Evidence that genomic and antigenomic RNA self-cleaving elements from hepatitis delta virus have similar secondary structures

Sarah P. Rosenztein and Michael D. Been*

Department of Biochemistry, Box 3711, Duke University Medical Center, Durham, NC 27710, USA

Received April 18, 1991; Revised and Accepted July 6, 1991

ABSTRACT

The two sequences that define the self-cleaving elements from the genomic and antigenomic RNA of hepatitis delta virus were folded into secondary structures with similar features. Evidence in support of the two models was obtained from limited ribonuclease digestion of genomic and antigenomic RNA fragments containing the sequence 3' of the cleavage site. Under conditions where the rates of self-cleavage are enhanced by addition of 5 M urea (2–10 mM Mg^{2+} at 37°C), ribonucleases T1, U2, A and V1 generated digestion patterns consistent with the proposed RNA structures. The evidence for a relatively stable structure in urea when Mg^{2+} is present suggests that denaturant-enhanced rates of self-cleavage could result from destabilization of competing inactive structures.

INTRODUCTION

The genome of hepatitis delta virus (HDV) consists of a single-stranded circular RNA of approximately 1700 nucleotides (nt) which is 70% self-complementary and can form an unbranched rod-like structure under non-denaturing conditions (1–4). The HDV genomic and antigenomic RNAs have self-cleavage sites located in comparable regions of the two sequences (5). It is hypothesized that HDV replicates by a rolling circle mechanism, and that the self-cleavage reactions associated with both the genomic and antigenomic RNAs are important for this process (5–8). Similar proposals were made for the function of self-cleavage sites in the plant pathogenic RNAs (9–12). Self-cleavage of the HDV sequence requires a divalent cation (5–7) and we have shown that when the concentration of Mg^{2+} is above 0.5 mM, the self-cleavage reaction of the genomic strand of HDV can be enhanced by the addition of denaturants such as urea or formamide (13). The mechanism by which the denaturant enhances cleavage has not been established, although a tertiary interaction important for activity in denaturants has been identified (14).

With the identification of the sites of self-cleavage in HDV genomic and antigenomic RNA, it was evident that the flanking sequences would not fold into either the 'hammerhead' (10, 11) or 'hairpin' (15) motif associated with other self-cleavage sites (6, 7). Although secondary structures for the genomic RNA self-cleavage site have been proposed (7, 16), both structures would appear to be unique to the genomic sequence. Part of the difficulty in establishing the secondary structure is related to the assignment of the boundaries of the genomic and antigenomic self-cleaving elements (5). Recently, it has been shown that a minimal self-cleaving element consists of an 85 nt region of the genomic RNA; nucleotides –1 to +84, relative to the cleavage site, appear to be sufficient for efficient self-cleavage under a wide variety of conditions (17). Similar results were obtained with an antigenomic RNA sequence (14). If the nucleotide 5' to the cleavage site is viewed as part of the substrate, the 'active site' of this RNA must reside in the structure formed by folding the sequence 3' to the cleavage site. We have found that by using the 84 nt sequences 3' to the cleavage site, potential models for the secondary structure of the genomic and antigenomic RNAs of HDV can be generated which are very similar to each other. Two pairings that define a pseudoknot-like interaction have been tested in the antigenomic sequence using site-specific mutagenesis (14). Here we show that for both the genomic and antigenomic sequences, these models are consistent with ribonuclease probing of the RNA in 2–10 mM Mg^{2+} and 5 M urea at 37°C.

MATERIALS AND METHODS

Enzymes, reagents and plasmid DNA

T7 RNA polymerase was purified from an over-expressing clone provided by W. Studier (18). Restriction endonuclease *EcoRI* was a gift from P. Modrich (Duke University). Ribonuclease (RNase) T1 and U2 were obtained from Calbiochem, RNase V1 from Pharmacia and RNase A from Sigma. Other enzymes, nucleotides and ^{32}P -labeled nucleotides were purchased from commercial sources. The construction of plasmids used in this study, pSA1 (antigenomic sequence) and pSD106 (genomic sequence), has been described elsewhere (14, 17).

