To the Graduate Council:

I am submitting herewith a dissertation written by Paul Alan Kapke entitled "Molecular Cloning and Sequence Analysis of the Porcine Transmissible Gastroenteritis 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. Stuart Riggsby

Karl Sirotkin

Beth L. Mullin

Accepted for the Council:

Lou Winkel

Vice Provost
and Dean of the Graduate School

# MOLECULAR CLONING AND SEQUENCE ANALYSIS OF THE PORCINE TRANSMISSIBLE GASTROENTERITIS CORONAVIRUS GENOME

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Paul Alan Kapke

December 1985

## DEDICATION

This work is dedicated to Jackie and Nathan for their love and support.

ACKNOWLEDGEMENTS

The author wishes to acknowledge his appreciation to Dr. D. Brian for his guidance, patience and friendship.

I also wish to acknowledge the people, past and present, who served on my committee. They are Dr. A. Brown, Dr. W. Farkas, Dr. R. Courtney, Dr. L. Etkin, Dr. L. Boone, Dr. K. Sirotkin, and Dr. B. Mullin.

I wish to give special thanks to Dr. W. Stuart Riggsby who was the only committee member to faithfully serve on my committee for the entire 6 years (and was inadvertently omitted in the initial draft of my dissertation). His unerring perseverance and dedication in seeing my project through to its completion was greatly appreciated.

A special thanks also goes to my lab mates for their many discussions, companionship and many good memories.

ABSTRACT

Two approaches were taken to examine the genome structure of the porcine transmissible gastroenteritis coronavirus (TGEV). The first approach was designed to compare the degree of nucleotide sequence homology between the genomes of TGEV and the bovine enteric coronavirus (BCV), a member of an antigenically distinct coronavirus subgroup. For this study, the RNA genomes of TGEV and BCV were separately metabolically labeled with $^{32}$P-orthophosphate, and analyzed by two-dimensional T1 oligonucleotide fingerprinting. From an analysis of comigrating oligonucleotides in a mixing experiment, it was determined that TGEV and BCV share a sequence homology of approximately 91%. This degree of homology supports the notion that the two viruses, while antigenically dissimilar, arose from a common ancestor. It further raises the question of whether conserved sequences might be confined to restricted regions of the genome allowing identification of conserved functional domains. The second approach was designed to systematically examine the primary structure of the TGEV genome thereby allowing rigorous genetic analyses. For this approach, the 3' end of the 20 Kb TGEV genome was cloned and sequenced, and the properties of two genes were deduced and analyzed. TGEV genomic RNA was copied into cDNA after priming with oligo(dT) and the double stranded product was cloned into the Pst 1 site of the pUC9 vector. One clone of 2.0 Kb contained part of the poly(A) tail and was sequenced in its entirety using the chemical method of Maxam and Gilbert. Another clone of 0.7 Kb also contained part of the poly(A) tail and was sequenced in part to confirm

the primary structure of the most 3' end of the genome. Two potential nonoverlapping genes were identified within the 3' terminal 1663 bases from an examination of open reading frames. The first gene encodes a 382 amino acid protein of 43,426 molecular weight, that is the apparent nucleocapsid protein on the basis of size, chemical properties, and amino acid homology with other coronavirus nucleocapsid proteins. It is flanked on its 5' side by at least part of the matrix protein gene. The second gene encodes a hypothetical 78 amino acid protein of 9101 molecular weight that is hydrophobic at both ends. A 3' non-coding sequence of 273 bases was also determined and a sequence of 9 nucleotides near the poly A tail was found to be conserved for TGEV, the mouse hepatitis coronavirus, and the avian infectious bronchitis coronavirus.

TABLE OF CONTENTS


| PART | | PAGE |
|---|---|---|
| I. | BACKGROUND ON CORONAVIRUSES | 1 |
| | Introduction | 2 |
| | What is Known About Coronavirus Gene Structure and Replication Strategy as Determined from the Three Most Highly Studied Species, the Mouse Hepatitis Coronavirus, the Infectious Bronchitis Coronavirus, and the Porcine Transmissible Gastroenteritis Coronavirus. | 5 |
| | The Porcine Transmissible Gastroenteritis Coronavirus Possesses Some Unique Features That Make It Worthy of Detailed Study. | 14 |
| | The Virtues of a Molecular Biological Approach to the Study of Porcine Transmissible Gastroenteritis Coronavirus. | 18 |
| II. | NUCLEOTIDE SEQUENCE HOMOLOGY BETWEEN THE GENOMES OF THE ANTIGENICALLY UNRELATED PORCINE TRANSMISSIBLE GASTROENTERITIS CORONAVIRUS AND BOVINE CORONAVIRUS AS DETERMINED BY OLIGONUCLEOTIDE FINGERPRINTING | 21 |
| | Introduction | 22 |
| | Materials and Methods | 22 |
| | Results and Discussion. | 23 |
| III. | CLONING AND SEQUENCE ANALYSIS OF THE 3' END OF THE PORCINE TRANSMISSIBLE GASTROENTERITIS CORONAVIRUS GENOME: PRIMARY STRUCTURE OF THE NUCLEOCAPSID PROTEIN GENE | 29 |
| | Introduction | 30 |
| | Materials and Methods | 31 |
| | Virus and Cells | 31 |
| | Purification of Genomic RNA | 31 |
| | cDNA Cloning of the 3' End of the TGEV Genome | 33 |
| | Identification of Large Clones Containing 3'-Specific TGEV Sequences. | 34 |
| | Restriction Endonuclease Mapping | 35 |
| | DNA Sequencing and Sequence Analysis | 35 |
| | Results. | 36 |
| | cDNA Cloning and Sequencing of Two Clones from the 3' End of the Genome | 36 |
| | The Largest Open Reading Frame Predicts a Protein Having Properties Expected of the Nucleocapsid Protein | 38 |
| | A Second Open Reading Frame to the 3' Side of the Nucleocapsid Protein Gene Encodes a Hypothetical Protein of 9,101 Molecular Weight that is Hydrophobic at Both Ends | 41 |

PART PAGE

Discussion. . . . . 43

REFERENCES. . . . . 50

APPENDICES. . . . . 58

APPENDIX A. DETAILED PROCEDURE FOR THE MOLECULAR CLONING OF THE 3' END OF THE PORCINE TRANSMISSIBLE GASTROENTERITIS CORONAVIRUS GENOME . . . . . 59
Keys to Success in cDNA Cloning. . . . . 60
Preparation of TGEV Genomic RNA for cDNA Cloning . . 60
First and Second Strand cDNA Synthesis . . . . . 65
Homopolymer Tailing. . . . . 66
Transformation Conditions. . . . . 68
Identification of Clones Containing TGEV Specific Sequences . . . . . 71
Identification of Clones Containing Poly A Sequences. 77
Identification of 3'-Specific TGEV cDNA Clones. . . 78

APPENDIX B. DETAILED ANALYSIS OF CLONE FG5. . . . . 86

APPENDIX C. HANAHAN'S METHOD FOR TRANSFORMING E. COLI WITH PLASMIDS. . . . . 105

APPENDIX D. PROTOCOL FOR 3' END LABELING WITH REVERSE TRANSCRIPTASE. . . . . 110

VITA. . . . . 113

LIST OF TABLES


| TABLE | | PAGE |
|---|---|---|
| 1.1. | Some key features of the 11 families of animal RNA viruses | 3 |
| 1.2. | Coronavirus species with corresponding host animal and disease | 4 |
| 1.3. | Percentage composition and relative molar concentration of polyadenylated viral RNA species synthesized *in vivo* | 16 |
| 1.4. | Size of coronavirus intracellular RNA species | 17 |
| A.1. | Effect of insert/vector weight ratio on cloning efficiency | 69 |
| A.2. | Effect of total DNA concentration during annealing on cloning efficiency | 70 |
| A.3. | Effect of time of annealing on cloning efficiency. | 72 |
| A.4. | Summary of TGEV-specific cDNA clones and their names. | 85 |
| B.1. | Amino acid composition of TGEV nucleocapsid protein | 95 |
| B.2. | Amino acid composition of the hypothetical hydrophobic protein | 102 |

LIST OF FIGURES


| FIGURE | | PAGE |
|---|---|---|
| 1.1. | Schematic representation of the porcine transmissible gastroenteritis coronavirion | 7 |
| 1.2. | The nested set pattern of subgenomic RNA's found in MHV infected cells. | 9 |
| 1.3. | Replication and transcription scheme used by coronaviruses to synthesize the nested set of mRNA's. | 13 |
| 1.4. | Coronavirus family tree based on antigenic inter-relatedness. | 15 |
| 2.1. | Oligonucleotide fingerprint pattern of TGEV. | 24 |
| 2.2. | Oligonucleotide fingerprint pattern of BCV | 25 |
| 2.3. | Oligonucleotide fingerprint pattern of TGEV and BCV | 26 |
| 2.4. | Schematic diagram of mixed fingerprints of TGEV and BCV. | 28 |
| 3.1. | Restriction endonuclease map and sequencing strategy for TGEV cDNA clones J21 and FG5 | 37 |
| 3.2. | The primary nucleotide sequence of FG5 and deduced amino acid sequences for a portion of the matrix glycoprotein (bases 20-337 in the second reading frame), the nucleocapsid protein (base 353 through 1498 in the second reading frame), and the hypothetical hydrophobic protein (bases 1507 through 1740 in the first reading frame) | 39 |
| 3.3. | Schematic diagram of possible open reading frames obtained when translating the FG5 nucleotide sequence as either virus-sense RNA or virus complement-sense RNA | 40 |
| 3.4. | Plot of hydrophobic region (areas below the line) and hydrophilic regions (areas above the lines) for the TGEV hydrophobic protein | 42 |
| 3.5. | Amino acid sequence homologies among IBV, MHV JHM and TGEV for a 68 amino acid conserved region in the nucleocapsid protein gene | 45 |
| 3.6. | Amino acid sequence homology between the 3' terminal portion of the M (E1) protein of MHV A59 and the deduced M protein of TGEV | 47 |

FIGURE PAGE

A.1. Isopycnic sedimentation of TGEV used for preparation of virion RNA. . . . . 62

A.2. Rate zonal sedimentation of TGEV genomic RNA . . . . . 64

A.3. Proof of first and second strand synthesis. . . . . 67

A.4. Colony hybridization of a collection of white colonies. . 75

A.5. Rapid screening to determine insert size . . . . . 76

A.6. Electrophoresis of putative poly(A)-containing plasmids in preparation for hybridization against $^{32}$P-labeled poly(dT) . . . . . 79

A.7. Southern blot used to identify poly(A)-containing clones . 80

A.8. Electrophoresis of selected TGEV clones in preparation for hybridization against the poly A containing clone J21 . 81

A.9. Southern hybridization of selected TGEV clones with nick-translated purified J21 insert . . . . . 82

A.10. Flow diagram showing efficiency of cloning protocol at key steps . . . . . 83

B.1. Diagram of pUC9 vector used in cloning TGEV, with key features labeled. . . . . 88

B.2. Smith and Birnstiel gel . . . . . 89

B.3. Original restriction map of clone FG5 as determined by the method of Smith and Birnstiel . . . . . 90

B.4. Restriction map of clone FG5 as modified from nucleotide sequence data. . . . . 91

B.5. Schematic diagram of protocol utilized for breakdown of DNA . . . . . 93

B.6. Autoradiograph of a DNA sequencing gel . . . . . 94

B.7. Diagon plot comparing amino acid homologies of TGEV and nucleocapsid proteins . . . . . 96

B.8. Diagon plot comparing amino acid homologies of TGEV and MHV-A59. . . . . 97

FIGURE PAGE

B.9. Amino acid homology shared between TGEV nucleocapsid protein and the M2 matrix protein of human influenza B . 99

B.10. Diagon plot comparing the nucleic acid sequences that code for the nucleocapsid proteins of TGEV and IBV . . . . 100

B.11. Diagon plot comparing the nucleic acid sequences that code for the nucleocapsid protein of TGEV and MHV-A59 . . . 101

B.12. Amino acid homology shared between the hypothetical protein of TGEV and the hypothetical protein of mouse mammary tumor virus . . . . . 103

# PART I

# BACKGROUND ON CORONAVIRUSES

## A. INTRODUCTION

Coronaviruses are one of eleven animal RNA viral families (Table 1.1). The coronaviruses are characterized by large bulbous peplomers attached along the circumference of the virus, giving the virus the appearance of having a "corona" in electron micrographs. They are pleomorphic with a helical nucleocapsid and the total diameter of the virion ranges from 80 to 120 nm (McIntosh, K., 1974). The genomic RNA is single-stranded, nonsegmented, infectious, and polyadenylated (Brian et al., 1980; Guy and Brian, 1979; Lai and Stohlman, 1978; Lai et al., 1981; Wege et al., 1978). They have several properties that are either unique or shared with only a few other viral families. They are one of only three virus families that bud through internal membranes. Coronaviruses and flaviviruses bud through the endoplasmic reticulum (Holmes and Behnke, 1981; Matsumura et al., 1971). Bunyaviruses bud into the golgi (Kuismanen et al., 1982). Coronaviruses are only one of four recognized positive stranded RNA virus families (Baltimore, 1971). Coronaviruses are the only animal RNA viruses that replicate via multiple intracellular (more than two) mRNAs (Stern and Kennedy, 1980a,b; Spaan et al., 1981).

Coronaviruses are responsible for a wide variety of diseases and they exhibit many different tissue tropisms (Table 1.2). Their natural host range includes man, pigs, cattle, dogs, rats, mice, turkeys, chickens, cats, and horses. (Two excellent coronavirus reviews are Siddell et al., 1982 and Wege et al., 1982.) Human strains appear to be responsible for about 20% of all human colds are implicated in human

Table 1.1. Some key features of the 11 families of animal RNA viruses

| Family | RNA Polarity | Number of Genome Segments | Envelope | Site of Budding[a] |
|---|---|---|---|---|
| Picornavirus | + | 1 | No | - |
| Alphavirus | + | 1 | Yes | CM |
| Flavivirus | + | 1 | Yes | ER |
| Coronavirus | + | 1 | Yes | ER |
| Myxovirus | - | 8 | Yes | CM |
| Paramyxovirus | - | 1 | Yes | CM |
| Rhabdovirus | - | 1 | Yes | CM |
| Bunyavirus | - | 3 | Yes | G |
| Arenavirus | - | 3 | Yes | CM |
| Retrovirus | - | 2 | Yes | CM |
| Reovirus | (double stranded) | 10 | No | - |

[a]CM = cytoplasmic membrane, ER = endoplasmic reticulum, G = golgi.

Table 1.2. Coronavirus species with corresonding host animal and disease[a]

| Coronavirus Species | Host Animal | Disease |
|---|---|---|
| Avian infectious bronchitis virus | Chicken | Upper and lower respiratory tract infection |
| Mouse hepatitis virus | Mouse | Hepatitis, encephalitis |
| Transmissible gastro-enteritis virus | Pig | Gastroenteritis |
| Hemagglutinating encephalomyelitis virus | Pig | Encephalomyelitis "wasting and vomiting disease" |
| Human coronavirus (respiratory) | Man | Upper respiratory tract infection |
| Rat coronavirus | Rat | Lower respiratory tract infection |
| Sialodacryadenitis virus | Rat | Sialodacryadenitis |
| Turkey bluecomb disease virus | Turkey | Gastroenteritis "bluecomb disease" |
| Canine coronavirus | Dog | Gastroenteritis |
| Bovine coronavirus | Calf | Gastroenteritis |
| Feline infectious peritonitis virus | Cat | Infectious peritonitis |
| Equine coronavirus | Foal | Gastroenteritis |

[a]Adapted from Wege et al., 1982.

viral gastroenteritis. In addition, they may cause such slowly developing diseases as endemic nephropathogy, (Apostolov, K. et al., 1975) and multiple sclerosis (Burks et al., 1980). The diseases they cause in food animals are of commercial importance because they result in a high incidence of mortality in young animals.

The inherent difficulties in growing coronaviruses to high titers have caused coronavirus research to lag behind that of other virus families.

There are many fundamental questions that need to be answered about coronaviruses and many molecular biological approaches can now be utilized to answer these questions. Some of the questions that still need to be answered are the following: What exactly is the structural composition of the virion? How many genes are there? Is there temporal control of gene expression and if so how is it regulated? What are the mechanisms of persistence? What proteins are involved in immunity? How can coronaviruses be best controlled medically?

## What is Known About Coronavirus Gene Structure and Replication Strategy as Determined from the Three Most Highly Studied Species, the Mouse Hepatitis Coronavirus, the Infectious Bronchitis Coronavirus, and the Porcine Transmissible Gastroenteritis Coronavirus

Our understanding of the basic molecular biology of coronaviruses, compared to many other animal virus families, is very limited at this time. Most of the work on coronaviruses has been carried out using the murine hepatitis virus (MHV) (primarily two different strains), the avian infectious bronchitis virus (IBV) strains, or the porcine transmissible gastroenteritis coronavirus (TGEV).

The virion contains one large (60S) polyadenylated, non-segmented, capped, infectious RNA molecule which by definition is positive stranded (Baltimore, 1974; Brian et al., 1980; Lai and Stohlman, 1978; Lomniczi and Kennedy, 1977; Stern and Kennedy, 1980). The coronavirion contains at least three major structural proteins (Figure 1.1) (Garwes and Peacock, 1975; Brian et al., 1980; Moreau and Brian, unpublished). The virus possesses a 50,000-60,000 dalton phosphoprotein that coats the RNA. It also has two glycoproteins: one of about 200,000 daltons that forms the large bulbous peplomers seen under the electron microscope, and a small 25,000-29,000 dalton transmembrane (matrix) glycoprotein.

Coronaviruses replicate via a unique mode of replication. For MHV, seven (six for IBV and possibly ten for TGEV) viral specific mRNA's are produced in infected cells (Dennis and Brian, 1982; Spaan et al., 1981; Stern and Kennedy, 1890a,b; Wege et al., 1981). The molecular weights of the viral intracellular polyadenylated RNA species add up to a sum that far exceeds the molecular weight of the virion RNA indicating that there must be redundant sequences present in the messenger RNA's. In a classic experiment first done with IBV (Stern and Kennedy, 1980a,b) and later with MHV (Spaan et al., 1982; Wege et al., 1981) it was demonstrated that the viral intracellular mRNA's formed a "nested" set. mRNA's in vivo labeled with $^{32}P$ were isolated and analyzed by two-dimensional T1 fingerprinting. All of the RNA's were related. The oligonucleotides found in each mRNA were also found in the next larger mRNA species indicating that each mRNA species contained all of the genetic content of the next smaller RNA plus some

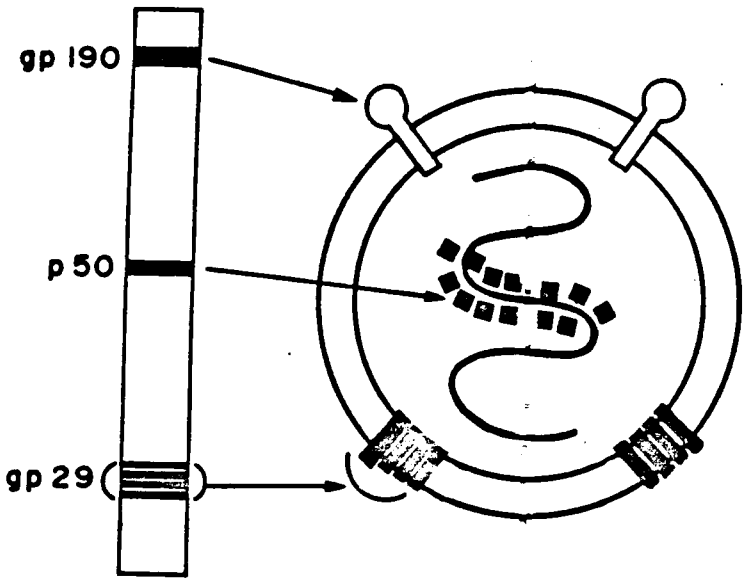



Figure 1.1. Schematic representation of the porcine transmissible gastroenteritis coronavirion. The data for this model come from the studies of Garwes and Pocock (Garwes and Pocock, 1975), Brian et al. (Brian et al., 1980) and Moreau (Moreau, Master's Thesis, University of Tennessee, 1982). The four major molecular components are (1) a 200 Kd peplomeric glycoprotein, (2) a 50 Kd phosphorylated nucleocapsid protein, (3) a 29 Kd transmembrane (matrix) glycoprotein, and, (4) a 7 Md (20 Kb) single-stranded, nonsegmented, 5' capped, 3' polyadenylated, infectious RNA genome.

additional information. Furthermore, the experiments demonstrated that each mRNA shared a common 3' end. The nested set pattern is shown in Figure 1.2.

The subgenomic RNA's are made in different molar ratios and in constant relative amounts throughout the replication cycle (Leibowitz et al., 1981; Spaan et al., 1981; Stern and Kennedy, 1980a; Wege et al., 1981). In vitro translational studies with some of the subgenomic RNA species have shown that the RNA's behave as if they are monocistronic. Despite several genes being represented in a subgenomic RNA, only the gene at the immediate 5' terminus is translated (Leibowitz and Weiss, 1981; Leibowitz et al., 1982; Rottier et al., 1981; Siddell et al., 1980). This pattern of replication apparently holds true for TGEV as well because of the intracellular poly $A^+$ RNA pattern (Dennis and Brian, 1982).

Two virally encoded RNA-dependent RNA polymerase activities are detected within MHV-infected cells (Brayton et al., 1982, 1984). One can be detected by one hour post-infection the other by six hours post-infection. They are distinguished by their cation requirement and pH optimum. The one detected by one hour post-infection copies the parental positive stranded RNA genome into a full length negative strand. The polymerase activity that appears by six hours post-infection copies the negative strand into the multiple positive stranded mRNA's in addition to making full length genomic RNA to be packaged into the virion (Brayton et al., 1982; Brayton et al., 1984; Dennis and Brian, 1982).

