Sequences of the Nucleocapsid Genes from Two Strains of Avian Infectious Bronchitis Virus

By M. E. G. BOURSNEll,∗ M. M. Binns, I. J. FOULDS and T. D. K. BROWN

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

(Accepted 2 November 1984)

SUMMARY

cDNAs prepared from viral genomic RNA purified from two strains of infectious bronchitis virus (IBV) (Beaudette and M41) have been cloned into pBR322. Three of these clones, which contain the complete sequences of mRNA A for both strains, except for the leader sequences which are only present on the subgenomic messenger RNAs, have been sequenced using the dideoxy method. The sequences are similar for both strains, each containing a single long open reading frame of 1227 bases which predicts a polypeptide of molecular weight approximately 45000. The genome position and size of this predicted polypeptide are consistent with it being the gene for the nucleocapsid protein. The amino acid sequence shows considerable homology with those of the nucleocapsids of murine hepatitis virus strains A59 and JHM. The major difference between the sequences determined for the two IBV strains is that the 3' non-coding region of the Beaudette strain contains a 184 base segment which is not present in the M41 strain.

INTRODUCTION

Coronaviruses are enveloped viruses with a positive-stranded RNA genome of 15 to 20 kilobases (Siddell et al., 1983). The virion contains three major protein structures: the spikes or surface projections, the membrane protein, and, associated with the RNA, the nucleocapsid protein (Sturman et al., 1980; Cavanagh, 1981; Siddell et al., 1983). In infected cells a number of subgenomic mRNA species are produced which form a nested set, with a common 3' terminus, but extending to different lengths in the 5' direction (Stern & Kennedy, 1980a; Leibowitz et al., 1981). It appears that, in general, only the 5' proximal region of each RNA is translated (Stern & Kennedy, 1980b; Rottier et al., 1981; Siddell, 1983; Stern & Sefton, 1984). In the case of avian infectious bronchitis virus (IBV), and of the murine coronavirus murine hepatitis virus (MHV), the smallest mRNA is the most abundant and codes for a phosphorylated (Stohlman & Lai, 1979; Siddell et al., 1981; Stohlman et al., 1983) unglycosylated (Sturman, 1977; Macnaughton et al., 1977) polypeptide of 50000 to 60000 mol. wt. (Siddell et al., 1980; Stern & Sefton, 1984). This is the only viral protein unaffected by Pronase treatment of intact virions, and is therefore most likely to be the nucleocapsid polypeptide (Sturman, 1977). It is also approximately the same size as the core proteins of several other enveloped RNA virus families such as myxoviruses, paramyxoviruses and rhabdoviruses, all of which have nucleocapsids with helical symmetry (Lenard & Compans, 1974). The nucleocapsid of coronaviruses also appears to have helical symmetry (Oshiro, 1973).

Avian infectious bronchitis virus causes economically important disease in the domestic fowl. However, the Massachusetts-derived strain Beaudette (Beaudette & Hudson, 1937), commonly used in the laboratory, has had over 250 passages in eggs and is no longer pathogenic for chickens. In contrast, the Massachusetts strain M41 has had over 400 passages in chickens and remains pathogenic (Geilhausen et al., 1972; Collins & Alexander, 1980).

In this paper we present the sequences of the nucleocapsid genes of the IBV strains Beaudette and M41 which have been obtained from cDNA clones of virion RNA.
METHODS

cDNA cloning. Three cDNA clones, derived from RNA isolated from gradient-purified virus, have been used for sequencing. For Beaudette the oligo(dT)-primed clone C5-322 was used, which includes the poly(A) tail derived from the 3' terminus of the viral genome and extends for 1930 base pairs in the 5' direction. The production of this clone was exactly as described in Brown & Boursnell (1984). For M41 two clones, C41-81 and 196, have been used. C41-81 was produced as described in Brown & Boursnell (1984). The clone 196 was made by the method of Gubler & Hoffman (1983). It was shown to include the poly(A) sequences from the 3' terminus of the viral RNA by probing Southern blots of PstI-digested plasmid DNA with γ-32P kinase-labelled poly(U) (Brown & Boursnell, 1984). The positions of these clones and of the mRNAs are shown in Fig. 1.

Recloning into M13 for DNA sequencing. For C5-322 the viral insert, which had been cloned into the PstI site of pBR322, was isolated from acrylamide gels (Maxam & Gilbert, 1980) and recloned into SmaI-cut, phosphatase-treated, M13mp10 replicative form DNA (Amersham) after digestion with DNase I (Anderson, 1981) or the restriction enzymes AluI and Rsal, and into BamHI-digested M13mp9 after digestion with Sau3AI. For the M41 clones the plasmids were digested with DNase I and recloned into M13mp10 without isolation of the viral inserts. Clones containing viral sequences were then identified by transferring bacterial colonies grown from colourless plaques on to nitrocellulose filters and probing with radiolabelled viral probes (Grunstein & Hogness, 1975). Two probes were used: γ-32P kinase-labelled alkali-degraded viral RNA, and α-32P-labelled single-stranded cDNA, reverse-transcribed from viral RNA using calf thymus primers. Using these two probes single-stranded phages containing inserts from both the plus and minus strands of viral sequences were selected.

