Nucleotide sequences of the F, L and G protein genes of two non-A/non-B avian pneumoviruses (APV) reveal a novel APV subgroup

Marie-Hélène Bayon-Auboyer, Claire Arnauld, Didier Toquin and Nicolas Eterradossi

Unité de Virologie, Immunologie et Parasitologie Aviaires et Cunicoles and Laboratoire de Biologie Moléculaire, Agence Française de Sécurité Sanitaire des Aliments (AFSSA), BP 53, 22440 Ploufragan, France

Sequence analysis was performed of all or part of the genes encoding the fusion (F), polymerase (L) and attachment (G) proteins of two French non-A/non-B avian pneumovirus (APV) isolates (Fr/85/1 and Fr/85/2). The two isolates shared at least 99.7% nt and 99.0% aa sequence identity. Comparison with the F genes from subgroup A, subgroup B or Colorado APVs revealed nt and aa identities of 70.0–80.5% and 77.6–97.2%, respectively, with the L gene sharing 76.1% nt and 85.3% aa identity with that of a subgroup A isolate. The Fr/85/1 and Fr/85/2 G genes comprised 1185 nt, encoding a protein of 389 aa. Common features with subgroup A and subgroup B G proteins included an amino-terminal membrane anchor, a high serine and threonine content, conservation of cysteine residues and a single extracellular region of highly conserved sequence proposed to be the functional domain involved in virus attachment to cellular receptors. However, the Fr/85/1 and Fr/85/2 G sequences shared at best 56.6% nt and 31.2% aa identity with subgroup A and B APVs, whereas these isolates share 38.3% aa identity. Phylogenetic analysis of the F, G and L genes of pneumoviruses suggested that isolates Fr/85/1 and Fr/85/2 belong to a previously unrecognized APV subgroup, tentatively named D. G-based oligonucleotide primers were defined for the specific molecular identification of subgroup D. These are the first G protein sequences of non-A/non-B APVs to be determined.

Introduction

Turkey rhinotracheitis virus (TRTV) or avian pneumovirus (APV), the aetiological agent of turkey rhinotracheitis (TRT), is a member of the proposed new genus Metapneumovirus (Pringle, 1998) within the sub-family Pneumovirinae of the family Paramyxoviridae (Cavanagh & Barrett, 1988; Collins & Gough, 1988; Collins et al., 1986; Giraud et al., 1986; McDougall & Cook, 1986; Yu et al., 1991). APV was assigned to this genus because its negative-stranded RNA genome contains eight genes arranged in a different order from the ten genes of mammalian pneumoviruses such as the respiratory syncytial viruses (RSV) (Collins et al., 1990; Ling et al., 1992; Yu et al., 1992a, b; Randhawa et al., 1997). APV causes acute respiratory tract infections in turkeys, resulting in egg drop in breeder turkeys, and is responsible for major economic losses (reviewed in Naylor & Jones, 1993; Alexander, 1997).

One of the major surface glycoproteins of APV is the attachment (G) protein, which, by analogy with RSV, has been proposed to be responsible for virus binding to its cell receptor. The APV G gene and its predicted protein have several features in common with their RSV counterparts. Both G proteins are type II glycoproteins and both the RSV G (Gruber & Levine, 1985; Lambert, 1988; Satake et al., 1985; Wertz et al., 1985) and APV G (Cavanagh & Barrett, 1988) proteins are heavily O-glycosylated. The APV and RSV G genes show little nucleotide identity but their products have similar overall amino acid content, with high proportions of serine, threonine and proline residues, although differences also exist, as the RSV G is shorter than its APV counterpart and contains only four extracellular cysteine residues (Juhasz & Easton, 1994; Ling et al., 1992; Wertz et al., 1985). In both RSV and APV, the G protein is the most variable protein and is a major target for neutralizing antibodies.
In human RSV (HRSV), differences in reciprocal cross-neutralization and nucleotide sequencing of the G gene supported the definition of two HRSV subgroups, termed A and B (Anderson et al., 1985; Cane et al., 1991; Mufson et al., 1985; Sullender et al., 1991). The G proteins from HRSV subgroups A and B share 53% aa identity (Johnson et al., 1987). In TRTV, differences in neutralization patterns, cross-ELISA and reactivity with monoclonal antibodies also supported the definition of two antigenic groups (Collins et al., 1993; Cook et al., 1993; Hafez, 1992; Toquin et al., 1992). Nucleotide sequencing of the G gene led to the definition of two virus subgroups, which share only 38% aa identity. These subgroups were termed A and B by analogy with the HRSV nomenclature (Juhasz & Easton, 1994) and were shown recently to be consistent with the previously described antigenic groups (Bayon-Auboyer et al., 1999). Until late 1996, all known APV isolates belonged to either the A or the B subgroup. However, the first APV isolate from North America (Colorado strain, APV/CO) (Senne et al., 1997) was shown to cross-neutralize poorly with subgroup A and B viruses, suggesting that it may belong to a different serotype, or to a new subgroup proposed as C (Cook et al., 1998). Sequencing of the APV/CO M protein confirmed the unique nature of this isolate compared with European subgroup A and B viruses (Seal, 1998). Two other non-A/non-B APVs, isolates Fr/85/1 and Fr/85/2, were isolated in France in 1985. The G genes of these isolates could not be characterized by using specific RT–PCR procedures designed for the identification of APV subgroups A and B (Bayon-Auboyer et al., 1999) and the viruses were shown to be antigenically distinct both from subgroups A and B and from the APV/CO virus (Toquin et al., 2000).

