

GENE 1020

Sequencing of coronavirus IBV genomic RNA: a 195-base open reading frame encoded by mRNA B

(Avian infectious bronchitis virus; cDNA clones; recombinant DNA)

M.E.G. Bournnell and T.D.K. Brown

Houghton Poultry Research Station, Houghton, Huntingdon, Cambs PE17 2DA (U.K.) Tel. (0480) 64101

(Received November 24th, 1983)

(Revision received March 1st, 1984)

(Accepted March 15th, 1984)

SUMMARY

DNA sequencing of genomic cDNA clones of avian infectious bronchitis virus (IBV) has been carried out. 770 bases have been determined which include genomic sequences spanning the 5' termini of the two smallest mRNAs of the 3'-coterminal "nested" set: mRNA A and mRNA B. This region contains the complete coding sequences for mRNA B which are additional to those present in mRNA A. Two open reading frames are present, predicting proteins of M_r s 7500 and 9500.

INTRODUCTION

Avian IBV, in common with other coronaviruses, has a single-stranded, polyadenylated, infectious RNA genome approx. 20 kb in length (Stern and Kennedy, 1980a). In infected cells multiple subgenomic positive-stranded RNAs are produced (Siddell et al., 1983). For IBV and MHV these have been shown to consist of a 3'-coterminal "nested" set (Stern and Kennedy, 1980a; Lai et al., 1981; Leibowitz et al., 1981). For both IBV and MHV the messenger function of these subgenomic RNAs has been demonstrated (Rottier et al., 1981; Stern et al., 1982; Siddell, 1983). However, certain differences of genome organisation have become apparent between

these two coronaviruses. First, MHV has six major subgenomic RNAs whereas IBV has only five (Stern and Kennedy, 1980b; Lai et al., 1981). Second, the coding function of the various messenger RNAs is different. Both IBV and MHV contain three main structural polypeptides: the nucleocapsid, the membrane (E1), and the spike or peplomer (E2) polypeptides (Cavanagh, 1981; Siddell et al., 1983). In both systems the smallest RNA codes for the nucleocapsid protein but in MHV the next smallest RNA codes for the membrane polypeptide, whereas in IBV the membrane polypeptide is coded for by the third smallest RNA (Siddell et al., 1980; Siddell, 1983; Rottier et al., 1981; Stern et al., 1982; Stern and Sefton, 1984). These differences are summarised in Fig. 1.

The organisation of the messenger RNAs and *in vitro* translation studies have led to the hypothesis that the 5'-most sequences of each mRNA, which are not present in the next smallest mRNA, contain

Abbreviations: bp, base pairs; IBV, infectious bronchitis virus; kb, kilobases or kilobase pairs; MHV, murine hepatitis virus; mRNA, messenger RNA; ORF, open reading frame.

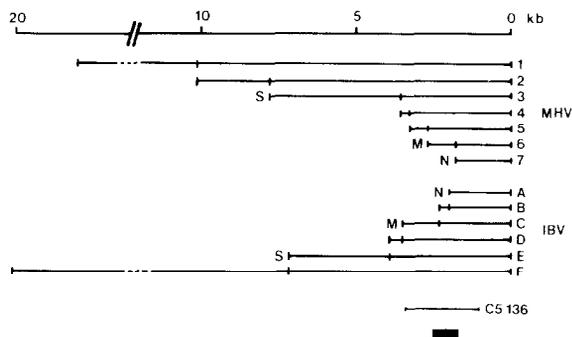


Fig. 1. Sequence organisation of MHV and IBV mRNAs. For MHV, mRNA 1 is the same length as the genomic RNA. For IBV, mRNA F is the same length as the genomic RNA. For IBV the approximate sizes of RNAs A, B, C, D and E are 2, 2.4, 3.4, 4.1 and 7.8 kb (see MATERIALS AND METHODS, section e). The coding assignments are shown at the left-hand end of each mRNA. N, nucleocapsid; M, membrane or E1 polypeptide; S, spike or peplomer polypeptide. The heavy bar shows the region of sequence presented in this paper. Also shown is the position of clone C5.136.

the complete coding sequences for the major protein product produced by that messenger species (Stern and Kennedy, 1980b; Lai et al., 1981). However, since the only coronavirus sequences published are those of the smallest RNA species (Armstrong et al., 1983; Skinner and Siddell, 1983) it has not been possible to examine this hypothesis at the RNA sequence level. RNA sequence data from this region might also enable us to predict the properties and thus aid the identification of possible polypeptides coded for by mRNA B.

