

Automated design of hammerhead ribozymes and validation by targeting the PABPN1 gene transcript

Nawwaf Kharma^{1,*}, Luc Varin², Aida Abu-Baker³, Jonathan Ouellet⁴, Sabine Najeh⁴, Mohammad-Reza Ehdaevand², Gabriel Belmonte¹, Anas Ambri¹, Guy Rouleau³ and Jonathan Perreault^{4,*}

¹Electrical & Computer Eng. Dept., Concordia University, 1455 boul. de Maisonneuve O., Montreal, QC, H3G 1M8, Canada, ²Biology Department, Concordia University, 7141 rue Sherbrooke O., Montreal, QC, H4B 1R6, Canada, ³Montreal Neurological Hospital and Institute, 3801 University Street, Montreal, QC, H3A 2B4, Canada and ⁴INRS - Institut Armand-Frappier, 531 boulevard des Prairies, Laval, QC, H7V 1B7, Canada

Received June 30, 2014; Revised October 10, 2015; Accepted October 12, 2015

ABSTRACT

We present a new publicly accessible web-service, RiboSoft, which implements a comprehensive hammerhead ribozyme design procedure. It accepts as input a target sequence (and some design parameters) then generates a set of ranked hammerhead ribozymes, which target the input sequence. This paper describes the implemented procedure, which takes into consideration multiple objectives leading to a multi-objective ranking of the computer-generated ribozymes. Many ribozymes were assayed and validated, including four ribozymes targeting the transcript of a disease-causing gene (a mutant version of PABPN1). These four ribozymes were successfully tested *in vitro* and *in vivo*, for their ability to cleave the targeted transcript. The wet-lab positive results of the test are presented here demonstrating the real-world potential of both hammerhead ribozymes and RiboSoft. RiboSoft is freely available at the website <http://ribosoft.fungalgenomics.ca/ribosoft/>.

INTRODUCTION

There are many uses for hammerhead ribozymes (hhRz), as they are short catalytic RNA molecules which anneal to and cleave other RNA molecules—typically transcripts of genes—leading to the knockdown of gene expression. There are other means of gene suppression via transcript cleavage, and they include the widely used RNAi (1). RNAi is effective at low concentrations, can be delivered via multiple pathways, allow for tissue specific expression and are generally accepted to be highly specific and non-toxic (2). Nev-

ertheless, ribozymes (3), including the hhRz (the focus of this paper) (4), have relatively simple catalytic domains and can be used to target specifically RNA. They can potentially discriminate between single-base polymorphisms, target introns and sub-cellular compartments, do not require helper proteins to function and hence, can be tested *in vitro* (2). They are also easily modifiable and have been modified in ways that allow for more efficient (5,6), more specific (7,8) as well as allosteric applications involving RNA (9), and small molecules (10) or protein triggers (11). Finally, ribozymes can work in organisms that lack the typical RNAi mechanism found in eukaryotes, such as bacteria (12) or even in simple solutions *in vitro*. For these reasons, hhRz are seen as general and malleable means of transcript modification and gene therapy (13), with applications in biosensors (10) and synthetic biology as well (14).

The hhRz is one of several structural motifs of ribozymes that can cause both cis- and trans-cleavage of RNA molecules with certain properties (3). The model of the hhRz, consists of three largely independent parts: stem I, stem II (including the catalytic core) and stem III. A trans-cleaving ribozyme annealed to a target substrate strand is shown in Figure 1. Stems I and III are chosen to be complementary to the substrate strand, which must include an NUH triplet. Stem II is found next to the catalytic core's conserved GAAA and CUGANGA. Generally speaking, a designer would pick an empirically proven stem II and catalytic core, then vary the two 'arms' (call them I and III) so they would form stems I and III, with pre-specified characteristics, with the targeted substrate.

Both target specificity and turnover (re-use) rate of a hhRz depend on the length and sequence of its two arms. Optimal turnover requires that product dissociation not be slow. For long substrates with extensive secondary structure, such is the case with most mRNAs, the optimal flank-

*To whom correspondence should be addressed. Tel: +450 687 5010 (Ext 4411); Fax: +450 686 5301; Email: jonathan.perreault@iaf.inrs.ca
Correspondence may also be addressed to Nawwaf Kharma. Tel: +1-514-848-2424 x3117; Fax: +514-8482802; Email: kharma@ece.concordia.ca
Present address: Jonathan Ouellet, Department of Chemistry and Physics, Monmouth University, 400 Cedar Avenue, West Long Branch, NJ, 07764, USA.

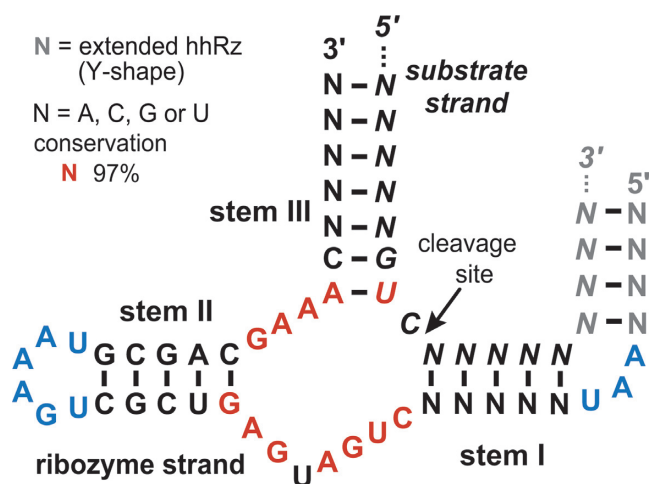


Figure 1. Model of trans-acting hammerhead ribozyme used by RiboSoft. Nucleotides in red correspond to the conserved catalytic core, other nucleotides may vary. Nucleotides in blue are important for the tertiary interaction between stems I and II. The sequence of stem II pictured here corresponds to the optional extended hhRz model (39) used by RiboSoft and includes the extra nucleotides in gray as well.

ing arms may be longer than 10 nt, but even in this case, increasing the arms too much reduces turnaround rate greatly.

This paper introduces a web-service that makes it easy for interested parties—not only ribozyme experts—to (i) design hhRz that target a particular gene, or any RNA sequence, and (ii) make an informed choice as to which of the possible ribozymes are best for a particular application. All this while automatically taking in account the multiple variables that determine ribozyme efficacy (target structure, ribozyme-target annealing and ribozyme structure, including the tertiary interaction of so-called extended ribozymes). In addition, we have used this tool to design hhRz, which were tested for efficient cleavage of three transcripts, as well as *in vivo* for the PABPN1 gene, the root cause of OPMD.

OPMD (Oculopharyngeal muscular dystrophy) is a late onset (around 50 years old) muscle disease with occasional involvement of the nervous system that can lead to obligatory wheelchair usage (15,16). Genetically, diagnosis requires detection of an expansion of a GCN trinucleotide repeat in the first exon of PABPN1. Normal alleles contain 10 GCN trinucleotide repeats; autosomal dominant alleles range in size from 12 to 17 GCN repeats; autosomal recessive alleles comprise 11 GCN repeats (16).

Spinocerebellar ataxia type 3 (SCA3), like OPMD, is a trinucleotide repeat neurodegenerative disease caused by the polyglutamine (versus polyalanine) expansion in the deubiquitinating enzyme, Ataxin-3 (*ATXN3*). Using the RNA interference mechanism, Rodríguez-Lebrón, E *et al.* made microRNA mimics that targeted the 3'-untranslated region of a human *ATXN3* expressed in the cerebellum of transgenic mice (17). This resulted in silencing of the human gene and cleared the abnormal nuclear accumulation of the mutant protein. Most OPMD-afflicted patients have a wild type PABPN1 allele and a mutant (disease-causing) PABPN1 allele, so targeting the 3'-UTR of the mRNA of PABPN1 (with microRNAs or ribozymes) will suppress

both alleles. Hu *et al.* (18) demonstrated that the targeting of the trinucleotide repeats can *preferentially* inhibit mutant *ATXN3*. However, an ideal solution is one that would *exclusively* impact the mutant allele or *compensate* for the knock-down effect on the wild type allele. This could be achieved by cleaving both alleles with a hhRz and providing a replacement mRNA with sequence modification rendering it immune to cleavage. HhRz were shown to efficiently suppress beta-secretase gene expression in HEK293 (human embryonic kidney) cells (19). In the sequel, we show that RiboSoft can generate the designs of active hhRz that were proven effective *in vitro* and *in vivo* against PABPN1 sequence. We also tested a dozen ribozymes that proved capable of cleaving RFP or a 16S rRNA fragment *in vitro*. Finally, we show that the combined use of ribozymes can achieve near complete depletion of the target RNAs.

MATERIALS AND METHODS

RiboSoft usage

Before the design algorithm can run, valid input from the user must be entered. The required input takes the form of (i) a target nucleotide sequence, which may be provided by: direct entry into a text box, uploading a pure text (.txt) or FASTA format file, or typing in a GenBank sequence number (Figure 2); and (ii) a set of design parameter values (Figure 3).

The design parameters are: (i) the target environment: whether it is *in vitro* or *in vivo*, and if it is *in vivo*, then the name or code of the intended organism; (ii) environment properties: temperature as well as oligomer, sodium and magnesium ion concentrations; (iii) potential cut-sites (e.g. GUC); (iv) ribozyme structure: minimal or extended (i.e. involving loop-loop interaction, designated as wishbone-shaped) (5); (v) advanced options: the minimum and maximum lengths of helices I and III (the targeting arms of the ribozyme), the presence or absence of a T7 promoter (usually used for *in vitro* experiments) and whether specificity should take into consideration off-target sites that allow annealing only or annealing and cleavage by the designed ribozyme. After the parameters are chosen, they are displayed on a summary page (Figure 4) before requesting the user's email to submit the ribozyme design request.

A target sequence must be provided, but most of the design parameters have default values, which are acceptable for most applications. The design algorithm then initiates the actual design and optimization procedure. This procedure has six steps, which starts with the generation of a large number of candidate ribozymes, assessing them using various measures of quality, including (but not limited to) target cut-site accessibility and specificity to the target sequence as well as proper folding of the ribozyme structure. The procedure concludes with the multi-objective assessment and ranking of resulting candidate ribozymes (Figure 5). The algorithms for each part of the procedure are explained below.

Candidate generation

In principle, a hhRz can cleave, with different efficiency levels (measured by different studies with varying results



Step 1 - Selecting the sequence

Search for sequence number.
You may also find your sequence [here](#)

(example: M73307)

Or

Select a file with your sequence to upload.
(TXT or FASTA format)

Paste your sequence here...

Figure 2. Snapshot of the input screen (step 1) of RiboSoft: asking the user for a sequence, a file or a sequence code.

[20,21]), at locations on RNA strands with an NUH motif. An RNA strand may have any number of different cut-sites within it, repeated any number of times. The most widely targeted cut-site is GUC, as it is thought to be the most susceptible to hhRz cleavage. For all cut-sites, all possible candidates with perfectly hybridizing arms, of lengths between the minimum and maximum values specified by the user, are generated. The total number of candidates is calculable using the formula:

$$\approx (\text{arm-length range})^2 \cdot (\text{number of cut-sites}) \quad (1)$$

This implies that the number of candidates is also proportional to the target sequence length, since more cut-sites will be found. To provide a rough estimate of the number of candidates that are typically generated by RiboSoft, with a conservative length of 3–10 for the hybridizing arms and two cut-site types, a sequence of 3000 nucleotides yields about 100 cut-sites for two cut-site types for a total of 6400 ribozyme candidates.

