Molecular Microbiology (1988) 2(4), 497–505

The integral membrane protein from a virulent isolate of transmissible gastroenteritis virus: molecular characterization, sequence and expression in *Escherichia coli*

P. Britton,* R. S. Cármenes,† K. W. Page and D. J. Garwes
Division of Microbiology, AFRC Institute for Animal Health, Compton Laboratory, Compton, Newbury, Berkshire RG16 0NN, UK.

Summary

Subgenomic mRNA from a virulent isolate of porcine transmissible gastroenteritis virus (TGEV) was used to produce cDNA clones. Part of a new clone and a previously reported clone were sequenced and used to construct the viral gene for integral membrane protein. A single open reading frame (ORF) encoding a polypeptide of 262 amino acids, relative molecular mass (M_r) 29459, was identified. The positive identification of the polypeptide as the integral membrane protein was demonstrated by the production in *E. coli* of a chimaeric protein comprising most of the ORF encoding the M_r 29459 polypeptide and β-galactosidase. The chimaeric protein reacted with a specific monoclonal antibody to viral integral membrane protein and antibodies raised against the chimaeric protein immune precipitated the viral protein. Comparison with the sequence of an avirulent isolate indicates amino acid residues that may be important in pathogenicity.

Introduction

TGEV is a coronavirus that causes gastroenteritis in pigs, resulting in a high mortality in neonates. TGEV belongs to the family Coronaviridae, a large group of pleomorphic enveloped viruses with a positive-stranded RNA genome. The coronavirus proteins are expressed from a 'nested' set of subgenomic mRNAs which have common 3′ termini but different 5′ extensions. The region of each mRNA responsible for the expression of a protein appears to correspond to the 5′-terminal region that is absent on the preceding smaller species. Mouse hepatitis virus (MHV) and infectious bronchitis virus (IBV) mRNA species contain identical short non-coding sequences at their 5′ ends which appear to be joined to the sequences encoding the viral genes by discontinuous transcription. A consensus sequence identified upstream of each gene/ORF 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; Shieh *et al.*, 1987). It has been postulated that a heptameric sequence, ACTAAAC (Britton *et al.*, 1988), or a hexameric sequence, CTAAAC (Kapke and Brian, 1986; Rasschaert *et al.*, 1987a; 1987b), may be involved in the binding of the TGEV RNA polymerase-leader complex for the transcription of the mRNA species from the negative form of the RNA genome.

The TGEV virion contains three major structural polypeptides: a surface glycoprotein (spike or peplomer protein) with a monomeric M_r of 200000, a glycosylated integral membrane protein observed as a series of polypeptides of M_r 28000–31000 and a basic phosphorylated protein (the nucleoprotein) of M_r 47000 associated with the viral genomic RNA (Garwes and Pocock, 1975). TGEV-infected cells have, in addition to the genomic RNA, six species of subgenomic mRNA (Britton *et al.*, 1986). Expression and sequencing studies have shown that the two smaller species (1.7kb and 0.7kb) contain the nucleoprotein gene and a gene encoding a polypeptide of M_r 9000 (Britton *et al.*, 1986; Kapke and Brian, 1986; Rasschaert *et al.*, 1987b; Britton *et al.*, 1988). The mRNA species of 2.6kb and the largest mRNA species (8–11.2kb) appear to encode the integral membrane protein and the peplomer protein (Britton *et al.*, 1986; Jacobs *et al.*, 1986; Kapke *et al.*, 1987; Laude *et al.*, 1987; Rasschaert and Laude, 1987; Rasschaert *et al.*, 1987a; 1987b). The other mRNA species of 3kb and 3.9kb (Britton *et al.*, 1986; Jacobs *et al.*, 1986) have had no product assigned to them from either infected cells or virions. Jacobs *et al.* (1986) identified a product of M_r 24000 from *in vitro* translation of the 3.9kb mRNA species (Purdue strain) in rabbit reticulocyte lysate.

The integral membrane proteins from other coronaviruses, including IBV (Boursnell *et al.*, 1984), MHV (Armstrong *et al.*, 1984; Pfleiderer *et al.*, 1986), bovine coronavirus ((BCV) Lapps *et al.*, 1987) and the avirulent Purdue strain of TGEV (Kapke *et al.*, 1987; Laude *et al.*, 1987) have been sequenced. The *N*-terminal region of the

Received 7 March, 1988. †Present address: Departamento de Bioquimica, Universidad de Oviedo, Oviedo, Spain. *For correspondence.

proteins from MHV and BCV contain *O*-linked glycans in which the sugar moieties are joined to serine and threonine residues (Niemann and Klenk, 1981; Niemann *et al.*, 1984; Lapps *et al.*, 1987). The integral membrane proteins from TGEV (Garwes *et al.*, 1984; Jacobs *et al.*, 1986) and IBV (Stern and Sefton, 1982a) have been shown to have *N*-linked glycans. The integral membrane proteins from MHV, IBV and BCV do not appear to have an *N*-terminal cleaved signal sequence (Rottier *et al.*, 1986; Stern and Sefton, 1982b; Lapps *et al.*, 1987). However, Laude *et al.* (1987) demonstrated that the integral membrane protein from the Purdue strain of TGEV contained a cleaved 17-amino-acid signal sequence. A model for the membrane topology of the integral membrane protein for coronaviruses has been postulated from a combination of biochemical and primary sequence data from MHV (Armstrong *et al.*, 1984; Rottier *et al.*, 1986).

Here we describe the cloning and sequence of the integral membrane protein gene of a virulent field isolate of TGEV, provide experimental confirmation that the gene encodes the integral membrane protein and compare its primary structure with that of three other coronaviruses.

Results

Cloning from TGEV mRNA species

TGEV poly(A)-containing mRNA species were isolated from virus-infected LLC-PK1 cells and used for the synthesis of cDNA. The production of plasmid pF4F-36 was described by Britton *et al.* (1988). Plasmid pTG22 was produced using a synthetic oligonucleotide as primer for cDNA synthesis. Plasmid pF4F-36 was found to hybridize, by Northern blot analysis, to all the TGEV mRNA species except the 0.7 kb species, postulated to express a polypeptide of M_r 9000 found at the 3′ end of the viral genome (Britton *et al.*, 1988). Plasmid pTG22 did not hybridize to either the 0.7 kb or 1.7 kb mRNA species, indicating that it originated within the 2.6 kb mRNA species previously shown, by *in vitro* translation in rabbit reticulocyte lysates, to express the TGEV integral membrane protein (Britton *et al.*, 1986; Jacobs *et al.*, 1986). From the size and position of the cDNA insert on the TGEV genome, pF4F-36 was deduced to contain about 80% of the integral membrane protein. Plasmid pTG22 is 2.5 kbp long and from its restriction map was shown to overlap pF4F-36 (Fig. 1) and extend 4.6 kb into the TGEV genome, thus completing the cloning of the integral membrane protein gene.

