



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

## The Amino-Terminal Signal Peptide on the Porcine Transmissible Gastroenteritis Coronavirus Matrix Protein Is Not an Absolute Requirement for Membrane Translocation and Glycosylation

PAUL A. KAPKE,\* FRANK Y. T. TUNG,† BRENDA G. HOGUE,†<sup>1</sup> DAVID A. BRIAN,†<sup>2</sup>  
ROGER D. WOODS,\* AND RONALD WESLEY\*

\*U.S. Department of Agriculture, Agricultural Research Service, National Animal Disease Center, P.O. Box 70 Ames, Iowa 50010, and  
†Department of Microbiology, The University of Tennessee, Knoxville, Tennessee 37996-0845

Received December 18, 1987; accepted April 6, 1988

cDNA clones mapping within the first 2601 bases of the 3' end of the porcine transmissible gastroenteritis coronavirus (TGEV) genome were sequenced by the method of Maxam and Gilbert and an open reading frame yielding a protein having properties of the matrix (M or E1) protein was identified. It is positioned at the 5' side of the nucleocapsid (N) gene from which it is separated by an intergenic stretch of 12 bases. The deduced M protein comprises 262 amino acids, has a molecular weight of 29,544, is moderately hydrophobic, and has a net charge of +7 at neutral pH. Thirty-four percent of its amino acid sequence is homologous with the M protein of the bovine coronavirus (BCV), 32% with that of the mouse hepatitis coronavirus (MHV), and 19% with that of the avian infectious bronchitis coronavirus (IBV). Judging from alignment with the BCV, MHV, and IBV M proteins, the amino terminus of the TGEV M protein extends 54 amino acids from the virion envelope which compares with only 28 for BCV, 26 for MHV, and 21 for IBV. Eleven of the sixteen amino-terminal amino acids are hydrophobic and the positions of charged amino acids around this sequence suggest that the first 16 amino acids comprise a potentially cleavable signal peptide for membrane insertion. A similar sequence is not found in the M proteins of BCV, MHV, or IBV. When mRNA from infected cells, or RNA prepared by *in vitro* transcription of the reconstructed M gene, was translated *in vitro* in the presence of microsomes, the M protein became translocated and glycosylated. When a protein without the amino-terminal signal peptide was made by translating a truncated version of the M gene transcript, some translocation and glycosylation also occurred suggesting that the amino-terminal signal peptide on the TGEV M protein is not an absolute requirement for membrane translocation. Interestingly, the amino-terminal peptide did not appear to be cleaved during *in vitro* translation in the presence of microsomes suggesting that a step in virion assembly may be required for proper exposure of the cleavage site to the signal peptidase. © 1988 Academic Press, Inc.

### INTRODUCTION

The porcine transmissible gastroenteritis coronavirus (TGEV) comprises three major structural proteins: an internal nucleocapsid phosphoprotein (N) of 43 kDa and two glycosylated envelope proteins, one of 29 kDa (a matrix-like protein, M or E1) and one of 200 kDa (the peplomeric, P, or E2 protein) (Brian *et al.*, 1983; Garwes and Pocock, 1975; Kapke and Brian, 1986; Wesley and Woods, 1986). While the 200-kDa P glycoprotein is demonstrably important in stimulating neutralizing antibody (Garwes *et al.*, 1978), the 29-kDa M glycoprotein may also be important, especially if complement is part of the virus-antibody reaction (Woods *et al.*, 1987).

To investigate the role of individual viral proteins in virus replication and in induction of immunity, we have prepared cDNA clones beginning from the polyadenylated 3' end of the TGEV genome and examined the

sequences of potential genes (Kapke and Brian, 1986). Within the first (3') 2000 bases, we deduced, from an examination of open reading frames, a noncoding region of 276 bases, and genes for a 9101 mol wt hypothetical hydrophobic polypeptide, a 43,426 mol wt nucleocapsid protein, and part of a matrix protein, arranged in that order from the 3' end of the 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 synthesis in the preparation of additional genomic cDNA clones. Several cDNA clones 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 gene 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 infec-

<sup>1</sup> Present address: Department of Microbiology and Immunology, University of California School of Medicine, Los Angeles, CA 90024.

<sup>2</sup> To whom requests for reprints should be addressed.

tious bronchitis virus, but also predicted an unexpected, potentially cleavable amino-terminal signal peptide that makes the TGEV M protein strikingly different. In this study, we have confirmed our preliminary report of the M gene sequence (Kapke *et al.*, 1987), and have examined the behavior of the amino-terminal peptide during synthesis of the M protein.

## MATERIALS AND METHODS

### Cells 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, 1986) except that the synthetic oligodeoxynucleotide 5'TTAGAAGTTTAGTTA3' was used as primer for first-strand cDNA synthesis. The primer was synthesized by the phosphoramidite 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 <sup>32</sup>P by nick-translation.

### DNA sequencing and sequence analyses

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 Microgenie system, October 1986 version (Beckman Instruments, Inc.). For protein homology determinations among the coronavirus M proteins, the "protein alignment" program of the Beckman Microgenie was used with MAXDIST set at 1000. For protein homology searches 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.

### Reconstruction of the M gene and synthesis of transcripts *in vitro*

To reconstruct the full-length M gene, clones FT36 and C4 were digested with *Accl* and the small frag-

ment of FT36 (which contained the first 782 bases of the sequence shown in Fig. 2) and the large fragment of C4 (which contained bases 783 through 934, an oligo(dC) tail of 13 bases, and the rest of the pUC9 sequence) were ligated and used to transform *Escherichia coli* strain 294. The insert was removed from this plasmid using *Pst*I, digested with *Bsp*1286 to remove the 5' 114 bases, and blunt-ended with mung bean nuclease. The fragment now contained 13 bases upstream (5'-ward) from the CTAAAC presumed intergenic sequence (or 22 bases upstream from the presumed M gene start codon) and extended to the end of the 13-base C-tail which begins 9 bases downstream from the M gene stop codon (bases 922–924 in Fig. 2). The blunt-ended fragment was ligated into the *Sma*I site of the pGEM3 vector (Promega Biotec) and the orientation yielding a sense-strand RNA by transcription with SP6 polymerase was chosen. The construct was designated pGEM-M-1. Transcripts were prepared from *Eco*RI-cut plasmid using the SP6 transcription system marketed by Promega Biotec.

