Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

VIROLOGY 206, 817–822 (1995)

Complete Sequence (20 Kilobases) of the Polyprotein-Encoding Gene 1 of Transmissible Gastroenteritis Virus

JEAN-FRANCOIS ELEOUET, DENIS RASSCHAERT, PATRICK LAMBERT, LAURENT LEVY, PATRICE VENDE, and HUBERT LAUDE[1]

Unité de Virologie et Immunologie Moléculaires, Institut National de la Recherche Agronomique, 78350 Jouy-en-Josas, France

Received July 22, 1994; accepted November 21, 1994

The entire nucleotide sequence of cloned cDNAs containing the 5′-untranslated region and gene 1 of Purdue-115 strain of transmissible gastroenteritis virus (TGEV) was determined. This completes the sequence of the TGEV genome, which is 28,579 nucleotides long. The gene 1 is composed of two large open reading frames, ORF1a and ORF1b, which contain 4017 and 2698 codons, respectively (stop excluded). A brief, three-codon-long ORF is present upstream of ORF1a. ORF1b overlaps ORF1a by 43 bases in the (−1) reading frame. *In vitro* experiments indicated that translation of the ORF1a/b polyprotein involves an efficient ribosomal frameshifting activity, as previously shown for other coronaviruses. Analysis of the predicted ORF1a and ORF1b translation products revealed that the putative functional domains identified in infectious bronchitis virus (IBV), mouse hepatitis virus (MHV) and human coronavirus 229E (HCV 229E) are all present in TGEV. The amino-terminal half of the ORF1a product exhibits greater divergence than the carboxy-terminal half, including within the TGEV/HCV229E pair. The ORF1b protein is overall highly conserved among the above four coronaviruses, except a divergent region situated near the carboxy terminus.

INTRODUCTION

Transmissible gastroenteritis virus (TGEV), an enteropathogenic virus causing a fatal diarrhea in newborn piglets, is a member of the Coronavirus genus. This genus comprises enveloped viruses with a genome consisting of a large single-stranded, positive-sense RNA molecule. In infected cells, full-length and five to seven subgenomic RNAs are synthesized by a viral RNA-dependent polymerase assumed to be coded by the 5′-most gene (gene1)(review: Lai, 1990). Complete sequence of the genomic RNA has been determined for three members: infectious bronchitis virus or IBV (Boursnell *et al.*, 1987), mouse hepatitis virus or MHV strains A59 (Pachuk *et al.*, 1989) and JHM (Lee *et al.*, 1991), and human coronavirus 229E or HCV 229E (Herold *et al.*, 1993). The unusual length of the coronavirus genome (about 3×10^4 nucleotides) is mainly accounted for by the size of the gene 1 which occupies more than two-thirds of the genome. This region contains two large open reading frames, ORF1a and ORF1b, from which two products are translated. The latter is synthesized via a (−1) ribosomal frameshifting mechanism, facilitated by a pseudoknot structure (Brierley *et al.*, 1989; Bredenbeek *et al.*, 1990; Lee *et al.*, 1991; Herold *et al.*, 1993). Both ORF1a and ORF1a/b products are assumed to be processed in a number of functional subunits by at least two viral proteinases (Gorbalenya *et al.*, 1989; Baker *et al.*, 1993). Although several cleavage products have been identified for MHV and IBV (Denison *et al.*, 1991, 1992; Liu *et al.*, 1994), the structure and functions of the mature products are largely unknown.

The TGEV genome region encoding the previously observed RNA-dependent RNA polymerase activity (Dennis and Brian, 1982) remains undetermined. The nucleotide sequence of 8.3-kb extending from the 3′ end has been determined (Kapke and Brian, 1986; Rasschaert *et al.*, 1987; Rasschaert and Laude, 1987; Laude *et al.*, 1987). This region encodes the nucleocapsid protein N, the membrane (M) and spike (S) glycoproteins (Laude *et al.*, 1986), and a minor membrane protein called sM (Godet *et al.*, 1992). Additionally, there are three open reading frames, ORF3a, ORF3b, and ORF7, the functions of which remain undetermined. A membrane-associated polypeptide (7K) detected in infected cells has been assigned to ORF7 (Tung *et al.*, 1992). These putative nonstructural protein genes as well as the four structural protein genes are expressed through the synthesis of six to seven subgenomic RNAs, depending on the virus strain (Jacobs *et al.*, 1986; Rasschaert *et al.*, 1987; Wesley *et al.*, 1989). The remaining part of the genome was assumed to contain no other genetic element than the gene 1, since the S protein was assigned to mRNA 2. Indeed, sequencing of the 930 nucleotides to the 5′ side of S gene revealed an ORF with a translation product 50% identical with that of IBV ORF1b (Wesley, 1990).

