



Transmissible gastroenteritis virus : genome and messenger RNA sequence
by Quentin Boyd Reuer

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in
Microbiology

Montana State University

© Copyright by Quentin Boyd Reuer (1988)

Abstract:

The genetic structure of the pathogenic Miller strain of transmissible gastroenteritis virus (TGEV) was studied at the molecular level. Subgenomic RNAs 6 and 7 and the 3' 7.3 kb of the viral genome were reverse transcribed into cDNA. Complementary DNA clones were mapped; maps suggested that RNA 7 was a subset of RNA 6, and the maps of both subgenomic RNAs were identical to the map of the 3' region of the virion cDNA. Restriction fragments of the cDNA clones were sequenced. Common 5' leader sequences were found in RNA 6- and RNA 7-specific cDNAs but not in the corresponding region of virion cDNA. The gene encoding the matrix (E1) protein of TGEV (Miller strain) was found in the virion RNA and RNA 6 nucleotide sequence. The 29.4 kd primary product of this gene possessed a 17-residue hydrophobic leader peptide. Hydrophilicity analysis of the protein revealed internal membrane-spanning regions, an amphiphilic C-terminal half, and a hydrophilic C-terminus.

A gene encoding the nucleocapsid (N) protein of TGEV (Miller strain) was found in the virion RNA, RNA 6, and RNA 7 nucleotide sequences. The predicted molecular weight of the serine-rich polypeptide was 43.4 kd. Clusters of charged residues were found over the entire amino acid sequence of N. The sequence of virion cDNA contained the 3' 3183 bases of the TGEV (Miller strain) peplomer (E2) gene. Downstream of the E2 gene were open reading frames that may represent the coding regions of TGEV (Miller strain) RNAs 4a, 4b, and 5. Data obtained during this research suggested that TGEV RNAs form a nested set and that the primary products of the RNAs are encoded by the 5' region not found in the next smaller RNA. The presence of 5' leader sequences in RNA 6- and RNA 7-specific cDNA not found in virion cDNA indicates that TGEV subgenomic RNAs may be transcribed by a leader—primed discontinuous process. Analysis of the primary structure of TGEV (Miller strain) structural proteins demonstrates that the virulent strain differs markedly from the attenuated Purdue strain of TGEV. These data will be useful in the development of safe, effective vaccines against porcine transmissible gastroenteritis.

TRANSMISSIBLE GASTROENTERITIS VIRUS: GENOME
AND MESSENGER RNA SEQUENCE

by

Quentin Boyd Reuer

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Microbiology

MONTANA STATE UNIVERSITY
Bozeman, Montana

January, 1988

APPROVAL

of a thesis submitted by

Quentin Boyd Reuer

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

4 January 1988
Date

[Signature]
Chairperson, Graduate Committee

Approved for the Major Department

4 January 1988
Date

Norman L. Reed
Head, Major Department

Approved for the College of Graduate Studies

1-12-88
Date

MB Maloe
Graduate Dean

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under the rules of the Library. I further agree that the copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this thesis should be referred to University Microfilms International, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute copies of the dissertation in and from microfilm and the right to reproduce and distribute by abstract in any format."

Signature



Date

January 4, 1988

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	vi
LIST OF FIGURES.....	vii
ABSTRACT.....	ix
INTRODUCTION.....	1
Biology and Biochemistry of Coronaviruses.....	2
Transmissible Gastroenteritis: Epizootic and Enzootic.....	8
Transmissible Gastroenteritis Virus Proteins....	13
Transmissible Gastroenteritis Virus RNAs.....	16
Goals and Experimental Design.....	18
MATERIALS AND METHODS.....	20
Chemicals, Media, and Buffers.....	20
Virus Strains and Cell Lines.....	24
Cloning and Sequencing Vectors.....	25
Virus Stocks.....	25
Plaque Assay.....	26
Organic Extraction and Recovery of Nucleic Acids.....	26
Virion RNA Production.....	27
Isolation of Polyadenylated Intracellular RNAs.....	29
Urea-Agarose Gel Electrophoresis.....	31
In vitro Translation of Gel-Purified mRNAs.....	32
Immunoprecipitation of Virus-Specific Proteins..	32
SDS-Polyacrylamide Gel Electrophoresis.....	33
Complementary DNA Synthesis.....	34
Hybridization Analysis of cDNA Clones.....	37
Subcloning of cDNA Restriction Fragments.....	39
Sequencing of TGEV-Specific cDNA.....	40

TABLE OF CONTENTS (continued)

	Page
RESULTS.....	43
Production of TGEV Virion RNA.....	43
Synthesis of cDNAs.....	45
Sequencing of cDNAs Representative of TGEV (Miller strain) Virion RNA.....	46
Time Course of Viral RNA Synthesis.....	65
Urea-Agarose Gel Electrophoresis of TGEV mRNAs..	67
Cloning of TGEV RNA 6 and RNA 7.....	72
DISCUSSION.....	84
TGEV (Miller strain) Virion RNA.....	85
TGEV (Miller strain) RNA 6 and RNA 7.....	90
TGEV (Miller strain) Gene Transcription.....	97
CONCLUSIONS.....	101
LITERATURE CITED.....	102

LIST OF TABLES

Table	Page
1. Antigenic cross-reactivity among coronaviruses..	3
2. Composition and pH of buffers and mixtures.....	21
3. Abbreviations for amino acids.....	48

LIST OF FIGURES

Figure	Page
1. Profile of isokinetic 10-30% sucrose gradient used in purification of TGEV (Miller strain) virion RNA.....	44
2. Restriction endonuclease map of TGEV (Miller strain) virion cDNA clones 150, 1561, 1563, and 141 and strategy used in nucleotide sequence determination.....	47
3. Partial nucleotide sequence of TGEV (Miller strain) virion cDNA.....	49
4. Hydrophilicity plot of the precursor to the matrix (E1) protein of TGEV (Miller strain)	59
5. Hydrophilicity plot of the nucleocapsid (N) protein of TGEV (Miller strain).....	61
6. Hydrophilicity profile of the potential product of the open reading frame extending from base 6868 to base 7101 of the virion cDNA sequence given in Figure 3.....	63
7. Positions of amino acid substitutions in the TGEV (Miller strain) peplomer (E2), matrix (E1), and nucleocapsid (N) protein amino acid sequences compared to the primary structures of TGEV (Purdue strain) structural proteins.....	64
8. Time course of TGEV RNA synthesis.....	66
9. Profile of intracellular RNAs from TGEV (Miller strain)-infected ST cells and ST cell ribosomal RNA electrophoresed in urea-agarose.....	68

LIST OF FIGURES (continued)

Figure	Page
10. Densitometric analysis of TGEV (Miller strain) translation products separated in SDS-10% polyacrylamide gels.....	70
11. Densitometric analysis of TGEV (Miller strain)-specific translation products separated in SDS-10% polyacrylamide gels...	71
12. Restriction endonuclease map of TGEV (Miller strain) RNA 6-specific cDNA and strategy used in nucleotide sequence determination..	74
13. Nucleotide sequence of TGEV (Miller strain) RNA 6-specific cDNA.....	75
14. Restriction endonuclease map of TGEV (Miller strain) RNA 7-specific cDNA and strategy used in determination of nucleotide sequence.....	80
15. Nucleotide sequence of TGEV (Miller strain) RNA 7-specific cDNA.....	81
16. Proposed mechanism of TGEV gene transcription..	98

ABSTRACT

The genetic structure of the pathogenic Miller strain of transmissible gastroenteritis virus (TGEV) was studied at the molecular level. Subgenomic RNAs 6 and 7 and the 3' 7.3 kb of the viral genome were reverse transcribed into cDNA. Complementary DNA clones were mapped; maps suggested that RNA 7 was a subset of RNA 6, and the maps of both subgenomic RNAs were identical to the map of the 3' region of the virion cDNA. Restriction fragments of the cDNA clones were sequenced. Common 5' leader sequences were found in RNA 6- and RNA 7-specific cDNAs but not in the corresponding region of virion cDNA. The gene encoding the matrix (E1) protein of TGEV (Miller strain) was found in the virion RNA and RNA 6 nucleotide sequence. The 29.4 kd primary product of this gene possessed a 17-residue hydrophobic leader peptide. Hydrophilicity analysis of the protein revealed internal membrane-spanning regions, an amphiphilic C-terminal half, and a hydrophilic C-terminus. A gene encoding the nucleocapsid (N) protein of TGEV (Miller strain) was found in the virion RNA, RNA 6, and RNA 7 nucleotide sequences. The predicted molecular weight of the serine-rich polypeptide was 43.4 kd. Clusters of charged residues were found over the entire amino acid sequence of N. The sequence of virion cDNA contained the 3' 3183 bases of the TGEV (Miller strain) peplomer (E2) gene. Downstream of the E2 gene were open reading frames that may represent the coding regions of TGEV (Miller strain) RNAs 4a, 4b, and 5. Data obtained during this research suggested that TGEV RNAs form a nested set and that the primary products of the RNAs are encoded by the 5' region not found in the next smaller RNA. The presence of 5' leader sequences in RNA 6- and RNA 7-specific cDNA not found in virion cDNA indicates that TGEV subgenomic RNAs may be transcribed by a leader-primed discontinuous process. Analysis of the primary structure of TGEV (Miller strain) structural proteins demonstrates that the virulent strain differs markedly from the attenuated Purdue strain of TGEV. These data will be useful in the development of safe, effective vaccines against porcine transmissible gastroenteritis.

INTRODUCTION

The development of safe and effective vaccines has been a goal of molecular biologists since the advent of recombinant DNA technology. Many types of virus preparations and means of inoculation have been used in attempts to provide immunity to viral diseases. Exposure to virulent viruses and inoculation with killed virus or attenuated virus vaccines have been used, although disease development, inadequate protection, and the reversion of attenuated viruses to pathogenic forms have been dangers inherent in the use of these preparations. Risks associated with immunization might be reduced or eliminated upon employment of subunit vaccines. In order to develop reliable viral subunit vaccines, the biological and biochemical characteristics of virulent and attenuated forms of the virus must be studied. Once the genetic organization of the virus and the molecular basis of its pathogenicity are understood, genetically altered viruses that induce a protective immune response but fail to cause disease might then be constructed. The purpose of my research was to study at the molecular level the genetic

structure of the Miller strain of transmissible gastroenteritis virus (TGEV), a pathogen of swine.

Biology and Biochemistry of Coronaviruses

Members of the family Coronaviridae are spherical, pleomorphic particles 60-220 nanometers in diameter which bear characteristic club-shaped surface projections. The corona-like appearance of these projections led to the creation of the family Coronaviridae by the International Committee on the Taxonomy of Viruses in 1975 (91). Coronaviruses have been placed in antigenic groups on the basis of cross-reactivity in serological tests (Table 1) (59,76,88,95). The mammalian coronaviruses fall into two groups, while the avian coronaviruses, infectious bronchitis virus (IBV) and turkey coronavirus (TCV), compose the remaining two groups. Several strains of two coronaviruses, IBV of chickens and murine hepatitis virus (MHV), have been intensively studied and serve as models of this family.

Coronaviruses are widespread pathogens of many species of mammals and birds and cause acute and chronic diseases. Targets of infection include the gastrointestinal tract, respiratory system, liver and nervous system. Marked tissue tropism is characteristic of the coronaviruses. Investigations of the epidemiology and pathogenesis of

Table 1. Antigenic cross-reactivity among coronaviruses

Antigenic group	Virus*	Host
I	HCV-229E	Human
	TGEV	Pig
	CCV	Dog
	FECV	Cat
	FIPV	Cat
II	HCV-OC43	Human
	MHV	Mouse
	HEV	Pig
	BCV	Cow
	RbCV	Rabbit
III	IBV	Chicken
IV	TCV	Turkey
Unclassified	HECV	Human

* Abbreviations: HCV-229E, human respiratory coronavirus; TGEV, transmissible gastroenteritis virus; CCV, canine coronavirus; FECV, feline enteric coronavirus; FIPV, feline infectious peritonitis virus; HCV-OC43, human respiratory coronavirus; MHV, mouse hepatitis virus; HEV, hemagglutinating encephalomyelitis virus; BCV, bovine coronavirus; RbCV, rabbit coronavirus; IBV, infectious bronchitis virus; TCV, turkey coronavirus; HECV, human enteric coronavirus.

The table was developed with references 59,76,88,95.

coronavirus infections have been impeded by the difficulty of isolating coronaviruses from diseased hosts.

Studies of coronaviruses have demonstrated several unique features in RNA transcription, protein composition, and virus assembly (73,75,76,88). Coronaviruses multiply

exclusively in the cytoplasm of infected cells (98). Unlike many other types of viruses, coronaviruses do not induce rapid inhibition of host cell macromolecular synthesis. Productive infection of a susceptible cell by a coronavirus usually results in cell death due to fusion or lysis, although persistent coronavirus infections can readily be established in vitro and in vivo. Some persistently infected cells synthesize viral antigens and release infectious virus particles.

Coronavirions assemble by budding at internal membranes of host cells (16). Coronavirions contain an envelope derived from the endoplasmic reticulum and Golgi apparatus of host cells. Virions are released from cells by fusion of post-Golgi vesicles with the plasma membrane (53,88,91,95).

The genome of coronaviruses is a single-stranded, polyadenylated, colinear RNA of 6-8 megadaltons (62,76,88). Virion RNA is infectious and is believed to encode an RNA-dependent RNA polymerase. Negative-sense copies of the genome are synthesized by this enzyme in infected cells, and from these templates genomic and subgenomic RNAs are transcribed.

MHV and IBV produce 5-6 polyadenylated subgenomic mRNAs in infected cells (43,72,88). These RNAs are made in unequal amounts (48,80,94). Regulatory mechanisms controlling coronavirus mRNA synthesis have not been

identified. MHV and IBV mRNAs form a nested set; the RNAs possess common 3' termini and extend for different lengths in the 5' direction. Only the 5'-terminal region not found in smaller species of mRNA is translated (88,91). Coding assignments of the mRNAs of several coronaviruses have been established (35,72,82). A common 5' leader sequence at least 70 nucleotides in length has been found in the genomic and subgenomic RNAs of MHV (43,78,79). UV transcriptional mapping studies have shown that the synthesis of each MHV subgenomic RNA is initiated independently, suggesting that the RNAs are not spliced from larger precursors (34). Rather, the leader sequence is joined to the body sequence of mRNAs by discontinuous transcription (78). In this process, the leader sequence may serve as a primer for transcription of the mRNA body sequence by a virus-specific RNA-dependent RNA polymerase (43,78,79).

Coronavirus particles contain from 3 to 7 structural proteins, which can be grouped into 3 functional classes (76,88). The nucleocapsid (N) protein is a phosphorylated, basic polypeptide of 45-60 kd that encapsidates the virion RNA to form a long, flexible structure with helical symmetry (12,48,75,88). In intact virions, N protein is resistant to treatment with bromelain (12) and pronase (93), indicating that it is located internally. N protein is the only virion protein to display significant

phosphorylation (83). In vitro studies by Siddell et al. have detected a protein kinase activity associated with the coronavirus JHM virion that specifically phosphorylates the virion nucleocapsid protein (74).

A second class of structural polypeptides is a heterogenous transmembrane glycoprotein species, E1, or matrix (M) protein, of 25-35 kd (1,7,12,49,75,92). This protein has been intensively studied in several coronaviruses and appears to possess three domains: a glycosylated hydrophilic region that extends outside the viral envelope, a hydrophobic region which extends through the viral membrane, and a third domain that possibly interacts with viral RNA within virus particles. E1 is not transported to the plasma membrane of infected cells as are other viral glycoproteins, but accumulates in the Golgi apparatus. The protein may bind the nucleocapsid to the viral envelope as the coronavirion buds, a function in common with that of the nonglycosylated matrix proteins of orthomyxo-, paramyxo-, and rhabdoviruses. Although E1 is a membrane protein, in MHV and IBV it lacks a signal peptide that is cleaved following translocation. Instead, the matrix proteins of these coronaviruses may be inserted into membranes by recognition of the internal hydrophobic region (65). E1 is glycosylated to varying degrees; this results in the protein appearing as multiple bands in sodium dodecyl sulfate-polyacrylamide gels of infected cell lysate

or dissociated virions. The N-terminal region of E1 proteins of MHV and bovine coronavirus (BCV) possess oligo- and polysaccharides O-linked to serine and threonine residues (56,58). O-linked glycosylation, unusual among viral glycoproteins, takes place at the Golgi apparatus of infected cells. Antibodies to E1 are capable of neutralizing viral infectivity only in the presence of complement (15).

The third class of polypeptides is a large (125-220 kd) complex glycoprotein, E2, which forms the characteristic surface projections of coronaviruses (7,12,13,21,28,40, 49,70,92,93). These projections, or peplomers, are responsible for attachment of coronaviruses to cells, induction of neutralizing antibodies, and the fusion of infected cells into syncytia. The biological activity of peplomers determines the virulence and tissue tropism of coronaviruses. Almost all of E2 is located outside the viral membrane; only a small anchor region of the peplomer is embedded in the viral envelope. The molecule is transported to the plasma membrane of infected cells. Cells displaying E2 on their surface are susceptible to cell-mediated cytotoxicity. The peplomer of IBV is composed of two or three copies each of glycoproteins S1 (90 kd) and S2 (84 kd) (13). A C-terminal hydrophobic domain of S2 secures the peplomer to the viral envelope. S1, the target of neutralizing and

hemagglutinating antibodies, is non-covalently linked to S2. S1 is necessary for infectivity of the virus but need not be present for attachment of the virion to cells. Comparison of the amino acid sequences of different IBV strains suggests that the neutralization epitopes are located near the S1 N terminus.

Bovine coronavirus (BCV) virions possess two different peplomeric glycoproteins (40). A 190 kd peplomer glycoprotein appears to be composed of 100 and 120 kd subunits. A second, smaller peplomer, a dimer of 65 kd glycopeptides joined by disulfide bonds, is responsible for the hemagglutinating activity of the virus. Both peplomers can elicit the production of BCV-neutralizing antibodies.

Cleavage of the peplomeric glycoprotein may be required for coronavirus infectivity and cytopathic effects (31,88,89). Trypsinization of peplomers in vitro increases the infectivity of virus particles and dramatically increases the yield of infectious virus from cells infected with trypsinized stock virus (84,87). Cells that make infectious coronaviruses without added trypsin possibly cleave peplomer proteins with cellular proteases in the Golgi apparatus or at the plasma membrane.

Transmissible Gastroenteritis: Epizootic and Enzootic

Transmissible gastroenteritis (TGE) is a disease of swine first reported in Indiana in 1946, and it is now

found in most swine-producing countries of the world (29). The loss of animals by swine producers makes TGE economically important. The cause of the disease, transmissible gastroenteritis virus (TGEV), is a porcine coronavirus that infects and destroys the absorptive epithelial cells of the small intestine. TGEV can be transmitted to a swine unit in a number of ways, including contaminated fomites, birds, and recently recovered swine that appear healthy but are still shedding the virus. Ingestion is the normal route of exposure to the pathogen. TGEV virions display resistance to low pH, trypsin, and bile, making it possible for the virus to maintain infectivity during passage through the alimentary tract. Two forms of TGE, epizootic and enzootic have been described (29).

Epizootic TGE affects swine of all ages; morbidity is virtually 100% in exposed herds. The severity of the disease is greatest in newborn pigs which, during epidemics, can suffer mortality rates of up to 100% (5,29). Newborn animals experience vomiting, severe diarrhea, and subsequent dehydration resulting in rapid weight loss. Extreme thirst is characteristic of TGE in nursing swine. Upon autopsy, nursing piglets display villus atrophy in the jejunum and ileum and undigested milk curds throughout the gastrointestinal tract. Fluorescent antibody tests often reveal the presence of viral antigens in villus epithelial

cells. Laboratory diagnosis may also include electron microscopic examination of small intestinal tissues for the presence of coronavirus particles. There is no practical method of treating young pigs; replacement fluid therapy has been successful in treating laboratory pigs, but the method is labor-intensive and not applicable in large swine units. Clinical signs in older animals include anorexia and profuse, watery diarrhea. Lactating sows may experience agalactia. The incubation period of TGEV prior to the onset of clinical signs in infected animals is 18-24 hours. Pigs under 7 days of age usually die 2-7 days after clinical signs appear. The rate of mortality decreases as the age of swine at the time of infection increases. Mortality in 2 to 3 week old pigs approaches 20-30%, while only 3-4% of weaned pigs and fewer than 1% of adult pigs die as a result of the disease.

Enzootic TGE occurs in swine units that practice continuous or nearly continuous farrowing. Previous infection of the herds with TGEV results in establishment of adequate immunity to the virus. Later this immunity declines, and a non-explosive form of TGE develops. Pigs are usually 6 days of age or older when stricken with diarrhea, and not all pigs in the litter may be affected. Vomiting is not always present, and the rate of mortality is often low. Sows may provide some lactogenic immunity; if the level of immunity is high enough, nursing pigs are

protected until weaning. Agalactia experienced by some sows may prevent transfer of adequate colostral and milk antibody to the young. Enzootic TGE is not always recognized by pork producers and veterinarians familiar with the explosive epizootic form of the disease.

Pigs that recover from TGEV infection can shed the virus from the lungs for more than 4 months after the initial infection. Porcine alveolar macrophages are capable of supporting TGEV replication, as demonstrated by positive immunofluorescence, infectious virus release, and interferon synthesis (45). Also, TGEV may be maintained in a latent state in cells of the intestinal villi. Persistently infected animals can then shed virulent TGEV in fecal material.

There is currently no effective vaccine against TGEV. Natural infection of sows with virulent TGEV leads to production of secretory antibodies capable of neutralizing the virus, but intramuscular inoculation of the pathogenic form does not. Induction of lactogenic immunity may depend upon the route of immunization. Ingestion of TGEV results in the infection of the intestinal epithelium (5); it is possible that macrophages in nearby Peyer's patches break down the virus and present viral antigens to migrating T lymphocytes, which then pass these antigens to lymph nodes near secretory glands. By this process large amounts of protective secretory IgA can be produced and provided to

suckling pigs in milk and colostrum. Passive protection provided by immune sows usually prevents the majority of piglets in a litter from developing TGE (6,10,26,95,99). In contrast, intramuscular administration of the virus stimulates production of circulating antibodies, largely of the IgG class. Although the IgG may neutralize TGEV particles, very little of this antibody is found in the milk of sows immunized intramuscularly. Oral or intramuscular administration of attenuated TGEV particles results in secretion of virus-specific antibodies, but this response is not adequately protective against infection with virulent TGEV. Also, an inherent danger in the use of attenuated viruses as vaccines is the possibility of reversion of the viruses to pathogenic forms. TGEV subunit vaccines containing purified viral protein from virulent or attenuated strains have also failed to protect vaccinated animals (23). A greater understanding of the molecular basis of TGEV pathogenicity is necessary if a protective vaccine is to be developed.

The attenuated Purdue strain of TGEV has been studied by several laboratories. This strain was developed by repeated passage of virulent TGEV through cell culture. Although it was produced for use as a vaccine, the Purdue strain is often used in laboratory studies of TGEV because it replicates to a higher titer in vitro than does the low-passage Miller strain. However, because the Purdue strain

is attenuated, its properties may not accurately reflect those of the virulent virus.

Transmissible Gastroenteritis Virus Proteins

TGEV contains three major structural polypeptides (21). The nucleocapsid (N) protein is a phosphorylated molecule of approximately 50 kilodaltons that is associated with TGEV genomic RNA (21). The nucleotide sequence of the N protein gene of the Purdue strain has been determined by Kapke and Brian (38). TGEV is not of the antigenic subgroup of the coronaviruses MHV and IBV, but the predicted amino acid sequence of TGEV (Purdue strain) N protein shows an overall homology of 26 and 27% with IBV and the neurotropic JHM strain of MHV, respectively. A conserved 68 amino acid region was found to be shared by the three viruses. This region is more basic than the overall nucleocapsid protein, and may interact with genomic RNA. Other regions of the N proteins were found to share structural characteristics even though amino acid sequences differed, suggesting the existence of additional conserved functional domains.

