



HHS Public Access

Author manuscript

Nat Chem Biol. Author manuscript; available in PMC 2016 February 01.

Published in final edited form as:

Nat Chem Biol. 2015 August ; 11(8): 606–610. doi:10.1038/nchembio.1846.

New classes of self-cleaving ribozymes revealed by comparative genomics analysis

Zasha Weinberg^{1,*}, Peter B. Kim^{2,*}, Tony H. Chen³, Sanshu Li^{1,2}, Kimberly A. Harris^{1,2}, Christina E. Lünse², and Ronald R. Breaker^{1,2,3}

¹Howard Hughes Medical Institute, Yale University, Box 208103, New Haven, CT 06520-8103, USA

²Department of Molecular, Cellular and Developmental Biology, Yale University, Box 208103, New Haven, CT 06520-8103, USA

³Department of Molecular Biophysics and Biochemistry, Yale University, Box 208103, New Haven, CT 06520-8103, USA

Abstract

Enzymes made of RNA catalyze reactions that are essential for protein synthesis and RNA processing. However, such natural ribozymes are exceedingly rare, as evident by the fact that the discovery rate for new classes has dropped to one per decade from about one per year during the 1980s. Indeed, only 11 distinct ribozyme classes have been experimentally validated to date. Recently, we recognized that self-cleaving ribozymes frequently associate with certain types of genes from bacteria. Herein this synteny was exploited to identify divergent architectures for two previously known ribozyme classes and to discover additional noncoding RNA motifs that are self-cleaving RNA candidates. Three new self-cleaving classes, named twister sister, pistol and hatchet, have been identified from this collection, suggesting that even more ribozymes remain hidden in modern cells.

The RNA World theory¹ is based on the notion that contemporary life is derived from organisms that exploited numerous and functionally diverse ribozymes before the emergence of proteins. Many of the roles once played by catalytic RNAs presumably diminished over time due to competition from protein enzymes. A few ribozyme classes that perform critical biochemical transformations such as ribosome-mediated peptide bond formation² and various RNA processing reactions^{3–5} persisted either because their roles

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence should be addressed to R.R.B. (ronald.breaker@yale.edu).

*These authors contributed equally to this work.

AUTHOR CONTRIBUTIONS

Z.W. conducted all bioinformatics searches and created the RNA consensus models. Validation studies were conducted for twister sister by P.B.K. and T.H.C., for pistol by K.A.H., for hatchet by S.L., hammerhead by P.B.K., and HDV by P.B.K. and S.L. P.B.K., T.H.C. and C.E.L. screened additional non-cleaving RNAs. R.R.B. worked with all authors to plan experiments, interpret data and write the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

could not easily be replaced or because RNA is particularly well suited to perform these tasks¹. However, modern natural ribozymes are exceedingly rare, as evident by the fact that the discovery rate for new classes has dropped to one per decade from about one per year during the 1980s.

Of the 11 previously validated ribozyme classes, six are self-cleaving. Three self-cleaving classes, hammerhead⁶, HDV⁷ and twister⁸, have thousands of representatives in living systems. Interestingly, only a few of these representatives have been linked to biological roles, such as rolling-circle replication of RNA pathogens⁹, processing of repetitive RNA sequences¹⁰, and metabolite-dependent gene regulation¹¹. Therefore, most self-cleaving ribozyme representatives have unknown utility, and much remains to be learned about the involvement of self-cleaving ribozymes in cellular function. Given the relative importance of known ribozymes to information processing, gene expression, and genomic integrity, the discovery of additional classes provides opportunities to advance our understanding of modern biochemical processes, to gain further insights into RNA structures, and to shed light on the possible diversity of RNA World functions.

Unfortunately, the pace of novel ribozyme class discovery has been exceedingly slow over the last 25 years, and all confirmed discoveries were made serendipitously and not while searching for ribozymes. One example of a purposeful search for self-cleaving ribozymes in humans using a biochemical selection method¹² yielded a variant of the HDV self-cleaving ribozyme class and three other natural ribozyme candidates. Recently, we noted that many hammerhead and twister ribozymes commonly reside within a few kilobases of each other and likewise near certain protein-coding genes⁸. Although the biological basis for this association remains mysterious, we hypothesized that other self-cleaving ribozyme classes might also appear in the vicinity of these genetic elements and that a computational search strategy might reveal new ribozyme classes.

In the current study, a search for conserved RNA structures nearby to these genetic elements yielded a ribozyme class that we called “twister sister” because it has vague similarities in sequence and secondary structure to twister ribozymes. However, the two ribozyme classes cleave at different sites, and therefore the significance of the sequence and structural similarities will require further investigation. The search also revealed variants of hammerhead and HDV ribozymes, as well as additional conserved RNA structures that did not self-cleave *in vitro*. Similarly, we performed another search using an expanded set of genetic elements and genomic sequences, and detected additional novel ribozyme classes, which we named “hatchet” and “pistol”.

RESULTS

Identification of ribozyme candidates

To select promising genomic locations for our search, we first enumerated genetic elements that are often located within 6 kilobases of twister or hammerhead self-cleaving ribozymes. Protein domains encoded by these nearby genes were identified using the Conserved Domain Database (CDD)¹³, and by using the JackHMMER program¹⁴ (see **Online Methods**, Supplementary Results, Supplementary Table 1). We calculated the frequency of

these ribozymes in the vicinity of each conserved protein domain or RNA class (Supplementary Table 2), and chose a total of six RNA classes (three permutations of twister and hammerhead ribozymes) and 16 conserved protein domains (Supplementary Table 3). We speculated that the intergenic regions (IGRs) nearby to these elements (available at <http://breaker.research.yale.edu/ts>), totaling ~7 million base pairs, are likely to be enriched for self-cleaving ribozymes and therefore we used them in our search.

