'Hairpin' catalytic RNA model: evidence for helices and sequence requirement for substrate RNA

Arnold Hampel*, Richard Tritz, Margaret Hicks and Phillip Cruz[1]
Department of Biological Sciences, Northern Illinois University, DeKalb, IL 60115 and [1]Department of Biochemistry and Kettering-Scott Magnetic Resonance Center, Kettering Medical Center, Wright State University, Dayton, OH 45435, USA

Received September 28, 1989; Revised and Accepted November 16, 1989

ABSTRACT

We have identified the catalytic domain within the sequence of the negative strand of the satellite RNA of tobacco ringspot virus. Minimum energy RNA folding calculations predict a two dimensional model with four major helical regions which are supported by mutagenesis experiments. This model for the catalytic complex consists of a 50 base catalytic RNA and a 14 base substrate RNA folded together in a type of hairpin two dimensional structure. Part of the recognition region between the catalyst and substrate is two helices of 6 bases and 4 bases respectively. Catalytic activity remains when the bases in these two helices are changed but base pairing is maintained. Thus an appropriately engineered 'hairpin' catalyst is capable of cleaving heterologous RNA.

INTRODUCTION

The catalytic center of the 359 base negative strand of the satellite RNA of tobacco ringspot virus [(−)sTRSV] (1) has been identified and reaction parameters determined (2). The catalytic center was shown to consist of two minimal sequences of satellite RNA: a catalytic RNA with 50 satellite bases and a substrate RNA with 14 satellite bases which cleaved to form the corresponding 5′ fragment and 3′ fragment. The reaction was truly catalytic since the 50 base RNA had multiple substrate cleavage events and was not consumed during the course of the reaction and a linear relationship was seen between reaction rate and catalytic RNA concentration (2).

The temperature optimum of the catalytic reaction is near 37°C and the reaction can be carried out under very mild conditions of salt, pH and temperature (2). At 37°C, pH 7.5 40mM Tris, 2mM spermidine and 12mM $MgCl_2$ the reaction has a $K_m = 0.03\mu M$ and $k_{cat} = 2.1/min$. This gives an enzyme efficiency value, k_{cat}/K_m, which is very favorable for a catalytic RNA reacting under such mild conditions of temperature, salt and pH (3,4). The reaction rate was over one half as fast even at 4mM $MgCl_2$ which gave overall conditions near physiological (2).

We propose a two-dimensional 'hairpin' catalytic RNA model for the catalytic center of (−)sTRSV. The model is predicted from minimum energy RNA folding calculations and certain features are supported by mutagenesis. This is a new catalytic RNA motif not previously seen and very different from the 'hammerhead' model proposed for other satellite and viroid RNAs (4,5); the well characterized Tetrahymena rRNA intervening sequence ribozyme of Cech (3); and the M1 catalytic RNA of ribonuclease P which catalyzes tRNA processing (6,7). It also appears to be different from the human hepatitis delta viral RNA (8) since no sequence homology exists in the region of these two catalytic centers (9).

The proposed hairpin model has four predicted helical regions which are supported by mutagenesis. Base pair mismatch mutants were catalytically inactive and mutants with restored but different base pairs were active. Two of these helices are between the substrate RNA and the catalytic RNA. These two helices contain 10 base pairs, all of which have been changed and all have catalytic activity as long as base pairing is maintained. This shows we can cleave a target sequence in a substrate RNA as long as we maintain base pairing in these 10 base pairs. Furthermore, we extended the length of the 6 bp helix to 10 bp and cleavage still occurred.

The remaining four bases of the substrate are in a proposed 4 base loop having the sequence AGUC. We have tested substrate sequences with the A changed to G, U or C and all were cleaved at full activity. The GUC is required, however. This means the catalytic RNA can be engineered to cleave a target in a substrate RNA as long as the target sequence contains a GUC. This could potentially be used to cleave very large RNA molecules which have an exposed 14 base target sequence. If the catalytic RNA could be shown to be active *in vivo*, this would result in possible down regulation of genes be they organismal or viral.

Cleavage of non-native substrates by an engineered RNA catalyst was first shown by Cech and colleagues using the Tetrahymena ribozyme (10, 11) and later for the 'hammerhead' catalytic RNA (4, 12, 13).

