Leader–mRNA Junction Sequences Are Unique for Each Subgenomic mRNA Species in the Bovine Coronavirus and Remain So Throughout Persistent Infection[1]

MARTIN A. HOFMANN,[2] RUEY-YI CHANG, SEULAH KU, AND DAVID A. BRIAN[3]

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

The common leader sequence on bovine coronavirus subgenomic mRNAs and genome was determined. To examine leader–mRNA junction sequences on subgenomic mRNAs, specific oligodeoxynucleotide sets were used in a polymerase chain reaction to amplify junction sequences from either the positive-strand mRNA (eight of nine total identified species) or the negative-strand anti-mRNA (six of the nine species), and sequenced. The mRNA species studied were those for the N, M, S, and HE structural proteins and the 9.5-, 12.7-, 4.8-, and 4.9-kDa putative nonstructural proteins. By defining the leader–mRNA junction sequence as the sequence between (i) the point of mismatch between the leader and genome and (ii) the 3' end of the consensus heptameric intergenic sequence [(U/A)C(U/C)AAAC)], or its variant, a unique junction sequence was found for each subgenomic mRNA species studied. In one instance (mRNA for the 12.7-kDa protein) the predicted intergenic sequence UCCAAAC was not part of the junction region, and in its place was the nonconforming sequence GGTAGAC that occurs just 15 nt downstream in the genome. Leader–mRNA junction sequences found after 296 days of persistent infection were the same as those found during acute infection (<18 hr postinfection). These data indicate that, in contrast to the closely related mouse hepatitis virus, the bovine coronavirus maintains a stable leader–mRNA junction sequence for each mRNA. Interestingly, this stability may be related to the fact that a UCUAA sequence element, postulated by others to be a regulator of the leader–mRNA fusion event, occurs only once within the 3' flanking sequence of the genomic leader donor and once at intergenic sites in the bovine coronavirus genome, whereas it occurs two to four times at these sites in the mouse hepatitis coronavirus. © 1993 Academic Press, Inc.

INTRODUCTION

It is well established that the 3' coterminal subgenomic mRNAs of coronaviruses have a common 5' terminal leader sequence of approximately 65 to 90 nucleotides (the length depends on the species of coronavirus) which is not encoded by a colinear region on the genome, but rather by the 5' end of the genome (for reviews see Lai, 1990; and Spaan *et al.*, 1988). The leader on each subgenomic mRNA makes up only a portion of the 5' untranslated region (Fig. 1) which, in turn, differs in length for each mRNA species (Table 1). Assuming that subgenomic mRNA molecules are derived from full-length genome during virus replication, then a mechanism must exist whereby a copy of the genomic leader becomes fused within the 5' untranslated region to the subgenomic mRNAs during virus replication. For each of several models that have been proposed to explain the fusion event (described below), a conserved heptameric intergenic sequence occurring in the genome at varying distances upstream from the start codon of the various genes is thought to be an important sequence element determining the fusion site, since the point of fusion in the 5' untranslated region of the mRNA is within or just upstream of the consensus intergenic sequence.

The three major hypotheses that have been put forward to explain the leader–mRNA fusion process are the following: (i) The discontinuous transcription hypothesis states that the polymerase discontinues synthesis after transcribing leader from the 3' end of the minus strand, and jumps to a new transcription site at conserved intergenic heptameric sequences on the antigenome to continue synthesis of the mRNA (Baric *et al.*, 1983; Lai, 1988; Spaan *et al.*, 1983). This model has been slightly modified to become the leader-priming hypothesis which states that leader is made in excess, becomes free, and serves to prime transcription at the intergenic sites on the minus strand *in cis* or *in trans* (Makino *et al.*, 1986). (ii) The minus-strand splicing hypothesis states that full-length minus-strand anti-genomes become spliced either *in cis* or *in trans* to mRNA-length minus-strand molecules that then serve as templates for synthesis of mRNAs (Sawicki and Sawicki, 1990). (iii) The plus-strand splicing hypothesis states that progeny genomes become directly spliced

---

[1] Sequence data from this article have been deposited with GenBank Data Library under Accession Number M62375.

[2] Present address: Institut für Viruskrankheiten und Immunoprophylaxe, Hagenaustrasse 74, CH-4025 Basel, Switzerland.

[3] To whom correspondence and reprint requests should be addressed.

A

leader   consensus heptameric intergenic sequence

5'- GAUUGUG-(N$_{50}$)-UUAUAA|UCUAAAC| --------------------- ? 3'

B

point of mismatch with genome

5'- GAUUGUG-(N$_{50}$)-UUAUA AUCUAAAC ——————— |AUG| mRNA ORF → 3'
                    | | | | | | | |
3'---——————————UAGAUUUG——————— |UAC|————————----5'
       antigenome

C

leader-mRNA junction sequence

5'- GAUUGUG-(N$_{50}$)-UUAUA|AUCUAAAC|————— |AUG| mRNA ORF → 3'
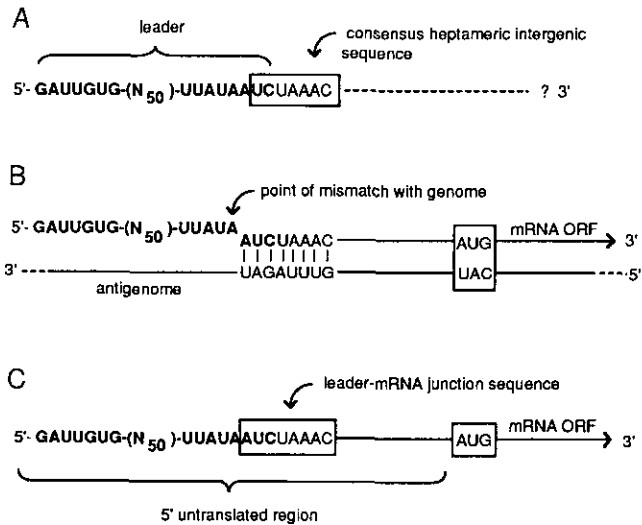
5' untranslated region

