AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 16, Number 8, 2000, pp. 731–740
Mary Ann Liebert, Inc.

Chromosomal Distribution and Coding Capacity of the Human Endogenous Retrovirus HERV-W Family

CÉCILE VOISSET,[1] OLIVIER BOUTON,[1] FRÉDÉRIC BEDIN,[1] LAURENT DURET,[2] BERNARD MANDRAND,[1] FRANÇOIS MALLET,[1] and GLAUCIA PARANHOS-BACCALÀ[1]

ABSTRACT

Some genomic elements of the multicopy HERV-W endogenous retroviral family have been previously identified in databases. One of them, located on chromosome 7, contains a single complete open reading frame (ORF) putatively encoding an envelope protein. We have experimentally investigated the genomic complexity and coding capacity of the HERV-W family. The human haploid genome contains at least 70, 100, and 30 HERV-W-related *gag, pro,* and *env* regions, respectively, widely and heterogeneously dispersed among chromosomes. Using *in vitro* transcription–translation procedures, three putative HERV-W *gag, pro,* and *env* ORFs were detected on chromosomes 3, 6, and 7, respectively, and their sequences analyzed. A 363 amino acid *gag* ORF containing matrix and carboxy-terminal truncated capsid domains encoded a putative 45-kDa protein. No *gag–pro* ORF was found, but a *pro* sequence containing a DTG active site was detected. Finally, the previously described 538 amino acid HERV-W *env* ORF, located on chromosome 7, was shown to be unique and encoded a putative 80-kDa glycosylated protein. Proteins of molecular mass identical to the one obtained by an *in vitro* transcription–translation procedure were detected in human placenta, using anti HERV-W Gag- and Env-specific antibodies. The absence of an HERV-W replication-competent provirus versus the existence of HERV-W-related Gag and Env proteins in healthy human placenta is discussed with respect to particle formation, physiology, and pathology.

INTRODUCTION

HUMAN ENDOGENOUS RETROVIRUSES (HERVs) are an integral part of the human genome and are consequently inherited as stable Mendelian genes. HERV sequences were integrated in the ancestor germ line millions of years ago,[1] and most contemporary HERV sequences are unable to encode functional proteins, because of acquisition of multiple stop codons and frameshift mutations. However, some HERV proviruses have retained open reading frames,[2–5] and encode proteins[6] that can be occasionally involved in the formation of retroviral particles.[7,8] The HERV-W family has been identified on the basis of the isolation of placental cDNA revealed by multiple sclerosis-associated retrovirus (MSRV) probes[9]. MSRV *pol* sequences were originally detected in autoimmune disease samples, i.e., multiple sclerosis[10] and rheumatoid arthritis.[11] The HERV-W multicopy family, whose active elements entered the germ line of the human ancestor nearly 25 million years ago,[12] is relatively complex. Most of the identified genomes contain large deletions, although one full-length non-replication-competent provirus was identified,[9] which includes a functional U3 sequence and a complete *env* open reading frame (ORF). Together, these observations suggest that some contemporary HERV-W elements may play a physiological role and/or could be associated with some pathological contexts.

The knowledge of the coding capacity of the HERV-W family is a prerequisite to study hypothetical physiopathological roles. As these roles could be driven by isolated proteins or by retroviral particles issued from a single replication-competent unit or assembled via a *trans*-complementation process, we followed a strategy based on the detection and subsequent characterization of predetermined subregions of the retroviral genome on isolated human chromosomes. We have focused on Gag, Gag–Pro, and Env proteins, as Gag and Gag–Pro en-

[1]Unité Mixte 103 CNRS-bioMérieux, Ecole Normale Supérieure de Lyon, 69364 Lyon, Cédex 07, France
[2]Laboratoire de Biométrie Génétique et Biologie des Populations, UMR CNRS 5558, Université Claude Bernard-Lyon 1, 69622 Villeurbanne Cedex, France.

dogenous retroviral proteins could direct virus-like particles budding. Moreover, Env incorporation with complete transmembranes (TM) protein and surface (SU) protein could confer receptor recognition to virus-like particles, and may be involved in membrane fusion.

Here, we describe a widespread and heterogeneous distribution of these three subregions among human chromosomes. Using the polymerase chain reaction (PCR) coupled with *in vitro* transcription–translation, *gag, pro,* and *env* ORFs were detected and further isolated from different chromosomes. Finally, in correlation with HERV-W mRNA expression,[9] HERV-W Gag- and Env-related proteins were detected in healthy human placenta, using HERV-W anti-Gag and anti-Env specific antibodies.

MATERIALS AND METHODS

Source of human isolated chromosomes

The human/rodent monochromosomal NIGMS somatic hybrid mapping panel 2, described by H.L. Drwinga *et al.* and B.L. Dubois *et al.*, was obtained from the Coriell Institute (Camden, NJ).

Probe definition and labeling

The Pgag-Cl2 (1100 bp), Penv-SU-C15 (880 bp), Penv-TM (530 bp), and Ppro-E (364 bp) probes corresponded to PCR fragments of clone Cl2 (AF123881) for *gag*, clone C15 (AF072505) for *env*, and clone E for *pro* (AF009668). After agarose gel electrophoresis separation, gel-isolated PCR amplification products Ppro-E, Pgag-Cl2, Penv-SU-C15, and Penv-TM were extracted and labeled with α-^{32}P by the random-priming method (Ready to Go DNA labeling kit; Pharmacia Biotech, Uppsala, Sweden). Nonincorporated labeled nucleotides were eliminated with a G-50 Quick Spin column (Boehringer GmbH, Mannheim, Germany).

Primer definition and PCR amplification conditions

The primers used for the amplification of *gag, pro,* and *env* HERV-W regions are described in Table 1. The nucleotide sequence of the forward primer presented a T7 RNA polymerase promotor, a spacer, a translation-initiating sequence, and the putative ATG start codon, in frame with the MSRV/HERV-W specific sequence. The sequence upstream to the stop codon of the reverse primer permitted stabilization of the transcribed RNA with a poly(A) tail. Reaction mixtures contained 40 pmol of each primer, a 25 mM concentration of each dNTP (Pharmacia), 2.5 mM $MgCl_2$, 2.5 U of *Taq* polymerase in standard PCR buffer (Perkin-Elmer, Norwalk, CT), and 300 ng of DNA in a final volume of 100 μl. PCR cycling conditions were as follows: 3 min at 94°C; 30 cycles, each consisting of 1 min at 94°C, 1 min at annealing temperature, and elongation at 72°C; followed by 7 min at 72°C. The annealing temperature and the elongation time were 55°C and 3 min, respectively, for gag5′–gag3′ primers, gag5′–pro3′ primers, and env5′–env3′ primers; 55°C and 2 min, respectively, for pro5′–pro3′ primers; and 60°C and 2 min, respectively, for U5gag–gag3′ primers. All the PCR amplifications were performed on an MJ Research (Watertown, MA) PTC-200 apparatus, a Peltier thermal cycler. PCR products were analyzed on 1% Tris–borate–EDTA (TBE) agarose gel, Southern blotted to verify the amplification specificity, and directly used in the *in vitro* transcription–translation assay.[13]

