STRUCTURE AND EXPRESSION OF THE BOVINE CORONAVIRUS HEMAGGLUTININ PROTEIN

Thomas E. Kienzle, Sushma Abraham, Brenda G. Hogue[1]
and David A. Brian

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

INTRODUCTION

cDNA clones prepared from genomic RNA of the Mebus strain of bovine coronavirus (BCV) were sequenced to reveal the hemagglutinin (H) gene of 1,272 bases that predicts a 47,700 mol. wt. apoprotein of 424 amino acids. The H gene mapped on the immediate 5′ side of the peplomer gene. The H protein sequence revealed a putative N-terminal signal peptide of 18 amino acids, 9 potential glycosylation sites, 14 cysteine residues, and a potential C-terminal anchor region of 26 amino acids. When transcripts of the gene were translated *in vitro* in the presence of microsomes, signal cleavage, glycosylation, and membrane anchorage were observed, but not disulfide-linked dimerization. Translation of a truncated mRNA having no sequence for the C-terminal anchor resulted in a nonanchored, intraluminal (intramicrosomal) protein. When the H protein was expressed in cells in the absence of other coronaviral proteins, it became glycosylated, dimerized, and transported to the cell surface. The BCV hemagglutinin protein, therefore, is a type 1 glycoprotein that contains all the information it needs for signal cleavage, glycosylation, disulfide-linked dimerization, and transport to the cell surface.

MATERIALS AND METHODS

cDNA Cloning of BCV Genomic RNA

The Mebus strain of BCV was grown on human rectal tumor (HRT) cells and purified as previously described (Lapps *et. al.*, 1987). cDNA cloning was done essentially as described (Lapps *et. al.*, 1987), except that random 5-mer oligodeoxynucleotides (Pharmacia) were used as primers for first strand synthesis to generate several clones, one of which was I1 (Fig. 1A). A synthetically made primer (5′ ATTATGACCGCACACC 3′) was used to extend genomic sequence 5′-ward from clone I1, and this was used to generate several clones one of which was LA6 (Fig. 1A). Clones were selected by colony hybridization to randomly primed cDNA prepared from

[1]Current address: Department of Microbiology and Immunology, UCLA School of Medicine, Center for Health Sciences, Los Angeles, CA 90024

Coronaviruses and Their Diseases
Edited by D. Cavanagh and T.D.K. Brown
Plenum Press, New York, 1990

genomic RNA, and clones were mapped relative to one another and to the 3′ end of the genome using a matrix spot hybridization technique.

DNA Sequencing and Sequence Analysis

Dideoxy sequencing was used throughout. The ends of clones I1 and LA6 were sequenced using universal primers for pUC vectors, and the clones were ligated at the Hae 3 site (Fig. 1A) to form a continuous sequence mapping in the region of the H gene. The reconstructed sequence was subcloned into pGEM3Z (Promega) and a 3′ and 5′ nested set of subclones was generated using exonuclease III and S1 nuclease (Henikoff, 1984). Subclones were sequenced using universal primers for the pGEM vector, and by synthetic primers that were made complementary to regions within the H gene. Sequences were analyzed with the aid of the Microgenie program (Beckman).

Expression Analyses

A construct of the H gene beginning 15 bases upstream (i.e., beginning with the CTAAAC intergenic sequence) from the putative ATG start codon (an exonuclease III/S1-generated subclone of the LA6/I1 construct in pGEM4Z) and under the control of the Sp6 polymerase promoter, was made and named pHSp6 (Fig. 1B). The same insert was also subcloned into pGEM3Z (using the Eco R1 and Hind III sites within the multiple cloning region) such that the H gene was under the control of the T7 polymerase promoter. This was named pHT7 (Fig. 1C).

