NUCLEOTIDE SEQUENCE OF THE E2-PEPLOMER PROTEIN GENE AND PARTIAL NUCLEOTIDE SEQUENCE OF THE UPSTREAM POLYMERASE GENE OF TRANSMISSIBLE GASTROENTERITIS VIRUS (MILLER STRAIN)

Ronald D. Wesley

United States Department of Agriculture
Agricultural Research Service
National Animal Disease Center, Ames, Iowa, USA

ABSTRACT

The E2-peplomer protein gene of the virulent Miller strain of transmissible gastroenteritis virus (TGEV) was sequenced from cDNA clones and compared to the E2 gene sequence of the avirulent Purdue strain. Sequence comparisons indicate that most amino acid differences occur in the N-terminal half of the E2-peplomer which represents the most exposed region of the protein. In addition, analysis of an incompletely sequenced open reading frame (ORF) to the immediate 5'side of the E2 gene indicates extensive sequence homology with the infectious bronchitis virus (IBV) F2 gene which is thought to encode a RNA polymerase.

INTRODUCTION

Transmissible gastroenteritis virus (TGEV) is an economically important pathogenic coronavirus of swine. It is an enveloped RNA virus composed of 3 major structural proteins (E2, E1, and N), and a 23.6 kilobase (kb) positive-stranded genome. At least 5 classes of neutralizing epitopes occur on the E2-peplomer glycoprotein while complement-dependent neutralization can be demonstrated only with E1 glycoprotein-specific monoclonal antibodies.[1-4] Nucleotide sequencing studies have focused on the avirulent Purdue 115 strain in which 8.3 kb at the 3' end of the genome have been sequenced.[5-7] Base sequence comparisons of the Purdue and Miller strains in the region between the E1 and E2 genes have shown that the Miller TGEV genome contains an additional intergenic recognition sequence and a larger ORF, perhaps, encoding a nonstructural protein.[8] In order to further determine genetic differences between these 2 strains, the E2 gene of the virulent Miller strain was sequenced. In addition, analysis of an incomplete ORF upstream of the E2 gene indicates extensive sequence homology with a presumed IBV polymerase gene.[9]

MATERIALS AND METHODS

Virus

A working stock of the virulent Miller strain of TGEV was prepared as described previously.[10] For the isolation of genomic RNA, the virus was

Coronaviruses and Their Diseases
Edited by D. Cavanagh and T.D.K. Brown
Plenum Press, New York, 1990

first plaque-picked 3 times on swine testicular (ST) cells before virus purification and subsequent RNA isolation.[8] The plaque-picked virus remained lethal for neonatal piglets.

Cloning and DNA Sequencing

cDNA was prepared from TGEV genomic RNA and cloned into λgt11. First and second strand syntheses were carried out using calf thymus DNA oligodeoxynucleotides as primers and a cDNA synthesis kit (Amersham Corp., Arlington Heights, IL). EcoR1 linkers were added to blunt-ended, double-stranded cDNA. The cDNA was then ligated to EcoR1 cut λgt11 and packaged *in vitro* (Stratagene, La Jolla, CA).[32] Lambda phage containing viral inserts were identified by hybridization to ^{32}P-labeled cDNA prepared from genomic RNA.

To facilitate cDNA sequencing, viral inserts that hybridized to specific mRNAs were subcloned into the EcoR1 site of the multipurpose pBluescript phagemid vector (Stratagene, La Jolla, CA). Stepwise unidirectional deletions were constructed[11] and sequenced by the dideoxy chain-termination method.[12] Programs for computer analysis of the DNA sequence were purchased from DNASTAR (Madison, WI).

