The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology

François Mallet*,†, Olivier Bouton*, Sarah Prudhomme*, Valérie Cheynet*, Guy Oriol*, Bertrand Bonnaud*, Gérard Lucotte†, Laurent Duret§, and Bernard Mandrand*

*Unite´ Mixte de Recherche 2142, Centre National de la Recherche Scientifique–bioMérieux, IFR128 BioSciences Lyon-Gerland, Ecole Normale Superieure de Lyon, 46 Allée d’Italie, 69364 Lyon Cedex 07, France; †Centre de Neurogen´etique Mole´culaire, 44 Rue Monge, 75005 Paris, France, and §Laboratoire de Biome´trie et Phyliologie Evolutive, Unite´ Mixte de Recherche Centre National de la Recherche Scientifique 5558, Universite Claude Bernard, Lyon 1, 43 Boulevard du 11 November 1918, 69622 Villeurbanne Cedex, France

Edited by John M. Coffin, Tufts University School of Medicine, Boston, MA, and approved December 2, 2003 (received for review September 9, 2003)

The definitive demonstration of a role for a recently acquired gene is a difficult task, requiring exhaustive genetic investigations and functional analysis. The situation is indeed much more complicated when facing multicopy gene families, because most or portions of the gene are conserved among the hundred copies of the family. This is the case for the ERVWE1 locus of the human endogenous retrovirus W family (HERV-W, which encodes an envelope glycoprotein (syncytin) likely involved in trophoblast differentiation. Here we describe, in 155 individuals, the positional conservation of this locus and the preservation of the envelope ORF. Sequencing of the critical elements of the ERVWE1 provirus showed a striking conservation among the 48 alleles of 24 individuals, including the LTR elements involved in the transcriptional machinery, the splice sites involved in the maturation of subgenomic Env mRNA, and the Env ORF. The functionality and tissue specificity of the 5’ LTR were demonstrated, as well as the fusogenic activity of the envelope polymorphic variants. Such functions were also shown to be preserved in the orthologous loci isolated from chimpanzee, gorilla, orangutan, and gibbon. This functional preservation among humans and during evolution strongly argued for the involvement of this recently acquired retroviral envelope glycoprotein in hominoid placental physiology.

We molecularly characterized the human endogenous retrovirus W family (HERV-W) family (1) by screening a placental cDNA library with a polymerase (pol) probe derived from a retroviral sequence named multiple-sclerosis associated retrovirus (MSRV) (2) isolated from biological samples from multiple sclerosis patients. Surprisingly, this multicopy family contained a unique proviral locus, located on chromosome 7, flanked by two intact LTRs, and which retained a 538-aa envelope ORF that exhibited all of the characteristic features of the precursor polypeptide of classical retroviral envelope proteins (1, 3). The presence of inactivating mutations in the gag and pol genes of this provirus, termed ERVWE1 (OMIM 604659), led us to propose that Env function has been selectively preserved (1).

Although HERV-W mRNA expression was detected in various physiological (1, 4) and pathological (2, 5, 6) contexts, HERV-W transcript pattern observed in placenta correlates with a LTR5′-U3 driven transcription coupled with a specific splicing strategy compatible with the ERVWE1 locus (1). Env expression was evidenced in placenta (3) at the syncytiotrophoblast layer (7, 8) covering the chorionic villi, bathed by maternal blood in the intervillous spaces, which is the site of nutrient exchanges and of the synthesis of hormones required for fetal growth and development (9, 10). We and others recently demonstrated that the product of the HERV-W env gene is a highly fusogenic membrane glycoprotein (4, 7), inducing the formation of syncytiotrophoblast upon interaction with the type D mammalian retrovirus fusogen receptor (7). The in vitro fusogenic activity of Env HERV-W and its in vivo localization suggested that this glycoprotein may have a critical role in human placental development steps involving the syncytiotrophoblast layer. Such a hypothesis was supported by the observations that an anti-Env-W polyclonal antibody was able to inhibit heterologous fusion between a BeWo cell line and COS reporter cells (4) and that anti-ERVWE1 antisens oligonucleotides were able to inhibit primary human trophoblast cell fusion and differentiation (8).

