Published online 20 November 2018

Nucleic Acids Research, 2019, Vol. 47, No. 1 **29–42**
doi: 10.1093/nar/gky1118

Design of highly active double-pseudoknotted ribozymes: a combined computational and experimental study

Ryota Yamagami[1,2], Mohammad Kayedkhordeh[3], David H. Mathews[3,4,*] and Philip C. Bevilacqua[1,2,5,*]

[1]Department of Chemistry, Pennsylvania State University, University Park, PA 16802, USA, [2]Center for RNA Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA, [3]Department of Biochemistry & Biophysics and Center for RNA Biology, University of Rochester Medical Center, Rochester, New York, NY 14642, USA, [4]Department of Biostatistics & Computational Biology, University of Rochester Medical Center, Rochester, New York, NY 14642, USA and [5]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA

Received August 26, 2018; Revised October 22, 2018; Editorial Decision October 23, 2018; Accepted October 24, 2018

ABSTRACT

Design of RNA sequences that adopt functional folds establishes principles of RNA folding and applications in biotechnology. Inverse folding for RNAs, which allows computational design of sequences that adopt specific structures, can be utilized for unveiling RNA functions and developing genetic tools in synthetic biology. Although many algorithms for inverse RNA folding have been developed, the pseudoknot, which plays a key role in folding of ribozymes and riboswitches, is not addressed in most algorithms. For the few algorithms that attempt to predict pseudoknot-containing ribozymes, self-cleavage activity has not been tested. Herein, we design double-pseudoknot HDV ribozymes using an inverse RNA folding algorithm and test their kinetic mechanisms experimentally. More than 90% of the positively designed ribozymes possess self-cleaving activity, whereas more than 70% of negative control ribozymes, which are predicted to fold to the necessary structure but with low fidelity, do not possess it. Kinetic and mutation analyses reveal that these RNAs cleave site-specifically and with the same mechanism as the WT ribozyme. Most ribozymes react just 50- to 80-fold slower than the WT ribozyme, and this rate can be improved to near WT by modification of a junction. Thus, fast-cleaving functional ribozymes with multiple pseudoknots can be designed computationally.

INTRODUCTION

RNA is a biopolymer that can form complex structures that allows it to possess functions essential to life. Myriad RNA-related genetic tools have been designed such as programmed RNA binding to target DNA and mRNA to regulate transcription and translation (1,2). *In vitro* systematic evolution of ligands by exponential enrichment (SELEX) (3,4) approaches have been utilized for generating novel RNAs that have desired functions for genetic tools. However, SELEX experiments require expertise and take a relatively long time to gain evolved RNAs and perform next-gen sequencing. Moreover, there is a need to design RNAs that adopt a desired fold to carry out a function for biotechnology or medicinal applications. Therefore, computational algorithms for rapid, rational, and accurate RNA design are being developed. The inverse folding problem, which identifies a sequence that will fold to a given structure or set of structures, has been extensively used for design of protein sequences and introduced to RNA sequences (Figure 1) (5–18). There are many algorithms for inverse RNA folding, which are well-described in a recent review (19). Notably, complete computational design of a a self-cleaving hammerhead ribozyme using RNAiFold was reported, where the designed ribozyme has sequence-specific cleaving activity with the catalytic mechanism of a native hammerhead ribozyme (20). Although pioneering, this algorithm does not currently allow design of sequences that adopt a pseudoknot fold.

The pseudoknot is essential to nearly all functional RNA structures including ribozymes, riboswitches, rRNA, RNase P RNA, telomerase RNA, tmRNA and several viral RNAs) (21–26). Generally, the pseudoknot involves a stem-loop reaching over and interacting with another RNA

*To whom correspondence should be addressed. Tel: +1 814 863 3812; Fax: +1 814 865 2927; Email: pcb5@psu.edu
Correspondence may also be addressed to David H. Mathews. Tel: +1 585 275 1734; Fax: +1 585 275 6007; Email: David_Mathews@urmc.rochester.edu

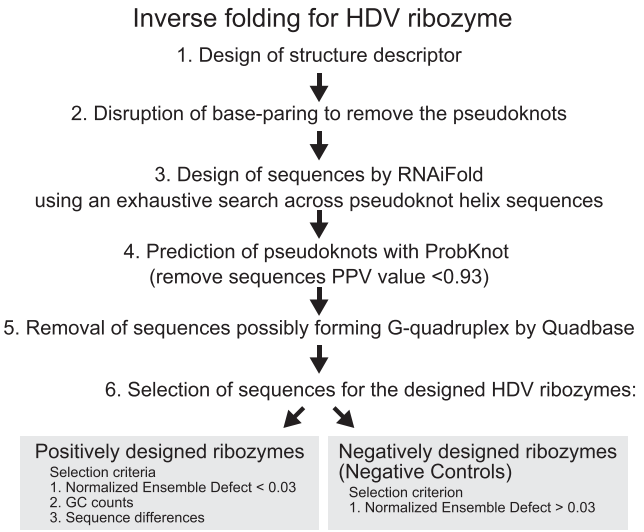

Figure 1. Workflow for computational RNA design for HDV ribozyme.

segment via its loop or a stem defect. Pseudoknots are critical to RNA function because they allow an RNA to fold into more compact/complicated structures (26–28). For instance, pseudoknot structure leads to functional domains of riboswitches and self-cleaving ribozymes, stabilization of RNA structures and programmed ribosome frame shifting (21,23,26,29,30). Topologically a pseudoknot consists of two helices that are non-nested, i.e. if the sequence of nucleotides is placed on a circle and base pairs drawn as lines, the pseudoknots require that the base pairs cross. A pseudoknot can be defined formally as more than 2 bp, i-j and i'-j', such that nucleotide i appears before i', i' before j, and j before j' in the primary RNA sequence (*i.e.* i<i'<j<j'). Despite the critical importance of pseudoknot structures for RNA function, most RNA inverse folding algorithms do not consider pseudoknot-containing RNAs (19). For the few inverse folding algorithms, such as AntaRNA, MODENA and Nanofolder, that do consider pseudoknots (10,11,31), there is no experimental evidence that the predicted RNAs are active.

We present herein a method for inverse RNA folding for the functional double-pseudoknot HDV ribozyme. The HDV and HDV-like ribozymes are self-cleaving ribozymes that are conserved in the three domains of life (32,33) and that contribute to rolling-circle genome replication, gene regulation and retrotransposon processing (34,35). They undergo site-specific cleavage at a rate $\sim 10^6$ over background. HDV ribozymes are comprised of 5 stems (P1, P1.1, P2, P3 and P4) and 2 linker regions (J1/2 and J4/2) (Figure 2A). There are double-pseudoknot pairings in the P1/P2 and P1.1/L3 regions that are required to fold into the active form (29), and a recent study revealed that the J1/2 region adjusts the topology of the catalytic core (36). We chose the HDV ribozyme as a model RNA because it has a double pseudoknot, has been extensively characterized mechanistically (37–41) and has a high-resolution crystal structure available (PDB ID: 3NKB) (29). Kinetics, thermodynamics and structure analyses performed herein reveal that our designed ribozymes have self-cleavage activity that is sequence-specific, shares the same mechanism as wild-type ribozymes, and is only ~50- to 80-fold slower. Moreover, redesign of a junction region leads to catalytic activity that is comparable to the native HDV ribozyme.

MATERIALS AND METHODS

Computation

Computational design of sequences. Computational RNA design was performed following the workflow in Figure 1. Central to the procedure is RNAiFold, a software package for the design of sequences that will fold to a user-specified secondary structure. RNAiFold systematically and efficiently identifies all sequences that will fold with lowest free energy to the specified secondary structure, using the Vienna RNA package to predict structures (42). RNAiFold is not currently able to include pseudoknots in the structure, however. Therefore, we developed a procedure in which we first identify pseudoknot-free sequences with RNAiFold, and post-process the output of RNAiFold to select the subset of sequences that will also form the pseudoknot. The pseudoknot is removed from the structure descriptor used by RNAiFold and the nucleotide pairs in the sequence are temporarily forced to be single stranded. We ran RNAiFold separately for all possible combinations (8192, see below) of Watson-Crick pairs in the pseudoknot. To post-process, we checked the predicted structure for each sequence using ProbKnot (28), which can predict secondary structures with pseudoknots.

We used a structure descriptor of the HDV ribozyme based on a prior comparative analysis of sequences (33) with minor changes (Figure 2A). For input to RNAiFold, the base pairs in P1 (nucleotides 1–7 and 29–35) and P1.1 (nucleotides 20–21 and 36–37) regions were removed. To later include these pairs, which are needed for function, we generated all 8192 sequences that can form all Watson–Crick pairs. (To save computational time, we did not include GU pairs, except for G1-Y35.) The total number of generated pairs is therefore equal to $2 \times 4^6 = 8192$, where there are 6 Watson–Crick base pairing positions and Y35 is either C or U. Next, RNAiFold version 2.1 (13,43) was run locally for each of these 8192 pseudoknot sequences, but constraining these nucleotides in P1 to be unpaired. RNAiFold was downloaded from http://bioinformatics.bc.edu/clotelab/RNAiFold. We used default parameters in RNAiFold version 2.1, except for MAXsol (maximum number of solutions), which was set to 1 000 000. For the version of Vienna Fold, we used version 2.1.7, and we used the nearest neighbor parameters from 2004 (44). For a sequence of this length, however, the calculation time to generate these many sequences that fold with lowest free energy to the specified structure was prohibitive. RNAiFold also returned a large number of excellent candidate sequences. Therefore, we ended execution for each calculation at 6 h (on a Intel(R) Xeon(R) CPU E5-2695 v2 @ 2.40GHz processor using a single core), and used the sequences that were generated in that time. We note that, because we truncated the sequence search by RNAiFold, we could have used the -RandomAssignment flag to randomize the search order and therefore diversify the set of sequences that were generated (13).

