# DNA SEQUENCING STUDIES OF GENOMIC cDNA CLONES OF AVIAN INFECTIOUS BRONCHITIS VIRUS

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

Houghton Poultry Research Station
Houghton, Huntingdon
Cambridgeshire, England

## INTRODUCTION

Avian infectious bronchitis virus (IBV) has a positive stranded RNA genome about 20 kilobases in length. The virus particle contains three major structural proteins, the nucleocapsid protein (which is associated with the RNA genome), the matrix or membrane protein and the spike or surface projection protein [1]. Infection with IBV results in the synthesis of six major polyadenylated messenger RNAs [2]. One of these is equal in size to the genomic RNA and the other five form a 3' coterminal ('nested') set, with the sequences from each RNA present in all the larger RNA species [3]. These mRNAs have been named A,B,C,D,E and F, RNA A being the smallest. In vitro translation studies [4], have shown that RNA A directs the synthesis of the viral nucleocapsid protein and RNA C directs the synthesis of the viral matrix polypeptide. Unpublished results of Stern and Sefton show that an unglycosylated form of the viral spike precursor can be synthesised by in vitro translation of RNA E.

Work on murine hepatitis viruses has shown that sequences from the 5' end of the genomic RNA are fused to the 5' end of each messenger RNA [5,6]. It is thought that the RNA polymerase may transcribe a short sequence from the 3' end of the negative stranded template and then jump to specific recognition sequences in the main part of the template which would then form the start of each messenger RNA. Thus each messenger RNA consists of 'leader'



P. J. M. Rottier et al. (eds.), *Molecular Biology and Pathogenesis of Coronaviruses*

and 'body' sequences which originate from non-contiguous regions of the genomic RNA. The protein-coding region of each mRNA is thought to consist of that part of the 'body' of the message which is not present in smaller mRNAs. At the moment it is not clear whether the same mechanism operates in IBV. Analysis of RNase $T_1$ oligonucleotides[3], gives no indication of leader sequences, but this evidence is not conclusive.

DNA sequencing of clones complementary to genomic IBV RNA has been undertaken to determine the sequence of the viral RNA at the point corresponding to the 5' ends of messenger RNAs A and B. Combined with S1 mapping data to locate the boundaries of the 'bodies' of the messenger RNAs it should be possible to find out if there are any recognition signals in the genomic RNA which might direct the fusion process.

## METHODS

### Isolation and analysis of cDNA clones

Double-stranded cDNA was synthesised from viral[+] genomic RNA using an oligo-dT primer and AMV reverse transcriptase for the first strand and reverse transcriptase again for the second strand. This was treated with S1 nuclease and tailed with dC residues. The tailed cDNA was annealed to dG-tailed PstI digest pAT153 and used to transform E.coli HB101 to tetracycline resistance. Ampicillin-sensitive colonies were picked and grown, the plasmid DNA isolated and analysed by restriction endonuclease digestion. Restriction sites were mapped in 25 of the clones and this enabled them to be fitted together into a continuous map, 3.3kb in length.

### Orientation and confirmation of viral origin

Clones were digested with PstI and run on agarose gels and the DNA was transferred to nitrocellulose by the Southern blot method[7]. Duplicate blots were hybridised with a) polynucleotide kinase-labelled IBV genomic RNA fragments, b) nick-translated pAT153 and c) kinase-labelled poly-U. The IBV probe was to confirm that the DNA inserted into the PstI site of the plasmid is a copy of viral sequences. The pAT153 probe was to confirm that no plasmid sequences were present in the putative 'viral' band. The poly-U probe was to identify which clones had a copy of the poly-A sequences present on the 3' end of the viral genome. This last blot was repeated with other restriction digests of the appropriate clones to identify which end of the inserted DNA had the poly-A sequences.

---

[+] IBV Beaudette strain

S1 mapping

The PstI/PvuII fragment (see figure 1b) of the cDNA clone used for DNA sequencing (C5.136) was purified on a polyacrylamide gel and 5' labelled using $32P_{\gamma}ATP$ and polynucleotide kinase. Total cytoplasmic RNA was prepared from IBV infected chick kidney cells using an NP40 lysis procedure. S1 mapping was carried out using essentially the modification of the Berk and Sharp procedure,[8] as described by Weaver and Weissman [9]. Protected 32P-labelled DNA fragments were analysed on 5% non-denaturing polyacrylamide gels. Sizes of the protected fragments were estimated from their positions relative to those of end-labelled Hae III fragments of $\phi$X174 RF DNA.