Nevertheless, as previously highlighted (7, 8, 11), there is a huge gap between the demonstration of the fusogenic function and a definitive proven physiological role. The ERVWE1 locus is a young element, as the HERV-W ancestor active element entered the primate lineage 25–40 million years ago (12). Our working hypothesis is that this element has been selected during primate evolution to play a role in placentation. However, it is also possible that this locus is a still-active pseudogene in the process of extinction, and that its activity in the syncytiotrophoblast formation is simply a remnant of its original function for the propagation of retroviruses, but that has no real physiological role. To rule out this latter hypothesis, we have designed a genetic and functional analysis of the ERVWE1 locus. We show in this article that the ERVWE1 locus is functionally preserved in the human population and in the identified orthologous locus of chimpanzee, gorilla, orangutan, and gibbon.

Materials and Methods

DNA Samples. The human rodent monochromosomal National Institute of General Medical Sciences somatic hybrid mapping panel 2 was obtained from the Coriell Institute (Camden, NJ). Stored human DNA from 180 individuals were used, including 59 Caucasians, three Asians, two Africans, five Metis, and one Ashkenazi. Thirty samples were derived from multiple sclerosis patients, including 22 familial forms (13). Ninety-three individuals were male, and 87 individuals female, as determined by PCR amplification of a segment of the X–Y homologous gene amelogenin (14). Animal DNAs from three chimpanzees (Pan troglodytes), three gorillas (Gorilla gorilla), two orangutans (Pongo pygmaeus), three gibbons (Hylobates pileatus) (one DNA per species from Quantum Biotechnologies Inc, Canada), six rhesus monkeys (Macaca mulatta), and one guinea baboon (Papio papio) were used.

Long-Distance (LD) PCR. The ERVWE1 locus was amplified by using sense primer U6198R, 5′-CAAGACGGCTTGAGATA-CAGCAATTATC-3′ and antisense primer L6186R, 5′-GCACCCCTATGTTGTTCTTGGC-3′ as illustrated in

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: HERV, human endogenous retrovirus; LD-PCR, long-distance PCR.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos.: human full-length ERVWE1 locus (4 entries), AY101582–AY101585; human 5′ and 3′ partial ERVWE1 locus (88 entries), AF520477–AF520564; and chimpanzees, gorilla, orangutan, and gibbon full-length ERVWE1 locus (8 entries), AY101586–AY101593].

1To whom correspondence should be addressed. E-mail: francois.mallet@ens-lyon.fr.

© 2004 by The National Academy of Sciences of the USA
The mutations corresponding to the four observed human polymorphic variants were introduced into the pCMV-Env plasmid (7) using a QuickChange mutagenesis kit (Stratagene) and verified by sequencing. This produced five expression vectors named according to the residues localized in each polymorphic site, i.e., VRS, VRN, VQ, aRS, and VQ. Similar ape ERVWE1 Env-containing fragments were cloned into the pCMV vector. The envelopes fusion efficiencies were determined as described (7). The overall experiment was repeated five times, and differences between envelopes fusion properties were assessed by using the Kruskal-Wallis test (StatView II, Abacus Concepts, Berkeley, CA). Receptor-dependent fusion activity was determined by using HEK-293 cells and X-CRDR cell lines as described (7). Detailed protocols are included in Appendix. 

