RNA (1995), *1*:1061–1070. Cambridge University Press. Printed in the USA.

Optimal self-cleavage activity of the hepatitis delta virus RNA is dependent on a homopurine base pair in the ribozyme core

MICHAEL D. BEEN and ANNE T. PERROTTA
Department of Biochemistry, Duke University Medical School, Durham, North Carolina 27710

ABSTRACT

A non-Watson–Crick G·G interaction within the core region of the hepatitis delta virus (HDV) antigenomic ribozyme is required for optimal rates of self-cleavage activity. Base substitutions for either one or both G's revealed that full activity was obtained only when both G's were replaced with A's. At those positions, substitutions that generate potential Watson–Crick, G·U, heteropurine, or homopyrimidine combinations resulted in dramatically lower cleavage activity. A homopurine symmetric base pair, of the same type identified in the high-affinity binding site of the HIV RRE, is most consistent with this data. Additional features shared between the antigenomic ribozyme and the Rev binding site in the vicinity of the homopurine pairs suggest some structural similarity for this region of the two RNAs and a possible motif associated with this homopurine interaction. Evidence for a homopurine pair at the equivalent position in a modified form of the HDV genomic ribozyme was also found. With the postulated symmetric pairing scheme, large distortions in the nucleotide conformation, the sugar–phosphate backbone, or both would be necessary to accommodate this interaction at the end of a helix; we hypothesize that this distortion is critical to the structure of the active site of the ribozyme and it is stabilized by the homopurine base pair.

Keywords: catalytic RNA; HDV; non-Watson–Crick; RNA structure

INTRODUCTION

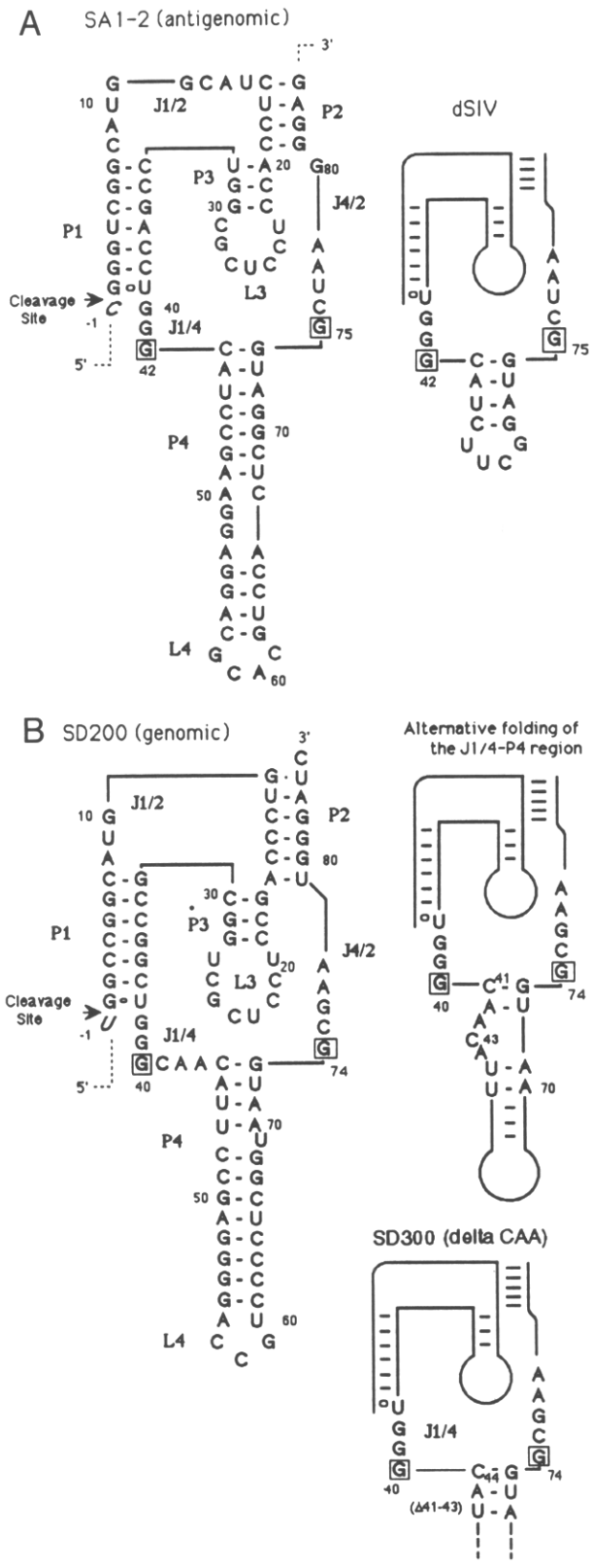
The self-cleaving RNA sequences (ribozymes) found in the genomic and antigenomic RNAs of hepatitis delta virus (HDV) (Kuo et al., 1988; Sharmeen et al., 1988; Wu et al., 1989) adopt a structural motif (Perrotta & Been, 1991; Rosenstein & Been, 1991) distinct from the motifs previously defined for the hammerhead and hairpin ribozymes (Symons, 1992; Been, 1994). In the HDV ribozymes, the cleavage site is located at the 5′ end of the sequence defining the self-cleaving element (Perrotta & Been, 1990, 1991; Been et al., 1992). Therefore, with the exception of a single ribonucleotide 5′ to the cleavage site, sequences essential for catalyzing the cleavage reaction are 3′ of the cleavage site. These sequences form a secondary structure containing four paired regions (P1–P4) in which the combination of P1 and P2 form a type of pseudoknot (Fig. 1) (Perrotta & Been, 1991). The strongest support for these models is from mutagenesis studies in both the antigenomic and genomic sequences (Perrotta & Been, 1991, 1993; Been et al., 1992; Kumar et al., 1992, 1993; Thill et al., 1993) and by the predictive value of the secondary structure in the generation of active *trans*-acting ribozymes with novel sequences (Been et al., 1992).

