1096–1100 Nucleic Acids Research, 1994, Vol. 22, No. 6

Mutagenesis of the hairpin ribozyme

Pamela Anderson, Joseph Monforte[1], Richard Tritz, Steven Nesbitt, John Hearst[2] and Arnold Hampel*

Departments of Biological Sciences and Chemistry, Northern Illinois University, DeKalb, IL 60115, [1]SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025 and [2]Department of Chemistry, University of California, Berkeley, CA 94720, USA

Received September 28, 1993; Revised and Accepted January 22, 1994

ABSTRACT

Extensive *in vitro* mutagenesis studies have been performed on the hairpin ribozyme and substrate in an effort to refine the overall secondary structure of the molecule and provide further insight into what elements are essential for activity. A secondary structure consisting of four helices and five loop regions remains the basic model as originally proposed. Two helices, helix 1 and 2, form between the substrate and ribozyme while helices 3 and 4 are within the ribozyme itself. Our results suggest that helices 3 and 4 are smaller than previously proposed, consisting of four base pairs and three base pairs respectively. Helix 4 can be extended without loss of activity and loop 3 at the closed end of the hairpin model can be varied in sequence with retention of activity. There is an unpaired nucleotide between helices 2 and 3 consisting of a single A base, suggesting the opportunity for flexibility within the tertiary structure at this point. Comparisons are made between the new data and previously published mutagenesis and phylogenetic data. Substrate targeting rules require base pairing between helices 1 and 2 with cleavage (*) occurring in a preferred 5′(g/c/u)n*guc3′ sequence of the substrate.

INTRODUCTION

The hairpin ribozyme and its substrate are the minimal catalytic and substrate sequences of the catalytic center of the negative strand of the satellite RNA of tobacco ringspot virus (1). The original native ribozyme is 50nt long and it cleaves a corresponding 14nt substrate. This reaction has been characterized biochemically (2) and shown to occur in trans (2,3). It is truly catalytic in that the ribozyme turns over, and the reaction proceeds very efficiently with a K_m of 30nM and a k_{cat} of 2.1/min under mild conditions (pH 7.5, 37°C in the presence of Mg^{++}) (2).

Results of previous limited mutagenesis, secondary structure computer modeling and phylogenetic studies produced an early picture of the secondary structure of the hairpin ribozyme/substrate complex (4,5). It consisted of four helical regions separated by single stranded or internal loop sequences. A key element of the structure was the location of the cleavage site within the internal loop sequence of the substrate, 5a*guc8 (Fig 1). Cleavage occurred between a5 and g6 of this four nucleotide internal loop sequence. Mutants assayed for cleavage activity showed that a5 preceeding the cleavage site was completely variable while the guc sequence was preferred. The bases flanking this n*guc sequence were shown to be involved in base pairs and binding of the substrate to the ribozyme. These two flanking helices were named 1 and 2 with helix 1 beginning at the 5′ end of the ribozyme. It was further shown that a number of heterologous substrate sequences could be cleaved by an engineered ribozyme which base paired to them (4).

Previous mutagenesis showed the existence of base pairs in two additional proposed helices of the ribozyme itself and were named helices 3 and 4 (4). The extent of these helices, and the consequent size of the connecting proposed single stranded loop regions, were unknown at that time. Here we present mutagenesis data providing further refinement of the secondary structure and some insight into which nucleotides might have important roles in the overall function of the ribozyme. These results, based on cleavage of the substrate itself, complement those using in vitro selection of active hairpin ribozymes by sequential RNA-catalyzed cleavage and ligation reactions and direct mutagenesis (6–10).

