Sequence of the membrane protein gene from avian coronavirus IBV

M.E.G. Boursnell, T.D.K. Brown and M.M. Binns

Houghton Poultry Research Station, Houghton, Huntingdon, Cambs PE17 2DA, U.K.

Summary

cDNA clones prepared from genomic RNA of coronavirus IBV have been sequenced. The nucleotide sequence for the complete 5' region of mRNA C, which is not present in mRNAs A and B, has been determined. A sequence of 1224 bases is presented which contains a long open reading frame predicting a polypeptide of molecular weight 25 443. This is in agreement with the molecular weight of 23 000 reported for the unglycosylated form of the membrane polypeptide.

IBV, membrane protein, DNA sequencing

Introduction

Avian infectious bronchitis virus (IBV) is a member of the family Coronaviridae. The coronaviruses are large enveloped viruses with positive-stranded RNA genomes (Siddell et al., 1983). The pleomorphic virus particle is generally spherical in shape and contains three major protein structures: the membrane protein, the nucleocapsid protein and the surface projections which form the distinctive 'corona' (Cavanagh, 1981). The membrane or M protein comprises a polypeptide of molecular weight 23 000 (23k) which is glycosylated to different extents to form glycopolypeptides of molecular weights ranging from 26k to 34k (Stern et al., 1982; Stern and Sefton, 1982b; Cavanagh, 1983). The oligosaccharides of the M polypeptide are of the high mannose type and are linked to the polypeptide by *N*-glycosidic linkages (Stern and Sefton, 1982b; Cavanagh, 1983). The major portion of the membrane protein appears to be embedded in the virion membrane (Cavanagh, 1981) with about 20–40% projecting outside the lipid envelope. Work on the membrane protein of the murine coronavirus MHV-A59 suggests that part of the molecule may also project from the inner surface of the membrane (Sturman et al., 1980).

Six major species of RNA have been observed in IBV-infected cells (Stern and Kennedy, 1980a). These include RNA F. which is the same length as the viral genome, and five smaller species. RNAs A–E, RNA A being the smallest. These mRNAs consist of a 3'-coterminal 'nested' set with each RNA species containing all the sequences present in the smaller RNAs in addition to some 'unique' sequences at its 5' end (Stern and Kennedy, 1980b). In vitro translation studies of fractionated and gel-purified mRNAs from IBV have shown that mRNAs A and C code for the nucleocapsid and membrane polypeptides respectively and that RNA E codes for a precursor polypeptide of the surface projection or spike (Stern et al., 1982; Stern and Sefton, 1984). The sizes of these primary translation products correspond well to the lengths of the 'unique' sequences at the 5' end of each mRNA. Thus it is probable that only the 5' portion of each messenger species is translated.

In this paper we report the DNA sequence of a cloned cDNA copy of IBV genomic RNA in the region corresponding to the 5' end of messenger RNA C. Translation of the sequence predicts a polypeptide of molecular weight 25.4k which is in agreement with the molecular weight of 23k reported for the unglycosylated form of the membrane polypeptide (Stern et al., 1982).

Materials and Methods

Production of oligo(dT)-primed cDNA clones from genomic RNA

The preparation of oligo(dT)-primed cDNA clones has been previously described (Brown and Boursnell, 1984). Briefly, virion RNA was isolated from IBV strain Beaudette grown in embryonated eggs. cDNA was produced by oligo(dT)-primed reverse transcription of the RNA, followed by self-primed reverse transcription to generate the second strand. S1-treated cDNA was dC-tailed using terminal transferase, annealed to dG-tailed pAT153 (Twigg and Sherratt, 1980) and transformed into E. coli HB101. Ampicillin-sensitive colonies were selected for further characterisation. Viral clones were identified as described by hybridisation with a probe prepared by polynucleotide kinase labelling of alkali-treated, full-length IBV genomic RNA. Restriction sites were mapped on a series of clones and this enabled construction of a continuous map, 3.3 kb in length. Hybridisation with a kinase-labelled poly(U) probe has confirmed that these clones include the poly(A) sequences at the 3' terminus of the viral genome.

