



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

## Sequence and Expression Analysis of Potential Nonstructural Proteins of 4.9, 4.8, 12.7, and 9.5 kDa Encoded between the Spike and Membrane Protein Genes of the Bovine Coronavirus

SUSHMA ABRAHAM, THOMAS E. KIENZLE,<sup>1</sup> WILLIAM E. LAPPS,<sup>2</sup> AND DAVID A. BRIAN<sup>3</sup>

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

*Received February 16, 1990; accepted April 3, 1990*

The nucleotide sequence between the spike and membrane protein genes in the bovine coronavirus (BCV) genome was determined by sequencing cDNA clones of the genome, and open reading frames potentially encoding proteins of 4.9, 4.8, 12.7, and 9.5 kDa, in that order, were identified. The 4.9- and 4.8-kDa proteins appear to be vestiges of an 11-kDa protein for which a single nucleotide deletion event in the central part of the gene gave rise to a stop codon. The consensus CYAAAC sequence precedes the 4.9-, 12.7-, and 9.5-kDa ORFs and predicts that transcription will start from each of these sites. Northern analyses using sequence-specific probes and oligo(dT)-selected RNA demonstrated that the predicted transcripts are made, and that these correspond to mRNAs 4, 5, and 5-1. BCV mRNA 4 appears to be a counterpart to mouse hepatitis virus (MHV) mRNA 4 which, in the MHV JHM strain, encodes the putative 15.2-kDa nonstructural protein. BCV mRNAs 5 and 5-1 appear to be used for the synthesis of the 12.7- and 9.5-kDa proteins, respectively, which demonstrates a pattern of expression strikingly different from that utilized by MHV. MHV makes its homologs of the 12.7- and 9.5-kDa proteins from the single mRNA 5. *In vitro* translation analyses demonstrated that the BCV 9.5-kDa protein, unlike its MHV counterpart, is poorly made from downstream initiation of translation. Thus, from a comparison between BCV and MHV we find evolutionary evidence for the importance of the CYAAAC sequence in regulating coronavirus transcription. © 1990 Academic Press, Inc.

### INTRODUCTION

The structural proteins of coronaviruses are encoded at the 3' end of the single-stranded, positive-strand RNA genome. In order from the 3' end they are the nucleocapsid (N) protein, the multispinning, integral membrane (M) or matrix protein, the spike (S) or peplomer protein, and, in the case of the hemagglutinating bovine coronavirus, the hemagglutinin-esterase (HE) protein (see Spaan *et al.*, 1988, for review; Kienzle *et al.*, 1990; Parker *et al.*, 1989) (Fig. 1). The remainder of the genome encodes the large RNA-dependent RNA polymerase molecule (Bournsnell *et al.*, 1987; Pachuk *et al.*, 1989) and possibly other nonstructural proteins (Cox *et al.*, 1989). Interspersed among the structural protein genes are large open reading frames (ORFs) potentially encoding nonstructural or minor structural proteins (referred to as nonstructural proteins throughout this paper). The number and position of these ORFs, however, differ among coronavirus species. In avian infectious bronchitis virus (IBV) there are five ORFs, three (for 6.7-, 7.4-, and 12.4-kDa proteins) residing between

the S and M genes and two (for 7.5- and 9.5-kDa proteins) residing between the M and N genes (Bournsnell and Brown, 1984; Bournsnell *et al.*, 1985). In the porcine transmissible gastroenteritis coronavirus (TGEV), there are four ORFs, three (for 7.7-, 27.7-, and 9.2-kDa proteins) residing between the S and M genes and one (for a 9.1-kDa protein) residing at the 3' side of the N gene (Kapke and Brian, 1986; Kapke *et al.*, 1988). In the mouse hepatitis coronavirus (MHV) there are four ORFs, three (for 15.2-, 12.4-, and 10.2-kDa proteins in the JHM strain or 11.7-, 13-, and 9.6-kDa proteins in the A59 strain) residing between the S and M genes (Budzilowicz and Weiss, 1987; Skinner and Sidell, 1985; Skinner *et al.*, 1985; Weiss, personal communications) and one (for a 23-kDa protein) residing within the N gene [A59 strain (Armstrong *et al.*, 1983)]. We have reported the nucleotide sequence for the N, M, S, and HE genes of bovine coronavirus (BCV) from which we have learned that there is much amino acid sequence identity with the N, M, S, and HE homologs in MHV A59 (70, 86, 70, and 60%, respectively) (Abraham *et al.*, 1990; Kienzle *et al.*, 1990; Lapps *et al.*, 1987).

To further characterize the genes of BCV, the genome sequence between the S and M genes was determined, and four ORFs encoding potential proteins of 4.9, 4.8, 12.7 and 9.6 kDa were found. Of these,

<sup>1</sup> Present address: Department of Molecular Biology and Microbiology, Case Western Reserve University, Cleveland, OH 44106.

<sup>2</sup> Present address: DNX Corporation, One President Street, Athens, OH 45701-2979.

<sup>3</sup> To whom requests for reprints should be addressed.

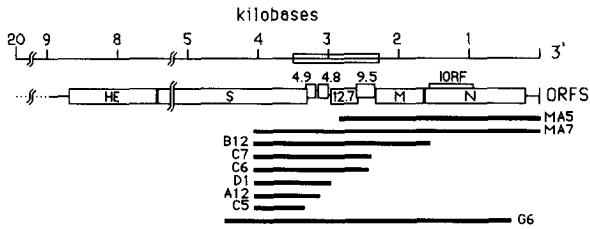


Fig. 1. Gene map of the BCV genome and cDNA clone positions. The boxed region of the genome represents the sequence shown in Fig. 2.

transcripts appear to be made beginning with ORFs for the 4.9-, 12.7-, and 9.5-kDa proteins (mRNAs 4, 5, and 5-1) revealing remarkable differences in transcriptional patterns between BCV and MHV, since in MHV both the 12.7- and 9.5-kDa homologs are translated from a single mRNA molecule (mRNA 5) (Budzilowicz and Weiss, 1987; Leibowitz *et al.*, 1988).

