

To the Graduate Council:

I am submitting herewith a dissertation written by William Lapps entitled "Molecular Cloning and Sequence Analysis of the 3' End of the Bovine Enteric Coronavirus Genome." 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:

W. S. Riggs

Karl Smith

Wen-Kuei Yang

Accepted for the Council:

Cowminkel

Vice Provost
and Dean of the Graduate School

MOLECULAR CLONING AND SEQUENCE ANALYSIS OF
THE 3' END OF THE BOVINE ENTERIC
CORONAVIRUS GENOME

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

William Lapps

March 1986

DEDICATED
TO
Michele and Christopher
for their love,
support, and
perseverance

ABSTRACT

The structure of the bovine enteric coronavirus (BCV) genome was examined using two experimental approaches. The first approach compared the degree of nucleotide sequences homology between the antigenically related BCV and human respiratory coronavirus OC43 (HCV OC43) by two-dimensional T_1 oligonucleotide fingerprinting. From an analysis of comigrating oligonucleotides in a mixing experiment, it was determined that BCV and HCV OC43 share a sequence homology of approximately 96%. This degree of homology supports the notion that the two viruses are, in fact, not identical, even though they share a very similar protein profile, and are closely related antigenically. The second approach examined the primary structure of the 3' end of the BCV genome. The 3' end of the 20 Kb BCV genome was cDNA cloned and DNA sequenced, and the properties of two genes and a potential third gene were deduced and analyzed from an examination of open reading frames. One clone of 2.16 Kb was sequenced in its entirety and other clones of 4.1, 2.8, and 1.08 were sequenced in part, using the chemical method of Maxam and Gilbert. The largest gene (N) begins at base 1634 from the 3' end of the genome and encodes a 448 amino acid protein of 49,379 molecular weight. It is the nucleocapsid protein gene on the basis of size, chemical properties, and extensive amino acid sequence homology with the murine coronavirus nucleocapsid protein. A hypothetical gene (IORF) identified from a second open reading frame internal to the N gene, encodes a 207 amino acid protein of 23,057 molecular weight beginning at base 1573. The third gene (M) begins at

base 2336 from the 3' end of the genome and encodes a 230 amino acid protein of 26,376 molecular weight. It is the matrix protein gene on the basis of its size, hydrophobicity pattern, and extensive amino acid sequence homology with the matrix protein of the murine coronavirus. Both nucleotide and amino acid sequence data show BCV to be closely related to the murine hepatitis coronavirus and very distantly related to the porcine transmissible gastroenteritis coronavirus and the avian infectious bronchitis coronavirus.

TABLE OF CONTENTS

CHAPTER	PAGE
I. REVIEW OF CORONAVIRUS STRUCTURE AND REPLICATION.	1
Introduction	2
Structural Properties of the Bovine Enteric	
Coronavirus.	6
Genome Structure	6
Protein Structure	6
Summary of Coronavirus Replicational Strategy	10
Sequence Analyses of Coronavirus Nucleic Acids.	17
Sequence Analysis of Mouse Hepatitis Virus Nucleic	
Acid Sequences	17
Sequence Analysis of Avian Infectious Bronchitis	
Coronavirus Nucleic Acid Sequences	20
Rationale for Molecular Biological Studies on the Bovine	
Enteric Coronavirus	24
II. OLIGONUCLEOTIDE FINGERPRINTS OF ANTIGENICALLY RELATED	
BOVINE CORONAVIRUS AND HUMAN CORONAVIRUS OC43.	27
Introduction	28
Materials and Methods	29
Cells and Viruses.	29
Radiolabeling and Purification of Virus	29
Analysis of RNA	30
Results.	31
Virion RNA	31

Genome Sequence Homology as Determined by

Oligonucleotide Fingerprinting. 31

Properties of the Virion 0.68 Md Species 33

Discussion. 35

III. CLONING AND SEQUENCE ANALYSIS OF THE BOVINE CORONAVIRUS

NUCLEOCAPSID PROTEIN GENE 37

Introduction 38

Materials and Methods 40

Virus and Cells 40

Purification of Genomic RNA 40

cDNA Cloning of the 3' End of the BCV Genome. 41

Identification of Large Inserts Containing 3'-Specific

BCV Sequences 43

Restriction Endonuclease Mapping. 44

DNA Sequencing and Sequence Analysis 44

Results. 44

cDNA Cloning and Sequencing of Three Clones from the

3' End of the Genome 44

The Largest Open Reading Frame Predicts a Protein

Having Properties of the Nucleocapsid Protein. 50

A Large Open Reading Frame Found Internal to the N

Gene, but in a Different Reading Frame, Predicts a

207 Amino Acid Protein Having a Molecular Weight of

23,057. 51

Discussion. 51

IV. CLONING AND SEQUENCE ANALYSIS OF THE BOVINE CORONAVIRUS

MATRIX PROTEIN GENE	57
Introduction	58
Materials and Methods	59
cDNA Cloning and Identification of BCV 3'-Specific	
Clones.	59
DNA Sequencing and Sequence Analysis	59
Results.	60
Discussion.	64
REFERENCES	66
APPENDICES	75

APPENDIX A. DETAILED METHODS USED IN THE MOLECULAR CLONING

OF THE BOVINE CORONAVIRUS GENOMIC RNA	75
Virus Production in Tissue Culture Cells.	76
Purification of the Bovine Coronavirus by Isopycnic	
Sedimentation	77
Purification of Genomic RNA by Rate-Zonal Sedimentation.	79
First Strand DNA Synthesis	80
Second Strand DNA Synthesis	84
Homopolymer Tailing of Double-Stranded cDNA.	85
Preparation of Vector DNA.	88
Annealing and Transformation.	89

APPENDIX B. SIZE AND SEQUENCE CHARACTERIZATION OF BOVINE

CORONAVIRUS SPECIFIC CLONES.	94
Initial Identification of Clones	95
Identification of Bovine Coronavirus-Specific Clones. .	98
Use of Radioactive Probes.	100
Isolation and Size Screening of Plasmids with BCV-	
Specific Inserts	100
Restriction Enzyme Site Mapping.	102
Sequence Analysis of 3' End Specific Clones: DNA End	
Labeling and Sequencing.	108
Isolation of DNA Fragments from Gels	109
DNA Strand Separation	112
Sequence Analyses of Clones	114

APPENDIX C. HANAHAN'S METHOD FOR TRANSFORMING E. COLI WITH

PLASMIDS	133
VITA	138

LIST OF TABLES

TABLE		PAGE
I.1.	Listing of the most characterized coronaviruses and their associated diseases	3
I.2.	Relative abundance of virus-specific RNAs synthesized between 6.1 and 6.6 h postinfection in MHV A59-infected Sac(-) cells	12
A.1.	Optimal annealing conditions for transformation by homopolymer-tailed ds cDNA and vector	90
B.1.	Summary of clones analyzed	107
B.2.	Flow diagram of the Maxam and Gilbert technique	110
B.3.	Urea-polyacrylamide sequencing gel formulas.	111
B.4.	Amino acid composition of the BCV nucleocapsid protein	119
B.5.	Amino acid composition of the BCV matrix protein protein	120
B.6.	Predicted amino acid composition from the internal open reading frame (IORF) found in BCV	121

LIST OF FIGURES

FIGURE	PAGE
I.1. Tentative coronavirus family tree based on antigenic interrelatedness.	5
I.2. Model of the bovine coronavirion	7
I.3. Virion proteins of the human coronavirus OC43 bovine coronavirus, and murine hepatitis virus	8
I.4. Scheme showing interrelatedness between the genome and mRNA's of the mouse hepatitis coronavirus	13
I.5. Proposed binding of leader RNA to the negative-strand template	15
I.6. Genome organization of the mouse hepatitis virus.	18
I.7. Genome organization of the avian infectious bronchitis virus	21
II.1. Agarose gel electrophoresis of virion RNA after glyoxal-dimethylsulfoxide denaturation	32
II.2. Oligonucleotide fingerprints of BCV and HCV OC43 genomic RNAs.	34
III.1. Restriction endonuclease map and sequencing strategy for BCV cDNA clones CB9, MN3, and MA5	46
III.2. The primary nucleotide sequence of clone MN3 and the deduced amino acid sequences for the nucleocapsid phosphoprotein and a portion of the matrix glycoprotein.	48
III.3. Schematic diagram of open reading frames from clone MN3	49

FIGURE	PAGE
III.4. Nucleotide and amino acid sequence of the internal open reading frame (IORF)	52
III.5. Positions of serines on the N protein of MHV, BCV, IBV and TGEV.	55
IV.1. Restriction endonuclease map and sequencing strategy of BCV cDNA clones MN3, MA5, and MA7.	61
IV.2. Nucleotide and amino acid sequence of the bovine coronavirus matrix protein	62
IV.3. Schematic diagram of open reading frames that locate the bovine coronavirus matrix protein.	63
A.1. Purification of bovine coronavirus by isopycnic sedimentation	78
A.2. Rate-zonal sedimentation of bovine coronaviral RNA for purification of full-length genomic RNA.	81
A.3. Agarose-urea gel of purified genomic RNA	82
A.4. Time course for adding homopolymer tails of dCMP to the 3' ends of Pst I cut pBR322.	87
A.5. Plasmid cloning vectors pUC8 and pUC9 showing opposite orientations of the multiple cloning regions	93
B.1. Colony hybridization using random-primed bovine coronavirus-specific probe	101
B.2. Size analysis of plasmids from potential clones	103
B.3. Sau3AI fragment sizes from clones MA7 and CB11	105
B.4. Restriction endonuclease map of MN3 obtained by the Smith and Birnstiel technique	106

B.5.	Double stranded DNA fragment separation on an acrylamide gel	.116
B.6.	Strand separation of isolated DNA fragments	.117
B.7.	Maxam and Gilbert sequencing of strand separated molecules.	.118
B.8.	Predicted alignment of homologous amino acids for the matrix glycoproteins of BCV and MHV-A59.	.122
B.9.	Predicted alignment of homologous nucleotides for the matrix glycoprotein genes of BCV and MHV-A59	.123
B.10.	Amino acid homologies of the BCV and MHV-A59 matrix glycoproteins compared by diagon plotting	.124
B.11.	Nucleotide homologies of the BCV and MHV-A59 matrix glycoprotein genes compared by diagon plotting	.125
B.12.	Amino acid homologies of the BCV and MHV-A59 nucleocapsid protein compared by diagon plotting	.126
B.13.	Amino acid homologies of the BCV and MHV-JHM nucleocapsid proteins compared by diagon plotting.	.127
B.14.	Amino acid homologies of the BCV and TGEV nucleocapsid proteins compared by diagon plotting.	.128
B.15.	Amino acid homologies of the BCV and IBV nucleocapsid proteins compared by diagon plotting.	.129
B.16.	Nucleotide homologies of the BCV and MHV-A59 nucleocapsid protein genes compared by diagon plotting	.130
B.17.	Nucleotide homologies of the BCV and MHV-JHM nucleocapsid protein genes compared by diagon plotting	.131

B.18. Alignment of homologous nucleotides in the internal open reading frames of BCV and MHV-A59.	.132
--	------

CHAPTER I

REVIEW OF CORONAVIRUS

STRUCTURE AND

REPLICATION

INTRODUCTION

Coronaviruses are one of eleven animal RNA virus families. They were originally grouped together on the basis of electron microscopic appearance, sensitivity to lipid solvents and the presence of an RNA genome (Tyrell et al., 1975). Coronaviruses have an average diameter of 100 nm including the characteristic glycoprotein projection of 12-24 nm. The virus appears pleomorphic by negative staining and spherical when freeze dried (Roseto et al., 1982). The spherical form of the virus may better approximate the natural form (Mebus et al., 1972). Coronavirus particles have a buoyant density of 1.16-1.23 g/cm³ in sucrose and 1.23-1.24 g/cm³ in CsCl (Tyrell et al., 1978).

Recent studies on the replicational strategy of coronaviruses have established that they are a unique family from a molecular biological point of view as well as from a structural point of view. The genomes of all well studied coronaviruses are single-stranded, non-segmented, polyadenylated infectious RNA molecule of approximately 15 to 20 Kb in length. Coronavirus are therefore the third family of positive-stranded viruses following the picornaviruses and togaviruses. The coronavirus genome is the largest single-stranded RNA genome known. The protein structure and the replicational strategy of coronaviruses are further discussed below.

A listing of the most characterized coronaviruses, their natural hosts and associated diseases, are in Table I.1. When the antigenic relationships among coronaviruses are studied, four antigenic subgroups can be found (Figure I.1). It can be noted that the turkey coronavirus

Table I.1. Listing of the most characterized coronaviruses and their associated diseases.^a

Virus	Abbreviation	Natural		Predominant Clinical Disease
		Host		
Avian infectious bronchitis virus	IBV (AIBV)	Chicken		Respiratory disease
Bovine enteric coronavirus	BCV ^b	Calf		Diarrhea
Canine coronavirus	CCV	Dog		Diarrhea
Feline infectious peritonitis virus	FIPV	Cat		Peritonitis, multiple organ inflammation
Human respiratory coronavirus OC43	HCV-OC43	Human		Common cold
Human respiratory coronavirus 229E	HCV-229E	Human		Common cold
Murine (mouse) hepatitis virus	MHV	Mouse		Encephalomyelitis, hepatitis, diarrhea
Porcine transmissible gastroenteritis virus	TGEV	Pig		Diarrhea
Porcine hemagglutinating encephalomyelitis virus	HEV	Pig		Vomiting and wasting, encephalomyelitis
Parker's rat coronavirus	RCV	Rat		Pneumonitis, rhinitis
Rat sialodacryoadenitis virus	SDAV	Rat		Adenitis of salivary gland
Turkey coronavirus	TCV	Turkey		Enteritis

^aFrom the review by Wege et al., 1982.

^bThe star of this dissertation and the unofficial coronavirus of the 1982 World's Fair.

comprises one subgroup, the chicken respiratory coronavirus comprises a second subgroup, and the remaining two subgroups are comprised entirely of mammalian coronaviruses.

The subject of this dissertation is a detailed analysis of the genome structure of the bovine enteric coronavirus. Several fundamental questions prompted this series of studies and briefly stated they are the following: (1) What is the evolutionary relationship among the coronaviruses as determined from an analysis of genetic sequences? Is there evidence that they arose from one ancestor? (2) What is the total number of genes encoded by the coronavirus genome? What are the products of these genes and what are the functions of these gene products in coronavirus replication? What is the arrangement of the genes in the genome? Are the number, identification and arrangement of genes the same among all coronaviruses? (3) What can the primary sequence of the genome tell us about the replication strategy of coronaviruses? What can it tell us about genetic regulatory mechanisms? (4) For the economically important pathogenic bovine enteric coronavirus, can the genes of the proteins responsible for protective immunity be cloned and prepared for use in genetically engineered vaccines?

Many of these questions were not answered at all. Some were answered only in part. For certain, the experimental approaches employed in this dissertation begin to answer some of the questions and lay the groundwork for further answers to come.

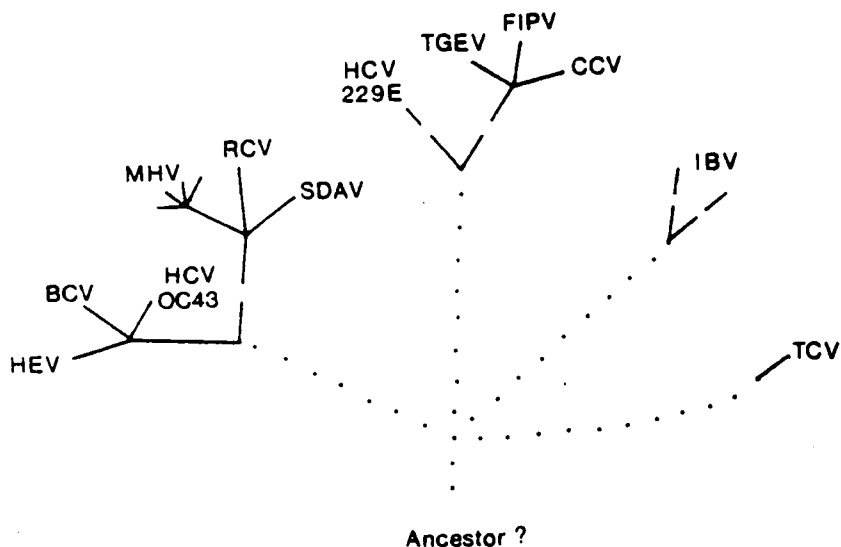


Figure I.1. Tentative coronavirus family tree based on antigenic interrelatedness. Data for this figure comes primarily from the review by Wege et al. (Wege et al., 1982). Solid lines indicate evidence for cross neutralizing activity; dashed lines indicate immunological evidence (other than neutralization) for cross reactivity; dotted lines indicate possible relatedness with little or no supporting data. Viruses include: the porcine hemagglutinating encephalomyelitis virus (HEV), bovine coronavirus (BCV), human coronavirus OC43 (HCV OC43), murine hepatitis virus (MHV), rat coronavirus (RCV), rat sialodacryoadenitis virus (SDAV), human coronavirus 229E (HCV 229E), porcine transmissible gastroenteritis virus (TGEV), feline infectious peritonitis virus (FIPV), canine coronavirus (CCV), avian infectious bronchitis virus (IBV), and turkey bluecomb disease coronavirus (TCV).

What follows in the remainder of the literature review is specific information about coronavirus structure and replication that lays the groundwork for my experiments with the bovine enteric coronavirus.

STRUCTURAL PROPERTIES OF THE BOVINE ENTERIC CORONAVIRUS

Genome Structure

At the time my studies were initiated, it was known that the bovine coronavirus contained a single-stranded, nonsegmented, polyadenylated RNA molecule as its genome (Guy and Brian, 1979). Although infectiousness of the molecule has not yet been demonstrated, its structure conforms with that of the prototypic avian infectious bronchitis virus and the extensively studied mouse hepatitis virus.

Protein Structure

Coronaviruses are described in recent reviews as having the general property of possessing three major structural proteins: a peplomeric glycoprotein of approximately 200 Kd, a nucleocapsid phosphoprotein of approximately 50 Kd, and a matrix glycoprotein of approximately 26 Kd (Siddell et al., 1982).

Detailed studies from our laboratory over the last five years have shown that the bovine enteric coronavirus contains still a fourth major structural protein, an external glycoprotein of approximately 140 Kd that apparently is the virion hemagglutinin (Figure I.2; King and Brian, 1982; King et al., 1985). Immunological studies demonstrate that the hemagglutinin protein is missing from the mouse hepatitis

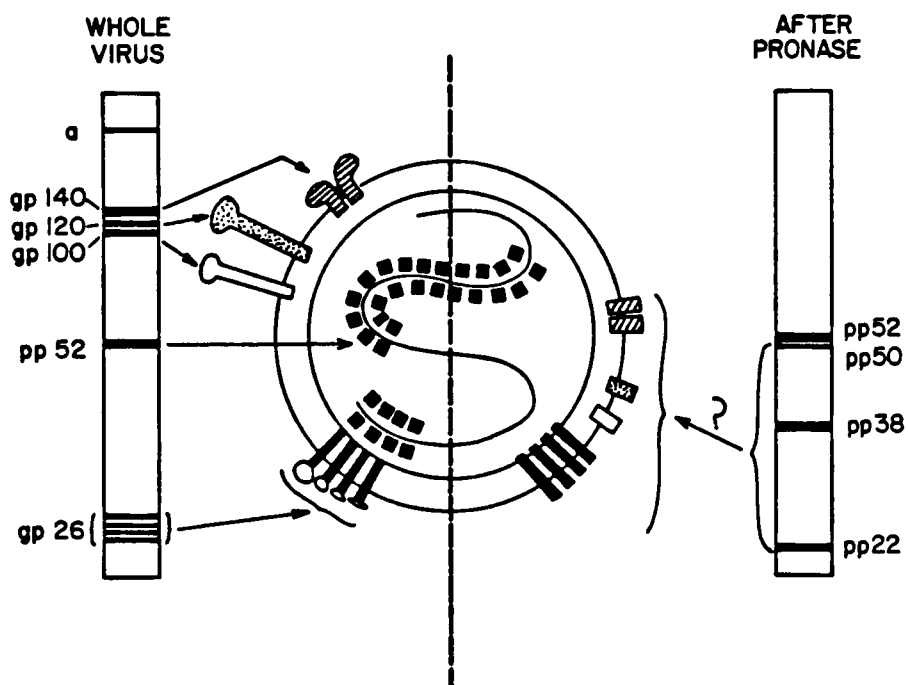


Figure I.2. Model of the bovine coronavirion. This model was proposed by King and Brian, 1982. The schematized electrophoretic profile on the left shows the protein profile of BCV electrophoresed in the absence of 2-mercaptoethanol. These protein species are described in the text. The schematized electrophoretic profile on the right is from particles that were treated with protease and then gradient purified.

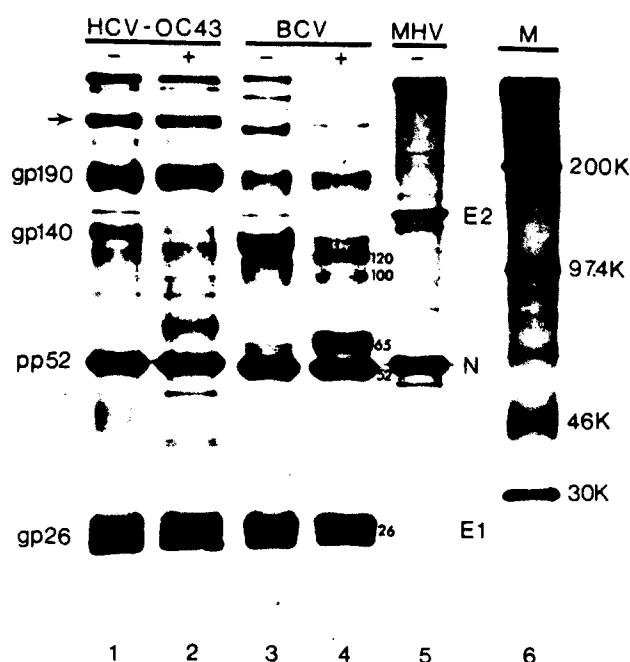


Figure I.3. Virion proteins of the human coronavirus OC43 bovine coronavirus, and murine hepatitis virus. These were described and compared previously (Hogue et al., 1984). Virus was labeled in vivo with [^3H] aa, 20 uCi/ml, purified, and analyzed by SDS-polyacrylamide gel electrophoresis. Sample treatment buffer was 4% SDS, 125 mM Tris Cl (pH 6.8), 5 M urea. Samples in lanes 2 and 4 also contained 2% 2-mercaptoethanol. The major protein species for HCV OC43 are indicated as gp190, gp130, pp55 and gp26. The equivalent species for BCV are gp120-100, gp140, pp52 and gp26. E1, N, and E2 refer the major envelope and nuclear proteins of MHV as described by Sturman and Holmes (1982).

coronavirus (Figure I.3; Hogue et al., 1984). These studies further demonstrate which proteins between BCV and MHV are antigenically homologous and demonstrate that the hemagglutinating human respiratory coronavirus OC43 is antigenically and structurally very similar to BCV (Figure I.3; Hogue et al., 1984).

The bovine coronavirus structural proteins can therefore be described in the following way: (1) The large peplomeric glycoprotein of 190 Kd exists on the virion primarily as its proteolytically cleaved subunits of 120 Kd and 100 Kd. The sugar is linked by N linkage. The 190 Kd peplomeric protein is therefore homologous to the 180 Kd E2 protein of MHV (Sturman and Holmes, 1977; B. Hogue, personal communication). (2) The hemagglutinin protein is a disulfide-linked dimer of 65 Kd subunits. The subunits are synthesized as 40 Kd protein subunits and the glycosylation is apparently totally of the N-linked type (Hogue, personal communication). (3) The phosphorylated nucleocapsid protein is 52 Kd and is contained totally within the virion envelope. It is undoubtedly closely associated with the virion RNA. (4) The glycosylated 26 Kd matrix protein contains only O-linked sugars (Nieman et al., 1984; Holmes et al., 1981; B. Hogue, personal communication) and is intimately associated with the virion envelope (Rottier et al., 1984).

The 190 Kd, 140 Kd, 52 Kd and 26 Kd species are present in molar ratios of 1, 4, 25 and 25 respectively (King and Brian, 1982) and the 1 to 1 ratio of 52 Kd and 26 Kd suggests that there may be a physical interaction between these molecules.

No nonstructural proteins or minor structural proteins of BCV have yet been characterized.

SUMMARY OF CORONAVIRUS REPLICATIONAL STRATEGY

The replication of coronaviruses has been studied in detail using the murine hepatitis viruses. Since the mouse hepatitis virus and bovine coronavirus are closely related, the description of the replication strategy for MHV will likely apply to the bovine coronavirus as well.

The replication of coronaviruses start with the entrance of the virus into the cell by either viropexis or envelope-membrane fusion (Siddell et al., 1982). It is not known at this time how nucleocapsids are uncoated in the cytoplasm. The nucleus of the cell is not needed for replication, since murine coronaviruses have been shown to replicate in enucleated cells (Wilhelmsen, 1981).

The first step in the replication of coronaviruses is probably the translation of the genomic RNA as a message to produce virus specific RNA-dependent RNA polymerase. This polymerase is found in infected cells and is not associated with purified virus. It was first described for the transmissible gastroenteritis virus (Dennis and Brian, 1981). The polymerase activity was later described for the murine hepatitis virus system (Mahy et al., 1983). Murine hepatitis viruses may also have two polymerase activities in infected cells (Brayton et al., 1982). The first (early) RNA polymerase is hypothesized to make full-length virus strand RNA (Lai et al., 1982b). It is unknown at this time whether or not a primer is needed for the

synthesis of the full-length minus strand. Full-length minus strand RNA is then used as a template for synthesis of full-length genomic RNA and messenger RNA molecules by use of the second (later) polymerase. The two separate polymerase activities may be function the same molecule whose specificity has been altered by an added subunit. The mechanisms involved in transcription have been the focus of a large number of studies. One prominent question is why coronavirus messages should be present in a constant, unequimolar distribution throughout the replication cycle. Table I.2 illustrates the relative abundance of virus-specific RNA found in MHV-A59-infected cells. There is a great deal of evidence suggesting that synthesis of coronavirus mRNA molecules requires a primer. UV transcriptional mapping of mRNA made in MHV A59-infected cells indicated target sizes equal to the size of mRNA molecules (Jacobs et al., 1981). These experiments also indicated that mRNA molecules are not formed from splicing or processing of a large common precursor. They do not rule out the existence of small negative-stranded templates that would serve to make the mRNA molecules, however. Small negative-strand templates are ruled out because the only negative-strand sense viral RNA found in infected cells is genome length (Lai et al., 1982b). Figure I.4 illustrates a scheme showing the interrelatedness between the genome and mRNAs of the mouse hepatitis coronavirus (Siddell et al., 1982). The messenger RNAs of coronaviruses form a nested set (Stern and Kennedy, 1980a; Stern and Kennedy, 1980b; Siddell et al., 1982). In other words, the nucleotide sequence of any given message is contained within the 3' end of next larger message. The 3' end of all the mRNAs is the same. The 5' end

Table I.2. Relative abundance of virus-specific RNAs synthesized between 6.1 and 6.6 h postinfection in MHV A59-infected Sac(-) cells.^a

RNA	Mol. wt. (x 10 ⁶)	cpm (%)	Relative molarity (%)
1	5.6	37.4	10.4
2	4.0	12.0	4.7
3	3.0	8.2	4.2
4	1.4	5.9	6.5
5	1.2	10.1	13.1
6	0.9	9.1	15.7
7	0.6	17.5	45.5

^a[³H] uridine-labeled, virus-specific RNA species were separated by agarose gel electrophoresis (Jacobs et al., 1981). The individual bands were excised from the gel and counted to determine the fraction of the label present in each RNA. The relative molarity of each RNA species was calculated.

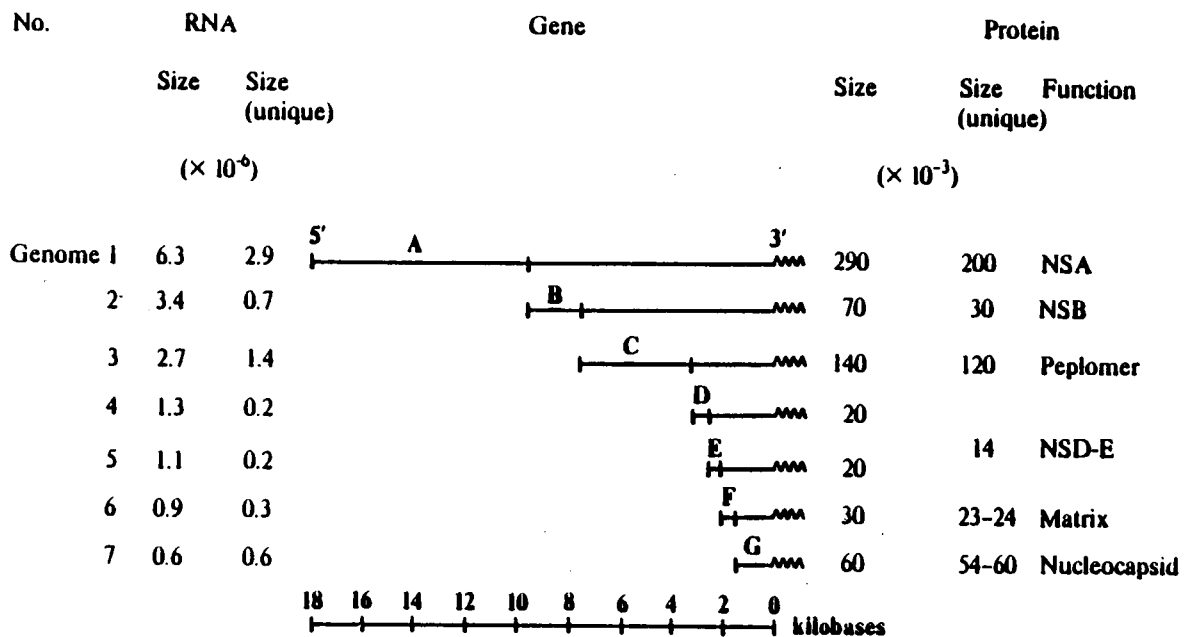


Figure I.4. Scheme showing interrelatedness between the genome and mRNA's of the mouse hepatitis coronavirus. This was adapted from Siddell et al. (1982). This scheme describes the sizes of the mRNA found in MHV-infected cells. It also illustrates that the messages are a nested set, having the same 3' end. The unique 5'-most coding regions for each mRNA (which is the only part that is translated) are designated by letters A through G. The proteins encoded by these mRNAs are: nonstructural protein A (NSA), nonstructural protein B (NSB), peplomer (or spike protein), nonstructural proteins D and E (NSD and NSE), matrix (or membrane protein), nucleocapsid protein.

of each mRNA appears to be the only region that is translated (Siddell, 1983). The translated portion of the messages is represented by the letters A through G on Figure I.4. The size of the "unique" region of each mRNA is also given in the third column of Figure I.4.

