1314–1321 *Nucleic Acids Research, 1996, Vol. 24, No. 7*

Core sequences and a cleavage site wobble pair required for HDV antigenomic ribozyme self-cleavage

Anne T. Perrotta and Michael D. Been*

Department of Biochemistry, Box 3711, Duke University Medical Center, Durham, NC 27707, USA

Received November 30, 1995; Revised and Accepted February 12, 1996

ABSTRACT

The secondary structures proposed for the *cis*-acting hepatitis delta virus (HDV) ribozymes contain four duplex regions, three sequences joining the duplexes and two hairpin loops. The core and active site of the ribozyme could be formed by portions of the joining sequences, J1/4 and J4/2, together with one of the hairpin loops, L3. To establish the core region and define essential bases within this putative active site 28 single base changes at 15 positions were made and tested for effects on ribozyme cleavage. At 14 of the 15 positions all of the changes resulted in detectable decreased rates of cleavage. At seven of the positions one or more of the changes resulted in a 500-fold or greater decrease in the observed rate constant for cleavage. Mutations that resulted in 10^3^-fold effects were found in all three regions hypothesized to form the core. At the cleavage site substitutions of the cytosine 5′ of the site of cleavage did not provide strong support for a sequence-specific interaction involving this nucleotide. In contrast, an A·C combination was the most effective substitution for a potential G·U pair 3′ of the cleavage site, suggesting a requirement for a wobble pair at that position.

INTRODUCTION

The self-cleaving RNA sequences (ribozymes) found in the genomic and antigenomic RNAs of hepatitis delta virus (HDV) (1–3) adopt a novel structural motif (4,5) distinct from the previously defined hammerhead and hairpin ribozymes found predominantly in plant pathogenic RNAs (6–8). In the HDV ribozymes, which are about twice the size of a minimal hammerhead ribozyme, the cleavage site is located at the 5′ end of the sequence defining the self-cleaving element (3,9,10). Consequently, all but one nucleotide of the self-cleaving sequence is 3′ of the cleavage site.

A secondary structure containing four paired regions (P1–P4) forming a pseudoknot (Fig. 1) is supported by several lines of evidence. First, the genomic and antigenomic ribozymes display variations in sequence that maintain the potential for base pairing in the four duplex regions of the proposed structures (4,5). More significant, however, a dependence for base pairing in these proposed duplex regions is demonstrated by the effect of base changes and compensatory base changes on the self-cleavage activity in both the antigenomic (5,9,11) and genomic sequences (12). In those studies mutations that would be predicted to destabilize pairing in P1, P2 and P3 reduced the rate or extent of self-cleavage in both the antigenomic (9) and genomic sequences (12). In the genomic sequence deletions in P4 that left part of the stem intact have resulted in decreased, but not total loss of, activity (12,13). The complete deletion of P4 resulted in a 100-fold decrease in activity in the antigenomic ribozyme, but shortening P4 to 4 bp and replacement of loop L4 with a stabilizing UUCG tetraloop sequence enhanced self-cleavage activity (9). In that case it could be demonstrated that pairing in the remaining part of P4 contributed to activity. These data have been interpreted as evidence for a structural role of P4 (9,11). In a bimolecular reaction in which the 5′-side of P1 is provided as substrate to the ribozyme *in trans* it was demonstrated that base pairing in P1 provides specificity in substrate cleavage (14). Thus there is support for functioning of the base paired regions in specifying the cleavage site and in stabilizing an essential ribozyme structure. Much less information is available for those regions of unknown structure which are represented in the model as single-stranded but nevertheless are likely to participate in interactions essential for formation of the active site of the ribozyme.

There are three single-stranded joining sequences (J1/2, J1/4 and J4/2) and two hairpin loops (L3 and L4) in these structures. That active forms of the ribozyme can be generated from two fragments in which J1/2 is deleted (3,9,14,15) or where L4 is interrupted or deleted (9,11,16,17) suggests that L4 and J1/2 are unlikely to participate directly in forming the core structure. In this paper we focus on J1/4, J4/2 and L3 and present evidence from site-specific mutagenesis demonstrating that the identity of bases at most of the positions in those sequences contribute to full activity of the antigenomic ribozyme. These data on the antigenomic ribozyme are in close agreement with a careful mutagenesis study of the J1/4 and J4/2 regions of the genomic ribozyme (18) and do not contradict other studies of the genomic ribozyme sequence (12,13,19,20). Together the mutagenesis results provide compelling evidence that the active sites of the two ribozymes are very similar. In addition, evaluation of mutations and compensatory changes at the cleavage site provides evidence for a wobble base pair requirement at the cleavage site.

