Virus Research, 25 (1992) 213–222

VIRUS 00820

Comparative analyses of the nucleocapsid genes of several strains of infectious bronchitis virus and other coronaviruses

Anna K. Williams, Li Wang, Loyd W. Sneed and Ellen W. Collisson

Department of Veterinary Pathobiology, College of Veterinary Medicine, Texas A&M University, College Station, TX 77843, USA

(Received 8 April 1992; revision received and accepted 4 June 1992)

Summary

The natural sequence variations of the nucleocapsid genes of the Gray, Arkansas99 (Ark99), and Holland52 (Holl52) strains of infectious bronchitis virus (IBV) were determined. These were compared with previously published sequencing data of other IBV strains, as well as other coronaviruses, in order to correlate the serological and evolutionary relationship of coronaviruses. IBV nucleotide sequence alignment shows that overall the sequences are highly conserved, with homologies from 91.1 to 96.5%. However, there are also two regions (730 to 800 and 1138 to 1166) that appear to be even more highly conserved. Overall, the nucleocapsid protein is highly variable both in size and composition between coronavirus major antigenic groups but is conserved within these groups. A phylogenetic tree of the nucleocapsid protein of various coronaviruses indicates that the coronaviruses fall into distinct groups that correspond to the three major antigenic groups; however, a phylogenetic tree of the IBV nucleocapsid shows that this does not hold true for the type specific antigenic groups of IBV.

Nucleocapsid gene; Infectious bronchitis; Coronavirus

Avian infectious bronchitis virus (IBV) is a highly contagious pathogen of chickens. IBV infection is characterized by tracheal rales, coughing and sneezing, and can also cause lesions in the reproductive tract, kidneys, and other organs

Correspondence to: E.W. Collisson, Department of Veterinary Pathobiology, College of Veterinary Medicine, Texas A&M University, College Station, TX 77843, USA.

(Darbyshire et al., 1979). IBV was the first coronavirus to be described (Schalk and Hawn, 1931), and many strains varying widely in virulence, serotype and pathotype have been identified (Cumming, 1963; Darbyshire et al., 1979). For example, Beaudette is a highly attenuated strain of IBV, Holl52 is a vaccine strain, Gray and a Japanese strain (KB8523) are both known to be nephropathogenic but have also been associated with respiratory disease, and Ark99 and Mass41 are generally thought to result in severe respiratory disease in the absence of gross kidney lesions (Beaudette and Hudson, 1937; Cumming, 1963; Darbyshire et al., 1979; Johnson and Marquardt, 1986; Jungherr et al., 1956; Sutou et al., 1988; Winterfield and Hitchener, 1962).

IBV is the prototype of the family *Coronaviridae* and was the first coronavirus to be completely sequenced. The genome is 27.6kb in length and consists of single-stranded RNA of positive polarity with a polyA tail (Boursnell et al., 1987). The infectious virion has at least three structural proteins (Siddell et al., 1983; Stern et al., 1982); the glycosylated spike protein, which is responsible for the corona-like appearance of the virions, the membrane protein, and the nucleocapsid protein which is associated with the genome and is encoded by a gene at the 3′ end of the genome.

Coronaviruses have been shown to fall into three antigenic groups. Human coronavirus (HCV) −229E, transmissable gastroenteritis virus (TGEV), canine coronavirus (CCV), feline enteric coronavirus (FECV), and feline infectious peritonitis virus (FIPV) belong to group I, HCV-OC43, murine hepatitis virus (MHV), turkey coronavirus (TCV) and bovine coronavirus (BCV) to group II, and IBV alone to group III (Dea et al., 1990). Although serotyping is generally based on the spike protein, the nucleocapsid and membrane proteins are also highly immunogenic. The nucleocapsid protein is produced in especially large quantities during replication (Sneed et al., 1989) and has also been shown to be recognized by T-cell hybridomas in association with the MHC II (Boots et al., 1991). As such, the nucleocapsid protein may be a useful component of a synthetic vaccine. The nucleocapsid protein of coronaviruses is closely associated with the RNA genome, has been shown to be involved in transcription of MHV, and may be involved in translation (Compton et al., 1987).