DNA sequence determination

Plasmid DNA was prepared from E.coli by a modification of the method of Holmes and Quigley [10]. DNA restriction fragments, 3' end-labelled with 32P-dNTP using Klenow polymerase, were sequenced essentially as described by Maxam and Gilbert[11]. The depurination reaction was carried out in 2% diphenylamine, 66% formic acid, 1mM EDTA for 10 minutes at 20°C, followed by 3-fold dilution in water, three ether extractions and lyophilisation. Piperidine hydrolysis was done as described[11]. For sequencing some regions of the DNA, restriction digests of the 'viral' insert were recloned into the plasmid pUC9, which contains the cloning sites corresponding to those in the M13 vector M13mp9. Sequence data were stored and analysed on an Apple IIe microcomputer using the programs of Larson and Messing[12].

RESULTS

One clone, C5.136, which represented from 1kb to 3.3kb from the 3' end of the viral genome was chosen for DNA sequencing. A preliminary sequence from only one DNA strand has been obtained from 1600 bases in this region. This sequence is located as shown in figure 1a. In figure 1b the arrows indicate the direction and amount of sequence information obtained from individual restriction enzyme cleavages. The DNA sequence from this region is shown in figure 2. Figure 3 shows the positions of stop and start codons in the three reading frames, the positions of the main open reading frames and the estimated positions of the ends of messenger RNAs A and B that were obtained from S1 mapping experiments using cDNA clones derived from genomic RNA.

The results of an S1 mapping experiment in which varying quantities of total cytoplasmic RNA from IBV infected chick kidney

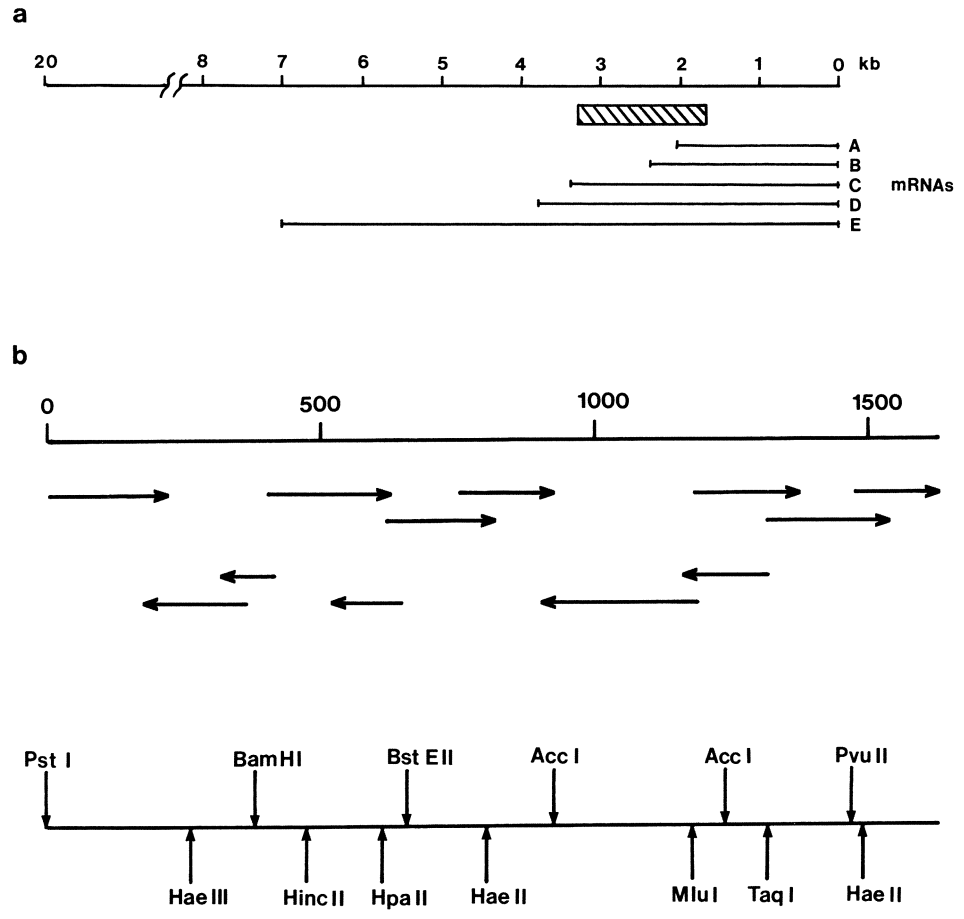


Figure 1. a) Map showing the region of IBV for which cDNA clones have been sequenced (hatched box). b) Restriction map of the 1600 base region which has been sequenced. Arrows indicate direction and extent of sequence information from individual restriction sites.