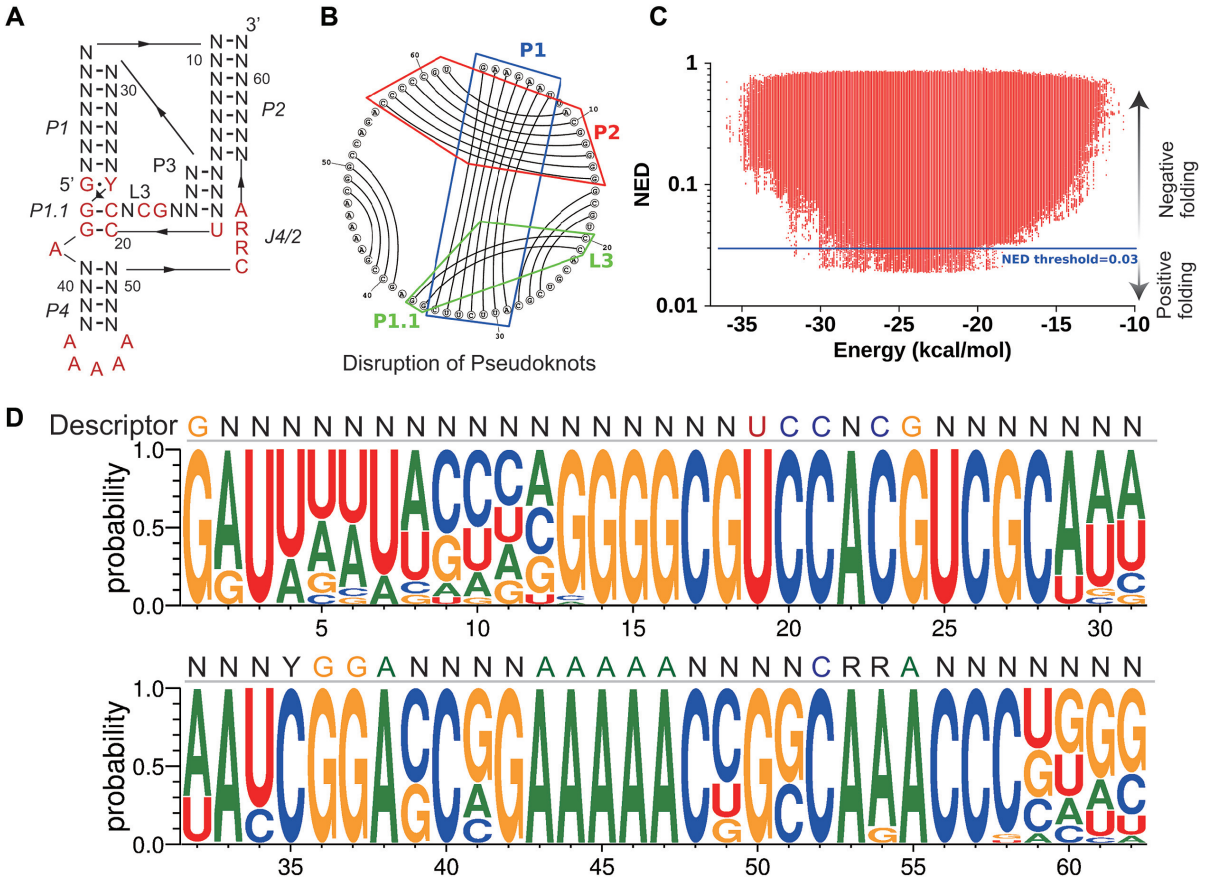

Figure 2. Designed RNAs have sequence differences in P1, P2 and P4 stems. (**A**) Structure descriptor for the inverse RNA folding. Fixed nucleotides are red. (**B**) Pseudoknot interactions in P1 (blue box) and between P1.1 and L3 (green box) were disrupted by randomizing the seed sequence. (**C**) Distribution of the designed RNAs plotted on NED versus folding free energy. Threshold value was set at 0.03 (blue line). (**D**) Nucleotide probability in all positively designed sequences analyzed by WebLogo 3. The primary sequence of the structure descriptor is provided in the top line.

Next, we used two filters to remove sequences that might not fold to the target structure. We used ProbKnot (with default parameters) to predict the structure for each sequence with all base pairing constraints removed (28). We required that over 93% of the Probknot-predicted base pairs, including the pseudoknots, be in the desired structure. Next, we identified sequences that could possibly fold into G-quadruplexes using Quadbase2, and removed those sequences (45). We used the default parameters of low stringency (Min stem length = 2, Min loop length = 1, Max loop length = 12) configuration in QuadBase2 and chose the greedy search algorithm option. All sequences are provided in Supplementary File 1.

Selection of candidate RNAs for biochemical analyses. After the computational design described in the previous section, candidate sequences for experiments were selected on the basis of the following screening criteria: normalized ensemble defect (NED) value, GC content and sequence difference. An NED value defines a length-normalized average number of incorrectly paired nucleotides at equilibrium evaluated over the ensemble of secondary structures for a given sequence (9). Because many of the sequences from RNAiFold were GC-rich, we clustered the data according to GC-richness. Specifically, designed sequences were classified into six subgroups depending on GC content in stem regions. We use the term 'GC-count' to designate the number of GC base pairs in an RNA, where there are 23 bp in total. We selected GC-counts of 14–19 for further consideration. The natural CPEB3 ribozyme has 15 GC base pairs of 24 bp in total. To pick up RNAs that do not have high sequence identity, phylogenetic trees were created (Supplementary Figure S1A) for each of the following GC-count subgroups GC14, GC15, GC16, GC17 and GC18. Specifically, a FASTA-formatted sequence file was analyzed by CLUSTAL W (web server at http://www.genome.jp/tools-bin/clustalw) with default parameters. A rooted phylogenetic tree (UPGMA) was created using the server and grouped into 10 branches for each GC count (see Supplementary Figure S1A). A sequence was randomly selected from each branch (total 10 sequences). The obtained sequences were analysed by WebLogo 3 (http://weblogo.berkeley.edu/) (46) with default parameters (Supplementary Figure S1B). In the case of GC19, there were only three designed sequences. For experiments, two candidate sequences were selected from each of the six GC-count subgroups on the basis of lowest and second lowest NED for a total of 12 candidate sequences (Supplementary Figure S1C).

Experiments

Preparation of RNA transcripts. RNA was obtained by *in vitro* transcription. DNA templates are provided in Supplementary File 2, where 'F' is for forward and 'R' is for reverse, the underlined sequences correspond to the T7 promoter and lower case shows 5′ leader sequences of the ribozymes. RNA sequences are also provided in Supplementary File 2. Transcription was conducted in a reaction mixture of 40 mM Tris–HCl (pH 7.5), 2 mM DTT, 1 mM spermidine, 2.5 mM NTPs, 25 mM $MgCl_2$ and 20 ng/μl of template DNA at 37°C for 4 h. Resultant transcripts were purified by 10% denaturing polyacrylamide gelelectrophoresis (PAGE) (8.3 M urea) and recovered by ethanol precipitation. The dry pellet was dissolved in water and stored at −20°C. To prepare full-length internally labeled precursor RNAs for self-cleavage assays, NTPs and $MgCl_2$ concentrations were reduced to 0.6 and 4.4 mM, respectively, to increase internal labeling efficiency and decrease self-cleavage during transcription. In addition, fresh [α-^{32}P]-GTP (Perkin-Elmer) was added to 10 μCi. Transcription at 37°C was performed for just 2 h to reduce self-cleavage. In case of CPEB3 ribozyme, 10 μM blocking oligonucleotide 5′-d(GTG GCC CCC TGT TAT C) was added into the transcription buffer to further inhibit self-cleavage.

Self-cleaving assay. Internally labeled RNAs were renatured at 85°C for 2 min in reaction buffer containing 25 mM Tris–HCl (pH 7.5), 140 mM KCl and 0.001% sodium dodecyl sulphate (final concentrations), and cooled to room temperature for 10 min. Self-cleaving reaction was initiated by addition of $MgCl_2$ to a final concentration of 50 mM. The reaction was quenched by 2× formamide buffer (70% formamide, 50 mM ethylenediaminetetraacetic acid (EDTA), 0.02% xylene cyanol, 0.02% bromophenol blue) at specific time intervals. Reaction products were fractionated by 10% denaturing PAGE (8.3 M urea) and visualized by a PhosphorImager (Typhoon 650; GE Healthcare). Fraction cleaved (f_{cleaved}) of internally labeled RNA is defined as

$$f_{\text{cleaved}} = \frac{I_{5'-\text{cleaved}} + I_{3'-\text{cleaved}}}{I_{5'-\text{cleaved}} + I_{3'-\text{cleaved}} + I_{\text{intact}}} \quad (1)$$

where $I_{5'\text{-cleaved}}$ is band intensity of 5′-cleaved RNA, $I_{3'\text{-cleaved}}$ is band intensity of 3′-cleaved RNA and I_{intact} is band intensity of uncleaved RNA. The observed first-order rate constants for the ribozymes (k_{obs}) were determined by single exponential curve fitting of f_{cleaved} versus time plots to

$$f_{cleaved} = A + Be^{-k_{obs}t} \quad (2)$$

where A is the fraction of ribozyme cleaved at completion, −B is the amplitude of the observable phase, k_{obs} is the observed first-order rate constant for ribozymes self-cleaving for the non-burst phase, and t is time. Self-cleaving assays were performed in duplicate, with each assay analyzed in duplicate (n = 2 experiments × 2 analyses = 4), or performed in triplicate (n = 3).

For rate-pH profile experiments, MES was used for pH 5.5, 6.0 and 6.5, and Tris–HCl for pH 7.5. The reaction was performed in the presence of 50 mM $MgCl_2$ at 37°C and quenched at specific time points out to 21 h. Although we tested the activity at pH 8.4 and 9.5, the RNAs were degraded during the course of the reaction and so the data were not used.

Monitoring of thermal denaturation by UV absorbance. To prepare RNA for thermal denaturation, 0.5 OD/ml (final concentration) of the RNAs were renatured at 95°C for 3 min in melt buffer containing 10 mM sodium cacodylate (pH 7.0) and 140 mM KCl, and cooled to room temperature for 10 min. Next, $MgCl_2$ was added into the mixture to 0, 2, 10, 25 or 50 mM, incubated at 55°C for 3 min, and cooled to room temperature for 10 min. Renatured samples were then centrifuged at 14 000 rpm for 10 min to degas. Samples were melted on an OLIS spectrophotometer with a data point acquired every 0.5°C and a heating rate of 0.5°C/min at 230–300 nm at a temperature range from 5°C to 95°C. Experiments were independently performed in duplicate. Data were smoothed by taking average of a data point with an 11-point window, and then the derivative was taken at every 0.5°C. The melting temperature (T_{M}) of each RNA in degrees Celsius was fit by a nonlinear Marquardt algorithm in Kaleidagraph using Equation (3)

$$f\ (T) = \frac{(m_u T + b_u) + (m_f T + b_f)\, e^{\left[\frac{\Delta H}{R}\right]\left[\frac{1}{T_m+273.15} - \frac{1}{T+273.15}\right]}}{1 + e^{\left[\frac{\Delta H}{R}\right]\left[\frac{1}{T_m+273.15} - \frac{1}{T+273.15}\right]}} \quad (3)$$

where m_u and m_f are the slopes of the upper (unfolded) and lower (folded) baselines, b_u and b_f are the y-intercepts of the upper and lower baselines, and ΔH is the enthalpy of folding in kcal·mol^{-1}, as described previously (47). R is the gas constant of 0.00198 kcal·K^{-1}·mol^{-1}. Data from the two independent melts were averaged, and errors are the range of two-independent experiments.

