling them to act as insertional mutagens. Other intrinsic properties of retroviral elements relevant for their tumorigenic potential reside in their sequence information. These include the potential immunosuppressive activity of p15E envelope-related proteins and the ability of retroviral LTRs to influence transcription of adjacent cellular genes. Besides the enlightenment of a possible contribution of retroviral elements to the evolutionary versatility of the human genome, the possible role of human endogenous retroviral sequences in pathogenesis is currently a subject of great interest.

References

1. Boller K, Frank H, Löwer J, Löwer R, Kurth R (1983) Structural organization of unique retrovirus-like particles budding from human teratocarcinoma cell lines. J Gen Virol 64:2549–2559
2. Bonner TI, O'Connell C, Cohen M (1982) Cloned endogenous retroviral sequences from human DNA. Proc Natl Acad Sci USA 79:4709–4713
3. Brack-Werner R, Barton DE, Werner T, Foellmer BE, Leib-Mösch C, Francke U, Erfle V, Hehlmann R (1989) Human SSAV-related endogenous retroviral element: LTR-like sequence and chromosomal localization to 18q21. Genomics 4, 68–75
4. Callahan R, Drohan W, Tronick S, Schlom J (1982) Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. Proc Natl Acad Sci USA 79:5503–5507
5. Callahan R, Chiu IM, Wong JF, Tronick SR, Roe BA, Aaronson SA (1985) A new class of endogenous human retroviral genomes. Science 228:1208–1211
6. Canaani E, Dreazen O, Klar A, Rechavi G, Ram D, Cohen JB, Givol D (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
7. Chattopadhyay SK, Cloyd MW, Linemeyer DL, Lander MR, Rands E, Lowy DR (1982) Cellular origin and role of mink cell focus-forming viruses in murine thymic lymphomas. Nature 295:25–31
8. Chen HR, Barker WC (1984) Nucleotide sequences of the retroviral long terminal repeats and their adjacent regions. Nucleic Acids Res 12:1767–1779
9. Cohen M, Powers M, O'Connell C, Kato N (1985) The nucleotide sequence of the env gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA. Virology 147:449–458
10. Deen KC, Sweet RW (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
11. Deka N, Willard CR, Wong E, Schmid CW (1988) Human transposon-like element insert at a preferred target site. Nucleic Acids Res 16:1143–1151
12. Devare SG, Reddy EP, Law JD, Robbins KC, Aaronson SA (1983) Nucleotide sequence of the simian sarcoma virus genome: demonstration that its acquired cellular sequences encode the transforming gene product $p28^{sis}$. Proc Natl Acad Sci USA 80:731–735
13. Evans LH, Cloyd MW (1985) Friend and Moloney murine leukemia viruses specifically recombine with different endogenous retroviral sequences to generate mink cell focus-forming viruses. Proc Natl Acad Sci 82:459–463
14. Fung YK, Lewis WG, Crittenden LB, Kung HJ (1983) Activation of the cellular oncogene c-erbB by LTR insertion: molecular basis for induction of erythroblastosis by avian leukosis virus. Cell 33:357–368
15. Gattoni-Celli S, Kirsch K, Kalled S, Isselbacher KJ (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
16. Goodwin RG, Rottman FM, Callaghan T, Kung H-J, Maroney PA, Nilsen TW (1986) c-erbB activation in avian leukosis virus-induced erythroblastosis: multiple epidermal growth factor receptor mRNAs are generated by alternative RNA processing. Mol Cell Biol 9:3128–3133
17. Hayward WS, Neel BG, Astrin SM (1981) Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis. Nature 290:475–480
18. Hehlmann R (1976) RNA-tumorviruses and human cancer. Current Topics in Microbiology and Immunology Vol 73, Springer Verlag, Berlin-Heidelberg-New York:141–215
19. Hehlmann R, Schetters H, Erfle V, Leib-Mösch C (1983) Detection and biochemical characterization of antigens in human leukemic sera that crossreact with primate C-type viral proteins (Mr 30000). Cancer Res 43:392–399
20. Hehlmann R, Erfle V, Schetters H, Luz A, Rohmer H, Schreiber MA, Pralle H, Essers U, Weber W (1984) Antigens and circulating immune complexes related to the primate retroviral glycoprotein SiSVgp70. Cancer 54:2927–2935
21. Hehlmann R, Schetters H, Leib-Mösch C, Erfle V (1984) Current understanding of

