



Published in final edited form as:

*Nat Chem Biol.* 2014 January ; 10(1): 56–60. doi:10.1038/nchembio.1386.

## A widespread self-cleaving ribozyme class is revealed by bioinformatics

Adam Roth<sup>1,2,\*</sup>, Zasha Weinberg<sup>1,2,\*</sup>, Andy G. Y. Chen<sup>1</sup>, Peter B. Kim<sup>1</sup>, Tyler D. Ames<sup>1,2,†</sup>, and Ronald R. Breaker<sup>1,2,3,‡</sup>

Adam Roth: adam.roth@yale.edu; Zasha Weinberg: zasha.weinberg@yale.edu; Andy G. Y. Chen: andy.chen@yale.edu; Peter B. Kim: peter.kim@yale.edu; Tyler D. Ames: tames@phosplatin.com

<sup>1</sup>Department of Molecular, Cellular and Developmental Biology, Yale University, Box 208103, New Haven, CT 06520-8103, USA

<sup>2</sup>Howard Hughes Medical Institute, Yale University, Box 208103, New Haven, CT 06520-8103, USA

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

### Abstract

Ribozymes are noncoding RNAs that promote chemical transformations with rate enhancements approaching those of protein enzymes. Although ribozymes are likely to have been abundant during the RNA world era, only ten classes are known to exist among contemporary organisms. We report the discovery and analysis of an additional self-cleaving ribozyme class, called twister, which is present in many species of bacteria and eukarya. Nearly 2700 twister ribozymes were identified that conform to a secondary structure consensus that is small yet complex, with three stems conjoined by internal and terminal loops. Two pseudoknots provide tertiary structure contacts that are critical for catalytic activity. The twister ribozyme motif provides another example of a natural RNA catalyst and calls attention to the potentially varied biological roles of this and other classes of widely distributed self-cleaving RNAs.

---

Contemporary biological catalysis is dominated by protein enzymes, but some classes of natural RNA molecules also possess catalytic activity<sup>1</sup>. Ribozymes are presumed to have been far more abundant during the RNA world era<sup>2</sup>, but their numbers are likely to have diminished drastically during the course of evolution due to competition from protein catalysts. Nonetheless, of the relatively few known natural ribozyme classes, some have

---

Users may view, print, copy, download and 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](http://www.nature.com/authors/editorial_policies/license.html#terms)

‡Corresponding author: Dr. Ronald R. Breaker, Tel: (203) 432-9389, Fax: (203) 432-0753, ronald.breaker@yale.edu.

\*These authors contributed equally to this work.

†Current Address: Phosplatin Therapeutics, 1350 Avenue of the Americas, New York, NY 10019, USA.

### Author contributions

Z.W. and T.D.A. conducted the bioinformatics analyses, A.R., A.G.Y.C. and P.B.K. conducted the biochemical and genetic analyses, R.R.B. prepared the manuscript and all authors interpreted data, designed experiments, and critically reviewed the text and figures.

### Competing financial interests

The authors declare no competing financial interests.

phylogenetically broad distributions and play fundamental roles in protein synthesis and RNA processing. Atomic resolution structures of the large ribosomal subunit have revealed that the peptidyl transfer reaction is RNA-catalyzed<sup>3</sup>, and processing of the 5' ends of tRNAs and certain other RNA substrates is mediated almost universally by the ribonucleoprotein complex Ribonuclease P, which contains a catalytic RNA subunit<sup>4,5</sup>. In addition, there are two widely distributed classes of self-splicing ribozymes<sup>6,7</sup>, and structural and functional similarities between one of these (group II intron) and the RNA core of the spliceosome suggest that eukaryotic splicing is also mediated by RNA catalysis<sup>8</sup>.

There are also five known classes of small ribozymes that catalyze site-specific self-cleavage<sup>9</sup>. Some of these ribozymes were initially discovered as components of satellite RNAs<sup>10–12</sup>, where the processing functions of the catalytic domains are essential for satellite replication via rolling circle mechanisms. More recent searches have revealed that some of these self-cleaving RNAs are distributed far more broadly, however. The hammerhead<sup>13–16</sup> and hepatitis delta virus (HDV)<sup>17</sup> ribozyme classes have been identified in multiple domains of life, where they might function in processes such as retrotransposition and genetic control<sup>18–22</sup>. The *glmS* ribozyme, which is a specialized self-cleaving RNA that serves as a genetic control element<sup>23</sup>, occurs commonly in Gram-positive bacteria<sup>24</sup>. Typically residing in the 5' UTR of the mRNA encoding glucosamine 6-phosphate synthase, the *glmS* ribozyme employs a metabolic product of this synthase, glucosamine 6-phosphate, as a coenzyme in the self-cleavage reaction<sup>25–27</sup>, and thus operates as a component of a negative feedback loop.

Given the broad distribution of natural self-cleaving ribozymes, and given that many artificial classes have been identified using *in vitro* selection methods<sup>28,29</sup>, it is likely that additional natural self-cleaving ribozyme classes are yet to be discovered. We used bioinformatics to identify the twister motif, a doubly pseudoknotted RNA structure that represents the sixth natural class of small self-cleaving ribozymes. Like hammerhead and HDV ribozymes, twister RNAs occur in diverse eukaryotic and bacterial species, suggesting potentially varied biological roles.

## RESULTS

### Identification of a widespread structured RNA motif

Using a bioinformatics pipeline<sup>30</sup>, we identified a small, highly conserved RNA motif (Fig. 1) that occurs commonly in the bacterial class *Clostridia* as well as in diverse eukaryotic species (Supplementary Results, Supplementary Table 1, Supplementary Data Set 1). We termed this novel motif “twister” because its conserved secondary structure resembles the ancient Egyptian hieroglyph “twisted flax”. Although twister RNAs are comparable in size and structural complexity to riboswitches<sup>31</sup>, their genetic contexts often bear striking similarities to those of hammerhead ribozymes (Fig. 2), which suggested that twister RNAs and hammerhead ribozymes could be functionally interchangeable. For example, individual hammerhead ribozyme sequences frequently occur in proximity to one another as well as to members of the twister class of RNAs. Also, many examples of the twister motif are circularly permuted, with the termini for a given representative contained within stem P1 (Fig. 1a, type P1), stem P3 (Fig. 1b, type P3), or stem P5 (Fig. 1c, type P5). Permutations of

this type are common among hammerhead ribozymes but, curiously, have never been observed for any of the approximately two dozen known riboswitch classes. Moreover, like hammerhead ribozymes<sup>13–16</sup>, twister RNAs are present in some eukaryotes, often in large numbers.

