J. gen. Virol. (1987), **68**, 1687–1693. *Printed in Great Britain*

Key words: *TGEV/E1 protein sequence/signal peptide*

Sequence and N-terminal Processing of the Transmembrane Protein E1 of the Coronavirus Transmissible Gastroenteritis Virus

By HUBERT LAUDE,[1]* DENIS RASSCHAERT[1] AND JEAN-CLAUDE HUET[2]

Institut National de la Recherche Agronomique,[1] Station de Recherches de Virologie et d'Immunologie, F-78850 Thiverval-Grignon and [2]Laboratoire d'Etude des Protéines, F-78000 Versailles, France

(*Accepted 26 February 1987*)

SUMMARY

Sequencing of part of a clone from a transmissible gastroenteritis virus genome cDNA library led to the identification of the gene encoding the E1 matrix protein. The amino acid sequence of the primary translation product predicts a polypeptide of 262 residues which shares many features with the previously characterized murine hepatitis virus and infectious bronchitis virus E1 proteins. However, N-terminal amino acid sequencing revealed that a putative signal peptide of 17 residues was absent in the virion-associated polypeptide. The predicted mol. wt. of the mature unglycosylated product, 27800, is in agreement with the experimental M_r value.

INTRODUCTION

Transmissible gastroenteritis (TGE) is a highly contagious disease of pigs causing high mortality in neonates. The causal agent (TGE virus, TGEV) belongs to the family Coronaviridae, a group of enveloped viruses with a large, positive-stranded RNA genome (for review, see Siddell *et al.*, 1983). Coronavirus-encoded information is expressed in the cell through a nested set of subgenomic mRNAs with common 3′-terminal sequences. The coding part of each mRNA corresponds approximately to the 5′-terminal sequences that are absent in the next smaller species. Mouse hepatitis virus (MHV) and infectious bronchitis virus (IBV) mRNAs contain at the 5′ end a short non-coding sequence joined to the body sequences by discontinuous transcription; a consensus sequence identified at each intergenic region may act as a binding site for the RNA polymerase–leader complex (Spaan *et al.*, 1983; Brown *et al.*, 1984; Lai *et al.*, 1984; Budzilowicz *et al.*, 1985).

TGEV contains three major structural polypeptides: the peplomer glycoprotein E2 (200K to 220K), which forms the distinctive surface projections, the transmembrane or matrix protein E1 (29K ± 1K), and the nucleoprotein N (47K ± 1K), in which a single infectious RNA molecule of 20 kb or more is embedded (Garwes *et al.*, 1976; Laude *et al.*, 1986; Brian *et al.*, 1980). In TGEV-infected cells, five species of subgenomic mRNA have been characterized, *in vitro* translation of which has allowed partial coding assignment (Hu *et al.*, 1984; Jacobs *et al.*, 1986; Rasschaert *et al.*, 1987).

The matrix protein has been the subject of intensive studies in other coronaviruses. The N-terminal regions of E1 proteins of MHV and bovine coronavirus (BCV) bear sugar chains *O*-linked to Ser or Thr residues (Niemann & Klenk, 1981; Niemann *et al.*, 1984), an unusual feature among viral glycoproteins. Also, unlike the majority of integral membrane proteins, both MHV and IBV E1 proteins lack a cleaved signal peptide (Rottier *et al.*, 1984; Stern & Sefton, 1982*a*). The restriction of E1 to internal membranes seems to determine the assembly of the coronavirus particles in the lumen of the endoplasmic reticulum (Holmes *et al.*, 1981; Tooze *et al.*, 1985). A model of the membrane topology of E1 has recently emerged from a combination of biochemical data and analysis of its primary structure (Armstrong *et al.*, 1984; Rottier *et al.*,

1986). The amino half is formed of a transmembrane helix with two hairpin structures, whereas the carboxy half is closely adjacent to the inner surface of the viral envelope and is largely resistant to proteolysis. At each end, two short hydrophilic segments are assumed to project from either side of the membrane.

In an attempt to elucidate the functional domains of the TGEV surface glycoproteins, a cDNA library of the TGEV genome has been created in this laboratory. In this paper, we report the sequence of part of a clone encoding the E1 gene, present information about the N-terminal processing of its product and compare the features predicted from its primary structure with those of the other E1 coronavirus proteins.

