Sequence Motifs Involved in the Regulation of Discontinuous Coronavirus Subgenomic RNA Synthesis

Sonia Zúñiga, Isabel Sola, Sara Alonso, and Luis Enjuanes*

Centro Nacional de Biotecnología, CSIC, Department of Molecular and Cell Biology, Campus Universidad Autónoma, Cantoblanco, 28049 Madrid, Spain

Received 14 July 2003/Accepted 1 October 2003

Transmissible gastroenteritis virus (TGEV) is a member of the Coronaviridae family, included in the Nidovirales order (10). TGEV is an enveloped virus with a single-stranded, positive-sense 28.5-kb RNA genome (28) for which infectious cDNA clones have been engineered (1, 12, 41). About the 5’ two-thirds of the entire RNA comprises open reading frames (ORFs) 1a and 1ab encoding the replicase (rep) genes. A leader sequence of 93 nucleotides (nt), derived from the 5’ end of the genome, is fused to the 5’ end of the leader (TRS-L). The conserved sequence of this TRS, the CS-L, is located in an appropriate sequence context (2). The concept was reinforced by demonstrating in arterivirus that the CS included in the sgRNA was derived from the CS preceding each gene (CS-B) and not from the CS present at the 3’ end of the leader sequence (CS-L) (26, 38) (Fig. 1). According to this model of discontinuous sgRNA synthesis during production of the negative strand, the TRS-B acts as a slow-down and detaching signal for the transcription complex.

Transcription regulation is probably a multifactor process in which three factors may have a relevant role: (i) base pairing between the TRS-L and the nascent negative strand, (ii) proximity to the 3’ end of the genome, and (iii) RNA-protein and protein-protein interactions within TRSs.

The synthesis of a negative sgRNA is most probably mediated by a direct base-pairing interaction between the nascent negative body TRS (cTRS-B) and the 3’ end of the leader (TRS-L). The conserved sequence of this TRS, the CS-L, is probably exposed in a stem-loop at the 5’ end of the viral genome both in TGEV (S. Alonso, I. Sola, S. Zúñiga, and L. Enjuanes, unpublished) and in equine arteritis virus (EAV) (26, 38), although this RNA structure has not been experimentally proven.

Proximity to the 3’ end of the genome probably influences the relative amount of sgRNAs, because the polymerase complex finds less slow-down and detaching signals during small negative sgRNA synthesis. Therefore, in principle, these
RNAs could be the most abundant. Although this is the case in the order Mononegavirales (15, 39) and, in general, in coronaviruses, the relative amounts of coronavirus mRNAs are not strictly related to their proximity to the viral 3' end (28, 37). Therefore, other factors may also regulate coronavirus transcription.

The interaction of RNA with viral and cellular proteins is probably involved in coronavirus transcription. The discontinuous synthesis of the negative RNA strand resembles a high-frequency copy-choice RNA recombination (3, 21, 26), in which the TRS-B (donor) and TRS-L (acceptor) sequences, located in distal domains in the RNA primary structure, are probably brought into physical proximity by RNA-protein and protein-protein interactions (Fig. 1C).

In arterivirus, base pairing between the leader CS and the negative-sense body CS (cCS-B) has been implicated in transcription, although the roles of other factors, such as relative TRS position in the genome and secondary structure, have led to less clear conclusions (25–27).

In this report, the role of CS sequences in coronavirus transcription is analyzed for the first time by using TGEV full-length genomes constructed with an infectious cDNA clone (1). The role of each nucleotide within the leader and body CSs has been studied by introducing point mutations in these sequences. A key strategy in these studies has been analysis of gene 3a transcription, because this gene is nonessential for TGEV replication (36). Therefore, infectious virus was rescued for all gene 3a CS-B mutants, allowing subsequent analysis. We show in the studies reported here that the presence of the highly conserved CS was associated with sgRNA production and high virus titers, but that this sequence was not essential for sgRNA synthesis when the TRS-L to cTRS-B duplex formation involved a high release of free energy ($\Delta G$). In fact, the genome positions in which a negative sgRNA most frequently fused to the leader could be predicted in silico by determining the identity between the TRS-L and sequence domains of the genome. To this end, a computer-based program has been developed to assess the strength of base pairing between body and leader TRS that successfully predicts the authentic products as well as novel, mutant-derived sgRNAs. In addition, it has been shown that nucleotide substitutions in the canonical CS led to the use of alternative noncanonical CSs, providing that sequences flanking the CS-L were also flanking the CS-B, leading to a favorable $\Delta G$ in duplex formation between TRS-L and cTRS-B. It has also been shown that during the synthesis of TGEV negative sgRNAs, template switching always took place after copying the canonical or noncanonical CS sequence, supporting the finding that coronavirus RNA discontinuous synthesis takes place during production of the negative strand. A three-step mechanism has been proposed as a working model for coronavirus mRNA transcription.