The active site of the ribozyme is likely to include regions of undefined structure represented as single stranded in the secondary structure models. There are three single-stranded regions and two hairpin loops in these structures; however, an active form of the ribozyme can be generated from two fragments in which the J1/2 is deleted (Been et al., 1992; Puttaraju et al., 1993) or where hairpin loop 4 is interrupted or deleted (Branch & Robertson, 1991; Been et al., 1992; Wu et al., 1992; Perrotta & Been, 1993). Therefore, L4 and J1/2 are hypothesized not to participate directly in forming the active site and efforts have been focused on defining the contributions of J1/4, J4/2, and L3 sequences to the core structure.

Mutagenesis of the core region of the HDV ribozymes (Suh et al., 1993; Thill et al., 1993; Tanner et al., 1994)

Reprint requests to: Michael D. Been, Department of Biochemistry, Duke University Medical School, Durham, North Carolina 27710.

suggests that sequences in these regions are important for activity but has not identified interactions beyond those already implied by the secondary structure. Interactions in the core could consist of contacts between two or more bases, or a variety of contacts between bases, sugars, and phosphates. Many such interactions will not be revealed in mutagenesis experiments. However, in this report, we demonstrate evidence for one such interaction: a purine–purine base pair with properties similar to that of the well-established homopurine base pair in the high-affinity Rev binding site (Bartel et al., 1991; Giver et al., 1993; Battiste et al., 1994; Peterson et al., 1994).



RESULTS

Mutations at G42 or G75 in the antigenomic ribozyme reduce self-cleavage activity

Two versions of the antigenomic ribozyme sequence (SA1-2 and dSIV) were used in the course of these studies (Fig. 1A). Both contain the wild-type core sequences of the antigenomic ribozyme. SA1-2 more closely resembles the wild-type antigenomic ribozyme in sequence and size with only minor sequence changes in the nonessential L4 (Perrotta & Been, 1991). dSIV (Been et al., 1992) is a shortened version of SA1-2, where much of the nonessential stem-loop 4 (P4 and L4) was deleted. Both cleave rapidly at 37 °C in 10 mM Mg^{2+} (4.7/min and 11/min for SA1-2 and dSIV, respectively), and those cleavage rates are enhanced two- to threefold in the presence of a denaturant such as 5 M urea or 10 M formamide (Table 1).

Base changes made in the J4/2 and the J1/4 regions suggested a possible interaction at the end of P4. In SA1-2, G75 was first changed to either A or C and activity of precursor RNA was screened in a 5-min reaction in either the absence or presence of formamide (Fig. 2). In this assay, it is clear that the G to C change interfered with cleavage activity. The mutant with the G to A change did cleave, but, in the absence of denaturant, there was less cleavage than with SA1-2. The G75A mutant is less active than it may appear from this experiment because the 5-min time point is considerably longer than the half-life of either SA1-2 or G75A. When precursor RNAs containing the base changes were purified and first-order rate constants for self-cleavage were determined, it was clear that both changes affected activity (Table 1). The G to A change caused a small but significant 7-fold decrease in the rate of cleavage relative to SA1-2, and the G to C mutation

FIGURE 1. Secondary structure of the ribozymes used in this study. **A:** Sequence and secondary structures of the antigenomic ribozymes SA1-2 and dSIV. For SA1-2, the full sequence from −1 to +84 of the SA1-2 is shown. For dSIV, in which P4 was deleted, only sequence of P4 and flanking regions are shown. **B:** Sequence and secondary structure of the genomic ribozymes SD200 and SD300. The full sequence, −1 to +84 is shown along with a partial schematic showing an alternative pairing that places part of the J1/4 sequence in P4. SD300 differs from SD200 in that 3 nt (C41A42A43) are deleted.

TABLE 1. Cleavage activity of antigenomic ribozymes at 37° in the absence and presence of formamide.

	Mg^{2+} [a]		Mg^{2+}+F [b]	
Ribozyme/mutation	Rate constant (min^{-1})	% Cleaved [c]	Rate constant (min^{-1})	% Cleaved
SA1-2	4.7	53	15	55
SA1-2/G75A	0.64	43	1.0	70
SA1-2/G75C	0.007	20	0.025	20
DSIV	11	72	19	90
DSIV/G42A	0.79	98	1.5	98
DSIV/G42U	0.040	80	—	
DSIV/G42C	0.004	50	0.012	40
DSIV/G75C	0.002	38	0.006	40
DSIV/G42C:G75C	0.017	70	0.005	73
DSIV/G75A	0.73	97	1.6	88
DSIV/G42C:G75A	0.96	96	[1.0] [d]	
DSIV/G42A:G75C	0.18	90	[0.16]	
DSIV/G42A:G75A	7.2	88	16	91
DSIV/C43G	6.7	90	0.24	90
DSIV/G74C	0.050	86	0.001	85
DSIV/C43G:G74C	0.34	83	0.58	92

[a] 11 mM $MgCl_2$, 1 mM EDTA, and 40 mM Tris-HCl, pH 8.0, at 37°.
[b] 11 mM $MgCl_2$, 1 mM EDTA, 40 mM Tris-HCl, pH 8.0, and 10 M formamide at 37°.
[c] Percent cleavage is the end point of the reaction, essentially the percentage of the precursor that cleaved with the rate constant given.
[d] Values in brackets are from one determination; for these two mutants, the formamide appeared to have no effect.

resulted in more than a 500-fold reduction in the rate constant for cleavage. A possible explanation for the large loss of activity is suggested in the secondary structure; if a G to C change at position 75 favored the formation of an additional base pair at the end of P4 (75 with G42), a stable Watson–Crick base pair at that position may interfere with activity. To facilitate the interpretation of additional mutations, a faster cleaving form of the antigenomic ribozyme, dSIV, was used. The rationale for making this change was that it is more likely that a mutation affecting catalysis will have an observable effect in a fast-cleaving ribozyme. Changes at position 75 were introduced into dSIV and the effect of G75A and G75C mutations were essentially the same as seen in SA1-2 (Table 1). In the dSIV background, G75C cleaves 5,000 times slower, and G75A about 15 times slower than the unmodified sequence. Again, these results could be interpreted to mean that changes that introduce stable base pairs at the end of P4 between nt 42 and 75 interfered with activity. If so, the same effect would be seen with mutations at position 42. Therefore, additional mutations were made in dSIV at G42, changing it to either A, C, or U. The G to A change had the smallest effect on cleavage rate (down 14-fold), the G to C the largest effect (down more than 1,000-fold), and the G to U mutation an intermediate effect (down 200–300-fold). Thus, to a large extent, the effect of specific base changes at position 42 mimic the same base changes at position 75. This is suggestive of, but not evidence for, an interaction between positions 42 and 75 in the wild-type sequence.