In the present work, we determined the sequences of portions of the F and L genes of the Fr/85/1 and Fr/85/2 isolates and the entire sequences of their G genes. These sequences were compared with those of subgroup A and B and Colorado APVs. The results demonstrate that Fr/85/1 and Fr/85/2 are similar though genetically distinct from all prototype APVs, although the G proteins of Fr/85/1 and Fr/85/2 share a number of features with those of subgroup A and B viruses, including a short conserved extracellular stretch that may be critical for the function of the G protein. The analyses provide the first nucleotide sequence for the G protein of a non-A/non-B APV and define a novel APV subgroup, which we term subgroup D.

**Methods**

- **Viruses.** APV isolates Fr/85/1 and Fr/85/2 were isolated in France in 1985 from TRT-affected turkey flocks. No epidemiological relationship has been documented between the flocks. Propagation of the viruses and their preliminary antigenic and genomic characterization have been reported previously (Bayon-Auboyer et al., 1999; Toquin et al., 1999; Table 1).

- **Strategy for G gene amplification.** No oligonucleotide primers used to characterize subgroup A or B APVs were able to amplify the G gene of the highest dilution of serum resulting in complete neutralization of 100 mean TCID₅₀ of virus. Neutralizing titre (log₁₀) of the expected size.

**Table 1. Antigenic and genetic comparison of isolates Fr/85/1 and Fr/85/2 with subgroup A and B APVs**

<table>
<thead>
<tr>
<th>Virus*</th>
<th>Subgroup A</th>
<th>Subgroup B</th>
<th>Subgroup A</th>
<th>Subgroup B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fr/85/1</td>
<td>0.219 0.203</td>
<td>0.893 0.941</td>
<td>0.989 0.773</td>
<td>0.355 0.310</td>
</tr>
<tr>
<td>Fr/85/2</td>
<td>0.209 0.239</td>
<td>0.944 0.773</td>
<td>0.989 0.773</td>
<td>0.355 0.310</td>
</tr>
<tr>
<td>Fr/85/1/2</td>
<td>0.200 0.203</td>
<td>0.941 0.941</td>
<td>0.989 0.773</td>
<td>0.355 0.310</td>
</tr>
</tbody>
</table>

*Origin of virus strains as reported by Bayon-Auboyer et al. (1999).*

†Sera were prepared from 7-week-old specific-pathogen-free turkeys maintained under isolation, 17 days after nasal inoculation of the virus being studied. Mean results from at least five individual sera are shown.

‡ELISA reactivity measured as described by Bayon-Auboyer et al. (1999).§Neutralizing titre (log₁₀) of the highest dilution of serum resulting in complete neutralization of 100 mean TCID₅₀ of virus, calculated as described in Toquin et al. (1999).††Amplified product of the expected size.

- **Table 1. Antigenic and genetic comparison of isolates Fr/85/1 and Fr/85/2 with subgroup A and B APVs**

<table>
<thead>
<tr>
<th>Subgroup A</th>
<th>Subgroup B</th>
<th>Subgroup A</th>
<th>Subgroup B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fr/85/1/2</td>
<td>0.200 0.203</td>
<td>0.941 0.941</td>
<td>0.989 0.773</td>
</tr>
<tr>
<td>Fr/85/1</td>
<td>0.219 0.203</td>
<td>0.893 0.941</td>
<td>0.989 0.773</td>
</tr>
<tr>
<td>Fr/85/2</td>
<td>0.209 0.239</td>
<td>0.944 0.773</td>
<td>0.989 0.773</td>
</tr>
</tbody>
</table>
genes of isolates Fr/85/1 and Fr/85/2 (Bâyon-Auboyer et al., 1999). The G genes were amplified from APV full-length genomic or antigeneic RNA by using primers specific for flanking genes.

