3-D models of the antigenomic ribozyme of the hepatitis delta agent with eight new contacts suggested by sequence analysis of 188 cDNA clones

Andrea D. Branch^{1,2,3,*} and Jana A. Polaskova^{1,2,4}

¹Division of Liver Diseases, Department of Medicine, ²Division of Abdominal Organ Transplantation, Department of Surgery and ³Department of Microbiology, Mount Sinai Medical Center, One Gustave L. Levy Place, New York, NY 10029, USA and ⁴Department of Biochemistry, Cornell University Medical College, 1300 York Avenue, New York, NY 10021, USA

Received May 30, 1995; Revised and Accepted September 14, 1995

ABSTRACT

We mapped 359 mutations at 25 positions in synthetic variants of the antigenomic ribozyme of the hepatitis delta agent by analyzing the sequences of 188 cDNA clones. These data were used to identify three features of the ribozyme: highly conserved nucleotides, positions with restricted nucleotide substitutions and three-dimensional relationships between nucleotides. The distribution of mutations at the 25 positions was as follows: G-11 (the eleventh nucleotide from the cleavage site) was mutated in 56 clones; G-12 in 36; U-15 in 33; C-13 in 26; G-28 in 23; C-27 in 21; C-29 in 19; U-26 in 17; C-18 in 14; A-14 in 13; C-16 in 13; C-19 in 12; U-17 in 11; A-20 in 10; G-42 in 9; G-40 in 7; G-41 in 7; C-24 in 6; U-32 in 6; U-23 in 5; C-25 in 4; C-21 in 3; G-30 in 3; G-31 in 3; C-22 in 1. All clones containing a mutation at C-25 had an A at this position, suggesting that the extra cyclic amino group present in adenine and cytosine may function during the cleavage event. Mutations at certain positions were common in simple clones (containing only one or two mutations), while mutations at other positions were over-represented in more complex clones. Both compensatory base changes and co-mutational frequencies were used to identify eight pairs of nucleotides which may interact with each other: G-11 and C-18, G-12 and C-27, C-13 and G-28, C-21 and U-23/C-24, C-21 and G-30, U-23 and G-31/U-32, C-24 and G-30, C-27 and G-42. These pairs, which involve some of the most conserved positions in the molecule, suggest interactions among nucleotides previously depicted in open-loop structures. The newly proposed points of contact between pairs of nucleotides are compatible with both the axehead and pseudoknot secondary structural models and were combined with previously proposed Watson-Crick base paired helices to produce two three dimensional models. In both of these, C-25 and C-76 are placed near the cleavage site.

INTRODUCTION

The delta hepatitis agent (1,2), like several other circular RNA pathogens (3), appears to use ribozymes (4-6) to complete its RNA to RNA rolling circle replication cycle (7-10). The delta self-cleavage structures are relatively small ribozymes. Each is ~90 nucleotides in length (4,5,11-13). In part because they evolved to function in human cells and have been re-engineered to carry out *trans*-cleavage (14,15), the delta ribozymes have attracted interest as potential ribocutical agents. Both comparative sequence analysis and a variety of biochemical studies indicate that the genomic and antigenomic ribozymes have closely related, but not identical, structures. Several models for each ribozyme have been proposed, tested, refined (14,16-35) and reviewed (13). Despite the progress that has been made, more work is needed to establish the details of the active structures. The very high resistance of the delta antigenomic ribozyme to inactivation by denaturing agents, such as 18 M formamide (36), suggests that it has a very stable structure and may contain many bonds which have not yet been described. Furthermore, since most studies have focused on the genomic ribozyme, the antigenomic ribozyme is in particular need of further investigation.

As revealed by the *trans*-cleavage hammerhead studies of Uhlenbeck (37) and Haseloff and Gerlach (38), it is possible to separate a self-cleavage structure into a ribozyme domain and a target (substrate) domain in different ways. Depending upon how the division is made, larger or smaller portions of the catalytic core may reside in the substrate. We previously demonstrated that the self-cleavage structures of delta genomic and antigenomic RNA can each be divided into a relatively small (~25 base-long) *trans*-active domain and a larger (~65 base-long) substrate domain containing the phosphodiester bond which is the target of *trans*-cleavage (14). Such substrates allow a major portion of the delta self-cleavage structures to be explored through mutational analysis. In the current study, mutations were introduced at 25 positions in the antigenomic substrate. A mixed population of these mutant molecules were incubated with the *trans*-active domain in the presence of magnesium ions. Molecules which

* To whom correspondence should be addressed

cleaved under these conditions were cloned and sequenced. Technically, the molecules undergoing cleavage are substrates because they contain the phosphodiester bond which is broken. However, because they also contain nucleotides which comprise part of the active site of the antigenomic self-cleavage structure, we refer to them as 'active ribozyme variants' in the remainder of this report to focus attention on this aspect of their structure.

