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 151, 41-49 (1986)

Sequence Analysis of the Porcine Transmissible Gastroenteritis Coronavirus Nucleocapsid Protein Gene

PAUL A. KAPKE[1] AND DAVID A. BRIAN[2]

Department of Microbiology, The University of Tennessee, Knoxville, Tennessee 37996-0845

Received December 12, 1985; accepted January 28, 1986

The 3′ end of the 20-kb genome of the Purdue strain of porcine transmissible gastroenteritis coronavirus (TGEV) was copied into cDNA after priming with oligo(dT) and the double-stranded product was cloned into the *Pst*I site of the pUC9 vector. One clone of 2.0-kb contained part of the poly(A) tail and was sequenced in its entirety using the chemical method of Maxam and Gilbert. Another clone of 0.7 kb also contained part of the poly(A) tail and was sequenced in part to confirm the primary structure of the most 3′ end of the genome. Two potential, nonoverlapping genes were identified within the 3′-terminal 1663-base sequence from an examination of open reading frames. The first gene encodes a 382-amino acid protein of 43,426 mol wt, that is the apparent nucleocapsid protein on the basis of size, chemical properties, and amino acid sequence homology with other coronavirus nucleocapsid proteins. It is flanked on its 5′ side by at least part of the matrix protein gene. The second encodes a hypothetical 78-amino acid protein of 9101 mol wt that is hydrophobic at both ends. A 3′-proximal noncoding sequence of 276 bases was also determined and a conserved stretch of 9 nucleotides near the poly(A) tail was found to be common among TGEV, the mouse hepatitis coronavirus, and the avian infectious bronchitis coronavirus.

INTRODUCTION

The genome of the porcine transmissible gastroenteritis coronavirus (TGEV) has been shown to be a single-stranded, nonsegmented, polyadenylated, infectious RNA molecule of 6.8×10^6 mol wt or approximately 20 kb in length (Brian *et al.*, 1980). The total number of genes encoded by the TGEV genome, however, has not yet been determined. The genome codes for at least four unique polypeptides on the basis of existing protein data. The virion is comprised of three major structural proteins: a 200-kd peplomeric glycoprotein, a 29-kd membrane-associated matrix glycoprotein, and an internal phosphorylated nucleocapsid protein that measures from 46 to 50 kd (Garwes and Pocock, 1975; Moreau and Brian, unpublished). These proteins alone would account for only approximately 8.4 kb of coding information. In addition, the virus synthesizes at least one nonstructural protein during its replication, an RNA-dependent RNA polymerase, the size of which is not yet known (Dennis and Brian, 1982). During replication, TGEV produces nine species of subgenome-size polyadenylated RNA molecules each of which may function as a separate mRNA (Dennis and Brian, 1982), assuming that the 3′ coterminal "nested set" arrangement described for the mRNAs of mouse hepatitis virus (MHV; Lai, *et al.*, 1981; Leibowitz *et al.*, 1982; Rottier *et al.*, 1981) and the avian infectious bronchitis coronavirus (IBV; Stern and Kennedy, 1980) is also true for TGEV. From this information, TGEV may code for as many as 10 different protein species.

One powerful approach for determining the number of potential genes in an RNA virus genome is to examine the primary nucleotide structure and deduce the identity of genes from an examination of open reading frames. In this paper we describe

[1] Present address: National Animal Disease Center, Ames, Ia. 50010

[2] To whom requests for reprints should be addressed.

0042-6822/86 $3.00

experiments that begin to examine the TGEV genome by this approach. cDNA clones were prepared from the 3′ terminal 10% of the polyadenylated genome and were sequenced. Two potential genes were identified within the first 1663 bases. One gene encodes a protein of 382 amino acids which is the apparent nucleocapsid protein on the basis of size, chemical properties, and significant amino acid sequence homology with other coronavirus nucleocapsid proteins. This gene is flanked on its immediate 5′ side by at least part of the matrix protein gene. The second gene lies to the 3′ side of the nucleocapsid protein gene and encodes a hypothetical protein of 78 amino acids that is hydrophobic at both ends. A 3′ noncoding sequence of 276 bases sharing a 9-base conserved sequence near the poly A tail with other coronaviruses was also determined.

MATERIALS AND METHODS

Virus and cells. The Purdue strain of TGEV was plaque purified and grown on the swine testicle (ST) cell line as previously described (Brian *et al.*, 1980).