**Partial sequencing of the F and L genes.** RNA was extracted from the supernatant of APV-infected Vero cells by using the RNasey Mini kit (Qiagen). Primers Fd and Fr (Bâyon-Auboyer et al., 1999) were used for RT–PCR amplification of the F gene, as reported previously, with HotStart Taq DNA polymerase (Qiagen) at the PCR step. The resulting 365 bp product, corresponding to the region encoding amino acids 294–401 of F, was sequenced in both directions with an automated ABI 373 DNA sequencer, AmpliTag DNA polymerase FS, the ABI Prism Dye Terminator cycle sequencing kit (all from Perkin Elmer Applied Biosystems) and PCR primers. The partial sequence of the L gene (353 bp, amplifying the region encoding as 371–472) was determined similarly by using primers L+1168 (5′ TCCGACATCCTATGGTAAAGA 3′) and L−1520 (5′ TTGGGAGGAGATAATGCTATTACCTT 3′) (PCR annealing at 53.7 °C), which were selected following alignment of the published nucleotide sequences for APV strain C/UL/4.1 (accession no. AP06312) and HRSV strain A2 (accession no. P38887) in regions encoding the amino acid blocks that are conserved between the polyomers of negative-stranded RNA viruses (Tordo et al., 1988; Poch et al., 1990).

**Cloning and sequencing of the Fr/85/1 G gene.** Virus RNA was extracted as described previously. Primers specific for the Fr/85/1 G gene were selected in the F and L genes. A 4.2 kb RT–PCR product encompassing the Fr/85/1 G gene was generated by using a cDNA obtained with primer L−1457 (5′ AATTCTGTACAAACATTCAAG 3′) and the primer pair F +1255 (5′ GTGATACTATAACATAGA 3′) and L−1207 (5′ CATTCCATGCTTCTCTTC 3′) with the Expand High Fidelity PCR system (Boehringer Mannheim). The PCR included a precycle step at 94 °C for 2 min followed by 30 cycles of denaturation at 94 °C for 15 s, annealing for 60 s at 54 °C and extension at 68 °C for 4 min. For the last 20 cycles, the duration of the extension step was increased by 5 s with each cycle (e.g. cycle 11 has an additional 5 s, cycle 12 has an additional 10 s and so on). A final 10 min extension step at 72 °C was performed. The amplified DNA was cloned with the pMOSBlue blunt-ended cloning kit (Amerham Pharmacia Biotech). Bacterial colonies were screened by determining the size of the inserts, their orientation and the corresponding positions in the APV genome by PCR amplification and sequencing of the insert extremities with two vector-specific primers, T7 (5′ TAATACGACTCACTATAG 3′) and U19 (5′ GTTTCGACAGCTGA 3′). Three clones containing 2.7–3.8 kb inserts encompassing the full-length of the G gene were selected for sequencing by ‘genome walking’. Both positive and negative strands were sequenced in their entirety.

**Direct sequencing of Fr/85/1 and Fr/85/2 PCR products.** The sequence of the cloned Fr/85/1 gene was confirmed from three independent full-length PCR products derived from PCR primer pairs F+1321 (5′ ATCAaGGGaGACaACaCGTTGAAA 3′)/L−1207 (annealing at 54 °C) or SH +513 (5′ CAAaGATaTGCCGTCAAC 3′)/L−59 (5′ TAAaGTCCTGAAATGTTACC 3′) (annealing at 53.8 °C). The complete Fr/85/2 G gene sequence was determined from two full-length RT–PCR products obtained either with the L−1457 RT primer and the SH +433 (5′ AGACCTACTGAAAATGTT 3′)/L−146 (5′ GCTATAGGCCGTTACATTTAC 3′) PCR primers (annealing at 53.8 °C) or with the SH +433 RT primer and the SH +513/L−146 PCR primers.

**Submission of the sequences.** All of the sequences reported in this paper are from cDNA. The F, L and G sequences of Fr/85/1 and Fr/85/2 have been deposited in the EMBL, GenBank and DDBJ databases with the accession numbers AJ400728, AJ400729, AJ251085, AJ400730, AJ400731 and AJ288946, respectively.

**Analysis of nucleotide and amino acid sequences.** The nucleotide and deduced amino acid sequence data were compared by using the CLUSTAL W program (Thompson et al., 1994). Hydrophobicity profiles were determined by using the TopPred 2 software (von Heijne, 1992). Glycosylation and myristylation sites were predicted by using the Prosite motif database and the Sequence Motif Search Service available online at the Genome Net WWW server (Institute for Chemical Research, Kyoto University; http://www.MOTIF.genome.ad.jp) and by using the NetOGlyc 2.0 software available online at the Center for Biological Sequence Analysis (Department of Biotechnology, Technical University of Denmark; http://genome.cbs.dtu.dk).