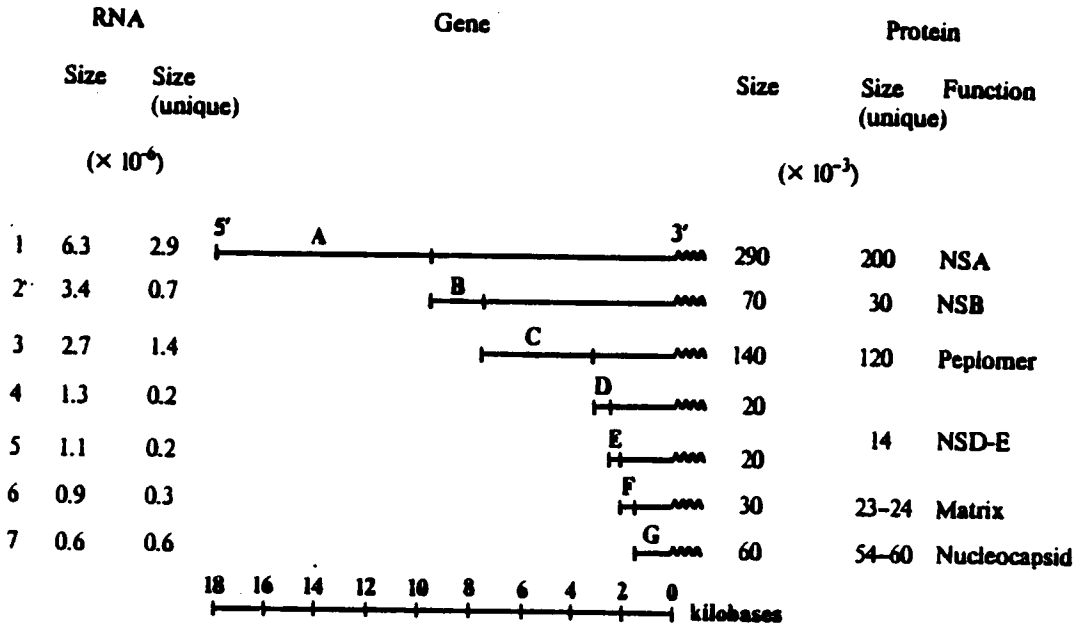



Figure 1.2. The nested set pattern of subgenomic RNA's found in MHV-infected cells. The translational products for RNA species 1, 2, 3, 4, 5, 6 and 7 are nonstructural protein A (NSA), nonstructural protein B (NSB), peplomeric protein, nonstructural protein D (NSD), nonstructural protein E (NSE), matrix protein, and nucleocapsid protein, respectively.

T1 oligonucleotide fingerprints of MHV intracellular mRNA's suggest a novel mode of replication for coronaviruses. In addition to the 3' coterminal nested set pattern for the mRNA's, evidence accumulated that supported the idea that all of the mRNA's had a common 5' terminal sequence as well. During T1 oligonucleotide fingerprinting experimentation, it was discovered that one spot, oligonucleotide number ten, did not map within the 3' terminal nested set (Lai et al., 1981). Oligonucleotide ten and the sequences around oligonucleotide ten consistently mapped at the 5' end of each mRNA (Lai et al., 1981; Lai et al., 1983). This suggested each mRNA contained a leader segment shared in common with all of the viral mRNA's. Oligonucleotide nineteen which contains the junction between the leader and main body of the mRNA (Lai et al., 1982) was sequenced and an oligomer made to a portion of it. This was used as a primer to sequence back into the leader itself. Leader coding segments are located only at the 5' end of the genomic RNA. Free leader sequences can be detected in infected cells (Baric et al., 1985). The leader sequences range in size from seventy to eighty-two nucleotides in length. Larger RNA's containing leader sequences are also detected (Baric et al., 1985). Ultraviolet transcriptional mapping using MHV or IBV reveals the target size of the template used for each mRNA is proportional to the size of the mRNA. The current, most favored model is that leader RNA's are transcribed from the 3' end of the minus strand, become detached as part of a leader-polymerase complex, and reattach at an "intergenic" sequence downstream on the minus strand, and serve as a primer for continued transcription (Figure 1.3; Lai et al., 1984).

Recent evidence has been obtained for high frequency of recombination occurring in coronaviruses (Keck et al., 1985). Using two distinguishable temperature sensitive mutants of MHV, cells were coinfected at the non-permissive temperature, and wild type phenotypes of the virus were recovered and the messages oligonucleotide fingerprinted. The proposed mechanism for coronavirus recombination is as follows. The polymerase transcribes the negative strand and pauses at certain places on the template. The polymerase and transcript can dissociate from the template and reassociate with either the same template or a different template and continue copying the viral genome. In this manner the polymerase could copy only sequences coding for the wildtype phenotype off of two separate templates. Thus, the two models, one for mechanism of leader-primed transcription, and one for the mechanism of high frequency recombination, are compatible.

While priming is not unique to viral systems, the proposed mechanism for coronaviruses is unique to viral systems. Influenza, the first discovered case of viral priming, is a negative stranded virus. The virus packages the polymerase and enzymes it needs to replicate its RNA. Influenza viruses use newly synthesized RNA II transcripts from the cell to prime influenza directed RNA synthesis (Herz et al., 1981). Influenza virus replication is sensitive to alpha amanitin, a potent inhibitor of polymerase II, and requires continuous functioning of this enzyme. A viral endonuclease cuts the cellular messages ten to thirteen bases from the cap structure (Plotch et al., 1981). The capped RNA fragments are covalently linked to a viral protein. There are no leader sequence requirements for hydrogen bonding between the

scavengered leader sequence and the 3' end of the influenza genome, only the terminal A presumably forms a hydrogen bond with the influenza terminal U. It was recently discovered that bunyavirus, also negative stranded, uses cellular capped RNA's; however, unlike influenza, priming occurs in the cytoplasm and does not require continuous polymerase II activity (Patterson et al., 1984). In contrast, coronaviruses code for their own primer, the leader sequence, which can then interact with the intergenic sequences by hybridization. The intergenic sequences of the first three genes located at the 3' end of the MHV genome are complimentary to an eleven base sequence found in the leader sequence (Budziloqicz et al., 1985).

The model for the replication strategy used by the coronaviruses is summarized in Figure 1.3. The infecting parental strand of RNA is first translated to code for a viral polymerase. This polymerase then copies the positive strand to yield a full length negative stranded RNA. At some point a second polymerase is produced. The second polymerase then makes many copies of the leader sequence which can subsequently hybridize with intergenic sequences, probably as a polymerase-leader complex, and prime both mRNA synthesis and progeny full length positive strands for encapsidation. Many questions remain unanswered, however. How are the first and second polymerase activities related? Is the same protein responsible for both activities? Are the two activities the result of a modification of one protein? If so, what is the modification? What controls the level of mRNA synthesis? Why are some mRNA's made in twenty times the molar amount of other mRNA's? What is the mechanism of recognition between

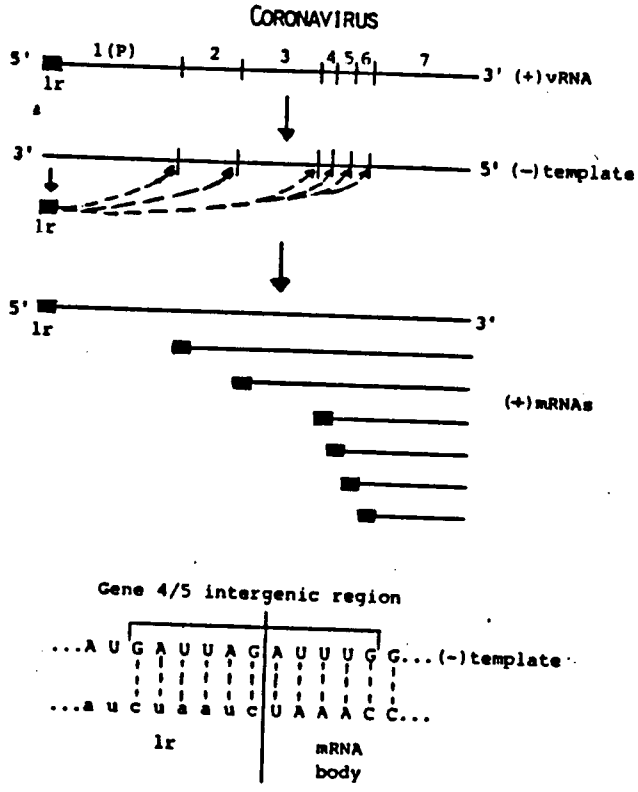



Figure 1.3. Replication and transcription scheme used by coronaviruses to synthesize a nested set of mRNA's. An early polymerase copies genomic "+" strand to full length "-" strands. A second polymerase then copies + strands off of the minus template. A leader (solid box) is synthesized from the 3' end of the negative strand of genomic RNA. This leader sequence dissociates with its polymerase and reassociates at other sites on the negative strand template priming continued synthesis of an mRNA. The sequence shown at the bottom of the figure shows the intergenic sequence found between genes 4 and 5 of the mouse hepatitis coronavirus. See text for discussion.

the polymerase and RNA template? Is leader RNA (which serves as primer) always a part of a polymerase complex?

The Porcine Transmissible Gastroenteritis Coronavirus Possesses Some Unique Features That Make It Worthy of Detailed Study

TGEV is already the most characterized member of a third extensively diverged group of the coronavirus family (Figure 1.4) and should be further characterized for an ultimate molecular understanding of coronavirus evolution. Coronaviruses are a heterogeneous family from a structural point of view. Antigenic studies alone show them to be diverse. On the basis of antigenic relatedness, determined by cross neutralization and cross immunofluorescence, antigenic differences among coronaviruses can be documented and a coronavirus family tree can be drawn (Figure 1.4 taken from the review by Brian et al., 1983). TGEV is closely related to the feline, canine, and human 229E coronaviruses, but none of these has been studied in the same detail as TGEV, nor have they yet been proven feasible for molecular studies.

TGEV has a larger number (ten) of intracellular polyadenylated mRNA species than either MHV (seven) or IBV (six) and may, therefore, code for additional proteins (Tables 1.3 and 1.4, from Dennis and Brian, 1982). Only a detailed study can answer this question.

Lastly, TGEV is medically important and a detailed understanding of its molecular structure may lead to a superior vaccine of recombinant DNA design. This is discussed in the next section below.

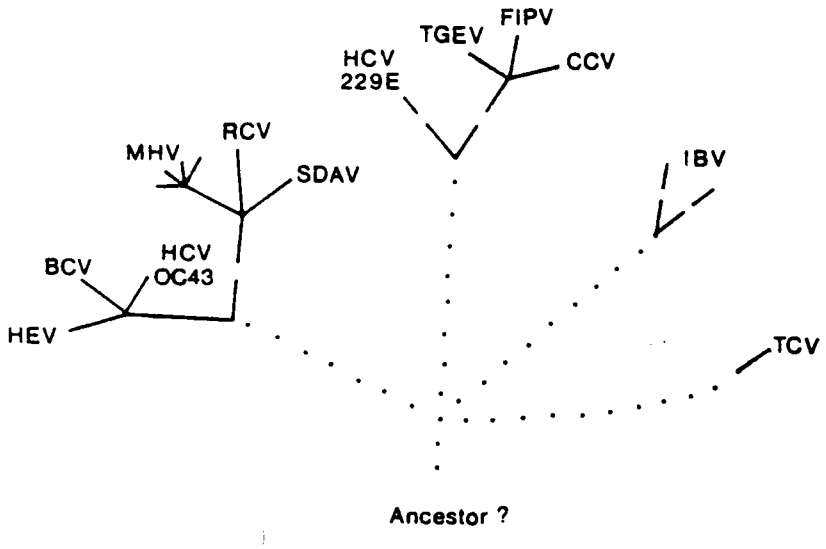



Figure 1.4. Coronavirus family tree based on antigenic interrelatedness. Data for this figure comes primarily from the review by Wege et al. (Wege et al., 1982). Solid lines indicate evidence for cross neutralizing activity; dashed lines indicate immunological evidence (other than neutralization) for 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 sialodacrytoadenitis 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).

Table 1.3. Percentage composition and relative molar concentration of polyadenylated viral RNA species synthesized *in vivo*[a]

| RNA species (mol wt x $10^{-6}$) | Percentage composition[b] | Relative molar concn. |
|---|---|---|
| 6.8 | 28.3 ± 2.5 | 2.6 |
| 6.2 | 9.8 ± 1.7 | 1 |
| 3.15 | 8.5 ± 3.8 | 1.7 |
| 1.40 | 5.6 ± 1.0 | 2.5 |
| 1.05 | 3.1 ± 0.1 | 1.9 |
| 0.94 | 9.0 ± 0.6 | 6.1 |
| 0.66 | 27.2 ± 2.5 | 26.1 |
| 0.39 | 1.3 ± 0.2 | 2.1 |
| 0.34 | 2.2 ± 0.8 | 4.1 |
| 0.24 | 5.0 ± 2.8 | 13.2 |

[a]Radiolabeled, polyadenylated viral RNA was extracted from infected cells and analyzed by denaturing gel electrophoresis as described in Dennis and Brian, 1980.

[b]The percentage composition was determined from densitometric tracings of fluorograms from three separate experiments. Heterogeneous radioactive material within the gel was excluded from the quantitative analysis.

Table 1.4. Size of coronavirus intracellular RNA species

| Virus | RNA species (mol wt x $10^{-6}$) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| TGEV[a] | 6.8, | 6.2, | 3.15, | 1.40, | 1.05, | 0.94, | 0.66, | 0.39, | 0.34, | 0.24 |
| Mouse hepatitis virus[b] | 5.6, | 4.0, | 3.0, | 1.4, | 1.2, | 0.9, | 0.8, | --, | --, | -- |
| Avian infectious bronchitis virus[c] | 5.9, | --, | 2.6, | 1.5, | 1.3, | 0.9, | 0.6 | --, | --, | -- |

[a]From Dennis and Brian, 1982; Spaan et al., 1981; Stern and Kennedy, 1980.

[b]From Spaan et al.

[c]From Stern and Kennedy.

## The Virtues of a Molecular Biological Approach to the Study of Porcine Transmissible Gastroenteritis Coronavirus

The advent of recombinant DNA technology has radically altered our ability to understand events on the molecular level. As a consequence of this, it has allowed us to control events at the molecular level, especially at the level of molecular genetics, to control the outcome of gene expression. This ability has expanded the entire spectrum of scientific ingenuity and opened many potential avenues for biomedical investigation. The commercial availability of restriction enzymes and other specialized reagents such as cloning vectors, the establishment of data banks for protein and nucleic acid sequences, and the drop in cost of microcomputers such that virtually every lab can have access to one, have all helped in the revolution.

The motivation for much of this Ph.D. dissertation comes from the anticipated application of recombinant DNA technology to the ultimate understanding and control of coronavirus infections. In essence, cDNA cloning and nucleic acid sequencing was chosen as an experimental approach for studying TGEV for the following four reasons. 1. Cloning and sequencing is the most rigorous means available to determine the exact size of the virus genes, their number, and location on the genome. 2. Clones developed from the project can be used to examine replication events in the cell. 3. Clones developed have a potential use in diagnosis of TGEV infections or monitoring viral persistence in the animal. 4. Clones have a potential usefulness in making a vaccine through the use of recombinant DNA technology.

Precedence exists for success in each one of these categories from other virus families. For example in the first of these categories, with the use of cloned sequences of the respiratory syncytial virus RNA, it has been possible to identify ten viral genes, when heretofore, only 8 viral genes could be identified from a direct analysis of viral proteins or genetic complementation groups (Collins et al., 1983). Also, a study of Pinchinde virus, an arenavirus, has lead to the striking discovery of ambisense RNA sequences (Auprin et al., 1984). Ambisense sequences are those that code for more than one protein off of complementary strands of the same RNA sequence. In the case of Pinchinde virus, one protein is translated from the virion-sense strand and one protein is translated from the virion-complementary-sense strand of the S (small) RNA. Thus, the virion sense strand and the complimentary virion sense strands both act as messages. In addition, numerous examples of overlapping reading frames have been found among RNA viruses using recombinant DNA techniques (Bishop et al., 1982).

Most immediately, cloned TGEV sequences will be useful in the study of coronavirus replication. Cloning and sequencing the entire genome of TGEV will enable us to: 1. Look for open reading frames on both viral sense strand and complimentary viral sense strand to locate all potential genes coded for by TGEV. 2. Develop hybridization probes for specific virus mRNAs. 3. Look for possible RNA sequences that may play an important control role in replication of TGEV. Additionally, cloned sequences will be immediately useful in studying coronavirus evolutionary relationships (i.e. homologous sequences

between TGEV and other coronaviruses) and developing clinically useful diagnostic hybridization probes.

Two potential strategies for the use of recombinant DNA technology in the long-term control virus infections are the following. (1) Viruses can be specifically engineered to destroy their pathogenicity. That is, an avirulent virus can be made and used for a vaccine. Relevant to this is the finding that increased neurovirulence of the Sabine vaccine is associated with a single nucleotide change in the non-coding region of the type 3 polio genome (Evans et al., 1985). It is possible to clone virus structural genes into appropriate vectors and express only one portion of a given protein to stimulate immunity without the chance of infecting the host with the parent pathogen. Such recombinant vaccines are becoming more frequent (for review see Lasky, 1984). (2) Chemotherapeutic approaches could be designed that interfere with virus replication. Eventually it may be useful to develop recombinant constructs that interfere with virus replication once the infection is underway to yield a protective effect. For example, anti-sense RNA may possibly be made in constituitive or inducible expression vectors (Izant and Weintraub, 1984, 1985) and thereby turn off specific genes. Control of virus infections may therefore become therapeutic (clinical treatment) as well as prophylactic (prevention by vaccination).

# PART II

# NUCLEOTIDE SEQUENCE HOMOLOGY BETWEEN THE GENOMES OF THE ANTIGENICALLY UNRELATED PORCINE TRANSMISSIBLE GASTROENTERITIS CORONAVIRUS AND BOVINE CORONAVIRUS AS DETERMINED BY OLIGONUCLEOTIDE FINGERPRINTING

## A. INTRODUCTION

The porcine transmissible gastroenteritis (TGEV) and the bovine enteric coronavirus (BCV) are both members of the coronavirus family. Both cause severe gastroenteritis in their normal host. Each belongs to an antigenically distinct subgroup of the coronavirus family (Pedersen et al., 1978) based on immunofluorescence and cross neutralization data. A study of nucleotide sequence homology was undertaken to determine the degree of relatedness between the two viruses. The degree of nucleotide sequence homology was determined from 2-dimensional electrophoresis of RNase-$T_1$ generated oligonucleotides. The theoretical basis for determining sequence homologies by this method is described by Aaronson et al. (Aaronson et al., 1982).

## B. MATERIALS AND METHODS

The Purdue strain of TGEV was plaque purified and stocks were prepared as previously described (Brian et al., 1980). TGEV was grown on confluent monolayers of ST cells in 150 $cm^2$ flasks (Corning). Cells were infected with TGEV at a multiplicity of approximately 5 PFU/cell and after 1 hour adsorption, cells were rinsed with Earle balanced salt solution and refed with phosphate-free minimum essential medium containing 20% fetal calf serum and 50 uCi of $^{32}P$-orthophosphate per ml. Virus was harvested at 18 hpi when 75 to 100% of the cells demonstrated a cytopathic effect.

The Mebus strain of BCV was grown and viral RNA was $^{32}P$-labeled as previously described (Lapps and Brian, 1985).

Supernatant fluids containing labeled virus were clarified by centrifugation at 10,000 x g for 10 min. and virus was concentrated by sedimentation at 90,000 x g for 1.5 h at 4°C onto a 60% sucrose (wt/wt) cushion. Virus was then sedimented isopynically in 18 ml linear gradients 20-60% sucrose (wt/wt) at 90,000 x g for 4 h at 4°C. Virus was located by radioactivity, diluted with an equal volume of TMEN (10 mM Tris maleate [pH 6.0]--100 mM NaCl--1 mM EDTA), and was pelleted at 150,000 x g for two hours in a Sorvall AH650 rotor. Sucrose solutions were made in TMEN.

Genomic 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)--1% SDS--0.5 mg proteinase K per ml, incubating for 0.5 h at 37°C, and extracting twice with an equal volume of a mixture of 48% phenol--48% chloroform--2% isoamyl alcohol. Sodium acetate (0.1 volume) and 100% ethanol (2 volumes) was added to the RNA and the RNA allowed to precipitate at -80°C.

RNA T1 oligonucleotide fingerprint analysis was performed according to the method of DeWachter and Fiers (DeWachter and Fiers, 1972) as modified by Clewley et al. (Clewley et al., 1977).

## C. RESULTS AND DISCUSSION

Figure 2.1 and Figure 2.2 shows the two dimensional fingerprint pattern produced by TGEV and BCV genomic RNA respectively. Figure 2.3 shows a composite pattern that is produced when the two samples are




Figure 2.1. Oligonucleotide fingerprint pattern of TGEV.




Figure 2.2. Oligonucleotide fingerprint pattern of BCV.




Figure 2.3. Oligonucleotide fingerprint pattern of TGEV and BCV mix.

mixed and electrophoresed on the same gel. Figure 2.4 is a schematic diagram showing the oligonucleotides that are unique to each virus and the ones that migrate in common. There are 33 well resolved oligonucleotides that can be identified for each TGEV and BCV. Of these, 10 are common to both.

The ten oligonucleotides that are in common represent a comigration of 30% of the large oligonucleotides. Assuming that sequence homology demonstrated in the large oligonucleotides is representative of the remaining genome, we can conclude that there is a nucleotide sequence homology of approximately 91%. This is based on a theoretical analysis of T1 oligonucleotide fingerprinting by Aaronson et al. (Aaronson et al., 1982). The sensitivity of T1 oligonucleotide fingerprinting was examined in great detail by Aaronson et al. using a computer simulation of the technique. They showed that a 1%, 5%, and 10% variation in the sequences were the result of two sequences sharing only 85%, 50%, or 25% of their large oligonucleotides respectively.

A genome sequence homology of 91% between TGEV and BCV supports the notion that the two viruses, while antigenically very dissimilar, arose from a common ancestor.