MATERIALS AND METHODS

Cells and viruses. Baby hamster kidney cells (BHK-21) stably transformed with the gene coding for the porcine aminopeptidase N (BHK-pAPN) (6) were
grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 5% fetal calf serum (FCS) and G418 (1.5 mg/ml) as a selection agent. Viruses were grown in swine testis (ST) cells (20).

**Plasmid constructs.** TGEV cDNAs with point mutations in the leader and body CS were generated by overlapping PCR. To get leader CS mutants, the plasmid pBAC-TGEV(SrfI-NheI), which bears nt 1 to 15062 from the TGEV genome (GenBank accession no. AJ271965) except a ClaI-ClaI fragment (nt 4417 to 9615) (1), was used as template. Overlapping PCR fragments with point mutations were amplified by using the oligonucleotides described in Table 1. The final PCR product (2,415 bp), amplified with outer oligonucleotides Oli 5′/H11032I and Oli 3′/H11032D, was digested with SfiI and ApaLI and cloned into the same restriction sites of plasmid pBAC-TGEV(SrfI-NheI). To introduce mutations in the TGEV

---

### Table 1. Oligonucleotides used for site-directed mutagenesis

<table>
<thead>
<tr>
<th>Mutant or other</th>
<th>Oligonucleotide&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Oligonucleotide sequence 5′ → 3′&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-mutant oligonucleotide</td>
<td>Oli 5′I</td>
<td>CGGGAAATTTCGATGTAAGCTGTCAAACAC</td>
</tr>
<tr>
<td>L-C1G</td>
<td>MutC94G-VS</td>
<td>CAACCTCGAATACAGGAAATATT</td>
</tr>
<tr>
<td>L-U2G</td>
<td>MutU95G-VS</td>
<td>CAACCTCGAAGAAAGGAAATATT</td>
</tr>
<tr>
<td>L-A3C</td>
<td>MutA96C-VS</td>
<td>CAACCTCGAATACAGGAAATATT</td>
</tr>
<tr>
<td>L-A4C</td>
<td>MutA97C-VS</td>
<td>CAACCTCGAATACAGGAAATATT</td>
</tr>
<tr>
<td>L-A5C</td>
<td>MutA98C-VS</td>
<td>CAACCTCGAATACAGGAAATATT</td>
</tr>
<tr>
<td>L-C6G</td>
<td>MutC99G-VS</td>
<td>CAACCTCGAATACAGGAAATATT</td>
</tr>
<tr>
<td>L-C1U</td>
<td>MutC94U-VS</td>
<td>CAACCTCGAATACAGGAAATATT</td>
</tr>
<tr>
<td>L-A3U</td>
<td>MutA96U-VS</td>
<td>CAACCTCGAATACAGGAAATATT</td>
</tr>
<tr>
<td>L-A4U</td>
<td>MutA97U-VS</td>
<td>CAACCTCGAATACAGGAAATATT</td>
</tr>
<tr>
<td>L-A5U</td>
<td>MutA98U-VS</td>
<td>CAACCTCGAATACAGGAAATATT</td>
</tr>
<tr>
<td>L-C6U</td>
<td>MutC99U-VS</td>
<td>CAACCTCGAATACAGGAAATATT</td>
</tr>
<tr>
<td>B-mutant oligonucleotide</td>
<td>S-3839-VS</td>
<td>GTTGCAACTAGTTCGAGTC</td>
</tr>
<tr>
<td>B-C1G</td>
<td>MutC24798G-VS</td>
<td>TTTAAGAAACTTACGAGTC</td>
</tr>
<tr>
<td>B-U2G</td>
<td>MutU24799G-VS</td>
<td>TTTAAGAAACTTACGAGTC</td>
</tr>
<tr>
<td>B-A3C</td>
<td>MutA24800C-VS</td>
<td>TTTAAGAAACTTACGAGTC</td>
</tr>
<tr>
<td>B-A4C</td>
<td>MutA24801C-VS</td>
<td>TTTAAGAAACTTACGAGTC</td>
</tr>
<tr>
<td>B-A5C</td>
<td>MutA24802C-VS</td>
<td>TTTAAGAAACTTACGAGTC</td>
</tr>
<tr>
<td>B-C6G</td>
<td>MutC24803G-VS</td>
<td>TTTAAGAAACTTACGAGTC</td>
</tr>
</tbody>
</table>

---

<sup>a</sup> Virus names were derived from leader CS mutants (L) or CS-3a mutants (B) and indicate the nucleotide substitution and its position in the CS.

<sup>b</sup> Oligonucleotides including the punctual mutations are named “Mut” and indicate the nucleotide substitution and its position at the TGEV genome. VS, virus sense; RS, reverse sense.

<sup>c</sup> The mutated nucleotide is shown in boldface. Restriction endonuclease sites used for cloning are shown in italics (EcoRI, GAATTC; SpeI, ACTAGT). CS and cCS are underlined.
infectious cDNA, SfiI-ClaI fragment (5,277 bp) from pBAC-TGEV(SflI-NheI) with the corresponding mutation was cloned into the same sites of pBAC-TGEV(StuI-AvrII) with the corresponding AvrII digestion product (2,900 bp) from pSL(SflI). To introduce mutations in the TGEV infectious cDNA, AvrII digestion product (2,900 bp) from pSL(AvrII-AvrII) with the corresponding mutation was cloned into the same sites of pBAC-TGEV(ClaI). To obtain the full-length TGEV cDNA, the toxic ClaI-ClaI fragment (5,198 bp) was introduced as previously described (1).