Annealing temperature

A trans-cleaving hhRz may be viewed as a three-part construct: a fixed largely double-stranded central segment (core and stem II), plus two single-stranded arms (stems I and III) up- and down-stream of the core. It is those two arms that provide the selectivity that permit the ribozyme to strongly anneal to a particular site on a target RNA molecule amongst myriad RNA strands in the medium and hence, cleave that target specifically.

The shorter an arm is, the lower the degree of selectivity and strength of annealing (annealing temperature) between the arm and the target RNA. In converse, the longer an arm is the more selective the annealing and the more heat is re-

quired to de-anneal the arm from the target RNA. Thus, it may appear that the solution is to automatically give every hhRz very long arms, but the catch is that longer single-stranded arms make it more likely that an arm would anneal to itself and/or with the other arm (partially or fully). Also, ribozymes with very long arms will anneal so strongly to their target that they will not turnover to anneal to new targets after cleavage.

In brief, the annealing temperature of a hhRz may be approximated by the sum of the annealing temperatures of the two arms of the ribozyme. All environmental conditions (e.g. ionic concentrations) being equal, the annealing temperature is a function of the nucleotide make-up and lengths of the arms of a ribozyme. A number of formulas have been used and compared (22,23) to calculate the annealing temperature of RNA (and DNA) molecules, including the fairly standard base-stacking approach employed by our algorithm (adapted from <http://www.basic.northwestern.edu/biotools/oligocalc.html> (24)).

Since calculation of the annealing temperature is computationally simple (relative to RNA folding) we use this calculation in a discriminatory manner to weed out all candidate hhRz with arms that are too short or too long. In fact, the range of acceptable annealing temperature we use is from -5°C to 60°C above the temperature of the intended environment for each for each arm separately. It is acceptable for a single arm to have a low T_m if the other arm makes up for it with a highly above average T_m . The exact values were determined by empirically testing many ranges to avoid eliminating perfectly sound sequences.

For a given cut-site on a target RNA strand, the sequence of any arm of a hhRz is dictated by the RNA sequence around the cut-site, as a hhRz should anneal perfectly to its target. Hence, every hhRz with target-matching arms and

Step 2 - Design Options

Please select the options that will be applied during the design of the ribozyme.

Regions for Target Design ?

- 5'UTR
- Open Reading Frame (ORF)
- 3'UTR

Target Environment ?

- In-vitro
- In-vivo
- Human ▼

Environment Properties ?

- Temperature: 37 °C
- Na⁺ Concentration: 150 mM
- Mg²⁺ Concentration: 0 mM
- Oligomer Concentration: 200 nM

Potential Cut-sites ?

- Preferred cut-site(s):

GUC
 CUC
 AUC
 UUC
 ...

Ribozyme Shape ?

- Minimal (before 2003)
- Wishbone (discovered in 2003)

Advanced Options - Arms Length ?

Figure 3. Partial snapshot of the input screen (step 2) for RiboSoft: allowing for user specification of various parameters.

Step 5 - Results

Design Summary

Design Option	Value
Sequence Length	371
Target Region	3', ORF, 5'
Target Environment	In-vivo
Temperature	37 °C
Na ⁺ Concentration	150 mM
Mg ²⁺ Concentration	1 mM
Oligomer Concentration	200 mM
Cut-sites	GUC
Fold Shape	Basic
Helix I Length	[5, 10]
Helix III Length	[5, 10]
T7 promoter used?	No
Specificity	Cleavage only

Figure 4. Snapshot of the input screen (step 3) for RiboSoft: confirmation of user input.

an overall annealing temperature within the pre-specified range is an acceptable hhRz candidate generated by this phase of the design process.

Ribozyme structure

Every hhRz candidate surviving the annealing temperature phase is evaluated for its structure. This is a measure of how open (single-stranded) the arms are, which makes it easier for a ribozyme to find and anneal to its intended target sequence and hence, perform its trans-cleavage function. The Sfold program (25) is used on the server side to fold all the candidates; it uses a statistical Boltzmann ensemble model to generate possible secondary structures for the candidate RNAs. Sfold reports 10 secondary structures, each with its own probability. This probability is used as a weight when considering the structure of a ribozyme.

RiboSoft gets the quality of one or many secondary structures for one hhRz with an algorithm that uses the sum of the melting temperatures of all continuous stretches of double-stranded RNA as an (inverse) measure of the openness of the two arms of the ribozyme. We allow up to one mismatch in each 'continuous' stretch of double-stranded RNA. All of the temperature values used are in Kelvin, as cancellation may erroneously occur with negative temperatures. A pseudo-code description can be found in supplementary information.

For each secondary structure of the 10 secondary structures associated with a given hhRz, we compute one overall melting temperature. Naturally, continuous pairs in the hhRz (Helix-II) will not be counted, since they are meant to be there. However, a penalty of 50 degrees is added to any structure lacking the hammerhead core, i.e. where the conserved bases of the core are base paired or where Helix-II is not formed. Since the resultant temperature here is associated with only one of 10 possible structures, a weighted sum (Q_f) of all 10 temperatures is calculated, for each hhRz candidate, using the formula:

$$Q_f = \sum_{i \in Q} f_s \cdot Q_s \quad (2)$$

where Q_s is the melting temperature of one secondary structure; f_s is the probability associated with that secondary structure; Q is the set of all 10 structures. Naturally, lower values for Q_f are preferred and as such, this measure of quality should be minimized.

Target accessibility

Accessibility is a measure of how easy it is to access the target. Targets that fold unto themselves strongly in the region around and including a cut-site will reduce the efficiency of a ribozyme designed for that cut-site.

Two accessibility terms are computed here. One term, *cut-site inaccessibility*, inversely measures the openness of a target cut-site, to a particular hhRz candidate with specific arms' lengths. This is reflected by the annealing temperature of the region of the target cut-site, the region to which the arms of a hhRz must anneal. The other term, *disruption energy* ($\Delta G_{\text{disruption}}$) is based on research done by (26), and is in fact: the free energy cost to melt any local secondary structure at the target site, making it accessible to hhRz binding.

Design Results

Show entries Filter by Cut-site Number:

Cut-site Number	Sequence	Melting Temperature (Left) (°C)	Melting Temperature (Right)(°C)	Accessibility 1	Accessibility 2	Ribozyme Shape Quality	Weighted off-target hits	Overall Rank
GUC4	CAUGCUGCUCUGAU GAGGCCGAAAGGCC GAAACUCUUUCUG	32.97	32.97	1	1	1	2.94	1
GUC4	CAUGCUGCUCUGAU GAGGCCGAAAGGCC GAAACUCUUUCUG	32.97	36.97	0.994	1	0.626	2.94	1
GUC3	UGAUCCACUCUGAU GAGGCCGAAAGGCC GAAACUGGCCUAGG	30.97	40.97	0	0.653	0.59	8.58	1
GUC3	UGAUCCACUCUGAU GAGGCCGAAAGGCC GAAACUGGCCU	30.97	30.97	0.018	0.653	0.771	8.58	2
GUC3	UGAUCCACUCUGAU GAGGCCGAAAGGCC GAAACUGGCCUA	30.97	32.97	0.015	0.653	0.522	8.58	2
GUC3	UGAUCCACUCUGAU GAGGCCGAAAGGCC GAAACUGGCCUAG	30.97	36.97	0.007	0.653	0.452	8.58	2
GUC1	GAGGAAGUCCUGAU GAGGCCGAAAGGCC GAAACUCUGUGU	32.97	38.97	0.272	0.402	0.483	11.2	1
GUC1	GAGGAAGUCCUGAU GAGGCCGAAAGGCC GAAACUCUGUGU	32.97	32.97	0.577	0.402	0.473	11.2	2
GUC1	GAGGAAGUCCUGAU GAGGCCGAAAGGCC GAAACUCUGUU	32.97	34.97	0.282	0.402	0.464	11.2	2
GUC1	GAGGAAGUCCUGAU GAGGCCGAAAGGCC GAAACUCUGUG	32.97	30.97	0.577	0.402	0.465	11.2	3

Showing 1 to 10 of 14 entries ◀ Previous Next ▶

Figure 5. Sample output from RiboSoft. The cut-site number (starting at 0 for the first cut-site) and corresponding ribozyme sequence is shown for each candidate. All the parameters corresponding to this candidate are shown on the same row, Tm of both arms, accessibility (from 0, worst, to 1, the best), ribozyme shape (also from 0 to 1) and the overall rank. Off-target hits are shown only when the ribozyme was selected for *in vivo* use, 1 normally means that only the targeted gene is hit, higher numbers imply potential non-specific activity.

The computation of cut-site inaccessibility is done per candidate hhRz, as different ribozymes targeting the same cut-site will have different arms' lengths. The substrate is also folded using Sfold, which returns 10 secondary structures for the cut-site region, hence 10 annealing temperature values are computed for that region, and these are summed using the probabilities of the corresponding secondary structures as weights. The result of this calculation is the raw cut-site accessibility value (both *cut-site inaccessibility* and *disruption energy* are succinctly described in *pseudocode*, supplementary information).

The calculation of disruption energy is done per cut-site. The reason for this is the significant computational cost associated with the computation of the secondary structure for a long RNA strand; typical gene transcripts are one to several thousand bases long. We use 23 bases for both arms' binding sites and hence compute $\Delta G_{\text{disruption}}$. This allows the evaluation of the structures likely to compete most strongly with hhRz binding. Since Sfold returns 10 possible secondary structures for the target transcript, we will also have 10 $\Delta G_{\text{disruption}}$ values. These values are used to calcu-

late a weighted linear sum in a fashion identical to what was done for cut-site accessibility.

As with ribozyme structure, cut-site inaccessibility and disruption energy should be minimized. In other words, RiboSoft looks both at the ease with which the ribozyme can bind to the target site, as well as the structure of hhRz and target.

Specificity assessment

For *in vivo* applications it is important that a given ribozyme affects the targeted transcript and no other. Typically, a hhRz would affect a transcript by cleaving it. It is possible, however, for a hhRz to anneal to a transcript and hence, lead to the transcript's destruction or to a reduction in its translational efficiency, without cleaving it. This is part of what is known as off-target effects. Hence, the 'advanced options' section of RiboSoft's input pages allow the user to check for either off-target hits that would potentially result in cleavage or *all* off-target hits, including those that do not lead to cleavage.

The NCBI BLAST web-service is queried with a high threshold (sensitivity) using the region around the targeted cut-site corresponding to the maximal possible lengths of hhRz arms. The results are weighted based on the number of matched pairs, thus a perfect match yields a weight of one (the targeted transcript). It should be noted that RNAs with accession numbers starting with XM or XR (RNA predicted sequences) are ignored in the count, but are still reported in the results' page.

The program will only return 0 if the RNA sequence does not exist, since there is not necessarily *a priori* knowledge of the transcript being targeted (users can enter arbitrary sequences directly). Optimally, the value of specificity is 1, entailing a perfect match to only one cut-site on the targeted transcript. Higher specificity values are detrimental to the quality of the hhRz candidate, and as such this measure of quality should be minimized (*pseudo-code* in supplementary information briefly describes this).

