Journal of General Virology (1997), **78**, 1731–1744. Printed in Great Britain

Structure and genomic organization of a novel human endogenous retrovirus family: HERV-K (HML-6)

Patrik Medstrand,[1,2] Dixie L. Mager,[3,4] Hong Yin,[1,2] Ursula Dietrich[5] and Jonas Blomberg[2]

[1] Department of Medical Microbiology, Section of Virology, Lund University, Sölvegatan 23, S-223 62 Lund, Sweden

[2] Department of Infectious Diseases and Clinical Microbiology, Section of Virology, Uppsala University, Dag Hammarskjölds väg 17, S-752 37 Uppsala, Sweden

[3] Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, BC, Canada V5Z 1L3

[4] Department of Medical Genetics, University of British Colombia, Vancouver, BC, Canada V6T 1Z1

[5] Georg-Speyer-Haus, Paul-Erlich-Str. 42–44, D-60596 Frankfurt, Germany

Prototypic elements of a novel human endogenous retrovirus (HERV) family were identified and cloned from a human genomic library by the use of a *pol* fragment, HML-6, related to type A and type B retroviruses and class II HERVs. Out of 39 *pol*-hybridizing clones, five contained structures of full-length retroviral proviruses, with regions showing similarity to *gag*, *pol* and *env*, flanked by long terminal repeats (LTRs). Restriction mapping and partial sequence analysis of each full-length clone revealed few conserved restriction sites among HML-6 genomes, and about 20% sequence divergence over the reverse transcriptase region sequenced, suggesting that HML-6 constitutes a heterogeneous, but distinct family of elements belonging to the HERV-K superfamily. Sequence analysis of two clones, HML-6p and HML-6.17, revealed a lysine (K) tRNA UUU primer-binding site, and 40–68% nucleotide sequence similarity to LTR, *gag*, *pro*, *pol* and *env* regions of type B retroviruses and class II HERVs. HERV-K (HML-6) elements are present at about 30–40 copies per haploid genome. The HML-6 LTRs contain putative progesterone-responsive elements, which may be involved in the regulation of HML-6 expression. Furthermore, there are about 50 additional solitary HML-6 LTRs per haploid genome. Such LTRs were integrated within the *pol* region of two clones belonging to the same HML-6 family, indicating that some site preference may be involved in HERV integration.

Introduction

The human genome contains a variety of elements with structures similar to mammalian retroviruses. Based on similarity in the polymerase region, two major classes of elements are recognized in humans – class I human endogenous retroviruses (HERVs) and class II HERVs (Callahan, 1988). Class II HERVs include the HERV-K family, which is present at about 50 copies per haploid human genome (Ono *et al.*, 1986). The well-characterized HERV-K10 clone is a 9·2 kb full-length provirus and contains a primer-binding site (PBS) for lysine tRNA, but the *gag* and *env* regions are disrupted by mutations (Ono *et al.*, 1986). However, other HERV-K elements encode non-defective full-length *gag* and *env* (Löwer *et al.*, 1993; Mueller-Lantzsch *et al.*, 1993). HERV-K and related genomes have been detected in retroviral particles, in some cases in association with reverse transcriptase (RT) activity (Boller *et al.*, 1993; Löwer *et al.*, 1987, 1993; Seifarth *et al.*, 1995; Patience *et al.*, 1996). Indeed, there is ample evidence that HERV-K encodes the human teratocarcinoma-derived virus particles (HTDV; for a review see Löwer *et al.*, 1996).

Class II HERVs are also represented by a variety of related sequences in humans, for example, the NMWV1 to 9 sequences (Franklin *et al.*, 1988), and the HML-1 to -6 sequences (Medstrand & Blomberg, 1993). One of the elements, NMWV4, is a 6 kb provirus containing 430 bp LTRs (May &

Author for correspondence: Patrik Medstrand. Correspondence to be sent to the Lund University address. Fax +46 46 189117.
e-mail Patrik.Medstrand@mmb.lu.se

GenBank accession numbers of the sequences reported in this paper are: U60268 (HML-6.17 5′ LTR and downstream sequence), U60269 (HML-6.17 *pol–env*–3′ LTR), U60270 (HML-6.17 solitary LTR), U60271 (HML-6.15 RT), U60272 (HML-6.17 RT), U60273 (HML-6.26 RT), U60274 (HML-6.29 RT) and U86698 (HML-6p *gag–pro–pol*).

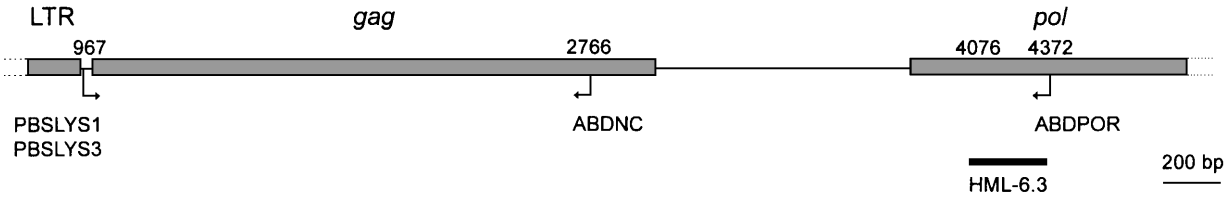

Fig. 1. Localization of PCR primers (shown by arrows) used to characterize HML-6 elements. Part of the LTR–*gag*–*pol* region of a putative class II HERV genome is shown. The starting positions of the primer sequences corresponding to HERV-K10 (Ono *et al*., 1986) are shown above each primer. The black bar represents the position of the HML-6.3 *pol* fragment, corresponding to position 4076–4372 of HERV-K10. Primer sequences were as follows: PBSLYS1, ATTTGGCGCCCAACGTGGGGC; PBSLYS3, ATTTGGCGCCCGAACAGGGAC; ABDNC, TG(C/T)TITAATTGTGGTCA, where I indicates an inosine; ABDPOR, CA(AG)TGGAAAGTTTTACCACAAGGAATG (Medstrand *et al*., 1992). All primers are shown in sense, but a complement of ABDNC and ABDPOR was used.

Westley, 1986). The heterogeneity of this class of HERVs has been illustrated by the differences in the restriction maps of the NMWV1 to 9 elements (Franklin *et al*., 1988), and of the NMWV4 LTRs, compared to those of the HERV-K family (May & Westley, 1986; Ono *et al*., 1986). A novel HERV family was recently discovered to be responsible for the size variation of human complement C4 long and short genes (Dangel *et al*., 1994; Tassabehji *et al*., 1994). This clone, HERV-K (C4), is a 6·4 kb truncated provirus, and represents a distinct class II HERV family comprising about 10–50 copies per haploid genome.

Many HERVs are transcriptionally active (Wilkinson *et al*., 1994), and biosynthesis of retroviral gene products might interfere with cellular functions in *trans*. In addition, HERVs contain *cis*-acting signals within their LTRs, which may influence the transcription of cellular genes. The involvement of HERVs in such regulation is supported by experimental data (for review see Wilkinson *et al*., 1994), for example, in the tissue-specific expression of the human salivary amylase gene (Ting *et al*., 1992). Evidence of particle formation involving HERVs other than HTDV/HERV-K comes from the human mammary carcinoma cell line, T47D (Seifarth *et al*., 1995; Patience *et al*., 1996). These particles contain RNA of different HERV classes. However, the nature and origin of these particles are still obscure. In addition, HERVs or other retroelements such as L1 sequences may be involved in the ongoing dispersion of *Alu* sequences (Wallace *et al*., 1991), and in the formation of processed pseudogenes (Wilkinson *et al*., 1994), by providing RTs needed for the propagation of these elements.

In a previous report, retroviral RT-encoding sequences were identified (Medstrand & Blomberg, 1993). Here, we further characterize one group of these retrovirus-like sequences, HML-6. HML-6 sequences have been described as 244 bp *pol* sequences, displaying 55–60% nucleotide sequence identity to HERV-K10. The present study was undertaken to gain insight into the structure, phylogeny and possible biological significance of these sequences. In this study, we show that the HML-6 elements display typical provirus structures, and belong to a distinct class II HERV family.

Methods

■ **Southern blot analysis.** All subsequent standard protocols and solutions were as described by Ausubel *et al*. (1987), unless otherwise stated. Genomic DNA was digested to completion with restriction enzymes, and fragments were separated by electrophoresis on 0·6% agarose gels. Southern hybridizations were done on nylon membranes as described previously (Medstrand & Blomberg, 1993). Membranes were washed 2 × 20 min in 2 × SSPE, 0·1% (w/v) SDS at room temperature, and 2 × 40 min in 1·2 × SSPE, 0·1% SDS at 60 °C (for the *pol* probe), and 0·4 × SSPE, 0·1% SDS at 65 °C for the *env* and LTR probes (see below).

■ **Probes and library screening.** The *pol* probe (clone HML-6.3) was a 298 bp fragment located at the 5′ end of the RT domain (Medstrand & Blomberg, 1993) (Fig. 1). Fragments containing HML-6 *gag*, *env* and LTR regions were isolated during this investigation (see Results). The *gag* probe was a 1·8 kb PCR fragment, encompassing a region between the PBS site and the 3′-located nucleocapsid (NC) region of *gag* (Fig. 1), the *env* probe was a 0·8 kb *Apa*I–*Eco*RI fragment containing only the *env* region of the 1·2 kb *pol*–*env* fragment of clone HML-6.17, and the LTR probe was a 0·8 kb *Eco*RI fragment of clone HML-6.17. Two oligonucleotides, LTR1 and LTR2 (Fig. 2), were specified as described in Results.

