Nucleotide and predicted amino acid sequence analysis of the ovine respiratory syncytial virus non-structural 1C and 1B genes and the small hydrophobic protein gene

H. Alansari* and L. N. D. Potgieter

Department of Environmental Practice, College of Veterinary Medicine, University of Tennessee, Knoxville, Tennessee 37901, U.S.A.

Respiratory syncytial virus (RSV) infects humans and cattle causing serious respiratory tract disease in both. The genome of the human and bovine RSV (HRSV and BRSV) codes for two non-structural and eight structural proteins. RSV has also been isolated from naturally infected sheep, but the genome of the ovine RSV (ORSV) has not been characterized and nor has the virus host range been identified. Here we report on the cloning and sequencing of the two non-structural 1C and 1B genes and of the small hydrophobic (SH) protein gene of the ORSV. The nucleotide identity of the ORSV 1C gene to those of subgroups A and B of HRSV was 70% and 65% respectively whereas the predicted amino acid identity was 68% and 69% respectively. The ORSV 1B gene had a 70% and 72% nucleotide identity with that of subgroups A and B of the HRSV respectively, and 79% predicted amino acid identity with both HRSV subgroups. The identity level of these two ORSV 1C and 1B genes to those of the BRSV could not be determined since these two BRSV genes have not been sequenced. The SH protein of RSV is a structural protein expressed on the surface of infected cells. In common with HRSV and BRSV subgroups, the ORSV SH had the three proposed domains with the C-terminal domain least conserved among the viruses. In the latter domain, the ORSV SH gene had a nucleotide identity of 68 to 69% and 47 to 51% with those of the BRSV and HRSV, respectively, and a predicted amino acid identity of 56% and 33 to 47% with those of BRSV and HRSV, respectively. Defining the genes and their products should help determine which genes and gene products are most suitable for use as diagnostic tools or as vaccine candidates.

Respiratory syncytial virus (RSV) is an important cause of respiratory tract disease in humans and cattle (Kimman & Westenbrink, 1990). The negative-sense single-stranded RNA genome is transcribed into two non-structural and eight structural proteins, and the genes encoding the 1C and 1B non-structural proteins are first in the order of viral transcription (Collins et al., 1984). The small hydrophobic (SH) protein of RSV, formerly known as the 1A protein, is a structural protein expressed on the surface of infected cells, and therefore is probably a target for the host's immune response to the virus. It has been proposed that the polypeptide predicted by the major open reading frame (ORF) of the SH gene contains three domains, the relatively conserved N-terminal and central hydrophobic domains, and a variable C-terminal domain. The SH protein may be oriented on infected cell surfaces so that the C-terminal domain occurs extracellularly, as suggested by the finding that antibodies against a synthetic peptide representing the 12 C-terminal amino acids bind to the SH protein when reacted with intact RSV-infected cells (Olmsted & Collins, 1989). Synthetic peptides representing amino acids 45 to 62 and 51 to 60 of strain A2 of HRSV SH protein contain epitopes recognized by RSV-specific T and B lymphocytes respectively (Nicholas et al., 1988). As with the G glycoprotein gene of RSV, although to a lesser degree, the SH gene and its encoded protein differ significantly between different human (HRSV) and bovine (BRSV) strains (Anderson et al., 1992; Collins et al., 1990; Samal & Zamora, 1991).

RSV has also been isolated from naturally infected sheep but the host range of this strain has not been identified nor has its molecular biology been characterized. In an initial effort to characterize the ORSV genome, we cloned and sequenced the two non-structural
base sequences and deduced amino acid sequences to
gene sequences of the two human subgroups, and that
ORSV SH gene and its encoded protein are distinct from
those of strains of other known RSV subgroups.

The ORSV (WSU 83-1578) was isolated originally
from a naturally infected sheep with rhinitis (LeaMaster
et al., 1983) and kindly provided by Dr Howard
Lehmkuhl (National Animal Disease Laboratory, Ames,
Iowa, U.S.A.) Total RNA was extracted from infected
cells as described (Chomczynski & Sacchi, 1987) and
polyadenylated mRNA was isolated with a column of
oligo(dT). A cDNA library was generated and PCR-
amplified by the 3' RACE protocol as described
(Frohman et al., 1989). A 17-mer oligo(dT) attached
to an adaptor at its 5' end was used for priming poly(A)
mRNA and synthesizing cDNA. The adaptor contained
the XhoI, SalI and ClaI restriction endonuclease
recognition sequences. Two primers were used for PCR
amplification: one had the nine nucleotides common to
most RSV mRNA 5' ends (GGGGCAAAT) and was
attached to an adaptor, consisting of the recognition
sequences for PstI and EcoRI endonucleases; the other
contained only the recognition sequences for XhoI, SalI
and ClaI. In addition, the two primers also had two extra
G residues each at their 5' ends.

The DNA products were cloned into a linearized
vector (pAMP; a derivative of pSPORT) according to
manufacturer's recommendations (CLONEAMP system;
BRL) and the recombinant plasmids were screened
by PCR (Goueli & Ahmed, 1991). The various inserts in
the recombinant plasmids were sequenced using the
dideoxynucleotide sequencing method (Sanger
et al., 1977) and the identity and similarity levels with other
known RSV gene sequences were determined using the
Wisconsin Genetics Computer Group program. Three
independently derived clones from three different PCR
reactions for each gene were used to determine the
consensus sequence.

