

Engineering a family of synthetic splicing ribozymes

Austin J. Che^{1,2,*} and Thomas F. Knight Jr^{1,2}

¹Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar St, Cambridge, MA 02139 and ²Ginkgo BioWorks, 7 Tide St, Boston, MA 02210, USA

Received November 29, 2009; Revised February 16, 2010; Accepted March 4, 2010

ABSTRACT

Controlling RNA splicing opens up possibilities for the synthetic biologist. The *Tetrahymena* ribozyme is a model group I self-splicing ribozyme that has been shown to be useful in synthetic circuits. To create additional splicing ribozymes that can function in synthetic circuits, we generated synthetic ribozyme variants by rationally mutating the *Tetrahymena* ribozyme. We present an alignment visualization for the ribozyme termed as *structure information diagram* that is similar to a sequence logo but with alignment data mapped on to secondary structure information. Using the alignment data and known biochemical information about the *Tetrahymena* ribozyme, we designed synthetic ribozymes with different primary sequences without altering the secondary structure. One synthetic ribozyme with 110nt mutated retained 12% splicing efficiency *in vivo*. The results indicate that our biochemical understanding of the ribozyme is accurate enough to engineer a family of active splicing ribozymes with similar secondary structure but different primary sequences.

INTRODUCTION

RNA splicing is common in eukaryotes but not found in the simpler bacterial systems that are often used when engineering synthetic biological systems. Being able to control and engineer RNA splicing would increase the number of ways we can regulate gene expression. Group I splicing ribozymes are self-contained RNA splicing elements that can be engineered in bacterial systems (1). To test our understanding and to expand the family of usable ribozymes, we sought to engineer new splicing ribozymes.

Schultes and Bartel (2) showed that synthetic ribozymes could be designed to fall on a neutral path between two unrelated ribozymes. Each step on the path changed <2 nt and preserved ribozyme activity. Along the path was one sequence that could adopt both ribozyme folds. Thus,

ribozyme folding is highly flexible and relatively independent of the primary sequence. Group I splicing ribozymes also typically have low primary sequence conservation, but they fold into a similar secondary structure (3,4). To take advantage of this sequence flexibility, we designed new splicing ribozymes to have low primary sequence identity but high secondary and tertiary structural identity.

We used the *Tetrahymena* ribozyme as the model splicing ribozyme (Figure 1). Although the *Tetrahymena* ribozyme itself is 413 nt, the standard numbering labels the guanosine added in the first step of splicing as 1, so the bases are numbered from 2 to 414. We use nucleotides 28–414 from the wild-type ribozyme. Helical double-stranded regions are numbered sequentially with a P (paired region) and helical loops are labeled with L (loop region). To reduce ambiguity, we label domains with a D. D₂, D_{4–6}, D_{3,7,8} and D₉ are the four major domains that form the ribozyme. For example, D₂ consists of P2 and P2.1. D_{4–6} consists of all of the P4, P5 P5abc, P6 and P6ab helices.

MATERIALS AND METHODS

Sequence alignment analysis

To understand the importance of each base in the ribozyme, we analyzed an alignment of 837 group IC1 ribozymes (the subgroup containing the *Tetrahymena* ribozyme) from the Group I Intron Sequence and Structure Database (GISSD) (5). The alignment was processed to make *structure information diagrams*, similar to sequence and structure logos (6,7), but instead of mapping information content on to a linear 'logo', bases are drawn as a secondary structure. The information content is not represented by the height of the base but rather by its color.

The total information $I(i)$ at position i in the alignment is calculated as

$$I(i) = \left(1 - \frac{n(i,-)}{n(i,-) + \sum_{b \in B} n(i,b)} \right) \left(\sum_{b \in B} J(i,b) \right)$$

*To whom correspondence should be addressed. Tel: +1 234 542 3965; Email: austin@csail.mit.edu