METHODS

Minimum energy calculations were based on the program FOLD from the University of Wisconsin Genetics Computer Group (14,15). Refinements and modifications of the minimum energy

* To whom correspondence should be addressed

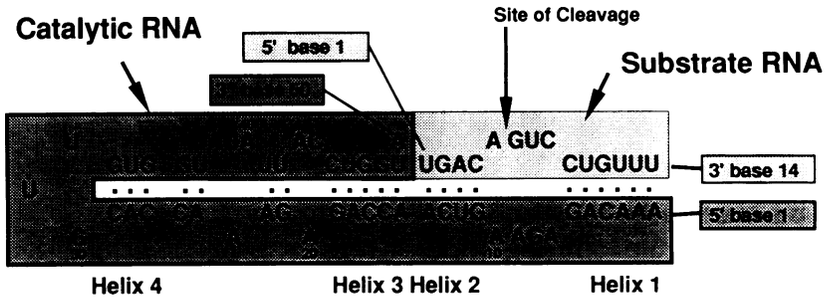


Figure 1. The 'hairpin' catalytic RNA model. Two molecules are shown folded: catalytic RNA which contains 50 bases of (−)sTRSV sequence and substrate RNA which contains 14 bases of satellite sequence (2). The model is based on minimum energy calculations (15−18) and mutagenesis results (Figs 2−4).

