Virus Research, 13 (1989) 87–100
Elsevier

VRR 00499

Nucleotide sequence of coronavirus TGEV genomic RNA: evidence for 3 mRNA species between the peplomer and matrix protein genes

Ronald D. Wesley, Andrew K. Cheung, David D. Michael and Roger D. Woods

USDA, Agricultural Research Service, National Animal Disease Center, P.O. Box 70, Ames, Iowa, U.S.A.

(Accepted 30 January 1989)

Summary

The region of the TGEV genome between the E1-matrix protein gene and the E2-peplomer protein gene has been sequenced from a cDNA clone. The consensus recognition sequence, 5′$^{AA}_{TT}$CTAAAC was found upstream from 3 large open reading frames. In coronaviruses these homologous recognition sequences are involved in the initiation of transcription suggesting that there are 3 mRNA species in this region of the TGEV genome. Northern blot analysis and nuclease S1 mapping confirmed the presence of 3 mRNA species between mRNA 3 encoding the E2-peplomer protein and mRNA 6 encoding the E1-matrix protein. The 5′ regions of these 3 mRNAs encode potential polypeptides of predicted molecular weight; 7859, 27744 and 9287, respectively. The potential translation product of ORF B (27744 Da) is considerably larger than previously reported and could be difficult to distinguish by size from the E1-matrix protein.

Coronavirus; TGEV; RNA sequencing

Introduction

Transmissible gastroenteritis virus (TGEV) is an economically important coronavirus of swine that produces an often fatal diarrhea especially in nursing pigs

Correspondence to: R.D. Wesley, National Animal Disease Center, P.O. Box 70, Ames, IA 50010, U.S.A.

less than 2 weeks of age (Saif and Bohl, 1986). Like other members of the family Coronaviridae, TGE virions consist of 3 major structural proteins; the E1-matrix and E2-peplomer surface glycoproteins and a phosphorylated nucleocapsid protein (N) that associates with the 23.6 kilobase (kb), non-segmented, positive-stranded RNA genome (Garwes and Pocock, 1975; Brian et al., 1984; Hu et al., 1984; Jacobs et al., 1986; Wesley and Woods, 1986). The nucleotide and deduced amino acid sequences of these structural proteins have been determined for the avirulent Purdue 115 strain of TGEV (Kapke and Brian, 1986; Laude et al., 1987; Rasschaert and Laude, 1987).

In the replication of TGEV, as with other Coronaviridae, transcription proceeds via a discontinuous, nested-set mechanism in which several distinct mRNA species of subgenomic size are synthesized (Hu et al., 1984; Jacobs et al., 1986; Rasschaert et al., 1987). These mRNAs share the co-terminal 3′ polyadenylated end of the TGEV genome and extend for different lengths in the 5′ direction. This transcription mechanism has been well documented in 2 other coronaviruses, murine hepatitis virus (MHV) and infectious bronchitis virus (IBV) (Lai et al., 1983; Spaan et al., 1983; Brown et al., 1984). In addition, the 5′ end of each subgenomic mRNA contains a short RNA leader sequence, derived from the 5′ end of the genome, that primes transcription. This leader sequence is 72 bases long in the case of MHV and approximately 60 bases in length for IBV. The leader and the body sequences of each subgenomic mRNA are joined by a discontinuous transcription mechanism. The freely dissociated leader sequence binds to an intergenic recognition sequence and serves as a primer for the transcription of subgenomic mRNAs (Lai, 1986). Thus each subgenomic mRNA contains a homologous recognition sequence approximately 60–70 bases downstream from the actual 5′ end. The core recognition sequence for MHV is 5′AATC$^{T}_{C}$AAAC and for IBV it is 5′CT$^{T}_{G}$AACAA. Furthermore, nucleotide sequences that flank the core homologous sequence are also important in leader RNA-primed transcription (Boursnell et al., 1987).

In general, the translated region of each coronavirus subgenomic mRNA is the 5′-most open reading frame (ORF) that is not present in the smaller subgenomic mRNAs. This is the case for mRNAs that encode the coronavirus structural genes. For TGEV, mRNAs 3, 6 and 7 have been shown by in vitro translation studies to code for the E2-peplomer, E1-matrix and nucleocapsid proteins, respectively (Jacobs et al., 1986). In some instances for subgenomic mRNAs that do not appear to code for any of the major virion structural proteins, internal initiation of translation is thought to occur. These include mRNA D for IBV and mRNA 5 for MHV for which a 12.4 kDa IBV polypeptide and a 10.2 kDa MHV polypeptide have been shown to be translated in infected cells (Skinner et al., 1985; Smith et al., 1987). In the case of TGEV, 2 mRNA species have been identified by Northern hybridization to be present between the E2-peplomer and E1-matrix structural genes (Hu et al., 1984; Jacobs et al., 1986; Rasschaert et al., 1987). The larger of these 2 mRNAs, designated mRNA 4 by Jacobs et al. (1986) and designated mRNA 3 by Rasschaert et al. (1987), coded for 2 non-overlapping ORFs that were separated by a non-coding intervening region of 334 bases containing only a single termination codon 267 bases upstream from the start of the second ORF. The next smaller TGEV mRNA

transcript contained only one unique 5′ coding sequence for a single hydrophobic polypeptide.