MATERIALS AND METHODS

Manufacture of DNA templates

DNA templates for the mutant antigenomic delta ribozyme variants were synthesized under the supervision of Mr Scott Geromanos in the Microchemistry Core Facility at the Rockefeller Research Laboratories of the Memorial Sloan-Kettering Cancer Center. Templates were 84 nucleotides in length and contained mutations at 25 positions. At these 25 positions, 82% of the DNA molecules contained the template for the wild-type nucleotide, while 6% of the molecules contained the template for each of the other three nucleotides. To synthesize DNA containing mutations, four specially-prepared cocktails were used, in addition to the usual precursors. Resulting template DNAs had the following sequence: 5'-3' GCGTCCTCCTTCGGATGWW-WAGGTCGGXWWYWXYYXYYZYXXYZYWWATGC-CGACCCGAAGAGGCTATAGTGAGTCGTATTA. In this sequence, W represents positions in which C occurred in 82% of the templates and A, G and U each occurred in 6% of the templates; Y represents positions in which G occurred in 82% of the templates and A, C and T each occurred in 6% of the templates; X represents positions in which A occurred in 82% of the templates and C, G and T each occurred in 6% of the templates; Z represents positions in which T occurred in 82% of the templates and C, A and G each occurred in 6% of the templates. Thus, $\sim 0.7\%$ (0.82^{25}) of the DNA molecules were templates for the wild-type sequence.

Selection, reverse transcription and cloning of active ribozyme variants

DNA templates were purified by gel electrophoresis and then transcribed into RNA under the conditions described by Milligan *et al.* (39) in the presence of [α - 32 P]GTP. RNA was purified by electrophoresis in a 10% polyacrylamide gel containing 7 M urea. To cleave active ribozyme variants, 7 pmol of these molecules were mixed with 10 pmol of the delta antigenomic *trans*-cleaving domain, described previously (14), in a 12 μ l reaction also containing 5 mM MgCl₂ and 5 mM HEPES, pH 7.5 and were incubated at 50°C for 40 min. Products were fractionated by electrophoresis in a 10% polyacrylamide gel containing 7 M urea. Molecules migrating to the position of the 3' cleavage product, described previously (14), were eluted from the gel.

Two PCR primers were synthesized and purified by gel electrophoresis. One of these primers, 5'-CTGCAGCGTCCTCC-TTCGG-3', is complementary to the 3' terminal portion of the antigenomic ribozyme variants and was used in the reverse transcriptase reaction as well as for PCR amplification. The other primer, which terminated with the sequence 5'-GGGTCGGC-AT-3', was used for PCR only. For cDNA synthesis, reactions were incubated at 42°C for 30 min with SuperScript RNaseH⁻ Reverse Transcriptase, using conditions specified by the supplier (Gibco BRL). At the end of the reaction, RNase H (Gibco BRL) was added

Figure 1. Experimental design.

and reactions were incubated for 10 min. Perkin-Elmer GeneAmp PCR Core Reagents including AmpliWax PCR and AmpliTaq DNA Polymerase were used to perform 30 rounds of amplification. The resulting double-stranded DNA inserts were then phosphorylated by incubation with T4 polynucleotide kinase (Promega) and ATP and ligated into pSP64 (Promega), which had been previously cleaved by *Sma*I (Promega), treated with calf intestinal alkaline phosphatase (Promega) and purified by gel electrophoresis. DNA was then used to transform MAX Efficiency DH10B Competent Cells, according to the protocol of the supplier (Gibco). Individual ampicillin-resistant colonies were selected and grown overnight in 3 ml of media containing 50 μ g/ml of filter-sterilized ampicillin (Sigma). Plasmid DNA minipreps, prepared as specified by Yie *et al.* (40), were analyzed using a Sequenase Version 2.0 DNA Sequencing Kit (United States Biochemical) and a single-stranded DNA primer (5'-GATT-TAGGTGACACTATAG-3'; Promega). 210 clones containing inserts of the appropriate size were selected for detailed analysis. The sequence of each clone was read independently by at least two individuals. One clone was discarded because its sequence could not be read with certainty; 21 clones were set aside because one or two nucleotides were deleted. The 188 clones containing no deletions or insertions were selected for further analysis.

RESULTS

Experimental design

To generate and select active ribozyme variants, the experimental design outlined in Figure 1 was employed. Using the customized phosphoramidite cocktails described in the Materials and Methods section, template DNAs were synthesized with muta-

Figure 2. Three proposed models of the delta antigenomic and genomic ribozymes. The boxed nucleotides identify the 25 antigenomic nucleotide positions mutated in this study in the axehead model (14), (A) the pseudoknot model (16), (B) and a recent model of Gottlieb and colleagues (35), (C). Corresponding genomic structures appear in (D)–(F), (see refs 14,16,29). In the antigenomic axehead model, the self-cleavage structure is divided into two domains, illustrating the two types of RNA molecules we used for *trans*-cleavage reactions.

tions at 25 positions. The sites of the mutated nucleotides are shown in Figure 2A–C, which depicts the antigenomic delta ribozyme in three different configurations, or models. For comparison, the delta genomic ribozyme is also presented as it appears in three models (Fig. 2D–F). Nucleotides are numbered consecutively beginning with the first nucleotide after the cleavage site. The DNA templates were synthesized so that they contained the promoter sequence for bacteriophage T7 RNA polymerase and thus could be transcribed *in vitro* under the conditions described by Milligan *et al.* (39).

To identify active delta ribozyme variants, gel-purified RNA transcripts were incubated with magnesium ions and the *trans*-cleaving delta antigenomic RNA. These two domains are presented in Figure 2A, which depicts the axehead configuration of the delta antigenomic ribozyme. Following incubation with the *trans*-cleaving domain, cleaved (active) variants were separated from uncleaved (inactive) variants by gel electrophoresis and then were eluted from gels, copied by reverse transcriptase,

amplified by PCR, cloned and sequenced (as described in the Materials and Methods section).