**Phylogeny.** Alignment of the APV and mammalian pneumovirus nucleotide sequences was performed as described above. Sites including only gaps introduced by the outgroup (Newcastle disease virus, strain La Sota; accession no. AF077761) were eliminated from the resulting multiple alignment. The final DNA alignment was analysed with the PHYLIP package version 3.52c (Felsenstein, 1993). Briefly, the SEQBOOT program was run to generate 100 datasets that are randomly resampled versions of the previously aligned sequences. Each dataset was analysed by using the neighbour-joining method with Kimura’s modified two-parameter distance (NEIGHBOR program). The resulting 100 phylogenetic trees were used to compute a consensus tree according to the CONSENSE program with the ‘majority rule’ criteria. Bootstrap values on the consensus tree are percentages based on the 100 bootstrap iterations. In such trees, the lengths of branches have no particular meaning.

**Fr/85/1 and Fr/85/2-specific PCR primer pair.** The Fr/85/1-specific upper and lower primers G +50 (5′ GCGATGGCCGATTTAAGAAA 3′) and G−1005 (5′ CCCCTTAAACACGTCTT 3′) were designed on the G nucleotide regions that exhibit maximum heterogeneity between APVs. These primers were used in RT–PCR as described previously (Bâyon-Auboyer et al., 1999) with an annealing step at 57 °C. The viruses tested included 21 previously described APV isolates (Bâyon-Auboyer et al., 1999), the Colorado strain APV/CO (kindly provided by Dr D. Senne, National Veterinary Services Laboratories, USDA, Ames, IA, USA) and type 1 avian paramyxoviruses (aPMV-1) La Sota (Sotasec vaccine, Merial, France) and HB1 (HB1 Nobilis, Intervet), aPMV-2 Cal/Cal/Yucc/56, aPMV-3 Pk/Neths/449/75 and an influenza virus (H2N2/Ty/Brit/075/87). These viruses were propagated and prepared as described previously (Bâyon-Auboyer et al., 1999).

**Results**

**Partial sequencing of the F and L genes.**

A total of 323 nt of the F1 part of the F gene was sequenced in both Fr/85/1 and Fr/85/2 (Fig. 1). The two isolates were identical and comparison with the homologous regions in the APV isolates UK/3B/85 (subgroup A; Yu et al., 1991), APV France/PR8602/86 (subgroup B; accession no. Y14293) and Colorado (non-A/non-B; accession no. AF085528) revealed 73.7, 80.5 and 70.0% nt identity, respectively. Comparison of the deduced 107 aa sequences from the same viruses revealed 89.7, 97.2 and 77.6% identity. The sequenced region encompassed seven cysteine residues and one potential glycosylation
A total of 306 nt was determined from the Fr/85/1 L gene (Fig. 1). The G199→A nucleotide change was the only change observed in Fr/85/2 (99.7% nt identity). It resulted in the change of an alanine to a threonine residue at aa 64 of the deduced amino acid sequence. The partial L genes of Fr/85/1 and Fr/85/2 shared 76–8% nt identity with the CVL/14.1 virus (accession no. APU65312). This corresponded to 85.3% and 84.3% aa identity.

Sequence of the Fr/85/1 and Fr/85/2 G genes

The Fr/85/1 virus G gene was 1185 nt in length from the first nucleotide of the gene start to the last nucleotide of the stop codon. This was six and 75 bases shorter, respectively,
A novel avian pneumovirus subgroup

Fig. 2. Alignment of the predicted amino acid sequence of the attachment protein G of APV strain Fr/85/1 with those of CVL/14.1 (subgroup A) and 2119 (subgroup B). The alignment was done by the CLUSTAL W method and has been presented so as to give the optimum alignment of the three amino acid sequences. Gaps (dots) introduced into the sequence to optimize the alignment are not taken into account in the arbitrary numbering. Proposed intracellular, transmembrane and extracellular domains are indicated above the sequences. Asterisks indicate the positions of conserved amino acid residues. The 15 aa conserved stretch is enclosed in an open box. Predicted sites of N-glycosylation are enclosed in shaded boxes, predicted sites of N-myristylation are underlined (also see Results). Conserved serine and threonine residues (potential sites of O-glycosylation) are indicated by filled circles (●), additional potential conserved sites of O-glycosylation due to S/T exchange are indicated by open circles (○) and the 19 conserved cysteines residues are indicated by filled diamonds (◆).

than the G genes of APV strains CVL/14.1 and 2119. The Fr/85/1 G gene showed 56-6 and 54-2% overall identity, respectively, to the G genes of strains CVL/14.1 and 2119. However, a 166 base region (nt 275–441 for Fr/85/1 and CVL/14.1, nt 278–444 for 2119) exhibited 74-3–79-0% nt identity.

