Diversity of human endogenous retrovirus class II-like sequences

Marie-Louise Andersson,1 Mats Lindeskog,1 Patrik Medstrand,1 Bruce Westley,2 Felicity May2 and Jonas Blomberg3

1University of Lund, Section of Virology, Department of Medical Microbiology, Solvegatan 23, S-223 62 Lund, Sweden
2University of Newcastle upon Tyne, Department of Molecular Pathology, Newcastle upon Tyne NE1 4LP, UK
3University of Uppsala, Section of Virology, Department of Infectious Diseases and Clinical Microbiology, Dag Hammarskjölds väg 17, S-752 37 Uppsala, Sweden

Class II human endogenous retroviruses (HERVs), often referred to as mouse mammary tumour virus (MMTV)-like or HERV-K elements, have similarities to several animal infectious retroviruses. Single clones from each of nine class II HERV groups (NMWV 1 to NMWV 9), isolated from a human breast cancer cell genomic library, were sequenced over a 244 bp stretch of the conserved reverse transcriptase region. These sequences were aligned to related exogenous and endogenous retroviruses and a phylogenetic tree was constructed. Sequences with more than 80% identity were considered as members of one group and we report here that the class II HERV family consists of at least ten groups. Three of the sequenced clones, from groups NMWV 3, 7 and 9, could not be related to any other previously identified elements and constituted their own groups. NMWV 8 had no similarity to any retroviral sequences in the sequenced region and is so far considered to be non-retroviral.

Human endogenous retroviruses (HERVs), with similarities to exogenous retroviruses known to cause disease in animals, constitute approximately 0.1–0.6% of the human genome (Leib-Mosch et al., 1990) and are probably relics of ancient infections and germline integrations. A large group of HERVs, the class II family (Callahan, 1988), are related to type A, B, D and avian type C retroviruses on the basis of their pol gene sequences. Several observations indicate that certain members of this family retain some of their original retroviral functions: sequences are expressed in several tissues and cell lines (reviewed in Wilkinson et al., 1994), some encode enzymatically functional proteins (reviewed in Wilkinson et al., 1994; Löwer et al., 1996; Kitamura et al., 1996), some encode particles released from teratocarcinoma cell lines, and gag proteins have been detected in the cytoplasm of human testicular tumours (reviewed in Löwer et al., 1996), and class II HERV transcripts have been identified in particles released from a breast cancer cell line (Seifarth et al., 1995). Moreover, a class II HERV was recently isolated from pancreatic islets and serum from type I diabetes patients (Conrad et al., 1997). The 5’ end of the env gene of this element was shown to encode an MHC II-dependent superantigen (SAG) and the authors suggest that expression of this endogenous SAG may lead to autoimmune cell destruction (Conrad et al., 1997).

Class II HERVs have been identified by various methods: by screening human genomic libraries with probes derived from mouse mammary tumour virus (MMTV) or mouse intracisternal A-particle (MIAP) (reviewed in Wilkinson et al., 1994), by using PCR (Medstrand & Blomberg, 1993; Seifarth et al., 1995; Conrad et al., 1997) or in association with characterization of other cellular genes (Yu, 1991). In two of the studies several distantly related clones were isolated and classified into different groups. First, nine NMWV groups (NMWV 1 to NMWV 9) were isolated by low-stringency hybridization of an MMTV gag–pol probe to a breast cancer cell genomic library (Franklin et al., 1988). Later, six human MMTV-like groups (HML-1 to HML-6) were detected in DNA and RNA from normal leukocytes by PCR and RT–PCR with primers from conserved regions of pol (Medstrand & Blomberg, 1993).