Sequence analysis

The sequences of cDNA clones representing synthetic variants of the antigenomic ribozyme, are given in Figure 3 and its legend. Each clone is a composite, containing two portions supplied by the PCR primers and a central region, 35 nucleotides in length, which was copied from RNA by reverse transcriptase. The wild-type sequence of these 35 nucleotides is presented in Figure 4. This central region contains the 25 nucleotides targeted for mutagenesis and an additional 10 nucleotides. We began by carrying out a quality control analysis on these 10 positions, which were not the targets of mutagenesis.

Among the 188 clones which contain no deletions or insertions, a total of three mutations were found in these 10 positions, indicating a maximum artifactual mutation rate of 3 in 1880

nucleotides (0.16%). The actual rate of artifactual mutations was probably lower than 0.16% for the following reason. Mutations at these 10 positions could arise during each of several procedures: DNA template synthesis, RNA transcription, reverse transcription, PCR amplification, cloning, sequencing. Only mutations arising from reverse transcription, PCR amplification, cloning and sequencing were created after cleavage occurred and thus constitute mutations which we detect but which were not actually present in active molecules. Unfortunately, it is not possible to distinguish between base changes (resulting from errors in DNA template synthesis and RNA transcription) that were present in active ribozyme variants from mutations introduced after cleavage (by errors during reverse transcription or PCR amplification). However, the low incidence of mutation in the 10 control positions indicates that the vast majority of the mutations we detected in the 25 targeted positions were present in active ribozyme variants at the time of cleavage.

Of the 188 clones, 16 (8.5%) contained the wild-type nucleotide at each of the 25 positions targeted for mutagenesis, while only ~0.7% of the DNA templates contained the wild-type sequence (see Materials and Methods section). Therefore, the process used to select active molecules enriched the wild-type sequence ~10-fold. The 172 clones containing mutations were divided into 25 groups, so that there was one group for each of the 25 positions targeted for mutagenesis. At certain positions all mutant clones contained the same mutant nucleotide; at other positions only two of the three possible mutant nucleotides were found. As indicated in Table 1, C-25 was mutated in four clones and in each case was replaced by A-25. This result may indicate that position 25 has a requirement for the extra cyclic amino group present on adenine and cytosine.

Table 1. Restricted nucleotide substitutions

The three clones with a mutation at C-21 all contained U-21.
The one clone with a mutation at C-22 contained U-22.
The four clones with a mutation at C-25 all contained A-25.
The three clones with a mutation at G-31 all contained A-31.
None of the 12 clones mutated at C-19 contained G-19.
None of the 9 clones mutated at A-20 contained U-20.
None of the 5 clones mutated at U-23 contained C-23.
None of the 6 clones mutated at C-24 contained A-24.

Mutations were not evenly distributed among the 25 positions targeted for mutagenesis. Rather, some positions were mutated in many more clones than others (see Fig. 4). For example, the G at position 11 was replaced by an A, C or U in 56 of the 188 clones, while the C at position 22 was replaced by a different nucleotide in only one of the 188 clones. Such differences provide the basis for the distinction made throughout this report between highly conserved and poorly conserved bases. In nature, however, all 25 bases are highly conserved. Within these 25 bases, only a single mutation (at C-13) has been reported in sequenced isolates of the delta agent (41). During the replication cycle, the ribozyme sequences are constrained by factors in addition to those directly related to their catalytic capability; such as their ability to function as templates for replication, which may impose sequence restrictions (42); ability to form the rodlike structure (9); and their ability to generate the UV-sensitive tertiary structure (43,44)

proposed as a control element for the genomic ribozyme (45,46). In our current *in vitro* study, in which none of these factors had an effect, the ranking of nucleotides from the least conserved to the most conserved was as follows: (the number of clones with a mutation at a particular position is given in parenthesis): G-11 (56); G-12 (36); U-15 (33); C-13 (26); G-28 (23); C-27 (21); C-29 (19); U-26 (17); C-18 (14); A-14 (13); C-16 (13); C-19 (12); U-17 (11); A-20 (10); G-42 (9); G-40 (7); G-41 (7); C-24 (6); U-32 (6); U-23 (5); C-25 (4); C-21 (3); G-30 (3); G-31 (3); C-22 (1). In well characterized RNAs, such as tRNA (47), nucleotides contributing to tertiary structure are known to be highly conserved. In ribozymes, conserved bases include both nucleotides playing critical structural roles and those required for cleavage. Thus, in our study, we expect an inverse relationship between the number of constraints placed on a given nucleotide and the number of mutant clones in which it appears.

While mutations at certain positions were commonly found in clones containing only one or two mutations, mutations at other positions were over-represented in clones containing a larger than average number of mutations (see asterisks in Fig. 4). When building three dimensional models of the ribozyme, these positions, C-21, U-23, C-24, C-25, U-26, C-27, G-30, G-31, G-40, G-41 and G-42 were placed either in the interior of the molecule and/or close to the cleavage site (see Fig. 5). Complex rearrangements may be needed to accommodate changes at these positions.