Construction of the TGEV integral membrane protein gene

A 0.44 kbp *Nsi*I fragment from pTG22, containing restriction sites at one end that overlapped with pF4F-36, was purified and ligated to a 0.5 kbp *Nsi*I-*Bal*I fragment from pF4F-36 to produce a 0.94 kbp contiguous piece of TGEV cDNA (Fig. 2). The 0.94 kbp fragment was digested using *Hgi*AI to produce a 0.83 kbp fragment containing the complete integral membrane protein gene. The *Hgi*AI site, 22 bp upstream of the integral membrane protein gene initiation site, deduced from sequence data, was end-repaired using the Klenow fragment of *E. coli* DNA polymerase I, and *Bam*HI linkers were added. The resulting *Bam*HI cassette was cloned into pBR322 as described by Britton *et al.* (1988) for the nucleoprotein gene. A

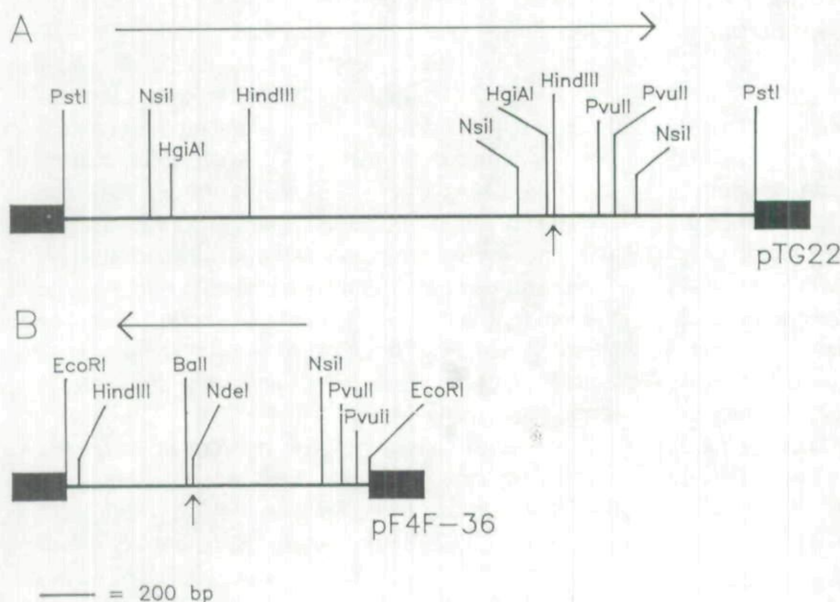

Fig. 1. Restriction endonuclease map of plasmids pTG22 (A) and pF4F-36 (B). The thick black lines represent DNA from the vector pUC9. The large arrows show the direction of the TGEV cDNA inserts, as compared to the viral genome, when mapped in pUC9. The small arrows illustrate (A) the start of the integral membrane protein gene, and (B) the start of the nucleoprotein gene.

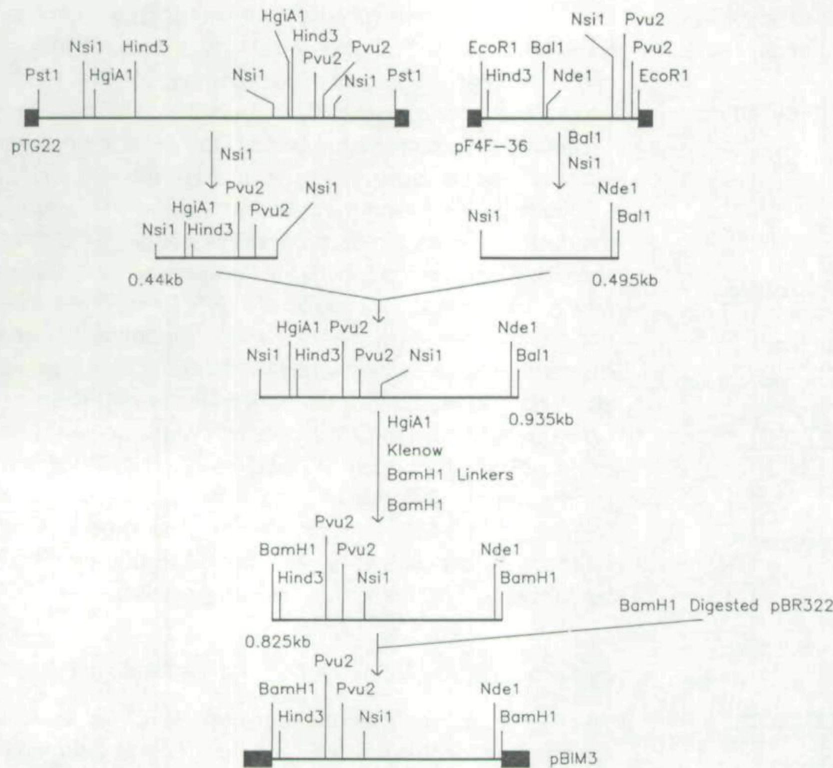

Fig. 2. Schematic representation of the construction of the TGEV integral membrane protein gene *Bam*HI cassette. The thin lines represents the TGEV cDNA and the thick line represents part of pUC9 DNA, for PTG22 and pF4F-36, and pBR322 for pBIM3.

recombinant plasmid, pBIM3 (Fig. 2), consisting of the 0.83 kbp *Bam*HI fragment in pBR322, was used as a source of the integral membrane protein gene.