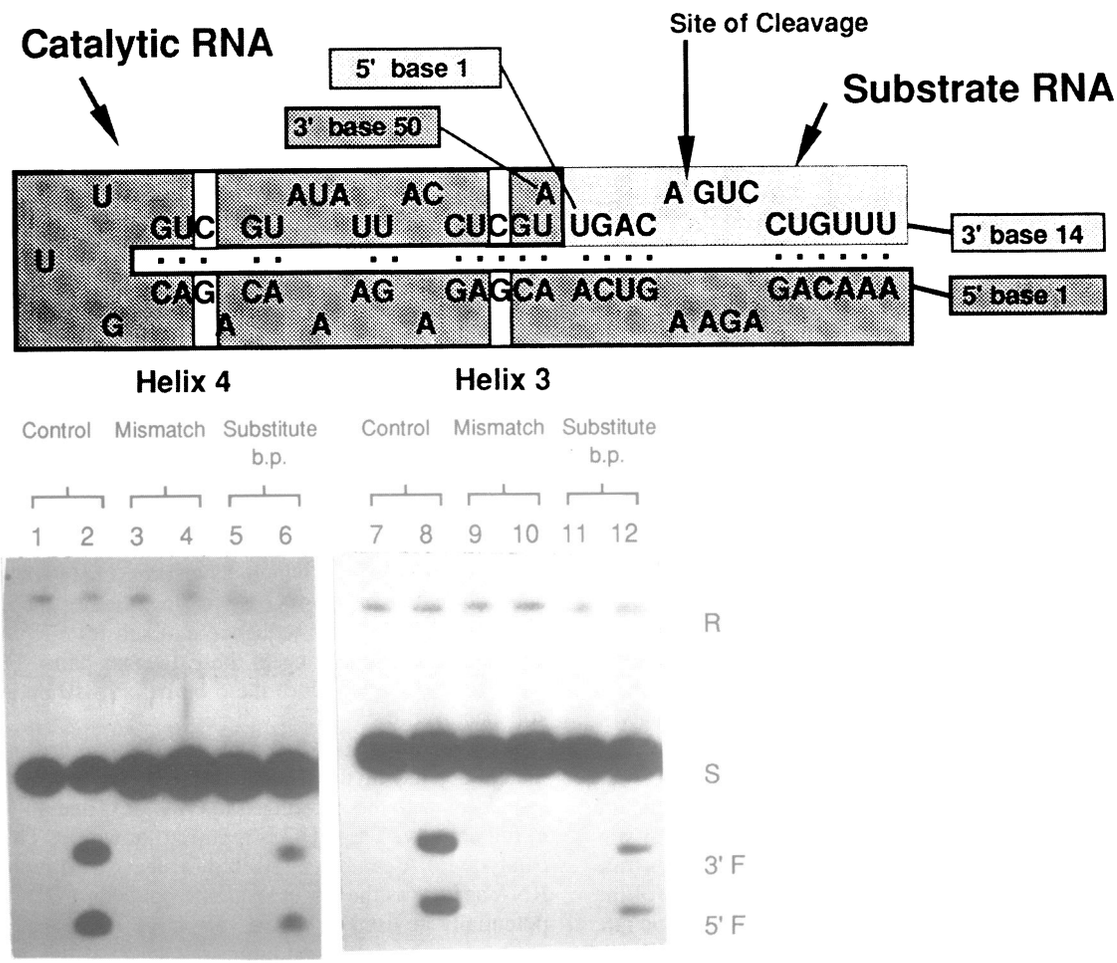


Figure 2. Mutagenesis of bases in helices 4 and 3 and the resulting catalytic activity. All substrate RNAs have an additional GCG vector base sequence at their 5′ end and the catalytic RNAs all have an additional 'G' at their 5′ end. For helix 4, a G→C base mutation in base #35 of the catalytic RNA has no catalytic activity, lanes 3 and 4. The double mutant, G35→C and base C27→G has restored catalytic activity, lanes 5 and 6. For helix 3, the catalytic RNA single base mutant at base G47→C was inactive, lanes 9 and 10, while the double mutant, with the second mutation C17→G, had restored activity, lanes 11 and 12. Lanes 1, 3, 5, 7, 9 and 11 are at zero time and lanes 2, 4, 6, 8, 10 and 12 are 15′ incubation under cleavage conditions; [catalytic RNA] = 0.0065μM; [substrate RNA] = 0.17μM; cleavage at 37°C. 'R' is catalytic RNA, 'S' is substrate RNA, '5′F' is the 5′ cleavage fragment and '3′F' is 3′ cleavage fragment. The control, lanes 1 and 2, is cleavage of the native substrate RNA by the native catalytic RNA sequences given Fig. 1.

structure predicted by this program used more recent results (16,17).

RNAs having the desired sequences were transcribed using the T7 RNA polymerase transcription method on synthetic DNA templates double stranded at the promoter site by the methods of Milligan (18). Synthetic DNA was made using phosphoramidite chemistry on the Northern Illinois University DNA synthesizer (Applied Biosystems 381A). Transcriptions were carried out as previously described (2) using alpha ^{32}P labelled CTP. The cleavage reactions were done by incubating

