www.sciencemag.org/cgi/content/full/326/5955/953/DC1

Supporting Online Material for

Widespread Occurrence of Self-Cleaving Ribozymes

Chiu-Ho T. Webb, Nathan J. Riccitelli, Dana J. Ruminski, Andrej Lupták*

*To whom correspondence should be addressed. E-mail: aluptak@uci.edu

Published 13 November 2009, *Science* **326**, 953 (2009)
DOI: 10.1126/science.1178084

This PDF file includes:

Materials and Methods
SOM Text
Figs. S1 to S19
Tables S1 and S2
References

Widespread Occurrence of Self-Cleaving Ribozymes

Chiu-Ho T. Webb,[1+] Nathan J. Riccitelli,[2+] Dana J. Ruminski,[3] Andrej Lupták[1,2,3,*]

[1]Departments of Molecular Biology & Biochemistry,

[2]Chemistry and [3]Pharmaceutical Sciences,

University of California, Irvine, CA 92697 USA.

*To whom correspondence should be addressed. E-mail: aluptak@uci.edu

[+]These authors contributed equally to this work.

On-line MS ALL TEXT: 2500 words	

This PDF file includes

- Abstract
- Online manuscript
- Figs. S1to S19
- Tables S1 and S2
- Materials and Methods
- References

Abstract

Hepatitis delta virus (HDV) and cytoplasimic polyadenylation element binding protein 3 (*CPEB3*) ribozymes form a family of self-cleaving RNAs characterized by a conserved nested double-pseudoknot and minimal sequence conservation. Secondary structure-based searches were used to identify sequences capable of forming this fold and their self-cleavage activity was confirmed in vitro. Active sequences were uncovered in several marine organisms, two nematodes, an arthropod, a bacterium, and an insect virus, often in multiple sequence families and copies. Sequence searches based on identified ribozymes showed that plants, fungi, and a unicellular eukaryote also harbor the ribozymes. In *Anopheles gambiae* the ribozymes were found differentially expressed and self-cleaved at basic developmental stages. Our results indicate that HDV-like ribozymes are abundant in nature and suggest that self-cleaving RNAs may play a variety of biological roles.

Self-cleaving ribozymes are involved in maturation of RNA viroids and virusoids, and satellite DNA transcripts. They include a number of motifs, including the hammerhead, hepatitis delta virus (HDV), hairpin, and the *Neurospora* Varkud satellite (*S1-5*). In addition, the *GlmS* ribozyme has been discovered in many bacteria *(S6)*, hammerhead ribozymes have been identified in plant, and mammalian genomes (*S7, 8*), and an HDV-like ribozyme was found in the second intron of the mammalian *CPEB3* gene (*S9*). The conservation of the *CPEB3* ribozyme among mammals suggests that HDV-like ribozymes might be widely distributed. Because structured functional RNAs have conserved base-pairing regions, but their sequences are not necessarily conserved (*S10, 11*), we hypothesized that a genomic search for sequences capable of forming the HDV and *CPEB3* secondary structure would yield sequences capable of catalysis.

The HDV and *CPEB3* ribozyme core consists of five paired regions forming two coaxially stacked helices that are further organized into a nested double pseudoknot structure linked via two single-stranded regions, J1/2 and J4/2 (fig. S1) (*S9, 12-14*). J4/2, along with L3 and P1.1, forms the core of the ribozyme, containing both the active-site cytosine and a conserved adenosine involved in an A-minor tertiary interaction with the P3 helix (*S12, 13*). In vitro selections have shown that these strands tolerate mutations so long as the secondary structure, the A-minor interaction, and the active-site cytosine of the ribozyme are preserved (*S15, 16*). In total, six nucleotides (nt) are invariant on the sequence level whereas about 60 nt are required to form the minimal structure.

Based on these data, a series of secondary structure descriptors were created for the paired, tertiary, and catalytic elements. Only few positions in the sequence were specified (fig. S1) (*S17*), including the catalytic cytosine residue, a purine-pyrmidine base pair at the base of the P1 helix, and a G•U wobble pair in L3 that forms the base of the P3 helix in the genomic HDV ribozyme (*S12, 13*). The descriptor was used to scan available genomic data for sequences capable of forming the prescribed fold and candidate sequences were tested for in vitro self-scission (*S18*). Self-cleavage activity was initially confirmed in six eukaryotes, one bacterium, and one virus (Fig. 1, table S1, and figs. S1 to S18). Mutation of the proposed active-site cytosine residue vastly diminished in vitro activity, supporting the residue's central role in the cleavage mechanism (figs. S2, S7, S18). All six eukaryotes were found to harbor multiple copies, and in three of them (*Anopheles*

gambiae, *Strongylocentrotus purpuratus* and *Branchiostoma floridae*) the candidate sequences can be grouped into multiple families representing different versions of the HDV and *CPEB3* motif. The identified ribozymes were later used to search partially assembled genomes and GenBank Expressed Sequence Tag database (dbEST), where additional ribozymes were found in several more eukaryotes, including plants, fungi, fish, insects, a tapeworm, and a unicellular ciliate (table S2 and fig. S19).

Two sequence families were identified in the African mosquito *A. gambiae* (drz-Agam-1 and -2; Fig. 1 and figs. S2 to S5) (*S19*). Representative sequences from each family were tested in vitro with self-cleavage rate constants ranging from sec^{-1} to hr^{-1} at ambient temperature and 1 mM Mg^{2+} (table S1). Rapid amplification of cDNA ends (5´ RACE) experiments performed on total RNA extracts from various developmental stages and sexes of *A. gambiae* showed that the sequenced 5´ ends map to the computationally predicted and in vitro verified self-cleavage sites (fig. S2) (*S17*), demonstrating that the ribozymes are expressed and self-cleave in vivo. Reverse transcription followed by quantitative polymerase chain reaction (RT-qPCR) revealed the highest expression of the drz-Agam-1 family to occur in adult male mosquitoes (figs. S2 to S4). However, these ribozymes remain largely uncleaved, whereas in larvae, pupae and adult females, the majority of the ribozymes were isolated in self-cleaved form, indicating that both the expression and the cleavage state of the ribozymes are regulated in vivo. Similar differential expression and self-cleavage were observed for the drz-Agam-2 family of ribozymes (Fig. 1 and fig. S5).

The drz-Agam-2 ribozymes (Fig. 1, figs. S1 to S5) are the largest identified HDV-like ribozymes, containing long sequences in the J1/2 and P4 regions. While a variable P4 has previously been shown to maintain ribozyme activity (*S20*), a long, structured J1/2 strand has not been observed. In contrast, the 63-nt drz-Bflo-2, one of two ribozymes found from the lancelet *B. floridae* (drz-Bflo-1 and -2, figs. S1, S10 and S11), is the smallest ribozyme found to date and, together with drz-Spur-1, most closely resembles the mammalian *CPEB3* ribozyme core sequence (*S9*).

Four unique sequence families were isolated in *S. purpuratus* (drz-Spur-1 to 4; figs. S6 to S9). All families are present in multiple copies, with drz-Spur-1 appearing 22 times throughout the genome. It maps to a variety of regions, including the first exon of a predicted non-LTR-like reverse transcriptase, an intron-exon

junction of an ependymin-related protein precursor, and the 3´ UTR of a predicted transmembrane protein containing HEAT repeats (fig. S6). A blastula-stage EST (CD324081) possessing the self-cleaved form of a drz-Spur-1 ribozyme indicates that the ribozyme is expressed and self-cleaves in vivo. Sequence searches based on drz-Spur-1 revealed a self-cleaved sequence in ESTs (e.g. AM565620) from the gastrula and pluteus stage of another sea urchin, *Paracentrotus lividus* (table S2). The little skate, *Leucoraja erinacea*, also harbors a ribozyme with similar core, however the sequence lacks the conserved guanosine of L3, possibly rendering it inactive (table S2).

While a 7 base-pair (bp) P1 helix was specified in the secondary structure searches, drz-Spur-3 and a number of drz-Spur-1 copies were found with a P1 helix of 6 bp. A significant fraction of drz-Spur-3 cleaves within minutes at ambient temperature and 1 mM Mg^{2+}, and high-resolution analysis of the cleavage products confirmed that the shorter P1 helix is sufficient to build a stable HDV and *CPEB3*-like ribozyme (fig. S8).

In both *Caenorhabditis japonica* and *Pristionchus pacificus* ribozymes were found dispersed throughout the organism's genome in many copies. Drz-Ppac-1 is present in 32 copies, and drz-Cjap-1 or fragments thereof take up approximately $1/10000^{th}$ of the organism's genome (figs. S12 to S15). In both organisms, all the copies appear intergenically and the flanking regions exhibit varying levels of sequence conservation between the two nematodes. In *P. pacificus* both regions upstream and downstream of the ribozymes are conserved, while in *C. japonica* only the 3´ region is conserved, suggesting that the self-cleaved form of the ribozyme was inserted into different genomic loci. Two similar sequences of *P. pacificus* were found to self-cleave with surprisingly different rate constants: drz-Ppac-1-1 shows appreciable scission in minutes at ambient temperature in 1 mM Mg^{2+}, while drz-Ppac-1-2 cleaves in hours in 10 mM Mg^{2+} at 37°C (figs. S14 and S15). A naturally occurring variant in which the active-site cytosine is replaced by adenosine was found to be inactive.

A single ribozyme family was identified in lamprey eel *Petromyzon marinus* (drz-Pmar-1). It is one of the smallest and fastest ribozymes tested and maps to the second intron of a zinc finger AN1-type domain 2 protein and to intergenic regions in twelve other contigs. Some of the sequence variants contain mutations of the active-site cytosine (to U), imperfect pairing of the P1 region, or other variations that most likely render them inactive (fig. S16). We found similar sequences in spiny dogfish shark, *Squalus acanthias*, as well as in

coelacanth *Latimeria menadoensis* (table S2), suggesting that this ribozyme sequence may be common among fish species.

Outside of eukaryotes we discovered the ribozyme in *Faecalibacterium prausnitzii*, a human gut bacterium, and *Chilo* Iridescent Virus (CIV, small iridescent insect virus type 6). In *F. prausnitzii*, the drz-Fpra-1 cleavage site maps 106 nt upstream of a phosphoglucosamine mutase (*GlmM*) start codon and 38 nt downstream of a putative cytochrome d ubiquinol oxidase subunit II gene (fig. S17). The two genes may be part of a polycistronic mRNA bisected by the ribozyme that self-cleaves with a rate constant of 0.47 hr^{-1} in 10 mM Mg^{2+} at 37ºC. Interestingly, the genome of *F. prausnitzii* contains another copy of the *GlmM* gene and its 5′ UTR, including most of the ribozyme. However, the first six nt of the ribozyme are not copied, and so the P1 helix cannot be formed. Because this copy starts only six nt from the ribozyme cleavage site, we speculate that the self-cleaved form of the ribozyme and the downstream *GlmM* mRNA were retrotransposed into another site in the genome and, during the process, the first six nt of the ribozyme were lost, perhaps as a result of exonuclease digestion.

In CIV, the drz-CIV-1 cleavage site maps 144 nt upstream of the DNA-dependent RNA polymerase (largest subunit) start codon, several nt downstream of the predicted transcription start site, and 120 nt downstream of a stop codon for a DNA binding protein gene (fig. S18). Because of its proximity to the predicted transcription start site, the ribozyme may serve to inactivate the mRNA translation by cleaving off the 5′ cap once the expression of this *immediate-early* viral gene (*S21*) is no longer necessary, or in cap-independent translation where the ribozyme may act as an internal ribosomal entry site (IRES), a role similar to one suggested for the *CPEB3* ribozyme (*S22*).

A search for sequences similar to drz-Agam-1 ribozymes revealed multiple putative ribozymes with almost identical nt in the P3-L3 regions but divergent sequences elsewhere (table S2). These ribozymes map to other insects, including another mosquito, *Anopheles funestus* (EST CD578134), a pea aphid, *Acyrthosiphon pisum*, and *Rhodinus prolixus*, a vector of Chagas disease parasite *Trypanosoma cruzi*, where it maps to the beginning of many ESTs isolated from the insect's central nervous system (e.g. FG545964). Similar ribozymes are found in several fungi, including *Ajellomyces capsultans* (EST XM_001543659), *Neurospora crassa* (EST

GE968064, possibly in self-cleaved form), and *Trichoderma atroviride*, where we confirmed its in vitro self-cleavage activity (fig. S19). In addition, transcripts with similar putative ribozyme core sequences map to ESTs from artichoke *Cynara scolymus*, sunflower *Helianthus annus*, the Pacific abalone *Haliotis discus*, and tapeworm *Moniezia expansa* (table S2). Finally, the genome of *Diplonema papillatum*, a free-living marine diplonemid, harbors an efficient ribozyme (fig. S19) that bisects an RNA transcript containing an upstream splice-leader sequence and downstream 5S ribosomal RNA, suggesting that the ribozyme may be involved in processing of non-coding RNAs.

Our results indicate that HDV-like ribozymes, and likely ribozymes in general, are widely distributed in nature. Although alignment-based sequence searching generally meets with little success in identifying these catalytic RNAs, a structure-based approach is effective. We note that with the large increase of available genomic information and the continued refinement of alignment algorithms, sequence searching based on the mammalian *CPEB3* ribozyme can now identify drz-Spur-1 and drz-Bflo-1, as well as a related sequence in the Atlantic cod *Gadus morhua* larvae (EST FG272465, table S2), where it may represent the self-cleaved form of a *CPEB3*-like ribozyme. This suggests that the *CPEB3*-like sequences, not just the HDV and *CPEB3* fold, are distributed throughout non-mammalian genomes and that the ribozyme is older than previously thought (*S9*). Retrotransposition likely contributed to this spread, and it is likely that self-cleaving ribozymes have a variety of biological roles.

Figures.

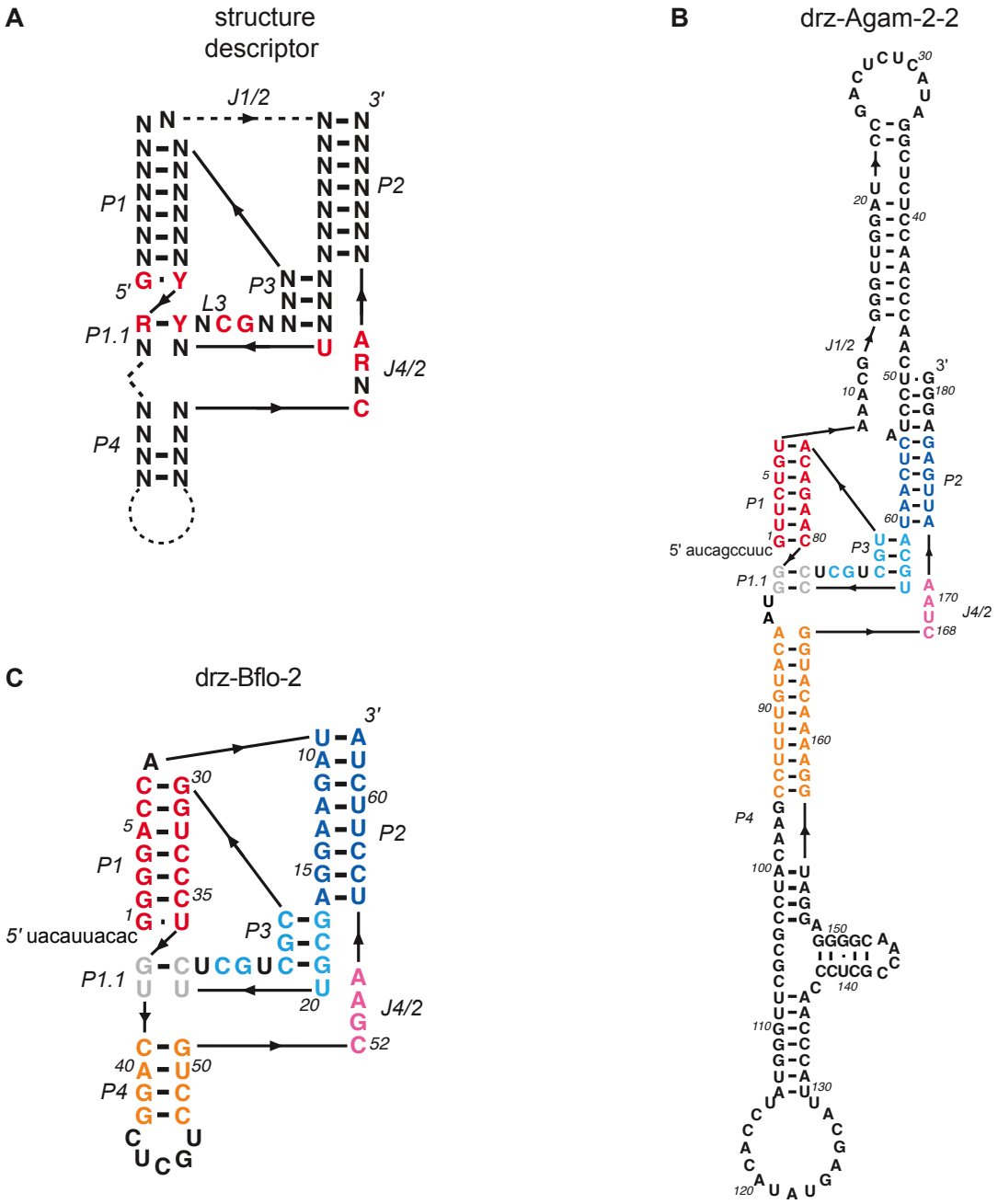


fig. S1. Secondary structure of HDV and *CPEB3*-like ribozymes. (**A**) An example of a structure descriptor for the self-cleaved HDV and *CPEB3* fold. P1 through P4 represent base-paired regions where co-variation is required; J1/2 and J4/2 are single-stranded regions. Typically one mismatch, including G.U wobble pairs, was allowed to form in P1 and P2. Conserved positions are indicated in red; N, any nucleotide; R, purine; Y, pyrimidine. Dashed lines represent variable-length sequence insertions. Solid lines represent direct connections without nucleotide insertions. (**B**) *A. gambiae* drz-Agam-2-2 and (**C**) *B. floridae* drz-Bflo-2 ribozymes, the largest and smallest confirmed ribozymes. Small letters correspond to the lead sequence that is cleaved off by the ribozyme. Core elements are colored by region. Red indicates P1; dark blue, P2; light blue, P3; grey, P1.1; orange, P4; pink, J4/2.