### Twister RNAs self-cleave *in vitro*

Given these observations, we speculated that twister RNAs represent a previously undiscovered class of self-cleaving ribozymes. This hypothesis was examined by conducting *in vitro* transcriptions using wild-type (WT) or mutant DNA templates corresponding to twister representatives from four eukaryotic and bacterial species. Indeed, all four WT twister RNAs appeared to undergo quantitative self-cleavage during transcription *in vitro* (Supplementary Fig. 1). In contrast, DNA templates with mutations in highly conserved nucleotides yielded products that matched the sizes expected for uncleaved RNAs. These results indicated that at least some of the conserved nucleotides among twister RNAs are important for self-cleaving ribozyme function.

To confirm ribozyme activity *in vitro*, we generated bimolecular constructs of various twister representatives and monitored the cleavage of a “substrate” RNA in the presence of an appropriate “enzyme” RNA (Fig. 3, Supplementary Fig. 2). For example, we created a bimolecular ribozyme complex corresponding to a twister RNA from the parasitic wasp *Nasonia vitripennis* by splitting the original contiguous RNA within L3 (Fig. 3a). The resulting substrate domain was rapidly cleaved only in the presence of the enzyme domain and Mg<sup>2+</sup> (Fig. 3b). The reaction yielded a 5' cleavage product with a terminal 2',3'-cyclic phosphate group, and a 3' cleavage product with a 5' hydroxyl group as determined by electrophoretic mobility (Fig. 3c) and mass spectrum analysis (Supplementary Fig. 3).

The ribozyme products are consistent with those expected from cleavage by internal phosphoester transfer due to attack of the 2' oxygen of U5 on the adjacent phosphorus atom, with subsequent departure of the 5' oxygen of A6. This general mechanism (Supplementary Fig. 4) for twister catalysis is identical to that used by members of the other five classes of self-cleaving ribozymes discovered previously<sup>9</sup>. As expected for a ribozyme that uses this mechanism, no cleavage product was detected when the nucleotide corresponding to U5 of the substrate lacked the 2' oxygen nucleophile (Supplementary Fig. 5). Also consistent with this mechanism, ligation products were generated during prolonged incubations of individually purified cleavage fragments, indicating that the reaction is reversible (Supplementary Fig. 6).

### Evidence for twister ribozyme activity *in vivo*

To provide evidence that twister ribozymes self-cleave *in vivo*, we performed reverse transcription and polymerase chain reaction (RT-PCR) amplifications using total RNA from *N. vitripennis* adults (Supplementary Fig. 7a, b). For three of the four twister ribozymes that were analyzed in this organism, we did not detect RT-PCR amplification products using DNA primers that bracketed the entire ribozyme sequence. In contrast, the use of primer pairs targeting sequences 3' to the cleavage site generated amplification products of the expected size for each of the twister representatives. These findings indicate that the selected

twister RNAs are expressed in *N. vitripennis* adults, and are consistent with robust cleavage activity *in vivo* for at least three of these examples. RT-PCR analyses of two twister ribozyme representatives from rice (*Oryza sativa* Japonica) yielded results similar to those obtained with *N. vitripennis* (Supplementary Fig. 7c). Interestingly, the detection of *N. vitripennis* twister 4 RNA in uncleaved form (Supplementary Fig. 7b) suggests that the half-lives of certain twister ribozymes can be extended under some conditions. Twister 4 RNA exhibited robust self-cleavage during transcription *in vitro* (Supplementary Fig. 7d), so the detection of the corresponding uncleaved transcript using RT-PCR appears not simply to reflect intrinsically low activity of this twister representative.

### Biochemical evidence for the secondary structure model

The importance of various sequences and substructures of twister ribozymes was evaluated *in vitro* using mutant versions of a bimolecular construct based on a twister representative from the bacterium *Clostridium bolteae*. Consistent with previous results, mutations of highly conserved nucleotides decreased cleavage activity. Furthermore, disruptions of proposed base-pairing interactions individually within P1, P2, P4, and the two pseudoknots resulted in drastic reductions in catalytic activity (Supplementary Fig. 8). In contrast, twister RNA constructs carrying compensatory mutations designed to restore base-pairing within these substructures recovered at least modest cleavage activity. These findings indicate that the formation of the predicted stems is critical for ribozyme activity, whereas the precise sequences of these substructures are generally less important.

### Effects of pH and metal ions on twister ribozyme activity

The amount of substrate processed by the *N. vitripennis* bimolecular ribozyme construct was approximately 50% in 15 seconds at 23°C, but longer incubation times did not increase this yield without additional thermal denaturation steps. This complex probably forms alternative structures that, barring additional denaturation, cannot refold to adopt the active structure on a short timescale. To avoid this problem, we prepared several bimolecular constructs based on twister sequences from other species and identified RNAs that are less prone to misfolding.

We observed efficient substrate processing (Fig. 4a and Supplementary Fig. 9) using a bimolecular construct corresponding to a type P3 twister representative, which was identified among sequences derived from uncultivated organisms (environmental sequences) (Supplementary Figs. 1a and 5a). The  $k_{\text{obs}}$  value for this construct under standard reaction conditions approached  $10 \text{ min}^{-1}$ , thereby precluding accurate rate constant determinations using manual methods alone. Consequently, in order to obtain reliable measurements over a range of  $\text{Mg}^{2+}$  concentrations, we reduced  $k_{\text{obs}}$  values by employing a suboptimal reaction buffer adjusted to pH 5.5. Rate constants for this twister representative were sharply dependent on  $\text{Mg}^{2+}$  concentration, displaying a slope of  $\sim 4$  on a log-log plot, before approaching a plateau above 1 mM  $\text{Mg}^{2+}$  (Fig. 4b). We used an analogous approach to assess the effects of pH on the rate constant of this ribozyme. Reaction buffers containing sub-physiological concentrations (50  $\mu\text{M}$ ) of  $\text{Mg}^{2+}$  permitted accurate determinations of  $k_{\text{obs}}$  values in the plateau phase above pH  $\sim 6.5$  (Fig. 4b). Below this value, rate constants decreased as pH was reduced, with an approximate slope of two on a log-log plot. If the

effects of pH and  $Mg^{2+}$  concentration on rate enhancements are independent of one another (as has been observed with the hammerhead ribozyme<sup>32</sup>), the  $k_{obs}$  for this twister representative under simulated physiological conditions (pH 7.4, 1 mM  $MgCl_2$ ) is estimated to approach  $\sim 1000 \text{ min}^{-1}$ .