In Fr/85/1, the SH–G and G–L non-coding regions (from the stop codon of the promotor-proximal gene to the start codon of the next) were 86 and 106 bases long, respectively. The SH–G gene junction included (i) 35 nt preceding the transcription termination/polyadenylation signal of the SH gene (AGTGAATAAAA), (ii) 24 nt corresponding to the true intergenic region, (iii) the G gene-start signal (GGGACCAAGT) and (iv) 6 nt preceding the first translated ATG codon. The transcription termination/polyadenylation signal of the G gene exhibited a 6 nt overlap with the G coding sequence (AGTTAATACAAA, with TAA corresponding to the G stop codon) and was followed by a 90 nt intergenic region preceding the GGGACCAGT start sequence of the L gene.

Over the same genomic region, the Fr/82/2 sequence differed by only two nucleotide changes (T1650 → C and G1050 → A) from the Fr/85/1 sequence.

Structure and properties of the Fr/85/1 and Fr/85/2 G proteins

The Fr/85/1 and Fr/85/2 genes contained a large open reading frame (ORF1) and a second, smaller ORF, ORF2. ORF1 encoded a 389 aa protein, shorter than the G proteins of CVL/14.1 (391 aa) and 2119 (414 aa) (Fig. 2). The ORF1 protein had a predicted molecular mass of 41-9 kDa and an isoelectric point of 8-9 at pH 7. Its predicted hydrophobicity profile included an amino-terminal hydrophilic region (aa 1–27) followed by a short hydrophobic area (aa 28–48), with the remaining carboxy-terminal sequence of the protein being...
Table 2. Amino acid sequence identity between the three putative domains of the G proteins of the APV isolates Fr/85/1 (non-A/non-B type), CVL/14.1 (type A) and 2119 (type B)

Percentage identities are shown. The domains of G are abbreviated as IC (intracellular), TM (transmembrane) and EC (extracellular). Numbers in parentheses are percentage nucleotide sequence identities.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>IC</th>
<th>TM</th>
<th>Overall</th>
<th>15 aa region</th>
<th>Entire ORF1</th>
<th>ORF2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fr/85/1–CVL/14.1</td>
<td>29-6</td>
<td>42-9</td>
<td>28-7</td>
<td>73-3 (77-8)</td>
<td>29-5 (56-6)</td>
<td>56-5</td>
</tr>
<tr>
<td>Fr/85/1–2119</td>
<td>32-3</td>
<td>36-0</td>
<td>29-8</td>
<td>93-3 (80-0)</td>
<td>31-2 (54-2)</td>
<td>55-2</td>
</tr>
<tr>
<td>CVL/14.1–2119</td>
<td>25-8</td>
<td>36-0</td>
<td>35-1</td>
<td>80-0 (91-1)</td>
<td>38-0 (57-0)</td>
<td>56-7</td>
</tr>
</tbody>
</table>

mainly hydrophilic with only a short hydrophobic stretch of 21 residues (aa 122–142). This overall organization was consistent with that of an anchored type II membrane protein and corresponded well with the regions in the G proteins of strains CVL/14.1 and 2119 proposed previously as cytoplasmic (aa 1–27 for CVL/14.1, 1–31 for 2119), transmembrane (aa 28–48 for CVL/14.1, 32–52 for 2119) and extracellular domains (Juhasz & Easton, 1994; Ling et al., 1992). The short hydrophobic stretch within the extracellular domain was conserved in strains CVL/14.1 (aa 119–139) and 2119 (aa 120–140), with the CVL/14.1 G protein containing an additional extracellular hydrophobic region (aa 200–220). The predicted amino acid sequence identities between the Fr/85/1 G protein and those of CVL/14.1 and 2119 are shown in Table 2. Within each of the three proposed domains, the maximum identity between Fr/85/1 and CVL/14.1 or 2119 was 32-3% (with 2119) in the putative cytoplasmic portion, 42-9% (with CVL/14.1) in the transmembrane region and 29-8% (with 2119) in the extracellular domain. However, a short region encompassing 15 residues of the extracellular domain was highly conserved (aa 128–142 for CVL/14.1 and Fr/85/1 and 129–143 for 2119), with the lowest identity being 73-3% (between Fr/85/1 and CVL/14.1).