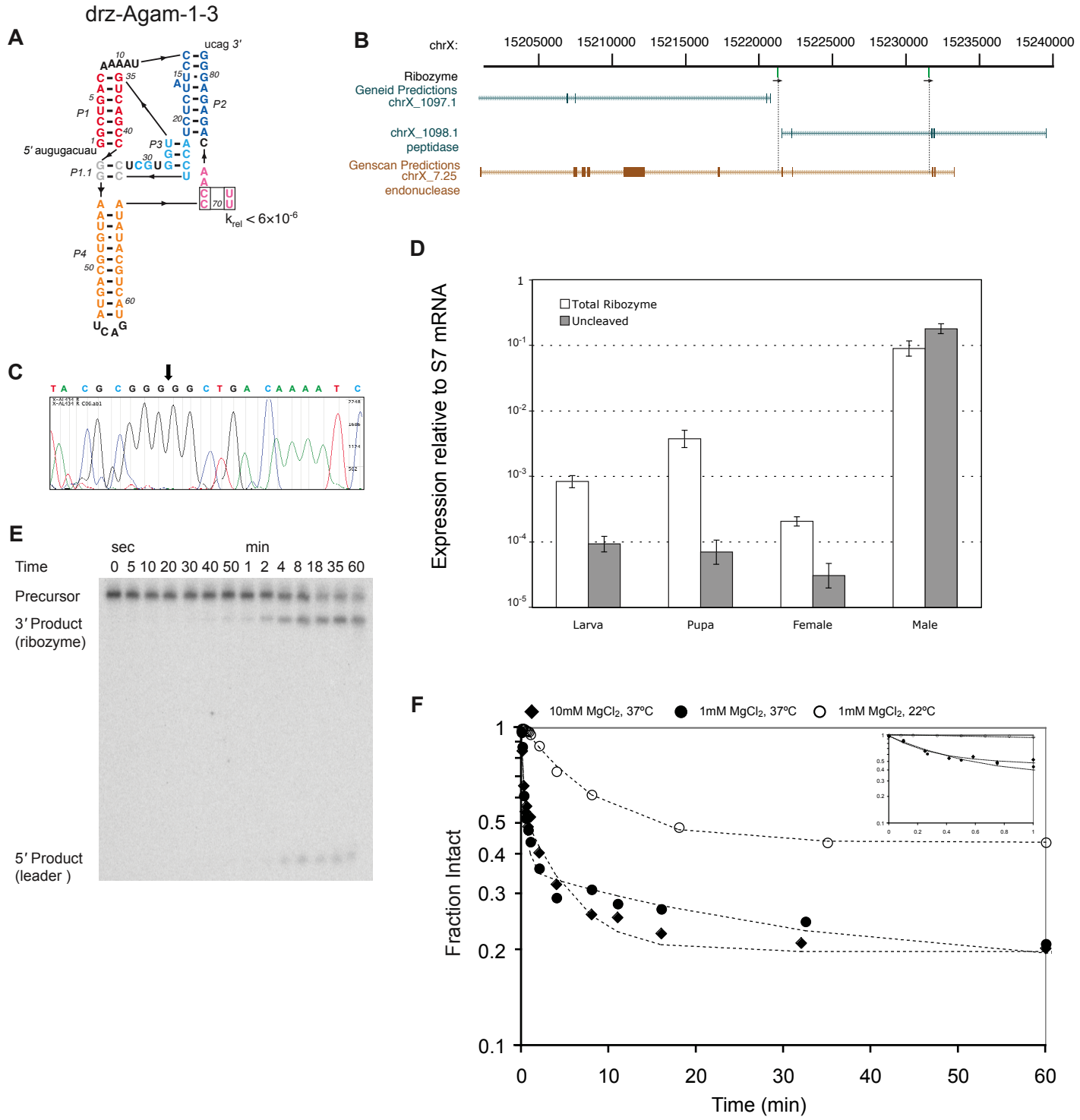
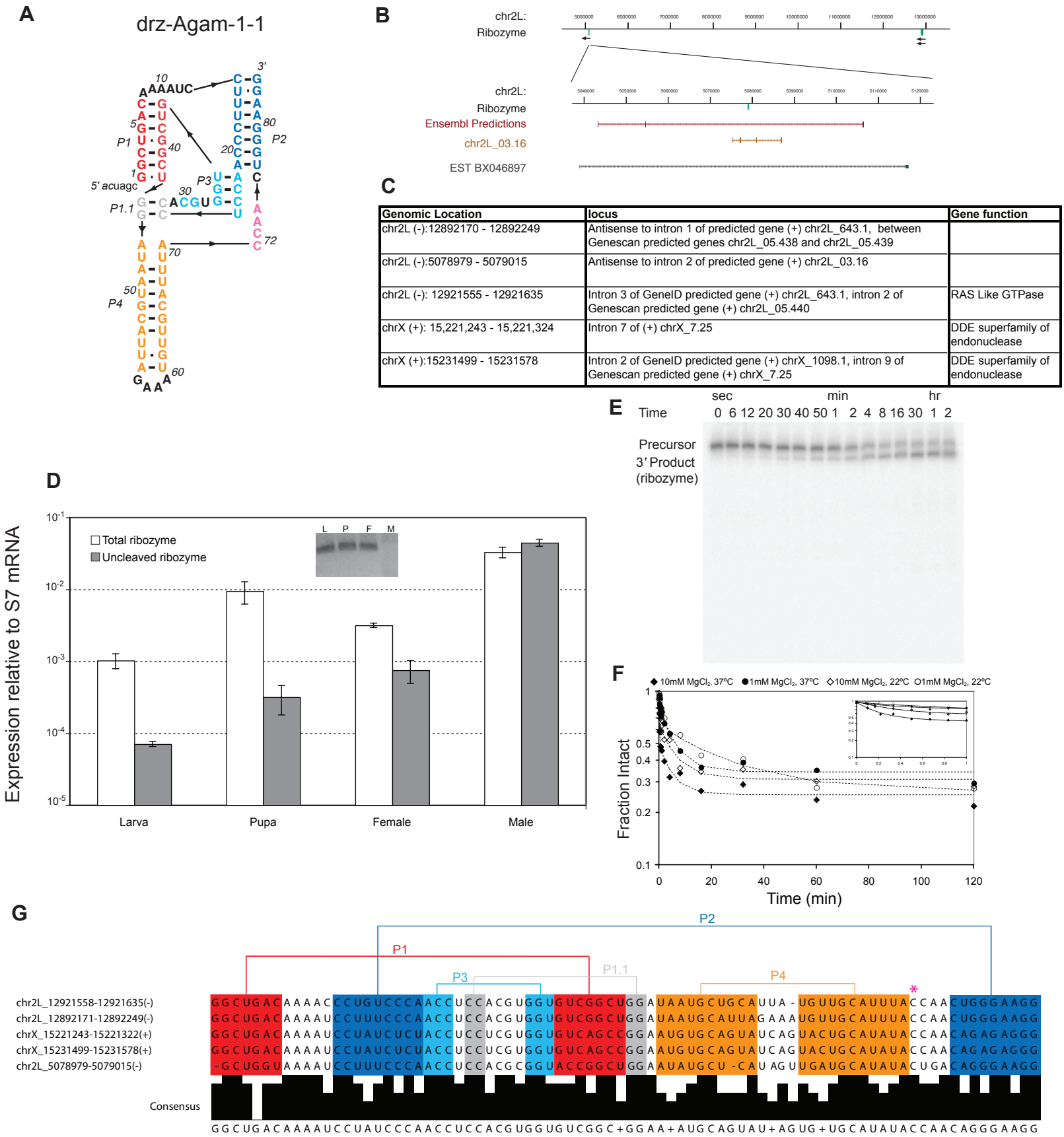


fig. S2. *A. gambiae* drz-Agam-1-3 ribozyme. (**A**) Secondary structure of the ribozyme. Color coding is the same as in Fig. S1. (**B**) Genomic loci of two ribozyme copies that map to introns of a predicted endonuclease gene. (**C**) Sequence of the 5' RACE product, showing the 3' end of trans-templating oligonucleotide following by 5' end of ribozyme sequence (*S17*). Arrow indicates the beginning of the ribozyme sequence, which matches the predicted cleavage site. **(D)** Quantitative RT-PCR of total and uncleaved ribozyme in different life stages of the insect. Expression levels are on logarithmic scale and are normalized to S7 mRNA. **(E)** In vitro self-cleavage activity of the ^{32}P-labeled ribozyme incubated in 1 mM $MgCl_2$ at 22ºC. **(F)** Semi-logarithmic graph of ribozyme activity at different conditions.



Genomic Location	locus	Gene function
chr2L (-):12892170 - 12892249	Antisense to intron 1 of predicted gene (+) chr2L_643.1, between Genescan predicted genes chr2L_05.438 and chr2L_05.439	
chr2L (-):5078979 - 5079015	Antisense to intron 2 of predicted gene (+) chr2L_03.16	
chr2L (-): 12921555 - 12921635	Intron 3 of GeneID predicted gene (+) chr2L_643.1, intron 2 of Genescan predicted gene (+) chr2L_05.440	RAS Like GTPase
chrX (+): 15,221,243 - 15,221,324	Intron 7 of (+) chrX_7.25	DDE superfamily of endonuclease
chrX (+):15231499 - 15231578	Intron 2 of GeneID predicted gene (+) chrX_1098.1, intron 9 of Genescan predicted gene (+) chrX_7.25	DDE superfamily of endonuclease

fig. S3. Secondary structure and activity of drz-Agam-1-1. (**A**) *A. gambiae* drz-Agam-1-1 secondary structure. (**B**) Genomic loci of drz-Agam-1-1 on chromosome 2L. (**C**) Genomic locations of the drz-Agam-1 family as well as their association with protein coding regions. (**D**) Quantitative RT-PCR of total ribozyme and uncleaved ribozyme from different life stages of the insect. Expression levels are on a logarithmic scale and normalized to S7 mRNA. Inset: Agarose gel electrophoresis of 5′ RACE from tissues isolated from different life cycle stages and sex. (**E**) Denaturing PAGE of self-cleavage at 1 mM Mg^{2+}, 37°C. (**F**) Semi-logarithmic graph of the in vitro ribozyme self-cleavage activity at various Mg^{2+} concentrations and temperatures (inset: fast kinetics). (**G**) Sequence alignment of the drz-Agam-1 family with highlighted regions indicating areas of conserved secondary structure and the proposed active site cytosine indicated by asterisk.

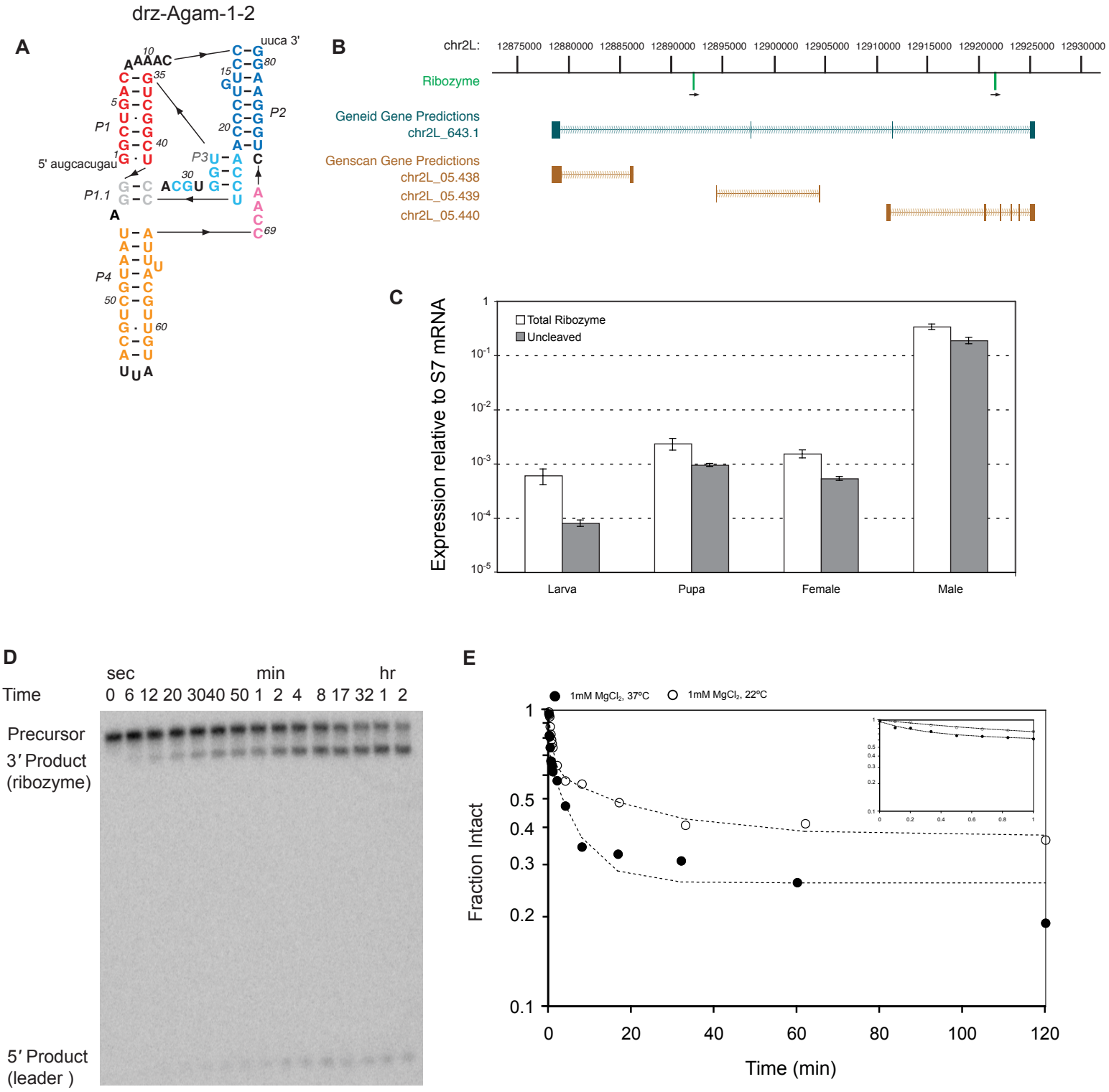


fig. S4. Secondary structure and activity of drz-Agam-1-2. (**A**) *A. gambiae* drz-Agam-1-2 secondary structure. (**B**) Genomic loci, where it maps to the third intron of a predicted RAS-like GTPase. (**C**) Quantitative RT-PCR of total ribozyme and uncleaved ribozyme from different life stages of the insect. Expression levels are on a logarithmic scale and normalized to S7 mRNA. (**D**) Denaturing PAGE of self-cleavage at 1 mM Mg^{2+}, 37°C. (**E**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures (inset: fast kinetics).

