JOURNAL OF VIROLOGY, Sept. 2011, p. 8968–8980 Vol. 85, No. 17
0022-538X/11/$12.00 doi:10.1128/JVI.00869-11

Gene N Proximal and Distal RNA Motifs Regulate Coronavirus Nucleocapsid mRNA Transcription▽

Pedro A. Mateos-Gómez, Sonia Zuñiga, Lorena Palacio, Luis Enjuanes,* and Isabel Sola

Department of Molecular and Cell Biology, Centro Nacional de Biotecnología, Darwin 3, Campus de la Universidad Autonoma de Madrid, 28049 Madrid, Spain

Received 29 April 2011/Accepted 20 June 2011

Coronavirus subgenomic mRNA (sgmRNA) transcription requires a discontinuous RNA synthesis mechanism driven by the transcription-regulating sequences (TRSs), located at the 3′ end of the genomic leader (TRS-L) and also preceding each gene (TRS-B). In transmissible gastroenteritis virus (TGEV), the free energy of TRS-L and cTRS-B (complement of TRS-B) duplex formation is one of the factors regulating the transcription of sgmRNAs. In addition, N gene sgmRNA transcription is controlled by a transcription-regulating motif, including a long-distance RNA-RNA interaction between complementary proximal and distal elements. The extension of complementarity between these two sequences increased N gene transcription. An active domain, a novel essential component of the transcription-regulating motif, has been identified. The active domain primary sequence was necessary for its activity. Relocation of the active domain upstream of the N gene TRS core sequence in the absence of the proximal and distal elements also enhanced sgmRNA N transcription. According to the proposed working model for N gene transcriptional activation, the long-distance RNA-RNA interaction relocates the distant active domain in close proximity with the N gene TRS, which probably increases the frequency of template switching during the synthesis of negative RNA. The transcription-regulating motif has been optimized to a minimal sequence showing a 4-fold activity increase in relation to the native RNA motif. Full-length TGEV infectious viruses were generated with the optimized transcription-regulating motif, which enhanced by 5-fold the transcription of the 3a gene and can be used in expression vectors based in coronavirus genomes.

Transmissible gastroenteritis virus (TGEV) is a member of the *Coronaviridae* family included in the *Nidovirales* order, which also comprises the *Arteriviridae* and *Roniviridae* families (3, 6) (see http://talk.ictvonline.org/media/g/vertebrate-2008/default.aspx). TGEV has a positive-sense single-stranded RNA genome of 28.5 kb. The 5′ two-thirds of the TGEV genome comprises open reading frames 1a and 1ab, which encode the replicase proteins. The 3′ third of the genome encodes structural and group accessory proteins in the order 5′-S-3a-3b-E-M-N-7-3′. The generation of subgenomic mRNAs (sgmRNAs) is a common strategy in many plus-stranded RNA viruses to regulate the expression of viral proteins encoded at the 3′ end of the genome (5, 6, 17, 22). To express structural and accessory genes, coronaviruses (CoVs) form a nested set of coterminal mRNAs that share the same 5′ and 3′ ends with the genome. CoV transcription includes a discontinuous RNA synthesis step during the production of the minus-stranded RNAs. These RNAs are used as templates to produce the plus-stranded sgmRNAs (22, 28, 35). This discontinuous process is guided by the transcription-regulating sequences (TRSs), located at the 3′ end of the genomic leader sequence (TRS-L) and also preceding each gene (body TRSs [TRS-Bs]). TRSs include a conserved core sequence (CS) which, in the case of TGEV, is 5′-CUAAAC-3′, and variable 5′- and 3′-flanking sequences that are essential for efficient sgmRNA production (28, 35).

According to the CoV transcription model with the widest experimental support (24, 25, 27, 28, 35), TRS-Bs lead to the formation of dynamic complexes with the TRS-L associated with viral and cellular proteins. These complexes most likely promote the stop of elongation of the minus-stranded RNA at each TRS-B and the template switch to the TRS-L. The base-pairing between the cTRS-B (in the nascent minus strand) and TRS-L (in the plus-strand genome) is a highly significant factor among those contributing to the regulation of CoV transcription and the production of sgmRNAs of different sizes (28, 35). CoV transcription is controlled by other factors, such as the proximity of the gene to the 3′ end of the genome, which increases the synthesis of the 3′-proximal sgmRNAs (22). It has also been suggested that protein-protein or RNA-protein interactions are involved in CoV transcription regulation (5, 27). Another factor that regulates CoV transcription is based on a long-distance RNA-RNA interaction (21). A good correlation was established between the amount of each sgmRNA produced during TGEV infection and the base-pairing score between TRS-L and the corresponding cTRS-B, with the clear exception of sgmRNA N. In this case, the base-pairing score between TRS-L and cTRS-N was one of the lowest, whereas sgmRNA N was the most abundant mRNA (21), indicating that additional factors must be involved in the regulation of N gene transcription. Interestingly, two 9-nucleotide (nt) motifs with complementary sequences were identified 5′ upstream of the N gene. The proximal element (pE; 5′-AUUACAUAU-3′) was located 7 nt upstream of CS-N, whereas the distal element (dE; 5′-AUAUGUAAU-3′), unique in the TGEV genome,

* Corresponding author. Mailing address: Department of Molecular and Cell Biology, Centro Nacional de Biotecnología, Darwin 3, Campus de la Universidad Autonoma de Madrid, 28049 Madrid, Spain. Phone: 34-91-585 4555. Fax: 34-91-585 4506. E-mail: L.Enjuanes@cnb.csic.es.

▽ Published ahead of print on 29 June 2011.

was located 449 nt upstream of CS-N. The requirement of the complementarity between these proximal and distal elements has been documented (21). This was the first time that a long-distance RNA-RNA interaction, such as those regulating sgmRNA transcription in tombusviruses by a premature termination mechanism (13), was associated with transcriptional regulation in the *Nidovirales* order. The proximal and distal elements and their relative positions in the viral genome are conserved within the α1 genus of CoVs. The most 3′ gene in the CoV α1 genus genome is gene 7. In contrast, in other CoVs and arteriviruses, the N gene is located at the 3′ end of the genome; this position would favor its transcription (21). The described long-distance RNA-RNA interaction between the proximal and distal elements most probably creates a structural motif that stops the transcription complex progress at the TRS-N, promoting the nascent minus-stranded RNA template switch to the TRS-L, leading to an increase of sgmRNA N levels (21). This unique RNA motif may contribute to increase the amount of N protein produced. This protein is an abundant structural CoV protein and is also required for CoV transcription (5, 34).

In this report, it is shown that the amount of sgmRNA N is directly proportional to the extent of complementarity between proximal and distal RNA sequences located 5′ upstream of the N gene TRS and inversely correlated with the distance between proximal and distal elements. A novel essential element of the sgmRNA N transcription-regulating motif has been identified, the active domain, consisting of a 173-nt region at the 5′ flank of the distal element. The active domain sequence and genome location are mainly conserved in the CoV α1 genus. The relocation of the active domain upstream of CS-N, in the absence of proximal and distal elements, also enhanced sgmRNA N transcription. The primary sequence of the active domain was essential for N gene transcription enhancement, suggesting that this domain might be involved in sequence-dependent RNA-RNA or RNA-protein interactions. An optimized transcription-regulating motif was engineered with a reduced length (250 nt) and 4-fold-higher activity than that of the native transcription-regulating motif. This transcription-regulating motif also enhanced the transcription of an alternative gene that, in addition, could be located at a different genome position. Infectious TGEVs were engineered with the optimized transcription-regulating motif enhancing the transcription of the 3a gene up to 5-fold. These RNA motifs might be useful for enhancing transcription of heterologous genes in virus-derived vectors.

MATERIALS AND METHODS

Cells and viruses. Baby hamster kidney (BHK) cells stably transformed with the porcine amino peptidase N gene (pAPN) (4) and with the Sindbis virus replicon pSINrep1 (7) expressing TGEV N protein (BHK-N) were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 5% fetal calf serum (FCS), G418 (1.5 mg/ml), and puromycin (5 μg/ml) as selection agents for pAPN and pSINrep1, respectively. Recombinant TGEVs were grown in swine testis (ST) cells (19). ST cells were grown in DMEM supplemented with 10% FCS. Virus titration was performed on ST cell monolayers as previously described (11).

Transfection and recovery of infectious TGEVs from cDNA clones. BHK-N cells were grown to 95% confluence on 35-mm-diameter plates and transfected with 4 μg of each cDNA encoding TGEV replicon or infectious viruses, representing on average 100 molecules per cell, by using 12 μg of Lipofectamine 2000 (Invitrogen) according to the manufacturer's specifications. The conditions in transfection experiments were strictly controlled: (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); (iii) cDNA was purified using a large-construct kit (Qiagen), including an exonuclease treatment to remove bacterial DNA contamination and damaged plasmids, thus providing ultrapure DNA plasmid for transfection. For recovery of infectious TGEVs from cDNA clones, transfected cells were plated over a confluent ST cell monolayer (28). Recombinant viruses were harvested from cell supernatant and cloned by four consecutive plaque purification steps.

