A comparison of the sequences of segment A of four infectious bursal disease virus strains and identification of a variable region in VP2

C. D. Bayliss, 1* U. Spies, 2 K. Shaw, 1 R. W. Peters, 1 A. Papageorgiou, 2† H. Müller 2 and M. E. G. Boursnell 1

1 Institute for Animal Health, Houghton Laboratory, Houghton, Huntingdon, Cambridgeshire PE17 2DA, U.K.
2 Institut für Virologie, Justus-Liebig-Universität Giessen, Frankfurter Strasse 107, D-6300 Giessen, F.R.G.

The nucleotide sequences of the large open reading frame (ORF) from segment A of three European strains of infectious bursal disease virus (IBDV) have been determined using cDNA clones. This ORF of 3036 nucleotides encodes the virion proteins as a polyprotein in the following order: VP2, VP4, VP3. The nucleotide sequences determined have been compared to each other and to the published sequence of an Australian strain. The four strains are closely related, the greatest difference between two strains being 7.7% at the nucleotide level and 2.7% at the amino acid level. Comparisons show that there is a tight cluster of amino acid changes in the virion protein VP2. This variable region corresponded to the region where binding of a neutralizing monoclonal antibody has previously been mapped. A region in the centre of the segment, corresponding to the N terminal of VP4, was found to be completely conserved. Amino acid changes were spread fairly evenly through VP4 and there was no indication of a variable region as found in VP2.

Introduction

Infectious bursal disease virus (IBDV) is a member of the Birnaviridae (Dobos et al., 1979; Müller et al., 1979; Kibenge et al., 1988). These viruses are characterized by a genome that consists of two segments (A and B) of double-stranded RNA and by non-enveloped icosahedral virus particles of about 60 nm in diameter. Four viral proteins have been recognized: VP1, 90K; VP2b, 37K to 40K; VP3, 32K to 35K; and VP4, 24K to 29K (Dobos, 1979; Fahey et al., 1985a). VP2b is processed from a precursor (VP2a) of 41K to 54K. VP1, which is thought to be the viral polymerase, is encoded by the smaller segment, B, whereas the other three proteins are encoded by segment A (Azad et al., 1985). Segment A contains one large open reading frame (ORF) that encodes a 110K polyprotein which is cleaved to release the three viral proteins whose order is NH2–VP2a–VP4–VP3–COOH (Hudson et al., 1986). VP2b and VP3 are the major structural proteins of the viral particles; VP1 and VP4 are present in small amounts (Dobos, 1979; Müller & Becht, 1982).

IBDV causes an acute infection of young chickens which is manifested either as immunodepression or as a clinical illness, depending on the age of the birds. Two serotypes are recognized with the virus neutralization test (McFerran et al., 1980). The serotype 1 strains vary in their degree of pathogenicity, whereas the serotype 2 strains are all non-pathogenic. The serotype 1 strains also have a great variation in their ability to cross-neutralize, some cross-reactions being as low as 19% (McFerran et al., 1980; Jackwood & Saif, 1987).

Neutralizing antibodies are elicited by VP2 (Becht et al., 1988; Fahey et al., 1989). VP3 has been used to induce antibodies but these were only very weakly neutralizing (Fahey et al., 1985b). Monoclonal antibodies (MAbs) have been raised against VP2 and VP3, but only those reacting to VP2 have the ability to neutralize the virus (Azad et al., 1987; Becht et al., 1988; Snyder et al., 1988). One neutralizing MAb recognition site has been mapped to a specific region of VP2 between amino acids 206 and 350 (Azad et al., 1987). This epitope is denatured by SDS and is therefore conformational. These studies with MAbs have begun to elucidate the antigenic structures of the virus and the basis of the variation between strains. In order to increase this understanding, three serotype 1 strains of IBDV have been sequenced and comparisons made between them and the sequence of a previously published Australian strain, 002-73 (Hudson et al., 1986). The three strains were a virulent British isolate, 52/70 (Bygrave &
Faragher, 1970), a British attenuated vaccine strain, PBG 98, (Baxendale, 1976) and a German strain, Cul (Nick et al., 1976). These sequence comparisons showed that overall the viruses are very closely related but they revealed a variable region in VP2 which may be responsible for the antigenic variation observed with these viruses.

Methods

Virus strains. The three serotype 1 strains were a virulent British strain 52/70 (Bygrave & Faragher, 1970) supplied by the Central Veterinary Laboratory, Weybridge, U.K., a British vaccine strain PBG 98 (kindly supplied by Mr W. Baxendale, Intervet, U.K.) and a German strain Cul (Nick et al., 1976).