Hammerhead ribozymes

The hammerhead DNA templates, recommended by RiboSoft for PABPN1, were prepared by 5 cycles of polymerase chain reaction (PCR) using Taq DNA polymerase and 1 μ M of the T7 promoter 5'-TAATACGACTCACTATAGCG-3' with 1 μ M of each oligonucleotide (sequences complementary to the T7 promoter are underlined):

5'-GGCCTGGAGTTTCGTCGCATTTC
GCGACTCATCAGTGAGGTTAACTGGAG
CCTGCGCTATAGTGAGTCGTATTA-3', 5'-
TATCAAAGCTCGAGTTTCGTCGCATTTC
AGCGACTCATCAGAGGGATTAGATGGAG
GAAGACGCTATAGTGAGTCGTATTA-3',
5'-AGATGAATATGAGTTTCGTCGCATTTC
CAGCGACTCATCAGCACCTTTACCAG
GCAATGCGCTATAGTGAGTCGTATTA-3',
5'-GGGTCGCGTTTCGTCGCATTTCAG
CGACTCATCAGTACAGTTAGGGCC
GGCGCTATAGTGAGTCGTATTA-3'.corresponding

respectively to the constructions for Yz144, Yz363, Yz437 and Yz867, with an annealing temperature of 55°C with 200 μ M dNTP, 2 mM MgCl₂, 120 mM Tris-HCl pH 8.8, 10 mM KCl, 10 mM (NH₄)₂SO₄ and 0.1% Triton X-100. The sizes of the DNA templates were confirmed on 1.5% agarose gel. Full annotated ribozyme sequences, as well as the target RNAs, are available in supplementary materials.

PABPN1 gene

The PABPN1 gene used to create by PCR the DNA template for transcription was the D323 clone used in (16). The first PCR product was made with the oligonucleotide 5'-GACTACGGGAACGGCCTGGAGTC-3' starting 129 nucleotides downstream of the start codon and with the oligonucleotide 5'-AAAGGGAACAAAAGCTGGAG-3' 147 nucleotides downstream of the stop codon (as reverse complement); covering 936 nucleotides. To generate a full-length PCR product, 20% betaine was incorporated to the PCR reaction (same reaction conditions as above) for 30 cycles at an annealing temperature of 51°C. A second PCR

reaction was performed on the first PCR product with identical conditions, except for the first oligonucleotide being replaced by the oligonucleotide 5'-TAATACGACTCACTATAGCGACTACGGGAACGGCCTGGAGTC-3' to incorporate the T7 promoter to the DNA template used for transcription. The size of both DNA templates were confirmed on 0.7% agarose gel. The PCR product was purified with an EZ-10 Spin Column PCR Products Purification Kit (Bio Basic Inc.) and precipitated in ethanol with NaOAc.

RNA synthesis of hammerhead ribozymes

Ribozymes were synthesized by *in vitro* run-off transcription in 100 μ l reactions using 20 pmol of double-stranded DNA template, 80 mM HEPES-KOH pH 7.5, 24 mM MgCl₂, 2 mM spermidine, 40 mM DTT, 3.0 mM of each NTP, 0.01 U of yeast pyrophosphatase and 2 μ g of purified T7 RNA polymerase at 37°C for 3 h (27). The reactions were stopped by adding 5 U of DNase (RNase free) and incubating at 37°C for 30 min. The mixtures were extracted twice with both phenol and chloroform, and the nucleic acid precipitated with ethanol. After dissolution in equal volumes of water and formamide dye buffer (95% formamide, 10 mM EDTA, 0.025% bromophenol blue and 0.025% xylene cyanol) the ribozymes were fractionated by 10% denaturing (8 M urea) polyacrylamide gel electrophoresis (PAGE; 19:1 ratio of acrylamide to bisacrylamide) in buffer containing 45 mM Tris-borate pH 7.5 and 1 mM EDTA. The reaction products were visualized by UV shadowing and the bands corresponding to the ribozymes cut out. The transcripts were eluted from the gel slices overnight at room temperature in a solution containing 0.1% SDS, 10 mM EDTA and 0.5 M ammonium acetate. The purified RNA was precipitated by the addition of 0.1 volume of 2 M sodium acetate pH 4.5 and 2.5 volumes of ethanol. After washing in 70% ethanol and drying, the pellets were resuspended in ultra-pure water, and the quantity of RNA determined by spectrophotometry at 260 nm.

RNA synthesis of PABPN1 gene

The full-length PABPN1 gene could not be transcribed efficiently. Various conditions and constructions were tested until a construction with a truncated 5'-end was successfully transcribed at a reasonable amount. The 'standard' transcription protocol described for the ribozymes was however inefficient for this gene. Transcripts were obtained only with the HiScribe™ T7 *in vitro* Transcription Kit (New England Biolabs) where 50 pmol of the column-purified and precipitated DNA template was in presence of 20% betaine and the kit's supplied buffer, High Molecular Weight solution, rNTPs and T7 RNA polymerase enzyme. The transcription reaction proceeded for 4h30 at 42°C and was stopped by the addition of 5 U of DNase (RNase free) and incubating at 37°C for 45 min. Followed by phenol-chloroform extraction and precipitation, the RNA was purified on 5% 8 M urea PAGE. Elution, precipitation and concentration determination were performed as stated above.

RNA labeling

Purified PABPN1 RNA (6 pmol) was 5' dephosphorylated with 5 U antarctic phosphatase in 50 mM Bis-Tris-Propane-HCl pH 6.0, 1 mM MgCl₂ and 0.1 mM ZnCl₂ for 30 min at 37°C followed by heat inactivation 5 min at 65°C. The RNA was radioactively 5'-³²P-labeled using 10 U of T4 polynucleotide kinase (NEB) with 20 μCi of [γ-³²P]-ATP (3000 Ci/mmol; Perkin-Elmer) for 1 h at 37°C in 70 mM Tris-HCl pH 7.6, 10 mM MgCl₂ and 5 mM DTT. The labeled RNA was gel purified on 5% 8 M urea PAGE, eluted, precipitated as stated above except for the RNA bands being identified by radioactivity rather than UV shadowing.

Kinetic measurements of ribozyme cleavage

Trace amount of [5'-³²P]-radioactively labeled PABPN1 RNA and 10 nM of ribozyme (10 nM of each ribozymes when in combination) were separately pre-incubated 10 min at 37°C in the presence of cleavage buffer with ionic strength roughly similar to physiological conditions (50 mM Tris pH 7.5, 100 μM MgCl₂, 100 mM NaCl and 25 mM KCl). The cleavage reaction was initiated by adding an equal volume of target to the ribozyme. Aliquots (1.5 μl) were removed at intervals and the reaction terminated by addition to a 5 μl chilled mixture containing 95% (v/v) formamide, 10 mM EDTA and electrophoresis dyes. Ribozyme and cleavage products were separated by electrophoresis in a dual concentration 8 M urea 19:1 polyacrylamide gel with a 4% top two-third and 20% bottom third. Gels were exposed to storage phosphor screens, and quantified by phosphorimaging. Proportion of cleavage was calculated by dividing the intensity of bands corresponding to ribozyme cleavage products, divided by the sum of all bands in the lane (including the full length RNA). Progress curves were fitted by nonlinear regression analysis to single exponential function.

Construction of plasmids for ribozyme expression in mammalian cells

The same four hhRz against PABPN1 were used, namely Yz144, Yz363, Yz437 and Yz867. Templates containing these hammerhead catalytic sequences were synthesized chemically. The sequences of these four ribozymes are:

- Yz144: 5' CAGGCTCCAGTTTAACCTCACTGATG AGTCGCTGAAATGCGACGAACTCCAGGCC 3'
- Yz363: 5' TCTTCCTCCATCTAATCCCTCTGATG AGTCGCTGAAATGCGACGAACTCGAGCTT TGATA 3'
- Yz437: 5' GCATTGCCTGGTAAAGGTGCTGATGA GTCGCTGAAATGCGACGAACTCATATTCATC T 3'
- Yz867: 5' CCGGCCCTAACTGTACTGATGAGTCG CTGAAATGCGACGAAACGCGACCC 3'

PCR products were digested with KpnI and Csp45I restriction endonucleases and were ligated with pUC-KE-tRNA-CTE plasmids that had been digested with both KpnI and Csp45I obtained kindly from Nawrot B. (19). The DNA sequence of every plasmid was verified using Sanger

sequencing with the PCR primer P7 (5'-CGCCAGGGTT TTCCAGTCACGAC-3').

Cell culture and transfection

Human transformed primary embryonal kidney cell lines HEK293 were cultured in Dulbecco's modification of Eagle's medium (Invitrogen), supplemented with 10% foetal bovine serum, G418 (200 μg/ml) was used as selection antibiotic for ribozyme-transfected HEK293 cells. Cells were grown in monolayer, at 37 °C in cell culture incubator with 5% CO₂ in 6-well plates and transfected at 70–80% confluence. Transfection with 2 μg of the ribozyme containing plasmid was carried out using the Jet prime reagent (Polyplus-transfection Inc., France) according to the manufacturer's protocol. One week later, selected G418 cells were collected for RNA and protein extractions in parallel.

Western blotting

Total proteins were extracted in SUB buffer at different time points (containing 8M urea, 2% mercaptoethanol and 0.5% SDS), and electrophoresed in 12% SDS-PAGE gels. Proteins were blotted into nitrocellulose membrane and probed with specific antibodies against PABPN1 (1:2000) (Abcam, USA). Parallel samples were probed using monoclonal actin antibody (Chemicon International, USA) to confirm the equal loading of lysates between lanes. After incubation with specific secondary HRP-conjugated antibodies, the membranes were revealed using the western blot chemiluminescence reagent plus kit (NEN Life Sciences Products, Boston, MA, USA).

Quantification of the band intensity was performed using the ImageJ software. Values were normalized by β-actin. Quantification of PABPN1 protein level was determined by western blot using the ImageJ software, normalized to actin and presented as fold change relative to the level of PABPN1 in untransfected HEK293T cells.

RESULTS

Fitness evaluation

The quality of a candidate ribozyme is reflected in all five terms computed in sections above. It should be noted that the raw values returned by some of those quality measures, are not intuitively comprehensible. For instance, greater values for ribozyme structure indicate worse structures. In addition, it is often not clear what the lower and upper bounds are, for a given measure. For these reasons, the following equation is used to normalize the values returned by most of the quality measures:

$$M_{normalized} = \frac{(M_{original} - M_{min})}{M_{max} - M_{min}} \quad (3)$$

where M_{min} is the smallest value for that measure among all candidates, and vice versa for M_{max} . $M_{original}$ is the original (raw) value for that measure, while $M_{normalized}$ is that value after normalization. Also, if for a given measure larger values are detrimental, the following equation is applied to the

output, so that 1 represents the best possible value and 0 the worst:

$$M_{adjusted} = 1 - M_{normalized} \quad (4)$$

$M_{normalized}$ is the normalized value for that measure, while $M_{adjusted}$ is that value inverted, and is the final value used by the multi-objective optimization procedure outlined below. Since melting temperature and specificity can be accurately judged by the user, they are not normalized or adjusted.

Depending on the application, the designer will have a greater interest in certain parameters versus others. For example, specificity or lack of off-target effects may be of great value to some *in vivo* applications, but would probably be of no concern to most *in vitro* applications. Hence, it is not possible to identify beforehand, and for all possible applications of a hhRz, what the relative importance of the various terms should be. Even for a particular well-defined application, it is sometimes impossible for a user to spell-out numerically the relative importance of the various quality terms. As such, we adopted a multi-objective optimization approach to the ranking of the ribozymes.