The second virion protein is the matrix glycoprotein E1. Using cDNA sequence data, Laude et al. predicted that the primary translation product of the TGEV (Purdue strain) E1 gene is 262 amino acids long with a molecular weight of 29.6 kd (47). A 17 amino acid leader peptide is removed

from E1 during maturation of the polypeptide. This leader sequence may direct passage of E1 into internal membranes of TGEV-infected cells, a means of localization different than that of MHV and IBV matrix proteins. Comparison of the nucleotide sequence of the TGEV (Purdue strain) E1 gene with the E1 genes of MHV (strain A59) and IBV revealed no significant homology. However, the amino acid sequences showed homologies of 38% (TGEV-MHV) and 27% (TGEV-IBV). Three potential membrane-spanning regions were found in the amino acid sequences of the three E1 polypeptides, but only one of these regions displays an equal degree of homology among the coronaviruses studied. The variance may be due to functional differentiation between the three hydrophobic segments. Binding of the hydrophilic region of E1 by antibodies in the presence of complement may result in virus neutralization (15). Two potential sites of N-glycosylation are present in this area, only one of which may be accessible to glycosylation while E1 is associated with the endoplasmic reticulum (47).

The largest TGEV structural protein is the 195-220 kd peplomer glycoprotein, E2. TGEV (Purdue strain) sequence data suggests the molecular weight of the primary translation product of the E2 gene to be 158 kd (60). The carbohydrate moiety is thus approximately 25% of the total molecular size of the polypeptide, a level in agreement with that reported for the IBV peplomer (3). E2 of the

Purdue strain is largely hydrophobic (60). Hydrophobic residues are concentrated in the core of the peplomer. An extremely hydrophobic sequence of 45 residues near the C terminus is the region of the peplomer presumed to anchor the molecule in the viral envelope. This segment has a much higher ratio (24.5%) of cysteine residues than the molecule as a whole (3.4%); this appears to be a distinctive feature of the coronaviruses (3,60). The peplomer domain immediately exterior to the viral envelope contains an eight residue segment that is perfectly conserved in TGEV (Purdue strain) and IBV (3,60). In both viruses, this region of E2 is preceded by N-glycosylation sites. Hydrophilic segments appear to be concentrated in the carboxyl half of E2. Overall homology between the amino acid sequences of IBV and TGEV (Purdue strain) peplomers is 32.3%. Regions of homology are concentrated in the carboxyl half of the molecule, while the amino half shows considerable divergence.

Possibly all complement-independent TGEV-neutralizing antibodies are directed against the peplomer (37,46). Laude et al. (46) used monoclonal antibodies to construct a map of the antigenic determinants of the TGEV peplomer protein. In this model, TGEV peplomers possess four major antigenic sites, fewer than expected of such a large protein. The neutralization-mediating domain is composed of two of these sites; both sites possess a common epitope.

Also, the sites are conserved among the different strains of TGEV tested. Laude's data suggested that the immunodominant site of the peplomer might reside within the neutralization-mediating region.

Transmissible Gastroenteritis Virus RNAs

Six viral mRNA species were detected by Jacobs et al. in porcine cells infected with the Purdue strain of TGEV (35). Their size, as determined by SDS-PAGE, was 23.6 kb (RNA 1), 8.4 kb (RNA 3), 3.8 kb (RNA 4), 3.0 kb (RNA 5), 2.6 kb (RNA 6), and 1.9 kb (RNA 7). The RNAs were translated in vitro. RNA 7 was shown to encode the nucleocapsid protein, while RNA 6 encoded an unglycosylated precursor of the matrix protein. A 24 kd nonstructural protein was the primary translation product of RNA 4. Translation of RNA 3 resulted in 130 and 250 kd proteins and smaller molecules that could be precipitated with a monoclonal antibody directed against the peplomer. No virus-specific translation product was identified for RNA 5.

The intracellular RNAs of swine testicle (ST) cells infected with the virulent Miller strain of TGEV were studied by Andreas Luder (personal communication). Seven virus-specific RNAs were detected. Sizes of the RNAs were predicted by their rate of migration in agarose following denaturation with glyoxal. The predicted lengths of the

RNAs were: 23.0 kb (RNA 1), 8.5 kb (RNA 3), 3.9 kb (RNA 4a), 3.6 kb (RNA 4b), 2.9 kb (RNA 5), 2.6 kb (RNA 6), and 1.8 kb (RNA 7). The full length minus-strand copy of the genome is considered to be RNA 2. Jacobs et al. did not detect the 3.6 kb RNA during their study of TGEV (Purdue strain) RNAs (35). However, the method of RNA numbering used by Jacobs et al. is also used in this thesis.

The kinetics of TGEV (Miller strain) RNA synthesis are not well understood. A proposed mechanism of coronavirus replication suggests two main phases of RNA synthesis (82). Following virion RNA-directed synthesis of an RNA-dependent RNA polymerase, a full-length minus-strand RNA is transcribed from the viral genome. Transcription of subgenomic RNAs from this template could comprise the first phase of virus-specific RNA synthesis. The second phase would occur later in the cycle of infection, when full-length virion RNA is transcribed from the minus-strand template. This may occur just before packaging of the genome and subsequent release of virions from infected cells. ST cells infected with TGEV (Miller strain) reach their maximum yield of virus at 18 h post-infection (67).

Past research has not conclusively demonstrated that TGEV produces a nested set of RNAs in infected cells. It is not known if TGEV mRNAs are transcribed by a leader-primed mechanism similar to that of MHV. Immediately preceeding the E1-encoding and N-encoding regions of TGEV

(Purdue stain) virion RNA, AACTAAAC sequences have been found (47). Laude et al. assumed these consensus sequences to be the start of mRNA transcripts, but did not assign the sequences a function in gene transcription.

In vitro translation studies have provided proof of the messenger function of TGEV intracellular RNAs (35), and identification of translation products based on electrophoretic migration and recognition by TGEV-specific antibodies suggests that the virus possesses a mechanism of expression similar to that of coronaviruses MHV and IBV. This has not been confirmed by determination of the primary structure of TGEV subgenomic mRNAs.

Goals and Experimental Design

The goal of my research was to determine the primary structure of the genome and RNAs 6 and 7 of the pathogenic Miller strain of transmissible gastroenteritis virus (TGEV). I did this by cloning and sequencing DNA copies of these RNAs. I used the sequence data to compare the structure of the virion RNA to that of the two subgenomic RNAs. Sequence data was used to identify open reading frames and predict the amino acid sequences, glycosylation sites, and functions of potential gene products. From a comparison of these results to information obtained from studies of the attenuated Purdue strain of TGEV, I

endeavored to increase our understanding of the molecular basis of TGEV pathogenicity.

MATERIALS AND METHODS

Chemicals, Media, and Buffers

Reagent grade liquid organic chemicals were obtained from J. T. Baker Chemical Co. Other chemicals and reagents were obtained from Sigma Chemical Co. unless otherwise stated in the text. Radioisotopes were purchased from New England Nuclear Corp. Calf intestinal phosphatase, T4 DNA ligase, Escherichia coli DNA polymerase I, terminal transferase, and Klenow fragment were obtained from Bethesda Research Laboratories and Promega. Cell culture media were obtained from Irvine Scientific, and sera were purchased from Hyclone Laboratories.

Cell cultures were maintained in Dulbecco's Modified Eagle's (DME) medium supplemented with 10% (vol/vol) calf serum (DME-10). Infection of cells with TGEV was done in DME supplemented with 2% (vol/vol) fetal bovine serum, 10 mM 3-(N-morpholino)propanesulfonic acid (MOPS), 10 mM N-tris-(hydroxymethyl)methyl-2-aminoethanesulfonic acid (TES), and 10 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES) (DME-2).

The composition and pH of buffers and reaction mixtures are shown in Table 2.

Table 2. Composition and pH of buffers and mixtures.

Buffer	Composition	pH
Annealing buffer	10 mM Tris-HCl, 2 mM EDTA, 150 mM NaCl	7.6
Chase solution	2 mM dCTP, 2 mM dATP, 2 mM dGTP, 2 mM dTTP	
Chloroform	96% chloroform (vol/vol), 4% isopentanol (vol/vol)	
CIP mix	100 mM glycine, 1 mM MgCl ₂ , 1 mM ZnCl ₂ , 1 unit/ul calf intestinal phosphatase (CIP)	10.5
Citrate-urea gel buffer	25 mM citric acid, 9 M urea	3.0
Citrate-urea sample buffer	10 mM citric acid, 6 M urea, 15% (wt/vol) sucrose, 0.005% (wt/vol) bromophenol blue	3.0
Hybridization buffer	50% (vol/vol) formamide, 5X SSPE, 0.4% (vol/vol) SDS, 200 ug/ml calf thymus DNA	
Klenow 10X buffer	100 mM Tris-HCl, 500 mM NaCl	7.5
Klenow d/ddATP mix	300 µM ddATP, 33 µM dCTP, 33 µM dTTP, 33 µM dGTP, 50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl ₂ , 1 mM dithiothreitol (DTT)	7.5
Klenow d/ddCTP mix	66 µM ddCTP, 1.66 µM dCTP, 33 µM dTTP, 33 µM dGTP, 50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl ₂ , 1 mM DTT	7.5
Klenow d/ddGTP mix	66 µM ddGTP, 33 µM dCTP, 33 µM dTTP, 1.66 µM dGTP, 50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl ₂ , 1 mM DTT	7.5

Table 2, continued.

Buffer	Composition	pH
Klenow d/ddTTP mix	117 μ M ddTTP, 33 μ M dCTP, 1.66 μ M dTTP, 33 μ M dGTP, 50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl ₂ , 1 mM DTT	7.5
Ligation buffer	66 mM Tris-HCl, 5 mM MgCl ₂ , 5 mM DTT, 1 mM ATP, 4 units/ml T4 DNA ligase	7.5
NET	10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA	7.6
Nick translation mix	50 mM Tris-HCl, 10 mM MgSO ₄ , 0.1 mM DTT, 500 ug/ml BSA, 1 nM dTTP, 1 nM dGTP, 1 nM dCTP, 2 nM [alpha- ³² P]-deoxyadenosine 5'-triphosphate (NEG-012A), 0.1 ug/ml DNase I, 0.1 units/ul <u>E. coli</u> DNA pol I	7.2
NTE	50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA	7.2
Oligo(dT) elution buffer	10 mM Tris-HCl, 1 mM EDTA, 0.05% (vol/vol) SDS	7.5
Oligo(dT) high-salt buffer	20 mM Tris-HCl, 500 mM NaCl, 1 mM EDTA	7.6
Oligo(dT) low-salt buffer	20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA	7.6
Oligo(dT) washing buffer	100 mM NaOH, 5 mM EDTA	
Phenol	63% phenol (vol/vol), 37% (vol/vol) 50 mM TE (pH 8.0), 7 mM 8-hydroxy quinoline	
PNE	30 mM piperazine-N,N'-bis [2-ethane-sulfonic acid] (PIPES), 100 mM NaCl, 1 mM EDTA	6.0

Table 2, continued.

Buffer	Composition	pH
Reverse transcription mix	50 mM Tris-HCl, 10 mM MgCl ₂ , 10 mM DTT, 4 mM Na pyrophosphate, 1.25 mM dGTP, 1.25 mM dCTP, 1.25 mM dATP, 1.25 mM dTTP, 0.5 units/ul RNasin, 3 units/ul reverse transcriptase (Life Sciences)	8.3
RIP buffer	50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, 0.2% (vol/vol) NP40, 0.05% (vol/vol) SDS, 1% (vol/vol) Aprotinin, 0.02% (wt/vol) Na azide	7.4
RT 10X buffer	340 mM Tris-HCl, 500 mM NaCl, 60 mM MgCl ₂ , 50 mM DTT	8.3
RT d/ddATP mix	3.6 μM ddATP, 250 μM dCTP, 250 μM dTTP, 250 μM dGTP, 50 mM NaCl, 34 mM Tris-HCl, 6 mM MgCl ₂ , 5 mM DTT	8.3
RT d/ddCTP mix	100 μM ddCTP, 250 μM dCTP, 250 μM dTTP, 250 μM dGTP, 50 mM NaCl, 34 mM Tris-HCl, 6 mM MgCl ₂ , 5 mM DTT	8.3
RT d/ddGTP mix	50 μM ddGTP, 250 μM dCTP, 250 μM dTTP, 250 μM dGTP, 50 mM NaCl, 34 mM Tris-HCl, 6 mM MgCl ₂ , 5 mM DTT	8.3
RT d/ddTTP mix	200 μM ddTTP, 250 μM dCTP, 250 μM dTTP, 250 μM dGTP, 50 mM NaCl, 34 mM Tris-HCl, 6 mM MgCl ₂ , 5 mM DTT	8.3
SDS-PAGE sample buffer	120 mM Tris-PO ₄ , 1% (vol/vol) SDS, 40% (vol/vol) glycerol, 0.02% (wt/vol) phenol red	6.7

Table 2, continued.

Buffer	Composition	pH
Second-strand synthesis mix	20 mM Tris-HCl, 5 mM MgCl ₂ , 10 mM (NH ₄) ₂ SO ₄ , 100 mM KCl, 50 ug/ml BSA, 40 uM dGTP, 40 uM dCTP, 40 uM dATP, 40 uM dTTP, 8.5 units/ml RNase H, 230 units/ml Klenow fragment of <i>E. coli</i> pol I	7.5
SSC	150 mM NaCl, 15 mM Na citrate	7.0
SSPE	50 mM NaPO ₄ , 900 mM NaCl, 5 mM EDTA	7.7
Stop solution	90% (vol/vol) formamide, 20 mM EDTA, 0.3% (wt/vol) bromophenol blue, 0.3% (wt/vol) xylene cyanol	
TAE	40 mM Tris-acetate, 2 mM EDTA	8.0
TBE	89 mM Tris-borate, 89 mM boric acid	8.3
10 mM TE	10 mM Tris base, 1 mM EDTA	8.0
50 mM TE	50 mM Tris base, 50 mM EDTA	8.0
Terminal transferase mix	200 mM K cacodylate, 2 mM MnCl ₂ , 1 uM DTT, 1 mM dCTP, 1 unit/u ^l terminal transferase	6.9

Virus Strains and Cell Lines

The virulent Miller strain of transmissible gastroenteritis virus (TGEV) was obtained from American Type Culture Collection (ATCC VR743-1W) in the form of porcine intestinal washings and cloned twice by plaque purification. Amplification of TGEV was carried out in a

continuous swine testicle (ST) cell line established by McClurkin and Norman (52) and obtained from Dr. David Brian. TGEV used in the experiments had been passed 8 to 10 times in ST cells after plaque purification.

Cloning and Sequencing Vectors

PstI-digested, oligo(dG)-tailed pBR322 used as the cDNA cloning vector in the following experiments was purchased from Bethesda Research Laboratories. Riboprobe Gemini sequencing vector plasmids were obtained from Promega.

Virus Stocks

ST cells were grown to 90% confluency in 10 cm plastic dishes (Nunc) and were infected with TGEV at a multiplicity of infection of 3 to 5 in DME-2. Following adsorption of the virus for one h, the inoculum was aspirated and replaced with 5 ml DME 2. The infected cells were incubated at 37°C until approximately 75% of the cells had lysed. The plates were scraped and the medium was collected and freeze-thawed once at -70°C prior to sonication for 120 s in a Heat Systems Sonicator (model W-225R) using a cup probe at 75% power. The lysates were clarified by centrifugation at 1200 x g for 5 min and stored at -70°C.

Plaque Assay

TGEV stocks were titered by plaque assay on ST cell monolayers in plastic six-well dishes (Nunc). Monolayers were infected with 0.4 ml of serial 10-fold dilutions of virus in DME-2. After an adsorption period of one h at 37°C, the inoculum was removed and replaced with 3 ml of DME-2 containing 0.75% (wt/vol) agarose (type II, Sigma). Plaque assays were incubated at 37°C until plaques became visible. Cells were fixed by adding one ml 2% (vol/vol) glutaraldehyde to each well and incubating at room temperature for at least three hours. The agarose was removed, the plates were allowed to dry, and plaques were counted. Titers were recorded as plaque-forming units per ml (PFU/ml).

Organic Extraction and Recovery of Nucleic Acids

Nucleic acids were purified by extracting twice with one volume phenol and one volume chloroform and once with one volume chloroform. Phenol as referred to in this text is equivalent to 63% (vol/vol) phenol, 37% (vol/vol) 50 mM TE (pH 8.0), 7 mM 8-hydroxy quinoline. Chloroform as referred to in this text is equivalent to 96% (vol/vol) chloroform, 4% (vol/vol) isopentanol. Aqueous and organic phases were separated by centrifugation at 4,000 x g for 5 min at room temperature.

DNA was precipitated from the aqueous phase by addition of one-fifth volume saturated ammonium acetate and two volumes 95% (vol/vol) ethanol. RNAs were precipitated by addition of one-fifth volume 5 M NaCl and 2.5 volumes 95% ethanol. RNA precipitations were incubated for at least one h at -20°C . Nucleic acids were pelleted by centrifugation at $16,300 \times g$ for 30 min at 4°C . Pelleted material was washed with 70% (vol/vol) ethanol and dried under vacuum.

Virion RNA Production

ST cell monolayers in 10 cm plastic dishes were infected at an MOI of 3 to 5 with TGEV in DME-2. After an adsorption period of 1 h at 37°C , an additional 3 ml DME-2 was added to each dish. The plates were incubated at 37°C . To 3 of 10 dishes, 100 mCi $[5,6-^3\text{H}]$ -uridine (NET-367) was added at 5 h post-infection. Alternatively, $[^{32}\text{P}]$ -labeled virion RNA was produced by adding 375 μCi $[^{32}\text{P}]$ -phosphoric acid (NEX-053) to each dish. When 75% of the cells had lysed, the plates were scraped with a rubber policeman and the cell lysate was collected. Following one freeze-thaw cycle at -70°C , the medium was sonicated for 120 s and clarified as described above. Virions were precipitated by the addition of 5.9 g NaCl and 50 ml 50% (wt/vol) polyethylene glycol (PEG) (mol. wt. 3350, Sigma cat. no. P3640) per 200 ml clarified lysate to give a final

concentration of 10% PEG and 2.4% NaCl (90). Following a 2 h incubation of the mixture at 4°C, the precipitate was collected by centrifugation at 10,400 x g for 45 min at 4°C. Pellets were dissolved in PNE buffer and transferred to 8.8 cm x 1.5 cm polyallomer centrifuge tubes (Seton Scientific). The suspensions were underlaid with 3.5 ml 10% (wt/vol) potassium tartrate (KT) in PNE buffer and 2 ml 33% (wt/vol) KT in PNE buffer. Centrifugation of the discontinuous gradients was carried out in a Beckman SW41Ti rotor at 160,000 x g for 150 min at 4°C. Virus particles were collected from the interface between the 10% KT and 33% KT in PNE buffer pads. The virus suspension was diluted ten-fold with PNE buffer and virions were pelleted by ultracentrifugation at 160,000 x g for 75 min at 4°C. Pellets were drained and dissolved in 2 ml NET buffer supplemented with 100 µl 200 mM vanadyl ribonucleoside complexes (Bethesda Research Laboratories), 10 µl 20 mg/ml proteinase K, 200 µl 10% (wt/vol) sodium dodecyl sulfate (SDS), 160 µl 5 M NaCl, and 100 µl 200 mM EDTA. Following a 30 min incubation at 37°C the RNA was phenol-chloroform extracted. RNA pellets were resuspended in 200 µl NET buffer. The material was loaded onto 10-30% continuous sucrose gradients in NET buffer supplemented with 0.2% (vol/vol) SDS. Following ultracentrifugation of the gradients at 160,000 x g for 3 h at 20°C, 400 µl fractions were collected with a peristaltic pump from the bottom of

the gradients. Twenty μ l aliquots from each fraction were transferred to Whatman GFC glass fiber filters. The filters were dried at room temperature and placed in scintillation vials. Three ml scintillation fluid [5 g 2,5-diphenyloxazole (PPO) per liter xylene] was added to each vial and the samples were counted in a Packard LSC 460CD liquid scintillation counter. Gradient profiles were plotted using the count data and from these plots the fractions containing TGEV genome-length RNA were identified. RNA was recovered from these fractions by ethanol precipitation. RNA pellets were resuspended in NET buffer and the purity and concentration of the RNA was determined by measuring the absorbance of the suspensions at wavelengths of 260 and 280 nm in a Gilford 2600 spectrophotometer. RNA was reprecipitated and stored at -70°C .

Isolation of Polyadenylated Intracellular RNAs

Monolayers of ST cells in 10 cm plastic dishes were infected at an MOI of 5 with TGEV in DME-2. After adsorption for one h at 37°C the inoculum was aspirated and replaced with 5 ml DME-2 per dish. At six h post-infection the DME-2 was aspirated and replaced with 3 ml fresh DME-2 containing 2.5 $\mu\text{g/ml}$ actinomycin D. Three of 10 dishes received 100 μCi [5,6- ^3H]-uridine (NET-367). The incubation was continued at 37°C to 11 h post-infection.

The medium was then aspirated and one ml ice cold NTE buffer containing 0.5% (vol/vol) Nonidet P40 (NP40) (Particle Data Laboratories Ltd.) was added to each dish. Dishes were held on ice for 5 min and the cell lysates were collected. Dishes were rinsed with an additional one ml volume of NTE buffer and the rinses were pooled with the lysates. Following clarification, the lysates were treated with proteinase K, precipitated, and recovered as described above.

Polyadenylated RNA was isolated by oligodeoxythymidylic acid (oligo (dT)) cellulose chromatography. Columns consisting of approximately 2 ml oligo(dT) cellulose (type 2, Collaborative Research, Inc.) in a pasteur pipet were washed with 3 volumes of distilled water, 3 volumes oligo(dT) washing buffer, and additional distilled water as needed to bring the pH of the effluent below 8.0. RNA was resuspended in 400 μ l distilled water and denatured by heating at 65°C for 5 min. An equal volume of 2X high-salt oligo(dT) loading buffer was added and the suspension was passed through the column. The eluate was collected, heated at 65°C for 5 min, and reapplied to the column. The column was washed with 8 ml high-salt loading buffer and 4 ml low-salt loading buffer. Polyadenylated RNA was eluted from the column by addition of 3 ml elution buffer. Eluted RNA was recovered as described above.

Urea-Agarose Gel Electrophoresis

TGEV subgenomic RNAs were purified by urea-agarose gel electrophoresis as described by Rosen et al. (63) with minor modifications. Three-fold concentrated agarose (2.1 g agarose in 70 ml distilled water) was autoclaved and added to 140 ml citrate-urea buffer and horizontal gels (8.8 cm x 25.4 cm) were poured in a Studier apparatus (85) (10 cm x 39.5 cm) in a refrigerated room. RNA was resuspended in citrate-urea sample buffer, held at room temperature for 5 min, and loaded onto the urea-agarose gels. Ribosomal RNAs from ST cells labeled with [5,6-³H]-uridine (NET-367) were used as molecular weight markers. Electrophoresis was carried out at 50 volts for 16 h at room temperature. Lanes were then sliced at 0.4 cm intervals, and slices were placed in scintillation vials. Three ml aqueous scintillation fluid [33% (vol/vol) Triton X-100 (Research Products International Corp.), 66.5% (vol/vol) xylene, 0.5% (wt/vol) PPO] was added to each vial and the samples were counted using the preset tritium channel. Count data were used to plot gel profiles of lanes containing TGEV-ST RNA or ST ribosomal RNA. Migration distances of the rRNAs were used to estimate the location of TGEV subgenomic mRNAs. Gel slices from parallel lanes predicted to contain TGEV mRNAs of interest were suspended in NET buffer and melted by heating at 65°C

for 5 min. Agarose slurries were extracted three times with chloroform and RNA was recovered from the aqueous fractions as described above.

In vitro Translation of Gel-Purified mRNAs

The identities of RNAs extracted from urea-agarose were confirmed by in vitro translation. RNA (1 to 2 μ g) in 8 μ l distilled water was added to a microcentrifuge tube containing 35 μ l rabbit reticulocyte lysate (Promega), 1 μ l methionine-free amino acid mixture (Promega), 5 μ l (50 uCi) L-[³⁵S]-methionine (NEG-009A), and 1 μ l (30 U) RNasin (Promega). Following incubation at 30°C for 2 h the reactions were terminated by freezing at -70°C.