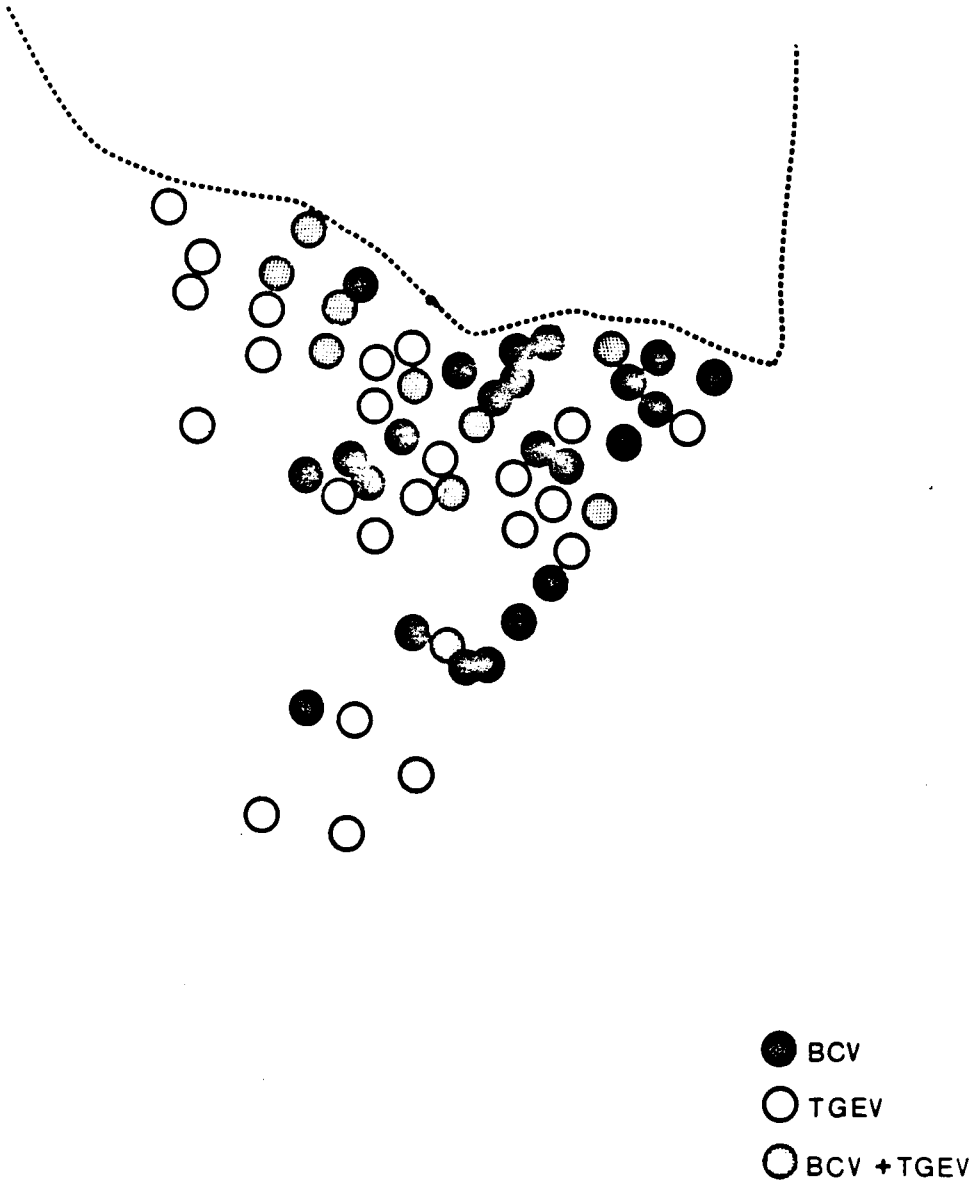



Figure 2.4. Schematic diagram of mixed fingerprints of TGEV and BCV.

# PART III

## CLONING AND SEQUENCE ANALYSIS OF THE 3' END OF THE PORCINE TRANSMISSIBLE GASTROENTERITIS CORONAVIRUS GENOME: PRIMARY STRUCTURE OF THE NUCLEOCAPSID PROTEIN GENE

## A. INTRODUCTION

The genome of the porcine transmissible gastroenteritis coronavirus (TGEV) has been shown to be a single-stranded, non-segmented, polyadenylated, infectious RNA molecule of $6.8 \times 10^6$ mol. wt. or approximately 20 Kb in length (Brian, et al., 1980). The total number of genes encoded by the TGEV genome, however, has not yet been determined. The genome codes for at least four unique polypeptides on the basis of existing protein data. The virion is comprised of three major structural proteins: a 200 Kd peplomeric glycoprotein, a 29 Kd membrane-associated matrix glycoprotein, and an internal phosphorylated nucleocapsid protein that measures from 46 to 50 Kd (Garwes and Pocock, 1975; Moreau and Brian, unpublished). These alone would account for only approximately 8.4 Kb of coding information. In addition, the virus synthesizes at least one non-structural protein during its replication, an RNA-dependent RNA polymerase, the size of which is not yet known (Dennis and Brian, 1982). During replication, TGEV produces 9 species of subgenome-size polyadenylated RNA molecules (Dennis and Brian, 1982). Assuming these molecules lie in a 3' co-terminal "nested set" arrangement, as described for the mRNAs of mouse hepatitis virus (MHV; Lai et al., 1981; Leibowitz et al., 1982; Rottier et al., 1981) and the avian infectious bronchitis coronavirus (IBV; Stern and Kennedy, 1980), and assuming each functions as a monocistronic mRNA, then TGEV may code for as many as 10 different protein species.

One powerful approach for determining the number of potential genes in an RNA virus genome is to examine the primary nucleotide structure and deduce the identity of genes from an examination of open reading frames. In this paper we describe experiments that begin to examine the TGEV genome by this approach. cDNA clones were prepared from the 3' end of the polyadenylated genome and were sequenced. Two potential genes were identified within the first 1663 bases. One gene encodes a protein of 382 amino acids which is the apparent nucleocapsid protein on the basis of size, chemical properties, and significant amino acid sequence homology with other coronavirus nucleocapsid proteins. This gene is flanked on its immediate 5' side by at least part of the matrix protein gene. The second gene encodes a hypothetical protein of 78 amino acids that is hydrophobic at both ends. A 3' non-coding sequence of 276 bases sharing a 10 base conserved sequence near the poly A tail with other coronaviruses was also determined.

## B. MATERIALS AND METHODS

### Virus and Cells

The Purdue strain of porcine transmissible gastroenteritis coronavirus (TGEV) was plaque purified and grown on the swine testicle (ST) cell line as previously described (Brian et al., 1980).

### Purification of Genomic RNA

Virus was purified from clarified supernatant fluids as previously described (Brian et al., 1980) except that all sucrose solutions were

made up in TMEN (10 mM Tris maleate [pH 6.0]--100 mM NaCl--1 mM EDTA). Viral RNA in one out of ten flasks was radiolabeled in order to follow the purification of the RNA. For these flasks, infected cells were refed with phosphate-free medium containing 1% fetal calf serum and 400 uCi $^{32}$P-orthophosphate (ICN) per ml. 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)--1% SDS--0.5 mg proteinase K per ml, incubating for 0.5 h at 37°C, and extracting twice with an equal volume of a mixture of 48% phenol--48% chloroform--2% isoamyl alcohol. RNA was precipitated at -80°C after adding 0.1 volume of 2 M sodium acetate and 2 volumes of 100% ethanol. Because small molecular weight RNA species are found in some preparations of purified coronavirion RNA, full-length genomic RNA to be used for cDNA cloning and making probe for colony screening was selected by rate-zonal sedimentation on preformed linear gradients of 15 to 30% sucrose (wt/wt) made up in TNE--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 Cerenkov radiation. Only RNA sedimenting with a sedimentation coefficient of 50S or greater, as determined by reference to sedimentation of mammalian 28S and 18S ribosomal RNA in a parallel gradient, was recovered by ethanol precipitation and used in the experiments described below.

cDNA Cloning of the 3' End of the TGEV Genome

TGEV 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 10 mM Tris hydrochloride (pH 8.3)--10 mM $MgCl_2$--10 mM DTT--2 mM dCTP--2 mM dTTP--2 mM dATP--2 mM dCTP--0.1 uCi [$^{32}$P] dCTP (3000 Ci/mmole, ICN)--50 pmoles oligo $dT^{12-18}$--6 ug TGEV RNA--30 U RNAsin--10 U reverse transcriptase, for 1 hour at 42°C, and the reaction was stopped by adding 2 ul 0.5 M EDTA. Nucleic acids were phenol extracted and the ethanol precipitated after the addition of 0.5 volume of 7.5 M ammonium acetate.

Second strand synthesis was carried out in a reaction volume of 100 ul containing 20 mM Tris hydrochloride (pH 7.5)--5 mM $MgCl_2$--10 mM $(NH_4)_2SO_4$--100 mM KCl--0.15 mM NAD--50 mg/ml BSA--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 the first strand reaction. The reaction was incubated at 12°C for 1 hour, then at 22°C for 1 hour. The reaction was stopped by adding 4 ul 0.5 M EDTA and reaction products were phenol extracted and fractionated on a Sephadex G50 column (80 mm x 0.25 mm). 0.5 ml fractions were collected and ds cDNA was pooled and ethanol precipitated.

Double stranded cDNA was homopolymer tailed essentially by the method of Roychoudhury and Wu (Roychoudhury and Wu, 1980). The following were added together, in order: 3 ul $dCTP^{32}$ (>800 Ci/mmole), 20 ul 10X cacodylate buffer (1.4 mK cacodylate 0.3 M Tris hydrochloride [pH 7.6], 4 ul 5 mM DTT, 3 ul 10 mM dCTP, 2 ul 100 mM $CoCl_2$, 1.5 ul terminal deoxynucleotide transferase (at least 8

units/pmole 3' end), $H_2O$ to 200 ul final volume. The reaction was carried out at 12°C for 1.5 min then stopped by adding 10 ul 0.5 M EDTA. This reaction resulted in an average of 15 cytidine residues added per 3' end of dsDNA, the optimal number for annealing and transformation (Peacock, 1981).

C-tailed ds cDNA was annealed to G-tailed, Pst-1-linearized pUC9 vector (PL Biochemicals) for 4 h at 58°C in a 50 ul volume of buffer containing 10 mM Tris hydrochloride (pH 7.5)--150 mM NaCl--1 mM EDTA. The total concentration of DNA used was less than 0.5 ug/ml and the optimal insert:vector ratio was 1:1 on a mass basis. E. coli strain JM103 was transformed using the method of Hanahan (Hanahan, 1983). Cells containing inserts 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, HAWP) and probed with random-primed cDNA copied from TGEV genomic RNA.

Identification of Large Clones Containing 3'-Specific TGEV Sequences

$^{32}P$-labeled, random primed cDNA used for colony hybridization was synthesized as described above for the oligo(dT)-primed reaction except that oligo(dT) was replaced by 50 pg of sheared calf thymus DNA. Probe was alkali treated to hydrolyze the RNA and then used for colony hybridization as described by Maniatis (Maniatis, 1982). Colonies yielding a strong signal were analyzed for insert size by electrophoresis in agarose gels (Kado and Liu, 1982). Inserts of 0.2 to 2.0 Kb (the largest) were further analyzed by Southern hybridization

with $^{32}$P-labeled poly(dT) to detect poly(dA) content and by cross hybridization with nick-translated inserts to detect overlapping sequences. $^{32}$P-labeled poly(dT) probe was prepared as described above for the oligo(dT)-primed reaction 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 lysozyme lysis and cesium chloride centrifugation as described by Maniatis (Maniatis, 1982), and restriction endonuclease mapping was done essentially as described by Smith and Bernstiel (Smith and Bernstiel, 1979) 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 obtained as part of the Beckman Microgenie program, March 1985 version (Beckman Instruments, Inc.).

## C. RESULTS

### cDNA Cloning and Sequencing of Two Clones From the 3' End of the Genome

Starting material for cDNA cloning was approximately 6 ug of rate-zonally purified genomic RNA obtained from 400 ml of tissue culture medium. An estimated 200 ng of ds cDNA was obtained, as determined by radiolabel incorporation during second strand synthesis, and from this approximately 1500 white colonies were obtained. By colony screening 200 colonies gave a strong signal to $^{32}$P-labeled cDNA prepared from genomic RNA, and of these, 13 had inserts of 200 to 2000 bases as determined by agarose gel electrophoresis of supercoiled plasmids, and were further analyzed by restriction enzyme analysis and poly A content. The largest clone, FG5, did not react under stringent hybridization conditions by Southern blotting to $^{32}$P-labeled oligo(dT), but did cross hybridize in Southern blot analysis with several other smaller clones that did react strongly with oligo(dT). One of these, J21, a clone of 700 bases, was sequenced in part to determine the primary structure of the extreme 3' end of the genome.

The orientation of clones FG5 and J21 with respect to the virus genome and the restriction enzyme sites used for sequencing are illustrated in Figure 3.1. This orientation assumes polyadenylation at only the 3' end of the genome, and is based on the precedent of the documented 3' polyadenylation site in the avian infectious bronchitis virus and mouse hepatitis virus genomes (Stern and Kennedy, 1980; Lai et al., 1981). The strategy used for sequencing is described in the legend to Figure 3.1. Over 96% of the sequence containing the two

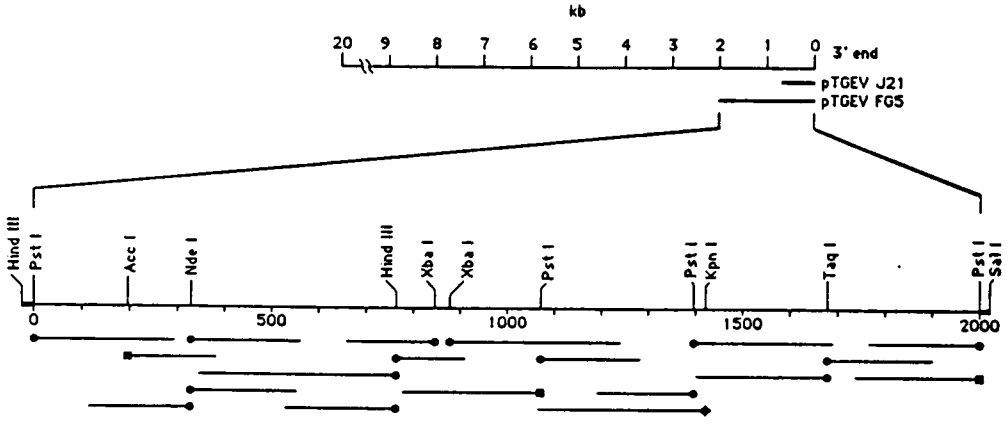



Figure 3.1. Restriction endonuclease map and sequencing strategy for TGEV cDNA clones J21 and FG5. The internal Hind III, Taq I, Pst I, and Xba I sites, derived by restriction endonuclease mapping, and the Hind III and Sal I sites in the multiple cloning region of the pUC9 vector, were the sites used for initial DNA sequencing. Internal Acc I, Kpn I and Nde I sites were identified from sequence data and were used to complete the sequencing. ■ indicates sites labeled at the 5' end using polynucleotide kinase. ● indicates sites labeled at the 3' end using reverse transcriptase and the appropriately labeled deoxynucleotide triphosphate. ◆ indicates site labeled at the 3' end using dideoxy A and terminal transferase.

complete genes we report was determined either by sequencing both strands or by repeated sequencing of the same strand using different methods of end labeling. Some of the sequences were derived from subclones of FG5 made from the Pst I restriction sites.

The total sequence of FG5 is illustrated in Figure 3.2. Sequences from J21 that overlap with FG5 are identical to those of FG5 except that the total length of the polyadenylated tail is 15 for the J21 clone, and 6 for the FG5 clone.

The entire nucleotide sequence was translated in all possible reading frames and only translation of the virion-sense strand revealed open reading frames of greater than 120 bases that are preceded by a termination codon and contain an appropriate initiator methionine codon (Figure 3.3). Of these, only the largest two open reading frames are evaluated below.

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

The largest open reading frame extends from base 353 to base 1498 and predicts a 382 amino acid protein of 43,426 molecular weight. The only TGEV structural or nonstructural protein described to date that approaches this size is the phosphorylated nucleocapsid protein that measures 46 to 50 Kd by SDS polyacrylamide electrophoresis (Garwes and Pocock, 1975; Moreau and Brian, unpublished). The protein has two properties that are strikingly similar to the nucleocapsid proteins of MHV and IBV (Armstrong et al., 1983; Skinner and Siddell, 1983; Boursnell et al., 1985). Firstly, it is rich in serine. Thirty-nine (10%) of the residues are serine making it the most abundant amino

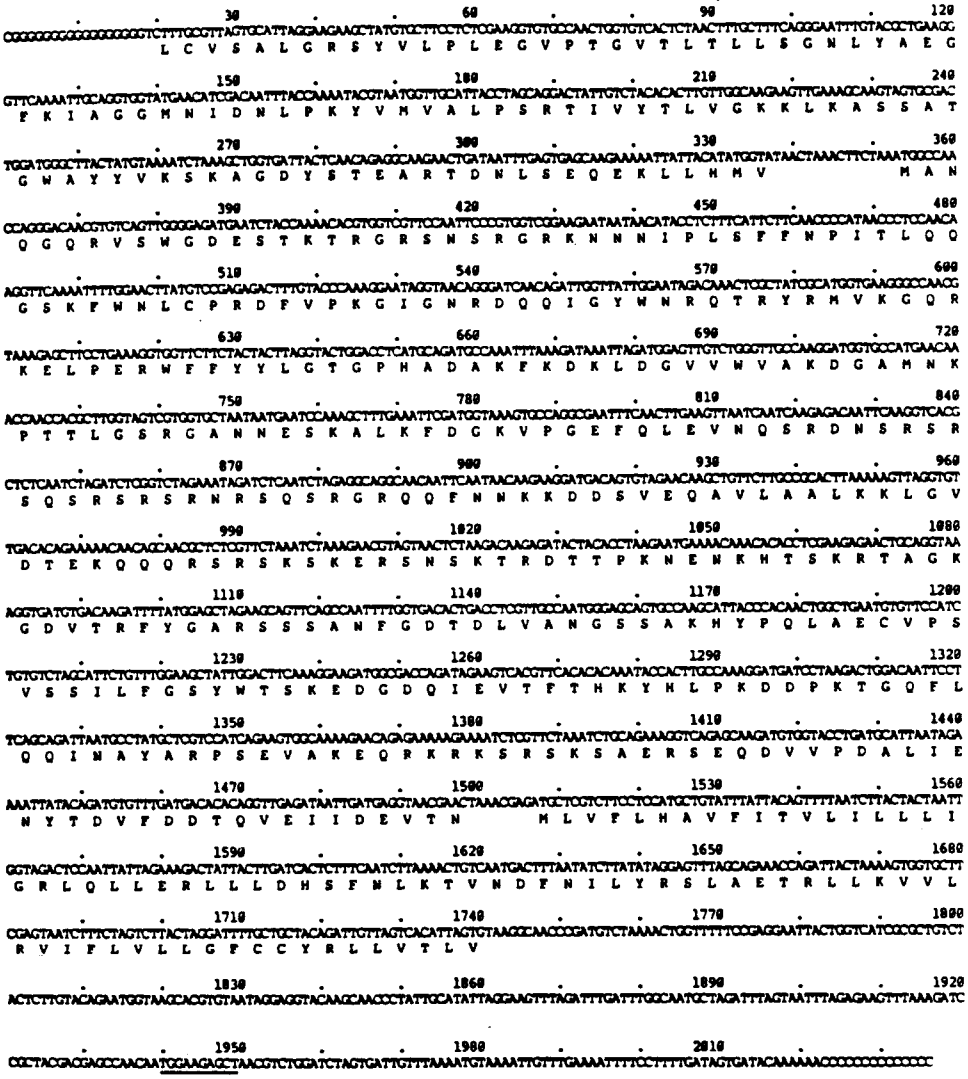


Figure 3.2. The primary nucleotide sequence of clone FG5 and the deduced amino acid sequences for a portion of the matrix glycoprotein (bases 20 through 337 in the second reading frame), the nucleocapsid protein (bases 353 through 1498 in the second reading frame), and the hypothetical hydrophobic protein (bases 1507 through 1740 in the first reading frame). A 10 base sequence highly conserved among coronaviruses is underlined beginning at base 1940.

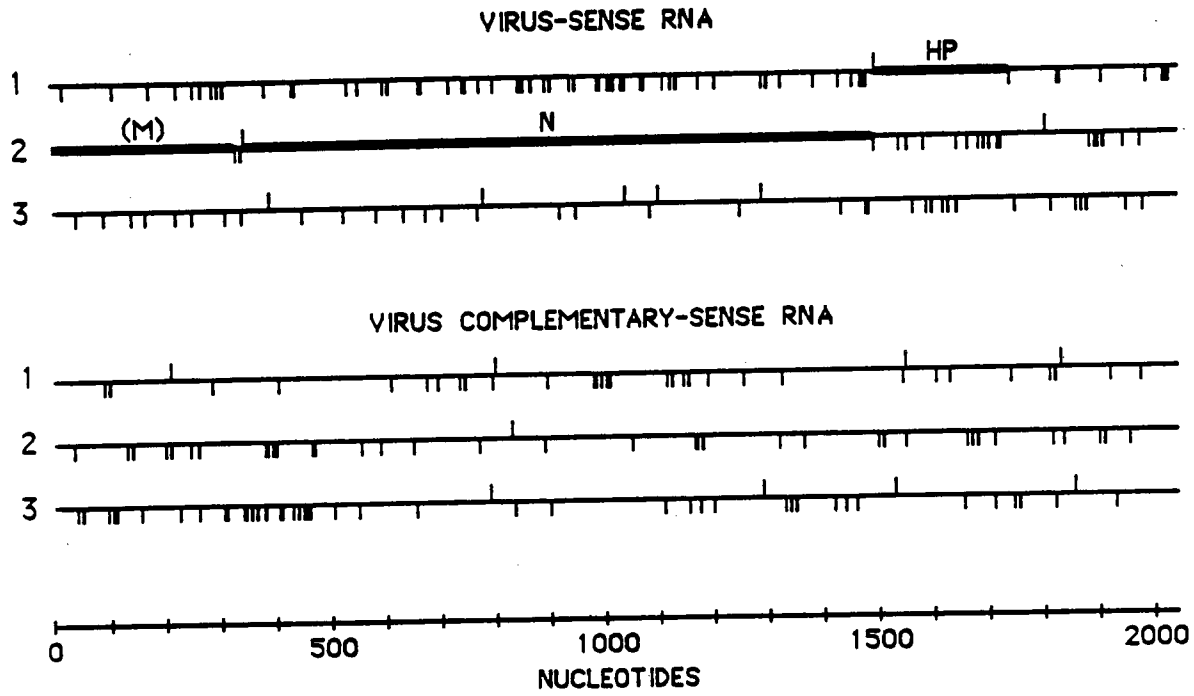



Figure 3.3. Schematic diagram of possible open reading frames obtained when translating the FG5 nucleotide sequence as either virus-sense (ie, virion-sense) RNA or virus complementary-sense (ie, virion-complementary-sense) RNA.

acid. Assuming this protein is phosphorylated at serine residues as is the MHV A59 protein (Stohlman and Lai, 1979), then a high level of phosphorylation might explain the 3-6 Kd difference between the predicted and measured molecular weights. Secondly, the protein is basic, a property expected of nucleic acid-binding proteins. Sixty-nine (18%) of the amino acids are basic whereas only 46 (12%) are acidic, giving the protein a net charge of +23 at neutral pH.