Plasmid constructs. cDNAs of TGEV-derived replicons and infectious viruses (1, 2) were generated by PCR-directed mutagenesis. To generate dE-153-158, dE-133-158, dE-113-158, dE-45-158, dE-20-158, dE-6-158, dE-173-95, dE-173-45, dE-173-20, dE-173-6, and dE-103-20 mutant replicons, the plasmid pBAC-TGEV (2), containing the TGEV genome (GenBank accession no. AJ271965), was used as template with the specific oligonucleotides shown in Table 1. To construct 2+dE+2, 4+dE+4, and 6+dE+6 mutant replicons, two overlapping PCR fragments were obtained by using as template the pBAC-TGEV and specific oligonucleotides (Table 1). All these mutated fragments generated by PCR contained AvrII sites at both ends to be introduced into the same site of TRS-N-ΔdE mutant cDNA (21). The mutant replicons pE-75, pE-45, and pE-20 were generated by two overlapping PCR fragments, using as templates the pBAC-TGEV and the dE-173-20 plasmids, respectively, with specific oligonucleotides (Table 1), to obtain a final AvrII-AscI DNA product. These AvrII-AscI fragments were introduced into the same sites of pBAC-REP-1 (1). REP-TRS-N-3a mutant was generated with two overlapping PCR fragments, one containing the proximal TRS-N sequence (from nucleotide −48 to ATG of gene N) and the other one the 3a gene, using as template the pBAC-TGEV and specific oligonucleotides (Table 1). The resulting DNA product with AvrII and PacI sites at 5′ and 3′ ends, respectively, was cloned into an intermediate plasmid that contained the other PCR fragment with the CS-N (from nucleotide −12 to ATG of N gene) and N gene sequence with PacI and AscI sites at the 5′ and 3′ ends, respectively, generated from the template pBAC-TGEV and specific oligonucleotides (Table 1). Finally, the AvrII-AscI fragment was introduced into AvrII-AscI sites of pBAC-REP-1 to obtain the REP-TRS-N-3a mutant. To generate the REP-TRM-3a mutant replicon, the distal TRS-N sequence (173 nt plus dE plus 20 nt) was inserted into the AvrII site of the REP-TRS-N-3a mutant cDNA. The REP-pE-3a-AD-dE and REP-3a-AD-dE mutants were generated with two overlapping PCR fragments using as template the pBAC-TGEV and specific oligonucleotides (Table 1). To generate REP-pE-3a-AD-dE, pE-CS-N (from nucleotide −48 to ATG of the N gene) and 3a gene sequences were added to the distal element and its flanking sequences (173 nt plus dE plus 20 nt). In the case of REP-3a-AD-dE, CS-N (from nucleotide −19 to ATG of the N gene) and 3a gene sequence were added to the distal element and its flanking sequences (173 nt plus dE plus 20 nt). Both DNA products contained AvrII and PacI sites at the ends. These products were introduced into an intermediate plasmid, containing CS-N and N gene sequence with PacI and AscI sites. The final AvrII-AscI DNA products were introduced into AvrII-AscI sites of pBAC-REP-1, leading to the new mutant replicons REP-pE-3a-AD-dE and REP-3a-AD-dE. The AD-TRS-N mutant was constructed with specific oligonucleotides (Table 1) from two overlapping PCR fragments, one including the active domain sequence and another one containing the CS-N and N gene sequences. The resulting product was introduced into the AvrII and AscI sites of pBAC-REP-1. For the construction of mutants dE-173-20-A, -B, and -C, the sequences including the active domain variations and the distal element were synthesized *de novo* (GENEART), including AvrII sites at both ends. Synthetic gene sequences were introduced into the same site of the TRS-N-ΔdE mutant (21). The cDNAs used to obtain mutant infectious viruses were constructed using an intermediate plasmid with the AvrII-AvrII fragment of TGEV (nucleotides 22,973 to 25,873). PCR fragments containing the sequence of the optimized transcription-regulating motif and those with mutations in the proximal and distal elements were joined to the 3a gene sequence by overlapping PCRs using as template pBAC-TGEV and specific oligonucleotides (Table 1). These fragments with BmgBI sites at both sides were introduced into the same sites of the intermediate plasmid. Then, mutated TGEV AvrII-AvrII fragments were used to replace the wild-type region of pBAC-TGEV.

RNA analysis by quantitative RT-PCR. Total intracellular RNA was extracted at 24 h posttransfection (p.t.) from transfected BHK-N cells or at 16 h postinfection from ST cells infected with mutant TGEVs. RNAs were purified with the RNeasy minikit (Qiagen) according to the manufacturer's specifications. To remove transfected DNA from samples for 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 the RNeasy minikit (Qiagen). cDNAs were synthesized at 37°C for 2 h with the

TABLE 1. Oligonucleotides used for directed mutagenesis

Mutant(s)	Oligonucleotide	5′→3′ sequence[a]
dE-153-158	dE 13 VS	**CTATACC**ATATGTAAT**AATTTT**CTTTAGTATTGCAGGTGCAATTGTT
dE-133-158	dE 14 VS	TTCCTAGGCTGTGCTACAATATGGAAGACC
dE-113-158	dE 200 AvrII VS	TTCCTAGGCCTCAATTCAGCTGGTTCGTGTATG
dE-45-158	dE 100 AvrII VS	TTCCTAGGGGCTCTTACGATTTTTAATGCATAC
dE-20-158	dE 50 AvrII VS	TTCCTAGGTCGGAATACCAAGTGTCCAG
dE-6-158	dE-173-20 AvrII VS	TTCCTAGGGTCCAGATATGTAATGTTCGGCTTT
dE-153-158, dE-133-158, dE-113-158, dE-45-158, dE-20-158, dE-6-158, 2+dE+2, 4+dE+4, 6+dE+6	Rep Mut 3 RS	AACCTAGGCATAGCTTCTTCCTAATGCACTAACGCAAAG
dE-173-95	dE 200 AvrII RS	AACCTAGGAAGACTTAGTCCTTCTGTACAACTG
dE-173-45	dE 100 AvrII RS	AACCTAGGCAGAGTACAAATGTAACAATTGCAC
dE-173-20	dE 50 AvrII RS	AACCTAGGCTGCAATACTAAAGCCGAAC
dE-173-6	dE 20 AvrII RS	AACCTAGGCCGAACATTACATATCTGGACACTT
dE-103-20	dE 16 RS	AACCTAGGCTGCAATACTAAAGCCGAACATTACATATTTATAAGC ATTTTAATGCC
2+dE+2	2+dE+2 RS	AGCCGA**TT**ATTACATAT**GG**GGACACTTGGTATTCC
	2+dE+2 VS	TGTCC**CC**ATATGTAAT**AA**TCGGCTTTAGTATTGCA
4+dE+4	4+dE+4 RS	AGCC**AATT**ATTACATAT**GGTA**ACACTTGGTATTCC
	4+dE+4 VS	TGT**TACC**ATATGTAAT**AATT**GGCTTTAGTATTGCA
6+dE+6	6+dE+6 RS	AG**AAAATT**ATTACATAT**GGTATAG**CTTGGTATTCCGAGTAT
	6+dE+6 VS	**CTATACC**ATATGTAAT**AATTTT**CTTTAGTATTGCAGGTGCAATTGTT
dE-173-95, dE-173-45, dE-173-20, dE-173-6, dE-103-20, 2+dE+2, 4+dE+4, 6+dE+6, pE75, pE45, pE20 AD-TRS-N	Rep Mut 3 VS	TTCCTAGGTGGAACTTCAGCTGGTCTATAATATTGATC
pE75	TRS-N1 RS	AATTTTTCTTGCTCACTCAAATTATCAGTTCTTGCCTCTGTTGAGT AATCACCAGCTTTAGATTTTACATAGTAACTGCAATACTAAAGC CGAAC
pE45	TRS-N2 RS	AATTTTTCTTGCTCACTCAAATTATCAGTTCTTGCCTCTGTTGAGC TGCAATACTAAAGCCGAAC
pE20	TRS-N3 RS	AATTTTTCTTGCTCACTCAACTGCAATACTAAAGCCGAAC
pE75, pE45, pE20	pE N VS	TTGAGTGAGCAAGAAAAATTATTA
pE75, pE45, pE20, AD-TRS-N	3′ N AscI RS	TTGGCGCGCCTTAGTTCGTTACCTCATCAATTATC
REP-TRS-N-3a	Rep 120 N AvrII VS	TTCCTAGGTTGAAAGCAAGTAGTGCGACTGG
	Rep Mut 3a RS	GTAAATGGATTTGACAATGTCCATTTAGAAGTTTAGTTA
REP-TRM-3a	Rep 5′3a VS	ACATATGGTATAACTAAACTTCTAAATGGACATTGTCAAA
	3′-3a PacI RS	TTTTAATTAACTAGGAAACGTCATAGGTATGGTCT
REP-pE-3a-AD-dE	AvrII pE VS	TTCCTAGGTTGAGTGAGCAAGAAAAATTATTACATATGG
REP-3a-AD-dE	AvrII CS VS	TTCCTAGGGGTATAACTAAACTTCTAAATGGACATTG
REP-TRS-N-3a	5′ PacI CS-N VS	TTTTAATTAACTAAACTTCTAAATGGCCAACCAGG
REP-TRM-3a	3′ N AscI RS	TTGGCGCGCCTTAGTTCGTTACCTCATCAATTATC
REP-pE-3a-AD-dE	3′3a+5′mENH RS	TTAAACAACTATATGACTATTGACTTCTTC
REP-3a-AD-dE	PacI 3′ dE RS	AATTAATTAACTGCAATACTAAAGCCGAACATTAC
REP-3a-AD-dE	3′3a+5′mENH VS	GAAGAAGTCAATAGTCATATAGTTGTTTAATGGAACTTCAGCTGG TCTATAATATTG
AD-TRS-N	dE-, pE- RS	GGTTGGCCATTTAGAAGTTTAGTTATACCGTCTGGACACTTGGTA TTCCGAGTATGC
	dE-, pE- VS	GGTATAACTAAACTTCTAAATGGCCAACC
TRM-3a	3′mENH+5′3a(AUG) RS	GGATTTGACAATGTCCATTTAGAAGTTTAGTTATACCATATGTAA TAATTTTTCTTGC
TRM(19)3a	3′mENH(19)+′3a(AUG) RS	GGATTTGACAATGTCCATTTAGAAGTTTAGTTATACCATATGTAA TAATTTTTCTTGCTCACTCAACTGCAATACTAAAGCAAATTATT ACATATGGTATCACTTGGTATTCCGAGTATG
TRM*3a	3′mENH*+5′3a(AUG) RS	GGATTTGACAATGTCCATTTAGAAGTTTAGTTATACC**TTAAAGTTA** AATTTTTCTTGCTCACTCAACTGCAATACTAAAGCCGAAC**TAACT TTAA**CTGGACACTTGGTATTCCGAGTATG
TRM-3a, TRM(19)3a, TRM*3a	5′3a(AUG) VS	GGTATAACTAAACTTCTAAATGGACATTGTCAAATCCATTTACAC ATCCG
	3a-AvrII RS	TTCCTAGGTTAAAGTTTGTACTACGGTAC
TRM*3a, TRM-3a, TRM(19)3a	BmgBI-S.end+5′En VS	AACACGTCCATTAATGGAACTTCAGCTGGTCTATAATATTGATCG

[a] The mutated nucleotides are shown in bold. Restriction sites are underlined. AvrII, CCTAGG; AscI, GGCGCGCC; PacI, TTAATTAA; BmgBI, CACGTC.

