Complete sequence of enzootic nasal tumor virus, a retrovirus associated with transmissible intranasal tumors of sheep

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
J Virol

Publisher Rights Statement:
Copyright © 1999, American Society for Microbiology

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Complete Sequence of Enzootic Nasal Tumor Virus, a Retrovirus Associated with Transmissible Intranasal Tumors of Sheep

CHRISTINA COUSENS,1,2* ESMERALDA MINGUION,3 ROBERT G. DALZIEL,2 AURORA ORTIN,3 MERCEDES GARCIA,4 JANE PARK,1 LORENZO GONZALEZ,4 J. MICHAEL SHARP,1 AND MARCELO DE LAS HERAS1

Moredun Research Institute1 and Royal (Dick) School of Veterinary Studies, University of Edinburgh,2 Edinburgh, United Kingdom, and Facultad de Veterinaria, Universidad de Zaragoza, Zaragoza3 and Servicio de Investigacion y Mejora Agraria, Derio,4 Spain

Received 9 September 1998/Accepted 3 February 1999

The sequence of the complete genome of ovine enzootic nasal tumor virus, an exogenous retrovirus associated exclusively with contagious intranasal tumors of sheep, was determined. The genome is 7,434 nucleotides long and exhibits a genetic organization characteristic of type B and D oncviruses. Enzootic nasal tumor virus is closely related to the Jaagsiekte sheep retrovirus and to sheep endogenous retroviruses.

Animal models of retrovirus-induced tumors of the lymphoreticular system are numerous and have contributed greatly to our understanding of oncogenesis (reviewed in reference 10). However, animal models of retrovirus-associated epithelial tumors are restricted to mouse mammary tumors, sheep pulmonary adenomatosis (SPA), and enzootic nasal tumors (ENT) of sheep and goats. Of these, ENT and SPA are also important for economic and animal welfare considerations in affected flocks. Both have long incubation periods, but once clinical signs become apparent the diseases are invariably fatal.

There is compelling evidence that SPA (also known as Jaagsiekte ovine pulmonary carcinoma) is caused by Jaagsiekte sheep retrovirus (JSRV) (reviewed in reference 15), although classical approaches to investigate the presence of an etiological role for this virus continue to be hampered by the lack of a cell culture system to propagate JSRV. The relationship between ENT and its associated retrovirus, enzootic nasal tumor virus (ENTV), is less well characterized. The presence of the retrovirus was first shown by electron microscopy in intranasal tumors of sheep (7, 23, 37) and goats (4, 5, 35) and in nasal fluid produced by these tumors. Antigenic cross-reactivity between ENT and its associated retrovirus, enzootic nasal tumor virus (ENTV), is less well characterized. The presence of the retrovirus was first shown by electron microscopy in intranasal tumors of sheep (7, 23, 37) and goats (4, 5, 35) and in nasal fluid produced by these tumors. Antigenic cross-reactivity between ENTV and JSRV, SERVs, and other retroviruses. This was finally resolved by the identification of small areas of sequence difference between JSRV and ENTV (1, 26). A similar approach is required to develop sensitive ENTV-specific detection techniques in order to elucidate the role of ENTV in disease. This is particularly important because ENTV, like JSRV, does not replicate in tissue culture.

Here we report the cloning and sequencing of the complete ENTV genome from an ovine case of ENT and investigate its relationship to JSRV, SERVs, and other retroviruses. This data is a vital step in the study of this animal model of epithelial cell-derived tumors and opens the way to further studies to elucidate the pathogenesis and oncogenic mechanisms of this virus.

MATERIALS AND METHODS

Source of material. Nasal fluid from a sheep with naturally occurring ENT was collected, clarified, and concentrated as described previously (6). The resultant pellet was resuspended as a 100-fold concentrate in TNE buffer (0.01 M Tris-Cl [pH 7.5], 0.1 M NaCl, 0.001 M EDTA). Samples of ENT and kidney were collected from this sheep at necropsy and snap frozen in liquid nitrogen. Tissue samples were stored at −70°C. DNA was extracted from these tissues by the method of Wu et al. (36). The sheep was verified to be negative for SPA by macroscopic and microscopic pathologic testing at necropsy.

Density gradient fractionation. ENTV was purified from 200 μl of 100-fold-concentrated nasal fluid by isopycnic centrifugation on 20 to 55% (wt/wt) sucrose gradients (16). Fractions (0.5 ml) were collected, and fractions with densities between 1.17 and 1.19 g/ml were pooled, diluted in TNE buffer and centrifuged at 100,000 × g for 1 h at 4°C. The pellet was resuspended in 0.5 ml of solution D (4 M guanidinium thiocyanate, 25 mM sodium citrate [pH 7.0], 0.5% [wt/vol] sarcosyl, 0.1 M 2-mercaptoethanol) for RNA preparation.