Although the concensus sequence around the AUG initiator codon for the TGEV nucleocapsid protein (UAAAUGG) is not among the most favored for translation initiation sites that have been described, it is not without precedent (Kozak, 1983). This AUG, therefore, probably identifies the authentic beginning of the TGEV nucleocapsid gene since the sequence from this point leftward to the end of clone FG5, except for a 12 base intergenic sequence, reveals an open reading frame coding for a protein sharing extensive regions of amino acid homology with the small matrix glycoprotein (E1) of the mouse hepatitis virus A59 (Armstrong et al., 1984; discussed below).

## A Second Open Reading Frame to the 3' Side of the Nucleocapsid Protein Gene Encodes a Hypothetical Protein of 9,101 Molecular Weight that is Hydrophobic at Both Ends

An open reading frame beginning at base 1507 and extending through base 1740 encodes a hypothetical 78 amino acid protein of 9,101 molecular weight (Figure 3.2). A hydrophobicity analysis of the protein reveals that it is hydrophobic for a stretch of approximately 25 amino acids at each end and it is hydrophilic in its central region (Figure 3.4). There are 8 basic amino acids and 4 acidic amino acids

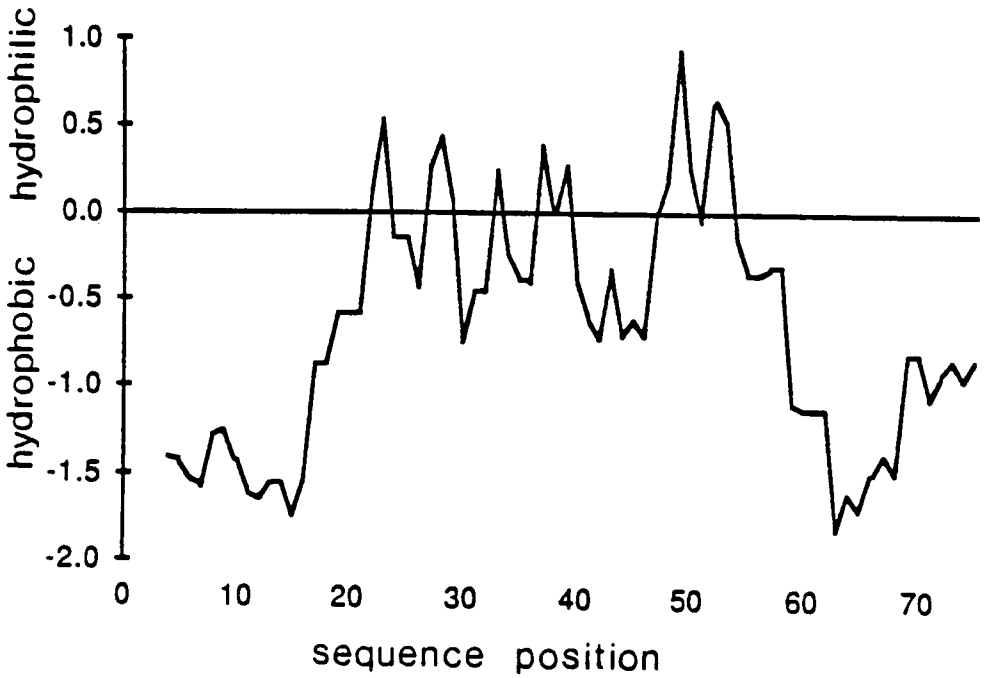



Figure 3.4. Plot of hydrophobic regions (areas below the line) and hydrophilic regions (areas above the line) for the TGEV hydrophobic protein. The method of Hopp and Woods (Hopp and Woods, 1981) was used and the plot was made using a window of 7 amino acids.

giving the protein a net +4 charge at neutral pH. Basic and acidic amino acids are distributed evenly throughout the central hydrophilic region, but 4 basic amino acids and no acidic ones are found among the 27 amino acids at the carboxy terminus. There is yet no direct evidence for this protein.

## D. DISCUSSION

We present the primary nucleotide sequence for the TGEV nucleocapsid protein (N) gene and the deduced amino acid sequence for the protein. This is the fourth coronavirus N gene to be sequenced and such information allows one to firstly, firmly conclude that TGEV shares an ancestral relationship with MHV and IBV, and secondly, to identify potentially functional domains on the N protein by examining conserved structures among the diverged viruses. The first two coronavirus N gene sequences to be described are those of the closely related JHM and A59 strains of MHV (Armstrong et al., 1983; Skinner et al., 1983) and between these an overall homology of 94% was found for both the nucleotide and amino acid sequences, reflecting the antigenic similarities between the viruses. Interestingly, the antigenically distinct avian infectious bronchitis virus shows no N gene nucleotide sequence homology with MHV above the 25% that would be expected from a random distribution of nucleotides, yet shares an overall amino acid sequence homology of 26% (Boursnell et al., 1985). Furthermore, there is a stretch of 67 amino acids within the amino terminal one-third of the protein that shows a sequence homology of 51% between the viruses suggesting that a strong selective pressure exists for a specific

functional group defined by this sequence. This interesting pattern repeats itself in the structure of the TGEV N protein. Although TGEV shows no antigenic relatedness to either MHV or IBV (Pederson et al., 1980), and no N gene nucleotide homology with either MHV or IBV, it does show an overall amino acid homology of 27 and 26% with MHV (JHM) and IBV respectively. Furthermore, the conserved 67 amino acid region is also found in TGEV (becoming 68 positions when TGEV is compared; Figure 3.5). This conserved region is slightly more basic than the overall nucleocapsid protein and therefore may function as a site of interaction with genomic RNA.

Other regions in the N proteins of the three viruses share structural similarities in the absence of a common primary structure suggesting the existence of other conserved functional domains. Although the N proteins are different lengths (382 amino acids for TGEV, 455 for MHV, and 409 for IBV), when the three are aligned by the 68 amino acid conserved sequence, the following structural similarities are observed. (1) 4 cluster groups containing 2-10 serine residues are found in parallel with TGEV amino acid positions 20-40, 150-190, 260-300, and 340-360. Other smaller serine clusters are found in MHV. In all three viruses, regions of 10-40 amino acid stretches can be found that are void of serines. (2) 3 cluster groups of 5-29 basic amino acid residues are found in parallel with TGEV amino acid positions 0-30, 150-260, and 330-350. (3) A cluster of 9-11 acidic amino acid residues is found within the last 32 amino acids at the carboxy terminus.

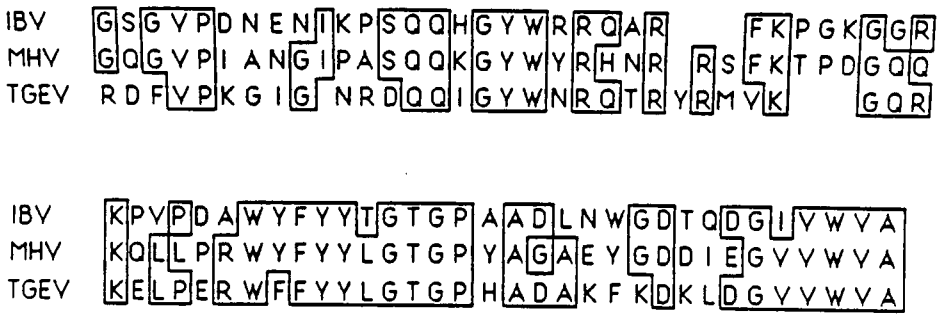



Figure 3.5. Amino acid sequence homologies among IBV, MHV JHM and TGEV for a 68 amino acid conserved region in the nucleocapsid protein gene. The sequence starts at amino acid 53 from the initiator codon for IBV (Boursnell et al., 1985), amino acid 86 for MHV JHM (Skinner and Siddell, 1983) and amino acid 53 for TGEV. Identical amino acids boxed in.

Although TGEV would appear to be equally diverged from IBV and MHV on the basis of amino acid sequence, TGEV more closely resembles MHV in its genome arrangement. Firstly, like MHV, the N gene for TGEV is flanked on its 5' side by the matrix protein (M or E1) gene, whereas for IBV, two genes, derived from overlapping reading frames and encoding putative nonstructural proteins, lie between the matrix protein (E1) gene and the nucleocapsid gene (Armstrong et al., 1984; Boursnell and Brown, 1984). That the matrix protein gene for TGEV lies to the immediate 5' side of the nucleocapsid gene is based on amino acid sequence homology. Of the 105 amino acids deduced for the TGEV matrix protein sequence, 31% are perfectly homologous and another 15% are conservative differences (Figure 3.6). Secondly, the number of nucleotides separating the M (E1) and N genes is close, 14 for MHV, and 12 for TGEV, and these match perfectly for a stretch of 8 bases:

```
MHV A59  T C T A A A C T T T A A G G
TGEV       C T A A A C T T C T A A
```

No direct evidence exists for the hydrophobic protein encoded by base 1507 through 1740. Certainly no gene of this description is found in a similar region of the genome for MHV or IBV, and such a protein may prove to be unique to TGEV. Genes encoding small hydrophobic proteins in MHV and IBV have been described, however (Boursnell and Brown, 1984; Skinner et al., 1985), but their hydrophobicity is only at one end, and no sequence homology is found between them and the TGEV hydrophobic protein. A protein showing the hydrophobicity pattern illustrated in Figure 3.4 would most certainly be membrane bound with both terminal domains embedded in the membrane and the hydrophilic

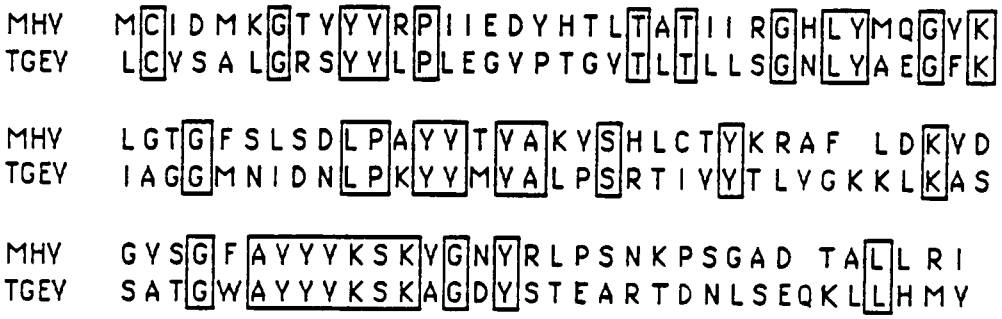



Figure 3.6. Amino acid sequence homology between the 3' terminal portion of the M (E1) protein of MHV A59 (Armstrong et al., 1984) and the deduced M protein of TGEV. The homology between the two sequences was aligned visually from the region of perfect match (AVYVKSK) beginning at amino acid 126 of the MHV sequence. From this alignment, the two termination codons are in phase if a single codon gap is inserted between amino acids 222 and 223 of the MHV sequence.

central region exposed. The hydrophilic portion of the protein could either be exposed to the luminal surface or the cytoplasmic side of the ER depending on its mode of synthesis. In either case the protein may serve as an anchor or signal for other proteins.

The potential for a small molecular weight protein in TGEV is of special interest since three small polyadenylated, putative messenger RNA's for TGEV have been identified (Dennis and Brian, 1982), and these messages have not been reported for MHV or IBV. Whether the hydrophobic protein would be translated from a smaller message of this type is pure conjecture at this time. Assuming TGEV replicates by the concensus scheme proposed for the replication of MHV and IBV, namely that (1) all messages show a 3' nested set arrangement, (2) there is a (approximately) 70 base primer leader sequence positioned at the 5' end of each message, and (3) only the 5' terminal ORF of each message is translated, then the 0.24 Md message described by Dennis and Brian may well be a message for the hydrophobic protein. From the known sequence (Figure 3.2), such a message would be 0.20 Md, well within the error limits for molecular weight determination by electrophoresis of such a polyadenylated species. There are two structural features that favor the plausability of functional hydrophobic protein gene. (1) The intergenic sequence preceding the HP gene, inclusive of the N gene terminal amino acid codon and stop codon, contains a 6 base sequence, CTAAAC, that is in common with part of the intergenic sequence preceding the N gene for both MHV and TGEV described above, and may play a role in the initiation of mRNA transcription (Budzilowicz, 1985). (2) The seven base sequence, GAGAUGC, at the imitation site of

the hydrophobic protein is a favored pattern among eukaryotic initiation sequences (Kozak, 1983).

Assuming that the gene for the hydrophobic protein is real then the 3' terminal noncoding sequence would be a total of 276 bases, exclusive of the poly A tail, and would be the shortest noncoding sequence of those identified for coronaviruses. The significance of the noncoding region is not completely known although it undoubtedly functions as an attachment region for the polymerase to initiate synthesis of the negative strand RNA. One possible site that may be critical for recognition or binding is a 10 base sequence, GGGAAGAGCT, that is conserved between IBV (found 81 bases from the 3' end) and MHV (82 bases from the end). With the exception of the first base, a T instead of G, TGEV shares an identical sequence beginning 77 bases from the 3' end (Figure 3.2).

# 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.

Apostolov, K., P. Spasic, and N. Bojanic. 1975. Evidence of a viral etiology in endemic (Balkan) nephropathy. Lancet 2:1271.

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.

AVMA Council on Research. 1974. Justifications for veterinary animal health research. Am. J. Vet. Res. 35: 875-887.

Baltimore, D. 1971. Expression of animal virus genomes. Bact. Rev. 35: 235-241.

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

Baric, R. S., S. A. Stohlman, M. K. Razani, and M. M. C. Lai. 1985. Characterization of leader-related small RNA's in coronavirus-infected cells: Further evidence for leader-primed mechanism of transcription. Virus Research 3: 19-33.

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.

Boursnell, M. E. G., T. D. K. Brown, and M. M. Binns. 1984. Sequence of the membrane protein gene from avian coronavirus IBV. Virus Research 1: 303-313.

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.

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

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, Saskatoon, Saskatchewan, Canada.

Briedis, D. J., R. A. Lamb, and P. W. Choppin. 1982. Sequence of RNA 7 of the influenza B virus genome: Partial amino acid homology between the membrane proteins (M1) of influenza A and B viruses and conservation of a second open reading frame. Virology 116:581-588.

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.

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.

Burks, J. S., B. 1. DeVald, L. D. Jankovsky, and J. C. Gerdes. 1980. Two coronavirus isolated from central nervous system tissue of two multiple sclerosis patients. Science. 209:933-934.

Caul, E. O., and S. K. R. Clarke. 1975. Coronavirus propagated from patients with non-bacterial gastroenteritis. Lancet 2: 953.

Caul, E. O., and S. I. Egglestone. 1977. Further studies on human enteric coronavirus. Arch. Virol. 54: 107-117.

Clewley, J. P., and D. H. L. Bishop. 1979. Assignment of the large oligonucleotides of vesicular stomatitis virus to the N, NS, M, G, and L genes and oligonucleotide gene ordering within the L gene. J. Virol. 30: 116-123.

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.

Collins, P. L., Y. T. Huang, and G. W. Wertz. 1984. Nucleotide sequence of the gene encoding the fusion (F) glycoprotein of human respiratory syncytial virus. Proc. Natl. Acad. Sci. 81: 7683-7687.

Collins, P. L., Y. T. Huang, and G. W. Wertz. 1984. Identification of a tenth mRNA of respiratory syncytial virus and assignment of polypeptides to the 10 viral genes. J. Virol. 49: 572-578.

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 $^{32}$P-labeled RNA. Anal. Biochem. 49: 184-197.

Donehorner, L. A., A. L. Huang, and G. L. Hager. 1981. Regulatory and coding potential of the mouse mammary tumor virus long terminal redundancy. J. Virol. 37:226-238.

Dugaiczyk, A., H. W. Boyer, and H. M. Goodman. 1975. Ligation of Eco Rl endonuclease-generated DNA fragments into linear and circular structures. J. Mol. Biol. 96:171-184.

Garwes, D. J., and D. H. Pocock. 1975. The polypeptide structure of transmissible gastroenteritis virus. J. Gen. Virol. 29: 25-34.

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.

Hanahan, D. 1983. Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. 166: 557-580.

Hanahan, D., and M. Meselson. 1980. A protocol for high density screening of plasmids in X1776. Gene 10:63-.

Heidecker, G., and J. Messing. 1983. Sequence analysis of zein cDNAs obtaind by an efficient mRNA method. Nuc. Acids Res. 11: 4891-4906.

Herz, C., E. Stovnezer, and R. M. Krug. 1981. Influenza virus, an RNA virus, synthesizes its messenger RNA in the nucleus of infected cells. Cell 26: 391-400.

Holmes, K. V. and J. N. Behnke. 1981. Evolution of a coronavirus during persistent infection in vitro. Adv. In Exp. Biol. and Med. 142: 287-299.

Hopp, T. P. and K. R. Woods. 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U.S.A. 78: 3824-3828.

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

Hu, S., J. Bruszewski, T. Boone, and L. Souza. 1984. Cloning and expression of the surface glycoprotein gp195 of porcine transmissible gastroenteritis virus. In "Modern Approaches to Vaccines". Edited by R. M. Chanock and R. A. Lerner. Cold Spring Harbor Laboratory, pp. 219-223.

Jacobs, L., W. J. M. Spaan, M. C. Horzinek, and B. A. M. Van der Zeijst. 1981. The synthesis of the subgenomic mRNA's of mouse hepatitis virus is initiated independently: evidence from UV transcription mapping. J. Virol. 39: 401-406.

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

Keck, J., R. Baric, J. Egbert, J. Leibowitz, S. Stohlman, and M. M. C. Lai. Murine coronaviruses: recombination between nonsegmented RNA genomes. The American Society for Microbiology, 1985 Annual Meeting. University of New Mexico, Albuquerque.

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

Kuismanen, E., K. Hedman, J. Saraste, and R. F. Petterson. 1982. Uukuniemi virus maturation: accumulation of virus particles and viral antigens in the Golgi complex. Mol. Cell. Biol. 2: 1444-1458.

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., J. O. Fleming, S. A. Stohlman, and K. Fujiwara. 1983. Genetic heterogeneity of murine coronaviruses. Arch. Virol. 78: 167-175.

Lai, M. M. C., C. D. Patton, R. S. Baric, and S. A. Stohlman. 1983. 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.

Lai, M. M. C., and S. A. Stohlman. 1978. RNA of mouse hepatitis virus. J. Virol. 26: 236-242.

Lai, M. M. C., and S. A. Stohlman. 1981. Comparative analysis of RNA genomes of mouse hepatitis viruses. J. Virol. 38:661-670.

Lasky, L., and J. Obijeski. 1984. Recombinant DNA vaccines. In "Concepts in Viral Pathogenesis". Notkins, A. and M. Oldstone, eds. pp. 366-373, Springer-Verlag, New York.

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

Leibowitz, J. L. and S. R. Weiss. 1981. Murine coronavirus RNA. Adv. In Exp. Biol. and Med. 142: 227-244.

Macrina, F. L., D. J. Kopecko, K. R. Jones, D. J. Ayers, and S. M. McCowen. 1978. A multiple plasmid-containing Escherichia coli strain: Convenient source of size reference plasmid molecules. Plasmid 1:417-420.

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

Matsumura, T., V. Stollar, and R. W. Schlesinger. 1971. Studies on the nature of dengue viruses. V. Structure and development of dengue virus in Vero cells. Virology 46: 344-355.

Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74:560-564.

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.

Moreau, K. M.S. Thesis, University of Tennessee.

Patterson, J. L., B. Halloway, and B. Kolakofsky. 1984. La Crosse virions contain a primer-stimulated RNA polymerase and a methylated cap-dependent endonuclease. J. Virol. 52: 215-222.

Peacock, S. L., C. M. McIver, and J. J. Monohan. 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 infections peritonitis virus to coronaviruses of other species. Arch. Virol. 58: 45-53.

Plotch, S. J., M. Bouloy, I. Ulmanen, and R. M. Krug. 1981. A unique cap ($m^7$GpppXm)-dependent influenza virion endonuclease cleaves capped RNA's to generate the primers that initiate viral RNA transcription. Cell 23: 847-858.

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

Rottier, P. J. M., M. C. Horzinek, and B. A. M. van der Zeijst. 1981. Translation of three mouse hepatitis virus (MHV-A59) subgenomic RNAs in Xenopus laevis oocytes. J. Virol. 38: 20-26.

Roychoudhury, R. and R. Wu. 1980. Terminal transferase-catalyzed addition of nucleotides to the 3' termini of DNA. Methods in Enzymology 65: 43-62.

Siddell, S. 1983. Coronavirus JHM: coding assignments of subjgenomic 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.

Siddell, S., H. Wege, A. Barthel, and V. ter Meulen. 1980. Coronavirus JHM; Cell-free synthesis of structural protein p60. J. Virol. 33: 10-17.

Siddell, S., H. Wege, A. Barthel, and V. ter Meulen. 1981. Intracellular protein synthesis and the in vitro translation of coronavirus JHM mRNA. Adv. in Exp. Biol. and Med. 142: 193-208.

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.

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., P. J. M. Rottier, M. C. Horzinek, and B. A. M. Van der Zeijst. 1981. Isolation and identification of virus-specific mRNAs in cells infected with mouse hepatitis virus (MHV-A59). Virology 108: 424-434.

Spaan, W. J. M., P. J. M. Rottier, M. C. Horzinek, and B. A. M. Van der Zeijst. 1982. Sequence relationships between the genome and the intracellular RNA species 1, 3, 6, and 7 of mouse hepatitis strain A59. J. Virol. 42: 432-439.

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: biogenesis of avian infectious bronchitis virus virion proteins. J. Virol. 44:974-803.

