

VRR 00368

The nucleotide sequence of the peplomer gene of porcine transmissible gastroenteritis virus (TGEV): comparison with the sequence of the peplomer protein of feline infectious peritonitis virus (FIPV)

Liesbeth Jacobs, Raoul de Groot, Bernard A.M. van der Zeijst¹,
Marian C. Horzinek and Willy Spaan

*Institute of Virology and ¹ Division of Bacteriology, Veterinary Faculty, State University,
Utrecht, The Netherlands*

(Accepted for publication 5 August 1987)

Summary

The amino acid sequence of the peplomer protein of transmissible gastroenteritis virus (TGEV) has been derived from the cloned cDNA sequence. The gene encodes a protein of 1447 amino acids with a molecular weight of 159 574. Comparison with the primary structure of the peplomer protein of feline infectious peritonitis virus (FIPV) (de Groot et al., 1987b) revealed one domain, from amino acids 1 to 274, in which the nucleotide homology was 39%, whereas in the second domain (from residues 275 to 1447) the homology was 93%.

Coronavirus TGEV; Sequence, peplomer protein; TGEV, FIPV

Introduction

Transmissible gastroenteritis virus (TGEV) causes an acute and mostly fatal enteric disease in newborn piglets. It belongs to the Coronaviridae, a family of enveloped positive-stranded RNA viruses. The virion contains three structural proteins: a surface (peplomer) glycoprotein (E2: 195 kDa), a transmembrane matrix

Correspondence to: W. Spaan, Institute of Virology, Veterinary Faculty, State University, Yalelaan 1, 3508 TD Utrecht, The Netherlands.

glycoprotein (E1: 28 kDa) and a nucleocapsid protein (N: 48 kDa; Garwes et al., 1978; Brian et al., 1980; Jacobs et al., 1986). In TGEV-infected cells five subgenomic mRNA species are synthesized. By *in vitro* translation, it has been shown that mRNA3 encodes the E2 protein (Hu et al., 1984; Jacobs et al., 1986). The peplomer protein of coronaviruses is responsible for virus attachment to the cell and membrane fusion (for a review see Sturman and Holmes, 1983). It has been shown recently that neutralizing monoclonal antibodies are directed against the E2 protein of TGEV and that the main neutralizing epitopes are conserved among TGEV strains (Laude et al., 1986; Delmas et al., 1986; Jimenez et al., 1986).

TGEV and feline infectious peritonitis virus (FIPV) are serologically closely related (Pederson et al., 1978; Horzinek et al., 1982), to the degree that monoclonal antibodies raised against TGEV react with FIPV. In this report we present the nucleotide sequence of the TGEV peplomer gene and the deduced amino acid sequence. The results show that the peplomer proteins of TGEV and FIPV (de Groot et al., 1987a) are closely related also at the nucleotide level.

Materials and Methods

Virus, cells and RNA isolation

The Purdue strain of TGEV was plaque-purified and grown on PD5 cells as described previously (Jacobs et al., 1986). Poly (A +) selected RNA was isolated from infected cells and fractionated by isokinetic sucrose gradient centrifugation as described previously (Jacobs et al., 1986).

cDNA synthesis and cloning

cDNA was prepared using a sucrose gradient fraction enriched for mRNA3 as template and calf thymus pentanucleotides as primers. First and second strand synthesis was carried out essentially as described by Gubler and Hoffman (1983); for details see Niesters et al. (1986). Double-stranded cDNA was dC-tailed using 4U terminal transferase (Amersham, England) for 30 s (Maniatis et al., 1982), annealed to G-tailed *Pst*I-cleaved pUC9 vector (Pharmacia, Uppsala, Sweden) and used for transformation (Hanahan, 1983) of *E. coli* strain JM 109 (Yanish-Perron et al., 1985). Colonies containing viral inserts were identified by hybridization using nick-translated FIPV-E2 probes (de Groot et al., 1987a).