* To whom correspondence should be addressed

Preparation of 3' cleavage product RNA

Plasmids pSA1 and pSD106 were digested with restriction endonucleases *Bam*HI and *Eco*RI, respectively. The conditions for preparing runoff transcripts were: 20 µg/ml linear plasmid DNA, 40 mM Tris-HCl (pH 7.5), 5 mM dithiothreitol (DTT), 2 mM spermidine, 300 units T7 RNA polymerase/µg DNA, 15 mM MgCl₂ and ribonucleoside triphosphates at 1 mM each. After 60 min at 37°C the reaction was stopped with EDTA and the products fractionated on a 6% polyacrylamide (29:1 acrylamide:bisacrylamide) gel containing 7 M urea, 0.05 M Tris-Borate (pH 8.4) and 0.5 mM EDTA. Approximately 40–50% of the RNA cleaved during transcription. Product RNA (containing the sequence 3' to the cleavage site) was located by UV shadowing, and eluted from the gel in 0.1% sodium dodecyl sulfate/10 mM EDTA at 4°C. SD106 3' cleavage product RNA contained genomic sequence to position 85 3' of the site of cleavage and an additional 18 nt of vector derived sequence (17). SA1 3' cleavage product RNA contained antigenomic sequence to position 87 3' of the cleavage site and an additional 18 nt of vector derived sequence (14).

5' end-labeling of RNA transcripts

Product RNA was labeled at the 5' end in a 10 µl reaction containing 25 pmol RNA, 45 pmol [γ -³²P]-ATP (7000 Ci/mmol), 50 mM Tris-HCl (pH 8.9), 10 mM MgCl₂, 5 mM DTT and 10–15 units of T4 polynucleotide kinase, which was incubated 2 hours on ice (0–4°C) or 1 hour at 37°C, and terminated by addition of EDTA. The products were fractionated on a 6% polyacrylamide/7 M urea gel and the labeled RNA located by autoradiography.

Reactions with ribonucleases

Structure probing reactions (10 µl volume) using RNase T1, U2 or A contained 125,000 cpm (Cerenkov) of 5' end-labeled 3' cleavage product RNA which was preincubated at 37°C for 5 min followed by addition of 1 µl of the diluted ribonuclease (final concentrations are given in the figure legends). Reactions with RNase T1 and RNase A contained 40 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.2 mg/ml yeast tRNA and 0, 3 or 11 mM MgCl₂ with 0 or 5 M urea. In addition, some buffers contained 0.1 mM or 0.5 mM MgCl₂ and were as above except no EDTA was present (see figure legends). RNase U2 reaction buffers were as described for RNase T1 and A except that 20 mM sodium acetate (pH 5.4) was used in place of the Tris-HCl. Reaction buffers for RNase V1 contained 25 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 200 mM NaCl and 0.2 mg/ml tRNA with 0 or 5 M urea. After 15 min at 37°C, an equal volume of 25 mM EDTA/24 M formamide/0.02% each bromphenol blue and xylene cyanol was added and the reactions were put on crushed dry ice. RNase V1 reactions were performed similarly, except that reactions were stopped after either 1 or 2 min.

Sequencing markers were generated by partial digestion of the end-labeled RNAs with ribonuclease T1, U2 or A according to Donis-Keller *et al.*, except that the pH was reduced to 3.5 for the T1 and U2 reactions (19). Alkaline hydrolysis ladders were generated by heating the RNA at pH 9.0 (19) for 5 min at 100°C.

Gel electrophoresis

Sequencing gels contained 0.1 M Tris-Borate (pH 8.4), 1 mM EDTA, and either 8.3 M urea and 12% polyacrylamide (29:1 acrylamide:bisacrylamide) or 7 M urea and 20% polyacrylamide (29:1). X-ray film was exposed to gels at –70°C overnight; an intensifying screen was used.

RESULTS AND DISCUSSION

Potential secondary structures

Using the boundaries of the genomic self-cleaving domain from HDV (17) as a guide, we examined the genomic and antigenomic sequences from nt position –1 to 84, relative to the cleavage site (5), for a potential secondary structure consistent with both sequences. Runs of 5 or more bases in common were identified and aligned, after which the sequences were folded by visual inspection in an effort to maximize similarities in pairing of those sequences. However, we placed emphasis on the possibility that the additional nucleotides at the 3' end of the cleavage domain, which are required for efficient cleavage of the genomic sequence in the presence of denaturant (17), might extend a shorter pairing which is otherwise of marginal stability. Two candidates for such an interaction were apparent in the genomic sequence; one was a pairing between nucleotides 80–84 and 59–63, the other was between nucleotides 79–84 and 11–16 (Stem II, Figure 1A). The latter combination was considered more likely because the potential to form a strong stem and loop structure involving nucleotides 44–73 (Stem IV, Figure 1A) would appear to exclude the former. Nucleotides 81–84 of the antigenomic sequence could form a similar, but non-identical pairing, with nucleotides 16–19 (Stem II, Figure 1B). The structures, as shown (Figure 1A and 1B), contain four paired regions or stems (I–IV), two hairpin loops at the ends of stem III and stem IV, a bulge loop between stem I and stem III, and two additional unpaired regions joining stem I to stem IV and stem IV to stem II. Identical sequences and pairings are boxed to facilitate visual comparison and to draw attention to overall similarities and differences. An alternative pairing for stem IV of the antigenomic sequence is shown (Figure 1B, left), however, the ribonuclease data presented below is less consistent with that version than with the stem in the contiguous structure (Figure 1B, right). Evidence for stems