From oligonucleotide maps of the mRNAs, a common oligonucleotide was found at the 5' end of each mRNA molecule (Lai et al., 1983). This oligonucleotide is present at the 5' end of genomic RNA as well, but not internally within the genome. This is the first direct evidence that a leader sequence exists in the mRNAs of coronaviruses.

Presumably the 3' end of the negative-stranded genome size template is transcribed to produce a 72 base pair leader (Baric et al., 1985; Lai et al., 1984). This positive-stranded leader is probably released from the template along with the RNA polymerase and helps to prime the synthesis of the mRNAs. The intercistronic regions (regions between open reading frames or regions before the start methionine codon of a coding sequence) contain a common sequence that allows base pairing at the 3' end of the leader sequence (Budzilowicz et al., 1985). This hypothetical binding is shown in Figure I.5. Presumably the six to eight base pairs which are shown in the figure, with the help of proposed protein interaction, serve to align the primer for mRNA transcription. The base paired regions shown in Figure I.5 are comparisons between the sequence of the negative strand template (derived from clones of the positive-stranded genomic RNA) and the known sequences of the 5' noncoding regions for previously sequenced mRNAs (Armstrong et al., 1983; Armstrong et al., 1984; Niemann et al., 1984; Spaan et al., 1984). This research has shown that the leader is

Gene 5	3' UAAUACAAUGAUAGAUUUUGGAGUAG 5' negative strand template	
	5' aguuuaaaucuaaUCUAAACCUCAUC 3' mRNA	
Gene 6	CCUACUAAUAGAUUAGGUUUUGUAAUAC negative strand template	
	aguuuaaaucuaaUCCAAACAUUAUG mRNA	
Gene 7	CAACUCUUGGAUUAGAUUUUGAAAUCCUAC negative strand template	
	aguuuaaaucuaaUCUAAACUUUAAGGAUG mRNA	

Figure I.5. Proposed binding of leader RNA to the negative-strand template. This model was proposed by Budzilowicz et al. (1985). The sequences at the 5' noncoding region of the mRNAs (on the negative strand template) were derived from cDNA clones of genomic RNA. The lower case letters represent the 3' end of the leader RNA. Base pairing between the leader and the genome is indicated by vertical lines. The underlined nucleotides are conserved at intergenic regions but are probably not used by the leader.

fused to the coding regions of the mRNAs. Heteroduplex mapping of full length cDNA clones of some of the messages (6 and 7) to genomic RNA further corroborates the location of the leader sequence at the five prime end of the genomic RNA.

Coronavirus mRNAs are not produced in equimolar amounts (Siddell et al., 1982; Budzillowicz et al., 1985). Mouse hepatitis virus mRNAs have been detected in infected cells in widely different concentrations which appear constant throughout infection. For example the molar ratios of mRNAs 7, 6, and 5 (which are depicted in Figure I.4) are 100:32:11 respectively (Budzillowicz et al., 1985). It is proposed that the amount of homology between the leader sequences and the intergenic sequences (where they base pair) is one of the factors that controls the relative frequency of transcription initiation (Budzillowicz et al., 1985). Figure I.5 shows the base pairing interactions that are possible for three of the intergenic regions. The intergenic regions show relative sequence homology in the following order: 7>6>5. This corresponds to the molar ratios of the mRNAs produced. This evidence suggests that the control of transcription is at the level of initiation.

In infected cells, 10% of progeny virion RNA and mRNA are associated with polysomes and are released from polysomes by treatment with EDTA (Siddell et al., 1982). The rest of the virion RNA is associated with an EDTA-resistant structure which is probably the nucleocapsid protein. It is presumed that translation of the nucleocapsid and nonstructural proteins occur on free polyribosome

(Siddell et al., 1982). The envelope glycoproteins have been shown to be translated on membrane bound polysomes (Siddell et al., 1982).

Coronavirus virions appear to be formed by budding of the RNA and nucleocapsid protein complexes through membranes into the lumen of the rough endoplasmic reticulum (Siddell et al., 1982). Coronaviruses do not bud through cytoplasmic membranes (Holmes et al., 1984). This is probably due to the restricted intracellular transport of the matrix glycoprotein (E1 or gp26 in the BCoV system). The budding process is proposed to be mediated by interactions between the nucleocapsid and the matrix glycoproteins (Sturman et al., 1980). After budding, coronaviruses are transported to the Golgi and smooth walled cytoplasmic vesicles (Siddell et al., 1982). Virus particles are then released by cell lysis or fusion of the cytoplasmic vesicle with the plasma membrane.

SEQUENCE ANALYSES OF CORONAVIRUS NUCLEIC ACIDS

Sequence Analysis of Mouse Hepatitis Virus Nucleic Acid Sequences

The coding regions for MHV mRNAs 4, 5, 6 and 7 have been cloned and sequenced (Skinner and Siddell, 1985; Skinner et al., 1985; Skinner and Siddell, 1983; Armstrong et al., 1984; Armstrong et al., 1983). Figure I.6 shows the arrangements of the open reading frames relative to the messages and to genomic RNA for MHV-JHM coronavirus (Skinner et al., 1985).

The open reading frame (ORF) marked N designates the sequence that encodes the nucleocapsid protein gene. This sequence has been determined for both MHV-A59 and JHM strains (Armstrong et al., 1983;

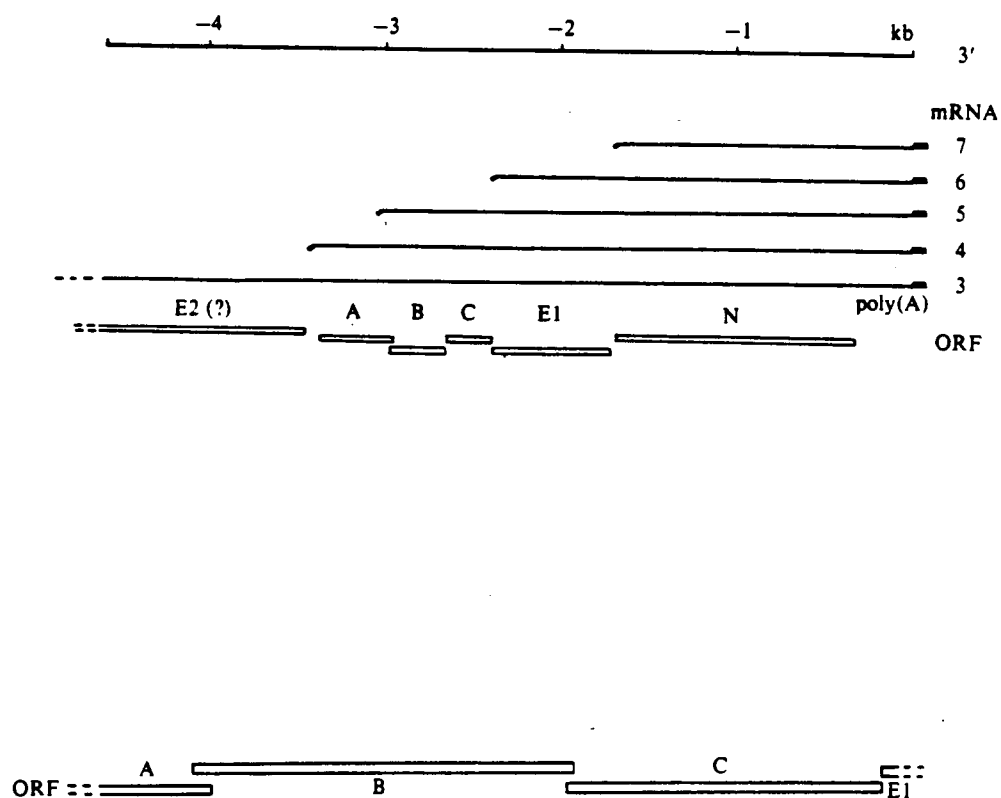


Figure I.6. Genome organization of the mouse hepatitis virus. This is adapted from Skinner et al. (1985). The abbreviations are: N for the nucleocapsid protein gene, E1 for the matrix or membrane glycoprotein gene, E2 for the peplomer protein gene, A, B and C for nonstructural protein genes. The expansion at the bottom illustrates the regions 16, 5, and 1 base overlaps between the open reading frames of A and B, B and C, and C and E1 respectively.

Skinner and Siddell, 1983). The cloned N sequence for MHV-JHM was shown to contain the nucleocapsid protein information by hybrid arrest of translation. This confirmed earlier in vitro translation data. The predicted protein size for this gene is 49,600, and the protein has a net positive charge.

Sequence analysis for the 5' unique region of mRNA 6 which encodes the E1 membrane glycoprotein (or gp26) is published for only the MHV A59 strain (Armstrong et al., 1984). The nucleotide sequence predicts a protein having a high percentage of hydrophobic amino acids. There is also no apparent cleavage site for a signal peptide on the basis of consensus sequence information. The amino terminus contains two serine residues which would be predicted from the known O-glycosylation site(s) in the protein.

The open reading frames labeled B and C in Figure I.6 describe the potentially dicistronic unique region of mRNA 5 (Skinner et al., 1985). These proteins are probably nonstructural and there is preliminary evidence to say that the ORF labeled C is used in vivo. ORF's B and C encode a 12,400 molecular weight basic protein and a 10,200 molecular weight neutral protein respectively. Sequence analyses suggest that the ORFs A, B, C, and E, overlap as shown in Figure I.6.

The ORF designated A in Figure I.6 is a nonstructural protein of about 14,000 molecular weight found in infected cells (Skinner and Siddell, 1985). The predicted amino acid sequence shows a large hydrophobic amino terminus and a basic carboxy terminus.

The intergenic regions between genes 5, 6, and 7 of MHV A59 have been identified and sequenced from clones of the viral genomic RNA

(Budzilowicz et al., 1985). Comparison of intergenic regions in clones derived from the genome and in clones derived from mRNAs allows identification of the sequences involved in joining leader RNA molecules to the body of the mRNA. This is shown in Figure I.5.

Sequence Analysis of Avian Infectious Bronchitis Coronavirus Nucleic Acid Sequences

As seen in Figure I.1 (page 5), IBV is distantly related to MHV (Brian et al., 1983). For IBV a large part of the genomic nucleotide sequence has been obtained and is discussed here in reference to the information we have for MHV. The organization of IBV open reading frames and mRNAs relative to the IBV genome is shown in Figure I.7 (Binns et al., 1985; Boursnell et al., 1985). Figure I.7 illustrates the genome organization for the first 4 kilobases of a 20 kilobase genome, beginning from the 3' polyadenylated end. The first difference between MHV and IBV that becomes immediately apparent is that IBV has only five subgenomic mRNAs compared to six for MHV. A second difference is that the second smallest mRNA (mRNA B) does not code for the matrix protein as does the second smallest message for MHV. For IBV, the third smallest mRNA (mRNA C) encodes the matrix protein. Thirdly, the genome size of IBV is approximately 20 kilobases compared to an estimated 18 kilobases for MHV. Despite the differences in gene arrangement on the genome between IBV and MHV, the 3' nested set pattern holds true for IBV (Stern and Kennedy, 1980b).

The sequences which have been completed to date for the IBV family include the nucleocapsid gene from two different strains of virus, the

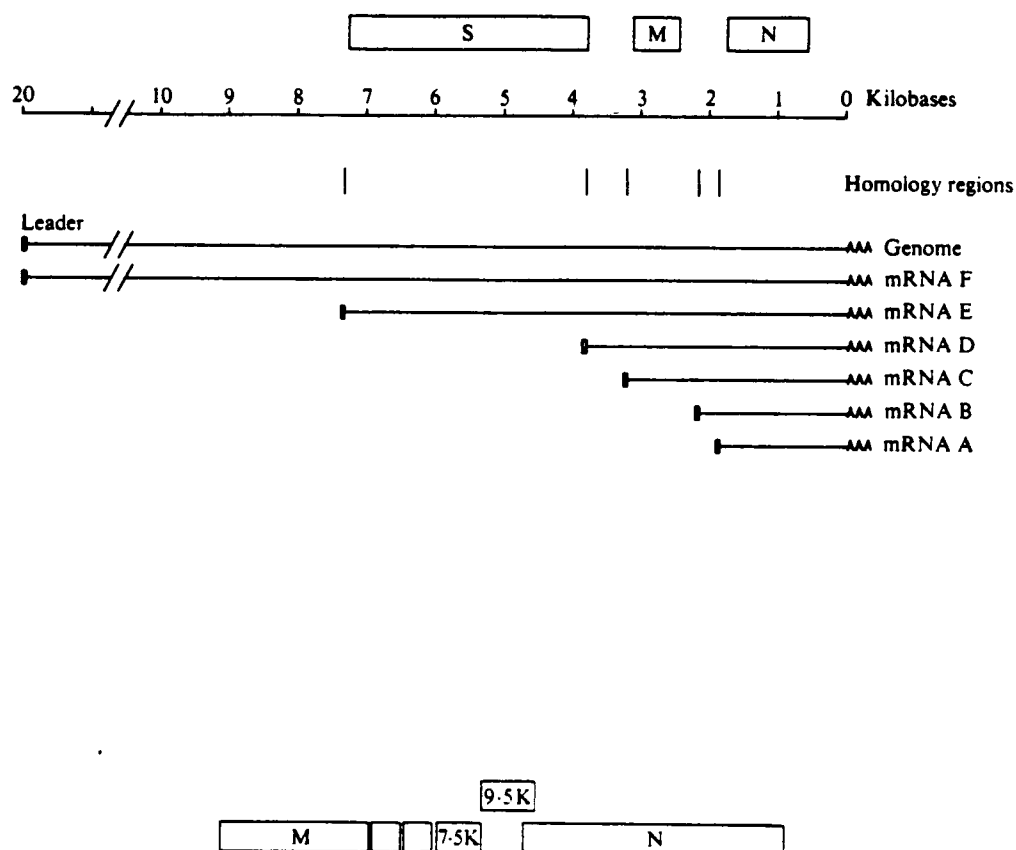


Figure I.7. Genome organization of the avian infectious bronchitis virus. This was adapted from previous figures (Binns et al., 1985; Boursnell et al., 1985). The arrangement of the subgenomic mRNAs as they map on the genome is depicted. The coding regions are designated as follows: S for spike protein gene, M for the matrix protein gene, N for the nucleocapsid protein gene, and unnamed boxes show possible coding regions for nonstructural proteins.

two open reading frames of mRNA B (for nonstructural proteins), the spike (or peplomer) protein gene, and the leader sequence by primer extension on mRNA A (Bournsnell et al., 1985; Bournsnell and Brown, 1984; Binns et al., 1985; Brown et al., 1984). The strategy for sequencing IBV differs from that used for MHV. The genes for IBV were identified by analyzing cDNA clones of genomic RNA not mRNA. This is a feasible approach because splicing does not occur in coronavirus transcription (Jacobs et al., 1982; Siddell et al., 1982).

The 3' most coding region on the genome (as seen in Figure I.7) is the nucleocapsid protein gene which has been sequenced for both the Beaudette and M41 strains (Bournsnell et al., 1985). Both sequences predict a protein with a molecular weight about 45,000. There is no nucleotide sequence homology between MHV and IBV, yet there is a short region (66 amino acids) of amino acid sequence homology between the nucleocapsid genes of the MHV and IBV. In spite of a lack of an overall amino acid sequence homology there is a parallel pattern in the clustering of serine residues along the nucleocapsid protein (Bournsnell and Brown, 1984).

The coding regions of mRNA B have been analyzed from cDNA clones of genomic RNA (Bournsnell and Brown, 1984). The sequence results predict two polypeptides with molecular weights of 7500 and 9500. These two open reading frames overlap by four bases. By results of codon bias analysis it seems likely that the first (most 5') open reading frame (7500 molecular weight) is the major translation product of mRNA B. The results of codon bias also predicted that it is unlikely that the 9500 molecular weight protein is real.

The gene for the spike or peplomer protein was also cDNA cloned from the genomic RNA (Binns et al., 1985). The location on the genome is indicated in Figure I.7. Sequence analysis predicts a polypeptide with a molecular weight of 127,006. This protein is found glycosylated and cleaved into two glycoproteins of 90,000 and 84,000 molecular weight in vivo (Binns et al., 1985). The predicted molecular weight agrees with the measured molecular weight when one considers the mass contributed by 28 sugar chains bound to the glycosylation site indicated from the amino acid sequence. The sequence analysis has also identified the putative signal and anchor sequences.

Sequence analysis of the intergenic regions between mRNAs A and B was made from cDNA clones prepared from genomic RNA (Brown and Boursnell, 1984). The location of these regions on the genome was confirmed by S1 mapping. There is no intergenic nucleotide sequence homology between IBV and MHV.

The leader sequence for IBV has been determined by primer extension of mRNA A (Brown et al., 1984). This research has identified a 60 base leader sequence and the location of the fusion point of the leader to the coding sequence of the mRNA. The existence of the leader sequence in the IBV family was not indicated by oligonucleotide fingerprint analysis as it was for MHV (Brown et al., 1984; Lai et al., 1983b).

RATIONALE FOR MOLECULAR BIOLOGICAL STUDIES ON THE BOVINE ENTERIC CORONAVIRUS

There are four reasons that justify molecular biological studies on the bovine enteric coronavirus and they can be described as follows:

(1) The BCV proteins responsible for inducing protective immunity and their corresponding genes need to be identified and characterized such that they can be employed in potentially useful vaccines manufactured using recombinant DNA technology. The bovine enteric coronavirus 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, 1978). Vaccines in use today, those developed by tissue culture attenuation of field strains of virus, are of dubious value.

The enteritis caused by BCV is severe to the neonate and so new approaches to develop immunity in the neonate, perhaps by induction of immunity prior to birth, remains a challenge. During infection with BCV, three factors are important in the pathogenesis of the disease and underline the challenge for developing effective vaccines to protect the young before birth or immediately after birth (Mebus et al., 1975). (a) BCV infects preferentially the immature villus epithelium. (b) BCV causes atrophy of the villous and colonic ridges causing a reduced surface area for absorption of food. (c) There is a continuous reinfection of the intestinal epithelium.

There remains, therefore, the challenge to develop (a) effective vaccines and (b) new vaccination approaches for the prevention of this disease.

(2) Cloned genes of BCV can serve as hybridization probes that will be useful for several purposes:

(a) Development of diagnostic tests for the rapid diagnosis of acute clinical infections. This will serve as an important model for the rapid diagnosis of human enteric viral infections, especially putative coronaviral infections from which human coronaviruses cannot be successfully cultured in vitro (Resta et al., 1985). It will also be useful for epidemiological studies in which a BCV zoonosis is implicated (Storz and Rott, 1981).

(b) Testing the hypothesis that coronaviruses related to BCV, such as the human coronavirus OC43 (Lapps and Brian, 1985), can cause chronic diseases in humans and animals as a result of persistent infections. The hepatitis and demyelinating encephalitis diseases in mice caused by the mouse hepatitis virus strains are the result of persistent infections (Wege et al., 1982). It is known that bovine coronavirus can be shed from clinically normal animals (Crouch et al., 1985) suggesting that BCV, at least in some animals, is capable of establishing persistent infections. BCV hybridization probes can be used in experimental systems ranging from tissue culture to animals to examine the state of coronavirus infection in the tissue. One such experimental animal system will be marmosets in which a high incidence of chronic colitis has been described (Russel et al., 1984). In these

animals, a coronavirus-like particle antigenically related to the bovine coronavirus was found.

(c) Ultimately the role of specific coronavirus genes in maintaining the persistently infected state can be examined. Such studies will probe the mechanism(s) of persistent coronavirus infections.

(3) Cloned genes can be used to derive primary sequence information that can be used to examine the evolutionary relationships among coronaviruses. It is striking, for example, that BCoV and MHV can be closely related antigenically but be totally different with regard to the hemagglutinin protein. Does the hemagglutinin gene represent a sequence acquired by BCoV or one that was lost by MHV? The identification of a hemagglutinin protein in coronaviruses, was first accomplished for the bovine coronavirus (King et al., 1985). The synthesis and glycosylation structure of the BCoV hemagglutinin are well characterized (Hogue, personal communication). BCoV, therefore, remains the best hemagglutinating coronavirus with which to examine the structure of the hemagglutinin gene.

(4) Nucleotide sequence information will yield information on transcriptional and regulational control mechanisms. A thorough comparative study of the detailed nucleotide structures between BCoV and the closely related MHV, and between BCoV and the distantly related transmissible gastroenteritis coronavirus (Kapke, personal communication) and avian infectious bronchitis virus (Brown and Boursnell, 1985) will likely yield clues as to important structure-function relationships.

CHAPTER II

OLIGONUCLEOTIDE FINGERPRINTS OF

ANTIGENICALLY RELATED BOVINE

CORONAVIRUS AND HUMAN

CORONAVIRUS OC43

INTRODUCTION

The bovine enteric coronavirus (BCV) causes a severe enteritis in young calves (Mebus, 1978). The human respiratory coronavirus OC43 (HCV OC43) causes only a mild upper respiratory disease in humans of all ages (McIntosh et al., 1969). Recent evidence suggests that these viruses are structurally closely related. Both agglutinate erythrocytes from mice, rats, and chickens (Kapikian et al., 1972; McIntosh, 1974; Mebus et al., 1973), and both share antigens as determined by immunofluorescence (Pedersen et al., 1978). No biological or structural data exists to rigorously differentiate between these viruses. Because a significant fraction of the human population (7 to 69%) both in Europe and the United States carry antibodies that are able to neutralize, immunoprecipitate, and indirectly immunofluoresce the bovine coronavirus (Storz and Rott, 1981), the possibility exists that the bovine coronavirus and the human coronavirus OC43 are two names for the same virus with zoonotic potential.

In this report we describe studies designed to compare genome sequences between the two viruses. The bovine coronavirus (Mebus strain) and the human respiratory coronavirus OC43, when grown on the same human cell line, share a sequence homology of greater than 96% as determined by T_1 oligonucleotide fingerprinting but each possess unique oligonucleotides and are thus distinctly different viruses. Both viruses incorporated a significant amount of subgenomic 0.68 Md

RNA late in the replication cycle and this may have implications for viral RNA sequence analyses.

MATERIALS AND METHODS

Cells and Viruses

The human rectal adenocarcinoma cell line HRT-18 (Tompkins et al., 1974) was obtained from J. Laporte, France, and was grown as monolayers in Dulbecco Modified Eagle Medium containing 50 ug gentamicin per ml and 5% fetal calf serum (Sterile Systems, Logan, Utah) heat inactivated at 56°C for 30 min. The Mebus strain of BCV (King and Brian, 1982; Mebus et al., 1973) which had undergone approximately 60 passages in tissue culture was cloned by four successive isolations from single plaques. The human respiratory coronavirus OC43 which had undergone seven passages in human embryonic tracheal organ culture and 15 passages in suckling mouse brain (McIntosh et al., 1969) was obtained from S. Weiss, University of Pennsylvania, Philadelphia, Pennsylvania, and was cloned by two successive isolations from single plaques. For each virus, a clone was passaged twice at a multiplicity of <0.1 PFU per cell, and then viral stocks were prepared from passages 3 through 6 by infecting cells at a multiplicity of approximately 0.5 PFU per cell. Viral titers ranging from 10^7 PFU per ml for OC43 to 10^8 PFU per ml for BCV were obtained in stock virus preparations.

Radiolabeling and Purification of Virus

Confluent monolayers of cells grown in 150 cm² flasks were infected with a multiplicity of 1-5 PFU per cell, rinsed, and fed with

20 mls per flask of medium containing 10% normal phosphate concentration, 20% fetal calf serum and 50 uCi of ^{32}Pi per ml. Infected cells were incubated 72-96 h at 37°C and virus was purified from supernatant fluids as described previously (Brian et al., 1980) except that sucrose was dissolved in TMEN (50 mM Tris-acid maleate [pH 6], 0.1 M NaCl, 1 mM EDTA).

Analysis of RNA

RNA was extracted from purified virus using SDS, proteinase K and phenol as described previously (Brian et al., 1980).

For electrophoretic analysis virion RNA was denatured in glyoxal and dimethylsulfoxide, and electrophoresed on 1% agarose gels in a vertical slab apparatus of 10 cm x 14 cm x 3 mm dimensions using the method of McMaster and Carmichael (McMaster and Carmichael, 1977). Gels were dehydrated in two successive 30 min baths of 100% methanol, compressed by blotting to a thickness of <1 mm, and exposed to X-Omat film.

For T_1 -oligonucleotide fingerprint analysis virion genomic RNA species were isolated by rate zonal sedimentation on preformed linear gradients of 30 to 15% (wt/wt) sucrose containing 0.1% SDS, 0.02 M Tris-hydrochloride, pH 7.4, 0.1 M NaCl, and 0.001 M EDTA, for 1.5 h at 110,000 x g in a Sorvall 650 rotor, at 25°C. Fractions (0.2 ml) were collected and the distribution of radioactivity was determined by Cerenkov radiation. RNA was precipitated at -20°C for 16 h by adding 2 volumes of 100% ethanol and 0.1 volume of 2 M sodium acetate. RNA was digested to completion with RNase T1 and electrophoresed in two

dimensions on polyacrylamide gels according to the method of DeWachter and Fiers (1972) as modified by Clewley et al. (1977).

RESULTS

Virion RNA

Experiments in which BCV RNA was labeled with [³H]-uridine for the first 48 h postinfection and electrophoretically analyzed on formaldehyde-agarose gels revealed only a high molecular weight species that comigrated with porcine transmissible gastroenteritis virus genomic RNA (Brian et al., 1980; data not shown). BCV genomic RNA therefore has a molecular weight of approximately 6.8×10^6 . When virion RNAs from both BCV and OC43 were labeled in vivo with ³²P-orthophosphate for 72 to 96 h postinfection and analyzed on denaturing glyoxal-agarose gels, approximately 55% from each virus migrated as a 6.8×10^6 M.W. species (Figure II.1). Approximately 10% migrated as a distinct species with an electrophoretic mobility close to that of 18S ribosomal RNA and therefore had an apparent molecular weight of 0.68×10^6 (Figure II.1). Approximately 15% of the RNA from each virus migrated as heterogeneous small molecular weight RNA between 4S and 18S in size. The remainder of the RNA was heterogeneous and was distributed throughout the gel.

Genome Sequence Homology as Determined by Oligonucleotide

Fingerprinting

To measure the degree of relatedness between BCV and HCV OC43, high molecular weight genomic RNA was isolated for each virus by rate

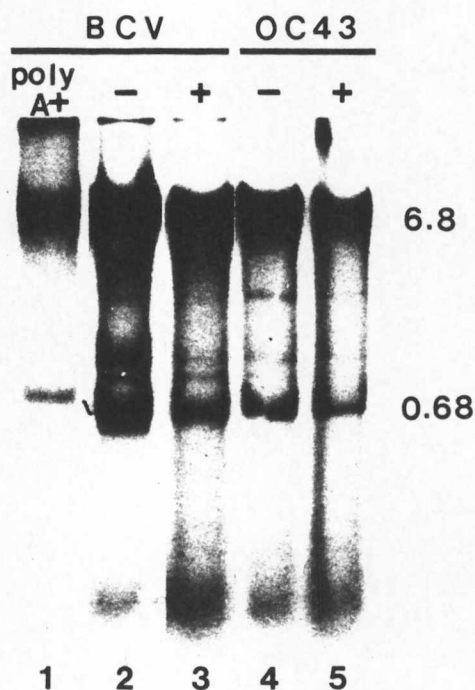


Figure II.1. Agarose gel electrophoresis of virion RNA after glyoxal-dimethylsulfoxide denaturation. Lanes: 1, BCV RNA that bound to oligo(dT)cellulose (Type 7, PL Biochemicals, Inc.) under conditions of 0.12 M NaCl, 0.01 M Tris HCl (pH 7.5), 1mM EDTA, 0.2% SDS and eluted with water; 2, BCV RNA from purified virus, 3, BCV RNA from virus treated with 20 ug bovine pancreatic ribonuclease A (Sigma) per ml for 1.5 h at 37°C prior to virus purification; 4, OC43 RNA from purified virus; 5, HCV OC43 RNA from virus treated with 20 ug bovine pancreatic ribonuclease A (Sigma) per ml for 1.5 h at 37°C prior to virus purification. 6.8 and 0.68 refer to molecular weight $\times 10^{-6}$. Percentage composition was determined by slicing the radioactive bands from the agarose, using the autoradiogram as a template, and counting the gel slices in Scintiverse (Fisher).

zonal sedimentation in sucrose gradients, fingerprinted separately, and then as a mixture using an equal number of counts (Figure II.2).

Inspection revealed a high degree of similarity between the two viruses consistent with earlier reports of close antigenic relatedness (Gerna et al., 1981; Pedersen et al., 1978) and our own recent work (Hogue et al., 1984). Of the 27 large resolvable oligonucleotides for BCV, 9 were unique to BCV, and of the 26 large resolvable oligonucleotides for HCV OC43, 8 were unique to HCV OC43. 18 oligonucleotides were in common and represent a comigration of 66-69% of the large oligonucleotides. Assuming that sequence homology demonstrated in the large oligonucleotides is representative of the remaining genome, we conclude there is a common nucleotide sequence of greater than 96% between the two virus genomes based on computer simulation of T₁ oligonucleotide mapping (Aaronson et al., 1982).

Properties of the Virion 0.68 Md Species

Incubating virus with pancreatic ribonuclease prior to purification, under conditions that completely destroy the integrity of ribosomal RNA, failed to destroy the 0.68 Md RNA (Figure II.1). From this we conclude the species is intravirion. The 0.68 Md RNA from BCV annealed to oligo(dT) cellulose indicating that it is polyadenylated (Figure II.1) and annealed to ³²P-labeled cloned cDNA representing the 3' terminal 1.19 Kb sequence of BCV genomic RNA (data not shown) indicating that this RNA is either an encapsidated subgenomic virus mRNA or possibly a defective genome of the type described for defective interfering particles of other positive-strand viruses.