METHODS

cDNA cloning. The preparation of cDNA clones will be described in more detail in a subsequent communication. cDNA was obtained from purified genomic RNA (TGEV Purdue-115 strain) using oligo-$(T)_{12-18}$ (Pharmacia) as primer and reverse transcriptase (Stehelin, Basel, Switzerland). RNase T2-treated cDNA–RNA hybrids were dC-tailed and inserted in *Pst*I-cut dG-tailed pBR322 (Bethesda Research Laboratories) (Zain *et al.*, 1979; Van der Werf *et al.*, 1981). *Escherichia coli* RR1 cells were transfected with this material. An insert (5 kb) covering the complete E1 coding region was located in the clone pTG2.15.

DNA sequencing. Sonicated fragments of pTG2.15 were subcloned into *Sma*I-cut M13mp18 vector (Deininger, 1983). Sequencing was performed by Sanger's dideoxy technique using [^{35}S]dATP (New England Nuclear) as the label and the reaction products were analysed in buffer gradient gels. Each strand of DNA was sequenced at least three times.

Sequence analysis. Sequences were analysed using the program of Queen & Korn (1984), marketed as part of the Microgenie program (March 1985 version, Beckman), developed for the IBM PC-XT microcomputer. The program, utilizing the hydrophilicity values of Hopp & Woods (1981), the mean fractional area exposed values of Rose *et al.* (1985), and the flexibility values (predicted Bnorm. data) of Karplus & Schulz (1985) was written in Apple Basic (F. Borras-Cuesta & H. Laude, unpublished data).

Isolation of the E1 polypeptide and partial amino acid analysis. Purified virion polypeptides were resolved by SDS–PAGE as described by Laude *et al.* (1986). After localization by gel slice staining, the protein band was excised from the gel and placed in an electroelution chamber (Isco). The protein was eluted for 24 h in 50 mM-$NaHCO_3$ + 0·1% SDS. The reservoir buffer was subsequently changed to 10 mM-$NaHCO_3$ + 0·01% SDS and electrodialysis was performed overnight (Hunkapiller *et al.*, 1983). About 100 pmol of protein was subjected to N-terminal amino acid sequencing. Sequential Edman degradation was done on a 'gas phase' Applied Biosystems 470A apparatus with its dedicated on-line PTH amino acid analyser 120A.

RESULTS

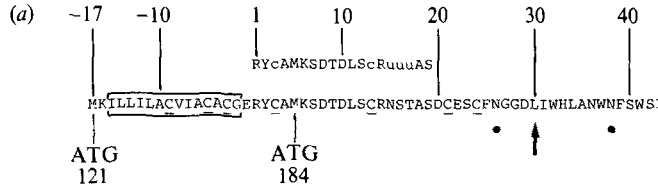
A long open reading frame (ORF) of 867 bases yielding a protein with the properties of E1, was identified on clone pTG2.15 of TGEV cDNA. The 5′ end of this ORF mapped at 2·48 kb from the 3′ end of the genome (Rasschaert *et al.*, 1987). According to its length and position, the E1 ORF corresponded to the 'unique' region of the mRNA 5 within the set of viral RNAs characterized by Northern blot analysis (data not shown).

Inspection of the nucleotide sequence displayed in Fig. 1 revealed the presence, near each extremity, of two AACTAAAC sequences, which were assumed to be the start of the mRNA transcripts 5 (E1-encoding) and 6 (N-encoding) respectively. Therefore, although the ORF extended 23 codons upstream from the first consensus sequence, it was postulated that initiation of translation on mRNA 5 should occur downstream at either of the two proximal ATGs available. The ATG adjacent to the consensus sequence is followed by a characteristically hydrophobic stretch of amino acids, which may possibly act as a signal peptide for the translocation of E1. Alternatively, translation might start at the next ATG codon (position 184), yielding a product of 241 residues, still slightly larger in size than the matrix proteins of other coronaviruses. However, the second ATG lies in a less favourable context than the first for initiation of translation (Kozak, 1983).