In this paper we report the complete nucleotide sequence of the gene 1 of TGEV and a comparative

[1] To whom reprint requests should be addressed at INRA, Laboratoire de Virologie et Immunologie Moléculaires, 78352 Jouy-en-Josas Cedex, France.

0042-6822/95 $6.00

analysis of the structural organization of the ORF1a/b predicted translation products with those of HCV 229E, IBV, and MHV.

MATERIALS AND METHODS

Virus and cells

The high cell passage Purdue-115 strain of TGEV was used as a virus source. Virus propagation and isolation of genomic RNA from purified virions were performed as described (Laude *et al.*, 1986; Rasschaert *et al.*, 1987).

cDNA cloning and sequencing

Strategy and protocol were as reported previously (Rasschaert *et al.*, 1987). Briefly, purified genomic RNA was copied by reverse transcriptase using the oligo(dT) 12–18 ("2–" clone set; see Fig. 1), an S gene-specific 30-mer primer (Rasschaert and Laude, 1987) (6–47 clone), or a 2–40 clone-specific primer (5′-ATGTATGCT-GTGAACAAAAT-3′); "P–" clone set. RNase T2-treated cDNA–RNA hybrids were dC-tailed and inserted in *Pst*I-cut dG-tailed pBR322. Transfection of *Escherichia coli* RR1 and selection of recombinant clones were performed following standard methods. DNA sequencing by Sanger's chain termination method and sequence analysis were accomplished as previously described (Rasschaert *et al.*, 1987). For sequencing of the PCR-amplified region 1 (see below), PCR products were sonicated and subcloned into M13mp19. PCR-amplified regions 2 and 3 were sequenced directly after purification on Quiaquick (Quiagen) columns. Sequencing was made using an Applied Biosystems Model 373A automated DNA sequencer. Each cDNA clone was sequenced on both strands.

cDNA synthesis and amplification by polymerase chain reaction (PCR)

Random-primed cDNA synthesis was done using 10 μg of total cytoplasmic RNA extracted from TGEV-infected ST cells by the guanidium thyocianate method (Chomczynsky and Sacchini, 1987). One-tenth of the resulting cDNA was used for PCR together with *Taq* Polymerase (5U, Promega) and 100 ng of each of the following upstream and downstream oligonucleotides: region 1 (7522–10727): 5′-GTTGGTGATGAATTCACTTCAT-3′, 5′-CATACTAGCTGCAGCCTGC-3′; region 2 (12249–12751): 5′-TGTGTTGTCTGTGGTTGTTGGC-3′, 5′-GCAAGCACC-TACTGTCACGAG-3′; region 3 (18148–18533): 5′-CCA-CACACCAGCTTATGATAG-3′, 5′-CGTTCGAGACACATA-CGCTGG-3′. PCRs were done by 30 cycles of denaturation (94°, 30 sec), annealing (55°, 30 sec), and extension (72°, 10 min for region 1, 1 min for regions 2 and 3), using a Perkin–Elmer Cetus DNA thermal cycler. For region 1, cDNA was phenol-extracted, ethanol-precipitated, and sonicated before subcloning.

Primer extension reaction

^{32}P-labeled oligonucleotide 755 (5′-TTCGAGTTGGTG-TCCGAAG-3′) (0.5 pmol) together with TGEV-purified genomic RNA (10 μg) were denatured in a total volume of 15 μl of water/hydroxymethylmercury (2/1,v/v) at room temperature for 10 min. After incubation at 37° for 10 min, the samples were adjusted to first-strand cDNA buffer in a total volume of 40 μl, MMLV reverse transcriptase (30 units) was added, and the mixture incubated at 42° for 1 hr. Samples were purified by passage through a Sephadex G-50 spun column, phenol extracted, and ethanol precipitated.

Sequence data analysis

Sequences were analyzed using the Microgenie program (March 1985, Beckman version) and the UWGCG (University of Wisconsin, Genetic Computer Group) sequence analysis software package. Multiple sequence analysis and optimal alignments were performed with either the GCG program GAP (default settings) or the Clustal V packages.

Ribosomal frameshifting analysis

A 2144-b p *Pst*I–*Pst*I cDNA fragment, corresponding to nucleotides 11,182–13,326 of the genomic RNA and containing the frameshifting region, was subcloned into a pTZ plasmid (Pharmacia) at the *Pst*I site, downstream of a T7 RNA polymerase promoter, resulting in the pTZRF plasmid. The first AUG is localized 4 nucleotides downstream of the *Pst*I site and is in frame with ORF1a. Correct orientation was determined by restriction enzyme digestion analysis. The pTZRF plasmid DNA was linearized with *Bgl*II (nucleotide 12925 of the genomic RNA) and transcribed *in vitro* using T7 RNA polymerase (Boehringer Mannheim). *In vitro* transcribed, capped RNAs, were translated in a rabbit reticulocyte lysate in the presence of [^{35}S]methionine. Products were visualized by 10% SDS–polyacrylamide gel and autoradiography. Frameshifting efficiency was estimated from the autoradiography using a Gelman DCD 16 scanner.

