J. gen. Virol. (1988), **69**, 621–629. *Printed in Great Britain*

Key words: *infectious bronchitis virus/nucleotide sequence/evolution*

Evolution of Avian Coronavirus IBV: Sequence of the Matrix Glycoprotein Gene and Intergenic Region of Several Serotypes

By DAVID CAVANAGH* AND PHILIP J. DAVIS

AFRC Institute for Animal Disease Research, Houghton Laboratory, Houghton, Huntingdon, Cambridgeshire PE17 2DA, U.K.

(*Accepted 12 November 1987*)

SUMMARY

We have sequenced 200 to 240 bases of the matrix (M) glycoprotein gene of 23 strains of infectious bronchitis virus (IBV) representing the A (D207), B (D3896), C (D3128), D (D212), Massachusetts (Mass), UK11 and UK12 serotypes. The bases examined code for the external, hydrophilic region and the first membrane-embedded hydrophobic region of M, both regions comprising approximately 20 amino acids. As predicted from protein M_r studies the A/D and B/C serotypes had two and one potential glycosylation sites respectively. This variation appeared to derive from a combination of base substitutions and deletions/insertions. The glycosylation sequence Asn-Cys-Thr was highly conserved. Overall, the exposed part of M exhibited a fourfold greater extent of amino acid variation than did the membrane-embedded sequence. The transcription-associated homology region sequence (CUUAACAA) in the 5′ intergenic region was identical in all strains but there was considerable variation as to its location. The M gene of UK12 appeared to have evolved from a group A-like M gene by a two stage process involving a base substitution in the intergenic region which generated a new AUG translation start codon followed by deletion of the original AUG. Isolate UK11 closely resembled Mass strains in the intergenic region but was dissimilar from all strains in the protein coding region. The M sequences of serotypes B and C were identical and those of the A and D serotypes very similar. These results are discussed in relation to recent sequencing of part of the spike glycoprotein gene of some of these strains and the discovery of *in vitro* recombination of murine hepatitis coronavirus.

INTRODUCTION

Infectious bronchitis virus (IBV) is a major cause of disease in the domestic fowl and is economically very important to the poultry industry. In the years 1978 to 1983 isolates of four new serotypes, defined on the basis of neutralization tests, were isolated in The Netherlands and the U.K. from diseased birds that had been vaccinated against IBV (Davelaar *et al.*, 1983; Cook, 1984). At that time IBV vaccines in Europe contained only strains of the Massachusetts (Mass) serotype.

The virus has, in addition to the large spike (S) glycoprotein which induces neutralizing antibodies (Cavanagh *et al.*, 1984), a smaller membrane or matrix (M) protein. Up to about 20 of its approx. 225 amino acids are believed to protrude at the outer membrane surface (Boursnell *et al.*, 1984; Cavanagh *et al.*, 1986*a*). Unlike the M proteins of several other viruses, that of coronaviruses is glycosylated. In the case of IBV strain Beaudette (Mass serotype), the M gene of which has been cloned and sequenced (Boursnell *et al.*, 1984), the 30000 (30K) M_r M glycopolypeptide has two glycans *N*-linked to asparagine residues at positions 2 and 5 from the N terminus of the mature M glycopolypeptide (Cavanagh, 1983*a*).

Electrophoretic analysis of the polypeptides of the recent isolates revealed that the serotypes had an M glycopolypeptide of either 27K or 30K, from which it was concluded that the numbers of glycans present were one and two respectively (Cavanagh & Davis, 1987). We sought to

confirm this by sequencing that part of the M gene, present in the plus-stranded virion RNA, which codes for 40 or so amino acids in the N-terminal half of the polypeptide. Such sequencing would also indicate whether a simple base substitution had led to the loss or gain of a glycosylation site.

We have also obtained sequence data from part of the non-coding (intergenic) region upstream from the translation initiation codon. Cloning and sequencing of IBV Beaudette genes (Brown & Boursnell, 1984; Brown *et al.*, 1984; Boursnell *et al.*, 1985) have shown that within the intergenic region to the 5′ side of each open reading frame (ORF) for which there is a corresponding mRNA there is a common sequence CUUAACAA. This sequence, termed the 'homology region', is believed to be involved in a recognition process between a transcriptase-associated leader sequence, derived from the 5′ terminus of the genome, which results in the generation of mRNAs (Lai *et al.*, 1983; Brown *et al.*, 1984). It was of interest to compare IBV strains isolated in Europe and the U.S.A. to determine the extent to which the composition and location of the homology region was conserved.

Finally it was considered that the sequencing of some 200 bases of the M genes would help reveal relationships between IBV isolates. Such data would be a contribution to the goal of establishing the nature of the evolution of IBV and understanding its epidemiology.