Immunoprecipitation of Virus-Specific Proteins

Products of in vitro translation were immuno-precipitated with TGEV-specific polyclonal ascitic fluid using a modification of previously described procedures (8,67). Fifty μ l of translation mixture was diluted eleven-fold in RIP buffer. Fifty μ l of hyperimmune ascitic fluid was added and the mixtures were incubated for 1 h at 0°C. Immune complexes were precipitated with 50 μ l 10% (vol/vol) heat- and formalin-fixed Staphylococcus aureus (Cowan) cells (39) by incubation at 0°C for 1 h and pelleted by centrifugation at 6,500 x g for 20 s. Pellets were washed 5 times with ice cold RIP buffer. Bound

proteins were eluted with 20 μ l 1% (wt/vol) SDS, 0.02 M dithiothreitol (DTT) for 15 min at room temperature and 5 min at 60°C. S. aureus cells were removed by centrifugation at 6,500 x g for 20 s. The supernatant fluids were mixed with 20 μ l SDS-polyacrylamide gel electrophoresis (PAGE) sample buffer and stored at -20°C.

SDS-Polyacrylamide Gel Electrophoresis

Immunoprecipitated proteins were analyzed by the Laemmli method of PAGE (42). Proteins were denatured prior to electrophoresis by heating for 5 min at 65°C. Electrophoresis was in 10% (wt/vol) polyacrylamide slab gels. Gels were fixed in 10% (vol/vol) trichloroacetic acid and proteins were stained with Coomassie brilliant blue G (18). To aid in detection of labeled proteins, gels were enhanced for 30 min in Fluoro-Hance (Research Products International Corp.) according to the manufacturer's directions. Gels were dried onto Whatman 3MM paper and exposed to preflashed Kodak XAR-2 x-ray film.

The molecular weights of translation products were determined from their distance of migration in the gels relative to those of standard proteins of known molecular weight. GEL, a computer program obtained from Dr. Brian Fristensky and based on the work of Schaffer and Sederoff (68), was used in this analysis. The standard proteins were bovine serum albumin (BSA) (66 kd), ovalbumin (45 kd),

β -glyceraldehyde-3-phosphate dehydrogenase (36 kd), carbonic anhydrase (29 kd), and trypsin inhibitor (20.1 kd).

Complementary DNA Synthesis

Synthesis of first-strand cDNA was carried out according to the protocol of Gubler and Hoffman (25). Three μ g virion or subgenomic RNA and 4 μ g oligo(dT) or one to 3 μ g priming cDNA restriction fragment were added to a reaction mixture containing 40 μ l reverse transcription mix and incubated for 60 min at 42°C. Reactions were terminated by addition of 100 μ l 50 mM TE buffer. Reaction products were phenol-chloroform extracted and ethanol precipitated.

First-strand cDNA reverse transcribed from virion RNA was rendered double-stranded in a reaction mixture containing 100 μ l of second strand synthesis mix and incubated at 12°C for 60 min and 22°C for 60 min. To stop the reaction, EDTA was added to a final concentration of 20 mM. Products were phenol-chloroform extracted and recovered by ethanol precipitation. First-strand cDNA copies of TGEV subgenomic RNAs were not subjected to second strand synthesis (47).

Oligodeoxycytidylate tails were added to double-stranded cDNA or cDNA:RNA hybrids in 50 μ l of terminal transferase mix as described by Michelson and Orkin (54).

Reaction mixtures were incubated at 30°C for 10 min. Homopolymeric tail addition was terminated by addition of 2.5 µl 200 mM EDTA and 100 µl 10 mM TE. Tailed molecules were recovered by ethanol precipitation.

Oligo(dC)-tailed double-stranded cDNA or cDNA:RNA hybrids were annealed to PstI-digested, oligo(dG)-tailed pBR322. A vector:cDNA molar ratio of approximately 50:1 was used. Annealing was done in 50 µl annealing buffer incubated at 65°C for 15 min and 58°C for 90 min. E. coli (strain JM109 or DH5) cells rendered competent by the method of Hanahan (27) were transformed with the recombinant molecules and plated on Luria agar containing 10 µg/ml tetracycline. Colonies that developed during a 24 h incubation at 37°C were patched on Luria agar supplemented with tetracycline (10 µg/ml) or ampicillin (100 µg/ml).

Isolates resistant to tetracycline but susceptible to ampicillin were subjected to small-scale plasmid analysis (4), as follows. Isolates were inoculated into 12 ml Luria broth containing 10 µg/ml tetracycline. Cultures were incubated at 37°C with agitation for approximately 16 h and chilled in an ice bath. Cells were pelleted at 3,000 x g for 5 min at 4°C and resuspended in 650 µl 10 mM TE buffer. The cell suspensions were transferred to 1.5 ml microcentrifuge tubes (Treff Lab). Cells were pelleted at 15,600 x g for 15 s and resuspended in 120 µl ice cold 20%

sucrose (wt/vol) in 50 mM TE buffer. The mixtures were freeze-thawed once and 20 μ l of 10 mg/ml lysozyme in 50 mM TE buffer was added. Following incubation at 0°C for 30 to 60 min, 400 μ l 1% (wt/vol) SDS, 0.2 N NaOH was added and the tubes were incubated at 50°C for 60 min and 0°C for 10 min. The mixtures were neutralized by addition of 200 μ l 3 M potassium acetate (pH 4.8); incubation at 0°C was continued for an additional 40 min. Debris was cleared from the preparations by centrifugation at 15,600 x g for 15 s. An equal volume of isopropanol was added to the supernatant fluids, the tubes were held at room temperature for 30 min, and plasmid DNA was pelleted by centrifugation at 15,600 x g for 10 min. Pellets were washed with 70% ethanol, dried under vacuum, and resuspended in 200 μ l distilled water. Ten μ l of this volume were run in horizontal gels consisting of 0.8% agarose (wt/vol) in TAE buffer. The lengths of cDNA inserts were predicted upon comparison of migration of the recombinant plasmids to that of unrestricted vector.

Restriction maps of these inserts were constructed as follows. Recombinant plasmids as large as vector dimers were phenol-chloroform extracted and recovered by ethanol precipitation. The plasmids were digested with the restriction endonucleases HindIII, KpnI, PstI, PvuII, XbaI, and XhoI (Boehringer Mannheim) and subjected to electrophoresis in 0.8% agarose in TAE. Bacteriophage

lambda DNA digested with restriction endonuclease HindIII was used as standards in these gels. The lengths of the lambda DNA restriction fragments were 23,130 bp, 9416 bp, 6682 bp, 4373 bp, 2322 bp, 2027 bp, and 564 bp. The lengths of cDNA restriction fragments were determined using the computer program GEL (68). Restriction maps were constructed by inspection.

Hybridization Analysis of cDNA Clones

Cloned cDNAs were screened for TGEV-specific sequences by colony (24) or slot blot hybridization. In colony hybridization studies, discs of Zeta-Probe nylon membrane (Bio-Rad) were placed on Luria agar supplemented with 10 µg/ml tetracycline. Bacteria from tetracycline-resistant, ampicillin-susceptible colonies were patched onto the discs with sterile toothpicks. The plates were incubated at 37°C for 8 h and bacterial cells on the filters were lysed by placing the membranes on Whatman 3MM paper saturated with 0.5 N NaOH for 5 min. Filters were neutralized on 3MM paper saturated with 1 M Tris-HCl (pH 8.0) for 5 min and were then incubated on 3MM paper saturated with 1 M Tris-HCl (pH 8.0), 1.5 M NaCl for 5 min, washed in 2X SSC and baked under vacuum at 80°C for 90 min.

A Hybri-Slot filtration manifold (Bethesda Research Laboratories) was used to carry out slot blot hybridizations. Plasmid DNA obtained from small-scale

plasmid isolations was denatured in 0.4 N NaOH at 70°C for one h and passed through the manifold onto nylon membranes that had been pre-wet with distilled water and 0.4 N NaOH. Membranes carrying plasmid DNA were dried at room temperature.

DNA copies of TGEV (Miller strain) RNA 6 and RNA 7 were characterized by the Southern blot procedure (77). Restricted plasmids were fractionated in 0.8% (wt/vol) agarose in TAE gels. Gels were run at 30 volts for 4 h. Following staining with ethidium bromide the gels were photographed using Polaroid Type 55 Land film. DNA was transferred to Zeta-Probe nylon membranes (Bio-Rad) by the method of Reed and Mann (61). Following DNA transfer, the membranes were baked at 80°C for one h under vacuum.

Colony and slot blots were probed with [³²P]-labeled TGEV virion RNA prepared as described above. Nylon membranes were prehybridized in 5 ml per filter of hybridization buffer at 42°C for one h. The radiolabeled RNA was denatured in 0.1 N NaOH at 70°C for 5 min and added to the prehybridization solution. Hybridization was carried out at 42°C for 15 h. Filters were washed in 2X SSPE, 0.1% (vol/vol) SDS at 68°C for one h and 1X SSPE, 0.1% SDS at 68°C for one h prior to exposure to x-ray film.

Southern blots were probed with labeled restriction fragments of virion RNA-specific cDNA. The fragments were separated and recovered as previously described and were

radiolabeled by nick translation (51) in a reaction mixture containing 50 μ l nick translation mix and incubated at 16°C for one h. The reaction was terminated by addition of 2 μ l 0.5 M EDTA and the nick-translated DNA was separated from unincorporated dNTPs by chromatography through a column containing 2 ml Sephadex G-50 in 10 mM TE. Radiolabeled DNA was recovered by ethanol precipitation. The nylon membranes were prehybridized and radiolabeled DNA fragments denatured in 0.1 N NaOH were added to the buffer. Hybridization conditions and washes were as described above.

Subcloning of cDNA Restriction Fragments

Selected cDNA restriction fragments were subcloned into pGEM sequencing vectors. The sequencing strategy was planned such that restriction fragments were overlapping. The sequence of regions not overlapped by other fragments was determined using at least two clones. Recombinant plasmids were restricted and loaded onto 0.8% low-gelling temperature agarose in TAE buffer gels. Gels were run at 40 volts for 2 h, stained with ethidium bromide, and viewed using an ultraviolet transilluminator. Gel slices containing cDNA fragments were excised, suspended in 400 μ l NET, and melted at 65°C for 5 min. Agarose was removed by extracting the suspensions three times with phenol/chloroform and once with chloroform. Restriction fragments were

recovered by ethanol precipitation.

Sequencing vector plasmids were restricted, gel purified as described above, and dephosphorylated with calf intestinal phosphatase (CIP), as follows. Plasmid DNA was resuspended in 50 μ l CIP mix. The reactions were incubated at 37°C for 15 min and 56°C for 15 min. An additional 50 units CIP were added and the incubations were repeated. The CIP was then inactivated by addition of 1 μ l 10% (vol/vol) diethylpyrocarbonate and incubation at 68°C for 10 min. The dephosphorylated vectors were phenol-chloroform extracted and precipitated with ethanol.

Complementary DNA restriction fragments were ligated to dephosphorylated sequencing vectors in 25 μ l ligation mix. Ligation mixtures were incubated overnight at 18°C. Competent E. coli (strain JM109) cells were transformed with the products and plated on Luria agar containing 100 μ g/ml ampicillin. Plasmids of ampicillin-resistant transformants were characterized by small-scale plasmid analysis.

Sequencing of TGEV-Specific cDNA

The GemSeq K/RT Sequencing System (Promega), which employs the dideoxy method of chain termination (66), was used to determine the nucleotide sequence of subcloned restriction fragments. One to 2 μ g plasmid DNA was resuspended in 20 μ l of distilled water in a 1.5 ml

microcentrifuge tube. Two μ l 2 M NaOH, 2 mM EDTA was added and the tubes were incubated for 5 min at room temperature. Reactions were neutralized by addition of 3 μ l 3 M sodium acetate (pH 5.0) and 7 μ l distilled water. Denatured DNA was precipitated by addition of 75 μ l absolute ethanol and incubated at -20°C for 15 min. DNA was pelleted by centrifugation at $15,600 \times g$ for 10 min. The pellets were washed with 70% ethanol and dried under vacuum. Dried pellets were resuspended in 6 μ l distilled water, 1 μ l 10X reverse transcriptase (RT) or Klenow buffer, and 3 μ l 10 ng/ml SP6 or T7 promoter/primer. Annealing mixtures were incubated at 37°C for 90 min. Five units of Klenow fragment or avian myeloblastosis virus RT and 5 μ l [^{35}S]-deoxyadenosine 5'-[alpha-thio]triphosphate (500 Ci/mmol) (NEG-034S) were added and 3 μ l of this radiolabel/primer/template mixture was added to 1.5 ml microcentrifuge tubes containing 3 μ l deoxy/dideoxy (d/dd) CTP, d/ddATP, d/ddGTP, or d/ddTTP. The reactions were incubated at 37°C (Klenow) or 42°C (AMV RT) for 20 min. One μ l chase solution was added and the incubation was continued for an additional 20 min. Alternatively, primers were extended for 10 to 30 min in the absence of dideoxy nucleotides, followed by addition of 3 μ l d/ddNTPs and incubation at 37°C (Klenow) or 42°C (AMV RT) for 20 min. The concentration of dNTPs in the extension mixture was 50 μM . All reactions were terminated by addition of 5 μ l stop solution and stored at -70°C .

Reaction mixtures were electrophoresed on 32 cm X 38 cm 8 M urea, 6 or 8% polyacrylamide sequencing gels (15 or 20 ml acrylamide:N,N'-methylene-bis-acrylamide::39:1, respectively, 10 ml 10X TBE buffer, 30 ml distilled water, 50 g urea, 1 ml 10% (w/v) ammonium persulfate, and 20 μ l N,N,N',N'-tetramethylethylenediamine) in TBE buffer at 1100 volts. To maximize the amount of sequence information obtained from a single sequencing reaction, some samples were loaded into adjacent wells when the bromphenol blue marker dye of the first (double-loading) or second (triple-loading) load had run off the gel. Gels were fixed in 10% (vol/vol) methanol, 10% (vol/vol) acetic acid for 15 min and dried onto Whatman 3MM paper. Kodak XAR-2 film was exposed to the gels for 120 to 168 h prior to development according to the manufacturer's instructions. Sequencing data were analyzed on an IBM personal computer using the programs of Mount et al. (55) and Fristensky et al. (20). The hydrophilicity plots were constructed with a computer program written by Dr. James Etchison that incorporated the algorithm of Hopp and Woods (33).

RESULTS

Production of TGEV Virion RNA

Full-length TGEV virion RNA was purified for use as a template in cDNA synthesis primed by a cDNA restriction fragment representative of TGEV (Miller strain) genomic RNA. ST cells were infected with TGEV, and the infections were allowed to proceed until 75% of the cells had lysed. Virions recovered from the cell lysate by PEG precipitation were purified by centrifugation to equilibrium in a discontinuous KT gradient. Coronavirions have a density in KT of 1.18 g/ml. A narrow band of virus particles formed at the interface between the 10% KT (density = 1.07) and 33% KT (density = 1.22) pads. Virus particles were recovered from the interface and the genomic RNA was partially purified by phenol-chloroform extraction. Virion RNA was purified from the extracted material by sedimentation velocity centrifugation in an isokinetic sucrose gradient. The results obtained are shown in Figure 1. Virion RNA sedimented in the lower portion of the gradient. The material sedimenting near the top of the gradient consisted of nonspecific RNAs. The yield of TGEV virion RNA, as determined by spectrophotometry, was

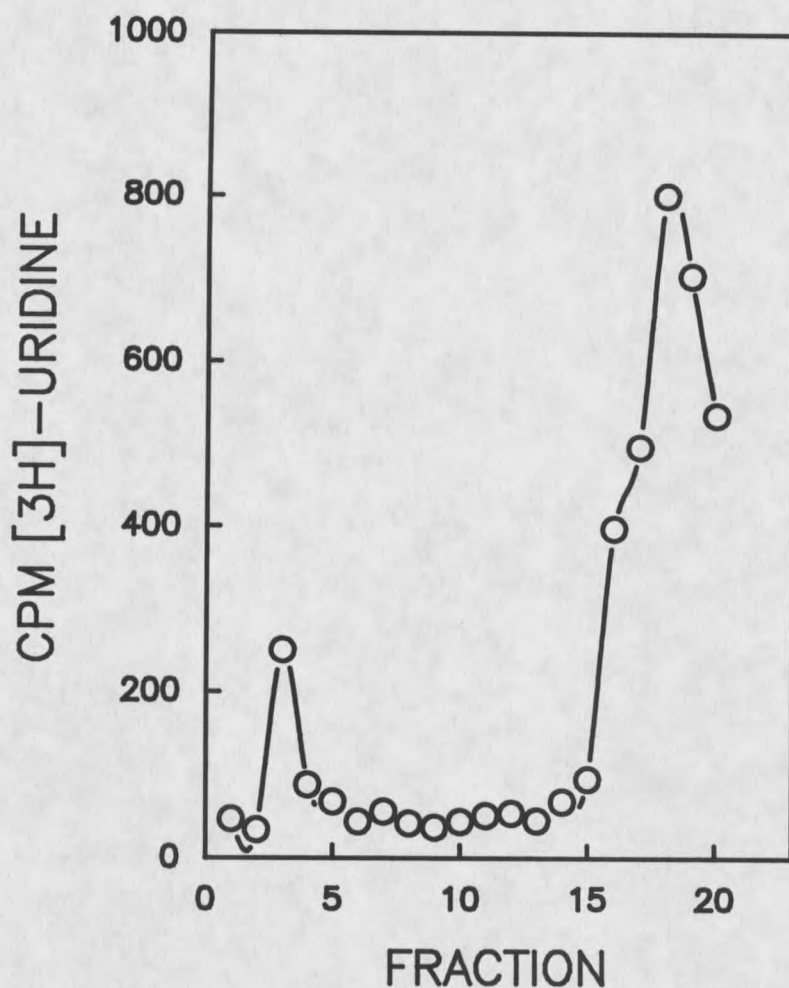


Figure 1. Profile of isokinetic 10-30% sucrose gradient used in purification of TGEV (Miller strain) virion RNA. Fractions were collected from the bottom of the gradient. The solid line represents the profile of $[^3\text{H}]$ -uridine-labeled RNAs from TGEV (Miller strain)-infected ST cells. Data are expressed as counts per minute (CPM).

approximately 0.2 µg per 10 cm dish of infected ST cells. The yield of full-length virion RNA was reduced when infections were incubated until all ST cells had lysed.

Synthesis of cDNAs

A restriction fragment from an oligo(dT)-primed clone (clone 141, provided by Andreas Luder) of polyadenylated TGEV (Miller strain) virion RNA was used to prime synthesis of first-strand virion RNA-specific cDNA. The 889 bp PvuII-HindIII fragment of clone 141, which corresponds to bases 5698 to 6583 of the sequence given in Figure 3, was gel purified and heat-denatured prior to addition to purified TGEV virion RNA. First- and second-strand complementary DNA synthesis was carried out as described in Materials and Methods. Double-stranded cDNA was tailed and inserted into pBR322. The authenticity of the cDNA inserts was confirmed by their hybridization to [³²P]-labeled virion RNA in slot-blot experiments (data not shown). Virus-specific complementary DNA inserts up to 5100 bp in length were obtained. Cloned inserts were mapped with restriction endonucleases and the sizes of the resulting fragments were estimated by comparison of their migration in agarose gels to that of standard DNA fragments. A restriction map of the region of TGEV virion RNA represented by these clones was constructed (Figure 2) and compared to the maps produced from oligo(dT)-primed clones

of virion RNA. The maps of both groups of clones were consistent.

Sequencing of cDNAs Representative of
TGEV (Miller strain) Virion RNA

Restriction fragments of clones obtained by directed first-strand synthesis (clones 1561 and 1563) and clones produced by oligo(dT)-primed first-strand synthesis (clones 141 and 150, provided by Andreas Luder) were subcloned in Riboprobe Gemini vectors and sequenced by the method of Sanger (66). The arrangement of subcloned fragments was one in which all cleaved restriction sites were present within overlapping fragments of cDNA. The subcloning strategy is illustrated in Figure 2. Subcloned restriction fragments were sequenced from both ends using primers complementary to the SP6 and T7 promoters flanking the Riboprobe Gemini multiple cloning regions. Primer extension in the absence of dideoxy nucleotides and multiple loading of sequencing gels made possible complete sequencing of restriction fragments up to 1200 bases in length.

The sequence of the 3' 7325 bases of TGEV (Miller strain) virion cDNA was determined (Figure 3). Included within this sequence were 5 complete ORFs and a portion of the peplomer-encoding ORF. The amino acid sequences of proteins encoded by the major open reading frames (ORFs) are listed beneath the corresponding nucleotide sequence.

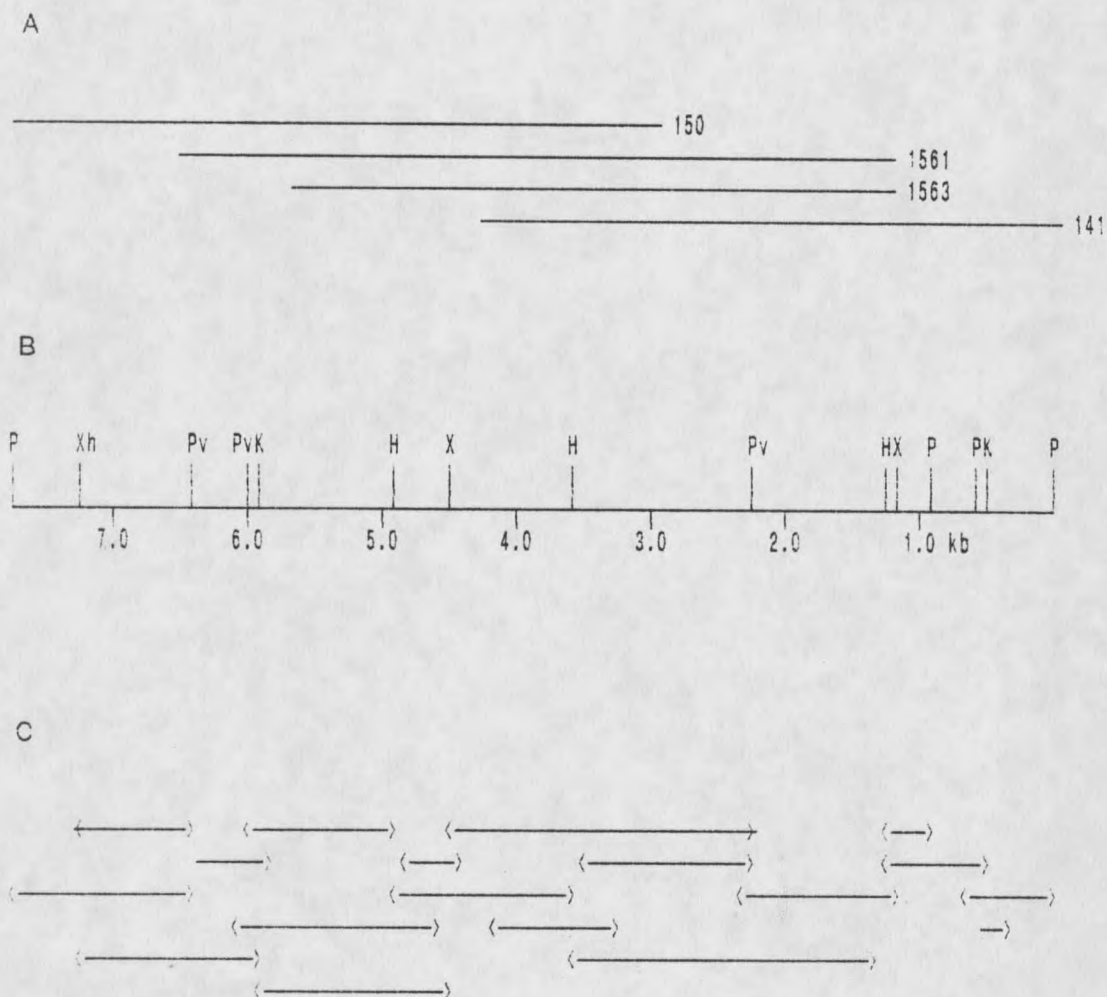


Figure 2. Restriction endonuclease map of TGEV (Miller strain) virion cDNA clones 150, 1561, 1563, and 141 and strategy used in nucleotide sequence determination. A. Virion cDNA clones 150, 1561, 1563, and 141 were mapped. B. Restriction map of the virion cDNA clones. C. Restriction fragments of the clones that were subcloned in Riboprobe Gemini sequencing vectors. Arrows indicate direction in which fragments were sequenced. Restriction enzymes: XhoI (Xh), PvuII (Pv), KpnI (K), HindIII (H), XbaI (X), PstI (P).

Table 3. Abbreviations for amino acids.

Amino acid	Three-letter abbreviation	One-letter symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

The amino acids and the 3-letter and one-letter codes are listed in Table 3.