## MATERIALS AND METHODS

*Cloning and sequence analysis of the region between the spike and membrane protein genes.* Growth of BCV and preparation of virus stocks were as previously described (King and Brian, 1982; Lapps *et al.*, 1987). cDNA cloning of the 3' end of the BCV genome for the Mebus strain of BCV and identification of clone MA7 that represents the 3' proximal 4.2 kb of the genome have been described (Lapps *et al.*, 1987) (Fig. 1). To sequence the 5'-terminal 1.5 kb of clone MA7, the 5'-terminal 3.73-kb *Pst*I fragment of clone MA7 was subcloned into the *Hind*III site of the pUC19 vector (Pharmacia), and a 5' nested set of deletion subclones of this was generated by the method of Henikoff (1984) and sequenced. Subclones with inserts ranging from 0.8 to 2.7 kb were selected and designated C5, A12, D1, C6, C7, and B12 (Fig. 1). Sequencing was done by the chemical method of Maxam and Gilbert (1980) starting from the *Sal*I site in the multiple cloning region of the vector after a 3' fill-in reaction using reverse transcriptase and [ $\alpha$ - $^{32}$ P]dNTP or after 5' end labeling using polynucleotide kinase and [ $\gamma$ - $^{32}$ P]dNTP. Sequences were analyzed with the aid of the Microgenie program (version 5.0) from Beckman Instruments (Queen and Korn, 1984).

*Northern analyses.* Freshly confluent HRT cells were infected with a m.o.i. of approximately 5 and incubated for 9 hr. RNA was extracted from uninfected or infected cells ( $4.4 \times 10^9$  cells from one 850-cm<sup>2</sup> roller bottle per batch) by the use of guanidinium isothiocyanate as described by Lizardi (1983). RNA was pelleted through CsCl, extracted with chloroform-1-butanol, and ethanol precipitated, and poly(A)-containing RNA was se-

lected by oligo(dT)-cellulose chromatography (Dennis and Brian, 1982). Oligo(dT)-selected RNA from one batch of cells was dissolved in 100  $\mu$ l water and 5  $\mu$ l of this was electrophoresed per lane for Northern analysis. RNA was electrophoresed in 1% agarose gels in the presence of 2.2 M formaldehyde as previously described (Sethna *et al.*, 1989), capillary blotted onto Nytran membrane (Schleicher and Schuell) using 20 $\times$  SSC (1 $\times$  SSC is 0.15 M NaCl-0.015 M sodium citrate) for 24 hr, and crosslinked to the membrane with ultraviolet light (Khandjian, 1986). Hybridization was done as described by Thomas (1980) using probes radiolabeled by nick-translation (approximately  $2 \times 10^7$  Cerenkov cpm for a membrane of 130 cm<sup>2</sup>), and blots were washed in 1 $\times$  SSC-0.1% SDS at room temperature for 30 min, then in 0.1 $\times$  SSC-0.1% SDS at 60 $^\circ$  for 45 min.

cDNA clones from four regions of the genome were used as probes. These were clone MA5 (in pUC9; Fig. 1), which represents the 3'-terminal 2.8 kb of the genome and was designated "3'-end probe"; clone HIN2 (a subclone of C7, in pGEM-4), which contains bases 995 through 1249 (of the sequence shown in Fig. 2) and was designated "9.5 probe"; clone EA13.8 (a subclone of B12, in pGEM-4) which contains bases 658 through 868 and was designated "12.7 probe"; and clone A12 (in pUC19), which includes bases 1 through 548 and was designated "4.9 probe." Probes were prepared by nick-translating entire insert-containing plasmid DNA in the presence of [ $\alpha$ - $^{32}$ P]dNTP.

*Construction of plasmids and expression analyses of the 12.7- and 9.5-kDa proteins.* For testing the translatability of the 12.7- and 9.5-kDa ORFs in the upstream position, and the 9.5-kDa ORF in the downstream position, constructs were made in the pGEM-4 vector such that transcripts containing the tandem sequence 5'-12.7-kDa ORF-9.5-kDa ORF-M ORF-3' or the sequence 5'-9.5-kDa ORF-M ORF-3' were made under the control of the T7 RNA polymerase promoter. A plasmid with the 12.7-kDa ORF in the upstream position was made by subcloning the blunt-ended 1.8-kb *Acc*I fragment from clone B12 (from base 630 in Fig. 2 to the 3' end of clone B12) into the *Hin*clI site of the pGEM-4 vector. The resulting clone, pT7-12.7-9.5-M, yielded sense transcripts with T7 polymerase. A plasmid with the 9.5-kDa ORF in the upstream position was prepared by removing a *Hin*clI fragment from clone pT7-12.7-9.5-M that begins in the polylinker region of the vector and ends 20 bases upstream from the 9.5-kDa ORF. Transcription of this plasmid was under control of the T7 promoter and the clone was designated pT7-9.5-M. pT7-12.7-9.5-M and pT7-9.5-M plasmids were linearized with *Eco*RI, and transcripts were prepared using T7 polymerase (Krieg and Melton, 1984).

