Molecular biology of transmissible gastroenteritis virus

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

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

ABSTRACT

Laude, H., Rasschaert, D., Delmas, B., Godet, M., Gelfi, J. and Charley, B., 1990. Molecular biology of transmissible gastroenteritis virus. *Vet. Microbiol.*, 23: 147–154.

The causative agent (TGEV) of porcine transmissible gastroenteritis belongs to the Coronaviridae, a family of enveloped viruses with a positive, single-stranded RNA genome. Important progress has recently been made concerning the molecular biology of TGEV. The research work of our group has been focused on two main aspects: genome structure and functional domains of the envelope proteins.

TGEV genomic RNA is organised into seven regions. The sequence of six of them, i.e. the 3′ most 8300 nucleotides, has been established from cDNA clones. Three genes encoding the structural proteins, the peplomer protein E2, the transmembrane protein E1 and the nucleoprotein, have been identified. Additional open reading frames allowed for the prediction of four non-structural polypeptides, the role of which remains to be discovered. The remaining part of the genome (estimated length 20 kb) is thought to encode the polymerase. Expression of TGEV genes involves the production of six subgenomic mRNAs, which together with the virion RNA, form a 3′ terminal nested set.

The peplomer glycoprotein E2 (220 kDa) is 1431 residues long and highly glycosylated. Several domains were identified, including a C-terminal anchoring region and at least four major antigenic sites, which cluster in the amino half part of the molecule. Two sites containing most of the critical neutralisation determinants are highly conserved among TGEV strains. The glycoprotein E1 (29kDa) is mostly embedded in the membrane and plays a crucial role in the virion architecture. However, a short N-terminal domain protruding out of the particle mediates complement-dependent neutralisation, and induces alpha interferon synthesis, likely through a direct interaction with the lymphocyte membrane.

INTRODUCTION

Transmissible gastroenteritis is an acute enteric disease of swine, most often fatal to newborn piglets. Its causative agent (TGEV) belongs to the Coronaviridae, a family of enveloped viruses with a large, single-stranded RNA of

---

*To whom reprint requests should be addressed.

positive polarity as a genome. The virions are constructed from three major structural proteins: a phosphoprotein (N), which encapsidates the genomic RNA, an integral membrane protein (M or E1) and a large glycoprotein (S or E2), which forms the viral projections (review: Spaan et al., 1988).

In spite of its economic importance, TGEV has been one of the least studied coronaviruses, as compared to infectious bronchitis virus (IBV) and murine hepatitis virus (MHV), which have been the objects of sustained research. Up to 1980, only limited information was available on the polypeptide constitution and the genome characteristics of TGEV (Garwes et al., 1976; Brian et al., 1980). During the last few years there has been a substantial accumulation of experimental molecular data. The scope of this paper is to review briefly current information about the organisation of the TGEV genome and the functions associated with the structural domains of the envelope glycoproteins.

GENOME STRUCTURE

Partial sequence analysis of cloned virion RNA and characterisation of virus-specific intracellular RNAs have led to a proposed model for TGEV genome organisation and expression which is depicted in Fig. 1. The sequence of the 3'-most 8300 nucleotides of the genomic RNA of Purdue-115 cell-adapted strain has been determined; it encompasses the whole part of the genome which is expressed through subgenomic RNAs (Rasschaert et al., 1987). The remaining part, of the genome (approximately 20 kilobases long), is thought to encode the viral replicase/transcriptase, as shown in the case of IBV (Boursnell et al., 1987). The three largest genes sequenced, encoding the structural proteins, are oriented according to the consensus gene order of coronaviruses: 5'-E2-E1-N-3'. Four additional major open reading frames (ORFs), poten-
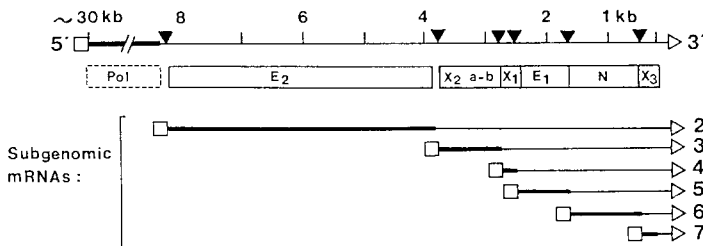

Fig. 1. The organisation and expression of TGEV genome. The genome is represented by the upper horizontal line, which is graduated in kilobases from 3' to 5'. Open reading frames are shown below as open rectangles. The corresponding subgenomic mRNAs species are represented in the bottom half of the figure. Symbols: ▼, conserved intergenic sequence, ▷, poly (A) tract; □, leader sequence.

tially coding for non-structural polypeptides, can be deduced from available sequence data. The number and distribution of the latter differ from that of both MHV and IBV.