Production of specific oligonucleotide-primed cDNA clones

A specific oligonucleotide primer, 13 bases long, complementary to a suitable sequence approximately 300 bases from the 5' end of the existing oligo(dT)-primed clones, was synthesised using the phosphotriester method (Gait et al., 1982). This was used to prime reverse transcription on IBV strain Beaudette genomic RNA. Reverse transcription of approximately 50 μg of IBV genomic RNA was carried out in a volume of 50 μl containing 140 mM KCl, 100 mM Tris-HCl pH 8.3, 10 mM MgCl$_2$, 1 mM dATP, dGTP, dTTP, 0.5 mM dCTP, 10 μCi [α-^{32}P]dCTP, 4 mM dithiothreitol, 800 ng of specific oligonucleotide primer, 60 units of human placental

ribonuclease inhibitor, and 80 units of AMV reverse transcriptase for 2 h at 42°C. EDTA was added to 20 mM and the reaction mixture extracted twice with phenol/chloroform. It was passed over a column of Sephadex G-100 equilibrated in 10 mM Tris-HCl pH 7.5, 1 mM EDTA. The excluded fractions were pooled and ethanol precipitated. The cDNA/RNA hybrids were further fractionated on a Sepharose CL4B column equilibrated in 0.3 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA. Excluded fractions were pooled and ethanol precipitated. This material was dC-tailed (Roychoudhury and Wu, 1980). The tailed cDNA was annealed with dG-tailed pBR322 and transformed (Hanahan, 1983) into *E. coli* strain LE392. Viral clones were again identified by hybridisation with a kinase-labelled viral probe, and the overlap with the existing clones confirmed by DNA sequencing after recloning into pUC9 (Messing and Vieria, 1982). These additional clones enabled construction of a continuous map extending 3520 bases from the 3' end of the viral genome.

Formaldehyde–agarose gel analysis of IBV mRNAs

1.5% formaldehyde–agarose gels were run essentially as described by Maniatis et al. (1982). Total RNA samples from IBV-infected chick kidney cell cultures were run overnight at 60 V on 16 cm vertical gels. IBV mRNAs were detected by blotting onto nitrocellulose and probing with nick-translated cloned IBV sequences (Maniatis et al., 1982). Molecular weights were calculated by comparison with the mobilities of DNA restriction fragments and *E. coli* and chicken ribosomal RNAs.

DNA sequence determination

Sequencing was carried out essentially as described by Maxam and Gilbert (1980). Plasmid DNA was prepared by a modification of the method of Holmes and Quigley (1981). For sequencing some regions of the DNA, restriction digests of the clones, or of the viral insert, were recloned into the plasmid pUC9 allowing sequencing from adjacent vector restriction sites (Messing and Vieria, 1982). Fragments recloned included Alu I digests of the C5.136 insert, and the complete inserts of the two clones C5.136 and 142. This latter procedure allowed sequencing in from the ends of both clones. DNA restriction fragments were 3' end-labelled with [α-^{32}P]dNTPs using Klenow polymerase, or 5' end-labelled with [γ-^{32}P]ATP using T4 polynucleotide kinase. The two labelled ends were separated by digestion with a second restriction enzyme, electrophoresis on 3.5% or 5% polyacrylamide gels and extraction of the fragments as described (Maxam and Gilbert, 1980). The depurination reaction was carried out in 66% formic acid for 10 min at 20°C, after which the samples were treated in the same way as the pyrimidine reaction. The products of the sequencing reactions were analysed on 0.3 mm, 8% polyacrylamide sequencing gels. Additional sequencing was carried out by the dideoxy method of Sanger et al. (1977), after recloning of Bam H1 digests of C5.136 into the filamentous phage M13mp9, using the Amersham M13 cloning and sequencing kit (Amersham International). [α-^{32}P]dATP was used in the sequencing reactions and the products were analysed on 0.3 mm, 6% polyacrylamide sequencing gels. All sequencing gels were dried down onto silane-treated glass plates as described by Garoff and Ansorge (1981). Sequence data were stored and analysed on an Apple IIe microcomputer

using the programs of Larson and Messing (1983). The hydrophilicity plot from these programs was used with the hydrophilicity values of Kyte and Doolittle (1982).

Results

1224 base pairs of DNA sequence have been determined from two clones: C5.136 and 142. The positions of these two clones and of the region of sequence presented here are shown in Fig. 1. In Fig. 2a the arrows show the direction and amount of sequence information determined from individual restriction sites. 97% of the sequence has been determined on both strands, all restriction sites used for the sequencing have been 'sequenced through' from other sites, and a large section has been sequenced by both enzymatic and chemical methods. Fig. 2a also shows restriction sites used in the sequencing. Sites marked above the line were used for Maxam and Gilbert sequencing directly from the clones. Sequencing from the Alu I sites marked below the line was carried out on fragments recloned into pUC9 (see Materials and Methods). Also shown below the line are the ends of the clones from which sequencing was carried out after recloning into pUC9.

