ANALYSIS OF THE NUCLEOCAPSID PROTEIN OF IBV AND ITS ASSOCIATION WITH A HYPERVARIABLE REGION IN THE 3' END OF THE GENOME

A Dissertation

by

ANNA K. WILLIAMS

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 1993

Major Subject: Veterinary Microbiology

ANALYSIS OF THE NUCLEOCAPSID PROTEIN OF IBV AND ITS ASSOCIATION WITH A HYPERVARIABLE REGION IN THE 3' END OF THE GENOME

A Dissertation

by

ANNA K.WILLIAMS

Approved as to style and content by:

Ellen W. Collisson
(Chair of committee)

Barbara Doughty
(Member)

Van G. Wilson
(Member)

William K. Read
(Head of Department)

Allison R. Ficht
(Member)

May 1993

Major Subject: Veterinary Microbiology

ABSTRACT

Analysis of the Nucleocapsid Protein of IBV and its Association with a Hypervariable Region in the 3' End of the Genome.

(May 1993)

Anna K. Williams, B.S. University of Reading, U. K.

Chair of Advisory Committee: Dr. Ellen Collisson

The natural sequence variations of the nucleocapsid genes and 3' non-coding regions of the Gray, Ark99 and Holl52 strains of IBV were determined and compared to the Mass41, Beau and KB8523 strains. The nucleocapsid protein was highly conserved and also shared conserved amino acids with other coronaviruses. In contrast, the 3' non-coding region contained a U-rich hypervariable region (HVR) immediately downstream of the nucleocapsid gene. This U-rich sequence was absent from 3' non-coding region of the Mass41 strain, and was also highly variable, especially in comparison to the highly conserved 3' non-coding region downstream of this sequence. Computer analyses of the sequences within and adjacent to this HVR showed that the 3' end of the genome was highly conserved downstream of this region, with 94.3 to 97.8% similarity. However, the similarities for the HVR ranged from 53.2% between Holl52 and Ark99, to 92.8% between Beaudette and Gray. The flanking sequences upstream and downstream of the HVR were conserved and also formed mirrored images. The association of the IBV nucleocapsid protein with RNA from portions of the 3' end of the genome was examined by northwestern blot analysis and gel shift assays.

The present studies have shown that the nucleocapsid protein was able to bind with a high affinity to RNA corresponding to the 3' end of the IBV genome. All the positive and negative sense IBV-derived RNAs used in this study were capable of binding to the nucleocapsid protein, and apparently all the positive sense IBV RNAs had similar affinity for the nucleocapsid protein.

ACKNOWLEDGEMENT

This dissertation is dedicated to my parents David and Valerie Williams, my husband Kent, and my daughter Jennifer. I am deeply grateful to my parents and Kent for their support and encouragement during my graduate training. I am also very grateful to Jennifer for being so patient with me for the last 5 months as I finished my research and wrote this dissertation.

I would like to express my appreciation and gratitude to the following people for their help in the successful completion of this dissertation.

Dr Ellen Collisson for her guidance during my training and research.

Dr Van Wilson, Dr Allison Ficht and Dr Barbara Doughty for their help and advice.

Dr Scott Davis for his help with the phylogenetic trees.

Robert Bauder for computer assistance.

Fellow students: Michele Hardy and Michael Dilling for their help and moral support, Scott Bearden and Li Wang for always being able to offer a few helpful tips.

Peggy Billingsley for her help and advice.

Betty Suehs for her help over the last 6 years and also for proof-reading this thesis.

The staff of Biomedical Communications for their help with the figures in this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Avian infectious bronchitis virus (IBV), the prototype of the *Coronaviridae*, causes an acute, highly contagious respiratory disease in chickens. Infection with IBV results in morbidity and sometimes mortality, but although significant economic losses result from the reduction in egg quality and production in laying flocks, the greatest losses result from condemnation of broilers (Cunningham, 1970; King & Cavanagh, 1991). The disease was first described by Schalk and Hawn in 1931, the causative agent was found to be a virus in 1936 (Beach and Schalm) and this virus was first cultivated in embryonating chicken eggs in 1937 (Beaudette and Hudson).

Members of the Coronaviridae family were first classified as a group on the basis of negative stains examined by electron microscopy (Berry *et al.*, 1964; Almeida *et al.*, 1967). Large club-shaped surface projections, also called spikes or peplomers, create the distinct crown-like or corona appearance observed with the electron microscope. The virions are roughly spherical and somewhat pleomorphic, with an estimated diameter of 90-200nm and spikes of 20nm (Davis *et al.*, 1979; Sturman & Holmes, 1983). A bilipid envelope of cell membrane origin surrounds the nucleocapsid of the virion making the virion sensitive to solvents such as chloroform (Otsuki *et al.*, 1979).

The format of this dissertation follows that of *Virology*.

Two glycosylated proteins are integrated into the envelope membrane of IBV: the large multimeric spike protein (S) and the membrane protein (M) (Cavanagh, 1981). This envelope surrounds a helical nucleocapsid consisting of an RNA genome associated with a number of copies of the third major structural protein, the nucleocapsid protein (N) (Schochetman *et al.*, 1977; Oshiro, 1973).

The IBV genome consists of single stranded, polyadenylated RNA (Shochetman *et al.*, 1977), is 27.6kb in length and was the first Coronavirus to be completely sequenced (Boursnell *et al.*, 1987). The genomic RNA is infectious as it can act directly in the cell as a viral mRNA, initiate replication in the absence of viral proteins and result in production of viable virus progeny (Schochetman *et al.*, 1977; Lomniczi, 1977). In addition to the complete genome which functions as a mRNA for putative polymerase activity, infected cells have five subgenomic mRNAs which form a 3' co-terminal nested set; that is, the entire sequence of each species is found in the larger species (Stern *et al.*, 1982; Boursnell *et al.*, 1987). These mRNAs are numbered from 1 to 6 with RNA 1 being the largest or genomic RNA and RNA 6 being the smallest sub-genomic RNA. A common leader sequence of about 60 bases is found at the 5' end of each sub-genomic mRNA, as well as the genomic RNA (Brown *et al.*, 1984).

In most cases, the IBV proteins are produced by translation of the open reading frame (ORF) located at the 5' end of each of the mRNAs (Stern and Kennedy, 1980). RNA 1, the genomic RNA, codes for the polymerase-associated proteins, RNA 2 encodes the spike protein, RNA 4, the membrane protein, and RNA 6, the nucleocapsid protein (Fig. 1-1). However, RNA 3 has 3 ORFs and the third ORF has been found to encode a

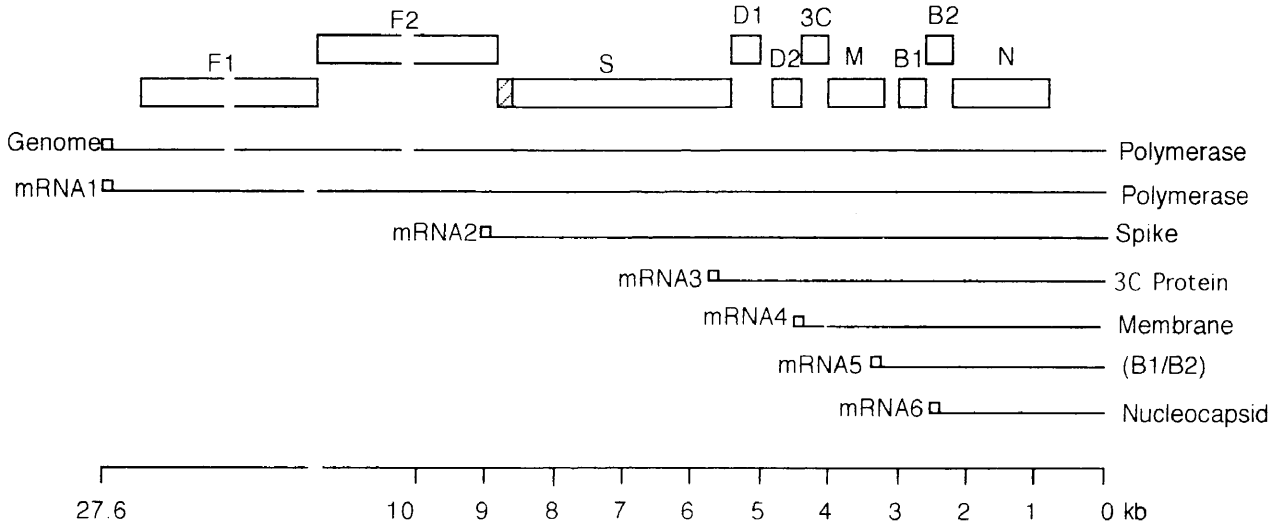


Fig. 1-1 Schematic representation of the open reading frames of IBV mRNAs. The small box on the 5' end of each RNA represents the leader sequence.

membrane-associated protein of 12.4kD that is found in infected cells (Smith *et al.*, 1990). RNA 5 contains 2 ORFs but no proteins corresponding to these ORFs have been found in either the virion or infected cells (Boursnell and Brown, 1984).

Viral replication takes place under the control of a viral RNA-dependent RNA polymerase, which is produced as a large polyprotein by translation of the genomic RNA upon entry of the virus into the cell. This polymerase (or a subunit of the polyprotein) then presumably recognizes the 3' end of the genomic RNA and initiates transcription, resulting in the production of either full-length negative sense RNA, or negative sense subgenomic RNAs, from which mRNAs can be transcribed (Boursnell *et al.*, 1987; Sethna *et al.*, 1989). The leader sequence found at the 5' end of genomic RNA is thought to play a role in transcription. Excess copies of the leader sequence are made from the full length negative RNA and are then able to recognize a sequence upstream of the ORFs and act as a primer for transcription.

The IBV nucleocapsid protein is a highly basic, phosphorylated protein of 409 amino acids and relative molecular weight of approximately 50kD. The nucleocapsid protein of coronaviruses is closely associated with the RNA genome, has been shown to be involved in transcription of murine hepatitis virus (MHV) (Nakanaga *et al.*, 1986) and may also be involved in translation (Compton *et al.*, 1987). A 140kD trimer of the nucleocapsid protein of MHV has been found in non-reduced preparations of purified virions suggesting the presence of disulphide linked multimers (Robbins *et al.*, 1986).

Although coronaviruses are the only positive sense RNA viruses known to have a helical nucleocapsid, several negative sense RNA viruses have helical nucleocapsids that are important in regulating viral transcription (Arnheiter *et al.*, 1985; Banerjee, 1987). The nucleocapsid protein of MHV has been found to be bound to cytosol and membrane-bound small leader RNAs, as well as to transcription complexes. Also, anti-N monoclonal antibodies precipitate both full length and subgenomic mRNAs as well as replicative intermediate (RI) RNAs (Baric *et al.*, 1988a). Baric *et al.* (1988b) have also shown a specific nucleocapsid protein recognition sequence exists in the small leader-containing RNAs, suggesting that a trans-acting leader RNA-N complex is involved with mRNA synthesis.

It is known that several copies of the nucleocapsid protein of IBV are associated with the RNA (Schochetman *et al.*, 1977; Oshiro, 1973) and, as transcription of the negative template is initiated at the 3' end of the genomic RNA, it seems possible that this is another region where the nucleocapsid protein binds and plays a role in transcription.

Coronaviruses have been shown to fall into three antigenic groups. Human coronavirus (HCV)-229E, transmissible 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.*, 1991). Although antigenicity is generally based on the spike protein, the nucleocapsid protein is also highly immunogenic and is produced in large quantities during replication (Cavanagh, 1983; Sneed *et al.*, 1989). It has also been shown that the nucleocapsid protein, in association with the MHC II, is

recognized by T-cell hybridomas and that there is some antigenic variation among nucleocapsid proteins of IBV (Boots *et al.*, 1991). As such, the nucleocapsid protein may be a useful component of a synthetic vaccine.

The present study has focused on the investigation of the conservation of the IBV nucleocapsid gene and its relationship to other coronaviruses, the sequence variations and resulting secondary structures of the IBV 3' non-coding region, and the binding of the nucleocapsid protein to certain sequences in the 3' end of the IBV genome.

CHAPTER II

COMPARATIVE ANALYSIS OF THE NUCLEOCAPSID GENES OF SEVERAL STRAINS OF IBV AND OTHER CORONAVIRUSES

Introduction

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 and kidneys (Darbyshire *et al.*, 1979). IBV was the first coronavirus to be described (Schalk & Hawn, 1931), and many strains varying widely in virulence, serotype and pathotype have been identified (Cumming, 1963; Darbyshire *et al.*, 1979). Beaudette is a highly attenuated strain of IBV, Holland52 (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 (Arkansas 99) and Mass41 are generally thought to result in severe respiratory disease in the absence of gross kidney lesions. (Beaudette & Hudson, 1937; Cumming, 1963; Darbyshire *et al.*, 1979; Johnson & Marquardt, 1986; Jungherr *et al.*, 1956; Sutou *et al.*, 1988; Winterfield & Hitchener, 1962).

IBV is the prototype of the family Coronaviridae and was the first coronavirus to be completely sequenced. The IBV 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 based on serotype. Human coronavirus-229E (HCV-229E), transmissible 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 protein is also highly immunogenic and is produced in large quantities during replication (Sneed *et al.*, 1989; Cavanagh, 1983). It has also been shown that the nucleocapsid protein in association with the MHC II is recognized by T-cell hybridomas and that there is some antigenic variation among nucleocapsid proteins of IBV (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 also 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 determine 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.

Materials and Methods

Viral Preparation

The viral isolates used in this study were propagated at 37°C for 36-48h following allantoic sac inoculation into 11-day old specific pathogen free embryonating chicken eggs (ECE). The eggs were then refrigerated at 4°C for 5h, and allantoic fluid collected and clarified by centrifugation at 16K x g for 20 min. Virus was precipitated overnight with 8% polyethylene-glycol, 2.33% NaCl, and concentrated by centrifugation at 16K x g for 20 min. Virus pellets were reconstituted in 1/50 vol 20mM Tris, 1mM EDTA, 100mM NaCl, pH7.4 and banded on a 30-50% glycerol/potassium tartrate gradient for 2h at 100K x g. After concentrating by ultracentrifugation at 100K x g for 1h, virus pellets were resuspended in 0.1M NaOAc, pH5.2, and the virions were disrupted with 1mg/ml proteinase K and 1% SDS at 37°C for 1h. 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.

Cloning of the Gray, Ark99 and Holl52 Strains

Purified RNA was pelleted and dried in a Savant vacuum drier. RNA was resuspended in 7µl deionized H_2O (dH_2O) and denatured in 6mM methyl mercuric hydroxide for 10 min at room temperature. First strand

cDNA synthesis was carried out using 5 units of AMV reverse transcriptase (BMB, Indianapolis, In) at 42°C for 2h in the presence of 14mM 2-mercaptoethanol (2ME), 50mM Tris pH8.3, 50mM KCl, 8mM $MgCl_2$, 0.8mM dNTP, 20µg/ml of 3' end primer (5' GGATCCGCTCTAACTCTATACTA GCCTAT 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 3h. 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' GAATTCCCGCGTGTACC TCTCTAG TA 3') in the presence of 4mM dNTP, 1x Taq buffer (1.5mM $MgCl_2$, 50mM KCl,10mM Tris-HCl, pH8) and 0.5µl Taq polymerase (Perkin Elmer, Norwalk, Ct.). The PCR was performed with a primary denaturation step at 95°C for 3 min, annealing at 50°C for 30s, 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, Ohio).

Computer Analysis

Resulting sequences were analyzed using the Genbank 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. IBV sequences were aligned and displayed using Pileup and Publish, with Gray set as the reference sequence. A hydrophobicity plot of the nucleocapsid protein of IBV was determined using the program Pepplot. Other coronavirus sequences were also aligned using Pileup, with the output being displayed either as a sequence alignment using Publish, or as a phylogenetic tree using the Macintosh program PAUP, version 3.0.

Results

Comparison of the nucleotide and amino acid sequences of the nucleocapsid protein of various strains of IBV

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 obtained and compared to other coronavirus sequences. At least two clones of each strain were obtained and these were sequenced in both directions. Fig. 2-1 shows the nucleotide sequences of the Gray, Ark99, and Holl52 nucleocapsid genes 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 2-1 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,

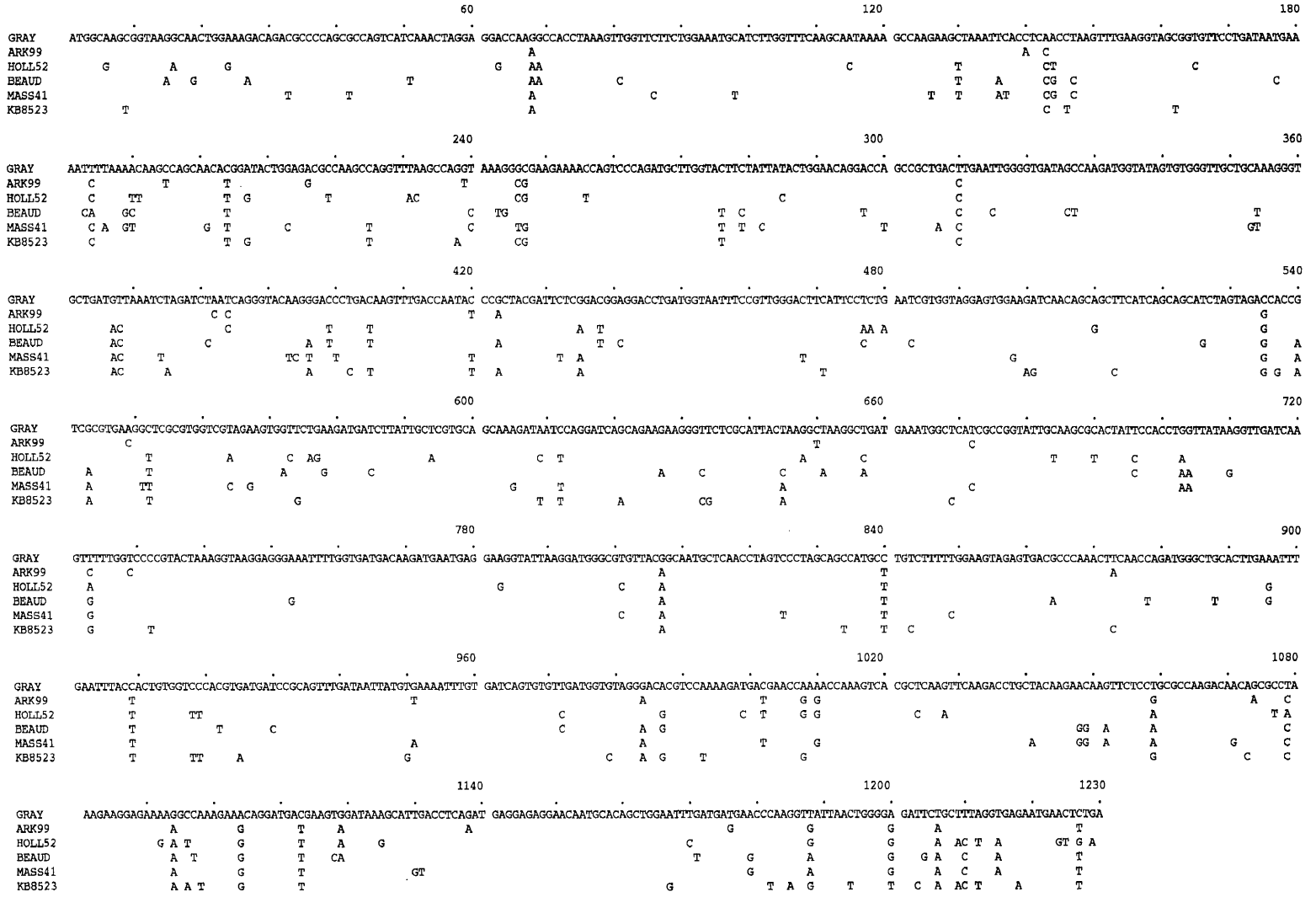


Fig. 2-1 Nucleotide sequence comparison of the nucleocapsid genes of six strains of IBV.

