Journal of General Virology (1990), 71, 1313–1323. *Printed in Great Britain*

Four major antigenic sites of the coronavirus transmissible gastroenteritis virus are located on the amino-terminal half of spike glycoprotein S

Bernard Delmas, Denis Rasschaert, Murielle Godet, Jacqueline Gelfi and Hubert Laude*

Laboratoire de Virologie et d'Immunologie Moléculaires, Institut National de la Recherche Agronomique, Centre de Recherches de Jouy-en-Josas, 78350 Jouy-en-Josas, France

Four major antigenic sites have been delineated on the spike protein (S) of the porcine enteric coronavirus transmissible gastroenteritis virus (TGEV) in previous topological studies using monoclonal antibodies (MAbs). Correlation of these sites with the physical structure of the protein was achieved by use of different approaches. Recombinant pEX plasmids directing the synthesis of various fused S polypeptides were constructed. A hybrid protein containing nine S-specific residues (363 to 371) was shown to express site C epitopes. The other sites were localized through study of the antigenic activity of fragments generated by controlled cleavage of the native protein with different endopeptidases. Two identified cleavage products of 26K and 13K, immunoreactive to site A–B- and site D-specific MAbs respectively, could be aligned on the S primary structure according to N-terminal sequence data. This led us to propose that the major neutralization domain A–B is contained in a region of approximately 200 residues with residue 506 as its N boundary. Similarly, site D epitopes should be located within a stretch of 130 residues, starting at 82 residues from the N terminus. Point mutations identified by direct RNA sequencing of neutralization-resistant mutants were consistent with the proposed location of these sites.

Introduction

Transmissible gastroenteritis virus (TGEV) causes an acute and usually fatal enteric disease in newborn piglets. Together with the murine hepatitis virus (MHV) and infectious bronchitis virus (IBV), TGEV is one of the most well studied coronaviruses, a family of enveloped positive-stranded RNA viruses (for review, see Spaan *et al.*, 1988). The organization of the TGEV genome has been established (Rasschaert *et al.*, 1987) as well as the sequence of the genes encoding the structural and most of the non-structural proteins (Kapke & Brian, 1986; Laude *et al.*, 1987; Rasschaert & Laude, 1987; Rasschaert *et al.*, 1987; Jacobs *et al.*, 1987). TGEV virions contain three proteins, a nucleocapsid protein and two envelope glycoproteins, M (29K) and S (220K) (Garwes & Pocock, 1975; Laude *et al.*, 1986). The spike protein S is 1431 residues long, highly glycosylated, with a membrane-anchoring domain near its carboxy-terminus and a globular domain assumed to correspond to its amino-terminal half (Rasschaert & Laude, 1987). Competition studies using monoclonal antibodies (MAbs) has led to the prediction of at least four main antigenic sites (Delmas *et al.*, 1986; Garwes *et al.*, 1987; Correa *et al.*, 1988).

Major antigenic sites, designated A, B, C and D, have been delineated using the library of MAbs established in our laboratory. Most of the epitopes critical for neutralization were found to be clustered in the immunodominant A and B sites, but the other two sites contained subsidiary neutralization epitopes. The A, B and D sites have been found to be highly conserved among TGEV strains, whereas antigenic variation was observed on site C epitopes (Laude *et al.*, 1986; Delmas *et al.*, 1986). In order to gain information on the antigenic structure of S at the molecular level, but also to investigate possible vital functions of distinct epitopes, we have undertaken a systematic mapping of the antibody-binding regions. The present study reports the localization of four antigenic sites on the primary structure of S by combined approaches including production of antigenically active fragments of the protein by bacterial expression or proteolytic treatment, and analysis of epitope mutants.

Methods

Viruses and MAbs. The high-passage Purdue-115 strain of TGEV was propagated in the pig kidney cell line PD5 as reported (Laude *et al.*, 1986). Infected monolayers were maintained in Dulbecco's modified Eagle's medium (DMEM) buffered with 25 mM-PIPES pH 6·8, and

supplemented with 5% newborn calf serum (NCS). The selection of epitope mutants resistant to neutralization has already been reported (Delmas *et al.*, 1986). The 23 anti-TGEV S MAbs used in this study are listed in Table 1. Their characteristics in terms of polypeptide specificity (S and/or S′) and neutralizing activity have been reported (Laude *et al.*, 1986). Two neutralizing MAbs directed against site C (CVI-TGEV-57.24 and -57.57) were provided by A. van Nieuwstadt (Central Veterinary Institute, Lelystad, The Netherlands).

Construction of pEX clones. The cDNA clones of the TGEV S protein in pBR322 have been described (Rasschaert & Laude, 1987). cDNA fragments excised by restriction enzymes as specified in Fig. 1 were resolved in 1% low melting point agarose, then purified using the Geneclean kit (New England Biolabs). The fragments were ligated into compatible sites of the pEX1, pEX2 or pEX3 expression vectors (Stanley & Luzio, 1984) as indicated. The 5′ protruding ends of inserts and vectors were filled using Klenow DNA polymerase. *Escherichia coli* strain MC1061 carrying plasmid pCI (Km^R, CIts 857) was used as the pEX host. Recombinant plasmids were introduced into cells via the Hanahan procedure. Screening of the library was done by hybridization to inserts excised from relevant cDNA clones, followed by restriction analysis. DNA manipulations were performed following standard protocols (Maniatis *et al.*, 1982).

Construction of pEX sublibraries. The random sublibrary designated pEXS was generated by sonication (disintegrator MK2-150W; MSE) of the insert from one epitope-expressing clone; resulting fragments (150 ± 50 nucleotide long) were repaired with the Klenow fragment and then ligated into the *Sma*I-cut, calf intestine phosphatase-treated pEX2 plasmid. The pEXB sublibrary was obtained by resection of a linearized plasmid using Bal 31 exonuclease ('slow form', IBI, 0·25 units per μg of DNA, 30 °C for 50 min in the manufacturer's buffer), followed by self ligation.

Immunoscreening of bacterial colonies. Colonies transferred to a nitrocellulose membrane (BA85, Schleicher & Schuell) were incubated at 25 °C overnight for plasmid amplification; then expression of the pEX fusion gene was induced by shifting to 42 °C for 2 h (Stanley & Luzio, 1984). Cell lysis was performed by holding the membranes in a $CHCl_3$-saturated atmosphere. Positive colonies were visualized according to Huynh *et al.* (1985), using 1:100 diluted ascites fluids and ^{125}I-labelled Protein A (Amersham, 2×10^6 c.p.m. per disk).

Isolation of expression products and Western blot. Expression was induced as above in an exponential culture (optical density at 600 nm of 0·5). The *cro*–β-galactosidase fusion proteins were extracted as Triton X-100-insoluble pellets (Stanley & Luzio, 1984). SDS-solubilized proteins corresponding to 0·5 ml of culture were resolved by 10% PAGE, then electroblotted using a Milliblot SDE system (Millipore). MAb binding was visualized as in the colony blot test.