Plasmid DNA was isolated from positive recombinants and analyzed by restriction enzyme mapping.

Sequence analysis

After digestion with several restriction enzymes the DNA fragments were separated by agarose gel electrophoresis, isolated by binding to NA-45 paper (Schleicher and Schuell, Dassel, W. Germany) and recloned in bacteriophage M13, mp18 and

mp19 vectors (Yanish-Perron et al., 1985). Single-stranded DNA was isolated and sequenced using the dideoxynucleotide chain termination procedure of Sanger et al. (1977). Data were analyzed using the computer programs of Staden (1982).

Results

Screening and sequencing TGEV E2 specific clones

TGEV and FIPV are serologically closely related and we assumed a nucleotide homology between both viruses within the E2 gene. Two FIPV E2 specific probes representing the 5' and 3' ends of the FIPV E2 gene (de Groot et al., 1987b) hybridized with RNA3, the mRNA for E2 (Jacobs et al., 1986) but not with smaller subgenomic RNAs of TGEV. They were subsequently used to select TGEV E2 specific recombinants (Fig. 1). Twenty-nine cDNA clones hybridized to one probe, the two clones pA6 and pB9 to both FIPV-probes. Finally three clones were selected for recloning into bacteriophage M13 (pB1, pA6 and pB9 respectively) and used for sequence analysis (Fig. 1).

The sequencing strategy is depicted in Fig. 1. With the exception of the extreme 5'-end the sequence was determined using at least two different cDNA clones. To avoid extensive subcloning into M13, synthetic oligonucleotides were used to prime the sequence reactions. The nucleotide sequence is given in Fig. 2.

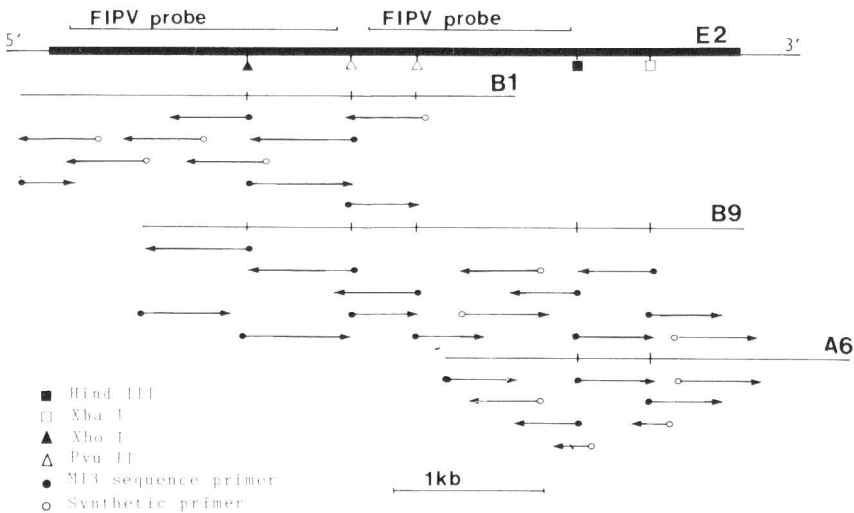


Fig. 1. The restriction sites and sequence strategy of the TGEV peplomer gene. The position of recombinant cDNA clones and the two FIPV probes used for screening the cDNA library are indicated.

