Long-Distance RNA-RNA Interactions in the Coronavirus Genome Form High-Order Structures Promoting Discontinuous RNA Synthesis during Transcription

Pedro A. Mateos-Gomez, Lucia Morales, Sonia Zuñiga, Luis Enjuanes and Isabel Sola

Published Ahead of Print 10 October 2012.

Updated information and services can be found at:
http://jvi.asm.org/content/87/1/177

These include:

REFERENCES
This article cites 35 articles, 16 of which can be accessed free at:
http://jvi.asm.org/content/87/1/177#ref-list-1

CONTENT ALERTS
Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»
Long-Distance RNA-RNA Interactions in the Coronavirus Genome Form High-Order Structures Promoting Discontinuous RNA Synthesis during Transcription

Pedro A. Mateos-Gomez, Lucia Morales, Sonia Zuñiga, Luis Enjuanes, Isabel Sola

Coronavirus (CoV) transcription requires a high-frequency recombination process that links newly synthesized minus-strand subgenomic RNA copies to the leader region, which is present only once, at the 5′ end of the genome. This discontinuous RNA synthesis step is based on the complementarity between the transcription-regulating sequences (TRSs) at the leader region and those preceding each gene in the nascent minus-strand RNA. Furthermore, the template switch requires the physical proximity of RNA genome domains located between 20,000 and 30,000 nucleotides apart. In this report, it is shown that the efficacy of this recombination step is promoted by novel additional long-distance RNA-RNA interactions between RNA motifs located close to the TRSs controlling the expression of each gene and their complementary sequences mapping close to the 5′ end of the genome. These interactions would bring together the motifs involved in the recombination process. This finding indicates that the formation of high-order RNA structures in the CoV genome is necessary to control the expression of at least the viral N gene. The requirement of these long-distance interactions for transcription was shown by the engineering of CoV replicons in which the complementarity between the newly identified sequences was disrupted. Furthermore, disruption of complementarity in mutant viruses led to mutations that restored complementarity, wild-type transcription levels, and viral titers by passage in cell cultures. The relevance of these high-order structures for virus transcription is reinforced by the phylogenetic conservation of the involved RNA motifs in CoVs.
Table 1 Oligonucleotides used for quantitative RT-PCR analysis

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Forward primer name*</th>
<th>Forward primer sequence (5′ → 3′)</th>
<th>Reverse primer name*</th>
<th>Reverse primer sequence (5′ → 3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gRNA</td>
<td>RT-REP-VS</td>
<td>TTCTTGTGACAAAACATACGGTGAA</td>
<td>RT-REP-RS</td>
<td>CTAAGCAGCTGGTGTAGACATCTTT</td>
</tr>
<tr>
<td>N sgRNA</td>
<td>Ldrt-VS</td>
<td>CGTGCGCTATATCTCTTTCATTTACTAAG</td>
<td>N(82)-RS</td>
<td>TCCTCCGAGCAGGGTTAAG</td>
</tr>
<tr>
<td>sgRNA 7</td>
<td>Ldrt-VS</td>
<td>CGTGCGCTATATCTCTTTCATTTACTAAG</td>
<td>7(38)-RS</td>
<td>AAAATGTAATATACAGCATGGAGG</td>
</tr>
</tbody>
</table>

*The hybridization sites of the oligonucleotides within the TGEV genome are as follows: for RT-REP-VS, nt 4,829 to 4,853; for RT-REP-RS, nt 4,884 to 4,909; for Ldrt-VS, nt 25 to 56; for N(82)-RS, nt 26,986 to 27,004; and for 7(38)-RS, nt 28,086 to 28,114.