Generation of S–S′ proteolytic fragments. Radiolabelled cytosol samples were prepared as described (Laude *et al.*, 1986) with some modifications. Confluent PD5 cell monolayers in Petri dishes (10 cm diam., Costar) were infected (or mock-infected) at a multiplicity of 50 p.f.u. per cell and maintained at 38 °C in 10 ml cystine-depleted DMEM with 2% NCS. At 3 h post-infection (p.i.) 7 ml of medium was discarded and [^{35}S]cysteine (> 600 Ci/mmol, Amersham) was added at 100 μCi/ml. At 8 to 9 h p.i., the medium was discarded, the cell sheets were rinsed, and then scraped into 4·5 ml chilled radioimmunoprecipitation assay (RIPA) lysis buffer (10 mM-Tris–HCl pH 8, 2% Triton X-100, 0·15 M-NaCl, 0·6 M-KCl, 0·5 mM-$MgCl_2$, 10^3 kallikrein inhibitor units/ml aprotinin). After 1 h, this material was ultracentrifuged for 1 h at 35000 r.p.m. in a 50Ti rotor (Beckman). The supernatants were stored at −70 °C. Proteolytic digestions were performed by incubating 50 μl of cytosol at 37 °C for 2 h with the following volumes of enzyme solution (10 mg/ml): 3 μl of TLCK α-chymotrypsin (Sigma), 3 μl of collagenase from *Clostridium histolyticum* (Boehringer Mannheim), 1 μl of TPCK trypsin (Sigma) or 3 μl of V8 protease (Miles Laboratories). The reactions were terminated by adding 40 mg/ml bovine serum albumin (fraction 5, Boehringer Mannheim). TPCK and TLCK (2 mg/ml, Sigma) were also added to stop chymotrypsin and trypsin digestion, respectively.

Immunoprecipitation assays. Aliquots of 50 μl of digested products were incubated for 2 h at 37 °C with 5 μl of undiluted hybridoma ascites fluid. Antigen–antibody complexes were bound to 80 μl of a 1:1 slurry of Protein A–Sepharose 4B (Pharmacia) under gentle agitation for 1 h at room temperature. The beads were washed six times with 1 ml of RIPA buffer and once with Tris buffer, then treated for 2 min at 100 °C in Laemmli's denaturing buffer plus 5% 2-mercaptoethanol, and centrifuged. Resulting supernatants were subjected to electrophoresis on a 9 to 20% polyacrylamide gradient gel which was then processed for fluorography. ^{14}C-methylated protein M_r standards (Amersham) were used.

Amino-terminal sequencing of proteolytic fragments. Tritium-labelled proteolytic fragments were prepared as described above with the following modifications. Infected monolayers were labelled using 1 mCi/ml [^3H]valine (CEA France; 40 Ci/mmol), or 0·5 mCi/ml [^3H]threonine (Amersham; 20 Ci/mmol) or 0·5 mCi/ml [^3H]isoleucine (Amersham; 100 Ci/mmol). Additional inhibitors were used following chymotrypsin treatment: soybean trypsin inhibitor (Sigma; 3 mg/ml), PMSF (Sigma; 0·2 mM) and aprotinin (2×10^3 kallikrein inhibitor units/ml). Immunoprecipitation was done by adding 100 μl of ascites fluid to 15 ml of digested cytosol (obtained by pooling the RIPA cell extracts from three Petri dishes), followed by 400 μl of Protein A–Sepharose beads. After a 2 h agitation, the beads were washed with 100 ml of RIPA buffer in eight steps, heated for 2 min in 200 μl Laemmli's buffer and centrifuged. PAGE of resulting supernatants was performed. A piece of gel containing the fragment was excised after localization of its position by reference to that of M_r standard proteins stained with Coomassie blue (Pharmacia; 14K to 94K), and then electroeluted as described previously (Laude *et al.*, 1987). The location of tritiated residues on recovered material was determined by automated gas-phase analysis on an Applied Biosystems 470A apparatus. Lactoglobulin (200 pmol) was included as a non-labelled control in order to check the sequential Edman degradation. Radioactivity from eluted fractions was counted by liquid scintillation (Minaxi 4000, Packard).

Dideoxynucleotide sequencing of supercoiled plasmid DNA and of genomic RNA. Sequencing of double-stranded DNA templates was done according to Lim & Pène (1988). Plasmid DNA prepared by the alkali-SDS method was precipitated by 6·5% polyethylene glycol 8000 plus 0·8 M-NaCl (Lin *et al.*, 1985), and then denatured by NaOH. Two oligonucleotides (151 and 190) complementary to pEX DNA sequences were used as primers. Reverse transcription of genomic RNA was performed as originally described by Zimmern & Kaesberg (1978). RNA matrixes were obtained by proteinase K–SDS treatment of virions semi-purified by ultracentrifugation through a 25% glycerol cushion, followed by phenol–chloroform extraction. After precipitation with 2 M-LiCl, RNA was resuspended in water and heat-denatured (100 °C for 1 min). Priming was done using 20-mer oligonucleotides complementary to the TGEV S sequence (Rasschaert *et al.*, 1987). The elongation reaction was performed at 42 °C for 20 min using 1 to 2 μg RNA per sequence, avian myeloblastosis virus reverse transcriptase (Appligène), [^{35}S]dATP (Amersham) and a deoxynucleotide : dideoxynucleotide ratio of 3·75:1, 5:1, 6:1 and 12·5:1 respectively, for nucleotides T, C, G and A. The samples were analysed on a 6% acrylamide–8 M-urea gel. The synthetic oligonucleotides were synthesized on a Biosearch 8600 apparatus.

Results

Subcloning of S cDNA fragments into pEX vector and identification of clones expressing epitopes

The inducible pEX bacterial expression system was used to produce fragments of TGEV S as a C-terminal extension of *cro–β*-galactosidase. A library was created by insertion of seven partially overlapping cDNA fragments into compatible pEX plasmids. The fragments were between 500 and 1500 nucleotides in size and covered the entire S gene, except for 105 nucleotides at the 5′ end (Fig. 1). Six to 10 recombinant clones from each construction were examined for expression of a chimeric protein after heat induction. Then two or three of the clones that produced the highest level of fusion protein were subjected to colony blot immunoscreening with our panel of anti-S MAbs. None of the MAbs directed against A, B and D sites or unrelated epitopes was found to react with any of the expression products. In contrast, MAbs 3b.5, 10.4 and 11.20, all defining site C, recognized the expression products relative to fragments 1, 3 and 4 (data not shown).

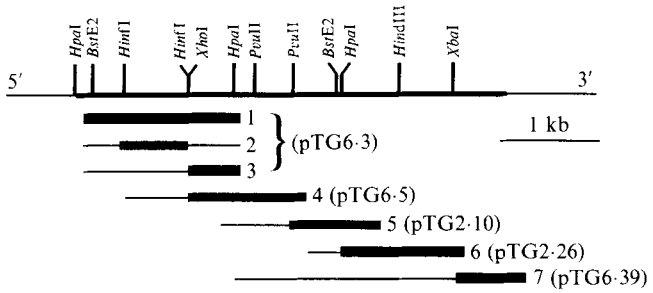

Fig. 1. Fragments of TGEV S cDNA subcloned in the pEX expression vector. The fragments, numbered 1 to 7, are viewed as solid bars below the restriction map of the TGEV S gene. They were produced from five pTG clones indicated in brackets (TGEV library in pBR322; Rasschaert & Laude, 1987). Fragments 1 (total insert, *Pst*I digest) and 2 (*Hin*fI digest) were ligated into *Pst*I-cut pEX1 and *Bam*HI-cut pEX3, respectively. Other fragments (3 and 4, *Hin*fI–*Pst*I digest; 5, *Pvu*II–*Pst*I digest; 6, *Hpa*I–*Pst*I digest; 7, *Xba*I–*Pst*I digest) were ligated into the appropriate *Bam*HI–*Pst*I-cut pEX.