Several observations indicated that neither the non-bridging phosphate oxygen atoms nor the 5' oxyanion leaving group at the cleavage site of twister RNAs form inner-sphere contacts with divalent metal ions. Most of the divalent metal ions that were tested fully supported ribozyme cleavage (Fig. 4c), indicating a lack of cation selectivity. This finding is more consistent with structural rather than catalytic roles for divalent metal ions. Moreover, cobalt hexammine, which mimics fully-hydrated  $Mg^{2+}$ , and elevated concentrations of various monovalent cations also promoted modest activity of the twister ribozyme derived from environmental sequences (Supplementary Fig. 10a, b). Finally, substrate RNAs with phosphorothioate modifications at the cleavage site were cleaved with comparable speed and to a similar extent in the presence of  $Mg^{2+}$  or  $Mn^{2+}$  (Supplementary Fig. 11). Collectively, these data suggest that twister ribozymes do not require the direct participation of divalent metal ions in the cleavage mechanism.

## DISCUSSION

The kinetic characteristics of twister ribozymes can be used to make additional predictions about the manner in which members of this class cleave RNA. There are four general catalytic strategies that can be used by enzymes to promote RNA cleavage by internal phosphoester transfer<sup>33–35</sup>. These strategies are: orienting the reactive atoms for in-line nucleophilic attack ( $\alpha$ ), neutralizing the negative charge on a non-bridging oxygen of the cleavage site phosphate ( $\beta$ ), deprotonating the 2' oxygen nucleophile ( $\gamma$ ), and neutralizing the developing negative charge on the 5' oxygen leaving group ( $\delta$ ) (Supplementary Fig. 4).

Twister ribozymes must employ  $\alpha$  catalysis to orient the labile linkage for in-line nucleophilic attack. Also, the rate constants increase sharply between pH 5 and pH 6.3, and one component of this pH dependence is likely due to  $\gamma$  catalysis, which would ostensibly be optimized under physiological (neutral) conditions. The majority of RNA-cleaving ribozymes and deoxyribozymes created by using directed evolution appear to use a combination of  $\alpha$  and  $\gamma$  catalysis only<sup>34</sup>. However, enzymes employing just these two strategies cannot generate a maximal  $k_{obs}$  of more than  $\sim 2 \text{ min}^{-1}$  under the standard reaction conditions used in this study, which is far below what is observed for twister ribozymes.

Thus, twister ribozymes are predicted either to exploit more than just  $\alpha$  and  $\gamma$  catalysis, or to employ a different combination of two or more of the four possible catalytic strategies to attain such high catalytic speeds. Regardless of the precise catalytic strategies used by twister ribozymes, it is clear that all six classes of natural self-cleaving RNAs exceed the  $\alpha\gamma$  speed limit for phosphoester transfer<sup>33–35</sup>. This strongly suggests that rapid strand cleavage is required for most ribozyme contexts, and that this selection pressure has resulted in ribozyme catalytic cores capable of exploiting more than just in-line orientation and 2' hydroxyl deprotonation.

Of the six known classes (including twister) of self-cleaving ribozymes, three appear to have clear and consistent biological functions. The metabolite-dependent *glmS* ribozyme participates in genetic control by serving as the key component in a negative feedback loop<sup>23</sup>. There are hundreds of known representatives of the *glmS* ribozyme class in bacteria, typically residing in the 5' untranslated regions of genes required for glucosamine 6-phosphate production<sup>24</sup>. Self-cleavage of these ribozymes is triggered by direct interactions with glucosamine 6-phosphate, resulting in decreased expression of the downstream genes<sup>23,36</sup>. The hairpin ribozyme class, of which only three examples are known<sup>37</sup>, appears to be important for processing the multimeric genomes of viral satellite RNAs. Similarly, the single known example of the *Neurospora* VS ribozyme class is presumed to be important for processing the abundant Varkud satellite transcript present in the mitochondria of this fungus<sup>38</sup>.

In contrast, there are many thousands of examples of hammerhead, HDV, and twister ribozymes located in many different genetic contexts and in organisms from different domains of life. This highlights a key question: what are the biological functions of these abundant self-cleaving ribozymes? Some hammerhead and HDV ribozymes have been shown or predicted to be important for processing selfish RNAs<sup>18,39-41</sup>, and these representatives are likely to cleave immediately upon transcription. However, additional biological roles have been proposed for self-cleaving ribozymes<sup>15,42</sup>. It has also been suggested that ribozyme-mediated regulation of gene expression could be effected through factor-dependent self-cleavage<sup>19-22</sup>, akin to the metabolite-dependent control of an allosteric self-splicing RNA<sup>43</sup>. If so, there could be control mechanisms and regulatory signals for self-cleaving ribozymes embedded within the noncoding regions of many transcripts.

## ONLINE METHODS

### Bioinformatic analysis of RNA sequences

We identified the type-P1 twister ribozyme structure using a previously established protocol<sup>44</sup> that was applied to Clostridiales bacteria present in RefSeq<sup>45</sup> version 44. The genes nearby to these RNAs seemed reminiscent to us of hammerhead-associated genes from a previous study<sup>15</sup>. We hypothesized that the RNA might function as a self-cleaving ribozyme and that permuted forms also might exist. We detected the permuted forms by manually permuting our multiple-sequence alignment of type-P1 twister ribozymes to the anticipated structure of type-P3 and type-P5 forms and using these to conduct homology searches.

In addition to previously described homology search strategies<sup>46</sup> that were applied to all three permuted forms of twister ribozymes, we searched manually defined patterns using DARN!<sup>47</sup>. We sought homologs of twister in all sequences in RefSeq<sup>45</sup> version 41, the “microbial” (bacterial and archaeal) subset of RefSeq version 44, the environmental sequences used previously<sup>46</sup>, and additional environmental data from IMG/M<sup>48</sup> and assembled Human Microbiome Project sequences<sup>49</sup>. These databases contain roughly 10<sup>11</sup> base pairs of genomic or metagenomic sequences.

To identify possible hairpins located upstream of twister ribozymes (P0 in Fig. 1a), we ran RNAfold<sup>50</sup> on the 30 nucleotides 5' to each type P1 ribozyme, and selected the 3'-most hairpin (if any). We anticipate that many of these predicted hairpins might not be biologically relevant, since even random sequences can commonly form hairpins. Therefore stems that were weak, that contained relatively large terminal loops (more than 8 nucleotides) or that were far from the P1 stem of the ribozyme (more than 4 nucleotides) were generally discarded, except in some circumstances when their terminal loops conformed to the stable tetraloop patterns GNRA, UNCG or CUUG. There were virtually no good candidates for P0 hairpins among eukaryotic twister sequences, suggesting that twister ribozymes in eukaryotes lack this element. In view of this observation, we never predicted a P0 hairpin in eukaryotic twister sequences.