This approach (summarized in supplementary material and described in full in [28]) allows us to present the various candidates to users, as members of various Pareto optimal fronts. In other words, all ribozymes ranked 1 correspond to an optimal choice between the parameters used by RiboSoft such that choosing another candidate ribozyme with one of these parameters 'improved' would actually lead to a less efficient ribozyme due to other parameters being adversely affected. This lesser candidate would be ranked as 2 with other such candidates 'only one value away from the optimum', similarly, ribozymes ranked 3 would be less optimal than 2, and so on until all candidates not rejected by RiboSoft cut-offs are ranked. In order for the user to make a final decision as to the best hhRz for his application: both rank and the various quality measures are included in a table of all hhRz candidates. The table can be sorted, by the user, according to any of the individual measures as well as overall rank. This table constitutes the output of the web service (RiboSoft) implementing the algorithm (Figure 5). Additional details (including *pseudo-code* and examples of results) are provided in supplementary information.

Single ribozyme cleavage and additive effect of combinations

In addition to specifying and devising a comprehensive hhRz design algorithm, then implementing it as web service, we have also used the service for a particular gene silencing application. This involved the PABPN1 gene, whose mutation results in a muscular dystrophy disease known as OPMD. Before proceeding to *in vivo* trials in various model organisms, which is a costly affair, we proceeded to test the validity of ribozymes generated by RiboSoft, *in vitro*, for their efficiency in cleaving the transcript (mRNA) of the PABPN1 gene. The following two sections describe in detail the wet lab experimental methods followed to test the ribozymes and the results of these tests. The results were positive and provide preliminary evidence of the value of the new web service.

The *in vitro* efficiency of cleavage by the designed hhRz of the PABPN1 transcript is shown in Figure 6. The exper-

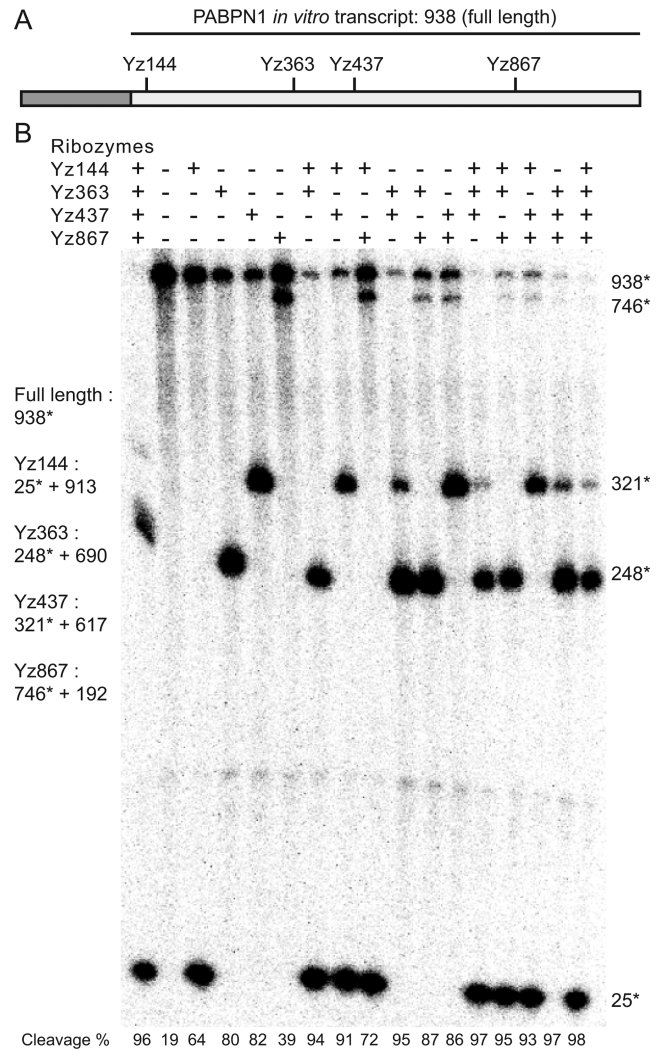


Figure 6. Single and combined ribozyme cleavage. (A) The PABPN1 transcript is illustrated. Gray box corresponds to the region that includes the mutations, but is not included in the transcript for this assay because we were unable to fully transcribe it *in vitro*. White box corresponds to the portion of the transcript used for the assay, with cleavage sites indicated. (B) Cleavage assay. The 938 bases substrate labeled in 5' was incubated 40 min at 37°C, with 100 nM of the indicated ribozymes, in 50 mM Tris, pH 7.5, 100 mM NaCl and 100 μM MgCl₂, except for lane 1, which was in 0 MgCl₂. Reactions were loaded on a 4% (top)/ 20% (bottom) denaturing PAGE. Numbers with asterisks correspond to the labeled fragment. The efficiency of cleavage, represented as a percentage of the original uncleaved transcript, is shown at the bottom of the figure. With the exception of the first lane, the results are arranged, from left to right to reflect an increasing number of participating ribozymes. Naturally, whatever remains of the original transcript is represented by the bands at the top of the figure, while the various fragments resulting from cleavage are spread all over the gel below. The expected lengths of the transcript fragments are listed to the left of the gel, while the actual lengths of the detected fragments are listed on the right at the proper location. This gel is a representative experiment of three assays.

iment measured the efficiency of cleavage of all single, double, triple and quadruple combinations of four ribozymes. We tested the efficiency of the ribozymes in conditions similar to those often used to assay ribozymes intended for use in mammalian cells (37°C, 100 mM Na⁺ and 100 μM Mg²⁺). The only exception being the very first lane in Fig-

ure 6, where a combination of all four ribozymes was used at zero magnesium ion concentration.

The results indicate several important facts: (i) all single ribozymes cleave the PABPN1 transcript, with an efficiency of cleavage ranging between 39.3% and 81.5%. (ii) A greater number of ribozymes invariably cleaves more efficiently than a smaller number of ribozymes; the average cleavage percentage for 1, 2, 3 and 4 ribozymes is: 66.15%, 87.65%, 95.575% and 97.7%. (iii) Oddly enough, a combination of all four ribozymes in the absence of magnesium ions also cleaved very efficiently (96.6%) even if magnesium is thought to be essential for efficient cleavage by hhRz. (iv) The greatest increase in the efficiency of cleavage resulted from using two instead of one ribozyme. This result confirms that the design of the ribozymes, amongst many generated by RiboSoft, were correct, in that they led to functional ribozymes.

To ensure the universal applicability of RiboSoft, we assayed 15 additional hhRz designed by our software against two very different targets, a RFP mRNA and a portion of a 16S ribosomal RNA, the latter is notably structured and likely much less accessible than the average mRNA. All the ribozymes that were ranked 1 against these targets could cleave similarly to the results obtained for PABPN1-hhRzs, except for one inactive ribozyme, as well as most ribozymes with ranks as low as 7 (Supplementary Figures S1, S2, Table S1 and data not shown). However, for two ribozymes that had poor rankings, and interestingly had an *accessibility* 2 measure of zero, no significant cleavage was observed (Supplementary Table S1). Two hhRz based on the same cut-sites, but designed manually, did not cleave either, providing additional evidence that 'good ranking hhRz' designed by RiboSoft warrant active ribozymes. As with PABPN1, the combination of ribozymes led to nearly complete degradation of the substrate RNA (Supplementary Figure S1 and S2).

Single and combined ribozymes kinetic reaction results

The progress of the cleavage reaction over time, for all combinations of the four ribozymes, is presented in Figure 7. The purpose of these data is to see and mathematically model the rate of the cleavage reaction. This allows us to identify the time it takes the reaction to reach its steady-state. The plots provide a complete picture of the dynamic relationship between the number of ribozymes (participating in a reaction) and the rate of cleavage over time. The efficiency of cleavage at each time step can be seen in Figure 7 as well as the constants of the cleavage rate (k_{obs}).

It can be seen from the plots that both the rate of the cleavage reaction and the steady-state value of cleavage efficiency have increased due to the presence of an additional ribozyme. The two-ribozyme combination was able to cleave about 43% of the transcript in 10 min, while the single ribozyme cleaved about 15% in the same time period. Also, the two ribozyme combination had a steady-state cleavage efficiency of ~90% a clear improvement over the steady-state of the single ribozyme of ~30%. Similarly, the combination of three ribozymes increased further the cleavage efficiency of the RNA substrate. Indeed, only 5 min was required to cleave 44% of the target RNA, half as much time

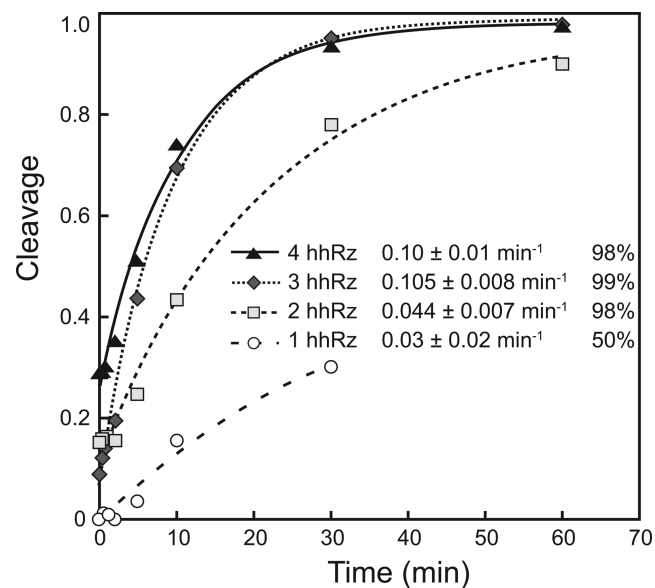


Figure 7. Single and combined ribozyme kinetic. The rate of cleavage of one, two, three and four combined ribozymes is illustrated. Cleavage was quantified and plotted to determine rate constants, which are indicated within the graph together with the final percentage of cleavage: open circle, RNA substrate cleaved by a single ribozyme (Yz144); by two ribozymes (Yz144 and Yz363); three ribozymes (Yz144, Yz363 and Yz437); and with all ribozymes. Nine time points are taken at 0 s, 20 s, 40 s, 1 min, 2 min, 5 min, 10 min, 30 min and 1 h. These results are representative of two assays.

as with two ribozymes, which achieved 43% cleavage in 10 min. In spite of this, three ribozymes presented only a minor improvement over the already highly efficient combination of two ribozymes, 90% cleavage. No measurable improvement was obtained with four ribozymes compared to three.

It is recommended that one uses two proven hhRz in gene silencing applications; three or more hhRz is not advisable, as that increases the cost of construction and delivery without considerably enhancing the steady-state value of cleavage efficiency. Also, expressing many ribozymes increases the likelihood of unwanted side-effects such as off-target hits and toxicity for *in vivo* applications.