Double CS-3a and CS-B mutants were obtained by introducing SfiI-ApuLI fragment from pBAC-TGEV(SflI-NheI) plasmid with the leader mutation into the same restriction sites of pBAC-TGEV(StuI-AvrII) bearing the corresponding CS-3a mutation. The plasmid containing the full-length TGEV cDNA with point mutations was then generated as previously described.

All cloning steps were checked by sequencing the PCR-amplified fragments and cloning junctions.

**TABLE 2. Reverse oligonucleotides used for RT-PCR analysis of RNA from rTGEV-infected cells**

<table>
<thead>
<tr>
<th>sgRNA</th>
<th>Primer</th>
<th>Sequence (5’→3’)</th>
<th>Amplicon size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genomic</td>
<td></td>
<td>TCTTTCGATGCAATCCAA</td>
<td>473</td>
</tr>
<tr>
<td>mRNA-S</td>
<td>5’-449</td>
<td>TAACTGGCACATACCTCC</td>
<td>499</td>
</tr>
<tr>
<td>mRNA-3a</td>
<td>3’-169-49</td>
<td>CAATAATGGAGAGACCAA</td>
<td>295</td>
</tr>
<tr>
<td>mRNA-3b</td>
<td>X2B-112</td>
<td>TTACATACACAAAGATGTC</td>
<td>458</td>
</tr>
<tr>
<td>mRNA-E</td>
<td>IGS</td>
<td>CAGTCGACACGGCTGCGGCCGCGCCTTATTGTCAAGCC</td>
<td>393</td>
</tr>
<tr>
<td>mRNA-M</td>
<td>M.415RS</td>
<td>AGACCCCAAGAGTTTGCTTC</td>
<td>530</td>
</tr>
<tr>
<td>mRNA-N</td>
<td>N-268RS</td>
<td>GGTCCGGTTAATCTAGTAGTAAGAACC</td>
<td>386</td>
</tr>
<tr>
<td>mRNA-7</td>
<td>7(213)RS</td>
<td>TCCTGAGCAGAAAATCC</td>
<td>302</td>
</tr>
</tbody>
</table>

**RESULTS**

Relevance of base pairing between the CS-L and cCS-B in coronavirus transcription. To study the relevance of the base pairing between CS-L and cCS-B, each of the 6 nt was substi-
tuted in the CS-L or in the gene 3a CS (CS-3a). Nucleotide changes within CS-3a, in principle, should only affect the synthesis of mRNA-3a. In contrast, nucleotide substitutions in the CS-L would have a pleiotropic effect on the synthesis of all mRNAs. Four groups of mutant TGEV infectious cDNA clones were generated (Fig. 2): (i) CS-B mutants, replacing each base of CS-3a by nucleotides that do not allow base pairing of the cCS-B with the CS-L (Fig. 2A); (ii) CS-L mutants with changes identical to those introduced in the CS-B mutants (Fig. 2B); (iii) CS-L mutants with changes allowing non-Watson-Crick base pairing with the body cCS (non-Watson-Crick mutants [C]). Virus titers (PFU per milliliter) obtained for the passage 0 supernatant are indicated in the figure.

Viruses were recovered from all CS-3a mutants, with titers similar to those obtained with the wild-type TGEV cDNA (Fig. 2A), as expected, since gene 3a is nonessential. In contrast, no virus was recovered from cDNAs when CS-L nt 1 to 3 were changed in the single or double mutants (Fig. 2B and D). Nucleotide substitutions in CS-L positions 4 to 6 led to the recovery of infectious recombinant TGEV (rTGEV) with titers up to 105-fold lower than the parental ones. Leader and double mutants showed the same behavior (Fig. 2B and D), as expected, since leader mutations affected the synthesis of all sgRNAs.

Interestingly, infectious rTGEV was recovered from all non-Watson-Crick leader mutants, with titers ranging from wild-type levels, like those obtained for L-C6U mutant, to 105-fold lower for the L-C1U mutant (Fig. 2C). Overall, these data indicated the requirement of base pairing between CS-L and cCS-B for sgRNA synthesis.