RESULTS AND DISCUSSION

Derivation and characterization of TGEV gene 1 cDNA clones

Thirteen clones, derived from three subsequent cDNA libraries (6, 2, and P), allowed the sequencing of 84% of gene 1 with 70% of the sequence being determined from at least two independent clones (Fig. 1). A region comprised between nucleotides 7522 and 10,727 from the 5′ end of the genomic RNA, not represented in the cDNA libraries, was amplified by RT-PCR (Fig. 1). The resulting 4-kb fragment was sequenced after shotgun in M13 phage. Two other regions, corresponding to nucleotides 12,249–12,751 (pseudoknot region) and 18,148–

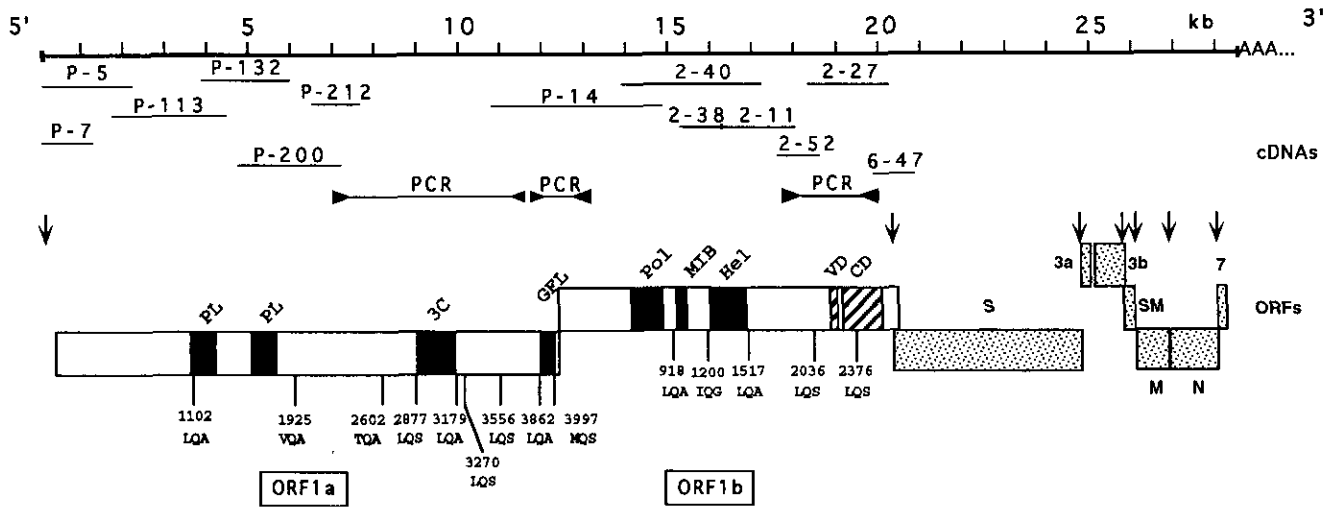

FIG. 1. TGEV genomic organization and position of cDNA clones used to determine the gene 1 sequence. Vertical arrows indicate functional intergenic consensus sequences. The position and size of the cDNA clones are shown. The double-head arrows indicate the regions amplified by RT-PCR. The major ORFs are represented as boxes in the 0, −1, and −2 reading frames with putative functional domains of the ORF1 product as black boxes. PL, papain-like protease; 3CL, 3C-like protease; GFL, growth factor/receptor-like; Pol, polymerase; MIB, metal ion binding; Hel, helicase; VD, ORF1b variable domain; CD, conserved domain (see text). Positions of putative 3C-like cleavage sites are indicated in amino acids from the ORF1a and ORF1b N terminus.

18,533 (ORF1b divergent region, see below), both covered by a single cDNA clone, were amplified by RT-PCR and sequenced directly using the same primers. The sequence data, combined with those previously reported (Rasschaert *et al.*, 1987), revealed that the genomic RNA of TGEV (Purdue-115 strain) is 28,579 nucleotides long. This sequence has been deposited with the EMBL database under Accession No. Z34093.