METHODS

Virus strains. Isolates obtained in The Netherlands (prefixed D; 1978 to 1979) and in the United Kingdom (prefixed UK; 1981 to 1983) have been described previously (Cavanagh & Davis, 1987) and comprise serotype A [strains D207, UK1, UK2 (also referred to as 6/82 in some publications), UK5, UK6 and 7/82], serotype B (strains D3896 and UK8), serotype C (strains D3128 and UK9), serotype D (strains D212 and D1466) and strain D274 which is serologically related to both A and B serotypes (Davelaar *et al.*, 1983; Kusters *et al.*, 1987). The serotypes designated A to D do not correspond directly to the A to D serotypes of Kusters *et al.* (1987). Strains UK11 and UK12 were isolated in the U.K. in 1984. On the basis of neutralization tests they are distinct from one another and from the other serotypes examined in this study (Cook & Huggins, 1986). The remaining strains were of the Massachusetts (Mass) serotype and included IBV Beaudette and M41, isolated in the U.S.A. in 1935 and 1941 respectively, five vaccine strains, H52, H120 and Bronchimmune (Smith Kline, Stevenage, U.K.) MM (Salsbury Laboratories, Southampton, U.K.) and IBVAX (IVAZ Poultry Vaccines, Padova, Italy) the original isolates of which had been obtained in The Netherlands or the U.S.A. between the late 1950s and mid-1970s, and one U.K. field isolate, possibly a reisolation of a vaccine strain, obtained in 1980.

Production of virion RNA. Virus was grown in embryonating eggs and purified essentially as described by Cavanagh (1983*b*); the final sucrose gradient step was omitted in many cases. RNA was extracted (Brown & Boursnell, 1984) and stored at −70 °C as an ethanol precipitate. The amount of RNA used for sequencing was equivalent to the yield of purified virus from one or two eggs.

RNA sequencing. A synthetic oligonucleotide JMM (GCACCATAACACTATC) complementary to base positions 219 to 234 of the UK2 M gene sequence (Binns *et al.*, 1986*a*) was used to prime the reverse transcription of IBV virion RNA. IBV RNA and 10 to 20 ng of JMM in 4 μl of 5 mM-Tris–HCl pH 8·3 were heated at 80 °C for 4 min, the mixture was cooled on ice and then mixed with 2·5 μl of reverse transcriptase (RTase) buffer (800 mM-Tris–HCl pH 8·3, 1·12 M-KCl, 80 mM-$MgCl_2$ and 163 mM-2-mercaptoethanol), 1·5 μl of water and 2·0 μl of 10 mM-Tris–HCl pH 8·3 containing 10 μCi of [α-^{32}P]dATP (sp. act. approx. 3000 Ci/mmol; Amersham) to form the hybridization mix. RTase mixture contained 40 μM-dATP, 820 μM-dCTP, -dGTP and -dTTP, 200 units/ml of AMV RTase (Super RT; Anglian Biotechnology, Colchester, U.K.) and one of 3·2 μM-dideoxy-ATP (ddATP), 100 μM-ddCTP, 120 μM-ddGTP or 120 μM-ddTTP. The sequencing reaction was initiated by mixing 2 μl volumes of hybridization mix with 2 μl of each of the RTase mixes. After incubation at 42 °C for 15 min 4 μl of formamide dye mix (formamide containing 0·03% of each of xylene cyanol and bromophenol blue and 20 mM-EDTA) was added and then the mixture was heated at 100 °C for 2 min. Each sample was analysed on two 6% polyacrylamide gels. These were 50 cm long, 0·3 mm thick and made as described by Maniatis *et al.* (1982). One gel was run at 35 W constant power until the bromphenol blue had reached the bottom of the gel and the other gel was run for 1 h after the xylene cyanol had migrated out of the gel. After fixation in 10% acetic acid and methanol the gels were dried under vacuum onto Whatman 3MM paper and exposed to Kodak XAR5 X-ray film.

RESULTS

The first 20 or so N-terminal amino acid residues of M form a sequence of hydrophilic character which is believed to be exposed at the outer membrane surface of the virus (Boursnell *et al.*, 1984; Cavanagh *et al.*, 1986*a*). Approximately the next 80 residues form three

Serotype	Strain	M_r of M	Amino acid sequence
			** *
UK12	1 strain	ND	MVGNNTTNCTLGTEQAVQ-FKEY-LFVTAFLLFLTILLQYGYAT-
			* *
A	6 strains	30K	MDNTTNCTLGTEQAVQLFKEYNLFVTAFLLFLTILLQYGYATR
			* *
A/B	D274	30K	MDNTTNCTLGTEQAVQLFKEYNLFVTAFLLFLTILLQYGYATR
			* *
D	D212	29K	MANTTNCTLGTEQAVQLFKEYNLFVTAFLLFLTILLQYGYATR
			* *
D	D1466	30K	MDNATNCTLGTEQAVQLFKEYNLFVTAFLLFLTILLQYGYATR
			*
B and C	4 strains	27K	MTENCTLDTEQAVQLFKEYNLFITAFLLFLTILLQYGYATR
			*
UK11	1 strain	ND	MENCTLDAEQAVQLFKDY-LFITAFLLFLTILLQYGYATR
			* *
Mass	7 strains	30K	MSNETNCTLDFEQSVELFKEYNLFITAFLLFLTIILQYGYATR
			* *
Mass	Beaudette	30K	MPNETNCTLDFEQSVQLFKEYNLFITAFLLFLTIILQYGYATR
			+++++++ ++ + + + + +
			---------- Hydrophobic sequence