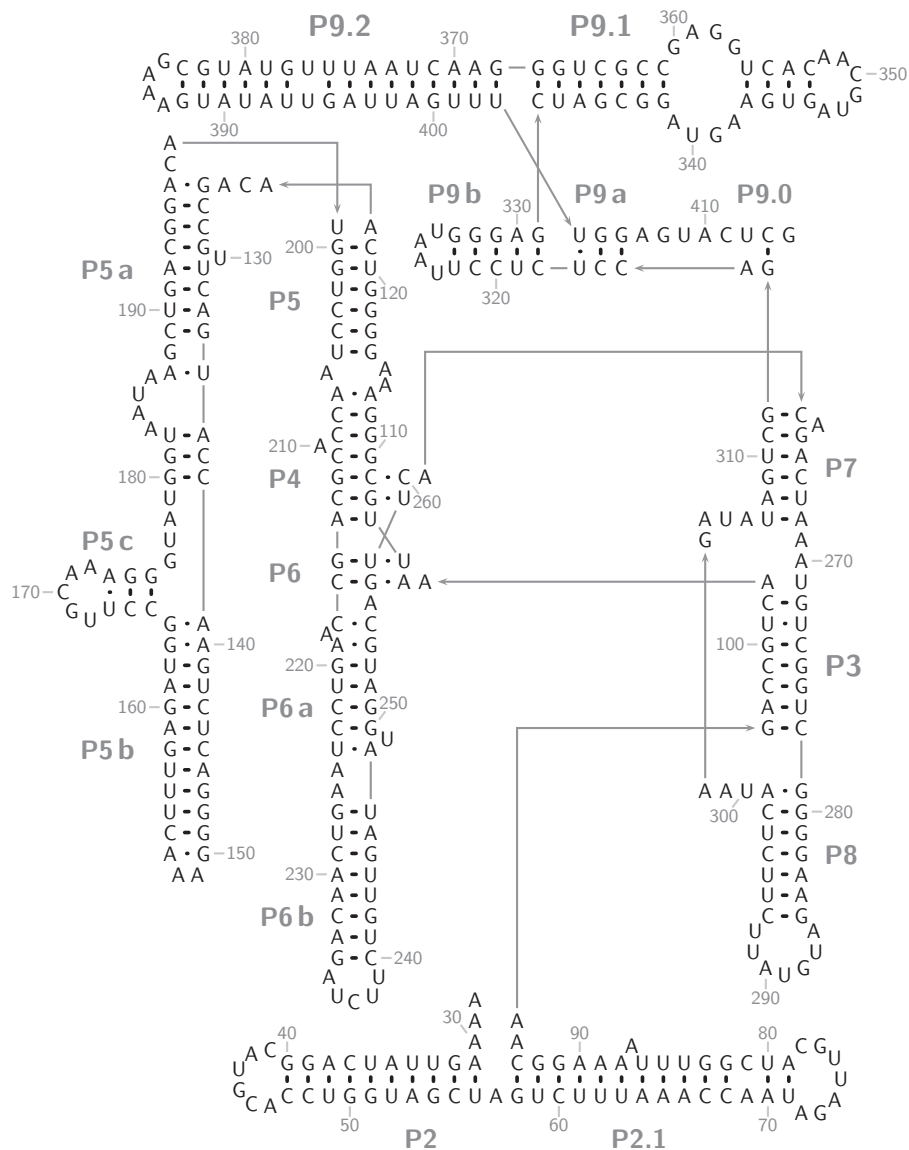


Figure 1. The secondary structure of the *Tetrahymena* ribozyme consists of a series of paired helical regions. The canonical numbering is based on the wild-type ribozyme.

where $B = \{A, C, G, U\}$, $n(i, -)$ is the number of sequences containing a gap at position i , $n(i, b)$ is the number of sequences containing base b at position i , and

$$J(i, b) = f(i, b) \cdot \log_2 \frac{f(i, b)}{0.25}$$

$$f(i, b) = \frac{n(i, b)}{\sum_{b \in B} n(i, b)}$$

The 0.25 indicates that all four bases are expected to occur with equal frequency. In calculating base frequencies, gaps are ignored, but the total information is reduced by the frequency of gaps (7). The color of a base is determined by $f(i, b) \cdot I(i)$, which is between 0 and 2 bits. If $J(i, b)$ is negative, the base is displayed upside down to indicate that it occurs less than expected (6).

In a sequence logo, bases at each position are stacked in order of increasing frequencies (7). To reduce visual clutter, a structure information diagram only shows one base or gap at every position. However, multiple structure information diagrams can represent all the information in a sequence logo. Figure 2 shows the most frequent base (or gap) at each position in the alignment, mapped on to the secondary structure of the *Tetrahymena* ribozyme. If a gap occurs most frequently at a position, it is represented with a dash. A black dot is shown at positions where the base has <0.1 bits of information. We could similarly show the second most frequent base (or gap) at each position (Supplementary Figure S1). Although additional diagrams could be used to show even less frequent bases, no position contains more than ~ 0.1 bits of information in these other diagrams (data not shown).