RESULTS

Figure 1 shows the location at the 3'end of the TGEV genome of the cDNA clones, pRP1 and pRP3, derived from the virulent Miller strain of TGEV that were used to obtain DNA sequences of the E2-peplomer gene and the 930 bases to the 5'side of the E2 gene. cDNA clone pRP1, 4256 base pairs (bp) in length, includes approximately 3/4 of the E2 gene sequence plus the 930 bases upstream. cDNA clone pRP3, 3232 bp in length, overlaps cDNA pRP1 by 72 bases, and contains the remaining 1/4 of the E2 gene sequence, the genetic region between the E2 and E1 genes and the entire E1 gene sequence except for the last 12 bases at the 3'end. For DNA sequencing of these clones overlapping unidirectional deletions were constructed by the exonuclease III/S1 nuclease method and double-stranded plasmid DNAs were sequenced by the dideoxynucleotide method.

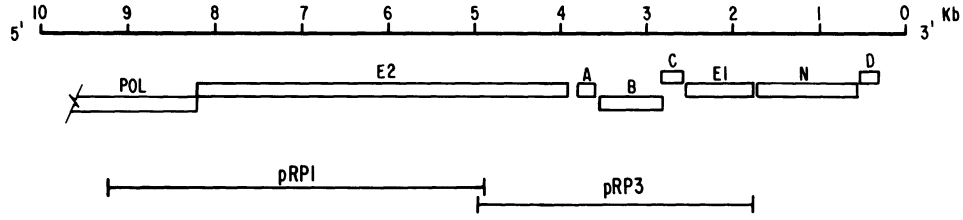

Fig. 1. Schematic diagram of the genomic organization of TGEV. The location at the 3'end of the genome of cDNA clones, pRP1 and pRP3, used for sequencing the gene encoding the E2-peplomer glycoprotein and the upstream POL ORF are shown. The relationship between the major structural protein genes E2, E1, and N and other large ORFs, A, B, C, D that might encode either non-structural or minor structural proteins are indicated. The POL ORF is part of a gene encoding a RNA polymerase.