Purification of virus and viral RNA. Strain 52/70 was purified from homogenates of infected bursae. The homogenates were extracted with arklone (trichlorotrifluoroethane) and the supernatant was layered onto a 60% sucrose cushion. After centrifugation at 16000 r.p.m. for 1 h the virus was collected from the top of the cushion and dialysed. The virus was then centrifuged to equilibrium in a 20 to 60% sucrose gradient. The peak fractions were identified and dialysed to remove the sucrose. The virus particles were digested with 1 mg/ml proteinase K by the method of Azad et al. (1985) and extracted with phenol. Finally the RNA was precipitated from the aqueous phase by addition of 2:5 volumes of ethanol. Strain PBG 98 was grown in chick embryo fibroblasts (CEF) and was harvested from the supernatant by ultracentrifugation at 17500 r.p.m. for 2.5 h. The virus was treated with 0.125 mg/ml RNase A for 30 min at 37 °C to remove cellular RNA and then the virus was purified by centrifugation on two caesium chloride step gradients (Müller et al., 1986). Viral RNA was extracted as above. The Cul strain was also grown in CEFs and viral particles were purified and the RNA extracted as described by Müller et al. (1986).

cDNA synthesis. The dsRNA was denatured by boiling or by treatment with dimethyl sulphoxide (DMSO). Methyl mercury II hydroxide was added to the denatured RNA and random primers were used to prime first strand synthesis, as described by Azad et al. (1985). For strains 52/70 and PBG 98, cDNA synthesis was carried out as described by Gubler & Hoffman (1983), using a cDNA synthesis kit (Amersham). For Cul, cDNA was also synthesized using a cDNA synthesis kit (Pharmacia).

Cloning. The cDNA of 52/70 was cloned by tailing with dCTP and annealing to PstI-cut dG-tailed pBR322 (Bethesda Research Laboratories). PBG 98 cDNA was cloned into pUX1 using a cDNA cloning kit (Amersham). Prior to cloning the cDNA was fractionated on columns to remove material below 600 bp in length, as described in the kit. Cul cDNA was cloned using an EcoRI adaptor, as described in the Pharmacia cDNA synthesis kit, into the EcoRI site of pBSII (Stratagene). Colonies were probed with viral RNA labelled with [32P]ATP using polynucleotide kinase, or with the DNA inserts from IBDV clones isolated from low gelling temperature agarose gels and labelled with [32P]dCTP by random priming with hexanucleotide primers (Feinberg & Vogelstein, 1983).

Sequencing. The clones of 52/70 and PBG 98 were subeloned into the Smal site of M13mp10 by sonication, end repair and blunt end ligation. M13 clones were пробed to identify those containing the insert and sequenced by the dideoxynucleotide chain termination method (Sanger et al., 1977). For sequencing of the Cul cDNA clones, controlled unidirectional deletions were made by the exonuclease III/nung bean nuclease method (Putney et al., 1981; Guo et al., 1983; Henikoff, 1984) and the subsequent plasmid clones were sequenced directly using T7 DNA polymerase (Tabor & Richardson, 1987). The sequences of 52/70 were extended by primer extension and Maxam & Gilbert sequencing as described by Brown et al. (1984). Sequences were analysed on a MicroVAX 3600 computer system using the programs of Staden (1986), as well as the computer programs PRINTDIFF which was written for Fig. 1 and 2 and PROCOMP for Fig. 3.

Results

The cloning of all the IBDV strains employed random primers to construct the cDNA clones. This protocol did not result in the isolation of clones which extended to the ends of the segment. The sequences at the 5′ end of the coding strand of one of the strains (52/70) were therefore determined by primer extension and Maxam & Gilbert sequencing as described in Methods. The last two bases could not be determined unambiguously and are therefore represented by an N. These 5′ sequences revealed that the sequence obtained from clones of the 52/70 strain extended from positions 61 to 3063, position 1 being defined as the beginning of the segment (Fig. 1). Sequence obtained from the clones of PBG 98 extended from positions 187 to 3191, and that from the Cul clones extended from positions 36 to 3250. The previously determined sequence of the Australian strain 002/73 (Hudson et al., 1986) extends from positions 104 to 3232. In order to obtain 52/70 sequences covering the complete large ORF, primer extension and Maxam & Gilbert sequencing was again carried out and sequences for 52/70 were extended to position 3190. Hence, Fig. 1 shows all the sequences up to position 3190. The sequence data for the Cul and 52/70 strains include the whole of the large ORF of segment A. The sequence data for the PBG 98 strain covers over 98% of this ORF, 18 amino acids being missing at the 5′ end.