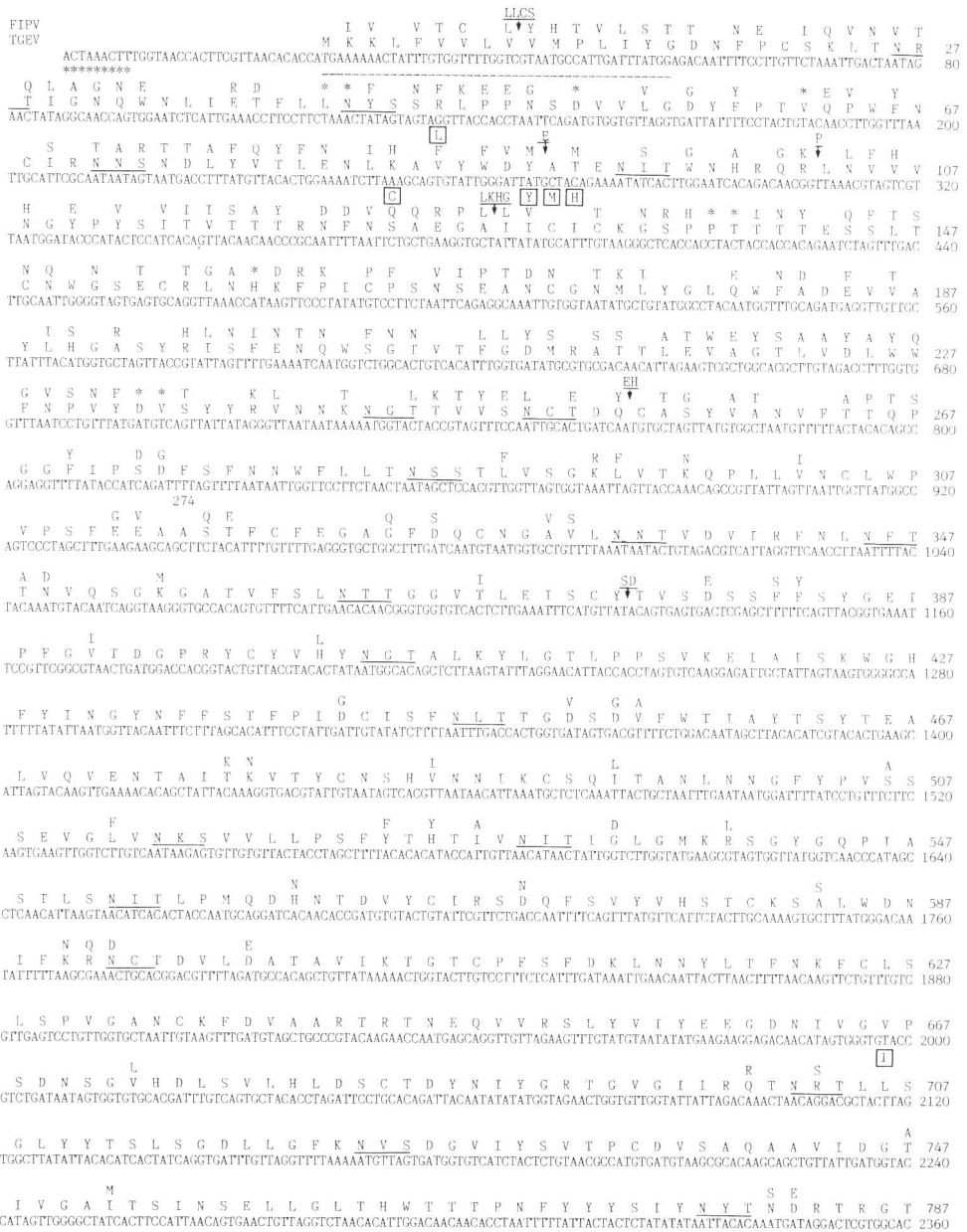


Fig. 2. (A) Nucleotide and predicted amino acid sequences of the peplomer protein of TGEV and comparison with the protein sequence of FIPV; only sequence differences are indicated; an arrow indicates the position of an insertion in the FIPV sequence. Amino acid differences with the TGEV E2 sequence of Raschaert and Laude (1987) are boxed. The repeating sequences in the intergenic boundaries are indicated by asterisks. Potential glycosylation sites (Asn-X-Ser or Asn-X-Thr except when X = Pro) are underlined. Hydrophobic region (position 1389-1411) and the signal peptide (position 1-14) are indicated by a broken line. (B) Positions of amino acid substitutions (vertical bars) in the FIPV peplomer gene compared to the peplomer gene of TGEV. The arrow indicates residue 274.