The consensus features of twister ribozymes presented in Fig. 1a were calculated based solely on type-P1 twister ribozymes, as depicted in the figure. Nucleotide frequencies and covariation, as well as a semi-automated drawing were the output of the R2R computer program<sup>51</sup>.

### Analysis of genes associated with self-cleaving ribozymes

Predictions of protein-coding genes were obtained as described previously<sup>44</sup>, with more recently obtained sequences annotated using metagenemark<sup>52</sup>. Conserved domains in the predicted proteins were predicted using the Conserved Domain Database<sup>53</sup> version 2.25 as described previously<sup>46</sup>.

We then defined an initial set of genes near each hammerhead<sup>9</sup>, HDV<sup>9</sup> and twister ribozyme. These are all of the self-cleaving ribozymes currently known to be present in bacteria, except for *glmS* ribozymes<sup>9</sup>. *glmS* ribozymes were not analyzed because their gene associations are well understood and dissimilar to those of the other self-cleaving ribozymes. Bacterial hammerhead ribozyme locations were obtained by using alignments from a previous study<sup>15</sup> as queries for the cmsearch program<sup>54</sup>. HDV ribozymes were predicted using a bacterial HDV ribozyme<sup>17</sup> in *Faecalibacterium prausnitzii* as an initial query for manually iterated searches using cmsearch.

Genes were collected that are located within 6 kb of a self-cleaving ribozyme and that are at least 200 bp from the edge of a sequence fragment (Fig. 2). The latter criterion was imposed because many metagenome sequences are fragmented due to limited coverage, and truncated protein-coding genes near fragment edges could complicate the analysis.

The occurrence frequencies reported might be underestimated because genetic elements with nearby truncations in a metagenomic contig were counted, but such truncations might eliminate some ribozyme sequences that are present in the genome. Although bacterial HDV ribozymes are often located nearby to other HDV ribozymes, they otherwise exhibit gene associations that are different from the other ribozyme classes examined. Therefore, HDV ribozymes are not included in our comparison (Fig. 2). The analysis of permuted forms other than twister type P1 and hammerhead type II ribozymes was complicated by the fact that fewer representatives are found in bacteria. Therefore, such permuted forms were not included in our comparison.

For the vertical axis (Fig. 2), we use the fraction of genetic elements residing within 6 kb of the twister or hammerhead RNAs to avoid including very common domains (such as cd00156 with 686,759 representatives). Sequences corresponding to such domains are likely to reside near those corresponding to one of these RNAs simply due to the abundance of the conserved domains, rather than to a functional association with the RNAs.

The protein domains whose positions are most highly associated with the designated twister or hammerhead RNAs represent 12 of 45,719 conserved domains in the database used. As a means to gauge the relative abundances of these domains, we note that the least- and most-common of these 12 associated domains occur 705 and 4,139 times in proteins, respectively, and 7,108 domains of the 45,719 occur between 705 and 4,139 times.

Note that genetic elements commonly associated with twister RNAs are hammerhead ribozymes, and *vice versa*. (Further, hammerhead ribozyme sequences sometimes occur multiply and in proximity to one another, and this is also true for twister ribozymes.) The frequent associations of twister and hammerhead RNAs suggest that the functions of these two RNA classes are interchangeable.

### Assays of self-cleavage during transcription

Double-stranded DNA templates encoding twister RNAs from *O. sativa* [NC\_008404.1 (sequence accession number)/1825405-1825354 (coordinates, 5' to 3')], *S. mansoni* (NW\_003038506.1/256125-256056), *R. intestinalis* (NZ\_ABYJ02000020.1/1803-1854), and environmental sequence (ASMM170b\_GJFD58A01ENAB6/114-41; this sequence was obtained from IMG/M<sup>48</sup> under taxonid 2100351011) were generated by extending the synthetic oligodeoxynucleotide 5'-TAATACGACTCACTATAGG (containing the T7 promoter) on the appropriate antisense DNA, or by simultaneously extending partially complementary oligodeoxynucleotides (Sigma-Aldrich and Integrated DNA Technologies) using SuperScript II reverse transcriptase (Invitrogen) according to the manufacturer's instructions. Transcriptions *in vitro* were performed as previously described<sup>55</sup>. Internally <sup>32</sup>P-labeled products were then 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

Bimolecular twister ribozyme constructs were based on type P1 representatives from *Nasonia vitripennis* (NW\_001815459.1/1079284-1079348), *Nematostella vectensis* (NW\_001834173.1/63183-63244), *Oryza sativa* (NC\_008404.1/26415927-26416006), and *Clostridium bolteae* (NZ\_ABCC02000048.1/9855-9790), and on a type P3 example from environmental sequence (coordinates provided above). Substrate and enzyme strands of the bimolecular complexes were individually synthesized, resulting in the elimination of L3 sequences from parent ribozymes of type P1, and in the elimination of the L1 sequence from the type P3 parent.

Substrate RNAs, including those containing deoxyuridine and phosphorothioate modifications, were purchased from Integrated DNA Technologies and 5'-end-labeled with

$\gamma$ -<sup>32</sup>P [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 transcriptions *in vitro*, also as above. Before being added to cleavage reactions, all RNAs were purified by denaturing PAGE, eluted from gel slices in 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 and contained 30 mM HEPES (pH 7.5 at 23°C), 100 mM KCl, 20 mM MgCl<sub>2</sub>, 5 nM 5' <sup>32</sup>P-labeled substrate and 100 nM unlabeled enzyme RNA, unless otherwise indicated. Substrate and enzyme RNAs were combined, heated to 80°C for 1 min, and cooled to room temperature before cleavage reactions were initiated with MgCl<sub>2</sub>. 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 detected as above.

### Ligation assay

The twister ribozyme derived from environmental sequence was transcribed *in vitro* as described above. The internally <sup>32</sup>P-labeled cleavage products generated during this step were individually purified by denaturing PAGE, eluted and concentrated also as described above. Ligation reactions contained the 5' and 3' cleavage products (500 nM and 2.5 μM, respectively) and 5 mM MgCl<sub>2</sub>, but were otherwise prepared exactly as the bimolecular twister ribozyme cleavage assays described above. Reaction products were separated using denaturing 10% PAGE and detected and quantified as described elsewhere in Online Methods.