The ORSV 1C gene was 524 bases long [excluding the
poly(A) tail], with 55 and 61 bases of non-coding
sequences at the 5' and 3' ends respectively, and only
four nucleotides shorter than the HRSV gene. The major
ORF of the ORSV 1C gene predicted a polypeptide of
136 amino acids with a calculated relative molecular
mass of 14960 compared to 139 amino acids and a
calculated \( M_r \) of 15290 for that of the HRSV strains. The
nucleotide identity of the ORSV 1C to those of human
subgroups A and B was 70% and 65% respectively and
the predicted amino acid identity was 68% and 69% respectively (Fig. 1). The nucleotide and amino acid

identity levels for this gene and its product between the
two human subgroups are 78% and 87% respectively
(Johnson & Collins, 1989).

The ORSV 1B gene was 490 nucleotides long excluding
the poly(A) tail, and only nine nucleotides shorter than the
Corresponding HRSV gene (Johnson & Collins, 1989). However, the lengths of the major ORF of the
HRSV and ORSV 1B genes were the same (Fig. 2). The
ORSV 1B gene major ORF predicted a polypeptide of
124 amino acids with a predicted \( M_r \) of 13640. The
ORSV 1B nucleotide identity was 70% and 72% to
those of HRSV subgroups A and B respectively and
shared a 79% amino acid identity with those from both
HRSV subgroups.

The BRSV 1C and 1B gene sequences have not been
reported and therefore we were unable to determine the identity of these ORSV genes to those of any BRSV
strain. It is plausible that the ORSV 1C and 1B genes
may be more similar to those of BRSV than to those of
HRSV. The glycoprotein G gene is the most divergent
among RSV strains. The ORSV G gene had greater
identity to that of the BRSV than to those of the HRSV
subgroup, but ORSV may constitute a ruminant sub-
group different from that containing BRSV strains based
on analysis of the gene and its encoded glycoprotein by
Western blotting, RNase A mismatch studies and base
sequencing (Alansari & Potgieter, 1993).
The ORSV SH gene was 464 bases long. The major ORF is predicted to encode a 73 amino acid polypeptide with an \( M_r \) of 8030. As in the human and bovine strains, the predicted polypeptide from the ORSV SH gene may have three domains, a conserved N-terminal (amino acids 1 to 13), a central hydrophobic (amino acids 14 to 41) and a C-terminal (residues 42 to the end) domain. The latter domain varied most among the strains (Fig. 3). They were at positions 2, 3 and 63 among the ORSV SH gene. The presence of these glycosylation sites may be a requirement for the polylactosaminoglycan modification of the glycosylated SH protein to produce the more complex glycosylated SH protein (Anderson et al., 1989), is also present at the same position in the ORSV gene and therefore might be used also to generate a similar protein in cells infected with ORSV.

Information on the antigenic identity and diversity among RSV strains is important for guiding vaccine development and might also provide insight into the structure, function and evolution of RSV and its gene products. The high level of strain identity of certain genes can be of diagnostic value by allowing their use as targets for probes to detect infection (Johnson & Collins, 1989). Whether the 1B and 1C genes constitute good candidates for identifying ruminant isolates awaits sequencing of BRSV 1C and 1B genes.

Table 1. Overall nucleotide (N) and amino acid (AA) identities of the SH protein between different RSV strains

<table>
<thead>
<tr>
<th>Strains compared*</th>
<th>N/AA identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRSV-A</td>
<td>HRSV-B</td>
</tr>
<tr>
<td>HRSV-A</td>
<td>BRSV 391-2</td>
</tr>
<tr>
<td>HRSV-A</td>
<td>BRSV A51908</td>
</tr>
<tr>
<td>HRSV-B</td>
<td>BRSV 391-2</td>
</tr>
<tr>
<td>HRSV-B</td>
<td>BRSV A51908</td>
</tr>
<tr>
<td>BRSV 391-2</td>
<td>BRSV A51908</td>
</tr>
<tr>
<td>ORSV</td>
<td>HRSV-A</td>
</tr>
<tr>
<td>ORSV</td>
<td>HRSV-B</td>
</tr>
<tr>
<td>ORSV</td>
<td>BRSV 391-2</td>
</tr>
<tr>
<td>ORSV</td>
<td>BRSV A51908</td>
</tr>
</tbody>
</table>

*HRSV-A and HRSV-B are human RSV subgroup A and B respectively. BRSV 391-2 and A51908 are bovine RSV strains. ORSV is ovine RSV WSU 83-1578.

Table 2. Nucleotide (N) and amino acid (AA) similarity/identity between ORSV SH and other RSV strains in different domains

<table>
<thead>
<tr>
<th>Domain*</th>
<th>BRSV (A51908)</th>
<th>BRSV (391-2)</th>
<th>HRSV-A (A2)</th>
<th>HRSV-B (18537)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>N</td>
<td>69</td>
<td>68</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>65/56</td>
<td>65/56</td>
<td>47/34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>47/34</td>
<td>33/25</td>
</tr>
<tr>
<td>CHD</td>
<td>N</td>
<td>78</td>
<td>74</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>82/71</td>
<td>78/64</td>
<td>82/68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>78/64</td>
<td></td>
</tr>
<tr>
<td>NTD</td>
<td>N</td>
<td>72</td>
<td>72</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>85/69</td>
<td>69/54</td>
<td>92/77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>92/84</td>
<td></td>
</tr>
</tbody>
</table>

*ED, Ectodomain; CHD, central hydrophobic domain; NTD, N-terminal domain.
† Similarity and identity are the same when nucleotides are compared.
References


JOHNSON, P. R. & COLLINS, P. L. (1989). The 3B (NS2), 1C (NS1) and N proteins of human respiratory syncytial virus (RSV) of antigenic subgroups A and B: sequence conservation and divergence within RSV genomic RNA. *Journal of General Virology* 70, 1539-1547.


(Received 16 July 1993; Accepted 21 September 1993)