In this paper we present evidence for 3 mRNA species between the E1 and E2 structural genes of a virulent strain of TGEV (Miller strain). Nucleotide sequence analysis revealed 3 large ORFs each preceded by an appropriate octanucleotide recognition sequence which could direct transcription initiation.

Materials and Methods

Cells and virus

Swine testicular (ST) cells (McClurkin and Norman, 1966) were grown in modified Eagle's MEM (Gibco) supplemented with fetal bovine serum (10%), sodium bicarbonate (0.22%), lactalbumin hydrolysate (0.25%), sodium pyruvate (0.01%), and gentamicin sulfate (50 μg/ml).

The virulent Miller strain of TGEV was kindly provided by Dr. L. Saif, Ohio Agricultural Research and Development Center, Wooster, Ohio, U.S.A. A working stock of this strain as homogenized intestinal contents of 3-day-old piglets was prepared as described previously (Wesley et al., 1988). Infectious intestinal content was plated directly onto ST cells for the synthesis of unadapted-gut virus intracellular RNAs. For the isolation of genomic RNA, the virus was first plaque-picked 3 times on ST cells before virus purification and subsequent RNA isolation. The plaque-picked virus remained lethal for neonatal piglets.

Purification of genomic RNA

Plaque-purified virus was isolated from clarified supernatant fluids as previously described (Wesley and Woods, 1986). Genomic RNA was extracted by dissolving the purified virus pellet in 0.4 ml TNE (0.02 M Tris-HCl, pH 9.0, 0.1 M NaCl, 0.001 M EDTA). The resuspended pellet was then disrupted in 1% SDS, 625 μg/ml proteinase K and incubated at 37 °C for 5 min. The genomic RNA was extracted once with an equal volume of TNE saturated phenol, once with chloroform-isoamyl alcohol (24 : 1), and concentrated by ethanol precipitation at −20 °C.

cDNA synthesis and cloning

cDNA was prepared from TGEV genomic RNA and cloned into the λgt11 expression vector (Huynh et al., 1985). First and second strand synthesis were carried out using calf thymus DNA oligodesoxynucleotides as primers and a cDNA synthesis kit (Amersham Corp, Arlington Heights, IL). *Eco*R1 linkers were added to blunt-ended, double-stranded cDNA. The cDNA was then ligated to *Eco*R1 cut λgt11 and packaged in vitro (Stratagene, La Jolla, CA). Lambda phage containing viral inserts were identified by hybridization to ^{32}P-labeled cDNA prepared from genomic RNA.

DNA sequencing and sequence analysis

To facilitate cDNA sequencing, viral inserts that hybridized to specific mRNAs were subcloned into the *Eco*R1 site of the multipurpose pBluescript phagemid vector (Stratagene, La Jolla, CA). Stepwise unidirectional deletions were constructed at both ends of the viral insert using the exonuclease III/S1 nuclease method as described by Henikoff (1984). These sequentially deleted plasmids were particularly useful for double-stranded DNA sequencing by the dideoxy chain-termination method (Sanger et al., 1977) because the primer binding site becomes juxtaposed to overlapping regions of new DNA sequences. Programs for computer analysis of the DNA sequence were purchased from DNASTAR (Madison, WI).

Northern blot hybridization

Intracellular polyadenylated RNA from TGEV-infected cells was denatured with glyoxal and dimethylsulfoxide (60 min at 50 °C) as described by Maniatis et al. (1982). The RNA samples, electrophoresed in 1% agarose gels, were blotted onto GeneScreen nylon membranes (New England Nuclear Corp., Boston, MA). UV crosslinking was used to covalently bind the RNA to the membrane filters (Church and Gilbert, 1984). Prehybridization was carried out for 3 h at 65 °C in 6 × SSC (SSC is 0.15 M NaCl, 0.015 M Na citrate, pH 7.0), 5 × Denhardt's solution (0.1% Ficoll, 0.1% polyvinyl pyrrolidone, 0.1% bovine serum albumin), 0.5% SDS, and 100 μg/ml of sonicated denatured salmon sperm DNA (Maniatis et al., 1982). Hybridization was carried out at 65 °C for 18 h in fresh prehybridization solution containing nick-translated [^{32}P]cDNA. A cDNA clone, pFG5 from the extreme 3′ end of the genome, was kindly provided by Dr. P. Kapke, National Animal Disease Center, Ames, Iowa, USA. After incubation, filters were washed with 3 changes of 2 × SSC, 0.1% SDS at room temperature, followed by 2 changes of 1 × SSC, 0.1% SDS at 65 °C for 60 min. Dried filters were exposed to Kodak XAR-2 film at −70 °C with an intensifying screen for 1 to 4 days.