MATERIALS AND METHODS

RNA was prepared by transcription using either [α-^{32}P]CTP or [α-^{32}P]UTP from synthetic DNA templates containing the T7 promoter and purified on 15% polyacrylamide/7M urea gels (2,11). The native hairpin ribozyme and substrate (Fig. 1) were mutated according to that numbering scheme. Cleavage reactions were carried out as before (2) by mixing ribozyme and substrate at 37°C in 12mM $MgCl_2$, 2mM spermidine, 40mM Tris, pH 7.5, for time points ranging from 0.5–240min; a zero time point was included. Concentrations of substrate and ribozyme used varied over a range of concentrations from 10–500nM for the entire study. Reaction products were analyzed on denaturing acrylamide gels, identified by autoradiography, cut out and quantitated by scintillation counting. Cleavage rate was the ratio (radioactivity of the products)/[(radioactivity of the products) +

*To whom correspondence should be addressed

(remaining substrate)] for a given time point. For each mutation experiment, the cleavage rate was determined simultaneously for the unmutated native ribozyme/substrate. Percent activity was the cleavage rate of the mutant compared with that of the unmutated native ribozyme/substrate with the parameters of concentration and time normalized. The ratio (the cleavage rate of the mutant)/(cleavage rate of the unmutated native ribozyme and substrate) was multiplied by 100 to determine percent activity. Side-by-side comparisons were made between relevant mutations (i.e., if the effect of an alternative base pair was being determined, the mismatched sequences were assayed simultaneously). The 5′ termini of ribozymes and substrates were GG, GGG or GCG. For each experiment the mutant and unmutated ribozymes and substrates were at equivalent concentrations. (*) Mutants (Table 1) had an additional 3′ terminal sequence of CGCCUUG-ACUUCUAG for the ribozyme and cguccucacggacucac for the substrate. All mutant sequences were compared with unmutated sequences having equivalent 5′ and 3′ terminal sequences.

For the sake of clarity, lower case letters were used for identification of substrate bases and upper case letters for ribozyme bases.

RESULTS AND DISCUSSION

Mutations in the native sequence of the hairpin ribozyme and substrate (Fig. 1) gave a range of effects on catalytic activity as measured by the ability of the ribozyme to cleave the substrate (Table 1). The percent activity values are the cleavage rates for the mutation relative to those for the native sequence. The influence nucleotide changes had on catalytic activity ranged from none (100% of wild type catalytic activity) to no detectable catalytic activity. None of these mutations gave activities significantly greater than that of the native form, which may not be surprising because the hairpin ribozyme has very high inherent catalytic efficiency when compared with other catalytic RNAs (2). Secondary structural features of the ribozyme can be deduced by mutation with potential base pairs identified by compensatory mutational analysis. A base pair was deduced to occur in the active structure if mismatches had reduced or no activity and alternative base pairs had restored activity.

Previous mutagenesis in our laboratory showed a four base 5aguc8 substrate loop (loop 5 of the ribozyme/substrate complex). It retained full cleavage activity if a5 was any of the four RNA bases but lost activity if the 6guc8 sequence was mutated as follows: g6→a; u7→a; and c8→g (4). This finding was further confirmed by additional mutants in the 6guc8 region, which showed either no activity or significantly reduced activity (Table 1), with one notable exception, c8→u gave activity 25% of the native substrate. These results are consistent with, but represent an expansion of those obtained by others using the cleavage ligation reaction (6, 10).

Earlier studies (4) determined that helices 1 and 2 are interactions between the ribozyme and the substrate and that heterologous RNA could be cleaved as long as base pairing was maintained between these two helices. The existence of helix 2 was further substantiated by using mismatch and alternate base pair mutagenesis of the two interior base pairs of a different form of the hairpin ribozyme (12, 13). Helix 1 could be extended or shortened to obtain an optimal cleavage rate depending on its sequence (4). When heterologous RNA was cleaved by engineered ribozymes all six positions in helix 1 and the first three positions in helix 2 (positions 1, 2 and 3 of the substrate) have been changed to all four bases, either singly or in combination, with resulting cleavage. This was not true for position 4 of the substrate. An a4 resulted in no cleavage by the corresponding G11→U ribozyme, but substrates with c4, g4 or u4 were cleaved by the corresponding base paired ribozyme. In the case of u4 the corresponding ribozyme could have either an A11 or G11 and cleavage occurred. Therefore, the helix 2 base pair G11:c4 can be b:V (14), where b indicates c, u, or g, and V indicates G, A or C. This result is consistent with that seen by others (8). These results define 4bn*guc8 as preferred sequence for the substrate. This information can be used for targeting experiments. Targeting experiments were carried out against HIV-1 using the hairpin ribozyme. The ribozyme to the HIV-1 substrate had excellent kinetic parameters in vitro, and was successful in lowering expression of HIV-1 in vivo (15, 16)