Materials and Methods

Cells and viruses. Baby hamster kidney (BHK) cells stably transformed with the porcine aminopeptidase N gene (pPAN) (22) and with the Sindbis virus replicon pSINrep1 (23) expressing TGEV N protein (BHK-N) were grown in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 5% fetal calf serum (FCS) and with G418 (1.5 mg/ml) and puromycin (5 μg/ml) as selection agents for pPAN and pSINrep1, respectively. Recombinant TGEVs were grown in swine testis (ST) cells (24). ST cells were grown in DMEM supplemented with 10% fetal bovine serum (FBS). TGEVs were obtained from an infective clone with the sequence of the isolated TGEV PUR46-MAD (25) and the spike protein of strain PTV (Purdue type strain) (26) by use of a reverse-genetic system. Virus titration was performed on ST cell monolayers as described previously (27).

Reverse-genetic system. cDNAs encoding TGEV replicons and replicons were constructed using pBAC-TGEV (25) (Genbank accession number A1271965), which includes the whole TGEV genome sequence, and the replicon pBAC-REP1 (28). The pBAC vectors were transfected into BHK-N cells (described above). Cellular RNA polymerase II transcribes virus or replicon genomes from pBAC vectors, and viral RNA is then translated, initiating the viral cycle. To rescue infectious viruses from cDNA clones, transfected BHK-N cells were treated with trypsin at 6 h posttransfection (hpt) and plated over confluent ST cell monolayers.

Transfection. BHK-N cells were grown to 95% confluence on 35-mm-diameter plates and then transfected with 4 μg of cDNA encoding a TGEV replicon or infectious virus, representing 100 molecules per cell, on average, by using 12 μg of Lipofectamine 2000 (Invitrogen) according to the manufacturer’s specifications. The conditions in transfection experiments were strictly controlled, as follows: (i) the same number of cells per well was seeded (5 × 10^5 cells/well), (ii) the same amount of cDNA was always transfected (100 molecules per cell), and (iii) cDNA was purified using a large-contract kit (Qiagen) including an exonuclease treatment to remove bacterial DNA contamination and damaged plasmids, thus providing an ultrapure DNA plasmid for transfection.

DNase treatment of RNAs from transfection experiments. To remove transfected cDNA-containing replicon genomes from samples before quantitative reverse transcription-PCR (qRT-PCR) analysis, 7 μg of each RNA in 100 μl was treated with 20 U of DNase I (Roche) for 30 min at 37°C. DNA-free RNAs were repurified using an RNeasy miniprep kit (Qiagen).

RNA analysis by quantitative RT-PCR. Total intracellular RNA was extracted from transfected BHK-N cells at 24 hpt or from ST cells infected with recombinant TGEVs at 16 h postinfection. RNAs were purified with an RNeasy miniprep kit (Qiagen) according to the manufacturer’s specifications. To remove the transfected replica DNA prior to qRT-PCR analysis, each RNA sample was treated with DNase I (Roche). cDNAs were synthesized at 37°C for 2 h with MultiScribe reverse transcriptase (high-capacity cDNA reverse transcription kit; Applied Biosystems). Real-time qRT-PCR was used for quantitative analysis of genomic and subgenomic RNAs from infectious TGEV and TGEV-derived replicons. Oligonucleotides used for quantitative PCRs (Table 1) were designed with Primer Express software. SYBR green PCR master mix (Applied Biosystems) was used in the PCR step, according to the manufacturer’s specifications. TaqMan assays were used to quantify the β-glucuronidase mRNA.
### Table 2: Oligonucleotides used for site-directed mutagenesis

