

Sequence Analysis of the Leader RNA of Two Porcine Coronaviruses: Transmissible Gastroenteritis Virus and Porcine Respiratory Coronavirus

KEVIN W. PAGE,¹ PAUL BRITTON,¹ AND MICHAEL E. G. BOURSNELL²
A.F.R.C., Institute for Animal Health, ¹Compton Laboratory, Compton, Newbury, Berkshire;
²Houghton Laboratory, Houghton, Huntingdon, Cambridgeshire

Received January 18, 1990

Accepted March 14, 1990

Requests for reprints should be addressed to Paul Britton, A.F.R.C., Institute for Animal Health, Compton Laboratory, Compton, Newbury, Berkshire RG16 0NN, United Kingdom

Key words: transmissible gastroenteritis virus, porcine respiratory coronavirus, leader sequence, primer extension

Abstract

The leader RNA sequence was determined for two pig coronaviruses, transmissible gastroenteritis virus (TGEV), and porcine respiratory coronavirus (PRCV). Primer extension, of a synthetic oligonucleotide complementary to the 5' end of the nucleoprotein gene of TGEV was used to produce a single-stranded DNA copy of the leader RNA from the nucleoprotein mRNA species from TGEV and PRCV, the sequences of which were determined by Maxam and Gilbert cleavage. Northern blot analysis, using a synthetic oligonucleotide complementary to the leader RNA, showed that the leader RNA sequence was present on all of the subgenomic mRNA species. The porcine coronavirus leader RNA sequences were compared to each other and to published coronavirus leader RNA se-

The nucleotide sequence data reported in this paper have been submitted to the EMBL/Genbank/ DDBJ nucleotide sequence databases and have been assigned the accession numbers X52157, X52668.

quences. Sequence homologies and secondary structure similarities were identified that may play a role in the biological function of these RNA sequences.

Introduction

Transmissible gastroenteritis virus (TGEV) and porcine respiratory coronavirus (PRCV) belong to the family Coronaviridae, a large group of pleomorphic enveloped viruses with a positive-stranded RNA genome. TGEV causes gastroenteritis in pigs, resulting in a high mortality in neonates (1). PRCV was isolated in several European countries between 1984 and 1986 (2-4), does not cause diarrhea, and has been shown to replicate in the respiratory tract with little or no clinical signs, but is very similar antigenically and serologically to TGEV (2,4). Virions from both viruses contain two envelope glycoproteins of relative molecular mass (M_r) 200,000 (spike) and M_r 28,000-31,000 (membrane protein) and a phosphorylated nucleoprotein of M_r 47,000. cDNA probes to the structural protein genes of TGEV hybridized to the appropriate mRNA species of PRCV, suggesting a high degree of homology at the RNA level (unpublished data).

Coronavirus proteins are expressed from a "nested" set of subgenomic mRNAs with common 3' termini but different 5' extensions. The sequence of each mRNA that is translated to produce viral proteins appears to correspond to the 5'-terminal region that is absent on the preceding smaller mRNA species. It has been shown for the coronaviruses, mouse hepatitis virus (MHV) and infectious bronchitis virus (IBV), the subgenomic mRNA species possess short "leader sequences" at their 5' ends. These sequences are not transcribed as a contiguous mRNA species, but are derived from the 5' end of the genomic RNA and are probably joined to the 5' end of each mRNA by a process of discontinuous transcription (5-9). The leader sequence appears to be produced by a mechanism termed *leader-primed transcription*, in which the leader RNA is transcribed independently, dissociated from the template, and then binds to the template (negative-sense strand) at specific transcriptional start sites (10,11). The mechanism appears to involve the recognition of consensus sequences identified on the genomic RNA at those points corresponding to the 5' ends of the subgenomic mRNAs. These consensus sequences may act as a binding site for the RNA polymerase-leader complex (7-9, 12-14). It has been previously postulated that a heptameric sequence, ACTAAAC (15-17), or a hexameric sequence, CTAAAC (18-20), may be involved in the binding of the TGEV RNA polymerase leader.

In this paper we describe the elucidation of the leader RNA sequences from the porcine coronaviruses TGEV and PRCV, the first leader sequence to be described from the TGEV serogroup of coronaviruses. Comparison of the leader RNAs of TGEV and PRCV with published leader RNAs of other coronaviruses was used to identify areas of conserved sequence and potential secondary structure that may be involved in the transcription of coronavirus subgenomic mRNA species.