Figure 1. Potential secondary structures for HDV genomic (A) and antigenomic RNA (B) sequences used in this study. The sequences are numbered 5' to 3' from the cleavage site. Watson-Crick basepairs are indicated by dashes, a G-U pair by a dot, and the continuity of the sequence by solid lines or boxes. The boxed nucleotides are regions of exact similarity between the two sequences. An alternative potential pairing for stem IV of the antigenomic sequence is also shown. Two changes to the natural HDV sequence were made in the synthetic antigenomic sequence (14): A60 was changed to a U and base C59 was deleted (triangle).

I and II in the antigenomic sequence, obtained by site specific base changes, is reported elsewhere (14), and there is similar evidence for stem III (A. Perrotta and M.B., unpublished data).

Probing the secondary structure of the 3' cleavage product with ribonucleases

Strategy and rationale. The secondary structure of the RNA associated with the genomic and antigenomic self-cleaving elements can be examined using ribonucleases that are sensitive to RNA structure (20, 21). Self-cleavage of the HDV sequences requires Mg^{2+} or another divalent cation (5–7), and the rate and extent of cleavage can be enhanced by adding a denaturant (5 M urea or 10 M formamide) to the reaction (13). It would therefore be most informative to examine the structure under those same conditions. It was also of interest to study the effect of Mg^{2+} and urea concentration on those aspects of the structure that may be revealed in the ribonuclease probing experiments. However, an obvious problem arises in examining the structure of a self-cleaving RNA under conditions where it rapidly cleaves. Therefore, rather than using end-labeled precursors in the following ribonuclease probing experiments, the 3' self-cleavage product was 5' end-labeled and utilized. We felt this was justified

because the catalytic domain would appear to reside entirely in the sequence 3' to the cleavage site (14, 17). The usefulness of this approach was established with structural studies of the group I intron from *Tetrahymena thermophila* pre-rRNA (22, 23).

Ribonuclease T1. RNase T1 (24) cuts after unpaired guanosine residues (20). The 5' end-labeled RNAs were digested with RNase T1 in 0 and 5 M urea with increasing Mg^{2+} concentrations. The genomic HDV sequence is rich in G residues, however, under those conditions where self-cleavage activity is strongly enhanced, 2 or 10 mM Mg^{2+} and 5 M urea (13), only G25 and G59 were readily susceptible to cutting by T1 (Figure 2, lanes 18–21). In the proposed model, these nucleotides are in hairpin loops III and IV, respectively. The two weak cut sites at G39 and G40 would fall in the unpaired region leading to stem IV in the proposed structure. In 5 M urea, the effect of magnesium on the structure was blatantly apparent at the 3' end of the RNA: in the absence of magnesium, G residues 73, 74, 76 and 80–82 were very susceptible to RNase T1, however, when magnesium was present, these bases became protected. In the proposed model, G80–82 are involved in the formation of stem II. Potential discrepancies with the proposed structure were the weak T1 cuts after G29 and G31, both of which are shown in paired regions (stem III and I). These residues appeared to be more susceptible to T1 cutting in the absence of urea (compare lanes 9–12 with 18–21, Figure 2).

With the antigenomic 3' cleavage product RNA, RNase T1 also cut at only certain G residues, although the difference between major and minor hits was less evident than with the genomic sequence (Figure 3). In the presence of 2 or 10 mM

Figure 2. Ribonuclease T1 probing of the genomic sequence. Reactions with RNase T1 contained either no urea (lanes 5–12) or 5 M urea (lanes 14–21). Except as noted below, each reaction contained 1 mM EDTA and either 0, 3, or 11 mM $MgCl_2$ to give approximate free Mg^{2+} concentrations as indicated in the figure. Reactions shown in lanes 7, 8, 16 and 17 contained 0.1 mM $MgCl_2$ with no EDTA present. For each set of conditions, two different RNase T1 concentrations were used: 5×10^{-5} units/ μ l (L) or 10×10^{-5} units/ μ l (H) (lanes 5–12); 1×10^{-2} units/ μ l (L) or 2×10^{-2} units/ μ l (H) (lanes 14–19); 0.5×10^{-2} units/ μ l (L) or 1×10^{-2} units/ μ l (H) (lanes 20 and 21). Random cleavage of the RNA was generated by alkaline hydrolysis (OH; lanes 2, 13 and 22). Sequencing markers were: lane 1, no enzyme (C); lane 3, G ladder (T1); lane 4, A ladder (U2). Samples were heated to 95°C for 2 min and analyzed by electrophoresis on a 12% polyacrylamide gel containing 8.3 M urea.

