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

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

Virus Research, 18 (1990) 71–80
Elsevier

VIRUS 00620

Sequence of the S gene from a virulent British field isolate of transmissible gastroenteritis virus

Paul Britton and Kevin W. Page *

A.F.R.C., Institute for Animal Health, Compton Laboratory, Compton, Newbury, Berkshire, RG16 0NN, U.K.

(Accepted 5 September 1990)

Summary

Subgenomic mRNA from a virulent field isolate of porcine transmissible gastroenteritis virus (TGEV), strain FS772/70, was used to produce cDNA. The cDNA from three overlapping clones was sequenced by the chain termination method and two open reading frames (ORFs) were identified. The largest ORF, 4350bp, encoded a polypeptide of 1449 amino acids of relative molecular mass (M_r) 159811, which contained 33 potential N-linked glycosylation sites, a cysteine-rich region, and a potential transmembrane region. The C-terminal half of this ORF showed homology to the S proteins of four other coronaviruses. The other ORF consisted of the 3′-end of a gene with homology to the carboxyl terminus of the F2 subunit of infectious bronchitis virus (IBV) RNA polymerase.

Transmissible gastroenteritis virus; Coronavirus; Spike; Porcine; cDNA

TGEV causes gastroenteritis in pigs resulting in a high mortality and morbidity in neonates and belongs to the family Coronaviridae, a group of enveloped viruses with a positive-stranded RNA genome. Coronavirus virions contain two membrane glycoproteins spike (S), membrane (M) and a basic phosphorylated nucleoprotein (N). The TGEV S protein has a monomeric M_r of 200,000 and has been shown to

Correspondence to: P. Britton, A.F.R.C., Institute for Animal Health, Compton Laboratory, Compton, Newbury, Berkshire, RG16 0NN, U.K.

* *Present address:* MRC, Institute of Virology, University of Glasgow, Church Street, Glasgow, G11 5JR, U.K.

0168-1702/90/$03.50 © 1990 Elsevier Science Publishers B.V. (Biomedical Division)

elicit the production of neutralising antibodies (Jimenez et al., 1986; Laude et al., 1986; Garwes et al., 1987) which are able to confer some protection on suckling piglets (Garwes et al., 1978/79). Coronavirus S proteins either exist as a single polypeptide, TGEV, feline infectious peritonitis virus (FIPV), canine coronavirus (CCV) and porcine respiratory coronavirus (PRCV) or as two subunits, IBV, mouse hepatitis virus (MHV), bovine coronavirus (BCV) and haemagglutinating encephalomyelitis virus (HEV). Cavanagh (1983) showed that IBV S2 anchors the protein to the viral envelope via a short carboxyl-terminal hydrophobic domain and that S1 is non-covalently attached to S2. Removal of the S1 subunit abolished infectivity but not attachment to cells (Cavanagh and Davis, 1986). In this paper, the nucleotide sequence of cloned cDNA corresponding to the S protein gene, from a virulent British field isolate of TGEV (strain FS772/70), was determined and the deduced amino acid sequence of the S protein was compared to the avirulent Purdue strain of TGEV.

TGEV poly(A)$^+$ mRNA species were isolated from virus infected LLC-PK1 cells (Britton et al., 1987) and used for the synthesis of cDNA (Britton et al., 1988b). Transformants containing TGEV cDNA were identified by colony hybridisation, using two ^{32}P-labelled TGEV cDNA fragments H11 (derived from the S gene) and H26 (derived from the ORF-3b gene, formerly ORF 2) as described by Britton et al. (1989). Plasmid pTG30 contained cDNA (3.4 kbp) which hybridised to both cDNA fragments and plasmid pTG25 contained cDNA (3.2 kbp) which hybridised to H11. Plasmid pTG30 overlapped pTG11 (Britton et al., 1989) and pTG25 in turn overlapped pTG30. A third plasmid, pTG47, contained cDNA (2.5 kbp) that overlapped pTG25 by 250 bp, but did not hybridise to H11 or H26. The cDNA from plasmids pTG30, pTG25 and pTG47 extended 10.5 kb into the viral genome, from the 3′-end, indicating that the S gene should be contained within the three plasmids.