To reconstruct the truncated version of the M gene (i.e., the M gene with no N-terminal signal peptide), clone pGEM-M-1 was cut with *Sph*I, which cuts within the multiple cloning region of the pGEM vector and between bases 174 and 175 in Fig. 2, and the resulting large fragment was isolated, religated, and designated pGEM-M-2. Translation of pGEM-M-2 transcripts allows initiation at the second AUG downstream from the CTAAAC intergenic sequence (beginning at base 200 in Fig. 2), i.e., at the fifth amino acid downstream from the potential peptidase cleavage site between glycine and lysine (von Heijne, 1986).

The sequence at the 5' junction of the insert (virus sense) and the vector were confirmed by sequencing for both the pGEM-M-1 and pGEM-M-2 constructs.

### Isolation of viral mRNA

Cells were grown to confluency in 850-cm<sup>2</sup> roller bottles (Falcon) and infected with a multiplicity of infection of approximately 10. At 6 hr p.i., cells were rinsed twice with Earl's balanced salt solution (EBSS), scraped from the bottle, transferred to a 50-ml conical tip polypropylene tube, and pelleted. Cells from five roller bottles yielded a pellet of approximately 5 ml and constituted one batch for RNA extraction. Cells were lysed by the addition of 5 vol (25 ml to a 5-ml cell pellet) of lysis buffer containing 10 mM Tris-HCl (pH 7.0), 10 mM NaCl, 5 mM MgCl<sub>2</sub>, 1% NP-40(v/v) at room temperature followed by vigorous vortexing for 10 sec. Nuclei were removed immediately by centrifugation at 4K × *g*, 5 min, and to the supernatant was added 0.2 vol of 10% SDS(w/v) in water and 4 mg proteinase K crystals. The solution was incubated 30 min at 37° and

extracted with an equal volume of phenol–chloroform–isoamyl alcohol (24:24:1), and RNA was precipitated with 2.2 vol of ethanol after adding 0.1 vol 3 M Na acetate. RNA from a 5-ml cell pellet was dissolved in water and polyadenylated RNA was selected by oligo(dT)-cellulose chromatography using a binding buffer of 0.5 M NaCl, 0.1 M Tris–HCl (pH 7.5), 0.2% SDS (w/v), and an elution buffer of 0.1 M Tris–HCl (pH 7.5), 0.2% SDS (w/v). Polyadenylated RNA was ethanol precipitated with Na acetate, dissolved in 100  $\mu$ l water, ethanol precipitated without salt, dissolved in 100  $\mu$ l water, and 5  $\mu$ l of this solution was used in a 50- $\mu$ l translation reaction.

### *In vitro* translation

*In vitro* translation was done using a wheat germ system (Amersham) in a 50  $\mu$ l reaction volume that contained 25  $\mu$ l wheat germ extract, 1  $\mu$ l 1 mM amino acid solution deficient in methionine (Promega Biotec), 3  $\mu$ l 1 M KAc (to make a final K<sup>+</sup> concentration of 1.17 mM), 2  $\mu$ l RNasin (Promega Biotec), 4  $\mu$ l microsomes (Amersham, Promega Biotec, or as a gift from Dr. Peter Walter, University of California School of Medicine, San Francisco) or 4  $\mu$ l microsome blank solution when microsomes were left out, 5  $\mu$ l 300  $\mu$ M octanoyl–asparagine–leucine–threonine or 5  $\mu$ l water when the tripeptide was left out, 5  $\mu$ l [<sup>35</sup>S]-methionine (>800 Ci/mmol, New England Nuclear), 5  $\mu$ l RNA. Octanoyl–asparagine–leucine–threonine, a competitive inhibitor of N-linked glycosylation (Lau *et al.*, 1983), was a kind gift from Dr. Fred Naider, City University of New York, and was prepared as a 300  $\mu$ M stock in an aqueous solution containing 25% dimethyl sulfoxide. Translations were done for 1 hr at 25°. Sodium carbonate treatment of microsomes followed the procedure of Fujiki *et al.* (1982). Deglycosylation of translation products was done with N-glycanase (Genzyme Corp.) or with endoglycosidase H (ICN) using methods recommended by the manufacturers. Immunoprecipitates were prepared as described by Anderson and Blobel (1983) except that iodoacetamide was not used to block SH groups prior to electrophoresis. Five microliters porcine hyperimmune TGEV-specific serum was used per 50  $\mu$ l translation volume, and protein A–Sepharose CL-4B (Pharmacia) was used to adsorb the immunoprecipitates. Porcine hyperimmune anti-TGEV serum was produced in a specific pathogen-free gilt (gilt 53) and was a kind gift from Dr. Lorant Kemeny, National Animal Disease Center (Ames, IA) Kemeny, 1976). Preprolactin mRNA was generated with SP6 polymerase from cloned cDNA kindly given to us by Drs. William Hansen and Peter Walter, University of California School of Medicine (San Francisco, CA)