Conserved RNA structures were then identified with a computational pipeline^{15,16} that first employed BLAST to identify IGR groups that are presumably homologous, and then employed the CMfinder program¹⁵ to predict conserved secondary structures. We manually analyzed the resulting predictions to refine the secondary structure model for each group of homologous sequences. Predictions were evaluated based on support from covarying mutations, which are known to provide powerful evidence for base pairing in natural RNAs^{17–19}.

We predicted fifteen distinct RNA motifs with conserved secondary structures (Fig. 1, Supplementary Fig. 1). Alignments and information on nearby genes are provided for all motifs (Supplementary Data Sets 1 and 2). Among these were unusual variants of hammerhead and HDV ribozymes (Fig. 1a, Fig. 1b, Supplementary Fig. 2), and a motif we named “twister sister” because it has vague sequence and structural similarities with a recently discovered self-cleaving ribozyme called twister⁸ (Fig. 1c). The genetic elements actually involved in detecting each confirmed ribozyme class are available in Supplementary Table 4. General properties of motifs that did not self-cleave in our experiments are listed in Supplementary Table 5.

Three ribozyme candidates self-cleave *in vitro*

To assess self-cleaving ribozyme activity, we conducted *in vitro* transcriptions using wild-type (WT) or various mutant DNA templates of each candidate. Representatives of all three new motifs undergo efficient self-cleavage during transcription (Supplementary Fig. 2). Twelve additional motifs (Supplementary Fig. 1) did not cleave *in vitro* (data not shown), and were not further pursued.

The hammerhead ribozymes uncovered in our study are variants of the type I architecture, wherein stem I is not covalently closed. These unusual variants are predicted to form stem II with only a single base-pair, which is a feature seen in only one hammerhead ribozyme representative reported previously⁶. Moreover, the loop of stem II forms a long-distance pseudoknot with the loop of an additional hairpin formed by the 5' region of the motif. Despite these distinct structural features, the variant hammerhead ribozymes retain the normal site of RNA cleavage used by all other hammerhead ribozymes examined to date (Fig. 1a, Supplementary Fig. 3).

The HDV ribozyme variants, which had not previously been detected in bacterial metagenomes, often carry an E-loop structure²⁰ at the base of P4. This arrangement is present in only one previously reported HDV ribozyme²¹. A bacterial representative carrying the E-loop sub-structure was found to undergo self-cleavage at the same location as other more-typical HDV ribozymes (Supplementary Fig. 4), suggesting that the newly-found

variants form the same active site using an alternative structural feature. Eukaryotic HDV variants with similar sequence and structure features were found by homology search, and also by an independent bioinformatics analysis of fungal genomes (S.L. and R.R.B., unpublished data).

Biochemical properties of twister sister ribozymes

The “twister sister” motif, newly revealed by our search, remotely resembles twister ribozymes chiefly because some twister RNAs have P1 through P5 stems in an arrangement similar to twister sister, and because of similarities in the nucleotides in the P4 terminal loop. However, we find no evidence of pseudoknot formation via Watson-Crick base pairing, which occurs in all known twister ribozymes. In addition, there is poor correspondence among many of the most highly conserved nucleotides in each of the two motifs. Given these observations, it was not immediately clear if twister sister RNAs would self-cleave. Even if they function as ribozymes, twister sister RNAs could either represent a distinct ribozyme class or simply be a highly divergent form of twister ribozymes.

A twister sister construct, called TS-1 and based on a microbial metagenomic DNA source, was engineered to function as a bimolecular complex with separate enzyme and substrate strands (Fig. 2a). The substrate strand of this complex was cleaved in the presence of Mg^{2+} only when the enzyme strand is present (Fig. 2b). Surprisingly, the cleavage site was on the right side of the internal loop linking P1 and P2 (Fig. 2c). Likewise another twister sister ribozyme construct called TS-2 from different microbial metagenomic DNA was cleaved in this same location (Supplementary Fig. 5). This cleavage site between nucleotide C13 and A14 is on the opposite side relative to the internal loop cleaved by twister. The distinct cleavage site, along with the sequence and structural differences noted above, provided initial evidence that twister sister RNAs might represent a separate ribozyme class from twister.

The TS-1 cleavage site was confirmed by mass spectrometry (Fig. 2d), which also revealed that the ribozyme reaction yields 5' cleavage product with a terminal 2',3'-cyclic phosphate group, and a 3' cleavage product with a 5' hydroxyl group. These reaction products result via a mechanism wherein the 2' oxygen of C13 attacks the adjacent phosphorus atom, with subsequent departure of the 5' oxygen of A14 (Supplementary Fig. 6). As expected for a ribozyme that uses this general phosphoester transfer mechanism, the cleavage reaction cannot proceed when the nucleotide corresponding to C14 of the substrate lacks the 2' oxygen nucleophile (Supplementary Fig. 7). This mechanism for RNA strand scission is identical to all six previously characterized self-cleaving ribozyme classes⁵.

To further assess the characteristics of twister sister ribozymes, we established additional biochemical properties of the TS-1 bimolecular complex. With each increase in pH unit, there was a ~10-fold increase in TS-1 cleavage activity, with a plateau near pH 7 (Fig. 2e). The simplest explanation for this pH-activity profile is that the ribozyme shifts the pK_a of the 2' hydroxyl group of C13 to approximately 7, which otherwise has a pK_a value of ~13.7 (ref. 21). An oxyanion at the 2' position will function as a far more powerful nucleophile compared to the 2' hydroxyl group, and therefore the rate constant should increase in proportion to the probability that the 2' position is deprotonated, but only if the chemical

step of the reaction is rate limiting. If this explanation is true, then ribozyme activity does not continue to increase beyond pH 7 because the 2' oxygen of C13 is already near fully deprotonated at values approaching and exceeding physiological pH.