<table>
<thead>
<tr>
<th>Mutant(s)</th>
<th>Oligonucleotide</th>
<th>Sequence (5' → 3')a</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD-ΔA-B’12, AD-ΔA-B’9, AD-ΔA-B’4, AD-ΔA-B’3, AD-ΔA-C’, AD-ΔA, AD-ΔB, AD-ΔC</td>
<td>Rep Mut 3 VS</td>
<td>TTTCTAGGTTGGAGACTCTGGCTCATATAATTAGTAC</td>
</tr>
<tr>
<td>AD-ΔA</td>
<td>ΔA-RS</td>
<td>CTCAGGCTCGAATCTAAAGGGCGAGCTTTATGCGGTATACCGTATGCG</td>
</tr>
<tr>
<td>AD-ΔB</td>
<td>ΔB-RS</td>
<td>CTCAGGCTCGAATCTAAAGGGCGAGCTTTATGCGGTATACCGTATGCG</td>
</tr>
<tr>
<td>AD-ΔC</td>
<td>ΔC-RS</td>
<td>GCCACTATATCTGAGCTTATATGAGGCAACATATACGTCATTGG</td>
</tr>
<tr>
<td>AD-ΔA-C’</td>
<td>3’AD-ΔA-C’-RS</td>
<td>CTTAGGCTGCAATCTAAAGGGCGAGCTTTATGCGGTATACCGTATGCG</td>
</tr>
<tr>
<td>AD-ΔA-B’12, AD-ΔA-B’9</td>
<td>3’AD-ΔA-B’12-RS</td>
<td>CTTAGGCTGCAATCTAAAGGGCGAGCTTTATGCGGTATACCGTATGCG</td>
</tr>
<tr>
<td>AD-ΔA-B’9</td>
<td>3’AD-ΔA-B’9-RS</td>
<td>CTTAGGCTGCAATCTAAAGGGCGAGCTTTATGCGGTATACCGTATGCG</td>
</tr>
<tr>
<td>AD-ΔA-B’4</td>
<td>3’AD-ΔA-B’4-RS</td>
<td>CTTAGGCTGCAATCTAAAGGGCGAGCTTTATGCGGTATACCGTATGCG</td>
</tr>
<tr>
<td>AD-ΔA-B’3</td>
<td>3’AD-ΔA-B’3-RS</td>
<td>CTTAGGCTGCAATCTAAAGGGCGAGCTTTATGCGGTATACCGTATGCG</td>
</tr>
<tr>
<td>cB-218*/ΔB, cB-218*/B</td>
<td>218-VS</td>
<td>CGTGCGACTGGTGGTTCTGAGGCTCGATCTCTCTC</td>
</tr>
<tr>
<td>cB-218*/ΔB, cB-218*/B</td>
<td>218-RS</td>
<td>ATAGGAGGCCAACCTCATCGACACCG</td>
</tr>
<tr>
<td>cB-477*/ΔB, cB-477*/B, cB-477*/ΔB, cB-477*/B-14</td>
<td>477-VS</td>
<td>CTGTCGACCTGGTCCGAGTATGCATTAAAATAGCCAG</td>
</tr>
<tr>
<td>cB-477*/ΔB, cB-477*/B, cB-477*/ΔB, cB-477*/B-14</td>
<td>477-RS</td>
<td>GCAATGATCTGGCAGGCTCAAAAGCCGAACATTACATATCTGGACACTTGGTATTCCGAGTATG</td>
</tr>
<tr>
<td>rTGEV-cB*</td>
<td>5’-482 1mut-VS</td>
<td>CTCGTGCGACTGGTGGTTCTGAGGCTCGATCTCTCTC</td>
</tr>
<tr>
<td>rTGEV-B*</td>
<td>3’-AvrII-VS</td>
<td>TGACGTTCCTCGAGGCTCAAGTCGTCATAG</td>
</tr>
<tr>
<td>AD-26418-1mut-RS</td>
<td>TACAGGCACTCGAATCTAAAGGGCGAGCTTTATGCGGTATACCGTATGCG</td>
<td></td>
</tr>
</tbody>
</table>