* To whom correspondence should be addressed

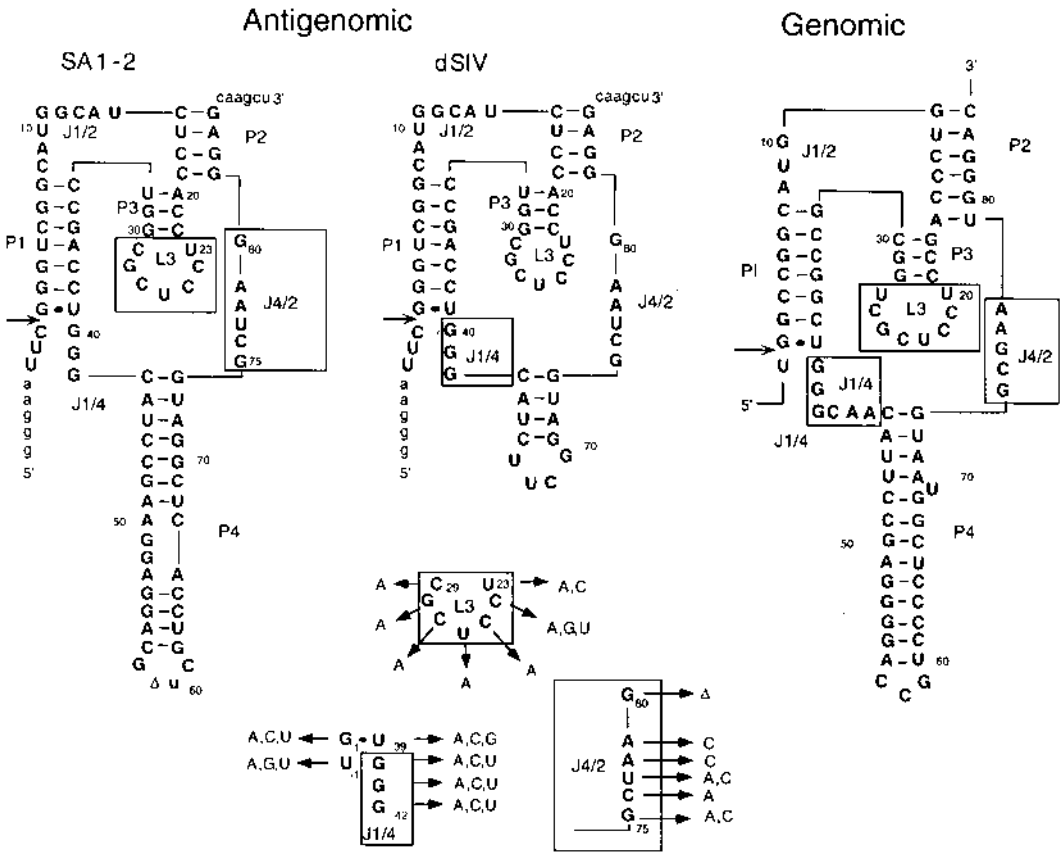

Figure 1. Secondary structures of the HDV antigenomic and genomic *cis*-acting ribozymes (5). Boxed regions in SA1-2 and dSIV indicate the sequences under consideration in the two antigenomic ribozymes used in this study. The base changes that were tested are indicated below the secondary structure. For comparison the equivalent areas of the genomic ribozyme are also boxed.

MATERIALS AND METHODS

Enzymes, chemicals and oligonucleotides

T7 RNA polymerase was purified by M.Puttaraju from an over-expressing clone provided by W.Studier (21). Modified T7 DNA polymerase (Sequenase) was purchased from US Biochemical (Cleveland). Restriction endonucleases, nucleotides, ^{32}P-labeled nucleotides and chemical reagents were purchased from commercial suppliers. Oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer (Department of Botany, Duke University, Durham, NC).

Plasmids and construction of mutants in the antigenomic ribozymes

Two versions of the wild-type antigenomic ribozyme sequence were used. Both sequences were cloned into the T7 promoter-containing phagemid vector pTZ18U. The parent plasmid pSA1-2 was constructed with a synthetic version of the sequence of the antigenomic ribozyme inserted downstream of a T7 promoter (5) and was nearly identical to the wild-type antigenomic self-cleaving sequence except for minor changes that introduced a restriction recognition site in the sequence constituting stem–loop 4. The ribozyme in pdSIV (9) contained a shortened P4 with only 4 bp and a UUCG hairpin tetraloop. The mutants of pSA1-2 and pdSIV were generated by oligonucleotide-directed mutagenesis using a uracil-containing single-stranded form of the plasmid as the template (22,23). The mutagenic oligonucleotides generated a mismatch when annealed to the strand of DNA of the same sense as the transcript. Plasmids with the mutation were identified by sequencing miniprep DNA using primer extension with Sequenase (US Biochemical) and dideoxynucleotide chain terminators. Following a second round of transformation plasmid DNA was prepared from overnight cultures and purified by CsCl equilibrium density centrifugation in the presence of ethidium bromide (24). All purified plasmid DNA was again sequenced before use as templates in transcriptions.

Transcriptions

Plasmid DNA was linearized by digestion with *Hin*dIII, extracted with phenol and chloroform, precipitated with ethanol and transcribed in 0.05 ml reactions containing 40 mM Tris–HCl, pH 7.5, 15 mM $MgCl_2$, 5 mM dithiothreitol, 2 mM spermidine, ribonucleoside triphosphates at 1 mM each, 0.05 mCi [α-^{32}P]CTP, 2.5 µg linear plasmid DNA and 300 U T7 RNA polymerase. Incubation was for 60 min at 37°C, EDTA was added to 50 mM, formamide to 50% (v/v) and the RNA was fractionated by electrophoresis on a 6% (w/v) polyacrylamide gel containing 7 M urea. RNA was located by autoradiography, excised, eluted overnight at 4°C (in 1 mM EDTA, 0.1% w/v SDS) and recovered by ethanol precipitation. The self-cleavage reaction

requires only low levels of a divalent cation and typically transcripts SA1-2 and dSIV cleaved extensively during synthesis under the above conditions. To increase the fractional yield of uncleaved RNA an oligodeoxyribonucleotide complementary to the 5′-region of the ribozyme RNA was added (2 pmol/µl) during transcription to inhibit self-cleavage. Although this resulted in a lower overall yield of RNA, it increased the ratio of uncleaved to cleaved RNA. Most of the mutated ribozymes cleaved slowly, so a blocking oligodeoxyribonucleotide was not needed in the transcription to obtain precursor RNA in high yields.

