To the Graduate Council:

I am submitting herewith a dissertation written by Frank Yao-Tsung Tung entitled "Molecular Cloning and Sequence Analysis of the Porcine Transmissible Gastroenteritis Coronavirus Matrix and Peplomer Protein Genes." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy with a major in Life Sciences.

David A. Brian

David A. Brian, Major Professor

We have read this dissertation
and recommend its acceptance:

Jerry P. Weir

Wen-Hsiung Yang

M. S. Riggs

Accepted for the Council:

Vice Provost
and Dean of The Graduate School

MOLECULAR CLONING AND SEQUENCE ANALYSIS OF THE PORCINE TRANSMISSIBLE GASTROENTERITIS CORONAVIRUS MATRIX AND PEPLOMER PROTEIN GENES

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Frank Yao-Tsung Tung

December 1987

This dissertation is dedicated to my parents and my wife, for their love and encouragement, and to my children, Kuang-Tsung and Jack, for their smiles and cooperation.

ACKNOWLEDGMENTS

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

I also appreciate the suggestions, discussions, and encouragement provided by my committee members: Dr. W. S. Riggsby, Dr. J. P. Weir and Dr. W. K. Yang.

I would also like to thank my lab mates, both past and present: Brenda Hogue, Sushma Abraham, Tom Kienzle and Joanne Maki for their discussions and humor.

ABSTRACT

The porcine transmissible gastroenteritis coronavirus (TGEV) is a major cause of acute gastroenteritis in pigs and serves as an important model for coronavirus-induced gastroenteritis in other animals including humans. In order to understand the mechanisms of viral pathogensis and develop a possible recombinant vaccine, the genome of TGEV was molecularly cloned and genes for two of the three major viral structural proteins, the matrix protein and the peplomer protein, were studied in detail.

A copy DNA library was prepared from viral genomic RNA by using an oligodeoxynucleotide primer that is complementary to intergenic regions throughout the 20 kilobase genome. One hundred and thirteen clones were obtained and 64 were found to map within the 3' terminal 12 kilobases. Twelve clones were mapped at the region between 1.5 Kb and 8.5 Kb from 3' end of the TGEV genome and were sequenced in part by the chemical method. Two genes for the matrix and peplomer protein were found within this sequence.

The deduced amino acid sequence for the matrix protein is 262 amino acids long giving the protein a molecular weight of 29,544. The matrix protein is unusual among coronavirus matrix proteins in that it has a hydrophobic amino terminal signal peptide of 17 amino acids in addition to 3 internal hydrophobic domains that apparently also serve as signals for membrane translocation.

The deduced amino acid sequence for the peplomer protein is 1,447 amino acids long giving the protein a molecular weight of 160,345. The peplomer protein has an N-terminal signal peptide, a 19 amino acid

transmembrane anchor region beginning 58 residues from its C-terminus, and 32 potential N-glycosylation sites throughout the molecule.

The peplomer protein gene was subcloned into the vaccinia virus vector in preparation for its use as vector to study its biological functions and as an experimental vaccine.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

REVIEW OF THE PORCINE TRANSMISSIBLE GASTROENTERITIS CORONAVIRUS

Introduction

Coronaviruses are one of twelve families of animal RNA viruses. They are distinguished by their shape and by the molecular biology of their replication (Siddell et al., 1982). Morphologically they are medium-sized (100 nm diameter), enveloped viruses with large bulous-shaped peplomers that project 12-20 nm from the envelope surface, giving them the apperance of having a "corona." It is a distinct appearance and therefore they are given the name coronavirus. Structurally they possess a single-stranded, non-segmented, polyadenylated and capped, infectious RNA genome of 20-28 Kb, and three major structural proteins (Siddell et al., 1982). They are positive-stranded RNA viruses and have the largest genome known of any single-stranded RNA virus. The structural features and the molecular biology of coronavirus replication are discussed in further detail below.

From a medical standpoint alone coronaviruses deserve detailed study. Currently there are 16 members in the coronavirus family (Table 1.1). They cause a wide variety of severe acute diseases in animals and man, most commonly of the respiratory or gastrointestinal tracts (Wege et al, 1982), and these, for the most part, are still not controlled by vaccination. In addition to causing acute infection in cell cultures, they also readily cause persistent infections in cell

Table 1.1 Coronavirus species with corresponding host animal and disease

Coronavirus Species	Host Animal	Disease
Avian infectious bronchitis virus[a]	Chichen	Upper and lower respiratory tract
Mouse enteric coronavirus[b]	Mouse	Gastroenteritis infection
Mouse hepatitis virus encephalitis[a]	Mouse	Hepatitis
Tansmissible gastro-enterititis virus[a]	Pig	Gastroenteritis
Hemagglutinating encephalomyelitis virus[a]	Pig	Encephalomyelitis
Human respiratory coronavirus[a]	Man	Upper respiratory tract infection
Human enteric coronavirus[a]	Man	Gastroenteritis
Rabbit coronavirus[a]	Rabbit	Gastroenteritis Myocarditis
Rat Coronavirus[a]	Rat	Lower respiratory tract infection
Sialodacryoadenitis virus[a]	Rat	Sialodacryoadenitis
Turkey bluecomb disease virus[a]	Turkey	Gastroenteritis
Canine coronavirus[a]	Dog	Gastroenteritis
Bovine coronavirus[a]	Calf	Gastroenteritis
Feline enteric coronavirus[a]	Cat	Gastroenteritis
Feline infectious peritonitis virus[a]	Cat	Infectious peritonitis
Equine coronavirus[a]	Foal	Gastroenteritis

[a]Adapted from Sturman and Holmes, 1983.

[b]Hierholzer et al., 1979.

cultures and in some laboratory animals. Some strains of mouse hepatitis coronavirus (MHV), for example, readily cause persistent infections with a concurrent severe chronic disease in mice (Stohlman and Weiner, 1981; Wege et al., 1982). They may cause persistent infection and chronic disease in humen and other animals.

Since coronaviruses classified as a separate family in 1968, the avian infectious bronchitis virus (IBV) has been most popularly accepted as the prototypic coronavirus (Almeida et al., 1968). The earlist studies on coronavirus morphology and genome structure were done with IBV (Becker et al., 1967). The description of a coronavirus as having an infectious RNA genome and a 3' nested set structure for mRNAs was first accomplished for IBV (Stern and Kennedy, 1980a). Even the first complete coronavirus genome sequence, just recently completed (Boursnell et al., 1987), was done with IBV. IBV was recognized early as a severe respiratory disease in chickens and an early effort was made to develop a vaccine. The vaccine used for IBV was a modified live virus vaccine that has been extremely effective and is used all over the world for controlling avian infectious bronchitis (Cavanagh, et al., 1984).

Unfortunately, this same pattern for vaccine success was not realized for the second most important coronavirus pathogen, the porcine transmissible gastroenteritis coronavirus, TGEV. TGEV is a contagious enteric viral disease of swine. The disease occurs in all age groups, but causes highest mortality (80-100%) in piglets under two weeks of age (Saif et al., 1986). TGEV infections have a world-wide distribution. Serologic surveys indicate that 19-54% of sampled

swine herds in Europe and North America are seropositive for TGEV antibodies (Saif et al., 1986). TGEV viral infection is considered one of the major causes of economic loss for the swine industry, an estimated annunal loss of 200 million dollars in the United States alone. There are several commercial vaccines availabale, but none of them provides sufficient protection for nursing piglets from TGEV viral infection (Saif et al., 1986).

Antigenic Relationship of TGEV to other Coronaviruses

As can be seem in Fig. 1.1, TGEV belongs to one of four major antigenic subgroups of coronaviruses, two of which are exclusively avian coronaviruses and two of which are exclusively mammalian coronaviruses. TGEV is closely related antigenically to the feline infectious peritonitis virus (FIPV) and the canine enteric coronavirus (CCV), but more distantly related antigenically to the human respiratory coronavirus 229E (Pedersen et al., 1978). It is interesting to note that for pigs and humans there exist coronaviruses in both of the two major antigenic subgroups of mammalian coronaviruses. The porcine hemagglutinating encephalomyelitis virus (HEV) and the human respiratory coronavirus strain OC43 (HCV-OC43) are closely related antigenically to each other and to the well-studied mouse hepatitis coronavirus and the bovine enteric coronavirus. TGEV and HEV, therefore, if sharing a common ancestry, are distantly diverged coronaviruses.

The antigenic relationships among TGEV, FIPV, CCV, and HCV-229E were determined by in vitro studies using virus neutralization and

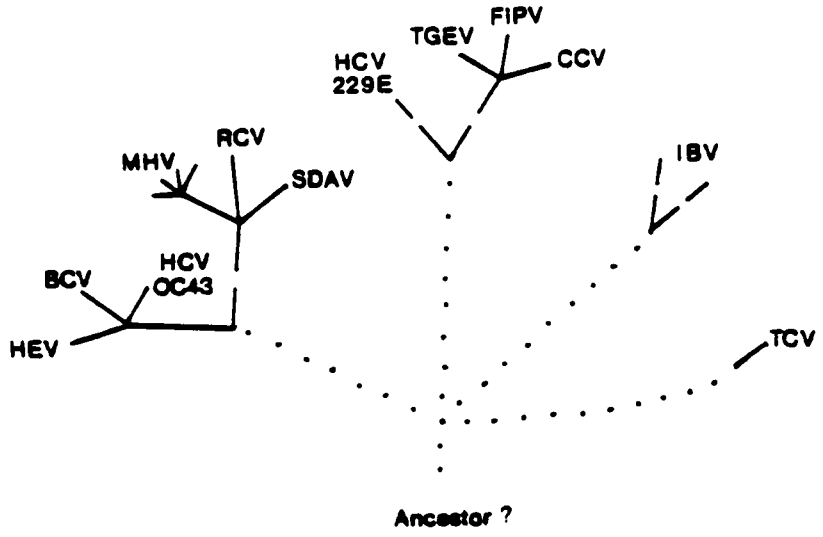


Fig. 1.1 Tentative coronavirus family tree based on antigenic interrelatedness. Data for this figure comes primarily from the review by Wege, et al., 1982. Solid lines indicate evidence for cross neutralizing activity; dashed line indicate immunological evidence (other than neutralization) for crossreactivity; dotted lines indicate possible relatedness with little or no supporting data. Viruses include the porcine hemagglutinating encephalomyetitis virus (HEV), bovine coronavirus (BCV), human coronavirus OC43 (HCV OC43), murine hepatitis virus (MHV), rat coronavirus (RCV), rat sialodacryoadenitis virus (SDAV), human coronavirus 229E (HCV 229E), porcine transmissible gastroenteritis virus (TGEV), feline infectious peritonitis virus (FIPV), canine coronavirus (CCV), avian infectious bronchitis virus (IBV), and turkey bluecomb disease coronavirus (TCV).

immunofluorescence tests (Pedersen et al., 1978). Antiserum to TGEV reacted strongly with FIPV, and vice versa. Antiserum to both of these viruses had barely detectable to weak antigenic cross-reactivity with HCV-229E, and antiserum to HCV-229E had weak to moderate cross-reactivity with FIPV and TGEV. In the case of CCV, however, the reactivity was largely in one direction only, since antiserum to CCV showed no reactivity with TGEV or FIPV. It was concluded from these studies that TGEV, FIPV, HCV 229E and CCV share antigens to varying degrees, but all of these viruses appear to be antigenically unrelated to other coronaviruses (Pedersen et al., 1978). In the neutralization tests, Reynolds et al. (1980) have shown that neutralization titers of the homologous serum-virus relationship were greater than that of heterologous. These results demonstrated that the three viruses TGEV, FIPV and CCV, can be distinguished serologically by employing two-way cross-neutralization test (Reynolds et al., 1980).

The antigenic relationships among TGEV, FIPV, and CCV have been studied at the level of virus structural polypeptides by SDS-PAGE analysis (Horzinek et al., 1982). Feline anti-FIPV antibody recognized three structural proteins of TGEV (peplomer, matrix, and nuceocapsid) in an immunoprecipitation assay. In immunoblotting studies, however, when TGEV, FIPV and CCV viral proteins were reacted with either homologous or heterologous antisera, differences in intensity of binding were observed. Furthermore, the matrix proteins of the three viruses showed different molecular weights suggesting there are some substantial mutational difference among the viruses. Horzinek et al., (1982) concluded from their experiments that FIPV,

TGEV, and CCV shared some common antigenic determinants even though there is evidence of evolutionary divergence.

A study of the antigenic relationship among the many isolates of TGEV from around the world raises a fundamental question regarding the structure of the TGEV surface proteins. The observation is that all TGEV isolates are of one serotype, namely that antiserum to one will neutralize any of the other (Kemeny, 1976). The question is how can neutralizing epitopes be conserved in a virus that apparently would mutate rapidly because of its RNA genome (Holland, 1982)? The conclusion by Kemney that there is only one serotype of TGEV is based on cross-neutralization studies done in vitro. Ten plaque purified isolates of TGEV, comprising eight American isolates, one Japanese isolate and one British isolate were indistinguishable by means of reciprocal plaque reduction neutralization analysis (Kemeny, 1976). In vitro cross-neutralization studies have also demonstrated a two-way antigenic relationship between the attenuated (Purdue 115) and virulent (Miller) strains of TGEV (Saif, unpublished data). There is, however, an indication that differences between strains may be distinguished in the animal. Pigs orally inoculated with attenuated TGEV had significantly enhanced cell-mediated immune response to attenuated TGEV, but not to virulent TGEV (Wan et al., 1986). Further studies are needed to examine the antigenic and structural differerces between the attenuated and virulent strains of TGEV and to map these differences at the genetic level. Such studies can now begin with the development of sensitive monoclonal antibodies and a variety of cDNA probes (Jimenez et al., 1986; Shockley et al., 1987).

Recently, a porcine respiratory, non-enteric virus has been isolated in pigs as well as cell culture (Pensaert and Vergote, 1986). This isolate causes an infection of the lungs and does not replicate in the enteric tract. The infection, either experimental or in the field, resulted in the formation of antibodies which neutralize enteric TGEV. This virus could be a new TGEV-related porcine respiratory coronavirus or TGEV itself which has totally lost its tropism for the enteric tract. This finding also underlines the interesting notion that TGEV neutralizing epitopes are conserved while mutations affecting other important biological properties arise at a seemingly high frequency. It is therefore important to understand the detailed structure of TGEV surface proteins and the nature of the conserved epitopes involved in virus neutralization.

Structure and Function of TGEV Proteins

TGEV contains three major structural proteins (Garwes et al., 1975, 1976; Moreau, 1982) (see Fig 1.2). The peplomer (P or E2) is glycosylated and has a molecular weight of 200 kilodaltons (Garwes et al., 1975; Horzinek et al., 1982; Moreau, 1982). Part of the peplomer protein is embeded in the viral envelope but a majority of it protrudes to form the surface projection. The peplomer protein is the major virus component that elicits neutralizing antibodies (Garwes et al., 1975, 1976). The nucleocapsid protein (N) is phosphorylated and has a molecular weight of 48-50K (Garwes et al., 1975; Horzinek et al., 1982; Moreau, 1982). By sequence analysis it is 43.4K. It is associated with the viral RNA genome. The matrix or transmembrane

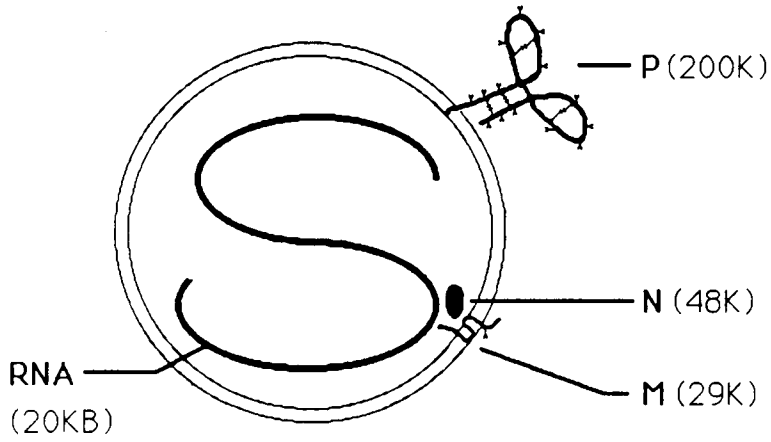


Fig. 1.2 Schematic representation of the porcine transmissible gastroentertitis coronavirus. The data for this model comes from the studies of Brian et al., (1980) and Moreau (Master's Thesis, University of Tennessee, 1982). The four major molecular components are (1) a 200 Kd peplomer glycoprotein, (2) a 48 Kd phosphorylated nucleocapsid protein, (3) a 29 Kd transmembrane (matrix) glycoprotein, and (4) a 6.8 Md (20 Kb) single-stranded, nonsegmented, 3'polyadenylated infectious RNA genome.

protein (M) is also glycosylated and has an electrophoretically determined molecular weight of 25-33K (Horzinek et al., 1982; Moreau, 1982). TGEV has been shown at times to possess two minor polypeptides whose fuctions are unknown (Garwes et al., 1975). They have molecular weights of 105k and 80.5K and they may be proteolytic products of P, analogous to the 90 Kd proteins of MHV (Sturman et al., 1980). Recently, a 17 Kd antigenic polypeptide was identified in TGEV infected cells (Wesly and Woods, 1986). Its function is unknown.

Differences in molecular weights for the P protein have been reported. The molecular weight of the P protein was estimated to be 200 Kd by SDS-PAGE analysis of a plaque purified strain of TGEV (Strain FS772/70) propagated in primary pig kidney cells (Garwes et al., 1975). The molecular weights of P protein identified from purified TGEV (Purdue strain) which was propagated in a PD5 pig kidney cell line were 280 Kd and 240 Kd (Jacobs et al., 1986). The molecular weight of the P protein was slightly different (195Kd) when intracellular viral proteins were analyzed in SDS-PAGE (Jacobs et al., 1986). In our laboratory the molecular weights were determined to be 200 Kd and 170 Kd respectively for virion and intracellular forms (Moreau, 1982). In the presence of tunicamycin, which is an inhibitor of N-glycosylation, the molecular weight of the P protein was 160 kd (Jacobs et al., 1986). Hu et al. (1984) reported that the molecular weight of the P protein of Miller virulent strain of TGEV was 195 Kd and became 145 Kd with the removal of the sugar moiety by endoglycosidase H treatment . Using radioimmunoprecipitation of intracellular viral polypeptides (Purdue strain) with monoclonal

antibodies, Laude et al (1986) identified a 220 Kd protein which was recognized by neutralizing monoclonal antibodies and a 175 Kd protein which was recognized only by nonneutralizing monoclonal antibodies. The 175 Kd protein was characterized as a precusor to the P protein. The differences reported for the molecular weight of virion P protein could be due to differences in the sugar composition and to variations in the treatment of the samples prior to gel electrophoresis.

TGEV Genome Structure and Replication Strategy

The TGEV genomic RNA is a single-stranded, non-segmented, polyadenylated and infectious (therefore positive-stranded) molecule of 6.8×10^6 daltons (Brian et al., 1980). Information on the replication mechanism of coronaviruses has been drived mostly from studies on the mouse hepatitis virus and avian infectious bronchitis virus. During infection, the coronaviral genomic RNA is first copied by a viral polymerase to form one genome-size negative strand RNA (Dennis and Brian, 1982; Lai et al., 1982). This negative strand RNA is then transcribed to form a nested set of seven subgenomic RNAs (probably 9 mRNAS in TGEV), all overlapping in sequence with the 3' end of the genomic RNA (Lai et al.,1981; Spaan et al., 1982; Dennis and Brian, 1982) (see Fig 1.3). The mRNAs contain a common 5' leader sequence of about 70 nucleotides that is probably transcribed from the 3' end of the negative strand RNA template (Lai et al., 1984). Ultraviolet light mapping studies have virtually eliminated the possibility of RNA splicing as a mechanism for the formation of at least the most abundant species (Jacobs et al., 1981). The most

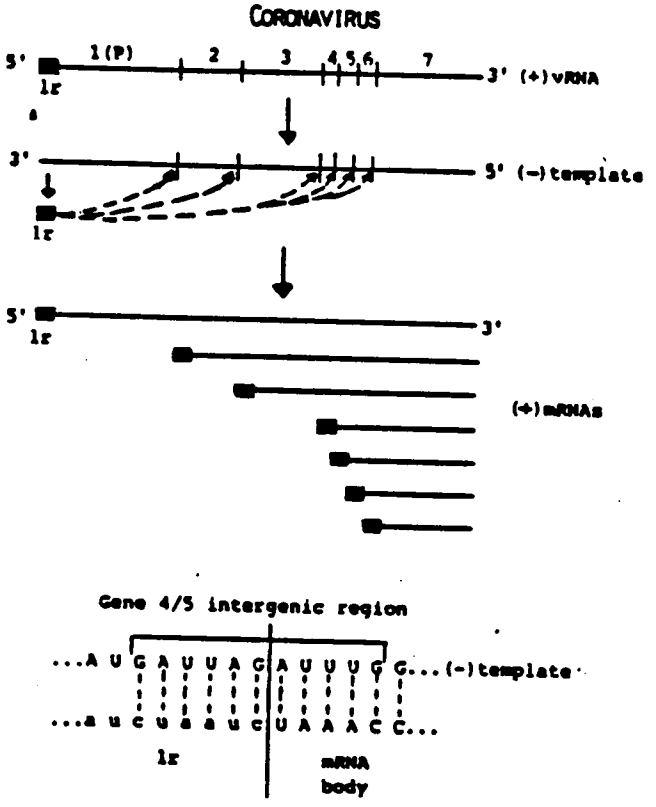


Fig. 1.3 Replication and transcription scheme used by coronaviruses. An early polymerase copies genomic "+" strand to full length "-" strand. A leader (solid box) is synthesized from the 3'end of the negative strand of genomic RNA. This leader sequence dissociates with its polymerase and reassociates at other sites on the negative strand template priming continued synthesis of mRNAs. The sequence shown at the bottom of the figure shows the intergenic sequence found between genes 4 and 5 of the mouse hepatitis coronavirus. This figue comes from Krug (1985).

likely alternative explanation is that the leader RNA acts as a primer at various sites along the negative strand RNA to initiate synthesis of each mRNA (Siddell et al., 1982). Indeed, this mechanism has been well documented for MHV (Lai et al., 1984) and coronavirus leader-primed transcription may be the first example of such a mechanism that is also used in other eukaryotic systems (Krug, 1985).

In MHV, the intergenic regions preceding three of the mRNAs share a common 11 nucleotide sequence, the first 6 nucleotides of which are complementary to the 3' terminus of the leader RNA (Budzilowicz et al., 1985). By itself, this short stretch of hydrogen bonding is probably not sufficient to promote primer binding. It may be that the leader RNA binds to the conserved region as part of a complex containing the viral polymerase.

RNA-dependent RNA polymerase activity in infected cells was first described by Dennis and Brian (Dennis and Brian, 1981, 1982) for TGEV and later by others for MHV (Brayton et al., 1982; Mahy et al., 1983). Further studies demonstrated the temporal appearance of two different peaks of RNA polymerase activity in MHV infected cells (Brayton et al., 1982; Lai et al., 1982b). The RNA product of the early polymerase activity at one hour post infection is a negative strand of RNA complementary to the genomic RNA (Brayton et al., 1984; Lai et al.,1982). Late polymerase activity (5-6 hours post infection) resulted in the production of primarily positive stranded mRNA species (Brayton at al., 1982 and 1984). The clear separation of two polymerase activities, however, has not been confirmed by other laboratories working with MHV, and in fact, only one polymerase

activity is found in a careful study of MHV (Sawicki and Sawicki, 1986).

The proteins responsible for the polymerase activity have not been identified. In the IBV, two large open reading frames (441 Kd and 300 Kd) mapping at the 5' end of genome are potential genes for the polymerase(s) (Boursnell et al., 1987).