Nuclease S1 analysis

Hybridization of [^{32}P]UTP-labeled, single-stranded RNA probes (1×10^5 cpm) was carried out with 3 μg of unlabeled poly(A) RNA (Maniatis et al., 1982). The hybridization buffer was 40 mM PIPES (pH 6.4), 1 mM EDTA (pH 8.0), 0.4 M NaCl, and 80% formamide. Samples containing the labeled and unlabeled RNAs were heated at 85 °C for 10 min and hybridized overnight at 50 °C. The annealed samples were digested for 30 min at 37 °C with 500 U/ml of S1 nuclease (BRL, Gaithersburg, MD). The protected RNA fragments were resolved on 1.4% agarose gels after chemical and heat denaturation (Cheung, 1988). Agarose gels were dried and exposed to Kodak XAR-2 film at −70 °C with an intensifying screen.

Results

Northern blot analysis

Poly(A) RNA was extracted from TGEV-infected cells at 9 h post-infection. Since TGEV synthesizes a nested-set of mRNA transcripts, at 2.0 kb cDNA probe, pFG5, which is located at the extreme 3′ end of the TGEV genome (Kapke and Brian, 1986), was used to demonstrate TGEV specific mRNAs. Fig. 1 shows a Northern blot of intracellular RNAs extracted from cells infected with either unadapted or plaque-purified TGEV. Each virus produced at least 6 discrete mRNA bands with calculated sizes in kb of 8.2 (RNA 3), 4.0 (RNA 4a), 3.7 (RNA 4b), 3.1 (RNA 5), 2.8 (RNA 6), 2.0 (RNA 7). The RNAs are numbered according to Jacobs

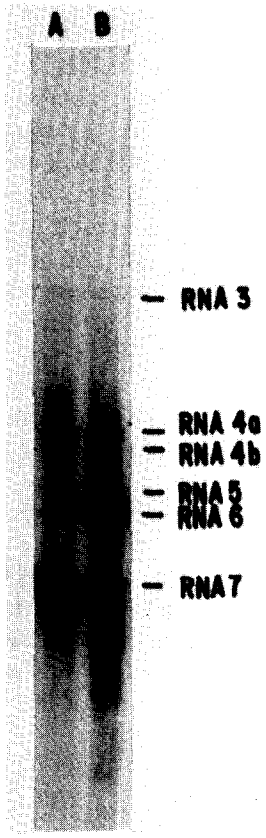

Fig. 1. Northern blots of intracellular poly(A)-containing RNA from TGEV-infected cells. The blotted gels were probed with nick-translated cDNA from the 3′ end of the TGEV genome. Six discrete RNA bands in cells infected with both the unadapted-gut virus (A) and the plaque-picked virus (B) were found. RNA 3 is the mRNA species encoding the E2-peplomer protein; RNA 6 encodes the E1-matrix protein and RNA 7 encodes the nucleocapsid protein.

et al. (1986). Although qualitatively identical, quantitative differences were apparent in RNAs 5 and 6 between the unadapted and the plaque-purified viruses. The genomic RNA (RNA 1) was not visualized by Northern blotting, perhaps, due to its large size and poor transfer to the nylon membrane. On the other hand, inadequate transfer could not account for the weak RNA 3 signal, the mRNA encoding the E2-peplomer protein, because 9.5 kb and 7.5 kb RNA markers on the same gel transferred efficiently (data not shown). This indicated that mRNA 3 is synthesized to a lesser extent than the other smaller transcripts. Our Northern blot analysis identified 3 distinct transcripts between the E2-peplomer mRNA (RNA 3) and the E1-matrix mRNA (RNA 6) of TGEV.

Cloning, sequencing and computer analysis of cDNA

To characterize these mRNA species further, the genetic region between the 2 virus surface structural genes E1 and E2 was sequenced. Genomic RNA was prepared from purified virus and was copied into cDNA after priming with random oligodeoxynucleotide primers (Binns et al., 1985). A cDNA library was prepared in the lambda expression vector λgt11. One recombinant, RP3, containing a 3.2 kb insert that hybridized to intracellular RNAs 3 through 6 by Northern blot analysis was selected for sequencing. This insert was subcloned into the pBluescript vector in order to facilitate double-stranded plasmid DNA sequencing.

The nucleotide sequence of the genetic region extending from the 3′ end of the E2-peplomer gene up to and including the 5′ end of the E1-matrix gene is illustrated in Fig. 2. Three major ORFs, A, B, C, that most probably represent the 5′ coding sequences of mRNAs 4a, 4b, and 5 are illustrated in Fig. 3. ORFs A and B are non-overlapping whereas ORFs B and C overlap by 11 bases. Each of the ORFs is preceded by the octamer $5'^{AA}_{TT}$CTAAAC 3′. This octameric sequence also precedes the genes for TGEV structural proteins E2, E1, N, and it also precedes a hypothetical hydrophobic polypeptide located downstream from the N gene (Kapke and Brian, 1986). These octameric sequences mark the 5′ boundary for each mRNA body sequence and appear to function in the initiation of leader RNA-primed transcription. The number and location of these recognition sites between the E1 and E2 structural genes are consistent with the mRNA species observed in Northern blots (Fig. 1). The proximity of the TGEV recognition sequences to the downstream AUG initiation codons is summarized in Table 1. Each of the 7 octameric sequences is followed immediately downstream by either 3 pyrimidines or 3 purines. No other $5'^{AA}_{TT}$CTAAAC occur in the coding sequences of the E1 and E2 structural genes for the Miller strain of TGEV (data not shown).