A search was made for initiation and termination codons in the three possible reading frames. These initiation and termination signals are shown in Fig. 2b with the main open reading frames marked. Also shown is the 5' terminus of mRNA B as determined by S1 mapping (Brown and Boursnell, 1984). S1 mapping of the 5' terminus of mRNA C, using clone 142, proved very difficult, probably due to the small size of the protected fragment. The lack of suitable restriction sites precluded the use of a procedure involving 5' end labelling. The approximate position of the 5'
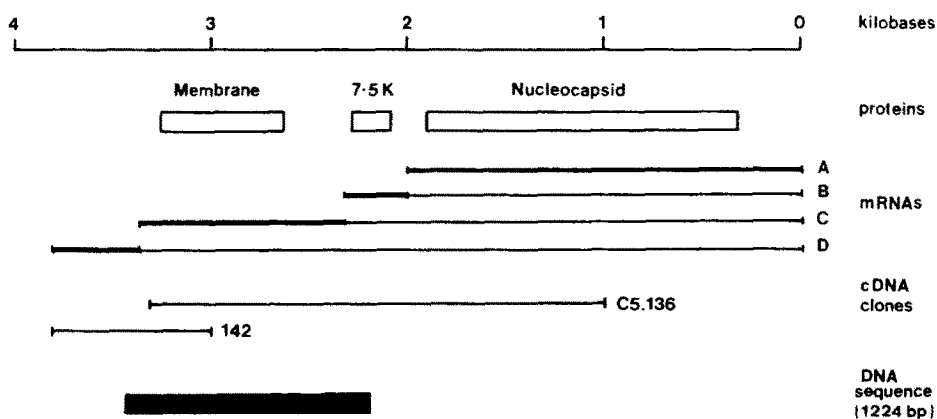

Fig. 1. The first four kilobases from the 3' end of the IBV genome, showing mRNAs A–D and the proteins which they encode. The heavy lines on the mRNAs show the regions not present in the next smallest RNA. Also shown are the positions of the two cDNA clones sequenced. The region of sequence presented in this paper is marked with a black bar.
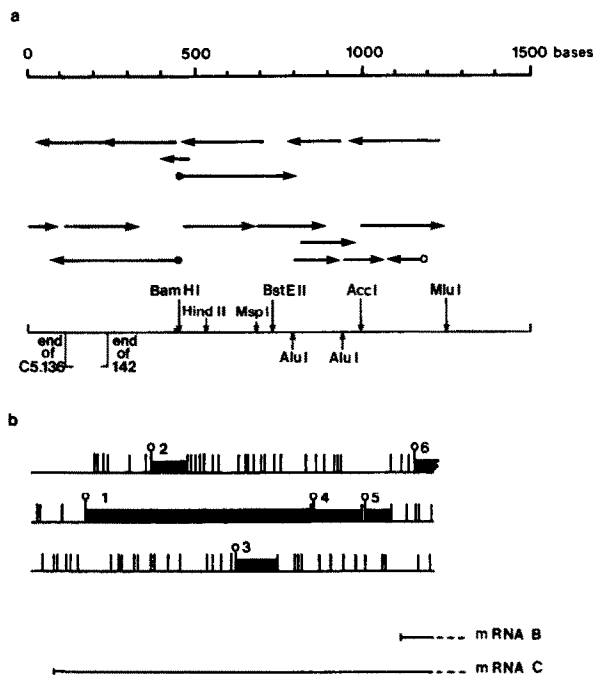
Fig. 2. (a) Direction and extent of sequence information obtained from individual restriction sites. Arrows starting with solid circles indicate DNA sequenced by the dideoxy method. Open circles indicate Maxam and Gilbert sequencing using 5' end-labelled DNA, and plain arrows indicate Maxam and Gilbert sequencing using 3' end-labelled DNA. Restriction sites used for sequencing are shown (see Results). (b) The locations of termination codons (vertical bars) and potential initiation codons (bars with open circles on top) in the three possible translational reading frames. The heavy black lines show the main open reading frames. Also shown are the positions of the 5' termini of the bodies of mRNAs B and C (see text).

terminus of mRNA C was therefore determined by mRNA length measurements on formaldehyde–agarose gels. These measurements give an estimated length of 3400 bases. Coronavirus mRNAs have a 5' leader sequence which is fused to the 'body' of the message during transcription (Lai et al., 1983; Baric et al., 1983; Spaan et al,, 1983). In the case of mRNA C, therefore, the length of the IBV leader sequence, approximately 60 bases (Brown and Boursnell, manuscript in preparation), has been subtracted from the measured length of the mRNA to give the length of the 'body' of the message which is present on genomic RNA.