5′ Region of the genome

The sequence of the 5′ end of the genomic RNA was obtained from cDNA clones P-5 and P-7. To confirm that the very 5′ end of the genomic RNA was included, a primer extension analysis was performed (Fig. 2A). The extension products comigrated with an adenine residue present in the P-5 and P-7 clone sequences. Thus, the 5′ end sequence of the TGEV Purdue-strain appears to be 5′-**AC**UUUUAAAG-3′, which differs from those previously published for the Purdue strain (5′-UUUUUAAAG-3′; Sethna *et al.*, 1991) and for the FS772/70 strain (5′-**CC**UUUUUAAAG-3′; Page *et al.*, 1990). The sequence of the first four nucleotides 5′-ACUU-3′ is identical to the 5′ ends of IBV and HCV 229E (Brown *et al.*, 1986; Herold *et al.*, 1993).

The predicted AUG initiation codon of ORF1a begins at nucleotide 315 and it is preceded by an in-frame three codons ORF which starts at nucleotide 117 (Fig. 2B). The AUG initiator codon of this "mini ORF" is in a less optimal context for efficient translation than that of ORF1a, in which an adenine is present at position −3 (Kozak, 1986). TGEV "mini ORF" is 3 codons long, instead of 8 codons for MHV and 11 codons for IBV and HCV 229E (Brown *et al.*, 1986; Soe *et al.*, 1987; Boursnell *et al.*, 1987; Pachuk *et al.*, 1989; Herold *et al.*, 1993). The yeast GCN4 gene and cauliflower mosaic virus 35S RNA provide two examples of upstream ORF (uORF) that suppress translation of a major cistron in a manner that is both regulated and independent of the amino acid sequence of the encoded peptide (reviewed by Geballe and Morris, 1994). The extreme brevity of the uORF present in TGEV led us to speculate that the regulation of ORF1 translation thought to occur in coronaviruses might also operate via a sequence-independent uORF.

Putative functional domains of ORF1

The predicted ORF1a extends from nucleotides 315 to 12,365, resulting in a 4017-codon ORF (stop excluded). Pairwise comparison of the amino acid sequence of TGEV ORF1a with those of MHV, IBV, and HCV 229E showed that the amino-terminal half of the predicted ORF1a polypeptide is more divergent than the carboxy-terminal half. For the TGEV/HCV229E pair, which includes the two most closely related of these viruses, the values are 53% of similarity (32% of identity) for the N-terminal half and 70% of similarity (51% identity) for the C-terminal half. Moreover, a sequence extending from amino acid 1435 to amino acid 1602 in HCV 229E ORF1a is not present in TGEV and partly accounts for the longer ORF1a of HCV 229E (4085 codons stop excluded). Two papain-like (PL) domains (amino acids 1080–1273 and 1575–1770), a 3C-like (3CL) protease domain (2877–3179), and a growth factor/receptor-like (GFL) domain (3862–4005) are predicted in TGEV ORF1a product (Fig. 1), as previously described for other coronaviruses (Gorbalenya *et al.*, 1989; Lee *et al.*, 1991; Herold *et al.*, 1993). The 3C-like and growth factor/receptor-like domains