In this paper we report the nucleotide sequence of a cloned cDNA copy of IBV genomic RNA in the region corresponding to the 5' end of mRNA B. The sequence shows that the 5'-most sequences of mRNA B could code for a hydrophobic 7.5-kDa protein.

MATERIALS AND METHODS

(a) Cloning of IBV genomic RNA

The preparation of cDNA clones has been previously described (Brown and Bournell, 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 nuclease-treated cDNA was dC-tailed using terminal transferase, annealed to dG-tailed *Pst*I-cleaved pAT153 (Twigg and Sherratt, 1980) and transformed into *Escherichia coli* HB101. Ampicillin-sensitive colonies were selected for further characterisation.

(b) Characterisation of cDNA clones

Viral clones were identified 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. That these included the poly(A) sequences at 3'-terminus of the viral genome was confirmed by hybridisation with a kinase-labelled poly(U) probe.

(c) 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). M_r s were calculated by comparison with the mobilities of DNA restriction fragments and *E. coli* and chicken ribosomal RNAs.

(d) DNA sequence determination

Plasmid DNA was prepared by a modification of the method of Holmes and Quigley, 1981. DNA restriction fragments, 3' end-labelled with [α - 32 P]dNTPs using Klenow polymerase or 5' end-labelled with [γ - 32 P]ATP using T4 polynucleotide kinase, were sequenced essentially as described by 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. For sequencing some regions of the DNA, restriction digests of the viral insert were recloned into the plasmid pUC9 allowing sequencing from adjacent

vector restriction sites (Messing and Vieira, 1982). Sequence data were stored and analysed on an Apple IIe microcomputer using the programs of Larson and Messing (1983) and on a VAX 11/780 mini-computer using the programs of Staden (1984).

RESULTS

770 bp of DNA sequence from one IBV genomic clone, C5.136 (Brown and Bournnell, 1984) have been determined. This sequence corresponds to the genomic RNA sequence stretching from 2.40 kb to 1.63 kb from the 3' end of the viral genome. In Fig. 2a the arrows show the direction and extent of DNA sequence information obtained from individual restriction enzyme cleavage sites. Fig. 2b shows positions of restriction sites used in the sequencing.

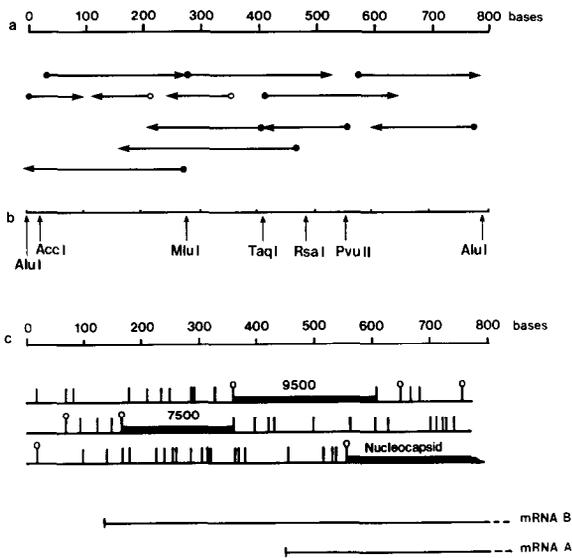


Fig. 2. Sequencing strategy for cloned IBV cDNA (clone C5.136). (a) Arrows show the direction and extent of sequence information obtained from individual restriction sites. Arrows starting with solid circles indicate sequencing of DNA 3' end-labelled with Klenow polymerase. Arrows starting with open circles indicate sequencing of DNA 5' end-labelled with polynucleotide kinase. (b) A map of restriction sites used in the sequencing. (c) 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 ORFs: M_r 7500, M_r 9500, and the putative nucleocapsid ORF. Also shown are the positions of the 5' ends of mRNAs A and B as determined by S1 nuclease mapping (see Fig. 3).

95% of the sequence has been determined on both strands, and on each strand most regions have been sequenced more than once from different restriction sites.