Table 2-1 Percent homologies of the nucleic acid sequences of the nucleocapsid genes of six strains of IBV.

	Gray	Ark99	Holl52	Beaud	Mass41	KB8523
Gray	100	96.5	92.8	92.6	93.1	96.3
Ark99		100	93.3	92.6	93.6	94.1
Holl52			100	91.1	90.9	91.8
Beaud				100	93.2	91.2
Mass41					100	92.0
KB8523						100

Table 2-2 Percent homologies of the amino acid sequences of the nucleocapsid protein of six strains of IBV.

	Gray	Ark99	Holl52	Beaud	Mass41	KB8523
Gray	100	95.1	94.9	94.9	97.1	97.1
Ark99		100	95.4	94.6	95.9	95.9
Holl52			100	94.4	95.4	95.4
Beaud				100	96.3	96.6
Mass41					100	98.3
KB8523						100

consists of 1227 bases, contains Kozac's consensus sequence at the AUG start codon (Kozac, 1986) and codes for a basic protein of 409 amino acids. Two way comparisons among the nucleocapsid proteins of these six 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 2-2). The predicted molecular weights are between 51.5 and 52.4kD, and they have isoelectric points from 10.35 to 10.47. Although the overall nature of the nucleocapsid protein is highly basic, there are also three acidic regions with isoelectric points of 4.46, 3.94, and 3.58 (Fig. 2-2). In addition, it is mainly a hydrophilic protein with a cluster of hydrophobic residues from positions 270 to 284 (Fig. 2-2).

Coronavirus nucleocapsid proteins

Unlike BCV, TGEV, FECV, HCV-OC43 and HCV-229E, the IBV nucleocapsid gene does not appear to have a large internal ORF. The alignment of the amino acid sequences of the six strains of IBV shows that there is a large region of absolute identity between residues 238 and 293 (Fig. 2-3) which corresponds to one of the basic regions. It also appears that the protein is more conserved in the central region than at the ends.

We have characterized the nucleocapsid proteins of various coronaviruses in order to locate highly conserved or highly variable regions. A representative of each of the coronavirus antigenic groups is shown in the alignment in Fig. 2-4 (Cruciere & Lapporte, 1988; Kamahora *et al.*, 1989; Kapke & Brian, 1986; Lapps *et al.*, 1987; Parker & Masters, 1990; Siddell, 1990; Vennema *et al.*, 1991; Verbeek & Tijssen, 1991). The nucleocapsid protein varies in size as well as composition from one coronavirus group to

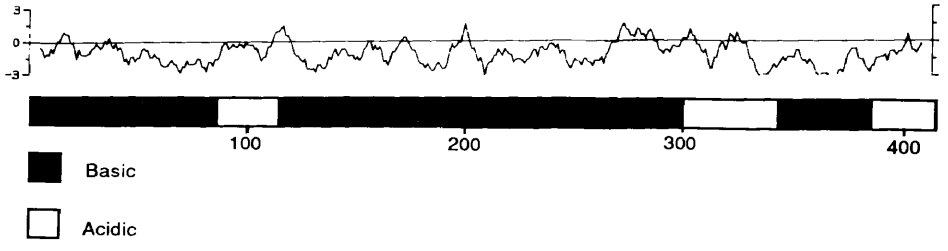


Fig. 2-2 Diagrammatic representation of the hydrophobic regions of the nucleocapsid protein of IBV (upper panel) and the acidic and basic regions of the protein (lower panel).

```
                .         .         .         .         .         .
GRAY     MASGKATGKTDAPAPVIKLGGPRPPKVGSSGNASWFQAIKAKKLNSPQPKFEGSGVPDNE   60
ARK99                      K                       HP
HOLL52        T            K                  L     P
BEAUD          A           K                       T P
MASS41                     K          V         N   I P
KB8523                     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. 2-3 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.

another. Group I viral genomes have the smallest nucleocapsid protein with 378 to 389 residues and group II have the largest with 449 to 455 residues. There are also gaps interspersed throughout all the sequences in the alignment. The amino acid residue numbers on the alignment in Fig. 2-3 do not correspond to the numbers in Fig. 2-4, because the variation among the sequences results in the insertion of gaps in specific sequences.

A plot of the similarity of all 17 protein sequences studied indicates that there are two regions, corresponding to residues 129 to 138 and 147 to 153, with greater than average homology (Fig 2-5). In addition, the sequence alignment shows several completely conserved residues, at positions 91, 110, 129, 130, 133, 134, 135, 136, 146, 151, 152, 153, 180, 200, 300, 320, 329, 330, 331 and 360 (Fig 2-4). There is also variation among the different species within the antigenic subgroups, such as is seen between TGEV and HCV-229E, and with MHV compared with BCV, TCV and HCV-OC43 (Fig. 2-4). This is reflected in the plot in Fig. 2-5 where, for example, there is a very low similarity score between residues 265 to 300 due to a number of residues present only in one member of antigenic group I, that is, HCV-229E. As with the IBV strains alone, there appears to be overall less similarity at the ends of the protein than in the middle.

Phylogeny

The phylogenetic tree in Fig. 2-6 shows the relationship of these viruses to each other based on their nucleocapsid genes and it clearly indicates that the amino acid sequences are highly conserved within the

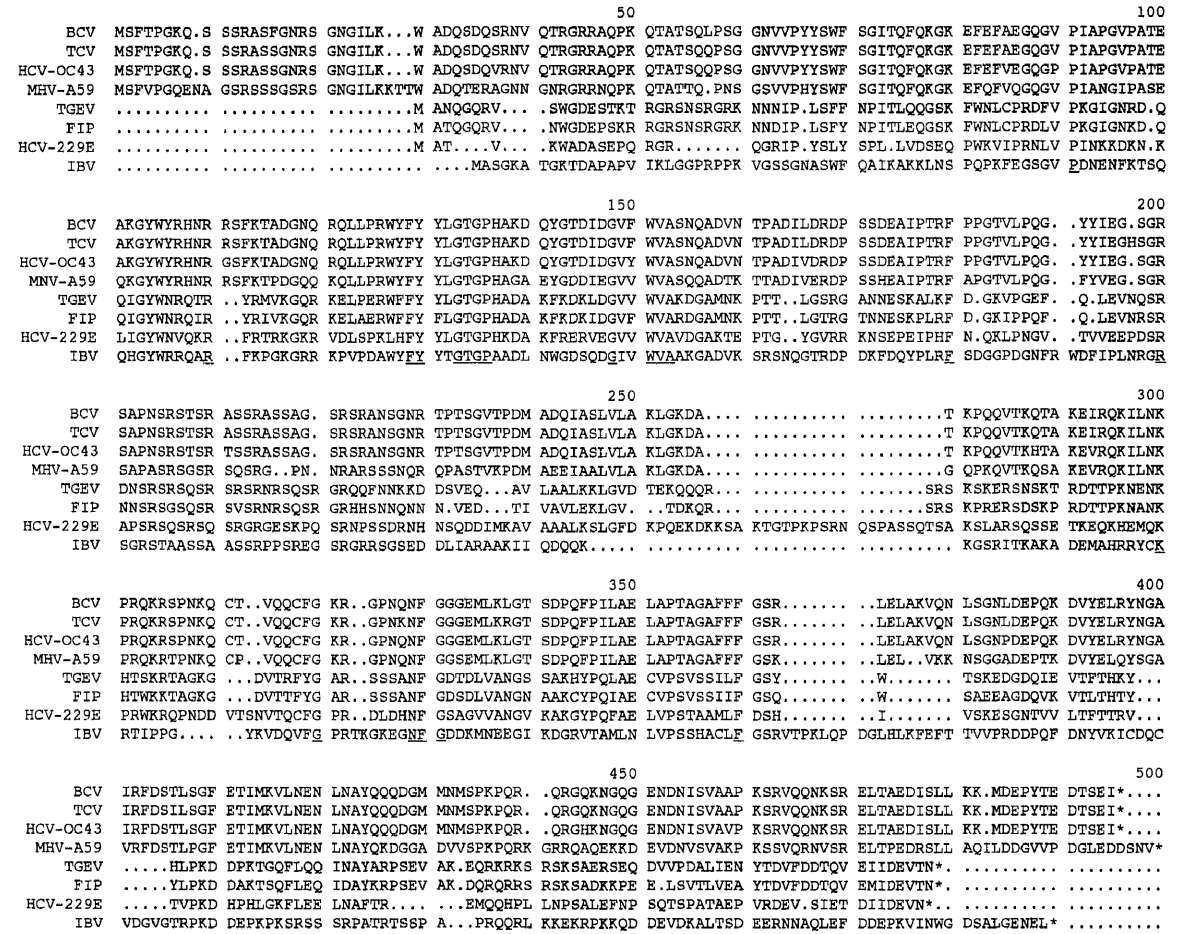


Fig. 2-4 Amino acid comparison of the nucleocapsid proteins of representatives of the major antigenic groups of the coronaviridae.

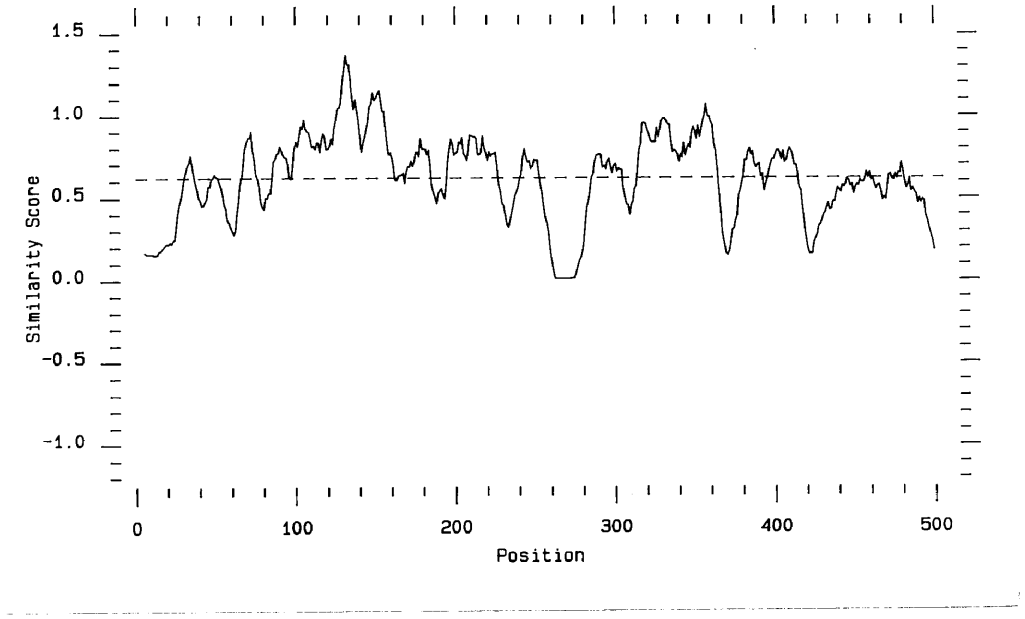


Fig 2-5. Similarity plot of the nucleocapsid protein amino acid sequence from 17 coronaviruses.

antigenic groups. Overall sequence similarities classify IBV strains with the other coronaviruses but the degree of divergence indicates they represent a distinct group from the others. The other coronaviruses fall into two branches, the first has fairly tight homology and includes MHV, BCV, TCV and HCV-OC43, whereas the other appears to have diverged earlier with HCV-229E being separated from TGEV and FIP. According to the phylogenetic tree of IBV (Fig. 2-6) Gray and Ark99 are more closely related to each other than to those strains of the Mass serotype. This suggests that Gray and Ark99 may have diverged more recently from a common lineage.

Discussion

The nucleocapsid protein of coronaviruses is a highly immunogenic protein that may be involved in stimulating cytotoxic T lymphocytes, and is produced in large amounts relative to the other structural proteins in the cells (Sneed *et al.*, 1989; Boots *et al.*, 1991; Cavanagh, 1983). 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. The nucleocapsid protein is variable among the different coronaviruses, but is generally highly conserved among strains within a species, such as IBV in chickens. Although little variation was found among the IBV nucleocapsid proteins, according to the amino acid sequences, there is more variation at the ends of the protein than in the middle. This also appears to be true of the other members of the coronavirus family. 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.

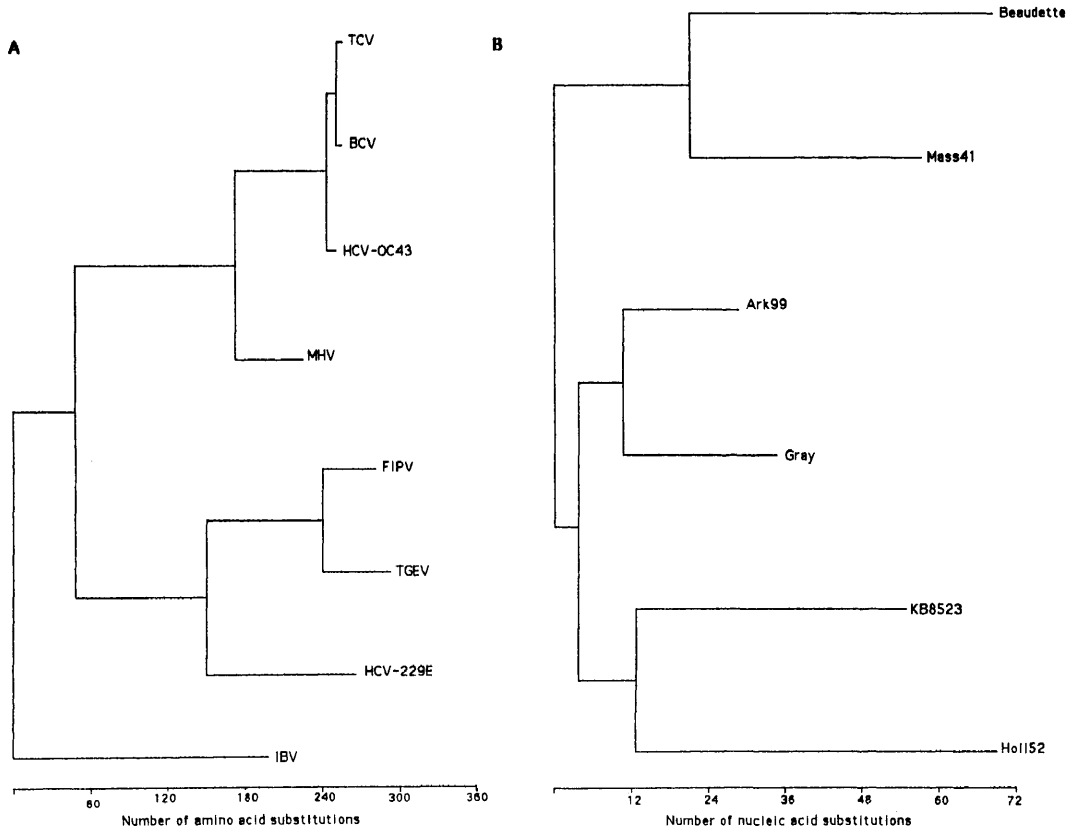


Fig. 2-6 Phylogenetic trees of the nucleocapsid protein of A) 8 coronavirus strains and B) 6 IBV strains. Of the IBV strains, Beaudette, Mass41, Holl52 and KB8523 are in the same serogroup and Ark99 and Gray are each in distinct serogroups.

However, all appear to have originally derived from the Mass serotype with Gray and Ark99 having evolved later. This correlates with the observation that the Gray and Ark99 strains have both appeared more recently in the U.S, and they are more closely related to each other than to the other strains of IBV.

Serological studies have shown that coronaviruses fall into three antigenic groups (Dea *et al.*, 1990). Corresponding coronavirus groups can also be distinguished in the phylogenetic tree in Fig. 2-6. 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 earlier than 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 group of its own, and considering the amino acid sequence probably diverged earlier from the members of group II and III.

Although the spike protein is highly variable (Cavanagh, 1991), it is 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 antigenic groups. Thus, classification of coronaviruses 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 & Davis, 1988), it would seem to be occurring within the 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.

Regions of strong homology among these different nucleocapsid proteins suggest a functional importance in viral replication that even over time and evolution are very sensitive to change. It is interesting that the acidic region nearest the amino terminal of IBV includes two regions which are also very highly conserved among all the coronaviruses. The first region has six out of ten amino acids completely conserved (129 FY(Y/F)(L/T)GTGP(H/A)(A/K) 136), and the two previous residues are at least conserved in their hydrophobic nature; Leu-His in HCV-229E, Trp-His in TGEV and FIP, and Trp-Tyr in all the other sequences. The second region has four out of seven amino acids completely conserved (147 (D/E)G(V/I)(F/Y/V)WVA 153). However, the functional significance of these residues is not known.

The nucleocapsid protein of coronaviruses is known to be associated with the viral RNA and subgenomic RNAs and also to be an important part of the replication machinery (Compton *et al.*, 1987). As yet, no regions of the nucleocapsid protein have been assigned specific functions, but the basic regions are most likely involved in nucleic acid recognition or binding. However, there are no regions homologous to any known consensus RNA binding proteins. Functional assignments to specific regions of the nucleocapsid protein may provide the background for developing antiviral reagents that target critical residues.

CHAPTER III

ANALYSIS OF A HYPERVARIABLE REGION IN THE 3' NON-CODING END OF THE INFECTIOUS BRONCHITIS VIRUS GENOME

Introduction

Avian infectious bronchitis virus (IBV) is a highly contagious pathogen of chickens with many strains varying widely in virulence, serotype, and pathotype (Darbyshire *et al*, 1979; Hopkins,1974). IBV was the first coronavirus to be completely sequenced and is the prototype of the family *Coronaviridae* (Boursnell *et al*,1987). The IBV genome is 27.6kb in length and consists of single-stranded RNA of positive polarity with a polyA tail (Schochetman *et al*, 1977). Viral replication takes place under the control of a viral RNA-dependent RNA polymerase produced by translation of the genomic RNA. The polymerase presumably binds to the 3' end of the genomic RNA to initiate transcription, resulting in the production of negative sense RNA, either full length or subgenomic length, from which mRNAs can be transcribed (Boursnell *et al.*, 1987, Sethna *et al.*, 1989).

Previous studies on infectious bronchitis virus (IBV) cDNA have identified a T rich region of about 184 bases in the 3' non-coding terminus of both the U.S. prototype strain (Beaudette) and a Japanese strain (KB8523), that was not present in an antigenically closely related U.S. strain, Massachusetts (Mass) 41 (Boursnell *et al.*, 1985; Sutou *et al.*, 1988). Since the 3' non-coding region of the IBV genome should be important in the

recognition and/or binding of polymerase, we chose to examine the sequences of this region of the genome. RNA from IBV strains Holl52, ArkDPI (embryonating chicken egg passages) 11 and 75 and Mass41 were screened by slot blot analysis in order to determine whether or not a U-rich region could be detected using a probe corresponding to the nucleocapsid gene of IBV. The cDNA sequences of the 3' non-coding regions of three additional strains of IBV, Gray, Arkansas (Ark) 99 and Holland (Holl) 52, were determined and compared to the sequences of the Beaudette, KB8523 and Mass41 strains.

Materials and Methods

Viral Preparation

The viral isolates used in this study were propagated at 37°C for 36-48h following allantoic sac inoculation of 11-day old SPAFAS eggs. The eggs were then refrigerated at 4°C for 5h, and allantoic fluid collected and clarified by centrifugation at 16K x g for 20 min. Virus was precipitated overnight with 8% polyethylene-glycol, 2.33% NaCl, and concentrated by centrifugation at 16K x g for 20 min. Viral pellets were reconstituted in 1/50 vol 20mM Tris, 1mM EDTA, 100mM NaCl, pH7.4 and banded on a 30-50% glycerol/potassium tartrate gradient for 2h at 100K x g.

After concentrating by ultracentrifugation at 100K x g for 1h, virus pellets were resuspended in 0.1M NaOAc, pH5.2, and the virions were disrupted in 1mg/ml proteinase K and 1% SDS at 37°C for 1h. 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 while the remaining RNA was reprecipitated with ethanol.

Dot blot of four strains of IBV

A 40 base synthetic oligonucleotide was synthesized that was complementary to the middle of the Ark99 U-rich region, and was labelled at the 3' end using ^{32}PdCTP and terminal deoxy transferase (Sambrook *et al.*,1989). Genomic RNA from Mass41, Holl52, Arkansas DPI-11 and ArkDPI-75 (passage 11 and passage 75 of a strain from the same serogroup as ark99) was baked on to a nitrocellulose membrane at 80^{o}C. The RNA was applied in four-fold dilutions, starting with 1μg, and dot blot hybridization was performed at 42^{o}C overnight in the presence of formamide (Sambrook *et al.*,1989).

Hybridization studies with this Ark99 derived probe detected similar sequences in genomic RNA from the Ark DPI11 and Ark DPI75 strains, but not in Mass41 or Holl52 RNA, indicating, as expected, the presence of a similar U rich sequence in Ark DPI (Fig. 3-1). However, this probe also did not hybridize to Holl52, suggesting either that this strain did not contain this region or that it was too dissimilar to be recognized by the Ark99-derived probe. The cDNA sequences of two additional strains, Gray and Holl52, were then examined and a comparable region was found in both.

Cloning of the Gray, Ark99 and Holl52 Strains

Purified RNA was pelleted and dried in a Savant vacuum drier. RNA was resuspended in 7μl of deionized H_2O (dH_2O) and denatured in 6mM methyl mercuric hydroxide for 10 min at room temperature. First strand cDNA

synthesis was carried out using 5 units of AMV reverse transcriptase (BMB, Indianapolis, In) at 42°C for 2h in the presence of 14mM 2-mercaptoethanol (2ME), 50mM Tris pH8.3, 50mM KCl, 8mM $MgCl_2$, 0.8mM dNTP, 20µg/ml of 3' end primer (5' GGATCCGCTCTAACTCTATACTAGC CTAT 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 3h. The cDNA was then neutralized with HCl, concentrated by ethanol precipitation, and amplified using the polymerase chain reaction (PCR).

The PCR reaction contained the same 3' end primer as above, and a primer from the upstream intergenic region (5' GAATTCCCGCGTGTACCTC TCTAG TA 3') together with 4mM dNTP, 1x Taq buffer (1.5mM $MgCl_2$, 50mM KCl,10mM Tris-HCl, pH8) and 0.5µl *Taq* polymerase (Perkin Elmer, Norwalk, Ct.). The PCR was denaturated at 95°C for 3 min, then annealed at 50°C for 30s, and the sequence extended at 72°C for 1 min. Subsequent denaturation steps were 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 the Sequenase kit (USB, Cleveland, Ohio).

Computer Analysis

Resulting sequences were analyzed using the Genbank programs. Nucleotide sequences were processed using the Seqed program and comparisons of nucleotide sequences were made using Gap and Bestfit. IBV sequences were aligned using Pileup with Gray set as the reference

sequence. Output from the Pileup program was also displayed as a similarity plot using Plotsimilarity, or as a phylogenetic tree using the Macintosh program PAUP, version 3.0. Predicted secondary structures of the 3' non-coding region were derived from Fold and Squiggles.

Results

3' hypervariable region

The 3' non-coding regions of three strains of IBV were cloned and sequenced in order to investigate the functional importance of this U-rich sequence within the IBV group of coronaviruses and its extent in nature. A similar U-rich region of 184 bases was found in the sequence of the Ark99 genome immediately downstream of the nucleocapsid gene where it had been described in the Mass related strains.

A 40 base probe was synthesized that was complementary to the middle of this Ark99 region. Hybridization studies with the Ark99 derived probe detected similar sequences in genomic RNA from the Ark99 and ArkDPI strains, but not in Mass41 RNA, indicating, as expected, the presence of a similar U rich sequence in ArkDPI-11 and ArkDPI-75 (Fig. 3-1). However, this probe also did not hybridize to Holl52, suggesting either that this strain did not contain this region or that it was too dissimilar to be recognized by the Ark99-derived probe. After sequencing two additional strains, Gray and Holl52, a comparable region was found in both strains.

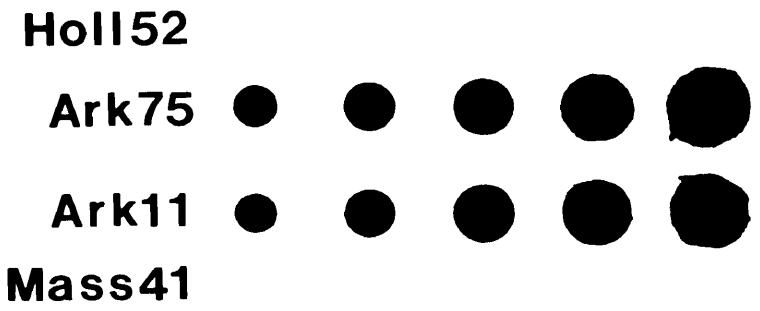


Fig. 3-1 RNA dot blot of 4 strains of IBV: Holl52, ArkDPI-11, ArkPI-75 and Mass41, with a cDNA probe from a portion of the T rich region in the 3' non-coding region of Ark99. RNA samples were applied to the membrane in four fold dilutions (from right to left).

Nucleotide sequences of the 3' non-coding region of IBV

The nucleotide sequences of the 3' non-coding regions of the Ark99, Gray and Holl52 strains of IBV were compared with the published sequences for Mass41, Beaudette and a Japanese strain, KB8523 (Boursnell *et al.*, 1985; Sutou *et al.*, 1988). Fig. 3-2 shows the alignment of these sequences and demonstrates the high degree of similarity among the strains.

However, the U rich region upstream of the 3' non-coding region, that was absent from Mass41, was found in Holl52, as well as Gray and Ark99, and was described as the 3' hypervariable region (HVR) because of the extent of its variation among strains. This HVR ranged from 184 (Gray) to 196 (Holl52) nucleotides in length and the nucleotide sequence homologies among the strains containing this sequence ranged from 53.2% to 92.8%, with the Holl52 strain showing the greatest divergence (Table 3-1). The greater variability of the Holl52 strain compared to the other strains in this region immediately downstream of the nucleocapsid gene would explain why this region was not detected with the Ark99 probe (Fig. 3- 1).

In comparison, the non-coding region downstream of this HVR was highly conserved, ranging from 94.3% to 97.5% identity (Table 3-2). Fig. 3-3 shows a plot of the similarity of the nucleotide sequences of the 3' non-coding regions of all six strains of IBV. The HVR is clearly seen immediately after the nucleocapsid gene and can be compared to the remaining 3' non-coding region which has very high similarity scores. This HVR could be considered either an insertion or a deletion since it is not present in Mass41. However, the higher U content in comparison to the rest of the IBV genome suggests that it is of exogenous origin (Boursnell *et al.*, 1985). The fact that it

Fig. 3-2 Nucleotide sequence comparison of the 3' non-coding region of six strains of IBV. Nucleotides are numbered from the stop codon for the nucleocapsid gene which is indicated by * * *. The hypervariable region starts immediately after the stop codon and continues to base 213. Conserved bases in this region are underlined.