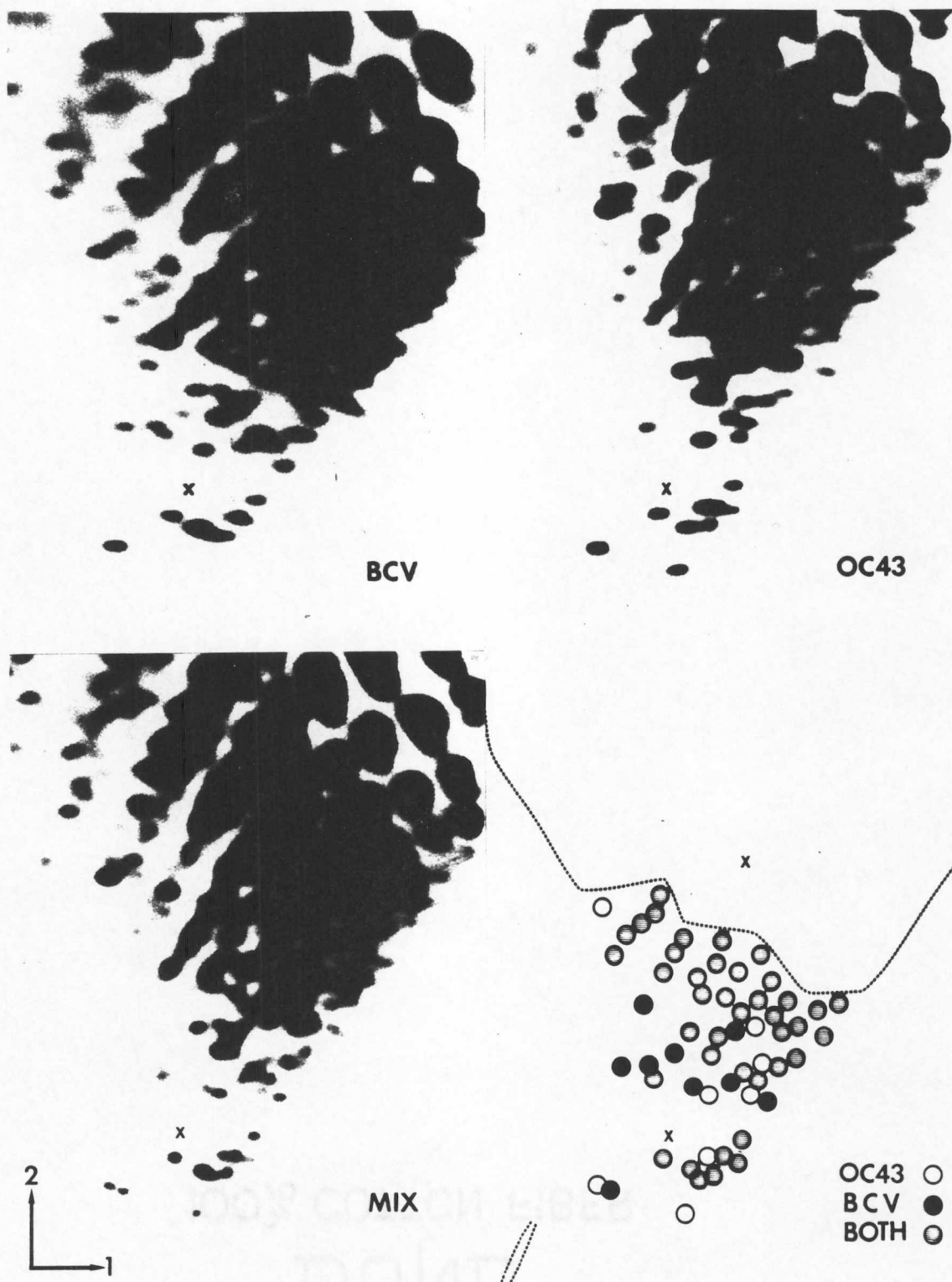


Figure II.2. Oligonucleotide fingerprints of BCV and HCV OC43 genomic RNAs. 1 to 3×10^5 Cerenkov counts were used per gel. Spots common to both viruses, as determined by a mixing experiment, are hatched in the schematic diagram.

DISCUSSION

Our studies demonstrate that BCV (Mebus strain) and HCV OC43 can both be grown on the human rectal adenocarcinoma cell line (HRT-18). Other strains of BCV, including an isolate from France (Laporte and Bobulesco, 1981), the LY-138 strain (Hajer and Storz, 1979) and a primary isolate from the University of Tennessee, College of Veterinary Medicine (our unpublished observations), and HCV OC43-like viruses (Macnaughton, 1982) can also grow on these cells without adaptation. Because the HRT-18 cells retain many properties of differentiated tissue cells (Tompkins et al., 1974), these observations are consistent with the notion that there is no rigid species restriction on the growth of these coronaviruses and that they may have zoonotic potential. BCV reportedly caused diarrhea in one investigator and the virus was reisolated from this person (Storz and Rott, 1981).

Our studies further establish that BCV (Mebus strain) and HCV OC43, although similar, are not identical viruses. The cytopathic effects caused by BCV and HCV OC43 are not identical (data not shown). BCV causes a foamy appearance of the cell cytoplasm and a clumping of infected cells and HCV OC43 causes a disintegration of the cell into small pieces. Both cause small (1-2 mm) opaque plaques under agar and both can be detected by hemadsorption following agar removal. No difference in size can be detected between the genomes of BCV and HCV OC43 by one-dimensional gel electrophoresis of glycosylated RNA. The genome for both viruses therefore measures approximately 6.8×10^6 mol. wt. since the genome of BCV comigrates with the 6.8×10^6 mol.

wt. genome of the porcine transmissible gastroenteritis virus in agarose gels after being denatured with formaldehyde (Brian et al., 1980; data not shown).

The comparative analysis of oligonucleotide fingerprints between the BCV and HCV OC43 genome RNAs suggests they have a sequence divergence of 3-4%. The exact degree of divergence cannot be determined using this technique since the large oligonucleotides probably represent less than 10% of the entire genome, but clearly these are not identical viruses. A sequence divergence of 3-4% is consistent with a distinct difference in the behavior of 3 of the 4 homologous proteins we observe between the two viruses (Hogue et al., 1984).

In preparations of BCV and HCV OC43 labeled with ^{32}P orthophosphate for 72-96 h we found a 0.68×10^6 mol. wt. species representing 10% of total virion RNA. For BCV we have shown this to be intravirion, polyadenylated, and subgenomic. Its function is unknown. It was not detected in earlier studies on BCV that employed ^3H -uridine for labeling periods of less than 48 h (Guy and Brian, 1979). Since our studies here use labeling periods of greater than 48 h it is possible that the 0.68×10^6 mol. wt. species becomes incorporated only late in infection and may represent adventitious encapsidation of a viral subgenomic messenger RNA of the type described for several coronaviruses (Dennis and Brian, 1982; Lai et al., 1981; Stern and Kennedy, 1980). Further studies are needed to confirm this interpretation.

CHAPTER III

CLONING AND SEQUENCE ANALYSIS OF THE BOVINE CORONAVIRUS NUCLEOCAPSID PROTEIN GENE

INTRODUCTION

The bovine enteric coronavirus (BCV) is one cause of severe neonatal enteritis in calves and is responsible for approximately one quarter of all deaths due to this disease (House, 1978). Vaccines produced from tissue culture attenuated strains of virus are of dubious value. Before attempting to develop vaccines by recombinant DNA that may have improved usefulness it is imperative that the genes and gene products responsible for inducing protective immunity be thoroughly characterized. Toward this end, and for the purpose of determining the function of individual proteins in virus replication, we have begun to clone and sequence the BCV genome.

BCV is known to possess a single-stranded, nonsegmented, polyadenylated RNA genome of approximately 20 Kb (Guy and Brian, 1979; Lapps and Brian, 1985). The total number of genes encoded by the genome is not known, but presumably, because of its close antigenic relatedness to the well-characterized mouse hepatitis coronavirus, BCV will be similar to MHV in the number and arrangement of genes on the genome. For MHV, the pattern, from 3' to 5', is a nucleocapsid protein (N) gene of approximately 1.6 Kb, a matrix protein (M or E1) gene of approximately 0.7 Kb, two overlapping genes for nonstructural proteins D and E, together a sequence of approximately 1.0 Kb, a peplomeric protein (P or E2) gene of approximately 4 Kb, a nonstructural protein B gene of approximately 2 Kb, and a nonstructural protein A (presumably the polymerase) gene of approximately 8 Kb (Siddell et al., 1982). One striking dissimilarity between BCV and MHV, however, is the possession

by BCV of a fourth major structural protein, the 140 Kd hemagglutinin protein that is comprised of two disulfide-linked subunits of 65 Kd each (Hogue et al., 1984). This protein is absent on MHV. Questions therefore arise concerning the origin and function of this protein for virus replication. Is it an important protein for imparting protective immunity to the animal? At this time we have special interest in identifying the gene for this protein and its location in the genome.

In this paper we describe experiments that begin to examine the BCV genome by cDNA cloning and DNA sequencing to rigorously characterize its structure. This experimental approach also allows us to deduce the primary structure of the protein products and to predict some of their properties. cDNA clones were prepared from the 3' end of the polyadenylated genome and were sequenced. One gene identified as the nucleocapsid protein (N) gene was encoded within the first 163 nucleotides. It identified a protein of 448 amino acids having a calculated molecular weight of 49,379 and extensive sequence homology with the N protein of mouse hepatitis virus. A second potential gene was identified internally within the N gene, but in a different reading frame, and began at base 1473 from the 3' end. This gene potentially encodes a peptide of 207 amino acids having a molecular weight of 23,057. The N gene is flanked on its immediate 5' side by part of the matrix protein gene and in this manner shares the same relative position in the genome as does the N gene for the mouse hepatitis virus.

MATERIALS AND METHODS

Virus and Cells

The Mebus strain of bovine coronavirus (BCV) was plaque purified and grown on the human rectal tumor (HRT) cell line as previously described (Lapps and Brian, 1985).

Purification of Genomic RNA

Virus was purified from clarified supernatant fluids as previously described (Lapps and Brian, 1985). Viral RNA in one out of 10 flasks was labeled with ^3H -uridine (400 Ci/mmmole, ICN), 20 uCi/ml, in order to follow purification of the RNA. Viral RNA was extracted by dissolving the virus pellet in 0.5 ml TNE (10 mM Tris hydrochloride [pH 7.5], 100 mM NaCl, 1 mM EDTA), 2% SDS, 0.5 mg proteinase K per ml for 10 min at 37°C, and extracting once with an equal volume of 50% phenol, 48% chloroform, 2% isoamyl-alcohol, a procedure shown to allow extraction of polyadenylated RNA species (Perry et al., 1974). RNA was precipitated at -80°C after adding 0.1 volume 3M sodium acetate and 3 volumes of absolute ethanol. Because subgenomic RNA species are incorporated into BCV (Lapps and Brian, 1985), full-length genomic RNA to be used for cDNA cloning and making probe for colony screening was size selected. Size selection was done by rate-zonal sedimentation on preformed linear gradients of 15% to 30% sucrose (wt/wt) made up in TNE 7.5-0.1% SDS. RNA was dissolved in water and sedimented 1.5 h at 110,000 X g, 25°C, on 5 ml gradients. 0.2 ml fractions were collected and the distribution of radioactivity was determined by scintillation spectrophotometry on 5 ul aliquots of each fraction. Only RNA sedimenting faster than mammalian 28S ribosomal RNA was recovered by

ethanol precipitation. RNA was dissolved in water and reprecipitated with ethanol and sodium acetate twice before being used in the experiments described below. RNA was quantitated by spotting 1 ul amounts of dissolved RNA onto agar containing ethidium bromide (Maniatis, 1982).

cDNA Cloning of the 3' End of the BCV Genome

BCV genomic RNA was cloned using a modified method of Gubler and Hoffman (Gubler and Hoffman, 1983). First strand synthesis was carried out in a reaction volume of 50 ul containing 50 mM Tris hydrochloride (pH 8.1), 148 mM KCl, 8 mM MgCl₂, 1 mM DTT, 2 mM dATP, 2 mM dCTP, 2 mM dGTP, 2 mM dTTP, 10 uCi [³²P] dCTP (3000 Ci/mmmole, ICN), 15 units RNasin (Promega), 50 pmoles oligo dT₁₂₋₁₈, 10 ug BCV RNA, 20 U reverse transcriptase (Seikagaku), for 1 h at 37°C, and the reaction was stopped by adding 2.5 ul 0.5 M EDTA. The BCV RNA was heated to 100°C for 5 min and quick cooled to 37°C immediately before adding it to the reaction. Nucleic acids were extracted with phenol-chloroform-isoamyl alcohol and ethanol precipitated after the addition of 0.5 volume of 7.5 M ammonium acetate. The nucleic acid pellet was dried only briefly (30 sec in a Savant Speed Vac Concentrator) before dissolving for use in second strand synthesis.

Second strand synthesis was carried out in a reaction volume of 100 ul containing 20 mM Tris hydrochloride (pH 7.5), 5 mM MgCl₂, 10 mM (NH₄)₂SO₄, 10 mM KCl, 0.15 mM B-NAD-40 um dNTP's, 8.5 U/ml E. coli RNase H, 230 U/ml DNA polymerase I, 10 U/ml DNA ligase, all of the product from first strand reaction. All components excepting enzymes

were incubated 7 min at 37°C to assure all was in solution, and then cooled to 12°C. The reaction, after adding enzyme, was incubated at 12°C for 1 h, then at 22°C for 1 h. The reaction was stopped by adding 20 ul 0.5 M EDTA and products were extracted with an equal volume of phenol-chloroform-isoamyl alcohol and ethanol precipitated. The precipitate was dissolved in 50 ul water and precipitated with 3 volumes of ethanol after adding 0.5 vol of 7.5 M ammonium acetate, and the process repeated. The purpose of the repeated ethanol precipitations was to remove unreacted dNTP's. The total quantity of product was estimated from the amount of radiolabeled first strand that remained.

Double stranded cDNA was homopolymer tailed essentially by the method of Roychoudhury and Wu (Roychoudhury and Wu, 1980). The following were added to the dried DNA in order: 20 ul 10X cacodylate-tris buffer (1.4 M K cacodylate, 0.3 M Tris hydrochloride [pH 7.6], 4 ul 5 mM DTT, 3 ul 10 mM dCTP, 162 ul H₂O, 2 ul 100 mM CoCl₂, 50 uCi ³²P-dCTP (>3000 Ci/mole, ICN) in 5 ul, 16 units of TdT in 2 ul. The reaction was carried out at 37°C for 1 min then stopped by adding 10 ul 0.5 M EDTA and 2 ul SDS, and the nucleic acid was extracted with phenol-chloroform-isoamyl alcohol and ethanol precipitated twice. This reaction resulted in an average of 15 CMP residues added per 3' end of dsDNA, the optimal number for annealing and transformation (Peacock, 1981).

C-tailed dsDNA was annealed to G-tailed, Pst-I-linearized vector (PL Biochemicals) for 2 h at 58°C in a 30 ul volume of buffer containing 10 mM Tris hydrochloride (pH 7.6), 150 mM NaCl, 1 mM EDTA

(Peacock et al., 1981). The total concentration of DNA used was less than 0.1 ug/ml and the optimal insert:vector ratio was 1:10 on a mass basis. The optimal insert:vector ratio was determined empirically in a pilot experiment. In designing the pilot experiment, it was assumed that the average dsDNA molecule was 1 Kb.

E. coli strain JM103 was transformed by the method of Hanahan (Hanahan, 1983). Cells containing recombinant plasmids were observed as white colonies on YT agar plates that contained 100 ug ampicillin per ml, 1 mM IPTG, 0.004% X-gal (Heidecker and Messing, 1983). Recombinant colonies were transferred to nitrocellulose (Millipore, HATF) and probed with random-primed cDNA copied from BCV genomic RNA.

Identification of Large Inserts Containing 3'-Specific BCV Sequences

³²P-labeled, random-primed cDNA used for colony hybridization was synthesized as described above for the oligo dT-primed reaction except that dNTP concentrations were 2.5 uM each and 0.2 ug RNA were used, and oligo(dT) was replaced by 20 ug fragmented calf thymus DNA. After 1 h incubation, dNTP concentrations made 2 mM each and incubation was continued for 15 minutes. Probe was alkali treated to hydrolyze the RNA and was used for colony hybridization as described by Maniatis (Maniatis, 1982). Colonies yielding a strong signal were analyzed for plasmid size (Kado and Liu, 1981). Inserts of 1.0 to 4.0 Kb (the largest) were further analyzed by Southern hybridization with ³²P-labeled poly(dT) to detect poly(dA) content. ³²P-labeled poly(dT) probe was prepared as described above for the oligo(dT)-primed

reaction for the synthesis of first-strand cDNA except that 50 pmoles oligo(dT)·poly(rA) (PL Biochemicals) replaced the RNA.

Alkali-treated ^{32}P -poly(dT) probe was incubated for hybridization at 37°C for 12 h, then at 20°C for 36 h, and blots were washed in 2X SSC, 0.1% SDS at 20°C.

Restriction Endonuclease Mapping

Plasmid was purified by alkaline lysis and cesium chloride centrifugation as described by Maniatis (Maniatis et al., 1982). Restriction endonuclease mapping was done essentially as described by Smith and Bernstiel (Smith and Bernstiel, 1976) using plasmids that were labeled at the Sal I site within the multiple cloning linker region.

DNA Sequencing and Sequence Analysis

Restriction fragments end-labeled with ^{32}P were isolated and sequenced by the method of Maxam and Gilbert (Maxam and Gilbert, 1980). Sequences were analyzed with the aid of the program developed by Queen and Korn (Queen and Korn, 1984) and sequence homologies were searched against Genbank, both marketed as part of the Beckman Microgenie program, March 1985 version (Beckman Instruments, Inc.).

RESULTS

cDNA Cloning and Sequencing of Three Clones from the 3' End of the Genome

Starting material for cDNA cloning was approximately 3 ug of rate-zonally purified genomic RNA obtained from 500 ml of tissue

culture medium. An estimated 70 ng of ds cDNA was obtained, as determined by radiolabel incorporation during second strand synthesis, and from this 670 white colonies were obtained. By colony screening 89 colonies gave a strong signal to ^{32}P -labeled cDNA prepared from genomic RNA, and of these, 9 had inserts of 1.2 to 4.1 kilobases as determined by agarose gel electrophoresis of supercoiled plasmids, and were further analyzed by restriction enzyme analysis and poly A content. Only one of the clones, an insert of 1.2 kilobases identified as clone CB9, reacted strongly under hybridization conditions by Southern blotting to ^{32}P -labeled oligo(dT). Three other clones identified as MN3 (2.1 Kb), MA5 (2.8 Kb) and MA7 (4.1 Kb) were found to contain sequences that overlap with CB9 on the basis of hybridization and overlapping restriction endonuclease maps (data not shown). Clone MN3 was chosen for complete sequencing. Clone CB9 was sequenced in part to confirm the extreme 3' end of the genome and to confirm other regions within clone MN3. Clone MA5 was sequenced in part to confirm regions of the MN3 clone.

The orientation of clones CB9, MN3, and MA5 in reference to the 20 Kb virus genome and the restriction enzyme sites used for sequencing are illustrated in Figure III.1. Our orientation presumes polyadenylation at only the 3' end of the genome and this is based on the precedent of the documented 3' polyadenylation site in the avian infectious bronchitis virus and mouse hepatitis virus genomes (Lai and Stohlman, 1981; Stern and Kennedy, 1980). The strategy used for sequencing is described in the legend to Figure III.1. Greater than 98% of the sequence containing the complete N gene was determined

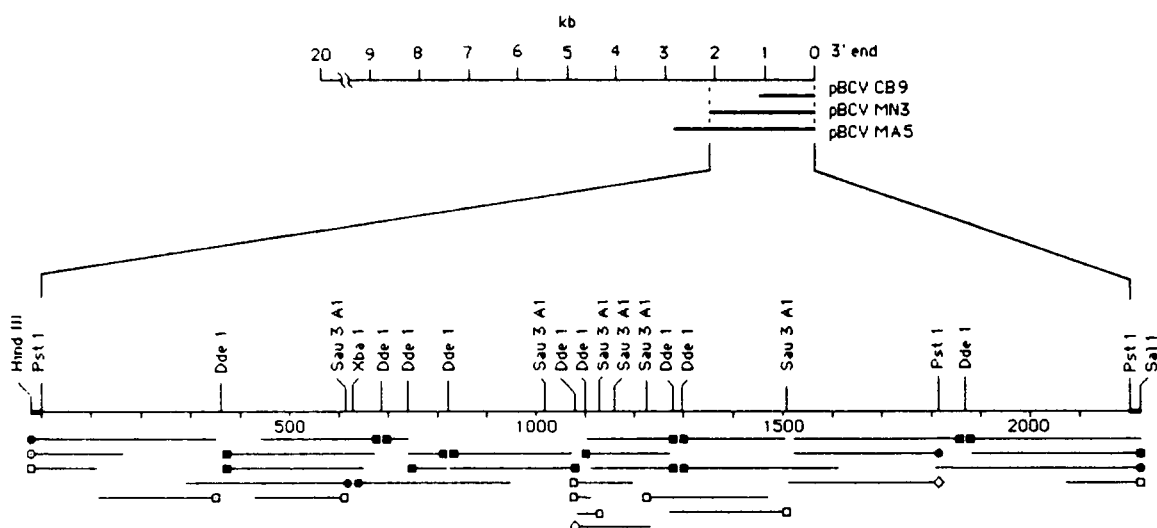


Figure III.1. Restriction endonuclease map and sequencing strategy for BCV cDNA clones CB9, MN3 and MA5. The orientation of clones in the pUC9 vector are as shown (except CB9 which is inverted, ie., the poly A end is next to the Hind III site of the vector). pUC9 sequences are indicated as a bold line at the ends of the sequences. The internal Dde I, Pst I, Sau 3A1, and Xba I sites were derived by restriction endonuclease mapping, and the Hind III and Sal I sites in the multiple cloning region of the pUC9 vector were the sites used for initial DNA sequences. ■, □, and ◻ indicate sites labeled at the 5' end using polynucleotide kinase for clones MN3, CB9, and MA5 respectively. ●, ○, and ⊙ indicate sites labeled at the 3' end using reverse transcriptase, and the appropriate labeled deoxynucleotide triphosphate, for clones MN3, CB9, and MA5 respectively. ◇ indicates sites labeled at the 3' end using radiolabeled cordycepin and terminal transferase for a subclone of MN3. Uniquely labeled molecules for sequencing were obtained by gel recovery after restriction endonuclease treatment of an end labeled fragment or by strand separation.

either by sequencing both strands of the cloned DNA, or by repeated sequencing of the same strand using different methods of end labeling. Some of the sequences were confirmed from subclones of MN3.

The total sequence of MN3 is illustrated in Figure III.2. Sequences of CB9 that overlap with MN3 are identical with those of MN3 except that the total length of the poly A tail is 70 for CB9 and 0 for MN3.

All possible translation products were determined for both virus-sense RNA and virus complementary-sense RNA because of the precedent that some single-strand RNA virus genomes are of ambisense polarity (Auperin et al., 1984) (Figure III.3). RNA complementary to coronavirion-sense RNA could therefore theoretically function as mRNA. Figure III.3 illustrates the existence of two large open reading frames labeled N (for nucleocapsid protein) and IORF (for "internal" open reading frame) when virus-sense RNA is translated in all three reading frames, considering that an open reading frame is preceded by a termination codon and is initiated by an appropriate initiator codon. A third region labeled M (for matrix protein) is apparently a continuation of an open reading frame since there is no stop codon within or preceeding this sequence present. From the virus complementary-sense RNA, two open reading frames of greater than 250 bases exist, beginning at approximate nucleotide positions 1250 and 1880 in the first reading frame.

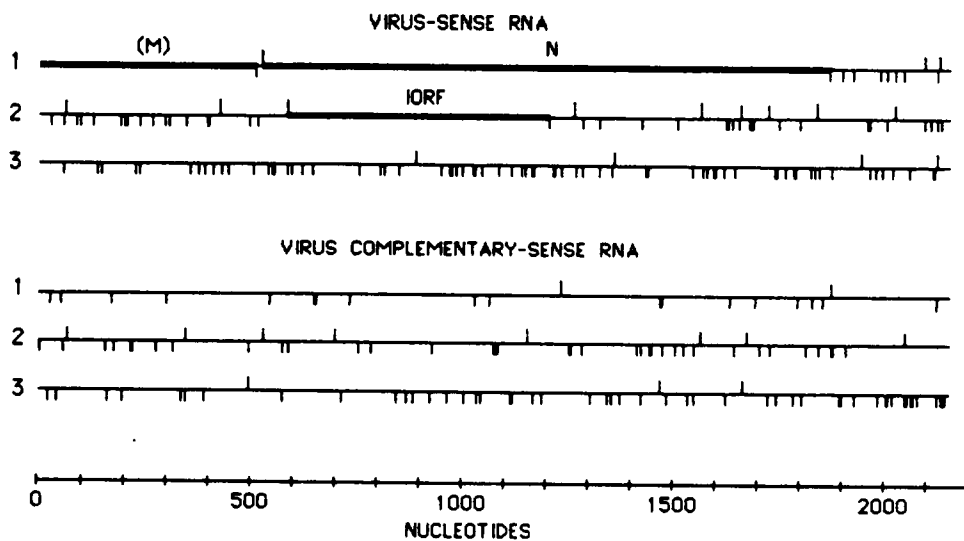


Figure III.3. Schematic diagram of open reading frames from clone MN3. Virus-sense open reading frames are from the virion sense RNA. Virus complementary-sense RNA are translations from negative-strand sense RNA. The 5' to 3' orientation is always left to right.

The Largest Open Reading Frame Predicts a Protein Having Properties of the Nucleocapsid Protein

The largest open reading frame extends from base 531 to base 1874 and predicts a 448 protein of 49,379 molecular weight. We conclude this is the coding sequence for the nucleocapsid protein (N) for the following reasons: (i) The only BCV protein described to date that approaches this size is the 52 Kd phosphorylated nucleocapsid protein (King and Brian, 1983). (ii) The predicted protein would be basic, a property expected of nucleic acid-binding proteins. Fifty-nine (13%) of the amino acids are basic whereas 43 (10%) are acidic, giving the protein a net charge of +16 at neutral pH. (iii) The amino acids encoded by this sequence share extensive (70%) sequence homology with the N protein of the closely related mouse hepatitis virus strains A59 and JHM (Armstrong et al., 1983; Skinner and Siddell, 1983).

The N gene for BCV shares other properties with the N gene of MHV. (i) It is flanked on its 5' side by the gene for the M protein, based on amino acid sequence homologies for this protein (data not shown). It is flanked on its 3' side by a noncoding region of 210 bases, similar to the 294 bases for MHV JHM. (ii) The intergenic region between the N and M genes, UAAACUUUAAGG, beginning with the termination codon of the M gene, is exactly the same. (iii) The consensus sequence at the initiator codon for the N gene, AGGAUGU, a preferred sequence for translation initiation by Kozak's rules (Kozak, 1983), is the same.

A Large Open Reading Frame Found Internal to the N Gene, but in a
Different Reading Frame, Predicts a 207 Amino Acid Protein Having a
Molecular Weight of 23,057

The nucleotide sequence from base 778 through base 1498 in the second reading frame encodes a 207 amino acid protein of 23,057 (Figure III.4). The protein is hypothetical since we yet have no proof for its existence. The protein is uncharged at neutral pH but is moderately hydrophobic since 79 (38%) of its amino acids are hydrophobic. The hydrophobic amino acids are spread somewhat evenly throughout the protein except at the carboxyl terminus where there are enough to make this part of the protein a potential membrane anchor region. Twenty-seven of the terminal 41 amino acids (66%) are hydrophobic. The consensus sequence GUAAUGG surrounding the initiator codon is one that is commonly used. It is found at the initiation site of 18% of all eukaryotic mRNAs catalogued (Kozak, 1983).

DISCUSSION

We present the primary nucleotide sequence for the BCV nucleocapsid protein (N) gene and the deduced amino acid sequence for the protein. This is the sixth coronavirus N gene to be sequenced and such sequence information allows one to examine potential evolutionary patterns among coronaviruses as well as to identify possible functional domains on the protein. An examination of nucleotide sequences near the ends the N gene also allows identification of possible regulatory signals involved in coronavirus RNA transcription and replication.

```

      619              649              679              709
ATGECATGCTTAAGTGGGGGATCAGTGGACCAATCTAGAAATGTTCAAAACAGGGGTAGAGAGCTCAACCCAAAGCAAACTGCTACTTCTCACTACCATCAGGAGGGAAATGTTGTAC
H A S L S G P I S P T N L E M F K P G V E E L N P S K L L L L S Y H Q E G H L Y

      739              769              799              829
CCTACTATTCTTGGTCTCTGGAACTACTCAGTTTCAAAAAGGAAGGAGTTTGAATTTCCAGAGGGACAGGTGTGCTTATTCCACAGGAGTCCAGCTACTGAACTAAAGGGTACT
P T I L G S L E L L S F K K E R S L N L Q R D K V C L L H Q E S Q L L K L R G T

      859              889              919              949
GGTACAGACACACAGAGGTTCTTTTAAACAGGGGATGCCAACAGGGTCAACTGCTGCCACGATGGTATTTTACTATCTTGGACAGGACCCCATGCCAAAGACAGTATGCCACGG
G T D T T D V L L K Q P H A T S V M C C H D G I F T I L E Q D R H P K T S H A P

      979              1009              1039              1069
ATATTGAAGGAGTCTTCTGGGTGCTAGTAAACAGGCTGATGTCAATACCCCGGCTGACATTTCTGGATGGGACCCAAAGTAGCGATGAGGCTATTTCCAGTAGGTTTCCGGCTGGCACGG
I L T E S S G S L V T R L H S I P R L T F S I G T Q V A H R L F R L G F R L A R

      1099              1129              1159              1189
TACTCCCTCAGGGTTACTATATTGAAGGCTCAGGAGGCTCTGCTCTAATTCAGATCTACTTCACCCCATCCAGTAGAGGCTCTAGTGCAGGATCGGCTAGTAGAGCCAAATCTGGCA
Y S L R V T I L K A Q E G L L L I P D L L H A H P V E P L V Q D R V V E P I L A

ACAGAACCCCTACCTCTGGTGTA
T E P L P L V U

```

Figure III.4. Nucleotide and amino acid sequence of the internal open reading frame (IORF).