Materials and Methods

Preparation of viral RNA

Confluent cultures of a pig kidney cell line LLC-PK1 were infected with a virulent British field isolate of TGEV strain FS772/70 or a British isolate of PRCV strain 86/137004 at a MOI of 1–10 PFU per cell. After 2 hr at 37°C, the inoculum was removed and replaced with medium containing 1 µg/ml actinomycin D to inhibit host-cell RNA synthesis (21). After a further 2-hr incubation, 25 µCi of [5,6-³H]uridine (Amersham International plc, TRK.410, 35-50 Ci/mM) was added per culture bottle and the cells were incubated for a further 5 hr. The cells were lysed with guanidinium thiocyanate, the RNA pelleted through 5.7 M cesium chloride and poly(A)-containing RNA isolated by poly(U) Sepharose affinity chromatography, as described previously (21).

Synthesis of oligonucleotide primers

Two oligonucleotides were synthesized by the phosphoramidite method using an Applied Biosystem 381A synthesizer. One oligonucleotide, oligo 38 (5'-TGGATT-CATCCCCCAACTA-3'), was complementary to the nucleoprotein gene 22 bp downstream from the initiation ATG codon (15), as shown in Fig. 1, and was used for primer extension. The second oligonucleotide, oligo 58 (5'-AGAGATA-TAGCCACGCTACTACTTTAC-3'), was complementary to the 5' end of the leader RNA (Fig. 1) and was used for Northern blot analysis of viral mRNA.

Primer extension of nucleoprotein mRNA

Gel-purified oligo 38 (500 ng) was 5'-end-labeled (22) using 20 U of T₄ polynucleotide kinase (Gibco-BRL, Paisley) and 20 µCi [γ -³²P]ATP (Amersham International plc, PB 10168, 3000 ci/mM). Poly(A)-containing RNA (1.5 µg) isolated from TGEV- and PRCV-infected cells was resuspended in water and heated at 60°C for 3 min. A further incubation was carried out using the two mRNA preparations in 27 µl reaction volumes containing 40 U of RNasin (Promega Biotec, Liverpool), 50 mM Tris-HCl (pH 8.3), 10 mM MgCl₂, 35 mM KCl, 30 mM 2-mercaptoethanol, 3 mM dithiothreitol, 4 mM dNTPs, 5'-end-labeled oligo 38 (120 ng), and 21 U of AMV reverse transcriptase (Super-RT, Anglian Biotech Ltd, Colchester) for 90 min at 42°C. Formamide dye (80% formamide, 10 mM NaOH, 1 mM EDTA, 0.1% xylene cyanol blue, 0.1% bromophenol blue) was added and the mixture boiled for 3 min and electrophoresed on a 40 cm buffer gradient sequencing gel (23). The wet gel was autoradiographed for 1 hr to locate the primer-extended products, which were excised from the gel. The labeled fragments were

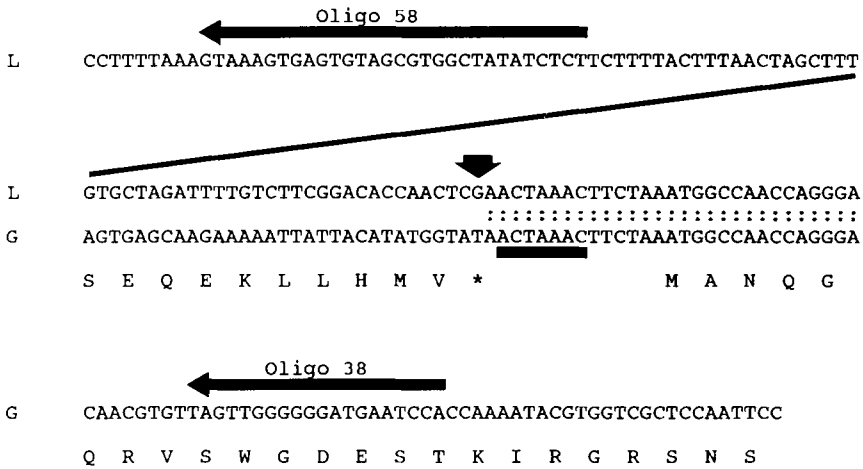


Fig. 1. Alignment of sequences from the 5' end of TGEV nucleoprotein mRNA (L) and the corresponding region of the TGEV genome (G). The positions of homology between the genome sequence and the leader sequence are shown; the point of divergence between the sequences is identified by an arrow. The positions of the synthetic oligonucleotides oligo 38, used for primer extension, and oligo 58, used to probe viral RNA for the presence of the leader RNA sequence, are shown. The direction of the arrows indicate that both of the oligonucleotides were the complement of the sequences shown. The amino acid sequences below the genomic sequence represent the carboxyl terminus of the membrane protein (16) and the amino terminus of the nucleoprotein (15). The postulated TGEV RNA polymerase-leader complex binding site or consensus sequence is underlined.

eluted from the polyacrylamide gel and chemically cleaved (24). Samples of the cleaved products from each of the primer extended products were electrophoresed on 6% polyacrylamide gels at 35 W constant power for two different lengths of time.

