ENTERIC CORONAVIRUS TGEV : MAPPING OF FOUR MAJOR ANTIGENIC DETERMINANTS IN THE AMINO HALF OF PEPLIMER PROTEIN E 2

B. Delmas, M. Godet, J. Gelfi, D. Rasschaert, H. Laude

I.N.R.A.

Laboratoire de Virologie et d'Immunologie Moléculaires
Centre de Recherches de Jouy-en-Josas. France

INTRODUCTION

Transmissible gastroenteritis virus (TGEV) is an enteropathogenic coronavirus of swine which induces an acute diarrhea syndrome especially severe in newborn animals less than two weeks of age (15). The organisation of the TGEV genome has been established as well as the sequences of the genes encoding the structural and non structural proteins, except the polymerase (7, 9, 13, 14). TGEV virions are made of three proteins, a nucleoprotein (N, 47K) and 2 envelope glycoproteins : M (29K), an integral membrane protein, and S (220K), which forms the surface projections. The peplimer protein S is a 1431 amino acid long, highly glycosylated polypeptide, with a membrane anchoring domain near its carboxy terminus (14). This protein is responsible for the induction of neutralising antibodies (4, 8) and is presumably involved in the recognition of target cells. A minimum of four major antigenic sites have been delineated using monoclonal antibodies (1, 3, 6).

The correlation of these sites with the primary structure of the molecule might contribute to the identification of functional domains of importance with regard to the immunogenicity or pathogenicity of the virus. A combination of several approaches has enabled us to localise the four main antigenic sites A, B, C and D defined previously using our hybridoma library (3). The present paper briefly reports the main findings of this study.

MATERIALS AND METHODS

Viruses : The high passage Purdue-115 strain was used as a virus source. The propagation of the virus and the selection of neutralisation escape mutant have been described (3).

Monoclonal antibodies : The characteristics of the 23 anti TGEV-S MAbs used in this study have been reported (8).

Characterisation of immunoreactive proteolytic fragments : (^{35}S) cysteine-labelled cytosols were prepared as described (8), with some modifications (multiplicity of infection 50 PFU, 8 hr labelling at 3 h p.i., 100 $\mu\text{Ci/ml}$). Proteolytic digestions were performed 2 hr at 37°C using one of these

endopeptidases : α -chymotrypsin, collagenase (*clostridium perfringens*), trypsin and V8 protease (200 to 600 $\mu\text{g/ml}$). Immunoprecipitations were performed as reported using hybridoma ascites fluids (5 μl for 50 μl of digested material) and protein A-Sepharose (8). Immunoreactive fragments were resolved on a 9-20% polyacrylamide gradient gel then fluorographed.

Amino-terminal sequencing of proteolytic fragments : Radiolabelled fragments were prepared as above using one of the following tritiated precursors : valine, threonine, isoleucine (0.5-1 mCi/ml). Appropriate fragments were excised from the gel, electroeluted, then subjected to microsequencing by Edman degradation in an automated gas-phase apparatus.

Production of bacterial fusion proteins : The pEX expression vectors were used to produce S fragments as a C-terminal extension of Cro- β -galactosidase, under control of the thermoinducible promotor λP_T (16). The S-cDNA clones used were from a previous study (13). All experiments were carried out in the *E. coli* strain MC1061 transformed with the plasmid pCI. DNA constructs were made using standard protocols. DNA fragments and plasmids were purified using the Geneclean kit (NEB). A random sublibrary was produced from one of the obtained DNA inserts by sonication. A second sublibrary was generated following Bal31 nuclease resection ("slow form", IBI) of a linearised plasmid. Screening of epitope expressing transformed bacteria was performed by colony blot procedure (16) using 1:100 diluted MAb ascites fluids and (125I) protein A. The immunoreactivity of fusion proteins from positive colonies was examined after lysis of bacterial cells by lysozyme + Triton X100, SDS-PAGE resolution and Western blotting.