The Fr/85/1 G protein had a high serine and threonine content (24%), similar to those of CVL/14.1 and 2119 (23-5% and 24-6%, respectively; Juhasz & Easton, 1994). Most (91-4%) of the serine and threonine residues were in the proposed extracellular domain of the Fr/85/1 G protein, with only 12 strictly conserved in the three aligned APVs, but 12 more possible O-glycosylation sites were conserved due to threonine-serine exchange (Fig. 2). The G protein of strain Fr/85/2 contained one more potential site for O-glycosylation than did that of strain Fr/85/1, due to the presence of a serine residue at position 212. The Fr/85/1 G protein contained 9-5% proline residues versus 6-6% in CVL/14.1 and 8-5% in 2119. Seven of these residues were conserved in the three APVs (Fig. 2). The Fr/85/1 G protein also contained three potential sites for N-glycosylation (Hubbard & Ivatt, 1981) compared with four and five in strains CVL/14.1 and 2119. These sites were not conserved in the three APVs, but were all located in the extracellular domain. Only one predicted N-myristylation site, located in the transmembrane domain (aa 38–43) of Fr/85/1, was conserved among the three APV isolates. As this conserved site is the nearest to the amino terminus of the protein, it is also the most likely to be utilized. Indeed, as myristylation occurs in amino-terminal glycine residues, other predicted sites imply post-translational cleavages that are unlikely to occur in the APV G extracellular domain. Finally, 20 cysteine residues were found in the Fr/85/1 G protein, of which 19 in the extracellular domain were conserved in the three APVs. Three of these residues were clustered in the highly conserved extracellular sequence of 15 amino acids.

Comparison of the Fr/85/1 and Fr/85/2 G protein sequences with those of mammalian pneumoviruses revealed that the RSV G proteins were at least 91 aa shorter than that of Fr/85/1. Pairwise and global alignments did not detect any region of statistically significant sequence relatedness between these proteins (the maximum identity, between Fr/85/1 and HRSV A, was only 17-4%) and the extracellular domain of the RSV G proteins did not contain the short conserved amino acid motif observed in the G proteins of the six APVs included in the phylogenetic study.

ORF2 in the G gene started at nt 215 and has the capacity to encode a 90 aa polypeptide. No gene start or gene end signals could be identified in its flanking nucleotide sequences.

Phylogeny

Fig. 3 shows phylogenetic trees generated from the F, G and L sequences. In both the F- and G-based consensus trees, the mammalian and avian pneumoviruses separated as distinct clusters with high bootstrap values ranging from 87 to 99%.
However, significant subclustering was apparent in both trees within the APV group. In the F-based tree, only the subgroup B APVs clustered together in 100% of the bootstrap-generated trees. The subgroup A, Colorado, Fr/85/1 and Fr/85/2 viruses were only weakly related to the subgroup B cluster, as demonstrated by bootstrap values lower than 50%. In the G-based consensus tree (Fig. 3), the atypical character of Fr/85/1 and Fr/85/2 was even more apparent, as these isolates branched off from other APVs in 100% of the bootstrap-generated trees. Subgroup B APVs again showed a close relationship (100% bootstrap). In the L-based consensus tree, mammalian pneumoviruses were closely related, but the APVs
appeared more heterogeneous than in the F- and G-based trees, as the Fr/85/1 and Fr/85/2 viruses, which clustered together in 100% of the trees, branched with subgroup A CVL/14.1 virus in only 56% of the bootstrap-generated trees.

**RT–PCR detection of Fr/85/1 and Fr/85/2**

Primers G + 50 and G−1005 amplified a 956 bp DNA fragment from both Fr/85/1 and Fr/85/2. The sequence of the PCR product was confirmed as identical to that of the previously sequenced Fr/85/1 and Fr/85/2 G genes. Neither the other APVs, including the Colorado virus, nor the non-APV respiratory viruses yielded any amplification product of the expected size with these primers, whereas all APVs were successfully detected by using previously reported N-based primers (data not shown; Bâyon-Auboyer et al., 1999).

**Discussion**

Previously published data indicated that the APV isolates Fr/85/1 and Fr/85/2 were antigenically and molecularly different from several previously described APVs (Bâyon-Auboyer et al., 1999; Toquin et al., 1999). The present study aimed to assess whether isolates Fr/85/1 and Fr/85/2 exhibited extensive variation of their G proteins and to establish their genetic relatedness to other avian and mammalian pneumoviruses.