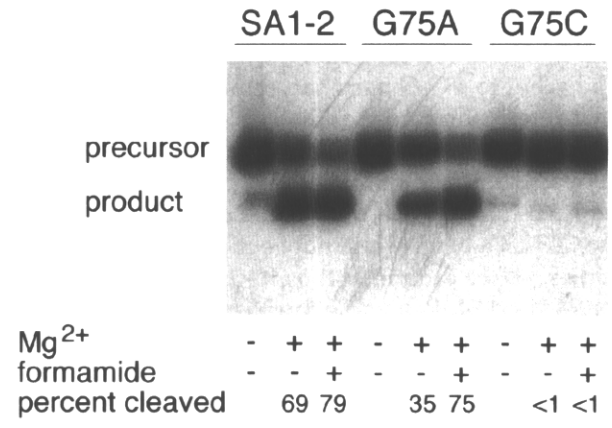


FIGURE 2. Cleavage activity of SA1-2 antigenomic ribozyme variants with base changes at position 75. Precursor RNAs (internally labeled with ^{32}P) SA1-2 (wt), SA1-2/G75A, or SA1-2/G75C, are compared for the extents of cleavage after 5 min at 37° in either 10 mM Mg^{2+} or 10 mM Mg^{2+} with 10 M formamide. Products were fractionated on a urea-containing polyacrylamide gel and a portion of the autoradiogram is shown. Quantitation was done with a PhosphorImager.

Evidence for a G·G base pair at the end of P4 in the antigenomic ribozyme

Sequence requirements at these positions were further investigated, and the hypothesis that single G to C mutations at position 42 or 75 were detrimental for activ-

ity due to the negative effects of a stable Watson–Crick base pair was tested. Changing both positions to C resulted in a variant that cleaved slightly faster than either single mutant (Table 1), but much slower (>500-fold) than the wild-type sequence. The low activity of this double mutant ruled out the simple explanation that the requirement is solely for a mismatch at the end of P4. It did, however, leave open the possibility of other interactions, including a specific pairing between guanosines at positions 42 and 75. With additional mismatches at positions 42 and 75 (Table 1), it was found that both the G42C and G75C single-base-change mutants could be partially rescued with a G to A change at positions 75 and 42, respectively. One of these, G42C:G75A, cleaved about 240 times faster than G42C, and the other, G42A:G75C, cleaved 90 times faster than G75C. However, G42C:G75A was not significantly faster than the G75A single mutation, and the G42A:G75C double mutant was slower than the G42A single. These results suggest a possible interaction between positions 42 and 75, but did not provide the evidence normally used as criteria for identifying a compensatory base change.

Evidence that G42 and G75 interact in a specific structure comes from the effect observed with a double mutant containing the homopurine combination, G42A:G75A. Replacing both G's with A's restored cleavage activity to wild-type levels (Table 1); this was seen in the presence of formamide as well as in its absence (Fig. 3A,B). This result strongly indicates an interaction between position 42 and 75 because the double mutant cleaved faster than either single mutant. Thus, although some non-Watson–Crick combinations at positions 42 and 75 did work better than the two Watson–Crick combinations tested, only the homopurine combination gave wild-type levels of activity. Most consistent with this finding would be a requirement for a symmetric (*trans*) G·G or A·A pair (Fig. 4; structures III and I, numbered according to Saenger, 1984), because, in that arrangement, the base pairs are isomorphic. Evidence for such a homopurine base pair has been found by in vitro selection and NMR spectroscopy for the HIV Rev binding site (Bartel et al., 1991; Giver et al., 1993; Battiste et al., 1994; Peterson et al., 1994), where it plays an essential role in recognition and binding of the RNA by the Rev protein. We hypothesize that G42 and G75 may form a very similar pairing in the antigenomic HDV ribozyme.

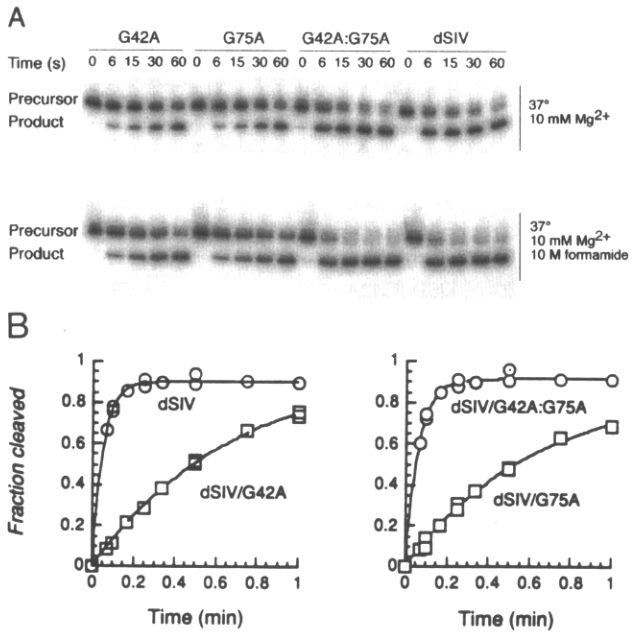


FIGURE 3. Time courses for cleavage of variants of dSIV demonstrating restored rates of cleavage for the compensatory purine base changes. **A:** Internally labeled precursor RNAs were incubated under the conditions indicated in the figure. Aliquots were quenched at the times indicated and fractionated by electrophoresis in polyacrylamide gels. A portion of the autoradiogram is shown. **B:** Cleavage of dSIV and variants; G42A, G75A, and G42A:G75A in the presence of formamide were quantified by PhosphorImager analysis and the data was fit to the equation for a first-order reaction with a variable end point. Data for the mutants with single base changes included a time point at 30 min.

