**Probing the sequence and structure of *in vitro* synthesized
antisense and target RNAs from the replication control system
of plasmid pMV158**

Celeste López-Aguilar and Gloria del Solar*

Molecular Microbiology and Infection Biology Department. Centro de
Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas.
Ramiro de Maeztu, 9, 28040 Madrid, Spain

*Corresponding author: Gloria del solar. Email-address: gdelsolar@cib.csic.es
Tel.: 34-91 837 3112; Fax: 34-91 536 0432;

Keywords: antisense RNA, target RNA, *in vitro* transcription, T7RNA
polymerase, RNA structure

ABSTRACT

Antisense RNAII is a replication control element encoded by promiscuous plasmid pMV158. RNAII binds to its complementary sequence in the *copG-repB* mRNA, thus inhibiting translation of the replication initiator *repB* gene. In order to initiate the biochemical characterization of the pMV158 antisense RNA-mediated control system, conditions for *in vitro* transcription by T7RNA polymerase were set up that yielded large amounts of antisense and target run-off products able to bind to each other. The run-off antisense transcript was expected, and confirmed, to span the entire RNAII as synthesized by the bacterial RNA polymerase, including the intrinsic transcription terminator at its 3'-terminus. On the other hand, two different target transcripts, mRNA₆₀ and mRNA₈₀, were produced, characterized and tested for efficient binding to the antisense product. The mRNA₆₀ and mRNA₈₀ run-off transcripts supposedly spanned 60 and 80 nucleotides, respectively, on the *copG-repB* mRNA and lacked terminator-like structures at their 3'-termini. Probing of the sequence and conformation of the main products, along with modeling of their secondary structures, showed that both target transcripts were actually longer-than-expected, and contained a 3'-terminal hairpin wherein the extra nucleotides base-paired to the expected 3'-terminus of the corresponding run-off transcript. These longer products were proposed to arise from the RNA-dependent polymerizing activity of T7RNA polymerase on correct run-off transcripts primed by extremely short 3'-selfcomplementarity. Seizing of the target mRNA sequence complementary to the 5'-terminus of RNAII in a stable 3'-terminal hairpin generated by this activity seemed to cause a 3-fold decrease in the efficiency of binding to the antisense RNA.

1. Introduction

Understanding the mechanistic basis that underlie inter- or intramolecular RNA interactions requires precise information on the actual sequence and dynamic secondary structure of the interacting RNA molecules or regions. Chemical or enzymatic probing of paired and unpaired RNA regions provides useful partial information on the most populated structural configurations, although fails to reveal the actual base pairing patterns. Therefore, the establishment of accurate RNA secondary structure models requires the combination of probing experiments and computational predictions (Washietl et al., 2012). *In vitro* transcription by the RNA polymerase (RNAP) of the bacteriophage T7 (T7RNAP) is a general choice method to produce large amounts of RNA for biochemical and functional analysis, structural studies and therapeutic applications (Beckert and Masquida, 2011; Kao et al., 1999; Moll et al., 2004). RNAs synthesized by T7RNAP during *in vitro* run-off transcription consist of the expected full-length (N, N+1 and N+2) RNA species accompanied by variable percentages of abortive transcripts as well as longer-than-expected products. These longer products have been reported to arise from the enzyme's RNA-dependent RNA polymerase activity elongating correct transcripts that have a self-complementary 3'-end and, therefore, can form inter- or intramolecular primed RNA templates (Cazenave and Uhlenbeck, 1994; Triana-Alonso et al., 1995). Longer-than-expected transcripts may represent even more than 70% of the total products of *in vitro* run-off transcription with T7RNAP (Triana-Alonso et al., 1995), hence compromising the identification and purification of the desired RNA. Summarizing, two main sources of uncertainty may arise when studying RNA-RNA interactions based on the predicted structure of transcripts synthesized *in vitro* by T7RNAP. One is the accuracy of the prediction itself, the other being the sequence (and hence, the structure) of the 3'-end of the transcripts.

Control of the replication, and hence of the copy number, of rolling-circle-replicating (RCR) plasmids is exerted by countertranscribed RNAs (ctRNAs) that, upon binding to the *rep* mRNA, inhibit expression of the replication initiator *rep* gene by different mechanisms (Alonso and Tailor, 1987; Kumar and Novick, 1985; Novick et al., 1989). Promiscuous RCR plasmid pMV158, and by

extension the members of the pMV158 replicon family, are singular in having a dual mechanism of replication control consisting of a transcriptional repressor protein and a ctRNA, both of which inhibit expression of the *rep* gene (del Solar et al., 1995; del Solar and Espinosa, 1992; del Solar et al., 1993). The ctRNA RNAII encoded by pMV158 binds to its complementary sequence in the *copG-repB* mRNA, therefore preventing translation of the replication initiator *repB* gene (del Solar et al., 1997; del Solar and Espinosa, 1992). The RNAII 5'-end has been determined by primer extension of total RNA isolated from *Streptococcus pneumoniae* cells, whereas the 3'-end has been mapped by *Escherichia coli* RNAP-mediated *in vitro* transcription experiments (del Solar et al., 1995; del Solar and Espinosa, 2001). The predicted structure of inferred 48-nt-long RNAII consists of single-stranded 5' and 3' tails and a hairpin (del Solar and Espinosa, 1992). The single hairpin together with the adjacent U-rich 3' tail compose a very efficient intrinsic transcription terminator (del Solar and Espinosa, 2001). We aim to characterize binding of antisense RNAII to its target in the *copG-repB* mRNA. To this end, we have set-up the experimental conditions that yielded sufficient amounts of the antisense RNA and of the target segments of the mRNA by T7RNAP-mediated *in vitro* transcription, and we have checked whether the sequence and structure of the obtained transcripts matched those expected. We have found that full-length run-off transcripts were obtained when synthesizing the antisense RNAs, a case where the 5'-end of the template DNA encodes an intrinsic terminator. In contrast, longer-than-expected products were obtained for the target RNA species, whose expected 3'-terminal regions were predicted to be either unpaired or forming non-terminator-like structures. In all cases, provided that the actual sequence of the transcript 3'-end was known, the *in silico* predicted structure matched quite well that observed.