In this study, we have analyzed the overall relatedness of the nucleotide and amino acid sequences of the nucleocapsid genes of the Gray, Ark99 and Holl52 strains of IBV, and the published sequences for Mass41, Beaudette and a Japanese strain, KB8523 (Boursnell et al., 1985; Sutou et al., 1988). In order to understand more about the serological and evolutionary relationship of coronaviruses and the physical and biochemical properties of the nucleocapsid protein, we have also compared the nucleocapsid protein of IBV with a number of coronaviruses from the other antigenic groups.

The viral isolates used in this study, previously described by Sneed et al. (1989), were propagated at 37°C for 36–48 h following allantoic sac inoculation into 11-day-old specific pathogen-free embryonating chicken eggs. The eggs were then refrigerated at 4°C for 5 h, and allantoic fluid collected and clarified by centrifugation at 10K rpm for 20 min. Virus was precipitated overnight with 8% polyethy-

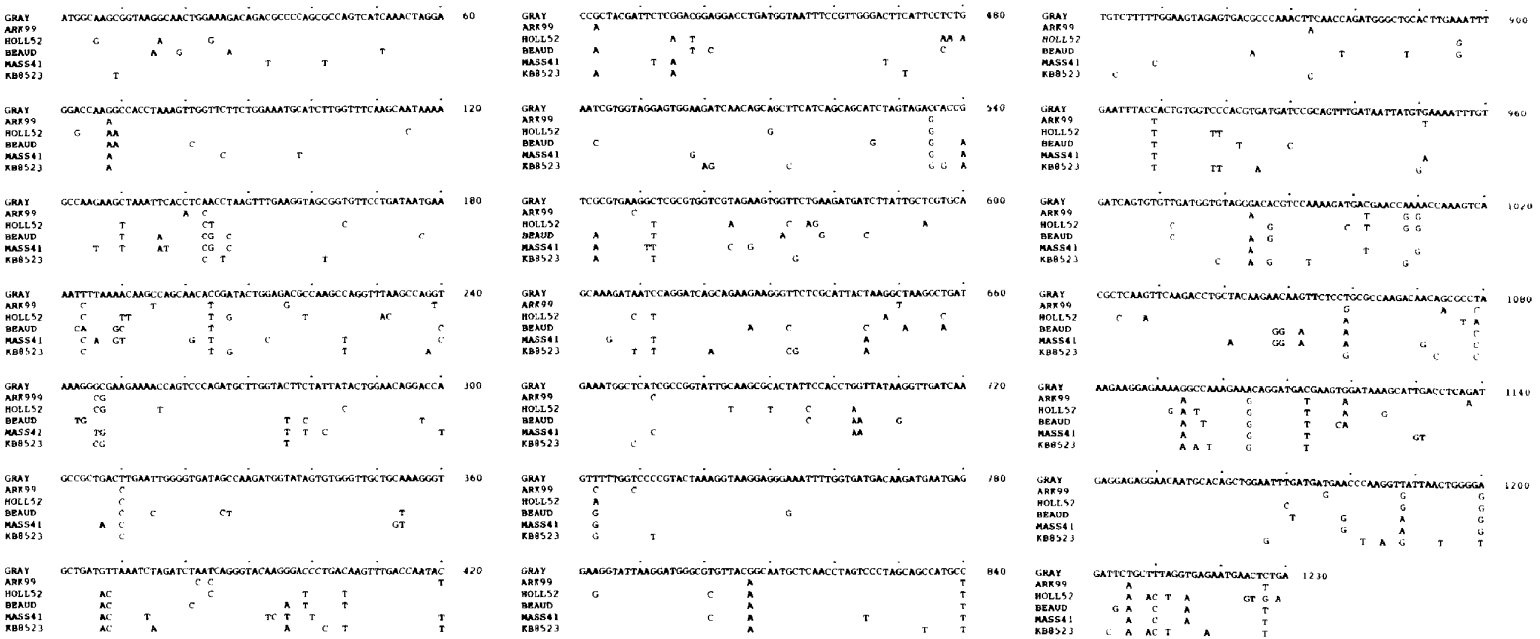

Fig. 1. Nucleotide sequence comparison of the nucleocapsid genes of six strains of IBV. Spaces indicate positions for which the nucleotides are identical to those of the Gray strain. Nucleotides are numbered from the first base of the nucleocapsid gene to the termination codon.