For *in vitro* expression analyses, pHSp6 was linearized with Hind III to yield full-length transcripts, or with Eco R5 to yield transcripts lacking the coding region for the C-terminal anchor, and transcribed with

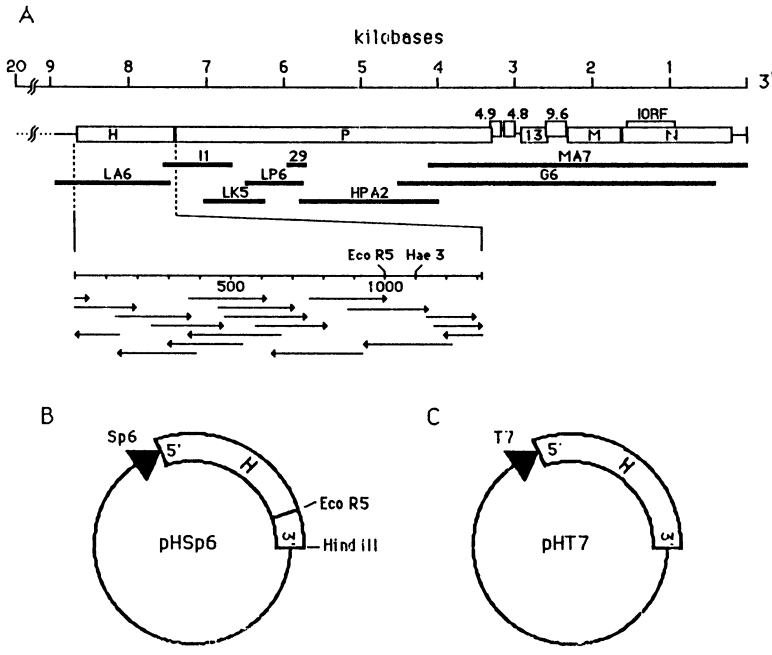

Fig 1. A. Gene map of the BCV genome, cDNA clone positions, and strategy for sequencing the H gene. B. Plasmid construct for generating H transcripts with Sp6 polymerase. C. Plasmid construct for generating H transcripts with T7 polymerase.

SP6 polymerase (Promega protocol for yielding capped transcripts). Translation was done in wheat germ extract (Promega) in the presence or absence of canine pancreatic microsomes (Promega). Protein was labeled with ^{35}S-methionine (ICN). Immunoprecipitates were made using polyclonal rabbit anti gp65 (H subunit) (Hogue *et. al.*, 1984) by the method of Anderson and Blobel (1983). Carbonate extractions at pH 11 were done by the method of Fujiki *et. al.*, (1982). Competitive inhibition of N-Linked glycosylation was accomplished with 30μM octanoyl-asparagine-leucine-threonine (a gift from Dr. F. Naider, City University of New York), in the microsomal-containing translation mixture (Lau *et. al.*, 1983). Polyacrylamide gel electrophoresis analyses were done as previously described (Hogue *et. al.*, 1984).

For *in vivo* expression and analysis by immunofluorescence, HRT cells were infected with a vaccinia virus recombinant (vTF7-3, moi of 30) that expresses T7 polymerase (Fuerst *et. al.*, 1987), transfected with pHT7 plasmid (1 μg per 1 cm^2) using Lipofectin Reagent (BRL) to mediate transfection, and prepared at 32h for internal or surface immunofluorescence using the method of Kaariainen *et. al.*, (1983). For