Northern Blot Analysis

TGEV and PRCV poly(A)-containing RNA was glyoxylated and separated on a 1% agarose gel (22). The RNA was transferred onto Biodyne A membranes (Pall P/N BNNG3R 1.2 μ m, Gallenkamp) in X20 SSC (X1 SSC = 0.15 M NaCl, 0.015 M trisodium citrate, pH 7.0) for 18 hr and baked at 80°C for 2 hr. The membrane was boiled in 50 mM Tris-HCl pH 8.0 for 5 min to remove glyoxal groups from the RNA and prehybridized in the presence of 50% formamide for 6 hr at 42°C (15). The viral mRNA species were hybridized with 32 P-labeled oligo 58 in the presence of 50% formamide for 18 hr at 42°C. The membrane was washed four times in X2 SSC containing 0.1% NaDodSO₄ for 15 min at room temperature and autoradiographed.

Results

Sequence of the nucleoprotein mRNA leader

Following primer extension, using oligo 38 at the 5' end of the nucleoprotein gene from the porcine coronaviruses TGEV and PRCV, labelled fragments of approximately 140 bases were produced and purified from gels. Larger molecular weight species were also observed (data not shown) in minor amounts, presumably corresponding to read-through sequences upstream of the nucleoprotein gene primed from the larger mRNA species. The nucleotide sequences of the two fragments, determined by chemical cleavage, were identical. The resulting nucleotide sequence of the TGEV leader RNA sequence is shown in relation to the TGEV nucleoprotein gene in Fig. 1. The leader RNA sequence diverges from the genomic sequence 15 bp upstream of the nucleoprotein gene, corresponding to the first nucleotide of the membrane protein gene stop codon (16), indicating a length of 91 nucleotides of unique sequence (Fig. 1). The 91 nucleotide leader sequence of TGEV and PRCV has a low content of G (18%) and C (20%), and a high A (22%) and T (40%) content, with 20% of the T residues grouped in three- to four-nucleotide motifs (Fig. 1). These values are similar to those observed from the TGEV genome so far sequenced, except that the values for A (30.5%) and T (32.1%) are more similar on the genome than on the leader sequence.

Analysis of the TGEV nucleoprotein nucleotide sequence (15) revealed a potential RNA polymerase-leader complex binding site. The site, ACTAAAC, is seven nucleotides upstream of the nucleoprotein initiation codon and has also been found to precede all the TGEV structural protein genes and two of the three potential genes shown to be at the 5' end of mRNA species (15–17). This consensus sequence is found two nucleotides downstream of the nucleotide where the leader RNA and TGEV genomic sequences diverge, indicating that this sequence is involved in the leader-primed transcription of TGEV mRNA molecules. As can be seen from Fig. 2, 4 of the 6 mRNA species from the FS772/70 strain of TGEV have the sequence AACTAAAC, of which the 5'-end adenosine residue is the next base down from the divergence point. In fact, the consensus sequence at the spike/ORF1-ORF2 gene junction has the sequence GAACTAAAC and at the NUC/ORF4 gene junction has the sequence CGAACTAAAC, indicating that the region of the leader sequence 5' to the homology motif, ACTAAAC, may vary between 89 and 91 nucleotides depending on the TGEV gene.

Computer analysis has also detected a homology between the leader RNA sequence and the 5' end of the negative strand (i.e., the reverse complement of the noncoding region at the 3' end of the positive strand). This is shown in Fig. 3. The nucleotides on the leader RNA sequence, bases 84–99, and on the negative strand, bases 136 to 152 counting from the first base after the poly(A) tail, have an overall homology of 82% and include the sequence CTAAC, which is part of the postulated TGEV RNA polymerase-leader complex binding site. This is

Gene Junction	Intergenic Sequences
POL/SPIKE	TAAGT <u>ACTAAACT</u> TTGGTAACCACTTCGTAAACACCA <u>CTG</u>
Spike/ORF1-ORF2	TTAAGA <u>ACTAAACT</u> TTCAAGTCATTACAGGTCCTG <u>TATG</u>
ORF2/ORF3	GGCGGTTCTAAACGAAATTGACTTAAAGAAGAAGAGGGAGACCGTACCT <u>ATG</u>
ORF3/MEM	GTTTGA <u>ACTAAACAAA</u> ATG
MEM/NUC	GGTATA <u>ACTAAACT</u> TCTAAATG
NUC/ORF4	TAACGA <u>ACTAAACG</u> ATG
LEADER	ACTCGA <u>ACTAAAC</u>
	: :
	87 99