In this report we have attempted a comprehensive study of the diversity of class II HERVs by sequencing one clone from each NMWV group in a pol region corresponding to the region covered in the HML clones. DNA from six of the nine NMWV clones, belonging to groups NMWV 1, 3, 4, 5, 6 and 7, could be amplified using the primer pair ABDPOL–
Fig. 1. For legend see facing page.
ABDPOR (Medstrand et al., 1992), corresponding to conserved regions within the reverse transcriptase of type A, B, D and avian type C retroviruses. DNA from NMWV 9 could be amplified using the 3′ primer ABDPOR and another 5′ primer, RT64N (5′ GATGCGCCATGCCGTGTGGTGTTNMNCARTGGC, where R represents A or G and M represents A or C), derived from a position 106 bp upstream of ABDPOL. All fragments amplified by PCR were cloned into plasmid vectors using the TA Cloning kit (Invitrogen) or the pT7 Blue T-vector Kit (Novagen). The remaining clones, NMWV 2 and NMWV 8, had to be treated separately. A 2-3 kbp NMWV 2 EcoRI fragment hybridizing to clone HML-5.1 (Medstrand & Blomberg, 1993) and a 0-8 kbp NMWV 8 BamHI fragment that has previously been shown to hybridize to MMTV gag-pol (Franklin et al., 1988) were cloned into Bluescript vectors (Stratagene). All clones were sequenced in both directions with M13 primers by using the Sequenase kit, version 2.0 (USB), and separating the reaction products on 6% denaturing polyacrylamide gels, or by using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit and an ABI PRISM 310 Genetic Analyser (Perkin-Elmer). The 0-8 kbp fragment of NMWV 8 had no similarity to any retroviral structure and was considered to be non-retroviral. It was therefore omitted from further analysis. However, it had features of a CpG island (data not shown) and as these regions may occur in conjunction with retroviral elements we cannot exclude the possibility that some other part of this clone may contain retroviral sequences.

The NMWV pol sequences were aligned to corresponding regions of other class II HERVs and type A, B, D and avian type C retroviruses using the PCGENE programs (Intelli-genes). The following sequences were used in the alignment (Fig. 1a): HML-1 to HML-6 (Medstrand & Blomberg, 1993), HERV-K10 (Ono et al., 1986), HERV-K(C4) (Yu, 1991), ERV-MLN (Seifarth et al., 1995), an exogenous human retrovirus (HRV-5) (Griffiths et al., 1997), MIAp (Mietz et al., 1987), MMTV (Moore et al., 1987), Jaagsiekte retrovirus (JSRV) (York et al., 1992), Mason–Pfizer monkey virus (MPMV) (Sonigo et al., 1986) and Rous sarcoma virus (RSV) (Schwartz et al., 1983). An unrooted tree was constructed (Fig. 1b) by using the NEIGHBOR and DRAWTREE programs from the PHYLIP package, kindly provided by J. Felsenstein (Dept of Genetics, University of Washington, Seattle, WA, USA). Sequences with nucleotide sequence identities of more than 80% were considered as belonging to one group and we concluded that the class II HERV family consists of at least ten groups. The sequenced clone from group NMWV 4 was only 78.9% identical to one HML-6 clone, HML-6.3, but was identical to HML-6.1. According to Medstrand & Blomberg (1993) and Medstrand et al. (1997), the HML-6 clones constitutes a distinct group, despite the large variation between them. Sequence identities between the groups varied from 50.8 to 77.9%. The NMWV 1 clone was identical to HERV-K10 (Ono et al., 1986) in the sequenced region and fell into the same group as HML-2 (Medstrand & Blomberg, 1993) (Table 1). The clones from NMWV 2 and NMWV 5 grouped together with HML-5 and HML-3, respectively, and NMWV 6 belonged to the same group as HML-1 (identical to clone HML-1.1). Three of the sequenced groups, NMWV 3, 7 and 9, were not related to any other previously identified elements and constituted their own groups.