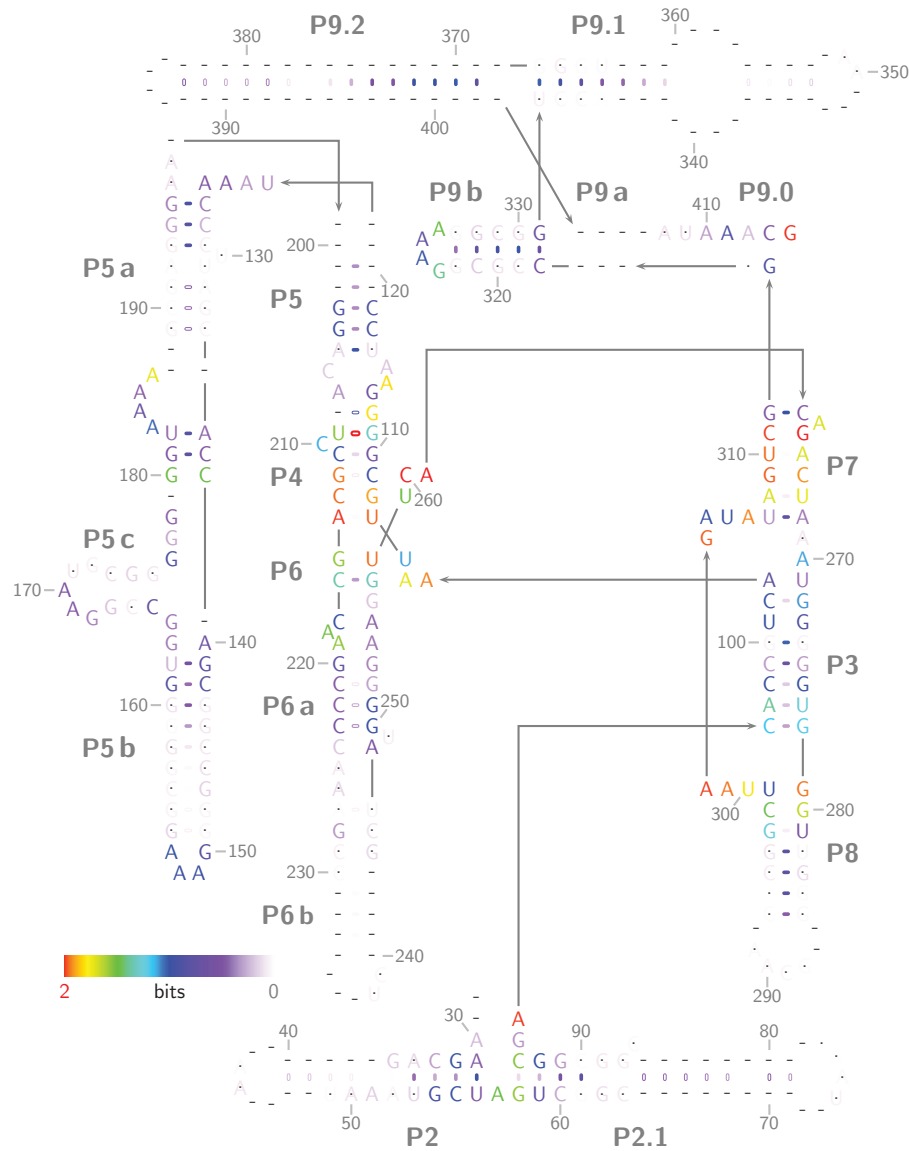


Figure 2. The structure information diagram from the group IC1 ribozyme alignment is mapped to the *Tetrahymena* secondary structure. At each position, the most common base (consensus) is shown. See text for details.

Just as in structure logos, base pairs can contain mutual information not found in the bases themselves (6). For example, even if all four bases are found equally at two positions (zero positional information), the bases at the two positions could co-vary to always base pair (high mutual information). The additional information from a base pair is calculated using the log-likelihood ratio of the observed to the expected frequency of base pairing. Both frequencies are calculated after eliminating sequences with gaps. The observed frequency of a base pairing is calculated as the number of sequences with the base pairing divided by the number of sequences that contain a non-gap base at one or both of the positions in the base pair. The expected frequency of base pairing between positions i and j is equal to $\sum f(i, b) \cdot f(j, c)$ for all combinations of bases b and c that can base pair (Watson-Crick plus G:U). As the base pairing for the *Tetrahymena* ribozyme

in the alignment sometimes differed from the reference pairing used (Figure 1), we only used the base pairs common to both.

The mutual information from base pairing is represented on the structure information diagram as the color of the base pairs and is drawn using the same scale as for the bases. Although the base pair information can be >2 bits, it is capped at 2 bits. If the actual frequency of a base pair is less than the expected frequency, then the base pair is drawn in outline form, rather than completely filled.