2. Materials and methods

2.1. DNA templates

Oligonucleotides (oligos) used to produce DNA templates for *in vitro* transcription were chemically synthesized and then purified from 12%

polyacrylamide (PAA) sequencing gels. DNA templates were generated by annealing and extension of oligos P-RNII (5'-GGAGGCCGGAGAATTGTAATACG-3') and RNII (5'-ATAAAAAAGCCGTGCTGGCAGGCACTGGCTAAAGTCAAACATTTCTTCCTATAGTGAGTCGTATTACAATTCTCCGGCCTCC-3') for the sequence encoding the antisense RNA, oligos mRNA (5'-GGAGGCCGGAGAATTGTAATACGACTCACTATAGGATAAAAAAGCCGTGCTGGCAGGCAC-3') and mRNA-60 (5'-TATAATATACCCAAGAAATGTTTGACTTTAGCCAGTGCCTGCCAGCACGGC-3') for the DNA encoding mRNA₆₀, and oligos mRNA and mRNA-80 (5'-CTTTTCTTTAGCCATAAAGTATAATATACCCAAGAAATGTTTGACTTTAGCCAGTGCCTGCCAGCACGGC-3') for the DNA encoding mRNA₈₀. Then, the template DNAs were PCR-amplified using the primer pairs P-RNII and P2-RNII (5'-ATAAAAAAGCCGTGCTGGCA-3') for the RNII template, P-RNII and mRNA-60R (5'-TATAATATACCCAAGAAATGTTTGAC-3') for the mRNA₆₀ template, and P-RNII and mRNA-80R (5'-CTTTTCTTTAGCCATAAAGTAT-3') as the primer pair for the mRNA₈₀ template. The PCR primer sets were chosen to generate DNA fragments harboring a functional T7RNAP promoter upstream of the sequences that encode the respective RNAs. To increase the yield of transcripts, two G residues were encoded at the transcription start sites (Moll et al., 2004). All amplified DNA fragments were purified from PAA gels before being used as templates for synthesis of the RNA molecules.

2.2. RNA synthesis

The *in vitro* run-off transcription assays were performed using T7RNAP (Roche) according to the manufacturer's protocol. Various concentrations of template DNA in the range of 2 to 30 ng/μl were tested. Optimal yield of the transcripts was obtained when ~4 ng/μl of template was used. After transcription, the template DNAs were removed by digestion with RNase-free DNase I (Boehringer). Samples were cleaned by filtration through Sephadex G-25 columns (GE-Healthcare) and successive extraction with phenol/chloroform and chloroform, followed by ethanol precipitation. RNAs were quantified from their absorbance at 260 nm in a NanoDrop™ spectrophotometer.

2.3. 5'-end labeling of RNA

RNAs (15 pmol) were dephosphorylated using calf intestinal alkaline phosphatase (Promega) following the manufacturer's indications. Then, the longest main RNA species were isolated from 12% PAA sequencing gels, extracted with phenol/chloroform and chloroform, and ethanol-precipitated, previous to the labeling of their 5' end with T4 polynucleotide kinase (PNK) (New England Biolabs) and [γ - 32 P]ATP according to the enzyme protocol. Labeled RNAs were re-isolated from 12% PAA sequencing gels as above.

2.4. RNA-RNA interaction assays

2.4.1. RNA-RNA binding

Previous to the binding reaction, the antisense and target RNAs were incubated separately at 65°C for 10 min and then at 37°C for 10 min, in order to obtain their native conformation. The binding mixtures (10 μ l) contained 0.2 nM of the 5'-end-labeled antisense RNA and various concentrations (0 to 150 nM) of the unlabeled target RNA (mRNA₆₀ or mRNA₈₀) in TMN buffer (100 mM NaCl, 20 mM Tris-OAc, 10 mM Mg(OAc)₂; (Persson et al., 1988). Binding was allowed for 30 min at 37°C, and terminated by adding an equal volume of loading buffer (glycerol 30%, tRNA 4%, xylene cyanol 0.4% and bromophenol blue 0.4%, in TMN buffer) and placing the mixture on ice. The samples were loaded on native 10% PAA gels. Electrophoresis was performed at 300 V for 2 h. Radioactive bands were visualized and quantified from fixed and dried gels using the storage phosphor technology, with the aid of a FLA-3000 (FUJIFILM) imaging system and the Quantity One software (Bio-Rad).

2.4.2. Time-course analysis of the RNA-RNA binding

Standard binding mixtures containing 0.2 nM of the 5'-end-labeled antisense RNA and 150 nM of the unlabeled target RNA in TMN buffer were prepared as above, and incubated at 37°C. Samples were withdrawn at different times within the first min of incubation, and the reaction was stopped by diluting the samples into an equal volume of gel application buffer (94%

formamide, 17 mM EDTA, 0,025% xylene cyanol and 0,025% bromophenol blue) on ethanol-dry ice. Samples were applied to polyacrylamide gels and analyzed as described above.

2.5. Partial enzymatic digests and chemical cleavage for structure analysis

Previous to any chemical or enzymatic treatment, RNAll, mRNA₆₀ and mRNA₈₀ were incubated separately at 65°C for 10 min and then at 37°C for 10 more min, in order to obtain their native conformation. The cleavage mixtures (10 µl) contained 1 µg of tRNA as carrier, and 25 femtomoles of any of the 5'-end-labeled RNA in TMN buffer.