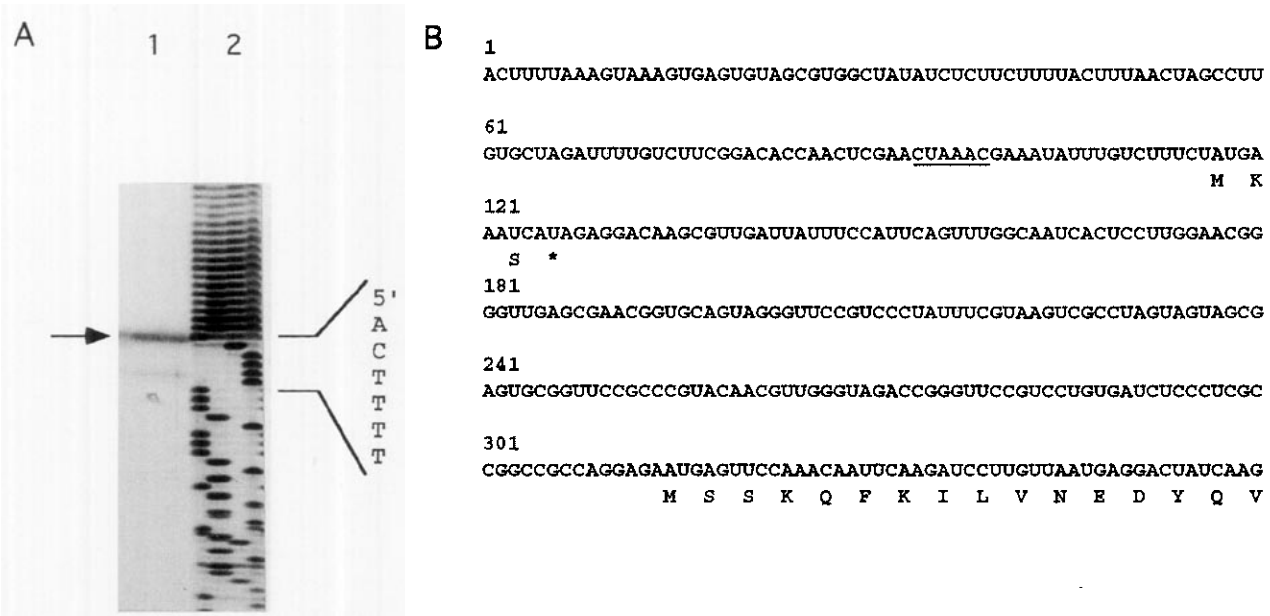

FIG. 2. 5′ End region of TGEV genomic RNA. (A) Primer extension products using 10 μg of genomic RNA (lane 1) were run next to sequencing reactions with cDNA clone P-5 (lane 2). Oligonucleotide 755 served as a primer for both extension and sequencing reactions. The band corresponding to the last TGEV-derived nucleotide is marked by an arrow. (B) Sequence of the 5′ region of the Purdue-115 strain genome. The leader sequence, ending at the consensus sequence "CUAAAC" (underlined) is 99 nucleotides long. The translation start site of the ORF1a protein is positioned at nucleotide 315. The upstream "mini ORF" (three codons) is shown.

show about 50% identity after optimal alignment. The C and H residues, identified as critical for the activity of the first PL domain of MHV (Baker *et al.*, 1993), are conserved. Two markedly hydrophobic regions predicted in ORF1a polyprotein are located at positions 2640–2980 and 3200–3450, on both sides of the 3C-like domain. Search for putative 3C-like cleavage sites (Gorbalenya *et al.*, 1989) in the TGEV gene 1 predicted product revealed 14 potential sites. The sites are clustered in the carboxy-terminal half of ORF1a (7 sites) and ORF1b (5 sites) proteins and the majority (10 sites) contains LQA/S triplets (Fig. 1). They are generally found at comparable positions for each of the four coronaviruses examined with only slight variations of sequence (Gorbalenya *et al.*, 1989; Lee *et al.*, 1991; Denison *et al.*, 1991). A putative 3C-like cleavage site (LQA) present in the first papain-like domain could be a target for a regulatory mechanism (inactivation) of this protease activity. LQA, LQS, and TQA motifs are not found in any of the other ORFs identified in the TGEV sequence.

Pairwise comparison of the amino acid sequence of TGEV ORF1b with those of MHV, IBV, and HCV 229E confirmed that ORF1b is more conserved among coronaviruses than ORF1a with 72% identity between TGEV and HCV 229E proteins and 54% identity for TGEV and IBV or TGEV and MHV proteins. However, we found a non-co-linear region over a background of otherwise high identity (Fig. 3). This region, extending from nucleotide 18828 to nucleotide 19006 for TGEV, is variable both in length and sequence, including within it the TGEV-HCV 229E pair. For MHV, a packaging signal has been identified in this region (Fosmire *et al.*, 1992). Analysis of the TGEV and HCV 229E RNA sequences in this region revealed no predicted secondary structure closely resembling that described for MHV. Similar comparison made for MHV and IBV (Fosmire *et al.*, 1992; Penzes *et al.*, 1994) led to speculate that the encapsidation signal may be differently located among coronaviruses. The TGEV ORF1b predicted polyprotein contains putative RNA polymerase (Pol) domain (amino acids 530–832), metal ion binding (MIB) domain (amino acids 920–995), and helicase (Hel) domain (amino acids 1198–1297) (Fig. 1). These three domains, together with the carboxy-terminal region of ORF1b (amino acids 2267–2581 for TGEV), show significant amino acid conservation with the torovirus Berne virus (BEV) and the arterivirus equine arteritis virus (EAV), which both belong to the coronavirus superfamily (Den Boon *et al.*, 1991; Snijder *et al.*, 1991).