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

Sturman, L. S., and K. V. Holmes. 1983. The molecular biology of coronaviruses. Adv. Virus Res. 28: in press.

Tomizawa, J., T. Itoh, G. Selzer, and T. Som. 1981. Inhibition of col E1 RNA primer formation by a plasmid-specified small RNA. Proc. Natl. Acad. Sci. USA 78: 1421-1425.

Tomizawa, J., and T. Itoh. 1982. The importance of RNA secondary structure in col E1 primer formation. Cell 31: 575-583.

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

Wege, H., S. Siddell, M. Sturm, and V. ter Meulen. 1981. Coronavirus JHM: Characterization of intracellular viral RNA. J. Gen. Virol. 54: 213-217.

Wege, H., A. Muller, and V. ter Meulen. 1978. Genomic RNA of the murine coronavirus JHM. J. Gen. Virol. 41: 217-227.

Wheeler, D. A., J. S. Butel, D. Medina, R. D. Cardiff, and G. L. Hager. 1983. Transcription of mouse mammary tumor virus: Identification of a candidate mRNA for the long terminal repeat gene product. J. Virol. 46:42-49.

# APPENDICES

# APPENDIX A

## DETAILED PROCEDURE FOR THE MOLECULAR CLONING OF THE 3' END OF THE PORCINE TRANSMISSIBLE GASTROENTERITIS CORONAVIRUS GENOME

## KEYS TO SUCCESS IN cDNA CLONING

Our laboratory was the first at the University of Tennessee, Knoxville, to clone cDNA and to completely sequence a gene. The keys to our success in cloning cDNA were four in number and were discovered after we initially attempted other methods. (1) Homopolymer tailing (with CTP) was done using the conditions of Roychoudhury and Wu, being careful to add the reaction components in order. Lorie Lorenz advised that precipitations (of Cobalt [?]) may result if the proper order is not followed. (2) The simple method of Gubler and Hoffman was used. (3) The transformation conditions of Hanahan were used. (4) Transformation with a high concentration of DNA (annealed vector plus insert) is not necessarily good and may be detrimental. At high DNA concentrations, background (uncut) plasmids cotransform with reannealed plasmids containing inserts. Since they "outgrow" the recombinant plasmids, recombinant plasmids are not identifiable.

## PREPARATION OF TGEV GENOMIC RNA FOR cDNA CLONING

The Purdue high titer strain of porcine transmissible gastroenteritis coronavirus (TGEV) was plaque purified and stocks were prepared as previously described (Brian et al., 1980). To grow virus, confluent monolayers of ST cells in 150 cm flasks were infected with virus at a multiplicity of approximately 5 PFU/cell. After one hour adsorption cells were rinsed with Earle balanced salt solution and refed with Dulbecco's medium containing 1% fetal calf serum (Sterile Systems, Inc.). At 18 hours post infection, when 75 to 100% of the

cells were showing cytopathic effect, supernatant fluids were clarified by centrifugation at 10,000 x g for 10 min., and stored at -80°C. Viral RNA in one out of ten flasks was radiolabeled in order to follow the purification of the RNA. RNA was either labeled with $^{32}$P-orthophosphate or $^{3}$H-uridine. For $^{32}$P-labeled RNA, infected cells were refed with phosphate-free, minimum essential medium containing 1% fetal calf serum and 40 uCi $^{32}$P-orthophosphate (ICN) per ml. If $^{3}$H-uridine was used to label the RNA, infected cells were refed with Dulbecco's medium containing 1% fetal calf serum and 100 uCi per flask $^{3}$H-uridine (10 uCi/ml, ICN). Virus was purified from clarified supernatant fluids by first sedimentation at 90,000 x g for 1.5 hour onto a cushion of 60% sucrose (wt/wt), and then by sedimentation to buoyant density at 90,000 x g for 4 hour at 4°C in 18 ml continuous gradients of 20 to 60% sucrose (wt/wt) using a Sorvall AH-27 rotor. Fractionated isopycnic gradients of radiolabeled TGEV virions typically gave a profile like that seen in Figure A.1. If the virus was labeled with $^{3}$H-uridine, 5 ul aliquots were spotted on Whatman paper and TCA precipitated. The Whatman paper was then counted in 3 ml aqueous scintillation fluor (Scintiverse, Fisher). If the virus was labeled with $^{32}$P-orthophosphate, samples were counted by Cerenkov radiation. Virus was found to band at a buoyant density of 1.18 to 1.20 g/cc (Figure A.1), the position shown for infectious virus (Moreau, K., M.S. Thesis, Univ. of Tennessee). Collected virus was diluted with an equal volume of TMEN (10 mM Tris maleate [pH 6.0)--100 mM NaCl--1 mM EDTA) and pelleted at 150,000 x g for two hours in a Sorvall AH650 rotor. Sucrose solutions were made in TMEN.

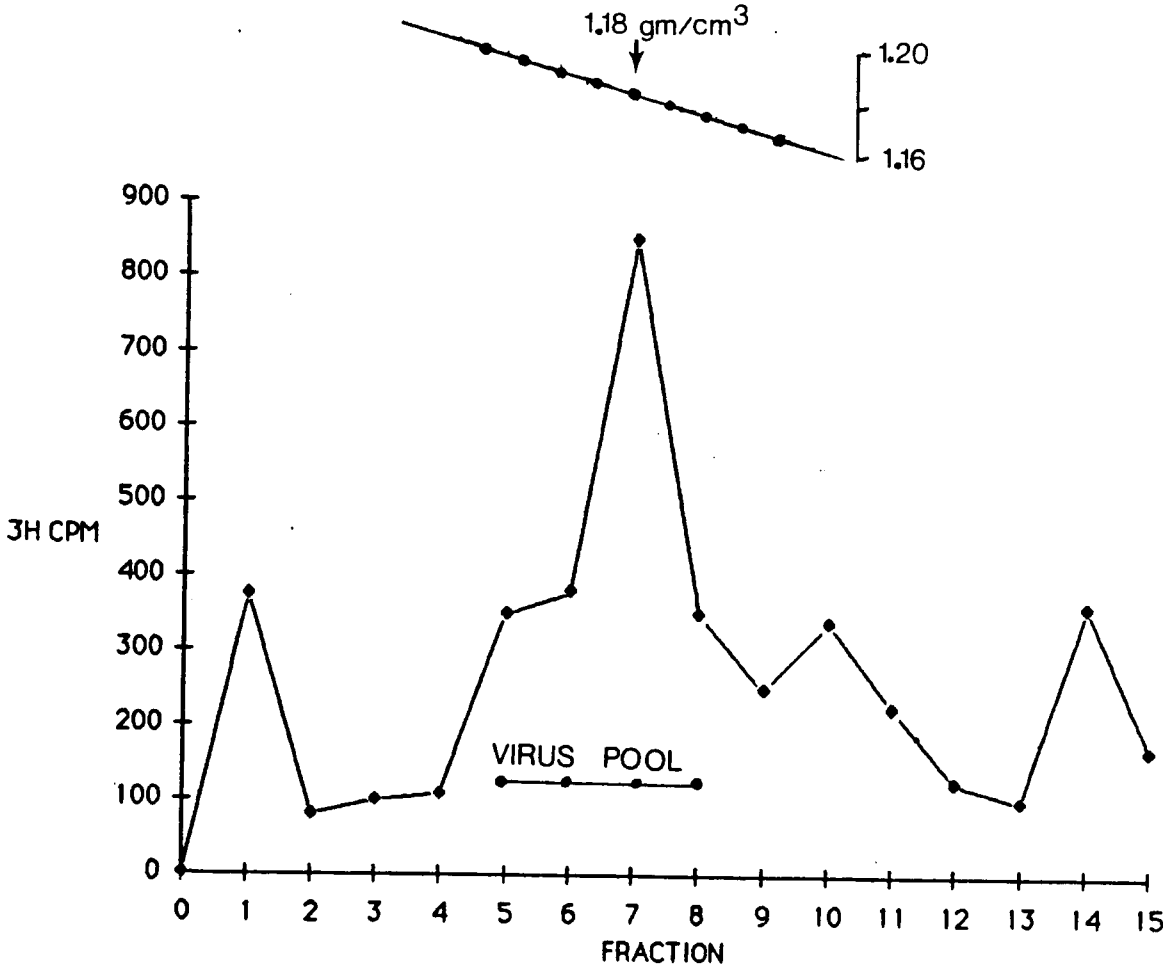



Figure A.1. Isopycnic sedimentation of TGEV used for preparation of virion RNA. A mixture of unlabeled TGEV and TGEV labeled with $^{3}$H-uridine was sedimented isopycnically in a 20-60% linear sucrose gradient as described in the text. Virus banding with a buoyant density of 1.17 to 1.20 gm/cc (fractions 5 through 8) was collected and pelleted for RNA extraction.

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)--1% SDS--0.5 mg proteinase K/ml, incubating for 0.5 hr at 37°C, and extracting twice with an equal volume of a mixture of 48% phenol--48% chloroform--2% isoamyl alcohol. RNA was precipitated at -80°C after adding 0.1 volume of 2 M sodium acetate and 2 volumes of 100% ethanol. Because small molecular weight RNA species are found in some preparations of purified coronavirion RNA, full-length genomic RNA was selected by rate-zonal sedimentation on preformed linear sucrose gradients of 15 to 30% sucrose (wt/wt) made up in TNE--0.1% SDS. RNA was dissolved in water and sedimented 1.5 hours at 110,000 x g, 25°C, on 5 ml gradients. Fractions (0.2 ml) were collected by dripping from the bottom of the tube and the distribution of radioactivity was determined by scintillation spectrophotometry ($^{3}H$) or by Cerenkov radiation ($^{32}P$) (Figure A.2). Only RNA sedimenting with a sedimentation coefficient of 50S or greater, as determined by reference to the sedimentation rate of mammalian 28S and 18S ribosomal RNA in a parallel gradient, was used. RNA was precipitated at -20°C after the addition of 0.1 volume of 2 M sodium acetate and 2 volumes of 100% ethanol.

In some preparations, TGEV genomic RNA appeared partially degraded and a continuous slope beginning from the bottom of the gradient and rising toward the top of the tube was observed. This appeared to be a function of virus harvest time. Virus harvested late (i.e., >18 h when there was CPE in 100% of the cells) contained RNA that was either 1) more degraded or 2) contained more cellular RNA species. Consequently,

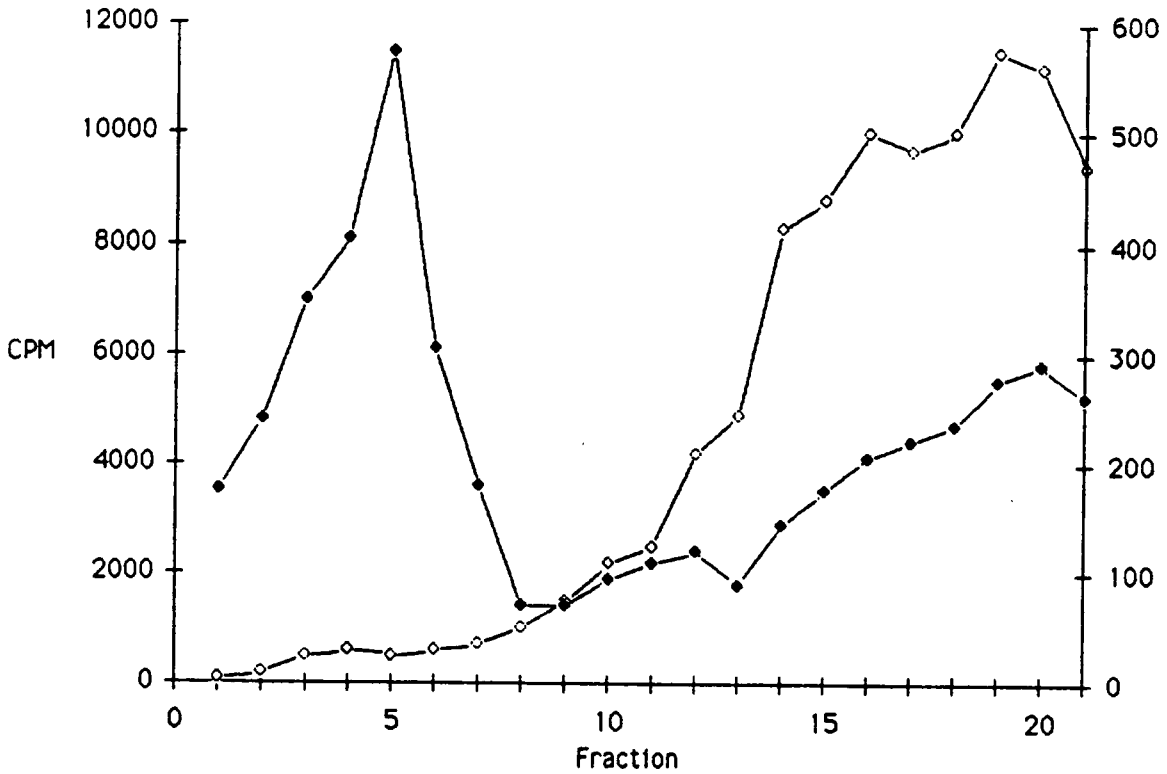



Figure A.2. Rate zonal sedimentation of TGEV genomic RNA. Pooled TGEV genomic RNA was harvested as indicated and used as starting material for cDNA cloning. $^{3}$H-labeled TGEV RNA was sedimented on a rate-zonal gradient as described in the text. $^{3}$H-labeled Maden-Darby bovine kidney cell 28S and 18S rRNA was sedimented on a parallel gradient and used for sedimentation rate markers. Sedimentation is from right to left.

for these studies, virus was harvested early (<18 h) after infection.

## FIRST AND SECOND STRAND cDNA SYNTHESIS

TGEV 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 10 mM Tris hydrochloride (pH 8.3)--10 mM $MgCl_2$--10 mM DTT--2 mM dGTP--2 mM dTTP--2 mM dATP--2 mM dCTP--0.1 uCi ($^{32}P$) dCTP (3,000 Ci/mmole, ICN)--50 pmoles oligo dT 12-18, 6 ug TGEV high molecular weight RNA, 30 U RNAsin--10 U reverse transcriptase, for 1 hour at 42°C. The reaction was stopped by adding 2 ul of 0.5 M EDTA and the nucleic acids were phenol extracted and then precipitated for a minimum of 30 min. on dry ice after the addition of 0.5 volume of 7.5 M ammonium acetate and 2.5 volumes of 100% ethanol. The precipitate was recovered by sedimentation at 12,000 x g for 20 min.

Second strand synthesis was carried out in a reaction volume of 100 ul containing 20 mM Tris hydrochloride (pH 7.5)--5 mM $MgCl_2$--10 mM ammonium sulfate--100 mM KCl--0.15 mM NAD--50 mg/ml BSA--40 uM dNTP's--8.5 u/ml *E. coli* RNAse H--230 U/ml DNA polymerase I--10 u/ml DNA ligase--all the product from the first strand reaction. The reaction was incubated at 12°C for one hour, then at 22°C for one hour. The reaction was stopped by adding 4 ul of 0.5 M EDTA and the reaction products were phenol extracted, passed over a Sephadex G-50 spun column, and the dscDNA was ethanol precipitated.

To test that conditions for first and second strand synthesis were correct, both reactions were carried out with rabbit globin mRNA such that the products were radiolabeled with $^{32}P$. The results of this testing are shown in Figure A.3.

## HOMOPOLYMER TAILING

Homopolymer tailing was carried out using the conditions described by Roychoudhury and Wu (Roychoudhury and Wu, 1980). Care was taken to avoid using sodium acetate for precipitation of nucleic acids since the terminal deoxynucleotide transferase enzyme is inhibited by sodium ions. Ammonium acetate was the salt used for precipitation of the nucleic acids prior to tailing.

To add the proper tail length of 15 dCMP residues per 3' end, (Peacock et al., 1981), an estimate was made of the total amount of dscDNA product, and all of the product was assumed to be 1 Kb in length. The amount of dscDNA made from 6 ug of genomic RNA was estimated from the amount of radiolabel in the dscDNA product that had been incorporated during first strand synthesis. During first strand synthesis, $^{32}P$- dCMP was incorporated at a specific activity of 2220 cpm/pmole. This was equivalent to 555 cpm/pmole total NMP molecules incorporated or 555 cpm/310 pg of incorporated NMP (using 310 daltons as the average mass of a NMP molecule). Since a total of 200,000 cpm was found in the dscDNA product, then the mass of DNA in the first strand was calculated to be 111 ng ([310 pg ÷ 555 cpm] x [200,000 cpm]) and the total mass of dscDNA is therefore 200 ng.




Figure A.3. Proof of first and second strand synthesis. Results of a preliminary experiment showing complete first and second strand synthesis of globin mRNA using the protocol of Gubler and Hoffman (Gubler and Hoffman, 1983). Proof of first strand synthesis was shown by a $^{32}$P-labeled sscDNA product (lane 1). 200 ng of rabbit globin mRNA was used for first strand synthesis in the presence of $^{32}$P-labeled- dCTP. Proof of second strand synthesis was shown by a $^{32}$P-labeled dscDNA product (lane 2). In this experiment, 400 ng of rabbit globin mRNA was used to make first strand cDNA in the absence of any label, and all the first strand product was used for second strand synthesis in the presence of $^{32}$P-labeled dNTPs.

Assuming all the product is 1 Kb (M.W. = 620,000) in size, then there is 0.322 pmole dscDNA product.

## TRANSFORMATION CONDITIONS

Bacterial transformation was performed using the method of Hanahan (Hanahan, D. J., 1983).

The cloning conditions recommended by Gubler and Hoffman (Gubler and Hoffman, 1983) were used throughout. Bacterial transformations were carried out exactly as described by Hanahan (Hanahan, 1983). Transformation efficiencies for pUC9 supercoiled were typically $10^8$ to $10^9$ transformations/ug pUC9.

Of the many parameters that are important for optimizing cloning efficiency, only three were examined in pilot experiments using TGEV cDNA and pUC9 vector. These parameters were (1) the vector to cDNA ratio, (2) the total DNA concentration, and (3) the length of time used for annealing.

Table A.1. shows the results of varying cDNA to vector ratios on transformation efficiency. Although a mixima was not obtained in this experiment, it is obvious that of the cDNA ratios used, the smallest ratio gave the highest efficiency. cDNA to vector weight ratios were therefore kept at less than 1 throughout all remaining cloning attempts.

Table A.2 shows the results of varying the total DNA concentration on transformation efficiency. Efficiency of transformation was 2-fold higher using the more dilute conditions, and a total DNA concentration

Table A.1. Effect of insert/vector weight ratio on cloning efficiency[a]

| ng cDNA | ng cDNA/ ng vector | # white colonies[b] | # white colonies/ ng cDNA |
|---|---|---|---|
| 1 | 0.43 | 13 | 13 |
| 2 | 0.86 | 10 | 5 |
| 5 | 2.15 | 14 | 2.8 |
| 10 | 4.29 | 18 | 1.8 |

[a]All experiments were carried out in triplicate. Annealing was done in a volume of 30 ul using 2.33 ng dg-tailed pUC9 as the vector.

[b]The number of blue colonies, indicating background uncut pUC9 as a contaminant, should have remained constant throughout. It did not remain constant. It varied $\pm$10% of the mean value, so therefore an error rate of at least 10% can be expected in these results.

Table A.2. Effect of total DNA concentration during annealing on cloning efficiency[a]

| Total DNA Concentration[b] (ug/ul) | # white colonies | # white colonies/ ng cDNA |
|---|---|---|
| 0.14 | 10 | 5 |
| 0.08 | 70 | 14 |

[a]All volumes are an average of three independent platings.

[b]The mass ratio of cDNA/vector DNA was 0.7 throughout.

of less than 0.1 ug/ul was therefore used throughout all remaining cloning attempts.

Since initial data showed that annealing ought to be carried out under relatively dilute conditions, and since ligation theory (Dugaiczyk, A., 1975) shows that the time for annealing increases with decreasing DNA concentration, the time for annealing was also investigated as a parameter for annealing and transformation. Table A.3 shows the effect of annealing time on cloning efficiency. Increasing the time of annealing from two to four hours increased cloning efficiency, but no advantage was gained by increasing the annealing time to nineteen hours. An annealing time of 4 hours was therefore used throughout all remaining cloning attempts.

Annealed DNA was either used immediately or stored at -20°C for later use. The annealed DNA was not precipitated. It was felt that recovery of annealed DNA, even with the use of carrier tRNA during precipitation, would not be high enough to warrant a chance of loss.

## IDENTIFICATION OF CLONES CONTAINING TGEV SPECIFIC SEQUENCES

Following annealing and transformation of JM 103 with the chimeric plasmids, white transformants were picked and placed in linear arrays on nitrocellulose, grown, and replica plated as described in Maniatis (Maniatis, 1982). A substantial number (40%) of the colonies that were white initially slowly turned blue after several weeks. Some colonies that turned blue still contained cDNA inserts. For storage, colonies were frozen at -20°C after first placing the nitrocellulose filters on

Table A.3. Effect of time of annealing on cloning efficiency[a]

| Annealing time (hours) | # white colonies | # white colonies/ ng cDNA |
|---|---|---|
| 2 | 12 | 12 |
| 4 | 19 | 19 |
| 19 | 19 | 19 |

[a]Conditions used for annealing are given in the text. 5 ng of cDNA and 7 ng of G-tailed pUC9 were annealed for the times indicated in a total volume of 150 ul. All values are an average of three independent platings.

medium containing 15% glycerol and growing the bacteria at 37°C for one hour. The bacteria stored at -20°C in this manner remained viable for at least two years.

Colony blots were lysed using a modified method of Hanahan and Meselson (Hanahan, D. and M. Meselson, 1980). The nitrocellulose filters were removed from the plate and placed in a puddle of the lysis buffer (0.5 M NaOH) on non absorbant sheets. After 5 minutes it was air dried and placed on lysis buffer again for another 5 minutes. Filters were then placed in 1 M Tris hydrochloride (pH 8.0) for five minutes, air dried, treated again for five minutes in 1 M Tris hydrochloride (pH 8.0), air dried, and placed in 1 M Tris hydrochloride (pH 8.0) with 1.5 M NaCl for 3 minutes. Filters were finally air dried, and baked in vacuo at 70°C for one to two hours. Filters that were not used immediately were placed between two pieces of Whatman #3 paper and heat-sealed in a plastic bag. Filters were kept up to two years in this manner and still were able to bind probe.