Partial digestion reactions were performed with RNA-Grade RNases (including RNase A, T1 and V1) from Ambion. Incubations were carried out in TMN buffer at 37°C for 5 min with 0.01 pg/µl RNase A or 0.00014 U/µl RNase T1, or for 30 min with 0.01 U/µl RNase V1. To obtain the T1 ladder, incubation with 0.01 U/µl of this enzyme was carried out at 55°C in RNA-Sequencing buffer (provided by Ambion). In order to produce the V1 ladder, the 3'-phosphate of the T1 ladder products was removed by treatment with T4 PNK as described (Brown and Bevilacqua, 2005).

Alkaline ladders were generated by incubation of 25 femtomoles of each RNA at 95°C for 5 min in 5 µl of Alkaline-Hydrolysis buffer (supplemented with 1 µg of tRNA) according to the Ambion's protocol.

All reactions were stopped by the addition of 20 µl Inactivation/Precipitation buffer (Ambion) (as described by the manufacturers). The RNA samples were precipitated and dissolved in 2 µl of sterile water plus 5 µl of loading buffer (95% formamide, 18 mM EDTA, 0.025% each of SDS, xylene cyanol and bromophenol blue).

When cleavage was induced by Pb²⁺ ions, 3 mM lead acetate (Sigma Aldrich) was used at 37°C for 1 min. Reactions were stopped by adding 100 mM EDTA, and the RNA samples were precipitated and treated as described previously for enzymatic digestions. A control of untreated RNA was always included to detect unspecific cuts on the RNA samples.

To detect the cleavage positions, the samples were denatured at 90°C for 3 min, and products were separated on high resolution denaturing 15% PAA gels. Electrophoresis was carried out in TBE buffer at a constant 45 W for 2 h. To determine the structure near the 3'-end of RNA, 8% PAA gels were used and the running time was extended to 4-5 h. Cleavage positions were identified by running in parallel RNase T1, RNase V1 and alkaline ladders. Chemical and enzymatic cleavage products other than those arising from digestion with RNase V1 contain 3'-phosphate ends, and hence their electrophoretic mobility was referred to the T1 ladder. On the other hand, the V1 ladder was used as reference for migration of the 3'-OH products arising from RNase V1 digestion. After electrophoresis, gels were dried and images of radioactive bands were visualized and quantified using the storage phosphor technology, with the aid of a FLA-3000 (FUJIFILM) imaging system and the Quantity One software (Bio-Rad).

The results of the chemical and enzymatic cleavage of the RNAs were interpreted according to the reported sequence- and configuration-specificities of the probing agents. Pb^{2+} cleaves preferentially after pyrimidine residues within single-stranded regions, loops and bulges (Chevalier et al., 2009; Gornicki et al., 1989; Sobczak et al., 2010). RNase V1 is known to cut double-stranded and stacked single-stranded regions of the RNA molecules (Chevalier et al., 2009; Lockard and Kumar, 1981), although recently it has been reported to recognize mainly U residues in both single- and double-stranded RNAs (Sobczak et al., 2010). Finally, single-strand-specific RNases T1 and A cleave after unpaired G and pyrimidine residues, respectively (Chevalier et al., 2009).

2.6. Construction of mRNA₆₀ and mRNA₈₀ cDNA libraries

2.6.1. 3'-RACE-PCR

To analyze the frequency of the different 3'-extended mRNA₆₀ and mRNA₈₀ species, a rapid amplification of cDNA ends was performed. The first stage was the production of cDNAs with a known "anchor primer" sequence incorporated at the 3'-end. For this purpose, dephosphorylated RNAs molecules (see above) were relaxed at 65°C for 5 min, and then ligation was performed in a final volume of 20 µl by incubation with T4 RNA ligase (Promega) at 4°C

overnight. The 5'-adenylated oligo (rApp-CTGTAGGCACCATCAAT-idT) (Integrated DNA technologies) was used as 3'-adapter to avoid circularization or multimerization of the RNAs in the pool (Donaire et al., 2009). Ligation products were purified by phenol-chloroform extraction and then cDNAs were obtained by reverse transcription with SuperScript II RT (Invitrogen) according to the manufacturer's protocol. The second stage is a PCR step with AmpliTaq-Gold polymerase (applied Biosystems) to amplify specific sequences from the pool of cDNAs using the gene specific primer 3'-RACE mRNA-F (5'-GGATAAAAAAAGCCGTGCT-3') and the anchor primer 3'-RACE mRNA-R (5'-ATTGATGGTGCTACAG-3') that anneals at the 3'-end incorporated in the first stage. The PCR products were re-amplified using the same PCR conditions. Gene-specific 3'-RACE products were purified from 10% PAA gels as described previously for RNA.

2.6.2. Cloning of 3'-RACE products

Purified DNA molecules were cloned into pGEM-T Easy Vector (Promega) under conditions described by the manufacturers. A blue-white screen in the presence of X-gal plates allowed the detection of successful ligations. Recombinant plasmid DNA was obtained from individual transformant colonies and automated sequenced using M13F and M13R commercialized primers.

2.7. RNA secondary structure prediction

RNA structures were predicted by using the Mfold (3.2) software (<http://mfold.bioinfo.rpi.edu>; (Zuker, 2003)).