No base pairs between the substrate loop and the opposite 7A-GAA10 loop 1 of the ribozyme could be shown. We also found the two exterior A nucleotides of AGAA loop 1 could be changed with little effect on activity while the two interior GA nucleotides (G8 and A9) were required for activity. When G8 was changed to either U or C, activity was lost. The mutation A9→U was also inactive.

Phylogenetic comparisons of the negative strand of the satellite of arabis mosaic virus, a sequence predicted to be catalytic (5), with that of the negative strand of the satellite RNA of tobacco ringspot virus sequence (17), from which the hairpin ribozyme was derived, revealed an interesting three base difference (c8→a, A7→C and A20→C). When each of these three mutations was

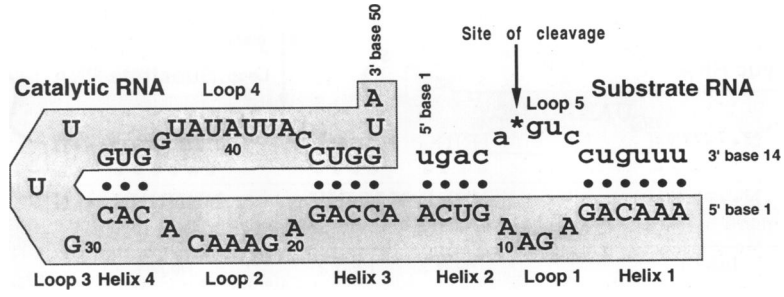

Figure 1. Two-dimensional model of the native sequence of hairpin ribozyme. The model contains four helices and five loops with the site of cleavage indicated. Substrate sequence is in lower case letters and ribozyme in upper-case letters.

made individually, in the native hairpin ribozyme the A7→C and A20→C mutations each gave nearly full activity, while the c8→a activity was very low. However, double mutations of c8→a/A7→C and c8→a/A20→C and the triple mutation of all three changed at once, significantly increased the low c8→a activity. While the reason for this increase is not known, it suggests some type of interaction might occur between c8 and A7 which are positioned directly opposite each other in the loops or that complementation needs to be retained in order to stabilize the internal loops. The reason for A20→C increasing activity was less clear from a simple two-dimensional model, and it may indicate a possible tertiary interaction.

In helix 3 the base pair C17:G47 was previously shown to likely occur by mismatch and alternate base pair mutagenesis (4). In addition we now show G19:C45 is also a likely base pair. However, the presumptive base pair C16:G48 at the beginning of the helix could not be similarly shown because both the C:C mismatch (G48→C) and the alternate base pair G16:C48 were

Table 1. Effect on catalytic activity of mutations made in the hairpin ribozyme and substrate