Nuclease S1 mapping

In order to demonstrate that the homologous recognition sequences immediately upstream from ORFs A and B are functional, we used S1 nuclease analysis to locate the 5′ end of the body sequences for subgenomic mRNAs 4a and 4b. A ^{32}P-labeled RNA probe, complementary to the positive-stranded viral genome and subgenomic

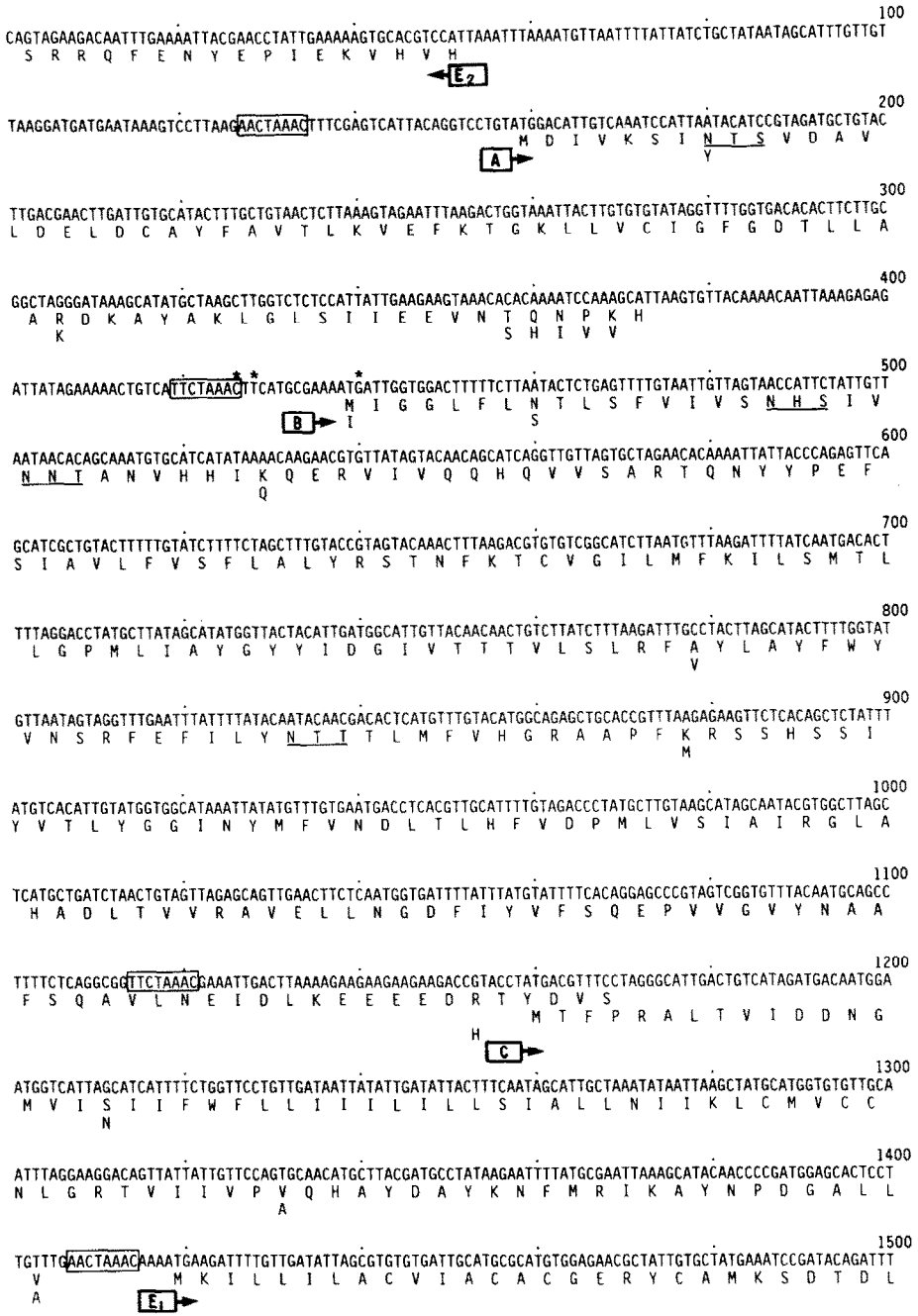

Fig. 2. The nucleotide sequence of the TGEV genomic region flanked by the genes for the E1-matrix and E2-peplomer proteins. The amino acid sequences for the 3 main ORFs A, B, and C are shown as single letter amino acid codes. Amino acids below the major sequence are substitutions found in the Purdue 115 strain (Rasschaert et al., 1987). The octanucleotide recognition sequence preceding each ORF is boxed. The potential N-glycosylation sites in ORFs A and B are underlined.