For hybridization, blots were rehydrated by floating the filter on the surface of 6 X SSC and were prewashed to remove loose debris by agitation in a large volume of prewashing solution (50 mM Tris hydrochloride [pH 8.0]--1 M NaCl--1 mM EDTA--0.1% SDS) in a glass baking dish. Filters were transferred to a plastic heat-sealable bag and prehybridized in 20 ml of prehybridization solution (6 X SSC--0.5% SDS--5X Denhardt's--100 ug/ml denatured salmon sperm DNA) at 68°C for 4 hours.

Probe used to identify TGEV transformants was prepared in reverse transcriptase buffer using the same conditions as described for first

strand synthesis except dCTP was ten fold less and 100 uCi of $^{32}$P-dCTP was used. 500 ng of high molecular weight TGEV RNA was primed with 50 ug of sheared calf thymus DNA. The reaction was stopped with 4 ul of 0.5 M EDTA, denatured with alkali (4 ul 1 M NaOH room temperature overnight, and neutralized with 25 ul 1 M Tris hydrochloride [pH 8.0] and 4 ul 1 M HCl), passed over a G-50 column and added to the blots. The blots were hybridized with at least ten million cpm (Cerenkov) for 48 hours at 68°C. Following hybridization, filters were washed in five changes of 2 X SSC at room temperature, then washed in 0.1 X SSC with 0.5% SDS at 68°C for two hours. The solution was changed and the filters washed for an additional hour at 68°C, allowed to air dry, and put on film. Figure A.4 shows the autoradiograph of one of the colony blots using TGEV randomly primed probe. Radioactive tape was placed on the edge of the filters in an assymetrical fashion for orientation. Colonies showing strong hybridization signals were selected and screened to determine insert size by the method of Kado and Liu (Kado and Liu, 1981).

Approximately 200 colonies were screened by rapid plasmid isolation and gel electrophoresis. Figure A.5 shows a photograph of one such gel stained with ethidium bromide. From this gel, clone FG5 was estimated to have an insert of at least 2 Kb. Clone FG5 was the clone with the largest TGEV insert identified. Clones containing inserts of 200 bases or larger, based on the migration of the supercoiled form of the plasmids, were selected for further analyses. All clones selected for further analyses were first subcloned into DH1. This was done for two reasons. 1. JM103 which was used for initial

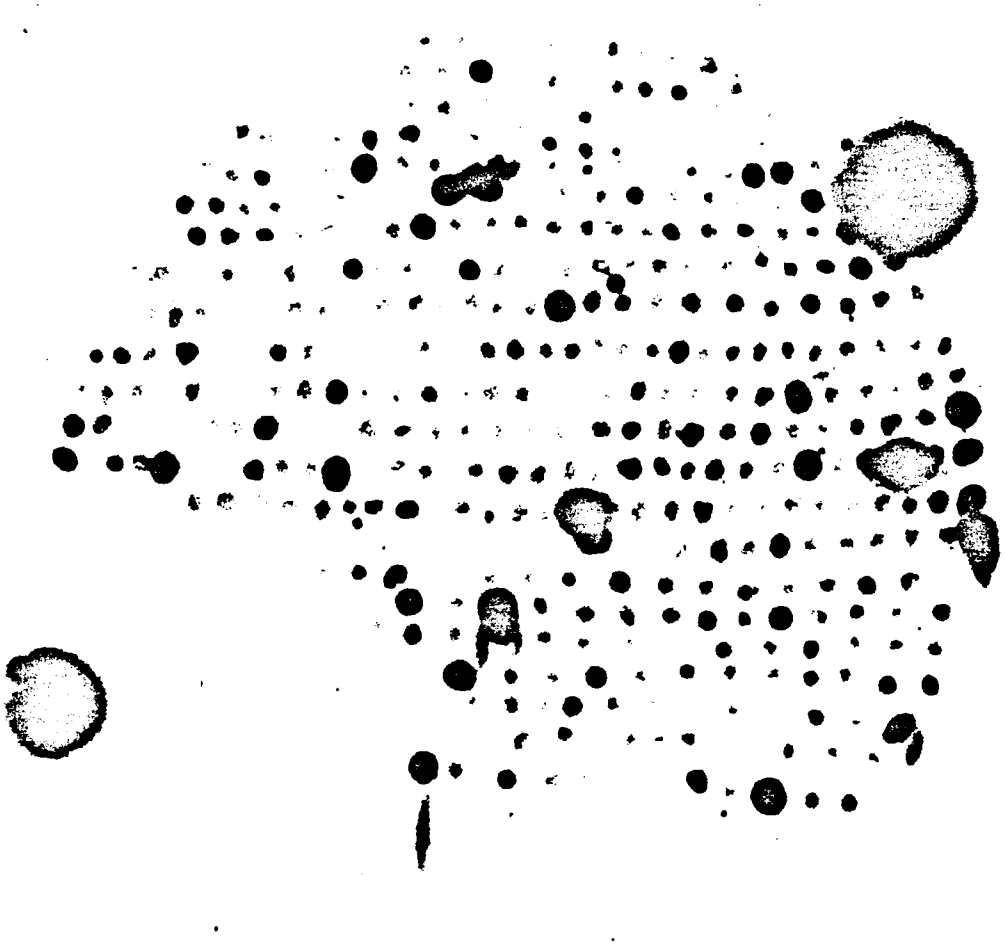


Figure A.4. Colony hybridization of a collection of white colonies. Hybridization probe was $^{32}$P-labeled cDNA prepared in a randomly primed reaction using size-selected TGEV genomic RNA as the template.

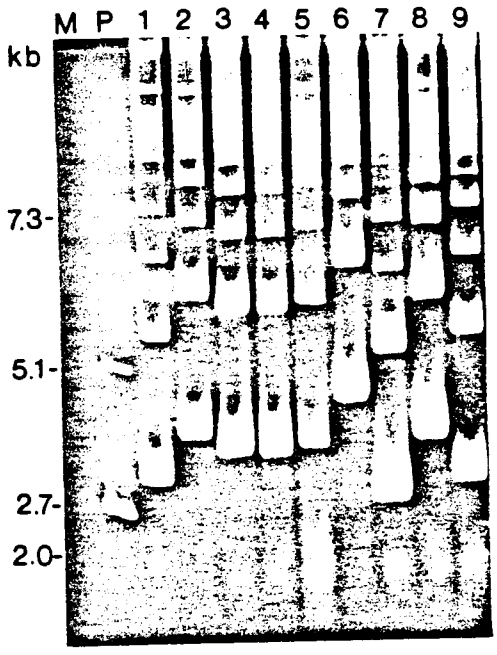



Figure A.5. Rapid screening to determine insert size. Electrophoresis of supercoiled plasmids from clones yielding a positive hybridization signal with random-primed TGEV cDNA probes were screened for insert size. Colonies were grown in LB broth containing ampicillin, and plasmid DNA was isolated and analyzed by the method of Kado and Liu (Kado and Liu, 1981). Lane M: E. coli plasmids from strain V517 (Macrima et al., 1978) that serve as molecular weight markers. Lane P: pUC9 supercoiled plasmid isolated by the method of Kado and Liu. Lane 6: clone FG5.

transformation contains a cryptic plasmid. This would interfere with the isolation of pUC9 with insert on CsCl gradients. 2. DH1 was chosen because it is rec $A^-$ and the plasmid is less likely to suffer recombinational events. Gels were blotted onto nitrocellulose using the technique of Southern (Southern, 1975) in order to screen for clones containing poly A sequences.

## IDENTIFICATION OF CLONES CONTAINING POLY A SEQUENCES

500 ng of a poly(rA):oligo(dT) template was copied using reverse transcriptase and the conditions were those used for first strand synthesis except for the following changes: dTTP was the only nucleotide used in the reaction. The dTTP concentration was 50 uM (rather than 2 mM). 150 uCi of $^{32}P$ dTTP was used to make the template hot. After incubating the reaction at 37°C for one hour, the reaction was stopped by adding 4 ul of 0.5 M EDTA. The RNA template was hydrolyzed with alkali overnight (4 ul of 1 M NaOH), neutralized (25 ul of 1 M Tris hydrochloride, then 4 ul of 1 N HCl), and passed over a Sephadex G-50 column (5 ml). Three drop (0.5 ml) fractions were collected and counted by Cerenkov radiation. The product was pooled and added directly to the prehybridization fluid.

Blots were hybridized with the $^{32}P$ poly(dT) probe for 24 hours at 37°C then for 48 hours at room temperature. Following hybridization, blots were washed in three successive changes of 6 X SSC at 20°C.

Out of a total of 15 clones selected for further study on the basis of size, two clones (J21 and J22) were identified as containing poly A inserts (Figures A.6 and A.7, and data not shown).

Poly A containing clones (J21 and J22) and selected large clones (N10, C7, H15, I32, FG5, FQ-2, and V14) were subcloned into E. coli strain DH1 and maxi preparations were made and plasmids purified by CsCl centrifugation.

## IDENTIFICATION OF 3'-SPECIFIC TGEV cDNA CLONES

For identification of 3' specific clones of the TGEV genome, the presumption was made that the genome is polyadenylated only at the 3' end. This in turn was based upon the documented 3' polyadenylation site in the genome of the mouse hepatitis coronavirus (Lai, 1983) and the avian infectious bronchitis coronavirus (Stern and Kennedy, 1980a,b). The 8 largest clones were selected, and isolated inserts were probed using the purified 700 insert from J21 that had been shown to be polyadenylated. Clones were analyzed by Southern hybridization to both supercoiled plasmid containing the insert and to insert above removed by Pst 1 digestion (Figures A.8 and A.9). When Pst 1 cut G-tailed vector was annealed to C-tailed cDNA, the Pst 1 site was reconstructed on each side of the insert. This allowed the plasmid to be cut with Pst 1 releasing the insert from the vector.

Figure A.10 is a flow diagram of the cloning procedure. Starting wtih 6 ug of TGEV high molecular weight RNA, the efficiency of cloning at every step is described.

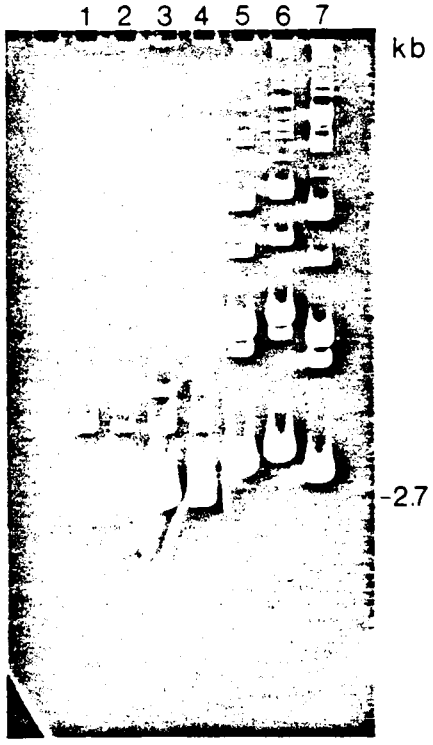



Figure A.6. Electrophoresis of putative poly(A)-containing plasmids in preparation for hybridization against $^{32}$P-labeled poly(dT). Agarose gel electrophoresis of supercoiled plasmids from clones suspected, from preliminary experiments, to contain poly A. Lanes 1 and 2: control plasmids known to contain poly A (clones of the bovine coronavirus genome). Lanes 3 and 4: pUC9 Lane 5: N10 Lane 6: J21 Lane 7: J22.

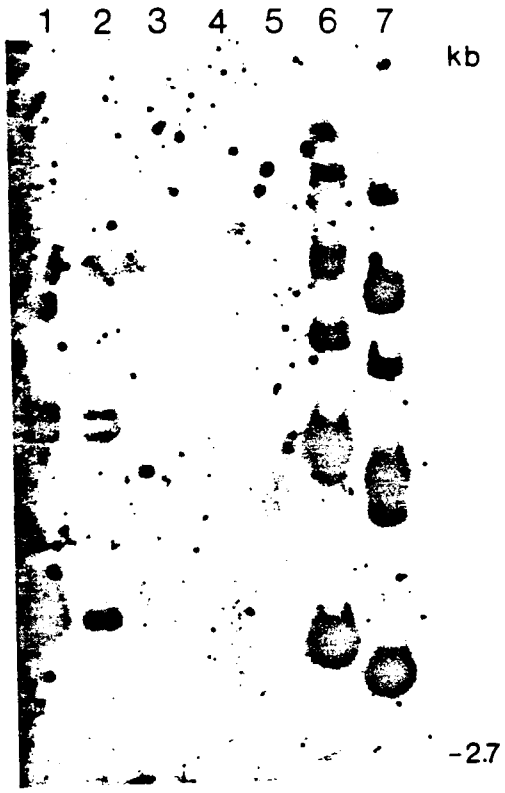



Figure A.7. Southern blot used to identify poly(A)-containing clones. Southern blot of gel shown in Figure A.6 using $^{32}$P-labeled poly(dT) to identify poly(dA) containing sequences. Lanes 1 and 2: control poly A containing plasmid. Lanes 3 and 4: pUC9 control. Lane 5: N10. Lane 6: J21. Lane 7: J22.

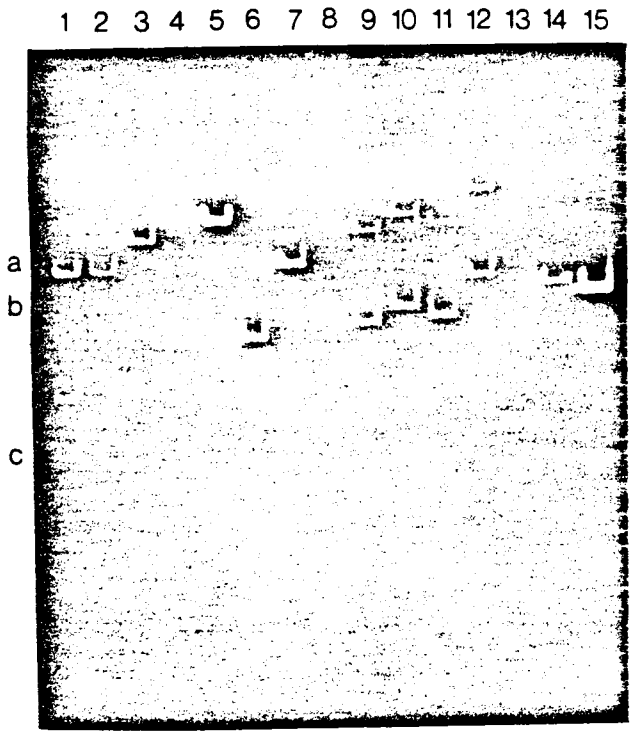



Figure A.8. Electrophoresis of selected TGEV clones in preparation for hybridization against the poly A containing clone J21. Lane 1: Pst 1 cut FG5. Lane 2: Pst 1 cut FQ-2. Lane 3: Hind III cut C7. Lane 4: Hind III cut H15. Lane 5: Hind III cut I32. Lane 6: supercoiled pUC9. Lane 7: supercoiled FG5. Lane 8: supercoiled FQ-2. Lane 9: supercoiled J22. Lane 10: supercoiled J21. Lane 11: supercoiled N10. Lane 12: supercoiled V14. Lane 13: supercoiled H15. Lane 14: supercoiled C7. Lane 15: supercoiled I32.

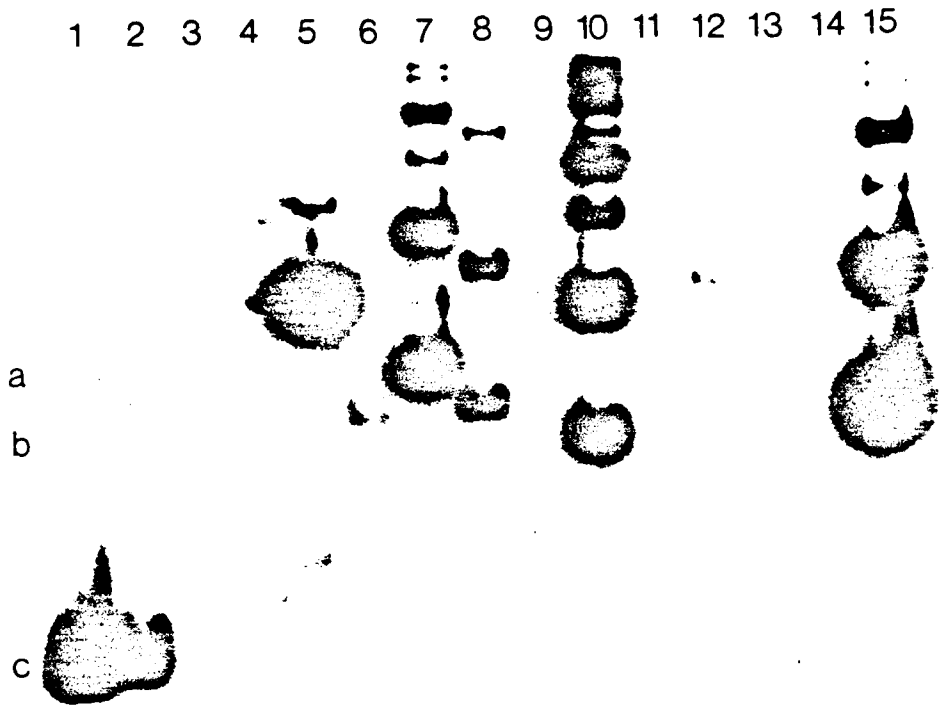



Figure A.9. Southern hybridization of selected TGEV clones with nick-translated purified J21 insert. The gel shown in Figure A.8 was blotted onto nitrocellulose and hybridized against $^{32}$P-labeled J21 insert. Note hybridization of clone FG5 in lane 7 and the linearized clone FG5 in lane 1.

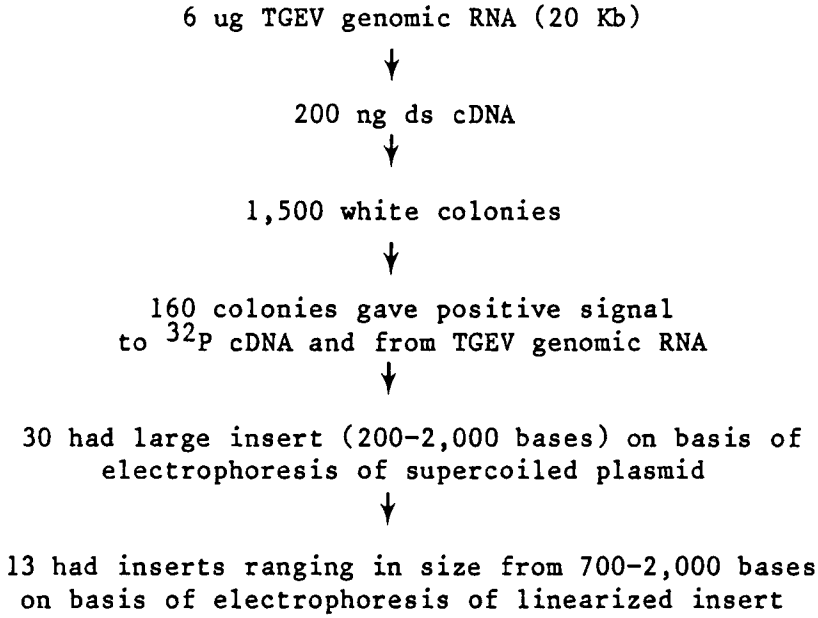



Figure A.10. Flow diagram showing efficiency of cloning protocol at key steps.

Table A.4 is a summary of the 13 largest clones that were selected and studied further. They range in size from 0.4 Kb to 2.0 Kb. Some of the clones may possibly represent TGEV sequences farther toward the 5' end of the genome than are represented in clone FG5. For instance, clone C7 has two internal Hind III sites and clone H15 has one internal Hind III site. Hind III is a restriction enzyme that has a six base recognition sequence. On a purely statistical basis one would expect a Hind III site only once every 4096 bases. The internal Hind III site in clone FG5 cannot be one of the same sites found in either clone C7 or clone H15 since nick-translated fragments of clone FG5 did not cross hybridize with either C7 or H15.

Table A.4. Summary of TGEV-specific cDNA clones and their names

| Clone name | Clone size[a] | poly A containing | Hybridizes with J21 | Number of Internal Sites | | Subcloned into DH1 |
|---|---|---|---|---|---|---|
| | | | | Pst 1 | Hind III | |
| FG5 | 2.1 Kb | ND | + | 2 | 1 | Yes |
| H15 | 1.85 | - | - | 0 | 1 | Yes |
| V14 | 1.5 | - | - | 0 | 0 | Yes |
| I32 | 1.4 | - | - | 3 (?) | 0 | Yes |
| C7 | 1.3 | - | - | 0 | 2 | Yes |
| FQ-2 | 1.1 | - | + | 1 | - | Yes |
| RRAC8 | 0.6-1 | ND | ND | ND | ND | No |
| K33 | 0.6-1 | ND | ND | ND | ND | No |
| RRA3W | 0.6-1 | ND | ND | ND | ND | No |
| B1 | 0.6-1 | ND | ND | ND | ND | No |
| J21 | 0.7 | + | + | 0 | 0 | Yes |
| N10 | 0.4 | - | - | 0 | 0 | Yes |
| J22 | 0.4 | + | + | 0 | 0 | Yes |

[a]Determined by agarose gel electrophoresis of inserts removed the plasmid with Pst 1.

