Nucleotide sequence analysis of a matrix and small hydrophobic protein dicistronic mRNA of bovine respiratory syncytial virus demonstrates extensive sequence divergence of the small hydrophobic protein from that of human respiratory syncytial virus

Siba K. Samal* and Miguel Zamora

College of Veterinary Medicine, University of Maryland, College Park, Maryland 20742, U.S.A.

The nucleotide sequences of the matrix (M) and small hydrophobic (SH) proteins of bovine respiratory syncytial virus (BRSV) have been determined from a dicistronic mRNA. Comparison of these sequences with the corresponding published sequences of human respiratory syncytial virus (HRSV) revealed extensive overall homology at both the nucleotide and amino acid levels in the M protein, but low overall homology at both the nucleotide and amino acid levels in the SH protein. There was only 16 to 22% identity between the BRSV SH protein and the HRSV SH proteins at the C terminus. There were also an additional eight amino acids at the C terminus of BRSV. Despite the low level of identity, there were similarities in the predicted hydropathy profiles of BRSV and HRSV SH proteins. The transcription start and stop signals, which are conserved among HRSV mRNAs, were also identified in the M–SH dicistronic mRNA of BRSV. In addition, the intergenic sequence for the M–SH gene junction of BRSV was determined.

Bovine respiratory syncytial virus (BRSV) is an important aetiological agent of respiratory disease in cattle world-wide (Paccaud & Jacquier, 1970; Bohlender et al., 1982; Stott & Taylor, 1985; Collins et al., 1988). It is classified as a member of the pneumovirus genus of the Paramyxoviridae family, which also includes human respiratory syncytial virus (HRSV). Although BRSV and HRSV are closely related antigenically and cause similar diseases, they are different in their host range (Matumoto et al., 1974) and in their reactivity pattern to monoclonal antibodies (Mufson et al., 1985; Òrvell et al., 1987). No antigenic cross-reactivity was found between the attachment surface glycoprotein G of BRSV and that of HRSV (Ôrvell et al., 1987). It is known that, like that of HRSV, the genome of BRSV codes for 10 different species of mRNA. Nine HRSV mRNAs have been sequenced completely. Ten BRSV-specific proteins corresponding to 10 HRSV proteins also have been identified (Lerch et al., 1989; Mallipeddi et al., 1990).

To understand the relationship between HRSV and BRSV at the molecular level, we have undertaken cDNA cloning and nucleotide sequencing of different mRNA species of BRSV for comparison with published data for HRSV. Recently, nucleotide sequences for the attachment (G), fusion (F) and nucleocapsid (N) protein mRNAs of BRSV have been reported (Lerch et al., 1990, 1991; Samal et al., 1991). There is 30%, 80% and 93% identity at the amino acid level between the G, F and N proteins of BRSV and HRSV, respectively. In this communication, we present the nucleotide and predicted amino acid sequences for the matrix (M) protein and the small hydrophobic (SH) protein of BRSV and demonstrate that the SH proteins are highly divergent between BRSV and HRSV. In addition, we present the intergenic sequence (M–SH) from a dicistronic mRNA derived from readthrough of the M and SH protein genes.

Construction of a cDNA library derived from mRNA isolated from cells infected with the A51908 strain of BRSV has been described previously (Samal et al., 1991). Clones containing the M and SH mRNA nucleotide sequences were identified by in vitro translation of mRNAs obtained by hybrid selection of randomly selected cDNA clones (Ricciardi et al., 1979). The nucleotide sequence of the polytranscript mRNA coding for the M and SH proteins was derived from two independent clones, A564 and A22, by the dideoxynucleotide chain termination method (Sanger et al., 1977). Both strands of the region from nucleotides 975 to 1200 were sequenced. Clone A564 [1398 bp exclusive of the poly(A) tract] started at nucleotide 37 in Fig. 1, spanned the M–SH intergenic region and ended at the translation start codon.
Fig. 1. Alignment of the nucleotide and amino acid sequences of the M-SH mRNA of BRSV (A51908) with those of HRSV (A2) (Collins & Wertz, 1985; Satake & Venkatesan, 1984). Gaps, indicated by dashes, were used to maximize the sequence identity of HRSV (A2) with BRSV (A51908). Transcription start and stop consensus sequences are highlighted.
Table 1. Summary of the percentage identity, at both amino acid and nucleotide levels, of the M protein and SH protein from BRSV (A51908) and HRSV (A2 and 18537)