3. Results and discussion

3.1. Designing and production of antisense and target transcripts

Run-off transcripts mimicking pMV158-encoded antisense RNAII as well as two target RNAs that include its complementary sequence on the *copG-repB* mRNA (Fig. 1) were synthesized *in vitro* using T7RNAP and double-stranded

DNA templates (see Materials and methods). To obtain high yields of the transcripts, two extra G residues were encoded at the 5'-end of the RNA products (Figs. 2, 3 and 4). *In vitro* synthesized RNAII was expected to consist otherwise of the presumed nucleotide sequence of the antisense RNA, from the 5'-end determined for pneumococcal RNAII (del Solar et al., 1995) to the 3'-end mapped by *E. coli* RNAP-mediated transcription (del Solar and Espinosa, 2001) (Figs. 1 and 2A). In turn, the *in vitro* synthesized mRNA₆₀ and mRNA₈₀ targets supposedly spanned 60 and 80 nucleotides, respectively, on the *copG-repB* mRNA, starting at the position 5'-adjacent to the *copG* gene stop codon (Fig. 1). The theoretical 3'-end of mRNA₆₀ was located 5 nt upstream from the start codon of the *repB* gene, whereas that of mRNA₈₀ corresponded to the first position of codon 6 of *repB*. Therefore, both mRNA₆₀ and mRNA₈₀ comprised the entire sequence complementary to antisense RNAII (Figs. 1, 3A and 4A). We chose to produce transcripts of the *copG-repB* intergenic region as targets of RNAII because their structure might best resemble that of the actual target *in vivo*, where the absence of translation through this region would allow the local secondary structures to persist in the growing mRNA.

3.2. *In silico* prediction of the secondary structure of the expected antisense and target run-off transcripts

The predicted structure of the expected antisense RNA synthesized *in vitro* matched that previously reported for biological RNAII (del Solar and Espinosa, 1992; del Solar and Espinosa, 2001). It consisted of a 5' AU-rich tail followed by a GC-rich stable hairpin and a 3' poly(U) tail, the latter two elements constituting the intrinsic transcription terminator (Fig. 2A).

On the other hand, the run-off target transcripts, mRNA₆₀ and mRNA₈₀, were predicted to share the structure of their common 5'-moieties and to differ in their 3'-termini (Figs. 3A and 4A). The predicted structure of both mRNA targets consisted of a hairpin complementary to the single one in RNAII preceded by an A-rich 5' tail. Whereas the 3'-terminus of mRNA₆₀ was predicted to be exclusively single-stranded, that of mRNA₈₀ theoretically contained a rather long although weakly base-paired hairpin ($\Delta G = -6.0$) involving the sequence complementary to the 5'-terminus of RNAII (Figs. 3A and 4A).

3.3. Chemical and enzymatic probing of the structure of *in vitro* synthesized antisense and target RNAs

The structure in solution of *in vitro* synthesized antisense and target RNAs was mapped by chemical and enzymatic partial cleavage of the corresponding 5'-end-labeled transcripts. The main RNA product of each transcription reaction, which in all cases comprised more than 80% of the total transcripts, was gel-purified both before and after 5'-end labeling. The purified RNA species were treated with Pb^{2+} ions, or with RNase T1, A or V1, under experimental conditions that yielded an average of one cut per RNA molecule.

Taking into account the RNA sequence- and configuration-specificities of the chemical and enzymatic probing agents (see Materials and methods), the results obtained allowed us to confirm that an antisense RNA of the expected sequence and the predicted structure was the main product of *in vitro* transcription (Fig. 2). Thus, the structure in solution of pMV158-encoded antisense RNAII consists of a single hairpin with a 6-nt terminal loop and a small bulge, which is flanked by an AU-rich 5' tail and a U-rich 3' tail. Given the small size of RNAII, verification of the secondary structure model predicted with the Mfold program (Zuker, 2003) was strongly expected.

Probing of the sequence and structure of *in vitro* synthesized mRNA₆₀ and mRNA₈₀ revealed that the purified products were populated with 3'-overextended transcripts that were up to 10 and 8 nucleotides longer than expected, respectively (Figs. 3B and 4B). The structure probed for the 5'-moiety of either target RNA fitted quite well the predicted model, and confirmed the formation of a hairpin complementary to that of the antisense RNA, which was preceded by an A-rich single-stranded 5' tail that was the counterpart of the U-tract of the RNAII intrinsic terminator. In contrast, the structure inferred for the 3'-region of either target transcript diverged substantially from the predicted one, and showed a double-stranded 3'-terminus that involved all the extra nucleotides of each actual RNA (Figs. 3B and 4B). Notably, nucleotides 45 to 57 of mRNA₈₀, which were located in the 5'-arm of a hypothetical hairpin in the expected transcript (Fig. 4A), appeared as single-stranded in the actual RNA product (Fig. 4B and C). The contrary was seen in mRNA₆₀, wherein

nucleotides 45 to 50, which were expected to be unpaired (Fig. 3A), proved to be double-stranded (Fig. 3B and C). It is worth noting that the sequence between residues 45 and 50, whose conformation in either of the actual target transcripts differed from that expected, is complementary to the 5'-end of antisense RNAII (Figs. 3 and 4).

3.4. Analysis of the self-priming mechanism for longer-than-expected product synthesis and structure of the actual target transcripts

To gain information on the variety and relative abundance of the target RNA species arising from *in vitro* transcription, the purified main products were characterized by cloning and sequencing of the 3'-RACE products (see Materials and methods).