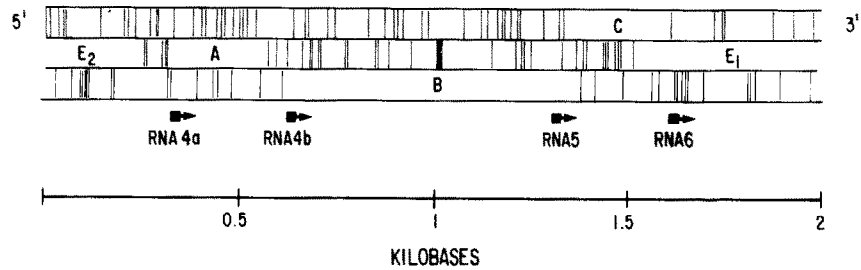

Fig. 3. The location of termination codons (vertical bars) in the three possible translational reading frames are shown for the cDNA sequence of Fig. 2. The main ORFs A, B, and C have predicted translation products of molecular weights of 7859, 27744 and 9287, respectively. The filled boxes (■) are positioned at the octameric recognition sequences for each mRNA species and the arrow indicates the direction of transcription.

mRNAs, was transcribed from the T7 promoter of Bluescript plasmid pB180. Plasmid pB180 contains 2171 TGEV specific nucleotides that extend from 1096 bases into the E2-peplomer gene to the 3′ end of ORF B (Fig. 2, nucleotide No. 1129). This probe was hybridized to poly(A)-containing RNA from uninfected and TGEV-infected ST cells, and then digested with nuclease S1, and the protected

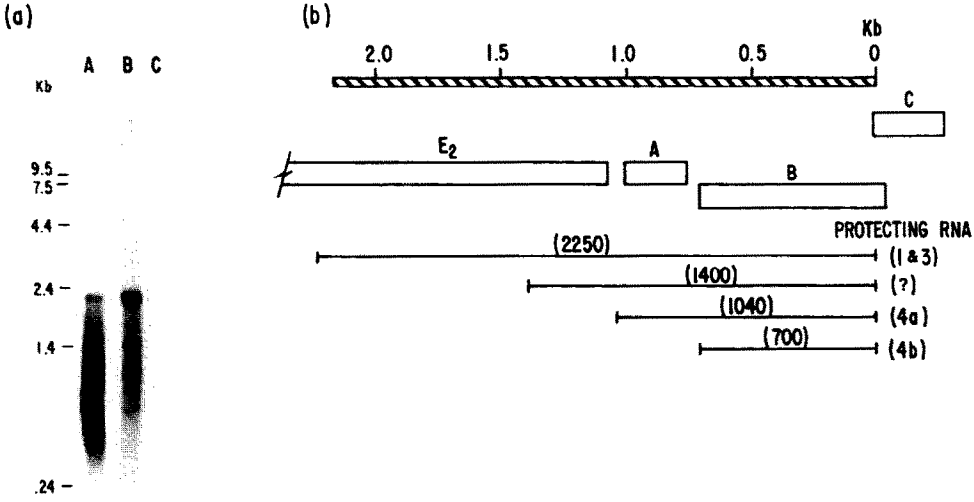

Fig. 4. (a) Nuclease S1 analysis to determine the location of the 5′ end of the body sequences for mRNA species 4a and 4b. A 2171 base, ^{32}P-labeled RNA probe was hybridized to poly(A) RNA extracted from either uninfected ST cells or ST cells infected with TGEV, followed by nuclease S1 digestion and agarose gel electrophoresis. The size in kb of known RNA markers are indicated. Lane A; ^{32}P-labeled, input probe. Lane B; RNA fragments protected by poly(A) RNA from infected ST cells. Lane C; ST cell control poly(A)-containing RNA. (b) Diagramatic representation of the S1 nuclease protection experiment. The hatched bar indicates the size and genomic location of the transcribed probe. The measured length and 5′ ends of the protected RNA fragments are shown below the indicated ORFs. Two RNA fragments, 1040 and 700 nucleotides in length, were protected by subgenomic mRNAs 4a and 4b.

RNA fragments were resolved by agarose gel electrophoresis (Fig. 4a). No RNA bands were detected from the control poly(A) RNA of uninfected ST cells. Four protected RNAs measuring 2250, 1400, 1040, and 700 nucleotides were observed with poly(A) RNA of TGEV-infected cells. The largest RNA fragment (2250 nucleotides) was protected from nuclease S1 digestion by annealing to the genomic RNA and subgenomic RNA 3. The 1400 nucleotide protected fragment was derived from an unidentified RNA species. In fact, using several different size input probes and nuclease S1 analysis, we have consistently observed this RNA species that extends only a few hundred nucleotides into the E2 gene. The protected RNA fragments of 1040 and 700 nucleotides were generated from subgenomic RNAs 4a and 4b, respectively, and therefore, the 5′ boundary of these mRNA species map upstream of ORFs A and B (Fig. 4b).

Properties of potential polypeptides

The deduced primary translation products for ORFs A, B and C are shown in Fig. 2. The presumed primary translation product of ORF A at the 5′ end of RNA 4a, contains 72 amino acids with a predicted molecular weight of 7859. One Asn near the N-terminal end of this polypeptide could function as a possible N-glycosylation site (Asn-X-Ser or Asn-X-Thr, where X is not proline). This polypeptide has 31 hydrophilic amino acids and is slightly acidic with a charge of −2 at pH 7.0.

