JOURNAL OF VIROLOGY, Aug. 2001, p. 7362–7374 Vol. 75, No. 16
0022-538X/01/$04.00+0 DOI: 10.1128/JVI.75.16.7362–7374.2001
Copyright © 2001, American Society for Microbiology. All Rights Reserved.

Downstream Sequences Influence the Choice between a Naturally Occurring Noncanonical and Closely Positioned Upstream Canonical Heptameric Fusion Motif during Bovine Coronavirus Subgenomic mRNA Synthesis

AYKUT OZDARENDELI, SEULAH KU, SYLVIE ROCHAT, GWYN D. WILLIAMS, SAVITHRA D. SENANAYAKE, AND DAVID A. BRIAN*

Department of Microbiology, University of Tennessee, College of Veterinary Medicine, Knoxville, Tennessee 37996-0845

Received 20 February 2001/Accepted 16 May 2001

Mechanisms leading to subgenomic mRNA (sgmRNA) synthesis in coronaviruses are poorly understood but are known to involve a heptameric signaling motif, originally called the intergenic sequence. The intergenic sequence is the presumed crossover region (fusion site) for RNA-dependent RNA polymerase (RdRp) during discontinuous transcription, a process leading to sgmRNAs that are both 5′ and 3′ coterminal. In the bovine coronavirus, the major fusion site for synthesis of mRNA 5 (GGUAGAC) does not conform to the canonical motif (UC[U,C]AAAC) at three positions (underlined), yet it lies just 14 nucleotides downstream from such a sequence (UCCAAAC). The infrequently used canonical sequence, by computer prediction, is buried within the stem of a stable hairpin (−17.2 kcal/mol). Here we document the existence of this stem by enzyme probing and examine its influence and that of neighboring sequences on the unusual choice of fusion sites by analyzing transcripts made in vivo from mutated defective interfering RNA constructs. We learned that (i) mutations that were predicted to unfold the stem-loop in various ways did not switch RdRp crossover to the upstream canonical site, (ii) a totally nonconforming downstream motif resulted in no measurable transcription from either site, (iii) the canonical upstream site does not function ectopically to lend competence to the downstream noncanonical site, and (iv) altering flanking sequences downstream of the downstream noncanonical motif in ways that diminish sequence similarity with the virus genome 5′ end caused a dramatic switch to the upstream canonical site. These results show that sequence elements downstream of the noncanonical site can dramatically influence the choice of fusion sites for synthesis of mRNA 5 and are interpreted as being most consistent with a mechanism of similarity-assisted RdRp strand switching during minus-strand synthesis.

Coronaviruses and arteriviruses, both members of the *Nidovirus* order of plus-strand RNA animal viruses, appear unique among RNA viruses in their use of a discontinuous transcription step during synthesis of subgenomic mRNAs (10, 14, 28, 48, 54). In both groups of viruses, the transcription pathway ultimately yields a 3′ coterminal nested set of subgenomic mRNAs that are also 5′ coterminal with the virus genome. The common 5′-terminal sequence, called the "leader," encoded at the genome 5′ terminus, makes up only a portion of the 5′ untranslated region in the genome and in each subgenomic mRNA (sgmRNA) species. In general, translation occurs most abundantly from the 5′-most open reading frame (ORF) on each sgmRNA. When originally described, the leader was postulated to become fused with the sgmRNA species by a leader-priming mechanism wherein the RdRp undergoes a copy choice jump on the virus genome-length minus-strand template during plus-strand synthesis (4). The jump in this model would occur for each sgmRNA molecule synthesized, and a postulated 3′→5′ exonuclease would trim the large primer (80 to 140 nucleotides [nt]), termed free leader, down to size (72 nt in mouse hepatitis virus [MHV]) (3). In the leader-priming model, double-stranded sgmRNA-length forms found in coronavirus-infected cells (5, 41, 43, 47) are postulated dead-end products resulting from minus-strand RNA synthesis on sgmRNA templates (23). A recent alternative model for coronavirus transcription postulates an RdRp jump during minus-strand synthesis wherein the intergenic sequence (IS) would exert an attenuating effect on the RdRp, resulting in a donor to (genomic) leader-containing acceptor template switch at the sites of RdRp pausing (40, 42). In this model, sgmRNA minus strands (47) possessing a 3′ antileader sequence (46) and a 5′ oligo(U) (18) would be generated by a copy choice mechanism and then serve as templates (5, 40, 41, 43, 47) for multiple rounds of sgmRNA synthesis. The IS in this case would "promote" formation of the 3′ end of the minus-strand templates for sgmRNA synthesis, and there would be no need to postulate an exonuclease for trimming of leader precursors. Since coronavirus sgmRNA molecules cannot yet be experimentally demonstrated to serve as templates for the generation of new rounds of sgmRNA (12, 35, 37), they cannot be considered replicons as was postulated at the time of discovery of the subgenome-length RNA minus-strand RNAs and double-stranded forms (20, 47). Furthermore, if the second model of sgmRNA synthesis proves correct and there is no bona fide replication of sgmRNAs, then the sgmRNA-length double-stranded forms cannot be considered replicative intermediates in the fullest sense but rather must be seen as transcriptive intermediates of unique character that remain to be fully characterized (40).

Fundamental to both models of discontinuous transcription

* Corresponding author. Mailing address: Department of Microbiology, University of Tennessee, Knoxville, TN 37996-0845. Phone: (865) 974-4030. Fax: (865) 974-4007. E-mail: dbrian@utk.edu.

is the question of what directs the RdRp to undergo the copy choice strand transfer. It was noted early on that a heptameric IS found in the genome near the transcription start site of each sgmRNA was also found just downstream of the genomic leader sequence at the 5′ end of the genome and that the plus strand of one could potentially base pair with the minus strand of the other (4, 9, 49). This led to the notion that base pairing within the IS was a mechanistic feature of the polymerase crossover event, and names such as transcription-associated sequence and transcription-regulating sequence have also been applied to this element (17, 38, 55, 56). The requirement for base pairing during transcription has been recently formally proven for arteriviruses by experiments wherein base pairing was fully manipulated in an infectious genomic clone (55). Transcription rates were controlled by manipulating only the base pairing between these two intergenic elements. Numerous studies with MHV defective interfering (DI) RNAs and the placement of IS elements and various amounts of flanking sequence within the DI RNA have demonstrated that the IS alone is not always sufficient for abundant transcription from that site (2, 22, 24, 25, 33, 34, 35, 53, 56). In addition, both the greater context of the IS location and the quality of flanking sequences can influence the strength of the IS for transcription initiation. It has been suggested that flanking sequences may contribute through base pairing to aid in similarity-assisted homologous recombination by an RdRp copy choice mechanism (see reference 8 and references therein). However, the rules that would allow prediction of IS strength have not yet been deciphered. It remains unknown what factors determine which molecules (among the genome and sgmRNAs) can become donors and which can become acceptors for the polymerase jump during discontinuous transcription and what factors influence the direction, site, and frequency of the jump. In one extreme case, abundant transcription from sites within the foreign green fluorescent protein gene experimentally placed into the coronavirus genome has so far appeared unexplainable by a simple base-pairing hypothesis (15).

In an earlier study of bovine coronavirus (BCoV) transcription, it was noted that the leader fusion motif (GGUAGAC) for sgmRNA 5, the mRNA species predicted to synthesize a gene product with a mass of 12.7 kDa, only partially conforms to the consensus IS (UC[U,C]AAAC) (19) (nonconforming sequences are underlined). Furthermore, it lies just 14 nt downstream from such a fully conforming canonical heptameric UCCAAAC sequence that, curiously, is rarely used and is found within the stem of a predicted stable (−17.2 kcal/mol) hairpin. Here we have documented the existence of the stem and have investigated potential structural determinants of the unusual choice between these two potential fusion sites. We found that placement of the 199-nt-long transcription-initiating region followed by a 92-nt-long reporter into a BCoV DI RNA led to the generation of sgmRNA transcripts from the noncanonical downstream IS, as in the virus genome, which has allowed us to carry out mutagenesis studies on the two IS motifs and their flanking sequences. Our results indicate that sequences downstream of the noncanonical IS motif can exert a stronger influence on the RdRp choice between the two sites than does the apparent secondary structural context of the upstream canonical IS. We conclude that these features are most consistent with a model of sequence similarity-assisted, polymerase copy choice strand switching during minus-strand synthesis.

MATERIALS AND METHODS

Virus and cells. A DI RNA-free stock of the Mebus strain of BCoV at 4.5×10^8 PFU/ml was prepared and used as a helper virus (12). The human rectal tumor cell line HRT-18 (29, 52) was used in all experiments.

Plasmid constructs. Construction of pGEM3Zf(−) (Promega)-based pDrep 1 (Fig. 1B) has been previously described (12). In the complete set of experiments described here, and in all for which data are shown, the 92-nt herpes simplex virus type 1 (HSV-1) glycoprotein D (gD) epitope-encoding sequence (6, 58) was used as the reporter (Fig. 1B and D). In preliminary experiments and for construction of some of the HSV-1-gD-containing mutants, constructs containing a 42-nt HIV-V3 epitope reporter (32) were used. All oligonucleotides used in plasmid construction are shown in Table 1, and all molecular manipulations followed standard protocols (39). To mutate pDrep1 such that it carries the 199-nt-long mRNA 5 transcription-initiating region (i.e., a region containing both the canonical and noncanonical IS sites, beginning 68 nt upstream from the canonical IS and continuing through the first 16 codons of the 12.7-kDa protein ORF [Fig. 1A to C]) and the 42-nt HIV-V3 reporter, thus making pDrepIS12.7V3, oligonucleotides 12.7V3-3′(+) and 12.7-5′(−) were used together with the BCoV genomic cDNA clone pMA5 DNA (29) in a PCR to make a 252-nt product that was trimmed to a fragment of 240 nt with *Nsi*I and cloned into the single *Nsi*I(1665) site of pDrep 1 DNA. The PCR product was also digested with *Nsi*I and *Bam*HI and cloned into *Nsi*I/*Bam*HI-linearized pGEM3Zf(−) (Promega) to make pGEM3Z12.7, a construct used to facilitate subsequent constructions. To prepare pDrepIS12.7V3-mut1, oligonucleotides 12.7V3-3′(+) and M1(−) were used together with pGEM3Z12.7 DNA to make a 252-nt PCR product that was trimmed to a fragment of 240 nt with *Nsi*I and cloned into *Nsi*I-linearized pDrep 1 DNA. pDrepIS12.7V3-mut2 was similarly constructed, except that oligonucleotides 12.7V3-3′(+) and M2(−) were used in the PCR. To construct pDrepIS12.7V3-mut3, overlap PCR mutagenesis was done with oligonucleotides 12.7V3-3′(+), M3(−), and pDrepIS12.7V3 DNA in the first reaction, oligonucleotides M3(+), 12.7-5′(−), and pDrepIS12.7V3 DNA in the second reaction, and oligonucleotides 12.7 V3-3′(+) and 12.5-5′(−) and the products of the first two reactions in a third reaction to make a 252-nt product that was trimmed to a fragment of 240 nt with *Nsi*I and cloned into *Nsi*I-linearized pDrep1. To construct pDrepIS12.7V3-mut4, oligonucleotides 12.7V3-3′(+) and 12.7-5′(−) were used in a PCR with pGEM3ZIS12.7 DNA to make a 252-nt product that was trimmed to a fragment of 190 nt with *Rsa*I and *Nsi*I, ligated at the single *Nsi*I junction with *Nsi*I-linearized pDrep1 DNA, filled in at the unligated ends with T4 polymerase, and ligated at the blunt-ended junctions.