```
                                                                                                                    100
Ark99       GGATTCAGCT TTAGGTGAGA ATGAACTTTG A.G..... .......TAA CATAATGGAC CTGCTGCATT CTTTGGTAC. ........AT TTTGTTAGCC
Gray        AGATTCTGCT TTAGGTGAGA ATGAACTCTG AGG..... .......TAA CATAATGGAC CTGCTGCATT TTTTGGTAC. ........AT TTTGTTAGCC
Beaudette   GGATGCAGCT CTAGGAGAGA ATGAACTTTG A.G..... .......TAA CATAATGGAC CTGTTG..TT TCCTGGTAC. ........AT TTTGTTAAAC
KB8523      TGACTCAGCA CTTGGTGAAA ATGAACTTTG A.T..... .......TAA CATAATGGAC TTGCTGCACT TTCTCTGGC. ........AT TTCTTTGTGT
Mass41      GGATTCAGCT CTAGGAGAGA ATGAACTTTG AGT..... .......... .......... .......... .......... .......... ..........
Holl52      GGATTCAGCA CTTGGAGAGA ATGAGTTGTA AAGCTAGATT TCCAACTTAA CATCATGGAC GTGCGTATGC TGTTTTTCCC TACTATAGAC TTCTTAGCAT
                                      ** *
                                                                                                                    200
Ark99       ATTATTTCTG TGCTTTCCTA TCAATCATTA CAGGCATTGA TTGTGATCAT GTGCAATATT TAAGTTTCTT TTGGTTGCTT TTTGCTTGTT GTATTGTTGC
Gray        ACTATTTTTG TGCTTTCCTA TCAATTATTA CAGGCATTGA TTGTGATTAT GTGCAATATT TAAGCTTCTT TTGGTTGCTT TTTGCTTGTT GTGTTGTTGC
Beaudette   ACTATTTCTG TGCTTTCCTA TCAATTATTA CAGGCATTGA TTGTGATTAT GTTCAATACT TAAGCTTCTT TTGGTTGCTT TTTGCTTATT GTATTGTTGC
KB8523      TTTGTTTTTG TAGATTTCTC ACAATTATTA TTGG.AATGA TTGTGATTAT GTGCAGTATC TTAGTTGGTT TAATTATCTT GTTGCAGTTT GTGTTCTTTC
Mass41      .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Holl52      ATTATTTTTT GCTATTTGTA TGGTTTATTA CAGGTGAAGA TTGTATGTAT TTGTTGTACA CTCGTATGTT CTATATTATG TTTTCTGTAG TTGTTATTAG
                                                                                                                    300
Ark99       TGTGCTTTTT ATTATTGTGA TTCTCATTAG TTTGTTTTAT CGTAGAAGTT CAATAGTAAG AGTTAAGGGA GATAGGCATG TAGCTTGATT ACCTACATGT
Gray        TGTGCTTTTT ATTATTGTCA TTCTC...AG TTTGCTTTAT TGTAGAAGTT CAATAGTAAG AGTTAAGGAA GATAGGCATG TAGCTT.AGC ACCTACATGT
Beaudette   TGTGCTTTTT ATTGTTGTGA TTCTCATTAG TTTGCTTTAT CGTAGAAATT CAATAGTAAG AGTTAAGGAA GATAGGCATG TAGCTTGATT ACCTACATGT
KB8523      TGTGCTTATT GTTTTTATGC TGCTTTATTC TTTGTTATAT CGTAGAAGTT ...TATTGAA AGTTAAGGAA GATAGGCATG TAGCTTGATT ACCTACATGT
Mass41      .......... .......... .......... .......... ....AAAATT CAATAGTAAG AGTTAAGGAA GATAGGCATG TAGCTTGATT ACCTACATGT
Holl52      TGTTGTTCTT GTTCTTACTC TACTGTTCTC TTTTCTTTAT TTTAGAGTAT CAATA...AG AATCAAGGAA GATAGGCATG TAGTTTGATT ACCTACATGT
                                                                                                                    400
Ark99       CTATCGCCAG GGAAATGTCT AATCTGTCTA CTTAGTAGCC TGGAAACGAA CGGTAGACCC TTAGATTTTA ATTTAGTTTA ATTTTTAGTT TAGTTTAAGT
Gray        CTATCGCCAG GGAAATGTCT AATCTGTCTA CTTAGTAGCC TGGAAACGAA CGGTAGACCC TTAGATTTTA ATTTAGTTTA ATTTTTAGTT TAGTTTAAGT
Beaudette   GTATCGCCAG GGAAATGTCT AATCTGTCTA CTTAGTAGCC TGGAAACGAA CCGTAGACCC TTAGATTTTA ATTTAGTTTA ATTTTTAGTT TAGTTTAAGT
KB8523      CTATCGCCAG GGAAATGTCT AATCTGTCTA CTTAGTAGCC TGGAAATGAA CGGTAGACCC TTAGATTTTA ATTTAGTTTA ATTTTTAGTT TAGTTTAAGT
Mass41      CTATCGCCAG GGAAATGTCT AATTTGTCTA CTTAGTAGCC TGGAAACGAA CGGTAGACCC TTAGATTTTA ATTTAGTTTA ATTTTTAGTT TAGTTTAAGT
Holl52      CTATCGCCAG GGAAATGTCT AATCTGTCTA CTTAGTAGCC TGGAAACGAA CGGTAGACCC TTAGATTTTA ATTTAGTTTA ATTTTTAGTT TAGTTTAAGT
                                                                                                                    500
Ark99       TAGTTTAGAG TAGGTATAAA GATGCCAGTG CCGAGGCCAC GCGGAGTACG ATCGAGGGTA CAGCACTAGG ACGCCCACTA GGGGAAGAGC TAAATTTTAG
Gray        TAGTTTAGAG TAGGTATAAA GATGCCAGTG CCAGGGCCAC GCGGAGTACG ATCGAGGGTA CAGCACCAGG ACGCCCATTA GGGGAAGAGC TAAATTTTAG
Beaudette   TAGTTTAGAG TAGGTATAAA GATGCCAGTG CCGGGGCCAC GCGGAGTACG ATCGAGGGTA CAGCACTAGG ACGCCCATTA GGGGAAGAGC TAAATTTTAG
KB8523      TAGTTTAGAG TAGGTATAAA GATGCCAGTG CCGGGGCCAC GCGGAGTACG ATCGAGGGTA CAGCACTAGG ACGCCCACTA GGGCAAGAGC TAAATTTTAG
Mass41      TAGTTTAGAG TAGGTATAAA GATGCCAGTG CCGGGGCCAC GCGGAGTACG ACCGAGGGTA CAGCACTAGG ACGCCCATTA GGGGAAGAGC TAAATTTTAG
Holl52      TAGTTTAGAG TAGGTATAAA GAAGCCAGTG CCGGGGCCAC GCGGAGTACG ATCGAGGGTA CAGCACTAGG ACGCCCACTA GAGGAAGAGC TAAATTTTAG
                                                                                                                    563
Ark99       TTTAAGTTAA GTTTAATTTA CTAAGTATAG TTAAAATTTA TAGGCTAGTA TAGAGTTAGA GC.
Gray        TTTAAGTTAA GTTTAATTGG CTAAGTATAG TTAAAATTTA TAGGCTAGTA TAGAGTTAGA GC.
Beaudette   TTTAAGTTAA GTTTAATTGG CTAAGTATAG TTAAAATTTA TAGGCTAGTA TAGAGTTAGA GC.
KB8523      TTTAAGTTAA GTTTAATTGG CTAAGTATAG TTAAAATTTA TAGGCTAGTA TAGAGTTAGA GCA
Mass41      TTTAAGTTAA GTTTAATTGG CTATGTATAG TTAAAATTTA TAGGCTAGTA TAGAGTTAGA GC.
Holl52      TTTAAGTTAA GTTTAATTGG CTAAGTATAG TTAAAATTTA TAGGCTAGTA TAGAGTTAGA GC.
```

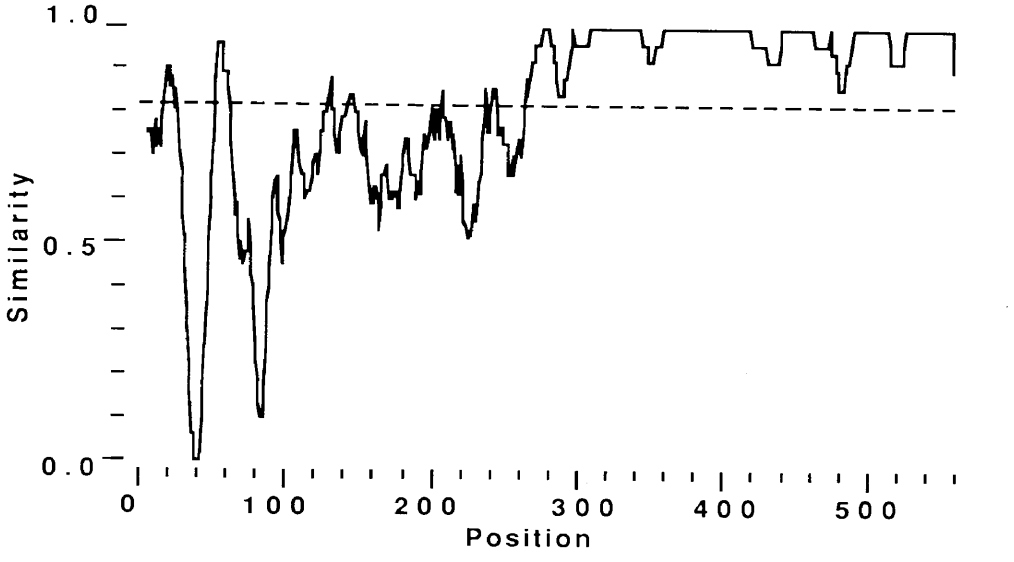


Fig. 3-3 Similarity plot of the 3' non-coding region of six strains of IBV. This plot starts from the end of the nucleocapsid gene and covers the 3' HVR and conserved 3' end.

Table 3-1. Percent homologies of the nucleic acid sequences of the 3' HVR in 5 strains of IBV.

	Ark99	**Gray**	**Beau**	**Holl52**	**KB8523**
Ark99	100	91.8	91.8	53.2	68.6
Gray		100	92.8	57.9	67.6
Beau			100	57.1	66.7
Holl52				100	60.0
KB8523					100

Table 3-2. Percent homologies of the 3' non-coding region downstream of the HVR of 6 strains of IBV.

	Ark99	**Gray**	**Beau**	**Holl52**	**KB8523**	**Mass41**
Ark99	100	97.2	97.5	94.7	97.2	96.3
Gray		100	97.8	94.3	96.8	96.6
Beau			100	94.7	97.2	97.5
Holl52				100	95.3	94.3
KB8523					100	94.7
Mass41						100

has at least 53.2% identity among various strains suggests a common ancestry for this sequence.

If the IBV genome acquired the HVR as exogenous RNA, this region could be a site for recombination. Sequences upstream and downstream of the HVR were analyzed to identify any distinctive patterns which would suggest that this region is of functional significance, such as being involved in recombination. Not only are the sequences flanking the 3' HVR highly conserved, the 5' flanking sequence also resembles the cutting site (5' TTAA 3') of retrovirus DNA and transposons Tn10 and Tn3 (Shoemaker *et al.*, 1980). It was of special interest that a unique pattern, that may be of functional importance, was observed in the oligonucleotides flanking the HVR. Several nucleotides downstream of this 3' HVR were mirrored in the sequences upstream. That is, the same sequences were seen in reverse orientation on either side of the 3' HVR. An example can be seen in the Gray strain where the sequence TGAGAATG appears 5' of the HVR and the sequence GTAAGAGT is in the 3' flanking sequence (Table 3-3). The function of these mirrored sequences is unknown; however, they may be important in preventing folding of the RNA in this region, as mirrored sequences should prevent complementary base pairing.

The variability of this HVR as compared to the conservation of the remaining 3' non-coding sequences suggests that the primary sequence itself may not be as functionally important as its influence on the secondary structure of the 3' end of the genome. The impact of the 3' HVR on RNA secondary structure was examined by determining the predicted secondary structure of the 3' non-coding region of each of these six IBV strains. Putative secondary structures, which can be seen in Fig. 3-4, show that the 3' non-

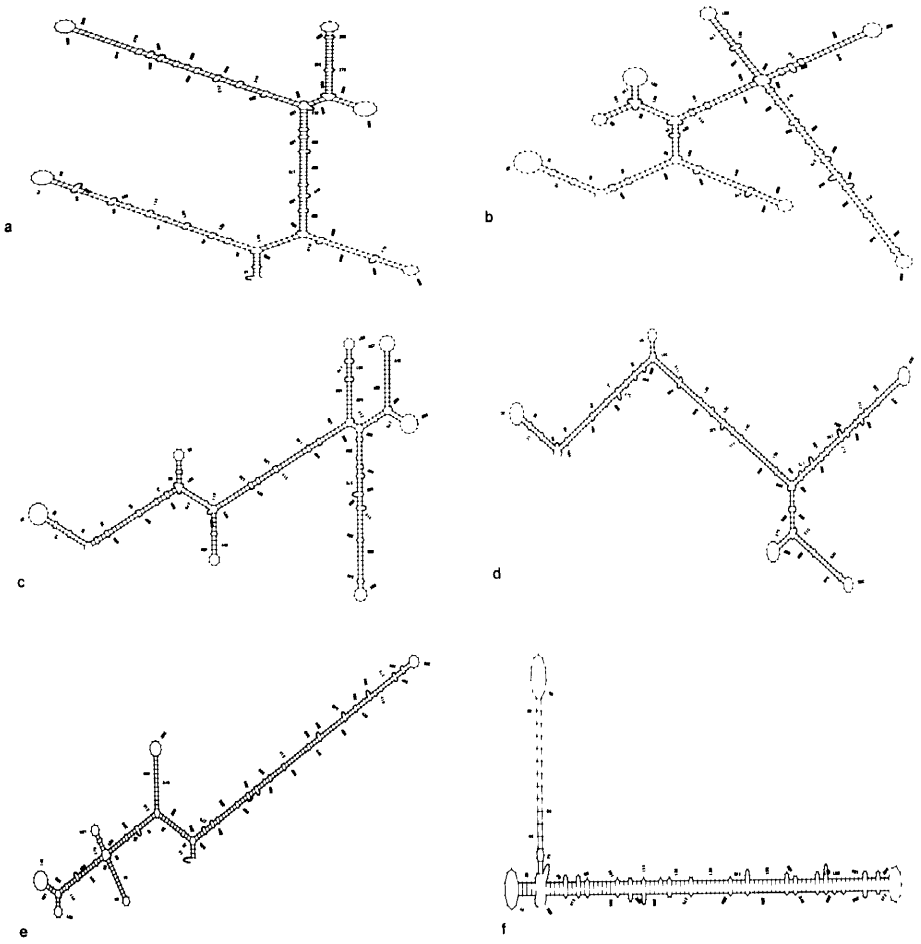


Fig. 3-4 Putative secondary structures of the 3' non-coding regions of six strains of IBV. a) Holl52 b) Gray c) Ark99 d) Beaudette e) KB8523 f) Mass41.

Table 3-3. Mirrored sequences flanking the 3' hypervariable region. |*| indicates the position of the insert.

ARK99	**TGAGAATGAAC**TTT**GA**A|*|**AGTTCAA**TA**GTAAGAGT**
GRAY	**TGAGAATGAA**CTC**TGA**GG|*|**AGT**TC**AA**TA**GTAAGAGT**
BEAUD	A**GAGAATGAACTT**TGAG|*|AA**TTCAA**T**AGTAAGAG**TT
KB8523	**GAA**AA**TGAA**C**TTTGA**T|*|**AGTT**TATTGA**AAGTTAAG**
MASS41	**GAGAATGAACTT**TGAGT|*|AAAA**TTCAA**TA**GTAAGAG**T
HOLL52	**GGA**G**AGAA**TGAGTTG**TAA**|*|GTATC**AATAAGA**ATCA**AGG**

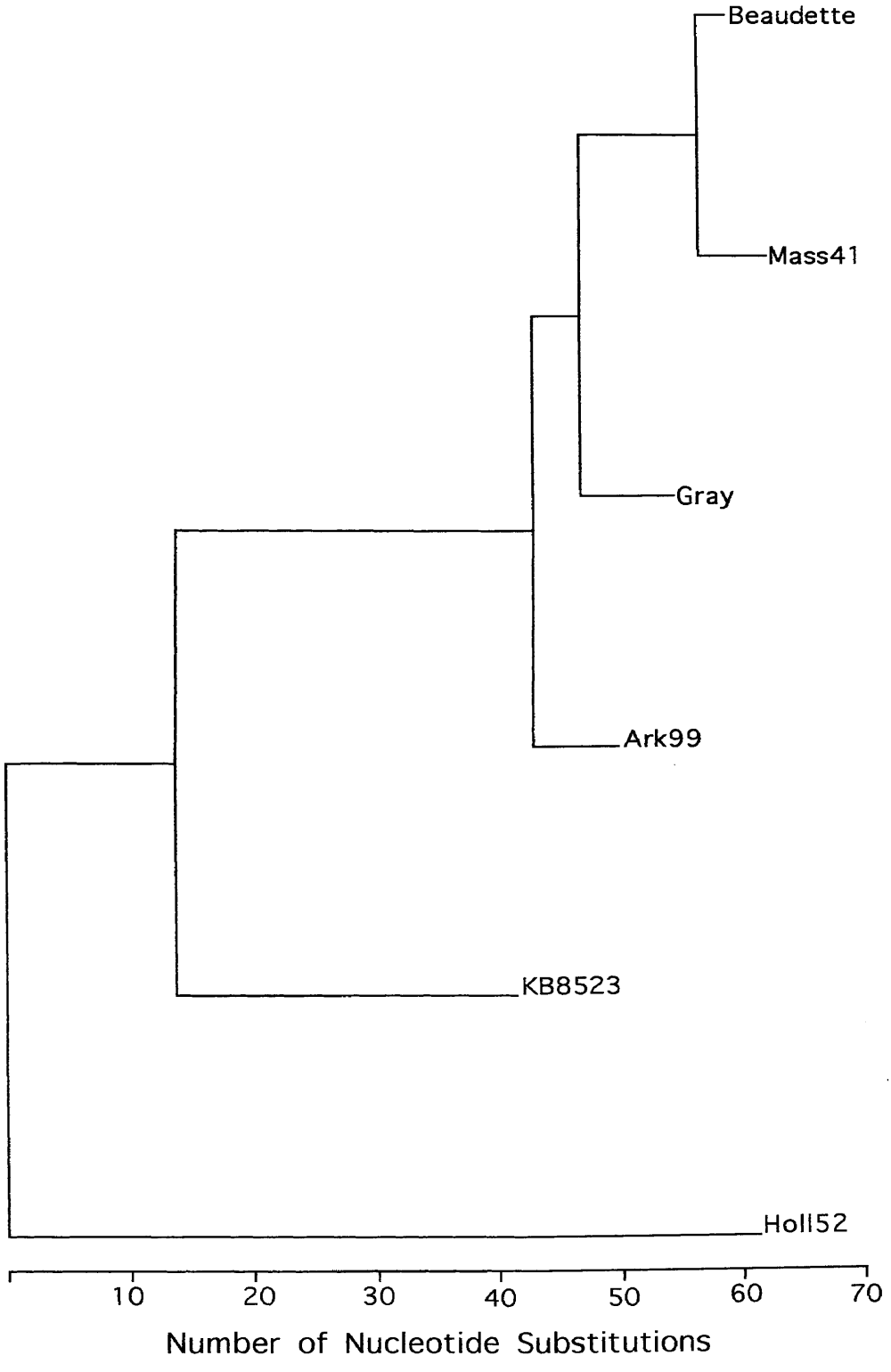


Fig. 3-5 Phylogenetic tree of six strains of IBV based on the sequences of their 3' non-coding regions.

coding region of the Mass41 strain, without the 3' HVR, has fewer branches with hairpin loops than the other strains. Although this program automatically aligns the ends of the input sequence and the rest of the genome is ignored, it still gives an idea of how the secondary structure may be affected by the presence of the 3' HVR.

Phylogeny

The sequence relationships of the 3' non-coding regions were examined with a phylogenetic tree (Fig. 3-5). The Holl52 and KB8523 strains were shown to be more distantly related to the other strains. The 3' non-coding sequences of Mass41, in spite of the absence of the HVR, appeared to be closely related to the Beaudette, and to a lesser extent to the Gray and Ark99 strains. The program used to determine this tree counts the absence of this whole HVR in Mass41 as the result of one recombination event, so only the remainder of the 3' non-coding region of the Mass41 strain was considered for the construction of the phylogenetic tree.

Discussion

There is considerable evidence for recombination in the Coronaviridae family, presumably resulting from polymerase pausing, detachment and homologous or heterologous reattachment during RNA synthesis. (Banner *et al.*, 1990; Baric *et al.*, 1990; Keck *et al.*, 1988a; Keck *et al.*, 1988b; Kusters *et al.*, 1990, Makino *et al.*, 1986; van der Most *et al.*, 1992; Lai, 1992). Recombination may also be responsible for the emergence of the nonenteropathogenic porcine respiratory coronavirus

(PRCV). PRCV is thought to have been derived from transmissible gastroenteritis virus (TGEV) by a deletion of a region of approximately 240 amino acids in the S protein. This deletion has resulted in the absence of two antigenic sites in PRCV isolates and is the basis for their differentiation from the enteropathogenic TGEV (Sanchez *et al.*, 1992).

The location of the IBV 3' HVR is conserved at a position immediately downstream from the stop codon of the nucleocapsid gene. Unlike the case with the porcine coronaviruses, there is no evidence as yet to show that this HVR has any effect on virulence or pathogenesis. However the sequences flanking the 3' end of the HVR are similar to the tetrameric (TTCC) or heptameric (AGTTTCC) repeats found within the regions deleted from the PRCVs.

There are no detectable protein products from the 3' HVR and no methionines with a preceding putative consensus ribosomal binding site. Since it is completely absent in the Mass41 strain, it would appear that it is not an absolute requirement for viral replication, at least for Mass41. Any specific function of this HVR apparently does not dictate that the sequences be highly conserved but its function may depend on the secondary structure of the RNA. The variability of this HVR as compared to the conservation of the remaining 3' non-coding region suggests that the primary sequence itself may not be as functionally important as its influence on the secondary structure of the 3' end of the genome. It is possible that the presence of the HVR may increase the complexity of the secondary structure of the RNA which could in turn affect its association with replication machinery, such as RNA binding proteins.

CHAPTER IV

EXPRESSION OF THE NUCLEOCAPSID GENE OF IBV AS A HISTIDINE FUSION PROTEIN IN E. COLI

Introduction

The nucleocapsid protein of IBV is one of the three major structural proteins of the virion. This protein is internally localized, makes up approximately 40% of the total viral protein, and is thought to be involved in the induction of protective immunity (Cavanagh, 1983; Boots *et al.*, 1992). Such protection may be induced by generating cytotoxic T cells (Lutticken *et al.*,1988), and also by generating T helper cell responses that augment the activity of B cells in the production of viral-neutralising antibody specific for the spike protein (Boots *et al.*, 1992). Therefore, the nucleocapsid protein is a potential target for vaccine development, especially as it is much more highly conserved than the spike protein, which is responsible for the induction of neutralising antibodies.