MultiScribe reverse transcriptase (high-capacity cDNA reverse transcription kit; Applied Biosystems). Specific oligonucleotides were used to obtain cDNAs from viral sequences. Real-time RT-PCR was used for quantitative analysis of genomic and subgenomic RNAs from infectious TGEV and TGEV-derived replicons. Oligonucleotides used for quantitative PCRs (Table 2) 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. Detection was performed with an ABI Prism 7000 sequence detection system (Applied Biosystems). Data were analyzed with ABI Prism 7000 SDS version 1.2.3 software. The relative quantifications were performed using the $2^{-\Delta\Delta Ct}$ method, which compares cycle threshold (C_T) values (16). For each mutant sequence, two independent replicons were constructed. Each of these constructs was analyzed in two independent transfections, and RNA from each transfection experiment was analyzed twice by qRT-PCR. To minimize transfection variability, each pair of data used for comparison came from the same transfection and qRT-PCR experiment.

RNA analysis by Northern blotting. Total intracellular RNA was extracted at 16 h postinfection from virus-infected ST cells by using the RNeasy minikit

TABLE 2. Oligonucleotides used for quantitative RT-PCR analysis

Amplicon	Forward primer[a]		Reverse primer[a]	
	Name	Sequence (5′→3′)	Name	Sequence (5′→3′)
gRNA	RT-REP-VS	TTCTTTTGACAAAACATACGGTGAA	RT-REP-RS	CTAGGCAACTGGTTTGTAACATCTTT
sgmRNA-N	Ldrt-VS	CGTGGCTATATCTCTTCTTTTACTTTAACTAG	N(82)-RS	TCTTCCGACCACGGGAATT
sgmRNA-3a	Ldrt-VS	CGTGGCTATATCTCTTCTTTTACTTTAACTAG	rt3a-RS	ATCAAGTTCGTCAAGTACAGCATCTAC
sgmRNA-S	L-CS1-VS	CCAACTCGAACTAAACTTTGGTAACC	L-CS1-RS	TCAATGGCATTACGACCAAAAC
sgmRNA-M	Ldrt-VS	CGTGGCTATATCTCTTCTTTTACTTTAACTAG	mRNAM-RS	GCATGCAATCACACACGCTAA
sgmRNA-7	Ldrt-VS	CGTGGCTATATCTCTTCTTTTACTTTAACTAG	7(38)-RS	AAAACTGTAATAAATACAGCATGGAGGAA

[a] The hybridization sites of the oligonucleotides within the TGEV genome were as follows: RT-REP-VS (nt 4829 to 4853), RT-REP-RS (nt 4884 to 4909), Ldrt-VS (nt 25 to 56), N(82)-RS (nt 26,986 to 27,004); rt3a-RS (nt 24,863 to 24,889), L-CS1-VS (hybridizes into the leader-body fusion of the sgmRNA-S, including nt 84 to 99 and nt 20,339 to 20,348), L-CS1-RS (nt 20,383 to 20,404), mRNAM-RS (nt 26,140 to 26,160), 7(38)-RS (nt 28,086 to 28,114).

(Qiagen) according to the manufacturer's instructions. Northern blotting was performed as previously described (28). The 3′-untranslated region (UTR)-specific single-stranded DNA probe used for detection was complementary to nt 28,300 to 28,544 of the TGEV strain PUR46-MAD genome (23).

***In silico* analysis.** Potential base-paring score calculations were performed as previously described (35). ΔG calculations were performed using the two-state hybridization server (http://www.bioinfo.rpi.edu/applications/hybrid/twostate.php) (18). Secondary structure predictions were performed using the Mfold web server for nucleic acid folding and hybridization prediction (http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi) (33). The analysis of sequences was performed using DNASTAR Lasergene software 7.0. Comparison of M gene RNA sequences was performed using ClustalW2/EBI (http://www.ebi.ac.uk/Tools/clustalw2/index.html).

RESULTS

Effect of the extent of complementarity between the proximal and distal elements on N gene transcription enhancement. We previously demonstrated that the reduction in the extent of complementarity between proximal and distal elements correlated with a decrease in the transcriptional activation of sgmRNA N (21). In order to analyze whether an increase in the extent of complementarity between proximal and distal elements led to a further increase in sgmRNA N transcription, the nucleotides flanking the distal element were mutated to make them complementary to those nucleotides flanking the proximal element. A set of three mutants was engineered by reverse genetics using a TGEV-derived replicon (E2-TRS-N) (21). This replicon included the N gene sequence preceded by a proximal TRS-N (nt −1 to −120 in relation to the start codon of gene N), which contained the CS-N and the proximal element. Preceding the proximal TRS-N, the distal region, including nt 136 to 490 of the M gene coding sequence, which contains the distal element, was inserted (Fig. 1A). The E2-TRS-N replicon efficiently enhanced sgmRNA N transcription, as previously shown (21). A mutant replicon including four point mutations, two at each side of the distal element (2+dE+2 mutant) increased in 4 nucleotides the complementarity between the proximal and distal elements at homologous positions. With these modifications the complementarity between the proximal and distal elements was increased from 9 nucleotides in the wild-type context (E2-TRS-N) to up to 13 nucleotides in the 2+dE+2 mutant (Fig. 1A). In other engineered replicons (4+dE+4 and 6+dE+6 mutants), 4 or 6 point mutations were introduced at the 5′- and 3′-flanking positions of the distal element, respectively, which increased the complementarity between the proximal and distal elements up to 17 and 21 nucleotides, respectively (Fig. 1A). Since efficient RNA synthesis in TGEV is associated with the presence of the viral nucleoprotein (34) and N protein expression could be affected in these mutants, the analysis of N gene transcription was performed in BHK cells expressing the N protein in *trans* (BHK-N) (1), to exclude any potential variation in transcription due to the availability of N protein. BHK-N cells were transfected with the cDNAs encoding mutant replicons, and the level of intracellular sgmRNA N was analyzed by qRT-PCR. The forward primer used for qRT-PCR analysis of sgmRNA N hybridized with the viral leader sequence, while the reverse primer hybridized within the N gene coding sequence (Table 2). These oligonucleotides specifically detected sgmRNA N synthesized from the TGEV-derived replicon, since mRNA N expressed from the Sindbis virus plasmid in BHK-N cells did not include the TGEV leader sequence. The amount of sgmRNA N in relation to that of genomic RNA (gRNA) was determined. This sgmRNA N/gRNA ratio in the reference control, the TRS-N-ΔdE replicon, was defined as 1. In this control replicon the distal element was absent (TRS-N-ΔdE), and the levels of N gene transcription were regulated by proximal TRS-N sequences, including the proximal element, but not by the interaction between the proximal and distal elements, leading to basal transcriptional levels of sgmRNA N (21). As a positive control, the replicon E2-TRS-N was used (21). An increase of the complementarity between the proximal and distal elements in the new mutants led to a significant increase in sgmRNA N levels compared to those of the replicon E2-TRS-N with the wild-type sequences (Fig. 1B). A linear correlation was established between the proximal and distal element extent of complementarity and the accumulation of sgmRNA N, reaching a maximum for the 4+dE+4 mutant (Fig. 1B). A further increase in complementarity from the 4+dE+4 to 6+dE+6 construct did not produce a significant increase in sgmRNA N transcription (Fig. 1B). Since the number of complementary nucleotides is proportional to the ΔG of the RNA-RNA interaction, this result showed that the decrease in the free energy associated with the distal and proximal elements interaction, that is, the increase in stability, had a positive effect on N gene transcription enhancement.

Relevance of distal element flanking sequences for N gene transcription enhancement. The distal element was previously shown to be essential for N gene transcription enhancement (21) in a context that included 5′- and 3′-flanking sequences of 173 and 158 nt, respectively (21). To determine the minimal sequences flanking the distal element required for transcriptional activation of sgmRNA N, in a first approach, six 5′ deletion mutants were generated, maintaining constant the 158

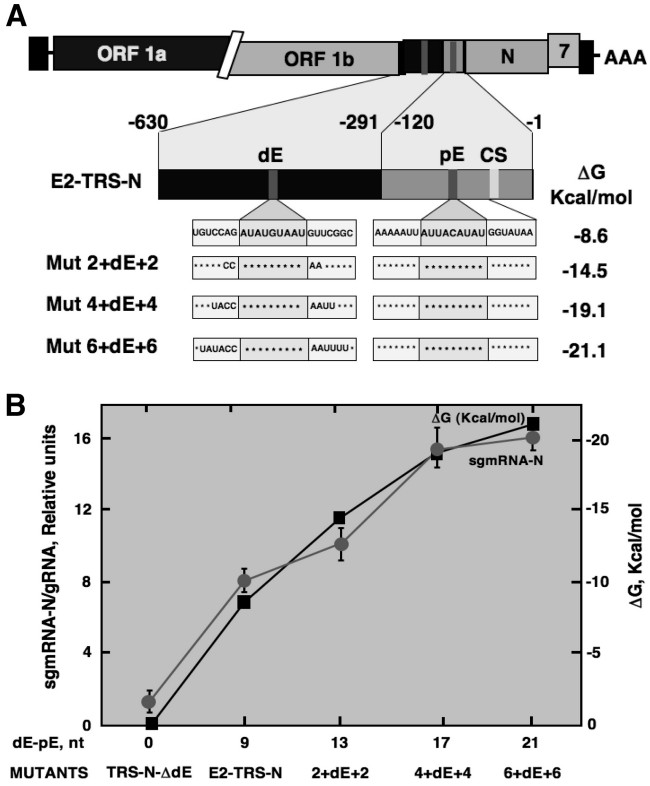

FIG. 1. Effect of the complementarity extent between the proximal and distal elements on transcriptional activity. (A) Scheme showing the genetic structure of the E2-TRS-N mutant. Regulatory sequences preceding the N gene are indicated by a number that corresponds to the position of the nucleotide in relation to the first base of the N gene start codon. The position −630 represents nucleotide 136 of the M gene coding sequence and nt 26,281 of the TGEV genome. The position −291 represents nucleotide 490 of the M gene and nt 26,620 of the genome. The position −120 represents nucleotide 682 of the M gene and nt 26,803 of the genome. The position −1 represents the nucleotide preceding the N gene ATG and nt 26,922 of the genome (23). The seven nucleotides immediately flanking both sides of the proximal (pE) and distal (dE) elements are also shown. The 3′-flanking nucleotides of the proximal element are just upstream of the CS-N, and the last nucleotides, UAA, represent the stop codon of the M gene. In the lower part, the names of the mutants with extended complementarity between proximal and distal elements are shown. The nucleotide changes introduced in the mutants are indicated in the boxes shown below the bar, where the relative positions of proximal and distal elements are shown. Asterisks represent nucleotides identical to those in the E2-TRS-N mutant. To the right of these boxes, the Δ*G* values associated with base-pairing between proximal and distal elements are indicated. (B) On the *x* axis, the numbers of complementary nucleotides between proximal and distal elements are indicated. Below these numbers, the names of the corresponding mutants are provided. The black line shows the stability of the proximal and distal element interaction (Δ*G*). The gray line shows the accumulation of sgmRNA N of each mutant, expressed as sgmRNA N/gRNA in relation to the TRS-N-ΔdE reference replicon (lacking the distal element and then having 0 complementary nt), which represented 1. The data are the averages of four independent transfection experiments. Quantitative RT-PCR analysis was performed in duplicate in each case. Error bars represent the standard deviations.