Sequencing of TGEV cDNA

Restriction fragments obtained from plasmids pF4F-36 and pBIM3 were subcloned into M13mp vectors and sequenced in both directions. The cDNA was translated in all six reading frames and translation in the virus sense strand revealed an ORF of 786 bp. The corresponding DNA sequence, 21 bp from the 5′ end of the ORF to the start of the TGEV nucleoprotein gene, present in pF4F-36 (Britton *et al.*, 1988), is illustrated in Fig. 3. Two other ORFs were found in the viral sense strand: one was composed of 20 amino acids positioned between nucleotides 57–116, and the other comprised 43 amino acids between nucleotides 141–269, within the integral membrane protein gene. Three ORFs were identified, in the complementary strand: one of 45 amino acids (nucleotides 78–212), one of 62 amino acids (nucleotides 246–431), and another of 33 amino acids (nucleotides 519–617), within the integral membrane protein gene. None of the internal ORFs were preceded by the potential RNA polymerase-leader complex binding site.

The 786 bp ORF, initiating from the ATG at position 22, overlapped both cDNA inserts and encoded a polypeptide of 262 amino acids with a M_r 29 459. The 5′ end of this ORF mapped at 2.5 kb from the 3′ end of the viral genome and, from its length and position, corresponded to the unique region present in the 2.7 kb mRNA species. The difference between molecular weights of the deduced polypeptide and those found in infected cells and virions is due to post-translational processing. The integral membrane protein is often found as a series of polypeptides when analysed by SDS-polyacrylamide gel electrophoresis.

The nucleotide sequence (Fig. 3) revealed the presence, near the 5′ end of the 786 bp ORF, of the heptameric sequence, ACTAAAC, also found 5′ of the ATG sequences at the start of the TGEV nucleoprotein gene and the ORF encoding the polypeptide of M_r 9000 (Britton *et al.*, 1988). The sequence context, CAAA<u>ATG</u>A, about the first ATG of the 786 bp ORF, downstream of the ACTAAAC sequence, is favourable for initiation by eukaryotic ribosomes ((CC) ACC<u>ATG</u>G; Kozak, 1983; 1986). A second ATG, found 60 bp from the first ATG, lies in a less favourable context.

The M_r 29 459 polypeptide, encoded by the 786 bp ORF, contains about 46% hydrophobic amino acid residues spread over five domains. The sequence is about 96% identical to that described by Kapke *et al.* (1987) and Laude *et al.* (1987), who sequenced the integral membrane protein gene from the avirulent Purdue strain of TGEV and postulated that the protein was the TGEV integral membrane protein because of its homology to the

integral membrane protein genes sequenced from IBV (Boursnell *et al.*, 1984) and MHV (Armstrong *et al.*, 1984).

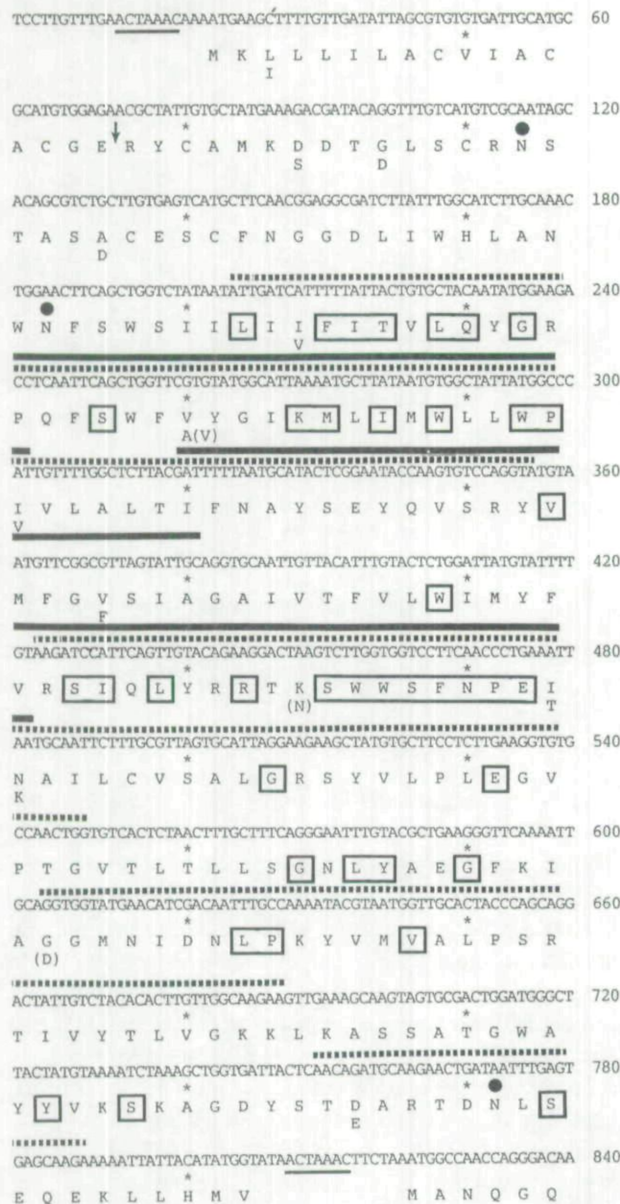

Fig. 3. The nucleotide sequence of the TGEV integral membrane protein gene. The amino acid sequences are the integral membrane protein (nucleotides 22–807) and the start of the nucleoprotein gene (nucleotides 823–840). The ACTAAAC consensus sequence preceding each gene is underlined. Amino acids below the major sequence are substitutions found in the Purdue strain (unbracketed by Kapke *et al.*, 1987 and bracketed by Laude *et al.*, 1987). The boxed amino acids are identical residues found between IBV (Boursnell *et al.*, 1984), MHV (Armstrong *et al.*, 1984) and BCV (Lapps *et al.*, 1987). The small arrow at the *N*-terminus of the protein is the signal peptidase cleavage site, as identified by Laude *et al.* (1987). The black dots above asparagine residues indicate potential *N*-glycosylation sites. The broken lines refer to areas of homology between the four integral membrane proteins derived by dot-matrix analysis (Fig. 5). The thick black line represents the postulated membrane-spanning regions. These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00560.

Laude *et al.* (1987) identified a 17-amino-acid sequence at the *N*-terminal end of the protein as a putative signal sequence that is lost upon incorporation of the TGEV integral membrane protein into cell membranes. The same 17-amino-acid sequence, except for a conservative substitution of an isoleucine for a leucine residue at amino acid residue 3 resulting from a change in the first base of the codon, is encoded from our cDNA sequence. Cleavage of the leader sequence results in a polypeptide of 245 amino acids with a predicted M_r of 27 712, of which 44% of the residues are hydrophobic; this polypeptide has an overall charge of +5 at neutral pH (identical to the Purdue strain but distributed over different amino acids). The size of the cleaved product is in agreement with a polypeptide of M_r 26 000 found by Garwes *et al.* (1984) for the expression of TGEV mRNA species in rabbit reticulocyte lysates or in TGEV-infected cells in the presence of tunicamycin; and with a polypeptide of M_r 25 000 identified by Jabobs *et al.* (1986) using the same analyses.