The partial ORF potentially encoded the C-terminal 1061 amino acids of the peplomer protein. The nucleotide sequence of this region was 96.9% homologous with the corresponding region of the peplomer gene of TGEV (Purdue strain) (60). Most differences in nucleotide sequence were located in the 5' portion of the incomplete ORF.

Differences in primary structure between the C-terminal portion of the TGEV (Miller strain) peplomer protein and

Figure 3. Partial nucleotide sequence of TGEV (Miller strain) virion cDNA.

20	40	60
GAGTGA	CTCGAGCTTTTTCA	GATACCGTGAAATACCGTTT
S D S S F F	R Y R E I P F	F V T E R Q R
80	100	120
TTACTGTTACGTACAATATA	ATGGCAGAGCTCTTAAGTAT	TTAGGAAAATTAGCACCTAG
Y C Y V Q Y	N G R A L K Y	L G K L A P S
140	160	180
TGTCAAGGAGATTGCTATTA	GTAATGGGGCCATTTTTAT	ATTAATGGTGACAATAATTT
V K E I A I	S K W G H F Y	I N G D N N F
200	220	240
TAGCACATTTCTATTGAAT	GTATATCTTTTAATTTGACC	ACTGGGGATAGTGACGTTTT
S T F P I E	C I S F N L T	T G D S D V F
260	280	300
CTGGACAATAGCTTACACAT	CGTACACTGAAGCATTAGTA	CAAGTTGAAAACACCGCTAT
W T I A Y T	S Y T E A L V	Q V E N T A I
320	340	360
TTCAAAGGTGACCTATTGTA	AAAGTCACGTTAATAACATT	AAATGCTCTCAAATAACTGC
S K V T Y C	K S H V N N I	K C S Q I T A
380	400	420
TAATTTGAATAAAGGATTTT	ATCCTGTTTCATCAAGTGAA	GTTGGTCTAGTCAATAAAAG
N L N K G F	Y P V S S S E	V G L V N K S
440	460	480
TGTTGTTTTACTACCTAGCT	TTTACACACATACAATTGTT	AACATAACTATTGGGCTTGG
V V L L P S	F Y T H T I V	N I T I G L G
500	520	540
TAAGAAGCGTAGTGGTTATG	GTCAACCCATAGCCTCAACA	TTAAGTAACATCACACTACC
K K R S G Y	G Q P I A S T	L S N I T L P
560	580	600
AATGCAGGATCATAACACCG	ATGTATACTGTATACGTTCT	GACCAATTTTCAGTTTATGT
M Q D H N T	D V Y C I R S	D Q F S V Y V
620	640	660
TCAATCTACTTGCAAAAGTG	CTTTATGGGACAATATTTTT	AAACGAAACTGCACGGACGT
Q S T C K S	A L W D N I F	K R N C T D V
680	700	720
TTTAGATGCCACAGCTGTTA	TAAAACTGGTACTTGTCT	TTCTCATTTGATAAATTGAA
L D A T A V	I K T G T C P	F S F D K L N
740	760	780
CAATTACTTAACTTTTAACA	AATTATGTTTGTGTTGAGT	CCTGTTGGGGCAAATTGTAA
N Y L T F N	K L C L S L S	P V G A N C K
800	820	840
GTTTGATGTAGCTGGCCGTA	CAAGAACCAATGAACAGGTT	GTTAGAAGTTTGTATGTAAT
F D V A G R	T R T N E Q V	V R S L Y V I
860	880	900
ATATGAAGAAGGGGACAACA	TAGTGGGTGTACCGTCTGAT	AATAGTGGTGTGCACGATTT
Y E E G D N	I V G V P S D	N S G V H D L

Figure 3, continued.

920	940	960
GTCAGTGTCTCCCTAGATT	CCTGCACAGATTACAATATA	TATGGTAGAACTGGTGTGG
S V L P L D	S C T D Y N I	Y G R T G V G
980	1000	1020
TATTATTAGGCAAATAACA	GAACGCTAATTAGTGGCTTA	TATTACACATCACTATCAGG
I I R Q T N	R T L I S G L	Y Y T S L S G
1040	1060	1080
TGATTTGTTAGGTTTTAAAA	ATGTTAGTGATGGTGTCAAT	TACTCTGAAACGCCATGTGA
D L L G F K	N V S D G V I	Y S E T P C D
1100	1120	1140
TGTAAGCGCACAAGCAGCTG	TTATTGATGGTACCATAGTT	GGGGCTATCACTTCCATTAA
V S A Q A A	V I D G T I V	G A I T S I N
1160	1180	1200
CAGTGAAGTGTAGGTCTAA	CACATTGGACAACAACACCT	AATTTTTATTACTACTCTAT
S E L L G L	T H W T T T P	N F Y Y Y S I
1220	1240	1260
ATATAATTACACAAATGATA	GGACTCGTGGCACTGCAATT	GACAGTAATGATGTTGATTG
Y N Y T N D	R T R G T A I	D S N D V D C
1280	1300	1320
TGAACCTGTCATAACCTATT	CTAACATAGGTGTTTGTA	AAATGGTGCTTTTGT
E P V I T Y	S N I G V C K	N G A F V F I
1340	1360	1380
TAACGTCACACATTTTGATG	GAGACGTGCAACCAATTAGC	ACTGGTAATGTCACGATACC
N V T H F D	G D V Q P I S	T G N V T I P
1400	1420	1440
TACAACTTTACCATATCCG	TGCAAGTCGAATATATTCAG	GTTTACACCACCCAGTGTC
T N F T I S	V Q V E Y I Q	V Y T T P V S
1460	1480	1500
AATAGACTGTTCAAGATATG	TTTGTAAATGGTAACCCTAGA	TGTAACAAATTGTTAACACA
I D C S R Y	V C N G N P R	C N K L L T Q
1520	1540	1560
ATACGTTTCTGCATGTCAAA	CTATTGAACAAGCACTTGCA	ATGGGTGCCAGACTTGAAAA
Y V S A C Q	T I E Q A L A	M G A R L E N
1580	1600	1620
CATGGAGGTTGATTCCATGT	TGTTTGTCTGAAAATGCC	CTTAAATTTGCATCTGTAGA
M E V D S M	L F V S E N A	L K F A S V E
1640	1660	1680
AGCATTCAATAGTTCAGAAA	CTTTAGAACCTATTTACAAA	GAATGGCCTAATATAGGTGG
A F N S S E	T L E P I Y K	E W P N I G G
1700	1720	1740
TTCTTGGCTAGAAGGTCTAA	AATACATACTTCCCTCCCAT	AATAGCAAACGTAAGTATCG
S W L E G L	K Y I L P S H	N S K R K Y R
1760	1780	1800
TTCAGCTATAGAAGACTTGC	TTTTTGATAAGGTAGTAACA	TCTGGTTTAGGTACAGTAGA
S A I E D L	L F D K V V T	S G L G T V D

Figure 3, continued.

1820	1840	1860
TGAAGATTATAAACGTTGTA	CAGGTGGTTATGACATAGCT	GACTTAGTATGTGCTCAATA
E D Y K R C	T G G Y D I A	D L V C A Q Y
1880	1900	1920
CTATAATGGCATCATGGTGC	TACCTGGTGTGGCTAATGCT	GACAAAATAACTATGTACAC
Y N G I M V	L P G V A N A	D K I T M Y T
1940	1960	1980
AGCATCCCTTGCAGGTGGTA	TAACATTAGGTGCACTTGGT	GGAGGCGCCGTGGCTATACC
A S L A G G	I T L G A L G	G G A V A I P
2000	2020	2040
TTTTGCAGTAGCAGTTCAGG	CTAGACTTAATTATGTTGCT	CTCCAAACTGATGTATTGAA
F A V A V Q	A R L N Y V A	L Q T D V L N
2060	2080	2100
CAAAAACCAGCAGATTCTGG	CTAGTGCTTTCAATCAAGCT	ATTGGTAAAATTACACAGTC
K N Q Q I L	A S A F N Q A	I G K I T Q S
2120	2140	2160
ATTTGGTAAGGTTAATGATG	CTATACATCAAACATCACGA	GGTCTAGCTACTGTTGCTAA
F G K V N D	A I H Q T S R	G L A T V A K
2180	2200	2220
AGCATTGGCAAAAGTGCAAG	ATGTTGTCAAAATACAAGGG	CAAGCTTTAAGCCACCTAAC
A L A K V Q	D V V K I Q G	Q A L S H L T
2240	2260	2280
AGAACAATTGCAAAATAATT	TCCAAGCCATTAGTAGTTCT	ATTAGTGACATTTATAATAG
E Q L Q N N	F Q A I S S S	I S D I Y N R
2300	2320	2340
GCTTGACGAATTGAGTGCTG	ATGCACAAGTTGACAGGCTG	ATCACAGGAAGACTTACAGC
L D E L S A	D A Q V D R L	I T G R L T A
2360	2380	2400
ACTTAATGCATTTGTGTCTC	AGACTCTAACCAGACAAGCG	GAGGTTAGGGCTAGTAGACA
L N A F V S	Q T L T R Q A	E V R A S R Q
2420	2440	2460
ACTTGCCAAAGACAAGGTTA	ATGAATGCGTTAGGTCTCAG	TCTCAGAGATTTCGGATTCTG
L A K D K V	N E C V R S Q	S Q R F G F C
2480	2500	2520
TGGTAATGGTACACATTTGT	TTTCACTCGCAAATGCAGCA	CCAAATGGCATGATTTTCTT
G N G T H L	F S L A N A A	P N G M I F F
2540	2560	2580
TCACACAGTGCTATTACCAA	CCGCTTATGAAACTGTGACT	GCTTGGCCAGGTATTTGTGC
H T V L L P	T A Y E T V T	A W P G I C A
2600	2620	2640
TTCAGATGGTGATCGGACTT	TTGGACTTGTGTTAAAGAT	GTCCAGTTGACTTTGTTTCG
S D G D R T	F G L V V K D	V Q L T L F R
2660	2680	2700
TAATCTAGATGACAAGTTAT	ATTTGACCCCCAGAACAATG	TATCAGCCTAGAGTAGCAAC
N L D D K L	Y L T P R T M	Y Q P R V A T

Figure 3, continued.

2720	2740	2760
TAGTTCAGACTTTGTTTCATA	TTGAAGGGTGCATGTGCTA	TTTGTTAATGCAACTGTAAG
S S D F V H I E G C D V L	F V N A T V S	
2780	2800	2820
TGATTTGCCAAGTATTATAC	CTGATTATATTGAAATTAAT	CAGACTGTTCAAGACATATT
D L P S I I P D Y I E I N	Q T V Q D I L	
2840	2860	2880
AGAAAATTTTAGACCAAATT	GGACTGTACCAGAGTTGACA	TTTGACATTTTTAAAGCAAC
E N F R P N W T V P E L T	F D I F K A T	
2900	2920	2940
CTATTTAAACCTGACTGGTG	AAATTGATGACTTTGAATTT	AGGTCAGAAAAGCTACATAA
Y L N L T G E I D D F E F	R S E K L H N	
2960	2980	3000
CACCACTGTCGAACTTGCAA	TTCTCATTGACAACATTAAC	AATACATTAATCAATCTTGA
T T V E L A I L I D N I N	N T L I N L E	
3020	3040	3060
ATGGCTCAATAGAATTGAAA	CCTATGTAAAAATGGCCTTGG	TATGTGTGGCTACTAATAGG
W L N R I E T Y V K W P W	Y V W L L I G	
3080	3100	3120
CTTAGTAGTAATATTTTTCGA	TACCATTACTGCTATTTTTCG	TGTTGTAGTACAGGTTGCTG
L V V I F C I P L L L F C	C C S T G C C	
3140	3160	3180
TGGATGCATAGGTTGTTTAG	GAAGTTGTTGTCACTCTATA	TGTAGTAGAAGACAATTTGA
G C I G C L G S C C H S I	C S R R Q F E	
3200	3220	3240
AAATTACGAACCAATAGAAA	AAGTGCACGTCATTAAATT	TAAAAAATATTAATTCTTAT
N Y E P I E K V H V H		
3260	3280	3300
CATCTGCTAATAATAGCAGT	TGTTTCTGCTAGAGAAATTT	TGTTAAGGATGATGAATAAA
3320	3340	3360
GTCTTTAAGAACTAAACTTA	CGAGTCATTACAGGTCCTGT	ATGGACATTGTCAAATCCAA
		M D I V K S N
3380	3400	3420
TTTACACATCCGTAGATGCT	GTA TAGACGAACTTGTTTG	TGCATACTTTGCTGTAACAG
L H I R R C C T R R T C L	C I L C C N R	
3440	3460	3480
AATCTGCAGGCATCGTGGTG	TCACGCTCGTCGTTTGGTAT	GGCTTCATTGAGCTCCGGTT
I C R H R G V T L V V W Y	G F I E L R F	
3500	3520	3540
CCCAACGATCAAGGCGAGTT	ACATGATCCCCATGTTGTG	CAAAAAAGCGGTTCGTTCCT
P T I K A S Y M I P H V V	Q K S G S F L	
3560	3580	3600
TGGTCTCCGATCGTTTGTG	AGAAGAAGTTGGCCGAGTG	TTATCACTCATGGTTATGGC
R S S D R C Q K K L A A V	L S L M V M A	

Figure 3, continued.

3620	3640	3660
AGCACTGCATTATTCTCTAC	TGAAGCTTTTCGAGCTCGGG	CAACCATTCTATACACTGGG
A L H Y S L	L K L F E L G	Q P F Y T L G
3680	3700	3720
GGGAGGGTTTCCTGATTGGA	TGTATGTTTTGAAAAAGCTA	CTTCGTCAACATCAGCATCC
G G F P D W	M Y V L K K L	L R Q H Q H P
3740	3760	3780
GTGCGTGGGTCAAATTTATA	AAGAACAGCTGGGGCAAAG	CGGGTTGCAAGACCTGTCTG
C V G Q I Y	K E Q L G Q K	R V A R P V W
3800	3820	3840
GCAACAGTCGGGCGCAAACA	AGTCTTTCGGGGTACATGGA	TGCGCGGACCATCTGGCATT
Q Q S G A N	K S F G V H G	C A D H L A F
3860	3880	3900
CCTCAATTCACCCTGGGGCG	TGAAGGCCTCAGCAGCCAT	CGATCATTTCTTGTAATTA
L N S P W G	V E G L S S H	R S F L V N Y
3920	3940	3960
TCCACTAGTGCTTAGTACAC	ATTACGTTGCACGTGCATAC	TCTCAGAGTTCGAGATATAC
P L V L S T	H Y V A R A Y	S Q S S R Y T
3980	4000	4020
AGTGTACATACACCGTATTT	AATATACACACTTACTGGAG	TGCAATTTAAACATCTGGG
V Y I H R I		
4040	4060	4080
ATACGAACATTGTATCGTTA	TGCACCGAATCGAGTACGAC	TAGATTGACATTCAATCTGC
4100	4120	4140
GTGCAACTGTGAAGAGTTAC	GACTGAAAATAAATACTATA	AAGAAGGTCGTCGAGTTCTG
4160	4180	4200
ATCTAATCTAAATGTCTATC	GTAATCTTGAGGTCCTTCGG	AAACGGGCCCCAAAGTCCTCG
M S I V I L R S F G	N G P K V L	
4220	4240	4260
CCTCAAGTCACACAGAGTCT	GACCTACTTCTAAAGCGGTA	CGAAACACGGCGGTGGGATC
A S S H T E S	D L L L K R Y	E T R R W D
4280	4300	4320
GAGAGTGTGTTCACTTCT	TACCGATTCCGACAGTTTGT	CTACCACTCTGGTCTGAACC
R E C V H H F	Y R F R Q F V	Y H S G L N
4340	4360	4380
ACGCCCAGGTCCACAAAAAC	ATTCTCTTCACAAGTCCTCA	CTCTTTCTGGACTAGGGGCA
H A Q V H K N	I L F T S P H	S F W T R G
4400	4420	4440
ATCTAAATGGCCTTATTCTC	TTGTGGTACACGTCCATGGT	CAGTCCGAATGGTTGAGCAC
N L N G L I L	L W Y T S M V	S P N G
4460	4480	4500
GAAAACTGCCTGATATGTGC	TCTTTTCTATCTCTCCTTCA	TTCTAAACGGGCCACTGTCT
4520	4540	4560
TGGTTTAAACATGAAAGGCC	CGGGATTGGTCTCGACATCG	TGGATCAAAGATCTGAAAA

Figure 3, continued.

4580	4600	4620
GTATCCTGCATAATGTGTTT	CGAAGACATGATACAGACCA	AAACATTCTCTTCACAAGT
4640	4660	4680
CCTCCACTCTTTCTGGACTA	GGGGCGTCTTCGATGGCCTT	ATTCTCTTGTGGTACACGTC
	M A L	F S C G T R
4700	4720	4740
CATGGTCACCGATGGGTTCA	GCACGCAAACTGCCTTGAT	ATGTGCTCTTCAACAGCTGG
P W S P M G S	A R K T A L I	C A L Q Q L
4760	4780	4800
ATACGACGATTCGTACGTTA	TGACGAATCCGAGTACGAGT	AGATTGACATCAATCTCGTC
D T (T I R T L		
4820	4840	4860
AACTGAAGAGTTACCACTA	AAATAAATACATGAAAACCA	TGCCTATTAGAATATTATGC
4880	4900	4920
GGGTAAAAACATAAAACCCC	GATGGAGCACTCCTTACTAG	<u>AACTAAACAAAATGAAAATT</u>
		M K I
4940	4960	4980
TTGTTAATATTAGCGTGTGT	GATTGCATGCGCATGTGGAG	AACGATATTGTGCTATGAAA
L L I L A C V	I A C A C G	E R Y C A M K
5000	5020	5040
TCAGATACAGATTTGTCATG	TCGCAATAGTACAGCGTCTG	ATTGTGAGTCATGCTTCAAC
S D T D L S C	R N S T A S	D C E S C F N
5060	5080	5100
GGAGGCGATCTTATATGGCA	TCTATCAAACCTGGAACCTCA	GCTGGTCTATAATATTGATC
G G D L I W H	L S N W N F	S W S I I L I
5120	5140	5160
GTTTTTATCACTGTGCTACA	ATATGGAAGACCTCAATTAA	GCTGGTTCGTGTATGGCATT
V F I T V L Q	Y G R P Q L	S W F V Y G I
5180	5200	5220
AAAATGCTTATAATGTGGCT	TTTATGGCCCGTTGTTTTGG	CTCTTACGATTTTTTAATGCA
K M L I M W L	L W P V V L	A L T I P N A
5240	5260	5280
TACTCGGAATATCAGCTGTC	CAGATATGTAATGTTGCGCT	TTAGTATTGCAGGTGCAATA
Y S E Y Q L S	R Y V M F G	F S I A G A I
5300	5320	5340
GTTACATTTGTACTCTGGAT	TATGTATTTTGTAAAGGTCCA	TTCAGTTGTACAGAAGGACT
V T F V L W I	M Y F V R S	I Q L Y R R T
5360	5380	5400
AACTCTTGGTGGTCTTTCAA	CCCTGAAACTAAAGCAATTC	TTTGCGTTAGTGCATTAGGA
N S W W S F N	P E T K A I	L C V S A L G
5420	5440	5460
AGGAGCTATGTGCTACCTCT	CGAAGGGGTGCCAACTGGTG	TCACTCTAACTTTGCTTTCA
R S Y V L P L	E G V P T G	V T L T L L S

Figure 3, continued.

5480	5500	5520
GGGAATTTGTACGCAGGAGG	GTTCAAAATTGCTGGTGGTA	TGAACATCGACAATTTACCA
G N L Y A G G	F K I A G G	M N I D N L P
5540	5560	5580
AAATACGTAATGGTTGCATT	ACCTATCAGGACTATTGTCT	ACACACTAGTTGGCAAGAAG
K Y V M V A L	P I R T I V	Y T L V G K K
5600	5620	5640
TTGAAAGCAAGTATTGCGAC	TGGGTGGGCTTACTATGTAA	AATCTAAAGCTGGGGATTAC
L K A S I A T	G W A Y Y V	L S K A G D Y
5660	5680	5700
TCAACAGAGGCAAGAAGTGA	TAATTTAAGTGAGCAAAAGA	AATTATTACATATGGTATAA
S T E A R S D	N L S E Q K	K L L H M V
5720	5740	5760
CTAAACTTTCTTAATGGCCA	ACCAGGGACAACGTGTCAGT	TGGGGAGATGAATCTACCAA
M A	N Q G Q R V S	W G D E S T K
5780	5800	5820
AACACGTGGTCGTTCCAATT	CCCGTGGTCGGAAGAATAAT	AACATACCTCTTTCATTCTT
T R G R S N	S R G R K N N	N I P L S F F
5840	5860	5880
CAACCCCATAAACCCTCCAAC	AAGATTCAAAATTTTGAAC	TTATGTCCGAGAGACTTTGT
N P I T L Q	Q D S K F W N	L C P R D F V
5900	5920	5940
CCCAAAGGAATAGGTAACA	GGGATCAACAGATTGGTTAT	TGGAATAGACAAACTCGCTA
P K G I G N	R D Q Q I G Y	W N R Q T R Y
5960	5980	6000
TGCGATGGTGAAGGGCCAAC	GTAAAGAGCTTCCTGAAAGG	TGGTTCTTCTACTACTTAGG
R M V K G Q	R K E L P E R	W F F Y Y L G
6020	6040	6060
TACTGGACCTCATGCAGATG	CCAAATTTAAAGATAAATTT	GATGGAGTTGTCTGGGTTGC
T G P H A D	A K F K D K F	D G V V W V A
6080	6100	6120
CAAGGATGGTGCCATGAACA	AACCAACCACGCTAGGAAGT	CGTGGTGCTAATAATGAATC
K D G A M N	K P T T L G S	R G A N N E S
6140	6160	6180
CAAAGCTTTGAAATTCGATG	GTAAAGTGCCAGGCGAATTT	CAACTTGAAGTTAATCAATC
K A L K F D	G K V P G E F	Q L E V N Q S
6200	6220	6240
AAGGGACAATTCAAGGTCAC	GCTCTCAATCTAGATCTCGG	TCTAGAAATAGATCTCAATC
R D N S R S	R S Q S R S R	S R N R S Q S
6260	6280	6300
TAGAGGCAGGCAACAATTCA	ATAACAAGAAGGATGACAGT	GTAGAACAAGCTGTTCTTGC
R G R Q Q F	N N K K D D S	V E Q A V L A

Figure 3, continued.

6320	6340	6360
CGCACTTAAAAAGTTAGGTG	TTGACACAGAAAAACAACAG	CAACGCTCTCGTTCTAAATC
A L K K L G	V D T E K Q	Q R S R S K S
6380	6400	6420
TAAAGAACGTAGTAACTCTA	AGACAAGAGAAACTACACCT	AAGAATGAAAACAAACACAC
K E R S N S	K T R E T T P	K N E N K H T
6440	6460	6480
CTCGAAGAGAACTGCAGGTA	AAGGTGATGTGACAAGATTT	TATGGAGCTAGAAGCAGTTC
S K R T A G	K G D V T R F	Y G A R S S S
6500	6520	6540
AGCCAATTTTGGTGACACTG	ACCTCGTTGCCAATGGGAGC	ACTGCCAAGCATTACCCACA
A N F G D T	D L V A N G S	T A K H Y P Q
6560	6580	6600
ACTGGCTGAATGTGTTCCAT	CTGTGTCTAGCATTCTGTTT	GGAAGCTATTGGACTTCAAA
L A E C V P	S V S S I L F	G S Y W T S K
6620	6640	6660
GGAAGATGGCGACCAGATAG	AAGTCACGTTACACACAAA	TACCACTTGCCAAAGGATGA
E D G D Q I	E V T F T H K	Y H L P K D D
6680	6700	6720
TCCTAAGACTGGACAATTCC	TTCAGCAGATTAATGCCTAT	GCTCGTCCATCAGAAGTGGC
P K T G Q F	L Q Q I N A Y	A R P S E V A
6740	6760	6780
AAAAGAACAGAGTAAAAGAA	AATCTCGTTCTAAATCTGCA	GAAAGGTCAGAGCAAGATGT
K E Q S K R	K S R S K S A	E R S E Q D V
6800	6820	6840
GGTACCTGATGCATTAATAG	AAAATTATACAGAAGTGTTT	GATGACACACAGGTTGAGAT
V P D A L I	E N Y T E V F	D D T Q V E I
6860	6880	6900
AATTGATGAGGTAACGAACT	AAACAAGATGCTCGTCTTCC	TCCATGCTGTATTTATTACA
I D E V T N	M L V F	L H A V F I T
6920	6940	6960
GTTTAAATCTTACTACTAAT	TGGTAGACTCCAATTATTAG	AAAGACTATTACTTGATCAC
V L I L L L I	G R L Q L L	E R L L L D H
6980	7000	7020
TCTTTCAATCTTAAACTGT	CAATGACTTTAATATCTTAT	ATAGGAGTTTTGCAGAAACC
S F N L K T V	N D F N I L	Y R S F A E T
7040	7060	7080
AGATTACTAAAAGTGGTGCT	TCGAGTAATCTTTCTAGTCT	TACTAGGATTTTGCTGCTAC
R L L K V V L	R V I F L V	L L G F C C Y
7100	7120	7140
AGATTGTTAGTCACCTTAGT	GTAAGGCAACCCGATACTAT	ACTACACTTTTAGCTACCAA
R L L V T L V		

Figure 3, continued.