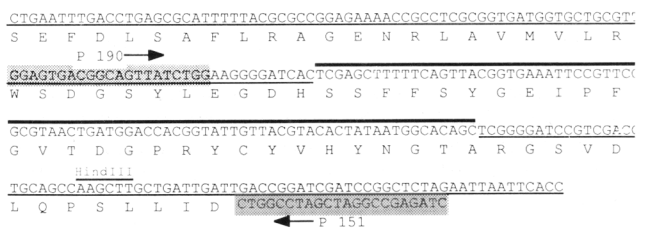

Fig. 2. Partial nucleotide sequence of pEXS8. The S-specific sequence of 89 nucleotides (overlined) fused to the pEX sequence (underlined) is shown, together with the predicted amino acid sequence of the resulting fusion protein. The pEX-specific primers 151 and 190 used for direct sequencing of plasmid DNA are indicated (primer 190 is specific for the truncated pEX; see text).

Generation of pEX sublibraries for delineating site C

Random fragments averaging 150 nucleotides produced by sonication of the insert excised from pEXG16 (fragment 3 in pEX3) by *Sma*I–*Hin*dIII digestion were subcloned into a *Sma*I-cut dephosphorylated pEX2 plasmid. The resulting sublibrary was screened against MAb 3b.5. Direct supercoiled plasmid sequencing of 10 of the positive clones was performed using oligonucleotide 151 as a 3′ primer. The shortest insert, which was carried by plasmid pEXS8, was 89 nucleotides long and encoded 28 S-specific amino acids (Fig. 2).

PAGE analysis of the pEXS8 expression product revealed an M_r value of 33K to 35K (Fig. 3*b*, lane 2), whereas the predicted M_r of the *cro–β*-galactosidase chain is 117K to 119K (Stanley & Luzio, 1984). Upon re-examination of the nucleotide sequence, it appeared that the S-specific sequence was bound 2495 bp upstream of the 3′ end of the *lacZ′* gene. The occurrence of a large deletion, most likely during the vector dephosphorylation step, thus explains the unexpected M_r of the pEXS8 expression product. This was confirmed by sequencing the opposite strand with primer 190 which is specific for the truncated pEX (Fig. 2).

Sequencing of the other nine random inserts, which ranged in size from 129 to 259 bp, revealed a striking feature: the S cDNA sequences were all 5′ coterminal together with that of the pEXS8 insert (data not shown). This suggested that the deduced N-terminal residues common to these 10 fusion proteins corresponded to one limit of site C. In order to determine the C-terminal limit of the antibody-binding site, we produced a second sublibrary by nuclease Bal 31 resection of *Hin*dIII-cut pEXS8 plasmid (*Hin*dIII was a unique restriction site located downstream of the S sequence; see Fig. 2). A series of plasmids from clones positive in colony blot immunoscreening were sequenced as above using primer 190. Thus the S cDNA sequences identified as the three shortest, encoded nine or 10 amino acids (Fig. 3*a*). The immunoreactivity towards MAb 3b.5 of the related expression products was confirmed by Western blotting (Fig. 3*b*). The immunoreactivity of these three fusion proteins towards the panel of site C-specific MAbs was also compared in a colony blot test (Fig. 3*c*). MAbs 57.24 and 57.57 exhibited a reactivity similar to that of 3b.5, whereas the reactivity of MAbs 10.4 and 11.20 was weaker as found also with the fusion protein of 28 residues.

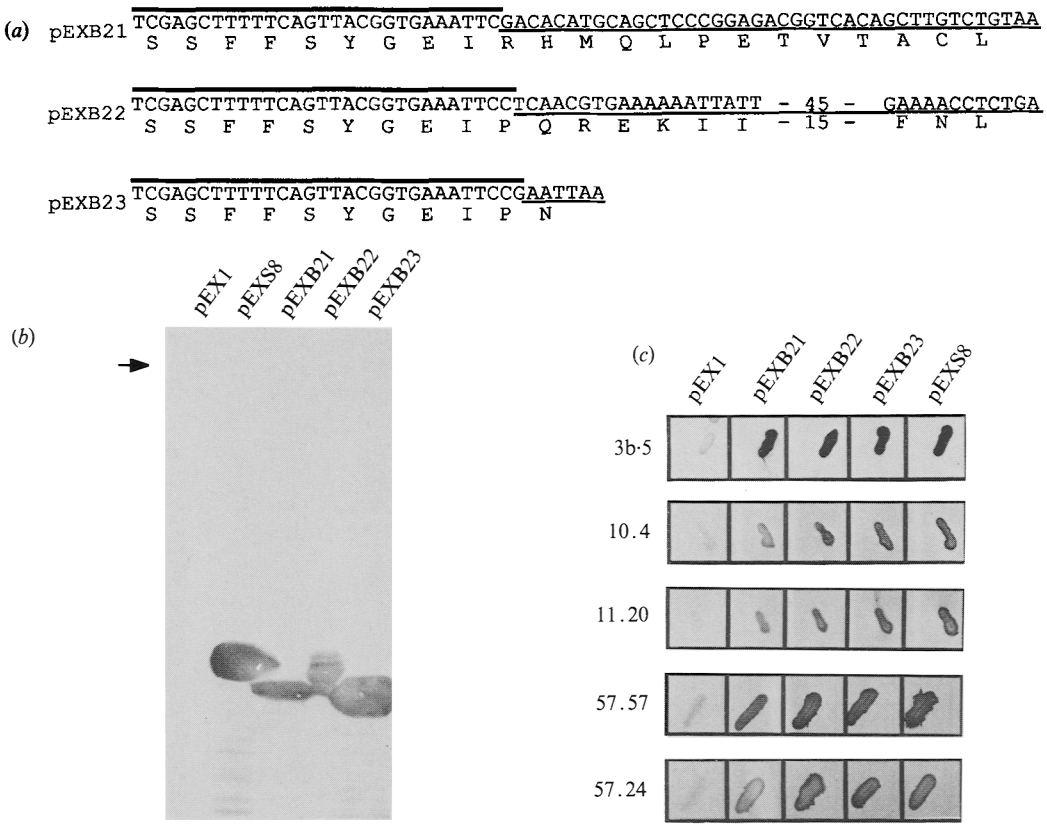

Fig. 3. Sequence and immunoreactivity of the fusion protein specified by three clones generated by Bal 31 resection of pEXS8. (*a*) Partial nucleotide sequence of pEXB21, 22 and 23 plasmids and predicted amino acid sequence of the C terminus of the encoded fusion proteins. The S-specific amino acid sequences are underlined. (*b*) Western blotting analysis with MAb 3b.5 of the pEX fusion proteins. The observed variation in M_r is related to the fact that non-identical stop codons were created in Bal 31-resected plasmids. Arrow shows the position of the original *cro*–*β*-galactosidase protein (118K). (*c*) Immunoreactivity of the indicated pEX proteins towards S site C-specific MAbs analysed in a colony blot test.