<table>
<thead>
<tr>
<th></th>
<th>(A51908)/(A2)</th>
<th>(A51908)/(A2)</th>
<th>(A2)/(18537)</th>
<th>(A2)/(18537)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M protein</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amino acid</td>
<td>89</td>
<td>38</td>
<td>36</td>
<td>76</td>
</tr>
<tr>
<td>Nucleotide</td>
<td>74</td>
<td>64</td>
<td>68</td>
<td>91</td>
</tr>
<tr>
<td>Overall</td>
<td>80</td>
<td>51</td>
<td>37</td>
<td>32</td>
</tr>
<tr>
<td>Coding region</td>
<td>51</td>
<td>22</td>
<td>32</td>
<td>54</td>
</tr>
<tr>
<td>3' end*</td>
<td>32</td>
<td>16</td>
<td>41</td>
<td>52</td>
</tr>
<tr>
<td><strong>SH protein</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amino acid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>50</td>
<td>51</td>
<td>49</td>
<td>78</td>
</tr>
<tr>
<td>Coding region</td>
<td>53</td>
<td>53</td>
<td>39</td>
<td>52</td>
</tr>
<tr>
<td>3' end*</td>
<td>52</td>
<td>22</td>
<td>41</td>
<td>54</td>
</tr>
<tr>
<td><strong>Intergenic region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nucleotide</td>
<td>37</td>
<td>37</td>
<td>32</td>
<td>54</td>
</tr>
</tbody>
</table>

* Non-coding region excluding the transcription start or stop consensus signals.
† Data from Collins et al. (1990).
‡ Amino end (amino acids 1 to 22).
§ Hydrophobic domain (amino acids 23 to 41).
¶ Ectodomain (amino acids 42 to 73).

The complete sequence of M–SH dicistronic mRNA is shown in Fig. 1. The M–SH dicistronic mRNA contains 1453 nucleotides, excluding the poly(A) tail at the 3' end. The dicistronic mRNA contains two major open reading frames (nucleotides 11 to 779 and 1056 to 1274) which encode the M and SH proteins, respectively, as they have a number of features common to the corresponding proteins of HRSV (Satake et al., 1984; Collins et al., 1985).

The nucleotide sequence corresponding to the BRSV M mRNA contains 938 nucleotides from the transcription start sequence to the stop sequence. The nucleotide sequence of the BRSV M mRNA has a 74% overall nucleotide identity with the previously published (Satake & Venkatesan, 1984) HRSV strain A2 mRNA sequence (Table 1, Fig. 1). The overall lengths are similar (938 and 948 nucleotides, respectively); the 3' non-translated region of the BRSV M mRNA is eight nucleotides shorter than the corresponding HRSV sequence. The start sequence of the BRSV M mRNA (nucleotides 1 to 10) contains a single nucleotide difference (NGGGCAAATA), excluding the 5'-terminal nucleotide indicated by 'N', compared with the HRSV M start signal (Satake & Venkatesan, 1984). The M mRNA consensus stop signal (Collins et al., 1986) was identified as the sequence from nucleotides 936 to 948. As in HRSV, there is no untranslated region, other than the start sequence, at the 5' end of the M mRNA of BRSV. The coding regions of M mRNA are more highly conserved than the 3' non-translated regions (80% compared to 51% identity, respectively; Table 1, Fig. 1). This is in agreement with the observation that non-translated regions are relatively non-conserved between viruses of the two HRSV subgroups, with the exception of conserved start and stop sequences (Johnson & Collins, 1988).

