

## Role of Nucleotides Immediately Flanking the Transcription-Regulating Sequence Core in Coronavirus Subgenomic mRNA Synthesis

Isabel Sola,<sup>†</sup> José L. Moreno,<sup>†</sup> Sonia Zúñiga, Sara Alonso,  
and Luis Enjuanes\*

Department of Molecular and Cell Biology, Centro Nacional de Biotecnología,  
CSIC, Campus Universidad Autónoma, Cantoblanco, Madrid, Spain

Received 13 July 2004/Accepted 23 September 2004

**The generation of subgenomic mRNAs in coronavirus involves a discontinuous mechanism of transcription by which the common leader sequence, derived from the genome 5' terminus, is fused to the 5' end of the mRNA coding sequence (body). Transcription-regulating sequences (TRSs) precede each gene and include a conserved core sequence (CS) surrounded by relatively variable sequences (5' TRS and 3' TRS). Regulation of transcription in coronaviruses has been studied by reverse-genetics analysis of the sequences immediately flanking a unique CS in the *Transmissible gastroenteritis virus* genome (CS-S2), located inside the S gene, that does not lead to detectable amounts of the corresponding mRNA, in spite of its canonical sequence. The transcriptional inactivity of CS-S2 was genome position independent. The presence of a canonical CS was not sufficient to drive transcription, but subgenomic synthesis requires a minimum base pairing between the leader TRS (TRS-L) and the complement of the body TRS (cTRS-B) provided by the CS and its adjacent nucleotides. A good correlation was observed between the free energy of TRS-L and cTRS-B duplex formation and the levels of subgenomic mRNA S2, demonstrating that base pairing between the leader and body beyond the CS is a determinant regulation factor in coronavirus transcription. In TRS mutants with increasing complementarity between TRS-L and cTRS-B, a tendency to reach a plateau in  $\Delta G$  values was observed, suggesting that a more precise definition of the TRS limits might be proposed, specifically that it consists of the central CS and around 4 nucleotides flanking 5' and 3' the CS. Sequences downstream of the CS exert a stronger influence on the template-switching decision according to a model of polymerase strand transfer and template switching during minus-strand synthesis.**

*Transmissible gastroenteritis virus* (TGEV) is a member of the *Coronaviridae* family, included in the *Nidovirales* order (7). TGEV is an enveloped virus with a single-stranded, positive-sense 28.5-kb RNA genome (27). About the 5' two-thirds of the entire RNA comprises open reading frames (ORFs) 1a and 1ab, which encode the replicase (*rep*). The 3' one-third of the genome includes the genes encoding structural and nonstructural proteins (5'-S-3a-3b-E-M-N-7-3'). Engineering of the TGEV genome to study fundamental viral processes, such as transcription, has been possible by the construction of TGEV infectious cDNA clones (1, 8, 40).

Coronavirus transcription, and in general transcription in the *Nidovirales* order, is an RNA-dependent RNA synthesis that includes a discontinuous step during the synthesis of subgenomic mRNAs (sgmRNAs) (16, 30). This transcription process ultimately generates a nested set of sgmRNAs that are 5'- and 3'-coterminal with the virus genome. The common 5'-terminal leader sequence of 93 nucleotides (nt), derived from the genome 5' terminus, is fused to the 5' end of the mRNA coding sequence (body) by a discontinuous transcription mechanism. Sequences preceding each gene represent

signals for the discontinuous transcription of sgRNAs. These are the transcription-regulating sequences (TRSs) that include a conserved core sequence (CS; 5'-CUAAAC-3'), identical in all TGEV genes (the CS of the body sequence [CS-B]), and the 5'- and 3'-end-flanking sequences (5' TRS and 3' TRS, respectively) that regulate transcription (2). Since this CS sequence is also found at the 3' end of the leader sequence (CS-L), it may base pair with the nascent minus strand complementary to each CS-B (cCS-B). In fact, the requirement for base pairing during transcription has been formally demonstrated to occur in arteriviruses (25, 38) and coronaviruses (44) by experiments in which base pairing between CS-L and the complement of CS-B was engineered in infectious genomic cDNAs. Subgenomic RNA (sgRNA) synthesis in CS-L and CS-B mutants was regulated by changing only the base pairing between these two elements. Moreover, alternative mRNAs were synthesized in TGEV from noncanonical CSs, provided that their flanking sequences extended complementarity with TRS-L (34, 44). In this report, the role in transcription of nucleotides immediately flanking the CS-B has been analyzed using infectious genomic TGEV cDNAs. Base pairing between leader sequences and the nascent negative RNA strand beyond the canonical CS sequence (5'-CUAAAC-3') has been shown in this report to be a determinant factor in coronavirus transcriptional regulation.

Although two major models have been proposed to explain the discontinuous transcription in *Nidovirales* (16, 30), current experimental data favor the model of discontinuous transcrip-

\* Corresponding author. Mailing address: Department of Molecular and Cell Biology, Centro Nacional de Biotecnología, CSIC, Campus Universidad Autónoma, Cantoblanco, 28049 Madrid, Spain. Phone: 34-91-585 4555 Fax: 34-91-585 4915. E-mail: L.Enjuanes@cnb.uam.es.

<sup>†</sup> I.S. and J.L.M. contributed equally to this work.

tion during negative-strand synthesis (28, 29, 31, 32). This concept was reinforced by demonstrating for arterivirus and coronavirus that the CS included in the sgmRNA was derived from the CS preceding each gene and not from the CS present at the 3' end of the leader sequence (25, 38, 44). In this model, the TRS-B acts as an attenuation and dissociation signal for the transcription complex during the synthesis of the RNA negative strand. Template switching at the sites of RNA-dependent RNA polymerase (RdRp) pausing resembles a high-frequency similarity-assisted copy choice RNA recombination (3, 20, 23) in which the noncontiguous TRS-B and TRS-L sequences are probably brought into physical proximity by RNA-protein and protein-protein interactions (44). Also in this model, the nascent negative RNA with the TRS complement at its 3' end (cTRS-B) (donor molecule) dissociates from the genomic-RNA (gRNA) template at the CS domain to join the TRS-L (acceptor) and resume synthesis of sgRNA. This sgRNA serves as a template for multiple rounds of sgmRNA synthesis. It has been experimentally proven that CS in TRS-L is exposed in a stem-loop in arteriviruses (35) and is, probably, also exposed in a similar secondary structure in coronaviruses (I. Sola, S. Alonso, S. Zúñiga, and L. Enjuanes, unpublished data), as shown by RNA structure predictions, making the TRS-L accessible as an acceptor during template switching.

Transcription in coronavirus is probably regulated by two main factors: (i) base pairing between the TRS-L and the nascent negative strand and (ii) RNA-protein and protein-protein interactions involving TRS sequences and viral and cellular proteins. The proximity of body TRSs to the 3' end of the genome probably influences the relative amount of sgmRNA, because the RdRp finds less attenuation and dissociation TRS-B signals during the synthesis of sgRNAs of smaller size. This is the case for other viruses that produce multiple sgRNAs (19) and, in general, for coronaviruses. Nevertheless, a perfect correlation between mRNA abundance and the relative position of TRS in the genome could not strictly be established (27, 37). The relevance in transcription of other potential factors, such as relative TRS position in the genome and the TRS-B secondary structure, could not be confirmed either for arteriviruses or for coronaviruses (21, 22, 24).

Previous studies of TGEV have shown that the canonical CS was nonessential for the generation of subgenomic mRNAs, but its presence led to transcription levels at least  $10^3$ -fold higher than those in its absence (44). It was also shown that the synthesis of sgmRNAs proceeds only when the CS is located in an appropriate context (2). That seems to be the case with CS-S2, a canonical CS sequence within the S gene (152 nt downstream of CS-S1, the CS leading to the synthesis of mRNA S coding for S protein) that does not lead to the synthesis of any detectable amount of the corresponding sgmRNA (2), probably because it is located in a nonfavorable sequence context.

In a previous work, the role of CS nucleotides (5'-CUAAA C-3') in coronavirus transcription was analyzed (44). In this report, the role in transcription of four nucleotides immediately flanking the CS both at the 5' and the 3' end has been studied using as a model the transcriptionally inactive canonical CS-S2. The rationale for selecting 5' and 3' TRS flanking sequences consisting of four nucleotides comes from the results of an *in silico* analysis (see below) showing that to predict

both viral mRNAs and alternative mRNAs at noncanonical junction sites, an optimal TRS-L should include the CS plus four nucleotides flanking the CS both at the 5' and 3' ends. Furthermore, these predictions have been supported by experimental data. Using TGEV infectious genomic cDNAs, we have shown that CS-S2 inactivity is a genome position-independent phenomenon. A good correlation between mRNA S2 levels and the free energy ( $\Delta G$ ) of duplex formation between TRS-L and cTRS-S2 was observed in mutants that extended complementarity with TRS-L, indicating that this base pairing during the synthesis of nascent RNA is a key factor for transcription regulation in coronavirus, leading to detectable transcription levels only when a minimum threshold is reached. It has also been shown that for similar  $\Delta G$  values, mutations extending complementarity with TRS-L by the 3' TRS sequences led to larger amounts of mRNA-S2 than mutations at the 5' TRS region, indicating that sequences downstream of the CS exert a stronger influence on the template-switching decision. This observation is most consistent with a model of polymerase strand transfer and template switching during minus-strand synthesis in which the CS is probably the major nascent chain detachment signal.