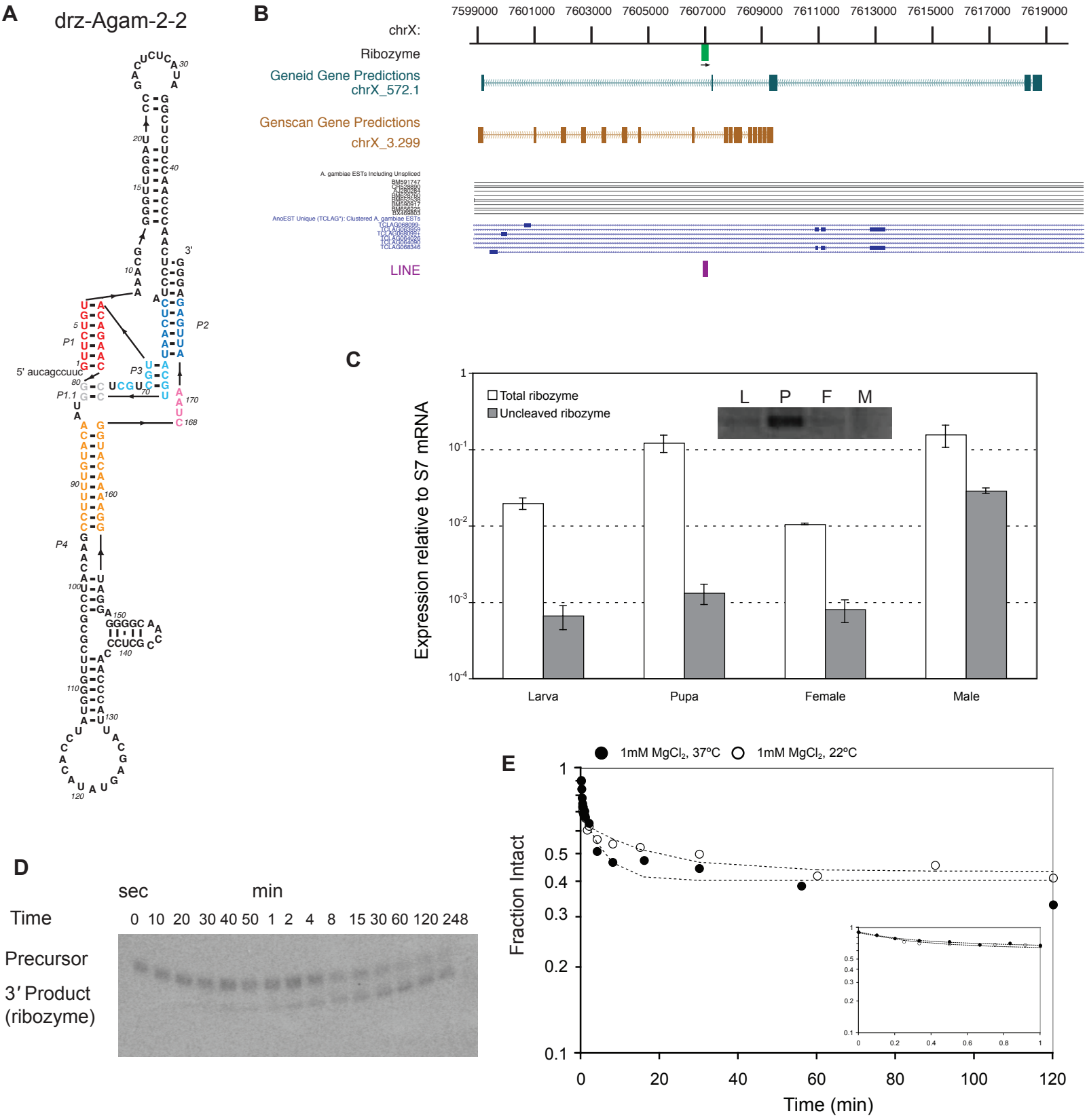
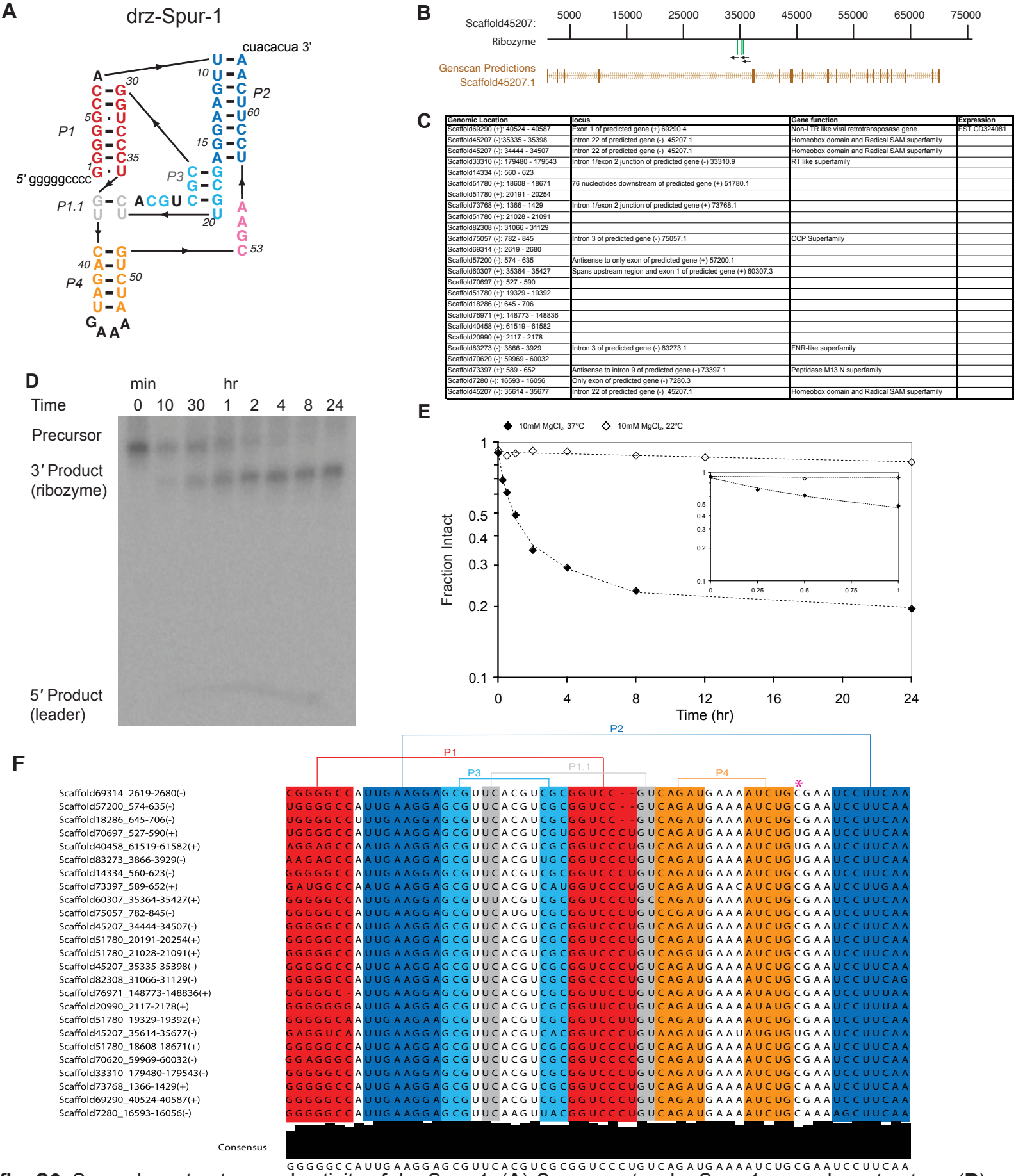
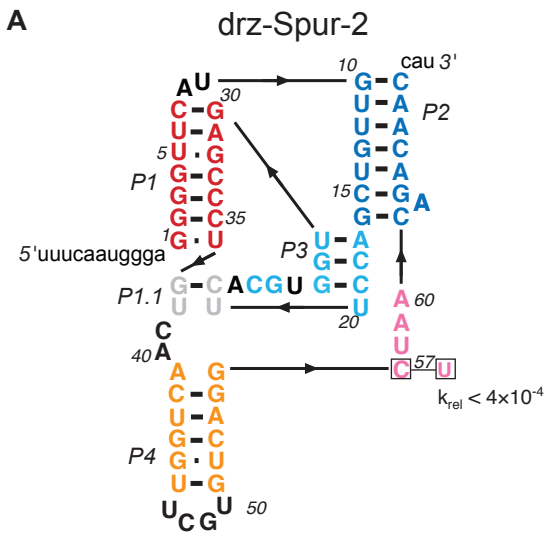


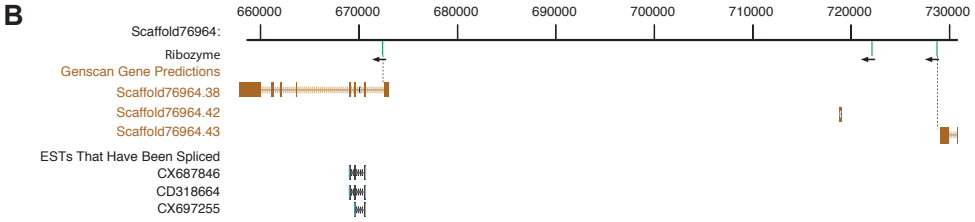
fig. S5. Secondary structure and activity of drz-Agam-2-2. (**A**) *A. gambiae* drz-Agam-2-2 secondary structure. (**B**) Genomic loci, where it maps to introns of two predicted genes. (**C**) Quantitative RT-PCR of total ribozyme and uncleaved ribozyme from different life stages of the insect. Expression levels are on a logarithmic scale and normalized to S7 mRNA. Inset: Agarose gel electrophoresis of 5' RACE from tissue isolated from different life cycle stages and sex. (**D**) Denaturing PAGE of self-cleavage at 1 mM Mg^{2+}, 22°C. (**E**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures (inset: early time points).



Genomic Location	locus	Gene function	Expression
Scaffold69290 (+): 40524 - 40587	Exon 1 of predicted gene (+) 69290.4	Non-LTR like viral retrotransposase gene	EST CD324081
Scaffold45207 (-):35335 - 35398	Intron 22 of predicted gene (-) 45207.1	Homeobox domain and Radical SAM superfamily	
Scaffold45207 (-): 34444 - 34507	Intron 22 of predicted gene (-) 45207.1	Homeobox domain and Radical SAM superfamily	
Scaffold33310 (-): 179480 - 179543	Intron 1/exon 2 junction of predicted gene (-) 33310.9	RT like superfamily	
Scaffold14334 (-): 560 - 623			
Scaffold51780 (+): 18608 - 18671	76 nucleotides downstream of predicted gene (+) 51780.1		
Scaffold51780 (+): 20191 - 20254			
Scaffold73768 (+): 1366 - 1429	Intron 1/exon 2 junction of predicted gene (+) 73768.1		
Scaffold51780 (+): 21028 - 21091			
Scaffold82308 (-): 31066 - 31129			
Scaffold75057 (-): 782 - 845	Intron 3 of predicted gene (-) 75057.1	CCP Superfamily	
Scaffold69314 (-): 2619 - 2680			
Scaffold57200 (-): 574 - 635	Antisense to only exon of predicted gene (+) 57200.1		
Scaffold60307 (+): 35364 - 35427	Spans upstream region and exon 1 of predicted gene (+) 60307.3		
Scaffold70697 (+): 527 - 590			
Scaffold51780 (+): 19329 - 19392			
Scaffold18286 (-): 645 - 706			
Scaffold76971 (+): 148773 - 148836			
Scaffold40458 (+): 61519 - 61582			
Scaffold20990 (+): 2117 - 2178			
Scaffold83273 (-): 3866 - 3929	Intron 3 of predicted gene (-) 83273.1	FNR-like superfamily	
Scaffold70620 (-): 59969 - 60032			
Scaffold73397 (+): 589 - 652	Antisense to intron 9 of predicted gene (-) 73397.1	Peptidase M13 N superfamily	
Scaffold7280 (-): 16593 - 16056	Only exon of predicted gene (-) 7280.3		
Scaffold45207 (-): 35614 - 35677	Intron 22 of predicted gene (-) 45207.1	Homeobox domain and Radical SAM superfamily	

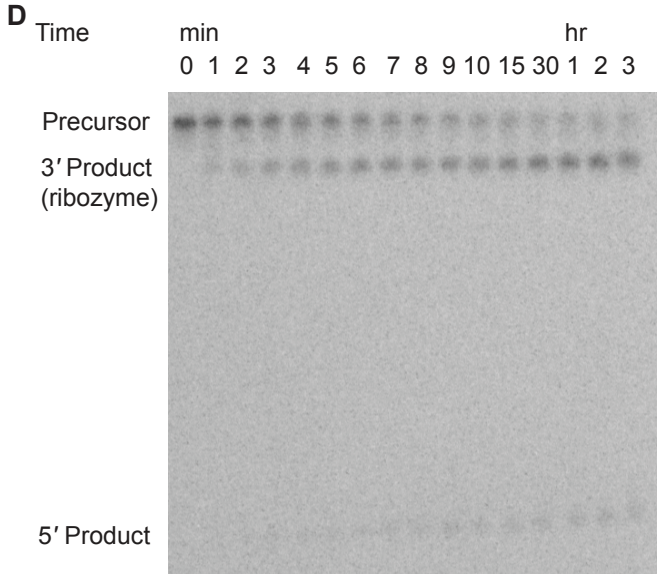
fig. S6. Secondary structure and activity of drz-Spur-1. (**A**) *S. purpuratus* drz-Spur-1 secondary structure. (**B**) Genomic loci of drz-Spur-1. (**C**) Genomic location of the other members of the drz-Spur-1 family and their association with protein coding regions and expression. (**D**) Denaturing PAGE of self-cleavage at 10 mM Mg^{2+}, 37°C. (**E**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures. (**F**) Sequence alignment of the drz-Spur-1 family with highlighted regions indicating areas of conserved secondary structure and the active site C indicated by asterisk. Several of these sequences contain a C to U mutation at this position, likely rendering them inactive.

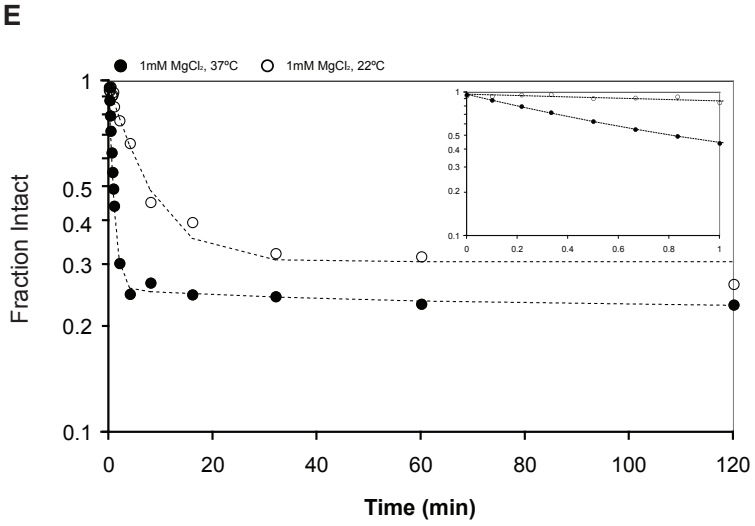




C

Genomic Location	locus	Gene function
Scaffold76964 (-): 722083 - 722149		
Scaffold76964 (-): 672367 - 672430	Intron 1 of predicted gene (-) 76964.38	RNAse PH 3' Exoribonuclease Superfamily
Scaffold55219 (-): 232549 - 232613		
Scaffold31218 (-): 1529 - 1593		





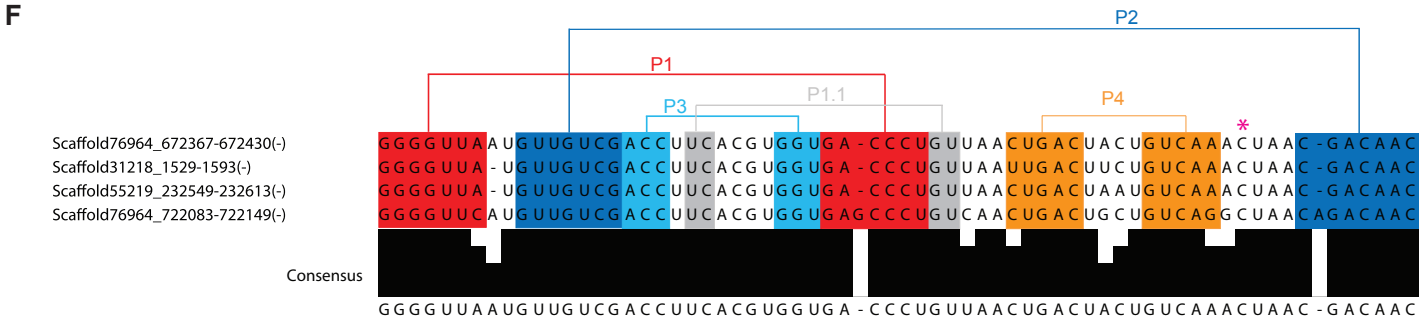
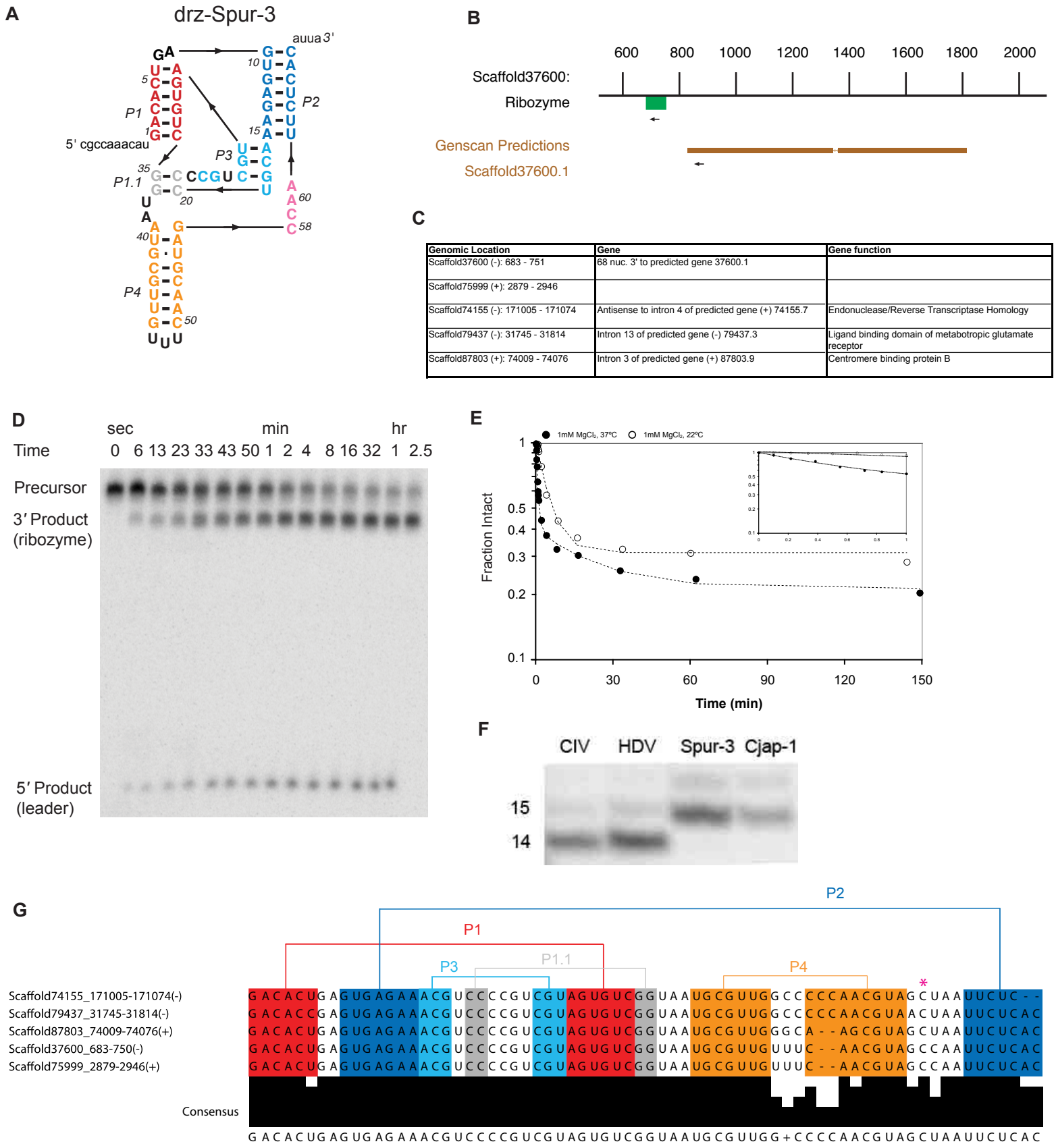
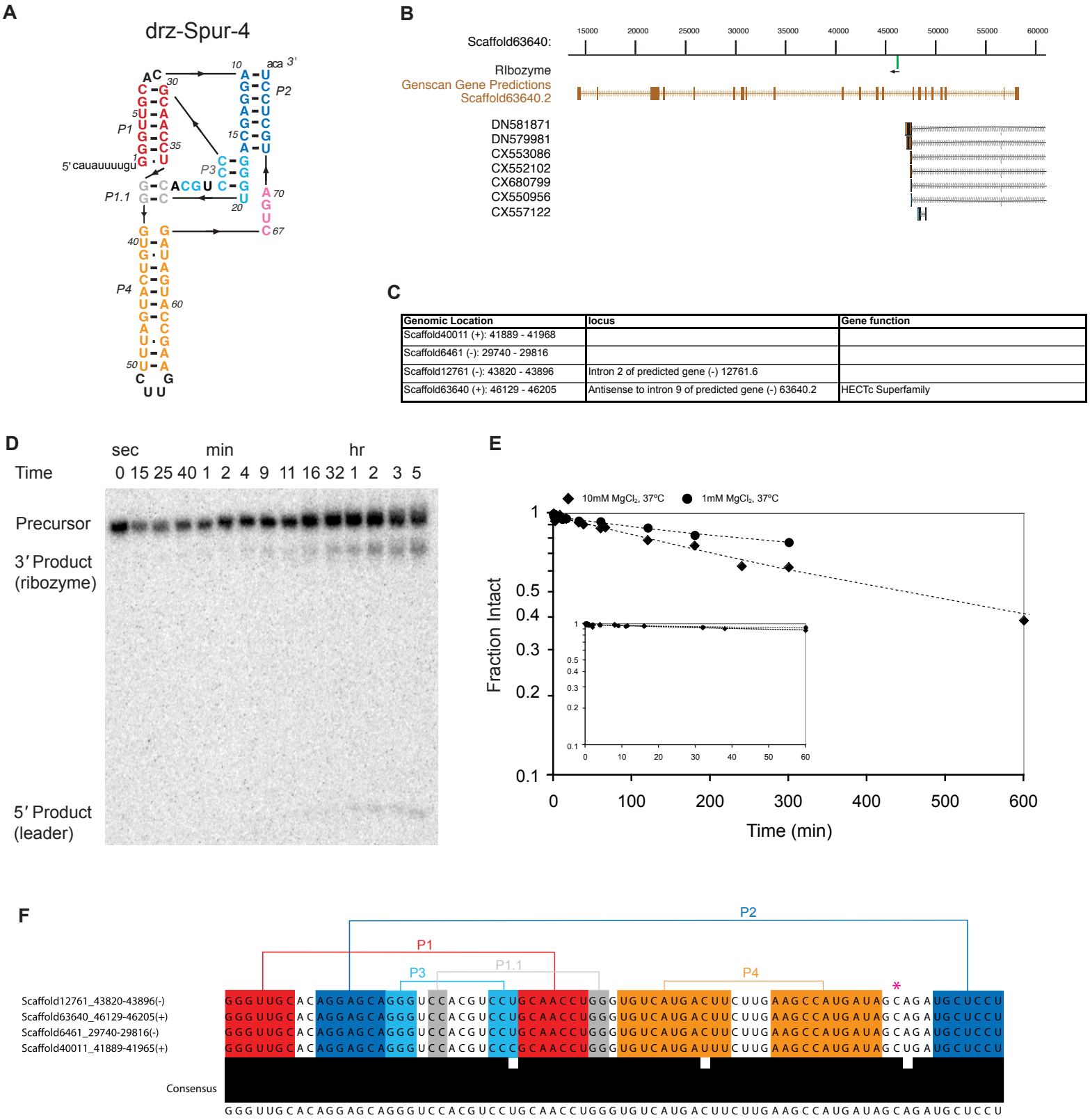