### Cleavage site mapping

The cleavage reaction contained ~100 nM 5' <sup>32</sup>P-labeled *N. vitripennis* substrate RNA and ~500 nM enzyme RNA, and was incubated at 23°C for 30 min in the presence of 5 mM MgCl<sub>2</sub> under otherwise standard conditions. Subsequently, aliquots of the reaction products were incubated for 3 h at 23°C under neutral conditions or in the presence of 25 mM HCl. The acid-treated sample was then neutralized with NaOH, while the no-acid control received the same molar quantity of NaCl. To prepare RNA ladders, the radiolabeled substrate was partially digested in the presence of 25 mM sodium citrate (pH 5.0 at 23°C), 4 M urea, 0.6 mM EDTA, and 0.2 U/μL RNase T1 (from *Aspergillus oryzae*; Boehringer Mannheim) for 11 min at 55°C, or in the presence of 50 mM Na<sub>2</sub>CO<sub>3</sub> (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 as above.

### Mass spectrum analysis of ribozyme cleavage products

A 50 μL reaction containing 20 mM HEPES (pH 7.5 at 23°C), 100 mM NaCl, 20 mM MgCl<sub>2</sub>, and substrate and enzyme RNAs at concentrations of 3 μM and 4 μM, respectively (corresponding to the *C. bolteae* bimolecular construct; Supplementary Fig. 8a), was incubated for 1 min at 37°C. Following precipitation with ethanol and centrifugation, the reaction products were resuspended in 10 μL water and subjected to monoisotopic (exact mass) spectrometry (Novatia LLC).

### Rate constant ( $k_{\text{obs}}$ ) measurements

Excepting variations in  $\text{Mg}^{2+}$  concentration and pH, reactions were performed under standard conditions using the bimolecular twister construct derived from environmental sequence. Cleavage reactions for the  $\text{Mg}^{2+}$  profile were terminated with 1.5 volumes of a solution containing 75% formamide, 125 mM EDTA, 0.05% xylene cyanol and 0.05% bromophenol blue. Cleavage levels were quantified using ImageQuant software (Molecular Dynamics). Apparent first order rate constants were determined by nonlinear curve fitting using MATLAB (MathWorks), with the equation  $f_c = (f_\infty - f_0) * e^{-k_{\text{obs}}t} + f_\infty$ , where  $f_c$ ,  $f_\infty$  and  $f_0$  are the fractions cleaved at time = t, upon exhaustive incubation, and at t = 0, respectively, and  $k_{\text{obs}}$  is the apparent first order rate constant.

### RT-PCR amplification

*N. vitripennis* pupae were obtained from Ward's Natural Science. Newly emerged *N. vitripennis* adults and leaves from *O. sativa* Japonica seedlings were pulverized under liquid nitrogen with a mortar and pestle, and genomic DNA and total RNA were isolated using TRIzol (Ambion) according to the manufacturer's instructions. RNA samples were treated with RQ1 DNase (Promega), and reactions were either terminated with RQ1 DNase Stop Solution as suggested by the manufacturer or extracted successively with phenol/chloroform (1:1) and chloroform/isoamyl alcohol (24:1) and concentrated by ethanol precipitation. Following a 1 min denaturation step at 80°C, 1.5 µg of DNase-treated RNA and 75 pmol of random hexamer primers (Applied Biosystems) were added to 25 µL reactions containing SuperScript II reverse transcriptase (Invitrogen) or to mock reactions that lacked reverse transcriptase, and these reactions were incubated (under conditions prescribed by the manufacturer) successively at room temperature for 15 min, 37°C for 15 min, and 42°C for 1 h. Subsequently, the genomic DNA preparations and the products of the reverse transcription reactions were used as templates in PCR amplifications with gene-specific primers (Supplementary Fig. 7e), using either standard or touchdown<sup>56</sup> protocols. PCR products were analyzed on 3% agarose gels and stained with ethidium bromide.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

We thank M. Murdock for performing the twister ribozyme ligation assay, P. McCown and other members of the Breaker laboratory for helpful discussions, R. Knight for alerting us to useful metagenome datasets, S. L. Tausta and T. Nelson for providing us with rice seeds, and N. Carriero and R. Bjornson for assistance with the Yale Life Sciences High Performance Computing Center (NIH grant RR19895-02). This work was supported by NIH (GM022778) and by the Howard Hughes Medical Institute.

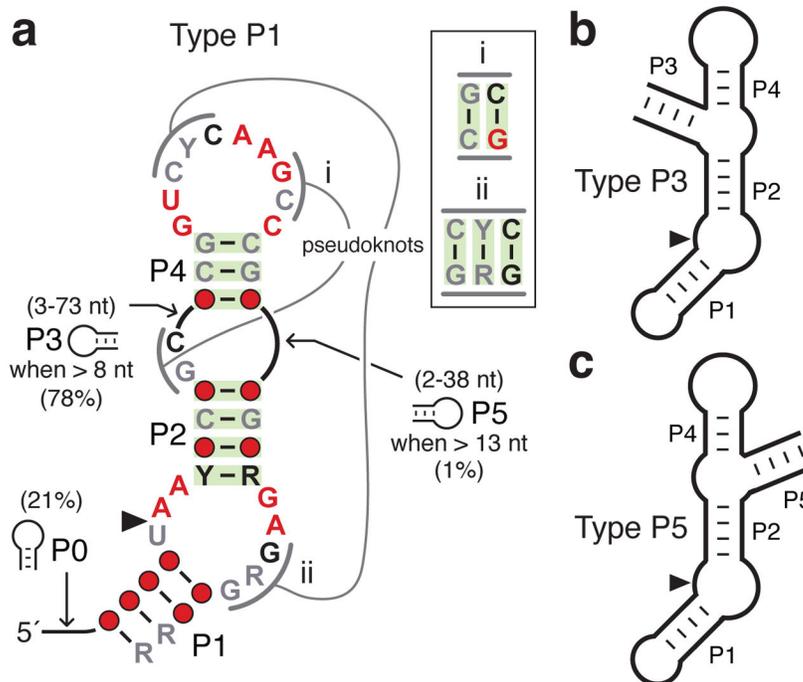
### References

1. Cech TR. Ribozymes, the first 20 years. *Biochem Soc Trans.* 2002; 30:1162–1166. [PubMed: 12440996]
2. 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]