Purification of genomic RNA. Virus was purified from clarified supernatant fluids as previously described (Brian *et al.*, 1980) except that all sucrose solutions were made up in TMEN (10 m*M* Tris-maleate, pH 6.0, 100 m*M* NaCl, 1 m*M* EDTA). Viral RNA in 1 of 10 flasks was radiolabeled in order to follow the purification of the RNA. For these radiolabeled cultures, infected cells were refed with phosphate-free medium containing 1% fetal calf serum and 40 μCi [^{32}P]orthophosphate (ICN) per milliliter. Viral RNA was extracted by dissolving the virus pellet in 0.5 ml TNE (10 m*M* Tris-hydrochloride, pH 7.5, 100 m*M* NaCl, 1 m*M* EDTA) containing 1% SDS and 0.5 mg proteinase K per milliliter, incubating for 0.5 hr at 37°, and extracting twice with an equal volume of a mixture of 50% phenol/48% chloroform/2% isoamyl alcohol. RNA was ethanol precipitated after adding 0.1 volume 2 *M* sodium acetate. Because small molecular weight RNA species are found in some preparations of purified coronavirion RNA, full-length genomic RNA to be used for cDNA cloning and making probe for colony screening was selected by rate-zonal sedimentation on preformed linear gradients of 15 to 30% sucrose (wt/wt) made up in TNE–0.1% SDS. RNA was dissolved in water and sedimented 1.5 hr at 110,000 *g*, 25°, on 5-ml gradients. Fractions of 0.2 ml were collected and the distribution of radioactivity was determined by Cerenkov counting. Only RNA sedimenting with a sedimentation coefficient of 50 S or greater, as determined by reference to sedimentation of mammalian 28 S and 18 S ribosomal RNA in a parallel gradient, was recovered by ethanol precipitation and used in the experiments described below.

cDNA cloning of the 3′ end of the TGEV genome. TGEV genomic RNA was cloned using a modified method of Gubler and Hoffman (1983). First strand synthesis was carried out in a reaction volume of 50 μl containing 50 m*M* Tris-hydrochloride, pH 8.3, 10 m*M* $MgCl_2$, 10 m*M* DTT, 2 m*M* dCTP, 2 m*M* dTTP, 2 m*M* dATP, 2 m*M* dCTP, 10 μCi [^{32}P]dCTP (3000 Ci/mmol, ICN), 50 pmol oligo $dT_{12\text{-}18}$, 6 μg TGEV RNA, 30 U RNAsin, 10 U reverse transcriptase (Seikagaku), for 1 hr at 42°, and the reaction was stopped by adding 2 μl 0.5 *M* EDTA. Nucleic acids were phenol-chloroform-isoamyl alcohol extracted and ethanol precipitated after the addition of 0.5 vol of 7.5 *M* ammonium acetate.

Second strand synthesis was carried out in a reaction volume of 100 μl containing 20 m*M* Tris-hydrochloride, pH 7.5, 5 m*M* $MgCl_2$, 10 m*M* $(NH_4)_2SO_4$, 100 m*M* KCl, 0.15 m*M* β-NAD, 50 mg/ml BSA, 40 μ*M* dNTPs, 8.5 U/ml *Escherichia coli* RNAse H, 230 U/ml DNA polymerase I, 10 U/ml DNA ligase, and all of the product from the first strand reaction. The reaction was incubated at 12° for 1 hr, then at 22° for 1 hr. The reaction was stopped by adding 4 μl 0.5 *M* EDTA and reaction products were phenol-chloroform-isoamyl alcohol extracted and fractionated on a Sephadex G50 spun column (Maniatis *et al.*, 1982), and the ds cDNA was ethanol precipitated.

Double-stranded cDNA was homopolymer tailed essentially by the method of Roychoudhury and Wu (1980). The follow-

ing were added together, in order: 3 μl dCTP^{32} (>800 Ci/mmol), 20 μl 10× cacodylate buffer (1.4 m*M* potassium cacodylate, 0.3 *M* Tris-hydrochloride, pH 7.6), 4 μl 5 m*M* DTT, 3 μl 10 m*M* dCTP, 2 μl 100 m*M* $CoCl_2$, 12 units terminal deoxynucleotide transferase (PL Biochemicals; at least 8 units/pmol 3′ end) in 1.5 μl, H_2O to 200 μl final volume. The reaction was carried out at 12° for 1.5 min then stopped by adding 10 μl 0.5 *M* EDTA. This reaction resulted in an average of 15 dCMP residues added per 3′ end of dsDNA, the optimal number for annealing and transformation (Peacock *et al.*, 1981).