Fig. 2. Comparison of the TGEV (strain FS772/70) gene junctions with the sequences immediately 5' of the consensus sequence and part of the TGEV leader RNA sequence. The positions of the ACTAAAC consensus sequences and any identical bases 5' to this sequence present on the leader RNA sequence are double underlined. The initiation codon of the gene immediately downstream of the consensus sequence is underlined. The sequences of the spike/ORF1-ORF2 and ORF2/ORF3 junctions are taken from (17); the ORF3/MEM junction is taken from (16); and the MEM/NUC and NUC/ORF4 junctions are taken from (15). The POL/SPIKE junction sequence is from unpublished work. POL = polymerase; MEM = M, 29 459 glycoprotein; NUC = nucleoprotein.

very similar to the observation for IBV (25) involving sequences present at the 5' end of the IBV genome, and on the IBV leader RNA sequences, with the 5' end of the IBV negative strand. The homology observed included the sequence CTTAAC, which is part of the postulated IBV RNA polymerase-leader complex binding site CT(T/G)AACAA.

Northern blot analysis of TGEV and PRCV mRNA subgenomic species

An oligonucleotide, oligo 58, was synthesised that was complementary to the 5' end of the TGEV and PRCV leader RNA sequences (Fig. 1). The oligonucleotide

TGEV GENOME (-)	115	AAATTACTAAA..TCTAGCATTG.CCAAATCAAATCTAAAC	152
		: : : : : : : : : : : :	
TGEV LEADER	62	TGCTAGATTTT.GTCTTCGGACA.CCAACTCGAA. <u>CTAAAC</u>	99

Fig. 3. Comparison between the reverse complement of the 3' end of the TGEV genome (i.e., the 5' end of the negative strand (15) and part of the TGEV leader RNA sequence. Colons show identical bases. The single dots in the sequences are padding characters inserted to achieve optimal alignment. Part of the postulated TGEV RNA polymerase-leader complex binding site sequence or consensus sequence is double underlined.

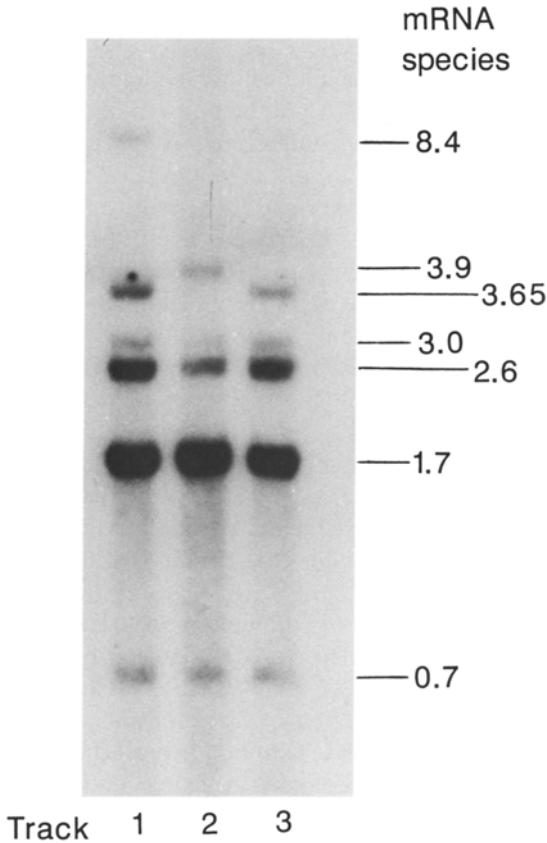


Fig. 4. Autoradiograph of the mRNA species from two strains of PRCV and TGEV northern blotted and probed with the 5' end-labeled synthetic oligonucleotide, oligo 58 (see Fig. 1). Track 1 PRCV is strain 86/135308, track 2 is TGEV strain FS772/70, and track 3 is PRCV strain 86/137004. The ORF1/ORF2 mRNA species from PRCV migrates faster than the corresponding TGEV species, relating to a size of 3.65 kb (34).

was end-labeled and used to probe TGEV and PRCV mRNA species that were Northern blotted onto Biodyne membranes. As can be seen from Fig. 4, the labeled probe hybridized to all of the TGEV and PRCV mRNA species. The intensity of the bands corresponding to labeled probe hybridized the spike mRNA species, and genomic RNA was lower than that observed for the smaller mRNA species due to less of these larger species being isolated from the poly(U) Sepharose column used in the isolation of mRNA. The fact that the probe hybridized to all of the mRNA species showed that the leader RNA sequence was present on the other RNA molecules of TGEV and both strains of PRCV was not unique to the nucleoprotein mRNA species.

