J. gen. Virol. (1989), **70**, 155–164. *Printed in Great Britain*

Key words: *coronavirus, bovine/nucleotide sequence/haemagglutinin*

Cloning and *in vitro* Expression of the Gene for the E3 Haemagglutinin Glycoprotein of Bovine Coronavirus

By MICHAEL D. PARKER,* GRAHAM J. COX, DIRK DEREGT,† DAVID R. FITZPATRICK AND LORNE A. BABIUK

Veterinary Infectious Disease Organization and Department of Veterinary Microbiology, University of Saskatchewan, 124 Veterinary Road, Saskatoon, Saskatchewan S7N 0W0, Canada

(*Accepted 4 October 1988*)

SUMMARY

A cDNA clone representing the gene for the E3 glycoprotein, the haemagglutinin, of bovine coronavirus was isolated from a plasmid cDNA library of the viral genome and sequenced. The gene is located immediately 5′ of the E2 glycoprotein gene on the viral genome. Nucleotide sequencing of the E3 gene predicts a polypeptide of 424 amino acids with an M_r of 47K. *In vitro* translation of mRNA transcribed from the cloned E3 gene yielded a polypeptide of M_r 45K, similar to that predicted from the nucleotide sequence. In the presence of microsomal membranes, the *in vitro* product was cotranslationally processed to a 62K polypeptide which comigrated on SDS–polyacrylamide gels with the E3 monomer (gp62) obtained from virus-infected cells. Both the 45K and 62K polypeptides were immunoprecipitated with E3-specific monoclonal antibodies, confirming the identity of the gene as that encoding the E3 glycoprotein. Finally, only monoclonal antibodies to the E3 protein inhibited haemagglutination by the virus thus confirming its identity as the haemagglutinin of bovine coronavirus.

INTRODUCTION

Bovine coronavirus (BCV) is a member of the family Coronaviridae, a group of enveloped, positive-stranded RNA viruses which infect human and animal hosts, causing a number of diverse diseases (Siddell *et al.*, 1983). In neonatal calves, BCV causes an acute and often fatal enteritis characterized by intestinal villous atrophy and diarrhoea (Mebus, 1978).

Coronaviruses multiply in the cytoplasm of infected cells and produce genome size (20 to 27 kb) and multiple subgenomic mRNAs which form a 3′ coterminal nested set (Siddell *et al.*, 1983; Jacobs *et al.*, 1986; de Groot *et al.*, 1987; Boursnell *et al.*, 1987; Keck *et al.*, 1988; Stern & Kennedy, 1980). In general, only the 'unique' 5′ end of each mRNA (sequences not contained in the smaller mRNAs) are translated.

The virions of most coronaviruses are composed of three structural proteins designated N, the nucleoprotein, E1, a small matrix glycoprotein, and E2, the glycoprotein which forms the large characteristic club-shaped peplomers (Siddell *et al.*, 1983). BCV and the other haemagglutinating mammalian coronaviruses, human respiratory coronavirus OC-43 and haemagglutinating encephalomyelitis virus of swine, contain a third glycoprotein designated E3 (Pocock & Garwes, 1977; Storz *et al.*, 1981; King & Brian, 1982; Hogue *et al.*, 1984; Hogue & Brian, 1986; Deregt *et al.*, 1987). The BCV E3 protein is a disulphide-linked dimer of 124K to 140K, reducible to 62K to 65K subunits. Recently we have identified the precursor to this protein in pulse–chase experiments as a 59K *N*-glycosylated polypeptide (monomer) (Deregt *et al.*, 1987). This protein rapidly dimerizes by intermolecular disulphide bridging before further glycosylation yields the mature form of the protein (gp124). These results indicated that the E3 protein was probably a dimer made of identical subunits. The apoprotein (monomer) has an M_r of 42·5K as determined

† Present address: Animal Diseases Research Institute, P.O. Box 640, Lethbridge, Alberta T1J 3Z4, Canada.

0000-8514

by studies utilizing tunicamycin, an inhibitor of *N*-linked glycosylation (Hogue & Brian, 1987). Thus, the sugars account for about 35% of the M_r of the protein.

Although a function for the E3 protein has not yet been determined, two important properties of this protein are known. First, E3 has been identified as the BCV haemagglutinin (King *et al.*, 1985). In this report we show that only monoclonal antibodies to the E3 protein inhibit virus haemagglutination thus confirming its identity as the haemagglutinin. Second, the E3 protein, in addition to the BCV E2 protein, can elicit the production of both *in vitro* and *in vivo* neutralizing antibodies (Deregt & Babiuk, 1987; D. Deregt *et al.*, unpublished data). Therefore, the available evidence indicates that this protein has an important role in BCV–cell interactions.

In this report we present the sequence of the unique E3 gene of BCV. The nucleotide sequence predicts a polypeptide with features characteristic of a membrane glycoprotein: a hydrophobic amino-terminal signal sequence and a carboxy-terminal membrane-anchoring domain. The cloned gene has also been utilized to direct the *in vitro* synthesis and glycosylation of a polypeptide which is recognized by E3-specific monoclonal antibodies. The results indicate that the E3 protein in its mature form is a homodimer.