TS-1 activity was also highly dependent on the concentration of divalent metal ion, and the steep increase in ribozyme rate constant only plateaued at Mg^{2+} concentrations above 1 mM (Fig. 2f). Divalent metal ions could strongly induce ribozyme activity either by serving important roles in the formation of the global structure of the ribozyme or by directly promoting catalysis at the active site. Both twister sister and twister ribozymes displayed similar dependencies on pH and Mg^{2+} (see ref. 8 for kinetic details of twister), suggesting that they might share some catalytic strategies (Supplementary Fig. 6) despite their differences in conserved sequences, structural features, and cleavage site. Although twister ribozymes also are strongly activated by the addition of Mg^{2+} or by a number of other divalent metal ions, initial finding suggests that these ribozymes do not employ metal ions in their active site^{8,24}. However, a recent x-ray structure model has been proposed wherein a Mg^{2+} ion interacts with the phosphate moiety at the cleavage site²⁵.

The maximum observed rate constant (k_{obs}) value for TS-1 was $\sim 5 \text{ min}^{-1}$ when both pH and Mg^{2+} are optimal. This is considerably slower than the twister ribozymes examined previously⁸, which have k_{obs} values estimated by extrapolation to be more than $1,000 \text{ min}^{-1}$. Interestingly, the pH profile and maximum rate constant for TS-1 are nearly identical to those of a collection²² of ribozymes and deoxyribozymes created previously by using directed evolution. It has been suggested that this previous collection of ribozymes and deoxyribozymes might use only two catalytic strategies to promote catalysis: orienting the labile linkage for in-line nucleophilic attack (called α , Supplementary Fig. 6) and promoting deprotonation of the 2' hydroxyl group (called γ). It is known that the maximum rate constant generated by γ catalysis is $\sim 0.02 \text{ min}^{-1}$, whereas α catalysis is estimated to provide an additional ~ 100 -fold increase in the rate constant²². If these two catalytic strategies are independent, and if their rate constant enhancements are multiplicative, this permits an estimate for a “speed limit” for catalysts that employ only α and γ catalytic strategies of $\sim 2 \text{ min}^{-1}$.

Given that the maximum k_{obs} value measured for the bimolecular TS-1 construct is only modestly greater than the estimated maximum k_{obs} for $\alpha\gamma$ ribozymes, we wondered if other twister sister ribozymes might achieve greater speeds. We tested two additional constructs called TS-3 and TS-4 (Supplementary Fig. 8). Both constructs exhibited much higher rate constants under suboptimal pH and Mg^{2+} ion concentrations, indicating that their rate constants would be far in excess of the TS-1 construct when tested under optimal reaction conditions (predicted as $>100 \text{ min}^{-1}$ for TS-3). These results demonstrate that at least some twister sister ribozymes can combine catalytic strategies to exceed the $\alpha\gamma$ speed limit.

To allow further comparisons of twister and twister sister, we consulted three recently reported atomic-resolution structures^{23–25}. These structures largely agree with the previously proposed secondary structure model for twister ribozymes⁸, and reveal the importance of highly conserved nucleotide positions. In one model, twister ribozymes appear to employ three catalytic strategies to achieve their high speeds²⁴: α and γ as

described above, as well as protonation of a non-bridging phosphate oxygen (β). However, the active sites in other structure models for twister ribozymes are either unclear²³ or different²⁵. As a result, additional studies will be needed to confirm the precise nature of the active site and the catalytic strategies used by twister ribozymes.

As noted above, the overall secondary structures of twister and twister sister exhibit vague similarities. However, there are differences in two places with important roles. First, the conserved nucleotide predicted to be responsible for β catalysis in twister ribozymes is part of a pseudoknot²⁴, which is presumed to be absent in twister sister ribozymes. If twister sister ribozymes employ β catalysis, it is not clear how they promote this catalytic strategy using a structure similar to twister ribozymes. Second, although the nucleotides in the loop of P4 are notably similar, key conserved nucleotides in the P1 and P2 stems of twister are not found in twister sister. Both motifs have highly, but not invariantly conserved A nucleotides immediately 3' to the cleavage site. However, the modest similarities in nucleotide sequence near the cleavage site could also be explained by chance. At least two explanations for the differences in conserved nucleotides are possible: either twister sister and twister ribozymes have distinct active sites that use a similar scaffold, or the different nucleotides in twister sister actually form the same geometry and tertiary contacts as those in twister ribozymes. An atomic-resolution structure of a twister sister ribozyme will help address the extent to which these ribozymes use similar structures and catalytic strategies to accelerate RNA phosphoester transfer.

As with many other self-cleaving ribozyme classes, twister sister catalytic activity was supported by divalent metal ions other than Mg^{2+} (Supplementary Fig. 9a). However, the TS-1 RNA responded differently to Sr^{2+} and Ni^{2+} compared to twister¹¹. Moreover, of five Group 1 monovalent cations tested at 1 M, only Li^+ resulted in observable ribozyme activity (Supplementary Fig. 9b), whereas all five cations support twister activity. Finally, cobalt hexamine, an analog of hydrated Mg^{2+} , did not induce twister sister activity (Supplementary Fig. 9c), but did support twister activity⁸. Together, these observations further suggest that there might be differences at the active sites formed by twister sister and other self-cleaving ribozyme classes that cause the distinct responses to these metal ions.

Identification of additional novel ribozymes

After completing the analyses described above, we reapplied our bioinformatics search strategy using a larger collection of ribozyme-associated gene classes and using additional bacterial DNA sequences from the rapidly accumulating genomic databases. Several novel motifs were found whose representatives did not cleave when assayed *in vitro*, and these will be detailed in a future report. However, this analysis also uncovered two additional novel self-cleaving ribozyme classes which we named pistol (Fig. 3) and hatchet (Fig. 4). Initial computational and biochemical analyses confirmed that these RNAs use distinct conserved sequences and structural features to accelerate RNA cleavage by more than 10 million fold (Supplementary Fig. 10) via an internal phosphoester transfer mechanism. Therefore, nine of the 14 known natural ribozyme classes promote rapid self-cleavage, which indicates that modern organisms make extensive use of RNA's capacity to form diverse structures and accelerate RNA strand scission^{26,27}.