fig. S7. Secondary structure and activity of drz-Spur-2. (**A**) *S. purpuratus* drz-Spur-2 secondary structure. Boxed nucleotides indicate active site C to U mutants with the relative rate constant shown. (**B**) Genomic loci of drz-Spur-2. (**C**) Genomic location of the other members of the drz-Spur-2 family as well as their association with protein coding regions. (**D**) Denaturing PAGE of self-cleavage at 1 mM Mg^{2+}, 22°C. (**E**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures (inset: fast kinetics). (**F**) Sequence alignment of the drz-Spur-2 family with highlighted regions indicating areas of conserved secondary structure and the active site C indicated by asterisk.



Genomic Location	Gene	Gene function
Scaffold37600 (-): 683 - 751	68 nuc. 3' to predicted gene 37600.1	
Scaffold75999 (+): 2879 - 2946		
Scaffold74155 (-): 171005 - 171074	Antisense to intron 4 of predicted gene (+) 74155.7	Endonuclease/Reverse Transcriptase Homology
Scaffold79437 (-): 31745 - 31814	Intron 13 of predicted gene (-) 79437.3	Ligand binding domain of metabotropic glutamate receptor
Scaffold87803 (+): 74009 - 74076	Intron 3 of predicted gene (+) 87803.9	Centromere binding protein B

fig. S8. Secondary structure and activity of drz-Spur-3. (**A**) *S. purpuratus* drz-Spur-3 secondary structure. (**B**) Genomic loci of drz-Spur-3. (**C**) Genomic location of the other members of the drz-Spur-3 family as well as their association with protein coding regions. (**D**) Denaturing PAGE of self-cleavage at 1 mM Mg^{2+}, 37°C. (**E**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures (inset: fast kinetics). (**F**) Denaturing PAGE sequencing gel of drz-Spur-3 leader sequence. A 15 nucleotide leader sequence corresponds to a 6 nucleotide P1 helix. (**G**) Sequence alignment of the drz-Spur-3 family with highlighted regions indicating areas of conserved secondary structure and the active site C indicated by asterisk.



Genomic Location	locus	Gene function
Scaffold40011 (+): 41889 - 41968		
Scaffold6461 (-): 29740 - 29816		
Scaffold12761 (-): 43820 - 43896	Intron 2 of predicted gene (-) 12761.6	
Scaffold63640 (+): 46129 - 46205	Antisense to intron 9 of predicted gene (-) 63640.2	HECTc Superfamily

fig. S9. Secondary structure and activity of drz-Spur-4. (**A**) *S. purpuratus* drz-Spur-4 secondary structure. (**B**) Genomic loci of drz-Spur-4. (**C**) Genomic location of the other members of the drz-Spur-4 family as well as their association with protein coding regions. (**D**) Denaturing PAGE of self-cleavage at 1 mM Mg^{2+}, 37°C. (**E**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures. (**F**) Sequence alignment of the drz-Spur-4 family with highlighted regions indicating areas of conserved secondary structure and the active site C indicated by asterisk.

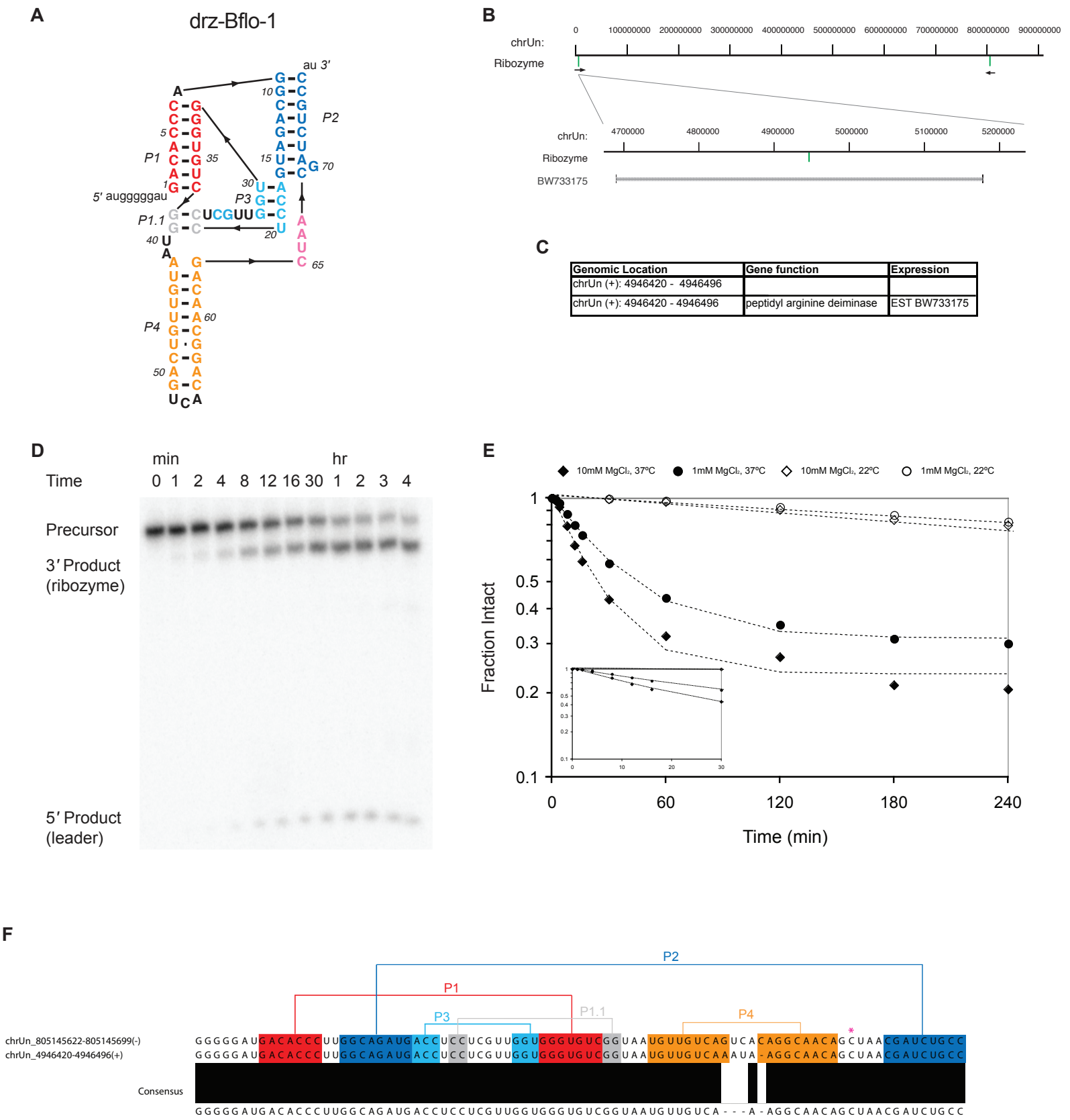
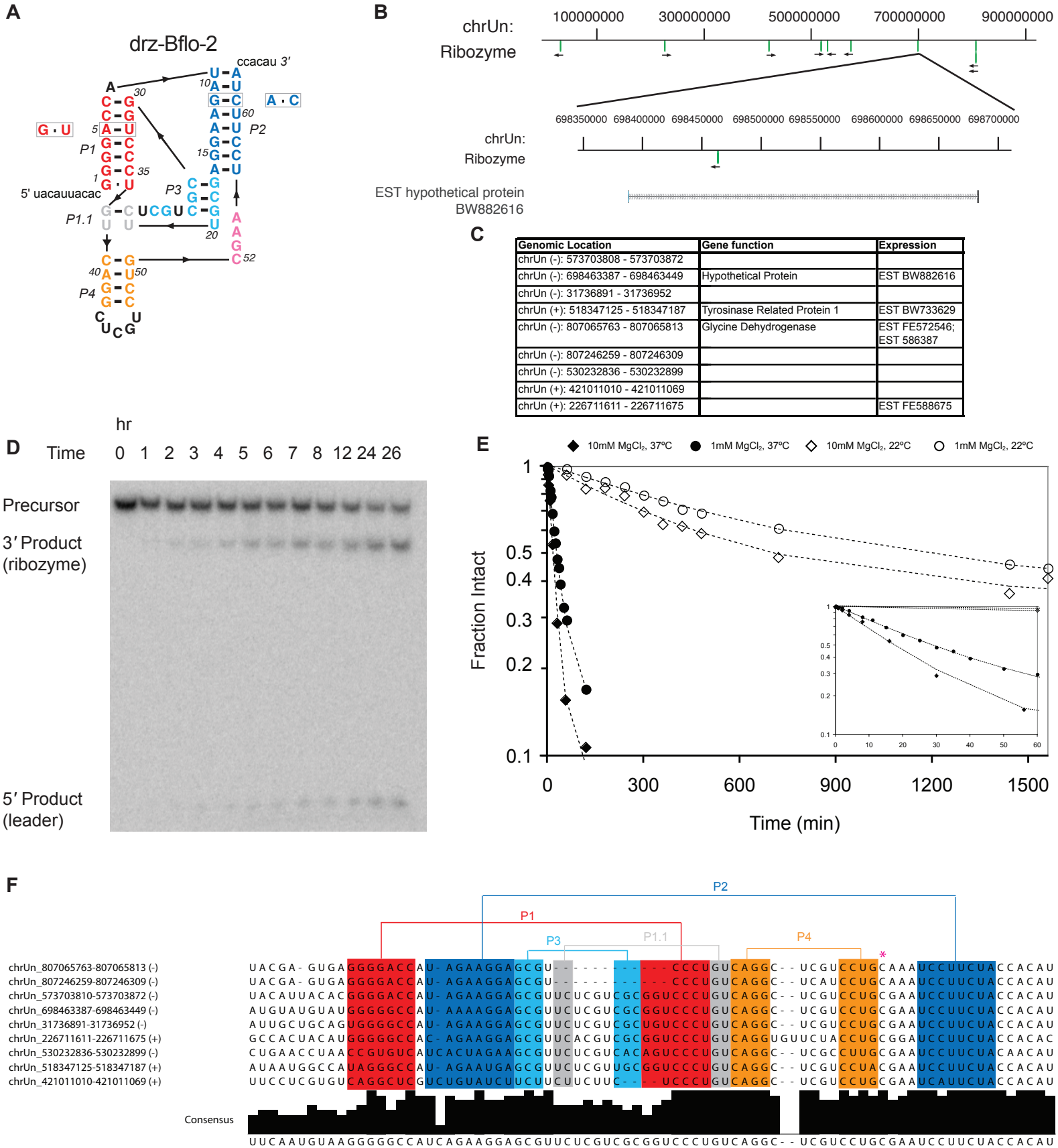


fig. S10. Secondary structure and activity of drz-Bfo-1 (**A**) *B. floridae* drz-Bflo-1 secondary structure. (**B**) Genomic loci of drz-Bflo-1. (**C**) Expression and location of the drz-Bflo-1 family. (**D**) Denaturing PAGE of self-cleavage at 10 mM Mg^{2+}, 37°C. (**E**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures (inset: early time points). (**F**) Sequence alignment of the drz-Bflo-1 family with highlighted regions indicating areas of conserved secondary structure and the active site C indicated by asterisk.



Genomic Location	Gene function	Expression
chrUn (-): 573703808 - 573703872		
chrUn (-): 698463387 - 698463449	Hypothetical Protein	EST BW882616
chrUn (-): 31736891 - 31736952		
chrUn (+): 518347125 - 518347187	Tyrosinase Related Protein 1	EST BW733629
chrUn (-): 807065763 - 807065813	Glycine Dehydrogenase	EST FE572546; EST 586387
chrUn (-): 807246259 - 807246309		
chrUn (-): 530232836 - 530232899		
chrUn (+): 421011010 - 421011069		
chrUn (+): 226711611 - 226711675		EST FE588675

fig. S11. Secondary structure and activity of drz-Bflo-2 (**A**) *B. floridae* drz-Bflo-2 secondary structure. Boxed nucleotides outside the structure indicate the sequence depicted in (**B**), while those within the structure indicate the tested construct. (**B**) Genomic loci of drz-Bflo-2. (**C**) Table of drz-Bflo-2 ribozyme copies. (**D**) Denaturing PAGE of self-cleavage at 1 mM Mg^{2+}, 22°C. (**E**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures (inset: early time points). (**F**) Sequence alignment of the drz-Bflo-2 family with highlighted regions indicating areas of conserved secondary structure and the active site C indicated by asterisk. Sequence conservation is observed in the region 3′ to the ribozyme but not 5′.