Nucleotide sequence analysis of the 3'-RACE products cloned in 10 individual transformants of the mRNA₈₀ cDNA library indicated that the vast majority of the transcripts would contain the 3'-extra sequence 5'-CCAUAAA-3', which could be inferred from every colony. Moreover, about 50% of the transcripts would also contain an additional G at the 3'-end, as inferred from the sequence of 5 out of the 10 colonies analyzed (Table 1). The length of this extra sequence (7-8 nucleotides) was consistent with the 3'-extension found by chemical and enzymatic probing of the mRNA₈₀ main products (Fig. 4B). These longer-than-expected products resulted most likely from the known RNA-dependent RNA polymerase activity of T7RNAP on inter- or intramolecular self-primed run-off transcripts of the correct size (Cazenave and Uhlenbeck, 1994; Triana-Alonso et al., 1995). Taking into account the sequence of the expected mRNA₈₀ transcript, we propose two alternative self-priming processes, either by intramolecular hairpin or by intermolecular dimer formation, leading to the 3'-overextended sequence found in the actual *in vitro* transcription products (Table 1).

The nucleotide sequence of the 3'-RACE products derived from mRNA₆₀ was analyzed in 25 individual clones of the corresponding cDNA library (Table 1). About 30% (8 out of 25) of the colonies arose from the correct run-off transcript without any 3'-extra sequence. The rest of the colonies arose from transcripts harboring one (13 out of 25) or two (4 out of 25) extra C at their 3'-

end (Table 1). These results contrasted with those obtained in the experiments of direct sequence and structure probing of the mRNA₆₀ transcripts, where the main products were populated with 10-11-nt longer-than-expected RNA species (Fig. 3B). These differences are unlikely to arise from discarding of the shortest RNA species during the purification that followed labeling of the transcription products (a missing step when performing RACE and subsequent cloning), since the purified transcripts represented the vast majority of the total products. It rather seems that non- or barely-extended mRNA₆₀ transcripts might be overrepresented among the templates for 3'-RACE, most likely due to the previously reported RNA ligase-dependent bias in small RNA representation in cDNA libraries (Hafner et al., 2011). Be that as it may, alternative inter- or intramolecular self-primed mRNA₆₀ template molecules can be proposed based on both the sequence of the 3'-end of the correct run-off transcript (Fig. 3A) and the first two extra residues suggested to be added according to the cDNA library sequencing (Table 1). Elongation by T7RNAP of either self-primed mRNA₆₀ template for 10-11 extra nucleotides would give rise to a longer-than-expected transcript of the observed size having a G residue in position 68, as supported by enzymatic probing of its sequence (see T1 ladder in Fig. 3B).

Computer predictions of the secondary structure of the inferred actual mRNA₆₀ and mRNA₈₀ transcripts were supported by chemical and enzymatic probing; that is, all the predicted single- and double-stranded sequences of either target mRNA were actually confirmed by structure probing (Figs. 3B, 3C, 4B and 4C). Hence, mRNA₆₀ and mRNA₈₀ main species were each shown to finish in a hairpin generated by pairing between the extra nucleotides and those located in the expected 3'-terminus of the correct run-off transcript (Figs. 3C and 4C).

The results obtained throughout this work from *in vitro* transcription experiments aimed to yield significant amounts of the antisense and target RNAs of the pMV158 replication control system, show that run-off transcripts of the expected size were only synthesized when the 5'-terminus of the template DNA included an intrinsic terminator, as was the case with antisense RNAII. The expected transcripts were also obtained when synthesizing antisense-RNA mutants that, in spite of having sequence changes in either the stem or the loop of the terminator hairpin, maintained a stable secondary structure (not shown).

Contrarily, transcripts lacking an intrinsic terminator structure at their 3'-end were over-elongated by the RNA-dependent RNA polymerase activity of T7RNAP upon self-priming of the correct run-off products. The presence of a long although weakly paired hairpin in the 3'-terminal region of the correct mRNA₈₀ transcript (Fig. 4A) did not prevent generation of the longer-than-expected products. These results agree with previous observations showing that only primed RNA templates that cannot form stable secondary structures at the 3'-end are converted into longer-than-expected products, whereas those forming stable hairpins are not (Triana-Alonso et al., 1995).

Although longer-than-expected transcripts have been reported to be especially abundant when transcribing short DNA templates (Nacheva and Berzal-Herranz, 2003), *in vitro* synthesis of the medium-size mRNA₆₀ and mRNA₈₀ transcripts is shown here to yield mainly 8-10-nt longer-than-expected products. The experimental conditions for T7RNAP-catalyzed synthesis of antisense RNA, mRNA₆₀ and mRNA₈₀ were those recommended by the enzyme manufacturer, which included the use of 1 mM final concentration of each NTP. This is also the NTP concentration recommended for preparing RNAs aimed at a variety of applications, including biochemical analysis (Beckert and Masquida, 2011). Therefore, the concentration of UTP in all these transcription assays was within the 0.2 to 1.0 mM range found by Triana-Alonso and coworkers to produce the highest yield and fraction of correct products, at least in the presence of 3.5-4 mM of the other NTP (Triana-Alonso et al., 1995). Decreasing the concentration of T7RNAP used in the transcription assays resulted in a drastic reduction in the total yield of products, which were not characterized further (not shown).