C-tailed ds cDNA was annealed to G-tailed, *Pst*I-linearized pUC9 vector (PL Biochemicals) for 4 hr at 58° in a 50-μl vol of buffer containing 10 m*M* Tris-hydrochloride, pH 7.5, 150 m*M* NaCl, 1 m*M* EDTA. The total concentration of DNA used was less than 0.5 μg/ml and the optimal insert:vector ratio was 1:1 on a mass basis. *E. coli* strain JM103 was transformed using the method of Hanahan (Hanahan, 1983). Cells containing inserts were observed as white colonies on YT agar plates that contained 100 μg ampicillin/ml, 1 m*M* IPTG, and 0.004% X-gal (Heidecker and Messing, 1983). Recombinant colonies were transferred to nitrocellulose (Millipore, HAWP) and probed with random-primed cDNA copied from TGEV genomic RNA.

Identification of large clones containing 3′-specific TGEV sequences. ^{32}P-labeled, random-primed cDNA used for colony hybridization was synthesized as described above for the oligo(dT)-primed reaction except that 0.2 μg of RNA was used and oligo(dT) was replaced by 20 μg of fragmented calf thymus DNA. Probe was alkali treated to hydrolyze the RNA and then was used for colony hybridization (Maniatis *et al.*, 1982). Colonies yielding a strong signal were analyzed for insert size by electrophoresis of plasmid DNA in agarose gels (Kado and Liu, 1981). Inserts of 0.2 to 2.0 kb (the largest) were further analyzed by Southern hybridization with ^{32}P-labeled poly(dT) to detect poly(dA) content and by cross-hybridization with nick-translated inserts to detect overlapping sequences. ^{32}P-labeled poly(dT) probe was prepared as described above for the oligo(dT)-primed reaction except that 50 pmol oligo(dT)·poly(rA) (PL Biochemicals) replaced the RNA. Alkali-treated ^{32}P-poly(dT) probe was incubated for hybridization at 37° for 12 hr then at 20° for 36 hr, and blots were washed in 2× SSC, 0.1% SDS at 20°.

Restriction endonuclease mapping. Plasmid was purified by lysozyme lysis and cesium chloride centrifugation (Maniatis *et al.*, 1982), and restriction endonuclease mapping was done essentially as described by Smith and Bernstiel (1976) using plasmids that were labeled at the *Sal*I site within the multiple cloning linker region.

DNA sequencing and sequence analysis. Restriction fragments end labeled with ^{32}P were isolated and sequenced by the method of Maxam and Gilbert (1980). Sequences were analyzed with the aid of the program developed by Queen and Korn (1984) and sequence homologies were searched against Genbank, both marketed as part of the Beckman Microgenie program, March 1985 version (Beckman Instruments, Inc.).

RESULTS

cDNA cloning and sequencing of two clones from the 3′ end of the genome. Starting material for cDNA cloning was approximately 6 μg of rate-zonally purified genomic RNA obtained from 400 ml of tissue culture medium. An estimated 200 ng of ds cDNA was obtained, as determined by radiolabel incorporation during second strand synthesis, and from this approximately 2000 white colonies were obtained. By colony screening 200 colonies gave a strong signal to ^{32}P-labeled cDNA prepared from genomic RNA, and of these, 13 had inserts of 200 to 2000 bases as determined by agarose gel electrophoresis of supercoiled plasmids, and were further analyzed by restriction enzyme analysis and poly (A) content. The largest clone of 2000 bases, FG5, did not react by Southern blotting to ^{32}P-labeled oligo(dT), but did cross-hybridize in Southern blot analysis with several other smaller clones that did react strongly with oligo(dT). One of these, J21, a clone of 700 bases, was sequenced in part

to determine the primary structure of the extreme 3′ end of the genome.

The orientation of clones FG5 and J21 in reference to the virus genome and the restriction enzyme sites used for sequencing are illustrated in Fig. 1. Our orientation presumes polyadenylation at only the 3′ end of the genome and this, in turn, is based on the precedent of the documented 3′ polyadenylation site in the avian infectious bronchitis virus and mouse hepatitis virus genomes (Lai *et al.*, 1981; Stern and Kennedy, 1980). The strategy used for sequencing is described in the legend to Fig. 1. Over 96% of the sequence containing the two complete genes we report was determined either by sequencing both strands or by repeated sequencing of the same strand using different methods of end labeling. Some of the sequences were derived from subclones of FG5 made from the *Pst*I restriction sites.

The total sequence of FG5 is illustrated in Fig. 2. Sequences from J21 that overlap with FG5 are identical to those of FG5 except that the total length of the polyadenylate tail is 15 bases for the J21 clone, and 6 for the FG5 clone.

The entire nucleotide sequence was translated in all possible reading frames and only translation of the virus-sense strand revealed open reading frames of greater than 120 bases that are preceded by a termination codon and contain an appropriate initiator methionine codon (Fig. 3). Of these, only the two largest open reading frames are evaluated below.