nt at the 3′-flanking side of the distal element (Fig. 2A). The newly generated replicons included 153, 133, 113, 45, 20, or 6 nt at the distal element 5′-flanking side (mutants dE-153-158, dE-133-158, dE-113-158, dE-45-158, dE-20-158, and dE-6-158, respectively) (Fig. 2A). BHK-N cells were transfected with the cDNAs encoding mutant replicons, and the levels of intracellular sgmRNA N were analyzed by qRT-PCR. Replicons dE-153-158, dE-133-158, and dE-113-158 led to a 5- to 6.75-fold increase, whereas the positive control, the E2-TRS-N mutant, with 173 nt flanking the 5′ side of the distal element, showed around an 8-fold increase above reference transcription levels of the TRS-N-ΔdE replicon (Fig. 2A). In contrast, mutants with smaller distal element 5′-flanking sequences (dE-45-158, dE-20-158, and dE-6-158 mutants) showed significantly reduced transcriptional activity, similar to that of the TRS-N-ΔdE replicon (Fig. 2A). This result indicated that the interaction between proximal and distal elements by itself was not enough to enhance the transcription of the N gene. In addition, the proximal distal element 5′-flanking sequences (45 nt) were not sufficient for N gene transcriptional enhancement, and the more distant sequences on 5′ flanks of the distal element (nt 173 to 45) were essential for the increase of the transcriptional activity (Fig. 2A).

To evaluate the relevance of the distal element 3′-flanking sequences in N gene transcription, four deletion mutants were generated, maintaining constant the 173 nt of the distal element 5′-flanking sequences and including 95, 45, 20, or 6 nt of the distal element 3′-flanking sequence (mutants dE-173-95, dE-173-45, dE-173-20, and dE-173-6, respectively) (Fig. 2B). All these 3′ deletion mutants maintained or even increased sgmRNA N transcriptional activity compared to the E2-TRS-N positive control (Fig. 2B). The activity of the dE-173-6 mutant was similar to that of the E2-TRS-N positive control, indicating that the distal element 3′-flanking sequences consisting of only 6 nt were enough for sgmRNA N transcriptional activation. In contrast, the other 3′ deletion mutants, including 20, 45, or 95 nt of the distal element 3′ flanking sequences, enhanced sgmRNA N transcription more efficiently than the E2-TRS-N positive control, indicating that the distal element 3′-flanking sequences were not essential for activity. The deletion of the distal element 3′-flanking sequences did not affect the structure predicted by the Mfold program for mutant replicons at either the level of distal and proximal element interaction or the secondary structure predicted for the 5′ flanking sequences of the distal element (Fig. 3B). Therefore, the deletion of sequences between the proximal and distal elements had a positive effect on N gene transcription enhancement.

We have shown that distant sequences on the 5′ flank of the distal element are essential to enhance N gene transcription (Fig. 2A). To analyze whether these distant sequences (nt −173 to −45 with respect to the distal element) alone would maintain transcriptional enhancement, the dE-103-20 mutant replicon, including the most-5′ 103 nt, was constructed (Fig. 3). The transcription activity of the dE-103-20 mutant was similar to that of the TRS-N-ΔdE replicon (Fig. 3), suggesting that the 173-nt complete region on the 5′ flank of the distal element was also required to enhance N gene transcription. This sequence was named the active domain (AD). Deletion of distal element 5′-flanking sequences in the dE-103-20 mutant replicon did not affect the distal and proximal elements interaction, according to the secondary structure predictions. Since in this mutant replicon the active domain is partially deleted (Fig. 3), its secondary structure is not maintained (Fig. 3C).

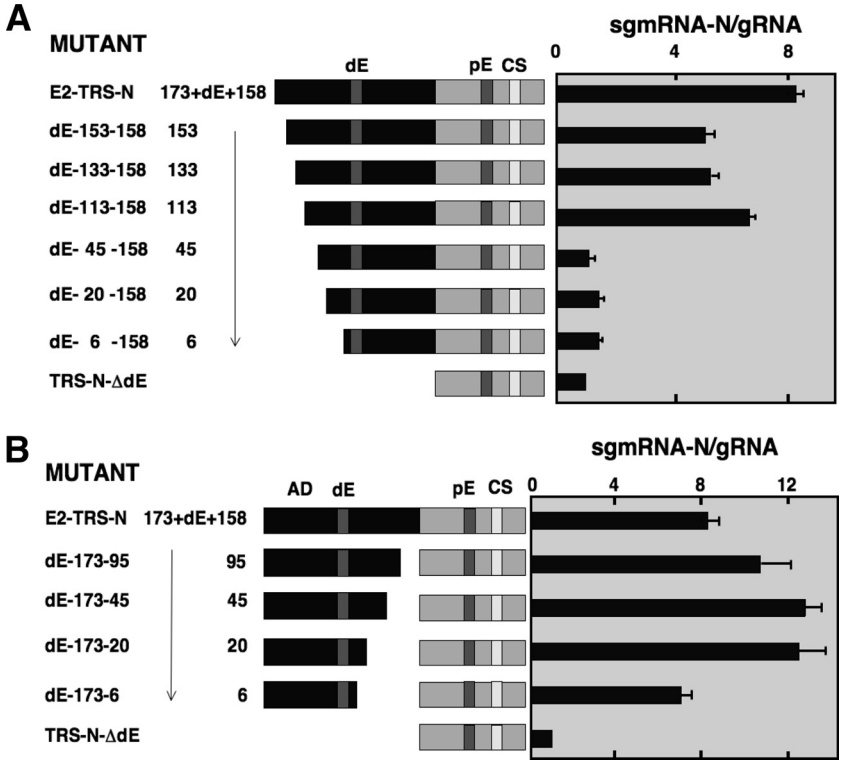

FIG. 2. Relevance of flanking sequences of the distal element on N gene transcription enhancement. To the left is a scheme showing the name and genetic structure of the mutants. Boxes represent the regulatory sequences preceding the N gene. To the right are qRT-PCR analysis results of the sgmRNA N relative amount (sgmRNA N/gRNA) expressed in relation to the TRS-N-ΔdE reference replicon, which represents 1. E2-TRS-N, positive control. (A) The lengths of the 5′-flanking sequences of the distal element are indicated by the numbers. The arrow represents sequences that are identical to those in the E2-TRS-N mutant. (B) The lengths of the 3′-flanking sequences of the distal element are indicated by numbers. The data are the averages of four independent transfection experiments. Quantitative RT-PCR analysis was performed in duplicate in each case. Error bars represent the standard deviations. The AD corresponds to 173 nt upstream of the distal element.

Relevance of the proximal element 5′-flanking sequences in N gene transcription enhancement. It has been observed that the deletion of the sequences between the distal and proximal elements increased N gene transcriptional activity (Fig. 2B). In order to bring into closer proximity the proximal and distal elements, three new deletion mutants in the proximal element 5′-flanking sequences were engineered (Fig. 4). The new replicon mutants pE-75, pE-45, and pE-20 conserved the same distal element 5′- and 3′-flanking sequences of the dE-173-20 replicon (173 and 20 nt at the 5′- and 3′-flanking sequences, respectively). In these mutants the proximal element 5′-flanking sequences were reduced to 75, 45, and 20 nt, respectively (Fig. 4). BHK-N cells were transfected with the cDNAs encoding mutant replicons, and the levels of intracellular sgmRNA N were analyzed by qRT-PCR. The three mutants (pE-75, pE-45, and pE-20) showed the same N gene transcriptional activity, with a 16-fold increase above the reference levels in the TRS-N-ΔdE replicon. This activity was higher than that of the dE-173-20 mutant, which showed a 12.6-fold increase (Fig. 4). These results indicated that the proximal element 5′-flanking sequences are not required for N gene transcription enhancement. Furthermore, it was confirmed that the deletion of the sequences between the proximal and distal elements had a positive effect on N gene transcription enhancement. The results also indicated that maximal transcriptional activity for the N gene was reached with the mutant replicon pE-75, in the sense that further reductions in the distance between proximal and distal elements (pE-45 and pE-20) did not lead to an increase in N gene transcription enhancement.

The region of the genome (642 nt upstream of the ATG of N gene) that includes the proximal and distal elements together with their flanking sequences responsible for the specific transcriptional regulation of the N gene was named the transcription-regulating motif (TRM) (Fig. 5A). Integrating all previous results, it can be concluded that an optimized transcription-regulating motif (TRM^{opt}), with a minimum size of 250 nt (pE20 mutant), was engineered using the native TRM (642 nt) by the deletion of genome nt 26483 to 26874 located within the 3′ end of the M gene, between proximal and distal elements (Fig. 5A). This optimized RNA motif led to transcription levels 4-fold higher than that of the original transcription-regulating motif sequence (21). The RNA secondary structure of the active domain and the interaction between the proximal and the distal elements were maintained in the optimized transcription-regulating motif, according to the Mfold predictions (Fig. 5B).