Figure 3. Ribonuclease T1 probing of the antigenomic sequence. Reactions were as described in Figure 2 except that lanes 7, 8, 16 and 17 contained 0.5 mM $MgCl_2$ instead of 0.1 mM (no EDTA present). Samples were analyzed as described in Figure 2.

Mg^{2+} and 5 M urea (Figure 3, lanes 18–21) T1 cut at nt 11 and 12 (which would fall in the single-stranded region between stem I and stem II, Figure 1B), at nt 28 (hairpin loop III), at nt 51 and 52 (the interior loop in stem IV) and at nt 58 (hairpin loop IV). In the absence of urea there was light cutting at nt 40–42 (unpaired region between stem I and IV), however, digestion at these positions was somewhat suppressed in the presence of urea. Suppression of T1 cuts at G's at the 3' end of the antigenomic RNA (nt 80–82) by magnesium was not as dramatic as in the genomic RNA, but was still apparent.

The sequence of both the genomic and antigenomic self-cleaving elements is G-rich (~35% G) and the lack of T1 cuts at most of these positions in 5 M urea and Mg^{2+} indicates that the RNA is highly structured. If the secondary structures are correct, the lack of strong cut sites after G's which are shown as unpaired in the model could be due to either unidentified tertiary interactions or to inaccessibility of those positions to the ribonuclease. The positions and a qualitative summary of T1 and other RNase cut sites in the genomic and antigenomic sequences in urea and Mg^{2+} were tabulated (Table I).

Figure 4. Ribonuclease U2 probing of the antigenomic sequence. Reactions with RNase U2 contained either no urea (lanes 5–8) or 5 M urea (lanes 10–13). Except as noted below, each reaction contained 1 mM EDTA and either 0, 3 or 11 mM $MgCl_2$ to give approximate free Mg^{2+} concentrations as indicated in the figure. Reactions shown in lanes 6 and 11 contained 0.5 mM $MgCl_2$ with no EDTA present in the reaction. The final concentration of RNase U2 in each reaction was 5×10^{-4} units/ μ l (lanes 5–8) or 5×10^{-3} units/ μ l (lanes 10–13). Lanes 1–4 were the same as in Figure 2; additional hydrolysis ladders were run in lanes 9 and 14. Samples were analyzed as described for Figure 2.

Ribonuclease U2. RNase U2 (25) shows specificity for cutting after unpaired adenosine residues although some unpaired guanosine residues are also accessible (21, 26). The antigenomic 3' product RNA was probed with this enzyme (Figure 4). As was seen with RNase T1, very few residues which correspond to paired regions in the proposed structure were readily cut with RNase U2 in urea when Mg^{2+} was present. Cuts at A9 and A14 would fall in the single-stranded region between stems I and III (Figure 1B) and were not suppressed by Mg^{2+} (Figure 4, lanes 10–13). However, the susceptibility to cutting should not be overestimated, as the RNA in some of those lanes is overdigested. In the presence of urea, there were several positions at which cutting by U2 was suppressed by the addition of Mg^{2+} . An obvious one was A20 (stem III), but others occurred at A44, A49, and A50 (stem IV). A56 (hairpin loop IV) appeared to be susceptible to attack in urea under all Mg^{2+} concentrations suggesting that it may not pair with U63. In the construction of the antigenomic sequence, a nucleotide from hairpin loop IV was deleted (14); that change may tend to destabilize that structure.

Two discrepancies with the model could be seen with cuts at A36 (stem I), and A83 (stem II), both of which would be in proposed paired regions. There is independent evidence from

Figure 5. Ribonuclease A probing of the antigenomic sequence. Reactions with RNase A contained either no urea (lanes 6–11) or 5 M urea (lanes 13–18). Each reaction contained 1 mM EDTA and either 0, 3 or 11 mM $MgCl_2$ to give approximate free Mg^{2+} concentrations as indicated in the figure. For each set of conditions, two different RNase A concentrations were used: 1×10^{-5} μ g/ μ l (L) or 2.5×10^{-5} μ g/ μ l (H) (lanes 6–9, 13 and 14); 5×10^{-5} μ g/ μ l (L) or 1×10^{-4} μ g/ μ l (H) (lanes 10, 11 and 15–18). Lanes 1–4 were the same as in Figure 2; lane 5 contained markers from RNase A digestion (C, U ladder) and lanes 12 and 19 contained hydrolysis ladders. Samples were analyzed as described for Figure 2.

mutagenesis of the antigenomic sequence to indicate that pairing between A36 and U4, and A83 and U17 contributes to optimal self-cleavage activity (14). While these two examples do not invalidate the general usefulness of ribonuclease probing, they do point out the importance of multiple approaches to secondary structure determination. It is possible that the potential juxtaposition of stems I, II, and III, in an unusual structure, could contribute to nuclease sensitivity in this region.