Results

Positional Conservation of the ERVWE1 Locus Containing a Preserved env ORF. Primers designed to demonstrate the positional conservation of the ERVWE1 locus (Fig. 1a) were evaluated in LD-PCR using human-rodent monochromosomal DNA. A 10-kb band can be detected only on chromosome 7 as expected (Fig. 1b). The same protocol was applied to 180 individuals, and 155 individuals were found positive as partially illustrated in Fig. 1c. The 25 ERVWE1-negative DNA were also LD-PCR-negative by using the 16.2-kb mitochondrial genome and the 13.5-kb human β-globin gene cluster control sequences, showing

Cell Lines. The b30 BeWo cell line was grown in F-12K medium (Life Technologies) containing 10% FCS and fungizone. Other cell lines were as follows: LC5, human lung fibroblasts; HeLa, human epithelioid carcinoma cells (ATCC CCL2); TELCeB6 rhabdomyosarcoma cells (7); XC, rat sarcoma cells (ATCC CCL-165); and XC-RDR cells, XC cells expressing the RDR receptor (8). LC5, HeLa, TELCeB6, XC, and XC-RDR cells were cultured in DMEM (Life Technologies) (without sodium pyruvate for LC5) supplemented with 10% FCS (5% FCS for TELCeB6) and fungizone.

LTR Luciferase Assays. LTRs were PCR amplified and cloned into pGL3-basic plasmid (Promega). HERVWE1 [5] and HERVWE1 [3] clones corresponded to nucleotide -35/+309 and -25/+311 of the human ERVWE1 5' and 3' LTRs, respectively. HERV-WHS14 [3] and HERV-W8 [3] (plasmid pBL-W8, ref. 18) clones corresponded to nucleotides -34/+375 and -16/+325 of 3' LTRs derived from a provirus located on chromosome 14 (GenBank accession number AL365295) and from an mRNA expressed in a breast cancer cell line, respectively. pGL3-basic and parental pBL-basic plasmids (control) exhibited similar background activity in all tested cell lines (data not shown). ERVWE1 5' and 3' ape cloned sequences corresponded exactly to their human counterparts. A total of 0.7 μg of firefly luciferase LTR-ERV-W plasmid and 0.035 μg of Renilla luciferase pRL-TK plasmid (standardization) were transiently transfected in choriocarcinoma b30 BeWo and fibroblastic LC5 cell lines by using Lipofectamine Plus (Life Technologies) method in 12-well plate format. Forskoline (50 μM) was added after 12 h to induce BeWo differentiation. All experiments were performed in triplicate. Promoter activities were calculated as [firefly/Renilla] × 100 luciferase activities by using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions.

Envelope Fusion Assays. The mutations corresponding to the four observed human polymorphic variants were introduced into the pCMV-env plasmid (7) using a QuickChange mutagenesis kit (Stratagene) and verified by sequencing. This produced five expression vectors named according to the residues localized in each polymorphic site, i.e., VRS, VRN, VQ, aRS, and VQ. Similar ape ERVWE1 env-containing fragments were cloned into the pCMV vector. The envelopes fusion efficiencies were determined as described (7). The overall experiment was repeated five times, and differences between envelopes fusion properties were assessed by using the Kruskal-Wallis test (StatView II, Abacus Concepts, Berkeley, CA). Receptor-dependent fusion activity was determined by using HEK-293 cells and X-CRDR cell lines as described (7). Detailed protocols are included in Appendix.
that those DNA preparations were not suitable for LD-PCR purpose (data not shown). Thus, the ERVWE1 locus is present in 100% of the 155 analyzable individuals. The conservation of the envelope ORF in all 155 ERVWE1 LD-PCR products was evaluated by an in vitro transcription–translation assay subsequent to a specific env gene nested PCR step (Fig. 1a). A full-length ORF was observed in all 155 individuals, as partially illustrated in Fig. 1d. Because the 5’ primer contained the initiation codon, the presence of the env natural ATG was confirmed by direct sequencing of the env nested-PCR product of the 155 individuals. The sequence was remarkably conserved, as not a single mutation was observed within a 20/20 interval surrounding the ATG codon (AACT AAAATCATAAATC– of the 155 individuals. The sequence was remarkably conserved, illustrated in Fig. 1.