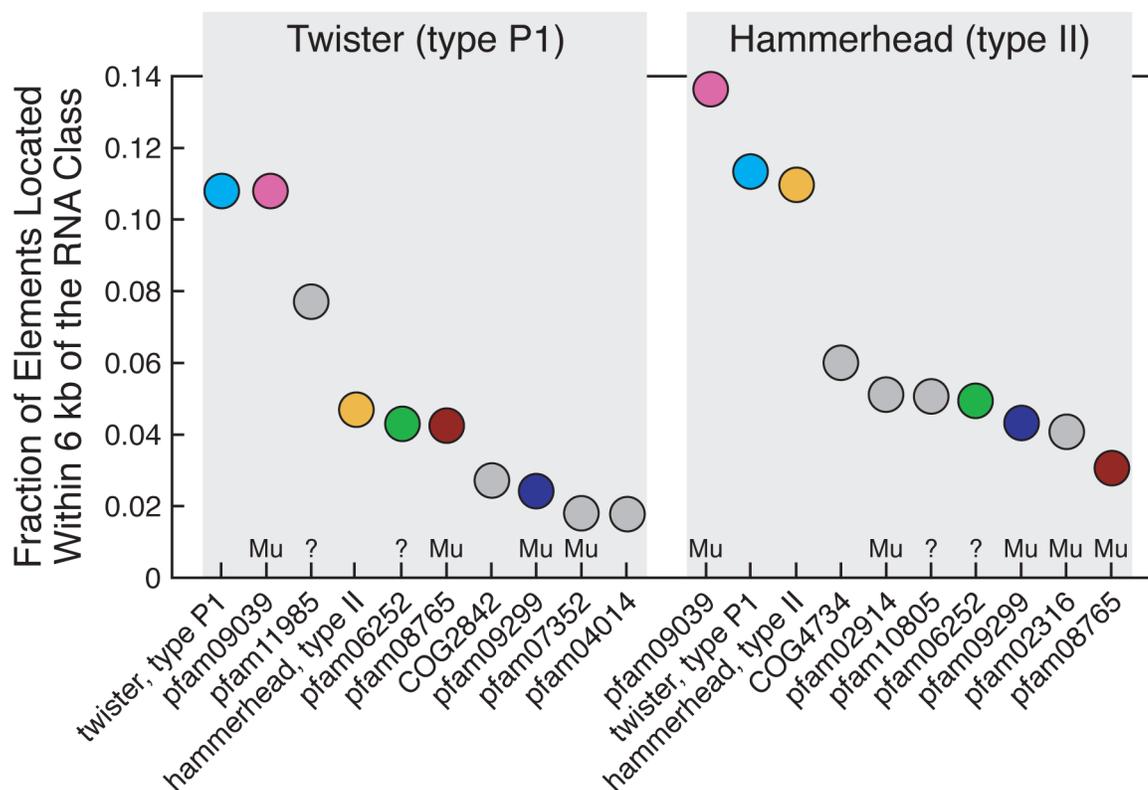
3. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. *Science*. 2000; 289:920–930. [PubMed: 10937990]
4. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. The RNA moiety of Ribonuclease P is the catalytic subunit of the enzyme. *Cell*. 1983; 35:849–857. [PubMed: 6197186]
5. Mondragón A. Structural studies of RNase P. *Annu Rev Biophys*. 2013; 42:537–557. [PubMed: 23654306]
6. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell*. 1982; 31:147–157. [PubMed: 6297745]
7. Peebles CL, Perlman PS, Mecklenburg KL, Petrillo ML, Tabor JH, Jarrell KA, Cheng HL. A self-splicing RNA excises an intron lariat. *Cell*. 1986; 44:213–223. [PubMed: 3510741]
8. Lambowitz AM, Zimmerly S. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol*. 2011; 3:a003616. [PubMed: 20463000]
9. Ferré-D'Amaré AR, Scott WG. Small self-cleaving ribozymes. *Cold Spring Harb Perspect Biol*. 2010; 2:a003574. [PubMed: 20843979]
10. Prody GA, Bakos JT, Buzayan JM, Schneider IR, Bruening G. Autolytic processing of dimeric plant virus satellite RNA. *Science*. 1986; 231:1577–1580. [PubMed: 17833317]
11. Buzayan JM, Gerlach WL, Bruening G. Non-enzymatic cleavage and ligation of RNAs complementary to a plant virus satellite RNA. *Nature*. 1986; 323:349–353.
12. Sharmeen L, Kuo MY, Dinter-Gottlieb G, Taylor J. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J Virol*. 1988; 62:2674–2679. [PubMed: 2455816]
13. De la Peña M, Garcia-Robles I. Ubiquitous presence of the hammerhead ribozyme motif along the tree of life. *RNA*. 2011; 16:1943–1950. [PubMed: 20705646]
14. Jimenez RM, Delwart E, Luptak A. Structure-based search reveals hammerhead ribozymes in the human microbiome. *J Biol Chem*. 2011; 286:7737–7743. [PubMed: 21257745]
15. Perreault J, Weinberg Z, Roth A, Popescu O, Chartrand P, Ferbeyre G, Breaker RR. Identification of hammerhead ribozymes in all domains of life reveals novel structural variations. *PLoS Comput Biol*. 2011; 7:e1002031. [PubMed: 21573207]
16. Seehafer C, Kalweit A, Steger G, Gräf S, Hammann C. From alpaca to zebrafish: Hammerhead ribozymes wherever you look. *RNA*. 2011; 17:21–26. [PubMed: 21081661]
17. Webb CH, Riccitelli NJ, Ruminski DJ, Luptak A. Widespread occurrence of self-cleaving ribozymes. *Science*. 2009; 326:953. [PubMed: 19965505]
18. Webb CHT, Lupták A. HDV-like self-cleaving ribozymes. *RNA Biol*. 2011; 8:719–727. [PubMed: 21734469]
19. Hammann C, Luptak A, Perreault J, De La Peña M. The ubiquitous hammerhead ribozyme. *RNA*. 2012; 18:871–885. [PubMed: 22454536]
20. Martick M, Horan LH, Noller HF, Scott WG. A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature*. 2008; 454:899–902. [PubMed: 18615019]
21. Hammann C, Westhof E. The unforeseeable hammerhead ribozyme. *F1000 Biology Reports*. 2009; 1:6. [PubMed: 20948624]
22. 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]
23. 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]
24. McCown PJ, Roth A, Breaker RR. An expanded collection and refined consensus model of *glmS* ribozymes. *RNA*. 2011; 17:728–736. [PubMed: 21367971]
25. McCarthy TJ, Plog MA, Floy SA, Jansen JA, Soukup JK, Soukup GA. Ligand requirements for *glmS* ribozyme self-cleavage. *Chem Biol*. 2005; 12:1221–1226. [PubMed: 16298301]
26. Klein DJ, Ferré-D'Amaré AR. Structural basis of *glmS* ribozyme activation by glucosamine-6-phosphate. *Science*. 2006; 313:1752–1756. [PubMed: 16990543]
27. Cochrane JC, Lipchok SV, Strobel SA. Structural investigation of the *glmS* ribozyme bound to its catalytic cofactor. *Chem Biol*. 2007; 14:97–105. [PubMed: 17196404]