Sequencing procedures : Point mutations of escape mutants were localised by direct RNA sequencing of portions of the S gene. RNA matrixes were obtained by proteinase K-SDS treatment of virions semi-purified through a glycerol cushion, phenol-chloroform extraction and precipitation by LiCl 2M. (20-mer) oligonucleotides complementary to TGEV S gene sequence were used to prime reverse transcription of the RNA in the presence of a mixture of deoxy/dideoxynucleotides (1-2 μg RNA per sequence, AMV reverse transcriptase, (^{35}S)dATP). Determination of the S encoding sequences in the inserts from sublibraries was achieved by supercoiled dideoxy-sequencing. Plasmids prepared according to the alkali-SDS method were precipitated by PEG-NaCl and denatured by NaOH as reported (11). Appropriate primers homologous to pEX sequences were used to prime the elongation reaction.

RESULTS

Localisation of the site C

Seven partially overlapping cDNA restriction fragments covering the entire S gene except 110 nucleotides at the 5' end were subcloned into appropriate pEX vectors. Clones expressing a β -Gal chimaeric protein were then subjected to colony blot immunoscreening towards our panel of anti-S monoclonal antibodies (MAbs). None of the expression products was found to react with any of the MAbs directed against A, B and D sites or unrelated epitopes. In contrast, 3 overlapping expression products were recognised by each of the 3 MAbs defining the site C. One of the positive clones had a 0.7Kb long insert with its 5' end corresponding to the unique Xho I site of the S sequence. In order to localise the site C more accurately, a random pEX library was created using sonicated fragments - about 100 base pair long - of the latter insert. The S nucleotide sequence of 10 of the epitope-expressing clones obtained in this way was determined by direct sequencing using appropriate synthetic primers. As a striking feature, all the deduced S-specific amino acid sequences possessed the same N-terminus, thus delineating the amino-

terminal limit of the site C. Fine mapping of the carboxy-terminal limit was achieved by analysing a third pEX library generated by Bal31 exonuclease resection of the smallest - 89 base pair long - random insert. As a final result, a fusion β -Gal protein having 9 amino acid residues was characterised, which was recognised by all site C MAbs in Western blotting. According to these data, the site C would stretch from residue 363 to 371 of the mature S protein (Fig. 1). Furthermore, 2 point mutations, leading to a non conservative amino acid change at position 367 and 368, were identified by direct RNA sequencing of mutants resistant to neutralisation by site C MAbs 3^b.5 and 10.4, respectively.

Localisation of the sites A and B

Their localisation was accomplished through the following two-step approach : i) fragmentation analysis of native S protein ; ii) microsequencing of the smallest fragment retaining immunoreactivity. S fragments were generated by controlled proteolytic cleavage using four separate endopeptidases (see Materials and Methods) then subjected to immunoprecipitation with anti-S MAbs. In general, the resulting profiles of cleavage products were typical of a given antigenic site. Fragmentation and competition experiments were thus in good agreement ; however, the sites A and B could not be distinguished in these experiments, which can be related to the fact that these sites share a common epitope, 48.1 (3).

Collagenase digestion led to the characterisation of a 26K fragment (CO-26K), which retained a strong reactivity towards all site A-B MAbs. Its mobility was identical in reductive and non reductive conditions, thus establishing that the antibody binding site was on the fragment itself. Then 3H-threonine or 3H-isoleucine labelled CO-26K precipitated by MAb 48.1 were purified and subjected to N-terminal sequencing. Determination of the position of the radioactive residues allowed us to align unambiguously the N-terminal extremity of CO-26K with the S amino acid sequence (Fig. 1). Its carboxy limit could be predicted on the basis of the fragment size and the presence of a potential collagenase cleavage site. According to the data, the A-B epitopes must be located within a $\sim$ 200 residue long segment starting at serine 506 of mature S.