Partial microsequencing of the mature E1 polypeptide was performed to confirm the site of translation initiation (Fig. 2). The N-terminal residues thus identified were found in perfect agreement with the predicted sequence up to position 14, beyond which the sequencing process was perturbed. This led us to conclude that: (i) translation of E1 cannot be initiated at the second ATG codon (position 184) and (ii) a hydrophobic peptide specified by the first 17 codons of the gene was not present in virion-associated E1.

```
                                30                                  60                                  90
GCA ATT TAG GAA GGA CAG TTA TTA TTG TTC CAG CGC AAC ATG CTT ACG ATG CCT ATA AGA ATT TTA TGC GAA TTA AAG CAT ACA ACC CCG
Ala Ile End Glu Gly Gln Leu Leu Leu Phe Gln Arg Asn Met Leu Thr Met Pro Ile Arg Ile Leu Cys Glu Leu Lys His Thr Thr Pro

                               120                                 150                                 180
ATG GAG CAC TCC TTG CTT GAA CTA AAC AAA ATG AAG ATT TTG TTA ATA TTA GCG TGT GTG ATT GCA TGC GCA TGT GGA GAA CGC TAT TGT
Met Glu His Ser Leu Leu Glu Leu Asn Lys Met Lys Ile Leu Leu Ile Leu Ala Cys Val Ile Ala Cys Ala Cys Gly Glu Arg Tyr Cys

                               210                                 240                                 270
GCT ATG AAA TCC GAT ACA GAT TTG TCA TGT CGC AAT AGT ACA GCG TCT GAT TGT GAG TCA TGC TTC AAC GGA GGC GAT CTT ATT TGG CAT
Ala Met Lys Ser Asp Thr Asp Leu Ser Cys Arg Asn Ser Thr Ala Ser Asp Cys Glu Ser Cys Phe Asn Gly Gly Asp Leu Ile Trp His
                                                ●
                               300                                 330                                 360
CTT GCA AAC TGG AAC TTC AGC TGG TCT ATA ATA TTG ATC GTT TTT ATA ACT GTG CTA CAA TAT GGA AGA CCT CAA TTC AGC TGG TTC GTG
Leu Ala Asn Trp Asn Phe Ser Trp Ser Ile Ile Leu Ile Val Phe Ile Thr Val Leu Gln Tyr Gly Arg Pro Gln Phe Ser Trp Phe Val
                ●
                               390                                 420                                 450
TAT GGC ATT AAA ATG CTT ATA ATG TGG CTA TTA TGG CCC GTT GTT TTG GCT CTT ACG ATT TTT AAT GCA TAC TCG GAA TAC CAA GTG TCC
Tyr Gly Ile Lys Met Leu Ile Met Trp Leu Leu Trp Pro Val Val Leu Ala Leu Thr Ile Phe Asn Ala Tyr Ser Glu Tyr Gln Val Ser

                               480                                 510                                 540
AGA TAT GTA ATG TTC GGC TTT AGT ATT GCA GGT GCA ATT GTT ACA TTT GTA CTC TGG ATT ATG TAT TTT GTA AGA TCC ATT CAG TTG TAC
Arg Tyr Val Met Phe Gly Phe Ser Ile Ala Gly Ala Ile Val Thr Phe Val Leu Trp Ile Met Tyr Phe Val Arg Ser Ile Gln Leu Tyr

                               570                                 600                                 630
AGA AGG ACT AAC TCT TGG TGG TCT TTC AAC CCT GAA ACT AAA GCA ATT CTT TGC GTT AGT GCA TTA GGA AGA AGC TAT GTG CTT CCT CTC
Arg Arg Thr Asn Ser Trp Trp Ser Phe Asn Pro Glu Thr Lys Ala Ile Leu Cys Val Ser Ala Leu Gly Arg Ser Tyr Val Leu Pro Leu

                               660                                 690                                 720
GAA GGT GTG CCA ACT GGT GTC ACT CTA ACT TTG CTT TCA GGG AAT TTG TAC GCT GAA GGG TTC AAA ATT GCA GAT GGT ATG AAC ATC GAC
Glu Gly Val Pro Thr Gly Val Thr Leu Thr Leu Leu Ser Gly Asn Leu Tyr Ala Glu Gly Phe Lys Ile Ala Asp Gly Met Asn Ile Asp

                               750                                 780                                 810
AAT TTA CCA AAA TAC GTA ATG GTT GCA TTA CCT AGC AGG ACT ATT GTC TAC ACA CTT GTT GGC AAG AAG TTG AAA GCA AGT AGT GCG ACT
Asn Leu Pro Lys Tyr Val Met Val Ala Leu Pro Ser Arg Thr Ile Val Tyr Thr Leu Val Gly Lys Lys Leu Lys Ala Ser Ser Ala Thr

                               840                                 870                                 900
GGA TGG GCT TAC TAT GTA AAA TCT AAA GCT GGT GAT TAC TCA ACA GAG GCA AGA ACT GAT AAT TTG AGT GAG CAA GAA AAA TTA TTA CAT
Gly Trp Ala Tyr Tyr Val Lys Ser Lys Ala Gly Asp Tyr Ser Thr Glu Ala Arg Thr Asp Asn Leu Ser Glu Gln Glu Lys Leu Leu His
                                                                    ●
                               930
ATG GTA TAA CTA AAC TTC TAA ATG GCC AAC CAG GGA CAA CGT GTC AGT TGG GGA GAT
Met Val                     Met Ala Asn Gln Gly Gln Arg Val Ser Trp Gly Asp
```