The human genomic library (Goodchild *et al*., 1995) contains partially digested 15–20 kb *Sau*3A fragments of DNA from a human female in phage vector λ GEM-12 (Promega). The genomic library was lifted using standard procedures (Sambrook *et al*., 1989), screened with the *pol* probe (see below), and washed at medium stringency in 1·2 × SSPE, 0·1% SDS at 60 °C. Hybridizations with the *pol* probe to 25 HML-1 to -6 clones were used to determine the specific washing conditions (Medstrand & Blomberg, 1993). Only HML-6 *pol* clones were detected at this stringency.

Individual clones detected from the library screening were plaque-purified (Sambrook *et al*., 1989), and λ DNA was digested with restriction enzymes, separated on 0·7% agarose gels and transferred to nylon membranes. Southern hybridizations and washings were done as for the genomic blots (see above), but in 0·4 × SSPE, 0·1% SDS at 65 °C for the *gag*, *env* and LTR probes. The two oligonucleotides, LTR1 and LTR2, were washed for 2 × 15 min in 5 × SSPE, 0·1% SDS at room temperature, and in an identical solution for 2 × 15 min at 55 °C. All fragments were labelled with [α-^{32}P]dCTP using a random priming protocol (Sambrook *et al*., 1989), and oligonucleotides were tailed with [α-^{32}P]dCTP using terminal transferase (Eschenfeldt *et al*., 1987).

■ **PCR.** PCR reactions were performed as described by Medstrand & Blomberg (1993). Each reaction contained 1 U *Taq* DNA polymerase (Perkin Elmer Cetus) and 100 ng of each primer, and was carried out with

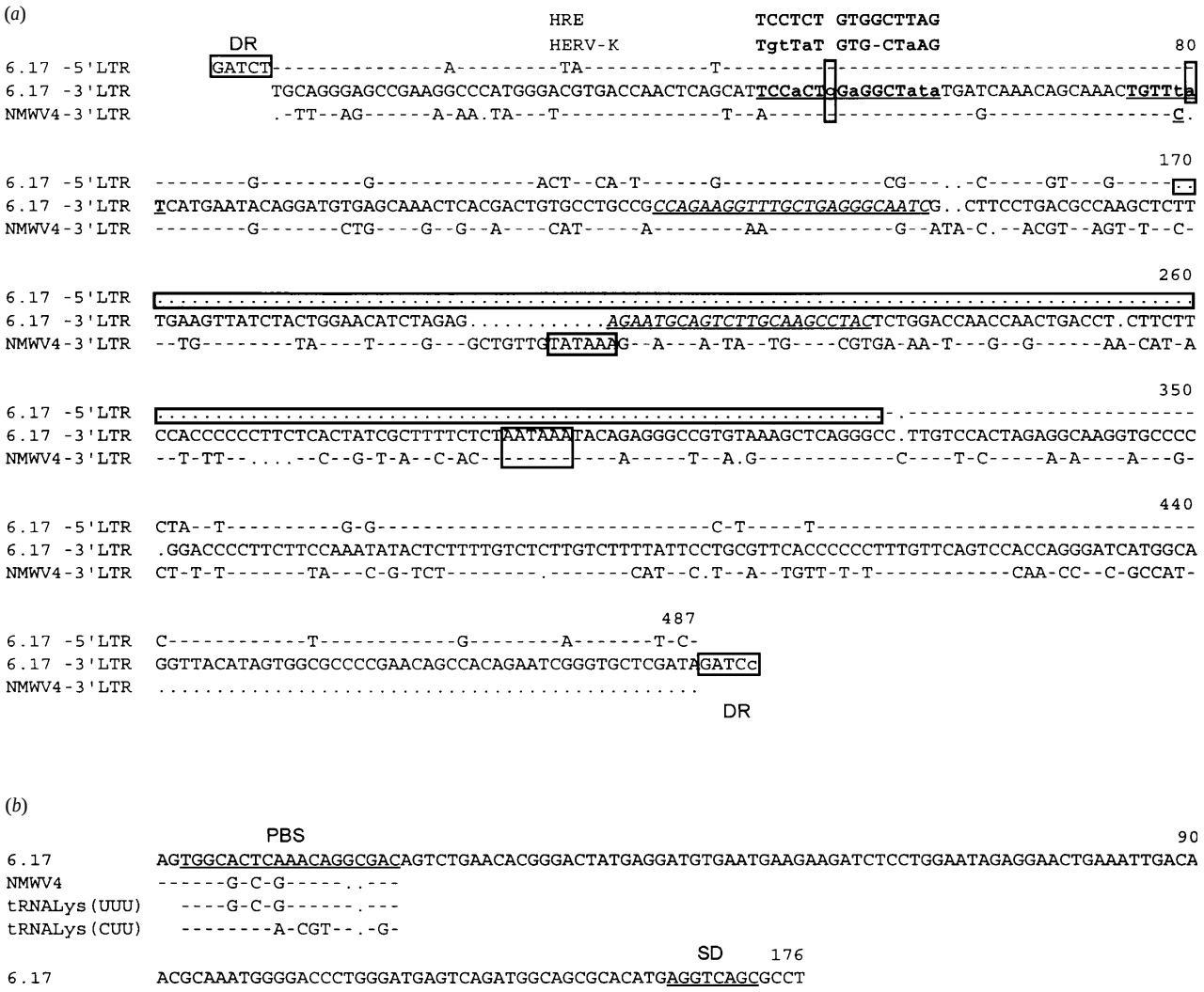

Fig. 2. (*a*) Nucleotide sequences of the 5′ and 3′ LTRs of clone HML-6.17 compared with the NMWV4 3′ LTR (May & Westley, 1986). Dashes indicate residues identical to HML-6.17 3′ LTR, and gaps are shown by dots. Cellular direct repeats (DR) are indicated and mismatches between the 5′ and 3′ DR are indicated in lower-case in boxes. The positions corresponding to the progesterone-like-responsive elements are underlined and in bold type. Above the first motif are the sequences of a hormone-responsive element (HRE) (Von der Ahe *et al*., 1985) and HERV-K (Ono *et al*., 1986) as indicated. A box in the HML-6 sequence indicates an insertion, and lower-case letters indicate mismatches to the sequence of HRE. The second motif corresponds to a partial GRE of MMTV, TGTT(AC)T (Scheidereit & Beato, 1984). The TATA box (TATAAA) and polyadenylation signal (AATAAA) are boxed. The major deletion of the HML-6.17 5′ LTR is indicated by framed dots. Numbers to the right refer to positions in the consensus alignment. The locations of the LTR oligonucleotides used are underlined and in italics. Sequences of the oligonucleotides used were: for LTR1, CCAGA(A/T)GG(A/T)TTGCTGAGGGCAATT; and for LTR2, AG(A/G)A(A/T)-GCA(A/G)TCATGCGAGCCTGCAATT. Both are shown in sense, but a complement of each was used. (*b*) Nucleotide sequence immediately downstream of the HML-6.17 5′ LTR. The PBS and a sequence corresponding to a splice donor sequence (SD) [AGGT(AG)AGT; Shapiro & Senapathy, 1987] are underlined. The PBS of HML-6.17 is compared to the downstream sequence of the 5′ LTR of NMWV4, and the last 18 complementary nucleotides of two rat lysine tRNAs with UUU or CUU anticodons (Sprinzl *et al*., 1985). Identities to HML-6.17 are shown by dashes, and gaps are shown by dots.

the following parameters on a DNA Thermal Cycler (Perkin Elmer Cetus): 5 min at 95 °C, followed by 35 cycles of 30 s at 95 °C, annealing for 30 s at 46 °C, elongation for 2 min at 72 °C and a final extension of 5 min at 72 °C, for primers PBLYS3/ABDNC. For amplification of the 3·4 kb *gag–pol* fragment (see text), a protocol as described by Kidd *et al*. (1990) was used. In this case, PCR was performed with 200 ng of each primer in combinations of PBLYS1/ABDPOR or PBLYS3/ABDPOR, 2 U of *Taq* DNA polymerase and 1 μg human DNA with the following parameters: 5 min at 95 °C, followed by 35 cycles of 1 min at 95 °C, annealing for 1 min at 60 °C, elongation for 5 min at 72 °C and a final extension of 10 min at 72 °C. PCR products were analysed on 1–1·5% agarose gels.