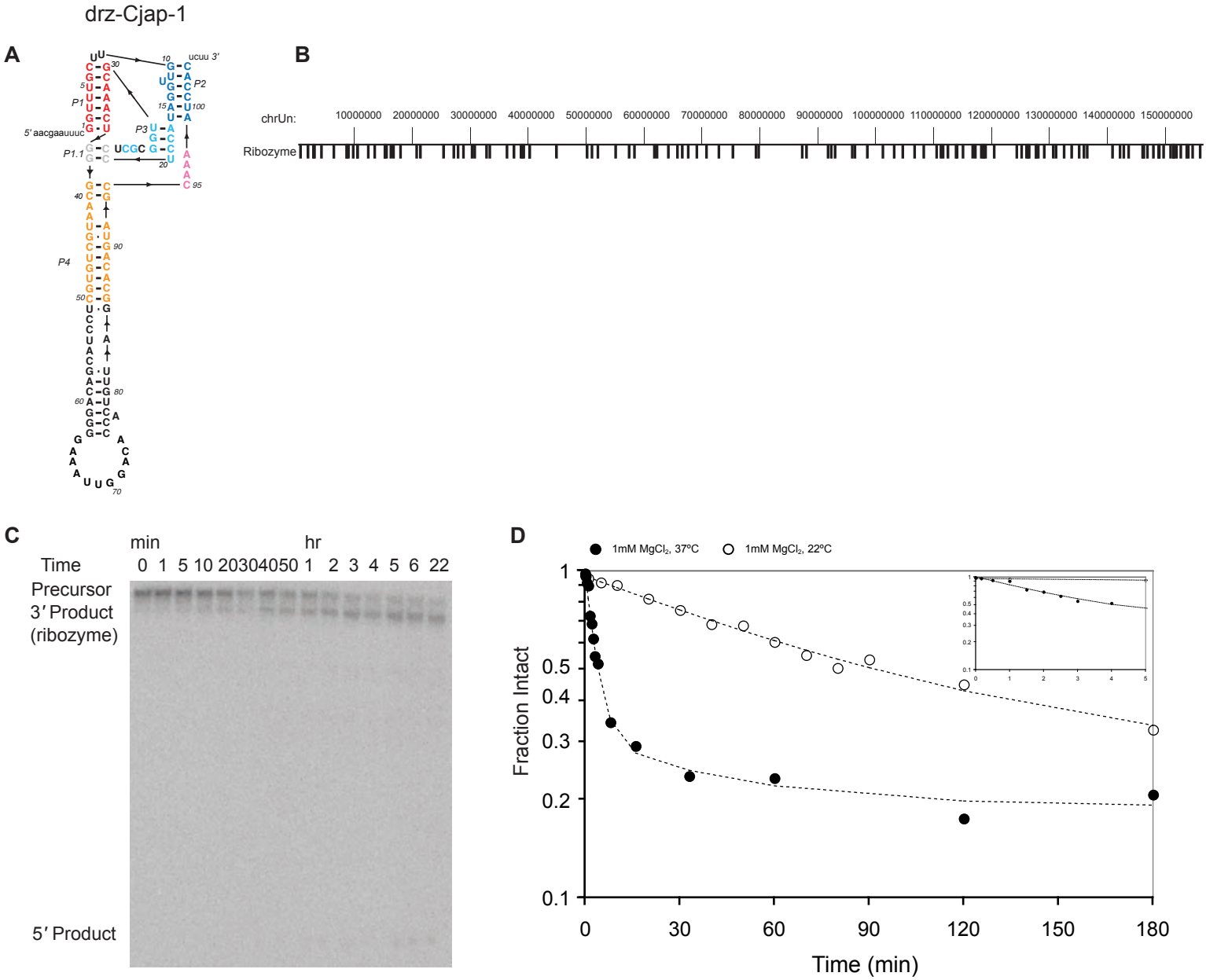


fig. S12. Secondary structure and activity of drz-Cjap-1. (**A**) *C. japonica* drz-Cjap-1 secondary structure. (**B**) Genomic locations of the 122 ribozyme copies. Red bands indicate partial matches with the drz-Cjap-1 construct, while green indicates perfect or near perfect hits. (**C**) Denaturing PAGE of self-cleavage at 1 mM Mg^{2+}, 22°C. (**D**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures (inset: early time points).

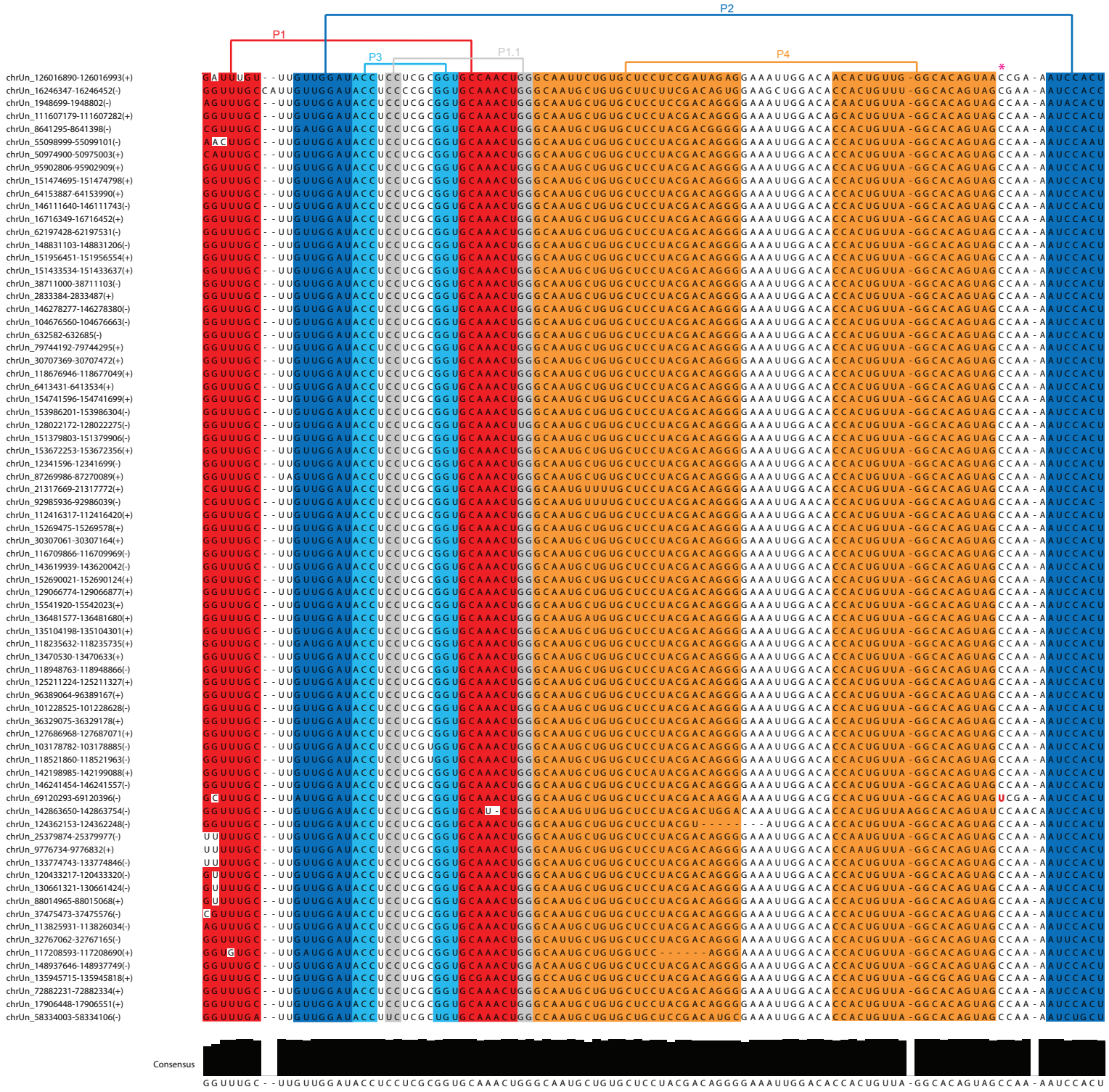


fig. S13. Sequence alignment of top sequence hits from a UCSC BLAT search of the drz-Cjap-1 ribozyme. Regions of conserved secondary structure are colored and mutations to the P1 helix that would likely prevent activity indicated by white blocks. In one sequence the active site C, as indicated by the asterisk, is mutated to U (pink), likely rendering that sequence inactive.

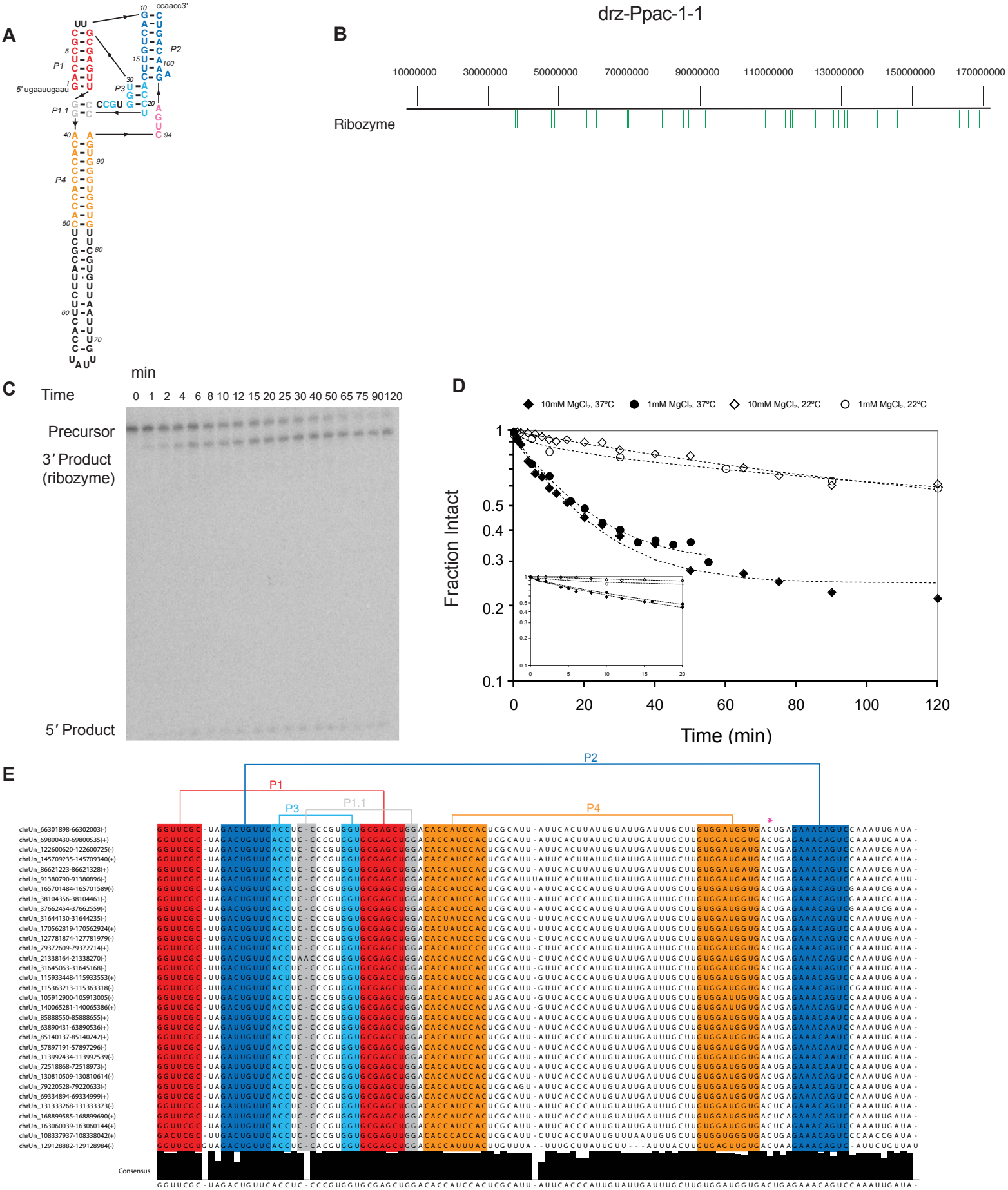


fig. S14. Secondary structure and activity of drz-Ppac-1-1 (**A**) *P. pacificus* drz-Ppac-1-1 secondary structure. (**B**) Genomic loci of drz-Ppac-1 ribozyme family. (**C**) Denaturing PAGE of self-cleavage at 10 mM Mg^{2+}, 37°C. (**D**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures (inset: fast kinetics). (**E**) Sequence alignment of the drz-Ppac-1 family with highlighted regions indicating areas of conserved secondary structure and the active site C indicated by asterisk. In several sequences the active site C is mutated to A.

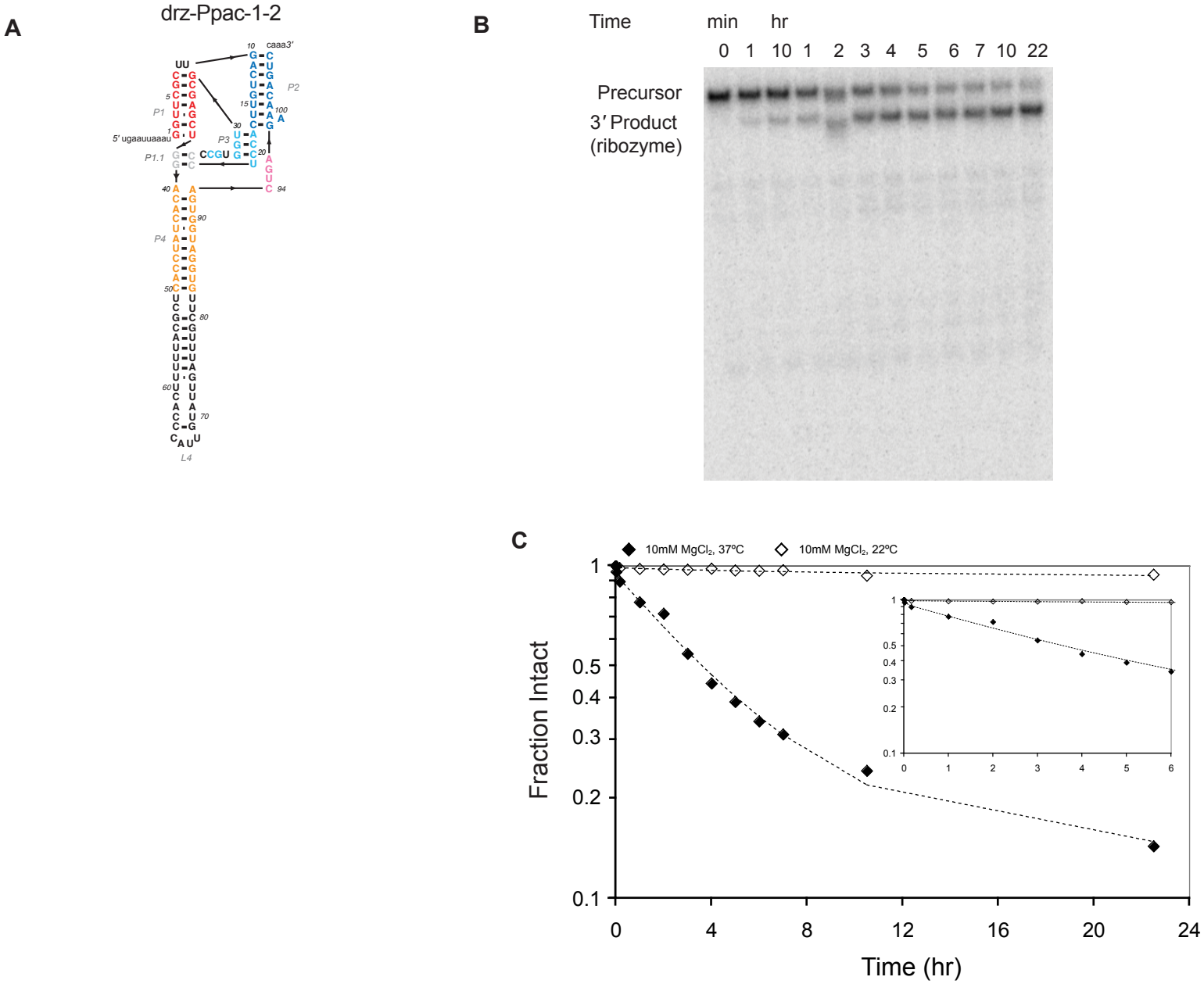
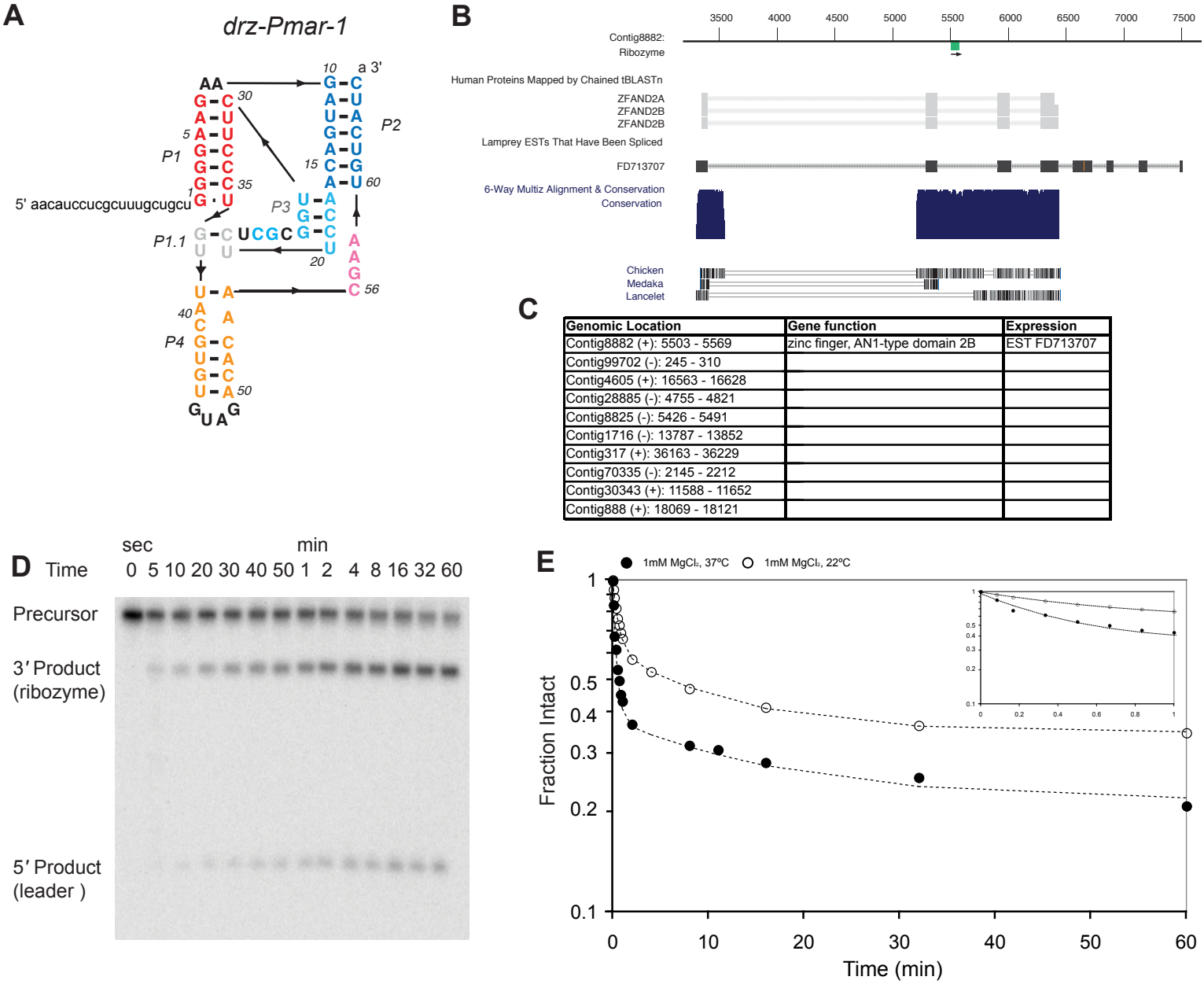


fig. S15. Secondary structure and activity of drz-Ppac-1-2. (**A**) *P. pacificus* drz-Ppac-1-2 secondary structure. (**B**) Denaturing PAGE of self-cleavage at 10 mM Mg^{2+}, 37°C. (**C**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures.