Homology to other coronavirus leader RNA sequences

The two porcine coronavirus leader sequences were identical, indicating that the two viruses probably use the same RNA polymerase-leader complex binding site, ACTAAAC, for the synthesis of subgenomic mRNA species. The SEQHP comparison program of the Los Alamos (26) package was used to compare the leader RNA sequences determined in this paper and those published for five other coronaviruses belonging to two different serogroups. The sequences were compared from the 5' ends to the point of divergence from the genomic sequences. The percentage homologies, Table 1, were expressed as the number of bases matched to the longer of the two sequences being compared. The homology of the leader sequences fell into three groups. Leader RNAs from coronaviruses belonging to different serological groups had homologies in the region of 35–40%. Serologically related viruses like human coronavirus (HCV) (strain OC43) and MHV (strains A59 and JHM) have about 60% homology. The third group involved different strains of MHV, A59, and JHM, which showed a homology of 91%. This observation indicates that TGEV and PRCV, which have a homology of 100%, are probably different strains of the same virus or that PRCV has very recently diverged from TGEV.

In order to identify common areas of homology, the leader RNA sequences from seven coronaviruses were aligned. As can be seen from Fig. 5, these fell into two groups. One group consists of MHV (strains A59 and JHM) with HCV (OC43), which have a fairly high degree of homology along their lengths. The other group consists of TGEV and PRCV (not shown on the diagram) with HCV (229E) and IBV, which have high homologies at their 3' ends and areas of homology at their 5' ends. There are good homologies towards the 3' ends, involving the postulated RNA polymerase-leader complex binding sites and sequences upstream of these sites, between the groups, but very little if any homology between the 5' ends.

Table 1. Comparison of coronavirus leader RNA sequences

	Percentage homology of leader RNA sequences ^a					
	PRCV (86/137004)	HCV (229E)	HCV (OC43)	MHV (A59)	MHV (JHM)	IBV (Beaudette)
TGEV (FS772/70)	100	36	39	35	39	40
HCV (229E)		100	45	42	41	52
HCV (OC43)			100	62	67	38
MHV (A59)				100	91	49
MHV (JHM)					100	44
IBV (Beaudette)						100

^aThe percentage homologies were determined from the 5' end of the leader RNA sequences to the points of divergence from the genomic RNA sequence. All values are calculated using the longest sequence of the pairs.


```

229E  CTAAAGT.....ACCTTATCTATCTACAAATAG. AAAAGTTGC...TTT.T..TAGACTTTGTGTCTACTT.....CTAAAC
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
TGEV  AGT.GAGGTAG...CGTGGCTATATCTCTCTTTTACTTTAACTAGC...TTTGTGCTAGATTTTGT.....CTTCGGACACCAACTCGAACTAAAC
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
IBV   ACTTAAAG.ATAGATATTAATATATATCT.....ATTAC....ACTAGC...CTTGGCTAGATTTTTAA.....CTT.....AAC.AAAA
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OC43  TTGTGAGCGAAGTTG.CGTGCGTGCAT.CCCGCTCCACTGAT...CTCTTGTAGATCTTTTT.....CTAAT...CTAA.T....CTAAAT
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
A59   TATAAGAGTGATTGGCGTCCGCTACGTACCCTCT.CAACTCTAAAACCTTTG.TAG.TTTAAAT.....CTAAT.....CTAAAC
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
JHM   TATAAGAGTGATTGGCGTCCGCTACGTACCCTCT.CTACTCTAAAACCTTTG.TAG.TTTAAAT.....CTAAT...CTAA.T....CTAAAC
      **  **  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

```

Fig. 5. Alignment of leader RNA sequences from nucleotide 14 in the TGEV sequence with those from other coronaviruses. PRCV is not included on the diagram, as it is identical to TGEV. Colons between pairs of sequences show identical bases. The top row of asterisks marks positions in which four or more bases are identical, and the bottom row of asterisks marks bases completely conserved between all seven sequences. The dots within the sequences are padding characters inserted to achieve optimal alignment. The sequences of the leader RNA sequences were taken from TGEV (FS772/70) and PRCV (86/137004), this paper; HCV, strain 229E (31) and strain OC43 (35); MHV, strain A59 (7) and strain JHM (13); avian, IBV strain Beaudette (9,25).

Prediction of secondary structures

As seen from Fig. 5 simple alignment did not reveal very much information about the homologies of the leader RNA sequences from the different coronaviruses, except at the 3' ends involving the consensus sequences. In order to identify any potential similarities in these sequences, the secondary structure of the RNA sequences in Fig. 5 were analyzed. Potential secondary structures of the leader RNA sequences were determined using the computer program FOLD (27) from the UWGCG DNA analysis programs (28). The coordinates determined by the FOLD program were displayed graphically using the UWGCG program SQUIGGLER. The potential secondary structures obtained were compared and, as can be seen from Fig. 6, the overall shape of these sequences are very similar, except for the avian coronavirus IBV. All the molecules appear to be composed of two stem-loop structures. The two MHV molecules are very similar in shape and, as seen from Fig. 5 and Table 1, are very homologous, 91%, at base sequence. The secondary structures of the coronavirus leader RNA sequences are probably influenced by their biological function, which results in the similarity of these potential structures.