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). Further information on this RNA-protein interaction may help with the development of either a vaccine or anti-viral agents specific for all serotypes of IBV.

With this goal in mind, the nucleocapsid gene of IBV was sub-cloned into the pQE8 expression vector (Qiagen, Chatsworth, CA) and subsequently produced in *E. Coli* with a tag of six histidine (6 x His) residues

(Fig. 4-1). This Qiaexpress system allows for the expression of recombinant proteins in *E. coli* which may then be purified by a single, rapid chromatography step based on the affinity of the 6 x His tag for a Ni^{2+}-NTA column (Fig. 4-2a). The recombinant protein may then be easily eluted by competition with imidazole which binds to the Ni-NTA and displaces the histidine tagged protein (Fig. 4-2b).

Materials and Methods

PCR amplification of the nucleocapsid gene

The nucleocapsid gene of the Gray strain of IBV was cloned and sequenced (Chapter II). Specific primers spanning the nucleocapsid gene were chosen to perform PCR amplification of the nucleocapsid gene of IBV, in order to subclone it into the correct open reading frame (orf) of the pQE8 vector. PCR was performed with the upstream primer 5' TCTGGATCCATGG CAAGCGGTAAGGC 3' and the downstream primer 5' GTTGGATCCGAGTT CATTCTCACCTAAAGC 3', in the presence of 4mM dNTP, 1x Taq buffer (1.5mM $MgCl_2$, 50mM KCl,10mM Tris-HCl, pH8) and 0.5µl *Taq* polymerase (Perkin Elmer, Norwalk, Ct.). The PCR was performed with a primary denaturation step at 95°C for 3 min, annealing at 50°C for 30s, 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.

The PCR primers were designed with BamH1 recognition sequences at their 5' ends. This enabled 5µg of the resulting PCR product to be digested with 2units of BamH1 (Promega corporation, Madison, WI) in buffer

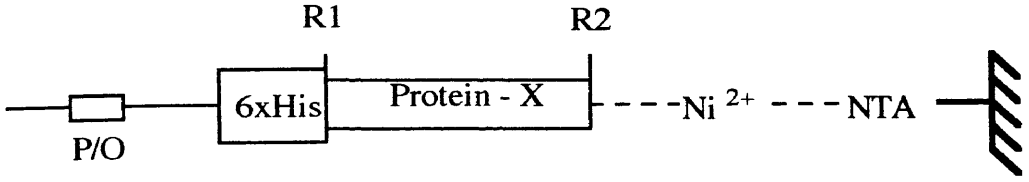


Fig. 4-1 Diagrammatic representation of the nucleocapsid gene in the QIAexpress system. The nucleocapsid gene was subcloned into the BamH1 site of pQE8. P/O represents the promotor and two lac operators.

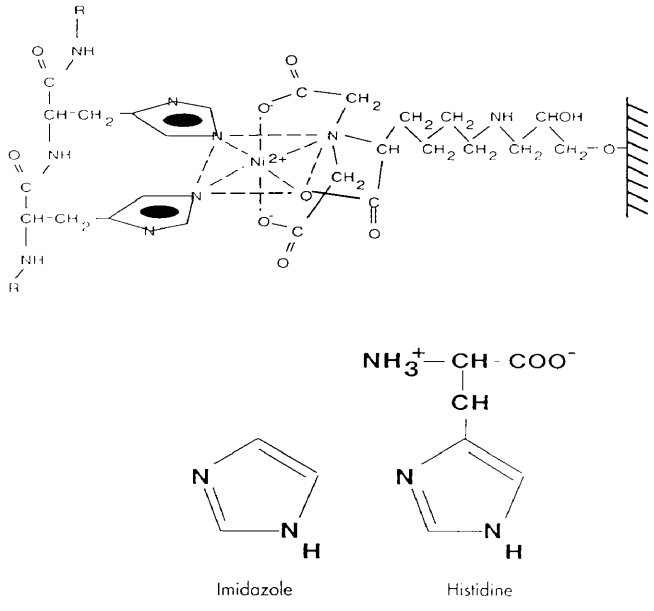


Fig. 4-2 Model for the binding of neighboring 6 x His residues to Ni-NTA resin. NTA associates with 4 of the 6 Ni^{2+} spheres leaving 2 to bind to the histidine residues (above), and structures of imidazole and histidine (below).

C (50mM NaCl, 10mM Tris-HCl, pH7.9, 10mM $MgCl_2$, 1mM DTT) at 37°C for 2h. Following the addition of a 1/10 volume 10X loading buffer, the reaction solutions were loaded onto a 0.8% low melting point agarose gel and electrophoresed at 60V to purify the digested nucleocapsid gene PCR product. The nucleocapsid DNA band was excised and the agarose melted at 56°C in 10mM Tris-HCl, pH 7.6, 1mM EDTA, pH 8.0 (TE). Following phenol extraction, the DNA was precipitated with 0.3M NaOAc, pH5.2 and two volumes of ethanol at -20°C for 2h.

Preparation of cells for electroporation

M15/pRep4 cells were grown overnight at 37°C with shaking in 2ml Luria Bertani (LB) media (10g/L Bacto tryptone, 5g/L Bacto yeast extract, 10g/L NaCl, pH 7) supplemented with 25μg/ml kanamycin. 200ml LB media supplemented with 25μg/ml kanamycin were then inoculated with this 2ml culture and shaken at 37°C for 3h. The cells were spun at 6K x g for 5min and the pellet was resuspended in 200ml cold dH_2O. This step was repeated twice with the cells being resuspended in 100ml cold dH_2O then 15ml 10% glycerol in dH_2O. Following a final spin of 20min at 10K x g, the cells were resuspended in 400μl 10% glycerol in dH_2O, and 40μl were aliquoted into pre-chilled microfuge tubes. These cells were then stored at -70°C until they were needed.

Cloning into the pQE8 vector

Two hundred ng of pQE8 were digested with BamH1 and the fragments gel purified as described above. A ligation reaction was set up overnight at 11°C with 10ng nucleocapsid DNA and 5ng pQE8 DNA, in the

presence of 1 unit T4 DNA ligase (Promega, Madison, WI), 300mM Tris, pH7.8, 100mM $MgCl_2$, 100mM DTT and 10mM ATP. These ligation products were then used to transform M15/pRep4 cells. Cells were thawed on ice and 1µl of the ligation product was added to 40µl cell aliquots. The cells were subjected to electroporation at 12.5kV/cm for 5msecs (BTX, San Diego, CA). Following the addition of 1.5ml LB the cells were grown at 37°C for 1h with shaking before spreading 50µl onto plates containing 25µg/ml kanamycin and 100µg/ml ampicillin. Plates were incubated at 37°C overnight and recombinant colonies selected for screening the following morning.

Recombinant colonies were streaked onto LB plates containing 25µg/ml kanamycin and 100µg/ml ampicillin, and incubated at 37°C overnight. The bacteria were collected using blunt toothpicks and resuspended in 80µl STE buffer (100mM NaCl, 20mM Tris, pH7.5 and 10mM EDTA). DNA was extracted after mixing with 80µl phenol: chloroform: isoamyl (25:24:1) and 0.25µg RNase was added to the DNA fraction (top layer). One µl of DNA was examined on a 1% agarose gel and the remainder was precipitated with 0.3M NaOAc, pH5.2, and 2 volumes of ethanol in a dry ice-ethanol bath for 5min. DNA was collected following microcentrifugation at 12,000g for 10min.

BamH1 digestion was performed as described above and clones that appeared to have an insert were then sequenced using Sequenase (USB, Cleveland, Ohio) in order to confirm that the correct sequence was inserted into the vector and that the open reading frame was intact. Positive clones were screened for nucleocapsid protein production.

Rapid screening of small scale expression cultures

Single colonies from positive transformants were grown at 37°C with shaking in individual 10ml cultures (LB, 100μg/ml ampicillin, 25μg/ml kanamycin). When the OD_{600} reached 0.7-0.9, IPTG was added to a final concentration of 2mM and cultivation continued for another 5h. Cells were collected from 1ml of the culture by centrifugation for 1.5min (Eppendorf microcentrifuge) and boiled for 7min in 200μl sample buffer (3% 2-ME, 3% SDS, and 0.5% bromophenol blue). After cooling, 10μl were applied to a 12.5% SDS polyacrylamide gel (Sambrook *et al.*, 1989) and subjected to electrophoresis at 200v (Bio-Rad mini PAGE, Bio-Rad, Richmond, CA) until the bromophenol blue dye front reached the bottom of the gel. The proteins were then visualized by staining with Coomassie blue (Sambrook *et al.*, 1989).

Determination of expressed recombinant protein solubility and compartmentalization

Positive clones were chosen to express recombinant nucleocapsid protein. In order to choose the best purification protocol using non-denaturing buffers, the localization of the recombinant protein was determined. Cells were grown and induced in 100ml cultures of LB media as described above along with a control culture of untransformed M15/pRep4 cells. Just before induction, a 1ml sample was taken from each of the cultures for use as an uninduced control on SDS-PAGE. These cells were resuspended in 50μl SDS-PAGE buffer and kept on ice. Before harvesting the remaining cells, a second sample of 0.5ml was taken as the induced

control, the cells were collected as before, resuspended in 100µl SDS-PAGE buffer and kept on ice.

The induced cultures were divided into 2 aliquots, and the cells harvested by centrifugation at 4000 x g for 10min. The supernatants were discarded and one cell pellet was resuspended in 5ml sonication buffer (50mM NaH_2PO_4, pH8, 300mM NaCl), supplemented with 1mg/ml lysozyme, and kept on ice for 5 min before processing for cytosolic localization. The other cell pellet was resuspended in 10ml of 30mM TrisHCl, pH8, 20% sucrose, in order to check for periplasmic localization.

The cells examined for cytosolic localization were frozen in a dry-ice ethanol bath and thawed in cold water before sonicating for 3min. Following centrifugation at 10,000 x g for 20min, the supernatant was decanted and saved on ice. This fraction contained the soluble proteins. Meanwhile, the cell pellet was resuspended in sonication buffer to suspend the insoluble fraction.

The samples examined for periplasmic localization were adjusted to 1mM EDTA and incubated at room temperature for 10min on a rocking platform. Following centrifugation at 8000 x g for 10min, the supernatant was removed and the cell pellet resuspended in 10ml ice cold 5mM $MgSO_4$. After shaking for 10min in an ice-water bath, the cells were centrifuged as above and the supernatant, or osmotic shock fluid, decanted and saved.

These three fractions of each of the two samples and the cell control were then analyzed by SDS-PAGE. Five µl of 2 x SDS-PAGE sample buffer was added to 5µl of the crude extracts of insoluble and soluble protein, and 10µl of SDS-PAGE sample buffer were added to 10µl of the osmotic shock fluid. These samples were boiled, along with the uninduced and induced

cell samples, for 5min. After microcentrifugation for 1min, 20µl of each of the samples were loaded onto an SDS-PAGE gel as described above.

Extraction and purification of the recombinant nucleocapsid protein under non-denaturing conditions

A positive clone was grown and induced in a 1L culture as described above. The cells were harvested by centrifugation at 4,000 x g for 20min, then resuspended in 6ml sonication buffer supplemented with 1mg/ml lysozyme. The sample was incubated for 5min on ice, then 0.66ml 3M NaCl was added and the sample incubated on ice for another 5min. The cells were sonicated for 5min on ice, centrifuged at 10,000 x g for 30min and the supernatant collected.

The supernatant was loaded onto a large (8ml) Ni^{2+}-NTA column (Qiagen, Chatsworth, CA) pre-equilibrated with sonication buffer. The column was washed with 10 volumes of sonication buffer supplemented with 20mM imidazole to reduce the background of cellular proteins. The recombinant nucleocapsid protein was eluted with sonication buffer, pH 7, supplemented with 250mM imidazole. Two ml fractions from the column were collected using a fraction collector connected to a UV monitor (Biorad) to determine when the protein was eluted. Ten µl of each of these fractions was then analyzed by SDS-PAGE and a duplicate gel was transferred to nitrocellulose membrane for western blotting analysis with IBV polyclonal antisera. Fractions containing the recombinant protein were concentrated using an Amicon centriprep 10 concentrator (Amicon, Beverly, MA) and duplicate samples were run on a SDS-PAGE for staining with Coomassie or transferred to nitrocellulose for western blotting with anti-IBV antisera.

Results

Protein expression

In order to study the interaction of IBV RNA with the nucleocapsid protein, the gene was sub-cloned into the pQE8 expression vector which can be used for the expression of large amounts of the purified protein. Out of a number of recombinant colonies screened, two were chosen that upon sequencing showed the nucleocapsid gene to be inserted into the vector with the ORF intact. These two clones, #3 and #68, were then analyzed in small scale expression cultures to confirm that the recombinant protein was synthesized upon induction with IPTG.

Duplicate samples of the small scale expression cultures from the two clones and a cell control were analyzed by SDS-PAGE (Fig. 4-3). The recombinant nucleocapsid protein is indicated by an arrow and corresponds to the expected MW of the nucleocapsid protein, which is approximately 50kD. Even against the background of cellular proteins, this band could be clearly seen in the induced cultures of clones #3 and #68, and is absent from the cell controls. Clone #68 expressed the nucleocapsid protein more efficiently than clone #3.

Solubility and localization of the recombinant protein.

Before expressing the nucleocapsid protein on a preparative scale under non-denaturing conditions, the solubility and compartmentalization of the recombinant protein needed to be determined. Three fractions of each of the samples representing the insoluble protein matter, the cytoplasmic

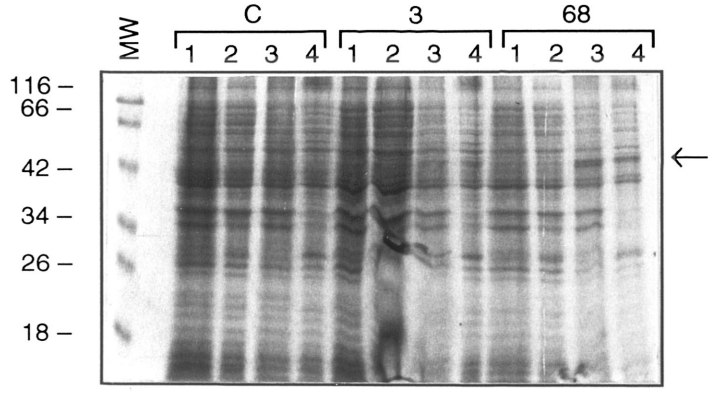


Fig. 4-3 Crude protein preparations of uninduced and induced cultures of clones #3 and #68 and a cell control. The recombinant protein is indicated by an arrow and is present in the induced lanes of #3 and #68.

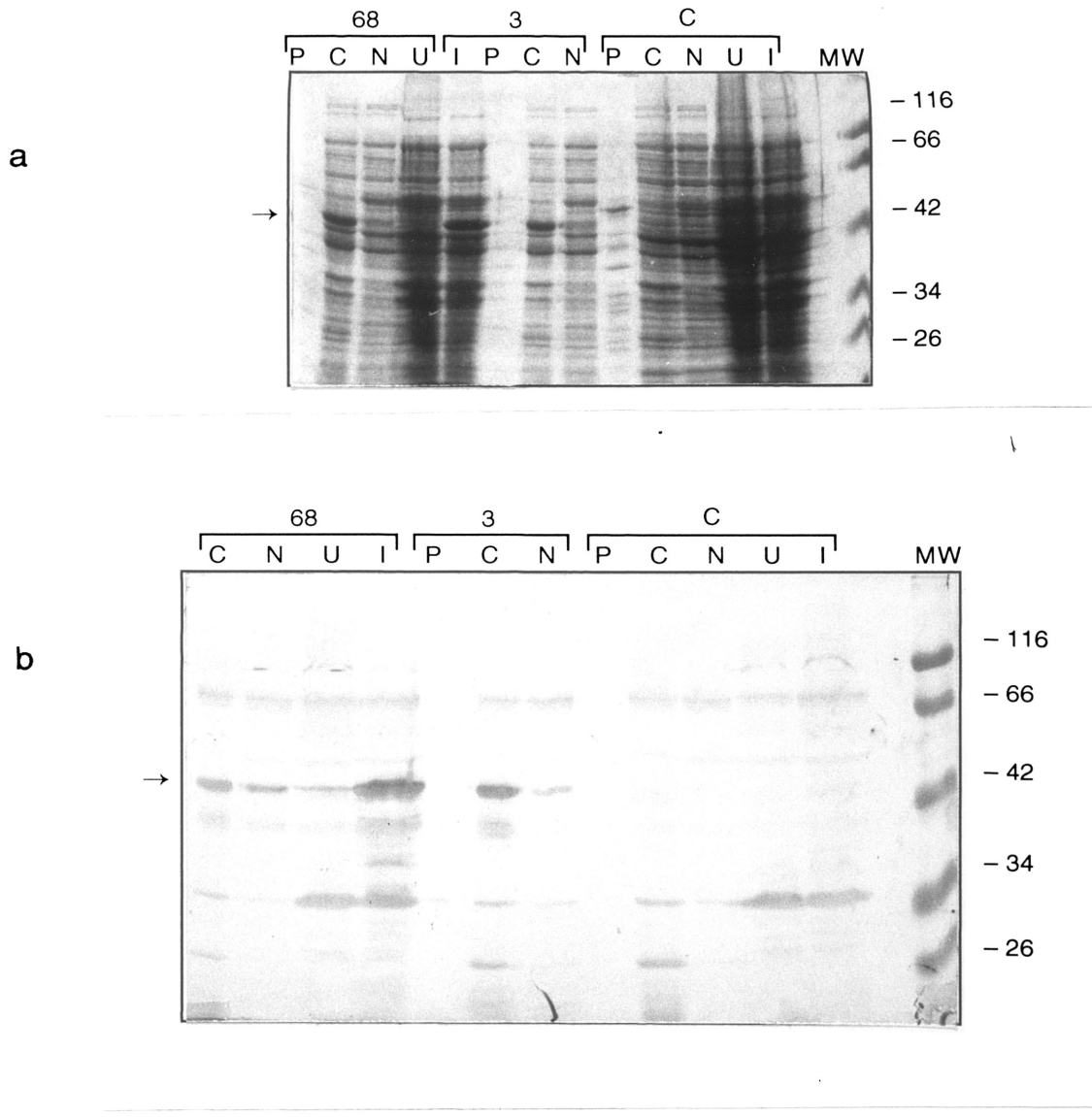


Fig. 4-4 a) SDS-PAGE of cellular fractions collected to determine the cellular localization and solubility of the recombinant protein. Recombinant protein is indicated by an arrow and is present in the cytosolic and insoluble fractions of #68. b) represents a parallel western blot of (a) showing the nucleocapsid protein in all lanes of the two clones except for the periplasmic fraction lane. Recombinant protein is indicated by an arrow.