Mutation	Percent Activity	Mutation	Percent Activity
No mutation (wild type)	100	A20→C	81
substrate		22-AAA-24→CGU	NC
g6→a	NC	A24→G/U37→C	NC
g6→c	NC	C25→G/G36→C	NC
g6→u	NC	*del A26	NC
u7→a	7	C27→G	NC
u7→c	1	C27→G/G35→C	10
u7→g	2	A28→U/U34→A	113
c8→a	2	C29→G/G33→C	12
c8→g	1	C29→G/G35→C	NC
c8→u	25	*29-CGU-31→GGUC (GUUA) GACC	NC
substrate:ribozyme		30-GUU-32→UUCG	NC
c4→g/G11→C	12	30-GUU-32→GGAC (UUCG) GUCC	116
c4→a/G11→U	NC	*30-GUU-32→GGUC (GUUA) GACC	100
c4→u/G11→A	40	*30-GUU-32→GGUC (GUUA) GACC/del U31 cut U32	NC
c4→u/G11→G	32	G33→C	10
g6→c/U39→G	NC	U34→A	11
u7→g/G8→U	NC	G35→C	NC
u7→a/A9→U	NC	U37→A/A43→U	NC
c8→g/G8→C	NC	*A38→G	2
c8→a/A7→C	14	*A38→U	NC
c8→a/A20→C	7	U39→C	100
c8→a/A20→C:A7→C	9	U39→G	100
ribozyme		*A40→U	3
A7→G	100	*A40→G	3
A7→C	100	*U41→C	25
A7→C/A20→C	104	*U42→C	3
G8→C	NC	*del U42	NC
G8→U	5	A43→U	NC
A9→U	NC	G47→C	NC
A10→G	61	G48→C	NC
A15→U/U49→A	115	del U49:del A50	38
C16→G/G48→C	NC	U49→A	80
C17→G	NC	U49→C	60
C17→G/G47→C	21	del A50	49
G19→C	1	A50→G	75
G19→C/C45→G	25		
*G21→C	58		

All values are rates of cleavage relative to wild type (non-mutated) expressed percent activity of wild type. NC is no cleavage. Numbering of bases is according to Fig. 1.

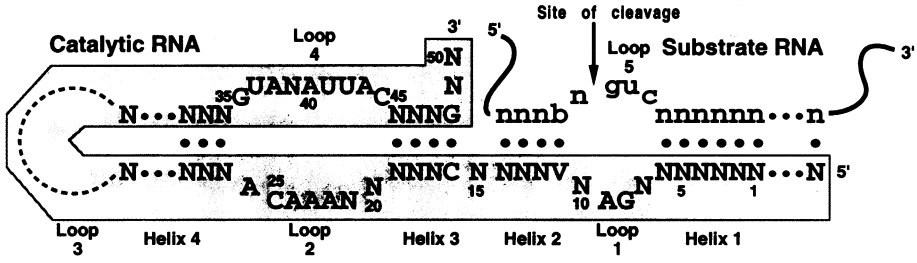

Figure. 2. A two-dimensional model summarizing the mutagenesis obtained in this study for hairpin ribozyme. Four helices and five loops with noninvariant bases indicated by N are shown. The b base is c, u, or g and the V base is G, A or C (13). Numbering is the same as in Fig. 1.

inactive. Because this G and C are opposite each other in this position by virtue of an experimentally supported adjacent base pair, they are probably base paired as well. It should be noted that anomolies can occur because of the potential for entirely novel alternate structures to form. Perhaps the identity of the bases must be maintained in this position, possibly for specific tertiary interactions that are not identified in this study. The A18:U46 base pair likely exists since it is located between two experimentally supported G:C base pairs. This data and that discussed below support a four base pair helix 3, which is an advancement over the previous model, which proposed a five base pair helix 3 (4, 9).

The existence of helix 4 is supported by the observed loss of activity of the G:G and C:C mismatches for C27:G35 and restoration of activity to 10% of the control level for compensatory mutation G27:C35 (4, Table 1). The adjacent potential base pair A28:U34 also probably exists because the alternate base pair U28:A34 was 100% active and had 10 times the activity of the mismatch (Table 1). Significantly, the A:A mismatch did show activity at a level of 11% of the native control, indicating that either the A:A interaction is partly stable or this helix need not be fully formed for activity to occur.

