DEDUCED AMINO ACID SEQUENCE AND POTENTIAL O-GLYCOSYLATION SITES FOR THE
BOVINE CORONAVIRUS MATRIX PROTEIN

William Lapps, Brenda G. Hogue, and David A. Brian

Department of Microbiology
The University of Tennessee
Knoxville, Tennessee 37996-0845

ABSTRACT

The nucleotide sequence of the matrix (M) protein gene of the bovine coronavirus (BCV) was determined by sequencing cDNA clones derived from genomic RNA. The gene was found to map at the 5' side of the nucleocapsid protein gene and its sequence predicts a protein of 230 amino acids having a molecular weight of 26,376. The BCV M protein shares extensive sequence homology with the matrix protein of the mouse hepatitis coronavirus (MHV) but differs notably in the amino terminal region external to the virion envelope where BCV apparently uses at least two of its six potential O-glycosylation sites.

INTRODUCTION

The bovine enteric coronavirus (BCV) is comprised of 4 major structural proteins, an internal nucleocapsid phosphoprotein (N) of 52 Kd, a glycosylated peplomeric envelope glycoprotein (P or E2) of 200 Kd (found on virions primarily as separate subunits of 100 and 120 Kd), a glycosylated hemagglutinin envelope glycoprotein (H) of 140 Kd (composed of disulfide-linked subunits of 65 Kd), and a glycosylated matrix-like envelope glycoprotein (M or E1) (Hogue et al., 1984; King and Brian, 1983). The role of each of these proteins in inducing protective immunity is not yet known, but presumably each of the envelope glycoproteins is important.

To investigate the role of individual viral proteins in virus replication and in induction of immunity, we have prepared cDNA clones beginning from the 3' end of the BCV genome and examined the sequences of potential genes. Within the first (3') 2163 bases we found a noncoding region of 291 bases, a gene for a 49,379 mol. wt. nucleocapsid protein and part of the gene for the matrix protein, arranged in that order from the 3' end of the genome (Lapps and Brian, submitted). To complete the sequencing of the M gene we sequenced a region of a clone that extends to 2.8 Kb from the 3' end of the genome.

MATERIALS AND METHODS

Cells and Virus

The Mebus strain of the bovine coronavirus (BCV) was plaque purified and grown on the human rectal tumor cell line as previously described (Hogue et al., 1984; Lapps and Brian, 1985).

Radiolabeling, Growth and Purification of Virus

Confluent monolayers of cells were infected with a multiplicity of 1 PFU/cell. Viral proteins were labeled by adding 400 μ Ci [3 H]-labeled essential amino acids (150–200 mCi/mg, ICN) per 150 cm² flask in medium containing 10% normal essential amino acid concentration. Viral glycoproteins were labeled by adding 400 μ Ci [3 H]-labeled glucosamine (5–15 Ci/mmol, ICN) per 150 cm² flask. Virus was purified by isopycnic sedimentation as previously described (Hogue et al., 1984).

Polyacrylamide Gel Electrophoresis and Immunoblotting

The procedures used for polyacrylamide gel electrophoresis and immunoblotting have been described (Hogue et al., 1984). For examining intracellular proteins, whole cell lysates were prepared by scraping washed cells into cold phosphate buffered saline, pelleting, and lysing the cells in sterile distilled water (100 μ l per cells from 60 mm dish) by bath sonication for 10 seconds. Lysates were stored at -80°C until use. Lysates were mixed with an equal volume of 2X Laemmli sample treatment buffer (Laemmli, 1970) containing 5 M urea before electrophoresis.

Inhibitors

Tunicamycin (Sigma) used at a final concentration of 1.2 or 12 μ M, and monensin (Calbiochem) used at a final concentration of 1.0 μ M were added to the cells immediately after virus adsorption and were incubated with the cells until the time of cell lysate preparation, 24 h p.i.

cDNA Cloning of BCV Genomic RNA

cDNA cloning and identification of three 3'-specific clones have been described in detail (Lapps and Brian, submitted). Essentially cDNA copied from size-selected genomic RNA was made double stranded and the double stranded product was C-tailed and cloned into the Pst I site of the pUC9 vector using a modified procedure of Gubler and Hoffman (1983). Clones containing virus specific inserts were selected by colony hybridization with 32 P-labeled cDNA prepared from size-selected genomic RNA, and 3' specificity of the clones was determined by Southern hybridization of insert DNA with 32 P-labeled poly(dT) probe.