Self-cleavage reactions

Radiolabeled precursor RNA was preincubated at 37°C for 5 min in the cleavage cocktail minus Mg^{2+} and the cleavage reactions were initiated by addition of $MgCl_2$ (37°C); final conditions were 40 mM Tris–HCl, pH 8.0, 1 mM EDTA, 11 mM $MgCl_2$ and ~5–50 nM RNA. For reactions that contained 10 M formamide the formamide was included in the preincubation. The kinetics of cleavage were followed by removing and mixing 5 µl aliquots with 5 or 10 µl formamide–dye mix containing 25 or 50 mM EDTA to quench the reaction. The precursor and product were separated by gel electrophoresis under denaturing conditions (6% polyacrylamide gels containing 7 M urea, 0.05 M Tris–borate, pH 8.3, 0.5 mM EDTA). The relative amounts of precursor and 3′-cleavage product were quantified by analysis in a phosphorimager (Molecular Dynamics). The fraction cleaved (F) was calculated as $(counts_{product})/(counts_{precursor} + counts_{product})$. The 3′-fragment contained 95–97% of the label for the precursors used in this study; the 5′-product migrates off these gels and was not included directly in the analysis, because correcting the data for this small difference had no effect on values obtained for the rate constants. The first order rate constant (k) and end point (m) were obtained by fitting the data to $F = m \times (1 - e^{-kt})$. The end points seen in reactions with purified precursors most likely did not reflect a true equilibrium between the cleaved and uncleaved forms, since the extent of cleavage can vary for different methods of preparation of the precursor. In addition, the extent of cleavage seen with purified precursor is routinely less than the extent of cleavage observed during transcription. More likely, the end point represents a combination of contaminating species that co-migrate with the precursor during gel purification and non-cleaving conformers of the ribozymes. For the faster cleaving sequences the calculated end points were compared with data from time points taken at 5–60 min (10–200 × $t_{½}$). At these extended time points the experimentally determined end point often exceeded the calculated value by 5–10%, suggesting that there may be a slow cleaving precursor population in those preparations. Correction for these differences did not result in significant changes in estimated rate constants, so all ribozyme sequences were analyzed as though they contained a single major species that cleaved with simple first order kinetics. For the two precursors containing a wild-type core sequence, SA1-2 and dSIV, the reaction was complete after 1 min (k = ~5/min and ~10/min respectively) and because the earliest time points were taken at 4 or 6 s values for rate constants >5/min were most likely minimal estimates. Thus, while reproducibility is good (±10–20%), we may have underestimated the fastest reactions. For the mutated precursor RNAs, which cleaved more slowly, reproducibility was often within ±10%. All rate constants reported were the average of at least three independent determinations.

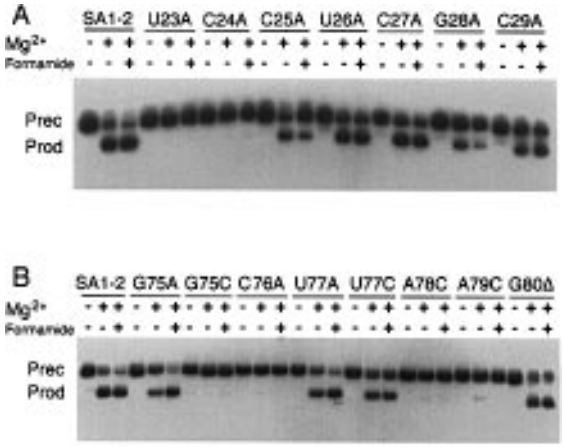

Figure 2. Cleavage of ribozymes with base changes in L3 (**A**) and J4/2 (**B**). The effect of individual base changes is illustrated by the extent of cleavage in a 5 min reaction at 37°C in the absence or presence of 10 M formamide. The conditions for these reactions were the same as for the kinetic studies and are described in Materials and Methods. Precursor and 3′-product were separated by electrophoresis in a polyacrylamide gel under denaturing conditions and only the portion of the gel containing precursor and 3′-product is shown. The 5′-product ran off the gel.

RESULTS

Antigenomic ribozymes with wild-type core sequences

Two versions of the antigenomic ribozyme sequence (SA1-2 and dSIV) were used as the starting sequences in these studies; each contained the same wild-type core sequence (Fig. 1). SA1-2 closely resembled the wild-type antigenomic ribozyme sequence with only a minor change in the non-essential L4 sequence (5). dSIV (9) was a shortened version of SA1-2. Both cleaved rapidly at 37°C in 10 mM Mg^{2+} (4.7/min and 11/min for SA1-2 and dSIV respectively), but dSIV cleaved slightly faster under all conditions tested (9). The genomic and antigenomic ribozymes cleaved in the presence of moderate to high levels of denaturants (5,10,25,26) and both SA1-2 and dSIV precursors cleaved two to three times faster when 10 M formamide was included in the reaction under the standard conditions (Table 1 and 2 respectively).

For the studies described below specific base changes were made in either the SA1-2 or dSIV ribozyme sequence. Precursor RNA containing the base change was purified and a first order rate constant was determined; comparisons of activity were made with respect to the parental wild-type core-containing precursor (SA1-2 or dSIV). In those instances where the same base change was made in both ribozymes the effect on activity was essentially the same.

Hairpin loop 3 (UCCUCGC, positions 23–29)

Hairpin 3 contains a 7 nt loop and a 3 bp stem which is contiguous in one strand with stem 1 and in the other strand with stem 2. In both HDV ribozymes a similar pyrimidine-rich sequence forms the loop of hairpin 3. However, the genomic ribozyme has an additional 3′ U, making it 8 nt in length. As a conserved sequence and structure this loop is a candidate for part of the active site. Initially each base from position 23 to 29 was changed individually to adenine and the effect on cleavage examined in the absence and presence of formamide (Fig. 2A). In this screen (5 min at 37°C)

mutations at either U23 or C24, in the 5′-side of L3, resulted in a large decrease in cleavage activity. Time courses of cleavage activity of these two mutants using gel-purified precursor revealed that the rate of cleavage decreased >10^3-fold (Table 1). With changes at the other positions (nt 25–29) the effects were less dramatic (Fig. 2A), but again time courses revealed that the rate constants for cleavage had decreased (Table 1). In this region the smallest effect was seen with U26A (2-fold decrease) and the largest effect was with G28A (50-fold decrease). The data suggested that the sequence of the 2 nt on the 5′-side of L3 were more critical than the 5 nt on the 3′-side, therefore, other bases were tested at positions 23 and 24 to determine to what extent the identity of the base substitution affected activity (Table 1). It was found that a U→C substitution at position 23 (U23C) still had a large negative effect on cleavage rate (10^3-fold), whereas at position 24 a C→U change (C24U) had a relatively small effect (~4-fold) (Table 1). A large decrease in activity (10^3-fold) was associated with a C24G mutation, indicating a possible preference for pyrimidine at position 24.

