

To the Graduate Council:

I am submitting herewith a dissertation written by Sushma E. Abraham entitled "Molecular Characterization of the Genes for the Spike Protein and Four Putative Nonstructural Proteins of Bovine Coronavirus." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Microbiology.

David A. Brian

David A. Brian, Major Professor

We have read this dissertation
and recommend its acceptance:

Jerry P. Weir

N. S. Rye

Beth C. Mulline

Accepted for the Council:

Curminkel

Vice Provost
and Dean of The Graduate School

MOLECULAR CHARACTERIZATION OF THE GENES FOR THE
SPIKE PROTEIN AND FOUR PUTATIVE NONSTRUCTURAL
PROTEINS OF BOVINE CORONAVIRUS

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Sushma E. Abraham

December 1989

This dissertation is dedicated to my grandmothers,
E. Abraham and E. Mathew, and to my grandaunt, Dr. Liza
Chacko, whose lives have been a source of inspiration to me.

ACKNOWLEDGEMENTS

I take this opportunity to express my utmost appreciation to my advisor Dr. David Brian for his constant guidance, support, encouragement, patience and friendship. Thanks Doc for everything!

I wish to thank my committee members: Dr. Stuart Riggsby, Dr. Jerry Weir and Dr. Beth Mullin for their valuable discussions, suggestions and personal interest during the course of this study.

I wish to thank Prakash Rao Sista and Nirupama Deshmane Sista who were instrumental in encouraging me to come to this University and for their continued support and encouragement.

I wish to thank my friends: Anil Joshi, Ajit Kamath, Phiroze Sethna, Manjiri Sethna, Arun Sharma, Andrew Ogram, Sujata DasGupta and Madhuri Narkhede who have been extremely tolerant and supportive towards me.

Lastly, I wish to express my gratitude to my lab mates both past and present: Brenda Hogue, Linda Shockley, Frank Tung, Selena Hung, Tom Kienzle, Huaiyu Mi, Joanne Maki, Phiroze Sethna, Martin Hofmann, Sally Mitchell and Chris Chambers for their friendship, laughter and above all for a great working environment.

ABSTRACT

Bovine coronavirus possesses four structural proteins, the spike, hemagglutinin-esterase, matrix and nucleocapsid proteins. The genes for the structural proteins all map within the 3' proximal 9 kilobases of a 20 kilobase, single stranded RNA genome. The genes for the hemagglutinin-esterase, matrix and nucleocapsid proteins have been cDNA cloned, physically mapped on the genome and sequenced. In this study, five kilobases of the genome mapping between the hemagglutinin-esterase and matrix protein genes and encoding genes for the spike and four potential nonstructural proteins was characterized at the molecular level.

A partial cDNA genomic library was made and the spike protein gene was identified and sequenced. The spike protein gene is 4,089 bases, lies on the 3' side of the hemagglutinin-esterase protein gene, and predicts a protein of 150,825 daltons. An internal lys-arg-arg-ser-arg-arg sequence predicts a protease cleavage site between amino acid 768 and 769 and amino terminal amino acid sequencing of the virion derived 100 kilodalton subunit confirmed the location of this site. The virion 120 and 100 kilodalton spike subunits were therefore established to be the amino- and carboxy-terminal subunits. RNA hybridization analyses

using a sequence specific probe demonstrated that the spike protein is encoded on mRNA 3.

Between the spike and matrix protein genes were found four open reading frames that encode in order 5' to 3', potential nonstructural proteins of 4.9, 4.8, 12.7 and 9.5 kilodaltons. A consensus sequence CYAAAC that is thought to play a role in coronavirus transcription precedes the open reading frames for the 4.9, 12.7 and 9.5 kilodalton proteins and predicts that transcripts from each of these sites would be made. RNA hybridization analyses using sequence specific probes confirmed this prediction and established the three transcripts to be mRNAs 3a, 4, and 5 respectively. Translation of in vitro synthesized transcripts of the bovine coronavirus 9.5 kilodalton protein gene was done in an attempt to clarify differences that were observed between the bovine coronavirus and the antigenically related mouse hepatitis coronavirus with respect to the expression of this protein.

TABLE OF CONTENTS

CHAPTER	PAGE
1 LITERATURE REVIEW	1
A. MOLECULAR AND CELLULAR BIOLOGY OF CORONAVIRUSES	1
Introduction	1
Coronaviruse Genome	8
Coronavirus Proteins	
a. Structural Proteins	8
b. Nonstructural Proteins	13
Adsorption, Penetration and Uncoating of Coronaviruses	14
RNA Replication, Transcription and Translation in Coronaviruses	16
Assembly and Release of Coronaviruses	22
B. FOCUS OF THE RESEARCH DISCUSSED IN THIS DISSERTATION	24
Bovine Coronavirus.	24
C. RATIONALE FOR THE STUDY OF THE SPIKE PROTEIN GENE AND POTENTIAL NONSTRUCTURAL PROTEIN GENES OF BCV.	25
2. GENERATION OF A PARTIAL cDNA LIBRARY OF THE BCV GENOME	29
Introduction	29
Materials and Methods	29
Cells and Virus	29
Growth and Purification of Virus.	30
RNA Purification	30
cDNA Cloning of the BCV Genome	31
Plasmid Transformation of E. coli JM83	32
Identification of Colonies Containing Viral Inserts	33
Isolation and Size Screening of Plasmids with BCV specific Inserts	34
Results and Discussion	34

CHAPTER

PAGE

3.	DEDUCED SEQUENCE OF THE BOVINE CORONAVIRUS SPIKE PROTEIN AND IDENTIFICATION OF THE INTERNAL PROTEOLYTIC CLEAVAGE SITE	40
	Introduction	40
	Materials and Methods	42
	cDNA cloning of BCV Genomic RNA .	42
	DNA Sequencing	43
	RNA Extraction for Northern Blot Analysis	43
	Electrophoresis and Northern Blot Analysis of BCV RNA	47
	Amino Acid Sequencing of the Subunits of the Spike Protein .	49
	Results	49
	Deduced Amino Acid Sequence of the Spike Protein	49
	Terminal Amino Acid Sequencing Confirms the Predicted Internal Protease Cleavage Site and Identifies the Virion 120K and 100K Subunits as S1 and S2, Respectively	59
	S mRNA is the Fourth Largest Virus Specific mRNA in Infected Cells	59
	Discussion	62
4.	SEQUENCE AND EXPRESSION ANALYSIS OF PUTATIVE NONSTRUCTURAL PROTEINS OF 4.9K, 4.8K, 12.7K, AND 9.5K ENCODED BETWEEN THE SPIKE AND MATRIX PROTEIN GENES OF THE BOVINE CORONAVIRUS	70
	Introduction	70
	Materials and Methods	71
	Generation of Deletion Clones of MA7 for Sequencing the Region Between the Spike and Matrix Protein Genes	71
	End Labelling of the DNA for Maxam and Gilbert Sequencing .	74
	DNA Sequencing	74
	Preparation of RNA from BCV Infected Cells for Northern Analysis	74

CHAPTER	PAGE
4. (Continued)	
Generation of Specific Probes for Northern Analysis	75
Construction of pGEM Recombinant Vectors	76
In Vitro Transcription.	78
In Vitro Translation	78
Immunoprecipitation	79
Immunofluorescence	80
Results	81
Nucleotide Sequence Between the Spike and the Matrix Protein Genes of BCV	81
Analysis of the Open Reading Frames Potentially Encoding 4.9K and 4.8K Proteins	84
Analysis of the Open Reading Frame Potentially Encoding a 12.7K Protein	86
Analysis of the Open Reading Frame Potentially Encoding a 9.5K Protein	89
Separate mRNAs are Made for the 4.9K, 12.7K and 9.5K Open Reading Frames in BCV Infected Cells	92
In Vitro Translation Analysis of the BCV 9.5K Protein Demonstrates that it is Not Made From Downstream (Internal) Initiation of Translation . . .	96
Discussion	101
Speculation on the Functions of Nonstructural Proteins of Coronaviruses	105
REFERENCES	108
APPENDIX	124
Codon Usage for the BCV Spike Protein . . .	124
Amino Acid Composition of the BCV Spike Protein	125
Codon Usage for the Putative BCV 4.9K Protein.	126

CHAPTER	PAGE
APPENDIX (continued)	
Amino Acid Composition of the Putative BCV 4.9K Protein	127
Codon Usage for the Putative BCV 4.8K Protein.	128
Amino Acid Composition of the Putative BCV 4.8K Protein.	129
Codon Usage for the Putative BCV 12.7K Protein.	130
Amino acid Composition of the Putative BCV 12.7K Protein	131
Codon Usage for the Putative BCV 9.5K Protein	132
Amino Acid Composition of the Putative BCV 9.5K Protein	133
Preparation of Miniprep Plasmid DNA For Sequencing With Sequenase	134
VITA	136

LIST OF TABLES

TABLE		PAGE
1.1	Characteristics of RNA virus families . . .	2
1.2	Coronaviruses and their related diseases . .	3
1.3	Antigenic relationships among coronaviruses	5
1.4	Consensus intergenic sequences of MHV-A59, BCV, FIPV, TGEV, and IBV	20
2.1	Estimated sizes of cDNA clones generated with random or specific primers	36
3.1	Subclones used to sequence the spike protein gene	44
3.2	Primers used to sequence the spike protein gene	45
3.3	Features of coronavirus spike proteins obtained from deduced amino acid sequences .	55
4.1	Transcripts encoded between the spike and matrix protein genes in BCV and MHV	95

LIST OF FIGURES

FIGURE	PAGE
1.1 Schematic representation of the structure of BCV and TGEV	7
1.2 Gene map of the genomes of TGEV, BCV, MHV and IBV	15
1.3 Model for leader primed RNA transcription. .	19
1.4 A model summarizing the events in the replication, transcription and assembly of a representative coronavirus, mouse hepatitis virus	23
2.1 Gene map of the first 8,955 bases of the BCV genome and position of cDNA clones . . .	39
3.1 Gene map of the BCV genome and the sequencing strategy for the spike protein gene	46
3.2 Nucleotide sequence of the S gene and its deduced amino acid sequence	51
3.3 Hydrophobicity plot of the BCV spike protein	56
3.4 Northern blot analysis identifying mRNA3 as the transcript from the S gene	60
3.5 Comparison of the amino acid sequence of BCV and MHV A59 spike proteins by diagonal plotting	64
3.6 Comparison of the amino acid sequence of BCV and MHV JHM spike proteins by diagonal plotting	65
3.7 Structural comparison of S proteins of MHV JHM, MHV A59 and BCV	66
4.1 Map positions of deletion clones of cDNA clone MA7 generated to sequence the region between the spike and matrix protein genes .	73

FIGURE	PAGE
4.2 Schematic representation of recombinant plasmid AB6	77
4.3 Schematic representation of recombinant plasmid AB9.6	77
4.4 Nucleotide sequence and deduced amino acid sequence of the region between the spike and matrix protein genes of BCV	82
4.5 Map position of the consensus intergenic sequences in the 3' terminal 3.5 kb of the BCV genome	85
4.6 Hydrophobicity plot of the putative BCV 4.9K protein	87
4.7 Hydrophobicity plot of the putative BCV 4.8K protein	88
4.8 Hydrophobicity plot of the putative BCV 12.7K protein	90
4.9 Hydrophobicity plot of the putative BCV 9.5K protein	91
4.10 Northern blot analysis identifying separate mRNAs for the 4.9K, 12.7K and 9.5K open reading frames	93
4.11 Surface and internal fluorescence of the 9.5K protein in BCV infected cells	97
4.12 In vitro translation of the 12.7K and 9.5K proteins from in vitro generated transcripts, using T7 polymerase in wheat germ extracts	100

CHAPTER 1

LITERATURE REVIEW

A. MOLECULAR AND CELLULAR BIOLOGY OF CORONAVIRUSES

Introduction

Coronaviruses belong to one of 16 families of RNA viruses (Fenner, 1987) (Table 1.1). They were first recognized and morphologically defined as a group by Tyrrell and coworkers (1968; 1975; 1978). The prefix corona was chosen because of the crown-like appearance of projections on the surface of the virions (Tyrrell, 1968). Morphologically, coronaviruses can be described as spherical, medium sized enveloped particles measuring 100-150 nm in diameter (Roseto, 1982).

There are now sixteen recognized members in the family and most are known to cause acute diseases of the gastroenteric or respiratory tracts. Table 1.2 lists the coronaviruses, their natural hosts and the diseases they cause. It can be seen from Table 1.2 that coronaviruses infect a wide range of vertebrate animals and cause a variety of diseases in multiple organ systems. The members of the coronavirus family earlier had been grouped into four subgroups on the basis of their antigenic relatedness (Wege, 1982). The avian coronaviruses formed two

a

Table 1.1 CHARACTERISTICS OF RNA VIRUS FAMILIES

Family	Genome ^b		Envelope	Site of ^c virus budding
	Nature	Size(kb)		
Picornaviridae	ss,1,(+)	7.5-8.5	-	-
Calciviridae	ss,1,(+)	7.9	-	-
Togaviridae	ss,1,(+)	12	+	PM
Flaviviridae	ss,1,(+)	12	+	ER,Golgi
Orthomyxoviridae	ss,8,(-)	13.6	+	PM
Paramyxoviridae	ss,1,(-)	18-20	+	PM
Coronaviridae	ss,1,(+)	20-30	+	ER,Golgi
Arenaviridae	ss,2,(-)	10-14	+	PM
Bunyaviridae	ss,3,(-)	14-17	+	Golgi
Retroviridae	ss,1,(+)	2X(3-9)	+	PM
Rhabdoviridae	ss,1,(-)	13-16	+	PM
Reoviridae	ds,10-12	17-22	-	ER
Filoviridae	ss,1,(-)	12.7	+	PM
Birnaviridae	ds,2	7	-	-
Toroviridae	ss,1,(+)	20	+	?
Astroviridae	ss,1,(?)	8	-	-

^a Adapted from Dubois-Dalcq et al., 1984 and Fenner et al., 1987

^b All molecules linear although circular forms of togaviruses have been found; ss=single-stranded; ds=double-stranded; 1 to 12=number of molecules in genome; (+) or (-)=sense of single stranded nucleic acid.

^c ER=Endoplasmic reticulum; PM=Plasma membrane

Table 1.2 CORONAVIRUSES AND THEIR RELATED DISEASES^a

Common Name of Virus	Designation	Natural Host	Disease
Avian infectious bronchitis virus	IBV	Chicken	Tracheobronchitis nephritis, oviduct hypoplasia
Turkey coronavirus	TCV	Turkey	Enteritis (blue comb disease)
Mouse hepatitis virus	MHV	Mouse	Hepatitis, encephalomyelitis, enteritis, interstitial pneumonia
Rat coronavirus	RCV	Rat	Pneumonia, Rhino-tracheitis
Sialodacryoadenitis virus	SDAV	Rat	Sialodacryo- adenitis, kerato- conjunctivitis, rhino-tracheitis
Transmissible gastro- enteritis virus	TGEV	Pig	Gastroenteritis
Hemagglutinating encephalomyelitis virus	HEV	Pig	Encephalomyelitis, gastroenteritis (vomiting and wasting disease)
Porcine epidemic diarrhoea virus	PED	Pig	Enteritis
Bovine coronavirus	BCV	Bovine	Gastroenteritis
Human coronavirus	HCV	Human	Common cold
Human enteric coronavirus	HEV	Human	Gastroenteritis
Canine coronavirus	CCV	Dog	Gastroenteritis
Feline infectious	FIPV	Cat	Meningo- encephalitis, panophthalmitis, peritonitis, pleuritis, pneumonia, wasting disease, vasculitis (immune complex disease), disseminated granulomatous disease.
Feline enteric coronavirus	FECV	Cat	Enteritis
Pleural effusion disease virus	RbCV	Rabbit	Pleuritis, myocarditis
Rabbit enteric coronavirus	RbECV	Rabbit	Enteritis

^a Adapted from Sturman and Holmes, 1983.

subgroups and the mammalian coronaviruses formed two subgroups. Recently, however, Dea et al., (1989) (personal communication) have shown that the turkey coronavirus is antigenically related to the bovine coronavirus and probably belongs to one of the mammalian subgroups and is no longer a separate subgroup (Table 1.3).

Most of the information that is available on the molecular and cellular biology of coronaviruses is derived from studies on the infectious bronchitis virus (IBV) and mouse hepatitis virus (MHV). These two viruses have been intensely studied and information obtained from these studies is used to illustrate the major features of coronavirus structure and replication. Some recent fundamental observations have been made with the porcine transmissible gastroenteritis virus (TGEV), however, and these will be discussed below. It is therefore assumed that basic features described for one coronavirus are universal to all. Some viruses have features that are not common to all coronaviruses. For example, the bovine coronavirus (BCV) (King and Brian, 1982), human coronavirus OC43 (Hogue and Brian, 1986), porcine hemagglutinating encephalomyelitis virus, diarrhoea virus of infant mice (Sugiyama et al., 1986) are found to contain the hemagglutinin-esterase glycoprotein which is

Table 1.3 ANTIGENIC RELATIONSHIPS AMONG CORONAVIRUSES^a

<u>Group 1</u>	<u>Group 2</u>
Human Coronavirus 229E	Mouse Hepatitis Virus
Porcine Transmissible Gastroenteritis Virus	Rat coronavirus
Canine Coronavirus	Sialodacryoadenitis Virus of Rats
Feline Infectious Peritonitis Virus	Human Coronavirus OC43
	Bovine Coronavirus
	Hemagglutinating Encephalomyelitis Virus of swine
	Turkey Coronavirus ^b
<u>Group 3</u>	
Infectious Bronchitis Virus	

^a adapted from Wege et al., 1982. Based on the serological techniques of neutralization, crossimmunofluorescence and crossimmunoprecipitation.

^b Dea et al., personal communication (IV International Coronavirus Meeting, Cambridge, England).

unique to this subgroup of coronaviruses. Figure 1.1 shows a schematic representation of the structure of two coronaviruses (BCV and TGEV) presently studied in our laboratory.

Coronaviruses grow in a wide variety of cells and their infection is often but not always, accompanied by cytopathic changes, leading to disintegration or in some cases syncytium formation (Tyrrell et al., 1978). There have been no quantitative studies on host cell DNA, RNA or protein synthesis in infected cells, but MHV does exhibit a strong inhibition of host cell protein synthesis (Hilton et al., 1986). The mechanisms for this inhibition are not known. BCV, on the other hand, causes very few cytopathic effect and 80-90% of the cells survive the acute infection to become persistently infected (a carrier culture). Replication like that for other positive strand RNA viruses takes place in the cytoplasm of infected cells. This conclusion is supported by (1) electron microscopic and immunofluorescence studies which show coronavirus proteins in the cytoplasm (McIntosh et al., 1974), (2) growth of coronaviruses in the presence of Actinomycin D and inhibitors of DNA metabolism (McIntosh, 1974), and, (3) replication of MHV JHM in enucleated cells (Brayton et al., 1981).

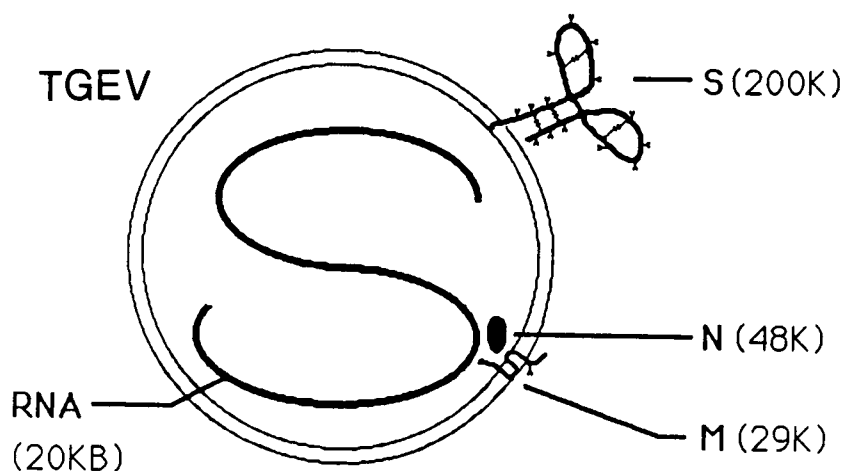
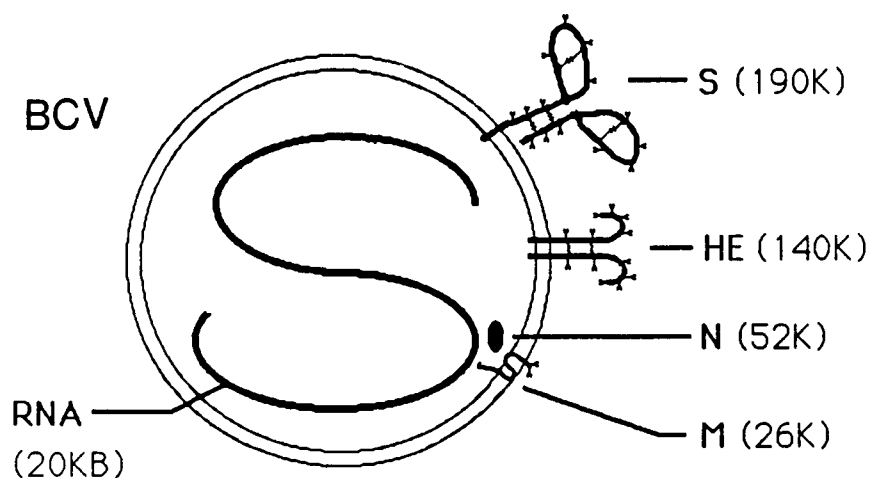


Fig 1.1 Schematic representation of structure of the bovine coronavirus (BCV) and porcine transmissible gastroenteritis virus (TGEV). The data for BCV structure is from King and Brian (1982) and TGEV structure from Moreau (Masters thesis, 1982). Four major structural proteins are S=Spike; HE=hemagglutinin-esterase; M=matrix; N=Nucleocapsid. The genome is a single stranded, nonsegmented 3'polyadenylated RNA.

Coronavirus genome

The genome of coronaviruses is a single-strand RNA molecule with a molecular mass of $5.4-6.9 \times 10^6$ daltons, that was originally estimated to be 16-20 kb in length (Siddell et al., 1982; Sturman and Holmes, 1983). The genome of IBV has been completely sequenced and is 27 kb in length (Boursnell, 1987). Mapped clones of MHV genome suggest that it is 34 kb in length (Pachuk et al., 1989). This makes the coronavirus genome the largest single stranded RNA genome described for any virus.

The genomic RNA has been shown to be infectious for IBV, MHV and TGEV. Therefore coronaviruses are plus strand RNA viruses (Baltimore, 1971; Sturman and Holmes, 1983). The 5'end of the genomic RNA is capped and the 3'end is polyadenylated (Schochetman et al., 1977; Guy and Brian, 1979; Sturman and Holmes, 1983; Siddell, 1982a).

Coronavirus proteins

a. Structural proteins

Three major structural proteins are present on the virion of all coronaviruses, the nucleocapsid (N) protein, the matrix (M) glycoprotein and a large spike or peplomer (S) protein (Stern and Sefton 1982a; Siddell, 1982). A subset of coronaviruses, those that hemagglutinate,

contain a fourth glycoprotein which is the hemagglutinin-esterase protein (HE) (King and Brian, 1982; Hogue et al., 1986; Sugiyama et al., 1986)

Nucleocapsid protein

The 50K-60K nucleocapsid protein (N) is a basic, phosphorylated protein (Stohlman and Lai, 1979) and is found associated with the positive stranded genomic RNA. Cloning and sequencing of the N gene of MHV A59 (Armstrong et al., 1983), MHV JHM (Skinner et al., 1983), IBV (Bourisnell et al., 1985b), TGEV (Kapke and Brian, 1986) and BCV (Lapps et al., 1987) has made it possible to deduce the primary structure of the nucleocapsid protein. Recent work has shown the involvement of the nucleocapsid protein in transcription. Antibodies to the nucleocapsid protein were shown to inhibit transcription in vitro (Compton et al., 1987). More recently Stohlman et al, (1988) reported the presence of N binding sites on MHV leader RNA which is part of the transcription complex. However the exact nature of the interaction between N and the transcription complex is not known. The binding of N to leader RNA probably has an important function in viral assembly since leader RNA sequences are present at the 5' end of the genome.

Matrix protein

The matrix protein (M) is a 26K-28K integral membrane glycoprotein. The gene for the matrix protein has been cloned and sequenced for many of the coronaviruses including MHV A59 (Armstrong et al., 1984), MHV JHM (Pfleiderer et al., 1986), IBV (Bourisnell et al., 1984), BCV (Lapps et al., 1987) and TGEV (Laude et al., 1987; Kapke et al., 1987). In infected cells M protein is found to accumulate within the rough endoplasmic reticulum (RER) and Golgi apparatus (Holmes, 1981b). It has been suggested that the matrix protein determines the site of virus budding (Tooze et al., 1984). The M proteins of MHV, IBV and BCV are interesting membrane proteins in that they do not possess an N-terminal signal sequence but they require the signal recognition particle for membrane insertion (Rottier et al., 1984, 1985). Internal membrane spanning domains have been found to function as signal sequences in these proteins (Machamer and Rose, 1987). However the absence of an N-terminal signal peptide is not universal among coronaviruses since a putative N-terminal signal peptide has been identified on the TGEV M protein that does not appear to be an absolute requirement for membrane translocation of the TGEV M protein (Laude et al., 1987; Kapke et al., 1988). Glycosylation of the matrix protein of BCV (Lapps et al., 1987) and MHV (Holmes et al., 1981a; Rottier et al., 1981b; Niemann et al., 1984)

differs from that of most viral glycoproteins in that glycosylation of the M protein is not inhibited by tunicamycin suggesting the presence of O-linked sugars. In IBV and TGEV, however, N-linked sugars are present on the M protein (Cavanagh et al., 1983a; Jacobs et al., 1986).