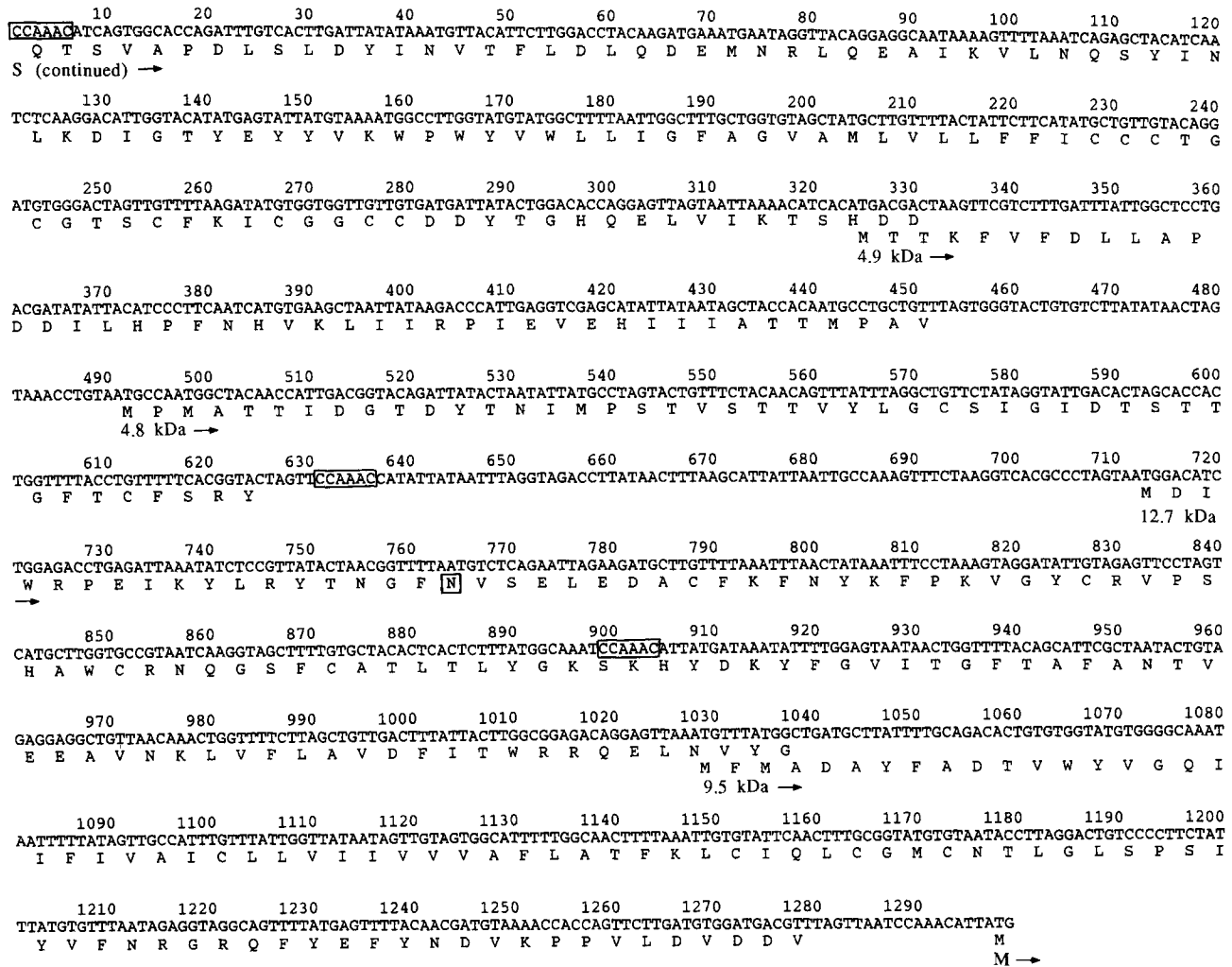


FIG. 2. Nucleotide sequence between the S and M genes and deduced amino acid sequence of the potential 4.9-, 4.8-, 12.7-, and 9.5-kDa proteins. The nucleotide sequence shown begins with the CYAAAC consensus sequence, which starts 331 bases upstream from the TAA stop codon of the S gene [3301 bases from the poly(A) tail] and ends with the ATG initiation codon of the M gene. Consensus CYAAAC sequences are boxed. Open reading frames are identified. The potential N-linked glycosylation site in the 12.7-kDa ORF is boxed. The GenBank accession number of the nucleotide sequence is M31054.

For *in vitro* translation, approximately 1  $\mu$ g of transcript was translated in a wheat germ cell-free lysate (Promega) using 1 mCi/ml [ $^{35}$ S]cysteine (>800 Ci/mole; ICN) as the radiolabeled precursor. Products either were analyzed directly by SDS-polyacrylamide gel electrophoresis in a 20% gel using the protocol of Giulian *et al.* (1985) or were immunoprecipitated first using the protocol of Anderson and Blobel (1983).

For *in vivo* expression analysis, cells were infected with a m.o.i. of 5, incubated for 48 hr, and examined for cytoplasmic immunofluorescence after fixation with methanol or, for surface immunofluorescence, after fixation with 4% paraformaldehyde (Kaariainen *et al.*, 1983).

Monospecific, polyclonal rabbit antibody prepared against the MHV A59 9.6-kDa protein that had been

expressed in *Escherichia coli* (Leibowitz *et al.*, 1988) was a kind gift from Dr. J. Leibowitz, University of Texas Health Science Center, Houston, and Dr. S. Weiss, University of Pennsylvania, Philadelphia, and was used for both immunoprecipitation and immunofluorescence studies.

## RESULTS

*Nucleotide sequence of the region between the spike and membrane protein genes identifies open reading frames for potential proteins of 4.9, 4.8, 12.7, and 9.5 kDa.* Analysis of the structural protein genes of BCV (Abraham *et al.*, 1990; Kienzie *et al.*, 1990; Lapps *et al.*, 1987; Parker *et al.*, 1989) established that they are colinear with the homologous genes of the MHV

genome. Between the S and M genes of BCV there are 962 bases that, when translated by computer, yield open reading frames for proteins of 4.9, 4.8, 12.7, and 9.5 kDa, in that order (Fig. 2). The 4.9 and 4.8 ORFs appear to be 5'-proximal tandem ORFs on mRNA 4, whereas the 12.7- and 9.5-kDa ORFs appear as 5'-proximal ORFs on mRNAs 5 and 5-1, respectively (see below). All four ORFs are therefore potentially expressed as proteins in infected cells.