The nucleotide sequence of the cDNA in plasmids pTG30, pTG25 and pTG47 was determined in both directions by the di-deoxy method using [α-^{35}S]dATP and Sequenase ™ (United States Biochemical Corporation). Analysis of sequence data corresponding to 6120 bp (Fig. 1) revealed one complete ORF of 4350 bp, nucleotides 1600–5949, and an incomplete ORF of 1602 bp, nucleotides 2–1603, which terminated at a TGA stop codon and overlapped the 4350 bp ORF by four nucleotides. The 1602 bp incomplete ORF encoded a contiguous polypeptide of 533 amino acids potentially forming the 3′-end of a larger gene. Comparison of the 533 amino acid sequence, using a dot-matrix analysis program (DIAGON; Staden, 1982), with the last 752 amino acids at the carboxyl terminus of the F2 subunit of the IBV polymerase (Boursnell et al., 1987) revealed very good homology (Fig. 2), suggesting that this sequence was the 3′-end of the TGEV polymerase gene.

The 4350 bp ORF (Fig. 1), initiating from the ATG at position 1600, encoded a polypeptide of 1449 amino acids with a M_r 159,811 and was preceded, 26 bp 5′ to the ATG, by the potential TGEV RNA polymerase-leader complex binding site ACTAAAC. The 5′-end of this ORF mapped 8180 nucleotides from the start of the poly (A) tail indicating that it corresponded to the S gene present on the 5′ coding region of mRNA 2. Comparison of the S genes from FS772/70 and two indepen-