METHODS

Molecular cloning of the BCV E3 gene. The purification of the Quebec strain of BCV (Dea *et al.*, 1980) by polyethylene glycol precipitation and gradient centrifugation was performed as described previously (Deregt & Babiuk, 1987). Genomic RNA was prepared by phenol:chloroform:isoamyl alcohol (50:48:2) extraction and ethanol precipitation. Synthesis of double-stranded cDNA was carried out essentially as described (Gubler & Hoffman, 1983; Maniatis *et al.*, 1982). After addition of *Bam*HI linkers and digestion with *Bam*HI restriction endonuclease, cDNA of 3000 bp and greater was purified by agarose gel electrophoresis, ligated into *Bam*HI-cleaved pTZ19R (Pharmacia) and used to transform *Escherichia coli* strain DH1. Recombinants were identified by *in situ* hybridization (Grunstein & Hogness, 1975) using a ^{32}P-labelled cDNA probe prepared by randomly primed reverse transcription of viral genomic RNA.

DNA sequencing. Sequencing was by the dideoxy termination method (Sanger *et al.*, 1977) after generation of a series of overlapping deletions (Henikoff, 1984) and transformation of *E. coli* JM105. Sequence analysis was carried out using the DNA/protein analysis system of International Biotechnologies.

Northern blotting and hybridization. MDBK cells, grown in MEM supplemented with 10% foetal bovine serum, were infected with BCV at a multiplicity of 1 p.f.u./cell. After 24 h at 37 °C, the monolayers were scraped into phosphate-buffered saline and pelleted at 1000 *g* for 5 min. The cells were lysed by vortexing in a minimal volume of 50 mM-Tris–HCl pH 8·0, 100 mM-KCl, 5 mM-EDTA, 1% NP40 and 0·5% sodium deoxycholate. Insoluble material was removed by centrifugation for 1 min at 12000 *g* and the supernatant was phenol-extracted and ethanol-precipitated. Polyadenylated RNA was prepared by oligo(dT)–cellulose chromatography (Aviv & Leder, 1972). The RNA was electrophoresed on 1% formaldehyde–agarose gels (Lehrach *et al.*, 1977) and electroblotted onto Zeta-probe membranes (Bio-Rad). Radiolabelled probes were prepared by random priming DNA synthesis with the Klenow fragment of DNA polymerase I using denatured virus-specific restriction fragments as template. Hybridization was carried out at 42 °C for 24 h in 50% formamide, 5 × SSC (1 × SSC is 0·15 M-NaCl, 0·015 M-sodium citrate), 5 × Denhardt's solution and 150 μg/ml yeast tRNA. The blots were washed for 1 h at 60 °C in 0·1 × SSC and 0·1% SDS and radiographed.

In vitro *transcription and translation of the BCV E3 gene.* A cassette containing nucleotides 53 to 1350 of the sequence in Fig. 3 was prepared by controlled exonuclease III (Henikoff, 1984) digestion of the 5′ *Bam*HI–*Bgl*II fragment of clone pCVA12H (Fig. 1). After addition of *Bam*HI linkers, the cassette was inserted into the *Bam*HI site of pTZ19R and the nucleotide sequences of the termini of the gene cassette were determined. m^{7}GpppA-capped mRNA was produced by the method of Melton *et al.* (1984) with T7 RNA polymerase after cleavage of the plasmid at the *Acc*I site in the polylinker. The RNA transcripts were translated in rabbit reticulocyte lysates containing 500 μCi/ml [^{35}S]methionine (Amersham; >800 Ci/mmol). The products were immunoprecipitated with a pool of three monoclonal antibodies (BD-98C, KC4-3, KD9-40) specific for the BCV E3 glycoprotein (Deregt & Babiuk, 1987) and analysed by electrophoresis on 13% acrylamide: *N,N′*-diallyltartardiamide (30:1·4) gels using the buffer system of Laemmli (1970).

Haemagglutination and haemagglutination inhibition assay. Haemagglutination of erythrocytes from BALB/c mice by BCV was performed as described previously (King & Brian, 1982). For haemagglutination inhibition testing, 20 haemagglutinating units of virus were mixed with an equal volume of twofold serial dilutions of each monoclonal antibody. After 1 h at room temperature, an equal volume of 0·4% (v/v) erythrocytes in phosphate-buffered saline was added and held at room temperature for 45 min. Haemagglutination inhibition titres are expressed as the highest dilution of monoclonal antibody at which inhibition of haemagglutination was complete.