In order to confirm this finding, we have searched the location of mutations in the S gene of relevant escape mutants. The sequence encoding the totality of the 26K region was determined by direct sequencing of genomic RNA. Mutants resistant to neutralisation by MAb 20.9 (site A) or 48.1 (sites A-B) were analysed. Each mutant had a single nucleotide change resulting in a non conservative amino acid substitution in the 26K region (Fig. 1).

Localisation of the D site

The approach followed for the localisation of this site was similar to that described above. Chymotrypsin digestion experiments have been shown to produce a 13K fragment (CT-13K) which was strongly recognised by all site D MAbs. Microsequencing of labelled CT-13K isolated using MAb 40.1 identified a unique sequence of valine, which allowed the determination of Asn 82 as the amino limit of the fragment. One of the 3 aromatic residues found at position 200-212 is candidate as the C-terminus of CT-13K (Fig. 1). The evidence for the location of site D was strengthened by direct RNA sequencing of MAb 40.1 escape mutants, which identified single amino acid changes at position 145, 147 and 149.

The presence of CT-13K as a minor band was observed in the chymotrypsin profiles obtained with most of the other anti-S MAbs, including site A, B and C MAbs, but not with anti-E1 or N MAbs. Partial amino

1 50
 | |
 DNFPCSKLTNRTIGNQWNLIETFLNYSRLPPNSDVVLGDYFPTVQPWF
 NCIRNNSNDLYVTLENLKALYWDYATENITWNHRQRLNVVVNGYPYSITV
 TTRNFNSAEGAIIICKGSPPTTTTESSLTCNWGSECLNHKFPICPSN Site D
 SEANCGNMLYGLQWFADEVVAYLHGASYRISFENQWSGTVTFGDMRATTL (CT-13K)
 EVAGTLVDLWFWNPVYDVSYYRVNNKNGTTVVSNCCTDQCASYVANVFTTQ
 PGGFIPSDFSFNNWFLLTNSSTLVSGKLVTKQPLLNVCLWPVPSFEEAAS
 TFCFEGAGFDQCGAVLNNTVDVIRFNLNFTTNVQSGKGATVFSLNTTGG
 VTLEISCYTVSDSSFFSYGEIPFGVTDGPRYCYVHYNGTALKYLGLTPPS Site C
 VKEIAISKWGHFYINGYNFFSTFPIDCISFNLTGDSDFWTIAYTSYTE
 ALVQVENTAITKVTYCNSHVNNIKCSQITANLNGFYVPSSEVGLVNKS
 VVLLPSFYTHITIVNITIGLGMKRSGYGQPIASTLSNITLPMQDHNTDVYC
 IRSDQFSVYVHSTCKSALWDNIFKRNCTDVLDATAVIKTGTCPFSEFDKLN Site A-B
 NYLTFNKFCLSLSPVGANCKFDVAARTRTNEQVVRSLYVIIEEGDNIVGV (Co-26K)
 PSDNSGVHDLSQLHLDSCDYNIYGRTVGVIIRQTNRTLISGLYYTSLSG
 DLLGFKNVSDGVIYSVTPCDVSAQAVIDGTIVGAITSINSELLGLTHWT
 TTPNFYYSIYNYTNDRTG.....
 751 1431

Fig. 1. The amino acid sequence of the amino half of TGEV S is shown (data from 14 ; potential N-glycosylation sites overlined). The underlines and box indicate the proposed location of the antigenic sites. Radioactive residues detected by microsequencing: *. Point mutations identified in neutralisation escape mutants: ●.

sequencing of CT-13K fragment prepared using MAb 20.9 (site A) also revealed Asn 82 as its N-terminus.

DISCUSSION

In this study, several approaches have been developed in order to correlate major antigenic sites of TGEV peplomer protein S with its primary structure. By combining proteolytic digestion of the native protein and microsequencing of immunoreactive cleavage products we were able to localise the sites A, B and D, whereas bacterial expression of large fragments of S failed to produce fusion proteins expressing these sites. A fourth antigenic site, C, could be mapped accurately through prokaryotic expression of short S fragments. The proposed location of these sites was confirmed by identification of point mutations in the S gene of relevant escape mutants.