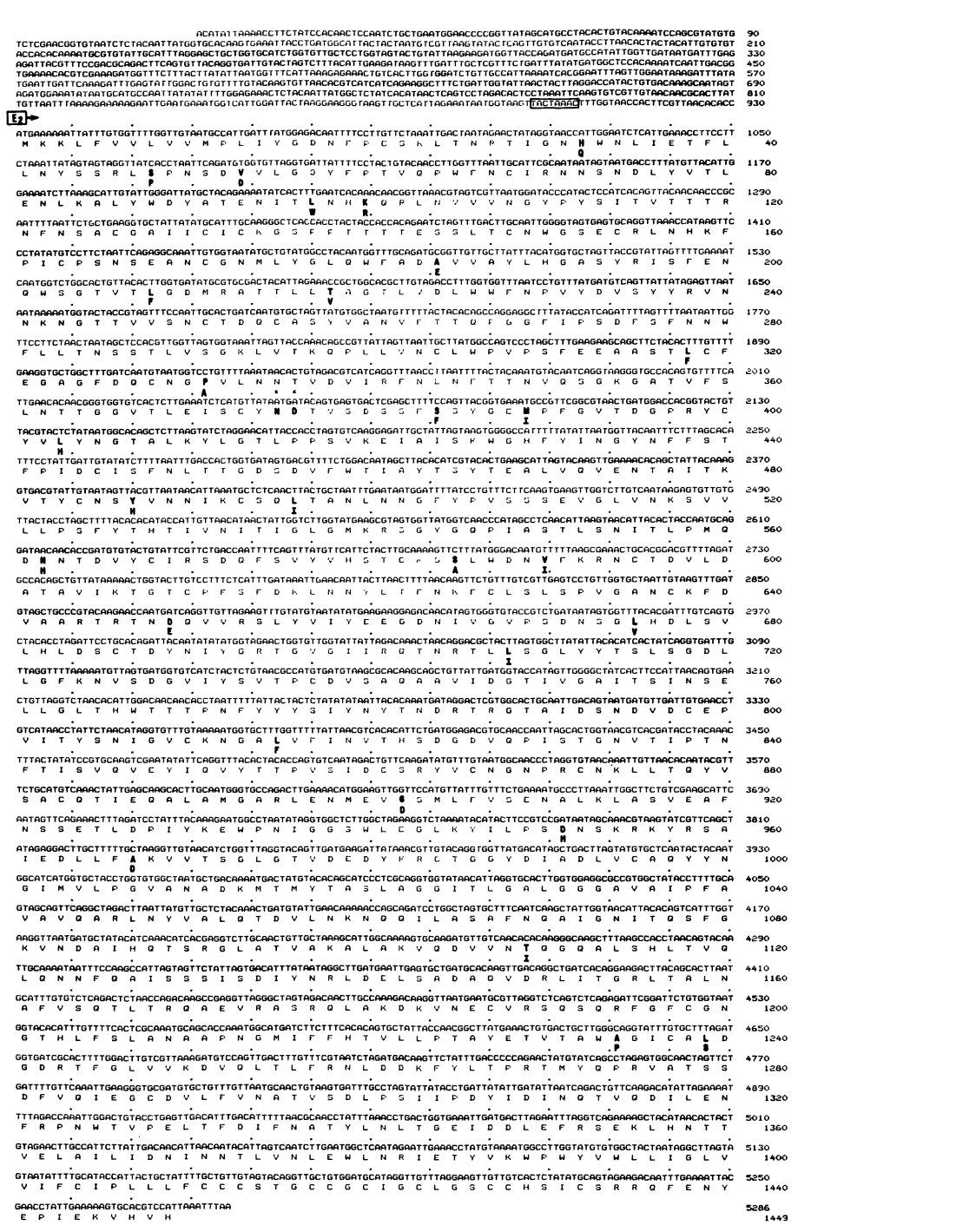

Fig. 2. The nucleotide sequence and predicted amino acid sequence of the E2-peplomer protein of TGEV. In addition, the base sequence (930 nucleotides) to the 5'side of the E2 gene is included. Amino acid differences between the Miller and Purdue strains are shown in bold letters. For the Miller strain, two extra amino acids are added at positions 375 and 376 that establish an additional potential glycosylation site.

The E2 gene sequence and the deduced amino acid sequence for TGEV, Miller strain, are shown in Figure 2 along with the 930 bases to the 5'side of the E2 gene. The nucleotide sequence was determined from both strands of the cDNA clones. The E2 gene primary translation product consists of 1449 amino acid residues, 2 residues longer than the E2 protein of the Purdue strain due to a 6 base insert at positions 2053-2058 (Fig. 2). There is 98% homology between the Miller strain and the Purdue strain E2 proteins at both the nucleotide and the amino acid levels. Between the 2 strains, there are 72 nucleotide differences in the E2 gene that resulted in the 30 amino acid changes shown as bold letters in Figure 2.

A single large ORF, overlapping only the initiation methionine of the E2 gene is found in the 930 bases to the 5'side of the E2 gene. Presumably, this is only a part of a much larger ORF. The amino acid translation product of this ORF is given in Figure 3 and compared with the IBV F2 ORF that is also located immediately 5' of the IBV E2 gene. Extensive sequence homology, 51% identity and 24% conservative amino acid changes, exists between the partial amino acid sequence determined for the TGEV ORF and the C-terminal 312 amino acid residues of the presumed IBV polymerase protein. Thus the base sequences immediately upstream from the E2 gene apparently code for a large RNA dependent-RNA polymerase in both IBV and TGEV.