DNA sequencing. Sequencing was carried out by the dideoxy method (Sanger et al., 1977; Bankier & Barrell, 1983). Additional sequence information was obtained by reverse sequencing of some of the M13 clones (Hong, 1981). [α-32P]dATP (Amersham) was used in the sequencing reactions and buffer gradient gels were used to analyse the reaction products (Biggin et al., 1983).

Computer analysis of sequence data. Sequence data were read directly into a BBC microcomputer using a sonic digitizer (Graf/Bar. Science Accessories Corporation) and the computer program READGEL, which is a version of GELIN (Staden, 1984a). Data were analysed on a VAX 11/750 minicomputer using the programs of Staden (1982b, 1984b). The program DIAGON (Staden, 1982a) was used for the matrix comparison with the MHV JHM nucleocapsid sequence. For comparison of the amino acid sequences a span length of 21 was used. The 'proportional' score of 234, above which a point is plotted, was selected so that only matches significant at the 1% level would be displayed.

RESULTS

The region of sequence presented in this paper is 1836 bases, from the CTTAACAAA homology sequence which occurs at the junction of the leader and body of mRNA A (Brown et al., 1984) to the poly(A) sequence at the 3' end. Thus, the whole of the body of mRNA A is presented for each strain. For the Beaudette clone C5-322 this region has been completely sequenced on both strands. For M41 this region, covered by two clones, has also been completely sequenced on both strands, although since they overlap to some extent this did not entail complete double-stranded sequencing of both clones. For M41 the region covered is only 1652 bases since neither M41 clone contains a sequence of 184 bases that is present in the Beaudette clones. In Fig. 2 the top line of sequence is IBV Beaudette, with the bases that are different in M41 marked underneath. A translation of the main open reading frame is also shown with amino acids altered in M41 indicated.

The sequences from both strains contain a single long open reading frame of 1227 bases, which predicts a polypeptide of molecular weight 45032 for Beaudette and 45214 for M41. These predicted molecular weights agree fairly well with the estimated molecular weight of 51000 (51K) of a polypeptide produced by translation in vitro of mRNA A (Stern & Sefton, 1984) and the molecular weights (50K to 54K) (Metcalf & Sefton, 1977; Nagy & Lomniczi, 1979; Collins & Alexander, 1980; Cavanagh, 1981; Stern et al., 1982) reported for the nucleocapsid polypeptide from IBV virions. It should be noted that the predicted molecular weight does not take into account the phosphorylation of the polypeptide, which in the case of the MHV JHM nucleocapsid increases the apparent molecular weight of the polypeptide from 57K to 60K (Stohlman et al., 1983). The sequences around the putative initiation codon, GUGAUGG, are common among sequences flanking functional eukaryotic initiation codons (Kozak, 1983). The
coding sequences for the two strains are very similar. There are 87 single base changes, the majority of which are purine/purine or pyrimidine/pyrimidine changes. These base changes result in 23 altered amino acids, many of which are alterations of a conservative nature.

Examination of the amino acid sequence shows that it predicts a polypeptide which is enriched in basic residues with an overall positive charge at neutral pH of 19 for Beaudette and 20 for M41. These basic residues are clustered in distinct regions; for example, in the Beaudette sequence they are concentrated in regions from bases 727 to 791 and from 1167 to 1205. The C-terminus of the polypeptide is, however, enriched in acidic residues, with 14 residues out of 47 being aspartic or glutamic acids. The Beaudette and M41 amino acid sequences also contain 28 and 32 serine residues respectively, for which the clustering is even more marked. In the Beaudette sequence for example there is a stretch of 28 amino acids from bases 594 to 677 which contains nine out of the total number of 28 serine residues.

Outside the predicted coding region the major difference between the two strains is the 184 base stretch of sequence present in Beaudette and absent in M41. That this is not a cloning artefact is confirmed by two observations. In the case of M41 the deletion is present in the two independently derived clones. In the case of Beaudette eight independently derived clones have restriction maps which show that they have this stretch of sequence, all containing a HindIII site at position 1436, which is within the 184 base region (data not shown). The deletion in M41 occurs only four bases after the 3' end of the large 1227 base open reading frame.

