Sequence analysis of M2 mRNA of bovine respiratory syncytial virus obtained from an F–M2 dicistronic mRNA suggests structural homology with that of human respiratory syncytial virus

Miguel Zamora and Siba K. Samal*

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

The nucleotide sequences of the F and M2 mRNAs of strain A51908 of bovine respiratory syncytial virus (BRSV) were determined by sequencing cDNA of an intracellular dicistronic mRNA. Comparison of the F mRNA sequence with those of other BRSV strains showed that there was extensive sequence identity at both the nucleotide (95% identity) and amino acid (94% identity) levels. Alignment of the nucleotide and encoded amino acid sequences of M2 mRNA of BRSV with those of human respiratory syncytial virus (HRSV) M2 mRNA showed 69% identity at the nucleotide level and 80% identity at the amino acid level. The general features of BRSV F and M2 proteins are similar to those described previously for the HRSV proteins. The M2 mRNA of BRSV also contained a second internal, overlapping open reading frame (ORF) similar to one reported for HRSV. The predicted products of the second ORFs of BRSV and HRSV shared 43% amino acid identity. As described for HRSV, the 3'-terminal end of the M2 mRNA overlaps with the 5' end of the L gene by 68 nucleotides. The identity between the N-terminal regions of the L proteins of BRSV and HRSV is 75%. In addition, the intergenic sequence of the F–M2 gene junction of BRSV was determined.

Bovine respiratory syncytial virus (BRSV) belongs to the genus Pneumovirus in the family Paramyxoviridae, which also includes human respiratory syncytial virus (HRSV). BRSV is an important aetiological agent of respiratory tract disease in cattle (Bohlender et al., 1982; Collins et al., 1988; Paccaud & Jacquier, 1970; Stott & Taylor, 1985). It has been shown recently that BRSV and HRSV are similar in gene and protein composition, and are antigenically related at the level of the F, M, N and P proteins (Lerch et al., 1989; Mallipeddi et al., 1990), with the major antigenic difference between BRSV and HRSV being in the G protein (Lerch et al., 1989; Òrvell et al., 1987; Taylor et al., 1984). The genome of RSV, a negative-sense ssRNA molecule, encodes 10 genes: glycoproteins G (attachment protein) and F (fusion protein); phosphoprotein (P), major nucleocapsid (N) protein, matrix proteins M and M2, small hydrophobic (SH) protein, large (L) protein, and non-structural proteins 1B and 1C. The nucleotide sequence of all 10 mRNAs of HRSV has been determined (Galinski, 1991). However, the nucleotide and deduced amino acid sequences of only the G, F, N, M and SH genes of BRSV have been reported (Lerch et al., 1990, 1991; Samal & Zamora, 1991; Samal et al., 1991; Walravens et al., 1990). The amino acid sequences of these BRSV proteins have been compared to those of their counterparts in HRSV. The G and SH proteins showed very low similarity (30% and 38%, respectively), whereas the N, M and F proteins were highly conserved (80 to 93% identity).

To understand the relationship between HRSV and BRSV at the molecular level, we have undertaken the cDNA cloning and nucleotide sequencing of different mRNA species of BRSV for comparison with the corresponding published sequences of HRSV. To date, we have obtained the nucleotide sequence of the mRNAs encoding the N, M, SH (Samal et al., 1991; Samal & Zamora, 1991), P and G proteins (S. K. Mallipeddi, M. Zamora, M. K. Pastey & S. K. Samal, unpublished results). In this communication, we have extended our sequence analysis to the F and M2 proteins of BRSV (A51908 strain), and present for the first time the nucleotide and deduced amino acid sequences of the M2 protein gene of BRSV. In addition, the nucleotide sequences of the F–M2 intergenic region and the M2–L overlapping region were determined from an intracellular dicistronic mRNA.
Construction of a cDNA library derived from intracellular mRNAs isolated from cells infected with the A51908 strain of BRSV has been described (Mallipedi et al., 1990). The cDNA library was screened with an HRSV F gene cDNA clone (kindly provided by Dr P. L. Collins). One of the clones that reacted with the HRSV F gene, A162, had an insert of approximately 3000 bp and was selected for nucleotide sequencing. Clone A162 [2907 nucleotides exclusive of a poly(A) tail] contained the entire F and M2 coding regions. In addition, it contained the 5' and 3' untranslated regions of each gene plus the F–M2 intergenic region. To determine the frequency of F–M2 dicistronic mRNAs, clone A162 was used to screen a cDNA library, resulting in the isolation of six cDNA clones which contained only the F gene, as determined by nucleotide sequencing. This result suggests that both the M2 mRNA and F-M2 polytranscripts observed previously (Samal & Zamora, 1991) are very rare, in contrast to the high frequency of M–SH polymerase chain reaction (PCR) amplification of M2 mRNA using the oligonucleotide GCAAATATGT-N as the forward primer.