Non-denaturing PAGE. Cleaved RNAs were obtained from self-cleaved products of precursor RNAs. The cleaved RNAs are ready to use for 5′ labeling since they have 5′ OH. The RNAs were 5′-end labeled with [γ-^{32}P]-ATP (Perkin-Elmer) using T4 polynucleotide kinase (New England BioLabs) as per the manufacturer's instructions, and purified by 10% PAGE (8.3 M urea). The RNAs were renatured at 95°C for 2 min in THE buffer (pH 7.5) containing 34 mM Tris-base, 66 mM HEPES-acid and 1 mM EDTA, and cooled to room temperature for 10 min. Since 1 mM EDTA binds 1 mM Mg^{2+}, 3 mM $MgCl_2$ was then added into the mixture so that the final concentration of $Mg^{2+}{}_{\text{free}}$ is 2 mM. The RNAs were then renatured at 55°C for 3 min, and cooled to room temperature for 10 min. After the renaturation, 10% glycerol (final concentration) was added into the RNA solution and the RNAs were immediately loaded onto 10% non-denaturing gel (10% acrylamide-bisacrylamide (29:1), 1× of THE buffer and 3 mM total $MgCl_2$) with running at 10 W at 5∼8°C. Upon completion of electrophoresis, the gel was dried and the RNAs were visualized by PhosphorImager. Fraction full-folded RNA (f_{F}) was defined as

$$f_F = \frac{\text{maximum mobility RNA count}}{\text{input RNA count (entire lane intensity)}} \quad (4)$$

In-line probing experiment. The 5′-end labeled RNAs (∼230 kcpm) were incubated with in-line probing (ILP) buffer containing 50 mM Tris–HCl (pH 8.3), 200 mM KCl

and 2 mM $MgCl_2$ in 10 μl of reaction mixture at 37°C for 0, 24 and 47 h. The ILP reaction was stopped by addition of 2× formamide buffer. RNA alkaline ladder and RNase T1 ladder were loaded as standard ladders. Then, 2 μl of the sample was loaded onto a 10% polyacrylamide (8.3 M urea) sequencing gel and fractionated at 40 W for approximately 2 h. Experiments were performed in triplicate. Data at 0 and 24 h time points were analyzed by ImageQuant.

RESULTS

Computation

The computational part of our study involved two stages. First, we designed sequences that have the potential to adopt the double-pseudoknotted structure of the HDV ribozyme. This provided thousands of sequences. Second, we used filters to select those candidates most likely to adopt the final structure. These two stages are described in the first two sections of the 'Results' section.

Approach for inverse RNA folding and structure descriptor for HDV ribozyme. Our strategy for the inverse folding is summarized in Figure 1. To begin (Figure 1, step1), we prepared a descriptor that contains pseudoknots as shown in Figure 2A, following previous sequence comparison (33,48). The base pairing G1-Y35 in P1, as well as C20-G37 and C21-G36 in P1.1, where the pseudoknot base pairing is required to form the catalytic center, were fixed in sequence. The conserved U19C23G24 sequence in L3 is also fixed since U19 and G24 form reverse wobble base-pairing that is crucial for the catalysis (29,49,50). The sequences C52R53R54A55 in J4/2 were used for the structure descriptor. The C52 in designed ribozymes corresponds to C75 in the HDV ribozyme (our structure descriptor has a shorter P4 stem than the HDV ribozyme) and acts as a general acid in the catalytic mechanism (37). We also fixed five successive nucleotides in L4 loop as As to avoid alternative folds. We further inserted an A (A38) between P1.1 and P4 to prevent incorrect G-C base paring between the Gs of P1.1 and C52.

We next used RNAiFold to identify sequences that fold to the descriptor with the lowest free energy. Because RNAiFold does not allow pseudoknots (Figure 1, step2), we disrupted the pseudoknots (specifically base pairs in P1 and P1.1) in the descriptor (Figure 2B) (13). We generated all possible sequences that would replace the disrupted pairs with base pairs, a total of 8192, and ran RNAiFold for each of these sequences, specified as sequence constraints in RNAiFold. This generated sequences that fold to the pseudoknot-free descriptor with lowest free energy. In total, 62 099 216 sequences were generated in this way.

The sequence design outputs vast numbers of ribozyme sequences. To favor correct folding, sequences were filtered in three ways: First, the normalized ensemble defect (NED) (defined in the Materials and Methods) was calculated for each sequence (Figure 1, step 3 and Figure 2C). Next, the structure for each sequence was predicted using ProbKnot (28) (Figure 1, step 4), which is able to predict pseudoknots. Only sequences with the correctly predicted structures (>93% of predicted pairs matching the expected structure) were passed to the final filter. This filter was a heuristic to ensure we had a large set of sequences and that these sequences would each fold largely to the correct structure. In the third filter (Figure 1, step 5), sequences that were predicted by Quadbase2 (45) to form G-quadruplexes were removed. There were 13 552 designed RNAs remaining after the steps in the preceding section. Note that these sequences are thought to contain both active and inactive ribozymes. We thus conducted further filtration (Figure 1, step 6). In 13 552 sequences, we set a threshold value as 0.03. We regard the sequences (NED < 0.03) as positively designed ribozymes whereas the sequence (NED > 0.03) as negatively designed ribozymes. The negatively designed ribozymes are negative controls that are predicted to fold with lowest free energy to the correct structure, but with low fidelity, as quantified by NED. We show nucleotide probability of positively designed ribozymes in Figure 2D. We note that this set of sequences was limited by truncating the RNAiFold calculation times to 6 h, so the sequences we chose are not representative of the full space of sequence solutions. The sequences were divided into subgroups according to GC count (Figure 1, step 6.2) (see Materials and Methods). We then aligned the sequences and generated phylogenetic trees for each GC count to look for unique classes of ribozymes. We identified 10 branches on the phylogenetic trees and picked one sequence from each in an effort to disfavor sequences that are highly similar (Figure 1, step 6.3 and Supplementary Figure S1). Final candidates for positively designed ribozymes (called "group 1") were selected based on the two lowest NED values to give 2 sequences for every GC count, or 12 sequences in total. An additional group of eight sequences with high NED (group 2, Figure 1, step 6.1) were selected to serve as negatively designed, i.e. negative control, ribozymes (Supplementary Figure S2). We named the designed ribozymes as Designed HDV Ribozymes (DHRz). Finally, we used the NCBI nucleotide BLAST web interface to test if these designed sequences appear in the nucleotide collection databases (51). We did not find any significant similarity to our designed sequences, indicating that the resultant sequences are unique.

Experiments

Designed ribozyme has self-cleavage activity and specificity. To evaluate our RNA design, we tested the activity of our computationally designed ribozymes experimentally. We first characterized the DHRz 2511 (Figure 3A). An upstream sequence gggagcgccaaacau that was derived from drz-spur-3, a native HDV-like ribozyme (33), was attached to the 5′-end of the designed RNA (Figure 3A). Ribozyme activity was detected after a 60 min reaction only in the background of high Mg^{2+} (50 mM) (Supplementary Figure S3A). To see if the ribozyme possesses a similar cleaving mechanism as native HDV ribozyme, we prepared a C52G variant in which C52 (corresponding to C75 in native HDV ribozyme) was substituted by G (Figure 3A). Self-cleaving activity was completely abolished in the C52G variant, tested out to 3 h at all temperatures (37, 50 and 60°C) (Figure 3B Compare lane sets 1 and 3, and Supplementary Figure S3B). This supports a similar mechanism as wild-type HDV ribozyme. Moreover, similar to the HDV ribozyme (52), activity in the C52G variant was rescued

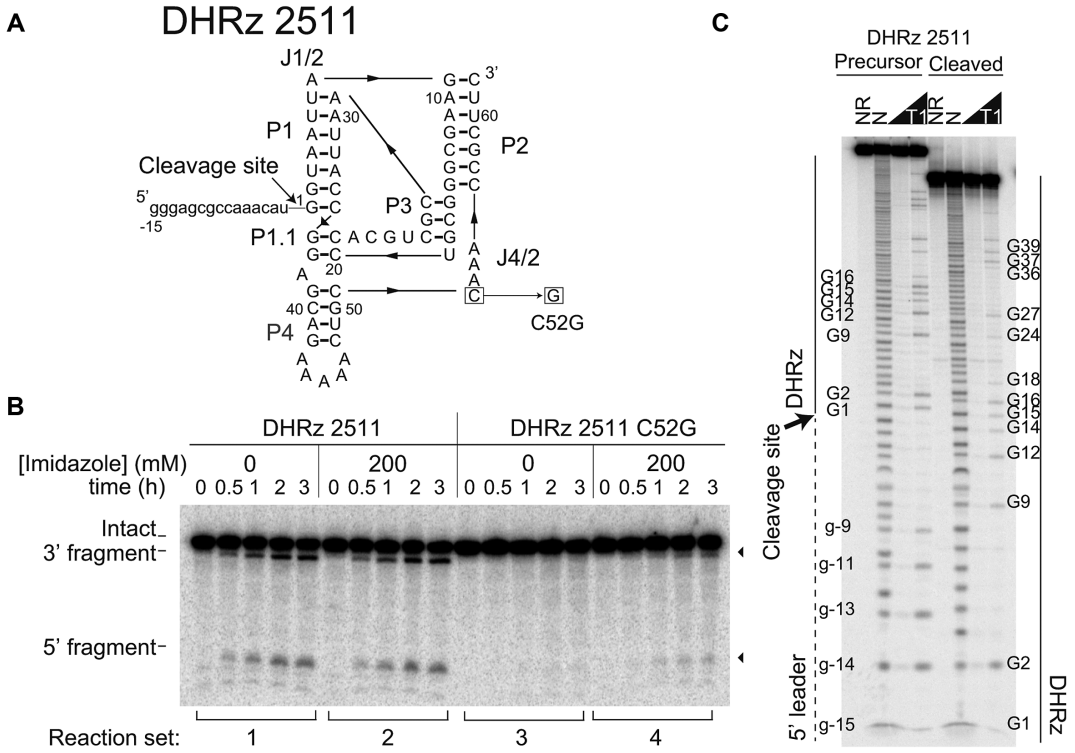

Figure 3. Designed ribozyme DHRz 2511 possesses catalytic activity with a native HDV ribozyme-like mechanism. (**A**) Secondary structure of DHRz 2511. The precursor sequence from drz-spur-3 was attached to 5′ end of the designed ribozyme. The C52G variant is indicated. (**B**) Self-cleaving activity of the ribozyme. Catalytic activity was tested in 50 mM Mg^{2+} and absence/presence of 200 mM imidazole. The reaction is divided into four sets: ±200 mM imidazole with wild-type or C52 variant. (**C**) Self-cleavage site of the ribozyme. Left half: Precursor ribozyme. Right half: Cleaved ribozyme. 'NR': No reaction, 'N': Alkaline ladder, 'T1': RNase T1 ladder. Reactions are at 37 °C.

by 200 mM imidazole, supporting that C52 acts as a general acid (Figure 3B). To determine the cleavage site of the ribozyme, digestion pattern of RNase T1 (G-specific) was compared between two RNAs: one from precursor DHRz 2511 and the other from cleaved DHRz 2511 that was obtained from the cleaved product after self-cleaving (Figure 3C). In the precursor ribozyme, the T1 ladder starts at G-15 in the 5′-leader sequence, whereas in the cleaved ribozyme, the T1 ladder starts at G1, indicating that the designed RNA cleaved the precursor sequence between G-1 and G1. This shows that DHRz 2511 has the same specificity for self-cleavage as the native HDV ribozyme. Overall, the DHRz 2511 ribozyme appears to have a similar mechanism and specificity as the native HDV ribozyme, albeit at a reduced rate.