The 1224 base pairs of sequence are shown in Fig. 3 with a translation of the largest open reading frame. This open reading frame (ORF) of 225 amino acids predicts a polypeptide of molecular weight 25 443. Two smaller ORFs, 4 and 5 (see Fig. 2b), which follow on directly from the large ORF in the same reading frame, predict polypeptides of 5896 and 3055 daltons respectively. There are also two small ORFs, 2 and 3, present within the sequences of the largest ORF. The ORF starting with the ATG codon at position 1147 is the start of the putative 7500 dalton polypeptide encoded by mRNA B (Brown and Boursnell, manuscript submitted).

```
                15            30            45            60            75
TTATTTTGGTATACATGGGTAGTAATTCCAGGAGCTAAGGGTACAGCCTTTGTATACAAGTATACATATGGTAGA


                90           105           120           135           150
AAACTTAACAATCCGGAATTAGAAGCAGTTATTGTTAACGAGTTTCCTAAGAACGGTTGGAATAATAAAAATCCA


               165           180           195           210           225
GCAAATTTTCAAGATGCCCAACGAGACAAATTGTACTCTTGACTTTGAACAGTCAGTTCAGCTTTTTAAAGAGTA
              METProAsnGluThrAsnCysThrLeuAspPheGluGlnSerValGlnLeuPheLysGluTyr
                   •             •
               240           255           270           285           300
TAATTTATTTATAACTGCATTCTTGTTGTTCTTAACCATAATACTTCAGTATGGCTATGCAACAAGAAGTAAGGT
AsnLeuPheIleThrAlaPheLeuLeuPheLeuThrIleIleLeuGlnTyrGlyTyrAlaThrArgSerLysVal


               315           330           345           360           375
TATTTATACACTGAAAATGATAGTGTTATGGTGCTTTTGGCCCCTTAACATTGCAGTAGGTGTAATTTCATGTAC
IleTyrThrLeuLysMETIleValLeuTrpCysPheTrpProLeuAsnIleAlaValGlyValIleSerCysThr


               390           405           420           435           450
ATACCCACCAAACACAGGAGGTCTTGTCGCAGCGATAATACTTACAGTGTTTGCGTGTCTGTCTTTTGTAGGTTA
TyrProProAsnThrGlyGlyLeuValAlaAlaIleIleLeuThrValPheAlaCysLeuSerPheValGlyTyr


               465           480           495           510           525
TTGGATCCAGAGTATTAGACTCTTTAAGCGGTGTAGGTCATGGTGGTCATTTAATCCAGAATCTAATGCCGTAGG
TrpIleGlnSerIleArgLeuPheLysArgCysArgSerTrpTrpSerPheAsnProGluSerAsnAlaValGly


               540           555           570           585           600
TTCAATACTCCTAACTAATGGTCAACAATGTAATTTTGCTATAGAGAGTGTGCCAATGGTGCTTTCTCCAATTAT
SerIleLeuLeuThrAsnGlyGlnGlnCysAsnPheAlaIleGluSerValProMETValLeuSerProIleIle


               615           630           645           660           675
AAAGAATGGTGTTCTTTATTGTGAGGGTCAGTGGCTTGCTAAGTGTGAACCAGACCACTTGCCTAAAGATATATT
LysAsnGlyValLeuTyrCysGluGlyGlnTrpLeuAlaLysCysGluProAspHisLeuProLysAspIlePhe


               690           705           720           735           750
TGTTTGTACACCGGATAGACGTAATATCTACCGTATGGTGCAGAAATATACTGGTGACCAAAGCGGAAATAAGAA
ValCysThrProAspArgArgAsnIleTyrArgMETValGlnLysTyrThrGlyAspGlnSerGlyAsnLysLys


               765           780           795           810           825
AAGGTTTGCTACGTTTGTCTATGCAAAGCAGTCAGTAGATACTGGCGAGCTAGAAAGTGTAGCAACAGGAGGAAG
ArgPheAlaThrPheValTyrAlaLysGlnSerValAspThrGlyGluLeuGluSerValAlaThrGlyGlySer


               840           855           870           885           900
TAGTCTTTACACATAAATGTGTGTGTGTAGAGAGTATTTAAAATTATTCTTTAATAGCGCCTCTGTTTTAAGAGC
SerLeuTyrThr***


               915           930           945           960           975
GCATAAGAGTATTTATTTTGAGGATACTAATATAAATCCTCTTTGTTTTATACTCTCCTTTCAAGAGCTATTAAC


               990          1005          1020          1035          1050
GGTGTTACCTTTCAAGTAGATAATGGAAAAGTCTACTACGAAGGAACACCAGTTTTCCAAAAAGGTTGTTGTAGG


              1065          1080          1095          1110          1125
ATGTGGTCCAATTATAAGAAAGAATAATTGAACCACCTACTACACTTATTTTTATAAGAGGTGTTTTACTTAACA


              1140          1155          1170          1185          1200
AAAACTTAACAAATACGGACGATGAAATGGCTGACTAGTTTTGGAAGAGCAGTTATTTCTTGTTATAAATCCCTA
                          METLysTrpLeuThrSerPheGlyArgAlaValIleSerCysTyrLysSerLeu


              1215
CTATTAACTCAACTTAGAGTGTTA
LeuLeuThrGlnLeuArgValLeu
```