The largest open reading frame predicts a protein having properties expected of the nucleocapsid protein. The largest open reading frame extends from base 353 to base 1498 and predicts a 382-amino acid protein of 43,426 mol wt. The only TGEV structural or nonstructural protein described to date that approaches this size is the phosphorylated nucleocapsid protein that measures 46 to 50 kd by SDS-polyacrylamide electrophoresis (Garwes and Pocock, 1975; Moreau and Brian, unpublished). The protein has two properties that are strikingly similar to the nucleocapsid proteins of MHV and IBV (Armstrong *et al.*, 1983; Boursnell *et al.*, 1985; Skinner and Siddell, 1983). First, it is rich in serine. Thirty-nine (10%)of the residues are serine making it the most abundant amino acid. Assuming this protein is phosphorylated at serine residues, as is the MHV A59 protein (Stohlman and Lai, 1979), then a high

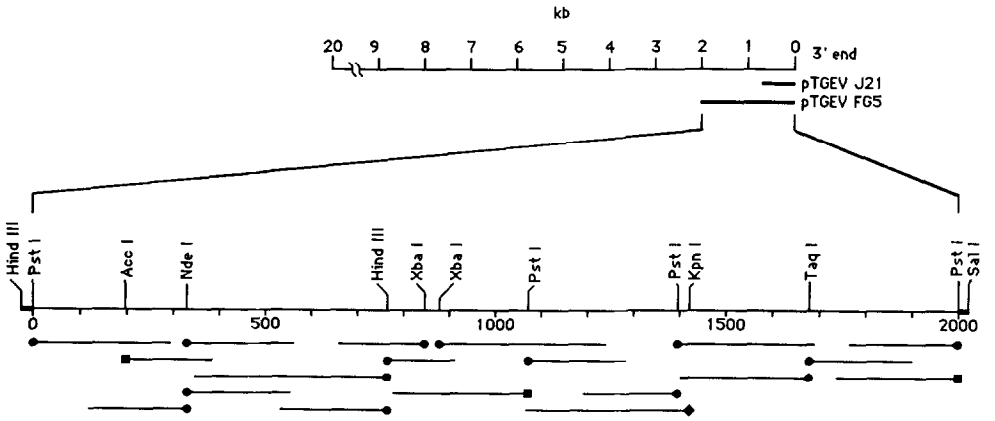

FIG. 1. Restriction endonuclease map and sequencing strategy for TGEV cDNA clones J21 and FG5. The internal *Hin*dIII, *Taq*I, *Pst*I, and *Xba*I sites, derived by restriction endonuclease mapping, and the *Hin*dIII and *Sal*I sites in the multiple cloning region of the pUC9 vector, were the sites used for initial DNA sequencing. Internal *Acc*I, *Kpn*I, and *Nde*I sites were identified from sequence data and were used to complete the sequencing. ■ Indicates sites labeled at the 5′ end using polynucleotide kinase. ● Indicates sites labeled at the 3′ end using reverse transcriptase and the appropriately labeled deoxynucleotide triphosphate. ♦ Indicates site labeled at the 3′ end using dideoxy A and terminal transferase.