DISCUSSION

The complete nucleotide sequence and the predicted amino acid sequence of the TGEV E2-peplomer gene (Miller strain) and a partial sequence of a presumed polymerase gene immediately to the 5'side of the E2 gene have been determined. The E2-peplomer gene of the virulent Miller strain shares 98% sequence homology at both the nucleotide and amino acid sequence levels with the E2 gene of the Purdue strain.[7] An additional 6 extra bases, positions 2053 to 2058 in the Miller strain E2 sequence, increase the length of the peplomer protein by 2 amino acids to 1449 residues and gives rise to a new potential glycosylation site. The amino acid differences between these strains do not occur randomly within the peplomer glycoprotein. Instead, over 75% of the amino acid changes occur in the N-terminal half of the peplomer which comprises the exposed club-shaped portion of the peplomer on top of a stalk. Only 7 amino acid differences occur in the C-terminal half of the peplomer which encompasses the stalk structure and the membrane anchoring domain. This feature of N-terminal amino acid variation in the peplomer glycoprotein has also been observed in mouse hepatitis virus (MHV) and IBV.[13,14] Further in the S1 protein of IBV which is the N-terminal cleavage product of the IBV peplomer protein, there are 2 regions of high amino acid variability; one of which has been shown to be associated with virus neutralization.[15,16] Similarly, most of the amino acid changes occur in the N-terminal half of the TGEV peplomer gene, however, no clustering of amino acid variation indicative of highly variable regions were apparent. The 30 amino acid changes in the E2 protein of the Miller and Purdue strains apparently are neutral substitutions since neutralizing monoclonal antibodies, representing 5 different noncompeting sites, were unable to distinguish between these 2 TGEV strains.

Coronaviruses have homologous recognition sequences that are involved in the initiation of transcription and the joining of a RNA leader sequence onto the 5'end of each subgenomic mRNA. The TGEV recognition sequence (A/T A/T)CTAAAC occurs 27 bases upstream from the E2 gene ATG start codon in both the Miller and Purdue strains. In the Purdue strain, a second (A/T A/T)CTAAAC sequence occurs 119 nucleotides into the E2 gene that might also function as an initiation site for a subgenomic mRNA. However, because of a C to T substitution at base 1056 (Fig. 2), the Miller strain E2 gene does not contain this second recognition sequence nor do any further

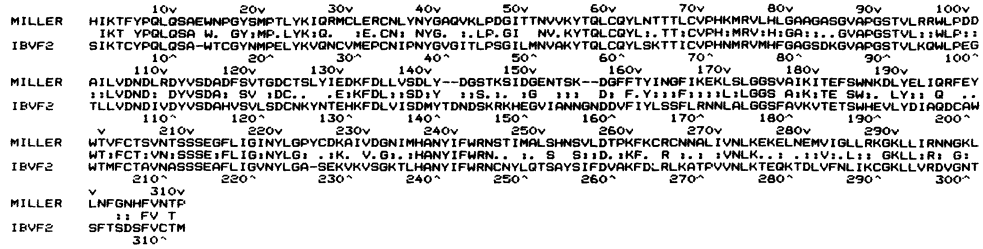

Fig. 3. Comparison between predicted amino acid sequences of IBV F2 and TGEV. The IBV sequence represents the C-terminal 312 amino acid residues of the F2 ORF. The TGEV amino acid sequences is derived from the 930 bases immediately to the 5'side of the E2-peplomer gene. Identical amino acid residues are indicated. A colon indicates a conservative amino acid substitute. A dash indicates a deleted amino acid residue in order to achieve optional alignment.

(A/T A/T)CTAAAC sequences occur in the rest of the E2 gene. Interestingly, the sequence (A/T A/T)CTAAAT occurs 3 times within 150 nucleotides on either side of the E2 gene homologous recognition sequence and does not appear again in the rest of the E2 gene sequence.

MHV encodes a potential nonstructural protein (30-35 Kd) to the 5'side of the E2-peplomer gene that is the translation product of an additional subgenomic mRNA (RNA 2).[17,18] Analyses of TGEV subgenomic RNAs do not reveal a RNA larger than the E2 subgenomic mRNA but smaller than total genomic RNA.[6,8,19,20] In IBV replication, the E2 mRNA is the largest subgenomic mRNA. Large ORFs, upstream of the IBV E2 gene, are thought to encode potential polymerase genes F1 and F2, the latter overlapping the peplomer E2 gene sequence.[9] In this regard, the TGEV genome arrangement resembles more closely IBV than MHV. The evidence that a TGEV polymerase gene occurs to the immediate 5'side of the E2 gene is based on amino acid sequence homology at the C-terminus of the F2 polymerase enzyme of IBV (Fig. 3). Of the 310 amino acids deduced for the TGEV polymerase protein, 51% are identical to the C-terminus of IBV F2 and another 24% represent conservative amino acid changes.

