Structural Characterization of the Rous Sarcoma Virus RNA Stability Element[∇][†]

Jason E. Weil, Michalis Hadjithomas, and Karen L. Beemon*

Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218

In eukaryotic cells, an mRNA bearing a premature termination codon (PTC) or an abnormally long 3′ untranslated region (UTR) is often degraded by the nonsense-mediated mRNA decay (NMD) pathway. Despite the presence of a 5- to 7-kb 3′ UTR, unspliced retroviral RNA escapes this degradation. We previously identified the Rous sarcoma virus (RSV) stability element (RSE), an RNA element downstream of the *gag* natural translation termination codon that prevents degradation of the unspliced viral RNA. Insertion of this element downstream of a PTC in the RSV *gag* gene also inhibits NMD. Using partial RNase digestion and selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) chemistry, we determined the secondary structure of this element. Incorporating RNase and SHAPE data into structural prediction programs definitively shows that the RSE contains an AU-rich stretch of about 30 single-stranded nucleotides near the 5′ end and two substantial stem-loop structures. The overall secondary structure of the RSE appears to be conserved among 20 different avian retroviruses. The structural aspects of this element will serve as a tool in the future design of *cis* mutants in addressing the mechanism of stabilization.

Organisms have developed multiple ways to regulate the abundance and quality of their gene products. The RNA stability and level of protein production of many mRNAs are dictated by elements located in their untranslated regions (UTRs). Binding sites for translation-inhibiting microRNAs and AU-rich element binding proteins that determine RNA stability are often located in 3′ UTRs (4, 12). Additionally, when the 3′ UTR is abnormally long, the mRNA is often rapidly turned over by the cellular nonsense-mediated mRNA decay (NMD) pathway (7, 26, 30). Retroviruses have developed mechanisms to splice only a fraction of their primary RNA transcripts and to export and translate both spliced and unspliced mRNAs (5). The resulting unspliced *gag* mRNA has a very long 3′ UTR (>5 kb) (32), yet it is stable, with a half-life of greater than 7 h (31), suggesting that it somehow evades cellular mRNA surveillance mechanisms.

We are studying the mechanism by which the Rous sarcoma virus (RSV) unspliced mRNA is immune to NMD. Previous work has elucidated a *cis*-acting RNA sequence, called the RSV stability element (RSE), downstream of the *gag* termination codon (36). When the RSE is deleted, the ensuing RNA degradation requires RNA translation and the critical NMD factor Upf1, implicating the NMD pathway (36). While premature termination codons (PTCs) in the RSV *gag* gene also trigger NMD (2, 3, 19), insertion of the RSE downstream of the PTC stabilizes the RNA (36). Thus, when the RSE is downstream of a termination codon, it defines the termination event as correct. The RSE is a novel element, and its charac-terization may teach us more about the relationship between translation termination and NMD.

One trigger of NMD in *Saccharomyces cerevisiae*, *Drosophila*, and human mRNAs is the presence of a long 3′ UTR (6, 14, 16). An mRNA with a long 3′ UTR may undergo NMD because the poly(A) tail and corresponding poly(A) binding protein are distant from the terminating ribosome (1, 6); however, a recent study questions whether this is the full story (22). Another recent experiment showed that when the 3′ UTR of human immunoglobulin-μ mRNA was artificially extended from 0.3 to 1.6 kb, the mRNA became unstable (7). If a *cis*-acting complementary sequence was then added to "fold back" the RNA so that the poly(A) tail was brought into close proximity to the site of translation termination, the RNA became stable (11).

The average length of a 3′ UTR in yeast is on the order of 200 nucleotides (nt) (17, 27), whereas chickens and humans generally have 3′ UTRs on the order of 400 to 600 nt and 800 to 1,000 nt, respectively (8, 27, 39). It is expected that a chicken cell will naturally identify the RSV unspliced RNA as aberrant due to its long 3′ UTR. We hypothesize that the RSE exists in the viral RNA to overcome the destabilization caused by such a long UTR. This RNA element might cause the RNA to naturally fold in a way to position the poly(A) tail proximal to the termination event, similar to the experiment done by Eberle et al. (11). Or perhaps this RNA element recruits a stabilizing protein to promote efficient/proper termination at the *gag* natural stop codon. The RSE RNA might even interact directly with the terminating ribosome to promote termination, a concept similar to the process by which an internal ribosome entry site leads to translation initiation (28).

Previous work defined the RSE as a unique genetic element (36). Here, we sought to determine the structure that the RSE RNA forms. To this end, we examined the physical properties of a functional, 294-nt-long RSE fragment, located 110 nt downstream of the *gag* termination codon. This sequence has

* Corresponding author. Mailing address: Department of Biology, The Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218. Phone: (410) 516-4693. Fax: (410) 516-5213. E-mail: KLB@jhu .edu.

been shown to provide significant stabilization when inserted after a PTC (36). It does not contain the pseudoknot structure involved in the ribosomal frameshift at the *gag-pol* junction (18, 20, 32). We analyzed the RSE RNA element by partial RNase digestion, selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) (37), and analytical ultracentrifugation.

Partial RNase digestion can provide some information as to which nucleotides are not base paired, but it is restricted by the nucleotide specificities and steric hindrances of the RNase proteins. On the other hand, SHAPE chemistry is a technique whereby the 2′ OH of the ribose sugar of a flexible nucleotide (presumably a nucleotide not involved in secondary or tertiary interactions) is acylated by addition of a small molecule, such as *N*-methylisatoic anhydride or 1-methyl-7-nitroisatoic anhydride (1M7) (25, 37). The reactivity of the 2′ OH is extremely sensitive to its proximity to the phosphodiester linkage at the 3′ OH. Flexible (unconstrained) nucleotides allow the 2′ OH to sample more conformations and therefore exist in states that allow acylation by 1M7, whereas constrained nucleotides do not have this ability (37). The modified RNA is then analyzed by primer extension using reverse transcriptase.

While examining features of the secondary structure of the RSE RNA by RNase digestion and SHAPE, we consistently observed a long, single-stranded AU-rich region and two well-defined stem-loop structures. The analytical ultracentrifugation analysis showed that this RNA molecule exists as a single species and has a molecular weight consistent with that of a monomer. This is a crucial result because we observed anomalous mobility of the RSE RNA on denaturing polyacrylamide gels, a result often attributed to multimerization of the RNA. Additionally, we compared the sequence of the RSE to sequences from similar avian retroviruses, noting whether variant nucleotides were consistent with the two-dimensional structural model that we report. We found that the majority of variant nucleotides preserve the predicted secondary structure of this region and that there is a higher propensity for C/U variation within the middle of the RSE fragment. We present the secondary structure of the RSE and conclude that it is conserved among avian retroviruses.

MATERIALS AND METHODS

In vitro transcription. All RSV nucleotides correspond to the NCBI sequence with accession number NC_001407. The full-length 10.8 Prague C RSV plasmid (24) was used as a template for PCR with *Taq* DNA polymerase (New England Biolabs). Template for in vitro transcription of the RSE RNA was synthesized by PCR amplification with primers containing a T7 promoter. For the RSE C fragment (RSV nt 2597 to 2885), the forward oligonucleotide was 5′-TAATAC GACTCACTATAGGGTAGCGCTAACGCAATTAGTGG-3′, and the reverse oligonucleotide was 5′-TAGTAAATGCAAAAGCTTCGCG-3′. The underlined portion of the oligonucleotides corresponds to the sequence of the T7 promoter. T7 transcription reactions were carried out as described in LeBlanc and Beemon (19). The resulting in vitro transcript corresponds to three guanine residues (artifact of T7 promoter) followed by RSV nt 2597 to 2885. An extra TA dinucleotide is found at the 3′ end that causes the reverse RNA product to begin with a stop codon. The final in vitro RSE transcript is therefore 294 nt long. Template to make the fragment termed C +203, which is 103 nt longer at the 5′ end and 100 nt longer at the 3′ end, was amplified using the forward oligonucleotide 5′-TAATACGACTCACTATAGGGCTGTTCTCACTGTTGCGC-3′ and the reverse oligonucleotide 5′-CACTACCAACTGACAGATAGTGGG-3′ (the sequence of the T7 promoter is underlined).