■ **Sequence analysis.** Restriction fragments were subcloned into plasmid vectors, and sequencing was done on either overlapping ExoIII-deleted subclones, generated according to the protocol of the Erase-a-Base kit (Promega), or on subcloned restriction fragments of 300 bases or fewer, by using the dideoxy termination method with Sequenase version 2.0 as per the manufacturer's protocol (USB). All sequences were

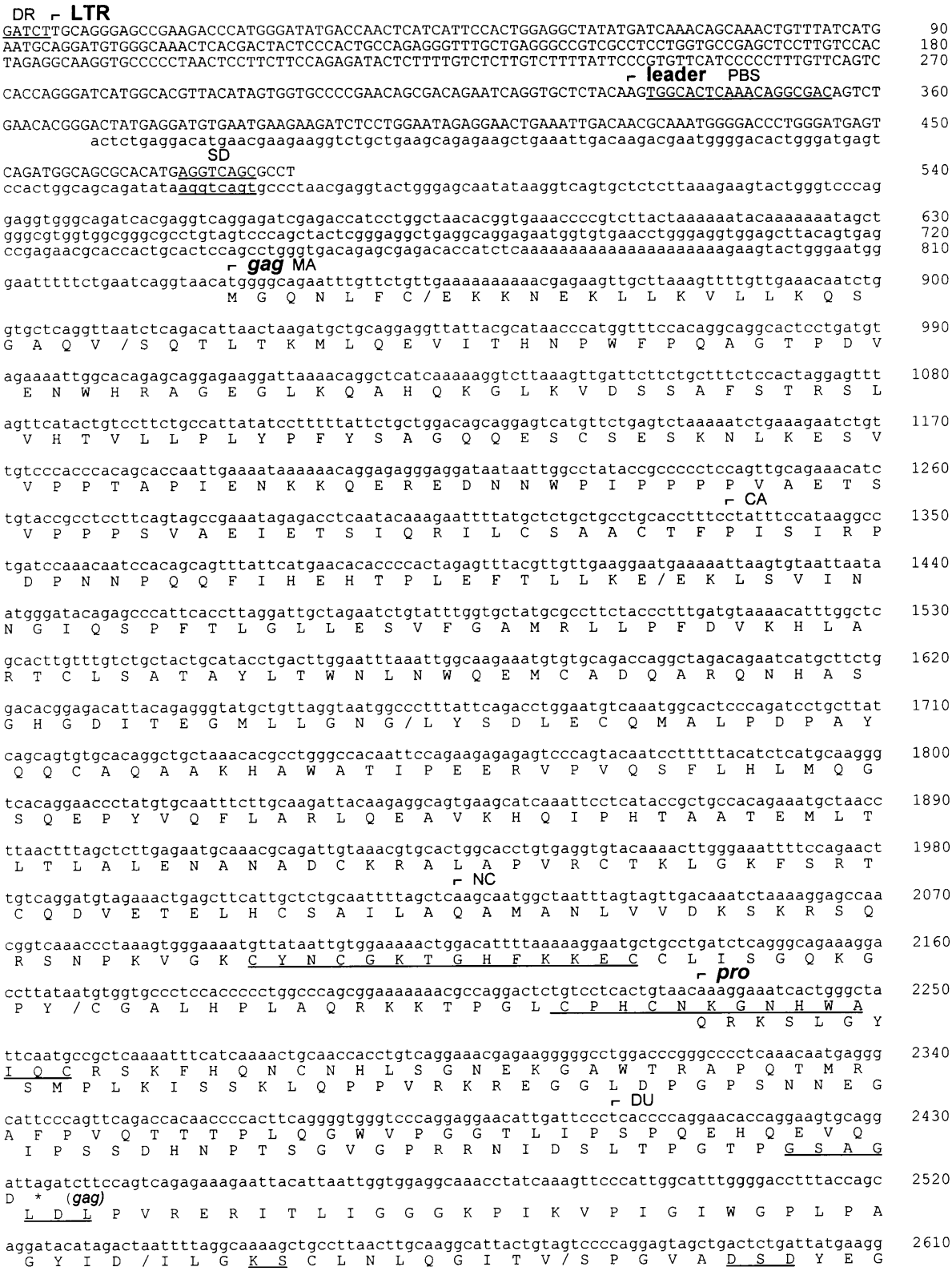

Fig. 3. For legend see p. 1736

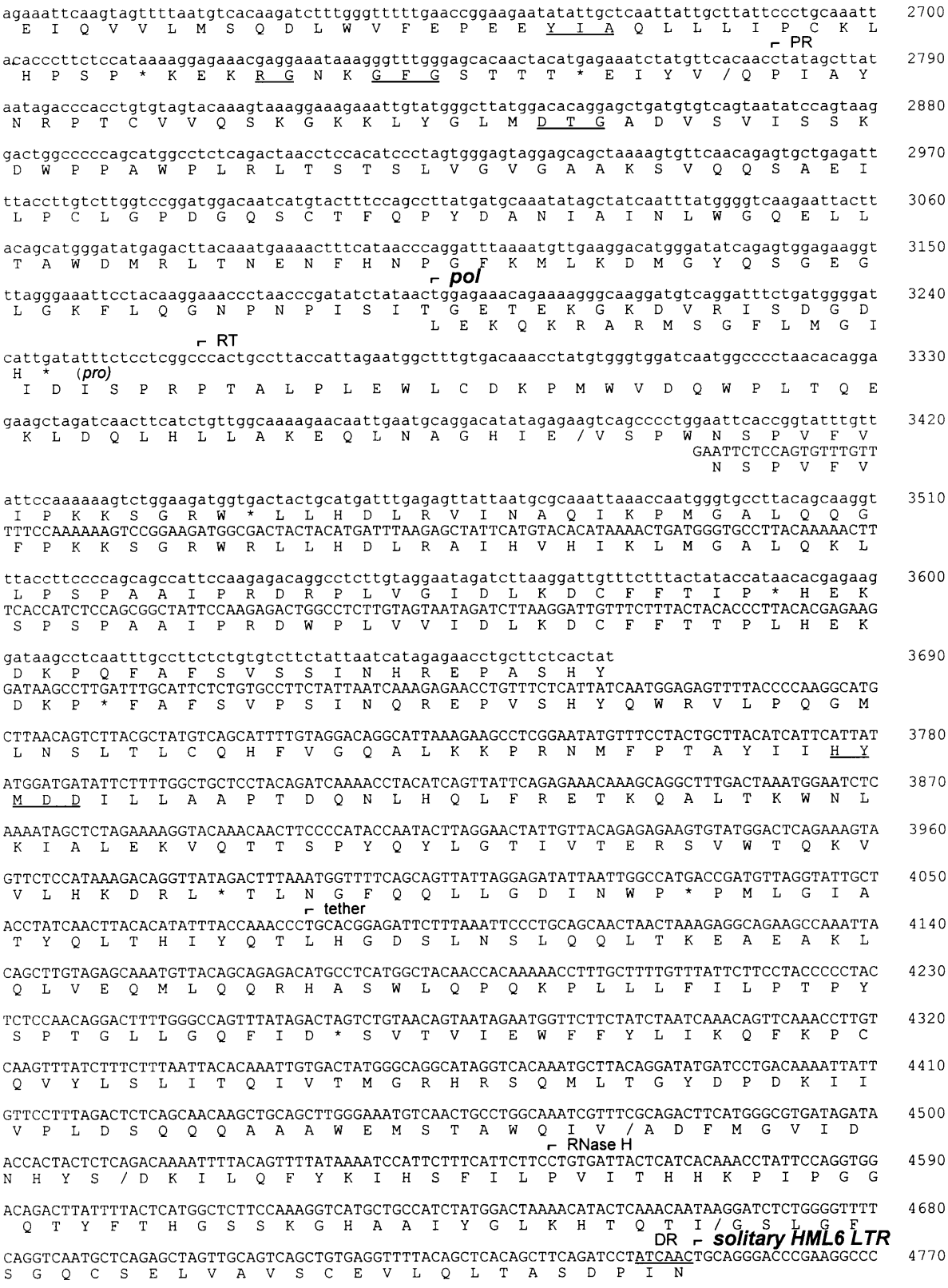

Fig. 3. For legend see p. 1736

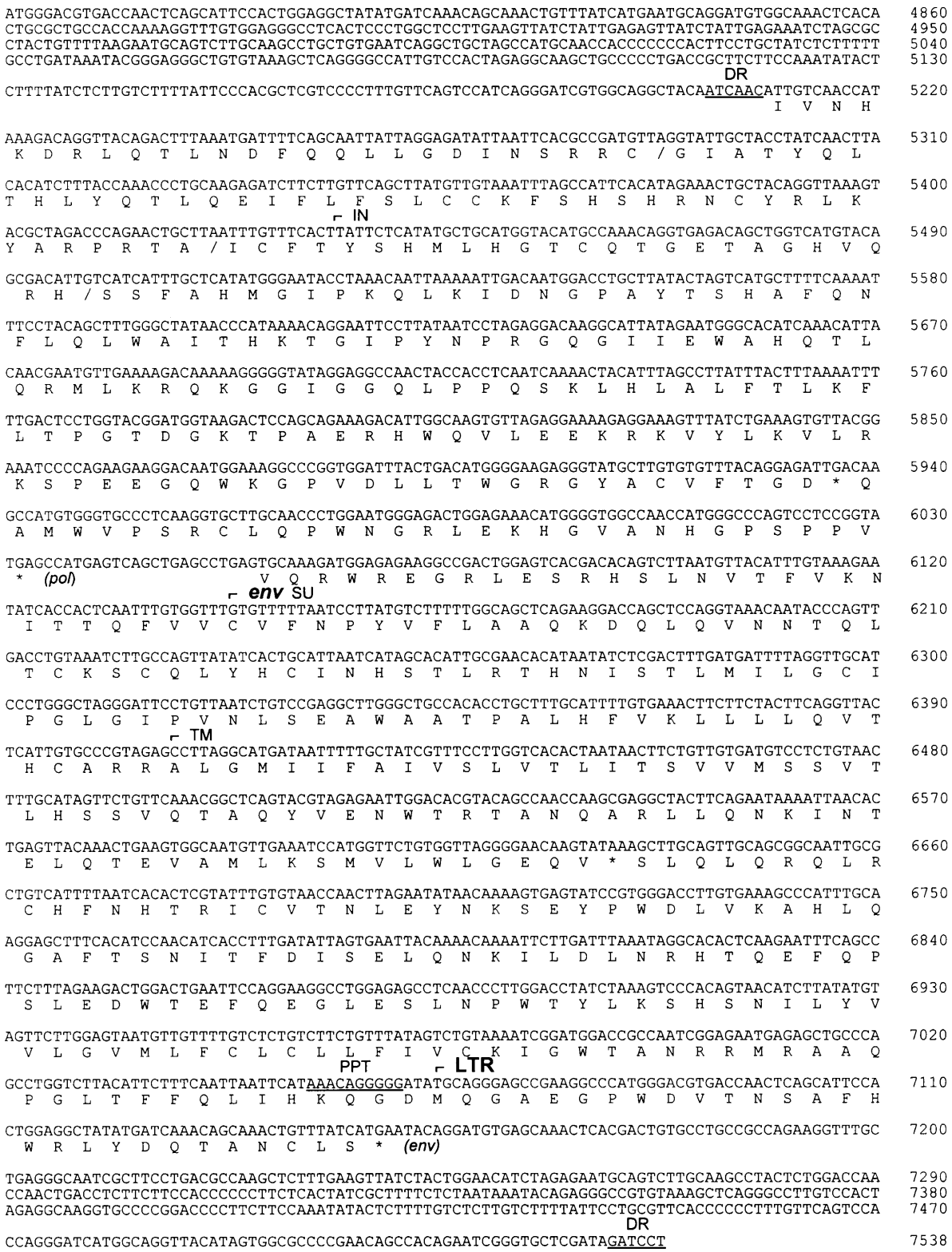

Fig. 3. Nucleotide sequence of a composite HERV-K (HML-6) genome. Amino acid sequences were translated from the nucleotide sequence and correspond to sequences encoded by the major retroviral reading frames (Gag, Pro, Pol and Env), where the start of the individual domains was decided from sequence alignments. Insertions/deletions causing a frameshift are indicated with a solidus and stops are shown with asterisks. The nucleotide sequence derived from clone HML-6.17 is shown in upper-case and encompasses positions 1–480 (5′ LTR and the leader), and positions 3400–7538 (*pol, env* and 3′ LTR). The nucleotide sequence derived from clone HML-6p is shown in lower-case (positions 369–3663; the leader region, *gag, pro*

determined on both strands, and analysed by the PC-GENE software (Intelligenetics). The BLAST programs (Altschul *et al.*, 1990) were used for searching the GenBank, EMBL and EST (Expressed Sequence Tags) databases.

Results

Characterization of HML-6 elements

HML-6 elements have previously only been described for the *pol* region (Medstrand & Blomberg, 1993). To facilitate the screening and characterization of genomic clones, we wanted to isolate a putative HML-6 *gag* region. A PCR-based screening of human DNA resulted in expected 3·4 kb products when amplifications were performed with a *pol* primer, ABDPOR, in combination with primers complementary to one of two versions of lysine (CUU or UUU) PBSs, PBSLYS1 or PBSLYS3, respectively (Fig. 1). Cloning of these products, and subsequent hybridization using the HML-6 *pol* probe, revealed that HML-6 elements had a PBS complementary to lysine UUU. An expected 1·8 kb *gag* fragment was subsequently isolated by using PCR with primer PBSLYS3 in combination with primer ABDNC (Fig. 1). Sequence analysis of one 3·4 kb *gag–pol* clone, HML-6p, revealed a close relationship to previously described HML-6 sequences in the *pol* region (see below and Figs 3 and 7).

Thus, HML-6 elements have the same PBS as the previously isolated NMWV4 HERV clone (May & Westley, 1986), from which the LTR sequences have been determined. Sequences with similarity to the LTRs of NMWV4 were found in five EST sequences (accession numbers R27971, R42279, R65589, T11275 and T99775) and two GenBank entries (CHPMHCDRD and MACMHCCDRB) reported by Mayer *et al.* (1993). Sequence alignment revealed 60–80% identity to the NMWV4 LTRs. Sequences for two oligonucleotides were chosen from conserved regions, and were used in parallel with the HML-6 *gag* probe to characterize HML-6 *pol*-hybridizing genomic clones.

Isolation of HML-6 genomic clones