Immunoprecipitation profiles of S proteolytic fragments

Partial cleavages of the protein were performed with four different endopeptidases. The resulting fragments were subjected to immunoprecipitation using anti-S–S′ or anti-S′ MAbs. S′ (175K, formerly called E′2), designates the underglycosylated precursor of S (220K); the two species are present in similar amounts in infected cell extracts (Laude *et al.*, 1986). Four main profiles were characterized following collagenase cleavage (Fig. 4). Profile 1, observed with site A- or B-specific MAbs, was distinguished by the presence of a prominent 26K fragment (CO-26K), together with fragments of higher M_r (135/200K and 62/72K). Profile 2 was obtained with site D-specific MAbs and also with the unrelated MAbs 44.4, 76.2 and 6.179. All these MAbs strongly precipitated three groups of fragments with M_r values 36/43K, 62/72K and 130/200K. Site C-specific MAbs and MAb 67.9 precipitated most of these fragments weakly (profile 3). The presence of 22K and 65K bands (profile 4) was typical for four out of six of the tested S′-specific MAbs. Moreover, these MAbs precipitated high M_r material (> 250K) which was never observed with uncleaved S material and was likely to represent aggregated fragments. An overall view of the profiles generated by collagenase and by the other assayed proteases (see below) is presented in Table 1.

The chymotrypsin digestion profiles obtained with the S-specific MAbs showed essentially quantitative dissimilarities (Fig. 5). As a main feature, a small fragment of 13K (CT-13K) strongly reacted with all site D-specific MAbs. In particular, MAbs 78.17 and 69.21 precipitated CT-13K as a major fragment. Most of the fragments from CT digestion were precipitated weakly or not at all, by site A-, B- and C-specific MAbs. Again, the S′-specific MAbs yielded a unique profile composed of 21/28K, 50K and 72/86K species and of aggregated material (data not shown). Controlled trypsin digestion produced three 'nested' profiles: 90/175K bands were precipitated by site A-, B-, C- and D-specific MAbs, an

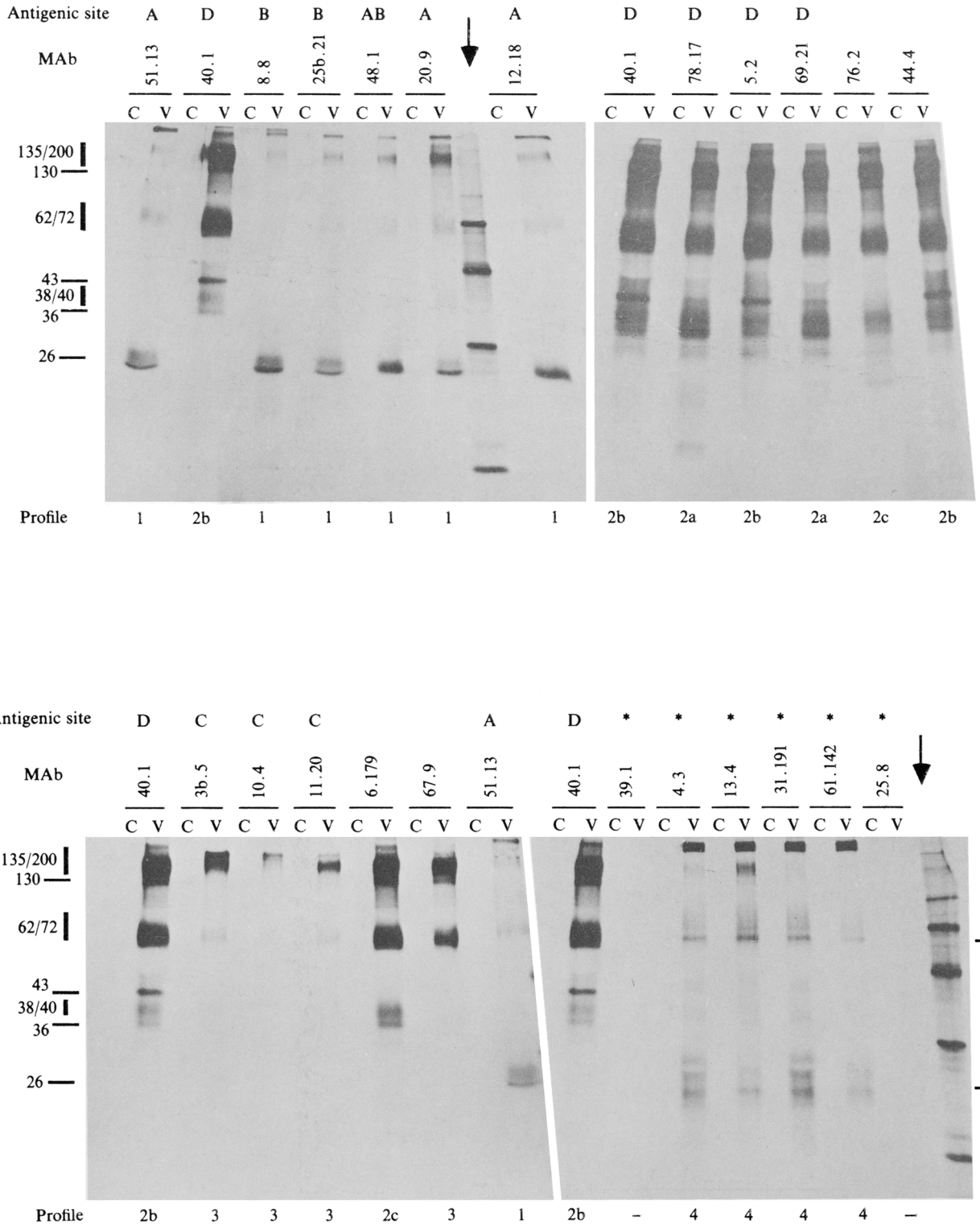

Fig. 4. Immunoreactivity analysis of collagenase-generated fragments of the S protein. [^{35}S]Cysteine-labelled extracts from infected (V) or uninfected (C) cultures were incubated with collagenase (0·6 mg/ml) and immunoprecipitated using the indicated MAbs (S′-specific MAbs are denoted by an asterisk). Immune complexes were analysed in SDS–9/20% gradient PAGE in reducing conditions. An autoradiograph of the fluorographed gel is shown. The relative M_r ($\times 10^{-3}$) were determined by reference to ^{14}C-methylated marker proteins (lanes shown by arrows). The immunoprecipitation profiles are identified by a number below each lane (minor differences in the pattern led to the definition of subprofiles: 2a, 2b, etc.).