Transcriptional regulation of another gene by the transcription-regulating motif and relevance of the relative positions of transcription-regulating motif elements. In TGEV the N gene is expressed under the control of the native transcription-regulating motif (21). In this study the influence of the transcrip-

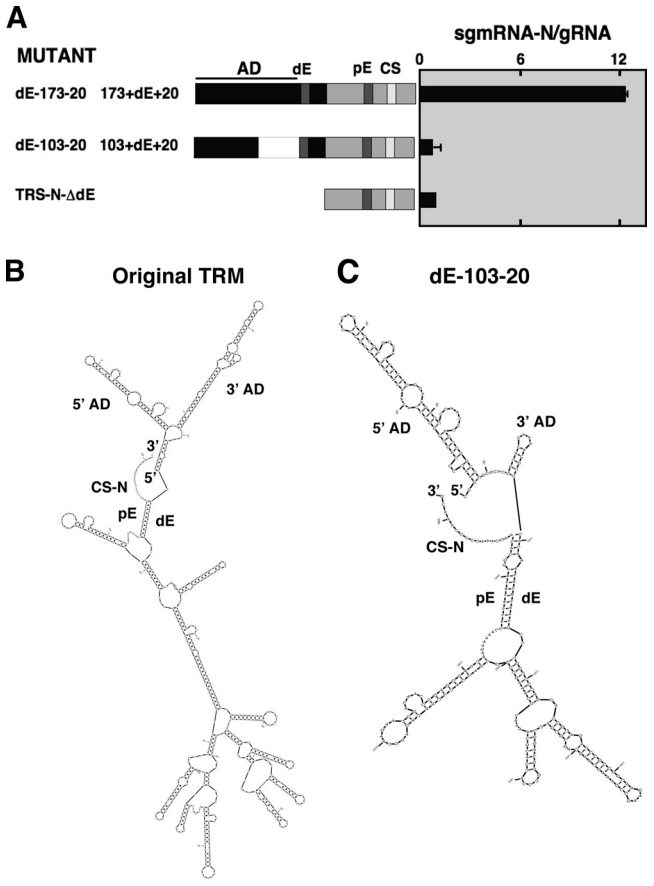

FIG. 3. Relevance of distal 5′-flanking sequences of the distal element on N gene transcription enhancement. (A) To the left, the scheme shows the names and genetic structures of the mutants, including the 5′-flanking sequences of the distal element preceding the N gene. To the right, qRT-PCR analysis of the sgmRNA N relative amount (sgmRNA N/gRNA) is expressed in relation to the TRS-N-ΔdE reference replicon, which represented 1. dE-173-20, positive control. The data are the averages of four independent transfection experiments. Quantitative RT-PCR analysis was performed in duplicate in each case. Error bars represent the standard deviations. (B) Scheme showing the Mfold RNA secondary structure prediction and stability of the original transcription-regulating motif, between nt 26281 and nt 26923 of the TGEV genome. 5′ AD, hairpin located at the 5′ end of the active domain; 3′ AD, hairpin located at the 3′ end of the active domain; pE-dE, RNA-RNA interaction between the proximal and distal elements; CS-N, conserved core sequence included within the TRS-N. ΔG is −169.41 kcal/mol, the free energy associated with the predicted RNA secondary structure. (C) Scheme showing the Mfold RNA secondary structure prediction and stability of the dE-103-20 mutant transcription-regulating motif. ΔG is −68.58 kcal/mol.

tion-regulating motif on the expression of gene 3a was analyzed. In addition, the relevance of the relative positions of the transcription-regulating motif elements (proximal and distal elements and the active domain) on transcription enhancement was studied. To these ends, four new mutants were constructed (Fig. 6). In mutant REP-TRM^{opt}-3a, the 3a gene was preceded by the optimized transcription-regulating motif sequences. As a control, the REP-TRS-N-3a mutant, lacking the active domain and the distal element, was constructed (Fig. 6A). BHK-N cells were transfected with the cDNAs encoding mutant replicons, and the levels of intracellular sgmRNA N were analyzed by qRT-PCR. The

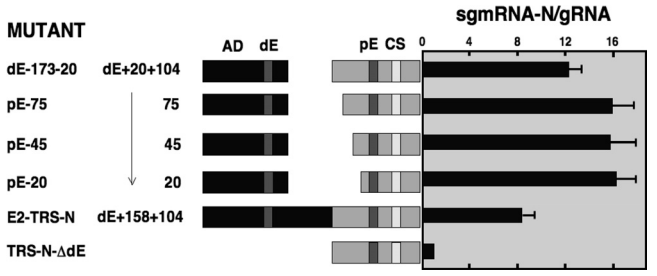

FIG. 4. Relevance of 5′-flanking sequences of the proximal element on N gene transcription enhancement. To the left is a scheme showing the names and genetic structures of the mutants and the lengths of the sequences on the 5′ flanks of the proximal element preceding the N gene. The arrow represents the sequences that are identical to those in the E2-TRS-N mutant. To the right are qRT-PCR analysis results of the sgmRNA N relative amount (sgmRNA N/gRNA) expressed in relation to the TRS-N-ΔdE reference replicon, which represented 1. E2-TRS-N, positive control. The data are the averages of four independent transfection experiments. Quantitative RT-PCR analysis was performed in duplicate in each case. Error bars represent the standard deviations.

REP-TRM^{opt}-3a mutant showed an efficient 3a gene transcription enhancement of around 13-fold, indicating that the optimized transcription-regulating motif increased the expression of the 3a gene (Fig. 6B).

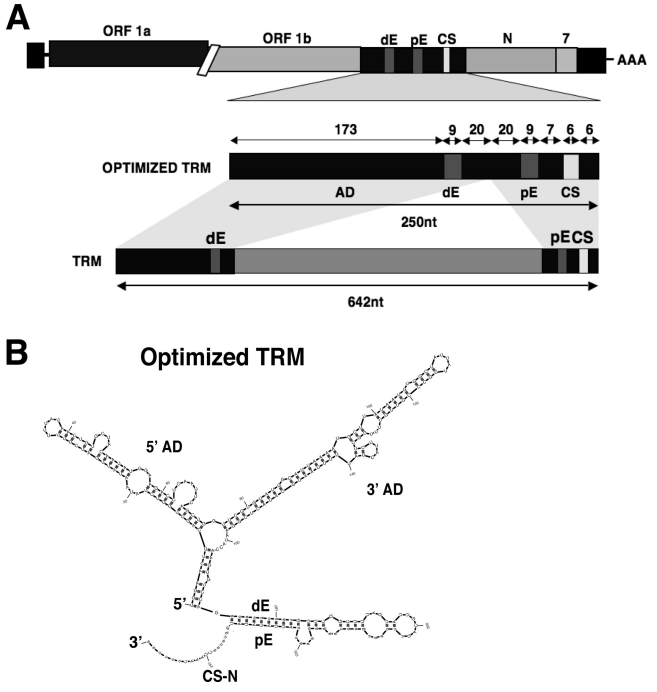

FIG. 5. Optimized transcription-regulating motif. (A) The upper line of the panel represents a replicon with the optimized TRM preceding the N gene. In the middle, a scheme of the optimized TRM is shown. The lengths (in nt) and the elements contained in the optimized TRM are indicated above and below the line, respectively. The line at the bottom represents the original sequence of the transcription-regulating motif. The sequences of the native TRM conserved in the optimized TRM are indicated by shadowed areas. The sequence removed in the optimized TRM corresponds to nucleotides 26,483 to 26,874 of the TGEV genome. (B) Scheme showing the Mfold RNA secondary structure prediction and stability of the optimized transcription regulating motif. ΔG is −54.50 kcal/mol.

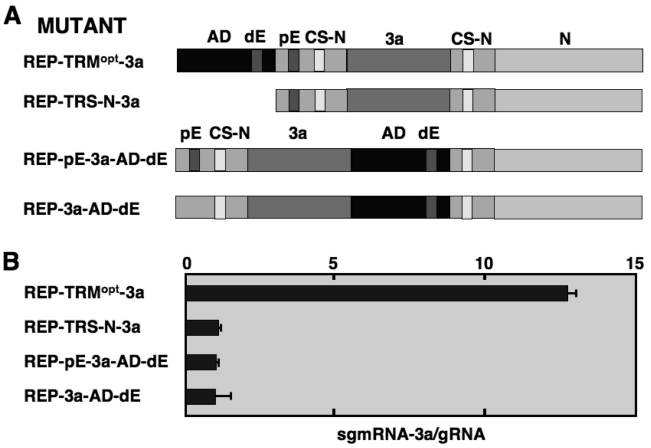

FIG. 6. Influence of the relative positions of the active domain and proximal and distal elements on transcriptional activation. (A) Scheme showing the names and the genetic structures of the mutants. Letters on top of the boxes represent the names of the genes and the elements of the transcription-regulating motif. (B) qRT-PCR analysis results for sgmRNA-3a relative to gRNA in each mutant, expressed in relation to the reference replicon REP-TRS-N-3a, which represented 1. The REP-TRM^{opt}-3a mutant included the optimized TRM controlling the expression of the 3a gene. The data are the averages of four independent transfection experiments. Quantitative RT-PCR analysis was performed in duplicate in each case. Error bars represent the standard deviations.

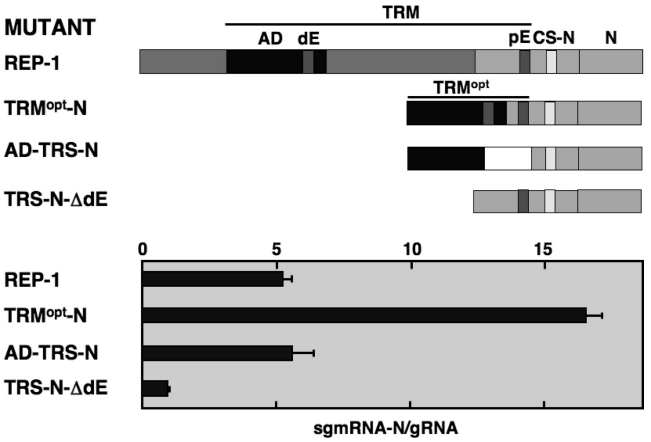

FIG. 7. Relevance of the active domain in N gene transcription enhancement. In the upper panel is a scheme showing the names and genetic structures of the mutants, including the regulatory sequences preceding the N gene. In the lower panel are qRT-PCR analysis results of the sgmRNA N relative amount (sgmRNA N/gRNA) expressed in relation to the reference replicon TRS-N-ΔdE, which represented 1. The data are the averages of four independent transfection experiments. Quantitative RT-PCR analysis was performed in duplicate in each case. Error bars represent the standard deviations.

To study whether the active domain and the distal element have to be located upstream of the gene and the proximal element, mutant REP-pE-3a-AD-dE was constructed. In this case the 3a gene was preceded by the N gene-proximal TRS, including the proximal element. The sequences comprising the active domain and distal element were relocated downstream of the 3a gene. Mutant REP-3a-AD-dE, designed as a negative control, was derived from REP-pE-3a-AD-dE by deleting the proximal element preceding the 3a gene (Fig. 6A). The REP-pE-3a-AD-dE mutant showed background 3a gene transcription levels, similar to those of the negative control, REP-3a-AD-dE (Fig. 6B), indicating that the active domain and distal element had to be located upstream of the transcriptionally regulated gene.