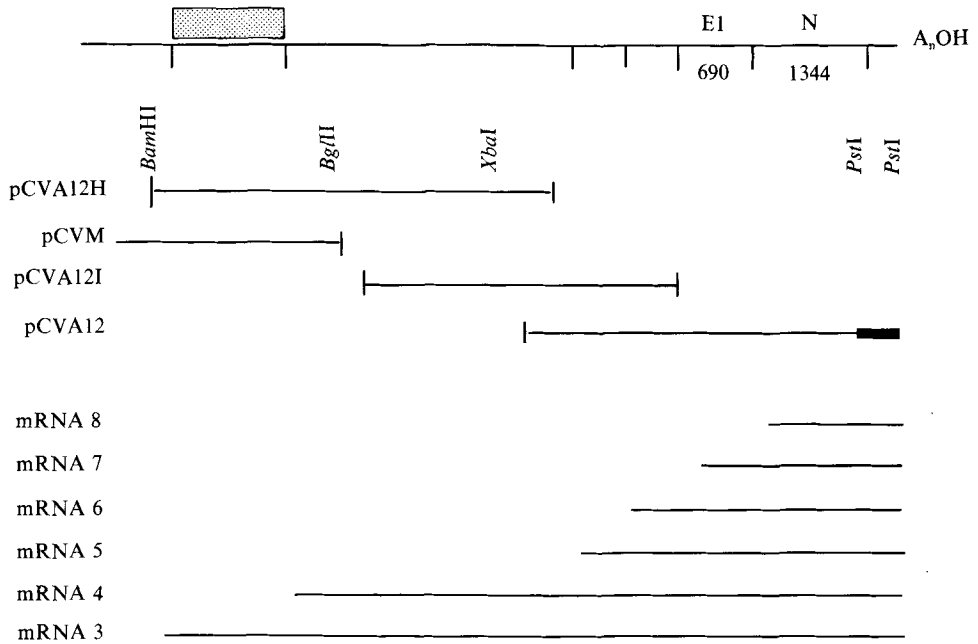

Fig. 1. Organization of the 3′-terminal 8500 nucleotides of the BCV genome. The positions of the N and E1 genes are indicated. A partial restriction map and relative location of representative cDNA clones are indicated. The stippled region of viral genome indicates the sequence shown in Fig. 3.

RESULTS

Cloning and sequencing of the BCV E3 gene

Restriction mapping and nucleotide sequencing of a series of large overlapping cDNA clones yielded a linear map representing approximately 10000 nucleotides of the BCV genome (Fig. 1). In order to determine the proximity of these sequences to the 3′ end of the virion genome, a Northern blot of poly(A)$^+$ RNA from BCV-infected MDBK cells was probed with a 300 bp *Pst*I fragment from the 3′ end of clone pCVA12 (Fig. 1; solid boxed area). Eight species of RNA were detected with this probe (Fig. 2, lane 1), the largest of which was greater than 25 kb in length and apparently the genomic equivalent (Fig. 2, compare lanes 1 and 2). These results indicated that the linear map shown in Fig. 1 extends to nearly the 3′ end of the viral genome given the nested nature of the eight intracellular BCV RNAs (Keck *et al.*, 1988).

Several additional RNAs were occasionally detected (Fig. 2, lane 1) but their origin is not known. They were not likely to be due to the presence of defective interfering particles as analysis of the virion RNA (Fig. 2, lane 2) indicated that a single species of RNA was present in virus used for infection of cells.

Based upon alignment of the genomic RNA and the restriction map shown in Fig. 1, it was predicted that clones pCVA12H and pCVM would contain cDNA sequences representing the unique sequences of mRNAs 3 and 4. RNA 3, previously designated RNA 2a, has been suggested to code for the BCV E3 haemagglutinin protein (Keck *et al.*, 1988) because it is absent from cells infected with murine hepatitis virus (MHV), a closely related coronavirus which lacks a protein analogous to the BCV E3 protein. As predicted, pCVA12H hybridized only to BCV mRNAs 1 to 4 when used to probe RNA from infected cells (Fig. 2, lane 3). The two bands at the positions of the 28S and 18S rRNAs are artefacts due to the large amounts of RNA loaded onto the gel.

The cDNA inserts in pCVA12H and pCVM were sequenced completely in both directions. On the basis of a comparison with the sequence of the MHV-JHM E2 gene (Schmidt *et al.*,

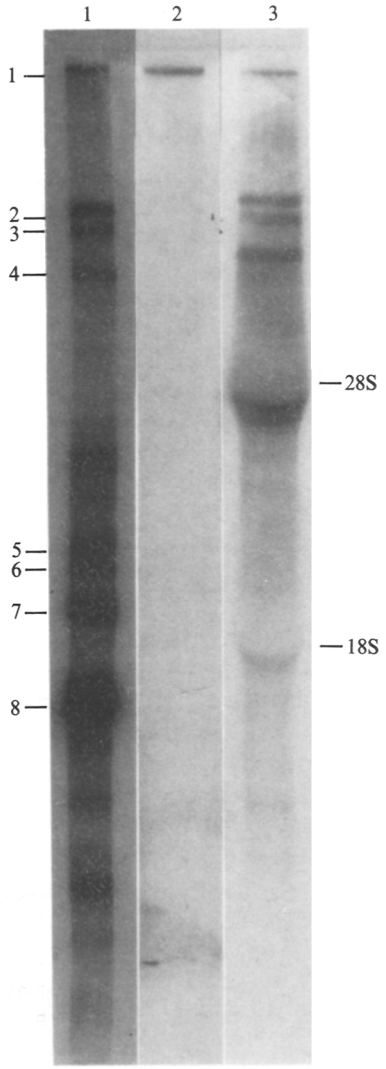

Fig. 2. Northern blot of bovine coronavirus RNA. RNA was electrophoresed on 1% agarose–formaldehyde gels and electroblotted onto nylon membranes. Lane 1, poly(A)$^+$ RNA from BCV-infected MDBK cells probed with the 3′ *Pst*I fragment of clone pCVA12; lane 2, purified virion RNA probed as in lane 1; lane 3, total intracellular RNA from BCV-infected cells probed with clone pCVA12H. The location of 28S and 18S rRNA markers is indicated.

1987), the sequences at the 3′ end of pCVA12H were found to represent most of the sequence of the BCV E2 gene (M. D. Parker, unpublished data), thus confirming that mRNA 4 encodes the BCV E2 precursor polypeptide. Extending upstream of the E2 gene, in the position predicted to contain the unique sequences of mRNA 3, were an additional 1350 nucleotides, which contained a large open reading frame (ORF) of 1272 nucleotides. This ORF terminated 14 nucleotides upstream from the E2 gene and encoded a polypeptide of 424 amino acids with an M_r of 47·6K (Fig. 3).