Genomic Location	Gene function	Expression
Contig8882 (+): 5503 - 5569	zinc finger, AN1-type domain 2B	EST FD713707
Contig99702 (-): 245 - 310		
Contig4605 (+): 16563 - 16628		
Contig28885 (-): 4755 - 4821		
Contig8825 (-): 5426 - 5491		
Contig1716 (-): 13787 - 13852		
Contig317 (+): 36163 - 36229		
Contig70335 (-): 2145 - 2212		
Contig30343 (+): 11588 - 11652		
Contig888 (+): 18069 - 18121		

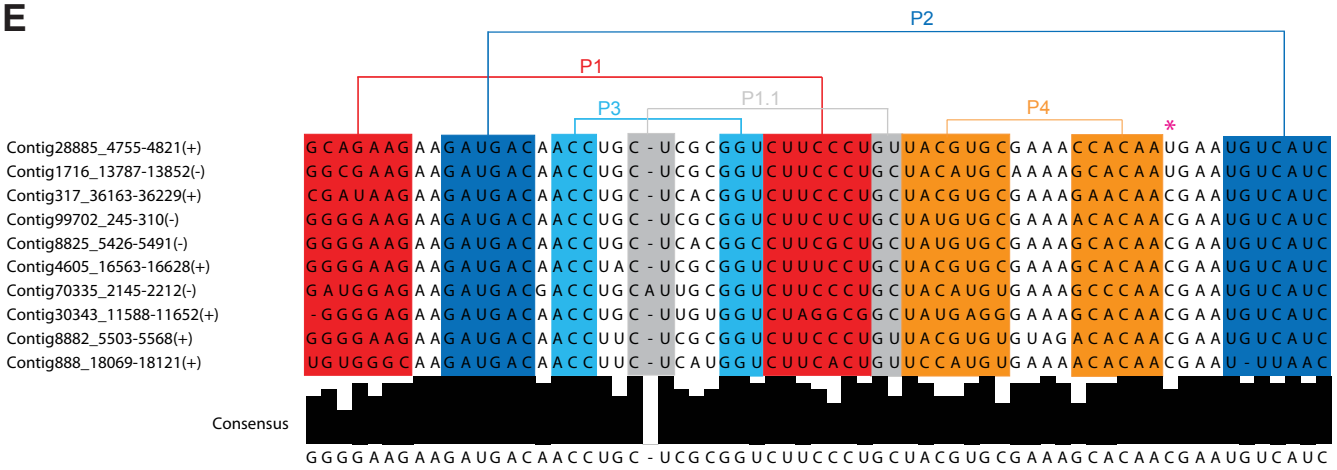


fig. S16. Secondary structure and activity of drz-Pmar-1 (**A**) *P. marinus* drz-Pmar-1 secondary structure. (**B**) Genomic locus of the intragenic copy of drz-Pmar-1. (**C**) Table listing other genomic locations of drz-Pmar-1. (**D**) Denaturing PAGE of self-cleavage at 1 mM Mg^{2+}, 22°C. (**E**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures (inset: fast kinetics). (**F**) Sequence alignment of the drz-Pmar-1 family with highlighted regions indicating areas of conserved secondary structure and the active site C indicated by asterisk. Those sequences with C to U mutations at this position are likely inactive.

drz-Fpra-1

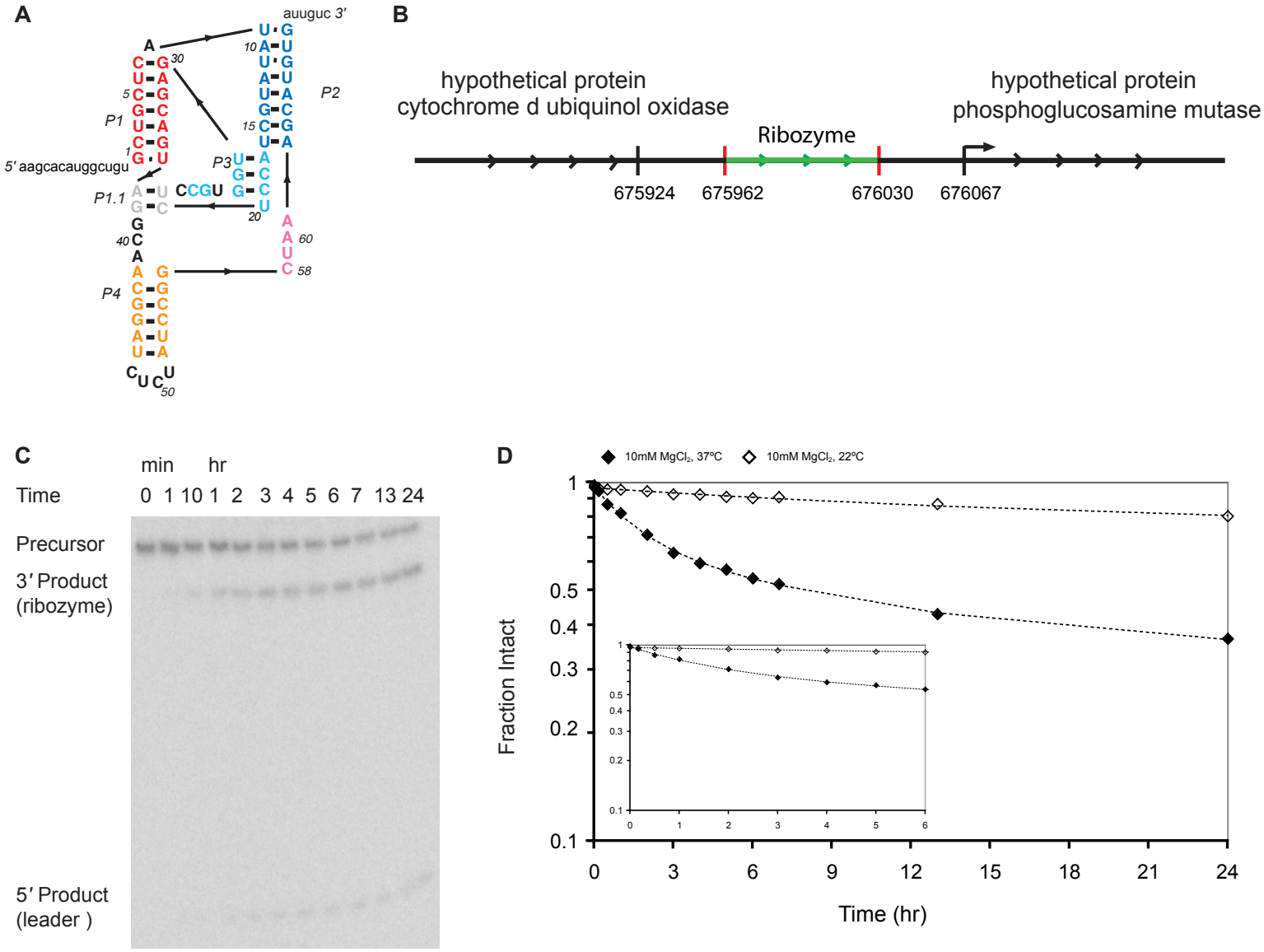


fig. S17. Secondary structure and activity of drz-Fpra-1. (**A**) *F. prausnitzii* drz-Fpra-1 secondary structure. (**B**) Drz-Fpra-1 genomic locus, showing its position between two genes. (**C**) Denaturing PAGE of self-cleavage at 10 mM Mg^{2+}, 37°C. (**D**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures.

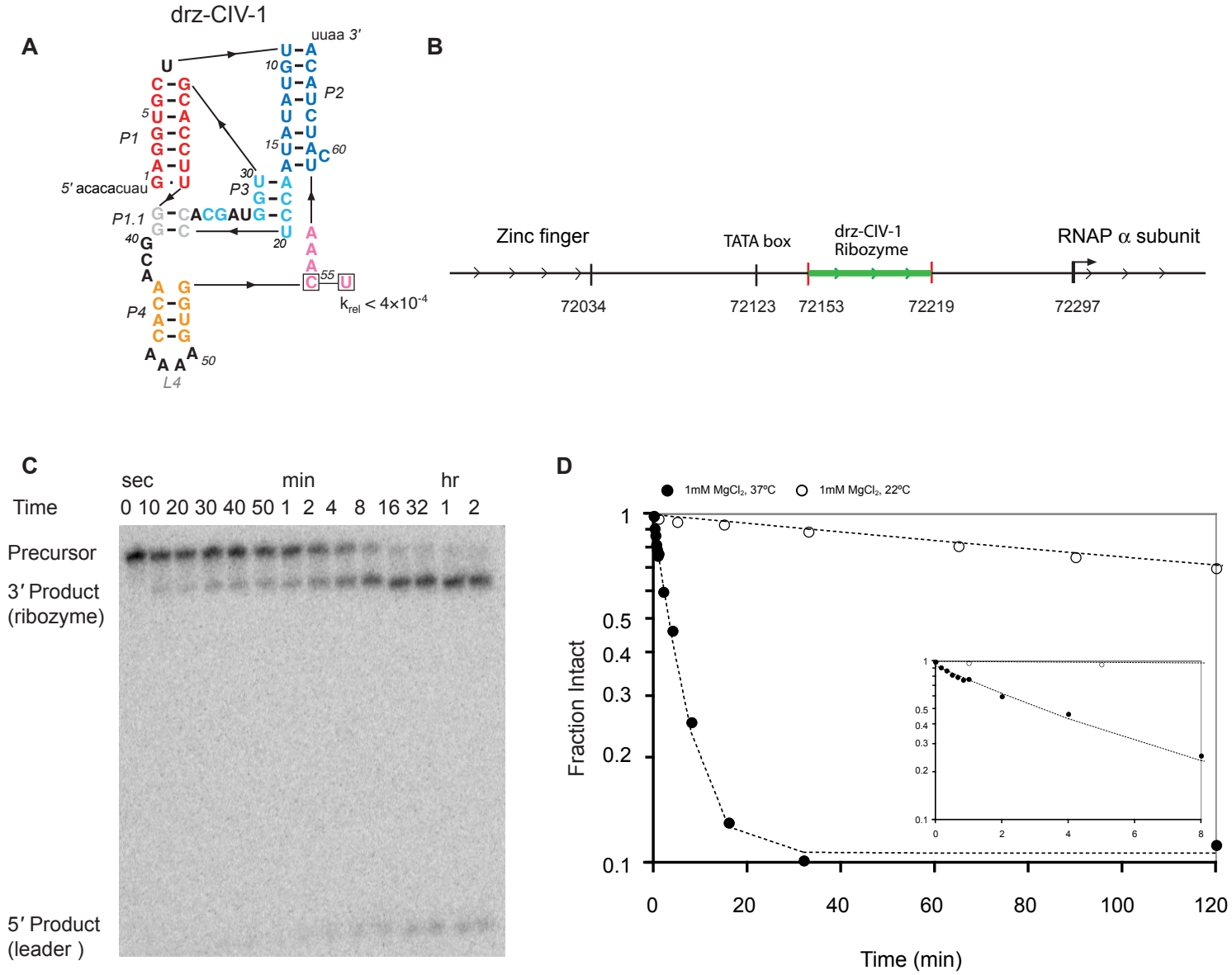


fig. S18. Secondary structure and activity of drz-CIV-1. (**A**) *Chilo* Iridescent Virus drz-CIV-1 secondary structure. Boxed nucleotides indicate active site C to U mutants with the relative rate constant shown. (**B**) Denaturing PAGE of self-cleavage at 1 mM Mg^{2+}, 37°C. (**C**) Semi-logarithmic graph of the in vitro self-cleavage activity of the ribozyme at various Mg^{2+} concentrations and temperatures (inset: early time points).

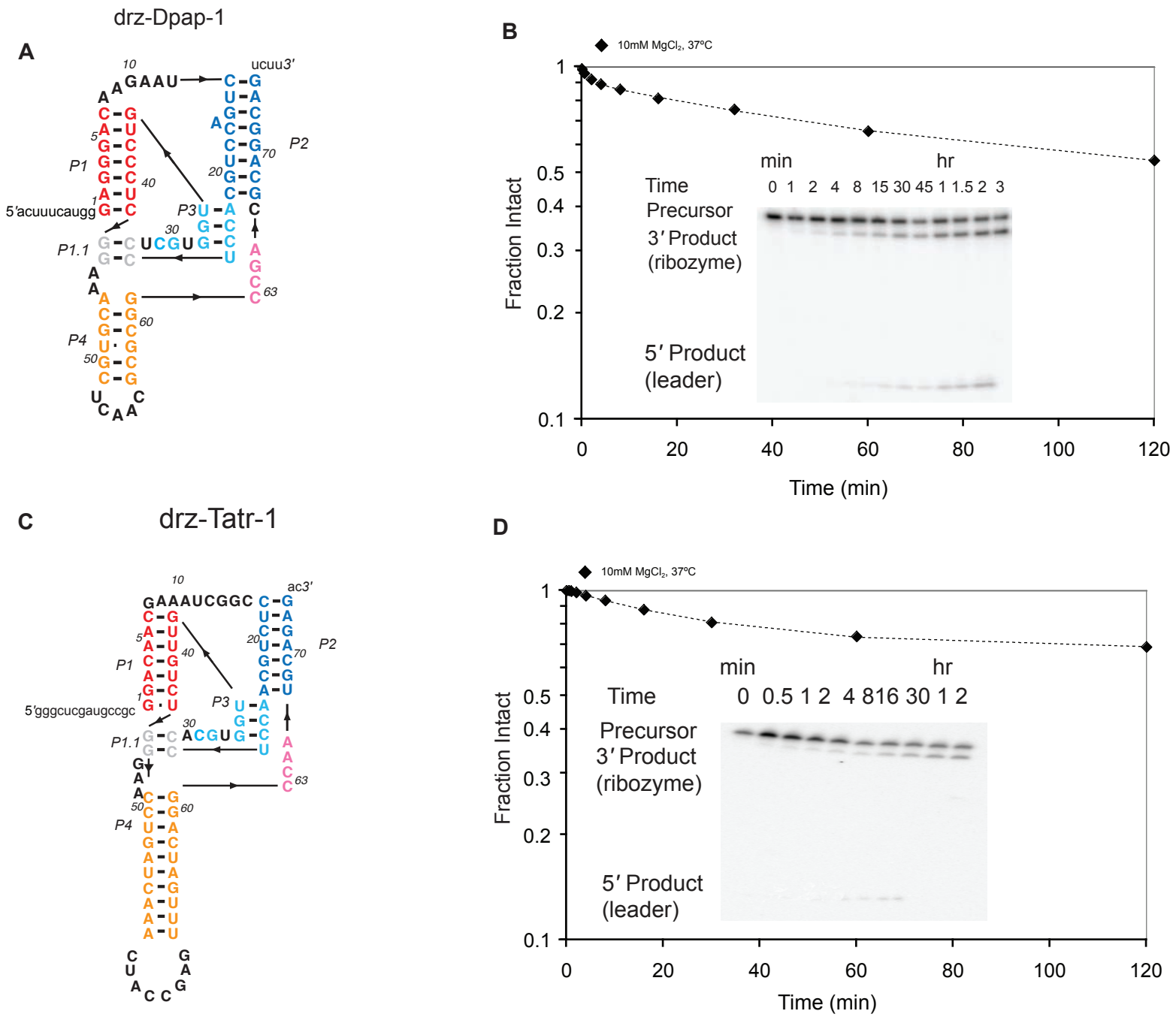


fig. S19. Secondary structure and activity of drz-Dpap-1 and drz-Tatr-1. (**A**) *D. papillatum* drz-Dpap-1 secondary structure. (**B**) Semi-logarithmic graph and denaturing PAGE of self-cleavage at 10 mM Mg^{2+}, 37°C (inset). (**C**) *T. atroviride* drz-Tatr-1 secondary structure. (**D**) Semi-logarithmic graph and denaturing PAGE of self-cleavage at 10 mM Mg^{2+}, 37°C (inset).