^V
 A I D S N D F D C E P V I T Y S N I G V C K N G A ^L F V F I N V T H S D G D V Q P 827
 TGCAATGACAGTAAATGATTTGATTTGTAACCGTTCACCAACTTATCCATATCCGTCGAAGTCCGAATATATTCAGGTTTACACCTACACCACTGTCAATAGACTGTTCAAGATATGTTGTAATGGTAA 2600

I S T G N V T I P T N F T I S V Q V E Y I Q V Y T T P V S I D C S R Y V C N G N 867
 AAATGCGCTTAAATGAGCATCTGTTGAAGGATTCAGTATGTCAGAACTTTAGACCTATTTACAAGAAATGCGCTAAATATAGGGGTTCTTGGCTACAGGTCCTAAAATACATACTTCC 2840

P R C N K L L T Q Y V S A C Q T I E Q A L A M G A R L E N M E V D S M L F V S E 907
 CCGTAGGTTGTAACAAATGTTTAAACAGATACTTTCTGCGATGTCAAACTATGAGCAAGGACCTGCAATGGGTGCCAGACTTGAAAAATGGAGGTTGATTCATGTTGTTTCTGTA 2720

N A L K L A S V E A F N S S E T L D P I Y K E W P N I G G S W L E G L K Y I L P 947
 AAATGCGCTTAAATGAGCATCTGTTGAAGGATTCAGTATGTCAGAACTTTAGACCTATTTACAAGAAATGCGCTAAATATAGGGGTTCTTGGCTACAGGTCCTAAAATACATACTTCC 2840

S H N S K R K Y R S A T E D L L F D K V V T S G L G T V D E D Y K R C T G G Y D 987
 TGCCATAATAGCAACCTAAATGATGTTGAGCTATAGGAGCTTGGCTTTTGTAAAGGTTTGAACATCTGGTTTGGTACAGTTGATGAAGATATAAAGCTGTGACAGGTTGTTATGA 2960

I A D L V C A Q Y Y N G I M V L P G V A N A D E M T M Y T A S L A G G I T L G A 1027
 CATAGCTGACTTATGATGTCGATAGCTATAATGCGCATGATGTCGCTACCTGGTGGCTAAATGCTGACAAAATGACTGTTACACAGCATCCGCTGGCAGTGGTATAACATGAGGTCG 3080

L G G G A V A I P P F A V A V Q A R L N Y V A L Q T D V L N K N Q Q I L A S A F N 1067
 ACTGCGTGGAGGCGCGCTGCTATACCTTTTTCAGCTAGCAGTTCAGCGCTAGACTTAAATGATGTCCTCTACAAAATGATGTAFTGAACAAAACCCAGCAGATTCCTGGCTAGGCTTTCGA 3200

Q A I G N I T Q S F G K V N D A I H Q T S R G L A T V A K A L A K V Q D V V N I 1107
 TCAAGCTATGTTAACAATGACAGCTTATGCTAAGGTTAATGATGCTATAGACTCAAAATCAGCAGGTTGCTGCTACTGTTGCTAAGCAATGCGCAAAAGTGAAGATGTTGTAACAT 3320

Q G Q A L S H L T V Q L Q N N F Q A T S S S I S D T Y N R L D E L S A D A Q V D 1147
 ACAAGGCGAGCTTTAAGCCACCTAACACGTAGCAATGCAAAAATAATTCAGAGCCATGATAGTTCTCAATGAGCATTTACAATAGGCTTGAAGAAATGAGTGGTATGCACAGTTGA 3340

R L I T T G R L T A I N A F V S Q T L T R Q A E V R A S R Q L A K D K V V E C V R 1187
 CAGGCTGATCAGCAAGACTTACAGCACCTAATGATGTTGCTGACAGCTTAACCAAGACAAGCGGAGGTTAGGGCTAGTAGACAACCTGCGCAAGACAAGGTTAATGAATGGCTAG 3560