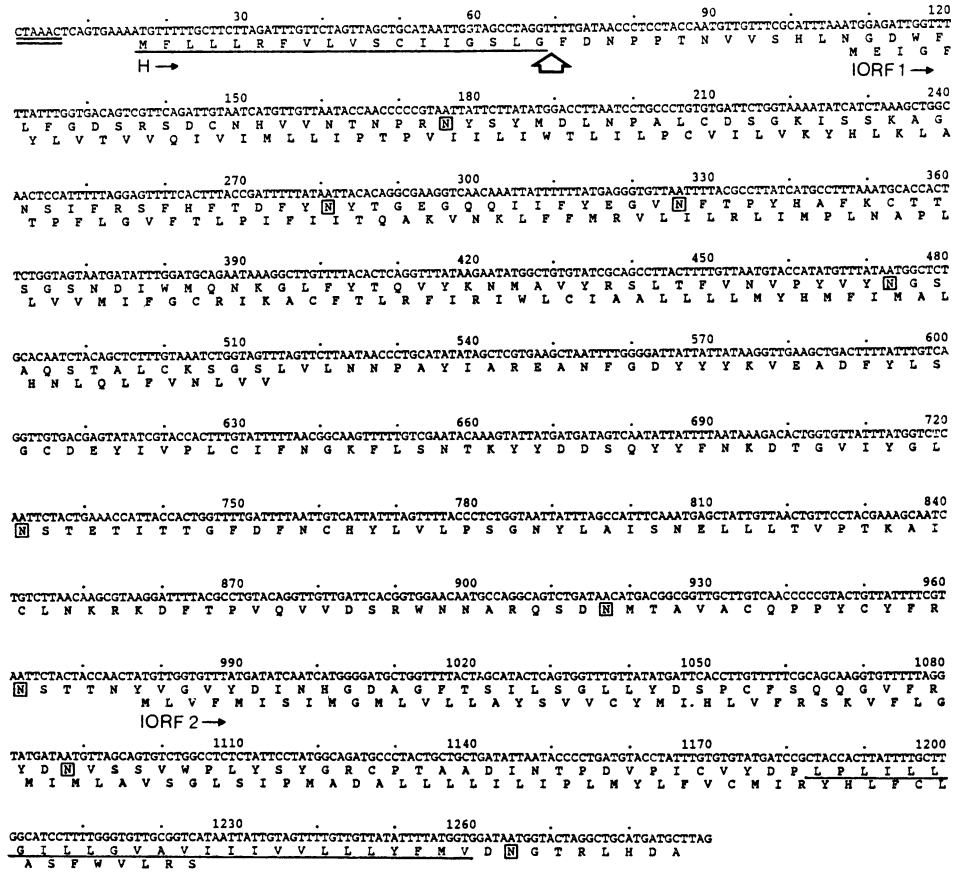

Fig. 2. Sequence of the H gene and deduced amino acid sequences of H, IORF1, and IORF2. The nucleotide sequence begins with the CTAAAC intergenic sequence 15 bases upstream from the presumed start condon of the H gene. The N-terminal signal peptide and presumed C-terminal anchor sequences are underlined. The signal cleavage site is indicated by an arrow. Potential N-linked glycosylation sites (NXS or NXT, where X ≠ P) are boxed.

analysis by immunoprecipitation, CV-1 cells were infected with vaccinia vTF7-3 (moi of 2), and transfected with pHT7 (0.2 μg per cm^2) using Lipofectin reagent, incubated 16 h at 37°, and labeled for 1 hr with 100 μCi each of ^{35}S-cysteine (Amersham) and ^{35}S-methionine (ICN). Immunoprecipitation was done on cell lysates as described above.

RESULTS

Based on the facts that (1) the hemagglutinating mammalian coronaviruses BCV and HCV OC43 have the H protein, while the antigenically related MHV A59 does not (Hogue *et al.*, 1984), and, (2) BCV has an additional mRNA species (species 2a) that the nonhemagglutinating MHV does not have (Keck *et. al.*, 1988), the H gene was predicted to be approximately 1.2 Kb in length and to lie on the 5′ side of the peplomer protein gene (Hogue *et. al.*, 1989; Keck *et. al.*, 1988). The genome map position and sequences within clone I1 suggested that it contained the 5′ end of the BCV peplomer (P) gene (because of sequence similarities with the MHV P gene), the intergenic CTAAAC sequence preceding the P gene, and part of an open reading frame immediately 5′-ward of the P gene. The open reading frame immediately 5′ of the P gene was tentatively identified as the H gene. To find the 5′ end of the putative H ORF, cDNA cloning was done with a primer sequence from the 5′ end of clone I1 and a 1.6 Kb clone, LA6, was obtained (Fig. 1A). Partial sequencing suggested that it