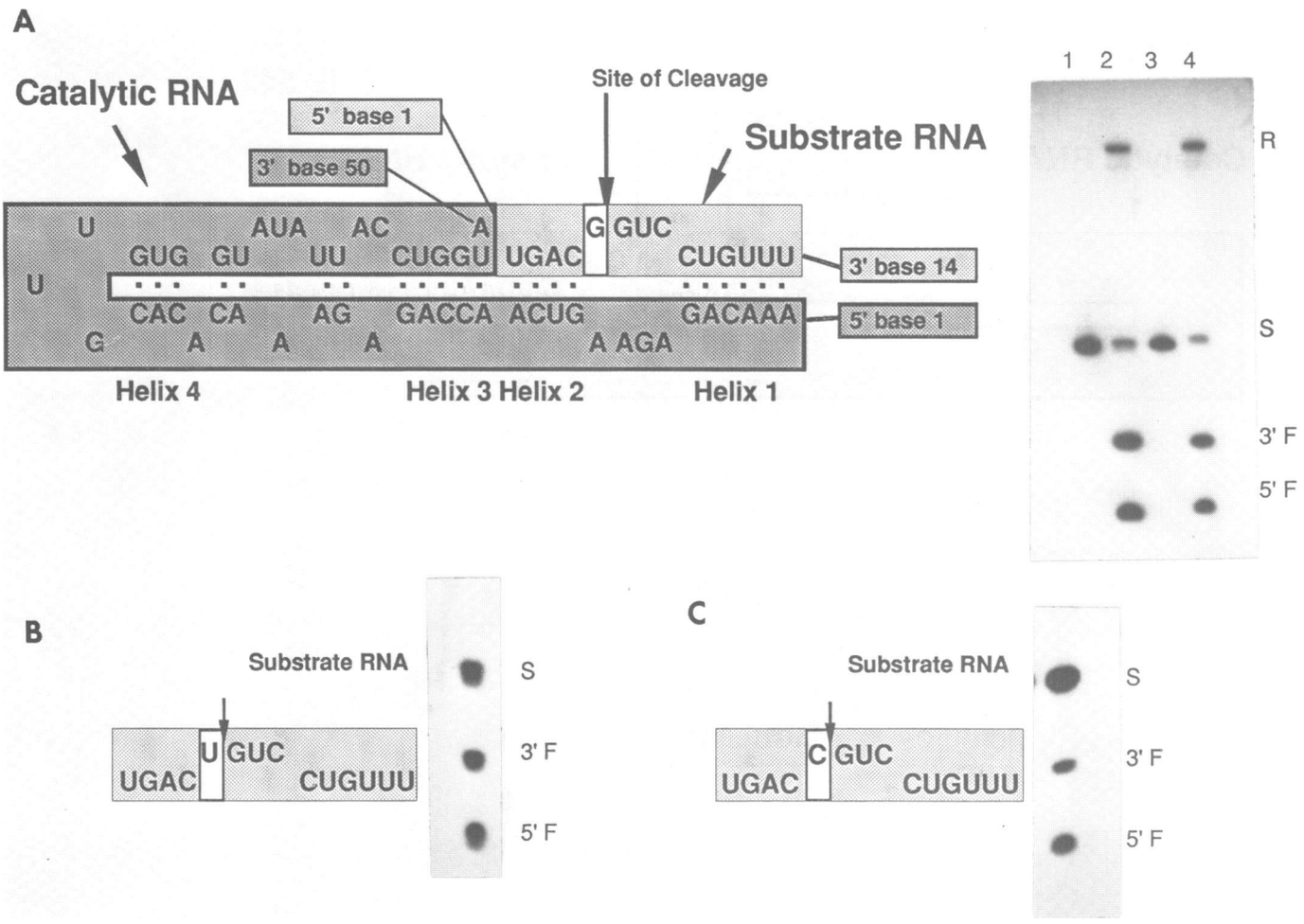


Figure 3. Cleavage of substrate RNA having base changes in the AGUC loop by catalytic RNA. Substitutions were made in the 'A' base of the substrate RNA AGUC loop sequence of Fig. 1. All substrate RNA had an additional GCG at the 5′ terminus and catalytic RNA has an additional 'G'. (**A**) A substrate RNA with a 'G' base substitution, GGUC substrate RNA, was cleaved by catalytic RNA, lanes 3 and 4. Control, lanes 1 and 2, was cleavage of native, AGUC, substrate RNA by catalytic RNA as shown in Fig. 1. Cleavage conditions were [control substrate RNA] = 0.4μM; [GGUC substrate RNA] = 0.2μM; [catalytic RNA] = 0.016μM; 37°C for 40min. Lanes 1 and 3 contain only substrate RNA with no catalytic RNA and lanes 2 and 4 contain substrate RNA and catalytic RNA. (**B**) Cleavage of substrate RNA with 'U' base substitution by catalytic RNA of Fig. 1. [substrate RNA] = 0.12 μM; [catalytic RNA] = 0.0065μM; 37°C for 60 min; catalytic RNA was unlabelled. (**C**) Cleavage of substrate RNA with 'C' base substitution by catalytic RNA of Fig 1. [substrate RNA] = 0.08μM; [catalytic RNA] = 0.0065μM; 37°C for 60 min; catalytic RNA was unlabelled.