S Q S Q R F G F C G N G T H L F S L A N A A P N G M I F E H T V L L P T A Y E T 1227
 GTCGACGTCGAGAGATTCGGATTCGCTGTAATGCTACACATTTGCTTCTGCTGCGAAATGCAAGCACCACCAATGGCATGATTTCTTTCAGACAGCTGCTATTCACCAAGGCTTATGAAAC 3680

V T A W P G I C A S D G D R T F G L V V K D V Q L T L F R N L D D K E F Y L F P R 1267
 TGTGACTGCTGCGCGGTTATTTGCTGTCAGATGTAATGCGACCTTTGGACTTGTGCTTAAAGATGTCAGCTGACTTGTGTTGCTAGATGACAGTTCATATTCGACCCCGAG 3800

T M Y Q P R V A T S S D F V Q I E G C D V L F V N A T V S D L P S I I P D Y I D 1307
 AAGTATGATCAGCCTGAGTTGCAACTGACTTCTGACTTTGTTGTTCAAAATGAAAGGTCGATGCTGCTTGTGTTAATGCAACTGTAAGTGAATTCGCTAGTATATACCTGATATATGA 3920

I Y Q T V Q D I L E N F R P N W T V P E L T F D I F N A T Y L N L T G E T D D L 1347
 TATTAATCAGACTGTTCAAGACAATATAGAAAATTTAGACCAAAATGAGCTGTACCTGAGTTGACATTTGACATTTTAAAGGCAAGCTATTTAAAGCTGAGTGGTGAATTTGATGACTT 4040

E F R S E K L H N T T V E L A I L I D N T Y N T L V N L E W L N R I E T Y V K W 1387
 AGAATTTAGTTCAGAAAAGCTACATAACCCACCTGATAGAACTTGCATTTGACACAAATTAACAAATACATTAAGTCAATGCTGAAATGCGTCAATAGAAAATGAAAACCTATGTAATAATG 4160

P W Y V W I L I G L V V I F C I P L L L F C C C S T G C C G C I G C L G S C C H 1427
 GCGCTGCTATGTTGCTACTAATAGGCTTAGTAGTAATPTTTPNGATACCAATFACCTGCTAATTTGCTGTTGATGACAGGTTGCTGTTGGATGCAATAGGTTGTTAGGAAGTTGTTGCTCA 4280

S I C S R R Q P E N Y E P I E K V H V H * 1447
 CFCATATGATGATGAAAGACAAATTTGAAAATTAGGAACTTGA AAAAGTCCAGTCCGATTAATTTAAAAAGTTAATTCATCACTGCTATAATAGCAGTTGTTCTGCTAGAGAAAT 4400

TTTGTTAAGGATGATGAATAAAGCTTTTAAAGCACTAACTT ***** 4441

B

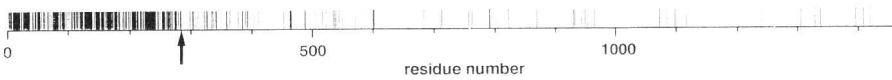


TABLE 1
HOMOLOGY BETWEEN THE E2 GENES OF FIPV AND TGEV.

	Residues 1-274 (%)			Residues 275-1447 (%)		
Amino acid homology	30			94		
Nucleotide homology	39			93		
	position codon			position codon		
Observed nucleotide substitutions (%)	1	2	3	1	2	3
	20	19	22	1	1	5

Amino acid sequence

An open reading frame (ORF) with the potential to encode a protein of 1447 amino acids was found. A putative signal sequence is present at the amino terminal end (residues 1-14; Fig. 2). To identify and characterize hydrophobic segments which may penetrate the lipid bilayer we have used the algorithm of Eisenberg et al. (1984). In addition to the signal sequence a segment with a mean hydrophobicity of 0.91 was detected from position 1389-1411 (Fig. 2). This part of the protein may be responsible for the anchoring of the protein in the lipid bilayer of the virion. Thirty-two potential glycosylation sites have been found (Asn-X-Ser or Asn-X-Thy, X = Pro) clustered in the amino terminal end of the protein and the carboxyl terminal end immediately upstream of the transmembrane anchor (Fig. 2). Many cysteine residues are located downstream of the transmembrane anchor.