7160	7180	7200
TCTAAATTAAGACGTCTACC	ACAGGTGCTGTTTGAAGGAG	GGTTTGTACCGATCAGACCT
7220	7240	7260
CTCTTTTCCTTTGGGGAAGT	GTAGAGTCGAGCATCACCGA	TGCTGTTTAGAGGGCCTTAA
7280	7300	7320
ATCTGGACAATGTTAACGGG	TAATAGGACGACAACTGCGG	CGTGAAGAGCTTGATGTAG
CCACA		

The consensus nucleotide sequences (see text) are underlined. Amino acids encoded by open reading frames are listed beneath the nucleotide sequence.

the corresponding region of the TGEV (Purdue strain) peplomer protein are illustrated in Figure 7.

An ORF 639 bases in length was found 126 bases downstream of the termination codon of the E2 gene. This ORF potentially encodes a polypeptide 213 amino acids in length, with a predicted molecular weight of 24.4 kd. One hundred seventy-two bases past the termination codon of this ORF was the ATG codon of a 282-base ORF that potentially encodes a 94 amino acid peptide. The location of these ORFs in the viral genome and the predicted lengths of the TGEV subgenomic RNAs suggest that the 639 base and 282 base ORFs may be found in the unique regions of TGEV (Miller strain) RNA 4a and RNA 5, respectively.

A long ORF of 786 bases extended from base 4831 through base 5697 of the predicted virion RNA sequence. Five ATG codons were present in the first 83 bases of this ORF. The polypeptide encoded using the fourth of these initiation codons was 262 amino acids in length with a molecular weight of 29.4 kd, a value close to the published molecular weight of the TGEV (Purdue strain) matrix protein (47). A hydrophilicity profile of the amino acid sequence was plotted (Figure 4). The profile was constructed with a running average of hydrophilicity taken over pentapeptides. Five regions of the polypeptide can be delineated: a hydrophobic signal peptide 17 residues in length (amino acids 1-17, encoded by bases 4912-4962), an exposed hydrophilic segment (amino acids 18-59, encoded by bases 4963-5088) containing two N-glycosylation sites (Asn-Ser-Thr encoded by bases 5005-5013 and Asn-Phe-Ser encoded by bases 5074-5082), three segments that may be incorporated into the viral envelope (amino acids 58-69, encoded by bases 5083-5118, amino acids 76-100, encoded by bases 5137-5211, and amino acids 123-139, encoded by bases 5277-5328), an amphiphilic C-terminal half (amino acids 140-262, encoded by bases 5329-5697) that interacts with the cytoplasmic face of the viral envelope, and a hydrophilic, protruding C-terminus (amino acids 238-257, encoded by bases 5623-5682) that contains a possible site of N-glycosylation (Asn-Leu-Ser, encoded by bases 5662-5670).

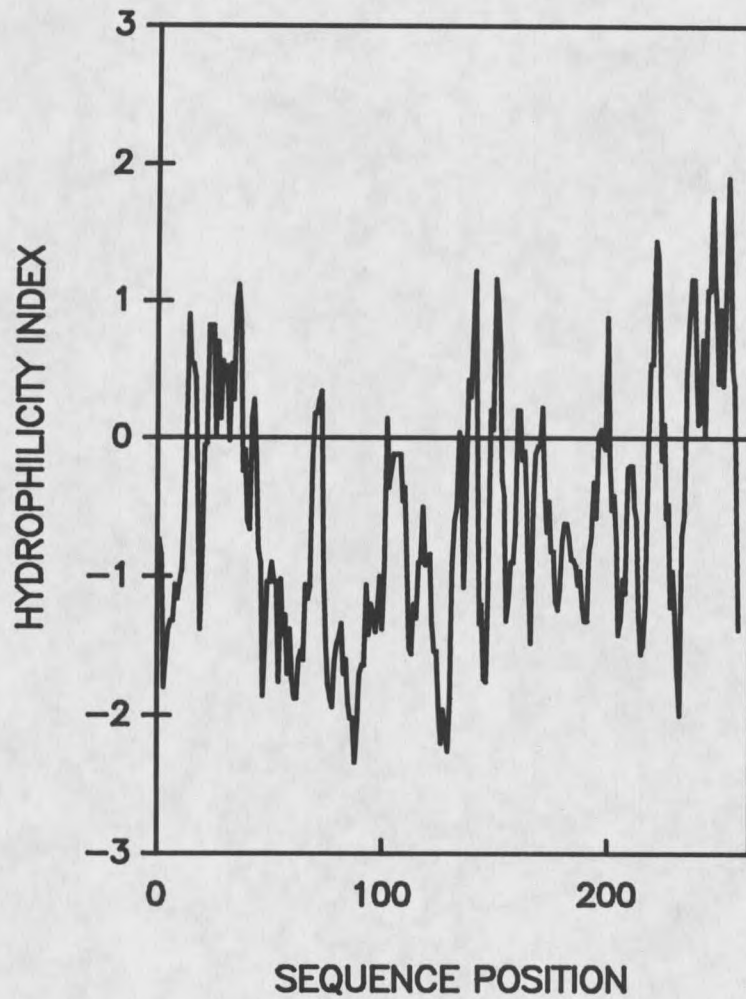


Figure 4. Hydrophilicity plot of the precursor to the matrix (E1) protein of TGEV (Miller strain). A running average was taken over pentapeptides using the hydrophilicity values of Hopp and Woods (33).

The first 17 amino acids encoded by the E1 ORF display a degree of hydrophobicity similar to that of eukaryotic signal peptides. Serine and threonine residues, potential but apparently unused sites of O-glycosylation, are present throughout the entire TGEV (Miller strain) E1 amino acid sequence. Fifteen amino acid differences exist between the matrix proteins of the Miller and Purdue strains of TGEV. The location of these differences is illustrated in Figure 7.

Seventeen bases downstream of the E1 ORF was the initiation codon of an 1146-base ORF that encodes a 382-residue protein. This ORF is 97.7% homologous in nucleotide sequence to the ORF that encodes the N protein of TGEV (Purdue strain) (38). The molecular weight of the protein encoded by the ORF of the Miller strain genome is 43,421 daltons, a value very close to the molecular weight of 43,426 daltons predicted by Kapke and Brian for the N protein of TGEV (Purdue strain). Thirty-nine of the 382 amino acids composing the Miller strain N protein are serine, the targets of phosphorylation in the MHV-A59 N protein. The changes in charge as well as molecular size due to phosphorylation of the TGEV N protein may explain the difference between the molecular weight predicted by the amino acid sequence of the protein and that predicted upon examination of the migration of virion-derived N protein in denaturing polyacrylamide gels. Although the

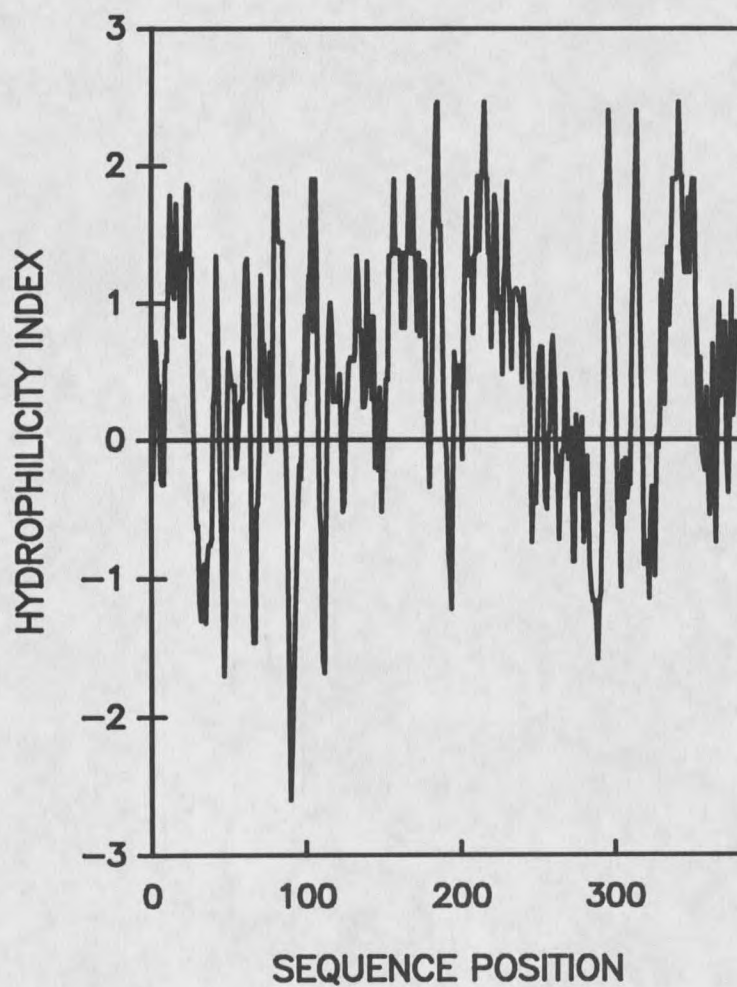


Figure 5. Hydrophilicity plot of the nucleocapsid (N) protein of TGEV (Miller strain). A running average was taken over pentapeptides using the hydrophilicity values of Hopp and Woods (33).

Purdue strain N polypeptide is of the same length as the protein encoded by the ORF found in clones of TGEV (Miller strain) virion RNA, six amino acid differences exist between the two. These differences are scattered over the entire amino acid sequence of N (Figure 7). A hydrophilicity profile revealed that the potential product of the TGEV (Miller strain) N gene contains clusters of charged residues along its entire length (Figure 5), a property of other coronavirus nucleocapsid proteins.

The last ORF of significant size in the virion RNA sequence was found to extend from base 6868 through base 7101 of the nucleotide sequence given in Figure 3. The initiation codon of this ORF lies six bases past the termination codon of the nucleocapsid gene. The coding capacity of this ORF is a 78-residue peptide of 9104 daltons. Twenty-three (29.5%) of the residues are leucine. Approximately twenty residues at each end of the peptide form hydrophobic regions, while the middle portion of the protein contains hydrophilic stretches (Figure 6). Overall, the protein contains 8 basic amino acids and 4 acidic amino acids, giving it a net positive charge at neutral pH. Each of the hydrophobic terminal sequences contain a basic residue; the remainder of the charged amino acids are scattered evenly throughout the hydrophilic central region.

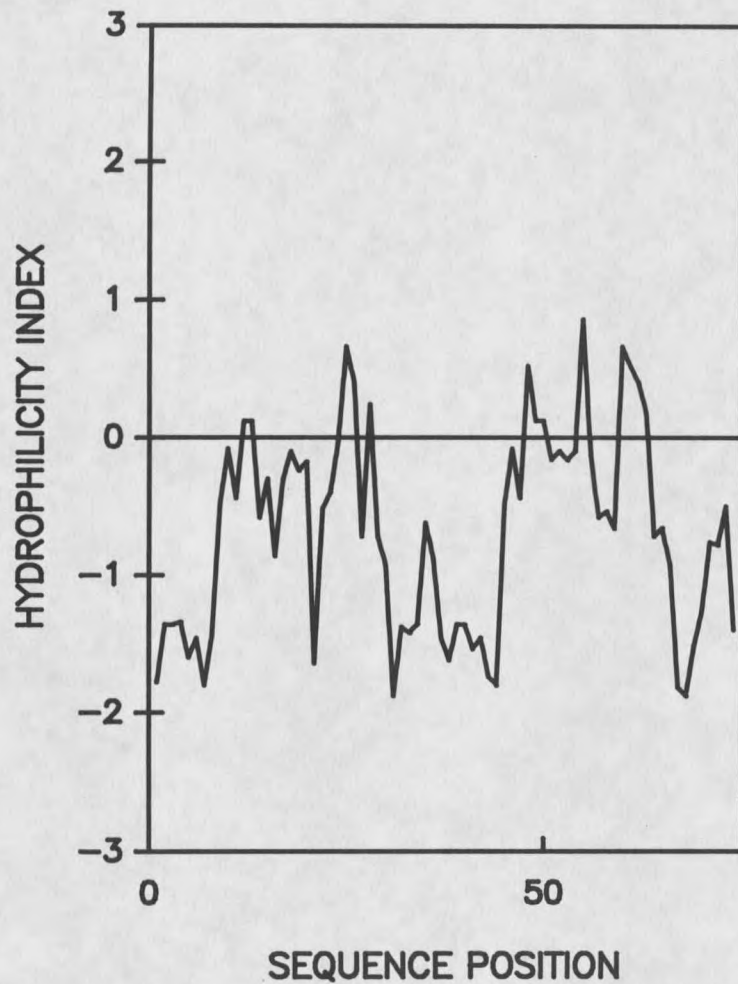
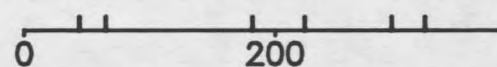
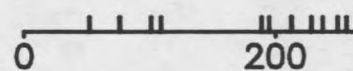
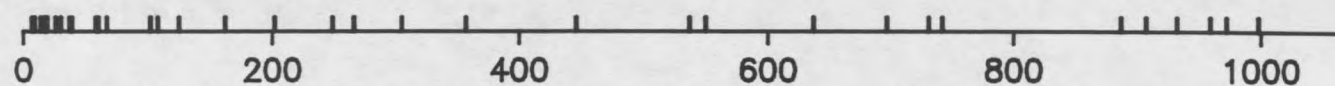


Figure 6. Hydrophilicity profile of the potential product of the open reading frame extending from base 6868 to base 7101 of the virion cDNA sequence given in Figure 3. A running average was taken over pentapeptides using the hydrophilicity values of Hopp and Woods (33).



RESIDUE

Figure 7. Positions of amino acid substitutions (vertical bars) in the TGEV (Miller strain) peplomer (E2), top, matrix (E1), middle, and nucleocapsid (N) protein, bottom, amino acid sequences compared to the primary structures of TGEV (Purdue strain) structural proteins.

The consensus sequence AACTAAAC or AATCTAAA precedes all of the ORFs described above. Several other ORFs encoding polypeptides ranging from 11 to 881 amino acids in length were found within, overlapping, or outside these ORFs. The location and/or length of these ORFs and the absence of consensus sequences suggested they were not primary protein-encoding units of TGEV subgenomic RNAs.

A 108-base ORF began 219 bases downstream of the ORF predicted to make up the protein-coding region of RNA 5. No consensus sequence was found immediately upstream from the 108-base ORF.

A restriction map of the 3' region of TGEV (Miller strain) virion RNA was constructed using the nucleotide sequence data. The map is consistent with maps constructed using data obtained from electrophoretic analysis of restricted cDNAs (Figure 2).

Time Course of Viral RNA Synthesis

The kinetics of TGEV (Miller strain) RNA synthesis were determined (Figure 8). Confluent monolayers of swine testicle cells in 6 cm dishes were infected with TGEV at an MOI of 5. At various times post-infection, DME-2 was removed from the cultures and replaced with 2 ml fresh DME-2 supplemented with 2.5 µg/ml actinomycin D. Twenty minutes after replacement of the medium, [³H]-uridine was added to 10 µCi/ml. At the times indicated in Figure 8,

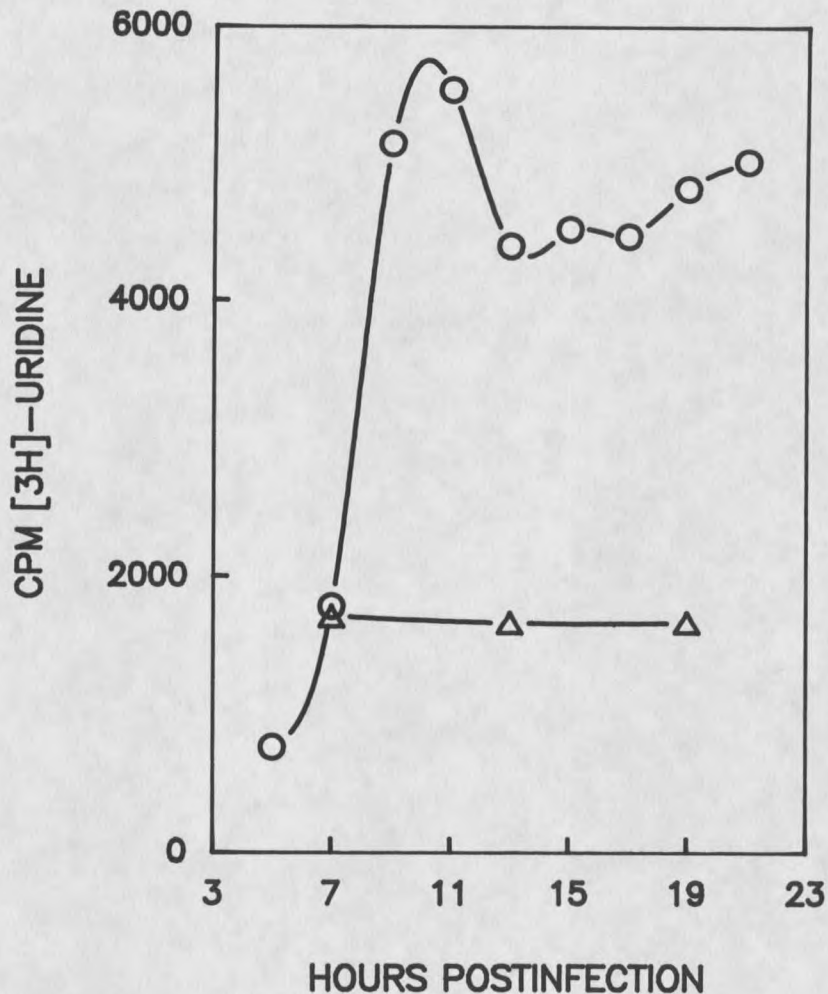


Figure 8. Time course of TGEV RNA synthesis. At the indicated times, ST cell monolayers were lysed and the amount of [^3H]-uridine incorporated into RNA was determined by measuring trichloroacetic acid-precipitable counts in the cell lysate. Open circles represent counts from TGEV (Miller strain)-infected cells, and open triangles represent counts from mock-infected cells. Data are expressed as counts per min (CPM) per 20 μl of lysate.

TGEV-specific RNA synthesis peaked at 10 to 12 h post-infection and decreased by 14 h post-infection. Subgenomic mRNAs are the majority of TGEV RNAs synthesized prior to 12 h post-infection (Andreas Luder, personal communication). Virion RNA synthesis is responsible for the second period of increase in [^3H]-uridine uptake, which occurs from 17 to 21 h post-infection. This interval corresponds to an increase in the quantity of genome-length viral RNA late in the virus multiplication cycle (Andreas Luder, personal communication). Cell lysis occurs concomitant with the second increase in RNA synthesis.

Urea-Agarose Gel Electrophoresis of TGEV mRNAs

Polyadenylated RNA 6 and RNA 7 from the Miller strain of TGEV were isolated for use as templates in oligo(dT)-primed cDNA synthesis. Polyadenylated RNA was separated from nonpolyadenylated RNA by chromatography on oligo(dT)-cellulose. Approximately 18% of labeled intracellular RNA from TGEV-infected, actinomycin D-treated ST cells was polyadenylated, based on the degree of binding to oligo(dT)-cellulose.

Polyadenylated TGEV RNAs were separated by electrophoresis in urea-agarose gels. The profiles of electrophoresed intracellular RNAs from TGEV (Miller strain)-infected ST cells and ST rRNA standards are shown in Figure 9. Peaks corresponding to molecules the

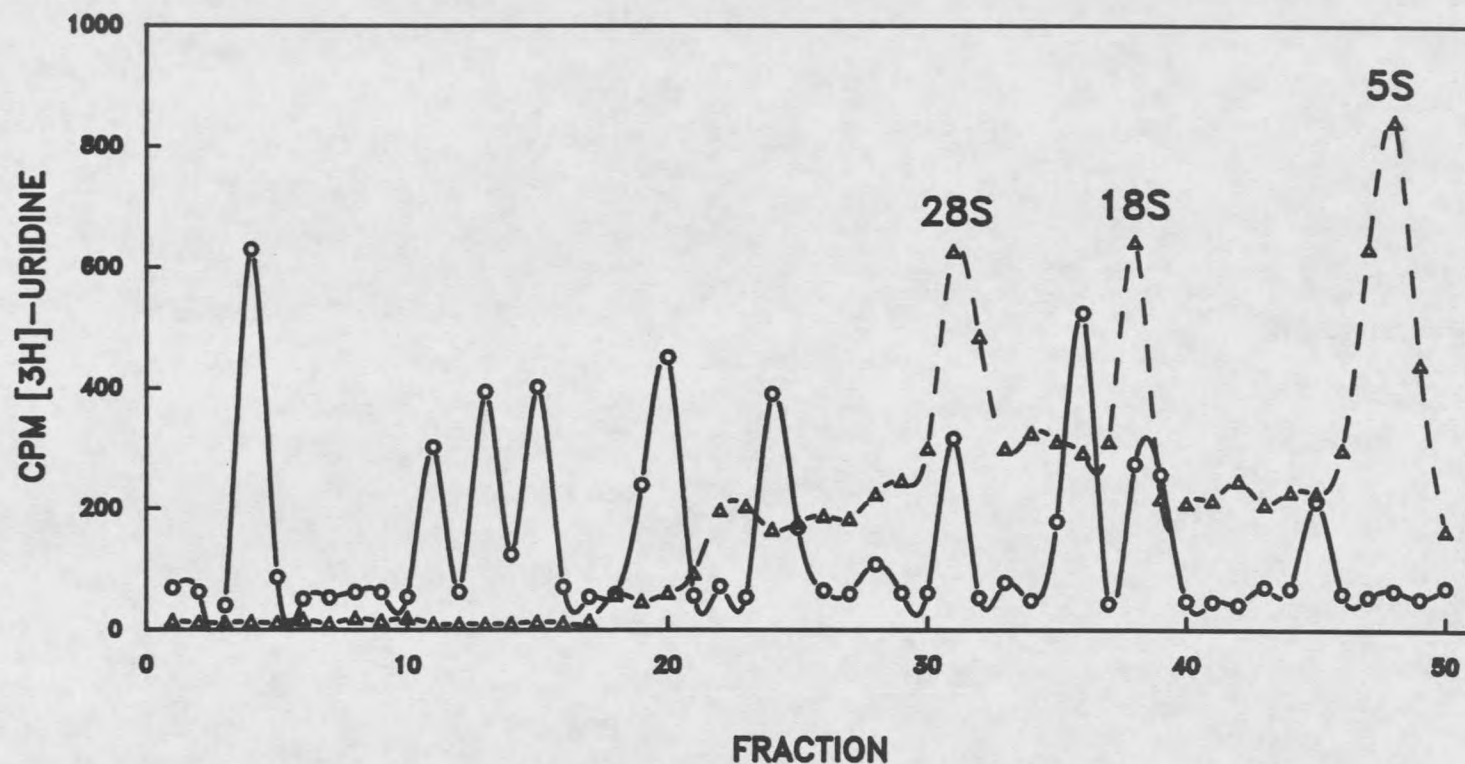


Figure 9. Profile of intracellular RNAs from TGEV (Miller strain)-infected ST cells (open circles) and ST cell ribosomal RNAs (open triangles) electrophoresed in urea-agarose. Gel slices were placed in aqueous scintillation fluid prior to measurement of [^3H]-uridine levels.

approximate sizes of RNA 6 (fractions 34-35) and RNA 7 (fractions 38-39), as estimated by the migration of 4.8 kb 28S rRNA (fraction 36) and 1.9 kb 18S rRNA (fractions 38 and 39), were evident. The appropriate gel slices from parallel lanes containing unlabeled polyadenylated RNA were pooled, diluted, and melted. Electrophoresed RNA was recovered from the molten agarose by extracting the slurries twice with chloroform.