Global Sequence Analysis of the ERVWE1 Locus. The sequencing and analysis strategy is summarized in Fig. 1e. The 10-kb PCR-derived sequences obtained from two individuals presented exactly the same structure as the previously identified env-containing HERV-W provirus (1) contained in human bacterial artificial chromosome clone RG083M05 from 7q21 to 7q22 (GenBank accession no AC007566), i.e., LTR–2kb insert–gag–pol–env–LTR, gag and pol genes containing many stop codons. The comparative analysis of the 48 sequences obtained from all 24 individuals have shown the presence of few polymorphic sites exhibiting an heterogeneous distribution along the locus (see below). The global mean divergence between the four full-length proviruses was ~0.05%. The local mean divergence calculated with the 48 partial sequences ranged from 0.005% to 0.2% for the 5’ LTRs and the 3’ LTRs, respectively, and was 0.04% for the env ORF, similar to the overall provirus.

Transcriptional Elements Are Functionally Preserved. The 5’ LTR was particularly conserved as a single mutation appeared in the U5 region in one of the 48 sequences (Fig. 7, which is published as supporting information on the PNAS web site), i.e., 47 sequences were strictly identical (e.g., isolate 132 allele A). Conversely, nine different alleles were observed for the 48 sequenced 3’ LTR, with variable frequencies: 18 of 48 (e.g., isolate 37B), 14 (e.g., isolate 24A), 6 (e.g., isolate 22B), 3 (e.g., isolates 83A and 37A), and 1 (isolates 132A, 71A, 80B, 132B). Most of these polymorphic sites relied within the U5 subdomain of the 3’ LTR.

The alignment of the two ERVWE1 5’ LTRs with the nine ERVWE1 3’ LTRs, the 3’ LTR of env containing clone PH74 derived from a placental cDNA library (HERV-WPH74), the 3’ LTR of an mRNA expressed in a breast cancer cell line (HERV-W8), and another HERV-W LTR locus from chromosome clone containing the ERVWE1 locus (bold letters), the PH74 placental cDNA clone containing a full-length env ORF (underlined), and two other HERV-W LTRs (HERV-WHS14 [3], HERV-WB [3]) neighbor-joining trees were based on 244 nucleotides, and all-gap removal was used. Bootstrap values for 500 replicated trees are indicated. (b) Relative activities of ERVWE1 5’ (HERVWE1 [5] and 3’ (HERVWE1 [3]) LTRs, and other HERV-W 3’ LTRs (HERV WHS14 [3], HERV-WB [3]) in choriocarcinoma BeWo b30 cell line in absence (gray bar) or presence (black bar) of forskolin and in lung fibroblastic LC5 cell line (white bar). Control, basic vector without LTR.

Env Polymorphic Variants Are Functionally Preserved. Within the env ORF, five mutations were detected including one synonymous and four non synonymous mutations (Fig. 3a). This corresponded to five envelope polymorphic variants named according to the residues present in the four polymorphic positions, i.e., VRSS, VRN3, VQ53, ARSS, Vq5f, representing 67%, 25%, 4%, 2%, and 2% of the sequenced population. Two major alleles VRSS (Fig. 8, which is published as supporting information on the PNAS web site) and VRN3 were present in all of the individuals of the sequenced population. The VRSS predominant form was identical to the one we previously identified in a placental cDNA.
Polymorphism and functional analysis of the ERVWE1 envelope in the human population. (a) Schematic drawing of the Env glycoprotein indicating the location of the polymorphic sites and the nature and frequency of the amino acids. SU, surface protein; TM, transmembrane protein; fp, fusion peptide; ir, immunosuppressive region; TR, carboxy-transmembrane region. (b) Formation of syncytia by ERVWE1 envelopes. TELCeB6 cells expressing the aRSS-VRSS, and VqSf-VRNS. No skewed allele distribution was observed in the whole locus of the five ERVWE1 orthologs. The observed fusions were detected in the whole locus of the five ERVWE1 orthologs. Interestingly, none of these gaps occurred within the env ORF, whose length was strictly constant. The even distribution of the gaps along the other regions of the locus suggested that there was a strong selective pressure to maintain the env ORF intact. Moreover, the env ATG context remained identical to the one observed in all human samples. Splice sites previously identified (1) were also remarkably conserved, notably the DS1 and AS2 donor and acceptor sites involved in env mRNA processing (Fig. 4).