To construct pDrepIS12.7gD(pre), identical to pDrepIS12.7V3 except that it carries the 92-nt HSVgD reporter in place of the 42-nt V3 reporter, a three-way ligation was done with (i) the 4,851-nt *Bst* 1107 I(1864)/*Hin*dIII (in vector) fragment from pDrepIS12.7V3, (ii) the 595-bp *Hin*dIII (in vector)/*Nsi*I(1905) fragment from pDrep IS12.7V3, and (iii) the 92-bp *Bam*HI (blunted with mung bean nuclease)/*Pst*I (*Nsi*I-compatible) fragment of HSV gD epitope-encoding pJB2 (a pUC-19 vector containing the sequence encoding amino acids 26 through 51 of the HSV-1-gD ORF as a *Bam*HI/*Pst*I fragment (6) (Fig. 1D) (a kind gift from J. Bowen). To correct a missing A in pDrepIS12.7gD(pre) at position 1858 (also missing in all pDrepIS12.7V3 constructs, resulting in an out-of-frame reporter with the gene 5 ORF) and to correct a spontaneous C-to-T mutation affecting the fifth amino acid position in the HSV gD epitope (resulting in an unwanted alanine (GCG)-to-valine (GTG) change (Fig. 1D), thus forming pDrepIS12.7gD, megaprimer PCR mutagenesis (21) was used. For this, oligonucleotide 12.7gD(+), 12.7 5′(−), and pDrepIS12.7gD(pre) DNAs were used in a PCR to make a product of 228 nt that was used with oligonucleotide BCV3′end(+) and pDrepIS12.7(pre) DNAs in a second PCR to make a product of 707 nt that was trimmed to a fragment of 280 nt and cloned into *Bam*HI(1667)/*Eco*RI(1947)-linearized pDrepIS12.7gD(pre) DNA. To construct pDrepIS12.7gD-mut1, the 287-nt mutation 1-containing *Spe*I (1441)/*Spe*I (1728) fragment from pDrepIS12.7V3-mut1 was ligated into the equivalent sites of pDrepIS12.7gD. pDrepDIS12.7gD-mut2 was similarly constructed, except that the *Spe*I/*Spe*I fragment came from pDrepIS12.7V3-mut2. To construct pDrepIS12.7gD-mut3, overlap mutagenesis was done with oligonucleotide M3(+), GpD4(+), and pDrepIS12.7gD DNAs in the first reaction, oligonucleotide M3(−), 12.7-5′(−), and pDrepIS12.7gD DNAs in the second reaction, and oligonucleotides pgD4(+) and 12.7-5′(−) and the products of the first two reactions in a third reaction to make a 262-nt product that was trimmed to a fragment of 198 nt with *Bam*HI and *Kpn*I and cloned into *Bam*HI(1667)/*Kpn*I(1865)-linearized

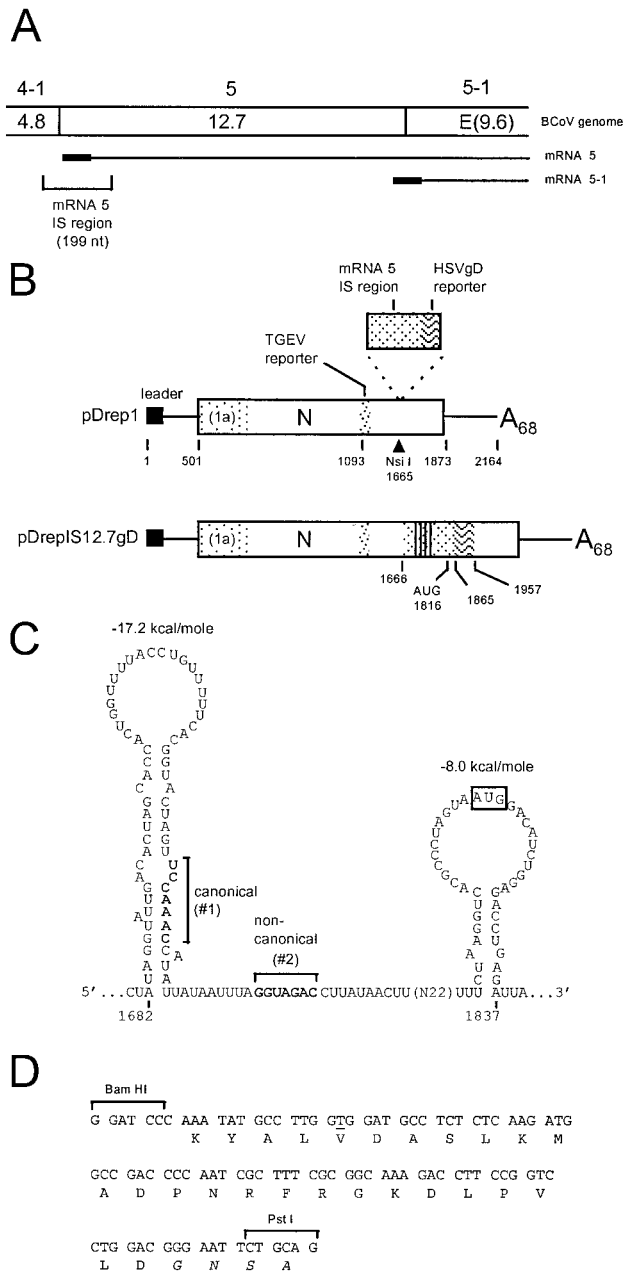

FIG. 1. Expression vector used for mutational analysis of transcription from the canonical site (site 1) and noncanonical site (site 2) for bovine coronavirus mRNA 5. (A) Schematic depiction of the genomic origin of the 199-nt IS region for mRNA 5. The positions of genes 4–1, 5, and 5–1 are shown. The leader sequence on sgmRNA is indicated by the filled box. (B) Modification of the cloned BCoV DI RNA, pDrep1, to contain the 199-nt IS region for gene 5 and the 92-nt HSVgD reporter sequence. Base positions in the respective DI RNA sequences are noted. In pDrepIS12.7gD, the ORF starting at base 501 is interrupted by four stop codons, indicated by vertical lines between bases 1690 and 1801. The ORF beginning with the 12.7-kDa protein start codon at base 1816 is contiguous with the HSVgD reporter and the 3′-terminal portion of the N gene. (C) Secondary structure in the regions of the canonical (site 1) and noncanonical (site 2) fusion sites for mRNA 5 as predicted by the Tinoco algorithm. (D) Sequence of the HSV gD epitope-encoding DNA from which the 92-nt *Bam*HI/*Pst*I fragment was obtained and used as a reporter. The epitope is comprised of amino acids 26 through 51 of the HSV gD protein (identified as nonitalicized letters).

pDrepIS12.7gD. pDrepIS12.7gD-mutants 5, 6, 7, 8, 9, 10, 12, 13, 16, 17, 47, and 48 were constructed identically to mutant 3 except that oligonucleotides M5(+) and M5(−), M6(+) and M6(−), M7(+) and M7(−), M8(+) and M8(−), M9(+) and M9(−), M10(+) and M10(−), M12(+) and M12(−), M13(+) and M13(−), M16(+) and M16(−), M17(+) and M17(−), M47(+) and M47(−), and M48(+) and M48(−) were used in the first and second reactions, respectively. pDrepIS12.7gD-mutants 11, 14, and 15 were prepared by overlap mutagenesis (21), as were mutants 10, 6, and 6, respectively, except that the plasmid DNAs used in the overlap PCR were, respectively, pDrepIS12.7gD-mut9, pDrepIS12.7-mut12, and pDrepIS12.7-mut5. To construct pDrepIS12.7gD-mut4, oligonucleotide 12.7-5′(−), Rev(+) (which binds within the vector), and pDrepIS12.7gD DNAs were used in a PCR to make a 950-nt product that was digested with *Spe*I (which cuts at the 3′ side of the loop at nt 1728 in pDrepIS12.7gD), blunt-ended with mung bean nuclease, digested to a fragment of 780 nt with *Hin*dIII, and cloned into *Bam*HI-digested, mung bean nuclease-blunt ended, *Hin*dIII-digested pDIS12.7gD DNA. To construct pDrepIS12.7gD-mut20, pDrepIS12.7gD-mut3 was digested with *Spe*I, blunt ended with mung bean nuclease, and digested with *Kpn*I to form a 140-nt fragment that was cloned into mung bean nuclease-blunt-ended/*Kpn*-linearized pDrepIS12.7gD DNA.

To construct pGEM4ZIS12.7EP, from which T7 RNA polymerase-generated transcripts were made for RNA enzyme probing, the 198-nt *Bam*HI/*Kpn*I fragment from pDrepIS12.7gD was ligated into *Bam*HI/*Kpn*I-linearized pGEM4Z (Promega) to form pGEM4ZIS12.7. Thirty nucleotides of vector sequence in the multiple cloning region of pGEM4ZIS12.7 was removed by digestion with *Hin*dIII and *Bam*HI, filling in of the vector ends with DNA polymerase I Klenow fragment, and ligation of the ends to form pGEM4ZIS12.7EP.

Enzyme structure probing of RNA. The protocol for enzyme structure probing of RNA described previously (11) with modifications (16, 27, 50) was used. For in vitro synthesis of RNA, 10 μg of *Eco*RI-linearized, mung bean nuclease blunt-ended pGEM4ZIS12.7EP DNA was transcribed with 80 U of T7 RNA polymerase (Promega) in a 100-μl reaction mixture. The resultant 204-nt-long transcript included 11 nt of vector sequence at its 5′ terminus. The product was treated with RNase-free DNase (Promega), extracted with phenol-chloroform, then chloroform, chromatographed through a Biospin 6 column (Bio-Rad), spectrophotometrically quantitated, and stored in water at −20°C. Forty micrograms of RNA was heat denatured and renatured in a 400-μl reaction volume containing 30 mM Tris HCl (pH 7.5)–20 mM $MgCl_2$–300 mM KCl by heating to 65°C for 3 min and slow cooling (0.5 h) to 35°C. two micrograms of RNA was incubated in a 100-μl reaction volume containing 30 mM Tris HCl (pH 7.5)–20 mM $MgCl_2$–300 mM KCl, 5 μg of tRNA, and 0.0001, 0.001, 0.01, 0.1, or 0.5 U of RNase CV_1 (Kemotex Bio Ltd., Tallin, Estonia) or 0.1, 0.5, or 1.0 U of RNase T_1 (GIBCO). Reactions were performed at 25°C for 15 min and terminated by the addition of 150 μl of 0.5 M sodium acetate. RNA was extracted with phenol-chloroform, then chloroform, ethanol precipitated, redissolved, and used for primer extension with 5′-end-labeled minus-strand-binding oligonucleotide M13(−). Equal amounts of undigested RNA were used to generate a sequencing ladder with the same end-labeled oligonucleotide. The extended products were analyzed on a DNA sequencing gel of 6% polyacrylamide.