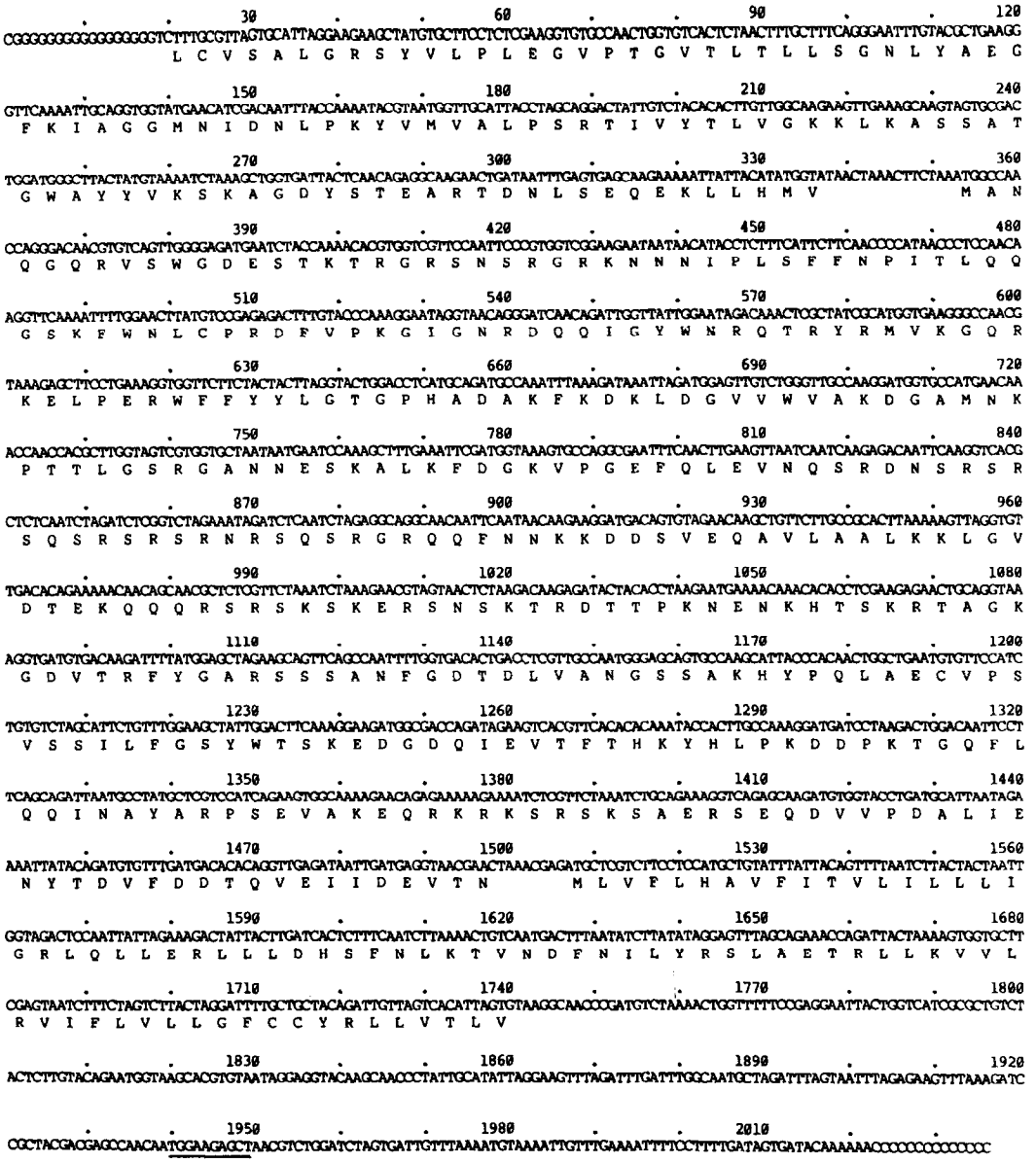

FIG. 2. The primary nucleotide sequence of clone FG5 and the deduced amino acid sequences for a portion of the matrix glycoprotein (bases 20 through 337 in the second reading frame), the nucleocapsid protein (bases 353 through 1498 in the second reading frame), and the hypothetical hydrophobic protein (bases 1507 through 1740 in the first reading frame). A 10-base sequence highly conserved among coronaviruses is underlined beginning at base 1940.

level of phosphorylation might explain the 3- to 6-kd difference between the predicted and measured molecular weights. Second, the protein is basic, a property expected of nucleic acid-binding proteins. Sixty-nine (18%) of the amino acids are basic whereas only 46 (12%) are acidic, giving the protein a net charge of +23 at neutral pH.

Although the consensus sequence around the AUG initiator codon for the TGEV nucleocapsid protein (UAAAUGG) is not among the most favored for translation

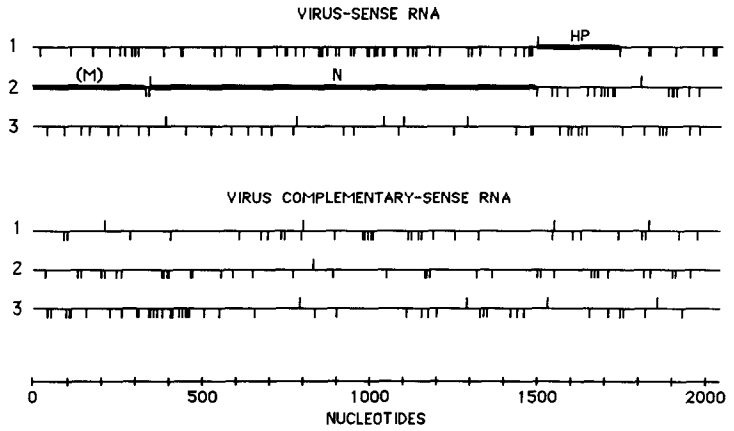

FIG. 3. Schematic diagram of possible open reading frames obtained when translating the FG5 nucleotide sequence as either virus-sense RNA or virus complementary-sense RNA. Vertical bars above the line represent the first methionine codon that could serve as the initiation site for translation. Vertical bars below the line represent termination codons. (M), partial sequence of the matrix protein gene. N, sequence of the nucleocapsid protein gene. HP, sequence of a hypothetical protein gene.