A homopurine base pair associated motif?

In addition to the homopurine requirement, there are noticeable similarities between the core of the antigenomic ribozyme and the Rev binding site (Fig. 5). First, finding that the non-homopurine mismatch G42C:G75A cleaved faster than G42A:G75C parallels in vitro selection, which revealed that a C·A combination can replace the homopurine pair in the Rev binding site (Giver et al., 1993). Giver et al. (1993) point out that the C·A mismatch can form a base pair that would approximate the nucleotide arrangement in the homopurine pairing (Fig. 4), and that the nonequivalence of the C·A and A·C interactions could reflect the requirement for a specific configuration of the 3′ purine nucleotide. Second, in the Rev binding site, the homopurine base pair is adjacent to a C-G pair that terminates the helix; the in vitro selection studies indicate a specific requirement for this C-G pair (Bartel et al., 1991). Likewise, in the HDV antigenomic ribozyme, P4 contains a C-G base pair (C43-G74) where J1/4 and J4/2 come together at P4. The C43-G74 pairing had not been tested in our earlier studies, so we examined the effect of changes at positions 43 and 74. Reversing the base pair (C43G:G74C) resulted in a lower rate of cleavage activity (Table 1); it was of intermediate activity relative to the C43G mutant (cleaved only slightly slower than the wild-type sequence) and the G74C mutant (a more dramatic decrease in the rate of cleavage) (Table 1). Such results do not lend themselves to simple interpretations, but may indicate a sequence requirement at position 74, superimposed on a requirement for a base pair. It is possible that a single C to G base change at position 43

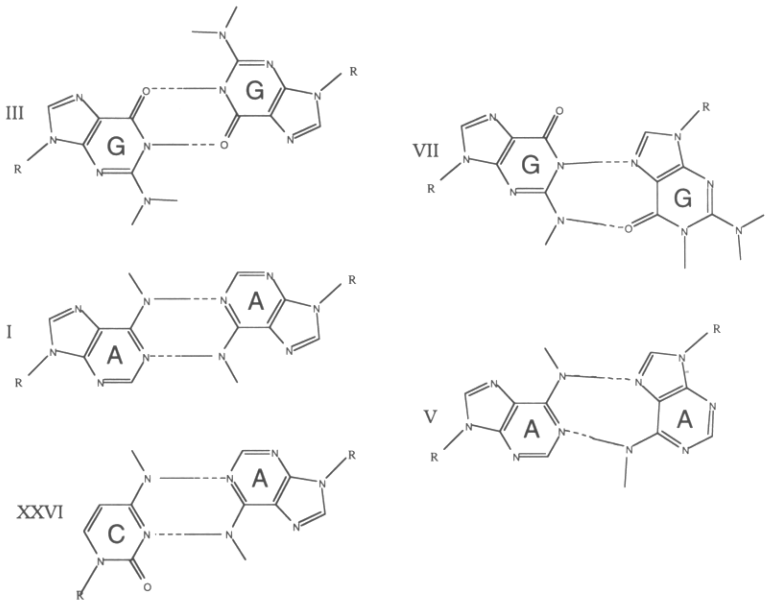


FIGURE 4. Potential base-pairing schemes at the end of P4. Pairings between G·G, A·A, and C·A are numbered according to Figure 6-1 in Saenger (1984).

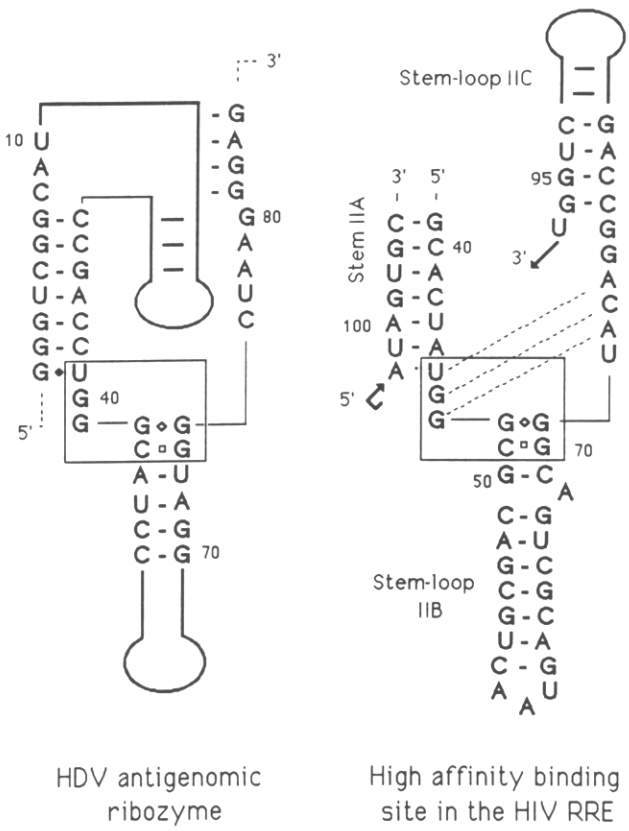


FIGURE 5. Comparison of secondary structures of the antigenomic ribozyme and the high-affinity Rev binding-site region of the HIV-1 RRE. The RRE sequence is redrawn from Tiley et al. (1992) and includes information from recent studies (Bartel et al., 1991; Giver et al., 1993; Battiste et al., 1994; Leclerc et al., 1994; Peterson et al., 1994). Dark arrows on the RRE sequence denotes continuity between nt 96 at the 3′ end of stem-loop IIC and nt 97 at the 5′ end of stem IIA. Thin dashed lines are additional base pairing interactions identified in the above references.

was accommodated as a mismatch and was not sufficiently destabilizing to have an effect on the structure or the cleavage reaction. In the presence of 10 M formamide, the ribozyme with the compensatory change (C43G:G74C) did cleave faster than the two single mutations (Table 1). This result suggests that, under conditions that reveal effects of destabilizing alterations, a Watson–Crick base pair did contribute to ribozyme structure and activity. Together, the data are consistent with both a base pair and a sequence requirement at the end of P4. Finally, in both the ribozyme and the RRE, the sequence UGG precedes the 5′ G of the homopurine pair.