Fig. 2c shows the positions of the initiation and termination codons in the three reading frames and the positions of the 5' ends of mRNAs A and B as determined by S1 nuclease mapping (Brown and Bournnell, 1984). It should be noted that S1 nuclease mapping will determine the 5' end of the "body" of the mRNA (see DISCUSSION). The DNA sequence of 770 nucleotides, with a translation of the three main ORFs, is shown in Fig. 3. The lines below the sequence at positions 131–165 and 440–466 indicate

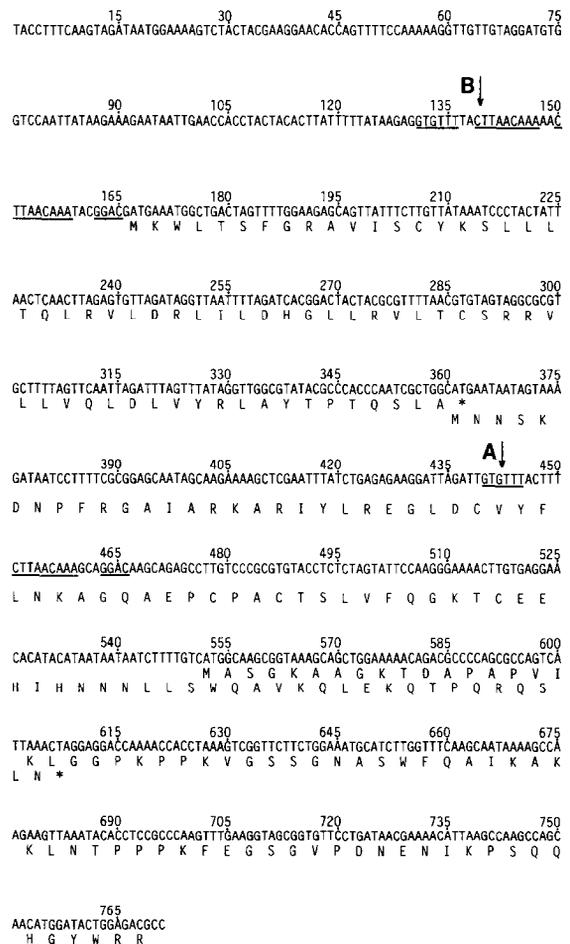


Fig. 3. 770-bp sequence of part of IBV cDNA clone C5.136. A translation of the three main ORFs is shown in single-letter amino acid code. Termination codons for these ORFs are shown as asterisks. Arrows above the sequence show the 5' ends of mRNAs A and B as determined by S1 nuclease mapping. Lines below the sequence at these points indicate the regions of homology which occur at the 5' ends of mRNAs A and B.

the regions of homology which occur at the 5' end of mRNAs A and B and the arrows above the sequence at these points show the 5' ends of the bodies of the mRNAs as determined by S1 nuclease mapping (Brown and Bournsnel, 1984).

In the genomic sequence immediately 5'-wards of the end of mRNA B, there are no ORFs longer than 81 bases. About 30 bases into mRNA B, at position 167, there is an AUG codon, followed by an ORF of 195 bases. This predicts a protein of M_r 7500. The coding sequences for this putative protein are entirely contained in that part of mRNA B which is not present in mRNA A. One base before the termination codon for this M_r 7500 ORF, at position 361, there is an initiation codon followed by an open reading frame which could code for a protein of M_r 9500. This ORF is not contained within the "unique" sequences of mRNA B and overlaps considerably with mRNA A. In mRNA A, 100 bases from the 5' terminus, at position 552, is the beginning of an ORF which extends beyond the limits of the sequenced region. This is likely to be the start of the gene coding for the nucleocapsid protein. A comparison of this partial sequence to the sequences published for the nucleocapsid proteins of MHV-A59 and MHV-JHM7 (Armstrong et al., 1983; Skinner and Siddell, 1983) reveals no significant homology at the RNA or protein level. Although some homology might be expected it should be noted that no serological cross-reaction between IBV and any of the mammalian coronaviruses has been reported (Siddell et al., 1983).

DISCUSSION

The region of the IBV sequence presented in this paper contains the 5' ends, on the viral genome, of mRNAs A and B. The messenger RNAs of coronavirus IBV are probably transcribed from non-contiguous regions of the viral genome. Work on the murine coronavirus MHV has shown that a common 5' leader sequence, originating from the extreme 5' end of the viral genome, is fused to the body of each messenger RNA (Lai et al., 1982; 1983; Spaan et al., 1983; Baric et al., 1983). It is likely that the IBV messenger RNAs have the same structure. However, whether or not leader sequences are present, S1 nuclease mapping experiments have shown that the

ends of the bodies of mRNAs A and B lie at positions 139 and 443 respectively (Brown and Bournsnel, 1984) as shown in Fig. 3.