**DISCUSSION**

The main open reading frame present in these sequences predicts a polypeptide whose size and genome position is consistent with it being the gene for the IBV nucleocapsid polypeptide. The basic nature of this predicted polypeptide would also be expected for an RNA-associated protein. It is very interesting to compare the sequences presented here with sequences published for the nucleocapsid genes of another coronavirus, MHV (Armstrong *et al.*, 1983; Skinner & Siddell, 1983). A comparison of the RNA sequences of IBV Beaudette and MHV JHM (Skinner & Siddell, 1983) has been made using the matrix comparison computer program DIAGON (Staden, 1982a). The RNA sequences appear to be unrelated for most of the sequence with a few
Fig. 2. DNA sequence of 1836 bases of IBV Beaudette cDNA clone C5-322. Beneath the sequence are marked those bases which are different in the M41 clones. Above the sequence is a translation of the main open reading frame in single-letter amino acid code. Where the amino acids differ between the strains the two residues are shown separated by a slash, the Beaudette amino acid being the first. The region of sequence present in Beaudette but not in M41 is underlined with a dashed line.
short stretches of homology. However, a comparison of the predicted amino acid sequences of the polypeptides reveals distinct similarities. Fig. 3 shows these results. A point is plotted at any position where the similarity between two stretches of sequence is above the 1% level of significance. The program not only looks for identical amino acids but also for residues with similar properties. Diagonal lines of points therefore represent regions of high homology. The highest degree of homology, represented by the most prominent diagonal line on the diagram, occurs between amino acids 50 and 150 in the IBV sequence (85 and 185 in the MHV sequence) and there are several other smaller regions. Fig. 4 shows the striking degree of amino acid homology which occurs in part of this region. Another interesting similarity between the two nucleocapsid sequences lies in the distribution of serine residues. The major cluster of serine residues from bases 594 to 677 is mirrored almost exactly in the MHV JHM sequence (although this is not otherwise a region of very high homology between the two sequences), and several smaller clusters of serines also have their counterparts in MHV. This is interesting, since it is known that the nucleocapsid protein of MHV is phosphorylated specifically at serine residues (Stohlman & Lai, 1979; Siddell et al., 1981).
The amino acid differences between IBV Beaudette and IBV M41 are not distributed evenly along the sequence but are restricted to certain regions. Similarly, the amino acid differences between MHV JHM and MHV A59 are clustered and the main region of change in MHV (from amino acids 136 to 162) is paralleled very closely by a major group of altered amino acids in IBV from amino acids 103 to 134. It is probable that these represent part of the nucleocapsid polypeptide which is less important to the function of the molecule.

Although there is very little homology between the RNA sequences of IBV and MHV in the coding sequences and the 5′ non-coding sequences, at one position in the 3′ non-coding sequences there is one very interesting homology (Stern, 1983). This is a 10 base match, GGGGAAGAGCT, which occurs at almost exactly the same position in the two viruses, i.e. 81 bases from the 3′ end for IBV and 82 bases from the 3′ end for MHV. Elsewhere in the 3′ non-coding sequences there is essentially no homology between IBV and MHV. Such a striking homology at a particular position strongly suggests some function, possibly in replication of the putative negative-strand replicative intermediate. This sequence does not occur in the IBV Beaudette leader sequence (Brown et al., 1984) but some homology between the leader and the reverse complement of the 3′ end of the IBV genome (i.e. the 5′ end of the putative negative strand) can be seen. In particular, nine bases from the 5′ end of the negative strand and 24 bases from the end of the leader there is a region in which 10 bases out of 11 match.

It is likely that most of the 3′ non-coding sequence serves an important function, as it is very highly conserved between Beaudette and M41 (three changes out of 315 bases or 99-0%). It also contains many repeated sequences, with the sequence AGTTTA occurring six times, and the 10 base sequence TTAGTTTAAA occurring three times. Bearing in mind the apparent importance of the 3′ non-coding region, it is interesting to speculate on the origin of the 184 base sequence present in Beaudette and absent in M41. The base composition of this stretch of sequence is distinct from the rest of the IBV sequence. It has a very high T content of 50%, which is constant along its length, whereas that part of the 3′ non-coding region present in both strains has a variable base composition which overall is very similar to that for IBV clones sequenced to date. When these differences are considered, and in view of the high homology between the strains in the rest of the 3′ non-coding region, it seems most likely that this sequence is an extra piece of RNA inserted into the genome at some time in the history of the Beaudette strain.

We are grateful to Penny Gatter, Bridgette Britton, Philip Green and Anne Foulds for excellent technical assistance. We wish to thank Dr Martin Bishop, Department of Zoology, Cambridge University, for donating a copy of his computer program READGEL. 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


Sequence of IBV nucleocapsid gene


*(Received 11 October 1984)*