protein fraction and the periplasmic protein fraction were collected. Samples of these protein fractions, along with uninduced and induced samples, were run on an SDS-PAGE and also transferred to nitrocellulose for western blot analysis with polyclonal anti-IBV antibody. The recombinant protein is indicated by an arrow and was present in both the insoluble and cytoplasmic fractions, but not in the periplasmic fraction (Fig. 4-4a). The fact that the protein was present in the cytoplasmic fraction indicated that it was soluble, and its presence in the insoluble fraction was probably due to incomplete lysis of the cells. For future expression cultures, the cells were sonicated for 5min instead of 3min, and most of the protein was then seen in the cytoplasmic fraction.

Although there was some background non-specific protein recognition, the western blot of these fractions clearly showed the nucleocapsid protein in the cytoplasmic and insoluble fractions of clones #3 and #68, as well as in the induced samples of #68 (Fig. 4-4b). There was no binding of anti-IBV antibody to proteins at the corresponding position in the cell control lanes.

Purification of the recombinant protein

Having determined the localization of the recombinant protein, large quantities of the protein were produced for purification by Ni-NTA affinity chromatography. UV monitoring of the fractions collected from the column showed large amounts of protein released from the column very soon after the addition of 250mM imidazole. The gel of these fractions (Fig. 4-5a) shows the nucleocapsid protein was released in eluate fractions 2 through

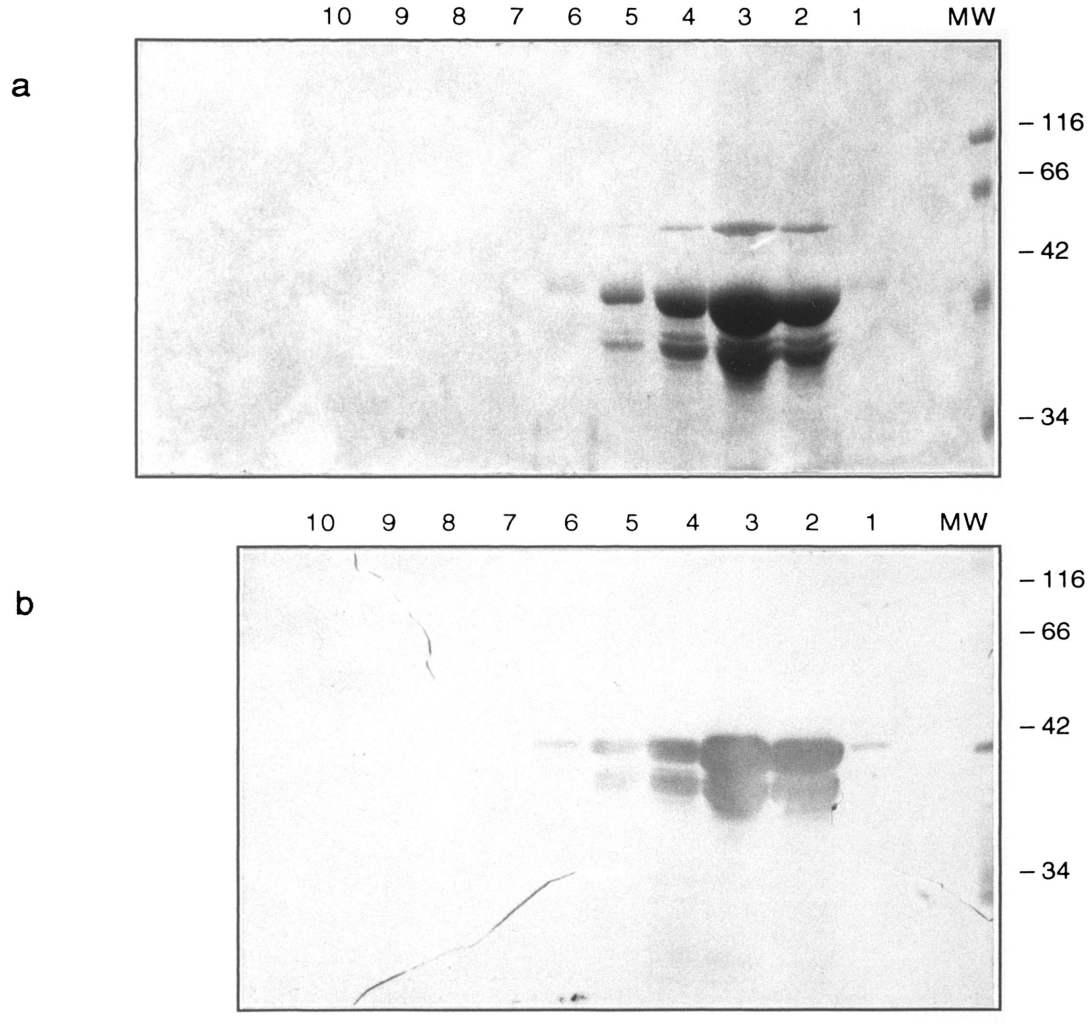


Fig. 4-5 Elution fractions collected following the addition of 250mM Imidazole to the Ni-NTA column: a) SDS-PAGE, b) Western blot. Recombinant protein is indicated by an arrow, and is seen in fractions 2 to 6 on both the gel and the western blot.

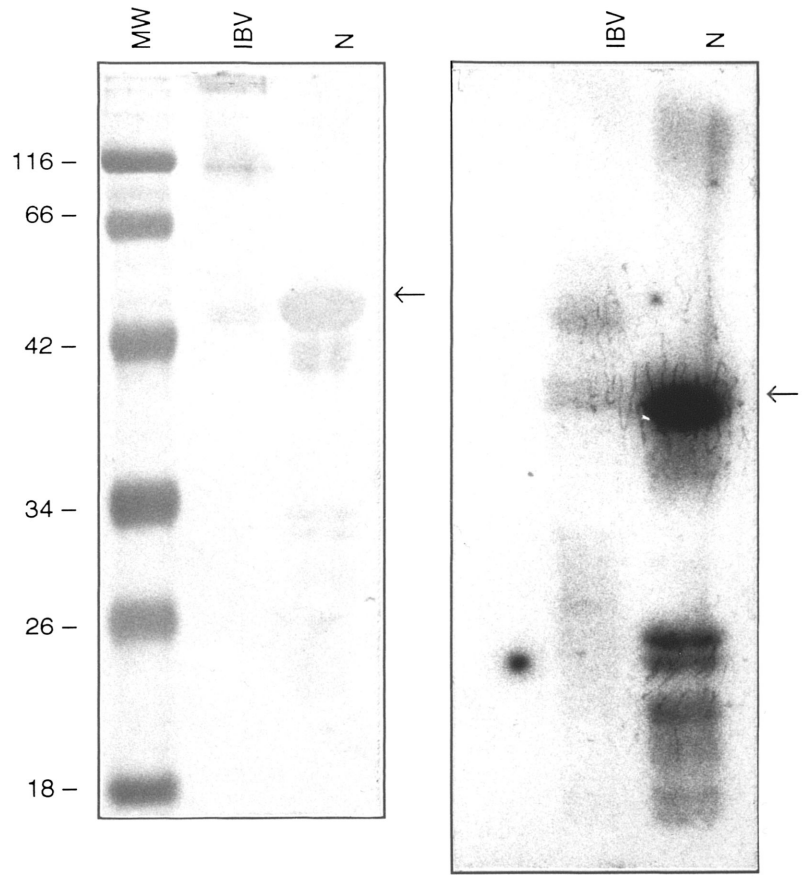


Fig. 4-6 Confirmation that purified and concentrated recombinant protein was the nucleocapsid protein by SDS-PAGE (left) and Western blot (right). Nucleocapsid protein bands are indicated by an arrow.

6, and the presence and localization of the nucleocapsid protein was confirmed by the western blot seen in Fig. 4-5b.

For future experiments, the nucleocapsid protein was concentrated with a centriprep 10 concentrator (Amicon). To ensure the protein was not lost during the concentration step, a fraction was analyzed by SDS-PAGE (Fig. 4-6a), and the presence of the nucleocapsid protein confirmed by western blot analysis (Fig. 4-6b). The recombinant nucleocapsid band also co-migrated with the band presumed to be the nucleocapsid protein in a sample of whole IBV. In addition to the nucleocapsid protein, there were also several smaller bands seen on the gel in the recombinant protein lane, but the fact that these were recognized by the anti-IBV antisera suggested that they were simply proteins produced by premature termination of translation. Since the 6 x His tag is at the amino terminal of the protein, any such premature termination products would be absorbed onto the column and thus co-purify with the full length nucleocapsid protein.

In order to determine the efficiency and specificity of the Ni-NTA affinity chromatography, another preparative scale expressicn was performed. Fractions corresponding to the flow-through, (the proteins not absorbed onto the column matrix) the washes, and the eluate following addition of 250mM imidazole, were collected and analyzed by SDS-PAGE (Fig. 4-7). The nucleocapsid protein is indicated by an arrow, and this band was clearly seen in the induced #68 before chromatography, and also in the eluate and in the #68 control (the sample already checked by western blotting in Fig. 4-6b). The lane corresponding to the flow-through contained large amounts of cellular protein, but not nucleocapsid protein, passing

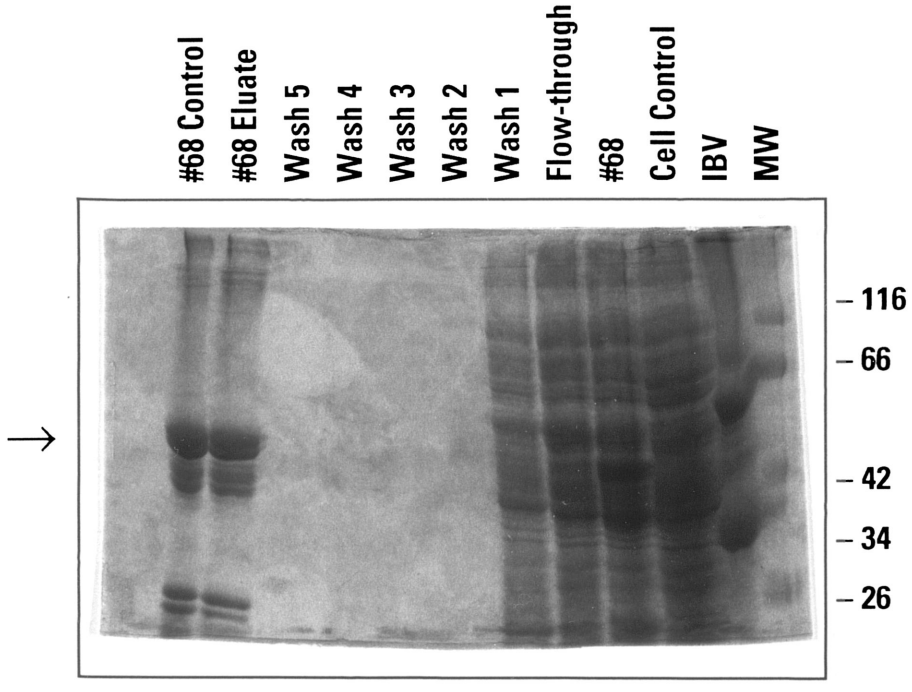


Fig. 4-7 SDS-PAGE of fractions collected from the Ni-NTA column during purification of the nucleocapsid protein. Recombinant protein is indicated by an arrow.

immediately through the column. Although some cell proteins did remain in the column, these were rapidly removed with the addition of low amounts of imidazole in the wash buffer, as these proteins had a very low affinity for the column. The later wash fractions were very clean, suggesting that any proteins remaining in the column had a high affinity for the Nickel by virtue of their 6 x his tag. This ensured that the eluate upon addition of high levels of imidazole was reasonably pure.

Discussion

Large quantities of purified protein were needed to analyse the interaction of the IBV nucleocapsid protein with RNA. For this purpose, the nucleocapsid gene was sub-cloned into the pQE8 vector, expressed in *E. coli,* and purified on a Ni-NTA column by virtue of a six residue histidine tag at the amino terminus. Analysis of the purified protein by western blot shows the sample to be contaminated only with smaller proteins presumed to be a result of premature termination of translation. As these proteins correspond to portions of the nucleocapsid protein, we presumed they would not interfere with our future experiments. However, these protein products could be eliminated by the expression of the protein in the pQE12 vector which contains the histidine tag at the carboxy terminal so that only full length protein products would be absorbed onto the column. The problem with placing a histidine tag at the carboxy terminal of the protein is that it is more likely to affect the folding of the protein and also its solubility in *E. coli* (The QIAexpressionist, QIAgen, 1992). Therefore, a histidine tag at the amino

terminal, which is recommended for purification of the protein under non-denaturing conditions, remains the method of choice for these experiments.

Because in nature the nucleocapsid protein of IBV is phosphorylated (Siddel *et al.*, 1981; Stohlman & Lai, 1979; Sturman *et al.*, 1980), future experiments should include the determination of the phosphorlyation status of the nucleocapsid protein in *E. coli.*. In the infected cell it is thought that the nucleocapsid protein is phosphorylated by a virion-associated protein kinase (Siddell *et al.*, 1981a), and although the effect of this phosphorylation on replication is unknown, it could affect the *in vitro* interaction between N and E1 in the human coronavirus JHM (Sturman *et al.*, 1980).

Although *E. coli* contains protein kinases, these probably do not specifically phosphorylate foreign proteins. In which case, if the nucleocapsid protein produced in M15/pREP4 cells is phosphorylated, it is probably due to non-specific phosphorylation which may not be the same as its phosphorylation in eukaryotic cells. However, recombinant-derived nucleocapsid protein produced in a bacterial expression system has been shown to be capable of inducing a cellular immune response to IBV, and is recognized by anti-IBV antisera, suggesting that the protein has some biological activity regardless of its phosphorylation (Boots *et al.*, 1992). This recombinant protein was then used to investigate the association of the IBV nucleocapsid protein with RNA from the 3' end of the IBV genome.

CHAPTER V

BINDING OF THE IBV NUCLEOCAPSID PROTEIN TO RNA SEQUENCES SPECIFIC FOR THE 3' END OF THE IBV GENOME

Introduction

The IBV nucleocapsid protein is a highly basic, phosphorylated protein of 409 amino acids and a relative molecular weight of approximately 50kD. A 140kD trimer of the nucleocapsid protein of MHV has been found in non-reduced preparations of purified virions suggesting the presence of disulphide linked multimers (Robbins *et al.*, 1986). The nucleocapsid protein of coronaviruses is closely associated with the RNA genome, has been shown to be involved in transcription of murine hepatitis virus (MHV) (Nakanaga *et al.*, 1986) and may also be involved in translation (Compton *et al.*, 1987). Although coronaviruses are the only positive sense RNA viruses known to have a helical nucleocapsid, several negative sense RNA viruses have helical nucleocapsids that are important in regulating viral transcription (Arnheiter *et al.*, 1985; Banerjee, 1987).

Since the IBV genome is infective and codes for polymerase function, release of the nucleocapsid into the cytoplasm of the infected cell is followed by the translation of the polymerase genes from genomic RNA (Stern & Kennedy, 1980a; Stern & Kennedy, 1980b). Two overlapping reading frames at the 5' end of the genome are capable of coding for two large polyproteins which are presumably involved in polymerase function (Boursnell *et al.*, 1987; Brierley *et al.*, 1987). Translation of the first ORF (F1),

at the 5' end, results in a putative protein of 441 residues. The second putative polyprotein encoded by this gene complex would be a fusion protein with the amino terminus being in common with the 441 residue polyprotein from the F1 ORF. The unique carboxyl 300 residues are transcribed from the F2 ORF in a -1 frame relative to the F1 (Brierley *et al.*, 1987). The F2 is translated as a result of the ribosomes slipping back one nucleotide before the F1 stop codon and thus reading through this stop codon and continuing into the F2 ORF (Brierley *et al.*, 1987; Brierley *et al.*, 1989).

A distinguishing characteristic of coronaviruses is their transcription strategy. Following primary translation, polymerase enzymes direct the transcription of a negative-stranded RNA template used for the transcription of the five subgenomic mRNAs (Baric *et al.*, 1983; Lai *et al.*, 1982b; Lai *et al.*, 1987). The subgenomic mRNAs are used as mRNAs for translation of the remaining proteins, in particular, the structural proteins required for virion assembly. It has also been demonstrated that subgenomic minus-strand RNA molecules are present in cells infected with the coronaviruses TGEV and MHV (Sethna *et al.*, 1989; Sawicki & Sawicki, 1990). These subgenomic minus-strands may also serve as template for the corresponding subgenomic mRNAs.

The unique mechanism for producing subgenomic RNAs used as messages for protein synthesis is thought to be mediated through the presence of the leader sequences found upstream of each message (Baric *et al.*, 1987; Baric *et al.*, 1985; Makino *et al.*, 1986b). The leader sequence is found only in the 5' non-coding terminus of the genome (Brown *et al.*, 1984; Boursnell *et al.*, 1989). Excess copies of free leader fragments are made

during replication, and subgenomic RNA species contain a fusion RNA with leader sequence at the 5' terminus. The leader fragments are thought to serve as primers for the transcription of the subgenomic RNAs (Lai, 1990).

Leader sequence fragments corresponding to this region are produced from the 3' end of the negative template (Lai *et al.*, 1982; Spaan *et al.*, 1983). Priming of subgenomic RNA transcription is mediated through the binding of the leader to a complementary consensus sequences, CTT(G)AACAA, found at the 5' end of each individual gene. The consensus sequence forms the junction between the body of the mRNA and the primer, that is the leader fragment (Spaan *et al.*, 1983; Baric *et al.*, 1983; Baric *et al.*, 1985; Makino *et al.*, 1986b).

The nucleocapsid protein of MHV has been found to associate with the cytosol and membrane-bound small leader RNAs, as well as to transcription complexes. Also, anti-nucleocapsid monoclonal antibodies precipitate both full length and subgenomic mRNAs as well replicative intermediate (RI) RNAs (Baric *et al.*, 1988a). Baric *et al,* (1988b) have also shown that a specific nucleocapsid protein recognition sequence exists in the small leader-containing RNAs, suggesting that a trans-acting leader RNA-N complex is involved with mRNA synthesis.

It is known that a number of copies of the nucleocapsid protein of IBV are associated with the RNA (Schochetman *et al.*, 1977; Oshiro, 1973) and, as transcription of the negative template is initiated at the 3' end of the genomic RNA, it seems possible that this is another region where the nucleocapsid protein should bind and thus play a role in transcription. In order to investigate the association of the nucleocapsid protein with the 3' end of the genome, recombinant nucleocapsid protein was produced in *E.*