The *pol* sequence HML-6.3 was used as a probe (Fig. 1) to screen 8×10^5 plaques under specific hybridization conditions. A total of 39 positive clones was identified. Digestion of plaque-purified DNA from each clone with *Eco*RI indicated that 25 had unique restriction patterns. The screening results were compatible with the results obtained in a genomic Southern hybridization using the same probe, when about 25–30 discrete bands at different hybridizing intensities in the range of 0·6–18 kb were recognized (data not shown).

In an attempt to identify full-length HML-6 clones, oligonucleotides LTR1 and LTR2 were hybridized to the 25 immobilized HML-6 clones. One clone with two potential LTR regions was identified by probe LTR1, which recognized the 0·8 and 4·1 kb *Eco*RI fragments of clone HML-6.17. To elucidate whether this clone represented a full-length HML-6 genome, the nucleotide sequences of the LTR-hybridizing fragments were determined.

The sequences of the 3′ and 5′ LTRs, derived from the 0·8 and 4·1 kb *Eco*RI fragments of clone HML-6.17, respectively, are shown in Fig. 2(*a*) in alignment with the NMWV4 3′ LTR. The 3′ LTR is 468 bp long and the 5′ LTR is 329 bp long. A deletion of 156 bp removed a central part of the 5′ LTR, as compared to the 3′ LTR and the NMWV4 LTR. The two HML-6.17 LTRs contained an additional terminal region of 47 bp not present in the NMWV4 LTRs. There is 90% identity between the two HML-6.17 LTRs (on either side of the internal deletion), but only 68% identity between the 3′ LTRs of clone HML-6.17 and NMWV4.

Motifs corresponding to a TATA box and a polyadenylation signal are present on NMWV4 LTRs (May & Westley, 1986). In the 5′ LTR of clone HML-6.17, the 156 bp deletion spans both regions, and the TATA box was also deleted in the HML-6.17 3′ LTR. Parts of a glucocorticoid-responsive element (GRE) (Scheidereit & Beato, 1984), followed directly by a sequence that is similar to an enhancer core, are present on HERV-K LTRs (Ono *et al.*, 1986). The juxtaposition of these two elements creates a sequence with similarity to the binding site of a progesterone-receptor complex (Von der Ahe *et al.*, 1985), which may account for the steroid-inducible HERV-K transcription seen in T47D cells (Ono *et al.*, 1987). A similar motif was present on the HML-6 LTRs (Fig. 2*a*). A second, partial GRE is located on the NMWV4 LTR (indicated in Fig. 2*a*).

A PBS sequence motif sharing identity with 15 out of 18 bp complementary to a lysine tRNA with a UUU anticodon [the same as for mouse mammary tumour virus (MMTV) and NMWV4] was present immediately downstream from the 5′ LTR (Fig. 2*b*), whereas the same region was less complementary (13 out of 18 bp) to the lysine tRNA with a CUU anticodon found in HERV-K clones and HERV-K (C4), or to any other known tRNA (Sprinzl *et al.*, 1985). At position 135 downstream of the 5′ LTR, a sequence motif reminiscent of the splice donor consensus was found (Fig. 2*b*).

Identification of full-length clones

To identify other potential full-length HML-6 clones, the 0·8 kb LTR fragment of clone HML-6.17 was hybridized

and *pol*). Overlapping HML-6.17/HML-6p sequences are aligned (positions 369–480 and positions 3400–3663). Underlined amino acid sequence motifs represent the DNA-binding zinc-finger motifs in NC; the residues of the active site in DU; the active site of the protease in PR; and the residues of the presumed active site of the polymerase in RT. DR, direct repeat; PPT, polypurine tract; MA, matrix; CA, capsid; NC, nucleocapsid; DU, dUTPase; PR, protease; RT, reverse transcriptase; IN, integrase; SU, surface; TM, transmembrane.