#### MATERIALS AND METHODS

**Cells and viruses.** Baby hamster kidney (BHK-21) cells stably transformed with the gene coding for porcine aminopeptidase N (6) were grown in Dulbecco's modified Eagle's medium supplemented with 5% fetal calf serum and G418 (1.5 mg/ml) as a selection agent. Mutant viruses obtained in this work were grown in swine testis (ST) cells (18) and titrated as previously described (12).

**Plasmid constructs.** TGEV cDNAs including mutated sequences of the inactive TRS-S2 transcription unit in the place previously occupied by the nonessential genes 3a and 3b were generated by PCR-directed mutagenesis. The 918-bp BmgBI-BlnI fragment, including the 3a gene and the first 320 nt of the 3b gene, was removed from intermediate plasmid pSL-TGEV-AvrII, comprising nt 22965 to 25865 from the TGEV genome (GenBank accession no. AJ271965), and replaced by mutant TRS-S2 transcription units consisting of CS-S2 and 30 nt from both the 5' TRS and 3' TRS, including a series of nucleotide substitutions that extend complementarity with TRS-L (Fig. 1). An 80-bp product corresponding to wild-type (wt) or mutant TRS-S2 transcription units was generated by PCR with the oligonucleotides described in Table 1, which included BmgBI and BlnI restriction sites, using the plasmid pSL-SC11-3EMN/C8-BGH (8) as the template. PCR products were digested with BmgBI and BlnI and cloned into the same restriction sites of plasmid pSL-TGEV-AvrII, leading to plasmid pSL-TGEV-AvrII- $\Delta$ 3-TRS-S2wt and the collection of TRS-S2 mutants M1 to M10. To introduce mutations in the TGEV infectious cDNA, the 1,982-bp AvrII-AvrII fragment from plasmid pSL-TGEV-AvrII- $\Delta$ 3-TRS-S2 with the corresponding mutations was inserted into the same sites of plasmid pBAC-TGEV $\Delta$ Cl<sub>a</sub> (nt 22976 and 25867 of the TGEV genome), leading to plasmid pBAC-TGEV $\Delta$ Cl<sub>a</sub>- $\Delta$ 3-TRS-S2. To obtain the full-length TGEV-cDNA, the toxic ClaI-ClaI fragment (5,198 bp) was introduced as previously described (1). All cloning steps were checked by sequencing the PCR-amplified fragments and cloning junctions.

**Transfection and recovery of infectious TGEV from cDNA clones.** BHK-porcine aminopeptidase N cells were grown to confluence in 35-mm-diameter plates and transfected with 4  $\mu$ g of the full-length TGEV-cDNA clone and 12  $\mu$ l of Lipofectamine 2000 (Invitrogen) according to the manufacturer's specifications. The estimated transfection efficiency of the TGEV cDNA using this system was around 20% in all cases. Cells were incubated at 37°C for 6 h, and then the transfection medium was discarded, 200  $\mu$ l of trypsin-EDTA was added, and trypsinized cells were plated over a confluent ST monolayer grown in a 35-mm-diameter plate. After a 2-day incubation period, the cell supernatants were harvested and passaged three times on fresh ST cell monolayers. After three passages, mutant viruses were cloned by three plaque purification steps. Recombinant TGEVs (rTGEVs) were grown and titrated as previously described (12).

**RNA analysis by Northern blotting.** Total intracellular RNA was extracted at 16 h postinfection (hpi) from virus-infected ST cells using the RNeasy mini kit (QIAGEN) according to the manufacturer's instructions. RNAs were separated in denaturing 1% agarose-2.2 M formaldehyde gels and blotted onto positively

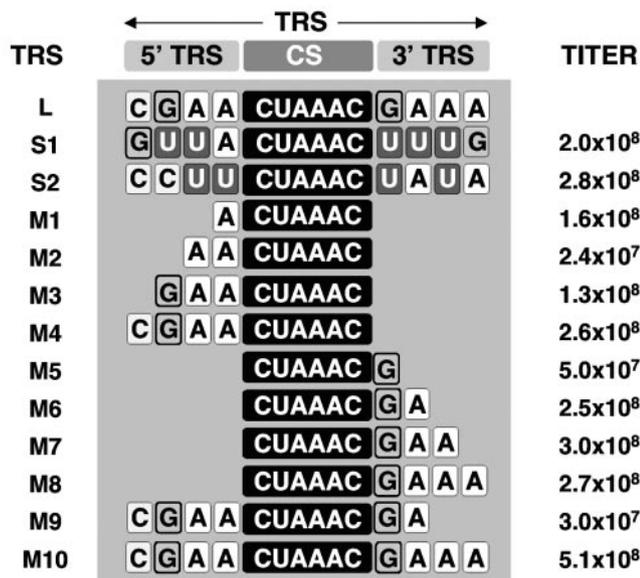


FIG. 1. Sequences of TRS-S2 mutants that extend the complementarity with the leader TRS. Mutations introduced into the nucleotides immediately flanking the CS-S2 are shown in the figure (mutants M1 to M10). The right column shows the titers of each mutant when it was grown on ST cells after three plaque purification steps. L, sequences adjacent to the CS-L in the leader region; S1, sequences adjacent to the native CS-S1; S2, sequences adjacent to the native CS-S2. Virus titers (PFU per milliliter) are indicated in the right column.

charged nylon membranes (BrightStar-Plus; Ambion) as described previously (2). The 3'-untranslated region-specific single-stranded DNA probe was complementary to nt 28300 to 28544 of the TGEV strain PUR46-MAD genome (27). Probe labeling was performed with the BrightStar psoralen-biotin nonisotopic labeling kit (Ambion), and Northern hybridizations were performed according to the manufacturer's instructions. Detection was done with the BrightStar BioDetect kit (Ambion).

**RNA analysis by RT-PCR.** RNA analysis of the rTGEVs was performed by reverse transcription-PCR (RT-PCR). Total intracellular RNA was extracted at 16 hpi from rTGEV-infected cells as previously described. cDNAs were synthesized at 50°C for 1 h with avian myeloblastosis virus reverse transcriptase (Reflactase) (Active Motif) and the antisense primer CS-RS (5'-ATCACCATTGA

GAAGTTCAACTGCT-3'), complementary to nt 375 to 351 of ORF 3b. The cDNA generated was used as a template for specific PCR amplification of mRNAs using the reverse primer CS-RS and the forward primer SP (5'-GTGA GTGTAGCGTGGCTATATCTCTTC-3'), complementary to nt 15 to 39 of the TGEV leader sequence. For RT-PCR amplification of genomic sequences in the regions of the 3a and 3b genes, oligonucleotides CS-RS and S-4310-VS (5'-AT TACGAACCAATTGAAAAAGTGC-3'), complementary to nt 4316 to 4339 of gene S, were used. RT-PCR products were separated by electrophoresis in 2% agarose gels, purified, and used for direct sequencing with the same oligonucleotides used for PCR.

Real-time RT-PCR was used for quantitative analysis of gRNA (used as an endogenous standard) and mRNA-S2 from different rTGEVs with mutated TRS-S2 sequences. Oligonucleotides used for RT and PCRs, described in Table 2, were designed by the Primer Express software. Forward oligonucleotide mRNACS1 JS4 was used to amplify mRNA-S1. Forward oligonucleotide mRNACS2 JS3 amplified mRNA-S2 from mutants M1, M2, M3, and M4. Forward oligonucleotide mRNACS2 JS7 was used to amplify mRNA-S2 from mutants M5 to M10. SYBR-green PCR master mix (Applied Biosystems) was used in the PCR step, according to the manufacturer's specifications. Detection was performed with an ABI PRISM 7000 sequence detection system (Applied Biosystems). Data were analyzed with ABI PRISM 7000 SDS version 1.0 software.

**In silico analysis.** Potential base pairing score calculations were done with the LALIGN program at the public ISREC LALIGN server (<http://www.ch.embnet.org/>), a local alignment tool that implements the algorithm of Huang and Miller (11), as previously described (44). Free-energy calculations were performed using the two-state hybridization server (<http://www.bioinfo.rpi.edu/applications/mfold/>) (17). In silico analysis was performed with a TRS-L, including the CS and four nucleotides from both the 5' TRS and 3' TRS, and the TGEV genomic sequences around the TRS-S2 insertion site. Secondary-structure predictions were done using the M-fold Zuker algorithm (42).