Fig. 3. 1224 bases of DNA sequence from IBV genomic cDNA clones. A translation of the main open reading frame in mRNA C, and part of the first open reading frame in mRNA B are shown. DNA sequence underlined shows the homologies at the 5′ termini of mRNAs B and C. The amino acids underlined show the three potential membrane-spanning regions in the membrane polypeptide (see Discussion), and the black dots beneath two amino acids show the potential N-linked glycosylation points.

Discussion

The organisation of the messenger RNAs of coronaviruses and in vitro translation studies have led to the hypothesis that it is only the 'unique' 5' sequences of each mRNA which are translated (Stern and Kennedy, 1980b; Lai et al., 1981). The sequence shown in Fig. 3 predicts a large open reading frame (ORF 1) of 225 amino acids. If translated this would code for a protein of molecular weight 25.4k. Gel-purified mRNA C, translated in vitro, gives a 23k polypeptide identical in size to the unglycosylated form of the membrane polypeptide (Stern et al., 1982; Stern and Sefton, 1984). Hence, the data is entirely consistent with this open reading frame being the coding'sequence for the membrane polypeptide.

The sequence also predicts four other open reading frames (ORFs) longer than 25 amino acids. Two of these, ORFs 2 and 3, of 32 and 45 amino acids, respectively, lie within the coding sequences for the membrane polypeptide. Examination of the sequences flanking the initiation codon for the membrane polypeptide gene shows that they conform to one of the sequences preferred for functional eucaryotic initiation codons (Kozak, 1983). Because of this, and because initiation at anything other than the first AUG codon is known to be very rare (Kozak, 1983), it seems probable that the 25.4k product of the 225 amino acid open reading frame, ORF 1, is the only polypeptide produced from mRNA C. It is possible, however, that there are minor mRNA species, intermediate in size between mRNAs B and C, which would enable a product to be translated from one or both of the ORFs which are present between the end of the membrane gene and mRNA B.

85 bases before the initiation codon for the membrane polypeptide gene is the sequence CTTAACAA which also occurs 101 bases before the first open reading frame of mRNA A (probably the nucleocapsid gene) and 28 bases before 7.5k open reading frame encoded by mRNA B (Boursnell and Brown, manuscript submitted). The point at which this sequence occurs is also known to mark the 5' termini of the 'bodies' of mRNAs A and B (Brown and Boursnell, 1984). Thus it seems likely that this sequence also represents the 5' terminus of mRNA C, and its position relative to the estimated position of the end of the body of mRNA C helps to confirm this. It is interesting to note that there appears to be greater homology between the sequences present at the 5' termini of mRNAs A and B than to that which we are postulating occurs at the end of mRNA C (see Fig. 4). There is, however, a 12-base homology, AAAACTTAACAA, between the sequences at the 5' termini of mRNAs B and C.

The amino acid sequence of the membrane polypeptide reveals several interesting features. It is known that the carbohydrate side chains of the glycosylated forms of

A :   GGATTAGATTGTGTTTACTTTCTTAACAAAGCAGGACAAGCAGAGCCTT

B :   TATAAGAGGTGTTTTACTTAACAAAAACTTAACAAATACGGACGATGAAA

C :   TATACATATGGTAGAAAACTTAACAATCCGGAATTAGAAGCAGTTA

Fig. 4. Nucleotide sequences at the 5' termini of mRNAs A, B and C. Some sequence homologies are underlined.

the 23k polypeptide are N-linked (Stern and Sefton, 1982b; Cavanagh, 1983). There are only two sites, Asn-X-Thr, in the amino acid sequence, at which N-linked glycosylation can occur (Hubbard and Ivatt, 1981). These are at the NH_2-terminus at amino acids 3 and 6. This agrees well with the data of Stern et al. (1982) who have shown that glycosylation of the 23k polypeptide occurs at the NH_2-terminus.