28. Williams KP, Ciafré S, Tocchini-Valentini GP. Selection of novel Mg<sup>2+</sup>-dependent self-cleaving ribozymes. *EMBO J.* 1995; 14:4551–4557. [PubMed: 7556098]
29. Tang J, Breaker RR. Structural diversity of self-cleaving ribozymes. *Proc Natl Acad Sci USA.* 2000; 97:5784–5789. [PubMed: 10823936]
30. Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, Breaker RR. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol.* 2010; 11:R31. [PubMed: 20230605]
31. Breaker RR. Prospects for riboswitch discovery and analysis. *Mol Cell.* 2011; 43:867–879. [PubMed: 21925376]
32. Canny MD, Jucker FM, Kellogg E, Khvorova A, Jayasena SD, Pardi A. Fast cleavage kinetics of a natural hammerhead ribozyme. *J Am Chem Soc.* 2004; 126:10848–10849. [PubMed: 15339162]
33. Emilsson GM, Nakamura S, Roth A, Breaker RR. Ribozyme speed limits. *RNA.* 2003; 9:907–918. [PubMed: 12869701]
34. Breaker RR, Emilsson GM, Lazarev D, Nakamura S, Puskarz IJ, Roth A, Sudarsan N. A common speed limit for RNA-cleaving ribozymes and deoxyribozymes. *RNA.* 2003; 9:949–957. [PubMed: 12869706]
35. Cochrane JC, Strobel SA. Catalytic strategies of self-cleaving ribozymes. *Acc Chem Res.* 2008; 41:1027–1035. [PubMed: 18652494]
36. Collins JA, Irnov I, Baker S, Winkler WC. Mechanism of mRNA destabilization by the *glmS* ribozyme. *Genes Dev.* 2007; 21:3356–3368. [PubMed: 18079181]
37. DeYoung MB, Siwkowski AM, Lian Y, Hampel A. Catalytic properties of hairpin ribozymes derived from Chicory Yellow Mottle Virus and Arabis Mosaic Virus satellite RNAs. *Biochemistry.* 1995; 34:15785–15791. [PubMed: 7495810]
38. 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]
39. Flores R, Gas M-E, Molina-Serrano D, Nohales M-Á, Carbonell A, Gago S, De la Peña M, Daròs J-A. Viroid replication: Rolling-circles, enzymes and ribozymes. *Viruses.* 2009; 1:317–334. [PubMed: 21994552]
40. Eickbush DG, Eickbush TH. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. *Mol Cell Biol.* 2010; 30:3142–3150. [PubMed: 20421411]
41. Sánchez-Luque FJ, López 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]
42. Ruminski DJ, Webb CHT, Riccitelli NJ, Lupták 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]
43. Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR. An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science.* 2010; 329:845–848. [PubMed: 20705859]
44. Weinberg Z, et al. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol.* 2010; 11:R31. [PubMed: 20230605]
45. 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:D61–65. [PubMed: 17130148]
46. Baker JL, et al. Widespread genetic switches and toxicity resistance proteins for fluoride. *Science.* 2012; 335:233–235. [PubMed: 22194412]
47. Zytnicki M, Gaspin C, Schiex T. DARN! A weighted constraint solver for RNA motif localization. *Constraints.* 2008; 13:91–109.
48. Markowitz VM, et al. IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* 2008; 36:D534–538. [PubMed: 17932063]
49. Methé BA, et al. A framework for human microbiome research. *Nature.* 2012; 486:215–221. [PubMed: 22699610]
50. Hofacker IL. RNA secondary structure analysis using the Vienna RNA package. *Current protocols in bioinformatics.* 2009; Chapter 12(Unit 12.2)

51. Weinberg Z, Breaker RR. R2R—software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics*. 2011; 12:3. [PubMed: 21205310]
52. Zhu W, Lomsadze A, Borodovsky M. *Ab initio* gene identification in metagenomic sequences. *Nucleic Acids Res*. 2010; 38:e132. [PubMed: 20403810]
53. 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]
54. Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: inference of RNA alignments. *Bioinformatics*. 2009; 25:1335–1337. [PubMed: 19307242]
55. Roth A, Nahvi A, Lee M, Jona I, Breaker RR. Characteristics of the *glmS* ribozyme suggest only structural roles for divalent metal ions. *RNA*. 2006; 12:607–619. [PubMed: 16484375]
56. Korbie DJ, Mattick JS. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols*. 2008; 3:1452–1456. [PubMed: 18772872]
57. Scott EC, Uhlenbeck OC. A re-investigation of the thio effect at the hammerhead cleavage site. *Nucleic Acids Res*. 1999; 27:479–484. [PubMed: 9862968]