RESULTS

**Effect of genome position on the lack of transcriptional activity of TGEV TRS-S2, including the canonical CS-S2.** To study whether the transcriptional inactivity of the canonical CS-S2 located at nt 120 of the S gene (2) was dependent on genome position, a transcription unit (TRS-S2) consisting of the central CS flanked both 5' and 3' by 30 nt from the native TRS of CS-S2 was introduced at different positions in the TGEV genome. TRS-S2 was inserted by replacing nonessential genes 3a and 3b (nt 24708 to 25691 of PUR46-MAD). The place occupied by genes 3a and 3b was previously shown to be a very stable site for the insertion of heterologous sequences

TABLE 1. Oligonucleotides used for site-directed mutagenesis

Mutant	Oligonucleotide <sup>a</sup>	Oligonucleotide sequence 5'→3' <sup>b</sup>
TRS-S1	S1-Bmg VS2 S1-BlpI RS	GGACGTCACGTCCATTAAAAGTTGCTCATTAGAAATAATGG GTCCGGCGCTAAGCTCATGGTGTGTTAACGAAGTGG
TRS-S2wt	S2-Bmg VS2 S2-BlpI RS	GGACGTCACGTCCATTAAAACCAGTGGAATCTCATTGAAAC GTCCGGCGCTAAGCATCTGAATTAGGTGGTAACTAC
M1	S2-Bmg 5'A VS	GGACGTCACGTCCATTAAAACCAGTGGAATCTCATTGAAACCTTCTACTAAAC
M2	S2-Bmg 5'AA VS	GGACGTCACGTCCATTAAAACCAGTGGAATCTCATTGAAACCTTCCAATAAAC
M3	S2-Bmg 5'GAA VS	GGACGTCACGTCCATTAAAACCAGTGGAATCTCATTGAAACCTTGGAACTAAAC
M4	S2-Bmg 5'CGAA VS	GGACGTCACGTCCATTAAAACCAGTGGAATCTCATTGAAACCTTCGAACTAAAC
M5	S2-BlpI 3'G RS	GTCCGGCGCTAAGCATCTGAATTAGGTGGTAACTACTACCTAGCCTTTAG
M6	S2-BlpI 3'GA RS	GTCCGGCGCTAAGCATCTGAATTAGGTGGTAACTACTACTACTACTCGTTTAG
M7	S2-BlpI 3'GAA RS	GTCCGGCGCTAAGCATCTGAATTAGGTGGTAACTACTACTACTACTCGTTTAG
M8	S2-BlpI 3'GAAA RS	GTCCGGCGCTAAGCATCTGAATTAGGTGGTAACTACTACTACTACTCGTTTAG
M9	S2-Bmg CGAA + GA VS	GGACGTCACGTCCATTAAAACCAGTGGAATCTCATTGAAACCTTCGAACTAAACGATAG
M10	S2-Bmg CGAA + GAAA VS	GGACGTCACGTCCATTAAAACCAGTGGAATCTCATTGAAACCTTCGAACTAAACGAAAGTA

<sup>a</sup> Oligonucleotides used to introduce the mutations begin with "S2" and indicate the restriction site included, the nucleotide substitutions, and the position of the substitutions 5' or 3' of the CS.

<sup>b</sup> The mutated nucleotides are shown in boldface. Restriction endonuclease sites used for cloning are underlined (BmgBI, CACGTC; BlpI, GCTAAGC).

TABLE 2. Oligonucleotides used for RT-PCR analysis of RNA from rTGEV-infected cells

Amplicon	Size (bp)	Forward primer 5'→3'	Reverse primer 5'→3'
Virus	80	TTCTTTTGACAAAACATACGGTGAA	CTAGGCAACTGGTTTGTAACATCTTT
mRNACS1 JS4	177	GTCTTCGGACACCAACTCGAACTAAACTAT	CCGCCTGAGAAAAGGCTGCATTG
mRNACS2 JS3	177	TCTTCGGACACCAACTCGAACTAAACGAAAGT	CCGCCTGAGAAAAGGCTGCATTG
mRNACS2 JS7	176	GTCTTCGGACACCAACTCGAACTAAACTA	CCGCCTGAGAAAAGGCTGCATTG

(34). As a positive control, a similar transcription unit (TRS-S1), derived from that preceding TGEV ORF S, that includes the CS motif (CS-S1) and 30 nt from the 5' and 3' CS flanking sequences, was inserted at the position of nonessential genes 3a and 3b. Viruses were recovered from these mutant TGEV infectious cDNA clones, with titers similar to those obtained with the wild-type TGEV cDNA (Fig. 1). Conventional and real-time RT-PCR analysis with specific primers for the detection of both mRNAs generated from CS-S1 and CS-S2 were performed. Only the mRNA synthesized from the TRS-S1 was detected, while the potential mRNA that could have been generated from TRS-S2 at the engineered cloning site was not detected (Fig. 2B). These results indicated that mRNA synthesis from CS-S2 was not affected by genome position and that the transcriptional inactivity of CS-S2 was independent of the sequences distantly flanking the CS.

**Transcriptional activity of TRS-S2 mutants.** Since TRS-S2 remained transcriptionally inactive at different genome locations, we postulated that the sequences immediately flanking the CS (30 nt from the 5' and 3' native TRSs of CS-S2) should be responsible for this inactivity. It has been shown that sequences flanking the CS motif in the TGEV genome modulate sgmRNA synthesis principally by contributing to the extent of base pairing with the TRS-L (44). To study whether base-pairing extension between TRS-L and the complementary sequences of TRS-S2 could account for the transcriptional activity of CS-S2, a collection of TGEV infectious cDNA clones was generated. These mutant TRS-S2 transcription units gradually increased complementarity with TRS-L and were inserted at the site of nonessential genes 3a and 3b. An *in silico* analysis method providing the potential base-pairing score of sequence domains complementary to genomic RNA with the TRS-L showed that to predict both viral mRNAs and alternative mRNAs at noncanonical junction sites, an optimal TRS-L should include the CS plus four nucleotides flanking the CS both at the 5' and the 3' end (44). This result suggested that during coronavirus transcription, base pairing between TRS-L and cTRS-B concerns a limited number of nucleotides around the CS and that base pairing with more-distal sequences would be less relevant. The TRS-S2 sequence selected for the mutational analysis consisted of the central CS flanked by four nucleotides derived from the 5' TRS-S2 and another four nucleotides from the 3' TRS-S2 (5'-CCUUCUAAACUAUA-3'). TGEV mutants were constructed with substitutions either in the four nucleotides immediately flanking the CS at the 5' end (M1, M2, M3, and M4), at the 3' end (M5, M6, M7, and M8), or at both ends (M9 and M10) (Fig. 1). The transcriptional unit TRS-S2, including each mutation, was inserted immediately downstream of the S gene, at the site previously occupied by nonessential genes 3a and 3b. Infectious recombinant viruses were recovered from all TRS-S2 mutants, with

titers similar to those obtained with the wt TGEV cDNA (Fig. 1). Swine testis cells were infected with rTGEVs, and intracellular RNA was extracted at 16 hpi. Sequences from genomic RNA at the region of TRS-S2 insertion were amplified by RT-PCR with oligonucleotides indicated in Materials and Methods. Sequencing of these products showed that the nucleotide substitutions introduced within the TRS-S2 sequence were stably maintained during virus passages. RT-PCR analysis with primers specific to detect mRNA-S2 (Fig. 2A), the mRNA generated from the engineered TRS-S2, showed a single sgmRNA in the virus containing the wt TRS-S2, while in all TRS-S2 mutants, two amplification products were observed. These products were detected in all TGEV mutants, although with different relative intensities, and also appeared in the virus including the TRS-S1 transcription unit, with sequences from the TRS preceding the S gene (Fig. 2B). Relative amounts of sgmRNA-S2 synthesized by the mutants seemed to be related to the potential base pairing between TRS-L and cTRS-S2 (see below for a precise quantification by real-time RT-PCR). Sequencing of these cDNAs showed that the smaller-size mRNA (mRNA-S2) corresponded to the subgenomic mRNA generated by the fusion of leader sequences to the canonical CS in TRS-S2 mutants (Fig. 2C). In the TRS-S1 mutant, the mRNA species with the smallest size corresponded to the sgmRNA generated at the CS-S1 site, as expected. The larger mRNA (mRNA-3a.2) was generated in all analyzed viruses by a leader-to-body joining at a noncanonical site located at the 3' end of ORF S, 57 nt upstream of the CS-S2 that replaced genes 3a and 3b (Fig. 2C). The junction site in the alternative mRNA-3a.2 showed an extended identity with TRS-L in sequences immediately flanking the noncanonical CS, including the 5'-CGAA-3' and 5'-GAAA-3' motifs at the 5' and 3' ends of the CS, respectively. The alternative mRNA-3a.2 was previously detected in TGEV mutants with nucleotide substitutions either in the CS-L or the CS-3a (44).

