SURFACE GLYCOPROTEINS OF TRANSMISSIBLE GASTROENTERITIS VIRUS : FUNCTIONS AND GENE SEQUENCE

Denis Rasschaert, Bernard Delmas, Bernard Charley, Jeanne Grosclaude, Jacqueline Gelfi and Hubert Laude*

Institut National de la Recherche Agronomique
Station de Recherches de Virologie et d'Immunologie
78850 Thiveral-Grignon, France

INTRODUCTION

Transmissible gastroenteritis virus (TGEV) causes acute and highly contagious diarrheal syndrome in pigs, most often fatal for animals aged less than 2 weeks. Despite the intensive research efforts, no safe and efficient vaccine is available against this major porcine pathogen. The tropism of the causative agent to the differenciated enterocytes covering the intestine villi is responsible for the observed enteric disorders. It is assumed that the natural protection conferred by mothers, infected with wild virus, on their suckling piglets is due to the presence in the milk of mainly IgA type neutralizing antibodies induced by the virus peplomers. (see 1 for a review).

The glycoprotein E2 forming the TGEV peplomers differs from those of IBV and MHV by its significantly larger size (200-220K) and by the absence of cleavage into 2 subunits. Two other major structural polypeptides, the nucleoprotein (N,47K) and the transmembrane or matrix protein (E1,29K) have been identified (2, 3, 4). The E2 protein is reasonably the best candidate as the protective antigen. However, the E1 protein could also have a role with regard to the host defense, since its NH2 terminal part is presumably exposed out of the viral membrane.

We provide here informations regarding the primary structures of both these membrane glycoproteins, as deduced from the nucleotide sequencing, and we discuss them in the light of their respective antigenic and biological properties.

MATERIALS AND METHODS

The high passage Purdue-115 strain was used as a virus source throughout the experiments. Derivation of hybridomas, competition assays and selection of neutralization-resistant E2 mutants were done as published (4, 5). Interferon induction from lymphocytes by virus-

* Presented the paper (to whom reprint request should be sent).

M. M. C. Lai et al. (eds.), *Coronaviruses*
© Plenum Press, New York 1987

infected, glutaraldehyde treated cells and antiviral activity measurement were performed as described previously (6, 7). El epitope mutants were obtained by survivor selection experiments ; newborn rabbit serum (1:100) was used as a source of complement in El-Mabs mediated neutralization tests.

cDNA cloning of TGEV genomic RNA was accomplished by a method derived from previously reported procedures (8, 9) ; in essence, cDNA-RNA hybrids have been directly inserted in PSTl-cut PBR322 plasmid, after C:G tailing. DNA sequence was done by the shotgun Sanger's method in M13 phage. All details will be communicated in separate publications.

MAPPING THE E2 AND El GENES ON THE TGEV GENOME

The Fig. 1 summarizes the present state of our knowledge concerning the organization of the TGEV genome, 8.5 kilobases of which has been sequenced from the 3'end. A consensus sequence CTAAAC was found upstream from each transcription unit. This sequence is identical to that of MHV and is likely to correspond to the binding site of the leader-polymerase complex (10 and ref. herewith). We have detected only 5 subgenomic RNAs in infected cells, in full agreement with the findings of two other groups (11, 12). The size distribution of this set of mRNAs correlated well with the position and lenght of every potential gene. The large open reading frame (ORF) of 4341 bases yielding a protein exhibiting properties of E2 protein and another 867 bases long ORF predicting for a coronavirus matrix-like protein were identified at 8.22 kb and 2.48 kb of the genomes 3'end, respectively. Nothern blot analysis established that E2 ORF corresponded to the mRNA species 2 and El ORF to the mRNA species 5.