Southern blot analysis of human chromosomal somatic cell hybrid DNA and PCR products

The BIOS monochromosomal somatic cell hybrid blot (Quantum Bioprobe, Montreuil, France), contains 5 μg of *Eco*RI-digested DNA from 24 somatic hybrid cell lines and 3 control DNA (human, mouse, and hamster), transferred by capillary blotting to a positively charged membrane. After 4 hr of prehybridization (in 5× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate], 1× Denhardt's, 0.1% sodium dodecyl sul-

TABLE 1. PRIMER SEQUENCES USED IN PCR AMPLIFICATION AND *in Vitro* TRANSCRIPTION–TRANSLATION ASSAYS

Primer	*Sequence*[a]	*Nucleotide position*[b]	*Accession number*[b]
U5gag	5′-TGT CCG CTG TGC TCC TGA TC-3′	199–218	AF123880
Gag5′	5′-TTT GGT AAT ACG ACT CAC TAT AGG GCA GCC ACC ATG GGA AAC GTT CCC CCC GAG-3′	592–612	AF123880
Gag3′	5′-TTT TTT TTT TTT TTT TTT TCA GGC TGC GCC AGT GTC CAG GAG AC-3′	101–123	AF009666
Pro5′	5′-TTT GGT AAT ACG ACT CAC TAT AGG GCA GCC ACC ATG GGG TGC CTG GGG CAA GCA CCA GCC-3′	18–41	AF009668
Pro3′	5′-TTT TTT TTT TTT TTT TTT TCA AAG AGG GTA CTT TTG GTA GGG AAA GG-3′	476–504	AF009668
Env5′	5′-CAT AAT ACG ACT CAC TAT AGG GAG ACC ATG GCC CTC CCT TAT CAT-3′	760–779	AF072506
Env3′	5′-TTT TTT TTT TTT TCT CTA ACT GCT TCC TGC TGA ATT GG-3′	2356–2380	AF072506

[a]The putative ATG start codon and the stop codon are underlined, on the forward and backward primers, respectively.
[b]The nucleotide position corresponded to the parental clone whose accession number is provided.

fate [SDS], 50% formamide, 20 mM Tris-HCl [pH 7.5] herring sperm DNA [0.1 mg/ml]), nylon membranes were hybridized (in 5× SSC, 1× Denhardt's, 0.1% SDS, 50% formamide, 20 mM Tris-HCl [pH 7.5], herring sperm DNA [0.1 mg/ml] 5% dextran sulfate) overnight at 42°C, with a DNA probe that is ^{32}P labled. Southern blot membranes were washed twice in 2× SSC–0.2% SDS for 15 min at room temperature, twice in 0.2× SSC–0.2% SDS for 15 min at 45°C, and twice in 0.2× SSC–0.2% SDS for 15 min at room temperature. After washing, blots were exposed to X-ray film at −80°C.

For sensitive detection of amplification product, 3 μl of PCR products *gag, pro, gag–pro,* and *env* were separated by agarose gel electrophoresis and transferred onto nylon membranes (Hybond-N+; Amersham Pharmacia, Freiburg, Germany) in 0.4 M NaOH, as described above. PCR products blots were washed in a solution of 2× SSC–0.2% SDS for 15 min at room temperature, twice in 0.2× SSC–0.1% SDS for 15 min at 65°C, twice in 0.1× SSC–0.1% SDS for 15 min at 65°C, and twice in 0.1× SSC–0.1% SDS for 30 min at room temperature.

Identification of ORFs using an in vitro *transcription–translation assay*

The presence of *gag*, *pro* and *env* ORFs was analyzed by an *in vitro* transcription–translation procedure.[13] Four microliters of nonpurified PCR products, visible on ethidium bromide-stained agarose gel, was transcribed and translated for 120 min at 30°C in a TNT-T7-coupled reticulocyte lysate system (Promega, Madison, WI) as recommended by the manufacturer. Newly synthesized proteins were labeled with [^{35}S]methionine (Amersham Life Science, Arlington Heights, IL). After incubation, 1× buffer (50 mM Tris-H_3PO_4 [pH 6.8], 1% SDS, 4% glycerin, bromophenol blue [1 mg/ml] 1% 2-mercaptoethanol) was added. Eight microliters of each sample was separated by SDS-polyacrylamide gel electrophoresis. Gag, Gag–Pro, Pro, and Env *in vitro* transcription–translation products were separated in 15, 12, 15, and 12% SDS-polyacrylamide gels, respectively. Acrylamide gels containing [^{35}S]methionine-labeled proteins were subsequently vacuum dried at 80°C, and finally exposed to x-ray film at room temperature.

Sequencing and sequence analysis

PCR products of interest were cloned into the pCR2.1-TOPO vector (InVitrogen, San Diego, CA), and sequenced from both directions with a Big Dye deoxyterminator cycle sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA), in an Applied Biosystems model 373A or 377 automated DNA sequencer, as recommended by the manufacturer. GenBank (release 113, August 1999) queries were performed with the BLAST[14] family of programs, including tblastn, which compares a protein query sequence against the nucleotide sequence database dynamically translated in all reading frames. Nucleic acid and protein alignments, percentage of similarities, and prediction of protein posttranslational events was done with GeneWorks 2.5.1 software (Oxford Molecular Ltd., Oxford, UK).

Western blot

To generate antibodies, HERV-W *gag* and *env* genes were expressed in *Escherichia coli* and purified by affinity chromatography on metal-chelate columns (Qiagen, Chatsworth, CA). AntiGag F(ab′)$_2$ purified polyclonal antibodies were obtained after immunization of rabbits with a recombinant protein derived from HERV-W ΔMA-ΔCA Gag Cl2 clone consisting of an open reading frame of 359 amino acids corresponding to an N-terminal truncated matrix domain (MA) and a C-terminal truncated capsid domain (CA).[15] A nonimmunized rabbit serum F(ab′)$_2$ purified antibody was used as control. An anti-Env monoclonal antibody was obtained after immunizing mice with recombinant protein derived from the HERV-W ΔSU-ΔTM Env C15 clone consisting of an open reading frame of 379 amino acids corresponding to an N-terminal truncated surface domain and a C-terminal truncated transmembrane domain.[15] Placental protein extract (200 μg; Clontech, Palo Alto, CA) was loaded per well. Proteins were separated on 12% SDS-polyacrylamide gels, and were subsequently transferred to a polyvinylidene difluoride (PVDF) membrane (Millipore, Bedford, MA). Specific indirect immunodetection of HERV-W-related proteins in the human placenta was performed with 1:125,000 diluted anti-Gag polyclonal F(ab′)$_2$ and the anti-Env monoclonal antibody, overnight at 4°C, and several washes were performed in 1× TBS–0.05% Tween 20. Chemiluminescent detection was performed with 1:2000 diluted peroxidase-conjugated secondary antibodies, for 1 hr at room temperature.