Addition of denaturants to cleavage reactions exaggerate the negative effect that mismatched base pairs in duplex regions had on cleavage rates (5,9); presumably in those cases it further destabilized the duplex containing the mismatch. With the L3 mutants the effect of formamide on cleavage rates varied with the particular change. With C24U, U26A, C27A and C29A there was a moderate (~2- to 3-fold) increase in the rate of cleavage when 10 M formamide was included in the reaction, about the same fold effect as was seen with SA1-2. If denaturant stimulated the rate of cleavage because it facilitated refolding from an inactive structure, then this result would suggest that folding may still be partially rate limiting for those mutants. The rate of cleavage of mutant C25A decreased 10-fold in the presence of formamide and a smaller decrease in activity was seen in formamide with the G→A change at position 28. Decreases were also seen with mutations at position 23, but these were already severely affected in the absence of formamide, so the magnitude of the effect is difficult to quantify. While there is no straightforward interpretation of the denaturant effects, decreased activity may indicate that a critical structure was at least partially disrupted in several of the L3 mutants.

Table 1.

Nucleotide[a]	Mutation	k[b] (% cleaved)	k_{rel}[c]	k_f[d] (% cleaved)
SA1-2	Wild-type	4.7/min (53%)	1	15/min (55%)
23	U23A	0.001 (75%)[e]	2.1×10^{-4}	<0.001
	U23C	0.003 (75%)[e]	6.4×10^{-4}	0.001 (75%)[e]
24	C24A	0.002 (75%)[e]	4.3×10^{-4}	0.003 (75%)[e]
	C24G	<0.001	$<2 \times 10^{-4}$	<0.001
	C24U	1.2 (57%)	2.6×10^{-1}	2.3 (59%)
25	C25A	0.50 (60%)	1.1×10^{-1}	0.05 (60%)
26	U26A	2.6 (60%)	5.5×10^{-1}	7.4 (60%)
27	C27A	0.59 (60%)	1.2×10^{-1}	1.3 (60%)
28	G28A	0.096 (60%)	2.0×10^{-2}	0.045 (60%)
29	C29A	1.0 (40%)	2.1×10^{-1}	2.2 (50%)
75	G75A	0.64 (43%)	1.4×10^{-1}	1.0 (70%)
	G75C	0.007 (20%)	1.5×10^{-3}	0.025 (20%)
76	C76A	0.011 (20%)	2.3×10^{-3}	0.003 (20%)
77	U77A	6.8 (50%)	1.4	8.8 (72%)
	U77C	1.9 (51%)	4.0×10^{-1}	1.5 (51%)
78	A78C	0.006 (62%)	1.3×10^{-3}	0.003 (54%)
79	A79C	0.018 (57%)	3.8×10^{-3}	0.004 (35%)
80	ΔG80	2.8 (55%)	6.0×10^{-1}	0.62 (67%)
–1	C-1A	2.7 (79%)	5.7×10^{-1}	0.84 (80%)
	C-1G	0.30 (88%)	6.4×10^{-2}	0.30 (99%)
	C-1U	4.9 (79%)	1.0	3.8 (78%)
1	G1A	0.33 (68%)	7.0×10^{-2}	0.22 (74%)
	G1C	0.027 (92%)	5.7×10^{-3}	0.07 (22%)
	G1U	0.090 (79%)	1.9×10^{-2}	0.04 (86%)
39	U39A	0.050 (33%)	1.1×10^{-2}	0.10 (62%)
	U39C	1.9 (64%)	4.0×10^{-1}	1.5 (64%)
	U39G	0.070 (22%)	1.5×10^{-2}	0.04 (47%)
	G1A:U39C	2.6 (52%)	5.5×10^{-1}	0.59 (73%)
	G1C:U39G	0.22 (78%)	4.7×10^{-2}	0.14 (80%)
	G1U:U39A	0.030 (79%)	6.4×10^{-3}	0.04 (40%)

[a]Nucleotide position.

[b]Cleavage in 10 mM Mg^{2+} at 37°C. First order rate constant (per min) and extent of cleavage expressed as a percentage of precursor cleaved (see Materials and Methods).

[c]Relative activity. Rate constant divided by 4.7/min.

[d]Cleavage with addition of 10 M formamide. First order rate constant (per min) and extent of cleavage expressed as a percentage of precursor cleaved (see Materials and Methods)

[e]The curve fitting program would not fit data obtained for this mutant to the rate equation given in Materials and Methods unless the end point was fixed. Therefore, the end point was set to a reasonable, though somewhat arbitrary, value.

In total, these data suggest that L3 provides an essential structure for activity, either as part of the active site or through interactions with other sequences in the ribozyme. This was in contrast to the situation with P4 and L4, where no effect was seen when those regions were mutated or deleted. It is very likely that the structure of L3 could be important for optimal cleavage activity and it is possible that several nucleotides in L3 (U23, C24 and G28) may have specific functions in the cleavage reaction or in stabilizing a structural feature essential for catalysis.