DISCUSSION

Our discovery of three novel ribozyme classes in one study demonstrates that regions near ribozyme-associated genes in bacteria and bacteriophages are fruitful places to search for novel classes. It remains unclear, however, why these genes frequently associate with self-cleaving ribozymes. In most instances, the functions of the protein products of these ribozyme-associated genes are unknown. However, it was previously noted that many genetic elements associated with self-cleaving ribozymes are typically carried by Mu-like phages⁸. Therefore certain phages might have particular need for RNA processing by self-cleaving ribozymes.

Twelve of 15 motifs identified in our original search did not exhibit self-cleavage activity during preparation by *in vitro* transcription. The reasons for the lack of ribozyme activity might be different for each motif. Some might function as ribozymes, but lack some critical component (e.g., an important domain was missing from the RNA molecule) or condition (e.g., a required metal ion or other cofactor) when tested outside of their natural environment. Some candidate motifs lack some of the structural features that are typical of self-cleaving ribozymes. For example, some candidates are unusually large compared to all known self-cleaving ribozyme classes, or they consist of only simple hairpin loops, whereas known natural self-cleaving ribozymes have more complex structures.

Perhaps some of the inactive motifs act as targets for endoribonuclease proteins, providing an alternate mechanism for cleavage. Some motifs may not even function as structured RNAs, despite evidence for structure formation via sequence covariation. The biologically relevant nucleic acid structure could be formed by single-stranded DNA, rather than its corresponding RNA transcript. Since many genes associated with self-cleaving ribozymes are typical of Mu-like phages, it is possible that Mu-like phages have a preference for RNA-based solutions to biological challenges such as gene regulation and RNA processing. If true, some of the 12 candidate motifs might serve other functions rather than promoting site-specific RNA cleavage.

The mystery regarding the biological utility of self-cleaving ribozymes is likely to deepen as more classes probably remain to be discovered in nature, and many more representatives of the known classes are certain to exist in organisms that have yet to have their genomes sequenced. Targeted computational searches should enable these ribozyme discoveries to be made more rapidly in the future. Moreover, each new ribozyme class provides another type of catalytic RNA for detailed dissection. Given their biological sources, these natural ribozymes will have had to compete effectively with protein-based enzymes to persist through evolution. Consequently, the catalytic activities of these RNAs should be highly refined compared to the many classes that have been created by directed evolution methods. A combination of biochemical and biophysical analyses on these natural ribozymes therefore should reveal much about how RNA molecules can promote rapid chemical transformations without the aid of protein factors.

ONLINE METHODS

Genome sequences and annotation

Our analysis used sequences in the bacterial and archaeal section of RefSeq²⁸ version 56, and various environmental sequences collected from IMG/M²⁹, the Human Microbiome Project³⁰, MG-RAST³¹, CAMERA³² or GenBank³³. The locations of genes were retrieved from RefSeq or IMG/M annotations, or predicted using MetaGene³⁴ or MetaGeneMark³⁵. Conserved domains were annotated using the Conserved Domain Database¹³ version 2.25, using a previously described procedure³⁶. The locations of twister ribozymes were predicted based on homology searches we conducted with Infernal³⁷ version 1.1, and these RNAs were used to generate the twister ribozyme consensus diagram. Other RNAs were predicted using Rfam³⁸, tRNAscan-SE³⁹, CRT⁴⁰ and rRNA prediction⁴¹. Rho-independent transcription terminators were predicted using RNie⁴². Consensus diagrams were generated using R2R⁴³.

Genetic elements associated with self-cleaving ribozymes

As in a previous study⁸, we collected genes located within 6 kilobases (Kb) of a twister or hammerhead ribozyme where each gene is at least 200 base pairs from the end of a sequence fragment. As before, HDV ribozymes and *glmS* ribozymes were not considered because their gene associations did not resemble that of twister or hammerhead ribozymes, while other self-cleaving ribozyme classes are not known in bacteria⁸.

Some of the genes in the above collection encoded proteins that did not contain a predicted domain in the Conserved Domain Database, yet were clearly homologous with other ribozyme-associated proteins in the collection. To group together related proteins, we adopted a strategy based on the JackHMMER program¹⁴. JackHMMER uses a single query protein to find homologs, and then uses the alignment of the homologs to the original query to conduct a more refined search for additional homologs, repeating this process in multiple rounds.

We ran JackHMMER searches on each of the proteins collected nearby to ribozymes above, and initially ran searches against the set of ribozyme-associated proteins. Proteins whose JackHMMER searches predicted a set of fewer than 20 homologous proteins were discarded. When the sets of homologs output from the JackHMMER searches of two query proteins overlapped, the protein with the smaller set was discarded, or an arbitrary protein was discarded if the homolog sets were equal in size. We performed an additional JackHMMER search for each of the remaining proteins against the full set of predicted proteins in our sequence database. The sets of predicted homologs arising from each of these searches were treated as a conserved domain and called SCRAP1-67 (SCRAP is Self-Cleaving Ribozyme Associated Protein) (Supplementary Table 1).

Detection of conserved RNAs

We used the conserved ribozyme-associated elements to identify putative non-coding sequences that were expected to be enriched for self-cleaving ribozymes. To evaluate the utility of each conserved element, we calculated two frequencies. For hammerhead and then

twister ribozymes, we calculated the frequency with which a ribozyme is within 6 Kb of representatives of the conserved element (Supplementary Table 2). These two frequencies can be viewed as estimates of how likely it is that a self-cleaving ribozyme will be present around examples of the conserved element.

We did not calculate p-values, primarily because of the difficulty in modeling correlations between metagenomic sequence fragments (caused by evolutionary relationships) that would otherwise distort p-value statistics. Moreover, the frequencies calculated are well suited to address the main goal of selecting the most promising genomic locations for discovering new ribozymes.

We manually selected conserved elements that had high frequencies for both hammerhead and twister ribozymes, as we anticipated that high numbers with both ribozyme classes would indicate a more robust association. We also observed that conserved elements and their associated self-cleaving ribozymes were almost always encoded on the same DNA template strand, and therefore we did not consider the complementary strands to be enriched for self-cleaving ribozymes. In the case of some conserved elements, it appeared that the ribozymes were usually located closer than 6 Kb to the element, so we often selected a shorter distance. In all, 22 conserved elements were selected (Supplementary Table 3).