Designed ribozymes demonstrate high-accuracy. We next tested the prediction accuracy of our inverse folding algorithm. We prepared the 12 ribozymes that scored NED < 0.03 (i.e. positively designed ribozymes) and 10 ribozymes that scored NED ≥ 0.03 (i.e. negatively designed ribozymes) (Figure 2C). The negatively designed ribozymes tested were not expected to fold into the catalytic structure, as there are low free energy structures that compete with the catalytic structure. Cleavage activity in the various designed ribozymes was tested at 37°C in the background of 50 mM Mg^{2+}. We found that 11 out of 12 positively designed ribozymes have self-cleaving activity, whereas only 3 out of 10 negatively designed RNAs do (Supplementary Figure S4). Thus, accuracy of positive selection is >90% (i.e. low false positives) while accuracy of negative control is ∼70% (i.e. low false negatives). These tests indicate that our computational design algorithm is successful in selecting active ribozymes.

Self-cleaving activity of designed ribozymes is less than that of native HDV ribozyme. We next compared self-cleavage activity between designed ribozymes and WT human CPEB3 ribozyme at 37°C (Figure 4). Self-cleavage for DHRz 2511 and CPEB3 ribozyme had monophasic behavior with first order rate constants of 0.0021 ± 0.0002 min^{-1} (50 mM Mg^{2+}) and 0.10 ± 0.027 min^{-1} (2 mM Mg^{2+}), respectively, showing that while the designed ribozyme cleaves at a rate ∼10^4 above background (53), it still reacts ∼50-fold slower than the CPEB3 ribozyme even at different Mg^{2+} concentrations (Figure 4A and Table 1). We also tested higher temperatures of 50°C and 60°C, however the reaction was not improved (Supplementary Figure S3B). We were curious if this kinetic property was observed in the other designed ribozymes. Thus, we repeated the assay using 14 different designed ribozymes (Figure 4B). Rate constants for the designed ribozymes lie in a narrow range of 0.001 ∼ 0.003 min^{-1}, and significantly slower than wild-type CPEB3 ribozyme. Slower rates were likely to be not related to the NED value and GC counts as shown by lack of a trend in Figure 4B (Table 1). To address the slower rates, we focused on J1/2, as such joining regions often play critical role in RNA folding (54). Indeed, the nucleotides in J1/2 have vari-

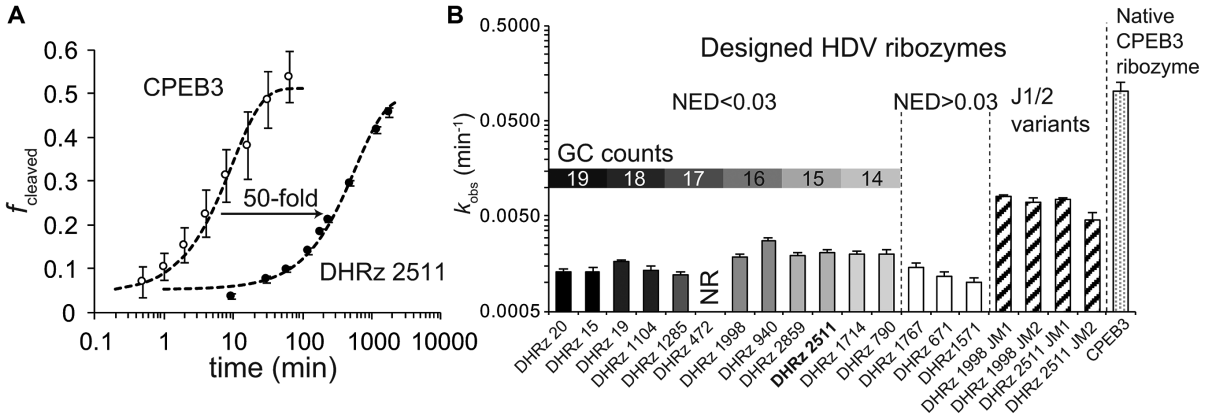

Figure 4. Cleavage activity of designed HDV ribozymes is less than native HDV ribozyme. (**A**) Plot of $f_{cleaved}$ versus time for DHRz 2511 (Closed circle) and CPEB3 ribozyme (Open circle). Data for CPEB3 are re-plotted from our previous study (59). Errors are S.D. ($n = 4$). Activity in CPEB3 and DHRz 2511 ribozyme were measured in backgrounds of 2 mM and 50 mM Mg^{2+}, respectively. (**B**) Rate constants for designed ribozymes, including J1/2 variants, and the CPEB3 ribozyme (far right). 'NR' indicates catalytic activity was not detectable. Errors are S.D. ($n = 4$).

ant lengths in naturally occurring HDV-like ribozymes (33). Recent work by Webb *et al.* revealed that nucleotides in extended J1/2 regions can base pair and form a P1.2 pairing that promotes catalytic formation. They also showed that extended single strand nucleotides in J1/2 stimulate self-cleavage activity in a nucleotide-length-dependent fashion (36). Our designed ribozymes possess only a single adenosine in J1/2. We thus prepared J1/2 variant ribozymes (JM) in which either one (JM1) or four additional As (JM2) were inserted into J1/2, and tested their self-cleaving activities (Figure 4B). The insertion stimulated rates 3.9- and 3.6-fold for JM1 and JM2, respectively. The joining regions are thus a critical part of design.

Designed RNAs fold non-cooperatively. Prior studies on proteins indicate that computationally designed proteins can lack cooperativity in their folding (55). We thus evaluated RNA folding cooperativity and structural compactness of our designed ribozymes (Figure 5). Thermal denaturation was carried out in 140 mM KCl with concentrations of Mg^{2+} ranging from 0 to 50 mM (Figure 5A–D and Supplementary Figure S5). Unfolding transition peaks were broad in 0–50 mM Mg^{2+}, shifting to higher temperature at high Mg^{2+} concentrations (Figure 5A–D). Note that $dA_{260\,nm}/dT$ values increase at high temperature (>80°C) in the high concentrations of Mg^{2+} (25 and 50 mM), which is likely due to non-specific cleavage of the RNA. As compared to wild-type CPEB3 ribozyme, the unfolding of the designed ribozymes is broad with multiple peaks, except for DHRz 2859, implicating that most designed RNAs unfold non-cooperatively with populated unfolding intermediates (Figure 5A–C and Supplementary Figure S5). As with the CPEB3 ribozyme, the DHRz 2859 shows cooperative unfolding transition with a single- and narrow-peak, although the activity in the DHRz 2859 is ∼60-fold slower than that of the CPEB3 ribozyme. These findings suggest that, at least in this research, we did not achieve a rational design of RNA that folds cooperatively using RNAiFold, similar to what has been reported for computationally designed proteins (55). Additional factors may thus need to be included in the computational design of RNAs to attain folding cooperativity. Our results also show that catalytic activity of the designed ribozymes does not correlate with extent of folding cooperativity.

We then performed non-denaturing PAGE to assess structural compactness of the designed ribozymes (Figure 5E). We used the CPEB3 ribozyme and a variant DHRz 2511 in which the pseudoknot interactions were disrupted, denoted 'DP', for comparison. The DP variant is designed to not form base-pairing in P1.1 and P2 stems, where C20C21 and C56CGCUUC62 were substituted by G20G21 and G56GCGAAG62, respectively (Supplementary File 2). The mobilities of the designed ribozymes are nearly identical to each other and to the wild-type, while the mobility of the DP variant is slower (Figure 5E). This holds in the presence of just 2 mM $Mg^{2+}{}_{free}$. Fraction of full folded RNA ($f_{full\ folded\ RNA}$) for most designed ribozymes is also similar to that of the CPEB3 ribozyme (Figure 5F). Similar native gel behavior suggests designed ribozymes share a similar topological structure with the CPEB3 ribozyme.

Weaker protonation for the general acid cytosine at position 52 in J4/2 in the designed ribozymes decreases the rate constants. After testing compactness in the previous section, we tested local structure of the ribozyme. We compared in-line reactivity between the designed ribozymes and the CPEB3 ribozyme. ILP provides a readout of mobile regions (loop, bulges and junctions) in RNA (47,56). We observe similar cleavage patterns in the designed ribozymes and the CPEB3 ribozyme, which correlate with the expected secondary structure; such specific cleavage patterns are not observed in the DP variant (Figure 6A). This result strongly suggests that our designed ribozymes fold into the double pseudoknot structure, consistent with the non-denaturing PAGE (Figure 5E). Interestingly, normalized in-line reactivity in J4/2 (especially A53) is significantly higher in the designed ribozymes than the CPEB3 ribozyme (Figure 6B and C), suggesting that the 2′-OH of the A53 in the designed ribozyme is solvent exposed. We thus hypothesized that incorrect formation of J4/2 perturbs formation of the catalytic core in the designed ribozyme. This fits the notion