Partial RNase digestion. RSE RNA was transcribed as described above. RNA to be kinased was treated with calf intestinal phosphatase (New England Biolabs)

and eluted from a G-50 spin column (GE Healthcare). Aliquots of the RNA were either 5′ end labeled with [γ-^{32}P]ATP or 3′ end labeled using 5′ [^{32}P]pCp (Perkin Elmer) and T4 RNA ligase (New England Biolabs). These RNAs were resuspended in 100 μl of water. Approximately 40,000 cpm of RNA was brought up to 20 μl in 1× refolding buffer (6 mM MgCl$_2$, 30 mM Tris-HCl, pH 7.8, 300 mM KCl) and boiled at 95°C for 5 min. These RNAs were allowed to cool to 25°C in a beaker of water starting at 65°C. The volume of each sample was brought to 40 μl and digested either without RNase or with the following amounts of RNase: RNase T$_1$ (Calbiochem), 0.05, 0.005, or 0.0005 units; RNase A (Calbiochem), 1 μg/ml, 0.2 μg/ml, or 0.1 μg/ml; and RNase V$_1$ (Pharmacia Biotech), 1.44, 0.72, or 0.36 units. These were digested for 40 min at 25°C and then extracted with phenol-chloroform-isoamyl alcohol and precipitated. They were electrophoresed on both 8% and 15% polyacrylamide gels containing 8 M urea.

SHAPE chemistry. Oligonucleotides (IDT technologies) used in primer extension had the following sequences: primer 1, 5′-CGAAGACAGGTGTGTTCC-3′; primer 2, 5′-GGAACAAGCTTGGCG-3′, primer 3, 5′-GTAAATGCAAAA GCTTCGCG-3′; and primer 4, 5′-CCTTCCATTGGAATCTTCG-3′. Primer 4 binds downstream of the RSE, and was used only with C +203 RNA. These oligonucleotides were 5′ end labeled using T4 polynucleotide kinase (New England Biolabs).

For SHAPE reactions, 1 pmole of in vitro transcribed RSE RNA was resuspended in 5 μl of H$_2$O. This was boiled for 2 min and then placed on ice. Three microliters of 3× refolding buffer (333 mM HEPES, pH 7.9, 333 mM NaCl, and 16.65 mM MgCl$_2$) was added, and the RNA was incubated at 37°C for 20 min. Either 1 μl or 0.5 μl of a 65 mM stock of the compound 1M7 in dimethyl sulfoxide was added to bring total volume to 9 μl (25). The 1M7 was a gift from Kevin Weeks, University of North Carolina, Chapel Hill, NC. This compound was allowed to react for >70 s (>5 half-lives) at room temperature. The reaction mixture was then ethanol precipitated and resuspended in 10.75 μl of H$_2$O, and 0.2 pmol (0.25 μl) of primer 1, 2, 3, or 4 was added. The reaction mixture was incubated at 65°C for 5 min, followed by incubation at 37°C for 5 min to allow primer annealing. The extension mixture consisted of 5 μl of 5× buffer, 2 μl of the deoxynucleoside triphosphates (20 mM each), 0.5 μl of RNasin (Promega), 0.25 μl of dithiothreitol (0.1 M), and 5.75 μl of H$_2$O. This mixture was incubated at 57°C for 1 min, and then 0.5 μl of Superscript III (Invitrogen) reverse transcriptase was added. This reaction mixture was incubated for 20 min at 57°C. The reaction mixture was then extracted with phenol-chloroform-isoamyl alcohol and precipitated. The primer extension reaction mixture was electrophoresed on an 8 M urea–15% polyacrylamide gel. Polyacrylamide gels were exposed to film overnight or a phosphor storage screen for 4 to 16 h and then developed or imaged on a Typhoon 9410 Phosphorimager (Amersham).

Analytical ultracentrifugation. RSE RNA was synthesized in vitro as described above. The RNA was denatured by heating 200 to 600 μl to 95°C for 2 min in refolding buffer (80 mM NaCl and 10 mM MOPS [morpholinepropanesulfonic acid], pH 7.0). MgCl$_2$ was added to a final concentration of 0.5 mM, and the RNA was renatured by cooling to room temperature. For the sedimentation equilibrium experiments, three concentrations of RNA (8.3, 20, and 28.3 μg/ml) were loaded in six-sector cells at a volume of 110 μl. Data were collected using the absorption optics of a Beckman XL-I analytical ultracentrifuge at 260 nm and 20°C and rotor speeds of 9,000, 10,500, and 12,000 rpm in a Ti-60 rotor. Equilibrium was confirmed by subtracting consecutive scans at 2-h intervals. The data were analyzed using the Origin-based data analysis software for Beckman XL-A/XL-I (Beckman Instruments, Beckman Coulter, Fullerton, CA). A calculated partial specific volume of RNA of 0.5691 ml/g (34) was used for the analysis. The density for the buffer was calculated to be 1.002 g/ml using SEDNTERP (http://www.bbri.org/rasmb/rasmb.html).

For the velocity sedimentations experiments, 400 μl of sample and 420 μl of buffer were loaded in double-sector cells. The loaded cells and rotor (Ti-60) were allowed to equilibrate to 20°C. Data were collected at 260 nm with a loading concentration of 31.7 μg/ml at rotor speeds of 35,000 rpm and 38,000 rpm. Analysis was carried out with the Van Holde-Weischet and time derivative methods using the Ultrascan program (10).

Alignment of avian retroviral RSE sequences. Twenty different avian retroviral RSE sequences were aligned using the ClustalW program at the European Bioinformatics Institute website (http://www.ebi.ac.uk/clustalw/) (15). The names and accession numbers of the viruses used are as follows: RSV, NC_001407; RSV (duck adapted), X68524; avian leukosis virus (ALV) strain EV-1, AY013303; ALV-RSA, M37980; ALV ADOL-7501, AY027920; myeloblastosis-associated virus type 2, L10924; RSV Schmidt-Ruppin D, D10652; Rous-associated virus type 2, D21237; ALV, AB112960; ALV-LR9, AY350569; RSV Schmidt-Ruppin B, AF052428; ALV HPRS103 (subgroup J), Z46390; ALV strain NX0101, DQ115805; ALV strain PDRC-1039, EU070900; ALV strain PDRC-3249, EU070902; ALV strain PDRC-
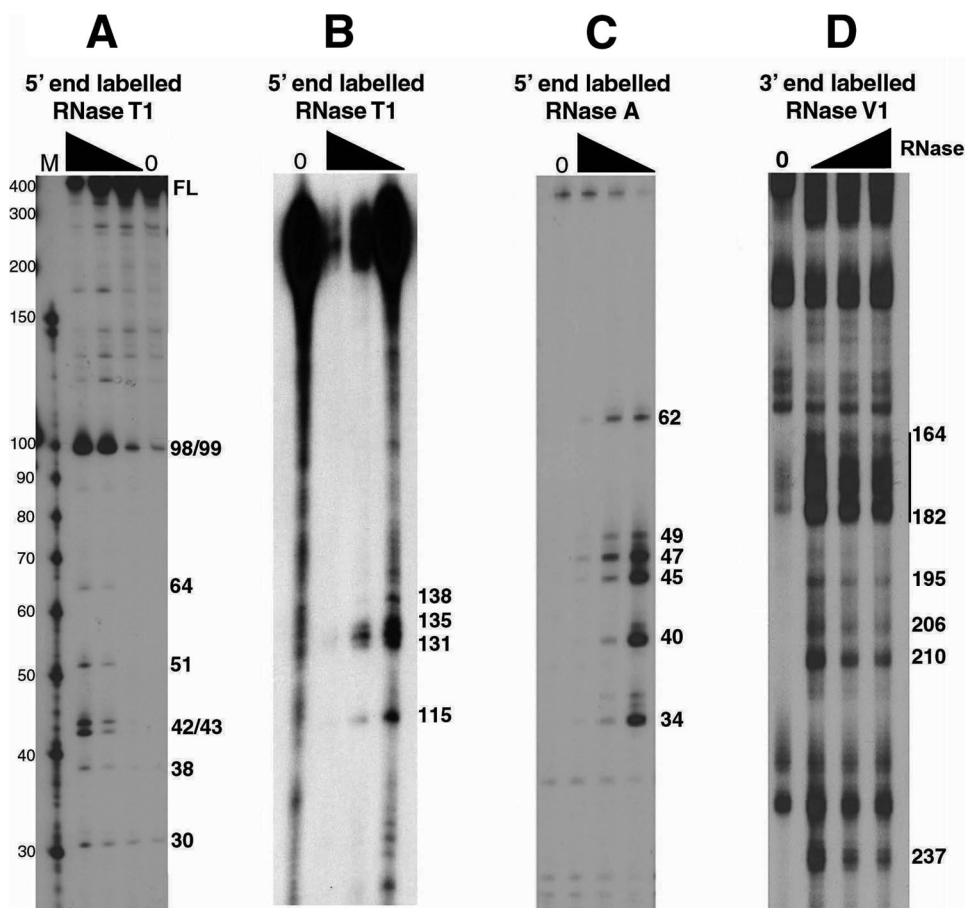
FIG. 1. Partial RNase digestion shows single- and double-stranded regions of RSE RNA. (A and B) The 5′-end-labeled RSE RNA (294 nt) was digested with increasing amounts of RNase T$_1$. The resulting cleavage products were run on a 15% (A) or 8% (B) polyacrylamide–8 M urea gel. Note that in panel A the 294-nt RSE full-length (FL) RNA comigrates with a marker of approximately 400 nt. (C) The 5′-end-labeled RSE RNA was digested with increasing amounts of RNase A and resolved on an 8 M urea–15% polyacrylamide gel. (D) The 3′-end-labeled RSE RNA was digested with RNase V$_1$ and resolved on an 8 M urea–8% polyacrylamide gel. Numbers on the right correspond to the RSE nucleotide that is cleaved by RNase. Nucleotide position 4 of the RSE RNA corresponds to nt 2597 of the complete RSV genomic RNA sequence. The black triangle indicates increasing RNase concentrations. Lanes 0, RNA alone (no RNase added).