Relationship between CS-L and CS-3a sequences and sgRNA levels. It was postulated that synthesis of negative sgRNAs is mediated by direct base pairing between the TRS-L and the cTRS-B. This being the case, the CS-L and CS-3a sequences should modulate sgRNA-3a levels. To determine whether this was the case, the pattern of sgRNA synthesis produced by different rTGEVs with CS point mutations was analyzed by Northern blotting (Fig. 3). Nucleotide substitutions within the first 3 nt of CS-L led to no virus rescue, and it was not possible to analyze the sgRNA pattern. To evaluate sgRNA synthesis by Northern blot analysis, because mutations in CS-L sequence positions 4 to 6 considerably reduce sgRNA production, the multiplicity of infection (MOI) and the amount of total RNA from the leader and double mutants loaded in the gel were increased in order to obtain similar levels of viral RNA (Fig. 3). The viral sgRNA pattern for the wild-type virus was the expected one, but new bands were identified in all CS mutants (Fig. 3). Some of these unexpected bands were amplified by RT-PCR and sequenced, corresponding to alternative sgRNAs for the S, 3a, and N genes. These data indicated that changes in the CS-L or CS-B opened new base-pairing possibilities throughout the genome, leading to the generation of alternative sgRNAs.

The sgRNA pattern for CS-3a mutants was also analyzed by RT-PCR. After ST cell infection with rTGEVs, total RNA was extracted, and genomic sequences from gene 3a were amplified by RT-PCR with oligonucleotides S-3839-VS and 3a-169-RS (Table 1 and Fig. 4A). Sequencing of these RT-PCR products showed that the nucleotide substitutions introduced within CS-3a were stably maintained during virus passage.

Using primers specific for mRNA-3a detection (SP and 3a-
169-RS, Materials and Methods and Table 2) a single sgRNA was detected in the wild-type virus, while in all CS-B mutants, three RT-PCR amplification bands were observed (Fig. 4B). This pattern was the same in the six mutants, although the relative band intensities were different. Moreover, sequencing of these cDNAs revealed that mRNA-3a.1 corresponded to the wild-type mRNA-3a, generated by a leader-to-body junction site within the CS-3a. The mRNA-3a.2 was generated in all CS-B mutants from a leader-to-body fusion site inside ORF S, 121 nt upstream of CS-3a. The third band (mRNA-3a.3) arose from a junction site 64 nt downstream of CS-3a, inside gene 3a.

Sequencing of the leader-to-body junction sites in the three sgRNA-3a species showed that there was an extended identity between TRS-L and gRNA in sequence domains around the noncanonical CSs used (Fig. 4C). Interestingly, all of the
mutations introduced in the CS-3a appeared in the mRNA-3a.1, including a substitution in the first CS-B nucleotide (B-C1G mutant), indicating that at least the whole-body CS was copied before template transfer. Nevertheless, because an extended upstream sequence identity is observed between the CS-L and CS-3a flanking sequences, the strand transfer point could not be accurately established. Even for the B-C1G mutation that remained in the mRNA-3a.1 sequence, strand transfer could happen in any of the 5′-GAA-3′ nucleotides upstream of CS-3a. However, in mutants B-A3C (Fig. 4C) and B-A5C (data not shown), template transfer had to occur at the A nucleotide, preceding GAA sequence upstream CS-3a, because the mRNA-3a.1 included the sequence 5′-AGAAACUA AAC-3′ (Fig. 4C) derived from the gRNA sequence. The identity between leader and body sequences was frequently extended by including all or part of the sequence 5′-GAA-3′, at either the CS 5′ or 3′ end, or at both ends (Fig. 4C), suggesting that template switching during transcription required high complementarity between TRS-L and cTRS-B.

The transcription pattern in CS-3a mutants of proximal (gene E) or distal upstream (gene S) or downstream (gene 7) TGEV genes was analyzed by RT-PCR using specific oligonucleotides (Table 2), and no alteration was observed in the relative synthesis of these TGEV mRNAs (Fig. 5). These data suggested that the template switch was dependent on the nature of local sequences and was not influenced by sequences mapping 5′ or 3′ downstream.

Relationship between potential base pairing of the TRS-L with nascent negative RNA sequences and template transfer. Termination of negative sgRNA synthesis seems to take place at sequence domains with high complementarity with the TRS-L. This complementarity would be the consequence of an identity between the TRS-L and sequences mapping throughout the genome. To determine whether a high identity score would promote template switching during the synthesis of viral negative RNA, an in silico approach was used that was based on a local alignment algorithm (14) that estimates the identity between the genomic RNA and the TRS-L, comprising the CS (5′-CUAAC-3′) plus 3, 4, or 5 nt flanking the CS both at the 5′ and 3′ ends. In the case of 5 nt flanking the CS at both ends, the sequence considered was 5′-TCGAACTAAACGAAAT-3′ (the CS sequence is in boldface). In these three cases, the patterns of sequence domains with high identity were similar, differing only quantitatively.

Base-pairing scores throughout the 5′ two-thirds of the genome were very low (below a value of 35), except at the TRS-L, which obviously has the maximum base pairing score (a value of 70) (data not shown). Interestingly, potential base pairing throughout the one-third 3′ end of the genome, encoding the structural and nonstructural proteins, showed that the sequences with highest local identity correlated with template transfer sites leading to generation of the standard TGEV mRNAs (Fig. 6A). Intermediate values of local complementarity (between 32 and 40) were associated with the generation of sgRNAs alternative to those generated by template transfer at positions of canonical CS-Bs. In contrast, no sgRNAs were detected at sequence positions with a low potential base-pairing score (data not shown), suggesting a dominant role for the complementarity between TRS-L and cTRS-B in the control of sgRNA levels.