Fig. 1. Part of the sequence (957 bases) from the TGEV genomic cDNA clone pTG2.15. The main ORF, encoding the E1 protein, and part of the ORF corresponding to the adjacent nucleocapsid gene (broken line) are translated in the three letter amino acid code. The two consensus sequences are boxed. Proximal ATG codons are underlined, stop codons are overlined. Dots beneath Asn residues indicate *N*-glycosylation signals (Asn-X-Ser or Asn-X-Thr). The signal peptide-like sequence is underlined.



(*b*)	Amino acid	Arg	Tyr	(c)	Ala	Met	Lys	Ser	Asp	Thr	Asp	Leu	Ser	(c)	Arg
	pmol PTH	90	131		115	98	107	39	66	23	34	34	39		34

Fig. 2. Proximal amino acid data of TGEV E1 polypeptide. (*a*) Residues identified by partial N-terminal microsequencing of the virion protein are aligned over the predicted sequence of the E1 precursor. The uncharged region of the putative signal peptide is boxed. The arrow indicates the proposed site of entry of E1 into the lipid membrane. (*b*) Quantity of PTH derivative measured for the first 14 residues. Symbols: c, Cys suspected as no PTH derivative was detected; u, residue undetermined; ●, as in Fig. 1.

DISCUSSION

A cDNA cloned from TGEV genomic RNA was sequenced in the region corresponding to the coding part of mRNA 5 (Fig. 1). A single long ORF was found, which is shown to direct the synthesis of a coronavirus matrix-like protein. Only one additional ORF of more than 20 amino acids was detected, which was 43 residues long and within the E1 gene (position 240). Partial microsequencing of the virion-associated E1 polypeptide unambiguously established that the first residue was the Arg specified by the CGC codon at position 172 in Fig. 1. Hence it can be deduced that the single ATG codon available between the CGC codon and the upstream consensus sequence CTAAAC should be the functional initiation codon of mRNA 5. It predicts a primary translation product of 262 amino acids with mol. wt. 29·6 K, slightly higher than the M_r 25K reported for the *in vitro* translation product of mRNA 5 (Jacobs *et al.*, 1986; referred to as mRNA 6 in their study).

The first 17 residues predicted from the nucleotide sequence were found to be lacking in the mature protein (Fig. 2). This oligopeptide actually fulfils the criteria of an eukaryotic signal peptide, namely a net charge immediately after the N-terminus and high degree of hydrophobicity of the 14 residue long uncharged region (McGeoch, 1985; Von Heijne, 1986). A striking feature is the presence of a triple Ala–Cys repeat, which also occurs in the signal peptide of rat oxytocin (see McGeoch, 1985). Also, the cleavage site appeared to be located between Gly and Arg, as predicted by the 75 to 80% accurate weight-matrix approach of Von Heijne (1986).

These results indicate that TGEV E1 matures through the removal of a 17 amino acid leader peptide. Accordingly, the final product is 245 amino acids long, and has a predicted mol. wt. of 27780 in the unglycosylated form; it is basic (with five net charges at neutral pH) and 44% of the residues are hydrophobic. This finding is in contrast with that reported for the matrix proteins of two other coronaviruses, MHV and IBV (Rottier *et al.*, 1986; Armstrong *et al.*, 1984; Stern & Sefton, 1982*a*; Boursnell *et al.*, 1984). It has been proposed that the matrix proteins of the latter are inserted into the membrane by the recognition of an internal transmembrane region as a signal sequence (Rottier *et al.*, 1985). Incidentally, in the IBV E1 sequence (Boursnell *et al.*, 1984), the 22 in-frame codons between the consensus sequence and the initiation codon predict numerous hydrophobic residues, which might be the remnant of an ancestral signal peptide.