Table 1. Kinetic parameters for designed and CPEB3 ribozymes

Ribozyme	Remarks	GC count	k_{obs} (min^{-1})	Relative activity to DHRz 2511
DHRz 20	NED < 0.03	19	0.0013 ± 0.0001	0.6
DHRz 15	NED < 0.03	19	0.0013 ± 0.0001	0.6
DHRz 19	NED < 0.03	18	0.0017 ± 0.0001	0.8
DHRz 1104	NED < 0.03	18	0.0013 ± 0.0002	0.6
DHRz 1285	NED < 0.03	17	0.0012 ± 0.0001	0.6
DHRz 472	NED < 0.03	17	N.D.	
DHRz 1998	NED < 0.03	16	0.0019 ± 0.0001	0.9
DHRz 940	NED < 0.03	16	0.0028 ± 0.0001	1.3
DHRz 2859	NED < 0.03	15	0.0019 ± 0.0001	0.9
DHRz 2511	**NED < 0.03**	**15**	**0.0021 ± 0.0002**	**1.0**
DHRz 1714	NED < 0.03	14	0.0020 ± 0.0001	1.0
DHRz 790	NED < 0.03	14	0.0020 ± 0.0002	1.0
DHRz 671	NED > 0.03	21	0.0014 ± 0.0002	0.7
DHRz 1767	NED > 0.03	17	0.0012 ± 0.0001	0.6
DHRz 1571	NED > 0.03	14	0.0010 ± 0.0001	0.5
DHRz 1998 JM1	A insertion in J1/2	16	0.0081 ± 0.0003	3.9
DHRz 1998 JM2	4As insertion in J1/2	16	0.0070 ± 0.0007	3.3
DHRz 2511 JM1	A insertion in J1/2	15	0.0075 ± 0.0004	3.6
DHRz 2511 JM2	4As insertion in J1/2	15	0.0045 ± 0.0009	2.2
DHRz 2511 M1	Domain substitution	12	0.028 ± 0.002	14
DHRz 2511 M2	Domain substitution	16	0.0017 ± 0.0003	0.8
DHRz 2511 M3	Domain substitution	15	0.0023 ± 0.0001	1.1
DHRz 2859 M1	Domain substitution	12	0.025 ± 0.007	13
DHRz 1714 M1	Domain substitution	11	0.038 ± 0.012	19
CPEB3	Native HDV-like ribozyme	15	0.10 ± 0.027[a]	50

N.D.: not detectable due to slow reaction. Errors are S.D. ($n = 3$ or 4). Negatively designed ribozymes (NED > 0.03) in which the activity is not detectable, are not listed. Reaction condition is pH7.5 and 50 mM Mg^{2+}. Activity relative to DHRz 2511.
[a]The k_{obs} was calculated from our previous work, where the reaction was tested in the background of 2 mM Mg^{2+} at 37°C (59).

that joining regions play critical roles in RNA folding, as discussed above for J1/2 and demonstrated elsewhere (54).

To test the importance of J4/2, we prepared three variants of DHRz 2511 where the sequence in specific regions was swapped with the analogous sequence in the CPEB3 ribozyme: M1, where J4/2 and portions of P4 and P2 were replaced; M2, where P1 and P2 were replaced; and M3 where P4 was replaced. The self-cleaving activities of the three J4/2 variants were then tested (Figure 7A and B). The rate constant for the M1 increased 14-fold as compared with DHRz 2511, whereas rate constants for M2 and M3 did not change significantly (Figure 7C and D). These findings are consistent with above ILP results. We tested other M1 variants (DHRz 2859 M1 and DHRz 1714 M1), which also increased the ribozyme activity at 13- and 19-fold, respectively (Supplementary Figure S6).

Next, we characterized the active M1 variant of DHRz 2511. The melting temperature of the DHRz 2511 M1 was decreased (~6°C) as compared with DHRz 2511, presumably due to low GC count (12 versus 15 GC counts) (Supplementary Figure S7 and Table 2). The ΔH for DHRz 2511 M1 is similar to that for DHRz 2511, and consistently smaller than the ΔH for CPEB3 ribozyme, consistent with non-cooperative unfolding (Table 2, last column). We tested also the thermal stability of the other M1 ribozymes. The DHRz 2859 M1 loses sharpness of the unfolding transition (Supplementary Figure S7D–F and Table 2), and the melting temperature of the ribozyme is decreased, as in DHRz 2511 M1. Finally, the DHRz 1714 M1 improves RNA folding cooperativity (Supplementary Figure S7G–I and Table 2). The mutation in J4/2 therefore affects RNA folding, with both positive and negative effects, which can change the population of misfolded ribozymes (Supplementary Figure S7). We also tested in-line reactivity in J4/2 in M1. Insertion of the CPEB3 sequence led to reduced, wild-type-like ILP reactivity in J4/2, consistent with native folding in this region (Supplementary Figure S8).

We hypothesized that the aberrant structure in J4/2 in the designed ribozymes might affect protonation of the general acid. To address this, we measured rate-pH profiles for the DHRzs and M1 (Supplementary Figure S9). At the pH range of 5.5–6.5, the logarithm of the rate constant for M1 increases with a slope of ~1.1, reflecting the fraction of the protonated $C52^+$, consistent with the native HDV ribozyme (37). In contrast, the rate constants for the DHRzs increase only slightly with pH, with slopes of 0.3–0.5, consistent with folding defects and kinetic complexity (Supplementary Figure S9). In sum, catalysis in the designed ribozyme is strongly influenced by the sequence in J4/2, which affects the protonation status of C75 at physiological pH.

DISCUSSION

In this study, we developed and implemented a computational RNA design for the pseudoknot-containing ribozymes. Our design strategy output over thirteen thousand sequences, which were selected based on the criteria of NED values, GC counts, and sequence difference. Experimental self-cleaving assays revealed accuracy of our RNA design: Most positively designed RNAs that we tested have self-cleaving activities (Figure 4 and Supplementary Figure S4), and the designed RNA was inactivated by the C75 counterpart C52G mutation and restored in the presence of imidazole, as previously described for the HDV ri-

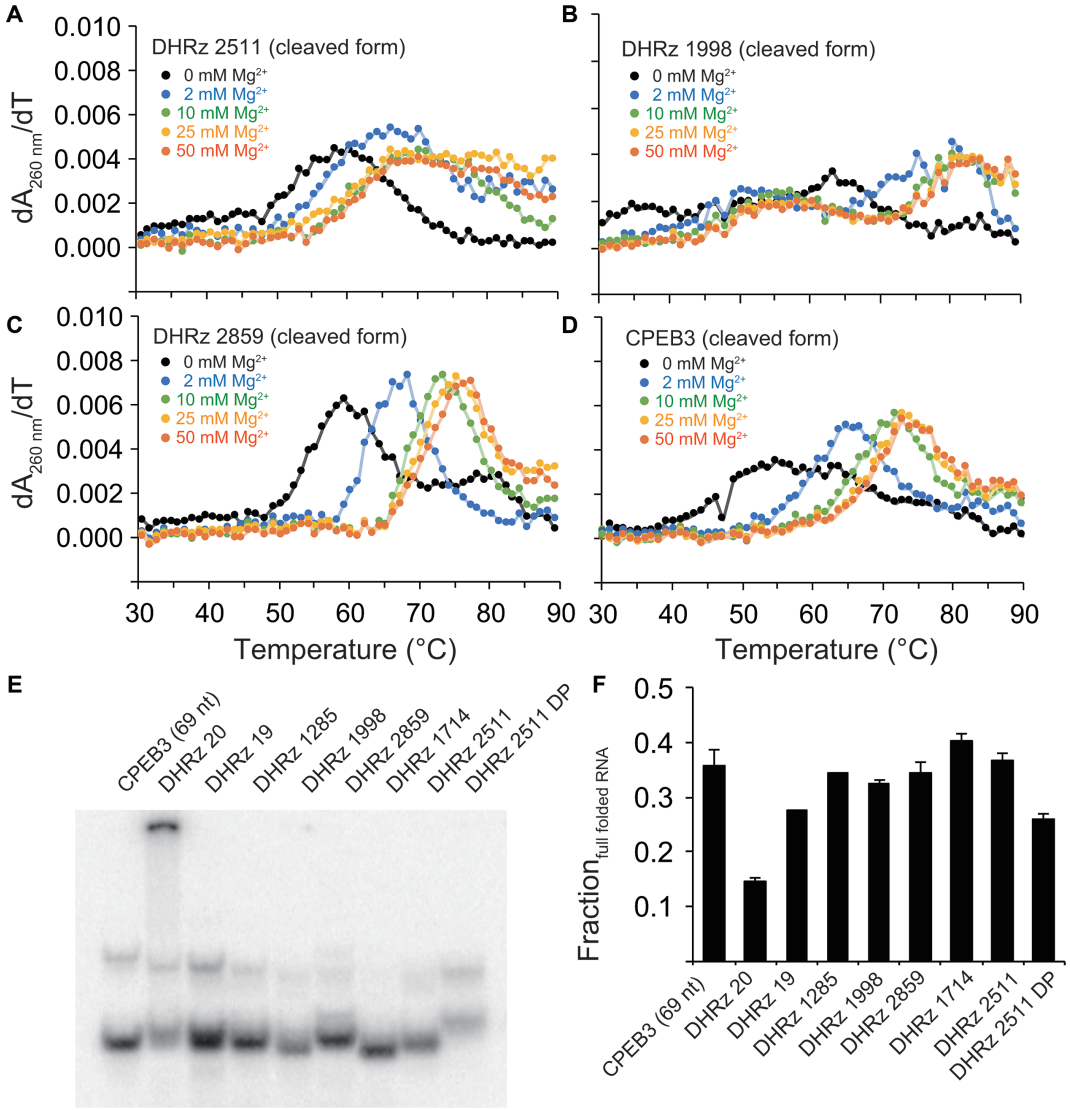

Figure 5. Designed ribozymes unfold non-cooperatively but are compact. (**A**–**D**) Denaturation curves of designed ribozymes and CPEB3 ribozyme measured in the background of 0 (black), 2 (blue), 10 (green), 25 (orange) and 50 mM Mg^{2+} (red). Cleaved ribozymes were used for the experiments. See Supplementary Figure S5 for additional RNAs. (**E**) Global folding of the ribozymes compared by non-denaturing PAGE. (**F**) Fraction $_{\text{full folded RNA}}$ calculated from non-denaturing PAGE. Errors are range of two-independent experiments.