A hydropathicity profile (Fig. 5a), which shows the local hydrophobic tendencies (positive values) or hydrophilic tendencies (negative values) of a polypeptide chain (Kyte and Doolittle, 1982) shows that there are several hydrophobic regions in the membrane polypeptide. In particular there are three stretches where the mean hydrophobicities taken over 19 amino acids greatly exceed the lowest value generally associated with membrane-spanning polypeptides (Kyte and Doolittle, 1982). These regions, which are underlined in Fig. 3, each consist of a contiguous length of at least 20 uncharged amino acids which are highly enriched in hydrophobic residues. It is possible that each of these regions spans the membrane once, with the short hydrophilic sections separating them being at the surfaces of the membrane. However, since these hydrophilic sections are only very short, it may be that all the hydrophobic NH_2-terminal region is buried in the membrane with only the hydrophilic COOH-terminal exposed.

It is known that the membrane proteins of the 23k family do not undergo any major post-translational proteolytic processing (Stern and Sefton, 1982a), and indeed there is no apparent hydrophobic signal sequence at the NH_2-terminus which might be cleaved off. However, it is possible that one of the internal hydrophobic
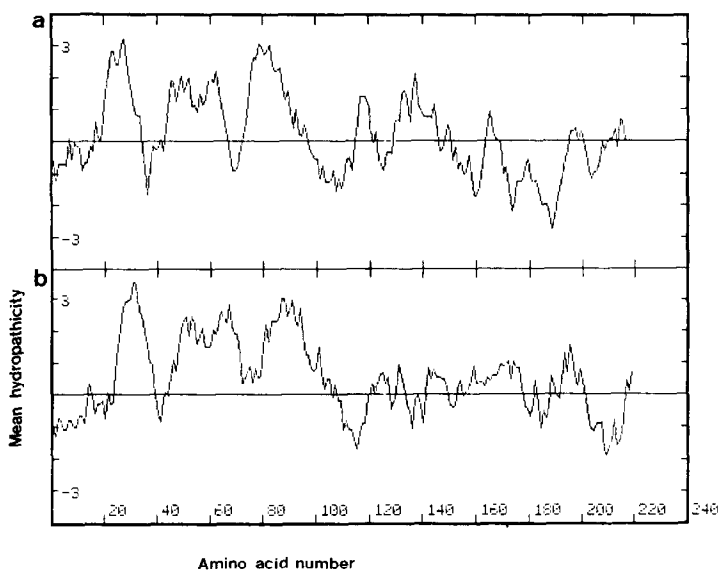

Fig. 5. Hydropathicity profile of the predicted amino acid sequences of (a) the IBV membrane polypeptide, and (b) the E1 polypeptide of MHV-A59. Positive values indicate hydrophobic regions and negative values hydrophilic regions. Each point is the mean hydropathicity of a span of nine residues, plotted at the first residue of the span.

regions acts as a signal sequence as may be the case for chicken ovalbumin (Lingappa et al., 1979).

Treatment of the viral particle with the protease bromelain reduces the molecular weight of all the glycosylated forms of the membrane polypeptide to a single polypeptide whose molecular weight is approximately 1k less than the size of the unglycosylated membrane polypeptide (D. Cavanagh, personal communication). Thus, the protease appears to cleave at least the first six amino acids with their associated variable carbohydrate side chains from the NH$_2$-terminus of the membrane protein. The fact that this is achieved by protease treatment of largely intact virus particles suggests that the carbohydrate side chains, and hence the NH$_2$-terminus, are located on the outside of the viral membrane. This would have been expected by analogy with the membrane (E1) polypeptide of MHV and with other viral glycoproteins (Sturman and Holmes, 1977; Rose and Gallione, 1981).

The 75 or so amino acids of the COOH-terminal region of the polypeptide form a hydrophilic region which may protrude from the inner surface of the viral envelope. For MHV-A59 Sturman et al. (1980) have shown that there is some interaction between the membrane (E1) protein and the viral RNA. If this is also the case for IBV it is interesting to note that of the 18 basic residues in the membrane polypeptide, 14 of them are in the COOH-terminal half of the molecule. It is possible that these basic residues are involved in binding to the RNA.