RNA preparation and reverse transcription. Total RNA was extracted from pooled sucrose gradient fractions by the acid guanidinium thiocyanate-phenol-chloroform method (2), without adding carrier RNA or treating with DNase. The RNA was resuspended in 15 μl of 1 mM EDTA. cDNA was synthesized by random-primed or oligo(dT)-primed reverse transcription with the Moloney murine leukemia virus (H-) ribocline cDNA synthesis system (Promega). The presence of ENTV cDNA was confirmed by gag-PCR (3).

Construction of cDNA libraries. EcoRI linkers were ligated to the cDNA by using the Ribocline EcoRI adapter ligation system (Promega). Excess adapters were removed from the ligation product with Sephadryl-400 as directed by the manufacturer. All of the cDNA-linker product was ligated to lambda arms by using the Zap II EcoRI-calf intestinal alkaline phosphatase cloning kit as described by the manufacturer (Stratagene). Gigapack II Gold packaging extract (Stratagene) was used for packaging. Recombinant clones were selected initially with JSRV clones as probes (pJS382 [38], pJS107 [38], and pLG1 [26]) and later by using ENTV clones from the library. The selected clones were excised by using plating strain SOLR, which converts lambda Zap II to Bluescript SK(−) plasmid and allows for blue-white screening for inserts.

PCR. Gaps in the sequence were filled by PCR with primers chosen from the available ENTV sequence data. The positions of the PCR fragments relative to the ENTV genome are shown in Fig. 1. In each case, PCR products were pooled from at least four independent reactions, gel purified, and cloned into pGEMT (Promega) or pCR2.1 (Invitrogen) as specified by the manufacturers. Product a was amplified from cDNA made from ENTV RNA prepared as described above. Products b, c, and d were obtained by amplification of ENTV provirus from nasal tumor DNA.

(i) PCR a. Product a was amplified from ENTV cDNA with primers ATGG GACGCCCTACAACCTGG (molecules [nt] 3113 to 3131 [Fig. 2]) and CACG AAGCTTCAAGGGAAGAAC (reverse of nt 3542 to 3572 [Fig. 2]). The PCR was carried out in 2.5 mM MgCl2–10 mM Tris (pH 9.2)–50 mM KCl–20 mM (NH4)2SO4–200 μM each deoxynucleoside triphosphate (dNTP)–0.125 μM each
A total of 39 clones were selected by screening of the ENTV cDNA libraries. Sequencing of these clones generated more than 75% of the ENTV genome. The remaining gaps, as identified by alignment with the JSRV genome, were generated by PCR (Fig. 1). Gap a in the pol gene was filled by PCR on cDNA from virus purified from nasal fluid from an animal with ENT. The sequencing gaps remaining at either end of the virus were amplified by PCR from nasal tumor genomic DNA, taking advantage of the fact that the long terminal repeat (LTR) is duplicated in the provirus (product b in Fig. 1). ENTV-specific U3 primers were designed by selecting regions of difference between the clones from the library and the published sequences of endogenous sheep retroviruses (1, 26) and used to amplify the region from U3 across U5 into gag. The LTR sequence obtained from these clones allowed the identification of a unique region in the ENTV LTR, from which more efficient exogenous specific PCR primers, PL3 and PL4, were chosen. These primers, together with primers from within pro or pol, respectively, allowed amplification from tumor DNA of products c and d, which included the two remaining sequence gaps. Control reactions, with kidney DNA from the same animal as the template, verified that the PCR amplified only exogenous ENTV and not endogenous sequences. We have previously reported the presence of two restriction enzyme sites, PsrI and AatII, within gag which are specific for ENTV (3). These occur at bases 1691 to 1696 and 1757 to 1762, respectively, in the full-length sequence of ENTV (Fig. 2). No JSRV or endogenous sequences that have these specific sites have been reported. Both these sites were present in all the clones spanning this region, which is further evidence that these clones were not derived from endogenous sequences. Furthermore, none of the clones contained the JSRV-specific Scal site in gag (26).

The complete sequence of ovine ENTV is shown in Fig. 2. The genome is 7,434 nt long. It has a genetic organization typical of type B and D retroviruses in that it has gag, pro, pol, and env genes (Fig. 3). No further open reading frames are present in ENTV. Comparison with other retroviruses suggest that pro and pol are expressed as gag-pro and gag-pro-pol precursors generated by ribosomal frameshifting and that env transcripts are generated by splicing.