Spike or Peplomer protein

The spike (S) protein is 150K-200K and forms the large peplomers or spikes on the virion envelope which give coronaviruses their characteristic crowned appearance. The gene for the peplomer protein has been cloned and sequenced in MHV A59 (Luytjes et al., 1987), MHV JHM (Schmidt et al., 1987), TGEV (Rasschaert et al., 1987; Tung, PhD disseration, 1987), and feline infectious peritonitis virus (FIPV) (deGroot et al., 1987). The deduced amino acid sequence of the S protein from the nucleotide sequence, showed the presence of a putative signal sequence at the N-terminus and a hydrophobic domain (potential membrane anchor) at the C-terminus characterisitic of membrane proteins (Garoff, 1985). The S protein is present on the virion as cleaved subunits in IBV (Cavanagh et al., 1983a, b; Stern and Sefton, 1982b), MHV (Sturman et al., 1985) and BCoV (King and Brian, 1982). However cleavage is not seen in TGEV (Rasschaert et al., 1987), FIPV (deGroot et al., 1987) and canine coronavirus (CCV) (Garwes et al., 1981b). N-linked

glycosylation is present on the S protein and is required for virus infectivity (Holmes et al., 1981b). The protein is acylated by palmitic acid at or near the hydrophobic domain (Schmidt, 1982). Acylation is postulated to have a role in membrane anchorage (Schlesinger, 1981) and is a common feature of envelope glycoproteins including the VSV G protein, hemagglutinin (HA) of influenza, fusion (F) protein of Newcastle disease virus, and hemagglutinin/neuraminidase (HN) protein of Simian virus 5 (Veit et al., 1989).

The S protein mediates the binding of the virions to receptors on the plasma membrane and is responsible for the induction of at least some of the neutralizing antibodies (Garwes et al., 1978; Deregt et al., 1987; 1989a, 1989b). A majority of the S protein is retained in the RER and incorporated into virus particles at this site, although some of it does get transported to the cell surface of infected cells which makes the cells susceptible to cytotoxic T cells (Haspel et al., 1981). It is also found to induce cell fusion thereby promoting virus entry and dissemination (Sturman and Holmes 1983).

Hemagglutinin-esterase protein

The hemagglutinin-esterase protein (HE) is 140K and is present in BCoV (King and Brian, 1982; King et al., 1985; Deregt et al., 1987), porcine hemagglutinating

encephalomyelitis virus (HEV), human coronavirus OC43 (HCV OC43) (Hogue and Brian, 1986) and diarrhoea virus of infant mice (Sugiyama, 1986). It is present on the virion as a disulfide linked homodimer of 65K subunits in BCV and has been shown to be involved in the binding of the virus to the cell (Vlasak et al., 1988a). It has been shown to stimulate neutralizing antibodies (Deregt et al., 1987; 1989). The HE protein possesses acetyl esterase activity very similar to that observed in influenza C viruses which cleaves acetyl residues from sialic acid on cellular receptors (Vlasak et al., 1988a; Vlasak et al., 1988b). The acetyl esterase activity (receptor destroying activity) may facilitate virus entry or release and may also promote the spread of the virus by inactivating its own receptor.

b. Nonstructural proteins of coronaviruses

The genome of coronaviruses also encodes potential nonstructural proteins. However, very little is known about the number and function of these proteins. The RNA dependent RNA polymerase is one of the nonstructural proteins that has been identified in TGEV and MHV infected cells (Dennis and Brian, 1981, 1982; Brayton et al., 1982; Mahy, 1983). The gene for the RNA dependent RNA polymerase has been identified in IBV (Bournsnell et al., 1987) and in MHV A59 (Pachuk et al., 1989). A number of open reading frames have been identified between the spike

and matrix protein genes of coronaviruses that can encode putative nonstructural proteins or proteins that form minor structural components of the virion (Figure 1.2). mRNAs that potentially encode these proteins are found to function as polycistronic mRNAs (MHV mRNA 5 and mRNA D of IBV) unlike the mRNAs for the structural proteins which function as monocistronic mRNAs (Budzilowicz et al., 1987; Skinner et al., 1985; Bournsnel et al., 1985a). A 15K protein in MHV JHM infected cells (Ebner et al., 1988), a 9.6K protein in MHV A59 infected cells (Leibowitz et al., 1988) and a 12.4K protein in IBV infected cells (Smith, 1987) have been detected as products of some of these genes.

Adsorption, penetration and uncoating of coronaviruses

Relatively very little is known about the early events in coronavirus replication. The binding of coronaviruses to cells appears to be via the spike and hemagglutinin-esterase glycoproteins, since antibodies to these proteins inhibit virus infectivity (Garwes, 1978; Deregt et al., 1987; Deregt et al., 1989a, b). Although detailed studies of coronavirus penetration and uncoating have not been performed, several studies suggest two possible mechanisms. Coronavirus uptake in some studies appears to be via the the endocytic pathway (Patterson, 1976) as with

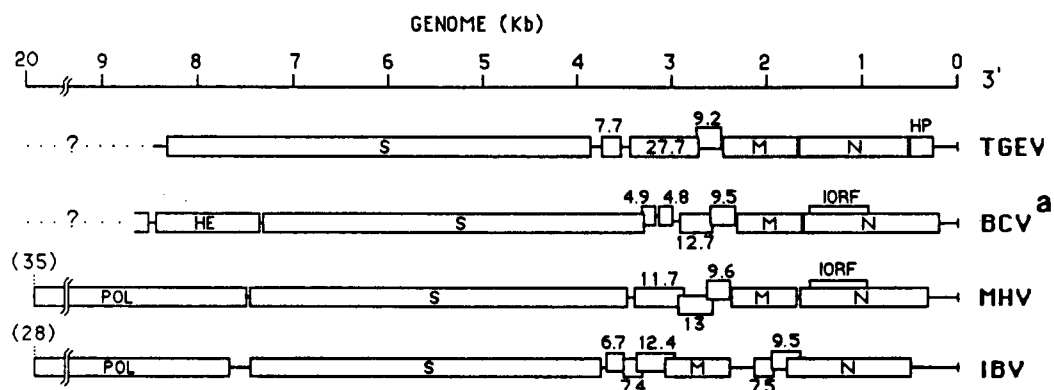


Fig 1.2 Gene map of the genome of porcine transmissible gastroenteritis virus (TGEV), bovine coronavirus (BCV), mouse hepatitis virus (MHV), infectious bronchitis virus (IBV). The open boxes represent open reading frames. N=nucleocapsid, M=matrix, S=spike, HE=hemagglutinin-esterase, POL=RNA polymerase. The numbers represent the predicted molecular weights of the potential nonstructural proteins encoded by the open reading frames.
^a open reading frames between HE and M of BCV were identified in this study.

Semliki forest virus (Helenius, 1980a) and vesicular stomatitis virus (Helenius, 1980b). In other studies coronaviruses appear to enter the cell by fusion with the cell membrane (Sturman et al., 1983).

RNA replication, transcription and translation in coronaviruses

RNA dependent RNA polymerase activity in cells infected with TGEV and MHV was first detected by Dennis and Brian (1981, 1982) and Mahy et al, (1983), respectively. Two different peaks of polymerase activity in MHV infected cells was reported (Brayton et al., 1984). The early polymerase activity was postulated to result in the synthesis of the negative strand antigenome RNA (Brayton et al., 1984; Lai et al., 1982b). The negative strand was then postulated to serve as the template for full length positive strand genomic RNA and subgenomic mRNA synthesis using the late RNA polymerase. The pattern of negative strand RNA being first made by the early polymerase activity and the positive strand RNA being made by the late polymerase activity has been disputed by Sawicki and Sawicki (1986). They have shown that both positive and negative strand RNA are detected early in infection, after which the synthesis of the negative strand declines and positive strand synthesis continues.

Stern and Kennedy (1980a, 1984) showed that IBV infection resulted in the production of a nested set of subgenomic mRNAs with common 3'ends. T1 oligonucleotide mapping demonstrated that the intracellular RNA 1 (Stern and Kennedy, 1980b), the largest RNA, corresponded to the genomic RNA of virions and that all intracellular subgenomic mRNAs formed a nested set. That is, the nucleotide sequence of any given message is contained within the 3'end of the next larger message. The 3'end of all the mRNAs is the same. The T1 oligonucleotide data also showed that the genomic RNA was of the same polarity as the subgenomic mRNAs since they shared the same oligonucleotides (Lai et al., 1981; Spaan et al., 1982). The 3' coterminal set of mRNAs was later shown to be present in MHV JHM and BCoV (Keck et al., 1988), by using specific cDNA probes from the 3'end of the genome. The mRNAs are polyadenylated at the 3'end and capped at the 5' end. The mRNAs have been designated 1 through 7 beginning with genomic RNA as 1 and the smallest RNA as 7 in BCoV (Keck et al., 1988).

An unusual feature of coronavirus mRNAs is the presence of approximately 70 nucleotides (leader RNA) at the 5'end of each mRNA and the genomic RNA (Lai et al., 1983, 1984; Spaan et al., 1988). Presently, there is no evidence that leader RNA on each mRNA is derived by conventional splicing. Baric et al, (1983, 1985, 1987), have provided

evidence which suggests that mRNAs are made by a leader-primed transcription mechanism. This model proposes that a leader RNA species is synthesized from the 3'end of the negative strand RNA antigenome, detaches from the template and later binds to the same or different templates at the initiation sites of various subgenomic mRNAs (Lai, 1986; Lai et al., 1987). The leader RNA then serves as a primer for mRNA transcription. A schematic representation of this model is shown in Figure 1.3. Free leader has been shown by Baric et al., (1985) to be present in the cytoplasm of infected cells. There is sequence similarity of 7-18 nucleotides that occurs between the 3'end of the leader RNA and the initiation sites of the various mRNAs. Several initiation sites have been identified by comparing the sequences of the 5'ends of the mRNAs and the corresponding regions of the genome (Spaan et al., 1983; Bredenbeek et al., 1987). Initiation sites have also been recognized by homology searches and their positions on the genome (Bournsnell et al., 1984; Binns et al., 1985; Bournsnell et al., 1985b; Bredenbeek et al., 1986, 1987; Budzillowicz and Weiss, 1985; Kapke and Brian, 1986; Lapps et al., 1987; Schmidt et al., 1987; Rasschaert et al., 1987; Luytjes et al., 1987). These initiation sites or intergenic sequences are found to be repeated at the boundaries of genes. Table 1.4 shows the consensus

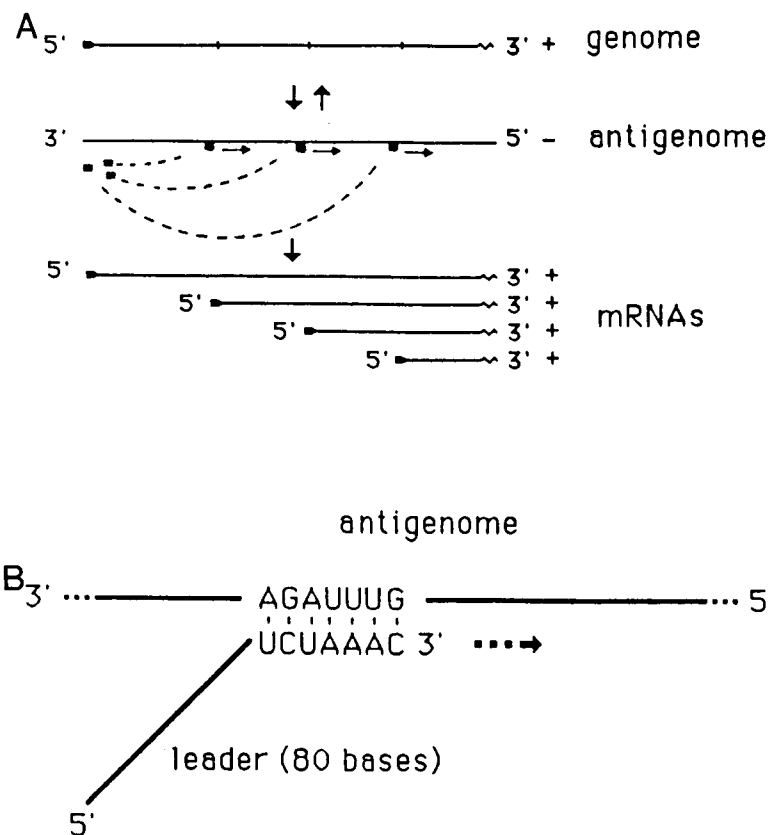


Fig 1.3 A. Model for leader primed RNA transcription. Leader RNA (-) is transcribed from the 3' end of the antigenome (negative strand), detaches, and with the RNA polymerase, binds to different consensus intergenic sequences on the antigenome and initiates transcription of the mRNAs. Leader RNA is present on the 5' end of the genome and on mRNAs.

B. Intergenic site (enlarged) at which leader RNA binds on the antigenome to initiate mRNA synthesis.

Table 1.4 CONSENSUS INTERGENIC SEQUENCES OF MHV-A59, BCV, FIPV, TGEV, IBV^a

MHV-A59	AUCYAAAC
BCV	AUCYAAAC
FIPV	AACYAAAC
TGEV	AACYAAAC
IBV	CUUAACAA

^asequences on the genome involved in the initiation of subgenomic mRNA synthesis in MHV (mouse hepatitis virus), BCV (bovine coronavirus), FIPV (feline infectious peritonitis virus), TGEV (porcine transmissible gastroenteritis virus), IBV (infectious bronchitis virus) (Spaan et al., 1988).

intergenic sequences involved in the initiation of synthesis of subgenomic mRNAs of MHV, BCV, FIPV, TGEV and IBV.

Expression of the genome of coronaviruses involves the production of eight mRNAs in BCV, six in MHV and TGEV and five in FIPV and IBV (review by Spaan et al., 1988). mRNA coding assignments for the structural proteins of MHV was established by cell free translation experiments by Siddell et al, (1981, 1983), by Leibowitz and Weiss (1982), and by translation in *Xenopus laevis* oocytes by Rottier (1981a). These experiments demonstrated that the mRNAs though structurally polycistronic were functionally monocistronic. Only the 5' open reading frame unique to the specific mRNA was translated. All the subgenomic mRNAs, as mentioned above, have 5' and 3' termini the same as the genome, and one fundamental question is why they do not undergo replication the same as genome. This question assumes that the 3' terminal untranslated sequence on the positive strand, and the 3' terminal antileader sequence on the negative strand contain the polymerase recognition signals for replication. Recently Sethna et al, (1989) have shown a negative strand counterpart for each subgenomic mRNA in TGEV-infected cells and a similar observation has been made in BCV-infected cells (M. Hofmann, unpublished

data). These negative strand subgenomic mRNAs have been postulated to have a role in mRNA amplification and the presence of mRNA replicons has been hypothesized.

Assembly and release of coronaviruses

Figure 1.4 summarizes the events in genome replication, transcription, translation and assembly of MHV. Coronaviruses mature by budding into the lumen of the rough endoplasmic reticulum (RER) or Golgi cisternae. The nucleocapsid protein is found associated with the genomic RNA to form the helical nucleocapsid. Stohlman et al., (1988) have recently shown a specific interaction of the nucleocapsid protein with the leader RNA, however no encapsidation signals have yet been identified. Virions form when nucleocapsids align near intracellular membranes containing virus glycoproteins (Massalski et al., 1981; Tooze et al., 1984). The phosphorylated nucleocapsid protein is associated with cellular membranes via strong noncovalent interaction presumably through the M protein which is found as an integral protein in the RER (Stohlman et al., 1983; Sturman et al., 1980). The S protein is evidently not required for virus assembly and release since MHV virions devoid of the S protein are released from cells treated with tunicamycin which inhibits the synthesis of S protein by inhibiting glycosylation (Holmes et al., 1981b). Virions are released from membranes by pinching off into

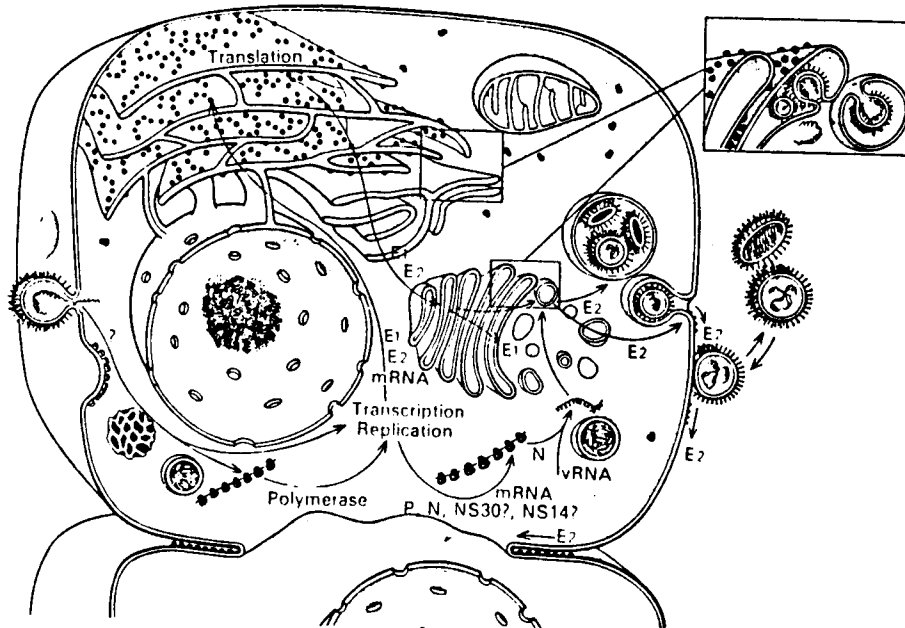


Fig 1.4 A model summarizing the events in the replication, transcription and assembly of a representative coronavirus, mouse hepatitis virus. E1, matrix protein; E2, spike protein; N, nucleocapsid protein; NS, nonstructural proteins; vRNA, viral RNA (Taken from Dubois-Dalcq et al., 1984).

the lumen of the RER or Golgi apparatus (Tooze et al., 1984; Dubois-Dalcq and Holmes, 1984) and large numbers of virions are found to accumulate within the lumen of the RER. Following their release into the RER and Golgi, coronavirions appear to escape from the cell within smooth walled vesicles that migrate to the cell membrane and fuse with it to release the virions externally (Oshiro, 1971). Budding from the endoplasmic reticulum has also been observed for Bunyaviruses, Flaviviruses and Rotaviruses. Table 1.1 shows the site at which virus budding takes place in different RNA virus families.

B. FOCUS OF THE RESEARCH DISCUSSED IN THIS DISSERTATION

The subject of this dissertation is a detailed analysis of a 5 kb region of the BCV genome that lies between the matrix protein and the hemagglutinin protein genes.

Specifically this study accomplishes the following:

1. Cloning and sequence analysis of the spike protein gene.
2. Sequence analysis and characterization of the region between the spike protein gene and the matrix protein gene which potentially encodes four nonstructural proteins.

Bovine coronavirus (BCV)

BCV was first isolated and characterized from a calf with diarrhoea (Mebus et al., 1978). It remains one of

the major causes of neonatal bovine enteritis and is responsible for approximately 25% of all deaths due to this disease in the developed countries (House et al., 1978). The genome of BCV is a single-stranded RNA with a polyadenylated 3' end (Guy and Brian, 1979). The structural proteins of BCV have been described by King et al, (1982) and include the fourth glycoprotein, the hemagglutinin-esterase protein. BCV has served as the prototypic hemagglutinating coronavirus for the study of the hemagglutinin-esterase protein (King et al., 1985; Hogue et al., 1988; Vlasak et al., 1988a,b). Prior to this study, part of the BCV genome containing the nucleocapsid protein gene and the matrix protein gene had been cloned and sequenced (Lapps et al., 1987). Hogue et al, (1984) showed that the bovine coronavirus and the mouse hepatitis virus are antigenically related and often in this study, comparisons between these two viruses are made.

C. RATIONALE FOR THE STUDY OF THE SPIKE PROTEIN GENE AND POTENTIAL NONSTRUCTURAL PROTEIN GENES OF BCV

The overall objective of our laboratory is to understand the molecular biology and pathogenesis of BCV. Towards this direction, the study of the spike protein gene and potential nonstructural protein genes was undertaken.

The spike protein of BCV is a 200K protein that is glycosylated with N-linked sugars and is present on the virion as cleaved subunits of 120K and 100K (King, 1982; Deregt et al., 1987). The 100K subunit apparently serves to anchor the 120K subunit to the virion. The gene for the peplomer protein of BCV has not been mapped or characterized, nor is there any information on the primary structure of the protein. This study was initiated to understand the structure of the protein and to begin to correlate it with its function.

It would be necessary to characterize the spike protein gene to begin to answer some of the following questions:

1. Is the primary structure of the BCV spike protein different from that of other coronaviruses and do these differences, if any, play an important role in the functional properties of the protein?
2. Can the proteolytic cleavage site of the spike protein be identified?
3. Can the mRNA for the spike protein be identified?
4. Can the spike protein be used as a recombinant vaccine, since it has been shown to induce neutralizing antibodies?
3. Is the synthesis and processing of the spike protein dependent on any other viral protein(s)?
4. What role does the spike protein play in the pathogenesis of the virus?

The number of nonstructural proteins present in BCV-infected cells is not known. cDNA cloning and sequence analysis of the BCV genome will help in the identification of open reading frames that could potentially encode proteins that are minor virion components or nonstructural proteins. The region of the genome between the spike and matrix protein genes in coronaviruses IBV, MHV JHM, MHV A59 and TGEV has been shown to contain open reading frames that could encode nonstructural proteins (Figure 1.2). The role of these proteins in the replication or pathogenesis of the virus is not known. With antibodies against these proteins or against peptides made from these proteins it may be possible to identify their functions. This study will help to answer some of the following questions:

1. How many nonstructural or minor structural protein genes are present in this region of the genome?
2. What putative functions can be identified from the deduced primary structure of the proteins?
3. Are these proteins made in BCV infected cells?
4. Can the proteins be localized within the virus infected cells and can a correlation be made with regards to their function?
5. Is the synthesis and processing of these proteins dependent on any other viral proteins?
6. Can the proteins when expressed induce neutralizing antibody and cytotoxic T cells, as observed when

nonstructural proteins of other viruses are expressed (Kees et al., 1984; Martin et al., 1988).

7. What role do these proteins play in the pathogenesis of the virus?

Some of these questions will be answered in this study, while others will still remain unanswered. However, the information obtained from this study will make it possible to design in vitro and in vivo experiments for detailed structure-function analyses of the BCV genome.

CHAPTER 2

GENERATION OF A PARTIAL cDNA LIBRARY OF THE BCV GENOME

Introduction

One of the initial steps required to study the molecular biology of BCV methodically and in detail was to cDNA clone and sequence the BCV genome. cDNA cloning and sequence analyses would enable the characterization of the genome, the identification of the genes and consequently the proteins they encode. In our laboratory B. Lapps (PhD dissertation, 1986) initiated studies on the characterization of the BCV genome using this approach. The 3' end of the BCV genome was cDNA cloned and the open reading frames for the nucleocapsid and the matrix proteins were characterized. These genes were found to lie within the first 2.5kb from the 3' end of the BCV genome (Lapps, 1987). In this study the cDNA cloning and generation of a partial cDNA library of the BCV genome is described.

Materials and Methods

Cells and virus

The Mebus strain of BCV was grown on the human rectal tumor cell line, HRT 18 (Tomkins, 1974). Cells were grown as monolayers in Dulbeccos modified minimal essential

medium (DMEM) containing 50 micrograms/ml gentamicin (Sigma) and 10% fetal bovine serum (Sterile Systems) (Hogue et al., 1984).

Growth and purification of virus

Confluent monolayers of cells in 150cm² flasks were infected with a MOI of 5. After 1.5 hours of adsorption at 37⁰C, the inoculum was removed and replaced with 15 ml of medium containing 1% fetal bovine serum. Supernatant fluid was removed after 48 hours and clarified at 800 X g for 10 minutes at 4⁰C. Virus was concentrated by centrifugation at 90,000 X g for 2 hours at 4⁰C onto a cushion of 60% wt/wt sucrose in TMEN (50mM Tris maleate (pH6), 0.1M NaCl, 1mM EDTA). Concentrated virus was isopycnically centrifuged in a continuous 60-20% (wt/wt) 18 ml sucrose gradient in TMEN for 4.5 hours at 90,000 X g at 4⁰C. Gradients were fractionated by using an ISCO gradient fractionator. The virus was diluted in TMEN and pelleted at 90,000 X g for 2.5 hours at 4⁰C

RNA purification

The viral pellet obtained from 500 ml of culture supernatant was resuspended in 500 microliters of RNase free TNE (0.01M Tris-HCl (pH7.5), 0.1M NaCl, 1mM EDTA) containing 2% sodium dodecyl sulfate (SDS) and 0.5mg Proteinase K (Boehringer Manneheim Biochemicals),

incubated at 37⁰C for 1 hour and the RNA was phenol extracted and ethanol precipitated (Maniatis, 1982).

cDNA cloning of the BCV genome

BCV genomic RNA was cloned using the method described by Gubler and Hoffman (1983). Random oligonucleotides (5-mers) (Pharmacia) were used as primers for first strand synthesis. The molar ratio of primer to genomic RNA used was 10,000:1 (Dowling, 1983). First strand synthesis was carried out in a 50 microliter reaction volume containing 50mM Tris-HCl (pH8.1), 148mM KCl, 8mM MgCl₂, 1mM DTT, 2mM each of the four dNTPs, 15 micrograms of 5-mer primer, 15 units of reverse transcriptase (Seikagaku) and 15 units of RNasin. Prior to the addition of the enzymes, the mixture was heated at 80⁰C for 5 minutes and then cooled to 37⁰C, so as to denature the RNA and anneal the primers. The reaction mixture was incubated at 37⁰C for one hour. Reaction products were extracted with phenol and ethanol precipitated in the presence of 2M ammonium acetate.

Second strand synthesis was carried out as described by Gubler and Hoffman (1983) in a 100 microliter reaction volume containing 20mM Tris HCl (pH7.5), 5mM MgCl₂, 10mM (NH₄)₂SO₄, 10mM KCl, 0.15mM NAD, 0.04mM dNTPs, 8.5 units/ml of RNase H (Boehringer Manneheim Biochemicals), 230 units/ml of DNA polymerase 1 (New England Biolabs). The reaction mixture was incubated at 12⁰C for 1 hour and

then at 22⁰C for 1 hour. Reaction products were extracted with phenol and ethanol precipitated with 2M ammonium acetate.

The double strand cDNA was homopolymer tailed by the method of Roy Choudhary and Wu (1980) in a 100 microliter reaction volume. The following were added to the DNA, 10 microliters of 10X Tris cacodylate buffer (1.4M K. cacodylate, 0.3M Tris HCl (pH7.6)), 0.5mMdCTP, 1mM CoCl₂ and volume made upto 100 microliters with water. The reaction mixture was incubated for 10 minutes at 37⁰C and 8 units of terminal deoxynucleotide transferase was added and incubated at 37⁰C for 1 minute. An average of 15-17 C residues/minute was added under these conditions. C-tailed double strand cDNA was annealed to G-tailed PstI linearized pUC9 vector (PL Biochemicals) according to the procedure described by Peacock (1981).

cDNA was also generated using specific primers for first strand cDNA synthesis, in this case the primer to genome ratio used was 100:1.