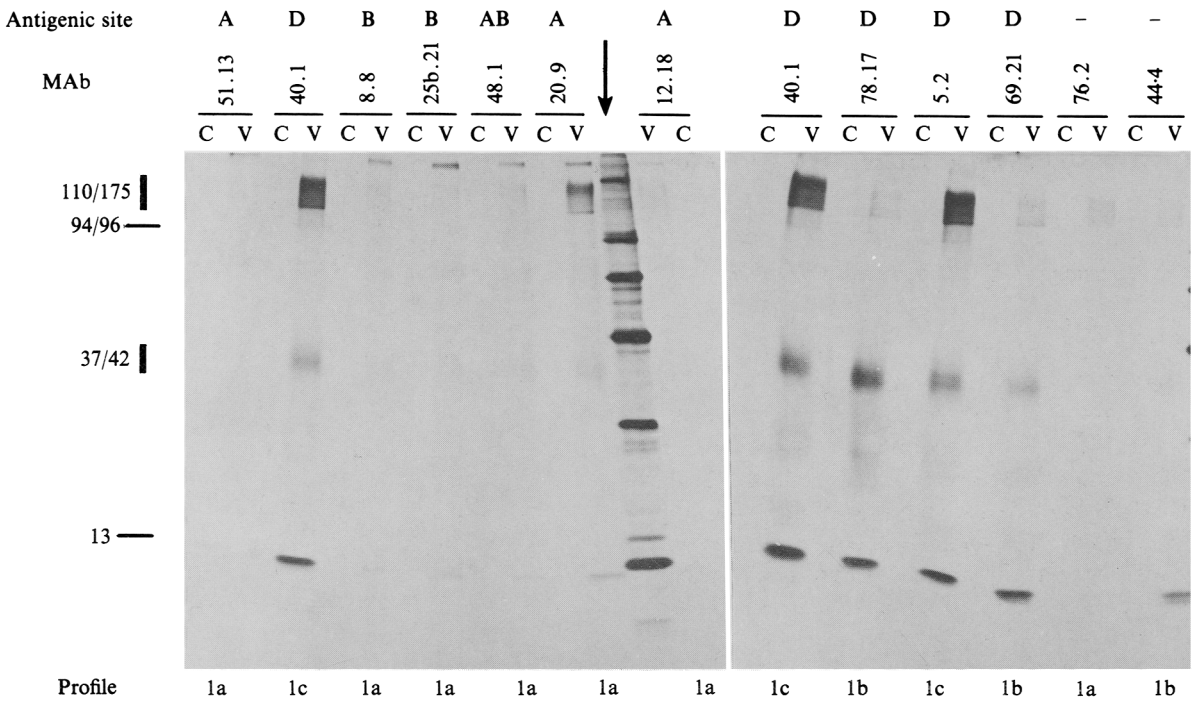

Fig. 5. Immunoreactivity analysis of chymotrypsin-generated fragments of the S protein. Legend as in Fig. 4 (only subprofiles were defined).

Table 1. *Immunoprecipitation patterns of TGEV S protein proteolytic fragments*

Specificity	MAb designation	Antigenic site	Neutralizing activity†	Protease*			
				CO	CT	V8	TR
Anti-S + S′	51.13	A	++	1‡	1a	1	1
	12.18	A	++	1	1a	1	1
	20.9	A	++	1	1a	1	1
	48.1	A–B	++	1	1a	1	1
	8.8	B	++	1	1a	1	1
	25b.21	B	++	1	1a	1	1
	78.17	D	–	2a	1b	2a	2
	69.21	D	–	2a	1b	2a	2
	40.1	D	++	2b	1c	2b	2
	5.2	D	+	2b	1c	2b	2
	3b.5	C	+	3	1a	2b	3
	10.4	C	+	3	–§	2c	3
	11.20	C	–	3	1a	2c	3
	76.2		–	2c	1a	3	ND‖
	44.4		–	2b	1b	2d	2
	6.179		–	2c	1c	2b	ND
	67.9		–	3	–	–	ND
Anti-S′	31.191		–	4	2	ND	ND
	4.3		–	4	ND	ND	ND
	13.4		–	4	2	ND	ND
	61.142		–	4	2	ND	ND
	39.1		–	–	–	ND	ND
	25.8		–	–	–	ND	ND

* Abbreviations: CO, collagenase; CT, chymotrypsin; V8, protease V8; TR, trypsin.
† ++, Strong neutralization; +, neutralization; –, no neutralization.
‡ Number refers to precipitation pattern shown for respective enzymes (partial data in Fig. 1 and 2).
§ No profile was defined because reactivity between antibody and fragment was too weak.
‖ ND, Not determined.

80K band by site C- and D-specific MAbs, and a 45K band by D-specific MAbs only. Two main profiles were defined using V8 protease: the major fragments were 45/49K, 68/72K with site A- and B-specific MAbs, and 36/47K, 135/200K with the other S–S′ MAbs, except MAb 76.2 which produced an intermediate profile (data not shown).

N-terminal sequencing of low M_r reactive fragments

The above experiments enabled us to characterize two S cleavage products, CO-26K and CT-13K, which potentially retained a substantial reactivity towards MAbs defining sites A–B or D, respectively. The effect of non-reducing conditions on the mobility of these fragments was examined in order to determine whether or not they could be linked through a disulphide bridge to a larger immunoprecipitated fragment. No change was detected, thus establishing that the antibody-binding site was on the fragment itself in both cases (data not shown). Hence, it was of interest to orientate CO-26K and CT-13K on the primary structure of S.

Radiosequencing proved to be necessary for determination of their N-terminal extremity owing to the limiting amount of cleavage product. Analysis of the [^{3}H]threonine- or [^{3}H]isoleucine-labelled CO-26K fragment precipitated by MAb 48.1 (site A–B) identified Thr residues at positions 4, 6 and 11, and Ile residues at positions 7, 10, 12 and 25 (Fig. 6*a* and *b*). Sequencing of the CT-13K fragment precipitated by MAb 40.1 (site D) revealed Val residues at positions 8, 9, 10, 14 and 19 (Fig. 6*c*). However, the yield of residue 14 was lower than that of residue 19, indicating the presence of more than one fragment. Indeed, reverse-phase HPLC revealed that the material consisted of two species in the ratio of 3:1. Sequencing of the major species no longer identified a Val residue at position 14 (data not shown). As mentioned above, a CT-13K fragment also reacted weakly with many non-site D-specific MAbs. Analysis of the labelled CT-13K fragment prepared with MAb 20.9 (site A) also revealed Val residues at positions 8, 9, 10 and 19 (Fig. 6*d*).

Sequence analysis of epitope mutants

Identification of amino acid changes conferring neutralization resistance to previously selected independent epitope mutants was a way to identify precisely the structure of the different antigenic sites. The sequence of the gene portions encoding all of the CO-26K and CT-13K regions and of about 200 nucleotides covering the region predicted to encode site C were determined by reverse transcription of the genomic RNA using appropriate DNA primers. The predicted amino acid sequence