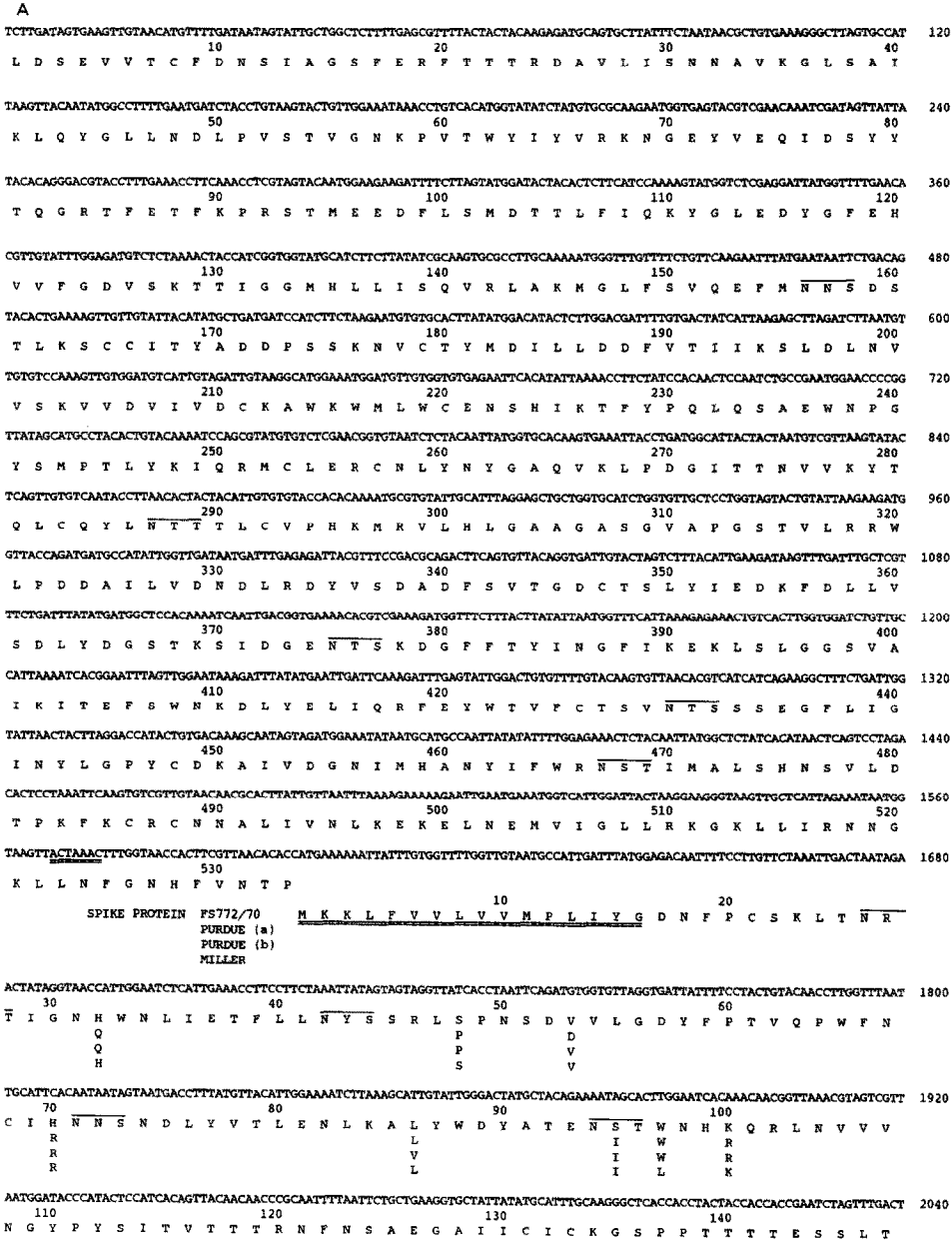

Fig. 1. The nucleotide and deduced amino acid sequences of the carboxyl-terminus of the polymerase, S and ORF-3a genes from TGEV FS772/70. The ACTAAAC sequences upstream of TGEV genes are double underlined. Amino acids below the FS772/70 S sequence are substitutions found in the avirulent Purdue-115 (a) Rasschaert and Laude, (1987), (b) Jacobs et al., (1987) and the Miller (Wesley, 1990) strains. The double underlined sequence at the beginning of the gene is the predicted N-terminal signal sequence of the S protein. Potential N-glycosylation sites (NXT or NXS) are identified by a line above the sequence. The lines under the main amino acid sequence refer to regions of homology between the TGEV, FIPV, MHV, BCV and IBV S proteins. The double underlined sequence at the carboxyl-terminus of the S protein had the highest $\langle H \rangle$ and $\langle \mu_H \rangle$ values (Eisenberg et al., 1984) indicating it is the potential transmembrane region. These sequence data will appear in the EMBL/GenBank/DDBJ nucleotide Database under the accession number X53128.

dent sequences of the avirulent Purdue strains of TGEV (Jacobs et al., 1987; Rasschaert and Laude, 1987) revealed 97% homology in both the nucleic and amino acid sequences. However, a six base pair insert within the FS772/70 sequence at nucleotide position 2722 (Fig. 1) resulted in the insertion of the amino acids