Fig. 1. Deduced amino acid sequence of the exposed hydrophilic and first membrane-embedded hydrophobic regions of the M glycoprotein of several serotypes of IBV. Where the identity of an amino acid residue could not be determined unequivocally its location is marked by a short line (–). The asparagine (N) residue of potential glycosylation sites is marked (*) and the NCT glycosylation site common to all isolates is solidly underlined (——). The hydrophobic sequence is indicated by a dashed line (– – –). Positions at which the amino acid of at least two serotypes differ are shown by crosses (+). ND, M_r of glycosylated M not determined.

hydrophobic sequences which are considered to span the membrane three times, the carboxy-terminal half of the protein being essentially on the internal side of the membrane (Boursnell *et al.*, 1984; Rottier *et al.*, 1986). In this communication the nucleotides coding for the N-terminal hydrophilic and the first hydrophobic regions were sequenced.

Number of glycosylation sites

Inspection of the deduced amino acid sequences revealed that serotypes A, D and Mass have two potential glycosylation sites on M while the B and C serotypes have only one (Fig. 1). The glycosylation site present in the B and C serotypes was analogous to that site of the Mass, A and D serotypes which was furthest from the N terminus. Indeed this common glycosylation site had a highly conserved sequence NCT (part of a conserved tetrapeptide NCTL) while there was amino acid variation in the flanking sequences. We have also sequenced part of the M gene of IBV isolates UK11 and UK12. Strain UK11 had only one glycosylation site whereas UK12 had three potential glycosylation sites. We do not know whether this additional site is used.

Amino acid differences between serotypes

The variation in the number of potential glycosylation sites was not simply a result of base substitutions leading to the loss or gain of an asparagine residue or of a change in the amino acid two residues downstream from the asparagine, which has to be either threonine or serine for *N*-linked glycosylation to occur. Rather those strains with only one glycosylation site had a shorter N terminus than those with two sites (Fig. 1). Strain UK12 had the longest N terminus which, as reported below, was a result of the extension of the ORF into the previously non-coding region of the M gene. Although some of the differences between the M proteins are explained in terms of the acquisition or deletion of bases there were also amino acid differences arising from base substitutions in the RNA sequence. Comparison of serotype A with the other serotypes reveals that there were many more differences in the hydrophilic sequence than in the hydrophobic region (Table 1). Indeed overall, and considering only those amino acid differences that can be attributed to base substitutions in the RNA, there were fourfold more amino acid differences in the hydrophilic than hydrophobic sequences.

Table 1. *Differences in the amino acid composition of the hydrophilic and the first hydrophobic region of the M glycoprotein of serotype A compared with other IBV serotypes*

Serotype A compared with serotype/strain	Hydrophilic region	Hydrophobic region
D274	0 (0)*	0 (0)
D212	1 (5)	0 (0)
D1466	1 (5)	0 (0)
UK12	4 (17)	0 (0)
B and C	5 (23)	1 (5)
Mass	6 (27)	2 (10)
UK11	8 (35)	1 (5)

* Number of differences with % in parentheses.

Sequence and location of the homology sequence

Sometimes there were bands of equal intensity (cross-bands) in each of the four sequencing tracks, preventing unequivocal identification of the base at that location. These positions are marked by an X in the figures and were disregarded when isolates were being compared. Fig. 2 shows the consensus sequences of part of the intergenic regions of the M genes presented such that the homology regions (CUUAACAA) are aligned. The AUG translation initiation codon follows immediately after the 3′-terminal base shown in Fig. 2. The octanucleotide homology region was completely conserved among all the strains examined and there was much homology between the flanking sequences of different serotypes. The most striking difference was the variability in the number of bases to the 3′ side of the homology region. Thus the Mass strains had 23, 24 and 36 bases more than the B/C, A/D and UK12 serotypes respectively, the additional bases being adjacent to the ORF. Thus the homology sequence was conserved with respect to its sequence but not to its location in relation to the translation initiation codon.

Extension of the ORF by mutation within the intergenic region

A probable evolutionary relationship between isolate UK12 and a strain with an M gene resembling that of the A serotype was deduced when the nucleotide sequences of both the non-coding and coding regions were analysed simultaneously (Fig. 3). Comparison of the base sequence of UK12 and the A serotype revealed that the first 12 non-coding nucleotides of group A were within and constituted the first 12 bases of the ORF of the M gene of UK12. The probable sequence of events was the substitution of a U for a C within the intergenic region of the M gene of a group A strain, resulting in the generation of an AUG translation initiation codon in frame with the existing ORF. Subsequently the original AUG codon, together with the GAU codon at its 3′ side, was deleted. The net result was an M protein with two additional amino acid residues at the N terminus compared to the A/D strains.