# APPENDIX B

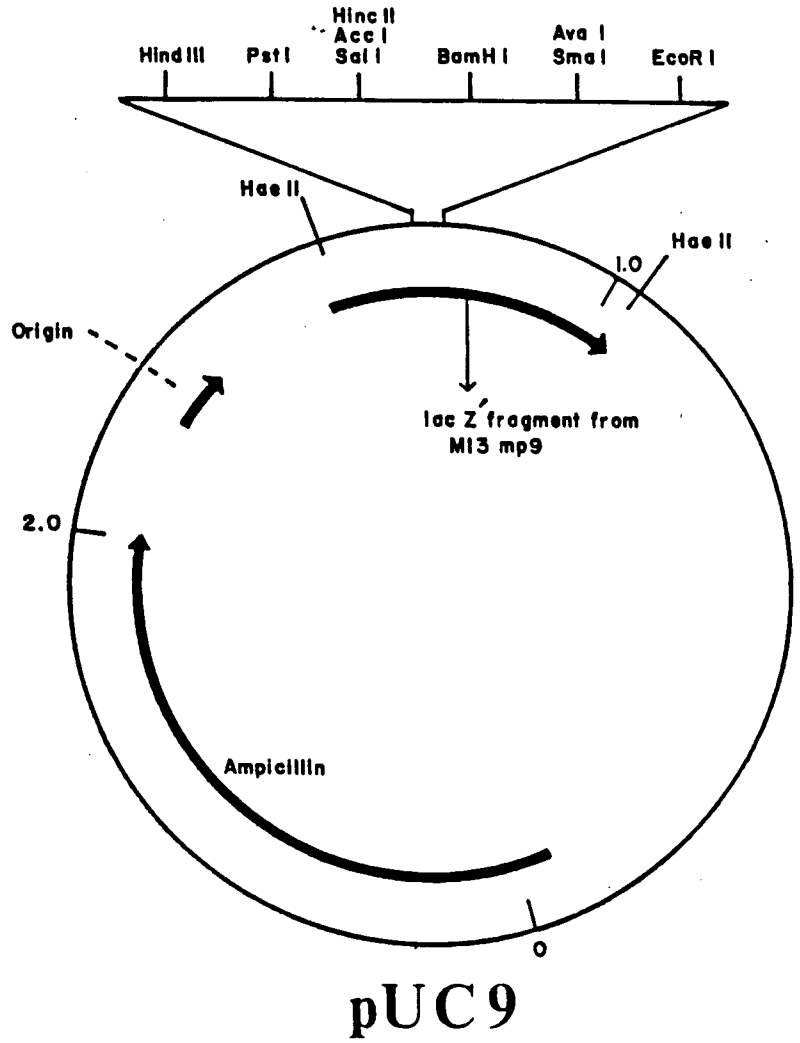
## DETAILED ANALYSIS OF CLONE FG5

## DETAILED ANALYSES OF CLONE FG5

The essential results of sequencing clones J21 and FG5 are discussed in chapter III: "Cloning and sequence analysis of the 3' end of the porcine transmissible gastroenteritis coronavirus: Primary structure of the nucleocapsid protein gene." Detailed analyses that lead to conclusions given in chapter III, as well as additional speculations concerning the significance of these findings, are given below.

Figure B.1 shows the key features of pUC9 and its multiple cloning site which was the vector used for cloning TGEV sequences. TGEV cDNA was cloned into the Pst 1 site.

To make a restriction map of clone FG5, the plasmid was cut with the restriction enzyme Sal I, end labeled with polynucleotide kinase and $-^{32}$P-dCTP, and recut with the restriction enzyme Bam Hl to remove the label from one end of the molecule (note the Bam Hl site in Figure B.1). The DNA was then digested with limited amounts of selected restriction enzymes. Aliquots from a given reaction were removed at various times, and the products separated on an agarose gel. Figure B.2 shows an autoradiogram of one such gel. The labeled DNA molecule gets digested to successively smaller fragments. The sizes of the fragments are determine by comparing their mobility with labeled fragments of lambda DNA digested with restriction endonuclease Hind III, and Phi X 174 DNA digested with restriction endonuclease Hae III. Figure B.3 is the restriction map for clone FG5 generated by the Smith and Bernstiel procedure just described. Figure B.4 is a restriction




pUC9

Figure B.1. Diagram of pUC9 vector used in cloning TGEV, with key features labeled.

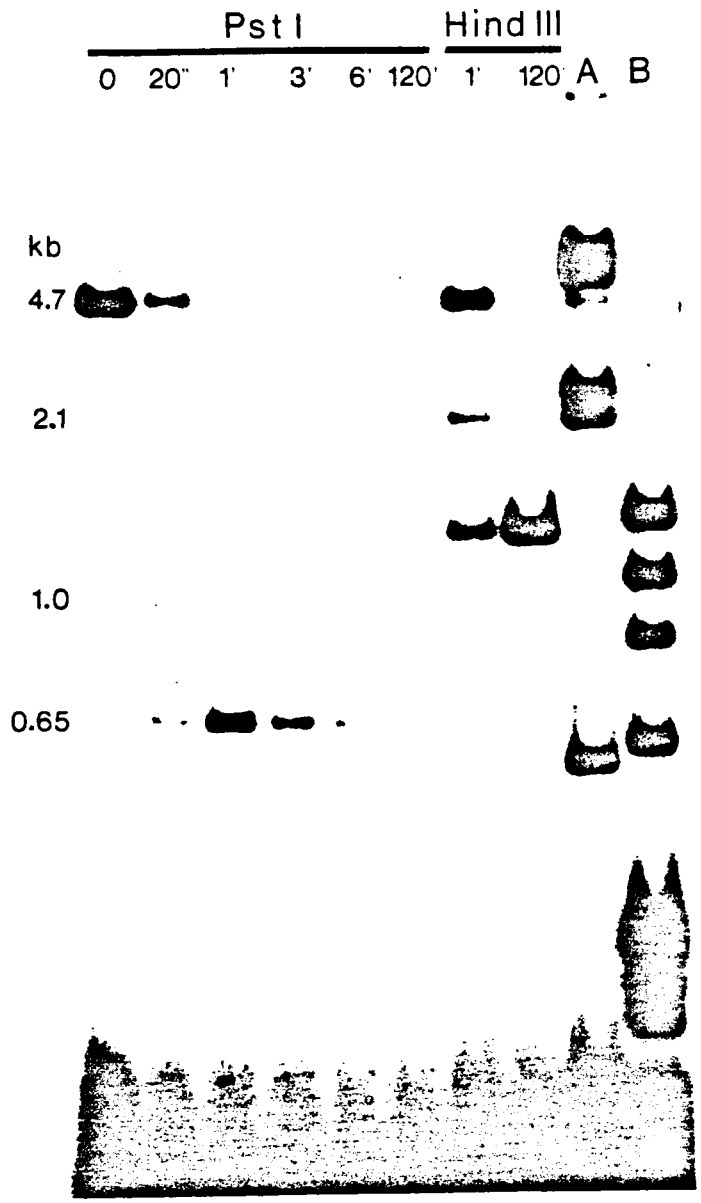



Figure B.2. Smith and Birnstiel gel. One of several gels used to generate a restriction endonuclease map of FG5 by the method of Smith and Birnstiel. Clone FG5 was cut at Sal I in the multiple cloning region and the DNA was labeled using polynucleotide kinase. The DNA was then cut with Bam HI, a site 5 bases removed from the Sal I site. This produced a DNA fragment labeled on only one end proximal to the cloned insert. The DNA was digested with limited amounts of restriction enzymes Pst 1 or Hind III. Aliquots were removed and electrophoresed on a 2% agarose gel. Lambda Hind III and Phi X174 Hae III fragments were labeled and electrophoresed in lanes A and B respectively as molecular weight standards.

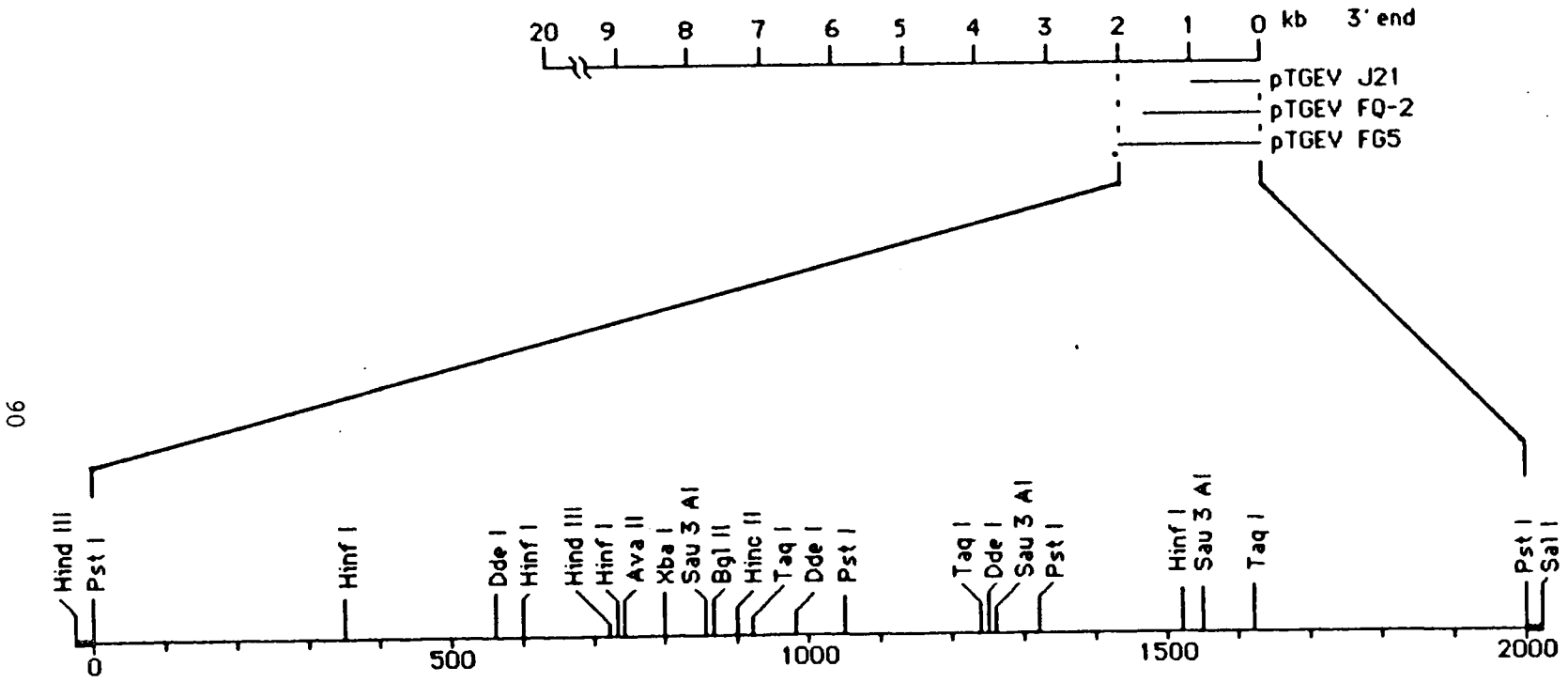



Figure B.3. Original restriction map of clone FG5 as determined by the method of Smith and Birnstiel. Restriction endonuclease map of the 3' proximal 2 Kb of the TGEV genome as determined by the method of Smith and Birnstiel. The TGEV genome which is represented by the line at the top, is 20 Kb in size. The relationship between clone J21 and clone FG5 is shown by the solid lines under the TGEV genome. The expanded line at the bottom of the figure is the restriction endonuclease map for clone FG5.

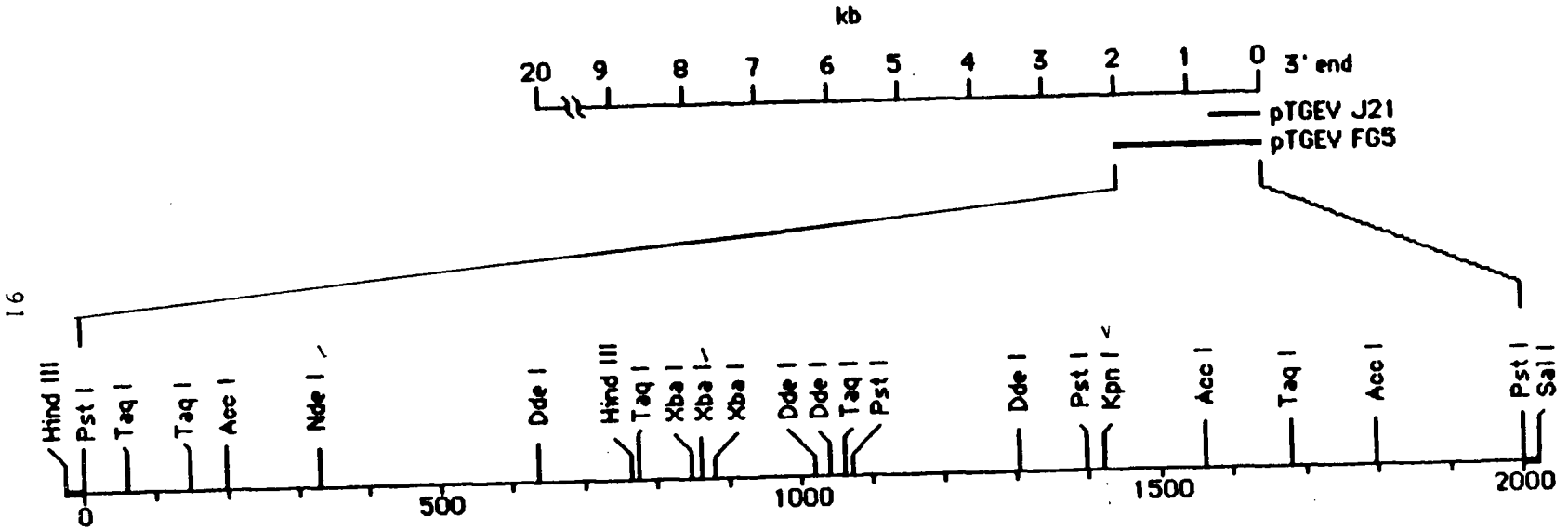



Figure B.4. Restriction map of clone FG5 as modified from nucleotide sequence data. Restriction endonuclease map of the 3' proximal 2 Kb of the TGEV genome. This is a revised restriction endonuclease map of clone FG5 based upon primary sequence data. The map only depicts restriction enzyme sites of those enzymes used to sequence clone FG5.

map of clone FG5 based on the data generated by sequencing the clone. The maps are similar but not identical. For instance, there are, surprisingly, three Xba I sites within a twenty base stretch beginning at base 848. This is clearly beyond the resolving power of the agarose gels employed and is undoubtedly the best example in the world (!) of why one must take care to sequence through any given restriction enzyme site.

Clone FG5 and portions of clone J21 were sequenced using the method of Maxam and Gilbert (Maxam and Gilbert, 1980) as modified by Wen Yang (breakdown at G, T, and C) and Gail Wertz (breakdown at A). Figure B.5 is a schematic diagram of the protocol utilized for the breakdown of DNA. Guanosine was modified with dimethyl sulfate for 90 seconds. Adenine was modified by adding 98% formic acid for 6 minutes. Thymidine was modified by hydrazine for 10 minutes. Cytosine was modified with hydrazine for 14 minutes. Following modification, samples were ethanol precipitated twice, rinsed twice with 80% ethanol, and broken down by piperidine at 90°C for 30 minutes. After breakdown, samples were lyophilized and stored at -20°C until use. Figure B.6 depicts one portion of a sequencing gel. The sequence shown is derived from the Hind III internal site of FG5. The fragment was labeled by a 3' fill-in reaction, and the sequence was read toward the 5' end of the clone.

Table B.1 shows the amino acid composition of the nucleocapsid protein as deduced from the primary nucleic acid sequence.

Figures B.7 and B.8 are diagon plots examining the relatedness of the amino acid sequences for the nucleocapsid proteins of TGEV, IBV,

REACTIONS FOR MAXAM AND GILBERT SEQUENCING[a]. D. BRIAN'S LAB.

| Date | Code | | Sample | |
|---|---|---|---|---|
| | A > G | G | T & C | C |
| | 5 ul DNA | 5 ul DNA* | 10 ul DNA* | 5 ul DNA* |
| | 25 ul 95% Formic Acid | 5 ul $H_2O$ | | |
| | 5-6 min | 200 ul DMS buffer | 10 ul $H_2O$ | 15 ul 5M NaCl |
| | 20°C | 0°C | 0°C | 0°C |
| | 200 ml Hydrazine Stop | 1 ul DMS, votex | 30 ul Hydrazine, vortex | |
| | | | 8-10 min. | 10-15 min. |
| | | | 20°C, 5 minutes | |
| | | 20°C, 90 seconds | 200 ul Hydrazine stop (0°C) | |
| | | 50 ul DMS stop (0°C) | | |
| | | 800 ul EtOH | | |
| | | -80°C, 8 min | | |
| | | Spin, 8 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 (5 minutes) | | |
| | | 100 ul, 1M piperidine (add 200 ul to 2 ml $H_2O$ 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 $H_2O$, vortex, spin, freeze-dry (overnight) | | |
| | | 50 ul $H_2O$, 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 $H_2O$ | | |
| | | Load 2 ul onto sequencing gel with the order of G, A/G, T/C, C | | |

[a]These are the reaction conditons used by P. Kapke.

Figure B.5. Schematic diagram of protocol utilized for breakdown of DNA.

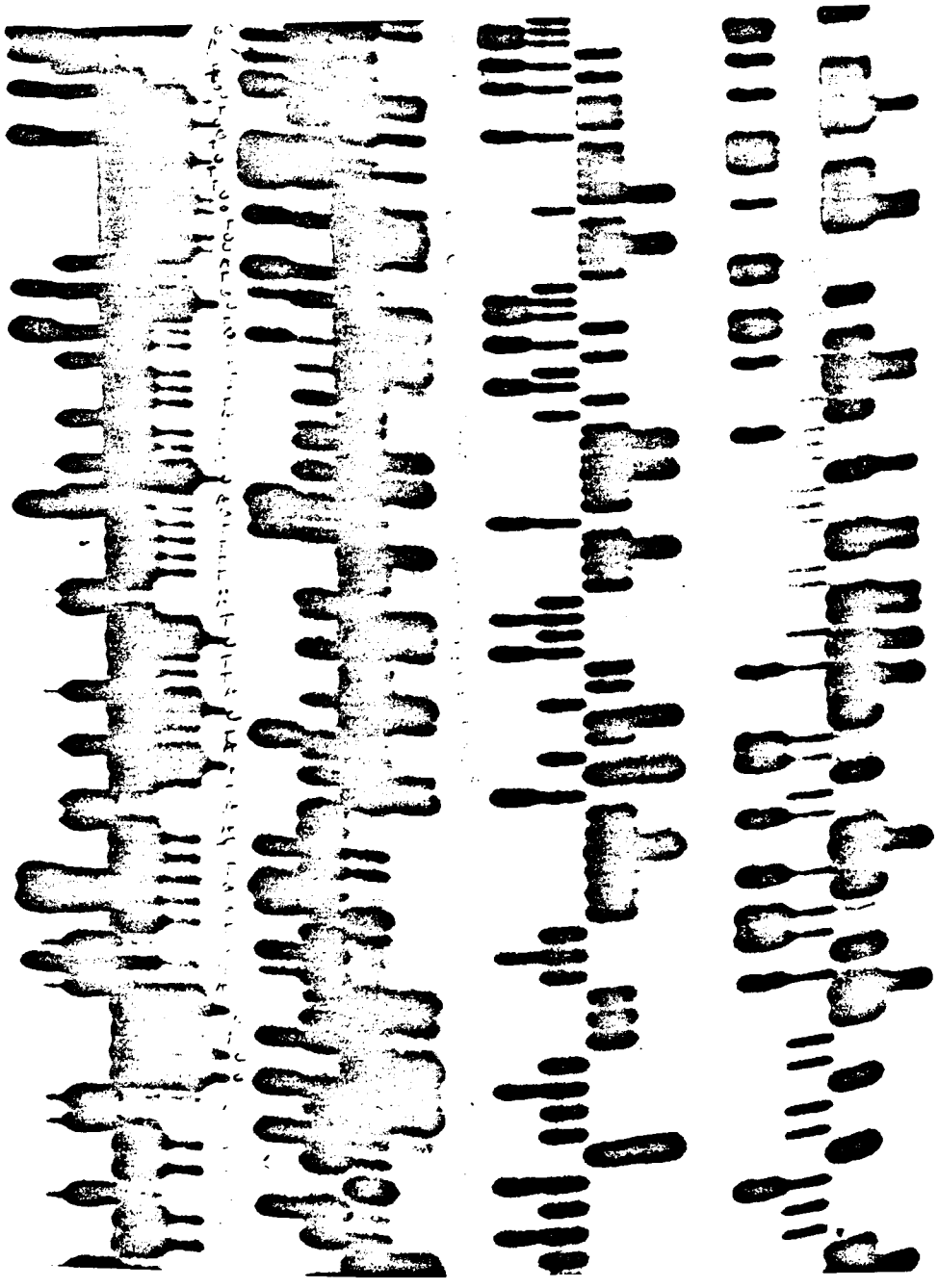


Figure B.6. Autoradiograph of a DNA sequencing gel. A section of clone FG5 sequenced by the method of Maxam and Gilbert is shown above. Clone FG5 was cut with restriction endonuclease Hind III, labeled by a 3' fill in reaction, and recut with Dde I. The labeled fragments were separated on a 1% agarose gel prior to DNA degradation and sequencing.

Table B.1. Amino acid composition of TGEV nucleocapsid protein. This information was derived from the primary nucleic acid sequence.

| | | | | | |
|---|---|---|---|---|---|
| Ala | 21 | (5.5) | Leu | 18 | ( 4.7) |
| Arg | 34 | (8.9) | Lys | 35 | ( 9.2) |
| Asn | 25 | (6.5) | Met | 3 | ( 0.8) |
| Asp | 26 | (6.8) | Phe | 16 | ( 4.2) |
| Cys | 2 | (0.5) | Pro | 15 | ( 3.9) |
| Gln | 26 | (6.8) | Ser | 39 | (10.2) |
| Glu | 20 | (5.2) | Thr | 22 | ( 5.8) |
| Gly | 27 | (7.1) | Trp | 6 | ( 1.6) |
| His | 5 | (1.3) | Tyr | 10 | ( 2.6) |
| Ile | 10 | (2.6) | Val | 22 | ( 5.8) |
| End | 0 | (0.0) | | | |
| Acidic | (Asp + Glu) | | | 46 | (12.0) |
| Basic | (Arg + Lys) | | | 69 | (18.1) |
| Aromatic | (Phe + Trp + Tyr) | | | 32 | ( 8.4) |
| Hydrophobic | (Aromatic + Ile + Leu + Met + Val) | | | 85 | (22.3) |

Molecular Weight = 43,426.