Northern assay for DI RNA replication and DI RNA-encoded sgmRNA synthesis. The Northern assay was performed essentially as previously described (12, 47). Briefly, cells at 80% confluency (10^6 cells) in 35-mm-diameter dishes were infected with BCoV at a multiplicity of 10 PFU per cell and transfected with 600 ng of transcript at 1 h postinfection (hpi). For passage of progeny virus, supernatant fluids were harvested at 48 hpi, and 500 μl was used to directly infect freshly confluent cells in a 35-mm dish. RNA extracted by the NP-40–proteinase K digestion method (approximately 10 μg per plate) was stored as an ethanol precipitate, and 2.5 μg per lane was used for electrophoresis in a formaldehyde-agarose gel. Approximately 1 ng of transcript was loaded per lane when used as a marker. Northern blots were probed with oligonucleotide gpD4(+), 5′-end labeled with ^{32}P to specific activities ranging from 1.5×10^5 to 3.5×10^5 cpm/pmol (Cerenkov counts), and exposed to Kodak XAR-5 film in the presence of an intensifying screen for 6 h to 5 days at −80°C.

Sequence analysis of progeny mRNAs. For direct sequencing of asymmetrically amplified cDNA, the procedure as described by Hofmann et al. (19) was used. For this, oligonucleotides 12.7gD(+) and leader(−) were used for reverse transcription-PCR (RT-PCR) with RNA extracted at 6 hpi from cells infected with the first-passage virus following transfection, and radiolabeled oligonucleotide 12.7gD(+) was used for sequencing.

For sequencing cDNA clones of progeny mRNA species, RNA was extracted at 6 hpi from cells infected with first-passage virus, and oligonucleotides 5′gD(+) and leader(−) were used for RT-PCR. Amplified fragments were cloned with the TOPO XL PCR cloning kit (Invitrogen), and dideoxynucleotide sequencing was done on purified DNA using oligonucleotide leader(−).

TABLE 1. Oligonucleotides used in this study

Oligonucleotide[a]	Polarity	Sequence (5′→3′)[b]	Binding region in pDrepIS12.7gD
Leader(−)	+	GAGCGATTTGCGTGCGTGCATCCCGC	7–32
NSI(−)	+	GTACACTTTCAGGTTTTGAG	1610–1629
BCV3′end(+)	−	TCGGCAATTACTTCCGCAAG	2345–2364
12.7-5′(−)	+	CCAATGCATGGATCCGGCTGTTCTATAGG	1660–1686
12.7V3-3′(+)	−	CCAATGCATCTGTTGTATAGAATGCTCTTCCTGGTCCTATATGTATACCGTTAGTATAACG	
12.7V3(+)	−	AGAATGCTCTTCCTGG	
12.7gD(+)	−	CATCCGCCAAGGCATATTTGGTACCGTTAGTA	1854–1885
5′gD(+)	−	GAGAGAGGCATCCGCCAAGGCATATTTG	1865–1893
GPD4(+)	−	CGATTCGGGTCGGCCATCTT	1894–1913
M1(−)	+	CCAATGCATGGATCCGGCTGTTCATAACCAAAACACACTAGCACC	1660–1702
M2(−)	+	CCAATGCATGGATCCGGCTGTTCATAACCAAAACACTGATCCTCCACTGG	1660–1707
M3(−)	+	CCATATTATAATTTACAGCTCACTTATAACTTTAAGC	1740–1776
M5(−)	+	GTTTTTCACGGTACTAGTTCTAAACCATATTATAATTTAG	1716–1755
M6(−)	+	AATTTAGGTAGACCAATATTCTTTAAGCATTATT	1748–1782
M7(−)	+	AATTTAGGTAGACGGTAGACCTTTAAGCATTATT	1748–1782
M8(−)	+	AATTTAGGTAGACCCCGCGGCTTTAAGCATTATT	1748–1782
M9(−)	+	GTTTTTCACGGTACTAGTGGTAGACCATATTATAATTTAGG	1716–1756
M10(−)	+	CCATATTATAATTTATCCAAACCTTATAACTTTAAGC	1740–1776
M12(−)	+	GTTTTTCACGGTACTAGTAGGTTTGCATATTATAATTTAG	1716–1755
M13(−)	+	GGTCACGCCCTAGTATTGGACATCTGGAGACCTG	1801–1834
M16(−)	+	CCATATTATAATTTAGGTAGACATTATGATAAATATTTTGGAGTAATGCCAAAGTTTC	1740–1797
M17(−)	+	CCATATTATAATTTAGGTAGACATTATGAGTAGTGTAACTACACCAGGCCAAAGTTTC	1740–1797
M47(−)	+	CCAAAGTTTCTAAGGGTAGACCCTAGTAATGGAC	1788–1821
M48(−)	+	CCAAAGTTTCTAAGGGGTAGACCTAGTAATGGAC	1788–1821
Rev(+)	−		

[a] The positive and negative symbols in the oligonucleotide names indicate the polarity of the nucleic acid to which the oligonucleotide anneals. Oligonucleotides M1(+), M2(+), M3(+), M5(+), M6(+), M7(+), M9(+), M10(+), M12(+), M13(+), M16(+), M17(+), M48(+), and M49(+) possess base sequences complementary to their respective minus-strand counterparts.

[b] Underlined bases represent mutated sites or restriction endonuclease sites used in cloning.

Synthetic oligonucleotides and accession numbers. The oligonucleotides used in this study are described in Table 1, and GenBank accession numbers for the sequences studied are M62375, M16620, and M30612 (1, 12, 29).

RESULTS

The upstream infrequently used canonical IS (site 1) for mRNA 5 synthesis is buried within the stem of a stem-loop as deduced from enzyme structure probing. Data showing that BCoV uses a downstream noncanonical IS as a fusion site for synthesis of mRNA 5 were derived from the sequencing of asymmetrically amplified PCR products of leader-mRNA body junction sequences from both the positive- and negative-strand RNA templates (19). Inasmuch as base pairing between the IS regions and the analogous region at the 3′ end of the 5′ genomic leader is an essential feature of the RdRp crossover event in discontinuous transcription (assuming the story is the same for coronaviruses as for arteriviruses [55]), accessibility of one strand for the other would be a presumed requirement. Curiously, the BCoV genome sequence in the region of the potential upstream IS fusion site for synthesis of mRNA 5 by both the Tinoco (51) and Zuker (57) algorithms is predicted to be within the stem of a stable stem-loop structure (the predicted Tinoco structure is shown in Fig. 1C, and the Zuker structure is shown in Fig. 2A). It therefore seemed that the helical region might inhibit the use of the upstream site for leader fusion, perhaps by impeding base pairing between the plus- and minus-strand elements. To test for the existence of the predicted helical region, enzyme structure probing was done on the isolated 199-nt-long IS-containing region (Fig. 2A and B). Whereas the double-stranded regions for the whole transcript identified by enzyme structure probing were in general more consistent with the structure predicted by Zuker than by Tinoco, the helical region surrounding site 1 was as predicted by both algorithms. That is, the bases immediately upstream of site 1 and the CAAAC within site 1 are part of a helical region as indicated by strong reactions with the single-strand-specific and double-strand-specific enzymes. The region immediately downstream of site 1 for a distance of 13 nt reacted as a single-stranded region as predicted by both algorithms. Site 2 appeared to be part of a double-stranded structure as well, although there is less agreement between the predicted and probed structures for this element. The sequence for a stretch of 16 nt downstream of site 2 appears to be mostly in a helical configuration.

Placement of the entire mRNA 5 wild-type (wt) IS region (199 nt) and a reporter sequence (92 nt) into the BCoV DI RNA led to transcription patterns indistinguishable from those directed by the BCoV genome. To test whether the helical region surrounding the canonical IS is an important factor in determining the use of site 1, we used the BCoV DI RNA system described earlier by Chang et al. (13) to examine the effects of mutations on subgenomic mRNA expression. This DI RNA has been successfully used to study subgenomic mRNA expression from a different set of ISs (26). When transcripts of pDrepIS12.7gD, the pDrep1 plasmid modified to carry the 199-nt IS region for the 12.7-kDa protein and the 92-nt HSV gD epitope reporter, were transfected into helper virus-infected cells, replication of the DI RNA genome in these and in cells infected with progeny virus appeared unimpaired relative to that of pDrep1 when evaluated by Northern