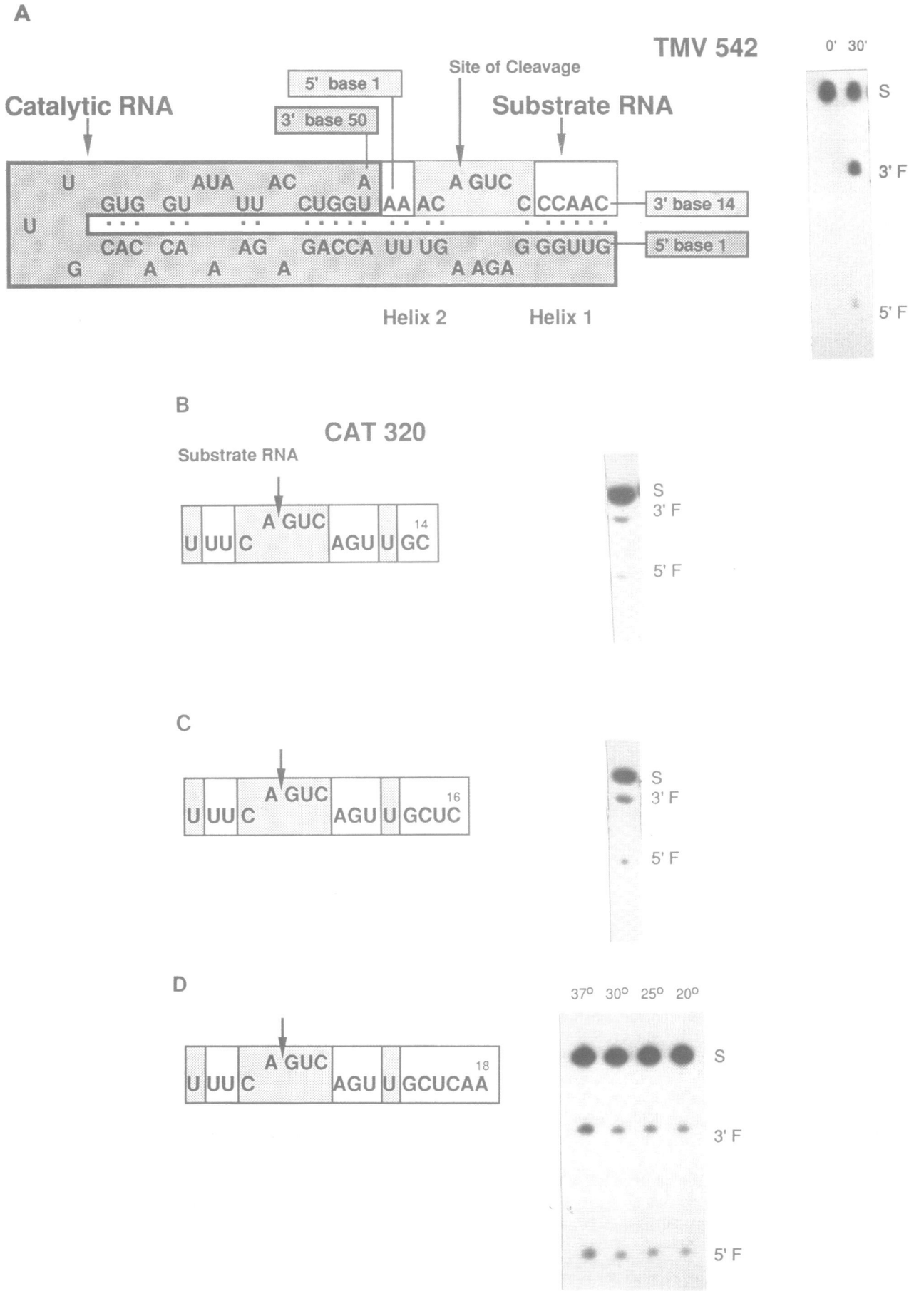
catalytic RNA with substrate RNA at the desired temperature in 12mM $MgCl_2$, 2mM spermidine, 40mM Tris pH7.5; reaction terminated with 7M urea; the products were run out on 20% acrylamide 7M urea gels using appropriate standards; and detected by autoradiography. The control substrate RNA and catalytic RNA were sequenced by direct RNA sequencing methods and end groups determined on the 5′ and 3′ cleavage products to unequivocally identify them (2). All sequences were as predicted from the sequences of the synthetic DNA templates and therefore the remaining mutant RNAs were not directly sequenced but presumed to be correctly transcribed from the known DNA template.

Note that for each figure, both the catalytic RNA and substrate RNA have additional vector sequence which is given in the figure legend but not shown in the figure drawing. The numbering of bases in the figures refers to the 50 base catalytic RNA and the 14 base substrate RNA having the original minimal (−)sTRSV sequences previously determined (2).

RESULTS AND DISCUSSION

Fig. 1 shows the proposed 'hairpin' two dimensional model for the 50 base catalytic RNA and the 14 base substrate RNA sequences found in (−)sTRSV (2). The model is characterized by a closed loop in the catalytic RNA and an open ended stem region with four major helices identified. Helices 1 and 2 are between the catalytic RNA and the substrate region. The two remaining primary helical regions, helices 3 and 4 are within the sequence of the catalytic RNA alone. This model has a free energy of helix formation of −12.5 kcal/mole at 37°C according to published parameters (15−18).