Discussion

This paper presents evidence that the nucleoprotein mRNA species of TGEV and the closely related porcine respiratory variant of TGEV, PRCV, contain an identical leader RNA sequence of about 91 nucleotides. Sequencing studies on TGEV have shown that the heptameric sequence ACTAAAC occurs on the genome upstream of the genes and is believed to be the binding site for the leader

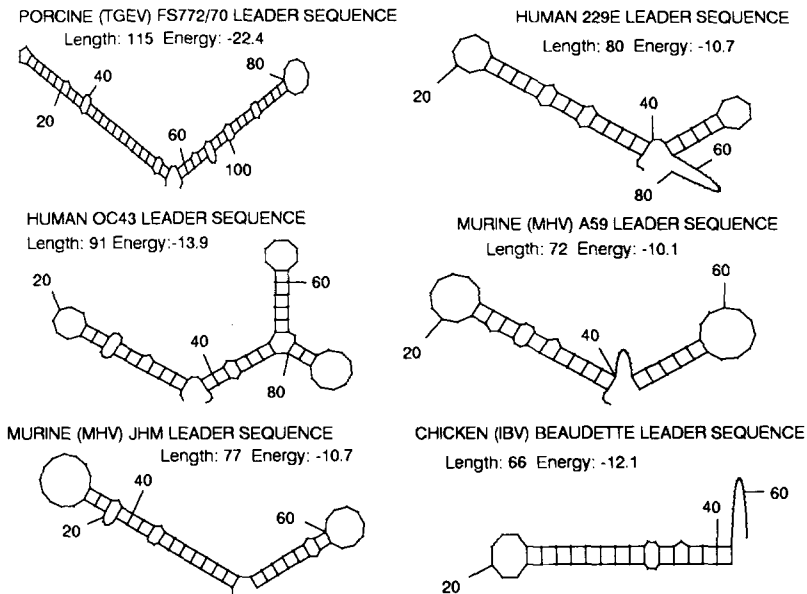


Fig. 6. Comparison of the predicted secondary structures of the coronavirus leader RNA sequence as determined by the UWGCG programs FOLD and SQUIGGLES (28). The leader RNA sequences used were as described in Fig. 5. The numbers refer to the positions of the nucleotides.

RNA primer. Northern blot analysis, Fig. 4, of TGEV strain FS772/70 and PRCV strains 86/137004 and 86/135308 RNA, probed with a complementary oligonucleotide to the leader RNA identified on the nucleoprotein mRNA species of these coronaviruses, showed that all of the subgenomic mRNA molecules contained the common leader RNA sequence. These observations support the finding observed for MHV that the synthesis of coronavirus mRNA species is primed from the negative-sense strand using a small RNA molecule derived from the 5' end of the genomic RNA. This mechanism has been termed *leader-primed transcription* and involves not only the leader RNA primer, but also consensus sequences along the genome found upstream of the genes, which act as binding sites for the leader RNA primer.

Comparison of TGEV and PRCV viral products has shown very little difference between the two coronaviruses, and until recently it was impossible to differentiate between the two viruses using antisera. PRCV is fully neutralized by antisera prepared against TGEV, and the majority of monoclonal antibodies (MAbs) raised against TGEV virion proteins cross-react with PRCV. However, MAbs, raised against antigenic determinants of the spike protein from either the virulent British isolate FS772/70 (29) or the avirulent Purdue strain of TGEV (30) have been identified that do not recognize PRCV. These observations and the fact that the leader RNA sequences from TGEV and PRCV are identical supports the evidence

that the two viruses are very similar and that PRCV may have evolved as a TGEV variant.

Comparison of the TGEV leader RNA sequence with the genomic sequence upstream of the nucleoprotein indicates that the length of the unique sequence of the leader sequence is 91 nucleotides. The point of divergence is two bases upstream of the ACTAAAC sequence, supporting the evidence that the TGEV RNA polymerase-leader complex binding site is ACTAAAC. Four out of the six mRNA species from the FS772/70 strain of TGEV have the sequence AACTAAAC, and the 5'-end adenosine residue is the next base down from the divergence point in the nucleoprotein mRNA (Fig. 2). The differences in the homologies between the leader RNA and sequences upstream of the consensus sequence on the genomic RNA may play a role in the levels of transcription of a particular mRNA species. The mRNA species of 3.0 kb has been shown to have an open reading frame at the 5' end encoding a potential polypeptide of M_r 9200 (17). This particular mRNA does not have the heptameric consensus sequence but has the hexameric CTAAAC sequence, and it is interesting to note that it is the least abundant TGEV mRNA species (observed from TGEV mRNA in total cell lysates). Hybridization of oligo 58 to the 3.0-kb mRNA species showed that this species does contain the TGEV leader RNA, confirming that it is a true mRNA species, even though it is the only TGEV species not to have the heptameric consensus sequence.