The frameshifting region

The ribosomal frameshifting mechanism described for translation of ORF1 IBV, MHV, and HCV 229E depends on at least two elements, a tertiary structure termed pseudoknot and a ribosomal slippage site UUUAAAC where frameshifting occurs. Two models based on experimental studies have been proposed for tertiary structure of the pseudoknot, involving two (for IBV; Brierley *et al.*, 1989) or three ("elaborated pseudoknot," for HCV 229E; Herold and Siddell, 1993) stem structures. A structure prediction of this region drawn by analogy to that predicted for HCV 229E is shown in Fig. 4A. To evaluate the efficiency of TGEV frameshifting, we made coupled *in vitro* transcription/translation experiments from a *Bgl*II-

```
HCV(2097-2358)   AKRKVGLTPPLSILKNLGVVATYKFVLWDYEAERPLTSFTKSVCGYTDFA--EDVCTCYD
IBV(2110-2361)   AKRNIRTLPNNRILKGLGVDVTNGFVIWDYANQTPLYRNTVKVCAYTDIE-PNGLVVLYD
MHV(2117-2404)   AERSIRPHPELKLFRSSNIHVCWNHVLWDYAKDSVFCSSTYKVCKYTDLQCIESLNVLFD
TGEV(2096-2348)  AKRKLGLTPPLTILRNLGVVATYKFVLWDYEAERPFSNFTKQVCSYTDLD--SEVVTCFD
                 * * .   *   ...   .      *.***  .  .   *  ** ***.     .   .*

HCV              NSIQGSYERFTLSTNAVLFSATAVKTGGKSLPAIK-LNFGMLNGNAIATVKSEDGNIKNI
IBV              DRY-GDYQSFLAADNAVLVSTQCYK-----------RYSYVEIPSNLLVQNG-------
MHV              GRDNGALEAFKKCRNGVYINTTKI----KSLSMIKGPQRADLNGVVVEKVGDSDVEF---
TGEV             NSIAGSFERFTTTRDAVLISNNAVK----GLSAIK-LQYGLLNDLPVSTVGN-----KPV
                     *  . *    ..*  .                    . ..       *

HCV              NWF---------VYVRKDGKPVDHY----------------------DGFYTQGRNLQD
IBV              --MPLKDGANLYVYKRVNGAFVTL----------------------PNTINTQGRSYET
MHV              -WFAMRRDGDDVIFSRTGSLEPSHYRSPQGNPGGNRVGDLSGNEALARGTIFTQSRFLSS
TGEV             TWY---------IYVRKNGEYVEQI----------------------DSYYTQGRTFET
                            .. *  .                              **.*

HCV              FLPRSTMEEDFLNMDIGVFIQKYGLEDFNFEHVVYGDVSKTTLGGLHLLISQVRLSKMGI
IBV              FEPRSDIERDFLAMSEESFVERYG-KDLGLQHILYGEVDKPQLGGLHTVIGMYRLLRANK
MHV              FSPRSEMEKDFMDLDEDVFIAKYSLQDYAFEHVVYGSFNQKIIGGLHLLIGLARRPKKSN
TGEV             FKPRSTMEEDFLSMDTTLFIQKYGLEDYGFEHVVFGDVSKTTIGGMHLLISQVRLAKMGL
                 * *** .* **. .    *. .*.  *  ..*...*   .  .**.* .*.  *  .

HCV              LKAEEFVAASDITLKCCTVTYLNDPSSKTVCTYMDLLLDDFVSVLKSLDLTVV---SKVH
IBV              LNAKSVTNSDSDVMQNYFVL-SDNGSYKQVCTVVDLLLDDFLELLRNILKEYGTNKSKVV
MHV              LVIQEFVPY-DSSIHSYFITDENSGSSESVCTVIDLLLDDFVDIVKSLNLKCV---SKVV
TGEV             FSVQEFMNNSDSTLKSCCITYADDPSSKNVCTYMDILLDDFVTIIKSLDLNVV---SKVV
                 .          .   .  .  *   *** .*.*****. .....        ***
```

FIG. 3. Multiple amino acid sequence alignments between the four coronaviruses TGEV, HCV 229E, IBV, and MHV-JHM in the ORF1b divergent region with the Clustal V program. Stars correspond to a 4/4 amino acid conservation; points correspond to a minimum of 2/4 amino acid conservation.