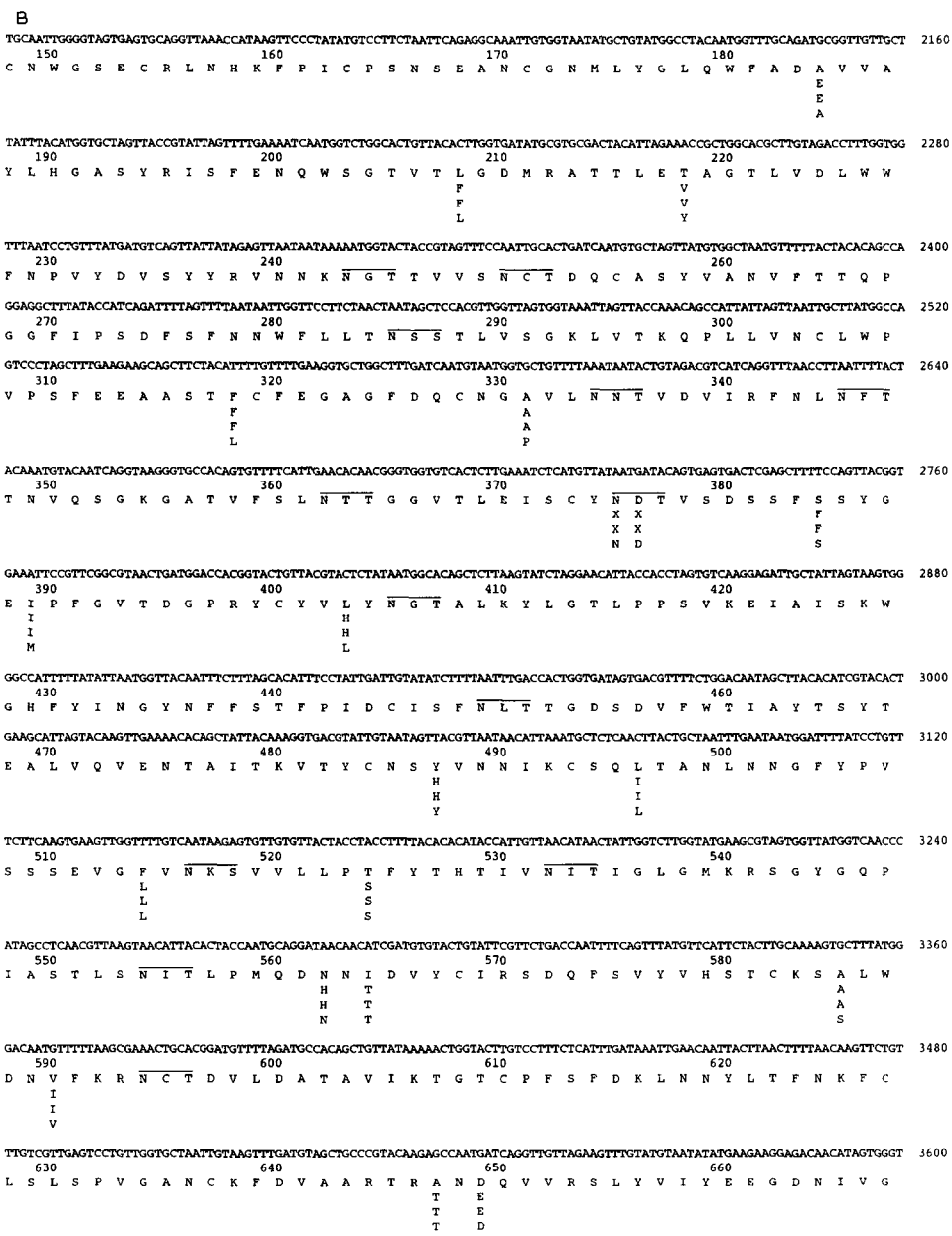

Fig. 1 (continued).

asparagine and aspartic acid giving an extra potential N-linked glycosylation site. This insertion has also been identified, at the same position, within the S gene sequence of the virulent Miller strain of TGEV (Wesley, 1990). Comparison of the deduced amino acid sequences from the avirulent Purdue strain and the virulent FS772/70 strain revealed 35 amino acid substitutions, though not all were the same for both Purdue sequences (Fig. 1). Conservative amino acid substitutions accounted for 23 of the 35 substitutions. It is interesting to note that out of the 44 amino acid substitutions between all four TGEV sequences 19 involved amino acids that were