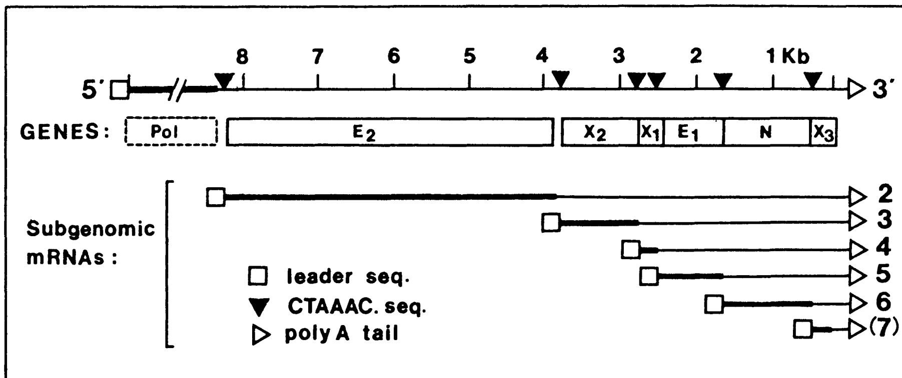

Fig. 1 Genomic organization and gene expression of TGEV. Genes are indicated with boxes. The corresponding subgenomic mRNAs species are represented in the bottom half of the fig. (Existence of a 5' leader seq. has not been formally proved with TGEV ; whether the 3' proximal potential gene X3 is expressed or not is still conjectural).

STRUCTURAL FEATURES OF E2 PROTEIN

The maturation of IBV peplomer protein proceeds through the cleavage of a leader peptide (13). Similarly, the 5' end of the TGEV E2 ORF shows the presence of an hydrophobic aminoacid sequence which fulfills the criteria of an eukaryotic signal sequence (14). Although yet to be confirmed by partial sequencing of the mature product, this indicates that the translation of E2 starts at the first Met codon, which is located 25 bases downstream from the consensus sequence. Accordingly, the E2 precursor molecule would contain 1447 residues (127 acidic, 93 basic, 520 hydrophobic), with m.w. 160 K. If most of the 32 potential N-glycosylation sites found in the sequence are functional, the m.w. of the mature product would closely approach the 220 K value determined by gel electrophoresis (4).

Several interesting features arise while comparing with the IBV S protein, the only coronavirus peplomer sequence yet reported in the litterature (13) : 1) The E2 protein is quite larger in size (285 more residues). 2) A dot matrix analysis reveals that the amino half parts of the two sequences are higly divergent. In fact, all the homology regions are clustered in the carboxy half part. This part of the TGEV E2 sequence is displayed in Fig. 2. 3) A typical structure, assumed to be responsible for anchoring of the peplomer into the virus enveloppe, is present at the carboxyterminus. It comprises of an hydrophobic segment of 45 aminoacids highly rich in cysteine, followed by a highly charged segment of 16 aminoacids, which supposedly protudes inside the virion. 4) A sequence DRTRG is present in the E2 protein at approximately the same position where the S1/S2 cleavage site has been located in the IBV S protein. This basic sequence remind the connecting peptide found in cleavable virus spike proteins, including that of IBV (15).