In vitro translation of the recovered material was done to demonstrate that the RNAs were functional after recovery from urea-agarose, and also to confirm the identities of the RNAs by analysis of their gene products. Labeled translation products were immunoprecipitated with polyclonal murine anti-TGEV antiserum. The antiserum had been preadsorbed on methanol-fixed ST cell monolayers to reduce the immunoprecipitation of cellular proteins. Identification of translation products was based on immunologic recognition of viral proteins by TGEV-specific antiserum and the electrophoretic migration of these proteins in denaturing polyacrylamide gels as compared to standard proteins of known molecular weight (Figures 10 and 11). Slices predicted to contain mRNA 6 yielded RNA that encoded primarily a 48.5 kd polypeptide, equivalent to the TGEV N protein (38), and a 26 kd protein, equivalent to the unglycosylated form of E1 (47). The large amount of the 48.5 kd species may be due both to read-through translation

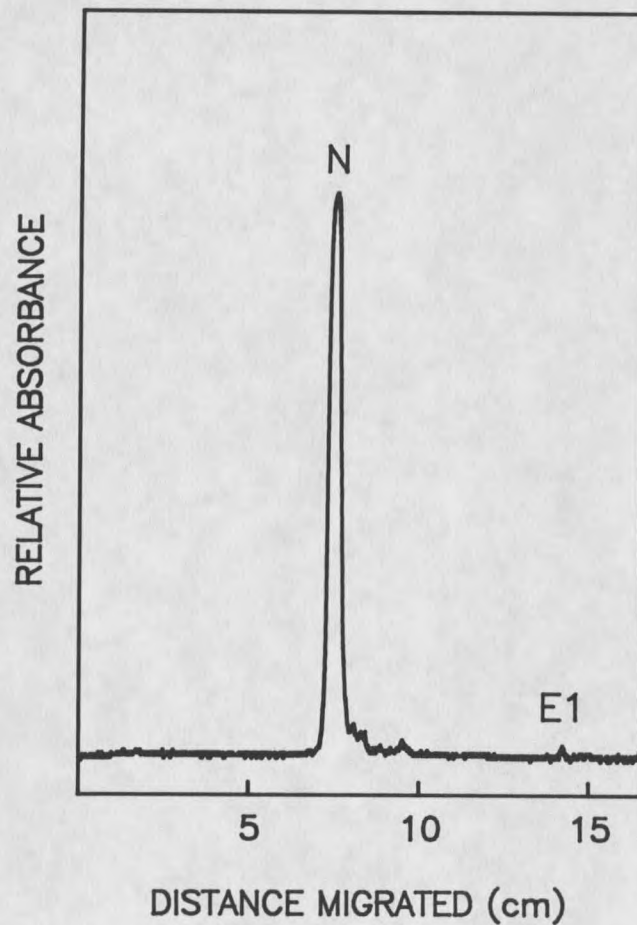


Figure 10. Densitometric analysis of TGEV (Miller strain) translation products separated in an SDS-10% polyacrylamide gel. The proteins are products of RNAs extracted from gel slices predicted to contain RNA 6. The peaks corresponding to the viral matrix (E1) and nucleocapsid (N) proteins are labeled.

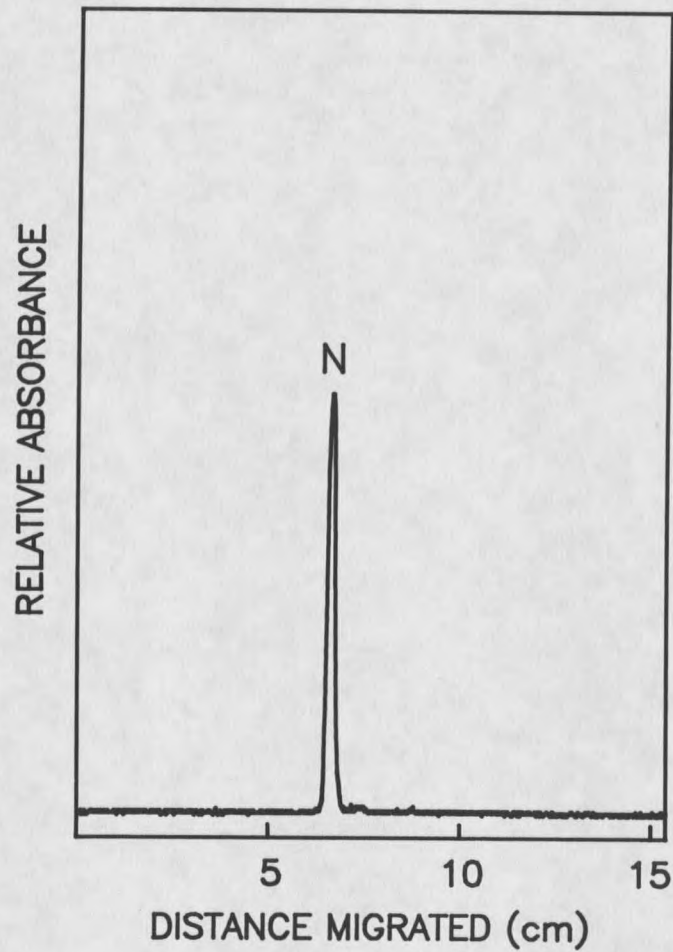


Figure 11. Densitometric analysis of TGEV (Miller strain) translation products separated in SDS-10% polyacrylamide gels. The proteins are products of RNAs extracted from gel slices predicted to contain RNA 7. The peak corresponding to the viral nucleocapsid (N) protein is labeled.

of RNA 6 and contamination of RNA 6 by RNA 7. RNA from the gel slices presumed to contain RNA 7 encoded a predominant product of 48.5 kd, a value very close to the published molecular weight of the TGEV (Purdue strain) nucleocapsid protein (38). Successful translation of the gel-purified RNAs indicated that these molecules were suitable templates for reverse transcription.

Cloning of TGEV RNA 6 and RNA 7

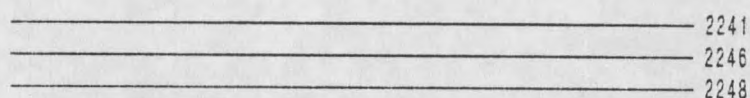
Complementary DNA copies of RNA 6 and RNA 7 were synthesized using oligo(dT) as the primer. RNA:cDNA hybrids were inserted into the plasmid pBR322. The size of inserts was estimated by the rate of migration of the recombinant plasmids in agarose gels relative to controls. Plasmids of the size class predicted to contain full length clones of RNA 7 were approximately 20 times more numerous than plasmids with cDNA inserts the length of RNA 6. Plasmids of appropriate size were characterized in hybridization experiments, as follows. Plasmids were digested with the restriction endonuclease PstI, electrophoresed, and transferred to Nytran membranes by the method of Reed and Mann (61). The blotted fragments were probed with labeled restriction fragments of a cDNA (clone 141) representative of the 3' end of TGEV (Miller strain) virion RNA. The 1.6 kb HindIII-PvuII fragment of clone 141 (Figure 3, bases 3624-5082) hybridized to the 5'-most PstI

fragment of RNA 6-specific clones, while the 0.6 kb HindIII-KpnI (Figure 3, bases 6124-6785) fragment hybridized to all PstI fragments of DNA copies of RNA 6 and RNA 7.

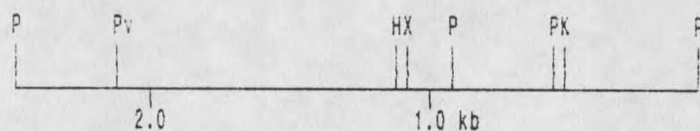
The longest of the RNA 6-specific cDNA clones was approximately 2.5 kb in length. Lengths varied according to the progression of first-strand cDNA synthesis. Restriction fragments of the longest cDNA inserts were subcloned in Riboprobe Gemini plasmids and sequenced. The map of RNA 6-specific cDNA and the fragments sequenced are given in Figure 12. The nucleotide sequence of RNA 6-specific cDNA and amino acids encoded by the major ORFs are illustrated in Figure 13. The first 35 bases of this clone were not found in the corresponding region of virion RNA-specific cDNA, while the sequence of the remainder of the clone was identical with that of DNA copied from the 3' end of TGEV (Miller strain) virion RNA. The 35-base stretch not found in the virion RNA-specific cDNA may represent all or part of a leader sequence derived from the 5' end of the viral genome that primes transcription of RNA 6.

An 867-base ORF was found in the 5' portion of RNA 6 and the corresponding region of virion RNA. An 8-nucleotide consensus sequence, AACTAAAC, was found near each end of this putative gene. The first of the potential start codons was found immediately downstream from this consensus sequence. The 5' end of this ORF corresponded to

A



B



C

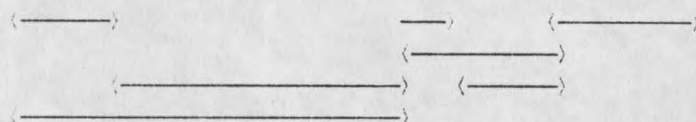


Figure 12. Restriction endonuclease map of TGEV (Miller strain) RNA 6-specific cDNA and strategy used in nucleotide sequence determination. A. RNA 6-specific cDNA clones. B. Map of RNA 6-specific cDNA. C. Restriction fragments of clones that were subcloned in Riboprobe Gemini sequencing vectors. Arrows indicate direction in which fragments were sequenced. Restriction enzymes: PvuII (Pv), HindIII (H), XbaI (X), PstI (P), KpnI (K).

base 4912 of the virion cDNA sequence illustrated in Figure 3. The putative translation product of the ORF is a polypeptide with the properties of E1. Two ATG codons

Figure 13. Nucleotide sequence of TGEV (Miller strain) RNA 6-specific cDNA.

20	40	60
TAAAACTCTTGGTAGTTTAA	ATCTAATCTA <u>ACTAAACAAA</u>	ATGAAAATTTTGTTAATATT
		M K I L L I L
80	100	120
AGCGTGTGTGATTGCATGCG	CATGTGGAGAACGATATTGT	GCTATGAAATCAGATACAGA
A C V I A C	A C G E R Y C	A M K S D T D
140	160	180
TTTGTTCATGTCGCAATAGTA	CAGCGTCTGATTGTGAGTCA	TGCTTCAACGGAGGCGATCT
L S C R N S	T A S D C E S	C F N G G D L
200	220	240
TATATGGCATCTATCAAAC	GGAACCTTCAGCTGGTCTATA	ATATTGATCGTTTTTATCAC
I W H L S N	W N F S W S I	I L I V F I T
260	280	300
TGTGCTACAATATGGAAGAC	CTCAATTAAGCTGGTTCGTG	TATGGCATTAAATGCTTAT
V L Q Y G R	P Q L S W F V	Y G I K M L I
320	340	360
AATGTGGCTTTTATGGCCCG	TTGTTTTGGCTCTTACGATT	TTTAATGCATACTCGGAATA
M W L L W P	V V L A L T I	P N A Y S E Y
380	400	420
TCAGCTGTCCAGATATGTAA	TGTTCCGGCTTTAGTATTGCA	GGTGCAATAGTTACATTTGT
Q L S R Y V	M F G F S I A	G A I V T F V
440	460	480
ACTCTGGATTATGTATTTTG	TAAGGTCCATTCAAGTTGTAC	AGAAGGACTAACTCTTGGTG
L W I M Y F	V R S I Q L Y	R R T N S W W
500	520	540
GTCTTTCAACCCTGAAACTA	AAGCAATTCTTTCGCTTAGT	GCATTAGGAAGGAGCTATGT
S F N P E T	K A I L C V S	A L G R S Y V
560	580	600
GCTACCTCTCGAAGGGGTGC	CAACTGGTGTCACTCTAACT	TTGCTTTTCAGGGAATTTGTA
L P L E G V	P T G V T L T	L L S G N L Y
620	640	660
CGCAGGAGGGTTCAAAATTG	CTGGTGGTATGAACATCGAC	AATTTACCAAATACGTAAT
A G G F K I	A G G M N I D	N L P K Y V M
680	700	720
GGTTGCATTACCTATCAGGA	CTATTGTCTACACACTAGTT	GGCAAGAAGTTGAAAGCAAG
V A L P I R	T I V Y T L V	G K K L K A S
740	760	780
TATTGCGACTGGGTGGGCTT	ACTATGTAAAATCTAAAGCT	GGGGATTACTCAACAGAGGC
I A T G W A	Y Y V L S K A	G D Y S T E A
800	820	840
AAGAAGTGATAATTTAAGTG	AGCAAAAGAAATTATTACAT	ATGGTATA <u>ACTAAACT</u> TTTCT
R S D N L S	E Q K K L L H	M V
860	880	900
TAATGGCCAACCAGGGACAA	CGTGTCAAGTTGGGGAGATGA	ATCTACCAAAACACGTGGTC
M A N Q G Q	R V S W G D E	S T K T R G

Figure 13, continued.

920	940	960
GTTCCAATTCCCGTGGTCCG	AAGAATAATAACATACCTCT	TTCATTCTTCAACCCCATAA
R S N S R G R	K N N N I P L	S F F N P I
980	1000	1020
CCCTCCAACAAGATTCAAAA	TTTTGGAAC TTATGTCCGAG	AGACTTTGTACCCAAAGGAA
T L Q Q D S K	F W N L C P R	D F V P K G
1040	1060	1080
TAGGTAACAGGGATCAACAG	ATTGGTTATTGGAATAGACA	AACTCGCTATCGCATGGTGA
I G N R D Q Q	I G Y W N R Q	T R Y R M V
1100	1120	1140
AGGGCCAACGTAAAGAGCTT	CCTGAAAGGTGGTTCTTCTA	CTACTTAGGTACTGGACCTC
K G Q R K E L	P E R W F F Y	Y L G T G P
1160	1180	1200
ATGCAGATGCCAAATTTAAA	GATAAATTTGATGGAGTTGT	CTGGGTTGCCAAGGATGGTG
H A D A K F K	D K F D G V V	W V A K D G
1220	1240	1260
CCATGAACAAACCAACCACG	CTAGGAAGTCGTGGTGCTAA	TAATGAATCCAAAGCTTTGA
A M N K P T T	L G S R G A N	N E S K A L
1280	1300	1320
AATTCGATGGTAAAGTGCCA	GGCGAATTTCAACTTGAAGT	TAATCAATCAAGGGACAATT
K F D G K V P	G E F Q L E V	N Q S R D N
1340	1360	1380
CAAGGTCACGCTCTCAATCT	AGATCTCGGTCTAGAAATAG	ATCTCAATCTAGAGGCAGGC
S R S R S Q S	R S R S R N R	S Q S R G R
1400	1420	1440
AACAATTCAATAACAAGAAG	GATGACAGTGTAGAACAAGC	TGTTCTTGCCGCACTTAAAA
Q Q F N N K K	D D S V E Q A	V L A A L K
1460	1480	1500
AGTTAGGTGTTGACACAGAA	AAACAACAGCAACGCTCTCG	TTCTAAATCTAAAGAACGTA
K L G V D T E	K Q Q Q R S R	S K S K E R
1520	1540	1560
GTAACCTAAGACAAGAGAA	ACTACACCTAAGAATGAAAA	CAAACACACCTCGAAGAGAA
S N S K T R E	T T P K N E N	K H T S K R
1580	1600	1620
CTGCAGGTAAAGGTGATGTG	ACAAGATTTTATGGAGCTAG	AAGCAGTTCAGCCAATTTTG
T A G K G D V	T R F Y G A R	S S S A N F
1640	1660	1680
GTGACACTGACCTCGTTGCC	AATGGGAGCACTGCCAAGCA	TTACCCACAACCTGGCTGAAT
G D T D L V A	N G S T A K H	Y P Q L A E
1700	1720	1740
GTGTTCCATCTGTGTCTAGC	ATTCTGTTTGAAGCTATTG	GACTTCAAAGGAAGATGGCG
C V P S V S S	I L F G S Y W	T S K E D G
1760	1780	1800
ACCAGATAGAAGTCACGTTT	ACACACAAATACCACTTGCC	AAAGGATGATCCTAAGACTG
D Q I E V T F	T H K Y H L P	K D D P K T

Figure 13, continued.

1820	1840	1860
GACAATTCCTTCAGCAGATT	AATGCCTATGCTCGTCCATC	AGAAGTGGCAAAAGAACAGA
G Q F L Q Q I	N A Y A R P S	E V A K E Q
1880	1900	1920
GTAAAAGAAAATCTCGTTCT	AAATCTGCAGAAAGGTCAGA	GCAAGATGTGGTACCTGATG
S K R K S R S	K S A E R S E	Q D V V P D
1940	1960	1980
CATTAATAGAAAATTATACA	GAAGTGTTTGATGACACACA	GGTTGAGATAATTGATGAGG
A L I E N Y T	E V F D D T Q	V E I I D E
2000	2020	2040
TAACGA <u>ACTAAACA</u> AGATGC	TCGTCTTCCTCCATGCTGTA	TTTATTACAGTTTAACTCT
V T N	M L V F L H A V	F I T V L I L
2060	2080	2100
ACTACTAATTGGTAGACTCC	AATTATTAGAAAGACTATTA	CTTGATCACTCTTTCAATCT
L L I G R L	Q L L E R L L	L D H S F N L
2120	2140	2160
TAAAACTGTCAATGACTTTA	ATATCTTATATAGGAGTTTT	GCAGAAACCAGATTACTAAA
K T V N D F	N I L Y R S F	A E T R L L K
2180	2200	2220
AGTGGTGCTTCGAGTAATCT	TTCTAGTCTTACTAGGATTT	TGCTGCTACAGATTGTTAGT
V V L R V I	F L V L L G F	C C Y R L L V
2240	2260	2280
CACCTTAGTGTAAGGCAACC	CGATACTATACTACACTTTT	AGCTACCAATCTAAATTAAG
T L V		
2300	2320	2340
ACGTCTACCACAGGTGCTGT	TTGAAGGAGGGTTTGTACCG	ATCAGACCTCTCTTTTCCTT
2360	2380	2400
TGGGGAAGTGTAGAGTCGAG	CATCACCGATGCTGTTTAGA	GGGCCTTAAATCTGGACAAT
2420	2440	2460
GTTAACGGGTAATAGGACGA	CAACTGCGGCGTGAAGAGC	TTGATGTAGCCACATTCTCC
AAAAAAAAAAAAAAAA		

The consensus nucleotide sequences (see text) are underlined. Amino acids encoded by the matrix (E1) protein gene (bases 41-826), the nucleocapsid (N) protein gene (bases 843-1988), and a 234-base open reading frame are listed beneath the nucleotide sequence.

were found in the extreme 5' portion of the gene. Following this codon is a sequence that encodes a stretch of 17 hydrophobic amino acids. The hydrophobic peptide may serve as a leader sequence that translocates E1 to the endoplasmic reticulum of TGEV-infected cells. The second ATG codon closely follows the sequence encoding the 17 hydrophobic residues, but microsequencing of virion-associated E1 has determined that this codon is not the site at which translation of E1 is initiated (47). Also, the hydrophobic amino acid sequence was not present in virion-associated E1 (47), indicating that the leader peptide is removed during processing of the primary translation product. The predicted molecular weight of the mature unglycosylated protein from which the hydrophobic leader peptide has been cleaved is 27.7 kd.

Three potential sites of N-glycosylation were detected in the amino acid sequence of E1. Two of these sites are near the amino terminus of the protein, while the third is located 12 residues from the carboxyl terminus of the polypeptide. Many potential sites of O-glycosylation are present in the predicted amino acid sequence, but this type of linkage has not been detected in the matrix glycoprotein of TGEV.

A second ORF representative of the sequence that encodes the nucleocapsid (N) protein was found downstream

of the E1 gene. The start codon for this gene is located 13 bases past the termination codon of the E1-encoding ORF.

Only one additional ORF more than 20 amino acids in length was found within RNA 6-specific cDNA; this ORF has a potential product of 78 amino acids.

RNA 7 of the Miller strain of TGEV was copied into DNA; the longest of the cDNA clones was approximately 1.7 kb in length. A restriction map of RNA 7-specific cDNA and the strategy employed in sequence determination are illustrated in Figure 14. The nucleotide sequence of RNA 7-specific cDNA is given in Figure 15. All of the sequence but the first 54 bases of the 5' end were found in the virion cDNA sequence from position 5698 to position 7325 (Figure 3) and in the RNA 6-specific cDNA sequence from position 833 to the polyadenylic acid tail. The clones contained an ORF of 1146 bases that encoded a basic polypeptide 382 residues in length. This protein has the properties of coronavirus nucleocapsid proteins, and is described above. In vitro translation of urea-agarose gel-purified TGEV (Miller strain) RNA indicated that a protein the predicted size of N is the primary translation product of RNA 7 (Figure 10).

The cDNA clones of RNA 7 contained an additional ORF of 234 bases downstream of the ORF that encodes the 43.3 kD N protein; the ATG codon of this ORF is found 24 bases past the termination codon of the N gene. This ORF begins at base 1225 of the RNA 7-specific clone and extends through

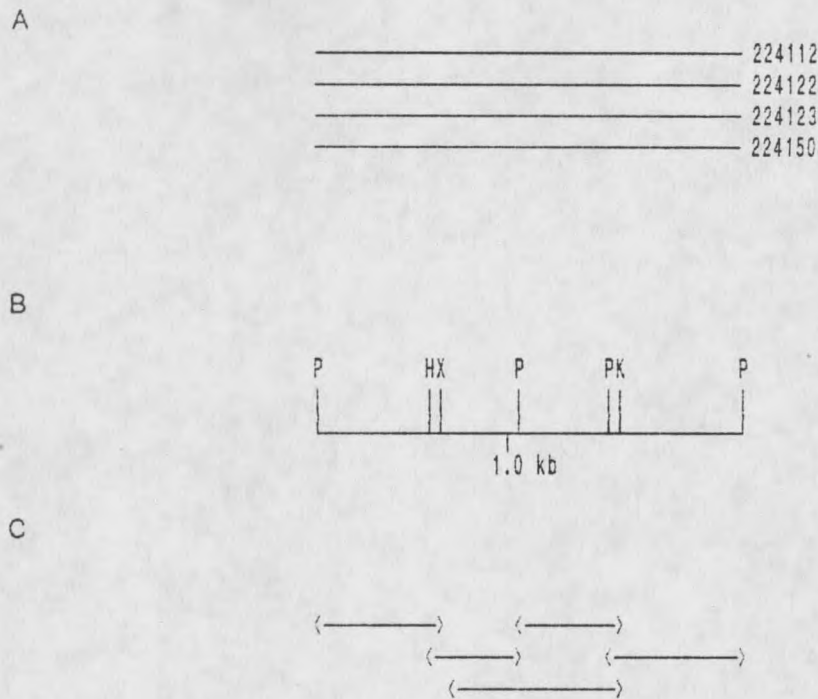


Figure 14. Restriction endonuclease map of TGEV (Miller strain) RNA 7-specific cDNA and strategy used in determination of nucleotide sequence. A. RNA 7-specific cDNA clones used in mapping and subcloning experiments. B. Restriction map of RNA 7-specific cDNA. C. Restriction fragments that were subcloned in the multiple cloning region of Riboprobe Gemini sequencing vectors. Arrows indicate direction in which fragments were sequenced. Restriction enzymes: PstI (P), HindIII (H), XbaI (X), KpnI (K).

Figure 15. Nucleotide sequence of TGEV (Miller strain) RNA 7-specific cDNA.