The complete nucleotide sequence [excluding the poly(A) tail] of the F–M2 dicistronic mRNA, and the deduced amino acid sequences for the F and M2 proteins (Lerch et al., 1991; Walravens et al., 1990). Nucleotides 14 to 1729 encode a 572 amino acid protein that was identified as the F protein by comparison with other BRSV F proteins (Lerch et al., 1991; Walravens et al., 1990). Nucleotides 1955 to 2512 encode a 572 residue protein that was identified as the M2 protein by comparison with other BRSV M2 proteins (Baybutt & Pringle, 1987; Collins & Wertz, 1985; Elango et al., 1985).

The nucleotide sequence corresponding to the F mRNA is 1892 nucleotides in length from the initiation to termination consensus sequences, which were identified at positions 1 to 9 and 1882 to 1892 (Fig. 1), respectively, by comparison with other BRSV mRNA sequences (Samal & Zamora, 1991; Samal et al., 1991). The 5'-terminal nucleotide was determined by primer extension and PCR amplification of the G–F intergenic region (unpublished results). The 5' untranslated region is identical to the sequence at the 5' end of the F mRNA of the 391-2 strain of BRSV (Lerch et al., 1991).

Alignment of the nucleotide sequences of BRSV strains A51908, 391-2 (Lerch et al., 1991) and RB94 (Walravens et al., 1990) showed 97% identity in the coding regions. This degree of identity is similar to that...
observed in the F genes of HRSV subgroup A (Lopez et al., 1988). It is interesting to note that within bovine strains the identity (97%) in the coding region of the F mRNA sequence is similar to that in the 3' end untranslated region (84%). The BRSV strains used in this comparison have different origins; A51908 and 391-2 are American strains isolated about 10 years apart, and RB-94 is an isolate from Europe. The same observation has been reported for the F mRNA of HRSV subgroup A. However, there was a significant decrease in identity in the coding and untranslated regions when F mRNAs from different antigenic groups were compared, e.g. BRSV and HRSV (subgroup A or B), and HRSV subgroup A and subgroup B. Interestingly, when the 3' non-coding regions of the P mRNAs of A51908 and FS-1 strains of BRSV were aligned, they were found to be 93% identical (unpublished results). It was thought previously that the untranslated regions of RSV mRNAs were very divergent due to a lack of selective pressure. This conclusion was drawn from comparisons made between strains from different antigenic subgroups (BRSV and HRSV, HRSV subgroup A and B). Although the amount of data is limited, the high degree of identity observed in the 3' non-coding regions of the F and P genes of BRSV suggests that separation into different antigenic subgroups occurred very early in their evolution, and since then BRSV strains have undergone very little variation. It will be interesting to see whether this observation also applies to HRSV strains.

The nucleotide sequence encoding the F protein of BRSV (A51908) contains a single ORF of 1716 nucleotides (nucleotides 14 to 1729) (Fig. 1). The predicted protein is 572 residues long, compared to 574 residues for the F proteins of BRSV strains RB94 (Walravens et al., 1990) and 391-2 (Lerch et al., 1991). Alignment of the amino acid sequences of the F proteins from BRSV strains A51908, RB94 and 391-2 showed an overall identity of 94.3%. The F1 region showed 96% identity among all three proteins, whereas the F2 fragment was 91% conserved. These values are very similar to those observed for the F proteins within the same antigenic subgroup of HRSV (Lopez et al., 1988).

The general features of the F protein of the A51908 strain of BRSV are identical to those of the F proteins of other BRSV strains (Lerch et al., 1991; Walravens et al., 1990). The only difference is that four (positions 27, 120, 498 and 567; Fig. 1) of the five possible N-glycosylation sites are well conserved in all three BRSV F proteins. The site located at position 70 is present only in strains A51908 and RB-94, not in strain 391-2. The C-terminal region of the F2 subunit (residues 100 to 129) has been described as a highly variable domain in HRSV strains, as well as between HRSV and BRSV (Walravens et al., 1990). However, this region is highly conserved among strains of BRSV. This observation is in contrast to the suggestion that this area might form a strain-specific epitope for HRSV strains (Johnson & Collins, 1988).

The nucleotide sequence corresponding to the BRSV M2 mRNA is 958 nucleotides long from the initiation to the termination consensus sequences. The coding region of the M2 mRNA extends between nucleotides 1955 and 2512 (Fig. 1). By comparison with the M2 mRNA sequence of HRSV (Collins & Wertz, 1985), the initiation and termination sequences were identified as nucleotides 1946 to 1954 and 2894 to 2905, respectively, and were found to be identical to those of HRSV M2 mRNA (Collins & Wertz, 1985). As in HRSV, there is a 5' untranslated region in the M2 mRNA of BRSV. The 3' untranslated region is 395 nucleotides in length, compared to 365 nucleotides in HRSV. The identity between BRSV and HRSV in the 3' non-coding region is 52%, significantly lower than the 78% identity calculated for the coding region (see below).