Comparison of the TGEV and FIPV nucleotide and amino acid sequences

Both the nucleotide and predicted amino acid sequences of the E2 genes of TGEV and FIPV were compared. Data on the homology of the E2 genes of TGEV and FIPV are given in Table 1; the differences on the amino acid level are indicated in Fig. 2. Only 74 amino acid substitutions are found downstream of position 274 (homology of 94%); some of the mutations are clustered (Fig. 2B). There is a striking difference between E2 protein of FIPV and TGEV at the N-terminal part (residues 1-274), where a homology of only 30% is found.

Discussion

An ORF in the E2 gene of TGEV of 1447 amino acids was identified. The calculated mol. wt. of 159,574 is in good agreement with the size of the protein found in tunicamycin-treated cells or after in vitro translation of mRNA3 (Hu et al., 1984; Jacobs et al., 1986). The M_r of the peplomer protein is approximately 200 000, suggesting that most of the 30 potential glycosylation sites (Fig. 2) carry carbohydrate side chains, each of which adds about 2000 to the mol. wt. of the protein (Neuberger

et al., 1972). Recently Rasschaert et al., 1987 have also obtained the TGEV E2 sequence and only 7 amino acid differences have been found (Fig. 2).

The ORF is flanked by a repeat of the 9 nucleotides 5'-ACTAAACTT-3' (Fig. 2). The same repeat has been found at the boundaries of the E2 gene of FIPV (de Groot et al., 1987b) as well as upstream of the TGEV nucleocapsid gene (Kapke et al., 1986). Similar sequence homologies have been found adjacent to the ORFs in the genomes of mouse hepatitis virus (MHV) and infectious bronchitis virus (IBV). Presumably these sequences are recognition signals used in the discontinuous transcription mechanism of coronaviruses (Spaan et al., 1983; Brown and Bournsnel, 1984). The amino acid sequence derived from the E2 gene has the characteristic features of coronavirus peplomer proteins with regard to the signal sequence, membrane anchor and the distribution of the glycosylation sites (Binns et al., 1985; Schmidt et al., 1987); however the TGEV and FIPV peplomer proteins lack proteolytic cleavage sites, which are present in IBV and MHV.

Cavanagh, (1983) has proposed a model for the coronaviral peplomer in which the C-terminal half of the protein forms its stalk and the N-terminal half of the protein its bulbous part. Recently, de Groot et al., (1987a) have postulated a model in which a coiled-coil structure forms the connection between the globular part of the peplomer protein and the viral membrane. This model is based on the occurrence of heptad repeats, i.e., a periodicity (a-b-c-d-e-f-g) in which amino acids a and d are hydrophobic. Heptad repeats are indicative of a coiled-coil structure in which the hydrophobic residues form the interface between interlocking α -helices. Two such repeats were detected in E2 of FIPV (amino acid residues 1067-1149 and 1334-1380). There are three amino acid differences between TGEV and FIPV in the first region and none in the second. All amino acid substitutions leave the heptad repeat intact.