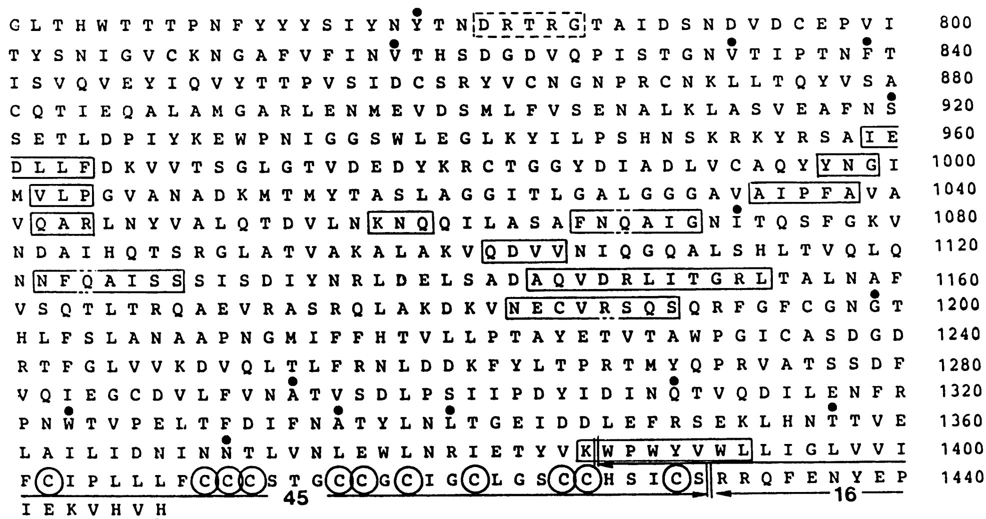

Fig. 2 Aminoacid sequence of the carboxy half part of the TGEV peplomer protein E2, showing the anchor region (underlined) and the regions of stringent homology with the IBV S protein (boxed). N-glycosylation sites : (●)

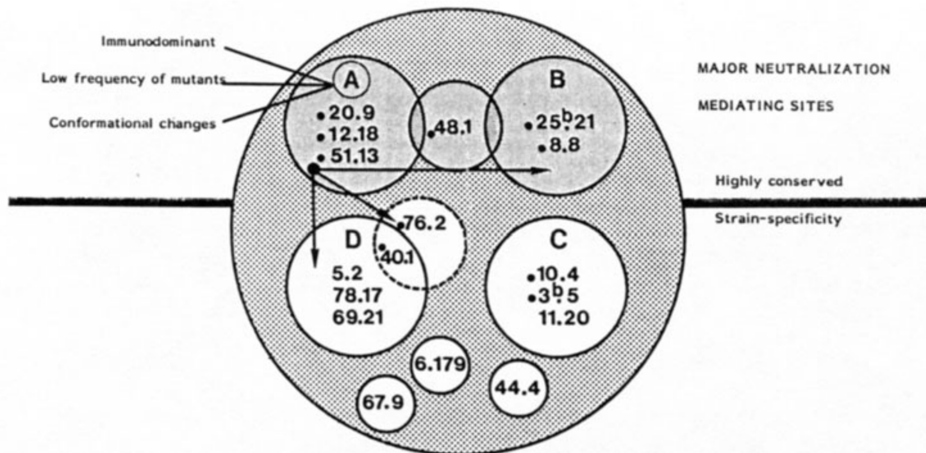

Fig. 3 Model for the antigenic and functional domains of TGEV peplomer. Neutralizing Mabs : (●) Arrows indicate the conformational changes induced on other sites by fixation of Mabs on site A.

DOMAINS IN THE E2 PROTEIN

A schematic map of the main antigenic and functional sites identified on the TGEV E2 protein is presented in Fig. 3. Four major antigenic sites, viz. A,B,C and D, were delineated by means of competition assays using monoclonal antibodies (Mabs). Part of the epitope linkages thus evidenced were confirmed by an analysis of neutralization co-resistance within a pannel of escape mutants (5).

The following facts were observed : 1) The sites A and B contained epitopes exclusively recognized by strongly neutralizing Mabs. 2) The few epitopes found to be related to strain specifity mapped out of the A-B domain. 3) The site A, in particular, might play essential biological functions (see 5 for details).

An analysis of the antigenic reactivity of fragments generated by proteolytic or chemical cleavage of E2 has been undertaken. Preliminary results tend to strengthen the model proposed above. For instance, collagenase digestion of E2 produced a 30 K polypeptide which was immunoprecipitated by Mabs mapping in sites A and B, but not in site D. Besides, this approach allowed us to caracterize a 14 K segment wich was still reactive towards neutralizing Mabs.

