Complete sequence of the G glycoprotein gene of avian metapneumovirus subgroup C and identification of a divergent domain in the predicted protein

Dhanasekaran Govindarajan, Abdul S. Yunus and Siba K. Samal

Virginia-Maryland Regional College of Veterinary Medicine, University of Maryland, 8075 Greenmead Drive, College Park, MD 20742, USA

The complete nucleotide sequences of the attachment glycoprotein (G) genes of three strains of avian metapneumovirus subgroup C (AMPV-C) were determined from the viral genomic and mRNAs. The G gene of AMPV-C was 1798 nt (1015 nt longer than previously reported) and the derived polypeptide had 585 aa. The deduced amino acid sequence of the predicted G protein of AMPV-C strain Colorado (AMPV-CO) showed 21–25 % amino acid identity to the G proteins of human metapneumoviruses, but only 14–16 % amino acid identity to those of other AMPV subgroups. The predicted G protein of AMPV-CO showed 98 and 81 % amino acid identity to those of AMPV-C strains Mn-1a and Mn-2a, respectively, indicating considerable sequence variation in the G proteins of AMPV-C isolates. Comparison of the G protein sequences of AMPV-CO and Mn-2a identified a highly divergent domain (48 % amino acid identity) at aa 300–450.

Avian metapneumovirus (AMPV) causes an acute respiratory disease in turkeys and is also associated with 'swollen head syndrome' in chickens (McDouggall & Cook, 1986; Wilding et al., 1986; Wyeth et al., 1987; Buys et al., 1989). AMPV subgroups A and B (AMPV-A and -B, respectively) were originally defined based on nucleotide sequence divergence in the attachment glycoprotein (G) gene (Juhasz & Easton, 1994) and antigenic differences (Toquin et al., 1992; Eterradossi et al., 1995; Bäyon-Auboyer et al., 1999). AMPV was first isolated in the USA in 1996 from commercial turkeys in Colorado (Kleven, 1997; Cook et al., 1999). The US isolates of AMPV were found to be genetically and antigenically different from AMPV-A and -B (Seal et al., 2000) and hence were designated AMPV-C. Subsequently, Bäyon-Auboyer et al. (2000) isolated a new subgroup of AMPV that was different from all three known subgroups and was tentatively designated AMPV-D. The mammalian counterpart of AMPV, known as human metapneumovirus (HMPV), was recently isolated from children suffering from acute respiratory infections (van den Hoogen et al., 2001). Interestingly, several reports have shown that AMPV-C has a closer resemblance to HMPV than to other AMPV subgroups (van den Hoogen et al., 2002; Yunus et al., 2003; Toquin et al., 2003; Govindarajan & Samal, 2004). AMPV is a member of the genus Metapneumovirus in the subfamily Pneumovirinae of the family Paramyxoviridae (Pringle, 1998). The virus contains a non-segmented, single-stranded, negative-sense RNA genome with the gene order 3′-N–P–M–F–M2–SH–G–L–tr-5′ (Ling & Pringle, 1988; Yu et al., 1992). The nucleotide sequences of all eight genes of AMPV-A have been determined (Randhawa et al., 1997 and references therein). All of the genes except for the L gene of AMPV-B have also been sequenced (Jacobs et al., 2003 and references therein). Nucleotide sequences of all eight genes of AMPV-C have also been reported (Seal, 1998; Seal et al., 2000; Dar et al., 2001, 2003; Toquin et al., 2003; Yunus et al., 2003; Govindarajan & Samal, 2004). However, there have been conflicting reports on the presence or absence of the SH gene and the length of the G gene of AMPV-C strain Colorado (AMPV-CO). Alvarez et al. (2003) and Jacobs et al. (2003) reported the absence of the SH gene, whilst we and others (Toquin et al., 2003; Yunus et al., 2003) have reported the nucleotide and deduced amino acid sequence of the SH gene of AMPV-CO. Similarly, the G gene of AMPV-CO was reported to be 1321 nt long with a predicted polypeptide of 435 aa (Alvarez et al., 2003; GenBank accession no. AY579780) or 783 nt long with a predicted polypeptide of 252 aa (Toquin et al., 2003; accession no. AJ457967). In both of these studies, G gene sequences were obtained from the genomic RNAs of AMPV-CO. In our study, G gene sequences were obtained by sequencing the genomic RNAs and mRNAs of three different strains of AMPV-C. Our results showed...
that the G gene of AMPV-C was 1798 nt long and that the derived polypeptide had 585 aa. Alignment of the G gene sequences of three strains of AMPV-C indicated the existence of considerable sequence variation in the G genes of AMPV-C isolates.