Ribozyme design

From the alignment and information known about each base in the ribozyme (Supplementary Table S2), we generated a map of positions in the ribozyme where the identity of the base is likely unimportant. ‘Harmless’ bases are defined as positions having a total information content

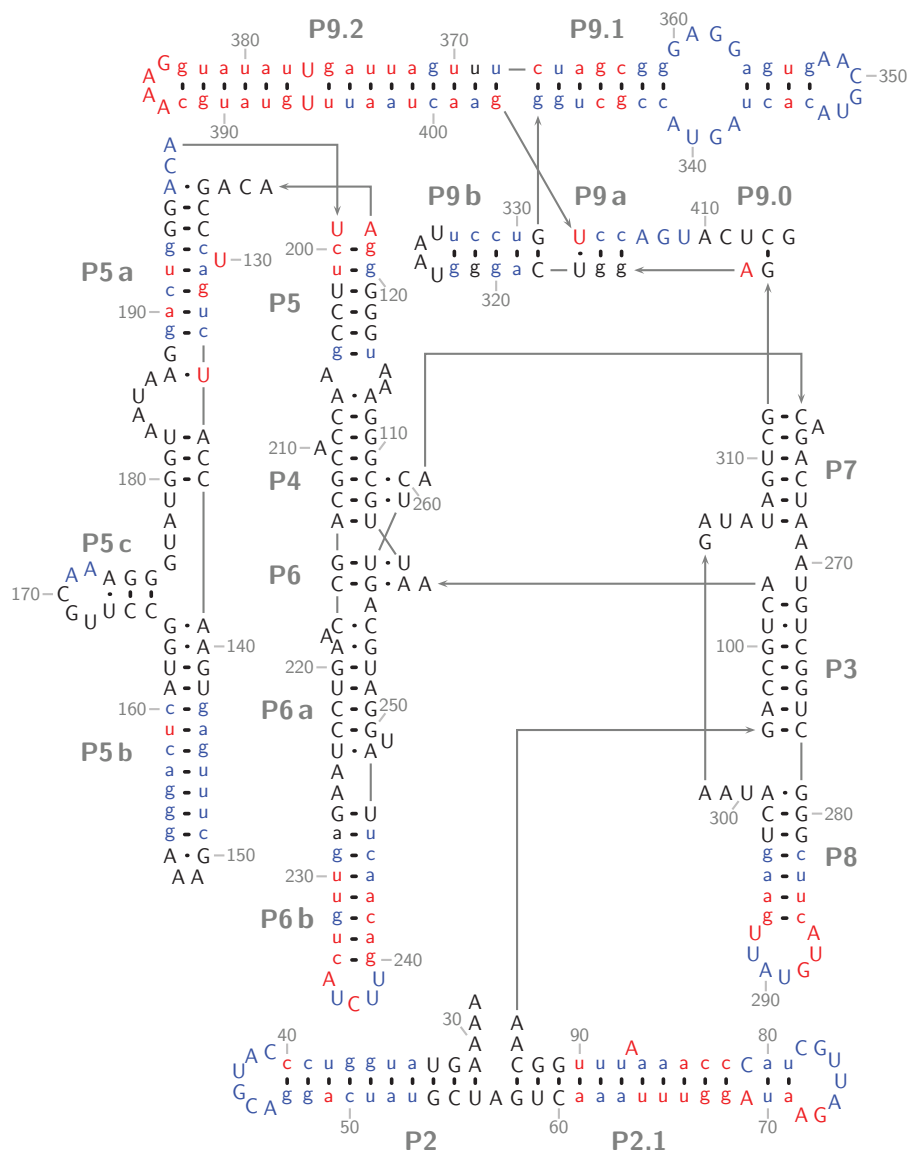


Figure 3. The sequence of a designed ribozyme is shown. Red bases indicate 'harmless' positions and blue bases indicate 'likely mutable' positions (as defined in text). Lowercase letters indicate bases that were swapped from the wild-type ribozyme sequence in Figure 1.

<0.05 and have no known tertiary interactions with other positions. 'Likely mutable' bases have a maximum total information content of 0.25, may have some tertiary interactions, but can likely be changed with some care or additional experimentation.

To use a relatively conservative approach, we only made changes by swapping bases within a base pair. These swaps should maintain the secondary structure while changing the primary sequence. Figure 3 shows the design of a ribozyme, containing 150 base changes, which is 39% of the sequence.

Experimental methods

All ribozymes were tested using *in vivo* splicing assays. The *Tetrahymena* ribozyme in previously tested *cis*-splicing systems was replaced with ribozyme variants in order to

quantify the splicing efficiency of different ribozymes. All constructs were cloned in pSB1A3, a pUC19 derived plasmid, transformed into Top10 (DH10B) *Escherichia coli* and verified by sequencing.