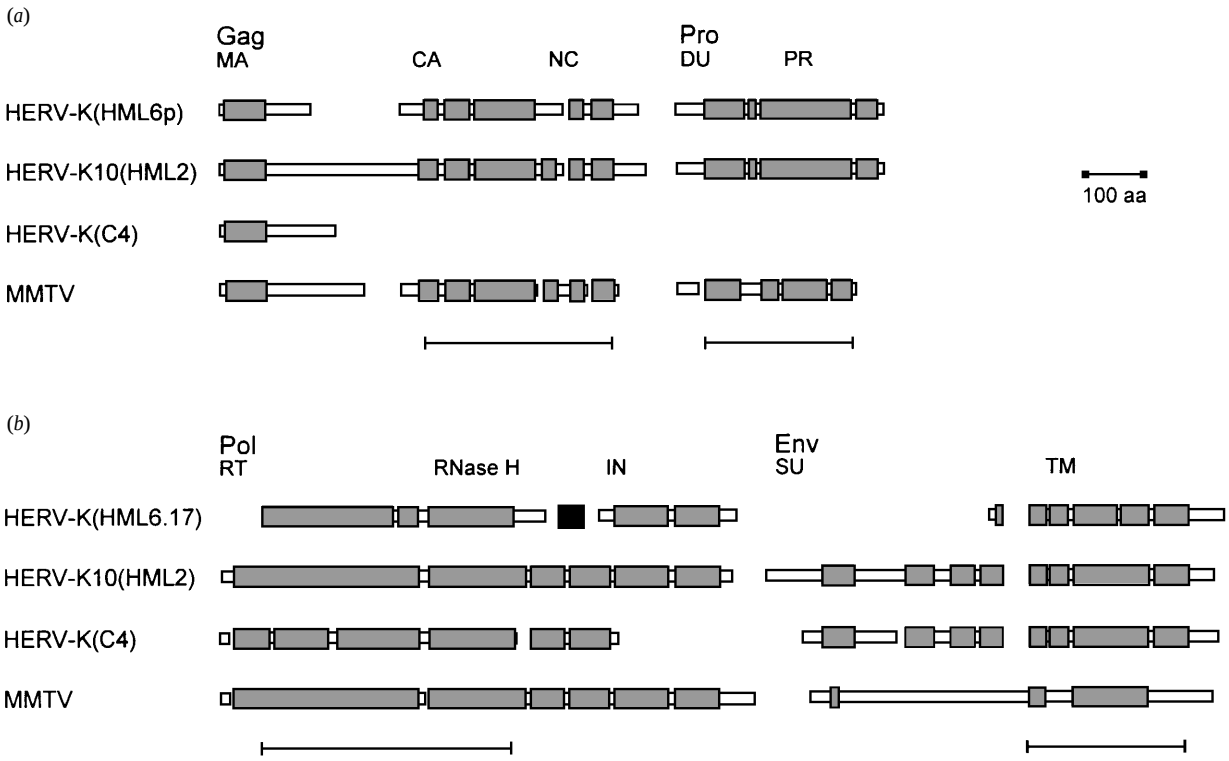

Fig. 4. Schematic representation of amino acid sequence similarity in (*a*) Gag and Pro, and (*b*) Pol and Env between HERV-K (HML-6p)/HERV-K (HML-6.17), HERV-K10 (Ono *et al*., 1986), HERV-K (C4) (Dangel *et al*., 1994) and MMTV (Moore *et al*., 1987). Shaded boxes represent stretches of similar amino acids between sequences, whereas open bars indicate regions of low similarity. Alignments and statistical scorings were done with the MACAW software (National Center for Biotechnology Information, Bethseda, Md., USA), implementing the algorithm of Karlin & Altschul (1990). The sections used for calculating amino acid sequence similarity (Table 1) between sequences are indicated below each alignment. A solid box in HML-6.17 indicates the position of the HML-6 solitary LTR. Sequence alignments are available by request from the author. For a list of abbreviations see Fig. 3.

against the 25 unique HML-6 clones. Five additional clones with two hybridizing *Eco*RI fragments were identified (clones HML-6.4, -6.15, -6.26, -6.29 and -6.36). The LTR probe recognized, besides the expected 0·8 and 4·1 kb fragments of clone HML-6.17, an additional *Eco*RI fragment of 2·3 kb, the same size as detected by the *pol* probe. A restriction map of clone HML-6.17 (see Fig. 5) revealed a 1·2 kb *Eco*RI fragment situated between the 2·3 kb *pol* LTR and 0·8 kb LTR-hybridizing fragments. Sequence analysis of the 2·3 kb subclone showed that an LTR was present within the *pol*-encoding region of clone HML-6.17 (see below). This LTR had 78% similarity to the 3′ LTR of clone HML-6.17, and 70% similarity to the NMWV4 3′ LTR, and had flanking direct repeats on either side. Such solitary LTRs have been described for several HERV families (Mager & Goodchild, 1989; Leib-Mösch *et al*., 1993; Dangel *et al*., 1994).

HML-6 sequences and similarity to other retroviruses

A composite HML-6 genome showing the continuous sequences of the 2·3 kb (*pol*), 1·2 kb (*env*) and 0·8 kb (3′ LTR) *Eco*RI fragments of clone HML-6.17, preceded by the 3·4 kb *gag–pol* sequence of clone HML-6p and the 5′ LTR of clone HML-6.17, is shown in Fig. 3. The sequence derived from clone HML-6p overlaps clone HML-6.17 with 111 bp in the leader region and 262 bp in the *pol* region. The overlapping regions display 80 and 87% nucleotide sequence identity, respectively, indicating that they belong to the same element family (see below and Fig. 7). In this prototype genome, the *gag* region starts 589 bp downstream of the 5′ LTR, and is followed by sequences encoding the *pro*, *pol* and *env* regions. Typical retrovirus motifs are present in the coding regions (underlined in Fig. 3): for example, the NC region of *gag* encodes the two DNA-binding zinc-fingers; the *pro* region encodes conserved motifs of the active sites of dUTPase (DU) and protease (PR); and the *pol* region encodes the motif corresponding to the presumed active site of (RT). A schematic representation derived from amino acid sequence alignments of HERV-K (HML-6), HERV-K10, HERV-K (C4) and MMTV shows similar regions in the four retroviral elements (Fig. 4). The genome organization is the same as for clone HERV-K10 and type B and D retroviruses, and is probably of the same origin due to the presence of DU at the same position. HERV-K (HML-6p) shows Gag and Pro domains similar to those in the other elements, but has a short matrix protein (MA) in comparison to

Table 1. Similarity of HERV-K (HML-6) to other retroviruses

The table shows the percentage amino acid sequence identity of retroviral sequences to HERV-K (HML-6p) in Gag and Pro, and to HERV-K (HML-6.17) in Pol and Env. The highest score for each domain is shown in bold type. Identities were calculated over the positions indicated in Fig. 4.

	HERV-K10	HERV-K (C4)	MMTV*	IAPm†	MPMV‡
Gag	**33·9**	–	29·8	29·9	32·5
Pro	**39·8**	–	36·9	29·2	35·7
Pol	**49·1**	41·8	42·6	40·3	44·8
Env	**30·1**	22·7	28·3	–	11·4

* MMTV, mouse mammary tumour virus (Moore *et al.*, 1987).
† IAPm, mouse intracisternal A-particle (Mietz *et al.*, 1987).
‡ MPMV, Mason–Pfizer monkey virus (Sonigo *et al.*, 1986).

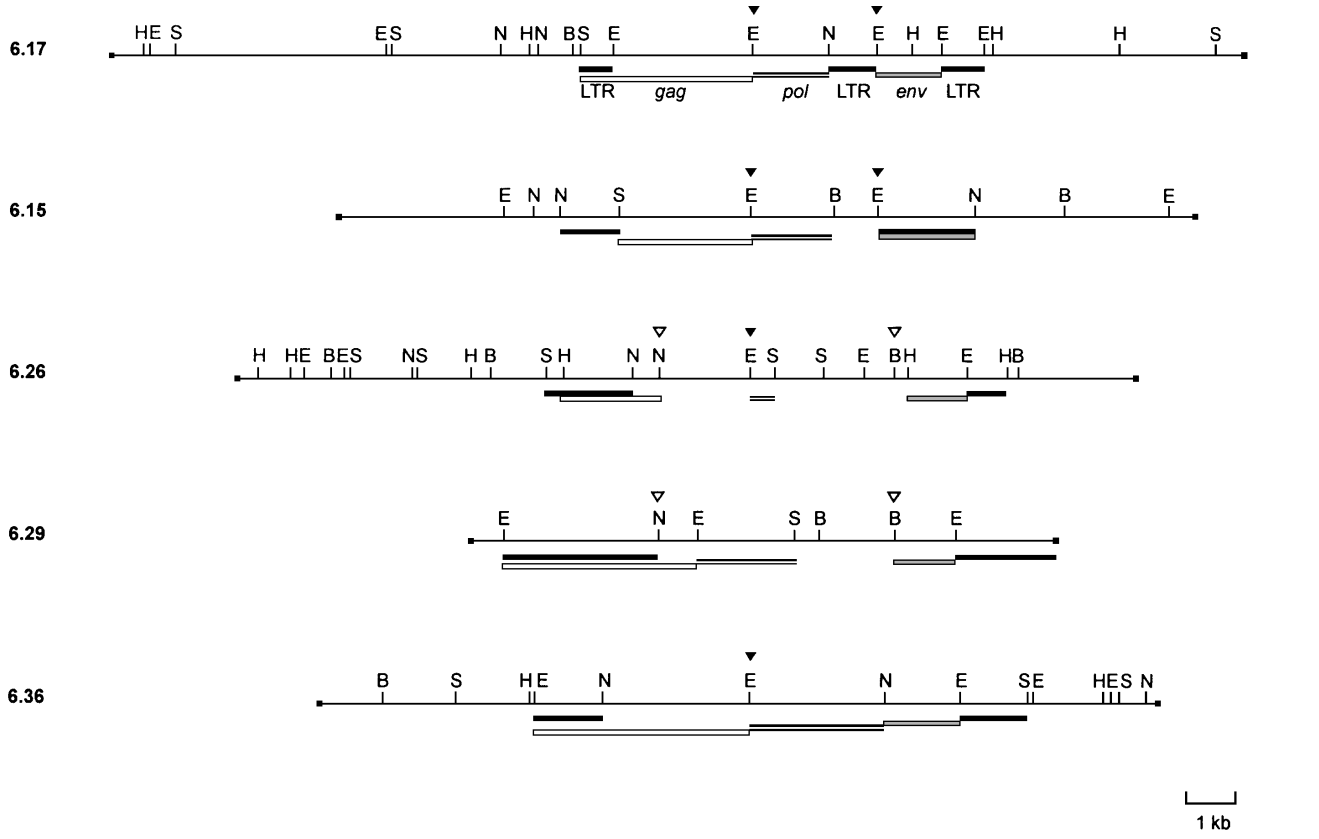

Fig. 5. Restriction maps of HERV-K (HML-6) genomes showing the restriction sites for *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), *Nhe*I (N) and *Sst*I (S). HML clone names are shown to the left of each map. *Hin*dIII restriction sites were not determined for clones HML-6.15 and HML-6.29. Common restriction sites between clones are indicated either with open or solid arrowheads. The orientation of the clones was determined on the basis of hybridizing fragments using the probes shown below the map of clone HML-6.17. The bars below each map correspond to restriction fragments that hybridize with the respective probes. The solid boxes on either side of each restriction map indicate the cloning cassette of the λGEM-12 arms.