Of the two ORFs which are present at the 5' end of mRNA B, the M_r 7500 ORF seems to be the most likely candidate for translation in vivo. The location of the coding sequence for the putative M_r 7500 polypeptide fits in well with the hypothesis that the major polypeptide product of each mRNA is translated from those 5' sequences not present in the next smallest RNA. If the M_r 9500 polypeptide were translated then this would no longer be true, since its coding region stretches well into mRNA A. Furthermore, the RNA sequences flanking the initiation codons for the putative M_r 7500 and nucleocapsid genes correspond well to those preferred for functional eukaryotic initiation codons (Kozak, 1983). However, the sequence GNNAUGA around the AUG codon at the start of the M_r 9500 ORF is rare in this context. In addition the AUG codon at the start of the M_r 7500 ORF is the first initiation codon to occur in the body of mRNA B; since initiation of translation at anything other than the first AUG codon is known to be rare (Kozak, 1983) this suggests that this ORF codes for the major product of mRNA B.

The amino acid sequence of the putative M_r 7500 polypeptide shows it to be hydrophobic in nature and to have an unusual composition in that 26% (17 out of 65) of its residues are leucine. Of the six possible triplets coding for leucine, one (UUA) is used 8 times out of 17. This unusual composition and codon bias, which would not be expected from a chance ORF, suggests that this polypeptide is translated in vivo. A computer analysis of the sequences presented here has been carried out using the program ANALYSEQ (Staden, 1984). This program uses certain criteria to select one of the three reading frames as being the most likely protein coding frame. Although better suited to analysing large ORFs, it is interesting to note that a search based on looking for codon biases above those expected from the base composition selects 85% of the M_r 7500 ORF as the most likely coding frame. The 15% of codons not selected are not in a single block, which might have suggested a sequencing error leading to an artificial frameshift. All of the nucleocapsid which has so far been sequenced was selected as the most likely coding frame, but none of the M_r 9500 ORF.

We have carried out *in vitro* translation of total and poly(A) + RNA populations from IBV-infected chick kidney cell cultures using a rabbit reticulocyte lysate system. However, analysis of the products on 10–18% polyacrylamide gradient gels containing urea could not resolve any small polypeptides due to high background in the relevant low- M_r range. A similar problem has been found by Stern and Sefton (1984; Stern, D.F., personal communication) who have carried out *in vitro* translation studies of gel-purified and gradient-fractionated mRNAs and have identified no major specific product from mRNAs B or D.

At present, therefore, it is not possible to say definitively whether either of these polypeptides is produced. In a search for small structural polypeptides, a [^3H]leucine-labelled preparation of virus has been analysed on a 12.5% polyacrylamide tube gel (Cavanagh, D., personal communication), using the phosphate buffer system of Swank and Munkres (1971). This showed that there were three detectable polypeptides of apparent M_r s: 16000, 12000, and 10000. The percentage of the total counts in the gel accounted for by these polypeptides was <1%. Thus, if either of these ORFs codes for a structural polypeptide, it must only be present in very small quantities. It is also possible that they could code for non-structural polypeptides. These would be difficult to identify in IBV-infected cells by pulse-labelling techniques without using immunoprecipitation to lower the background of host-cell incorporation which is poorly shut off by IBV infection. The availability of sequence data for these putative polypeptides, however, opens up the possibility of using immunoprecipitation with antisera prepared against synthetic oligopeptides to search for the presence of these polypeptides in IBV-infected cells.

ACKNOWLEDGEMENTS

We are grateful to Penny Gatter, Anne Foulds and Ian Foulds for excellent technical assistance. This research was carried out under Research Contract No. GBI-2-011-UK of the Biomolecular Engineering Programme of the Commission of the European Communities.