The taxonomy of HERVs is complex and different classification systems exist in parallel. In addition to the system based on identities in the pol sequence, another system exists based on the tRNA molecule used by retroviruses as a primer in replication. All class II HERVs so far sequenced in their primer binding site (PBS) regions have PBSS complementary to lysine (K) tRNA molecules and have accordingly been designated HERV-K. By naming the groups according to both the pol region and the tRNA molecule, we propose an integrated nomenclature system (Table 1). Using the BLAST program we searched GenBank with the pol fragments presented here and identified possible PBSSs. As the sequenced NMWV 1 clone was identical to HERV-K10, we assumed them to have identical PBSSs, i.e. corresponding to the lysine (CUU) anticodon. Two cosmid clones with regions similar to the NMWV 2 pol region were found in GenBank (accession nos Z955437 and U69569). However, they had large deletions and no PBSSs could be identified. The long terminal repeats (LTRs) of NMWV 4 have been sequenced (May & Westley, 1986) and searching GenBank with the 242 bp stretch of NMWV 4 pol identified a cosmid clone derived from chromosome 19 (accession no. AC003005), containing an identical pol region, that was 99% identical within the 426 bp LTRs. In addition, the two elements had almost identical PBSSs, differing in only one nucleotide, both corresponding to the lysine (UUU) tRNA, which differs from the (CUU) found in other elements. The clones from groups NMWV 5 and NMWV 6 both had similarities to cosmid clones in the sequenced region (NMWV 5 to accession nos AC002464 and Z83841) and NMWV 6 to

Fig. 1. (a) Alignment of nucleotide sequences from a conserved region of pol from the NMWV and HML groups, HERV-K10, HRV-5, ERV-MLN, HERV-K(C4), MIAp, MMTV, MPMV, JSRV and RSV. Sources of sequences are indicated in the text. Hyphens denote identities and dots denote deletions in the sequences. The length of each sequenced clone is indicated. The amino acid sequences translated from the nucleotide sequences in (a). All sequences are compared with HERV-K10. Hyphens indicate identical amino acids and dots indicate gaps. The upper row of asterisks denotes well-conserved sites (at least 15 of 17 amino acids identical) in HERVs (the NMWV and HML groups, HERV-K10, ERV-MLN and HERV-K(C4)) whereas the lower row of asterisks denotes conserved sites (6 of 6 amino acids identical) in five exogenous retroviruses, MMTV, MPMV, JSRV, RSV and HRV-5, and MIAp. Frameworks are indicated by a slash (/) and stop codons by X.
Table 1. Suggested classification of HERVs according to the sequence of a conserved region of pol

The table presents a way to combine the different nomenclatures; we hereby present this alternative to the HERV research community for consideration. **Not known.**

<table>
<thead>
<tr>
<th>Suggested group name</th>
<th>NMWV clones</th>
<th>HML clones</th>
<th>Other elements</th>
<th>PBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>HERV-K(10.HML-2)</td>
<td>NMWV 1</td>
<td>HML-2.1 to HML-2.6</td>
<td>HERV-K10, HLM-2, HM16</td>
<td>Lysine (CUU)†</td>
</tr>
<tr>
<td>HERV-K(HML-5)*</td>
<td>NMWV 2</td>
<td>HML-5.1</td>
<td>–</td>
<td>NK</td>
</tr>
<tr>
<td>HERV-K(NMWV3.HML-8)*</td>
<td>NMWV 3</td>
<td>–</td>
<td>–</td>
<td>NK</td>
</tr>
<tr>
<td>HERV-K(NMWV4.HML-6)</td>
<td>NMWV 4</td>
<td>HML-6.1 to HML-6.3</td>
<td>HERV-K(HML-6p), HERV-K(HML-6.17),</td>
<td>Lysine (UUU)‡</td>
</tr>
<tr>
<td>HERV-K(HML-3)</td>
<td>NMWV 5</td>
<td>HML-3.1 to HML-3.6</td>
<td>–</td>
<td>Lysine (CUU)‡</td>
</tr>
<tr>
<td>HERV-K(HML-1)</td>
<td>NMWV 6</td>
<td>HML-1.1 to HML-1.3</td>
<td>–</td>
<td>Lysine (CUU)‡</td>
</tr>
<tr>
<td>HERV-K(NMWV7.HML-7)*</td>
<td>NMWV 7</td>
<td>–</td>
<td>–</td>
<td>NK</td>
</tr>
<tr>
<td>HERV-K(NMWV9.HML-9)*</td>
<td>NMWV 9</td>
<td>–</td>
<td>–</td>
<td>NK</td>
</tr>
<tr>
<td>HERV-K(HML-4)*</td>
<td>–</td>
<td>HML-4.1</td>
<td>ERV-MLN</td>
<td>NK</td>
</tr>
<tr>
<td>HERV-K(C4.HML-10)</td>
<td>–</td>
<td>–</td>
<td>HERV-K(C4)</td>
<td>Lysine (CUU)</td>
</tr>
</tbody>
</table>

* HERV-K designation is tentative, as no PBS has been determined.
† According to the PBS of HERV-K10.
‡ Derived from FASTA searches of GenBank using the 244 bp pol fragment of NMWV.