Overall, the identified coding regions of TGEV genome are densely packed, yet there are almost no overlaps. Most of the intergenic region consists of 0-15 bases. The largest non-coding region, located upstream from the 3'-most gene (N), is about 280 nucleotides long exclusive of the polyA stretch (Kapke and Brian, 1986; Rasschaert et al., 1987).

GENE EXPRESSION STRATEGY

Expression of coronavirus genomic information involves the synthesis of a negative strand which in turn serves as a template for the production of one genome-sized and several subgenomic RNA species (review: Spaan et al., 1988). Six subgenomic mRNAs are synthesized in non equimolar amounts in TGEV-infected cells (Jacobs et al., 1986; Rasschaert et al., 1987). They form, together with the full-length genome RNA, a 3' coterminal nested set, a distinctive feature of coronaviridae. The synthesis of these molecules has been demonstrated to occur through a discontinuous transcription process, in which a short leader sequence (~ 70 bases) is joined to the body of each RNA (review: Lai, 1986). The nucleotide composition of the leader sequence remains to be determined in the case of TGEV. A consensus sequence $5'^{AA}_{UU}$CUAAAC3', found in each intergenic region of TGEV, is assumed to act as a transcription-initiation signal for the leader-polymerase complex. The translationally active part of each mRNA corresponds to the segment which is not redundant with smaller species. Except for mRNA 3, all TGEV subgenomic RNAs appear to be functionally monocistronic.

Non-structural polypeptides

Characterisation of the non-structural polypeptides of TGEV is incomplete, and most of the available information comes from the sequence data. The translation of four polypeptides is potentially allowed by mRNA 3,4 and 7 (Rasschaert et al., 1987). Two usable reading frames have been identified on mRNA 3: X2a (71 codons) and X2b (254 codons). In vitro translation of mRNA 3 gave rise to a 24 kDa polypeptide (Jacobs et al., 1986), presumably encoded by X2b. A highly hydrophobic polypeptide of 9 kDa has been predicted from ORF X1 present on mRNA 4 (Rasschaert et al., 1987). None of the above species has been characterised as yet in TGEV-infected cells. However, a 14 kDa polypeptide has been shown to be expressed (Wesley and Woods, 1986), which reacted with an antiserum raised against a synthetic peptide derived from the ORF X3 found on mRNA 7 (Garwes et al., 1988). As for other coronaviruses, the biological role of these non-structural polypeptides remains to be elucidated.

STRUCTURAL FEATURES OF PEPLOMER PROTEIN E2

The glycoprotein E2 forms the typical petal-shaped, 20 nm long peplomers projecting outward from the virion's surface. In addition to its probable role in determining the host cell tropism, E2 is assumed to have a membrane fusion activity which enables the viral nucleoprotein to enter the cytoplasm. Also, E2 bears major B lymphocyte antigenic determinants (Laude et al., 1986; Jimenez et al., 1986) which are likely to play a crucial role in promoting the adaptive immunity. In particular, W2 is the sole structural protein capable of eliciting the production of neutralising antibodies, which are able to confer protection to sucking piglets (Garwes et al., 1979).

E2 is initially synthesized as a molecule of 1447 residues (Rasschaert and Laude, 1987), significantly larger than the IBV and MHV peplomer protein. After cleavage of a short hydrophobic sequence acting as a secretory signal peptide, the polypeptide chain is 1431 residues long, with a predicted Mw of 158 kDa. Mannose-rich carbohydrate side-chains are cotranslationally added to most of the 32 potential Asn-linked glycosylation sites present on the apoprotein, thus giving the intracellular species designated E'2 (Mw: 175 kDa). After an additional glycosylation step within the Golgi apparatus (Delmas and Laude, unpublished results), the mature protein (220 kDa) is incorporated into the virions.

Analysis of the primary sequence of TGEV E2 has led to the identification of several functional domains (Fig. 2). Near the C-terminus, a hydrophobic segment of about 20 residues is assumed to anchor the protein into the lipid membrane. The internal domain, about 35 residues long, is divided into 2 parts: a stretch with a high ratio of cysteine residues (50%), which are putative fatty acylation sites, and a hydrophilic stretch, likely protruding inside the virions. The ectodomain thus forms the essential part of the protein. At 1/4-1/5 from the carboxy-end, an amphipatic alpha helix 8 nm long is pre-
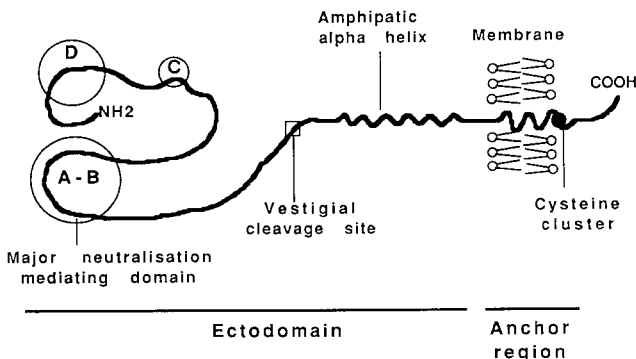