To seek evidence of contact between pairs of nucleotides, both compensatory base changes and co-mutational frequencies were taken into account (see Table 2). In this analysis, each Watson-Crick compensatory base change was identified and given the value of two; each G-U substitution, the value of one. Inspection of our sequencing data revealed that mutations at certain nucleotides occurred more frequently in clones containing mutations at certain other nucleotides. To identify pairs of nucleotides with a high incidence of co-mutation, a multi-step process was used. First, clones containing single point mutations were eliminated from the analysis, since they could not contain examples of co-mutation. Secondly, the distribution of mutant nucleotides was determined for the remaining 119 clones. Thirdly, as illustrated for the case of G-11, we counted the number of mutations outside G-11, but present in clones in which G-11 was also mutated. Clones with mutations at G-11 contain a total of 72 mutations at other nucleotides. Using this number, 72, as a starting point, the number of mutations expected at each of the other 24 positions was determined (based on the distribution of mutations in the 119 clones). The number of mutations actually present at each of the other 24 positions was divided by the number expected to be present if the distribution of these mutations were identical to the distribution of mutations in the total population. This process, carried out by application of a Quattro Pro 5.0 spreadsheet, yielded the 'co-mutation value'. For example, clones with mutations at G-11 contain 12 mutations at G-12. Based on the distribution of mutations in the 119 clones, 9.7 of any 72 mutations would be expected to occur at G-12. The value of 12 was divided by 9.7 to produce a co-mutation value, 1.2, which was added to the value of the compensatory base changes, 0, in this case. Since this sum, 1.2, was low, we conclude that our data provide no evidence of an interaction between G-11 and G-12.

Only 20 of the 300 nucleotide pairs had a sum value >7, which we used as an arbitrary cut-off point. From this group of 20, the

Step	Class	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	40	41	42
22	187	G	G	C	A	U	C	U	C	A	C	U	C	U	C	U	C	G	C	G	U	U	G	G	G	G
21	3										U	U	U													
	15					C	U	C					U	A				A								
	354												U									A	C			
30	354					C	U					U											A	C		
	86																					C				
	82					C												A				U	U			C
31	60																									
	106			U																		U				U
	562																							A	A	
25	82					C																U	U			C
	106	A																								
	534	U				A																				
	564			U																				A	A	C
23	15							C				U		A	A											
	201			U		G								A	A											
	203	A						A						A	G											
	533			A				C							G											
	562														G										A	C
24	67		A																							
	98																									
	356					C	U					U														
	113		U																							U
	246					G																				
	341	A							A																	U
40	249			A					U																	A
	82					C																				C
	185																									U
	182																									U
	220		A																							U
	246			G																						U
41	511																									A
	46	C																								C
	127																									C
	206	A	C																							C
	10																									U
	113		U			G																				U
	552		U																							U
32	8		U																							A
	7			U																						A
	259																									A
	531			A																						A
	82					G																				A
	562																									C
	30		A																							G
42	21	A	U																							A
	54																									A
	183							G																		A
	553			U																						A
	347		U																							A
	16							G																		U
	100			U																						U
	205		U		U																					U
20	66											C	C	C	C											U
	247	C																								U
	357					A		A																		U
	8		U																							U
	8																									U

[illegible][illegible][illegible]

Figure 4. Frequency of mutations. The total number of clones containing a mutation at a particular position is given. In addition, the number of clones with one, two, three, four or five mutations is indicated. Positions which contain more than the average number of mutations in mutated clones are marked by an asterisk.

additional criteria presented in the legend to Table 2, were used to identify eight pairs considered to have the greatest potential significance. Models were built in which these eight pairs of nucleotides were placed close enough to each other to permit contact. As illustrated in Figure 5, the newly proposed points of contact were compatible with the helices of both the axehead model (Fig. 5A) and the pseudoknot model (Fig. 5B).

When data about the incidence of a mutation at a particular site, the complexity of the clones bearing mutations at that site and nucleotide restrictions were taken into account, the 25 mutated nucleotides fell into several groups:

(i) Nucleotides G-11, G-12, C-13 and U-15 all occur near the 5' end of the 25 targeted positions; were the four most frequently mutated nucleotides; and were mutated in simple clones (clones with only one or two mutations), suggesting that they may be on the outer surface of the molecule. An association between G-11 and C-18 was suggested by the presence of compensatory base changes in four clones (see Table 2). This interaction is compatible with the pseudoknot model in which C-18 base pairs with G-83, for example, if a base triple occurs. G-12 and C-13 had contacts with two more distant nucleotides, C-27 and G-28. Standard antiparallel Watson-Crick base pairing between both

G-12 and C-27 and C-13 and G-28 is not possible; the bonds involved in these contacts cannot be determined from these data. The analogous region of the genomic ribozyme is shorter than that of the antigenomic ribozyme (see Fig. 2); however, a similar contact may be possible for the genomic ribozyme because it contains U-27, a nucleotide not present in the analogous region of the antigenomic ribozyme.

(ii) Nucleotide A-14 is more highly conserved than its neighbors; however, no new interactions involving this nucleotide were suggested by our data. This position was not over-represented in clones with complex changes. It may be on the outside of the molecule, possibly at a bend in the nucleotide chain.