Table 2. Profiles of thermal denaturation

Ribozyme	GC count (/23bp)	NED	Folding ΔG^a (kcal/mol)	Tm[b] (°C)	ΔH^c (kcal/mol)	$\Delta H/\Delta H_{CPEB3}$
DHRz 20	19	0.0201	−29.5	67.8 ± 0.7	−37.2 ± 1.7	0.6
DHRz 19	18	0.0235	−28.0	71.8 ± 0.4	−33.5 ± 0.9	0.5
DHRz 1285	17	0.0199	−25.7	65.6 ± 1.7	−20.5 ± 11.3	0.3
DHRz 1998	16	0.0206	−23.4	71.6 ± 5.4	−16.6 ± 4.2	0.3
DHRz 2859	15	0.0208	−22.7	67.0 ± 0.1	−68.1 ± 7.4	1.1
DHRz 2511	15	0.0237	−21.4	64.7 ± 1.4	−45.3 ± 2.8	0.7
DHRz 1714	14	0.0204	−19.5	62.8 ± 0.9	−16.7 ± 0.5	0.3
DHRz 2511 M1	12			58.4 ± 0.2	−38.5 ± 0.7	0.6
DHRz 2859 M1	12			57.1 ± 0.1	−44.1 ± 1.6	0.7
DHRz 1714 M1	11			50.3 ± 0.2	−47.9 ± 4.8	0.8
CPEB3	15[d]			65.6 ± 0.5	−62.9 ± 0.8	1.0

Tm values were calculated from the 2 mM Mg^{2+} condition. Errors are range of two-independent experiments ($n = 2$).
[a]Folding free energy for each sequence was estimated by RNAiFold.
[b]Tm was estimated from the UV melting experiments.
[c]ΔH values were calculated from the UV melting experiments.
[d]CPEB3 possesses 24 bp in total.

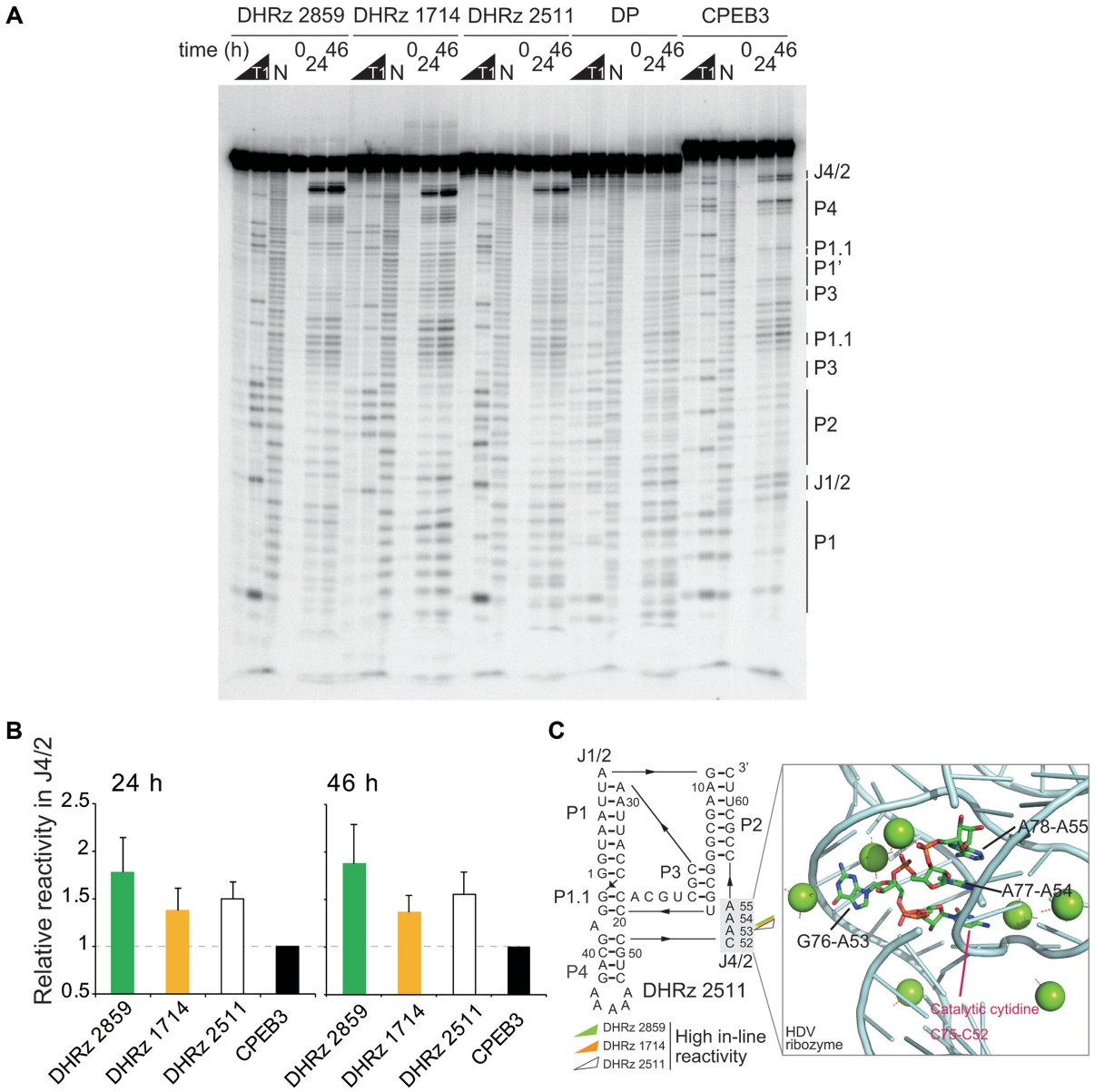

Figure 6. In-line reactivity in J4/2 in designed RNAs suggests structural flexibility. (**A**) ILP on designed ribozymes and CPEB3 ribozyme. RNase T1 ladders (G) and alkaline hydrolysis ladders (OH^-) are denoted by 'T1' and 'N'. DP is the pseudoknot-disrupted variant. (**B**) In-line reactivity in J4/2 relative to the CPEB3 ribozyme. Errors are S.D. ($n = 3$). (**C**) Strongest cleavage sites. Close-up for the J4/2 loop (PDB ID: 3NKB). 2′-OH of the G76 (=A53 in designed ribozyme) has high solvent accessibility. Magnesium ions are represented by the green spheres.

bozyme (52) (Figure 3B). These observations strongly support the conclusion that the self-cleaving mechanism of the designed ribozymes coincide with the native HDV ribozyme (i.e. the designed RNA have an ability to fold into the native-like catalytic structure). Furthermore, our ILP experiment supports the conclusion that these designed ribozymes can form the double pseudoknot structure, as the cleavage patterns of the RNAs were similar to those of the CPEB3 ribozyme (Figure 6A). Domain mutation, ILP and rate-pH profile analyses revealed that J4/2 is important to protonate C75 in HDV ribozyme (Figures 6 and 7; Supplementary Figures S6, 8 and 9).

We propose a possible mechanism of the insufficient protonation of C52 in the designed ribozymes (Figure 8). In the native HDV ribozyme, the general acid C75 is just 3.6 Å away from the scissile phosphate of G1, while G76 is flipped out to solvent side, and A77 interacts with A78 through π–π stacking (closest distance between these two base is 3.2 Å) (29). Furthermore, the ribose-phosphate backbone of A77–A78 nucleotides binds a magnesium ion (Figure 8A). This unique secondary structure probably moves C75 up into the catalytic cavity in the wild-type ribozyme, providing efficient protonation of the general acid. In contrast, in the designed ribozymes, high ILP reactivity suggests that A53, the counterpart of G76, is pushed out to the solvent

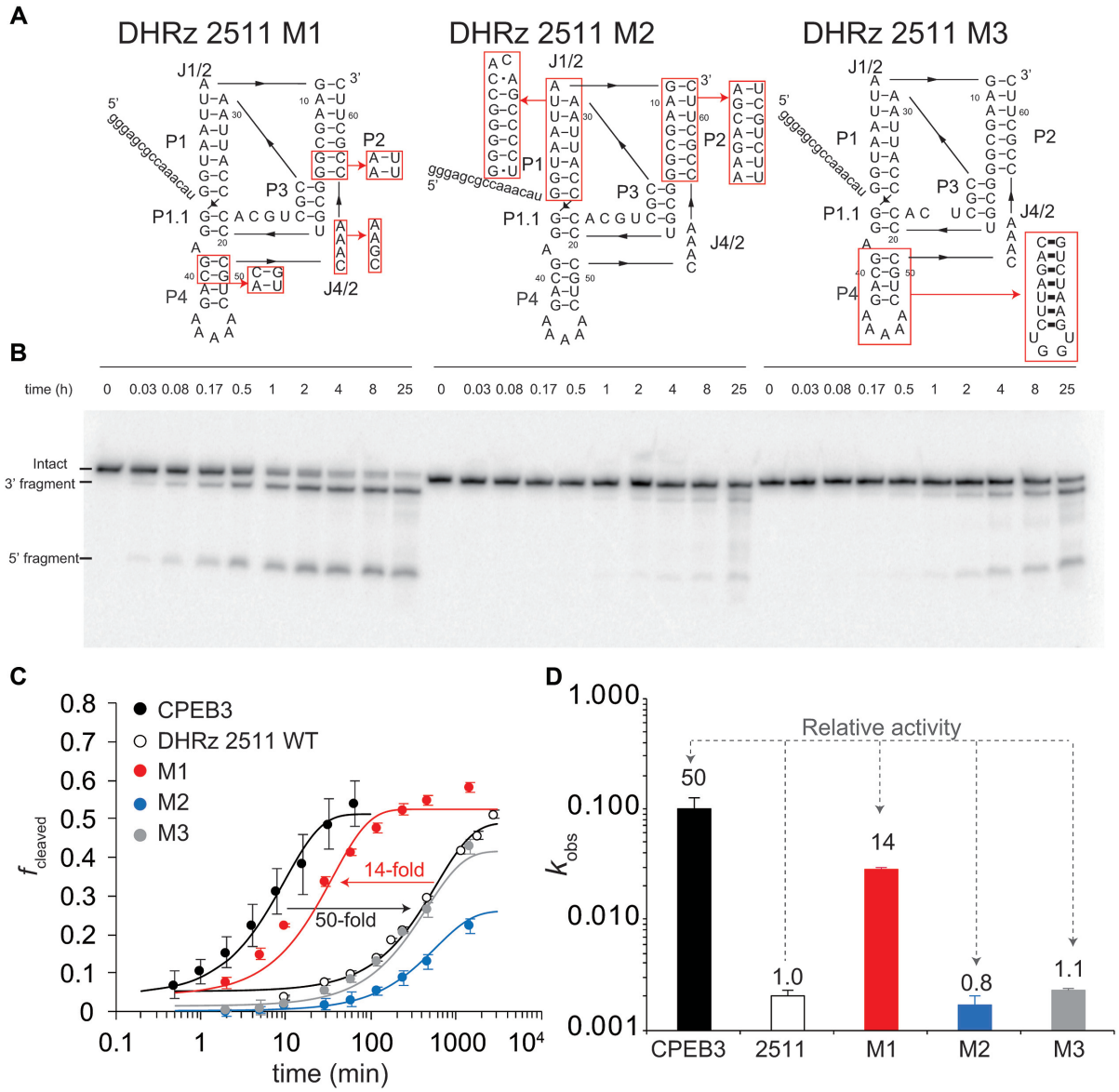

Figure 7. Substitution of J4/2 in the designed CPEB3 ribozyme sequence significantly increases catalytic activity. (**A**) DHRz 2511 M1 in which J4/2 and portions of P2 and P4 sequences were substituted with sequences from the CPEB3 ribozyme. DHRz 2511 M2 in which P1 and P2 sequences were substituted with sequences from the CPEB3 ribozyme. DHRz 2511 M3 in which P4 sequence was substituted with sequence from the CPEB3 ribozyme. (**B**) Self-cleaving assay for the three domain substituted variants. (**C**) Scatter graph for the variants (M1: red, M2: blue and M3: gray). The same data for the CPEB3 ribozyme and the unmodified DHRz 2511 in Figure 4 were plotted for this graph. Data were analyzed by single exponential curve fitting. Errors are S.D. ($n = 4$). (**D**) Plot of k_{obs} values. Relative activity (given at top of bars) was calculated relative to unmodified DHRz 2511. Errors are S.D. ($n = 4$).

side and might interact with a free Mg^{2+}, which is located nearby in J4/2 (29,57). The immediately adjacent general acid C52 could move with A53, perturbing protonation of the cytosine (Figure 8B). In the future, structural information and/or sequence constraints in joining regions should be taken into account in the structure descriptor, similar to recently described (54), in order to design fully functional ribozymes.