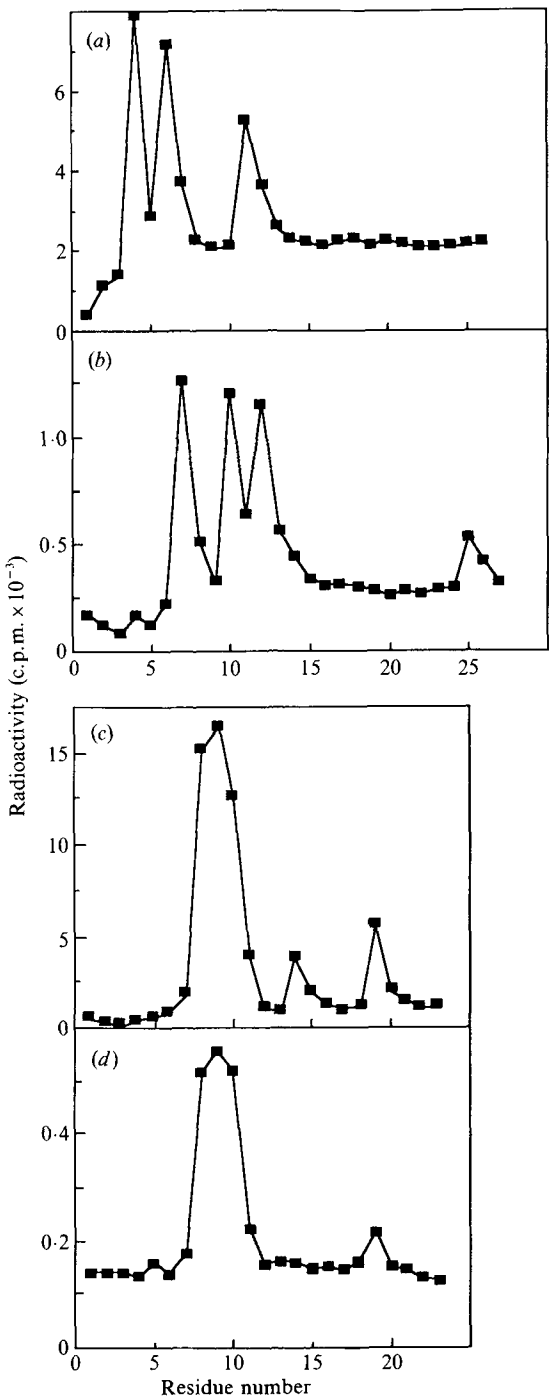

Fig. 6. Partial N-terminal sequencing of S proteolytic fragments. Immunoprecipitated ^{3}H-labelled fragments were isolated after resolution in SDS–PAGE. Radioactivity recovered at each cycle is shown. (*a* and *b*) CO-26K fragment prepared with MAb 48.1; (*a*) threonine-labelled CO-26K, 6×10^5 c.p.m.; (*b*) isoleucine-labelled CO-26K, 2×10^5 c.p.m. (*c* and *d*) Valine-labelled CT-13K fragment prepared using (*c*) MAb 40.1, $1{\cdot}5 \times 10^6$ c.p.m. or (*d*) MAb 20.9, 1×10^5 c.p.m.

Table 2. *Predicted amino acid changes of TGEV S epitope mutants*

Selector MAb	Site	Mutant	Nucleotide substitution	Amino acid change	Position
20.9	A	1,2,3,5,6	GAC→AAC	Asp→Asn	570
48.1	A–B	1,3	ACA→CCA	Thr→Pro	533
		4,2	ACA→AAA	Thr→Lys	533
		5	ACA→GCA	Thr→Ala	533
3b.5	C	1,2	AGU→AUU	Ser→Ile	367
10.4	C	1,2	UAC→UGC	Tyr→Cys	368
40.1	D	3	UGU→CGU	Cys→Thr	147
		1,9	UCU→CCU	Ser→Pro	149

for the parental virus agreed completely with that determined for cDNA clones (Rasschaert & Laude, 1987), except for Leu instead of Ile at position 690 (26K region). As shown in Table 2, each mutant analysed contained a single nucleotide substitution specifying one non-conservative amino acid change.

Discussion

Although there are several published reports dealing with the physical mapping of B cell-related epitopes on coronavirus MHV and IBV spike proteins (Makino *et al.*, 1987; Cavanagh *et al.*, 1988; Talbot *et al.*, 1988; Lenstra *et al.*, 1989; Luytjes *et al.*, 1989), the emerging picture is still largely incomplete. The present study provides molecular data on the antigenic structure of the TGEV S protein. Our findings establish that four previously delineated major antigenic sites (Delmas *et al.*, 1986) are all located in the amino-terminal half of the polypeptide chain.

In the first series of experiments, large restriction fragments or small random fragments of the S protein were expressed in *E. coli* using the pEX system. Only epitopes belonging to site C could be detected by this approach. Subsequently, accurate mapping of this site could be achieved through subcloning in pEX of Bal 31-resected cDNA fragments. A hybrid protein containing nine amino acid residues was shown to be recognized by all the available five site C-specific MAbs (Fig. 3). From this result it was concluded that the sequence stretching from residue 363 to 371 of the mature protein is an essential component of site C. There is thus a good correlation between these data and those obtained through a peptide scanning approach (W. Posthumus *et al.*, IV International Coronavirus Symposium, Cambridge, July 1989). The existence of a potential antigenic site at this position was predicted by only one of the five assayed antigenicity algorithms, i.e. the accessibility analysis (Janin, 1979), which gave a high score for the first five residues. Besides, no ordered secondary structure was predicted in the 40 surrounding residues (Gibrat *et al.*, 1987), which suggests that site C is located on an extended loop structure protruding at the spike surface.

Use of the pEX system in the case of MHV and IBV proteins (Luytjes *et al.*, 1989; Lenstra *et al.*, 1989) enabled the identification of potentially linear epitopes in the amino-terminal region of the S2 subunit, but not in the S1 (amino) subunit, in contrast to our findings on TGEV. The successful expression of site C as a bacterial fusion protein is consistent with other observations which suggest that it is composed of largely conformation-independent epitopes: glycosidase treatment affected binding of site C-specific MAbs to S protein moderately, in contrast to site A-, B- and D-specific MAbs (B. Delmas, unpublished results), and a high proportion of site C-specific hybridomas were isolated using a Western blotted, partially denatured antigen for screening (A. Van Nieuwstadt, personal communication).

Additional information about the structure of site C was provided by comparison of the homologous sequences available for selected mutants of Purdue virus, for one TGEV variant strain and for one strain of feline infectious peritonitis virus (FIPV). It appears from the data in Table 3 that each of the three residues Phe(366)-Ser-Tyr(368) is critical for neutralization of infectivity. Despite adjacent amino acid changes, mutants 3b.5 and 10.4 did not exhibit functional reciprocity (i.e. 3b.5 mutants were still recognized by MAb 10.4). Also neutralization resistance was not always associated with loss of antibody-binding (e.g. 10.4 mutants). Finally, residues 365 to 368 might not be crucial for binding by non-neutralizing MAb 11.20 (which does not discriminate the FIPV strain). Altogether, the above findings lend support to the view that the sequence of nine amino acids corresponding to site C comprises several potentially linear epitopes having distinct functional properties. This is not surprising in the light of recent studies which show that only four to five residues determine specificity and binding energy in epitopes (Geysen *et al.*, 1988).

Table 3. *TGEV S protein amino acid changes* affecting the expression of site C epitopes†*

		MAb				
		3b.5		10.4		11.20
		N‡	IF§	N	IF	IF
363	371					
S S F F S Y G E I	TGEV Purdue	+	+	+	+	+
– – – – I – – – –	TGEV mutant 3b.5	–	–	+	+	+
– – – – – C – – –	TGEV mutant 10.4	–	+	–	+	+
– – – S – – – – –	TGEV FS772/70	–	–	–	–	+
– – S Y – – – – –	FIPV 79-1146	–	–	–	–	+

* Sequence data: selected mutants of Purdue strain, this study; British TGEV strain FS772/70, P. Britton, personal communication; FIPV strain 79-1146, de Groot *et al.* (1987).

† Serological data: Delmas *et al.* (1986); J. Gelfi, unpublished results; MAbs 57.57 and 57.24 had a reactivity pattern similar to that of MAb 3b.5.