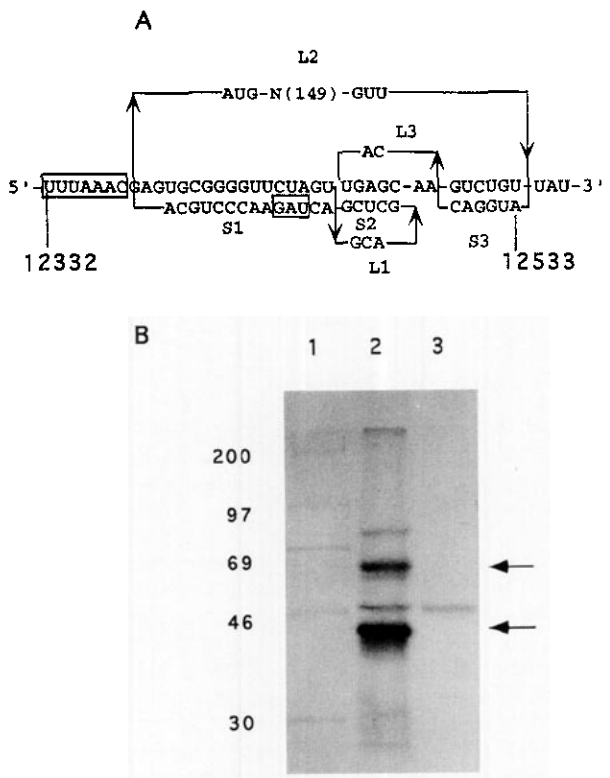

FIG. 4. Analysis of the TGEV RNA-mediated ribosomal frameshifting. (A) Possible structure of the pseudoknot region by analogy to that proposed for HCV 229E (Herold and Siddell, 1993). S, stem; L, loop. The putative ribosomal slippery site UUUAAAC and the ORF1a stop codon are boxed. (B) SDS–PAGE analysis of *in vitro*-translated products from the pTZRF plasmid. Lane 1, molecular weight markers; lane 2, pTZRF/*Bgl*II RNA; lane 3, no RNA. The arrows indicate the 42K and 64K major species.

linearized pTZRF plasmid containing the ORF1a/ORF1b overlap region. The resulting translation products corresponded to the ones expected with (63K; 577-amino-acid-long) or without (43K; 390-amino-acid-long) a frameshifting event (Fig. 4B). The efficiency of the frameshifting was estimated to be 20%. A slight band migrating around 80K might correspond to a longer, 138-amino-acid-long product (78K) translated from nondigested plasmid, as this plasmid contains a stop codon 411 nucleotides downstream from the *Bgl*II site.

With the newly available sequence, TGEV is the fourth coronavirus to be completely sequenced, and the second in the genetic subset which also includes feline coronavirus (FIPV), canine coronavirus (CCV), HCV 229E, and porcine epidemic diarrhea virus (PEDV) (see Duarte *et al.*, 1993). It is hoped that this data will not only facilitate future investigations on the molecular pathogenesis of TGEV but also contribute to elucidation of the structure and function of the coronavirus ORF1a- and ORF1b-derived products.

ACKNOWLEDGMENT

We thank J. Herold (Würzburg) for his assistance in predicting a TGEV "elaborated" pseudoknot structure.

REFERENCES

Baker, S. C., Yokomori, K., Dong, S., Carlisle, R., Gorbalenya, A. E., Koonin, E. V., and Lai, M. M. C. (1993). Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus. *J. Virol.* **67**, 6056–6063.

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.

Bredenbeek, P. J., Pachuk, C. J., Noten, A. F. H., Charite, J., Luytjes, W., Weiss, S. R., and Spaan, W. J. M. (1990). The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59: A highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Res.* **18,** 1825–1832.

Brierley, I., Digard, P., and Inglis, S. C. (1989). Characterization of an efficient coronavirus ribosomal frameshifting signal: Requirement for an RNA pseudoknot. *Cell* **57,** 537–547.

Brown, J. D. K., Boursnell, M. E. G., Binns, M. M., and Tomley, F. M. (1986). Cloning and sequencing of the 5′ terminal sequences from avian infectious bronchitis virus genomic RNA. *J. Gen. Virol.* **67,** 221–228.

Chomczynski, P., and Sacchini, N. (1987). Single-step method of RNA isolation by acid guanidium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162,** 156–159.

Denison, M. R., Zoltick, P. W., Leibowitz, J. L., Pachuk, C. J., and Weiss, S. R. (1991). Identification of polypeptides encoded in open reading frame 1b of the putative polymerase gene of the murine coronavirus mouse hepatitis virus A59. *J. Virol.* **65,** 3076–3082.

Denison, M. R., Zoltick, P. W., Leibowitz, J. L., Hughes, S. A., Giangreco, B., Olson, A. L., Perlman, S., Leibowitz, J. L., and Weiss, S. R. (1992). Intracellular processing of the N-terminal ORF1a proteins of the coronavirus MHV-A59 requires multiple proteolytic events. *Virology* **189,** 274–284.

Dennis, D. E., and Brian, D. A. (1982). RNA-dependent RNA polymerase activity in coronavirus-infected cells. *J. Virol.* **42,** 153–164.

Den Boon, J. A., Snijder, E. J., Chirnside, E. D., De Vries, A. A. F., Horzinek, M. C., and Spaan, W. J. M. (1991). Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. *J. Virol.* **65,** 2910–2920.

Duarte, M., Gelfi, J., Lambert, P., Rasschaert, D., and Laude, H. (1993). Genome organization of porcine epidemic diarrhea virus. *In* "Coronaviruses, Molecular Biology and Virus–Host Interactions" (H. Laude and J.-F. Vautherot, Eds.), Vol. 342, pp. 55–60. Plenum, New York.