Major antigenic sites have been mapped in the carboxy-terminal subunit of both MHV and IBV peplomer proteins (for references see H. Laude, this volume). In the case of TGEV, however, the four main antigenic sites delineated using our MAb library all cluster in the amino-terminal half of S, which is assumed to correspond to the globular part of the peplomer (Fig. 2). Such a picture is closer of that reported for several virus glycoproteins, like influenza virus (17).

The site C appears to be composed of potentially linear epitopes, lying in a short sequence of 9 contiguous amino acid residues (363-371). This site is defined by moderately or non neutralising MAbs, and exhibits a slight antigenic variation (3, 8). Thus, the British isolate FS 772/70 is not recognised by neutralising MAbs 3^b.5 and 10.4. Comparison of its sequence (P. Britton, personal communication) with that of Purdue strain actually revealed amino acid divergence at position 366, within the antibody binding site. Besides, two adjacent amino acid changes (365-366) have been noted in the homologous sequence of the feline coronavirus FIPV (strain 79-1146 ; 2), which does not react with any of our site C MAbs. These findings, together with those derived from the analysis of selected epitope mutants (see results) indicate that the residues Phe-Ser-Tyr 368 are critical for neutralisation by MAb 3^b.5 and 10.4, and Phe 365 for binding by MAb 11.20.

Fig. 2. Schematic view of the location of four major antigenic sites on the TGEV S polypeptide chain. The approximate position of residues critical for antibody binding is marked by an open circle. The minimal distance between A-B and D sites is 290 amino acids.

In contrast to site C, sites A, B and D have been shown to be highly conserved among TGEV strains. The spatially related sites A and B constitute the major neutralisation-mediating domain of S, and are immunodominant in the natural host species. Microsequencing of a collagenase 26K fragment specifically recognised by site A and B MAbs led us to conclude that this domain is contained within a ~ 200 residue long region with residue 506 as its amino-terminal limit. The positions of amino acid substitutions conferring neutralisation resistance were consistent with this location. The site D is not a major neutralisation site as only one (40.1) of the delineating MAbs is highly neutralising. Substantial evidence for the location of this site was obtained by alignment with the S sequence of a chymotrypsin 13K fragment, which was strongly - but not specifically - recognised by the relevant MAbs, and by identification of point mutations present on MAb 40.1 escape mutant.

Since there is no overlap between the 13K region (site D), the 26K region (A-B domain), and the 9 amino acid sequence (site C), the observed weak reactivity of non site D MAbs with the 13K cleavage product is an intriguing feature. The absence of binding when using anti-E1 or anti N MAbs indicated that this effect is S-specific. One explanation would be that the antibody binding sites for A, B and C MAbs also involve sequences of the 13K region. If so, all these epitopes would be of the assembled type. Alternatively, a non-covalent interaction of some stability might exist between the 13K region and another part of the S molecule ; as a consequence, the 13K fragment would be coprecipitated via larger S fragments.

The localisation of several major antigenic sites of TGEV S provides a molecular basis for the understanding of the relationships between TGEV and closely related viruses. We have shown in a separate study (10) that all site A and B MAbs do react with both FIPV and PRCV (a newly recognised respiratory variant of TGEV : 12), whereas site D MAbs exhibit no cross-reactivity. An optimal alignment of the predicted amino acid sequence corresponding to 26K and 13K regions of TGEV and FIPV revealed a level of homology of 92 and 25%, respectively. The clustering of most of the critical neutralisation epitopes within the highly homologous and immunodominant 26K region thus explains the high level of cross-neutralisation observed between the members of coronavirus group I, also including canine CCV (5). Such a conservation of the major neutralisation mediating domain raises the question of its biological function. On the other hand, it is tempting to speculate that determinants essential to the expression of the tropism or host specificity of this group of coronaviruses resides within the non conserved N terminal region of S, where the site D has been mapped.