In TGEV four main antigenic epitopes (A, B, C and D) have been found (Delmas et al., 1986). Neutralization relevant epitopes A and B are highly conserved in TGEV (Jimenez et al., 1986, Delmas et al., 1986) and monoclonal antibodies recognizing both epitopes also neutralized FIPV infectivity (data not shown). Hence these conserved epitopes should be located downstream of position 274, where FIPV and TGEV possess a homology of 94%. Neutralization epitopes in antigenic sites C and D of TGEV are less well conserved (Delmas et al., 1986). Three neutralizing monoclonal antibodies recognizing sites C and D were not able to neutralize FIPV infectivity. Probably these epitopes are situated where amino acid substitutions between FIPV and TGEV are clustered. In the absence of selection, mutations will occur randomly. Downstream of position 274 there is a preference for mutations in the third position of a codon (Table 1), where only 28% of the nucleotide substitutions result in a amino acid replacement (Masotoshi and Gojobozi, 1986). In this part of the protein there seems to be a selection for conservation of the amino acid sequence, rather than for antigenic variation.

The high sequence divergence at the N-terminus was unexpected and is not caused by selection of neutralizing antibodies. A possible explanation would be that FIPV arose by recombination of TGEV with a related virus - related, because there is still a homology of 30% in residues 1-274. In vitro, a high frequency of

recombination has been described for murine corona viruses (Makino et al., 1986; Lai et al., 1985).

We have compared the 274 amino acids using the FASTP program (Lipman and Pearson, 1985) with the NBRF sequence library and with the MHV and IBV structural proteins. Excepting the 30% homology with FIPV E2 protein, no homologies have been found. However, sequences of other related coronaviruses such as canine coronavirus are not available yet.

Acknowledgements

This study was supported by a research grant from Duphar B.V., Weesp, The Netherlands. We thank Renée Ter Haar for the FIPV neutralisation experiments, Hans Lenstra for the comparison of the E2 amino acid sequence with the NBRF library, and for many helpful discussions, Diane Ruysendael for synthesis of the oligonucleotides, and Willem Luytjes for the advice on computer analysis. We also thank Dr. H. Laude for sending us the TGEV E2 sequence data prior to publication.

References

- Binns, M.M., Boursnell, M.E.G., Cavanagh, D., Pappin, D.J.C. and Brown T.D.K. (1985) Cloning and sequencing of the gene encoding the spike protein of the coronavirus IBV. *J. Gen. Virol.* 66, 719–726.
- Brian, D.A., Dennis, D.E. and Guy, J.S. (1980) Genome of porcine transmissible gastroenteritis virus. *J. Virol.* 34, 410–415.
- Brown, T.D.K. and Boursnell, M.E.G. (1984) Avian infectious bronchitis virus genomic RNA contains sequence homologies at the intergenic boundaries. *Virus Res.* 1, 15–24.
- Cavanagh, D. (1983) Coronavirus IBV: Structural Characterization of the spike protein. *J. Gen. Virol.*, 64, 2577–2583.
- Delmas, B., Guelfi, J and Laude, H. (1986) Antigenic structure of transmissible gastroenteritis virus. II. Domains in the peplomer glycoprotein. *J. Gen. Virol.* 67, 1405–1418.
- Eisenberg, D., Schwarz, E., Komazomy, M. and Wall, R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* 179, 125–142.
- Garwes, D.J., Lucas, M.H., Higgins, D.A., Pike B.V. and Cartwright, S.T. (1978/1979) Antigenicity of structural components from porcine transmissible gastroenteritis virus. *Vet. Microbiol.* 3, 179–190.
- De Groot, R.J., Luytjes, W., Horzinek, M.C., van der Zeijst, B.A.M., Spaan, W.J.M. and Lenstra, J.A. (1987a) Evidence for a coiled-coil structure in the spike proteins of coronaviruses. *J. Mol. Biol.* 197, in press.
- De Groot, R.J., Maduro, J., Lenstra, J.A., Horzinek, M.C., van der Zeijst, B.A.M., and Spaan, W.J.M. (1987b) cDNA cloning and sequence analysis of the gene encoding the peplomer protein of feline infectious peritonitis virus. *J. Gen. Virol.* 68, in press.
- Gubler, U. and Hoffman, B.J. (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.
- Horzinek, M.C., Lutz, H. and Pedersen, N.C. (1982) Antigenic relationships among homologous structural polypeptides of porcine, feline and canine coronaviruses. *Infect. and Immunity* 37, 1148–1155.

