Deduced amino acid sequence of the fusion glycoprotein of turkey rhinotracheitis virus has greater identity with that of human respiratory syncytial virus, a pneumovirus, than that of paramyxoviruses and morbilliviruses

Q. Yu, 1 P. J. Davis, 1 T. Barrett, 2 M. M. Binns, 1† M. E. G. Boursnell1‡ and D. Cavanagh 1*

1Institute for Animal Health, Division of Molecular Biology, Houghton Laboratory, Houghton, Huntingdon, Cambridgeshire PE17 2DA and 2Pirbright Laboratory, Pirbright, Woking, Surrey GU24 0NF, U.K.

The sequence of the fusion (F) glycoprotein of turkey rhinotracheitis virus (TRTV) has been deduced from cDNA clones derived from oligo(dT)-selected infected cell RNA. The protein consists of 538 amino acids, the F2 and F1 subunits containing 102 (including the F2–F1 connecting peptide RRRR) and 436 residues, respectively. Each subunit has one potential N-linked glycosylation site. The protein has 38 to 39% amino acid identity with the F protein of respiratory syncytial virus (Pneumovirus genus) but only about half that with members of the other two genera (Paramyxovirus and Morbillivirus) in the Paramyxoviridae family. This is the first sequence evidence to support the view that TRTV is a pneumovirus, the first avian member of the genus to be described.

Introduction

The disease turkey rhinotracheitis was first described in 1978 in South Africa (Buys et al., 1989) and was recognized shortly afterwards in Europe (McDougall & Cook, 1986). The causative agent has been isolated and shown to be a virus with a morphology similar to that of members of the Paramyxoviridae which have a single-stranded, non-segmented, negative-sense RNA genome (McDougall & Cook, 1986; Wyeth et al., 1986; Jones et al., 1986; Picault et al., 1987; Buys et al., 1989) and has been named turkey rhinotracheitis virus (TRTV). Analyses of the TRTV-induced proteins (Collins & Gough, 1988; Ling & Pringle, 1988, 1989; Cavanagh & Barrett, 1988) and RNAs (Cavanagh & Barrett, 1988) suggested that of the three genera in this family (Paramyxovirus, Morbillivirus and Pneumovirus), TRTV most closely resembled the Pneumovirus genus, of which the type species and most comprehensively studied member is human respiratory syncytial (RS) virus (Collins et al., 1984a, 1986).

In this communication we present sequence data that strongly support the view that TRTV is indeed a pneumovirus, the first avian pneumovirus to be described. To our knowledge this is the first report of sequence data for TRTV.

Methods

Virus and cells: The UK/3B/85 strain of TRTV (McDougall & Cook, 1986) was used after the following passage history: isolation and passage in chick embryo tracheal organ cultures, 24 passages in chick kidney cells, four passages in CV-1 cells (ATCC No. CCL 70; Flow Laboratories) followed by threefold plaque purification in CV-1 cells and two further passages to produce seed stocks. One plaque-purified clone (RI.1) was subsequently passaged in either CV-1 or Vero cells (ATCC No. CCL 81; Flow Laboratories) to produce working stocks.

Preparation of TRTV-induced RNAs: Ten 75 cm² flasks of Vero cells were inoculated with 1 ml of TRTV working stock. After 72 h at 37 °C actinomycin D (4 µg/ml; Sigma) was added and 24 h later the cells were lysed in the presence of 1250 units of RNasin (Anglian Biotec) and RNA extracted by the SDS/proteinase K method (Maniatis et al., 1982) or by the guanidinium isothiocyanate method as described by Chomczynski & Sacchi (1987), except that isopropanol was substituted by ethanol. Poly(A)+ RNA was then selected using Hybond mAP affinity paper as described by the manufacturers (Amersham).