The deduced BRSV M protein (Fig. 1) contains 256 amino acids and has a calculated Mr, of 28713. This agrees closely with the apparent Mr, of 29000 determined earlier in SDS–PAGE (Mallipeddi et al., 1990). The BRSV M protein has 89% amino acid identity with the HRSV M protein; most of the differences are due to conservative amino acid substitutions. Computer analysis indicated that the M protein of BRSV is moderately basic and predicted a single hydrophobic region (residues 188 to 204). The same hydrophobic region also has been observed in the HRSV M protein (Satake & Venkatesan, 1984).

The nucleotide sequence corresponding to the BRSV SH mRNA is 566 nucleotides in length from the transcription start to stop sequences. The coding region of BRSV SH mRNA is 219 nucleotides in length (nucleotides 1057 to 1237) (Fig. 1), i.e. 24 nucleotides longer than the HRSV SH coding region. The start signal has been identified as nucleotides 975 to 983 by comparison with HRSV SH mRNA (Collins & Wertz, 1985; Collins et al., 1990). The 5' untranslated region is the same length (74 nucleotides) in BRSV and HRSV SH mRNAs. The 3' untranslated region is 153 nucleotides in length in BRSV SH mRNA compared to 121 nucleotides in HRSV SH mRNA. The nucleotide sequence identities between the SH mRNAs of BRSV and HRSV subgroups in coding, 5' untranslated and 3' untranslated regions are 53 to 60%, 39 to 51% and 32 to 41%, respectively (Table 1).

The BRSV SH mRNA has a major open reading frame with a predicted polypeptide product of 73 amino acids. The estimated Mr of the polypeptide is 8402.
The deduced amino acid sequence of the BRSV SH protein was aligned with the published amino acid sequences of the SH proteins of HRSV A2 strain (subgroup A) and HRSV 18537 strain (subgroup B) (Collins & Wertz, 1985; Collins et al., 1990) (Fig. 2). The BRSV SH protein contains an eight amino acid extension at the C-terminal end, compared with that of the SH proteins of the HRSV A2 and 18537 strains (Collins & Wertz, 1985; Collins et al., 1990).

Computer analysis predicted that the BRSV SH protein has a central hydrophobic core (amino acids 14 to 41) flanked by two lysine residues at positions 13 and 43 (Fig. 2). This feature is conserved in the SH proteins of HRSV A2 and 18537 strains (Collins et al., 1990; Olmsted & Collins, 1989). The BRSV SH protein hydrophobic core contains a potential membrane-spanning region (amino acids 20 to 40) similar to the one predicted for the SH proteins of HRSV A2 and 18537 strains (Collins & Wertz, 1985; Collins et al., 1990) (Fig. 3).

The amino acid identity between the BRSV SH protein and the HRSV SH proteins is surprisingly low (Table 1). The BRSV SH protein shares only 38% amino acid identity with the HRSV A2 SH protein and 36% amino acid identity with the HRSV 18537 SH protein. The amino acid identity between the N-terminal regions (amino acids 1 to 22) of the BRSV and HRSV SH proteins of either subgroup is significantly higher (>64%) than the identities between the central hydrophobic core (amino acids 23 to 41) (<37%) and the C-terminal regions (amino acids 42 to 73) (<22%) (Table 1). By contrast, the N-terminal and the hydrophobic domains of the A2 and 18537 strains of HRSV are well conserved and only the ectodomain is significantly different (Collins et al., 1990). It was interesting to note that the central region, residues 23 to 38, contains only one conserved residue in all three proteins; however, this region is flanked by highly conserved sequences. All negatively charged residues present in both HRSV SH proteins are conserved in the BRSV SH protein. In addition, positively charged residues Lys-13 and Lys-43 are also conserved in all three proteins (Fig. 2 and 3).