Nucleotide sequence comparisons between the serotypes

The starting point of this work was the analysis of the A, B, C and D serotypes which had been isolated in The Netherlands and the United Kingdom between 1978 and 1983. The Mass strains were included to help determine how close was the relationship of the field isolates to Mass serotype vaccine strains. The nucleotide sequences coding for the intergenic and protein coding sequences are shown in Fig. 2 and 4 and are summarized in Table 2. Within each serotype there was a high degree of homology.

As discussed above, the M gene of strain UK12 appears to have evolved from the M gene of a strain which resembled that of the A group. Further evidence for this claim is that apart from the single base change involved in the mutation of an ACG to give an AUG translation initiation codon and the deletion of six consecutive bases from UK12, there was only one other base difference between these strains, as shown in Fig. 3.

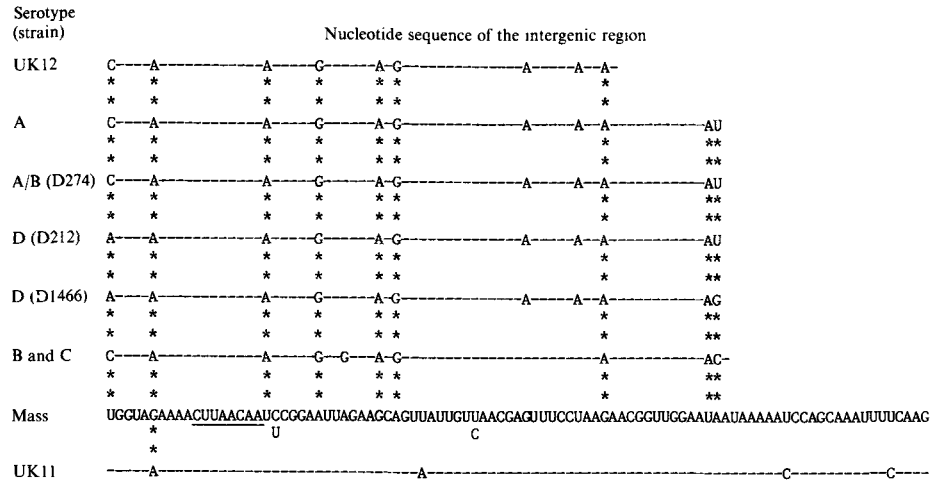

Fig. 2. Nucleotide sequence of part of the intergenic region. The sequences were aligned with reference to the homology region (CUUAACAA; underlined). The AUG translation initiation codon is situated immediately to the 3′ (right hand) side of the sequences shown. The consensus sequence for the Mass strains is shown in its entirety, with variations shown underneath. For the other serotypes only those bases which differ from those of the Mass strains are shown. The locations where all of the serotypes A, B, C, D and UK12 differ from Mass are marked (*).

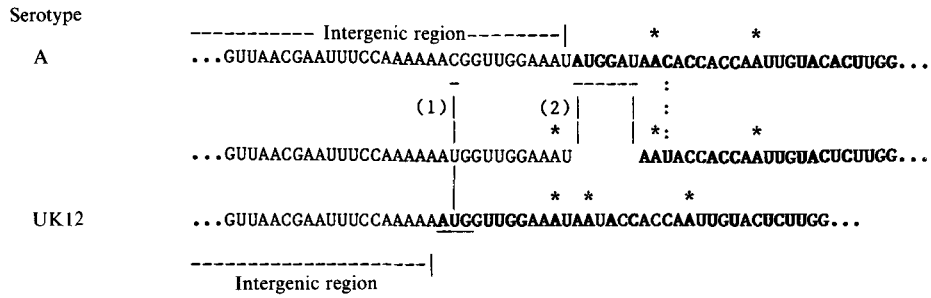

Fig. 3. Suggested evolution of the M gene of isolate UK12 from the M gene of a serotype A-like strain. Part of the intergenic sequence (normal type) and protein-coding sequence (bold type) are presented. It is proposed that initially (1) a C mutated to a U, thus creating a translation initiation codon (AUG; ——) in-frame with the ORF. Subsequently (2) the original translation initiation codon with three adjacent bases (---) was deleted. The bases coding for the asparagine residues to which glycans can potentially be attached are shown (*) and the location of one other base difference between the two serotypes (:).