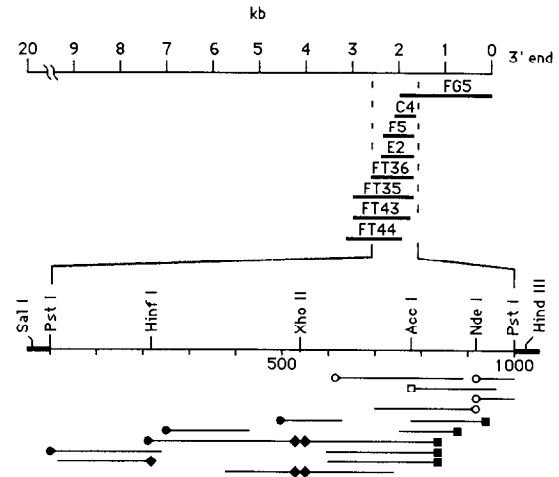


FIG. 1. Sequencing strategy used to derive the TGEV M gene sequence. cDNA clones FG5, C4, F5, E2, FT36, FT35, and FT44 were cloned into the *Pst*I site of vector pUC9 and were all found to be in the same orientation with respect to the virus genomic RNA illustrated at the top of the figure. That is, the 3' end of the insert is near the *Hind*III site in the multiple cloning region, and the 5' end is near the *Sal*I site in the multiple cloning region. FT43 was likewise cloned but was found to be in the opposite orientation. Nucleotide position 1 on the restriction map sequence is the first base at the 5' end (virus-sense) of the FT36 insert. ○ and □ indicate sites labeled on fragments of clone FG5 at the 3' end of DNA with reverse transcriptase and at the 5' end with polynucleotide kinase, respectively. ● 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 the *Hind*III site in the multiple cloning region of clones C4, F5, E2, FT36, and FT35. ◆ indicates 3' end-labeling with reverse transcriptase at the *Xho*II site in clones E2 and FT43, or at the *Hinf*I site on clone FT44.

and  $\beta$ -lactamase mRNA was obtained from Promega Biotec.

### *In vivo* labeling

For labeling intracellular M protein, ST cells in 60-mm plastic petri dishes were infected with a multiplicity of infection of approximately 10, incubated 1 hr, and refed, after rinsing, with 10 ml per dish of minimum essential medium containing 5% normal methionine concentration, 10% fetal calf serum (Sterile Systems), and 200  $\mu$ Ci [<sup>35</sup>S]methionine (Translabel, ICN). Where indicated, tunicamycin (Sigma) at a final concentration of 2  $\mu$ g/ml was included in the medium used for re-feeding. At 6 hr p.i., cells were rinsed with EBSS, scraped into a 15-ml conical tip polypropylene tube and pelleted, and lysate was prepared by adding 0.5 ml phosphate-buffered saline, 1% NP-40, 10 units Aprotinin (Sigma)/ml, and incubating the mixture at 25° for 30 min with frequent vortexing. Nuclei and cell debris were removed by centrifugation at 13,000 *g* for 5 min and 50  $\mu$ l cell lysate supernatant was used in an immunoprecipitation reaction as described above for

```

      30          60          90          120
CTATACATGGTGTGTTGCAATTTAGGAAGGACAGTTATTATTGTTCCAGCGCAACATGCTTACGATGCCTATAAGAATTTATGCGAATTAAGCATACAACCCCGATGGAGCACTCCTT
      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
GCTTGAACTAAACAAAATGAAGATTTGTTAATATTAGCGTGTGTGATTGCATGCGCATGTGGAGAACGCTATTGTGCTATGAAATCCGATACAGATTGTTCATGTCGCAATAGTACAGC
  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
GTCTGATTGTGAGTCATGCTTCAACGGAGGCGATCTTATTGGCATCTTGCAAACTGGAAGCTTCAGCTGGTCTATAATATTGATCGTTTTTATAACTGTGCTACAATATGGAAGACCTCA
  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
ATTACAGCTGGTTCGCGTATGGCATTAAAAATGCTTATAATGTGGCTATTATGGCCCGTGTGTTGGCTCTACGATTTTTAATGCATACTCGGAATACCAAGTGTCCAGATATGTAATGTT
  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
CGGCTTTAGTATTGTCAGGTGCAATTTGTTACTCTGGATTATGATTTTGTAGATCCATTCAGTTGTACAGAAGGACTAAGTCTTGGTGGTCTTTCAACCTGAAACTAAAGC
  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
AATCTTTGCGTTAGTGCATTAGGAAGAAGCTATGTGCTTCTCTCGAAGGTGTGCCAAGTGGTGTCACTCTAACTTTGCTTTCAGGGAATTTGTACGCTGAAGGGTTCAAAATTGCAGG
  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
TGGTATGAACATCGACAATTTACCAAAATACGTAATGGTTGCATTACCTAGCAGGACTATTGTCTACACACTTGTGGCAAGAAGTTGAAAGCAAGTAGTGCAGCTGGATGGGCTTACTA
  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
TGTAATACTAAAGCTGGTGTACTCAACAGAGGCAAGAACTGATAATTTGAGTGAGCAAGAAAATATTACATATGGTATAACTAAACTTCTAAATGGCCAACCCAGGGACAACGTGT
  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
CAGTTGGGGAGATGAATCTACCAAAACACGTGGTTCGTTCC
  S W G D E S T K T R G R S

```

Fig. 2. Nucleotide sequence of the TGEV M gene and deduced amino acid sequence for the protein. The nucleotide sequence comes from the part of the virus genome illustrated in Fig. 1. A continuous open reading frame beginning at nucleotide position 56 and continuing through nucleotide 922 is identified. The CTAAC intergenic sequences are underlined. The proposed amino terminus for the M protein is identified by an underlined methionine residue near base position 137.

products of a 50  $\mu$ l *in vitro* translation reaction. The M protein was immunoprecipitated with 10  $\mu$ l M-specific monoclonal antiserum (identified as 1A6; Woods *et al.*, 1987).

For labeling virion M protein, cells grown in 150-cm<sup>2</sup> plastic flasks were infected as described above, and incubated with 500  $\mu$ Ci [<sup>35</sup>S]methionine (Translabel, ICN) per flask. At 18 hr p.i., virus was purified from clarified supernatant fluids by isopycnic sedimentation in sucrose gradients as previously described (Brian *et al.*, 1980). Virion proteins were solubilized in 4% SDS, and M protein was immunoprecipitated with M-specific monoclonal antibody and deglycosylated with *N*-glycanase.