Preparation of cDNA: cDNA synthesis was performed using a cDNA synthesis kit [oligo(dT) priming; Amersham] and the poly(A)+ RNA from approximately two 75 cm² flasks of infected cells. Clones F8, F14 and F17 were prepared from this cDNA. After sequencing of clone F17, additional cDNA was produced from poly(A)+ RNA using an F gene-specific oligonucleotide primer, corresponding to nucleotides 531 to 548 in Fig. 2.
Cloning. In the first phase of cloning the cDNA obtained by oligo(dT) priming was tailed with dCTP and annealed to PstI-cut (dg) -tailed pBR322 (BRL) (Maniatis et al., 1982). This material was used to transform Escherichia coli RR1, selection was made for tetracycline resistance and ampicillin-sensitive colonies were transferred to nitrocellulose filters (Binns et al., 1985). The colonies were probed with poly(A)-selected RNA (from approximately one flask of TRTV-infected Vero cells) which had been labelled with ^32^P]ATP using polynucleotide kinase. The probe was mixed with non-poly(A)-selected RNA extracted from five roller bottles of non-infected Vero cells to diminish the number of Vero cell clones detected. Hybridization was performed overnight at 65 °C in 3 x SSC, 1 × Denhardt's solution, 10 μg/ml herring sperm DNA and 0.2% SDS (Maniatis et al., 1982) followed by autoradiography on Kodak XAR X-ray film. Plasmid DNA was prepared from the colonies which bound the probe, dotted onto nitrocellulose filters in duplicate and used for hybrid selection of RNA (Barrett & Mahy, 1984) from in vivo ^32^P-labelled infected and mock-infected cells (Cavanagh & Barrett, 1988). Bound RNA was eluted and analysed in agarose gels. Subsequently plasmids containing presumptive TRTV-specific inserts were prepared for sequencing. In the second phase of cloning cDNA prepared using an F gene-specific oligonucleotide primer was cloned into the EcoRV-cut phagemid pBluescript KS+ (Northumbria Biologicals).

**Sequencing.** Clones F8, F14 and F17 (Fig. 1) were subcloned into the Smal site of M13mp10 either after restriction with RsaI or sonication followed by end repair. Transfection was with E. coli TG1. Plaques containing TRTV-specific inserts were identified by hybridization with a ^32^P-labelled, random hexanucleotide-primed (Pharmacia) probe made from the F clones. Suitable M13 clones were sequenced by the dideoxynucleotide chain termination method (Sanger et al., 1977). Clone F37 (Fig. 1), in pBluescript, was sequenced following the generation of a nested set of deletions. The phagemid was digested with BstXI and XbaI (Boehringer Mannheim) and deletions were then generated by incubating 30 μl (approximately 10 μg) of phagemid with 30 μl of 2 × exonuclease III buffer and 1 μl (100 units) of exonuclease III (Northumbria Biologicals) at 30 °C. Samples (12 μl) were removed at 45 to 60 s intervals into 12 μl of 10 mM-Tris-HCl, 10 mM-EDTA pH 8.0 on ice. After heating at 65 °C for 10 min 225 μl of S1 nuclease buffer (33 mM-sodium acetate, 50 mM-sodium chloride, 0.03 mM-Mg sulphate pH 4.6) containing 36 units of S1 nuclease (Boehringer Mannheim) was added for 30 min at 37 °C. After phenol-chloroform extraction and ethanol precipitation the mutants were religated using 4 units of T4 DNA ligase overnight. E. coli JM109 cells were transformed with the mutants and minipreparations of phagemid were subsequently used for sequencing (Murphy & Kavanagh, 1988). Sequences were analysed on a MicroVAX 3600 computer system using the program FASTP and the Wisconsin program GAP.

**Results**

**Cloning of the TRTV F protein mRNA**

Clone F8 (Fig. 1) was sequenced following its hybridization with RNA from infected but not mock-infected cells. Comparison, using the program FASTP, of the deduced amino acid sequence of F8 (0.75 kb) with a database of sequences of the proteins of members of the Paramyxoviridae family revealed high amino acid identity to the F protein of RS virus and lower but significant relatedness to the F proteins of other members of the family. Given this relationship with the F proteins of paramyxoviruses clone F8 was then used to select clones F14 (1.15 kb) and F17 (1.59 kb) (Fig. 1). Since F17 did not completely represent the F gene a further clone (F37, 0.72 kb) was generated using an oligonucleotide primer corresponding to nucleotides 531 to 548 (Fig. 2). Clone F37 was selected by hybridization with a DNA fragment, corresponding to a region near the beginning of the F gene, which was generated from clone F17 using PstI. 75% of the F gene sequence was determined from two or three clones, the whole sequence was obtained from both strands and 94% of bases were determined at least six times, the remaining 6% being determined three or four times.