3246, EU070901; ALV strain SD0501, EF467236; ALV strain RSA, M37980; ALV strain MQNCSU, DQ365814; and ALV strain TymS_90, AB303223.

RESULTS

Partial RNase digestions suggest two stable stem-loop structures within the RSE. To determine single-stranded and helical regions of the RSE RNA, we radioactively labeled either the 5′ or 3′ end, folded the RNA, and then digested with limiting amounts of specific RNases. RNase T$_1$ specifically cuts after unpaired G residues, RNase A specifically cuts after unpaired C and U residues, and RNase V$_1$ preferentially cuts helical RNA. After partial digestion with one of these RNases, cleavage products were run on an 8 M urea-polyacrylamide gel (either 8% or 15% polyacrylamide). We observed several prominent bands by RNase T$_1$ digestion (Fig. 1A and B) and by RNase A digestion (Fig. 1C) indicating non-base paired regions. We also saw regions of the RNA highly susceptible to RNase V$_1$ digestion although single-nucleotide cuts were difficult to distinguish (Fig. 1D).

The results of the RNase T$_1$ and RNase A digestion exper-

iments were incorporated into an RNA secondary structure prediction using the programs RNAstructure (21) and Mfold (41). The programs predicted about 10 potential structures when constrained with the RNase T$_1$ and RNase A data. In contrast, when unconstrained, the programs predicted more than 30 potential structures with a similar free energy. Utilizing the RNase A and T$_1$ constraints, every predicted structure contained a very strong imperfect stem-loop between RSE nucleotides 162 to 216 (stem-loop 2). When this putative stem-loop was individually folded with the Mfold program, it had a predicted ΔG of -35.90 kcal/mol. Because this local structure has such a favorable free energy, it most likely forms readily and contributes to the overall folding of the RSE RNA. This imperfect stem-loop was also predicted in the most favorable predicted structures in the absence of any structural constraints. Interestingly, this stem-loop contains the nucleotides cleaved by RNase V$_1$ (Fig. 1D), providing further evidence for its existence. This stem-loop probably provides the strong structure that prevents the RNA from being fully denatured when run on an 8 M urea-polyacrylamide gel (Fig. 1A, full-

length [FL] RNA); constructs lacking this stem-loop did not run anomalously (data not shown).

With these same constraints (RNase A and T$_1$), a second, much smaller perfect stem-loop (stem-loop 1, nt 89 to 109) was also predicted in most structures but only in the most favorable of the unconstrained predictions. Interestingly, the most prominent RNase T$_1$ cleavage sites observed, even at the lowest RNase concentrations, are at the guanine residues found at positions 98 and 99 within the loop of this structure (Fig. 1A). This suggests that this loop is on the outer surface of the three-dimensional folded RNA and very accessible to the RNase. In conclusion, the RNase digestion results provided constraints that helped the structural prediction programs generate a two-dimensional structure and suggested the presence of at least two stem-loop structures.

Although the RNase digestion experiments helped to define these two stem-loops, the remainder of the molecule was poorly defined among the predictions of the RNAstructure and Mfold programs. In order to gain information on the rest of the molecule and to better define the nature of the stem-loop between nt 89 and 109, we used the technique of SHAPE chemistry (25, 37).

SHAPE chemistry reveals additional secondary structure features of the RSE. To analyze the secondary structure of the RSE by SHAPE chemistry, we denatured and refolded the RSE RNA and then treated the RNA with 1M7. Four different 5′-end-labeled primers were used for primer extension (Fig. 2A). Primer 4 is outside of the RSE region and can only be used with C +203 RNA. Using SHAPE chemistry, we found several specific primer extension stop sites in samples that had been treated with 1M7 (Fig. 2B to E). Primer 1 begins at nt 87 of the RSE and allows us to examine the 5′ end of the RNA at single-nucleotide resolution (Fig. 2B). Surprisingly, using this primer most nucleotides were modified by 1M7, indicating that a majority of the 5′ portion of the RNA consists of flexible nucleotides (nt 26 to 29, 32/33, and 39 to 53) and is therefore probably unstructured. In contrast, we observed some adjacent regions that were completely devoid of detectable modification (nt 54 to 60).

Primer 2 starts at nt 157 and, much like primer 1, generates several clear bands corresponding to modified nucleotides (Fig. 2C). However, many additional nucleotides were weakly modified. Using primer 2, we observed a strong natural reverse transcriptase pause site (indicated by a very strong signal in the lane lacking 1M7 treatment) (Fig. 2C) corresponding to nt 108. Presumably, this is because the polymerase is stopped at an incompletely denatured region of the RNA; this pause is at the base of the predicted stem-loop 1 (nt 89 to 109).

Finally, to resolve the structure at the 3′ end, we used primer 3 (Fig. 2D) and primer 4 (Fig. 2E). Using these primers, we again observed distinct regions of modification (nt 244 to 249, 256 to 261, and 276 to 288). There is some general background in the input lane (lane 0), but the major bands derived from primer 3 in the absence of 1M7 (nt 200 and nt 212) are within the strong, imperfect stem-loop 2 predicted from the RNase digestion experiments. Surprisingly, there also appear to be two natural pause sites in a predicted single-stranded region (nt 237 and 242). Interestingly, there is also a single RNase V$_1$ cut near this site. This leaves open the possibility that the RSE

forms additional structure in this region, or that this single stranded region is involved in a tertiary interaction.

Primer 4 was used only with the C +203 RNA transcript, and extension from primer 4 led to more background than from primer 3. However, the nucleotides corresponding to the small portion of the RSE covered by this transcript/primer are clearly modified or not modified. An additional primer complementary to sequence near stem-loop 2 was also used, but no extension products were observed. This is also consistent with a very strong structure (stem-loop 2) in this region.

SHAPE chemistry confirms RNase digestion results and leads to a two-dimensional structure of the RSE. Using the Mfold (41) and RNAstructure (21) programs with the RNA folding constraints based on our SHAPE and RNase (T$_1$ and A) data, we constructed a two-dimensional model describing the most likely structure of this RNA (Fig. 3). In order to prevent inaccuracy in our model due to tertiary interactions, we incorporated only information on single-stranded nucleotides. If a nucleotide is held in a rigid tertiary interaction, it might not be modified by 1M7 but is also not necessarily in a helical secondary structure. Although the observations of non-modified nucleotides and natural pause sites were not included in the structural prediction, they are depicted on Fig. 3, and aside from nt 237 and 242, it is clear that they fit the prediction.