The ORFs for the proteins of 4.9 and 4.8 kDa appear to have arisen from a single base deletion near the middle of an 11-kDa ORF sequence. This notion is derived from the observation that a nucleotide inserted anywhere between bases 452 and 490 in the sequence shown in Fig. 2, such that an open reading frame is retained, converts the 4.9- and 4.8-kDa ORFs into a single 11-kDa ORF. A nucleotide inserted between bases 452 and 453, for example, creates an 11-kDa ORF that shows significant sequence identity throughout its length with the 15.2-kDa ORF of MHV JHM (32%) (Skinner and Siddell, 1985), for which a protein product has been identified (Ebner *et al.*, 1988), and with the 11.7-kDa ORF of MHV A59 (34%) (S. Weiss, personal communication). To test the possibility that a sequencing error gave rise to only an apparent deletion in the BCV sequence, a separately derived clone, G6 (Fig. 1), was sequenced for this region of the genome. No sequence difference was found, thereby confirming the discontinuity of the 4.9- and 4.8-kDa ORFs on the genome. The properties of the 4.9- and 4.8-kDa ORFs are therefore discussed separately.

The 4.9-kDa ORF sequence overlaps by 8 bases at its beginning with the putative S gene (Abraham *et al.*, 1990). The length of the 5' untranslated region of mRNA 4 is not known, but the CYAAAC consensus sequence (which begins at base 1 in Fig. 2, 300 bases upstream from the spike protein stop codon) predicts that the 5' untranslated region will be approximately 395 bases in length (including the leader of an estimated 80 bases). The 43-amino-acid protein has a predicted molecular weight of 4911, and a net charge of  $-0.5$  at neutral pH. The hydrophobic N and C termini (Fig. 3) are unlikely to serve as signal peptide or transmembrane anchor regions because of insufficient length and hydrophobicity (Kyte and Doolittle, 1982; Von Heijne, 1985). Despite its amino acid sequence similarity to the amino-terminal portion of the MHV JHM 15.2-kDa protein, the 4.9-kDa ORF shows three contrasting features. For MHV, the CYAAAC consensus sequence begins 33 nucleotides downstream from the S protein stop codon, and the initiation codon for the 15.2-kDa protein begins 92 nucleotides downstream from this site (i.e., not within the S gene sequence) (Schmidt *et al.*, 1987). Finally, the N-terminal region of

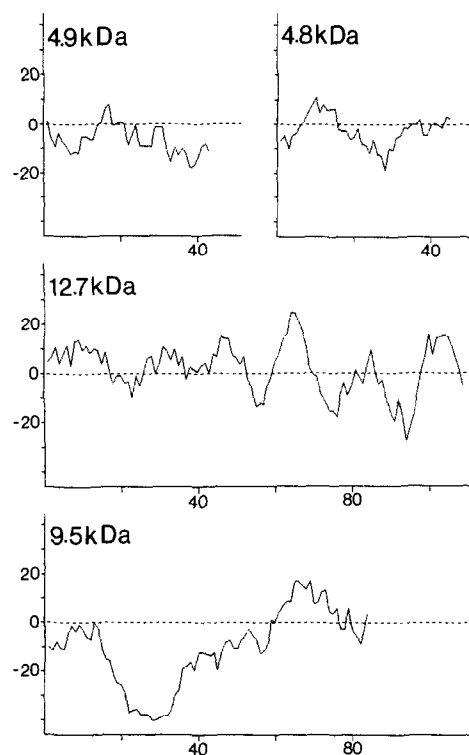


Fig. 3. Hydrophobicity plots of the potential 4.9-, 4.8-, 12.7-, and 9.5-kDa proteins. Hydrophobicity was determined by the method of Kyte and Doolittle (1982) using a 9-amino-acid window. Hydrophobic regions are plotted below the median line representing grand average hydrophobicity.

the MHV JHM 15.2-kDa protein contains a very long hydrophobic region (39 amino acids) of sufficient hydrophobicity to serve as a signal peptide or transmembrane anchor (Skinner and Siddell, 1985).

The 4.8-kDa ORF is predicted to begin approximately 570 bases downstream from the 5' end of mRNA 4. Of the two methionine codons near the beginning of this ORF (at positions 1 and 3), the second is in a more preferred context for initiation of translation although both are considered to be suboptimal (Kozak, 1989). The 4.8-kDa ORF predicts a 45-amino-acid protein having a molecular weight of 4823 and a net charge of  $-2$  at neutral pH. It possesses one central hydrophobic region of sufficient hydrophobicity but probably of insufficient length to be a transmembrane domain (Fig. 3). Like the C terminus of the MHV JHM 15.2-kDa protein, the 4.8-kDa ORF is threonine rich (representing 24% of its amino acids), but unlike the 15.2-kDa protein, it is not basic at its C terminus.

The 12.7-kDa ORF predicts a 109-amino-acid protein having a molecular weight of 12,749. The CCAAAC consensus sequence beginning at base 631 (Fig. 2) predicts that the transcript for the 12.7-kDa protein will have a 5' untranslated region of approximately 160 bases, including the leader. The deduced 12.7-

kDa protein has a net charge of +5 at neutral pH and is therefore basic, but there is no obvious clustering of basic residues in any part of the molecule. There is one potential asparagine-linked glycosylation site at amino acid position 18. A hydrophobicity plot (Fig. 3) illustrates that amino acids 86 through 99 are of sufficient length and hydrophobicity to be a transmembrane domain. The 12.7-kDa protein has amino acid sequence identities of 50 and 49%, respectively, with the 12.4- and 13-kDa proteins of MHV JHM and A59 (Budzilowicz and Weiss, 1987; Skinner *et al.*, 1985), and like these, has only one methionine and this is derived from the initiation codon.