initiation sites that have been described, it is not without precedent (Kozak, 1983). This AUG, therefore, probably identifies the authentic beginning of the TGEV nucleocapsid gene since the sequence from this point leftward to the end of clone FG5, except for a 12-base intergenic sequence, reveals an open reading frame coding for a protein sharing extensive regions of amino acid homology with the small matrix glycoprotein (M or E1) of the mouse hepatitis virus A59 (Armstrong *et al.*, 1984; discussed below).

A second open reading frame to the 3' side of the nucleocapsid protein gene encodes a hypothetical protein of 9101 mol wt that is hydrophobic at both ends. An open reading frame beginning at base 1507 and extending through base 1740 encodes a hypothetical 78 amino acid protein of 9101 mol wt (Fig. 2). A hydrophobicity analysis of the protein reveals that it is hydrophobic for a stretch of approximately 25 amino acids at each end and it is hydrophilic in its central region. There are eight basic amino acids and four acidic amino acids giving the protein a net +4 charge at neutral pH. Basic and acidic amino acids are distributed evenly throughout the central hydrophilic region, but 4 basic amino acids and no acidic ones are found among the 27 amino acids at the carboxy terminus. There is yet no direct evidence for this protein.

DISCUSSION

We present the primary nucleotide sequence for the TGEV nucleocapsid protein (N) gene and the deduced amino acid sequence for the protein. This is the first primary sequence data for a coronavirus in the antigenic subgroup to which TGEV belongs, and such information allows one to first, firmly conclude that TGEV shares an ancestral relationship with MHV and IBV, and second, to identify potentially functional domains on the N protein by examining conserved structures among the diverged viruses. The first two coronavirus N gene sequences to be described are those of the closely related JHM and A59 strains of MHV (Armstrong *et al.*, 1983; Skinner and Siddell, 1983) and between these an overall homology of 94% was found for both the nucleotide and amino acid sequences, reflecting the antigenic similarities between the viruses. Interestingly, the antigenically distinct avian infectious bronchitis virus shows no N gene nucleotide sequence homology with MHV, yet shares an overall amino acid sequence homology of 26% (Boursnell *et al.*, 1985). Furthermore,

there is a stretch of 67 amino acids within the amino terminal one-third of the protein that shows a sequence homology of 51% between the viruses (Boursnell *et al.*, 1985) suggesting that a strong selective pressure exists for a specific functional group defined by this sequence. This interesting pattern repeats itself in the structure of the TGEV N protein. Although TGEV shows no antigenic relatedness to either MHV or IBV (Pedersen *et al.*, 1978), and no N gene nucleotide homology with either MHV or IBV, it does show an overall amino acid homology of 27 and 26% with MHV (JHM) and IBV, respectively. Furthermore, the conserved 67 amino acid region is also found in TGEV (becoming 68 positions when TGEV is compared; Fig. 4). This conserved region is slightly more basic than the overall nucleocapsid protein and therefore may function as a site of interaction with genomic RNA.

Other regions in the N proteins of the three viruses share structural similarities in the absence of a common primary structure suggesting the existence of additional conserved functional domains. Although the N proteins are different lengths (382 amino acids for TGEV, 455 for MHV, and 409 for IBV), when the three are aligned by the 68-amino acid conserved sequence, the following structural similarities are observed. (i) Four cluster groups containing 2–10 serine residues are found in parallel with TGEV amino acid positions 20–40, 150–190, 260–300, and 340–360. Other smaller serine clusters are found in MHV. In all three viruses, regions of 10–40 amino acid stretches can be found that are void of serines. (ii) Three cluster groups of 5–29 basic amino acid residues are found in parallel with TGEV amino acid positions 0–30, 150–260, and 330–350. (iii) A cluster of 9–11 acidic amino acid residues is found within the last 32 amino acids at the carboxy terminus.

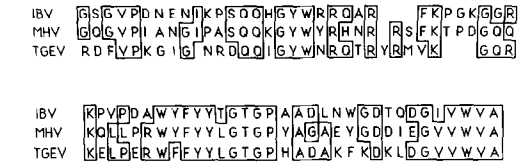

FIG. 4. Amino acid sequence homologies among IBV, MHV, JHM, and TGEV for a 68-amino acid conserved region in the nucleocapsid protein gene. The sequence starts at amino acid 53 from the initiator codon for IBV (Boursnell *et al.*, 1985), amino acid 86 for MHV JHM (Skinner and Siddell, 1983) and amino acid 53 for TGEV. Identical amino acids are boxed in.