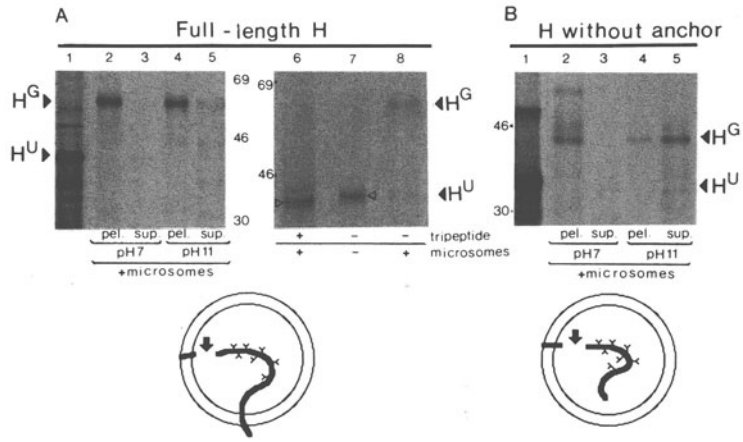

Fig. 3. Membrane anchorage and signal cleavage of the H protein. A. Full-length H subunit synthesized and glycosylated (H^G) in the presence of microsomes remains with the pelleted membrane after treatment at pH 11 (lane 4). Full-length H subunit undergoes signal cleavage, but not glycosylation, when translated in the presence of the tripeptide glycosylation inhibitor and microsomes (lane 6). This is schematically pictured in the lower left corner. B. The C-terminally truncated and glycosylated H subunit (H^G) is released from microsomes and is found mostly in the supernate when microsomes are treated at pH 11 (lane 5). This is schematically pictured in the lower right corner. All lanes represent translation products treated as indicated (lane 1 in parts A and B had no treatment) and electrophoresed in a 12% SDS polyacrylamide gel after reduction with 5% 2-mercaptoethanol.

probably contained the 5′ end of the H gene and a LA6/I1 construct was made and completely sequenced (Fig. 1A). An open reading frame identifying the H gene was obtained (Fig. 2).

The following features were found from an analysis of the potential BCV H gene (Fig. 2). (1) The ORF is preceded 15 bases upstream by a CTAAAC intergenic consensus sequence of the type that precedes other BCV genes (Lapps *et. al.*, 1987). (2) The H ORF is 1,272 nucleotides long, encoding a potential apoprotein of 424 amino acids and having a predicted mol. wt. of 47,700. This is near the size expected from studies on H synthesis and processing (Deregt *et. al.*, 1987; Hogue *et. al.*, 1989). (3) There are two strongly hydrophobic regions, one covering a stretch of 18 aa at the N terminus and a second covering a stretch of 26 aa near the C terminus. An N-terminal sequence of NH_2-F-D-N-P-P-T-N-V-V- on the virion H protein (Hogue *et. al.*, 1989) that matches with amino acids 19 through 27 (Fig. 2) indicates first that the H gene sequence is authentic, and second that the N-terminal hydrophobic region functions as a signal peptide for membrane translocation and is cleaved between amino acids 18 and 19. The C-terminal hydrophobic region may function as a stop transfer signal and protein anchor. (4) There are 9 potential N glycosylation sites consistent with the high level of N-linked glycosylation for this protein which would suggest that nearly all of these must be used (Hogue *et. al.*, 1989). (5) There are 14 cysteine residues, some of which must be used for interchain disulfide bonding (King and Brian, 1982). (6) There are two large internal open reading frames, the significance of which is not yet known. The first (IORF1), from nucleotide 107 to 514, and the second (IORF2) from nucleotide 977 to 1,225, potentially encode proteins of 15,741 and 9,514 respectively.

To confirm by a second method that the H ORF does, in fact, encode the hemagglutinin, and to study the synthesis, orientation, processing and cellular localization of the hemagglutinin in the absence of other coronaviral proteins, H gene expression was studied *in vitro* and *in vivo*.