a The mutated nucleotides are shown in bold. Restriction sites are underlined (AvrII, CCTAGG; EcoRI, GAATTC; and MfeI, CAATTG).
shaker for at least 5 h each time at 4°C. For RNA binding, 60 μl of solid matrix per RNA binding assay was first washed twice with solution BW (5 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 M NaCl) and then incubated with the biotinylated RNAs (8 μg) in BW solution for 30 min at room temperature. Immobilized RNAs were washed twice with H-BW. Five hundred micrograms of protein extract and different amounts of nonspecific competitor (0.5 or 1.25 μg tRNAs/μg protein) tRNA (baker’s yeast; Sigma) were incubated overnight in an orbital shaker at 4°C. After three washes with H-BW solution, the proteins were eluted with NuPAGE LDS sample buffer (Invitrogen) supplemented with 100 mM DTT for 10 min at room temperature. Proteins were resolved in NuPAGE 4 to 12% Bis-Tris gels at 180 V. With H-BW solution, the proteins were eluted with NuPAGE MOPS-SDS running buffer (Invitrogen) (2.5 mM morpholinepropanesulfonic acid [MOPS], 2.5 mM Tris base, 0.005% SDS, 0.05 mM EDTA, pH 7.7). Finally, gels were stained with Coomassie Simply Blue Safe stain (Invitrogen). Images were taken with Image Lab V3.0 (Bio-Rad).

**In silico analysis.** Potential base pairing score calculations were performed as previously described (19). ΔG calculations were performed using the two-state hybridization server (http://mfold.rna.albany.edu/?q=DINAMelt/Two-state-melting) (31). Secondary structure predictions were performed using the Mfold server for nucleic acid folding and hybridization prediction (http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form) (32). The analysis of sequences was performed using DNAStar Lasergene software 7.0. Comparison of RNA sequences was performed using ClustalW2/EBI (http://www.ebi.ac.uk/Tools/clustalw2/index.html).

To identify sequences complementary to the B RNA motif within the TGEV genome, DOT- PLOT MAKER v1.0 software was used (http://bioinfogg.cnb.csic.es/tools/dotplot/index).

**RESULTS**

**Structural requirements of the active domain for N gene transcription.** We previously described that the active domain, including two hairpins, is essential for N gene transcription in TGEV. Since only the most 3’ hairpin is necessary for N gene transcription (21), the term “active domain” refers to this hairpin, comprised of nt 26,356 to 26,438 of the TGEV genome (33) (Fig. 1B). To study the sequence and structural requirements of the active domain, its Mfold-predicted structure was divided into the apical (A), intermediate (B), and basal (C) regions (Fig. 2). Three mutant TGEV genome-derived replicons were constructed in order to analyze the contribution of each region to the active domain function. Regions A, B, and C were deleted in mutants AD-ΔA, AD-ΔB, and AD-ΔC, respectively (Fig. 2). BHK cells constitutively expressing the TGEV N protein (BHK-N cells) were transfected with the corresponding cDNAs, and N sgmRNA levels were determined twice by qRT-PCR. Error bars represent standard deviations.

**Relevance of active domain region C secondary structure for N gene transcription.** According to Mfold predictions, the active domain region C adopted a secondary structure consisting of a long and stable stem. In order to determine whether this secondary structure was required for active domain function, the sequence of region C was replaced by a nonrelated primary sequence adopting a similar secondary structure according to Mfold prediction (AD-ΔA-C’ replicon) (Fig. 3). The AD-ΔA-C’ mutant showed N sgmRNA levels similar to those of the replicon including the complete active domain, previously named dE-173-20 (21), indicating that region A is not necessary for active domain function (Fig. 2). In contrast, N sgmRNA levels in the AD-ΔB and -ΔC mutants were similar to those in the replicon TRS-N-ΔE, lacking the whole active domain (Fig. 2), indicating that regions B and C are required to maintain N gene transcription levels.