DNA Sequencing and Sequence Analyses

DNA sequencing was done by the method of Maxam and Gilbert (1981) and sequences were analyzed with the aid of the program developed by Queen and Korn (1984) as made available by Beckman Microgenie Software (Beckman Instruments, Inc.).

RESULTS

Evidence for Two O-linked Oligosaccharides Per M Molecule

The M protein (gp26) of BCV routinely electrophoretically migrated in SDS polyacrylamide gels as a group of predominantly three poorly resolved bands (Fig. 1; Hogue et al., 1984; King and Brian, 1982). Immunoblotting

with rabbit antiserum against gp26 failed to resolve the bands and furthermore demonstrated the tendency of this protein to self aggregate (Fig. 1). In a study designed to characterize the glycosylation of the three BCV glycoproteins, we learned that tunicamycin, an inhibitor of N-linked glycosylation, while it inhibited the glycosylation of gp 190 (i.e., the peplomer protein for which the subunits are gp120 and gp100) and gp140 (the hemagglutinin) (Hogue and Brian, this volume), it did not inhibit the synthesis or glycosylation of gp26 (Fig. 1). These results support the conclusions of Niemann and Klenk (1981) which state that, on the basis of carbohydrate labeling patterns, gp26 is glycosylated with O-linked oligosaccharides. Three distinct intracellular protein species with apparent molecular weights of 22, 24, and 26 kDa were identified by gp26 antiserum, and their sizes and relative amounts appeared not to be altered by tunicamycin (Fig. 1). Because the increments in molecular weight are approximately 2000, the difference contributed by one O-linked oligosaccharide (Klenk and Rott, 1981), we hypothesized that a polypeptide of approximately 22 kDa is the primary translation product and two chains of O-linked oligosaccharides are subsequently added. Such a hypothesis is supported by the appearance on glucosamine-labeled virus of only two species of labeled protein corresponding in molecular weight to the 24 and 26 kDa M species (Fig. 1). Since O-glycosylation occurs in the Golgi and

Fig. 1. Electrophoretic identification of different states of the M protein. Lane 1: purified virus immunoblotted with anti-gp26 serum. Lanes 2, 3 and 4: lysates of infected cells grown respectively in 12 uM, 1.2 uM or 0 tunicamycin (TM), electrophoresed on the same gel as lane 1, and immunoblotted with anti-gp26 serum. Lanes 5 and 6: purified virus radiolabeled with ³H-amino acids. Lanes 7 and 8: lysates of infected cells grown respectively in 0 or 1 uM monensin (MON) and immunoblotted with anti-gp26 serum. Lanes 9 and 10: purified virus radiolabeled with ³H-glucosamine. Samples treated with 2-mercaptoethanol (2-ME) are indicated. Polyacrylamide concentrations are 9% for lanes 1 through 4, and 8% for lanes 5 through 10.

is indirectly inhibited by monensin, a drug which destroys Golgi function (Niemann et al., 1982), we tested the effect of monensin on the appearance of the different M species. Monensin caused most of the M protein to remain in the 22 kDa form, further supporting the notion that M is glycosylated with O-linked oligosaccharides (Fig. 1).

The Deduced Amino Acid Sequence of the BCV M Protein Reveals a 230 Amino Acid Protein Having a Molecular Weight of 26,376 and Identifies 6 Potential O-glycosylation Sites

To examine the primary structure of the BCV M protein, three cDNA clones that come from the 3' end of the genome were sequenced in part as depicted in Fig. 2. The clones MN3, MA5 and MA7 include all of the N protein gene and extend 5'-ward from it. Clone MN3 was sequenced in its entirety to determine the N protein structure and was shown to contain some M gene sequence on the basis of homology with the MHV M (E1) gene (Fig. 3) (Lapps and Brian, submitted).

The 5' end of the sequence depicted in Fig. 3 begins at the third Dde I restriction endonuclease site of clone MA5 (Fig. 2). An examination of all possible open reading frames by computer analysis, reveals a large open reading frame in the first frame of virus-sense RNA (identified as M) that is immediately preceded by several stop codons (Fig. 4). Potential open reading frames that could encode peptides of 30 amino acids or greater are found in other reading frames of both virus-sense RNA and of virus complementary-sense RNA, but these were not evaluated further.