### Polyacrylamide gel electrophoresis

*In vitro* translation reaction products or immunoprecipitates on protein A-Sepharose CL-4B beads were diluted with an equal volume of 2 $\times$  Laemmli sample treatment buffer [1 $\times$  sample treatment buffer is 0.0625 M Tris-HCl (pH 6.8), 2% SDS, 10% glycerol, 5% 2-mercaptoethanol] that contained 5 M urea, heated 2 min at 100°, and electrophoresed using the method of Laemmli (1970).

## RESULTS

### Deduced amino acid sequence of the matrix protein

Seven clones, C4, E2, F5, FT35, FT36, FT43, and FT44, mapping in the positions illustrated in Fig. 1, were sequenced in part to extend the TGEV genomic sequence that was known from clones FG5 and J21 (Kapke and Brian, 1986). Clone FG5 maps at the extreme 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 homology with the M proteins of MHV (Armstrong *et al.*, 1984) and IBV (Bournsnel *et al.*, 1984). The sequencing strategy we used is described in Fig. 1.

The molecular weight of the glycosylated M protein has been estimated from electrophoretic migration patterns to be approximately 28 to 30 kDa (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 weight of 28 kDa or less for the unglycosylated protein. Surprisingly, the completed open reading frame of what we had 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 (Fig. 2). Since the M proteins of the mouse hepatitis virus

(Armstrong *et al.*, 1984; Tooze *et al.*, 1984) and bovine coronavirus (Lapps *et al.*, 1987) migrate in sodium dodecyl sulfate-containing polyacrylamide gels with elec-