In vitro expression of H revealed the following: (1) Unglycosylated H subunit (H^u) migrated as a protein of 43Kd whereas glycosylated H subunit (H^G) migrated as a 65Kd glycoprotein, similar to the size of subunits obtained from infected cells or virions (Fig. 3A, lanes 1 and 2,

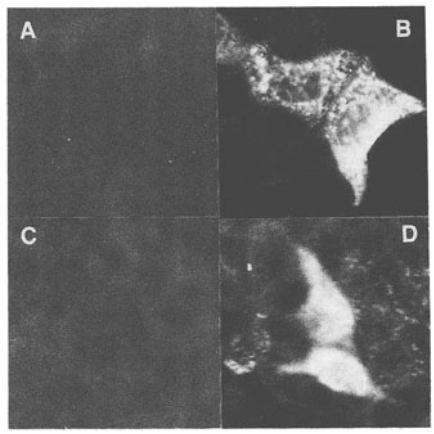

Fig. 4. Surface immunofluorescence of H protein. A. Uninfected HRT cells. B. BCV-infected cells. C. Vaccinia (vTF7-3)-infected cells. D. Cells infected with vaccinia vTF7-3 and transfected with pHT7 DNA.

and Hogue *et. al.*, 1989). (2) Both forms (H^u and H^G) were immunoprecipitable with H-specific rabbit antiserum (data not shown). (3) The H protein is almost entirely protected by protease when made in the presence of microsomes indicating that it is translocated into microsomes (data not shown). (4) H protein is not released from microsomes by carbonate (pH 11) extraction indicating that it is membrane anchored (Fig. 3A, lanes 2-5). (5) H protein made in the presence of microsomes and an inhibitor of glycosylation migrates as a protein approximately 2Kd smaller (Fig. 3A, lane 6) indicating that the signal peptide was cleaved. (6) H protein with its C-terminal hydrophobic sequence missing is not anchored in microsomal membranes but is secreted into the lumen of the microsomes (Fig. 3B, lanes 2-5), indicating that the C-terminal hydrophobic sequence serves as the membrane anchor. (7) No dimeric forms of H (either full-length or truncated) were found after *in vitro* translation.

In vivo expression of H revealed that it is found in the cytoplasm (data not shown) and on the external surface of the cell membrane (Fig. 4D). Immunoprecipitation of this product revealed a dimeric protein of 140 Kd reducible with 2-mercaptoethanol to subunits of 65 Kd (data not shown).

DISCUSSION

While this work was in progress, the sequence and *in vitro* expression of the H gene of the Quebec isolate of BCV was reported by Parker *et. al.* (1989). The Mebus gene sequence differs by only two bases from the Quebec sequence, and these are at base 322 where in our sequence it is G (rather than C) making amino acid 103 a valine (rather than leucine), and base 522 which is T (rather than A) resulting in no amino acid difference. The two large internal ORFs are present in both isolates.

Our results demonstrate for the first time that the BCV H protein is a type 1 glycoprotein and is not dependent upon other viral proteins for signal cleavage, dimer formation and transport to the cell surface. The scheme is summarized in Fig. 5.

A presumably incomplete form of the H gene is found in MHV A59 and maps on the 5′ side of the peplomer protein gene (Luytjes *et. al.*, 1988). The MHV A59 protein (ORF2), although apparently not expressed during infection, possesses the putative active site (FGDS) for esterase activity (Luytjes *et. al.*, 1988). Remarkable sequence homology between the MHV A59 ORF2 and the hemagglutinin of influenza virus C, which also possesses esterase activity, has suggested an evolutionary relationship between these two proteins (Luytjes, 1988; Vlasak *et. al.*, 1987).

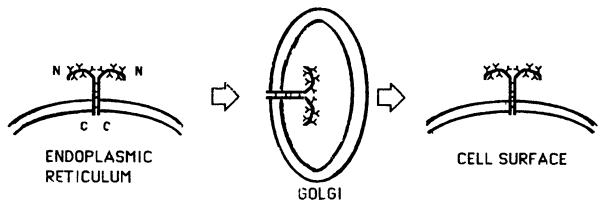

Fig. 5. Model illustrating the *in vivo* expression of BCV H protein in the absence of other coronaviral proteins. H is synthesized at the endoplasmic reticulum where it becomes translocated into the endoplasmic reticulum, becomes anchored at its C-terminus, loses the signal peptide from its N-terminus, becomes disulfide linked, and acquires sugars. It migrates through the Golgi and then to the plasma membrane.