FIG. 1. Schematic definition of the leader–mRNA junction sequence. (A) Postulated "free" leader containing the common leader sequence (65 nt for BCV, identified in bold type), the flanking consensus intergenic sequence, and extended sequence of unknown length. (B) Alignment of leader and remaining mRNA with the antigenome after synthesis of mRNA by the postulated leader-primed mechanism of transcription. BCV N mRNA is used here for illustration. In this case, the leader–mRNA junction sequence is AU-CUAAAC. (C) 5' end of the completed BCV N mRNA showing the leader sequence and the leader–mRNA junction sequence as components of the 5' untranslated region.

either *in cis* or *in trans* into subgenomic mRNA molecules (Baric *et al.,* 1983; Lai, 1988). Of these hypotheses, the leader-priming hypothesis has been used the most widely to explain the origin of leader-containing subgenomic mRNA molecules. With any one of the three models, however, it might be expected that the fusion event would be an imprecise one and that variant fusion junctions for any given mRNA species would be observed from a collection of imprecise fusion events. For the leader priming model, for example, it could be postulated that the free leader might misalign during the priming process and create a different leader–mRNA junction. For the splicing models, these variations could arise from alternative splice sites.

Recent studies on the mouse hepatitis virus (MHV) have shown that multiple leader–mRNA junction sequences do in fact exist for certain of the subgenomic mRNA species (Makino *et al.,* 1988). These findings were interpreted in light of the leader-priming model and explained by the existence of multiple 5'UCUAA sequence sets in the putative free leader that align in optional ways with multiple 3'AGAUU repeats occurring within intergenic regions on the antigenome (Makino and Lai, 1989; Baker and Lai, 1990). Thus, the resulting mRNA species had two, three, or four UCUAA repeats within the leader–mRNA junction sequence. It was further documented for the JHM strain of MHV that transcription of one subgenomic mRNA species,

mRNA 2b (the transcript for the HE protein), occurred when two UCUAA repeats existed in the genomic leader (and thus also the presumed free leader), but not when three existed (Shieh *et al.,* 1989). Other variations of this phenomenon have been described for MHV (LaMonica *et al.,* 1992). The number of UCUAA repeats was thus postulated to be part of a mechanism for regulating the rate of transcription initiation (LaMonica *et al.,* 1992; Shieh *et al.,* 1989).

We have sought to study the leader–mRNA junctions on bovine coronavirus (BCV) mRNAs for several reasons. (i) Sequence analyses may provide clues to the leader fusion mechanism. (ii) BCV is a close relative of MHV, on the basis of amino acid sequence similarities in the proteins (Abraham *et al.,* 1990a,b; Kienzle *et al.,* 1990; Lapps *et al.,* 1987), and a comparative analysis of leader and leader–mRNA junction sequences *may point out important conserved structural features* that might suggest leader function(s). This is especially important with regard to fusion sequence variations and the regulation of gene expression. We notice, for example, no variation in the levels of HE protein or HE mRNA (species 2b) production in different subclones or passages of BCV (Hogue *et al.,* 1989; Hofmann *et al.,* 1990; data not shown). During BCV replication both are invariably expressed at high levels. (iii) Highly variable junction sequences would challenge the notion of mRNA replication (Sethna *et al.,* 1989). We and others have recently shown that coronavirus subgenomic mRNAs have minus-strand counterparts that appear to be active in transcriptional or replicational complexes

TABLE 1

PROPERTY OF 5'-UNTRANSLATED REGIONS AND LEADER–mRNA JUNCTION SEQUENCES ON BOVINE CORONAVIRUS GENOME AND SUBGENOMIC mRNAs[a]

mRNA species	Length of 5'UTR (nt)	Leader–mRNA junction sequence
Genome	210	-TAAAC-
32K	ND	ND
HE	79	-CTAAAC-
S	70	-ATAATCTAAAC-
(4.9K)[b]	(no start codon)	(-CTAAGT-)
4.8K	75	-TAAAC-
12.7K	124	-TAGAC-
9.5K	193	-AATCCAAAC-
M	73	-TAATCCAAAC-
N	77	-ATCTAAAC-

[a] The 5' leader sequence of 65 nt is identical on all sequenced BCV RNA species. The leader–mRNA junction sequence is defined as that sequence which lies between (i) the point of mismatch between leader and genome, and (ii) the 3' end of the consensus intergenic sequence or its variant.

[b] Although the mRNA species for the 4.9-kDa protein was found, it is apparently nonfunctional since it contains no start codon (as explained in the text). ND, not determined.
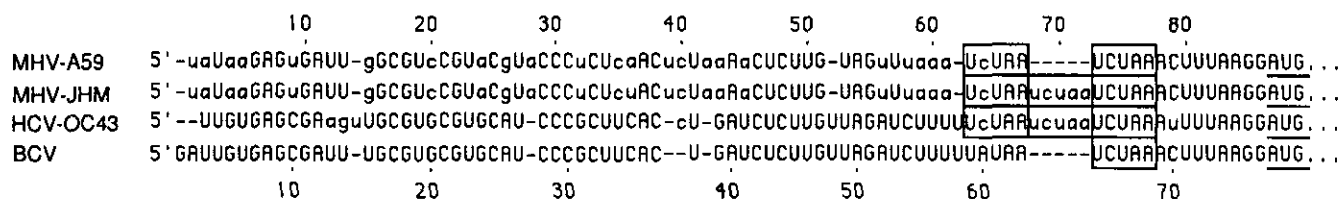
FIG. 2. Alignment of the leader-containing 5' untranslated regions on the N mRNAs for the bovine coronavirus and three of its close relatives. Numbers on top refer to the composite sequence, and on the bottom to the BCV sequence. Nucleotides that are identical to the BCV sequences are shown in upper case letters. The putatively critical UCUAA sequences are boxed, and the N start codons are underlined. (One apparently noncritical UCUAA set beginning at base position 39 in the MHV strains is not boxed.) The first 65 nt of the BCV sequence make up the BCV leader. The first 74 nt of the illustrated BCV sequence are identical to the first 74 nt on the BCV genome.