The N protein of BCV shows an overall amino acid sequence homology of 70% with the N protein of MHV A59 (Siddell, 1983; Armstrong et al., 1983), and therefore appears to be closely related to MHV. This degree of homology is consistent with antigenic data suggesting the viruses are closely related (Pedersen et al., 1978). BCV and MHV, therefore, appear to be relatively recently diverged viruses. The N protein of BCV, however, shows an amino acid sequence homology of only 27% with each the avian infectious bronchitis virus (Bournsell and Brown, 1984) or the porcine transmissible gastroenteritis coronavirus (Paul Kapke, personal communication) suggesting that BCV is more highly diverged from these viruses.

The degree of homology between the N amino acid sequences of BCV and MHV is not evenly distributed throughout the gene. There are 8 regions of 6 to 18 amino acid stretches, for example, that show less than 30% homology. These begin at amino acid positions 6, 26, 51, 204, 334, 386, 396, and 434 on the BCV sequence. Conversely there are regions of up to 31 amino acid stretches showing greater than 90% homology. The 68 amino acid highly conserved region among IBV, TGEV and MHV JHM, a region showing 51% homology beginning at amino acid number 86 in the MHV JHM sequence (P. Kapke, Ph.D. dissertation), shows a 78% homology between BCV and MHV JHM. Such regions of highly conserved sequences suggest that there are evolutionary pressures for conservation of specific protein function.

Such homologous regions become even more striking when one looks at the potential sites for protein modification. The phosphorylation sites of the MHV N protein are the serine residues (Lai and Stohlman,

1981). Presuming this holds true for all coronaviruses, then certain "hot spots" for phosphorylation become apparent when the N protein sequences are compared (Figure III.5). By aligning the N proteins of MHV, BCV, IBV and TGEV in reference to the conserved 66 amino acid region, 3 clusters of 3-12 serine residues in common among all viruses become apparent at BCV amino acid positions 40-70, 180-225, and 300-350. The major cluster region is at amino acid positions 180-225. Such regions may indicate sites for protein-nucleic acid or protein-protein interactions.

The position of the N gene in the genome relative to the M gene and the 3' noncoding region, the identical intergenic sequence between the M and N genes, and the identical translation consensus sequence surrounding the initiator codon for the N gene, all support the argument that BCV and MHV may be very recently diverged. The intergenic sequences in TGEV and IBV, for example, share no resemblance to each other, and TGEV shares only some resemblance to the BCV and MHV intergenic sequence. Since the intergenic sequence is proposed to be critical for the leader (primer)-polymerase complex to initiate mRNA transcription (Budzilowicz et al., 1985), we predict the leader sequences of BCV will be identical to that for MHV.

No direct evidence exists for the 23,057 molecular weight protein encoded by bases 778 through 1498 in the second reading frame. However, proteins of this size exist both in the virion and in virus infected cells (B. Hogue, personal communication). A favorable consensus sequence surrounding its initiation codon would suggest that the open reading frame is used. The hydrophobic domain positioned at

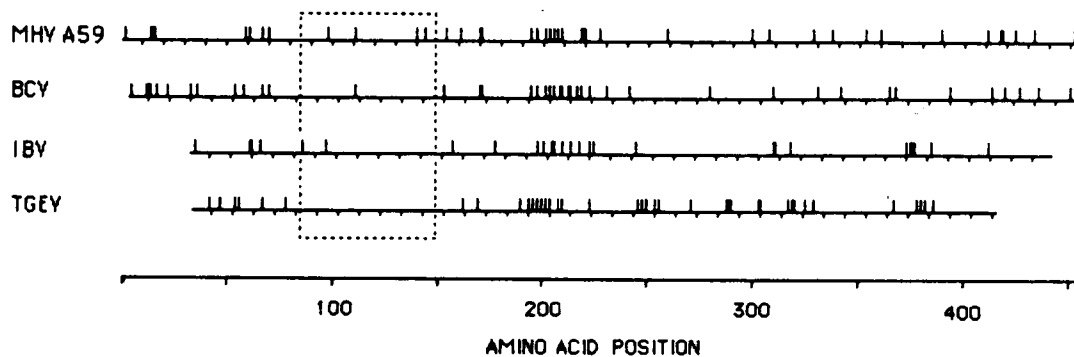


Figure III.5. Position of serines on the N protein of MHV, BCV, IBV and TGEV. The amino terminus is positioned on the left and the proteins are aligned by the 66 amino acid region of high homology indicated by the box. Serine positions are indicated by vertical bars.

the carboxy terminus is the most common structural arrangement for the hydrophobic domain for membrane attached proteins. A powerful approach to test the existence of this protein is to prepare synthetic oligopeptides from its sequence, raise antibodies against them, and directly attempt immunoprecipitation of the protein from infected cell lysates. This approach has worked to prove the existence of predicted measles virus proteins (Bellini et al., 1985).

The 3' terminal noncoding region of 210 bases in BCV, exclusive of the poly A tail, is the shortest noncoding sequence of those identified for coronaviruses thus far. The significance of the noncoding region is not completely known although it probably functions as an attachment region for the polymerase to initiate synthesis of the negative strand. One possible recognition site that is conserved among all coronaviruses sequenced to date is a sequence, GGGAAGAGCC, found beginning -79 basis from the beginning of the poly A tract (Figure 3.2). The homologous sequence begins at position -81 in IBV, -82 in MHV, and -77 in TGEV.

CHAPTER IV

CLONING AND SEQUENCE ANALYSIS

OF THE BOVINE CORONAVIRUS

MATRIX PROTEIN GENE

INTRODUCTION

The first viral protein shown to have O-linked glycosylation was the matrix protein of the bovine enteric coronavirus (Nieman and Klenk, 1980). Subsequently the matrix protein (also called E1) of the mouse hepatitis coronavirus was also shown to be glycosylated only by O-linkages to serine residues. More recently the herpes simplex virus glycoproteins (Olofsson et al., 1981; Wenske and Courtney, 1983) and the G protein of the respiratory syncytial virus (Collins and Wertz, 1985) were shown to have O-linked sugars.

The function of the O-linked sugars on the matrix protein is not yet known. Presumably they play a key role in targeting protein migration in infected cells during virus synthesis and assembly. The matrix protein gene for the mouse hepatitis virus has been cloned and sequenced (Armstrong et al., 1984). The protein is shown to have three strongly hydrophobic domains, and in direct experimentation has been shown to be intimately associated with the viral envelope, apparently crossing the lipid bilayer three times. The matrix protein is found associated with the endoplasmic reticulum but not the cytoplasmic membrane of infected cells. Since coronaviruses bud into the endoplasmic reticulum during assembly it is hypothesized that the matrix protein becomes first associated with the membrane of the endoplasmic reticulum, probably during translation, and serves as a signal for localizing nucleocapsid budding. The matrix protein and nucleocapsid protein in the bovine coronavirus virion exist in a molar ratio of 1:1 (King and Brian, 1982) and this ratio supports our

hypothesis that there is a direct interaction between these proteins. Perhaps the O-linked carbohydrates are involved in this interaction. One problematic feature in suggesting the role of O-linked sugars in this interaction is that the analogous matrix protein in the avian infectious bronchitis coronavirus contains only N-linked sugars (Stern and Sefton, 1982).

In this paper we describe the cloning and sequence analysis of the bovine coronavirus matrix protein gene. The deduced amino acid sequence reveals a homology of greater than 86% with the mouse hepatitis virus matrix protein. When amino acids of like properties are included, probably representing conservative codon mutations, the homology becomes greater than 92%. Such a conservation of sequence homology, especially as compared to a nucleocapsid homology of only 70% (perfect match only), suggests rigorous selection pressures for the M protein and implies an important function for this protein dependent upon its entire structure.

MATERIALS AND METHODS

cDNA Cloning and Identification of BCV 3'-Specific Clones

cDNA cloning, identification and characterization of 3'-specific BCV clones have been described in Chapter III.

DNA Sequencing and Sequence Analysis

DNA restriction mapping was done by the method of Smith and Bernstiel (Smith and Bernstiel, 1976), DNA sequencing was done by the chemical method (Maxam and Gilbert, 1980), and DNA and protein

sequences were analyzed using the program of Queen and Korn (1984) as described in Chapter III.

RESULTS

Sequence analysis of the entire 2.16 Kb 3'-specific cDNA clone MN3 of the BCV genome revealed the entire gene for the nucleocapsid (N) protein, the 3' noncoding region of the genome, the 3' poly A tail and a significant portion of the gene for the matrix protein (M) protein, as judged by amino acid homology with the M protein of MHV (Chapter III).

To complete the sequence of the M gene we have examined the nucleotide sequence of two additional clones that extend the primary structure of the genome toward its 5' end. Two clones that overlap in sequence with MN3 as determined by restriction map analysis but extend toward the 5' end of the genome are MA5 (2.8 Kb) and MA7 (4.0 Kb) (Figure IV.1). The entire MA5 sequence beginning from the third MA5 Dde restriction endonuclease site (sequence position 0 in Figure IV.1) to the beginning of the MN3 sequence (sequence position 286) was obtained by sequencing this region of MA5. The sequence from positions 283 to 915 (the Xba I site on MN3) is the same as shown for the left hand end of MN3 (Chapter III) and the sequencing strategy for this sequence is repeated here for completeness. The nucleotide sequence for map positions 0 to 915 of Figure IV.1 is shown in Figure IV.2. An examination of all possible open reading frames by computer analysis, Figure IV.3, reveals a large open reading frame, in reading frame 1 of virus-sense RNA, (identified as M) that is immediately preceded by

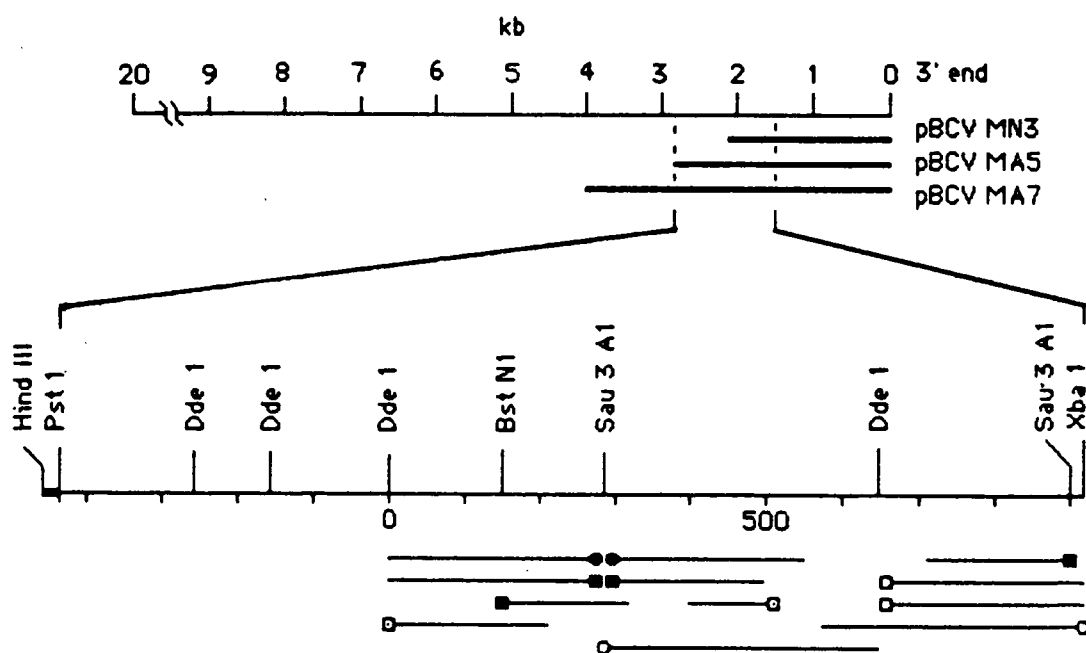


Figure IV.1. Restriction endonuclease map and sequencing strategy of BCV cDNA clones MN3, MA5, and MA7. The inserts from clones MN3, MA5 and MA7 are shown as they map along the genome. The orientation of these three inserts is the same in the vector pUC9. ■, □, and ○ indicate restriction sites that are 5' end labeled with polynucleotide kinase for clones MA5, MN3, and MA7 respectively. ● and ○ indicate sites that were 3' end labeled using reverse transcriptase and the appropriate labeled deoxynucleotide triphosphate for clones MA5 and MN3 respectively.

```

      30      60      90      120
AGTACTGTCCCTTCTATTATGTGTTAATAGAGGTAGGCAGTTTATGAGTTTTACAACGATGTA AACACCAGTTCTTGATGTGGATGACGTTTAGTTAATCCAACATTATGAGT
M S

      150      180      210      240
AGTGTAACTACACCAGCAACAGTTTACACCTGGACTGCTGATGAAGCTATTAAATTCCTAAAGGAATGGAACTTTCTTTGGGTATTATACTACTTTTATTACAATCATATTGCAATTT
S V T T P A P V Y T W T A D E A I K F L K E W N F S L G I I L L F I T I I L Q F

      270      300      330      360
GGATATACAAAGTCCAGTATGTTTGTATTATTAAGATGATCATTTTGTGGCTTATGTGGCCCTTACTATCATCTTAAC TATTTTCAATTGCGTGTATGCGTTGAATAATGTGTAT
G Y T S R S M F V Y V I K M I I L W L M W P L T I I L T I F N C V Y A L N N V Y

      390      420      450      480
CTTGGCTTTTCTATAGTTTCACTATAGTGGCCATTATCATGTGGATTGTGTATTTGTGAATAGTATCAGGTTGTTTATTAGAAGTGGAGTTGGTGGAGTTTCAACCCAGAAACAAAC
L G F S I V F T I V A I I M W I V Y F V N S I R L F I R T G S W W S F N P E T N

      510      540      570      600
AACTTGATGTGTATAGATATGAAGGGAAGGATGTATGTTAGGCCGATAATTGAGGACTAOCCTAACCTTACGGTCACAATAATACGTGGTCATCTTTACATGCAAGGTATAAACTAGGT
N L M C I D M K G R M Y V R P I I E D Y H T L T V T I I R G H L Y M Q G I K L G

      630      660      690      720
ACTGGCTATTCCTTGTTCAGATTTCACAGCTTATGTGACGTGTCTAAGGTCTCAACCTGCTCAAGTATAAGCGTGGTTTCTTGACAAGATAGGCGATAGTGTGTTTGTCTGTTTAT
T G Y S L S D L P A Y V T V A K V S H L L T Y K R G F L D K I G D T S G F A V Y

      750      780      810      840
GTTAAGTCCAAAGTCGTAATTACCGACTGCCATCAACCCAAAAGGGTCTGGCATGGACACCGCATTTGTTGAGAAATAATATCTAAACTTTAAGGATGTCTTTACTCTCTGTAAGCAA
V K S K V G N Y R L P S T Q K G S G M D T A L L R N N I M S F T P G K Q

      870      900
TCCAGTAGTAGAGGCTCTTTGGAAATCGTCTCTGGTAAATGGCATCCTTAAGTGGCCCGATCAGTCCGACCAATCT
S S S R A S F G N R S G N G I L K W A D Q S D Q S

```

Figure IV.2. Nucleotide and amino acid sequence of the bovine coronavirus matrix protein. The sequence is from the total region described in Figure IV.1 and includes the beginning of the nucleocapsid protein gene. Nucleotides bounding matrix protein are 115 to 804. The first nucleotide in the initiating methionine of the nucleocapsid protein gene is 817.

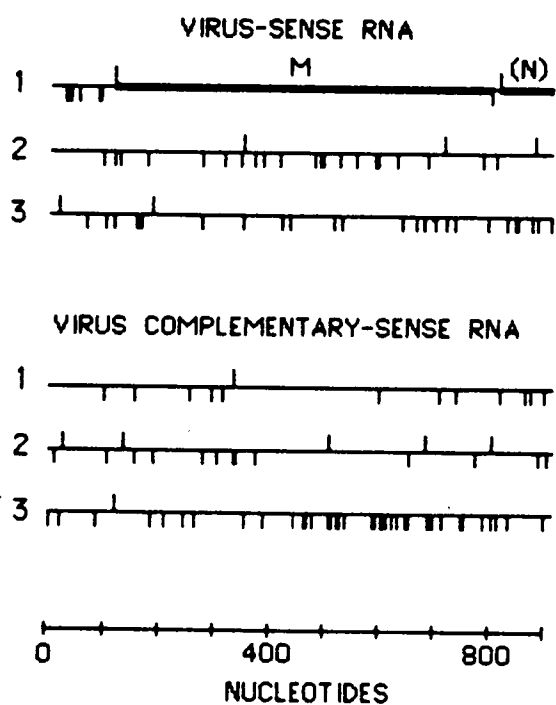


Figure IV.3. Schematic diagram of open reading frames that locate the bovine coronavirus matrix protein. Bars above the lines designate potential initiating methionine codons. Bars below the lines designate translation termination codons.

several stop codons. Potential open reading frames that would encode peptides of greater than 30 amino acids are found in reading frame 3 of virus-sense RNA (nucleotide position 180), and reading frames 1 (nucleotide position 330) and 2 (nucleotide position 510) of virus complementary-sense RNA were also identified (Figure 4.3).

Amino acid sequence homology examination reveals that the M protein of BCV is nearly identical to the M protein of MHV (Armstrong et al., 1984). The M protein of BCV is 230 amino acids long and that of MHV is 229. By maximum alignment, 198 of the amino acids (>86%) are the same, and another 15 (6%) represent conservative changes. The homology, including the conservative amino acids, is therefore >92%.

The protein is slightly basic having a net charge of +4 at neutral pH and is hydrophobic. One hundred thirteen (49%) of the amino acids are hydrophobic and the distribution of hydrophobic amino acids is identical to that for the MHV M protein.

DISCUSSION

Direct examination of the M protein of MHV reveals that it has no cleavable signal sequence, yet has the ability to become intimately associated with membranes either cotranslationally or posttranslationally (Armstrong et al., 1984). Furthermore, the protein becomes mostly embedded within the membrane with only a small portion of its amino terminus on the luminal side of the membrane (outside the virion) and a small portion of its carboxy terminus on the cytoplasmic side (inside the virion) (Armstrong et al., 1984). Since the protein is found only at the endoplasmic reticulum and is not transported to

the cytoplasmic membrane, as are many other viral membrane proteins, there are probably severe constraints on its structural possibilities. That is to say, it probably cannot tolerate evolutionary divergence in its structure and still retain its specialized function(s). The fact that the M proteins of BCV and MHV are highly conserved (86% perfect homology) while other proteins such as the nucleocapsid protein are more diverged (70% perfect homology) supports this hypothesis. The hypothesis also predicts a similarly high degree of sequence homology among the M proteins of BCV, MHV and the more highly diverged IBV and TGEV.

It is interesting to note that the N terminal O-linkage site for M of MHV A59 is Ser-Ser-Thr-Thr and is identical to the O-glycosylated amino terminus of M-type glycophorin A (Armstrong et al., 1984). The N-terminal sequence on BCV M protein is Ser-Ser-Val-Thr-Thr and is therefore not identical. Perhaps this is not the O-glycosylation site for BCV M protein. There are, however, no other internal sequences of Ser-Ser-Thr-Thr. There is one sequence of Asn-Phe-Ser, which is a potential N-linked glycosylation site. No N-linked glycosylation for this protein has been observed, however (Hogue, personal communication).

REFERENCES

REFERENCES

- Aaronson, R. P., J. F. Young, and P. Palese. 1982. Oligonucleotide mapping: evaluation of its sensitivity by computer simulation. *Nuc. Acids Res.* 10: 237-246.
- Armstrong, J., H. Niemann, S. Smeekens, P. Rottier, and G. Warren. 1984. Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. *Nature* 308: 751-752.
- Armstrong, J., S. Smeekens, and P. Rottier. 1983. Sequence of the nucleocapsid gene from murine coronavirus MHV-A59. *Nuc. Acids Res.* 11: 833-891.
- Auperin, D. D., V. Romanowski, M. Galinski, and D. H. L. Bishop. 1984. Sequence studies of Pichinde arenavirus S RNA indicate a novel coding strategy, an ambisense viral S RNA. *J. Virol.* 52: 897-904.
- 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 sequences on nascent chains. *J. Virol.* 48: 633-640.
- Bellini, W. J., G. Englund, S. Rozenblatt, H. Arnheiter, and C. D. Richardson. 1985. Measles virus P gene codes for two proteins. *J. Virol.* 53: 908-919.
- Berkner, K. L. and W. R. Folk. 1977. Polynucleotide kinase exchange reaction. *J. Biol. Chem.* 252: 3176-3184.
- Binns, M. M., M. E. G. Boursnell, D. Cavanagh, D. J. C. Pappin, and T. D. K. Brown. 1985. Cloning and sequencing of the gene encoding the spike protein of the coronavirus IBV. *J. Gen. Virol.* 66: 719-726.
- Boursnell, M. E. G., M. M. Binns, I. J. Foulds and T. D. K. Brown. 1985. Sequences of the nucleocapsid genes from two strains of avian infectious bronchitis virus. *J. Gen. Virol.* 66: 573-580.
- Boursnell, M. E. G. and T. D. K. Brown. 1984. Sequencing of coronavirus IBV genomic RNA: a 195-base open reading frame encoded by mRNA B. *Gene* 29: 87-92.
- Brayton, P. R., M. M. C. Lai, C. D. Patton, and S. A. Stohlman. 1982. Characterization of two RNA polymerase activities induced by mouse hepatitis virus. *J. Virol.* 42: 847-853.
- Brian, D. A., D. E. Dennis, and J. S. Guy. 1980. Genome of porcine transmissible gastroenteritis virus. *J. Virol.* 34: 410-415.

- Brian, D. A., B. Hogue, W. Lapps, B. Potts, and P. Kapke. 1983. Comparative Structure of Coronaviruses. In "Proceedings from the Fourth International Symposium on Neonatal Diarrhea." (Ed. S. D. Acres), University of Saskatoon, Saskatchewan, Canada.
- Brown, T. D. K., and M. E. G. Boursnell. 1984. Avian infectious bronchitis virus genomic RNA contains sequence homologies at the intergenic boundaries. *Virus Research* 1: 15-24.
- Brown, T. D. K., M. E. G. Boursnell, and M. M. Binns. 1984. A leader sequence is present on mRNA A of avian infectious bronchitis virus. *J. Gen. Virol.* 65: 1437-1442.
- Budzilowicz, C. J., S. P. Wilczynski, and 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 the viral mRNA leader sequence. *J. Virol.* 53: 834-840.
- Clewley, J., J. Gentsch, and D. H. L. Bishop. 1977. Three unique viral RNA species of snowshoe hare and La Crosse bunyaviruses. *J. Virol.* 22: 459-468.
- Collins, P. L., and G. W. Wertz. 1983. cDNA cloning and transcriptional mapping of none polyadenylated RNAs encoded by the genome of human respiratory syncytial virus. *Proc. Natl. Acad. Sci.* 80: 3208-3212.
- Crouch, C. F., H. Bielefeldt Ohman, T. C. Watts, and L. A. Babiuk. 1985. Chronic shedding of bovine enteric coronavirus antigen-antibody complexes by clinically normal cows. *J. Gen. Virol.* 66: 1489-1500.
- Dagert, M. and S. D. Ehrlich. 1979. Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene* 6: 23-28.
- Dennis, D. E., and D. A. Brian. 1981. Coronavirus cell-associated RNA-dependent RNA polymerase. In "Biochemistry and Biology of Coronaviruses". Adv. In Exp. Biol. and Med. 142: 155-170.
- Dennis, D. E., and D. A. Brian. 1982. RNA-dependent RNA polymerase activity in coronavirus-infected cells. *J. Virol.* 42: 153-164.
- DeWachter, R., and W. Fiers. 1972. Preparative two-dimensional polyacrylamide gel electrophoresis of ³²P-labeled RNA. *Anal. Biochem.* 49: 184-197.
- Dretzen, G., M. Bellard, P. Sassone-Corsi, and P. Chambon. 1981. A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. *Anal. Biochem.* 112: 295-298.

- Dugaiczky, A., H. W. Boyer, and H. M. Goodman. 1975. Ligation of EcoRI endonuclease-generated DNA fragments into linear and circular structures. *J. Mol. Biol.* 96: 171-184.
- Forsdyke, D. R. 1984. Purification of oligo dG-tailed Okayama-Berg linker DNA fragments by oligo dC-cellulose chromatography. *Anal. Biochem.* 137:143-145.
- Gerna, G., P. M. Cereda, M. G. Revello, E. Cattaneo, M. Battaglia, and M. Torsellini Gerna. 1981. Antigenic and biological relationships between human coronavirus OC43 and neonatal calf diarrhea coronavirus. *J. Gen. Virol.* 54: 91-102.
- Grunstein, M., and D. Hogness. 1975. Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci.* 72: 3961-3965.
- Gubler, U. and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* 25: 263-269.
- Guy, J. S., and D. A. Brian. 1979. Bovine coronavirus genome. *J. Virol.* 29: 293-300.
- Hajer, L., and J. Storz. 1979. Structural polypeptides of the enteropathogenic bovine coronavirus strain LY-138. *Arch. Virol.* 59: 47-57.
- Hanahan, D. and M. Meselson. 1980. Plasmid screening at high density. *Gene* 10: 63-67.
- Hanahan, D. 1983. Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* 166: 557-580.
- Heidecker, G., and J. Messing. 1983. Sequence analysis of zein cDNAs obtained by an efficient mRNA method. *Nuc. Acids Res.* 11: 4891-4906.
- Hogue, B. G., B. King, and D. A. Brian. 1984. Antigenic relationships among proteins of bovine coronavirus, human regulatory coronavirus OC43, and mouse hepatitis coronavirus A59. *J. Virol.* 51: 384-388.
- Holmes, K. V., E. W. Doller, and L. S. Sturman. 1981. Tunicamycin resistant glycosylation of a coronavirus glycoprotein: Demonstration of a novel type of viral glycoprotein. *Virology* 115: 334-344.
- Holmes, K. V., M. F. Frana, S. G. Robbins, and L. S. Sturman. 1984. Coronavirus maturation. pp. 37-52. In P. J. M. Rottier, B. A. M. van der Zeijst, W. J. M. Spaan and M. C. Horzinek (eds.). *Molecular biology and pathogenesis of coronaviruses*. Plenum Publishing Corp., New York.

House, J. A. 1978. Economic impact of rotavirus and other neonatal disease agents of animals. *J. Am. Vet. Med. Assoc.* 173:573-576.

Jacobs, L., W. J. M. Spaan, M. C. Horzinek, and B. A. M. van der Zeijst. 1981. Synthesis of subgenomic mRNAs of mouse hepatitis virus is initiated independently: Evidence from UV transcription mapping. *J. Virol.* 39: 401-406.

James, R. and R. A. Bradshaw. 1984. Strand separation of DNA fragments and their isolation from nondenaturing polyacrylamide gels. *Anal. Biochem.* 140: 456-458.

Kado, C. I., and S. T. Liu. 1981. Rapid procedure for detection and isolation of large and small plasmids. *J. Bact.* 145: 1365-1373.

Kapikian, A. Z., H. D. James, S. J. Kelly, L. M. King, A. L. Vaughn, and R. M. Chanock. 1972. Hemadsorption by coronavirus strain OC43. *Proceedings of the Society of Experimental Biology and Medicine* 139:179-186.

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

King, B., B. J. Potts, and D. A. Brian. 1985. Bovine coronavirus hemagglutinin protein. *Virus Res.* 2:53-59.

Kozak, M. 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev.* 47: 1-45.

Lai, M. M. C., P. R. Brayton, R. C. Armen, C. D. Patton, C. Pugh, and S. A. 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. M. C., R. S. Baric, P. R. Brayton, and S. A. Stohlman. 1984. Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proc. Natl. Acad. Sci. USA* 81:3626-3630.

Lai, M. M. C., C. D. Patton, R. S. Baric, and S. A. Stohlman. 1983b. Presence of leader sequences in the mRNA of mouse hepatitis virus. *J. Virol.* 46:1027-1033.

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

Laporte, J., and P. Bobulesco. 1981. Polypeptide structure of bovine enteric coronavirus: comparison between a wild strain purified from feces and a HRT 18 cell adapted strain. *In* "Biochemistry and Biology of Coronaviruses" (ed. V. ter Meulen, S. Sidell, and H. Wege). *Advances in Experimental Medicine and Biology* 142:171-179.

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

Leriche, A., D. Christophe, H. Brocas, and G. Vassart. 1983. Molecular cloning of complementary DNA: Preparation of a plasmid vector with low transformation background. Anal. Biochem. 129:249-252.

Lizardi, P. M. 1981. Binding and recovery of DNA and RNA using S & S NA-45 DEAE membranes. In Sequences by Schleicher and Scuell Application update No. 364, Keene, NH 03431

Macnaughton, M. R. 1982. Occurrence and frequency of coronavirus infections in humans as determined by enzyme-linked immunosorbent assay. Int. Immun. 38:419-423.

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

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A laboratory manual. Cold Spring Harbor Laboratory.

Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods in Enzymology 65: 499-560.

McIntosh, K. 1974. Coronaviruses: a comparative review. Curr. Top. Microbiol. Immunol. 63: 85-129.

McIntosh, K., A. Z. Kapikian, K. A. Hardison, J. W. Hortle, and R. M. Chanock. 1969. Antigenic relationships among the coronaviruses of man and between human and animal coronaviruses. J. Immunol. 102:1109-1118.

McMaster, G. K., and G. G. Carmichael. 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. USA 74: 4835-4838.

Mebus, C. A. 1978. Pathogenesis of coronavirus infection in calves. Am. J. Vet. Res. 173:631-632.

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

Mebus, C. A., E. L. Stair, M. B. Rhodes, and M. J. Twiehaus. 1973. Neonatal calf diarrhea: propagation, attenuation and characteristics of a coronavirus-like agent. Am. J. Vet. Res. 34: 145-150.

Messing, J., R. Crea, and P. H. Seeburg. 1981. A system for shotgun DNA sequencing. *Nucl. Acids Res.* 9: 309-321.

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

Niemann, H., G. Heisterberg-Moutsis, R. Geyer, H.-D. Klenk, and M. Wirth. 1984. Glycoprotein E1 of MHV-A59: Structure of the O-linked carbohydrates and construction of full length recombinant cDNA clones. pp. 201-213. In P. J. M. Rottier, B. A. M. van der Zeijst, W. J. M. Spaan, and M. C. Horzinek (eds.), *Molecular biology and pathogenesis of coronaviruses*. Plenum Publishing Corp., New York.