**Requirement of active domain region C secondary structure for N gene transcription.** According to Mfold predictions, the active domain region C adopted a secondary structure consisting of a long and stable stem. In order to determine whether this secondary structure was required for active domain function, the sequence of region C was replaced by a nonrelated primary sequence adopting a similar secondary structure according to Mfold prediction (AD-ΔA-C’ replicon) (Fig. 3). The AD-ΔA-C’ mutant showed N sgmRNA levels similar to those of the replicon including the complete active domain, previously named dE-173-20 (21), indicating that region A is not necessary for active domain function (Fig. 2). In contrast, N sgmRNA levels in the AD-ΔB and -ΔC mutants were similar to those in the replicon TRS-N-ΔE, lacking the whole active domain (Fig. 2), indicating that regions B and C are required to maintain N gene transcription levels. These results indicated that the active domain function required a region C with stable secondary structure but not a specific primary sequence.

**Relevance of active domain region B primary sequence for optimal N gene transcription.** The requirement for region B to maintain N gene transcription levels was shown above (Fig. 2). To analyze whether the region B requirement was dependent on its secondary structure or primary sequence, four new mutant replicons, containing 12-, 9-, 4-, and 3-nucleotide substitutions (AD-
ΔA-B’12, -B’9, -B’4, and -B’3, respectively), were constructed (Fig. 4). All replicons showed N sgRNA levels similar to those of the reference replicon TRS-N-ΔdE, with no active domain, and significantly lower than those of the positive control, AD-ΔA, which included the wild-type region B sequence (Fig. 4). The significant reductions in N sgRNA levels of the AD-ΔA-B’4 and AD-ΔA-B’3 mutants, which included mutations in the primary sequence of region B without affecting its secondary structure prediction, suggested that the primary sequence of the active domain region B, named the B motif (nt 26,408 to 26,421), is essential for maintaining N gene transcription levels. This requirement could be due either to the interaction of this RNA motif with viral and host proteins or to RNA-RNA interactions with other RNA sequences in the viral genome.

Involvement of the B motif included in the active domain in RNA-RNA interactions. The possibility of differential protein binding to the B motif with a native or mutated sequence was analyzed by RNA affinity chromatography. No proteins specifically interacting with the B motif with a native or mutated sequence were identified by mass spectrometry (Fig. 5). The alternative possibility of an interaction of the B motif with a cRNA sequence (cB) present within the viral genome was then analyzed using the program DOT-PLT MAKER v1.0. The criterion to select potential RNA-RNA interactions was an associated stability (ΔG) similar to or higher than that of the described interaction between proximal and distal elements (−8.6 kcal/mol) required to maintain N gene transcription (20). Two potential B motif interaction sites, located at the genome’s 5’ end (nt 218 to 229 and 477 to 486), were identified in the TGEV genome. To analyze the functional relevance of the postulated RNA-RNA interactions between the B and cB motifs, two mutant replicons were constructed (Fig. 6A and B). In the cB-218*/B mutant, the sequence of interest, located at nt 218 to 229 (5’-UCGUAAGUCGC-3’), was modified at 8 positions (underlined) to disrupt the complementarity with the active domain B motif: 5’-AACUAUCUCUGU-3’. Similarly, in replicon cB-477*/B, the original sequence, located at nt 477 to 486 (5’-GAUCGUAGGG-3’), was mutated to 5’-GACAG GAAAG-3’ (Fig. 6A and B). Since the cB-477 RNA motif was included within the viral ORF1a coding sequence, the mutations were designed to preserve the amino acid sequence (Fig. 6B). Furthermore, since mutations at the 5’ end of the genome could have a general effect on RNA synthesis, two reference replicons (218*/ΔB and cB-477*/ΔB) were constructed to lack the active domain and include the same mutations as those described above, at positions 218 to 229 and 477 to 486, respectively (Fig. 6A and B). After transfection into BHK-N cells, N sgRNA levels transcribed by replicon cB-218*/B, with a mutated sequence from nt 218 to 229, were similar to those of the positive control (Fig. 6C) and significantly lower than those of the reference replicons cB-218*/ΔB and wt/ΔB (Fig. 6C). Accordingly, mutations in the cB-218*/B+ replicon to restore the complementarity between the mutated cB-218* sequence and the B motif (the original sequence of the B motif [5’-GGCUUCUACGA-3’] was changed to 5’-GGCUCUUACGA-3’) did not maintain the original N sgRNA levels, indicating that the cB sequence motif in nt 218 to 229 is not required for N gene transcription.