Table 2. *Nucleotide differences between the sequence of part of the M gene and the intergenic region of serotype A and other IBV serotypes*

Serotype A compared with serotype/strain	Number of differences in each region				
	Intergenic	Hydrophilic	Hydrophobic	Total	%
D274	0 (0)*	0†	1	1 (1)	0·5 (0·5)‡
D212	1 (1)	1	0	2 (2)	1 (1)
D1466	2 (2)	1	0	3 (3)	1·5 (1·5)
B and C	4 (5)	14	11	29 (30)	15 (15)
Mass	11 (35)	15	9	35 (59)	17 (26)
UK11	12 (36)	22	9	43 (67)	21 (31)

* The figure in parentheses is the number of base differences including those bases present in the intergenic region of the Mass and UK11 strains but absent in the other strains (see Fig. 2). Serotypes B and C had one base more than serotype A in the intergenic region.

† These numbers include bases which were present in serotype A but absent from serotype B and UK11 strains.

‡ The figure in parentheses is the percentage of base differences including those bases present in the intergenic region of the Mass and UK11 strains but absent in other strains.

(*a*)

Serotype	Strain	Nucleotide sequence
Mass	4 strains	AUGUCCAACGAGACAAAUUGUACUCUUGACUUUGAACAGUCAGUUGAGCUUUUUAAAGAGUAUXAU
Mass	3 strains	----------A---------- ---------- ---------- ---------- ----------
Mass	Beaudette	---C---------- ---------- ---------- ---------- ----------C----------
UK11	1 strain	AU-GA------C--A------GC-------G-----C----------U------
B and C	4 strains	-UGACAGA---------UAC-------AG-----C----C-----G--A------
A	6 strains	---GAU---ACC--C----------GUAC------AG-----C---------G--A------
A/B	D274	---GAU---ACC--C----------GUAC------AG-----C---------G--A------
D	D212	---G-U---ACC--C----------GUAC------AG-----C---------G--A------
D	D1466	---GAU----CC--C----------GUAC------AG-----C---------G--A------

(*b*)

Serotype	Strain	Nucleotide sequence
Mass	4 strains	UUAUUUAUAACUGCAUUCUUGUUGUUCUUAACCAUAAUACUUCAGUAUGGCUAUGCAACAAGAAGUAAC
Mass	3 strains	---------- ---------- ---------- ---------- ----------U----------
Mass	Beaudette	---------- ---------- ---------- ---------- ---------- ----------
UK11	1 strain	C----------C------C-------UC--------C----------C----------U--G---C--
B and C	4 strains	----------C--C---C--------UC-U--U---C-U----------A--C-----U--G--CCGU
A	6 strains	C-----G----------C----------C-X----------A--C-----U--G--C---
A/B	D274	C-----G----------C----------C-U----------A--C-----U--G-UC
D	D212	C-----G----------C----------C-U----------A--C-----U--G--C---
D	D1466	C-----G----------C----------C-U----------A--C-----U--G--C---

Fig. 4. Comparison of the nucleotide sequence of four Mass strains coding for (*a*) the hydrophilic N terminus and (*b*) the hydrophobic, membrane-associated region of M. A short line (–) indicates that the nucleotide at that position is the same as in the group of four Mass strains, the complete sequence of which is given on the top line. Three vaccine strains had an identical sequence to M41. The second line shows the base differences of a group comprising two other vaccine strains and a field isolate. In (*a*) the first three nucleotides of each line are AUG, the translation initiation codon.

Isolate UK11 proved to be very interesting because its intergenic region closely resembled that of the Mass serotype (Fig. 2) while its coding region differed substantially from all the serotypes examined (Fig. 4). The most obvious similarity was that UK11 possessed the intergenic region bases of the Mass serotype that were absent from the other isolates examined (Fig. 2). In the non-coding region there were nine positions at which Mass strains differed from all the A, B, C and D serotypes. At eight of these positions UK11 was identical to the Mass serotype. There were only two (3%) differences between UK11 and Mass in the bases shared with the A to D serotypes compared with 14% for the A to D serotypes. In the non-coding bases that were unique to UK11 and Mass there were two base differences. When the base sequences of the coding regions were compared it was found that UK11 differed from all the serotypes examined, including Mass, by 19 to 22 bases, excluding those bases present in serotypes A to D and Mass but absent from UK11. Approximately half of these base differences had occurred in the region that coded for the membrane-associated amino acid sequence.