Plasmid transformation of E. coli JM83

Cells were made competent for plasmid uptake by the procedure of Hanahan (1983). This technique routinely yielded 1 X 10⁸ transformants per microgram of supercoiled plasmid DNA. After transformation, cells were plated onto medium that contained 100 micrograms/ml ampicillin, 1mM

isopropyl-1-thio-beta-D-galactosidase (IPTG), and 0.004% wt/vol 5-bromo-4-chloro-3-indoyl-beta galactoside (X-Gal). Recombinants (white) were distinguished from nonrecombinants (blue) on the basis of color (Messing, 1982).

Identification of colonies containing viral inserts

White colonies were picked and transferred onto selective indicator plates containing ampicillin, IPTG and X-Gal in a grid pattern. Colonies were grown overnight and transferred onto nitrocellulose filters (Schleicher and Schuell) by laying the filter on the colonies. Filters were removed, air dried, and in situ colony hybridization was done as described in Maniatis (1982). Briefly, colonies were lysed with alkali, neutralized, air dried and baked in a vacuum oven for two hours. Prehybridization was done at 65⁰C for 2-4 hours in prehybridization solution which contained 6X SSC (1X SSC contains 0.15M sodium chloride, 0.015M trisodium citrate (pH7.0)), 0.5% SDS, 5X Denhardt's solution, 100 micrograms/ml denatured salmon sperm DNA. Hybridization was done overnight (14-18 hours) at 65⁰C. The probe used was random-primed radiolabeled first strand cDNA generated from BCV genomic RNA. The RNA of the RNA-DNA hybrid was hydrolyzed with 0.12M NaOH at 37⁰C for two hours. An equal volume of 1M Tris HCl (pH8.0) was added and the free nucleotides separated from the DNA

using G-50 Sephadex spun column (Maniatis, 1982). The probe was heated at 100⁰C for 5 minutes and quick cooled on ice just before addition to the prehybridization solution. After hybridization the filters were washed twice with 2X SSC, 0.5% SDS for 20 minutes and finally with 0.1X SSC, 0.5% SDS at 55⁰C for 20 minutes. Filters were air dried and autoradiographed.

Isolation and size screening of plasmids with BCV specific inserts

Colonies identified as having viral inserts by hybridization with random primed BCV cDNA were grown in 3 ml of Luria Broth containing ampicillin to prepare plasmid DNA for characterization. Plasmid DNA was prepared by the alkali-lysis method, subjected to restriction endonuclease digestion, and analyzed by agarose gel electrophoresis (Maniatis, 1982). The sizes of the inserts were estimated using lamda DNA cut with HindIII and 1kb DNA ladder (BRL) as standards.

Results and Discussion

Initial efforts to cDNA clone the BCV genome (Lapps-PhD dissertation, 1986; Lapps et al., 1987) employed oligo dT as primer for cDNA synthesis and generated clones that were from the 3' end of the genome. Random pentamers were used as primers in this study with the objective to generate a

cDNA library of the entire BCV genome (Binns et al., 1985). This method yielded approximately 10,000 white colonies from three micrograms of total virion RNA.

The 10,000 white colonies that were obtained using random primers for cDNA synthesis were screened by colony hybridization with ^{32}P labeled first strand BCV cDNA as probe. 200 clones that contained viral inserts of greater than 0.3 kb were obtained. It was surprising that of the 10,000 white colonies only 200 colonies were found to contain viral specific inserts. The rest of the colonies probably contained very small inserts that could not be identified by hybridization using this screening procedure.

Plasmid DNA from the 200 cDNA clones was isolated and the size of inserts was estimated and found to range between 0.3 and 4.1 kb (Table 2.1).

The 200 cDNA clones were mapped by T. Kienzle to determine their position on the BCV genome. He determined that 50% of the clones mapped within the first 5kb of the 3' end of the BCV genome and another 50 clones mapped somewhere within the remaining 15 kb of the genome. These results indicated that the BCV genome is not evenly represented by the cDNA clones generated by random priming. Two possibilities are speculated to explain the skewing observed: (1) The secondary structure of the extracted genomic RNA may be such that the 3' end of the genome is more accessible for cDNA synthesis. (2) cDNA may be

Table 2.1 ESTIMATED SIZES OF cDNA CLONES GENERATED WITH RANDOM OR SPECIFIC PRIMERS

Clone Name	Size (kb)	Clone Name	Size (kb)	Clone Name	Size (kb)	Clone Name	Size (kb)
A1	0.8	B16	0.5	D5	2.1	F15	0.9
A2	1.0	B17	?	D6	1.6	F16	1.0
A3	0.4	B18	0.7	D7	0.4	F17	?
A4	1.0	B19	0.4	D8	1.2	G1	2.5
A5	1.0	B20	0.7	D9	0.6	G2	0.9
A6	0.6	B21	1.0	D10	1.8	G3	1.6
A7	0.6	B22	0.8	D11	0.5	G4	1.2
A8	0.6	B23	0.4	D12	1.3	G5	2.5
A9	1.0	B24	1.1	D13	?	G6	4.1
A10	1.2	B25	1.2	D14	2.0	G7	0.4
A11	1.5	B26	1.4	D15	1.0	G8	?
A12	0.5	B27	0.6	D16	1.5	G9	1.0
A13	1.0	B28	?	E1	0.6	G10	0.5
A14	0.7	B29	0.6	E2	0.8	G11	?
A15	1.0	B30	?	E3	0.7	G12	0.5
A16	?	C1	2.2	E4	1.6	G13	0.4
A17	0.8	C2	1.6	E5	0.4	G14	0.5
A18	0.5	C3	0.6	E6	1.2	G15	0.8
A19	2.2	C4	0.4	E7	0.9	G16	2.0
A20	0.6	C5	0.5	E8	0.4	G17	1.0
A21	0.8	C6	1.9	E9	0.7	G18	0.8
A22	1.0	C7	0.6	E10	1.4	H1	0.4
A23	2.9	C8	0.6	E11	0.7	H2	0.8
A24	0.7	C9	1.2	E12	0.4	H3	0.7
A25	1.0	C10	?	E13	?	H4	1.2
A26	?	C11	1.8	E14	0.4	H5	0.6
A27	1.5	C12	?	E15	0.6	H6	0.7
A28	0.6	C13	0.8	E16	0.5	H7	0.9
A29	1.2	C14	1.0	E17	0.5	H8	0.4
A30	0.6	C15	0.4	E18	0.7	H9	?
A31	?	C16	0.4	E19	?	H10	0.4
A32	1.3	C17	?	E20	0.5	H11	1.0
B1	1.8	C18	2.5	E21	0.7	H12	0.8
B2	1.3	C19	0.4	F1	0.7	H13	?
B3	?	C20	1.8	F2	0.9	H14	?
B4	0.7	C21	0.8	F3	?	H15	0.8
B5	?	C22	0.5	F4	?	H16	?
B6	?	C23	1.8	F5	0.9	H17	0.8
B7	?	C24	?	F6	0.4	H18	2.5
B8	0.7	C25	0.4	F7	0.9	H19	1.6
B9	0.8	C26	2.5	F8	2.4	I1	1.3
B10	1.0	C27	?	F9	0.5	I2	0.7
B11	0.8	C28	0.6	F10	0.5	I3	0.3
B12	?	D1	0.6	F11	0.5	I4	1.0
B13	?	D2	0.4	F12	0.5	I5	?
B14	?	D3	1.8	F13	1.7	I6	0.9
B15	?	D4	1.6	F14	?	I7	3.0

Table 2.1(continued)

Clone Name	Size (kb)	Clone Name	Size (kb)	Clone Name	Size (kb)	Clone Name	Size (kb)
I8	?	HP28	0.6	KT11	0.4*		
I9	?	HP29	0.6	KT12	0.4*		
I10	?	HP30	0.7	KS10	0.3*		
I11	2.5	HP31	0.8	KS11	? *		
I12	0.4	HP32	0.8	KS12	? *		
I13	?	HP33	0.5	KS13	0.4*		
I14	1.3	HP34	0.5	LK1	? *		
I15	1.6	HP35	1.4	LK2	? *		
I16	0.7	HP36	?	LK3	? *		
I17	0.5	HP37	?	LK4	? *		
I18	0.4	HP38	0.7	LK5	0.9*		
I19	0.4	HP39	0.7	LK6	0.9*		
I20	1.2	HP40	0.6	LP6	0.7		
I21	0.9	HP41	0.6	29	.15		
I22	0.4	HP42	1.3				
I23	1.8	HP43	0.9				
I24	0.4	HP44	1.5				
I25	2.5	HP45	1.3				
I26	1.3	HP46	4.1				
I27	1.3	HP47	0.7				
I28	2.4	HP48	0.4				
I29	0.7	HP49	?				
HP1	1.9	HP50	?				
HP2	1.9	HP51	?				
HP3	0.7	HP52	0.8				
HP4	1.2	HP53	0.7				
HP5	1.2	HP54	0.4				
HP6	0.7	HP55	0.8				
HP7	0.7	HP56	0.7				
HP8	?	HP57	?				
HP9	0.5	HP58	0.8				
HP10	1.0	HP59	0.7				
HP11	1.5	HP60	?				
HP12	1.0	HP61	?				
HP13	0.6	HP62	1.8				
HP14	?	HP63	0.9				
HP15	0.5	HP64	?				
HP16	0.6	HP65	1.0				
HP17	?	HP66	1.5				
HP18	0.6	HP67	3.0				
HP19	1.3	HP68	2.5				
HP20	?	HP69	?				
HP21	1.4	HP70	?				
HP22	1.0	HP71	0.5				
HP23	0.6	HP72	1.1*				
HP24	0.3	KT3	? *				
HP25	0.7	KT7	0.4*				
HP26	1.3	KT8	0.3*				
HP27	1.0	KT9	0.4*				

? size of clones not known

* clones generated by T. Kienzle

synthesized from large amounts of small mRNAs that are in the virion (B. Lapps et al., 1985; M. Hofmann, unpublished observation). Since the mRNAs form a 3' nested set with the genome, their cloning would demonstrate a skewing toward sequences found at the 3' end of the genome.

While mapping the cDNA clones generated by random priming gaps were encountered in which no overlapping clones could be found. For example, no clones were found to extend 5'-ward beyond clones HPa2 and I1 (Figure 2.1), yet clones upstream (ie., toward the 5' end of the genome) could be found. To clone these regions the 5' end of HPa2 and I1 were sequenced, specific primers were made, and these were used to prime first strand cDNA synthesis. cDNA clones LP6 and LK5 were generated using specific primer 5'AGATGTAATGGGGTCGG 3' and cDNA clone 29 was generated using specific primer 5' GTGGTTGGCAAGTACAAGGA 3'. cDNA clones that represented a sequence extending to a position 8.9 kb from the 3'end of the BCV genome were obtained (Figure 2.1). These clones were used in the identification of the spike protein gene discussed below and also for the identification of the hemagglutinin-esterase protein gene (T. Kienzle, PhD dissertation, 1989). cDNA clones that together spanned a 4 kb stretch and mapped somewhere between the 5' end of the genome and a point 8.9 kb from the 3' end of the genome were also generated.

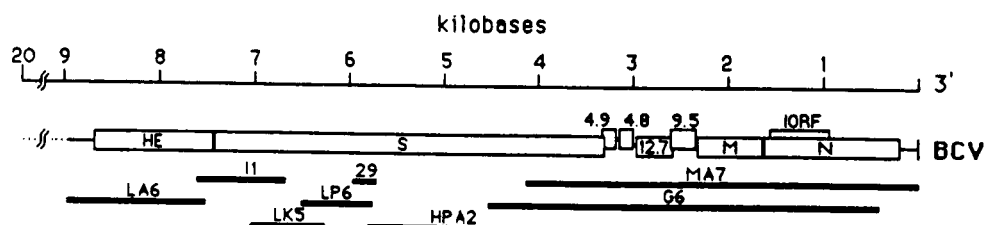


Fig 2.1 Gene map of the first 8,955 bases of the BCV genome and position of cDNA clones. Open boxes represent open reading frames, letters denote the structural proteins and numbers are predicted molecular weights of nonstructural proteins as described in Fig 1.2. cDNA clones are represented by heavy black horizontal lines. MA7 was generated using oligo-dT as primer (Lapps, PhD dissertation, 1986). cDNA clones G6, HPA2 and II were generated using 5-mer random primers, and LK5, LP6 and 29 were generated using specific primers.

CHAPTER 3

DEDUCED SEQUENCE OF THE BOVINE CORONAVIRUS SPIKE PROTEIN AND IDENTIFICATION OF THE INTERNAL PROTEOLYTIC CLEAVAGE SITE

Introduction

The bovine coronavirus is an important cause of neonatal calf diarrhoea (Mebus et al., 1978; House, 1979) and may also be the cause of winter dysentery in adult cattle (Saif et al., 1988). The mechanisms by which BCV causes disease and persistent infection (Crouch et al., 1985) are not understood, nor are current vaccines universally regarded as effective. In an effort to develop better strategies of prevention and therapeutics, the characterization of BCV proteins was initiated by determining their primary structure from cloned genome sequences.

BCV is known to be comprised of four major structural proteins (King and Brian, 1982). These are (1) a 200K spike (peplomer) glycoprotein (S), that exists on the virion as cleaved subunits of approximately 120K and 100K (S1 and S2), (2) a 140K glycoprotein (HE) that has both hemagglutinating (King et al., 1985) and esterase (Vlasak et al., 1988a, b) activities, and is comprised of two identical, disulfide linked 65K subunits (Deregt et al.,

1987; Hogue et al., 1989; Parker et al., 1987) (3) a 26K integral membrane glycoprotein (M) (Lapps et al., 1987) and (4) an internal phosphorylated nucleocapsid protein (N) (Lapps et al., 1987). Of these the S protein has been presumed to be the major structure by which coronaviruses attach to cells and initiate infection (reviewed by Spaan et al., 1988). The HE protein, however, has also been shown to bind to cells and initiate infection, and for BCV, the relative importance of these two proteins in initiating infection is not known. Both S and HE are probably important in inducing immunity since antibodies to each are known to neutralize virus infectivity in cell culture and in calves (Deregt and Babuik, 1987; Deregt et al., 1989a, b). S and HE, therefore, may both be useful in developing effective engineered vaccines against BCV.

The 3' terminal 8.9 kb of the BCV genome has been cloned and sequenced and the genes for the N and M proteins (Lapps et al., 1987), for four putative nonstructural proteins of 4.9K, 4.8K, 12.7K and 9.5K (Figure 1.1; Chapter 4), and for a 43K hemagglutinin-esterase protein (Kienzle, PhD dissertation, 1989) have been identified. This study (1) reports the cDNA cloning and sequence analysis of the S gene which resides between 3.3 kb and 7.4 kb from the 3' end of the genome, (2) confirms the location of the predicted internal cleavage site by amino terminal amino acid sequencing of the virion gp100 subunit, which also

estabilishes that gp120 and gp100 are the N-terminal S1 and C-terminal S2 subunits, respectively, (3) confirms that the S mRNA is the fourth largest virus specific mRNA species in infected cells, and (4) compares the deduced primary structure of the BCV spike protein with that of the deduced spike protein sequences of transmissible gastroenteritis virus (TGEV), feline infectious peritonitis virus (FIPV), infectious bronchitis virus (IBV) and the closely related mouse hepatitis virus (MHV). Regions that are conserved among the members of the three antigenic subgroups of coronavirus are identified and their role in the structure and function of the spike protein are discussed.

Materials and Methods

cDNA cloning of BCV genomic RNA

cDNA clones that contained the spike protein gene were identified from the partial BCV cDNA library that was generated (see Chapter 2 for details). cDNA clone MA7 was generated using oligo dT, cDNA clones G6, HPa2, and I1 were generated using random primers, cDNA clones LP6, LK5 and 29 were generated using specific primers (Chapter 2). These clones were used to obtain the spike protein gene sequence.

DNA sequencing

Some clones were sequenced by the chemical method of Maxam and Gilbert (1980) and some by the dideoxynucleotide chain termination method of Sanger (1983) as described by Kraft et al. (1988) using Sequenase enzyme (United States Biochemicals). For much of the sequencing, restriction endonuclease fragments were subcloned into the pGEM4Z vector (Promega) and forward and reverse sequencing primers for pGEM vectors were used. The list of the subclones made to sequence the spike protein gene is shown in Table 3.1. Sequence specific primers (Table 3.2) were also synthesized (University of Tennessee Analytical Facility) and used for sequencing within certain regions of the large subclones. Figure 3.1 shows the complete sequencing strategy used for sequencing the spike protein gene. Sequences were analyzed with the aid of the Microgenie program (version 5.0) from Beckman Instruments (Queen and Korn, 1984).

RNA extraction for Northern Blot analysis

Growth of HRT cells and infection with BCV was as described in Chapter 2 except that 250 cm² dishes were used and RNA was extracted at 8 hours post infection by the following protocol (Sethna et al., 1989). Cells were rinsed with 10 ml EBSS containing 10-20mM Hepes. An additional 10 ml of EBSS/Hepes was added, and the cells were scraped into a 50 ml conical tube (precooled on ice).

Table 3.1 SUBCLONES USED TO SEQUENCE THE SPIKE PROTEIN GENE

Name of subclone	Parent cDNA clone	Restriction enzymes used	Size of insert
C5XSp	MA7	Xba1-Spe1	0.4kb
GXP	G6	Xba1-Pst1	0.37kb
GHP	G6	Hpa1-Hind111	0.6kb
GHPac	G6	Hpa1-Acc1	0.7kb
GHPa	G6	Hpa1-Hpa1	1.0kb
HBS	HPa2	Bgl11-Spe1	0.6kb
H1	HPa2	Hind111-Hind111	0.3kb
H2	HPa2	Hind111-Hind111	0.38kb
Hac	HPa2	Acc1-Acc1	1.1kb
HSty	HPa2	Sty1-Sty1	1.1kb
HSp2	HPa2	Ssp1-Ssp1	0.4kb
HSp3	HPa2	Ssp1-Ssp1	0.3kb
HSp4	HPa2	Ssp1-EcoR1	0.2kb
IBE	I1	Bgl11-EcoR1	0.68kb
IBH	I1	Bgl11-Hind111	0.6kb
IME	LK5	Mlu1-EcoR1	0.6kb
IMH	LK5	Mlu1-Hind111	0.3kb
ISE	I1	Sty1-EcoR1	0.54kb
LKS	LK5	Ssp1-Ssp1	0.32kb
LPS	LP6	Ssp1-Pst1	0.4kb

Table 3.2 PRIMERS USED TO SEQUENCE THE SPIKE PROTEIN GENE

Name	Sequence
A	5' AACCGATAACCAAGTGGT 3'
B	5' ACCACTGGTTATCGGTT 3'
C	5' GTTACTATTGATTGTTC 3'
D	5' TTCGGCACCTCCAGTAC 3'
E	5' ACCATAGAGATTACCAT 3'
F	5' AGGATTAAACCTGCTAA 3'
G	5' CACTTAATCTCACTCAT 3'

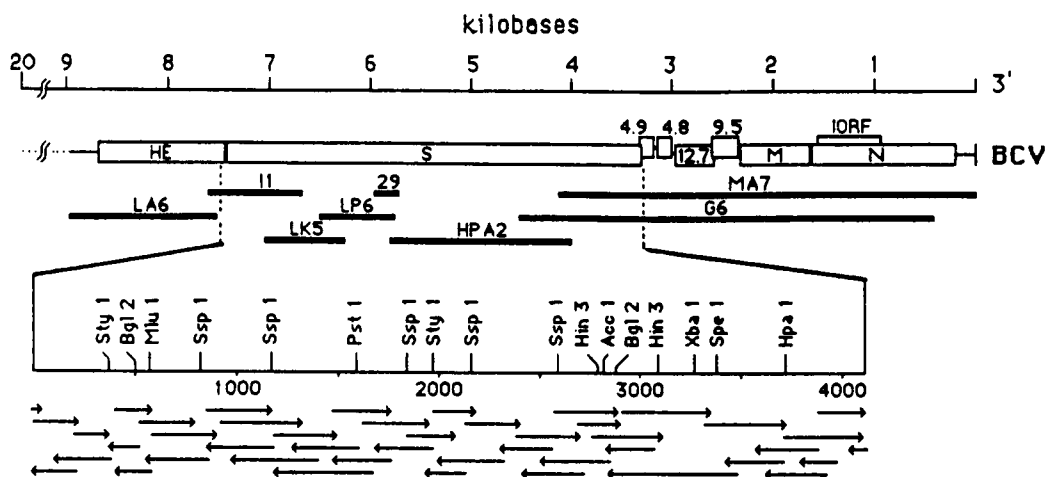


Fig 3.1 Gene map of the BCV genome and the sequencing strategy of the spike protein gene. Open boxes represent the open reading frames. The horizontal arrows in the enlarged region show the sequencing strategy. Restriction endonuclease sites used to generate the subclones listed in Table 3.1 are shown.

The plate was rinsed once with 5 ml EBSS/Hepes and added to the 50 ml tube. Cells were pelleted at 4,000 x g for 4 minutes at 4⁰C, the supernatant was discarded, and the cells were suspended in 3 ml ice-cold lysis buffer (10mM Tris-HCl(pH 7.0), 10mM NaCl, 5mM MgCl₂, 1% v/v NP 40, 10mM vanadyl ribonucleoside complex). This was vortexed for twenty seconds and nuclei were pelleted at 4,000 x g for 4 minutes at 4⁰C. The supernatant was gently, but quickly transferred to a 15 ml conical tube containing 0.2 volumes (0.6 ml) of 10% SDS. One mg proteinase K (Boehringer Mannheim Biochemicals) was added and the sample incubated at 37⁰C for one hour. The RNA was extracted with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) (equilibrated in 100mM Tris-HCl (pH8.0)) by vortexing three times and centrifuging for 5 minutes at 4,000 x g at room temperature. The aqueous phase was transferred to a new tube and the interphase reextracted with 1 ml lysis buffer without vanadyl ribonucleoside complex. The aqueous phases were combined and the RNA precipitated by the addition of one-tenth volume 3M sodium acetate and 2.2 volumes ethanol, mixing and storing at -20⁰C until used.

Electrophoresis and Northern analysis of BCV RNA

Ten micrograms per lane of total RNA was denatured in formaldehyde and formamide and electrophoresed in the presence of formaldehyde as described by Sethna et al,

(1989). Briefly, RNA was dissolved in 37 microliters 1X MOPS buffer (20mM morpholinopropane sulfonic acid (pH 7.0), 5mM sodium acetate, 1mM EDTA), 50% deionized formamide, 2.2M formaldehyde, denatured by heating at 65⁰C for 5 minutes, mixed with 8 microliters loading dye (50% glycerol, 0.25% bromophenol blue, 0.25% xylene cyanol, 1mM EDTA), and loaded onto a 1% agarose gel made in 1X MOPS buffer and 2.2M formaldehyde. Electrophoresis was carried out at 140 volts for 4 hours at room temperature in 1X MOPS buffer in the fume hood. The RNA was transferred to a Nytran membrane by using a Vacublot apparatus (LKB) with 20X SSC for 6 hours and UV cross linked for 2 minutes with 254 nm wavelength (Khandijian 1982). The membrane was prehybridized at 55⁰C for 4 hours in 5XSSC, 0.1% Ficoll, 0.1% Polyvinylpyrrolidone, 0.1% BSA, 50mM NaPO₄ pH 7.0, 1% SDS, 1000 micrograms/ml of sheared salmon sperm DNA and 50 micrograms/ml of tRNA.

The probe for the spike protein mRNA was the insert of cDNA clone LP6, the probe for the HE mRNA was a StyI-EcoR5 fragment representing bases 48 through 978 of the HE gene and the probe mapping on the 5'side of HE was a 600bp AccI-AatII fragment containing 400bp of clone LA6 (a sequence ending 101 bases upstream from the start codon of the HE gene)(Kienzle, PhD dissertation, 1989). The fragments were isolated, nick translated and used to probe the Northern blots containing RNA from BCV infected cells.

After hybridization the filters were washed twice with 2X SSC + 0.5% SDS at 25⁰C for 10 minutes and washed once with 0.1X SSC and 0.5% SDS at 55⁰C for 30 minutes. Filters were air dried and exposed to Kodak XAR-5 film at -80⁰C.

Amino acid sequencing of the subunits of the spike protein

The amino terminal ends of the purified gp120 and gp100 subunits were subjected to sequencing by the method of Matsudaira (1987) in his laboratory. Unlabelled BCV was purified by isopycnic sedimentation in sucrose gradients and the proteins were electrophoretically separated after reduction with 2-mercaptoethanol (King and Brian, 1982) and electroblotted (Hogue et al., 1984) onto polyvinylidene difluoride membrane (Matsudaira, 1987). The proteins were visualized by staining with Coomassie brilliant blue and the gp120 and gp100 bands were excised and shipped to Dr. Matsudaira (MIT) for analysis.

Results

Deduced amino acid sequence of the spike protein

Sequence analysis of the 3' end of the BCV genome for 3,400 bases has revealed a 3' noncoding region of 291 bases, and open reading frames for the N gene, the M gene (Lapps et al., 1987), and four genes for putative nonstructural proteins of 9.5, 12.7, 4.8 and 4.9K (Figure 1.2; Chapter 4). Complete sequencing of cDNA clone MA7

which extends 4.2 kb from the 3' end of the genome (Figure 2.1) revealed a continuous open reading frame located on the 5' side of the open reading frame for the 4.9K protein. The deduced sequence of the extended open reading frame demonstrated high amino acid sequence similarity to the C-terminal end of the antigenically related MHV A59 (Luytjes et al., 1987) and MHV JHM (Schmidt et al., 1987) S proteins, both antigenic homologues of the BCV S protein (Hogue et al., 1984). These data suggested that the S protein gene of BCV lies in the same relative map position on the genome as does the spike protein gene of MHV. To complete sequencing of the S gene, both strands of clones G6, HPa2, 29, LP6, LK5 and I1 were sequenced (Figure 3.1).