Expression of the 786 bp ORF as a gene fusion in E. coli

In order to prove that the polypeptide of M_r 29 459 encoded by the 786 bp ORF was the viral integral membrane protein, *Hin*dIII linkers were added to a 0.88 kbp *Fsp*I fragment from plasmid pBIM3 containing 95% of the complete gene. This was then ligated into the *Hin*dIII site of the *lacZ* gene in plasmid pUR290. The 0.88 kbp *Fsp*I fragment consisted of the TGEV integral membrane protein gene 39 bp downstream from the start of the gene to the *Fsp*I site 113 bp from the *Bam*HI site in pBR322. This resulted in removal of DNA encoding most of the amino acids from the signal sequence at the *N*-terminal end of the protein. The recombinant plasmid pURIM-2 expressed a chimaeric protein of M_r 140 000, upon induction with IPTG, which consisted of part of the TGEV integral membrane protein fused to the *C*-terminal end of β-galactosidase. A specific monoclonal antibody to the integral membrane protein from the Purdue strain of TGEV, 3BB3, reacted with the chimaeric protein by immunoblot analysis (Fig. 4A). The chimaeric protein was purified from SDS-polyacrylamide gels and used to raise antibodies in mice. The antibodies immune-precipitated the TGEV integral membrane protein, further confirming that the cDNA encoded the viral integral membrane protein gene (Fig. 4B).

Discussion

The complete integral membrane protein gene (from a British field isolate of TGEV) was cloned over two cDNA fragments and then sequenced. A major ORF of 786 bp was identified in the viral-sense strand. The initiation codon of the ORF was preceded by the heptameric

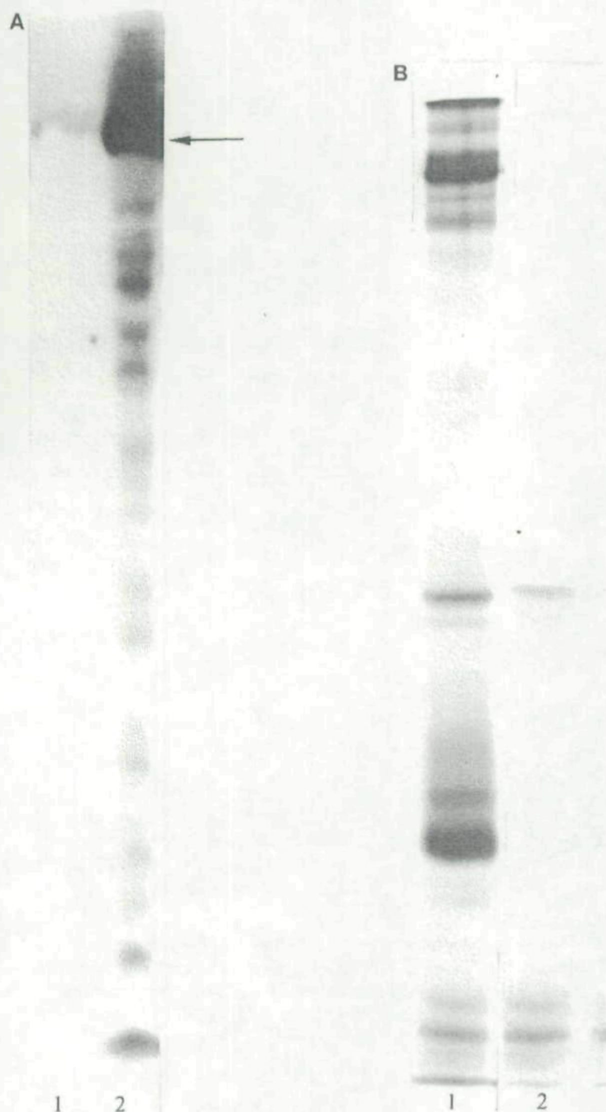

Fig. 4 A. Detection by immunoblotting of the chimaeric protein expressed in *E. coli* cells using the monoclonal antibody 3BB3. The cell extracts were: track 1, JM101(pUR290); and track 2, JM101(pURIM-2). The arrow shows the position of the chimaeric protein. The lower molecular weight polypetides are probably degradation products. B. Fluorograph of [^{35}S]-methionine-labelled TGEV polypeptides detected by immune precipitation using mouse anti-TGEV serum (track 1); control mouse serum (track 2); and mouse anti-chimaeric protein serum (track 3). The arrow shows the position of the integral membrane protein.

sequence, ACTAAAC, which was previously described as preceding both the TGEV nucleoprotein gene and an ORF encoding a polypeptide of M_r 9068 at the 3′ end of the viral genome. The 786 bp ORF encoded a primary translation product of 262 amino acids with a M_r 29459. The first 17 amino acids fulfil the criteria of being a eukaryotic signal sequence, having a net charge following the *N*-terminus and a hydrophobic uncharged region of 14 amino acids (McGeoch, 1985; Von Heijne, 1986). The cleavage site, identified by Laude *et al.* (1987) for the Purdue strain of TGEV, is located between a glutamic acid and an arginine residue. The signal sequence is predicted to fall between the glycine (16) and glutamic acid (17) residues (score = 10.2) by the weight-matrix rule of Von Heijne (1986), which predicts 75 to 80% of eukaryotic signal sequences.

The ORF encoding the polypeptide of M_r 29459 was confirmed as the TGEV integral membrane protein gene by constructing a fusion between most of the 786 bp ORF and the *E. coli lacZ* gene in a similar way to that described for the TGEV nucleoprotein gene (Britton *et al.*, 1987). The gene fusion resulted in the synthesis, in *E. coli* cells, of a polypeptide of M_r 140000 composed of β-galactosidase with most of the TGEV integral membrane protein gene attached to the *C*-terminal end. A monoclonal antibody, 3BB3, specific for TGEV integral membrane protein interacted with the chimaeric polypeptide, and antibodies raised against the chimaera precipitated the integral membrane protein from TGEV-infected cells. This confirmed that the 786 bp ORF, predicted to direct the synthesis of a polypeptide of M_r 29459, was the viral integral membrane protein gene.