Experiments with the genomic RNA using RNase U2 were done in a similar manner (data not shown). The major site of U2 cutting in this sequence was at A56, which would fall in hairpin loop IV of the proposed structure (Table I).

Ribonuclease A. RNase A will cut after unpaired pyrimidine residues and has a preference for cytidine residues, especially in the sequence context CpA (21). Cutting by RNase A in the antigenomic sequence (Figure 5) at C8, C13 and C19 may have been enhanced because each is followed by an A. However, the CpA sequence at position 43 was protected from RNase A in the presence of urea when Mg^{2+} was added (lanes 15–18). In the proposed structure, residues C8 and C13 are in the unpaired region between stems I and III while C19 is at one end of stem II, which may render it accessible to the ribonuclease. The extent of cutting by RNase A was difficult to control and several of the lanes shown in Figure 5 were overdigested. This resulted

Table I. Results of ribonuclease probing in 5 M urea and $MgCl_2$ at 37°C.

A: Genomic RNA					B: Antigenomic RNA									
nt	T1	U2	A	V1	nt	T1	U2	A	V1	nt	T1	U2	A	V1
G1	-	-		-	A43		+		-	G1	+b	-		-
G2	-	-		+	C44			++	-	G2	+b	-		-
C3			a	++	A45		-		-	G3	+b	-		+
C4			a	++	U46			-	-	U4			+b	++
G5	-	-		+	U47			-	-	C5			+b	++
G6	-	-		+	C48			-	-	G6	+b	-		+
C7			-	+	C49			-	++	G7	+b	+b		+
A8		+		+	G50	-	-		++	C8			++b	+
U9			-	-	A51				++	A9		++b		+
G10	-	-		-	G52	-	-		++	U10			+b	+
G11	-	-		+	G53	-	-		+	G11	++b	+b		-
U12			-	+	G54	-	-		-	G12	++b	+b		-
C13			-	+	G55	-	-		-	C13			++b	-
C14			-	++	A56		++		-	A14		++b		+
C15			-	++	C57			+	-	U15			-	+
A16		-		++	C58			+	-	C16			-	+
G17	-	-		-	G59	++	-		-	U17			-	++
C18			-	-	U60			+	-	C18			+b	++
C19			-	-	C61			-	-	C19			++b	+
U20			-	-	C62			-	-	A20		+/-		+
C21			-	-	C63			-	-	C21			+	+
C22			-	-	C64			-	-	C22			-	+
U23			-	-	U65			-	-	U23			-	+
C24			-	-	C66			-	-	C24			-	+
G25	++	+		+	G67	-	-		-	C25			-	+
C26			-	+	G68	-	-		+	U26			-	+
U27			-	+	U69			-	-	C27			+	-
G28	-	-		-	A70			-	-	G28	++	-		-
G29	+	-		-	A71			-	-	C29			+	-
C30			-	-	U72			-	-	G30	-	-		-
G31	+	-		-	G73	+	-		-	G31	-	-		-
C32			-	++	G74	+	-		-	U32			-	+
C33			-	++	C75			+	-	C33			-	++
G34	-	-		-	G76	+	-		-	C34			-	++
G35	-	-		++	A77		+		-	G35	+	-		-
C36			-	-	A78		+		-	A36		+		-
U37			-	-	U79			-	-	C37			-	+
G38	-	-		-	G80	+	-		-	C38			-	+
G39	+	-		-	G81	+	-		++	U39			-	-
G40	+	-		-	G82	+	-		-	G40	+	-		-
C41			++	-	A83		-		-	G41	+	-		-
A42		-		-	C84			-	-	G42	+	-		-

A qualitative summary of the sites of RNase cutting and protection is shown for RNase T1, U2, A and V1 under conditions of 5 M urea and Mg^{2+} . The results are from reactions performed as in Figures 2–6 and analyzed on 12% polyacrylamide gels containing 8.3 M urea or 20% polyacrylamide gels containing 7 M urea. Nucleotides (nt) were judged to be either not digested (–), weakly digested (+) or strongly digested (++) by the particular enzyme.

^aNot determined.

^bsmaller bands may be overrepresented due to overdigestion of the RNA under some conditions.

in secondary cutting, and the effect was most apparent in lanes 15 and 16 where extensive cutting has occurred after C5. Again, mutagenesis implicates C5 in a base-pairing interaction (14).

In the genomic RNA, RNase A generated strong cuts at C41 and C44, which would be in the region predicted to be unpaired between stems I and IV and at the base of stem IV, respectively. C57 and C58 (hairpin loop IV) were weakly cut by RNase A in the presence of magnesium but not in its absence (data not shown; see Table I).