Specific hhRzs are able to inhibit PABPN1 expression level

To test the hhRzs silencing effect, we used the human cell line HEK293T and transfected the cells with all different hhRzs individually. Stable clones of hhRzs were maintained in G418 selective media all the time during manipulations. We then performed western blot on cell extracts to test the effectiveness of hhRzs in knocking down the endogenous PABPN1 level in this cell line. Western blot results revealed hhRzs sequences: Yz144, and Yz437 as optimal candidates, as they showed high efficiency in knock-down PABPN1 (Figure 8A). We then transfected combinations of two hhRzs simultaneously in HEK293T cells and performed western blot to test if they are more effective in knocking down PABPN1 protein level. A combination of two hhRzs showed more reduction in PABPN1 protein level (Figure 8B) than individual hhRzs (Figure 8A), reducing expression down to 5% for the combinations of Yz363 with either Yz437 or Yz144. In addition to provid-

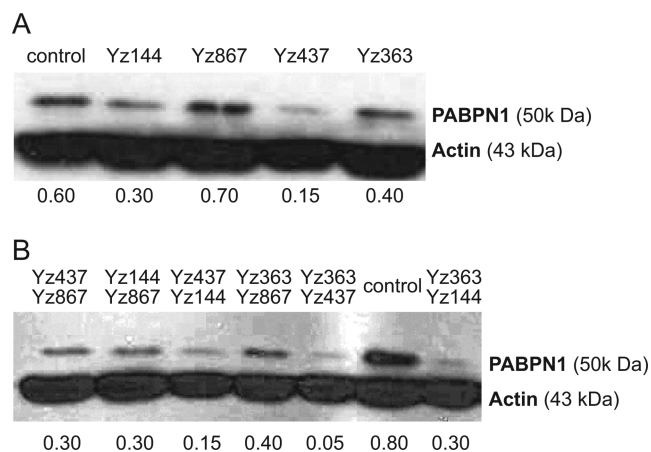


Figure 8. Ribozymes are effective in knocking down PABPN1. (A) Different individual ribozymes targeting PABPN1 transcript are able to knock down the level of PABPN1 protein. The endogenous expression of PABPN1 was analyzed by western blot. HEK 293T cells were transfected with the corresponding individual ribozymes, and selected with G418 for 5 days before protein extraction. (B) Combination of two ribozymes targeting PABPN1 transcript are more efficient in knocking down the level of PABPN1 protein. The endogenous expression of PABPN1 was analyzed by western blot. HEK 293 cells were transfected with the combined two indicated hhRzs. For (A) and (B), the ‘control’ lanes correspond to plasmid without ribozymes. Actin antibody was used to confirm equal loading. The values below the gel indicate PABPN1 protein signal intensities (quantified using ImageJ) after normalization to actin signal intensities and to normal PABPN1 protein levels (in cells that have not been transfected).

ing a strong promoter, the pUC-KE-tRNA-CTE plasmid includes a tRNA to help scaffold and export RNA to the cytoplasm, as well as the Constitutive Transport Element (CTE) which tethers an RNA helicase that helps unfold RNA and facilitates hhRz binding to its target mRNA at the same time. While concerns could arise from the use of ‘non-ribozyme’ sequence in the 5’ portion of the constructs, this extra sequence is expected to fold in a modular fashion independent from the hhRz. The excellent activity of the ribozymes against their target *in vivo* confirms that the ribozymes are still active.

DISCUSSION

This paper presents two main contributions to the world of ribozyme design and ribozyme application. The first part of the paper describes the algorithm and implementation of RiboSoft, a completely automated program (offered as a web service) for hhRz design. It can be used to design hhRz to target any RNA sequence, including the transcript of any gene. The program has clear and valuable advantages over existing hammerhead design services, related to usability, ribozyme quality assessment and multi-objective optimization. It unifies in a single tool the functionalities of targeting ribozyme design software that have been done in the last decade: it finds target sites, computes target accessibility, ribozyme structure validity and potential off-target effects. More importantly for the future community of users, it provides a user friendly interface with the flexibility required for any application: from *in vitro* assays, to gene targeting in *Escherichia coli*, in humans or in any desired species and computes all the complexity of the analysis to simply pro-

vide ribozymes that will likely be active over 90% of the time as demonstrated by our *in vitro* assays.

A number of software exist to assist in ribozyme design. Up to now, the best tool in that regard was probably Aladdin (SeArch computing tool for hAmmerhead ribozyme DesIgN) (<http://aladdin.ifc.cnr.it/aladdin>). This tool, which evolved from an earlier computational method (29,30), is handy and scientifically sound. However, RiboSoft has several advantages: it uses the latest research on target-site accessibility; and it has off-target specificity analysis. Off-target accessibility could be evaluated with RiboSubstrates (31), but the design of ribozymes would then require the use of two different tools, which makes automated design more complicated. RiboSoft also has the advantage of allowing specification of the nature of the environment, the type and size of the ribozyme and the nature of the application. Similarly, the evaluation of possible hhRzs is easier with RiboSoft than Aladdin, because the former provides multi-objective ranking and allows sorting ribozymes according to each parameter. Our experience with these parameters seem to point toward a stronger impact of ‘accessibility 2’ on the ability of a chosen ribozyme to cleave its target.

The other contribution is the design (by RiboSoft) of four hhRz targeting the transcript of the PABPN1 gene, and testing these ribozymes, *in vitro* and *in vivo*, for their cleavage efficiency. A mutated PABPN1 is the root cause of OPMD (oculopharyngeal muscular dystrophy) in humans, a late onset muscle disease associated with progressive ptosis of the eyelids, and dysphagia. The experiments validated the designs as they demonstrated cleavage by each of the four ribozymes. In spite of using single turnover conditions for *in vitro* assays, the results *in vivo* show that the ribozymes designed by RiboSoft lead to an efficient knockdown of PABPN1. This is likely due to the use of a strong promoter for ribozyme expression and higher stability of the structured ribozyme RNAs, as compared to the mRNA, which generally have low expression as compared to most non-coding RNAs. Overall, according to our results, we recommend using two hhRzs for researchers interested in targeting a gene for silencing.

Any upcoming gene targeting projects can benefit from RiboSoft which will easily provide multiple ribozyme designs against any transcript. This fact will allow anyone to select multiple ribozymes in order to attain the best effect possible. Indeed, as we have demonstrated, a dramatic additive effect is obtained by combining two ribozymes, and in the few cases where it would be insufficient, three ribozymes are likely to be enough, making RiboSoft all the more valuable.

There are two ways in which this work can be developed. One direction is that of expanding the computational design tool to include not just hhRz, as it does now, but other ribozymes, such as hairpin and HDV ribozymes. These ribozymes share the basic architecture of a hhRz, which is made of hybridizing arms responsible for the selective capability of the ribozyme and a largely conserved core. As such, many ribozymes can, in principle, be optimized using the same overall multi-objective approach employed by RiboSoft, which essentially takes into account the (i) structure of the ribozyme and exact sequence of its recognition arms;

(ii) structure of the region of the targeted cut-site and its exact sequence; (iii) off-target effects, for *in vivo* applications. We cannot see any conceptual or technical impediments to the customization of the general approach to the various ribozymes whose functionality is a function of these three (i–iii) sets of considerations.

In addition, riboswitch-ribozymes have been developed and their use could be greatly extended with proper design tools such as RiboSoft. A riboswitch is an RNA that turns genes on or off, when it binds to a ligand input or a combination of ligand inputs (32). Riboswitches can be combined with ribozymes so that the binding of their ligands trigger RNA cleavage (33,34). Engineered examples of riboswitch-ribozymes implementing Boolean functions include some that respond to small-molecule ligands (14). Allosteric ribozymes, or any other ribozyme, could be added to RiboSoft in the same way as hhRzs have been implemented. In the long run, users could choose from a list of ribozymes with different features to target genes. In principle, any allosteric ribozymes found via SELEX (35) could be implemented for future use, as long as their structures are known. Although it was not packaged as a user friendly software and was designed for self-cleaving ribozymes rather than for ribozymes capable of cleaving any target, recently developed algorithms allowed construction of switchable ribozymes based on the hhRz model and known aptamers (36). This is similar to what has been developed by the group of Ronald Breaker in 2005, where a purely computational approach was developed that found complementary RNA ligands that altered the conformation and hence activity of a hhRz (37). In fact, they tested *in vitro* computationally designed ribozymes that implemented four Boolean logic functions: AND, OR, NOT and YES. It is worth noting that only a precisely delineated part of the ribozyme need to be altered to give the whole ribozyme this unique allosteric capability; the rest of the ribozyme stays fixed. Both of these allosteric ribozymes could be designed by a RiboSoft 2.0.

Another direction for development involves the modification of the design of the resulting hhRz to test their gene silencing functionality, as well as unwanted effects, in living organisms. Selected ribozymes against PABPN1 appear to cleave their target both *in vitro* and *in vivo*, which is promising for future work aiming to replace the mutant transcripts with healthy mRNAs in human cell, thus advancing toward a therapeutic approach to OPMD. Indeed, even when it is theoretically possible to target them, the exact regions harboring mutations are not necessarily ideal cleavage sites. However, addition of a healthy copy with sequence modifications rendering it immune to cleavage, as previously suggested by others (38), allows the use of ribozymes for any type of genes. In fact, even if many deleterious mutations are known for numerous diseases, in such a model for gene therapy it would be possible to target defective genes without even knowing the exact mutations responsible of the disease.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Natural Sciences and Engineering Council of Canada (NSERC) [418240 to J.P.]. J.P. is a junior 1 FRQS research scholar. Funding for open access charge: Natural Sciences and Engineering Council of Canada (NSERC) [418240].
Conflict of interest statement. None declared.

REFERENCES

- Kim, D.H. and Rossi, J.J. (2007) Strategies for silencing human disease using RNA interference. *Nat. Rev. Genet.*, **8**, 173–184.
- Scherer, L.J. and Rossi, J.J. (2003) Approaches for the sequence-specific knockdown of mRNA. *Nat. Biotechnol.*, **21**, 1457–1465.
- Serganov, A. and Patel, D.J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.*, **8**, 776–790.
- Amarzguioui, M. and Prydz, H. (1998) Hammerhead ribozyme design and application. *Cell. Mol. Life Sci.*, **54**, 1175–1202.
- Khvorova, A., Lescoute, A., Westhof, E. and Jayasena, S.D. (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat. Struct. Biol.*, **10**, 708–712.
- De la Pena, M., Gago, S. and Flores, R. (2003) Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. *EMBO J.*, **22**, 5561–5570.
- Ohmichi, T. and Kool, E.T. (2000) The virtues of self-binding: high sequence specificity for RNA cleavage by self-processed hammerhead ribozymes. *Nucleic Acids Res.*, **28**, 776–783.
- Lévesque, D., Brière, F.P. and Perreault, J.P. (2007) A modern mode of activation for nucleic acid enzymes. *PLoS One*, **2**, e673
- Kuwabara, T., Warashina, M., Tanabe, T., Tani, K., Asano, S. and Taira, K. (1998) A novel allosterically trans-activated ribozyme, the maxizyme, with exceptional specificity *in vitro* and *in vivo*. *Mol. Cell*, **2**, 617–627.
- Breaker, R.R. (2002) Engineered allosteric ribozymes as biosensor components. *Curr. Opin. Biotechnol.*, **13**, 31–39.
- Vaish, N.K., Dong, F., Andrews, L., Schweppe, R.E., Ahn, N.G., Blatt, L. and Seiwert, S.D. (2002) Monitoring post-translational modification of proteins with allosteric ribozymes. *Nat. Biotechnol.*, **20**, 810–815.
- Rusk, N. (2012) Microbiology: Prokaryotic RNAi. *Nat. Methods*, **9**, 220–221.
- Lewin, S. and Hauswirth, W.W. (2001) Ribozyme gene therapy: applications for molecular medicine. *Trends Mol. Med.*, **7**, 221–228.
- Purnick, E.M.P. and Weiss, R. (2009) The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.*, **10**, 410–422.
- Brais, B. (2009) Oculopharyngeal muscular dystrophy: a polyalanine myopathy. *Curr. Neurol. Neurosci. Rep.*, **9**, 76–82.
- Brais, B., Bouchard, J.P., Xie, Y.G., Rochefort, D.L., Chrétien, N., Tomé, F.M., Lafrenière, R.G., Rommens, J.M., Uyama, E., Nohira, O. *et al.* (1998) Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nat. Genet.*, **18**, 164–167.
- Rodríguez-Lebrón, E., do Carmo Costa, M., Luna-Cancelon, K., Peron, T.M., Fischer, S., Boudreau, R.L., Davidson, B.L. and Paulson, H.L. (2013) Silencing mutant ATXN3 expression resolves molecular phenotypes in SCA3 transgenic mice. *Mol. Ther.*, **21**, 1909–1918.
- Hu, J., Matsui, M., Gagnon, K.T., Schwartz, J.C., Gabillet, S., Arar, K., Wu, J., Bezprozvanny, I. and Corey, D.R. (2009) Allele-specific silencing of mutant huntingtin and ataxin-3 genes by targeting expanded CAG repeats in mRNAs. *Nat. Biotechnol.*, **27**, 478–484.
- Nawrot, B., Antoszczyk, S., Maszewska, M., Kuwabara, T., Warashina, M., Taira, K. and Stec, W.J. (2003) Efficient inhibition of beta-secretase gene expression in HEK293 cells by tRNAVal-driven and CTE-helicase associated hammerhead ribozymes. *Eur. J. Biochem.*, **270**, 3962–3970.
- Zoumadakis, M. and Tabler, M. (1995) Comparative analysis of cleavage rates after systematic permutation of the NUX consensus target motif for hammerhead ribozymes. *Nucleic Acids Res.*, **23**, 1192–1196.
- Shimayama, T., Nishikawa, S. and Taira, K. (1995) Generality of the NUX rule: kinetic analysis of the results of systematic mutations in