The mRNAs of MHV, IBV and TGEV have been translated in vitro (Jacobs et al., 1986; Leibowitz et al., 1982; Rottier et al., 1981; Stern and Kennedy 1980b; Stern and Sefton, 1984; Siddell, 1980a, 1981). These studies demonstrated that each species, although structurally polycistronic, functions monocistronically to encode only one protein. Apparently only the 5' open reading frame of each mRNA is translated. In both MHV and IBV the smallest RNA codes for the nucleocapsid protein. In MHV the next smallest RNA encodes the matrix protein, whereas, it is the third smallest RNA in IBV which encodes the matrix protein (Siddell et al., 1980a; Siddell, 1983; Rottier et al., 1981; Stern at al., 1982; Stern and Sefton 1984). In TGEV, two potential, nonoverlapping genes were identified within the 3' 1663 bases (Kapke and Brian, 1986). From an examination of open reading frames (as shown in Fig 1.4), it was concluded that the first gene encodes the 382 amino acid nucleocapsid protein of 43,426 molecular weight. The second encodes a hypothetical 78 amino acid protein of 9101 molecular weight that is hydrophobic at both ends. The two structural genes mapping immediately to 5'-end of the nucleocapsid gene are the major subjects of this dissertation.

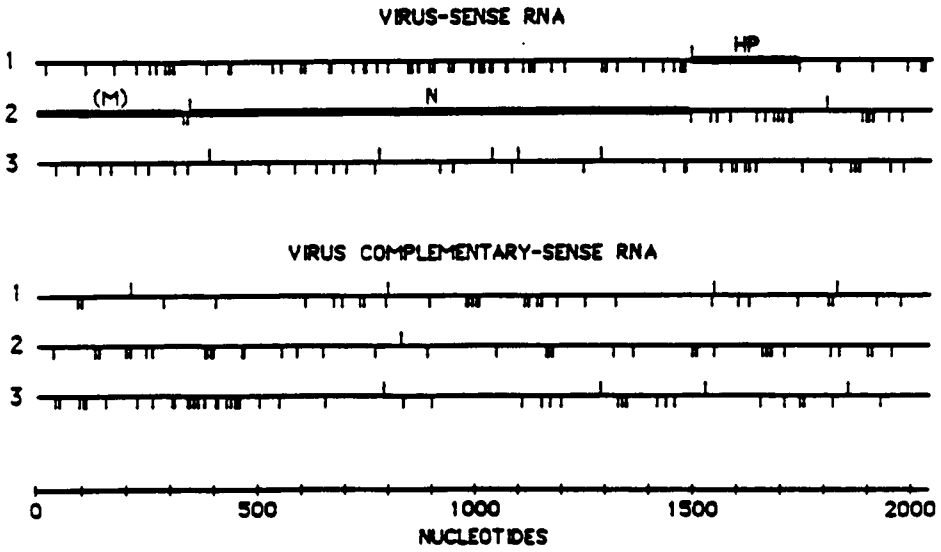


Fig. 1.4 Schematic diagram of possible open reading frames from the 3' end of the TGEV genome. Sequence obtained from clone FG5 was translated as either virus-sense or virus complementary-sense. Vertical bars above the line represent the first methionine codon that could serve as the initiation site for translation. Vertical bars below the line represent termination codons. (M), partial sequence of the matrix protein gene. N, sequence of the nucleocapsid protein gene. HP, sequence of a hypothetical protein gene.

Prospects for Development of Recombinant Vaccines Against TGEV Infection

The immune defense system fulfills two fuctions in connection with infectious disease: (1) Elimination of the infection, and, (2) establishing an immunity that prevents new infection with the same microbe. Immune protection against virus infection has been shown to be mediated by both humoral and cell-mediated immunity. Currently used vaccines in a majority of cases stimulate only effective antibody production, and thus the full power of the immune system has not been exploited. Because of the special importance of cell-mediated immunity in disease recovery and in the removal of persistently infected cells that are the source of virus shedding, it is important that better vaccines be developed.

Recombinant live vaccines offer the potential advantage of stimulating both parts of immune response, perhaps even more effectively than does the original pathogen. Animals vaccinated with live vaccinia recombinants expressing the surface antigen of the human hepatitis B virus (HBsAg) (Smith et al., 1983; Paoletti et al., 1984), the influenza hemagglutin protein (HA) (Smith et al., 1983; Panicali et al., 1983), the herpes simplex virus surface protein gD (Cremer et al., 1985), the vesicular stomatitis virus surface glycoprotein G (Mackett et al., 1985), and the plasmodial circumsporozoite surface protein (Moss, 1985), produce antibodies that react with or neutralize the corresponding virus or protozoa. Furthermore, mice inoculated with vaccinia virus recombinants that express influenza HA or nucleoprotein (N) gene were primed to produce

cytotoxic T cells that recognized these antigens (Yewdell et al., 1985). Thus both kinds of immune response were stimulated with the replicating vaccinia recombinants.

Vaccination with recombinant live viruses may prove to be an especially advantageous approach for the control of TGEV infection. For TGEV two major challenges exist: (1) to generate the production of secretory IgA in colostrum, and, (2) to generate immunity that would cure carrier animals of a persistent infection. It has been observed that natural infection of sows with virulent TGEV stimulates the production of effective secretory antibody (IgA), but intramuscular inoculation of the pathogen generally does not (Bohl and Saif, 1975). Since suckling pigs are protected only as a result of frequent ingestion of colostrum or milk that contains TGEV-neutralizing IgA antibodies, antigenic stimulation of the intestinal tract (or mucosal surface) of the sows is a prerequisite for confering lactogenic immunity to the piglets (Bohl and Saif 1975). Recently, a recombinant TK^- vaccinia virus containing a cloned copy of the influenza hemagglutinin gene was administrated to mice either intranasally or by scarification to stimulate antibody production (Small, 1985). The results demonstrated that vaccination by scarification stimulates the production of serum antibodies but not nasal antibody (local immunity) and gives protection of the lungs but not the nose from challenge with homologous virus. By constrast, intranasal immunization elicits both serum and nasal antibodies and protects both sites. Nasal antibody, however, was detected in mice that were immunized intranasally (Small, 1985). Bennink et al. (1984) have reported that the influenza-HA

vaccinia recombinant vaccine stimulate the production of cytotoxic T lymphocytes and these probably played an essential role in the recovery of mice vaccinated by scarification and challenged by intranasal inoculation. Because both secretory antibody and cytotoxic T cells can be induced by a replicating recombinant virus vaccine, both goals for controlling TGEV infections appear to be attainable.

Before developing a subunit vaccine for any viral disease it is crucial to identify the viral antigens that are essential for induction of either one or both kinds of immunity. In most cases this will probably be a surface protein that is involved in the step of virus adsorption to receptors on the target cells. Furthermore, such an antigen itself must be made by the recombinant vaccine in a form that is able to induce neutralizing antibody or cell-mediated immunity in animals and to confer protection.

It has been shown that a TGEV subviral component consisting mainly of peplomer protein could elicit a neutralizing immune response when injected into pigs (Garwes et al., 1979). This observation suggested that the peplomer protein is important for virus infectivity and may be important in stimulating protective immunity during recovery. The peplomer gene has been cloned and expressed in E. coli in the form of fusion proteins (Hu, et al., 1986). These proteins were isolated and used to immunize mice. All animals developed antibodies that cross rected with the TGEV virion but failed to neutralize the virus.

It also has been shown that peplomer protein purfied from detergent-dissociated TGE virus failed to elicit neutralization

(Jimenez et al., 1986). Both of these experiments suggest that the native comformation of peplomer protein is critical to elicit neutralizing antibody. Recently, preliminary experiments were done with a recombinant vaccinia virus carrying an undefined portion of the TGEV peplomer protein gene (Hu et al., 1986). These studies demonstrated that mice immunized by scarification did develop low titers (50 to 160) of neutralizing antibody, but no studies of disease prevention or clearing of persistent infection could be done since TGEV has not been shown to cause disease in mice. Because there was no primary sequence data, it is not clear what portion of the peplomer gene was cloned into vaccinia. Although many basic studies are still needed to investigate the mechanisms of immune recognition and antigen presentation in the stimulation of many kinds of protective immunity, the data mentioned above provide us with an a insight on how to develop an approach for the control of TGEV infection by vaccination.

Perhaps other expression vectors (bacterial or viral) would be more appropriate than vaccinia for expression of the peplomer or other TGEV structural protein genes in pigs. For example, pseudorabies virus, a pantropic virus in pigs, has been developed by the Upjohn company as an expression vector (Leonard Post, personal communication) and we are currently collaborating with them to express the TGEV M and P genes in the pseudorabies virus vector and use it as an experimental vaccine. The molecular cloning and characterization of the cloned genes for the two most important TGEV structural proteins for vaccine development is the major subject of this dissertation. This work should lay the foundation for studying the role of the M and P

proteins in both humoral and cell-mediated immunity against infection by TGEV.

CHAPTER 2

PREPARATION OF A COPY DNA LIBRARY FOR A MAJORITY OF THE PORCINE TRANSMISSIBLE GASTROENTERITIS CORONAVIRUS GENOME AND PHYSICAL MAPPING OF THE CLONES WITH RESPECT TO THE VIRUS GENOME

Introduction

The genome of the porcine transmissible gastroenteritis virus has been shown to be a single-stranded, non-segmented, polyadenylated, infectious RNA molecule of 6.8×10^6 Kd or approximately 20-28 Kb (Brian et al., 1980). A single-stranded RNA molecule of this size, the largest single-stranded RNA genome known for any animal virus, will undoubtedly possess much secondary structure and may therefore present problem when attempts to clone cDNAs representative of the entire genome are made.

Since the first reports on the molecular cloning of cDNAs (Efstratiadis et al., 1976; Rougeon and Mach, 1976), this technology has become more and more refined. Recent modifications have made it possible to clone and express RNAs that are available only in small amounts and that may have much secondary structure. Gubler and Hoffman (1983) developed a simple method for generating cDNA libraries from submicrogram quantities of mRNA. It combines classical first-strand synthesis with RNase H/DNA polymerase I-mediated second-strand synthsis. The method of Gubler and Hoffman was used successfully in our laboratory for generating cDNA clones that represent the 3'end of the TGEV genome for the first 2000 bases (Kapke and Brian, 1986). For this cloning procedure, oligo(dT)12-18 was used as a primer to begin

first-strand synthesis from the 3'poly(A) tail. To generate a cDNA library from the 20 Kb TGEV genome it is required that first strand cDNA synthesis be primed from various sites throughout the molecule. For this two approaches could be used: (1) random priming, or, (2) priming with TGEV-specific oligodeoxynucleotides that are complementary to sites throughtout the genome. For the studies described below we have chosen the second approach. From the studies of Kapke and Brian (1986), a sequence was found between the nucleocapsid protein gene and the matrix protein gene that was presumed to be a conserved intergenic sequence similar to those described for the mouse hepatitis coronavirus (Budzilowicz et al., 1985) and the avian infectious bronchitis virus (Brown and Binns, 1984). With the assumption that the mechanism of leader-primed mRNA synthesis occurs in TGEV replication as has been shown to occur in mouse hepatitis coronavirus replication (Lai et al., 1981), a conserved sequence (intergenic sequence) should serve to locate the regions between several if not all genes on the genome. To produce a cDNA library of the TGEV genome we prepared a synthetic oligodeoxy-nucleotide that was complementary to the intergenic regions between the N and M genes and used this to prime first strand cDNA synthesis from genomic RNA. By this method, a cDNA library for a large part of TGEV genome was generated and cloned sequences were mapped by a cross-hybridization method.

Materials and Methods

Cells and Virus

The Purdue strain of TGEV was plaque purified and stocks were prepared as previously described (Brian et al., 1980). To grow virus, swine testicle (ST) cell were infected with virus at a multiplicity of approxmately five PFU/cell. After one hour adsorption, cells were rinsed with Earle's balanced salt solution and refed with Dulbecco's medium containing 1% fetal calf serum (Sterile Systems, Inc.). At 18 hours post infection, supernatant fluids were clarified by centrifugation at 2,000xg for 10 minutes. Virus was concentrated from cell culture supernatant fluids by sedimentation onto a 3 ml cusion of 60% sucrose (wt/wt) in a 38 ml tube. For this, virus was sedimented at 90,000xg (26,000 rpm) in an SW27 rotor for 2 hours, and virus from several mls was concentrated by doing several successive spins in the same tubes. For this, 32 ml of spent fluid was removed and replaced with new cell culture supernatant fluid. Typically, virus from 1 liter of supernatant fluid could be concentrated to 20 ml by 5 successive spins using the 6-bucket SW 27 rotor (Beckman). Sixty percent sucrose (wt/wt) was made up in TMEN buffer (0.05 M Tris-maleate, pH6.0, 0.1M NaCl, 0.001M EDTA). The concentrated virus band along with the 60% cusion was mixed with an equal volume of TMEN to make a final sucrose concentration of less than 20%. Diluted virus was layered onto an 18 ml continous gradient of 20 to 60% (wt/wt) sucrose gradient in TMEN and centrifuged at 24,000 rpm for 4.5 hours at $4^{o}C$. The sucrose gradient was fractioned and the fractions containing virus (buoyant density 1.18-1.20 g/cm) were collected after

measuring the buoyant density with a refractometer (Kapke and Brian 1986). Viral fractions were pooled, diluted with TMEN to yield a final sucrose concentration less than 20%, and virus was pelleted at 49 K rpm on SW 50.1 rotor. To extract genomic RNA, the virus pellet obtained from up to one liter of supernatant fluids was resuspended in 0.5 ml of TNE (10 mMTris-HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA) containing 0.5% sodium dodecyl sulfate and 0.1 mg/ml of protease K. The mixture was incubated at 20^{o}C for 30 minutes followed by two consecutive extractions with a mixture of phenol, chloroform, isoamylalcohol (50:24:1). The RNA in the aqueous phase was ethanol precipitated at -80^{o}C after adding 0.1 volume 3M sodium acetate and three volumes of absolute ethanol. Virus from one liter of supernatant fluid (collected from 30,000 cm^{2} of infected cells, or approximately 40 flasks) yielded approximately 10 micrograms of genomic RNA.

Copy DNA Synthesis

A synthetic oligodeoxynucleotide of the sequence 5' TTAGAAGTTTAGTTA3' was synthesized by the phosphoramadite method (C. Naeve, St. Jude Hospital, Memphis), deprotected, and purified by electrophoresis on a 20% polyacrylamide gel in the presence of urea according the method of C. Naeve (See Appendix). The intergenic primer was used to prepare a cDNA library. In order to fill a gap in the cDNA sequence that existed at the 5'end of the peplomer (P) gene, a primer specific for this region of the genome was made and cDNA cloning was repeated. The primer near the 5' end of the P gene had the sequence 5' GAGCTGTGCCATTATTAGTGT3'.

cDNA cloning was accomplished by the method of Gubler and Hoffman (1983). First strand synthesis was carried out in a reaction volume of 50 microliters containing 50 mM Tris.HCl, pH 8.3, 10 mM DTT, 4 mM Na.pyrophosphate, 1.5 mM dNTPs, 100 mM KCl, RNasin 46U/100 ul (Promega Biotech), reverse transcriptase 100 U/100 microliter (Seikagaku America), 5 micrograms of primer. Approximately six micrograms of genomic RNA was heated at 90^{o}C for 90 seconds with the primer, then cooled on ice before adding first strand synthesis buffer. Synthetic oligodeoxynucleotide was used in a primer to genomic RNA molar ratio of 1000:1 for first strand cDNA synthesis. The first strand reaction mixture was completed, incubated at 42^{o}C for one hour and the reaction stopped by phenol extraction. The nucleic acids were ethanol precipitated after adding 0.1 volume of 3 M Na.acetate (pH 8.0).

Second strand synthesis was carried out in a reaction volume of 100 microliters that contained 20 mM Tris.HCl, pH 7.5, 5 mM MgCl2, 10 mM (NH4)2SO4, 100 mM KCl, 0.15 mM β-NAD, 50 micrograms/ml BSA, 40 uM dNTPs, 2.5 U/100 ul *E. coli* DNA ligase (New England Biolabs), 1 U/100 ul *E. coli* RNase H (Pharmacia) and 40 U/100 ul *E. coli* polymerase I (New England Biolabs). The first strand cDNA product was dissolved in water before adding the second strand synthesis buffer. The mixture was incubated at 12^{o}C for one hour then 22^{o}C for another hour. The reaction was stopped by adding 4 microliters of 0.5 M EDTA and the reaction product was phenol extracted, passed over a Sephadex G 50 spun column, and the double stranded cDNA was ethanol precipitated out of 2 M NH4acetate.

Homopolymer Tailing

Homopolyer tailing was carried out using the conditions described by Roychoudhury and Wu (1980). Optimal tail length for transformation of dG-dC homopolymer linked chimeras is 15 to 20 bases (Peacock et al., 1981). The 100 microliter reaction mixture contains the following: 140 mM K cacodylate pH6.9, 1 mM CoCl2, 1mM dCTP, 0.1 mM dithiothreitol, 22 U/100 ul terminal deoxynucleotidyl transferase (Pharmacia). An excess of dCTP was used so that it would not to be a rate limiting factor. Before adding double stranded cDNA, the mixture was heated at 37^{o}C for 5 minutes. The reaction time for tailing was determined by a pilot experiment using PstI-cut pBR322 DNA and $\alpha[^{32}P]$ dCTP. The incorporation curve of dCTP into one pmole of 3'ends of PstI-cut pBR322 is illustrated in Fig. 2.1. Radioactive alpha 32P (>3000 Ci/mmole) was used to monitor incorporation. Based on the radioactivity incorported approximately 10 bases were added for every one minute of reaction time. The incubation time for tailing was chosen to be two minutes since we desired a tail length of 20 bases and assumed we had 1 pmole 3'ends of double stranded cDNA. The reaction was done without radioactive tracer and was terminated by adding 4 microliters of 0.5 M EDTA and heating for 5 minutes at 68^{o}C. Double-stranded cDNA was ethanol precipitated and dissolved in 50 microliters of water.

Annealing and Transformation

The annealing reaction was done using the following conditions: ds cDNA was added to PstI site of the dG tailed pUC9 vector (Pharmacia) in 30 microliters of 10 mM Tris.HCl, pH 7.5, 1 mM EDTA,

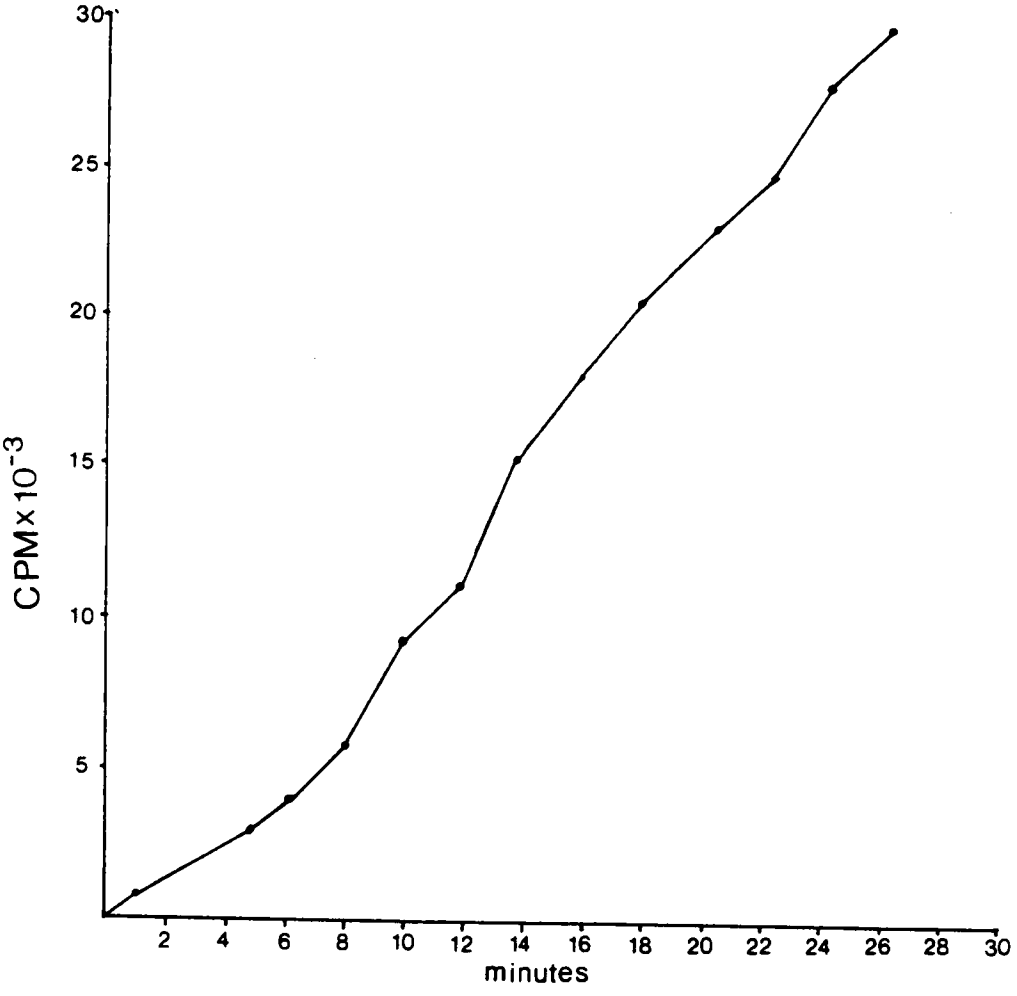


Fig. 2.1 Time course for homopolyer tailing of Pst1-cut pBR322. At each of the time points indicated an aliquot was removed and reaction was stopped by adding 1 microliter of 0.5 M EDTA. The sample were then spotted onto a DE 81 paper disc and washed to remove the unincorporated radioactive dCTP (Maniatis et al., 1982). Radioactivity was quantitated by scintillation counting.

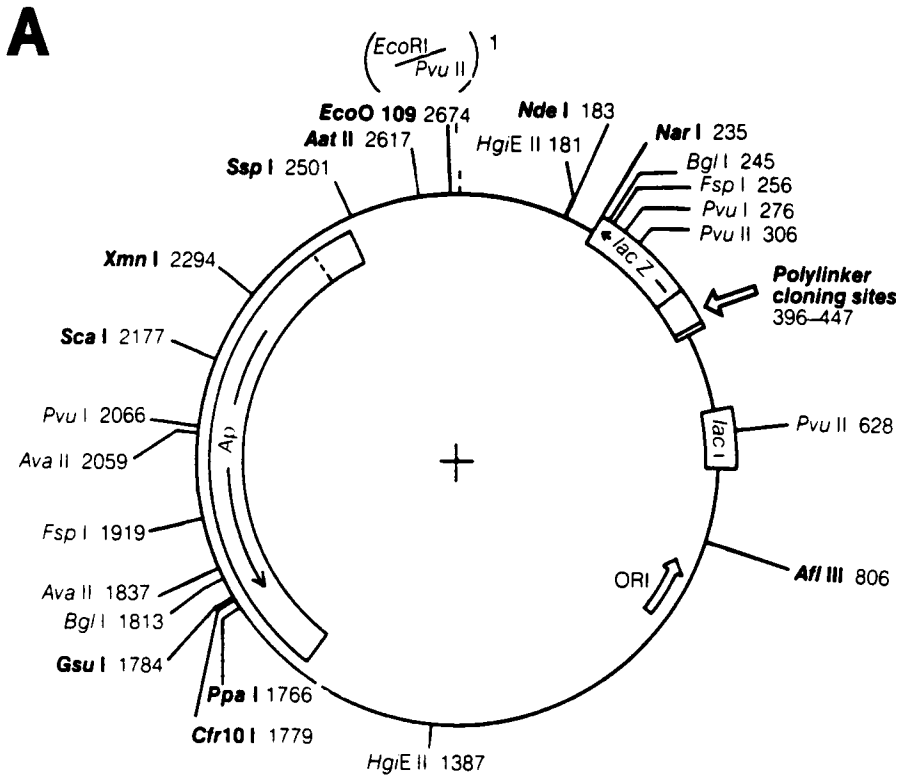
150 mM NaCl, and the mixture was incubated at 58°C for two hours (Peacock et al., 1981). The final concentration used for annealing was 0.5 micrograms/ml or less.