20	40	60
TCCCGTACGGTACCCCTCCT	CTACTCTAAACTCTTGGTA	GTTTAAATCTAATCTAACTA
80	100	120
<u>AACTTTCTTAATGGCCAACC</u>	AGGGACAACGTGTCAATTGG	GGAGATGAATCTACCAAAAC
M A N	Q G Q R V S W	G D E S T K T
140	160	180
ACGTGGTCGTTCCAATTCCC	GTGGTCGGAAGAATAATAAC	ATACCTCTTTCATTCTTCAA
R G R S N S	R G R K N N N	I P L S F F N
200	220	240
CCCCATAACCCTCCAACAAG	ATTCAAAATTTTGGAACTTA	TGTCCGAGAGACTTTGTACC
P I T L Q Q	D S K F W N L	C P R D F V P
260	280	300
CAAAGGAATAGGTAACAGGG	ATCAACAGATTGGTTATTGG	AATAGACAAACTCGCTATCG
K G I G N R	D Q Q I G Y W	N R Q T R Y R
320	340	360
CATGGTGAAGGGCCAACGTA	AAGAGCTTCCTGAAAGGTGG	TTCTTCTACTACTTAGGTAC
M V K G Q R	K E L P E R W	F F Y Y L G T
380	400	420
TGGACCTCATGCAGATGCCA	AATTTAAAGATAAATTTGAT	GGAGTTGTCTGGGTTGCCAA
G P H A D A	K F K D K F D	G V V W V A K
440	460	480
GGATGGTGCCATGAACAAAC	CAACCACGCTAGGAAGTCGT	GGTGCTAATAATGAATCCAA
D G A M N K	P T T L G S R	G A N N E S K
500	520	540
AGCTTTGAAATTCGATGGTA	AAGTGCCAGGCGAATTTCAA	CTTGAAGTTAATCAATCAAG
A L K F D G	K V P G E F Q	L E V N Q S R
560	580	600
GGACAATTCAAGGTCACGCT	CTCAATCTAGATCTCGGTCT	AGAAATAGATCTCAATCTAG
D N S R S R	S Q S R S R S	R N R S Q S R
620	640	660
AGGCAGGCAACAATTCAATA	ACAAGAAGGATGACAGTGTA	GAACAAGCTGTTCTTGCCGC
G R Q Q F N	N K K D D S V	E Q A V L A A
680	700	720
ACTTAAAAAGTTAGGTGTTG	ACACAGAAAAACAACAGCAA	CGCTCTCGTTCTAAATCTAA
L K K L G V	D T E K Q Q Q	R S R S K S K
740	760	780
AGAACGTAGTAACTCTAAG	CAAGAGAAACTACACCTAAG	AATGAAAACAAACACACCTC
E R S N S K	T R E T T P K	N E N K H T S
800	820	840
GAAGAGAACTGCAGGTAAAG	GTGATGTGACAAGATTTTAT	GGAGCTAGAAGCAGTTCAGC
K R T A G K	G D V T R F Y	G A R S S S A
860	880	900
CAATTTTGGTGACACTGACC	TCGTTGCCAATGGGAGCACT	GCCAAGCATTACCCACAAC
N F G D T D	L V A N G S T	A K H Y P Q L

Figure 15, continued.

920	940	960
GGCTGAATGTGTTCCATCTG	TGTCTAGCATTCTGTTTGA	AGCTATTGGACTTCAAAGGA
A E C V P S	V S S I L F G	S Y W T S K E
980	1000	1020
AGATGGCGACCAGATAGAAG	TCACGTTACACACAAATAC	CACTTGCCAAAGGATGATCC
D G D Q I E	V T F T H K Y	H L P K D D P
1040	1060	1080
TAAGACTGGACAATTCCTTC	AGCAGATTAATGCCTATGCT	CGTCCATCAGAAGTGGCAA
K T G Q F L	Q Q I N A Y A	R P S E V A K
1100	1120	1140
AGAACAGAGTAAAAGAAAAT	CTCGTTCTAAATCTGCAGAA	AGGTCAGAGCAAGATGTGGT
E Q S K R K	S R S K S A E	R S E Q D V V
1160	1180	1200
ACCTGATGCATTAATAGAAA	ATTATACAGAAGTGTGTTGAT	GACACACAGGTTGAGATAAT
P D A L I E	N Y T E V F D	D T Q V E I I
1220	1240	1260
TGATGAGGTAACGA <u>ACTAAA</u>	CAAGATGCTCGTCTTCCTCC	ATGCTGTATTTATTACAGTT
D E V T N	M L V F L	H A V F I T V
1280	1300	1320
TTAATCTTACTACTAATTGG	TAGACTCCAATTATTAGAAA	GACTATTACTTGATCACTCT
L I L L L I G	R L Q L L E	R L L L D H S
1340	1360	1380
TTCAATCTTAAACTGTCAA	TGACTTTAATATCTTATATA	GGAGTTTTGCAGAAACCAGA
F N L K T V N	D F N I L Y	R S F A E T R
1400	1420	1440
TTACTAAAAGTGGTGCTTCG	AGTAATCTTTCTAGTCTTAC	TAGGATTTTGCTGCTACAGA
L L K V V L R	V I F L V L	L G F C C Y R
1460	1480	1500
TTGTTAGTCACCTTAGTGTA	AGGCAACCCGATACTATACT	ACACTTTTAGCTACCAATCT
L L V T L V		
1520	1540	1560
AAATTAAGACGTCTACCACA	GGTGCTGTTTGAAGGAGGGT	TTGTACCGATCAGACCTCTC
1580	1600	1620
TTTTCTTTGGGGAAGTGTA	GAGTCGAGCATCACCGATGC	TGTTTAGAGGGCCTTAAATC
1640	1660	1680
TGGACAATGTTAACGGGTAA	TAGGACGACAAGTGC	GGAAGAGCTTGATGTAGCCA
1700		
CAGACGTCATTA	AAAAAAAAAAAAAAAAAAAA	

The consensus nucleotide sequences (see text) are underlined. Amino acids encoded by the nucleocapsid (N) protein gene (bases 71-1216) and a 234-base open reading frame (bases 1225-1458) are listed beneath the nucleotide sequence.

base 1458. A 78 amino acid protein of 9104 daltons is encoded by the putative gene. A hydrophilicity analysis of the gene product revealed that the first twenty and last 25 amino acids composed hydrophobic sequences (Figure 5). The central region is amphiphilic and contains both basic and acidic amino acids. Twenty-three (29.5%) of the amino acids in the protein are serine. There is no evidence supporting the existence of this peptide in virions or TGEV-infected cells.

Several shorter ORFs, encoding peptides 11 to 52 amino acids in length, are present within, overlapping, or past the 5' end of the ORF encoding N.

A noncoding region approximately 230 bases in length was found at the 3' end of TGEV (Miller strain) virion RNA and the two subgenomic RNAs sequenced in this study. This noncoding region immediately preceeds the poly A tails of the RNAs.

Comparison of the maps of RNA 6-specific cDNA and RNA 7-specific cDNA to each other and to the map of the 3' 7325 bases of virion cDNA suggested that RNA 7 is a subset of RNA 6, and the maps of both of these molecules were identical to the map of the 3' portion of the virion RNA.

DISCUSSION

Research on TGEV genetic structure has focused on virion RNA. These studies have revealed open reading frames within the genome and have made possible the determination of the primary structure of the peplomer (E2) (60), matrix (E1) (47), and nucleocapsid (N) (38) proteins of the attenuated Purdue strain of TGEV. None of these projects dealt with the subgenomic mRNAs of TGEV. Data presented in this thesis suggests a discontinuous model of gene transcription for TGEV that is similar to that proposed for MHV and IBV (9,43,73,75,78,79,82).

Following penetration of the virus into a cell, the virion RNA directs synthesis of an RNA-dependent RNA polymerase (88). A minus-strand copy of the viral genome is transcribed by this enzyme, and the minus-strand RNA is the template for subgenomic mRNA synthesis. Synthesis of the mRNAs makes up the majority of early RNA production. The kinetics of TGEV-specific RNA synthesis in ST cells were studied to determine the time of maximum synthesis of virus-specific subgenomic RNA. RNA was extracted from infected cells at this time for the preparation of subgenomic viral RNA. TGEV (Miller strain)-infected ST

cell cultures reach their maximum yield of virus at 18 h post-infection (67), when approximately 50 to 75% of the cells have lysed. At this time cell lysates were harvested, virus was partially purified, and virion RNA was extracted.

TGEV (Miller strain) Virion RNA

Complementary DNA was prepared from TGEV (Miller strain) genomic RNA extracted from partially purified virions. First-strand cDNA synthesis was primed by a restriction fragment from an oligo(dT)-primed, TGEV (Miller strain)-specific cDNA clone (clone 141, provided by Andreas Luder). The specificity of the primer extension clones was determined by Southern blotting. These clones, along with oligo(dT)-primed clones 141 and 150, obtained from Andreas Luder, were mapped. The restriction maps were used to develop a sequencing strategy in which restriction fragments subcloned in Riboprobe Gemini sequencing vectors were overlapping, eliminating the possibility of incomplete sequence analysis due to inadvertent omission of very small restriction fragments.

Complementary DNA clones, produced by both primer extension and oligo(dT)-primed cDNA synthesis, representing the 3' 7325 nucleotides of the TGEV (Miller strain) genome were sequenced. A portion of the ORF encoding the peplomer protein of the virus was at the 5' end of the sequence.

The sequence of the C-terminal 1061 amino acids of the peplomer protein was predicted from the nucleotide sequence data. Because the entire peplomer-encoding region was not represented by the cDNA clones studied, few conclusions concerning the structure of the Miller strain's protein spikes can be made. However, comparison of the partial amino acid sequence I obtained to the primary structure of the Purdue strain peplomer published by Rasschaert and Laude (60) revealed 33 amino acid differences between the two proteins. The degree of divergence was 3.39 substitutions per 100 amino acids. Most of these differences were found in the more N-terminal region of the incomplete sequence of the TGEV (Miller strain) peplomer. Rasschaert and Laude (60) found that most regions of homology between the peplomer of TGEV (Purdue strain) and that of IBV are clustered in the carboxyl halves of the molecules. Sequences in the amino halves were divergent. Comparison by Jacobs et al. (36) of the primary structure of the peplomer proteins of TGEV (Purdue strain) and feline infectious peritonitis virus (FIPV) found positions of amino acid substitutions to be more numerous in the amino halves of the molecules. Since the peplomer proteins of coronaviruses mediate binding of the virus to cells and determine many of the subsequent alterations in cell physiology and structure (17,30,37,95), changes due to attenuation may be concentrated in exposed regions of the

peplomer. However, the hydrophilicity profile of the TGEV (Purdue strain) peplomer protein displays few highly hydrophilic (exposed) segments in the amino half but several in the carboxyl half (60). Delmas et al. (17) have shown that four neutralization epitopes reside in the more amino terminal region of the peplomer protein of the Purdue strain of TGEV. The data presented here indicate divergence in amino acid sequence between the two strains in the more amino terminal region, but also in the more carboxyl region. This information may be useful in locating determinants of TGEV pathogenicity.

An ORF corresponding to the unique region of RNA 4a was found in the Miller strain virion cDNA sequence. The polypeptide encoded by this ORF is 212 residues in length and has a molecular weight of 24.4 kd. In the TGEV (Miller strain) genome, the consensus nucleotide sequence AACTAAAC was located 23 bases upstream of the potential ORF of RNA 4a, a distance greater than that which separates the consensus sequence from the E1 and N protein genes of RNA 6 and RNA 7, respectively. No function has been assigned to the RNA 4 gene product. The protein is probably nonstructural. In vitro translation of RNA 4 of the Purdue strain resulted in a 24 kd protein that was not detected in infected cells or purified virions (35). Also, the protein was not immunoprecipitated with antiserum directed against virion proteins.

A short ORF that may correspond to the protein-encoding region of RNA 5 was found in the virion RNA sequence. The 282-base ORF encodes a 94-residue peptide. No function has been assigned to the product of RNA 5. Jacobs et al. (35) detected no virus-specific translation product for the molecule. The 8-base consensus sequence, AACTAAAC, preceding the ORFs of RNAs 4a, 6, and 7 was not found in the region preceding the ORF of RNA 5, although a similar sequence, AATCTAAA, was found to overlap the ATG codon of the ORF. The AATCTAAA consensus sequence is identical to an 8-base sequence immediately upstream of the peplomer gene of MHV (strain JHM) (69). Consensus sequences may regulate initiation or level of transcription of the subgenomic mRNA. Alteration of this sequence, as in the case of RNA 5, may result in synthesis of a decreased amount of the mRNA. No differences in the sequence of consensus regions were found between the Miller and Purdue strains of TGEV.

The gene encoding the E1 matrix protein was identified in the sequence of virion RNA-specific cDNA. A single long ORF 786 bases in length encodes a 262-residue polypeptide with a molecular weight of 29.4 kd. Fifteen differences in primary structure were found between the matrix proteins of the Miller and Purdue strains of TGEV. The substitution ratio of 5.72 per 100 amino acids sequenced was the highest of the TGEV structural proteins. Present in the primary

translation product of the E1 gene is a hydrophobic stretch of 17 amino acids that may serve as a signal sequence for translocation of E1 to the endoplasmic reticulum of infected cells. Like many eukaryotic signal peptides, this oligopeptide has a charged N-terminal region followed by a stretch of uncharged residues (41). Microsequencing of virion-associated E1 by Laude et al. (47) indicates that this leader sequence is cleaved from E1 during maturation of the protein. This is in contrast to the matrix proteins of coronaviruses MHV and IBV. Matrix proteins of these viruses appear to be translocated to membranes by recognition of an internal hydrophobic region that is present in the virion-associated protein. Three N-glycosylation sites (Asn-X-Ser or Asn-X-Thr) were present in the amino acid sequence. All glycosylation of the TGEV E1 protein has been reported to be N-linked; no O-linked glycosylation, as occurs in the matrix proteins of MHV and BCV, has been discovered. However, clusters of serine and threonine residues, potential sites of O-linked glycosylation, were found in the predicted amino acid sequence of TGEV (Miller strain) E1.

The TGEV nucleocapsid protein contains many basic residues, a property expected of RNA-binding proteins (38). However, hydrophobic regions were detected throughout the entire amino acid sequence, including the amino terminus. The hydrophobic regions may play a role in the association

of the nucleocapsid with the viral envelope. The nucleocapsid protein displayed the least divergence in amino acid sequence between the Miller and Purdue strains of TGEV. Because the N protein is not exposed in intact virions, there is little selective pressure for variability in primary structure.

TGEV (Miller strain) RNA 6 and RNA 7

Because the mRNAs of coronaviruses MHV and IBV have been demonstrated to form a 3'-coterminal nested set of multiple species, "shotgun" cloning of bulk RNA from TGEV-infected cells was not a satisfactory means of obtaining DNA copies of subgenomic mRNAs. Premature termination of first-strand synthesis on a TGEV RNA 3 template, for example, could result in a cDNA clone the length of RNA 6 but lacking nucleotides present in its 5'-noncoding region. To determine if TGEV employs a leader-primed method of transcription, it was necessary to obtain full-length copies of RNA 6 and RNA 7 rather than truncated clones of larger mRNAs. To separate TGEV RNAs on the basis of size, urea-agarose gel electrophoresis was used. Eucaryotic 18S and 28S rRNA, molecules approximately 1900 and 4800 nucleotides in length, respectively, were markers useful in estimating the location of TGEV RNAs 6 and 7 in the denaturing gels. Extraction of the gel slices twice with chloroform was found to be an efficient method of

extracting the electrophoresed RNA. RNAs recovered from urea-agarose were sufficiently resolved and functional, as demonstrated by in vitro translation experiments carried out to confirm the identities and integrity of the gel-purified molecules (Figures 9 and 10). RNA from gel slices predicted to contain TGEV (Miller strain) RNA 6 produced protein products of 26 and 48.5 kd. The 26 kd protein was presumably the unglycosylated precursor of the E1 protein, while the 48.5 kd protein was probably nucleocapsid protein encoded by mRNA 7 not adequately separated from mRNA 6. Incomplete separation of mRNA 6 from mRNA 5 and mRNA 7 by isokinetic sucrose-gradient centrifugation has been reported by Jacobs et al. (35), and electrophoresis in urea-agarose of two RNAs differing in length by less than 800 bases may also have not provided resolution sufficient for complete separation. Alternatively, the 48.5 kd protein could be the product of read-through of an open reading frame coding for N in mRNA 6. The consensus sequence upstream of the N protein gene in RNA 7 also precedes the N protein gene of RNA 6. There is a one-base shift between the E1- and N-encoding ORFs of RNA 6, but both read-through translation and independent initiation of N gene translation may be possible. The 48.5 kd polypeptide was the predominant protein product of RNA recovered from the peak presumed to contain RNA 7. The nucleocapsid protein is the most abundant virus-specific

polypeptide in infected cells, and it was of no surprise to obtain much larger amounts of N than E1 in the translation experiments. These results support the coding assignments given by Jacobs et al. (35). The ability of the gel-purified molecules to encode complete proteins suggested there was little loss of structural integrity during electrophoresis and recovery, and that the RNA molecules would be suitable templates for reverse transcription. Oligo(dT) was used as the primer of first strand synthesis to obtain the 3' end of the polyadenylated mRNAs. RNA:cDNA hybrids, rather than double-stranded cDNAs, were inserted into vector plasmids by means of complementary homopolymeric tails. Elimination of second-strand cDNA synthesis reduced the chance of sequence loss due to incomplete synthesis of second-strand cDNA by E. coli DNA polymerase I. Also, the 3' end of the RNA templates was likely to extend past the 5' end of first-strand cDNA. Protruding 3' ends are favored substrates for homopolymeric tail addition by terminal deoxynucleotidyl transferase (54).

Two main size classes of cDNA inserts were obtained. Estimation of insert length by electrophoresis of the recombinant plasmids suggested that the groups were composed primarily of clones of TGEV (Miller strain) RNA 6 and RNA 7. The clones were further characterized in hybridization experiments. The 1.6 kb HindIII-PvuII

fragment of virion RNA-specific clone 141, which extends from 2092 to 3704 bases from the 3' end of the genome, was chosen as a probe because, if TGEV mRNAs form the nested set arrangement characteristic of coronaviruses MHV and IBV, it would hybridize to copies of mRNA 6 but not to copies of mRNA 7. The 662 bp HindIII-KpnI fragment of clone 141 was predicted to hybridize to clones of both mRNA 6 and mRNA 7. These predictions were confirmed upon analysis of Southern blots. The results were the first evidence that TGEV (Miller strain) mRNAs form a nested set. Restriction fragments from subgenomic RNA-specific cDNAs were subcloned in Riboprobe Gemini vectors for direct sequence determination. The nucleotide sequence of RNA 6- and RNA 7-specific cDNA was determined and from this information and virion cDNA sequence data the primary structure of the precursor to the matrix and nucleocapsid proteins encoded by the messages were predicted (Figures 11 and 13). Differences in sequence between the cDNAs of the Miller and Purdue strain were discovered; the possible influence of these differences on the pathogenicity of the virus is discussed below.

TGEV (Miller strain) RNA 6

RNA 6, which encodes the matrix protein E1 of TGEV, was copied into DNA clones up to 2483 bases in length. All but the first 35 bases of the nucleotide sequence derived from

these clones corresponded to the final 2431 bases of the predicted virion cDNA sequence. The 35-base leader sequence was not found in the portion of the viral genome sequenced in this study. The leader may be transcribed from the 5'-terminal portion of the TGEV genome, as is true of MHV and IBV leader sequences (9,43,78,79). Priming of transcription of the subgenomic RNAs during discontinuous transcription may be the function of the leaders. It is not known if dissociation of the polymerase/leader complex from the template takes place during this process. It seems likely that a sequence in the 3' portion of the leader sequence may anneal to regions of the virion RNA preceeding the ORFs of the subgenomic RNAs. Leader-primed transcription of the body sequences of the RNAs could then begin. Conservation of the AATAAAC consensus sequences in regions of the genome flanking the ORFs of RNAs 3, 4, 6, and 7 suggest that these sequences play a role in primer recognition. No similarity in sequence was found in the regions of the viral genome immediately preceeding the consensus sequences. The 5' noncoding regions of the viral RNAs may also regulate expression of gene products at the translational level, as differences in the location of the consensus sequence relative to ORFs was noted in the virion cDNA sequence. Differences in quantity of each TGEV subgenomic RNA have been described (48,94). The differences in quantity of the RNAs may act in concert with

differences in the 5' noncoding region in regulating the levels of viral protein produced in infected cells.

The mature matrix (E1) protein of TGEV (Miller strain), as predicted by cDNA sequence data, is 27.7 kd. The migration of virion-derived E1 in denaturing polyacrylamide gels suggested a molecular weight of 29 kd (67). That only 5 to 6% of the molecular weight of virion-associated E1 from the Miller strain of TGEV is carbohydrate is in marked contrast to the peplomer glycoprotein of TGEV. The carbohydrate moiety of the TGEV (Purdue strain) peplomer may account for approximately 27% of the molecule's total molecular size (60).

TGEV (Miller strain) RNA 7

The protein encoded by the long ORF of TGEV (Miller strain) RNA 7 is 382 amino acids long and its molecular weight as predicted from cDNA sequence data is 43.4 kd. This protein shares similarities with the N proteins of coronaviruses MHV and IBV, although the polypeptides are of different lengths (382 amino acids for TGEV, 455 for MHV, and 409 for IBV). First, it is serine-rich. Thirty-nine of 382 amino acids are serine. Serine residues are the probable sites of phosphorylation in the MHV-A59 N protein (83). If the same is true of the TGEV nucleocapsid protein, the change in molecular weight and charge of the molecule upon phosphorylation could explain the difference

between the molecular weight predicted by the amino acid sequence (43.4 kd) and that predicted by the comparative migration of virion-derived N protein in denaturing gels (54 kd) (67).

There is no evidence for the 9104 dalton protein potentially encoded by bases 1225 through 1458, an ORF outside the region of RNA 7 that encodes N. This protein has not been detected in translation experiments or analyses of intracellular virus-specific or virion proteins. Three small polyadenylated RNAs have been found in TGEV-infected cells (35, Andreas Luder, personal communication); RNAs analagous to these have not been reported for MHV- and IBV-infected cells. If all TGEV subgenomic RNAs form a 3'-coterminal nested set, one of these RNAs may be the molecule that encodes this protein. The consensus sequence AACTAAAC precedes the 234-base ORF, lending support to the possibility of independent transcription and subsequent translation of the gene. Also, a six-base sequence, AGAUGC, overlapping the initiation codon of the gene is a sequence found in eukaryotic initiation sequences (41).

The length and position of the ORF and comparison of the nucleotide sequence of mRNA 6 to that of the virion RNA and mRNA 7 established the location of the E1 ORF to be the region of mRNA 6 not found in mRNA 7.