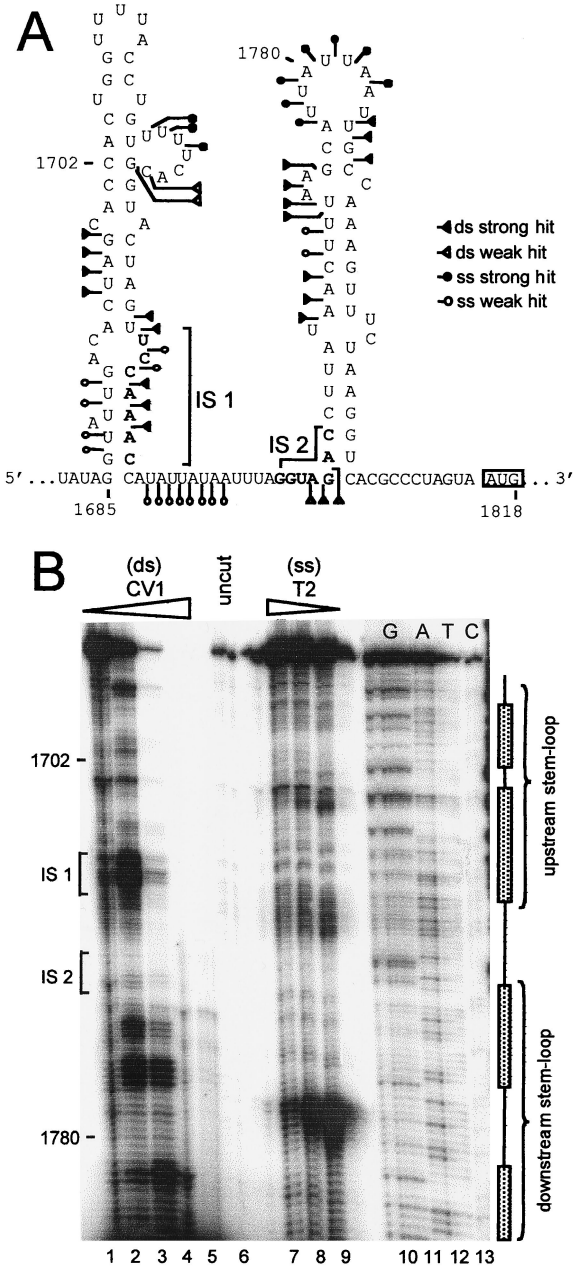

FIG. 2. Enzyme structure probing of the mRNA 5 IS region. (A) Predicted secondary structure of the mRNA 5 IS region by the Zuker algorithm and a summary of the single-stranded (ss) and double-stranded (ds) regions as determined by enzyme probing. The canonical upstream IS (site 1) and noncanonical downstream sequence (site 2) are shown in bold type. (B) Pattern of end-labeled extended primer after separation on a DNA sequencing gel. The 207-nt-long T7 RNA polymerase-generated transcript containing 193 nt of the 199-nt-long mRNA 5 IS region was treated with RNases as indicated, and a 5′ end-labeled oligonucleotide binding at the 3′ end of the transcript was used for primer extension. Lanes: 1 through 5, CV_1 digestion with 0.0001, 0.001, 0.01, 0.1, and 0.5 U/ml, respectively; 6, undigested RNA; 7 through 9, RNase T_2 digestion with 0.5, 0.1, and 1.0 U/ml, respectively; 10 through 13, sequencing ladder generated from the same transcript using the same primer. Base positions noted are those for the pDrepIS12.7 DI RNA. Positions of the IS1 and IS2 motifs are noted, as are the deduced stem-loop structures.

analysis with a probe specific for the DI RNA genome sequence (Fig. 3A, lanes 3 to 6 and 9 to 12, and data not shown). Furthermore, sgmRNA transcripts of the expected size were revealed by Northern analysis using an HSV-gD-specific radiolabeled probe (Fig. 3A, lanes 9 to 12), and sequencing of asymmetrically amplified cDNA derived from passage 1 virus showed the use of the noncanonical downstream (site 2) over the canonical promoter (site 1) as the site of leader fusion (Fig. 3B). The results were the same in preliminary experiments with the V3-containing RNA (data not shown). This picture, therefore, mimicked that observed for the virus genome (19). To evaluate this transcription pattern by examining individual transcripts, mRNA 5-specific cDNA products prepared from cells infected with passage 1 virus from pDrepIS12.7gD RNA were cloned and sequenced. These revealed that the predominant but not exclusive transcripts (8 of 10 clones) came from site 2, whereas 10% (1 in 10 clones) came from the upstream canonical site 1, and 10% (1 in 10 clones) came from a newly identified site, labeled site 3, just downstream of site 2 (Fig. 4). Thus, the DI RNA-derived construct appeared to mimic the virus genome with regard to the predominant transcription product from this region of the genome.

Mutations designed to make the canonical sequence (site 1) conform to the most common canonical motif (UCUAAAC), to unfold the upstream site-containing stem-loop, or to make the noncanonical downstream sequence (site 2) totally nonconforming failed to switch the leader fusion site from site 2 to site 1. To test the notion that the upstream canonical IS is not used because it fails to conform to the most common of the canonical motifs (UCUAAAC), site 1 in pDrepIS12.7gD was mutated to TCTAAAC to make mutant 5. Although mutant 5 showed wt levels of replication and sgmRNA synthesis as determined by Northern analysis, site 2 was still the primary fusion site used as determined by the sequencing of asymmetrically amplified cDNA (data not shown; Fig. 5; Table 2).

To test the notion that the upstream canonical motif is not used because it is buried within a stable stem and is therefore inaccessible for base pairing, three separate mutations that were predicted to unfold the stem were tested. These were present in mutants 1, 2, and 4 of pDrepIS12.7gD (Fig. 5). In mutant 1, nucleotide changes were made within the lower portion of the upstream side of the stem such that the downstream lower half of the stem would be expected to hold the UCCAAAC motif in a nonhelical region of the plus-strand RNA molecule. This prediction also holds true for the minus-strand equivalent of this structure. Transcripts of mutant 1 replicated as well as those of wt pDrep12.7ISgD, and an sgmRNA species was made, as evidenced by Northern analysis (Fig. 3A, lanes 33 to 36) and RT-PCR analysis (data not shown). Sequencing of asymmetrically amplified DNA, however, revealed that site 2 was still used as the fusion site for synthesis of sgmRNA (Fig. 5; Table 2). This was true as well in preliminary experiments with mutant 1 of the V3-containing plasmid (data not shown). Surprisingly, identical results were obtained for mutants 2 and 4 of the gD-containing constructs (Fig. 5 and Table 2; also data not shown) and in preliminary experiments with the V3 mutant 2- and mutant 4-containing constructs (data not shown). In mutant 2 the whole of the stem was predicted to unfold as a result of disrupted base pairing throughout the stem, and in mutant 4 the upstream side of the

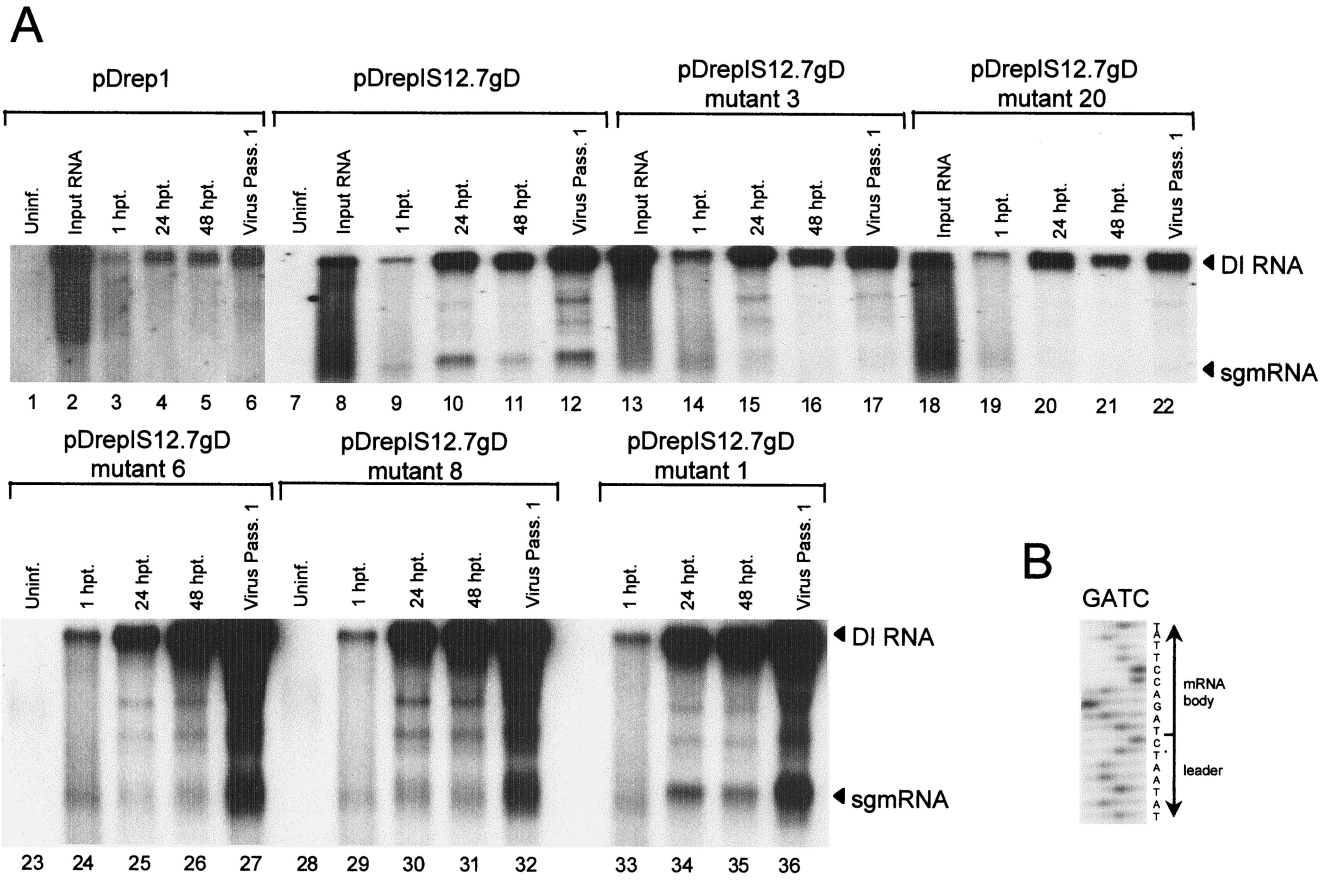

FIG. 3. Replication of DI RNA and synthesis of subgenomic mRNA transcripts. (A) Northern analysis showing the replication (accumulation) of transfected DI RNA transcripts and synthesis of sgmRNA at 1, 24, and 48 h postransfection (hpt) and in cells infected with passage 1 virus (Virus Pass. 1). Lanes: 3 through 6, RNA from pDrep1-transfected or VP1-infected cells probed with end-labeled TGEV reporter-detecting probe; 9 through 12, 14 through 17, 19 through 22, 24 through 27, and 29 through 36, RNA from cells transfected or infected as indicated were probed with end-labeled HSVgD reporter-detecting probe. Uninf., uninfected. (B) Sequence of asymmetrically amplified cDNA prepared from mRNA generated from pDrepIS12.7gD, virus passage 1 (panel A, lane 12). The junction of the leader and mRNA body is indicated.

stem and the entire loop were deleted. By both the Tinoco and Zuker algorithms, no new stable helical structures were predicted to arise in the region of site 1 as a result of these mutations. Thus, it appeared that the use of the upstream canonical site was not encouraged by an unfolding of the stem in which it is found.

To test the notion that site 2 is preventing the use of site 1 by locally domineering RdRp behavior, site 2 was made totally nonconforming by converting GGTAGAC in pDrepIS12.7gD to CAGCTCA, making mutant 3. Transcripts of mutant 3 underwent wt levels of replication as determined by Northern analysis, but surprisingly, by both Northern analysis and RT-PCR designed to amplify HSVgD sequence-containing sgmRNAs, there was no evidence of sgmRNA synthesis (Fig. 3A, lanes 14 to 17; RT-PCR data not shown). Thus, nonuse of site 1 cannot be attributed to a preemptive use of site 2 by the RdRp. When the nonconforming sequence in site 2 of mutant 3 was combined with the contextual changes at site 1 of mutant 4, making mutant 20, there was still no sgmRNA synthesis in the presence of wt levels of genome replication as determined by Northern analysis (Fig. 3A, lanes 19 to 22). Small amounts of sgmRNA synthesis for mutant 20 could be detected by RT-PCR, however, but no crossovers at sites 2 or 1 were detected, and sgmRNA synthesis occurred from a heretofore-unrecognized site beginning immediately downstream of site 3 (Fig. 5; Table 2). The IS in this instance, termed site 4, has the sequence UUUAAGC, in which five of the seven bases conform to the consensus IS motif (nonconforming bases are underlined).