Although RNase digestion data (Fig. 3) gave us some insight into the structure of this molecule, adding the data derived from SHAPE analysis (Fig. 3) helped us to refine the model. With constraints from SHAPE data alone, the RNAstructure program predicted six different structures, and Mfold returned four potential, similar structures. When we combined both the RNase and SHAPE data, RNAstructure predicted five different folds, and Mfold predicted two possible structures. The two most favorable RNAstructure predictions were highly similar to the two Mfold predictions and differed only in predicted base pairing in the unstructured 5′ terminal region. The main difference in predicted structures between the two programs was the proposal of the weak stem-loop between nucleotides 54 and 71 as well as a weak three-base-pair interaction (data not shown) predicted by RNAstructure. In contrast, a weak stem-loop between nt 71 and 86 was predicted by Mfold. As discussed below, the weak stem-loop between nt 54 and 71 fit our observations of nucleotides not modified during SHAPE (Fig. 2B, nt 54 to 60) and was the slightly more energetically favorable prediction. The model shown in Fig. 3 incorporates the above observations.

In our model, the first notable structural feature is the long single-stranded region between nt 25 and 53 (Fig. 3), corresponding to the large stretch of modified nucleotides observed using primer 1 (Fig. 2B). Interestingly, the small region of unmodified nucleotides in Fig. 2B corresponds to nt 54 to 60 in the structural model. Although these nt were not included as constraints in the structural prediction because they could be due to tertiary interactions, the structure shown in Fig. 3 has a partial stem-loop in this region. This is interesting in that nt 56 and 57 (both uracils) are not predicted to base pair with nt 68 and 69 (also both uracils). Although not depicted here, uracil can sometimes form weak interactions with other uracils (40), as our model suggests.

The next prominent feature is stem-loop 1, found between nt 89 and 109. This region contains 7 to 8 bp in the stem and 5 to
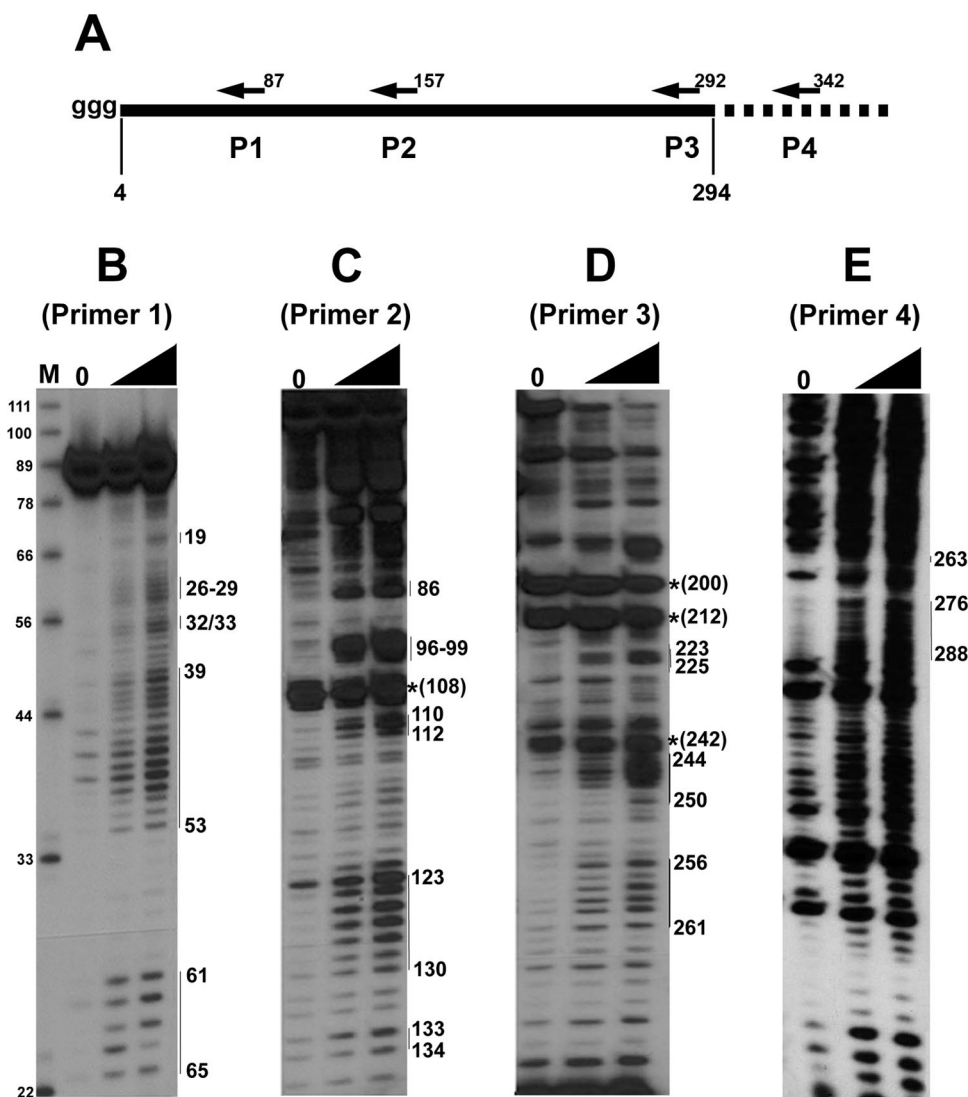
FIG. 2. SHAPE chemistry reveals structural features of the RSE. (A) A schematic of the in vitro transcribed RSE RNA used in structural studies. Nucleotide position 4 of the RSE RNA corresponds to nt 2597 of RSV. Folded, in vitro transcribed RSE RNA was selectively modified at single-stranded nucleotides by the compound 1M7. Modified nucleotides were detected by primer extension using primers (P) beginning at nt 87 (B), 157 (C), 292 (D), and 342 (E). Extension products were resolved on 12% or 15% polyacrylamide–8 M urea gels. The numbers on the right correspond to the modified RSE nucleotide. *, a nucleotide at which there is a natural pause site or break in the RNA. The RNA used for primer extension in panel E is approximately 100 nt longer at both the 5′ and 3′ ends. The black triangle indicates increasing 1M7 amounts. Lanes 0, RNA alone (no 1M7 added).

7 nt in the loop. The RNAstructure program often predicted this structure with only RNase digestion data. Additionally, partial RNase digestion of end-labeled RNA shows a prominent band corresponding to cleavage at the guanine residues in the loop of this structure (Fig. 4). The fact that this is often the most prominent band in the RNase digestion analysis suggests that RNase T_1 readily cuts this region. This observation led us to hypothesize that stem-loop 1 is on the external face of the folded RNA and therefore available for potential RNA-RNA or RNA-protein interactions. A strong natural pause site of the reverse transcriptase was observed at nt 108 but was not incorporated in the constraints for structural prediction. However, a natural pause site is consistent with a stable structure (stem-loop 1) at this position.