REFERENCES

- Armstrong, J., Smeekens, S. and Rottier, P.: Sequence of the nucleocapsid gene from murine coronavirus MHV-A59. *Nucl. Acids Res.* 11 (1983) 883–891.
- Baric, S., Stohlman, S. and Lai, M.M.C.: Characterization of replicative intermediate RNA of mouse hepatitis virus: presence of leader sequences on nascent chains. *J. Virol.* 48 (1983) 633–640.
- Brown, T.D.K. and Bournsnel, M.E.G.: Avian infectious bronchitis virus genomic RNA contains sequence homologies at the intergenic boundaries. *Virus Res.* 1 (1984) 15–24.
- Cavanagh, D.: Structural polypeptides of coronavirus IBV. *J. Gen. Virol.* 53 (1981) 93–103.
- Holmes, D.S. and Quigley, M.: A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* 114 (1981) 193–197.
- Kozak, M.: Comparison of initiation of protein synthesis in prokaryotes, eukaryotes and organelles. *Microbiol. Revs.* 47 (1983) 1–45.
- Lai, M.M.C., Brayton, P.R., Armen, R.C., Patton, C.D., Pugh, C. and Stohlman, S.A.: Mouse hepatitis virus A59: mRNA structure and genetic localization of the sequence divergence from hepatotropic strain MHV-3. *J. Virol.* 39 (1981) 823–834.
- Lai, M.M.C., Patton, C.D. and Stohlman, S.: Further characterisation of mRNAs of mouse hepatitis virus: presence of common 5'-end nucleotides. *J. Virol.* 41 (1982) 557–565.
- Lai, M.M.C., Patton, C.D., Baric, R.S. and Stohlman, S.: Presence of leader sequences in the mRNA of mouse hepatitis virus. *J. Virol.* 46 (1983) 1027–1033.
- Larson, R. and Messing, J.: Apple II computer software for DNA and protein sequence data. *DNA* 2 (1983) 31–35.
- Leibowitz, J.L., Wilhelmson, K.C. and Bond, C.W.: The virus-specific intracellular RNA species of two murine coronaviruses: MHV-A59 and MHV-JHM. *Virology* 114 (1981) 39–51.
- Maniatis, T., Fritsch, E.F. and Sambrook, J.: *Molecular Cloning – A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.
- Maxam, A.M. and Gilbert, W.: Sequencing end-labelled DNA with base-specific chemical cleavages, in Grossman, L. and Moldave, K. (Eds.), *Methods in Enzymology*, Vol. 65, Part 1, Academic Press, New York, 1980, pp. 499–560.
- Messing, J. and Vieira, J.: The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19 (1982) 259–268.
- Rottier, P., Spaan, W.J.M., Horzinek, M.C. and Van der Zeijst, B.A.M.: Translation of three mouse hepatitis virus strain A59 subgenomic RNAs in *Xenopus laevis* oocytes. *J. Virol.* 38 (1981) 20–26.
- Siddell, S.: Coronavirus JHM: coding assignments of subgenomic mRNAs. *J. Gen. Virol.* 64 (1983) 113–125.
- Siddell, S., Wege, H., Barthel, A. and Ter Meulen, V.: Coronavirus JHM: cell-free synthesis of structural protein p60. *J. Virol.* 33 (1980) 10–17.
- Siddell, S., Wege, H. and Ter Meulen, V.: The biology of coronaviruses. *J. Gen. Virol.* 64 (1983) 761–776.

- Skinner M. and Siddell, S.: Coronavirus JHM: nucleotide sequence of the mRNA that encodes nucleocapsid protein. *Nucl. Acids Res.* 11, (1983) 5045–5054.
- Spaan, W., Delius, H. Skinner, M., Armstrong, J., Rottier, P. Smeekens, S., Van der Zeijst, A.M. and Siddell, S.G.: Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO J.* 2 (1983) 1839–1844.
- Staden, R.: Graphic methods to determine the function of nucleic acid sequences. *Nucl. Acids Res.* 12 (1984) 521–538.
- Stern, D.F. and Kennedy, S.I.T.: Coronavirus multiplication strategy, I. Identification and characterisation of virus-specified RNA. *J. Virol.* 34 (1980a) 665–674.
- Stern, D.F. and Kennedy, S.I.T.: Coronavirus multiplication strategy, II. Mapping the avian infectious bronchitis virus intracellular RNA species to the genome. *J. Virol.* 36 (1980b) 440–449.
- Stern, D.F., Burgess, L. and Sefton, B.M.: Structural analysis of virion proteins of the avian coronavirus infectious bronchitis virus. *J. Virol.* 42 (1982) 208–219.
- Stern, D.F. and Sefton, B.M.: Coronavirus multiplication: locations of virion proteins on the avian infectious bronchitis virus genome. *J. Virol.* (1984) in press.
- Swank, R.T. and Munkres, K.D.: Molecular weight analysis of oligopeptides by electrophoresis in polyacrylamide gel with sodium dodecyl sulfate. *Anal. Biochem.* 39 (1971) 462–477.
- Twigg, A.J. and Sherratt, D.: Trans-complementable copy-number mutants of plasmid ColE1. *Nature* 283 (1980) 216–218.

Communicated by K. Chater.