DISCUSSION

We have previously proposed (Cavanagh & Davis, 1987) that IBV strains of serotypes A to D with an M glycoprotein of M_r 27K or 30K have one and two glycans respectively. Our M gene sequencing has, by identifying potential glycosylation sites, shown that this proposal is correct (Fig. 1). The N-terminal hydrophilic exposed sequence of M exhibited more than a fourfold greater amino acid variation than the first hydrophobic sequence (Fig. 1). The inference that the exposed part is the most variable part of the molecule is supported by a comparison of the complete sequence of the M gene of a serotype A strain with that of a Mass strain (Binns *et al.*, 1986*a*). The M genes differed at 18 (8%) of the residues, 28% of the amino acid differences being in the N-terminal hydrophilic sequence although this sequence accounts for only 9% of the total amino acids in M. The M glycoprotein of murine hepatitis coronavirus (MHV) has an 87% amino acid homology with bovine coronavirus (BCV) (Lapps *et al.*, 1987) and a 35% homology with IBV. Further inspection of the data of Lapps *et al.* (1987) reveals that 22% of the differences between M of BCV and MHV occur in the exposed part which accounts for only 10% of the total length of M. Expressed another way, there were differences at 29% of the amino acid residue positions within the N-terminal hydrophilic sequence. Thus the exposed part of M of mammalian coronaviruses would appear to share with that of avian IBV a hypervariability

with respect to the exposed part of the molecule. It is noteworthy that variation between the exposed M sequence of different serotypes of IBV is in some cases as high as the differences between the equivalent part of BCV and MHV (Table 1). This region of M would have a net negative charge at around neutrality for IBV (ranging from −1 to −3), BCV (−1) and MHV (−3). Whether the observed high degree of variation in the N terminus of M is of biological significance is not known. Approximately half of the amino acids have been conserved among the strains that we have examined. It would appear that variation of the other residues, especially those to the N-terminal side of the conserved glycosylation site, are readily tolerated and of little consequence with regard to the function of the protein. This variation may, however, have some relevance to protective immune responses. Although purified M protein of MHV (Hasony & MacNaughton, 1981) and IBV (Cavanagh *et al.*, 1984) did not induce detectable neutralizing antibodies, some monoclonal antibodies to the M protein of transmissible gastroenteritis coronavirus (Woods & Wesley, 1987) and MHV (Collins *et al.*, 1982) did neutralize virus in the presence of complement. Since so few amino acids of M are exposed at the virion surface a small number of residue changes might have a great effect on the specificity of the immune responses induced by M. The immune responses directed against the exposed part of M of a Mass vaccine strain, following vaccination, may operate with much reduced efficiency against a subsequently infecting IBV strain with a heterologous M protein. The loss or gain of a glycan could also have a profound effect on the antigenicity of the exposed part of M. Although the S glycoprotein of IBV is believed to be the prime inducer of protective immunity (Cavanagh *et al.*, 1986*b*) it is possible that immune responses to other IBV proteins might also have a role. The relative importance of these other immune responses might be greater when a chicken is infected by an IBV strain whose S glycoprotein is antigenically significantly different from that of the virus used to vaccinate the bird. In addition the capacity of a field strain which is heterologous with respect to both S and M, to break through the immunity of a vaccinated chicken may be greater if the birds have been improperly vaccinated, the immunity has declined, husbandry is poor, or a combination of all these factors.

Much of the sequence between the homology region and the ORF would appear not to be of great importance since it varies among all the serotypes that we have examined. The Mass serotype has twice as many such bases as strain UK12 (Fig. 2). However, even the latter strain has many more bases in this location than do MHV and BCV which have only three bases before the ORF (Armstrong *et al.*, 1984; Pfleiderer *et al.*, 1986; Lapps *et al.*, 1987). One consequence of a large number of bases to the 3′ side of the homology sequence is that there is a greater possiblility for mutation which might lead to elongation of the M protein, as was the case for strain UK12. There is virtually no scope for this to happen with the M genes MHV and BCV.

UK11 closely resembled Mass strains in the intergenic region but was substantially different in the protein-coding region. Comparison of group A strains with the B/C isolates also shows a higher degree of homology in the intergenic region than in the coding region (Table 2). These results cannot simply be explained in terms of selection of mutants in the coding region under the pressure of antibody or other immune responses since many of the base differences occurred in the sequence coding for the non-exposed, membrane-embedded part of M and because the majority of these base differences did not result in changes in amino acids. As illustrated in Fig. 2 the 36 intergenic bases adjacent to the beginning of the Mass and UK11 ORF are not essential. It might be expected, therefore, that mutations in this region would occur and be tolerated as frequently as in the hydrophobic polypeptide coding sequence.

It is noteworthy that with those strains in which the intergenic regions are more closely related than the protein-coding sequences (serotypes A and D compared with B and C; UK11 compared with Mass) there are deletions at the beginning of the ORF of some serotypes (B, C and UK11) with respect to others (A, D and Mass).