Two alternative mechanisms have been invoked to produce the 3'-self-complementary primed RNA templates that lead to over-elongated products: intermolecular dimer or intramolecular hairpin formation, and the results of different reports favored either hypothesis (Cazenave and Uhlenbeck, 1994; Nacheva and Berzal-Herranz, 2003; Triana-Alonso et al., 1995). Very little self-complementarity (2-4 base pairs) between residues located at the 3'-end and 1-8 nts apart from it seems to suffice for the mRNA₆₀ and mRNA₈₀ transcripts to form primed templates that will be elongated by T7RNAP (Table 1). Interestingly, the 3'-terminal AG of the correct mRNA₈₀ run-off transcript is

complementary to two additional CU dinucleotides that are farther from the 3'-end than the one used to generate the primed template (Fig. 4C). However, mRNA₈₀ priming by these alternative base-pairings, if occurs, does not succeed in yielding the longer-than-expected products. The fact that, in both mRNA₆₀ and mRNA₈₀, priming is achieved by as few as 2-4 base pairs between residues at the very 3'-end and a few nucleotides away rather suggests that the process occurs via intramolecular hairpin, so that complementary regions spaced by a small 'loop' cause the transcript to fold back on itself forming a short hairpin. Elongation of the primed RNA templates by T7RNAP would convert these short hairpins into longer stable secondary structures with GC-rich upper stem regions and AU-rich lower stems (Figs. 3C and 4C). These terminator-like structures would promote disruption of the elongation complexes (Triana-Alonso et al., 1995), hence leading to the release of the longer-than-expected target transcripts.

We have found that generation of the longer-than-expected mRNA₆₀ and mRNA₈₀ products was the main source of unpredictability in assigning the actual sequence and structure of the run-off transcripts synthesized *in vitro* by T7RNAP. In fact, once the sequence of the actual transcripts was provided, the Mfold program accurately predicted their structure (Figs. 3C and 4C).

3.5. Binding of *in vitro* synthesized antisense and target RNAs of the pMV158 replication control system

Due to the presence of the extra sequence, the region complementary to the 5' tail of RNAII at the 3'-moiety of the actual mRNA₈₀ transcript was free of intramolecular pairing (Fig. 4C), whereas that of actual mRNA₆₀ was partially base-paired to the extra nucleotides, forming the AU-rich lower stem of a rather stable hairpin ($\Delta G = -15.6$; Fig. 3C). As it has been previously mentioned, this contrasted with the structure predicted for the respective expected target transcripts (Figs. 3 and 4).

Since the accessibility of the mRNA target regions complementary to the single-stranded sequences of RNAII might be critical for binding of both RNAs, it was interesting to test the efficiency of binding of the antisense RNA to each target transcript. This was performed by determining both the rate of RNA-RNA

complex formation and the target concentration that yields half-maximal binding of antisense RNAII under the experimental conditions employed (see Materials and methods). The results (Fig. 5) showed that both mRNA targets could bind efficiently to the antisense RNA, irrespective of whether the sequence complementary to the 5' tail of RNAII was totally or partially single-stranded. Therefore, the presence in mRNA₆₀ of the 3'-terminal hairpin generated by the over-extension of the run-off transcript is shown here not to preclude RNAII binding. It should be noted, however, that mRNA₈₀ is about three times more efficient than mRNA₆₀ in binding to RNAII, as shown by the lower concentration of free mRNA₈₀ required to reach half-maximal binding of the antisense RNA, and by the higher association rate constant obtained when this target is used (Fig. 5C). This moderate difference in binding efficiency suggests that seizing of the sequence complementary to the 5'-terminus of RNAII in a stable structure of the target RNA (Fig. 3C) may impair to some degree the RNA-RNA interaction. Be it as it may, the *copG-repB* mRNA sequence complementary to the 5'-moiety of RNAII is not predicted to be involved in such a stable hairpin either in the growing or in the complete mRNA molecule (not shown), and the less stable stem-loop formed temporarily in the theoretical mRNA₈₀ would be unfolded by the ribosomes reading *repB*.

4. Conclusions

Folding models for antisense RNAII and its mRNA₆₀ and mRNA₈₀ targets were established by combining computational predictions of thermodynamically stable RNA secondary structures, sequence and structure probing of the transcripts, and cDNA library sequencing analyses.

In vitro synthesis of large amounts of these medium-size RNAs by T7RNAP-mediated run-off transcription resulted in full-length or longer-than-expected products depending on whether the expected transcripts harbored, or not, intrinsic terminator-like structures, respectively. Longer-than-expected products finished at a 3'-terminal quite stable hairpin, which was generated during extension of self-primed run-off transcripts by T7RNAP, and resembled intrinsic transcription terminators. Providing that the actual sequence of the transcript products was known, the Mfold program successfully predicted the

secondary structure of the RNAs. Seizing of target RNA sequences complementary to the 5'-terminus of RNAII in stable hairpins generated by over-extension of the transcripts seemed to impair binding to the antisense RNA to some extent, although did not prevent the interaction between both RNAs.

Acknowledgments

We are very grateful to Livia Donaire and César Llave for their help in 3'-RACE. We also thank J.A. Ruiz-Masó and Lorena Bordanaba for critical reading of the manuscript. This study was supported by the Spanish *Ministerio de Ciencia e Innovación* (Grant BFU2010-19597, PNEUMOTALK to GdS).