lene-glycol, 2.33% NaCl, and concentrated by centrifugation at 10K rpm for 20 min. Virus pellets were reconstituted in 1/50 vol 20 mM Tris, 1 mM EDTA, 100 mM NaCl, pH 7.4 and banded on a 30–50% glycerol/potassium tartrate gradient for 2 h at 28K rpm. After concentrating by ultracentrifugation at 28K rpm for 1 h, virus pellets were resuspended in 0.1 M NaOAc, pH 5.2, and the virions were disrupted with 1 mg/ml proteinase K and 1% SDS at 37°C for 1 h. The RNA was extracted first with phenol and then with chloroform/isoamyl alcohol followed by ethanol precipitation (Wang et al., 1988). Presence of IBV genomic RNA was confirmed by electrophoresis of a sample of the RNA on a 1% agarose gel and the remaining RNA was reprecipitated with ethanol.

Purified RNA was pelleted and dried in a Savant vacuum drier. RNA was resuspended in 7 μl deionised H_2O (dH_2O) and denatured in 6 mM methyl mercuric hydroxide for 10 min at room temperature. First strand cDNA synthesis was carried out using 5 units of AMV reverse transcriptase at 42°C for 2 h in the presence of 14 mM 2-mercaptoethanol (2ME), 50 mM Tris pH 8.3, 50 mM KCl, 8 mM $MgCl_2$, 0.8 mM dNTP, 20 μg/ml of 3′ end primer (5′GGATC-CGCTCTAACT CTATACTAGCCTAT 3′) and 20 units of RNasin (Promega, Madison, WI). Following phenol extraction and ethanol precipitation of the RNA/cDNA hybrid, the RNA was eliminated by exposure to highly alkaline conditions at 37°C for 3 h. The cDNA was then neutralized, concentrated by ethanol precipitation, and amplified using the polymerase chain reaction (PCR). PCR was performed with the same 3′ end primer as above and a primer from the upstream intergenic region (5′GAATTCCCGCGTGTACCTCTCTAGTA 3′) in the presence of 4 mM dNTP, 1 × Taq buffer (1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl, pH 8) and 0.5 μl *Taq* polymerase (Perkin Elmer, Norwalk, CT). The PCR was performed with a primary denaturation step at 95° for 3 min, annealing at 50°C for 30 s, extension at 72°C for 1 min, and subsequent denaturation steps at 95°C for 1 min. A total of 30 cycles were used with a final extension step of 3 min at 72°C. Resulting PCR products were cloned into the pCR1000 vector, utilizing the single T overhang at the insertion site (Invitrogen, San Diego, CA) and sequenced using Sequenase (USB, Cleveland, OH).

In order to analyze the natural variation among IBV nucleocapsid proteins and genes, sequencing data of the nucleocapsid genes of Ark99, Gray and Holl52 were

TABLE 1

Percent homologies of the nucleic acid and amino acid sequences of six strains of IBV

Amino acids	Nucleotides					
	Gray	Ark	Holl52	Beau	Mass41	KB8523
Gray		96.5	92.8	92.6	93.1	93.6
Ark	98.3		93.3	92.6	93.6	94.1
Holl52	96.6	96.3		91.1	90.9	91.8
Beau	95.4	95.4	94.4		93.2	91.2
Mass41	95.9	95.9	94.6	95.4		92.0
KB8523	97.1	97.1	94.9	94.9	95.1	

obtained and compared to other coronavirus sequences. Of the IBV strains analyzed, Beaudette, Mass41, Holl52 and KB8523 all belong to the same serotype, whereas Gray and Ark99 each belong to different serotypes. At least two clones of each strain were obtained and these were sequenced in both directions. Resulting sequences were analyzed using the University of Wisconsin Genetics Computer Group programs. Nucleotide sequences were processed using the Seqed program and the corresponding nucleocapsid protein amino acid sequences were derived by the Translate program. Comparisons of nucleotide and amino acid sequences were made using Gap and Bestfit. Sequences were aligned using Pileup, with the output being displayed either as a sequence alignment using the program Publish, or as a phylogenetic tree using the Macintosh program PAUP version 3.0.

The sequencing data in Fig. 1 include the entire nucleocapsid genes of Gray, Ark99 and Holl52 aligned with the published sequences of Beaudette, Mass41 and KB8523. This alignment indicates that overall the sequences are highly conserved with some regions showing no variation at all. Table 1 shows that the percent similarity of the nucleotide sequences encoding the nucleocapsid protein ranges from 90.9 between Mass41 and Holl52, to 96.5% between Ark99 and Gray.