AMPV-CO was obtained from the National Veterinary Services Laboratory (Ames, IA, USA). AMPV-C strains Mn-1a and Mn-2a were kindly provided by Sagar M. Goyal (University of Minnesota, St Paul, MN, USA). Virus propagation, RNA extraction and reverse transcription (RT) reactions were performed as described elsewhere (Govindarajan & Samal, 2004). Briefly, virus-infected cells were scraped into the medium and lysed by three cycles of alternate freezing and thawing. After initial clarification at 3000 g for 15 min, PEG 8000 (Sigma) was added to the cell lysate to a concentration of 10% and the lysate was incubated for 4 h at 4°C. The virus was pelleted at 4500 g for 30 min at 4°C and viral genomic RNA was extracted from the virus pellet by using TRIzol reagent (Invitrogen). Viral mRNAs were isolated from infected Vero cell lysates by using a polyA Spin mRNA Isolation kit (New England Biolabs).

RT-PCR of AMPV genomic RNA was performed by using a G gene-specific positive-sense primer, G-513 (5'-CAC-AAGCAATAGCACAACCTGACAAACCAACACAC-3'), and an L gene-specific negative-sense primer, L-713 (5' -GGC-CCTTGTAAATTGCTTCTAA-3'); the primer sequences were obtained from the published G gene sequence (Toquin et al., 2003). Although a single, sharply defined amplification product was not obtained from this PCR, the obtained product was cloned and subsequently sequenced. Surprisingly, sequence analysis of numerous clones did not yield a single consensus sequence; instead, consensus sequences were obtained at the 5' and 3' ends of the cDNA clones, whereas the internal regions varied in length. As numerous attempts to determine the complete G gene sequence from the viral genomic RNA proved unsuccessful, we decided to use the G mRNA as a template to determine the complete G gene sequence.

All RT reactions of mRNAs isolated from virus-infected cells were performed by using a ProtoScript First Strand cDNA Synthesis kit (New England Biolabs). Three separate RT reactions were performed for each virus, using an oligo(dT) primer and two G gene-specific reverse primers, G-1589 (5'-CAGTGCCGTCCAAAAACAT-3') and G-1640 (5'-CATCATAAGCAACCGGCGC-3'), which were designed based on the sequence obtained from viral genomic RNA. PCR was performed with Takara LA Taq polymerase and GC buffer II (TakaRa), using primers G-513 and G-1589. The following cycle parameters were used in the PCR: initial denaturation at 94°C for 1 min, 30 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 2 min, and a final elongation step of 72°C for 5 min. This PCR yielded a single 1-1 kb product. RT-PCR of mRNAs isolated from uninfected cells by using the same primers did not yield any product. The entire RT-PCR was performed three times, each time with a new viral mRNA preparation, and each time a single product of 1.1 kb was amplified. These products were either sequenced directly or were cloned and subsequently sequenced on an ABI 3100 DNA sequencer (Applied Biosystems). All sequencing procedures were performed in the presence of 5% (v/v) DMSO or 1 M betaine (final concentration) in the sequencing reaction mixture. Analysis of all sequences was performed by using the DNASTAR software program. The sequences obtained from RT reactions with genomic and mRNAs were assembled to yield the complete sequence of the G gene. The following G gene sequences were obtained from GenBank: AMPV-A (L34032), AMPV-B (accession no. L34031), AMPV-D (AJ251085), HMPV00-1 (AF371337) and Canadian isolates of HMPV, CAN98-75 (AY297748) and CAN97-83 (AY297749).