Comparison of the seven coronavirus leader RNA sequences against each other identified three groups (Table 1): non-serologically related viruses had about 35–40% homology; serologically related viruses had about 60% homology; viral strains had about 90–100% homology. However, TGEV and HCV (229E) have been placed in the same serological group, but have only 36% homology within their leader RNA sequences, suggesting that the two viruses are not particularly related. TGEV and HCV (229E) have been shown to have 46% homology at the amino acid level within their derived nucleoprotein sequences (31), whereas the homology between the derived nucleoprotein amino acid sequences for different viruses within the MHV serological group are between 80% and 98% homology. This indicates that the serological grouping of coronaviruses is not a particularly useful test, as similar epitopes may exist on the viral structural proteins. Comparisons of nucleic and amino acid sequences from the viruses will provide a more accurate method for grouping the viruses. It will be interesting to compare the leader sequences of bovine coronavirus (BCV), which is serologically related to HCV (OC43) and MHV (A59 and JHM), with feline infectious peritonitis virus (FIPV) and canine coronavirus (CCV), which are serologically related to TGEV, once their sequences have been determined.

The large variation in sequence length and content made the alignment of the different leader sequences difficult. However, alignment of the six different coronaviruses revealed that they fell into two groups. There appears to be some conservation of short sequence motifs between the seven leader sequences. Toward the 3' end of the sequences, a TAG motif is conserved in all the leaders,

followed by a string of Ts. In five out of seven of the sequences, this motif is TAGANNTT. About ten nucleotides downstream of this region is a conserved CT motif, which is followed by a series of nucleotides differing in number, depending on the coronavirus, followed by the postulated RNA polymerase-leader complex binding site. The largest number of nucleotides between the CT motif and the consensus sequence are found on TGEV and PRCV, the shortest is found on HCV (229E) and IBV. It is interesting to note that there is a five-base insert in MHV strain JHM when compared to MHV strain A59, which is also present in HCV (OC43) within this region. All the mammalian coronaviruses appear to have the motive CTAAAC, except HCV (OC43), which has CTAAAT. Recent sequence data suggest that coronaviruses FIPV and BCV have ACTAAAC as their mRNA consensus sequence. Upstream of the TAG motif there is an ACT motif occurring in six out of seven sequences. Toward the 5' end of the leader RNA sequences, the homologies are patchy and limited to short matches, occurring only between pairs of sequences.

The area upstream of the consensus sequence has been suggested to be involved in the binding of nucleoprotein to the leader RNA sequence at nucleotides 56–65 in MHV (32). It was suggested that mRNA species and genomic RNA form a complex with the nucleoprotein by the protein binding to or near the leader sequence attached to the RNA molecules (33). Secondary structure analysis of the leader RNA sequences showed that all the sequences except for IBV possess a putative double stem-loop structure (Fig. 6). In the case of the mammalian coronaviruses, the consensus sequences and upstream regions of homology are on the second stem-loop structure, leaving the possibility that the RNA-dependent RNA polymerase could interact with the first stem-loop structure. The IBV consensus sequence is present on the free 3' end of the single stem-loop structure, possibly leaving the single stem-loop structure to interact with the polymerase.

Acknowledgments

We thank Miss K. Mawditt, of this laboratory, for synthesizing oligos 38 and 58 and Dr. S. F. Cartwright, Central Veterinary Laboratory, Weybridge for PRCV strains 86/137004 and 86/135308. This work was supported by a research contract from the Biomolecular Engineering Programme of the Commission of the European Communities, contract No. BAP-0235-UK(HI).

References

1. Doyle, L.P. and Hutchings, L.M., *J Am Vet Med Assoc* 108, 257–259, 1946.
2. Pensaert, M.B., Callebaut, P.E., and Vergote, J., *Vet Q* 8, 257–260, 1986.
3. Brown, I. and Cartwright, S., *Vet Rec* 119, 282–283, 1986.
4. Pensaert, M.B., in Horzinek, M.C. (ed.) *Virus Infections of Vertebrates, Vol. 2*. Elsevier, Amsterdam, 1989, pp 154–165.