Because the concentration of ds cDNA was unknown, the optimal annealing conditions (insert/cDNA molar ratio) for homopolymer-tailed molecules to be used for transformation needed to be determined empirically. To find the maximum number of clones obtainable per weight input of cDNA, we titrated several annealing ratios. The results from the titration with different amounts of cDNA in the annealing reaction is shown in Table 2.1. The insert:cDNA molar ratio yielding the maximal transformation rate was used for the scale-up reaction. Competent cells were prepared by the methods of Hanahan (1983). This technique routinely yielded 10^7 to 10^8 transformants per microgram of supercoiled pUC9 plasmid DNA. The pUC9 vector used in this experiment contains the Lac Z gene along with the E. coli lac promoter and operator. When induced by IPTG, the Lac Z gene produces functional β-galactosidase alpha peptide. The expressing phenotype can be selected by blue color when transformed cells grow on X-gal plates in the presence of the IPTG inducer (Messing and Vieira 1982). Fig. 2.2 illustrates the important features of the pUC9 cloning vector. Since the multiple cloning site is located within the coding region of the Lac Z gene, when homopolyer tailed cDNA is inserted at PstI site the Lac Z gene become inactivated. The transformed cells carrying insert cDNA can be easily selected on X-gal plates as white colonies.

Table 2.1 Determination of the best insert: vector ratio for maximum number of transformants[a]

Experiment	dG-tailed pUC9 (ng)	dC-tailed cDNA (ul)	Number of White Colonies
1	12.5	1	5
2	12.5	2	3
3	12.5	3	50
4	12.5	0	0

[a]JM83 cells were transformed by the method of Hanahan (Hanahan, 1983).



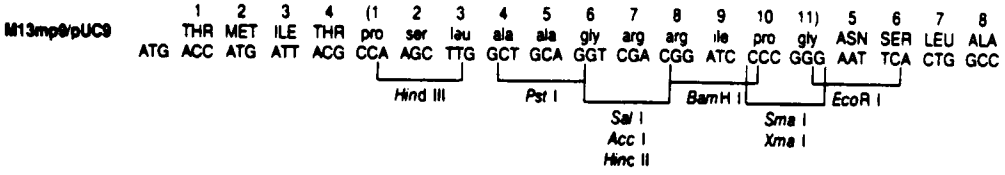


Fig. 2.2 Model of pUC9 cloning vector. A. Restriction map of cloning vector pUC9 (New England Biolabs, Catalog 1987). B. Multiple cloning region (polylinker cloning sites) in pUC9. cDNA was inserted at Pst I site in the multiple cloning region. The Hind III and Sal I sites were commonly used for end labeling insert DNA to be used for DNA sequencing by the Maxam and Gilbert method (see Chapters 3 and 4).

Identification of Virus-Specific Clones

To identify colonies that contained TGEV-specific inserts, random primed cDNA of TGEV genomic RNA was used as the probe for colony hybridization. Fragmented calf thymus DNA that was to be used as random primers was prepared using the technique described by Maniatis (1982). Reaction conditions for probe synthesis were the same as those used for first strand synthesis, except that the cold dNTP concentration were 15 mM, and the alpha 32P dCTP (>3000 Ci/mM, ICN) concentration was 2 uM. The reaction was incubated at 42^{o}C for two hours and stopped by adding 10 microliter of 0.5 M EDTA. The nucleic acids were extracted with phenol-chloroform-isoamyl alcohol. After ethanol precipitation the nucleic acids were treated with 150 mM NaOH at 80^{o}C for five minutes to destroy the RNA.

For colony hybridization, colonies grown on selective agar plates and were replica-plated onto nitrocellulose paper (BA85 Schleicher and Schuell). The bacterial colonies were lysed and nitocellulose paper was prepared for colony hybridization using the technique described in Maniatis (1982). The air-dried nitrocellulose was placed, colony side up, on a piece of Whatman 3MM paper saturated with 10% SDS for three minutes. The nitrocellulose was then placed on Whatman 3MM paper saturated with denaturation solution (0.5 M NaOH and 1.5 M NaCl) for five minutes. The colony blot with lysed colonies was then placed, colony side up, on a piece of Whatman 3MM paper to dry for 5 minutes. The dried colony blot was neutralized on 3MM paper saturated with 1.5 M NaCl and 0.5M Tris.HCl, pH 8.0, for 5 minutes. After neutralization the colony blot was transfered twice onto 3 MM paper

saturated with 2xSSC for 5 minutes, air-dried, sandwiched between two sheets of 3 MM paper, and baked in 70^{o}C vacuum oven for one hour. Colony blots were wetted with 6xSSC before prehybridization, and were prehybridized at 65^{o}C for at least one hour in prehybridization solution containing 6xSSC, 0.1% SDS, 5xDenhardt's solution and 100 microgram/ml denatured salmon sperm DNA. Following prehybridization the solution was removed by pipette leaving 50 microliter/cm2 in the bag. Denatured radiolabeled probe was added to the bag and sealed. Hybridization was done at 65^{o}C with rocking for at least 12 hours. The blot was washed initially in 2xSSC and 0.5% SDS for five minutes, and subsequent washes were carried out using varying dilutions of SSC depending on the stringency of washing required. After washing, the blot was dried and autoradiographed. Fig. 2.3 shows the results of a colony blot that was probed with TGEV random-primed cDNA.

Isolation and Size Screening of Clone Inserts

Colonies identified by the TGEV specific probe were isolated and grown in 3 ml L-B medium overnight in the presence of ampicillin. The plasmid was purified by the alkali lysis method (Maniatis, 1982) and analyzed by agarose gel electrophoresis. The sizes of insert cDNAs were estimated from the migration rate of supercoiled plasmids. Migration patterns were compared to the migration patterns of plasmids of known size including the pUC9 vector. Fig. 2.4 shows the results of agarose gel electrophoresis for some of the clones.

Fig. 2.3 Autoradiograph of colony hybridization using random-primed TGEV cDNA probe. All colonies were white colonies selected from several X-gal plates and placed in a grid pattern. Colonies were lysed and probed with TGEV-specific cDNA probe which was synthesized from size-selected genomic RNA.

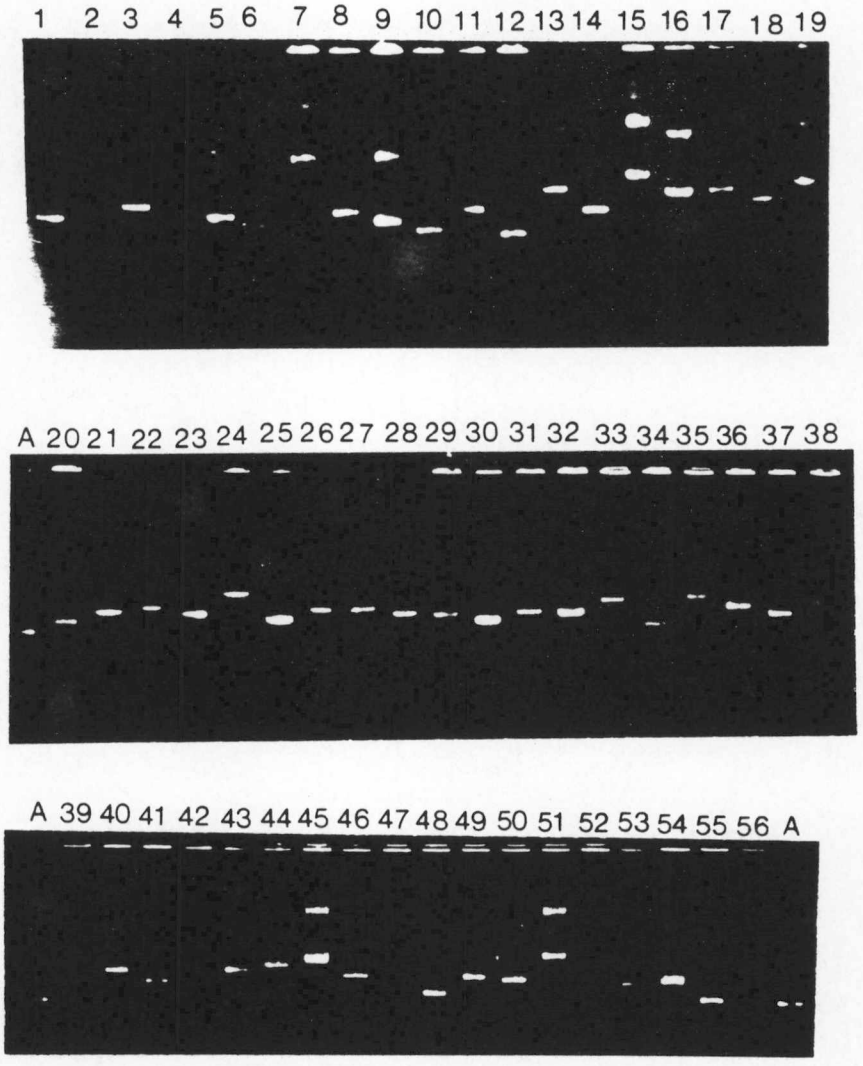


Fig. 2.4 Agarose gel electrophoresis of TGEV-specific clones. Mini-prep plasmid DNA isolated from colonies yielding a strong signal by hybridization to 32P-labeled TGEV-specific cDNA (Fig 2.3) was electrophoresed on a 1% agarose gel. Lane A is pUC9 DNA. The lowest band in each lane represents the supercoiled form. Clones were named by the number indicated (FT1-FT56).

Physical Mapping of Clones With Respect to the Virus Genome

A cross hybridization technique was used to map the clones relative to the virus genome and this technique was done as follows: (1) plasmid DNA isolated from a 3 ml overnight miniprep culture was dissolved in 50 ul water and 1 ul of each prep was spotted in a grid pattern on 150 mm diameter circular nitrocellulose sheets. The nitrocellulose used was printed with a 5mm grid pattern (Schleicher and Schuell). Several sheets were made for continuous mapping. (2) Nitrocellulose sheets spotted with plasmid DNA were treated with alkali, neutralized, and baked using the same procedures as that for colony hybridization (described above) except that the 10% SDS lysis step was ommited. (3) A given filter was hybridized with a specific insert DNA fragment labeled with 32P by nick translation. Insert fragments were removed from supercoiled plasmid with PstI, isolated by electrophoresis onto NA45 paper and eluted with a hight salt buffer (1M NaCl, 10 mM EDTA). In some cases when an internal restriction enzyme cutting site was found and the orientation determined, only a 5' fragment of the insert was isolated and used for identifying additional clones. The procedure for using the NA45 paper was that recommended by Schleicher and Schuell except that the paper was only presoaked in water and not alkali treated. Filter hybridization followed the procedure of Maniatis described above.

Results and Discussion

From approximately 3000 white colonies obtained using the intergenic primer, 113 clones were identified by the TGEV random-

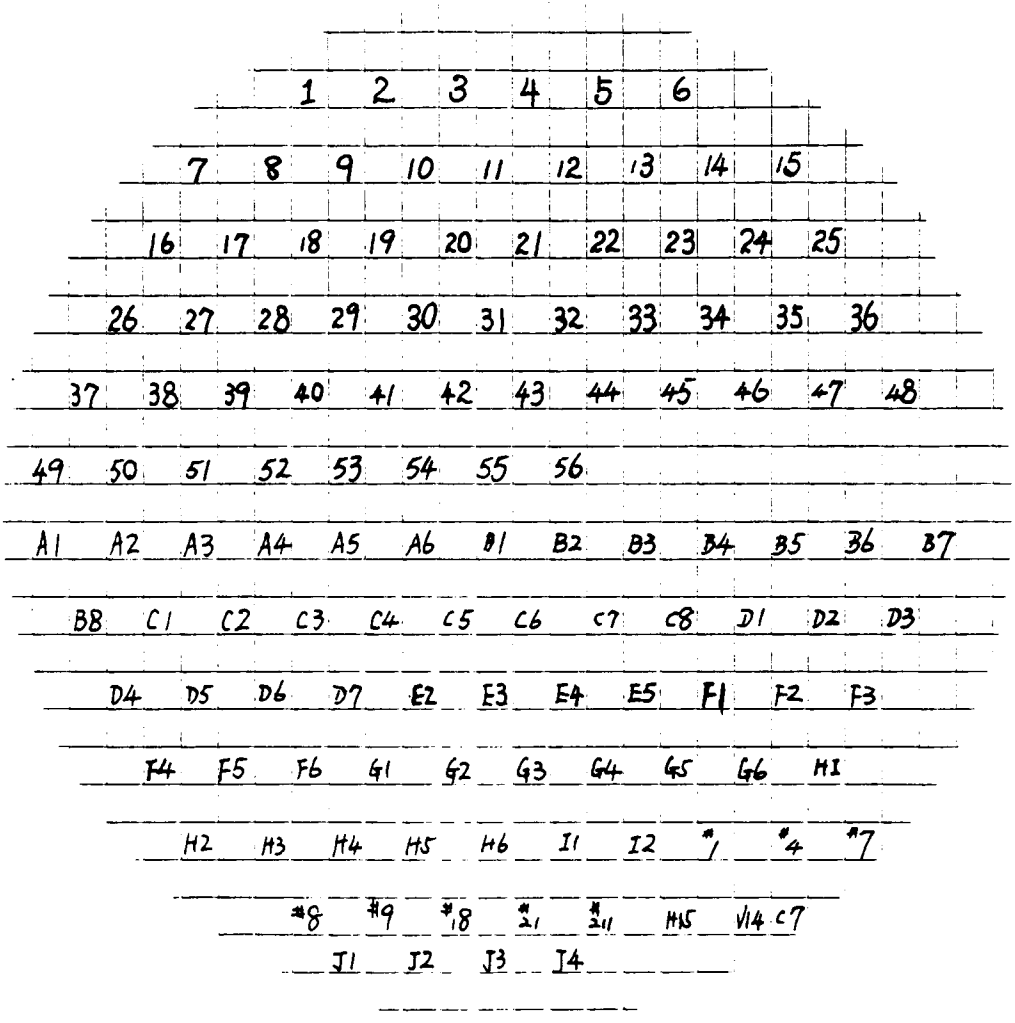


Fig. 2.5 Spotting pattern for all TGEV cDNA clones prepared with intergenic primer. Note: Clones 1-56, A1-A6, B1-B8, C1-C8, D1-D8, E2-E5, F1-F6, I1, I2, and J1-J4 use the prefix FT in their full names. Clones #1, #4, #7, #8, #9, #18, #21 and #211 were made by priming with a specific primer made from a sequence at the 5'end of clone FG5. Clones C7, H15, and V14 were made by P. Kapke using an oligo(dT) 12-18 primer, and were known not to map within the 3' 2 Kb of the genome.

Table 2.2. Summary of TGEV clones prepared with the intergenic primer[a]

Clone Name	Insert Size (Kb)[b]	Orientation	Clone Name	Insert Size (Kb)[b]	Orientation	Clone Name	Insert Size (Kb)[b]	Orientation
FT1	(<0.8)	ND	FT43	1.2	B	FTD7	1.5	A
FT2	(0.8)	ND	FT44	1.4	A	FTE1[d]	?	ND
FT3	(0.8)	ND	FT45	(1.5)	ND	FTE2[d]	0.6	A
FT4	(<0.8)	ND	FT46	(0.8)	ND	FTE3	?	ND
FT5	(<0.8)	ND	FT47	2.4	ND	FTE4	<0.8	ND
FT6	(0.3)	ND	FT48	(0.5)	ND	FTE5	(1.2)	ND
FT7	3.2	B	FT49	(0.9)	ND	FTF1	0.1	ND
FT8	(<0.8)	ND	FT50	0.8	A	FTF2	?	ND
FT9	(0.4)	ND	FT51	(2.0)	ND	FTF3	(2.0)	ND
FT10	(0.2)	ND	FT52	(2.0)	ND	FTF4[d]	(0.6)	ND
FT11	(1.0)	ND	FT53	(<0.8)	ND	FTF5[d]	0.6	A
FT12	(<0.8)	ND	FT54	(<0.8)	ND	FTF6	(0.6)	ND
FT13	(1.3)	ND	FT55	0.2	ND	FTG1	0.1	ND
FT14	(1.0)	ND	FT56	(0.8)	ND	FTG2	?	ND
FT15	(2.0)	ND	FTA1	2.0	ND	FTG3	0.1	ND
FT16	(1.0)	ND	FTA2	(1.4)	ND	FTG4	(0.2)	ND
FT17	(1.0)	ND	FTA3	(0.2)	ND	FTG5	0.1	ND
FT18	(0.8)	ND	FTA4	(0.6)	ND	FTG6	?	ND
FT19	1.5	B	FTA5	0.1	ND	FTH1	(1.2)	ND
FT20	(0.6)	ND	FTA6	(0.2)	ND	FTH2	?	ND
FT21	(1.2)	ND	FTB1	(1.2)	ND	FTH3	?	ND
FT22	(1.1)	ND	FTB2	2.0	ND	FTH4	(0.5)	ND
FT23	(1.0)	ND	FTB3	(0.8)	ND	FTH5	(0.5)	ND
FT24	(1.4)	A	FTB4	(0.8)	ND	FTH6	0.1	ND
FT25	(<0.8)	ND	FTB5	?	ND	FTI1	0.1	ND
FT26	(1.0)	ND	FTB6	?	ND	FTI2	?	ND
FT27	(1.0)	ND	FTB7	0.9	ND	FTJ1	(1.4)	ND
FT28	(0.8)	ND	FTB8	?	ND	FTJ2	(2.0)	ND
FT29	(<0.8)	ND	FTC1	?	ND	FTJ3	(1.0)	ND
FT30	(<0.8)	ND	FTC2	(0.6)	ND	FTJ4	0.1	ND
FT31	(<0.8)	ND	FTC3[d]	2.0	ND	(#1)	?	ND
FT32	(<0.8)	ND	FTC4[d]	0.4	A	(#4)	1.2	ND
FT33	(1.3)	ND	FTC5	0.7	ND	(#7)	1.5	ND
FT34	(<0.8)	ND	FTC6	2.0	ND	(#8)	?	ND
FT35	1.3	A	FTC7	2.0	ND	(#9)	?	ND

Table 2.2 (continued)

Clone Name	Insert Size (Kb)[b]	Orien-tation	Clone Name	Insert Size (Kb)[b]	Orien-tation	Clone Name	Insert Size (Kb)[b]	Orien-tation
FT36	0.8	A	FTC8	?	ND	(#18)	0.7	ND
FT37	0.7	ND	FTD1	(0.8)	ND	(#21)	0.7	ND
FT38	(0.8)	ND	FTD2	(0.5)	ND	(#211)	0.8	ND
FT39	1.4	ND	FTD3	2.0	ND	(H15)	1.8	ND
FT40	(0.8)	ND	FTD4	?	ND	(V14)	1.5	ND
FT41	(<0.8)	ND	FTD5	(2.0)	ND	(C7)	1.3	ND
FT42	(1.4)	ND	FTD6	?	ND			

[a]Order of listing is as they appear on the spotted filter (Fig. 2.7), left to right. Note, clones (#1), (#4), (#7), (#8), (#9), #18), (#21) and (#211) were made by priming with a specific primer made from a sequence at the 5′ end of FG5. Clones (C7), (H15), and (V14) were unmapped clones made by P Kapke using an oligo(dT)12-18 primer. In all cases, cDNA was cloned into G tailed pUC9 vector using host JM83.

[b]Insert size was estimated from electrophoresis of the Pst1 fragment in agarose gels. Insert size in parantheses was estimated from electrophoresis of supercoiled plasmid in agarose gels.

[c]"A" orientation has the 3′ end of the virus-sense strand near the Hind III site on the vector. "B" orientation is the oposite of A.

[d]FTC4, FTE2, and FTF5 were called C4, E2 and F5 in the early matrix gene sequence paper.

Table 2.3. Mapping of TGEV clones prepared with the intergenic primer[a]

Clone Name[b]	FG5	FTF5	FT44	FT39	FT7	FT50	(H15)	FT47
(FG5)	+	+						
FTC4	+	+						
FTE2	+	+	+					
FTF5	+	+	+					
FT4		+						
FT12		+						
FT27		+						
(#18)		+						
FT8		+	+					
FT14		+	+					
FT34		+	+					
FT35		+	+					
FT36		+	+					
FT43		+	+					
FT44		+	+					
FT53		+	+					
FT54		+	+					
FT41			+					
FT5			+	+				
FT25			+	+				
FT26			+	+				
FT30			+	+				
FT32			+	+				
FT38			+	+				
FT40			+	+				
FT42			+	+				
FT44			+	+				
FT45			+	+				
FT49			+	+				
FTB4			+	+				
FTE4			+	+				
FT39			+	+	+			
FT7				+	+			
FT19				+	+			
FT1					+			
FT13					+			
FT24					+			
FT28					+			
FT29					+			
FT31					+			
FTB3					+			
FTB7					+			
FTC5					+			
FTD7					+			
FT16						+		
(#211)						+		
(H15)						+	+	
FT17						+	+	
FT22						+	+	
(V14)						+	+	
FT50						+	+	
FTD3							+	
FTC7							+	
FT20[c]								+
FT21[c]								+
FT37[c]								+
FT47[c]								+
FTC2[c]								+

Table 2.3 (continued)

[a]Clones not listed remain unmapped. Purified inserts were used as probes.

[b]See Table 2.2 for a complete listing of clones and the significance of their designation.

[c]Note: These clones map somewhere between 12 Kb from the 3′ terminus and the 5′ terminus. We know not where.

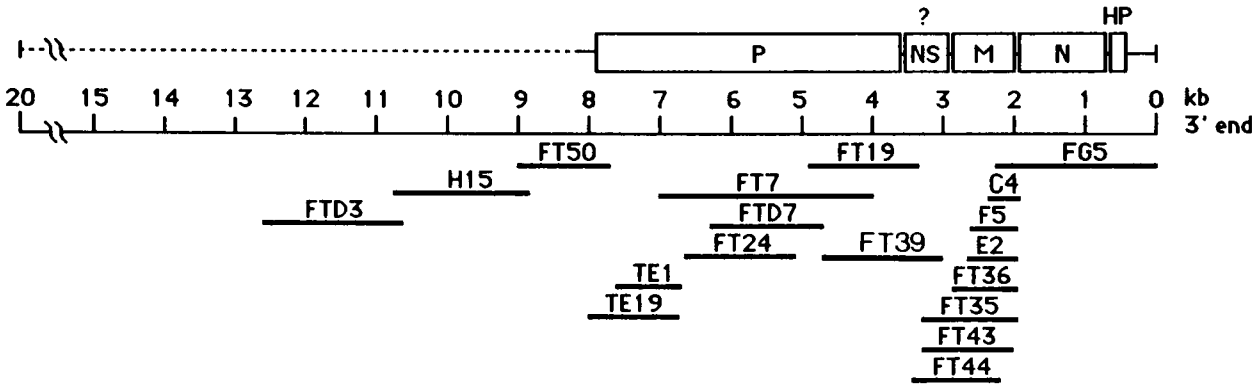


Fig. 2.6 Physical map of representative clones that lie within the first 13 Kb from 3'end of viral genome.

primed probe and selected for mapping. Fig. 2.5 shows the spotting pattern for mapping by the cross hybridization mathod. Table 2.2 lists all of the clones by name. From the 113 cDNA clones, 64 were found to yield a continuous sequence for approximately 12 Kb (60%) from the 3' end of the genome (Table 2.3 and Fig. 2.6). Forty nine clones remain to be mapped but they do not reside within the 3' terminal 12 Kb.