REFERENCES

1. I. Correa, G. Jimenez, C. Sune, M.J. Bullido & L. Enjuanes (1988). Antigenic structure of the E2 glycoprotein from transmissible gastroenteritis coronavirus. *Virus Res.* 10 : 77-94 .
2. R.J. De Groot, J. Maduro, J.A. Lenstra, M.C. Horzinek, B.A.M. Van der Zeijst & W.J.M. Spaan (1987). cDNA cloning and sequence analysis of the gene encoding the peplomer protein of feline infectious peritonitis virus. *J. Gen. Virol.* 68 : 2639-2646.
3. B. Delmas, J. Gelfi & H. Laude (1986). Antigenic structure of transmissible gastroenteritis virus. II. Domains in the peplomer glycoprotein. *J. Gen. Virol.* 67 : 1405-1418 .
4. D.J. Garwes, M.H. Lucas, D.A. Higgins, B.V. Pike & S.F. Cartwright (1978-1979). Antigenicity of structural components from porcine transmissible gastroenteritis virus. *Vet. Microbiol.* 3 : 179-190 .
5. M.C. Horzinek, H. Lutz & N.C. Pedersen (1982). Antigenic relationships among homologous structural polypeptides of porcine, feline and canine coronaviruses. *Infect. Immun.* 37 : 1148-1155 .

6. G. Jimenez, I. Correa, M.P. Melgosa, M.J. Bullido & L. Enjuanes (1986). Critical epitopes in transmissible gastroenteritis virus neutralization. *J. Virol.* 60 : 131-139 .
7. P.A. Kapke & D.A. Brian (1986). Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. *Virology* 151 : 41-49 .
8. H. Laude, J.M. Chapsal, J. Gelfi, S. Labiau & J. Grosclaude (1986). Antigenic structure of transmissible gastroenteritis virus. I. Properties of monoclonal antibodies directed against virion proteins. *J. Gen. Virol.* 67 : 119-130 .
9. H. Laude, D. Rasschaert & J.C. Huet (1987). Sequence and N-terminal processing of the transmembrane protein E1 of the coronavirus transmissible gastroenteritis virus. *J. Gen. Virol.* 68 : 1687-1693.
10. H. Laude, J. Gelfi, D. Rasschaert & B. Delmas (1988). Caractérisation antigénique du coronavirus respiratoire porcin à l'aide d'anticorps monoclonaux dirigés contre le virus de la gastro-entérite transmissible. *Journées Rech. Porcine en France* 20 : 89-94.
11. H.M. Lim & J.J. Pène (1988). Optimal conditions for supercoil DNA sequencing with *Escherichia coli* DNA polymerase I large fragment. *Gene_Anal. Techn.* 5 : 32-39.
12. M. Pensaert, P. Callebaut & J. Vergote (1986). Isolation of a porcine respiratory, non-enteric coronavirus related to transmissible gastroenteritis. *Vet. Quart.* 8 (3) : 257-261.
13. D. Rasschaert, J. Gelfi & H. Laude (1987). Enteric coronavirus TGEV : partial sequence of the genomic RNA, its organization and expression. *Biochimie* 69 : 591-600 .
14. D. Rasschaert & H. Laude (1987). The predicted primary structure of the peplomer protein E2 of the porcine coronavirus transmissible gastroenteritis virus. *J. Gen. Virol.* 68, 1883-1890 .
15. L.J. Saif & E.H. Bohl (1986). Transmissible gastroenteritis. In : *Diseases of swine* (sixth edition). pp. 255-274. Edited by A.D. Leman, B. Straw, R.D. Glock, W.L. Mengeling, R.H.C. Penny & E. Scholl. Iowa State University Press .
16. K.K. Stanley & J.P. Luzio (1984). Constructions of a new family of high efficiency bacterial expression vectors : identification of cDNA clones coding for human liver proteins. *EMBO J.* 3 : 1429-1434.
17. D.C. Wiley, I.A. Wilson & J.J. Skehel (1981). Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature* 289 : 373-378 .