Over 99% of the nucleotide sequence of the 52/70 clones was determined from at least two separate M13 clones, and 93% was sequenced on both strands; each base was determined a mean of six times. For the PBG 98 and Cul clones, because of the comparative nature of the analysis, the levels of duplication of sequence determination were lower. For PBG 98, 93% of the sequence was obtained from at least two separate M13 clones and in the case of Cul much of the sequence was determined from single plasmid clones.

The largest ORF, which encodes the 110K polypeptide, begins 132 bp from the end of segment A. There is an in-frame stop codon 36 bp 5′ to, or upstream of, the start codon of this ORF. In addition to this initiation codon at position 132, there are two other potential start codons upstream of the polypeptide ORF (Fig. 1). The first of these start codons, at position 66, begins a short ORF of 12 amino acids whose presence is conserved in the two sequences through this region (52/70 and Cul).
Fig. 1. The nucleotide sequences from segment A of four IBDV strains. Differences from the 52/70 strain are shown, identical bases being marked as a dash (-). Asterisks (*) denote sequences not determined, and the x marks a deletion in the Cul strain. Start codons (ATG) discussed in the text are marked with a filled circle (●). The AecI and Spel sites are marked. At position 2688 in 52/70 the base was found to be a G in two cDNA clones and an A in a third, and is here marked as an A, since it causes a change at amino acid position 853.
Fig. 2. The deduced amino acid sequences from the large ORF of four IBDV strains. Differences from the 52/70 strain are shown, identical amino acids being marked as a dash (-). Asterisks (*) denote a region where the sequence was not determined, and mark the stop codon at the end of the sequence. The AccI and SpeI sites which mark the boundary of the region where MAb 17/82 binds (Azad et al., 1987), are shown. Filled circles (●) mark the possible cleavage sites of VP2/VP4 and VP4/VP3. The two hydrophilic peaks in the AccI/SpeI region are underlined.
Table 1. Nucleotide and amino acid changes between four IBDV strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>52/70</th>
<th>PBG 98</th>
<th>Cul</th>
<th>002-73</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleotide changes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52/70</td>
<td>- (-)*</td>
<td>78 (2.6)</td>
<td>81 (2.6)</td>
<td>237 (7.7)</td>
</tr>
<tr>
<td>PBG 98</td>
<td>78 (2.6)</td>
<td>- (-)</td>
<td>9 (0.3)</td>
<td>229 (7.6)</td>
</tr>
<tr>
<td>Cul</td>
<td>81 (2.6)</td>
<td>9 (0.3)</td>
<td>- (-)</td>
<td>232 (7.5)</td>
</tr>
<tr>
<td>002-73</td>
<td>237 (7.7)</td>
<td>229 (7.6)</td>
<td>232 (7.5)</td>
<td>- (-)</td>
</tr>
<tr>
<td>Amino acid changes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52/70</td>
<td>- (-)†</td>
<td>17 (1.7)</td>
<td>15 (1.5)</td>
<td>27 (2.7)</td>
</tr>
<tr>
<td>PBG 98</td>
<td>17 (1.7)</td>
<td>- (-)</td>
<td>5 (0.5)</td>
<td>24 (2.4)</td>
</tr>
<tr>
<td>Cul</td>
<td>15 (1.5)</td>
<td>5 (0.5)</td>
<td>- (-)</td>
<td>23 (2.3)</td>
</tr>
<tr>
<td>002-73</td>
<td>27 (2.7)</td>
<td>24 (2.4)</td>
<td>23 (2.3)</td>
<td>- (-)</td>
</tr>
</tbody>
</table>

* Numbers in parentheses are the percentage differences in nucleotide changes for the IBDV strains.
† Numbers in parentheses are the percentage differences in amino acid changes for the IBDV strains.

The second start codon, at position 98, begins an ORF potentially encoding a 17K protein which is observed in all the sequences. The start codon at position 66 is in a favourable context for the initiation of translation (Kozak, 1987) and the initiation codon for the polyprotein is also in a fairly favourable context. The initiation sequence for the potential 17K protein at position 98 is poor and would be termed non-functional by the rules of Kozak (1987).