The upstream canonical IS does not function ectopically to cause transcription from the downstream noncanonical site, nor does the initiation codon for the 12.7-kDa protein influence IS usage. Since the five separate mutations described above that were made within the regions of site 1, the stem-loop, and site 2 failed to cause a switch from site 2 to site 1 as hypothesized, other factors within the IS-containing region were sought that might explain the heavy use of noncanonical site 2. How can it be that an IS in which three out of the seven nucleotides are nonconforming show such strong fusion activity? Three possibilities were tested. The first was that the upstream conforming IS is working at a distance to cause a polymerase strand transfer at the downstream site. To test this, site 1 was changed to a totally nonconforming sequence, AGGUUUG, creating mutant 12. Whereas transfected RNA

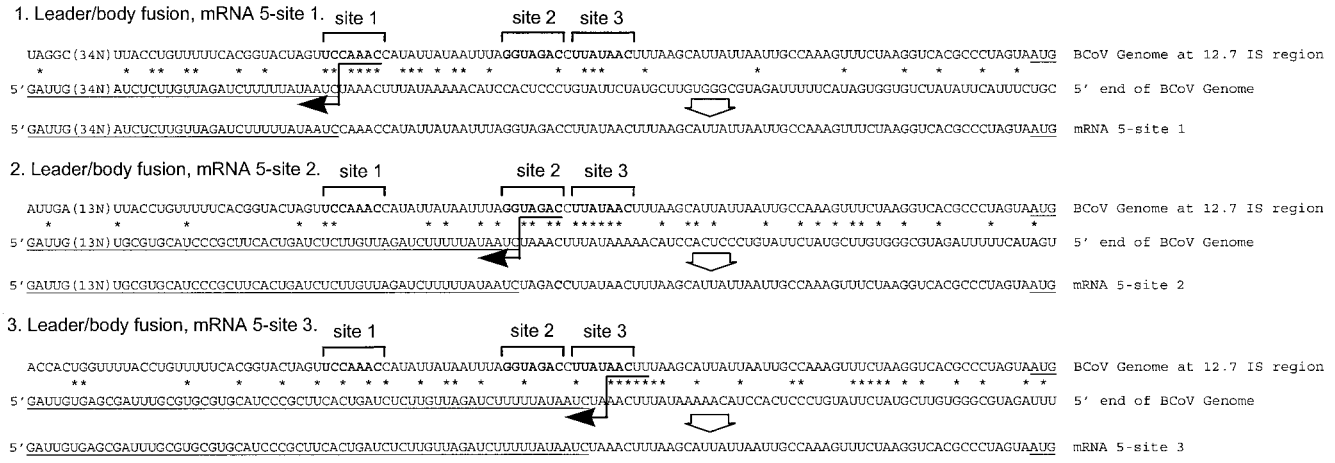

FIG. 4. The three types of wild-type fusion products found for mRNAs generated from pDrepIS12.7gD. Sequences were derived from cloned RT-PCR products of mRNA-leader fusion regions. Asterisks indicate positions of base identity with the aligned 5′ terminus of the virus genome.

of mutant 12 replicated at wt levels based on Northern analysis, transcription was also at wt levels, and the predominant fusion site was site 2, as learned from the sequence of asymmetrically amplified cDNA (Fig. 5; Table 2). Thus, site 1 is not working ectopically to deliver competence to site 2.

To test the possibility that the 12.7-kDa protein ORF start codon somehow gives direction for the synthesis of mRNA from site 2, the start codon was mutated from AUG to UUG, thus creating mutant 13. No differences in replication or sgmRNA synthesis patterns were observed between mutant 13 and wt pDrepIS12.7gD (data not shown), indicating that the start codon for the 12.7-kDa protein ORF plays no role in directing the transcription event (Fig. 5; Table 2). This is consistent with the results of preliminary experiments with pDrepIS12.7V3 and its mutants 1, 3, and 4, for which subgenomic transcripts were generated but none of which had a reporter ORF in-frame with the 12.7-kDa protein ORF (data not shown).

Mutations made downstream of the downstream noncanonical site 2 in DI RNA, designed to decrease sequence similarities with the genome 5′ end, caused leader fusion to take place at the upstream canonical site 1. To test the third possibility, that sequences flanking the downstream noncanonical IS contribute to RdRp strand switching at site 2 through a sequence similarity between putative donor and acceptor templates in this region, two other mutants were tested. The idea that flanking sequences might contribute to similarity-assisted strand transfer stems from two sets of observations: (i) earlier work of members of our group (13) and the work of others (60) showing that downstream flanking sequences near the genomic leader junction are important in directing polymerase crossover during high-frequency leader switching on genomic DI RNA (36) and (ii) the observation that a general clustering of 9 to 17 identical bases occurs within a 22-nt-long stretch immediately surrounding the eight ISs in the BCoV genome (Fig. 6). Especially noteworthy in the second case is the grouping of 5 to 7 nt within the downstream flanking 10 bases of the abundantly produced mRNAs 2, 2-1, 5, 5-1, 6, and 7.

Since the region of high-frequency strand crossover during DI RNA leader switching occurs in an 8-nt AU-rich sequence of perfect identity just downstream of the genomic leader (13),

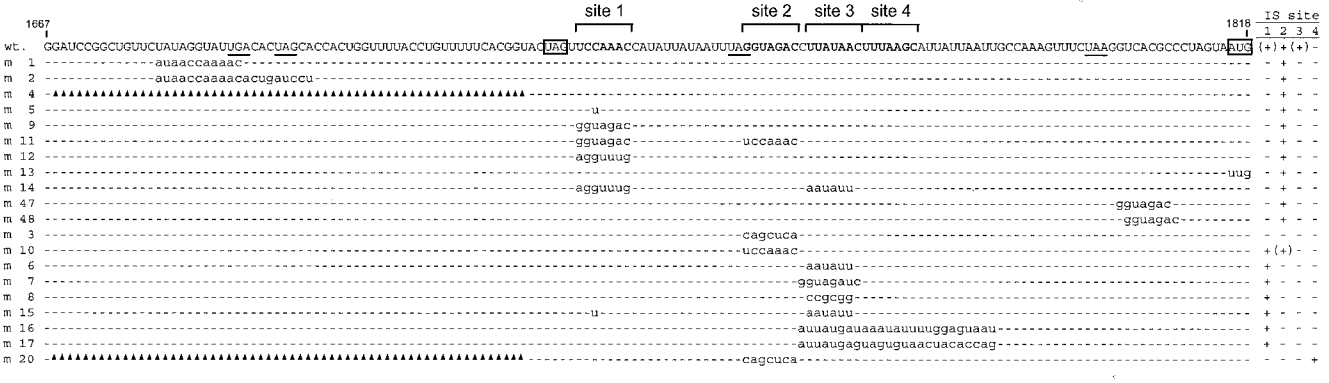

FIG. 5. Summary of mutations made within the reporter-expressing BCoV DI RNA engineered to synthesize mRNA from the mRNA 5 fusion sites (pDrepIS12.7gD). Sites 1, 2, 3, and 4 (in bold) are IS motifs found to function as fusion sites in either the wild type or mutant constructs as noted in the column at the right. Sites identified by parentheses are minor sites. The stop codon for gene 4–1 and the start codon for gene 5 are boxed. The four in-frame stop codons between the DI-RNA ORF and the 12.7-kDa protein ORF are underlined. Sequence numbering refers to base positions in pDrepIS12.7gD. m, mutant.

TABLE 2. Summary of mutation results[a]

Construct	No. of clones from intergenic site/total			Comments
	Site 1	Site 2	Site 3	
wt	1/10	8/10	1/10	By asymmetric RT-PCR sequencing, all were from site 2
Mut. 1		All		Asymmetric RT-PCR sequencing
Mut. 2		All		Asymmetric RT-PCR sequencing
Mut. 4		All		Asymmetric RT-PCR sequencing
Mut. 5		All		Asymmetric RT-PCR sequencing
Mut. 9		9/9		
Mut.10	3/8	2/8	1/8	1/8 crossed over at the ATG start site; 1/8 appeared to have crossed over onto mRNA 5 II (i.e., it had a UCCAAAC junction sequence)
Mut.11		5/6	1/6	
Mut.12		All		Asymmetric RT-PCR sequencing
Mut.13		All		Asymmetric RT-PCR sequencing
Mut.14		1/2		1/2 appeared to have crossed over between sites 1 and 2 onto a sgmRNA template (i.e., it had a UCCAAAC junction sequence)
Mut.47	1/6	5/6		
Mut.48		5/6	1/6	
Mut. 3				No sgmRNAs
Mut. 6	8/11		1/11	1/11 from sgmRNA 61 nt upstream of site 1; 2/11 came from within the site 3 region and may have crossed over onto a sgmRNA (i.e., they had junction sequences of UCCAAAC)
Mut. 7	3/3			
Mut. 8	4/4			
Mut.15	7/7			
Mut.16	5/6			1/6 had a leader and nonviral sequence
Mut.17	6/10			3/10 were 9.6 mRNA sequence; 1/10 had a leader and nonviral sequence
Mut.20				4/5 crossed over at TTTAAGC immediately downstream of site 3

[a] All RNA used for sequence analyses was derived from cells infected with first-passage virus. Mut., mutant.

we first chose to test a mutant that maintained AU richness but in which base identity was disrupted in the 6 nt mapping just downstream of site 2. For this we made mutant 6, in which AAUAUU replaced the wt sequence UUAUAA (Fig. 5). Mutant 6 underwent replication and supported sgmRNA synthesis (Fig. 3A, lanes 24 to 27), but, surprisingly, no transcripts were made from site 2 and almost all were made from the upstream canonical site 1. Of the 11 clones sequenced, 8 came from site 1, 1 from a novel site located 61 nt upstream from site 1, and 2 from novel sites located at positions of 1 and 2 nt downstream of site 2, respectively (Table 2 and described below). Curiously, the two clones coming from within site 3 had the fusion sequence UCCAAAC, which suggests the use of a leader template other than that on the virus genome (see below). Thus, changing the downstream flanking sequence of the downstream noncanonical IS in the context of the wt upstream stem resulted in a switch in fusion sites from site 2 to site 1.