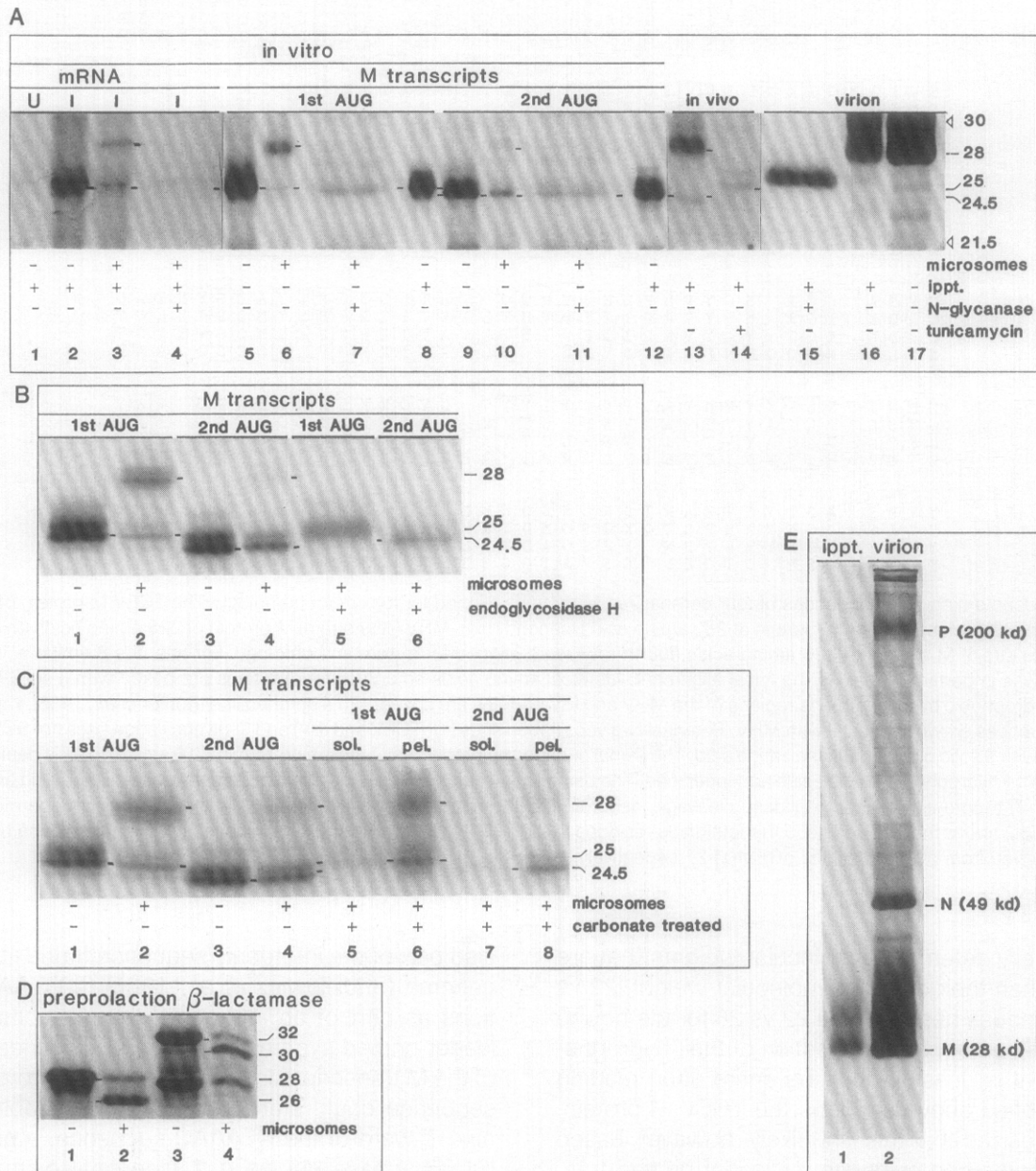


Fig. 3. Translocation and glycosylation of the M protein with and without its amino-terminal peptide. (A) Comparison of M protein from *in vitro* and *in vivo* synthesis and from the virion. Lanes 1 through 12, mRNAs isolated from uninfected (U) or infected (I) cells, and *in vitro* transcripts prepared from the reconstructed full-length M gene (pGEM-M-1 or 1st AUG) or the truncated M gene (pGEM-M-2 or 2nd AUG), were translated *in vitro* and treated as noted. For immunoprecipitation, porcine hyperimmune antiserum was used. Lanes 13 and 14, M protein was immunoprecipitated from cell lysates with monoclonal antibody. Cell lysates were prepared as noted. Lanes 15 and 16, M protein was immunoprecipitated from virion proteins with monoclonal antibody. Immunoprecipitates were treated as noted. Lane 17, proteins from purified TGEV were electrophoresed and the position of the 28-kDa protein is identified. Lanes 16 and 17 are a longer exposure of part of lanes 1 and 2 used in (E). Lanes 1 through 4, 13, and 14 were overexposed to identify species in relatively lower abundance. (B) Endoglycosidase H treatment of the *in vitro* translation products of the reconstructed full-length M gene (1st AUG) and of the truncated M gene (2nd AUG). (C) Sodium carbonate treatment of microsomes containing the translation products of the reconstructed full-length M gene (1st AUG) and of the truncated M gene (2nd AUG). (D) Controls used for detecting signal peptidase activity in the microsomes employed. Positions of the uncleaved and cleaved forms of preprolactin and  $\beta$ -lactamase are shown. (E) Immunoprecipitation showing the specificity of the monoclonal antibody 1A6. All gels were 10% polyacrylamide.  $^{14}$ C-Radiolabeled protein molecular weight markers were obtained from New England Nuclear.

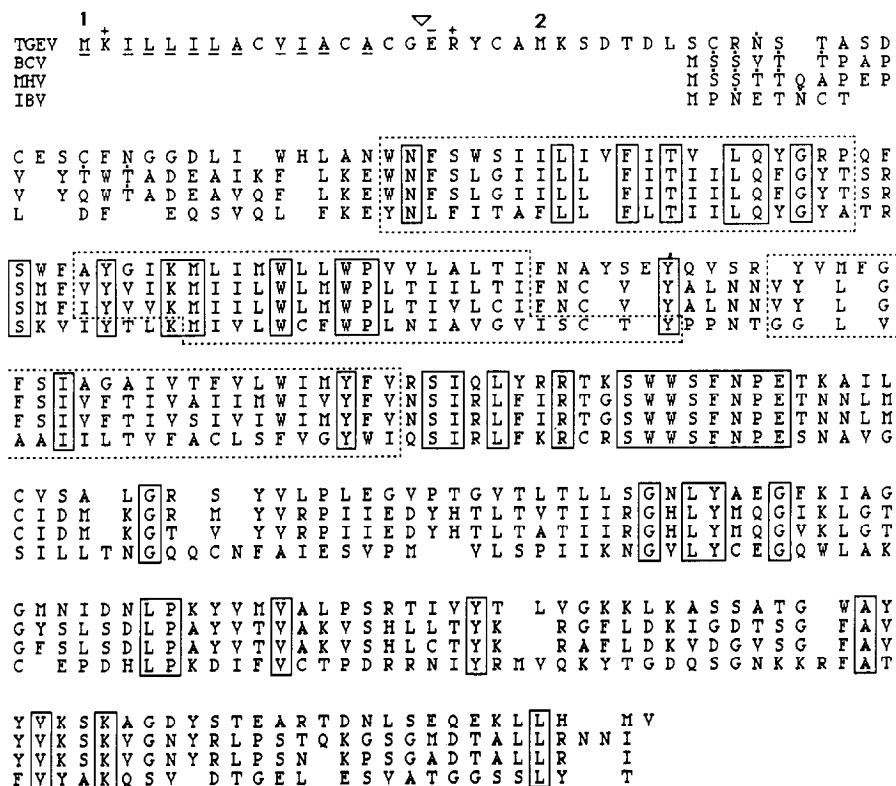


Fig. 4. Deduced amino acid sequences of four coronavirus M proteins aligned for maximum homology. The TGEV M protein of 262 amino acids (top row), BCV (Mebus strain) M protein of 230 amino acids (second row), MHV (A59 strain) M protein of 228 amino acids (third row), and IBV (Beaudette strain) M protein 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. Identical amino acids among all four proteins are boxed with a solid line. The 21 amino acid hydrophobic transmembrane regions of the MHV and IBV proteins are boxed with a hatched line (Rottier *et al.*, 1986), with the TGEV and BCV sequences drawn in parallel with MHV. 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 positions 2 and 18 and the acidic amino acid at position 17 are indicated. The potential site for peptidase cleavage, according to the rules of Von Heinje (1986), is indicated by an open triangle. The numbers 1 and 2 above the first and second methionine residues in the TGEV M sequence indicate initiation sites for translation of the *in vitro* transcripts derived from pGEM-M-1 and pGEM-M-2, respectively.

trophoretically determined molecular weights 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.*, 1986b), 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. 2, which is part of a conserved intergenic sequence in the 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. 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 with the first available methionine 3'-ward of the CTAAAC sequence, a codon that begins at base 137 in Fig. 2. It could also start with the second, downstream, in-frame AUG codon beginning at base 200, but this is surrounded by a much less favorable Kozak consensus sequence (Kozak, 1983).

To test our hypothesis, we prepared transcripts identical to the postulated functional mRNA structure (transcripts from the reconstructed full-length M gene, construct pGEM-M-1, that would initiate translation at the first AUG downstream from the CTAAAC sequence) and compared sizes of the resulting translation products with those of mRNA isolated from infected cells. Figure 3A, lanes 2 and 5, illustrates that

M protein translated *in vitro* from cell-derived, poly(A)-selected mRNA and immunoprecipitated with TGEV-specific antibody had an electrophoretically determined molecular weight of 25K and comigrated with protein translated from the reconstructed full-length M gene. Protein translated from the reconstructed full-length M gene immunoprecipitated with the same antibody thus confirming its authenticity (Fig. 3A, lane 8). Furthermore, truncated M protein (generated from construct pGEM-M-2), although also antigenically authentic, migrated distinctly faster than full-length M but with a migration rate much less than expected for a 2-kDa molecular weight difference (Fig. 3A, lanes 9 and 12). The small difference in electrophoretic mobility between the two forms of M again is an ostensible function of the hydrophobic nature of the protein. The electrophoretically determined molecular weight of the truncated M polypeptide is 24.5K. Judging from the size of the various translation products, it is unlikely that initiation of translation *in vivo* starts at any place other than at the first AUG downstream from the CTAAAC intergenic sequence.