In Fig. 1 the sequences from the three strains cloned in this study are presented. The 52/70 sequence is used as the basic sequence in the comparison and the previously published 002-73 sequence (Hudson et al., 1986) is also presented for comparative purposes. Table 1 shows the differences between the strains in terms of numbers of changes and as percentages. These comparisons show that the strains are very closely related, the greatest difference being between 52/70 and 002-73 with 7.7% of bases being different. An intriguing feature of this comparison is that Cul and PBG 98 have only nine base changes between them (0.3%).

In Fig. 1, 2 and 3, namely the AccI (amino acid 206) and SpeI (350) sites. All the differences found in VP2 between the three European strains 52/70, Cul, and PBG 98 are in this variable region, and the majority (18/20) of the differences in VP2 between these European strains and the Australian strain 002-73 are also in this region.

Although the remaining differences between the strains are not highly clustered they are entirely confined to the right-hand end of the ORF between amino acids 677 and 992. In the central region of the ORF, the sequence is completely conserved between amino acids 333 and 676. The majority of this region probably encodes VP4. Fig. 2 and 3 show the most likely position of VP4 between amino acids 451 and 721, this estimate being based on size considerations and on the positions of dibasic residues which have been suggested as possible cleavage sites of the polyprotein (Jagadish et al., 1988).

It is interesting to note that these comparisons between IBDV strains show that the VP4 region is more conserved than the rest of segment A. In contrast when comparisons are made between the amino acid sequences of IBDV and a different birnavirus, infectious pancreatic necrosis virus (Duncan & Dobos, 1986), it is VP2 and VP3 that are more conserved, VP4 showing almost no homology between the two species.
Discussion

Hudson et al. (1986) have reported the nucleotide sequence of 3128 bp of segment A of an Australian strain (002-73) of IBDV. This includes the sequence of the polyprotein ORF, as well as 28 bases and 62 bases 5' and 3' to the ORF, respectively. In this paper we present the sequences from the polyprotein ORF of three IBDV strains, a British virulent strain (52/70), a British vaccine strain (PBG 98) and a German strain (Cul). In addition the complete 5' sequence of the coding strand of one strain (52/70) has been determined. This 5' sequence demonstrates that the polyprotein ORF does not extend upstream beyond the initiation codon identified by Hudson et al. (1986).

This upstream sequence has some interesting features which may be relevant to the regulation of translation of this virus. Before this is discussed it should be noted that the ends of the mRNA of this virus have not been mapped and that it is assumed that the mRNA is the same length as the coding strand of the genomic dsRNA. This assumption is based on evidence that the virus replicates by a strand displacement mechanism (Spies et al., 1987) and that the separated strands of the genomic dsRNA can be translated in vitro (Azad et al., 1985). The first feature of interest is the length of the 5' non-coding region which is 132 bp. This is much longer than those of other dsRNA viruses such as reoviruses, bluetongue and rotaviruses. The presence of two initiation codons upstream of the polyprotein start codon is also of interest, the first being at the start of a small ORF which potentially encodes a 12 amino acid polypeptide and the second potentially initiating translation of a 17K polypeptide, whose sequences would overlap those of the polyprotein ORF. Overlapping ORFs encoding two proteins have been found in one segment of a reovirus (Ernst & Shatkin, 1985). In addition upstream ORFs have been shown to modify the translation of a downstream gene, on bicistronic mRNAs, in other viral systems (Sedman et al., 1989). Hence, it is possible that in IBDV these ORFs may modify the translation of the polyprotein ORF. The apparent conservation of the positions of the two upstream ORFs between these four IBDV strains also indicates that they may be functional and that their products may have a role in the life cycle of the virus. Although a 16K polypeptide has been observed after in vitro translation of segment A (Azad et al., 1985), a protein of this size has not been observed in infected cells and its poor context for initiation of translation (Kozak, 1987) implies that it may not be expressed. This has yet to be resolved.

The sequence comparisons show that these viruses are closely related, with between 92.3% and 99.7% conservation at the nucleotide level (Table 1). Table 1 also shows that the 52/70 strain is more closely related to the other European isolates, Cul and PBG 98, than to the geographically isolated Australian strain 002-73. This result is most clearly demonstrated in the top line of Table 1, where the number of nucleotide differences between 52/70 and the other strains is shown. The high level of similarity of PBG 98 and Cul is difficult to explain. PBG 98 was derived from a mildly pathogenic strain 4/68, which was isolated in the U.K. in 1968, by adaptation to tissue culture (W. Baxendale, personal communication) and was marketed as a vaccine from 1974. Cul was isolated from infected chickens in West Germany in 1975 (Nick et al., 1976). One obvious possibility is that Cul could be a derivative of PBG 98 if the chickens from which it was isolated had been vaccinated with PBG 98; however, there is no record of the use of the PBG 98 vaccine in Germany.