Sequence in the J4/2 in the positively designed ribozymes is almost fixed to CAAA (C52: 100%, A53: 100%, A54: ∼85% and A55: 100%) even though we used a structure descriptor where the J4/2 was allowed to be CRRA (Figure 2A and D), where 'R' is A or G. On the other hand, the J4/2 sequences in the negatively designed ribozymes show all four sequences (CAAA, CGAA, CAGA, and CGGA) where the nucleotide probability of CAAA is C52: 100%, A53: ∼70%, A54: ∼60% and A55: 100% (Supplementary Figure S2b). Note that our calculation by RNAiFold was limited to only 6 h, which did not provide sufficient time to fully explore the sequence space; therefore, we cannot conclude that these sequences represent the full solution space of sequences. However, it is clear that NED filtration led to CAAA sequence in the positively designed ribozymes. Rejection of G is likely because Gs in position 53 and 54 of J4/2 allow formation of base pairs to C20 and C21, which violate the descriptor. Additionally, these spurious GC pairs

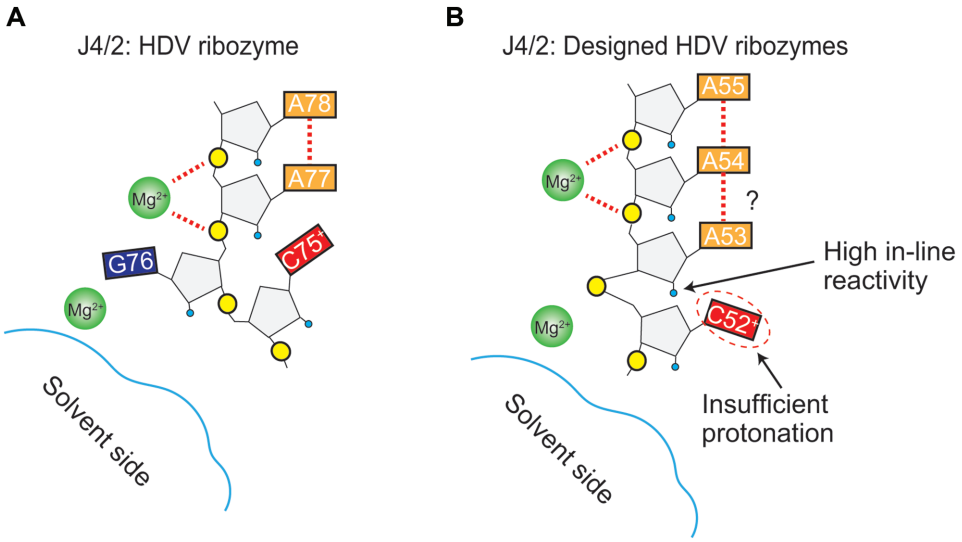

Figure 8. Proposed model of weaker protonation of C52 in designed RNAs. (**A**) J4/2 in the HDV ribozyme illustrated based on the crystal structure (PDB ID: 3NKB). Protonated $C75^+$ acts as general acid. G76 is flipped out to solvent side, and A77 is stacked with A78 via π–π interaction. A magnesium ion binds with the phosphates of both A77 and A78. (**B**) J4/2 in the designed RNAs is proposed to be structurally labile. C52 and A53 fluctuate in the catalytic cavity, and the 2′OH of A53 is oriented to solvent side, which could lead this region to have high in-line reactive for A53 and insufficient protonation for C52.

drive formation of a U19-A54 base pair, which also violates the descriptor and increases the ensemble defect of the sequence. Hence, G's in J4/2 were likely to be removed in our selection of lowest NED sequences. To test this in detail, we separated the CRRA constraint into four cases: CGGA, CGAA, CAGA and CAAA. Then, we selected randomly a subset of 1000 pseudoknot sequence constraints for each of the four cases from the original 8192 total constraints used originally. The results are shown in Supplementary Figure S10. Sequences with CGGA in J4/2 have the highest NED values (Supplementary Figure S10A) and sequences with CAAA in J4/2 have the lowest NED values (Supplementary Figure S10D). As a control, we replaced GG in CGGA-constrained sequences with AA and overall the NED values decreased (Supplementary Figure S11).

In our study, we used an HDV ribozyme descriptor that has a standard secondary structure, including a P4 stem. However, HDV-like ribozymes can have diverse sequence length and elements. For example, Riccitelli *et al.* (58) found that genomic minimal HDV-like ribozymes lacking the P4 stem have a similar metal dependency as the full length HDV ribozyme. The P4 stem is unlikely to interfere with our computational design since it does not have any tertiary interactions. Such minimal ribozymes, without extraneous nucleotides, should be able to reduce calculation time in future applications.

We have only one inactive sequence (DHRz 472) out of 12 positively designed ribozymes, and three active sequences (DHRz 1571, 1767 and 671) out of 10 negatively designed ribozymes. Although we have not clarified exact reasons why these RNAs are inactive/active, we have some hypotheses. For example, the DHRz 472 has two palindromic sequences in P1 and P2, and these palindromic regions possibly allow it to form intermolecular dimer structure, which diminishes the activity. Removing palindromes could be a future filter. On the other hand, DHRz 671 has a relatively high NED value (0.32) but has a low minimal free energy of −34 kcal/mol (Supplementary File 1) and relatively high PPV value (76.47%) in the negatively designed ribozymes. Therefore, the DHRz 671 might be able to fold in the precise structure. We cannot rationally explain why the DHRz 1581 and 1767 have cleavage activity. To address these hypotheses, further investigations are necessary.

Our biochemical analyses allowed us to greatly improve the rate in our designed ribozyme. The earlier work of Dotu *et. al.* succeeded to computationally design type III hammerhead ribozyme (HH) using RNAiFold, where the activity in the designed 10 HH sequences were tested, and the observed cleavage rates were in a broad range from 0.0027 min^{-1} to 0.25 min^{-1} (20). The fastest HH, which was just 10-fold less than wild-type HH, was selected with the constraint of the cleavage site of GUC, based on previous work on the efficiency of cleavage sites. Furthermore, they created a modular HH within a larger riboswitch structure, which showed a rate comparable to wild-type HH (20). Thus, structural stability, as suggested by the work of Dotu *et al.*, as well as consideration of a junctions' properties, as suggested by our work, may be complementary ways to facilitate design of RNAs.

Natural RNAs appear to fold cooperatively (47,59). This includes tRNAs and the natural CPEB3 ribozyme (Figure 5D). Gain of activity with J1/2 and J4/2 does not correlate with gain in folding cooperativity, similar to what has been seen in designed proteins (55). In the protein design study, the authors suggested that there may be a feature of natural selection that leads to folding cooperativity that is missing in current computer design algorithms. This could be similarly true for functional RNA. A goal for future studies is to identify this missing feature and to implement it into design algorithms in an effort to design cooperative folding RNAs. Finally, our approach herein considered only RNA folding thermodynamics. Co-transcriptional folding is known to be

important in many RNA folding processes *in vivo*. Future studies could use design algorithms to test whether such pathways result in even better folders.

CONCLUSION

In this study, we established a method for designing a pseudoknotted HDV ribozyme. RNAiFold (13,43) generated over 10 000 sequences based on the structure descriptor of HDV ribozymes. To accomplish the design, the two pseudoknot interactions were disrupted for the calculation, and all possible pseudoknot helices were systematically tested. Then, ProbKnot (28) and Quadbase2 (45) filter out those sequences that are not likely to fold into the precise structure or that potentially form G-quadruplex structure, respectively. Further selection by NED value is necessary to obtain positively/negatively designed ribozymes. Kinetics and ILP experiments reveal that junctions are critical for the catalysis and substitution of the junctions with native ribozyme sequence dramatically improves catalytic activity. This rational design is applicable to other pseudoknotted functional RNAs such as riboswitches and self-cleaving ribozymes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Kyle Messina for kind gifts of pH buffer for the rate-pH experiments, and the members of the Bevilacqua laboratory for valuable discussion.

FUNDING

NIH [R01-GM110237, R35-GM127064 to P.C.B., R01-GM076485 to D.H.M.]; Uehara Memorial Foundation (to R.Y.). Funding for open access charge: NIH [R35-GM127064].
Conflict of interest statement. None declared.