To test whether decreasing sequence similarity with the genome 5′ end in this downstream flanking region is, independent of its AU richness, important for this result, the 6-nt downstream flanking sequence of promoter 2 was made GC rich by replacing nucleotides UUAUAA in the wt sequence with CCGCGG at site 3, creating mutant 8. Transcripts of mutant 8 underwent replication and supported sgmRNA synthesis (Fig. 3A, lanes 29 to 32), and in a pattern almost like that of mutant 6, all of the sgmRNAs came from site 1 (Fig. 5; Table 2). Thus, a decrease in base similarity in the downstream flanking region of site 2 with the analogous region of the genome 5′ end, whether it be AU or GC rich, discouraged the use of fusion site 2 and encouraged the use of site 1. Nucleotide sequence similarities in the downstream flanking region with the putative genomic leader template, therefore, appeared in this instance to be a decisive factor in determining where the RdRp will switch strands.

In a separate set of experiments, the GGUAGAC noncanonical heptameric sequence, which occurs only once in the entire BCoV genome (K. Nixon and D. Brian, unpublished data), was placed into pDrepIS12.7gD at different positions to determine if this motif alone could function independently as a strong fusion site. For this, GGUAGAC was positioned at site 1 and at sites beginning at base positions 7, 47, and 48 nt downstream of site 2 by mutating the natural bases in these regions, thus forming mutants 9, 7, 47, and 48, respectively (Fig. 5). Transcription patterns from mutants 9, 47, and 48 were very similar to those with wt pDrepIS12.7gD, wherein nine of nine sequenced clones for mutant 9 came from site 2, five of six clones for mutant 47 came from site 2 and one came from site 1, and five of six clones for mutant 48 came from site 2 and 1 from site 3 (Table 2). Likewise, clones from mutant 11, which combined the mutations of mutants 9 and 10, also came predominantly (five of six clones) from site 2. Mutant 7, however, did not function as a crossover site but rather, as with mutants 6 and 8, caused the crossover to shift to the canonical motif at site 1. Thus, the GGUAGAC motif does not appear to function as an independent fusion site in all sequence contexts. Furthermore, since the results for mutant 7 are the same as those for mutants 6 and 8, we would postulate that the same mechanism is involved. That is, with mutant 7 there is a resultant decrease in nucleotide sequence similarity with the immediate downstream flanking sequence of the putative genomic leader template.

Can the use of the 5′ termini of sgmRNA molecules 5-1 and 6 be induced as acceptors for the RdRp jump by making the

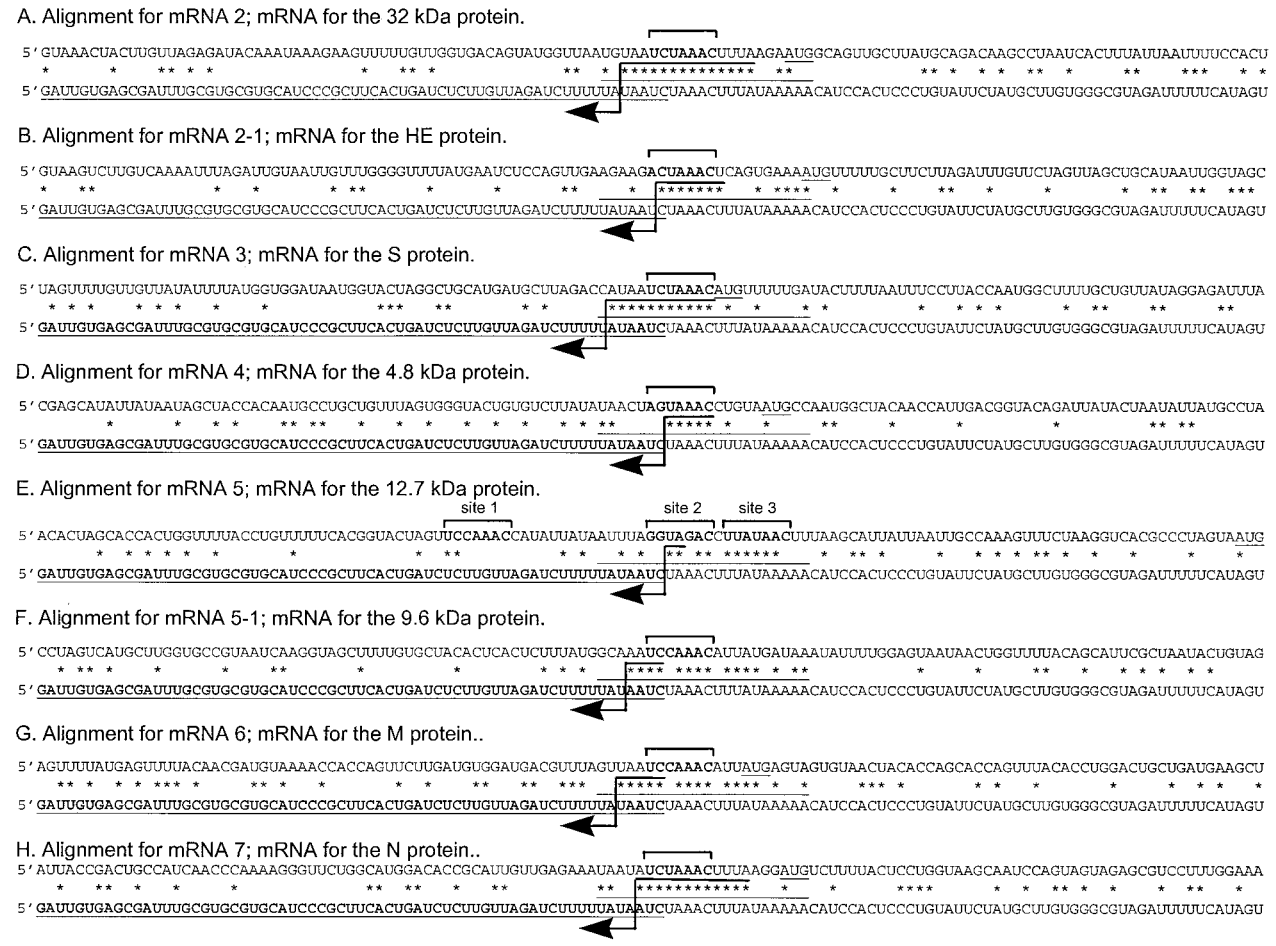

FIG. 6. Alignment of identified fusion sites in the BCoV genome with the 5′ end of the virus genome and identity of clustered base similarities within the 22-nt region immediately surrounding the ISs. The 22-nt region (underlined) includes the heptameric IS (in bold type) and the immediate 5 upstream and 10 downstream nucleotides. Postulated RdRp strand switching during minus-strand synthesis is indicated by the arrow. Strand switching theoretically could occur just upstream of any of the consecutive bases overlined by the tail of the arrow. For example, for mRNA 2, strand switching could occur at any of the 14 sites (bases) and yield the same sgmRNA species.

downstream flanking 25 nt in DI RNA identical to the 5′-proximal sequences on these sgmRNA species? Inasmuch as the coronavirus genome alone is sufficient to initiate infection (7, 44, 59), the 5′-terminal genomic leader sequence must initially be the only potential acceptor molecule for the RdRp jump were it to happen during minus-strand synthesis (42). Soon after infection is established, however, sgmRNAs become much more abundant than virus genome and are theoretically available as acceptor molecules for the RdRp jump during minus-strand synthesis (20, 47). The comigration of genome-size and subgenome-size replicative intermediates in fractionated replication complexes (45) is consistent with this idea, although it is not yet known if the two types of double-stranded intermediates (genomic and subgenomic) are physically associated within a common RNA-synthesizing complex. Intriguingly, four separate clones in our experimental results had leader mRNA body junction sequences of UCCAAACC, which is a result (i.e., a C in the third position, underlined) that is difficult to explain by a copy choice event involving only the leader junction sequence on the genomic leader (a sequence of UCUAAAC), since the C in the third position also could not have been donated by the genomic donor strand. One example is depicted in Fig. 7A which comes from clone #8 of mutant 14. The only plausible explanation, barring a PCR artifact, is that the RdRp jumped to a molecule of mRNA 5-site 1 (i.e., the mRNA from gene 5 using site 1 for the 12.7-kDa protein; Fig. 4), mRNA 5-1 (mRNA for the E protein; Fig. 6), or mRNA 6 (mRNA for the M protein; Fig. 6), all of which have UCCAAAC as the leader junction sequence. Alternatively, since minus-strand copies of these mRNA species exist (19, 20) that could in theory have served as templates for synthesis of mRNA plus-strand copies by way of a mispriming at the 3′ terminus of the minus strand by the gD primer in the RT reaction, artifactual clones could have arisen. Subsequent amplification in this case could have then been carried out with the intended 3′ and 5′ primers to ultimately yield an artifactual clone. A second alternative explanation is that the C in the third position arose from a PCR-generated mutation.

To try to experimentally induce a copy choice crossover that would utilize a leader (and possibly leader junction) template

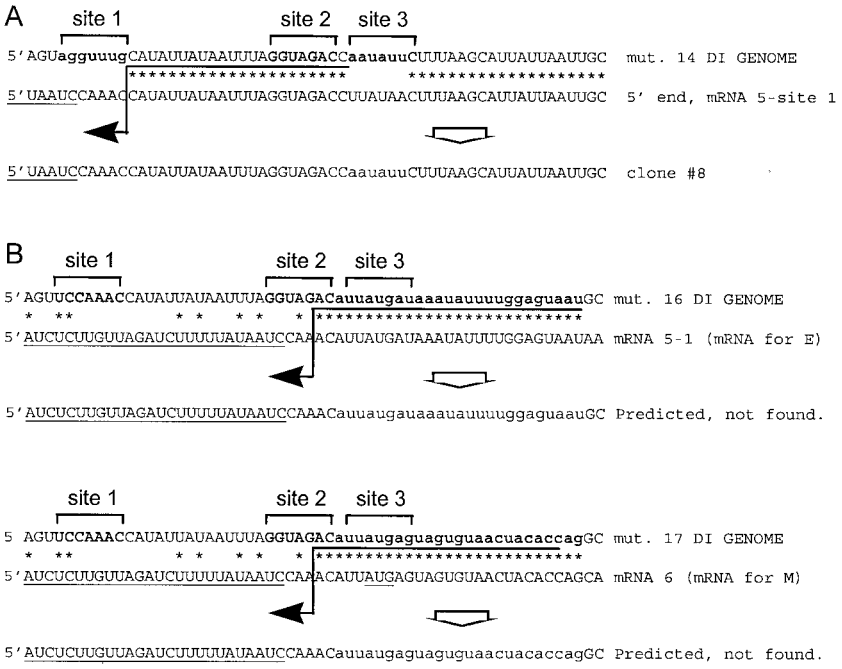

FIG. 7. Possible use of sgmRNAs as acceptor molecules for the RdRp crossover. (A) sgmRNA 5 site 1 could have theoretically served as an acceptor molecule for the RdRp jump from mutant 14 DI RNA, creating the reporter-containing recombinant clone #8 as depicted. The virus genomic leader with its junction sequence of UCUAAAC could not have given rise to clone #8 with a junction sequence of UCCAAAC as depicted. (B) The sgmRNA molecules predicted to arise from DI RNA mutants 16 and 17. In mutant 16, a region of high identity with the 5′ end of mRNA 5–1 (mRNA for E protein) and in mutant 17, a region of high identity with the 5′ end of mRNA 6 (mRNA for M protein) were made and tested for leader fusion. The crossover could have occurred anywhere within the region of continuous asterisks to generate the predicted transcripts.

source on sgmRNA, molecules of two designs were used. Mutants 16 and 17 were made in which regions of 25 nt in length just downstream of site 2 maintained an identity of 100% with the 5′-end regions of mRNAs 5-1 and 6, respectively (Fig. 7B). The rationale was that with enhanced sequence similarity, a leader-generating RdRp jump with these mRNAs serving as receptor species at site 2 would also be enhanced. Whereas replication and sgmRNA synthesis from both mutants were at wt levels as determined by Northern analysis, the only sgmRNAs found were those resulting from a fusion at site 1. Thus, induced fusion with mRNAs 5-1 and 6 was not accomplished by these mutations, and a movement of crossover from site 2 to site 1 in reporter-containing mRNA 5 molecules was obtained. Since the base changes in mutants 16 and 17 had the result of decreasing sequence similarities with the virus genome 5′ end, the switch to site 1 may have shared mechanistic features with the process observed for mutants 6, 7, and 8.