The total sequence for the putative S ORF extended to a position 7.4 kb from the 3' end of the genome and contained 4,089 bases (Figure 3.2). It is concluded that this ORF encodes the S protein since it potentially encodes a 1,363 amino acid protein with a molecular weight of 150,825 the approximate size of the unglycosylated spike precursor (Deregt et al., 1987), and because its deduced amino acid sequence shows extensive similarity throughout with the S proteins of both strains of MHV. Five other open reading frames ranging in size from 34 to 66 amino acid were also found within the S gene sequence in the plus one open reading frame, but their significance is not known at this

Fig 3.2 Nucleotide sequence of the S gene and its deduced amino acid sequence. The nucleotide sequence shown begins with a TAG termination codon of the HE gene (underlined) 17 bases upstream of the presumed S start site, and ends with the TAA termination codon of the S protein. The first three amino acids of the putative 4.9K protein are shown beginning at base position 4,099. Consensus intergenic sequences are boxed. The presumed amino terminal signal peptide and carboxy-terminal anchor sequences are underlined. Potential N-linked glycosylation sites (NXS or NXT, where X≠P) are boxed. The proteolytic cleavage site separating S1 and S2 is identified with an arrow. (-) and (.) below the amino acids are sequences absent in MHVJHM and MHVA59 respectively.

52

53

time. Two features of the putative S open reading frame suggest that it would be a functional gene. First, it is preceded immediately upstream (beginning at base 12 in Figure 3.2) by the consensus CYAAAC intergenic sequences thought to play a role in leader priming of coronavirus transcription. This sequence is found to precede other BCV structural protein genes that are known to be transcribed (Lapps et al., 1987). The sequence is also found three times within the S ORF, beginning at position 817, 1,667 and 3,776, but it is not established that transcripts initiate at any of these sites. Second, the presumed initiation codon for translation (AAACATGT) is in a moderately favorable context for translation (Kozak, 1984).

Five features of the deduced BCV S protein reflect the properties of five other coronavirus spike proteins that have been characterized to date from nucleotide sequence data (Table 3.3) (Binns et al., 1985; Binns et al., 1986; Jacobs et al., 1987; Kusters et al., 1989; Luytjes et al., 1987; Niesters et al., 1986; Rasschaert and Laude, 1987; Schmidt et al., 1987). (1) There is an N-terminal hydrophobic stretch of amino acids (note arrow A hydrophobicity plot Figure 3.3) which predicts a signal peptide with a cleavage site between amino acids 17 and 18 (von Heijne, 1985). (2) There are 19 potential asparagine-linked glycosylation sites that could give rise

Table 3.3 FEATURES OF CORONAVIRUS SPIKE PROTEINS OBTAINED FROM DEDUCED AMINO ACID SEQUENCES.

	BCV ^a	MHVA59 ^b	MHVJHM ^c	IBV ^c	TGEV ^e	FIPV ^f
Size of gene	4.08kb	3.97kb	3.7kb	3.48kb	4.3kb	4.35kb
Number of amino acids	1361	1324	1235	1162	1447	1452
Predicted mol.wt.	150K	146K	136K	127K	158K	160K
Size of subunits	85.6K(S1) 65.1K(S2)	79K(S1) 66K(S2)	70K(S1) 66K(S2)	64K(S1) 61K(S2)	-	-
Site of Trypsin cleavage	KRRSRR	RRAHR	RRARR	RRFRR	None	None
Length of putative signal sequence	17	17	17	18	16	18
Number of cysteine residues in the hydrophilic C-terminal sequence	8	10	10	7	11	10
Length of putative anchor sequence	17	27	34	44	45	21
Number of potential N-glycosylation sites	19	20	21	28	32	35

^a from this study

^b from Luytjes et al., 1987

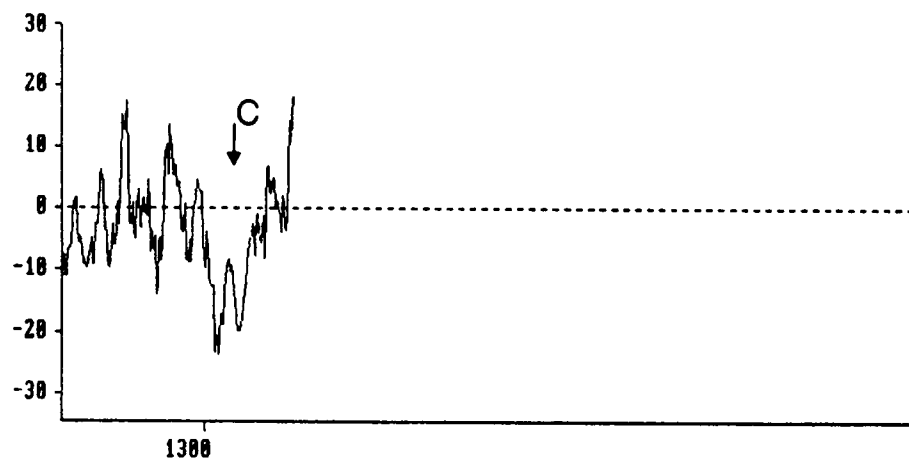
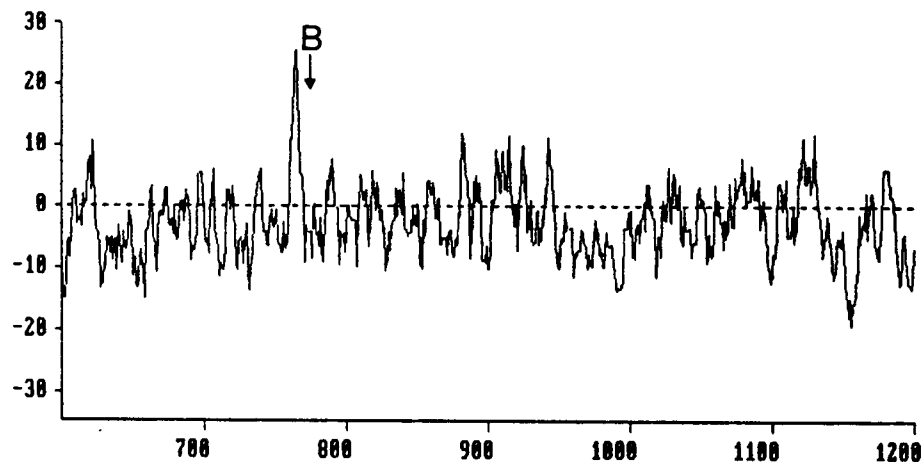
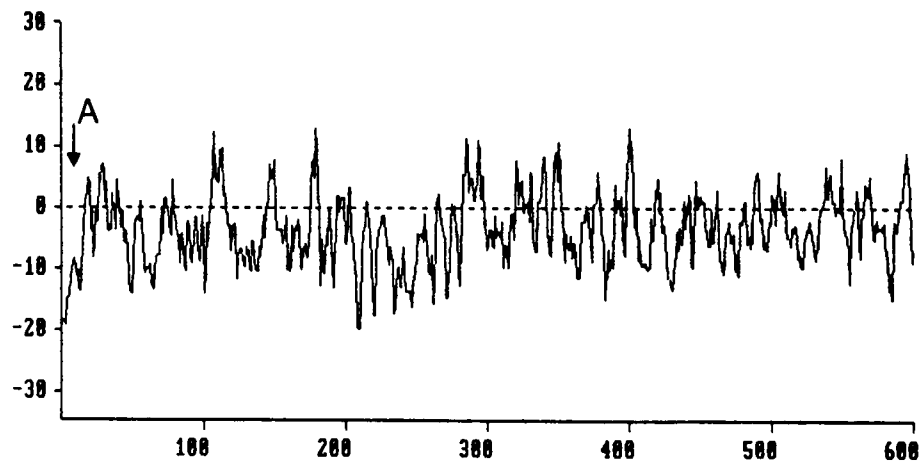
^c from Schmidt et al., 1985

^d from Binns et al., 1987

^e from Rasschaert et al., 1987

^f from DeGroot et al., 1987

Fig 3.3 Hydrophobicity plot of the BCV spike protein. Analysis was performed using the method of Kyte and Doolittle (1982) with a 9 amino acid window. Positive values represent hydrophilic regions and negative values represent hydrophobic regions. Arrow A denotes the N-terminal putative signal sequence, Arrow B the internal protease cleavage site, and Arrow C, the C-terminal putative anchor sequence.



to the only kind of glycosylation demonstrated for this protein (Hogue and Brian, unpublished data; Deregt et al., 1987). (3) There is a hydrophobic stretch of 17 amino acids near the C terminus that could serve as a stop transfer sequence (note arrow C hydrophobicity plot Figure 3.3). (4) There is a stretch of 8 amino acids on the immediate N-terminal side of the predicted anchor sequence (K-W-P-W-Y-V-W-L, beginning with amino acid 1,305) that is identical in all coronavirus S proteins sequenced to date. (5) There is a cysteine rich C-terminus of 8 amino acids that is probably in the intravirion domain and predicted to be the site of acylation. In common with MHV and IBV viruses, but not in common with TGEV and FIPV, is also an internal sequence of basic amino acids that, in the case of MHV and IBV, lies on the immediate N-terminal side of the protease cleavage site (Cavanagh et al., 1986; Luytjes et al., 1987). In BCV the sequence K-R-R-S-R-R beginning with amino acid 763, and, on the basis of the pattern in MHV and IBV, predicts a cleavage between amino acids 768 and 769 (note arrow in Figure 3.2 and arrow B in Figure 3.3). Cleavage at this point would divide the unglycosylated S protein into an N-terminal segment of 85,690 daltons (S1) and a C-terminal segment of 65,153 daltons (S2).

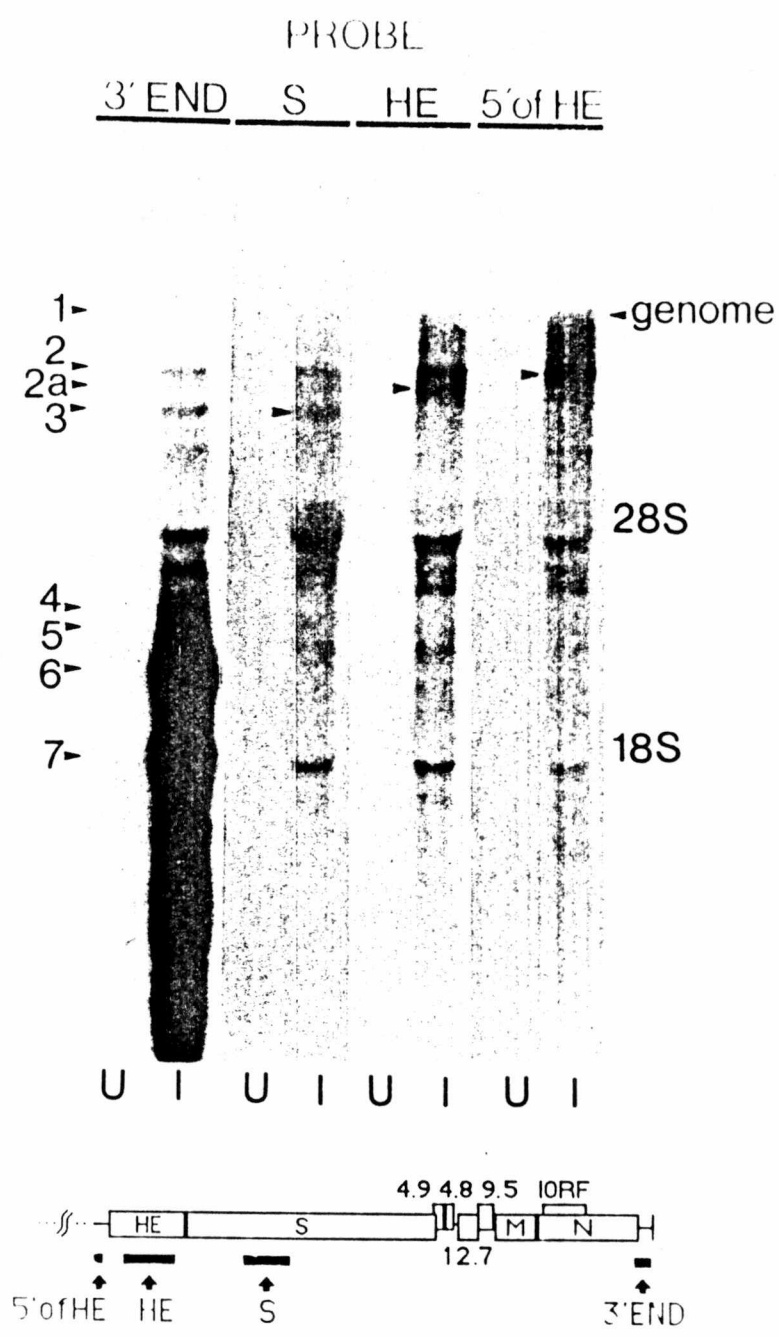
Terminal amino acid sequencing confirms the predicted internal protease cleavage site and identifies the virion 120K and 100K subunits as S1 and S2 respectively

No amino acid sequence could be obtained from the 120K subunit, possibly because of N-terminal blockage. The N-terminal sequence of the 100K subunit could be obtained, however, and was determined to be X-I-T-T-G-Y-X-F, identifying the first amino acids downstream from the predicted cleavage site (see arrow in Figure 3.2). These results confirm the predicted internal cleavage site and establish that the 120K subunit is S1 and the 100K subunit is S2.

S mRNA is the fourth largest virus specific mRNA in infected cells

Nine species of polyadenylated mRNAs have been identified in BCV infected cells. These were named 1, 2, 2a, 3, 3a, 4, 5, 6, 7 (beginning with the largest). mRNAs 3a, 4, 5, 6, 7 encode open reading frames for 4.9K, 12.7K, 9.5K, M and N proteins (see Chapter 4). mRNA2a has been recently shown to encode the open reading frame for the HE protein (T. Kienzle, PhD dissertation, 1989). Northern analysis of total RNA from BCV infected cells hybridized separately with nick translated probes for the S (base 899 through 1,638), HE and for a sequence 5' to HE ORFs is shown in Figure 3.4. The S specific sequence hybridized to

Fig 3.4 Northern blot analysis identifying mRNA3 as the transcript from the S gene. Cytoplasmic RNA from infected and uninfected cells was probed with fragments of cDNA clones as described in Materials and Methods. The map position of the cloned fragments is illustrated below. U=RNA from uninfected cells. I=RNA from infected cells.



mRNA 3 and larger species, the HE sequence identified mRNA 2a and larger species, and the sequence 5' of the HE ORF identified mRNA 2 and genome, as would be predicted from the 3' nested set pattern of mRNAs. A sequence from the very 3' end of the genome was used as a probe to identify all virus specific mRNAs. The exact nature of bands migrating as displaced molecules just below the positions of 28S and 18S ribosomal RNA and those migrating between the ribosomal RNA markers in lanes probed with S, HE, and 5' of HE probes is not known. On the basis of size they may represent transcripts initiating from within the S gene sequence, but based on their hybridization to the probes used, their structure would be very unusual and would not fit the 3' nested set pattern. It is believed that they represent specific degradation products of the large BCV mRNA species, however, since coronavirus mRNAs are known to have half lives of only a few hours (Keck et al., 1988; Sethna et al., 1989).

Discussion

Analysis of the deduced S protein sequence of BCV establishes that it is structurally very similar to the S proteins of MHV JHM (Schmidt et al., 1987) and A59 (Luytjes et al., 1987), although neither of these viruses has been shown to possess HE as a second peplomeric surface glycoprotein. Recently, some strains of JHM have been

shown to synthesize an HE subunit, but no dimeric peplomer structure or enzyme activity was demonstrated (Shieh et al., 1989). Therefore, no striking deficiency in the BCV S protein, such as a large deletion, that would suggest a loss of functional capacity correlating with the presence of the HE glycoprotein is observed. In fact by comparison with the BCV S protein deletions appear to have occurred in the MHV S protein (see Figure 3.5 and 3.6 dot matrix comparison). In JHM, a contiguous sequence of 138 amino acids is missing from the S1 subunit (Figure 3.2 and 3.7). A missing sequence mapping in the same region of the molecule is also found in A59, but is only 50 bases in length, and is discontinuous (Figure 3.2 and 3.7). The function of the additional sequence in the BCV S1 subunit is not known, but one possibility is that it interacts with the HE glycoprotein. No electron microscopic or chemical data exist, however, to suggest that S and HE do physically and chemically interact (Bridger et al., 1978; King and Brian, 1982; King et al., 1985). It is interesting to note that the entire region is especially rich in cysteine residues and contains 26% of the total cysteines in the BCV S protein (Figure 3.7). This suggests that this part of the molecule may be important for intramolecular or intermolecular disulfide linkages.

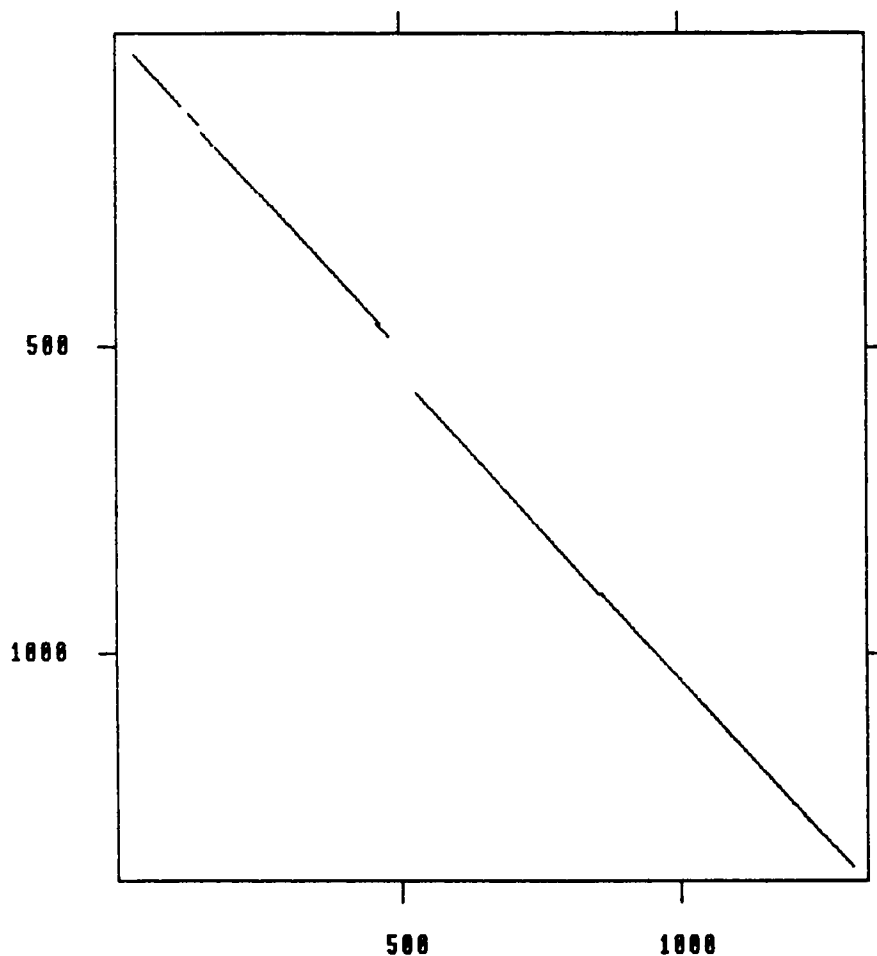


Fig 3.5 Comparison of the amino acid sequence of BCV and MHV A59 spike proteins by diagon plotting. BCV amino acid sequence is plotted on the Y axis and MHV A59 on the X axis. A dot is placed where there is 50% amino acid matches between the two proteins in a stretch of 24 amino acids.

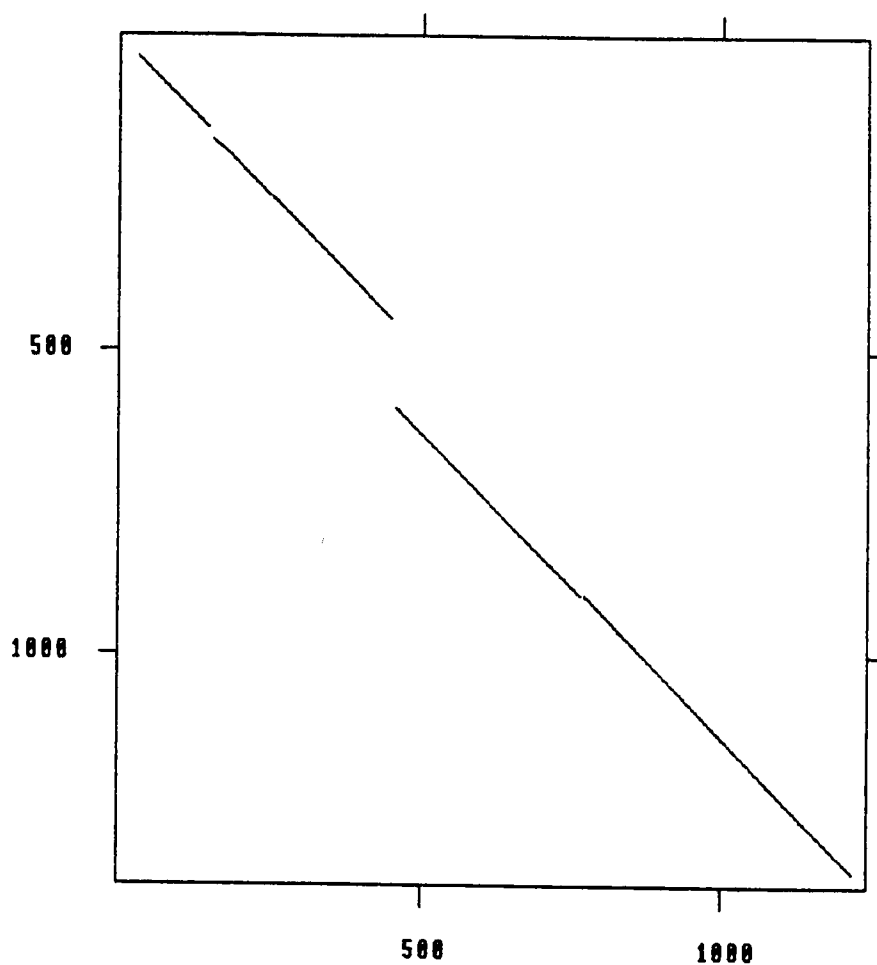


Fig 3.6 Comparison of the amino acid sequence of BCV and MHV JHM spike proteins by diagonal plotting. BCV amino acid sequence is plotted on the Y axis and MHV JHM on the X axis. A dot is placed where there is 50% amino acid matches between the two proteins in a stretch of 24 amino acids.

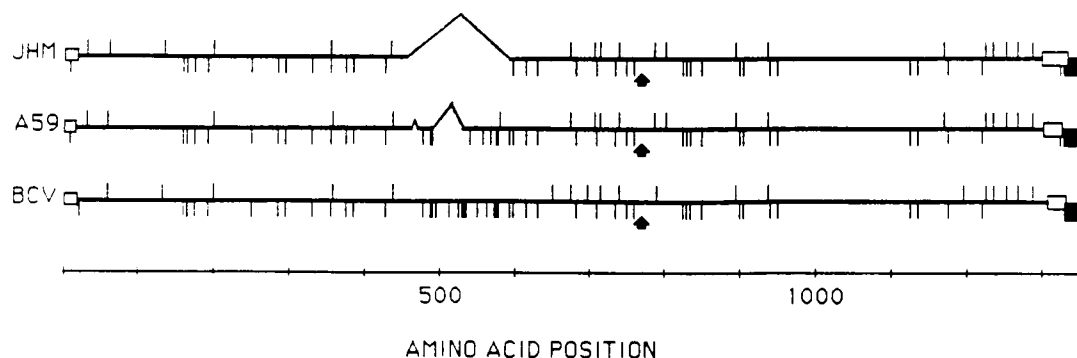


Fig 3.7 Structural comparison of S proteins of MHV JHM, MHV A59, and BCV. Sequences are aligned for maximum homology. A sequence found in BCV but not found in MHV JHM or MHV A59 is expressed as a gap (thin line) in MHV sequences. Putative N-terminal signal peptides and C-terminal anchor sequences are boxed. Vertical lines above the sequence indicate potential asparagine linked glycosylation positions, and below the sequence, cysteine positions. The identified (BCV, MHV A59) and putative (MHV JHM) proteolytic cleavage sites are identified by arrows.

The BCV and MHV S proteins show remarkable amino acid sequence identity suggesting that these viruses are recently diverged. After aligning the amino acid sequences for maximum homology, and excluding the large gap in the MHV sequences, the S1 subunits of JHM and A59 show 62% and 60% identity with BCV, and the S2 subunits show 75% and 74% respectively. Throughout the S protein, 41 of 56 cysteine positions and 13 of 19 potential N-linked glycosylation sites are conserved. The internal proteolytic cleavage position (not yet confirmed for JHM) is also conserved (Figure 3.7). Several heptad repeat sequence patterns (a, b, c, d, e, f and g, where a and d are apolar amino acid residues) described in the S2 subunit of MHV that are thought to contribute to a coiled coil structure in the S stem (DeGroot et al., 1987a, 1987c) are also found in the BCV S2 subunit. The pattern of greater amino acid sequence divergence in the S1 subunit is consistent with the model of Cavanagh (1983b) and DeGroot et al, (1987a) which suggests that the S1 subunit comprises the exposed bulbous structure of the spike and probably contains most (Cavanagh et al., 1988), but not all (Makino et al., 1987; Talbot et al., 1988), of the neutralizable antigenic sites. The spike knob is thus the most likely part to undergo changes from immunologic selective pressure.

The BCV S protein, like those of MHV and IBV, (Binns et al., 1985; Binns et al., 1986; Kusters et al., 1989; Niesters et al., 1986), but unlike those of TGEV (Jacobs et al., 1987; Rasschaert and Laude, 1987; Tung and Brian, unpublished data) and FIPV (Jacobs et al., 1987), undergoes proteolytic cleavage. Fusion of cells in culture is one biological activity associated with cleavage of the MHV S protein (Sturman et al., 1985). Despite its extensive similarity with the MHV S protein, the BCV S protein shows no fusion activity. It is not clear why BCV and MHV behave so differently in their fusogenic properties, but functional evaluation of sequence differences near the cleavage sites of these viruses may aid in elucidating the mechanism by which MHV causes fusion. This is especially interesting since hydrophobic regions common at cleavage sites on fusion proteins of paramyxoviruses and myxoviruses, are absent in the MHV S protein (Luytjes et al., 1987) and a different mechanism for fusion may be involved.

Important questions remain concerning the relative roles of S and HE glycoproteins in both BCV structure and replication. Do S and HE physically interact with each other on the cell surface? Do both S and HE interact with cell receptors to start infection? If so what is the importance of each? Does the additional sequence present in the BCV S protein determine organ and tissue specificity?