Helix 4 can be extended, and loop 3 does not have an invariant sequence. We replaced loop 3, 30GUU32, with the common and very stable RNA tetraloop stem loop sequence GG-AC(UUCG)GUCC (18). This change resulted in a 4bp extension of helix 4 and replacement of the GUU sequence of loop 3 with a UUCG sequence. Catalytic activity was fully retained, which was also the case when a different tetraloop was used (Table 1), further confirming the existence and nature of helix 4 as shown (Fig. 1). The tetraloop was also tested in combination with the A26 deletion and no activity was seen, as was the case for the native ribozyme with this same deletion. In both cases the deletion would have promoted formation of two additional base pairs in helix 4, C25:G36 and A24:U37. The sequence 29CGU31 was replaced with the stable helix/loop sequence GGUC(GUUA)G-ACC to give a ribozyme with a helix 4 extension containing a potential alternative base-pairing scheme. No activity was detected, indicating that no such alternative structure promotes cleavage and that helix 4 is as shown in Fig. 1.

In the molecule where helix 4 is extended and loop 3 replaced by the tetraloop stem loop sequence above, GGAC(UUCG)-GUCC, the potentil base pair C29:G33 is located between two experimentally supported base pairs of helix 4 and the 4bp tetraloop helix, making the C29:G33 base pair likely to exist in the interior of such a 7bp helix. In the native sequence it would constitute the third base pair in helix 4, although both the mismatch G33→C and alternate base pair C29:G33→G29:C33 for this site were partly active and had equivalent levels of activity. This data therefore supports helix 4 as a three base pair helix, a result supported by cleavage ligation selection experiments (9).

Loops 2 and 4 are much larger than originally predicted by computer modeling and energy minimization. In addition, many of these bases are required for activity (Table 1). In loop 2 the three A bases 22AAA24, when mutated to 22CGU24, gave an inactive ribozyme, indicating that one or all of these bases are essential. The double mutation of C25→G/G36→C was inactive, providing no support for a base pair between these two nucleotides. When A26 was deleted, activity was lost, showing that when this base is removed, the resulting ribozyme is inactive. Among the active mutations in loop 2 was G21→C which had a cleavage rate about 58% of wild type. This mutation was also significant because it converted the AGAA sequence in loop 2 to ACAA. Loop 1 also has an AGAA sequence, but when it was mutated to ACAA (G8→C mutation) activity was totally lost, indicating that the two sequences are likely functionally different.

Other active mutations in loop 2 were A20→C and the double mutation A20→C/A7→C. These mutations are significant because the potentially self-cleaving RNA from the negative strand of the satellite of arabis mosaic virus (19) has both these same base alterations, and the negative strand of the smaller satellite of chicory yellow mottle virus (sCYMV-S1) (20) has a G20 and C7 sequence. Thus, both the phylogenetic and mutational data support variability at this position as does cleavage ligation selection (9)

Loop 4 likely contains nine nucleotides, G36-C44, some of which are essential while others can be readily varied. From Table 1, mutation of G35→C, A38→U, A43→U gave no cleavage, while detectable, but greatly reduced, activity was observed for A38→G, A 40→U,G, U41→C, and U42→C indicating that these bases are either essential or very important. Full catalytic activity was retained for the U39→C,G mutation. Previously published cleavage/ligation experiments identified U39→C,A mutations as permissible (8). These previously published results combined with the present results show variability at this position.

An unpaired nucleotide between helices 2 and 3 consisting of a single 'A' base at position 15 was shown. The A15/U49 potential base pair in this region may exist but is not necessary, since neither the A:A nor the A:C mismatch significantly reduced activity. Such an unpaired nucleotide could be significant in the ultimate three-dimensional folding of the ribozyme, because it more readily allows the molecule to fold back on itself. Psoralen cross-linking experiments performed by Monforte (21) suggest that folding about such an unpaired nucleotide is an essential feature in the active ribozyme/substrate structure.