RESULTS

Chromosomal distribution of HERV-W gag, pro, *and* env *regions*

The preliminary identification of the presence of each *gag, pro,* and *env* subregion was achieved by Southern blot hybridization of membranes containing human–rodent monochromosomal DNA, using Pgag-Cl2, Ppro-E, and Penv-SU-C15 probes (Fig. 1). No hybridization signal was observed using Pgag-Cl2, Ppro-E, and Penv-SU-C15 probes on mouse and hamster DNA. Southern blot analysis of monochromosomal DNA allowed evaluation of the HERV-W distribution complexity on each human chromosome, and the heterogeneous distribution of HERV-W sequences among chromosomes. The hybridization patterns illustrated in Fig. 1 revealed about 70, 100, and 30 fragments containing HERV-W *gag, pro,* and *env*-SU related sequences, respectively. Nonuniformity in band intensity may reflect differences in the number of repeat units that carry each chromosome, although differences in sequence composition cannot be excluded. Chromosomes 2, 16, and 22 lacked HERV-W *gag, pro,* and *env* sequences. HERV-W-related *gag* and *pro* sequences were found simultaneously on chromosomes 1, 3–15, 17–19, 21, and X and additional *pro* sequences were detected on chromosomes 20 and Y. HERV-W-related *env*-SU sequences were detected on chromosomes 1, 3–7, 9, 11–12, 14–15, 17–18, and X. Hybridization with an *env*-TM probe showed nearly the same distribution except for the absence of signal on chromosome 18 and two additional signals on 19 and 21 (data not shown). All four *gag, pro, env*-SU, and *env*-TM regions were found simultaneously on chromosomes 1, 3–7, 9, 11–12, 14–15, 17, and X, revealing the existence of several putative complete retroviral units, replication competent or not, an addition to the one previously described on chromosome 7.[9]

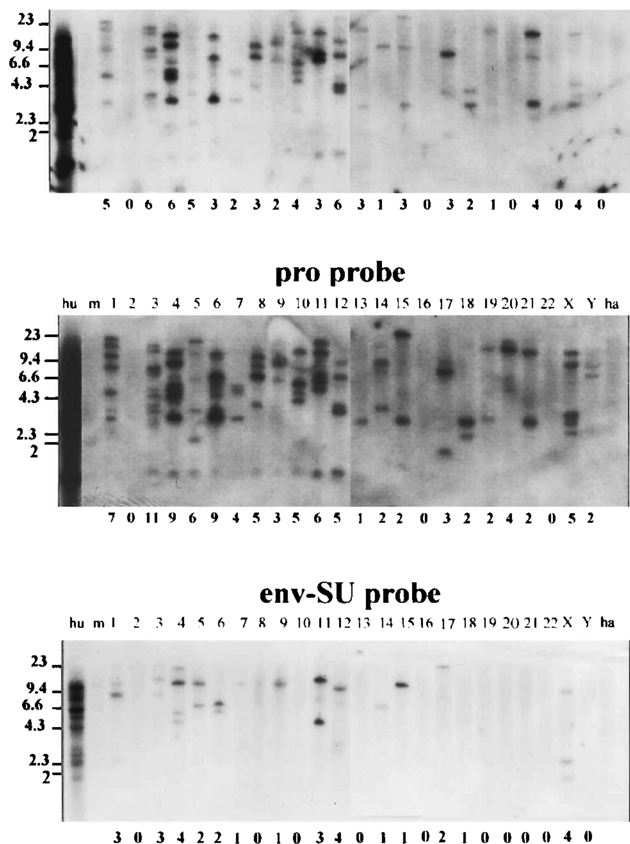

FIG. 1. Chromosomal distribution of HERV-W *gag*, *pro*, and *env*-related sequences by Southern blot analysis. *Eco*RI-digested monochromosomal DNA was hybridized with HERV-W Pgag-Cl2 (**gag probe**), Ppro-E (**pro probe**), and Penv-SU-C15 (**env-SU probe**) probes. Each of the 24 monochromosomal somatic cell hybrids used carried a single human chromosome, except for chromosome 4, which contains some DNA of chromosome 7, and chromosome 20, which contains some DNA of chromosomes 4 and 8. Human (hu), mouse (m), and hamster (ha) total DNA were included as controls. The upper lines indicate the chromosome numbers, and the lower lines indicate the approximate number of bands observed for each chromosomal DNA. Numbers to the left are molecular mass markers (in kilobases).

PCR amplification of HERV-W gag, pro, *and* env *sequences on human chromosomes*

To detect HERV-W *gag, pro,* and *env* open reading frames, DNA samples from isolated chromosomes were amplified by PCR, using HERV-W-specific primers. As translation of the retroviral protease enzyme generally results from a frameshift or a termination suppression event, both phenomena being unlikely in the *in vitro* transcription–translation assay, we chose to analyze the *gag* and *pro* regions independently. Furthermore, it cannot be excluded that some HERV-W sequences with large deletions within the *gag* region may have conserved protease-coding capacity. The gag5′ forward primer ATG initiation codon corresponded to the putative start codon observed in the CL17 clone (AF123880), and the reverse gag3′ primer corresponded to the DTG active site of the protease region, which was repeatedly observed in *gag–pro* regions obtained from a human DNA library (data not shown) regardless of any coding capacity (Table 1). An artificial initiation ATG codon was included for the protease gene PCR amplification within the pro5′ forward primer. The pro3′ reverse primer was complementary to the *pol* region (Table 1). env5′ and env3′ primers corresponded to the borders of the full-length HERV-W *env* region previously identified in the cl.PH74 clone (AF072506, Table 1).

gag, pro, and *env* sequences were PCR amplified and HERV-W specificity was confirmed by Southern blot hybridization (Table 2). Rodent controls from mouse and hamster DNA, as well as the water control, were negative after probe hybridization. All chromosomal DNA presenting no hybridization signal by Southern blot analysis did not yield any PCR product, as expected. Most of the chromosomes presenting a positive hybridization signal using HERV-W *gag, pro,* and *env*-SU probes in Southern blot analysis produced positive PCR amplification signals, although some failed to yield the expected PCR products.