The non-coding regions nearby to the selected conserved elements totaled 6.7 Mb. To identify conserved RNA motifs within these non-coding regions, we used a previously published method¹⁶. After manual analysis of the results, we identified 15 putative conserved RNA structures. By looking at consensus structures, we determined that three of these motifs corresponded to or had some aspects in common with known structural classes of self-cleaving ribozymes, *i.e.*, hammerhead, HDV and twister ribozymes. Eukaryotic homologs of motifs were found only for the HDV-like motif. Although we observed that many likely homologs of the HDV-like motif seemed defective, a similar phenomenon was observed with some eukaryotic hammerhead ribozymes⁶. We generally avoided these defective sequences, and did not attempt to find a comprehensive set of HDV ribozymes. However, we did include some HDV ribozymes that appeared to be truncated on their 5' ends, as has previously been observed in some cases⁴².

Evaluation of ribozyme self-cleavage during transcription

Experiments were conducted on two twister sister representatives (TS-1 and TS-2) identified in a human gut metagenome⁴⁵, found respectively in nucleotides 361 to 522 in sequence accession “scaffold1830_2_V1.CD-8” and the reverse complement of nucleotides 574 to 689 in sequence “scaffold909_4_MH0022.” Double-stranded DNA templates encoding these RNAs were generated by extending the synthetic DNA 5'-TAATACGACTCACTATAGG (containing the T7 RNA polymerase promoter sequence) on the appropriate template DNA by using *Taq* DNA polymerase. *In vitro* transcription reactions were performed as previously described⁸. Internally ³²P-labeled products were separated using denaturing (8 M urea) 15% polyacrylamide gel electrophoresis (PAGE) and detected with a Typhoon Trio+ Variable Mode Imager (GE Healthcare).

Cleavage assays for bimolecular ribozyme constructs

Substrate and enzyme RNAs for bimolecular complexes as designated were individually synthesized, either by *in vitro* transcription or by solid-phase chemical synthesis (Sigma-Aldrich), resulting in the elimination of loop sequences that otherwise join P5. Substrate RNAs, including those containing the 2'-deoxycytosine modification, were purchased from Sigma-Aldrich and 5' radiolabeled using γ -³²P [ATP] and T4 polynucleotide kinase (New England Biolabs) according to the manufacturer's instructions. Double-stranded DNAs encoding enzyme RNAs were prepared as described above and used as templates for *in vitro* transcription reactions, also as described above. Before being added to cleavage reactions, all RNAs were purified by denaturing PAGE, the appropriate product bands were eluted from gel slices using 10 mM Tris-HCl (pH 7.5 at 23°C), 200 mM NaCl and 1 mM EDTA, and concentrated by precipitation with ethanol.

Bimolecular cleavage reactions were incubated at room temperature in standard reaction conditions: 30 mM HEPES (pH 7.5 at 23°C), 100 mM KCl, 20 mM MgCl₂. Reactions contained 50 nM ³²P-labeled substrate and 1000 nM unlabeled enzyme RNA, unless otherwise indicated. Substrate and enzyme RNAs were combined in reaction buffer lacking magnesium, heated to 80°C for 1 min, and cooled to 23°C before cleavage reactions were initiated by the addition of MgCl₂. Unless otherwise indicated, reactions were halted by adding an equal volume of stop solution (90% formamide, 50 mM EDTA, 0.05% xylene cyanol and 0.05% bromophenol blue). Reaction products were separated using denaturing 20% PAGE, and then were imaged as described above and quantified by using ImageQuaNT software (GE Healthcare Life Sciences).

Cleavage site mapping

Reaction for TS-1 ribozyme cleavage site mapping contained ~100 nM 5' ³²P-labeled substrate RNA and ~500 nM enzyme RNA, and was incubated at 23°C for 30 min in the presence of 5 mM MgCl₂ under otherwise standard conditions. To prepare the RNA marker lanes, the radiolabeled substrate was partially digested with RNase T1 nuclease [25 mM sodium citrate (pH 5.0 at 23°C), 4 M urea, 0.6 mM EDTA, and 0.2 U per L RNase T1 (from *Aspergillus oryzae*; Boehringer Mannheim) for 11 min at 55°C] or with alkali [50 mM Na₂CO₃ (pH 9.0 at 23°C) and 1 mM EDTA for 7 min at 90°C]. Samples were mixed with equal volumes of a urea-containing gel loading buffer and analyzed by denaturing 20% PAGE.

Similarly, the bimolecular pistol ribozyme construct was based on a representative from *A. putredinis* (reverse complement of nucleotides 466281 to 466361 in RefSeq accession NZ_ABFK02000017.1). The enzyme strand was obtained by *in vitro* transcription from synthetic DNA oligonucleotides (Sigma-Aldrich) that were made double stranded as described above. The substrate strand RNA was chemically synthesized (Sigma-Aldrich), 5' ³²P-labeled and purified also as described above. The same methods were used to generate marker lanes and to conduct the ribozyme assay as noted above.

Likewise, the bimolecular hatchet ribozyme construct was based on a representative from unclassified environmental sequences (reverse complement of nucleotides 23144 to 23308

in the previously mentioned human gut metagenome accession scaffold115765_3_MH0070). Both the enzyme and substrate strands were chemically synthesized (Sigma-Aldrich). The substrate strand RNA was 5' ³²P-labeled and purified also as described above. The same methods were used to generate marker lanes and to conduct the ribozyme assay as noted above.

Mass spectrum analysis of ribozyme cleavage products

A 50 μ L reaction containing 30 mM HEPES (pH 7.5 at 23°C), 100 mM KCl, 5 mM MgCl₂, 4 μ M substrate and 4 μ M enzyme RNAs (the bimolecular TS-1 construct, Fig. 2A) was incubated for 60 min at 23°C. Following precipitation with ethanol and sedimentation by centrifugation, the reaction products were resuspended in 20 μ L deionized H₂O and subjected to monoisotopic (exact-mass) spectrometry (Novatia LLC).