Although TGEV would appear to be equally diverged from IBV and MHV on the basis of amino acid sequence, TGEV more closely resembles MHV in its genome arrangement. Firstly, like MHV, the N gene for TGEV is flanked on its 5′ side by the matrix protein (M or E1) gene, whereas for IBV, two genes, derived from overlapping reading frames and encoding hypothetical proteins of unknown function, lie between the M and N genes (Armstrong *et al.*, 1984; Boursnell and Brown, 1984). Our conclusion that the M gene for TGEV lies to the immediate 5′ side of the nucleocapsid gene is based on amino acid sequence homology with the M protein of MHV A59. Of the 105 amino acids deduced for the TGEV matrix protein sequence, 31% are perfectly homologous and another 15% are conservative differences (Fig. 2 and data not shown). Secondly, the number of nucleotides separating the M (E1) and N genes is close, 14 for MHV and 12 for TGEV, and these match perfectly for a stretch of 8 bases:

```
MHV A59 T C T A A A C T T T A A G G
TGEV       C T A A A C T T C T A A
```

Since part of this sequence may play a role in primer recognition for transcription (Brown and Boursnell, 1984; Budzilowicz *et al.*, 1985), some common features between the leader molecules of MHV and TGEV may be anticipated.

No direct evidence exists for the hydrophobic protein encoded by base 1507 through 1740. Genes encoding small hydrophobic proteins in MHV and IBV have been described, however (Boursnell and Brown, 1984; Skinner *et al.*, 1985), but their hydrophobicity is only at one end, they map at an entirely different region in the genome, and no sequence homology is found between them and the TGEV hydrophobic

protein. Regarding this open reading frame, it is noteworthy that three small polyadenylated, putative messenger RNAs have been identified in TGEV-infected cells that have not been reported for MHV or IBV (Dennis and Brian, 1982). Assuming TGEV replicates by the consensus scheme proposed for the replication of MHV and IBV, namely that all messages have a 3'-coterminal nested set arrangement (as suggested by preliminary experiments with TGEV [Hu *et al.*, 1984]), then one of the small messages described by Dennis and Brian may be the message for the hydrophobic protein. From the known sequence (Fig. 2), such a message would be 0.20 Md. Two structural features favor the plausability of this being a functional hydrophobic protein gene. (i) The intergenic sequence preceding the gene, inclusive of the N gene stop codon, contains a 6-base sequence, CTAAAC, that is in common with part of the intergenic sequence preceding the N gene for both MHV and TGEV described above, and may play a role in the initiation of mRNA transcription (Budzilowicz *et al.*, 1985). (ii) The 7-base sequence, GAGAUGC, at the initiation site of the hydrophobic protein is a favored pattern among eukaryotic initiation sequences (Kozak, 1983).

Assuming that the gene for the hydrophobic protein is real then the 3' terminal noncoding sequence would be a total of 276 bases, exclusive of the poly (A) tail, and would be the shortest noncoding sequence of those identified for coronaviruses. The significance of the noncoding region is not completely known although it undoubtedly functions as an attachment region for the polymerase to initiate synthesis of the negative strand RNA. One possible site that may be critical for recognition or binding is a 10-base sequence, GGGAAGAGCT, that is conserved between IBV (found 81 bases from the 3' end) and MHV (82 bases from the end). With the exception of the first base, a T instead of G, TGEV shares an identical sequence beginning 77 bases from the 3' end (Fig. 2).

ACKNOWLEDGMENTS

We thank William Lapps for many helpful discussions. This work was supported by Public Health Service Grant AI-14367 from the National Institute of Allergy and Infectious Diseases, by Grant 82-CRSR-2-1090 from the U.S. Department of Agriculture, and in part by a grant from the National Foundation for Ileitis and Colitis, Inc.

REFERENCES

ARMSTRONG, J., NIEMANN, H., SMEEKENS, S., ROTTIER, P., and WARREN, G. (1984). Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. *Nature (London)* **308,** 751–752.

ARMSTRONG, J., SMEEKENS, S., and ROTTIER, P. (1983). Sequence of the nucleocapsid gene from murine coronavirus MHV-A59. *Nucleic Acids Res.* **11,** 833–891.

BOURSNELL, M. E. G., BINNS, M. M., FOULDS, I. J., and BROWN, T. D. K. (1985). Sequences of the nucleocapsid genes from two strains of avian infectious bronchitis virus. *J. Gen. Virol.* **66,** 573–580.