Fig. 2. Schematic drawing of the main functional domains identified on the TGEV peplomer protein E2.

dicted, which is thought to be responsible for the association of 2 or 3 E2 subunits through a coiled-coil interaction, as shown for the influenza HA protein. Such a structure may explain the elongated shape of the peplomer's stem (De Groot et al., 1987). Approximately at the middle of the polypeptide chain a sequence DRTRG might correspond to a vestigial site of cleavage, whereas a functional cleavage sequence gives rise to a heterodimer peplomer protein in IBV and MHV. The amino half of E2 molecule has an overall hydrophobic profile, which reflects the existence of a tightly packed core in the globular part of the peplomer.

ANTIGENIC STRUCTURE OF E2 PROTEIN

Four main antigenic sites, designated A, B, C and D, have been delineated on the TGEV E2 protein by means of competition assay using monoclonal antibodies (MAbs) raised against the Purdue strain (Delmas et al., 1986). Most of the critical neutralisation determinants are clustered in a discrete area comprising the two topologically related sites A and B. Whereas the A, B and D sites are highly conserved among TGEV strains from different origins, the C site exhibits slight variation. Prokaryotic expression of E2 gene fragments allowed the characterisation of the latter as a continuous epitope of 10 residues. Controlled proteolytic digestion of E2 has been shown to generate two reactive fragments of 26 kDa and 14 kDa recognised by site A-B and D MAbs, respectively. Partial amino acid sequencing permitted the localisation of these fragments in two distinct regions of the E2 primary structure (Delmas et al., 1990). Furthermore, amino acid substitutions capable of conferring neutralisation resistance were identified in each region by direct RNA sequencing of the genome of site A or D antigenic mutants (Laude et al., unpublished results). All four antigenic sites appeared to be located in the amino-half part of E2 (Fig. 2).

Analysis of the antigenic profile of a newly recognised porcine respiratory coronavirus, closely related to TGEV (Pensaert et al., 1986), has revealed important epitope changes within the site D, but non in the A-B domain, a situation similar to that shown for the feline FIPV and canine CCV (Laude et al., 1988b), two other members of coronavirus antigenic group I. This observation is consistant with the fact that the site D lies in the last 250 N-terminal residues of E2, i.e. the region which is the most divergent between TGEV and FIPV (Jacobs et al., 1987). Hence, it can be speculated that determinants important in modulating the virus tropism might reside in this domain of E2. Different approaches are currently being used in the laboratory to address this question.

TRANSMEMBRANE PROTEIN E1

Mostly embedded within the lipid envelope, the glycoprotein E1 is as-sumed to play an essential role in virion architecture. Moreover, the restric-tion of E1 to endoplasmic reticulum membranes has been shown to deter-mine the assembly site of the coronavirus particles. Sequence data of the TGEV E1 gene predict a primary translation product of 262 residues. In con-trast to both IBV and MHV, the translocation of TGEV E1 into the mem-brane appears to take place through the cleavage of a 17 residue N-terminal signal peptide (Laude et al., 1987). The mature E1 protein, of Mw 29 kDa, carries a unique mannose-rich Asn-linked side-chain; minor species of E1 (up to 36 kDa) have more complex glycans (Delmas and Laude, unpublished results).

Inspection of TGEV E1 amino acid sequence reveals several well-defined structural domains (Laude et al., 1987). At the N-terminus, a highly hydro-philic stretch of about 40 residues containing the glycosylation site, consti-tutes the ectodomain of the protein. The latter is significantly larger and, due to the presence of four cysteine residues, potentially more complex in its sec-ondary structure than the exposed domain in IBV and MHV. It is followed by three hydrophobic, membrane-spanning segments. The remaining half of the polypeptide comprises an amphiphilic region presumably associated with the inner part of the membrane, and an hydrophilic C-terminus. The above findings indicate that the topology of TGEV E1 within the membrane should be essentially similar to that previously proposed for MHV and IBV (Rottier et al., 1986).

E1 PROTEIN AND INTERFERON INDUCTION

Inactivated TGEV particles or glutaraldehyde-fixed infected cells have been shown to trigger an early and strong synthesis of alpha-type interferon (IFN) in non-immune porcine or human peripheral blood lymphocytes. This activ-ity appeared to be blocked by two of four MAbs directed against glycoprotein E1, whereas MAbs specific for the other structural proteins were not effective in blocking the interferon induction (Charley and Laude, 1988). A series of epitope mutants were selected on the basis of their resistance to complement-mediated neutralisation towards the IFN-induction blocking MAbs. All the relevant amino acid substitutions identified by sequencing of mutant E1 genes were localised within the last 22 N-terminal residues. Strikingly, two muta-tions were shown to markedly affect the IFN-inducing capacity (Laude et al., 1988a and unpublished results).