DISCUSSION

The importance of flanking sequences in coronavirus sgmRNA synthesis. In this study, we have analyzed the question of how it is that a noncanonical IS motif can be chosen over a closely positioned upstream canonical site during transcription in the BCoV in the generation of sgmRNAs. The study differs from others using a DI RNA reporting system in that here two naturally occurring, closely positioned IS motifs yielding an unusual transcription pattern were used for mutational studies. We have confirmed previous reports on the properties of coronavirus ISs by showing that flanking sequences can critically influence the strength of a given fusion site (2, 22, 24, 25, 34, 35, 53, 56). We have extended these observations in the following ways. (i) We have shown that sequences downstream (in the 3′ direction) of the IS motif in the genome can exert a decisive influence on which of two nearby upstream promoters will be used, even overpowering sequence similarities within the heptameric IS itself. In many ways, this parallels the influences on leader switching at the 5′ end of the DI RNA genome for which sequences downstream of the heptameric motif, primarily AU rich in BCoV (13) and MHV (60), were shown to be a region of crossover for the RdRp during leader switching (13). (ii) We present evidence suggesting that secondary structure surrounding an IS motif might not be a primary determinant in the choice of that site for leader fusion (discussed further below). (iii) We present evidence supporting the notion (42) that sgmRNAs rarely if ever serve as receptor templates for the RdRp jump and are thus, as a consequence, rarely if ever indirectly the source of leader sequence on sgmRNAs as it has been postulated (13, 26). If sgmRNAs were to serve as templates, we would first expect a large number of cloned products in our experiments to be recombinant molecules between the HSVgD reporter and site 2-specific mRNA 5 leader junction sequences, since (i) wt site 2-specific mRNA 5 molecules are the predominant mRNA 5 species in virus-infected cells (19), and (ii) extensive base pairing would exist between these templates. None were found. Second, in experiments with mutants 16 and 17, no

HSVgD-containing clones looking like recombinants with these messages were found despite an extensive region (25 nt) of base identity with mRNAs 5-1 (encoding the E protein) and 6 (encoding the M protein).

Does secondary structure influence the polymerase crossover site? The surprise to us in these studies was that we were unable by relieving the putative helical structure surrounding the upstream site to cause a switch in crossover to this site. The speculation that this would happen was based on the knowledge that base pairing within the intergenic region is part of the signal promoting the crossover event, as shown in several studies with coronavirus DI RNAs (17, 22, 35, 56) and with the infectious cDNA clone for the arterivirus genome (55). Presumably, since the IS is within the loop of a stem-loop in the 3′ flanking region of the BCoV genomic leader (13) and in the analogous region near the arterivirus leader (55), a linear region of the RNA would facilitate the base pairing. The linear nature of the upstream IS in mutants 1, 2, and 4, however, is only a prediction, and physical mapping studies may reveal differences from these predictions. Sequences in the region of the upstream canonical site no doubt play an important role in the choice of fusion sites, however, since the combined set of mutations in mutant 20 caused a whole new junction sequence to appear downstream of site 3, recognized as site 4. Clearly, what determines the availability of nucleotides for base pairing remains to be fully determined.

How do flanking sequence similarities contribute to the decision of where the RdRp will jump? We think the data reported here are most consistent with the model of similarity-assisted RdRp strand transfer (RNA recombination) during minus-strand synthesis (8, 40, 42), and the figures throughout the paper are drawn with this model in mind. We think this model is consistent with the one presented by Chang et al. (13), wherein the RdRp jump takes place during minus-strand synthesis in the process of leader switching on DI RNA, and base pairing downstream of the heptameric IS contributes to, and in some cases solely determines, the polymerase strand switching event. In the model of Chang et al., the RdRp crossover can take place within or very near the heptameric IS, and the only potential templates for leader exist on genome or sgmRNAs, since RNase protection experiments showed no evidence of a "free" leader. As drawn in Fig. 6, the jump during transcription (or the events leading up to transcription) could take place anywhere within the region of the solid asterisks, which in the cases of mRNAs 2 and 7 could extend 4 bases downstream of the IS, or in the cases of mRNAs 2, 3, 5-1, 6, and 7 could extend 3, 4, 2, 3, and 1 base upstream of the IS, respectively. Thus, in the case of leader switching on the DI RNA 5′ end, an extension of the model of Sawicki and Sawicki (40, 42), the RdRp jump could occur well downstream of the heptameric IS in a region of AU richness (8, 13), whereas in the RdRp jump at site 2 for mRNA 5 synthesis, the jump occurs within the IS region but is decisively influenced by the homology with the AU-rich downstream flanking sequence. It can be envisioned, therefore, consistent with the Sawicki model, that the nascent minus strand made during minus-strand synthesis is temporarily encouraged to separate from its template strand (i.e., breathe) and switch to an analogous region on another molecule. Our data support the model of the Sawickis (42) too in that few, if any, sgmRNAs appear to serve as acceptor molecules for the RdRp during the jump. More work is required, however, to determine the origin of the sgmRNAs that appeared to have been derived from an RdRp jump to another sgmRNA (Fig. 7A).

A number of recent studies have suggested that the ISs in the plus or minus strand sense, in addition to being regions of base pairing, are also motifs to which viral or cellular proteins attach to bring the points of fusion into close proximity, thus facilitating an RdRp jump (references 30 and 31 and references therein). In this regard, the cellular protein Hn RNP-A1 has been demonstrated to bind to the 5′UUUAG3′ motif, a motif found within the minus-sense ISs in MHV (31). Inasmuch as BCoV and MHV share the same consensus ISs in the minus strand (GUUUA/GGA), the same mechanism might be postulated to extend to BCoV. If the proposed protein binding mechanism is correct, then the experiments presented here would suggest that the underlined bases within the sequence 5′ GUCUACC3 would also suffice as an Hn RNP-A1 binding site. Likewise, the UUA sequence in the minus-sense strand of the IS for a recently described noncanonical site for MHV sgmRNA synthesis (61) should also suffice. These remain to be shown.

What is the biological significance of a functional but noncanonical IS motif in coronavirus transcription? Alternative, noncanonical fusion sites have been identified in the equine arteritis arterivirus and have been shown, by mutational analysis of the infectious cloned genome, to be important in the regulation of gene expression for optimal viral growth (38). At this time, since no comparable infectious clone for BCoV exists, we are unable to examine the question by the same experimental approach. We note, however, that the use of the downstream noncanonical site is, although preferred, an alternative to the canonical site and thus may be playing a heretofore undetermined regulatory role important to the BCoV life cycle. We also note that gene 5 of HECV 4408F92, a close relative of the BCoV (62), uses an identical downstream noncanonical site (H.-Y. Wu, A. Ozdarendeli, and D. A. Brian, unpublished data), which suggests an evolutionary pressure for retention of the noncanonical transcription motif in these two viruses.

ACKNOWLEDGMENTS

We thank Jennifer O'Connor, Cary Brown, Ruey-Yi Chang, Kimberley Nixon, Shamila Raman, Hung-Yi Wu, and David Hacker for valuable discussions. We thank Sharmila Raman for incorporating the noted modifications in the enzyme structure probing protocol.

This work was supported by Public Health Service grant AI14267 from the National Institute of Allergies and Infectious Diseases, grant 92-37204-8046 from the U.S. Department of Agriculture, and by funds from the University of Tennessee, College of Veterinary Medicine, Center of Excellence Program for Livestock Diseases and Human Health. A.O. was supported in part by a stipend grant from the Turkiye Bilim Teknik ve Arastirma Kurumu (TUBITAK), Government of Turkey, and S.R. was supported by a stipend grant from the Swiss National Science Foundation.

REFERENCES

1. **Abraham, S., T. E. Kienzle, W. Lapps, and D. A. Brian.** 1990. 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.
2. **An, S., and S. Makino.** 1998. Characterization of coronavirus cis-acting RNA elements and the transcription step affecting its transcription efficiency. Virology **243:**198–207.
3. **Baric, R. S., C.-K. Shieh, S. A. Stohlman, and M. M. C. Lai.** 1987. Analysis