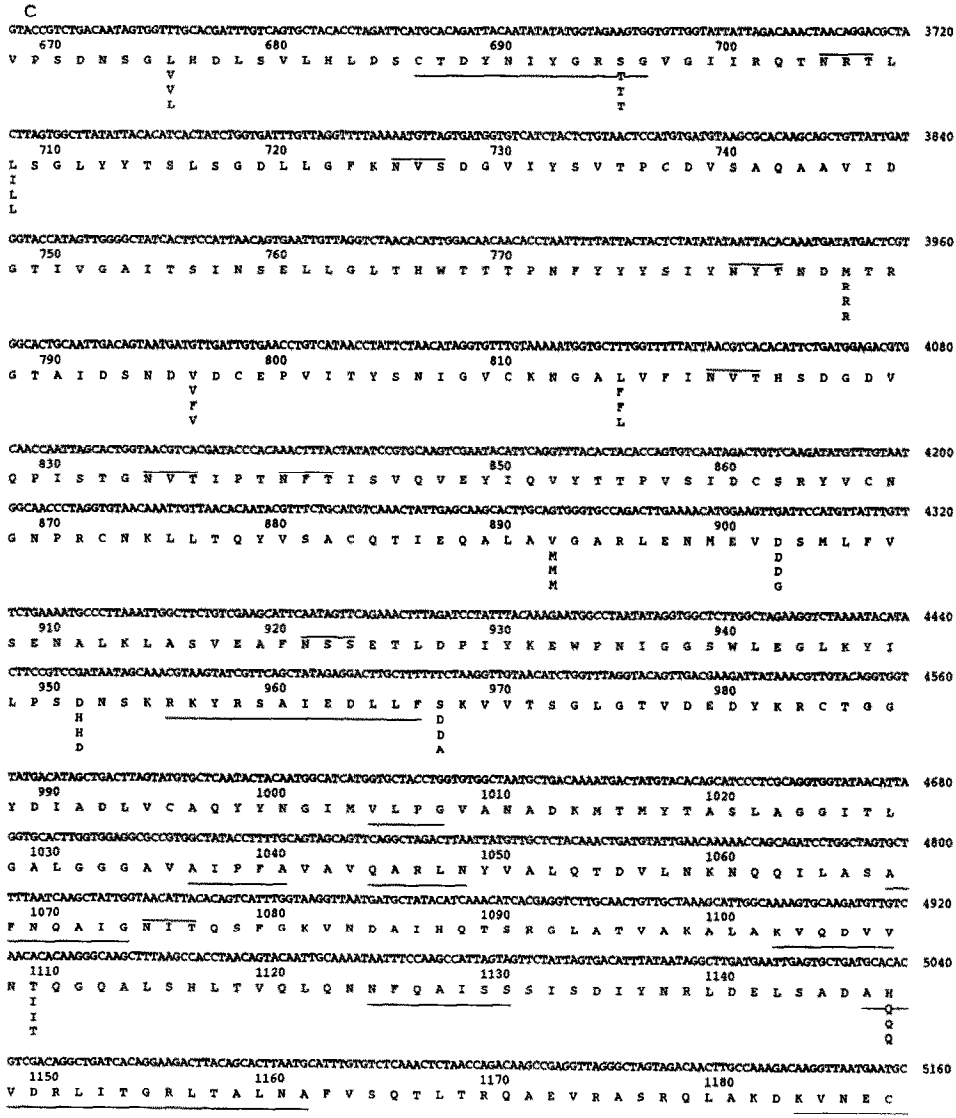

Fig. 1 (continued).

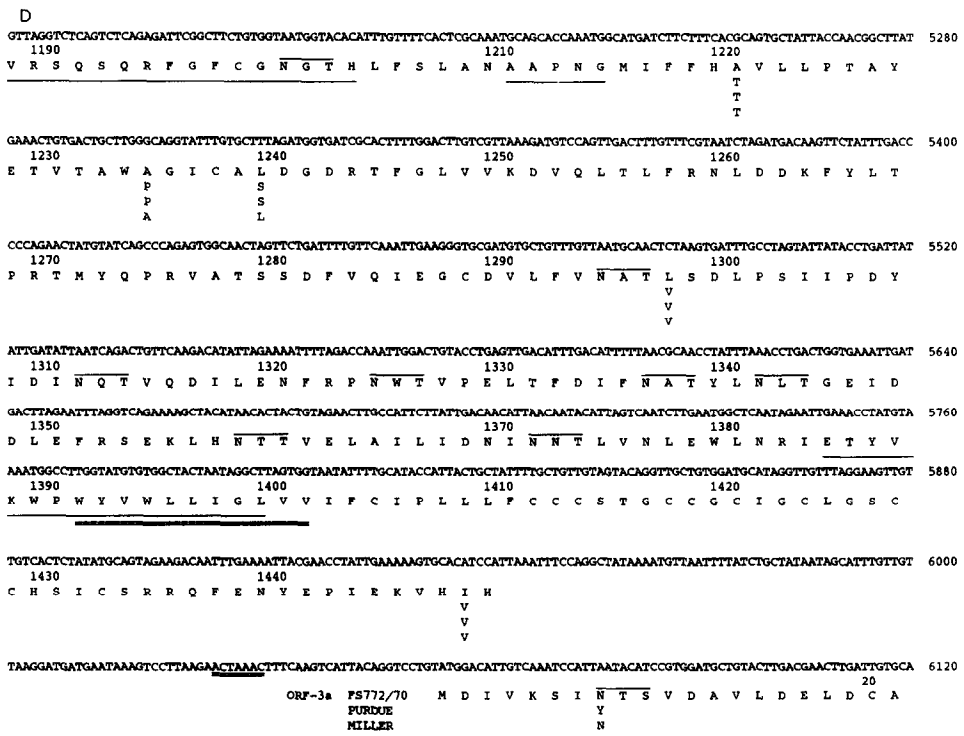

Fig. 1 (continued).

identical for the two virulent strains of TGEV but different from those found in the two independent Purdue sequences.