Comparison of the sequence of the IBV membrane protein and the E1 protein of MHV-A59 (Peter Rottier, personal communication) at the RNA sequence level reveals essentially no homology. However, comparison of the predicted amino acid sequences (Fig. 6) shows a remarkable degree of homology especially in the amino

1
```
        MPNETNCTLDFEQSVQLFKEVNLFITAFLLFLTIILQMGYATRSK
MSSTTQAPEPVYQWTADEAVQFLKEWNFSLGIILLFITIILQFGYTSRSM
```

51
```
VIYTLKMIMLWCFWPLNIAVGVISCTVPPNTGGLVAAIILTVFACLSFVG
FIYVVKMIILWLMWPLTIVLCIFNCVMALNNVYLGFSIVFTIVSIVIWIM
```

101
```
YWIQSIRLFKRCRSWWSFNPESNAVGSILLTNGQQCNFAIESVPMVLSPI
YFVNSIRLFIRTGSWWSFNPETNNLMCIDMKGTVYVRPIIEDYHTLTATI
```

151
```
IKNGVLYCEGQWLAKCEPDHLPKDIFVCTPDRRNIVRMVQKYTGDQSGNK
IRGHLYMQGVKLGTGFSLSDLPAYVTVAKVSHLCTYKRAFLDKVDGVSGF
```

201
```
KRFATFVYAKQSVDTGELESVATGGSSLYTB
AVYVKSKVGNYRLPSNKPSGADTALLRIB
```
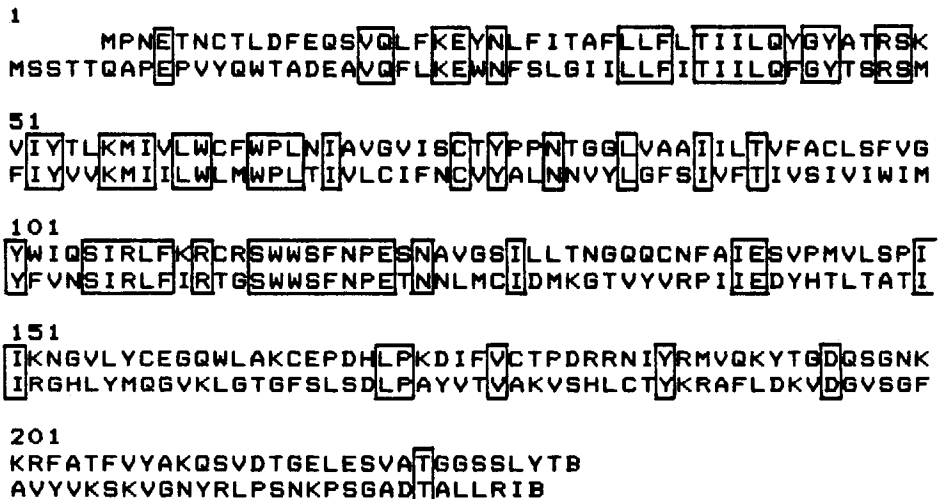
Fig. 6. Comparison of the amino acid sequences (single letter code) of the IBV membrane polypeptide (top sequence) and the E1 polypeptide of MHV-A59 (bottom sequence). Boxed regions show homologies between the two sequences.

terminal half of the molecule. In particular, the sequence within the first two potential membrane-spanning regions seem to be highly conserved, with 24 out of 44 amino acids the same and most of the differences involving pairs of amino acids with similar properties. A region of even greater homology occurs in the hydrophilic stretch of residues at position 104 (Fig. 6) in which 15 out of 19 residues are the same. These similarities at the amino acid level are strikingly reflected in the hydropathicity profiles (Fig. 5). This protein sequence conservation in the hydrophobic NH_2-terminal half of the polypeptide suggests a high degree of selective pressure for maintenance of the structure of the putative trans-membrane regions of coronavirus membrane proteins.

In conclusion then, the sequence presented here shows good agreement with the observed and expected properties of the membrane polypeptide of infectious bronchitis virus.

References

Baric, S., Stohlman, S. and Lai, M.M.C. (1983) Characterisation of replicative intermediate RNA of mouse hepatitis virus: presence of leader RNA sequences on nascent chains. J. Virol. 48, 633–640.

Brown, T.D.K. and Boursnell, M.E.G. (1984) Avian infectious bronchitis virus genomic RNA contains sequence homologies at the intergenic boundaries. Virus Res. 1, 15–24.

Cann, A.J., Stanway, G., Hauptmann, R., Minor, P.D., Schild, G.C., Clarke, L.D., Mountford, R.C. and Almond J.W. (1983) Poliovirus type 3 molecular cloning of the genome and nucleotide sequence of the region encoding the protease and polymerase proteins. Nucleic Acids Res. 11, 1267–1281.

Cavanagh, D. (1981) Structural polypeptides of coronavirus IBV. J. Gen. Virol. 53, 93–103.