Figure 6. Ribonuclease V1 probing of the genomic (A) or antigenomic (B) sequence. Reactions with RNase V1 contained either no urea (lanes 5 and 6) or 5 M urea (lanes 7 and 8). The incubation times were 1 or 2 min, as indicated. The final concentration of RNase V1 used in all reactions was 0.01 units/ μ l. Lanes 1–4 were as described in Figure 2. Samples were analyzed on a 12% polyacrylamide gel containing 8.3 M urea (top) and a 20% gel containing 7 M urea (bottom). The asterisks (*) indicate two G's in the sequence that are overlapping on the different percentage gels. Note that the V1 products contain a 3' hydroxyl rather than phosphate; this results in a decreased mobility relative to the sequencing markers.

Ribonuclease V1. Ribonuclease V1 from cobra venom (27) recognizes portions of the sugar phosphate backbone that are in somewhat helical conformations (28) and therefore can be used to detect residues that are either base-paired or stacked. RNase V1 cutting of the antigenomic 3' product RNA was limited to discrete regions (Figure 6B). Residues which would correspond to both the 5' and 3' sides of stem I were susceptible to the enzyme (lanes 7 and 8), with the major sites at nt positions 4, 5, 33 and 34 and minor sites at nt.3, 6, 7, 37 and 38. Sequences involved in the proposed stem II interaction were also cut by RNase V1, specifically at nt 15–19 and nt 84, with strong cuts at nt 17 and 18. As with stem II, residues which are proposed to generate stem III were susceptible to cutting with this enzyme (nt 19–21 and nt 32). If stem-loop III exists, the V1 cuts at residues 23–26 could mean that the bases in the loop are stacked or paired, and accessible to RNase V1. Sequences that may form stem IV were not as accessible to the enzyme as other regions of the RNA; however, there were some RNase V1 sites at nt 44–48, and weaker cuts at nt 72 and 73.

With the genomic RNA fragment, strong RNase V1 cuts at nt 3, 4, 32, 33 and 35 (Figure 6A, lanes 7 and 8) would be consistent with stem I. This was also the case for stem II, with cuts at nt 11–16 and also at nt 81. The sequence forming the proposed stem III pairing was not susceptible to the enzyme under the conditions used, while that corresponding to stem IV had strong cuts at residues 49–52.

Evaluation of the proposed structures

The method of choice for generating and testing RNA secondary structure is a phylogenetic comparison of homologous RNAs (11, 29–33). However, in this case, when applied to only two RNAs for which the sequence similarities could be due to a property of the structure of the genome (internal complementarity), such

Figure 7. Major cut sites of ribonucleases T1, U2, A and V1 on SA1 and SD106 3' cleavage product RNA in 5 M urea and Mg^{2+} . The results from experiments shown in Figures 2–6 and others not shown are summarized on the proposed secondary structures for the genomic (A) and antigenomic (B) RNAs. Only those RNase sensitive sites qualitatively judged to be strongly cut under the conditions stated are shown.

an approach would not be convincing. Having identified minimal self-cleaving elements for both the genomic and antigenomic sequences which are identical in size and relation to the cleavage site (14, 17), it seemed reasonable that a similar structure might exist. The structures shown (Figure 1) were generated by visually aligning and pairing common sequences, although related structures have since been generated with the Zuker folding program (34) (data not shown). The program generated structures that contained pairings similar to those shown in Figure 1 but differed in that stem II, as a tertiary interaction, was not predicted. The bases involved in that interaction were predicted to be either unpaired or involved in alternative interactions.

Two experimental approaches have been used to test the models. Site specific mutagenesis of an antigenomic sequence demonstrated a role for pairing of specific bases in stems I and II (14) and in stem III (A. Perrotta and M.B., unpublished results). The approach we used here was ribonuclease probing. Both methods have their limitations. Mutagenesis combined with compensatory changes can be used to test specific basepairing interactions only if the function of the RNA is not dependent on the identity of the basepair. Another problem arises in identifying the original functional structure when compensatory mutations compatible with alternative structures or alignments of the sequences are both functional (35). With ribonuclease probing, the data must be interpreted cautiously because, as documented with tRNAs (36), factors affecting specificity of these enzymes are not completely understood. In addition, since the ribonucleases may not have access to all residues which are unpaired, one cannot draw specific conclusions about the structure at sites which are not cut. Nevertheless, ribonuclease probing allowed the examination of the general structure of both the genomic and antigenomic sequences and also revealed structural changes induced by denaturants and Mg^{2+} .