A G·G pair in the genomic ribozyme

The genomic ribozyme differs from the antigenomic sequence in the J1/4 region and, because of those differences, it is not apparent if a similar G·G pair forms in that RNA. A pairing equivalent to the antigenomic G42·G75 would be predicted to involve G40 and G74 in the genomic sequence. If this is the case, there would be a 3-nt insertion (CAA) between G40 and C44, that is, assuming C44 forms the C-G pair at the end of P4 (Fig. 1B). Although there is an alternative way to pair P4 so that the genomic and antigenomic J1/4 region is the same (Branch & Robertson, 1991; Been et al., 1992) (Fig. 1B, structure on right), a natural isolate with sequence variation in this region (Casey et al., 1993) is more compatible with the genomic structure as originally proposed (Fig. 1B, structure on left) (Perrotta & Been, 1991).

The possibility of a homopurine pair was tested in two versions of the genomic ribozyme, one with the CAA

sequence (SD200) and one without (SD300) (Fig. 1B). In the sequence that more closely resembles the wild-type sequence, SD200, G to C mutations at either position 40 or position 74 (SD200/G40C, SD200/G74C, and SD200/G40C:G74C) resulted in dramatic loss of activity (Table 2), an effect very similar to that seen in the antigenomic sequence. However, the effect of the G to A changes at positions 40 and 74 was ambiguous. The G40A mutation had little or no effect, whereas the G74A mutation resulted in a 10-fold reduction in rate of cleavage. However, the G40A:G74A double mutant cleaves faster than SD200/G74A, but slower than the SD200/G40A (Table 2). Adding formamide to the reactions further slowed the rate of SD200/G40A:G74A. Thus, in SD200, the mutations used to test for the existence of a G·G pair neither supported nor ruled out the homopurine interaction. With regard to the effect of denaturant, we note that, unlike the two antigenomic constructs, 10 M formamide did not have much effect either way on the cleavage rates of SD200 or SD300.

An inspection of the secondary structure suggested a possible explanation for some of the above results. The G74A mutation, which reduced activity, may allow the purine at position 74 to form an A·A homopurine pair with A43 or A42 rather than with the purine at position 40. If so, this alternative homopurine pair at the end of P4 would disrupt the normal structure and could distort the active site. If an A43·A74 interaction were stable or favored over the A40·A74 interaction, then that would explain why the double mutant did not restore activity. To test this explanation, the same G to A mutations were made in SD300, the ribozyme lacking the CAA sequence between G40 and C44. The deletion had little effect on cleavage activity; SD300 cleaved approximately two- to threefold faster than SD200 (Table 2). In the SD300 ribozyme, single mutations at positions 40 and 74 reduced activity, whereas the double mutant (SD300/G40A:G74A) has full activity in both the absence and presence of formamide (Table 2). Thus, we conclude that it is very likely that a homopurine interaction does form in the genomic ribozyme, but it may be difficult to detect in the wild-type core sequence by this approach.

DISCUSSION

The secondary structure of the HDV ribozymes defines a region of "single-stranded" residues that may form the catalytic core of the ribozyme. Extensive mutagenesis suggests that there are essential sequences in this region (Kumar et al., 1992; Tanner et al., 1994); however, in the mutagenesis approach to defining structure, the information gained from simple loss of activity is of limited value unless a set of compensatory mutations can be found. We have used site-specific mutagenesis to reveal a functionally important interaction between two purines (G42 and G75) in the antigenomic ribozyme core region proximal to P4. Two mutations, G42C and G75C, had very large effects, almost completely knocking out activity. Either one, G42C or G75C, could extend P4 into the core region with Watson–Crick base pairs, suggesting that introduction of a stable Watson–Crick pair at this position disrupted a structure essential for forming the active site. However, the loss of activity was also seen with several mismatches, ruling out the simple interpretation that only non-Watson–Crick combinations function at that position. A clue to the nature of the interaction was revealed when full cleavage activity was restored by replacing both G's

TABLE 2. Cleavage activity of genomic ribozymes at 37° in the absence and presence of formamide.

	Mg^{2+}[a]		Mg^{2+}+F[b]	
Ribozyme/mutation	Rate constant (min^{-1})	% Cleaved[c]	Rate constant (min^{-1})	% Cleaved
SD200	6.8	54	6.3	73
SD200/G40C	[0.022][d]		[0.044]	
SD200/G74C	[0.0054]		[0.007]	
SD200/G40C:G74C	[0.0029]		[0.0001]	
SD200/G40A	6.1	58	5.4	75
SD200/G74A	0.50	67	0.16	64
SD200/G40A:G74A	1.5	58	0.40	70
SD300 (-CAA)	18	72	15	80
SD300/G40A	5.4	79	2.8	74
SD300/G74A	1.2	86	0.85	83
SD300/G40A:G74A	14	74	16	83

[a] 11 mM $MgCl_2$, 1 mM EDTA, and 40 mM Tris-HCl, pH 8.0, at 37°.
[b] 11 mM $MgCl_2$, 1 mM EDTA, 40 mM Tris-HCl, pH 8.0, and 10 M formamide at 37°.
[c] Percent cleavage is the end point of the reaction, essentially the percentage of the precursor that cleaved with the rate constant given.
[d] Values in brackets are averages of two determinations.

with A's. In addition to suggesting specific structures, this result argues against the possibility that exocyclic groups of either G42 or G75 are critical components of the ribozyme active site. Saenger (1984) describes seven possible homopurine pairs that are able to form two hydrogen bonds. The data were most consistent with a symmetric G·G pairing scheme (Fig. 4, III) at that position because the double G42A:G75A mutation restored full activity, and the symmetric A·A pair (Fig. 4, I) can form a structure isomorphic with the G·G pair (Saenger, 1984; Bartel et al., 1991). In the symmetric pairing scheme, the axis of symmetry is perpendicular to the plane of the base pair, with two N1-amino hydrogen bonds for the A·A pair and two N1-carbonyl hydrogen bonds for the G·G pair (Fig. 4, III and I). However, an alternative pairing scheme is not ruled out by the mutagenesis (Fig. 4, VII and V). Asymmetric homopurine pairings involving N1-N7 and amino-carbonyl hydrogen bonds for the G·G pair and amino-N7 and N1-amino hydrogen bonds for the A·A pair, although not isomorphic, are similar enough that it may be possible to substitute one for the other. Arguing against this alternative scheme, both guanosines (genomic and antigenomic RNAs) are protected from kethoxal addition when the RNA is folded into an active structure in the presence of Mg^{2+} at 37 °C (A. Perrotta & M. Been, unpubl. data). This result can be interpreted as support of the symmetric pairing scheme because, in the asymmetric pairing, one of the guanosines has both the N1 and N2 positions exposed and thus could be modified by kethoxal, provided these positions are not involved in other contacts. Another combination that worked, though less well than the homopurines, was the C·A combination (G42C:G75A), whereas reversing the orientation (G42A:G75C) resulted in decreased activity.