The largest ORF, B, potentially gives rise to a primary translation product of 244 amino acids derived from the translated 5′ end of RNA 4b. The predicted molecular weight is 27744. This protein is strongly hydrophobic with 107 hydrophobic to 72 polar residues, and it has a net charge of −2 at pH 7.0. There are 3 Asn residues in the proper context for N-glycosylation and only one Cys residue. The C-terminal end of this protein is hydrophilic; however, none of the potential N-glycosylation sites are associated with this end.

The potential polypeptide of ORF C at the 5′ translable end of mRNA 5 overlaps the C-terminal portion of the large ORF B protein by 11 base pairs. ORF C encodes a potential polypeptide of molecular weight 9287 (82 amino acid residues). It is a basic protein (isoelectric point 8.33) that is strongly hydrophobic. There are 6 strongly basic residues and 4 strongly acidic residues yielding a net charge of +2 at neutral pH, and it contains no potential N-glycosylation sites.

Discussion

We are sequencing the virulent Miller strain of TGEV in order to find differences from attenuated TGEV strains that might correlate with increased virus virulence. A 3.2 kb cDNA, pRP3, synthesized from genomic RNA was cloned and sequenced. This clone extends from 1096 bases upstream from the end of the E2-peplomer gene to 773 bases into the E1-matrix gene. Stop code analysis in the genetic region between the E1 and E2 structural protein genes showed 3 ORFs with the potential coding information densely packed. Intergenic non-coding regions of 105 bases, 66

TABLE 1

Nucleotide context of intergenic initiation sequence for leader primed transcription of TGEV

Genomic sequence	No. of bases upstream from AUG initiation codon	Downstream gene product
TAAGT TACTAAAC TTT	24	E2 [a]
TTAAG AACTAAAC TTT	21	ORF A
TGTCA TTCTAAAC TTC	9	ORF B
GGCGG TTCTAAAC GAA	37	ORF C
GTTTG AACTAAAC AAA	3	E1
GGTAT AACTAAAC TTC	6	N [b]
TAACG AACTAAAC GAG	3	X3 [b]

[a] Genomic context of Purdue 115 strain (Rasschaert and Laude, 1987) confirmed in our laboratory for the Miller strain.

[b] Genomic context of Purdue 115 strain (Kapke and Brian, 1986).

bases, and 13 bases respectively were present between the end of E2 and ORF A, ORFs A and B, and ORF C and the AUG translation initiation codon for the E1-matrix protein. The end of ORF B and the beginning of ORF C overlapped by 11 bases. Upstream from each of these ORFs was the putative TGEV recognition sequence, $5'^{AA}_{TT}$CTAAAC. This sequence marks the approximate location that a leader RNA-polymerase complex initiates the synthesis of the subgenomic mRNAs. This homologous sequence involved in discontinuous leader-primed transcription is also upstream from the genes for the 3 TGEV major structural proteins, and it occurs again upstream from a postulated hydrophobic polypeptide encoded near the 3′ end of the genome (Kapke and Brian, 1986; Britton et al., 1988). The seven TGEV recognition sequences, flanking sequences and their relationship to the first downstream AUG start codon are summarized in Table 1. Nuclease S1 protection and northern blot analyses of poly(A)-containing RNA extracted from TGEV infected cells confirmed the existence of 3 mRNA species between the E1 and E2 structural genes.

The nucleotide sequence for the region of the TGEV genome between the E1 and E2 structural protein genes was presented for the avirulent Purdue 115 strain (Rasschaert et al., 1987). Two nucleotide differences between the Miller and Purdue sequences at positions 426, and 440 (Figs. 2 and 5) alter the interpretation of these sequences and suggest the existence of the third mRNA species and a larger ORF than originally postulated. A single T to C base substitution at position 426 establishes an intergenic recognition sequence $5'^{AA}_{TT}$CTAAAC marking the approximate 5′ limit of the body sequence for mRNA 4b. Fourteen bases downstream from this homologous sequence another T to G base change generates an AUG translation initiation codon in frame with ORF B of 725 bases. ORF B encodes a protein of 244 amino acid residues (predicted molecular weight 27744) in contrast to the truncated X2b polypeptide of 165 amino acid residues predicted for the Purdue strain. Consistent with these findings, we have identified in Northern blots a mRNA

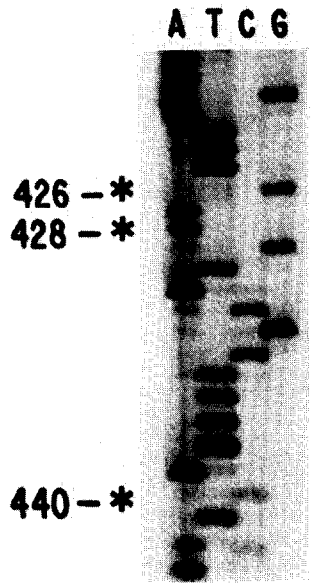

Fig. 5. A region of a 6% polyacrylamide sequencing gel showing the homologous recognition sequence and the translation initiation codon for ORF B. The negative sense strand was sequenced in this gel. The asterisks, indicating nucleotide differences between the Miller and the Purdue TGEV strains, correspond to positions 426, 428, and 440 of Fig. 2.

species (RNA 4b) with a unique 5′ region of the proper size to code for this polypeptide. In addition, in vitro translation of TGEV intracellular RNAs (Purdue strain) of this size have demonstrated a 25 kDa polypeptide (Jacobs et al., 1986). These data suggest that for TGEV each subgenomic mRNA species is functionally monocistronic at their 5′ translated ends and that internal initiation of protein synthesis does not occur.