Comparison of the M_r 29459 product, using a dot-matrix analysis program (DIAGON; Staden, 1982), with the integral membrane proteins from MHV, IBV and BCV in a pairwise manner revealed remarkable homology between the proteins (Fig. 5). There are four major areas of homology between the integral membrane proteins of TGEV (FS772/70), MHV (A59) and BCV (Mebus), whereas there are only two regions of homology between TGEV and IBV (Beaudette) as also seen between IBV and either MHV or BCV (Fig. 5). The homology between MHV and BCV is almost 100%, indicating that the two viruses probably share recent common ancestral evolution. From the homologies between the integral membrane proteins, this paper, and the nucleoproteins, Britton *et al.* (1988), it is clear that TGEV and MHV share a more recent ancestral evolution than either virus with IBV and it will be interesting to compare homologies with other coronaviruses once their gene sequences have been completed. Direct alignment of the integral membrane proteins from TGEV strains FS772/70 (virulent) and Purdue (avirulent) with MHV (A59), BCV (Mebus) and IBV (Beaudette) using the GAP program (Devereux *et al.*, 1984) identified the position of identical amino acids between different viruses or in some cases between all four viruses (Fig. 6), which are indicated on the deduced amino acid sequence of TGEV integral membrane protein (Fig. 3). The optimal alignment identified a perfectly conserved eight-amino-acid sequence, SWWSFNPE, between all four viruses (Figs 3 and 6). The alignment of the amino acid sequences allowed the identification of the three membrane-spanning regions for TGEV by deduction from the regions identified for MHV (Rottier *et al.*, 1986). These are shown as thick black lines

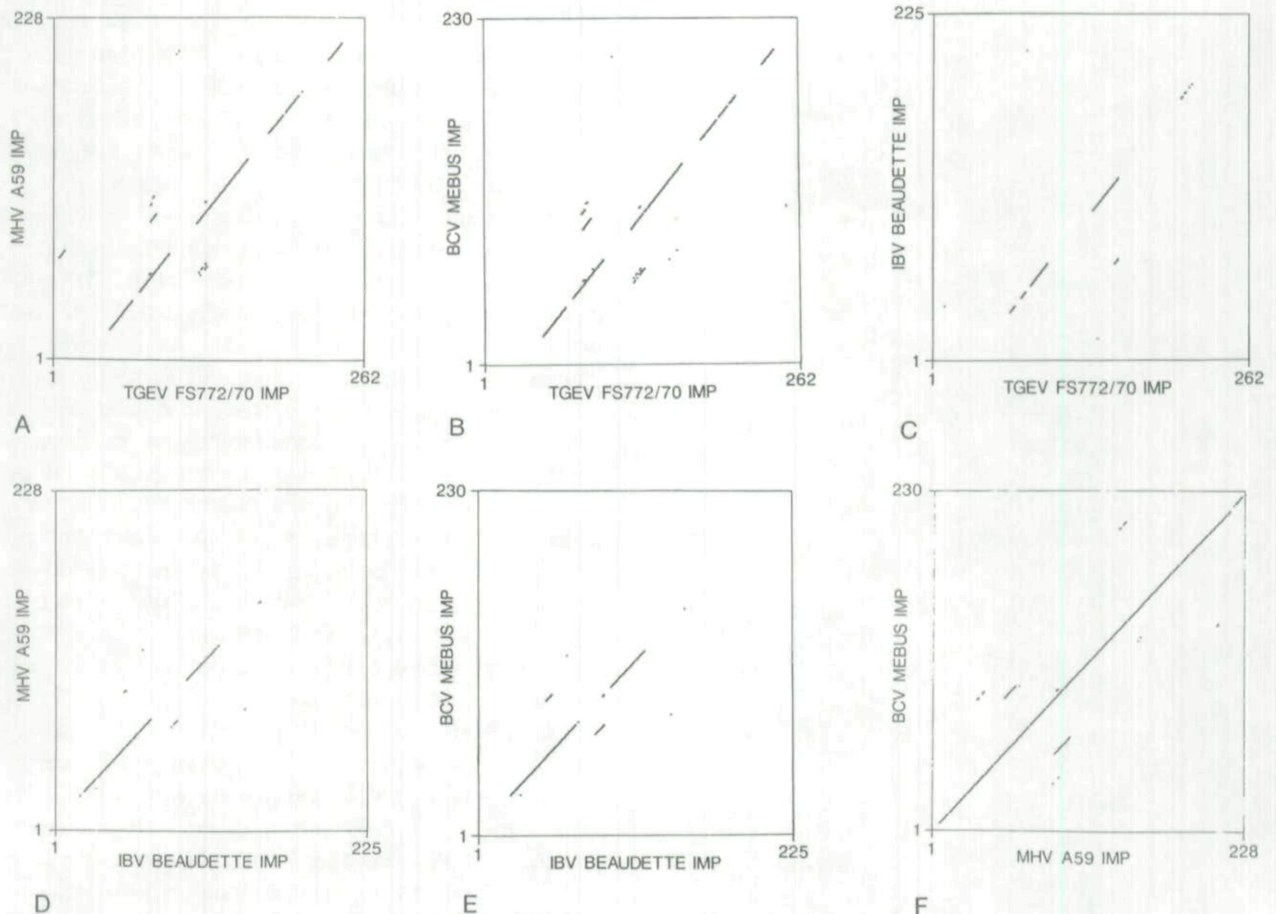

Fig. 5. Comparison of the sequences of coronavirus integral membrane proteins (from four different viruses) that fall into three antigenically distinct subgroups, using DIAGON (Staden, 1982) A, MHV A59 and TGEV FS772/70; B, BCV Mebus and TGEV FS772/70; C, IBV Beaudette and TGEV FS772/70; D, MHV A59 and IBV Beaudette; E, BCV Mebus and IBV Beaudette; F, BCV Mebus and MHV A59. MHV A59 (Armstrong *et al.*, 1984), IBV Beaudette (Boursnell *et al.*, 1984) and BCV Mebus (Lapps *et al.*, 1987). The comparisons used a window of 15 residues with a score of 180.

on Figs 3 and 6. The regions consist of very similar amino acids between TGEV and MHV, indicating that the regions are well conserved between viruses, with very few changes between strains of the same virus. Analysis of the hydropathicity (Kyte and Doolittle, 1982) of the TGEV integral membrane protein identified five hydrophobic regions. The first region, covering amino acids 1–17, is the putative signal sequence. The next three regions, covering amino acids 54–74, 80–100 and 114–134, are the three postulated membrane-spanning regions identified from Fig. 6. The fifth hydrophobic region, covering amino acids 150–220, is the one postulated to interact with the inside of the viral envelope. The remaining 42 residues at the *C*-terminus have an overall hydrophilic nature similar to the region seen in MHV, and may interact with the nucleoprotein. Thus it is possible to predict the membrane topology of the TGEV integral membrane protein from the DNA sequence data when compared to those of other coronaviruses.