The organization of the G genes of Fr/85/1 and Fr/85/2 is consistent with that found in other APVs. In the SH–G and G–L non-translated junctions, the Fr/85/1 and Fr/85/2 gene end sequences contain the sequence AGUUUA upstream of a poly(A) tract, as found in the transcription stop/polyadenylation signal of most APV genes (Jing et al., 1996; Ling et al., 1992; Yu et al., 1992a). The transcription stop/polyadenylation signal of the Fr/85/1 and Fr/85/2 G genes includes the UAA stop codon, as found in ORF2 of the M2 gene (Yu et al., 1992b) and in the G gene of CVL/14.1 (Ling et al., 1992). The intergenic regions preceding the gene start signals of the Fr/85/1 G and L genes are longer than their counterparts in CVL/14.1, which are 13 and 69 nt long, respectively (Yu et al., 1992a; Randhawa et al., 1996). This finding is similar to the low conservation of the non-translated sequences in the genome of HRSV isolates belonging to different subgroups, excluding the conserved gene start and end sequences (Johnson et al., 1987). The Fr/85/1 and Fr/85/2 G gene start signals (GGGACAAGT) and positions of the translation initiation codons are identical to those of most APV genes (Jing et al., 1996; Ling et al., 1992; Yu et al., 1991, 1992b). Interestingly, the gene start sequences of the Fr/85/1 and Fr/85/2 L genes (GGGACCAGT in both viruses) differ from the G gene start. Differences between the G and L gene starts have also been reported in strain CVL/14.1 (Randhawa et al., 1996); however, the Fr/85/1 and Fr/85/2 L gene starts differ from that of CVL/14.1 (AGGACCAAT) at two nucleotide positions. This could be an important difference between the two groups of isolates, as the L gene start sequence probably controls the expression of the polymerase and hence indirectly that of all other viral proteins.

Comparison of the G proteins of Fr/85/1 and Fr/85/2 encoded by ORF1 of the G gene with those of other APVs revealed at most 31.2% aa identity. However, all APV G proteins shared (i) a comparable length, (ii) an orientation typical of type II glycoproteins, (iii) a mucin-like composition (Jentoft, 1990), (iv) the potential for O- and N-glycosylation and for N-myristylation, (v) high conservation of cysteine residues that may contribute to protein secondary structure through the formation of disulphide bridges (Fig. 2) and (vi) a common hydrophobicity profile, with a single extracellular stretch of highly conserved hydrophobicity and sequence. This organization is similar to the structure of the G protein of mammalian RSV, the ectodomain of which contains a short central hydrophobic area located between two mucin-like regions (Langedijk et al., 1996). The hydrophobic region is highly conserved between HRSV strains belonging to the same subgroup and has been proposed to be involved in receptor binding (Cane et al., 1991; Johnson et al., 1987; Sullender et al., 1991). The secondary structure of this conserved area in RSV is constrained by four cysteine residues involved in two disulphide bridges, thus creating a loop-like structure that is critical for immunogenicity (Akerlind-Stopner et al., 1990; Trudel et al., 1991; Rueda et al., 1994; Langedijk et al., 1997; Simard et al., 1997). In contrast, the conserved 15 aa stretch of APV includes only three cysteines, but the 16 other extracellular conserved cysteines might also contribute to the secondary structure. As with HRSV, the 15 aa stretch is so highly conserved among APVs that it is tempting to speculate that it plays a role in one or more aspects of the function of the G protein, such as cell receptor-binding activity. Whether the extracellular cysteine residues and short hydrophobic region conserved in APVs are critical for binding properties or for antigenicity should be investigated further, as different structural requirements for attachment might represent another biological difference between the two closely related genera *Pneumovirus* and *Metapneumovirus*. In HRSV, antigenic changes have been shown to be related to mutations of the G protein such as changes in the protein-linked carbohydrates (Palomo et al., 1991) or changes in the G amino acid backbone itself due to premature termination mutations, to single amino acid substitutions, to frameshift mutations or to multiple A–G transitions that can lead to the loss of one or two cysteine residues in the conserved cysteine cluster (reviewed in Collins et al., 1996; Melero et al., 1997). Fig. 2 shows that premature termination and changes in conserved cysteine residues are not responsible for the atypical antigenicity of isolates Fr/85/1 and Fr/85/2, but some of the other mechanisms described above might be involved.

Interestingly, a second small ORF (ORF2) is present in the G coding sequence of the APVs. Although ORF2 has neither
gene start nor gene end signals, it is striking that it is more conserved than ORF1 (at least 55-2% aa identity versus at best 38% aa identity for ORF1). To our knowledge, there is no evidence that ORF2 encodes a functional protein, and such a conserved minor ORF has not been reported in the G gene of mammalian pneumoviruses.