Now that the S protein gene is cloned, strategies can be developed to answer these questions. The spike protein can now be used in an experimental recombinant vaccine in attempts to improve protective immunity against BCoV infection.

CHAPTER 4

SEQUENCE AND EXPRESSION ANALYSIS OF PUTATIVE NONSTRUCTURAL PROTEINS OF 4.9K, 4.8K, 12.7K, and 9.5K ENCODED BETWEEN THE SPIKE AND MATRIX PROTEIN GENES OF THE BOVINE CORONAVIRUS

Introduction

cDNA cloning and sequence analysis of the genomes of a variety of coronaviruses have identified the genes for the structural proteins at the 3' end of the genome (Armstrong et al., 1983; Bournsnel et al., 1984; Kapke et al., 1986; Lapps et al., 1987). In IBV (Bournsnel et al., 1987) and MHV A59 (Pachuk et al., 1989), the RNA dependent RNA polymerase gene has been identified and is encoded by the 5' end of the genome. Interspersed among the structural protein genes are open reading frames that could encode potential nonstructural proteins or proteins that form minor components of the virion (Bournsnel et al., 1985; Skinner et al., 1985; Budzilowicz and Weiss, 1987; Kapke et al., 1988). However the position and sizes of these open reading frames varies with each member of the coronavirus family and these are schematically represented in Figure 1.2 (Chapter 1).

Our laboratory has reported cDNA cloning and sequence analysis of four open reading frames of BCV encoding the

nucleocapsid, matrix, spike and hemagglutinin-esterase proteins (Lapps et al., 1987; Chapter 3; Kienzle PhD dissertation, 1989) and one internal open reading frame within the nucleocapsid protein gene (Lapps et al., 1987). To further characterize the BCV genome and identify all potential genes, the sequence between the spike and the matrix protein genes was determined. Open reading frames for potential proteins of 4.9K, 4.8K, 12.7K and 9.5K were found to be present in this region of the genome. Sequence and Northern analyses of these open reading frames led to the finding that there is one mRNA species (mRNA 3a) that encodes the 4.9K and 4.8K in tandem as 5' proximal ORFs and there are separate mRNAs for the putative 12.7K and 9.5K proteins (mRNA 4 and mRNA 5). The pattern for the expression of the 9.5K protein in BCV is in striking contrast to the MHV counterparts in which the 9.6K protein is synthesized from a downstream ORF of mRNA 5.

Materials and Methods

Generation of deletion clones of MA7 for sequencing the region between the spike and matrix protein genes

MA7, a cDNA clone described by Lapps (1987) (Figure 2.1, Chapter 2) extends to 4.2 kb from the 3' end of the BCV genome. Deletion clones of MA7 were made in order to sequence the region between the spike and matrix protein genes. To do this, the 3.73 kb PstI fragment from the 5'

end of MA7 was isolated by electrophoresis onto NA 45 paper (Schleicher and Schuell) (Lizardi et al., 1982), blunt ended with Klenow enzyme and ligated to Hind111 linkers (Maniatis, 1982). The fragment was digested with Hind111 restriction enzyme and ligated to pUC19 plasmid that had been digested with Hind111 and dephosphorylated (Maniatis, 1982). The ligated DNA was used to transform E. coli HB101 cells (Hanahan, 1983). One subclone was labeled DM3 and used to create deletions as described below.

Plasmid DNA from DM3 was prepared by the alkali lysis method (Maniatis, 1982) and digested with *SauI* and *PstI* which leaves 5' and 3' overhanging ends respectively, and subjected to unidirectional digestion with exonuclease 111 by the method of Henikoff (1984). Exonuclease 111 removes bases only from a 5' overhanging end and generates unidirectional deletions in the insert. The DNA was blunt ended with *S1* nuclease, ligated and used to transform E. coli JM83 cells (Hanahan, 1983). Plasmid DNA from the deletion clones was prepared by the alkali lysis method and the size of the inserts were estimated by agarose gel electrophoresis (Maniatis, 1982). Figure 4.1 shows the relative map positions of the deletion clones that were used for sequencing the region between the matrix and spike protein genes.

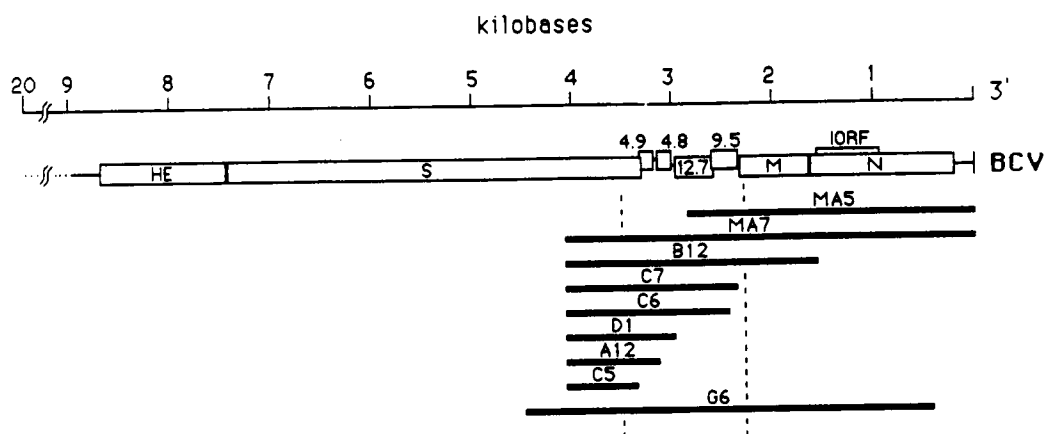


Fig 4.1 Map position of the deletion clones of cDNA clone MA7 generated to sequence the region between the spike and matrix protein genes. The region sequenced is represented between the dotted lines. cDNA clones MA7 and MA5 were generated using oligo dT as primer (Lapps, PhD dissertation, 1986) and G6 was generated using random primers (Chapter 2).

End labeling of the DNA for Maxam and Gilbert sequencing

Deletion clones shown in Figure 4.1 were used for DNA sequencing. Plasmid DNA was isolated and purified using cesium chloride centrifugation (Maniatis, 1982). The DNA was linearized with EcoR1 and then digested with Sal1 which leaves a 5' overhanging end and releases a 30 base pair fragment. The site generated by Sal1 was end labeled using a specific 3' fill in reaction with reverse transcriptase and α - ^{32}P -dNTP (Lapps, PhD dissertation, 1986). To sequence the opposite strand the plasmid was first digested with Sal 1 and the 5' end generated was end labeled using polynucleotide kinase and γ - ^{32}P -dNTP (Maniatis, 1982) and then digested with EcoR1. The end labeled plasmid was used for DNA sequencing.

DNA sequencing

The chemical method of Maxam and Gilbert (1980) for DNA sequencing was used. Sequence obtained was analysed by the Microgenie program Version 5.0 (Korn and Queen, 1984) marketed by Beckman.

Preparation of RNA from BCV infected cells for Northern analysis

Growth of HRT cells and infection with BCV was as described in the materials and methods of Chapter 2. RNA was extracted from uninfected and infected cells at 9 hours post infection using guanidinium isothiocyanate as

described by Lizardi (1983). RNA was pelleted through cesium chloride extracted with chloroform:1-butanol, ethanol precipitated, dissolved in water and poly A⁺ containing RNA was selected by oligo (dT)-cellulose chromatography (Dennis and Brian, 1982). 5 micrograms/lane of oligo dT selected RNA was used for Northern blot analysis as described in Chapter 3 and Sethna et al, (1989).

Generation of specific probes for Northern analysis

The 350 bp Pst1 fragment of cDNA clone MA5 (Figure 4.1; Lapps et al., 1987), which represents the 3' end of the BCV genome was nick translated and used as probe for the 3' end of the genome. Deletion clone C7 was cut with Hinc 11 and a 250bp fragment was isolated and cloned into the Hinc11 site of plasmid pGEM4Z and was used to transform E. coli JM83 cells. One of the recombinant clones was named Hin2 and contained sequences within the 9.5K ORF. The insert of Hin2 was nick translated and used as probe to identify the 9.5K mRNA. Clone AB4 (described below) was digested with EcoR1 and Alu1 and a 300 bp fragment isolated and cloned into the EcoR1 and Hinc 11 site of pGEM4Z. The recombinant plasmid was used to transform E. coli JM83 cells and yielded clone EA13.8 which contained sequences within the 12.7K ORF. The insert of clone EA13.8 was isolated and nick translated and used as probe for the

12.7K mRNA. The insert of deletion clone A12 (Figure 4.1) was isolated, nick translated and used as probe for the region 5' to the 12.7K open reading frame which contained 4.8K and 4.9K ORF sequences.

Construction of pGEM recombinant vectors

To make transcripts that contained the open reading frames for the 12.7K and 9.5K proteins, constructs were made in pGEM4Z vector which contained the T7 and SP6 promoters for in vitro transcription (Promega Biotec). A 1.8kb AccI fragment of B12 (a deletion clone of MA7, Figure 4.1) was isolated after agarose gel electrophoresis using NA45 paper (Schleicher and Schuell). The fragment was blunt ended with Klenow enzyme (Maniatis, 1982) and ligated into the SmaI site of pGEM4Z and used to transform E. coli JM83 cells (Hanahan, 1983). The 1.8kb fragment contained the open reading frames for the 12.7K, 9.5K and matrix proteins in tandem. Two clones in opposite orientations were obtained. Recombinant clone AB4 yielded transcripts from the SP6 promoter while clone AB6 yielded transcripts from the T7 promoter (Figure 4.2).

Clone AB9.6 (Figure 4.3) was made to generate transcripts that had the open reading frame for the 9.5K protein at the 5' proximal end. Plasmid DNA from clone AB6 was digested with HincII which removed part of the 12.7K ORF, the digested plasmid DNA was isolated, religated and

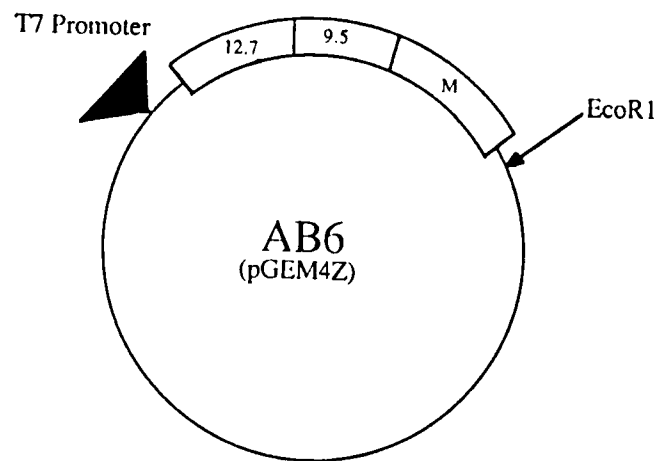


Fig 4.2 Schematic representation of recombinant plasmid AB6.

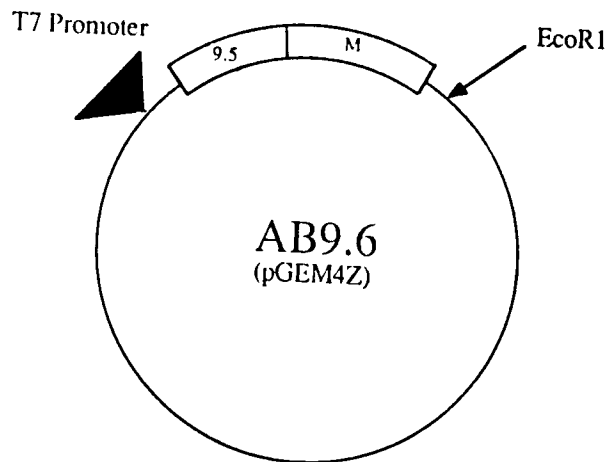


Fig 4.3 Schematic representation of recombinant plasmid AB9.6.

E. coli JM83 cells were transformed with it to yield the recombinant clone AB9.6 (Hanahan, 1983). The recombinant plasmid AB9.6 yielded transcripts with T7 polymerase.

In vitro transcription

Approximately 10 micrograms of plasmid DNA was linearized with EcoR1, phenol extracted, ethanol precipitated (Maniatis, 1982), and used as template for in vitro transcription. Transcripts were prepared with T7 polymerase according to the capping protocol #2 in Promega catalogue (1987) (Krieg and Melton, 1984). The reaction was carried out in a 50 microliter volume (containing 0.5mM ATP, CTP, UTP & 0.05mM GTP; 10mM DTT; 1X Transcription buffer; 5 micrograms BSA; 0.5mM GpppG; 50 units of T7 polymerase; 100 units of RNasin). The reaction mixture was incubated at 40⁰C for 60 minutes. The products were phenol extracted and ethanol precipitated, and the capped transcripts were used for in vitro translation.

In vitro Translation

In vitro translation of the capped transcripts was performed according to the manufacturer's protocol (Promega Biotec), in a 50 microliter reaction volume using 25 microliters of wheat germ extract, 1mCi/ml of ³⁵S cysteine (ICN-1000Ci/mMole), 1.17mM potassium acetate, 0.02mM amino acid solution (without cysteine), 100 units of RNasin. ³⁵S cysteine was used because the proteins were found to be low

in methionine. The reaction was performed at 25⁰C for 60 minutes and was stopped by the addition of an equal volume of 2X Laemmli standard treatment buffer (STB) (1XSTB contains 0.0625M Tris buffer pH6.8, 2%SDS, 10% glycerol and 10 micrograms bromophenol blue). An SDS-PAGE system for detecting low molecular weight proteins was used (Guilian and Graham-Hoeffer Catalogue, 1988). This gel system contained 20% acrylamide and 0.5% bis (cross linking agent) and 10% glycerol to slow diffusion and retard migration of the small molecular weight polypeptides. Samples were loaded onto the gel after boiling for three minutes in the presence of 5% 2-mercaptoethanol. Gels were fixed in 40% methanol and 10% acetic acid, dried and autoradiographed.

Immunoprecipitation

In vitro synthesized 9.5K protein of BCV was immunoprecipitated as described by Anderson and Bloebel (1983), using monospecific polyclonal rabbit antibody against MHV A59 9.6K fusion protein (Leibowitz, 1988), a gift from Drs. J. Leibowitz, University of Texas Health Center, Houston, Texas and S. Weiss, University of Pennsylvania, Philadelphia.

Immunofluorescence

Cells were prepared for immunofluorescence by the method of Kaarianen (1983). HRT cells were grown on 8 well Lab Tek tissue culture slides (1cm^2 per well) as described in Chapter 2. BCV Mebus strain was used to infect cells at a MOI of 10. Virus was adsorbed for one hour at 37°C , removed, and DMEM containing 1% fetal calf serum was added and the infection was allowed to continue for 48 hours. Cells were washed twice in phosphate buffered saline (PBS) and fixed with 4% paraformaldehyde (made in 0.1M sodium phosphate (pH 7.0)) for 10 minutes at room temperature followed by two washes with PBS. The slides were stored at 4°C under PBS/0.5% gelatin. Before immunofluorescence, cells were washed twice with PBS and twice with PBS/0.5% gelatin. For internal fluorescence, acetone:methanol (1:1) was added to each well and the slide was kept at -20°C for 10 minutes. The acetone/methanol was removed and cells were washed twice with PBS/0.5% gelatin. After this step the following procedure was used for both surface and internal fluorescence. 100 microliters of 1:300 dilution in PBS of antibody to MHV A59 9.6K protein was added to each well and incubated at 37°C for one hour. The antibody was removed and cells were washed five times with PBS/0.5% gelatin, 10 minutes per wash. 100 microliters of goat anti-rabbit IgG-FITC conjugate (diluted 1:100) was then added and incubated at 37°C for 30 minutes. Cells were

again washed five times with PBS/0.5% gelatin, 10 minutes per wash. The 8 well partition and rubber gasket were removed from the slide and cells were air dried and 10 microliters of PBS/50% glycerol was applied to each well. Coverslips were carefully placed on top and the cells viewed for fluorescence using a Lietz microscope.

Results

Nucleotide sequence between the spike and the matrix protein genes of BCV

From the strong similarities in the spike, matrix and nucleocapsid proteins of MHV and BCV, respectively, both at the level of antigenicity of the proteins (Hogue et al., 1984) and at the level of amino acid sequence (Lapps et al., 1987; Chapter 3), strong similarities between MHV and BCV in both the location and sequence of other genes were anticipated. The map positions of the BCV and MHV S, M, N genes relative to the 3' end of the genome is colinear. The nucleotide sequence between the genes for the spike and matrix protein of BCV is shown in Figure 4.4. This figure also shows the 3' end of the open reading frame for the spike protein gene. Between the spike and matrix protein genes of BCV are 962 bases that, when translated, yield open reading frames for 4.9K, 4.8K, 12.7K and 9.5K proteins. The open reading frames for the 12.7 and 9.5K

Fig 4.4 Nucleotide sequence and deduced amino acid sequence of the region between the spike and matrix protein genes of BCV. The sequence begins with the open reading frame for the 3' end of the spike protein gene (base 1-334), followed by the putative 4.9K, 4.8K, 12.7K and 9.5K proteins. The sequence ends with the first amino acid of the matrix protein. Consensus sequences and the potential N-linked glycosylation site are boxed.

10 20 30 40 50 60 70 80 90 100 110 120
 CCAAAATCAGTGGCACCAGATTGTTCACCTTGATTATATAATGTTTCACTTCTGGACCTACAGATGAATAGTTAGGTACAGGAGCAATAAAGTTTAAATCAGAGCTACATCAA
 Q T S V A P D L S L D Y I N V T F L D L Q D E M N R L Q E A I K V L N Q S Y I N
 S (continued) →

130 140 150 160 170 180 190 200 210 220 230 240
 TCTCAGGACATTTGGTACATATGAGTATTATGTAATAATGGCTTTAATTTGGCTTTGGCTGTAGCTATGCTTTGTTTACTTCTTCATATGCTGTGTGTACAGG
 L K D I G T Y E Y Y V K W P W Y V W L L I G F A G V A M L V L F F I C C C T G

250 260 270 280 290 300 310 320 330 340 350 360
 ATGTGGACTAGTTTAAAGATATGTGGTGGTTGTGTGATGATTATATCTGACACAGAGGTTAGTAAATTAATAACATCAGATGAGTCTGCTTTTGAATTTATGGCTCCTG
 C G T S C F K I C G G C C D D Y T G H Q E L V I K T S H D D
 M T T K F V F D L L A P
 4.9K →

370 380 390 400 410 420 430 440 450 460 470 480
 ACGATATATACATCCCTTCAATCATGTGAAGCTAATTAAGACCCATGAGGTGAGCATATTAATAGCTACCAATGCCCTGCTTTAGTGGTACTGTGTCTTATATACTAG
 D D I L H P F N H V K L I I R P I E V E H I I A T T M P A V

490 500 510 520 530 540 550 560 570 580 590 600
 TAACTGTAAATGCCAATGGCTACAACTAGGCTACAGATTACTAATATTATGCTAGTACTGTTTCTACACAGATTATTAGGCTGTCTTAGGTAATGACACTAGCACAC
 M P M A T T I D G T D Y T N I M P S T V T V L G C S I G I D T S T T
 4.8K →

610 620 630 640 650 660 670 680 690 700 710 720
 TGGTTTACCTGTTTTCACCGTACTAGTTTCCAAAGCATATTAATTTAGGTAGACCTTATACTTAAGCATTAATTAATGCCAAAGTTTCTAAGGTACAGCCCTAGTAATGACATC
 G F T C F S R Y
 M D I
 12.7K →

730 740 750 760 770 780 790 800 810 820 830 840
 TGGAGACCTGAGATTAAATATCTCGTTATACTAAACGGTTTAAATGTCTCAGAAATTAAGATGCTGTGTTTAAATTTAACTATAAATTTCTAAAGTAGGATAITGTAGAGTTCTTAGT
 W R P E I K Y L R Y T N G F N V S E L E D A C F K F N Y K F P K V G Y C R V P S
 →

850 860 870 880 890 900 910 920 930 940 950 960
 CATGCTTGTGCGGTAAATCAAGGTAGCTTTGTGCTACACTCCTTTATGTCGCAATCCAAAGCATTAATGATATAAATTTTGGAGTAATACTGGTTTACAGCAITTCGCTAATACTGTA
 H A W C R N Q G S F C A T L T L Y G K S K H Y D K Y F G V I T G F T A F A N T V

970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080
 GAGGAGCTGTTAACAACTGTTTCTTAGCTTTGCTGTTTACTTTATTTACCTGCGGAGACAGGAGTTAAATGTTTATGGCTGATGCTTTTTCAGACACTGTGTGTATGTGGGCAAT
 E E A V N K L V F L A V D F I T W R R Q E L N V Y G
 M F M A D A Y F A D T V W Y V G Q I
 9.5K →

1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
 AATTTTATAGTGGCAATTTGTTTATGTTTAAATGTTAGTGGCACTTTTGGCACTTTTAAATGTTGTTTCACTTTGGGATGTTGTTAATACCTTAGGACTGCCCCCTCTAT
 I F I V A I C L L V I I V V A F L A T F K L C I Q L C G M C N T L G L S P S I

1210 1220 1230 1240 1250 1260 1270 1280 1290
 TTATGTTTAAATAGAGGTAGGCACTTTTATGATGTTTACAAAGATGTAACCACTGTTTGTGATGGATGACCTTTAGTTAATCCAAACATATG
 Y V F N R G R Q F Y E F Y N D V K P P V L D V D V
 M
 M →

proteins overlap by 11 nucleotides while that of the 4.9K overlaps the spike protein gene by 8 bases. Two consensus intergenic sequences CYAAAC are present between the open reading frames for the spike and matrix proteins (122 bases upstream of the 9.5K ORF and 74 bases upstream of the 12.7K ORF) (Figure 4.4) and one intergenic sequence is present at the 3' end of the spike protein gene (316 bases upstream of the 4.9K ORF) (Figure 4.4). Similar intergenic sequences are present upstream of the nucleocapsid and matrix protein genes (Lapps, 1987). According to the leader primed hypothesis for mRNA synthesis, the intergenic sequences represent the start site for mRNA synthesis which therefore predicts a separate mRNA encoding an ORF for each of these proteins (4.9K and 4.8K in tandem, 12.7K and 9.5K) at its 5' end (Figure 4.5).

Analysis of the open reading frames potentially encoding the 4.9K and 4.8K proteins

The nucleotide sequence between 324 and 455 encodes a 43 amino acid protein with a predicted size of 4.9K that overlaps with the S gene by 23 bases. The initiation codon is in favorable context according to Kozak's rules (1984). The open reading frame for the 4.8K protein is present between nucleotide 491 and 628 and the initiation codon is in favorable context according to Kozaks rules (1984). The hydrophobicity plot (Kyte and Doolittle, 1982)

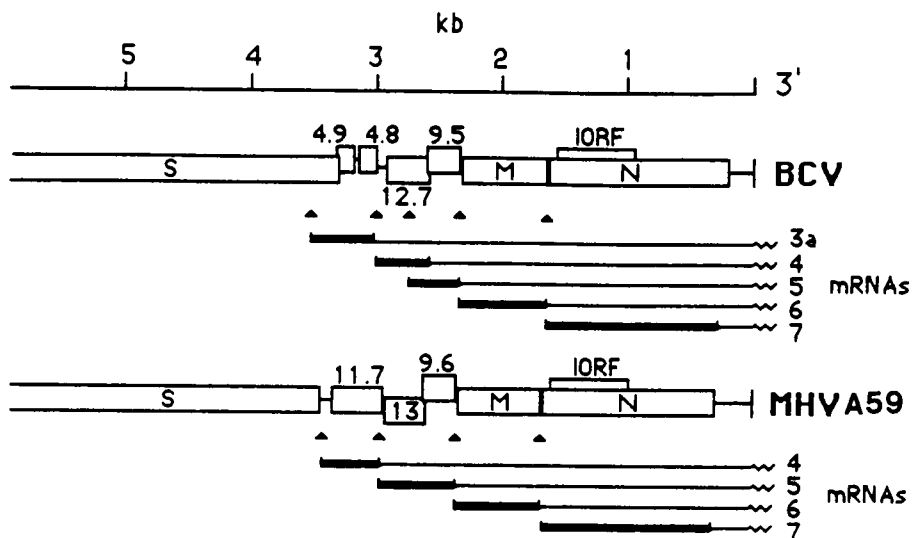


Fig 4.5 Map position of the consensus intergenic sequences in the 3' terminal 3.5 kb of the BCV and MHV A59 genome. The open boxes denote the open reading frames. (▲) represents the consensus intergenic sequences. Also shown are the different mRNAs, their corresponding start sites, and a bar indicating the 5' terminal open reading frame.

obtained from the deduced amino acid sequence is shown in Figure 4.6 and Figure 4.7. The 4.9K and 4.8K ORFs appear to be 5' proximal tandem ORFs on mRNA3a (see below) with the mRNA transcription start site at base 1 (Figure 4.4).

Analysis of the region encoding the open reading frames for the putative 4.9K and 4.8K proteins led to the observation that a single base addition between nucleotide 440 and 454 would yield an open reading frame that encodes a 11K protein. The 11K protein showed amino acid sequence similarity with the 11.7K protein of MHV A59 (34%) (Weiss et al., personal comm.) and with the 15.2K protein of MHV JHM (32%) (Skinner et al., 1985). The deletion of a base is not due to an error during cDNA synthesis or sequencing since this sequence was confirmed by sequencing a separate cDNA clone containing this region of the genome (G6 in Figure 4.1).