With the genomic ribozyme, interpretation of the data was less straightforward, but it still suggested a homopurine interaction. Clearly, G to C changes at either position reduced cleavage activity severely. However, with the unmodified genomic ribozyme (SD200), the G to A change at position 40 had a small effect. It is possible that the latter mutation had little effect because either the proposed homopurine pairing is not critical for stabilizing the structure of the active site when the CAA sequence is present, or a rate-limiting step in self cleavage is not affected despite disrupting the homopurine pair. Nevertheless, this could be only part of the explanation because the SD200/G74A had reduced activity, and the double mutant, SD200/G40A:G75A, did not restore full activity. Strong support for the equivalent homopurine pair in the genomic ribozyme required deleting the J1/4 CAA sequence and, as noted in the Results, it is possible that this modification was necessary to prevent a favorable but inactive alternative structure from forming in the presence of the G74A mutation. The CAA mutation on its own had little effect, increasing cleavage activity only threefold relative to SD200. It is worth noting that, although we observed a small increase in activity with the CAA deletion, Tanner et al. (1994) find that the same deletion in their genomic ribozyme resulted in a small decrease (4–14-fold) in the rate of cleavage. Also, they find that, in a wild-type background, the G40A mutant cleaves 4–50 times slower than the unmodified ribozyme, an effect we would have predicted but did not see in our construct. The G74A mutation was not tested in that study, but a G75C mutation had an effect similar to what we found. Although there are some minor sequence differences in 5′ and 3′ flanking sequences between the two constructs, there are no differences in the core region, and we therefore cannot easily account for our different results.

Evidence for other examples of G·G pairs exist for the splice-site interaction of nuclear pre-mRNA intron (Parker & Siliciano, 1993), and between positions 722 and 733 in the 16 S rRNA (*Escherichia coli* numbering) (Gutell & Woese, 1990). However, the biochemical data for the ribozyme G·G pair appears to parallel the data for the RRE, and the more detailed information available for the RRE may provide some insight into the possible structure of the ribozyme core. A G·G homopurine pair (Fig. 4) in the high-affinity Rev binding site of the RRE was first postulated on the basis of Rev binding sequence variants (A·A and C·A) isolated from randomized pools of RNA (Bartel et al., 1991), and evidence for that particular G·G pairing was obtained from NMR studies (Battiste et al., 1994; Peterson et al., 1994). The same symmetric pairing scheme is also most consistent with the data obtained for the antigenomic ribozyme. This pairing will distort the helical structure at the end of the P4 helix because, relative to a Watson–Crick pair, one of the two bases is flipped over. If the pair is to be accommodated into antiparallel strands at the end of a helix, one of the nucleotides will either have the glycosydic torsion angle in the *syn* conformation (Saenger, 1984; Bartel et al., 1991; Peterson et al., 1994), or, if both nucleotides are *anti*, the phosphate backbone will have to be distorted from the normal antiparallel structure (Battiste et al., 1994). Both scenarios have been suggested for the RRE structure based on NMR data. In the RRE, the formation of the homopurine pair is accompanied by the displacement of an adjacent pyrimidine (U72) out of the helix. In both the genomic and antigenomic ribozyme, there is an essential pyrimidine in the equivalent position (C75, genomic [Tanner et al., 1994] and C76, antigenomic [unpubl. data]) that could have a critical role in the cleavage reaction (Tanner et al., 1994). It is possible that one function of the G·G pair would be to position that cytosine in the active site. The similarity in terms of an adjacent and critical C-G pair has already been mentioned, and it is noteworthy that chemical probing of the genomic ribozyme (Kumar et al., 1994) may indicate possible

nonstandard interactions with the bases at the end of P4. Similar data is not available for the antigenomic ribozyme.

The loop region of the RRE has been shown to be highly structured in the peptide-bound form (Battiste et al., 1994; Peterson et al., 1994). For the ribozymes, it would appear from the secondary structure that there is potential for pairing between either G41 or G40 and C76 in the antigenomic sequence. Equivalent interactions are possible in the genomic sequence. In the antigenomic ribozyme, mutations at those positions dramatically lower activity, but potential compensatory changes G41C:G76G and G41A:G76U are completely inactive for self-cleavage. Likewise, base substitutions at G40 result in an inactive ribozyme and the G40A:C76U double substitution is inactive (A. Perrotta & M. Been, in prep.). This would suggest that if pairing occurs, the identity of the participating bases is important.

A three-dimensional model for the genomic ribozyme has been proposed that incorporates all the features of the secondary structure and brings together many of the essential bases in single-stranded regions of the putative core (L3, J1/4, and J4/2) (Tanner et al., 1994). In this model, however, the G40 and G74 are not close enough to form a base pair. A new model that takes this new constraint into account will have to make adjustments to the path taken by J1/4 and the J4/2 segments. Given the evidence that there may be some structural similarity between the HDV ribozyme core and the high-affinity Rev binding site, it now may be possible to use the physical data and current models available for the RRE RNA to refine models for the core structures of the HDV ribozymes.

MATERIALS AND METHODS

Enzymes, chemicals, and oligonucleotides

T7 RNA polymerase was purified from an overexpressing clone kindly provided by W. Studier (Davanloo et al., 1984). Modified T7 DNA polymerase (Sequenase) was purchased from Amersham. Restriction endonucleases, nucleotides, ^{32}P-labeled nucleotides, and chemical reagents were purchased from commercial suppliers. Oligodeoxynucleotides for mutagenesis were obtained from the in-house synthesis facility (Department of Botany, Duke University).