Analysis of the potential base pairing between the TRS-L and sequences in the gRNA complementary to the fusion site of gene 3a showed that the three peaks of higher identity score surrounding the CS-3a corresponded to the canonical and non-canonical leader-to-body junction sites found in all CS-B mutants, generating mRNAs 3a.1, 3a.2, and 3a.3 (Fig. 6A). In silico analysis of the potential base pairing within this sequence domain showed that the potential base-pairing patterns were almost identical for all CS-3a mutants (Fig. 6B). The highest TRS-L to TRS-3a identity corresponded to the junction site 3a.1. In contrast, TRS-L to TRS-3a identity decreased in this sequence domain in all CS-3a mutants and was very close to the value at the sgRNA-3a.3 leader-to-body junction site. In these cases, the highest base-pairing value corresponded to the junction site upstream of the CS-3a sequence, within ORF S, generating sgRNA-3a.2. These results could explain the generation of the same new sgRNA species in all of the body mutants, despite the nucleotide change introduced, and suggested an important role for the GAA CS flanking sequence in junction site selection, especially when a nucleotide substitution was introduced within the CS-3a.

Influence of CS-L to cCS-B duplex ΔG on sgRNA-3a levels. To study the influence of base pairing between the nascent negative sgRNA and the CS-L on sgRNA synthesis, mRNA-3a.1 levels were quantified in all CS-B mutants by real-time RT-PCR using specific oligonucleotides (Table 3) and the gRNA as an internal standard for mRNA evaluation. The concentration of mRNA-3a.1 in CS-B mutant viruses was expressed in relation to that of the wild type. The results showed a significant decrease in mRNA-3a.1 levels of up to 103-fold and a good correlation between mRNA-3a.1 concentration and duplex ΔG except for nucleotide substitutions at both the 5′ and 3′ ends of the CS (B-C1G and B-C6G mutants) that had a higher effect than expected on sgRNA levels (Fig. 7A). The additional decrease in the amount of mRNA-3a.1 in the B-C1G mutant could be due to the importance of this nucle-
otide to prime the synthesis of negative sgRNA after template switching. In addition, both the first and last CS nucleotides could play the extra role of stabilizing the formation of a duplex between the exposed CS-L and the cCS-B.

The mRNA-3a.1 levels were also quantified by real-time RT-PCR in the leader and double mutants with CS substitutions at positions 4 to 6 (data not shown). The amount of mRNA-3a.1 decreased in CS-L mutants at least 10-fold in relation to mRNA-3a.1 levels in the wild type. In mutants D-A4C and D-A5C, the amount of mRNA-3a.1 was similar to

FIG. 6. In silico analysis of the identity between TRS-L and the TGEV genome. As indicated in Materials and Methods, a continuous line graph was selected to facilitate visualization of the data. (A) Graphical plot of the potential base-pairing score versus the genome position. All peaks assigned to the viral CSs are indicated as the peaks corresponding to the new 3a sgRNA species. (B) Graphical plot of the potential base-pairing score versus the genome position around CS-3a. Each three-dimensional line represents either the wild-type (wt) situation or the body mutants. The peaks assigned to each 3a sgRNA species are indicated.
that obtained for the wild-type virus. However, mRNA-3a.1 levels were not restored in double mutant D-C6G and were at least 10^3-fold lower than that in the wild type (data not shown), reinforcing the possibility of an extra role for the nucleotide in the last position of the CS, such as the interaction of these sequences with regulatory proteins.

The amount of the alternative mRNA-3a species was also analyzed by real-time RT-PCR using specific oligonucleotides (Table 3). The level of mRNA-3a.2 in the CS-3a mutants did not change significantly when compared with that of the wild-type virus (Fig. 7B). The apparent discrepancy between the relative abundance of the mRNA-3a.1 and mRNA-3a.2 relative to the level of gRNA, expressed as [mRNA] in relative units.

![Figure 7](http://jvi.asm.org/content/82/12/988/F7.large.jpg)

**FIG. 7.** mRNA-3a quantification by real-time RT-PCR. (A) Amount of mRNA-3a.1, quantified by real-time RT-PCR, in the body mutants relative to the wild-type (wt) levels. Shown is a graphical representation of the ΔG (as −ΔG in kilocalories per mole) of the CS-L with cCS-B duplex and the relative amount of mRNA-3a.1 (represented as log [mRNA-3a.1] in relative units) for each virus. The data presented are the average of six independent experiments with duplicates in each case. Error bars represent the standard deviation in each case. (B) Graphical plot of the amounts of mRNA-3a.1 and mRNA-3a.2 relative to the level of gRNA, expressed as [mRNA] in relative units.

specific oligonucleotides was not possible because a duplication of the sequence appears at the leader-to-body fusion site.