Mapping began with the 3'specific clone FG5 (Kapke and Brian, 1986) and was continued by moving toward the 5'end of the genome. Fig. 2.5 shows two example autoradiographs of spot hybridization. The blot in panel A was probed with the entire insert of clone FT44. This probe lit-up clones 5, 8, 14, 25, 26, 30, 32, 34-36, 38-45, 53, 54, B4, E2, E4, and F5. The blot in panel B was probed with nick-translated insert from clone FT39, the largest insert identified by probling with FT44. By comparing blot A with blot B it can be seen that an additional more 5' mapping clone appeared. This is indicated by an arrow in blot B and the clone was identified as FT19. Clone FT19, therefore, maps toward 5'end of clone FT39. At the time of this writting, 12 KB from the 3'end of genome (approximately 60% of the genome) has been mapped. The properties of the mapped and unmapped clones are summarized in Table 2.2. Seven clones named C4, E2, F5, FT35, FT36, FT43, and FT44 mapping in the positions shown in Fig. 2.6, were sequenced in total or in part to extend the TGEV genomic sequence that was known from clone FG5 (Kapke and Brian 1986). Identification of one open reading frame as the matrix protein gene in this region was based on amino acid sequence homology with the matrix proteins of

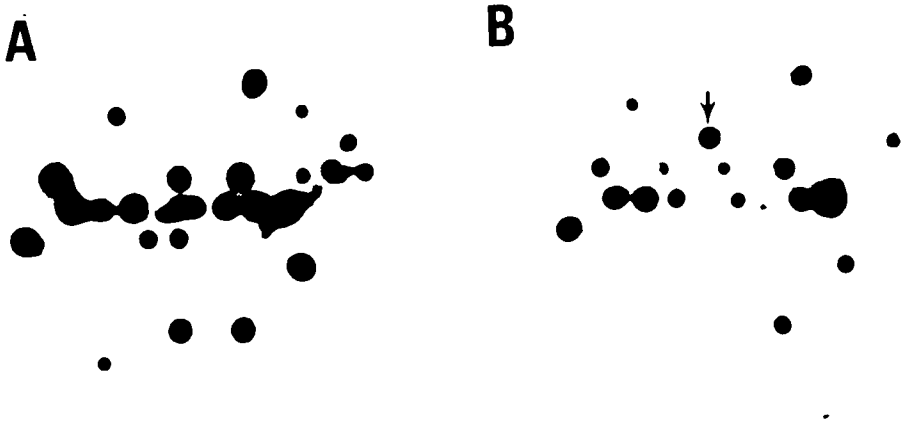


Fig. 2.7 Example autoradiographs of the spot hybridization technique used for mapping cDNA clones relative to the TGEV genome. Panel A: nitrocellulose paper spotted with DNA from all potential cDNA clones in a grid pattern was probed with 32 P-labeled insert from clone FT44. Panel B: Nitrocellulose paper, spotted with DNA in a pattern identical to that in panel A, was probed with insert from clone FT39 (a clone that has a 5' extention of FT44). A new clone, FT19, identified by the arrow in panel B, is now know to map as a 5' extention of clone FT39, note that several clones mapping to the 3' side of FT39 are missing in panel B.

MHV (Armstrong et al., 1984) and IBV (Boursnell et al., 1984). A detailed description of matrix protein gene and its deduced protein sequence is the subject of Chapter 3.

Three clones named FT44, FT39, and FT19 were shown to map at the 5'end of the matrix protein gene and sequence analysis shows them to lie between the matrix protein gene and peplomer protein gene. Sequencing of this region is incomplete at this time, but partial sequence analysis suggests that one or more open reading frames exist in this region. It is probable that this is the map portion of one or more nonstructural genes. Characterization of this (these) open reading frames is a continuing collaborative project with Dr. Paul Kapke.

Clones FT19, FT7, D7, FT24, in addition to clones TE1, TE19, were mapped at the region between 4 Kb to 8.2 Kb from the 3' end of viral genome. Clones TE1 and TE19 were prepared with a specific primer made from a sequence near the 5' end of clone FT8 (Table 2.4). All six clones were sequenced totally or in part to obtain the primary sequences of the peplomer protein gene and this information is the subject of Chapter 4.

Table 2.4 Summary of TGEV clones prepared with primer specific for the 5' end of the peplomer gene

Clone name[a]	Insert Size (bases)[b]	Orientation[c]
TE1	800	A
TE16	500	A
TE19	1,300	B

[a]Clones through TE26 were all <400 bases in size and therefore were not listed.

[b]As judged from migration of linearized fragment or isolated insert on agarose gels after double cutting with XhoI and PstI.

[c]Plasmids were double cut with XhoI and EcoR1. Clones appearing linearized only were in orientation B. Clones showing an isolated insert fragment were in orientation A.

CHAPTER 3

NUCLEOTIDE SEQUENCE ANALYSIS AND DEDUCED AMINO ACID SEQUENCE OF THE MATRIX PROTEIN GENE: EVIDENCE OF AN AMINO TERMINAL SIGNAL PEPTIDE

Introduction

The coronavirus matrix protein (M or E1) was described as an unsual virus glycoprotein when it was first characterized. Initial characterization was for the mouse hepatitis M protein which was described as "resident" protein that remained in the cell as a molecule embedded in the Golgi and endoplasmic reticulum membranes and failed to migrate to the cell surface as do most other virus glycoproteins (Holmes et al., 1981). The MHV M protein gene was the first coronavirus M protein to be cloned and sequenced (Armstrong et al., 1983), and its deduced amino acid sequence as well as other in vitro translation studies revealed that it possesses an internal hydrophobic signal sequences for membrane insertion rather than the common amino terminal signal peptide (Armstrong et al., 1984; Rottier et al., 1986). The signal that causes it to remain a resident protein is unknown. Presumably the M protein functions to initiate coronavirus budding into the endoplasmic reticulum (Tooze et al., 1984).

The M proteins of the avian infectious brochitis virus and the bovine coronavirus were the second and third coronavirus M proteins for which the sequence was determined (Boursnell et al., 1984; Lapps et al., 1987). For these M proteins there was also no

amino terminal signal peptide found and there was found to be much amino acid sequence homology with the MHV M protein and much evidence that there exists an internal signal sequence for membrane insertion.

In an initial report on the primary sequence of the 3' end of the TGEV genome for the first 2001 bases (Kapke and Brian, 1986) we reported a partial sequence for the TGEV M gene. To complete the nucleotide sequence of the M gene we prepared cDNA clones by priming with oligodeoxynucleotide sequences complementary to the TGEV genome. Assuming that a conserved intergenic sequence would be found in TGEV as has been found in the mouse hepatitis coronavirus (MHV) (Budzilowicz et al., 1985), and the avian infectious bronchitis coronavirus (IBV) (Brown and Boursnell, 1984), we prepared a synthetic oligodeoxynucleotide that is complementary to the TGEV intergenic sequence and used it as a primer for first-strand DNA synthsis in the preparation of additional genomic cDNA clones. Several cDNAs were thus prepared and seven that mapped within the first (3') 2601 bases were sequenced in part and another clone was sequenced completely to derive a potential sequence for the M protein. The nucleotide sequence predicted an M protein that shared many features with the M proteins of the mouse hepatitis virus, the bovine coronavirus and the avian infectious bronchitis virus, but also predicted an unexpected, potentially cleavable amino terminal signal peptide that makes the TGEV M protein strikingly different. The sequence analysis of the TGEV M gene and the properties of the deduced protein are described.

Materials and Methods

Cell and Virus

The Purdue strain of TGEV was grown on swine testicle (ST) cells as previously described (Kapke and Brian, 1986).

cDNA Cloning of TGEV Genomic RNA

Virus genomic RNA was prepared as previously described (Kapke and Brian, 1986). cDNA cloning was accomplished by the method of Gubler and Hoffman (1983) essentially as described (Kapke and Brian) except that the synthetic oligodeoxynucleotide 5'TTAGAAGTTTAGTTA3' was used as primer for first-strand cDNA synthesis. The primer was synthesized by the phosphoramadite method and was purified by polyacrylamide gel electrophoresis. Clones were selected by colony hybridization to random-primed cDNA prepared from size-selected genomic RNA (Kapke and Brian, 1986). Clones were initially physically mapped to obtain their approximate position on the genome by using a matrix cross-hybridization method in which plasmid DNA from individual clones was probed with purified inserts or segments of inserts that had been radiolabeled with 32P by nick-translation.

DNA Sequencing and Sequence Analysis

DNA sequencing was done by the chemical method of Maxam and Gilbert (1980) and sequence analyses were done with the aid of the computer programs developed by Queen and Korn (1984) and marketed as part of the Beckman Microgenic system, October, 1986 version (Beckman Instruments, Inc.). For protein homology determiantion

among the coronavirus M proteins, the "protein alignment" program of the Beckman Microgenic was used with MAXDIST set at 10000. For protein homology searchs among signal peptides, the "IFIND" program of the Intelligenetics system (Intelligenetics, Mountain View, CA) was used to search the National Biomedical Research Foundation Protein database.

Results

Nucleotide Sequence of the Matrix Protein Gene and Deduced Amino Acid Sequence of the Matrix Protein

Seven clones named C4, E2, F5, FT35, FT36, FT43, and FT44, mapping in the positions illustrated in Fig. 3.1, were sequence in part to extend the TGEV genomic sequence that was known from clone FG5 and J21 (Kapke and Brian, 1986). Clone FG5 maps at the extreem 3' end of the genome and contains the sequence for the hypothetical hydrophobic protein gene, the N gene and part of the M gene. Identification of the third open reading frame as the M gene sequence was based on regions of extensive amino acid sequence homology with the M proteins of MHV (Armstrong et al., 1984) and IBV (Boursnell, 1984). The sequencing strategy we used is described in Fig. 3.1.

The molecular weight of the glycosylated M protein has been estimated from electrophoretic migration patterns to be approixmately 28 Kd to 30 Kd (Brian et al.,1983; Garwes and Pocock, 1975; Wesley and Woods, 1986). We therefore anticipated that we would be able to deduce from the gene sequence a molecular

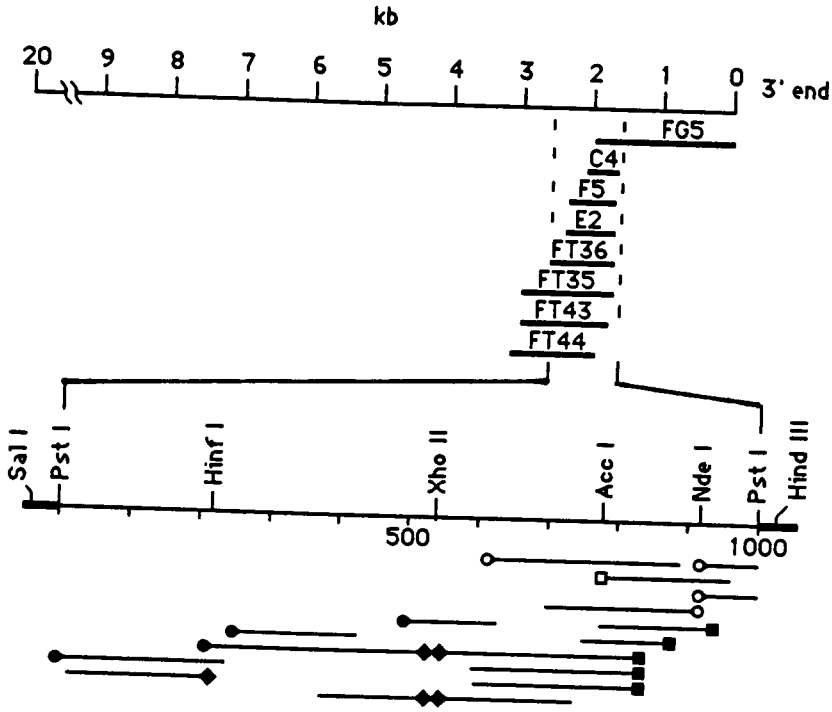
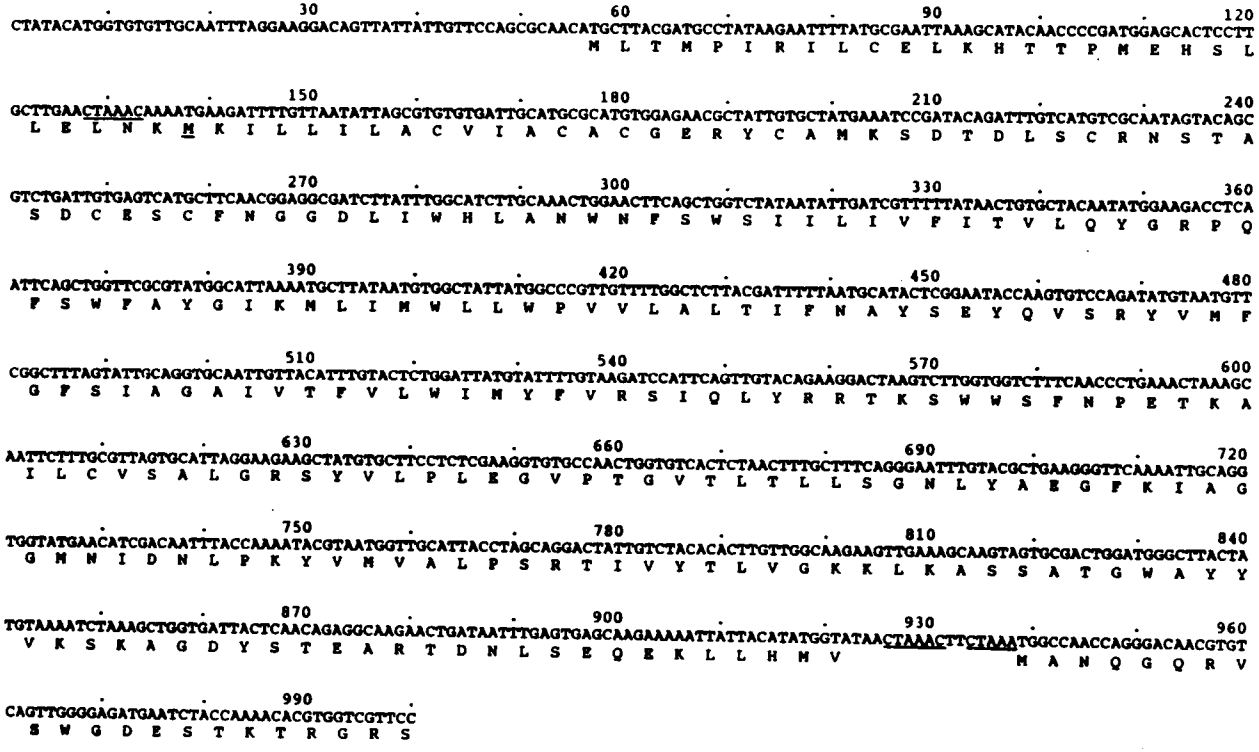


Fig. 3.1 Sequencing strategy used to derive the TGEV M gene sequence. cDNA clones FG5, C4, F5, E2, FT36, and FT35 were cloned into the Pstl site of vector pUC9 and were all found to be in the same orientation with respect to the virus genomic RNA illustrated at the top of the figure. FT43 was likewise cloned but was found to be in the opposite orientation. and indicate sites labeled on fragments of clone FG5 at 3'end of DNA with reverse transcriptase and at the 5'end with polynucleotide kinase, respectively (Kapke and Brian 1986). ● indicates 3' end labeling with reverse transcriptase at the Sal I site in the multiple cloning region of clones C4, F5, E2 and FT36. ■ indicates 3' end labeling with reverse transcriptase at HindIII site in the multiple cloning region of clone C4, F5, E2, FT36 and FT35. ◆ indicates 3' end labeling with reverse transcriptase at the Xho II site in clones E2 and FT43.

weight of 28 Kd or less for the unglycosylated protein. Surprisingly, the completed sequence identified earlier as part of the M gene (Kapke and Brian, 1986) yielded a protein of 289 amino acids having a molecular weight of greater than 32,000 (Figs. 3.2. and 3.3). Since the M proteins of the mouse hepatitis virus (Armstrong et al., 1984; Rottier et al., 1984) and the bovine coronavirus (Lapps et al., 1987) migrate in SDS-polyacryamide gels with electrophoretically determined molecular wights that are 11% less than their deduced molecular weight (23 vs 26 for the mouse hepatitis and 22 vs 26 for the bovine coronavirus), an apparent reflection of their hydrophobic nature, it is possible that the entire open reading frame identified above encodes the TGEV M protein. We hypothesized that this is unlikely, however, based on the documented evidence for leader-primed transcription in coronavirus replication (Makino et al., 1986 b), and on the existence of a primer binding-like intergenic sequence early in the open reading frame. The most probable site for initiation of transcription of the M message is suggested by the sequence CTAAAC beginning at base 128 in Fig. 3.2, which is part of a conserved intergenic sequence in TGEV genome. It is a sequence found in total and again in part between the M and N genes beginning at base 926 in Fig. 3.2, and also between the N and hypothetical hydrophobic protein genes (Kapke and Brian, 1986). It is also part of the intergenic sequence found in the MHV genome (Budzilowicz et al., 1985). If CTAAAC functions as part of an intergenic sequence that directs leader-primed synthesis and thereby defines the start of the M transcript for TGEV, then the M protein coding sequence could start

Fig. 3.2 Nucleotide sequence of the TGEV M gene and the deduced amino acid sequence for the protein. The nucleotide sequence comes from the part of the virus genome illustrated in Figure 3.1. A continous open reading frame beginning at nucleotide position 56 and continuing through nucleotide 922 is identified. The ACTAAAC intergenic sequences are underlined. The proposed amino terminus for the M protein is identified by an underlined methionine residue near base position 137.


30 60 90 120
CTATACATGGTGTGTTGCAATTTAGGAAGGACAGTTATTATTGTTCCAGCGCAACATGCTTACGATGCCTATAAGAATTTTATGCGAATTAAAGCATACAACCCCGATGGAGCACTCCTT
M L T M P I R I L C E L K H T T P M E H S L
150 180 210 240
GCTTGAACTAAACAAAATGAAGATTTTGTTAATATTAGCGTGTGTGATTGCATGCGCATGTGGAGAACGCTATTGTGCTATGAAATCCGATACAGATTTGTCATGTCGCAATAGTACAGC
L E L N K M K I L L I L A C V I A C A C G E R Y C A M K S D T D L S C R N S T A
270 300 330 360
GTCTGATTGTGAGTCATGCTTCAACGGAGGCGATCTTATTTGGCATCTTGCAAACTGGAACTTCAGCTGGTCTATAATATTGATCGTTTTTATAACTGTGCTACAATATGGAAGACCTCA
S D C E S C F N G G D L I W H L A N W N F S W S I I L I V F I T V L Q Y G R P Q
390 420 450 480
ATTCAGCTGGTTCGCGTATGGCATTAAAATGCTTATAATGTGGCTATTATGGCCCGTTGTTTTGGCTCTTACGATTTTTAATGCATACTCGGAATACCAAGTGTCCAGATATGTAATGTT
F S W F A Y G I K M L I M W L L W P V V L A L T I F N A Y S E Y Q V S R Y V M F
510 540 570 600
CGGCTTTAGTATTGCAGGTGCAATTGTTACATTTGTACTCTGGATTATGTATTTTGTAAGATCCATTCAGTTGTACAGAAGGACTAAGTCTTGGTGGTCTTTCAACCCTGAAACTAAAGC
G F S I A G A I V T F V L W I M Y F V R S I Q L Y R R T K S W W S F N P E T K A
630 660 690 720
AATTCTTTGCGTTAGTGCATTAGGAAGAAGCTATGTGCTTCCTCTCGAAGGTGTGCCAACTGGTGTCACTCTAACTTTGCTTTCAGGGAATTTGTACGCTGAAGGGTTCAAAATTGCAGG
I L C V S A L G R S Y V L P L E G V P T G V T L T L L S G N L Y A E G F K I A G
750 780 810 840
TGGTATGAACATCGACAATTTACCAAAATACGTAATGGTTGCATTACCTAGCAGGACTATTGTCTACACACTTGTTGGCAAGAAGTTGAAAGCAAGTAGTGCGACTGGATGGGCTTACTA
G M N I D N L P K Y V M V A L P S R T I V Y T L V G K K L K A S S A T G W A Y Y
870 900 930 960
TGTAAAATCTAAAGCTGGTGATTACTCAACAGAGGCAAGAACTGATAATTTGAGTGAGCAAGAAAAATTATTACATATGGTATAACTAAACTTCTAAATGGCCAACCAGGGACAACGTGT
V K S K A G D Y S T E A R T D N L S E Q E K L L H M V M A N Q G Q R V
990
CAGTTGGGGAGATGAATCTACCAAAACACGTGGTCGTTCC
S W G D E S T K T R G R S

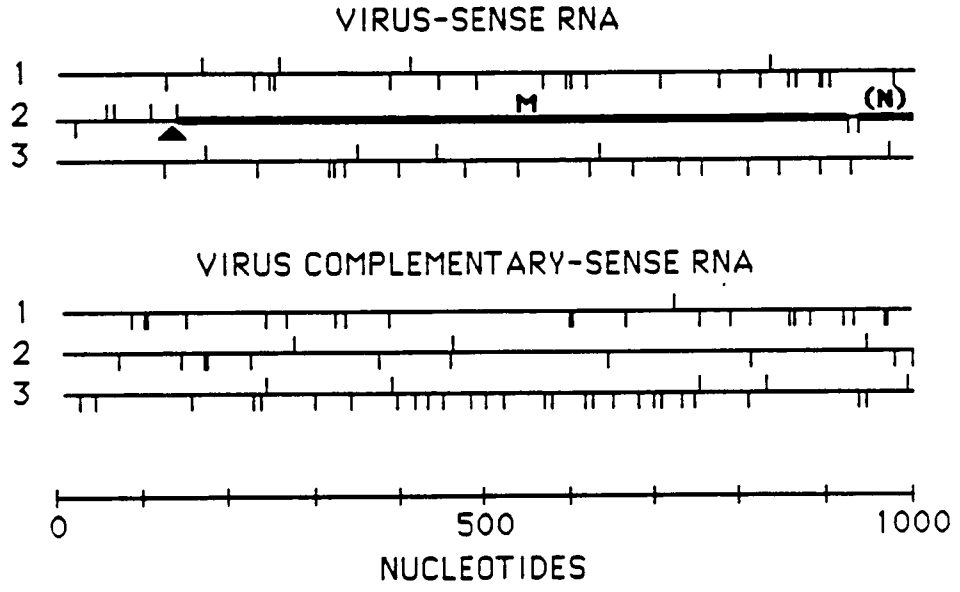


Fig. 3.3 Schematic diagram of possible open reading frames obtained when translating the nucleotide sequence illustrated in Fig. 3.2 as either virus-sense RNA or virus complementary-sense RNA in all three reading frames. Vertical bars above the line represent the first methionine codon that could serve as the initiation site for translation. In the M open reading frame the initiation methionine residue identified is the first one to follow the putative ACTAAAC intergenic sequence. The ACTAAAC intergenic sequence is identified with an arrowhead. Vertical bars below the line represent termination codons. M, open reading frame for the matrix protein. (N), partial open reading frame for the nucleocapsid protein.

with the first available methionine 3'-ward of the CTAAAC sequence, a codon that begins at base 137 in Fig. 3.2. It could also start with the second, downstream methionine, but this is a much less favorable Kozak consensus sequence (Kozak, 1983) for iniation of translation.