For base pair mutagenesis, we assayed activity using a *cis*-splicing module with LacZ α as a reporter. LacZ α is formed only upon successful splicing of the ribozyme. No background activity was detectable when an inactive G264A point mutant was used. Individual base pairs were swapped using site-directed mutagenesis. Some clones containing single mutations were also characterized. LacZ α activity for each mutant was measured and normalized to the non-mutated construct (Supplementary Methods).

For testing synthetic ribozyme variants, we used a *cis*-splicing module with GFP as a reporter. This ribozyme reporter system splits the codon for the Tyr66 fluorophore

of GFP such that there is no possibility of background GFP fluorescence in the absence of splicing. Synthetic ribozymes were made via DNA synthesis and PCR. To calculate splicing efficiency, the maximum GFP synthesis rate for the ribozyme variant was normalized to the wild-type ribozyme (Supplementary Methods).

RESULTS

Mutagenesis characterization

We swapped single base pairs in the *Tetrahymena* ribozyme and measured the relative splicing efficiency of the mutated ribozyme. The change in splicing efficiency indicates the importance of the bases beyond base pairing, such as additional stacking or tertiary interactions. We measured the efficiencies from nine base pair swaps, with most being in D₄₋₆. Supplementary Table S2 includes the efficiencies of tested variants. As expected, switching the guanosine binding site 264:311 destroyed activity. All other base pair swaps maintained activity. The only base pair swap that was found to be truly neutral on splicing efficiency was 116:205. Some single base mutations were generated incidentally during site-directed mutagenesis and were also characterized. All single base mutations were worse than the compensatory double mutation, indicating the importance of base pairing and the secondary structure over the primary sequence.

Synthetic ribozymes

We attempted to mutate the *Tetrahymena* ribozyme into new synthetic ribozymes with varied primary sequence. We determined conservative base swaps that were unlikely to affect the secondary structure or function of the ribozyme (Figure 3). To test individual domains of the ribozymes, we constructed different synthetic ribozyme combinations. Table 1 shows the splicing efficiencies for the synthetic ribozyme variants.

A G264A point mutant in the catalytic core served as a negative control and showed an order of magnitude lower activity than the least efficient synthetic ribozyme tested. Several variants (indicated with a star in Table 1) contained additional base swaps not present in Figure 3 and were generated before we gathered additional information indicating that those base swaps could be harmful. As could be expected, these starred variants showed relatively low splicing efficiencies.

Large portions of the D₂ and D₉ peripheral domains can be mutated with only a modest decrease in splicing efficiency. Comparing SZ11–SZ13, it appears that the base pairs 322:327 and 346:353 in D₉ contribute to splicing efficiency and should not be changed. SZ20 contains 110 mutations located in all regions except P5 and P5abc and still had easily measurable splicing activity. The synthetic P6b and P8 helices (SZ2–SZ4) were the least disruptive and even had a slight beneficial effect for ribozyme splicing.

Mutations in the P5abc region appear to be the most detrimental. SZ10 contained mutations throughout D₄₋₆ and had low splicing efficiency. Mutations in P5 (SZ1 and

Table 1. The splicing efficiencies for synthetic ribozyme variants were normalized to the wild-type ribozyme (SZ0)

Name	Regions changed	No. of nucleotides changed	Efficiency
SZ0	None	0	1.00 ± 0.01
-Control*	G264A	1	<0.003
SZ1	P5 – 116	4	0.85 ± 0.03
SZ2	P8	8	1.14 ± 0.06
SZ3	P6b – 228	12	1.09 ± 0.02
SZ4	P6b + SZ2	22	1.36 ± 0.10
SZ5	SZ6 – 70,71	30	0.70 ± 0.03
SZ6	D ₂	34	0.64 ± 0.05
SZ7*	P5 + P5abc – 116 + 135,166,167	34	0.02 ± 0.00
SZ8*	SZ7 + 122	36	0.01 ± 0.00
SZ9	SZ5 + SZ2	38	0.80 ± 0.01
SZ10*	D ₄₋₆ + 166	46	0.12 ± 0.01
SZ11	SZ12 – 322	60	0.56 ± 0.03
SZ12	SZ13 – 346	62	0.44 ± 0.04
SZ13	D ₉	64	0.28 ± 0.02
SZ14	SZ11 + SZ2	68	0.58 ± 0.01
SZ15	SZ12 + SZ2	70	0.25 ± 0.02
SZ16	SZ11 + P2	74	0.43 ± 0.02
SZ17	SZ18 – 343,344,345	84	0.13 ± 0.01
SZ18	SZ5 + SZ11	90	0.15 ± 0.01
SZ19	SZ18 + SZ2	98	0.13 ± 0.00
SZ20	SZ19 + SZ3	110	0.12 ± 0.01
SZ21*	SZ20 + SZ8	146	0.03 ± 0.01