The complete open reading frame (ORF) for the nucleocapsid gene of Gray, Ark99 and Holl52, as well as Beaudette, Mass41 and KB8523, consists of 1227 bases, contains Kozak's consensus sequence at the AUG start codon (Kozac, 1986) and codes for a basic protein of 409 amino acids. Unlike many of the other coronaviruses, the IBV nucleocapsid gene does not appear to have a large internal ORF. The alignment of the amino acid sequences shows that there is a large region of absolute identity between residues 238 and 293 (Fig. 2) which corresponds to a highly basic region. It also appears that the protein is more conserved in the central region than at the ends. Two-way comparisons among the nucleocapsid proteins of these 6 strains of IBV indicate that the similarities range from 94.4 to 98.3%, with Beaudette and Holl52 having the least similarity, and Gray and Ark99 the most (Table 1).

We have characterized the nucleocapsid proteins of various coronaviruses in order to locate highly conserved or highly variable regions (Cruciere and Lapporte, 1988; Kamahora et al., 1989; Kapke and Brian, 1986; Lapps et al., 1987; Parker and Masters, 1990; Siddell, 1990; Vennema et al., 1991; Verbeek and Tijssen, 1991). A comparison of the amino acid sequences of these nucleocapsid proteins indicates that they are highly conserved within their antigenic groups. The nucleocapsid protein varies in size as well as composition from one coronavirus group to another. Group I viral genomes have the smallest nucleocapsid protein with 378 to 389 residues and group II genomes have the largest with 449 to 455 residues.

The phylogenetic tree in Fig. 3 shows the relationship of these viruses to each other based on their nucleocapsid genes. Overall sequence similarities classify IBV with the other coronaviruses but the degree of divergence indicates that IBV clearly represents a distinct group. Serological studies have shown that coronaviruses fall into three antigenic groups (Dea et al., 1990). These antigenic groups are defined by cross-reactivity among the structural proteins. Corresponding coronavirus groups can also be distinguished in the phylogenetic tree in Fig. 3,

```
GRAY     MASGKATGKTDAPAPVIKLGGPRPPKVGSSGNASWFQAIKAKKLNSPQPKFEGSGVPDNE   60
ARK99                          K                        HP
HOLL52        T                K                  L     P
BEAUD          A               K                        T P
MASS41                         K           V        N   I P
KB8523    P                    K                        PL

GRAY     NFKTSQQHGYWRRQARFKPGKGRRKPVPDAWYFYYTGTGPAADLNWGDSQDGIVWVAAKG   120
ARK99     L                 V  G
HOLL52    L L             LQ    G  S
BEAUD     I P                   G                        T
MASS41    L S       C           G                    N              G
KB8523    L                  S  G

GRAY     ADVKSRSNQGTRDPDKFDQYPLRFSDGGPDGNFRWDFIPLNRGRSGRSTAASSAASSRPP   180
ARK99           H                                                    A
HOLL52    T                                          I               A
BEAUD     T                                                    A  A
MASS41    T F         S                                            A
KB8523    T                                                K       A

GRAY     SREGSRGRRSGSEDDLIARAAKIIQDQQKKGSRITKAKADEMAHRRYCKRTIPPGYKVDQ   240
ARK99     D                                    V
HOLL52    D         A                                        S
BEAUD               D G                                      N R
MASS41     V                    R                            N
KB8523              A                    A

GRAY     VFGPRTKGKEGNFGDDKMNEEGIKDGRVTAMLNLVPSSHACLFGSRVTPKLQPDGLHLKF   300
ARK99
HOLL52                                                           R
BEAUD                                                       L    R
MASS41
KB8523

GRAY     EFTTVVPRDDPQFDNYVKICDQCVDGVGTRPKDDEPKPKSRSSSRPATRTSSPAPRQQRL   360
ARK99                                    R                       P
HOLL52         S                         R    PN                 Q
BEAUD           C                                      GN         P
MASS41                                   R              GN         P
KB8523         S N        G         L                          P P

GRAY     KKEKRPKKQDDEVDKALTSDEERNNAQLEFDDEPKVINWGDSALGENEL*   410
ARK99        K              N           E
HOLL52       KS
BEAUD        KL       A                 Y          A
MASS41       K            C
KB8523       KT
```

Fig. 2. Amino acid comparison of the nucleocapsid proteins of six strains of IBV. Spaces indicate positions for which the amino acids are identical to those of the Gray strain. Conserved residues are underlined.