**Prediction of subgenomic mRNA synthesis by potential base pairing between the TRS-L and the nascent negative RNA sequences.** It was previously shown that potential base pairing between TRS-L and the nascent negative RNA sequences of TGEV was an important regulatory factor of discontinuous transcription (44). Potential base pairing was estimated using an *in silico* approach, based on a local alignment algorithm that analyzed the identity between TRS-L and genomic RNA sequences and predicted not only the TGEV sequences leading to the synthesis of structural and nonstructural TGEV mRNAs, but also the noncanonical sites involved in the generation of alternative mRNAs. Analysis of the potential base pairing between TRS-L and sequences complementary to gRNA in the region of insertion of TRS-S2 showed two peaks of high identity corresponding to the canonical and noncanonical leader-to-body junction sites found in all TRS-S2 mutants

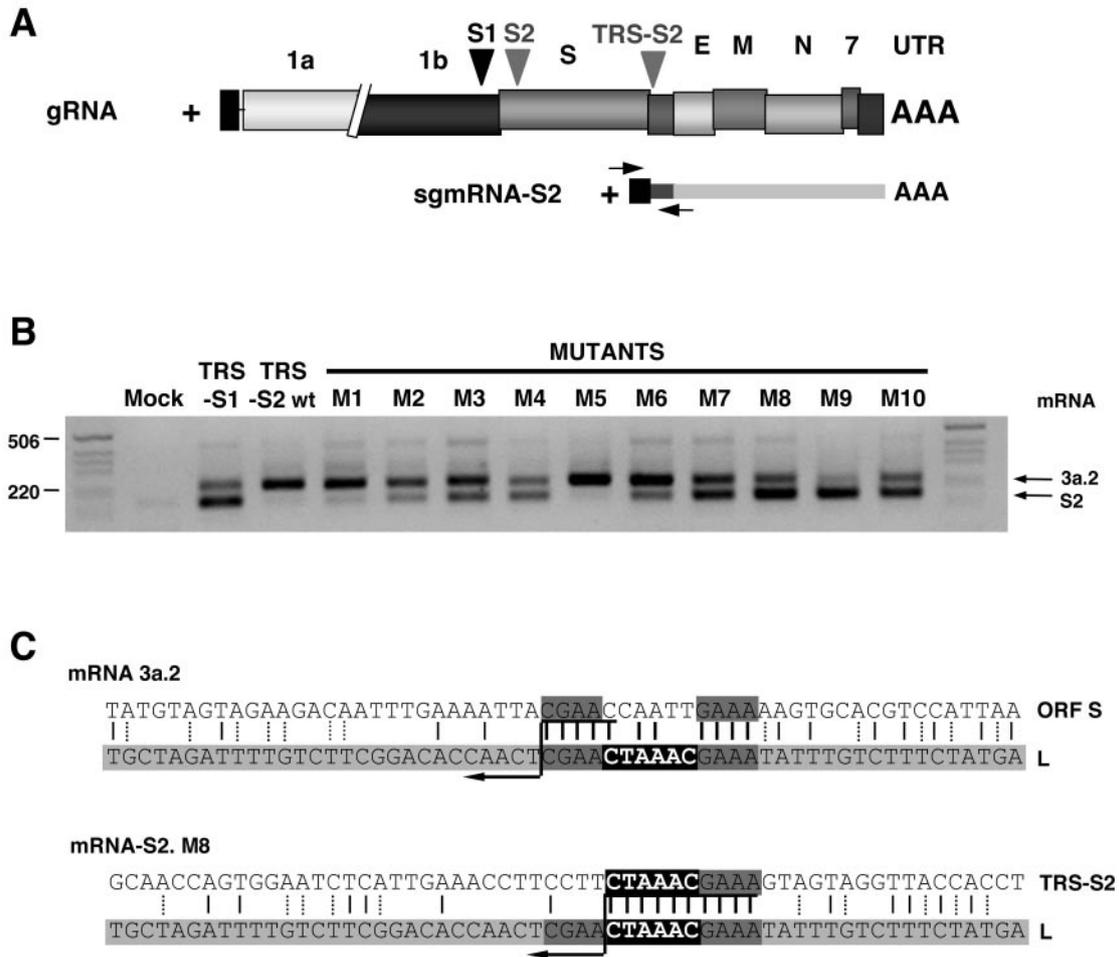


FIG. 2. Synthesis of sgmRNA-S2 by TRS-S2 mutants. (A) Diagram showing the insertion site in the TGEV genome (at the position of nonessential genes 3a and 3b) of the TRS-S2 transcription unit, which consists of CS-S2 and 30 nt from both the 5' TRS and the 3' TRS, including a series of nucleotide substitutions that extend the complementarity with the leader TRS. The structure of the expected sgmRNA-S2 is shown under the gRNA structure. Arrows indicate the approximate positions of primers used for RT-PCR. UTR, 3' untranslated region. (B) Specific detection by RT-PCR of mRNA-S2 in TRS-S2 mutants. The mRNA 3a.2 species corresponds to the alternative mRNA generated from a noncanonical leader-to-body junction located at the 3' end of ORF S, 57 nt upstream of the CS-S2 engineered at the 3ab site. TRS-S1, transcription unit containing the sequence of the active TRS that drives the synthesis of sgmRNA S; TRS-S2 wt, transcription module containing the wt sequence of the inactive TRS, including CS-S2. (C) Sequence analysis of the body-to-leader junction sites in the two mRNA species detected. As an example, the case of M8 is shown. The sequence on the top corresponds to gRNA in the fusion site. The bottom sequence in the light-gray box corresponds to the leader (L) sequence. The CS appears as white letters in a black box. CGAA and GAAA motifs are shown in dark-gray boxes. Vertical bars represent the identity between the sequences, with thick bars at the possible fusion site. Dotted vertical bars represent the possible non-Watson-Crick interactions. Crossover should occur in any nucleotides above the arrow.

that led to mRNAs S2 and 3a.2, respectively (Fig. 3). In the rTGEV mutant including the wt TRS-S2 sequence, the peak with the highest TRS-L identity to gRNA (score, 45) corresponded to the junction site 3a.2, whereas a low potential base pairing value (score, 32) was associated with the CS-S2 site. This result may explain the detection of a unique mRNA species generated at the most favorable junction site, 3a.2 (Fig. 2B). In contrast, for all engineered TRS-S2 mutants, in addition to the peak corresponding to the 3a.2 junction site, a second peak with increasing levels of identity between TRS-L and gRNA (score between 35 and 70) was also predicted. This peak corresponds to the CS-S2 junction site in mutants with different sequences immediately flanking CS-S2. This result may explain the detection of two mRNA species generated at

the 3a.2 and CS-S2 sites, with varied relative intensities according to the different values of potential base pairing at the CS-S2 junction site (Fig. 2B). In the TRS-S1 mutant, the smallest peak was associated with CS-S1 (score 35).

**Synthesis of viral mRNAs in TRS-S2 mutants.** To evaluate whether the insertion of mutant TRS-S2 transcriptional units in the TGEV genome had any effect on the transcription of viral mRNAs, intracellular RNA from rTGEV-infected cells was analyzed by Northern blotting. The viral sgmRNA pattern of all TRS-S2 mutants was similar to that observed in wt TGEV, with the exception of mRNA 3a, absent in TGEV mutants in which genes 3a and 3b had been replaced by TRS-S2 transcription units (Fig. 4), indicating that RNA viral expression was unaffected. sgmRNA-S2 was clearly detected in

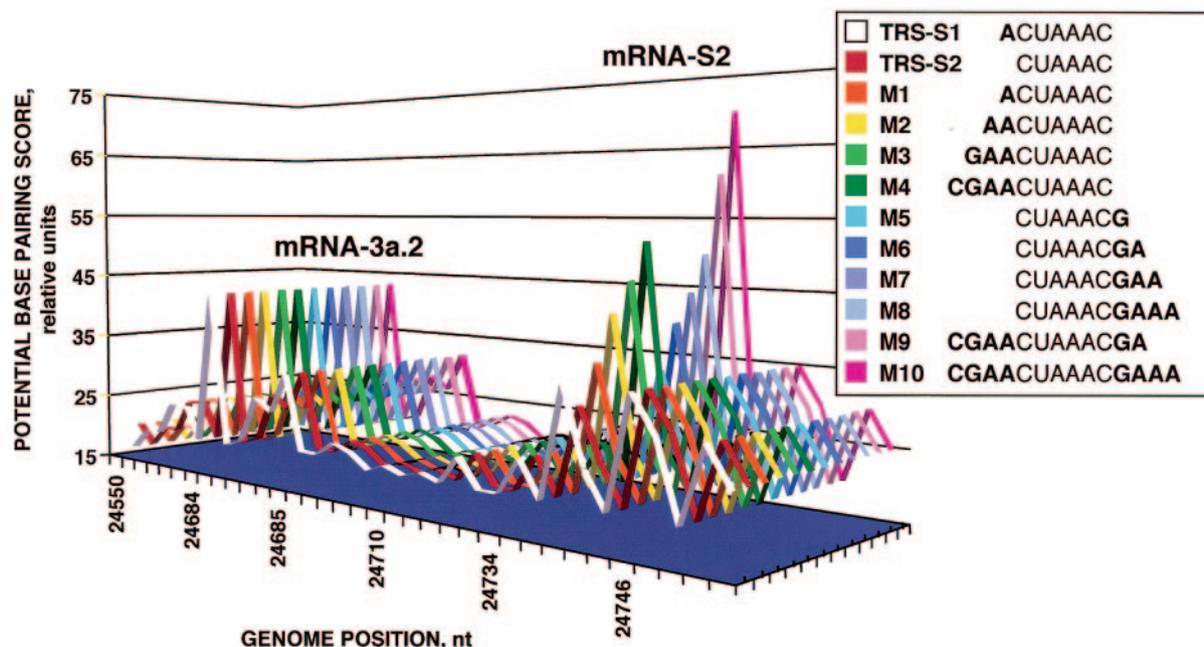


FIG. 3. In silico analysis of the identity between TRS-L (5'-CGAACUAAACGAAA-3') and the TGEV genomic sequences in the region of TRS-S2 insertion, the place previously occupied by genes 3a and 3b. A graphical plot of the potential base-pairing score versus the genome position is shown. Each of the three-dimensional lines represent the identity between TRS-L and the sequences corresponding to TRS-S1, TRS-S2, or the different TRS-S2 mutants with increasing complementarity with the leader (M1 to M10). The peaks assigned to the two sgRNA species are indicated.

mutants M9 and M10, which included nucleotide substitutions providing the highest complementarity with TRS-L. sgRNA-S2 was not detected either in wt TRS-S2, as expected, or in the TRS-S2 mutants, indicating that transcription levels in these mutants were below the sensitivity of this technique.