Fourteen of the 19 chromosomal DNAs that were positive by Southern blot analysis with the *gag* probe also gave positive amplification signals using the gag5′–gag3′ primer pair (Table 2). Depending on the chromosome, we observed heterogeneity of PCR product sizes of HERV-W *gag* amplified regions, corresponding to a 1.6- and/or a 1.1-kb fragment. The larger fragment corresponded to the expected size with respect to the sequences of the closest functional retroviruses BaEV (baboon endogenous virus) and GaLV (gibbon ape leukemia virus), which encode a 60-kDa Gag precursor. As five discrepancies were observed between Southern blot analysis and gag5′–gag3′ amplification, the corresponding DNA samples were amplified with gag5′–pro3′. No PCR product was obtained from chromosomes 5, 9, and 18 and, conversely, larger 2-kb fragments were amplified from chromosomes 14 and 19 (Table 2).

Of the 21 chromosomal DNA positive for *pro* probe hybridization, 19 led to positive amplification signals with the pro5′–pro3′ primer pair (Table 2). HERV-W *pro* PCR fragments were 0.5 kb long, compatible with the previously described MSRV protease region.[10] Further attempts to amplify the HERV-W *pro* region with gag5′ as a forward primer were unsuccessful on both chromosomes 9 and 17.

As for the 13 chromosomal DNAs detected by *env* probe hybridization, 9 gave positive PCR signals with the env5′–env3′ primer pair (Table 2). HERV-W *env* PCR fragments were 1.7 kb long as previously described for the human bacteria artificial chromosome (BAC) clone RG083M05 located on chromosome 7.[9] No complete *env* region was PCR amplified from chromosomes 1, 9, 11, and 15, whereas they contained both SU and TM *env* sequences, as shown by Southern blot hybridization. However, complete TM regions were detected on all four chromosomes by specific PCR amplification, using a forward ATG-containing primer and the env3′ reverse primer (data not shown).

PCR products obtained with gag5′–gag3′, pro5′–pro3′, and env5′–env3′ primer pairs were transcribed and translated *in vitro*. When *in vitro*-translated proteins were observed, the corresponding PCR products that exhibited the largest protein size

TABLE 2. CORRELATION OF SOUTHERN BLOT ANALYSIS WITH PCR AMPLIFICATION OF HERV-W *gag, pro,* AND *env* REGIONS ON HUMAN CHROMOSOMES

	gag		*pro*		*env*	
Chromosome number	*Southern blot*[a]	*PCR product (kb)*	*Southern blot*[a]	*PCR product (kb)*	*Southern blot*[a]	*PCR product (kb)*
1	+	1.6	+	0.5	+	—
2	—	—	—	—	—	—
3	+	1.6–1.1	+	0.5	+	1.7
4	+	1.6–1.1	+	0.5	+	1.7
5	+	—; (—[b])	+	0.5	+	1.7
6	+	1.6–1.1	+	0.5	+	1.7[c]
7	+	1.1	+	0.5	+	1.7
8	+	1.1[c]	+	0.5	—	—
9	+	—; (—[b])	+	—	+	—
10	+	1.1	+	0.5	—	—
11	+	1.6–1.1	+	0.5	+	—
12	+	1.6–1.1	+	0.5	+	1.7
13	+	1.1	+	0.5[c]	—	—
14	+	—; (2[b])	+	0.5	+	1.7
15	+	1.6	+	0.5	+	—
16	—	—	—	—	—	—
17	+	1.6[c]	+	—	+	1.7
18	+	—; (—[b])	+	0.5	+	—
19	+	—; (2[b])	+	0.5	—	—
20	—	—	+	0.5	—	—
21	+	1.6	+	0.5	—	—
22	—	—	—	—	—	—
X	+	1.1	+	0.5	+	1.7[c]
Y	—	—	+	0.5	—	—

[a]Southern blot analysis is depicted in Fig. 1.
[b]Additional PCR amplification using gag5′–pro3′ primer pair.
[c]PCR amplification products weakly revealed by Southern blot analysis, and not stained with ethidium bromide.

were cloned and sequenced. Moreover, as HERV-W mRNA expression was previously described to be restricted to placenta,[9] HERV-W protein expression was analyzed in this human tissue.

Coding capacity of the gag *region*

Although the 1.6-kb *gag* fragments on some chromosomes could potentially encode a 60-kDa Gag precursor, none of the chromosomes tested yielded such a large transcription–translation product (Fig. 2A). However, chromosomes 1, 3, 6, and 7 yielded HERV-W Gag proteins ranging form 17 to 45 kDa: chromosome 1 presented a pattern of three proteins of 28, 23, and 18 kDa; chromosome 3 yielded a major protein of approximately 45 kDa, and two minor proteins of 20 and 17 kDa; chromosome 6 had a major protein of 25 kDa, and two minor proteins of 18 and 17 kDa; chromosome 7 had a 22-kDa protein (Fig. 2A). To check whether the observed different bands represented products of mixed different PCR fragments, or resulted from the analytical process itself, *gag* PCR products were cloned and several clones were analyzed by the *in vitro* transcription–translation assay. The same patterns were observed with *gag* PCR products and 1.6-kb cloned PCR products for chromosomes 1, 3, and 6, and 1.1-kb PCR product for chromosome 7, suggesting that the different bands resulted from the analytical process itself (data not shown). No additional Gag protein was observed with this assay when using pro3′ as reverse primer on chromosomes 14 and 19 (data not shown), which contain HERV-W-related *gag* regions (Table 2).