The J4/2 sequence (positions 75–79)

The J4/2 region in the antigenomic (GCUAA) and genomic (GCGAA, positions 74–78) ribozymes are non-identical, but similar, in sequence. If the secondary structures are accurate, the antigenomic J4/2 sequence may contain an extra nucleotide (G80). The significance of this difference was examined by deleting G80 in the antigenomic sequence. While in the 5 min screen it appeared that deletion of G80 had very little effect (Fig. 2B), rate measurements revealed that the deletion actually resulted in a small decrease in activity and formamide sensitivity (Table 1). This suggested that G80 contributed in some fashion, perhaps to stability through interation with sequence in P3. No further alterations at position 80 were tested. At other positions in J4/2 several base changes had dramatic effects on self-cleavage activity (Fig. 2B). Most notably G75C, C76A, A78C and A79C resulted in large decreases in the rate of cleavage (250- to 800-fold down). A U77A mutation had no detectable effect on activity, although a U77C mutation did cause a small 2- to 3-fold decrease. The negligible effect seen with altering U77 is consistent with sequence variation at the equivalent position in the genomic ribozyme (G76 in the genomic sequence).

These data suggest that there are several key residues in this region of the ribozyme that are required for optimal activity. While it provided little insight into possible structures involving these nucleotides, it was noted that the G→C change at position 75 could result in an extra Watson–Crick G-C pair at the end of P4 (the C pairing with G42), thus extending the helix into the potential core region. The possibility that an extra Watson–Crick base pair at this position may disrupt the core structure was supported by recent evidence that a homopurine base pair at the end of P4 was required for full cleavage activity (27).

The J1/4 sequence (positions 40–42)

The genomic and antigenomic ribozymes each contain the sequence GGG connecting the 3′-side of P1 to the 5′-side of P4. The genomic sequence also contains an additional 3 nt (CAA) that have been shown to be non-essential (18). Changes in this run of G residues were made in the dSIV background, which differs from SA1-2 only in that P4 has been shortened. The numbering of nucleotides in dSIV is otherwise the same as for SA1-2. Shortening P4 resulted in a 2-fold higher rate of cleavage (9), indicating that there was probably no adverse effect on the self-cleaving reaction due to this deletion. All of the six possible individual base changes at positions 40 and 41 had dramatic effects on cleavage rates; most rates decreased by at least two orders of magnitude (Table 2). Clearly, the presence of guanine at positions 40 and 41 is important. Addition of formamide to these reactions did not stimulate cleavage of these mutants and cleavage rates in formamide were not quantified.

Relative to positions 40 and 41 the effects of base changes at position 42 were more sequence dependent. G42A, G42U and G42C were down 10-, 100- and 1000-fold respectively. As hypothesized above for position G75, one explanation for these effects was that the G→C change at position 42 allows P4 to form an additional Watson–Crick base pair (with G75) and it was this specific pairing that interfered with structure in the core sequences.

Possible base pairing between J1/4 and J4/2

The potential for Watson–Crick interactions between either G40 or G41 and C76 was tested in the dSIV ribozyme. At position 76 a C→G or U change resulted in complete loss of activity (Table 2). The potential for position C76 to pair with either G41 or G40 was tested with three double mutants, G41C:C76G, G41A:C76U and G40A:C76U. None of the double mutants showed any detectable cleavage activity in Mg^{2+} with or without formamide. Although these results were negative, the experiments did not rule out contacts across this region.

Sequence requirement for the nucleotide 5′ of the cleavage site

Immediately 5′ of the cleavage site a C in the antigenomic sequence and a U in the genomic sequence suggested a preference for a pyrimidine at position –1. This possibility was investigated by making mutations at the –1 position in the antigenomic sequence. Cleavage rates were found to be very similar with either C or U at position –1 (Table 1). With the addition of formamide, however, the C–1U change resulted in 2- to 3-fold slower cleavage, suggesting that an interaction had been disrupted with that change. With an A at position –1 there was a 2-fold decrease in the rate of cleavage and an additional 3-fold drop with the addition of formamide. A C–1G mutation resulted in the largest decrease relative to SA1-2, 15- or 50-fold in the absence or presence of formamide respectively. Thus when formamide (10 M) was included in the reaction there was a clear preference for C. In general it appeared that the rate of cleavage with substitutions at position –1 decreased in the order, C = U > A > G. It should be emphasized that the differences in rates seen with changes at –1 were very small relative to the effects seen with changes at G40, a position which could potentially pair with –1C. Even the slowest cleaving –1 variant (C–1G) had a $t_{½}$ of <3 min in the absence of denaturants and thus would cleave to completion in 15–30 min.

A wobble base pair at the cleavage site

The secondary structures of both the genomic and antigenomic ribozymes (Fig. 1) include the potential for a G·U base pair at the cleavage site. In both structures the guanosine at position 1 (3′ of the cleavage site) could pair with a U at the 3′-end of P1 (U39 antigenomic or U37 genomic). While a G·U wobble interaction at the end of an RNA helix would not be unusual, evidence for the nature of the base pair requirement at this position in these ribozymes was lacking. The proximity to the cleavage site suggested that a G·U pair at this position could be functionally significant. Changing G1 to an A resulted in a 14-fold decrease in cleavage rates, while changing it to a pyrimidine resulted in a larger (50- to 200-fold) decrease (Table 1). Changing U39 to a C resulted in only a 2-fold decrease, but changing it to a purine resulted in a 50- to 100-fold decrease in activity (Table 1). If the bases at positions 1 and 39 interact, these data indicate that purine·pyrimidine base pairs (G·U, G-C and A-U) might be preferable to several pyrimidine·pyrimidine and purine·purine combinations (C·U, U·U, G·A and G·G). Three additional mutations were made in which the potential G·U pair was replaced with C-G (G1C:U39G), U-A (G1U:U39A) and

A·C (G1A:U39C). The pyrimidine·purine orientations did not cleave well, but the A·C cleaved about eight times faster than the A-U and about half as fast as the wild-type G·U. The latter result suggest that the G1A mutation could be partially rescued with a U39C change. These data indicated the order of activity for those combinations that cleaved fastest was G·U > A·C ≥ G-C > A-U. This data was consistent with two requirements: a preference for G at position 1 and a purine·pyrimidine non-Watson–Crick base pair. The addition of 10 M formamide to the reaction had a negative effect (4- to 5-fold) on the rate of cleavage of the A·C mutant, but a negligible effect on the G-C mutant (Table 1). In formamide, because of the enhanced activity, there was a clear preference for the wild-type G·U combination. This suggested that while A·C can substitute for G·U at this site, it might be a less stable interaction. This would be consistent with an A·C wobble pair, isosteric with a G·U wobble, but with one rather than two hydrogen bonds (Fig.3). In support of this idea, there is preliminary evidence that the A·C mutant cleaves better in formamide at lower pH (pH 6.4; A.T.Perrotta and M.D.Been, unpublished results), where protonation of the adenine could allow formation of a second hydrogen bond (28).