Beginning 16 nucleotides upstream of the proposed initiation codon at nucleotides 62 to 64 for the E3 polypeptide is the sequence ACTAAAC, which is similar to a conserved intergenic

Fig. 3. Nucleotide sequence of the bovine coronavirus E3 gene. The amino acid sequence of the predicted polypeptide products of the E3 gene and the IORFs IORF1 and IORF2 are also shown. Asterisks indicate conserved intergenic sequences. Underlined Asn residues indicate possible sites for addition of *N*-linked oligosaccharides in the E3 polypeptide product.

```
                                                    *******
TAGATTGTAATTGTTTGGGGTTTTATGAATCTCCAGTTGAAGAAGACTAAACTCAGTGAAAATGTTTTTGCTTCTTAGATTTGTTCTAGTTAGC-94
                                                                  MetPheLeuLeuLeuArgPheValLeuValSer

TGCATAATTGGTAGCCTAGGTTTTGATAACCCTCCTACCAATGTTGTTTCGCATTTAAATGGAGATTGGTTTTTATTTGGTGACAGTCGTTCA-187
CysIleIleGlySerLeuGlyPheAspAsnProProThrAsnValValSerHisLeuAsnGlyAspTrpPheLeuPheGlyAspSerArgSer
                                                        IORF1  MetGluIleGlyPheTyrLeuValThrValValGl

GATTGTAATCATGTTGTTAATACCAACCCCCGTAATTATTCTTATATGGACCTTAATCCTGCCCTGTGTGATTCTGGTAAAATATCATCTAAA-280
AspCysAsnHisValValAsnThrAsnProArgAsnTyrSerTyrMetAspLeuAsnProAlaLeuCysAspSerGlyLysIleSerSerLys
nIleValIleMetLeuLeuIleProThrProValIleIleLeuIleTrpThrLeuIleLeuProCysValIleLeuValLysTyrHisLeuLy

GCTGGCAACTCCATTTTTAGGAGTTTTCACTTTACCGATTTTTATAATTACACAGGCGAAGGTCAACAAATTATTTTTTATGAGGGTCTTAAT-373
AlaGlyAsnSerIlePheArgSerPheHisPheThrAspPheTyrAsnTyrThrGlyGluGlyGlnGlnIleIlePheTyrGluGlyLeuAsn
sLeuAlaThrProPheLeuGlyValPheThrLeuProIlePheIleIleThrGlnAlaLysValAsnLysLeuPhePheMetArgValLeuIl

TTTACGCCTTATCATGCCTTTAAATGCACCACTTCTGGTAGTAATGATATTTGGATGCAGAATAAAGGCTTGTTTTACACTCAGGTTTATAAG-466
PheThrProTyrHisAlaPheLysCysThrThrSerGlySerAsnAspIleTrpMetGlnAsnLysGlyLeuPheTyrThrGlnValTyrLys
eLeuArgLeuIleMetProLeuAsnAlaProLeuLeuValValMetIlePheGlyCysArgIleLysAlaCysPheThrLeuArgPheIleAr

AATATGGCTGTGTATCGCAGCCTTACTTTTGTTAATGTACCATATGTTTATAATGGCTCTGCACAATCTACAGCTCTTTGTAAATCTGGTAGT-559
AsnMetAlaValTyrArgSerLeuThrPheValAsnValProTyrValTyrAsnGlySerAlaGlnSerThrAlaLeuCysLysSerGlySer
gIleTrpLeuCysIleAlaAlaLeuLeuLeuLeuMetTyrHisMetPheIleMetAlaLeuHisAsnLeuGlnLeuPheValAsnLeuValVa

TTAGTTCTAAATAACCCTGCATATATAGCTCGTGAAGCTAATTTTGGGGATTATTATTATAAGGTTGAAGCTGACTTTTATTTGTCAGGTTGT-652
LeuValLeuAsnAsnProAlaTyrIleAlaArgGluAlaAsnPheGlyAspTyrTyrTyrLysValGluAlaAspPheTyrLeuSerGlyCys
l---

GACGAGTATATCGTACCACTTTGTATTTTTAACGGCAAGTTTTTGTCGAATACAAAGTATTATGATGATAGTCAATATTATTTTAATAAAGAC-745
AspGluTyrIleValProLeuCysIlePheAsnGlyLysPheLeuSerAsnThrLysTyrTyrAspAspSerGlnTyrTyrPheAsnLysAsp

ACTGGTGTTATTTATGGTCTCAATTCTACTGAAACCATTACCACTGGTTTTGATTTTAATTGTCATTATTTAGTTTTACCCTCTGGTAATTAT-838
ThrGlyValIleTyrGlyLeuAsnSerThrGluThrIleThrThrGlyPheAspPheAsnCysHisTyrLeuValLeuProSerGlyAsnTyr

TTAGCCATTTCAAATGAGCTATTGTTAACTGTTCCTACGAAAGCAATCTGTCTTAACAAGCGTAAGGATTTTACGCCTGTACAGGTTGTTGAT-931
LeuAlaIleSerAsnGluLeuLeuLeuThrValProThrLysAlaIleCysLeuAsnLysArgLysAspPheThrProValGlnValValAsp

TCACGGTGGAACAATGCCAGGCAGTCTGATAACATGACGGCGGTTGCTTGTCAACCCCCGTACTGTTATTTTCGTAATTCTACTACCAACTAT-1024
SerArgTrpAsnAsnAlaArgGlnSerAspAsnMetThrAlaValAlaCysGlnProProTyrCysTyrPheArgAsnSerThrThrAsnTyr
                                                                                IORF2  Me

GTTGGTGTTTATGATATCAATCATGGGGATGCTGGTTTTACTAGCATACTCAGTGGTTTGTTATATGATTCACCTTGTTTTTCGCAGCAAGGT-1117
ValGlyValTyrAspIleAsnHisGlyAspAlaGlyPheThrSerIleLeuSerGlyLeuLeuTyrAspSerProCysPheSerGlnGlnGly
tLeuValPheMetIleSerIleMetGlyMetLeuValLeuLeuAlaTyrSerValValCysTyrMetIleHisLeuValPheArgSerLysVa

GTTTTTAGGTATGATAATGTTAGCAGTGTCTGGCCTCTCTATTCCTATGGCAGATGCCCTACTGCTGCTGATATTAATACCCCTGATGTACCT-1210
ValPheArgTyrAspAsnValSerSerValTrpProLeuTyrSerTyrGlyArgCysProThrAlaAlaAspIleAsnThrProAspValPro
lPheLeuGlyMetIleMetLeuAlaValSerGlyLeuSerIleProMetAlaAspAlaLeuLeuLeuLeuIleLeuIleProLeuMetTyrLe

ATTTGTGTGTATGATCCGCTACCACTTATTTTGCTTGGCATCCTTTTGGGTGTTGCGGTCATAATTATTGTAGTTTTGTTGTTATATTTTATG-1303
IleCysValTyrAspProLeuProLeuIleLeuLeuGlyIleLeuLeuGlyValAlaValIleIleIleValValLeuLeuLeuTyrPheMet
uPheValCysMetIleArgTyrHisLeuPheCysLeuAlaSerPheTrpValLeuArgSer---

                                                    *******
GTGGATAATGGTACTAGGCTGCATGATGCTTAGACCATAATCTAAAC-1350
ValAspAsnGlyThrArgLeuHisAspAla---
```