- the trinucleotide at the cleavage site of hammerhead ribozymes. *Biochemistry*, **34**, 3649–3654.
22. Panjkovich, A. and Melo, F. (2005) Comparison of different melting temperature calculation methods for short DNA sequences. *Bioinformatics*, **21**, 711–722.
 23. SantaLucia, J. and Hicks, D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.
 24. Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 25. Ding, Y., Chan, C.Y. and Lawrence, C.E. (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.*, **32**, W135–W141.
 26. Shao, Y., Wu, S., Chan, C.Y., Klapper, J.R., Schneider, E. and Ding, Y. (2007) A structural analysis of in vitro catalytic activities of hammerhead ribozymes. *BMC Bioinformatics*, **8**, 469.
 27. Milligan, J.F., Groebe, D.R., Witherall, G.W. and Uhlenbeck, O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, **15**, 8783–8798.
 28. Deb, K. (2001) *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, Inc., New York, NY, USA.
 29. Mercatanti, A., Lande, C. and Citti, L. (2012) A computational approach to predict suitable target sites for trans-acting minimal hammerhead ribozymes. *Methods Mol. Biol.*, **848**, 337–356.
 30. Markham, N.R. and Zuker, M. (2008) UNAFold: software for nucleic acid folding and hybridization. In: Keith, J.M. (ed.), *Bioinformatics, Volume II: Structure, Function and Applications*, Vol. **453**, in *Methods in Molecular Biology*. Humana Press, Totowa, NJ, 3–31.
 31. Lucier, J.F., Bergeron, L.J., Brière, F.P., Ouellette, R., Elela, S.A. and Perreault, J.P. (2006) RiboSubstrates: a web application addressing the cleavage specificities of ribozymes in designated genomes. *BMC Bioinformatics*, **7**, 480.
 32. Edwards, A.L. and Batey, R.T. (2010) Riboswitches: a common RNA regulatory element. *Nat. Educ.*, **3**, 9.
 33. Lee, E.R., Baker, J.L., Weinberg, Z., Sudarsan, N. and Breaker, R.R. (2010) An Allosteric Self-Splicing Ribozyme Triggered by a Bacterial Second Messenger. *Science*, **329**, 845–848.
 34. Breaker, R.R. (2002) Engineered allosteric ribozymes as biosensor components. *Curr. Opin. Biotechnol.*, **13**, 31–39.
 35. Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
 36. Shen, S., Rodrigo, G., Prakash, S., Majer, E., Landrain, T.E., Kirov, B., Daròs, J.A. and Jaramillo, A. (2015) Dynamic signal processing by ribozyme-mediated RNA circuits to control gene expression. *Nucleic Acids Res.*, **43**, 5158–5170.
 37. Penchovsky, R. and Breaker, R.R. (2005) Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes. *Nat. Biotechnol.*, **23**, 1424–1433.
 38. Millington-Ward, S., O'Neill, B., Tuohy, G., Al-Jandal, N., Kiang, A.S., Kenna, P.F., Palfi, A., Hayden, P., Mansergh, F., Kennan, A. *et al.* (1997) Strategies in vitro for gene therapies directed to dominant mutations. *Hum. Mol. Genet.*, **6**, 1415–1426.
 39. Saksmerprom, V., Roychowdhury-Saha, M., Jayasena, S., Khvorova, A. and Burke, D.H. (2004) Artificial tertiary motifs stabilize trans-cleaving hammerhead ribozymes under conditions of submillimolar divalent ions and high temperatures. *RNA*, **10**, 1916–1924.

Supplementary Information

for the manuscript

Automated Design and In Vitro Validation of Hammerhead Ribozymes Targeting the PABPN1 Gene Transcript

Nawwaf Kharma¹, Luc Varin², Aida Abu-Baker³, Jonathan Ouellet^{4,5}, Sabrine, Najeh⁴,
Mohammad-Reza Ehdaevand², Gabriel Belmonte¹, Anas Ambri¹, Guy Rouleau³ and Jonathan
Perreault⁴

¹ Electrical & Computer Eng. Dept., Concordia University, 1455 boul. de Maisonneuve O.,
Montreal, QC, H3G 1M8, Canada

² Biology Department, Concordia University, 7141 rue Sherbrooke O., Montreal, QC, H4B 1R6,
Canada

³ Montreal Neurological Hospital and Institute, 3801 University Street, Montreal, QC, H3A 2B4,
Canada.

⁴ INRS - Institut Armand-Frappier, 531 boulevard des Prairies, Laval, QC, H7V 1B7, Canada,

To whom correspondence should be addressed:

Phone: 514-8482424 x3117, fax: 514-8482802, e-mail: kharma@ece.concordia.ca or phone:

450-687-5010 x4411, fax: 450-686-5301, e-mail: jonathan.perreault@iaf.inrs.ca

⁵ Present address:

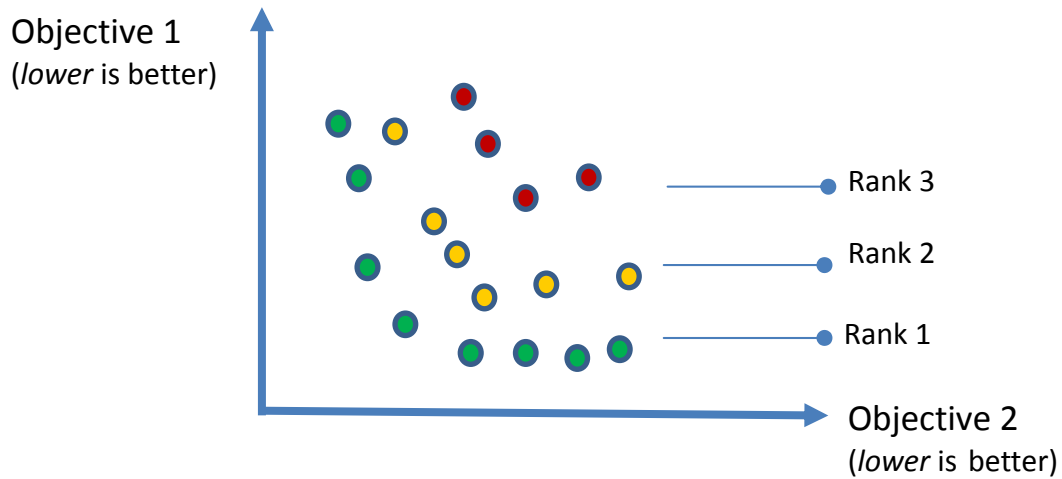
Jonathan Ouellet, Department of Chemistry and Physics, Monmouth University, 400 Cedar
Avenue, West Long Branch, NJ, 07764

Table of content

Multi-objective optimization and ranking	3
RiboSoft pseudo-code	4
Ribozyme Structure	4
Target Accessibility	4
Specificity Assessment	5
Fitness Evaluation	6
Supplementary methods	7
RFP and 16S Transcripts	7
Additional Ribozyme Transcripts	7
Kinetic measurements of ribozyme cleavage for RFP and 16S	8
Table S1	9
Figure legends	10
Figure S1	11
Figure S2	12
Sequences	13

Multi-objective optimization

The schematic below illustrates the principles of multi-objective optimization with “pareto-optimal” fronts.



- Ranks of potential solutions (hhRz designs) are based on the concept of domination & non-domination
 - Domination: solution X dominates another solution Y, if (1) Solution X is no worse than Y in all objectives AND (2) solution X is strictly better than Y in at least one objective.
 - Non-dominated solutions: If two solutions are compared, then the solutions are said to be non-dominated with respect to each other if and only if neither solution dominates the other.
- The set of those solutions that are non-dominated by any solution make-up the 'pareto-optimal' front, and are all assigned the same rank: 1
- If the rank 1 solutions are removed, the next front is comprised of solutions of rank 2
- This layered removal of solution 'fronts' is continued until either we don't need any more solutions or there are simply no solutions left
- The same principles apply independently of the number of objectives (dimensions) assessed

RiboSoft pseudo-code

The following pseudo-code provides brief explanations of the algorithms that evaluate each parameter of the ribozyme candidates.

Ribozyme Structure

```
sum = 0;

for each candidateStructure

    ContinuousPairs = FindContinuousBasePairs();

    for each ContinuousPair in ContinuousPairs

        sum += MeltingTemperature(ContinuousPair);

    end for each;

end for each;
```

A candidateStructure is a possible secondary structure of an hhRz. FindContinuousBasePairs will find all double-stranded sequences in the structure, such that there is at most one mismatch in each sequence. This list will be stored in ContinuousPairs. For each of these double stranded segments (or ContinuousPair), MeltingTemperature will be computed and added to sum. This sum represents the melting temperature of the whole candidateStructure.

Target Accessibility

Cut-site inaccessibility

```
sum = 0;

for each fold in targetFolds

    ContinuousPairs = FindContinuousBasePairs();

    for each ContinuousPair in ContinuousPairs

        sum += MeltingTemperature(ContinuousPair);

    end for each;

end for each;
```


The algorithm is identical to the one used for Ribozyme Structure explained earlier. Only the regions have changed. The regions are now the cut-site regions in targetFolds.

Disruption energy

```
for each cut-site
    constrainedFolds =
        FoldTargetWithOpenCutsite(target, cut-site);
    cut-site.disruptionE =
        targetFolds.LFE-constrainedFolds.LFE;
end for each;
```

The cut-site is the region of the target substrate that is complementary to the arms of the hhRz. The constrainedFolds represent secondary structures of the target such that the target is forced open at the cut-site. These secondary structures are obtained through the FoldTargetWithOpenCutsites operation. The disruptionE energy is the difference between two energies: targetFolds.LFE and constrainedFolds.LFE. targetFolds.LFE (where "LFE" is "Local Folding Energy"). The former is the free energy of the unconstrained folding of the target substrate, while the latter is the free energy of the target substrate, with the cut-site region forced open.