HERV-K10 and MMTV. HERV-K (HML-6.17) also displays conserved motifs in Pol, except for a region between the RNase H and integrase (IN) domains, where the HML-6 solitary LTR is located. As is evident in Fig. 4, most of the surface protein (SU) of HERV-K (HML-6.17) is missing, whereas the carboxy-terminal part of SU and the transmembrane (TM) glycoprotein are similar to the other retroviral elements.

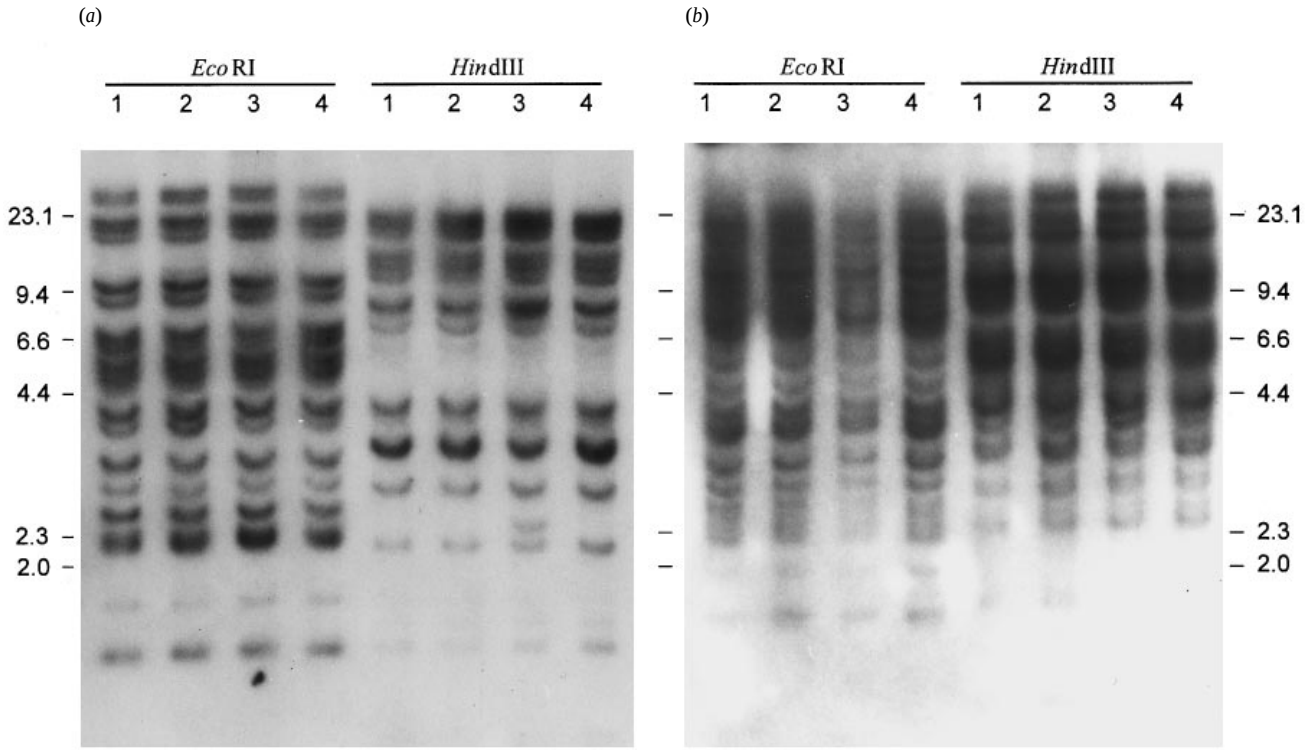

Fig. 6. Southern blot hybridizations of DNA from a human male (lanes 1), DNA from two different females (lanes 2 and 3), and placental DNA (lanes 4) digested with restriction enzymes indicated and hybridized with (*a*) the *env* probe, or (*b*) the LTR probe, and exposed to X-ray film for 36 h and 20 h, respectively. Sizes of co-migrated λ*Hin*dIII restriction fragments (kb) are indicated.

The Gag-, Pro- and Pol-encoding domains show similarity to those of type A, B and D and class II HERVs (29–49% amino acid identity), whereas the Env domain shows similarity to MMTV (type B) and the HERV-K10 and HERV-K (C4) elements (23–30% amino acid identity), but not to Mason–Pfizer monkey virus (MPMV; 11% identity). In all coding regions, the HML-6 elements share the greatest amino acid identities to clone HERV-K10 (Table 1). Despite the presence of retroviral structural protein-encoding regions, all contained deletions, nonsense and frameshift mutations. The TM domain of Env was the least defective, containing only one stop codon.

Genomic organization of HML-6 elements

For a comparison of the HML-6 clones isolated, clone HML-6.17 and five other clones that had two LTR-hybridizing *Eco*RI fragments which were identified by the LTR probe of clone HML-6.17 (HML-6.4, -6.15, -6.26, -6.29 and -6.36) were mapped with restriction enzymes. The restriction map of HML-6.4 showed that this clone, similar to clone HML-6.17, contained a solitary LTR in the *pol* region, flanked on either side by *gag*- and *env*-hybridizing fragments. A 3′ LTR was not present. Despite an LTR in the *pol* region of both HML-6.4 and HML-6.17, the restriction maps of the two differed both in the retroviral and flanking cellular DNA. A short region in *pol* displayed about 10% nucleotide sequence divergence between the two (not shown). These data indicate that the clones represent unique HML-6 proviruses. From the restriction maps of the remaining five clones (Fig. 5), the HML-6 elements evidently make up a rather diverse family of HERVs. Only a few restriction enzyme sites were common among the clones (as indicated in Fig. 5). All full-length elements showed a typical retrovirus structure, with hybridizing fragments in the order *gag*, *pol* and *env*, flanked by LTR-hybridizing fragments. The HML-6.17 5′ LTR made up the last 680 bp of the 4·1 kb *Eco*RI fragment, and the *pol*–*env*–LTR regions encompassed 3674 bp (excluding the solitary LTR). Since the intervening *gag*-hybridizing *Eco*RI fragment is about 2·6 kb, the size of HML-6.17 was estimated to be 6·9 kb.

The mapped clones all contained regions of similarity to *env*. To see whether this was a general feature of the HML-6 elements, Southern hybridization was done with the *env* probe. In addition, because two of the clones described here contained solitary LTRs within the *pol* region, we wanted to see to what degree such LTRs were dispersed in human DNA.

The *env* probe (Fig. 6*a*) resulted in a similar pattern of hybridizing fragments for each set of four enzyme-digested lanes of DNA. In the *Eco*RI digests, approximately 20–25 hybridizing fragments of different intensities were observed, whereas there were about 15 hybridizing *Hin*dIII fragments,

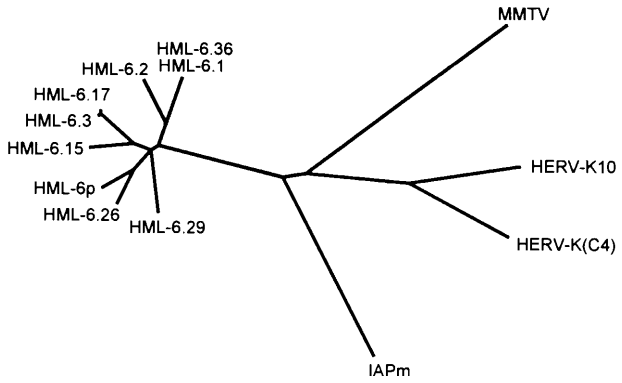

Fig. 7. Unrooted tree derived from alignment of the nucleotide sequences of each full-length HERV-K (HML-6) element. The published sequences of HERV-K10 (Ono *et al*., 1986; positions 4102–4345), HERV-K (C4) (Dangel *et al*., 1994; positions 1459–1679), MMTV (Moore *et al*., 1987; positions 4274–4517), IAPm (Mietz *et al*., 1987; positions 3450–3493), HML-6.1, HML-6.2 and HML-6.3 (Medstrand & Blomberg, 1993) were included for comparison. The tree was constructed by using the programs DNADIST, KITSCH and DRAWTREE of the PHYLIP program package. Sequence alignments are available by request from the author.

but these were slightly more intense. The same blot was hybridized with the *pol* probe. The number of bands was about the same as estimated for the *env* hybridization, and it is therefore likely that most HML-6 elements contain an *env* gene. The only variation seen, apart from the general hybridizing pattern with the *env* probe, was an additional hybridizing *Hin*dIII fragment of about 2·5 kb in DNA from one female (Fig. 6*a*).

The LTR probe was expected to cause a more intense hybridization on the blot due to the presence of two LTRs per HML-6 element. Hybridization with the LTR probe resulted in a more intense hybridization than with the *pol* and *env* probes (Fig. 6*b*), but due to the many hybridizing fragments it was not possible to determine the LTR copy number. To estimate this, we screened about $2{\cdot}5 \times 10^5$ plaques (2×10^5 plaques equal about one human genome) of the human genomic library with the LTR probe, which after medium-stringency washing conditions resulted in 140 LTR-hybridizing clones. Southern blot analyses with *pol* and *env* probes indicated that humans harbour about 30–40 *pol*–*env*-hybridizing HML-6 elements. Based on these results, we estimated that about 50 additional HML-6-related LTRs are present per human haploid genome.