coli using the pQE8 expression vector. Its association with RNA from three different regions of the 3' non-coding region was determined from northwestern blot analysis and gel-shift assays.

Materials and Methods

Production and purification of recombinant nucleocapsid protein

The nucleocapsid gene was cloned into the pQE8 expression vector, and the recombinant nucleocapsid protein purified by Ni-NTA affinity chromatography, as described in chapter IV. The concentrated protein was quantitated using the Bio Rad protein assay (Bio-Rad), and stored at -100°C in 10μl aliquots.

In vitro transcription of IBV strand-specific probes

The *in vitro* transcripts used in this study were derived from either pCR1000 (Invitrogen) or pGem3Z (Promega) plasmids, containing either the bacterial T7 (in pCR1000) or SP6 and T7 (in pGEM3Z) RNA polymerase initiation sites. Inserts were obtained by cDNA cloning of PCR products corresponding to the sequences depicted in Fig. 5-1. The PCR primers used to amplify these regions are listed in Table 5-1. Pcr was performed, and the resulting products either cloned into the PCR1000 vector by virtue of the T overhang resulting from the PCR (clones G2+, G3-, AB+, AB-, CD+, CD-), or into the pGem3Z vector following restriction enzyme digestion (clones $I^{+/-}$, $EF^{+/-}$), as described in chapter II. The PCR primers were designed with restriction enzyme recognition sequences at their 5' ends. This enabled 5μg of the resulting PCR product to be digested with 2units of either BamH1 or

Table 5-I. PCR primers used to amplify specific regions of the IBV cDNA for sub-cloning into transcription vectors. Restriction enzyme sites are shown in bold: GGATCC=BamH1, GAATCC=EcoR1, AAGCTT= HindIII.

Clone	Upstream primer	Downstream primer
G2+	**GAATTC**CCGCGTGTACCTCTCTAGTA	**GGATCC**GCTCTAACTCTATACTAGCCTAT
G3-	**GGATCC**GCTCTAACTCTATACTAGCCTAT	**GAATTC**CCGCGTGTACCTCTCTAGTA
I+	**GAATTC**CAGCTTTAGGTGAGAATGAAC	**GGATCC**AAGCTACATGCCTATCTCCC
I-	**GGATCC**AAGCTACATGCCTATCTCCC	**GAATTC**CAGCTTTAGGTGAGAATGAAC
AB+	**GAATTC**GGGAGATAGGCATGTAGCTT	**GGATCC**GCACTGGCATCTTTATACCTACT
AB-	**GGATCC**GCACTGGCATCTTTATACCTACT	**GAATTC**GGGAGATAGGCATGTAGCTT
CD+	**GGATCC**GAGTAGGTATAAAGATGCCAGTG	TAGAATACTC**AAGCTT**GCATGCCTGCCTGCAG
CD-	TAGAATACTC**AAGCTT**GCATGCCTGCCTGCAG	**GGATCC**GAGTAGGTATAAAGATGCCAGTG
EF+	**GAATTC**CGAACGGTAGACCCTTAAGATTTTC	**GGATCC**CCTAGTGGGTGTCCTAGTGTTGTA
EF-	**GGATCC**CCTAGTGGGTGTCCTAGTGTTGTA	**GAATTC**CGAACGGTAGACCCTTAAGATTTTC

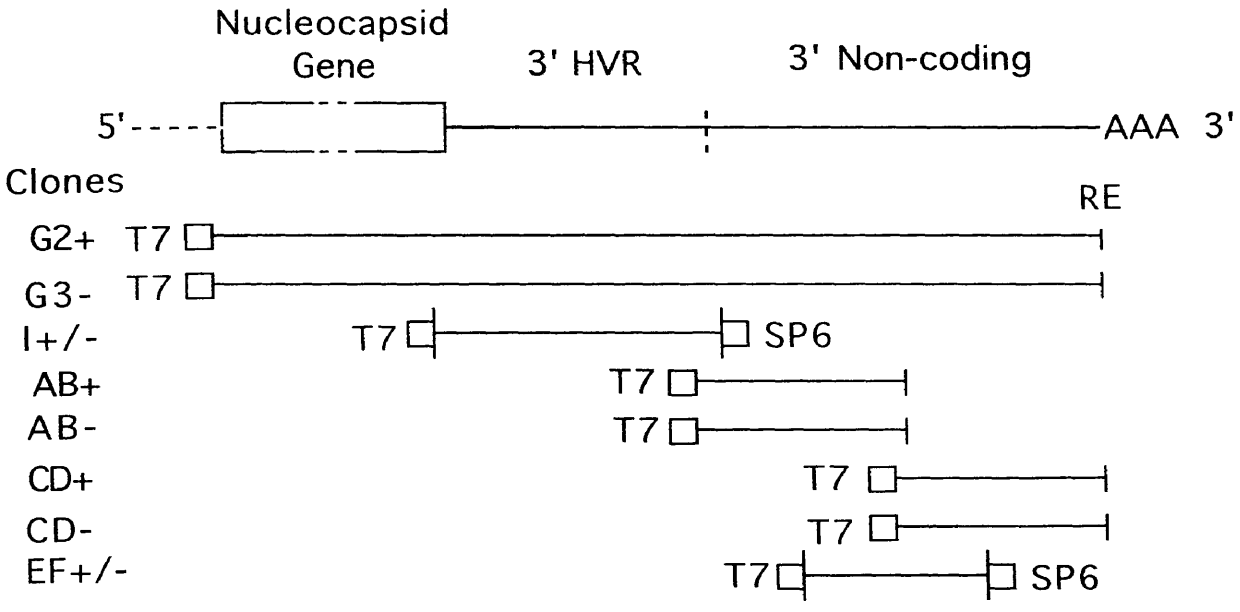


Fig. 5-1 Diagramatic representation of clones used for *in vitro* transcription of IBV-specific RNAs. T7 and SP6= recognition sequences for bacterial RNA polymerases. | = restriction enzyme recognition sites for linearization of plasmid DNA.

HindIII (Promega corporation, Madison, WI) in buffer C (50mM NaCl, 10mM Tris-HCl, pH7.9, 10mM $MgCl_2$, 1mM DTT) or EcoR1 (Promega) in buffer H (50mM NaCl, 90mM Tris-HCl, pH7.5, 10mM $MgCl_2$, 1mM DTT) at 37°C for 2h. Following the addition of 1/10 volume 10X loading buffer, the reactions were loaded onto a 0.8% low melting point agarose gel and electrophoresed at 60V to purify the digested PCR products. The DNA bands were excised and the agarose melted at 56°C in 10mM Tris-HCl, pH 7.6, 1mM EDTA, pH 8.0 (TE). Following phenol extraction, the DNA was precipitated with 0.3M NaOAc, pH5.2, and two volumes of ethanol at -20°C for 2h. Ligation reactions were set up as described in chapter IV and used to transform electroporation-competent JM109 cells (Stratagene, La Jolla, CA).

Clone G2+ contains the whole of the nucleocapsid gene and the 3' non-coding region, and G3- is complementary to G2+. $I^{+/-}$ are the positive and negative sense clones of the 3' hypervariable region. $AB^{+/-}$, $CD^{+/-}$ and $EF^{+/-}$are the positive and negative sense clones of overlapping regions of the 3' non-coding region (Fig. 5-1).

Linear templates were obtained by digestion of these clones with either BamH1 (G2, I+, AB+, CD-, EF+), EcoR1 (G3, I-, AB-, EF+) or HindIII (CD+) as previously described, and the products were collected by ethanol precipitation following phenol extraction. Transcription was carried out in 100µl of reaction buffer containing approximately 1.5µg of linearized template, 40mM Tri-HCl, pH7.5, 6mM $MgCl_2$, 2mM spermidine, 10mM NaCl, 100 units rRNasin (Promega) 0.5mM each of ATP, GTP, and UTP, 0.1mM CTP, 200µCi [α-^{32}P]CTP, and 20 units of either T7 or SP6 RNA polymerase (Promega, Madison, WI). Competitor (cold) RNA was made under the same conditions as the labelled RNA but 0.5mM CTP was used with no

radiolabelled CTP. Following transcription at 37°C for 2h, the DNA template was removed by incubation with 2.5units DNase1 for 15min at 37°C, and the RNA was extracted with an equal volume of phenol:chloroform:isoamyl alcohol and precipitated with 0.5 volume 7.5M ammonium acetate and 2 volumes ethanol at -20 overnight. Following centrifugation at 12,000 x g for 20min, the RNA was resuspended in 10µl RNAse-free dH_2O and quantitated using a spectrophotometer (Biorad). All RNAs were diluted in RNase-free dH_2O to a standard concentration of 1ng/µl.

Northwestern blot analysis

RNA sequences binding to the nucleocapsid protein were detected by northwestern blot analysis. 3.5µg of whole IBV proteins and purified nucleocapsid protein were separated by electrophoresis on 10% SDS-PAGE, and transferred to nitrocellulose as described in chapter IV. Before performing the binding assay, the blots were washed for 30min at room temperature in SBB buffer (10mM Tris, pH7, 50mM NaCl, 1mM EDTA, 0.02% bovine serum albumin, 0.02% Ficoll, and 0.02% polyvinyl pyrrolidone). Strips of the blots were placed in petri dishes containing 4ml SBB and 2 x 10^5 cpm of ^{32}P-labelled RNA (that is approximately 10ng RNA). Competitor RNA (bovine liver RNA) was added to a final concentration of 20µg/ml to remove non-specific binding. After 1h at room temperature with constant agitation, the strips were washed with three changes of SBB and dried, and the bound RNA was then visualized by autoradiography. Nucleocapsid protein was also localized on the nitrocellulose strips by western blot analysis using anti-IBV antisera (Sambrook *et al.*, 1989).

Gel-shift analysis

Specific protein-RNA interactions were also analyzed by a gel-shift assay. One ng of RNA and varying concentrations of nucleocapsid protein (0.3×10^{-2}mg/ml to 4mg/ml) were co-incubated for 20min at room temperature in 10µl of gel-shift buffer (25mM Hepes, 25mM EDTA, 150mM NaCl, 5mM DTT, 10% glycerol and 20units rRNasin). Following the addition of 1µl 10x sample buffer, the reactions were loaded onto a 0.5% agarose gel and run at 60V in 1XTBE until the bromophenol blue was approximately 1 inch from the end of the gel. The gels were placed into Ziploc bags, wrapped in aluminium foil together with a film, and visualized by autoradiography.

Competition assays were performed using a constant amount of nucleocapsid protein (4.2µg) and labelled RNA (1ng), with variable amounts of cold homologous or heterologous competitor RNA (0.4ng to 50ng). Buffer, dH_2O and RNAs were all mixed together first, before the addition of the proteins to allow the RNAs an equal opportunity to bind to the nucleocapsid protein. Cache valley virus (CVV) RNA was also used as a competitor to determine whether the nucleocapsid protein also had a high affinity for other viral RNAs. This RNA was produced from the transcription of BamH1 digested S segment cDNA in pGEM3Z, supplied by Shan-Ing Chung (Chung, 1992)

Results

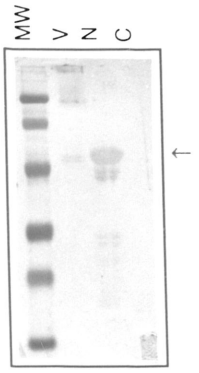
Northwestern and western blots

A northwestern assay was first used in order to determine which RNA fragments were able to associate with the IBV nucleocapsid protein. The recombinant nucleocapsid protein and IBV whole viral proteins (used as controls) were electrophoresed on SDS-PAGE and transferred to nitrocellulose. Fig. 5-2 shows the results of the northwestern and parallel western blot of these proteins. The position of the nucleocapsid protein, migrating at 50kD in both the whole virus and purified recombinant protein lanes, was verified by western blot analysis. RNA from clone G2+, corresponding to the nucleocapsid gene through to the 3' end of the IBV genome, bound to the nucleocapsid protein in the northwestern blot. However, this RNA was also found to bind to various other proteins in the virion (Fig. 5-2). For additional northwestern assays a $2x10^3$ molar excess of competitor RNA, that is 20μg/ml of bovine liver RNA, was used to compete out non-specific binding of IBV RNA to other proteins. In contrast, even at this high concentration of competitor RNA, all the positive sense clones (Fig. 5-3) and negative sense clones (Fig. 5-4) appeared to bind to the nucleocapsid protein with a greater affinity than the non-IBV RNAs.

Gel-shift analysis

A gel shift assay was developed to determine the relative affinity of each of the IBV RNAs for the nucleocapsid protein. As long as the protein concentration was relatively high, a huge complex that could not enter a 0.5% agarose gel was formed when any of the RNA fragments were bound

a



b

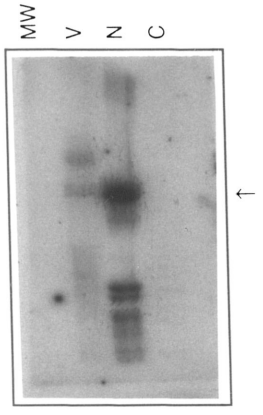


Fig. 5-2 Western blot and Northwestern blot of recombinant nucleocapsid protein. a) Western blot of whole IBV and recombinant nucleocapsid protein using polyclonal α-IBV. The nucleocapsid protein is migrating at approximately 50kD. b) Northwestern blot of whole IBV and recombinant nucleocapsid protein with ^{32}P labelled G2+ RNA as the probe.

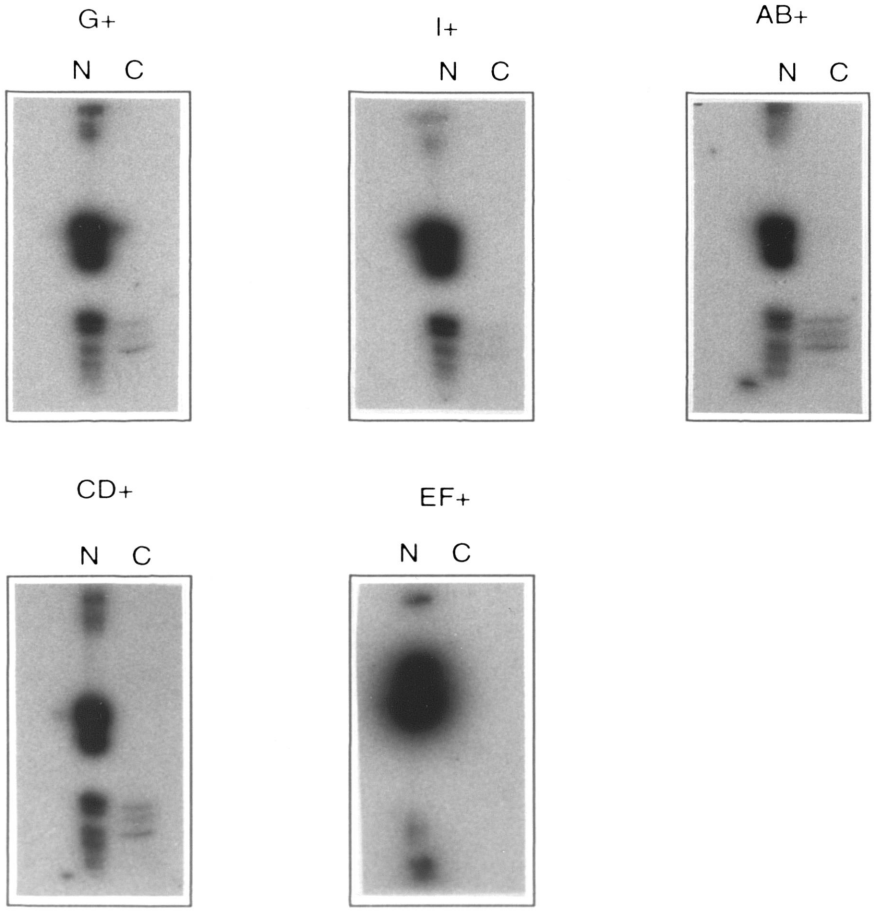


Fig. 5-3 Northwestern blot analysis of recombinant nucleocapsid protein and its binding to positive sense RNAs from the 3' end of the IBV genome. I+ RNA represents the 3' HVR, and AB+, CD+, & EF+ represent the remaining portions of the 3' non-coding region. N = Nucleocapsid, C = Cell control.

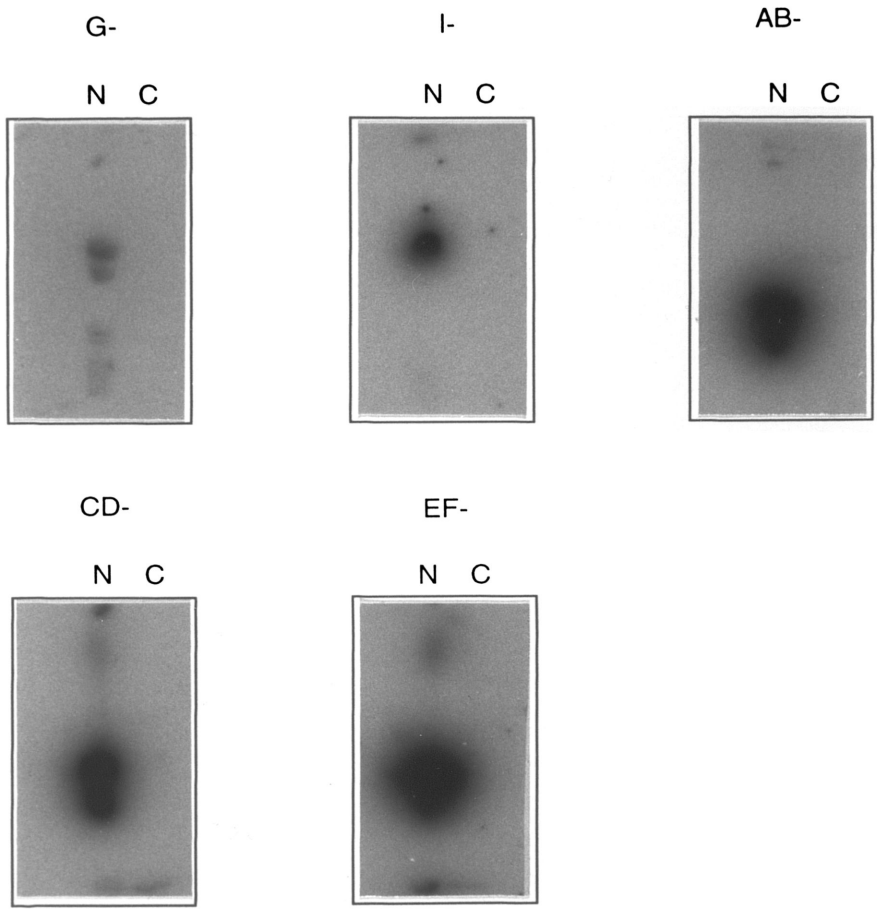


Fig. 5-4 Northwestern blot analysis of recombinant nucleocapsid protein and its binding to negative sense RNAs from the 3' end of the IBV genome. I- represents RNA complementary to the 3' HVR and AB-, CD- & EF- represent RNAs complementary to the 3' non-coding region. N = Nucleocapsid protein, C = Cell control.

to the nucleocapsid protein. However, upon titration of the protein, the complex was found to migrate into the gel as a smear, rather than as a discrete band, suggesting that either several copies of the nucleocapsid protein are able to bind to a single piece of RNA, or that once a protein molecule is bound to the RNA, other copies of the protein are then able to bind to the RNA-bound protein. Three examples of this are seen in Fig. 5-5 with RNA from clones G2+, I+ and CD+. The complexed RNA and protein can be seen remaining in the wells of the gel. As the protein concentration was decreased, less complex was observed in the wells while a smear was found in the middle of the gel and increasing amounts of free RNA at the bottom of the gel.

Competition assays were set up using the smaller positive sense RNAs (I+, AB+, CD+ and EF+) and non-IBV RNAs (CVV and bovine liver). Fig 5-6 shows gels of I+ RNA competing with other RNAs for the nucleocapsid protein. The amounts of protein and labelled I+ RNA were kept constant while increasing amounts of cold competitor RNA were added. Approximately 10 to 25 times as much cold competitor RNA from the IBV positive sense clones examined was needed to compete with the radio-labelled I+ RNA for binding to the nucleocapsid protein. That is, at these levels of competitor RNA, the majority of the labelled I+ RNA was free rather than complexed with the nucleocapsid protein. In contrast, the I+ RNA was not competed off the nucleocapsid protein by the CVV RNA or the bovine liver RNA, although the CVV RNA bound to the nucleocapsid protein in northwestern blots. Even with 50 times as much CVV or bovine liver RNA the I+ RNA remained complexed with the nucleocapsid protein. This shows a relatively high affinity of the I+ RNA for the nucleocapsid protein. A similar

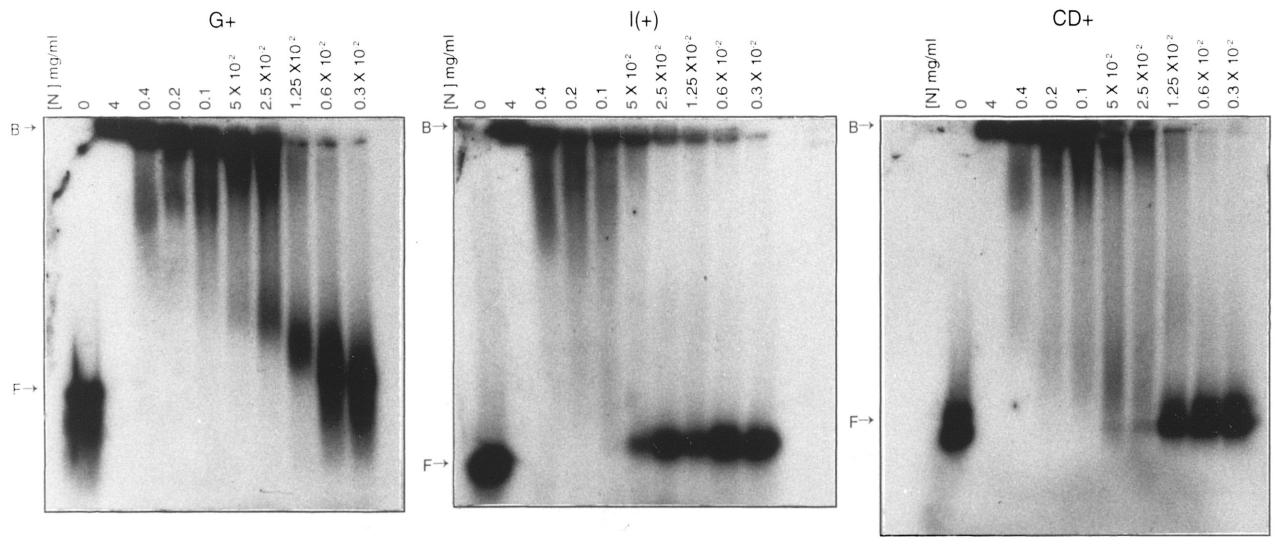


Fig. 5-5 Gel-shift assay demonstrating the binding of G+, I+, and CD+ RNAs to the nucleocapsid protein at variable protein concentrations. At high protein:RNA ratios the complex remaining in the wells is indicated by B. Free RNA is indicated by F.