The active domain sequence located just upstream of the N gene core TRS enhances transcription of the N gene. In order to analyze the relevance of the active domain in N gene transcription enhancement, a new mutant replicon, AD-TRS-N, in which the proximal and distal elements and the sequences between these motifs were deleted, was constructed, relocating the active domain sequence immediately preceding the CS of the N gene (Fig. 7). The position of the active domain sequence in this mutant mimicked the position that this domain would have in the presence of the interaction between the proximal and distal elements, that is, close to the 5′ side of the CS-N. BHK-N cells were transfected with the cDNAs encoding mutant replicons, and the levels of intracellular sgmRNA N were analyzed by qRT-PCR. Interestingly, the AD-TRS-N mutant with the proximal and distal elements deleted enhanced 5-fold N gene transcription compared to the expression of the reference replicon TRS-N-ΔdE, which lacks the active domain and distal element. The observed increase in transcription was similar to that of replicon REP-1, which contains the transcription-regulating motif present in the native virus (Fig. 7). Nevertheless, the transcriptional activity of the replicon missing the proximal and distal elements (AD-TRS-N mutant) was 3-fold lower than that of the pE20 mutant, containing the complete optimized transcription-regulating motif (Fig. 7). This result indicated not only that the physical location of the active domain sequence just upstream of CS-N was sufficient for transcription enhancement, but also that the interaction between the proximal and distal elements was required for optimal transcription enhancement.

Requirement of the primary RNA sequence of the active domain for transcriptional activation. Mfold RNA secondary structure predictions of the active domain showed that it probably adopts a stable secondary structure with two hairpins (Fig. 5B). When the hairpin at the 5′ side was deleted (dE-113-158 mutant [Fig. 2A]), 80% of the activity was maintained by the 3′ hairpin. To analyze the functional relevance of the active domain secondary structure, in mutant dE-173-20-A (Fig. 8A and B) the hairpin located at the 3′ side was replaced by another one with a different nucleotide sequence but with the same secondary structure and similar stability (Fig. 5B) of the corresponding wild-type region, according to Mfold predictions. In a second construct (mutant dE-173-20-B), the 3′ hairpin was replaced by another one with a different nucleotide sequence and secondary structure, but with similar length and stability, based on Mfold predictions (Fig. 8A and C). In an alternative construction (mutant dE-173-20-C), the 3′ hairpin was replaced by another one with similar stability, but without similarities in primary sequence, length, or structure (Fig. 8A and D). After replicon transfection into BHK-pAPN-N cells, sgmRNA N levels were analyzed. N gene transcription levels of mutants dE-173-20-A, -B, and -C were similar to those of the TRS-N-ΔdE reference replicon and 12-fold lower than those of the dE-173-20 positive control, including the wild-type se-

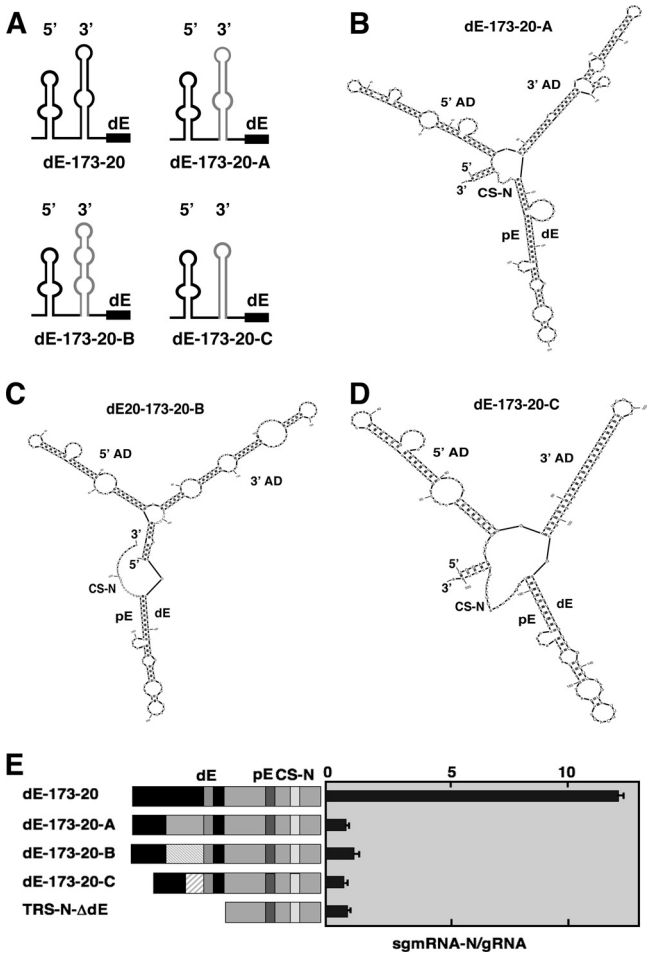

FIG. 8. Effect of the active domain secondary structure on transcriptional activation. (A) Scheme of the predicted secondary structures for the wild-type (dE-173-20) and mutant AD motifs (dE-173-20-A, dE-173-20-B, and dE-173-20-C). The nonmodified sequences are represented in black. Gray lines in the indicated mutants represent different versions of the AD 3′ hairpin. The distal element is also indicated. (B) Scheme showing the Mfold RNA secondary structure prediction and stability of the dE-173-20-A mutant transcription-regulating motif. ΔG is −57.80 kcal/mol. (C) Scheme showing the Mfold RNA secondary structure prediction and stability of the dE-173-20-B mutant transcription-regulating motif. ΔG is −59.50 kcal/mol. (D) Scheme showing the Mfold RNA secondary structure prediction and stability of the dE-173-20-C mutant transcription-regulating motif. ΔG is −76.39 kcal/mol. (E, left) Scheme showing the names and genetic structures of the mutants. Boxes represent the regulatory sequences preceding the N gene. The light boxes preceding the distal element represent different versions of the AD 3′ hairpin preserving (dE-173-20-A) or disrupting (dE-173-20-B and dE-173-20-C) the predicted RNA secondary structure. (Right) qRT-PCR analysis of the sgmRNA N relative amount (sgmRNA N/gRNA). TRS-N-ΔdE, reference replicon representing 1 relative unit. The data are the averages of four independent transfection experiments. Quantitative RT-PCR analysis was performed in duplicate in each case. Error bars represent the standard deviations.

quence of the active domain (Fig. 8E). The active domain mutant with the different nucleotide sequence that maintained a similar secondary structure (dE-173-20-A) completely lost its activity, indicating that the specific RNA primary sequence of the active domain was essential to maintain active domain activity.

Transcription enhancement of gene 3a by the transcription-regulating motif in a full-length TGEV infectious virus. The transcription enhancements shown above were obtained using TGEV-derived replicons. To analyze the effect of the optimized transcription-regulating motif in the context of a TGEV infectious virus, three different versions of the optimized transcription-regulating motif were introduced in a cDNA encoding the TGEV genome (Fig. 9A). Wild-type TRS-3a was replaced either by the optimized transcription-regulating motif (TGEV-TRM^{opt}-3a) or by a modified optimized transcription-regulating motif with extended complementarity (from nt 9 to 19) between the proximal and distal elements (TGEV-TRM$^{opt\text{-}19}$-3a) (Fig. 9B). In order to avoid potential interactions between the proximal and distal elements controlling the expression of gene 3a in the optimized transcription-regulating motif and the proximal and distal elements regulating the transcription of the N gene in the native TGEV (21), a recombinant TGEV-TRM^{opt}*-3a was engineered that included alternative complementary proximal and distal elements (Fig. 9B). Since the extent of the complementarity between the proximal and distal elements is relevant in transcriptional activation, the new proximal and distal elements were designed to conserve a ΔG associated with the base-pairing of the interaction similar to that in the wild-type virus. The wild-type 9-nt sequence of the distal element (5′-AUAUGUAAU-3′) in the optimized transcription-regulating motif was replaced by a new distal element (5′-UUAAAGUUA-3′) in TGEV-TRM^{opt}*-3a virus. At the same time, the proximal element was replaced by a new sequence (5′-UAACUUUAA-3′), complementary to that of the new distal element. The cDNAs containing the mutant genomes were transfected into BHK-N cells to rescue the infectious viruses. The three viruses were plaque purified four times. The region of the genome where the optimized transcription-regulating motif was introduced was sequenced. No deletions or nucleotide substitutions were identified in the optimized transcription-regulating motif region (data not shown). Therefore, the recombinant viruses were stable along passages in cell culture.

Northern blot analysis of viral RNAs showed similar sgmRNA expression patterns for the wild-type and mutant viruses (Fig. 9C), meaning that no alternative sgmRNAs were detected in mutant viruses. The relative amount of sgmRNA 3a/gRNA in the mutant virus was significantly increased compared to the wild-type virus. Quantitative RT-PCR confirmed that the levels of the sgmRNA-3a expressed by mutant viruses TGEV-TRM^{opt}-3a, TGEV-TRM$^{opt\text{-}19}$-3a, and TGEV-TRM^{opt}*-3a were 4- to 5-fold higher than those of the wild type. Additionally, the levels of sgmRNAs M, N, and 7, located downstream of gene 3a, showed a minor increase in relation to the wild-type virus, although this enhancement was not statistically significant. In contrast, the levels of sgmRNA S, located upstream of gene 3a, showed a slight reduction, which was also statistically irrelevant (Fig. 9D). These results indicated that only the sgmRNA 3a was significantly and specifically enhanced in mutant viruses (Fig. 9D). Therefore, the optimized transcription-regulating motif also increased the transcription of an alternative gene, located in a genome position different from that of gene N, in the