A GC-rich imperfect stem-loop found between nt 163 and

215 was termed stem-loop 2. As mentioned above, RNAstructure and Mfold predict this structure even without experimentally determined constraints. In fact, a primer (nt 221 to 237) downstream of this region either did not bind or could not extend, and we were subsequently unable to obtain SHAPE data with it. Likewise, it is impossible to garner any information from the longer extension products with primer 3 (which would cover this region), except for the very strong natural pause sites. These pauses appear to occur around nt 212 and 200. Again, reverse transcriptase pausing here is consistent with strong secondary structure in this region. We were, however, able to confirm the structure of this region by cleaving with RNase V_1. RNase V_1 is an enzyme that preferentially cuts after base-paired nucleotides. Despite slight degradation, when we labeled the 3′ end of the RSE and cut with RNase V_1,
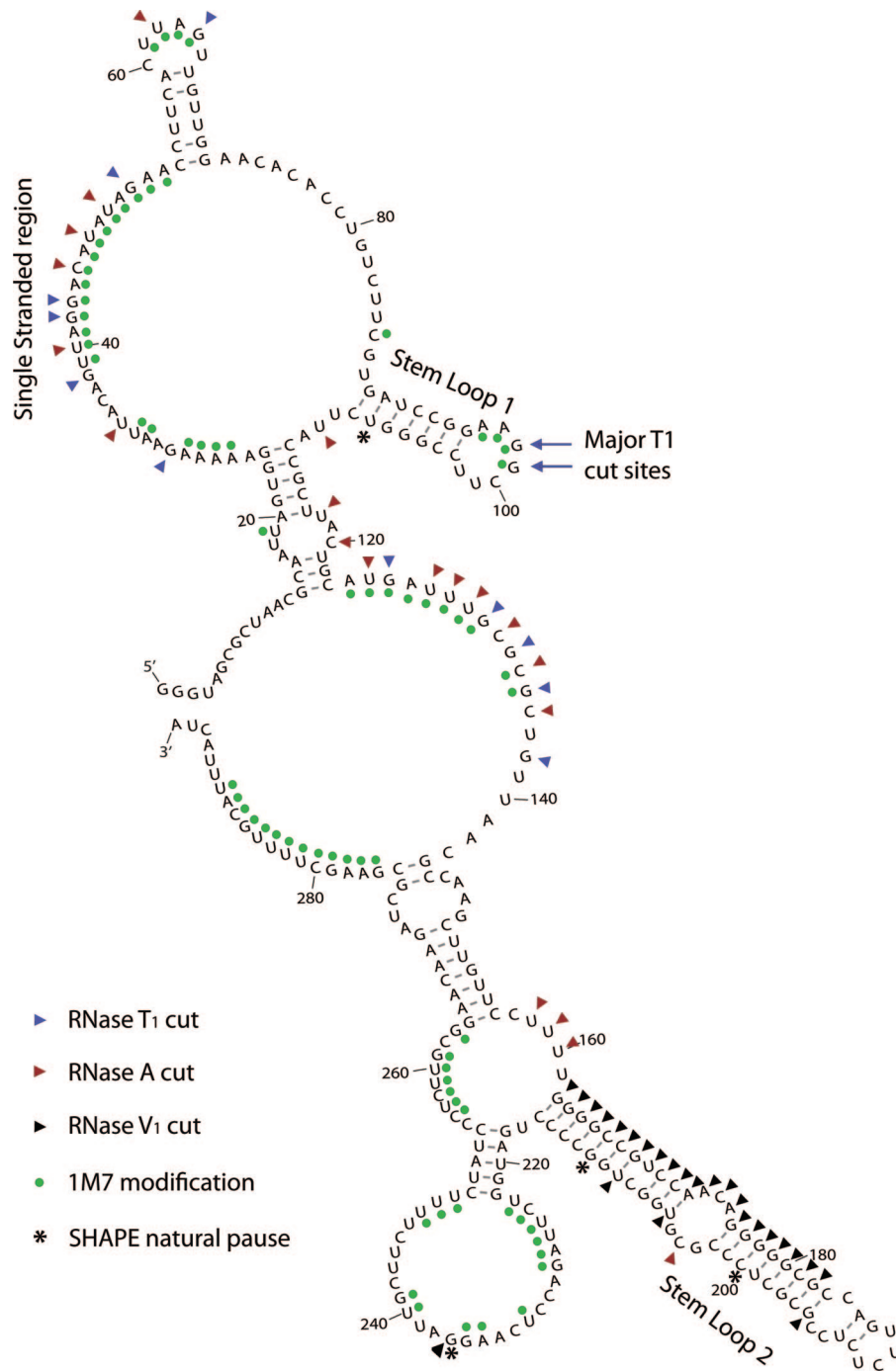
FIG. 3. Secondary structure model of the RSE RNA. RSE RNA corresponding to nt 2597 to 2885 (nt 4 to 292 in this figure) was folded in silico using the Mfold RNA structure prediction program (41). The RNase T_1 and A cuts as well as SHAPE modifications (1M7) shown in Fig. 1 and 2 were used as constraints for this prediction and are indicated on the structure. Natural pause sites and RNase V_1 were not included in the prediction but are indicated on this figure.

we observed several prominent bands that correspond to cuts within this predicted stem (Fig. 1D).

Finally, we transcribed RNA in vitro that was 203 nt longer than the RSE RNA (extra 103 nt upstream and 100 nt downstream). This RNA was termed C +203, and we did not observe any substantial differences between it and the shorter C fragment transcript in SHAPE experiments (data not shown). Additionally we used C +203 and primer 4 (nt 2935 to 2917) to

resolve the base pairing profile of nucleotides at the 3′ end of the RSE region (Fig. 2E). Although there is more background (Fig. 2E, lane 0) with this transcript than with the others examined, the small region of the RSE that is analyzed in this experiment is clearly modified.

Alignment of avian retroviral RNA sequences shows that variation between strains preserves secondary structure. The RSE region is highly conserved in most of the known avian
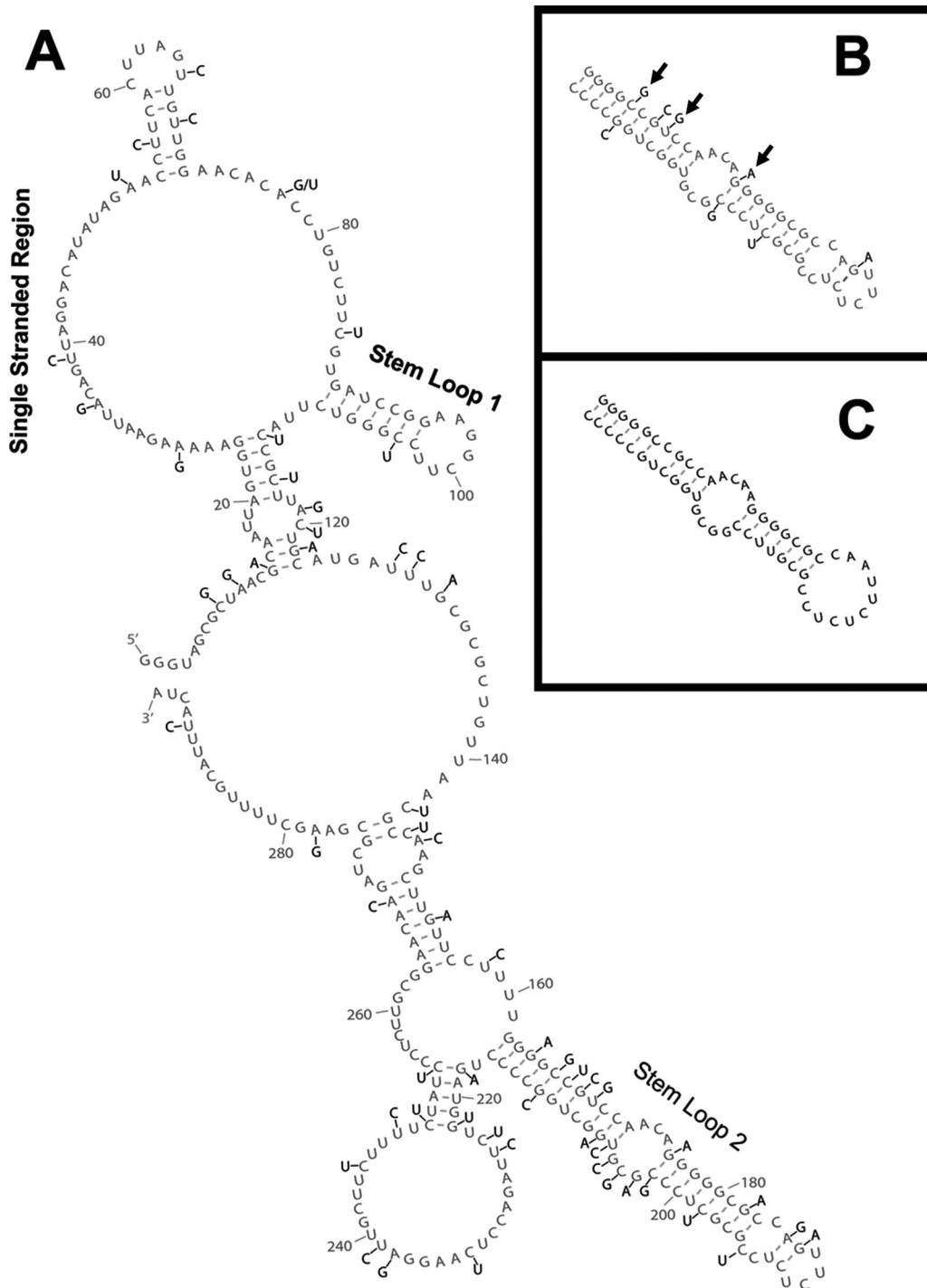
FIG. 4. Analysis of 20 retroviral genomes shows variant nucleotide positions within the RSE. Genomic sequences of 20 different avian retroviral strains were aligned, and variant nucleotide positions were noted. (A) A nucleotide position that varied in one or more of the 20 retroviral strains is indicated (black lettering) on the two-dimensional model from Fig. 3 (gray lettering). The three key structural features are indicated. (B) The RSE stem-loop 2 region is shown in gray, with variant nucleotides from the most divergent strain (ALV NX0101) shown in black. Arrows indicate a compensatory mutation, a new base pair, and a coordinated disruption of a base pair. (C) Mfold prediction of the stem-loop 2 region from ALV NX0101.

retroviral genomes. We aligned 20 avian retroviruses and found 58 variant (20.1%) nucleotide positions (a nucleotide that differed from the consensus alignment in one or more of the viruses) within the RSE region. In contrast, the frequency

of variant nucleotides throughout the entire *pol* gene was 13.7%. Although the overall mutation rate within the RSE is slightly higher, the nature of these variant nucleotides as well as their distribution throughout the RSE is very interesting.