The complete G gene of AMPV-CO was 1798 nt from the start to the end of the gene, encoding a predicted polypeptide of 585 aa. The length of the G gene determined by us was 1015 nt longer than that reported previously for the same AMPV-CO strain (Toquin et al., 2003). The G gene of AMPV-CO was significantly longer than those known for other metapneumoviruses (MPVs). It was 538–613 nt longer than the G genes of other AMPV subgroups and 1066–1087 nt longer those of HMPV. It is of interest to note that, although AMPV-CO showed genetic relatedness to HMPV with respect to most of its genes (van den Hoogen et al., 2002; Govindarajan & Samal, 2004), it differed greatly from HMPV with respect to the length and sequence of the G gene. To determine whether the lengths and sequences of the G genes were conserved among strains of AMPV-C, we cloned and sequenced the G genes of AMPV-C strains Mn-1a and Mn-2a. We observed that the gene length and the predicted protein length of G genes of Mn-1a and Mn-2a were exactly the same as those of AMPV-CO, but that considerable sequence variation existed between the two strains. The G gene of AMPV-CO, as observed for other MPVs, possessed the highly conserved gene-start (5'-GGGACAAGU-3', mRNA sense) and gene-end (5'-UAGUUAUUAAA-3') signals (Ling et al., 1992; Biacchesi et al., 2003). Apart from the major ORF, four potential secondary ORFs (ORF2, 146–1771 nt, 541 aa; ORF3, 155–1771 nt, 538 aa; ORF4, 167–1771 nt, 534 aa; and ORF5, 1312–1608 nt, 98 aa) were also observed in the G gene of AMPV-CO. The G genes of Mn-1a and Mn-2a also contained similar secondary ORFs. It would be interesting to determine whether these secondary ORFs yield additional protein products or secreted forms of the G protein.

The predicted molecular mass of the G protein of AMPV-CO was 58754 Da, having a net charge of 8-27 at neutral pH and an isoelectric point of 8.28. The G+C content of the G gene of AMPV-CO was 61 mol%, which was significantly higher than that of the G genes of other MPVs. The additional 1015 nt determined by us in the G gene of AMPV-CO had a much higher G+C content (73 mol%).
The amino acid composition of the deduced G protein of AMPV-CO was relatively similar to those of other AMPV G proteins (Table 1). The G protein of AMPV-CO contained 7.2% proline and 23.1% serine/threonine residues, a consistent feature of mucin-like glycoproteins. The predicted G proteins of the three AMPV-C strains contained three conserved, potential N-linked glycosylation sites (Fig. 1). Numerous sites for potential O-linked glycosylation were also observed on the G proteins of all AMPV-C strains. The G ORFs of the three AMPV-C strains contained 18 conserved cysteine residues, 17 of which were present in the extracellular domains. The G ORFs of

<table>
<thead>
<tr>
<th>Virus</th>
<th>Size of ORF (aa)</th>
<th>Molecular mass (kDa)</th>
<th>G+C content (mol%)</th>
<th>Proline (%)</th>
<th>Serine + threonine (%)</th>
<th>No. cysteine residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMPV-CO</td>
<td>585</td>
<td>58-8</td>
<td>61.7</td>
<td>7.2</td>
<td>23.1</td>
<td>18</td>
</tr>
<tr>
<td>Mn1a</td>
<td>585</td>
<td>58-3</td>
<td>62.9</td>
<td>7.2</td>
<td>22.2</td>
<td>18</td>
</tr>
<tr>
<td>Mn2a</td>
<td>585</td>
<td>61-6</td>
<td>54.0</td>
<td>7.7</td>
<td>27.4</td>
<td>18</td>
</tr>
<tr>
<td>AMPV-A</td>
<td>391</td>
<td>43-0</td>
<td>48.5</td>
<td>6.7</td>
<td>23.5</td>
<td>20</td>
</tr>
<tr>
<td>AMPV-B</td>
<td>414</td>
<td>44-6</td>
<td>49.5</td>
<td>8.5</td>
<td>24.6</td>
<td>20</td>
</tr>
<tr>
<td>AMPV-D</td>
<td>389</td>
<td>41-8</td>
<td>49.7</td>
<td>9.5</td>
<td>23.9</td>
<td>20</td>
</tr>
<tr>
<td>HMPV-001</td>
<td>236</td>
<td>25-8</td>
<td>47.0</td>
<td>8.5</td>
<td>33.9</td>
<td>1</td>
</tr>
<tr>
<td>CAN97-83</td>
<td>219</td>
<td>23-7</td>
<td>46-4</td>
<td>7.8</td>
<td>32.0</td>
<td>1</td>
</tr>
<tr>
<td>CAN98-75</td>
<td>236</td>
<td>25-5</td>
<td>46-6</td>
<td>5.5</td>
<td>33.9</td>
<td>2</td>
</tr>
</tbody>
</table>

Complete nucleotide sequence of the AMPV-C G gene
AMPV-A, -B and -D each contained 20 cysteine residues (Juhasz & Easton, 1994; Bäyon-Auboyer et al., 2000), whereas, remarkably, the G ORF of HMPV-001 contained only one cysteine residue (van den Hoogen et al., 2002).