PRIMARY STRUCTURE OF THE TGEV MATRIX PROTEIN E1

Fig. 4 shows the aminoacid data of part of the E1 ORF which extends downstream from the intergenic sequence CTAAAC, defining the start of the mRNA 5 transcript. Two Met residues (circled) are available near the aminoterminus. The first one is immediatly followed by a sequence displaying several carasteristics of a signal peptide (14). This is in contrast with that reported for MHV and IBV matrix proteins, both of which lack signal like sequence.

```
                          10                            20                            30  ↓           ↓      40
(M) K [I L L I L A C V I A C A C G] E R Y C A (M) K S D T D L S C R N S T A S D C E S

                          50                            60                            70                     80
C F N G G D L I W H L A N W N F S W S I I L I V F I T V L Q Y G R P Q F S W F V

                          90                           100                           110                    120
Y G I K M L I M W L L W P V V L A L T I F N A Y S E Y Q V S R Y V M F G F S I A

                         130                           140                           150                    160
G A I V T F V L W I M Y F V R S I Q L Y R R T N [S W W S F N P E] T K A I L C V S

                         170                           180                           190                    200
A L G R S Y V L P L E G V P T G V T L T L L S G N L Y A E G F K I A D G M N I D

                         210                           220                           230                    240
N L P K Y V M V A L P S R T I V Y T L V G K K L K A S S A T G W A Y Y V K S K A

                         250                           260
G D Y S T E A R T D N L S E Q E K L L H M V
```

Fig. 4 Aminoacid data of the genomic ORF encoding the TGEV E1 matrix protein. The two possible start sites for translation are circled. The two arrows indicate the relative position of the first residue of MHV and IBV respectively, as obtained by anchoring to the common 8 aminoacids sequon (around 150) then aligning the entire sequences.

Alternatively, translation of TGEV E1 may initiate at the next Met codon (position 22). In the latter case, the product would be 241 aminoacids long, i.e. slightly larger in size as compared to MHV (228) and IBV (225) (16,17). None of these two possibilities is favored by considering the context of each potential initiator codon. Additionnal experiments are going on in order to assess wether the insertion of TGEV E1 into the membrane leads or not to an N-terminal processing.

In any case, the m.w. of the mature product is envisaged between 27.8 k and 26.5 k. The carbohydrate content probably account for the difference with the 29 K value found for the E1 major species in infected cells. Two potential N-glycosylation sites are present in the aminoterminal region, but the accessibility of the second one is uncertain. As illustrated by the hydrophilicity profile shown in Fig. 5, the topology of E1 should essentially be similar to that proposed for MHV and IBV (18,19). The existence of three membrane spanning fragments is strongly suggested. Furthermore, the 8 aminoacids strech SWWSFNPE, perfectly conserved in the MHV and IBV E1 proteins, is also present in the TGEV E1 sequence.

POSSIBLE ROLE OF THE E1 PROTEIN IN INTERFERON INDUCTION

Our interest in E1 has been reinfored by the following intriging observations : peripheral blood lymphocytes of non immune pigs produce interferon (IFN) when cultured with fixed, TGEV-infected cells. The antiviral activity was of the α-type and averaged 10^3 i.u./ml by 8 to 16 hrs of incubation. On examining the effect of anti-TGEV Mabs in that system, it was found that two anti-E1 Mabs (25-22 and 49-22) were able to block the IFN induction, whereas other anti-E1 or anti-E2 Mabs relevant to sites A, B, C or D did not affect the synthesis.