**Influence of the TRS-L-cTRS-B duplex  $\Delta G$  on sgRNA-S2 synthesis.** To study the effect of base pairing between the nascent negative sgRNA and the TRS-L on the transcriptional activity of TRS-S2, mRNA-S2 levels were quantified in TRS-S2 mutants by real-time RT-PCR using specific oligonucleotides and the gRNA as an internal standard. The amount of mRNA-S2 in TRS-S2 mutants was expressed in relation to that of an rTGEV including wt TRS-S2 sequences (Fig. 5). Subgenomic mRNA-S2 was expressed at different levels in all TRS-S2 mutants, with relative abundances ranging from 10- to  $7 \times 10^4$ -fold the amount of mRNA-S2 produced from wt TRS-S2 sequences. Just by extending complementarity to TRS-L with a unique nucleotide immediately flanking the CS-S2 at either the 5' or the 3' end (mutant M1 or M5, respectively), mRNA-S2 expression was significantly increased, suggesting that a minimum complementarity of TRS-L to cTRS-S2 was necessary to promote the synthesis of the corresponding sgRNA from a canonical CS that was previously inactive due to the insufficient complementarity to TRS-L of its adjacent sequences. A good correlation was observed between the mRNA-S2 amount and the value of  $\Delta G$  for the duplex formation between the TRS-L and the cTRS-S2 in the mutants, confirming that this thermodynamic parameter is a decisive factor for template switching during the synthesis of coronavirus subgenomic RNAs. In general, for similar  $\Delta G$

values, mutants extending complementarity with TRS-L through nucleotides within the 3' TRS synthesized larger amounts of mRNA-S2 than mutants whose complementarity with TRS-L was extended within the 5' TRS (Fig. 5). As an example, M8 extended complementarity with TRS-L through the 5'-GAAA-3' motif within 3' TRS sequences (5'-CUAAA CGAAA-3') and expressed 15-fold more sgRNA-S2 than M4, with similar  $\Delta G$  values for duplex formation, but extended complementarity with TRS-L through the 5'-CGAA-3' motif within 5' TRS sequences (5'-CGAACUAAAC-3'). These results indicated that nucleotides adjacent to CS-S2 by the 3' region are more decisive for mRNA synthesis than nucleotides in the 5' region. The differential behavior of mutants that extended complementarity with TRS-L by the 5' TRS or 3' TRS nucleotides is also shown in the graphical representation of the  $\Delta G$  values for duplex formation between TRS-L and cTRS-B versus the relative amount of mRNA-S2 (Fig. 6). Two separate representations for 5' TRS and 3' TRS mutants were required, with the mRNA-S2 level scale significantly shifted, suggesting that complementarity with the TRS-L at the 3' TRS region was a stronger determinant for template switching than a similar complementarity at the 5' TRS. In both representations, a tendency to reach a plateau was observed, indicating that during discontinuous transcription, the  $\Delta G$  for duplex formation between TRS-L and cTRS-B varied towards a minimum value associated with maximum levels of mRNA synthesis. These data are compatible with the proposed transcription mechanism model postulating a template switch during synthesis of the negative strand and suggest that the CS sequence would behave as a detachment signal for the transcription

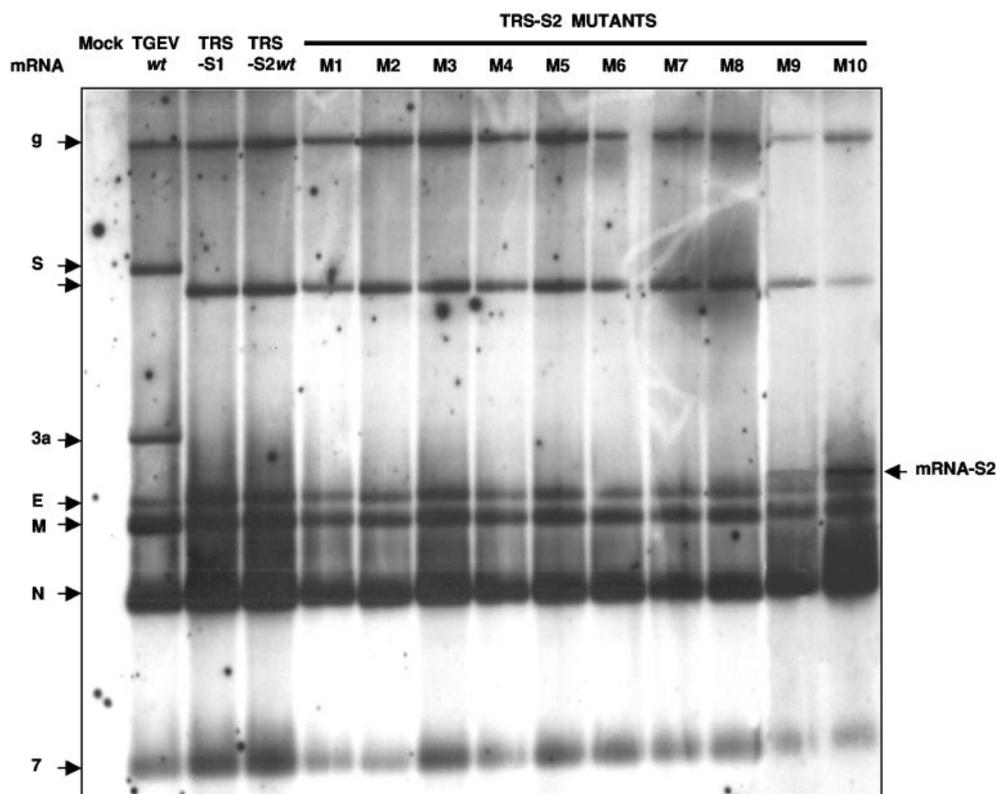


FIG. 4. Northern blot analysis of intracellular RNA from TRS-S2 mutants. ST cells were infected with rTGEVs at a multiplicity of infection of 3. Total RNA was extracted at 16 hpi and analyzed by Northern blotting with a probe complementary to the 3' end of the gRNA as described in Materials and Methods. The difference in the sizes of mRNA S and wt TGEV is due to the deletion of ORFs 3a and 3b (918 bp), which have been replaced by the TRS-S1 or TRS-S2 transcription unit (80 bp). Viral mRNAs are indicated on the left side of the figure, and the mRNA-S2 detected in mutants M9 and M10 is indicated on the right.

complex, providing that a minimum  $\Delta G$  in the duplex between TRS-L and cTRS-B is reached.

## DISCUSSION

Regulation of transcription in coronaviruses has been studied in this report by analyzing the unique CS in the TGEV genome (CS-S2) that does not lead to detectable amounts of the corresponding mRNA, in spite of its canonical sequence. This observation indicated that the presence of a CS was not sufficient to drive transcription, probably because of the sequences flanking the CS. CS-S2 is located within the S gene, 152 nt downstream of CS-S1, the CS leading to the synthesis of mRNA S1, which codes for S protein. To determine whether the transcriptional inactivity of CS-S2 was due to positional effects, the TRS-S2 cassette, including CS-S2 and its immediately flanking sequences (30 nt from both the 5' TRS and the 3' TRS) was inserted in distal positions in the TGEV genome (i) replacing the nonessential 3a and 3b genes and (ii) between the N and 7 genes at the 3' end of the genome (data not shown). The expected mRNA-S2 was not detected in any case, indicating that the transcriptional inactivity of CS-S2 was genome position independent and was determined by the adjacent 5' and 3' 30 nt. Furthermore, the lack of TRS-S2 activity was reversed simply by increasing the complementarity with TRS-L with a single nucleotide adjacent to the CS, suggesting

that base pairing between TRS-L and cTRS-B, and not the position of TRS-S2 within the genome, was the crucial factor for transcriptional activity. Position effects have been previously described for other coronavirus systems, such as mouse hepatitis virus (MHV) (9, 39) and TGEV (5), for which it was suggested that the TRS location probably has an impact on gene expression, especially in proximal promoter locations.