**Effect of leader CS mutants on sgRNA levels.** The introduction of nucleotide substitutions at CS-L could affect the potential base pairing between the TRS-L and cTRS-B of all TGEV genes, with the consequent reduction in sgRNA and virus production. Alternatively, the decrease in virus titers could also be due to an effect of CS-L nucleotide substitutions in the TRS-L secondary structure. The transcription model proposed in this article, like the one proposed for arterivirus (26, 38), postulates exposure of the CS-L in a stem-loop within the TRS-L. In agreement with this model, virus production was only observed in TGEV mutants with a CS-L presented as a single-strand RNA according to secondary structure predictions (19; data not shown).

Construction of rTGEVs with nucleotide substitutions not allowing base pairing with cCS-B at each CS-L position led to the rescue of infectious viruses when these mutations were introduced within positions 4 to 6 of the CS, but not in positions 1 to 3. Therefore, the analysis of the sgRNA generated after infection of cells was only possible in mutants with substitutions in positions 4 to 6. Total RNA from infected cells was analyzed by RT-PCR using specific oligonucleotides (Table 2) to amplify gRNA and mRNAs (Fig. 8). Nucleotide substitutions in CS-L positions 4 to 6 led to a reduction in virus titers higher than 10^4-fold in relation to wild-type virus (Fig. 8, bottom). rTGEV mRNAs could be clustered into two sets: one that in general led to a unique sgRNA (genes E, M, and N) and another leading to alternative sgRNAs (genes S, 3a, and 7). The sgRNA corresponding to gene 3b was only produced when the mismatch in the sixth nucleotide of the CS-B present.
in the parental TGEV strain, considered in this report as the wild-type strain (2, 40), was compensated for by the mutation introduced within the CS-L (mutant L-C6U).

The pattern for sgRNAs S, 3a, and 7 in the virus mutants differed from that in the wild-type virus (Fig. 8). Different mutations in the same CS-L nucleotide position led to different sgRNA species (for instance, compare sgRNAs in mutants L-A4C to L-A4U or L-C6G and L-C6U). These results were expected, since changes in the CS-L were creating new possibilities of base pairing with alternative sequences in the nascent negative sgRNA, leading to the formation of new duplexes that could result in novel template switches during negative-strand synthesis and the production of new sgRNA species.

All nucleotide substitutions introduced in the cDNA remained in the rescued virus genome (data not shown). Moreover, sequencing of 72 viral mRNA leader-to-body junction sites included in the sgRNAs identified (Fig. 8) showed that nucleotide substitutions within the CS-L did not appear in the mRNA sequence, confirming that the CS sequence in the mRNA came from CS-B (data not shown). These results strongly suggest that the template switch was produced during negative sgRNA synthesis.

Synthesis of alternative sgRNAs in viruses with nucleotide substitutions in CS-L. Mutations in CS-L led to the formation of at least five different sgRNA-S species, named mRNA-S.1 (wild type) to mRNA-S.5 (Fig. 9A). Some of these sgRNA species, such as mRNA-S.2 and mRNA-S.4, were indistinguishable in agarose gel electrophoresis because of their similar size. RT-PCR amplification and sequencing of leader-to-
body junction sites showed four new junction domains (extending nt 20291 to 20644 of the TGEV genome) leading to the synthesis of new sgRNA species (Fig. 9B). Extended complementarity between leader and body sequences, mediated by the 5’-GAA-3’ sequence involved in the TRS-L with cTRS-B base pairing with noncanonical junction sites, was possible in all cases (Fig. 9B). Most likely, this extended complementarity leads to a higher base-pairing score between the nascent negative RNA strand and the TRS-L, promoting a template switch in these sequence positions and production of the corresponding sgRNAs.

The TGEV sequence between nt 20438 and 20459 presented high identity to TRS-L (Fig. 6A), and different base pairings were possible, depending on the leader mutant. In fact, two sgRNAs were generated (mRNA-S.2 or mRNA-S.4) (Fig. 9B). Interestingly, a potential mRNA that would use the canonical sequence CS-S2 located 121 nt upstream of the first S gene nucleotide was not detected, suggesting that sequences flanking this internal CS are instrumental in sgRNA synthesis (2).

Nucleotide substitutions in the CS-L led to two gene 3a sgRNA species, named mRNA-3a.1 (wild type) and mRNA-3a.2 (Fig. 10A). These sgRNAs corresponded to the two larger gene 3a sgRNA species found in the corresponding CS-3a mutants. The potential base-pairing score between the mutated TRS-L and the TRS-B, leading to mRNA-3a.3 synthesis, was smaller in the mutants than in the wild-type virus (Fig. 10B). Furthermore, the estimated ΔG indicated that base pairing in this junction site was energetically disfavored, providing a justification for the lack of template switch and production of mRNA-3a.3.