of intracellular small RNAs of mouse hepatitis virus: evidence for discontinuous transcription. Virology **156:**342–354.
4. **Baric, R. S., S. A. Stohlman, and M. M. C. Lai.** 1983. Characterization of replicative intermediate RNA of mouse hepatitis virus: presence of leader RNA sequences on nascent chains. J. Virol. **48:**633–640.
5. **Baric, R. S., and B. Yount.** 2000. Subgenomic negative-strand RNAs function during mouse hepatitis virus infection. J. Virol. **74:**4039–4046.
6. **Bowen, J. C., O. Alpar, R. Phillpotts, I. S. Roberts, and M. R. W. Brown.** 1990. Preliminary studies on infection by attenuated *Salmonella* in guinea-pigs and on expression of herpes simplex virus. Res. Microbiol. **141:**873–877.
7. **Brian, D. A., D. E. Dennis, and J. S. Guy.** 1980. Genome of porcine transmissible gastroenteritis virus. J. Virol. **34:**410–415.
8. **Brian, D. A., and W. J. M. Spaan.** 1997. Recombination and coronavirus defective interfering RNAs. Semin.Virol. **8:**101–111.
9. **Budzilowicz, C. J., S. P. Wilczynski, and S. R. Weiss.** 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.
10. **Cavanagh, D.** 1997. Nidovirales: a new order comprising Coronaviridae and Arteriviridae. Arch. Virol. **142:**629–633.
11. **Chang, R.-Y., and D. A. Brian.** 1996. *cis* requirement for N-specific protein sequence in bovine coronavirus defective interfering RNA replication. J. Virol. **70:**2201–2207.
12. **Chang, R.-Y., M. A. Hofmann, P. B. Sethna, and D. A. Brian.** 1994. A *cis*-acting function for the coronavirus leader in defective interfering RNA replication. J. Virol. **68:**8223–8231.
13. **Chang, R.-Y., R. Krishnan, and D. A. Brian.** 1996. The UCUAAAC promoter motif is not required for high-frequency leader recombination in bovine coronavirus defective interfering RNA. J. Virol. **70:**2720–2729.
14. **de Vries, A. A. F., M. C. Horzinek, P. J. M. Rottier, and R. J. de Groot.** 1997. The genome organization of the Nidovirales: similarities and differences between arteri-, toro-, and coronaviruses. Semin. Virol. **8:**33–47.
15. **Fischer, F., C. F. Stegen, C. A. Koetzner, and P. S. Masters.** 1997. Analysis of a recombinant mouse hepatitis virus expressing a foreign gene reveals a novel aspect of coronavirus transcription. J. Virol. **71:**5148–5160.
16. **Hartshorne, T., and N. Agabian.** 1994. A common core structure for U3 small nucleolar RNAs. Nucleic Acids Res. **22:**3354–3364.
17. **Hiscox, J. A., K. L. Mawditt, D. Cavanagh, and P. Britton.** 1995. Investigation of the control of coronavirus subgenomic mRNA transcription by using T7-generated negative-sense RNA transcripts. J. Virol. **69:**6219–6227.
18. **Hofmann, M. A., and D. A. Brian.** 1991. The 5′ end of coronavirus minus-strand RNAs contains a short poly(U) tract. J. Virol. **65:**6331–6333.
19. **Hofmann, M. A., R.-Y. Chang, S. Ku, and D. A. Brian.** 1993. Leader-mRNA junction sequences are unique for each subgenomic mRNA species in the bovine coronavirus and remain so throughout persistent infection. Virology **196:**163–171.
20. **Hofmann, M. A., P. B. Sethna, and D. A. Brian.** 1990. Bovine coronavirus mRNA replication continues throughout persistent infection in cell culture. J. Virol. **64:**4108–4114.
21. **Horton, R. M., C. Zeling, N. H. Steffan, and L. R. Bease.** 1990. Gene splicing by overlap extension: tailor-made genes using polymerase chain reaction. BioTechniques **8:**528–535.
22. **Hsue, B., and P. S. Masters.** 1999. Insertion of a new transcriptional unit into the genome of mouse hepatitis virus. J. Virol. **73:**6128–6135.
23. **Jeong, Y. S., and S. Makino.** 1992. Mechanism of coronavirus transcription: duration of primary transcription initiation activity and effect of subgenomic RNA transcription on RNA replication. J. Virol. **66:**3339–3346.
24. **Jeong, Y. S., J. F. Repass, Y. N. Kim, S. M. Hwang, and S. Makino.** 1996. Coronavirus transcription mediated by sequences flanking the transcription consensus sequence. Virology **217:**311–322.
25. **Joo, M., and S. Makino.** 1995. The effect of two closely inserted transcription consensus sequences on coronavirus transcription. J. Virol. **69:**272–280.
26. **Krishnan, R., R.-Y. Chang, and D. A. Brian.** 1996. Tandem placement of a coronavirus promoter results in enhanced mRNA synthesis from the downstream-most initiation site. Virology **218:**400–405.
27. **Krol, A., and P. Carbon.** 1989. A guide for probing native small nuclear RNA and ribonucleoprotein structures. Methods Enzymol. **180:**212–227.
28. **Lai, M. M. C., and D. Cavanagh.** 1997. The molecular biology of coronaviruses. Adv. Virus Res. **48:**1–100.
29. **Lapps, W., B. G. Hogue, and D. A. Brian.** 1987. Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. Virology **157:**47–57.
30. **Li, H.-P., P. Huang, S. Park, and M. M. C. Lai.** 1999. Polypyrimidine tract-binding protein binds to the leader RNA of mouse hepatitis virus and serves as a regulator of viral transcription. J. Virol. **73:**772–777.
31. **Li, H.-P., X. Xhang, R. Duncan, L. Comai, and M. M. C. Lai.** 1997. Heterogeneous nuclear ribonucleoprotein A1 binds to the transcription-regulatory region of mouse hepatitis virus RNA. Proc. Natl. Acad. Sci. USA **94:**9544–9549.
32. **Li, S., V. Polonis, H. Isobe, H. Zaghouani, R. Guinea, T. Moran, C. Bona, and P. Palese.** 1993. Chimeric influenza virus induces neutralizing antibodies and cytotoxic T cells against human immunodeficiency virus type 1. J. Virol. **67:**6659–6666.
33. **Liao, C. L., and M. M. C. Lai.** 1994. Requirement of the 5′-end genomic sequence as an upstream *cis*-acting element for coronavirus subgenomic mRNA transcription. J. Virol. **68:**4727–4737.
34. **Makino, S., and M. Joo.** 1993. Effect of intergenic consensus sequence flanking sequences on coronavirus transcription. J. Virol. **67:**3304–3311.
35. **Makino, S., M. Joo, and J. K. Makino.** 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.
36. **Makino, S., and M. M. C. Lai.** 1989. High-frequency leader sequence switching during coronavirus defective interfering RNA replication. J. Virol. **63:**5285–5292.
37. **Masters, P. S., C. A. Koetzner, C. A. Kerr, and Y. Heo.** 1994. Optimization of targeted RNA recombination and mapping of a novel nucleocapsid gene mutation in the coronavirus mouse hepatitis virus. J. Virol. **68:**328–337.
38. **Pasternak, A. O., A. P. Gultyaev, W. J. M. Spaan, and E. J. Snijder.** 2000. Genetic manipulation of arterivirus alternative mRNA leader-body junction sites reveals tight regulation of structural protein expression. J. Virol. **74:**11642–11653.
39. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
40. **Sawicki, D. L., T. Wang, and S. G. Sawicki.** 2001. The RNA structures engaged in replication and transcription of the A59 strain of mouse hepatitis virus. J. Gen. Virol. **82:**385–396.
41. **Sawicki, S. G., and D. L. Sawicki.** 1990. Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in mRNA synthesis. J. Virol. **64:**1050–1056.
42. **Sawicki, S. G., and D. L. Sawicki.** 1995. Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands. Adv. Exp. Med. Biol. **380:**499–506.
43. **Schaad, M. C., and R. S. Baric.** 1994. Genetics of mouse hepatitis virus transcription: evidence that subgenomic negative strands are functional templates. J. Virol. **68:**8169–8179.
44. **Schochetman, G., R. H. Stevens, and R. W. Simpson.** 1977. Presence of infectious polyadenylated RNA in the coronavirus avian bronchitis virus. Virology **77:**772–782.
45. **Sethna, P. B., and D. B. Brian.** 1997. Coronavirus subgenomic and genomic minus-strand RNAs are found in N protein-deficient, membrane-protected replication complexes. J. Virol. **71:**7744–7749.
46. **Sethna, P. B., M. A. Hofmann, and D. A. Brian.** 1991. Minus-strand copies of replicating coronavirus mRNAs contain antileaders. J. Virol. **65:**320–325.
47. **Sethna, P. B., S.-L. Hung, and D. A. Brian.** 1989. Coronavirus subgenomic minus-strand RNA and the potential for mRNA replicons. Proc. Natl. Acad. Sci. USA **86:**5626–5630.
48. **Snijder, E. J., and J. J. M Meulenberg.** 1998. The molecular biology of arteriviruses. J. Gen. Virol. **79:**961–979.
49. **Spaan, W., H. Delius, M. Skinner, J. Armstrong, P. Rottier, S. Smeekens, B. A. M. Van der Zeijst, and S. G. Siddell.** 1983. Coronavirus mRNA synthesis involves fusion of noncontiguous sequences. EMBO J. **2:**1839–1844.
50. **Stern, S., D. Moazed, and H. F. Noller.** 1988. Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. Methods Enzymol. **164:**481–489.
51. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nat. New Biol. **246:**40–41.
52. **Tompkins, W. A. F., A. M. Watrach, J. D. Schmale, R. M. Schulze, and J. A. Harris.** 1974. Cultural and antigenic properties of newly established cell strains derived from adenocarcinomas of the human colon and rectum. J. Natl. Cancer Inst. **52:**1101–1106.
53. **van der Most, R. G., R. J. de Groot, and W. J. M. Spaan.** 1994. Subgenomic RNA synthesis directed by a synthetic defective interfering RNA of mouse hepatitis virus: a study of coronavirus transcription initiation. J. Virol. **65:**3656–3666.
54. **van der Most, R. G., and W. J. M. Spaan.** 1995. Coronavirus replication, transcription, and RNA recombination, p. 11–31. *In* S. G. Siddell (ed.), The Coronaviridae. Plenum Press, New York, N.Y.
55. **van Marle, G., J. D. Dobbe, A. P. Gultyaev, W. Luytjes, W. J. M. Spaan, and E. J. Snijder.** 1999. Arterivirus discontinuous mRNA transcription is guided by base-pairing between sense and antisense transcription-regulating sequences. Proc. Natl. Acad. Sci. USA **96:**12056–12061.
56. **van Marle, G., W. Luytjes, R. G. van der Most, T. van der Straaten, and W. J. M. Spaan.** 1995. Regulation of coronavirus mRNA transcription. J. Virol. **69:**7851–7856.
57. **Walter, A. E., D. H. Turner, J. Kim, M. H. Lyttle, P. Muller, D. H. Mathews, and M. Zuker.** 1994. Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. Proc. Natl. Acad. Sci. USA **91:**9218–9222.
58. **Watson, R. J., J. H. Weis, J. S. Salstrom, and L. W. Enquist.** 1982. Herpes simplex virus type-1 glycoprotein D gene: nucleotide sequence and expression in *Escherichia coli*. Science **218:**381–384.

59. **Wege, H., A. Muller, and V. ter Meulen.** 1978. Genomic RNA of the murine coronavirus JHM. J. Gen. Virol. **41:**217–277.
60. **Zhang, X., and M. M. C. Lai.** 1996. A 5′-proximal RNA sequence of murine coronavirus as a potential initiation site for genomic-length mRNA transcription. J. Virol. **70:**705–711.
61. **Zhang, X., and R. Liu.** 2000. Identification of a noncanonical signal for transcription of a novel subgenomic mRNA of mouse hepatitis virus: implication for the mechanism of coronavirus RNA transcription. Virology **278:** 75–85.
62. **Zhang, X. M., W. Herbst, K. G. Kousoulas, and J. Stortz.** 1994. Biological and genetic characterization of a hemagglutinating coronavirus isolated from a diarrhoeic child. J. Med. Virol. **44:**152–161.