Analysis of the FS772/70 S protein amino acid sequence identified a 16 amino acid N-terminal membrane signal sequence with the cleavage site between the glycine (16) and aspartic acid (17) residues (score = 7.98; von Heijne, 1986) and a C-terminal hydrophobic region corresponding to the transmembrane anchor region. There are 33 potential N-linked glycosylation sites (Asn-X-Thr or Asn-X-Ser), of which 27 (82%) are Asn-X-Thr, mainly distributed over the N-terminal half of the molecule but with a group of 7 potential sites between amino acid residues 1293-1374 (Fig. 1). Dotmatrix analysis showed that the polypeptide had homology, at the carboxyl half of the molecule, to the S proteins from the coronaviruses IBV Beaudette (Binns et al., 1985) and MHV (Schmidt et al., 1987) BCV (Parker et al., 1990) and HCV (Raabe et al., 1990) S proteins and is almost identical throughout its length, except for the first 260 amino acids, with FIPV 79-1149 (De Groot et al., 1987) S gene product. Analysis of the FS772/70 S amino acid sequence with the algorithm of Eisenberg et al. (1984), for membrane and surface protein sequences, revealed an 11 amino acid sequence WYVWLLIGLVV (Fig. 1) which had the highest mean hydrophobicity ($\langle H \rangle$), 0.92, and mean hydrophobic moment ($\langle \mu_H \rangle$), 0.106, suggesting its role as a potential transmembrane α-helix responsible for

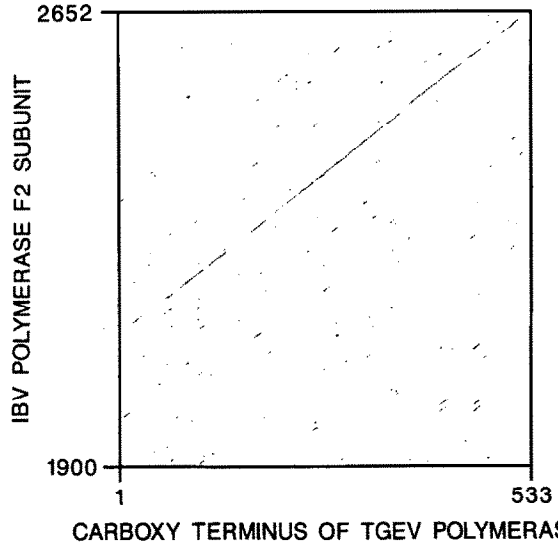

Fig. 2. Comparison of the deduced amino acid sequence derived from the 1602 bp incomplete ORF, nucleotides 2-1603, with the last 752 amino acids at the carboxyl-terminus of the F2 subunit of the IBV polymerase (Boursnell et al., 1987), by dot-matrix analysis [DIAGON; (Staden, 1982)]. The comparison used a span of 11 residues and a score of 136.

anchoring the TGEV S protein. The sequence, KWPWYVWL (residues 1388–1395; Fig. 1), flanking the membrane anchor region was conserved in TGEV, IBV, MHV, BCV, FIPV and HCV (except for one amino acid) indicating that it may be the site of entry into the virion membrane. One of the S protein subunits from both MHV and BCV has been shown to be acylated with palmitic acid (Schmidt, 1982). A cysteine rich region, residues 1404–1449 (11 cysteine residues out of the 46 amino acids), a distinctive feature of all coronavirus S proteins sequenced to date, follows the predicted transmembrane α-helix and may serve as a site for acylation with palmitic acid.

The amino acid substitutions identified between virulent and avirulent sequences may lead to the identification of differences between antigenic determinants. A region on the avirulent Purdue-115 S sequence, SSFFSYGEI (site C), has been identified as a linear monoclonal antibody (mAb) binding site (Delmas et al., 1990), with residues FSY critical for the binding of two neutralising mAbs. Posthumus et al. (1990) showed that amino acid substitutions in the sequence, FFSYGEI, confirmed this as the primary determinant for neutralising mAbs and showed that the two Phe and the Ser, Gly and Glu residues were critical for the binding of both neutralising and non-neutralising mAbs. Delmas et al. (1990) have shown that mAbs which react with the Purdue site C do not react with the S protein from FS772/70. Comparison of this epitope on the FS772/70 sequence identified that the second Phe residue on the Purdue sequence was replaced by a Ser residue on the FS772/70 sequence, according to the data of Posthumus et al. (1990) this would prevent the binding of the Purdue derived mAbs to this epitope in TGEV FS772/70. The