Mutagenesis experiments carried out on each of the four proposed helices shown in Fig. 1 support their existence by showing that base pairs exist within them. For helix 4 (Fig. 2) a mutation of G35→C created a C:C mismatch and was inactive, but the activity was restored by a second mutation, C27→G which formed a substitute base pair with the first mutation. Such a result


A
TMV 542
0' 30'
Site of Cleavage
5' base 1
3' base 50
Catalytic RNA
Substrate RNA
U AUA AC A A GUC
GUG GU UU CUGGU AA AC C CCAAC
3' base 14
U
CAC CA AG GACCA UU UG G GGUUG
5' base 1
G A A A A AGA
Helix 2
Helix 1
S
3' F
5' F
B
CAT 320
Substrate RNA
A GUC
U UU C AGU U GC
14
S
3' F
5' F
C
A GUC
U UU C AGU U GCUC
16
S
3' F
5' F
D
A GUC
U UU C AGU U GCUCAA
18
37° 30° 25° 20°
S
3' F
5' F

supports the existence of this base pair in helix 4 and is evidence for a helix (19). Helix 3 was shown similarly (Fig. 2). The mutation G47→C was inactive since a C:C mismatch would occur in helix 3, but the double mutant which formed a substitute base pair was catalytically active. A similar result was seen for the A:U base pair adjacent to this G:C pair. In every case the single base change mismatch mutation was a down mutation while the only up mutation was the double mutant with an alternate base pair.

This data excludes an alternate hairpin model with a different base pairing scheme at the loop end of the molecule and predicted to be only 2.2 kcal/mole less stable than the structure shown. This model had a predicted G35:C29 base pair and was further tested. In addition to the single mutant G35→C being inactive, the double mutant with both G35→C and C29→G was also inactive (data not shown).

The unpaired sequence of bases in the loop of the substrate sequence of the proposed model shown in Fig. 1 is AGUC. We mutagenized these bases to determine if they are required. The first 'A' base in this loop could be changed to either G, U, or C and full catalytic activity was maintained (Fig. 3). However, catalytic activity was destroyed (data not shown) when any base in the remaining GUC sequence, G→A, U→A, or C→G respectively, was changed. Thus the sequence requirement for the target site of cleavage by the 'hairpin' catalytic RNA is GUC.

An additional two base pairs could be proposed in the regions between helices 1 and 2 to lower the free energy an additional 3.4 kcal/mole, but mutagenesis studies do not support this. These potential base pairs would be between substrate RNA base U7 and catalytic RNA base A9 for one base pair and substrate RNA base C8 and catalytic RNA base G8 for the second potential base pair. Single mutation mismatches having substrate base changes, U7→A and C8→G respectively, were inactive as were single mutation catalytic RNA changes of G8→C and A9→U respectively. The double mutations having potential restored base pairs (substrate RNA U7→A with catalytic RNA A9→U as one double mutant and substrate RNA C8→G with catalytic RNA G8→C as another double mutant) were also catalytically inactive (data not shown).

In the model shown in Fig. 1, two helices, helix 1 and 2, are proposed to exist between the catalytic RNA and the substrate RNA. A substrate RNA with two base changes in helix 2 and four base changes in helix 1 was cleaved by an engineered catalytic RNA such that the base pairing was restored in these two helices with substitute base pairs (Fig. 4A). This substrate was not cleaved by catalytic RNA with mismatches in these two helices. A similar result was seen for a different substrate and corresponding catalytic RNA wherein a total of seven different base pairs were substituted in helices 1 and 2 (Fig. 4B). The temperature optimum for the reaction varies according to the sequence in these two helices.

It was also seen that helix 1 could be extended beyond the native 6 bp and catalytic activity maintained. For example catalytic activity was present when two base pairs are added to helix 1 to give a 16 base target site (Fig. 4C), four more base pairs to give an 18 base target site (Fig. 4D). A 20 base target site was also cleaved but much less efficiently. All these substrates were cleaved by a catalytic RNA which base paired to form the four base pairs in helix 2 and the base pairs in helix 1. In this case a longer helix 1 increased the temperature optimum of the reaction. The 14 and 16 base target sites were inactive at 37°C but the 18 base target site was cleaved at this temperature.

It is clear that different substrate RNA molecules can be cleaved by appropriately engineered catalytic RNA if base pairing is maintained between the substrate RNA and catalytic RNA in helices 1 and 2. The 14 base substrate RNA sequence in Fig. 4A is that found in the sequence of tobacco mosaic virus with the cleavage site being at base position #542 in the TMV sequence (20). Thus Fig. 4A shows cleavage of a target RNA sequence in the TMV sequence. Similarly the target sequences cleaved in Figs. 4 B, C, D are respectively 14, 16 and 18 base sequences found in the sequence of chloramphenicol acetyl transferase (CAT) mRNA with the site of cleavage being at base position #320 in the Tn9 sequence (21).