Measurements of observed rate constants (k_{obs})

Ribozyme cleavage assays for determining k_{obs} values were performed using the bimolecular TS-1 construct depicted in Fig. 2a. It is not known if the k_{obs} values reflect only the chemical step of the RNA cleavage reaction, or if there are slower structural transitions that limit the rate constants observed. All k_{obs} values were established under standard reaction conditions (see above), except that Mg²⁺ concentrations and pH conditions were varied as noted for each experiment. Cleavage reactions were terminated using a stop solution containing 90% formamide, 50 mM EDTA, 0.05% xylene cyanol and 0.05% bromophenol blue that was added in a volume equal to the ribozyme reaction mixture. The fraction of 5' ³²P-labeled substrate RNA cleaved over time was quantified after separation by denaturing PAGE as described above. Apparent first order rate constants were determined by nonlinear curve fitting using GraphPad Prism⁸ (GraphPad Software).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank A. Roth and other members of the Breaker laboratory for helpful discussions, and N. Carriero and R. Bjornson for assistance with the Yale Life Sciences High Performance Computing Center funded by the NIH (RR19895-02). P.B.K was supported by an NIH Genetics Training Grant (5T32GM007499). C. E. L. was supported by the Deutsche Forschungsgemeinschaft (LU1889/1-1). This work was also supported by an NIH grant (GM022778) to R.R.B and by the Howard Hughes Medical Institute.

References

1. Benner SA, Ellington AD, Tauer A. Modern metabolism as a palimpsest of the RNA world. *Proc Natl Acad Sci USA*. 1989; 86:7054–7058. [PubMed: 2476811]
2. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. *Science*. 2000; 289:920–930. [PubMed: 10937990]
3. Lambowitz, AM.; Caprara, MG.; Zimmerly, S.; Perlman, PS. Group I and Group II Ribozymes as RNPs: Clues to the Past and Guides to the Future. In: Gesteland, RF.; Cech, TR.; Atkins, JF., editors. *The RNA World*. 2. Vol. 37. Cold Spring Harbor Laboratory Press; Cold Spring Harbor: 1999. p. 451-483.
4. Ellis JC, Brown JW. The RNase P family. *RNA Biol*. 2009; 6:362–369. [PubMed: 19738420]

5. Ferre-D'Amare AR, Scott WG. Small self-cleaving ribozymes. *Cold Spring Harbor Perspect Biol.* 2010; 2:a003574.
6. Perreault J, et al. Identification of hammerhead ribozymes in all domains of life reveals novel structural variations. *PLoS Comput Biol.* 2011; 7:e1002031. [PubMed: 21573207]
7. Webb CHT, Lupták A. HDV-like self-cleaving ribozymes. *RNA Biol.* 2011; 8:719–727. [PubMed: 21734469]
8. Roth A, et al. A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat Chem Biol.* 2014; 10:56–60. [PubMed: 24240507]
9. Hutchins CJ, Rathjen PD, Forster AC, Symons RH. Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res.* 1986; 14:3627–3640. [PubMed: 3714492]
10. Saville BJ, Collins RA. A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria. *Cell.* 1990; 61:685–696. [PubMed: 2160856]
11. Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature.* 2004; 428:281–286. [PubMed: 15029187]
12. Salehi-Ashtiani K, Lupták A, Litovchick A, Szostak JW. A genome-wide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science.* 2006; 313:1788–1792. [PubMed: 16990549]
13. Marchler-Bauer A, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 2011; 39:D225–229. [PubMed: 21109532]
14. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics.* 2010; 11:431. [PubMed: 20718988]
15. Yao Z, Weinberg Z, Ruzzo WL. CMfinder--a covariance model based RNA motif finding algorithm. *Bioinformatics.* 2006; 22:445–452. [PubMed: 16357030]
16. Weinberg Z, et al. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol.* 2010; 11:R31. [PubMed: 20230605]
17. Pace NR, Thomas BC, CRW. Probing RNA Structure Function and History by Comparative Analysis. *Cold Spring Harb Perspect Biol.* 1999; 37:113–141.
18. Michel F, Westhof E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol.* 1990; 216:585–610. [PubMed: 2258934]
19. Weinberg Z, Perreault J, Meyer MM, Breaker RR. Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature.* 2009; 462:656–659. [PubMed: 19956260]
20. Westhof E. The amazing world of bacterial structured RNAs. *Genome Biol.* 2010; 11:108. [PubMed: 20236470]
21. Sanchez-Luque FJ, Lopez MC, Macias F, Alonso C, Thomas MC. Identification of an hepatitis delta virus-like ribozyme at the mRNA 5'-end of the L1Tc retrotransposon from *Trypanosoma cruzi*. *Nucleic Acids Res.* 2011; 39:8065–8077. [PubMed: 21724615]
22. Breaker RR, et al. A common speed limit for RNA-cleaving ribozymes and deoxyribozymes. *RNA.* 2003; 9:949–957. [PubMed: 12869706]
23. Liu Y, Wilson TJ, McPhee SA, Lilley DMJ. Crystal structure and mechanistic investigation of the twister ribozyme. *Nat Chem Biol.* 2014; 10:739–744. [PubMed: 25038788]
24. Eiler D, Wang J, Steitz TA. Structural basis for the fast self-cleavage reaction catalyzed by the twister ribozyme. *Proc Natl Acad Sci USA.* 2014; 111:1414571111
25. Ren A, et al. In-line alignment and Mg²⁺ coordination at the cleavage site of the env22 twister ribozyme. *Nat Commun.* 2014; 5:5534–5543. [PubMed: 25410397]
26. Tang J, Breaker RR. Structural diversity of self-cleaving ribozymes. *Proc Natl Acad Sci USA.* 2000; 97:5784–5789. [PubMed: 10823936]
27. Salehi-Ashtiani K, Szostak JW. In vitro evolution suggests multiple origins for the hammerhead ribozyme. *Nature.* 2001; 414:82–84. [PubMed: 11689947]
28. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 2007; 35
29. Markowitz VM, et al. IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res.* 2012; 40:D123–129. [PubMed: 22086953]