cells were annealed with a 5' labelled Pst I/Pvu II fragment (see figure 1b) are shown in figure 4. Three DNA fragments are protected by the RNA. The sequence organisation of the genomic clone C5.136 has been demonstrated to correspond to the 3' nested set shown for IBV mRNAs by Stern and Kennedy[3]. Labelled restriction fragments of the clone were used to probe IBV mRNAs

```
                                                        50
GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGAAGAA

                                        !--->       100  First ATG codon
CGGTTGGAATAATAAAAATCCAGCAAATTTTCAAGATGCCCAACGAGACA  in Matrix ORF

                                                    150
AATTGTACTCTTGACTTTGAACAGTCAGTTCAGCTTTTTAAAGAGTATAA

                                                    200
TTTATTTATAACTGCATTCTTGTTGTTCTTAACCATAATACTTCAGTATG

                                                    250
GCTATGCAACAAGAAGTAAGGTTATTTATACACTGAAAATGATAGTGTTA

                                                    300
TGGTGCTTTTGGCCCCTTAACATTGCAGTAGGTGTAATTTCATGTACATA

                                                    350
CCCACCAAACACAGGAGGTCTTGTCGCAGCGATAATACTTACAGTGTTTG

                                                    400
CGTGTCTGTCTTTTGTAGGTTATTGGATCCAGAGTATTAGACTCTTTAAG

                                                    450
CGGTGTAGGTCATGGTGGTCATTTAATCCAGAATCTAATGCCGTAGGTTC

                                                    500
AATACTCCTAACTAATGGTCAACAATGTAATTTTGCTATAGAGAGTGTGC

                                                    550
CAATGGTGCTTTCTCCAATTATAAAGAATGGTGTTCTTTATTGTGAGGGT

                                                    600
CAGTGGCTTGCTAAGTGTGAACCAGACCACTTGCCTAAAGATATATTTGT

                                                    650
TTGTACACCGGATAGACGTAATATCTACCGTATGGTGCAGAAATATACTG

                                                    700
GTGACCAAAGCGTAAATAATAAAAGGTTTGCTACGTCTATGCAAAGCAGT

 <---! ORF1                                         750  UAG stop codon
CAGTAGATACTGGCTAGCTATAAAGTGTAGCAACAGTAGTAAGTAGTCTT  at the end of
                                                         the Matrix gene
                                                    800
TACACATAAATGTGTGTGTGTATATAGTATTTAAAATTATTCTTTAATAG
```

Figure 2. DNA sequence of the cDNA clone C5.136. The positions of the open reading frames (ORFs) are shown. Asterisks indicate the end of the mRNAs as measured by S1 mapping. Lines over the sequence show regions of homology at the messenger termini. (continued on next page)

```
                                                        850
CGCCTCTGTTTTAAGAGCGCATAAGAGTATTTATTTTGAGGATACTAATA

                                                        900
TAAATCCTCTTTGTTTTATACTCTCCTTTCAAGAGCTATTAACGGTGTTA

                                                        950
CCTTTCAAGATAATGGAAAGTCTACTACGAAGGAACACCAGTTTTACAAA

                                                       1000
AAGGTTGTTGTAGGATGTGGTCCAATTATAAGAAAGAATAATTGAACCAC

                                  *                         * indicates end
CTACTACACTTATTTTTATAAGAGGTGTTTTACTTAACAAAAACTTAACA  of RNA B
                                  *
          !---> ORF2                                1100  Start codon for
AATACGGACGATGAAATGGCTGACTAGTTTTGGAAGAGCAGTTATTTCTT  ORF2

                                                       1150
GTTATAAATCCCTACTATTAACTCAACTTAGAGTGTTAGATAGGTTAATT

                                                       1200
TTAGATCACGGACTACTACGCGTTTTAACGTGTAGTAGGCGCGTGCTTTT

                                                       1250
AGTTCAATTAGATTTAGTTTATAGGTTGGCGTATACGCCCACCCAATCGC

    !---> ORF3                                        1300  Start codon for
TGGCATGAATAATAGTAAAGATAATCCTTTTCGCGGAGCAATAGCAAGAA  ORF3
    <--! ORF2                                             UGA stop codon
                                      *                   for ORF2
AAGCTCGAATTTATCTGAGAGAAGGATTAGATTGTGTTTACTTTCTTAAC  * indicates end
                                      *                   of RNA A
                                                       1400
AAAGCAGGACAAGCAGAGCCTTGTCCCGCGTGTACCTCTCTAGTATTCCA

                                                ORF4 !---> Start codon for
AGGGAAAACTTGTGAGGAACACATACATAATAATAATCTTTTGTCATGGC  ORF4
                                                           (Nucleocapsid?)
                                                       1500
AAGCGGTAAAGCAGCTGGAAAAACAGACGCCCCAGCGCCAGTCATTAAAC

<-! ORF3                                               1550  UAG stop codon
TAGGAGGACCAAAACCACCTAAAGTCGGTTCTTCTGGAAATGCATCTTGG  for ORF3

                                                       1600
TTTCAAGCAATAAAAGCCAAGAAGTTAAATACACCTCCGCCCAAGTTTGA


AGGTAGCGGTGTTCCTGATAAC
```