Nucleotides 1955 to 2512 (Fig. 1) of the dicistronic mRNA encode a protein that was identified as the M2 protein by comparison with the amino acid sequence of the M2 protein of HRSV (Collins et al., 1990; Collins & Wertz, 1985). The M2 protein of BRSV is 186 amino acids long, compared to 194 and 195 amino acids for the M2 proteins of HRSV subgroups A and B, respectively. The M2 proteins of BRSV and HRSV share 80% identity in amino acid sequence. The C-terminal region (residue 166 to the last residue) of the M2 protein of BRSV shows extensive divergence (33% identity) from that of HRSV subgroups A and B, not only in the amino acid sequence, but also in length (Fig. 2). However, the general features of the M2 proteins are conserved in both BRSV and HRSV. It has been shown recently that the M2 gene encodes the major target antigen for RSV-specific murine cytotoxic T lymphocytes (Nicholas et al., 1990; Openshaw et al., 1990). Based on the close similarity observed in the M2 proteins, it is reasonable to expect...
that the M2 proteins of BRSV and HRSV probably have similar roles.

Overlapping with the M2 coding region, a second ORF has been described in HRSV which encodes a polypeptide of 90 to 93 amino acids depending on the subgroup (Collins et al., 1990; Collins & Wertz, 1985). The M2 mRNA of BRSV also contains a second internal ORF, which overlaps with the amino acid sequence of the M2 protein by six residues and encodes a polypeptide of 95 amino acids. The polypeptide predicted to be encoded by the second ORF of BRSV shares 43% overall identity with that of HRSV subgroup A, and low (35%) identity with that of HRSV subgroup B (Fig. 3), assuming the first methionine codon is the translation initiation site. The overall identity between the second ORFs of subgroups A and B of HRSV is 62%. The hydrophilicity profiles of the polypeptides encoded by the second ORFs of BRSV and HRSV suggest that the polypeptides are significantly different (Fig. 4).

The intergenic sequence between the F and M2 genes extends from nucleotides 1893 to 1945 (Fig. 1), which is six nucleotides longer than the F-M2 intergenic sequence in HRSV. Alignment of the two sequences shows a very low degree of similarity (<25%), which is in agreement with the homology observed between other BRSV and HRSV intergenic sequences (Samal & Zamora, 1991).

The 5' end of the L gene of HRSV has been reported to overlap with the 3' end of the M2 gene (Collins et al., 1987). Overlapping has been shown to be a mechanism to regulate transcription of the L gene (Collins et al., 1987). As shown in Fig. 1, the proposed initiation sequence of the L gene of BRSV was found 67 nucleotides upstream of the termination sequence of the M2 gene. The overlap between the M2 and L genes of HRSV is 68 nucleotides (Collins et al., 1987), which suggests that M2-L overlap is a general mechanism for transcription regulation of the L gene in RSV. At the nucleotide level the identity in the overlapping region of BRSV and HRSV is 84%. The overlapping region encodes the first 20 amino acids of the L protein of BRSV (Fig. 1), which are 75% identical to that of HRSV. The proposed initiation sequence, 5' GGACAAAU 3' (mRNA sense), of the L gene of BRSV has two nucleotide differences compared to the consensus initiation sequence GGGGCAAAT (Collins et al., 1986). Both differences are localized within the four consecutive G residues: (i) the 5'-terminal G residue is missing in the initiation signal of the L gene of BRSV; (ii) the last G residue in the consensus sequences has been replaced by an A residue in the BRSV L gene initiation sequence, as described for HRSV (Collins et al., 1987). Since the 5' end of all the BRSV and HRSV mRNAs for which the 5'-terminal nucleotide has been determined experimentally begin with a G residue, it is
reasonable to assume that the 5′-terminal nucleotide of the L gene of BRSV is the first G residue of the proposed initiation sequence. These are the only two RSV genes reported that lack the canonical four G residues at the 5′ end of the mRNA.

In conclusion, the work presented here describes the nucleotide sequence of the F and M2 mRNA, and the F–M2 intergenic sequence of the A51908 strain of BRSV, plus the deduced amino acid sequences of the encoded proteins. In both cases, a high degree of similarity was observed when compared with the corresponding sequence of HRSV. The F protein of the A51908 strain of BRSV also showed a high degree of similarity to the F proteins of other BRSV strains. As described for HRSV, the 5′ end of the L gene of BRSV was found to overlap with the 3′ end of the M2 gene, suggesting that a similar transcription attenuation mechanism is used in BRSV.

We thank Dr Sashi B. Mohanty for his support and enthusiasm, Dr Peter L. Collins for providing a cDNA clone of the F gene of HRSV and Mr Daniel Rockemann for excellent technical assistance. This study was supported by grants from Maryland Agricultural Experiment Station and USDA grant no. 89-34116-4855. Approved as scientific article no. A6210, and contribution no. 8379 of the Maryland Agricultural Experiment Station.

References


(Received 27 August 1991; Accepted 19 November 1991)