Table S1. Organism	Common Name	Ribozyme name	Genomic location of reference sequence	Copy number	Gene	Gene function	Rate constant/hr-1 ([Mg^{2+}]/mM, T/°C)	Expression
Petromyzon marinus	sea lamprey	drz-Pmar-1	Contig8882 (+): 5503 – 5569	10		zinc finger, AN1-type domain 2B	90.0 ± 4.8 (1,22)	EST FD713707
Branchiostoma floridae	lancelet	drz-Bflo-1	chrUn (+): 4946420 – 4946496	2		peptidyl arginine deiminase	2.0 ± 0.2 (1,37)	EST BW733175
		drz-Bflo-2	chrUn (-): 573703808 – 573703872	4			0.11 ± 0.02(10,22)	EST BW882616
Pristionchus pacificus	round worm	drz-Ppac-1-1	chrUn (+): 108337937 – 108338042	32			1.2 ± 0.2 (1,22)	
		drz-Ppac-1-2	chrUn (-): 31644135 – 31644235				0.78 ± 0.1 (1, 37)	
Caenorhabditis japonica	round worm	drz-Cjap-1	chrUn (-): 153986201 – 153986304	122			20 ± 4 (1,37)	
Strongylocentrotus purpuratus	purple sea urchin	drz-Spur-1	Scaffold69290 (+): 40524 – 40595	22	Homolog resides in exon 1 of Predicted Gene (+) 69290.4	Non-LTR like viral retrotransposase	1.4 ± 0.3 (1,22)	EST CD324081
		drz-Spur-2	Scaffold76964 (-): 722083 – 722149	4	Homolog resides in intron 1 of predicted gene (-) 76964.38	RNAse PH 3' exoribonuclease	10.0 ± 0.2 (10,37)	
		drz-Spur-3	Scaffold37600 (-): 683 – 751	5	68 nts 3' to predicted gene 37600.1		12.0 ± 0.8 (1,22)	
		drz-Spur-4	Scaffold40011 (+): 41889 – 41968	4			0.08 ± 0.01 (1,37)	
Anopheles gambiae	African mosquito	drz-Agam-1-1	chr2L (-):12892170 - 12892249	2	Intron 1 of Geneid predicted gene chr2L_643(+), between Genescan predicted genes chr2L_05.438 and chr2L_05.439		29.0 ± 0.3 (1,22)	5' RACE, qPCR
		drz-Agam-1-2	chr2L (-): 12921555 – 12921635	1	Intron 3 of GeneID predicted gene (+) chr2L_643.1, intron 2 of Genescan predicted gene (+) chr2L_05.440		66 ± 2 (1,22)	5' RACE, qPCR
		drz-Agam-1-3	chrX (+): 15,221,243 – 15,221,324	2	Intron 7 of (+) chrX_7.25	DDE superfamliy	160 ± 11 (1,37)	5' RACE, qPCR
		drz-Agam-2-1	chr2R(+): 34860001 – 34860176	3	Intron 3 of GeneID predicted gene (+) chr2R_2558.1, intron 6 of Genescan predicted gene (+) chr2R_15.114	Group I glycosyl transferase	387 ± 63 (1,37)	5' RACE, qPCR
		drz-Agam-2-2	chrX(+): 7606912 – 7607088	1	Intron 4 of GeneID predicted gene (-) chrX_572.1, intron 8 of Genescan predicted gene (+) chrX_3.299	TspO MBR	49 ± 16 (1,22)	5' RACE, qPCR
Faecalibacterium prausnitzii		drz-Fpra-1	675962 – 676030	1	Cleavage site is 38 bp upstream of CydB and 106 bp downstream of GlmM hypothetical proteins.	CydB: cytochrome d ubiquinol oxidase, GlmM: phosphoglucosamine mutase	0.47 ± 0.18 (10,37)	
Chilo iridescent virus	insect iridescent virus type 6	drz-CIV-1	72153 – 72219	1	Cleavage site is 144 bp upstream of RNAP and 120 bp downstream of Zinc finger protein	DNA-dependent RNA polymerase	13.0 ± 0.2 (1,37)	
Diplonema Papillatum		drz-Dpap-1					107 ± 39 (10,37)	EST EC842608
Trichoderma atroviride		drz-Tatr-1					1.8 ± 0.28 (10,37)	EST GE274206

Table S1. Self-cleaving ribozymes from the HDV and *CPEB3* family with confirmed in vitro activity. Number of copies corresponds to the full-length sequences found in the genome, including potentially inactive ribozymes, for each subfamily, except in *P. pacificus* where two subfamilies (drz-Ppac-1-1 and -1-2) are included. Where multiple copies occur, only the sequence with confirmed in vitro activity is listed by its genomic location. Other genomic locations and sequence alignments are shown in the SOM. Gene corresponds to an example of genes where a given ribozyme (sub)family maps. Expression indicates any evidence for in vivo transcription. Rate constant corresponds to the self-cleavage rate constant under indicated conditions (faster of two is reported if the data were fit to a biexponential decay model), averaged over at least two experiments. Error is calculated as average deviation.

Seed sequence	Species	Common name/notes	Expression	leader	P1	J1/2	P2	P3	L3	P3'	P1'	J1.1/4	P4	L4	P4	J4/2	P2'	Comments
drz-Pmar-1	*Squalus acanthias*	Spiny Dogfish shark	(-)EST ES605876	gccgcua	GGGGUGG	AU	GUUGACA	GCC	UUCUUGC	GGC	CCACCCU	GU U	AUGUGAUC	UUA	GAUCACAA	CUAA	UGUCGGC	
drz-Pmar-1	*Latimeria menadoensis*	Indonesian coelacanth			GGGGUGG	A	AGAUGACA	GCU	UUCCCGC	AGC	CCACCCU	GU	UAAG	UGG	UUUA	CC**ACG**	ACUAAU	poor P2, J4/2
drz-Spur-1	*Leucoraja erinacea*	Little Skate	EST 48691621		GAGGGUC	AA	GGAAAGA	GCG	UUCA**C**U	CGC	GGCCCCU	GG	U	UCC	A	CCAA	UCUGUCC	no G in L3, short P4
drz-Spur-1	*Paracentrotus lividus*	common sea urchin	EST AM565620 s-c		GGGGUC	AC	AGAAAGA	GCC	UUCACGC	GGC	GAUCCCU	GU	CAGGCUG	UUUUC	CAGUUUG	C**GAU**	UCUUUCU	
drz-Spur-2	*Leucoraja erinacea*	Little skate	(-)EST EE992768		AGGGCUC	A	UGUUGUCG	ACC	UCCCCGU	GGU	GAGCCCU	GUC	ACUGAC	CCCA	GUCAGG	CAAA	CAAUAACA	
drz-Spur-2	*Paracentrotus lividus*	common sea urchin	EST AM562678 s-c		GGGAUUC	AA	GUUGUCG	ACC	UCCACGU	GGU	GAGUCCU	GUCC	ACUGGU	GUAA	ACCAGG	CGAA	CGGCAAC	
drz-Agam-1	*Anopheles funestus*	African malaria mosquito	EST CD578134		ACUGGC	AAAAU	CCGAUCCCC	GCC	UCCACGU	GGC	GCUGCU	GG	AUGUC GGUCUUGAUGAG	CCUCGUCAC	CUCAGUCAAGACC UAAC	CAAA	GGGACGG	short P1
drz-Agam-1	*Acyrthosiphon pisum*	pea aphid			GGUCGGC	AGA	GUCCUACCCCC	ACC	UCCACGU	GGU	GCCGACU	GGAA	ACGGAAU	UCCG	GUUCCGG	C**AAC**	GGGGGAGGCAC	
drz-Agam-1	*Rhodnius prolixus*	Triatomid bug, vector of Chagas parasite *Trypanosoma cruzi*	EST FG544597	g	GGGGGAC	AAAAAU	CUGAGCUCU	ACC	UCCUCGU	GGU	GCCCCCC	GGCAA	ACUGCUC	UUAU	GAGCAGG	CUAAC	AGAGCCAG	
	Helianthus annus	common wild sunflower	DY925396 s-c	gcggg	GGGCG**U**C	AGU	CCUACUCUGC	ACC	UCCUCGU	GGU	G**U**CGCCU	GGGA	ACCCUCUUUC	GCAA	GAAAGAGGAG	CCAA	GCAGAGAGG	mismatch in P1
	Cynara scolymus	globe artichoke	GE606767 s-c		GGCG**U**C	AGU	CCUACUCUGC	ACC	UCCUCGU	GGU	G**U**CGCCU	GGGA	ACCCUCUUUC	ACAA	GAAAGAGGAG	CCAA	GCAGAGAGG	mismatch in P1
	Trichoderma atroviride	cold tolerant fungus	EST GE274206	ucgaugccgc	GGACAAC	GAAAUCG	GCCUCUGCA	ACC	UCCACGU	GGU	GUUGUCU	GGGAA	CCUGAUCAAAACU	ACC	GAGUUUGAUCAGG	CCAA	UGCAGAGAC	
	Gibberella moniliformis	fungus, rice bakanae disease		ggggaggc	GGGCGGC	GAAGUCG	GACUCUGCA	ACC	UCCUCGU	GGU	GCCGCCU	GGGA	ACCCGGU	CAAU	ACCGGAG	CC**AG**	UGCAGGGAC	
	Magnaporthe grisea	rice blast fungus		cuuuguu	GGGUGAC	CAAU	GACUCUGG	ACC	UCCUCGU	GGU	GUCGCCU	GGGA	ACAGGUA	CAG	UACCAGU	CAA**C**	GCAGAGUC	
	Trichoderma virens	mycoparasitic fungus			AGUCGG**A**		CUCUGCA	ACC	UCCUCGU	GGU	G**C**CGCCU	GGAA	AUCUAGUCUC	GAC	GAAGGCCGAAG	CCAA	UGCAGAG	
	Podospora anserina	filamentous fungus			AGUCGG**A**		CUCUGUC	GCC	UCCUCGU	GGU	G**U**CGCCU	GGAA	ACUCGUCUCUCC	GA	GGAGAGCGAGG	CCAA	CGCAGAG	
	Aiellomyces capsulatus		XM_001543659, CV603133	cacuuaggau	GGGCUGC	GAAGU	CCGACUCUGCC	ACC	UCCUCGU	GGU	GCAGCCU	GGUA	ACAUCAUC	UAGU	GAUGAUGG	CCAA	GAGCAGAGUGG	
	Neosartorya fischeri	invasive fungal infection			GGGCUGC	GAAGU	CCGACUCUGCC	ACC	UCCUCGU	GGU	GCAGCCU	GGUA	ACAUCACC	CAAGU	GGUGAUGG	CCAA	GAGCAGAGCGG	
	Talaromyces stipitatus	filamentous fungus		g	GGCUGC	GAAGU	CCGACUCUGCC	ACC	C**C**UCGU	GGU	GCAGCU	GGUA	ACAUCACU	CAAGU	GGUGAUGG	CCAA	GAGCAGAGCGG	short P1
	Phaeosphaeria nodorum				GGGCAGC	AGAAGUCG	GACCCUUGC	GCC	UCCUCGU	GGU	GCUGCCU	GG	AUAGUCGACCGCUUAG	AAAU	UUGAGCGGGAGAUGAAC	CAAAU	GCUUGGGACGC	
	Alternaria brassicicola				GGGCGGU	AGAAGUCG	GACCCUUGC	GCC	UCCUCGU	GGU	GCCGCCU	GG	AUAGUCGACCAUCUCUUCCU	UUGGUU	GGGAGGGGGUGGGAGAUGAAAU	CU**AG**	GCUUGGGAC	
	Neurospora crassa	bread mold	EST GE968064	a	GGG**UC**GGC		CUCUGC	ACC	UCCACGC	GGU	GCCGCCC	GGGA	ACCUGAAUGCUG	AUUCCACCUUUUCGCGCCAACUUACUCAGACUC AUGUGCGAGAAAACAACUGGUUACUUCUUCGA	CAGCAUACAGAG	CAAAC	GCAGAG	bulge in P1
	Haliotis discus	red shell Pacific abalone	EST EE674746 s-c		GGCACUC	GU	GAAGAGCA	ACC	UCCUCGU	GGU	GAGUGUU	GGG	U	AAC	G	CUAA	UAGCUCUUU	short P4
	Moniezia expansa	sheep and goat tapeworm	EST FE942699	uacggccggg	GGUGGUC	AAA	CCUACUCUCA	ACC	UCCUCGU	GGU	GACCACC	GGUAA	UCCCAAGGG	CAA	CUCUGGGGG	CUAA	UGAGAGGGG	
	Diplonema papillatum	free living marine diplonemid	EST EC842608 s-c		GAGGGAC	AAG	AAUCUGACCUGC	ACC	UCCUCGU	GGU	GUCCCUC	GGAA	ACGUGC	UCAAC	GCGCGG	CCGAC	GCAGGCAGUCUU	
CPEB3	*Gadus morhua*	Atlantic cod	EST FG272465 s-c		GGGGAUC	A	UGGUAGAA	GCG	UUCACGUC	CGC	GAUCCCU	GU	CAGGUCU	UAAU	AAACCUG	CGAA	UUCUACCA	

Table S2. Putative HDV-like ribozymes identified by sequence alignment. Seed sequences are those ribozymes discovered by secondary-structure searching in this study or the mammalian *CPEB3* ribozyme. Expression lists any database evidence that the sequence is expressed; s-c, EST corresponds to the apparent self-cleaved form of the ribozyme; (-), ribozyme is antisense to the EST sequence, therefore the EST may not be evidence of ribozyme transcription. Where available from the genomic data or ESTs, the leader sequence is listed. The sequences are organized into the predicted structure elements of the HDV-like double pseudoknot. The color scheme of the structure elements is similar to Fig. 1, except P1.1 is not indicated explicitly, rather L3 and J1.1/4 are shaded. Sequence positions that interfere with the canonical secondary structure description are in bold and described in Comments.

Materials and Methods

Secondary structure-based searches

We used the RNABOB program (courtesy of S. Eddy, HHMI; ftp://selab.janelia.org/pub/software/rnabob/) to define the conserved sequence and structure elements of the HDV/*CPEB3* ribozymes (fig. S1) (*S11*). The following is an example of a structure descriptor that allows one mismatch in P1, P2, and P4 regions each. In the P2 region the mismatch can also be an insertion, i.e. either an insertion of any nucleotide is allowed in either strand, or, if both strands contain an insertion, a pair of any composition is allowed. "h" allows for G.U wobble pairs, while "r" only considers Watson-Crick base pairs. In this descriptor, at least one C-G base pair is required in P1.1 (specified in s3 and s4), P2 is six or seven base-pairs long, P4 is three base-pairs long, J1/4 is 1-151 nucleotides (nt) long, and L4 is 1-101 nt long. The active-site cytosine and the adenosine involved in an A-minor interaction with P3 are specified in s6 (*S12*).

```
h1 r2 s1 r3 s2 r4 r5 s3 r5' r2' h1' s4 h6 s5 h6' s6 r4' s7 r3'

h1 0:0 G:Y
r2 0:1 NNNNNN:NNNNNN TGCA
r3 0:0 NNN:NNN TGCA
r4 0:0 NNN:NNN TGCA
r5 0:0 NNN:NNN TGCA
h6 0:1 NNN:NNN

s1 0   N[150]
s2 0   *
s3 0   UYCNCG*Y
s4 0   GNN****
s5 0   N[100]
s6 0   NCNRA*
s7 0   *
```

Primer Sequences

drz-Agam-1-1

```
Ribozyme construct: 5' gggaacuagc
GGCUGACAAAAUCCUUUCCCAACCUCCACGUGGUGUCGGCUGGAUAAUGCAUUAGAAAUGUUGCAUUUACCAACUGGGAAgg

AL292: 5' TTCCCGCGAAATTAATACGACTCACTATAGGGAactagcGGCTGACAAAATCCTTTCCCAA

AL293: 5'
ccTTCCCAGTTGGTAAATGCAACATTTCTAATGCATTATCCAGCCGACACCACGTGGAGGTTGGGAAAGGATTTTGTCAGCC

AL301 (inhibitor): 5' ATAGGATTTTGTCAGCC
```

drz-Agam-1-2

Ribozyme construct: 5′ gggagaugcacugau
GGCUGACAAAACCCUGUCCCAACCUCCACGUGGUGUCGGCUGGAUAAUGCUGCAUUAUGUUGCAUUUACCAACUGGGAAGGuuca

AL368: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGATGCACTGAT GGCTGACAAAACCCTGTCCCAACCTCCACG

AL369: 5′ TGAACCTTCCCAGTTGGTAAATGCAACATAATGCAGCATTATCCA GCCGACACCACGTGGAGGTTGGGACAGGGT

AL301 (inhibitor): 5′ ATAGGATTTTGTCAGCC

drz-Agam-1-3

Ribozyme construct: 5′ gggagaaaugugacuau
GGCUGACAAAAUCCUAUCUCUACCUCCUCGUGGUGUCAGCCGGAAUGUGCAGUAUCAGUACUGCAUAUACCAACAGAGAGggucag

AL262: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGAAATGTG
ACTATGGCTGACAAAATCCTATCTCTACCTCCTCGTGGTGTCAGCC

AL263: 5′ CTGACCCTCTCTGTTGGTATATGCAGTACTGATACTGCACAT
TCCGGCTGACACCACGAGGAGGT

AL301 (inhibitor): 5′ ATAGGATTTTGTCAGCC

drz-Agam-2-1

Ribozyme construct: 5′ gggagguaagugau
GCUCUGCAAAUGGGGUAGGAGGCGAUGCCUCGUCCUCAUACCCAACUCCUAUUCCGGCACGUCCACGUCGUGCAGAGCGGUAACAUGCG
UUACUAGGGGUGCAAGAGCUCUUUUUGAGGAGGAGCUCUUUUUGCUGCACUAGUUGCAUCAGAUGGUAACGCAUGGCUAAGCCGGAAAG
GGG

AL441 (primer for PCR from genomic DNA): 5′
TTCCCGCGAAATTAATACGACTCACTATAGGGAGGTAAGTGATGCTCTGCAAAT

AL741 (primer for PCR from genomic DNA): 5′ CCCCTTTCCGGCTTAGCCATG

AL443: 5′ CCCATTTGCAGAGC

drz-Agam-2-2

Ribozyme construct: 5′ gggagaucagccuuc
GUUCUGUAAACGGGGUUGGAUCCGACUCUCAUAGGCUCUCCAACCCAACUCCUACUCAAUACGUCCUCGUCGUACAGAACGGUAACAUG
UUUUCCGAACAUCCGCGCUUGGGUAUACGAGUAUACACCUUACCCAACCCUCGCCAACGGGGAGGAUGGAAAACAUGGCUAAAUUGAGa
ggg

AL444 (primer for PCR from genomic DNA): 5′
TTCCCGCGAAATTAATACGACTCACTATAGGGAGATCAGCCTTCGTTCTGTAAAC