30. Human Microbiome Project C. A framework for human microbiome research. *Nature*. 2012; 486:215–221. [PubMed: 22699610]
31. Meyer F, et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. 2008; 9:386. [PubMed: 18803844]
32. Sun S, et al. Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. *Nucleic Acids Res*. 2011; 39:D546–551. [PubMed: 21045053]
33. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. *Nucleic Acids Res*. 2008; 36:D25–30. [PubMed: 18073190]
34. Noguchi H, Park J, Takagi T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res*. 2006; 34
35. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res*. 2010; 38:e132. [PubMed: 20403810]
36. Baker JL, et al. Widespread genetic switches and toxicity resistance proteins for fluoride. *Science*. 2012; 335:233–5. [PubMed: 22194412]
37. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 2013; 29:2933–2935. [PubMed: 24008419]
38. Burge SW, et al. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res*. 2013; 41:D226–232. [PubMed: 23125362]
39. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997; 25:955–964. [PubMed: 9023104]
40. Bland C, et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics*. 2007; 8:209. [PubMed: 17577412]
41. Huang Y, Gilna P, Li W. Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics*. 2009; 25:1338–1340. [PubMed: 19346323]
42. Gardner PP, Barquist L, Bateman A, Nawrocki EP, Weinberg Z. RNIE: genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res*. 2011; 39:5845–5852. [PubMed: 21478170]
43. Weinberg Z, Breaker RR. R2R--software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics*. 2011; 12:3. [PubMed: 21205310]
44. Ruminski DJ, Webb CH, Riccitelli NJ, Luptak A. Processing and translation initiation of non-long terminal repeat retrotransposons by hepatitis delta virus (HDV)-like self-cleaving ribozymes. *J Biol Chem*. 2011; 286:41286–41295. [PubMed: 21994949]
45. Qin J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010; 464:59–65. [PubMed: 20203603]

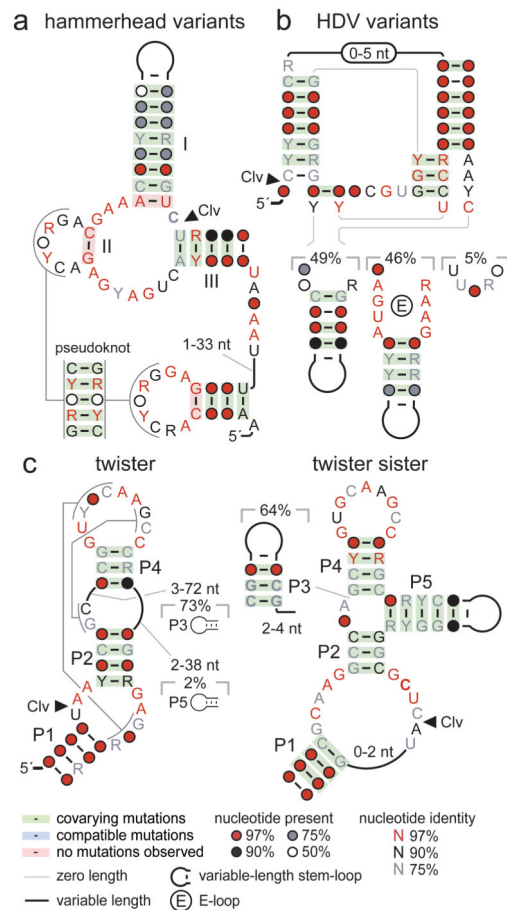


Figure 1. Self-cleaving ribozyme candidates

(a) Consensus sequence and secondary structure model for distinct variants of hammerhead ribozymes. R and Y represent purine and pyrimidine nucleotides, respectively. The site of ribozyme-mediated RNA cleavage (Clv) is identified by arrowhead. **(b)** Consensus sequence and secondary structure model for distinct variants of newly found variants of HDV ribozymes. **(c)** Consensus models for twister and twister sister ribozymes. Noncanonical base pairs and other additional structural interactions for twister ribozymes recently revealed by biophysical studies are not included.