Within *Nidovirales*, it has also been postulated that the RNA secondary structure in the TRS-B regions regulates transcription. Regions with a characteristic secondary structure were proposed either as polymerase-pausing signals during transcription (14) or as binding sites for host proteins (41). However, the role of the predicted TRS secondary structures could not be confirmed for the arterivirus equine arteritis virus (EAV) (22). Similar results were obtained with a bovine coronavirus defective interfering RNA system, in which the inactive canonical TRS of sgRNA5 was buried in the stem of a stable hairpin. Mutations predicted to unfold the stem and make the canonical TRS accessible for base pairing did not result in its transcriptional activation (21). Predictions of RNA secondary structures for all TGEV TRSs, including the inactive TRS-S2, by the M-fold algorithm (17, 43) did not show any differential features between functional and nonfunctional TRSs (data not shown). A similar analysis was performed with the TRS-S2 mutants described above, and no correlation between the pre-

dicted secondary structure and the different degrees of transcriptional activity could be established, indicating that the secondary structure surrounding a body CS motif might not be a strong determinant factor in transcription. According to the proposed model for coronavirus transcription, the helicase activity of the RdRp complex may unwind the double-stranded RNA structures during the synthesis of the negative RNA strand, making the RNA secondary structure of the body TRS not decisive in transcription.

In a previous work on TGEV TRSs, the effect of nucleotide substitutions within the conserved core sequence CS (5'-CUA AAC-3') of leader and body TRSs was analyzed (44). It was shown that the canonical body CS was not absolutely essential for the generation of sgmRNA, since alternative sgmRNAs generated from noncanonical CSs were detected in virus mutants. However, point mutations in the body CS nucleotides reduced by more than 10<sup>3</sup>-fold sgmRNA synthesis compared to that of the wt virus, confirming the requirement of complementarity between CS-L and cCS-B for transcription. This concept was reinforced by showing that sgmRNA synthesis was partially or completely restored by the introduction of nucleotide substitutions allowing the formation of non-Watson-Crick or Watson-Crick base pairs. The extent of sgmRNA synthesis was related to the base-pairing potential between CS-L and cCS-B, calculated as  $\Delta G$  values. From these results, it was proposed that the lack of transcription observed for the canonical CS-S2 could be explained by the relatively low potential base pairing and  $\Delta G$  value between the TRS-L and cTRS-B, as a consequence of the context sequence surrounding the CS. The results described in this report confirm and extend this hypothesis in which the  $\Delta G$  associated with duplex formation between TRS-L and cTRS-B is a determinant factor in the transcriptional regulation of the TGEV coronavirus. Using the TGEV genomic cDNA clone, the TRS-S2 cassette with the wt sequence or an alternative one including mutations that extended complementarity with TRS-L, were inserted at the site of nonessential genes 3a and 3b. We have shown that

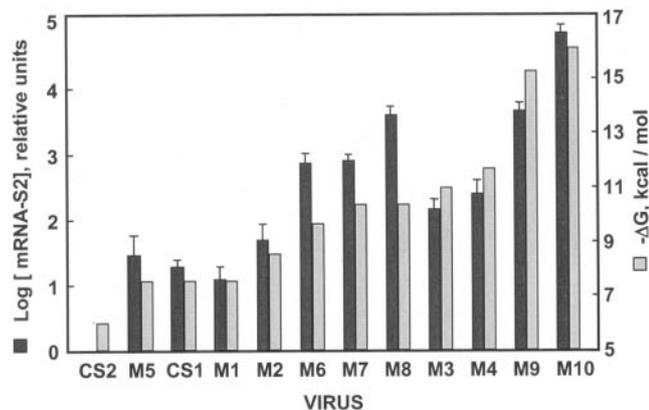


FIG. 5. Quantification of sgmRNA-S2 by real-time RT-PCR. Shown are the amounts of sgmRNA-S2 in TRS-S2 mutants relative to the wt mRNA-S2 level. The log levels of mRNA-S2 are represented. The data presented are the averages of results from six independent experiments carried out in triplicate in each case. Error bars represent standard deviations. The  $\Delta G$  value (shown as  $-\Delta G$ ) of the formation of the TRS-L-cTRS-S2 duplex is represented for each virus.

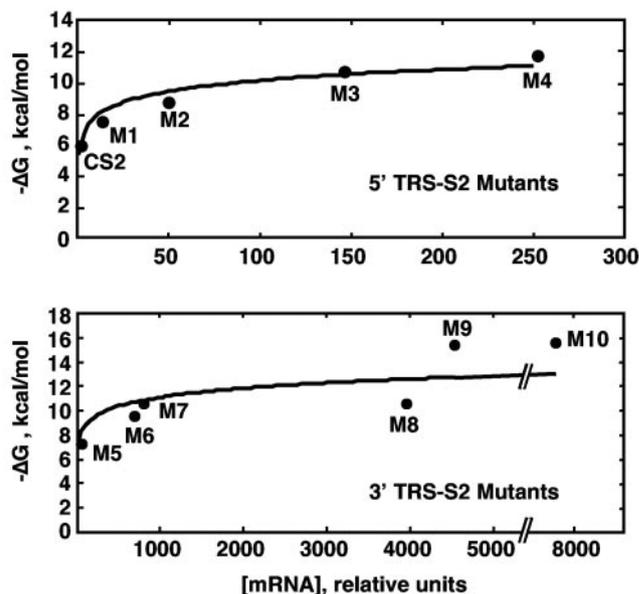


FIG. 6. Relation between mRNA-S2 transcription levels and the  $\Delta G$  of the formation of the TRS-L-TRS-B duplex. Shown is a graphical representation of the  $\Delta G$  of the formation of the TRS-L-cTRS-S2 duplex (shown as  $-\Delta G$ ) versus the amount of mRNA-S2 in TRS-S2 mutants, relative to the wt mRNA-S2 level. The upper curve represents the TRS-S2 mutants in which the complementarity with the TRS-L was extended by replacing nucleotides at the 5' TRS (mutants M1, M2, M3, and M4). The lower curve represents the mutants in which the complementarity with the TRS-L was extended by replacing the nucleotides at the 3' TRS (M5, M6, M7, and M8) or at both the 5' and 3' TRSs (M9 and M10). The scale for the relative amounts of mRNA in both representations is significantly shifted.

a minimal increase in the complementarity between TRS-L and cTRS-B by a single substitution in the nucleotides immediately flanking 5' or 3' the CS-S2 (M1 or M5, respectively), promoted the synthesis of detectable amounts of sgmRNA-S2, indicating that sgmRNA synthesis in coronaviruses requires a minimum thermodynamic stability in the TRS-L and cTRS-B duplex. This result implies that complementarity limited to CS-L and cCS-B is not sufficient to drive sgRNA synthesis (Fig. 7), as observed in the arterivirus EAV (26). Therefore, the nucleotides adjacent to the CS are decisive in the duplex formation between the leader and the nascent negative RNA. In fact, in all TGEV genes, complementarity with TRS-L includes at least one additional nucleotide adjacent to the CS (Fig. 7). In addition, mutant M1 (Fig. 1) extended complementarity with TRS-L by one adenine immediately upstream of the CS, generating a sequence similar to that of TRS-S1, which promotes the synthesis of sgmRNA S. A good correlation was observed between the  $\Delta G$  of the formation of the TRS-L and cTRS-B duplex and the relative amounts of sgmRNA-S2 in TRS-S2 mutants, confirming that the  $\Delta G$  value, as a measure of duplex stability, is a determinant factor during sgRNA synthesis. A similar conclusion was reached for the EAV arterivirus (26), suggesting common elements in the transcriptional regulation of both coronaviruses and arteriviruses, despite the relatively large evolutionary distance.

The mutational analysis leading to extended complementarity with TRS-L was restricted to the four nucleotides imme-

		← TRS →				
		5' TRS	CUAAAC	3' TRS		
		CS				
TRS						TRANSCRIPTION
L		C G A A	CUAAAC	G A A A		
S1		G U U A	CUAAAC	U U U G		+
S2		C C U U	CUAAAC	U A U A		-
3a		A G A A	CUAAAC	U U A C		+
3b		C A U U	CUAAAU	U C C A		-
E		G G U U	CUAAAC	G A A A		+
M		C G A A	CUAAAC	A A A A		+
N		A U A A	CUAAAC	U U C U		+
7		C G A A	CUAAAC	G A G A		+

FIG. 7. Identity between the TRS-L and TRS-Bs of all TGEV sgRNAs. The CS sequence is in white letters in a black box. White boxes highlight the identity in the sequences immediately flanking the CS both at the 5' and 3' ends. The transcriptional activity of each TRS is shown in the right column.

diately flanking upstream and downstream the CS (5'-CGAA CUAAACGAAA-3'). The motif 5'-CGAA-3' or 5'-GAAA-3' was present very frequently in the TGEV gRNA at the 5' or 3' end, respectively, of noncanonical CSs that were used as leader-to-body junction sites in the synthesis of alternative sgRNAs (34, 44), suggesting that an extended complementarity between TRS-L and cTRS-B within these sequences was decisive for a junction event to occur. Furthermore, when an *in silico* approach was used to estimate the potential base-pairing scores between gRNA and TRS-L, the best predictions of viral and alternative sgRNA synthesis were obtained with a TRS-L consisting of the sequence 5'-CGAACUAAACGAAA-3'. Extension of the TRS-L sequence did not improve the prediction of viral sgRNAs, and also failed to predict some alternative sgRNAs, suggesting that complementarity between TRS-L and cTRS-B during sgRNA synthesis is restricted to a limited region. Therefore, these predictions implied that extending duplex formation upstream or downstream of the indicated sequence of 14 nt apparently was not a determining factor for transcription. In TRS-S2 mutants, a tendency to reach a plateau in  $\Delta G$  values with the increase of sgRNA levels was observed. Since  $\Delta G$  values are directly proportional to the number of complementary nucleotides, that plateau would correspond to a limited extension in the TRS-L region involved in duplex formation with cTRS-B. This observation suggests that during the synthesis of sgRNA, only a restricted number of nucleotides within the TRS is relevant for the transcription complex. Since there is a correlation between  $\Delta G$  values and the amount of sgRNA, a limited sgRNA synthesis would also be reached. These results contribute to a more precise definition of the TRS limits in TGEV, consisting of the central CS and about four nucleotides flanking the CS 5' and 3'. However, future knowledge of