**Sequence of the TRTV F protein mRNA**

The sequence encompassed by clones F37 and F17 consisted of 1636 nucleotides [excluding the poly(A) tail at the 3' end of the mRNA] which encoded a long open reading frame (ORF) of 538 amino acid residues (M, 58728) (Fig. 2). The next longest ORF, starting at nucleotide 800 in Fig. 2, contained only 16 amino acids. There were three in-frame translation stop codons within the 45 nucleotides immediately preceding the proposed AUG translation start codon. These observations strongly indicated that the AUG at positions 13 to 15 in Fig. 2 was the start of the F protein ORF. The long ORF had a number of features consistent with those described for the F proteins of all three genera within the Paramyxoviridae: (i) an N-terminal hydrophobic sequence (Fig. 2, residues 1 to 18) which would be expected to function as a signal sequence for membrane translocation of the pre-propolypeptide; (ii) a highly basic sequence RRRR, at which cleavage of the propolypeptide could occur to generate a small N-terminal F2 subunit of 102 amino acids (inclusive of RRRR) and a large F1 subunit of 436 residues; (iii) a hydrophobic N terminus for the F1 subunit (Fig. 2, residues 103 to 125) characteristic of the fusion-associated moiety of the F proteins of paramyxoviruses; (iv) a hydrophobic sequence of 22 residues near the C terminus of the polypeptide (Fig. 2, residues 491 to
Fig. 2. Sequence of the TRTV F mRNA and comparison of the deduced amino acid sequence of the pre-propoplypeptide with that of RS virus strain A2 (Collins et al., 1984b). The TRTV sequence has been presented so as to give the optimum alignment of the two amino acid sequences. The putative site at which cleavage into the N-terminal F2 and C-terminal F1 subunits occurs is shown by a filled arrowhead adjacent to the basic connecting peptide RRRR, which has been underlined twice, in the TRTV sequence. The hydrophobic sequence (putative signal sequence) at the N terminus of F2 and the hydrophobic sequences at the N terminus of F1 created after cleavage and near the C terminus (membrane-spanning sequence) have been underlined once. Potential glycosylation sites (N-X-T/S) are indicated by bold underlining and cysteine residues by filled squares. Where the two viruses possess exactly the same amino acid is indicated by (:). Where the amino acid differences are conservative is indicated by (.). A (-) shows padding inserted to improve the alignment.

513) characteristic of membrane-spanning sequences; (v) a 25 residue hydrophilic sequence at the C terminus (Fig. 2, residues 514 to 538). As discussed below, the TRTV F protein had sequence similarity with the F protein of many members of the Paramyxoviridae family.

Upstream (at positions 1 to 10 in Fig. 2) from the presumed translation start codon (nucleotides 13 to 15 in Fig. 2) was the sequence GGGACAAGUA which is very similar to the sequence GGGGCAAAAUA which is located at the start of the RS virus F mRNA and is largely conserved in other RS virus genes (Collins et al., 1984b). A few nucleotides upstream from the poly(A) tail was the sequence AGUUA; an identical sequence is located at the termini of the F mRNAs of RS virus and pneumonia virus of mice (PVM) (Collins et al., 1984b; Chambers et al., 1990). Immediately preceding the sequence GGGACAAGUA (nucleotides 1 to 10 in Fig. 2) was the sequence AGUCAUAAAAAAAU. The first five nucleotides, AGUCA, resemble the AGUUA sequence near the end of the TRTV, RS virus and PVM F mRNAs while AAAAAAA is the complement of the signal for polyadenylation (four to seven U residues) at the end of RS virus genes (Collins et al., 1986; Chambers et al., 1990). This suggests that the sequence AGU-CAUAAAAAAA marks the end of the gene which precedes the TRTV F gene and that the intergenic sequence contains only two bases (UU). These observations, combined with the characteristics of the protein encoded by the long ORF, strongly indicated that the 1642 nucleotides shown in Fig. 2 [and a poly(A) tail] constituted the mRNA sequence of the F glycoprotein of TRTV.

Discussion

Since previous observations (see Introduction for references) had suggested that TRTV was a member of the family Paramyxoviridae, the sequence of TRTV clone F8 was first compared with a database of Paramyxoviridae sequences. This search revealed high amino acid identity with the F gene of the A2 strain of RS virus (Collins et al., 1984b). Subsequent comparison of the whole F protein sequence using the Wisconsin program GAP with some manual alterations revealed 40% amino acid identity (38% without manual alteration) with the F protein of RS virus strain A2 (Fig. 2).

Comparison with the F proteins of five paramyxoviruses, Newcastle disease virus (NDV), Sendai virus, human and bovine parainfluenza type 3 viruses and...
mumps virus (Chambers et al., 1986; Shioda et al., 1986; Miura et al., 1985; Spriggs et al., 1986; Suzu et al., 1987; Waxham et al., 1987) and two morbilliviruses, measles virus and rinderpest virus (Richardson et al., 1986; Buckland et al., 1987; Tsukiyama et al., 1988) using the GAP program without manual alteration revealed amino acid identities in the range of 20 to 24% compared with 38 to 39% for subtype A and B RS viruses (Collins et al., 1984b; Elango et al., 1985; Baybutt & Pringle, 1987; Lopez et al., 1988; Johnson & Collins, 1988) (Table 1). When conservative amino acid differences (Dayhoff et al., 1972) are taken into account, the relatedness of the RS virus and TRTV F proteins increases to 85%.