DISCUSSION

Self-cleavage activity of the HDV antigenomic ribozyme was sensitive to base changes in three regions that are represented as single-stranded in the secondary structures. These three regions, a hairpin loop (L3) and two joining sequences (J1/4 & J4/2), were hypothesized to form part of the catalytic core of the ribozyme. In terms of proximity to the site that is cleaved J1/4 has to be close to the cleavage site simply because it is continuous with the 3′-side of P1 (Fig. 1). Cross-linking studies revealed that parts of L3 and J4/2 are also within ~10 Å of the cleavage site phosphate (29). Thus we can conclude that portions of all three of these regions are physically close to the cleavage site phosphate in the folded RNA and, therefore, could contribute to catalysis. By characterizing the effect of single base changes on cleavage activity we have now shown that each of these regions contain sequence critical for cleavage activity.

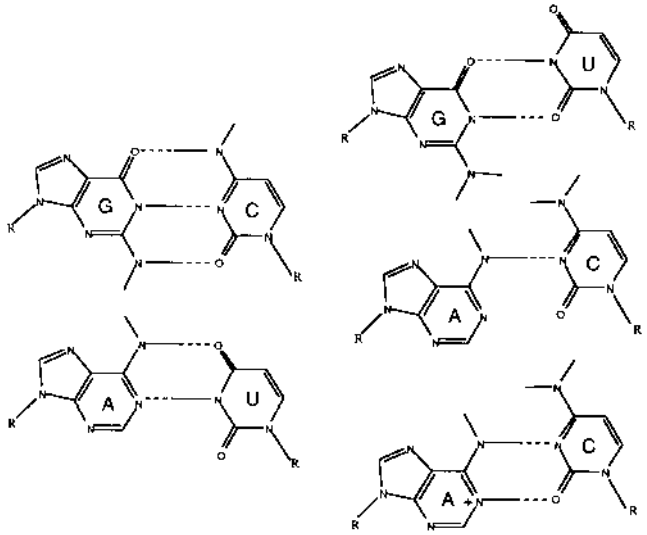

Figure 3. Potential G·U and A·C wobble pairs compared with Watson–Crick pairs.

Table 2.

Nucleotide[a]	Mutation	k[b] (% cleaved)	k_{rel}[c]	k_f[d] (% cleaved)
dSIV		11 (72%)	1	19 (90%)
40	G40A	0.024 (70%)	2.2×10^{-3}	nd[f]
	G40C	0.040 (60%)	3.6×10^{-3}	nd
	G40U	0.019 (70%)[e]	1.7×10^{-3}	nd
41	G41A	0.060 (60%)[e]	5.5×10^{-3}	nd
	G41C	0.030 (60%)[e]	2.7×10^{-3}	nd
	G41U	0.062 (60%)[e]	5.6×10^{-3}	nd
42	G42A	0.79 (98%)	7.1×10^{-2}	0.82 (80%)
	G42U	0.040 (80%)[e]	3.6×10^{-3}	nd
	G42C	0.004 (50%)	3.6×10^{-4}	0.012 (40%)
75	G75C	0.002 (38%)	1.8×10^{-4}	0.006 (40%)
	G75A	0.73 (97%)	6.6×10^{-2}	1.6 (88%)
76	C76G	<0.001	$< 1 \times 10^{-4}$	<0.001
	C76U	<0.001	$< 1 \times 10^{-4}$	<0.001
	G41A:C76U	<0.001	$<1 \times 10^{-4}$	<0.001
	G41C:C76G	<0.001	$<1 \times 10^{-4}$	<0.001
	G40A:C76U	<0.001	$<1 \times 10^{-4}$	<0.001

[a]Nucleotide position.
[b]Cleavage in 10 mM Mg^{2+} at 37°C. First order rate constant (per min) and extent of cleavage expressed as a percentage of precursor cleaved (see Materials and Methods).
[c]Relative activity. Rate constant divided by 11/min.
[d]Cleavage with addition of 10 M formamide. First order rate constant (per min) and extent of cleavage expressed as a percentage of precursor cleaved (see Materials and Methods)
[e]The curve fitting program would not fit data obtained for this mutant to the rate equation given in Materials and Methods unless the end point was fixed. Therefore, the end point was set to a reasonable, though somewhat arbitrary, value.
[f]nd, rate not determined.