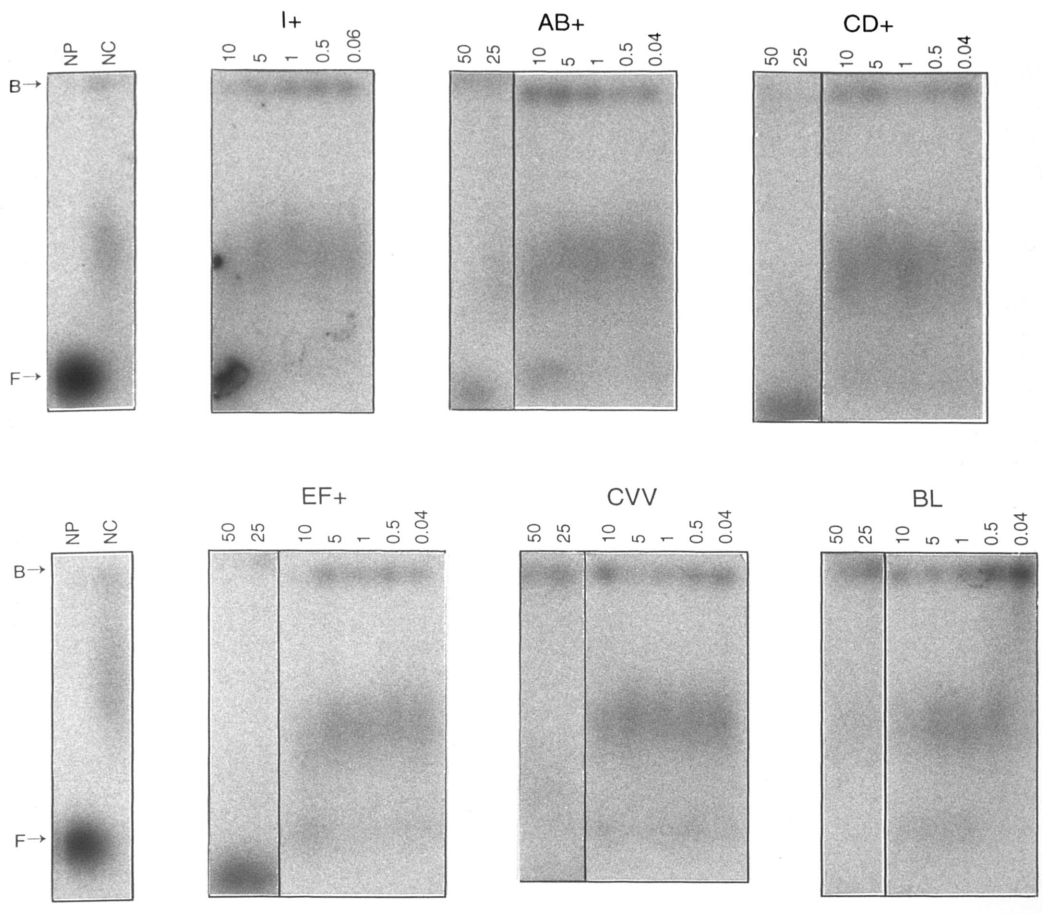


Fig. 5-6. Competition gel shift assay with ^{32}P labelled I+ RNA competing with other RNAs for nucleocapsid protein binding. Equal amounts of I+ RNA and protein were present in each lane while the cold competitor RNA concentration ranged from 0.4ng to 50ng.

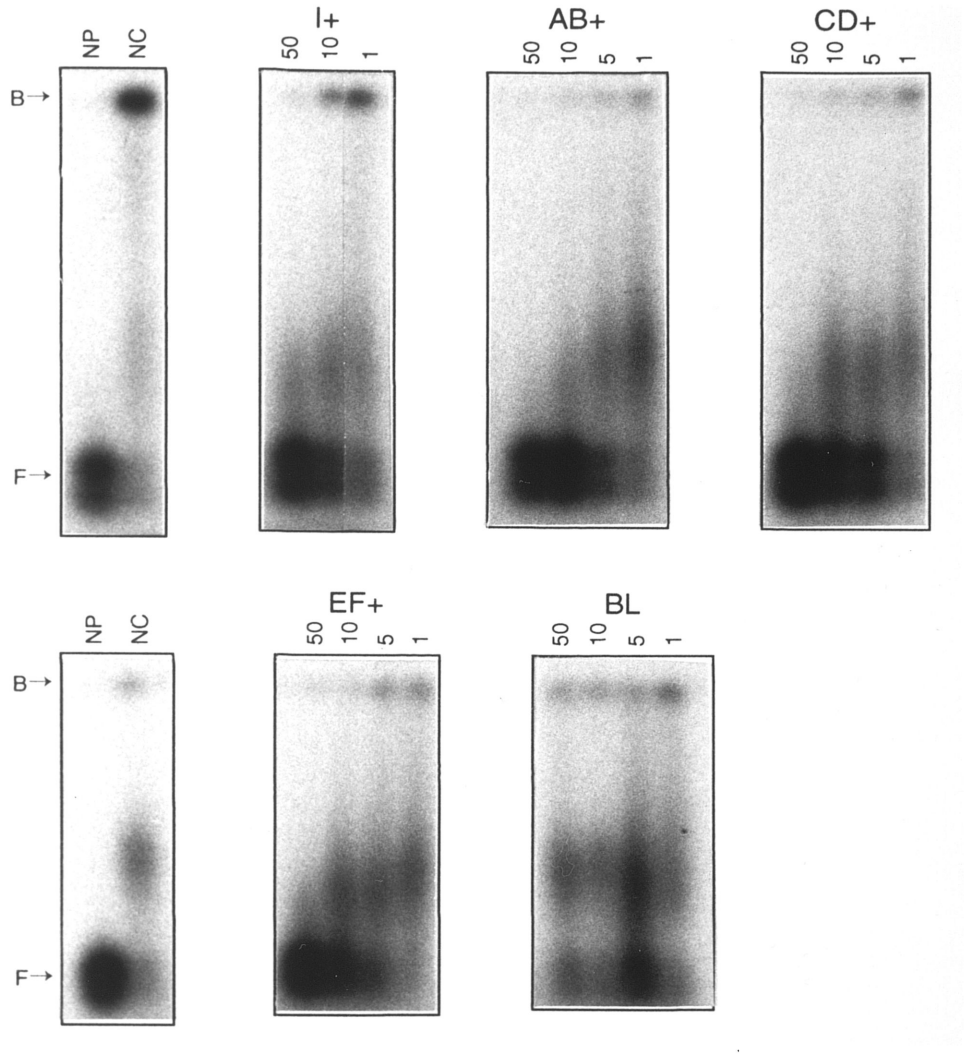


Fig. 5-7. Competition gel shift assay with CD+ RNA competing with other RNAs for nucleocapsid protein binding. Equal amounts of CD+ RNA and protein were present in each lane while the cold competitor RNA concentration ranged from 0.4ng to 50ng.

assay was performed using CD+ RNA, and this too shows a relatively high affinity of the CD+ RNA for the nucleocapsid protein, as it also is not competed off by the non-IBV RNAs, that is CVV and bovine liver RNA (Fig. 5-7). These two figures also demonstrated that within the sensitivity of the assay the IBV RNAs representing te 3' non-coding region were able to bindto the nucleocapsid protein with a similar affinity. That is, if one of the competitor RNAs had a greater affinity than the labelled RNA for the nucleocapsid protein, we would have seen the labelled protein displaced with much lower levels of competitor RNA.

Discussion

It has previously been demonstrated that the nucleocapsid protein of MHV is able to bind specifically to RNA corresponding to the leader sequence (Stohlman *et al.*, 1988). In this paper we have shown that the nucleocapsid protein is also able to bind with a relatively high affinity to RNA corresponding to the 3' end of the IBV genome. The northwestern assay was optimized using a larger clone corresponding to the nucleocapsid gene through the entire 3' non-coding region, and the gel shift assay was used to examine the protein interacting with fragments of the 3' non-coding region. We have shown that all the IBV RNAs were capable of binding to the nucleocapsid protein, and apparently with a similar affinity.

Because there are no immediately obvious consensus sequences repeated in each of these regions, the affinity of the RNA for the nucleocapsid protein may depend on the overall secondary structure of the RNA rather than a specific sequence. There are also no sequences of any

significant homology to other known protein binding sequences. The 3' HVR transcribed from clone I+ demonstrated a similar binding affinity for the nucleocapsid protein as the other IBV RNAs. It is not known if this 3' HVR was inserted into some strains or deleted from others, but it is obviously not required by all strains as it is absent in Mass41 (Boursnell *et al*, 1985). However, in spite of its distinctly different overall composition and because the 3' HVR clearly impacts the RNA secondary structure in this region of the genome, its affinity for the nucleocapsid protein supports the hypothesis that the nucleocapsid protein recognizes secondary structure rather than primary sequences.

The nature of the binding of the nucleocapsid protein to RNA is not clear, but at high protein:RNA ratios, a huge complex is formed that does not even migrate into a 0.5% agarose gel. Rather than forming discrete bands when the ratio is lowered, a smear was seen between the free and bound RNAs. This could be due to either several copies of the nucleocapsid protein binding at different positions along the RNA or specific binding of the nucleocapsid protein in a few locations, resulting in subsequent protein-protein interactions. Protein-protein interactions could result from folding of the complex bringing several proteins into close proximity to each other such that they can interact, or from co-operative binding of free nucleocapsid protein to RNA-bound protein. Co-operative binding is unlikely to be occurring as this usually results from the specific binding of protein to a nucleation site, thus enabling other proteins to bind. As the nucleocapsid protein bound to all the sequences in this study with a similar affinity, there is unlikely to be a nucleation site in this region of the genome. There is certainly a high affinity of the nucleocapsid protein for IBV RNA, but a

specific preferred region was not identified in the sequences at the 3' end of the genome. The nucleocapsid protein is presumably involved in packaging of the IBV RNA, and maybe the specific recognition of the leader sequence is important for packaging while the nucleocapsid protein associates non-specifically with the remaining RNA.

CHAPTER VI

DISCUSSION AND CONCLUSION

The nucleocapsid protein of coronaviruses is a major structural protein in the virion, plays a role in transcription and translation, and is involved in the induction of T cell mediated immune responses. (Cavanagh, 1983; Compton *et al.*, 1987 and Boots *et al.,* 1992). The nucleocapsid protein is variable among the different coronaviruses, but is generally highly conserved among strains within a species, such as IBV in chickens. The nucleocapsid protein sequences can also be used to distinguish between the major coronavirus antigenic groups (Dea *et al.*, 1990). The close correlation between antigenic grouping from serotyping and from nucleocapsid protein sequences implies that the spike and nucleocapsid proteins are evolving parallel to each other within distinct antigenic groups (Williams *et al.*, 1992).

The branching pattern of the IBV strains in the nucleocapsid protein phylogenetic tree is not exactly the same as that for the 3' non-coding region as different roots were selected for producing these trees. However, the overall grouping is similar as the Holl52 and KB8523 are shown to be more closely related to each other than to the other strains.

It is still not clear whether the 3' HVR is the result of an insertion or a deletion event. It is possible that the 3' HVR was present in the original IBV, from which the Holl52 would have been derived, and then KB8523 may have been the result of some deletions and mutations, with Ark99, Gray and Beau being produced from further deletions and/or mutations. In this case

the absence of the 3' HVR in the Mass41 strain would be explained by one large deletion or recombination event. Alternatively, Mass41 may have been derived from the original IBV strain without the 3' HVR, and either the other strains then picked up variations on this region or the Holl52 was the progenitor of those strains with this 3' HVR with various recombination or deletion events in this region leading to the emergence of the other strains. If the 3' HVR is the result of an insertion, its presence is most likely due to recombination with the host RNA. In order to verify this hypothesis a genomic library of chicken mRNA could be screened with probes from the IBV 3' HVR. Any positive clones would then have to be sequenced and compared to the 3' HVR.

Future experiments may also be performed to determine if the 3' HVR is present at this putative recombinational hot spot. Recombination might be experimentally reproduced by co-infecting chick embryo kidney (CEK) cells with the Mass41 strain of IBV, and transfecting with an expression vector containing the 5' non-coding region, the nucleocapsid gene and the 3' non-coding region of the Gray strain. Progeny virus would then be screened for the presence of the 3' HVR.

Regions of strong homology among these different nucleocapsid proteins suggest a functional importance in viral replication that even over time and evolution are very sensitive to change. It is interesting that the acidic region nearest the amino terminal of IBV includes two regions which are also very highly conserved among all the coronaviruses. The high degree of conservation of the nucleocapsid protein could also make it a valuable diagnostic probe.

In contrast to the high level of conservation of the nucleocapsid protein, the adjacent 3' region is highly variable. A high level of recombination has been reported during the replication of murine coronaviruses (Lai, 1992), and it is likely that the 3' HVR was inserted into the 3' non-coding region of many IBV strains as a result of this recombination. The location of the 3' HVR is conserved at a position immediately downstream from the stop codon of the nucleocapsid gene but there is no evidence as yet to show that this HVR has any effect on virulence or pathogenesis. In contrast, a deletion of approximately 240 amino acids from the S protein of TGEV is thought to be responsible for the emergence of PRCV. This deletion resulted in the absence of two antigenic sites in PRCV isolates and is the basis for their differentiation from the enteropathogenic TGEV (Sanchez *et al.*, 1992). However, the sequences flanking the 3' end of the IBV HVR are similar to the tetrameric (TTCC) or heptameric (AGTTTCC) repeats found within the regions deleted from the PRCVs. It is possible that the sequences contained in the 3' HVR are not essential for replication, but they may play an important role in determining the secondary structure of this region of the genome and thus influence its association with the replication machinery.

It has previously been demonstrated that the nucleocapsid protein of coronaviruses is able to bind specifically to RNA corresponding to the leader sequence (Stohlman *et al.*, 1988). The present studies have shown that the nucleocapsid protein is also able to bind with high affinity to RNA corresponding to the 3' end of the IBV genome. All the positive and negative sense IBV-derived RNAs used in this study were capable of binding to the

nucleocapsid protein, and apparently all the positive sense IBV RNAs had similar affinity for the nucleocapsid protein.

The promiscuous affinity of the nucleocapsid protein for IBV oligonucleotides, and the absence of obvious repeated consensus sequences in each of these regions, suggests that it is recognizing an overall secondary structure in the RNA, or sequences with a similar compostion, rather than a specific sequence. Additionally, there are no sequences of any significant homology to other known protein binding sequences, supporting our hypothesis that the protein is recognizing the RNA secondary structure.

Future experiments should address the potential for phosphorylation of a recombinant nucleocapsid protein produced in *E. coli*, because this may well influence the binding specificity of the nucleocapsid protein with certain RNA sequences. For example, it has been shown that the removal of a phosphate group from the avian retrovirus nucleocapsid protein pp12 decreases its affinity for viral RNA 100-fold (Fu *et. al.,* 1985). Experiments such as site-directed mutagenesis of the nucleocapsid protein gene could help to define regions important in the binding of the protein to RNA. Shorter RNA sequences should also be examined to determine whether the protein binds to specific residues or to a particular secondary structure.

The nucleocapsid protein plays many important roles in the coronavirus life cycle. It is an integral part of the replication cycle of the virus, presumably through its interaction with the RNA, and maybe also through association with other viral proteins. Therefore, further investigation of the association of the nucleocapsid protein with IBV RNA should help us to understand more about the coronavirus replication cycle, and ultimately aid

in development of anti-viral drugs and better control of IBV associated disease.

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.

Almeida, J. D., Berry, D. M., Cunningham, C. H., Hamre, D., Hofstad, M. S., Mallucci, L., McIntosh, K., and Tyrrell, D. A. J. (1968). Coronaviruses. *Nature* **220**, 650.

Arnheiter, H., Davis, N. L., Wertz, G., Schubert, N., and Lazzarini, R. A. (1985). Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell* **41**, 259-267.

Banerjee, A. (1987). Transcription and replication of rhabdoviruses. *Microbiol. Rev.* **51**, 66-87.

Banner, L. R., Keck, J. G., and Lai, M. M. C. (1990). A clustering of RNA recombination sites adjacent to a hypervariable region of the peplomer gene of murine coronavirus. *Virology* **175**, 548-555.

Baric, R. S., Fu, K., Schaad, M. C., and Stohlman, S. A. (1990). Establishing a genetic recombination map for murine coronavirus strain A59 complementation groups. *Virology* **177**, 646-656.

Baric, R. S., Nelson, G. W., Fleming, J. O., Deans, R. J., Keck, J. G., Casteel, N., and Stohlman, S. A. (1988). Interactions between coronavirus nucleocapsid protein and viral RNAs: Implications for viral transcription. *J. Virol.* **62**(11), 4280-4287.