Thus, the much greater similarity of the TRTV F protein to that of RS pneumovirus strengthens the view that TRTV is a member of the Pneumovirus genus. Another feature that links TRTV and RS virus is the distribution of cysteine residues. RS virus has a cluster of eight cysteines in a region of F1 where paramyxoviruses and morbilliviruses have only four (Morrison, 1988); TRTV also has eight residues at this site (Fig. 2). Overall, of the 17 cysteines present in F of TRTV, 16 were perfectly aligned with the 16 cysteines present in the RS virus F protein. Also aligned were the basic sequences, 

<table>
<thead>
<tr>
<th>Virus</th>
<th>Identical amino acids (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS (strain A2)</td>
<td>38</td>
</tr>
<tr>
<td>RS (strain B18537)</td>
<td>39</td>
</tr>
<tr>
<td>NDV</td>
<td>20</td>
</tr>
<tr>
<td>Sendai</td>
<td>24</td>
</tr>
<tr>
<td>Human parainfluenza</td>
<td>23</td>
</tr>
<tr>
<td>Bovine parainfluenza</td>
<td>22</td>
</tr>
<tr>
<td>Mumps</td>
<td>23</td>
</tr>
<tr>
<td>Measles</td>
<td>20</td>
</tr>
<tr>
<td>Rinderpest</td>
<td>20</td>
</tr>
</tbody>
</table>

shows that the putative fusion-inducing peptides of TRTV and SV5 have 16/23 residues (70%) in common, including a sequence of 17 residues of which 14 are identical (82%). In contrast, there is much lower similarity between SV5 and RS virus (Fig. 3).

One of the least conserved features relates to the number of glycosylation sites. Although both TRTV and RS virus have one glycan site in F1 they are situated over 100 residues apart. In RS virus F2 has four or five potential N-linked glycan sites whereas TRTV has only one site likely to be occupied by a glycan. A second site (NPS) exists within the N-terminal signal sequence. This location, plus the presence of a proline residue immediately following the asparagine residue suggests that this site is not glycosylated (Hubbard & Ivatt, 1981). The very similar incorporation of [3H]glucosamine into F1 and F2 of TRTV (Cavanagh & Barrett, 1988) supports the view that each subunit has one glycan.

In addition to fewer glycosylation sites, the F2 subunit of TRTV was shorter than that of RS virus; most notably there was a 22 residue deletion when compared with RS virus (Fig. 2). That this difference was not a cloning artefact was proven when clone F37 was generated using a specific oligonucleotide corresponding to nucleotides 531 to 548 in Fig. 2. This position was chosen for the primer so that the resulting clone would include the region of the TRTV F gene which, according to the sequencing of clone F17, had a large deletion in the F2 subunit when compared with the RS virus sequence.

Comparison with the mRNA start signals of RS virus genes indicates that the start signal for the TRTV F mRNA is GGGACAAGUA. This 11 nucleotide sequence has eight bases in common with the mRNA start signal of the RS virus F mRNA. In common with the F mRNAs of RS virus and SV5 (Paterson et al., 1984) the mRNA start signal is followed less than five nucleotides later by the translation start codon. This is in contrast to at least five paramyxoviruses (Chambers et al., 1986;
non-coding regions at either end of the mRNA and the translation start codon. The eight nucleotide sequence AGUUAUUU which precedes the poly(A) tract has seven residues in common with the mRNA termination signal of the RS virus F mRNA. Despite the great similarity in the composition of the mRNA termination signals, their location with respect to the translation stop codon is different. Whereas in the RS virus F mRNA there are 153 nucleotides between the two, in TRTV there is an overlap, i.e. the first base of the AGUUAUUU mRNA termination signal is also the third base of the UAA translation stop codon (Fig. 2).

In conclusion the data show that TRTV is very economical with respect to its F gene, having virtually no non-coding regions at either end of the mRNA and encoding a F2 subunit which is shorter than that of most members of the Paramyxoviridae family. The overall amino acid identity and the number and distribution of cysteine residues supports the inclusion of TRTV in the pneumovirus genus.

We are happy to acknowledge the excellent technical assistance of Rachel Purvis.

References


Veterinary Record 121, 135.
The nucleotide sequence of the mRNA encoding the fusion protein of measles virus (Edmonston strain): a comparison of fusion proteins from several different paramyxoviruses. 
Virology 155, 508–523.
Nucleic Acids Research 14, 1545–1563.

Fusion protein of avian pneumovirus TRTV

TSUKIYAMA, K., YOSHIIKAWA, Y. & YAMANOUCHI, K. (1988). Fusion glycoprotein (F) of rinderpest virus: entire nucleotide sequence of the mRNA, and several features of the F protein. 
Virology 164, 523–530.
Virology 159, 381–388.
Veterinary Record 119, 139.

(Received 24 August 1990; Accepted 12 October 1990)