Although the overall sequence similarity is high, the amino acid comparisons have revealed a highly variable region within VP2. VP2 has been identified as the protein that contains the neutralizing epitopes of IBDV by the fact that all of the neutralizing MAbs isolated bind to this protein (Azad et al., 1987; Becht et al., 1988). Azad et al. (1987) identified the binding site of one of these MAbs (17/82) by expressing gradually shortened clones of the whole segment of IBDV in Escherichia coli. The amino acid sequence from 206 to 350 was the shortest region identified to which the antibody would bind. Azad et al. (1988) showed that all their neutralizing MAbs competed for binding with the 17/82 MAb. Therefore it appears that this region of VP2 contains the major neutralization site of the virus. Cross-neutralization tests (McFerran et al., 1980; Jackwood & Saif, 1987) have been used to define antigenic variation of the virus. The fact that the variable region we have identified in VP2 coincides so accurately with the major neutralization site would suggest that the large number of amino acid changes seen here are an important cause of antigenic variation of the virus. However, limited data on the cross-neutralization of three of these strains (PBG 98, Cul and 002-73) shows that they will cross-neutralize (Wood et al., 1988) but to what extent is not clear. Thus it is uncertain how the amino acid changes found in this region affect cross-neutralization and therefore antigenic variation. This will need further definition, possibly by mapping the binding sites of other neutralizing MAbs.

Azad et al. (1987) found that if either end of the 206 to 350 region was deleted, the binding of the 17/82 MAb was abolished. Hydrophilicity profiles of this region show that there are two hydrophilic peaks at either end of this region, the larger peak being from amino acids 212 to 224 and the other from 314 to 324 (these peaks are underlined in Fig. 2). Although it might be expected that
the main antigenic sites would reside within these hydrophilic regions, it is interesting to note that the majority of amino acid variation in this region falls between them. Wood et al. (1988) have shown that the 17/82 MAb will neutralize the PBG 98 and Cul strains in vitro. It seems likely, therefore, that the binding site of 17/82 lies within those sequences which are conserved between these three strains, for example the hydrophilic peaks. The structure of this region of VP2 will be elucidated further as the binding site of MAb 17/82 is more accurately defined and as the binding sites of other MAbs are mapped.

The exact cleavage positions of the polyprotein are unknown but dibasic residues at positions 552 and 722 have been postulated as possible sites (Azad et al., 1987; Jagadish et al., 1988) on the basis of the known M values of the proteins. These have been marked on Fig. 2. Work on the cleavage of the polyprotein by VP4 (Jagadish et al., 1988), which is thought to be a protease, does not contradict these possible sites. The most striking feature of the distribution of changes in VP3 and VP4 is the fact that the N-terminal portion of VP4 is completely conserved. It may be that this part of the VP4 molecule is essential for correct activity of the VP4 protease and that it cannot be varied. VP3 and VP4 have never been conclusively identified as having a role in neutralization of IBDV and therefore the changes in these proteins may well not be involved in antigenic variation of IBDV.

Some of the observed differences between 52/70 and the other strains are likely to be connected with virulence determinants. 52/70 is a virulent strain which causes mortality of between 30% and 50%, whereas PBG 98 is apathogenic (Baxendale, 1976). Cul also appears to be more pathogenic than PBG 98, causing necrosis of bursal follicles (Lange et al., 1987), thus any changes in pathogenicity, between these two strains, caused by differences in the proteins encoded by segment A must be due to the five amino acid changes in VP2. At the moment, however, it is impossible to pinpoint the crucial differences. The possibility of a contribution to changes in pathogenicity by differences in segment B must also be considered (Müller, 1988). Attenuation of a pathogenic strain and subsequent sequencing, or sequencing of MAb-resistant variants, will be necessary to identify the exact sequences involved in antigenic variation and pathogenicity.

We would like to thank Teresa Berrill for excellent technical assistance, and Mr W. Baxendale for providing the PBG 98 strain of IBDV. C. D. B. was funded by an AFRC research studentship. U. S.’ work was supported by the Deutsche Forschungsgemeinschaft and is presented in partial fulfilment of the requirement for Dr. rer. nat. at the Justus-Liebig-Universität, Giessen.

References


(Received 13 October 1989; Accepted 15 December 1989)