For TGEV gene 7, only one alternative sgRNA was found in the L-C6G mutant and its double mutant, D-C6G (Fig. 10C). The new mRNA-7.2 was generated by a leader-to-body fusion near the ORF 7 3’ end, with high identity with the mutated TRS-L, as shown by RT-PCR amplification and sequencing of leader-to-body junction sites (Fig. 10D). Although this sgRNA could not be translated, because there is no ATG codon, progeny virus was produced, since gene 7 is not essential (22).

Overall, analysis of the alternative sgRNAs produced in virus with nucleotide substitutions in the CS-L indicated that production of novel sgRNAs was associated with the possibility of duplex formation between the TRS-L and cTRS-B with a high base-pairing score.

**DISCUSSION**

In this report, the mechanism of coronavirus transcription has been studied by reverse genetics using full-length genomes for the first time. We have shown that discontinuous coronavirus transcription takes place during the synthesis of a negative sgRNA, and most template switch sites can be predicted by estimating the potential base-pairing score and free energy of the duplex between TRS-L and cTRS-B, using an adapted computer-based program to assess the strength of base pairing between body and leader TRS. In addition, it has been shown that the body canonical core sequence 5’-CUAAAC-3’, although nonessential for the generation of sgRNA, promotes transcription levels by more than 10³-fold over sgRNAs associated with TRs without a canonical CS-B. The modification of a canonical CS led to the synthesis of alternative sgRNAs using noncanonical CSs. It has also been shown that the core sequence and flanking regions in the TGEV genome play a role in discontinuous transcription and modulate sgRNA levels principally by the extent of base pairing. The data obtained were compatible with a three-step mechanism for coronavirus transcription postulating first the formation of a complex between the TRS-L and the 3’ end of the genome and, in a second step, a base-pairing scanning to determine the complementarity between the nascent negative RNA chain and the TRS-L. If duplex formation is favored, in a third step, a template switch is produced, leading to generation of a negative sgRNA.

A key strategy in our study of transcription regulation was the selection of gene 3a, a nonessential gene for TGEV growth both in vitro and in vivo, since modification of this gene did not affect the recovery of mutant viruses (36).

The requirement of complementarity between the CS-L and the cCS-B for the synthesis of a negative sgRNA was reinforced by showing a reduction in the sgRNA synthesis associated with point mutations reducing complementarity between CS-L and cCS-B, and by demonstrating that, in general, sgRNA synthesis was partially or completely restored by the introduction of nucleotide substitutions allowing formation of non-Watson-Crick or Watson-Crick base pairs.

The extent of sgRNA synthesis was related to the free energy of the duplex between CS-L and cCS-B. The potential base-pairing score of sequence domains complementary to genomic RNA with the TRS-L ranged between 15 and 70. According to this score, the local sequence domains could be classified into domains with low (<35), and high (>35) base-pairing potential. In general, sequences with local base pairing of <35 led to no significant production of sgRNA; in contrast, local base pairing higher than 35 led to the synthesis of standard viral mRNAs. These findings validated the in silico analysis method. This method was also found reliable for the prediction of most sgRNAs synthesized by TGEV leader mutants and by other coronaviruses, such as HCoV 229E and BCoV (data not shown). The presence of a canonical CS within the TRS promoted higher sgRNA levels. Nevertheless, the presence of a canonical CS within the TGEV genome did not guarantee the synthesis of an sgRNA. The requirement of an appropriate sequence context was confirmed by showing that a 5’-CUAAAC-3’ sequence present 121 nt downstream of the gene S initiation codon (CS-S2) did not lead to synthesis of the corresponding sgRNA as a consequence of the 5’ and 3’ flanking TRS sequences (2). The lack of sgRNA synthesis could be explained by the relatively low potential base-pairing score and ΔG values (32 and −3.0 kcal/mol, respectively) between the corresponding TRS-L and cTRS-B, values lower than those estimated for the canonical CS-S1 used (35 and −4.3 kcal/mol, respectively).

The presence of TRS-L complementary sequences flanking noncanonical cCS-Bs led to the use of alternative TRS-B sequences. These sgRNAs, such as mRNAs 3a.3 and 7.2, although produced in significant amounts, were generally not translated into truncated proteins, because there was no initiation codon in their sequence. Nevertheless, in a minority of
cases, new sgRNAs could encode essential proteins, and the use of alternative noncanonical CSs could be a safeguard mechanism for virus survival. In fact, this could be the case of gene S alternative sgRNAs that could lead to the production of truncated S proteins, similar to that found in field variants of TGEV, such as the porcine respiratory coronavirus (PRCV) (4, 29). The synthesis of alternative sgRNAs using noncanonical CS as part of the viral life cycle has also been reported.
for other coronavirus (17, 23), in arterivirus (24), and in nidovirus-derived expression systems (7, 11, 36).

Nucleotide substitutions, not allowing base pairing with cCS-B, within the first 3 nt of the CS-L led to the inhibition of sgRNA synthesis and, consequently, to a failure in infectious virus production. In contrast, mutation of CS-L nt 4 to 6 promoted infectious virus recovery, although virus production was reduced by more than 10^4-fold. The higher restriction within the first 3 CS-L nt during template switch fits the discontinuous negative-strand synthesis model (18, 32), because extension of the negative sgRNA body sequences to add the cL sequence could not proceed in the absence of a complementarity between the 3' end of the growing negative strand and the TRS-L. However, alternative explanations, such as the effect of these nucleotide substitutions on key CS-L structural motifs, cannot be discarded.