Analysis of the open reading frame potentially encoding a 12.7K protein

This sequence contains 329 nucleotides and is found between nucleotides 712 and 1041 (Figure 4.4) and encodes a protein of 109 amino acids with a predicted size of 12.7K. The initiation codon is in a favorable context according to Kozak's rules (1984). The protein appears to be low in methionine content and is a basic protein with a net positive charge of +5 and contains one N-linked

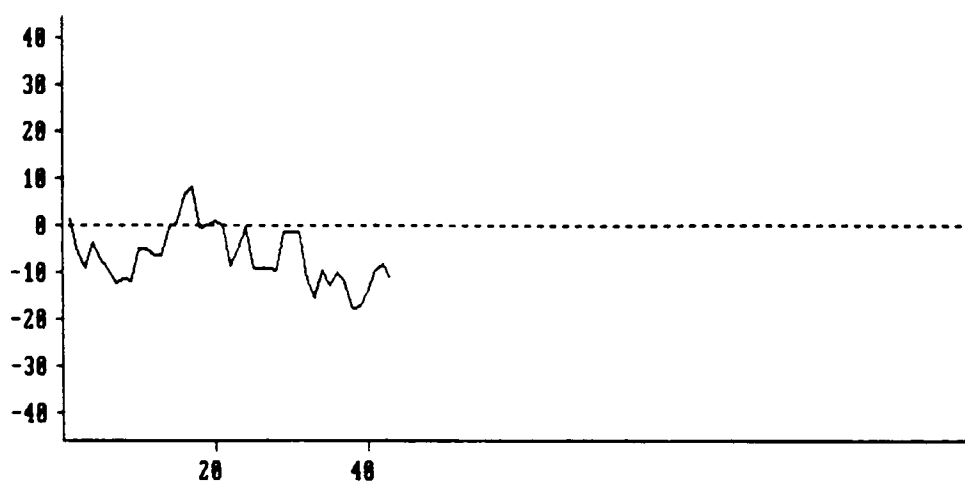


Fig 4.6 Hydrophobicity plot of the putative BCV 4.9K protein. Analysis was performed using the method of Kyte and Doolittle (1982) with a nine amino acid window. Positive values represent hydrophilic regions and negative values represent hydrophobic regions.

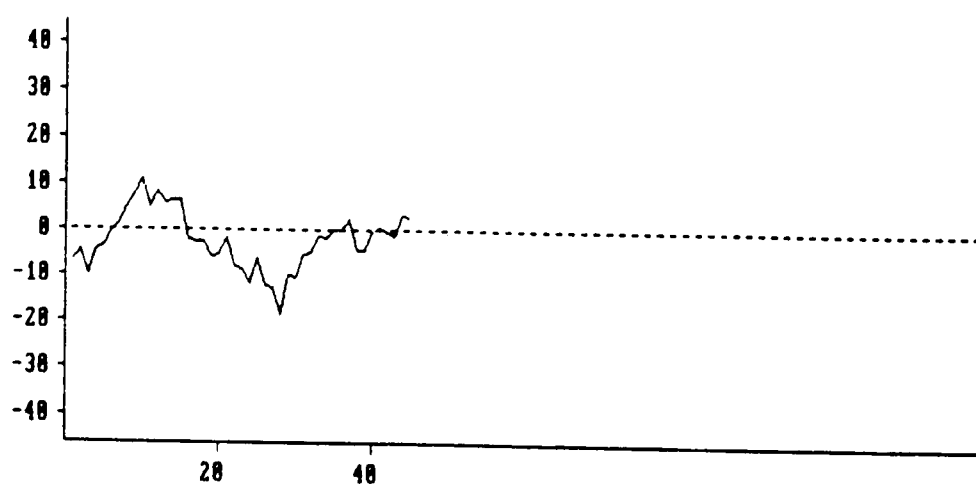


Fig 4.7 Hydrophobicity plot of the putative BCV 4.8K protein. Analysis was performed as described in the legend to Fig 4.6.

glycosylation site. The basic protein may function as a nucleic acid binding protein. The hydrophobicity plot (Figure 4.8) (Kyte and Doolittle, 1982) from the deduced amino acid sequence predicts the protein to be neither hydrophilic nor hydrophobic. However, there is a short hydrophobic region near the C-terminus of the protein which could potentially function as a membrane anchor.

Analysis of the open reading frame potentially encoding a 9.5K protein

An open reading frame extending between nucleotides 1,028 and 1,282 (Figure 4.4), potentially encodes a protein of 84 amino acids having a predicted size of 9.5K. There are two methionine residues that could be used as initiator codons. The second one is in a more favorable context for initiation of translation according to Kozak's rules (1984). The hydrophobicity plot according to Kyte and Doolittle (1982) (Figure 4.9) showed that more than half of the protein from the amino terminus is hydrophobic. This suggests that the protein is probably membrane associated. There is 68% amino acid sequence similarity between the 9.6K protein of MHV A59 and the putative 9.5K protein of BCV at the amino acid level. This property has been used for the detection of the protein in BCV infected cells (discussed below).

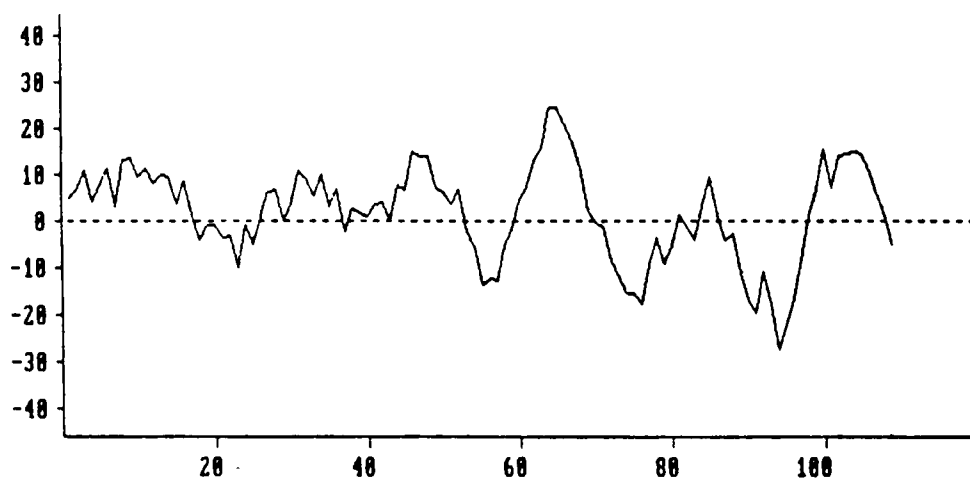


Fig 4.8 Hydrophobicity plot of the putative BCV 12.7K protein. Analysis was performed as described in the legend to Fig 4.6.

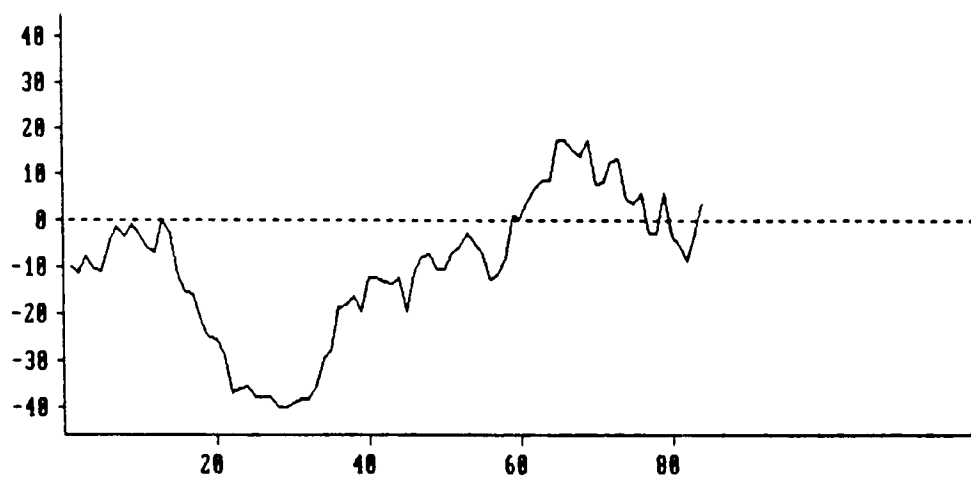


Fig 4.9 Hydrophobicity plot of the putative BCV 9.5K protein. Analysis was performed as described in the legend to Fig 4.6.

Separate mRNAs are made for the 4.9K, 12.7K and 9.5K open reading frames in BCV infected cells

To determine the presence of separate mRNAs for the open reading frames identified in this region of the genome, Northern blot analysis of poly A⁺ RNA from BCV infected cells was done using sequence specific probes. The results of the Northern analysis are shown in Figure 4.10. The probe for the 3' end of the genome hybridized to all mRNA species (1-7) consistent with the 3' coterminal nested set structure of coronavirus mRNAs. The probe specific to the 9.5K open reading frame (between bases 992 and 1,248, Figure 4.4) hybridized to mRNA 5 and all larger species, while the probe specific to the 12.7K open reading frame (between bases 656 and 868, Figure 4.4) hybridized to mRNA 4 and all larger species. The two probes hybridized to separate mRNAs implicating the presence of separate mRNAs for the 9.5K and 12.7K open reading frames, assuming a 3' nested set arrangement of all mRNAs. The probe specific to the region 5' to the 12.7K open reading frame (4.9K and 4.8K) hybridized to mRNA 3a and all species above it, this species of mRNA comigrates with the 28S rRNA and has been confirmed by M. Hofmann as a specific mRNA (unpublished observation).

Evaluating the expression of the putative nonstructural protein genes between S and M protein genes of BCV revealed

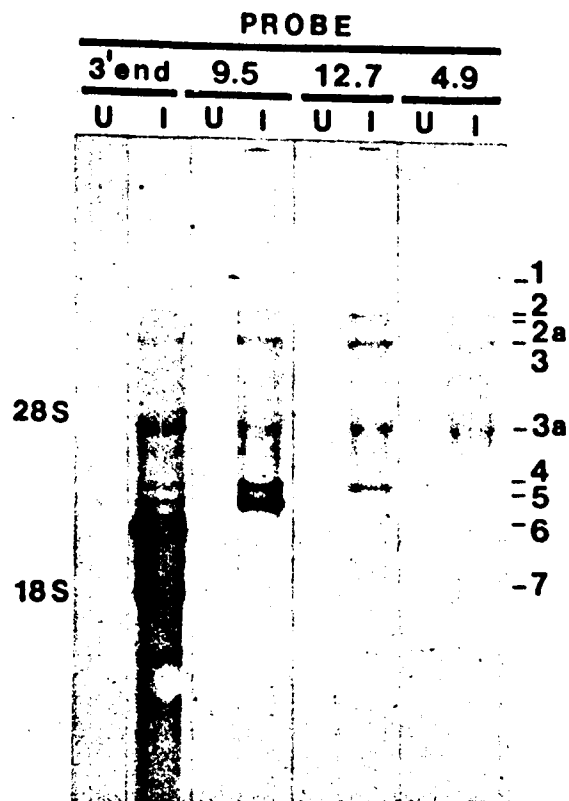


Fig 4.10 Northern blot analysis identifying separate mRNAs for the 4.9K, 12.7K and 9.5K open reading frames. Poly A⁺ RNA from BCV infected and uninfected cells was probed with sequence specific fragments as described in materials and methods. U=RNA from uninfected I=RNA from infected cells.

three striking differences from its close relative, MHV.

these are summarized as follows:

(1) The first of the three open reading frames in MHV (for 15.2K protein in JHM strain or 11.7K protein in A59 strain) is divided by an apparent base deletion into two smaller open reading frames in BCV.

(2) There is an intergenic sequence present in BCV upstream of each of the ORFs for the 4.9K, 12.7K and 9.5K proteins while in MHV (A59 and JHM strains) an intergenic sequence is present upstream of the ORFs for the 11.7K (in A59) or 15.2K (in JHM) and 13K (in A59) or 12.4K (in JHM) proteins (Figure 4.5).

This difference results in the entire region (four putative ORFs) in BCV being expressed on three transcripts whereas in MHV, the region (three putative ORFs) is expressed on two transcripts and is summarized in Table 4.1.

(3) The 9.5K protein of BCV is translated from its own mRNA whereas its MHV counterpart requires downstream initiation of translation (Table 4.1).

Of these, the third difference is the most striking and further analyses were done to determine whether BCV does in fact employ a different transcriptional and translational mechanism for the synthesis of the 9.5K protein. The following were done:

Table 4.1 TRANSCRIPTS ENCODED BETWEEN THE SPIKE AND MATRIX PROTEIN GENES IN BCV AND MHV.

mRNA	BCV	mRNA	MHV A59	mRNA	MHV JHM
	ORF(s) encoded at 5' end		ORF(s) encoded at 5' end		ORF(s) encoded at 5' end
3a	4.9K 4.8K	-	-	-	-
4	12.7K	4	11.7K	4	15.2K
5	9.5K	5	13.0K 9.6K	5	12.4K 10.2K

(1) It was confirmed by immunofluorescence that the 9.5K protein is expressed in BCV infected cells.

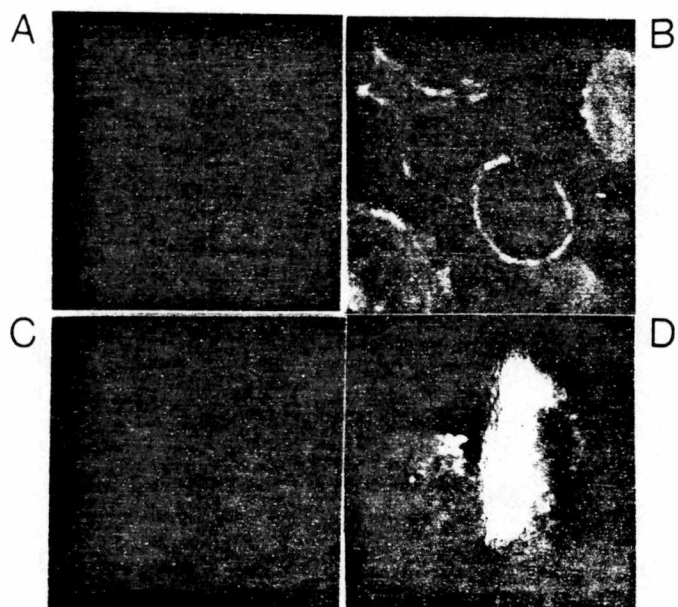
(2) In vitro translation of in vitro-generated transcripts confirmed that the BCV 9.5K protein does not translate from transcripts requiring downstream (internal) initiation of translation, but does translate well when it is the 5' proximal open reading frame. The identity of the 9.5K in vitro translated product was confirmed by immunoprecipitation with the MHV 9.6K antibody.

Since the BCV 9.5K protein was shown to be immunoprecipitated by MHV 9.6K specific antiserum (see below), this antiserum was used to detect the presence of the 9.5K protein in BCV infected cells by immunofluorescence. Internal fluorescence observed in BCV infected cells as compared to no fluorescence in uninfected cells indicated that the 9.5K protein is made in BCV infected cells (Figure 4.11). Furthermore, surface immunofluorescence indicates that the 9.5 K protein is transported to the cell surface and suggests that this is a membrane associated protein.

In vitro translation analysis of the BCV 9.5K protein demonstrates that it is not made by downstream (internal) initiation of translation

Sequence and Northern analyses provide evidence for the presence of two separate transcripts (mRNA4&5) encoding the

Fig 4.11 Surface and internal fluorescence of the 9.5K protein in BCV infected cells.
Surface fluorescence A=uninfected cells, B=BCV infected cells.
Internal fluorescence C=uninfected cells, D=BCV infected cells.



12.7K and 9.5K open reading frames respectively. Synthetically generated transcripts encoding these open reading frames were translated in wheat germ extracts to determine whether the 9.5K protein can be made by downstream initiation of translation. The transcript made from clone AB6 contained the open reading frames for the 12.7K, 9.5K and M proteins in tandem, while the transcript made from clone AB9.6 contained the 9.5K open reading frame as the 5' proximal open reading frame followed by the M open reading frame. Results of in vitro translation are shown in Figure 4.12. A 12.7K product but no 9.5K protein is made from the AB6 transcript whereas a 9.5K product was made from transcripts from clone AB9.6 (Figure 4.12, lanes 1 and 2). A band of approximately the size of the unglycosylated form of the matrix protein is seen in Figure 4.12 lane 2. However at this time the identity of this band has not been determined. These results demonstrate the absence of in vitro downstream initiation of translation of the BCV 9.5K protein.

Antibody to the MHV 9.6K protein immunoprecipitated the 9.5K product made in vitro establishing that it is an immunologically cross reactive protein (Figure 4.12, lane 3).

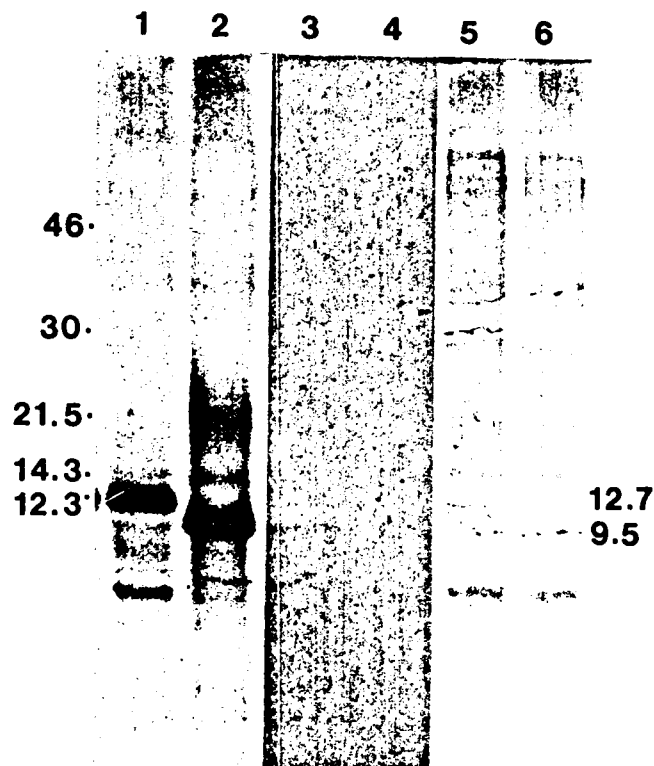


Fig 4.12 In vitro translation of the 12.7K and 9.5K proteins, from in vitro-generated transcripts using T7 polymerase, in wheat germ extracts. Lane 1, AB6 transcript which encodes 12.7K, 9.5K and matrix proteins in tandem. Lane 2, AB9.6 transcript which encodes 9.5K and matrix proteins in tandem. Lane 3, BCV 9.5K protein made in vitro and immunoprecipitated with the antibody to the MHVA59 9.6K protein. Lane 4, BCV 9.5K protein made in vitro and immunoprecipitated with preimmune serum. Lane 5, transcript from pGEM42 that contains no insert. Lane 6, no RNA added to the system.

Discussion

Similarities and differences between BCV and the antigenically related MHV were determined by nucleotide and deduced amino acid sequence analyses of open reading frames encoded between the spike and matrix protein genes of BCV.

Two small open reading frames encoding putative 4.9K and 4.8K proteins are present immediately downstream of the C-terminus of the S protein. Northern analysis indicates the presence of a mRNA (mRNA3a) that probably initiates at the intergenic sequence found upstream of the 4.9K initiation codon, thereby encoding the open reading frames for the 4.9K and 4.8K proteins at its 5' end and suggesting that the 4.8K protein is translated as a downstream open reading frame. The MHV counterparts 11.7K (A59) and 15.2K (JHM) proteins have the same relative map position on the genome as the BCV 4.9K and 4.8K proteins. These small proteins have not as yet been detected in the virus infected cell.

The deduced amino acid sequence of the putative BCV 12.7K protein compared with the sequence of its MHV counterparts 13K (A59) and 12.4K (JHM) proteins showed 48% amino acid sequence identity. All these proteins are basic and contain only one methionine, and have the same relative map position on the genome (Figure 4.5). These

proteins could probably have a similar function in the virus infected cell. However they have so far not been detected in the infected cell.

The putative BCV 9.5K protein showed remarkable sequence identity with MHV A59 9.6K and MHV JHM 10.2K proteins. After aligning the deduced amino acid sequence of these proteins there is 68% sequence identity at the amino acid level. The putative BCV 9.5K protein, MHV A59 9.6K protein (Budcillo et al., 1987), MHVJHM 10.2K (Skinner et al., 1985), IBV 12.4K protein (Bourisnell et al., 1985) and TGEV 9.2K protein (Wesley et al., 1989) were all found to have a hydrophobic amino terminus. All of these proteins are found to have the same relative map position on their respective genomes (Figure 1.2). The conserved hydrophobic pattern of these polypeptides suggest that each protein might perform an analogous function in the replication of coronaviruses.

The finding of separate mRNAs by Northern analyses for the 12.7K and 9.5K proteins in BCV, together with the results of the in vitro translation experiments which demonstrated no downstream initiation of translation for the synthesis of the 9.5K protein, suggest strongly that BCV employs a different transcriptional and translational strategy than does the 9.6K protein of MHV.

It is not clear why downstream initiation of translation is absent in BCV. The initiation codon for the

BCV 12.7K protein is in favorable context for initiation of translation (Kozak, 1984), while that of MHV A59 13K protein is not in optimal context (Kozak, 1984) for translation initiation. As proposed by Kozak (1986; 1987), the mechanisms that allow some mRNAs to direct the synthesis of more than one protein are (1) leaky scanning of ribosomes (2) an ability to establish reinitiation of translation. One can speculate that the presence of a suboptimal translation initiation codon in the MHV 13K protein allows the ribosomes (leaky scanning) to reach the start of the second cistron thereby allowing downstream initiation of translation. If this is true for MHV, the lack of downstream initiation in BCV could be attributed to the presence of the optimal initiation codon in BCV 12.7K protein which prevents ribosomes from reaching the second cistron. For reinitiation to take place the AUG of the second cistron should be preceded by the termination codon for the first open reading frame. Since the two open reading frames for the 13K and 9.6K proteins of MHV and their counterparts in BCV overlap, reinitiation probably is not a mechanism by which downstream initiation of translation takes place.

Another possible mechanism by which downstream initiation of translation in MHV takes place is by internal entry of ribosomes similar to picornavirus mRNA translation which does not use the ribosomal scanning

mechanism (Jang et al., 1989). However no consensus sequence for internal entry of ribosomes has yet been identified to determine if it is present in MHV and absent in BCV.

In this study, the antibody to the MHV A59 9.6K protein was shown to immunoprecipitate the 9.5K protein of BCV made in vitro supporting the fact that MHV and BCV are antigenically related (Hogue et al., 1984). The 9.5K protein was detected in BCV infected cells, but the full extent of its exact cellular localization was not determined. It was shown to be present both in the cytoplasm and on the surface of infected cells. There is no evidence as yet whether the 9.5K BCV protein is part of the virion. Similar small hydrophobic proteins have been identified in other RNA viruses, for example, the M2 protein of Influenza A (Zebedee et al., 1985), SH protein of SV5 (Hiebert et al., 1985) and the SH protein of mumps virus (Elango et al., 1989). The BCV 9.5K protein is comparable to the M2 protein of Influenza A virus in that both are hydrophobic and are found on the infected cell surface. It may, as a result, be a target for cytotoxic T cells.

Further work will be required to determine the significance of the different transcriptional and translational strategies utilized by BCV and MHV. It will now be possible to characterize the ORFs that have been

identified and the 9.5K protein can be further characterized to determine its function in BCV replication.

Speculation on the functions of nonstructural proteins of coronaviruses

The identification of the number and functions of nonstructural (NS) proteins in virus infected cells has been slow and difficult. The main reason for this is that they are synthesized in low amounts which makes their detection and purification very difficult. This contrasts with the structural proteins which are synthesized in relatively large amounts. Information obtained from sequence analysis of viral genomes has provided insight into genome structure and on the open reading frames encoded on the genome. Often knowledge of the primary structure of the predicted products helps in assigning putative functions. The availability of mutants or antibodies to the nonstructural proteins will aid in the identification of these proteins as well as their functions. Information on nonstructural proteins of other viruses are used to speculate functions for the nonstructural proteins of coronaviruses, some of which are discussed below.

1. Enzymatic function. RNA polymerase activity has been assigned to one of the nonstructural proteins of coronaviruses. Other enzymatic functions could be assigned

to NS proteins, for example, proteases that may be involved in the processing of proteins as observed in Sindbis virus (Ding et al., 1989), and nucleic acid modifying enzymes such as RNA methyl transferases for capping genome RNA and mRNAs (reported in Sindbis virus by Mi et al., 1989).

2. Scaffolding function. Nonstructural proteins have been shown to play a role in virion assembly, for example, in rotavirus assembly, nsp28 is involved in virus assembly (Meyer et al., 1989). During coronavirus assembly the RNA-nucleocapsid complex aligns at the rough endoplasmic reticulum (RER). The site of budding is presumably governed by the matrix protein in the RER, but so far no interaction between the nucleocapsid protein and the matrix protein has been identified. A nonstructural protein could mediate this interaction and trigger the budding event.

3. Molecular Chaperones. Nonstructural proteins could play a role in the synthesis and processing of other proteins, for example, in retaining the structural proteins at the RER until the virions are assembled (review Ellis and Hemmingsen, 1989).

4. Regulatory molecules. Nonstructural proteins could function as regulatory molecules such as those involved in (a) the switch from transcription to replication (reported in Sindbis virus by Sawicki and Sawicki, 1986), (b) determining the synthesis of other proteins, for example,

binding to the 5'noncoding region of mRNAs thereby enhancing or arresting translation, (c) stabilizing mRNAs, for example, as ssRNA binding proteins (Reovirus-Wiener et al., 1989).

REFERENCES

REFERENCES

- Anderson, D.J., G. Blobel. 1983. Immunoprecipitations of proteins from cell free translations. In *Methods in Enzymology*. (S. Fleischer and B. Fleischer) 96:111-120.
- Armstrong, J., S. Smeekeens, and P. Rottier. 1983. Sequence of the nucleocapsid gene from MHV A59. *Nucl. Acid. Res.* 11:833-891.
- Armstrong, J., H. Niemann, S. Smeekeens, P. Rottier, and G. Warren. 1984. Sequence and topology of a model intracellular membrane protein, E1 glycoprotein from a coronavirus. *Nature*. 308:751-752.
- Baltimore, D. 1971. Expression of animal virus genomes. *Bacteriol Rev.* 35: 235-241.
- Baric, R.S., S. A. Stohlman, and M. M. C. Lai. 1983. Characterization of replicative intermediate RNA of mouse hepatitis virus: Presence of leader RNA sequence on nascent chains. *J. Virol.* 48:633-640.
- Baric, R. S., S. A. Stohlman, M. K. Razavi, M. M. C. Lai. 1985. Characterization of leader related RNAs in coronavirus infected cells. Further evidence for leader primed mechanism of transcription. *Virus Res.* 3:19-33.
- Baric, R. S., C. K. Shieh, S. A. Stohlman, and M. M. C. Lai. 1987. Analysis of intracellular small RNAs of mouse hepatitis virus . Evidence for discontinuous transcription. *Virology*. 156:342-354.
- Binns, M. M., M. E. G. Boursnell, F. M. Tomley, T. D. K. Brown. 1986. Comparison of the spike precursor sequences of coronavirus IBV strains M41 and 6/82 with that of IBV Beaudette. *J. Gen. Virol.* 67: 2825-2831.
- Binns, M. M., M. E. G. Boursnell, F. M. Tomley, T. D. K. Brown. 1985. Cloning and sequencing of the gene encoding the spike protein of coronavirus IBV. *J. Gen. Virol.* 66:719-726.
- Boursnell, M. E. G., T. D. K. Brown, M. M. Binns. 1984. Sequence of the membrane protein gene from avian coronavirus IBV. *Virus Res.* 1:303-313.
- Boursnell, M. E. G., M. M. Binns, T. D. K. Brown. 1985a. Sequencing of the coronavirus IBV genomic RNA: three open

reading frames in the 5' unique region of mRNAD. J. Gen. Virol. 63:2253-2258.