The alignment of U3 domains of 5' and 3' LTRs revealed that both LTRs are distinct within each species (Fig. 9, which is published as supporting information on the PNAS web site). All LTRs contained a CAAT box and a TATA box, identical or slightly different from the human motifs, e.g., the TATA box of the 5' LTRs of both alleles of the gibbon was AATAAA as compared to the human sequence. The sites that were shown to discriminate between 5' and 3' human U3 did not systematically discriminate for each species, with the exception of the 3' last base of U3, which was an A vs. a G in 5' and 3' LTRs, respectively. Nevertheless, the functional analysis of all U3 elements revealed that the ape ERVWE1 5' LTRs were always more active in BeWo cells than the ERVWE1 3' LTRs, as observed for the human ERVWE1 locus (Fig. 5). Within the R region, the T-rich region, the polyadenylation signal, and the CA dinucleotide poly(A) site were conserved for all LTRs; two mutations located within the first 30 bases or R, which discriminated between the 5' and 3' human LTR, also specified the 3' R region in apes (data not shown).

Sequencing of the four ape env ORFs showed that all these animals were homoyzous for this protein. The similarities observed by pairwise alignment of Env proteins between human and chimpanzee, gorilla, orangutan, and gibbon were 96.9%, 97.3%, 99.7%, and 94.6%, respectively. Alignments of human and ape envelopes showed a high conservation of critical domains essential for classical retroviral envelope expression and function (Fig. 6a). The topology of the tree obtained by the
Comparative analysis of the ERVWE1 locus for 24 individuals showed a pattern of sequence variation that was variable along the proviral locus. The 5' LTR exhibited an unusually low polymorphism (one variable site every 18.0 kb) as compared to the one in 0.47 kb and one in 0.31 kb variability described for noncoding sequences and repeated sequences (20), respectively, suggesting that there has been a selective sweep of this region. Conversely, the variability of the 3' LTR (one in 0.5 kb) was typical of repeated sequences (20). The observed variability of the env ORF (one variable site every 2.2 kb) fell within the same range as the one in 1.08–2.00 kb variability described for human coding sequences (20), highlighting that the behavior of this gene of retroviral origin is similar to any essential cellular gene, as opposed to infectious retroviruses or more generally RNA-based organisms (21).

The identification of the ERVWE1 locus exclusively in hominoids may reflect two situations. First, the ERVWE1 locus had retrotransposed in the hominoid lineage after the split of hominoids from the Old World monkeys ~15–20 million years after the germ line insertion of the ancestral virus in the Catarrhini lineage (12). Second, this locus was already present in the ancestor of the Catarrhini lineage and was fixed in the hominoid lineage, but has degenerated in the Old World monkey lineage. Further analyses will be required to choose between these two hypotheses. Nevertheless, the structure conservation of both LTRs and the sequence conservation of the 5' and 3' proviral-cellular junctions, for all hominoid species, suggested that the ERVWE1 transposition process depended on a functional and specific HERV-W reverse transcriptase and not a LINE-mediated retrotransposition process as observed for other HERV-W elements (22, 23).

Although HERV-W expression was detected in autoimmune diseases (2, 5) and schizophrenia (6), no direct link between mRNA expression and (re)activation of a specific HERV-W locus (or a subset of loci) was demonstrated to date (24, 25).