RNA-protein interactions regulating the transcription process in the TRS region might modify the prediction of the TRS extension based strictly on the base pairing between TRS-L and cTRS-B. These observations are in agreement with previous results reported for MHV (36) and TGEV (2) minigenomes showing that sgRNA levels were not indefinitely increased by extending duplex formation between TRS-L and cTRS-B. This concept could be supported by the RNA secondary structure shown for the leader region of EAV and predicted for other coronaviruses (35). According to this prediction, TRS-L would be exposed in the loop of a hairpin structure. It has been proposed that extending complementarity with TRS-L would be relevant only for nucleotides residing within the loop and not for nucleotides base paired in the stem (26). RNA secondary-structure predictions for TGEV TRS-L show that the 5'-AACUAAA-3' sequence would be exposed in a small loop but that some adjacent nucleotides would participate in a 3-bp stem connecting the small loop with another large loop. Similar structures were predicted for other coronaviruses, in contrast to the unique large loop in the leader region of the EAV genome (35).

It has also been observed that for similar  $\Delta G$  values, mutants increasing complementarity with TRS-L by the 3' TRS-flanking nucleotides led to higher levels of mRNA-S2 than mutants that extended complementarity by means of the 5' TRS region, indicating that sequences downstream of the CS are more decisive in the RdRp choice during discontinuous sgRNA synthesis. A similar observation was derived from a bovine coronavirus DI system (21). This feature is most consistent with a model of sequence similarity-assisted, polymerase copy choice strand switching during minus-strand synthesis.

A fundamental aspect of the discontinuous transcription model is the identification of the mechanism promoting RdRp to undergo strand transfer and template switching. According to the model that postulates the RdRp jump during minus-strand synthesis, the CS sequence and most probably the TRS sequences exert an attenuating effect on the RdRp. At this postulated site of RdRp pausing, the transcription complex should evaluate the base pairing between the nascent negative RNA strand and the TRS-L. It can be postulated that the CS within the TRS acts as a dissociation signal from the genomic RNA template in order to stimulate a strand transfer to the leader region, providing that duplex formation between TRS-L and cTRS-B has reached a minimum  $\Delta G$ . This model would explain that, for similar  $\Delta G$  values, mutants extending complementarity with TRS-L at the 3' TRS synthesize larger amounts of mRNA-S2 than 5' TRS mutants. Since negative RNA synthesis proceeds from the 3' TRS to the CS and then to the 5' TRS, when the transcription complex reaches the CS, in 3' TRS mutants providing additional base pairing within the 3' TRS-B flank, nascent RNA will include an extended complementarity with the TRS-L, leading to a template switch at the CS site with a higher frequency. In contrast, in mutants with an increased complementarity at the 5' TRS, when the transcription complex arrives at the CS, the complementarity of the nascent negative chain with the TRS-L will be restricted to the six CS nucleotides, providing an insufficient  $\Delta G$  to promote strand transfer to the leader. When 5' TRS nucleotides are added to the nascent negative RNA, increasing complementarity to TRS-L, the dissociation

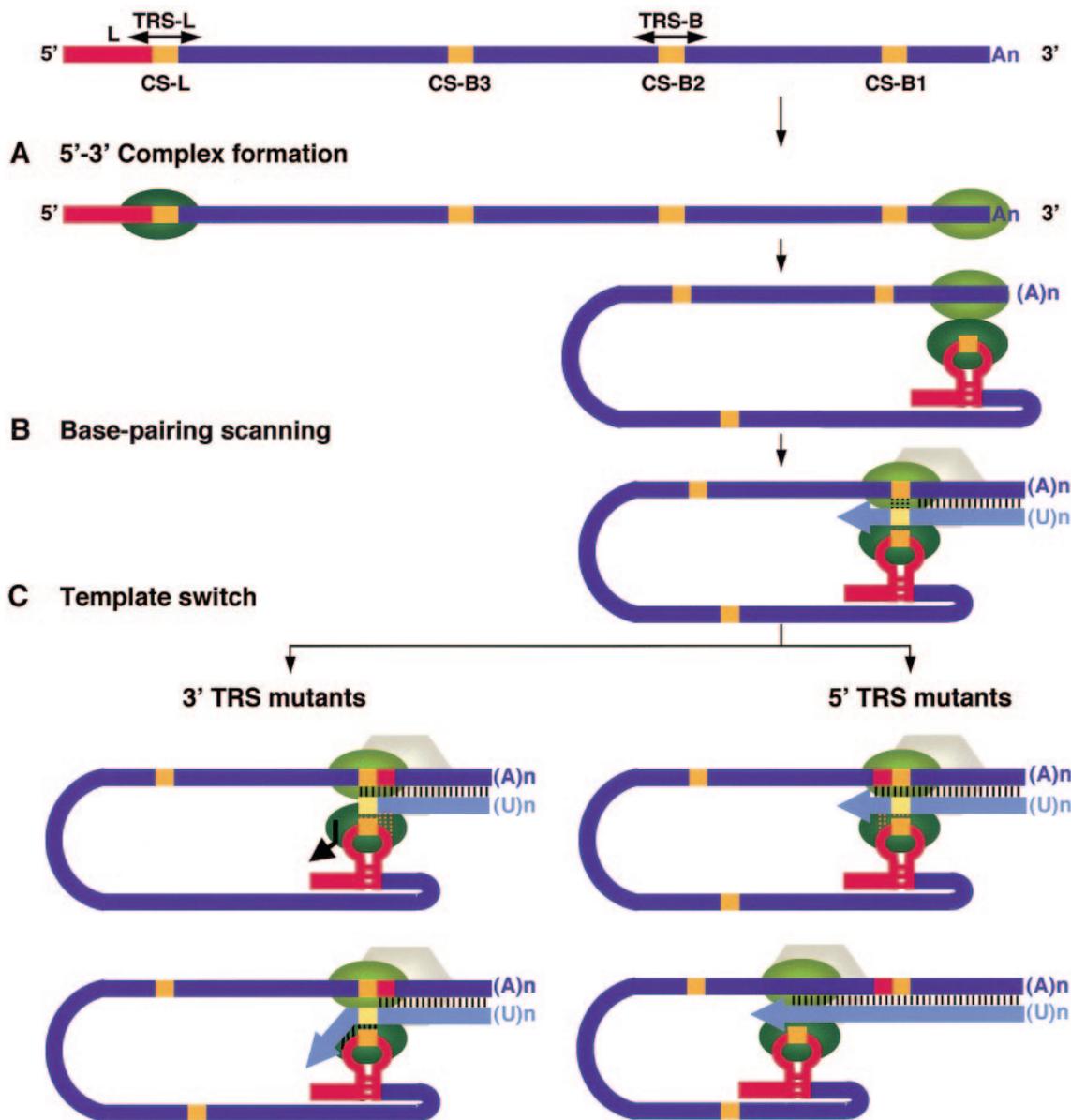


FIG. 8. Three-step working model of coronavirus transcription for subgenomic mRNA synthesis in TRS-S2 mutants extending complementarity with TRS-L either at the 3' TRS or 5' TRS. (A) 5'-3' complex formation step. Proteins binding the 5'- and 3'-end TGEV sequences are represented by the green ovals. The leader sequence, represented as the predicted secondary structure, is in red, and CS sequences are in yellow. (A)<sub>n</sub>, poly(A) tail. (B) Base-pairing scanning step. Negative-strand RNA is in a lighter color than positive-strand RNA. The replication complex is represented by the hexagon. Vertical dotted bars represent the base-pairing scanning by the TRS-L during synthesis of nascent mRNA at attenuation sites. Vertical solid bars indicate complementarity between gRNA and the nascent negative-strand. (U)<sub>n</sub>, poly(U) tail. (C) Template switch step. The thick arrow indicates the switch in the template made by the transcription complex to complete the synthesis of negative sgRNA.

tion signal CS will have been overcome, and strand transfer will be a less frequent event.

This hypothesis can be integrated into the previously proposed working model of coronavirus transcription (Fig. 8) (44), including three steps. The first step is the formation of 5'-end-3'-end complexes mediated by protein-RNA and protein-protein interactions, which locate the TRS-L in close proximity to sequences at the 3' end of genomic RNA. In the MHV system several host proteins, such as the heterogeneous nuclear ribonucleoprotein A1 and the polypyrimidine tract binding protein, probably involved in these interactions, have been identified (4, 10, 33).