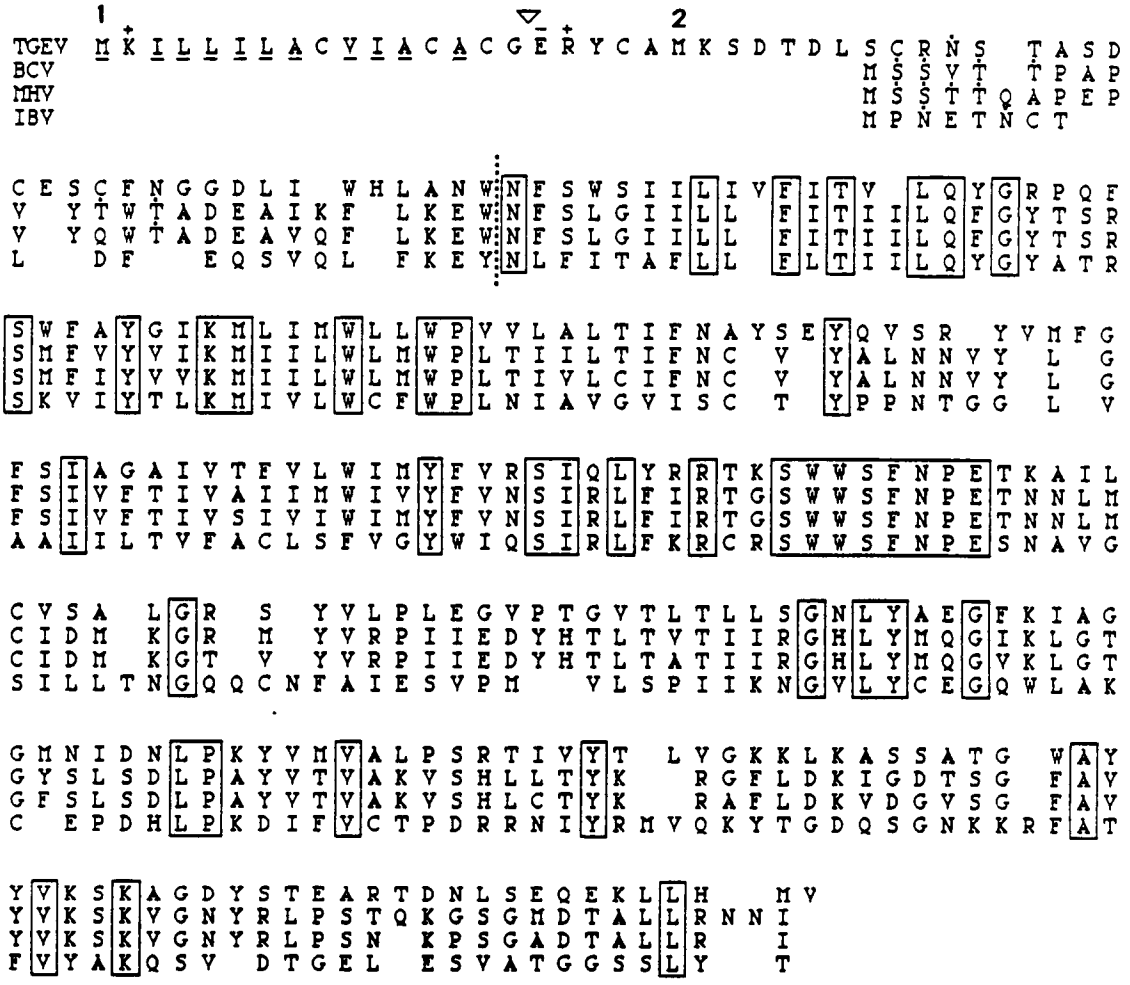
In a collaborative study, Dr. B. Hogue, University of California at Los Angeles, has reconstructed two versions of the M gene by joining clones FT36 and C4, prepared transcripts, translated each in an in vitro system, and compared the sizes of the protein products with those made when translating mRNAs isolated from the infected cell. The purpose was to estimate the real start site of the M gene by comparing the sizes of products from bona fide mRNA with reconstructed versions of mRNA possibilities. The two versions of the reconstructed M gene were (1) a gene that begins with the first ATG downstream from the CTAAAC intergenic sequence at base position 137 in Fig. 3.2 (long version of M), and (2) a gene that begins from the second ATG downstream from CTAAAC at base position 200 (truncated version of M). The results of these experiments demonstrated that mRNA from infected cells when translated in the absense of microsomes generates a product identical in size to that generated by the long version of the reconstructed M gene, and larger than that generated by the truncated version of the M gene (data not shown). These data strongly support our hypothesis that it is the the first ATG following the CTAAAC intergenic sequence that serves as the beginning of the M gene.

Using the first methionine codon downstream from the CTAAAC sequence as the initiaton site for translation, the deduced M protein is comprised of 262 amino acids and has a molecular weight of 29,544. It is moderately hydrophobic with 44% of its amino acids being hydrophobic, and is basic since it carries a net charge of +7 at neutral pH.

The Deduced Amino Acid Sequence of the TGEV M Protein Suggests the Existence of a Potentially Cleavable Amino Terminal Signal Peptide

A comparision of the amino acid sequence for the M proteins of TGEV, BCV, MHV, and IBV (Fig. 3.4) reveals several features that are shared among all four viruses, but also one feature for TGEV that is strikingly contrasting. Regions of high sequence homology are found among the proteins. Most notable is a stretch of 8 amino acids beginning at position 132 on the TGEV sequence that is identical for all four viruses. By computer analysis using a protein alignment function, 34%, 32%, and 19% of the TGEV protein sequence is homologous with that of BCV, MHV, and IBV, respectively. Furthermore, a hydrophobicity plot of the TGEV M protein shows three internal hyrophobic domains that align with similar domains in MHV (Fig. 3.5) (Rottier, 1985), BCV (Lapps et al., 1987), and IBV (Boursnell et al., 1984). This suggests that the topology of the four proteins is similar, that is, from its entrance into the virion membrane and as it extends towards the carboxy terminus, the protein spans the membrane 3 times and has a relatively hydrophilic intravirion carboxy terminal region (Rottier et al., 1986).

Fig. 3.4 Deduced amino acid sequence of four coronavirus M proteins aligned for maximum homology. The TGEV M protein of 262 amino acids (top row), BCV (Mebus strain) M protein of 230 amino (second row), MHV (A59 strain) M protein of 228 amino acids (third row), and IBV (Beaudette strain) M protein of 225 amino acids (fourth row) were aligned for maximum homology. For this, the 8 amino acid sequence SWWSFNPE was important in making the initial alignment. For computer analysis, the amino terminus of the TGEV M protein for a distance of 28 amino acids was not included in order to maximize the homology. The putative site for entrance of the protein into the virion envelope is indicated by a hatched vertical line between amino acids 26 and 27 on the MHV sequence. Potential N glycosylation sites (for TGEV and IBV) and O-glycosylation sites (for BCV and MHV) are indicated with a solid circle above the amino acid. The signal peptide-like properties of the amino terminal 16 amino acids is depicted. Within this region the 11 hydrophobic amino acids are underlined. The basic amino acids at position 2 and 18, and the acidic amino acid at position 17 are indicated. The potential site for peptidase cleavage, according to the rule of Von Heijne (1986), is indicated with an open triangle. The numbers 1 and 2 identify the first and second methionine residues in the potential TGEV M sequence.


1
2
TGEV
BCV
MHV
IBV

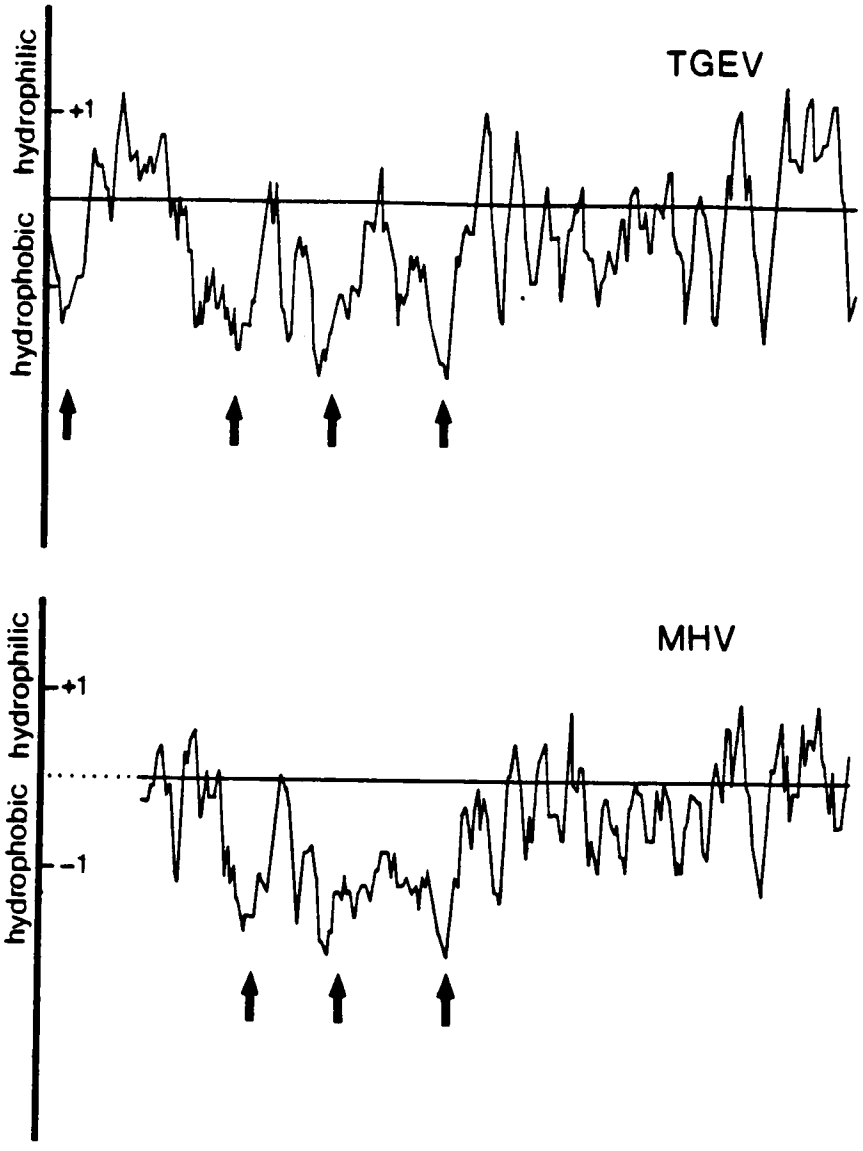


Fig. 3.5 Hydrophobicity plots for the TGEV and MHV proteins. The method of Hopp and Woods (1983) was used using a 7 amino acid window. Sequences were aligned at their carboxyl termini and no gaps were present in the sequence. Hydrophobic regions sufficient for membrane anchorage are indicated by dark vertical arrows. Note the hydrophobic amino terminus of the TGEV M protein.

The striking feature of the TGEV M protein is its much longer amino terminus that includes a sequence resembling a cleavable peptide for membrane insertion. Assuming a parallel structure for the M protein of the four viruses and assuming the MHV M protein enters the virion evelope at position 26 (Rottier et al., 1986), then the external amino terminal portion is 28 amino acids for BCV, 21 for IBV, and 54 for TGEV. Within the first 54 amino acids there is one asparagine at position 32 that has the appropriate surrounding sequence required for N-linked glycosylation (Hubbard and Ivatt, 1981). The kind of glycosylation shown for the TGEV M protein (Jacobs et al., 1986). Unlike the amino terminus of the MHV, BCV and IBV proteins, the TGEV M protein is hydrophobic for the first 16 amino acids. Eleven of the 16 terminal amino acids are hydrophobic, and amino acids at position 2, 17, and 18 are charged (Fig. 3.4). This sequence has the properties of a cleavable amino terminal signal peptide and peptidase cleavage would occur between amino acids 16 and 17 judged by "-3,-1 rule" (the cleavage site is 0; the residue in position -1 must be either Ala, Ser, Gly, Cys, Thr, or Gln; the residure in -3 must not be aromatic, charged, or large and polar; Von Heijne, 1986).

Discussion

The M protein of MHV was the first coronavirus M protein to be sequenced (Armstrong et al., 1985) and its topography with regard to membrane orientation and insertion has been carefully documented (Rottier et al., 1984; Rottier et al., 1986). It therefore serves as the prototypic coronavirus M protein to direct

the budding of virus into the rough endoplasmic reticulum and Golgi since it inserts into these membranes and is found at hightest concentrations there (Tooze et al., 1984). Presumably virus assembly is mediated through an interaction between the M protein (in the membrane) and the N protein (in the cytoplasm) or the RNA (in the cytoplasm), or both. The M protein of MHV does not have an amino terminal cleavable signal peptide for membrane translocation, but rather utilizes one or more of its three internal hydrophobic domains for signal recognition particle-directed membrane insertion (Rottier et al., 1985; Rottier et al., 1986). Because of their strong parallel structure with regard to biochemical properties, that is, three internal hydrophobic domains in similar positions and over all length, the BCV M protein (Lapps et al., 1987) and the IBV M protein (Boursnell et al., 1985) presumably have a topography very similar to that of the MHV M protein and a similar mechanism for membrane translocation.

It came as a surprise to us, therefore, that the deduced TGEV M protein has, in addition to the three internal hydrophobic signal domains, an amino terminal sequence having properties of a cleavable signal peptide for membrane insertion (Von Heijne, 1986). A nearly identical gene sequence has been reported for the same strain of TGEV and direct evidence for the peptidase cleavage site was show (Laude et al., 1987). At least two fundamental questions are therefore raised by the existence of an amino terminal signal peptide in the TGEV M protein: (i) What is the evolutionary origin of such a

sequence, and, (ii) what role does the sequence play for the TGEV M protein?

With regard to the evolutionary origin of the hydrophobic amino terminus, two possibilities can be entertained, assuming the four coronavirus, TGEV, BCV, MHV and IBV have a common evolutionary origin. (i) There was an amino terminal hydrophobic sequence in the primordial protein which was lost during the evolution of BCV, MHV and IBV since it is superfluous for membrane insertion and virus assembly. Mechanistically, the loss of a genetic sequence could be explained by the "jumping polymerase" hypothesized by Lai et al. (Makino et al., 1986a). The nucleotide sequence encoding the amino terminal hydrophobic sequence could have been eliminated by the polymerase as it copied the negative strand template. (ii) There was no amino terminal hydrophobic sequence in the primordial protein but the TGEV M protein aquired one during evolution. The polymerase, during replication of the TGEV genome, could have "jumped" to a minus strand RNA template that carried the sequence for the hydrophobic signal. While negative strand RNA of this kind is probably nonexistent or of low abundance in eucaryotic cells normally, it does exist in cells coinfected with another RNA virus. Conceivably the TGEV M protein could have acquired its amino terminal sequence by copying the negative strand of another RNA virus. In this regard, it is interesting that 6 of the first 8 amino acids in the TGEV sequence are identical to the VSV G protein signal sequence (Rose et al., 1981).

With regard to the function of the amino terminal hydrophobic sequence, we propose that it may aid in the translocation of the amino

terminus of the TGEV M protein through the endoplasmic reticulum as would other amino terminal signal peptides. It is interesting to note that even after removal of the amino terminal peptide, the TGEV M protein extends 12 amino acids farther from the envelope than does BCV, 11 amino acids farther than MHV, and 16 amino acids farther than IBV. Perhaps the amino terminal signal peptide, while not being an absolute requirement for translocation of the TGEV M protein because of its apparant internal signals for translocation, aids enough in the translocation of the additional external amino terminal sequence that it was evolutionarily selected.

Because intracellular forms of M do not have a cleaved signal (B. Hogue, personal communication), whereas the virion form of M does have a cleaved signal (Laude et al., 1987), we propose that the M protein with an uncleaved signal peptide would have the orientation depicted in Fig. 3.6. With this orientation, the cleavage site is buried in the membrane and fewer than 37 amino acids are exposed as a loop on the lumenal side of the endoplasmic reticulum. Once virion assembly begins, the cleavage site becomes exposed to the signal peptidase, cleavage occurs, and 37 amino acids, including a glycosylated asparagine, are left remaining on the virion surface. Because of the amino terminal hydrophobic sequence, the TGEV M protein may behave differently from its MHV, BCV or IBV counterparts with regard to intracellular trafficing.

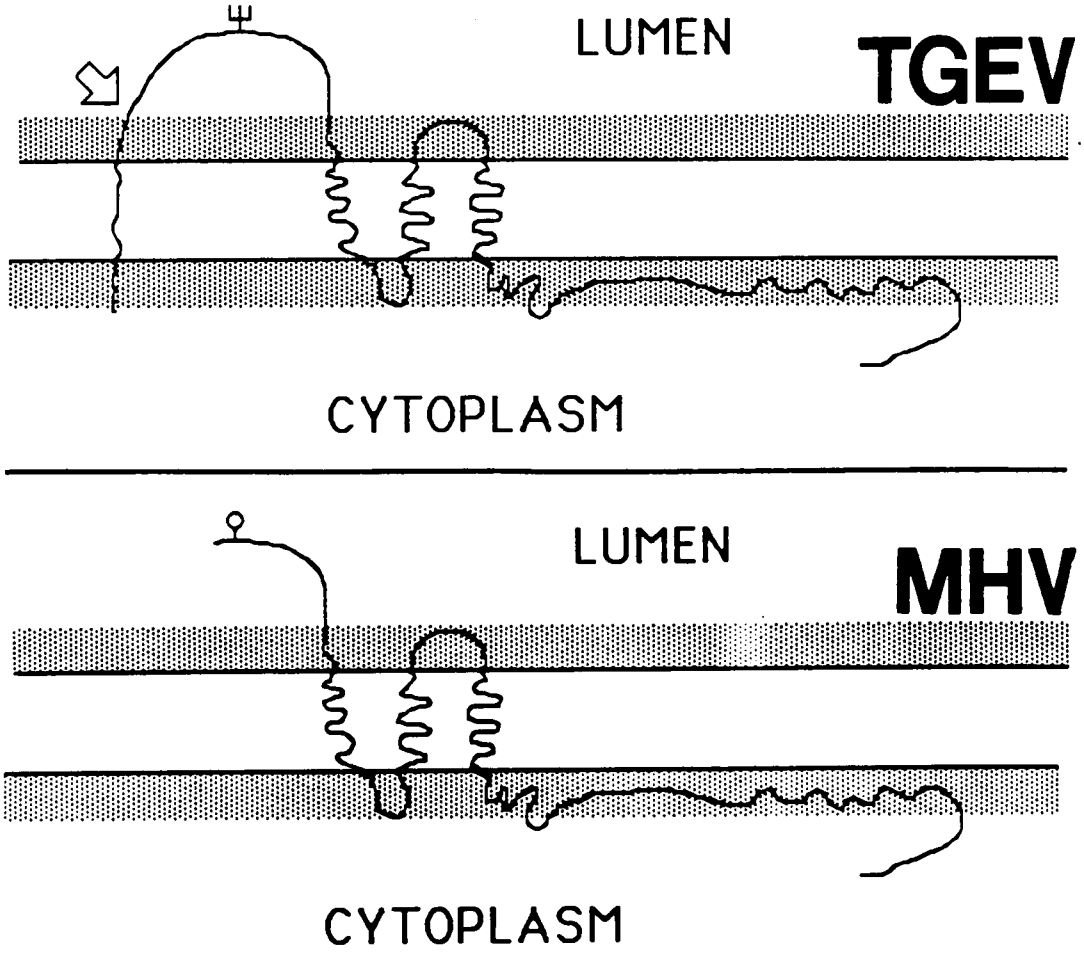


Fig. 3.6 Proposed topography of the TGEV M protein. The proposed topography of the TGEV M protein with an uncleaved amino terminal signal sequence is based on the proposed topographical arrangement of the MHV M protein (Rottier et al., 1986). Our model predicts that synthesis of the TGEV M protein is arrested after the amino terminal signal sequence (of about 16 amino acids) is made, and then proceeds after signal insertion into the membrane, probably all with the aid of a signal recognition particle. This would be unlike the MHV M protein which is not inserted until approximately 150 amino acids are synthesized. The TGEV M amino terminal signal is not cleaved, but remains in the orientation shown. The N glycosylation site (Ψ) and O-glycosylation site (♀) on the protein are shown.

CHAPTER 4

NUCLEOTIDE SEQUENCE ANALYSIS AND DEDUCED AMINO ACID SEQUENCE OF THE PEPLOMER PROTEIN GENE: CONSTRUCTION OF A VACCINA VIRUS-PEPLOMER RECOMBINANT

Introduction

Transmissible gastroenteritis coronavirus (TGEV) is one of the two most important members of the coronavirus family from a medical (economic) point of view. The other member is the avian infectious bronchitis coronavirus. TGEV causes a severe enteric disease resulting in a high mortality rate (up to 100%) in young piglets. Unlike the very successful vaccines that have been produced to control avian infectious bronchitis virus, vaccines that have been developed to control TGEV have been overwhelmingly unsuccessful (Saif et al., 1986). Current TGEV vaccines rely on the transmission of lactogenic immunity from vaccinated sows to nursing piglets and one problem with current vaccines is that they induce poor lactogenic immunity. A second and equally important problem is that vaccination does not cure a herd of carrier animals. It is not known how TGEV resides in a herd between disease outbreaks.

Because of the challenge for developing an effective TGEV vaccine in the face of the failure of traditional vaccine approaches, and to understand the role of individual TGEV proteins in coronavirus infection and immune response, we have undertaken a rigorous characterization of the TGEV structural protein genes. The cloned gene of a protectively immunogenic TGEV protein could perhaps be used

in a replicating virus expression vector that would stimulate good lactogenic immunity. Perhaps a vector of this type, now known to include the vaccinia virus (Small, 1985), or the pseudorabies virus (Leonard Post, personal communication), would present the TGEV protein to the host in a way that it would stimulate good protective immunity. TGEV contains three major structural proteins (Brian, et al., 1980; Garwes et al., 1975, 1976; Moreau, M.S. Thesis, 1982). These include a 180-200 kd glycosylated peplomer protein (P or E2), a 30 kd glycosylated matrix protein (M or E1), and a 50 kd phosphorylated nucleocapsid protein (N). The P protein has been shown to elicit the production of neutralizing antibodies (Laude et al., 1986; Jimerez et al., 1986) which are able to protect suckling piglets (Garwes et al., 1979).

In this communication we report the cDNA cloning and sequencing of the TGEV P gene and its deduced amino acid sequence. We also report the subcloning of the P gene into the vaccinia virus in preparation for its expression in this recombinant vector.

Materials and Methods

Virus and Cells

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

cDNA Cloning of TGEV Genomic RNA

Virus genomic RNA was prepared as previously described (Kapke and Brian, 1986). cDNA cloning was accomplished by the method of Gubler

and Hoffman essentially as described (Kapke and Brian, 1986) except that the intergenic primer 5'TTAGAAGTTTAGTTA3' was used to create a cDNA library of the genome and the site specific primer 5'GAGCTGTGCCATTATTAGTGT3' was used to clone a gap that existed at the 5' end of the P gene. The primers were synthesized by the phosphoramadite method and purified by polyacrylamide gel electrophoresis. Clones prepared using the intergenic primer were selected by colony hybridization to random-primed cDNA prepared from size-selected genomic RNA. These were initially physically mapped to obtain their approximate position on the genome by using a matrix cross-hybridization method in which plasmid DNA from individual clones was probed with purified inserts that had been radiolabeled with ^{32}P by nick-translation. Clones generated from the primer specific for the 5' end of the P gene were screened by colony hybridization to a nick-translated XhoI fragment from the 5' end of clone FT7.

DNA Sequencing and Sequence Analysis

DNA sequencing was done by the chemical method of Maxam and Gilbert (1982) and sequence analyses were done with the aid of the computer program developed by Queen and Korn (Queen and Korn, 1984) as described previously (Kapke et al., 1987).

Construction of Vaccinia-TGEV Peplomer Recombinant

Clone TE19 encoding the amino-terminal half of the P gene was digested with HpaI and XhoI. HpaI site cuts at a unique site 8 bases upstream of the AUG initiation codon of the P gene and XhoI cuts at a unique site on the open reading frame of the P gene. The incomplete

C-terminal end of the P gene in clone FT7 was replaced by the XhoI-EcoRI fragment isolated from clone FT19, which carried the intact C-terminal end of the P gene. The resulting clone TC5, obtained by joining FT7 and FT19, was linearized with XhoI and SmaI. The XhoI site is unique in the P gene and the SmaI site is in the multiple cloning region of the pUC9 vector.