RNA sequences cleaved to date by the appropriately engineered hairpin catalytic RNA were such that at one time or another, all bases in helices 1 and 2 were changed and the substrate was still cleaved by a catalytic RNA which base pairs to form the appropriate helix 1 and 2. Mismatches were inactive but alternate base pairs were catalytically active. One substrate had an 18 base target site, UGACAGUCCUGUUUUUUU, and when reacted with its corresponding catalytic RNA, the reaction had a K_m=0.03μM and k_{cat}=7.1/min. These catalytic parameters are the best for any catalytic RNA known to date under these mild conditions of 37°C, low salt and neutral pH.

Fig. 5 summarizes these mutagenesis studies and gives the sequence requirements for cleavage of heterologous RNA by an appropriately engineered catalytic RNA. It is possible to specifically cleave an RNA molecule 5′ of a GUC sequence and with flanking regions of variable sequence. The GUC sequences are simply identified, 'hairpin' catalytic RNA engineered such that it base pairs to the flanking sequences around the NGUC and cleavage of the target sequence occurs. It is necessary to determine the optimal length of helix 1 to obtain the best reaction at the temperature required. We have successfully used this strategy to cleave a large number of heterologous RNA target sequences (data not shown). During the course of these experiments all bases in helix 1 and 2 were changed in one sequence or another when cleavage occurred. It must be noted, however, that large variations in k_{cat} and K_m occur as a function of the sequence flanking the NGUC sequence. Larger RNAs with unpaired bases beyond the substrate recognition site, were cleaved as long as the target site near the GUC was exposed.

One possibly important target site we have cleaved is found in the transcript of the tat gene in the AIDS virus HIV-1. This sequence, 5′ UGGGUGUCGUCGACAUA 3′, is found near the splice acceptor site of tat exon 2. Cleavage occurred at the UGUC sequence where the 'U' is base #5366 of the HIV-1 RNA (22).

Figure 4. Cleavage of substrate RNA by catalytic RNA with base changes in helices 1 and 2. (**A**) Cleavage of an RNA substrate with two base changes in helix 2 and five base changes in helix 1 by an RNA catalyst which restored the base pairing. Conditions were [substrate RNA]= 0.04μM; [catalytic RNA]= 0.0025μM; 37°C for 30 min in lane 2, and zero time in lane 1. SB) Cleavage of an RNA substrate with different base substitutions in helices 1 and 2 by a catalytic RNA which restores base pairing. Only the substrate RNA is shown. Reaction conditions were [substrate RNA] = 0.05μM; [catalytic RNA] = 0.005μM; 16°C for 20min. The substrate RNA has one additional 'G' vector base at the 5′ terminus and the catalytic RNA has 'GA' at its 5′ terminus. (**C**) The same reaction as in Fig. 4B but helix 1 has been extended by two bases. This is now a 16 base recognition sequence. Conditions are the same except catalytic RNA has no 5′ vector bases. (**D**) Cleavage of a substrate RNA with 4 base extension to helix 1 to give an RNA target sequence of 18 bases. The substrate had an additional GCG vector sequence at the 5′ end and the catalytic RNA had additional GGG 5′ vector sequence. Reaction conditions were the same as in (A) except the time was 40min and the temperatures were at 20°C, 25°C, 30°C and 37°C as shown for the lanes.

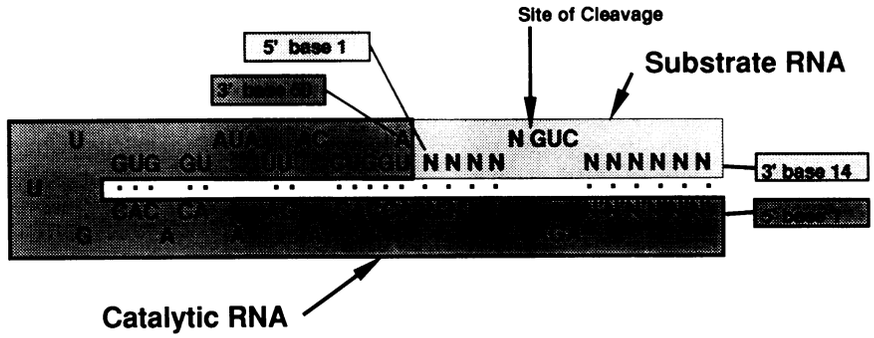


Figure 5. Summary of base requirements in the target sequence of the RNA substrate. The substrate RNA target site requires only a GUC sequence as long as the substrate RNA base pairs with the catalytic RNA, N:N'.