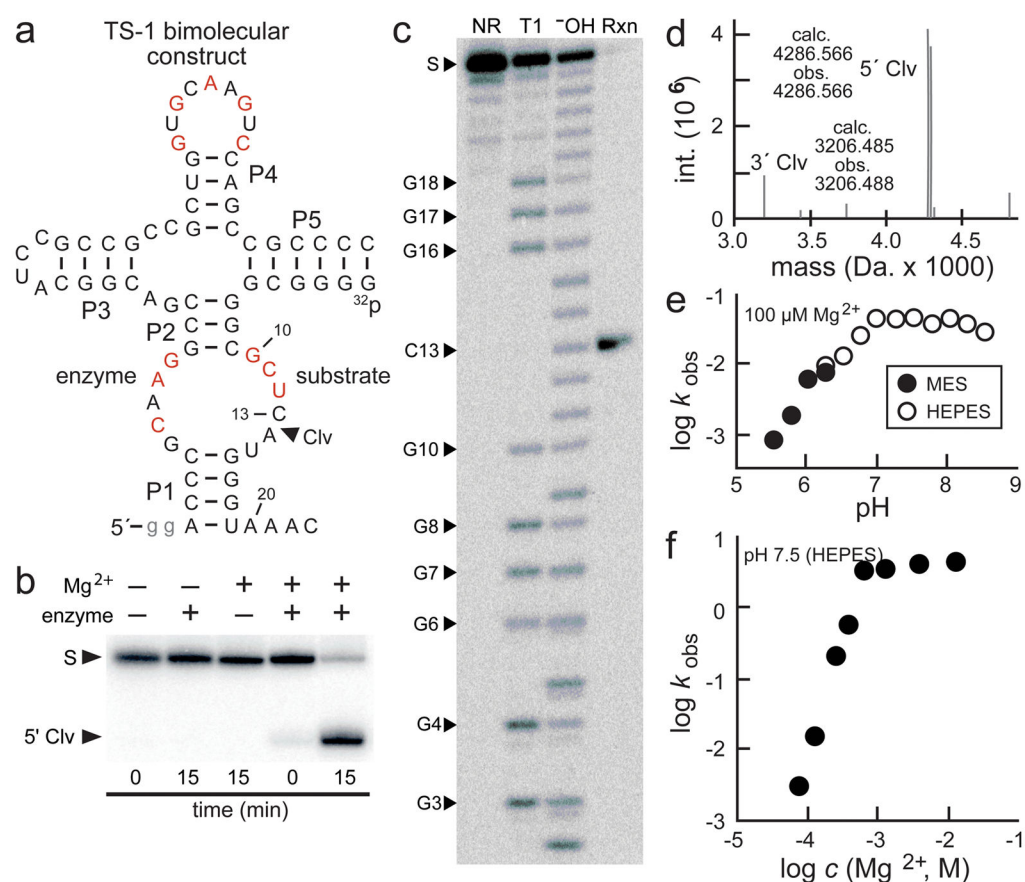


Figure 2. Activity of a bimolecular twister sister ribozyme

(a) Sequence and secondary structure model of a bimolecular construct derived from TS-1 (Supplementary Fig. 1). Highly conserved nucleotides are depicted in red and Clv designates the cleavage site. Lowercase letters identify non-native guanosine residues that were added to facilitate *in vitro* transcription. **(b)** PAGE separation of bimolecular TS-1 ribozyme assay products demonstrating the requirement for the enzyme strand and for divalent metal ions. S designates the 5' ³²P-labeled 23-nucleotide RNA substrate and 5' Clv identifies the cleavage product. Reactions were conducted as described (see **Online Methods**) with variations noted. **(c)** PAGE separation of ribozyme cleavage products. S designates the 5' ³²P-labeled 23-nucleotide RNA substrate, and C13 identifies the 5' ³²P-labeled fragment band produced by incubation with excess unlabeled ribozyme for 30 min (Rxn lane). NR, T1 and ⁻OH lanes designate no reaction, RNase T1 partial digestion (cleaves after G nucleotides), or partial alkaline digestion (cleaves all internucleotides), respectively. **(d)** Mass spectrum analysis of bimolecular TS-1 cleavage reaction products were examined by mass spectrometry (see **Online Methods**). The second largest peak near the 5' Clv annotation is the spontaneously formed opened version (observed mass, 4304.576) of the initial 2',3'-cyclic phosphate product. Intensity is abbreviated int. **(e)** The dependence of twister sister rate constants on pH. **(f)** The dependence of Mg²⁺ concentration on twister sister rate constants. A version of this figure containing full-length gel images is shown in Supplementary Figure 11.

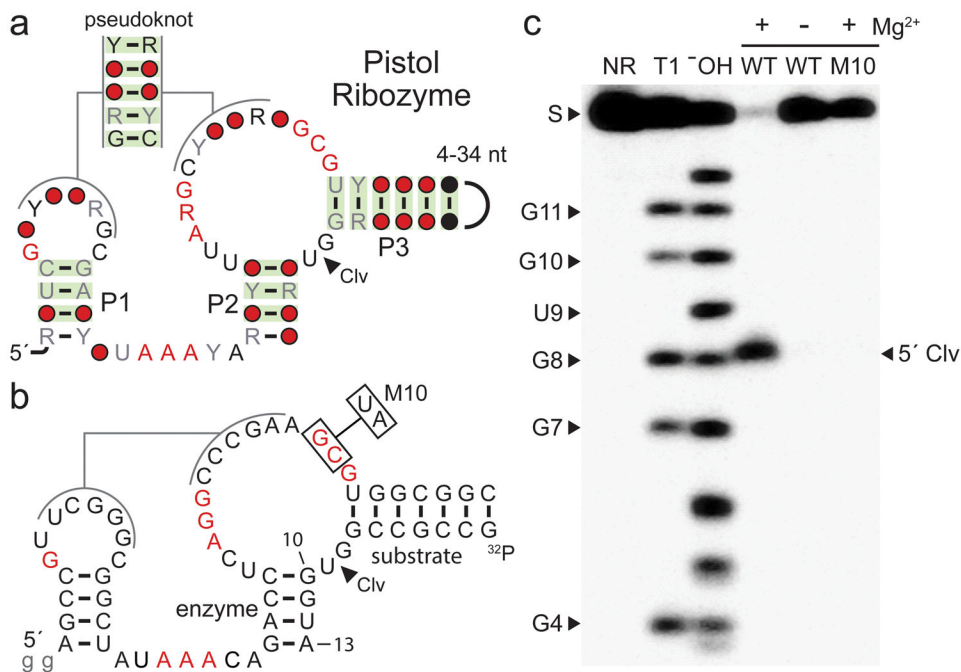


Figure 3. Structure and activity of pistol self-cleaving ribozymes

(a) Consensus sequence and secondary structure model for pistol self-cleaving ribozymes based on 449 unique examples. Annotations are as described in Fig 1a. (b) A bimolecular pistol ribozyme construct based on a representative from the bacterium *Aliistipes putredinis*. Annotations are as described in the legend for Fig. 2a. (c) Pistol ribozyme activity and cleavage site mapping of the *A. putredinis* bimolecular construct wherein the substrate RNA (S) was 5'-labeled with ³²P. Other annotations are as described in the legend to Fig. 2c. Trace amount of substrate was incubated with excess WT or M10 enzyme strand either with (+) or without (-) 20 mM MgCl₂. RNA cleavage products were separated by denaturing 20% PAGE. A version of this figure containing full-length gel images is shown in Supplementary Figure 12.