Sequence diversity among HML-6 elements

To estimate the diversity among HML-6 elements, the nucleotide sequence of a region within the RT-encoding domain of each full-length clone was determined. A tree depicting the nucleotide sequence divergence of the HML-6 clones is shown in Fig. 7, together with three previously described HML-6 RT sequences, HML-6.1 to -6.3 (Medstrand & Blomberg, 1993), and HERV-K10, HERV-K (C4), MMTV and mouse intracisternal A-particle (IAPm), to which the HML-6 elements displayed 54–63% nucleotide similarity. Although the HML-6 sequences diverged by as much as 20%, they made up a distinct collection of related sequences compared to the rest. The HML-6.36 RT sequence was identical to that of HML-6.1, and clone HML-6.17 differed by only one nucleotide from the HML-6.3 sequence. This single mismatch could have been generated during PCR by *Taq* polymerase (Gelfand & White, 1990). The remaining clones (HML-6p, -6.15, -6.26 and -6.29) were 84–87% similar to HML-6.17. All HML-6 clones had either stops or additional insertions/deletions disrupting their putative ORFs. In addition, the ORF of clone HML-6.29 was disrupted by insertion of an *Alu* element.

Discussion

We have in this study characterized a novel class II HERV family. All class II HERVs identified so far fall into distinct groups based on sequence similarity (Ono *et al*., 1986; Franklin *et al*., 1988; Medstrand & Blomberg, 1993; Dangel *et al*., 1994; Tassabehji *et al*., 1994). The initial approach used to characterize members of this family was a PCR-based screening method with primers corresponding to conserved retroviral *pol* regions (Medstrand & Blomberg, 1993). Here, initial analysis showed that HML-6 elements contain a PBS complementary to lysine tRNA UUU, and sequence analysis of clone HML-6p revealed retroviral *gag*, *pro* and *pol* regions. The full-length clone HML-6.17 revealed a typical proviral DNA structure, with LTR sequences and regions showing similarity to retroviral *pol* and *env*. Sequences of both these clones are related to type A, B and D mammalian retroviruses, showing highest conservation in the *pol* region (Table 1). However, termination codons and deletions disrupt the reading frames (Fig. 3). Similarity in the *env* region is a good indicator of relatedness and evolutionary history of retroviruses (McClure *et al*., 1988). Based upon this criterion, Wilkinson *et al*. (1994) assigned several type-C-related HERVs to the HERV-ERI superfamily. The presence of a sequence motif complementary to a lysine tRNA, a genome structure, and an *env* region similar to type B retroviruses suggests that HML-6 elements belong to the HERV-K superfamily.

HERV-K (HML-6.17) was estimated to be a 6·9 kb element. The *env* region of this clone was truncated in comparison to MMTV and HERV-K10. Restriction mapping of each full-length clone showed that all elements had a typical retrovirus structure (Fig. 5). This general structure is probably true for most HML-6 elements, due to the presence of approximately the same number of hybridizing *pol* and *env* fragments in the Southern blot analyses. Sequence analysis in the RT region from each clone revealed that members within the HML-6 family constitute a heterogeneous, but distinct family (Fig. 7). The sequence divergence was as much as 20% in RT among HML-6 elements, whereas identity to HERV-K10 and HERV-K (C4) did not exceed 65%. Therefore, differences found between the *pol* and *env* regions of HML-6 elements and those

of other HERVs suggest that HML-6 elements are distinct class II HERVs.

Sequencing of the two fragments containing putative LTRs confirmed the hybridization result. The two HML-6.17 LTRs were 329 and 468 bp long, respectively (Fig. 2*a*) and displayed 90% identity, the same diversity as the two NMWV4 LTRs (May & Westley, 1986). Differences between LTR sequences of endogenous retroviruses are well documented (Wilkinson *et al*., 1994), and the base changes observed probably arose after insertion of the provirus into the human genome. Based on a neutral divergence rate, base changes have been used to estimate the evolutionary age of proviruses (Mager & Freeman, 1995). GenBank searches resulted in the identification of NMWV4/HML-6 LTRs in the genomes of chimpanzee and rhesus macaque (*Macaca mulatta*) (Mayer *et al*., 1993). Mayer *et al*. (1993) also identified a region upstream of the LTR in the chimpanzee clone, *Patr-DRB6* LTR, which showed sequence similarity to MMTV *env*. Sequence comparison between this region in *Patr-DRB6* LTR and HML-6.17 *env* displayed as much as 90% identity. Evidently, *Patr-DRB6* LTR, NMWV4 and HML-6.17 belong to the same family of retroviral elements, despite relatively divergent LTRs. The presence of homologues of HML-6.17 LTRs in Old World monkeys (Mayer *et al*., 1993) and the 10% divergence between each LTR of NMWV4 and HML-6.17 (cf. Mager & Freeman, 1995), indicate that HML-6-like elements entered the primate lineage more than 30 million years ago.

A sequence motif corresponding to the putative GRE of HERV-K elements was identified in these LTRs. NMWV sequences are abundantly expressed in breast cancer cell lines (e.g. T47D) and in the placenta (Franklin *et al*., 1988). Expression of these HERV sequences may be influenced by hormones, as has been shown for HERV-K (Ono *et al*.,1987; Franklin *et al*., 1988). In addition, HML-6 sequences are differentially expressed in liver, lung (Medstrand & Blomberg, 1993) and PBMCs (Andersson *et al*., 1996), suggesting that some transcriptional control directs the observed differential tissue-specific expression (Andersson *et al*., 1996).

Studies on the distribution of retroelements have revealed that host-genome integration sites are not completely random (Sandmeyer *et al*., 1990). A preference for integration into transcribed regions in the genome or in clusters near other HERVs and transposable elements has been observed (Rogers, 1985; Taruscio & Manuelidis, 1991). However, different classes of retroelements show specific regional integration patterns (Taruscio & Manuelidis, 1991), suggesting that some integration site or target-specific preferences exist. An interesting example of non-random integration of retroelements into specific regions involves the primate haptoglobin locus, where three independent HERV-I insertions have occurred during the course of evolution (Maeda & Kim, 1990). In our study, we have found solitary LTRs of the same HML-6 family integrated within the genomes of HML-6.4 and HML-6.17, which suggests that some site preference, as in the haptoglobin region for HERV-I, may be involved in integration. The copy number for different retroelements varies from a few up to thousands per haploid genome (Li & Graur, 1991). The integration target-specificity may account for some of the copy number differences seen for different retroelements. Furthermore, insertion into previously integrated proviruses is probably non-deleterious for the host, and such selective integration may be due to a symbiotic host–virus relationship established over a long period of time.

In the case of the LTRs within the two HML-6 clones, a full-length HML-6 element presumably integrated and was followed by an intrachromosomal recombination event, resulting in a solitary LTR. The presence of host-cell-target direct repeats on either side of the solitary LTR in HML-6.17 supports this hypothesis (Mager & Goodchild, 1989). HERV-K10-related solitary LTRs are present at about 25 000 copies in the human genome (Leib-Mösch *et al*., 1993), whereas solitary LTRs of HERV-K (HML-6) (this study) and HERV-K (C4) (Dangel *et al*., 1994) families are present at much lower copy numbers.

Insertions of retroelements may alter gene function in the host by a variety of mechanisms and may result in disorders due to inactivation or promotion of cellular genes (Favor & Morawetz, 1992). By searching GenBank, we identified a solitary HML-6 LTR present in the 3′ end of a human MHC class I-like cDNA, *MICB*. *MICB* is transcribed at low levels in comparison to the related gene product *MICA*. The down-regulation was suggested to be the result of a disruption of the polyadenylation site of *MICB* (Bahram & Spies, 1996). The solitary HML-6 LTR is integrated in the polyadenylation site of this gene, thus explaining the disruption, and possibly the low abundance of *MICB* mRNA. However, alternative polyadenylation signals are located downstream of the original site. These are used but are apparently not as efficiently as the corresponding site in *MICA*. Other examples suggest that retroelements facilitate novel host-organism adaptations, where expression of cellular genes is controlled and modulated by retroelements (for review see Wilkinson *et al*., 1994). One mechanism is represented by the HML-6-related *Patr-DRB6* LTR (Mayer *et al*., 1993). Despite a deletion causing a frameshift in exon 1, *DRB6* is transcribed at low levels (Corell *et al*., 1991). Mayer *et al*. (1993) showed that transcription was initiated in an LTR located upstream of exon 1. Splicing from the LTR directly in-frame into exon 2 of *DRB6* led to the expression of this gene, where the first amino acids of the presumed polypeptide are encoded by the LTR. Mayer *et al*. (1993) also demonstrated that transcription initiation took place at a TATA region (corresponding to positions 50–60 in Fig. 2*a*) that was different to the presumed TATA box on the NMWV4 LTR (position 205, Fig. 2*a*), which indicates that alternative TATA boxes may direct transcription from deleted LTRs such as HML-6.17.

Several studies have shown that retroelements may act as mutagens resulting in disorders or as driving forces in genetic

variation. Investigations regarding the potential of HERVs in such events are emerging, and further studies on the biology and evolution of these elements will shed light on their involvement in genetical and pathological processes.

We thank Doug Freeman for helpful discussions, Mats Lindeskog for providing the PBS primers used and for helpful discussions, and Per Alm and Jolanta Juraszczyk for technical assistance. This work was in part supported by the European Commission, project GENE-CT930019; funds at the Medical Faculty of Lund; the Royal Physiographic Foundation; the Crafoord Foundation, Lund; the Österlund Foundation, Malmö; by a grant to P.M. from the Medical Research Council, Sweden; and in part by a grant to D.M. from the Medical Research Council of Canada.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.

Andersson, M.-L., Medstrand, P., Yin, H. & Blomberg, J. (1996). Differential expression of human endogenous retroviral sequences similar to mouse mammary tumor virus in normal peripheral blood mononuclear cells. *AIDS Research and Human Retroviruses* **12**, 833–840.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1987). *Current Protocols in Molecular Biology*. New York: John Wiley & Sons.