In contrast, the mutations introduced into replicon cB-477*/B (nt 477 to 486) reduced N sgRNA levels (Fig. 6C). Therefore, the cB-477 RNA motif at the 5’ end of the viral genome is involved in N gene transcriptional regulation. To confirm the RNA-RNA interaction between the cB-477 and B motifs, a new replicon (cB-477*/ΔB) was constructed by introducing point mutations in the B motif (5’-UCUU ACCGUU-3’) was changed to 5’-UUUUCUGU-3’) in order to restore the complementarity with the mutated cB-477* motif. In addition, the mutant replicon cB-477*/B’-14 was constructed, with the double objective of restoring and ex-
tending the complementarity between the cB-477 and B RNA motifs from nt 10 to 14 (the B motif sequence \([\text{5'-UGGCUCUUAC-3'}]\) was changed to \([\text{5'-UGAUCUUUCCUGUC-3'}]\)]. Gene transcription levels in both the cB-477*/B* and cB-477*/B*-14 mutants were even higher than those in the positive control (Fig. 6C), indicating that the restoration of the complementarity between the cB-477 (nt 477 to 486) and B (nt 26,412 to 26,421) RNA motifs recovered N gene transcription levels. Furthermore, these results indicated that the complementarity, not the primary sequences, of the cB-477 and B motifs was required for transcriptional regulation of N sgmRNA.

Relevance of RNA-RNA interaction between the cB motif at the 5' end of the genome and the B motif in the active domain in the viral infective cycle. To confirm that the RNA-RNA interaction between the cB-477 and B RNA motifs was relevant for the virus life cycle, two recombinant viruses, rTGEV-cB* and rTGEV-B*, were generated (Fig. 7A) and included one point mutation each, at positions 482 and 26,418, respectively, that decreased the complementarity between the cB and B RNA motifs without introducing amino acid changes. Viruses with the wild-type and mutated sequences were obtained after full-length cDNA transfection. A region extending 1,000 nt over the mutated positions of the recombinant viruses was sequenced after one virus passage in ST cells. Only the originally introduced mutations were observed in mutant viruses (Fig. 7B). The N sgmRNA levels and viral titers for both rTGEV-cB* and rTGEV-B* were significantly reduced (4-fold for the N sgmRNA level and 100-fold for the viral titer) in comparison to those of the wild type (Fig. 7C), whereas viral replication was not reduced (Fig. 7D). After 11 serial passages in cell culture, the sequences of interest were sequenced. The sequence at position 482 in rTGEV-cB* had reverted to the wild-type sequence (Fig. 7B), and more importantly, viral titers and N sgmRNA amounts were also restored to the wild-type levels (Fig. 7C). In the mutant virus rTGEV-B*, a double sequence (C/U) at position 480 was observed, corresponding to the wild-type nt (C) and to the new sequence (U), restoring the complementarity with a G→A mutation at position 26,418 of the rTGEV-B* genome. Although at passage 11 both sequences coexisted in the virus population, the C→U substitution was dominant by passage 18 (Fig. 7B). Interestingly, this new mutation at position 480 in rTGEV-B* restored the wild-type viral titers and N sgmRNA levels (Fig. 7C). These results showed the functional relevance of the RNA-RNA interaction between the B and cB RNA motifs in the TGEV infectious cycle.
Conservation of genomic RNA motifs involved in TGEV N gene transcription. The sequences of RNA motifs B and cB involved in the novel RNA-RNA interaction described above are conserved among the Alphacoronavirus I species, comprising TGEV, canine coronavirus (CCoV), feline coronavirus (FCoV) type I, and FCoV type II or feline infectious peritonitis virus (FIPV) (Fig. 8A). Additionally, the secondary structure of the 5'-end genomic region, including the cB motif at the apical loop, is conserved according to Mfold predictions (Fig. 8B), suggesting a functional conservation in these CoV species. Interestingly, in the cB RNA motifs of CCoV (nt 477 to 486) and FCoV types I and II (nt 474 to 483), a single nucleotide change was found in comparison with the TGEV sequence, replacing a G-U pair with an A-U pair (Fig. 8A). In FCoV, another nucleotide change was observed at position 26,271, replacing U with G and thus disrupting the G-U pair present in TGEV and CCoV (Fig. 8A). However, none of these nt changes significantly affected the stability of the RNA-RNA interaction (Fig. 8A). The sequence of the nsp1 coding region containing the cB motif was mainly conserved at the nucleotide and amino acid levels in Alphacoronavirus I species. A substitution (G→A) in the first nucleotide of the cB motif in the canine and feline CoVs with respect to TGEV led to an amino acid