This particular site is thought to be very exposed during transcription and it has been shown that antisense RNA made to this region is very effective in inhibiting the virus itself (23). This sequence cleaved as part of a larger RNA with an additional GCG at the 5' end and GC at the 3' end. Additional bases at the 3' end and 5' end have been used for other substrates and did not appear to affect the reaction. Experiments are in progress to cleave very large RNA transcripts having suitable target sites.

If these experiments are successful, the 'hairpin' catalytic RNA may be more efficient in destroying RNA transcripts than antisense RNA (24) since the kinetic parameters of this reaction are very favorable under physiological conditions. If 'hairpin' RNA catalysis can be shown *in vivo*, it would have the potential to be developed as a general down regulator of gene expression which may include viruses as well.

ACKNOWLEDGEMENTS

This work was supported in part by NSF grant # DMB-8817576, NIH grant R01 GM43718 and the Biotechnology Research Development Corporation to A.H. We thank Doug Turner and Olke Uhlenbeck for helpful discussions and Jeffrey Aspinall for computer searches and modeling.

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. Cech, T. and Bass, B. (1986) Ann. Rev. Biochem. **55**, 599–629.
4. Uhlenbeck, O. (1987) Nature **328**, 596–600.
5. Forster, A. and Symons, R. (1987) Cell **49**, 211–220.
6. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) Cell **35**, 849–857.
7. Waugh, D., Green, C. and Pace, N. (1989) Science **244**, 1569–1571.
8. Sharmeen, L., Kuo, M., Dinter-Gottlieb, G. and Taylor, J. (1988) J. Virology **62**, 2674–2679.
9. Wu, H., Lin, Y., Lin, F., Makino, S., Chang, M. and Lai, M. (1989) Proc. Nat. Acad. Sci. U.S.A. **86**, 1831–1835.
10. Zaug, A., Grosshans, C. and Cech, T. (1988) Biochemistry **27**, 8924–8931.
11. Zaug, A., Been, M. and Cech, T. (1986) Nature **324**, 429–433.
12. Koizumi, M., Iwai, S. and Ohtsuka, E. (1988) FEBS Lett. **228**, 228–230.
13. Haseloff, J. and Gerlach, W. (1988) Nature **334**, 585–591.
14. Zuker, M. and Stiegler, P. (1981) Nuc. Acids Res. **9**, 133–148.
15. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nuc. Acids Res. **12**, 387–395.
16. Freier, S., Kierzek, R., Jaeger, J., Sugimoto, N., Caruthers, M., Neilson, T. and Turner, D. (1986) Proc. Nat. Acad. Sci. U.S.A. **83**, 9373–9377.
17. Groebe, D. and Uhlenbeck, O. (1988) Nuc. Acids Res. **16**, 11725- 11735.
18. Milligan, J., Groebe, D., Witherell, G. and Uhlenbeck, O. (1987) Nuc. Acids Res. **15**, 8783–8798.
19. Fox, G. and Woese, C. (1975) Nature **256**, 505–507.
20. Goelet, P., Lomonossoff, G., Butler, P., Akam, M., Gait, M. and Karn, J. (1982) Proc. Nat. Acad. Sci. U.S.A. **79**, 5818–5822.
21. Alton, N. and Vapnek, D. (1979) Nature **282**, 864–869.
22. Muesing, M., Smith, D., Cabradilla, C., Benton, C., Lasky, L. and Capon, D. (1985) Nature **313**, 450–458.
23. Sarin, P., Agrawal, S., Civeira, M., Goodchild, J., Ikeuchi, T. and Zamecnik, P. (1988) Proc. Nat. Acad. Sci. U.S.A. **85**, 7448–7451.
24. *Antisense RNA and DNA*, ed. by Douglas A. Melton, Cold Spring Harbor Laboratory, 1988.