These findings provide strong evidence for the role of E1 as the effector and indicate that the interferogenic domain must be located within the ectodo-main of the protein. It is tempting to postulate that the interaction of E1 with

the lymphocyte membrane is the first of the events leading to the induction of IFN-coding genes. This newly recognised biological function of E1 may explain the high titre of circulating IFN observed in TGEV-infected piglets. It is likely that too little attention has been paid to such an alternate pathway of alpha IFN-induction by non-replicating viral structures.

CONCLUSION

Molecular studies of the highly enteropathogenic virus, TGEV, have produced relevant information. While increasing the knowledge of coronaviruses, these data have revealed some features that are unique to TGEV. Moreover, further development of these studies should have an important impact in several domains. Thus, the availability of cDNA copies of the genes coding for surface antigens provide the opportunity to analyse in more detail the mechanisms of protective immunity in TGEV infection and to design new generation vaccines. Elucidation of the molecular basis of pathogenesis would also be a fascinating area for future research.

REFERENCES

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.
Brian, D.A., Dennis, D.E. and Guy, J.S., 1980. Genome of porcine transmissible gastroenteritis virus. J. Virol., 34: 410–415.
Charley, B. and Laude, H., 1988. Induction of interferon alpha by transmissible gastroenteritis coronavirus: role of transmembrane glycoprotein E1. J. Virol., 62: 8–11.
De Groot, R.J., Luytjes, W., Horzinek, M.C., Van der Zeijst, B.A.M., Spaan, W.J.M. and Lenstra, J.A., 1987. Evidence for a coiled-coil structure in the spike proteins of coronaviruses. J. Mol. Biol., 196: 963–966.
Delmas, B., Gelfi, J. and Laude, H., 1986. Antigenic structure of transmissible gastroenteritis virus. II. Domains in the peplomer glycoprotein. J. Gen. Virol., 67: 1405–1418.
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 glycoproteins. J. Gen. Virol., 71 (in press).
Garwes, D.J., Pocock, D.H. and Pike, B.V., 1976. Isolation of subviral components from transmissible gastroenteritis virus. J. Gen. Virol., 32: 283–294.
Garwes, D.J., Lucas, M.H., Higgins, D.A., Pike, B.V. and Cartwright, S.F., 1979. Antigenicity of structural components from porcine transmissible gastroenteritis virus. Vet. Microbiol., 3: 179–190.
Garwes, D.J., Stewart, F. and Britton, P., 1988. Studies on the 14kD protein of porcine transmissible gastroenteritis virus. 111th meeting of the Society for General Microbiology, Warwick, 11–14 April 1988.
Jacobs, L., Van der Zeijst, B.A.M. and Horzinek, M.C., 1986. Characterization and translation of transmissible gastroenteritis Virus mRNAs. J. Virol., 57: 1010–1015.
Jacobs, L., De Groot, R., 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

Kapke, P.A. and Brian, D.A., 1986. Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. Virology, 151: 41–49.

Lai, M.M.C., 1986. Coronavirus leader-RNA-primed transcription: an alternative mechanism to RNA splicing. Bio Essays, 5: 257–260.

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.

Laude, H., Rasschaert, D. and Huet, J.C., 1987. Sequence and N-terminal processing of the transmembrane protein E1 of the coronavirus transmissible gastroenteritis virus. J. Gen. Virol., 68: 1687–1693.

Laude, H., Gelfi, J., Delmas, B., Rasschaert, D. and Charley, B., 1988a. Interferon inducing domain of a coronavirus glycoprotein: TGEV transmembrane protein E1. 111th Meeting of the Society for General Microbiology, Warwick, 11–14 April 1988.

Laude, H., Gelfi, J., Rasschaert, D. and Delmas, B., 1988b. Caractérisation antigénique du coronavirus respiratoire porcin à l'aide d'anticorps monoclonaux dirigés contre le virus de la Gastroentérite Transmissible. Journ. Rech. Porcine en France, 20: 89–94.

Pensaert, M., Callebaut, P. and Vergote, J., 1986. Isolation of a porcine respiratory, non enteric coronavirus related to transmissible gastroenteritis. The Vet. Quart., 8: 257–261.

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

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

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

Spaan, W., Cavanagh, D. and Horzinek, M.C., 1988. Coronaviruses: structure and genome expression. J. Gen. Virol., 69: 2939–2952.

Wesley, R.D. and Woods, R.D., 1986. Identification of a 17000 molecular weight antigenic polypeptide in transmissible gastroenteritis virus-infected cells. J. Gen. Virol., 67: 1419–1425.