As revealed by ribonuclease digestion, the Mg^{2+} had a dramatic effect on structure; this effect was exaggerated in the presence of 5 M urea (Figure 2). Details of the digestion patterns could not be reproduced by substituting polyamines (spermidine and spermine) or monovalent salts (NaCl) for Mg^{2+} (S.R., unpublished data). The major cut sites in 5 M urea and Mg^{2+} when compiled on the proposed structures (Figure 7) appear to be compatible with the secondary structures that we propose. If the paired elements of the structures are correct, the conspicuous lack of cutting within regions drawn as unpaired suggests a compact tertiary structure and additional unidentified interactions. Although sites of minor cutting could represent residues which are susceptible but less accessible to the ribonuclease, they could also result when a fraction of the RNA is folded into alternative structures, or, in those reactions which were overdigested, where secondary cutting has occurred. Therefore, the significance of minor cut sites should be interpreted with caution.

Are there other secondary structures consistent with the nuclease data? The possibility for an alternative pairing in the genomic sequence of nt 80–84 (stem II) with nt 59–63 (stem IV) had been considered. The extreme sensitivity of G59 to T1 makes this potential pairing unlikely. In the antigenomic sequence, the sensitivity of G52 (as well as G51) make the alternative arrangement shown in Figure 1 less probable. However, the exact alignment of sequences in stem IV may not be an important issue since the deletion of much of stem IV in the antigenomic sequence results in increased activity (A. Perrotta and M.B., unpublished data).

Potential secondary structures for the sequence at the genomic

self-cleavage site have been proposed by two other groups (7, 16). Both of those structures include sequences 5' to the cleavage site, so applying our nuclease data to those structures is difficult. However, in regions where the structures differ from the one we propose, some discrepancies with the data presented here appear to exist. For example, G25 is one of the two major RNase T1 cut sites, and in both of the other proposed structures, G25 is base-paired.

Some features of the secondary structures are worth emphasizing. Of the potential basepairs in both the genomic and antigenomic sequences, a large percentage involve G-C interactions (75% and 66%, respectively). The extensive G-C pairing may contribute to the overall stability of these structures. The combination of stems I, II and III would generate a pseudoknot-like structure (37) with a tertiary interaction involving a bulge loop sequence between stems I and III and the 3' end of the element. In that combination, stem III could have the potential to stack on either stem I or stem II. The potential for a G-U pair is found at the cleavage site, and while we have no direct evidence for this interaction, its position adjacent to the cleavage site could be important for optimal activity. In both sequences it may be possible to extend stem I to include the nucleotide 5' of the cleavage site with either a U-G (genomic) or a C-G (antigenomic) basepair involving the nucleotide at position -1. The G in either of these potential basepairs (genomic-G38; antigenomic-G40) is not readily cut by RNase T1, but this is in the absence of a nucleotide at position -1. Putting either G or A at the -1 position in the antigenomic sequence resulted in rates of cleavage of about 14% and 130% of the wild-type level, respectively (A. Perrotta and M.B., unpublished data), only indicating that a pyrimidine at -1 is not critical.

The self-cleaving RNA from HDV will cleave in relatively high concentrations of urea (≤ 8 M) or formamide (≤ 18 M); in fact, the rate of cleavage of certain transcripts of the genomic sequence is enhanced by moderate levels of these denaturants (5–7 M urea and 10–15 M formamide) (13). Although a kinetic mechanism that identifies the rate limiting step of the cleavage reaction in the absence of denaturants has not been established, one possible explanation for the denaturant enhanced rate of cleavage is that the denaturant facilitates the folding of the RNA into the active structure. This could occur by increasing the rate of interconversion between various structures, with the active form rapidly cleaving once it is formed. Alternatively, the denaturants could alter the equilibrium in favor of the active self-cleaving structure. Subtle differences in the patterns of fragments produced in the absence and presence of denaturants suggest that it could be the latter. In the presence of urea, there appear to be fewer minor hits relative to what is seen in the absence of urea (for example, compare lanes 12 and 13 with lanes 7 and 8 in Figure 4; also apparent in Figure 2). This could be interpreted to mean that, in the presence of Mg^{2+} and urea, fewer conformations of the folded RNA exist. If the denaturant acted to facilitate interconversion of various structures without changing the equilibrium distribution, one might expect to see either no change or perhaps even an increase in the number of residues that become accessible to the ribonuclease when urea was added. Regardless of the actual enhancement mechanism, it is clear that the RNA forms a highly structured molecule under these conditions of urea and Mg^{2+} . The combination of denaturant and Mg^{2+} may prove to be of some general use in examining the structures or folding properties of RNA.

ACKNOWLEDGEMENTS

We thank A. Kendall, I. Rosenstein and A. Perrotta for valuable comments on the manuscript and A. Perrotta for useful discussions of this work. This work was supported by a grant from the NIH (GM-40689). M.B. was supported by a Junior Faculty Research Award from the American Cancer Society (JFRA-233).