The 1.6-kb PCR product amplified from chromosome 3 DNA, which contained the 45-kDa largest ORF, was cloned and sequenced. The corresponding C3-6 clone contained a 1089-bp ORF with a calculated molecular mass of approximately 42 kDa. This *gag* sequence contained a full-length putative matrix (MA) protein and a carboxyterminal truncated putative capsid (CA) protein, lacking approximately 55 amino acids as compared with the protein sequences of the related BaEVand GaLV infectious retroviruses. As the gag5′ primer imposed the ATG start codon, a U5gag forward primer was designed to verify the presence of a natural ATG. The C3-37 clone obtained (Fig. 2B) presented 1572/1576 (1087/1089 for the ORF) identical nucleotides with the C3-6 clone. A natural ATG start codon was identified in an appropriate context (GCGATGG), although slightly different from the known sequence (A/G)CCATGG favorable for translation.[16] The termination codon was located near the carboxyterminal part of CA, at the same position observed in the C3-6 clone. A glycine codon located just after the ATG start codon indicated a potential myristylation site classically present at the amino-terminal part of MA.[17] Two potential glycosylation sites are present within the capsid C3-6 and C3-37 amino acid sequences. However, *in*

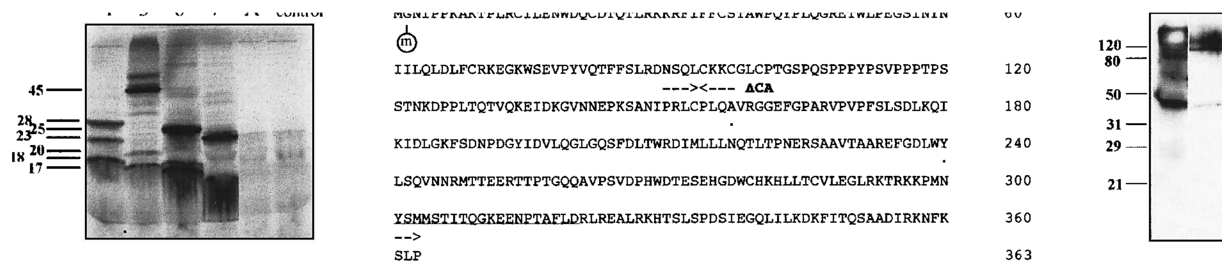

FIG. 2. Coding capacity of HERV-W *gag* region. (**A**) Identification of putative HERV-W *gag* open reading frames by *in vitro* transcription–translation assay. PCR products were *in vitro* transcribed and translated, and [^{35}S]methionine-labeled proteins were separated on SDS-polyacrylamide gels. The upper line indicates the chromosome numbers and the negative control (water control). Numbers to the left indicate the apparent molecular mass of the proteins (in kilodaltons). (**B**) Amino acid sequence of the ORF contained within the C3–37 HERV-W *gag* clone obtained by PCR amplification of human chromosome 3 DNA. Matrix (MA) and partial capsid (ΔCA) putative domains are indicated between arrows. The putative myristylation site ⓜ is indicated at the amino-terminal region of MA. The major homology region (MHR) is underlined and potential glycosylation sites are indicated by dots within the ΔCA domain. (**C**) Western blot analysis of healthy human placental protein extract, using anti-HERV-W Gag (Fab′)$_2$ purified polyclonal antibodies (lane Gag). A nonimmunized rabbit serum was used as negative control (lane C–). Numbers to the left indicate the apparent molecular mass of the proteins (in kilodaltons).

vitro translation in the presence of microsomal membranes failed to detect such a posttranslational maturation (data not shown). A tblastn query using C3-6/C3-37 translated ORF on nonredundant nucleic acid databases provided no additional information to support HERV-W *gag* coding capacities or illustrating chromosomal positive Southern blots versus negative PCR.

To determine if some HERV-W *gag* putative ORFs are expressed as proteins in human placenta, a specific purified polyclonal F(ab′)$_2$ anti-Gag antibody was used in Western blot analysis of extracted proteins from a full-term placenta. A band of about 45 kDa, similar to the *gag in vitro*-translated product derived from chromosome 3, was observed (Fig. 2C). High molecular weight proteins and weak 40-kDa proteins were revealed by the F(ab′)$_2$ nonimmunized rabbit negative control serum.

Coding capacity of the pro *region*

Figure 3A showed that the only HERV-W *pro* open reading frame was obtained on chromosome 6, as a 17-kDa protein. The protein observed is larger than the 12 kDa expected for this retroviral protease, because pro3′ primer was localized in the *pol* region, which is in the same reading frame as the *pro* gene.

The cloning of this HERV-W protease 500-bp fragment led to clone C6-19, which contained a 444-bp open reading frame encoding a putative 16-kDa protein containing a 12-kDa protease subdomain. This ORF exhibited characteristic features of retroviral protease, i.e., a DTG aspartic active site and a flap sequence (Fig. 3B). The different cloned PCR products, which showed some sequence divergences, could represent distinct protease loci (data not shown). A tblastn query showed that such a complete putative HERV-W *pro* ORF was previously sequenced on a human clone 134E15 of chromosome 6 (AL022067). This sequence exhibited 481/484 identical nucleic acids and 145/146 amino acids with the C6-19 clone (109/110 in the protease domain). Although a nonconding *gag* upstream sequence was found in the 134E15 clone, an in-frame TGA stop codon immediately preceded the protease frame along with an in-frame ATG codon located 18 bp upstream. Note that such a protease region was also found on chromosome 4 (AC005187), but was not amplified because of high variation in the downstream region, which included the reverse primer.

As *gag* and *pro* open reading frames have been obtained from chromosomes 3 and 6 respectively, gag5′ and pro3′ primers were used to amplify complete *gag–pro* regions on these chromosomes. However, the *in vitro* transcription–translation assay did not reveal any open reading frame in these PCR products (data not shown).

Coding capacity of the env *region*

As shown in Fig. 4A, a 60-kDa protein compatible with the size of a complete Env protein was exclusively detected on chromosome 7. Furthermore, this Env protein had exhibited a molecular mass of 80 kDa when glycosylated by *in vitro* translation in the presence of canine microsomes (data not shown), as previously observed for the Env protein encoded by the corresponding placental cDNA clone cl.PH74.[9] A truncated protein of 35 kDa encoded by chromosome 5 was the only other potential *env* gene product detected with the SU-associated natural ATG. However, three additional 22-kDa proteins corresponding to full-length TM ORFs were localized on chromosomes 1, 4, and X, using a forward TM-specific primer with an added ATG (data not shown).