viral etiology in leukemia. Recent Results in Cancer Res Vol 93, Springer Verlag, Berlin-Heidelberg-New York 93: 1–28
22. Hehlmann R, Brack-Werner R, Leib-Mösch C (1988) Human endogenous retroviruses. Leukemia 2 (12S): 167S–177S, 1988
23. Horn TM, Huebner K, Croce C, Callahan R (1986) Chromosomal locations of members of a family of novel endogenous human retroviral genomes. J Virol 58: 955–959
24. Irons RD, Stillman WS, Cloyd MW (1987) Selective activation of endogenous ecotropic retrovirus in hematopoietic tissues of B6C3F1 mice during the preleukemic phase of 1,3-butadiene exposure. Virology 161: 457–462
25. Janowski M, Merregaert J, Boniver J, Maisin JR (1985) Proviral genome of radiation leukemia virus: molecular cloning of biologically active proviral DNA and nucleotide sequence of its long terminal repeat. J Virol 55: 251–255
26. Johnson MS, McClure MA, Feng D-F, Gray J, Doolittle RF (1986) Computer analysis of retroviral pol genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. Proc Natl Acad Sci USA 83: 7648–7652
27. Kato N, Pfeifer-Ohlsson S, Kato M, Larson E, Rydnert J, Ohlsson R, Cohen M (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
28. Keydar I, Ohno T, Nayak R, Sweet R, Simoni F, Weiss F, Karby S, Mesa-Tejada R, Spiegelman S (1984) Properties of retrovirus-like particles produced by a human breast carcinoma cell line: immunological relationship with mouse mammary tumor virus proteins. Proc Natl Acad Sci USA 81: 4188–4192
29. Kröger B, Horak I (1987) Isolation of novel human retrovirus-related sequences by hybridization to synthetic oligonucleotides complementary to the $tRNA^{Pro}$ primer-binding site. J Virol 61: 2071–2075
30. Leib-Mösch C, Brack R, Werner T, Erfle V, Hehlmann R (1986) Isolation of an SSAV-related endogenous sequence from human DNA. Virology 155: 666–667
31. Leib-Mösch C, Schmidt J, Etzerodt M, Pedersen FS, Hehlmann R, Erfle V (1986) Oncogenic retrovirus from spontaneous murine osteomas. Virology 150: 96–105
32. Lesser J, Lasneret J, Canivet M, Emanoil-Ravier R, Peries J (1986) Simultaneous activation by 5-azacytidine of intracisternal R particles and murine intracisternal-A particle related sequences in Syrian hamster cells. Virology 155: 249–256
33. Löwer R, Löwer J, Frank H, Harzmann R, Kurth R (1984) Human teratocarcinomas cultured in vitro produce unique retrovirus-like viruses. J Gen Virol 65: 887–898
34. Mager DL, Henthorn PS (1984) Identification of a retrovirus-like repetitive element in human DNA. Proc Natl Acad Sci USA 81: 7510–7514
35. Mager DL, Freeman JD (1987) Human endogenous retrovirus-like genome with type C pol sequences and gag sequences related to human T-cell lymphotropic viruses. J Virol 61: 4060–4066
36. Martin MA, Bryan T, Rasheed S, Khan AS (1981) Identification and cloning of endogenous retroviral sequences present in human DNA. Proc Natl Acad Sci USA 78: 4892–4896
37. May FEB, Westley BR (1986) Structure of human retroviral sequence related to mouse mammary tumor virus. J Virol 60: 743–749
38. Mondal H, Hofschneider PH (1982) Isolation and characterization of retrovirus-like elements from normal human fetuses. Int J Cancer 30: 281–287
39. Moore R, Dixon M, Smith R, Peters G, Dickson C (1987) Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events are required for translation of gag and pol. J Virol 61: 480–490
40. Nusse R (1986) The activation of cellular oncogenes by retroviral insertion. TIG 2, 244–247
41. O'Brien SJ, Bonner TI, Cohen M, O'Connell C, Nash WG (1983) Mapping of an endogenous retroviral sequence to human chromosome 18. Nature 303: 74–77
42. O'Connell C, O'Brien S, Nash WG, Cohen M (1984) ERV3, a full-length human endogenous provirus: chromosomal localization and evolutionary relationships. Virology 138: 225–235
43. O'Connell CD, Cohen D (1984) The long terminal repeat sequences of a novel human endogenous retrovirus. Science 226: 1204–1206