**LTR.** The LTR of ENTV is 374 bases long and is bound by the inverted-repeat sequences GCAG (nt 122 to 125) and CTGC (nt 7172 to 7175) and flanked by the primer binding site (nt 127 to 142) and the polyuridine tract (nt 7154 to 7170). U3 is 249 bp, R is 13 bp, and U5 is 112 bp. U3 is defined as the region between R and the polypurine tract that contains the initiation site for synthesis of the positive strand of viral DNA. Like JSRV, ENTV uses tRNA1,2 initiator for synthesis of the positive strand of viral DNA. Although the ENTV LTR contains cis-acting sequences necessary for viral replication and regulatory signals for retroviral transcription. Although the ENTV LTR must interact with cellular factors, no recognized consensus sequences for transcription factor binding sites are identifiable. Experiments with expression constructs involving the LTR linked to a reporter gene must be performed before the functional regions of the ENTV LTR can be dissected.

**Leader sequence.** The untranslated region between the primer binding site and the start of gag is likely to contain the splice donor site for generation of subgenomic (env) transcripts, as well as the packaging signal which specifies incorporation of genome RNA into virions. Alignment with JSRV suggests that the splice donor site is after base 189.

**Viral proteins.** (i) gag. ENTV gag (nt 256 to 2094) encodes a 613-amino-acid (aa) polypeptide. The boundaries between the Gag proteins were assigned, as for JSRV (38), by alignment with MPMV, squirrel monkey retrovirus (SMRV), and mouse mammary tumor virus (MMTV) sequences. The ENTV capsid protein (CA) is encoded by nt 1027 to 1677, and the nucleic acid binding protein is encoded by nt 1678 to 1995. There are two sequences of Cys-X6-Cys-X7-His-X3-Cys, separated by 13 residues, in the nucleic acid binding protein. This is typical of a zinc finger domain and is thought to mediate nucleocapsid binding to the genomic RNA.

(ii) pro. The pro open reading frame (nt 1878 to 2858) encodes a 326-residue polypeptide. It is likely to be expressed
FIG. 2. Complete nucleotide sequence of the ENTV genome together with the deduced amino acid sequences of the open reading frames. The limits of U3, R, and U5 are indicated by arrows. The inverted repeats (IR), primer binding site (PBS), and polypurine tract (PPT) are indicated by underlines. The consensus protease cleavage site for env SU and TM is indicated by $\protease$.
FIG. 2—Continued.
as a gag-pro fusion polypeptide generated by ribosomal frameshifting, as has been shown for MMTV (17). The exact site and relative efficiency of this frameshifting event remain to be elucidated. The protein encoded by pro is thought to comprise two domains, a pseudoprotease (or protease-like) domain which exhibits a dUTPase activity, as has been suggested for other retroviruses by both computer analysis (22) and functional studies (9), and an active protease (Pr). The sequence Leu-Asp-Thr-Gly, indicative of the core amino acid sequence of cellular aspartyl protease, is encoded by nt 2469 to 2480.

(iii) pol. The pol open reading frame (nt 2825 to 5437) encodes a peptide of 870 aa giving a predicted molecular mass of 99 kDa. pol is likely to be translated as a fused gag-pro-pol polypeptide, again generated by ribosomal frameshifting. pol encodes the RT and integrase domains. The putative RT active site is the Tyr-Met-Asp-Asp sequence encoded by nt 3401 to 3412 in ENTV.

(iv) env. The env open reading frame (nt 5346 to 7196), at its start, overlaps pol, and the env transcript is likely to be generated by splicing. Like JSRV, ENTV contains two possible ATG start codons 7 aa apart, but the second is thought to be the initiation codon used since it is preceded by a predicted splice acceptor sequence. Retroviral env open reading frames encode a polypeptide precursor which is cleaved twice, first following the peptide signal sequence and second following an internal proteolytic cleavage site, which generates the surface (SU) and transmembrane (TM) envelope proteins. In ENTV, env encodes a polypeptide precursor of 617 aa (predicted molecular mass, 68 kDa). A hydrophobic region (encoded by nt 5526 to 5573) is thought to be the peptide signal sequence, and cleav-
TABLE 1. Percentage nucleotide identity and amino acid identity and similarity between ENTV and JSRV.