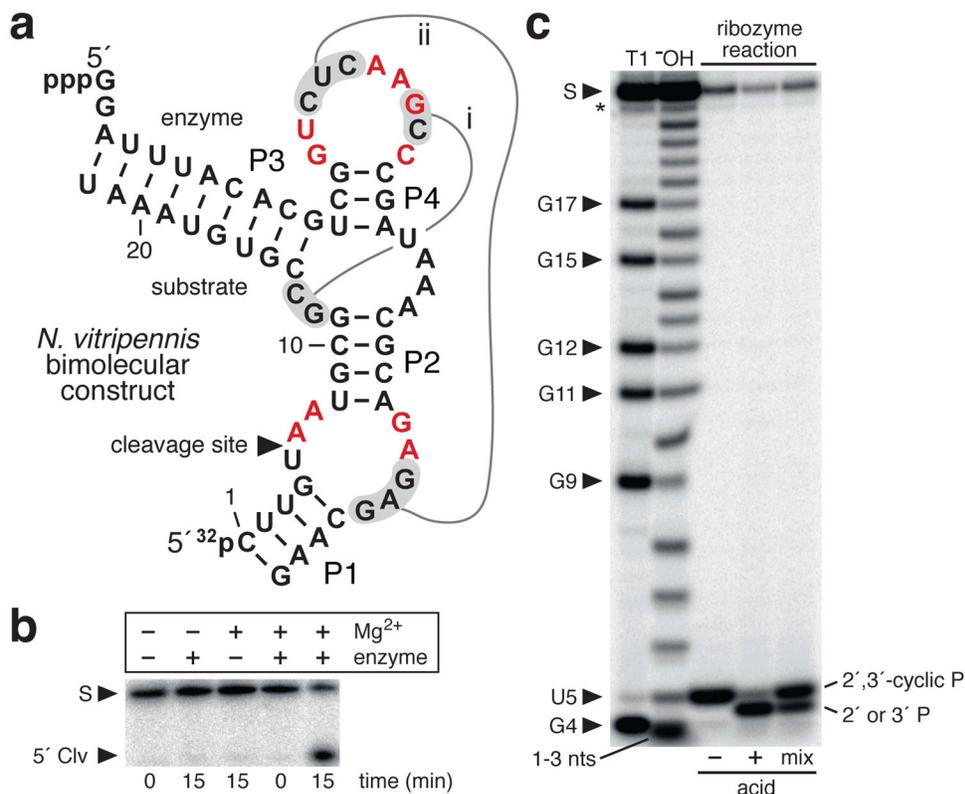


**Figure 1. Consensus sequence and secondary-structure model for twister self-cleaving ribozymes**  
**(a)** Detailed consensus model based on 2690 twister ribozymes depicted in its type P1 configuration, wherein the RNA chain begins and ends at the base of the P1 stem. The arrowhead identifies the cleavage site. Gray, black and red nucleotides designate conservation of at least 75, 90, and 97%, respectively; positions in which nucleotide identity is less conserved are represented by circles. Green shading denotes predicted base pairs supported by natural covariation. Notations *i* and *ii* identify predicted pseudoknots. Numbers in parentheses are the variable lengths of linker sequences that sometimes form stem structures as indicated. R and Y denote purine and pyrimidine, respectively. **(b, c)** The RNA chains of naturally occurring type P3 and type P5 configurations enter and depart the motif at the optional P3 or P5 stems, respectively.

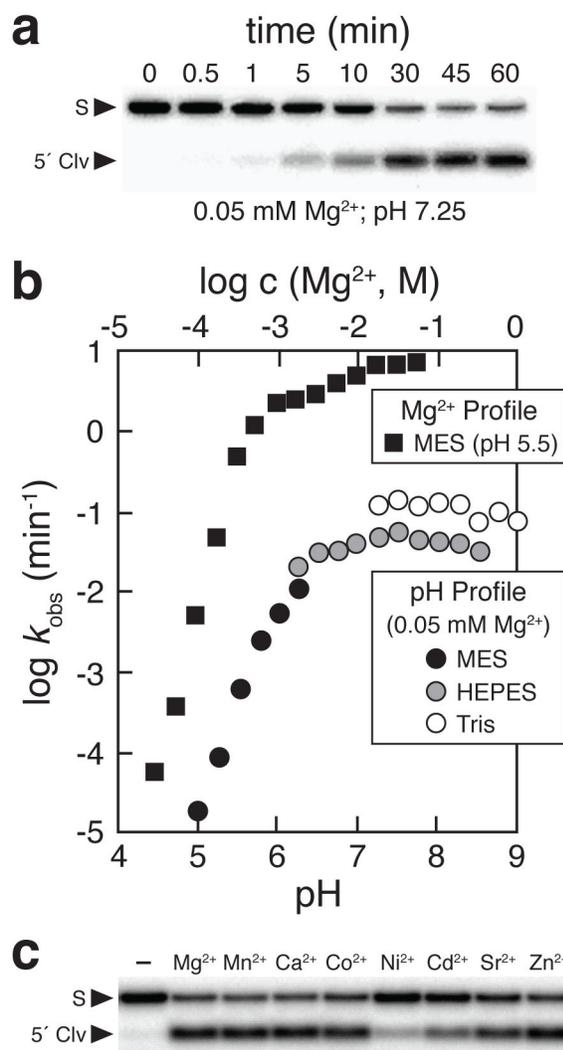


**Figure 2. Common associations between various genetic elements and twister or hammerhead RNAs**

Plot of the ten genetic elements most frequently associated with twister type P1 (left) or hammerhead type II (right) RNAs as ranked by the fraction of occurrences of the element located within 6 kilobases (kb) of these RNAs. Other permuted ribozyme forms were not analyzed (see Supplementary Methods). Conserved protein domains from the Protein Families (pfam) or Clusters of Orthologous Groups (COG) databases with a stated relationship to Mu bacteriophages are designated. Protein domains of unknown function are labeled with a question mark. Genetic elements associated with both ribozyme classes are assigned a color, whereas other elements are depicted in gray.



**Figure 3. Sequence, structure and activity of a twister ribozyme from *N. vitripennis***  
**(a)** Sequence and secondary structure of the bimolecular construct derived from a twister ribozyme from *N. vitripennis*. Red nucleotides correspond to the ten most highly conserved nucleotides of the consensus sequence; other annotations are as described in the legend to Fig. 1a. Non-native guanosine residues were added to the 5' end of the enzyme strand to facilitate transcription *in vitro*. **(b)** Activity of the *N. vitripennis* bimolecular ribozyme construct. Trace amounts of 5' <sup>32</sup>P-labeled 22-nt substrate RNA (S) were incubated for 0 or 15 min in the absence (–) or presence (+) of 20 mM Mg<sup>2+</sup> and the corresponding unlabeled RNA enzyme domain as indicated. 5' <sup>32</sup>P-labeled cleavage product (5' Clv) was separated from the uncleaved substrate by denaturing 20% PAGE. **(c)** For mapping the cleavage site, <sup>32</sup>P-labeled substrate (S) was partially digested with RNase T1 (T1; cleaves after G nucleotides) or with alkali (–OH), or was reacted with unlabeled enzyme RNA, followed by product separation using denaturing 20% PAGE. The products of the ribozyme reaction were loaded on the gel without (–) or with (+) prior acid treatment, or samples of each were combined (mix) to directly compare product mobilities. The asterisk denotes a band with a mobility that would be expected for a side product of RNA synthesis that is shorter by one nucleotide than the full-length substrate; this is not unusual for synthetic RNA preparations. A version of this figure containing full-length gel images is shown in Supplementary Figure 12.



**Figure 4. Kinetic characteristics of a twister ribozyme derived from an environmental DNA sequence**

(a) Representative time course for cleavage of a bimolecular twister ribozyme derived from an environmental sequence (Supplementary Fig. 5a). Other details are as described in the legend for Fig. 3b. (b) Effects of  $Mg^{2+}$  concentrations and pH on twister ribozyme rate constants. (c) Twister ribozyme activity assays in the presence of various divalent metal ions.  $5'$   $^{32}P$ -labeled substrate was incubated with excess enzyme RNA for 1 min in the absence (–) or presence of 1 mM of the divalent metal ions indicated. A version of this figure containing full-length gel images is shown in Supplementary Figure 13.