Since BCV H protein shares 60% amino acid sequence homology with the potential product of MHV A59 ORF2, possesses the FGDS sequence of the esterase active site (beginning at base 124 in Fig. 2), and demonstrates esterase activity (Vlasak *et. al.*, 1988), it too may be evolutionarily related to the influenza C virus hemagglutinin. Interestingly, the influenza C hemagglutinin protein is derived by cleavage of a large precursor molecule (Herrler *et. al.*, 1988) and in this sense it is evolutionarily diverged from the BCV H protein.

ACKNOWLEDGEMENTS

This work was supported by NIAID Grant AI14367 and by USDA Grant 82-CRSR-2-1090.

REFERENCES

Anderson, D. J., and G. Blobel. 1983. Immunoprecipitation of protein from cell-free translations. Meth. Enzymol., 96:111-120.

Deregt, D., M. Sabara, and L. A. Babiuk, 1987. Structural proteins of bovine coronavirus and their intracellular processing. J. Gene Virol., 68:2863-2877.

Fuerst, T. R., P. L. Earl, and B. Moss. 1987. Use of a hybrid vaccinia virus-T7 RNA polymerase system for expression of target genes. Mol. Cell. Biol., 7:2538-2544.

Fujiki, Y., A. Hubbard, S. Fowler, and P. Lasarow. 1982. Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. J. Cell Biol., 93:97-102.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene, 28:351-359.

Herrler, G., I. Durkop, H. Becht, and H. Klenk. 1988. The glycoprotein of influenza C virus is the hemagglutinin, esterase, and fusion factor. J. Gen. Virol., 69:839-846.

Hogue, B. G., T. E. Kienzle, and D. A. Brian. 1989. Synthesis and processing of the bovine enteric coronavirus hemagglutinin protein. J. Gen. Virol., 70:345-352.

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.

Kaariainen, L., I. Virtanen, J. Saraste, and S. Keranen. 1983. Transport of virus membrane glycoproteins, use of temperature-sensitive mutants and organelle-specific lectins. Meth. Enzymol., 96:453-465.

Keck, J. G., B. G. Hogue, D. A. Brian and M. M. C. Lai. 1988. Temporal regulation of bovine coronavirus RNA Synthesis. Virus Res., 9:343-356.

King, B., and D. A. Brian. 1982. Bovine coronavirus structural proteins. J. Virol., 42:700-707

King, B., B. J. Potts, and D. A. Brian. 1985. Bovine coronavirus hemagglutinin protein. Virus Res., 2:53-59.

Lapps, W., B. G. Hogue, and D. A. Brian. 1987. Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. Virology, 157:47-57.

Lau, J. T. V., J. Welply, P. Shenbagamurthi, F. Naider, and W. J. Lennarz. 1988. Substrate recognition of oligosaccharyl transferase: Inhibition of cotranslational glycosylation by acceptor peptides. J. Biol. Chem., 258:15,255-15,250.

Luytjes, W., P. Bredenbeck, A. Noten, M. Horzinek, and W. Spaan., 1988. Sequence of mouse hepatitis virus A59 mRNA2: indications for RNA recombination between coronaviruses and influenza C virus. Virol., 166:415-422.

Parker, M. D., G. J. Cox, D. Deregt, D. R. Fitzpatrick, and L. A. Babunk. 1989. Cloning and *in vitro* expression of the gene for the E3 hemagglutinin glycoprotein of bovine coronavirus. J. Gen. Virol., 70:155-164.

Vlasak, R., M. Krystal, M. Nacht, and P. Palese. 1987. The influenza C virus glycoprotein (HE) exhibits receptor-binding (hemagglutinin) and receptor-destroying (esterase) activities. Virol., 160:419-425.

Vlasak, R., W. Luytjes, J. Leider, W. Spaan, and P. Palese. 1988. The E3 protein of bovine coronavirus is a receptor-destroying enzyme with acetyl esterase activity. J. Virol., 62:4686-4690.