Boursnell, M. E. G., M. M. Binns, I. J. Foulds, T. D. K. Brown. 1985b. Sequences of the nucleocapsid genes from 2 strains of avian infectious bronchitis virus. J. Gen. Virol. 66:573-580.

Boursnell, M. E. G., T. D. K. Brown, I. J. Foulds, P. F. Green, F. M. Tomley, and M. M. Binns. 1987. Completion of the sequence of the genome of the coronavirus Avian Infectious Bronchitis Virus. J. Gen. Virol. 68:57-77.

Brayton, P. R., R. G. Ganges, S. A. Stohlgman. 1981. Host nuclear function and murine hepatitis virus replication. J. Gen. Virol. 56:457-460.

Brayton, P. R., M. M. C. Lai, C. D. Patton, S. A. Stohlgman. 1982. Characterization of two RNA polymerase activities induced by mouse hepatitis virus. J. Virol. 42:847-853.

Brayton, P. R., S. A. Stohlgman, M. M. C. Lai. 1984. Further characterization of mouse hepatitis virus RNA dependent RNA polymerase. Virology. 133:197-201.

Bredenbeek, P. J., J. F. Noten, J. A. Lestra, M. C. Horzinek, B. A. M. van der Zeijst, W. J. M. Spaan. 1986. The nucleotide sequence of the extreme 5' end of the avian coronavirus genome; implications for the discontinuous mRNA synthesis. Nucl. Acids. Res. 14:7806.

Bredenbeek, P. J., J. Charite, J. F. Noten, W. Luytjes, M. C. Horzinek, B. A. M. van der Zeijst, W. J. M. Spaan. 1987. Sequences involved in the replication of coronaviruses. Adv. Exp. Med. and Biol. 218:65-72.

Bridger, J. C., E. O. Caul, Egglestone. 1978. Replication of an enteric bovine coronavirus in intestinal organ cultures. Arch. Virol. 57:43-51.

Budzilowicz, C. J., S. P. Wilczynski, S. R. Weiss. 1985. Three intergenic regions of coronavirus mouse hepatitis virus strain A59 genome RNA contain a common nucleotide sequence that is homologous to the 3' end of viral mRNA leader RNA sequence. J. Virol. 53:834-840.

Budzilowicz, C. J., and S. R. Weiss. 1987. In vitro synthesis of two polypeptides from a nonstructural gene of coronavirus MHVA59 strain. Virology. 157:509-515.

- Cavanagh, D. 1983a. Coronavirus IBV: further evidence that the surface projections are associated with two glycopolypeptides. *J. Gen. Virol.* 64:1787-1791.
- Cavanagh, D. 1983b. Coronavirus IBV: structural characterization of the spike protein. *J. Gen. Virol.* 64:2577-2583.
- Cavanagh, D., P. J. Davis. 1986. Coronavirus IBV: removal of spike glycopolypeptide S1 by urea abolishes infectivity and hemagglutination but not attachment to cells. *J. Gen. Virol.* 67:1443-1448.
- Cavanagh, D., P.J. Davis, A. P. A. Mockett. 1988. Amino acids within hypervariable region 1 of avian coronavirus IBV (Massachusetts serotype) spike glycoprotein are associated with neutralization epitopes. *Virus Res.* 11: 141-150.
- Compton, J. R., D. B. Rogers, K. V. Holmes, D. Fertsch, J. Remenick, J. J. McGowan. 1987. In vitro replication of mouse hepatitis virus A59. *J. Virol.* 61:1814-1920.
- Crouch, C. F., H. Bielefeldt Ohman, T. C. Watts, L. A. Babuik. 1985. Chronic shedding of BCV antigen-antibody complexes by clinically normal cows. *J. Gen. Virol.* 66:1489-1500.
- DeGroot, R. J., J. A. Lenstra, W. Luytjes, H. G. M. Niesters, M. C. Horzinek, B. A. M. van der Zeijst, and W. J. M. Spaan. 1987a. Sequence and structure of the coronavirus peplomer protein. *Coronavirus. Adv. in Exp. Med and Biol.* 218:31-33.
- DeGroot, R. J., J. Maduro, J. A. Lenstra, M. C. Horzinek, B. A. M. van der Zeijst and W. J. M. Spaan. 1987b. cDNA cloning and sequence analysis of the gene encoding the peplomer protein of the feline infectious peritonitis virus. *J. Gen. Virol.* 68:2639-2646.
- DeGroot, R. J., W. Luytjes, M. C. Horzinek, B.A. M. van der Zeijst, W. J. M. Spaan, J. A. Lenstra. 1987c. Evidence for coiled coil structure in spike proteins of coronavirus. *J. Mol. Biol.* 196:963-966.
- Dennis, D. E., D. A. Brian. 1981. Coronavirus cell associated RNA dependent RNA polymerase. In *Biochemistry and Biology of coronaviruses*. (V. ter. Meulen, S. Siddell, H. Wege, eds) p 155-170. Plenum, New York.

- Dennis, D. E., D. A. Brian. 1982. RNA dependent RNA polymerase activity in coronavirus infected cells. *J. Virol.* 42:153-164.
- Deregt, D., M. Sabara, L. A. Babuik. 1987. Structural proteins of bovine coronavirus and their intracellular processing. *J. Gen. Virol.* 68:2863-2877.
- Deregt, D., L. A. Babuik. 1987. Monoclonal antibodies to bovine coronavirus: Characteristics and topographical mapping of neutralizing epitopes on E2 and E3 glycoproteins. *Virology.* 161:410-420.
- Deregt, D., M. D. Parker, G. C. Cox, L. A. Babuik. 1989a. Mapping of neutralizing epitopes to fragments of bovine coronavirus E2 protein by proteolysis of antigen-antibody complex. *J. Gen. Virol.* 70:647-658.
- Deregt, D., G. A. Gifford, M. K. Ijaz, T. C. Watts, J. E. Gichrist, D. M. Haines, L. A. Babuik. 1989b. Monoclonal antibodies to bovine coronavirus glycoproteins E2 and E3: demonstration of in vivo virus neutralizing activity. *J. Gen. Virol.* 70:993-998.
- Dowling, P., C. Giorgi, L. Roux, L. Dethlefsen, M. Galantowicz, B. Blumberg, D. Kolakofsky. 1983. Molecular cloning of the 3' proximal third of Sendai Virus genome. *Proc. Natl. Acad. Sci.* 80:5213-5216.
- Dubois-Dalcq, M., K. V. Holmes, B. Rentier. 1984. *Assembly of enveloped RNA viruses* (Springer Verlag Wien New York): 100-110.
- Ebner, D., T. Raabe, S. G. Siddell. 1988. Identification of the Coronavirus MHV JHM mRNA 4 product. *J. Gen. Virol.* 69:1041-1051.
- Elango, N., J. Kovamees, T. M. Varsanyi, E. Norby. 1989. mRNA sequence and deduced amino acid sequence of the mumps virus small hydrophobic protein gene. *J. Virol.* 63:1413-1415.
- Ellis, J. R., S. M. Hemmingsen. 1989. Molecular chaperones: proteins essential for biogenesis of some macromolecular structures. *Trends. Biochem. Sci.* 14:339-342.
- Fenner, F., P. A. Bachmann, E. P. J. Gibbs, F. A. Murphy, M. J. Studdert, D. O. White. 1987. *Veterinary Virology.* Academic Press Inc (London).

- Garoff, H. 1985. Using recombinant DNA techniques to study protein targeting in the eukaryotic cell. *Ann. Rev. Cell Biol.* 1:403-445.
- Garwes, D. J., M. H. Lucas, D. A. Higgins, B. V. Pike, S. F. Cartwright. 1978. Antigenicity of structural components from porcine transmissible gastroenteritis virus. *Vet. Microbiol.* 3:179-190.
- Garwes, D. J., D. J. Reynolds. 1981. The polypeptide structure of canine coronavirus and its relationship to porcine transmissible gastroenteritis virus. *J. Gen. Virol.* 52:153-157
- Gubler, U., B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* 25:263-269.
- Guillan, G., J. Graham. 1988. The electrophoretic separation of low molecular weight polypeptides in polyacrylamide gels. 1988. Hoeffer Scientific Instruments catalogue.
- Guy, J. S., D. A. Brian. 1979. Bovine coronavirus genome. *J. Virol.* 29:293-300.
- Hanahan, D., 1983. Studies on transformation of *E. coli*. with plasmids. *J. Mol. Biol.* 166:557-580.
- Haspel, M. V., K. V. Holmes, R. M. Welsh. 1981. Natural cell mediated cytotoxicity against mouse hepatitis infected cells. *Am. Soc. Microbiol. Ann. Meet. Abstract No. E9.*
- Helenius, A., J. Kartenbeck, K. Simons, E. Fries. 1980a. On the entry of Semliki forest virus into BHK21 cells. *J. Cell Biol.* 84:404-420.
- Helenius, A., M. Marsh, J. White. 1980b. The entry of viruses into animal cells. *Trends Biochem. Sci.* 5:104-106.
- Henikoff, S., 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene.* 335-359.
- Hiebert, S. W., R. G. Patterson, R. A. Lamb. 1985. Identification and predicted sequence of a previously unrecognized small hydrophobic protein SH of paramyxovirus Simian virus. *J. Virol.* 55:744-751.

- Hilton. A., L. Mizzen, G. Macintyre, S. Cheley, R. Anderson. 1986. Translational control in murine hepatitis virus infection. *J. Gen. Virol.* 67:923-932.
- Hogue, B., B. King, D. Brian. 1984. Antigenic relationships among proteins of Bovine coronavirus, Human respiratory coronavirus OC43, MHV-A59. *J. Virol.* 51:384-388.
- Hogue, B. 1986. PhD dissertation, University of Tennessee.
- Hogue, B. G., D. A. Brian. 1986. Structural proteins of human respiratory coronavirus OC43. *Virus Res.* 5:131-144.
- Hogue, B., T. Kienzle, D. Brian. 1989. Synthesis and processing of BCV hemagglutinin protein. *J. Gen. Virol.* 70:345-352.
- Holmes, K. V., E. W. Doller, L. S. Sturman. 1981a. Tunicamycin resistant glycosylation of a coronavirus glycoprotein: Demonstration of a novel type of viral glycoprotein. *Virology.* 115:334-344.
- Holmes, K. V., E. W. Doller, J. N. Behnke. 1981b. Analysis of the functions of coronavirus glycoproteins by differential inhibition of synthesis with tunicamycin. In *Biochemistry and Biology of coronaviruses* (ter Meulen, S. Siddell, H. Wege. eds) 133-42. New York: Plenum Press.
- House, J., 1978. Economic impact of Rotavirus and other neonatal disease agents of animals. *J. Am. Vet. Med. Assoc.* 173:573-576.
- Jacobs, L., B. A. M. van der Zeijst, M. C. Horzinek. 1986. Characterisation and translation of TGEV virus mRNAs. *J. Virol.* 57:1010-1015.
- Jacobs, L., R. DeGroot, B. A. M. Van der Zeijst, M. C. Horzinek, W. Spaan. 1987. Nucleotide sequence of the peplomer gene of porcine transmissible gastroenteritis virus (TGEV): comparison with feline infectious peritonitis virus (FIPV). *Virus Res.* 8:367-371.
- Jang, S. K., M. V. Davies, R. J. Kaufman, E. Wimmer. 1989. Initiation of protein synthesis by internal entry of ribosomes into the 5' nontranslated region of encephalomyocarditis virus RNA in vivo. *J. of Virol.* 63:1651-1660.
- Kaariainen, L., I. Virtanen, J. Saraste, S. Keranen. 1983. Transport of virus glycoproteins, use of temperature

sensitive mutants and organelle specific lectins. Meth. in Enzymol. 96:453-465.

Kapke, P. A., D. A. Brian. 1986. Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid gene. Virology. 151:41-49.

Kapke, P. A., F. Y. T. Tung, B. G. Hogue, D. A. Brian, R. D. Woods, R. Wesley. 1988. The amino-terminal signal peptide on the Porcine transmissible gastroenteritis coronavirus matrix protein is not an absolute requirement for membrane translocation and glycosylation. Virology. 165:367-376.

Kapke, P. A., F. Y. T. Tung, D. A. Brian. 1988. Nucleotide sequence between the peplomer and matrix protein genes of the porcine transmissible gastroenteritis coronavirus identifies three large open reading frames. Virus Genes. 2:3:293-294.

Keck, J. G., B. G. Hogue, D. A. Brian, M. M. C. Lai. 1988. Temporal regulation of Bovine coronavirus RNA synthesis. Virus. Res. 9:343-356.

Kees, U., P. Krammer. 1984. Most influenza A virus specific memory cytotoxic T-lymphocytes react with antigenic epitopes associated with internal viral determinants. J. Exp. Med. 151:365-377.

Khandjian, E. W., C. Meric. 1986. A procedure for northern blot analysis of native RNA. Anal. Biochem. 159:227-232.

Kienzle, T. E. 1989. PhD dissertation, University of Tennessee.

King, B., D. A. Brian. 1982. Bovine coronavirus structural proteins. J. Virol. 42:700-707.

King, B., B. J. Potts, D. A. Brian. 1985. BCV hemagglutinin protein. Virus Res. 2:53-59.

Kozak, M., 1984. Compilation and analysis of sequences upstream from the translation start site in eukaryotic mRNAs. Nucl. Acid. Res. 12:857-872.

Kozak, M. 1986. Regulation of protein synthesis in virus infected animal cells. Adv. in Virus. Res. 31:229-292.

Kozak, M. 1987. An analysis of 5' noncoding sequences from 699 vertebrate mRNAs. Nucl. Acid. Res. 15:8125-8148.

- Kraft, R., J. Tardiff, K. Krauter, L. Leinward. 1988. Using miniprep DNA for sequencing double strand template with sequenase. *Biotechniques*. 6:544-549.
- Krieg, P. A., D. A. Melton. 1984. Functional messenger RNAs are produced by SP6 in in vitro transcription of cloned cDNAs. *Nucl. Acid. Res.* 12:7057-7070.
- Kusters, J. G., H. G. M. Nieters, J.A. Lenstra, M. C. Horzinek and B. A. M. Van der Zeijst. 1989. Phylogeny of antigenic variants of avian coronavirus IBV. *Virology*. 169:217-221.
- Kyte, J., R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105-132.
- Laemmli, U. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 227:680-685.
- Lai, M., P. Brayton, R. Armen, C. Patton, C. Pugh, S. Stohlman. 1981. Mouse Hepatitis virus A59: mRNA structure and genetic localization in the sequence divergence from hepatotropic strain MHV-3. *J. Virol.* 39:823-834.
- Lai, M., C. Patton, S. Stohlman. 1982a. Further characterization of mRNAs of Mouse Hepatitis virus: presence of common 5' end nucleotides. *J. Virol.* 41:557-565.
- Lai, M., C. Patton, S. Stohlman. 1982b. Replication of Mouse Hepatitis virus: negative strand RNA and replicative form RNA are genome length. *J. Virol.* 44:487-492.
- Lai, M. M. C., C. D. Patton, R.S. Baric, S. A. Stohlman. 1983. Presence of leader RNA sequences in mRNA of mouse hepatitis virus. *J. Virol.* 46:1027-1033.
- Lai, M. M. C., R. S. Baric, P. R. Brayton, S. A. Stohlman. 1984. Characterisation of leader RNA sequences on the virion and mRNAs of Mouse Hepatitis virus, a cytoplasmic RNA virus. *Proc. Natl. Acad. Sci. U.S.A.* 81:3626-3630.
- Lai, M. M. C. 1986. Coronavirus leader primed transcription an alternative mechanism to RNA splicing. *Bioessays*. 5:257-260.
- Lai, M. M. C., S. Makino, L. Soe, C. Shieh, J. Keck, J. Fleming. 1987. Coronavirus: A jumping RNA transcription. *Cold Spring Harbor Sym. Quan. Biol.* 52:359-365.

Lapps, B., D. A. Brian. 1985. Oligonucleotide fingerprints of antigenically related bovine coronavirus and human coronavirus OC43. Arch. Virol. 86:101-108.

Lapps, W. 1986. PhD dissertation, University of Tennessee.

Lapps, W., B. G. Hogue, D. A. Brian. 1987. Sequence analysis of BCV nucleocapsid and matrix protein genes. Virology. 157:47-57.

Laude, H., D. Rasschaert, J. C. Huet. 1987. Sequence and N-terminal processing of the transmembrane protein E1 of the coronavirus transmissible gastroenteritis virus. J. Gen. Virol. 68:1687-1693.

Leibowitz, J. L., S. R. Weiss, E. Paavola, C. W. Bond. 1982. Cell free translation of murine coronavirus RNA. J. Virol. 43:905-913.

Leibowitz, J. L., S. Perlman, G. Weinstock, J. R. DeVries, C. Budzillowicz, S. R. Weiss. 1988. Detection of a Murine Coronavirus nonstructural protein encoded in a downstream open reading frame. Virology. 164:156-164.

Lizardi, P., 1982. Binding and recovery of DNA and RNA using S&S NA45 DEAE-membrane. Scheicher and Shuell Sequences Application and uptake.

Lizardi, P. M. 1983. methods for preparation of messenger RNA. Methods in Enzymology. 96:24-38.

Luytjes, W., L. S., Sturman, P. J. Bredenbeck, J. Charite, B. A. M. van der Zeijst, M. C. Horzinek, W. J. M. Spaan. 1987. Primary structure of the glycoprotein E2 of Coronavirus MHV A59 and identification of the trypsin cleavage site. Virology. 161:479-487.

Machamer, C. E., J. K. Rose. 1987. A specific transmembrane domain of a coronavirus E1 glycoprotein is required for its retention in the Golgi region. J. Cell. Biol. 105:1205-1214.

Mahy, B., S. Siddell, H. Wege, V. ter Meulen. 1983. RNA dependent RNA polymerase activity in murine coronavirus infected cells. J. Gen. Virol. 64:103-111.

Makino, S., J. O. Fleming, J. G. Keck, S. A. Stohlman, M. M. C. Lai. 1987. RNA recombination of coronaviruses: localization of neutralizing epitopes and neuropathogenic

determinants of carboxyl terminus polypeptides. Proc. Natl. Acad. Sci. U.S.A. 84:6567-6571.

Maniatis, T., E. Fritsch, J. Sambrook. 1982. Molecular cloning laboratory manual. Cold Spring Harbour Laboratory, Cold Spring Harbour. N.Y.

Martin, S., R. J. Courtney, G. Fowler, B. T. Rouse. 1988. Herpes simplex virus type specific cytotoxic T-lymphocytes recognise virus nonstructural proteins. J. Virol. 62:2265-2273.

Massalski, A., M. Coulter-Mackie, S. Dales. 1981. Assembly of mouse hepatitis virus strain JHM. In Biochemistry and Biology of Coronaviruses (V ter Meulen, S. Siddell, H. Wege. eds) 111-118. New York. Plenum Press.

Matsudaira, P., 1987. Sequence from picomole quantities of proteins electrophoretically transferred onto polyvinylidene difluoride membranes. J. Biol. Chem. 262:10035-10038.

Maxam and Gilbert, 1980. Sequencing end labeled DNA with base specific chemical cleavages. In Methods in Enzymology (L. Grossman and K. Moldave, eds) 65:499-560. Academic Press, Orlando, Fl.

McIntosh, K. 1974. Coronavirus. A comparative review. Curr. Top. Micro. Immunol. 63:85-129.

Mebus, C., L. Newman, E. Stair, Jr. 1975. Scanning electron, light and immunofluorescent microscopy of intestine of gastrobiotic calf infected with diarrhoeal coronavirus. Am. J. Vet. Res. 36:1719-1725.

Messing, J., J. Viera. 1982. A new pair of M13 vectors for selecting either DNA strand of double digest restriction fragments. Gene. 19:269-276.

Mi, S., R. Durbin, H. V. Huang, C. M. Rice, V. Stollar. 1989. Association of the Sindbis virus RNA methyl transferase activity with the nonstructural protein nsP1. Virology. 170:385-391.

Moreou, K. 1982. MS thesis, University of Tennessee.

Myer, J. C., C. C. Bergmann, R. Bellamy. 1989. Interaction of Rotavirus cores with the nonstructural glycoprotein NS28. Virology 171:98-107.

- Niemann, H., H. D. Klenk. 1981. Coronavirus glycoprotein E1, a new type of viral glycoprotein. *J. Mol. Biol.* 153:993-1010.
- Oshiro, L. S., J. Schreble, E. H. Lennette. 1978. Electron microscopic studies of a coronavirus. *J. Gen. Virol.* 12:161-168.
- Pachuk, C. J., P.J. Bredenbeek, P. W. Zottick, W. J. M. Spaan, S. R. Weiss. 1989. Molecular cloning of the gene encoding the putative polymerase of mouse hepatitis coronavirus strain A59. *Virology.* 171:141-148.
- Parker, M., G. Cox, D. Dereg, D. Fitzpatrick, L. Babuik. 1989. Cloning and in vitro expression of the gene for the E3 hemagglutinin glycoprotein of BCV. *J. Gen. Virol.* 70:155-164.
- Patterson, S., R. W. Bingham. 1976. Electron microscope observations on the entry of avian infectious bronchitis virus into susceptible cells. *Arch. Virol.* 53:267-276.
- Peacock, S. L., C. M. McIver, J. J. Monahan. 1981. Transformation of *E. coli* using homopolymer linked plasmid chimeras. *Biochim. Biophys. Acta.* 655:243-250.
- Pfleiderer, M., M. A. Skinner, S. G. Siddell. 1986. Coronavirus MHVJHM:nucleotide sequence of the mRNA that encodes the membrane protein. *Nucl. Acid. Res.* 14: 6338-
- Queen, C., K. J. Korn. 1984. A comprehensive sequence analysis program for the IBM personal computer. *Nucl. Acid. Res.* 12:581-599.
- Rasschaert, D., H. Laude. 1987. The predicted primary structure of the peplomer protein E2 of the Porcine transmissible enteric coronavirus. *J. Gen. Virol.* 68:1883-1890.
- Roseto, A., P. Boulesco, J. Laporte, J. Escarg, D. Gaches, J. Peries. 1982. Bovine enteric coronavirus structure as studied by a freeze drying technique. *J. Gen. Virol.* 63:241-245.
- Rottier, P. J. M., M. C. Horzinek, B. A. M. van der Zeijst. 1981a. Translation of the three Mouse Hepatitis virus (MHV A59) sub genomic RNAs in *Xenopus laevis* oocytes. *J. Virol.* 38:20-26.
- Rottier, P. J. M., M. C. Horzinek, B. A. M. van der Zeijst. 1981b. Viral protein synthesis in mouse hepatitis virus

strain A59-infected cells: effect of Tunicamycin. J. Virol. 40:350-357.

Rottier, P. J. M. D. Bradenburg, J. Armstrong, B. van der Zeijst, G. Warren. 1984. Assembly in vitro of a spanning membrane protein of the endoplasmic reticulum. The E1 glycoprotein of coronavirus mouse hepatitis virus A59. Proc. Natl. Acad. Sci. U.S.A. 81:1421-1425.

Rottier, P. J., J. Armstrong, D. I. Meyer. 1985. Signal recognition particle dependent insertion of coronavirus E1, an intracellular membrane glycoprotein. J. Biol. Chem. 260:4648-4652.

Roy Choudhary, R., R. Wu. 1980. Terminal transferase catalysed addition of nucleotides to the 3' termini of DNA. Meth. in Enzymol. 65:43-62.

Saif, L. J., D. R. Redman, K. V. Brock, E. M. Kohler, R. A. Heckert. 1988. Winter dysentery in adult dairy cattle: Detection of coronavirus in the faeces. Vet Res. 123:300-301.

Sanger, F., S. Nicklen, A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74:5463-5467.

Sawicki, S. G., D. L. Sawicki., 1986. Coronavirus minus strand RNA synthesis and effect of cycloheximide on coronavirus RNA synthesis. J. Virol. 57:328-334.

Sawicki, S. G., D. C. Sawicki, 1986b. The effect of overproduction of nonstructural proteins on alphavirus plus strand and minus strand RNA synthesis. Virology. 152:507-512.

Schlesinger, M. 1981. Proteolipids. Ann. Rev. Biochem. 50:193-203.

Schmidt, M. F. G. 1982. Acylation of viral spike glycoproteins: A feature of enveloped RNA viruses. Virology. 116:327-338.

Schmidt, I., M. Skinner, S. Siddell. 1987. Nucleotide sequence of the gene encoding the surface glycoprotein of coronavirus JHM. J. Gen. Virol. 68:47-56.

Schochetman, G., R. H. Stevens, R. W. Simpson. 1977. Presence of infectious polyadenylated RNA in the coronavirus Avian infectious bronchitis virus. Virology. 77:772-782.

Sethna, P. B., S. Hung, D. A. Brian. 1989. Coronavirus subgenomic replicons as a mechanism for mRNA amplification. *Proc. Natl. Acad. Sci. U.S.A.* 86:5626-5630.

Shieh, C. K., H. J. Lee, K. Yokomori, N. L. Monica, S. Makino, M. M. C. Lai. 1989. Identification of a new transcriptional initiation site and the corresponding functional gene 2b in the murine coronavirus RNA genome. *J. Virol.* 63:3729-3736.

Siddell, S. G., H. Wege, A. Barthel, V. ter Meulen. 1981. Intracellular protein synthesis and the in vitro translation of coronavirus JHM mRNA. In: ter Meulen, V., Siddell, S., Wege, H. (eds) *Biochemistry and Biology of coronaviruses*. Plenum Press. New York, p. 193-208.

Siddell, S. G., H. Wege, V. ter Meulen. 1982a. The structure and replication of coronaviruses. *Curr. Top. Microbiol. Immunol.* 99:131-163.

Siddell, S. G. 1982b. Coronavirus JHM. Tryptic peptide finger printing of virion proteins and intracellular polypeptides. *J. Gen. Virol.* 62:259-269.

Siddell, S., W. Wege, V. ter. Meulen. 1983. The biology of coronaviruses. *J. Gen. Virol.* 64:761-776.