The 9.5-kDa ORF predicts an 84-amino-acid protein with a molecular weight of 9543. The CCAAAC consensus sequence beginning 122 bases upstream from the first potential start codon (Fig. 2) predicts a separate transcript for this protein that would have a 5' untranslated region of approximately 205 bases. The first and third codons of the predicted protein are for methionine and it is unknown which of these initiates synthesis of the protein. The second methionine codon is in a more preferred context for initiation of translation although both are considered to be suboptimal (Kozak, 1989). Fifty-three percent of the amino acids in the 9.5-kDa protein are hydrophobic and these are concentrated between amino acids 17 and 62 (Fig. 3). They give rise to an extremely hydrophobic region containing few charged amino acids and this is a potential transmembrane domain. The C-terminal one-third of the protein is hydrophilic and has a net negative charge. The BCV 9.5-kDa protein has amino acid sequence identities of 65 and 62%, respectively, with the 10.2- and 9.6-kDa proteins of MHV JHM and A59 (Budzilowicz and Weiss, 1987; Skinner *et al.*, 1985).

*Northern hybridization analyses identify mRNA species 4, 5, and 5-1 which have ORFs for the 4.9/4.8-, 12.7-, and 9.5-kDa proteins at their respective 5'-proximal ends.* We have previously identified eight BCV-specific RNA species in infected cells by both metabolic labeling and Northern hybridization experiments, and these include the genome (species number 1) and seven putative subgenomic mRNAs (Keck *et al.*, 1988). To determine which of these might be transcripts beginning with the 4.9/4.8-, 12.7-, or 9.5-kDa genes, subclones containing sequences for the bodies (i.e., 5'-proximal ORF) of the three putative transcripts were used separately as radiolabeled probes in Northern analyses. A separate RNA species for each of these probes was identified between the S (mRNA 3) and M (mRNA 6) mRNA species (Fig. 4). The mRNA species having the 5'-terminal 4.9/4.8-kDa sequence is newly identified by these experiments and is named mRNA 4. [The previously named BCV species 4 and 5 (Keck *et*

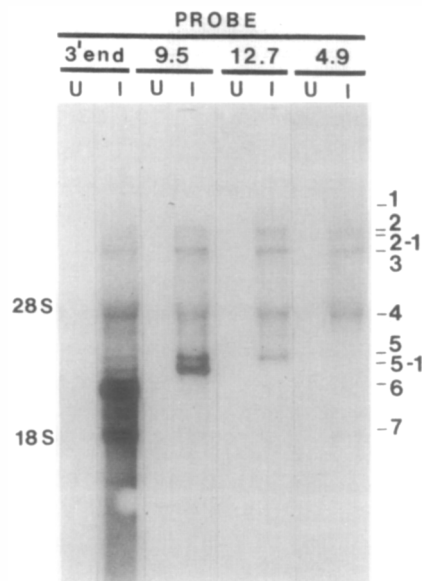


Fig. 4. Northern blot analysis identifying mRNAs 4, 5, and 5-1 as initiating with the 4.9-, 12.7-, and 9.5-kDa ORFs, respectively. Cytoplasmic RNA from infected and uninfected cells was probed with cDNA clones of defined sequence as described under Materials and Methods. U = RNA from uninfected cells. I = RNA from infected cells.

*et al.*, 1988) are renamed here as 5 and 5-1 to correspond to the homologous gene 5 products of MHV (Budzilowicz and Weiss, 1987; Skinner *et al.*, 1985).] mRNA 4 is obscured by 28 S ribosomal RNA in Northern analyses for which the RNA had not been first selected by oligo(dT)-cellulose chromatography (data not shown). A species migrating between mRNAs 6 and 7 (Fig. 4) is known to be a transient-defective RNA that is present in the inoculum stock used in these experiments (M. Hofmann and D. Brian, unpublished data). Thus, a total of nine BCV-specific RNA species (putative mRNAs) have now been identified by Northern analyses.