When the first methionine codon downstream from the CTAAAC sequence is used as the initiation site for translation, the deduced M protein comprises 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.

#### A potentially cleavable amino-terminal signal peptide is not an absolute requirement for membrane translocation and glycosylation

A comparison of the amino acid sequence for the M proteins of TGEV, BCV, MHV, and IBV (Fig. 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 hydrophobic domains that align with similar domains in BCV, MHV, and IBV (Fig. 4 and data not shown; Rottier *et al.*, 1984; 1986), BCV (Lapps *et al.*, 1987), and IBV (Bournsell *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 toward the carboxy terminus, the protein spans the membrane three times and has a relatively hydrophilic intravirion carboxy-terminal region (Rottier *et al.*, 1984, 1986).

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 proteins of the four viruses and assuming the MHV M protein enters the virion envelope 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 sixteen terminal amino acids are hydrophobic, and amino acids at positions 2, 17, and 18 are charged (Fig. 4). By inspection, this sequence has the properties of a cleavable amino-terminal signal peptide and from the “-3, -1” rule (von Heijne, 1986) peptidase cleavage would occur between amino acids 16 and 17.

Since an amino-terminal peptide is not required for membrane translocation and glycosylation of the M proteins of BCV, MHV, and IBV, we examined what effect the peptide had on the translocation of the TGEV M protein. Both forms of the reconstructed M gene generate the asparagine glycosylation site when translated, pGEM-M-1 at amino acid position number 32 and pGEM-M-2 at amino acid position number 11 of their respective translation products (Fig. 4). Both forms of transcripts were therefore translated in the presence of microsomes known to have glycosylating activity. Figures 3A, lanes 5, 6, 9, and 10, 3B, lanes 1, 2, 3, and 4, and 3C, lanes 1, 2, 3, and 4, illustrate that,

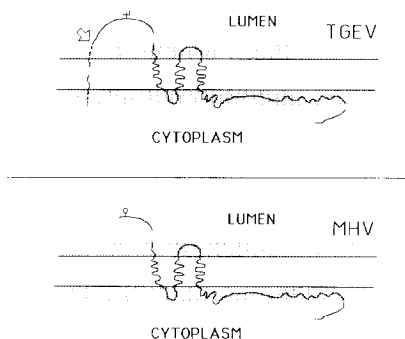


Fig. 5. Proposed topology of the TGEV M protein. The proposed topology of the TGEV M protein with an uncleaved amino-terminal signal peptide is based on the topographical arrangement of the MHV M protein (Rottier *et al.*, 1986). Our model predicts that the amino-terminal signal peptide aids in the translocation of an especially long amino terminus, but is not removed by signal peptidase until some step in virus assembly is achieved. The TGEV M protein, like the MHV M protein, contains internal signals for translocation of the three internal transmembrane regions of the protein. The N-glycosylation site ( $\psi$ ) and O-glycosylation site ( $\phi$ ) on the proteins are shown.



whereas the full-length reconstructed M gene (1st AUG) yielded a 25-kDa protein that was mostly glycosylated, 65–88% as determined from optical density tracings, the truncated M gene (2nd AUG) yielded a protein that was also glycosylated but to a far lesser extent, only 10–30%. The glycosylated products of both the 1st AUG and 2nd AUG transcripts were approximately 28 kDa (Fig. 3A, lanes 6 and 10), the same size as the glycosylated product from viral mRNA translation (Fig. 3A, lane 3) and glycosylated virion M (Fig. 3A, lanes 16 and 17). Glycosylation was confirmed to have taken place since the products from *in vitro* translation migrated again with the unglycosylated polypeptides after they had been digested with *N*-glycanase (Fig. 3A, lanes 7 and 11) or endoglycosidase H (Fig. 3B, lanes 5 and 6). These results demonstrate that although the asparagine glycosylation site of both the full-length and truncated forms of M protein became translocated to the luminal side of the microsome, translocation with the amino-terminal peptide present was far more efficient. Furthermore, translocation of the amino terminus appeared to be specifically enhanced by the peptide since carbonate treatment showed both full-length and truncated forms of the protein as a whole to be equally membrane embedded (Fig. 3C, lanes 5, 6, 7, and 8), presumably as a result of internal translocation signals of the kind described for the MHV and IBV M proteins (Rottier *et al.*, 1985; Machamer and Rose, 1987). Following carbonate treatment greater than 85% of both the full-length and truncated forms of the M protein remained membrane bound as determined by optical density tracing of the autoradiogram.

#### Cleavage of the amino-terminal peptide did not appear to occur during *in vitro* translation

To test for amino-terminal peptide cleavage, viral mRNA from infected cells and transcripts of the cloned M gene were translated in the presence of microsomes known to contain signal peptidase activity (Fig. 3D). In initial experiments, a tripeptide serving as a competitive inhibitor of asparagine-linked glycosylation (Lau *et al.*, 1983) was incubated with the microsomal-translation mixture in order to inhibit concurrent glycosylation that would otherwise obscure the results of peptide cleavage. Although inhibition of glycosylation was never complete, at no time was there evidence of peptide cleavage (data not shown). To use a second approach, mRNA and transcripts were translated in the presence of microsomes and the products were deglycosylated with either *N*-glycanase or endoglycosidase H, and sizes were compared by electrophoresis (Figs. 3A, lanes 4, 7, and 11, and 3B, lanes 5 and 6). Interestingly, the sizes of the deglycosylated

products, as described above, appeared to be no smaller than the polypeptides synthesized without microsomes. There appeared, therefore, to be no amino-terminal peptide removal during *in vitro* translation in the presence of microsomes.