FIG 6 Involvement of the active domain in an RNA-RNA interaction with the 5' end of the TGEV genome. (A) Schemes of the genetic structures of the mutant replicons. The sequence positions in the genome are indicated on the top line. Letters above the boxes indicate the proximal element (pE), distal element (dE), N gene (N), and B motif in the active domain (nt 26,412 to 26,421). The asterisks indicate that the sequence was mutated as described in the text (nt 218 to 229 and 477 to 486). (B) Depiction of the possible RNA-RNA interactions between the B motif in the active domain (nt 26,412 to 26,421) and the sequences at nt 218 to 229 or nt 477 to 486 in the different mutants. (C) Analysis by qRT-PCR of N sgRNA levels normalized to gRNA in relation to reference levels in replicons wt/ΔB, cB-218*/ΔB, and cB-477*/ΔB, which were set at 1. wt/B, positive control. Data are averages for four independent transfection experiments analyzed twice by qRT-PCR. Error bars represent standard deviations. (D) Analysis by qRT-PCR of sgRNA 7 levels normalized to gRNA in relation to reference levels in replicons wt/ΔB, cB-218*/ΔB, and cB-477*/ΔB, which were set at 1. wt/B, positive control. Data are averages for four independent transfection experiments analyzed twice by qRT-PCR. Error bars represent standard deviations.
change (Asp → Asn) in the nsp1 protein. In contrast, the coding sequence of the M protein containing the B motif was conserved at both the nucleotide and amino acid levels in Alphacoronavirus I species (Fig. 8A). Similar structures at similar genome positions have not been found either in other species of the Alphacoronavirus genus (human coronaviruses [HCoVs] 229E and NL63) or in Beta- and Gammacoronavirus species.