REFERENCES

- Wang, K.-S., Choo, Q.-L., Weiner, A.J., Ou, J.-H., Najarian, R.C., Thayer, R.M., Mullenbach, G.T., Denniston, K.J., Gerin, J.L. and Houghton, M. (1986) *Nature* **323**, 508–514.
- Kos, A., Dijkema, R., Amberg, A.C., van der Meide, P.H. and Schellekens, H. (1986) *Nature* **323**, 558–560.
- Makino, S., Chang, M.-F., Shieh, C.-K., Kamahora, T., Vannier, D.M., Govindarajan, S. and Lai, M.M.C. (1987) *Nature* **329**, 343–346.
- Kuo, M.Y.-P., Goldberg, J., Coates, L., Mason, W., Gerin, J. and Taylor, J. (1988) *J. Virol.* **62**, 1855–1861.
- Kuo, M.Y.-P., Sharmeen, L., Dinter-Gottlieb, G. and Taylor, J. (1988) *J. Virol.* **62**, 4439–4444.
- Sharmeen, L., Kuo, M.Y.-P., Dinter-Gottlieb, G. and Taylor, J. (1988) *J. Virol.* **62**, 2674–2679.
- Wu, H.-N., Lin, Y.-J., Lin, F.-P., Makino, S., Chang, M.-F. and Lai, M.M.C. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1831–1835.
- Taylor, J.M. (1990) *Cell* **61**, 371–373.
- Branch, A.D. and Robertson, H.D. (1984) *Science* **223**, 450–455.
- Hutchins, C.J., Rathjen, P.D., Forster, A.C. and Symons, R.H. (1986) *Nucl. Acids Res.* **14**, 3627–3640.
- Forster, A.C. and Symons, R.H. (1987) *Cell* **49**, 211–220.
- Symons, R.H. (1989) *TIBS* **14**, 445–450.
- Rosenstein, S.P. and Been, M.D. (1990) *Biochemistry* **29**, 8011–8016.
- Perrotta, A.T. and Been, M.D. (1991) *Nature* **350**, 434–436.
- Hampel, A., Tritz, R., Hicks, M. and Cruz, P. (1990) *Nucl. Acids Res.* **18**, 299–304.
- Belinsky, M.G. and Dinter-Gottlieb, G. (1991) *Nucl. Acids Res.* **19**, 559–564.
- Perrotta, A.T. and Been, M.D. (1990) *Nucl. Acids Res.* **18**, 6821–6827.
- Davanloo, P., Rosenberg, A.H., Dunn, J.J. and Studier, F.W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2035–2039.
- Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977) *Nucl. Acids Res.* **4**, 2527–2538.
- Ehresmann, C., Baudin, F., Mougé, M., Romby, P., Ebel, J.-P. and Ehresmann, B. (1987) *Nucl. Acids Res.* **15**, 9109–9128.
- Knapp, G. (1989) In Dahlberg, J.E. and Abelson, J.N. (eds.), *Methods in Enzymology*. Academic Press, Inc., San Diego, Vol. 180, pp. 192–212.
- Cech, T.R., Tanner, N.K., Tinoco, I., Jr., Weir, B.R., Zuker, M. and Perlman, P.S. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3903–3907.
- Inoue, T. and Cech, T.R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 648–652.
- Sato, K. and Egami, F. (1957) *J. Biochem. (Tokyo)* **44**, 753–767.
- Arima, T., Uchida, T. and Egami, F. (1968) *Biochem. J.* **106**, 601–607.
- Uchida, T., Terukatsu, A. and Egami, F. (1970) *J. Biochem. (Tokyo)* **67**, 91–102.
- Vasilenko, S.K. and Rait, V.K. (1975) *Biokhimiya* **40**, 578–583.
- Lowman, H.B. and Draper, D.E. (1986) *J. Biol. Chem.* **261**, 5396–5403.
- Fox, G.E. and Woese, C.R. (1975) *Nature* **256**, 505–507.
- Noller, H.F. and Woese, C.R. (1981) *Science* **212**, 403–411.
- Michel, F., Jacquier, A. and Dujon, B. (1982) *Biochimie* **64**, 867–881.
- Davies, R.W., Waring, R.B., Ray, J.A., Brown, T.A. and Scazzocchio, C. (1982) *Nature* **300**, 719–724.
- James, B.D., Olsen, G.J., Liu, J. and Pace, N.R. (1988) *Cell* **52**, 19–26.
- Devereux, J., Haeblerli, P. and Smithies, O. (1984) *Nucl. Acids Res.* **12**, 387–395.
- Been, M.D. and Cech, T.R. (1987) *Cell* **50**, 951–961.
- Swerdlow, H. and Guthrie, C. (1984) *J. Biol. Chem.* **259**, 5197–5207.
- Pleij, C.W.A. (1990) *TIBS* **15**, 143–147.