(iii) Nucleotides C-16, U-17, C-18, C-19 were mutated in between 12 and 14 clones and had no contacts we could identify, except for the interaction between C-18 and G-11 mentioned above. In the pseudoknot model, C-16, U-17, C-18 and C-19 provide one strand of helix II (see Fig. 2B). The other strand of helix II (G-81, G-82, A-83, G-84) was not mutated in this study and thus, we have no test of this helix. Furthermore, our cleavage reactions were not carried out in denaturants (see Materials and Methods). In 1991, Perotta and Been (16) showed that helix II is not required for efficient self-cleavage in the absence of 40%

Figure 3. Mutant antigenomic ribozyme variants. Clones were sorted by nucleotide position and are presented in sets, beginning with the smallest set, which has a single member, clone 167 (the only clone with a mutation at U-22). Mutations at positions 11–32 and 40–42 are presented. Blank spaces indicate wild-type nucleotides. Because of their large size and the need to conserve space, the sets of clones for positions 11, 12, 13 and 15 are not presented in their entirety in the Figure, but can be reconstructed. G-11 is mutated in the following clones, which appear in the Figure, Q, R, T, V, 2, 19, 21, 30, 34, 40, 43, 46, 61, 74, 76, 89, 99, 105, 154, 156, 165, 175, 181, 184, 189, 199, 203, 204, 208, 217, 226, 229, 238, 246, 247, 305, 341, 347, 359, 518, 519, 534, 550, 551, 563 and in the following additional clones, which contain a single mutation, 12(A-11), 28(C-11), 38(C-11), 42(U-11), 164(U-11), 179(C-11), 219(C-11) 240(A-11), 260(U-11), 505(U-11), 530(U-11), 532(C-11), 542(U-11) and clone 566, which contains two mutations, (U-11, G-13). G-12 is mutated in the following clones which appear in the Figure, B, H, V, 6, 21, 39, 51, 67, 75, 88, 100, 113, 139, 165, 181, 183, 184, 199, 205, 208, 220, 236, 246, 342, 348, 352, 502, 518, 519, 550, 551, 552, 553, 554, 564 and in the following additional clones which contain a single mutation, W(U-12), 308(C-12), 326(C-12). C-13 is mutated in the following clones which appear in the Figure, F, H, 2, 4, 24, 32, 35, 101, 154, 168, 184, 201, 213, 249, 531, 533, 544 and in the following additional clones which contain a single mutation, 81(A-13), 198(U-13), 206(U-13), 216(A-13), 525(U-13) and clone 566, which contains two mutations, (U-11, G-13). U-15 is mutated in the following clones which appear in the Figure, H, S, T, 5, 10, 40, 43, 82, 92, 114, 125, 134, 163, 169, 175, 184, 201, 204, 231, 248, 253, 342, 349, 352, 356, 357, 502, 564 and in the following additional clones which contain a single mutation, 29(A-15), 118(C-15), 121(G-15), 130(C-15), 556(G-15). All clones, except two, have the wild-type nucleotide in positions 33–39 and 43–45. In clone 128, C-37 is replaced by U-37 and A-44 is replaced by G-44. In clone 518, G-35 is replaced by A-35.

Figure 5. Position of the eight contact points identified through mutational analysis (a–h, see Table 2) of the delta antigenomic ribozyme. The axlehead model (A) and the pseudoknot model (B) of the delta antigenomic ribozyme have been re-drawn to accommodate the eight additional contacts suggested by these mutant studies. The 25 mutated positions are shaded. Newly proposed contacts are indicated by dashed lines, Watson–Crick base pairs are shown in open lines, arrows identify the orientation of C-25 and C-76 with respect to the cleavable phosphate at the 5' side of G-1.

formamide or 5 M urea, but only in their presence and demonstrated that in the absence of denaturants, nucleotides G-80 to G-84 can be deleted. Later studies showed that disruption of the pseudoknot in mutant SII3'(G81c,A82u) reduced the rate constant 15-fold, even in the absence of denaturant (23). While this reduction is significant, it is of much lower magnitude than those associated with other mutations examined in the same study, such as the mutation of C-21 to G-21, which reduced the cleavage rate constant ~5000-fold (23). [For further comparison, mutation of C-75 in the genomic ribozyme reduced the cleavage rate >500 000-fold (34)]. Particularly in the presence of denaturing agents, helix II may be important because it inhibits the formation of inactive structures.

(iv) Among the helix III nucleotides (see Fig. 2B) and the connecting loop, (a total of 13 nucleotides, bases A-20 to U-32), the most highly conserved nucleotides (C-21, C-22, U-23, C-24, C-25, G-30 and G-31) are interrupted by four much more frequently mutated nucleotides (U-26, C-27, G-28 and C-29). Unlike nucleotides G-28 and C-29, nucleotides U-26 and C-27 are over-represented in complex clones (see Fig. 4), suggesting that they may be in the interior of the molecule where they must conform to space constraints. No associations were found between these four nucleotides and the other nine nucleotides in this group. In contrast, as indicated in Table 2, many interactions were found among the other nine nucleotides.

While we did not find any evidence for helix III, clone 356 contained a potential compensatory base change involving C-21 and G-30, suggesting that these two nucleotides and adjacent nucleotides are close to each other in the active ribozyme, as