Figure 2. (continued) See caption on previous page.


Figure 3. Positions of termination codons (vertical bars) and potential initiation codons (bars with open circles on top) in the three reading frames. The heavy black lines show the positions of the main open reading frames (ORFs). The lines underneath show the positions of the mRNAs as determined by S1 mapping.

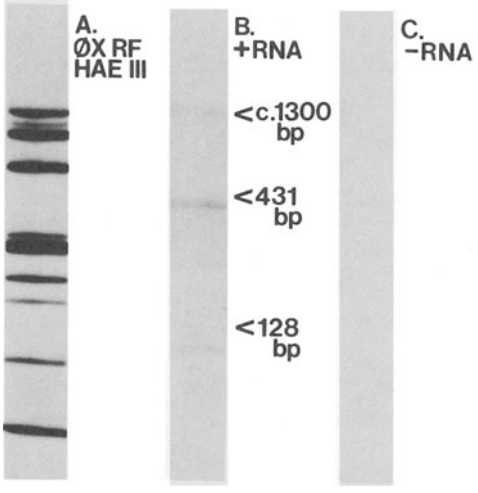


Figure 4. S1 mapping of the non-leader 5' termini of mRNAs A and B on cDNA clone C5.136 using a 5' $^{32}$P-labelled PstI/Pvu II fragment.

separated on formaldehyde gels (Brown & Boursnell, unpublished results). Given these data on sequence organisation the protected bands can be interpreted as mapping the distances of the 5' termini of the mRNAs A and B from the PvuII site in C5.136, excluding any leader sequences which may be joined to the mRNAs. The presence of a band at essentially the same length as the PstI/PvuII fragment also indicates that the terminus of the non-leader sequences of mRNA C is either at or lies beyond the 3.3kb covered by clone C5.136. The latter idea is supported by the observation in primer extension experiments using mRNAs of a major band which maps approximately 350 bases beyond the 5' end of clone C5.136 (Brown & Boursnell, unpublished results).

Knowledge of the positions of the 5' termini of the non-leader sequences of the mRNAs A and B prompted a search for homologies between the two boundary regions. Extensive sequence homologies were indeed observed (see figure 5). In particular a nine base stretch of sequence is common between the two regions, and is present twice at the end of mRNA B.

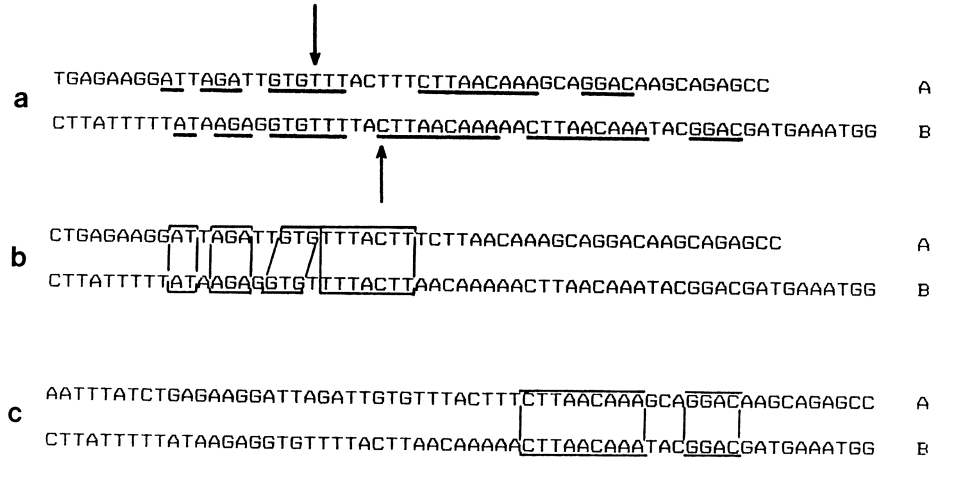


Figure 5. Sequences of cDNA at the 5' ends of the non-leader sequences of mRNAs A and B. a) The arrows show the position of the ends of the non leader sequences. The underlined regions show homologies between the two regions. b) and c) show two possible alignments of the two regions which give high degrees of homology.