virulent Miller strain of TGEV also has the Phe → Ser substitution preventing Purdue-derived mAbs to this epitope from binding. The two amino acid insertion found on FS772/70 and Miller sequences was only four amino acids upstream from this epitope and may either induce a conformational change in this region of the S protein or present an extra potential N-linked glycan, both of which could also interfere with antibody binding. If the Phe → Ser substitution and the two amino acid insertion are features of all virulent TGEV strains it is highly unlikely that this linear epitope will be of any value in vaccine regimes, incorporating avirulent strains, for protection against virulent TGEV infections. It is also significant that polyclonal sera from pigs inoculated with the Purdue strain of TGEV did not recognise this epitope (Posthumus et al., 1990).

S proteins from coronaviruses antigenically distinct from TGEV such as MHV, IBV, BCV and HEV are post-translationally cleaved. The S protein from TGEV and from coronaviruses antigenically related such as FIPV, CCV and the recently isolated PRCV are not cleaved. The sequence DMTRG is found at residue 784 on the FS772/70 sequence and is similar to the sequences DRTRG and ERTRG, found on the Purdue (Rasschaert and Laude, 1987) and FIPV 79–1146 (De Groot et al., 1987) S sequences and were postulated to be reminiscent of the S protein cleavage sites. Division of the TGEV S protein, at the DMTRG site, showed there were 20 potential N-glycosylation sites and 23 amino acid substitutions, of which 16 are conservative, within the amino terminal half when comparing the virulent and avirulent strains of TGEV. Comparison of the S proteins from TGEV, FIPV, MHV, BCV and IBV revealed sequences with high homology (Fig. 1) mainly on the carboxyl half of the S protein which may reflect constraints in forming the stalk structure and transmembrane anchor domain of the protein. The heptad repeat, identified on the S protein of the Purdue strain (Rasschaert and Laude, 1987), postulated to form an α-helical coiled-coil structure resulting in the elongated shape of the S protein was conserved on the Miller (Wesley, 1990) and FS772/70 sequences except for one amino acid substitution, threonine for isoleucine at position 1107 of the Purdue sequence. The low homologies observed at the amino terminal half of the coronavirus S proteins probably result from the different tropisms of the various coronaviruses and give rise to the different antigenic subgroups.

This paper, together with three other papers reporting the sequences of the N (Britton et al., 1988a), M (Britton et al., 1988b) and the genome between the S and M genes (Britton et al., 1989), provides a complete set of sequence data, from the poly (A)-tail to the 3′-end of the polymerase gene, for a virulent field isolate of TGEV. The order of the viral genes from the FS772/70 strain of TGEV is 5′-[Pol-S-3a-3b-4-M-N-7]-3′, where ORFs 3a, 3b and 4 (formerly ORFs 1, 2 and 3; Britton et al., 1989) are potential genes whose products have yet to be confirmed. ORF-7 (formerly ORF 4; Britton et al., 1988a), whose product has been detected in virus infected cells (Garwes et al., 1989) has high homology to a penultimate ORF in FIPV (De Groot et al., 1988). The identification of an incomplete ORF, 5′ to the TGEV S gene, showing homology to the F2 subunit of the IBV polymerase, indicated that there are no extra gene sequences in this region of the TGEV genome,

in contrast to other mammalian coronaviruses BCV and MHV where other genes or pseudogenes are present, possibly resulting from recombination events.

Acknowledgements

We would like to thank Miss K. Mawditt for synthesising the oligonucleotides used in sequencing the TGEV cDNA, Dr R. Wesley for providing the sequence of the S gene from the Miller strain of TGEV prior to publication and Dr D. Pocock for helpful comments on the manuscript. This research was supported by the Biomolecular Engineering Programme of the Commission of the European Communities Contract No. [BAP-0235-UK(HI)].

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.

Boursnell, M.E.G., Brown, T.D.K., Foulds, I.J., Green, P.F., Tomley, F.M. and Binns, M.M. (1987) Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. J. Gen. Virol. 68, 57–77.

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: M.M.C. Lai and S.A. Stohlman (Eds), Coronaviruses. Adv. Exp. Med. Biol., Vol. 218, pp. 55–64. Plenum Press, New York, London.

Britton, P., Carmenes, R.S., Page, K.W., Garwes, D.J. and Parra, F. (1988a) 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.