TABLE 1. Variation among avian retroviruses

Region[a]	No. of variant nucleotides/total no. of nucleotides (%)	No. of C/U changes/total no. of variant nucleotides (%)
pol	371/2,705 (14)	165/371 (44)
RSE	58/288 (20)	27/58 (47)
RSE without SL2	40/233 (17)	23/40 (58)
SL2 alone	18/55 (33)	4/18 (22)

[a] SL2, stem-loop 2.

For instance, we observe that stem-loop 2 makes up only 19% of the RSE but houses 33% of the variant nucleotide positions. Additionally, the type of nucleotide variation seems to differ between stem-loop 2 and the rest of the RSE. There is a substantially greater number of C/U transitions outside of stem-loop 2.

This abundance of C/U transition changes is relevant to our study in that these changes will often preserve RNA secondary (or tertiary) structure. For instance, a transition mutation of C to U (or vice versa) preserves base pairing with a G. We found that throughout the entire *pol* gene, 45% of variant nucleotides were transitions of C to U or U to C. The RSE has approximately this same ratio of C/U transitions (47%). However, if stem-loop 2 is excluded, 58% of variant nucleotides are C/U transitions. This is in stark contrast to stem-loop 2 alone, which consists of only 22% C/U transitions.

Figure 4A shows the two dimensional model of the RSE with all variant positions labeled. Nucleotides that varied in even one of the 20 viruses examined were marked on this figure. Table 1 shows a comparison of four different regions: all of *pol*, the RSE, the RSE with stem-loop 2 removed, and stem-loop 2 alone. The table lists the RSV region examined, the number and percentage of variant positions within the region, and the percentage of these that are C/U transitions (see Table S1 in the supplemental material for detailed information on all viruses examined).

The higher number of variant positions within stem-loop 2 is not a surprising result. From Fig. 3, it is clear that stem-loop 2 is quite GC rich, and the structure is therefore quite stable. In fact, Mfold analysis of stem-loop 2 alone (nt 162 to 216) predicts a ΔG of -35.9 kcal/mol. Slight mutation within this stem is highly unlikely to disrupt the folding of such a stable structure. Accordingly, when we carried out Mfold analysis on stem-loop 2 of all the different viral strains, the predicted ΔG values ranged from -26.8 kcal/mol to -39.2 kcal/mol, with all but two of them more favorable than -30.0 kcal/mol (see Table S2 in the supplemental material for a list of the predicted ΔG of stem-loop 2 from all 20 different strains).

To further address the hypothesis that stem-loop 2 forms despite these variant nucleotides, we closely examined ALV strain NX0101. This strain varied the most within stem-loop 2 compared to the RSV sequence. This strain still forms a similar stem-loop with a free energy of -32.1 kcal/mol despite the extensive variation in this region (8/55 positions). Figure 4B shows stem-loop 2 from the RSV RSE with the variant nucleotides from ALV NX0101 depicted, and Fig. 4C shows the predicted Mfold of the NX0101 variant. A few notable structural features such as the compensatory change from GC to CG (nt 166 and 212), the formation of a base pair between nt

169 and 209, and the disruption of both nucleotides in one base pair between nt 176 and 202 are denoted by arrows (Fig. 4B). These data suggest that the importance of stem-loop 2 may be its ability to form a robust hairpin structure rather than its primary nucleotide sequence.

Analytical ultracentrifugation suggests that the RSE RNA is a folded monomer. During the course of our secondary structure analysis, we observed that the RSE RNA ran anomalously in denaturing gels. A 294-nt RNA comigrated with an RNA marker of approximately 400 nt in a 15% polyacrylamide gel (Fig. 1A). In order to determine whether this was due to incomplete RNA denaturation or multimerization, we performed both equilibrium and velocity sedimentation experiments. Figure 5A shows the sedimentation equilibrium data collected at an RNA concentration of 20 μg/ml at three different rotor speeds. The resulting data were globally fit to a curve using a previously reported calculation of the partial specific volume of RNA (0.5691 ml/g) (34). Notably, all scans fit well to a curve generated by constraining the molecular mass at 94,173 Da (the calculated mass of the RSE RNA) (Fig. 5B) as opposed to the predicted curve of a molecule with a mass of 188,346 Da (calculated mass of an RSE dimer) (Fig. 5B). The raw data align almost perfectly with the curve predicted for a monomer. The experiment is quite reproducible in that the curves generated from three different speeds (Fig. 5A) and three different RNA concentrations (data not shown) are very similar. From this we can conclude that under the conditions of the analytical ultracentrifugation, the RSE RNA is a monomer in solution.

Additionally, we examined the hydrodynamic behavior of the molecule by velocity sedimentation. Van Holde-Weischet analysis, using the Ultrascan program (10), indicated that the apparent sedimentation coefficient values in the sample were distributed over a narrow range (Fig. 5C). This provides no evidence of multiple species, an observation consistent with the equilibrium data showing that the molecule is a monomer. The average sedimentation coefficient value was 7.4S, the same as that calculated using the time derivative method (Fig. 5C, inset) of the Ultrascan program. Comparing the frictional coefficient (f) value of the RSE RNA to that of an ideal sphere (f_0) having the same molecular weight, sedimentation coefficient value, and partial specific volume gives an indication of the level of compaction. The frictional coefficient ratio (f/f_0) of the RSE was 1.72, a value that is similar to that of the folded 16S rRNA (f/f_0 of 1.77) (33). We therefore conclude that under these conditions, the RSE exists as a folded monomer in solution.