Ribozyme changes are specified relative to Figure 3 and other ribozymes with any differences specified using the first base number of the base pair. For example, SZ6 - 70,71 indicates that the all base pair flips from SZ6 are present (all bases in the P2 and P2.1 helices use the design from Figure 3) except for the base pairs at positions 70 and 71 (70:80 and 71:79). The number of bases different between each ribozyme and the wild-type *Tetrahymena* sequence is shown, out of a total of 387 nt. The starred names indicate ribozymes with base changes not present in Figure 3. The complete ribozyme sequences are shown in the Supplementary Data.

swap of 116:205) and P6b appeared to be benign. Thus, the P5abc region is the likely cause for the inefficiency of the SZ10 ribozyme. Four of the base pairs in P5abc were individually swapped and all showed reasonably efficient splicing. Either one of the untested base pairs in P5abc is responsible for significantly affecting splicing or the mutations in combination have a deleterious effect. The most likely detrimental mutation is the C166:G174 base swap. Both of these bases may form alternate base pairs during the folding process (8) and should not have been mutated.

Figure 4 shows the splicing efficiency for the non-starred ribozymes dropped linearly relative to the number of nucleotides changed. Only one of the synthetic ribozymes did not follow the trend line. The ribozyme containing a synthetic P6b and P8 with 22 nt changed (SZ4) had significantly greater splicing efficiency than even the wild-type ribozyme. The starred ribozymes containing base swaps not present in Figure 3 would fall far off the trend line, indicating that they contain harmful base swaps.

DISCUSSION

Expanding the ribozyme family

One approach for obtaining a new splicing ribozyme is to use one of the many existing ribozymes in the family.

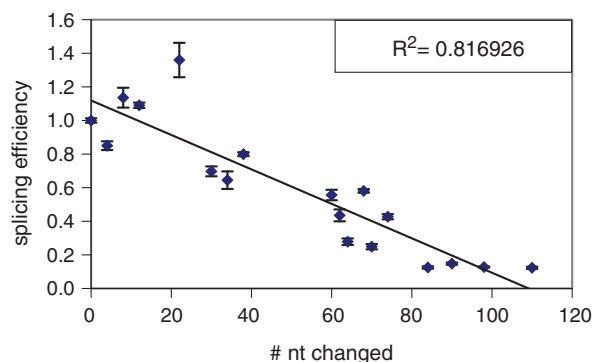


Figure 4. For the non-starred synthetic ribozymes in Table 1, the number of nucleotides changed is plotted versus splicing efficiency. The splicing efficiency dropped linearly as more nucleotides were changed.

Most of the ribozymes in the family were determined to be similar by sequence or structure alignment. Despite the large number of splicing ribozymes determined by alignment, only several have ever been experimentally characterized and may not function as a self-splicing ribozyme in a bacterial host. Other ribozymes that have been studied include the ribozymes from *Azoarcus* (9,10), *Pneumocystis* (11,12), *Didymium iridis* (DiGIR2) (13,14), and *Fuligo* (Fse.L569 and Fse.L1898) (13). We tested two uncharacterized sequences found in the alignment for their ability to function as self-splicing ribozymes, but neither ribozyme showed *cis*-splicing activity (Supplementary Data).

An alternative approach to expanding the number of usable ribozymes is to tweak an existing functional ribozyme. Previous work has shown the generation of a family of splicing ribozymes using random selection (15). By randomizing peripheral regions around a core P3–P7 catalytic domain, a large number of ribozymes with high activity can be selected.

Although a selection-based approach allows for generating a large number of active ribozymes, maintaining a desired base pairing and secondary structure is difficult. Thus, it would be difficult to generate ribozymes with a large number of changes in peripheral regions which depend on the secondary structure. Our approach uses rational design coupled with secondary structure information to simultaneously mutate a large number of bases in the ribozyme. In the process, we directly test the completeness of our knowledge about the structure of splicing ribozymes.

Structure information diagram

To help visualize sequence alignment information for large RNA structures, we developed the structure information diagram. The structure information diagram maps the information content found in sequence logos on to a secondary structure diagram to allow for a more natural visualization (Figure 2). The catalytic core and the non-conserved peripheral regions can be easily visualized.