in the infected cell (Hofmann et al., 1990; Sawicki and Sawicki, 1990; Sethna et al., 1989). If subgenomic mRNAs do undergo replication, then this would be one mechanism by which mRNAs could maintain a stable structure during persistent infection. Conversely, variations in leader–mRNA junction sequences over time would be difficult to explain if mRNAs were perpetuated only by a replication mechanism.

In experiments described here, the common leader sequence on the BCV subgenomic mRNAs and genome was determined, and by sequencing PCR amplified junctions, leader–mRNA junction sequences for eight of nine species of BCV subgenomic mRNAs were determined. The junction sequences on the anti-mRNAs (six of nine species) were also determined. We learned that the leader–mRNA junction sequence was unique for each subgenomic mRNA and remained so throughout persistent infection. The intergenic consensus sequence (U/A)C(U/C)AAAC occurs no more than once in any given BCV leader–mRNA junction sequence, and on one mRNA species does not occur at all. The leader–mRNA junction sequences were therefore remarkably stable in the bovine coronavirus and, by comparison with published data on the mouse coronavirus, this stability appears to be related to the singular occurrence of the UCUAA sequence element in the genomic leader 3' flanking sequence and at each intergenic region in the genome.

MATERIALS AND METHODS

Virus and cells and preparation of RNA from infected cells

The Mebus strain of BCV was grown on human rectal tumor (HRT) cells as previously described (King and Brian, 1982; Lapps et al., 1987). A persistent infection in cell culture was established, and RNA was extracted from acutely and persistently infected cells as previously described (Hofmann et al., 1990). In the current study, persistently infected cells had been carried for 74 passages (296 days).

Determination of the BCV leader sequence on the N, M, and S mRNAs and genomic RNA

To establish the BCV leader sequence on N mRNA (Fig. 2), oligodeoxynucleotide 1, which binds to a re-

gion beginning 34 nt downstream from the N translation initiation codon (Lapps et al., 1987) (Fig. 3), was 5' end-labeled with polynucleotide kinase and [γ-^{32}P]ATP (>3000 Ci/mmol; ICN Pharmaceuticals), and used as primer for an extension reaction on cytoplasmic RNA harvested at 6 hr postinfection (Hofmann et al., 1990). The extended primer was isolated by electrophoresis on a 12% polyacrylamide sequencing gel and sequenced by the chemical method (Maxam and Gilbert, 1980). This is essentially the same method used to determine the leader sequence of MHV (Lai et al., 1984), the avian infectious bronchitis virus (IBV) (Brown et al., 1984), and the porcine transmissible gastroenteritis virus (TGEV) (Sethna et al., 1991). To confirm the leader sequence, the experiment was done with extended products from the M mRNA template (Lapps et al., 1987). For this, oligodeoxynucleotide 2 which binds to a region beginning 19 nt downstream from the predicted M start codon (Fig. 3) was likewise used.

Because it was impossible with this experimental approach to unequivocally resolve the 5' terminal bases, a method was developed in which extended primers in separate experiments for each the N, M, and S mRNAs were ligated into head-to-tail multimers (dimers and trimers) with RNA ligase, and PCR-amplified molecules representing the joined regions were cloned and sequenced by the dideoxynucleotidyl chain-termination method (Hofmann and Brian, 1991a,b).