Specificity Assessment

```
Specificity = 0;
for each cut-site
    results = Queryblast(cut-site);
    for each result in results
        Weight(result);
        if result is XM or XR
            result.Weight = 0;
        end if;
```

```

        Specificity += result.Weight;
    end for each;
end for each;

```

Specificity is initialized to zero. For each cut-site, results will hold information obtained from querying BLAST on that cut-site. A cut-site is defined here as the region of the target that is complementary to the arms of the hhRz. Each BLAST result will be dealt with individually, even if two or more results are attributed to a single transcript. Weight will weigh each result based on the perfection of its match to the hhRz arms. These weighted results will be summed into Specificity. An XM or XR Blast hit will not be used in the specificity calculation.

Fitness Evaluation

```

ParetoRank (candidates, rank)
{
    for each pivot in candidates
        for each candidate in candidates
            if (Dominates (candidate, pivot))
                exit inner-loop,
                continue outer-loop -> try new pivot;
            end for each
            pivot.rank = rank;
        end for each
        remove all x in candidates with pivot.rank = rank;
        ParetoRank(candidates, rank - 1);
    }

```

The first time this operation is called, it is called using a list of all the candidates and a rank equal to 1. A pivot is a candidate that will be compared to every other candidate. Candidate A dominates candidate B if it is at least as good as B in all measures of quality

and better than B in at least one quality. If pivot is not dominated by any other as determined by *Dominates*, then its rank is set to rank. When all candidates have been evaluated, all those that have been assigned the current rank value are removed from the candidates list in this function. The function will be re-invoked with the *ParetoRank* call, with a decreased set of candidates and a decremented rank value.

Supplementary methods

RFP and 16S Transcripts

The RFP (Red Fluorescent Protein) gene was amplified by PCR using as a template the plasmid BBa_E1010, provided by Biobricks. The PCR reaction was made in one step using as the forward oligonucleotide 5'-TAATACGACTCACTATAGCGTCTAGA-3' containing the T7 promoter and the reverse oligonucleotide 5'-TTATTAAGCACCGGTGGAG-3' resulting in a PCR product of 726 nucleotides that was used as the DNA template for transcription. The reaction was made for 35 cycles at an annealing temperature of 51.3°C with the same conditions as other PCR reactions.

Another target RNA was prepared, a region of the 16S rDNA of *Microcystis aeruginosa* NIES 298 was amplified by PCR from genomic DNA using the forward primer 5'-TAATACGACTCACTATAGCGAGAGTTTGATCCTGGCTCAG-3', containing the T7 so that the template can be used for the transcription, and the reverse primer 5'-GCTTCGGCACGGCTCGGGTCGATA-3' to generate a DNA template of 822 nucleotides. The same reaction conditions were used at an annealing temperature of 48.9°C for 35 cycles. The size of each DNA template was confirmed using a 2% agarose gel. The DNA was precipitated in ethanol with NaOAc.

In both cases, the RNA was prepared as described for hammerhead ribozymes, except that it was purified on 6% PAGE.

Additional Ribozyme Transcripts

The hhRz designed against RFP and the 16S fragment were all prepared in the same way as the ribozymes against PABPN1 with the following oligonucleotides:

GUC2 (4) :

TGGGTCGTTTTCGTCGCATTTTCAGCGACTCATCAGCGTACTTAGAAGGTACCGCTATAGTGAGTCGTATTA

GUC4 (1) :

CATCCTGTTTTCGTCGCATTTTCAGCGACTCATCAGCCCGCTTAGTTCCAGCGCTATAGTGAGTCGTATTA

GUC4 (6) :

ACATCCTGTTTTCGTCGCATTTTCAGCGACTCATCAGCCCGCTTAAGTTCCGCTATAGTGAGTCGTATTA

GUC7 (1) :

CGTCCGACGGTTTTCGTCGCATTTTCAGCGACTCATCAGCGGTTTTAATGCAGAACGCTATAGTGAGTCGTATTA

GUC7 (4) :

CCGACGGTTTTCGTCGCATTTTCAGCGACTCATCAGCGGTTTTAATGCAGAAACGCTATAGTGAGTCGTATTA

GUC9 (1) :

AGACGGTGGTTTTCGTCGCATTTTCAGCGACTCATCAGACTACTTAGACGCTCGCTATAGTGAGTCGTATTA

GUC9 (7) :

ACGGTGGTTTTCGTCGCATTTTCAGCGACTCATCAGACTACTTAGACGCCGCTATAGTGAGTCGTATTA

GUC11 (1) :

GAAGGTCGTTTTCGTCGCATTTTCAGCGACTCATCAGACTCCTTAACCGGTGCTCGCTATAGTGAGTCGTATTA

GUC5 (m) :

GAAACTGTTTTCGTCGCATTTTCAGCGACTCATCAGCTTCTTACGGAATGCCGCTATAGTGAGTCGTATTA

GUC5 (3) :

TGAAACTGTTTTCGTCGCATTTTCAGCGACTCATCAGCTTCTTACGGAAGCGCTATAGTGAGTCGTATTA

GUC3 (1) :

AGGAAGGTTTTCGTCGCATTTTCAGCGACTCATCAGTTTGGTTAATTGTCGCTATAGTGAGTCGTATTA

GUC7 (5) :

AAGTCAGCTTGTTCGTCGCATTTTCAGCGACTCATCAGAAATCTTAAGGTTAACGCTATAGTGAGTCGTATTA

GUC7 (m) :

AAGTCTGCTGTTTTCGTCGCATTTTCAGCGACTCATCAGAAATCTTAAGGTTGCTCGCTATAGTGAGTCGTATTA

GUC7 (1) (A) :

CGTCCGACGGTTTTCGTCGCATTTTCAGCGACTCATCAGCGGTTTTAATGCAGACGCTATAGTGAGTCGTATTA

GUC2 (1) (A) :

TGAAGGTCGTTTTCGTCGCATTTTCAGCGACTCATCAGCGTACTTAGAAGGTCGCTATAGTGAGTCGTATTA

GUC9 (1) (A) :

AAGACGGTGGTTTTCGTCGCATTTTCAGCGACTCATCAGACTACTTAGACGCGCTATAGTGAGTCGTATTA

GUC11 (5) :

GAAGGTCGTTTTCGTCGCATTTTCAGCGACTCATCAGACTCCTTAACCGGCGCTATAGTGAGTCGTATTA

Kinetic measurements of ribozyme cleavage for RFP and 16S

To monitor the ribozyme cleavage kinetics, 5nM of each ribozyme was added to the cleavage buffer (as described the main text) in the presence of labelled RFP or 16S and in absence of $MgCl_2$. The first step was the incubation at 85°C for 1 minute to allow RNA folding, followed by the immediate transfer on ice for a few minutes. In the second step, 1µl of $MgCl_2$ was added to start the cleavage reaction and this moment was considered as time 0, after which aliquots of 2 µl each were taken at different times. All the rest is the same as described in the main text.

Table S1

Ribozyme (rank)	Tm left	Tm right	Accessibility 1	Accessibility 2	Shape quality	Cleavage	Target
2(4)	45.39	35.39	0.38	0.546	0.315	+	RFP
4(1)	49.39	29.39	0.519	0.519	0.708	++	RFP
4(6)	39.39	31.39	0.624	0.519	0.241	++++	RFP
5(3)	40.97	36.97	0.62	0	0.651	-	RFP
7(1)	43.39	43.39	0.268	0.227	0.968	++	RFP
7(4)	43.29	33.39	0.281	0.227	0.912	++	RFP
9(1)	39.39	37.39	0.684	0.226	0.525	++++	RFP
9(7)	37.39	31.39	0.684	0.226	0.466	+++	RFP
11(1)	52.07	33.39	0.499	0.309	0.536	-/+	RFP
11(5)	39.39	35.39	0.648	0.309	0.284	-/+	RFP
A 2(1)	39.39	35.39	0.763	0.5	0.976	-	RFP
A 7(1)	41.39	43.39	0.635	0.236	0.953	++	RFP
A 9(4)	37.39	39.39	0.605	0.212	0.551	+	RFP
3(1)	31.39	29.39	0.645	0.652	1	++++	16S
7(4)	41.39	33.39	0.338	0	0.653	-	16S
7(m)						-	16S
5(m)						-	RFP

The table shows the different parameters used by RiboSoft for ranking with the cleavage efficiency for comparison.

Ribozyme(rank) : the numbers like “2(4)” refer to:

The cut-site position (which is GUC by default), so “2” means the third GUC that exists in the target sequence because the counting begins from 0 (the first GUC is n° 0). The 4 which is in parenthesis is the rank of the ribozyme as determined by RiboSoft (1 being the best).

Tm: melting temperature for one ribozyme arm with the target. Because ribozymes require complementarity to form two stems with the target, both the left and the right Tm are shown.

Accessibility: to be able to hybridize with the template, the ribozyme should find an accessible sequence so the value of the accessibility that goes from 0 to 1 depends on the position of the GUC and the sequence on both sides of it that could be in a stem (so the value is closer to 0) or in a bulge (so the value is closer to 1) so the closer the value is to 1 the better the cleavage is supposed to be.

Shape quality : Each RNA folds in a specific secondary structure, but the hammerhead ribozyme has a consensus structure so the values of the shape quality are the prediction by RiboSoft of the ability for each putative ribozyme sequence to fold in the active ribozyme structure.

Cleavage: those values presented as – and + are the results of testing the ability of the ribozymes on cutting their template. These values reflect both maximum cleavage and speed of cleavage (from time course experiments, data not shown).

-, no detectable cleavage; -/+, ambiguous; +, low; ++, medium; +++, high; +++++, very high.

Note: the “A” written in front of some ribozyme names means “Anterior” and refers to a first set of ribozymes that were tested on RFP.

Figure S1. Some of the ribozymes designed by RiboSoft against RFP RNA were chosen to evaluate the effect of different parameters on the ribozyme activity. Thus, for many sites two ribozymes of different ranks were tested. All ribozymes were designed by RiboSoft (designated by their cut-site number with their ranking in parenthesis) except Rz 5m, which was designed manually (a ribozyme with a catalytic core identical to that used by RiboSoft, but with binding arms chosen manually to include mismatches with the target). Different combinations of 2 or 3 ribozymes were made and the cleavage efficiency was calculated for each reaction.

The 6% polyacrylamide gel shows labelled RFP RNA in the last lane, the higher band corresponds to the full length RFP (726* nt) with no ribozymes. In the following ten lanes, from 2(4) to 5(m), RFP RNA was incubated in presence of one ribozyme in every reaction where, for the ribozymes 2, 4, 7 and 9, RFP was cleaved and we can observe different bands corresponding to the ribozyme cut-sites respectively: 153* nt (site2), 229* nt (site4), 393* nt (site7) and 558* nt (site 9). In the 8th lane, it is not clear whether the ribozyme is active or not because the corresponding cut-site 11 results in an RFP fragment of 702* nucleotides which we cannot differentiate from the full length RFP (although the band at the top does appear to be slightly lower in this lane). The absence of cleavage for the cut-site 5 was expected as the accessibility 2 of the 5(3) is 0 so the ribozyme is not able to hybridize with the target.

The other lanes are combinations of 2 or 3 ribozymes where the full length RFP is not detectable in most cases, with the corresponding bands for cleaved fragments.

Band intensity was measured by using ImageQuant to calculate the cleavage percentage in each case (every ribozyme by itself and the different combinations).