The HRSV A2 SH protein has been reported to exist in a truncated form (SHt) (Olmsted & Collins, 1989). The observation that the first AUG is in an unfavourable sequence context (Kozak, 1987) led to the suggestion that the truncated form originates by translation initiation at the second methionine (Met-23) in the sequence (Olm-
sted & Collins, 1989). This is consistent with the observation that the truncated form of HRSV SH protein (18537 strain) is present at an undetectable level, if at all, due to the favourable sequence context of the first AUG (Collins et al., 1990). It is interesting to note that the first AUG of the BRSV SH protein mRNA is also in a favourable sequence context (UUACAUGAAC). Therefore, it is reasonable to expect that truncated forms of the BRSV SH protein, in which internal AUG triplets in the mRNA act as translation initiation codons, are either present at a very low level or are not present at all.

The M–SH intergenic region of BRSV is 25 nucleotides in length, which is 16 nucleotides longer than the corresponding HRSV A2 sequence. The nucleotide sequence of the BRSV M–SH intergenic region has only 37% nucleotide identity with the published HRSV A2 M–SH intergenic sequence (Collins & Wertz, 1985). This is in agreement with the observation that the intergenic sequences between any two genes are relatively unconserved regions between viruses of the two HRSV subgroups (Johnson & Collins, 1988). The low sequence identity observed in the M–SH intergenic region between BRSV and HRSV suggests that the intergenic sequences may not have any functional significance and probably act merely as ‘bridges’ between genes as has been suggested previously (Johnson & Collins, 1988).

The BRSV M–SH dicistronic mRNA contains two additional open reading frames. One (nucleotides 111 to 266) codes for a protein of 52 amino acids and overlaps with the M gene. Similarly, in the HRSV A2 M gene, a second open reading frame encoding a protein of 75 amino acids has been found (Satake & Venkatesan, 1984). Since these two second open reading frames are conserved neither at the sequence level nor in their relative positions in the genome, they probably do not play any role in virus replication. Another additional open reading frame in the BRSV M–SH dicistronic mRNA begins at nucleotide 1271 and ends at nucleotide 1423, and codes for a protein of 51 amino acids. No similar protein has been described for HRSV.

Finally, in order to determine the frequency of dicistronic M–SH transcripts, we identified 19 cDNA clones which reacted with a radiolabelled probe generated by random primer extension of clone A564. The identity of these M–SH cDNA clones was determined by Southern blot analysis using end-labelled oligonucleotides corresponding to either the SH or M gene. Six of the 19 cDNA clones (32%) reacted with both the SH and the M oligonucleotide probes, suggesting dicistronic cDNAs. The occurrence of high frequency readthrough of M–SH transcripts has also been reported for HRSV (Collins & Wertz, 1983, 1985). Of the 19 cDNA clones, eight and five were monocistronic M and SH genes, respectively.

In conclusion, the work presented here describes the sequence of an M–SH dicistronic mRNA of BRSV A51908 strain, their encoded proteins and the M–SH intergenic region. A high degree of relatedness between the M proteins of BRSV and HRSV was observed. This homology of the amino acid sequence suggests that the conservation in the M proteins is probably necessary for RSV replication and morphogenesis. The amino acid sequence of the BRSV SH protein shows a low level of overall identity with the HRSV SH proteins; however, the hydropathy profiles of the two proteins are similar, suggesting similar structural characteristics. There is only 16 to 22% identity between the BRSV SH protein and the HRSV SH proteins in the C terminus. The high degree of divergence between the ectodomain of the BRSV SH protein and that of the HRSV SH protein agrees with the hypothesis of Collins et al. (1990) that extracellular antigens are subject to selective pressure from host immunity, resulting in a greater amount of amino acid changes.

We thank Dr Sashi B. Mohanty for his support and enthusiasm and Daniel Rockemann for excellent technical assistance. This study was supported by grants from Maryland Agricultural Experiment Station and USDA grant no. 89-34116-4855. This study has been approved as scientific article no. A6118 and contribution no. 89-34116-4855 of the Maryland Agricultural Experiment Station.

References


(Received 6 December 1990; Accepted 20 March 1991)