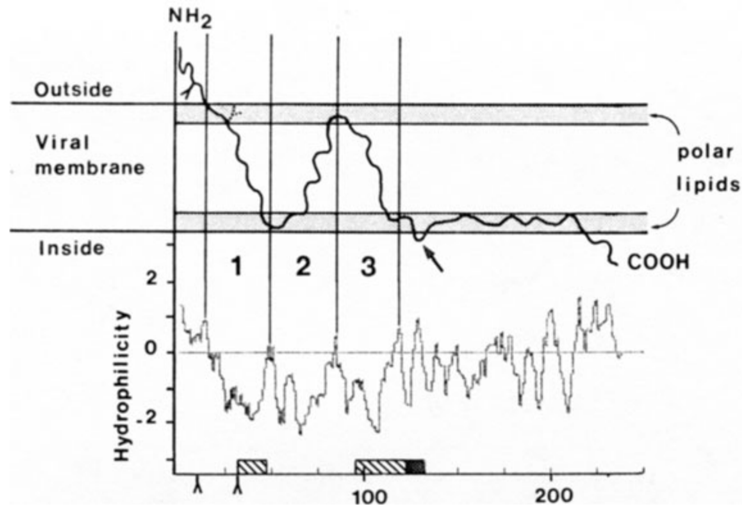

Fig. 5 Lower part : Predicted hydrophilicity profile (Hopp & Woods) of the TGEV E1 protein, starting from the second Met residue. Upper part : Possible membrane topology of E1 according to the model proposed by P. Rottier et al. (1986).

From these results it was tentatively concluded, that : 1) the protein E1 was able to commit the lymphocytes to synthetize IFN. 2) only specific epitopes were involved in the phenomenon. A possible approach to test this assumption was to isolate E1 epitope mutants. As illustrated in the schematic antigenic map of E1 (Fig. 6), the two IFN blocking Mabs were also able to strongly neutralize virus infectivity in the presence of complement, hence providing us the opportunity to select escape mutants. The IFN induction activity of the latters is currently being examined in the presence or absence of blocking Mabs.

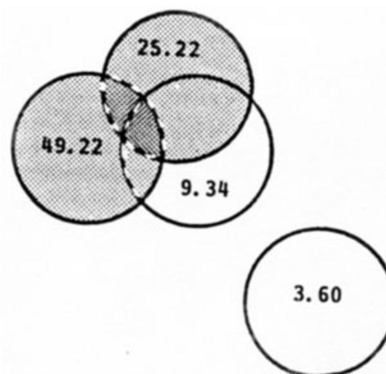

Fig. 6 Schematic antigenic map of TGEV E1 protein. The shadowed epitopes are those which i) block the IFN-induction, ii) mediate the complement dependant neutralization of virus infectivity. Broken line indicates an uncomplete overlapping of the epitopes as judged by competition assays.

CONCLUSIONS

The cloning and sequencing of the TGEV genes which encode the E2 and El proteins represent an important step towards a better understanding of the structure and function of these surface glycoproteins.

It is now well established that the peplomers of the coronaviruses bear essential antigenic and virulence determinants. A major goal for the coronavirologists is to orient the epitope maps available for TGEV and MHV E2 proteins on their primary structure deduced from the nucleotide sequence. Concerning TGEV, we have found that the major neutralization-mediating and immunodominant sites apparently cluster in a discrete area of the molecule, which is higly conserved among the TGEV strains. This domain may, however, involve structural parts distantly located on the primary E2 sequence. Direct RNA sequencing of the genome of the available escape mutants, expression in E-coli of large known or short random cDNA fragments, and fragmentation experiments of the protein are the approaches currently being put to an use to adress this question.

On the other hand, studies with TGEV have thrown the light on a previously unrecognized property of El, i.e. the capacity to trigger a relevant IFN synthesis in non-immune lymphocytes. Since structural features indicate that El is mainly buried in the membrane, the question arises wether the exposed N-terminal stretch (the exact residues number of which remains to be determined) mediates this activity, and if so, by which mechanism. If confirmed, these results would partly elucidate at the molecular level the origin of the marked IFN-α response observed in TGEV infected piglets (7).