Oloffson, S., J. Blomberg, and E. Lycke. 1981. O-glycosidic carbohydrate-peptide linkages of herpes simplex virus glycoproteins. *Arch. Virol.* 70: 321-329.

Peacock, S. L., C. M. McIver, and J. J. Monahan. 1981. Transformation of *E. coli* using homopolymer-linked plasmid chimeras. *Biochim. Biophys. Acta* 655: 243-250.

Pedersen, N. C., J. Ward, and W. L. Mengeling. 1978. Antigenic relationship of the feline infectious peritonitis virus to coronaviruses of other species. *Arch. Virol.* 58: 45-53.

Perry, R. P., J. Latorre, D. E. Kelly, and J. R. Greenberg. 1972. On the lability of poly(A) sequences during extraction of messenger RNA from polyribosomes. *Biochim. Biophys. Acta* 262: 220-226.

Queen, C. and Korn, L. J. 1984. A comprehensive sequence analysis program for the IBM personal computer. *Nucl. Acids Res.* 12: 581-599.

Resta, S., J. P. Luby, C. R. Rosenfield, and J. D. Siegel. 1985. Isolation and propagation of a human enteric coronavirus. *Science* 229: 978-981.

Retzel, E. F., M. S. Collett, and A. J. Faras. 1980. Enzymatic synthesis of deoxyribonucleic acid by the avian retrovirus reverse transcriptase in vitro: optimum conditions required for transcription of large ribonucleic acid templates. *Biochem.* 19: 513-518.

Roseto, A., P. Bobulesco, J. Laporte, J. Escarg, D. Gaches, and J. Peries. 1982. Bovine enteric coronavirus structure as studied by a freeze-drying technique. *J. Gen. Virol.* 63: 241-245.

Rottier, P., D. Brandenburg, J. Armstrong, B. Van Der Zeijst, and 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. USA* 81: 1421-1425.

- Roychoudhury, R. and R. Wu. 1980. Terminal transferase-catalyzed addition of nucleotides to the 3' termini of DNA. *Methods in Enzymology* 65: 43-62.
- Russell, R. G., D. A. Brian, A. Lenhard, L. N. D. Potgieter, D. Gillespie, and N. K. Clapp. 1984. Coronavirus-like particles and *Campylobacter* in marmosets with diarrhea and colitis. *Digestive Diseases and Sciences*. In press.
- Shapiro, D. J. 1981. Quantitative ethanol precipitation of nanogram quantities of DNA and RNA. *Anal. Biochem.* 110: 229-231.
- Siddell, S. 1983. Coronavirus JHM: coding assignments of subgenomic mRNAs. *J. Gen. Virol.* 64:113-125.
- Siddell, S., H. Wege, and V. ter Meulen. 1982. The structure and replication of coronaviruses. *Curr. Top. Microbiol. Immunol.* 99: 131-163.
- Skinner, M. A., D. Ebner, and S. G. Siddell. 1985. Coronavirus MHV-JHM mRNA 5 has a sequence arrangement which potentially allows translation of a second, downstream open reading frame. *J. Gen. Virol.* 66:581-592.
- Skinner, M. A., and S. G. Siddell. 1983. Coronavirus JHM: nucleotide sequence of the mRNA that encodes nucleocapsid protein. *Nucl. Acids Res.* 11: 5045-5054.
- Skinner, M. A. and S. G. Siddell. 1985. Coding sequence of coronavirus MHV-JHM mRNA 4. *J. Gen. Virol.* 66: 593-596.
- Smith, H. O. 1980. Recovery of DNA from gels. pp. 371-380. In Lawrence Grossman and Kilvie Moldave (eds.). *Methods in Enzymology*, Vol. 65, Nucleic Acids, Part I, Academic Press, NY.
- Smith, H. O., and M. L. Birnstiel. 1976. A simple method for DNA restriction site mapping. *Nucl. Acids Res.* 3: 2387-2398.
- Spaan, W. J. M., H. Delius, M. A. Skinner, J. Armstrong, P. J. W. Rottier, S. Sreekens, S. G. Siddell, and B. A. M. van der Zeijst. 1984. Transcription strategy of coronaviruses: fusion of non-contiguous sequences during mRNA synthesis. pp. 173-186. In P. J. M. Rottier, B. A. M. van der Zeijst, W. J. M. Spaan, and M. C. Horzinek (eds.). *Molecular biology and pathogenesis of coronaviruses*. Plenum Publishing Corp., New York.
- Stern, D. and S. I. T. Kennedy. 1980a. Coronavirus multiplication strategy. I. Identification and characterization of virus-specified RNA. *J. Virol.* 34: 665-674.

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

Storz, J., and R. Rott. 1981. Reactivity of antibodies in human serum with an enteropathogenic bovine coronavirus. *Med. Microbiol. Immunol.* 169: 169-178.

Sturman, J. S., and K. V. Holmes. 1977. Characterization of a coronavirus. II. Glycoproteins of the viral envelope: tryptic peptide analysis. *Virology* 77: 650-660.

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

Tompkins, W. A. F., A. M. Watrach, J. D. Schmale, R. M. Schulze, and J. A. Harris. 1974. Cultural and antigenic properties of newly established cell strains derived from adenocarcinomas of the human colon and rectum. *J. Natl. Canc. Inst.* 52: 1101-1106.

Tu, C.-P. D. and S. N. Cohen. 1980. 3' end labeling of DNA with ³²P cordycepin-5'-triphosphate. *Gene* 10: 177-183.

Tyrrell, D. A. J., J. D. Almeida, C. H. Cunningham, W. R. Dowdle, M. S. Hofstad, K. McIntosh, M. Tajima, L. Y. A. Zakstelskaya, B. C. Easterday, A. Kapikian, and R. W. Bingham. 1975. Coronaviridae. *Intervirology* 5: 76-82.

Tyrrell, D. A. J., D. J. Alexander, J. A. Almeida, C. H. Cunningham, B. C. Easterday, D. J. Garwes, J. C. Hierholzer, A. Kapikian, M. R. Macnaughton, and K. McIntosh. 1978. Coronaviridae: second report. *Intervirology* 10: 321-328.

Vieira, J. and J. Messing. 1982. The pUC plasmids on M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19: 269-276.

Wege, H., S. Siddell, and V ter Meulen. 1982. The biology and pathogenesis of coronaviruses. *Curr. Top. Microbiol. Immunol.* 99: 165-200.

Wenske, E. A. and R. J. Courtney. 1983. Glycosylation of herpes simplex virus type 1 gC in the presence of tunicamycin. *J. Virol.* 46: 297-301.

Winberg, G. and M.-L. Hammar-skjold. 1980. Isolation of DNA from agarose gels using DEAE-paper. Application to restriction site mapping of adenovirus type 16 DNA. *Nucl. Acids Res.* 8: 253-264.

APPENDIX A

DETAILED METHODS USED IN THE MOLECULAR

CLONING OF THE BOVINE CORONAVIRUS

GENOMIC RNA

VIRUS PRODUCTION IN TISSUE CULTURE CELLS

Stocks of plaque purified Mebus strain of BCV were prepared as previously described (Lapps and Brian, 1985). The virus was grown in human rectal tumor cells designated HRT-18 (Tomkins et al., 1974). Virus was harvested from 48-72 hours post infection. Cytopathic effect of BCV is very minimal and can be described as clear cytoplasmic vesicles in infected cells at 48 hours post infection. The infected cell supernatant fluid (containing the extracellular virus) was removed and monolayers were rinsed with a minimal amount of TMEN [10 mM Tris maleate (pH 6.0), 100 mM NaCl, 1 mM EDTA]. This ensures maximal recovery of extracellular virus which may be loosely bound to the outside of infected cells.

The presence of virus was monitored by hemagglutination, radiolabeling of RNA with ^3H -uridine or of RNA, lipids and protein with ^{32}P orthophosphate. Mouse erythrocytes were used for hemagglutination. The ^3H -uridine or ^{32}P was used to monitor purification of viral RNA. Typically, one flask out of every ten was radiolabeled to limit the amount of damage done by radiolabel on the cells. When ^3H -uridine was used, 100 uCi was added to a 150 cm² flask which contained 20 ml Dulbecco's minimal essential medium (DMEM), 5% fetal calf serum. When ^{32}P orthophosphate was used, cells were refed at 1.5 hours post infection with medium containing 10% of normal phosphate concentration and 2% fetal calf serum. Forty uCi ^{32}P orthophosphate was added per ml.

PURIFICATION OF THE BOVINE CORONAVIRUS BY ISOPYCNIC SEDIMENTATION

Supernatant fluids were clarified (to remove cellular debris) by centrifugation for 10 minutes at 10,000 rpm. Clarification was done in screw-capped Oak Ridge tubes using the JA20 rotor and Beckman J21C centrifuge. Clarified supernatant fluids were kept frozen at -80°C until use. Virus was concentrated by layering supernatant fluids onto a 2 ml cushion of 60% sucrose made in TMEN at 24,000 rpm in a AH-627 rotor for 2 hours. At the end of each spin, 32 ml of supernatant fluid could be removed and virus from new fluid could be sequentially concentrated in this manner. After concentration, virus mixed with the 2 ml of sucrose was diluted with TMEN to a sucrose concentration of less than 20%. The virus was layered onto an 18 ml continuous gradient of 20 to 60% wt/wt sucrose gradient and centrifuged at 24,000 rpm for 4.5 hours at 4°C . The sucrose gradient was fractionated and ^3H -uridine incorporation was determined by scintillation spectrophotometry of TCA precipitated 5 μl samples or ^{32}P -label was determined by Cherenkov radiation. The buoyant density of each fraction was determined from the refractive index of 5 μl aliquots on the Abbe refractometer (American Optical Corp.). Isopycnic purification of ^{32}P -labeled virus is shown in Figure A.1. Previous experiments have demonstrated that the peak of radioactive incorporation also corresponds to peak hemagglutination (data not shown). Peak fractions were pooled, diluted with TMEM to have final sucrose concentrations less than 20%, and virus was pelleted at 49,000 rpm on a SW50.1 rotor.

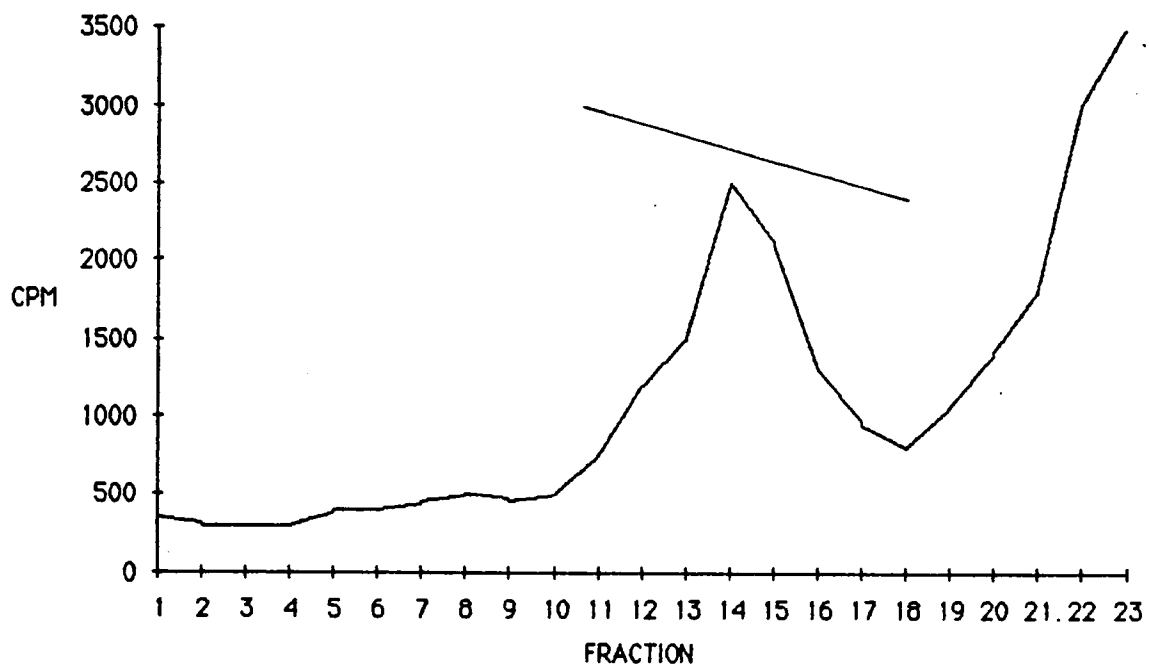


Figure A.1. Purification of bovine coronavirus by isopycnic sedimentation. ^{32}P -labeled virus was sedimented in an 18 ml 20-60% linear sucrose gradient. The virus peak banded between the buoyant densities of 1.17-1.20.

The virus pellet from up to ten 150 cm² flasks was resuspended in 0.5 ml of TNE [10 mM Tris HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA] containing 2% sodium dodecyl sulfate and 0.5 mg/ml of proteinase K. The mixture was incubated at 37°C for 10 minutes followed by two consecutive extractions with phenol, chloroform, isoamylalcohol (50:48:2). The rationale for phenol chloroform extraction comes from the work of Perry et al., 1972, describing the lability of poly(A) sequences during extraction in phenol in the absence of chloroform. The RNA was ethanol precipitated at -80°C after adding 0.1 volume 3M sodium acetate and three volumes of absolute ethanol.

PURIFICATION OF GENOMIC RNA BY RATE-ZONAL SEDIMENTATION

RNA from purified virus sometimes contains an 18S subgenomic species of RNA (Lapps and Brian, 1985). This 18S RNA could be a packaged mRNA species or a defective RNA genome. To increase chances of cloning only full-length genomic RNA, high molecular weight virion RNA was size selected on a continuous 5 ml gradient 10-30% wt/wt sucrose made in TNE, 1% SDS. Virion RNA was dissolved in 100 ul of water and layered onto a preformed gradient. The gradient was spun for 1.5 hours at 49,000 rpm at 25°C. ³H-uridine-labeled Maden-Darby bovine kidney cell ribosomal RNA (18 and 28S rRNA) served as a size marker and was sedimented in a parallel gradient. Gradients were fractionated by collecting 10 drops per fraction dripped from the bottom of the tube. Distribution of radioactivity was determined by Cherenkov counting for ³²P, or by scintillation counting of 5 ul aliquots in aqueous fluor for ³H. Results of a typical rate zonal

sedimentation are shown in Figure A.2. Virion RNA that was larger than the 28S marker RNA was ethanol precipitated at -80°C after adding 2.5 volumes of absolute ethanol and 0.1 volume of 3M sodium acetate. RNA was dissolved and precipitated in ethanol at least 2 more times before subsequent quantitation and use. This RNA served as the starting material for all cDNA cloning procedures, and for preparing virus-specific cDNA probes for colony detection of virus-specific inserts. RNA was quantitated by spotting small amounts onto 1% agar containing 0.5 $\mu\text{g/ml}$ of ethidium bromide (Maniatis et al., 1982). The approximate amount of RNA was determined by comparing virion RNA fluorescence to dilutions of globin mRNA of known concentrations.

Enrichment of the genomic RNA is demonstrated in Figure A.3. Lane 1 shows the presence of the 18S (or $.68 \times 10^6$ molecular weight) subgenomic RNA in a preparation of purified virus that was radiolabeled with ^{32}P orthophosphate. RNA was separated on a 1% agarose urea gel that contained iodoacetate as an RNase inhibitor. Lane 2 shows the high molecular weight virion RNA that was enriched by sedimentation.

FIRST STRAND DNA SYNTHESIS

BCV genomic RNA was estimated to be 20 kilobases in length. It was therefore desirable to copy RNA into the longest possible molecules of DNA before cloning. Reaction conditions for first strand synthesis were the same as described by Retzel (Retzel et al., 1980), in which optimal conditions for transcription of large RNA templates into DNA were given. First strand DNA reactions were carried out in a sterile

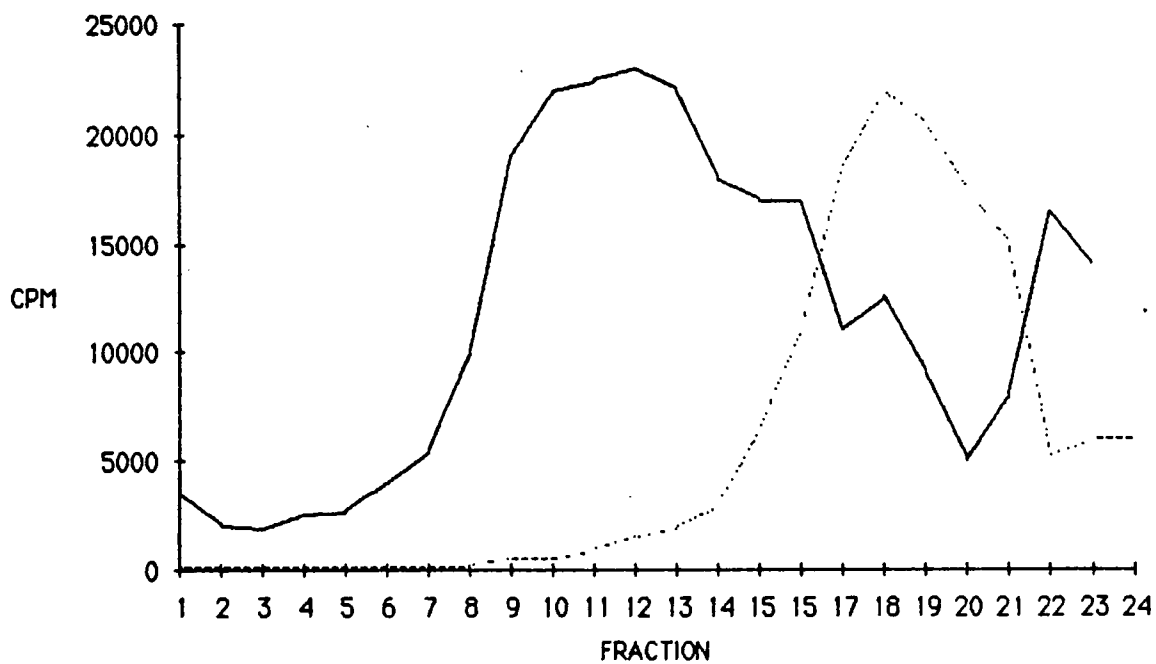


Figure A.2. Rate-zonal sedimentation of bovine coronaviral RNA for purification of full-length genomic RNA. RNA was prepared and sedimented as described in the text. Sedimentation is from right to left.



Figure A.3. Agarose-urea gel of purified genomic RNA.

gas-siliconized, 1.5 ml Eppendorf tube. The typical 50 ul reaction mixture consisted of: 8 mM MgCl₂, 148 mM KCl, 50 mM Tris·Cl (pH 8.1), 1 mM dithiothreitol, 2 mM each nucleotide triphosphate, 15 units RNasin, 25:1 molar ratio oligo(dT) to poly(A) RNA, 20 units of reverse transcriptase (Seikagaku) and trace alpha ³²P(d)CTP to monitor the polymerization.

The first control experiments were completed with rabbit B hemoglobin mRNA (P-L Biochemicals) as the test molecule. These reaction conditions give full length cDNA copies when oligo(dT) was used as the primer. The amount of total RNA copied was estimated from the amount of radioactive nucleotide incorporated. Typically 10 to 30% of the RNA was copied into cDNA. The results for first strand synthesis of coronavirus RNA were different from that obtained with globin mRNA for a couple of reasons. First, an equivalent mass of coronavirus RNA has less than one twentieth the number of poly A tails on which the oligo(dT) can prime first strand synthesis. Also, the size of first strand cDNA made is heterogeneous, ranging from 200 bases to about 8,000 bases (data not shown). This could be due to breakdown of the genomic RNA or early termination of the reverse transcriptase.

The first strand synthesis reaction was incubated at 37°C for one hour in all reverse transcriptase reactions. The reaction was then stopped by adding 2.5 ul of 0.5 M EDTA. The mixture was deproteinized by extraction with an equal volume of water-saturated PCI (phenol, chloroform, isoamyl alcohol [50:48:2]). The reaction mixture was vortexed for 30 seconds and the emulsion was separated by three minutes

centrifugation in the microfuge (Fisher) at 12,000 x g. The aqueous (upper) phase of the mixture was removed and the product was ethanol precipitated in a 1.5 ml gas-siliconized Eppendorf tube using one half volume of 7.5 M ammonium acetate and 3 volumes of absolute ethanol. The product of first-strand cDNA synthesis (RNA-DNA duplexes), was pelleted after first freezing the ethanol and salt-water mixture on dry ice snow for ten minutes. The tube was then centrifuged for ten minutes on the microfuge in the cold room. After removing the ethanol from the precipitate, the pellet was briefly dried (about 30 seconds) on the Savant Speed Vac Concentrator. This step was used to remove residual ethanol from the pellet which may inhibit enzyme reactions. The pellet was not dried to completion because anhydrous DNA is, on occasion, very difficult to get back into solution.

SECOND STRAND DNA SYNTHESIS

Second strand synthesis of cDNA was accomplished using the conditions described by Gubler and Hoffman (1983). The DNA-RNA pellet from first strand synthesis was dissolved in a 100 ul reaction mixture containing: 20 mM Tris·HCl (pH 7.5), 5 mM MgCl₂, 10 mM (NH₄)₂SO₄, 10 mM KCl, and 0.15 mM B-NAD. The reaction mixture was incubated at 37°C for seven minutes to ensure that the DNA-RNA pellet was in solution. The reaction mixture was then cooled to 12°C in a styrofoam water bath and 8.5 units/ml RNase H and 230 units/ml DNA polymerase I were added. The reaction was incubated at 12°C for 60 min and then at 22°C for another hour. The reaction was terminated by adding 20 ul of 0.5 M EDTA and 3 ul 10% SDS. The products were then

deproteinized by extraction with an equal volume of PCI, and the aqueous phase was ethanol precipitated twice using one half volume of 7.5 M ammonium acetate as the salt. Double ethanol precipitation in this manner removes most all of the unreacted nucleotide triphosphates from the product. The total quantity of product was determined from the remaining radioactivity.

HOMOPOLYMER TAILING OF DOUBLE-STRANDED cDNA

Homopolymer tailing of the 3' end of DNA molecules was first calibrated by tailing known quantities of 3' ends. 3' ends used were generated by cutting either plasmid pUC9 or pBR322 with the restriction enzyme PstI. Optimal tail length for transformation of (dG)-(dC) homopolymer-linked chimeras is 15 to 20 bases (Peacock et al., 1981).

Reaction conditions for tailing were based on personal communication from L. Lorenz. This procedure was modified from that described (Roychoudhury and Wu, 1980). DNA to be homopolymer-tailed was ethanol precipitated using ammonium acetate as the salt. Sodium acetate is avoided because sodium ions inhibit terminal deoxynucleotidyl transferase (TdT). The 200 μ l reaction mixture contains the following: 140 mM potassium cacodylate, 30 mM Tris·HCl (pH 7.6), 0.1 mM dithiothreitol, 150 μ M dCTP (or 30,000 pmoles of dCTP), 1 mM CoCl_2 , 2 pmole of 3' ends of DNA per 200 μ l, and 8 units of TdT per pmole of 3' end. Radioactive tracer as alpha ^{32}P dCTP was added as necessary. Reaction conditions used were for 2 pmoles of 3' DNA ends or less. A vast excess of dNTP is used so as not to be rate limiting in the reaction. This is an especially important

consideration for tailing double-stranded cDNA whose exact concentration is unknown.

The incorporation curve of dCTP into 1 pmole of 3' ends of PstI-cut pBR322 is shown in Figure A.4. Radioactive alpha ^{32}P dCTP was used to monitor incorporation. Based on radioactivity incorporated, this curve demonstrates that in one minute an average of twenty bases were added to each 3' end.

Curves for the incorporation of dCMP were reproducible in several experiments. Therefore, the amount of time chosen to tail double-strand cDNA was one minute. The reaction for tailing double-stranded cDNA was performed without radioactive tracer for two reasons: First, the molar quantity of ends in the reaction mix was unknown. Therefore, any calculation of incorporation based on radioactivity would be grossly error-prone. Secondly, for cDNA cloning, only the first-strand is labeled and used as an indicator of the amount of cDNA product formed, the efficiency of extraction, and the efficiency of ethanol precipitation. Addition of a radiolabel during second strand synthesis would, therefore, be counterproductive.

The tailing reaction for double-stranded cDNA was stopped by adding 10 μl 0.5 M EDTA and 2 μl 10% SDS. The reaction was then deproteinized by extracting with PCI. DNA in the aqueous phase was ethanol precipitated twice. The first precipitation was with sodium acetate because the aqueous volume was large. The second ethanol precipitation employed ammonium acetate to facilitate removal of unreacted triphosphates which are inefficiently precipitated when using this salt.

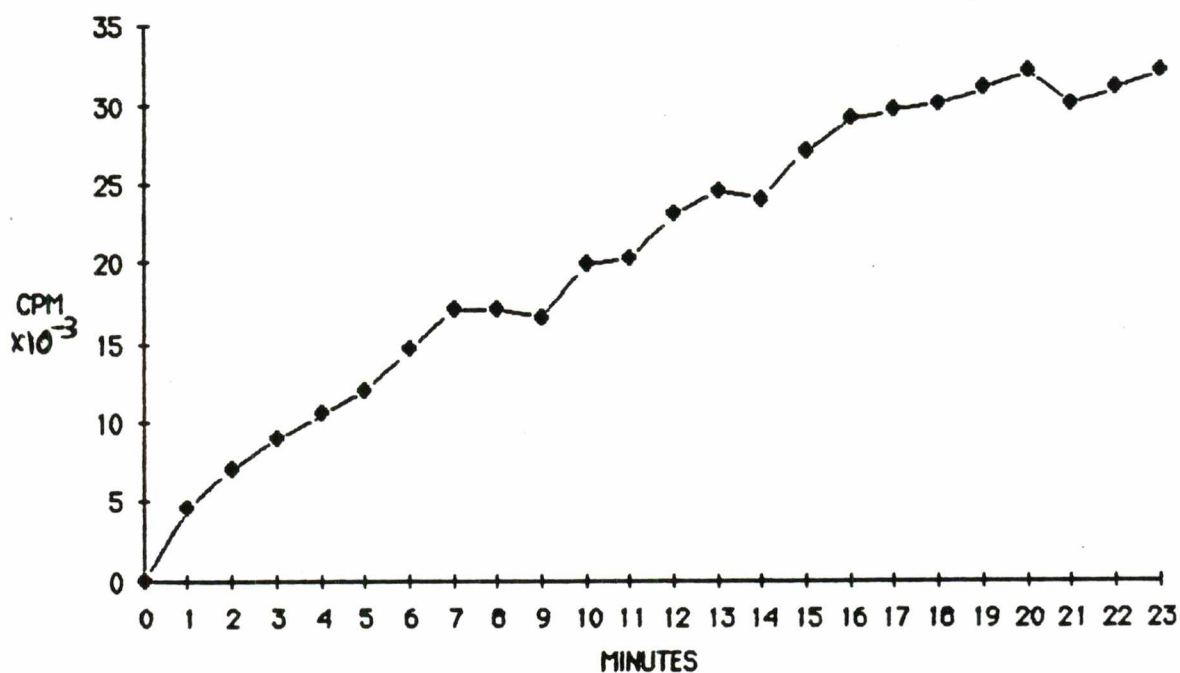


Figure A.4. Time course for adding homopolymer tails of dCMP to the 3' ends of Pst I cut pBR322. At each of the time points indicated, a 4 μ l sample was removed from the reaction mixture and added to 1 μ l 0.5 M EDTA. The samples were then spotted onto DE81 filter discs and washed (Maniatis et al., 1982) to remove the unincorporated nucleotide triphosphates. Radioactivity was quantitated by scintillation spectrophotometry.

PREPARATION OF VECTOR DNA

The plasmid vector pUC9 (Vieira and Messing, 1982) which contains multiple cloning sites was used to clone the dC-tailed cDNA by homopolymer chimera formation. Homopolymer dG-tailed pUC9 (tailed at the PstI site) was either prepared in this laboratory or purchased from P-L Biochemicals when available.

dG-tailed pUC9 was prepared in the laboratory and enriched to have low transformation background. Plasmid pUC9 was grown, extracted, and purified twice over a CsCl gradient to have minimal contaminating RNA. Plasmid was then cut to completion at the PstI site and homopolymer tailed using dGTP. Pilot experiments to determine the incorporation curve showed that (unlike the results obtained for dC-tailing) the incorporation of 15 bases at the 3' end of the DNA required 15 minutes at 10°C. The lower temperature was used to slow down the incorporation of bases and make the experiment technically feasible.

The cloning vector prepared in this manner contains very high background of wild type plasmid (Leriche et al., 1983). This may occur from uncut plasmids or cut plasmids which do not get homopolymer tailed. Thirty ug of vector DNA that was dG homopolymer tailed had a background of 287 transformants per nanogram. After enrichment of the cut and tailed vectors over an oligo(dC) column (Forsdyke, 1984), the background transformation dropped to the very acceptable level of three transformants per nanogram of prepared vector.

ANNEALING AND TRANSFORMATION

Optimal annealing conditions for homopolymer-tailed molecules to be used for transformation was studied by Peacock et al. (1981). Optimal tail length, temperature, and DNA concentrations were determined for an insert molecule of about one kilobase in length. The double-stranded cDNA from coronavirus genomic RNA is heterogeneous in size, but was assumed to have an average length of one kilobase. This size was used when estimating the total number coronavirus-dscDNA molecules made. Based on this assumption, an estimate of the J to I ratio can be made (Dugaiczky et al., 1975). The total DNA concentration is then chosen such that a compromise is made between having the maximum number of attainable clones (from the highest dilution) and having the minimum amount of experimental manipulation (from the lowest dilution).

Because the cDNA was heterogeneous in size, a titration was made of the optimal cDNA:vector weight ratios in the annealing reactions to obtain the maximum number of transformants per input cDNA (Gubler and Hoffman, 1983). The results from the titration of different amounts of cDNA into the annealing reaction is shown in Table A.1. For this experiment, the optimal tail length (15 bases), temperature, time, and total DNA concentrations were kept as the constants used by Peacock (Peacock et al., 1981; Gubler and Hoffman, 1983). The annealing buffer was also the same as described by Peacock et al. (1981).

Table A.1 shows that the optimal amount of cDNA annealed under the stated conditions was 0.35 ng and 7.0 ng. The optimal nature of these ratios was confirmed in later experiments (data not shown).