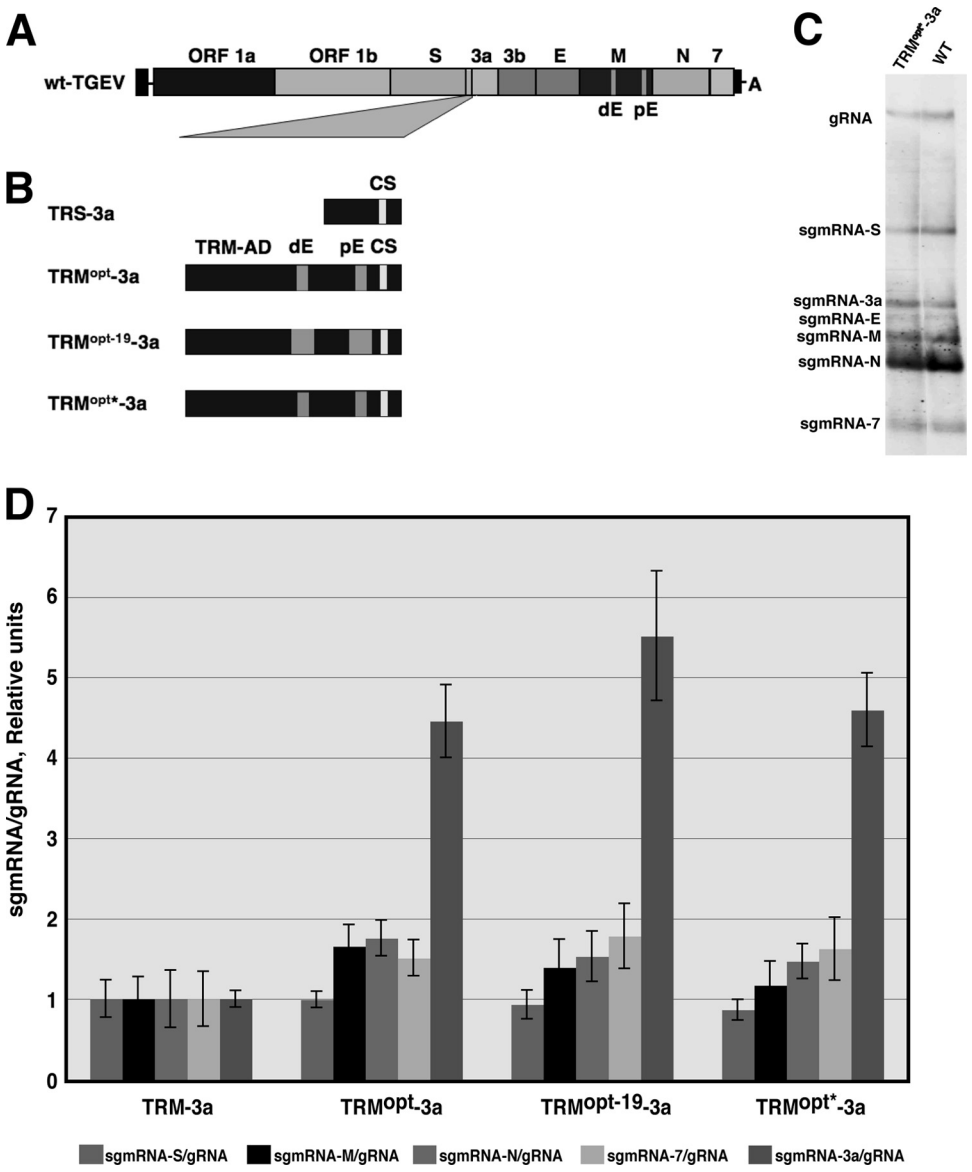

FIG. 9. Transcriptional activation of the 3a gene by the optimized transcription-regulating motif in TGEV infectious viruses. (A) Scheme showing the genetic structure of TGEV. Letters above the boxes represent the names of the genes. The proximal and distal elements regulating N gene transcription in wild-type (wt) TGEV are indicated. (B) Scheme showing the genetic structure of sequences regulating 3a gene transcription in wt (TRS-3a) and mutant viruses. TRM-AD, active domain in the TRM; TRM^{opt}-3a, optimized TRM regulating the transcription of the 3a gene; TRM^{opt-19}-3a, optimized TRM including 19 complementary nucleotides between the proximal and distal elements; TRM^{opt*}-3a, modified TRM^{opt} with alternative complementary sequences for the proximal and distal elements. (C) Analysis by Northern blotting of viral RNAs at 16 h postinfection from a wild-type virus and the TRM^{opt*}-3a mutant. Viral gRNA and sgmRNAs are indicated. (D) qRT-PCR analysis of sgmRNAs (S, M, N, 7, and 3a) relative amount in mutant viruses normalized for gRNA and in reference to levels in the wild-type virus. The data are the averages of four independent infection experiments evaluated twice. Quantitative RT-PCR analysis was performed in duplicate in each case. Error bars represent the standard deviations.

context of the infectious virus. No significant differences were found between the transcriptional activation in TRM^{opt}-3a and TRM^{opt*}-3a mutant viruses, indicating that proximal and distal elements of TRM^{opt}-3a did not interfere with the proximal and distal elements in the original transcription-regulating motif controlling the expression of the N gene. Additionally, a slight increase in transcription of gene 3a was observed in virus TGEV-TRM^{opt-19}-3a compared to that of TRM^{opt}-3a virus, as expected from the higher stability of the proximal and distal elements interactions.

DISCUSSION

CoV transcription is regulated by many factors (5), including the base-pairing between TRS-L and the complement of TRS-B in the nascent RNA, which represents the main driving force in the regulation of CoV sgmRNA transcription (28, 35).

We previously identified an additional transcription regulation mechanism specifically enhancing the transcription of the TGEV N gene (21). This second level of regulation, specifically affecting N gene expression, is mediated by a long-distance RNA-RNA interaction between two 9-nt complementary sequences (21). In this work, we extended this observation and provided evidence for novel elements essential for this transcription enhancement.

The relevance of the complementarity between proximal and distal elements was initially analyzed by disrupting this complementarity (21). In this study, the effect of extending the complementarity between proximal and distal elements up to 21 nt confirmed a direct correlation between the extent of complementarity and transcriptional enhancement. A novel observation of this study was the identification of a sequence flanking the 5′ side of the distal element that was essential for transcription enhancement. In contrast, the distal element 3′-flanking sequence and proximal element 5′-flanking sequences were not necessary. Furthermore, their deletion, which decreased the distance between proximal and distal elements, increased transcriptional enhancement. An optimized transcription-regulating motif was obtained, containing 250 nt, instead of the 642 nt of the native transcription-regulating motif (Fig. 5). The optimized transcription-regulating motif led to a 4-fold increase in transcription levels in relation to the native sequence in the context of the CoV replicon.

The relevance of the relative position of transcription-regulating motif elements (active domain and proximal and distal elements) on transcriptional activation has been shown. The active domain and the distal element need to be located upstream of the proximal element and of the gene that is being regulated by the transcription-regulating motif. In addition, the transcription-regulating motif activity was independent of its position within the replicon and the nature of the transcribed gene. When the optimized transcription-regulating motif was located preceding the 3a gene, in a position more distal from the 3′ end of the genome than that of the N gene, the activity of the transcription-regulating motif was essentially maintained, indicating that the enhancer activity was conserved at different genome positions.

Interestingly, the optimized transcription-regulating motif was also relocated within the genome of TGEV infectious viruses, and the transcriptional enhancement was still conserved. Recombinant full-length TGEV infectious viruses with the optimized transcription-regulating motif controlling the expression of the 3a gene were stable. Neither changes in the introduced sequence nor changes in the expression pattern of viral sgmRNAs were observed, as confirmed by quantitative analysis of S, M, N, and 7 sgmRNA levels (Fig. 9D). The significant 5-fold increase in transcription promoted by the optimized transcription-regulating motif in infectious viruses supports the potential use of the these transcription regulating sequences to enhance the transcription of foreign genes in CoV-derived expression vectors.

A novel observation of this study was the effect of the active domain by itself on transcription enhancement, in the absence of the proximal and distal elements. Transcription levels enhanced just by the active domain were similar to those observed with the original transcription-regulating motif. These levels could be increased 3-fold by using the optimized regulating sequences. These results suggest that the active domain was sufficient to enhance transcription when located immediately upstream of CS-N. Nevertheless, the long-distance RNA-RNA interaction between proximal and distal elements was required to reach maximum transcriptional activation, possibly because the interaction between the proximal and distal elements led to a more efficient conformation of the active domain to enhance the frequency of template switching of the nascent negative RNA to the leader TRS (Fig. 10).

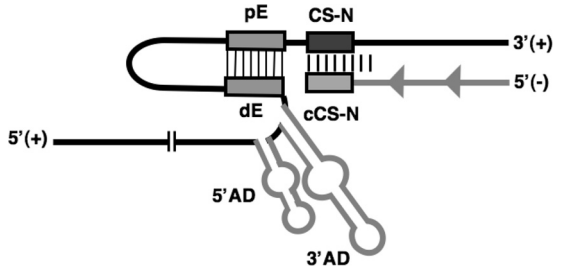

FIG. 10. Working model for transcriptional regulation of the N gene by the transcription-regulating motif. The TGEV genome is represented by a black line. The gray line below represents nascent minus-stranded RNA. RNA secondary structure prediction of the AD is represented by two gray hairpins. CS-N and cCS-N (complement of the CS-N) are represented by boxes. The 9-nt sequences of the proximal and distal elements are represented by boxes. The long-distance RNA-RNA interactions between the proximal and distal elements in the plus-stranded RNA are indicated by thin lines between the boxes. This interactions would relocate the AD in close proximity to the CS-N and contribute to stop the transcription complex during the synthesis of the minus-stranded RNA, therefore promoting the template switch to the TRS-L from the TRS-N. 5′AD, 5′ hairpin of the active domain; 3′AD, 3′ hairpin of the active domain.

The previously described proximal and distal elements were conserved in CoV α1 genus viruses (21). The active domain, a new component of the transcription-regulating motif, was also mainly conserved within this genus as evaluated by using the ClustalW2 program. The active domain is located in the same genome position (nt 136 to 332 of the gene M coding sequence) of genus α1 viruses TGEV, canine coronavirus (CCV), and feline infectious peritonitis virus (FIPV). Sequence similarities between the TGEV active domain and those of CCV and FIPV were 94% and 90%, respectively. The RNA secondary structure predictions of distal element 5′-flanking sequences in the deletion mutants showed that the constructs that had lost the enhancer activity (dE-45-158, dE-20-158, dE-6-158, and dE-103-20) did not include the 3′ hairpin of the active domain, while those that had preserved activity (dE-153-158, dE-133-158, and dE-113-158) did maintain this 3′ hairpin. These results confirmed that the active domain was required to enhance N gene transcription (Fig. 2). Based on Mfold predictions, the 3′ hairpin region of the active domain (113 nt) of TGEV is highly structured and is mainly conserved in members of CoV genus α1. Only two and three point mutations present in CCV and FIPV, respectively, disrupted base pairs in the RNA secondary structure compared to the TGEV active domain structure (data not shown). The phylogenetic conservation of transcription-regulating motif elements reinforces the relevance of this specific transcription-regulating mechanism. Interestingly, in CoV genus α1 viruses, the N gene was not located at the most 3′ position in the

genome, in contrast to other CoVs, in which the N gene is the last one in the genome, thus favoring production of the essential N protein. In α1 genus viruses, the most 3′ gene is gene 7, and a regulatory mechanism, such as the one described in this paper, was probably incorporated during evolution to improve N gene transcription (1, 10, 34).