Conversely, the ERVWE1 locus is the unique HERV-W locus for which a tissue specificity was proposed (1, 4). Interestingly, the U3 region of the 5' HERVWE1 LTR was the most well conserved in the human population, and comparison with the other HERV-W LTRs tested showed that it contained some specific signatures. The presence of such a combination of sites was apparently correlated with the HERVWE1 promoter efficiency and tropism, as the 5' HERVWE1 LTR was the most active promoter in BeWo cells. Nevertheless, among the nine sites that were shown to discriminate between 5' and 3' HERVWE1 U3 regions, only a single one was conserved during the whole hominoid evolution. In addition, because cloned sequences contained some residues upstream from U3 and a downstream R sequence providing the natural cap site, further analysis will be required to identify the determinants involved in the observed tropism. Nevertheless, it is noticeable that all active
ERVWE1 LTRs lacked a Sp1 site identified in a subset of inactive HERV-W LTRs and absent from another subset of active HERV-W LTRs (18).

All five human envelope polymorphic variants and four ape orthologous envelopes were found to be functional in an in vitro fusogenic assay, a process dependent on the recognition of the hASCT-2/RDR receptor. This was also true for the VNRS human mutant, which has been previously proposed to be involved in a receptor-independent fusion process (4), in contradiction with the apparent selectivity of trophoblast fusion, as highlighted elsewhere (11). The similar recognition of the human ASCT-2 receptor by all hominoid ERVWE1 envelopes suggested that the ape ASCT-2 receptors were closely related to the hASCT-2. Indeed, it would be interesting to isolate such hASCT-2 orthologous receptors with the potential of resolving the conflict between the coevolution of the receptor with the endogenous envelope and the introduction of hypervariability to avoid putative pathogenic retroviral infection (26, 27). In addition, the sequencing of the 48 HERVWE1 envelopes confirmed that the MSRV envelope (GenBank accession no. AF331500), which has been proposed to trigger a superantigen effect (28) involved in multiple sclerosis, was not encoded by the ERVWE1 locus. First, Env HERVWE1 presented an 81% identity with Env MSRV, and second, a 4-aa gap that was shown to be a unique signature of Env HERVWE1 among all HERV-W envelope sequences (data not shown) is absent from Env MSRV. The presence of inactivating mutations in the gpol and pol genes of the HERVWE1 provirus initially led us to propose that Env function has been selectively preserved (1). The demonstrations that the Env HERV-W glycoprotein was (i) fusogenic in vitro (4, 7), (ii) localized at the syncytiotrophoblast layer of the placenta (7, 8), and (iii) directly involved in primary human trophoblast cell fusion and differentiation (8) supported the hypothesis of a role in placenta formation. The functional preservation of the transcripational and posttranscriptional regulatory elements and of the fusogenic property of the Env glycoprotein of the HERVWE1 locus in the human population and in hominoids demonstrated that this recently acquired gene has become a bona fide gene. The corollary of the implication of this envelope in the formation of the continuous multinucleate syncytiotrophoblast is that a deregulation of Env expression could be associated with pathologies ranging from implantation to delivery. Interestingly, Env was poorly expressed in pathogens associated with hypoxia (29), hypoxia-impairing cytotrophoblast fusion, and differentiation in vitro (30). Because the majority of the Eutherians do not contain the young HERVWE1 locus, it will be challenging to ascertain whether this retroviral acquisition represents an additional adaptive or substitutive mechanism for placental morphogenesis (31). The recent finding of an in vitro fusogenic HERV-FRD envelope conserved during primate evolution (32) raises questions about the apparent redundancy of recently acquired genes sharing common function(s). This may thereby contribute to our understanding of the extreme diversity of the placental structures observed in nature (33).

We thank B. Rudkin for critical reading of the manuscript. We thank B. Mougin, S. Mazoyer, A. Blancher, J. Brosius, C. Coiffier, and G. Hunsmann for providing us with human and ape DNAs. We thank A. Ponge and A. Chedal for technical assistance and O. Séimonin at ESGS Cybergène for help in sequencing. We thank J. Strauss III for providing the b30 BcWo cell line and C. Leib Mösch for gift of the Lc5 cell line and of the pBl-W8 plasmid. S.P. is supported by a doctoral fellowship from the Ministère de la Recherche and Centre National de la Recherche Scientifique.