The nucleotide sequence of the C7-1 clone, corresponding to the HERV-W *env* ORF present on the isolated chromosome 7, had a single nucleotide difference inducing one amino acid change from the *env* region of the RG083MO5 BAC clone located on the long arm of chromosome 7. This ORF exhibited the characteristic features of the precursor polypeptide of retro-

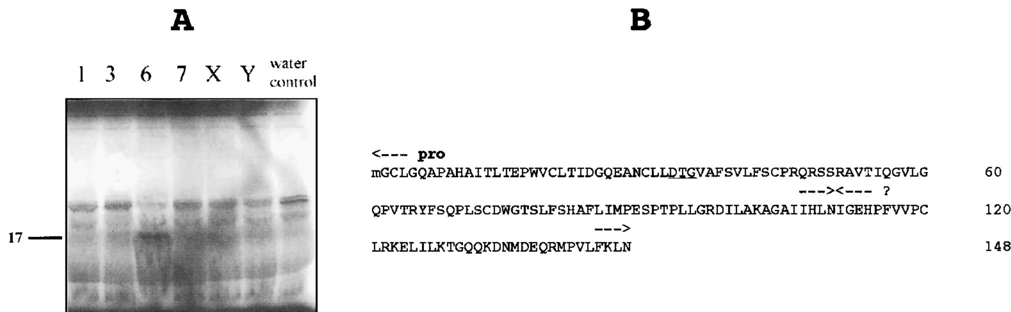

FIG. 3. Coding capacity of HERV-W *pro* region. (**A**) Identification of putative HER-W *pro* open reading frames by *in vitro* transcription–translation assay. PCR products were *in vitro* transcribed and translated, and [^{35}S] methionine-labeled proteins were separated on SDS-polyacrylamide gels. The upper line indicates the chromosome numbers and the negative control (water control). Numbers to the left indicate the apparent molecular mass of the proteins (in kilodaltons). (**B**) Amino acid sequence of the ORF contained within the C6-19 HERV-W *pro* clone obtained by PCR amplification of human chromosome 6 DNA. The initiation codon (m) was brought by the forward PCR primer. The Pro domain is indicated between arrows and the DTG active site is underlined.

viral envelope proteins as previously described for the placental ORF envelope encoded by the cl.PH74 spliced mRNA,[9] i.e., an amino-terminal leader peptide and a carboxy-terminal hydrophobic segment, a furin cleavage site separating the SU and the TM domains, a hydrophobic fusion domain, a putative immunosuppressive region within the TM subdomain, and several conserved potential glycosylation sites (Fig. 4B). It is interesting to note that the different cloned PCR products showed no sequence divergence, suggesting the presence of a unique *env* locus-containing ORF (data not shown)

To determine whether some HERV-W *env* putative ORF is expressed at the protein level in human placenta, a monoclonal anti-Env antibody was used in Western blot analysis. A major 80-kDa doublet was revealed (Fig. 4C), similar to the glycosylated *env in vitro*-translated product derived from chromosome 7.

DISCUSSION

Some HERV elements have retained coding capacity, although they have been present in the human genome for millions of years. Here, we report the chromosomal distribution of the HERV-W family and the persistence of a few ORFs, i.e., *gag, pro,* and *env* ORFs on human chromosomes 3, 6, and 7, respectively. The specific detection of HERV-W subregions by Southern blot on isolated human chromosomes confirmed the complexity of this family.[9] The number of identified HERV-W-related fragments, at least 100 for *pro*, 70 for *gag*, and 30 for *env* per haploid genome, is correlated with the increase in complexity from *env* to *gag* and *pro* regions previously described.[9] Furthermore, HERV-W elements are widely and heterogeneously distributed in the human genome, as half of the

FIG. 4. Coding capacity of HERV-W *env* region. (**A**) Identification of putative HERV-W *env* open reading frames by *in vitro* transcription–translation assay. PCR products were in vitro transcribed and translated, and [^{35}S] methionine-labeled proteins were separated on SDS-polyacrylamide gels. The upper line indicates the chromosome numbers and the cl.PH74[9] positive control (+). Numbers to the left indicate the apparent molecular mass of the proteins (in kilodaltons). (**B**) Amino acid sequence of the ORF contained within the C7-1 clone obtained by PCR amplification of human chromosome 7 DNA. The leader peptide (L) and the surface (SU) and transmembrane (TM) putative domains of Env are indicated between arrows. Potential glycosylation sites are indicated by dots. The hydrophobic fusion peptide and the carboxy-transmembrane region are singly and doubly underlined, respectively, and the putative immunosuppressive region is indicated by italics. (**C**) Western blot analysis of healthy human placental protein extract, using specific anti-Env monoclonal antibody. Numbers to the left indicate the apparent molecular mass of the proteins (in kilodaltons).

chromosomes carried several copies each of the *gag, pro,* and *env* HERV-W subregions (chromosomes 1, 3–7, 9, 11–12, 14–15, 17 and X) but some did not carry any of these HERV-W sequences (chromosomes 2, 16, and 22) while containing an isolated single HERV-W LTR.[9] A similar heterogeneous distribution of HERV elements has been observed for the HERV-K family[18] and the HERV-H family, which is organized in clusters on chromosomes 1, 7, and X.[19,20] In addition, some chromosomes carried only residual fragments of the structural genes of the HERV-W infectious ancestor, e.g., *pro* sequences alone on chromosomes 20 and Y and *env*-TM sequences lacking the associated *env*-SU domain on chromosomes 19 and 21. Such an isolated *env*-TM subregion had been previously found to be transcribed on a single RNA downstream from the GTP-binding protein RAB7.[9]

Some virus-like particles (VLPs), whose origin is currently unknown, have been described in healthy tissues, such as placenta and blood donor peripheral blood mononuclear cells,[21,22] and in autoimmune diseases such as multiple sclerosis and systemic lupus erythematosus,[23,24] and could be encoded by human endogenous retroviral sequences. The HERV-W family retained a few coding capacities, i.e., three ORFs were detected on different chromosomes, a 363 amino acid-encoding *gag* ORF located on chromosome 3, a 148 amino acid-encoding *pro* ORF on chromosome 6, and the previously described 538 amino acid-encoding *env* ORF on chromosome 7,[9] which was shown to be unique. The 45-kDa *gag* ORF, with a corresponding protein expressed in placenta, contained a complete MA domain and a nearly complete CA domain including the major homology region (MHR), but lacking nucleocapsid, as compared with the sequences of related infractious retroviruses such as GaLV and BaEV. A 148 amino acid-encoding ORF was identified as encoding a canonical 110 amino acid protease region with respect to classic retroviruses. As no *gag* ORF was detected upstream from the *pro* ORF, the absence of a natural ATG initiation codon draws into question the capacity of such a sequence to be translated. However, the presence within the same reading frame of an ATG followed by a stop codon immediately upstream from the protease sequence may suggest a termination suppression event. Furthermore, as chromosome 6 carries a high density of *pro* ORF sequences, it cannot be excluded that an undetected favorable context of initiation may exist, possibly resulting from a splicing event.