DISCUSSION

The RSE, an RNA regulatory element found in RSV RNA, allows the full-length viral transcript to evade the cellular NMD pathway (36). In this report we describe physical properties of the RSE. As determined by RNase digestion and SHAPE chemistry, this structured portion of the viral RNA contains two substantial stem-loops and a long single-stranded region. We found it useful to use both techniques since they addressed different features of the RNA. SHAPE chemistry was more accessible to most regions of the RNA structure. However, the strong stem-loop 2 could not be probed easily by SHAPE due to difficulties in reverse transcription in this highly structured region, but it was cut by RNase V_1. Additionally, we find that the RSE is well conserved among avian retroviruses
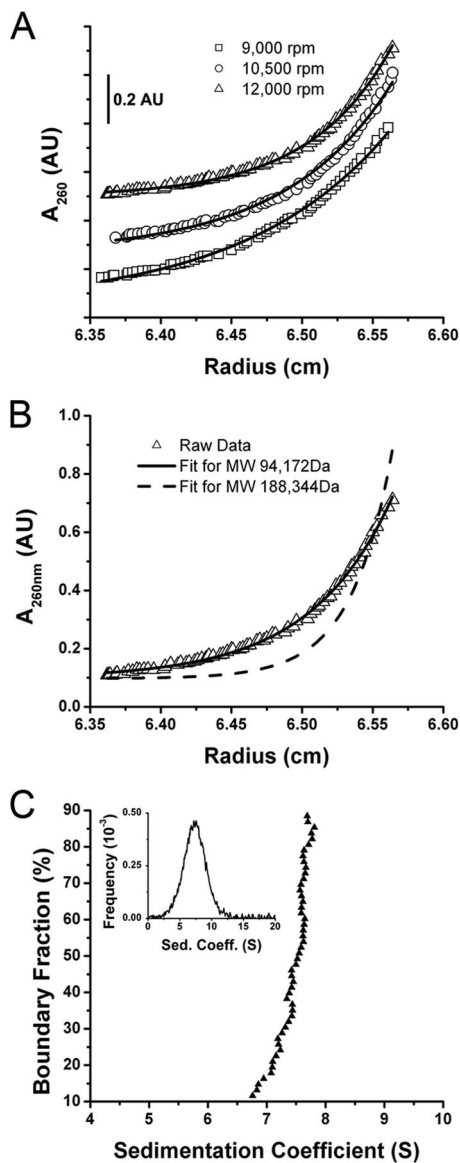
FIG. 5. Analytical ultracentrifugation of the RSE RNA. (A) Sedimentation equilibrium data were collected at three rotor speeds (9,000, 10,500, and 12,000 rpm) at 20°C with a loading concentration of 20 μg/ml monitored at 260 nm. Data plots were offset by 0.2 absorbance units (AU) for clarity. The solid line represents the predicted mobility of an RNA with a molecular mass of 94,173 Da (the mass calculated from the RSE RNA sequence) as determined by the Ultrascan software. (B) Comparison of curve fits to the 12,000-rpm data set from panel A. Using the Ultrascan software, the molecular mass was constrained to either the monomeric (solid line; 94,173 Da) or dimeric (dashed line; 188,346 Da) calculated mass, and both of these predicted curves were overlaid on the graph. (C) Van Holde-Weischet analysis of the velocity sedimentation data at 35,000 rpm and 20°C shows that the sample is primarily a single species. This plot was constructed by analyzing 50 equally spaced boundary fractions, spanning 10 to 90% of the boundary. The average S value determined was 7.4. (Inset) The same data analyzed using the time derivative method produced the same S value. MW, molecular weight; Sed Coeff, sedimentation coefficient.

and that variant nucleotides preserve the secondary structure. Moreover, there is a greater number of C/U variations outside of stem-loop 2 in the RSE. Many of these are in single-stranded regions, that may be involved in tertiary interactions

in the full-length viral RNA. This observation could potentially be used when searching for similar functional RNA elements in other viral and cellular RNAs.

The ultracentrifugation experiments (Fig. 5) suggest that the RSE RNA is a folded monomer in solution, with a frictional coefficient similar to that of the structured 16S rRNA (33). Its anomalous electrophoretic mobility in denaturing gels is therefore due to incomplete denaturation of the RNA, caused by the extremely stable stem-loop 2 because deletion of this stem-loop restored normal mobility (data not shown).

Our secondary structure analysis identified two well-defined stem-loop structures. Stem-loop 1 appears to be a relatively stable structure as evidenced by the natural pause site at nt 108. Additionally, this stem-loop was frequently predicted with only RNase digestion data and may be on the surface of the folded RNA since RNase T_1 so readily cleaves the G residues in the loop. The long and robust stem-loop 2 houses a substantial portion of the variant nucleotides seen in an alignment of different viruses. However, the high GC composition of this structure allows it to tolerate slight base pair disruptions but still fold properly. This suggests that the primary sequence information in stem-loop 2 may not be its contribution to RSE function. As mentioned above, RNA containing stem-loop 2 is very difficult to fully denature and contributes substantially to the overall folding of the RSE RNA. Since this functional fragment of the RSE does not contain the structured pseudoknot upstream involved in the ribosomal frameshift, stem-loop 2 may serve to help it rapidly and correctly fold. However, strong secondary structure alone is not sufficient for RNA stabilization; the reverse complement of the RSE RNA also has a predicted GC-rich stem-loop and an anomalous electrophoretic mobility (data not shown) but has no stabilizing function when placed after a PTC in *gag* (36). Thus, we think that the RSE requires a functional middle region and either the known pseudoknot at the 5′ end or stem-loop 2 at the 3′ end that act as nucleation points for RSE folding.

Secondary structure determination of human immunodeficiency virus type 1 (HIV-1) RNA by SHAPE chemistry demonstrated that the 5′ UTR noncoding regulatory region is more structured than the HIV *gag* coding region (38). The RSE is clearly not as structured as the 5′ regulatory regions of HIV, but this is not surprising since it is both a regulatory element and part of the *pol* gene coding region. It will be interesting to determine whether other retroviruses, such as HIV-1, have elements that function like the RSE. Endogenous retroviruses and retrotransposons are one class of RNA stabilized in human cells when the NMD pathway is blocked (23), possibly due to their loss of putative stability elements by mutation.

Because the RSE region is part of the *pol* gene, a high rate of natural sequence identity exists within this region among avian retroviruses. Despite their similarity, alignment of 20 avian retroviruses suggests several interesting facets of the derived secondary structure. The most common nucleotide change within the *pol* gene among the avian retroviruses examined is a transition of cytosine to uracil or vice versa. A change of cytosine to uracil preserves the ability to form a base pair with G. Within the RSE but outside of stem-loop 2, C/U transitions make up greater than one-half of all variant positions. These variants all preserve the predicted base pair when found in a double-stranded region; some of the variants are in

single-stranded regions. Conversely, stem-loop 2 seems to tolerate more nucleotide variation than the rest of the RSE. Greater than one-third of the variant nucleotides within the RSE are found within the 55-nt stem-loop 2, suggesting that the primary sequence of stem-loop 2 might not be necessary for RSE function. This is consistent with our proposal that stem-loop 2 serves as an anchor, allowing the rest of the RSE to fold properly. All but two of the 20 viruses examined retained a ΔG more favorable than -30.0 kcal/mol for the stem-loop 2 region. One of the strains (compare Fig. 4B and C) has ~15% variability within this region yet is still predicted to fold into a stable structure similar to that predicted for the Prague C RSV RSE.

Previous work showed that this RSE fragment has a necessary but not sufficient region between nt 1 and 142 (36). This region encompasses almost one-half of the RSE but contains only slightly more than one-third of the variant nucleotides (21/58). Also of note is the higher prevalence of C/U transitions within this region. Over a 100-nt span between RSE nt 40 and 140, C/U transitions make up 75% (10/15) of the variant nucleotides. This is in stark contrast to stem-loop 2 in which C/U transitions make up only ~25% of the variants. In several cases the C/U variation preserves a base pair. It is tempting to speculate that some of these C/U transitions found in single-stranded regions exist to preserve tertiary interactions with other portions of the viral RNA.

Using the DINAMelt server for the prediction of melting profiles for nucleic acids (accessible from www.MFOLD.bioinfo.rpi.edu), we identified a sequence shortly upstream of the RSV poly(A) site that might interact with the single-stranded face of the RSE between nt 37 and 52. Although preliminary gel shift assays show only a weak interaction in vitro (data not shown), future experiments will address the possibility of an RNA-RNA interaction in vivo. There is a precedent for long-range RNA-RNA interactions in the tombusvirus family; the tomato bushy stunt virus has a translation enhancer in its 3′ UTR that is brought into close proximity to the 5′ end of the RNA through direct RNA-RNA interactions (9, 13, 35). Future experiments will utilize both the two-dimensional structure and information from the sequence conservation described above to design *cis* mutants in hopes of elucidating the mechanism of action of the RSE.

It has become clear lately that there are several different signals that can invoke NMD (1, 6, 7, 29, 30). Here, we describe an RNA stability element that prevents decay of an RNA that might otherwise be expected to undergo NMD due to its abnormal 3′ UTR. We can envision several different mechanisms by which the RSE can act. Perhaps a termination-promoting protein binds to the single-stranded region or stem-loop 1. Perhaps the RSE can physically interact with the 3′ end of viral RNAs, thereby stabilizing termination. A third possibility is that the RSE physically interacts with the ribosome itself or with ribosome-associated factors. Future work will use the structure presented here to address these possible mechanisms.