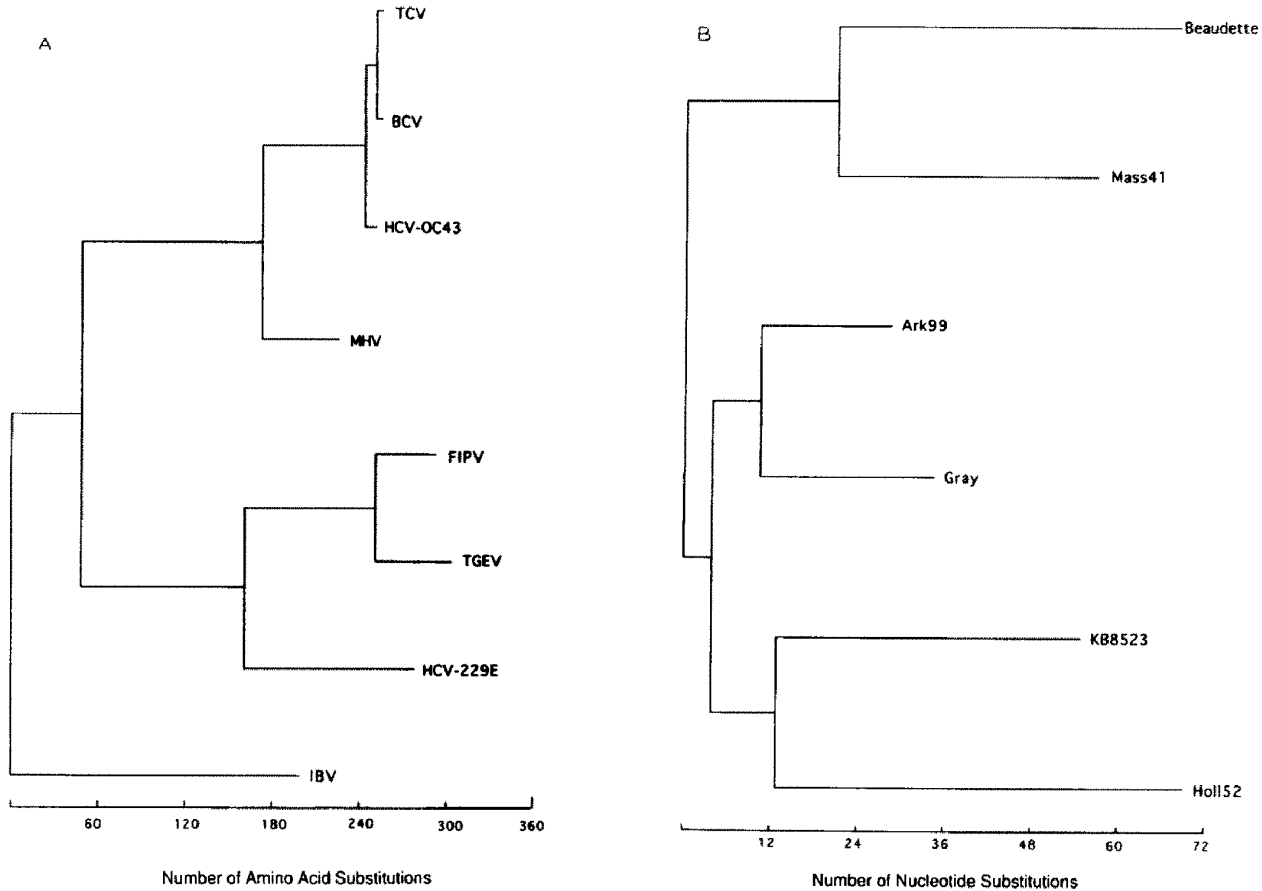

Fig. 3. Phylogenetic trees of the nucleocapsid protein of (A) 8 coronavirus strains and (B) 6 IBV strains. Of the IBV strains, Beaudette, Mass41, Holl52 and KB8523 are in the same serogroup and Ark99 and Gray are each in distinct serogroups.

suggesting a co-evolution of the nucleocapsid protein with the other components that define the antigenic groups. BCV, TCV and HCV-OC43 all appear to be genetically closely related when looking at the nucleocapsid protein and clearly fall into the same antigenic group as MHV, that is group II; however, MHV appears to have diverged away from the other members of group II. Within group I there appears to be a wide evolutionary distance separating HCV-229E from FIP and TGEV.