Our data indicate that template switching during synthesis of the negative strand takes place after termination of the complementarity between the 3' end of the nascent RNA strand (negative polarity) and the TRS-L. In this process, mismatches may be tolerated, providing that several complementary nucleotides between the TRS-L and cTRS-B upstream of the CS-B would be present. This conclusion is based on the sequence of the leader-to-body junctions in a collection of 72 sgRNAs generated after the introduction of point mutations within CS-L and CS-3a. For instance, nucleotide substitutions introduced in the first CS nucleotide (B-C1G) of gene 3a were transferred to the sgRNA, probably because the identity between the TRS-B and the TRS-L was extended 3 nt upstream of the CS (Fig. 11). Similarly, nucleotide substitutions in the first CS-L nucleotide (L-C1U) never appeared in the sgRNA sequence—except in the mRNA-E sequence, since in this case there was no immediately adjacent upstream complementarity (Fig. 11) (data not shown)—reinforcing the concept that template switching takes place at the end of the complementarity between the TRS-L with the cTRS-B.

The data presented are compatible with the working model of coronavirus transcription shown in Fig. 12. This model includes three steps in the selection of the gRNA sequence in which template switching takes place. The first step would involve the formation of 5'–end–3'–end complexes mediated by protein-RNA and protein-protein interactions, by which the TRS-L would be located in close proximity to sequences located at the 3' end of genomic RNA (Fig. 12A). In the mouse hepatitis virus (MHV), it has been shown that members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family, like hnRNP A1 and the polypyrimidine track binding protein (PTB), could interact either with the 5' or 3' ends of the viral genome and between them (5, 13, 35), suggesting that viral ends could be in close proximity during virus transcription and replication. This step is a requirement for the viability of the next step.

In the second base-pairing scanning step, the TRS-L would scan the nascent negative chain, looking for complementary sequence domains (Fig. 12B) leading to a high potential base-pairing score, associated with a favorable ΔG. This scanning has been postulated based on the observed relationship between the presence of a high potential base-pairing score between the TRS-B and the TRS-L and synthesis of sgRNA. This correlation implies that during the synthesis of the negative RNA, the nascent chain has to be screened by the TRS-L, probably partially exposed at the top of a stem-loop. If the complementarity is above a certain threshold, the third template switch step takes place (Fig. 12C) in a proportion of the nascent chains, and a copy of the gRNA leader is made, leading to termination of a negative sgRNA that will be used to generate an sgRNA with the same length. The existence of this step is required to explain the primary structure of a high number (>60) of sgRNAs described in this article. The transcription model postulated in this article reinforces and extends a previously proposed model for coronavirus and arterivirus (26, 27, 30, 32, 38).

The proposed coronavirus transcription mechanism implies a close interaction between TRS-L and each cTRS-B present in the gRNA. Therefore, there should be a restriction in the evolution of these TRSs, because changes in a given TRS-B would affect synthesis of the specific sgRNA. More importantly, changes within the TRS-L should have a pleiotropic effect on the synthesis of all viral mRNAs. Therefore, the degree of freedom in the evolution of these TRSs should be limited, particularly for essential genes. Nucleotide substitutions within a TRS should only be fixed if the sequences flanking a canonical or noncanonical CS-B compensate the decrease in the base pairing between cCS-B and CS-L. Therefore, the probability of a nucleotide substitution within a TRS-B should be lower than that of an average nucleotide substitution within virus RNA genomes (i.e., <1 × 10^-10) (8). Furthermore, the fixation of a nucleotide substitution within the TRS-L, particularly within the first 3 CS-L nt, would require the simultaneous incorporation of complementary mutations within the CS-B of at least the four essential TGEV genes encoding proteins S, E, M, and N. Therefore, this event should have an even lower probability (<1 × 10^-20 in TGEV). These theoretical considerations are supported in TGEV by the lack of...
CS-L evolution throughout more than 50 years of virus replication.

Multiple factors seem to regulate the transcription process (2, 26, 27). These factors would probably imply protein-RNA and protein-protein recognition, including viral and host cell components that will be the subject of future studies.

ACKNOWLEDGMENTS

We thank D. Sawicki, F. Almazán, J. Ortego, and D. Escors for critically reading the manuscript and helpful discussions. We are also grateful to J. C. Oliveros for the PERL script used in the in silico analysis and D. Dorado and V. Hernández for technical assistance.

This work was supported by grants from the Comisión Interminis-
terial 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., A.A., and Z.Z. received postdoctoral fellowships from the Community of Madrid and the European Union (Frame V, Key Action 2, Control of Infectious Disease Projects: QLRT-1999-00002, QLRT-1999-30739, and QLRT-2000-00874).

REFERENCES