Examination of the deduced amino acid sequence reveals that the M protein of BCV is approximately the same size as the M protein of MHV A59 and shares extensive homology with it (Armstrong et al., 1984). It is

Fig. 2. Restriction endonuclease map and sequencing strategy of BCV cDNA clones. Clones MN3, MA5 and MA7 were all found to be in the same orientation and are shown as they map with respect to the virus genome. ■, □, and ◻ indicate restriction sites that are 5' end labeled with polynucleotide kinase for clones MA5, MN3, and MA7 respectively. ● and ○ indicate sites that were 3' end labeled using reverse transcriptase and the appropriate labeled

comprised of 230 amino acids and has a calculated molecular weight of 26,376, that of MHV is 228 amino acids and has a calculated molecular weight of 26,030. By maximum alignment of the proteins, 200 of the amino acids (>86%) are the same, and another 16 (7%) represent conservative changes. The BCV M protein is slightly basic, having a net charge of +9 at neutral pH, and is hydrophobic. One hundred thirteen (49%) of the amino acids are hydrophobic and the distribution of hydrophobic amino acids is nearly identical to that for the MHV M protein.

DISCUSSION

Because BCV and MHV are antigenically related viruses, that is, monospecific antibodies recognize analogous proteins on each virus (Hogue et al., 1984), some amino acid sequence homology between their M proteins was expected. BCV and MHV have significant structural differences, however, (BCV has 4 major structural proteins and MHV 3) and differences in genome sequences are expected. Where does the gene for the additional BCV protein (the hemagglutinin, gpl40) lie, for example? Sequence differences within genes are also expected, especially for regions of external proteins carrying epitopes for neutralization, since there is no serological cross neutralization between BCV and MHV.

From an analysis of the primary sequence of the 3' 2,451 bases of the BCV genome we know that, at least for this region of the genome, there is a direct parallel between BCV and MHV. That is, from the 3' end toward the 5' end, in order, there is a poly A tract, a noncoding region, the N gene and the M gene. This order is not strictly observed for IBV or TGEV (Boursnell et al., 1984; Kapke and Brian, 1986). Although notable differences exist in the primary structure of the genome at the 3' end

```

      30      60      90      120
AGTACTGTGCCCTTCTATTTATGTGTTTATAGAGGTAGGCAGTTTATGAGTTTTCACACGATGTAAACACCAAGTCTCTGATGTGGATGACGTTTAGTTAATCCAACATTATGAGT
      M S

      150      180      210      240
AGTGTAACTACACCAACCAAGTTTACACCTGGACTGCTGATGAAGCTATTAAATTCCTAAAGGAATGGAACTTTCTTTGGGTATTATACTACTTTTATTACAATCATATGCAATTT
S V T T P A P V Y T W T A D E A I K F L K E W N F S L G I I L L F I T I I L Q F

      270      300      330      360
GGATATACAAAGTCGAGTAAGTTTGTATTATTAAGATGATCATTTTGTGGCTTATGTGGCCCTTACTATCATCTTAACATTTTCAATTGCGTGTATGCGTTGAATAATGTGTAT
G Y T S R S M F V Y V I K M I I L W L M W P L T I I L T I F N C V Y A L N N V Y

      390      420      450      480
CTTGGCTTTTCTATAGTTTTCACCTATAGTGGCCATTATCATGTGGATTGTGTTTGTGAATAGTATCAGGTGTGTTTATAGAACTGGAAGTGGTGGAGTTTCAACCCAGAAACAAAC
L G F S I V F T I V A I I M W I V Y F V N S I R L F I R T G S W W S F N P E T N

      510      540      570      600
AACTTGATGTGTATAGATATGAAGGGAAGGATGTATGTTAGGCCGATAATTGAGGACTACCATACCTTACGGTCACAATAATACGTGGTGCATCTTTACATGCAAGGTATAAAACAGGT
N L M C I D M K G R M Y V R P I I E D Y H T L T V T I I R G H L Y M Q G I K L G

      630      660      690      720
ACTGGCTATCTTTGTGACATTTGCCAGCTTATGTGACTGTCTGTAAGGTCTCACACCTGCTCAGGTATAAGCGTGGTTTCTTGACAAGATAGGCGATAGTGTGTTTGTCTGTTTAT
T G Y S L S D L P A Y V T V A K V S H L L T Y K R G F L D K I G D T S G F A V Y

      750      780      810      840
GTTAAGTCCAAAGTCGGTAATTACCGACTGCCATCAACCCAAAAGGGTCTGCGCATGGACACCGCATTTGTTGAGAAATAATCTCTAACTTTAAGGATGCTCTTTACTCTGGTAAGCAA
V K S K V G N Y R L P S T Q K G S G M D T A L L R N N I M S F T P G K Q

      870      900
TCCAGTAGTAGACGCTCTTTTGAAATCGTTCTGGTAATGGCATCTTTAAGTGGCGCATCAGTCGACCAATCT
S S S R A S F G N R S G N G I L K W A D Q S D Q S