Mutational analysis of the active domain showed that preserving RNA secondary structure and stability was not sufficient to maintain transcriptional activity and that the primary sequence was also necessary for its function. These results strongly suggested that the sequence of the active domain was required for a possible RNA-RNA or RNA-protein interaction essential to stop minus-stranded RNA synthesis, and promote the template switch. Further studies are being performed to identify the mechanism involved in this transcription enhancement. Transcription-regulating mechanisms similar to that described for the TGEV transcription-regulating motif have also been identified in other plus-stranded RNA viruses, such as tombusviruses. In tomato busy stunt virus (TBSV), the transcription of sgmRNA1 and sgmRNA2 is regulated by a long-distance RNA-RNA interaction (12). The complementarity between RNA sequences forms a higher-order RNA structure that modulates transcription (29). However, at variance with our results, in TBSV it has been demonstrated that a double-stranded RNA structure with a similar stability and an unrelated primary sequence can replace the native structure and maintain the transcription enhancement (29). These RNA structures would act as a physical barrier, stopping the transcription complex in tombusvirus, increasing the frequency of a premature termination of the minus-stranded RNA synthesis at a specific position (29).

Viral RNA synthesis relies on a diverse set of RNA sequences and structural motifs located throughout the genome (15). Knowledge of CoV genome tertiary structure is still incomplete, although the relevance of several short-distance RNA-RNA interactions involved in CoV RNA synthesis, such as the kissing-loop interaction within the 3′-UTR of bovine CoV (30) and mouse hepatitis virus, has been shown (8). For CoV discontinuous transcription, a long-distance interaction must occur between TRS-L and the TRS-B, most probably mediated by RNA-protein and RNA-RNA precomplexes bringing into proximity the TRS-L and each of the TRS-Bs (27). The regulatory role of long-distance RNA-RNA interactions is shown in many viral life cycles participating in translation, replication, genome circularization, or transcription (15, 20). The turnip crinkle virus (TCV) 3′-UTR is an interactive tertiary structure for which different RNA-RNA interactions cause changes that regulate the switch between translation and replication (32). In TBSV these sorts of interactions between a diverse array of sequences are involved in translation, replication, and transcription regulation and constitute a well-characterized viral system (31). The synthesis of sgmRNAs in different viral systems, and by distinct transcription mechanisms, frequently seems to involve long distance RNA-RNA interactions. In the potato virus X (PVX), with an internal initiation of transcription in a full-length minus-stranded genome, an octanucleotide sequence located within the 5′-UTR, interacts with central conserved sequences that precede each gene promoting sgmRNA synthesis (9). The mechanism of transcription by premature termination during the synthesis of the minus-stranded RNAs requires long-distance RNA-RNA interactions. As described above, this in *cis* RNA-RNA interaction operates in TBSV but also takes place during the expression of subgenomic RNA3 in Flock House virus (FHV) (14). An 11-nt sequence located just upstream of the RNA3 start site is complementary to another sequence mapping 1.5 kb upstream in the viral genome. The complementarity between these two sequence motifs, and not the primary sequence, was essential to increase transcription levels (14). Whereas, in red clover necrotic mosaic virus (RCNMV) an intermolecular interaction between two segments of the genome promotes the synthesis of the sgmRNA by premature termination (26). These examples and the novel results described for CoV illustrate the relevance of the long-distance RNA-RNA interactions and the diversity of viral processes in which they are involved.

ACKNOWLEDGMENTS

This work was supported by grants from the Ministry of Science and Innovation of Spain (BIO2007-60978 and PET2008-0310), the Community of Madrid (S-SAL-0185-2006), U.S. National Institutes of Health (ARRA-W000151845), and Pfizer Animal Health. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under the projects EMPERIE (EC grant agreement number 223498) and PoRRSCon (EC grant agreement number 245141). I.S. received a contract supported by grants from the Ministry of Science and Innovation of Spain (BIO2007-60978). P.A.M.-G. received a fellowship from the Ministry of Science and Innovation of Spain (BES-2008-001932).

We gratefully acknowledge C. M. Sánchez, M. González, and S. Ros for technical assistance.

REFERENCES

1. **Almazan, F., C. Galan, and L. Enjuanes.** 2004. The nucleoprotein is required for efficient coronavirus genome replication. J. Virol. **78:**12683–12688.
2. **Almazan, F., et al.** 2000. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. Proc. Natl. Acad. Sci. U. S. A. **97:**5516–5521.
3. **de Groot, R. J., et al.** 2010. Taxonomic structure of the Coronaviridae. *In* C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberg, and A. King (ed.), Virus taxonomy. International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA.
4. **Delmas, B., J. Gelfi, H. Sjöström, O. Noren, and H. Laude.** 1993. Further characterization of aminopeptidase-N as a receptor for coronaviruses. Adv. Exp. Med. Biol. **342:**293–298.
5. **Enjuanes, L., F. Almazan, I. Sola, and S. Zuniga.** 2006. Biochemical aspects of coronavirus replication and virus-host interaction. Annu. Rev. Microbiol. **60:**211–230.
6. **Enjuanes, L., et al.** 2008. The Nidovirales, p. 419–430. *In* B. W. J. Mahy, M. Van Regenmortel, P. Walker, and D. Majumder-Russell (ed.), Encyclopedia of virology, 3rd ed. Elsevier Ltd., Oxford, England.
7. **Frolov, I., et al.** 1996. Alphavirus-based expression vectors: strategies and applications. Proc. Natl. Acad. Sci. U. S. A. **93:**11371–11377.
8. **Goebel, S. J., B. Hsue, T. F. Dombrowski, and P. S. Masters.** 2004. Characterization of the RNA components of a putative molecular switch in the 3′ untranslated region of the murine coronavirus genome. J. Virol. **78:**669–682.
9. **Hu, B., N. Pillai-Nair, and C. Hemenway.** 2007. Long-distance RNA-RNA interactions between terminal elements and the same subset of internal elements on the potato virus X genome mediate minus- and plus-strand RNA synthesis. RNA **13:**267–280.
10. **Hurst, K. R., R. Ye, S. J. Goebel, P. Jayaraman, and P. S. Masters.** 2010. An interaction between the nucleocapsid protein and a component of the replicase-transcriptase complex is crucial for the infectivity of coronavirus genomic RNA. J. Virol. **84:**10276–10288.
11. **Jimenez, 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.
12. **Lin, H. X., and K. A. White.** 2004. A complex network of RNA-RNA interactions controls subgenomic mRNA transcription in a tombusvirus. EMBO J. **23:**3365–3374.
13. **Lin, H. X., W. Xu, and K. A. White.** 2007. A multicomponent RNA-based

control system regulates subgenomic mRNA transcription in a tombusvirus. J. Virol. **81:**2429–2439.
14. **Lindenbach, B. D., J. Y. Sgro, and P. Ahlquist.** 2002. Long-distance base pairing in flock house virus RNA1 regulates subgenomic RNA3 synthesis and RNA2 replication. J. Virol. **76:**3905–3919.
15. **Liu, Y., E. Wimmer, and A. V. Paul.** 2009. *cis*-Acting RNA elements in human and animal plus-strand RNA viruses. Biochim. Biophys. Acta **1789:** 495–517.
16. **Livak, K. J., and T. D. Schmittgen.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2(-\Delta\Delta C_T)$ method. Methods **25:**402–408.
17. **Masters, P. S.** 2006. The molecular biology of coronaviruses. Adv. Virus Res. **66:**193–292.
18. **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.
19. **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.
20. **Miller, W. A., and K. A. White.** 2006. Long-distance RNA-RNA interactions in plant virus gene expression and replication. Annu. Rev. Phytopathol. **44:**447–467.
21. **Moreno, J. L., S. Zuniga, L. Enjuanes, and I. Sola.** 2008. Identification of a coronavirus transcription enhancer. J. Virol. **82:**3882–3893.
22. **Pasternak, A. O., W. J. Spaan, and E. J. Snijder.** 2006. Nidovirus transcription: how to make sense? J. Gen. Virol. **87:**1403–1421.
23. **Penzes, Z., et al.** 2001. Complete genome sequence of transmissible gastroenteritis coronavirus PUR46-MAD clone and evolution of the Purdue virus cluster. Virus Genes **23:**105–118.
24. **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.
25. **Sawicki, S. G., and D. L. Sawicki.** 1995. Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands. Adv. Exp. Med. Biol. **380:**499–506.
26. **Sit, T. L., A. A. Vaewhongs, and S. A. Lommel.** 1998. RNA-mediated transactivation of transcription from a viral RNA. Science **281:**829–832.
27. **Sola, I., P. A. Mateos-Gomez, F. Almazan, S. Zuñiga, and L. Enjuanes.** 2011. RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. RNA Biol. **8:**237–248.
28. **Sola, I., J. L. Moreno, S. Zuñiga, S. Alonso, and L. Enjuanes.** 2005. Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. J. Virol. **79:**2506–2516.
29. **Wang, S., L. Mortazavi, and K. A. White.** 2008. Higher-order RNA structural requirements and small-molecule induction of tombusvirus subgenomic mRNA transcription. J. Virol. **82:**3864–3871.
30. **Williams, G. D., R.-Y. Chang, and D. A. Brian.** 1999. A phylogenetically conserved hairpin-type 3′ untranslated region pseudoknot functions in coronavirus RNA replication. J. Virol. **73:**8349–8355.
31. **Wu, B., et al.** 2009. A discontinuous RNA platform mediates RNA virus replication: building an integrated model for RNA-based regulation of viral processes. PLoS Pathog. **5:**e1000323.
32. **Yuan, X., K. Shi, A. Meskauskas, and A. E. Simon.** 2009. The 3′ end of Turnip crinkle virus contains a highly interactive structure including a translational enhancer that is disrupted by binding to the RNA-dependent RNA polymerase. RNA **15:**1849–1864.
33. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31:**3406–3415.
34. **Zuñiga, S., et al.** 2010. Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. J. Virol. **84:**2169–2175.
35. **Zuñiga, 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.