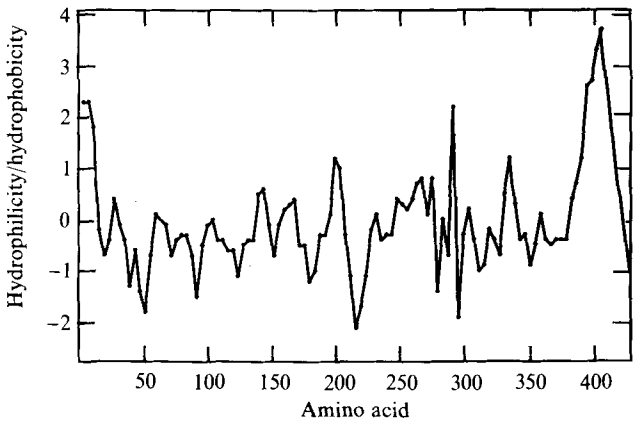

Fig. 4. Hydropathy plot of the predicted polypeptide of the E3 gene. The hydropathicity is calculated with a window of four amino acids. Positive values indicate hydrophobic regions and negative values indicate hydrophilic regions.

sequence upstream of the BCV N and E1 genes (Lapps *et al.*, 1987). A similar sequence, TCTAAAC, is also found beginning eight nucleotides downstream from the termination codon and immediately precedes the BCV E2 gene (M. D. Parker, unpublished data). The lack of other upstream and in-frame ATG codons further supports the idea that this is the initiation codon for the E3 polypeptide. Hydropathic analysis of the predicted polypeptide product indicated that the product had the characteristics of a membrane protein (Fig. 4). Immediately following the presumed initiation codon is a hydrophobic stretch of 15 amino acids which may function as a signal sequence for translocation of the polypeptide across the membranes of the rough endoplasmic reticulum. Another stretch of extreme hydrophobicity extends from amino acid 389 to 414 and may serve to anchor the polypeptide in the viral envelope. Beyond the second hydrophobic domain, the 10 carboxy-terminal amino acids form a more hydrophilic domain which may extend into the interior of the virion. In addition, there are nine possible sites (Asn–X–Ser/Thr) for the addition of *N*-linked oligosaccharides.

Within the large ORF extending from nucleotides 62 to 1334 are two additional shorter ORFs, ORF1 and ORF2, both potentially encoding extremely hydrophobic polypeptides. The first of the internal reading frames (IORF) begins at nucleotide 153 and terminates at nucleotide 561. The second begins at position 1023 and extends to 1272.

Identification of the gene product of mRNA 3 as the E3 polypeptide

In order to identify the product of the long ORF shown in Fig. 3, m^7GpppA-capped mRNA was produced *in vitro* and translated in a rabbit reticulocyte lysate. Immunoprecipitation of the product with a pool of E3-specific monoclonal antibodies yielded a polypeptide of approximately 45K (Fig. 5, lane 1). The presence of pancreatic microsomes during *in vitro* translation resulted in quantitative conversion of the 45K polypeptide to a product of approximately 62K (Fig. 5, lane 2), indicative of extensive post-translational processing, presumably the addition of oligosaccharides. The *in vitro* processed product migrated on SDS–PAGE gels to the same position as the E3 monomer from BCV-infected MDBK cells (Fig. 5, lane 3). The E3 monomer obtained from virus-infected cells appeared as a more heterogeneous band presumably due to additional post-translational processing of the oligosaccharide chains. These results demonstrate that the large ORF shown in Fig. 3 represents the E3 gene of bovine coronavirus.