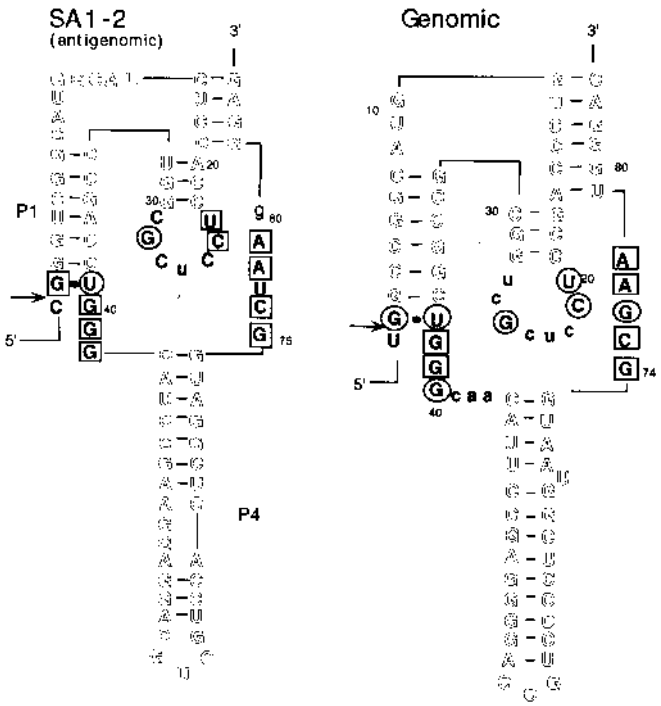

Figure 4. Summary of mutagenesis results. The secondary structure of the antigenomic ribozyme is redrawn with only the core sequences emphasized. The effect that one or more mutations at a particular position had on the rate of cleavage is denoted as follows: lower case, <2-fold decrease; upper case, 2- to 10-fold decrease; circled upper case, 10- to 100-fold decrease; boxed upper case, >100-fold decrease. For comparison the genomic ribozyme is also shown using the mutagenesis data of Tanner *et al.* (18) for J1/4 and J4/2, Wu *et al.* (13) for the cleavage site G·U pair and Kawakami*et al.* (19) for L3. For the G·U pair and L3 the activity is based on extent of cleavage after extensive incubation times, therefore, the rate of cleavage would probably be more severely affected than is indicated.

Several groups have assessed a variety of mutations in the equivalent core regions of the genomic ribozyme (9,13,18–20,30,31). Relative to the genomic ribozyme the antigenomic is shorter by 1 nt in L3, shorter by three in J1/4 and longer by one in J4/2. However, of the remaining nucleotides in these three regions there is only one base difference (antigenomic U77 and genomic G76) and mutagenesis in both ribozymes indicated that there was no strict base requirement at that position (18). Having similar data sets on the core sequences of different ribozymes is useful, because the ribozyme sequences forming the remainder of each structure are slightly different. Therefore, this makes it possible to distinguish between mutations that may specifically affect one ribozyme and those that are general to both and thus may identify positions that participate in a catalytic center common to both. Tanner *et al.* (18) have carried out a careful and thorough characterization of the effect of site-specific change in the genomic J1/2 and J4/2 core sequences. They started with a highly active minimal ribozyme similar in length to SA1-2 and have determined rates of cleavage for the mutants relative to the wild-type. Despite differences in the sequences and in the approaches of the various laboratories, there appears to be a striking correlation in the results. For the most part all of the mutagenesis data were consistent with the idea that the two ribozymes form similar core structures (Fig. 4). Our results with the antigenomic ribozyme closely parallel the results and conclusions reported by Tanner *et al.* (18) for J1/2 and J4/2 in the genomic ribozyme. Specifically, in each of the joining sequences we found that base changes at equivalent positions showed similar effects. In the J1/4 run of three G reidues the 5′-most G (genomic G38, antigenomic G40) was most critical and in J4/2 a cytosine (genomic C75, antigenomic C76) appeared to be most critical for activity. For the purpose of comparison it is somewhat difficult to directly use the results of Kumar *et al.* (20) and Kawakami *et al.* (19), because activity estimates were based on a screen that measured the extent of cleavage after a 2–3 h transcription reaction. This approach quickly identifies the most defective mutants, but will grossly over-estimate the activity of slow cleaving variants. Thus there was a correlation between the inactive genomic mutants in their studies and the least active mutants in the antigenomic RNA in our studies, but many of the slower cleaving mutants that we identified would have no phenotype in that assay. Nevertheless, for L3 our results did not conflict with those of Kawakami *et al.* (19) in that they saw the greatest loss of activity with mutations at U20 (antigenomic U23), C21 (antigenomic C24) and G25 (antigenomic G28). These data suggest that the sequences in these regions are conserved for function and that the two sequences may use a very similar arrangement of core nucleotides to form the catalytic center of the ribozyme (Fig. 4).

The secondary structures of the two ribozymes suggest that there is likely to be a G·U pair at the cleavage site. While G·U pairs are common in RNA secondary structures, especially at the ends of helices (32), this is the only naturally occuring G·U pair proposed in the current model of HDV ribozymes. Our data strongly suggest that a wobble pair at the cleavage site is important for cleavage activity. Neither the Watson–Crick base pairs nor several mismatches function as well at this position. Wu *et al.* (1993) found a similar effect for several base pair combination substitutions at this position in the genomic sequence. In that study, while an A·C combination at this position was not tested, G·U was preferable to G-C, which in turn was better than A-U. For the G·U pair there are two obvious differences from a Watson–Crick pair. Structural changes associated with displacement of the U into the major groove and the G into the minor groove could distort the sugar–phosphate backbone and may facilitate the cleavage reaction. While distortions to the backbone may be mimicked by an A·C wobble, G·U is unique because of the guanosine 2 amino group, which is displaced into the minor groove. The importance of the 2 amino group can be tested by substituting inosine for guanosine at that position (33). It is noteworthy that in an extensively studied class of ribozymes, the group I introns, there is a U·G base pair at the 5′ splice site. This feature is nearly invariant in all of the natural isolates (34) and, as with HDV ribozymes, the U·G pair can be replaced by C-G (35) or C·A (36) with only partial loss of activity *in vitro*. Recent studies on the effect of substituting C-G for the U·G pair in the *Tetrahymena* group I intron revealed that several aspects of the splicing reaction were affected, with the unifying theme being that docking of the P1 helix into the core of the ribozyme is less favorable with G-C (37). The relationship of the wobble pair to the site of cleavage, however, is different in these two cases. In the group I introns cleavage of the RNA occurs 3′ of the U in the U·G pair and a 3′ hydroxyl group is generated; in HDV ribozymes cleavage is 5′ of the G and a 5′ hydroxyl is generated. In addition, preliminary studies with a *trans*-acting HDV ribozyme indicate that G·U pairing is weaker than either G-C or A-U combinations (A.T.Per-

rotta and M.D.Been, unpublished results). The role of the G·U pair in the HDV ribozyme warrants further investigation.