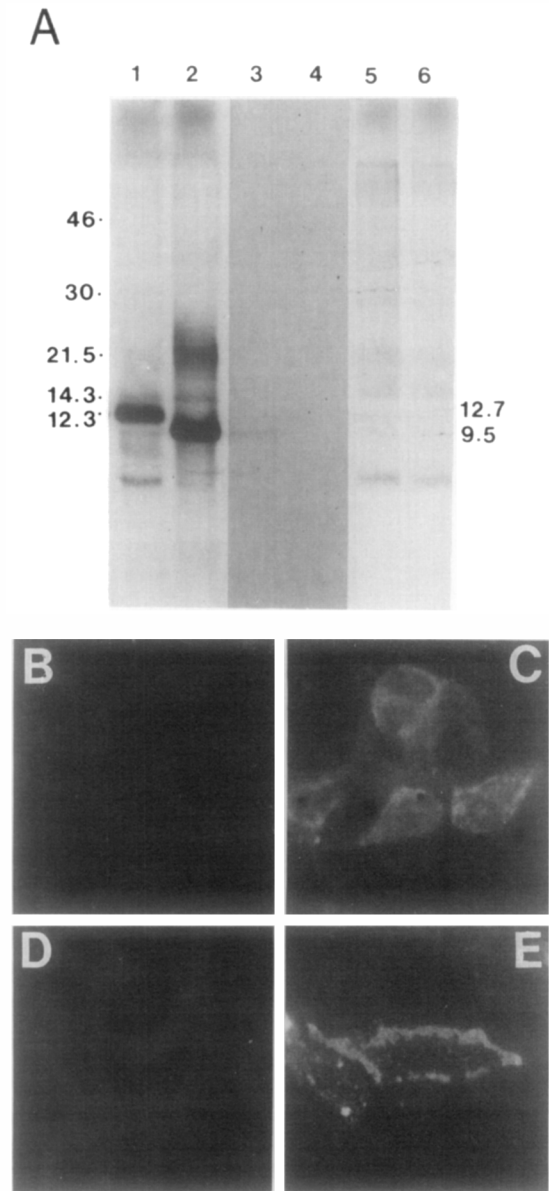
*Translation of transcripts made in vitro demonstrate that the BCV 9.5-kDa protein is readily made from an upstream ORF, and poorly made, if at all, from a downstream ORF.* From our analysis, the transcription pattern for synthesis of the BCV 12.7- and 9.5-kDa proteins is strikingly different from that for the MHV homologs. In MHV, one transcript, mRNA 5, is utilized for the synthesis of both the 13- and 9.6-kDa proteins (Budzilowicz and Weiss, 1987; Leibowitz *et al.*, 1988; Skinner *et al.*, 1985). Supporting this conclusion are the facts that for MHV (1) no transcripts with the 9.6-kDa gene as the 5'-terminal open reading frame are found, (2) no CAAAC consensus sequences are found upstream of the 9.6-kDa open reading frame (i.e., within the 13-kDa ORF), and (3) *in vitro* translation of a synthetic transcript having the 13- and 9.6-kDa open reading frames in tandem demonstrates that synthesis of the 9.6-kDa

downstream open reading frame is the preferred translation product. To analyze the significance of the differences between BCV and MHV regarding expression of the 9.5-kDa protein, we first established that the BCV 9.5-kDa protein is made during virus infection by seeking immunofluorescent labeling of infected cells with antiserum prepared against the MHV A59 9.6-kDa protein (Leibowitz *et al.*, 1988). Both internal and surface immunofluorescence patterns were found and they are similar to those in MHV-infected cells (Figs. 5C and E). The BCV 9.5-kDa protein is therefore made during virus infection. To test whether the BCV 9.5-kDa protein can be made from a downstream tandem transcript (i.e., a structure that mimics MHV mRNA 5), *in vitro* transcripts from pT7-12.7-9.5-M were translated and the results are shown in Fig. 5A, lane 1. The vast majority of product from the pT7-12.7-9.5-M transcript was a protein of 12.7 kDa, and essentially no product of 9.5 kDa was made. Neither was a product the size of M protein produced. To test the translatability of the 9.5-kDa ORF, transcripts from pT7-9.5-M (i.e., a structure that mimics BCV mRNA 5-1) were translated and abundant amounts of the 9.5-kDa protein were produced (Fig. 5A, lane 2). Interestingly, with this transcript moderate amounts of a 21.5-kDa protein were also made which could represent the membrane protein since the unglycosylated form of this protein has a molecular mass of 22 kDa (Lapps *et al.*, 1987). The identity of the 9.5-kDa protein product was confirmed by immunoprecipitation with MHV 9.6-kDa protein-specific antiserum (Fig. 5A, lane 3). There appears, therefore, to be little or no downstream initiation of 9.5-kDa protein synthesis in the BCV sequence, but good synthesis when the 9.5-kDa open reading frame is the 5'-proximal open reading frame.

## DISCUSSION

We have described genes for potential proteins of 4.9, 4.8, 12.7, and 9.5 kDa encoded between the S and M genes of BCV, and have demonstrated the existence of three mRNAs, species 4, 5, and 5-1, that potentially express these proteins in infected cells. Species 4 has not been described before and it appears to encode the 4.9- and 4.8-kDa open reading frames in tandem at its 5' end, suggesting that if the 4.8-kDa protein is made, it is translated from a downstream open reading frame.

One feature of transcripts for the putative nonstructural proteins was found to contrast sharply with transcripts for the structural proteins. Whereas mRNAs for the BCV HE, S, M, and N structural proteins have initiation codons beginning respectively 88, 82, 80, and 86 bases downstream from their 5' ends (assuming a



**Fig. 5.** *In vitro* synthesis and immunodetection of the 9.5-kDa protein. (A) Lane 1, *in vitro* translation of transcripts on which the 9.5-kDa protein comes from the second (downstream) ORF (i.e., transcripts of pT7-12.7-9.5-M). Lane 2, *in vitro* translation of transcripts on which the 9.5-kDa protein comes from the first (upstream) ORF (i.e., transcripts of pT7-9.5-M). Lane 3, immunoprecipitate of product made in lane 2. Lane 4, control for immunoprecipitation using nonimmune rabbit serum. Lane 5, *in vitro* translation of transcripts from the pGEM-4 vector only. Lane 6, *in vitro* translation with no RNA added. (B) Internal immunofluorescence of uninfected cells. (C) Internal immunofluorescence of BCV-infected cells. (D) Surface immunofluorescence of uninfected cells. (E) Surface immunofluorescence of BCV-infected cells.

leader sequence of 80 bases for BCV), mRNAs for the 4.9-, 4.8-, 12.7-, and 9.5-kDa nonstructural proteins have corresponding predicted sequences of 395, 570, 160, and 205 bases. The much longer 5' untranslated

region on the mRNAs for the nonstructural proteins suggests that a different strategy of translation may be used by the nonstructural proteins. A mechanism other than ribosomal scanning (Kozak, 1989), for example, a mechanism such as that utilized by picornaviruses in which downstream assembly of a ribosomal complex allows the bypassing of a very long (745-base) 5' untranslated region (Pelletier and Sonenberg, 1989), could aid in the synthesis of some coronaviral nonstructural proteins. Certainly, potential upstream methionine start sites (12 for the 4.9-kDa protein, 15 for the 4.8-kDa protein, 0 for the 12.7-kDa protein, and 1 for the 9.5-kDa protein) require bypassing during synthesis of the nonstructural proteins, whereas there is no such requirement during synthesis of the structural proteins since there are no potential start codons within the BCV leader (M. A. Hofmann and D. A. Brian, unpublished). This pattern is also seen by the downstream translation of the MHV JHM 10.2-kDa and the MHV A59 9.6 protein (Skinner *et al.*, 1985; Budzilowicz and Weiss, 1987).