Baric, R. S., Shieh, C.-K., Stohlman, S. A., and Lai, M. M. C. (1987). Analysis of intracellular small RNAs of mouse hepatitis virus: evidence for discontinuous transcription. *Virology* **156**, 342-354.

Baric, R. S., Stohlman, S. A., Razavi, M. K., and Lai, M. M. C. (1985). Characterization of leader-related small RNAs in coronavirus-infected cells: further evidence for leader-primed mechanism of transcription. *Virus Res.* **3**, 19-33.

Baric, R. S., Stohlman, S. A., and Lai, M. M. C. (1983). Characterization of replicative intermediate RNA of mouse hepatitis virus: presence of leader RNA sequences on nascent chains. *J. Virol.* **48**(3), 633-640.

Beaudette, F. R. and Hudson, C. B. (1937). Cultivation of the virus of infectious bronchitis. *J. Am. Vet. Med. Ass.* **90**, 51-60.

Berry, D. M., Cruickshank, J. G., Chu, H. P., and Wells, R. J. H. (1964). The structure of infectious bronchitis virus. *Virology* **23**, 403-407.

Boots, A. M. H., Benaissa-Trouw, B. J., Hesselink, W., Rijke, E., Schrier, C., and Hensen, E. J. (1992). Induction of anti-viral immune responses by

immunization with recombinant-DNA encoded avian coronavirus nucleocapsid protein. *Vaccine* **10**(2), 119-124.

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 recognizing the nucleocapsid protein of avian coronavirus IBV. *Immunol.* **72**, 10-14.

Boots, A. M. H., Kusters, J. G., Van Noort, J. M., Zwaagstra, K. A., Rijke, E., van der Zeijst, B. A. M., and Hensen, E. J. (1991). Localization of a T-cell epitope within the nucleocapsid protein of avian coronavirus. *Immunol.* **74**, 8-13.

Boursnell, M. E. G., Binns, M. M., Brown, T. D. K., Cavanagh, D., and Tomley, F. M. (1989). Molecular biology of avian infectious bronchitis virus. *Prog. Vet. Microbiol. Immunol.* **5**, 65-82.

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.

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., and Brown, T. D. K. (1984). Sequencing of coronavirus IBV genomic RNA: a 195 base open reading frame encoded by mRNA B. *Gene* **29**, 87-92.

Brierley, I., Digard, P., and Inglis, S. C. (1989). Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**, 537-547.

Brierley, I., Boursnell, M. E. G., Binns, M. M., Bilimoria, B., Block, V. C., Brown, T. D. K., and Inglis, S. C. (1987). An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J.* **6**, 3779-3785.

Brown, T. D. K., Boursnell, M. E. G., and Binns, M. M. (1984). A leader sequence is present on mRNA A of avian infectious bronchitis virus. *J. Gen. Virol.* **65**, 1437-1442.

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, Rauischholzhausen, Germany" pp147-153.

Cavanagh, D. (1983). Coronavirus IBV: structural characterization of the spike protein. *J. Gen. Virol.* **64**, 2577-2583.

Cavanagh, D. (1981). Structural polypeptides of coronavirus IBV. *J. Gen. Virol.* 53, 93-103.

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

Chung, S-I. (1992). Molecular and pathogenesis studies of cache valley virus: a common teratogenic agent in sheep. PhD Dissertation, Texas A&M University.

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.

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**, 123-128.

Cumming, R. B. (1963). infectious avian nephrosis (uraemia) in Australia. *Aust. Vet. J.* **39**, 145-147.

Cunningham, C. H. (1970). Avian infectious bronchitis. *Adv. Vet. Sci. Comp. Med.* **14**, 105-148.

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.

Davies, H. A. and Macnaughton, M. R. (1979). Comparison of the morphology of three coronaviruses. *Arch. Virol.* **59**, 25-33.

Dea, S. and Garzon, S. (1991). Identification of coronaviruses by the use of indirect protein A-gold immunoelectron microscopy. J. *Vet. Diagn. Invest.* **3**, 297-305.

Dea, S., Verbeek, A. J., and Tijssen, P. (1990). Antigenic and genomic relationships among turkey and bovine enteric coronaviruses. *J. Virol.* **64**(6), 3112-3118.

Fu, X., Phillips, N., Jentoft, J., Tuazon, P. T., Traugh, J. A., and Leis, J. (1985). Site-specific phosphorylation of avian retrovirus nucleocapsid protein pp12 regulates binding to viral RNA. *J. Biol. Chem.* **260**(17), 9941-9947.

Hopkins, S. R. (1974). Serologic comparisons of strains of infectious bronchitis using plaque-purified isolants. *Avian Dis.* **18**(2), 231-239.

Johnson, R. B. and Marquardt, W. W. (1976). Strains of infectious bronchitis virus on the Delmarva Peninsula and in Arkansas. *Avian Dis.* **20**, 382-386.

Johnson, R. B. and Marquardt, W. W. (1975). The neutralizing characteristics of strains of infectious bronchitis virus as measured by the constant virus variable serum method in chicken tracheal cultures. *Avian Dis.* **19**, 82-90.

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" pp203-209.

Kamahora, T., Soe, L. H. and Lai, M. 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 the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. *Virology* **151**, 41-49.

Keck, J. G., Matsushima, G. K., Makino, S., Fleming, J. O., Vannier, D. M., Stohlman, S. A., and Lai, M. M. C. (1988). In vivo RNA-RNA recombination of coronavirus in mouse brain. *J. Virol.* **62**, 1810-1813.

Keck, J. G., Soe, L. H., Makino, S., Stohlman, S. A., and Lai, M. M. C. (1988). RNA recombination of murine coronaviruses: recombination between fusion-positive mouse hepatitis virus A59 and fusion-negative mouse hepatitis virus 2. *J. Virol.* **62**(6), 1989-1998.

King, D.J. and Cavanagh, D. (1991). Infectious bronchitis. In: "Diseases of Poultry, ed. B.W. Calnek *et al.* Iowa State Press, Ames, Iowa" pp. 471-484.

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

Kusters, J. G., Jager, E. J., Niesters, H. G., and van der Zeijst, B. A. M. (1990). Sequence evidence for RNA recombination in field isolates of avian coronavirus infectious bronchitis virus. *Vaccine* **8**, 605-608.

Lai, M. M. C. (1992). RNA recombination in animal and plant viruses. *Microbiol. Rev.* **56**(1), 61-79.

Lai, M. M. C. (1990). Coronavirus: organization, replication, and expression of genome. *Ann. Rev. Microbiol.* **44**, 303-333.

Lai, M. M. C., Makino, S., Soe, L. H., Shieh, C. K., Keck, J. G., and Fleming, J. O. (1987). Coronavirus: A jumping RNA transcription. *Cold Spring Harbor Symp. Quant. Biol.* **52**, 359-365.

Lai, M. M. C., Patton, C. D., and Stohlman, S. A. (1982). Further characterization of mRNA's of mouse hepatitis virus: presence of common 5' end nucleotides. *J. Virol.* **41**(2), 557-565.

Lai, M. M. C., Patton, C. D. and Stohlman, S. A. (1982). Replication of mouse hepatitis virus: Negative-stranded RNA and replicative form RNA are of genome length. *J. Virol.* **44**(2), 487-492.

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.

Lomniczi, B. (1977). Biological properties of avian coronavirus RNA. *J. Gen. Virol.* **36**, 531-533.

Lutticken, D., Rijke, E. O., Loeffen, T., and Hesselink, W. (1988). Aspects of local immune response to IBV. In: "Proceedings of the 1 international symposium on Infectious Bronchitis. Rauischholzhausen, Germany" pp 171-181.

Makino, S., Keck, J. G., Stohlman, S. A., and Lai, M. M. C. (1986a). High-frequency RNA recombination of murine coronaviruses. *J. Virol.* **57**(3), 729-737.

Makino, S., Stohlman, S. A. and Lai, M. M. C. (1986b). Leader sequences of murine coronavirus mRNAs can be freely reassorted: evidence for the role of free leader RNA in transcription. *Biochem.* **83**, 4204-4208.

Nakanaga, K., Yamanouchi, K., Fujiwara, K. (1986). Protective effect of monoclonal antibodies on lethal mouse hepatitis virus infection in mice. *J. Virol.* **59**, 168-171.

Oshiro, L.S. (1973). Coronaviruses. In: "Ultrastructure of Animal Viruses and Bacteriophages: An Atlas, ed. A.J. Dalton and F. Haguenau. Academic Press, New York" pp. 331-343.

Otsuki, K., Yamamoto, H., and Tsubokura, M. (1979a). Studies on avian infectious bronchitis virus (IBV). I. Resistance of IBV to chemical and physical treatments. *Arch. Virol.* **60**, 25-32.

Otsuki, K., Maeda, J., Yamamoto, H., and Tsubokura, M. (1979b). Studies on avian infectious bronchitis virus (IBV). III. Interferon induction by and sensitivity to interferon of IBV. *Arch Virol* **60**, 249-255.

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.

Robbins, S.G., Frana, M.F., McGowen, J.J., Boyle, J.F., and Holmes, K.V. (1986). RNA-binding proteins of coronavirus MHV: detection of monomeric and multimeric N protein with an RNA overlay-protein blot assay. *Virology* **150**, 402-410.

Sambrook, S., Fritsch, E. F., and Maniatis, T. (1989). "Molecular cloning- a laboratory manual. 2nd edition. Cold Spring Harbor Laboratory Press".

Sanchez, C. M., Gebauer, F., Sune, C., Mendez, A., Dopazo, J., and Enjuanes, L. (1992). Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. *Virology* **190**, 92-105.

Sawicki, S. G. and Sawicki, D. L. (1986). Coronavirus minus-strand RNA synthesis and effect of cyclohexamide on coronavirus RNA synthesis. *J. Virol.* **57**(1), 328-334.

Schalk, A. F. and Hawn, M. C. (1931). An apparently new respiratory disease of baby chicks. *J. Am. Vet. Med. Assoc.* **78**, 413-422.

Schochetman, G., Stevens, R.H., and Simpson, R.W. (1977). Presence of infectious polyadenylated RNA in the coronavirus avian bronchitis virus. *Virology* **77**, 772-782.

Sethna, P. B., Hung, S.-L. and Brian, D. A. (1989). Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proc. Natl. Acad. Sci. USA* **86**, 5626-5630.

Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Sudha, W., and Baltimore, D. (1980). Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: Implications for retrovirus integration. *Proc. Natl. Acad. Sci. USA* **77**(7), 3932- 3936.

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.

Siddell, S. G., Barthel, A., and Ter Meulen, V. (1981). Coronavirus JHM: a virion-associated protein kinase. *J. Gen. Virol.* **52**, 235-243.

Smith, A. R., Boursnell, M. E. G., Binns, M. M., Brown, T. D. K., and Inglis, S. C. (1990). Identification of a new membrane-associated polypeptide specified by the coronavirus infectious bronchitis virus. *J. Gen. Virol.* **71**, 3-11.

Sneed, L. W., Butcher, G. D., Parr, R., Wang, L., 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.

Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B. A. M., and Siddell, S. G. (1983). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO J.* **2**, 1839-1844.

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.

Stern, D.F. and Sefton, B.M. (1982). Synthesis of coronavirus mRNAs: kinetics of inactivation of infectious bronchitis virus RNA synthesis by UV light. *J. Virol.* **42**(3), 755-759.

Stern, D. F. and Kennedy, S. I. T. (1980a). Coronavirus multiplication strategy. I. Identification and characterization of virus-specified RNA. *J. Virol.* **34**(3), 665-674.

Stern, D. F. and Kennedy, S. I. T. (1980b). Coronavirus multiplication strategy. II. Mapping the avian infectious bronchitis virus intracellular RNA species to the genome. *J. Virol.* **36**(2), 440-449.

Stohlman, S. A., Baric, R. S., Nelson, G. N., Soe, L. H., Welter, L. M., and Deans, R. J. (1988). Specific interaction between coronavirus leader RNA and nucleocapsid protein. *J. Virol.* **62**(11), 4288-4295.

Stohlman, S. A. and Lai, M. M. C. (1979). Phosphoproteins of murine hepatitis viruses. *J. Virol.* **32**(2), 672-675.

Sturman, L. S. and Holmes, K. V. (1983). The molecular biology of coronaviruses. *Adv. Virus Res.* **28**, 35-112.

Sturman, L. S., Holmes, K. V., and Behnke, J. (1980). Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. *J. Virol.* **33**(1), 449-462.

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. *Virology* **165**, 589-595.

van der Most, R. G., Heijnen, L., Spaan, W. J. M., and deGroot, R. J. (1992). Homologous RNA recombination allows efficient introduction of site-specific mutations into the genome of coronavirus MHV-A59 via synthetic co-replicating RNAs. *Nuc. Acids Res.* **20**(13), 3375-3381.

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 infectious peritonitis virus and immunogenicity of recombinant vaccinia viruses 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.

Williams, A. K., Wang, L, Sneed, L. W., and Collisson, E. W. (1992). Comparative analyses of the nucleocapsid genes of several strains of infectious bronchitis virus and other coronaviruses. *Virus Res.* **25**, 213-222.

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