The unique *env* ORF, with protein expression in the placenta, encodes an 80-kDa glycosylated protein containing complete SU and TM domains, as previously described.[9] The exclusive finding of *gag, pro,* and *env* ORFs on three different chromosomes suggests that the human genome does not contain an HERV-W replication-competent provirus. This was confirmed by a sequential multiprobe screening of a human DNA library (F. Besème, J.-L. Blond, O. Bouton, and F. Mallet, unpublished data). However, it would be of interest to determine whether virus-like particles may be formed either by *trans*-complementation events, as proposed for HTDV and HERV-K-T47D particles,[8,25] or by Gag precursor alone.[8,26,27] Nevertheless, the absence of the nucleocapsid domain, although not essential for the formation of virus-like particles,[28,29] would impair the specific incorporation of related RNA. In addition to these three ORFs identified by specific *in vitro* transcription-translation assays, it should be mentioned that the test did not succeed in amplifying all the elements detected by the Southern blot analysis. These discrepancies might be explained by mutations altering the hybridization efficiency of the primer, as observed for the *pro* region found on chromosome 4 (AC005187), or insertion of larger sequences hindering the PCR amplification. However, a databank analysis did not reveal any additional HERV-W-related *gag, pro,* or *env* ORF other than the one we experimentally observed, except for a *pro* region located on chromosome 4.

Placental protein expression is a common feature of HERV.[30] HERV-W *gag* and *env* mRNAs were indeed previously identified in the placenta, including a full-length cDNA encoding Env but no cDNA encoding Gag,[9] a situation that may result from the relative abundance of *env* and *gag* genes in the human genome, but also from a differential regulation process. Moreover, the protein expression of the other truncated *gag* and *env* ORFs found on chromosomes 1, 3, and 7 for Gag and on chromosome 5 for Env has not been detected in placental tissue, suggesting the expression of subsets of the HERV-W family in different tissues, as described at the mRNA level for the expression of HTDV/HERV-K,[30] ERV-9,[3] and HERV-R.[31,32] Thus it would be of interest to identify the promoter sequence driving the expression of the *gag* ORF located on chromosome 3 in order to compare it with the *env*-associated U3 promoter located on chromosome 7. The Env placental expression suggests some physiological functions related to pregnancy, such as a role in the creation of the syncytiotrophoblast layer of the placenta[33–35] and/or in suppressing the maternal immune response against the fetal allograft[33,35,36] as suggested originally for the envelope of ERV-3. However, the existence of an *env* ERV-3 knockout[37] showed that the polymorphism of the HERV-W *env* ORF should be addressed.

In pathological situations, Env and Gag HERV-W proteins may protect against exogenous retroviral infections by a receptor interference mechanism, as described for the mouse Fv-1 and Fv-4 mouse genes, respectively.[38] Gag may play a role in the regulation of the allergic immune response as observed for the rodent IgE-binding factor.[39] Env-driven alteration of the immune system was also proposed via the immunosuppressive region[40] or a superantigen-encoded region as proposed in type I diabetes,[41] although these latter data remain controversial.[42] Conversely, HERV-W expression may only represent a consequence of physiological or pathological events. Whatever the situation, the detection of HERV-W proteins in a physiological tolerogenic immune context such as pregnancy deserves further investigation.

ACKNOWLEDGMENTS

This study was supported by bioMérieux S.A. We thank David Griffiths and Robin Weiss (Wohl Virion Center, University College London, United Kingdom) and Doran Spencer (UMR 103 CNRS-bioMérieux, Lyon, France) for critical reading of the manuscript. Cécile Voisset was supported by a doctoral fellowship from the Fondation Mérieux.

SEQUENCE DATA

The sequences described in this article have been submitted to Genbank under the following accession numbers (HERV-W

chromosomal isolated clones): gag C3-37 (AF156961), pro C6-19 (AF156962), env C7-1 (AF156963).