REFERENCES

1. Isaacs,F.J., Dwyer,D.J. and Collins,J.J. (2006) RNA synthetic biology. *Nat. Biotechnol.*, **24**, 545–554.
2. Chappell,J., Watters,K.E., Takahashi,M.K. and Lucks,J.B. (2015) A renaissance in RNA synthetic biology: new mechanisms, applications and tools for the future. *Curr. Opin. Chem. Biol.*, **28**, 47–56.
3. Bartel,D.P. and Szostak,J.W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science*, **261**, 1411–1418.
4. Wilson,D.S. and Szostak,J.W. (1999) In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.*, **68**, 611–647.
5. Hofacker,I.L., Fontana,W., Stadler,P.F., Bonhoeffer,L.S., Tacker,M. and Schuster,P. (1994) Fast folding and comparison of RNA secondary structures. *Monatshefte für Chem.*, **125**, 1434–4475.
6. Andronescu,M., Fejes,A.P., Hutter,F., Hoos,H.H. and Condon,A. (2004) A new algorithm for RNA secondary structure design. *J. Mol. Biol.*, **336**, 607–624.
7. Dirks,R.M., Lin,M., Winfree,E. and Pierce,N.A. (2004) Paradigms for computational nucleic acid design. *Nucleic Acids Res.*, **32**, 1392–1403.
8. Busch,A. and Backofen,R. (2006) INFO-RNA–a fast approach to inverse RNA folding. *Bioinformatics*, **22**, 1823–1831.
9. Zadeh,J.N., Wolfe,B.R. and Pierce,N.A. (2011) Nucleic acid sequence design via efficient ensemble defect optimization. *J. Comput. Chem.*, **32**, 439–452.
10. Taneda,A. (2011) MODENA: a multi-objective RNA inverse folding. *Adv. Appl. Bioinform. Chem.*, **4**, 1–12.
11. Bindewald,E., Afonin,K., Jaeger,L. and Shapiro,B.A. (2011) Multistrand RNA secondary structure prediction and nanostructure design including pseudoknots. *ACS Nano*, **5**, 9542–9551.
12. Lyngsø,R.B., Anderson,J.W., Sizikova,E., Badugu,A., Hyland,T. and Hein,J. (2012) Frnakenstein: multiple target inverse RNA folding. *BMC Bioinformatics*, **13**, 260.
13. Garcia-Martin,J.A., Clote,P. and Dotu,I. (2013) RNAiFOLD: a constraint programming algorithm for RNA inverse folding and molecular design. *J. Bioinform. Comput. Biol.*, **11**, 1350001.
14. Weinbrand,L., Avihoo,A. and Barash,D. (2013) RNAfbinv: an interactive Java application for fragment-based design of RNA sequences. *Bioinformatics*, **29**, 2938–2940.
15. Esmaili-Taheri,A., Ganjtabesh,M. and Mohammad-Noori,M. (2014) Evolutionary solution for the RNA design problem. *Bioinformatics*, **30**, 1250–1258.
16. Kleinkauf,R., Mann,M. and Backofen,R. (2015) antaRNA: ant colony-based RNA sequence design. *Bioinformatics*, **31**, 3114–3121.
17. Drory Retwitzer,M., Reinharz,V., Ponty,Y., Waldispühl,J. and Barash,D. (2016) incaRNAfbinv: a web server for the fragment-based design of RNA sequences. *Nucleic Acids Res.*, **44**, W308–W314.
18. Kuhlman,B., Dantas,G., Ireton,G.C., Varani,G., Stoddard,B.L. and Baker,D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science*, **302**, 1364–1368.
19. Churkin,A., Retwitzer,M.D., Reinharz,V., Ponty,Y., Waldispühl,J. and Barash,D. (2018) Design of RNAs: comparing programs for inverse RNA folding. *Brief. Bioinform.*, **19**, 350–358.
20. Dotu,I., Garcia-Martin,J.A., Slinger,B.L., Mechery,V., Meyer,M.M. and Clote,P. (2014) Complete RNA inverse folding: computational design of functional hammerhead ribozymes. *Nucleic Acids Res.*, **42**, 11752–11762.
21. Zaug,A.J. and Cech,T.R. (1995) Analysis of the structure of Tetrahymena nuclear RNAs in vivo: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. *RNA*, **1**, 363–374.
22. Powers,T. and Noller,H.F. (1991) A functional pseudoknot in 16S ribosomal RNA. *EMBO J.*, **10**, 2203–2214.
23. Haas,E.S., Morse,D.P., Brown,J.W., Schmidt,F.J. and Pace,N.R. (1991) Long-range structure in ribonuclease P RNA. *Science*, **254**, 853–856.
24. Romero,D.P. and Blackburn,E.H. (1991) A conserved secondary structure for telomerase RNA. *Cell*, **67**, 343–353.
25. Nameki,N., Chattopadhyay,P., Himeno,H., Muto,A. and Kawai,G. (1999) An NMR and mutational analysis of an RNA pseudoknot of Escherichia coli tmRNA involved in trans-translation. *Nucleic Acids Res.*, **27**, 3667–3675.
26. Staple,D.W. and Butcher,S.E. (2005) Pseudoknots: RNA structures with diverse functions. *PLoS Biol.*, **3**, e213.
27. Butcher,S.E. and Pyle,A.M. (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.*, **44**, 1302–1311.
28. Bellaousov,S. and Mathews,D.H. (2010) ProbKnot: fast prediction of RNA secondary structure including pseudoknots. *RNA*, **16**, 1870–1880.
29. Chen,J.H., Yajima,R., Chadalavada,D.M., Chase,E., Bevilacqua,P.C. and Golden,B.L. (2010) A 1.9 Å crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage. *Biochemistry*, **49**, 6508–6518.
30. Wilkinson,S.R. and Been,M.D. (2005) A pseudoknot in the 3′ non-core region of the glmS ribozyme enhances self-cleavage activity. *RNA*, **11**, 1788–1794.
31. Kleinkauf,R., Houwaart,T., Backofen,R. and Mann,M. (2015) antaRNA–Multi-objective inverse folding of pseudoknot RNA using ant-colony optimization. *BMC Bioinformatics*, **16**, 389.
32. Salehi-Ashtiani,K., Lupták,A., Litovchick,A. and Szostak,J.W. (2006) A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science*, **313**, 1788–1792.
33. Webb,C.H., Riccitelli,N.J., Ruminski,D.J. and Lupták,A. (2009) Widespread occurrence of self-cleaving ribozymes. *Science*, **326**, 953.

34. Lazinski,D.W. and Taylor,J.M. (1995) Regulation of the hepatitis delta virus ribozymes: to cleave or not to cleave? *RNA*, **1**, 225–233.
35. Ruminski,D.J., Webb,C.H., Riccitelli,N.J. and Lupták,A. (2011) Processing and translation initiation of non-long terminal repeat retrotransposons by hepatitis delta virus (HDV)-like self-cleaving ribozymes. *J. Biol. Chem.*, **286**, 41286–41295.
36. Webb,C.H., Nguyen,D., Myszka,M. and Lupták,A. (2016) Topological constraints of structural elements in regulation of catalytic activity in HDV-like self-cleaving ribozymes. *Sci. Rep.*, **6**, 28179.
37. Nakano,S., Chadalavada,D.M. and Bevilacqua,P.C. (2000) General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science*, **287**, 1493–1497.
38. Gong,B., Chen,J.H., Chase,E., Chadalavada,D.M., Yajima,R., Golden,B.L., Bevilacqua,P.C. and Carey,P.R. (2007) Direct measurement of a pK(a) near neutrality for the catalytic cytosine in the genomic HDV ribozyme using Raman crystallography. *J. Am. Chem. Soc.*, **129**, 13335–13342.
39. Chadalavada,D.M., Gratton,E.A. and Bevilacqua,P.C. (2010) The human HDV-like CPEB3 ribozyme is intrinsically fast-reacting. *Biochemistry*, **49**, 5321–5330.
40. Thaplyal,P., Ganguly,A., Golden,B.L., Hammes-Schiffer,S. and Bevilacqua,P.C. (2013) Thio effects and an unconventional metal ion rescue in the genomic hepatitis delta virus ribozyme. *Biochemistry*, **52**, 6499–6514.
41. Thaplyal,P., Ganguly,A., Hammes-Schiffer,S. and Bevilacqua,P.C. (2015) Inverse thio effects in the hepatitis delta virus ribozyme reveal that the reaction pathway is controlled by metal ion charge density. *Biochemistry*, **54**, 2160–2175.
42. Lorenz,R., Bernhart,S.H., Höner Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
43. Garcia-Martin,J.A., Dotu,I. and Clote,P. (2015) RNAiFold 2.0: a web server and software to design custom and Rfam-based RNA molecules. *Nucleic Acids Res.*, **43**, W513–W521.
44. Mathews,D.H., Disney,M.D., Childs,J.L., Schroeder,S.J., Zuker,M. and Turner,D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 7287–7292.
45. Dhapola,P. and Chowdhury,S. (2016) QuadBase2: web server for multiplexed guanine quadruplex mining and visualization. *Nucleic Acids Res.*, **44**, W277–W283.
46. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
47. Leamy,K.A., Yennawar,N.H. and Bevilacqua,P.C. (2017) Cooperative RNA folding under cellular conditions arises from both tertiary structure stabilization and secondary structure destabilization. *Biochemistry*, **56**, 3422–3433.
48. Webb,C.H. and Lupták,A. (2011) HDV-like self-cleaving ribozymes. *RNA Biol.*, **8**, 719–727.
49. Chen,J., Ganguly,A., Miswan,Z., Hammes-Schiffer,S., Bevilacqua,P.C. and Golden,B.L. (2013) Identification of the catalytic Mg^{2+} ion in the hepatitis delta virus ribozyme. *Biochemistry*, **52**, 557–567.
50. Veeraraghavan,N., Ganguly,A., Chen,J.H., Bevilacqua,P.C., Hammes-Schiffer,S. and Golden,B.L. (2011) Metal binding motif in the active site of the HDV ribozyme binds divalent and monovalent ions. *Biochemistry*, **50**, 2672–2682.
51. Johnson,M., Zaretskaya,I., Raytselis,Y., Merezhuk,Y., McGinnis,S. and Madden,T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.*, **36**, W5–W9.
52. Perrotta,A.T., Shih,I. and Been,M.D. (1999) Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. *Science*, **286**, 123–126.
53. Li,Y. and Breaker,R.R. (1999) Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2′-Hydroxyl group. *J. Am. Chem. Soc.*, **121**, 5364–5372.
54. Denny,S.K., Bisaria,N., Yesselman,J.D., Das,R., Herschlag,D. and Greenleaf,W.J. (2018) High-throughput investigation of diverse junction elements in RNA tertiary folding. *Cell*, **174**, 377–390.
55. Watters,A.L., Deka,P., Corrent,C., Callender,D., Varani,G., Sosnick,T. and Baker,D. (2007) The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell*, **128**, 613–624.
56. Regulski,E.E. and Breaker,R.R. (2008) In-line probing analysis of riboswitches. *Methods Mol. Biol.*, **419**, 53–67.
57. Skilandat,M., Rowinska-Zyrek,M. and Sigel,R.K. (2016) Secondary structure confirmation and localization of Mg^{2+} ions in the mammalian CPEB3 ribozyme. *RNA*, **22**, 750–763.
58. Riccitelli,N.J., Delwart,E. and Lupták,A. (2014) Identification of minimal HDV-like ribozymes with unique divalent metal ion dependence in the human microbiome. *Biochemistry*, **53**, 1616–1626.
59. Yamagami,R., Bingaman,J.L., Frankel,E.A. and Bevilacqua,P.C. (2018) Cellular conditions of weakly chelated magnesium ions strongly promote RNA stability and catalysis. *Nat. Commun.*, **9**, 2149.