BOURSNELL, M. E. G., and BROWN, T. D. K. (1984). Sequencing of coronavirus IBV genomic RNA: A 195-base open reading frame encoded by mRNA B. *Gene* **29,** 87–92.

BRIAN, D. A., DENNIS, D. E., and GUY, J. S. (1980). Genome of porcine transmissible gastroenteritis virus. *J. Virol.* **34,** 410–415.

BROWN, T. D. K., and BOURSNELL, M. E. G. (1984). Avian infectious bronchitis virus genomic RNA contains sequence homologies at the intergenic boundaries. *Virus Res.* **1,** 15–24.

BUDZILOWICZ, C. J., WILCZYNSKI, S. P., and WEISS, S. R. (1985). Three intergenic regions of coronavirus mouse hepatitis virus strain A59 genome RNA contain a common nucleotide sequence that is homologous to the 3' end of the viral mRNA leader sequence. *J. Virol.* **53,** 834–840.

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

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

GUBLER, U., and HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25,** 263–269.

HANAHAN, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166,** 557–580.

HEIDECKER, G., and MESSING, J. (1983). Sequence analysis of zein cDNAs obtained by an efficient mRNA method. *Nucleic Acids Res.* **11,** 4891–4906.

HU, S., BRUSZEWSKI, J., BOONE, T., and SOUZA, L. (1984). Cloning and expression of the surface glycoprotein gp195 of porcine transmissible gastroenteritis virus. *In* "Modern Approaches to Vaccines" (R. M. Chanock and R. A. Lerner, eds.), pp. 219–223. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

KADO, C. I., and LIU, S. T. (1981). Rapid procedure for

detection and isolation of large and small plasmids. *J. Bacteriol.* **145,** 1365–1373.

KOZAK, M. (1983). Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev.* **47,** 1–45.

LAI, M. M. C., BRAYTON, P. R., ARMEN, R. C., PATTON, C. D., and STOHLMAN, S. A. (1981). Mouse hepatitis virus A59 messenger RNA structure and genetic localization of the sequence divergence from the hepatotropic strain MHV 3. *J. Virol.* **39,** 823–834.

LEIBOWITZ, J. L., WEISS, S. R., PAAVOLA, E., and BOND, C. W. (1982). Cell-free translation of murine coronavirus RNA. *J. Virol.* **43,** 905–913.

MANIATIS, T., FRITSCH, E. F., and SAMBROOK, J. (1982). "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MAXAM, A. M., and GILBERT, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *In* "Methods in Enzymology" (L. Grossman and K. Moldave, eds.), Vol. 65, pp. 499–560. Academic Press, Orlando, Fla.

PEACOCK, S. L., MCIVER, C. M., and MONOHAN, J. J. (1981). Transformation of *E. coli* using homopolymer-linked plasmid chimeras. *Biochim. Biophys. Acta* **655,** 243–250.

PEDERSEN, N. C., WARD, J., and MENGELING, W. L. (1978). Antigenic relationship of the feline infections peritonitis virus to coronaviruses of other species. *Arch. Virol.* **58,** 45–53.

QUEEN, C., and KORN, L. J. (1984). A comprehensive sequence analysis program for the IBM personal computer. *Nucleic Acids Res.* **12,** 581–599.

ROTTIER, P. J. M., HORZINEK, M. C., and VAN DER ZEIJST, B. A. M. (1981). Translation of three mouse hepatitis virus (MHV-A59) subgenomic RNAs in *Xenopus laevis* oocytes. *J. Virol.* **38,** 20–26.

ROYCHOUDHURY, R., and WU, R. (1980). Terminal transferase-catalyzed addition of nucleotides to the 3′ termini of DNA. *In* "Methods of Enzymology" (L. Grossman and K. Moldave, eds.), Vol. 65, pp. 43–62. Academic Press, Orlando, Fla.

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.

SKINNER, M. A., and SIDDELL, S. G. (1983). Coronavirus JHM: Nucleotide sequence of the mRNA that encodes nucleocapsid protein. *Nucleic Acids Res.* **11,** 5045–5054.

SMITH, H. O., and BIRNSTIEL, M. L. (1976). A simple method for DNA restriction site mapping. *Nucleic Acids Res.* **3,** 2387–2398.

STERN, D. F., and KENNEDY, S. I. T. (1980). Coronavirus multiplication strategy. II. Mapping the avian infectious bronchitis virus intracellular RNA species to the genome. *J. Virol.* **36,** 440–449.

STOHLMAN, S. A., and LAI, M. M. C. (1979). Phosphoproteins of murine hepatitis virus. *J. Virol.* **32,** 672–675.