Cavanagh, D. (1983) Coronavirus IBV glycopolypeptides: size of their polypeptide moieties and nature of their oligosaccharides. J. Gen. Virol. 64, 1187–1191.

Gait, M.J., Matthes, H.W.D., Singh, M., Sproat, B. and Titmas, R.C. (1982) Rapid synthesis of oligodeoxyribonucleotides VII. Solid phase synthesis of oligodeoxyribonucleotides by a continuous flow phosphotriester method on a kieselguhr-polyamide support. Nucleic Acids Res. 10, 6243–6254.

Garoff, H. and Ansorge, W. (1981) Improvements in DNA sequencing gels. Anal. Biochem. 115, 450–457.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. 166, 557–580.

Holmes, D.S. and Quigley, M. (1981) A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114, 193–197.

Hubbard, S.C. and Ivatt, R.J. (1981) Synthesis and processing of asparagine-linked oligosaccharides. Annu. Rev. Biochem. 50, 555–584.

Kozak, M. (1983) Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. Microbiol. Rev. 47, 1–45.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

Lai, M.M.C., Brayton, P.R., Armen, R.C., Patton, C.D., Pugh, C. and Stohlman, S.A. (1981) Mouse hepatitis virus A59: mRNA structure and genetic localization of the sequence divergence from hepatotropic strain MHV-3. J. Virol. 39, 823–834.

Lai, M.M.C., Patton, C.D., Baric, R.S. and Stohlman, S. (1983) Presence of leader sequences in the mRNA of mouse hepatitis virus. J. Virol. 46, 1027–1033.

Larson, R. and Messing, J. (1983) Apple II computer software for DNA and protein sequencing data. DNA 2, 31–35.

Lingappa, V.R., Lingappa, J.R. and Blobel, G. (1979) Chicken ovalbumin contains an internal signal sequence. Nature (London) 281, 117–121.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning – A Laboratory Manual, Cold Spring Harbor Laboratory, New York.

Maxam, A.M. and Gilbert, W. (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. In: Methods in Enzymology (Grossman, L. and Moldave, K., eds.), Vol. 65, Part 1, pp. 499–560. Academic Press, New York.

Messing, J. and Vieria, J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19, 259–268.

Rose, J.K. and Gallione, C.J. (1981) Nucleotide sequences of the mRNAs encoding the vesicular stomatitis virus G and M proteins determined from cDNA clones containing the complete coding regions. J. Virol. 39, 519–528.

Roychoudhury, R. and Wu, R. (1980) Terminal transferase-catalysed addition of nucleotides to the 3' termini of DNA. In: Methods in Enzymology (Grossman, L. and Moldave, K., eds.), Vol. 65, Part 1, pp. 43–62. Academic Press, New York.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467.

Siddell, S., Wege, H. and ter Meulen, V. (1983) The biology of coronaviruses. J. Gen. Virol. 64, 761–776.

Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B.A.M. and Siddell, S.G. (1983) Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. EMBO J. 2, 1839–1844.

Stern, D.F. and Kennedy, S.I.T. (1980a) Coronavirus multiplication strategy. I. Identification and characterisation of virus-specified RNA. J. Virol. 34, 665–674.

Stern, D.F. and Kennedy, S.I.T. (1980b) Coronavirus multiplication strategy. II. Mapping the avian infectious bronchitis virus intracellular RNA species to the genome. J. Virol. 36, 440–449.

Stern, D.F. and Sefton, B. (1982a) Coronavirus proteins: Biogenesis of avian infectious bronchitis virus virion proteins. J. Virol. 44, 794–803.

Stern, D.F. and Sefton, B. (1982b) Coronavirus proteins: structure and function of the oligosaccharides of the avian infectious bronchitis virus glycoproteins. J. Virol. 44, 804–812.

Stern, D.F. and Sefton, B.M. (1984) Coronavirus multiplication: the locations of genes coding for the virion proteins on the avian infectious bronchitis virus genome. J. Virol., in press.

Stern, D.F., Burgess, L. and Sefton, B.M. (1982) Structural analysis of virion proteins of the avian coronavirus infectious bronchitis virus. J. Virol. 42, 208–219.

Sturman, L.S. and Holmes, K.V. (1977) Characterization of a coronavirus. II. Glycoproteins of the viral envelope: tryptic peptide analysis. Virology 77, 650–660.

Sturman, L.S., Holmes, K.V. and Behnke, J. (1980) Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. J. Virol. 33, 449–462.

Twigg, A.J. and Sherratt, D. (1980) Trans-complementable copy-number mutants of plasmid ColE1. Nature (London) 283, 216–218.