In addition, the Miller and Purdue strains differ in that 2 large deletions are present in the Miller strain between the end of the E2 gene and ORF B. One, a 16 base deletion occurring in the non-coding region between E2 and ORF A, shortens this intergenic region from 121 to 105 nucleotides. The second is a 29 base deletion altering the C-terminus of the ORF A polypeptide by shifting the reading frame. This deletion occurs after base position 359 of the Purdue sequence (Rasschaert et al., 1987). The length of the ORF A polypeptide is increased by one amino acid residue, and then, the C-terminal 6 amino acid residues are different (Fig. 2).

Amino acid sequence homologies were sought among the predicted translation products of TGEV ORFs A, B and C and the polypeptides of MHV and IBV. For all three coronaviruses the predicted translation product of the ORF immediately 5′ of the E1-matrix protein gene showed similarities in hydropathicity distributions and amino acid sequence homologies. Fig. 6 shows a Kyte-Doolittle plot of N-terminally aligned potential products of TGEV ORF C (82 amino acid residues), the second downstream ORF of MHV mRNA 5 (88 amino acid residues) and the

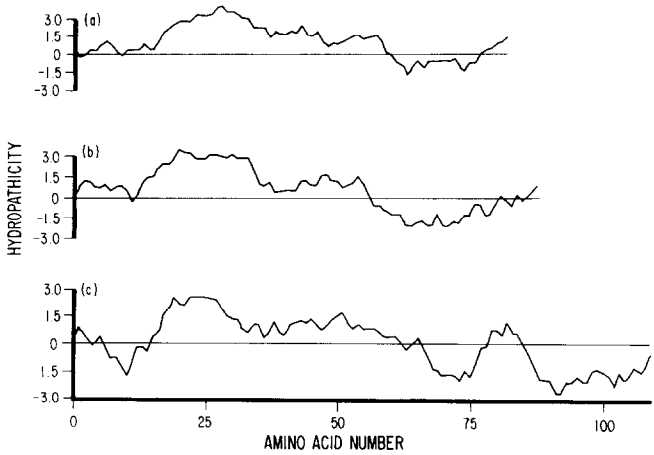

Fig. 6. Comparison of hydropathicity profiles (Kyte and Doolittle, 1982) for the TGEV, MHV, and IBV potential polypeptides encoded by the gene immediately upstream of the E1 protein gene. Each putative polypeptide was aligned at the N-terminus. The vertical scale is the average hydropathicity (+3 to −3) for a frame of 9 amino acids. Values above the midpoint line are hydrophobic and values below the line are hydrophilic. The hydropathicity plots are for the translation products of (a) ORF C of TGEV mRNA 5, (b) the second downstream ORF of MHV mRNA 5, and (c) the third downstream ORF of IBV mRNA D (Beaudette strain).

third downstream ORF of IBV mRNA D (109 amino acid residues). The additional length at the C-terminal hydrophilic end of the IBV polypeptide varies in different IBV serotypes (Cavanagh and Davis, 1988). Both the predicted MHV and IBV polypeptides have been detected in infected cell lysates (Skinner et al., 1985; Smith et al., 1987). The conserved hydrophobic patterns of these polypeptides and the potential translation product of TGEV ORF C suggest that each protein might perform analogous functions in the replication of coronaviruses. The amino acid sequence homologies were 23% between the predicted polypeptide of TGEV ORF C and the MHV polypeptide and 22% between the TGEV translation product and the polypeptide predicted for the IBV ORF.

In TGEV-infected cells the E1-matrix protein occurs as a heterogeneous series of bands (27–33 kDa) due to differing degrees of glycosylation (Garwes and Pocock, 1975; Brian et al., 1984; Hu et al., 1984). Sequencing studies predict a molecular weight of 27800 for the mature E1 protein (Laude et al., 1987). The postulated primary translation product encoded by ORF B, molecular weight 27744, is almost identical in size to E1, perhaps, preventing the resolution of these proteins on gels. At this time we cannot rule out the possibility that the ORF B protein may contribute to the observed E1 size heterogeneity. Even after N-linked glycosylation is blocked in TGEV-infected cells with tunicamycin, 2 protein bands remain in the E1 region of the gel (Hu et al., 1984; Jacobs et al., 1986). One of these protein bands is E1 but it is possible that the other could be the product of ORF B. In addition, some caution should be exercised in assigning monoclonal antibody specificity to the E1 protein simply on the basis of size because the E1-matrix protein and the

product encoded by ORF B could appear as similar proteins on SDS gels following immunoprecipitation.