To determine the leader sequence on the BCV genome, the Ndel site-containing oligonucleotide 5'CCTCCAAATCATATGGACGTGTATTC3' which binds to a region on genomic RNA beginning 456 nt downstream from the 5' end, and the EcoRI, KpnI site-containing oligonucleotide5'CGGAATTCGGTACCGATTGTGAGCGATTTGCGTGCG3' which binds to the first 22 nt of the antileader, were used in a PCR to generate a 495 nt fragment that was cloned and sequenced. Genomic RNA was obtained from a virus purified from the supernatant fluids of persistently-infected cells at 120 days postinfection (Hofmann et al., 1990). At this passage level, there was no detectable defective interfering RNA species in the cells or virus. The 5' end of the genome was originally obtained from a cloned and se-
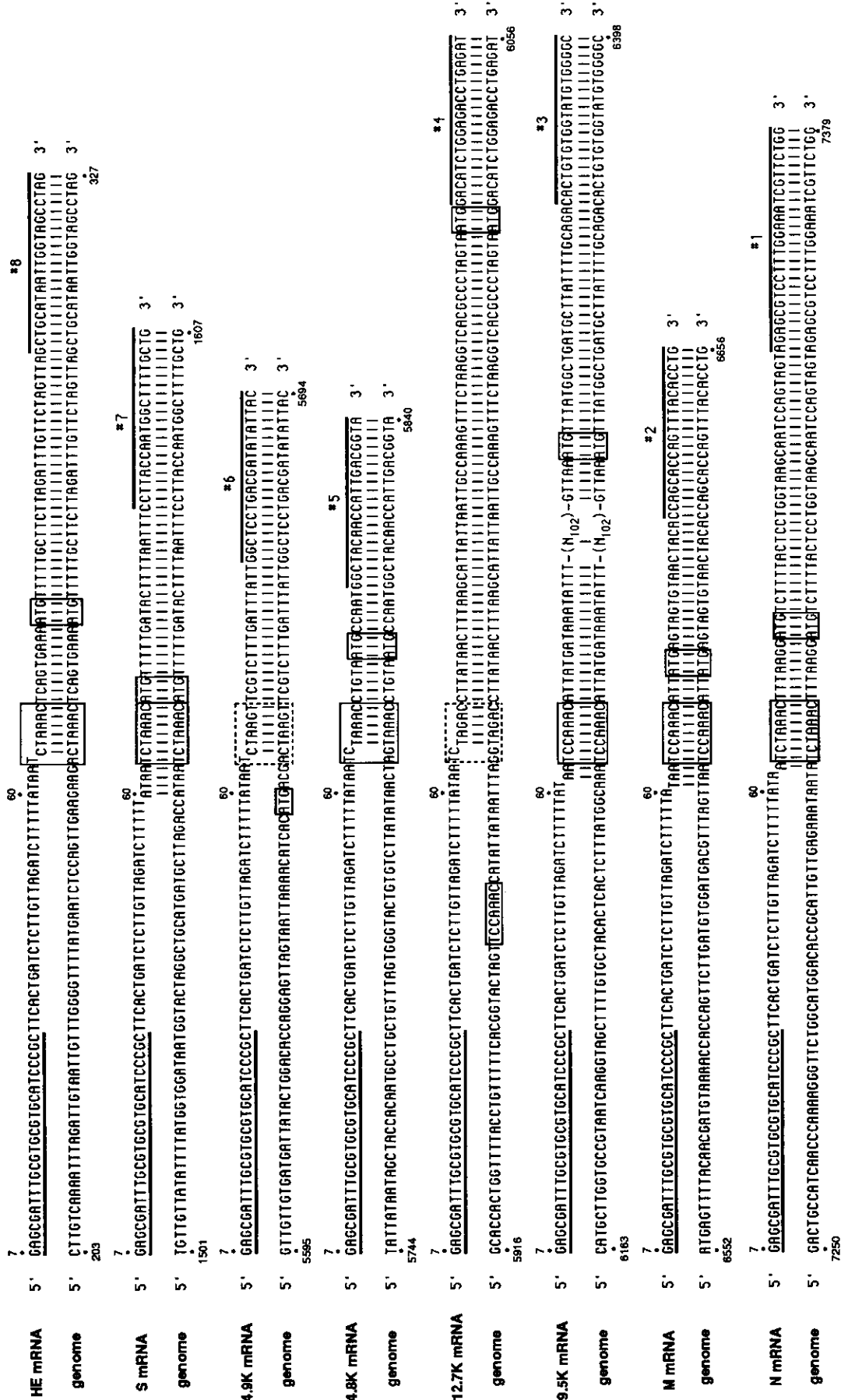
FIG. 3. Leader–mRNA junction and flanking sequences for eight of the nine bovine coronavirus subgenomic mRNAs. mRNA sequences are shown in alignment with the genome sequence. Numbering on the genome sequence begins at a position 8955 nt from the base of the 3' poly(A) tail. Underlined sequence identifies the region to which antileader-specific primer binds. Overlined sequences identify regions to which the mRNA-specific primers (1 through 8) bind. Regions boxed with a solid line are the putative intergenic sequences and start codons. Regions boxed with a dashed line were the unpredicted, but used, variants of the consensus intergenic sequence. Homologous nucleotides between the mRNA and genome are indicated with connecting lines.

quenced defective-interfering RNA which showed a 5′ terminal nucleotide sequence identity of 64% with the 5′ terminal 498 nt of the MHV genome (Chang et al., submitted for publication; Soe et al., 1987) and it was from this sequence that the NdeI site-containing oligonucleotide (described above) was obtained. The 5′ terminal 455 sequence of the genome is identical to that of the defective-interfering RNA. Since some of the 5′ 23 nt of the genomic leader may have been dictated by the antileader primer, this portion of the leader remains to be confirmed.

Asymmetric PCR amplification and sequencing of leader–mRNA junction sequences

For analysis of leader–mRNA junction sequences, oligodeoxynucleotides 1 and 2 (described above), and oligodeoxynucleotides 3 through 8 (described in Fig. 3) which bind respectively to regions within the open reading frames of the 9.5-, 12.7-, 4.8-, 4.9-kDa, S, and HE proteins, (Abraham et al., 1990a,b; Kienzle et al., 1990) were prepared and used for first-strand cDNA synthesis and for subsequent thermocycling reactions. A 26-mer oligodeoxynucleotide (5′GAGCGATTT-GCGTGCGTGCATCCCGC3′) which binds to nt 7 through 32 of the antileader (counting from the 3′ end of the antileader) was prepared and used as the second primer in all thermocycling reactions.

For first-strand cDNA synthesis from mRNA template, 75 μg of total cytoplasmic RNA (stored as an ethanol precipitate) was pelleted, washed with 80% ethanol, dried, dissolved in 12.25 μl water, denatured by heating at 70° for 5 min, and quick-cooled on ice. First-strand cDNA was synthesized in a 30-μl reaction mixture containing 12.25 μl (75 μg) RNA, 5 μl (~50 pmol) first-strand mRNA-specific primer, 3 μl 10× reverse transcriptase buffer (1× reverse transcriptase buffer = 50 mM Tris–HCl, pH 8.3, at 42°, 7 mM MgCl₂, 40 mM KCl, 1 mM DTT, and 0.1 mg BSA per ml), 6 μl of the four dNTPs (at 5 mM each), 0.75 μl (30 units) RNasin (Promega), and 3 μl (24 units) AMV reverse transcriptase (Promega), for 2 hr at 42°. The reverse transcriptase reaction was stopped by incubation at 75° for 15 min, and the RNA was removed by adding 2 μl RNase A (at 2 mg per ml; Sigma) and incubating the mixture at 37° for 1 hr. ssDNA was extracted once with phenol chloroform and once with chloroform.