References

- Alonso, J. C., Tailor, R. H., 1987. Initiation of plasmid pC194 replication and its control in *Bacillus subtilis*. *Mol. Gen. Genet.* 210, 476-484.
- Beckert, B., Masquida, B., 2011. Synthesis of RNA by in vitro transcription. *Methods Mol. Biol.* 703, 29-41.
- Brown, T. S., Bevilacqua, P. C., 2005. Method for assigning double-stranded RNA structures. *BioTechniques*. 38, 368-372.
- Cazenave, C., Uhlenbeck, O. C., 1994. RNA template-directed RNA synthesis by T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA.* 91, 6972-6976.
- Chevalier, C., Geissmann, T., Helfer, A. C., Romby, P., 2009. Probing mRNA structure and sRNA-mRNA interactions in bacteria using enzymes and lead(II). *Methods Mol. Biol.* 540, 215-232.
- del Solar, G., Acebo, P., Espinosa, M., 1995. Replication control of plasmid pLS1: efficient regulation of plasmid copy number is exerted by the combined action of two plasmid components, CopG and RNA II. *Mol. Microbiol.* 18, 913-924.
- del Solar, G., Acebo, P., Espinosa, M., 1997. Replication control of plasmid pLS1: the antisense RNA II and the compact *rnaII* region are involved in translational regulation of the initiator RepB synthesis. *Mol. Microbiol.* 23, 95-108.
- del Solar, G., Espinosa, M., 1992. The copy number of plasmid pLS1 is regulated by two trans-acting plasmid products: the antisense RNA II and the repressor protein, RepA. *Mol. Microbiol.* 6, 83-94.
- del Solar, G., Espinosa, M., 2001. *In vitro* analysis of the terminator T_{II} of the inhibitor antisense *rna II* gene from plasmid pMV158. *Plasmid.* 45, 75-87.
- del Solar, G., Moscoso, M., Espinosa, M., 1993. Rolling circle-replicating plasmids from Gram-positive and Gram-negative bacteria: a wall falls. *Mol. Microbiol.* 8, 789-796.
- Donaire, L., Wang, Y., González-Ibeas, D., Mayer, K. F., Aranda, M. A., et al., 2009. Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology.* 392, 203-2014.

- Gornicki, P., Baudin, F., Romby, P., Wiewiorowski, M., Kryzosiak, W., et al., 1989. Use of lead(II) to probe the structure of large RNAs. Conformation of the 3' terminal domain of *E. coli* 16S rRNA and its involvement in building the tRNA binding sites. *J. Biomol. Struct. Dyn.* 6, 971-984.
- Hafner, M., Renwick, N., Brown, M., Mihailovic, A., Holoch, D., et al., 2011. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA*. 17, 1697-1712.
- Kao, C., Zheng, M., Rüdisser, S., 1999. A simple and efficient method to reduce nontemplated nucleotide addition at the 3' terminus of RNAs transcribed by T7 RNA polymerase. *RNA*. 5, 1268-1272.
- Kumar, C. C., Novick, R. P., 1985. Plasmid pT181 replication is regulated by two countertranscripts. *Proc. Natl. Acad. Sci. USA*. 82, 638-642.
- Lockard, R. E., Kumar, A., 1981. Mapping tRNA structure in solution using double-strand-specific ribonuclease V1 from cobra venom. *Nucleic Acids Res.* 9, 5125-5140.
- Moll, P. R., Duschl, J., Richter, K., 2004. Optimized RNA amplification using T7-RNA-polymerase based *in vitro* transcription. *Anal. Biochem.* 334, 164-174.
- Nacheva, G. A., Berzal-Herranz, A., 2003. Preventing undesired RNA-primed RNA extension catalyzed by T7 RNA polymerase. *Eur. J. Biochem.* 270, 1458-1465.
- Novick, R. P., Iordanescu, S., Projan, S. J., Kornblum, J., Edelman, I., 1989. pT181 plasmid replication is regulated by a countertranscript-driven transcriptional attenuator. *Cell*. 59, 395-404.
- Persson, C., Wagner, E. G. H., Nordström, K., 1988. Control of replication of plasmid R1: kinetics of *in vitro* interaction between the antisense RNA, CopA, and its target, CopT. *EMBO J.* 7, 3279-3288.
- Sobczak, K., Michlewski, G., de Mezer, M., Krol, J., Krzyzosiak, W. J., 2010. Trinucleotide repeat system for sequence specificity analysis of RNA structure. *Anal. Biochem.* 402, 40-46.
- Triana-Alonso, F. J., Dabrowski, M., Wadzack, J., Nierhaus, K. H., 1995. Self-coded 3'-extension of run-off transcripts produces aberrant products during *in vitro* transcription with T7 RNA polymerase. *J. Biol. Chem.* 270, 6298-6307.
- Washietl, S., Hofacker, I. L., Stadler, P. F., Kellis, M., 2012. RNA folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction. *Nucleic Acids Res.* 40, 4261-4272.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406-3415.

FIGURE LEGENDS

Figure 1. Replication control region of pMV158. Schematic representation of the *copG-repB* operon transcribed from promoter P_{cr} . The nucleotide sequence of the region encompassing the end of *copG*, the entire *rnalI* and the beginning of *repB*, is shown amplified below. The relevant features of this region, namely the *copG* termination codon, the *repB* initiation codon, and the -35 and -10 boxes of promoter P_{ctlI} are underlined. The coding sequence of *rnalI* is shown in italics. Initiation site of antisense RNAII is indicated by an arrow pointing in the direction of the transcription. The palindrome sequence corresponding to the transcription terminator (TII) hairpin of RNAII is indicated by opposite arrows below the sequence.

Figure 2. Structure of *in vitro* synthesized RNAII. (A) Sequence and RNA folding prediction of the expected antisense run-off transcript by the Mfold program. The different regions are color-coded, and the RNA configuration is displayed as follows: continuous line, single-stranded sequence (ss); dotted line, double-stranded sequence (ds). Two Gs (labeled with +) were added at the 5'-end of the RNA sequence for efficient *in vitro* transcription. (B) Structure probing of antisense RNA by enzymatic and chemical cleavage. C: untreated RNA; OHL: alkaline hydrolysis ladder; Pb^{2+} : chemical cleavage; T1L: RNase T1 ladder; T1: RNase T1; A: RNase A; V1: RNase V1; V1L: dephosphorilated T1 ladder. The sequence of RNAII is shown on the left side of the gel. To facilitate the interpretation of the structure, the G residue positions are indicated on the right of V1L. As a summary of the probing (observed) results, a black line indicating the unpaired (continuous line) and paired (dashed line) sequences is shown on the left. For comparison, the predicted ss or ds conformation in the different regions of the expected transcript is also displayed on the left, maintaining the same color and line-style code as in A. (C) Antisense transcript folding model. The probed ss and ds sequences of the RNA (black line with the same line-style code as in A and B) are shown aligned with the predicted secondary structure of the transcript. Perfect match between the predicted and observed configuration of the RNA was observed. The ΔG value of the single hairpin is shown.