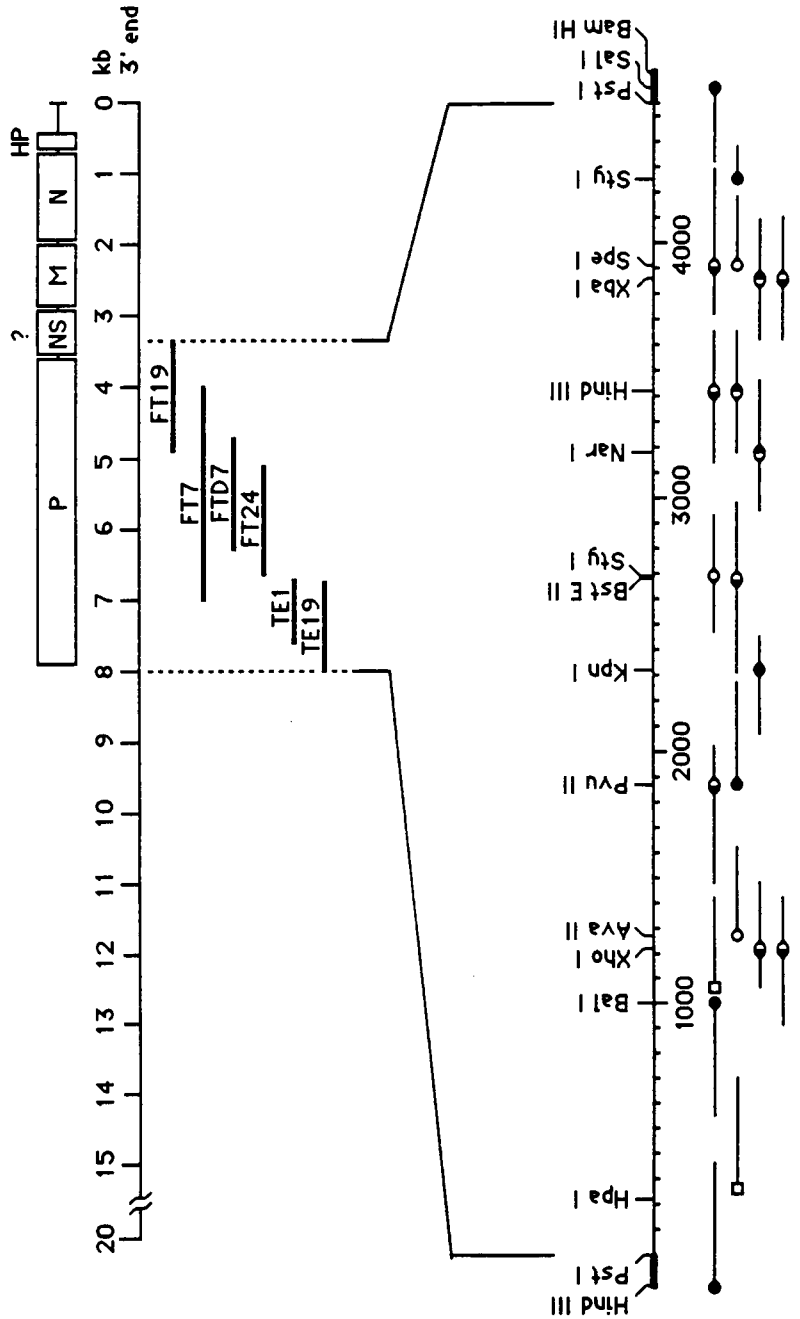
Both the HpaI-XhoI fragment of clone TE19 and the XhoI-SmaI fragment of clone TC5 were recovered by electrophoresis onto DEAE paper (NA45) from an agarose gel. These two isolated DNA fragments were ligated with SmaI-linealized, phosphatase-treated vaccinia virus expression vector pSC11 by using T4 DNA ligase. In this reaction, 80 ng of isolated HpaI-XhoI fragment and 200 ng of isolated Xho-Sma I fragment (a molar ratio of 1:1) were used. The final DNA concentration of the ligation mixture was less than 50 ug/ml and the molar ratio of insert to vector was set between 3 and 5. The restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. The ligated DNA mixture was used to transform E. col. JM83 (Hanahan, 1983). Recombinant clones carrying the peplomer gene (P) were identified by colony hybridization using P gene-specific, ^{32}P-labeled DNA isolated from clone FT7 insert as a probe.

Results

Sequencing of the Peplomer Protein Gene

The physical map of the six clones used to determine the sequence of the peplomer protein gene FT19, FT7, FTD7, FT24, TE1 and TE19 are shown relative to the virus genome in Fig. 4.1. The sequencing

Fig. 4.1 Sequencing strategy used to derive the TGEV P gene sequence. cDNA clones FT19, FT7, FTD7, FT24, TE1, and TE19 were cloned into the Pst I site of vector pUC9. Clones FTD7, FT24, and TE1 had their 3' end (virus sense) situated near the Hind III site on the vector. Clones FT19, FT7, and TE19 were of the opposite orientation. ● indicates sites labeled with kinase. ○ indicates sites labeled with reverse transcriptase by a 3' fill-in reaction. □ indicates 3' fill-in reaction using reverse transcriptase at the Hind III site on FT7 or at the Sal I site of TE1.


HP
?
P
NS
M
N
kb
3' end
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
20
FT19
FT7
FTD7
FT24
TE1
TE19
Hind III
Pst I
Hpa I
Bal I
Xho I
Ava II
Pvu II
Kpn I
Bst E II
Stu I
Nar I
Hind III
Xba I
Spe I
Stu I
Pst I
Sal I
Bam HI
1000
2000
3000
4000

strategy is likewise shown. Four of the clones used, FT19, FT7, FTD7 and FT24, were generated from the intergenic primer. Clones TE1 and TE19 were generated using the primer derived from the 5' end of clone FT7.

From sequencing the clones described in Fig. 4.1, either in total or in part, an open reading frame of 4341 bases was identified that we conclude to represent the gene for the peplomer protein (Fig. 4.2). This conclusion is based on three observations. (1) The deduced protein of 1447 amino acids long has a molecular weight of 160,345, which is close to the measured molecular weight of the unglycosylated P polypeptide identified by gel electrophoresis (Moreau, M.S. Thesis, 1982). (2) The gene maps in a position on the genome from which mRNA's encoding the P protein are generated, assuming there is a leader primed, nested set structure of TGEV mRNA molecules (Dennis and Brian, 1982, Jacobs et al., 1986). (3) There is extensive amino acid sequence homology with the P proteins of the mouse hepatitis virus (Schmidt et al., 1987) and the avian infectious bronchitis virus (Binns et al., 1986; Binns et al., 1985) (discussed below).

The probability that the ATG beginning at base position 88 in Fig. 4.2 is the beginning of the P gene is suggested strongly by a conserved intergenic sequence of ACTAAAC just 33 bases upstream from the ATG start site that could serve as a leader binding site for generation of the mRNA of approximately 8.4 kb (Dennis and Brian, 1982; Jacobs et al., 1986). An internal sequence of CTAAAC beginning at base 208 is also found and may also function as a sequence for leader priming. This apparently does not happen since it would

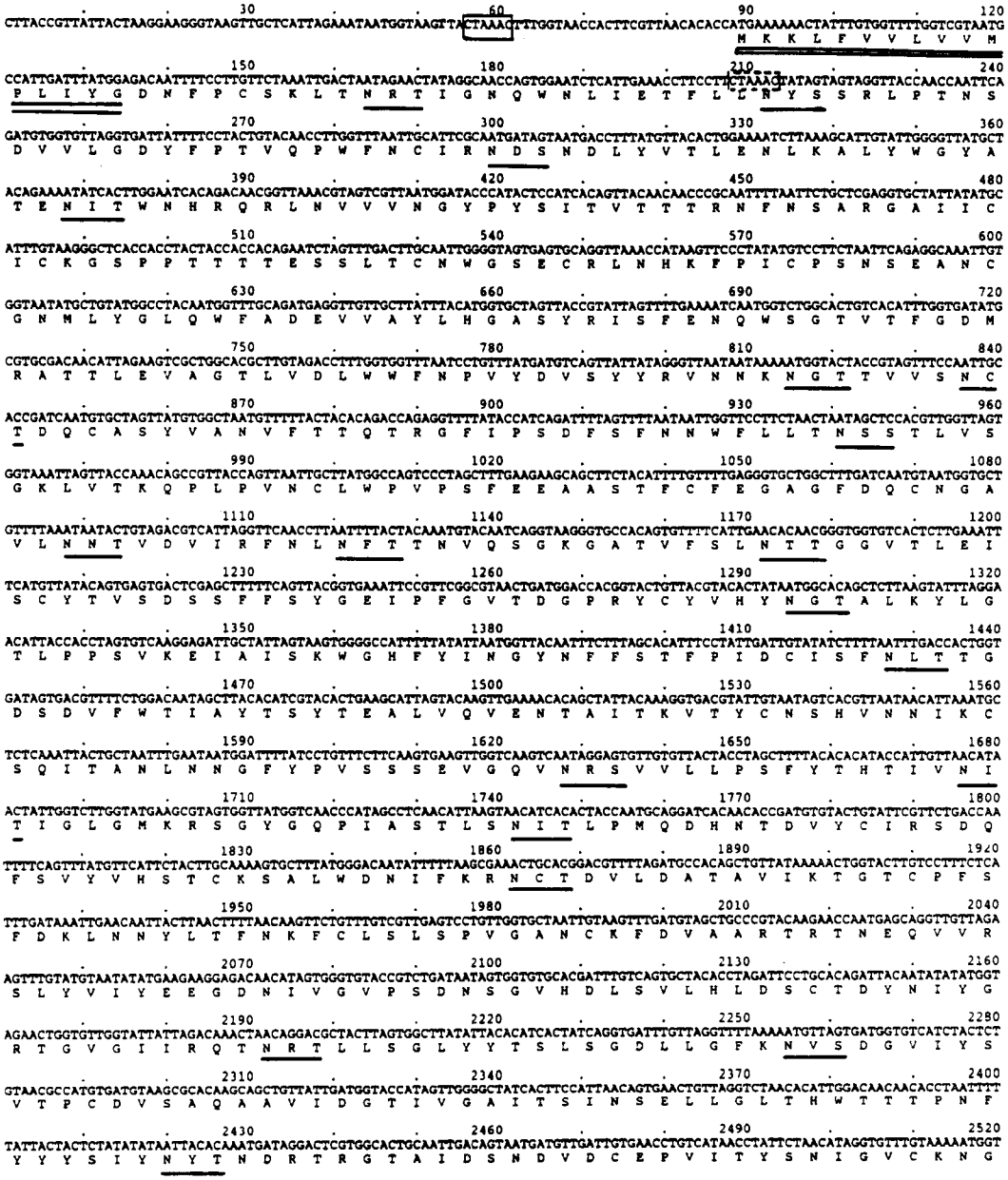


Fig. 4.2 The primary nucleotide sequence and the deduced amino acid sequence of TGEV peplomer gene. The intergenic concensus sequences are boxed. The potential N-glycosylation sites are underlined. The amino terminal signal peptide and membrane-anchoring region are indicated by open and closed lines respectively.

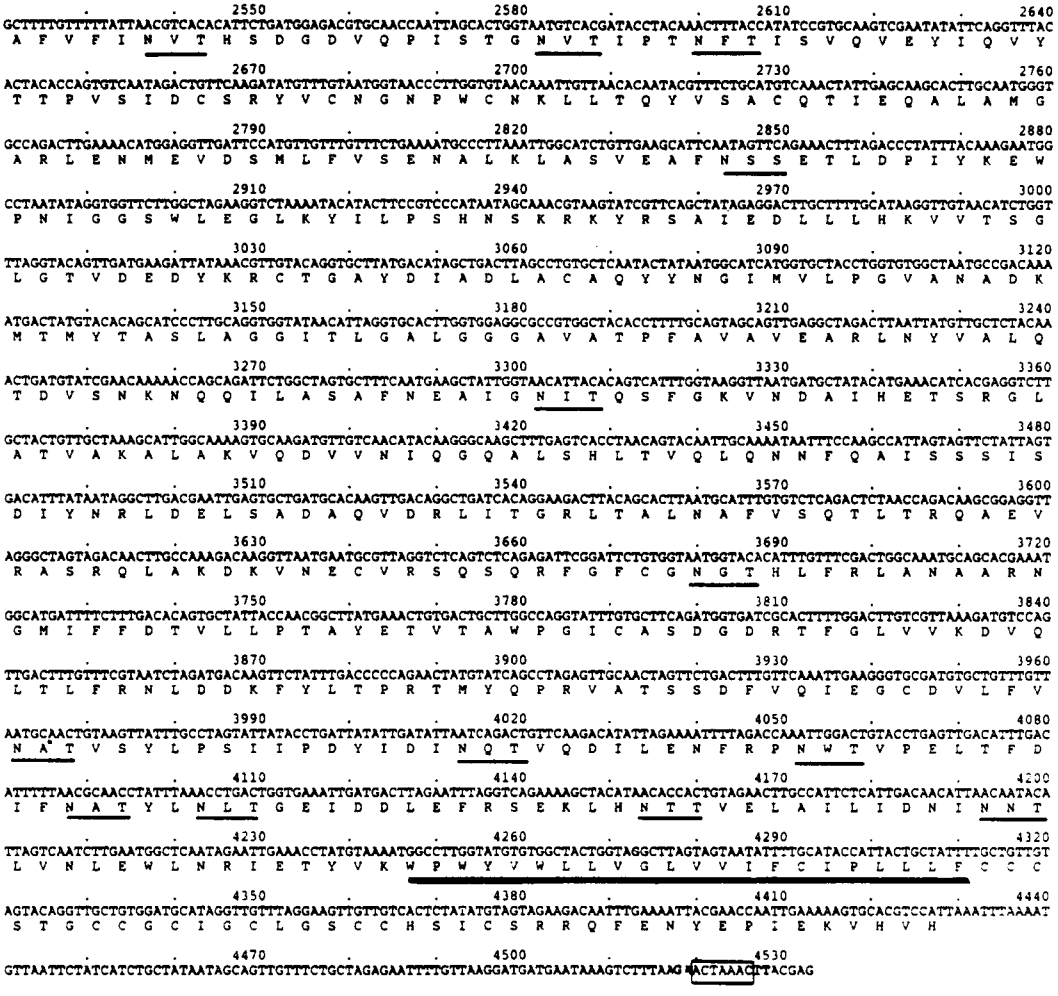


Fig. 4.2 (continued)

generate an mRNA of approximately 200 fewer bases (a molecular weight difference of approximately 100 kd) and no polyadenylated RNA species of this size is observed in TGEV-infected cells (Brian and Dennis, 1982; Jacobs et al., 1986). An amount of this RNA too small to be observed may be present, however. Use of the second (downstream) CTAAAC sequence as a primer binding site also seems unlikely because it would generate an mRNA with an unusually long 5' untranslated region since the first possible ATG start codon is 399 bases downstream from the CTAAAC sequence. A comparison of all intergenic sequences found so far in the TGEV genome is made in Table 4.1.

No other open reading frames of larger than 200 amino acids was found when translating the P gene sequence, or the sequence complementary to the P gene, in all other possible reading frames.

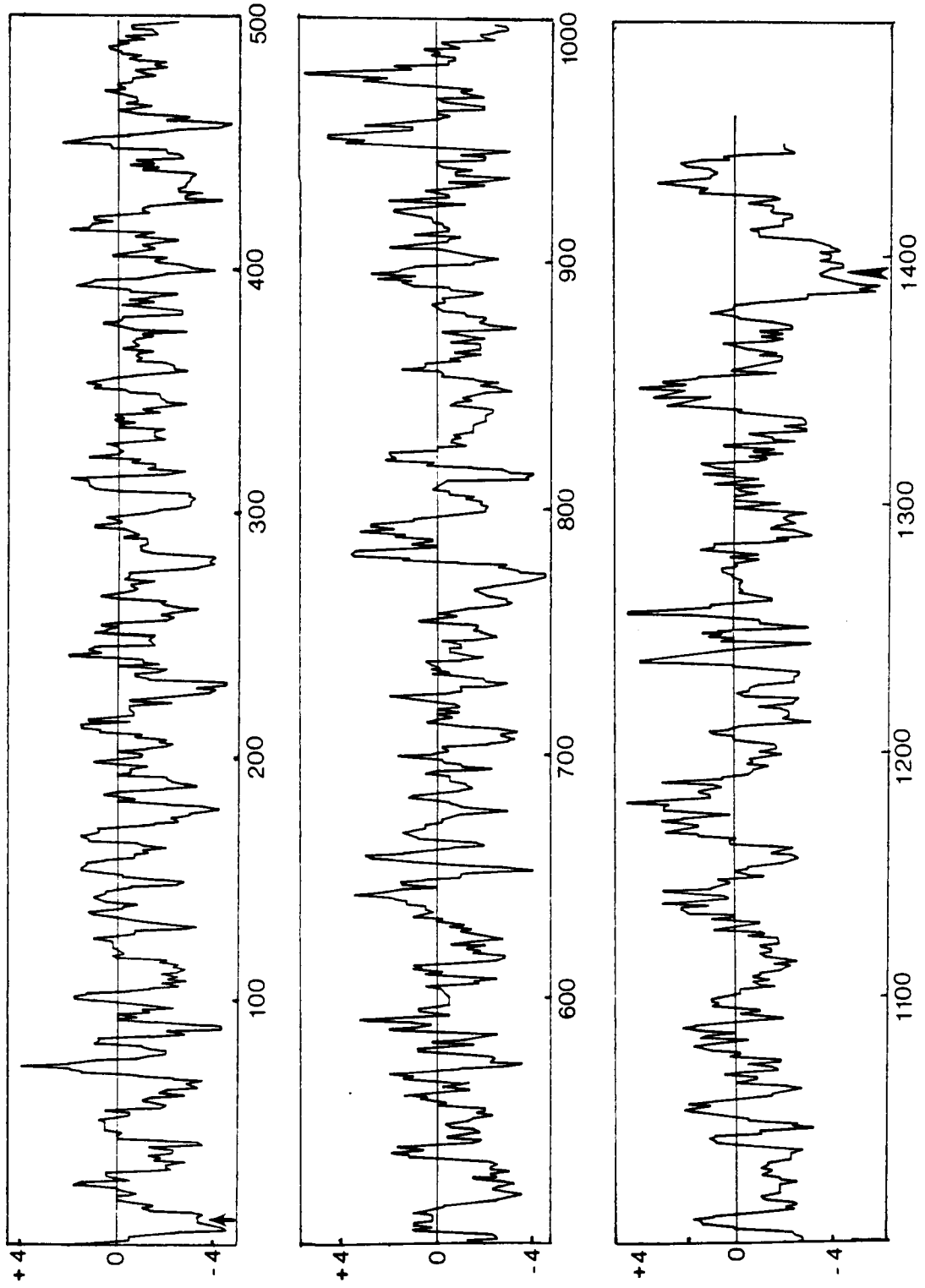
The 1447 amino acid peplomer protein is composed of 126 acidic amino acids and 96 basic amino acids and so therefore would have a net charge of -30 at neutral pH. 518 amino acids (35%) are hydrophobic. The deduced peplomer protein shows many properties that would be expected of a large virion glycoprotein. A hydrophobicity profile (Fig. 4.3) reveals several distinct regions of hydrophobicity but two of special significance. The first is a stretch of 16 amino acids at the amino terminus, 12 of which are hydrophobic, that are followed by a negatively charged aspartic acid at position 17. This sequence is typical of an N-terminal signal peptide for membrane translocation of a surface protein (von Heinje, 1986). From the common pattern of signal sequences giving the "-3,-1" rule (von Heinje, 1986), the predicted peptidase cleavage site would occur between glycine at position 16 and aspartic acid at position 17.

Table 4.1 Consensus Intergenic Sequences in TGEV Genome

Location	Sequence	Reference
N/HP	A ACTAAAC GA	Kapke and Brian, 1987
M/N	A ACTAAAC TT	Kapke and Brian, 1986
NS1/M	A ACTAAAC AA	Chapter 3
P/NS1	A ACTAAAC TT	this paper
?/P	T ACTAAAC TT	this paper

[a]Sequences that preceed the open reading frame and that are homologous. The absolutely conserved 7 base sequence is underlined and the flanking bases are shown.

Fig. 4.3 A hydrophobicity plot of the TGEV peplomer protein. The method of Hopp and Woods (1983) was used. Regions above line are hydrophilic, regions below the line are hydrophobic. A amino terminal signal peptide (arrow) and C-terminal anchorage region (arrowhead) are indicated.


+4
0
-4
100
200
300
400
500
+4
0
-4
600
700
800
900
1000
+4
0
-4
1100
1200
1300
1400

The second special hydrophobic domain begins 58 amino acids from the carboxy terminus and extends for 19 residues. This domain is typical of an anchor sequence on virion surface proteins (von Heijne, 1986) and may serve to anchor this peplomer protein. Throughout the protein sequence there are 32 potential asparagine glycosylation sites as indicated by the asparagine-X-threonine or asparagine-X-serine receptor sequences, where X can represent any amino acid but proline (Ivatt and Hubbard, 1981). Assuming that a high mannose sugar chain would contribute 2000 to the molecular weight (Klenk and Rott, 1981), then up to 20 of the potential N-glycosylation sites could be used to bring the final molecular weight of the protein to 200,000.

Construction of a Vaccinia-TGEV Peplomer Recombinant

Clones TE19, FT7 and FT19 were used to reconstruct the complete TGEV P gene and to make a vaccinia recombinant (Fig. 4.4). During subcloning, 73 positive clones of pSC11 recombinants were identified by colony hybridization using peplomer-specific sequence derived from clone FT7 (data not shown). Clones possessing inserts of the correct orientation for transcription from the vaccinia 7.5 promoter will be identified after digestion with restriction enzymes KpnI and PstI. Each enzyme cuts the peplomer gene and the pSC11 vector once respectively.

Discussion

During the course of this research the first four reports on the primary structure of coronavirus peplomer protein genes were published. These included two separate antigenic strains of avian

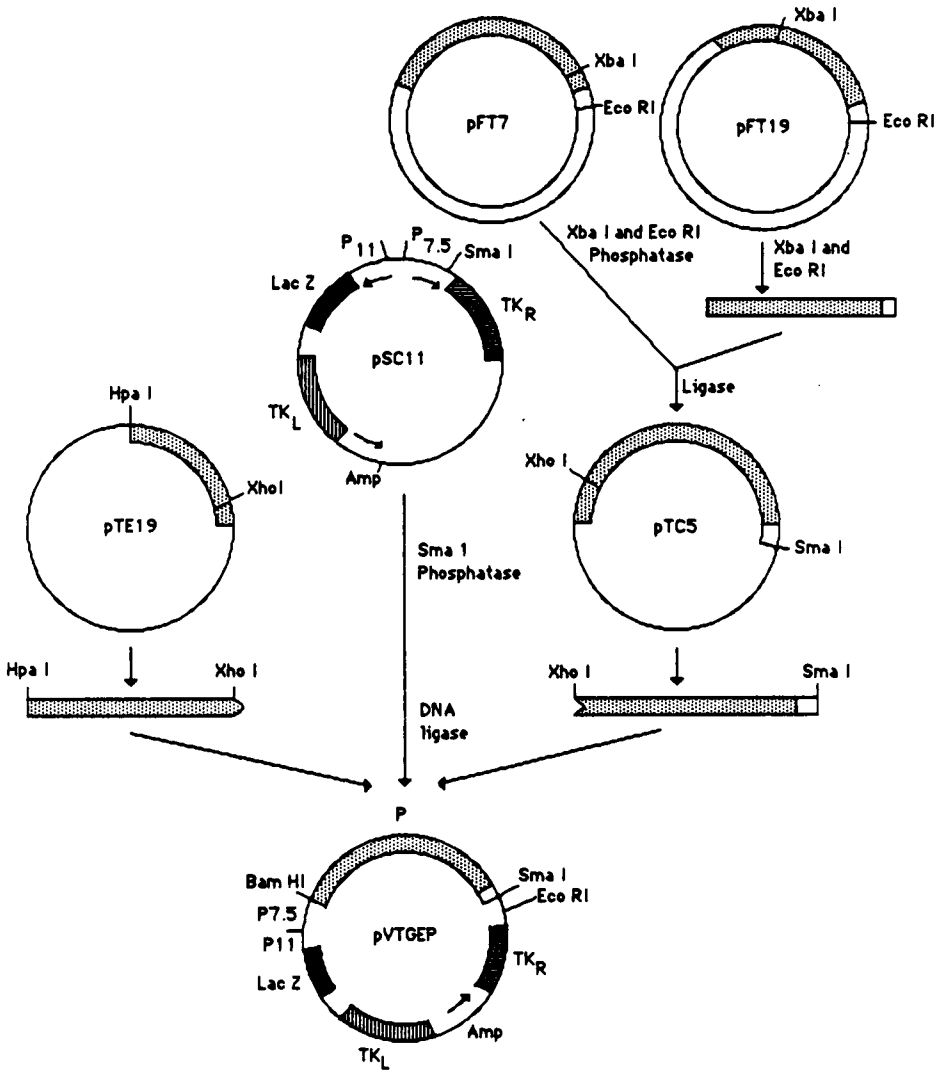


Fig. 4.4 Construction of the vaccinia-TGEV peplomer recombinant. Details are described in the Materials and Methods.

infectious bronchitis virus (IBV) (Binns et al., 1985; Binns, et al., 1986; Niesters et al., 1986), and the JHM strain of the mouse hepatitis virus (Schmidt et al., 1987). During the writing of this work the first paper describing the primary structure of the TGEV peplomer protein was published (Rasschaert and Laude, 1987). It happens also to be the Purdue strain of TGEV. A very careful comparison can therefore be made between the two TGEV sequences, and between TGEV and IBV and MHV, and conserved structures can be noted that may reflect upon functional domains of the protein.