One-dimensional peptide mapping of S1 (Cavanagh & Davis, 1987) showed that the S1 glycoproteins of D212 and D1466 were distinguishable. Indeed nucleotide sequencing has shown that these two strains, while they are of the same (D) serotype, have only about 50% nucleotide and amino acid sequence homology in the S1 part of the S gene (Niesters, 1987). However sequencing of D207 (Niesters, 1987) and UK2 (Binns *et al.*, 1986*b*), both in serotype A,

has shown that these strains have greater than 97% nucleotide and amino acid S1 sequence homology with D212 and only about 50% homology with D1466 (Niesters, 1987). It might be expected that D212 and serotype A strains would have very similar M genes. Indeed we detected only two nucleotide differences between that part of the M gene that we sequenced (Table 2). However the M gene of D1466 was also very similar to that of the A serotype, differing by only three bases (1·5%) which contrasts with the approximately 50% difference of the S1 gene. Perhaps the simplest, but not the only, interpretation of these results is that D212 and D1466 have diverged from a common ancestor with subsequent extensive mutation in the S1 gene but essentially conservation of the M gene. Such variation of S1 might be expected since experiments (Mockett *et al.*, 1984; Cavanagh *et al.*, 1986*b*) and comparisons of deduced amino acid sequences (Binns *et al.*, 1986*b*; Niesters, 1987) have indicated that S1 is the major inducer of neutralizing antibody. D274, which is serologically related to serotypes A and B, has only about 50% amino acid and nucleotide S1 sequence homology with serotype A strains but has greater than 98% homology with D1466 (Niesters, 1987). Of the bases sequenced, the M gene of D274 differed at only one position from that of the serotype A strains and at only two positions from D1466 (Fig. 4). It may be, therefore, that all the strains mentioned in this paragraph have diverged from a common ancestor, the M gene undergoing minimal variation, the S1 gene substantial variation, but with D1466 and D212 retaining a common neutralizing antibody-inducing determinant, and likewise for D274 and the A serotype strains. However, the M gene sequence is not highly conserved among IBV strains in general; on the contrary the M gene exhibits extensive variation (Table 2). It would appear unlikely, therefore, that two strains, e.g. D212 and D1466, could diverge from a common ancestor by random mutations such that the S1 part of the S gene differed at over 40% of bases while M remained virtually unchanged. An alternative interpretation of these findings which should be considered is that recombination has occurred, one recombinant having the M gene of one parent and the S gene, or the S1 part, from another. MHV can undergo recombination at high frequency *in vitro* (Keck *et al.*, 1987). Since live IBV vaccines are used extensively world-wide, antigenic variants abound and the virus can persist in chickens for weeks or months after primary infection of the respiratory tract (Cook, 1968; Alexander & Gough, 1978) there is therefore ample opportunity for recombination to occur *in vivo*. Further sequencing of M and S genes will help to clarify the evolution of IBV in general and the importance of recombination in particular.

We wish to thank the following people: Karen Coleman for technical assistance; Dr John McCauley for oligonucleotide JMM and, with Dr Sue Newton, for advice on RNA sequencing; Li Deshan for RNA from some IBV strains; Dr Jane Cook for virus isolates; Drs Matthew Binns, Mike Boursnell, Fiona Tomley and David Brown for advice at various times and Dr Bert Niesters for results in advance of publication.

REFERENCES

ALEXANDER, D. J. & GOUGH, R. E. (1978). A long-term study of the pathogenesis of infection of fowls with three strains of avian infectious bronchitis virus. *Research in Veterinary Science* **24**, 228–233.

ARMSTRONG, J., NIEMANN, H., SMEEKENS, S., ROTTIER, P. & WARREN, G. (1984). Sequence and topology of a model intracellular membrane protein, El glycoprotein, from a coronavirus. *Nature, London* **308**, 751–752.

BINNS, M. M., BOURSNELL, M. E. G., TOMLEY, F. M. & BROWN, T. D. K. (1986*a*). Nucleotide sequence encoding the membrane protein of the IBV strain 6/82. *Nucleic Acids Research* **14**, 5558.

BINNS, M. M., BOURSNELL, M. E. G., TOMLEY, F. M. & BROWN, T. D. K. (1986*b*). Comparison of the spike precursor sequences of coronavirus IBV strains M41 and 6/82 with that of IBV Beaudette. *Journal of General Virology* **67**, 2825–2831.

BOURSNELL, M. E. G., BROWN, T. D. K. & BINNS, M. M. (1984). Sequence of the membrane protein gene from avian coronavirus IBV. *Virus Research* **1**, 303–313.

BOURSNELL, M. E. G., BINNS, M. M. & BROWN, T. D. K. (1985). Sequencing of coronavirus IBV genomic RNA: three open reading frames in the 5′ 'unique' region of mRNA D. *Journal of General Virology* **66**, 2253–2258.

BROWN, T. D. K. & BOURSNELL, M. E. G. (1984). Avian infectious bronchitis virus genomic RNA contains sequence homologies at the intergenic boundaries. *Virus Research* **1**, 15–24.

BROWN, T. D. K., BOURSNELL, M. E. G. & BINNS, M. M. (1984). A leader sequence is present on mRNA A of avian infectious bronchitis virus. *Journal of General Virology* **65**, 1437–1442.

CAVANAGH, D. (1983*a*). Coronavirus IBV glycopolypeptides: size of their polypeptide moieties and nature of their oligosaccharides. *Journal of General Virology* **64**, 1187–1191.