Pairwise comparisons of the gene sequence of TGEV E1 with those of MHV and IBV at the DNA level revealed no significant homology. The amino acid sequences showed, in contrast, a remarkable homology. The homologies found by Dayoff's optimal alignment are 38% (TGEV–MHV), 30% (MHV–IBV) and 27% (TGEV–IBV). The main regions of homology are shown in Fig. 3. An eight amino acid section is perfectly conserved among the three viruses (residues 128 to 135, TGEV). Of the three potential membrane-spanning regions (thickly underlined), the second is well conserved within the MHV–IBV pair, and the third within the TGEV–MHV pair; only the first shows a nearly equal degree of homology within both pairs. This might be indicative of functional differentiation between the three segments.

The above findings suggested that the topology of TGEV E1 within the membrane might be essentially similar to that proposed for MHV and IBV (Armstrong *et al.*, 1984; Boursnell *et al.*, 1984; Rottier *et al.*, 1986). This is supported by the data presented in Fig. 4, where each profile corresponds to a computer-assisted prediction of the local tendency of the TGEV E1 polypeptide chain to hydrophilicity (Hopp & Woods, 1981), accessibility (Rose *et al.*, 1985) or mobility (Karplus & Schulz, 1985). By combining our data with those cited above, five regions can be delineated from the amino to the carboxy end: a signal peptide (-17 to -1), an exposed glycosylated segment (1 to 29), three lipid bilayer-incorporated segments (30 to 55, 66 to 86, 98 to 117), an amphiphilic C-terminal half supposedly associated to the cytoplasmic face of the membrane (118 to 228) and a protruding C terminus.

Unlike the MHV and BCV matrix proteins, glycosylation of TGEV E1 has been reported to be of the *N*-linked type (Garwes *et al.*, 1984; Jacobs *et al.*, 1986), as for IBV (Stern & Sefton, 1982*b*). Indeed, two potential *N*-glycosylation sites are available near the N terminus of the sequence (Fig. 2). The accessibility of the second site (Asn, 38) in the lumen of the endoplasmic reticulum is uncertain since the first putative membrane-spanning segment of TGEV E1 could extend virtually up to position 30, i.e. seven residues farther than the WNFS sequence, where