**DISCUSSION**

The sequence and secondary structure requirements of the active domain, a TGEV genome region essential for N gene transcription regulation, have been shown in this report. The mechanism of action of this motif implies a novel long-distance interaction between RNA motifs mapping almost 26,000 nt apart in the coronavirus genome. This interaction between the B motif in the active domain and the complementary cB motif next to the 5′ end of the genome was essential to maintain N gene transcription levels in both TGEV-derived replicons and infectious viruses (Fig. 6 and 7). In contrast, transcription of sgmRNA 7 in mutant replicons (Fig. 6) and viruses (data not shown) was not significantly affected by the novel RNA-RNA interaction, suggesting that the interaction between B and cB RNA motifs is required mainly for N gene transcription. These results led us to propose a new model for N gene transcription regulation. According to this model, the RNA-RNA interaction between the proximal and distal elements, located within the
transcription-regulating sequences of the N gene, relocates the active domain with a proper conformation to a site immediately preceding the N gene CS (Fig. 9). The B motif within the active domain interacts with the complementary cB motif sequence at the 5’ end of the genome (nt 477 to 486), bringing in close proximity the TRS-L and TRS-N sequences, which guide transcription by a recombination process (Fig. 9). The physical proximity of CS-L and CS-N sequences favors the template switch necessary for discontinuous transcription, increasing the frequency of this long-distance recombination. Our data represent the first evidence of a mechanism that brings into physical proximity the leader sequence at the 5’ end of the genome and the TRS preceding a gene close to the 3’ end, which map 26,000 nt apart in the coronavirus genome. We postulate that this network of RNA-RNA interactions drives the genome to adopt a high-order structure that facilitates discontinuous transcription in nidoviruses. The previously proposed working model of coronavirus transcription (17) includes three steps. The first step is the formation of precomplexes in the genomic RNA that locate the TRS-L in close proximity to the TRS-Bs at the 3’ end of the genome, promoting a stop of synthesis of the nascent minus-strand RNA and a high-frequency recombination. The novel finding reported in this work fits very well into the first step of the model, since the interaction described for the B and cB motifs locates the TRS-L in close proximity to the TRS-N. The second step is the formation of a complex among the TRS-L, the TRS-B, and the nascent cTRS-B. The complementarity between the TRS-L and the cTRS-B is evaluated, and if this complementarity is above a minimum threshold, a certain number of newly synthesized negative RNAs undergo a template switch to the TRS-L. The third step resumes the synthesis of RNA to obtain a complete negative-strand sgRNA.

Interestingly, in the nonrelated tombusvirus tomato bushy stunt virus (TBSV), RNA-RNA interactions relocate, just before each gene, a secondary structure that acts as a physical barrier mediating premature termination during the synthesis of the minus-strand subgenomic RNAs (7, 34). In TGEV, the secondary structure of the active domain and the high-order structure formed by the RNA-RNA interactions could also promote the slowdown and stop of the transcription complex at the CS-N. In addition to premature termination, CoV discontinuous transcription also requires a template switch, a recombination process guided by the homology between the TRS-L and the TRS-Bs. To achieve a high-frequency recombination during transcription, physical proximity between the leader sequence and the TRS-Bs would be extremely helpful. This proximity is probably promoted by the described long-distance RNA-RNA interaction between the active domain and the genome’s 5’ end. This paper shows evidence for the first time of a molecular mechanism that could bring into physical proximity sequences present in the exceptionally long coronavirus RNA genomes.

FIG 9 Model for N gene transcription regulation mediated by the formation of genome high-order structures. The black line represents the TGEV genome. The thicker line indicates the leader sequence in the 5’ end of the genome. The gray line with arrowheads represents the minus-polarity nascent RNA. The dotted gray line represents the copy of the leader that is added to the nascent RNA after the template switch. AD, active domain secondary structure prediction. The RNA-RNA interactions between the proximal and distal elements (pE, proximal element; dE, distal element) and between the B motif in the active domain (nt 26,412 to 26,421; B-M) and the cB motif in the 5’ end of the genome (nt 477 to 486; cB-M) are represented in the plus-strand genomic RNA. CS-L, conserved core sequence of the leader; CS-N, conserved core sequence of the N gene; cCS-N, copy of the CS-N. The arrow indicates the template switch to the leader sequence during discontinuous transcription.
that are fused during discontinuous transcription. Although RNA-RNA interactions leading to genomic high-order structures are relevant for essential viral processes in other RNA virus families (35–37), this work provides the first evidence for a long-distance (26,000 nt) RNA-RNA interaction regulating discontinuous transcription in nidoviruses.

ACKNOWLEDGMENTS

This work was supported by grants from the Ministry of Science and Innovation of Spain (BIO2010-16707 and PET2008-0310) and the European Community’s Seventh Framework Programme (FP7/2007-2013), under the projects EMPERIE (EC grant agreement 223498) and PoRRSCon (EC grant agreement 245141). P.A.M.-G. received a fellowship from the Ministry of Science and Innovation of Spain (BES-2008-001932).

We gratefully acknowledge C. M. Sanchez, M. Gonzalez, and S. Ros for technical assistance.

REFERENCES