TGEV (Miller strain) Gene Transcription

Evidence presented in this thesis suggests a mechanism of TGEV (Miller strain) gene transcription similar to that proposed for other coronaviruses. This mechanism is illustrated in Figure 16. RNA 6 and RNA 7 of TGEV (Miller strain) possess common 5' leader sequences not transcribed from regions of the virion RNA immediately upstream from the subgenomic RNA body sequences the genome contains. The data suggests that the leader sequences are transcribed from the 5' end of the TGEV genome. The leaders could serve as primers of subgenomic RNA synthesis, perhaps by annealing of a consensus sequence at the 3' end of the leader to complementary consensus sequences upstream of major ORFs in the genome-length negative-strand template. Laude et al. (47) assumed the consensus sequences to be the start of mRNA transcripts, but if TGEV replication involves discontinuous transcription it would be more accurate to say that discontinuous transcription, primed by leader sequences, begins at the consensus regions. The leader sequences would be the actual 5' ends of subgenomic RNAs. That the subgenomic RNAs of TGEV are not transcribed in equal amounts may be due to differences in efficiency of binding of the leader sequence to the various consensus regions in the negative-strand RNA. Secondary structure of

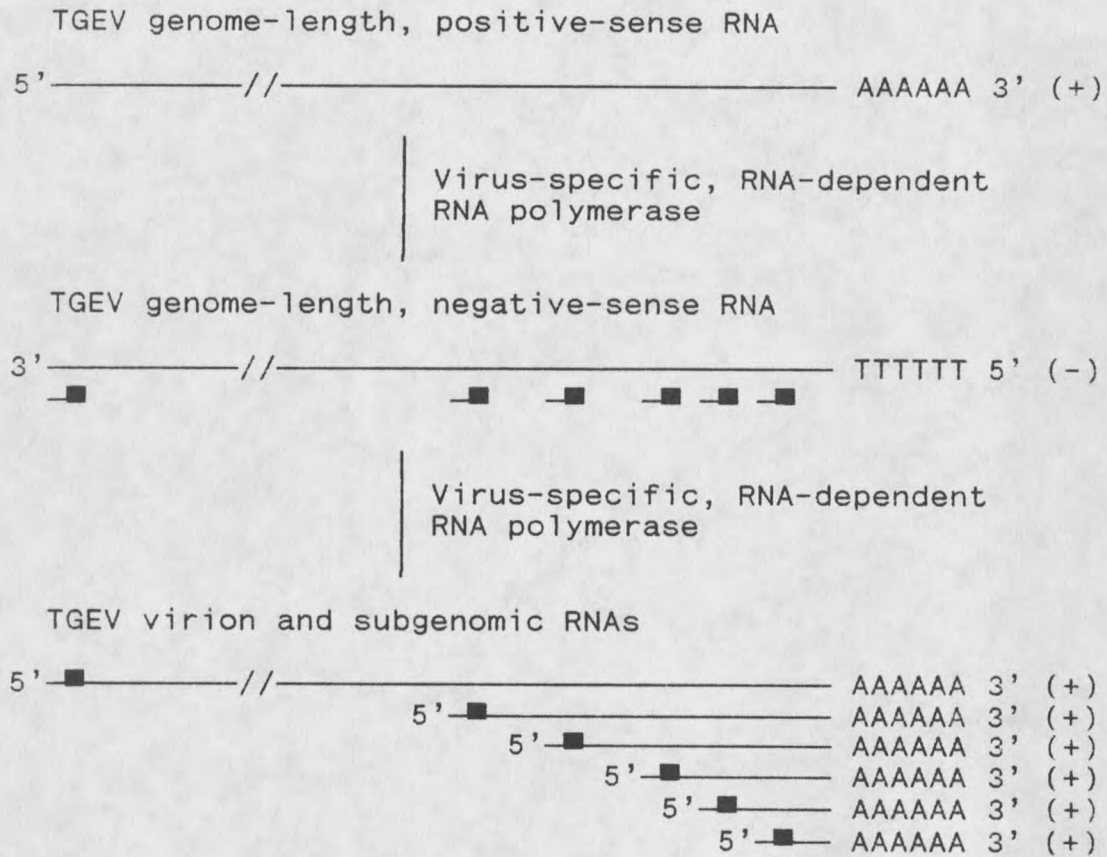


Figure 16. Proposed mechanism of TGEV gene transcription. A portion of a leader sequence transcribed from the 3' end of a negative-sense copy of the viral genome anneals to consensus sequences downstream and primes the synthesis of at least five subgenomic mRNAs as well as genome-length RNA. The RNAs form a 3'-coterminal nested set. The filled box represents the complement to the consensus sequence.

the negative strand, as well as slight differences in the consensus sequences, could determine this efficiency.

The sequence data presented above provides strong evidence that TGEV (Miller strain) subgenomic RNAs form a 3'-coterminial nested set. Primary protein products of RNA 6 and RNA 7 are encoded by the 5' unique regions of the messages.

An RNA approximately 500 bases in length that contains the 234-base ORF that follows the N gene has not been definitively identified, but its small size and possible low copy number in TGEV (Miller strain) infected cells may make detection difficult. The consensus sequence immediately upstream of the ORF provides a site for transcription of the small RNA.

Differences in the amino acid sequence of structural proteins were found between the virulent Miller strain and attenuated Purdue strain of TGEV. Primary structure was most divergent in the two membrane glycoproteins, E2 and E1. Because these proteins contain neutralization epitopes, variance in their primary structure may help the virus evade the immune system of the host. Most of the variability in the peplomer amino acid sequence was found near the N-terminus of the protein; the N-terminus is the location of both conserved and variable neutralization epitopes (46). Divergence in the E1 amino acid sequences of the two strains was more common in the amphiphilic C-

terminal half of the protein, a region predicted to be exposed on the surface of TGEV virions.

The results indicate that the low-passage, virulent Miller strain of TGEV used in this study differs from the high-passage Purdue strain used by other groups in studies of TGEV biology and biochemistry. It may be prudent to further examine the properties of the Miller strain, as it is less far removed from the TGEV that swine are likely to encounter. Immunogenic surface proteins of the Miller strain are likely to resemble more closely antigens of wild-type TGEV particles. Advancement toward production of a protective vaccine may be hastened by employment of TGEV (Miller strain) in future studies of the swine pathogen, rather than the attenuated Purdue strain.

CONCLUSIONS

The gene structure of the pathogenic Miller strain of transmissible gastroenteritis virus (TGEV) was studied. Nucleotide sequence data from two subgenomic RNAs, RNA 6 and RNA 7, and the 3' region of the viral genome was obtained. My research suggests that TGEV employs a leader-primed mechanism of discontinuous transcription by which a sequence transcribed from the 5' end of the viral genome recognizes and anneals to consensus sequences within the virion RNA. The annealed sequence may serve as a primer for transcription of a nested set of 3'-coterminial subgenomic mRNAs.

From the nucleotide sequence data open reading frames were identified and the primary structure of TGEV structural proteins was predicted. Nucleotide and amino acid sequence data and translational studies indicate that the protein-coding sequences of RNA 6 and RNA 7 are located in the 5'-terminal regions of the molecules. Substantial differences in the amino acid sequence of structural proteins were found between the pathogenic Miller strain and the attenuated Purdue strain.

The data collected in this study will be useful in locating determinants of pathogenicity in TGEV. Ultimately this research will lead to production of a safe, effective vaccine against porcine transmissible gastroenteritis.

LITERATURE CITED

1. 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.
2. Armstrong, J., S. Smeekens, P. Rottier, and B. van der Zeist. 1984. Cloning and sequencing the nucleocapsid and E1 genes of coronavirus MHV-A59. *Adv. Exp. Med. Biol.* 173:155-162.
3. Binns, M.M., M.E.G. Boursnell, D. Cavanagh, D.J.C. Pappin, and T.D.K. Brown. 1985. Cloning and sequencing of the gene encoding the spike protein of coronavirus IBV. *J. Gen. Virol.* 66:719-726.
4. Birnboim, H.C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7:1513-1523.
5. Bohl, E.H. 1981. Transmissible gastroenteritis, p. 195-208. In A.D. Leman, R.D. Glock, W.L. Mengeling, R.H.C. Penny, E. Scholl, and B. Straw (ed.), *Diseases of Swine*. Iowa State University Press, Ames.
6. Bohl, E.H., R.K.P. Gupta, L.W. McCloskey, and L. Saif. 1972. Immunology of transmissible gastroenteritis. *J. Am. Vet. Med. Assoc.* 160:543-548.
7. Bond, C.W., J.L. Leibowitz, and J.A. Robb. 1979. Pathogenic murine coronaviruses. II. Characterization of virus-specific proteins of murine coronaviruses JHMV and A59V. *Virology* 94:371-384.
8. Bond, C.W., K. Anderson, S. Goss, and L. Sardinia. 1981. Relatedness of virion and intracellular proteins of the murine coronaviruses JHM and A59. *Adv. Exp. Med. Biol.* 142:103-110.
9. Brown, T.D.K., M.E.G. Boursnell, and M.M. Binns. 1984. A leader sequence is present on mRNA A of avian infectious bronchitis virus. *J. Gen. Virol.* 65:1437-1442.
10. Butler, D.G., D.G. Gall, M.H. Kelly, and J.R. Hamilton. 1974. Transmissible gastroenteritis: Mechanism responsible for diarrhea in an acute viral enteritis in piglets. *J. Clin. Invest.* 53:1335-1342.
11. Caul, E.O., C.R. Ashley, M. Ferguson, and S.I. Egglestone. 1979. Preliminary studies on the isolation of coronavirus 229E nucleocapsids. *FEMS Microbiol. Lett.* 5:101-105.

12. Cavanaugh, D. 1981. Structural polypeptides of coronavirus IBV. *J. Gen. Virol.* 53:93-103.
13. Cavanaugh, D. 1983. Coronavirus IBV: Structural characterization of the spike protein. *J. Gen. Virol.* 64:2577-2583.
14. Cavanaugh, D. 1983. Structural characterization of IBV glycoproteins. *Adv. Exp. Med. Biol.* 173:95-108.
15. Collins, A.R., R.L. Knobler, H. Powell, and M.J. Buchmeier. 1982. Monoclonal antibodies to murine hepatitis virus-4 (strain JHM) define the viral glycoprotein responsible for attachment and cell-cell fusion. *Virology* 119:358-371.
16. David-Ferreira, J.F., and R.A. Manaker. 1965. An electron microscope study of the development of a mouse hepatitis virus in tissue culture cells. *J. Cell. Biol.* 24:57-78.
17. Delmas, B., J. Guelfi, and H. Laude. 1986. Antigenic structure of transmissible gastroenteritis virus. II. Domains in the peplomer glycoprotein. *J. Gen. Virol.* 67:1405-1418.
18. Fairbanks, G., T.L. Stock, and D.F.H. Wallach. 1971. Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. *Biochemistry* 10:2606-2617.
19. Frederick, G.T., E.H. Bohl, and R.F. Cross. 1976. Pathogenicity of an attenuated strain of transmissible gastroenteritis virus for newborn pigs. *Am. J. Vet. Res.* 37:165-169.
20. Fristensky, B., J. Lis, and R. Wu. 1982. Portable microcomputer software for nucleotide sequence analysis. *Nucleic Acids Res.* 10:6451-6463.
21. Garwes, D.J. and D.H. Pocock. 1975. The polypeptide structure of transmissible gastroenteritis virus. *J. Gen. Virol.* 29:25-34.
22. Garwes, D.J., D.H. Pocock, and B.V. Pike. 1976. Isolation of subviral components from transmissible gastroenteritis virus. *J. Gen. Virol.* 32:283-294.
23. Garwes, D.J., M.H. Lucas, D.A. Higgins, B.V. Pike, and S.F. Cartwright. 1978/1979. Antigenicity of structural components from porcine transmissible gastroenteritis virus. *Vet. Microbiol.* 3:179-190.

24. Grunstein, M., and D. Hogness. 1975. Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. U.S.A.* 72:3961-3965.
25. Gubler, U., and B.J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* 25:263-269.
26. Haelterman, E.O. 1972. On the pathogenesis of transmissible gastroenteritis of swine. *J. Am. Vet. Med. Assoc.* 160:534-539.
27. Hanahan, D. Studies on the transformation of *Escherichia coli* with plasmids. 1983. *J. Mol. Biol.* 166:557-580.
28. Hasony, H.J. and M.R. MacNaughton. 1981. Antigenicity of mouse hepatitis virus strain 3 subcomponents in C57 strain mice. *Arch. Virol.* 69:33-41.
29. Hogg, A. 1982. TGE: Epizootic and Enzootic. *Mod. Vet. Practice* 8:489-492.
30. Holmes, K.V., E.W. Doller, and J.N. Behnke. 1981. Analysis of the functions of coronavirus glycoproteins by differential inhibition of synthesis with tunicamycin. *Adv. Exp. Med. Biol.* 142:133-142.
31. Holmes, K.V., E.W. Doller, and L.S. Sturman. 1981. Tunicamycin resistant glycosylation of a coronavirus glycoprotein: Demonstration of a novel type of viral glycoprotein. *Virology* 115:334-344.
32. Holmes, K.V., M.F. Frana, S.G. Robbins, and L.S. Sturman. 1984. Coronavirus maturation. *Adv. Exp. Med. Biol.* 173:37-52.
33. 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.
34. Jacobs, L., W.J.M. Spaan, M. C. Horzinek, and B.A.M. van der Zeijst. 1981. Synthesis of subgenomic mRNAs of mouse hepatitis virus is initiated independently: Evidence from UV transcription mapping. *J. Virol.* 39:401-406.
35. Jacobs, L., B.A.M. van der Zeijst, and M.C. Horzinek. 1986. Characterization and translation of transmissible gastroenteritis virus mRNAs. *J. Virol.* 57:1010-1015.

36. Jacobs, L., R. de Groot, B.A.M. van der Zeijst, M.C. Horzinek, and W. Spaan. 1987. The nucleotide sequence of the peplomer gene of porcine transmissible gastroenteritis virus (TGEV): Comparison with the sequence of the peplomer protein of feline infectious peritonitis virus (FIPV). *Virus Research*. 8:363-371.
37. Jimenez, G., I. Correa, M.P. Melgosa, M.J. Bullido, and L. Enjuanes. 1986. Critical epitopes in transmissible gastroenteritis virus neutralization. *J. Virol.* 60:131-139.
38. Kapke, P.A., and D.A. Brian. 1986. Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. *Virology* 151:41-49.
39. Kessler, S.W. 1975. Rapid isolation of antigens from cells with a staphylococcal protein A-antibody adsorbent: Parameters of the interaction of antibody-antigen complexes with protein A. *J. Immunol.* 115:1617-1624.
40. King, B. and D.A. Brian. 1982. Bovine coronavirus structural proteins. *J. Virol.* 42:700-707.
41. Kozak, M. 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. *Microbiol. Rev.* 47:1-45.
42. Laemmli, U.K. and M. Favre. 1973. Maturation of the head of bacteriophage T4. I. DNA packaging events. *J. Mol. Biol.* 80:575-599.
43. Lai, M.M.C., C.D. Patton, and S.A. Stohman. 1982. Further characterization of mouse hepatitis virus: Presence of common 5'-end nucleotides. *J. Virol.* 41:557-565.
44. Lai, M.M.C., R.S. Baric, P.R. Brayton, and S.A. Stohman. 1984. Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proc. Natl. Acad. Sci. U.S.A.* 81:3626-3630.
45. Laude, H., B. Charley, and C. La Bonnardiere. 1984. Interactions of porcine enteric coronavirus TGEV with macrophages and lymphocytes. *Adv. Exp. Med. Biol.* 173:385-386.

46. Laude, H., J.-M. Chapsal, J. Guelfi, S. Laibiau, and J. Grosclaude. 1986. Antigenic structure of transmissible gastroenteritis virus. I. Properties of monoclonal antibodies directed against virion proteins. *J. Gen. Virol.* 67:119-130.
47. Laude, H., D. Rasschaert, and J. Huet. 1987. Sequence and N-terminal processing of the transmembrane protein E1 of the coronavirus transmissible gastroenteritis virus. *J. Gen. Virol.* 68:1687-1693.
48. Leibowitz, J.L., K.C. Wilhelmsen, and C.W. Bond. 1981. The virus-specific intracellular RNA species of two murine coronaviruses: MHV-A59 and MHV-JHM. *Virology* 114:29-51.
49. Lomniczi, B. and J. Morser. 1981. Polypeptides of infectious bronchitis virus. I. Polypeptides of the virion. *J. Gen. Virol.* 55:155-164.
50. Macnaughton, M.R., H.A. Davies, and M.V. Nermut. 1978. Ribonucleoprotein-like structures from coronavirus particles. *J. Gen. Virol.* 39:545-549.
51. Maniatis, T., A. Jeffrey, and D.G. Kleid. 1975. Nucleotide sequence of the rightward operator of phage lambda. *Proc. Natl. Acad. Sci. U.S.A.* 72:1184-41.
52. McClurkin, A.W. and J.O. Norman. 1966. Studies on transmissible gastroenteritis of swine. II. Selected characteristics of a cytopathogenic virus common to five isolates from transmissible gastroenteritis. *Can. J. Comp. Med. Vet.* 30:190-198.
53. McIntosh, K. 1974. Coronaviruses: A comparative review. *Curr. Topics Microbiol. Immunol.* 63:85-129.
54. Michelson, A.M., and S.H. Orkin. 1982. Characterization of the homopolymer tailing reaction catalyzed by terminal deoxynucleotidyl transferase. *J. Biol. Chem.* 257:14773-14782.
55. Mount, D.W., and B. Conrad. 1986. Improved programs for DNA and protein sequence analysis on the IBM personal computer and other computer systems. *Nucleic Acids Res.* 14:443-454.
56. Niemann, H. and H-D. Klenk. 1981. Coronavirus glycoprotein E1, a new type of viral glycoprotein. *J. Mol. Biol.* 153:993-1010.

57. Niemann, H., B. Boschek, D. Evans, M. Rosing, T. Tamura, and H.D. Klenk. 1982. Post translational glycosylation of coronavirus glycoprotein E1: Inhibition by monensin. *EMBO J.* 1:1499-1504.
58. Niemann, H., R. Geyer, H-D. Klenk, D. Linder, S. Stirm, and M. Wirth. 1984. The carbohydrates of mouse hepatitis virus (MHV) A59: Structures of the O-glycosidically linked oligosaccharides of glycoprotein E1. *EMBO J.* 3:665-670.
59. Pedersen, N.C., I. Ward, and W.L. Mengeling. 1978. Antigenic relationships of the feline infectious peritonitis virus to coronaviruses of other species. *Arch. Virol.* 58:45-53.
60. Rasschaert, D., and H. Laude. 1987. The predicted primary structure of the peplomer protein E2 of the porcine coronavirus transmissible gastroenteritis virus. *J. Gen. Virol.* 68:1883-1890.
61. Reed, K.C., and D.A. Mann. 1985. Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res.* 13:7207-7221.
62. Robb, J.A., and C.W. Bond. 1979. Coronaviridae. vol 14. p. 193-247. In H. Fraenkel-Conrat and R.R. Wagner (ed.), *Comprehensive Virology*. Plenum Press, New York.
63. Rosen, J.M., S.L.C. Woo, J.W. Holder, A.R. Means, and B.W. O'Malley. 1975. Preparation and preliminary characterization of purified ovalbumin messenger RNA from the hen oviduct. *Biochemistry* 14:69-78.
64. Rottier, P., D. Brandenburg, J. Armstrong, B.A.M. van der Zeijst, and G. Warren. 1984. Assembly in vitro of a spanning membrane protein of the endoplasmic reticulum: the E1 glycoprotein of coronavirus mouse hepatitis virus A59. *Proc. Natl. Acad. Sci. U.S.A.* 81:1421-1425.
65. Rottier, P., J. Armstrong, and D.I. Meyer. 1985. Signal recognition particle dependent insertion of coronavirus E1, an intracellular membrane glycoprotein. *J. Biol. Chem.* 260:4648-4652.
66. Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Biochemistry* 16:5463-5467.

67. Sardinia, L.M. 1985. Synthesis and processing of structural and intracellular proteins of two enteric coronaviruses. Ph.D. Thesis. Montana State University. 102 pp. University Microfilms, Ann Arbor, Mich.
68. Schaffer, H.E., and R.R. Sederoff. 1981. Improved estimation of DNA fragment lengths from agarose gels. *Anal. Biochem.* 115:113-122.
69. Schmidt, I., M. Skinner, and S. Siddell. 1987. Nucleotide sequence of the gene encoding the surface projection glycoprotein of coronavirus MHV-JHM. *J. Gen. Virol.* 68:47-56.
70. Schmidt, O.W. and G.E. Kenny. 1982. Polypeptides and function of antigens from human coronaviruses 229E and OC43. *Infect. Immun.* 35:515-522.
71. Shapiro, A.L., E. Viñuela, and J.V. Maizel. 1967. Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochem. Biophys. Res. Comm.* 28:815-820.
72. Siddell, S. 1983. Coronavirus JHM: Coding assignments of subgenomic mRNAs. *J. Gen. Virol.* 64:113-125.
73. Siddell, S.G., R. Anderson, D. Cavanagh, K. Fujiwara, H.D. Klenk, M.R. Macnaughton, M. Pensaert, S.A. Stohman, L. Sturman, and B.A.M. van der Zeijst. 1983. Coronaviridae. *Intervirology* 20:181-189.
74. Siddell, S.G., A. Barthel, and V. ter Meulen. 1981. Coronavirus JHM: A virion-associated protein kinase. *J. Gen. Virol.* 52:235-243.
75. Siddell, S.G., H. Wege, and V. ter Meulen. 1982. The structure and replication of coronaviruses. *Curr. Topics in Microbiol. Immunol.* 99:131-163.
76. Siddell, S., H. Wege, and V. ter Meulen. 1983. The biology of coronaviruses. *J. Gen. Virol.* 64:761-776.
77. Southern, E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.
78. Spaan, W.J.M., H. Delius, M. Skinner, J. Armstrong, P. Rottier, S. Smeekens, B.A.M. van der Zeijst, and S.G. Siddell. 1983. Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO J.* 2:1839-1844.

79. Spaan, W.J.M., H. Delius, M.S. Skinner, J. Armstrong, P. Rottier, S. Smeekens, S.G. Siddell, and B.A.M. van der Zeijst. 1984. Transcription strategy of coronaviruses: Fusion of non-contiguous sequences during mRNA synthesis. *Adv. Exp. Med. Biol.* 173:173-186.
80. Stern, D.F., and S.I.T. Kennedy. 1980. Coronavirus multiplication strategy. I. Identification and characterization of virus specified RNA. *J. Virol.* 34:665-674.
81. Stern, D.F. and B.M. Sefton. 1982. Coronavirus proteins: Biogenesis of avian infectious bronchitis virus virion particles. *J. Virol.* 44:794-803.
82. Stern, D.F., and B.M. Sefton. 1984. Coronavirus multiplication: The locations of genes for the virion proteins on the avian infectious bronchitis virus genome. *J. Virol.* 50:22-29.
83. Stohlman, S.A., and M.M.C. Lai. 1979. Phosphoproteins of murine hepatitis viruses. *J. Virol.* 32:672-675.
84. Storz, J., R. Rott, and G. Kaluza. 1981. Enhancement of plaque formation and cell fusion of an enteropathogenic coronavirus by trypsin treatment. *Infect. Immun.* 31:1214-1222.
85. Studier, F.W. 1973. Analysis of bacteriophage T7 early RNAs and proteins on slab gels. *J. Mol. Biol.* 79:237-248.
86. Sturman, L.S. 1981. The structure and behaviour of coronavirus A59 glycoproteins. *Adv. Exp. Med. Biol.* 142:1-18.
87. Sturman, L.S. and K.V. Holmes. 1977. Characterization of a coronavirus. II. Glycoproteins of the viral envelope: tryptic peptide analysis. *Virology* 77:650-660.
88. Sturman, L.S. and K.V. Holmes. 1983. The molecular biology of coronaviruses. *Adv. Virus Res.* 28:35-112.
89. Sturman, L.S., and K.V. Holmes. 1984. Proteolytic cleavage of peplomeric glycoprotein E2 of MHV yields two 90 K subunits and activates cell fusion. *Adv. Exp. Med. Biol.* 173:25-35.

90. Sturman, L.S., K.V. Holmes, and J. Behnke. 1980. Isolation of coronavirus envelope glycoproteins and interaction with the viral nucleocapsid. *J. Virol.* 33:449-462.
91. Tyrrell, D.A.J., J.D. Almeida, C.H. Cunningham, W.R. Dowdle, M.S. Hofstad, K. McIntosh, M. Tajima, L.Y.A. Zakstelskaya, B.C. Easterday, A. Kapikian, and R.W. Bingham. 1975. *Coronaviridae*. *Intervirology* 5:76-82.
92. Wadey, C.N. and E.G. Westaway. 1981. Structural proteins and glycoproteins of infectious bronchitis virus particles labeled during growth in chick embryo cells. *Intervirology* 15:19-27.
93. Wege, H., K. Nagashima, and V. ter Meulen. 1979. Structural polypeptides of the murine coronavirus JHM. *J. Gen. Virol.* 42:37-47.
94. Wege, H., S. Siddell, M. Sturm, and V. ter Meulen. 1981. Coronavirus JHM: Characterization of intracellular viral RNA. *J. Gen. Virol.* 54:213-217.
95. Wege, H., Siddell, and V. ter Meulen. 1982. The biology and pathogenesis of coronaviruses. *Curr. Topics Microbiol. Immunol.* 99:165-200.
96. Welsh, R.M., C.A. Biron, D.C. Parker, J.F. Bukowski, S. Habu, K. Okumura, M.V. Haspel, and K.V. Holmes. 1983. Regulation and role of natural cell mediated immunity during virus infection, p. 21-42. In F.A. Ennis (ed.), *Human Immunity to Viruses*. Academic Press, New York.
97. Welter, C.J., E. Laun, and H. Head. 1966. Transmissible gastroenteritis of swine: Properties of a vaccine and immunologic aspects in the sow and pig. *J. Am. Vet. Med. Assoc.* 149:1587.
98. Wilhelmsen, K.C., J.L. Leibowitz, C.W. Bond, and J.A. Robb. 1981. The replication of murine coronaviruses in enucleated cells. *Virology* 110:225-230.
99. Wood, E.N. 1979. Transmissible gastroenteritis and epidemic diarrhoea of pigs. *Brit. Vet. J.* 135:305-314.

MONTANA STATE UNIVERSITY LIBRARIES



3 1762 10049646 0