REFERENCES

1. D.J. Garwes, Coronavirus in animals. In : "Virus Infections of the Gastrointestinal Tract", pp 315-359. D.A.J. Tyrrel & A.Z. Kapikian eds. Marcell Dekker, New York - Basel. (1982).
2. D.J. Garwes, D.H. Pocock, The polypeptide structure of transmissible gastroenteritis virus. J. Gen. Virol. 29 : 25-34. (1975).
3. M. Horzinek, H. Lutz, N.C. Pedersen, Antigenic relationships among homologous structural polypeptides of porcine, feline and canine coronaviruses. Inf. Immun. 37 : 1148-1155. (1982).
4. H. Laude, J.M. Chapsal, J. Gelfi, S. Labiau, J. Grosclaude, Antigenic structure of transmissible gastroenteritis virus. I. Properties of monoclonal antibodies directed against virion proteins. J. Gen. Virol. 67 : 119-130. (1986).
5. B. Delmas, G. Gelfi, H. Laude, Antigenic structure of transmissible gastroenteritis virus. II. Domains in the peplomer glycoprotein. J. Gen. Virol. 67 : 1405-1418. (1986).
6. P. Lebon, M.J. Commoy-Chevalier, B. Robert-Galliot, C. Chany, Different mechanisms for alpha and beta interferon induction. Virology 119 : 504-507. (1982).
7. C. La Bonnardière, H. Laude, High interferon titer in newborn pig intestine during experimentally induced viral enteritis. Inf. Immun. 32 : 28-31 (1981).
8. S. Zain, J. Sambrook, R.J. Roberts, W. Keller, M. Fried, A.R. Dunn, Nucleotide sequence analysis of the leader segments in a cloned copy of adenovirus 2 fiber mRNA. Cell. 16 : 851-861. (1979).

9. S. Van der Werf, F. Bregegere, H. Kopecka, N. Kitamura, P.G. Rothberg, P. Kourilsky, E. Wimmer, M. Girard, P.N.A.S. 78 : 5983-5987. (1981).
10. C.J. Budzilowicz, S.P. Wilczynski, S.R. Weiss, Three intergenic regions of coronavirus mouse hepatitis virus strain A 59 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. (1985).
11. S. Hu, J. Bruszewski, T. Boone, L. Souza, Cloning and expressing of the surface glycoprotein gp 195 of porcine transmissible gastroenteritis virus. In : "Modern approaches to Vaccines", pp219-223. R.M. Chanock & L.A. Lerner eds. Cold. Spr. Harb. Lab. (1984).
12. L. Jacobs, B.A.M. Van der Zeijst, M.C. Horzinek, Characterization and translation of transmissible gastroenteritis virus mRNAs. J. Virol. 57 : 1010-1015. (1986).
13. M.M. Binns, M.E.G. Boursnell, D. Cavanagh, D.J.C. Pappin, T.D.K. Brown, Cloning and sequencing of the gene encoding the spike protein of the coronavirus IBV. J. Gen. Virol. 66 : 713-726. (1985).
14. D.J. Mc Geogh, On the predictive recognition of signal peptide sequences. Virus Research 3 : 271-286. (1985).
15. D. Cavanagh, P.J. Davis, D.J.C. Pappin, M.M. Binns, M.E.G. Boursnell, T.D.K. Brown, Coronavirus IBV : partial amino terminal sequencing of spike polypeptide S2 identifies the sequence Arg-Arg-Phe-Arg-Arg at the cleavage site of the spike precursor propolypeptide of IBV strains Beaudette and M41. Virus 4 : 133-143. (1986).
16. J. Armstrong, H. Niemann, S. Smeekens, P. Rottier, G. Warren, Sequence and topolology of a model intracellular membrane protein, El glycoprotein, from a coronavirus. Nature, 308 : 751-752. (1984).
17. M.E.G. Boursnell, T.D.K. Brown, M.M. Binns, Sequence of the membrane protein gene from avian coronavirus IBV. Virus Res. 1 : 303-313. (1984).
18. P. Rottier, G.M. Welling, S. Welling-Wester, H.G.M. Niesters, J.A. Lenstra, B.A.M. Van der Zeijst, Predicted membrane topology of the coronavirus protein El. Biochemistry 95 : 1335-1339. (1986).
19. D. Cavanagh, P.J. Davis, D.J.C. Pappin, Coronavirus IBV glycopeptides : locational studies using proteases and saponin, a membrane permeabilizer. Virus Res. 4 : 145-156. (1986).