indicated in the pseudoknot model. However, these data argue for a slightly different orientation of the bases in this highly conserved portion of the molecule. Several previously reported observations suggest that helix III may not be a simple Watson–Crick base paired helix. First, for example, neither Rosenstein and Been (17), Wu *et al.* (19), nor Kumar *et al.* (24) found ribonuclease V1 cuts in helix III. Secondly, replacing the genomic helix III (GCC/GGC) with CGU/ACG led to complete inactivation (28), although base pairing was maintained. Thirdly, disruption of genomic helix III by the introduction of an A·C mismatch did not reduce the activity relative to the wild-type sequence as dramatically as expected, leading Thill *et al.* to comment: 'Evidently the absolute stability of stem III is not important for reactivity, or there are other structural features within the RNA, such as, for example, helical stacking, that provide compensating stabilization (25)'. It is interesting to note that while helices II and III appear to be weaker in the antigenomic ribozyme than in the genomic ribozyme ACC/GGU versus GCC/CCG and CUCC/GGAG versus GUCCCA/UGG-GAC, the antigenomic ribozyme is highly resistant to denaturants, perhaps indicating that this ribozyme has a number of unrecognized tertiary interactions which confer stability. The importance of these interactions may underlie the very high degree of sequence conservation of nucleotides C-21, C-22, U-23, C-24, C-25, G-30 and G-31. Our information about permissible base substitutions in these positions is limited. However, as noted in Table 1, all four clones with a mutation at C-25 contained A-25, suggesting a possible requirement for the extra cyclic amino group (at position C6 in adenine and C4 in

cytosine). Studies of the genomic ribozyme indicate that there is a requirement for cytosine in this region of the molecule. Deletion of the genomic analog of C-25, C-22, reduced the relative activity of the ribozyme 520 000-fold (25). Mutation of C-21 led to reduced activity (30). Our data indicate that many interactions occur among nucleotides C-21 to G-31. It is not clear whether the antigenomic nucleotides U-26 to C-29 are in a helical configuration as suggested for the analogous genomic nucleotides by Tanner *et al.* (34); our data probably favor a more complex structure, for example, one that brings U-23 close to G-31 and U-32 (see Table 2). Our data suggest that antigenomic nucleotides U-23 to C-29 participate in many interactions and may form a more compact structure than the open-loop depicted previously (see Fig. 2). (v) The three G nucleotides, G-40, G-41 and G-42, which make up the junction going down to helix IV (see Fig. 2A and B), were mutated in six, seven and eight clones respectively. Our data suggest that G-42 is close to C-27.

Cleavage studies of clone 10 RNA and clone 553 RNA

RNAs with the sequences of clone 10 and clone 553 (see Fig. 3), which have mutations in both G-42 and C-27, were synthesized, purified by gel electrophoresis and tested for their ability to be cleaved. Reactions (in triplicate) were carried out for 10 min at 55°C in a solution containing 5 mM HEPES buffer, pH 7.5, 50 mM magnesium chloride. Products were fractionated by electrophoresis. The extent of cleavage was determined by the analysis of gel bands present in autoradiograms using a densitometer. Under these

conditions, ~30% of wild-type RNA cleaved, 5% of clone 553 RNA cleaved and 0.5% of clone 10 RNA cleaved. Thus, both clone 553 RNA and clone 10 RNA cleave to measurable extents. Clone 10 RNA contains U-42 in addition to four other mutations, including a potential compensatory base change involving position 27. Even minimal cleavage of clone 10 RNA is of interest considering the previous report by Wu and Huang that detectable cleavage was abolished by mutation of G-42 to U-42 in their construct; their assay conditions included a 2 h incubation at 50°C (21).

DISCUSSION

We mapped 359 mutations at 25 positions in synthetic variants of the antigenomic ribozyme of the hepatitis delta agent and used these data to identify highly conserved nucleotides and associations between nucleotides. From the eight pairs of potential interactions identified in this study, three proposals can be made about how the primary sequence of the antigenomic ribozyme folds into the three dimensional structure of the molecule. First, compensatory base changes between G-11 and C-18 suggest that the sequence folds back on itself in this region. Secondly, contacts between G-12 and C-27 and between C-27 and G-42 indicate that these three nucleotides are on the same side of the ribozyme. Thirdly, several of the highly conserved nucleotides, C-21, C-22, U-23, C-24, C-25, G-30 and G-31, interact with each other, probably forming a compact structure, rather than an open-loop. These proposals need to be tested through further investigation.

Table 2. Evidence for eight newly-proposed nucleotide-nucleotide contacts

Nucleotides	Compensatory base changes	Co-mutation value	Form of association
a. G-11 and C-18	Clone 2 (U-11 and A-18) Clone 19 (U-11 and A-18) Clone 89 (U-11 and G-18) Clone 217 (A-11 and U-18)	1.4	Base pair or proximity
b. G-12 and C-27	Clone 75 (C-12 and G-27) Clone 100 (U-12 and G-27) Clone 551 (C-12 and G-27) Clone 554 (U-12 and G-27)	1.6	Base pair or proximity
c. C-13 and G-28	Clone 154 (G-13 and U-28) Clone 184 (G-13 and C-28) Clone 544 (A-13 and U-28)	2.5	Base pair or proximity
d. C-21 and U-23	None possible	8.7	Proximity
C-21 and C-24	None possible	7.2	Proximity
e. C-21 and G-30	Clone 356 (U-21 and A-30)	14.5	Base pair or proximity
f. U-23 and G-31	None possible	8.4	Proximity
U-23 and U-32	None possible	4.2	Proximity
g. C-24 and G-30	Clone 98 (G-24 and C-30)	14.3	Base pair or proximity
h. C-27 and G-42	Clone 10 (A-27 and U-42) Clone 100 (G-27 and U-42) Clone 553 (U-27 and A-42)	3.4	Base pair or proximity

Watson-Crick compensatory base changes were identified and assigned the value of 2; G-U substitutions were assigned the value of 1. The 'co-mutation value' for each pair of nucleotides was calculated as described in the text and summed with the compensatory base change value. Pairs in which this sum exceeded the arbitrary cut-off value of 7 were evaluated further and excluded from Table 2 if they shared only a single pair of non-compensatory mutations, unless a neighboring, rarely mutated position also contained a mutation (see C-21 and U-23/C-24 and U-23 and G-31/U-32); if either member were so restricted that it was found as only one of the three possible alternative nucleotides; or if the association were based on a single pair of compensatory mutations occurring in a clone in which a different pair of compensatory mutations appeared to be more significant (see clone 356).