The coronavirus integral membrane protein is glycosylated and has been reported to be of the *N*-linked type for TGEV (Garwes *et al.*, 1984; Jacobs *et al.*, 1986) and IBV (Stern and Sefton, 1982a), whereas MHV and BCV are of the *O*-linked type (Niemann *et al.*, 1984; Lapps *et al.*, 1987). There are three potential *N*-glycosylation sites within the integral membrane protein: at asparagine positions 32, 55 and 251, as deduced from the DNA sequence for TGEV. The first two are at the amino terminal of the polypeptide (Fig. 3) although the second site (Asn-55) falls within the first predicted membrane-spanning region (Figs 3 and 6), as determined from alignment with MHV, and so it may not be glycosylated. The third potential glycosylation site is at the carboxy-terminal end of the polypeptide, predicted to be found inside the virion, and so it may not be glycosylated. The prediction that only the first *N*-glycosylation site is functional would give rise to a polypeptide of M_r 29000–30000, a size close to the major species identified by SDS polyacrylamide gel electrophoresis of

```
FS772     MKLLLILACVIACACGERYCAMKDDTGLSCRNS.TASACESCFN..GGDLIWHLANWNFS
PURDUE    MKILLILACVIACACGERYCAMKSDTDLSCRNS.TASDCESCFN..GGDLIWHLANWNFS
MHV       ..............................MSS.TTQAPEPVYQWTADEAVQFLKEWNFS
BCV       ..............................MSSVTTPAP..VYTWTADEAIKFLKEWNFS
IBV       ...................................MP.NETNCTLDFEQSVQLFKEYN..

Consensus ------------------------------S-T--A-E-------------LKEWNFS

FS772     WSIILIIFITV........LQYGRPQFSWFVYGIKMLIMWLLWPIVLALTIFNAYSEYQV
PURDUE    WSIILIVFITV........LQYGRPQFSWFAYGIKMLIMWLLWPVVLALTIFNAYSEYQV
MHV       LGIILL.FIT.......IILQFGYTSRSMFIYVVKMIILWLMWPLTIVLCIFNCV..YAL
BCV       lGIILL.FIT.......IILQFGYTSRSMFVYVIKMIILWLMWPLTIILTIFNCV..YAL
IBV       ....L..FITAFLLFLTIILQYGYATRSKVIYTLKMKILWCFWPLNIAVGVISCTYPPNT

Consensus --IIL--FIT.......IILQ-GY--RS-F-Y--KM-ILWL-WPL-I-L-IFNC---Y--

FS772     SRYVMFGVSIAGAIVTFVL.WIMYFVRSIQLYRRTKSWWSFNPEINAILCV.SALGRSYV
PURDUE    SRYVMFGFSIAGAIVTFVL.WIMYFVRSIQLYRRTKSWWSFNPETKAILCV.SALGRSYV
MHV       NN.VYLGFSIVFTIVSIVI.WIMYFVNSIRLFIRTGSWWSFNPETNNLMCI.DMKGTVYV
BCV       NN.VYLGFSIVFTIVAIIM.WIVYFVNSIRLFIRTGSWWSFNPETNNLMCI.DMKGRMYV
IBV       GGLVAAIILTVFACLSFVGYWI....QSIRLFKRCRSWWSFNPESNAVGSILLTNGQQCN

Consensus ---V--GFSIVF-IV--V--WI-YFV-SIRLF-RT-SWWSFNPETN---CI----G--YV

FS772     LP.LEGVPTGVTLTLLS.GNLYAEGFKIAG.GMNI.DNLPKYVMVALPSRTIVYTLVGKK
PURDUE    LP.LEGVPTGVTLTLLS.GNLYAEGFKIAG.GMNI.DNLPKYVMVALPSRTIVYTLVGKK
MHV       RPIIEDYHT.LTATIIR.GHLYMQGVKL.GTGFSLSD.LPAYVTVAKVSHLCTYKRAFLD
BCV       RPIIEDYHT.LTVTIIR.GHLYMQGIKL.GTGYSLSD.LPAYVTVAKVSHLLTYKRGFLD
IBV       FA.IESVPMVLS.PIIKNGVLYCEGQWLA.KC.EP.DHLPKDIFVCTPDRRNIYRMVQKY

Consensus -P-IE---T-LT-TII--G-LY--G-KL-G-G----D-LP-YV-VA--S----Y------

FS772     LKASSAT..GWAY..YVK.SKAGDYSTDARTDNLSEQEK.LLHMV.
PURDUE    LKASSAT..GWAY..YVK.SKAGDYSTEARTDNLSEQEK.LLHMV.
MHV       .KVDGVS..GFA..VYVK.SKVGNYRLPS.NKP.SGADTALLR..I
BCV       .KIGDTS..GFA..VYVK.SKVGNYRLPSTQKG.SGMDTALLRNNI
IBV       TGDQSGNKKRFATFVYAKQS.VDTGELESVATGGSSLYT.......

Consensus -K-------GFA--VYVK-SKVG-Y-L-S-----S---T-LL----
```

Fig. 6. Alignment of the integral membrane protein amino acid sequences from TGEV strains FS772/70 (virulent), Purdue (avirulent), MHV (A59), BCV (Mebus) and IBV (Beaudette) using the GAP program of Devereux *et al.* (1984). The dots within the sequences are 'padding characters' inserted to achieve optimal alignment. The consensus sequence is derived if all four types of virus contain the same residue at the aligned position. The thick black line shows the position of the membrane–spanning regions in MHV, as deduced by Rottier *et al.* (1986).

either TGEV-infected cells or purified virions, since the presence of an *N*-linked glycan may add up to 2000 to the molecular weight of a polypeptide (Klenk and Rott, 1980). Other species of the integral membrane protein observed on SDS gels from both TGEV-infected cells and virions may arise from heterogeneity in the oligosaccharide side chains and from incomplete glycosylation.