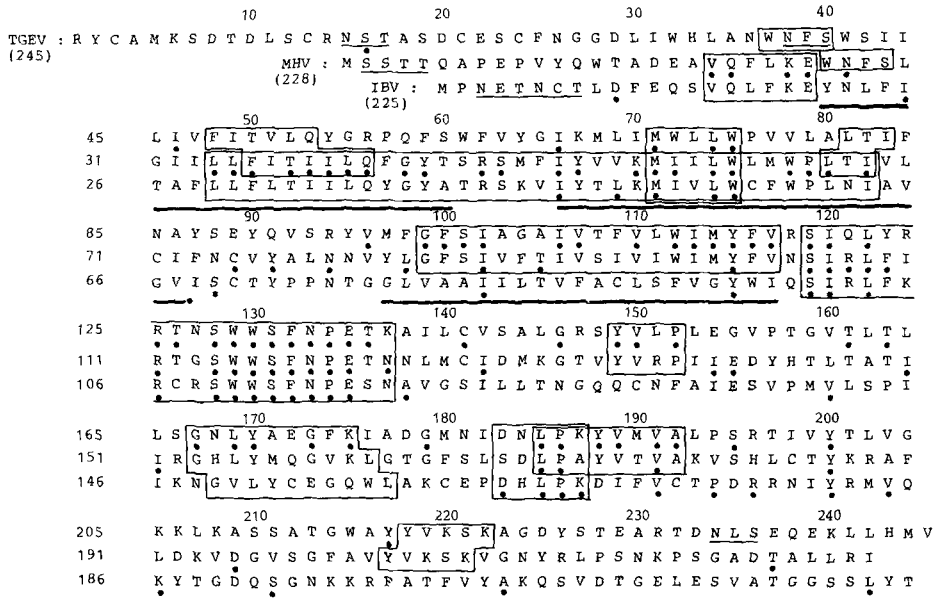


Fig. 3. Comparison of the amino acid sequences (single letter code) of the E1 polypeptides of TGEV, MHV (strain A59: Armstrong *et al.*, 1984) and IBV E1 (Beaudette strain: Boursnell *et al.*, 1984). Dots denote a match between two residues; matches between the TGEV and the IBV sequences are indicated beneath the latter. With the simple alignment used, the homology within each E1 pair is 23% (TGEV–MHV), 26% (MHV–IBV) and 15% (TGEV–IBV). Boxed regions show homologies $\geqslant 60\%$ between two or three sequences. Potential glycosylation sites are underlined. Thick bars indicate the three putative membrane-spanning segments (see text).

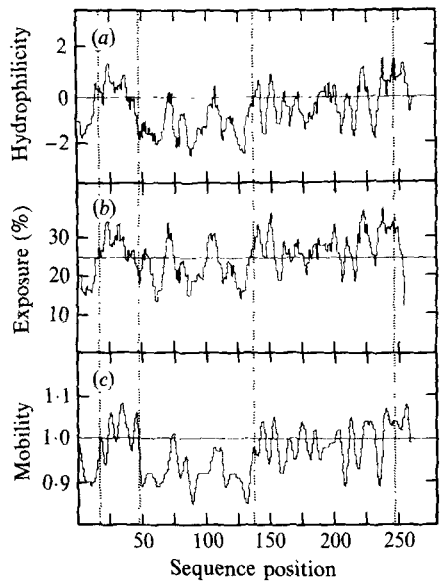


Fig. 4. Graphical output of three prediction methods applied to the amino acid sequence of TGEV E1 precursor polypeptide. (*a*) Hydrophilicity profile with a running average taken over a hexapeptide. (*b*) Exposure profile averaged as in (*a*); the straight line represents the mean area exposed calculated for the whole E1 chain. (*c*) Flexibility profile, with calculations made on a window of seven residues.

the MHV spanning segment is assumed to end (see Fig. 3). In contrast the side chain of Asn at residue 15 may be linked to an oligosaccharide residue, as suggested by the disturbance observed at this point during the N-terminal sequencing (see Fig. 2). The supposition that only the first *N*-glycosylation site is functional leads to an estimated mol. wt. of 29·5K to 30K, a value close to the M_r value of the E1 major species determined by electrophoresis (since the M_r of a carbohydrate-rich mannose chain is about 2K; Klenk & Rott, 1980). Minor E1 species consistently observed as bands migrating more slowly in SDS–PAGE (Garwes & Pocock, 1975; Hu *et al.*, 1984; Laude *et al.*, 1986; Jacobs *et al.*, 1986) might reflect a heterogeneity in the oligosaccharide chain rather than in the polypeptide chain (for example an oversized E1 polypeptide produced from mRNA 4, only 10% larger than mRNA 5). This viewpoint is supported by the fact that TGEV E1 yielded a single band after endoglycosidase H treatment (Hu *et al.*, 1984; B. Delmas & H. Laude, unpublished results).

To sum up, the E1 protein of TGEV shares many structural features with those of MHV and IBV. It is becoming clear, however, that a certain diversity may exist despite the constraints that are necessary to achieve the distinctive architecture of a coronavirus particle. This is true at least for the small hydrophilic region protruding out of the particle, to which no biological function has been assigned so far. Our results provide substantial evidence that TGEV E1 undergoes N-terminal processing and that its exposed NH_2 extremity may be significantly larger and possibly more complex in secondary structure than those of IBV and MHV. Protease digestion has been shown to remove an external glycopeptide of nine residues (IBV; Cavanagh *et al.*, 1986) and 2·5K (MHV; Rottier *et al.*, 1984). In comparison, the TGEV E1 external segment may approach 30 residues, including four cysteines (see Fig. 2*a* and 3). Recent investigations have suggested an involvement of TGEV E1 in the induction of interferon α from non-immune lymphocytes (see Rasschaert *et al.*, 1987; B. Charley & H. Laude, unpublished observation). An interesting possibility is raised that this previously unrecognized activity of E1 might proceed through the interaction of its NH_2 free tail with the lymphocyte surface. Additional experiments are currently underway in order to study this question.