Confirmation that the E3 protein is the BCV haemagglutinin

Previously, King *et al.* (1985) had identified the E3 protein (gp140/gp65) as the haemagglutinin based on the observation that bromelain treatment of BCV neither destroyed

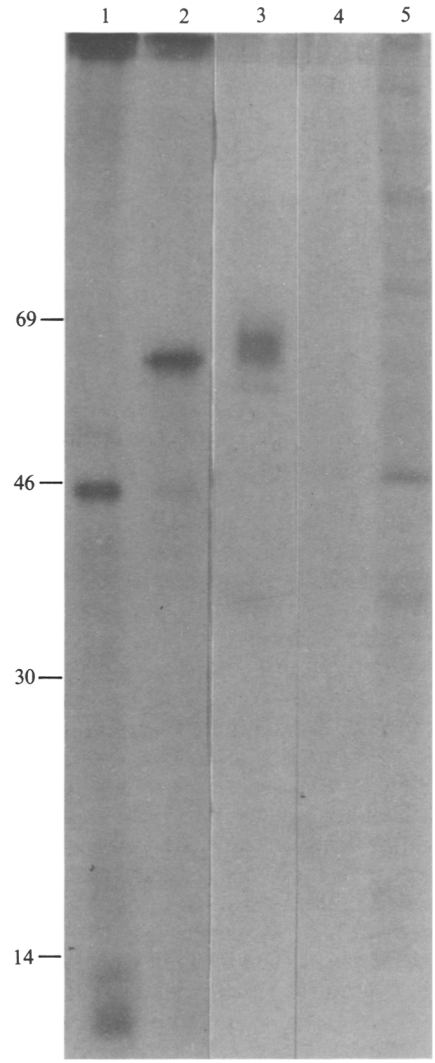

Fig. 5. *In vitro* synthesis of the product of the largest ORF of the BCV E3 gene. *In vitro* transcripts representing nucleotides 53 to 1350 of the cloned gene were translated in rabbit reticulocyte extracts and immunoprecipitated with pooled monoclonal antibodies. Lane 1, *in vitro* translation without added pancreatic microsomes; lane 2, *in vitro* translation in the presence of pancreatic microsomes; lane 3, E3 polypeptide immunoprecipitated from BCV-infected MDBK cells; lane 4, as in lane 1, immunoprecipitated with normal bovine serum; lane 5, immunoprecipitation of BCV-infected MDBK cell extracts with normal bovine serum. $M_r \times 10^{-3}$ indicated at the left.

Table 1. *Haemagglutination inhibition by monoclonal antibodies specific for bovine coronavirus*

Monoclonal antibody	Specificity/ antigenic group	Neutralizing titre	Haemagglutination inhibition titre
HC10-5	E3/A1	150000	2560
KD9-40	E3/A2	50000	2560
KC4-3	E3/B	1100	<20
BD9-8C	E3/C	1250	40
HF8-8	E2/A	130000	<60
HB10-4	E2/A	12500	<40
BB7-14	E2/B	125000	<40
Foetal bovine serum			<20

the integrity of the E3 protein nor eliminated the haemagglutinating activity of the virus whereas Pronase had the opposite effects. Consistent with this assignment, it was also pointed out that the non-haemagglutinating MHV A59 did not contain an analogous protein. However, serological evidence for this assignment has been lacking.

To determine whether any of our monoclonal antibodies with specificity for the BCV E2 or E3 proteins (Deregt & Babiuk, 1987) could inhibit haemagglutination of BALB/c erythrocytes by BCV, virus was incubated separately with these antibodies in a haemagglutination inhibition assay. The results show that only E3-specific monoclonal antibodies, HC10-5 and KD9-40 could significantly inhibit haemagglutination by BCV, thus confirming that the E3 protein is the BCV haemagglutinin (Table 1).

DISCUSSION

In this report, evidence has been presented that the sequence shown in Fig. 3 is that of the E3 glycoprotein monomer. First, the nucleotide sequence of the BCV E3 gene predicts a product with the typical hallmarks of a membrane protein; the presence of a putative hydrophobic amino-terminal signal sequence and a second hydrophobic region at the carboxy terminus which probably serves as a membrane anchoring domain (Garoff, 1985). Second, in the absence of microsomal membranes, *in vitro* translation of mRNA produced from the cloned gene yielded a 45K polypeptide. This is similar in size to that predicted from the deduced amino acid sequence (47K) and also similar to the M_r (42·5K) determined for the E3 apoprotein by studies utilizing tunicamycin, an inhibitor of *N*-linked glycosylation (Hogue & Brian, 1987). Third, when translated in the presence of microsomal membranes, mRNA produced from the cloned gene directed the synthesis of a 62K polypeptide. Processing of the 45K polypeptide to the 62K polypeptide probably involved addition of *N*-linked sugars. There are nine potential *N*-linked glycosylation sites in the sequence shown in Fig. 3, a sufficient number to accommodate the notion that this increase in M_r is due solely to glycosylation (Klenk & Rott, 1980, also see below). However, other forms of post-translational modification such as fatty acylation may also be present. Fourth, the 62K polypeptide produced in the presence of microsomal membranes had the same M_r as the E3 monomer (gp62) from BCV-infected cells. Finally, both 45K and 62K *in vitro* products obtained by translation of mRNA produced from the cloned gene could be immunoprecipitated by E3-specific monoclonal antibodies.