REFERENCES

1. Wilkinson DA, Mager DL, and Leong JAC: Endogenous human retroviruses. In: *The Retroviridae* (Levy JA, ed.). Plenum, New York, 1994, pp. 465–535.
2. Cohen M, Powers M, O'Connell C, and Kato N: The nucleotide sequence of the *env* gene from the human provirus ERV-3 and isolation and characterization of an ERV-3-specific cDNA. Virology 1985;147:449–458.
3. Lindeskog M, Medstrand P, and Blomberg J: Sequence variation of human endogenous retrovirus ERV-9-related elements in an *env* region corresponding to an immunosuppressive peptide: Transcription in normal and neoplastic cells. J Virol 1993;67:1122–1126.
4. Murphy VJ, Harrison LC, Rudert WA, Luppi P, Trucco M, Fierabracci A, Biro PA, and Bottazzo GF: Retroviral superantigens and type 1 diabetes mellitus. Cell 1998;95:9–11.
5. Ono M, Yasunaga T, Miyata T, and Ushikubo H: Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. J Virol 1986;60:589–598.
6. Banki K, Maceda J, Hurley E, Ablonczy E, Mattson D, Szegedy L, Hung C, and Perl A: Human T-cell lymphotropic virus (HTLV)-related endogenous sequences, HERS-1, encodes a 28 kDa protein: A possible autoantigen for HTLV-I Gag-reactive autoantibodies. Proc Natl Acad Sci USA 1992;89:1939–1943.
7. Boller K, König H, Sauter M, Mueller-Lantzsch N, Löwer R, Löwer J, and Kurth R: Evidence that HERV-K is the endogenous retrovirus sequence that encodes for the human teratocarcinoma-derived retrovirus HTDV. Virology 1993;196:349–353.
8. Tönjes RR, Boller K, Limbach C, Lugert R, and Kurth R: Characterization of human endogenous retrovirus type K virus-like particles generated from recombinant baculoviruses. Virology 1997; 233:280–291.
9. Blond JL, Besème F, Duret L, Bouton O, Bedin F, Perron H, Mandrand B, and Mallet F: Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family. J Virol 1999;73:1175–1185.
10. Perron H, Garson JA, Bedin F, Beseme F, Paranhos-Baccala G, Komurian-Komurian-Pradel F, Mallet F, Tuke PW, Voisset C, Blond JL, Lalande B, Seigneurin JM, and Mandrand B: Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. Proc Natl Acad Sci USA 1997;94: 7583–7588.
11. Gaudin P, Perron H, Favre G, Mandrand B, Juvin R, Marcel F, Besème F, Bedin F, Mallet F, Mougin B, Fauconnier J, Seigneurin JM, and Phelip X: Detection of retrovirus RNA in plasma from rheumatoid arthritis. Arthritis Rheum 1997;40:S245.
12. Voisset C, Blancher A, Perron H, Mandrand B, Mallet F, and Paranhos-Baccalà G: Phylogeny of a novel family of human endogenous retrovirus sequences, HERV-W, in humans and different primates. AIDS Res Hum Retrovirus. 1999;15:1529–1533.
13. Roest PA, Roberts RG, Sugino S, van Ommen GJ, and den Dunnen JT: Protein truncation test (PTT) for rapid detection of translation-terminating mutations. Hum Mol Genet 1993;2:1719–1721.
14. Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ: Basic local alignment search tool. J Mol Biol 1990;215:403–410.
15. Komurian-Komurian-Pradel F, Paranhos-Baccala G, Bedin F, Ounanian-Paraz A, Sodoyer M, Ott C, Rajoharison A, Garcia E, Mandrand B, and Perron H: Molecular cloning and characterization of overlapping sequences associated with retrovirus-like particles in multiple sclerosis. Virology 1999;260:1–9.
16. Kozak M: A consideration of alternative models for the initiation of translation in eukaryotes. Crit Rev Biochem Mol Biol 1992;27: 385–402.
17. Henderson LE, Krutzsch HC, and Oroszlan S: Myristyl amino-terminal acylation of murine retrovirus proteins: An unusual post-transcriptional protein modification. Proc Natl Acad Aci USA 1983;80:339–343.
18. Meese E, Gottert E, Zang KD, Sauter M, Schommer S, and Mueller-Lantzsch N: Human endogenous retroviral element K10 (HERV-K10): Chromosomal localization by somatic hybrid mapping and fluorescence in situ hybridization. Cytogenet Cell Genet 1996; 72:40–42.
19. Fraser C, Humphries R, and Mager D: Chromosomal distribution of the RTVL-H family of human endogenous retrovirus-like sequences. Genomics 1988;2:280–287.
20. Shiraishi M, Alitalo T, and Sekiya T: The chromosomal organization of the human endogenous retrovirus-like sequence HERV-H: Clustering of the HERV-H sequences in a 300-kb region close to the GRPR locus on the X chromosome. DNA Res 1996;3:425–429.
21. Kalter SS, Helmke RJ, Heberling RL, Panigel M, Fowler AK, Strickland JE, and Hellman A: C-type particles in normal human placentas. J Natl Cancer Inst 1973;50:1081–1084.
22. Dourmashkin RR, Bucher D, and Oxford JS: Small virus-like particles bud from the cell membranes of normal as well as HIV-infected human lymphoid cells. J Med Virol 1993;39:229–232.
23. Munch M, Moller-Larsen A, Christensen T, Morling N, Hansen HJ, and Haahr S: B-lymphoblastoid cell lines from multiple sclerosis patients and a healthy control producing a putative new human retrovirus and Epstein–Barr virus. Multiple Sclerosis 1995;1: 78–81.
24. Tamura N, Sekigawa I, Hashimoto H, Yamamoto N, and Kira S: Syncytial cell formation in vitro by type C retroviral particles in the systemic lupus erythematosus (SLE) lung. Clin Exp Immunol 1997;107:474–479.
25. Seifarth W, Baust C, Murr A, Skladny H, Krieg-Schneider F, Blusch J, Werner T, Hehlmann R, and Leib-Mösch C: Proviral structure, chromosomal location, and expression of HERV-K-T47D, a novel human endogenous retrovirus derived from T47D particles. J Virol 1998;72:8384–8391.
26. Korbmacher C, König H, Boller K, Löwer R, Kurth R, and Löwer J: The *gag* gene of the endogenous retrovirus HERV-K is sufficient for particle production. J Cancer Res Clin Oncol 1993; 119(Suppl. 1):S5.
27. Wills JW and Craven RC: Form, function, and use of retroviral Gag proteins. AIDS 1991;5:639–654.
28. Gonzalez SA, Affranchino JL, Gelderblom HR, and Burny A: Assembly of the matrix protein of simian immunodeficiency virus-like particles. Virology 1993;194:548–556.
29. Méric C and Goff SP: Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the Cys-His box of the nucleocapsid protein. J Virol 1989;63:1558–1568.
30. Löwer R, Löwer, J, and Durth R: The viruses in all of us: Characteristics and biological significance of human endogenous retroviral sequences. Proc Natl Acad Sci USA 1996;93:5177–5184.
31. Cohen M, Kato N, and Larsson E: ERV3 human endogenous provirus mRNA are expressed in normal and malignant tissues and cells, but not in choriocarcinoma tumor cells. J Cell Biochem 1988;36:121–128.
32. Kato N, Larsson E, and Cohen M: Absence of expression of a human-endogenous retrovirus is correlated with choriocarcinoma. Int J Cancer 1988;41:380–385.
33. Boyd MT, Bax CMR, Bax BE, Bloxam DL, and Weiss R: The human endogenous retrovirus ERV-3 is upregulated in differentiating placental trophoblast cells. Virology 1993;196;905–909.
34. Lin L, Xu B, and Rote NS: Expression of endogenous retrovirus ERV-3 induces differentiation in BeWo, a choriocarcinoma model of human placental trophoblast. Placenta 1999;20:109–118.

35. Venables PJW, Brookes SM, Griffiths D, Weiss RA, and Boyd MT: Abundance of an endogenous retroviral envelope protein in placental trophoblasts suggests a biological function. Virology 1995;211:589–592.
36. Villareal LP: On viruses, sex, and motherhood. J Virol 1997;72: 859–865.
37. De Parseval N and Heidmann T: Physiologicla knockout of the envelope gene of the single-copy ERV-3 human endogenous retrovirus in a fraction of the Caucasian population. J Virol 1998;72: 3442–3445.
38. Best S, Teissier PR, and Stoye JP: Endogenous retroviruses and the evolution of resistance to retroviral infection. Trends Microbiol 1997;5:313–318.
39. Moore KW, Jardieu P, Mietz JA, Trounstine ML, Kuff EL, Ishizaka K, and Martens CL: Rodent IgE-binding factor genes are members of an endogenous, retrovirus-like gene family. J Immunol 1986; 136:4283–4290.
40. Mangeney M and Heidmann T: Tumor cells expressing a retroviral envelope escape immune rejection in vivo. Proc Natl Acad Sci USA 1998;95:14920–14925.
41. Conrad B, Weissmahr RN, Böni J, Arcari R, Schüpbach J, and Mach B: A human endogenous retroviral superantigen as candicate autoimmune gene in type I diabetes. Cell 1997;90:303–313.
42. Löwer R, Tönjes RR, Boller K, Denner J, Kaiser B, Phelps RC, Löwer J, Kurth R, Badenhoop K, Donner H, Usadel KH, Miehke T, Lapatschek M, and Wagner H: Development of insulin-development diabetes mellitus does not depend on specific expression of the human endogenous retrovirus HERV-K. Cell 1998;95: 11–14.

Address reprint requests to:
François Mallet
Unité Mixte 103 CNRS-bioMérieux
Ecole Normale Supérieure de Lyon
46 allée d'Italie
69364 Lyon, Cédex 07, France

E-mail: fmallet@ens-bma.cnrs.fr