For second-strand cDNA synthesis and PCR amplification, a 50-μl PCR reaction mix containing 25.75 μl H₂O, 5 μl 10× PCR buffer (1× PCR buffer = 10 mM Tris–HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 0.01% gelatin, and 0.01% Triton X-100), 8 μl dNTPs (at 1.25 mM each), 5 μl of the antileader-specific primer (at 10 μM), 5 μl of the respective mRNA-specific primer (at 10 μM), 1 μl of first-strand synthesis mix, and 0.25 μl (1.25 units) Taq DNA polymerase (Promega) was prepared

and incubated for 30 cycles in the thermal cycler, 1 min at 94°, 1 min at 55°, and 1 min at 72° for each cycle.

For analysis of minus-strand RNA, the same procedures as described above for first and second-strand cDNA synthesis and PCR amplification were used, except the order of primer usage was reversed.

To examine the sizes of the amplified double-stranded DNA products and to obtain a small amount of purified product for asymmetric PCR amplification, 20 μl of the PCR reaction mix was electrophoresed in a 4% agarose gel (NuSieve, FMC) containing Tris–borate buffer and visualized by ethidium bromide staining. The predicted sizes of the amplified products were 129 bp for the N mRNA, 104 bp for M mRNA, 235 bp for the 9.6-kDa protein mRNA, 140 bp for the 12.7-kDa protein mRNA, 96 bp for the 4.9-kDa protein mRNA, 99 bp for the 4.9-kDa protein mRNA, 106 bp for the S mRNA, and 124 bp for the HE mRNA.

For asymmetric amplification of DNA by PCR in preparation for DNA sequencing, a modified method of Innis et al., (1988) was used. To a 50-μl reaction mix containing 42.2 μl water, 5 μl 10× PCR buffer, 0.8 μl dNTP's (at 1.25 mM each), 1 μl of the antileader-specific primer (the limiting primer; 0.1 μM), and 1 μl of the mRNA-specific primer (10 μM) was added a small inoculum of amplified DNA from the agarose gel. The DNA was transferred by stabbing the band with a micropipet tip and rinsing the tip in the asymmetric reaction mix. The reaction mix was overlaid with mineral oil and incubated for 40 cycles in a thermocycler, 30 sec at 94°, 1 min at 55°, and 1 min at 72°. The asymmetrically amplified single-stranded DNA was separated from unincorporated dNTPs by chromatography in a water-equilibrated Sephadex G25 Spin column, dried, and redissolved in 20 μl 1× PCR buffer.

For sequencing asymmetrically amplified DNA, the protocol of Innis et al. (1988; Hofmann and Brian, 1991b) for dideoxynucleotidyl DNA sequencing with Taq DNA polymerase was used. The 5′ end of the leader-specific primer was labeled with ^{32}P using polynucleotide kinase and [γ[32]P]ATP (>3000 Ci/mmol; ICN) and used in the extension reactions. DNA was electrophoresed on 8% polyacrylamide gels containing 50% urea.

RESULTS

BCV leader sequence on mRNAs and genome

The entire 77-nt untranslated region on the BCV N mRNA is shown in Fig. 2. A majority of the BCV 5′ untranslated region for the N and the M mRNAs was determined by chemical sequencing, but the 5′ terminal five nucleotides for these mRNA species as well as for the S mRNA were established by sequencing ligated head-to-tail cDNA products prepared from mRNA obtained during acute infection (Hofmann and

Brian, 1991a,b). From these data, and from the leader–mRNA sequence data on all but one of the subgenomic mRNAs described below, we conclude that the BCV common leader on subgenomic mRNAs is the first 65 nucleotides of the BCV sequence shown in Fig. 2. Furthermore, by sequencing a cloned 495-nt PCR fragment of the 5′ end of the genome, it was established that the genomic 5′ terminus is identical to the N mRNA 5′ terminus for nt 23 through 74, and probably also identical for nt 1–22 (the uncertainty arises because this is the region to which the primer bound) (Fig. 2).

Since the leader–mRNA fusion sites appear to be within or just upstream of a singular consensus heptameric intergenic sequence [(U/A)C(U/C)AAAC)], the leader–mRNA junction sequence was defined in this study as that region extending from the point of divergence between the leader and genome, to the 3′ end of the heptameric intergenic sequence (or its variant) (Fig. 1).

Leader–mRNA junction sequences are unique for each subgenomic mRNA species

The results of each sequenced junction for eight of nine species of subgenomic mRNA are shown in Fig. 3 and Table 1. The junction sequence on the subgenomic mRNA of the 32-kDa nonstructural protein described by Cox et al. (1989) was not studied since the sequence of this gene in the Mebus strain of BCV had not been determined.