```

Fig. 3. Nucleotide sequence of the BCV M gene and deduced amino acid sequence for the protein. The nucleotide sequence illustrated is from the total region described in Fig. 2 and includes the beginning of the nucleocapsid protein gene. Nucleotides bounding matrix protein are 115 to 804. The first nucleotide in the initiating methionine of the nucleocapsid protein gene is 817.

between BCV and MHV (Lapps and Brian, submitted) a remarkable conservation in M gene and protein sequence is found suggesting rigorous constraint on structures required for specific functions. Between BCV and MHV proteins there is a 76% homology at the nucleotide level, and 86% of the amino acid level. When amino acids of like properties are considered, the homology becomes 93%. By analogy, one would therefore expect the BCV M protein to share both topological (Rottier et al., 1986) and biological (Collins et al., 1982) properties with the mouse protein. That is to say, probably 28 amino acids are external to the virion envelope and are able to interact with antibody and complement to neutralize infectivity. These properties have not been studied for BCV, however.

Both BCV and MHV have been shown to possess O-glycosidically-linked oligosaccharides (Holmes et al., 1981; Niemann and Klenk, 1981; Niemann et al., 1984). We conclude from our data that BCV utilizes at least two of its 6 potential to add oligosaccharide side chains.

ACKNOWLEDGEMENTS

This work was supported by grant AI-14367 from the National Institute of Allergy and Infectious Diseases, by grant 82-CRSR-2-1090 from the U.S. Department of Agriculture, and in part by a grant from the National Foundation for Ileitis and Colitis, Inc.

Fig. 4. Schematic diagram of open reading frames that locate the bovine coronavirus matrix protein. Bars above the lines designate the first potential initiating methionine codons. Bars below the lines designate translation termination codons.

REFERENCES

- Armstrong, J., H. Niemann, S. Smeekens, P. Rottier, and G. Warren. 1984. Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. *Nature* 308:751-752.
- Collins, A. R., R. L. Knobler, H. Powell, and M. J. Buchmeier. 1982. Monoclonal antibodies to murine hepatitis virus-4 (strain JHM) define the viral glycoprotein responsible for attachment and cell-cell fusion. *Virology* 119:358-371.
- Gubler, U. and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* 25:263-269.
- Hogue, B. G., B. King, and D. A. Brian. 1984. Antigenic relationships among proteins of bovine coronavirus, human respiratory coronavirus OC43, and mouse hepatitis coronavirus A59. *J. Virol.* 51:384-388.
- Holmes, K. V., E. W. Doller, and L. S. Sturman. 1981. Tunicamycin resistant glycosylation of a coronavirus glycoprotein: demonstration of a novel type of viral glycoprotein. *Virology* 115:334-344.
- King, B., and D. A. Brian. 1982. Bovine coronavirus structural protein. *J. Virol.* 42:700-707.
- Klenk, H. D. and R. Rott. 1981. Cotranslational and posttranslational processing of viral glycoproteins. *Curr. Top. Microbiol. Immunol.* 90:19-48.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685.
- Lapps, W., and D. A. Brian. 1985. Oligonucleotide fingerprints of antigenically related bovine coronavirus and human coronavirus OC43. *Arch. Virol.* 86:101-108.
- Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Meth. Enzymol.* 65:499-560.
- Niemann, H., B. Boschek, D. Evans, M. Rosing, T. Tamura, and H.-D. Klenk. 1982. Post-translational glycosylation of coronavirus glycoprotein E1: Inhibition by monensin. *The EMBO Journal* 1:1499-1504.
- Niemann, H., G. Heisterberg-Moutsis, R. Geyer, H.-D. Klenk, and M. Wirth. 1984. Glycoprotein E1 of MHV-A59: Structure of the O-linked carbohydrates and construction of full length recombinant cDNA clones. *Adv. Exp. Biol. Med.* 173:201-213.
- Niemann, H., and H.-D. Klenk. 1981. Coronavirus glycoprotein E1, a new type of viral glycoprotein. *J. Mol. Biol.* 153:993-1010.
- Queen, C. and L. J. Korn. 1984. A comprehensive sequence analysis program for the IBM personal computer. *Nucl. Acids Res.* 12:581-599.
- Rottier, P., D. Brandenburg, J. Armstrong, B. Van Der Zeijst, and G. Warren. 1984. Assembly *in vitro* of a spanning membrane protein of the endoplasmic reticulum: The E1 glycoprotein of coronavirus mouse hepatitis virus A59. *Proc. Natl. Acad. Sci. USA* 81:1421-1425.