‡ Neutralization test.

§ Indirect immunofluorescence test.

Bacterial expression of several large fragments covering the entire S polypeptide (except the 19 N-terminal residues) failed to produce fusion proteins expressing sites A, B or D. The approximate location of the latter sites was determined through the following two-step approach, which is theoretically more appropriate for mapping conformation-dependent epitopes: immunoanalysis of proteolytic fragments generated from the native protein and microsequencing of the smallest fragment retaining substantial antigenic activity. The results obtained using four different endopeptidases were in agreement with the previously proposed epitope map, as MAbs defining the same antigenic site yielded a similar profile of cleavage products (Table 1). However, sites A and B could not be distinguished in the present study, consistent with the fact that potentially they share a common epitope (MAb 48.1). Also, these data indicated that site D was composed of two distinct subsites (see Table 1), a finding which seems to be functionally relevant because the two site D-specific MAbs capable of mediating virus neutralization belong to the same subsite. Finally, the fragmentation analysis strengthened our earlier assumption (Delmas *et al.*, 1986) that MAbs against the S′ species define a class of epitopes quite distinct and possibly distant from those recognized by S–S′ MAbs.

The above experiments allowed the identification of two antigenic cleavage products of limited size which subsequently could be orientated on the S polypeptide chain by partial amino acid sequencing (Fig. 6). Collagenase digestion produced a 26K fragment expressing all the epitopes belonging to the antigenic domain A–B. Comparison of the radiosequencing data with the amino acid sequence identified unequivocally the N-terminal residue of CO-26K as serine 506 of the mature S protein (Fig. 7). The cleavage appeared to occur just behind a proline, whereas the cleavage site of collagenase from *C. histolyticum* usually involves a proline at position ± 2; however, the substrate specificity of this enzyme has not yet been fully documented (Van Wart & Steinbrink, 1985). The C terminus of CO-26K was not determined but, on account of the size of the fragment and the presence of five potential *N*-linked glycosylation sites, it can be predicted to be located around the proline residue 718. Accordingly, antigenic sites A and B should be located within a stretch of about 200 residues with its carboxy terminus near the middle of the S chain. Further evidence was afforded by the localization of point mutations in the pertinent gene of escape mutants (Table 2). The changes conferring neutralization resistance to MAbs 20.9 (site A) and 48.1 (site A–B) were located within the N-terminal third of CO-26K (Fig. 7). The distance between the two sites of mutation, i.e. 37 residues, is consistent with our earlier observation that MAb 48.1 still recognized escape mutants of MAb 20.9 and vice versa. Escape mutations of site A have also been mapped in this region of S by another group (L. Enjuanes *et al.*, IV International Coronavirus Symposium, Cambridge, July 1989).

Radiosequencing also enabled us to localize unambiguously the chymotrypsin cleavage fragment CT-13K, which strongly reacted to all site D-specific MAbs. Its N terminus was found to correspond to the S residue Asn(82) (Fig. 7). The immediate upstream residue was tryptophan, an aromatic amino acid which frequently precedes the chymotrypsin cleavage site (Keil, 1987). The C terminus of CT-13K might be located around the triplet of aromatic residues, Trp-Trp-Phe, found at positions 210 to 212 as the fragment is approximately 130 residues long and devoid of an *N*-glycosylation site.

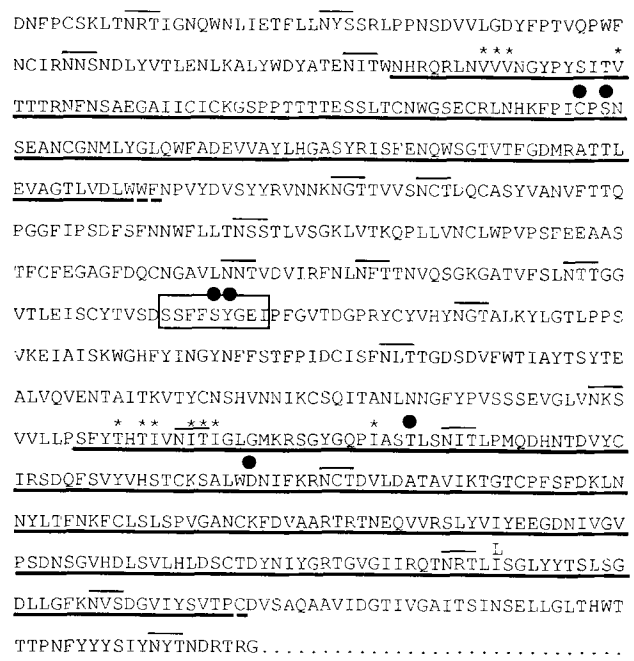

Fig. 7. Location of antigenic regions on the TGEV S primary structure. The amino acid sequence of the amino-half of the mature protein is shown (data from Rasschaert & Laude, 1987); potential glycosylation sites are overlined. The nine amino acid sequence expressing site C epitopes is boxed. The proposed locations of immunoreactive fragments CT-13K (site D) and CO-26K (sites A and B) are underlined. Valine, threonine and isoleucine residues identified by N-terminal radiosequencing are marked by an asterisk (*). The symbol (●) shows the position of single amino acid changes imparting neutralization resistance.

Thus, site D epitopes should be included in the amino acid stretch 82 to 212. The location of point mutations affecting the integrity of the epitope reacting with MAb 40.1 at positions 147 and 149 fully supports this conclusion. One amino acid change leading to neutralization resistance has also been found in the N-terminal region of the IBV S protein (Cavanagh *et al.*, 1988). Finally, our fragmentation analysis suggested that the unrelated epitopes recognized by MAbs 44.4 and 6.179 are also located in the region 82 to 212 of TGEV S.

Taken together, the above results led to the conclusion that the A–B domain, the D site and the C site occupy physically well distinct regions of the S molecule. Nevertheless, as mentioned in the Results section, the CT-13K fragment appeared to exhibit a weak reactivity towards non-site D-specific MAbs, in particular site A–B-specific MAbs. This might be explained by the involvement of residues of the CT-13K region in the antibody-binding site of A and B epitopes. Recent data suggesting that site A-specific MAbs recognize the N-terminal region of S (L. Enjuanes *et al.*, IV International Coronavirus Symposium, Cambridge, July 1989), brings some support to this hypothesis. It should be pointed out, however, that a CT-13K fragment was also weakly revealed by two of the three site C-specific MAbs and by two of the five S′-specific MAbs (data not shown). Therefore the possibility should not be ignored that a non-covalent interaction between the CT-13K region and another region of the S molecule might enable some CT-13K fragments to coprecipitate with large fragments.

In contrast to site C, sites A, B and D are extensively conserved among TGEV strains. Furthermore, all MAbs defining the major neutralization domain A–B have been shown to react with FIPV (strain 79-1146), whereas site D-specific MAbs, as well as two MAbs directed against unrelated epitopes positioned in the same region, exhibited no cross-reactivity (Delmas *et al.*, 1986; Laude *et al.*, 1988). The physical location of these sites provides a molecular basis for a better understanding of the antigenic relationship between TGEV and FIPV. An optimal alignment of their predicted amino acid sequences in the 26K (A–B) and 13K (D) regions revealed a homology level of 92 and 25%, respectively (sequence data from FIPV S; de Groot *et al.*, 1987). This led us to propose that the high level of cross-neutralization observed between the porcine, feline and canine coronaviruses (Horzinek *et al.*, 1982) is due to the clustering of most of the critical neutralization epitopes within the structurally conserved and immunodominant 26K region. Elucidating the biological function of the virus- and group-specific regions represents an enticing area for future research.