A most striking finding from our data is that the BCV 9.5-kDa protein, unlike its antigenic homolog in MHV, appears to be synthesized from a transcript on which it is in the 5'-terminal position. From a large body of coronavirus sequence information (see Spaan *et al.*, 1988, for review), a CYAAAC consensus sequence residing upstream (within the 12.7-kDa ORF) in BCV strongly predicts that a separate transcript for the 9.5-kDa protein will be made. The difference between BCV and MHV on this point suggests the possibility that coronavirus transcription start sites are of a pleiotropic nature. Did BCV evolve from a MHV-like progenitor and gain a CYAAAC sequence, and thus a new transcript, or did MHV lose the sequence and develop a compensating mechanism for synthesis of the 9.6-kDa protein (i.e., a downstream initiation site for translation)? It will be interesting to learn if such transcriptional start sites can be engineered. It may be that more than the minimal CYAAAC sequence is required for initiating transcription since within the S gene of BCV two other CYAAAC sequences have been found that apparently do not initiate transcription (Abraham *et al.*, 1990).

The functions of the putative 4.9-, 4.8-, 12.7-, and 9.5-kDa proteins are not known. Since some of the small coronavirus nonstructural proteins have both hydrophobic and highly charged domains, it has been proposed that they may serve a siting or anchoring function for structural proteins during virus assembly, or may maintain a membrane association of the viral polymerase during replication (Bournsnell and Brown, 1984; Bournsnell *et al.*, 1985; Leibowitz *et al.*, 1988; Skinner and Siddell, 1985; Skinner *et al.*, 1985). It is interesting to note that although significant regions of

amino acid sequence identity are found among the homologous structural proteins of evolutionarily divergent coronaviruses (BCV, IBV, and TGEV), no identity among the nonstructural proteins was found in our computer search except for a small region in the 9.2-kDa protein of TGEV (amino acids 40–55) that shares 43% sequence identity with amino acids 43–58 of the BCV 9.5-kDa protein. The region of homology is

```

BCV  | Q L C G M C N T L G L S P S |
      | | | | | | | | | | | |
TGEV | K L C M V C C N L G R T V I |

```

The 9.5-kDa protein of BCV, 9.6-kDa protein of MHV, and 9.2-kDa protein of TGEV may therefore serve a homologous function in virus replication.

By using cross-reacting rabbit antiserum against the MHV 9.6-kDa protein, we have shown that the BCV 9.5-kDa protein, like its MHV counterpart, is expressed on the surface of virus-infected cells as well as internally. It may therefore be an integral membrane protein as its hydrophobicity profile suggests. Because of its cell surface location, it may be an important immunogen to exploit for modulation of BCV infection. The role of this putative nonstructural protein in immune responses to BCV infection should therefore be investigated.

## ACKNOWLEDGMENTS

We thank Dr. S. Weiss for permission to cite unpublished results, and Dr. B. Hogue for many helpful discussions. This work was supported by Grant AI-14736 from the National Institute of Allergy and Infectious Diseases, and by Grant 82-CRSR-2-1090 from the U.S. Department of Agriculture. T.E.K. and W.L. were predoctoral trainees on Grant T32-AI07123 from the National Institutes of Health.

## REFERENCES

- ABRAHAM, S., KIENZLE, T. E., LAPPS, W., and BRIAN, D. A. (1990). Deduced sequence of the bovine coronavirus spike protein and identification of the internal proteolytic cleavage site. *Virology* **176**, 296–301.
- ANDERSON, D. J., and BLOBEL, G. (1983). Immunoprecipitation of proteins from cell-free translations. In "Methods in Enzymology" (S. Fleischer and B. Fleischer, Eds.), Vol. 96, pp. 111–120. Academic Press, Orlando, FL.
- ARMSTRONG, J., SMEEKENS, S., and ROTTIER, P. (1983). Sequence of the nucleocapsid gene from murine coronavirus MHV-A59. *Nucl. Acids Res.* **11**, 833–891.
- BOURNSNELL, M. E. G., BINNS, M. M., and BROWN, T. D. K. (1985). Sequencing of coronavirus IBV genomic RNA: Three open reading frames in the 5' "unique" region of mRNA D. *J. Gen. Virol.* **66**, 2253–2258.
- BOURNSNELL, M. E. G., and BROWN, T. D. K. (1984). Sequencing of coronavirus IBV genomic RNA: A 195-base open reading frame encoded by mRNA B. *Gene* **29**, 87–92.
- BOURNSNELL, M. E. G., BROWN, T. D. K., FOULDS, I. J., GREEN, P. F., TOMLEY, F. M., and BINNS, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *J. Gen. Virol.* **66**, 57–77.