Results of experiments to determine whether the N-terminal peptide is cleaved *in vivo* suggest that cleavage *in vivo* may be dependent upon the glycosylation of M or other glycoproteins. When infected cells were incubated in the presence of tunicamycin and radiolabeled M was immunoprecipitated from cell lysate using monoclonal antibody, only one major band was found and it migrated as an uncleaved protein of 25 kDa (Fig. 3A, lane 14). Similar results were obtained when polyclonal hyperimmune TGEV serum was used (data not shown). On the other hand, in the absence of tunicamycin, two forms of M were precipitated. These were a 28-kDa species, the size of fully glycosylated M, and a 24.5-kDa species, the size of M from which the N-terminal signal had been removed (Fig. 3A, lane 13). Glycosylation *in vivo* may have important consequences on the exposure of the peptidase cleavage site because of interactions between M and other viral components *in vivo*.

In light of a recent report demonstrating by amino-terminal amino acid sequencing that the virion form of TGEV M protein is indeed cleaved (Laude *et al.*, 1987), we included deglycosylated, radiolabeled virion M in our electrophoretic analysis (Fig. 3A, lane 15). The deglycosylated virion M did not migrate as a cleaved polypeptide but rather migrated with an apparent molecular weight of approximately 26 kDa suggesting that it had perhaps undergone a second as yet undetermined modification after becoming virion-associated.

## DISCUSSION

The M protein of MHV was the first coronavirus M protein to be sequenced (Armstrong *et al.*, 1984) and its topography with regard to membrane orientation and insertion has been carefully documented (Rottier *et al.*, 1984, 1986). It therefore serves as the prototypic coronavirus M protein to which others can be compared. The coronavirus M protein apparently functions to direct the budding of virus into the rough endoplasmic reticulum and the Golgi since it inserts into these membranes and is found at highest concentrations there (Holmes *et al.*, 1984; 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 membrane insertion (Rottier *et al.*, 1985). The same is also true for the IBV M protein (Machamer and Rose, 1987), and is probably true for the BCV M protein (Lapps *et al.*, 1987) which has a structure similar to that of the MHV M protein.

It came as a surprise to us, therefore, that the deduced TGEV M protein has, in addition to the three internal hydrophobic domains, an amino-terminal sequence that possesses the properties of a cleavable signal peptide for membrane insertion (von Heijne, 1986). The potentially cleavable N-terminal peptide was also revealed by the work of Laude *et al.* (1987) in which a nearly identical M gene sequence was reported for the same strain of TGEV. In their sequence, bases in positions 375, 468, and 720 of Fig. 2 were T, C, and A, respectively, making the corresponding amino acids at these codon positions Val, Asn, and Asp. Furthermore, they obtained direct evidence that the N-terminal signal peptide was cleaved from the virion-associated M protein. 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 coronaviruses, 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 dissociating-reassociating 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 acquired one during evolution. The polymerase during replication of the TGEV genome could have become dissociated and then reassociated with another minus-strand RNA template that carried the sequence for a 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 coinfecting 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 M

sequence are identical to the VSV G protein signal sequence (Rose and Gallione, 1981).

With regard to the function of the amino-terminal hydrophobic sequence, we propose that it aids in the translocation of the amino terminus of the TGEV M protein through the endoplasmic reticulum as would other amino-terminal signal peptides, but it is not an absolute requirement. The TGEV M protein without its amino-terminal signal peptide, as approximated by the translation product of the pGEM-M-2 construct which is 4 amino acids shorter than the polypeptide identified by the peptidase cleavage site (Laude *et al.*, 1987), can translocate apparently by the use of an internal signal(s) of the type described for the MHV and IBV M proteins (Rottier *et al.*, 1985; Machamer and Rose, 1987). 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, aids enough in the translocation of the additional external amino-terminal sequence that it was evolutionarily selected.

Since the amino-terminal peptide is apparently not removed from the M protein during *in vitro* translation in the presence of microsomes, but is removed from the M protein in the assembled virion (Laude *et al.*, 1987), we propose that its cleavage depends on the context in which it finds itself. Perhaps a step in virion assembly, for example, an interaction with another viral protein or viral RNA, may be required for cleavage to occur. Glycosylation of the M protein alone is apparently not a prerequisite for cleavage *in vitro* (Figs. 3A and B) but may be important in the *in vivo* context. Context is important for the cleavability of other signal peptides. For example, under certain constraints the potentially cleavable signal on the invariant (I $\gamma$ ) chain of class II histocompatibility antigens (Lipp and Dobberstein, 1986) is not cleaved. We further propose that the M protein with an uncleaved signal peptide would have the orientation depicted in Fig. 5. With this orientation, the cleavage site is buried in the membrane and fewer than 37 amino acids are exposed as a loop on the luminal 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 than its MHV, BCV, or IBV counterparts with regard to intracellular trafficking.

## ACKNOWLEDGMENTS

This work was supported by Grant AI-14367 from the National Institute of Allergy and Infectious Diseases, by Grant 82-CRSR-2-1090 from the U.S. Department of Agriculture, and in part by a grant from the National Foundation for Ileitis and Colitis, Inc.