We thank J. Gelfi for her excellent technical assistance and J. C. Pernollet (Head of Laboratoire d'Etude des Protéines, Versailles, France) for supporting the collaboration. Part of the results were presented at the Third International Coronavirus Symposium (Asilomar, September 1986).

REFERENCES

ARMSTRONG, J., NIEMANN, H., SMEEKENS, S., ROTTIER, P. & WARREN, G. (1984). Sequence and topology of a model intracellular membrane protein, E1 glycoprotein from a coronavirus. *Nature, London* **308**, 751–752.

BOURSNELL, M. E. G., BROWN, T. D. K. & BINNS, M. M. (1984). Sequence of the membrane protein gene from avian coronavirus IBV. *Virus Research* **1**, 303–313.

BRIAN, D. A., DOUGLAS, D. E. & GUY, J. S. (1980). Genome of porcine transmissible gastroenteritis virus. *Journal of Virology* **34**, 410–415.

BROWN, T. D. K., BOURSNELL, M. E. G. & BINNS, M. M. (1984). A leader sequence is present on mRNA A of avian infectious bronchitis virus. *Journal of General Virology* **65**, 1437–1442.

BUDZILOWICZ, C. J., WILCZYNSKI, S. P. & 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 3′ end of the viral mRNA leader sequence. *Journal of Virology* **53**, 834–840.

CAVANAGH, D., DAVIS, P. J. & PAPPIN, D. J. C. (1986). Coronavirus IBV glycopolypeptides: locational studies using proteases and saponin, a membrane permeabilizer. *Virus Research* **4**, 145–156.

DEININGER, P. L. (1983). Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Analytical Biochemistry* **129**, 216–223.

GARWES, D. J. & POCOCK, D. H. (1975). The polypeptide structure of transmissible gastroenteritis virus. *Journal of General Virology* **29**, 25–34.

GARWES, D. J., POCOCK, D. H. & PIKE, B. V. (1976). Isolation of subviral components from transmissible gastroenteritis virus. *Journal of General Virology* **32**, 283–294.

GARWES, D. J., BOUNTIFF, L., MILLSON, G. C. & ELLEMAN, C. J. (1984). Defective replication of porcine transmissible gastroenteritis virus in a continuous cell line. *Advances in Experimental Medicine and Biology* **173**, 79–93.

HOLMES, K. V., DOLLER, E. W. & STURMAN, L. S. (1981). Tunicamycin resistant glycosylation of coronavirus glycoprotein: demonstration of a novel type of viral glycoprotein. *Virology* **115**, 334–344.

HOPP, T. P. & WOODS, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences, U.S.A.* **78**, 3824–3828.

HU, S., BRUSZEWSKI, J., BOONE, T. & SOUZA, L. (1984). Cloning and expression of the surface glycoprotein gp 195 of porcine transmissible gastroenteritis virus. In *Modern Approaches to Vaccines*, pp. 219–223. Edited by R. M. Chanock & R. A. Lerner. New York: Cold Spring Harbor Laboratory.

HUNKAPILLER, M. W., LUJAN, E., OSTRANDER, F. & HOOD, L. E. (1983). Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. *Methods in Enzymology* **91**, 227–236.

JACOBS, L., VAN DER ZEIJST, B. A. M. & HORZINEK, M. C. (1986). Characterization and translation of transmissible gastroenteritis virus mRNAs. *Journal of Virology* **57**, 1–6.

KARPLUS, P. A. & SCHULZ, G. E. (1985). Prediction of chain flexibility in proteins, a tool for the selection of peptide antigens. *Naturwissenschaften* **72**, 212–214.

KLENK, H. D. & ROTT, R. (1980). Cotranslational and posttranslational processing of viral glycoproteins. *Current Topics in Microbiology and Immunology* **90**, 19–48.

KOZAK, M. (1983). Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. *Microbiological Reviews* **47**, 1–45.

LAI, M. M. C., BARIC, R. S., BRAYTON, P. R. & STOHLMAN, S. A. (1984). Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proceedings of the National Academy of Sciences, U.S.A.* **81**, 3626–3630.

LAUDE, H., CHAPSAL, J. M., GELFI, J., LABIAU, S. & GROSCLAUDE, J. (1986). Antigenic structure of transmissible gastroenteritis virus. I. Properties of monoclonal antibodies directed against virion proteins. *Journal of General Virology* **67**, 119–130.

