Sequence variability of the glycoprotein gene of bovine respiratory syncytial virus

Sanjay K. Mallipeddi and Siba K. Samal*

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

Sequence variation in the attachment glycoprotein G of bovine respiratory syncytial virus (BRSV) was determined. The nucleotide sequences of the G mRNAs of the A51908, VC464 and FS-1 strains of BRSV were compared with the published sequence of the BRSV strain 391-2. Nucleotide sequence alignment showed that overall they are highly conserved, with 90 to 97% identity. In addition, the coding region of strain A51908 was longer by 18 nucleotides at the 3' end. An 84 to 95% level of identity was observed among the deduced amino acid sequences of the G proteins of BRSV strains. A maximum divergence of 19% was found when the extracellular domains of the G proteins were compared. The level of diversity would be consistent, by analogy to the human respiratory syncytial viruses, with these BRSV strains forming a single subgroup.

Bovine respiratory syncytial virus (BRSV) is an important cause of respiratory tract infections in cattle. It belongs to the genus Pneumovirus of the Paramyxoviridae, which also includes human respiratory syncytial virus (HRSV). The proteins of these two viruses are antigenically related, except for the antigenically distinct attachment (G) protein (Lerch et al., 1990). Sequence comparisons between proteins from BRSV and HRSV indicate that the G and SH proteins are least conserved (30% and 38% identity respectively; Lerch et al., 1990; Samal & Zamora, 1991). HRSV isolates are divided into two antigenic subgroups, A and B, on the basis of their reactions with monoclonal antibodies (MAbs) to the G protein (Anderson et al., 1985; Mufson et al., 1985; Gimenez et al., 1986). Sequencing studies have also shown that the G protein of HRSV is less conserved than other proteins between strains of the same or different subgroups (Johnson et al., 1987; Sullender et al., 1990; Cane et al., 1991).

BRSV strains have not been characterized by G-specific MAbs, nor have any sequence comparisons of the G proteins of BRSV strains been reported. Thus, BRSV is thought to be monotypic. To examine the variability of BRSV isolates, we compared the sequences of the G proteins of four strains.

Strains of BRSV used in this study were A51908, 391-2, FS-1, and VC464. Strain A51908 was isolated in Maryland in 1975 (Mohanty et al., 1975); strain 391-2 was isolated in North Carolina in 1985 (Fetrow et al., 1985), strain FS-1 was isolated in Iowa in 1975 (Smith et al., 1975) and strain VC464 was isolated in Missouri in 1974 (Rosenquist et al., 1974). Construction of cDNA libraries from mRNAs isolated from infected Madin Darby bovine kidney cells has been described (Samal et al., 1991). Recombinant cDNA clones containing G mRNA sequences of BRSV strain A51908 were identified by Northern blot hybridization and by comparing the sequences to the previously published G gene sequence of the BRSV strain 391-2 (Lerch et al., 1990). cDNA clones of G genes of BRSV strains FS-1 and VC464 were identified by the colony hybridization technique with a probe made from the G gene of BRSV strain A51908.

The nucleotide sequence of the mRNA encoding the G protein of each strain of BRSV was derived from at least two independent cDNA clones, using the dideoxynucleotide chain termination method (Sanger et al., 1977). The nucleotide sequence of the 5' end of the G mRNA of A51908 strain was determined by PCR amplification of the SH–G intergenic region (Zamora & Samal, 1992). The G mRNA sequences of BRSV strains FS-1 and VC464 were determined from nucleotide positions 37 and 42, respectively.

The G mRNA of BRSV strain A51908 contained 835 nucleotides excluding the poly(A) tail (Fig. 1). The G mRNA sequence of strain A51908 was compared with those of strains FS-1, VC464 and 391-2 (Lerch et al., 1990). The levels of identity among the four different G mRNA sequences were 90 to 97% (Table 1). There were three nucleotide deletions in the 3' non-coding region of all BRSV strains sequenced compared to the corresponding region of strain 391-2. In strains 391-2, FS-1 and VC464, the stop codon was at positions 787 to 789,
whereas in strain A51908 it was at positions 805 to 807. Correspondingly, the 3' non-coding region of the G mRNA of strain A51908 was 18 nucleotides shorter than those of strains VC464, FS-1 and 391-2. The 3' non-coding region of BRSV G mRNA is 70 nucleotides (52 in strain A51908) shorter than the corresponding sequence of HRSV. Recently, sequence analysis of the carboxy-terminal third of the G protein gene of subgroup A HRSVs revealed a region prone to polymerase errors resulting in the insertion or deletion of adenosine residues in some runs of such residues (Cane et al., 1993). Sequences of BRSV G genes also had runs of adenosine residues in the region coding for the extracellular domain, but no enhanced mutation was observed in this region among BRSV strains.

The G mRNA of strain A51908 had a major open reading frame of 263 amino acids. The predicted relative molecular mass of this polypeptide is 28.9K. In vitro
expression of the G gene produced two unglycosylated forms of the G protein (28K and 23K). When the second methionine at position 48 was mutated to isoleucine, only one polypeptide of 28K was produced by in vitro expression, indicating that the 23K protein was produced by either leaky scanning or independent internal initiation (S. K. Mallipeddi & S. K. Samal, unpublished). The deduced G protein amino acid sequence of strain A51908 was compared with those of strains FS-1, VC464 and 391-2 (Fig. 2). The predicted G protein of strain A51908 was longer by six amino acids at the C terminus than that of the strain 391-2 (Lerch et al., 1990). However, cross-neutralization and cross-immunoprecipitation tests using polyclonal antibodies failed to show any detectable antigenic difference between these strains (data not shown). Our results indicate that the six amino acids from the C terminus of the BRSV G protein are probably not involved in neutralization. Similar results have been reported with HRSV, where C-terminal truncated forms of the G protein did not show any difference in neutralization (Olmsted et al., 1989), and this region has been shown to accommodate many changes without losing its function (Garcia-Barreno et al., 1990).

The levels of identity among the G proteins of the four BRSV strains were 84 to 95% (Table 1). However, between the prototype strains of subgroups A and B of HRSV, there was only 53% amino acid identity (Johnson et al., 1987). Within these subgroups, the G protein was also highly divergent with up to 20% amino acid variability between subgroup A strains (Cane et al., 1991) and up to 9% variability between subgroup B strains (Sullender et al., 1991). Since variability between the four BRSV strains sequenced was 5 to 16%, all appear to belong to a single subgroup. When the extracellular domains of the G proteins of BRSV strains were compared, there was a maximum divergence of 19% between A51908 and 391-2 or FS-1, and there was a minimum divergence of 6% between FS-1 and VC464. However, the cytoplasmic and transmembrane domains were highly conserved, with less than 5% divergence among BRSV strains.

The deduced amino acid sequences of the BRSV strains showed a high level (25%) of serine plus threonine content that are potential O-linked glycosylation sites. In addition, four or five potential N-linked glycosylation sites were found in all strains examined (Fig. 2). Only the potential N-linked glycosylation site at amino acid residue 251 was conserved in all BRSV strains and in a majority of subgroup A strains of HRSV (Cane et al., 1991); it was absent in subgroup B strains. This suggests that N-linked glycans at this site are probably important for structural and functional properties of the glycoprotein. There are five cysteine residues conserved in the G proteins of all strains, indicating similar structural features. A conserved 13-amino acid region of HRSV (Johnson et al., 1987) is boxed. (●) Cysteine residue; (*) end of each of the four proteins.
References


(Received 9 February 1993; Accepted 30 April 1993)