- BUDZILOWICZ, C. J., and WEISS, S. R. (1987). *In vitro* synthesis of two polypeptides from a nonstructural gene of coronavirus mouse hepatitis virus strain A59. *Virology* **157**, 509–515.
- COX, G. J., PARKER, M. D., and BABIUK, L. A. (1989). The sequence of cDNA of bovine coronavirus 32K nonstructural gene. *Nucl. Acids Res.* **17**, 5847.
- DENNIS, D. E., and BRIAN, D. A. (1982). RNA-dependent RNA polymerase activity in coronavirus-infected cells. *J. Virol.* **42**, 153–164.
- EBNER, D., RAABE, T., and SIDDELL, S. G. (1988). Identification of the coronavirus MHV-JHM mRNA 4 product. *J. Gen. Virol.* **69**, 1041–1050.
- GIULIAN, G. G., SHANAHAN, M. F., GRAHAM, J. M., and MOSS, R. L. (1985). Resolution of low molecular weight polypeptides in a non-urea sodium dodecyl sulfate slab polyacrylamide gel system. *Fed. Proc.* **44**, 686.
- HENIKOFF, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–359.
- KAARIANEN, L., VIRTANEN, I., SARASTE, J., and KERANEN, S. (1983). Transport of virus membrane glycoproteins, use of temperature-sensitive mutants and organelle-specific lectins. In "Methods in Enzymology" (S. Fleischer and B. Fleischer, Eds.), Vol. 96, pp. 453–465. Academic Press, Orlando, FL.
- KAPKE, P. A., and BRIAN, D. A. (1986). Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. *Virology* **151**, 41–49.
- KAPKE, P. A., TUNG, F. Y. T., and BRIAN, D. A. (1988). Nucleotide sequence between the peplomer and matrix protein genes of the porcine transmissible gastroenteritis coronavirus identifies three large open reading frames. *Virus Genes* **2**, 293–294.
- KECK, J. G., HOGUE, B. G., BRIAN, D. A., and LAI, M. M. C. (1988). Temporal regulation of bovine coronavirus RNA synthesis. *Virus Res.* **9**, 343–356.
- KHANDJIAN, E. W. (1986). UV crosslinking of RNA to nylon membrane enhances hybridization signals. *Mol. Biol. Rep.* **11**, 107–115.
- KIENZLE, T. E., ABRAHAM, S., HOGUE, B. G., and BRIAN, D. A. (1990). Structure and orientation of expressed bovine coronavirus hemagglutinin-esterase protein. *J. Virol.* **64**, 1834–1838.
- KING, B., and BRIAN, D. A. (1982). Bovine coronavirus structural proteins. *J. Virol.* **42**, 700–707.
- KOZAK, M. (1989). The scanning model for translation: An update. *J. Cell Biol.* **108**, 229–241.
- KRIEG, P. A., and MELTON, D. A. (1984). Functional messenger RNAs are produced by SP6 in *in vitro* transcription of cloned cDNAs. *Nucl. Acids Res.* **12**, 7057–7070.
- KYTE, J., and DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132.
- LAPPS, W., HOGUE, B. G., and BRIAN, D. A. (1987). Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. *Virology* **157**, 47–57.
- LEIBOWITZ, J. L., PERLMAN, S., WEINSTOCK, G., DEVRIES, J. R., BUDZILOWICZ, C., WEISSEMANN, J. M., and WEISS, S. R. (1988). Detection of a murine coronavirus nonstructural protein encoded in a downstream open reading frame. *Virology* **164**, 156–164.
- LIZARDI, P. M. (1983). Methods for the preparation of messenger RNA. In "Methods in Enzymology" (S. Fleischer and B. Fleischer, Eds.), Vol. 96, pp. 24–38. Academic Press, Orlando, FL.
- MANIATIS, T., FRITSCH, E. F., and SAMBROOK, J. (1982). "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- MAXAM, A. M., and GILBERT, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. In "Methods in Enzymology" (L. Grossman and K. Moldave, Eds.), Vol. 65, pp. 499–560. Academic Press, Orlando, FL.
- PACHUK, C. J., BREDENBEEK, P. J., ZOLTICK, P. W., SPAAN, W. J. M., and WEISS, S. R. (1989). Molecular cloning of the gene encoding the putative polymerase of mouse hepatitis coronavirus, strain A59. *Virology* **171**, 141–148.
- PARKER, M. D., COX, G. J., DEREGT, D., FITZPATRICK, D. R., and BABIUK, L. A. (1989). Cloning and *in vitro* expression of the gene for the E3 hemagglutinin glycoprotein of bovine coronavirus. *J. Gen. Virol.* **70**, 155–164.
- PELLETIER, J., and SONENBERG, N. (1989). Internal binding of eucaryotic ribosomes of poliovirus RNA: Translation in HeLa cell extracts. *J. Virol.* **63**, 441–444.
- QUEEN, C., and KORN, K. J. (1984). A comprehensive sequence analysis program for the IBM personal computer. *Nucl. Acids Res.* **12**, 581–599.
- SCHMIDT, I., SKINNER, M., and SIDDELL, S. (1987). Nucleotide sequence of the gene encoding the surface projection glycoprotein of coronavirus MHV-JHM. *J. Gen. Virol.* **68**, 47–56.
- SETHNA, P. B., HUNG, S.-L., and BRIAN, D. A. (1989). Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proc. Natl. Acad. Sci. USA* **86**, 5626–5630.
- SKINNER, M. A., EBNER, D., and SIDDELL, S. G. (1985). Coronavirus MHV-JHM mRNA 5 has a sequence arrangement which potentially allows translation of a second, downstream open reading frame. *J. Gen. Virol.* **66**, 581–592.
- SKINNER, M. A., and SIDDELL, S. G. (1985). Coding sequence of coronavirus MHV-JHM mRNA 4. *J. Gen. Virol.* **66**, 593–596.
- SPAAN, W., CAVANAGH, D., and HORZINEK, M. C. (1988). Coronaviruses: Structure and genome expression. *J. Gen. Virol.* **69**, 2939–2952.
- THOMAS, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* **77**, 5205–5209.
- VON HEIJNE, G. (1985). Signal sequences: The limits of variation. *J. Mol. Biol.* **184**, 99–105.