## REFERENCES

- ANDERSON, D. J., and BLOBEL, G. (1983). Immunoprecipitation of proteins from cell-free translations. In "Methods in Enzymology" (S. Fleischer and B. Fleischer, Eds.), Vol. 96, pp. 111-120. Academic Press, New York.
- ARMSTRONG, J., NIEMANN, H., SMEEKENS, S., ROTTIER, P., and WARREN, G. (1984). Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. *Nature (London)* **308**, 751-752.
- BOURNELL, M. E. G., BROWN, T. D. K., and BINNS, M. M. (1984). Sequence of the membrane protein gene from avian coronavirus IBV. *Virus Res.* **1**, 303-313.
- BRIAN, D. A., DENNIS, D. E., and GUY, J. S. (1980). Genome of porcine transmissible gastroenteritis virus. *J. Virol.* **34**, 410-415.
- BRIAN, D. A., HOGUE, B. G., LAPPS, W., POTTS, B. J., and KAPKE, P. A. (1983). Comparative structure of coronaviruses. In "Proceedings from the Fourth International Symposium of Neonatal Diarrhea" (S. D. Acres, Ed.), pp. 100-115. University of Saskatoon, Saskatoon, Saskatchewan, Canada.
- BROWN, T. D. K., and BOURNELL, M. E. G. (1984). Avian infectious bronchitis virus genomic RNA contains sequence homologies at the intergenic boundaries. *Virus Res.* **1**, 15-24.
- BUDZILOWICZ, C. J., WILCZYNSKI, S. P., and WEISS, S. R. (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.
- FUJIKI, Y., HUBBARD, A. L., FOWLER, S., and LAZAROW, P. B. (1982). Isolation of intracellular membranes by means of sodium carbonate treatment: Application to endoplasmic reticulum. *J. Cell Biol.* **93**, 97-102.
- GARWES, D. J., LUCAS, M. H., HIGGINS, D. A., SPIKE, B. V., and CARTWRIGHT, S. F. (1978). Antigenicity of structural components from porcine transmissible gastroenteritis virus. *Vet. Microbiol.* **3**, 179-190.
- GARWES, D. J., and POCOCK, D. H. (1975). The polypeptide structure of transmissible gastroenteritis virus. *J. Gen. Virol.* **29**, 25-34.
- GUBLER, U., and HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263-264.
- HOLMES, K. V., FRANA, M. F., ROBBINS, S. G., and STURMAN, L. S. (1984). Coronavirus maturation. *Adv. Exp. Biol. Med.* **173**, 37-52.
- HOPP, T. P., and WOODS, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* **78**, 3824-3828.
- HUBBARD, S. C., and IVATT, R. J. (1981). Synthesis and processing of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **50**, 555-583.
- JACOBS, L., VAN DER ZEIJST, B. A. M., and HORZINEK, M. C. (1986). Characterization and translation of transmissible gastroenteritis virus mRNAs. *J. Virol.* **57**, 1010-1015.
- KAPKE, P. A., and BRIAN, D. A. (1986). Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. *Virology* **151**, 41-49.
- KAPKE, P. A., TUNG, F. Y. C., BRIAN, D. A., WOODS, R. D., and WESLEY, R. (1987). Nucleotide sequence of the porcine transmissible gastroenteritis matrix protein. *Adv. Exp. Biol. Med.* **218**, 117-122.
- KEMENY, L. J. (1976). Antibody responses in pigs inoculated with transmissible gastroenteritis virus and cross-reactions among ten isolates. *Canad. 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.
- LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**, 680-685.
- LAPPS, W., HOGUE, B. G., and BRIAN, D. A. (1987). Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. *Virology* **157**, 47-57.
- LAU, J. T. Y., WELPLY, J. K., SHENBAGAMURTHI, P., NAIDER, F., and LENNARZ, W. J. (1983). Substrate recognition by oligosaccharyl transferase: Inhibition of cotranslational glycosylation by acceptor peptides. *J. Biol. Chem.* **258**, 15,255-15,260.
- LAUDE, H., RASSCHAERT, D., and HUET, J. C. (1987). Sequence and N-terminal processing of the transmembrane protein E1 of the coronavirus transmissible gastroenteritis virus. *J. Gen. Virol.* **68**, 1687-1693.
- LIPP, J., and DOBBERSTEIN, B. (1986). The membrane-spanning segment of invariant chain (I $\gamma$ ) contains a potentially cleavable signal sequence. *Cell* **46**, 1103-1112.
- MACHAMER, C. E., and ROSE, J. K. (1987). A specific transmembrane domain of a coronavirus E1 glycoprotein is required for its retention in the Golgi region. *J. Cell Biol.* **105**, 1205-1214.
- MAKINO, S., KECK, J. G., STOHLMAN, S. A., and LAI, M. M. C. (1986a). High frequency RNA recombination of murine coronaviruses. *J. Virol.* **57**, 729-737.
- MAKINO, S., STOHLMAN, S. A., and LAI, M. M. C. (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.
- 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.
- Queen, C., and Korn, K. J. (1984). A comprehensive sequence analysis program for the IBM personal computer. *Nucleic Acids Res.* **12**, 581-599.
- Rose, J. K., and Gallione, C. (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., ARMSTRONG, J., and MEYER, D. I. (1985). Signal recognition particle-dependent insertion of coronavirus E1, and intracellular membrane glycoprotein. *J. Biol. Chem.* **260**, 4648-4652.
- ROTTIER, P., BRANDENBURG, D., ARMSTRONG, J., VAN DER ZEIJST, B., and WARREN, G. (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., WELLING, G. W., WELLING-WEBSTER, S., NIESTERS, H. G. M., LENSTRA, J. A., and VAN DER ZEIJST, B. A. M. (1986). Predicted membrane topology of the coronavirus protein E1. *Biochemistry* **25**, 1335-1339.
- TOOZE, J., TOOZE, S., and WARREN, G. (1984). Replication of coronavirus MHV-A59 in sac<sup>-</sup> cells: Determination of the first site of budding of progeny virions. *Eur. J. Cell Biol.* **33**, 281-294.
- VON HEIJNE, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**, 4683-4690.
- WESLEY, R. D., and WOODS, R. D. (1986). Identification of a 17,000 molecular weight antigenic peptide in transmissible gastroenteritis virus. *J. Gen. Virol.* **67**, 1419-1425.
- WOODS, R. D., WESLEY, R. D., and KAPKE, P. A. (1987). Neutralization of transmissible gastroenteritis virus by complement-dependent monoclonal antibodies. *Amer. J. Vet. Res.* **49**, 300-304.