Alternatively, complexes between the TRS-L and the TRSs-B present along the coronavirus genome may have previously been formed based on sequence identity and protein-RNA interactions. The second step is scanning of the nascent negative chain by the TRS-L, which looks for complementary sequence domains leading to a favorable  $\Delta G$ . The third step is the template switch of the nascent negative RNA strand to the leader TRS to resume the synthesis of negative sgRNA, when complementarity is above a certain threshold.

Multiple factors seem to regulate the transcription process (2, 25, 26, 44). Complementarity between TRS-L and cTRS-B

has been confirmed in this work as a crucial regulating factor. The relative order of body TRSs in the viral genome may also determine the relative amounts of sgmRNAs (9, 13, 15, 39), and most probably viral and host components involved in protein-RNA and protein-protein recognition will also be decisive in transcription.

#### ACKNOWLEDGMENTS

We thank F. Almazán and D. Escors for critically reading the manuscript and for helpful discussions. We are also grateful to J. C. Oliveros for the PEARL script used in the *in silico* analysis and to Diana Dorado for technical assistance.

This work was supported by grants from the Comisión Interministerial de Ciencia y Tecnología (CICYT), la Consejería de Educación y Cultura de la Comunidad de Madrid, Fort-Dodge Veterinaria, and the European Communities (Frame V, Key Action 2, Control of Infectious Disease Projects). I.S., J.L.M., and S.Z. received fellowships from the Community of Madrid and the European Union (Frame V, Key Action 2, Control of Infectious Disease Projects QLRT-2000-00874, QLRT-2001-00825, and QLRT-2001-01050).

#### REFERENCES

- Almazán, F., J. M. González, Z. Péntzes, A. Izeta, E. Calvo, J. Plana-Durán, and L. Enjuanes. 2000. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc. Natl. Acad. Sci. USA* **97**: 5516–5521.
- Alonso, S., A. Izeta, I. Sola, and L. Enjuanes. 2002. Transcription regulatory sequences and mRNA expression levels in the coronavirus-transmissible gastroenteritis virus. *J. Virol.* **76**:1293–1308.
- Brian, D. A., and W. J. M. Spaan. 1997. Recombination and coronavirus defective interfering RNAs. *Semin. Virol.* **8**:101–111.
- Choi, K. S., P.-Y. Huang, and M. C. C. Lai. 2002. Polypyrimidine-tract-binding protein affects transcription but not translation of mouse hepatitis virus RNA. *Virology* **303**:58–68.
- Curtis, K. M., B. Yount, and R. S. Baric. 2002. Heterologous gene expression from transmissible gastroenteritis virus replicon particles. *J. Virol.* **76**:1422–1434.
- Delmas, B., J. Gelfi, H. Sjöström, O. Noren, and H. Laude. 1994. Further characterization of aminopeptidase-N as a receptor for coronaviruses. *Adv. Exp. Med. Biol.* **342**:293–298.
- Enjuanes, L., W. Spaan, E. Snijder, and D. Cavanagh. 2000b. *Nidovirales*, p. 827–834. In M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carsten, M. K. Estes, S. M. Lemon, D. J. McGeoch, J. Maniloff, M. A. Mayo, C. R. Pringle, and R. B. Wickner (ed.), *Virus taxonomy. Classification and nomenclature of viruses*. Academic Press, San Diego, Calif.
- González, J. M., Z. Péntzes, F. Almazán, E. Calvo, and L. Enjuanes. 2002. Stabilization of a full-length infectious cDNA clone of transmissible gastroenteritis coronavirus by the insertion of an intron. *J. Virol.* **76**:4655–4661.
- 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.
- Huang, P., and M. M. C. Lai. 2001. Heterogeneous nuclear ribonucleoprotein A1 binds to the 3′-untranslated region and mediates potential 5′-3′-end cross talks of mouse hepatitis virus RNA. *J. Virol.* **75**:5009–5017.
- Huang, X., and W. Miller. 1991. A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* **12**:337–357.
- Jiménez, G., I. Correa, M. P. Melgosa, M. J. Bullido, and L. Enjuanes. 1986. Critical epitopes in transmissible gastroenteritis virus neutralization. *J. Virol.* **60**:131–139.
- Joo, M., and S. Makino. 1995. The effect of two closely inserted transcription consensus sequences on coronavirus transcription. *J. Virol.* **69**:272–280.
- Konings, D. A. M., P. J. Bredenbeek, J. F. H. Noten, P. Hogeweg, and W. J. M. Spaan. 1988. Differential premature termination of transcription as a proposed mechanism for the regulation of coronavirus gene expression. *Nucleic Acids Res.* **16**:10849–10860.
- 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.
- Lai, M. M. C., and D. Cavanagh. 1997. The molecular biology of coronaviruses. *Adv. Virus Res.* **48**:1–100.
- Mathews, D. H., J. Sabina, M. Zuker, and D. H. Turner. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**:911–940.
- McClurkin, A. W., and J. O. Norman. 1966. Studies on transmissible gastroenteritis of swine. II. Selected characteristics of a cytopathogenic virus common to five isolates from transmissible gastroenteritis. *Can. J. Comp. Med. Vet. Sci.* **30**:190–198.
- Miller, W. A., and G. Koev. 2000. Synthesis of subgenomic RNAs by positive-strand RNA virus. *Virology* **273**:1–8.
- Nagy, P. D., and A. E. Simon. 1997. New insights into the mechanisms of RNA recombination. *Virology* **235**:1–9.
- Ozdarendeli, A., S. Ku, S. Rochat, S. D. Senanayake, and D. A. Brian. 2001. Downstream sequences influence the choice between a naturally occurring noncanonical and closely positioned upstream canonical heptameric fusion motif during bovine coronavirus subgenomic mRNA synthesis. *J. Virol.* **75**:7362–7374.
- Pasternak, A. O. 2003. Nidovirus transcription-regulating sequences. Ph.D. thesis. Leiden University, Leiden, The Netherlands.
- Pasternak, A. O., A. P. Gulyaev, W. J. 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.
- Pasternak, A. O., W. J. M. Spaan, and E. J. Snijder. 2004. Regulation of relative abundance of arterivirus subgenomic mRNAs. *J. Virol.* **78**:8102–8113.
- Pasternak, A. O., E. van den Born, W. J. M. Spaan, and E. J. Snijder. 2001. Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis. *EMBO J.* **20**:7220–7228.
- Pasternak, A. O., E. van den Born, W. J. M. Spaan, and E. J. Snijder. 2003. The stability of the duplex between sense and antisense transcription-regulating sequences is a crucial factor in arterivirus subgenomic mRNA synthesis. *J. Virol.* **77**:1175–1183.
- Péntzes, Z., J. M. González, E. Calvo, A. Izeta, C. Smerdou, A. Mendez, C. M. Sánchez, I. Sola, F. Almazán, and L. Enjuanes. 2001. Complete genome sequence of transmissible gastroenteritis coronavirus PUR46-MAD clone and evolution of the Purdue virus cluster. *Virus Genes* **23**:105–118.
- 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**:386–396.
- Sawicki, S. G., and D. L. Sawicki. 1990. Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. *J. Virol.* **64**:1050–1056.
- Sawicki, S. G., and D. L. Sawicki. 1998. A new model for coronavirus transcription. *Adv. Exp. Med. Biol.* **440**:215–220.
- Schaad, M., and R. S. Baric. 1994. Genetics of mouse hepatitis virus transcription: evidence that subgenomic negative strands are functional templates. *J. Virol.* **68**:8169–8179.
- Sethna, P. B., S.-L. Hung, and D. A. Brian. 1989. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proc. Natl. Acad. Sci. USA* **86**:5626–5630.
- Shi, S. T., P. Huang, H.-P. Li, and M. M. C. Lai. 2000. Heterogeneous nuclear ribonucleoprotein A1 regulates RNA synthesis of a cytoplasmic virus. *EMBO J.* **19**:4701–4711.
- Sola, I., S. Alonso, S. Zúñiga, M. Balach, J. Plana-Durán, and L. Enjuanes. 2003. Engineering transmissible gastroenteritis virus genome as an expression vector inducing latrogenic immunity. *J. Virol.* **77**:4357–4369.
- van den Born, E. 2004. Secondary structure and function of the 5′-proximal region of the equine arteritis virus RNA genome. *RNA* **10**:424–437.
- 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.* **68**: 3656–3666.
- 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.
- van Marle, G., J. C. Dobbe, A. P. Gulyaev, 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.
- 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.
- Yount, B., K. M. Curtis, and R. S. Baric. 2000. Strategy for systematic assembly of large RNA and DNA genomes: the transmissible gastroenteritis virus model. *J. Virol.* **74**:10600–10611.
- Yu, W., and J. L. Leibowitz. 1995. Specific binding of host cellular proteins to multiple sites within the 3′ end of mouse hepatitis virus genomic RNA. *J. Virol.* **69**:2016–2023.
- Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**:3406–3415.
- Zuker, M., D. H. Mathews, and D. H. Turner. 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. *NATO ASI Ser.* **3**:11–43.
- Zúñiga, S., I. Sola, S. Alonso, and L. Enjuanes. 2004. Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *J. Virol.* **78**:980–994.