The nucleocapsid protein of IBV is in a major antigenic group of its own, and considering the amino acid sequence probably diverged earlier from the members of group II and III. According to the phylogenetic tree of IBV, it appears that Gray and Ark99 have diverged further away from those strains belonging to the Mass serotype represented by Beaudette, Holl52, Mass41 and KB8523. This suggests that Gray and Ark99 may have diverged more recently from a common lineage. This correlates with the observation that the Gray and Ark99 strains have both appeared more recently in the USA, and they are more closely related to each other than to the other strains of IBV. However, the Holl52 and Beaudette strains have the lowest similarity scores (Table 1) even though they are in the same serotype, suggesting that recombination may be occurring within the IBV groups. Nucleotide sequencing of the spike gene has also indicated that recombination is occurring between serologically distinct IBV serotypes (personal observations, E.W. Collisson and L. Wang).

Although the coronavirus spike protein is highly variable (Cavanagh, 1991), it is often used as the basis for antigenic grouping by serological techniques (Dea et al., 1990). The close correlation between antigenic grouping from serotyping and from nucleocapsid sequences implies that the spike and nucleocapsid proteins are evolving parallel to each other within distinct major antigenic groups. Thus, classification of major coronavirus groups may actually be determined on the basis of either spike or nucleocapsid protein and presumably, this grouping could also apply to the other coronavirus genes. If recombination is naturally occurring as is expected (Cavanagh and Davis, 1988), it would seem to be occurring within these major antigenic groups, such as is seen with poliovirus (Agut et al., 1987). If this is the case, it would appear that the viability of the progeny of coronaviruses depends on specific interactions of more than one protein working in concert to maintain structural or replicative integrity of the virus.

The nucleocapsid protein of coronaviruses is a highly immunogenic protein, may be involved in stimulating cytotoxic T-lymphocytes, and is produced in large amounts relative to the other structural proteins in the cells (Boots et al., 1991; Sneed et al., 1989). Therefore, any highly conserved regions may be worth considering as potential vaccine components, especially if such determinants can be recognized by either B- or T-cells. Although the nucleocapsid protein is variable among the different coronaviruses, it is generally highly conserved among strains within a species, such as IBV in chickens. It also would seem that while antigenic drift is occurring in the major antigenic groups of coronaviruses, recombination is causing an antigenic shift in the individual IBV strains.

Acknowledgements

We thank Peggy M. Billingsley for her excellent technical assistance. This work was supported by the Southeastern Poultry and Egg Association, nos. 29 and 460; Formula Animal Health (section 1433), no. TEXO-6824 and TATRP no. 999902-063.

References

Agut, H., Kean, K.M., Bellocq, C., Fichot, O. and Girard, M. (1987) Intratypic recombination of polioviruses: evidence for multiple crossing-over sites on the viral genome. J. Virol. 61, 1722–1725.

Beaudette, F.R. and Hudson, C.B. (1937) Cultivation of the virus of infectious bronchitis. JAVMA 90, 51–60.

Boots, A.H.M., Van Lierop, M.J., Kusters, J.G., Van Kooten, P.J.S., Van Der Zeijst, P.J.S. and Hensen, E.J. (1991) MHC class II-restricted T-cell hybridomas recognising the nucleocapsid protein of avian coronavirus IBV. Immunol. 72, 10–14.

Boursnell, M.E.G., Binns, M.M., Foulds, I.J. and Brown, T.D.K. (1985) Sequences of the nucleocapsid genes from two strains of avian infectious bronchitis virus. J. Gen. Virol. 66, 573–580.

Boursnell, M.E.G., Brown, T.D.K., Foulds, I.J., Green, P.F., Tomley, F.M. and Binns, M.M. (1987) Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. J. Gen. Virol. 68, 57–77.

Cavanagh, D. (1991) Sequencing approach to IBV antigenic variation and epizootology. In E.F. Kaleta and U. Heffels-Redman (Eds), II. International symposium on infectious bronchitis, pp. 147–153. Rauischholzhausen, Germany.