<table>
<thead>
<tr>
<th>Region</th>
<th>Nucleic acid similarity (%)</th>
<th>Amino acid similarity (%)</th>
<th>Amino acid identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gag</td>
<td>89.7</td>
<td>96.7</td>
<td>95.0</td>
</tr>
<tr>
<td>pro</td>
<td>95.2</td>
<td>97.2</td>
<td>96.3</td>
</tr>
<tr>
<td>pol</td>
<td>92.3</td>
<td>98.0</td>
<td>96.9</td>
</tr>
<tr>
<td>env</td>
<td>85.5</td>
<td>95.2</td>
<td>89.0</td>
</tr>
<tr>
<td>SU</td>
<td>87.9</td>
<td>96.3</td>
<td>93.4</td>
</tr>
<tr>
<td>TM</td>
<td>81.9</td>
<td>93.2</td>
<td>81.9</td>
</tr>
<tr>
<td>LTR</td>
<td>75.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U5</td>
<td>82.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U3</td>
<td>72.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The predicted molecular mass of the env proteins from the sequence data alone are 34 kDa for SU and 27 kDa for TM. However, the actual molecular masses will be larger due to glycosylation. The motif Asn-X-Ser/Thr, which can serve as a site for N-linked glycosylation, is found nine times in SU and twice in TM.

Comparison with JSRV. As shown in Table 1, the genes of ENTV and JSRV are very similar, with an overall amino acid similarity of greater than 95%. The biggest differences occur at the 3′ end of env and in the U3 LTR region. In addition, the open reading frame ORFX postulated for JSRV contains two stop codons in ENTV and therefore does not appear to be an open reading frame in ENTV.

The predicted molecular mass of the env proteins from the sequence data alone are 34 kDa for SU and 27 kDa for TM. However, the actual molecular masses will be larger due to glycosylation. The motif Asn-X-Ser/Thr, which can serve as a site for N-linked glycosylation, is found nine times in SU and twice in TM.

Comparison with JSRV. As shown in Table 1, the genes of ENTV and JSRV are very similar, with an overall amino acid similarity of greater than 95%. The biggest differences occur at the 3′ end of env and in the U3 LTR region. In addition, the open reading frame ORFX postulated for JSRV contains two stop codons in ENTV and therefore does not appear to be an open reading frame in ENTV.

Although ENTV and JSRV are distinct viruses associated conservative, suggesting that there are some functional restraints on mutation of this protein.

Alignment of the LTR of ENTV with that of JSRV shows that they are quite similar (75.7% similarity), although JSRV has an insertion of 21 bp in U3 compared to ENTV. Several regions contain a cluster of base differences. These regions are also the most divergent from SERVs and for this reason were chosen for PCR primers. Interestingly, ENTV does not appear to contain the same regulatory sequences as JSRV. For example, the NF-1 site (TGCGAAGCCCA) seen in JSRV (38) (Fig. 4, bases 44 to 55) is interrupted by the insertion of a T in ENTV (TGCGAAGCCCA). Also, the 19-bp tandem repeat identified in JSRV (38) is shorter in ENTV (15 bp). Finally, the C/EBP sequence (Fig. 4, bases 211 to 218) is not found in ENTV.

Comparison with other retroviruses. Construction of a phylogenetic tree based on comparison of gag genes of a range of different retroviruses (Fig. 5A) shows that ENTV and JSRV group most closely with MPMV. However, if env sequences are compared (Fig. 5B), ENTV and JSRV are most similar to MMTV. The tree shown is for SU, but comparison of env TM regions generates a very similar tree.) These results confirm that ENTV has a type B envelope and a type D capsid.

Construction of a phylogenetic tree based on LTR sequences (data not shown) shows that ENTV is most closely related to JSRV and SERVs. These have only approximately 43% identity to the type D retroviruses examined (simian SRV-1 type D retrovirus [SRV-1] and MPMV) and less than 33% identity to any of the other retroviruses listed in Fig. 5. Although this degree of identity is in the same range as that for comparison between the gag and env genes compared above, the LTR sequences from diverse retroviruses do not generate a reliable alignment and do not aid the identification of functional regions of the LTR.

DISCUSSION
This is the first report of the full sequence of ENTV. Determination of the sequence has enabled comparisons to be made between ENTV and other viruses. This confirms that ENTV is a chimeric type B/D retrovirus and that it is distinct from the related virus JSRV and from the endogenous retroviral sequences found in the sheep genome. Work currently under way suggests that ovine ENTV is also distinct from caprine ENTV (25a). It is hoped that similarities and differences between members of this family of viruses may give an insight into the mechanism of cellular transformation and factors determining cellular tropism.