We thank J. C. Huet and J. C. Pernollet (Laboratoire d'Etude des Protéines, Versailles, France) for their helpful collaboration in N-terminal sequencing and A. Van Nieuwstadt (Central Veterinary Institute, Lelystad, The Netherlands) for providing 57.57 and 57.24 MAbs. Thanks are also due to A. M. Spite for typing the manuscript and to K. Rérat for revising the English manuscript. The contribution of the Commissariat à l'Energie Atomique (Saclay, France), who supplied labelled precursors in a suitable form for our experiments is gratefully acknowledged. Part of the results were presented at the IV International Coronavirus Symposium (Cambridge, U.K., July 1989).

References

CAVANAGH, D., DAVIS, P. J. & MOCKETT, A. P. A. (1988). Amino acids within hypervariable region 1 of avian coronavirus IBV (Massachusetts serotype) spike glycoprotein are associated with neutralization epitopes. *Virus Research* **11**, 141–150.

CORREA, I., JIMENEZ, G., SUNE, C., BULLIDO, M. J. & ENJUANES, L. (1988). Antigenic structure of the E2 glycoprotein from transmissible gastroenteritis coronavirus. *Virus Research* **10**, 77–94.

DE GROOT, R. J., MADURO, J., LENSTRA, J. A., HORZINEK, M. C., VAN DER ZEIJST, B. A. M. & SPAAN, W. J. M. (1987). cDNA cloning and sequence analysis of the gene encoding the peplomer protein of feline infectious peritonitis virus. *Journal of General Virology* **68**, 2639–2646.

DELMAS, B., GELFI, J. & LAUDE, H. (1986). Antigenic structure of transmissible gastroenteritis virus. II. Domains in the peplomer glycoprotein. *Journal of General Virology* **67**, 1405–1418.

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

GARWES, D. J., STEWART, F. & ELLEMAN, C. J. (1987). Identification of epitopes of immunological importance on the peplomer of porcine transmissible gastroenteritis virus. In *Coronaviruses*, pp. 509–515. Edited by M. M. C. Lai & S. Stohlman. New York & London: Plenum Press.

GEYSEN, H. M., MASON, T. J. & RODDA, S. J. (1988). Cognitive features of continuous antigenic determinants. *Journal of Molecular Recognition* **1**, 32–41.

GIBRAT, J. F., GARNIER, J. & ROBSON, B. (1987). Further developments of protein secondary structure prediction using information theory. New parameters and consideration of residue pairs. *Journal of Molecular Biology* **198**, 425–443.

HORZINEK, M. C., LUTZ, H. & PEDERSEN, N. C. (1982). Antigenic relationships among homologous structural polypeptides of porcine, feline and canine coronaviruses. *Infection and Immunity* **37**, 1148–1155.

JACOBS, L., DE GROOT, R., VAN DER ZEIJST, B. A. M., HORZINEK, M. C. & SPAAN, W. J. M. (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 Research* **8**, 363–371.

HUYNH, T. V., YOUNG, R. A. & DAVIS, R. W. (1985). Constructing and screening cDNA libraries in λgt10 and λgt11. In *DNA Cloning*, vol. **1**, pp. 49–78. Edited by D. M. Glover. Oxford & Washington, D.C.: IRL Press.

JANIN, J. (1979). Surface and inside volumes in globular proteins. *Nature, London* **277**, 491–492.

KAPKE, P. A. & BRIAN, D. A. (1986). Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. *Virology* **151**, 41–49.

KEIL, B. (1987). Proteolysis data bank: specificity of alpha-chymotrypsin from computation of protein cleavages. *Protein Sequence Data Analysis* **1**, 13–20.

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.

LAUDE, H., RASSCHAERT, D. & HUET, J.-C. (1987). Sequence and N-terminal processing of the transmembrane protein E1 of the coronavirus transmissible gastroenteritis virus. *Journal of General Virology* **68**, 1687–1693.

LAUDE, H., GELFI, J., RASSCHAERT, D. & DELMAS, B. (1988). Caractérisation antigénique du coronavirus respiratoire porcin à l'aide d'anticorps monoclonaux dirigés contre le virus de la gastroentérite transmissible. *Journées Recherche Porcine en France* **20**, 89–94.

LENSTRA, J. A., KUSTERS, J. G., KOCH, G. & VAN DER ZEIJST, B. A. M. (1989). Antigenicity of the peplomer protein of infectious bronchitis virus. *Molecular Immunology* **26**, 7–15.

LIM, H. M. & PÈNE, J. J. (1988). Optimal conditions for supercoil DNA sequencing with the *Escherichia coli* DNA polymerase I large fragment. *Gene Analysis Techniques* **5**, 32–39.

LIN, H. C., LEI, S. P. & WILCOX, G. (1985). An improved DNA sequencing strategy. *Analytical Biochemistry* **147**, 114–119.

LUYTJES, W., GEERTS, D., POSTHUMUS, W., MELOEN, R. & SPAAN, W. J. M. (1989). Amino acid sequence of a conserved neutralizing epitope of murine coronaviruses. *Journal of Virology* **63**, 1408–1412.

MAKINO, S., FLEMING, J. O., KECK, J. G., STOHLMAN, S. A. & LAI, M. M. C. (1987). RNA recombination of coronaviruses: localization of neutralizing epitopes and neuropathogenic determinants on the carboxyl terminus of peplomers. *Proceedings of the National Academy of Sciences, U.S.A.* **84**, 6567–6571.

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory.

RASSCHAERT, D. & LAUDE, H. (1987). The predicted primary structure of the peplomer protein E2 of the porcine coronavirus transmissible gastroenteritis virus. *Journal of General Virology* **68**, 1883–1890.

RASSCHAERT, D., GELFI, J. & LAUDE, H. (1987). Enteric coronavirus TGEV: partial sequence of the genomic RNA, its organization and expression. *Biochimie* **69**, 591–600.

SPAAN, W., CAVANAGH, D. & HORZINEK, M. C. (1988). Coronaviruses: structure and genome expression. *Journal of General Virology* **69**, 2939–2952.

STANLEY, K. K. & LUZIO, J. P. (1984). Construction of a new family of high efficiency bacterial expression vectors: identification of cDNA clones coding for human liver proteins. *EMBO Journal* **3**, 1429–1434.

TALBOT, P. J., DIONNE, G. & LACROIX, M. (1988). Vaccination against lethal coronavirus-induced encephalitis with a synthetic decapeptide homologous to a domain in the predicted peplomer stalk. *Journal of Virology* **62**, 3032–3036.

VAN WART, H. E. & STEINBRINK, D. R. (1985). Complementary substrate specificities of class I and class II collagenases from *Clostridium histolyticum*. *Biochemistry* **24**, 6520–6526.

ZIMMERN, D. & KAESBERG, P. (1978). 3′-terminal nucleotide sequence of encephalomyocarditis virus RNA determined by reverse transcriptase and chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **75**, 4257–4261.

(*Received 4 December 1989; Accepted 9 February 1990*)