Plasmids and construction of mutants

Two versions of the antigenomic ribozyme sequence were used. Both sequences are cloned into the T7 promoter containing phagemid vector pTZ18U. pSA1-2 (Perrotta & Been, 1991) was constructed with a synthetic version of the sequence of the antigenomic ribozyme inserted downstream of a T7 promoter; it is nearly identical to the wild-type antigenomic self-cleaving sequence except for minor changes that introduced a restriction recognition site in the sequence comprising stem-loop 4. For the ribozyme dSIV (Been et al., 1992), stem-loop 4 is shortened to four base pairs plus 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 (Kunkel et al., 1987; Vieira & Messing, 1987) as described previously (Perrotta & Been, 1991).

The plasmid containing the genomic ribozyme (pSD200) was prepared using PCR (Saiki et al., 1985) to amplify a region of HDV cDNA containing the genomic ribozyme sequence from a plasmid clone provided by Belinsky and Dinter-Gottlieb (1991). The PCR primers added restriction sites flanking the ribozyme sequence to facilitate cloning into a T7 promoter containing plasmid (pTZ18U), but did not alter sequence within the 85 nt that encompass the sequences essential for optimal ribozyme activity (Perrotta & Been, 1990). The full sequence of the resulting ribozyme (SD200) is: 5′ GGAAUU CUUA CCUGAU-GGCCGGCAUG GUCCCAGCCU CCU CGCUGGC GCCGGCUGGG CAACAUUCCG AGGGGA CCGU CCCCUCGGUA AUGGCGAAUG GGAUC, where the dash indicates the cleavage site. Deletions and base changes to the initial sequence were made as described above (Kunkel et al., 1987; Perrotta & Been, 1991).

Transcriptions

Template DNA was prepared by linearizing plasmid DNA with *Hin*d III (pSA1-2 and pdSIV) or *Bam*H I (pSD200 and pSD300). Typically, transcriptions were conducted 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 of [α^{32}P] CTP, 2.5 μg DNA, and 300 units of T7 RNA polymerase. Incubation was for 60 min at 37 °C, then EDTA was added to 50 mM, formamide to 50% (v/v), and the RNA was fractionated by electrophoresis on a 6% polyacrylamide gel containing 7 M urea. RNA was located by autoradiography and eluted from the gel slice. The self-cleavage reaction requires only low levels of a divalent cation, and, as a result, many of the transcripts cleave 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 the transcription reaction to block self-cleavage. We assume that the oligonucleotide anneals to the transcript and interferes with ribozyme folding. Although this resulted in a lower overall yield of RNA, it increased the ratio of uncleaved to cleaved RNA. For the slow-cleaving mutants, a blocking oligodeoxyribonucleotide was not needed 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 the addition of $MgCl_2$ (37 °C); final conditions were 40 mM Tris-HCl, pH 8.0, 1 mM EDTA, 11 mM $MgCl_2$, and approximately 5–50 nM RNA. For reactions that contained 10 M formamide, the formamide was included in the preincubation. The kinetics of cleavage was followed by removing and mixing 5-μL aliquots with 10 μL of formamide-dye mix containing 50 mM EDTA to quench the

reaction. The precursor and product were separated by gel electrophoresis under denaturing conditions (6% polyacrylamide gel containing 7 M urea, 0.05 M Tris-borate, pH 8.3, and 0.5 mM EDTA). The relative amounts of precursor and 3′ cleavage product were quantified by analysis on a PhosphorImager (Molecular Dynamics). The 5′ product migrates off these gels and was not included directly in the analysis, but a correction was made by multiplying counts in the 3′ product by the ratio of the number of labeled nucleotides in the precursor divided by the labeled nucleotides in the 3′ product. When the 5′ product contained less than 5% of the labeled nucleotides, this correction became insignificant and was not included. The fraction cleaved (F) was calculated as ($\text{counts}^{\text{product}}$)/($\text{counts}^{\text{precursor}} + \text{counts}^{\text{product}}$). The first-order rate constant, k, and endpoint, m, were obtained by fitting the data to the equation $F = m(1 - e^{(-kt)})$. For the faster cleaving sequences, the estimated endpoint was compared to data from time points taken at 5–60 min (10–200 times the half-life of the precursors). At these extended times, the experimentally determined end point often exceeded the value obtained in the above equation by 5–10%, suggesting that there may be fast- and slow-cleaving precursor populations in those preparations. Correction for these differences did not result in significant changes in estimated rate constants, so all ribozyme sequences were analyzed as if they contained a major species that cleaved with simple first-order kinetics. Thus, the rate constants given in the tables are for either the fastest cleaving species or represent some average of more than one fast cleaving species. For the two antigenomic precursors containing a wild-type core sequence, SA1-2 and dSIV, the reaction is essentially complete after 1 min ($k > 5$/min), and, because the earliest time points are taken at 4 or 6 s, values for rate constants greater than 5/min are most likely minimal estimates and subject to larger handling errors. Thus, although reproducibility was good (±10–20%, with different preparations of precursors), we might be underestimating the fastest reactions. For the mutated precursor RNAs that cleaved slower, reproducibility was ±10–20%. All rate constants reported are the average of at least three independent determinations, unless noted otherwise.

ACKNOWLEDGMENTS

We thank T. Wadkins, M. Puttaraju, and G. Wickham for comments on the manuscript, S. Towne for help with its preparation, and numerous colleagues with interests in RNA structure and function for their insights and suggestions. This work was supported by a grant from the NIH (GM47233).

Received October 6, 1995; returned for revision October 19, 1995; revised manuscript received November 7, 1995

REFERENCES

Bartel DP, Zapp ML, Green MR, Szostak JW. 1991. HIV-1 Rev regulation involves recognition of non-Watson–Crick base pairs in viral RNA. *Cell 67*:529–536.