5. Lai, M.M.C., Patton, C.D., and Stohlman, S.A., *J Virol* *41*, 557–565, 1982.
6. Lai, M.M.C., Patton, C.D., Baric, R.S., and Stohlman, S.A., *J Virol* *46*, 1027–1033, 1983.
7. Lai, M.M.C., Baric, R.S., Brayton, P.R., and Stohlman, S.A., *Proc Natl Acad Sci USA* *81*, 3626–3630, 1984.
8. Spaan, W.J.M., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B.A.M., and Siddell, S.G., *EMBO J* *2*, 1839–1844, 1983.
9. Brown, T.D.K., Bournsnel, M.E.G., and Binns, M.M., *J Gen Virol* *65*, 1437–1442, 1984.
10. Baric, R.S., Stohlman, S.A., and Lai, M.M.C., *J Virol* *48*, 633–640, 1983.
11. Makino, S., Stohlman, S.A., and Lai, M.M.C., *Proc Natl Acad Sci USA* *83*, 4202–4208, 1986.
12. Budzilowicz, C.J., Wilczynski, S.P., and Weiss, S.R., *J Virol* *53*, 834–840, 1985.
13. Shieh, C.-K., Soe, L.H., Makino, S., Chang, M.-F., Stohlman, S.A., and Lai, M.M.C., *Virology* *156*, 321–330, 1987.
14. Makino, S., Soe, L.H., Shieh, C.-K., and Lai, M.M.C., *J Virol* *62*, 3870–3873, 1988.
15. Britton, P., Carmenes, R.S., Page, K.W., Garwes, D.J., and Parra, F., *Mol Microbiol* *2*, 89–99, 1988.
16. Britton, P., Carmenes, R.S., Page, K.W., and Garwes, D.J., *Mol Microbiol* *2*, 497–505, 1988.
17. Britton, P., Lopez Otin C., Martin Alonso J.M., and Parra, F., *Arch Virol* *105*, 165–178, 1989.
18. Kapke, P.A. and Brian, D.A., *Virology* *151*, 41–49, 1986.
19. Rasschaert, D., Delmas, B., Charley, B., Grossclaude, J., Gelfi, J., and Laude, H., in Lai, M.M.C. and Stohlman, S.A. (eds) *Coronaviruses. Adv Exp Med Biol, Vol. 218*. Plenum Press, New York, 1987, pp 109–116.
20. Rasschaert, D., Gelfi, J., and Laude, H., *Biochimie* *69*, 591–600, 1987.
21. Britton, P., Garwes, D.J., Page, K., and Walmsley, J., in Lai, M.M.C. and Stohlman, S.A. (eds) *Coronaviruses. Adv Exp Med Biol, Vol. 218*. Plenum Press, New York, 1987, pp 55–64.
22. Maniatis, T., Fritsch, E.F., and Sambrook, J., *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, New York, 1982.
23. Biggin, M.D., Gibson, T.J., and Hong, G.F., *Proc Natl Acad Sci USA* *80*, 3963–3965, 1983.
24. Maxam, A.M. and Gilbert, W., *Methods Enzymol* *65*, 499–560, 1980.
25. Bournsnel, M.E.G., Brown, T.D.K., Foulds, I.J., Green, P.F., Tomely, F.M., and Binns, M.M., *J Gen Virol* *68*, 57–77, 1987.
26. Kanehisa, M.I., *Nucleic Acids Res* *10*, 183–196, 1982.
27. Zuker, M. and Stiegler, P., *Nucleic Acids Res* *9*, 133–148, 1981.
28. Devereux, J., Haerberli, P., and Smithies, O., *Nucleic Acids Res* *12*, 387–395, 1984.
29. Garwes, D.J., Stewart, F., Cartwright, S.F., and Brown, I., *Vet Rec* *122*, 86–87, 1988.
30. Callebaut, P., Correa, I., Pensaert, M., Jimenez, G., and Enjuanes, L., *J Gen Virol* *69*, 1725–1730, 1988.
31. Schreiber, S.S., Kamahora, T., and Lai, M.M.C., *Virology* *169*, 142–151, 1989.
32. Stohlman, S.S., Baric, R.S., Nelson, G.N., Soe, L.H., Welter, L.M., and Deans, R.J., *J Virol* *62*, 4288–4295, 1988.
33. Baric, R.S., Nelson, G.W., Fleming, J.O., Deans, R.J., Keck, J.G., Casteel, N., and Stohlman, S., *J Virol* *62*, 4280–4287, 1988.
34. Britton, P., Garwes, D.J., Page, K., and Stewart, F., in *Coronaviruses. Adv Exp Med Biol*, Plenum Press, New York, in press.
35. Kamahora, T., Soe, L.H., and Lai, M.M.C., *Virus Res* *12*, 1–9, 1989.