Fosmire, J. A., Hwang, K., and Makino, S. (1992). Identification and characterization of a coronavirus packaging signal. *J. Virol.* **66,** 3522–3530.

Geballe, A. P., and Morris, D. R. (1994). Initiation codons within 5′-leaders of mRNAs as regulators of translation. *Trends Biochem. Sci.* **19,** 159–164.

Godet, M., L'Haridon, R., Vautherot, J.-F., and Laude, H. (1992). TGEV coronavirus ORF-4 encodes a membrane protein that is incorporated into virions. *Virology* **188,** 666–675.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P., and Blinov, V. M. (1989). Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res.* **17,** 4846–4861.

Gorbalenya, A. E., and Koonin, E. V. (1989). Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Res.* **17,** 8413–8440.

Herold, J., Raabe, T., Schelle-Prinz, B., and Siddell, S. G. (1993). Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. *Virology* **195,** 680–691.

Herold, J., and Siddell, S. G. (1993). An "elaborated" pseudoknot is required for high frequency frameshifting during translation of HCV 229E polymerase mRNA. *Nucleic Acids Res.* **21,** 5838–5842.

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.

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

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eucaryotic ribosomes. *Cell* **44,** 283–292.

Lai, M. M. C. (1990). Coronavirus: Organization, replication and expression of genome. *Ann. Rev. Microbiol.* **44,** 303–333.

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.

Lee, H.-J., Shieh, C.-K., Gorbalenya, A. E., Koonin, E. V., La Monica, N., Tuler, J., Bagdzhadzhyan, A., and Lai, M. M. C. (1991). The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. *Virology* **180,** 567–582.

Liu, D. X., Brierley, I., Tibbles, K. W., and Brown, T. D. K. (1994). A 100 kilodaltons polypeptide encoded by open reading frame (ORF) 1b of the coronavirus infectious bronchitis virus is processed by ORF1a products. *J. Virol.* **68,** 5772–5780.

Pachuk, C. J., Bredenbeek, P. J., Zoltick, P. W., Spaan, W. J. M., and Weiss, S. (1989). Molecular cloning of the gene encoding the putative polymerase of mouse hepatitis coronavirus, strain A59. *Virology* **171,** 141–148.

Page, K. W., Britton, P., and Boursnell, M. E. G. (1990). Sequence analysis of the leader RNA of two porcine coronaviruses: Transmissible gastroenteritis virus and porcine respiratory coronavirus. *Virus Genes* **4,** 289–301.

Penzes, Z., Tibbles, K., Shaw, K., Britton, P., and Cavanagh, D. (1994). Characterization of a replicating and packaged defective RNA of avian coronavirus infectious bronchitis virus. *Virology* **203,** 286–293.

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 organisation and expression. *Biochimie* **69,** 591–600.

Sethna, P. B., Hofmann, M. A., and Brian, D. A. (1991). Minus-strand copies of replicating coronavirus mRNAs contain antileaders. *J. Virol.* **65,** 320–325.

Snijder, E. J., Den Boon, J. A., Horzinek, M. C., and Spaan, W. J. M. (1991). Comparison of the genome organization of toro- and coronaviruses: Evidence for two nonhomologous RNA recombination events during Berne virus evolution. *Virology* **180,** 448–452.

Soe, L. H., Shieh, C.-K., Baker, S. C., Chang, M.-F., and Lai, M. M. C. (1987). Sequence and translation of the murine coronavirus 5′-end genomic RNA reveals the N-terminal structure of the putative RNA polymerase. *J. Virol.* **61,** 3968–3976.

Tung, F. Y. T., Abraham, S., Sethna, M., Hung, S.-L., Sethna, P., Hogue, B. G., and Brian, D. A. (1992). The 9-kDa hydrophobic protein encoded at the 3′ end of the porcine transmissible gastroenteritis coronavirus genome is membrane-associated. *Virology* **186,** 676–683.

Wesley, R. D., Cheung, A. K., Michael, D. D., and Woods, R. (1989). Nucleotide sequence of coronavirus TGEV genomic RNA: Evidence for 3 mRNA species between the peplomer and matrix protein genes. *Virus Res.* **13,** 87–100.

Wesley, R. D. (1990). Nucleotide sequence of the E2-peplomer protein gene and partial nucleotide sequence of the upstream polymerase gene of transmissible gastroenteritis virus (Miller strain). In "Coronaviruses and Their Diseases" (D. Cavanagh and T. D. K. Brown, Eds.), Vol. 276, pp. 301–306. Plenum, New York.