Battiste JL, Tan R, Frankel AD, Williamson JR. 1994. Binding of an HIV Rev peptide to Rev responsive element RNA induces formation of purine–purine base pairs. *Biochemistry 33*:2741–2747.

Been MD. 1994. *Cis*- and *trans*-acting ribozymes from a human pathogen, hepatitis delta virus. *Trends Biochem Sci 19*:251–256.

Been MD, Perrotta AT, Rosenstein SP. 1992. Secondary structure of the self-cleaving RNA of hepatitis delta virus: Applications to catalytic RNA design. *Biochemistry 31*:11843–11852.

Belinsky MG, Dinter-Gottlieb G. 1991. Non-ribozyme sequences enhance self-cleavage of ribozymes derived from hepatitis delta virus. *Nucleic Acids Res 19*:559–564.

Branch AD, Robertson HD. 1991. Efficient *trans* cleavage and a common structural motif for the ribozymes of the human hepatitis δ agent. *Proc Natl Acad Sci USA 88*:10163–10167.

Casey JL, Brown TL, Colan EJ, Wignall FS, Gerin JL. 1993. A genotype of hepatitis D virus that occurs in northern South America. *Proc Natl Acad Sci USA 90*:9016–9020.

Davanloo P, Rosenberg AH, Dunn JJ, Studier FW. 1984. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc Natl Acad Sci USA 81*:2035–2039.

Giver L, Bartel D, Zapp M, Pawul A, Green M, Ellington AD. 1993. Selective optimization of the Rev-binding element of HIV-1. *Nucleic Acids Res 21*:5509–5516.

Gutell RR, Woese CR. 1990. Higher order structural elements in ribosomal RNAs: Pseudo-knots and the use of noncanonical pairs. *Proc Natl Acad Sci USA 87*:663–667.

Kumar PKR, Suh YA, Miyashiro H, Nishikawa F, Kawakami J, Taira K, Nishikawa S. 1992. Random mutations to evaluate the role of bases at two important single-stranded regions of genomic HDV ribozyme. *Nucleic Acids Res 20*:3919–3924.

Kumar PKR, Suh YA, Taira K, Nishikawa S. 1993. Point and compensation mutations to evaluate essential stem structures of genomic HDV ribozyme. *FASEB J 7*:124–129.

Kumar PKR, Taira K, Nishikawa S. 1994. Chemical probing studies of variants of the genomic hepatitis delta virus ribozyme by primer extension analysis. *Biochemistry 33*:583–592.

Kunkel TA, Roberts JD, Zakour RA. 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol 154*:367–382.

Kuo MYP, Sharmeen L, Dinter-Gottleib G, Taylor J. 1988. Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *J Virol 62*:4439–4444.

Leclerc F, Cedergren R, Ellington AD. 1994. A three-dimensional model of the Rev-binding element of HIV-1 derived from analyses of aptamers. *Nature Structural Biol 1*:293–299.

Parker R, Siliciano PG. 1993. Evidence for an essential non-Watson–Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature 361*:660–662.

Perrotta AT, Been MD. 1990. The self-cleaving domain from the genomic RNA of hepatitis delta virus: Sequence requirements and the effects of denaturant. *Nucleic Acids Res 18*:6821–6827.

Perrotta AT, Been MD. 1991. A psuedoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA. *Nature 350*:434–436.

Perrotta AT, Been MD. 1993. Assessment of disparate structural features in three models of the hepatitis delta virus ribozyme. *Nucleic Acids Res 21*:3959–3965.

Peterson RD, Bartel DP, Szostak JW, Horvath SJ, Feigon J. 1994. ^{1}H NMR studies of the high-affinity Rev binding site of the Rev responsive element of HIV-1 mRNA: Base pairing in the core binding element. *Biochemistry 33*:5357–5366.

Puttaraju M, Perotta AT, Been MD. 1993. A circular *trans*-acting hepatitis delta virus ribozyme. *Nucleic Acids Res 21*:4253–4258.

Rosenstein SP, Been MD. 1991. Evidence that genomic and antigenomic RNA self-cleaving elements from hepatitis delta virus have similar secondary structures. *Nucleic Acids Res 19*:5409–5416.

Saenger W. 1984. *Principles of nucleic acid structure*. New York: Springer-Verlag.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science 230*:1350–1354.

Sharmeen L, Kuo MYP, Dinter-Gottlieb G, Taylor J. 1988. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J Virol 62*:2674–2679.

Suh YA, Kumar PKR, Taira K, Nishikawa S. 1993. Self-cleavage activity of the genomic HDV ribozyme in the presence of various divalent metal ions. *Nucleic Acids Res 21*:3277–3280.

Symons RH. 1992. Small catalytic RNAs. *Annu Rev Biochem 61*:641–671.

Tanner NK, Schaff S, Thill G, Petit-Koskas E, Crain-Denoyelle AM, Westhof E. 1994. A three-dimensional model of hepatitis delta vi-

rus ribozyme based on biochemical and mutational analyses. *Curr Biol 4*:488–497.

Thill G, Vasseur M, Tanner NK. 1993. Structural and sequence elements required for the self-cleaving activity of the hepatitis delta virus ribozyme. *Biochemistry 32*:4254–4262.

Tiley LS, Malim MH, Tewary HK, Stockley PG, Cullen BR. 1992. Identification of a high-affinity RNA-binding site for the human immunodeficiency virus type 1 Rev protein. *Proc Natl Acad Sci USA 89*:758–762.

Vieira J, Messing J. 1987. Production of single-stranded plasmid DNA. *Methods Enzymol 153*:3–11.

Wu HN, Lin YJ, Lin FP, Makino S, Chang MF, Lai MMC. 1989. Human hepatitis δ virus RNA subfragments contain an autocleavage activity. *Proc Natl Acad Sci USA 86*:1831–1835.

Wu HN, Wang YJ, Hung CF, Lee HJ, Lai MMC. 1992. Sequence and structure of the catalytic RNA of hepatitis delta virus genomic RNA. *J Mol Biol 223*:233–245.

Optimal self-cleavage activity of the hepatitis delta virus RNA is dependent on a homopurine base pair in the ribozyme core.

M D Been and A T Perrotta

RNA 1995 1: 1061-1070