Table A.1. Optimal annealing conditions for transformation by homopolymer-tailed ds cDNA and vector.^a

C-tailed ds cDNA added per 30 ul reaction	Number of white colonies per 10 ul reaction mix	Efficiency (number of white colonies per ng ds cDNA)
0.07	0	0
0.175	0	0
0.35	4	34.3
0.7	8	34.3
1.05	6	17.1
1.4	1	2.1
2.1	6	8.6
3.5	8	6.9
7.0	11	4.7

^aReaction conditions were 7 ng of dG-tailed pUC9 vector in a volume of 30 ul, incubated for 2 hours at 58°C with increasing amounts of dC tailed ds cDNA. 10 ul of each annealing reaction was used for each transformation mix. The average number of blue colonies per ng of dG-tailed vector was 25.7. The transformation efficiency for this preparation of competent cells was 1.9×10^7 transformants per ug supercoiled pUC9. When both vector and cDNA, or just vector, were left out of the transformation, no colonies were observed.

Annealing reactions were scaled-up to larger volumes to anneal the rest of the ds cDNA that was made. After annealing, DNA in several reactions was concentrated by ethanol precipitation in Eppendorf tubes using the ultracentrifuge (Shapiro, 1981). The precedent for ethanol concentration of annealed chimeric molecules prior to transformation comes from the paper by Norgard et al. (Norgard et al., 1980). Efficiency of transformation, however, was lower when ethanol-concentrated DNA was used. It is unlikely that this is due to DNA loss during ethanol precipitation since previous experience with the technique by Shapiro using radiolabeled DNA showed no losses (data not shown). The loss in efficiency was best explained by having an excessively high DNA to cell ratio for transformation (Hanahan, 1983). The remainder of ds cDNA was therefore annealed and used without ethanol concentration.

Cells made competent for DNA uptake were prepared exactly by the technique of Hanahan (1983). This technique routinely yielded 1×10^7 to an occasional 10^8 transformants per μg of supercoiled plasmid. By comparison, the efficiency of transformation using the method of Dagert and Ehrlich (1979) was usually around 1×10^6 transformants per μg .

The pUC9 vector used for our studies (Vieira and Messing, 1982), contains the lac Z gene along with the E. coli lac operon promoter and operator. When turned on by galactose or the gratuitous inducer, IPTG (Isopropyl B-D-Thiogalactopyranoside), this gene produces the functional B-galactosidase alpha peptide. The alpha peptide is used to complement the defective gene in the E. coli host (Messing and Vieira,

host used was JM103 (Messing et al., 1981) that has the following genotype: lacpro, thi, str A, sup E, end A, sbc B15, hsd R4, F' tra D36, pro AB, lac I^q Z M15.

Figure A.5 depicts the important features of the cloning vector pUC9. The multiple cloning site of pUC9 is in the opposite orientation to that in pUC8. The multiple cloning site is also located in the coding region of the lac Z gene. If the pUC9 plasmid is used as an expression vector, one may obtain a fusion protein from the inserted coding sequence and the lac Z sequence. Homopolymer tailed (dC) cDNAs made in this study were inserted at the PstI site. The dG homopolymer tailing of PstI cut vector allows for the in vivo repair of the PstI recognition site at both ends of the inserted cDNA.

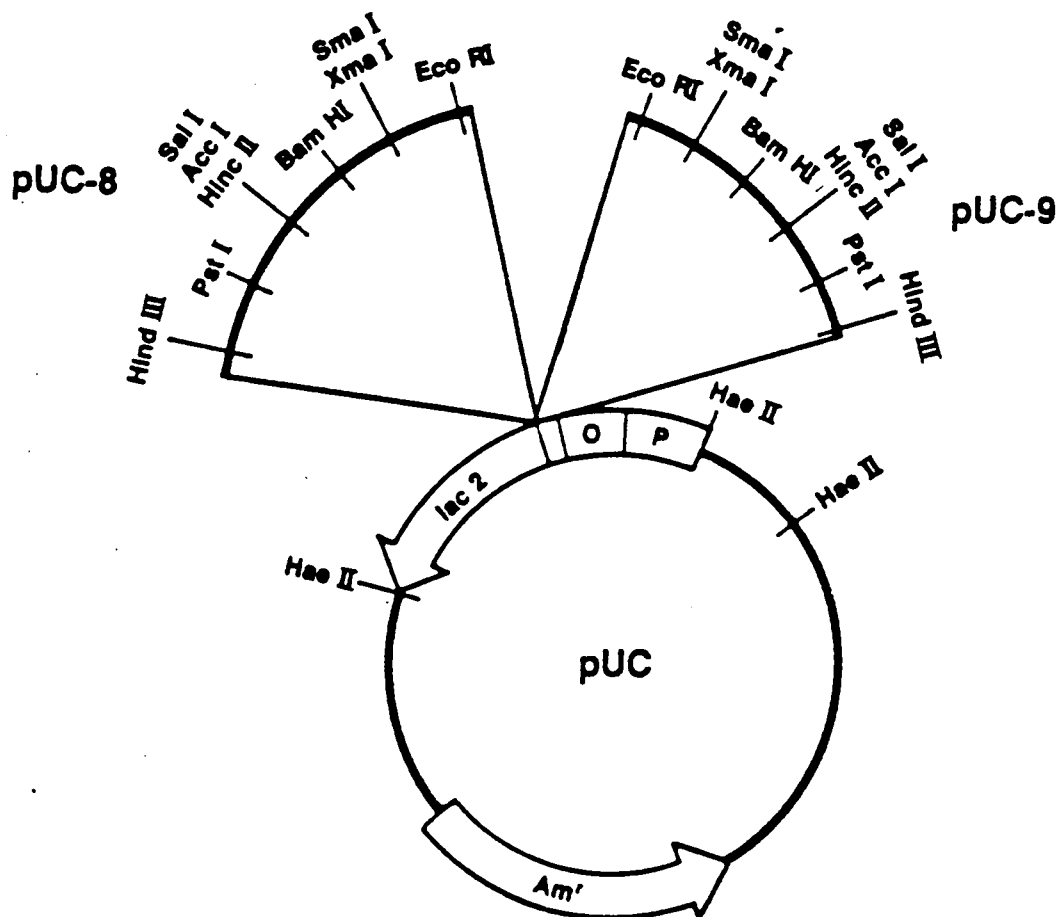


Figure A.5. Plasmid cloning vectors pUC8 and pUC9 showing opposite orientations of the multiple cloning regions. Many of the restriction endonuclease sites in the multiple cloning region were used to restriction map, subclone, and sequence BCV insert DNA.

APPENDIX B

SIZE AND SEQUENCE CHARACTERIZATION

OF BOVINE CORONAVIRUS

SPECIFIC CLONES

INITIAL IDENTIFICATION OF CLONES

After cells were transformed, they were plated on YT plates that contained 100 ug/ml ampicillin, 1 mM IPTG, and 0.004% w/v X gal. After overnight incubation on selective indicator plates, both blue and white colonies were obtained (Messing and Vieira, 1982; Messing et al., 1981). Blue colonies represent bacteria that contained the pUC9 plasmid encodes a functional ampicillin resistance gene and a functional B-galactosidase alpha-peptide gene. White colonies indicate bacteria that obtained the pUC9 plasmid with a non-functional B-galactosidase alpha-peptide. By design, this is usually the result of insertional inactivation of the lac Z gene within the multiple cloning site. An "in-frame" insertion into the multiple cloning site may give rise to a functional insert-alpha peptide fusion product, but this usually leads to weaker transcomplementation (Vieira and Messing, 1982) and therefore a less intense or slower developing blue indicator reaction. For this reason both white colonies and blue ones with atypical histochemical reactions were picked for further characterization.

For cloning BCV RNA, 3 ug of genomic RNA was transcribed into 70 ng of double-stranded cDNA. This yielded 670 white colonies that were picked for characterization by hybridization with random primed cDNA and poly(dT) probes.

The bacterial colony hybridization technique used was a hybrid of the techniques of Hanahan and Meselson (Hanahan and Meselson, 1980),

Grunstein and Hogness (1975), Maniatis et al. (1982), and personal communication from B. Lasker.

Colonies presumed to contain recombinant plasmids were picked using sterile toothpicks and inoculated in a grid pattern on sterile HATF nitrocellulose filters (Millipore Triton-Free) that were positioned on selective indicator plates. Plates were incubated to establish colonies of 3-4 mm diameter. Colonies were replica-plated onto nitrocellulose filters by a technique similar to that used by Hanahan and Meselson (1980). The master nitrocellulose was removed from the agarose plate and placed colony side up on a piece of sterile Whatman 3 MM paper. A sterile nitrocellulose filter that was prewetted by temporary contact with an agar plate, was placed on the master. Another sterile Whatman 3 MM paper was placed on top of the nitrocellulose sandwich and the stack was firmly pressed with a rubber blotting roller. The top 3 MM paper was then discarded and the nitrocellulose filters were keyed with three asymmetrically placed sterile ink spots on the edge. Filters were then separated and returned to the selective culture plates for incubation to establish 4 mm colonies. Some spotting pattern distortions are helpful for facilitating location of positive colonies after probing. Positive and negative control colonies were used as keys by inoculating X shapes or other recognizable obscene forms. This replica plating technique allows for high pattern reproducibility with very little colony spread or smearing.

By first growing the bacteria on selective agar plates supplemented with 5% glycerol, the pressed nitrocellulose filter

sandwiches can be stored in between several layers of Whatman 3 MM paper, sealed in a food pouch, with one damp paper for humidity and frozen at -80°C (Hanahan and Meselson, 1980).

The bacterial colonies were lysed and nitrocellulose filters were prepared for colony hybridization studies using the technique described in Maniatis et al. (1982). Nitrocellulose filters were placed, colony side up, on a piece of Whatman 3 MM filter paper saturated with 10% SDS for three minutes. Excess drops of 10% SDS solution were carefully scraped off the bottom of the nitrocellulose filter using the edge of the baking dish. The filter was then placed on filter paper saturated with denaturation solution, 0.5 M NaOH and 1.5 M NaCl, for 5 minutes. The colony blot with lysed colonies was then placed, colony side up, in a 60 mm sintered glass Buchner funnel which is fixed over a vacuum flask. Vacuum is applied to the filter for 30 sec drawing out excess liquid and limiting movement of the lysed colony debris. The colony blot filter is then neutralized on 3 MM paper saturated with 1.5 M NaCl and 0.5 M Tris pH 8.0 for 5 minutes. Application of vacuum on the Buchner funnel was repeated before the final treatment on 2X SSPE saturated 3 MM paper. The nitrocellulose colony blot was then transferred colony side up to a dry sheet of 3 MM paper and dried for 30 minutes. Colony blots were then sandwiched between two sheets of 3 MM paper and dried in an 80°C vacuum oven for two hours. Dried and baked blots were stored between 3 MM paper and wrapped in aluminum foil and sealed in plastic bags to keep out moisture until used for probing.

Hybridization of colony blots was done as described by Maniatis et al. (1982). Prehybridization and hybridization solutions did not contain formamide and the hybridization temperature was 68°C.

The filter was first wetted with 6X SSC by flotation and then was submerged for five minutes. The blot was washed in 50 mM Tris·Cl (pH 8.0), 1 M NaCl, 1 mM EDTA and 0.1% SDS at 37°C, with rocking, for 1-2 hours to remove loose bacterial debris. After washing, the filter was prehybridized for at least 2 hours in a volume of 0.2 ml/cm², in prehybridization fluid containing 6X SSC, 0.5% SDS, 5X Denhardt's solution and 100 ug/ml denatured salmon sperm DNA.

Prehybridization was done in a heat sealed plastic pouch. Following prehybridization, the solution was removed by pipette leaving 50 ul/cm² in the bag. Denatured radiolabeled probe, along with enough EDTA to make a final concentration of 10 mM, was added to the bag and sealed. Duration of hybridization was at least 24 hours at 68°C with rocking. The blot was washed initially in 2X SSC and 0.5% SDS for five minutes, and subsequent washes were carried out using various other dilutions of SSC depending on the stringency of washing required (Maniatis et al., 1982).

IDENTIFICATION OF BOVINE CORONAVIRUS-SPECIFIC CLONES

To identify colonies that contain bovine coronavirus specific inserts, random primed cDNA of BCV genomic RNA was used as the probe. Random primers were made using the technique described by Maniatis (1982). Starting with 1 gram of calf thymus DNA, 0.1 gram of oligodeoxynucleotide primer was obtained. The optimum mass ratio of

62.5:1 primer to template was then determined empirically. Reaction conditions for probe synthesis were the same as for first strand synthesis, except that the cold dNTP concentrations were 2.5 uM and alpha ³²P dCTP (ICN >4000 Ci/mM) concentration was 0.6 uM. The reaction was typically incubated 45 minutes at 37°C and then continued for 15 min after adding all four cold dNTPs to make a final concentration of 0.2 mM. The reaction was stopped by adding 3 ul 0.5 M EDTA and 2 ul 10% SDS, and the nucleic acids were extracted with phenol-chloroform-isoamyl alcohol. The aqueous phase from the extraction was layered onto a Sephadex G50 (or G80) (Sigma) column having a 3 ml bed volume. The column was prepared in a siliconized, baked, 5 ml glass pipette and had a siliconized glass wool plug. Prior to sample loading, the column was washed with 0.1 M NaCl, 20 mM Tris·Cl (pH 8.0), 2 mM EDTA, 0.5% SDS. SDS was used to eliminate loss of material on column by sticking.

DNA, however, was eluted using the same buffer containing no SDS. Three drop fractions were collected and the radioactive product was pooled for use as a probe.

To identify inserts sequences having a poly(A) tract, and hence representing the 3' end of the genome, a radioactive poly(dT) probe was used. Approximately 2.4 ug of oligo(dT)·poly(rA) (P-L Biochemicals) was used in a reverse transcription reaction that contained 10 uM cold dTTP and about 0.6 uM alpha ³²P dTTP. Probe synthesis was chased for 15 min. after making the reaction 8 uM with cold dTTP. Nucleic acid extraction and Sephadex G50 chromatography were the same as that described for randomly primed probe. The probe was then heat-denatured

or alkali treated to remove the RNA. The radioactive poly(dT) probe was used to probe Northern blots of plasmids extracted from several different clones.

USE OF RADIOACTIVE PROBES

The random-primed cDNA probes were denatured by boiling and added to the hybridization solutions. Colony blots were typically hybridized for 24 hours (Maniatis, 1982). Poly(dT) probe, however, was hybridized at 37°C for 24 hours and then at 20°C for another 24 hours before washing with 2X SSC at 20°C. The low temperature of hybridization and very low stringency for washing was necessary due to the low temperature of melting poly(dA)-poly(dT) hybrids. Figure B.1 shows the results of a colony blot that was probed with bovine coronavirus random-primed cDNA. Triangular radioactive tape strips on the filter serve to "key" the autoradiograph.

ISOLATION AND SIZE SCREENING OF PLASMIDS WITH BCV-SPECIFIC INSERTS

Colonies identified as have inserts by hybridization with random-primed BCV cDNA were lysed and analyzed directly by the method of Kado and Liu (1981) or were first grown in 5 ml overnight cultures, pelleted, and analyzed by the same method to measure the size of the supercoiled plasmid.

The use of an overnight culture is the preferred method for two reasons: (1) If bacterial growth was poor, consecutive spins can be done to obtain a pellet of reasonable size for extraction. (2) The bacteria can be kept for long term storage by mixing 0.85 ml of broth

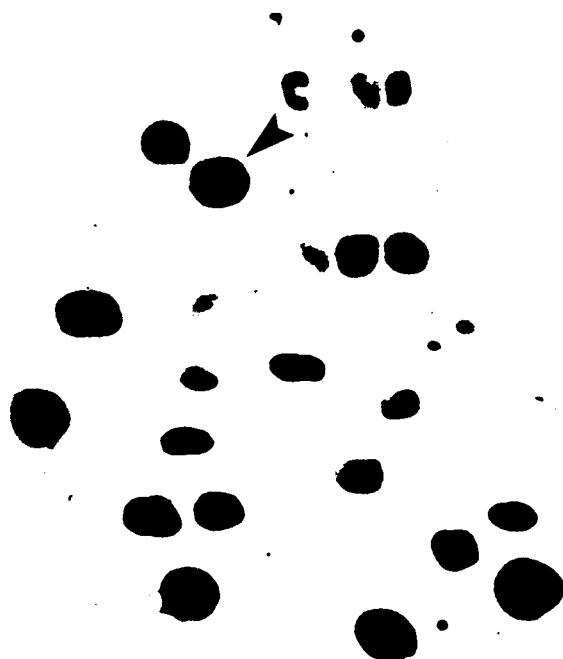


Figure B.1. Colony hybridization using random-primed bovine coronavirus-specific probe. The colony having the largest insert obtained, named pBCV MA7, is identified with an arrow.

culture with 0.15 ml sterile glycerol in a 1.5 ml polypropylene snap cap tube, and freezing at -80°C . Figure B.2 is a photograph of an ethidium bromide-stained agarose gel containing direct colony lysates prepared by the technique of Kado and Liu. Insert sizes were crudely estimated from migration rates of the supercoiled plasmids as compared to pUC9 alone, and to a pUC9 clone containing a 600 bp rabbit B hemaglobin insert.

Samples of plasmid DNA extracted by the method of Kado and Liu were also saved for restriction enzyme digestion and for transformation into the recA F^{-} host DH-1. Subcloning of plasmids was necessary because the host JM103 contains an F factor which copurifies with pUC9 during large scale preparation of plasmid over CsCl.

RESTRICTION ENZYME SITE MAPPING

From 670 white colonies, 89 had coronavirus specific sequences as judged by hybridization to the random primed probe. From these, the nine largest were picked for further analysis.

Initial size measurements of inserts was accomplished by electrophoresing inserts removed with restriction enzyme PstI. Many clones were found to have an internal PstI site, and all of these yielded a common PstI fragment of about 0.5 kilobases in length. This prompted the hypothesis that the internal PstI site mapped about 500 bases from the 3' end of the viral RNA. One would expect most of the clones to be derived from the 3' end of the viral RNA since oligo(dT) was used as the primer.

A detailed restriction endonuclease map of clones CB9, MN3, MA5, and MA7 was obtained using a variety of restriction enzymes and the method of Smith and Bernstiel (Figures B.3 and B.4). Detailed maps were used to determine regions of overlap among cloned inserts. The method of Smith and Bernstiel was performed by labeling one end of the insert fragment and analyzing partial digests collected at different times during a digestion reaction. The sizes of the resultant nested set of fragments were measured from an autoradiograph (Figure B.3).

Restriction enzyme mapping identified six clones that mapped as a nested set and are probably copies of the 3' end of the genome (Table B.1). Verification of this hypothesis was obtained by showing that one of the clones having the 500 bp fragment, detecting cloned plasmids CB9, also hybridized with the poly(dT) probe (Table B.1 and Figure B.4).

The original restriction map shown in Figure B.4 is different from the one constructed from nucleotide sequence data shown in Chapter III, for two reasons. First, only some of the known restriction enzyme sites (only those used for sequencing) are shown in the map of Chapter III. Secondly, the nucleotide sequence data revealed sites for enzymes that were not employed in the Smith and Birnstiel analysis but were used for sequencing. Some errors were corrected after sequencing. For example, two Dde I sites located approximately 40 base pairs apart on the genome were mapped as one site using the Smith and Birnstiel technique. This also very strongly strengthens the idea that sequencing in two directions from a restriction site is not sufficient

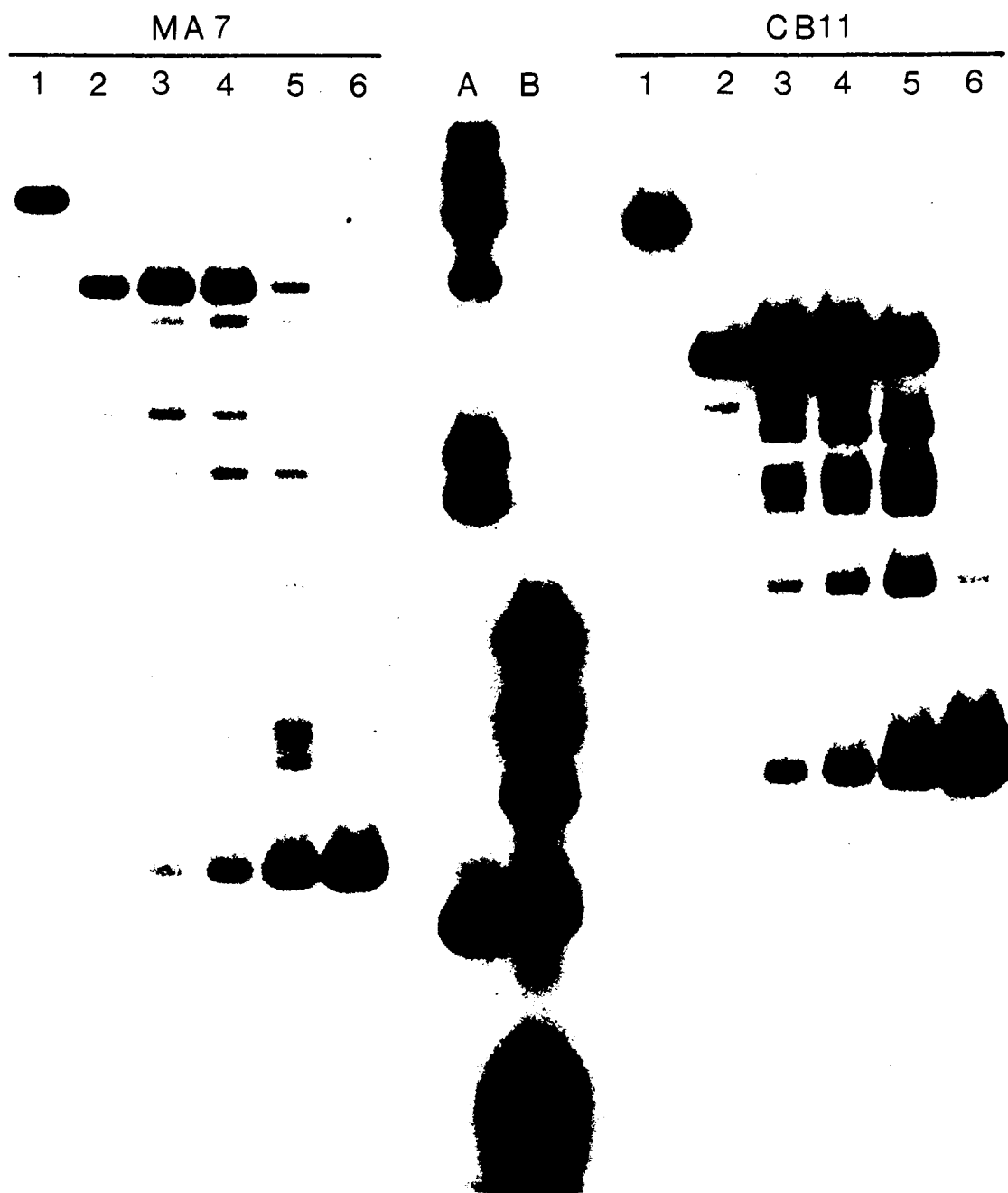


Figure B.3. Sau3AI fragment sizes from clones MA7 and CB11. Lane 1: plasmids linearized with Sal I and end labeled using polynucleotide kinase. Lane 2: labeled, linearized plasmids digested to completion with Hind III and Bam HI to produce a single end-labeled insert. Lanes 3, 4, 5, and 6 depict 2, 5, 20, and 60 minute digestions of the end labeled insert. Lane A: Lambda DNA digested with Hind III and end labeled with polynucleotide kinase. Lane B: Phi X174 digested with Hae III and end labeled with polynucleotide kinase.

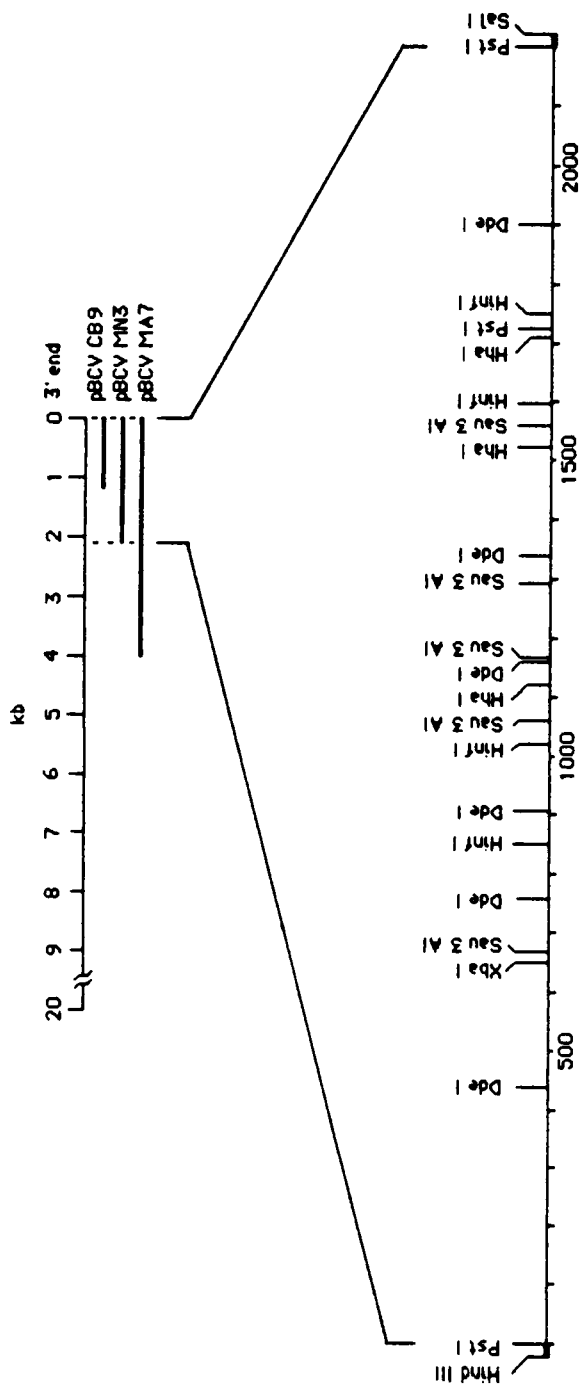


Figure B.4. Restriction endonuclease map of MN3 obtained by the Smith and Birnstiel technique. The top line labeled 1 through 20 designates the genomic RNA of BCV. Lines designated pBCV CB9, pBCV MN3 and pBCV MA7 represent the map location of clones CB9, MN3 and MA7. The expanded line at the bottom represents the sequence of clone MN3. Overlapping regions of CB9 and MA7 are indicated.

Table B.1. Summary of clones analyzed.^a

Clone	Insert Size	400 Base Pst I Fragment	Contains Poly A	Restriction Mapped	Chosen for Sequencing
MA7	4.14	yes	no	yes	yes
CB11	2.95	no	no	yes	no
MA5	2.9	yes	no	yes	yes
MA6	2.9	yes	no	N.D.	no
NL15	2.5 ^b	N.D.	yes	incomplete	no
NL1	2.5	N.D.	no	incomplete	no
MN3	2.25	yes	no	yes	yes
CB1	2.0	yes	no	incomplete	no
CB9	1.15	yes	yes	yes	yes

^aAll clones listed were subcloned into E. coli host DH-1. Insert sizes given were derived from the restriction analysis of inserts. The size of the 400 base Pst I fragment varies slightly from clone to clone and can be explained by the different lengths of the poly A tail. The fragment from clone CB9 is the largest and CB9 has the longest poly A tail (70 bases). Clone MA6 has the same restriction map as MA5, and it is therefore believed to be an identical clone. Clone CB11 may not lie within the MA7 sequence and may therefore come from a more 5' position in the genome.

to complete the sequence of the region. Another stretch of sequences must cover the region!

All clones having large coronavirus-specific inserts as judged by hybridization with random primed cDNA are depicted in Table B.1. Based on the results of the restriction mapping, four clones, MA7, MA5, MN3, and CB9 were chosen for sequencing.

SEQUENCE ANALYSIS OF 3' END SPECIFIC CLONES:

DNA END LABELING AND SEQUENCING

The Maxam and Gilbert technique requires that DNA be end labeled before chemical breakdown. End labeling was accomplished using three different techniques. (1) 3' single-stranded overhangs were labeled with alpha ^{32}P cordycepin-5'-triphosphate (Tu and Cohen, 1980). The reaction conditions used were the same as those described for homopolymer tailing with terminal deoxynucleotidyl transferase. Cordycepin-5'-monophosphate lacks a 3'OH group so that another nucleotide cannot be attached. (2) 5' single stranded overhangs were labeled using polynucleotide kinase in the exchange reaction (Berkner and Folk, 1977). This reaction exchanges unlabeled phosphate on the DNA with ^{32}P gamma phosphate on ATP. (3) 3' recessed ends were labeled by "filling in" at the 3' end using reverse transcriptase (Gail Wertz, personal communication). This reaction was the same as that used for first strand synthesis of cDNA except the only radioactive nucleotide added to the reaction was the one radioactive base.

DNA sequencing reactions were used as modified by Wen Yang (personal communications). The reaction conditions are described in

Table B.2. The time for the hydrazine reaction was further modified by me. When small quantities of DNA were used in these reactions, 20 ug of carrier tRNA was added at the ethanol precipitation steps. This minimized losses that sometimes occurred from DNA not precipitating. Formulas used for sequencing gels are listed in Table B.3.

ISOLATION OF DNA FRAGMENTS FROM GELS

Resolved large DNA fragments (greater than 1 kilobase pairs) were usually recovered from agarose gels by electroelution (H. O. Smith, 1980). DNA was electroeluted into a dialysis membrane-bound slot in a horizontal gel. Electroeluted DNA requires extensive "cleaning up" before it can be digested with enzyme. The clean-up steps usually included butanol extraction to remove ethidium bromide, two PCI extractions, and purification over a small ion exchange column (NACS, BRL).

DNA fragments were at times recovered by electrophoresis onto DEAE ion exchange membranes (Dretzen et al., 1981; Winberg and Hammarskjold, 1980; Lizardi, 1981). These techniques were best used to recover DNA fragments that are smaller than 1 Kb because recovery is good for these sizes. The capacity of the DEAE membranes is 7 ug/cm² for quantitative recovery. DNA recovered using the ion exchange membranes is much purer than that obtained using electroelution into the slot.