References

Binns, M.M., Boursnell, M.E.G., Foulds, I.J. and Brown, T.D.K. (1985) The use of a random priming procedure to generate cDNA libraries of infectious bronchitis virus, a large RNA virus. J. Virol. Methods 11, 265–269.

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., Hogue, B., Lapps, W., Potts, B. and Kapke, P. (1984) Comparative structure of coronaviruses. In: Proceedings, 4th International Symposium on Neonatal Diarrhea, pp. 100–116. Veterinary Infectious Disease Organization, Saskatoon.

Britton, P., Carmenes, R.S., Page, K.W., Garwes, D.J. and Parra, F. (1988) Sequence of the nucleoprotein gene from a virulent British field isolate of transmissible gastroenteritis virus and its expression in *Saccharomyces cerevisiae*. Mol. Microbiol. 2, 89–99.

Brown, T.D.K., Boursnell, M.E.G., Binns, M.M. and Tomley, F. (1984) Cloning and sequencing of 5′ terminal sequences from avian infectious bronchitis virus genomic RNA. J. Gen. Virol. 67, 221–228.

Cavanagh, D. and Davis, P.J. (1988) Evolution of avian coronavirus IBV: sequence of the matrix glycoprotein gene and intergenic region of several serotypes. J. Gen. Virol. 69, 621–629.

Cheung, A.K. (1988) Fine mapping of the immediate-early gene of the Indiana-Funkhauser strain of pseudorabies virus. J. Virol. 62, 4763–4766.

Church, G.M. and Gilbert, W. (1984) Genome sequencing. Proc. Natl. Acad. Sci., USA 81, 1991–1995.

Garwes, D.J. and Pocock, D.H. (1975) The polypeptide structure of transmissible gastroenteritis virus. J. Gen. Virol. 29, 25–34.

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

Hu, S., Bruszewski, J., Boone, T. and Souza, L. (1984) Cloning and expression of the surface glycoprotein gp 195 of porcine transmissible gastroenteritis virus. In: R.M. Chanock and R.A. Lerner (Eds.), Modern Approaches to Vaccines, pp. 219–223. Cold Spring Harbor Laboratory, New York.

Huynh, T.V., Young, R.A. and Davis, R.W. (1985) Constructing and screening cDNA libraries in λgt10 and λgt11. In: D. Glover (Ed.), DNA Cloning, Vol. I. A Practical Approach, pp. 49–78. IRL Press, Oxford.

Jacobs, L., Van Der Zeijst, B.A.M. and Horzinek, M.C. (1986) Characterization and translation of transmissible gastroenteritis mRNAs. J. Virol. 57, 1010–1015.

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

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

Lai, M.M.C. (1986) Coronavirus leader-RNA-primed transcription an alternative mechanism to RNA splicing. Bioessays 5, 257–260.

Lai, M.M.C., Patton, C.D., Baric, R.S. and Stohlman, S.A. (1983) Presence of leader sequences in the mRNA of mouse hepatitis virus. J. Virol. 46, 1027–1033.

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.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning; A Laboratory Manual. Cold Spring Harbor Laboratory, New York.

McClurkin, A.W. and Norman, J.O. (1966) Studies on transmissible gastroenteritis of swine. II. Selected characteristics of a cytopathogenic virus common to five isolates from transmissible gastroenteritis. Can. J. Comp. Med. Vet. Sci. 30, 190–198.

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.

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.

Saif, L.J. and Bohl, E.H. (1986) Transmissible gastroenteritis. In: A.D. Leman, B. Straw, R.D. Glock, W.L. Mengeling, R.H.C. Penny and E. Scholl (Eds.), Diseases of Swine, 6th edit., pp. 255–274. Iowa State Unit. Press, Ames, Iowa, U.S.A.

Skinner, M.A., Ebner, D. and Siddell, S.G. (1985) Coronavirus MHV-JHM mRNA 5 has a sequence arrangement which potentially allows translation of a second downstream open reading frame. J. Gen. Virol. 66, 581–592.

Smith, A.R., Boursnell, M.E.G., Binns, M.M., Brown, T.D.K. and Inglis, S.C. (1987) Identification of a new gene product encoded by mRNA D of infectious bronchitis virus. In: M.M.C. Lai and S.A. Stohlman (Eds.), Coronaviruses, Adv. Exp. Med. Biol., Vol. 218, pp. 47–54. Plenum Press, New York.

Spaan, W.J.M., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., Van der Zeijst, B.A.M. and Siddell, S. (1983) Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. Eur. Mol. Biol. Organ. J. 2, 1839–1844.

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

Wesley, R.D. and Woods, R.D. (1986) Identification of a 17,000 molecular weight antigenic polypeptide in transmissible gastroenteritis virus-infected cells. J. Gen. Virol. 67, 1419–1425.

Wesley, R.D., Woods, R.D., Correa, I. and Enjuanes, L. (1989). Lack of protection in vivo with neutralizing monoclonal antibodies to transmissible gastroenteritis virus. Vet. Microbiol. 18, 197–208.

(Received 26 October 1988; revision received 30 January 1989)