44. Ono M, Toh H, Miyata T, Awaya T (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
45. 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
46. Ono M, Yasunaga T, Miyata T, Ushikubo H (1986) Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. J Virol 60:589–598
47. Ono M, Kawakami M, Ushikubo H (1987) Stimulation of expression of the human endogenous retrovirus genome by female steroid hormones in human breast cancer cell line T47D. J Virol 61:2059–2062
48. Paulson KE, Deka N, Schmid CW, Misra R, Schindler CW, Rush MG, Kadyk L, Leinwand L (1985) A transposon-like element in human DNA. Nature 316:359–361
49. Query CC, Keene JD (1987) A human autoimmune protein associated with U1 RNA contains a region of homology that is cross-reactive with retroviral $p30^{gag}$ antigen. Cell 51:211–220
50. Rabson AB, Steele PE, Garon CF, Martin MA (1983) mRNA transcripts related to full-length endogenous retroviral DNA in human cells. Nature 306:604–607
51. Rabson AB, Hamagishi Y, Steele P, Tykocinske M, Martin MA (1985) Characterization of human endogenous retroviral envelope RNA transcripts. J Virol 56:176–182
52. Renan MJ, Reeves BR (1987) Chromosomal localization of human endogenous retroviral element ERV1 to 18q22→q23 by in situ hybridization. Cytogenet Cell Genet 44:167–170
53. Repaske R, O'Neill RR, Steele PE, Martin MA (1983) Characterization and partial nucleotide sequence of endogenous type C retrovirus segments in human chromosomal DNA. Proc Natl Acad Sci USA 80:678–682
54. Repaske R, Steele PE, O'Neill RR, Rabson AB, Martin MA (1985) Nucleotide sequence of a full-length human endogenous retroviral segment. J Virol 54:764–772
55. Schmidt J, Luz A, Erfle V (1988) Endogenous murine leukemia viruses: frequency of radiation-activation and novel pathogenic effects of viral isolates. Leukemia Res 12:393–403
56. Shen-Ong GLC, Morse III HC, Potter M, Mushinski JF (1986) Two modes of c-myb activation in virus-induced mouse myeloid tumors. Mol Cell Biol 6:380–392
57. Shinnick TM, Lerner RA, Sutcliffe JG (1981) Nucleotide sequence of Moloney murine leukaemia virus. Nature 293:543–548
58. Silver J, Rabson A, Bryan T, Willey R, Martin MA (1987) Human retroviral sequences on the Y chromosome. Mol Cell Biol 7:1559–1562
59. Snyderman R, Cianciolo GJ (1984) Immunosuppressive activity of the retroviral envelope protein P15E and its possible relationship to neoplasia. Immunol Today 5:240–244
60. Steele PE, Rabson AB, Bryan T, Martin MA (1984) Distinctive termini characterize two families of human endogenous retroviral sequences. Science 225:943–947
61. Steele PE, Martin MA, Rabson AB, Bryan T, O'Brien SJ (1986) Amplification and chromosomal dispersion of human endogenous retroviral sequences. J Virol 59:545–550
62. Sun L, Paulson KE, Schmid CW, Kadyk L, Leinwand L (1984) Non-Alu family interspersed repeats in human DNA and their transcriptional activity. Nucleic Acids Res 12:2669–2690
63. Suni J, Närvänen A, Wahlström T, Aho M, Pakkanen R, Vaheri A, Copeland T, Cohen M, Oroszlan S (1984) Human placental syncytiotrophoblastic M_r 75000 polypeptide defined by antibodies to a synthetic peptide based on a cloned human endogenous retroviral DNA sequence. Proc Natl Acad Sci USA 81:6197–6201
64. Trainor CD, Scott ML, Josephs SF, Fry KE, Reitz MS, Jr (1984) Nucleotide sequence of the large terminal repeat of two different strains of gibbon ape leukemia virus. Virology 137:201–205
65. Wahlström T, Närvänen A, Suni J, Pakkanen R, Lehtonen T, Saksela E, Vaheri A, Copeland T, Cohen M, Oroszlan S (1985) M_r 75000 protein, a tumor marker in renal adenocarcinoma, reacting with antibodies to a synthetic peptide based on

a cloned human endogenous retroviral nucleotide sequence. Int J Cancer 36: 379–382

66. Weiher H, König M, Gruss P (1983) Multiple point mutations affecting the simian virus 40 enhancer. Science 219: 626–631

67. Weiner AM, Deininger PL, Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Ann Rev Biochem 55: 631–661