DNA fragments from acrylamide gels were eluted by the diffusion technique described by Smith (1980). DNA bands from a 1.6 mm gel were cut out after autoradiographic location and syringed 3 times between 2-one ml syringes. The gel fragments were then incubated overnight at

Table B.2. Flow diagram of the Maxam and Gilbert technique.^a

Date	Code		Sample		A > C
	A > G	G	T & C	C	
10 ul DNA ^b	5 ul DNA ^b 5 ul H ₂ O 200 ul DMS buffer 0°C	10 ul DNA ^b 5 ul H ₂ O 15 ul 5M NaCl 0°C	5 ul DNA ^b	5 ul DNA ^b	5 ul DNA ^b 100 ul NaOH
10 ul H ₂ O 0°C					
2 ul pip. form. (6 ml H ₂ O, 250 ul formic acid, 10 ul piperidine) 37°C, 8 min Freeze dry	1 ul DMS, vortex 20°C, 90 seconds 50 ul DMS stop (0°C) 800 ul EtOH -80°C, 8 min Spin, 8 min	20°C, 9 minutes 200 ul Hydrazine stop 800 ul EtOH Spin, 8 min	30 ul Hydrazine, vortex		90°C 10 min.
	Remove supernatant with Pasteur pipette 300 ul 0.3 M NaOAc, redissolve DNA 800 ul EtOH -80°C, 8 min, Spin, 8 min				
	Rinse the pellet and vial with 100 ul 2 times ice-cold 80% EtOH, spin 5 seconds				
	Freeze, dry				
100 ul, 1M piperidine vortex, spin 2 seconds 90°C, 30 min, with caps clamped down. Short spin Punch holes in caps with flamed needle Freeze dry (2 hr)	100 ul, 1M piperidine (add 200 ul to 2 ml H ₂ O in 15 ml tube with vortexing) vortex, spin 2 seconds 90°C, 30 min, with caps clamped down. Short spin Punch holes in caps with flamed needle Freeze dry (2 hr)				
100 ul H ₂ O, vortex, spin, freeze-dry (overnight)					
50 ul H ₂ O, vortex, spin, freeze-dry (1 hr)					
10 ul loading buffer (80% formamide, 10 mM NaOH, 1 mM EDTA, 0.1% XC and 0.1% BPB)					
50°C, 10 min, vortex and spin (make sure everything is in solution)					
Heat at 90°C, 2 min and quick cool in ice H ₂ O					
Load 2 ul onto sequencing gel with the order of G, A/G, T/C, C					

^aAs modified and used by B. Lapps.

^b40-60 K cpm (Cerenkov) per lane.

Table B.3. Urea-polyacrylamide sequencing gel formulas.

Component	Amount Added for Gels of the Following Concentrations			
	20%	8%	6%	4.5%
40% acrylamide 19:1 ^a	75 ml	30 ml	22.5 ml	16.9 ml
Urea (Ultrapure, IBI)	75 g	75 g	75 g	75 g
5X TBE ^b	15 ml ^c	30 ml	30 ml	30 ml
water to 150 ml ^{d,e}				

^a40% acrylamide is 19:1 acrylamide to bisacrylamide (wt/wt).

^b5X TBE is 54 g Tris base, 27.5 g boric acid, 20 ml of 0.5 M EDTA, and water to a volume of 1000 ml.

^cFinal concentration of TBE in the 20% gel was 0.5X.

^dBe certain urea is completely in solution before adding water. To get urea completely in solution requires the addition of 26 ml, 33 ml, and 39 ml water to the 8%, 6% and 4.5% gel solutions respectively.

^eThe 150 ml volume is filtered by vacuum through a 0.4 um Nalgene syringe, and 1.05 ml--10% $\text{NH}_4\text{S}_2\text{O}_4$ (wt/vol) is added to begin polymerization. From this mixture, 10 ml is removed to form a gel plug. Five ul TEMED is added to the 10 ml aliquot to catalyze complete polymerization. To the remaining solution, 40 ul TEMED is added (35 ul in the case of the 20% gel).

37°C in a siliconized test tube with 1 ml of 0.5 M ammonium acetate, 5 mM EDTA, 0.1% SDS buffer. The slurry was then poured into a 5 ml syringe that had 1.5 cm³ of packed siliconized glass wool. A siliconized 1.5 ml Eppendorf tube was snapped onto the leur-lok tip, and the assembly, with the slurry added, was centrifuged at 2,000 rpm for 10 min on the bench top centrifuge. The eluate was PCI extracted and ethanol precipitated.

DNA STRAND SEPARATION

End labeled restriction fragments up to 700 base pairs in length were strand separated to obtain single-stranded molecules labeled only at one end for sequencing. This technique was especially useful for DNA fragments of 300 bases or smaller. Fragments of this size could be totally sequenced in both strands to clear up any ambiguities or gaps that may exist. The method used is that described by James and Bradshaw (1983), and modified by me. Insert DNA was recovered from agarose gels by electroelution. Insert fragments were then cut to completion with a frequently cutting enzyme. Fragments were then end-labeled and ethanol precipitated twice using ammonium acetate as the salt to remove the unreacted nucleotide triphosphates.

Fragments were separated by electrophoresis on a nondenaturing polyacrylamide gel. The gel composition, voltage and length of run were carefully recorded so as to be repeated exactly during strand separation electrophoresis. For strand separation, electrophoresis conditions were the same except that DNA was first denatured with urea, and the length of the run was increased such that the time would be

1.5- to 2-fold over what was required for the dye to reach the bottom of the gel. The ideal length of run for strand separation was the time required to electrophorese the double-stranded fragment to the bottom of the gel. The three fragments (one double-stranded form and two single-stranded forms) were then observed on the autoradiograph. Separated strands always run slower than the double-stranded DNA on acrylamide gels.

Separation of DNA strands can be done using two possible schemes. For a small number of fragments (1 to 4) all fragments can be denatured and strand separated in the same lane. Alternatively, a larger number of fragments must be separated and individually isolated before strand separation. Isolation of individual fragments is necessary since each double-stranded fragment generates three bands after denaturation, the two separated strands and the residual undenatured fragment. Four native DNA molecules therefore yield 12 bands on the strand separating gel and closely migrating bands are difficult to isolate. DNA was denatured by bringing the concentration of the salt free sample to 10 M urea. (This much urea causes considerable volume expansion.) DNA was heated to 65°C for 3 minutes, quick cooled and loaded immediately onto the gel, taking care that the sample was not cooled to the point that the urea precipitated.

Acrylamide nondenaturing gels were 10 to 15% acrylamide in TBE and had an acrylamide to bisacrylamide ratio of 60:1 (James and Bradshaw, 1984). The gel was 1.6 mm thick, 41 cm long and 38 cm wide. The 1.6 mm thickness allowed for a potentially large DNA mass and, more importantly, a large volume and surface area for the sample to enter

the gel. The sample size was kept as small as possible, usually around 100 ul.

Electrophoresis was carried out at 296 volts constant voltage for 12-18 hours for separation of double-stranded fragments. Strand separated DNA was electrophoresed for 24-36 hours. The electrophoresis buffer was changed every 12 hours to prevent depletion.

Control, undenatured DNA was run on a separate lane on the strand separating gels to determine which preparative bands are strand separated.

Figure B.5 shows electrophoretic separation of double-stranded DNA fragments. Figure B.6 shows a strand separation gel. An example of a sequencing gel, showing the sequences strand separated fragments 3T, 3B, 5T and 5B, is shown in Figure B.7. A large portion of all the sequences for insert MN3 was derived from this set of experiments.

SEQUENCE ANALYSES OF CLONES

All the sequences obtained from cDNA clones were analyzed by the program of Queen and Korn (1984). The following tables and figures describe the sequence analyses. Tables B.4, B.5, B.6 (pages 119-121), and Figures B.8, B.9, B.10, B.11, B.12, B.13, B.14, B.15, B.16, B.17, and B.18 (pages 122-132).

Sequence homologies for both nucleotide sequences and amino acid sequences were expressed as diagon plots. A diagon plot is produced by placing a dot on a matrix that shows the position of a short homologous sequence. A homologous sequence, for all the analyses described in the figures, is defined as a sequence of 20 units, 75% of which must match

perfectly, allowing a maximum of 4 deletions (MINMATCH 20, MINPER 75, LOOPLEN 4). The search for homologous regions was unlimited since the the maximum distance allowable between two starting members of a repeat was 999 units (MA x DIST 999). All sequences compared were shorter than this.

The conclusions from these analyses are discussed in Chapters II and III.

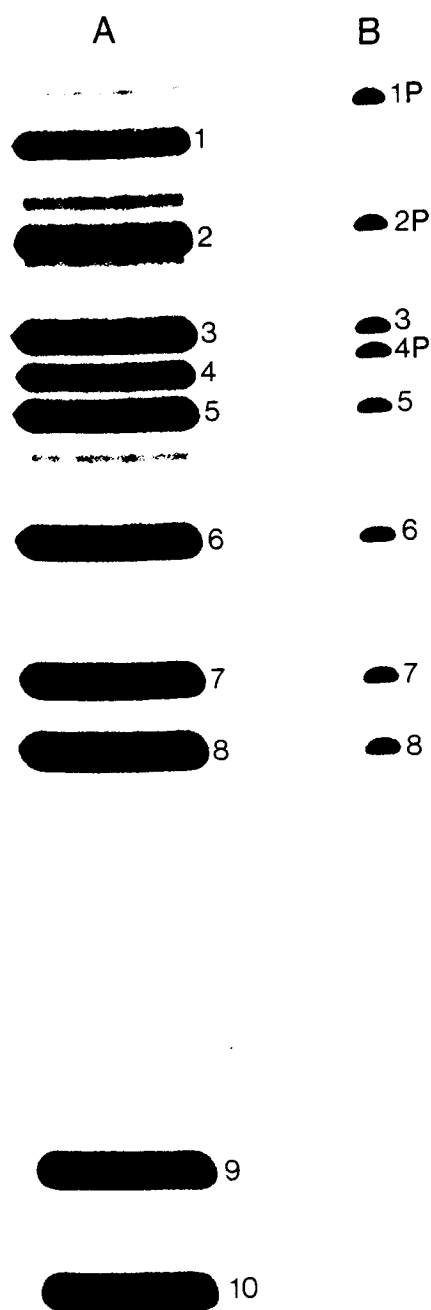


Figure B.5. Double stranded DNA fragment separation on an acrylamide gel. Lane A: Preparative fragments from insert MN3 cut with DdeI, labeled with polynucleotide kinase, and recut with enzyme Pst I. Fragments 1, 4 and most of 2 are double stranded DNA with label only at one end. Lane B: Control lane of DdeI DNA fragments of insert MN3 that are end labeled with polynucleotide kinase. Fragments 1P, 2P, and 4P have the recognition site for Pst I that is used in preparative lane A. This picture is from a 13.5% acrylamide gel that was autoradiographed for 10 minutes.

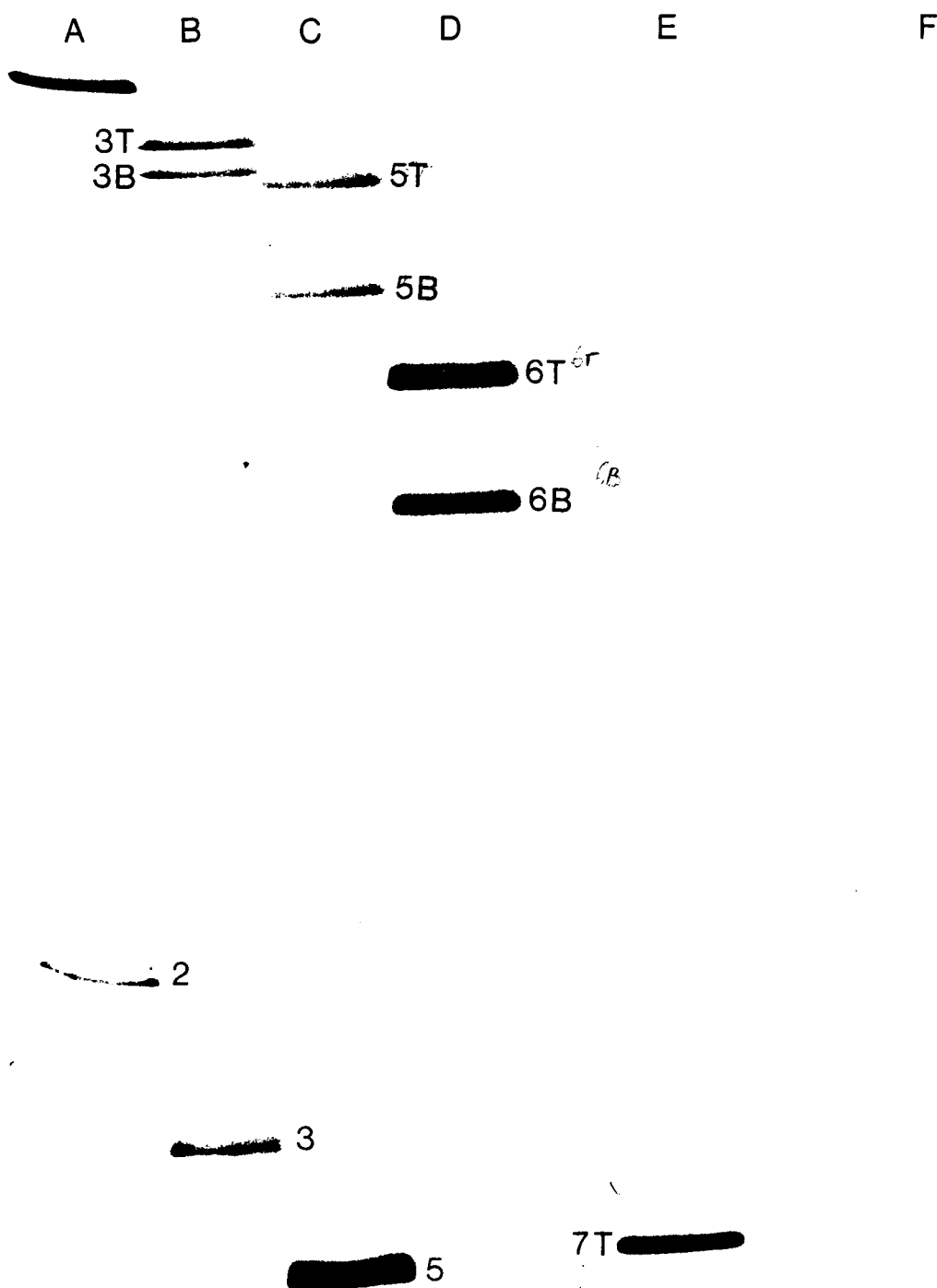


Figure B.6. Strand separation of isolated DNA fragments. The isolated, gel purified DNA fragments 2, 3, 5, 6, and 7 from the gel in Figure B.6 were denatured with 10 M urea and run on lanes A, B, C, D, and E respectively. Lane F shows predominantly undenatured control Dde I fragments of MN3. The bands marked 3T, 3B, 5T, 5B, 6T, 6B, 7T are all single-stranded, end-labeled DNA that were extracted and sequenced. The bands marked 2, 3, and 5 are the double-stranded molecules that failed to strand separate.

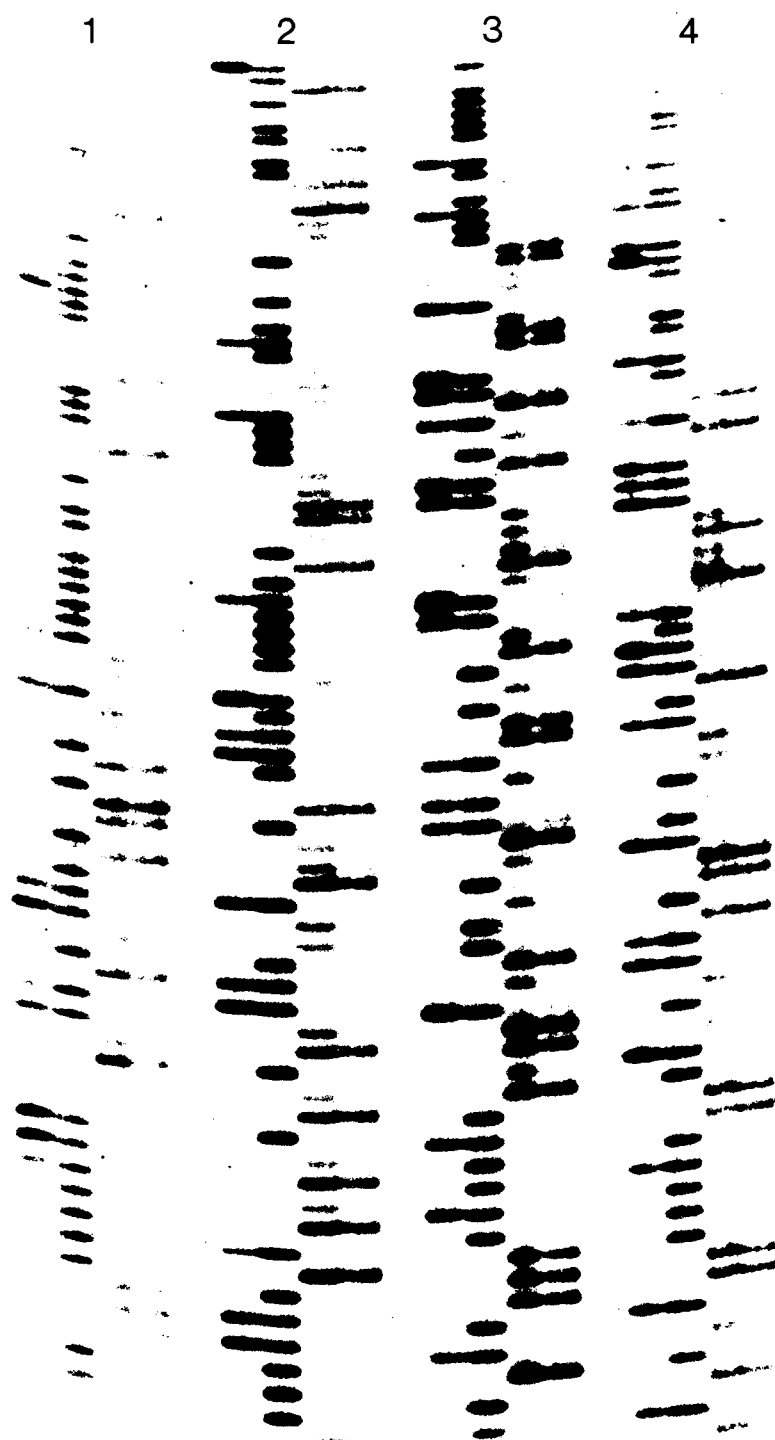


Figure B.7. Maxam and Gilbert sequencing of strand separated molecules. Lanes 1, 2, 3, and 4 shows the sequencing of isolated strands 3B, 3TG, 5B, 5T respectively. These isolated strands were shown in Figure B.7. The loading sequence for each lane is G, A, T and C.

Table B.4. Amino acid composition of the BCV nucleocapsid protein.
The total number of amino acids deduced from the open
reading frame is 448.

Ala	34	(7.6)	Leu	26	(5.8)
Arg	31	(6.9)	Lys	28	(6.3)
Asn	26	(5.8)	Met	8	(1.8)
Asp	23	(5.1)	Phe	18	(4.0)
Cys	2	(0.4)	Pro	30	(6.7)
Gln	37	(8.3)	Ser	42	(9.4)
Glu	20	(4.5)	Thr	27	(6.0)
Gly	41	(9.2)	Trp	5	(1.1)
His	2	(0.4)	Tyr	14	(3.1)
Ile	16	(3.6)	Val	18	(4.0)
End	0	(0.0)			
Acidic	(Asp + Glu)		43	(9.6)	
Basic	(Arg + Lys)		59	(13.2)	
Aromatic	(Phe + Trp + Tyr)		37	(8.3)	
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)		105	(23.4)	

Molecular Weight = 49379.

Table B.5. Amino acid composition of the BCV matrix protein protein.
The total number of amino acids deduced from the open
reading frame is 230.

Ala	9	(3.9)	Leu	24	(10.4)
Arg	9	(3.9)	Lys	11	(4.8)
Asn	11	(4.8)	Met	10	(4.3)
Asp	7	(3.0)	Phe	13	(5.7)
Cys	2	(0.9)	Pro	7	(3.0)
Gln	3	(1.3)	Ser	16	(7.0)
Glu	4	(1.7)	Thr	20	(8.7)
Gly	15	(6.5)	Trp	7	(3.0)
His	3	(1.3)	Tyr	14	(6.1)
Ile	27	(11.7)	Val	18	(7.8)
End	0	(0.0)			
Acidic	(Asp + Glu)			11	(4.8)
Basic	(Arg + Lys)			20	(8.7)
Aromatic	(Phe + Trp + Tyr)			34	(14.8)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)			113	(49.1)

Molecular Weight = 26376.

Table B.6. Predicted amino acid composition from the internal open reading frame (IORF) found in BCV. The number of amino acids represented in IORF is 207.

Ala	8	(3.9)	Leu	40	(19.3)
Arg	12	(5.8)	Lys	9	(4.3)
Asn	4	(1.9)	Met	8	(3.9)
Asp	7	(3.4)	Phe	6	(2.9)
Cys	3	(1.4)	Pro	15	(7.2)
Gln	9	(4.3)	Ser	17	(8.2)
Glu	13	(6.3)	Thr	15	(7.2)
Gly	11	(5.3)	Trp	0	(0.0)
His	5	(2.4)	Tyr	3	(1.4)
Ile	10	(4.8)	Val	12	(5.8)
End	0	(0.0)			
Acidic	(Asp + Glu)			20	(9.7)
Basic	(Arg + Lys)			21	(10.1)
Aromatic	(Phe + Trp + Tyr)			9	(4.3)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)			79	(38.2)

Molecular Weight = 23057.

```

1      M S S V T   T P A P V Y T W T A D E A I
      | | | | |   | | | | | | | | | | | |
1      M S S T T Q A P E P V Y Q W T A D E A V

20     K F L K E W N F S L G I I L L F I T I I
21     Q F L K E W N F S L G I I L L F I T I I

40     L Q F G Y T S R S M F V Y V I K M I I L
41     L Q F G Y T S R S M F I Y V V K M I I L

60     W L M W P L T I I L T I F N C V Y A L N
61     W L M W P L T I V L C I F N C V Y A L N

80     N V Y L G F S I V F T I V A I I M W I V
81     N V Y L G F S I V F T I V S I V I W I M

100    Y F V N S I R L F I R T G S W W S F N P
101    Y F V N S I R L F I R T G S W W S F N P

120    E T N N L M C I D M K G R M Y V R P I I
121    E T N N L M C I D M K G T V Y V R P I I

140    E D Y H T L T V T I I R G H L Y M Q G I
141    E D Y H T L T A T I I R G H L Y M Q G V

160    K L G T G Y S L S D L P A Y V T V A K V
161    K L G T G F S L S D L P A Y V T V A K V

180    S H L L T Y K R G F L D K I G D T S G F
181    S H L C T Y K R A F L D K V D G V S G F

200    A V Y V K S K V G N Y R L P S T Q K G S
201    A V Y V K S K V G N Y R L P S N   K P S

220    G M D T A L L R N N I
220    G A D T A L L R   I

```

Matches = 199 Mismatches = 28 Unmatched = 4
 Length = 231 Matches/length = 86.1 percent

Figure B.8. Predicted alignment of homologous amino acids for the matrix glycoproteins of BCV and MHV-A59. The upper string of amino acids is the BCV matrix protein amino acid sequence. The vertical lines represent homologous amino acids.

```

115   ATGAGTAGTGTAACT A  CAOCAGCACCACTTTACACCTGGACTGCTGATGAAGCTATT
      |||||
77    AUGAGUAGUACUACUCAGGCCCCAGAGCCCGUCUAUCAUUGGACGGCCGACGAGGCAGUU
      |||||
172   AAATTCCTAAAGGAATGGAACCTTTCTTTGGGTATTATACTACTTTTATTACAATCATA
      |||||
137   CAAUCCUUAAGGAAUGGAACUUCUCGUUGGGCAUUAUACUACUUAUUAUUAUUAUUAUUA
      |||||
232   TTGCAATTGGAATATACAAGTCGCAGTATGTTTGTATTATTAAGATGATCATTTTG
      |||||
197   CUACAGUUCGGUUAACAGAGCCGUAGCAUGUUUAUUUAUGUUGGAAAUGAUAUUCUUG
      |||||
292   TGGCTTATGTGGCCCTTACTATCATCTTAACTATTMTCAATTGCGTGTATGCGTTGAAT
      |||||
257   UGGUUAUUGUGGCCACUGACUAUUGUUUGUGUAUUUCAAUUGCGUGUAUGCGCUAAAU
      |||||
352   AATGTGTATCTTGGCTTTTCTATAGTTTCACTATAGTGGCCATTATCATGTGGATTGTG
      |||||
317   AAUGUGUAUCUUGGAUUUUCUAUAGUGUUUACUAUAGUGUCCAUGUAAUCUGGAUUAUG
      |||||

412   TATTTTGTGAATAGTATCAGGTGTGTTATTAGAACTGGAAGTGGTGGAGTTTCAACCCA
      |||||
377   UAUUUUGUUAUAUAGCAUAAGGUUGUUUAUCAGGACUGGUAGCUGGUGAGCUUACAACCC
      |||||
472   GAAACAAACAACTTGTATGTATAGATATGAAGGGAAGGATGTATGTTAGGCCGATAATT
      |||||
437   GAAACAAACAAUUAUUGUGUAUAGAUAUGAAAGGUACCGUGUAUGUUAAGACCAUUAUU
      |||||
532   GAGGACTACCATACCTTACGGTCACAATAACGTGGTCATCTTTACATGCAAGGTATA
      |||||
497   GAGGAUUAACCAUACACUAACAGCCACUAUUAUUCGUGGCCACCUUACAUGCAAGGUGUU
      |||||
592   AAAGTAGGTACTGGCTATTTCTTTGTCAGATTGCGCAGCTTATGTGACTGTGTGCTAAGGTC
      |||||
557   AAGCUAGGCACCGGUUUCUUCUUGUCUGACUUGCCCGCUUAUGUUAACAGUUGCUAAGGUG
      |||||
652   TCACACCTGCTCAGTATAAGCGTGGTTTCTTTGACAAGATAGCGGATAGTGGTTT
      |||||
617   UCACACCUUUGCACUUAUAAGCCGCGCAUUCUUAAGACAAGGUAAGCGGUUAGCGGUUUU
      |||||

712   GCTGTATTATGTTAAGTCCAAAGTCGGTAATTACCGACTGCCATCAACCCAAAGGGTTCT
      |||||
677   GCUGUUUAUGUGAAGUCCAGGUCGGAUUUAACCGACUGCCCUCAAACAAACCGAG  U
      |||||

772   GCCATGGACACCGCATTGTGTGAGAAATATATC
      |||
734   GGCCTGGACACCGCAUUGUUGAGAAUC
      |||

```

Matches = 547 Mismatches = 134 Unmatched = 12
 Length = 693 Matches/length = 78.9 percent

Figure B.9. Predicted alignment of homologous nucleotides for the matrix glycoprotein genes of BCV and MHV-A59. The upper string of nucleotides is the BCV matrix glycoprotein gene. The vertical lines represent homologous nucleotides between the two genes.

GP26AA CORRECT
mhv-a59 e1

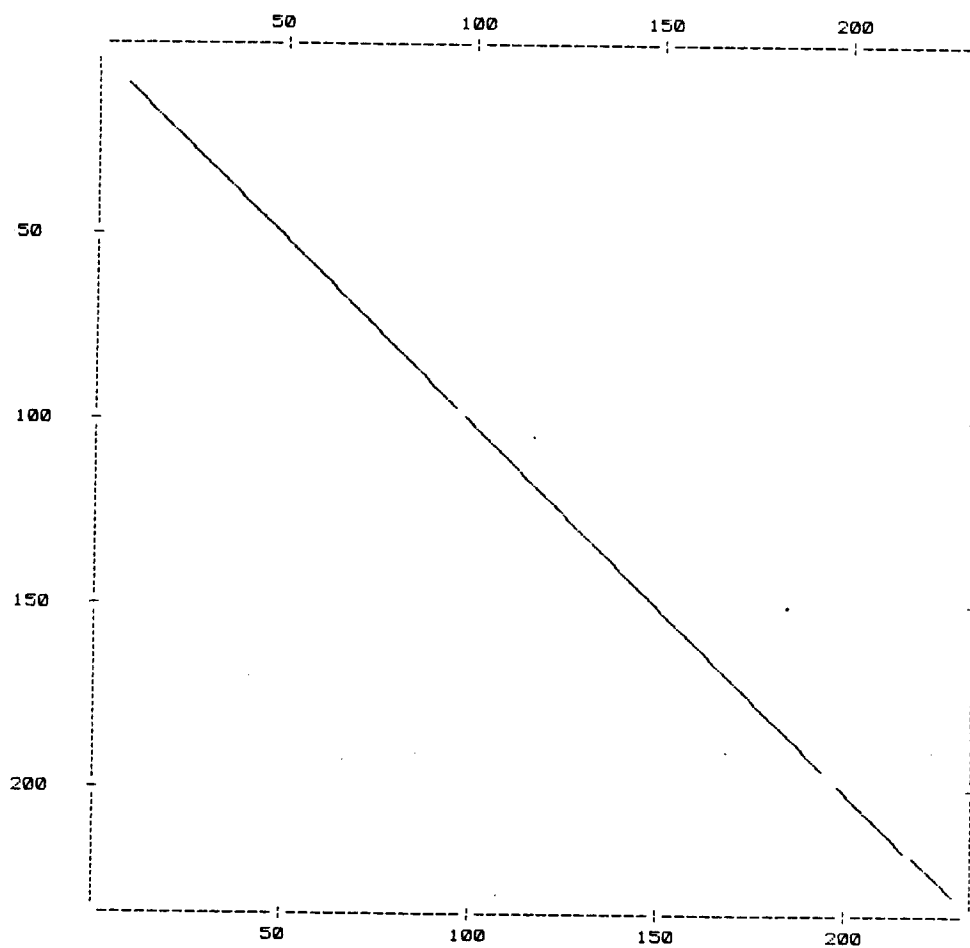


Figure B.10. Amino acid homologies of the BCV and MHV-A59 matrix glycoproteins compared by diagonal plotting.