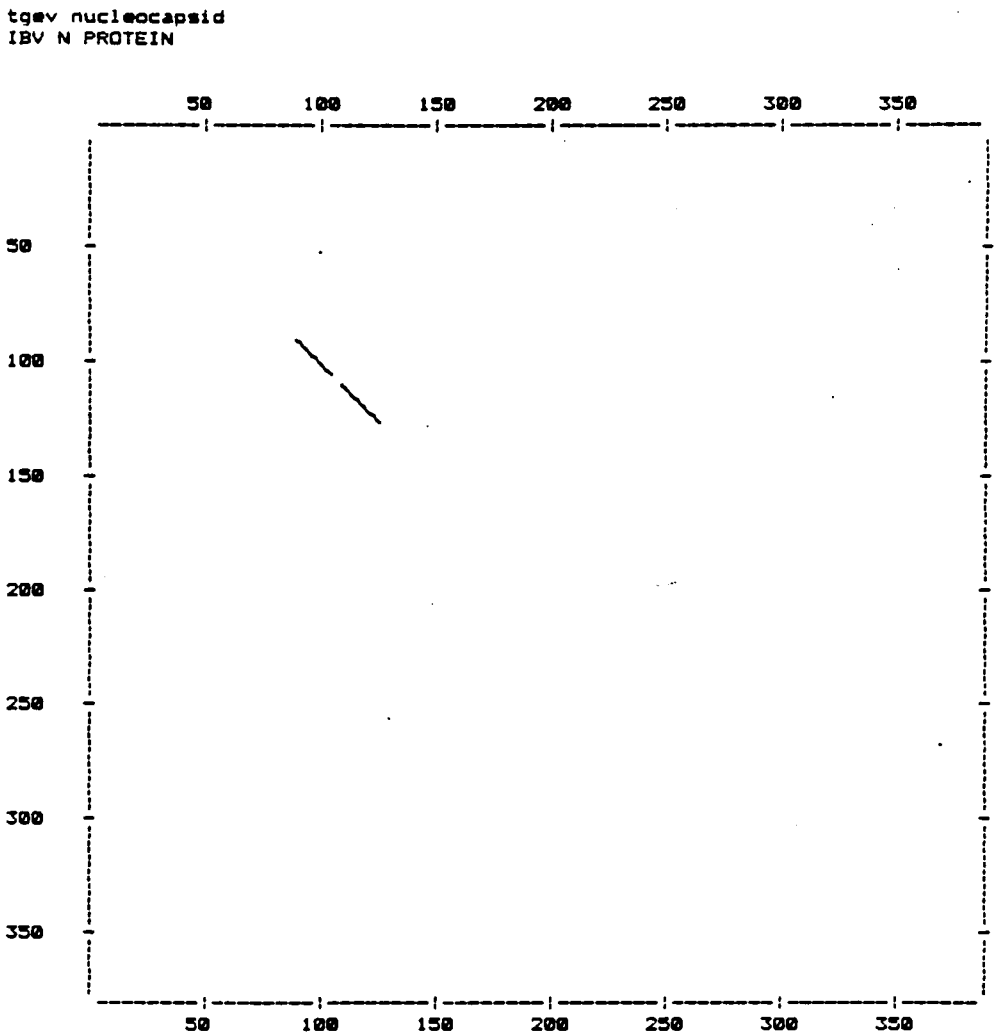



Figure B.7. Diagon plot comparing amino acid homologies of TGEV and IBV nucleocapsid proteins.

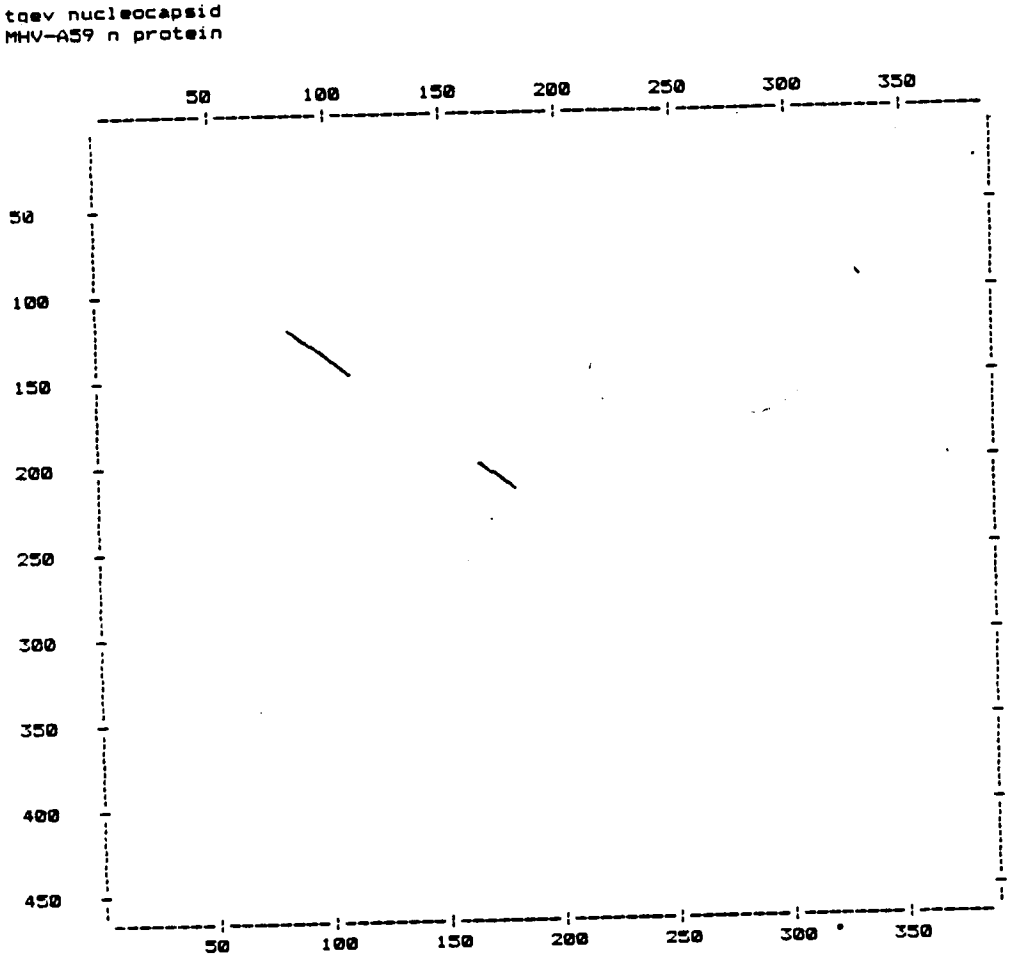



Figure B.8. Diagon plot comparing amino acid homologies of TGEV and MHV-A59.

and MHV-A59. If two sequences are closely related they will have significant regions showing colinear sequence homology. Looking at a span of 9 amino acids, and a minimum crition of 65% match (the parameters used in this analysis) it is obvious that there is some homology shared between TGEV and the two other coronaviruses IBV and MHV-A59. The region of homology between TGEV and IBV is in the area of high conservation discussed in Chapter III. The areas shared between TGEV and MHV-A59 are in the area of high conservation discussed in Chapter III and in one other serine rich region.

From a computer search of Genbank (Beckman Microgenie Program, March 1985 version), a homology was found between the nucleocapsid protein of TGEV and the hypothetical M2 protein of influenza B virus (Figure B.9). The significance of this homology is not known at this time.

Figures B.10 and B.11 are diagon plots comparing the nucleic acid sequences that code for the nucleocapsid protein of TGEV, IBV, and MHV-A59. Using the parameters of 65% minimum match and a span of 9 nucleotides, it is obvious that despite some amino acid homologies between TGEV and IBV, and TGEV and MHV-A59, there is no homology at the nucleotide level.

Table B.2 shows the amino acid composition of the hypothetical hydrophobic protein.

From a computer search of Genbank, a stretch of significant amino acid homology between the hypothetical hydrophobic protein (HHP) and a hypothetical protein described for the mouse mammary tumor virus (HP MMTV) was found (Figure B.12). The MMTV protein was deduced from DNA

```
tgev nucleocapsid
Probable matrix (M2) protein - Human influenza B virus (strain

212     S  R  S  K  S  K  E     R  S  N      221
52      S  R  S  K  S  K  E  W  R  R  N       62
                           *     *

Matches = 9          Length = 11          Matches/length =  81.8 percent
```

Figure B.9. Amino acid homology shared between TGEV nucleocapsid protein and the M2 matrix protein of human influenza B. The homology begins at amino acid number 212 in TGEV from the initiating methionine residue and amino acid number 52 in influenza from the initiating methionine residue.

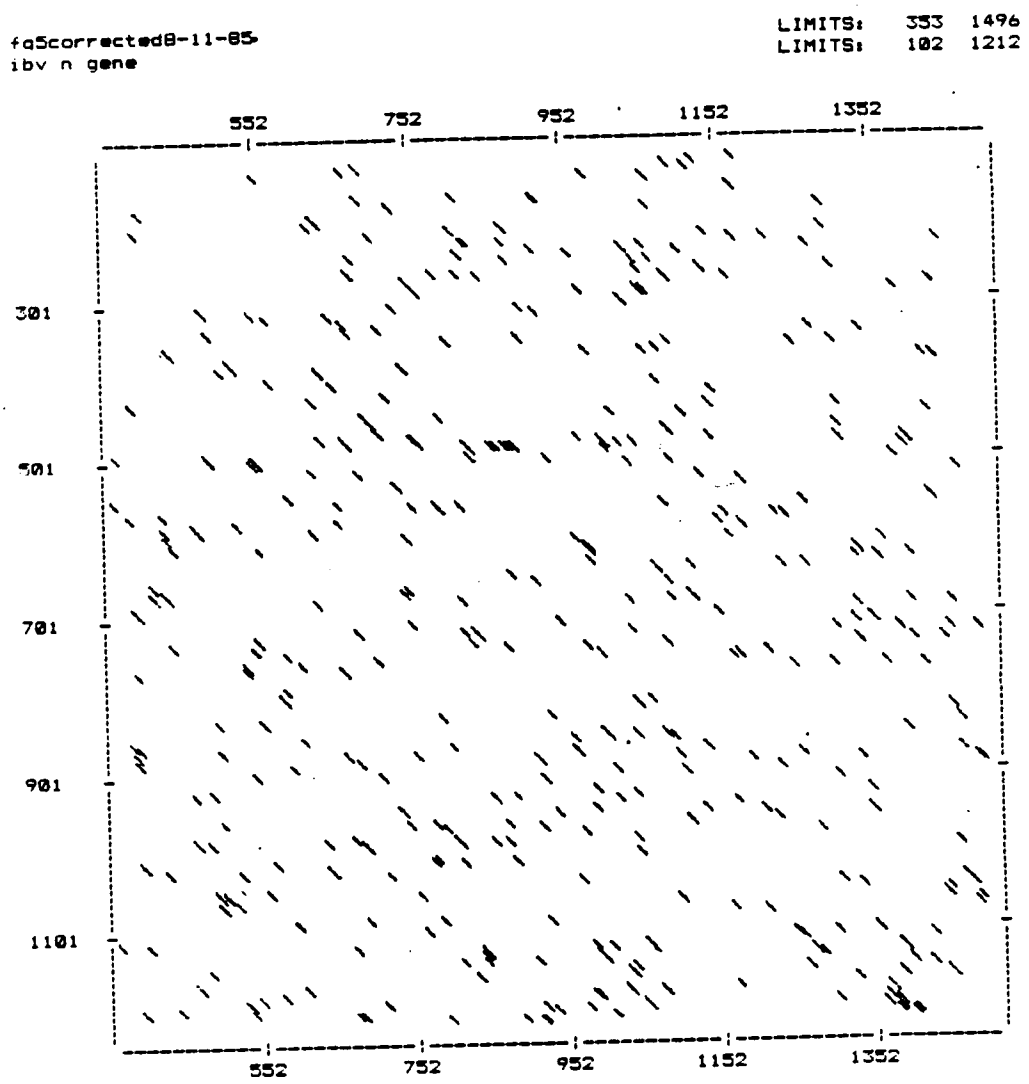



Figure B.10. Diagon plot comparing the nucleic acid sequences that code for the nucleocapsid proteins of TGEV and IBV.

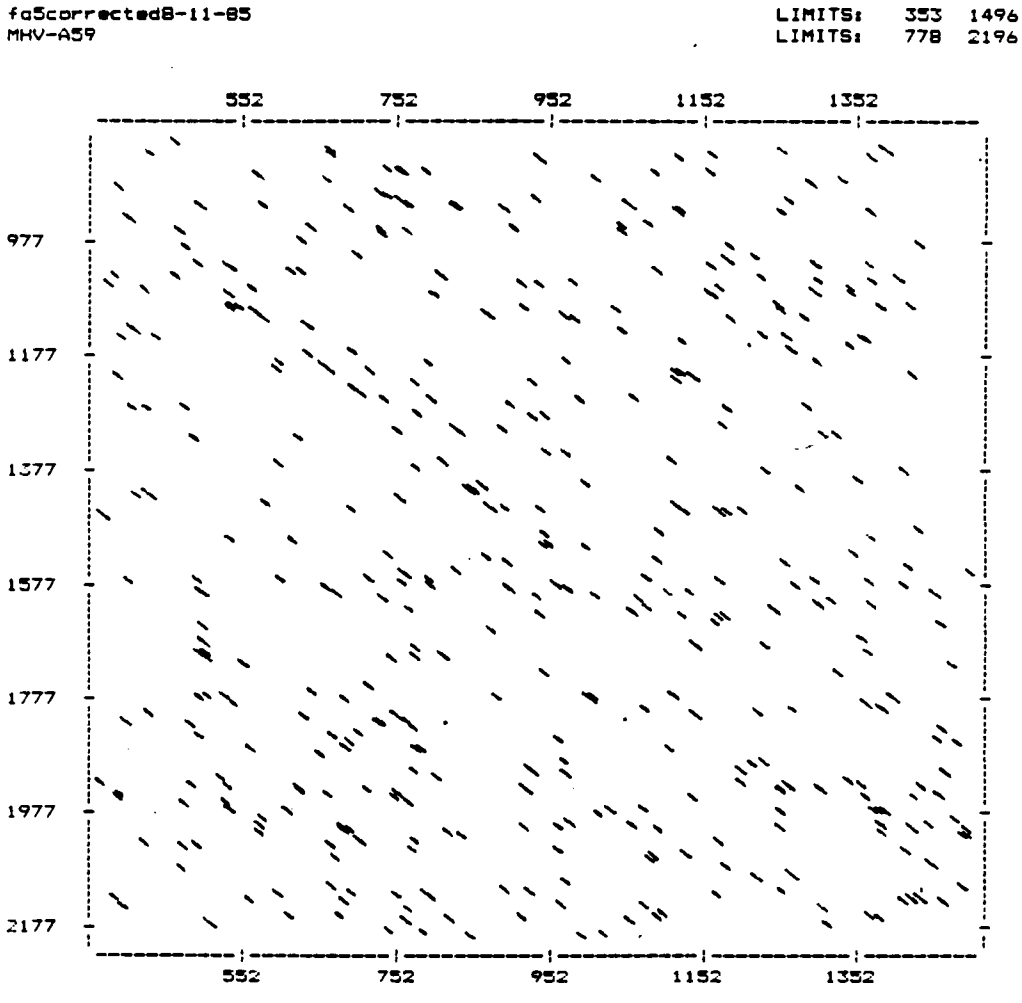



Figure B.11. Diagon plot comparing the nucleic acid sequences that code for the nucleocapsid proteins of TGEV and MHV-A59.

Table B.2. Amino acid composition of the hypothetical hydrophobic protein. An open reading frame (ORF) with a start codon was found between the nucleocapsid gene and the 3' end of the genome. This ORF could code for a novel protein 79 amino acids in size.

| | | | | | |
|---|---|---|---|---|---|
| Ala | 2 | (2.6) | Leu | 24 | (30.8) |
| Arg | 6 | (7.7) | Lys | 2 | ( 2.6) |
| Asn | 3 | (3.8) | Met | 1 | ( 1.3) |
| Asp | 2 | (2.6) | Phe | 6 | ( 7.7) |
| Cys | 2 | (2.6) | Pro | 0 | ( 0.0) |
| Gln | 1 | (1.3) | Ser | 2 | ( 2.6) |
| Glu | 2 | (2.6) | Thr | 4 | ( 5.1) |
| Gly | 2 | (2.6) | Trp | 0 | ( 0.0) |
| His | 2 | (2.6) | Tyr | 2 | ( 2.6) |
| Ile | 5 | (6.4) | Val | 10 | (12.8) |
| End | 0 | (0.0) | | | |
| Acidic | (Asp + Glu) | | | 4 | ( 5.1) |
| Basic | (Arg + Lys) | | | 8 | (10.3) |
| Aromatic | (Phe + Trp + Tyr) | | | 8 | (10.3) |
| Hydrophobic | (Aromatic + Ile + Leu + Met + Val) | | | 48 | (61.5) |

Molecular Weight = 4,101

hydrophobic protein
Hypothetical protein A-198 - Mouse mammary tumor virus

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 37 | T | V | N | D | F | N | I | L | Y | R | S | L | A | E | T | | R | L | 53 |
| 117 | T | V | N | G | Y | K | V | L | Y | R | S | L | P | F | R | E | R | L | 134 |
| | | | | * | * | * | * | | | | | | * | * | * | * | | | |

Matches = 10 Length = 18 Matches/length = 55.6 percent

Figure B.12. Amino acid sequence homology shared between the hypothetical protein of TGEV and the hypothetical protein of mouse mammary tumor virus. The amino acids shown in the top line represent the sequence from TGEV.

sequence analysis (Briedis et al., 1982). An mRNA species has been described that is thought to code for this protein but direct evidence for the protein in MMTV is still not available. The significance of the homology between the HHP for TGEV and the HP MMTV is not known at this time.

# 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/in$^2$ (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 $N_2$ 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 $MgCl_2 \cdot 6H_2O$ (f.w. 203)
12.0 g $MgSO_4$ (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 $MnCl_2 \cdot 4H_2O$, 10 mM $CaCl_2 \cdot 2H_2O$, 3 mM HA $CoCl_3$ ( 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 $MnCl_2 \cdot 4H_2O$ (f.w. 198)
1.47 g $CaCl_2 \cdot 2H_2O$ (f.w. 147)
0.80 g $[Co(NH_3)] Cl_3$ (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 $MnCl_2 \cdot 4H_2O$, 10 mM $CaCl_2 \cdot 2H_2O$, 3 mM $HACoCl_3$, 10% redistilled glycerol.)

0.98 g potassium acetate (m.w. = 98.1)
0.45 g KCl (m.w. 74.5)
8.9 g $MnCl_2 \cdot 4H_2O$ (f.w. 198)
1.47 g $CaCl_2 \cdot 2H_2O$ (f.w. 147)
0.80 g HA $\cdot$ $CoCl_3$ (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 x $10^7$ to 7 x $10^7$/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 ul volumes into Falcon 2059 tubes or screw-cap polypropylene tubes (e.g. Nunc), flash-frozen in solid $CO_2$/EtOH or liquid $N_2$, 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.

# APPENDIX D

## 3' END-LABELING WITH REVERSE TRANSCRIPTASE

# 3' END-LABELING WITH REVERSE TRANSCRIPTASE

This is a "fill-in" reaction described to us by Dr. Gail Wertz during her visit Oct. 1984. The object is to fill-in a 5' overhang using one hot base such that the DNA can be sequenced by the method of Maxam and Gilbert. For example:

1. 5'..AAGCTT..3' Hind III 5' AGCTT..3'
   3'..TTCGAA..5' A..5'

2. 5' AGCTT..3' R.T. 5 'AGCTT..3'
   A..5' ATP [ $-^{32}P$] A*A..5'

Notes:

1. Be careful of cutters that have repeats (e.g. AA) since you cannot tolerate a mixture of molecules some labeled with one base and some with two. Fill-in will not work to completion even with a cold chase.

2. R.T. will not fill in the last base in a cut.

3. Avoid using G. If necessary, fill in with cold G first and label with the next base.

Procedure:

1. Cut about 50 ug of plasmid with enzyme of choice.
   e.g. 50 ul plasmid
   10 ul enzyme
   8 ul 10 x buffer
   12 ul $H_2O$
   80 ul total

2. Heat at 65°C, 5' to inactivate first enzyme.

3. Add reverse transcriptase, about 4 ul of 15-25 u/ul from Life Sciences. Add 50-100 Ci x $^{32}P$ XTP, incubate 37°C, 30-40'. (Fill in the first base. No need for a cold chase.)

4. Heat at 65-70°C, 5' to inactivate reverse transcriptase.

5. Put through Sephadex G-50 spun column equilibrated in TE buffer.

6. Elute in 100 ul TE buffer.

7. Adjust salt and buffer for cutting with desired second enzyme.

8. Add second enzyme, digest 37°, 60!

9. Separate fragments on 6-8 polyacrylamide gel. (About 10 MA constant current, 10-12 hrs.)

10. Do a wet autoradiogram (10 min. exposure) and cut out band to elute electrophoretically or by crushing, or electroelute onto DE81 paper.

11. Clean up eluate on NACS (BRL) column. Ethanol precipitate 3x with 70% ethanol to remove salt. Wash pellet in 70% ethanol, lyophilize and sequence.

Paul's Modifications:

A. Usually 50 ug of plasmid was digested in 300 ul of buffer. Digests were either carried out simultaneously with two enzymes, such as Hind III and Sal I, or were carried out in sequential digests (digesting with the low salt restriction enzyme first, adjusting the salt concentration and then digesting with the high salt restriction enzyme). Unless the enzyme used in the first digest was known to have star activity, the enzyme used in the first digest was not removed by phenol extraction or heat inactivation. After digestion, sample was phenol extracted and ethanol precipitated.

B. After filling in with reverse transcriptase, the reaction was stopped by adding 4 ul 0.5 M EDTA. Loading dye was added directly to the sample just prior to electrophoresis.

# VITA

Paul Alan Kapke was born in Mason City, Iowa, September 23, 1952. He graduated from Mason City High School in 1971. He attended North Iowa Area Community College where he received an Associate of Arts degree in 1973. He attended the University of Iowa where he received a Bachelor of Science degree in Microbiology and a Bachelor of Science degree in General Science in 1975. He entered graduate school at the University of Kentucky in Lexington, Kentucky, where he received a Master of Science degree in Microbiology. In December, 1985 he received a Doctor of Philosophy degree in Microbiology, The University of Tennessee, Knoxville, Tennessee.

The author is a member of the American Society for Microbiology, and the American Association for the Advancement of Science.

He is married to the former Jacquelyn Nollsch of Iowa City, Iowa, and has a son, Nathan.