AL445 (primer for PCR from genomic DNA): 5′ CCCTCTCAATTTAGCCATGTTTT

AL446: 5′ CCCCGTTTACAGAAC

AL497 (leader and p1): 5′ TTTACAGAACGAAGG

drz-Pmar-1

Ribozyme construct: 5′ gggagaacauccucgcuuugcugcu
GGGGAAGAAGAUGACAACCUUCUCGCGGUCUUCCCUGUUACGUGUGUAGACACAACGAAUGUCAUCa

AL274: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGAACATCCTCGCTTTGCTGCT GGGGAAGAAGATGACAACCTTC

AL275: 5′ TGATGACATTCGTTGTGTCTACACACGTAACAGGGAAGACC
GCGAGAAGGTTGTCATCTTCTTCCCC

AL304 (inhibitor): 5′ TGTCATCTTCTTCCCC

drz-Bflo-1

Ribozyme construct: 5′ gggagaugggggau
GACACCCAGGCAGAUGACCUCCUCGUUGGUGGGUGUCGGUAAUGUUGUCAGUCACAGGCAACAGCUAACGAUCUGCCaugu

AL288: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGATGGGGGA
TGACACCCAGGCAGATGACCTCC

AL289: 5′ ATGGCAGATCGTTAGCTGTTGCCTGTGACTGACAACATTACCGAC
ACCCACCAACGAGGAGGTCATCTGCCTGGGTGTC

AL322 (inhibitor): 5′ ATCTGCCTGGGTGTC

drz-Bflo-2

Ribozyme construct: 5′ gggaguacauuacac
GGGGACCAUAGAAGGAGCGUUCUCGUCGCGGUCCCUGUCAGGCUCGUCCUGCGAAUCCUUCUAccacau

AL452: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGTACATTACACGGGGACCATAGAAGGAGCGTT

AL453: 5′ ATGTGGTAGAAGGATTCGCAGGACGAGCCTGACAGGGACCGCGACGAGAACGCTCCTTCTATGGTCCCC

AL454 (inhibitor): 5′ CTTCTATGGTCCCC

drz-Spur-1

Ribozyme construct: 5′ ggagggagggggcccc
GGGGGCCAUUGAAGGAGCGUUCACGUCGCGGUCCCUGUCAGAUGAAAAUCUGCGAAUCCUUCAAcuacacua

AL433: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGAGGGAGGGGGGCCCCGGGGGCCATTGAAGGAGCGT

AL419: 5′ TAGTGTAGTTGAAGGATTCGCAGATTTTCATCTGACAGGGACCGCGACGTGAACGCTCCTTCAATGGCCCCC

AL420 (inhibitor): 5′ CCTTCAATGGCCCCC

drz-Spur-1-2

Ribozyme construct: 5′ ggggggaggggcccc
GGGGGCCAUUGAAGGAGCGUUCACGUCGCGGUCCCUGUCAGGUGAAAAUCUGCGAAUCCUUCAacucaacua

AL233: 5′
TAGTTGAGTTGAAGGATTCGCAGATTTTCACCTGACAGGGACCGCGACGTGAACGCTCCTTCTGTGGCCCCCGGGGCCCTATAGTGAGT
CGTATTA

AL282: 5′ TTCCCGCGAAATTAATACGACTCACTATA GGGGGGAGGGGGGCCCCGGGGGCCATTGAAGGAGCGTTCACG

AL238: 5′ TAGTTGAGTTGAAGGATTCGCAGA

AL303 (inhibitor): 5′ CTTCAATGGCCCCC

drz-Spur-2

Ribozyme construct: 5′ gggaguuucaauggga
GGGGUUCAUGUUGUCGACCUUCACGUGGUGAGCCCUGUCAACUGACUGCUGUCAGGCUAACAGACAAccau

AL366: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGTTTCAATGGGA
GGGGTTCATGTTGTCGACCTTCACGTGGTGAGCCCT

AL361: 5′ ATGGTTGTCTGTTAGCCTGACAGCAGTCAGTTGACAGGGCTCACC
ACGTGAAGGT

AL362 (inhibitor): 5′ ACAACATGAACCCC

drz-Spur-3

Ribozyme construct: 5′ gggagcgccaaacau
GACACUGAGUGAGAAACGUCCCCGUCGUAGUGUCGGUAAUGCGUUGUUUCAACGUAGCCAAUUCUCACauua

AL367: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGCGCCAAACA TGACACTGAGTGAGAAACGT

AL364: 5′ TAATGTGAGAATTGGCTACGTTGAAACAACGCATTACCGACACTA
CGACGGGGACGTTTCTCACTCAGTGTCA

AL365 (inhibitor): 5′ CTCACTCAGTGTCA

drz-Spur-4

Ribozyme construct: 5′ gggagcauauuuugu
GGGUUGCACAGGAGCAGGGUCCACGUCCCGCAACCUGGGUGUCAUGAUUUCUUGAAGCCAUGAUAGCUGAUGCUCCuaca

AL297: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGCATATTTTGTGG
GTTGCACAGGAGCAGGGTCCACGTCCCGCAACCTGG

AL298: 5′ TGTAGGAGCATCAGCTATCATGGCTTCAAGAAATCATGACACCC
AGGTTGCGGGACGTGGA

AL323 (inhibitor): 5′ CTCCTGTGCAACCC

drz-Ppac-1-1

Ribozyme construct: 5′ gggagugaauugaau
GACUCGCUUGACUGUUCACCUCCCCGUGGUGCGAGUUGGACACCCACCACUCGCAUUCUUCACCUAUUGUUUAAUUGUGCUUGUGGUGG
GUGACUGAGAAACAGUCccaacc

AL330: 5′
TTCCCGCGAAATTAATACGACTCACTATAGGGAGTGAATTGAATGACTCGCTTGACTGTTCACCTCCCCGTGGTGCGAGTTGGACAC

AL331: 5′ GGTTGGGACTGTTTCTCAGTCACCCACCACAAGCACAATTAAACA
ATAGGTGAAGAATGCGAGTGGTGGGTGTCCAACTCGCACCACGG

AL332 (inhibitor): 5′ AGTCAAGCGAGTC

drz-Ppac-1-2

Ribozyme construct: 5′ gggagugaauuaaau
GGUUCGCUUGACUGUUCACCUCCCCGUGGUGCGAGCUGGACACUAUCCACUCGCAUUUUUCACCCAUUGUAUUGAUUUGCUUGUGGAUG
GUGACUGAGAACAGUCcaaa

AL333: 5′ TTCCCGCGAAATTAATACGACTCACTATA GGGAGTGAATTAAAT
GGTTCGCTTGACTGTTCACCTCCCCGTGGTGCGAGCTGGACACTA

AL334: 5′ GGTTTGGACTGTTTCTCAGTCACCATCCACAAGCAAATCAATACA
ATGGGTGAAAAATGCGAGTGGATAGTGTCCAGCTCGCACCAC

AL335 (inhibitor): 5′ AGTCAAGCGAACC

drz-Cjap-1

Ribozyme construct: 5′ gggagaacgaauuuc
GGUUUGCUUGUUGGAUACCUCCUCGCGGUGCAAACUGGGCAAUGCUGUGCUCCUACGACAGGGGAAAUU
GGACACCACUGUUAGGCACAGUAGCCAAAAUCCACucuu

AL400: 5′
TTCCCGCGAAATTAATACGACTCACTATAGGGAGAACGAATTTCGGTTTGCTTGTTGGATACCTCCTCGCGGTGCAAACTGG

AL401: 5′
AAGAGTGGATTTTGGCTACTGTGCCTAACAGTGGTGTCCAATTTCCCCTGTCGTAGGAGCACAGCATTGCCCAGTTTGCACCGCGAGGA
G

AL402 (inhibitor): 5′ CCAACAAGCAAACC

drz-Fpra-1

Ribozyme construct: 5′ gggagaagcacauggcugu
GCUGCUCAUAUAUGCUACCUCUCCGUGGUGAGCAGUAGGCAACGGAUCUCUAUCCGGCUAAAGCAUGUGauuguc

AL324: 5′
TTCCCGCGAAATTAATACGACTCACTATAGGGAGAAGCACATGGCTGTGCTGCTCATATATGCTACCTCTCCGTGGTGAGCAGT

AL325: 5′ GACAATCACATGCTTTAGCCGGATAGAGATCCGTTGCCTACTGCTCACCACGGAGAGGT

AL326 (inhibitor): 5′ GCATATATGAGCAGC

drz-CIV-1

Ribozyme construct: 5′ gggagacacacuau
GAGGUGCUUGUAGAUAACCUCCACGAUGGUGCACCUUGGGCAACACAAAAGUGGCAAAUCAUCUACAuuaa

AL391: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGACACACTATGAGGTGCTTGTAGATAACCTCCACGAT

AL392: 5′ TTAATGTAGATGATTTGCCACTTTTGTGTTGCCCAAGGTGCACCATCGTGGAGGTTATCTACAA

AL393 (inhibitor): 5′ ATCTACAAGCACCTC

drz-Dpap-1

Ribozyme construct: 5′ gggagacuuucaugg
GAGGGACAAGAAUCUGACCUGCACCUCCUCGUGGUGUCCCUCGGAAACGUGCUCAACGCGCGGCCGACGCAGGCAGucuu

AL494: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGACTTTCATGGGAGGGACAAGAATCTGACCTGCACC

AL495: 5′
AAGACTGCCTGCGTCGGCCGCGCGTTGAGCACGTTTCCGAGGGACACCACGAGGAGGTGCAGGTCAGATTCTTGTCCCTC

AL496 (inhibitor): 5′ CTTGTCCCTCCCATG

drz-Tatr-1

Ribozyme construct: 5′ gggcucgaugccgc
GGACAACGAAAUCGGCCUCUGCAACCUCCACGUGGUGUUGUCUGGGAACCUGAUCAAAACUACCGAGUUUGAUCAGGCCAAUGCAGAGa
c

```
AL498: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGGAGATGCCGCGGACAAC GAAATCGGCCTCTGCA

AL499: 5′
GTCTCTGCATTGGCCTGATCAAACTCGGTAGTTTTGATCAGGTTCCCAGACAACACCACGTGGAGGTTGCAGAGGCCGATTTCGTTGTC
C

AL500 (inhibitor): 5′ GTTGTCCGCGGC
```

DNA Template Preparation

Most constructs were prepared by mutual priming of two synthetic, PAGE-purified oligonucleotides. Primer extension was carried out in a 100 µl volume containing 0.2 mM of each of the four dNTPs, 500 pmol of each primer, and 10 unit of Taq DNA polymerase (Gift of Dr. G. Weiss, Department of Chemistry, UC Irvine). The primer-extension conditions were 94ºC for 1 min, followed by 2 cycles of 50ºC for 30 sec and 72ºC for 2 min.

RNA Transcription

RNA was transcribed at 37ºC for an hour in a 5 µl volume containing 10 mM of DTT, 2.5 mM each GTP, UTP and CTP, 250 µM ATP, 1.25 µCi [α-^{32}P]-ATP (Perkin Elmer, Waltham, MA), 7.75 mM $MgCl_2$, 20 µM of inhibitor oligo (specific for each construct), 1 unit of T7 RNA polymerase, and 0.5 pmole of DNA template.

Cleavage Kinetics

In vitro self-cleavage reactions were performed as previously described (*S9*). The denaturing polyacrylamide gel (PAGE) of self-cleavage products was exposed to phosphorimage screens (Molecular Dynamics) and analyzed by using Typhoon phosphorimager and ImageQuant software (GE Healthcare). Fraction intact was modeled using a biexponential decay and uncleaved residuals function, with the faster cleavage rates reported in Fig. 1 and Table S1.

Total RNA extraction

A. gambiae samples were gifted by the laboratory of Prof. Anthony James, University of California, Irvine. All samples were suspended in Tri-reagent (Invitrogen) and total RNA was extracted following the supplier's protocol. The extraction was performed twice to prevent isolation of genomic DNA. DNase I treatment was omitted to avoid ribozyme self-cleavage during the enzyme reaction. All RNA samples were stored at -80ºC.

5′ RACE

In vivo self-cleavage was confirmed using the Clontech SMART RACE cDNA Amplification Kit and pooled *A. gambiae* total RNA, as well as separate larva, pupae, adult female, and adult male total RNA. The supplier's protocol was followed with the exception that a random decamer primer was added during the First-Strand cDNA Synthesis. Outside gene-specific primers were used for primary 5′ RACE (drz-Agam-1-1: cgtttcgtcgtcaccttatttgacgtaaacc, drz-Agam-1-2: tcgcaacagtacggatgacgctgt, drz-Agam-1-3: cctgtcgatggattaatacggtcaacgg, drz-Agam-2-1: tttgccgttgaaggtaaggccgaga, drz-Agam-2-2: ttcgtgggtccatttgtccaagcag). Nested gene-specific primers were used for secondary 5′ RACE (drz-Agam-1-1: tggtggttgccaatctagtgccctagct, drz-Agam-1-2: caagcaaattgcatcaacgtgcaaa, drz-Agam-1-3: cagtctggtgaacattcaaacaccctcagtg, drz-Agam-2-1: tccagcactaatttgttggcctgtctcc, drz-Agam-2-2: ggatgagccaccggctcgagaatag). Products from both primary and secondary 5′ RACE PCR reactions with pooled total RNA were sequenced by Genewiz (La Jolla, CA).

Quantitative PCR

cDNA synthesis of 1 μg of larvae, pupae, adult female, and adult male total RNA was performed with random decamer primers using SuperScript III and ThermoScript reverse transcriptases (Invitrogen). Quantitative PCR was performed with forward primers for both total ribozyme and uncleaved ribozyme (drz-Agam-1-2: total ribozyme – gacaaaaccctgtcccaacctcca, uncleaved - tcgtttcgtacgttccggaagaattt; drz-Agam-1-3: total ribozyme - cgtggtgtcagccggaatgtg, uncleaved – ttttgcttcgtgctctaagcttcacaaa; drz-Agam-2-1: total ribozyme - tgcaaatggggtaggaggcgatg, uncleaved - caatcaatcaggatgttaacggcaaaaa) and nested gene specific primers designed for 5′ RACE PCR were used as reverse primers. Uncleaved primers amplify a given construct across the ribozyme cleavage site and total ribozyme primers amplify the cDNA downstream of the cleavage site, therefore, they include both cleaved and uncleaved ribozyme populations. The MiniOpticon Real-Time PCR Detection System (Bio-Rad Laboratories, Inc.) was used to perform and monitor the reactions for 45 cycles. 10 μl of iQ SYBR Green Supermix (Bio-Rad) was combined with 6 μl diluted cDNA and 4 μl 2.5 μM primer mixtures in each reaction. Each set of experiments was repeated three times and normalized to expression of S7 ribosomal protein mRNA (*S17*).

References

S1. A. C. Forster, R. H. Symons, *Cell* **50**, 9 (1987).
S2. A. Hampel, R. Tritz, *Biochemistry* **28**, 4929 (1989).
S3. B. J. Saville, R. A. Collins, *Cell* **61**, 685 (1990).
S4. L. Sharmeen, M. Y. Kuo, G. Dinter-Gottlieb, J. Taylor, *J. Virol.* **62**, 2674 (1988).
S5. H. N. Wu *et al.*, *Proc Natl Acad Sci U S A* **86**, 1831 (1989).
S6. W. C. Winkler, A. Nahvi, A. Roth, J. A. Collins, R. R. Breaker, *Nature* **428**, 281 (2004).
S7. R. Przybilski *et al.*, *Plant Cell* **17**, 1877 (2005).
S8. M. Martick, L. H. Horan, H. F. Noller, W. G. Scott, *Nature* **454**, 899 (2008).
S9. K. Salehi-Ashtiani, A. Luptak, A. Litovchick, J. W. Szostak, *Science* **313**, 1788 (2006).
S10. G. W. Fox, C. R. Woese, *Nature* **256**, 505 (1975).
S11. D. Gautheret, F. Major, R. Cedergren, *Comput Appl Biosci* **6**, 325 (1990).
S12. A. R. Ferre-D'Amare, K. Zhou, J. A. Doudna, *Nature* **395**, 567 (1998).
S13. A. Ke, K. Zhou, F. Ding, J. H. Cate, J. A. Doudna, *Nature* **429**, 201 (2004).
S14. T. S. Wadkins, A. T. Perrotta, A. R. Ferre-D'Amare, J. A. Doudna, M. D. Been, *RNA* **5**, 720 (1999).
S15. A. Nehdi, J. P. Perreault, *Nucleic Acids Res* **34**, 584 (2006).
S16. M. Legiewicz, A. Wichlacz, B. Brzezicha, J. Ciesiolka, *Nucleic Acids Res* **34**, 1270 (2006).
S17. Y. Ding, F. Ortelli, L. C. Rossiter, J. Hemingway, H. Ranson, *BMC Genomics* **4**, 35 (2003).
S18. Materials and methods are available as supporting material on Science Online.
S19. To introduce a systematic nomenclature for the ribozymes we propose to divide the name into parts representing the ribozyme type (drz – Delta-like ribozyme); organism in analogy to the rules established for restriction endonucleases, based on the binomial nomenclature (*Anopheles gambiae* – Agam); and sequence family and subfamily number in consecutive order (e.g. drz-Agam-2-1, drz-Agam-2-2). The virus sequences are named using the common acronym for the virus (drz-CIV-1).
S20. M. D. Been, A. T. Perrotta, S. P. Rosenstein, *Biochemistry* **31**, 11843 (1992).
S21. S. M. D'Costa, H. Yao, S. L. Bilimoria, *Arch Virol* **146**, 2165 (2001).
S22. A. Lupták, J. W. Szostak, in *Ribozymes and RNA catalysis,* D. M. J. Lilley, F. Eckstein, Eds. (Royal Society of Chemistry, Cambridge, UK, 2007).

S23. We thank the Lupták laboratory, G. Weiss, and J. Kieft for comments; A. James and O. Marinotti for *A. gambiae* samples. This work was supported by the University of California, Irivne. A.L. is a member of the Chao Family Comprehensive Cancer Center. The GenBank accession numbers for the confirmed ribozymes are BK006878 to BK006897.