All the ribozymes have a high cutting percentage (88%-98%) except the ribozymes 5(3) and the mismatch control which are not active. The combination of different ribozymes gives better results we can see that all the RNA is degraded when combining three ribozymes (cleavage goes from 97% to 100%). Note that the negative control (RFP alone) is not 0% because of the presence of some background. Also, many weaker bands can be observed likely because the full length RFP RNA was not perfectly pure and may have included a shorter fragment (which could also be cleaved), which is the most likely explanation for these bands, including the bands at the bottom corresponding roughly to a size of 30 nt.

Figure S2. As in Figure S1, three ribozymes were chosen to be tested on a fragment of the 16S RNA of *Microcystis aeruginosa*. The "16S" lane is the RNA in absence of ribozyme, the band represents the labelled target (full length 800*nt). In the other lanes, the same RNA is incubated in presence of ribozymes, with the same convention as explained in Figure S1. The ribozymes 7(m) and 7(4) have an "accessibility 2" of 0 and are not active, as for 5(m) and 5(3) in figure S1.

4(6) 4(6) 4(6) 4(6) 4(1) 4(1) 4(1) 4(1)

+ + + + + + + +

7(4) 7(4) 7(1) 7(1) 7(4) 7(4) 7(1) 7(1)

+ + + + + + + +

9(7) 9(1) 9(7) 9(1) 9(7) 9(1) 9(7) 9(1)

7(4) 7(4) 7(1) 7(1) 4(6) 4(6) 4(6) 4(6) 4(1) 4(1) 4(1) 4(1)

+ + + + + + + +

9(7) 9(1) 9(7) 9(1) 9(7) 9(1) 7(4) 7(1) 9(7) 9(1) 7(4) 7(1)

5 (m) 5(3) 11(1)9(7) 9(1) 7(4) 7(1) 4(6) 4(1) 2(4) RFP

Full length
726*

Rz 9:
558*+168

Rz 7:
393*+333

Rz 4:
229*+497

Rz 2:
153*+573

726*

558*

393*

229*

165*

153*

cleavage% 100 98 98 98 96 97 97 98 83 88 88 97 96 97 97 98 97 97 98 97 39 39 45 86 88 93 88 98 93 75 45



Supplementary Figure S2

Sequences

The target sequences used in this study, as well as the corresponding ribozymes, can be found below. All the potential “GUC” target sites have been annotated in red. Similarly, the different regions of the ribozymes (and corresponding regions on the target RNAs) have been highlighted in different colors to help visualize the ribozymes and cleavage sites: gray, stem 1; green, stem 3; turquoise, stem 2; yellow, conserved regions of CUGANGA and GAAA; finally, pink, nucleotides important for the tertiary interaction between stem I and stem II. For RFP, in cases where regions of complementarity with the ribozymes can vary in length depending on the ribozyme used for targeting, different shadings of the same colors are used.

PABPN1

Transcription product of the new target = 938 nt

```
144 5' GCGACUACGGGAAC GGCCUGGA GUCUGAGGAACUGGAGCCUGAGGAGCUGCUGCUGGAGCCCGAGCCGG
AGCCCGAGCCCGAAGAGGAGCCGCCCGGCCCGCCCGGGAGCUCGCGGGCCUGGGCCUGGUUCG
GGAGCCCCCGGCAGCCAAGAGGAGGAGGAGGAGCCGGGACUGGUCCGAGGGUGACCCGGGGGACGGCGCCAU
363 UGAGGACCCGGAGCUGGAAGC UAUCAAGCUCGA GUCAGGGAGAUGGAGGAAGAAGCUGAGAAGCUAAAAG
437 AGCUACAGAACGAGGUAGAGAAGC AGAUGAAUAUGA GUCCACCUCCAGGCAAUGCUGGCCCGGUGAUC AUG
UCCAUGAGGAGAAGAUGGAGGCUGAUGCCCCGUCCAUCUAUGUUGGCAAUGUGGACUAUGGUGCAACAGC
AGAAGAGCUGGAAGCUCACUUUCAUGGCUGUGGUUCAGUCAACCGUGUUACCAUACUGUGUGACAAAUUUA
GUGGCCAUCCCAAAGGUUUGCGUAUAUAGAGUUCUCAGACAAAGAGUCAGUGAGGACUUCUUGGCCUUA
GAUGAGUCCCUAUUUAGAGGAAGGCAAUCAAGGUGAUCCAAAACGAACCAACAGACCAGGCAUCAGCAC
AACAGACCGGGUUUCCACGAGCCCGUACCGCGCCCGGACCACCAACUACAACAGCUCCCGCUCUCGAU
867 UCUACAGUGGUUUUAACAGCAGGCCCCG GGGUCGC GUCUACAGGGGCCGGGCUAGAGCGACAUCAUGGUUAU
UCCCCUUAUAAAAAAGUGUGUAUUAGGAGGAGAGAGAGGAAAAAAGAGGAAAGNARGGAAAAA
RAAUUAAAAAAGAAAAACAGAAAAAUAUUAAGCGGCCGCCACCGCGGUGGAGCUC
CAGCUUUUGUCCCUUU 3'
```

YZ144

```
5' CAGGCUCCAGUUUAACCUCA CUGAUGA GUCGCUGAAAU GCGAC GAAA CUCCAGGCC 3'
```

YZ363

```
5' UCUUCCUCCAUCUAAUCCCU CUGAUGA GUCGCUGAAAU GCGAC GAAA CUCGAGCUUUGAUA 3'
```

oldYZ363

```
5' GGGUCUUCCUCCAUCUAAUCCCU CUGAUGA GUCGCUGAAAU GCGAC GAAA CUCGAGCUUUGAUA 3'
```

Yz437

```
5' GCAUUGCCUGGUAAAGGUG CUGAUGA GUCGCUGAAAU GCGAC GAAA CUCAUAUUCAUCU 3'
```

YZ867

```
5' CCGCCCUAACUGUA CUGAUGA GUCGCUGAAAU GCGAC GAAA CGCGACCC 3'
```

RFP-Expression Cassette

5' UCUAGAGAAAGAGGAGAAUACUAGAUGGCUUCCUCCGAAGACGUUAUCAAGAGUUCAUGCGUUUCA
AAGUUCGUAUGGAAGGUUCCGUUACGGUCACGAGUUCGAAAUCGAAGGUGAAGGUGAAGGUCGUA
CGAAGGUACCCAGACCGCUAAACUGAAAGUUACCAAAGGUGGUCGCGUGCCGUUCGCUUGGACAUCUCCG
UCCCCGCAGUUCAGUACGGUUCCAAAGCUUACGUUAAACACCCGGCUGACAUCCCGGACUACCUGAAAC
UGUCUUCUCCGGAAGGUUUCAAAUGGGAACGUGUUAUGAACUUCGAAGACGGUGGUGUUGUUACCGUUAC
CCAGGACUCCUCCUGCAAGACGGUGAGUUAUCUACAAAGUUAAACUGCGUGGUACCAACUUCGCUCC
GACGGUCGCUUUAUGCAGAAA AAAACCAUGGGUUGGGAAGCUUCCACCGAACGUAUGUACCCGGAAGACG
GUGCUCUGAAAGGUGAAAUCAAAAUGCGUCUGAAACUGAAAGACGGUGGUCACUACGACGCUGAAGUUAA
AACCACCUACAUGGCUAAAAACCGUUCAGCUGCCGGGUGCUUACAAAACCGACAUCAAACUGGACAUC
ACCUCACACAACGAAGACUACACCAUCGUUGAACAGUACGAACGUGCU GAAGGUCGUCACUCCACCGGUG
CUUAAUAA 3'

GUC2 (rank4)

5' GUACCUUCUAAGUACG CUGAUGA GUCGCU GAAAUGCGACGAAACGACCCA 3'

GUC4 (rank1)

5' CUGGAACUAAGCGGG CUGAUGA GUCGCU GAAAUGCGACGAAACAGGAUG 3'

GUC4 (rank6)

5' GAACUUAAGCGGG CUGAUGA GUCGCU GAAAUGCGACGAAACAGGAUGU 3'

GUC7 (rank1)

5' UUCUGCAUUA AAAACCG CUGAUGA GUCGCU GAAAUGCGACGAAACCGUCGGACG 3'

GUC7 (rank4)

5' UUUCUGCAUUA AAAACCG CUGAUGA GUCGCU GAAAUGCGACGAAACCGUCGG 3'

GUC9 (rank1)

5' AGCGUCUAAGUAGU CUGAUGA GUCGCU GAAAUGCGACGAAACACCGUCU 3'

GUC9 (rank7)

5' GCGUCUAAGUAGU CUGAUGA GUCGCU GAAAUGCGACGAAACACCGU 3'

GUC11 (rank1)

5' AGCACCGGUUAAGGAGU CUGAUGA GUCGCU GAAAUGCGACGAAACGACCUUC 3'

GUC5 (m)

5' GCAUCCGUAAAGGAAG CUGAUGA GUCGCU GAAAUGCGACGAAACAGUUUC 3'

GUC5 (4)

5' CUUCCGUAAAGGAAG CUGAUGA GUCGCU GAAAUGCGACGAAACAGUUUCA 3'

16S rRNA fragment from *Microcystis aeruginosa*

5' AGAGUUUGAUCCUGGCUCAGGAUGAACGCUGGGCGGCGUGCCUAACACAUGCAAGUCGAACGGGAAUCU
UCGGAUUCUAGUGGCGGACGGGUGAGUAACGCGUAAGAAUCUAACUUCAGGACGGGGACAACAGUUGGAA
ACGACUGCUAAUACCCGAUAUGCCGCGAGGUGAAACCUAUUUGGCCUGAAGAAGAGCUUGCGUCUGAUUA
GCUAGUUGGUGGGGUAAGAGCCUACCAAGGCGACGAUCAGUAGCUGGUCUGAGAGGAUGAGCAGCCACAC
UGGGACUGAGACACGGCCCAGACUCCUACGGGAGGCAGCAGUGGGGAAUUUUCCGCAAUGGGCGAAAGCC
UGACGGAGCAACGCCGCGUGAGGGAGGAAGGUCUUUGGAUUGUAAACCUCUUUUUCUAAGGAAGAAGUUC
UGACGGUACUUGAGGAAUCAGCCUCGGCUAACUCCGUGCCAGCAGCCGCGGUAUACGGGGGAGGCAAGC
GUUAUCCGGAAUUUUGGGCGUAAAGCGUCCGAGGUGGUCAGCCAGUCUGCUAAAUCAGGUUGCU
UAACGACCUAAAGGCGGUGGAAACUGGCAGACUAGAGAGCAGUAGGGGUAGCAGGAAUCCAGUGUAGC
GGUGAAAUGCGUAGAGAUUGGGAAGAACAUCGGUGGCGAAAGCGUGCUACUGGGCUGUAUCUGACACUCA
GGGACGAAAGCUAGGGGAGCGAAAGGGAUUAGAUACCCUUGUAGUCCUAGCCGUAACGAUGGAUACUAG
GCGUGGCUUGUAUCGACCCGAGCCGUGCCGAAGCU 3'

GUC7 (m)

5' UUAACCUUAAGAUUU CUGAUGA GUCGC UGAAAU GCGAC GAAA CAAGCUGACUU 3'

GUC7 (4)

5' AGCAACCUUAAGAUUU CUGAUGA GUCGC UGAAAU GCGAC GAAA CAGCAGACUU 3'

GUC3 (1)

5' ACAAUUAACCAAA CUGAUGA GUCGC UGAAAU GCGAC GAAA CCUUCU 3'