Bahram, S. & Spies, T. (1996). Nucleotide sequence of a human MHC class I *MICB* cDNA. *Immunogenetics* **43**, 230–233.

Boller, K., König, H., Sauter, M., Mueller-Lantzsch, N., Löwer, R., Löwer, J. & Kurth, R. (1993). Evidence that HERV-K is the endogenous retrovirus sequence that codes for the human teratocarcinoma-derived retrovirus HTDV. *Virology* **196**, 349–353.

Callahan, R. (1988). Two families of human endogenous retroviral genomes. Eukaryotic transposable elements as mutagenic agents. *Banbury Report* **30**, 91–100.

Corell, A., Martin-Villa, J. M., Morales, P., De Juan, M. D., Varela, P., Vicario, J. L., Martinez-Laso, J. & Arnaiz-Villena, A. (1991). Exon-2 nucleotide sequences, polymorphism and haplotype distribution of a new *HLA–DRB* gene: *HLA–DRB* sigma. *Molecular Immunology* **28**, 533–543.

Dangel, A. W., Mendoza, A. R., Baker, B. J., Daniel, C. M., Carroll, M. C., Wu, L.-C. & Yu, C. Y. (1994). The dichotomous size variation of human complement *C4* genes is mediated by a novel family of endogenous retroviruses, which also establishes species–species genomic patterns among Old World primates. *Immunogenetics* **40**, 425–436.

Eschenfeldt, W. H., Puskas, R. S. & Berger, S. L. (1987). Homopolymeric tailing. *Methods in Enzymology* **152**, 337–342.

Favor, J. & Morawetz, C. (1992). Insertional mutations in mammals and mammalian cells. *Mutation Research* **284**, 53–74.

Franklin, G. C., Chretien, S. C., Hanson, I. M., Rochefort, H., May, F. E. B. & Westley, B. R. (1988). Expression of human sequences related to those of mouse mammary tumor virus. *Journal of Virology* **62**, 1203–1210.

Gelfand, D. H. & White, T. J. (1990). Thermostable DNA polymerases. In *PCR Protocols – A Guide to Methods and Applications*, pp. 129–141. Edited by M. A. Innis, D. H. Gelfand, J. J. Sninsky & T. J. White. San Diego: Academic Press.

Goodchild, N. L., Freeman, J. D. & Mager, D. L. (1995). Spliced HERV-H endogenous retroviral sequences in human genomic DNA: evidence for amplification via retrotransposition. *Virology* **206**, 164–173.

Karlin, S. & Altschul, S. F. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proceedings of the National Academy of Sciences, USA* **87**, 2264–2268.

Kidd, A. H., Erasmus, M. J. & Tiemessen, C. T. (1990). Fiber sequence heterogeneity in subgroup F adenoviruses. *Virology* **179**, 139–150.

Leib-Mösch, C., Haltmeier, M., Werner, T., Geigl, E.-M., Brack-Werner, R., Francke, U., Erfle, V. & Hehlmann, R. (1993). Genomic distribution and transcription of solitary HERV-K LTRs. *Genomics* **18**, 261–269.

Li, W.-H. & Graur, D. (1991). Evolution by transposition. In *Fundamentals of Molecular Evolution*, pp. 172–203. Sunderland, Mass.: Sinauer Associates.

Löwer, J., Wondrak, E. M. & Kurth, R. (1987). Genome analysis and reverse transcriptase activity of human teratocarcinoma-derived retroviruses. *Journal of General Virology* **68**, 2807–2815.

Löwer, R., Boller, K., Hasenmaier, B., Korbmacher, C., Müller-Lantzsch, N., Löwer, J. & Kurth, R. (1993). Identification of human endogenous retroviruses with complex mRNA expression and particle formation. *Proceedings of the National Academy of Sciences, USA* **90**, 4480–4484.

Löwer, R., Löwer, J. & Kurth, R. (1996). The virus in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proceedings of the National Academy of Sciences, USA* **93**, 5177–5184.

McClure, M. A., Johnson, M. S. & Doolittle, R. F. (1988). Sequence comparison of retroviral proteins: relative rates of change and general phylogeny. *Proceedings of the National Academy of Sciences, USA* **85**, 2469–2473.

Maeda, N. & Kim, H.-S. (1990). Three independent insertions of retrovirus-like sequences in the haptoglobin gene cluster of primates. *Genomics* **8**, 671–683.

Mager, D. L. & Freeman, J. D. (1995). HERV-H endogenous retrovirus: presence in the New World branch but amplification in the Old World primate lineage. *Virology* **213**, 395–404.

Mager, D. L. & Goodchild, N. (1989). Homologous recombination between the LTRs of a human retrovirus-like element causes a 5 kb deletion in two siblings. *American Journal of Human Genetics* **45**, 848–854.

May, F. E. B. & Westley, B. R. (1986). Structure of a human retroviral sequence related to mouse mammary tumor virus. *Journal of Virology* **60**, 743–749.

Mayer, W. E., O'Huigin, C. & Klein, J. (1993). Resolution of the *HLA–DRB6* puzzle: a case of grafting a *de novo*-generated exon on an existing gene. *Proceedings of the National Academy of Sciences, USA* **90**, 10720–10724.

Medstrand, P. & Blomberg, J. (1993). Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retrovirus: differential transcription in normal human tissues. *Journal of Virology* **67**, 6778–6787.

Medstrand, P., Lindeskog, M. & Blomberg, J. (1992). Expression of human endogenous retroviral sequences in peripheral blood mononuclear cells of healthy individuals. *Journal of General Virology* **73**, 2463–2466.

Mietz, J. A., Grossman, Z., Lueders, K. K. & Kuff, E. L. (1987). Nucleotide sequence of a complete mouse intracisternal A-particle genome: relationship to known aspects of particle assembly and function. *Journal of Virology* **61**, 3020–3029.

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*. *Journal of Virology* **61**, 480–490.

Mueller-Lantzsch, N., Sauter, M., Weiskircher, A., Kramer, K., Best, K., Buck, M. & Grässer, F. (1993). Human endogenous retroviral element K10 (HERV-K10) encodes a full-length Gag homologous 73-kDa protein and a functional protease. *AIDS Research and Human Retroviruses* **9**, 343–350.

Ono, M., Yasunaga, T., Miyata, T. & Ushikubo, H. (1986). Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. *Journal of Virology* **60**, 589–598.

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. *Journal of Virology* **61**, 2059–2062.

Patience, C., Simpson, G. R., Colletta, A. A., Welch, H. M., Weiss, R. A. & Boyd, M. T. (1996). Human endogenous retrovirus expression and reverse transcriptase activity in the T47D mammary carcinoma cell line. *Journal of Virology* **70**, 2654–2657.

Rogers, J. (1985). The origin and evolution of retrotransposons. *International Review of Cytology* **93**, 187–279.

Sandmeyer, S., Hansen, L. & Chalker, D. (1990). Integration specificity of retrotransposons and retroviruses. *Annual Review of Genetics* **24**, 491–518.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning*: *A Laboratory Manual,* 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Scheidereit, C. & Beato, M. (1984). Contacts between hormone receptor and DNA double helix within a glucocorticoid regulatory element of mouse mammary tumor virus. *Proceedings of the National Academy of Sciences, USA* **82**, 3029–3033.

Seifarth, W., Skladny, H., Krieg-Schneider, F., Reichert, A., Hehlmann, R. & Leib-Mösch, C. (1995). Retrovirus-like particles released from human breast cancer cell line T47D display type B- and C-related endogenous retroviral sequences. *Journal of Virology* **69**, 6408–6416.

Shapiro, M. B. & Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Research* **15**, 7155–7174.

Sonigo, P., Barker, C., Hunter, E. & Wain-Hobson, G. (1986). Nucleotide sequence of Mason–Pfizer monkey virus: an immunosuppressive D-type retrovirus. *Cell* **45**, 377–385.

Sprinzl, M., Moll, J., Messner, F. & Hartmann, T. (1985). Compilation of tRNA sequences. *Nucleic Acids Research* **13**, 1–49.

Taruscio, D. & Manuelidis, L. (1991). Integration site preferences of endogenous retroviruses. *Chromosoma* **101**, 141–156.

Tassabehji, M., Strachan, T., Anderson, M., Campbell, R. D., Collier, S. & Lako, M. (1994). Identification of a novel family of human endogenous retroviruses and characterization of one family member, HERV-K (C4), located in the complement *C4* gene cluster. *Nucleic Acids Research* **22**, 5211–5217.

Ting, C.-N., Rosenberg, M. P., Snow, C. M., Samuelson, L. & Meisler, M. H. (1992). Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes & Development* **6**, 1457–1465.

Von der Ahe, D. S., Janich, S., Scheidereit, C., Renkawitz, G., Schutz, G. & Beato, M. (1985). Glucocorticoid and progesterone receptors bind the same site in two hormonally regulated promoters. *Nature* **313**, 706–709.

Wallace, M. R., Anderson, L. B., Saulino, A. M., Gregory, P. E., Glover, T. W. & Collins, F. S. (1991). A *de novo Alu* insertion results in neurofibromatosis type 1. *Nature* **353**, 864–866.

Wilkinson, D. A., Mager, D. L. & Leong, J. C. (1994). Endogenous human retroviruses. In *The Retroviridae,* vol. 3, pp. 465–535. Edited by J. Levy. New York: Plenum Press.

Received 8 November 1996; Accepted 5 February 1997