Base pair information is also represented in the structure information diagram. 110:209 is a high information base pair that occurs much less than expected by chance.

Many sequences in the alignment are missing base 208 so that base 110 pairs with base 210 instead. Thus, 110:209 is an unusual base pair found in *Tetrahymena* but not in many other ribozymes. Other high information base pairs are 262:312 and 116:205. 262:312 is almost equally a G:C or C:G base pair. Thus, even though the individual bases do not have high information content, the base pair is conserved. Similarly, at base pair 116:205, all of the pairings U:A, C:G and G:U occur with high frequencies.

These diagrams map the alignment on to the *Tetrahymena* structure, so only alignment positions for which the *Tetrahymena* ribozyme does not have a gap are shown. There were only several positions in the alignment with positive information content where the *Tetrahymena* ribozyme has a gap and the consensus base is not a gap (Supplementary Table 1). The limited number of such positions indicates that most conserved bases are present in the *Tetrahymena* ribozyme. The position with the highest information content, 207.1, indicates that many ribozymes contain an A between positions 207 and 208. When an A was inserted after position 207 in the *Tetrahymena* ribozyme, the ribozyme showed no splicing activity. Thus, there are limits to using sequence alignment to infer non-harmful changes that can be made to the ribozyme.

Experimental mutagenesis

Although alignments can provide useful information about functionally important bases, ultimately, experiments are needed to test and verify our understanding of the ribozyme. As an initial effort at understanding how to manipulate the ribozyme, we generated a small set of base pair swaps and characterized the change in splicing efficiency. Base pairs that can be swapped without significantly altering splicing efficiency would be good targets for future ribozyme engineering. Completing this work by measuring the effect from switching every base pair (around 125 total base pairs) is experimentally feasible and would help us better understand the core ribozyme.

Synthetic ribozymes

Previous work has shown that many single base pair flips completely inactivate the ribozyme (16). Our results show that the ribozyme can remain active over many changes, as long as the appropriate bases are changed. Very few bases of the primary sequence are strictly conserved in group I ribozymes. Even in the P7 catalytic core region, except for the guanosine binding site G264:C311, the primary sequence can be changed, whereas the secondary structure usually needs to be maintained (17,18). We designed synthetic ribozymes with new primary sequences while trying to maintain the secondary structure and splicing activity of the ribozyme.

Using the available information about each base in the ribozyme, we generated lists of harmless and likely mutable bases. Around half of the bases can likely be changed (Figure 3). To generate synthetic ribozymes, we switched base pairs and tested different groups of base swaps for splicing efficiency. Over 150 different bases

were individually changed and it is unexpected that so many bases can be changed without eliminating splicing activity. The SZ20 ribozyme with 110 nt simultaneously changed still showed a 12% splicing efficiency.

There is a difference in the free energies of stacked base pairs (e.g. GC:CG does not have the same energy as CG:GC). The fact that so many base pairs could be successfully switched is evidence of the ribozyme's robustness to folding conditions. However, an increased number of mutations generally reduced splicing efficiency. Base changes can affect the folding process in subtle ways that are currently unpredictable. One way to work around possible folding problems is to use mutagenesis and selection on the designed ribozymes to bring the efficiency back up.

Many more synthetic ribozyme variants could be generated. We did not attempt to mutate unpaired bases which is another source for generating many ribozyme variants. Some of the ribozyme domains can support more significant changes beyond base pair swaps. For example, ribozymes with inserted tags, coding sequences or other payloads could be useful. The peripheral loops of P6b and P8 may tolerate the addition of a significant amount of sequence. All of the ribozyme variants with base changes in P6b and P8, e.g. SZ4, showed improved or unchanged splicing efficiencies. We can also likely add sequence to the D₂ and D₉ domains. The P3, P4, P5, P6 and P7 helices form the catalytic core and should be manipulated with caution.

One region that is not well-understood is P5abc, which is not conserved but mutations in this region can strongly affect splicing efficiency. P5abc is found only in a small number of group I ribozymes, but is essential for the *Tetrahymena* ribozyme (19,20). The D₄₋₆ domain containing P5abc folds quickly and helps with the correct folding and assembly of the slow-folding D_{3,7,8} domain (21,22). P5abc likely helps in the folding process by stabilizing the ribozyme through tertiary interactions. Adding P5abc *in trans* can rescue splicing from ribozymes missing this domain (23). Destabilizing mutations in P5abc have been found to *increase* the rate of folding of D_{3,7,8} (24). If the ribozyme normally enters a kinetic trap, then destabilizing P5abc can allow escaping the kinetic trap, leading to a faster overall folding rate. However, all P5abc mutants here showed less efficient splicing. Clarifying the contribution of P5abc towards splicing would enable engineering this region of the ribozyme.