From our analyses, the following points emerge: (i) The leader–junction regions range in length from 5 nt for the mRNAs of the 12.7- and 4.8-kDa proteins to 11 nt for the S mRNA, and the leader–junction sequence is unique for each of the eight subgenomic mRNAs studied. (ii) The positions of five of the leader–mRNA junction sequences were as predicted from the positions of a (U/A)C(U/C)AAAC consensus sequence on the genome, from the proximity of these sequences to the putative AUG initiator codons on the deduced genes, and from the size of the mRNAs as judged by Northern analyses. These were for the mRNAs of the N, M, 9.5-kDa, S, and HE proteins (Abraham et al., 1990a,b; Kienzle et al., 1990; Lapps et al., 1987). (iii) Three of the eight junction regions, however, were different from those predicted. (a) The mRNA for the 12.7-kDa protein was found to be joined to the leader sequence 21 nt downstream from the predicted UCCAAAC sequence (at boxed position 5953 in Fig. 3), and joined at the unusual consensus variant GGUAGAC (Fig. 4). The resulting mRNA would be expected to be functional since the first AUG on the sequence predicts the start of the identified 12.7-kDa protein. (b) The mRNA having the 4.8-kDa ORF as the upstream ORF was not predicted earlier since no (U/A)C(U/C)AAAC consensus sequence was found near the pu-
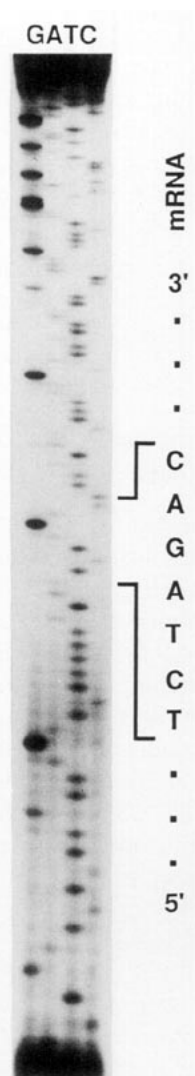

FIG. 4. The leader–mRNA junction sequence for the mRNA of the 12.7-kDa protein. The asymmetrically amplified DNA was sequenced and the leader–mRNA junction sequence is shown.

tative AUG initiation codon, and the 4.8- and 4.9-kDa ORFs together appeared to be the result of a point mutation disrupting an ancestral 11-kDa ORF of the kind found in MHV (Abraham et al., 1990). The 4.8-kDa ORF junction region appears to utilize AGUAAAC as the intergenic sequence which requires a mismatch in the second nucleotide of the consensus sequence. (c) Most surprising is an intergenic sequence near the beginning of the AUG for the 4.9-kDa protein mRNA. This junction did not utilize the predicted ACCAAAC consensus sequence beginning 323 nt upstream from the AUG, but rather used the ACUAAGU sequence beginning 4 nt downstream from the putative AUG. This transcript is therefore probably not used for protein synthesis since the first AUG in the transcript, even though it is in a good sequence context for initiation of translation (Kozak, 1991), is 102 nt downstream and would yield a peptide of only four amino acids.

To confirm the existence of the leader–mRNA junction sequences, oligodeoxynucleotides that would prime and amplify the minus-strand copy of each leader–mRNA junction (Hofmann et al., 1990) were used and the asymmetrically amplified DNA was sequenced. Abundant discrete PCR products were obtained from the minus strands for all but the 4.9- and 4.8-kDa protein mRNAs, and hence no DNA sequence was obtained from these two species. These results are consistent with our earlier RNA blotting analysis (Hofmann et al., 1990), and suggest that minus-strand copies of the specific 4.9- and 4.8-kDa mRNA species either do not exist or are present in very small numbers.

Leader–mRNA junction sequences remain unchanged after 296 days of persistent infection

To examine the leader–mRNA junction sequences after a long period of persistent infection, cytoplasmic RNA extracted from cells at 296 days (74 passages) postinfection was used in an identical manner to that described above to separately identify junction sequences on plus- and minus-strand subgenomic RNA species. Except for our failure to confirm sequences on minus-strand copies of mRNAs for the 4.9- and 4.8-kDa proteins, junction sequences were found to be identical to those described in Fig. 3.

DISCUSSION

We have established the sequence of the 5' untranslated region for the bovine coronavirus N and M subgenomic mRNAs by using a combination of three separate methods, direct chemical sequencing of an extended primer, sequencing of a cloned extended primer, and dideoxynucleotidyl sequencing of DNA asymmetrically amplified from an extended primer. By using the latter two methods, the 5' untranslated region on the S mRNA was also determined (Hofmann et al., submitted for publication; and Fig. 3). From these data, we conclude that the common leader on BCV genome and subgenomic mRNAs is 65 nt in length (Fig. 2). The BCV leader is now the seventh coronavirus leader to be sequenced, and its sequence shows more similarity to the leaders of the antigenically closely related human coronavirus HCV-OC43 (Kamahora et al., 1989), MHV-JHM (Shieh et al., 1987), and MHV-A59 (Lai et al., 1984), than to those of the more distantly related porcine transmissible gastroenteritis virus (Sethna et al., 1991), avian infectious bronchitis virus (Brown et al., 1984) and HCV-229E (Schreiber et al., 1989). The leader and flanking consensus sequences of the BCV-related coronavirus species can be compared by aligning the 5' untranslated regions of their respective N mRNAs (Fig. 2). From this it can be seen that the BCV leader has more identity with HCV-OC43 (85%) than with MHV-A59 or MHV-JHM (67 and 63%,

respectively), suggesting that BCV and HCV-OC43 share a common branch point during the evolution of this antigenic virus cluster. One notable feature is that whereas the pentanucleotide UCUAA is found as a sequentially repeated sequence set near the 3' end of the leaders of HCV-OC43 (three times), MHV-JHM (three times), and MHV-A59 (two times), it is present only once near the leader 3' end on BCV mRNAs and genome (the UAA portion actually flanks the leader) (Fig. 2).