The intracellular precursor to E3 is an *N*-glycosylated protein of 59K which is subsequently processed to the mature 62K form (Deregt *et al.*, 1987). With an average M_r of 2000 for *N*-linked chains of the high mannose type and 3000 for those of the complex type, we estimate that each 59K monomer would contain six or seven *N*-linked chains of the high mannose type assuming that the sugars on the 59K precursor polypeptide are the result of core glycosylation (Klenk & Rott, 1980). Each mature E3 monomer would then have an estimated three or four *N*-linked chains of the complex type. The presence of high mannose and complex oligosaccharides on the E3 monomer may explain the apparent heterogeneity of the E3 monomer immunoprecipitated from infected cells. Thus, seven of the nine potential glycosylation sites are predicted to be utilized for the addition of carbohydrates to the BCV E3 polypeptide.

From pulse–chase experiments, it was concluded that the E3 protein (gp124) was a homodimer (Deregt *et al.*, 1987). Hogue & Brian (1987) have also made this suggestion based upon their deglycosylation studies which gave no indication that two different subunits existed. From M_r considerations and the sequence of the E3 gene in Fig. 3, it is evident that the mature E3 protein is a disulphide-linked dimer of identical units. Viral proteins of this type appear to be rare. Other examples include the neuraminidase of influenza viruses and the haemagglutinin–neuraminidase protein of paramyxoviruses (Lazdins *et al.*, 1972; Varghese *et al.*, 1983; Markwell & Fox, 1980).

The E3 gene contains two IORFs in addition to the sequence encoding the E3 polypeptide. The existence of multiple ORFs in other coronavirus genes, including the N protein gene of BCV (Lapps *et al.*, 1987) and non-structural polypeptide genes of both infectious bronchitis virus (Boursnell *et al.*, 1985) and MHV (Budzilowicz *et al.*, 1985) have been reported previously. We have been able to produce a polypeptide *in vitro* encoded by the IORF extending from nucleotides 153 to 561, but the 15K product was not bound by bovine convalescent serum antibodies (M. D. Parker, unpublished data). Neither of the two internal reading frames is preceded by the consensus (T/A)CTAAAC sequence characteristic of the intergenic regions of

BCV, proposed to be involved in transcriptional control of coronaviruses (Baric *et al.*, 1987). Nor are the proposed initiation codons in favourable sequence contexts. Thus, the question as to whether either of these ORFs is expressed during infection remains unanswered.

The identity of the E3 protein as the BCV haemagglutinin was confirmed in that only E3-specific monoclonal antibodies were effective in inhibiting virus haemagglutination. The two E3-specific monoclonal antibodies that exhibited haemagglutination inhibition both recognized different epitopes within the same antigenic domain (A). They also exhibited the highest neutralization titres of our panel of E3 monoclonal antibodies (Deregt & Babiuk, 1987). This suggests that the site on the E3 protein responsible for interaction with red blood cell receptors may be identical with or in close proximity to a site which is involved in BCV infectivity.

The origin and function of the BCV haemagglutinin is presently unknown. The question arises as to what additional function is provided by the E3 glycoprotein in addition to those activities which are normally carried out by the E2 glycoprotein in coronaviruses which lack the E3 gene. The availability of both neutralizing monoclonal antibodies and the cloned genes for these proteins should aid in the identification of the role played by the E3 glycoprotein in the replication of BCV.

A search of 4248 amino acid sequences in the protein sequence database of the National Biomedical Research Foundation indicated that the E3 polypeptide of BCV bears significant similarity to the HA1 of type C influenza viruses (Nakada *et al.*, 1984; Pfeifer & Compans, 1984). Homologous amino acid sequences begin immediately following the putative signal sequence of the E3 protein and extend to just before the putative carboxy-terminal anchoring domain. This evidence of a common ancestral protein for the BCV E3 protein and the haemagglutinin (HA1) subunit for influenza viruses will be discussed further elsewhere.

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada and is published with permission of the Director of the Veterinary Infectious Disease Organization as article no. 72.

REFERENCES

AVIV, H. & LEDER, P. (1972). Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proceedings of the National Academy of Sciences, U.S.A.* **69**, 1408–1412.

BARIC, R. S., SHIEH, C., STOHLMAN, S. A. & LAI, M. M. C. (1987). Analysis of the intracellular small RNAs of mouse hepatitis virus: evidence for discontinuous transcription. *Virology* **156**, 342–354.

BOURSNELL, M. E. G., BINNS, M. M. & BROWN, T. D. K. (1985). Sequencing of coronavirus IBV genomic RNA: three open reading frames in the 5′ 'unique' region of mRNA D. *Journal of General Virology* **66**, 2253–2258.

BOURSNELL, M. E. G., BROWN, T. D. K., FOULDS, I. J., GREEN, P. F., TOMLEY, F. M. & BINNS, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *Journal of General Virology* **68**, 57–77,

BUDZILOWICZ, C. J., WILCZYNSKI, S. P. & WEISS, S. R. (1985). Three intergenic regions of coronavirus mouse hepatitis virus strain A59 genome RNA contain a common nucleotide sequence that is homologous to the 3′ end of the viral mRNA leader sequence. *Journal of Virology* **53**, 834–840.

DEA, S., ROY, R. S. & BEGIN, M. E. (1980). Bovine coronavirus isolation and cultivation in continuous cell lines. *American Journal of Veterinary Research* **41**, 30–38.