Britton, P., Carmenes, R.S., Page, K.W. and Garwes, D.J. (1988b) The integral membrane protein from a virulent isolate of transmissible gastroenteritis virus: molecular characterization, sequence and expression in Escherichia coli. Mol. Microbiol. 2, 497–505.

Britton, P., Lopez Otin, C., Martin Alonso, J.M. and Parra, F. (1989) Sequence of the coding regions from the 3.0 kb and 3.9 kb mRNA subgenomic species from a virulent isolate of transmissible gastroenteritis virus. Arch. Virol. 105, 165–178.

Cavanagh, D. (1983) Coronavirus IBV: structural characterization of the spike protein. J. Gen. Virol. 64, 2577–2583.

Cavanagh, D. and Davis, P.J. (1986) Coronavirus IBV: removal of spike glycopeptide S1 by urea abolishes infectivity and haemagglutination but not attachment to cells. J. Gen. Virol. 67, 1443–1448.

De Groot, R.J., Maduro, J., Lenstra, J.A., Horzinek, M.C., Van Der Zeijst, B.A.M. and Spaan, W.J.M. (1987) cDNA cloning and sequence analysis of the gene encoding the peplomer protein of feline infectious peritonitis virus. J. Gen. Virol. 68, 2639–2646.

De Groot, R.J., Andeweg, A.C., Horzinek, M.C. and Spaan, W.J.M. (1988) Sequence analysis of the 3′ end of the feline coronavirus FIPV79-1146 genome: comparison with the genome of porcine coronavirus TGEV reveals large insertions. Virology 167, 370–376.

Delmas, B., Rasschaert, D., Godet, M., Gelfi, J. and Laude, H. (1990) Four major antigenic sites of the coronavirus transmissible gastroenteritis virus are located on the amino-terminal half of spike glycoprotein S. J. Gen. Virol. 71, 1313–1323.

Eisenberg, D., Schwarz, E., Komaromy, 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.F. (1978/79) Antigenicity of structural components from porcine transmissible gastroenteritis virus. Vet. Microbiol. 3, 179–190.

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: M.M.C. Lai and S.A. Stohlman (Eds), Coronaviruses. Adv. Exp. Med. Biol., Vol. 218, pp. 509–515. Plenum Press, New York, London.

Garwes, G.J., Stewart, F. and Britton, P. (1989) The polypeptide of M_r 14000 of porcine transmissible gastroenteritis virus: gene assignment and intracellular location. J. Gen. Virol. 70, 2495–2499.

Jacobs, L., de Groot, A., van der Zeijst, B.A.M., Horzinek, M.C. and Spaan, W. (1987) 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). Virus Res. 8, 363–371.

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.

Laude, H., Chapsal, J.-M., Gelfi, 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.

Parker, M.D., Yoo, D., Cox, G.J. and Babiuk, L.A. (1990) Primary structure of the S peplomer gene of bovine coronavirus and surface expression in insect cells. J. Gen. Virol. 71, 263–270.

Posthumus, W.P.A., Lenstra, J.A., Schaaper, W.M.M., van Nieuwstadt, A.P., Enjuanes, L. and Meloen, R.H. (1990) Analysis and simulation of a neutralizing epitope of transmissible gastroenteritis virus. J. Virol. 64, 3304–3309.

Raabe, T., Schelle-Prinz, B. and Siddell, S.G. (1990) Nucleotide sequence of the gene encoding the spike glycoprotein of human coronavirus HCV 229E. J. Gen. Virol. 71, 1065–1073.

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.

Schmidt, I., Skinner, M. and Siddell, S. (1987) Nucleotide sequence of the gene encoding the surface projection glycoprotein of coronavirus MHV-JHM. J. Gen. Virol. 68, 47–56.

Schmidt, M.F. (1982) Acylation of viral spike glycoproteins: a feature of enveloped RNA viruses. Virology 116, 327–338.

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

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

Wesley, R.D. (1990) Nucleotide sequence of the E2-peplomer protein gene and partial nucleotide sequence of the upstream polymerase gene of transmissible gastroenteritis virus (Miller strain). In: D. Cavanagh and T.D.K. Brown (Eds), Coronaviruses and Their Diseases. Adv. Exp. Med. Biol., Vol. 276. Plenum Press, New York and London. (in press).

(Received 19 June 1990; revision received 29 August 1990)