BCVE1N
MHV-A59 GP26NUC

LIMITS: 115 804
LIMITS: 77 760

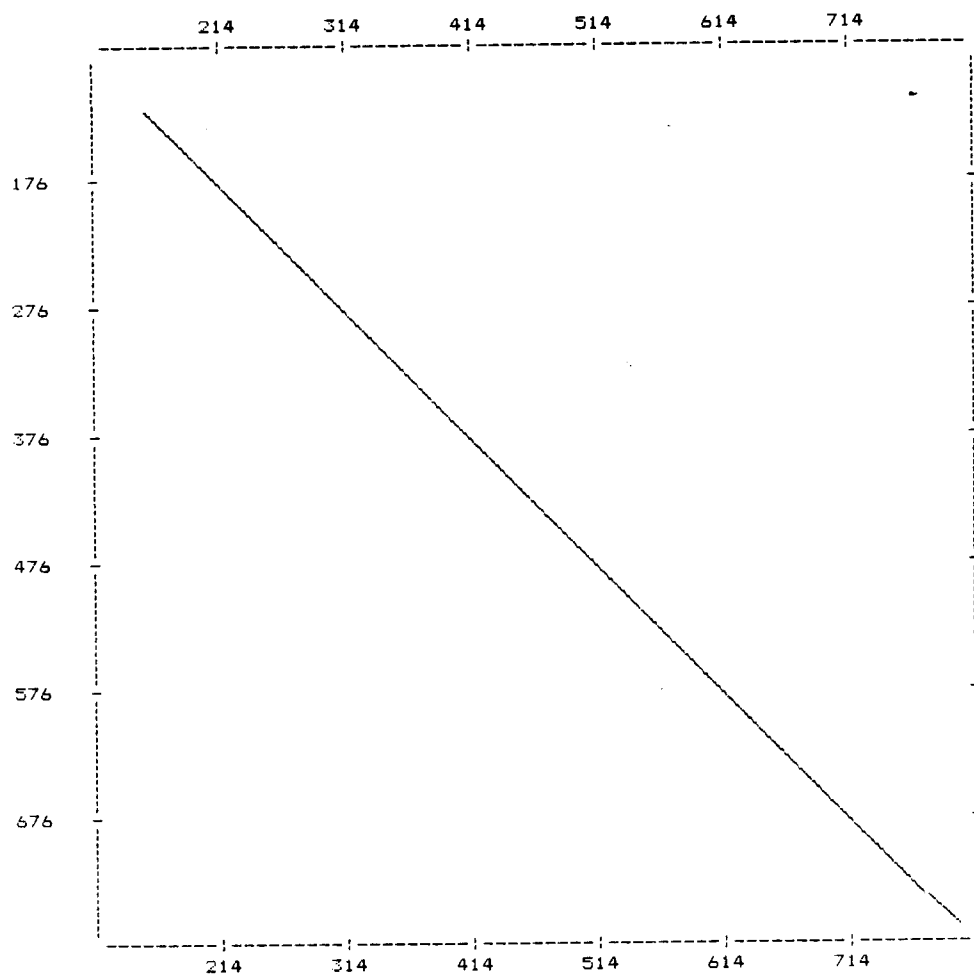


Figure B.11. Nucleotide homologies of the BCV and MHV-A59 matrix glycoprotein genes compared by diagonal plotting.

BCV N PROTEIN
MHV-A59 n protein

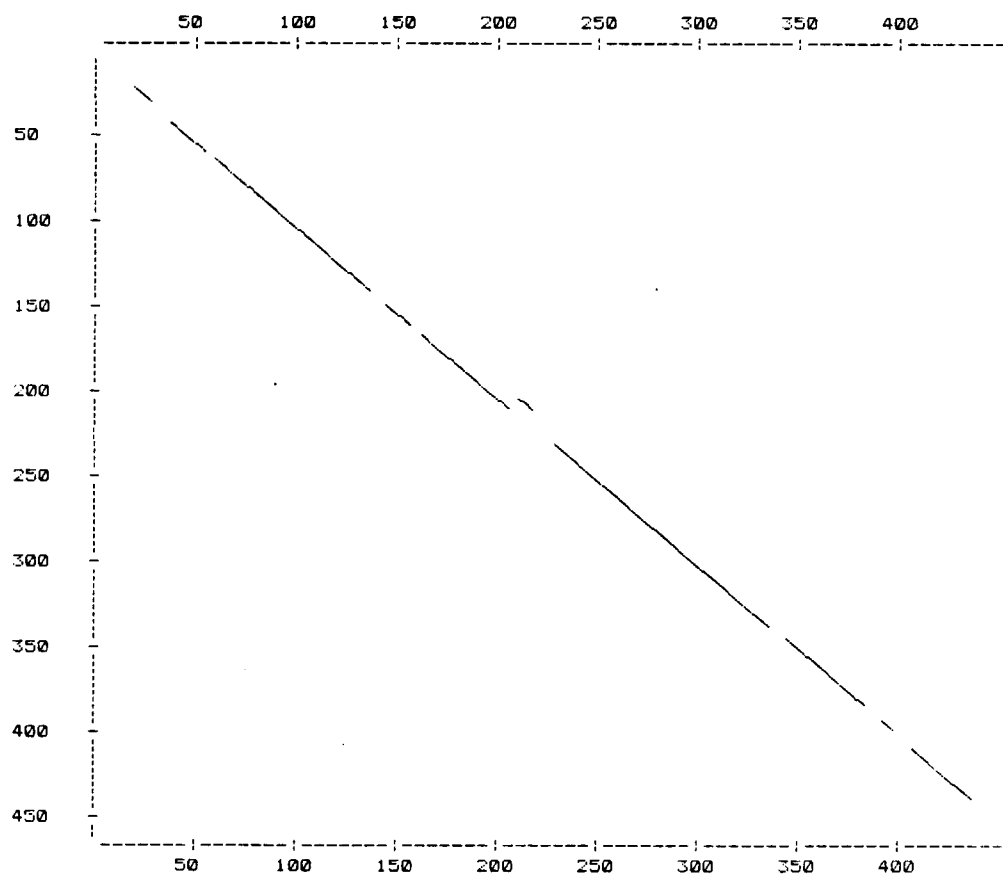


Figure B.12. Amino acid homologies of the BCV and MHV-A59 nucleocapsid protein compared by diagon plotting.

BCV N PROTEIN
 MHV-JHM n protein

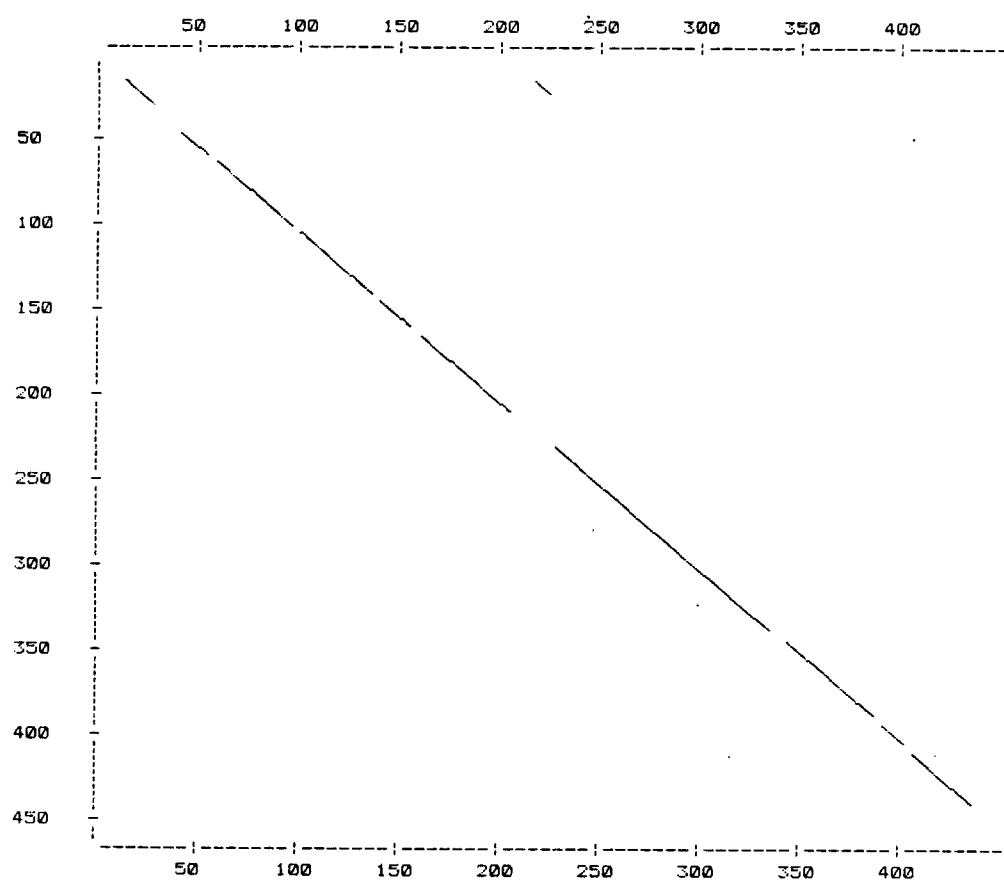


Figure B.13. Amino acid homologies of the BCV and MHV-JHM nucleocapsid proteins compared by diagon plotting.

BCV N PROTEIN
TGEV N PROTEIN

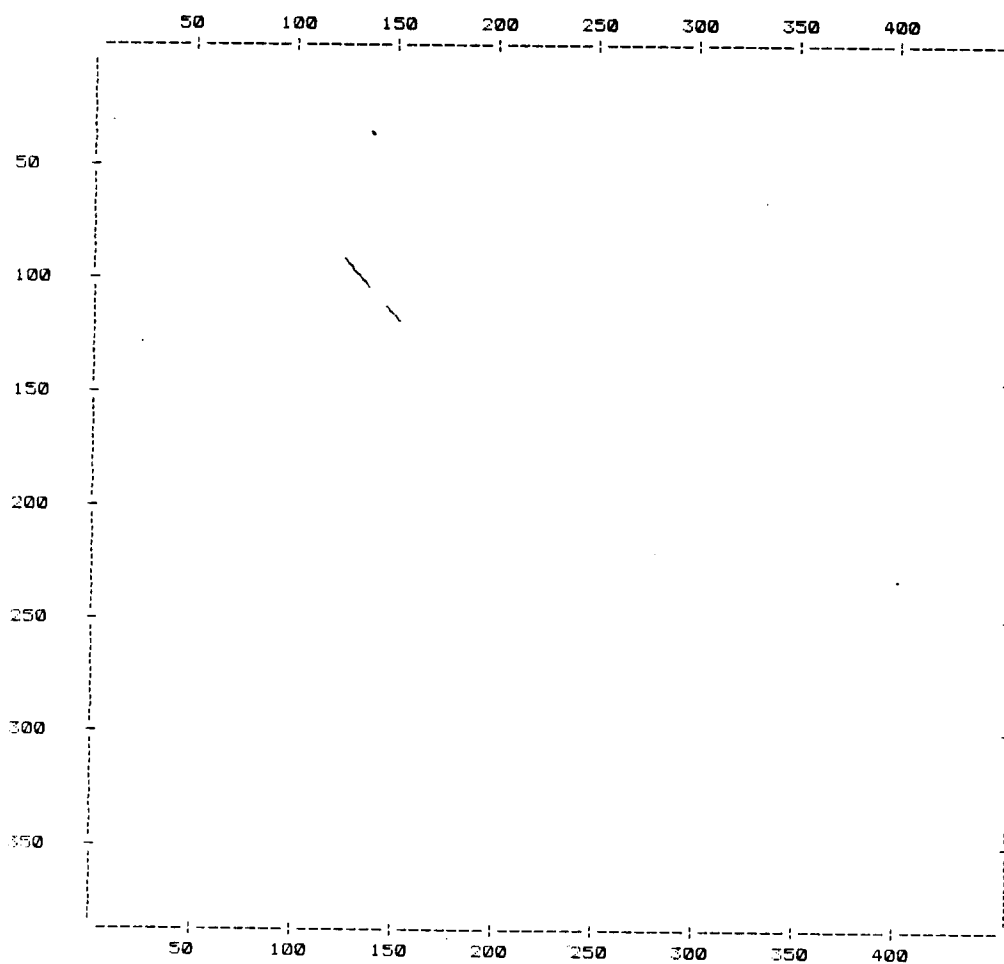


Figure B.14. Amino acid homologies of the BCV and TGEV nucleocapsid proteins compared by diagon plotting.

BCV N PROTEIN
IBV N PROTEIN

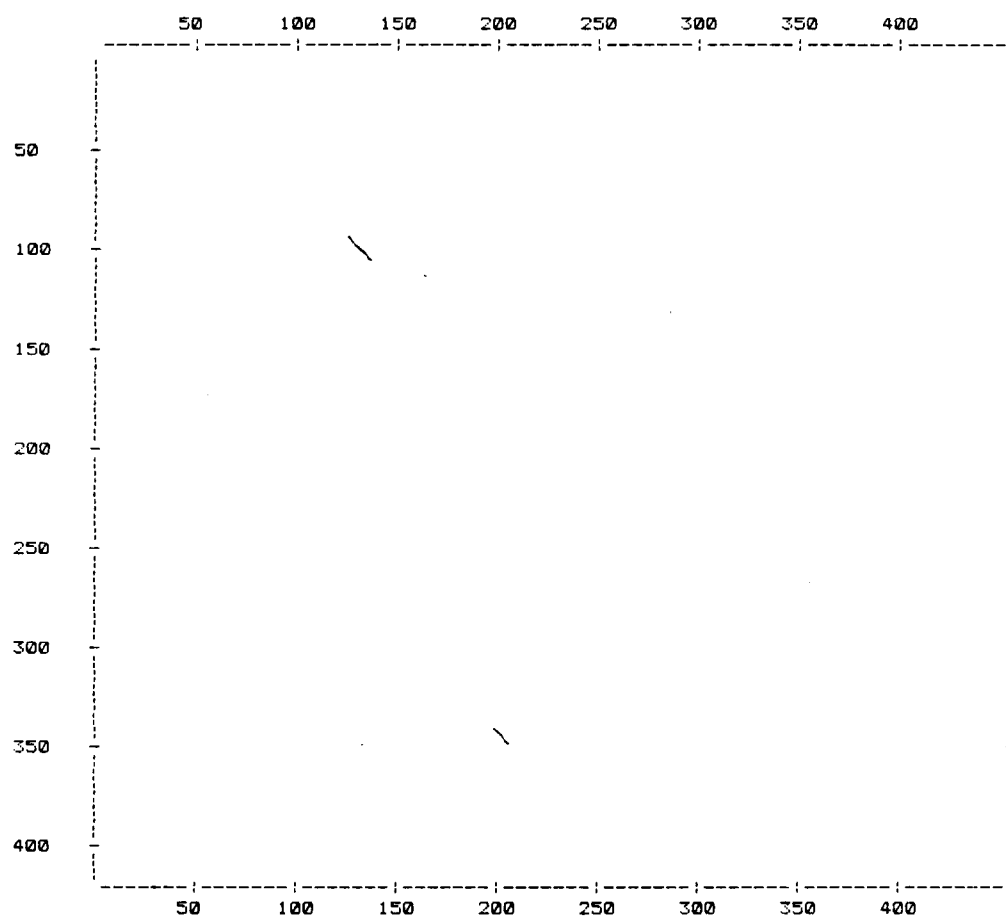


Figure B.15. Amino acid homologies of the BCV and IBV nucleocapsid proteins compared by diagon plotting.

BCVE1N
MHV-A59 GP26NUC

LIMITS: 817 2160
LIMITS: 778 2139

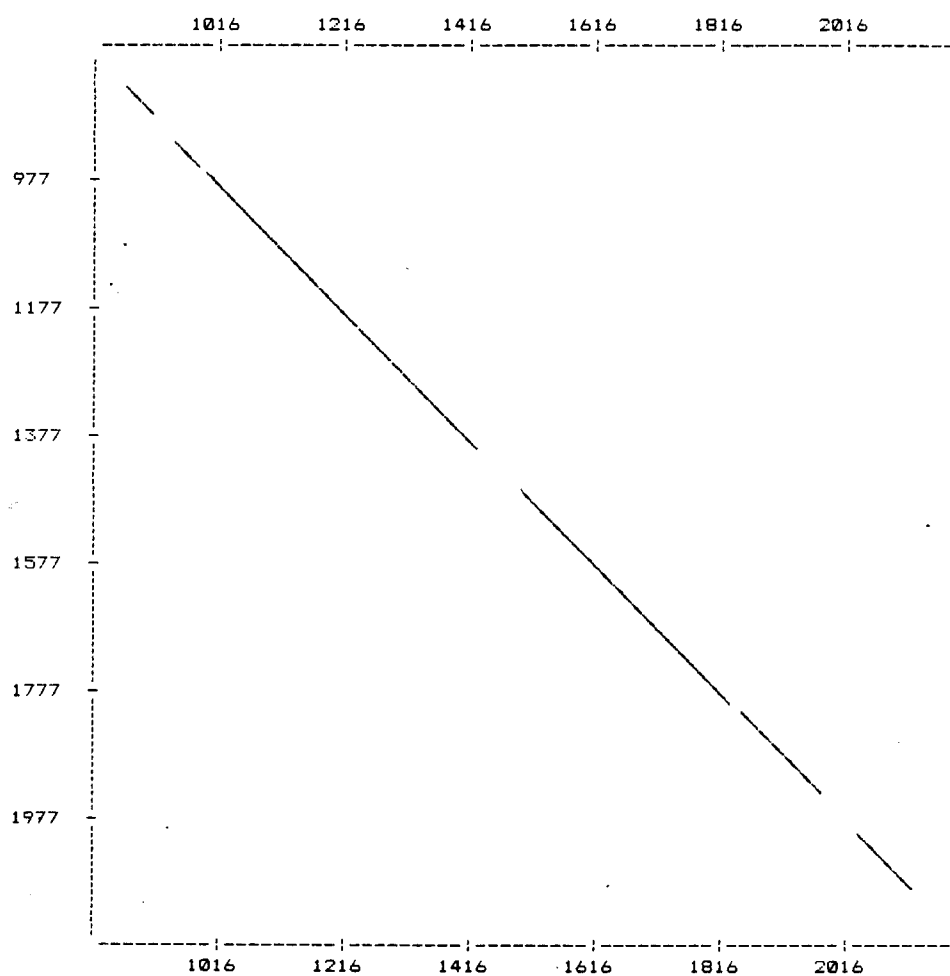


Figure B.16. Nucleotide homologies of the BCV and MHV-A59 nucleocapsid protein genes compared by diagon plotting.

BCVEIN
JHM NUCLEOCAPSID

LIMITS: 817 2160
LIMITS: 84 1448

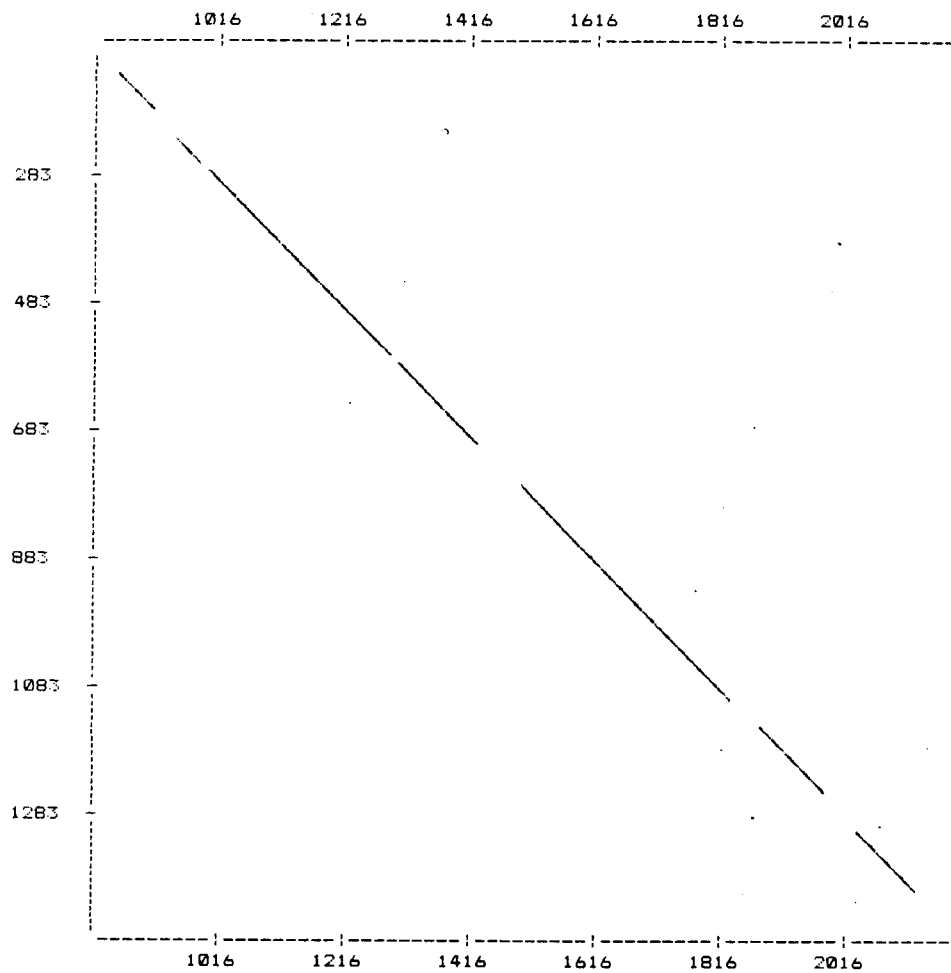


Figure B.17. Nucleotide homologies of the BCV and MHV-JHM nucleocapsid protein genes compared by diagonal plotting.

[illegible]

132

APPENDIX C

HANAHAN'S METHOD FOR TRANSFORMING E. COLI WITH PLASMIDS

(This procedure is summarized from Hanahan, D. 1983. Studies on transformation of E. coli with plasmids. J. Mol. Biol. 166:577-580. This paper should be read thoroughly. It is a classic.)

Materials:

YT Plates with Ampicillin, IPTG and Xgal (from Miller, J. H. 1974. Expt. in Mol. Gen. (CSHL) p. 433, but containing 100 ug/ml ampicillin, 1mM IPTG, and 0.004% w/v X-gal, for use with JM103 and Messing's pUC vectors as described by Heidecker G. and J. Messing. 1983. Nucl. Acids Res. 11:4891-4906. Note Maniatis cloning book p. 71 for preparation of ampicillin stock, and p. 290 for preparation of X-gal stock. Note BRL M13 cloning manual, p. 8 for preparation of IPTG. 1 liter makes 50-100 mm x 15 mm plates at 20 ml/plate.

Ampicillin stock: prepare a 25 mg/ml solution of the sodium salt of ampicillin in Milli-Q-filtered water. Sterilize by filtration through 0.2 um filter and store in aliquots at -20°C. For plates, allow autoclaved medium to cool to 55°C before adding ampicillin.

8 g Bacto tryptone
5 g Bacto yeast extract
5 g NaCl
15 g Agar
Milli-Q-filtered water to 1 liter. Autoclave 15 lb/in² (121°C, 256°F) 30 min. Cool to 55°C then add:
4 ml Ampicillin stock (25 mg/ml)
10 ml IPTG (dissolve 238 mg in 10 ml water [this is a 0.1 M solution since the m.w. of IPTG = 238] and filter sterilize through a 0.2 um filter on a syringe.)
2 ml XGal (dissolve 40 mg in 2 ml dimethylformamide. No filter sterilization is necessary. Store dimethyl-formamide under N₂ at -20°C.)

Ampicillin in plates goes bad after about 2 weeks. IPTG and XGal are good for 1 year (Messing, personal communication.)

SOB Medium

Part A (SOB without Mg⁺⁺):

20 g Bacto tryptone
5 g Bacto yeast extract
0.58 g NaCl (m.w. 58.4)
0.19 g KCl (m.w. 74.5)
MPF water to 1 liter autoclave

Part B (2 M Mg⁺⁺):

20.3 g MgCl₂ · 6H₂O (f.w. 203)
12.0 g MgSO₄ (m.w. 120)
MPF water to 100 ml. Filter sterilize.
This is 100x.

Add 1 ml sterile 2 M Mg⁺⁺ solution to 99 ml sterile SOB medium (minus Mg⁺⁺) to make complete SOB medium.

SOC Medium

99 ml SOB medium (complete) (sterile)
1 ml 2 M glucose stock (sterile)

2 M Glucose Stock

360 g glucose (m.w. 180)
MPF water to 1 liter. Filter sterilize. Store at 4°C.

1M K-MES (pH 6.3) (10 ml is used to prepare 1 liter of TFB.)

19.5 g MES (2-[N-Morpholino] ethane sulfonic acid) (f.w. 195)
Millipore-filtered water to 50 ml. Adjust pH to 6.3 with 1 M KOH.
Millipore-filtered water to 100 ml. Filter sterilize through prerinsed
0.22 um filter and store at -20°C.

TFB (Standard Transformation Buffer: 10 mM K-MES (pH 6.20), 100 mM-RbCl, 45 mM MnCl₂ · 4H₂O, 10 mM CaCl₂ · 2H₂O, 3 mM HA CoCl₃ (1 ml used/transformation)

10 ml 1 M MES adjusted to pH 6.3 with KOH
12.1 g RbCl (m.w. 121)
8.9 g MnCl₂ · 4H₂O (f.w. 198)
1.47 g CaCl₂ · 2H₂O (f.w. 147)
0.80 g [Co(NH₃)] Cl₃ (f.w. 267)
Millipore-filtered water to 1 liter. Filter sterilize through prerinsed 0.22 um filter and store at 4°C.

FSB (Transformation buffer for frozen storage of competent cells: 10 mM potassium acetate, 100 mM KCl, 45 mM MnCl₂ · 4H₂O, 10 mM CaCl₂ · 2H₂O, 3 mM HAcCoCl₃, 10% redistilled glycerol.)

0.98 g potassium acetate (m.w. = 98.1)
0.45 g KCl (m.w. 74.5)
8.9 g MnCl₂ · 4H₂O (f.w. 198)
1.47 g CaCl₂ · 2H₂O (f.w. 147)
0.80 g HA · CoCl₃ (f.w. 267)
Millipore-filtered water to 0.5 liter. Adjust pH to 6.4 with 0.1 M HCl. Then add water to liter. Filter sterilize through prerinsed 0.22 um filter and store at 4°C.

DMSO Open sealed, new bottle of DMSO, portion into 550 ul amounts in 0.5 ml Eppendorf polypropylene tubes, store at -20°C.

DDT Stock (2.25 M DTT, 40 mM potassium acetate [pH 6.0])

3.46 g DDT
0.4 ml 1 M potassium acetate (pH 7.0)
MPF water to 6 ml. Adjust pH to 6.0 with 0.1 N HCl.
MPF water to 10 ml. Filter sterilize. Store -20°C in 0.5 ml aliquots.
Thaw, use promptly, refreeze.

Procedure for standard transformation:

1. Scrape stock of frozen cells with sterile tip (sterile pasteur pipet works well). Spot onto LM* plate, when melted, spread for isolation. Incubate at 37°C overnight.
2. Pick colony (2.5 mm diam.) and disperse in 1 ml of SOB medium by moderate vortexing.
3. Use this to inoculate 10 to 30 ml SOB medium in a 300 ml flask (or 30 to 100 ml medium in a 1000 ml flask, etc.). Incubate at 37°C, 257 rev/min until density is 4×10^7 to 7×10^7 /ml (absorbance 550 nm = 0.45 to 0.55 for DH1 in 2-2.5 h).
4. Collect cells (2.5 ml per discrete transformation) into 50 ml polypropylene tubes (Falcon 2070), place on ice 10-15 min, and pellet at 2500 rpm, 12' at 4°C.
5. Resuspend cells in 1/3 vol. TFB by gentle vortexing, place on ice 10-15 min, pellet again 2500 rpm, 10' at 4°C.
6. Resuspend pellet in TFB at 1/12.5 of original cell volume (2.5 ml of the culture is concentrated to 200 ul for one discrete transformation; aliquot into 200 ul amounts if needed).
7. Add fresh DMSO to 3.5% (7 ul/200 ul), swirl, and leave on ice for 5 min.
8. Add DTT to 75 mM (7 ul stock/200 ul), swirl, and leave on ice for 10 min.
9. Add same vol. of DMSO again (7 ul/200 ul), incubate cells on ice for 5 min.
10. Samples (210 ul) are then placed into chilled 17 mm x 100 mm polypropylene tubes (Falcon 2059).
11. Add DNA ($\leq$ 10 ul) and swirl mixture, incubate on ice 30 min.
12. Heat pulse without agitation 90 sec. at 42°C and place on ice 1-2 min.
13. Add 800 ul SOC medium (20°C), incubate tubes at 37°C, 225 rpm in shaker, 1 h.
14. Pipet an appropriate fraction of the culture in a pool of SOB medium (100-200 ul) on an LM* plate (with appropriate antibiotic). Spread gently and minimally using a bent (L-shaped) pasteur pipette.
15. Incubate at 37°C to establish colonies.

*We use YT plates with Ampicillin, XGal and IPTG for use with pUC vectors and the M13 cloning system of Messing.

Procedure for frozen storage and use of frozen competent cells: Cells are prepared as described above except that FSB is used. DMSO is added twice and DTT is omitted. The cells are portioned out in 200 μ l volumes into Falcon 2059 tubes or screw-cap polypropylene tubes (e.g. Nunc), flash-frozen in solid CO₂/EtOH or liquid N₂, and placed at -80°C. To use, tubes are removed and thawed in air at 20°C, and, when just liquid, placed on ice for 10 min. DNA is added and the cells heat-pulsed and incubated as described above. Transformation efficiency is 10% of that for the standard transformation procedure.

VITA

William Lapps was born in Philadelphia, Pennsylvania on December 7, 1953. He attended elementary school in Newtown Square and graduated from Marple Newtown Senior High School. His first undergraduate year was spent at Alliance College in Cambridge Springs, Pennsylvania. In May of 1975 he graduated with a Bachelor of Arts degree in Biology, minoring in Russian language from Temple University, Philadelphia, Pennsylvania. During 1978 he had courses and training in histology and electron microscopy at Saint Joseph's University in Philadelphia which led to a position as electron microscopist in the Department of Neurology at University of Pennsylvania. In the fall of 1980 he accepted a teaching assistantship in the Department of Microbiology at The University of Tennessee, Knoxville. In March of 1986 he received a Doctor of Philosophy degree in Microbiology.

William Lapps is a member of the American Society of Microbiology. He was funded for two years as a National Institutes of Health Predoctoral Trainee.

William Lapps is a postdoctoral fellow in the Department of Molecular Genetics at Hoffmann La Roche, in Nutley, New Jersey.