The homology between the two strains of TGEV in the amino acid and nucleotide sequences is 98%. There are 11 amino acid substitutions between strain FS772/70 and the avirulent Purdue strain, as determined by Kapke *et al.* (1987), and 12 substitutions from the Purdue strain, as determined by Laude *et al.* (1987); see Figs 3 and 6. A comparison of the sequences of avirulent viruses is important in identifying regions that may be involved in pathogenicity. Any conformational change in the amino acid backbone may affect receptor binding affinities and give rise to different epitope sites. The data on the virulent strain studied here provide a framework for the identification of the amino acids which play a key role in these processes.

Experimental procedures

Preparation of viral RNA

TGEV mRNA was prepared from LLC-PK1 cells infected with TGEV strain FS772-70 and purified from other RNA species on poly(U) Sepharose, as described previously (Britton *et al.*, 1987; 1988).

Digestion and analysis of plasmid DNA

Standard recombinant DNA methods were used (Maniatis *et al.*, 1982) with enzymes purchased from New England Biolabs (CP Laboratories, Bishop's Stortford, UK) unless otherwise stated in the text. DNA fragments were isolated from agarose gels by freeze-phenol elution (Silhavy *et al.*, 1984). Ligation reactions were carried out as described by Britton *et al.* (1984). *E. coli* cells were routinely transformed using the RbCl method developed by V. Simanis (Hanahan, 1985). *E. coli* strain DH1 was used for routine plasmid construction and JM101 for expression of the chimaeric protein. *E. coli* transformants were selected on LB plates containing ampicillin (100 μg ml^{-1}). Vector DNA was routinely treated with alkaline phosphatase prior to ligation.

cDNA synthesis

cDNA synthesis was carried out as described by Britton *et al.* (1988) except that a synthetic oligonucleotide, 5′-GCCATTTA-GAAGTTTAGT-3′, was used to prime first-strand synthesis and 1200U of Moloney murine leukaemia virus (M-MLV) reverse transcriptase (BRL) was used for the production of cDNA. The primer corresponded to a complementary sequence 13bp upstream from the 5′ end of the nucleoprotein gene, derived from the sequence data reported by Britton *et al.* (1988), and was synthesized by the phosphoramidite method using an Applied Biosystems model 381A DNA synthesizer. Second-strand synthesis was carried out as described by Britton *et al.* (1988). Following second-strand synthesis, oligo-dC tails were added to the cDNA molecules using 25U terminal transferase at 15°C for 2 min. The dC-tailed cDNA molecules were annealed to oligo(dG)-tailed pUC9 (Pharmacia) at 65°C for 10 min. After 4h at 58°C, they were transformed into competent DH1 *E. coli* cells. Transformants containing TGEV cDNA were identified by colony hybridization, as described by Britton *et al.* (1988) using a [^{32}P]-labelled TGEV cDNA fragment that hydridized to all of the TGEV mRNA species larger than the 2.7 kb mRNA species.

Subcloning for M13 sequencing

Various restriction endonuclease fragments from plasmids pF4F-36 and pBIM3 deduced from restriction maps were cloned into M13mp vectors. DNA sequencing and sequence analysis were carried out as described by Britton *et al.* (1988).

Production of a β-galactosidase-TGEV integral membrane protein chimaera

A TGEV cDNA fragment containing 95% of the integral membrane protein gene, was ligated into the expression vector pUR290

(Ruther and Muller-Hill, 1983) and the resulting plasmid, pURIM-2, was transformed into JM101. Cells containing plasmid pURIM-2 were induced by isopropyl-β-D-thiogalactopyranoside (IPTG) to express the chimaeric protein identified by SDS gel electrophoresis as described by Britton *et al.* (1987).

Purification of chimaeric protein

Cultures of JM101(pURIM-2) were grown at 37°C in Luria broth to an absorbance of 0.5 at 680 nm, the chimaeric protein was induced by the addition of IPTG (1 mM), and the cells were grown for a further 3 h. The cells were collected by centrifugation and lysed by the addition of Laemmli sample buffer and incubation at 100°C. The cell lysate was electrophoresed into 6% preparative polyacrylamide gels. The protein bands were visualised by soaking the gels in 4 M sodium acetate, and the band corresponding to the chimaeric protein was excised from the gel, crushed and electroeluted from the acrylamide as described by Britton (1981). The purified chimaeric protein was used to immunize Balb/C mice, as described by Garwes *et al.* (1987).

Immunological analysis of the chimaeric protein

Immunoblot analysis was carried out as described by Britton *et al.* (1988) except that monoclonal antibody, 3BB3 (Jimenez *et al.*, 1986), raised against TGEV integral membrane protein was used as a probe. Immune precipitation was carried out as described by Britton *et al.* (1987) except that the formalin-fixed *Staphylococcus aureus* cells used were obtained from BRL (Immunoprecipitin, 560–9321SA).

Acknowledgements

This research was supported by the Biomolecular Engineering Programme of the Commission of the European Communities (Contract No. GB1-2-089-UK). R. S. Cármenes (from the Departamento de Bioquimica, Universidad de Oviedo, Spain) was supported by a Fleming Award from the Spanish Ministerio de Educacion y Ciencia and the British Council. We would like to thank Dr J. McCauley from the Pirbright Laboratory of this Institute for synthesizing the oligonucleotide primer, and Dr L. Enjuanes from the Centro de Biologia Molecular, Madrid for providing the monoclonal antibody (3BB3) against the integral membrane protein of the Purdue strain of TGEV.

References

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* **308:** 751–752.

Boursnell, 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.

Britton, P. (1981) Ph.D. Thesis, University of Edinburgh.

Britton, P., Murfitt, D., Parra, F., Jones-Mortimer, M.C., and Kornberg, H.L. (1982) Phosphotransferase-mediated regulation of carbohydrate utilisation in *Escherichia coli* K12: identification of the products of genes on the specialised transducing phages λ*iex(crr)* and λ*gsr(tgs)*. *EMBO J* **1:** 907–911.

Britton, P., Lee, L.G., Murfitt, D., Boronat, A., Jones-Mortimer, M.C., and Kornberg, H.L. (1984) Location and direction of transcription of the *ptsH* and *ptsI* genes on the *Escherichia coli* K12 genome. *J Gen Microbiol* **130:** 861–868.