## DISCUSSION

As can be seen from figure 1b, approximately 80% of the sequence presented here has not been verified by sequencing the complementary strand. As a result therefore it is in some ways only a preliminary sequence. More work is at present under way in order to obtain sequence information from both strands of the DNA throughout this region.

There are four open reading frames (ORFs) longer than 150 bases and these are shown in figure 3. ORF1 appears to be the protein coding region of messenger RNA C, which has been reported to direct the synthesis of the viral matrix protein [4]. Because the cDNA clone only extends 3.3kb from the 3' end of the virus, it is not possible to tell whether the first AUG codon in this stretch of sequence is in fact the initiation codon for the matrix gene. However, if this sequence were to be translated, it would code for a protein of molecular weight 23,500 daltons which is very similar to the molecular weight of the unglycosylated form of the matrix protein (23,000 daltons) as obtained by polyacrylamide gel electrophoresis [13, 14]. ORF2 corresponds to the part of messenger RNA B which is not present in RNA A and probably represents that region of RNA B which is translated. If it were translated it would code for a hydrophobic polypeptide of molecular weight 7,500 daltons. ORF4 starts within mRNA A and continues to the boundary of the region which has been sequenced. Since it has been reported that RNA A directs the synthesis of the viral nucleocapsid protein it is probable that this open reading frame represents the start of the nucleocapsid gene. ORF3 is more difficult to assign. Current thinking in the coronavirus field suggests that only the part of each messenger RNA which is not present in the next smallest RNA is translated. If this is true then ORF4 would represent a chance open reading frame which is not translated in vivo. This may be the case, and indeed it can be seen that other smaller open reading frames are present at various points in this sequence.

The significance of the sequence homologies found at the 5' termini of the non-leader sequences of mRNAs A and B represented in the genomic RNA clone remains unclear at the present time. It does however seem probable that they have a role in mRNA synthesis and that they might possibly act as recognition sites for the enzymes involved in generating the messenger RNAs. More detailed consideration of their role will however only become possible when the 5' termini of the messenger RNAs have been sequenced.

REFERENCES

1. D. Cavanagh. Structural polypeptides of coronavirus IBV. J. Gen. Virol. 53: 93 (1981).
2. D. F. Stern and S. I. T. Kennedy. Coronavirus multiplication strategy. I. Identification and characterisation of virus-specified RNA. J. Virol. 34: 665 (1980)
3. D. F. Stern and S. I. T. Kennedy. Coronavirus multiplication strategy. II. Mapping the avian infectious bronchitis virus intracellular RNA species to the genome. J. Virol. 36: 440 (1980).
4. D. F. Stern, L. Burgess and B. M. Sefton. Structural analysis of virion proteins of the avian coronavirus infectious bronchitis virus. J. Virol. 42: 208 (1982).
5. M. M. C. Lai, C. D. Patton and S. A. Stohlman. Further characterisation of mRNAs of Mouse Hepatitis Virus: presence of common 5'-end nucleotides. J. Virol. 41: 557 (1982).
6. W. J. M. Spaan, P. J. M. Rottier, M. C. Horzinek and B. A. M. van der Zeijst. Sequence relationships between the genome and the intracellular species 1, 3, 6, and 7 of mouse hepatitis strain A59. J. Virol. 42: 432 (1982).
7. E. Southern. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503 (1975).
8. A. J. Berk and P. A. Sharp. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell 12: 721 (1977).
9. R. F. Weaver and C. Weissman. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15S$\beta$-globin mRNA precursor and mature 10S$\beta$-globin mRNA have identical map coordinates. N. Acids Res. 7: 1175 (1979).
10. D.S. Holmes and M. Quigley. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114: 193 (1981).
11. A. M. Maxam and W. Gilbert. Sequencing end-labelled DNA with base-specific chemical cleavages. Meth. Enzymol. 65: 499 (1980).
12. R. Larson and J. Messing. Apple II software for M13 shotgun sequencing. N. Acids Res. 10: 39 (1982).
13. D. Cavanagh. Coronavirus IBV polypeptides: size of their polypeptide moieties and nature of their oligosaccharides. J. Gen. Virol. 64: 1187 (1983)
14. D. F. Stern and B. M. Sefton. Coronavirus proteins: structure and function of the oligosaccharides of the Avian Infectious Bronchitis Virus Glycoproteins. J. Virol.44: 804 (1982)