- Hu, S., Bruszewski, J., Boone, T. and Souza, L. (1984) Cloning and expression of the surface glycoprotein gp195 of porcine transmissible gastroenteritis virus. In: *Modern Approaches to Vaccines* (Chanock, R.M. and Lerner, R.A., eds.) pp. 219–223. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Jacobs, L., van der Zeijst B.A.M. and Horzinek, M.C. (1986) Characterisation and translation of transmissible gastroenteritis virus mRNAs. *J. Virol.* 57, 1010–1015.
- Jimenez, G., Correa, I., Melgosa, M.P., Bullido, M.J. and Enjuanes, L. (1986) Critical epitopes in transmissible gastroenteritis virus neutralisation. *J. Virol.* 60, 131–139.
- Kapke, P.A. and Brian, D.A. (1986) Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. *Virology* 151, 41–49.
- Lai, M.M.C., Baric, R.S., Makino, S., Keck, J.G., Egbert, J., Leibowitz, J.L. and Stohman, S.A. (1985) Recombination between nonsegmented RNA genomes of murine coronaviruses. *J. Virol.* 56, 449–456.
- Laude, H., Chapsal, J.M., Guelfi, J., Labiau, S. and Grosclaude, J. (1986) Antigenic structure of transmissible gastroenteritis virus. I. Properties of monoclonal antibodies directed against virion proteins. *J. Gen. Virol.* 67, 119–130.
- Lipman, D.J. and Pearson, W.R. (1985) Rapid and sensitive protein similarity searches. *Science* 227, 1435–1441.
- Makino, S., Keck, J.G., Stohman, S.A. and Lai, M.M.C. (1986) High frequency RNA recombination of murine coronaviruses. *J. Virol.* 57, 729–737.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular cloning. A laboratory manual.* Cold Spring Harbor Laboratory.
- Masatoshi, N. and Gojobozi, T. (1986) Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426.
- Neuberger, A., Gottschalk, A., Marshall, R.O. and Spiro, R.G. (1972) Carbohydrate peptide linkages in glycoproteins and methods for their elucidation. In: *The glycoproteins: their Composition, Structure and Function* (A. Gottschalk, ed.) pp. 450–490. Elsevier, Amsterdam.
- Niesters, H.G.M., Lenstra, J.A., Spaan, W.J.M., Zijderfeld, A.J., Bleumink-Pluym, N.M.C., Hong, F., Scharrenburg, G.J.M., Horzinek M.C. and van der Zeijst, B.A.M. (1986) The peplomer protein sequence of the M41 strain of coronavirus IBV and its comparison with Beaudette strains. *Vir. Res.* 5, 253–263.
- Pederson, N.C., Ward, J., Mengeling, W.L. (1978) Antigenic relationship of feline infectious peritonitis virus to coronaviruses of other species. *Arch. Virol.* 58, 45–53.
- Rasschaert, D. and Laude, H (1987) The predicted primary structure of the peplomer E2 of porcine coronavirus TGEV. *J. Gen. Virol.* 68, in press.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- Schmidt, J., Skinner, M., Siddell, S. (1987) Nucleotide sequence of the gene encoding the surface projection glycoprotein of coronavirus MHV-JHM. *J. Gen. Virol.* 68, 47–56.
- Spaan, W.J.M., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst B.A.M. and Siddell S.G. (1983) Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO. J.* 2, 1893–1844.
- Staden, R. (1982) Diagon: an interactive graphics program for comparing and aligning nucleic acid for amino acid sequences. *Nucl. Acid. Res.* 10, 2951–2961.
- Sturman, L.S. and Holmes, K.V. (1983) The molecular biology of coronaviruses. *Adv. Virus Res.* 28, 35–112.
- Yanish-Perron, C., Vierra, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC 19 vectors. *Gene* 33, 103–119.

(Received 16 April 1987; revision received 14 July 1987)