CAVANAGH, D. (1983*b*). Coronavirus IBV: further evidence that the surface projections are associated with two glycopolypeptides. *Journal of General Virology* **64**, 1787–1791.

CAVANAGH, D. & DAVIS, P. J. (1987). Coronavirus IBV: relationships among recent European isolates studied by limited proteolysis of the virion glycopolypeptides. *Avian Pathology* **16**, 1–13.

CAVANAGH, D., DARBYSHIRE, J. H., DAVIS, P. & PETERS, R. W. (1984). Induction of humoral neutralizing and haemagglutination-inhibiting antibody by the spike protein of avian infectious bronchitis virus. *Avian Pathology* **13**, 573–583.

CAVANAGH, D., DAVIS, P. J. & PAPPIN, D. J. C. (1986*a*). Coronavirus IBV glycopolypeptides: locational studies using proteases and saponin, a membrane permeabilizer. *Virus Research* **4**, 145–156.

CAVANAGH, D., DAVIS, P. J., DARBYSHIRE, J. H. & PETERS, R. W. (1986*b*). Coronavirus IBV: virus retaining spike glycopolypeptide S2 but not S1 is unable to induce virus-neutralizing or haemagglutination-inhibiting antibody, or induce chicken tracheal protection. *Journal of General Virology* **67**, 1435–1442.

COLLINS, A. R., KNOBLER, R. L., POWELL, H. & BUCHMEIER, M. J. (1982). Monoclonal antibodies to murine hepatitis virus (strain JHM) define the viral glycoprotein responsible for attachment and cell–cell fusion. *Virology* **119**, 358–371.

COOK, J. K. A. (1968). Duration of experimental infectious bronchitis in chickens. *Research in Veterinary Science* **9**, 506–514.

COOK, J. K. A. (1984). The classification of new serotypes of infectious bronchitis virus isolated from poultry flocks in Britain between 1981 and 1983. *Avian Pathology* **13**, 733–741.

COOK, J. K. A. & HUGGINS, M. B. (1986). Newly isolated serotypes of infectious bronchitis virus: their role in disease. *Avian Pathology* **15**, 129–138.

DAVELAAR, F. G., KOUWENHOVEN, B. & BURGER, A. G. (1983). Experience with vaccination against infectious bronchitis in broilers and significance of the vaccination against variant infectious bronchitis viruses in breeders and layers in The Netherlands. *Clínica Veterinaria* **106**, 7–11.

HASONY, H. J. & MACNAUGHTON, M. R. (1981). Antigenicity of mouse hepatitis virus strain 3 subcomponents in C57 strain mice. *Archives of Virology* **69**, 33–41.

KECK, J. G., STOHLMAN, S. A., SOE, L. H., MAKINO, S. & LAI, M. M. C. (1987). Multiple recombination sites at the 5′-end of murine coronavirus RNA. *Virology* **156**, 331–341.

KUSTERS, J. G., NIESTERS, H. G. M., BLEUMINK-PLUYM, N. M. C., DAVELAAR, F. G., HORZINEK, M. C. & VAN DER ZEIJST, B. A. M. (1987). Molecular epidemiology of infectious bronchitis virus in The Netherlands. *Journal of General Virology* **68**, 343–352.

LAI, M. M. C., PATTON, C. D., BARIC, R. S. & STOHLMAN, S. A. (1983). Presence of leader sequences in the mRNA of mouse hepatitis virus. *Journal of Virology* **46**, 1027–1033.

LAPPS, W., HOGUE, B. G. & BRIAN, D. A. (1987). Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. *Virology* **157**, 47–57.

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory.

MOCKETT, A. P. A., CAVANAGH, D. & BROWN, T. D. K. (1984). Monoclonal antibodies to the S1 spike and membrane proteins of avian infectious bronchitis coronavirus strain Massachusetts M41. *Journal of General Virology* **65**, 2281–2286.

NIESTERS, H. G. M. (1987). *Molecular epidemiology of infectious bronchitis virus*. Ph.D. thesis, University of Utrecht.

PFLEIDERER, M., SKINNER, M. A. & SIDDELL, S. G. (1986). Coronavirus MHV-JHM: nucleotide sequence of the mRNA that encodes the membrane protein. *Nucleic Acids Research* **14**, 6338.

ROTTIER, P. J. M., WELLING, G. W., WELLING-WESTER, S., NIESTERS, H. G. M., LENSTRA, J. A. & VAN DER ZEIJST, B. A. M. (1986). Predicted membrane topology of the coronavirus protein E1. *Biochemistry* **25**, 1335–1339.

WOODS, R. D. & WESLEY, R. D. (1987). Complement dependent virus neutralization of TGEV by anti-matrix monoclonal antibodies. In *Advances in Experimental Medicine and Biology*. Edited by M. M. C. Lai & S. A. Stohlman. London: Plenum Press (in press).

(*Received 10 July 1987*)