ACKNOWLEDGMENT

The author wishes to thank David Michael for excellent technical assistance.

REFERENCES

1. 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).

2. G. Jimenez, I. Correa, M. P. Melgosa, M. J. Bullido, and L. Enjuanes, Critical epitopes in transmissible gastroenteritis virus neutralization, J. Virol. 60:131-139 (1986).

3. R. D. Woods, R. D. Wesley, and P. A. Kapke, Neutralization of porcine transmissible gastroenteritis virus by complement-dependent monoclonal antibodies, Am. J. Vet. Res. 49:300-304 (1988).

4. R. D. Woods, R. D. Wesley, Unpublished result.

5. P. A. Kapke, and D. A. Brian, Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene, Virol. 151:41-49 (1986).

6. D. Rasschaert, J. Gelfi, and H. Lande, Enteric coronavirus TGEV: partial sequence of the genomic RNA, its organization and expression, Biochimie. 69:591-600 (1987).

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

8. R. D. Wesley, A. K. Cheung, D. D. Michael, and R. D. Woods, Nucleotide sequence of coronavirus TGEV genomic RNA: evidence for 3 mRNA species between the peplomer and matrix protein genes, Virus Res. 13:87-101. (1989).

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

10. R. D. Wesley, R. D. Woods, I. Correa, L. Enjuanes, Lack of protection in vivo with neutralizing monoclonal antibodies to transmissible gastroenteritis virus, Vet. Micro. 18:197-208 (1988).

11. S. Henikoff, Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing, Gene 28:351-359 (1984).

12. F. Sanger, S. Micklen, and A. R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. USA, 74:5463-5467 (1977).

13. W. Luytjes, L. S. Sturman, P. J. Bredenbeek, J. Charite, B. A. M. Van der Feijst, M. C. Horzinek, and W. J. M. Spaan, Primary structure of the glycoprotein E2 of coronavirus MHV-A59 and identification of the trypsin cleavage site, Virology 161:479-487 (1987).

14. J. G. Kusters, H. G. M. Niesters, J. A. Lenstra, M. C. Horzinek, and B. A. M. Van Der Feijst, Phylogeny of antigenic varients of avian coronavirus IBV, Virology 169:217-221 (1989).

15. H. G. M. Niesters, J. A. Lenstra, W. J. M. Spaan, A. J. Zijderveld, N. M. C. Bleumink-Pluym, F. Hong, G. J. M. van Scharrenburg, M. C. Horzinek, and B. A. M. van der Zeijst, The peplomer protein sequence of the M41 strain of coronavirus IBV and its comparison with Beaudett strains, Virus Res. 5:253-263 (1986).

16. D. Cavanagh, P. J. Davis, and A. P. A. Mockett, Amino acids within hypervariable region 1 of avian coronavirus IBV (Massachusetts serotype) spike glycoprotein are associated with neutralization epitopes, Virus Res. 11:141-150 (1988).

17. J. L. Leibowitz, S. R. Weiss, E. Paavola, and C. W. Bond, Cell-free translation of murine coronavirus RNA, J. Virol. 43:905-913 (1982).

18. S. Siddell, Coronavirus JHM: coding assignments of subgenomic mRNAs, J. Gen. Virol. 64:113-125 (1983).

19. S. Hu, J. Braszewski, T. Boone, and L. Souza, Cloning and expression of the surface glycoprotein gp195 of porcine transmissible gastroenteritis virus, in: "Modern Approaches to Vaccines," pp.219-223, Cold Spring Harbor Laboratory, New York (1984).

20. L. Jacobs, B. A. M. van der Zeijst, and M. C. Horzinek, Characterization and translation of transmissible gastroenteritis mRNAs, J. Virol. 57:1010-1015 (1986).