Britton, P., Garwes, D.J., Millson, G.C., Page, K., Bountiff, L., Stewart, F., and Walmsley, J. (1986) Towards a genetically-engineered vaccine against porcine transmissible gastroenteritis virus. In *Biomolecular Engineering in the European Community. Final Report*. Magnien, E. (ed.). The Netherlands: Martinus Nijhoff, pp. 301–313.

Britton, P., Garwes, D.J., Page, K., and Walmsley, J. (1987) Expression of porcine transmissible gastroenteritis virus genes in *E. coli* as β-galactosidase chimaeric proteins. In *Coronaviruses*. Lai, M.M.C., and Stohlman, S.A. (eds). New York: Plenum Press. *Adv Exp Med Biol* **218:** 55–64.

Britton, P., Cármenes, R.S., Page, K.W., Garwes, D.J., and Parra, F. (1988) Sequence of the nucleoprotein from a virulent British field isolate of transmissible gastroenteritis virus and its expression in *Saccharomyces cerevisiae*. *Mol Microbiol* **2:** 89–99.

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

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.

Dennis, D.E., and Brian, D.A. (1981) Coronavirus cell-associated RNA-dependent RNA polymerase. In *Biochemistry and Biology of Coronaviruses*. ter Meulen, V., Siddell, S., and Wege, H. (eds). New York: Plenum Press. *Adv Exp Med Biol* **142:** 155–170.

Devereux, J., Haeberli, P., and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl Acids Res* **12:** 387–395.

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

Garwes, D.J., Pocock, D.H., and Wijaszka, T.M. (1975) Identification of heat-dissociable RNA complexes in two porcine coronaviruses. *Nature* **257:** 508–510.

Garwes, D.J., Bountiff, L., Millson, G.C., and Elleman, C.J. (1984). Defective replication of porcine transmissible gastroenteritis virus in a continuous cell line. In *Molecular Biology and Pathogenesis of Coronaviruses*. Rottier, P.J.M., van der Zeijst, B.A.M., Spaan, W.J.M., and Horzinek, M.C. (eds). New York: Plenum Press. *Adv Exp Med Biol* **173:** 79–93.

Garwes, D.J., Stewart, F., and Elleman, C.J. (1987) Identification of epitopes of immunological importance on the peplomer of porcine transmissible gastroenteritis virus. In *Coronaviruses*. Lai, M.M.C., and Stohlman, S.A. (eds). New York: Plenum Press. *Adv Exp Med Biol* **218:** 509–515.

Hanahan, D. (1985) Techniques for transformation of *Escherichia coli*. In *DNA Cloning: A Practical Approach*. Vol. II. Glover, D.M. (ed.). Oxford: IRL Press, 109–135.

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.

Jimenez, G., Correa, I., Melgosa, M.P., Bullido, M.J., and Enjuanes, L. (1986) Critical epitopes in transmissible gastroenteritis virus neutralization. *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.

Kapke, P.A., Tung, F.Y.C., Brian, D.A., Woods, R.D., and Wesley, R. (1987) Nucleotide sequence of the porcine transmissible

gastroenteritis coronavirus matrix protein. In *Coronaviruses.* Lai, M.M.C., and Stohlman, S.A. (eds). New York: Plenum Press. *Adv Exp Med Biol* **218:** 117–122.

Klenk, H.D., and Rott, R. (1980) Cotranslational and posttranslational processing of viral glycoproteins. *Curr Top Microbiol Immunol* **90:** 19–48.

Kozak, M. (1983) Comparison of initiation of protein synthesis in prokaryotes, eukaryotes and organelles. *Microbiol Rev* **47:** 1–45.

Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44:** 283–292.

Kyte, J., and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157:** 105–132.

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

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.

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.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

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

Niemann, H., and Klenk, H.D. (1981) Coronavirus glycoprotein E1, a new type of viral glycoprotein. *J Mol Biol* **153:** 993–1010.

Niemann, H., Geyer, R., Klenk, H.D., Linder, D., Stirm, S., and Wirth, M. (1984) The carbohydrates of mouse hepatitis virus (MHV) A59: structures of the *O*-glycosidically linked oligosaccharides of glycoprotein E1. *EMBO J* **3:** 665–670.

Pfleiderer, M., Skinner, M.A., and Siddell, S.G. (1986) Coronavirus MHV-JHM: nucleotide sequence of the mRNA that encodes the membrane protein. *Nucl Acids Res* **14:** 6338.

Rasschaert, D., and Laude, H. (1987) The predicted structure of the peplomer protein E2 of the porcine coronavirus transmissible gastroenteritis virus. *J Gen Virol* **68:** 1883–1890.

Rasschaert, D., Delmas, B., Charley, B., Grossclaude, J., Gelfi, J., and Laude, H. (1987a) Surface glycoproteins of transmissible gastroenteritis virus: functions and gene sequence. In *Coronaviruses.* Lai, M.M.C., and Stohlman, S.A. (eds). New York: Plenum Press. *Adv Exp Med Biol* **218:** 109–116.

Rasschaert, D., Gelfi, J., and Laude, H. (1987b) Enteric coronavirus TGEV: partial sequence of the genomic RNA, its organisation and expression. *Biochimie* **69:** 591–600.

Rottier, P., Welling, G.W., Welling-Wester, S., Niesters, H.G., Lenstra, J.A., and van der Zeijst, B.A.M. (1986) Predicted membrane topology of the coronavirus protein E1. *Biochemistry* **25:** 1335–1339.

Ruther, U., and Muller-Hill, B. (1983) Easy identification of cDNA clones. *EMBO J* **2:** 1791–1794.

Shieh, C.-K., Soe, L.H., Makino, S., Chang, M-F., Stohlman, S.A., and Lai, M.M.C. (1987) The 5′-end sequence of the murine coronavirus genome: implications for multiple fusion sites in leader-primed transcription. *Virology* **156:** 321–330.

Silhavy, T.J., Berman, M.L., and Enquist, L.W. (1984) *Experiments With Gene Fusions.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

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:** 1839–1844.

Staden, R. (1982) An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucl Acids Res* **10:** 2951–2961.

Stern, D.F., and Sefton, B.M. (1982a) Coronavirus proteins: structure and function of the oligosaccharides of the avian infectious bronchitis virus glycoproteins. *J Virol* **44:** 804–812.

Stern, D.F., and Sefton, B.M. (1982b) Coronavirus proteins: biogenesis of avian infectious bronchitis virus virion proteins. *J Virol* **44:** 794–803.

Von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites. *Nucl Acids Res* **14:** 4683–4690.

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.