The results from this study are summarized in model form (Fig. 2). Shown are the four helices and five loops found in the ribozyme/substrate complex, along with mutable bases indicated by an N. Not all N bases shown have been changed to each of the four nucleotides in all cases; thus, the symbol is used to indicate that these bases can be altered. There may be combinations of base changes, as well as single base changes, within these variant positions that either are inactive or have much reduced activity. Furthermore, this study has not attempted to make every possible base change within the hairpin ribozyme, and therefore, many bases which are not shown as mutable, may indeed have some variation. Certain features of this model are supported by a variety of other data such as phylogenetic (5, 20), cleavage ligation selection and direct mutagenesis (4, 9, 10).

ACKNOWLEDGMENTS

This work was supported in part by NIH and the Biotechnology Research and Development Corporation to A.H. Support was obtained by the office of Basic Energy Sciences, Department of Energy, under contract No. DE-AC030-76SF00098 to J.H.

REFERENCES

1. Buzayan, J. Gerlach, W. and Bruening, G. (1986) *Nature*, 323, 349–352.
2. Hampel, A. and Tritz, R. (1989) *Biochemistry*, 28, 4929–4933).
3. Feldstein, P., Buzayan, J. and Bruening, G. (1989) *Gene*, 82, 53–61.

4. Hampel, A., Tritz, R. Hicks, M. and Cruz, P. (1990) *Nuc. Acids Res.*, 18, 299–304.
5. Haseloff, J. and Gerlach, W. (1989) *Gene*, 82, 43–52.
6. Chowrira, B., Berzal-Herranz, A. and Burke, J. (1991) *Nature*, 354, 320–322.
7. Berzal-Herranz, Simpson, J, and Burke, J. (1992) *Genes and Development*, 6, 129–134.
8. Simpson, J. Berzal-Herranz, A., Chowrira, B, Butcher, S. and Burke, J. (1993) *Genes and Development*, 7, 130–138.
9. Berzal-Herranz, A., Simpson, J., Chowrira, B., Butcher, S. and Burke, J. (1993) *EMBO J.* 12, 2567–2574.
10. Sekiguchi, A. Komatsu, Y., Koizumi, M. and Ohtsuka, E. (1991) *Nuc. Acids. Res.* 19, 6833–6838.
11. Milligan, J. Groebe, D., Witherall, G. and Uhlenbeck, O. (1987) *Nuc. Acids Res.*, 15, 8783–8798.
12. Feldstein, P., Buzayan, J., Van Tol, H. DeBear, J., Gough. G. Gilham, P. and Bruening, G. (1990) *Proc. Natl. Acad. Sci. USA* 87, 2623–2627.
13. Van Tol, H., Buzayan, J. and Bruening, G. (1991) *Virology*, 180, 23–24.
14. IUB Nomenclature. *Eur. J. Biochem.*, 150, 1–5 (1985).
15. Ojwang, J, Hampel, A., Looney, D., Wong-Staal, F. and Rappaport, J. (1992) *Proc. Natl. Acad. Sci., USA* 89, 10802–10806.
16. Yu, M., Yamada, O. Ojwang, J. Hampel, A. Rappaport, J. Looney, D. and Wong-Staal, F. (1993) *Proc. Natl. Acad. Sci. USA* 90, 6340–6344.
17. Buzayan, J., W. Gerlach, W., Bruening, G., Keese P. and Gould, A. (1986) *Virology*, 151, 186–191.
18. Cheong, C., Varani, G. and Tinoco, I. (1990) *Nature*, 346, 680–682.
19. Kaper, J., Tousignant, M. and Steger, G. (1988) *Biochem. Biophys. Res. Com.*, 154, 318–325.
20. Rubino, L., Tousignant, M., Steger, G. and Kaper, J. (1990) *J. Gen. Virology*, 71, 1897–1903.
21. Joseph Monforte, Ph.D. Thesis, University of California-Berkeley (1991).