Our analyses of the bovine coronavirus leader–mRNA junction sequences reveal first of all that a unique junction sequence exists for each mRNA, and secondly that the uniqueness is maintained throughout at least 296 days of persistent infection. Regardless of whether coronavirus mRNAs arise by only continuous de novo synthesis (transcription) from a larger-than-unit size template (Lai, 1988), or by a two stage process in which the leader–mRNA fusion event is followed by faithful mRNA replication (Hofmann et al., 1990; Sethna et al., 1989), the remarkable stability (i.e., invariance) of the BCV leader–mRNA junction sequences, in our view, is probably mechanistically related to the singular occurrence of the UCUAA sequence element in the putative leader donor sequence and at intergenic sites in the genome. This conclusion is based on (i) a comparison of BCV mRNA structure with the variable nature of leader–mRNA junction sequences in MHV which correlates with multiple UCUAA sequence sets in the putative free leader (donor) and at intergenic sites (Baker and Lai, 1990; Makino et al., 1988; Makino and Lai, 1989; Monica et al., 1992; Shieh et al., 1989), and (ii) recent in vitro mutagenesis studies in which an altered UCUAA intergenic sequence disrupted formation of subgenomic transcripts from a replicating defective-interfering RNA template (Joo and Makino, 1992; Makino et al., 1991). While the stability of the BCV leader–mRNA junction sequences in and of itself is consistent with the notion of mRNA replication, two other observations leave open the possibility that mRNAs arise only through a continuous transcriptional process requiring an accompanying leader fusion step. (i) The stability of the BCV leader–mRNA junction sequences contrasts sharply with a pattern of hypervariability of the 5' end of subgenomic mRNAs observed throughout establishment and maintenance of persistent infection (Hofmann et al., submitted for publication). One would expect that replication should faithfully copy the termini as well as internal (leader–mRNA) junction sequences. (ii) Direct evidence to date does not support the notion that BCV mRNA molecules undergo complete replication. Whereas a cloned subgenomic defective RNA molecule carrying a reporter sequence undergoes replication after transfection into infected cells, copies of cloned mRNA carrying the same reporter sequence do

not (Chang *et al.,* submitted for publication). Since minus-strand anti-mRNA molecules exist, they may serve as templates for transcription by the use of an internal promoter, however (in the same manner as antigenome), but in this case strict control of the fusion event would likewise be expected.

Although we have not clarified the mechanism of leader fusion onto mRNAs in these studies, two features of our data are difficult to reconcile with a leader-priming mechanism that depends on base pairing between the complementary heptameric consensus sequences on plus (free leader) and minus (antigenome or antimRNA) strands. The first is that the most favorable heptameric intergenic sequence upstream of the 12.7-kDa protein gene, UCCAAAC (which would allow a pairing of six bases), appears to have been bypassed in favor of the unusual GGUAGAC intergenic sequence (which allows a pairing of only four bases), occurring just 15 nucleotides downstream, for becoming part of the leader–mRNA junction. More than base pairing appears to have directed this alignment. The second feature difficult to reconcile is that the leader–mRNA junction sequences we find are not dictated by the putative free leader but rather by the genomic sequence. The leader priming hypothesis postulates that the intergenic consensus sequence on mRNAs will be derived from the free leader primer since the primer is extended during the transcription process. From our studies we learn that rather than the expected TAAAC becoming part of the leader–mRNA junction sequence (as predicted from the putative free leader sequence), CAAAC becomes part of the junction on the M and 9.5-kDa mRNAs, TAGAC on the 12.7-kDa mRNA, and TAAGT on the 4.9-kDa mRNA. A similar problem has also been noted for MHV (Makino *et al.,* 1988) which led Makino *et al.* to postulate a corrective proof-reading enzymatic action as part of the coronavirus transcriptase. In addition, a corollary of the base-pairing-based leader-priming hypothesis that states that the rate of mRNA synthesis will be directly proportional to the degree of base pairing at the intergenic consensus region (Budzilowicz *et al.,* 1985; Shieh *et al.,* 1987), is not fulfilled for BCV. Whereas the abundance of BCV mRNAs at the peak time of mRNA synthesis was measured to be inversely proportional to mRNA length (Hofmann *et al.,* 1990), namely N > M > 9.5-kDa > 12.7-kDa > S > HE, the degree of base pairing found within the leader–mRNA junction regions in this study predicts the relative mRNA abundances to be S > M > 9.5-kDa > N > 4.9-kDa = HE > 12.7-kDa = 4.8-kDa (Fig. 3). Base pairing outside the region of the leader–mRNA junction, however, may contribute to the putative priming event and these have not been considered in this prediction.

Although, in our view, the mechanism of leader fusion onto mRNA remains to be established, our data

lend support to the idea that the UCUAA sequence is important to the fusion process, and that the fusion event is strictly controlled when only a singular UCUAA sequence element is present near the fusion site either in the 3' flanking sequence of the putative leader donor or at the intergenic template. The bovine coronavirus appears to be a useful system with which to examine the details of this controlled event since the UCUAA set is present only once at these sites in this coronavirus.

ACKNOWLEDGMENTS

REFERENCES

ABRAHAM, S., KIENZLE, T. E., LAPPS, W. E., and BRIAN, D. A. (1990a). Deduced sequence of the bovine coronavirus spike protein and identification of the internal proteolytic cleavage site. *Virology* **176**, 296–301.

ABRAHAM, S., KIENZLE, T. E., LAPPS, W., and BRIAN, D. A. (1990b). Sequence and expression analysis of potential nonstructural proteins of 4.9, 4.8, 12.7, and 9.5 kilodaltons encoded between the spike and membrane protein genes of the bovine coronavirus. *Virology* **177**, 488–495.

BAKER, S. C., and LAI, M. M. C. (1990). An *in vitro* system for the leader-primed transcription of coronavirus mRNAs. *EMBO J.* **9**, 4173–4179.

BARIC, R. S., STOHLMAN, S. A., and LAI, M. M. C. (1983). Characterization of replicative intermediate RNA of mouse hepatitis virus: Presence of leader RNA sequences on nascent chains. *J. Virol.* **48**, 633–640.