From a comparative study of the peplomer protein structures, the following points emerge:

(1) The length of the P gene in our sequence is exactly the same as that reported by Rasschaert and Laude (1987), however, 36 bases of the 4341 total are different. Although the origin of the viruses used in both laboratories is the same, we do not know the passage histories. In our laboratory the virus had been plaque purified twice and passaged no more than 8 times prior to cDNA cloning. The passage history prior to our acquisition from Dr. E. Bohl (Brian et al., 1980) is unknown. Likewise we do not know the number of passages the virus had undergone prior to cDNA cloning in Dr. Laude's laboratory. Judging from the high frequency of point mutations that can be expected to survive during RNA virus replication, 10^{-4} changes for every base replication event (Holland et al., 1982), the virus would have to have undergone a theoretical minimum of approximately 80 replications. Considering that TGEV undergoes several rounds of replication during a single passage, only a few passages would be

required to generate the variance we observe between the Rasschert sequence and ours. It is also possible that a copying error could have occurred during reverse transcription in the preparation of the cDNA clones. The 36 bases that differ appear not to be clustered in any particular portion of the gene. The 36 base difference gives rise to 26 different amino acids. It is surprising that only 3 of these are conservative changes, i.e., changes to amino acids having similar properties.

Both our sequence and that of Rasschaert and Laude demonstrate that the partial sequence information previously reported by Hu et al. (1984) for the carboxy terminus of the TGEV genome is in error. Hu et al. reported that for cDNA clones of the Miller strain of TGEV the sequence GCCATGA occurs at the 3' terminus of the P gene. Our data shows that the sequence GCCTAGA occurs instead at 3.8 kb from the initiation codon.

(2) The TGEV P protein is larger than the MHV-JHM P protein by 212 amino acids, and larger than the IBV P protein, either the Beandette or M41 strains, by 285 amino acids. From a dot matrix homology search (Figs. 4.5 and 4.6), homology regions of 6 amino acids or greater showing 60% perfect homology or more were found only in the carboxy one-half of the protein. When the carboxy terminal half of the three proteins were closely aligned by these homology regions, 5 short regions of especially high homology could be found (Fig. 4.7). As a consequence of this close carboxy-terminal alignment, the differences in length of the P proteins among the three viruses must occur in the amino terminal portion of the proteins. Because of

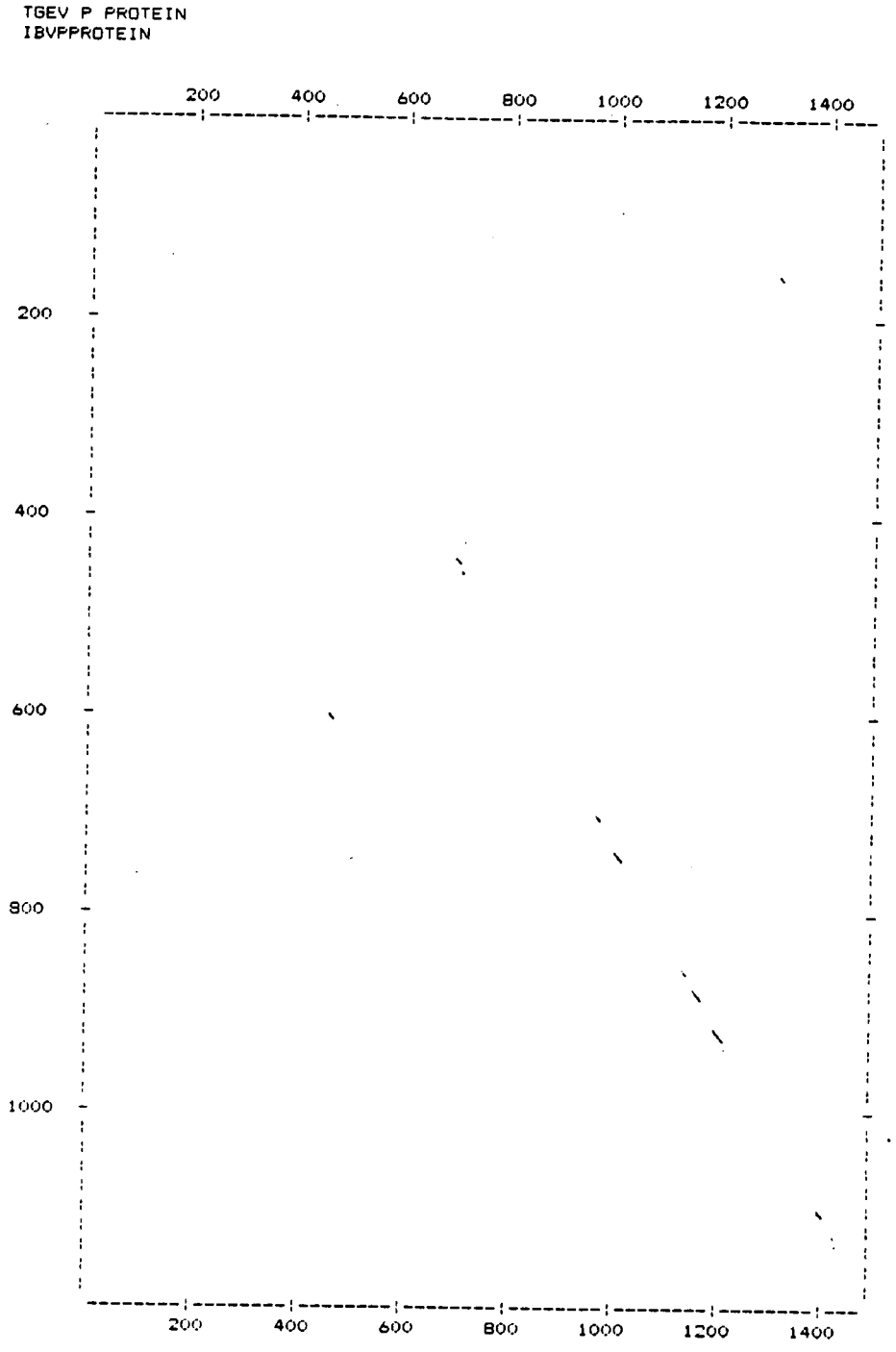


Fig. 4.5 Diagon plot comparing amino acid homologies of TGEV and IBV peplomer proteins. MINMATCH is set at 6 and MINPER is set at 75.

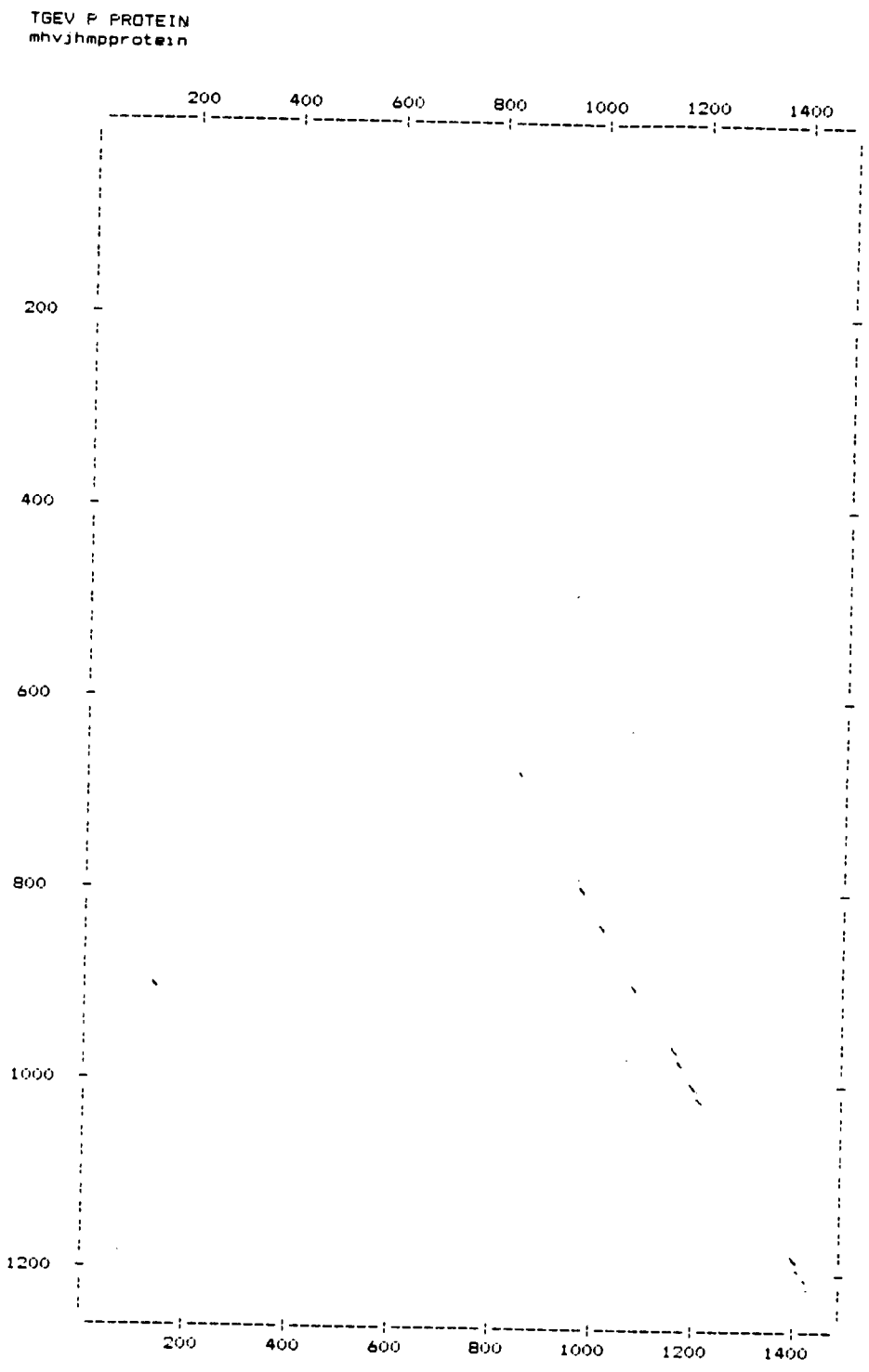


Fig. 4.6 Diagon plot comparing amino acid homologies of TGEV and MHV-JHM peplomer proteins. MINMATCH is set at 6 and MINPER is set at 75.

Region	Virus		Amino Acid Sequences	% homology with TGEV
1	TGEV	953	R K Y R S A I E D L L	100
	IBV	688	R K R S L I E D L L	81
	MHV	778	R G R S A I E D L L	81
2	TGEV	990	D L A C A Q Y Y N G I M V L P G V	100
	IBV	725	D L A C A R E Y V G L L V L P P I	70
	MHV	813	D L L C V Q S T N G I K V L P P V	60
3	TGEV	1144	A Q V D R L I T G R L T A L	100
	IBV	864	A Q V D R L I T G R L S S L	86
	MHV	952	A Q I D R L I N G R L T A L	93
4	TGEV	1181	K V N E C V R S Q S Q R	100
	IBV	1001	K I N E C V K S Q S I R	83
	MHV	989	K V N E C V K S Q T T R	83
5	TGEV	1384	Y V K W P W Y V W L L V G L	100
	IBV	1089	Y I K W P W Y V W L A I A F	64
	MHV	1172	Y V K W P W Y V W L L I G L	92

Fig. 4.7 Amino acid sequence homologies among TGEV, MHV and IBV for five conserved regions in the carboxy terminal region of the peplomer gene. The number in lane 3 indicates the first amino acid position. Identical amino acids are boxed in. The IBV sequences were obtained from Boursnell et al., (1985) and MHV were taken from Schmidt et al., (1987).

differences in overall lengths of the P proteins it is difficult to obtain a quantitative assessment of the degree of similarities over the whole protein length. If we align the carboxy terminal one-half of the sequence, however, we can obtain a sequence homology of 33% with the MHV-JHM protein and 32% with the IBV-Beandette protein for the carboxy terminal-half of the protein. In addition to the 5 regions of high sequence homology, there exist other regions of conserved structure. For example, in all three viruses there exists a C-rich region on the carboxy side of the putative anchor sequence that may play a role in acylation since acylation of the vesicular stomatitis virus G protein is believed to occur at cysteine residues in the vicinity of the hydrophobic transmembrane domain (Rose et al., 1984; McGee et al., 1984). Regions of high sequence conservation among these antigenically diverged viruses suggest there are important functions associated with them and the functions probably exert selective pressures for their retention during evolution.

(3) Much work remains to be done to correlate the tertiary structure of the peplomer protein to the primary structure. It is not known what portion of the molecule makes up the "knob" and which makes up the "stem." Likewise it is not know where the neutralizing epitopes reside. So far the literature reveals a discrepancy. Monoclonal antibodies to the S_1 segment of the IBV P protein (amino terminal half) are neutralizing (Niesters et al., 1986) whereas monoclonal antibodies to the S_2 segment of the MHV P protein (carboxy terminal portion) are neutralizing (Makino et al., 1987). This is difficult to reconcile in light of a high conservation of primary

sequences in a large portion of the proteins. Studies of this kind can now be directly approached with all three viruses since discrete portions of the gene can now be expressed independently. By using subclones of the TGEV P gene, for example, vaccinia virus recombinants could be made, antibody produced, and neutralizing epitopes examined by studying the neutralizing affects of the antiserum on wild type virus.

Finally, an examination of the secondary structure of the genome may suggest features important in the control of transcription. When the P gene was examined for stable stem and loop structures using a program that allows a maximum distance of 1000 bases for the loop, greater than 1,290 possible stem and loop structures were identified. Of these, only 28 have a free energy value of equal to or less than -10 K cal/mole. They range from -10 to -22 K cal/mole. Interestingly, one of the more stable structures begins at base 1 and ends at base 52, just two bases preceding the ACTAAAC intergenic sequence in front of the P gene (Fig. 4.8). Since a similar structure is at or near the intergenic sequences of the M and N genes for MHV and IBV (unpublished data) it may play a role in leader-primed transcription, perhaps by providing accessibility to the RNA polymerase binding site.

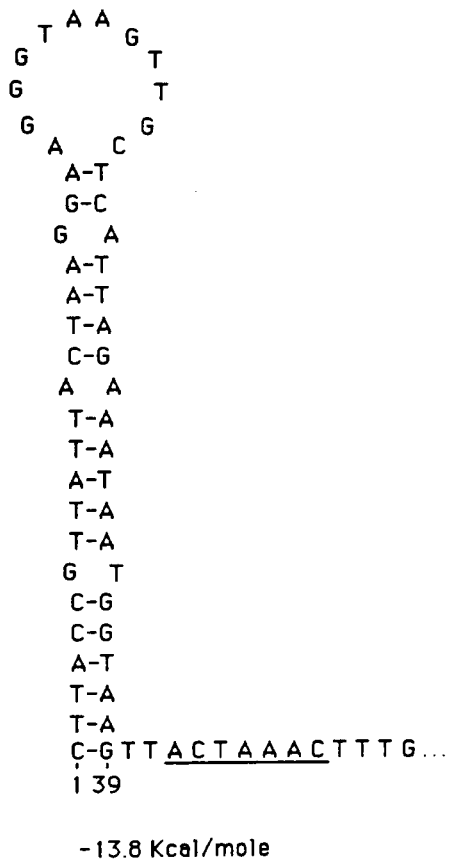


Fig. 4.8 Structure of potential stable stem-and-loop in the genome at the site of P transcript initiation. The conserved intergenic sequence just preceeding the P initiation codon is underlined.

CHAPTER 5

SUBCLONING OF THE HYDROPHOBIC PROTEIN GENE OF THE PORCINE TRANSMISSIBLE GASTROENTERITIS CORONAVIRUS INTO A PROKARYOTIC EXPRESSION VECTOR

Introduction

The total number of genes encoded by the coronavirus genome is not known. As a first approximation the number of genes has been estimated from the number of coronavirus-specific mRNA molecules that have been isolated from infected cells. For the avian infectious bronchitis virus (IBV) the total number of polyadenylated RNA molecules is 6 (Stern and Kennedy, 1980b) and for the mouse hepatitis virus (MHV) the total number is 7 (Spaan et al., 1981). Because only the 5' end of each molecule is translatable, it is estimated that the total nulmber of viral proteins to be 6 for IBV and 7 for MHV (Spaan et al., 1982).

Our laboratory has found the total number of polyadenylated RNA species in TGEV-infected cells to be 10. Their sizes and the proteins which they encode (if known) are as follows: 6.8×10^6 (~ 20 Kb), apparently polymerase; 6.2×10^6 (~18.1 Kb); 3.15×10^6 (~9.2 Kb), peplomer protein; 1.4×10^6 (~4.1 Kb); 1.05×10^6 (~3 Kb); 0.94×10^6 (~2,756 b) matrix protein; 0.66×10^6 (~ 1,935 b), nucleocapsid protein; 0.39×10^6 (~ 1,143 b); 0.34×10^6 (~ 997 b); and 0.24×10^6 (~ 703 b). Since we have fully sequenced the 3' 2601 bases of the TGEV genome and know that the matrix protein gene starts at base 2,564 from the 3' end (Chapter 3 of this dissertation), we know that our size estimates for

the mRNA species are quite accurate. If the 2,564 base sequence has added to it a 100 base poly (A) tail and a 70 base leader primer, then it becomes 2,730 bases, close to the 2,700 bases estimated by gel electrophoresis for the matrix protein mRNA species (Dennis and Brian, 1982).

What remains a puzzle is the origin and identity of the polyadenylated mRNA species of 1,143, 997, and 703 bases in length. If they are real (functional) mRNA species, and if the leader-primer mode of coronavirus mRNA synthesis is the only mechanism by which coronavirus mRNAs are generated, then they must come from within the 3' 2000 bases of the TGEV genome. From our sequence analysis of the 3' end of the TGEV genome we find an open reading frame beginning 523 bases from the 3' end of the genome that would encode a 78 amino acid hydrophobic protein of 9,101 mol. wt. (Kapke and Brian, 1986). The open reading frame is preceded by a ACTAAAC primer binding-like intergenic sequence (9 bases upstream) and the 7 base sequence GAGAUGC at the initiation site on the hydrophobic protein is a favored pattern among enkaryotic initiation sequences (Kozak, 1983). These suggest the protein is real and that one of the 3 small polyadenylated molecules serves as its message.

In this study we begin to determine if the 234 base open reading frame is real gene by subcloning it into a prokaryotic expression vector. Once made this protein can then be used to generate an antibody that can identify its counterpart in the infected cells. The subcloned sequence can then also be used as a probe to study the identity of the three small polyadenylated RNA species.

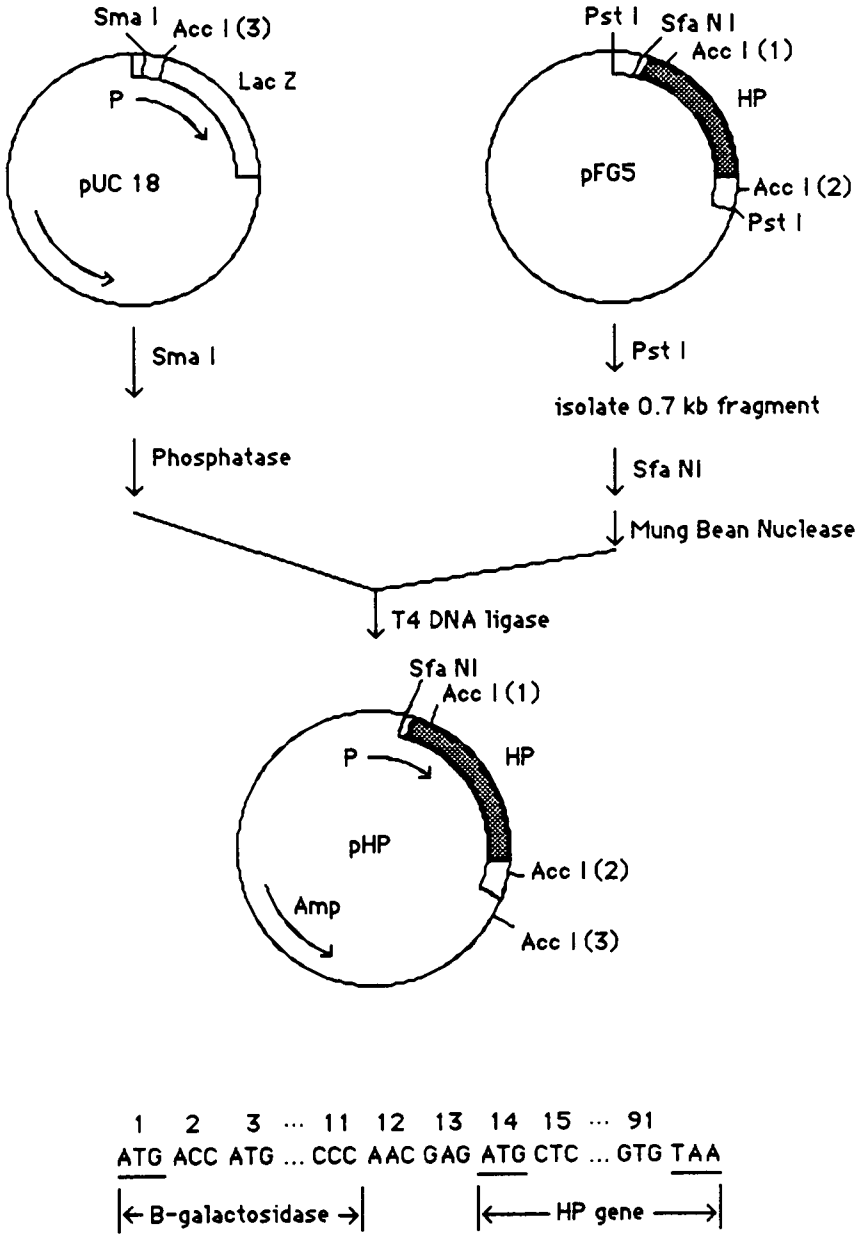
Materials and Methods

Construction of the Hypothetical Gene-Expression Vector Recombinant

Clone FG5 (Kapke and Brian, 1986) was grown in L-B medium and plasmid DNA was purified by the alkaline lysis procedure (Maniatis, 1982). DNA was digested with PstI and the 0.7 Kb PstI DNA fragment was isolated by DEAE-ion exchange paper (NA45) (Schleicher and Schuell). The isolated 0.7 Kb fragment was further restricted with SFNaI, which cuts 6 bases upstream from the AUG initiation codon of the hypothetical hydrophobic protein gene (see Fig. 5.1), and blunt ended with mung bean nuclease. The blunt-ended fragment was ligated with the SmaI-restricted, dephosphosylated pUC18 expression vector (Celeste et al., 1985). The expression recombinant clones were selected as white colonies on YT agar plates that contained 100 ug ampicillin per ml, 1 mM IPTG, 0.004% X-gal. Restriction endonuclease enzymes, T4 DNA ligase and mung bean nuclease were obtained from New England Biolabs.

Prokaryotic DNA-Directed Translation

Plasmid DNA was purified by the alkaline lysis method and CsCl centrifugation (Maniatis, 1982). The *in vitro* translation reaction was carried out using a protocol recommended by the commercial supplier of the prokaryotic DNA-directed translation kit (Amersham). For a reaction, the following were added together to make a premixture: 7.5 ul supplement solution, 3 ul amino acid mixture minus methionine, 2 ul L - [^{35}S] methionine (> 800 Ci/m mole, New England Nuclear) and approximately 2.5 ul DNA. The reaction was started by adding 5 ul



1 2 3 ··· 11 12 13 14 15 ··· 91
ATG ACC ATG ... CCC AAC GAG ATG CTC ... GTG TAA
|← B-galactosidase →| |←—— HP gene ——→|

Fig. 5.1 Construction of expression recombinant for hypothetical gene. For details see Materials and Methods.

cell extract and the reaction was continued at 37^{0} for 60 minutes. After incubation, 5 ul methionine chase solution was added to each tube and the mixture was incubated at 37^{0}C for another 5 minutes. The reaction was terminated by placing it on ice and the sample was heated 100^{0}C for 2 minutes in Leammli sample treatment buffer (Leammli, 1970). SDS polyacrylamide gel electrophoresis was performed as described by Leammli (1970).