A three-dimensional structure of the HDV genomic ribozyme has been proposed (18). This model incorporated established features of the secondary structure and brought residues identified by mutagenesis as essential in the core region together near the cleavage site phosphate. Because of extensive similarities in sequence and secondary structures between the two ribozymes in HDV we would expect the three-dimensional structures to be similar. The mutagenesis data on the antigenomic ribozyme appeared to be consistent with many features of the genomic ribozyme three-dimensional model. Until a physical structure is available, additional constraints, either from identifying compensatory mutants or from biochemical approaches, are necessary to further refine the details. When a physical model is available it should account for the mutagenesis data and together the two approaches will provide greater insight into RNA stucture and function.

ACKNOWLEDGEMENTS

We thank T.Wadkins and G.Wickham for helpful comments on the manuscript. This work was supported by a grant from the NIH (GM47233).

REFERENCES

1 Sharmeen,L., Kuo,M.Y.-P., Dinter-Gottlieb,G. and Taylor,J. (1988) *J. Virol.*, **62**, 2674–2679.
2 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.
3 Kuo,M.Y.-P., Sharmeen,L., Dinter-Gottleib,G. and Taylor,J. (1988) *J. Virol.*, **62**, 4439–4444.
4 Rosenstein,S.P. and Been,M.D. (1991) *Nucleic Acids Res*, **19**, 5409–5416.
5 Perrotta,A.T. and Been,M.D. (1991) *Nature*, **350**, 434–436.
6 Symons,R.H. (1989) *Trends Biochem. Sci.*, **14**, 445–450.
7 Symons,R.H. (1992) *Annu. Rev. Biochem.*, **61**, 641–671.
8 Been,M.D. (1994) *Trends Biochem. Sci.*, **19**, 251–256.
9 Been,M.D., Perrotta,A.T. and Rosenstein,S.P. (1992) *Biochemistry*, **31**, 11843–11852.
10 Perrotta,A.T. and Been,M.D. (1990) *Nucleic Acids Res.*, **18**, 6821–6827.
11 Perrotta,A.T. and Been,M.D. (1993) *Nucleic Acids Res.*, **21**, 3959–3965.
12 Thill,G., Vasseur,M. and Tanner,N.K. (1993) *Biochemistry*, **32**, 4254–4262.
13 Wu,H.-N., Lee,J.-Y., Huang,H.-W., Huang,Y.-S. and Hsueh,T.-G. (1993) *Nucleic Acids Res.*, **21**, 4193–4199.
14 Perrotta,A.T. and Been,M.D. (1992) *Biochemistry*, **31**, 16–21.
15 Puttaraju,M., Perrotta,A.T. and Been,M.D. (1993) *Nucleic Acids Res.*, **21**, 4253–4258.
16 Branch,A.D. and Robertson,H.D. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10163–10167.
17 Wu,H.-N., Wang,Y.-J., Hung,C.-F., Lee,H.-J. and Lai,M.M.C. (1992) *J. Mol. Biol.*, **223**, 233–245.
18 Tanner,N.K., Schaff,S., Thill,G., Petit-Koskas,E., Crain-Denoyelle,A.-M. and Westhof,E. (1994) *Curr. Biol.*, **4**, 488–497.
19 Kawakami,J., Kumar,P.K.R., Suh,Y.-A., Nishikawa,F., Kawakami,K., Taira,K., Ohtsuka,E. and Nishikawa,S. (1993) *Eur. J. Biochem.*, **217**, 29–36.
20 Kumar,P.K.R., Suh,Y.-A., Miyashiro,H., Nishikawa,F., Kawakami,J., Taira,K. and Nishikawa,S. (1992) *Nucleic Acids Res.*, **20**, 3919–3924.
21 Davanloo,P., Rosenberg,A.H., Dunn,J.J. and Studier,F.W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2035–2039.
22 Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
23 Vieira,J. and Messing,J. (1987) *Methods Enzymol.*, **153**, 3–11.
24 Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25 Rosenstein,S.P. and Been,M.D. (1990) *Biochemistry*, **29**, 8011–8016.
26 Belinsky,M.G. and Dinter-Gottlieb,G. (1991) *Nucleic Acids Res.*, **19**, 559–564.
27 Been,M.D. and Perrotta,A.T. (1996) *RNA*, in press.
28 Puglisi,J.D., Wyatt,J.R. and Tinoco,J.I. (1990) *Biochemistry*, **29**, 4215–4226.
29 Rosenstein,S. (1994) PhD Thesis. Duke University.
30 Kumar,P.K.R., Suh,Y.-A., Taira,K. and Nishikawa,S. (1993) *FASEB J.*, **7**, 124–129.
31 Wu,H.-N. and Huang,Z.-S. (1992) *Nucleic Acids Res.*, **20**, 5937–5941.
32 Gutell,R.R., Larsen,N. and Woese,C.R. (1994) *Microbiol. Rev.*, **58**, 10–26.
33 Strobel,S.A., Cech,T.R., Usman,N. and Beigelman,L. (1994) *Biochemistry*, **33**, 13824–13834.
34 Michel,F. and Westhof,E. (1990) *J. Mol. Biol.*, **216**, 585–610.
35 Barfod,E.T. and Cech,T.R. (1989) *Mol. Cell. Biol.*, **9**, 3657–3666.
36 Doudna,J.A., Cormack,B.P. and Szostak,J.W. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7402–7406.
37 Pyle,A.M., Moran,S., Strobel,S.A., Chapman,T., Turner,D.H. and Cech,T.R. (1994) *Biochemistry*, **33**, 13856–13863.