REFERENCES

1. **Amrani, N., R. Ganesan, S. Kervestin, D. A. Mangus, S. Ghosh, and A. Jacobson.** 2004. A faux 3′-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. Nature **432:**112–118.
2. **Barker, G. F., and K. Beemon.** 1991. Nonsense codons within the Rous sarcoma virus *gag* gene decrease the stability of unspliced viral RNA. Mol. Cell. Biol. **11:**2760–2768.
3. **Barker, G. F., and K. Beemon.** 1994. Rous sarcoma virus RNA stability requires an open reading frame in the *gag* gene and sequences downstream of the *gag-pol* junction. Mol. Cell. Biol. **14:**1986–1996.
4. **Barreau, C., L. Paillard, and H. B. Osbourne.** 2005. AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res. **33:**7138–7150.
5. **Beemon, K. L.** 2008. Retroviruses of birds, p. 455–459. *In* B. W. J. Mahy and M. H. V. Van Regenmortel (ed.), Encyclopedia of virology. Elsevier, Oxford, United Kingdom.
6. **Behm-Ansmant, I., I. Kashima, J. Rehwinkel, J. Sauliere, N. Wittkopp, and E. Izaurralde.** 2007. mRNA quality control: an ancient machinery recognizes and degrades mRNAs with nonsense codons. FEBS Lett. **581:**2845–2853.
7. **Buhler, M., S. Steiner, F. Mohn, A. Paillusson, and O. Muhlemann.** 2006. EJC-independent degradation of nonsense immunoglobulin-μ mRNA depends on 3′ UTR length. Nat. Struct. Mol. Biol. **13:**462–464.
8. **Caldwell, R. B., A. M. Kierzek, H. Arakawa, Y. Bezzubov, J. Zaim, P. Fiedler, S. Kutter, A. Blagodatski, D. Kostovska, M. Koter, J. Plachy, P. Carninci, Y. Hayashizaki, and J. M. Buerstedde.** 2005. Full-length cDNAs from chicken bursal lymphocytes to facilitate gene function analysis. Genome Biol. **6:**R6.
9. **Danthinne, X., J. Seurinck, F. Meulewaeter, M. Van Montagu, and M. Cornelissen.** 1993. The 3′ untranslated region of satellite tobacco necrosis virus RNA stimulates translation in vitro. Mol. Cell. Biol. **13:**3340–3349.
10. **Demeler, B.** 2005. UltraScan version 8.0. A comprehensive data analysis software package for analytical ultracentrifugation experiments. University of Texas Health Science Center, San Antonio, TX.
11. **Eberle, A. B., L. Stadler, H. Mathys, R. Z. Orozco, and O. Muhlemann.** 2008. Posttranscriptional gene regulation by spatial rearrangement of the 3′ untranslated region. PLoS Biol. **6:**e92.
12. **Eulalio, A., E. Huntzinger, and E. Izaurralde.** 2008. Getting to the root of miRNA-mediated gene silencing. **132:**9–14.
13. **Fabian, M. R., and K. A. White.** 2006. Analysis of a 3′-translation enhancer in a tombusvirus: a dynamic model for RNA-RNA interactions of mRNA termini. RNA **12:**1304–1314.
14. **Gatfield, D., L. Unterholzner, F. D. Ciccarelli, P. Bork, and E. Izaurralde.** 2003. Nonsense-mediated mRNA decay in *Drosophila*: at the intersection of the yeast and mammalian pathways. EMBO J. **22:**3960–3970.
15. **Higgins, D. G.** 1994. CLUSTAL V: multiple alignment of DNA and protein sequences. Methods Mol. Biol. **25:**307–318.
16. **Hilleren, P., and R. Parker.** 1999. Mechanisms of mRNA surveillance in eukaryotes. Annu. Rev. Genet. **33:**229–260.
17. **Hurowitz, E. H., and P. O. Brown.** 2003. Genome-wide analysis of mRNA lengths in *Saccharomyces cerevisiae*. Genome Biol. **5:**R2.
18. **Jacks, T., H. D. Madhani, F. R. Masairz, and H. E. Varmus.** 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. Cell **55:**447–458.
19. **LeBlanc, J. J., and K. L. Beemon.** 2004. Unspliced Rous sarcoma virus genomic RNAs are translated and subjected to nonsense-mediated mRNA decay before packaging. J. Virol. **78:**5139–5146.
20. **Marczinke, B., R. Fisher, M. Vidakovic, A. J. Bloys, and I. Brierley.** 1998. Secondary structure and mutational analysis of the ribosomal frameshift signal of Rous sarcoma virus. J. Mol. Biol. **284:**205–225.
21. **Mathews, D. H., M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner.** 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. USA **101:**7287–7292.
22. **Meaux, S., A. van Hoof, and K. E. Baker.** 2008. Nonsense-mediated mRNA decay in yeast does not require PAB1 or a poly(A) tail. Mol. Cell **29:**134–140.
23. **Mendell, J. T., N. A. Sharifi, J. L. Meyers, F. Martinez-Murillo, and H. C. Dietz.** 2004. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat. Genet. **36:**1073–1078.
24. **Meric, C., E., and P. F. Spahr.** 1986. Rous sarcoma virus nucleic acid-binding protein p12 is necessary for viral 70s RNA dimer formation and packaging. J. Virol. **60:**450–459.
25. **Mortimer, S. A., and K. M. Weeks.** 2007. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. J. Am. Chem. Soc. **129:**4144–4145.
26. **Muhlrad, D., and R. Parker.** 1999. Recognition of yeast mRNAs as "nonsense containing" leads to both inhibition of mRNA translation and mRNA

degradation: implications for the control of mRNA decapping. Mol. Biol. Cell **10:**3971–3978.

27. **Pesole, G., F. Mignone, C. Gissi, G. Grillo, F. Licciulli, and S. Liuni.** 2001. Structural and functional features of eukaryotic mRNA untranslated regions. Gene **276:**73–81.

28. **Pfingsten, J. S., and J. S. Kieft.** 2008. RNA structure-based ribosome recruitment: lessons from the *Dicistroviridae* intergenic region IRESes. **14:**1255–1263.

29. **Ruiz-Echevarria, M. J., and S. W. Peltz.** 2000. The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. Cell **101:**741–751.

30. **Singh, G., I. Rebbapragada, and J. Lykke-Andersen.** 2008. A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated decay. PLoS Biol. **6:**e111.

31. **Stoltzfus, C. M., K. Dimock, S. Horikami, and T. A. Fight.** 1983. Stabilities of avian sarcoma virus RNAs: comparison of subgenomic and genomic species with cellular mRNAs. J. Gen. Virol. **64:**2191–2202.

32. **Swanstrom, R., and J. W. Wills.** 1997. Synthesis, assembly, and processing of viral proteins, P. 263–334. *In* J. M. Coffin, S. H. Hughes and H. E. Varmus (ed.), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

33. **Tam, M. F., J. A. Dodd, and W. E. Hill.** 1981. Physical characteristics of 16 S rRNA under reconstitution conditions. J. Biol. Chem. *256:*6430–6434.

34. **Voss, N. R., and M. Gerstein.** 2005. Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. J. Mol. Biol. **346:**477–492.

35. **Wang, S., K. S. Browning, and W. A. Miller.** 1997. A viral sequence in the 3′-untranslated region mimics a 5′ cap in facilitating translation of uncapped mRNA. EMBO J. **16:**4107–4116.

36. **Weil, J. E., and K. L. Beemon.** 2006. A 3′ UTR sequence stabilizes termination codons in the unspliced RNA of Rous sarcoma virus. RNA. **12:**102–110.

37. **Wilkinson, K. A., E. J. Merino, and K. M. Weeks.** 2006. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat. Protoc. **1:**1610–1616.

38. **Wilkinson, K. A., R. J. Gorelick, S. M. Vasa, N. Guex, A Rein, D. H. Mathews, M. C. Giddings, and K. M. Weeks.** 2008. High throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. PLOS Biol. **6:**e96.

39. **Wren, J. D., E. Forgacs, J. W. Fondon 3rd, A. Pertsemlidis, S. Y. Cheng, T. Gallardo, R. S. Williams, R. V. Shohet, J. D. Minna, and H. R. Garner.** 2000. Repeat polymorphisms within gene regions: phenotypic and evolutionary implications. Am. J. Hum. Genet. **67:**345–356.

40. **Zhan, Y., and G. S. Rule.** 2005. Stable triplet of uracil-uracil base pairs in a small antisense RNA. J. Am. Chem. Soc. **127:**15714–15715.

41. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31:**3406–3415.