Figure 3. Structure of the *in vitro* synthesized mRNA₆₀ target transcript. (A) Sequence and predicted secondary structure of the mRNA₆₀ run-off transcript. The two extra Gs (+) are indicated. The mRNA₆₀ sequence complementary to antisense RNAII is shown. The different regions are color-coded, and the RNA configuration is displayed as in Fig. 2. (B) Structure probing of mRNA₆₀ by enzymatic and chemical cleavage. The different lanes are labeled as in Fig. 2. The sequence of mRNA₆₀ is indicated on the left side of the gel. The G residue positions are displayed on the right of V1L. The predicted and observed conformations of the different regions of the transcript are shown on the left with the same color and line-style code as in Fig. 2. The indicated portion of the gel is augmented to improve the observation of the cleavage pattern at the 3'-end of the RNA. G and U residue positions are shown on the right. A horizontal red line displays the 3'-end position of the expected 62-nt-long mRNA₆₀ run-off transcript. An extra G residue at position 68 is observed in the longer-than-expected product. (C) Folding model of the actual mRNA₆₀ run-off transcript.

The predicted secondary structure of the main longer transcript, as hinted from sequence probing in **B** and from the analysis of the mRNA₆₀ cDNA library, is displayed. The last base of the expected mRNA₆₀ transcript is shown in a green circle. The probed ss and ds sequences of the RNA (black line with the same line-style code as in **A** and **B**) match the predicted structure. The ΔG values of the hairpins in the mRNA₆₀ folding model are indicated.

Figure 4. Structure of the *in vitro* synthesized mRNA₈₀ target transcript.

(A) Sequence and predicted secondary structure of the mRNA₈₀ run-off transcript. The two extra Gs (+) are indicated. The mRNA₈₀ sequence complementary to antisense RNAII is shown. The different regions are color-coded, and the RNA configuration is displayed as in Fig. 2. **(B)** Structure probing of mRNA₈₀ by enzymatic and chemical cleavage. The different lanes are labeled as in Fig. 2. The sequence of mRNA₈₀ and the G residue positions are displayed on the left and right sides of the gel, respectively. The predicted and observed conformations of the different regions of the transcript are shown on the left with the same color and line-style code as in Fig. 2. The indicated portion of the gel is augmented, and G and U residue positions are shown on the right. A horizontal red line displays the 3'-end position of the expected 82-nt-long mRNA₈₀ run-off transcript. **(C)** Folding model of the actual mRNA₈₀ run-off transcript. The predicted secondary structure of the main longer-than-expected transcript, as inferred from the analysis of the mRNA₈₀ cDNA library, is displayed. The last base of the expected mRNA₈₀ transcript is shown in a green circle. The two additional CU dinucleotides that might base-pair with the 3'-terminal AG of the expected transcript are shadowed. The probed ss and ds sequences of the RNA (black line with the same line-style code as in **A** and **B**) match the predicted structure. The ΔG values of the hairpins in the predicted structures of expected and actual mRNA₈₀ are shown.

Figure 5. Efficiency of binding of the antisense RNA to the mRNA₆₀ and mRNA₈₀ target transcripts.

5'-end-labeled antisense RNA (0.2 nM) was incubated in the absence (-) or in the presence of unlabeled mRNA₆₀ (**A**) or mRNA₈₀ (**B**). Binding reactions shown on the left side of panels **A** and **B** contained increasing concentrations (50 to 150 nM) of the corresponding target mRNA, and were carried out at 37°C for 30 min. Time-course experiments displayed on the right side of panels **A** and **B** were performed in the presence of 150 nM of either target mRNA; samples were analyzed at various times within the first minute of incubation at 37°C. **(C)** The table shows the values of the association rate constant and the concentration of target mRNA for half-maximal binding obtained in the experiments of formation of complexes between antisense RNA and mRNA₆₀ (panel **A**) or mRNA₈₀ (panel **B**). Values are the mean of at least three independent assays.

Table 1

	frecuency of colonies	extra sequence	self-priming mechanism	
			intermolecular dimer	intramolecular hairpin
mRNA ₆₀	8/25	—	<div>5'...CUUGGGUAUAUUAUAcc3'</div> <div>3' ccAUAUAUAUAUGGGUUC...5'</div>	<div>5'...CUUGGGUAUAU</div> <div>3' ccAUAUA</div>
	13/25	c		
	4/25	cc		
mRNA ₈₀	5/10	ccauaaa	<div>5'...AUACUUUAUGGCUAAAGAAAAGccauaaag3'</div> <div>3' gaaauaccGAAAAGAAUUGGUAUUUCAUA...5'</div>	<div>5'...AUACUUUAUGGCUAAAG</div> <div>3' gaaauaccGAAAAA</div>
	5/10	ccauaaag		

Nucleotide sequence in the expected run-off transcripts is indicated by upper-case letters. The extra nucleotides, inferred from the cDNA libraries, are in lower-case letters. The extra sequence found in all individual colonies is shown in bold. Base-pairs proposed for self-priming are indicated with a dash. New base-pairs formed by the extra sequence are indicated with a dot.

Figure 1
[Click here to download high resolution image](#)

Figure 2
[Click here to download high resolution image](#)

Figure 3
[Click here to download high resolution image](#)

Figure 4
[Click here to download high resolution image](#)

Figure 5
[Click here to download high resolution image](#)