Although ENTV and JSRV are distinct viruses associated
with discrete pathologic findings, translation and alignment of the open reading frames suggests a very high degree of homology between them (Table 1). The similarity between the coding regions of these viruses is in agreement with the evidence that antisera against CA proteins of ENTV and JSRV are cross-reactive. Antisera against other proteins have not been tested, but the degree of sequence similarity between JSRV and ENTV genes suggests that the proteins will all be antigenically cross-reactive.

Results for the few regions of SERVs that have been sequenced suggest that there is also a high degree of similarity between any putative expressed endogenous sequences and ENTV or JSRV. Although these JSRV and ENTV related endogenous sequences have not been linked to any pathologic findings, they may play an important role in allowing these viruses to persist in the host and eventually induce tumors. By reverse transcription-PCR, SERV transcripts have been detected in all sheep tissues tested, but protein has not been detected, probably due to the limited sensitivity of detection methods for proteins compared to nucleic acids. If these SERV sequences are expressed during ontogeny, it is likely that sheep would become immunotolerant to them and therefore to the related exogenous viral proteins, as has been hypothesized previously (25, 38). The surprisingly small number of sequence variations observed in our clones from the population of ENTVs in a single case of ENT supports the hypothesis that this virus is not subject to immune pressure. The earlier sequence data from two regions of ENTV gag (3, 29) also contains few changes from the ENTV sequence presented in this paper.

A putative open reading frame named ORFX was described in JSRV (29), and it has been suggested that this may encode an oncogenic product. In ENTV, the region homologous to ORFX contains two stop codons and therefore is not an open reading frame. Since it is likely that these two viruses have the same oncogenic mechanism, we therefore argue against ORFX playing any role in cellular transformation.

Apart from ORFX, the main differences between ENTV and JSRV occur in the LTR and the env TM subunit, suggesting that one or both of these regions must be important in determining viral tropism. The role of the cytoplasmic domain of env is thought to be important in envelope processing (21) and in ensuring the incorporation of the envelope glycoproteins into the budding virus (8, 12). It is the SU region of the retroviral envelope that interacts with the cell surface receptor and therefore is largely responsible for directing cell tropism at the surface binding level. Therefore, it is surprising to find that it is the TM and intracytoplasmic domains of env which differ the most between these two viruses. The change in the 3' end of env could be of little significance compared to that of U3 and may be a by-product of a single recombinational event in the 3' end of the ancestral type B/D retrovirus, which led to the divergence of JSRV and ENTV. However, it has been shown that alterations in the TM cytoplasmic domain can alter the pathogenicity of retroviruses (20, 39), and so a role for this domain in affecting the pathogenicity of ENTV and JSRV cannot be ruled out.

While a number of transcription factor binding sites are identifiable in JSRV, they could not be shown in the ENTV LTR. The differences in the LTRs of these two viruses are likely to affect the replication potential and therefore the pathogenicity of these viruses and may reflect the presence of different transcription factors in the cells that they infect. Sequences in the LTR that regulate transcription in specific cell types may also contribute to the ability of the retrovirus to transform these same cell types, as has been shown for murine leukemia virus (34). Thus, an understanding of transcriptional control of ENTV may also contribute to elucidating the mechanism of viral oncogenesis. We are currently characterizing the LTR function of these viruses.

In conclusion, we have presented, for the first time, the complete sequence of ENTV. Analysis of the sequence places this virus as another member of a group of B/D-type retroviruses, which also includes JSRV and SERVs. The disease pathology associated with ENTV has some similarities to that associated with JSRV in that both are tumors derived from
secretory epithelial cells. However, the diseases are clearly independent and affect different target organs. A virus species is defined as "a polythetic class of viruses that constitutes a replicating lineage and occupies a particular ecological niche" (International Committee on the Taxonomy of Viruses). Thus, it can be argued that JSRV and ENTV are distinct viral species.

The main sequence differences between JSRV and ENTV reside in ORFX, the U3 LTR, and the 3' end of env. Whether it is these differences and/or more subtle sequence differences that direct the tropism of these viruses requires further investigation.

The availability of the sequence of ENTV is a significant step that opens the way for the development of ENTV-specific detection techniques, which are vital to the further study of the distribution and pathogenesis of this virus and to the development of strategies to eradicate ENT.

ACKNOWLEDGMENTS

This work was supported by CEC contract AIR3-CT94-084, BBSRC grant RO6361, SOAFED grant ROAME MR1/027/95, and Spanish grant CICYT AGF96-0535-C02-01.

We thank I. D. Bennett for the sequencing and P. Harrison for critical discussion of the manuscript.

REFERENCES