Z70280) and the identified PBSs were complementary to lysine (CUU) tRNA. No structures similar to NMWV 3, 7 or 9 were found in GenBank.

To examine the sequences over a larger region we aligned six NMWV-related elements found in GenBank (accession nos AC002464, Z83745, Z95437, U69569, Z70280 and AC003100) to HERV-K10, ERV-MLN and three exogenous retroviruses over a 244 and an extended 823 bp region of pol.

That a longer sequence is preferable for accurate comparisons, this supports the use of the 244 bp stretch as a marker for classification of class II HERVs.

This region of pol contains 244 bp in exogenous type A, B, D and avian type C retroviruses and most HERVs share that length. HERVs differing from it usually contain frameshifts and/or stop codons. The clone from group NMWV 2 was 234 bp in length and contained one frameshift and one stop codon. The clones from groups NMWV 3 and 4, of 240 and 242 bp, respectively, both had frameshifts. The remaining five NMWV clones contained open reading frames and, on the basis of this short stretch, could possibly code for proteins (Fig. 1c). The sequences of five exogenous retroviruses, MMTV, MPMV, JSRV, RSV and HRV-5, and the MIAP sequence shared 30 conserved amino acids (Fig. 1c) in this region of pol, while the 17 endogenous retroviruses shared 29 conserved amino acids. Nineteen of these residues were shared by both endogenous and exogenous retroviruses.

We have studied class II HERVs on the basis of a conserved region of pol previously used for classification of HERVs (Medstrand & Blomberg, 1993). Elements from this family of HERVs have previously been detected in the genomes of great apes and Old World monkeys but not in more distantly related species, which would indicate an integration of this family into the primate genome about 30 million years ago (reviewed in Wilkinson et al., 1994). The large diversity reported here also indicates that it is an old family. Assuming the 5' and 3' LTRs of a retroviral element to be identical at the time of integration, their divergence can be used to determine the relative age of the insertion (Dangel et al., 1995). The divergence of class II HERV LTRs determined so far ranges from 10% for NMWV 4 (May & Westley, 1986) and 9-11% for HERV-K(C4) (Dangel...
et al., 1995) to 0.3% for HERV-K10 (Ono et al., 1986). By searching GenBank, we determined that two NMWV elements (accession nos Z83841 and AC002464) have degrees of LTR divergence between the two extremes: 4.9 and 7.3%. Ten per cent divergence suggests an insertion about 30 million years ago (Mager & Freeman, 1995) while 0.3% would indicate a considerably more recent event. The large differences in divergence between elements suggest a continuous integration of these elements into the primate/human genome. Whether these integrations resulted from exogenous infections or endogenous re-integrations is still unknown.

Both large RNAs as well as smaller transcripts have been detected for the class II family in both normal and malignant tissues (reviewed in Wilkinson et al., 1994). By searching the dBEST databases, we detected class II HERVs mainly in expressed sequence tags (ESTs) derived from neuroepithelial dBEST databases, we detected class II HERVs mainly in endogenous re-integrations is still unknown. The nature of transcriptional control of these elements is still unknown. However, HERV-K10 has a hormone-responsive element (HRE) and its expression is stimulated by female steroids (Ono et al., 1987). Motifs similar to an HRE has been found in other class II elements (Medstrand et al., 1997), which may partly explain their differential expression. This family of HERVs is interesting both from an evolutionary point of view and with respect to its possible involvement in normal biological functions or dysfunctions.

This work was supported by funds from the Medical Faculty and University Hospital of Lund, the Crafoord, Alfred Osterlund and Royal Physiographic Foundations, the Lars Hiertas Memorial Foundation and the European Commission, project GENE-CT930019.

References


Received 11 May 1998; Accepted 25 August 1998