DE GROOT, R. J., TER HAAR, R. J., HORZINEK, M. C. & VAN DER ZEIJST, B. A. M. (1987). Intracellular RNAs of feline infectious peritonitis coronavirus strain 79-1146. *Journal of General Virology* **68**, 995–1002.

DEREGT, D. & BABIUK, L. A. (1987). Monoclonal antibodies to bovine coronavirus: characteristics and topographical mapping of neutralizing epitopes on the E2 and E3 glycoproteins. *Virology* **161**, 410–420.

DEREGT, D., SABARA, M. & BABIUK, L. A. (1987). Structural proteins of bovine coronavirus and their intracellular processing. *Journal of General Virology* **68**, 2863–2877.

GAROFF, H. (1985). Using recombinant DNA techniques to study protein targeting in the eukaryotic cell. *Annual Review of Cell Biology* **1**, 403–445.

GRUNSTEIN, M. & HOGNESS, D. S. (1975). Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proceedings of the National Academy of Sciences, U.S.A.* **72**, 3961–3965.

GUBLER, U. & HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269.

HENIKOFF, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–359.

HOGUE, B. G. & BRIAN, D. A. (1986). Structural proteins of human respiratory coronavirus, OC43. *Virus Research* **5**, 131–144.

HOGUE, B. G. & BRIAN, D. A. (1987). Glycosylation of the bovine coronavirus hemagglutinin. *Advances in Experimental Medicine and Biology* **218**, 131–136.

HOGUE, B. G., KING, B. & BRIAN, D. A. (1984). Antigenic relationships among proteins of bovine coronavirus, human respiratory coronavirus OC-43, and mouse hepatitis coronavirus A59. *Journal of Virology* **51**, 384–388.

JACOBS, L., VAN DER ZEIJST, B. A. M. & HORZINEK, M. C. (1986). Characterization and translation of transmissible gastroenteritis virus mRNAs. *Journal of Virology* **57**, 1–6.

KECK, J. G., HOGUE, B. G., BRIAN, D. A. & LAI, M. M. C. (1988). Temporal regulation of bovine coronavirus RNA synthesis. *Virus Research* **9**, 343–356.

KING, B. & BRIAN, D. A. (1982). Bovine coronavirus structural proteins. *Journal of Virology* **42**, 700–707.

KING, B., POTTS, B. J. & BRIAN, D. A. (1985). Bovine coronavirus hemagglutinin protein. *Virus Research* **2**, 53–59.

KLENK, H.-D. & ROTT, R. (1980). Cotranslational and post-translational processing of viral glycoproteins. *Current Topics in Microbiology and Immunology* **90**, 19–48.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, London* **227**, 680–685.

LAPPS, W., HOGUE, B. G. & BRIAN, D. A. (1987). Sequence analysis of the bovine coronavirus nucleocapsid and matrix proteins genes. *Virology* **157**, 47–57.

LAZDINS, I., HASLAM, E. A. & WHITE, D. O. (1972). The polypeptides of influenza virus. VI. Composition of the neuraminidase. *Virology* **49**, 758–765.

LEHRACH, H., DIAMOND, D., WOZNEY, J. M. & BOEDTKER, H. (1977). RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry* **16**, 4743–4748.

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual.* New York: Cold Spring Harbor Laboratory.

MARKWELL, M. A. K. & FOX, C. F. (1980). Protein–protein interactions within paramyxoviruses identified by native disulphide bonding or reversible chemical cross-linking. *Journal of Virology* **33**, 152–166.

MEBUS, C. A. (1978). Pathogenesis of coronaviral infection in calves. *Journal of the American Veterinary Medical Association* **173**, 631–632.

MELTON, D. A., KRIEG, P. A., REBAGLIATI, M. R., MANIATIS, T., ZIMM, K. & GREEN, M. R. (1984). Efficient in vitro synthesis of biologically active RNA and RNA probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Research* **12**, 7035–7056.

NAKADA, S., CREAGER, R. S., KRYSTAL, M., AARONSON, R. P. & PALESE, P. (1984). Influenza C virus hemagglutinin: comparison with influenza A and B virus hemagglutinins. *Journal of Virology* **50**, 118–124.

PFEIFER, J. B. & COMPANS, R. W. (1984). Structure of the influenza C glycoprotein as determined from cloned DNA. *Virus Research* **1**, 281–296.

POCOCK, D. H. & GARWES, D. J. (1977). The polypeptides of haemagglutinating encephalomyelitis virus and isolated subviral particles. *Journal of General Virology* **37**, 487–499.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

SCHMIDT, I., SKINNER, M. & SIDDELL, S. (1987). Nucleotide sequence of the gene encoding the surface projection glycoprotein of coronavirus MHV-JHM. *Journal of General Virology* **68**, 47–56.

SIDDELL, S., WEGE, H. & TER MEULEN, V. (1983). The biology of coronaviruses. *Journal of General Virology* **64**, 761–776.

STERN, D. F. & KENNEDY, S. I. T. (1980). Coronavirus multiplication strategy. II. Mapping the avian infectious bronchitis viral intracellular RNA species to the genome. *Journal of Virology* **36**, 440–449.

STORZ, J., KALUZA, G., NEIMANN, H. & ROTT, R. (1981). On enteropathogenic bovine coronavirus. *Advances in Experimental Medicine and Biology* **142**, 171–180.

VARGHESE, J. N., LAVER, W. G. & COLMAN, P. M. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2·9 Å resolution. *Nature, London* **303**, 35–40.

(Received 21 June 1988)