Despite the absence of a documented epidemiological relationship between the corresponding TRT outbreaks, the high nucleotide and amino acid sequence identity (99-7 and 99-0%, respectively) of Fr/85/1 and Fr/85/2 suggests that they are two isolates of the same APV strain. The strategy used for the amplification of their G genes was based on partial sequencing of the F and L genes. The phylogenetic study can thus be based on three genes, although nucleotide differences in the G gene were originally considered as the reference for the genetic definition of APV subgroups (Juhaz & Easton, 1994). The phylogenetic analysis produced consistent results irrespective of the gene analysed; the Fr/85/1 and Fr/85/2 viruses branched off from other APVs in the three bootstrap-generated consensus trees. This was most apparent with the G and L genes, where the highest amino acid sequence identities to previously sequenced APVs were only 31-2% and 76-1%, respectively. These levels of identity are at least as low as those recognized between the G genes of APVs belonging the A and B subgroups (38% aa identity; Juhaz & Easton, 1994) or between the G genes of pneumoviruses from different mammalian species [the ovine RSV (ORSV), bovine RSV (BRSV) and HRSV strains tested exhibit 29-7-33-9% aa identity] or the L genes of pneumoviruses from different mammalian species, such as HRSV A and BRSV (91-2% aa identity). The genetically distinct character of isolates Fr/85/1 and Fr/85/2 was also supported by partial analysis of the F gene (F1 subunit). The F1 subunit has been reported to be among the most conserved regions of the F ectodomain, with 86 and 91 % nt and aa sequence identity, respectively, between APV subgroups A and B (Naylor et al., 1998). However, in the same region, Fr/85/1 and Fr/85/2 exhibited at best 80-5 % nt identity to other APVs (French strain PR8602/86). In the F-based tree, Fr/85/1 and Fr/85/2 consequently appeared to be as weakly related to the cluster formed by APV subgroup B strains as was the Colorado isolate (bootstrap values lower than 50%). Interestingly, Fr/85/1 and Fr/85/2 still share 97-2% aa identity with subgroup B APV France/PR/8602/86, an indication that most nucleotide changes between these viruses are synonymous. This could be explained by Fr/85/1 and Fr/85/2 having evolved from a distant ancestor common to subgroup B strains, if nucleotide changes in the sequenced F1 region do not undergo positive selection leading to the accumulation of non-synonymous changes, as observed in the variable regions of the HRSV G molecule (Cane et al., 1991; Garcia et al., 1994).

On the basis of differences in antigenicity and in the F, G and L genes, we propose that isolates Fr/85/1 and Fr/85/2 are divergent enough to represent a previously unrecognized fourth subgroup of APV, which we term D. Whether subgroup D still causes field infections is unknown. In both the UK and France, recent TRTV isolates have all been shown to belong to subgroup B (Naylor et al., 1997; Bâyon-Auboyer et al., 1999). The new subgroup D-specific PCR primers will allow more APV isolates to be screened for more precise subgroup assignment. It is apparent from the present study and from previous results (Juhaz & Easton, 1994; Bâyon-Auboyer et al., 1999; Toquin et al., 1999) that APV isolates in continental Europe in the mid 1980s included three distinct virus subgroups, with the potential genetic diversity of APV being higher, since the Colorado virus has been proposed as subgroup C. In HRSV subgroup A, an antigenic drift induced by immune pressure has been proposed to play an epidemiologically important role, allowing more recent virus lineages to replace older ones in successive epidemics (Cane & Pringle, 1995; Melero et al., 1997; Garcia et al., 1994). In TRT, clinical signs induced by subgroup A, B or D APV are controlled experimentally by either subgroup A or subgroup B live-attenuated vaccines (Eterradossi et al., 1995; Toquin et al., 1996, 1999). However, the efficacy of these vaccines in controlling virus-shedding by turkeys experiencing heterologous or homologous challenges and the extent to which the antigenic and genetic variabilities of APV are influenced by vaccine-induced immune pressure or contribute to the spread of the disease remain to be established. The relationships between APV antigenicity and genetic diversity will be understood better once the G gene of the North American Colorado isolate, which has been proposed as subgroup C, has been sequenced.

This study was supported by EU Feder Funds. The authors thank Dr N. Tordo (Institut Pasteur, Paris), Drs V. Jestin and G. Kuntz-Simon (AFSSA-Ploufragan) and Dr F. Tekaia (Institut Pasteur, Paris) for their help with L primers, F primers, pMOSBlue cloning kit and phylogenetic analysis, respectively. Dr A. Easton (University of Warwick) is gratefully acknowledged for his critical review and help with the manuscript.

References


A novel avian pneumovirus subgroup


Received 5 May 2000; Accepted 14 July 2000