BROWN, T. D. K., BOURSNELL, M. E. G., and BINNS, M. M. (1984). A leader sequence is present on the mRNA A of avian infectious bronchitis virus. *J. Gen. Virol.* **65**, 1437–1442.

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

CHANG, R.-Y., HOFMANN, M. A., SETHNA, P. B., and BRIAN, D. A. Submitted for publication.

COX, C. J., PARKER, M. D., and BABUIK, L. A. (1989). The sequence of cDNA of bovine coronavirus 32K nonstructural gene. *Nucleic Acids Res.* **17**, 5847.

HOFMANN, M. A., and BRIAN, D. A. (1991a). A PCR-enhanced method for determining the 5-prime end sequence of messenger RNAs. *PCR Methods Appl.* **1**, 43–45.

HOFMANN, M. A., and BRIAN, D. A. (1991b). Sequencing PCR DNA amplified directly from a bacterial colony. *BioTechniques* **11**, 30–31.

HOFMANN, M. A., SENANAYAKE, S. D., and BRIAN, D. A. Submitted for publication.

HOFMANN, M. A., SETHNA, P. B., and BRIAN, D. A. (1990). Bovine coronavirus mRNA replication continues throughout persistent infection in cell culture. *J. Virol.* **64**, 4108–4114.

HOGUE, B. H., KIENZLE, T. E., and BRIAN, D. A. (1989). Synthesis and processing of the bovine enteric coronavirus hemagglutinin. *J. Gen. Virol.* **70**, 345–352.

INNIS, M. A., MYAMBO, K. B., GELFAND, D. H., and BROW, M. A. D.

(1988). DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl. Acad. Sci., USA* **85**, 9436–9440.

Joo, M., and Makino, S. (1992). Mutagenic analysis of the coronavirus intergenic consensus sequence. *J. Virol.* **66**, 6330–6337.

Kamahora, T., Soe, L. H., and Lai, M. M. C. (1989). Sequence analysis of nucleocapsid gene and leader RNA of human coronavirus OC43. *Virus Res.* **12**, 1–9.

Kienzle, T. E., Abraham, S., Hogue, B. G., and Brian, D. A. (1990). Structure and orientation of expressed bovine coronavirus hemagglutinin-esterase protein. *J. Virol.* **64**, 1834–1838.

King, B., and Brian, D. A. (1982). Bovine coronavirus structural proteins. *J. Virol.* **42**, 700–707.

Kozak, M. (1991). An analysis of vertebrate mRNA sequences: Intimations of translational control. *J. Cell. Biol.* **115**, 887–903.

Lai, M. M. C. (1988). Replication of coronavirus RNA. *In* "RNA Genetics" (E. Domingo, J. J. Holland, and P. Ahlquist, Eds.), Vol. I, pp. 115–136. CRC Press, Boca Raton, FL.

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

Lai, M. M. C., Baric, R. S., Brayton, P. R., and Stohlman, S. A. (1984). Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proc. Natl. Acad. Sci. USA* **81**, 3626–3630.

LaMonica, N. L., Yokomori, K., and Lai, M. M. C. (1992). Coronavirus mRNA synthesis: Identification of novel transcription initiation signals which are differentially regulated by different leader sequences. *Virology* **188**, 402–407.

Lapps, W., Hogue, B. G., and Brian, D. A. (1987). Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. *Virology* **157**, 47–57.

Makino, S., and Lai, M. M. C. (1989). Evolution of the 5'-end of genomic RNA of murine coronavirus during passages *in vitro. Virology* **169**, 227–232.

Makino, S., Joo, M., and Makino, J. K. (1991). A system for study of coronavirus mRNA synthesis: A regulated, expressed subgenomic defective interfering RNA results from intergenic site insertion. *J. Virol.* **65**, 6031–6041.

Makino, S., Soe, L. H., Shieh, C.-K., and Lai, M. M. C. (1988). Dis-

continuous transcription generates heterogeneity at the leader fusion sites of coronavirus mRNAs. *J. Virol.* **62**, 3870–3873.

Makino, S., Stohlman, S. A., and Lai, M. M. C. (1986). Leader sequences of murine coronavirus mRNAs can be freely reassorted: Evidence for the role of free leader RNA in transcription. *Proc. Natl. Acad. Sci. USA* **83**, 4204–4208.

Maxam, A., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 499–560.

Sawicki, S. G., and Sawicki, D. L. (1990). Coronavirus transcription: Subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. *J. Virol.* **64**, 1050–1056.

Schreiber, S. S., Kamahora, T., and Lai, M. M. C. (1989). Sequence analysis of the nucleocapsid protein gene of coronavirus 229E. *Virology* **169**, 142–151.

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.

Sethna, P. B., Hung, S.-L., and Brian, D. A. (1989). Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proc. Natl. Acad. Sci. USA* **86**, 5626–5630.

Shieh, C.-K., Lee, H.-J., Yokomori, K., Monica, N. L., Makino, S., and Lai, M. M. C. (1989). Identification of a new transcriptional initiation site and the corresponding functional gene 2b in the murine coronavirus RNA genome. *J. Virol.* **63**, 3729–3736.

Shieh, C.-K., Soe, L. H., Makino, S., Chang, M.-F., Stohlman, S. A., and Lai, M. M. C. (1987). The 5'-end sequence of the murine coronavirus genome: Implications for multiple fusion sites in leader-primed transcription. *Virology* **156**, 321–330.

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.

Spaan, W., Cavanagh, D., and Horzinek, M. C. (1988). Coronaviruses: Structure and genome expression. *J. Gen. Virol.* **69**, 2939–2952.

Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B. A. M., and Siddell, S. G. (1983). Coronavirus mRNA synthesis involves fusion of noncontiguous sequences. *EMBO J.* **2**, 1839–1844.