Gottlieb and colleagues recently presented a number of novel and interesting conformations resulting from 'structural compensation' (35). However, since these conformations involve sequence re-arrangements not present in the molecules we examined, discussion of our newly proposed interactions will focus on their impact on the pseudoknot model (16) and the axehead model (14). As illustrated in Figure 2, these two models share two Watson-Crick base paired helices (identified as helices I and IV in Fig. 2A and B). They differ in the interactions involving nucleotides at the 3' end of the ribozyme. In the pseudoknot model, these nucleotides comprise part of the pseudoknot (helix II) and in the axehead model, they bond with the sequence upstream from the cleavage site (see Fig. 2A). The pseudoknot model contains a three base-pair helix (helix III) composed of nucleotides drawn as part of a large open-loop structure in the axehead model. In the pseudoknot model of Perrotta and Been, three nucleotide chains separate C-76 from the cleavage site (see Fig. 2B); while in the axehead model of Branch and Robertson (14), C-76, a nucleotide essential for ribozyme function (22,34), lies opposite the cleavage site (see Fig. 2A).

The contacts we identified in this study can be incorporated into either the axehead model or the pseudoknot model, as illustrated by Figure 5A and B, respectively. Tanner and colleagues (34) developed a three-dimensional version of the pseudoknot model of the genomic delta ribozyme in which (genomic) nucleotides U-20, C-21 and C-75 are all placed near the cleavable phosphate, in accordance with studies demonstrating the importance of these bases (22,25,30,34). Our data are largely compatible with their model, the biggest difference being the absence in their model of a contact analogous to that between the loop nucleotide C-27 and the junction nucleotide G-42.

A future challenge for all models will be to identify the contacts made by nucleotide A-79 and, in particular, to explain the observation that while it is possible to delete all bases to the 3' side of A-79 (16), eliminating both helix II of the pseudoknot model and the interaction between the 5' upstream sequences and the 3' end of the ribozyme depicted in the axehead model, deletion of the next nucleotide, A-79, abolishes activity. In the primary sequence, A-79 is only a few bases away from C-76, a base which appears to play a crucial role in cleavage. Contacts between A-79 and other parts of the molecule may help to position C-76. Suh and colleagues (20) demonstrated that a genomic ribozyme 133 nucleotides in length retained activity when bases C-13 to C-30 were deleted, suggesting that A-78 (the genomic analog of A-79) makes at least some important contacts with nucleotides in helix I, helix IV or the G-rich junction element. Of these possibilities, an interaction with helix I appears to be most likely and has been built into our three-dimensional versions of the axehead model (see Fig. 5A) and the pseudoknot model (see Fig. 5B). Suh and colleagues (20) also found that a shortened ribozyme, lacking sequences from both the 5' and 3' ends of the 133-long molecule, did not retain activity when bases C-13 to C-30 were deleted, suggesting that either a very stable interaction between the 5' and 3' ends (of ribozymes such as the 133 base-long RNA) or an interaction involving nucleotides between C-13 and C-30 is sufficient to preserve activity. A variety of interactions may enhance the probability that (antigenomic) bases C-76 to A-79 assume the configuration required for cleavage. Identification of these interactions may provide further information about the overall-folding of the delta ribozyme.

ACKNOWLEDGEMENTS

We wish to thank Sylvia Genus, Susan Lee, Toby Lefkowitz, Olivia Neel and David Schreiber for outstanding technical support. This work was funded in part by the Liver Transplant Research Fund, Department of Surgery, The Mount Sinai School of Medicine and by NIH grants U01AI-31867 and R01AI-31067 to Dr Hugh Robertson (Cornell University Medical College) and NIDA Research Center Grant P50DA-05130 directed by Dr Mary Jeanne Kreek (Rockefeller University), project 7a to ADB.