McGEOCH, D. J. (1985). On the predictive recognition of signal peptide sequences. *Virus Research* **3**, 271–286.

NIEMANN, H. & KLENK, H. D. (1981). Coronavirus glycoprotein E1, a new type of viral glycoprotein. *Journal of Molecular Biology* **153**, 993–1010.

NIEMANN, H., GEYER, R., KLENK, H. D., LINDER, D., STIRM, S. & WIRTH, M. (1984). The carbohydrates of mouse hepatitis virus (MHV) A59: structures of the *O*-glycosidically linked oligosaccharides of glycoprotein E1. *EMBO Journal* **3**, 665–670.

QUEEN, C. & KORN, L. J. (1984). A comprehensive sequence analysis program for the IBM personal computer. *Nucleic Acids Research* **12**, 581–599.

RASSCHAERT, D., DELMAS, B., CHARLEY, B., GROSCLAUDE, J., GELFI, J. & LAUDE, H. (1987). Surface glycoproteins of transmissible gastroenteritis virus: functions and gene sequence. In *Biochemistry and Biology of Coronaviruses*. Edited by M. M. C. Lai & S. Stohlman. New York & London: Plenum Press (in press).

ROSE, G. D., GESELOWITZ, A. R., LESSER, G. J., LEE, R. H. & ZEHFUS, M. H. (1985). Hydrophobicity of amino acid residues in globular proteins. *Science* **229**, 834–838.

ROTTIER, P., BRADENBURG, D., ARMSTRONG, J., VAN DER ZEIJST, B. A. M. & WARREN, G. (1984). Assembly *in vitro* of a spanning membrane protein of the endoplasmic reticulum: the E1 glycoprotein of coronavirus mouse hepatitis virus A59. *Proceedings of the National Academy of Sciences, U.S.A.* **81**, 1421–1425.

ROTTIER, P., ARMSTRONG, J. & MEYER, D. I. (1985). Signal recognition particle dependent insertion of coronavirus E1, an intracellular membrane glycoprotein. *Journal of Biological Chemistry* **260**, 4648–4652.

ROTTIER, P., WELLING, G. W., WELLING-WESTER, S., NIESTERS, H. G. M., LENSTRA, J. A. & VAN DER ZEIJST, B. A. M. (1986). Predicted membrane topology of the coronavirus protein E1. *Biochemistry* **25**, 1335–1339.

SIDDELL, S., WEGE, H. & TER MEULEN, V. (1983). The biology of coronaviruses. *Journal of General Virology* **64**, 761–776.

SPAAN, W. J. M., DELIUS, H., SKINNER, M., ARMSTRONG, J., ROTTIER, P., SMEEKENS, S., VAN DER ZEIJST, B. A. M. & SIDDELL, S. G. (1983). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO Journal* **2**, 1839–1844.

STERN, D. F. & SEFTON, B. M. (1982*a*). Coronavirus proteins: biogenesis of avian infectious bronchitis virus virion proteins. *Journal of Virology* **44**, 794–803.

STERN, D. F. & SEFTON, B. M. (1982*b*). Coronavirus proteins: structure and function of the oligosaccharides of the avian infectious bronchitis virus glycoproteins. *Journal of Virology* **44**, 804–812.

TOOZE, J., TOOZE, S. A. & WARREN, G. (1985). Laminated cisternae of the rough endoplasmic reticulum induced by coronavirus MHV-A59 infection. *European Journal of Cell Biology* **36**, 108–115.

VAN DER WERF, S., BREGEGERE, F., KOPECKA, H., KITAMURA, N., ROTHBERG, P. G., KOURILSKY, P., WIMMER, E. & GIRARD, M. (1981). Molecular cloning of the genome of poliovirus type 1. *Proceedings of the National Academy of Sciences, U.S.A.* **78**, 5983–5987.

VON HEIJNE, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Research* **14**, 4683–4690.

ZAIN, S., SAMBROOK, J., ROBERTS, R. J., KELLER, W., FRIED, M. & DUNN, A. R. (1979). Nucleotide sequence analysis of the leader segments in a cloned copy of adenovirus 2 fiber mRNA. *Cell* **16**, 851–861.

(*Received 28 November 1986*)