Expression in E. coli and Immunoblotting

Bacteria were grown in L-B medium to one OD unit at A550. From this, one ml of culture was used to inoculate 2 ml of SOB medium (Hanahan, 1983) and the culture was grown for another 20 minutes before adding the inducer. Bacteria were induced for 30 minutes (or overnight) by adding 30 microliters 100 mM IPTG. Cells were pelleted and washed with 50 mM glucose, 1 mM Tris HCl, pH 8.0, 1 mM EDTA. Fifteen ul 2X Laemmli treatment buffer (0.0625 M Tris-HCl, pH 6.8, 4% SDS, 10% glycerol, 10% mercaptoethanol) was added to the cell pellet, mixed, and the mixture was heated at 100^{0}C for 5 minutes before loading onto the SDS- polyacrylamide gel.

For immunoblotting, a modified method of Towbin et al. (1979) was used (Hogue et al., 1985). Blots were probed with porcine hyperimmune serum against TGEV.

Results

Construction of the Hydrophobic Protein Gene-Expression Vector Recombinant

Recombinant clones carrying the hydrophobic protein gene were isolated as white colonies on X-gal plates. Plasmid DNA was purified

by the alkaline lysis method (Maniatis, 1982), and digested with AccI. Since there are two AccI sites at positions 62 and 297 in the 0.5 Kb SFNaI-PstI DNA fragment, and one in the multiple cloning region of the pUC18 vector, clones having the proper orientation for generating functional mRNA from the lac promoter on the HP gene would generate fragments of 235 bp, 251 bp, and 2700 bp (Fig. 5.2). Clones having the incorrect orientation would generate fragments of 75 bp, 235 bp and 3000 bp. Four clones having the correct orientation (pHP1, pHP11, pHP13, pHP14) and one having an incorrect orientation clone (pHP7) were identified.

Expression of the Hydrophobic Protein Gene

Figure 5.3 shows the results of proteins translated by a DNA-mediated translation assay. One protein band of approximately 30 Kd in size was identified in lanes containing DNA from the pHP1, pHP7, pHP11 and pHP14 clones. No protein band of this size was observed in the control lanes containing no added DNA, JM83 DNA, or pUC18 DNA.

Discussion

The results of the *in vitro* DNA-mediated translation assay reveal two important discrepancies requiring that further work be done to unequivocally establish that the 30 Kd protein product is the lac-inducible fusion protein we seek. (1) The control plasmid pUC18 did not produce the expected 10 Kd peptide product of the Lac gene. (2) the product observed is approximately 30 Kd, whereas the Lac Z-hydrophobic fusion protein should be approximately 10 Kd. A third discrepancy is that the pHP13 clone in the correct orientation did not

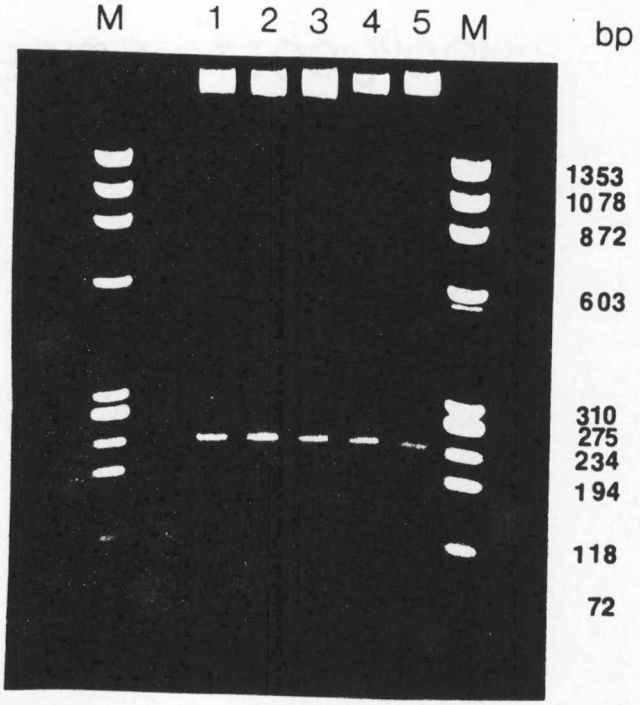


Fig. 5.2 Agarose gel electrophoresis of expression clones restricted with AccI. Lane 1, pHP14. Lane 2, pHP13. Lane 3, pHP11. Lane 4, pHP7. Lane 5, pPH1. Lane M, size marker of øx 174 DNA restricted with Hae III.

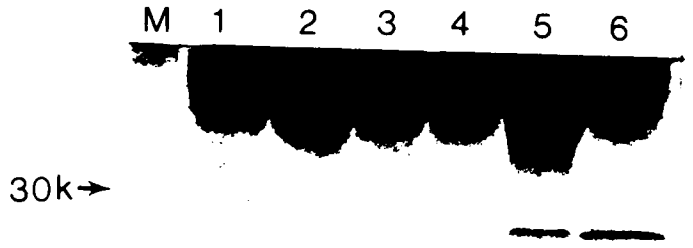


Fig. 5.3 SDS-PAGE of protein translated by a DNA mediated translation assay. Lane 1 blank. Lane 2, pUC18. Lane 3, pHP13. Lane 4, pHP14. Lane 5, pHP14. Lane 6, pHP7. M, mol. wt. marker.

generate a protein product, whereas the incorrectly oriented clone pHP7 did.

Two other experimental approaches were attempted to identify a fusion gene product. Both approaches used an immunodetection method that employed hyperimmune porcine anti-TGEV serum that is known to detect at least 12 protein species between 8 and 22 Kd in TGEV infected cells (unpublished data). On the first approach, anti-TGEV serum was used to screen colonies containing the hydrophobic protein gene subclones in a colony hydridization method using 125 I-labeled Staph A protein (Helfman et al., 1983). By this method no positive colonies were observed (data not shown). In the second approach, E. coli isolates carrying clones of the hydrophobic protein known to be in either the correct or incorrect orientation for transcription, were induced, and lysates were electrophoresed on SDS polyacrylamide gels, transferred to nitrocellulose membranes, and immunoblotted using 125 I Staph A for radiodetection. The result from clones pHP1, pHP11, pHP13, pHP14 and pHP7 and control bacteria were all the same. That is a 30 K protein was detected in all lanes, and no 10 K protein of expected sizes was found (data not shown). Interestingly, 5 of 30 other clones gave immuno reactive proteins of distinctly different size (approximately 20 Kd) that may be abberently expressed fusion product (data not shown).

One explanation of the 30 K protein is that it is a read-through product of B-galactosidase:hydrophobic protein fusion gene. It could also be that what we have subcloned produces an out-of-frame open reading frame at the junction region of this fusion gene. A detailed

DNA sequencing experiment is needed to confirm the junction sequences of this fusion gene.

A third possibility is that the fusion protein is unstable in the E. coli. Therefore, a more sensitive in vitro system is needed to express this hypothetical gene. Currently, we are subcloning the hydrophobic protein gene into pGEM vector (Promega), which can be used as an in vitro transcription system. mRNA can then be synthesized in large quantities and translated in an in vitro system and the product identified by immunodetection. By this approach we should be able to answer the question whether the hypothetical gene is real and lay the groundwork for the establishing the function of the protein.

LITERATURE CITED

LITERATURE CITED

Armstrong, J., H. Niemann, S. Smeekens, P. Rottier, and G. Warren. 1984. Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. Nature 308: 751-752.

Armstrong, J., S. Smeekens, and P. Rottier. 1983. Sequence of the nucleocapsid gene from murine coronavirus MHV-A59. Nuc. Acids Res. 11: 833-891.

Almeida, J. D., D. M. Berry, C. H. Cunningham, D. Hamre, M. S. Hofstad, L. Mullucci, K. McIntosh, D. A. J. Tyrrell. 1968. Coronavirusis. Nature (Lond.) 220: 650.

Becker, W. B., K. McIntosh, J. H. Dees, and R. M. Chanock. 1967. Morphogenesis of avian infectious brunchitis virus and a related Human Virus (Strain 229E) 5" 1019-1027.

Bennink, J. R., J. W. Yewdell, G. L. Smith, C. Moller, and B. Moss. 1984. Recombinant vaccinia virus primes and stimulates influenza hemagglutinin-specific cytotoxin T cells. Nature 311: 578.

Bohl, E. H., and L. Saif. 1975. Passive immunity in transmissible gastroenteritis of swine: Immunoglobulin characteristics of antibodies in milk after inoculating virus by different routes. Infection and Immunity 11: 23-32.

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

Boursnell, M. E. G., T. D. K. Brown, I. J. Foulds, P. F. Green, F. M. Tomley, and M. M. Binns. 1987. Completion of the sequence of the genoms of the coronavirus avian infectius bronchitis virus. J. Gen. Virol. 68: 57-77.

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

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

Brian, D. A., D. E. Dennis, and J. S. Guy. 1980. Genome of porcine transmissible gastroenteritis virus. J. Virol. 34: 410-415.

Brian, D. A., B. G. Hogue, W. Lapps, B. J. Potts, and P. A. Kapke. 1983. Comparative structure of coronaviruses. In "Proceedings from the Fourth International Symposium on Neonatal Diarrhea." (Ed. S. D. Acres) University of Saskatoon, Saskatoon, Saskatchewan, Canada. pp. 100-115.

Brown, T. D. K., and M. E. G. Boursnell. 1984. Avian infectious bronchitis virus genomic RNA contains sequence homologies at the intergenic boundaries. Virus Res. 1: 15-24.

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

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

Cavanagh, D., M. M. Binns, M. E. G. Boursnell, and T. D. Brown. 1984. Genetically engineered vaccine against avian infectious bronchitis virus with the advantages of current live and killed vaccines. In "Modern approaches to vaccines." Edited by R. M. Chanock and R. A. Lerner. Cold Spring Harbor Laboratory, pp. 215-218.

Celeste, Y. P., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-119.

Cremer, M. A., M. Mackett, C. Wohlenberg, A. L. Notkins, and B. Moss. 1985. Vaccinia virus recombinant expresing herpes simplex type 1 glyesprotein D prevents latent herpes in mice. Science 228: 737-739.

Efstratiadis, A., F. C. Kafatos, A. M. Maxam, and T. Maniatis. 1976. Enzymatic in vitro synthesis of globin genes. Cell 7: 279-288.

Garwes, D. J., M. H. Lucas, D. A. Higgins, B. V. Pike, and S. F. Cartwright. 1979/87. Antigenicity of structural components from porcine transmissible gastro-enteritis virus. Vet Microbiol. 3: 179-190.

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

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.

Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. Gene 25: 263-269.

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

Helfman, D. M., J. R. Feramisco, J. C. Fiddes, G. P. Thomas, and S. H. Hughes. 1983. Identification of clones that encode chicken tropomyosin by direct immunological screening of a cDNA expression library. Proc. Natl. Acad. Sci. 80: 31-35.

Hogue, B. G., B. King, and D. A. Brian. 1984. Antigenic relationships among proteins of bovine coronavirus, human regulatory coronavirus OC43, and mouse hepatitis coronavirus A59. J. Virol. 51: 384-388.

Holmes, K. V., M. F. Frana, S. G. Robbins, and L. S. Sturman. 1984. Coronavirus maturations. Adv. Exp. Med. and Biol. 173: 37-52.

Holland, J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. Vandepal. 1982. Rapid evolution of RNA gonomes. Science 215: 1577-1586.

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.

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

Horzinek, M. C., H. Lutz, and N. C. Pederson. 1982. Antigenic relationships among homologous structural polypeptides of porcine, felines, and canine carnaviruses. Infect. Immun. 37: 1148-115.

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

Hu, S., J. Bruszewski, R. Smalling, and J. K. Browne. 1986. Studies of TGEV spike protein gp195 expressed in E. coli and by a TGE-vaccinia virus recombinant. In "Immunobiology of Proteins and Peptides" III. M. Zouhair-Atassi and H. L. Cacvach (Eds). Plenum Publishing Corporation, pp. 63-82.

Hubbard, S. C., and R. J. Ivatt. 1981. Synthesis and processing of asparagine-linked oligosaccharides. Ann. Rev. Biochem. 50: 555-583.

Jacobs, L., W. J. M. Spann, 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.

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.

Jimenez, G., I. Correa, M. Melgosa, M. J. Bullido, and L. Enjuanes. 1986. Critical epitopes in transmissible gastroenteritis virus neutralization. J. of Virology 60: 131-139.

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

Kemeny, L. J. 1976. Antibody responses in pigs inoculated with transmissible gastroenteritis virus and cross reactions among ten isolates. Can. J. Comp. Med. 40: 209-214.

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

Krug, R. 1985. The role of RNA priming in viral and trypanosomal mRNA synthesis. Cell. 41: 651-652.

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

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

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

Laude, H., J. M. Chapsal, J. Geffi, S. Lobiau, 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.

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

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

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

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

Mackett, M., T. Y. Yilma, J. A. Rose, B. Moss. 1985. Vaccinia virus recombinants express vesiculum stomatitis virus genes and protectively immunize mice and cattle. Science 227: 433-435.

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

Makino, S., J. G. Keck, S. A. Stahlman, and M. M. C. Lai. 1986a. High frequency RNA recombination of murine coronaviruses. J. Virol. 57: 729-737.

Makino, S., S. A. Stohlman, and M. M. C. Lai. 1986b. Leader sequences of murine coronavirus mRNAs can be freely reassorted: Evidence for the role of free leader RNA in transcription. Proc. Natl. Acad. Sci. USA 83: 4204-4208.

Makino, S., J. O. Fleming, J. G. Keck, S. A. Stahlman, and M. M. C. Lai. 1987. RNA recombination of coronaviruses: Localization of neutralizing epitopes and neutropathogenic determinants on the carboxyl terminus of peplomers. Proc. Natl. Acad. Sci. USA 84: 6567-6571.

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

Maxam, A. M., and Gilbert, W. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. In "Methods in Enzymology" (L. Grossman and K. Moldave, eds.), vol 65, pp. 499-560. Academic Press, Orlando, FL.

Moreau, K. 1982. Master's Thesis, The University of Tennessee, Knoxville.

Moss, B. 1985. Replication of poxviruses. In Field, B. N., Chanock, R. M., Roizman, B. (eds.) Virology. Raven Press, New York, pp. 685-703.

Paoletti, E., B. R. Lipinskas, C. Samsanoff, S. Mercer, and D. Panicali. 1984. Construction of live vaccines using genetically engineered proxviruses: Biological activity of vaccinia virus recombinants expressing the hepatitis B virus surface antigen and the herpes simplex virus glycoprotein D. Proc. Natl. Acad. Sci. USA 81: 193-197.

Panicali, D., and E. Paoletti. 1982. Construction of proxviruses as cloning vectors; Insertion of the thymidine kinase gene of herpes simplex virus into the DNA of infectious vaccinia virus. Proc. Natl. Acad. Sci. USA 79: 4927-4931.

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

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

Pensaert, M., P. Callebaut, and J. Vergate. 1986. Isolation of a porcine respiratory,, non-enteric coronavirus related to transmissible gastroenteritis. The Veterinary Quartery. 8: 257-261.

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

Reynolds, D. J., D. J. Garwes, and S. Lucey. 1980. Differentiation of carrine coronavims and dorcine transmissible gastroenteritis virus by neutralization with canine, porcine and feline sera. Vt. Microbial. 5: 283-290.

Rose, J. K., and C. Gallione. 1981. Nucleotide sequences of the mRNAs encoding the VSV G and M proteins as determined from cDNA clones containing the complete coding regions. J. Virol. 39: 519-528.

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

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

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

Rottier, J. M., G. W. Welling, S. Welling-Webster, H. G. M. Niesters, J. A. Lanstra, and B. A. M. Van der Zeijst. 1986. Predicted membrane topology of the coronavirus protein E1. Biochemistry 25: 1335-1339.

Rougeon, F., and Mach, B. Stepwise biosynthesis in virto of globin genes from globin mRNA by DNA polymerase of avian myeloblastosis virus. Proc. Natl. Acad. Sci. USA 73 (1976) 3418-3422.

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

Saif, L. J., and E. H. Bahl. 1986. Transmissible gastroentestis. In A. D. Lemar, et al., Eds. Disease of Swine (6th Ed.) Iowa State University, Ames, IA.

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

Shockley, L. J., P. A. Kapke, W. Lapps, D. A. Brian, L. N. D. Potgieter, and R. Woods. 1987. Diagnosis of porcine and bovine enteric coronavirus infections using cloned cDNA probes. J. of clinical microbiology 25: 1591-1596.

Siddell, S. G. 1982. Coronavirus JHM: Trypic peptide fingerprinting of virion proteins and intracellular polypeptides. J. Gen. Virol. 62: 259-269.

Siddell, S. 1983. Coronavirus JHM: Coding assignments of subjgenomic mRNAs. J. Gen. Virol. 64: 113-125.

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

Siddell, S. G., A. Banthel, and V. ter Meulen. 1981. Coronavirus JHM. A virion-associated protein kinase. J. Gen. Virol. 52: 235-243.

Siddell, S. G., H. Ege, A. Banthel, and V. ter Meulen. 1980a. Coronavirus JHM. Cell-free synthesis of structural protein p60. J. Virol. 33: 10-17.

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

Smith, G. L., M. Mackett, and B. Moss. 1983. Construction and characterization of an infectious vaccinia virus recombinant that expresses the influenza virus infection in hamsters. Proc. Natl. Acad. Sci. USA 80: 7155-7159.

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

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

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

Stern, D. F., and B. M. Sefton. 1982a. Coronavirus proteins: Biogenesis of avian infectious bronchitis virus virion proteins. J. Virol. 44: 794-803.

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

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

Stohlman, S. A., and L. P. Weiner. 1981. Chronic untral nervous system demyclination in mice after JHM virus infection. Neurology 33: 38-44.

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.

Tooze, J., S. Tooze, and G. Warren. 1984. Replication of coronavirus MHV-A59 in sac$^-$ cells: Determination of the first site of budding of progeny virions. Eur. J. Cell. Biol. 33: 281-294.

Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. Proc. Natl. Acad. Sci. 76: 4350-4354.

von Heijne, G. 1986. Towards a comparative anatomy of N-terminal topogenic protein sequences. J. Mol. Biol. 189: 239-242.

von Heijne, G. 1986. A new method for prediciting singal sequence cleavage sites. Nucl. Acids Res. 14: 4683-4690.

Wan, S. K., S. Ram, L. T. Saif, and D. A. Weilnau. 1986. Cell-mediated immune response of pigs challenged with virulent and attenuated transmissible gastroenteritis virus. 67th Ann. Meeting Conf. Res. Workers in Animal Disease. p. 21.

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

Wesley, R. D., and R. D. Woods. 1986. Identification of a 17,000 molecular weight antigenic peptide in transmissible gastroenteritis virus. J. Gen. Virol. 67: 1419-1425.

Yewdell, J. W., J. R. Benrink, G. L. Smith, and B. Moss. 1985. Influenza A virus nucleoprotein is a major target antigen for cross-reactive anti-influenza A virus gytotoxic T lymphocytes. Proc. Natl. Acad. Sci. USA 82: 1785-1789.

APPENDIXES

APPENDIX A

OLIGONUCLEOTIDE PURIFICATION

This procedure was obtained from Dr. Clayton Naeve, St. Jude Hospital, Memphis, Tennessee. Oligodeoxynucleotides purifed by this procedure can be used for primer extension studies (sequencing) as well as for priming cDNA synthesis from specific sites. The method is used for purifying 50-80 A_{260} units (1.85mg-2.96mg) of crude material and yields usually 5-10 A_{260} units (0.185mg-0.37mg) purified oligodeoxynucleotide.

1. Complete deprotection of 5' end:

   Add 1 ml fresh concentrated NH_4OH to vial.
   Incubate at 55°C for 4-12 h.

2. Removal fo NH_4OH under vacuum.

   Dry down to approximately 200 ul in Speed-Vac.
   Add H_2O to 0.5 ml, dry completely.
   Wash with 100 ul 95% EtOH, dry completely.
   Resuspend in 30 ul H_2O.
   Add 70 ul deionized formamide (no dye).
   Heat to 100°C for 3 min, chill.
   Load gel.

3. Preparative gel (20% acrylamide - 7M urea - TBE buffer):

   Load gel with 1-2 A_{260} units/1 cm well; 1.5mm x 40 cm gel.
   Run until BPB is 20 cm from top (5-6 h at 600-900V). Run sample treatment buffer with dyes in flanking lanes.

4. Elution:

   Visualize the digonucleotide by UV absorbance using a handheld UV lamp (254 nm) over a fluorescence intensifying screen (Whatman PK1F).
   Cut out the dense band, crush with a silicone glass rod then add approximately 3 mls elution buffer (2 M NH4Ac).
   Elute overnight at 37°C on shaker.
   Recover supernatant fluid.
   Dilute 1:4 to make 0.5 M NH4Ac (12-16 ml vol).

5. DEAE column:

Load entire volume onto a 1 ml DEAE cellulose column equilibrated with 0.5 M NH4Ac.
Wash with approximately 4 ml H_2O.
Elute DNA with 2 washes of 1 ml each of 30% (V/V) triethylamine. PH to approximately 9 by bubbling with CO_2.
This final product becomes triethylammonium bicarbonate (TEAB).

6. Desalt:

Speed-Vac to dryness.
Resuspend in 0.5 ml H_2O.
Speed-Vac to dryness.
Store dry at -20°C or resuspend in H_2O to measure A_{260}.

APPENDIX B

THE AMINO ACID COMPOSITION OF THE TGEV MATRIX PROTEIN

TGEV M PROTEIN

The sequence contains 262 amino acids:

Ala	21	(8.0)	Leu	30	(11.5)
Arg	10	(3.8)	Lys	13	(5.0)
Asn	10	(3.8)	Met	9	(3.4)
Asp	7	(2.7)	Phe	12	(4.6)
Cys	8	(3.1)	Pro	7	(2.7)
Gln	5	(1.9)	Ser	23	(8.8)
Glu	9	(3.4)	Thr	15	(5.7)
Gly	17	(6.5)	Trp	10	(3.8)
His	2	(0.8)	Tyr	15	(5.7)
Ile	19	(7.3)	Val	20	(7.6)
End	0	(0.0)			

Acidic	(Asp + Glu)	16	(6.1)
Basic	(Arg + Lys)	23	(8.8)
Aromatic	(Phe + Trp + Tyr)	37	(14.1)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	115	(43.9)

Molecular Weight = 29544.

APPENDIX C

THE AMINO ACID COMPOSITION OF THE TGEV PEPLOMER PROTEIN

TGEV P PROTEIN

The sequence contains 1447 amino acids:

Ala	86	(5.9)	Leu	126	(8.7)
Arg	52	(3.6)	Lys	44	(3.0)
Asn	115	(7.9)	Met	14	(1.0)
Asp	73	(5.0)	Phe	70	(4.8)
Cys	49	(3.4)	Pro	51	(3.5)
Gln	48	(3.3)	Ser	113	(7.8)
Glu	53	(3.7)	Thr	130	(9.0)
Gly	93	(6.4)	Trp	24	(1.7)
His	22	(1.5)	Tyr	68	(4.7)
Ile	88	(6.1)	Val	128	(8.8)
End	0	(0.0)			

Acidic	(Asp + Glu)	126	(8.7)
Basic	(Arg + Lys)	96	(6.6)
Aromatic	(Phe + Trp + Tyr)	162	(11.2)
Hydrophobic	(Aromatic + Ile + Leu + Met + Val)	518	(35.8)

Molecular Weight = 160345.

VITA

Frank Yao-Tsung Tung was born in Taiwan, Republic of China, on February 6, 1958. He received a Bachelor of Science degree in animal science from Tunghai University, Taiwan, in 1980. From 1980 to 1982, Frank enrolled in the China Army ROTC program as medical officer to fulfill his military service.

In 1984 he received a Master of Science degree in Biochemistry at National Yangming Medical College. He began his doctoral program at The University of Tennessee, Knoxville, in the fall of 1984, and in December 1987 he received his doctorate in the Life Sciences interdisciplinary program of Cell, Molecular, and Development Biology. Frank has accepted a postdoctoral position in New England Primate Research Center in Harvard Medical School.