Hydropathy analysis of the putative G proteins of AMPV-CO, Mn-1a and Mn-2a showed characteristics of an anchored, type II membrane glycoprotein. The predicted hydrophobic profile of AMPV-CO G protein included an amino-terminal, hydrophilic intracellular domain (aa 1–31), a hydrophobic transmembrane domain (aa 32–54) and a hydrophilic extracellular domain (aa 55–585). Although the G protein of AMPV-CO showed structural and biochemical features similar to those of the HMPVs, it showed only 23, 25 and 21% amino acid identity to the corresponding proteins of HMPV-001, CAN97-83 and CAN98-75, respectively. However, the levels of amino acid identity to the G proteins of other AMPV subgroups were still lower: 14% to those of AMPV-A and -B and 16% to that of AMPV-D. Among the US strains, the G protein of AMPV-CO showed 98 and 81% amino acid identity to those of Mn-1a and Mn-2a, respectively. The predicted G proteins of Mn-1a and Mn-2a exhibited 79% amino acid identity between themselves. Sequence alignment of the G genes of the three AMPV-C strains revealed that Mn-1a and Mn-2a possessed 21 and 195 nt substitutions, resulting in 11 and 110 aa changes, respectively, in their deduced proteins. The majority of the changes in the predicted G protein of Mn-2a lay between aa 300 and 450, thus forming a hypervariable region with only 48% amino acid identity (Fig. 1).

Careful examination of the nucleotide sequence of the G gene of AMPV-CO published by Alvarez et al. (2003) revealed that it contained partial sequences of the SH and G genes as a single G ORF, which accounts for the absence of the SH gene as reported by the authors. Our results also showed that all 783 nt of the G gene described by Toquin et al. (2003) were contained in the G gene sequence determined by us, but an additional 1015 nt insertion was identified between nt 752 and 753 of the G gene. These additional 1015 nt contained a higher G+C content (73 mol%) than the rest of the G gene (61 mol%). We believe that these 1015 nt probably contributed to the structural complexity in the genome that prevented successful and complete RT-PCR of the genomic RNA in vitro. In our study, the use of the smaller-sized G mRNA (compared with that of the genomic RNA), the direction of the RT reaction and the use of GC buffer II (TaKaRa) were probably responsible for our ability to amplify the secondary structure region of the G gene successfully. This method will be useful in sequencing the G genes of other AMPV-C isolates. Although mechanisms such as gene splicing and gene rearrangement are not common in paramyxoviruses, it is possible that such mechanisms could occur in AMPV-CO, accounting for the differences in the G gene sequences reported previously (Alvarez et al., 2003; Toquin et al., 2003). However, whether these mechanisms actually occur needs further investigation.

The G protein of pneumoviruses is the major viral glycoprotein that is involved in the attachment of viral particles to the host cells and is also one of the major protective antigens. Contradictory reports on the length of the G gene sequence of AMPV-CO have caused problems for some time and, thus, the availability of the correct and complete G gene sequence should be of considerable help in diagnosis, as well as in the development of control strategies against AMPV infection in turkeys.

Acknowledgements

We are thankful to Subbiah Elankumaran, Daniel Rockemann and Peter Savage for their excellent technical assistance. We also thank Ireen Dryburgh-Barry for proofreading the manuscript. Our work was supported by USDA grant no. 2003-02176.

References


Toquin, D., de Boisseson, C., Beven, V., Senne, D. A. & Eterradossi, N. (2003). Subgroup C avian metapneumovirus (MPV) and the recently isolated human MPV exhibit a common organization but have extensive sequence divergence in their putative SH and G genes. *J Gen Virol* 84, 2169–2178.