Siddell, S. G. 1983a. Coronavirus JHM: coding assignments of subgenomic mRNAs. *J. Virol.* 64:113-125.

Skinner, M. A., S. G. Siddell. 1983. Coronavirus JHM: nucleotide sequence of the mRNA that encodes nucleocapsid protein. *Nucl. Acid. Res.* 11:5045-5054.

Skinner, M. A., D. Ebner, S. G. Siddell. 1985. Coronavirus MHV JHM mRNA5 has a sequence arrangement which potentially allows translation of a second downstream open reading frame. *J. Gen. Virol.* 66:581-582.

Smith, A. R., M. E. G. Boursnell, M. M. Binns, T. D. K. Brown, S. C. Inglis. 1987. *Adv. Exp. Med. and Biol.* 218:47-54.

Spaan, W., P. Rottier, M. Horzinek, B. A. M. van der Zeijst. 1981. Isolation and identification of virus specific mRNAs in cells infected with Mouse Hepatitis virus (MHV A59). *Virology.* 108:424-434.

Spaan, W., P. Rottier, M. Horzinek, B. A. M. van der Zeijst. 1982. Sequence relationships between the

genome and the intracellular RNA species 1,2,3,6,7 of Mouse Hepatitis strain A59. J. Virol. 42:432-439.

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

Spaan, W., D. Cavanagh, M. C. Horzinek. 1988. Coronaviruses: Structure and Genome expression. J. Gen. Virol. 69:2939-2952.

Stern, D., S. Kennedy. 1980a. Coronavirus multiplication strategy. I. Identification and characterization of virus specific RNA. J. Virol. 34:665-674.

Stern, D., S. Kennedy. 1980b. Coronavirus multiplication strategy. II. Mapping the Avian Infectious Bronchitis virus intracellular RNA species to the genome. J. Virol. 36:440-449.

Stern, D., S. Kennedy. 1982a. Coronavirus proteins: Biogenesis of Avian Infectious Bronchitis virus virion proteins. J. Virol. 44:794-803.

Stern, D. F., Sefton, B. M. 1982b. Coronavirus proteins: Structure and function of the oligosaccharides of the avian infectious bronchitis virus glycoproteins. J. Virol. 44:804-812.

Stern, D. F., B. M. Sefton. 1984. Coronavirus multiplication: The location of genes for the virion proteins on the avian infectious bronchitis virus genome. J. Virol. 50:22-29.

Stohlman, S. A., Lai, M. M. C. 1979. Phosphoproteins of murine hepatitis virus. J. Virol. 32:672-675.

Stohlman, L. S., J. O. Fleming, C. D. Patton, M. M. C. Lai. 1983. Synthesis and subcellular localization of the murine coronavirus nucleocapsid protein. Virology. 130:527-532.

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

n, L. S., K. V. Holmes, J. Behnke. 1980. Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. J. Virol. 33:449-462.

- Sturman, L. S. 1981. The structure and behaviour of coronavirus A59 glycoprotein: Biochemistry and Biology of Coronavirus. (V. ter meulen, S. Siddell, H. Wege. eds) Plenum Press. New York. 1-18.
- Sturman, L. S., K. V. Holmes. 1983. The molecular biology of coronaviruses. Adv. Virus. Res. 28:35-112.
- Sturman, L. S., C. S. Ricard, K. V. Holmes. 1985. Proteolytic cleavage of the E2 glycoprotein of murine coronavirus: activation of cell fusing activity of virions by trypsin and separation of two different 90K cleavage fragments. J. Virol. 56:904-911.
- Sugiyama, K., R. Ishikawa, N. Fukuhara. 1986. Structural polypeptides of the murine coronavirus DVIM. Archives of Virol. 89:245-254.
- Tomkins, W., A. Watrch, J. Schmate, R. Schulze and J. Harris. 1974. Cultural and antigenic properties of newly established cell strains derived from adenocarcinoma of human colon and rectum. J. Natl. Canc. Inst. 52:1101-1106.
- Talbot, P. J., G. Dionne, M. Lacroix. 1988. Vaccination against lethal coronavirus induced encephalitis with a synthetic decapeptide homologous to a domain in the predicted peplomer stalk. J. Virol. 62:3032-3036.
- Tooze, J., S. Tooze, G. Warren 1984. Replication of coronaviruses MHV A59 in Sac cells : determination of the first site of budding of progeny virus. Eur. J. Cell Biol. 33:281-293.
- Tung, F. 1987. PhD dissertation, University of Tennessee.
- Tyrrell, D., D. Alexander, J. Almeida, C. Cunningham, D. Hamre, M. Hofstad, L. Malluci, K. McIntosh. 1968. Coronaviruses. Nature. 220:650
- Tyrrell, D. A. J., J. D. Almeida, C. H. Cunningham, W. R. Dowde, M. W. Hofstad, K. McIntosh, M. Tajima, L. Y. Zastelskaya, B. C. Easterday, Kapikian, R. W. Bingham. 1975. Coronaviridae. Interviol. 5:76-82.
- Tyrrell, D. A. J., D. J. Alexander, J. D. Almeida, C. H. Cunningham, B. C. Easterday, D. J. Garwes, J. C. Hierholzer, A. Kapikian, M. R. Macnaughton, K. McIntosh. 1978. Coronaviridae. 2nd report. Interviol. 10:321-328.

- Veit, M., M. F. G. Schmidt, R. Rott. 1989. Different palmitoylation of paramyxovirus glycoproteins. *Virology*. 168:173-176.
- Vlasak, R., W. Luytjes, J. Leider, W. Spaan, P. Palese. 1988a. Human and bovine coronaviruses recognize sialic acid receptors similar to those of influenza C viruses. *Proc. Natl. Acad. Sci.* 85:4526-4529.
- Vlasak, R., Luytjes, W., J. Leider, W. Spaan, P. Palese. 1988b. The E3 protein of Bovine coronavirus is a receptor destroying enzyme with acetylcholinesterase activity. *J. Virol.* 62:4686-4690.
- Von Heijne, G. 1984. How signal sequences maintain cleavage specificity. *J. Mol. Biol.* 173:243-251.
- Wege, H., S. S. Siddell, V. ter Meulen. 1982. The Biology and pathogenesis of coronaviruses. *Curr. Top. microbiol. Immunol.* 99:165-200.
- Wesley, R. D., A. K. Cheung, D. D. Michael, R. D. Woods. 1989. Nucleotide sequence of coronavirus TGEV genomic RNA: evidence for 3 mRNAs species between the peplomer and matrix protein genes. *Virus Res.* 13:87-100.
- Wiener, J., J. A. Bartlett, W. K. Joklik. 1989. The sequences of Reovirus serotype 3. Genome segments M1 and M3 encoding the minor protein u2 and the Major nonstructural proteins uNS, respectively. *Virology* 169:293-304.
- Zebedee, S. L., C. D. Richardson. R. A. Lamb. 1985. Characterisation of the influenza virus M2 integral membrane protein and expression at the infected cell surface from cloned cDNA. *J. Virol.* 56:502-511.

APPENDIX

A.1 Codon usage for the BCV spike protein.

TTT Phe	46 (3.4)	TCT Ser	10 (0.7)	TAT Tyr	17 (1.2)	TGT Cys	21 (1.5)
TTC Phe	18 (1.3)	TCC Ser	5 (0.4)	TAC Tyr	21 (1.5)	TGC Cys	16 (1.2)
TTA Leu	79 (5.8)	TCA Ser	19 (1.4)	TAA End	32 (2.3)	TGA End	14 (1.0)
TTG Leu	49 (3.6)	TCG Ser	3 (0.2)	TAG End	37 (2.7)	TGG Trp	19 (1.4)
CTT Leu	42 (3.1)	CCT Pro	9 (0.7)	CAT His	22 (1.6)	CGT Arg	2 (0.1)
CTC Leu	19 (1.4)	CCC Pro	4 (0.3)	CAC His	20 (1.5)	CGC Arg	2 (0.1)
CTA Leu	65 (4.7)	CCA Pro	13 (0.9)	CAA Gln	12 (0.9)	CGA Arg	4 (0.3)
CTG Leu	44 (3.2)	CCG Pro	7 (0.5)	CAG Gln	25 (1.8)	CGG Arg	6 (0.4)
ATT Ile	61 (4.5)	ACT Thr	18 (1.3)	AAT Asn	19 (1.4)	AGT Ser	21 (1.5)
ATC Ile	23 (1.7)	ACC Thr	13 (0.9)	AAC Asn	18 (1.3)	AGC Ser	9 (0.7)
ATA Ile	60 (4.4)	ACA Thr	17 (1.2)	AAA Lys	23 (1.7)	AGA Arg	13 (0.9)
ATG Met	80 (5.8)	ACG Thr	4 (0.3)	AAG Lys	26 (1.9)	AGG Arg	19 (1.4)
GTT Val	39 (2.8)	GCT Ala	9 (0.7)	GAT Asp	5 (0.4)	GGT Gly	10 (0.7)
GTC Val	14 (1.0)	GCC Ala	7 (0.5)	GAC Asp	7 (0.5)	GGC Gly	7 (0.5)
GTA Val	49 (3.6)	GCA Ala	14 (1.0)	GAA Glu	6 (0.4)	GGA Gly	8 (0.6)
GTG Val	52 (3.8)	GCG Ala	5 (0.4)	GAG Glu	6 (0.4)	GGG Gly	5 (0.4)

A.2 Amino acid composition of the BCV spike protein.

The sequence contains 1363 amino acids:

Ala	78	(5.7)	Leu	127	(9.3)
Arg	35	(2.6)	Lys	56	(4.1)
Asn	108	(7.9)	Met	20	(1.5)
Asp	71	(5.2)	Phe	72	(5.3)
Cys	56	(4.1)	Pro	54	(4.0)
Gln	53	(3.9)	Ser	112	(8.2)
Glu	39	(2.9)	Thr	98	(7.2)
Gly	92	(6.7)	Trp	14	(1.0)
His	18	(1.3)	Tyr	75	(5.5)
Ile	79	(5.8)	Val	106	(7.8)
End	0	(0.0)			

Acidic	(Asp + Glu)	110	(8.1)
Basic	(Arg + Lys)	91	(6.7)
Aromatic	(Phe + Trp + Tyr)	161	(11.8)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	493	(36.2)

Molecular Weight = 150825.

Average pI = 5.95

A.3 Codon usage for the putative BCV 4.9K protein.

TTT Phe	1 (2.3)	TCT Ser	0 (0.0)	TAT Tyr	0 (0.0)	TGT Cys	0 (0.0)
TTC Phe	2 (4.5)	TCC Ser	0 (0.0)	TAC Tyr	0 (0.0)	TGC Cys	0 (0.0)
TTA Leu	2 (4.5)	TCA Ser	0 (0.0)	TAA End	0 (0.0)	TGA End	0 (0.0)
TTG Leu	1 (2.3)	TCG Ser	0 (0.0)	TAG End	1 (2.3)	TGG Trp	0 (0.0)
CTT Leu	0 (0.0)	CCT Pro	2 (4.5)	CAT His	3 (6.8)	CGT Arg	0 (0.0)
CTC Leu	0 (0.0)	CCC Pro	2 (4.5)	CAC His	0 (0.0)	CGC Arg	0 (0.0)
CTA Leu	1 (2.3)	CCA Pro	0 (0.0)	CAA Gln	0 (0.0)	CGA Arg	0 (0.0)
CTG Leu	0 (0.0)	CCG Pro	0 (0.0)	CAG Gln	0 (0.0)	CGG Arg	0 (0.0)
ATT Ile	3 (6.8)	ACT Thr	1 (2.3)	AAT Asn	1 (2.3)	AGT Ser	0 (0.0)
ATC Ile	0 (0.0)	ACC Thr	1 (2.3)	AAC Asn	0 (0.0)	AGC Ser	0 (0.0)
ATA Ile	4 (9.1)	ACA Thr	1 (2.3)	AAA Lys	0 (0.0)	AGA Arg	1 (2.3)
ATG Met	2 (4.5)	ACG Thr	1 (2.3)	AAG Lys	2 (4.5)	AGG Arg	0 (0.0)
GTT Val	1 (2.3)	GCT Ala	3 (6.8)	GAT Asp	2 (4.5)	GGT Gly	0 (0.0)
GTC Val	2 (4.5)	GCC Ala	0 (0.0)	GAC Asp	1 (2.3)	GGC Gly	0 (0.0)
GTA Val	0 (0.0)	GCA Ala	0 (0.0)	GAA Glu	0 (0.0)	GGA Gly	0 (0.0)
GTG Val	1 (2.3)	GCG Ala	0 (0.0)	GAG Glu	2 (4.5)	GGG Gly	0 (0.0)

A.4 Amino acid composition of the putative BCV 4.9K protein.

The sequence contains 43 amino acids:

Ala	3	(7.0)		Leu	4	(9.3)
Arg	1	(2.3)		Lys	2	(4.7)
Asn	1	(2.3)		Met	2	(4.7)
Asp	3	(7.0)		Phe	3	(7.0)
Cys	0	(0.0)		Pro	4	(9.3)
Gln	0	(0.0)		Ser	0	(0.0)
Glu	2	(4.7)		Thr	4	(9.3)
Gly	0	(0.0)		Trp	0	(0.0)
His	3	(7.0)		Tyr	0	(0.0)
Ile	7	(16.3)		Val	4	(9.3)
End	0	(0.0)				

Acidic	(Asp + Glu)	5	(11.6)
Basic	(Arg + Lys)	3	(7.0)
Aromatic	(Phe + Trp + Tyr)	3	(7.0)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	20	(46.5)

Molecular weight = 4911.

Average pI = 6.10

A.5 Codon usage for the putative BCV 4.8K protein.

TTT Phe	2 (4.3)	TCT Ser	2 (4.3)	TAT Tyr	2 (4.3)	TGT Cys	2 (4.3)
TTC Phe	0 (0.0)	TCC Ser	0 (0.0)	TAC Tyr	1 (2.2)	TGC Cys	0 (0.0)
TTA Leu	1 (2.2)	TCA Ser	1 (2.2)	TAA End	0 (0.0)	TGA End	0 (0.0)
TTG Leu	0 (0.0)	TCG Ser	0 (0.0)	TAG End	1 (2.2)	TGG Trp	0 (0.0)
CTT Leu	0 (0.0)	CCT Pro	1 (2.2)	CAT His	0 (0.0)	CGT Arg	0 (0.0)
CTC Leu	0 (0.0)	CCC Pro	0 (0.0)	CAC His	0 (0.0)	CGC Arg	0 (0.0)
CTA Leu	0 (0.0)	CCA Pro	1 (2.2)	CAA Gln	0 (0.0)	CGA Arg	0 (0.0)
CTG Leu	0 (0.0)	CCG Pro	0 (0.0)	CAG Gln	0 (0.0)	CGG Arg	1 (2.2)
ATT Ile	3 (6.5)	ACT Thr	4 (8.7)	AAT Asn	1 (2.2)	AGT Ser	1 (2.2)
ATC Ile	0 (0.0)	ACC Thr	3 (6.5)	AAC Asn	0 (0.0)	AGC Ser	1 (2.2)
ATA Ile	1 (2.2)	ACA Thr	4 (8.7)	AAA Lys	0 (0.0)	AGA Arg	0 (0.0)
ATG Met	3 (6.5)	ACG Thr	0 (0.0)	AAG Lys	0 (0.0)	AGG Arg	0 (0.0)
GTT Val	2 (4.3)	GCT Ala	1 (2.2)	GAT Asp	1 (2.2)	GGT Gly	3 (6.5)
GTC Val	0 (0.0)	GCC Ala	0 (0.0)	GAC Asp	2 (4.3)	GGC Gly	1 (2.2)
GTA Val	0 (0.0)	GCA Ala	0 (0.0)	GAA Glu	0 (0.0)	GGA Gly	0 (0.0)
GTG Val	0 (0.0)	GCG Ala	0 (0.0)	GAG Glu	0 (0.0)	GGG Gly	0 (0.0)

A.6 Amino acid composition of the putative BCV 4.8K protein.

The sequence contains 45 amino acids:

Ala	1	(2.2)	Leu	1	(2.2)
Arg	1	(2.2)	Lys	0	(0.0)
Asn	1	(2.2)	Met	3	(6.7)
Asp	3	(6.7)	Phe	2	(4.4)
Cys	2	(4.4)	Pro	2	(4.4)
Gln	0	(0.0)	Ser	5	(11.1)
Glu	0	(0.0)	Thr	11	(24.4)
Gly	4	(8.9)	Trp	0	(0.0)
His	0	(0.0)	Tyr	3	(6.7)
Ile	4	(8.9)	Val	2	(4.4)
End	0	(0.0)			

Acidic	(Asp + Glu)	3	(6.7)
Basic	(Arg + Lys)	1	(2.2)
Aromatic	(Phe + Trp + Tyr)	5	(11.1)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	15	(33.3)

Molecular Weight = 4823.

Average pI = 5.92

A.7 Codon usage for the putative BCV 12.7K protein.

TTT Phe	8 (7.3)	TCT Ser	0 (0.0)	TAT Tyr	8 (7.3)	TGT Cys	3 (2.7)
TTC Phe	2 (1.8)	TCC Ser	1 (0.9)	TAC Tyr	0 (0.0)	TGC Cys	1 (0.9)
TTA Leu	3 (2.7)	TCA Ser	1 (0.9)	TAA End	0 (0.0)	TGA End	1 (0.9)
TTG Leu	0 (0.0)	TCG Ser	0 (0.0)	TAG End	0 (0.0)	TGG Trp	3 (2.7)
CTT Leu	1 (0.9)	CCT Pro	3 (2.7)	CAT His	2 (1.8)	CGT Arg	2 (1.8)
CTC Leu	2 (1.8)	CCC Pro	0 (0.0)	CAC His	0 (0.0)	CGC Arg	0 (0.0)
CTA Leu	0 (0.0)	CCA Pro	0 (0.0)	CAA Gln	1 (0.9)	CGA Arg	0 (0.0)
CTG Leu	1 (0.9)	CCG Pro	0 (0.0)	CAG Gln	1 (0.9)	CGG Arg	1 (0.9)
ATT Ile	2 (1.8)	ACT Thr	5 (4.5)	AAT Asn	4 (3.6)	AGT Ser	1 (0.9)
ATC Ile	1 (0.9)	ACC Thr	0 (0.0)	AAC Asn	3 (2.7)	AGC Ser	1 (0.9)
ATA Ile	1 (0.9)	ACA Thr	2 (1.8)	AAA Lys	8 (7.3)	AGA Arg	3 (2.7)
ATG Met	1 (0.9)	ACG Thr	0 (0.0)	AAG Lys	0 (0.0)	AGG Arg	0 (0.0)
GTT Val	5 (4.5)	GCT Ala	6 (5.5)	GAT Asp	2 (1.8)	GGT Gly	3 (2.7)
GTC Val	1 (0.9)	GCC Ala	0 (0.0)	GAC Asp	2 (1.8)	GGC Gly	2 (1.8)
GTA Val	3 (2.7)	GCA Ala	1 (0.9)	GAA Glu	2 (1.8)	GGA Gly	2 (1.8)
GTG Val	0 (0.0)	GCG Ala	0 (0.0)	GAG Glu	4 (3.6)	GGG Gly	0 (0.0)

A.8 Amino acid composition of the putative 12.7k protein.

The sequence contains 109 amino acids:

Ala	8 (7.3)	Leu	7 (6.4)
Arg	6 (5.5)	Lys	8 (7.3)
Asn	7 (6.4)	Met	1 (0.9)
Asp	4 (3.7)	Phe	10 (9.2)
Cys	4 (3.7)	Pro	3 (2.8)
Gln	2 (1.8)	Ser	4 (3.7)
Glu	5 (4.6)	Thr	7 (6.4)
Gly	7 (6.4)	Trp	3 (2.8)
His	2 (1.8)	Tyr	8 (7.3)
Ile	4 (3.7)	Val	9 (8.3)
End	0 (0.0)		

Acidic	(Asp + Glu)	9 (8.3)
Basic	(Arg + Lys)	14 (12.8)
Aromatic	(Phe + Trp + Tyr)	21 (19.3)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	42 (38.5)

Molecular Weight = 12749.

Average pI = 6.23

A.9 Codon usage for the putative BCV 9.5K protein.

TTT Phe	8 (9.4)	TCT Ser	1 (1.2)	TAT Tyr	4 (4.7)	TGT Cys	3 (3.5)
TTC Phe	0 (0.0)	TCC Ser	1 (1.2)	TAC Tyr	1 (1.2)	TGC Cys	1 (1.2)
TTA Leu	2 (2.4)	TCA Ser	0 (0.0)	TAA End	0 (0.0)	TGA End	0 (0.0)
TTG Leu	3 (3.5)	TCG Ser	0 (0.0)	TAG End	1 (1.2)	TGG Trp	1 (1.2)
CTT Leu	2 (2.4)	CCT Pro	1 (1.2)	CAT His	0 (0.0)	CGT Arg	0 (0.0)
CTC Leu	0 (0.0)	CCC Pro	0 (0.0)	CAC His	0 (0.0)	CGC Arg	0 (0.0)
CTA Leu	0 (0.0)	CCA Pro	2 (2.4)	CAA Gln	2 (2.4)	CGA Arg	0 (0.0)
CTG Leu	1 (1.2)	CCG Pro	0 (0.0)	CAG Gln	1 (1.2)	CGG Arg	0 (0.0)
ATT Ile	4 (4.7)	ACT Thr	2 (2.4)	AAT Asn	2 (2.4)	AGT Ser	0 (0.0)
ATC Ile	0 (0.0)	ACC Thr	1 (1.2)	AAC Asn	1 (1.2)	AGC Ser	0 (0.0)
ATA Ile	4 (4.7)	ACA Thr	0 (0.0)	AAA Lys	2 (2.4)	AGA Arg	1 (1.2)
ATG Met	3 (3.5)	ACG Thr	0 (0.0)	AAG Lys	0 (0.0)	AGG Arg	1 (1.2)
GTT Val	5 (5.9)	GCT Ala	2 (2.4)	GAT Asp	4 (4.7)	GGT Gly	2 (2.4)
GTC Val	0 (0.0)	GCC Ala	1 (1.2)	GAC Asp	2 (2.4)	GGC Gly	0 (0.0)
GTA Val	2 (2.4)	GCA Ala	3 (3.5)	GAA Glu	0 (0.0)	GGA Gly	1 (1.2)
GTG Val	5 (5.9)	GCG Ala	0 (0.0)	GAG Glu	1 (1.2)	GGG Gly	1 (1.2)

A.10 Amino acid composition of the putative 9.5K protein.

The sequence contains 84 amino acids:

Ala	6	(7.1)	Leu	8	(9.5)
Arg	2	(2.4)	Lys	2	(2.4)
Asn	3	(3.6)	Met	3	(3.6)
Asp	6	(7.1)	Phe	8	(9.5)
Cys	4	(4.8)	Pro	3	(3.6)
Gln	3	(3.6)	Ser	2	(2.4)
Glu	1	(1.2)	Thr	3	(3.6)
Gly	4	(4.8)	Trp	1	(1.2)
His	0	(0.0)	Tyr	5	(6.0)
Ile	8	(9.5)	Val	12	(14.3)
End	0	(0.0)			

Acidic	(Asp + Glu)	7	(8.3)
Basic	(Arg + Lys)	4	(4.8)
Aromatic	(Phe + Trp + Tyr)	14	(16.7)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	45	(53.6)

Molecular Weight = 9543.

Average pI = 5.86

PREPARATION OF MINIPREP PLASMID DNA FOR SEQUENCING WITH SEQUENASE

This procedure is a modification of the protocol of Kraft et al., (1988).

1. Inoculate 3ml LB medium containing 100 micrograms/ml ampicillin with a single plasmid containing colony of E. coli strain JM83.
2. Shake overnight at 37⁰C
3. Centrifuge cells at 4⁰C for 5 minutes. Discard supernatant.
4. Resuspend cells in 100 microliters solution I (50mM glucose, 100mM EDTA, 25mM Tris-HCl (pH 8.0) by vortexing.
5. Incubate at room temperature 5 minutes.
6. Add 200 microliters freshly made 0.2N NaOH/1% SDS, mix, incubate at room temperature until solution becomes clear.
7. Incubate on ice 15 minutes.
8. Add 150 microliters 5M sodium acetate (pH 4.8), mix and incubate on ice 5 minutes.
9. Centrifuge at 4⁰C for 15 minutes.
10. Transfer supernatant to 1.5ml Eppendorf centrifuge tube.
11. Add 100 units RNase T1 and incubate at 37⁰ for 45 minutes.
12. Phenol/chloroform (1:1) extract.
13. Chloroform/isoamyl alcohol (24:1) extract.
14. Add 2.5 volumes ethanol, mix, and incubate on dry ice for 10 minutes.
15. Spin on microfuge at 4⁰C for 15 minutes to pellet DNA.
16. Discard supernatant, wash twice with 1ml 70% ethanol and dry under vacuum.
17. Dissolve DNA in 18.8 microliters water.

18. Remove 2 microliters for restriction enzyme digestion.
19. To the remaining 16.8 microliters add 3.2 microliters of 5M NaCl and mix.
20. Add 20 microliters 13% PEG (polyethylene glycol) and mix.
21. Incubate on ice 30 minutes.
22. Spin on microfuge at 4⁰C for 10 minutes, discard supernatant, wash pellet twice with 1ml 70% ethanol, and dry under vacuum.
23. Dissolve DNA in 18 microliters water.
24. Add 2 microliters 2N NaOH, mix and incubate at room temperature for 5 minutes.
25. Add 8 microliters 5M ammonium acetate (pH 7.5), mix, and add 100 microliters of ethanol.
26. Incubate on dry ice 5 minutes, spin on microfuge at 4⁰C for 15 minutes, and discard supernatant.
27. Wash DNA twice with 1ml 70% ethanol and dry under vacuum.

The DNA is now ready for sequencing using the protocol supplied with the Sequenase sequencing kit.

VITA

Sushma E. Abraham was born on September 5 1960 in Madras, India. She received her Bachelor of Science degree majoring in Chemistry, Botany and Zoology in 1981 from Bangalore University, India. She joined M.S. University of Baroda, India in 1981 and graduated with a Master of Science degree in Microbiology in 1983. From 1983 to 1985 she worked as a scientific assistant in the Department of Biochemistry at the Indian Institute of Science, Bangalore. In the Fall of 1985 she entered the PhD program in the Department of Microbiology at the University of Tennessee, Knoxville and received her PhD degree in December 1989.