REFERENCES

- Rizzetto, M., Canese, M.G., Arico, S., Crivelli, O., Trepo, C., Bonino, F. and Verme, G. (1977) *Gut*, **18**, 997-1003.
- Gerin, J.L., Purcell, R.H. and Rizzetto, M. (1989) *The Hepatitis Delta Virus*, Wiley-Liss, New York.
- Branch, A.D., Levine, B.J. and Robertson, H.D. (1990) *Semin. Virol.*, **1**, 143-152.
- Sharmeen, L., Kuo, M.Y.-P., Dinter-Gottlieb, G. and Taylor, J. (1988) *J. Virology*, **62**, 2674-2679.
- Kuo, M.Y.-P., Sharmeen, L., Dinter-Gottlieb, G. and Taylor, J. (1988) *J. Virology*, **62**, 4439-4444.
- Macnaughton, T.B., Wang, Y.-J. and Lai, M.M.C. (1993) *J. Virology*, **67**, 2228-2234.
- Branch, A.D. and Robertson, H.D. (1984) *Science*, **223**, 450-455.
- Chen, P.-J., Kalpana, G., Goldberg, J., Mason, W., Werner, B., Gerin, J. and Taylor, J. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8774-8778.
- 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.
- 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.
- Perrotta, A.T. and Been, M.D. (1990) *Nucleic Acids Res.*, **18**, 6821-6827.
- Been, M.D. (1994) *Trends Biosci.*, **19**, 251-256.
- Branch, A.D. and Robertson, H.D. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10 163-10 167.
- Perrotta, A.T. and Been, M.D. (1992) *Biochemistry*, **31**, 16-21.
- Perrotta, A.T. and Been, M.D. (1991) *Nature*, **350**, 434-436.
- Rosenstein, S.P. and Been, M.D. (1991) *Nucleic Acids Res.*, **19**, 5409-5416.
- Thill, G., Blumenfeld, M., Lescure, F. and Vasseur, M. (1991) *Nucleic Acids Res.*, **19**, 6519-6525.
- Wu, H.-N., Wang, Y.-J., Hung, C.-F., Lee, H.-J. and Lai, M.M.C. (1992) *J. Mol. Biol.*, **223**, 233-245.
- Suh, Y.-A., Kumar, P.K.R., Nishikawa, F., Kayano, E., Nakai, S., Odai, O., Uesugi, S., Taira, K. and Nishikawa, S. (1992) *Nucleic Acids Res.*, **20**, 747-753.
- Wu, H.-N. and Huang, Z.-S. (1992) *Nucleic Acids Res.*, **20**, 5937-5941.
- 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.
- Been, M.D., Perrotta, A.T. and Rosenstein, S.P. (1992) *Biochemistry*, **31**, 11 843-11 852.
- Kumar, P.K.R., Suh, Y.-A., Taira, K. and Nishikawa, S. (1993) *FASEB J.*, **7**, 124-129.
- Thill, G., Vasseur, M. and Tanner, N.K. (1993) *Biochemistry*, **32**, 4254-4262.
- Suh, Y.-A., Kumar, P.K.R., Kawakami, J., Nishikawa, F., Taira, K. and Nishikawa, S. (1993) *FEBS Lett.*, **326**, 158-162.
- Perrotta, A.T. and Been, M.D. (1993) *Nucleic Acids Res.*, **21**, 3959-3965.
- Wu, H.-N., Lee, J.-Y., Huang, H.-W., Huang, Y.-S. and Hsueh, T.-G. (1993) *Nucleic Acids Res.*, **21**, 4193-4199.
- Belinsky, M.G., Britton, E. and Dinter-Gottlieb, G. (1993) *FASEB J.*, **7**, 130-136.
- 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.
- Kumar, P.K.R., Taira, K. and Nishikawa, S. (1994) *Biochemistry*, **33**, 583-592.

- 32 Nishikawa,S., Kumar,P.K.R., Jeoung,Y.H., Kawakami,J., Nishikawa,F., Suh,Y.-A., Ohtsuka,E. and Taira,K. (1993) *Nucleic Acids Symposium Series*, **29**, 119–120.
- 33 Lee,B.-S., Wu,H.-N. and Huang,T.-H. (1993) *FEBS Lett.*, **324**, 296–300.
- 34 Tanner,N.K., Schaff,S., Thill,G., Petit-Koskas,E., Crain-Denoyelle,A.-M. and Westhof,E. (1994) *Current Biol.*, **4**, 488–498.
- 35 Gottlieb,P.A., Prasad,Y., Smith,J.B., Williams,A.P. and Dinter-Gottlieb,G. (1994) *Biochemistry*, **33**, 2802–2808.
- 36 Smith,J.B. and Dinter-Gottlieb,G. (1991) *Nucleic Acids Res.*, **19**, 1285–1289.
- 37 Uhlenbeck,O.C. (1987) *Nature*, **328**, 596–600.
- 38 Haseloff,J. and Gerlach,W.L. (1988) *Nature*, **334**, 585–591.
- 39 Milligan,J.F., Groebe,D.R., Witherell,G.W. and Uhlenbeck,O.C. (1987) *Nucleic Acids Res.*, **15**, 8783–8798.
- 40 Yie,Y., Wei,Z. and Tien,P. (1993) *Nucleic Acids Res.*, **21**, 361.
- 41 Casey,J.L., Brown,T.L., Colan,E.J., Wignall,F.S. and Gerin,J.L. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 9016–9020.
- 42 Branch,A.D., Lee,S., Neel,O.D. and Robertson,H.D. (1993) *Nucleic Acids Res.*, **21**, 3529–3535.
- 43 Branch,A.D. Benenfeld, J.B., Baroudy, B.M., Wells, F.V., Gerin, J.L. and Robertson, H.D. (1989) *Science*, **243**, 649–652.
- 44 Branch,A.D., Levine,B.J. and Polaskova,J.A. (1995) *Nucleic Acids Res.*, **23**, 491–499.
- 45 Branch,A.D., Levine,B.J., Baroudy,B.M., Buckler-White,A., Gerin,J.L. and Robertson,H.D. (1991) In Gerin,J.L., Purcell,R.H. and Rizzetto,M. (ed.), *The Hepatitis Delta Virus*. Wiley-Liss, New York, pp. 257–264.
- 46 Branch,A.D., Polaskova,J.A. and Schreiber,D.R. (1995) *Nucleic Acids Res.*, in press.
- 47 Rich, A. and RajBhandary, U.L. (1976) *Annu. Rev. Biochem.*, **45**, 805–860.