Cavanagh, D. and Davis, P.J. (1988) Evolution of avian coronavirus IBV: sequence of the avian coronavirus matrix glycoprotein gene and intergenic region of several serotypes. J. Gen. Virol. 69, 621–629.

Compton, S.R., Rogers, D.B., Holmes, K.V., Fertsch, D., Remenick, J. and McGowan, J.J. (1987) In vitro replication of mouse hepatitis virus strain A59. J. Virol. 61, 1814–1820.

Cumming, R.B. (1963) The etiology of uremia of chickens. Aust. Vet. J. 39, 145–147.

Cruciere, C. and Lapporte, J. (1988) Sequence and analysis of bovine enteric coronavirus (F15) genome. 1. Sequence of the gene coding for the nucleocapsid protein; analysis of the predicted protein. Ann. Inst. Pasteur 139, 123–138.

Darbyshire, J.H., Rowell, J.G., Cook, J.K.A. and Peters, R.W. (1979) Taxonomic studies on strains of avian infectious bronchitis virus using neutralization tests in tracheal organ cultures. Arch. Virol. 61, 227–238.

Dea, S., Verbeek, A.J. and Tijssen, P. (1990) Antigenic and genomic relationships among turkey and bovine enteric coronaviruses. J. Virol. 64, 3112–3118.

Johnson, R.B. and Marquardt, W.W. (1986) Strains of infectious bronchitis virus on the Delmarva Peninsula and in Arkansas. Avian Dis. 20, 382–386.

Jungherr, E.L., Chomiak, T.W. and Luginbuhl, R.E. (1956) Immunologic differences in strains of infectious bronchitis virus. In: Proceedings of the 60th annual meeting of United States Livestock Sanitary Association, pp. 203–209.

Kamahora, T., Soe, L.H. and Lai, M.C. (1989) Sequence analysis of nucleocapsid gene and leader RNA of human coronavirus OC43. Virus Res. 12, 1–9.

Kapke, P.A. and Brian, D.A. (1986) Sequence analysis of porcine transmissable gastroenteritis coronavirus nucleocapsid protein gene. Virology 151, 41–49.

Kozac, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44, 283–292.

Lapps, W., Hogue, B.G. and Brian, D.A. (1987) Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. Virology 157, 47–57.

Parker, M.M. and Masters, P.S. (1990) Sequence comparison of the N genes of five strains of the coronavirus mouse hepatitis virus suggests a three-domain structure for the nucleocapsid protein. Virology 179, 463–468.

Schalk, A.F. and Hawn, M.C. (1931) An apparently new respiratory disease of baby chicks. JAVMA 78, 418–422.

Siddell, S. (1990) Personal communication.

Siddell, S.G., Wege, H. and Ter Meulen, V. (1983) The biology of coronaviruses. J. Gen. Virol. 64, 761–776.

Sneed, L.W., Butcher, G.D., Wang, L., Parr, R. and Collisson, E.W. (1989) Comparisons of the structural proteins of avian infectious bronchitis virus as determined by Western blot analysis. Viral Immunol. 2(3), 221–227.

Stern, D.F., Burgess, L. and Sefton, B.M. (1982) Structural analysis of virion proteins of the avian coronavirus infectious bronchitis virus. J. Virol. 42, 208–219.

Sutou, S., Sato, S., Okabe, T., Nakai, M. and Sasaki, N. (1988) Cloning and sequencing of genes encoding structural proteins of avian infectious bronchitis virus. Virol. 165, 589–595.

Vennema, H., DeGroot R.J., Harbour, D.A., Horzinek, M.C. and Spaan, W.J.M. (1991) Primary structure of the membrane and nucleocapsid protein genes of feline infectous peritonitis virus and immunogenicity of recombinant vaccinia virus in kittens. Virology 181, 327–335.

Verbeek, A. and Tijssen, P. (1991) Sequence analysis of the turkey enteric coronavirus nucleocapsid and membrane protein genes: a close genomic relationship with bovine coronavirus. J. Gen. Virol. 72, 1659–1666.

Wang, L., Kemp, M.C., Roy, P. and Collisson, E.W. (1988) Tissue tropism and target cells of Bluetongue virus in the chicken embryo. J. Virol. 62, 887–893.

Winterfield, R.W. and Hitchner, S.B. (1962) Etiology of an infectious nephritis-nephrosis syndrome of chickens. Am. J. Vet. Res. 23, 1273–1279.