Engineering a minimal ribozyme would provide a scaffold for new synthetic ribozymes and test the limits of our ribozyme knowledge. Nearly 75% of the ribozyme can be deleted one section at a time without destroying activity *in vitro* (25). Deleting the entire D₉ domain, except P9.0, produces a ribozyme more active than wild-type. Deleting both P6b and D₉ is also more active than wild-type. Deleting both D₂ and D₉ or both P5 and D₉ maintains activity, whereas deleting both D₂ and P5 does not produce an active ribozyme. Using the available information, a minimal ribozyme should be relatively straightforward to design and test.

CONCLUSION

Synthetic ribozymes can give us better ribozymes. Some of the ribozymes generated here were more efficient at splicing than the wild-type ribozyme. Selection could likely produce even better ribozymes. The ribozymes were not characterized beyond their ability to perform one *cis*-splicing reaction. There are other possible reactions catalyzed by the ribozyme. When Williams *et al.* (26) selected new P5abc domains, they obtained ribozymes that could self-splice but were deficient in the 3'-hydrolysis reaction. As the 3'-hydrolysis reaction is an unproductive side reaction, ribozymes capable of splicing but unable to hydrolyze the 3'-exon would be an improvement. More work is needed to understand how to not only design equivalent ribozymes, but also to design better ribozymes.

Supplementary Table S2 collects information about each base in the *Tetrahymena* ribozyme from the literature and characterization experiments done on the ribozyme. Understanding the ribozyme core will facilitate its use as a standard and reusable component of engineered biological systems.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Defense Science and Engineering Graduate Fellowship (to A.J.C); Funding for open access charge: US National Science Foundation Synthetic Biology Engineering Research Center.

Conflict of interest statement. MIT provided research support and has filed a patent application on some aspects of this work.

REFERENCES

- Che, A.J. (2008) *Engineering RNA logic with synthetic splicing ribozymes PhD thesis*. Massachusetts Institute of Technology.
- Schultes, E.A. and Bartel, D.P. (2000) One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science*, **289**, 448–452.
- Cech, T.R. (1990) Self-splicing of group I introns. *Ann. Rev. Biochem.*, **59**, 543–568.
- Woodson, S.A. (2005) Structure and assembly of group I introns. *Curr. Opin. Struct. Biol.*, **15**, 324–330.
- Zhou, Y., Lu, C., Wu, Q.J., Wang, Y., Sun, Z.T., Deng, J.C. and Zhang, Y. (2008) GISSD: group I intron sequence and structure database. *Nucleic Acids Res.*, **36**, D31–D37.
- Gorodkin, J., Heyer, L.J., Brunak, S. and Stormo, G.D. (1997) Displaying the information contents of structural RNA alignments: the structure logos. *Comput. Appl. Biosci.*, **13**, 583–586.
- Schneider, T.D. and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.
- Wu, M. and Tinoco, I. (1998) RNA folding causes secondary structure rearrangement. *Proc. Natl Acad. Sci. USA*, **95**, 11555–11560.
- Hayden, E.J., Riley, C.A., Burton, A.S. and Lehman, N. (2005) RNA-directed construction of structurally complex and active ligase ribozymes through recombination. *RNA*, **11**, 1678–1687.

10. Riley, C.A. and Lehman, N. (2003) Generalized RNA-directed recombination of RNA. *Chem. Biol.*, **10**, 1233–1243.
11. Alexander, R.C., Baum, D.A. and Testa, S.M. (2005) 5' transcript replacement *in vitro* catalyzed by a group I intron-derived ribozyme. *Biochem.*, **44**, 7796–7804.
12. Baum, D.A., Sinha, J. and Testa, S.M. (2005) Molecular recognition in a trans excision-splicing ribozyme: non-Watson-Crick base pairs at the 5' splice site and omegaG at the 3' splice site can play a role in determining the binding register of reaction substrates. *Biochemistry*, **44**, 1067–1077.
13. Fiskaa, T., Lundblad, E.W., Henriksen, J.R., Johansen, S.D. and Einvik, C. (2006) RNA reprogramming of alpha-mannosidase mRNA sequences *in vitro* by myxomycete group IC1 and IE ribozymes. *FEBS J.*, **273**, 2789–2800.
14. Lundblad, E.W., Haugen, P. and Johansen, S.D. (2004) Trans-splicing of a mutated glycosylasparaginase mRNA sequence by a group I ribozyme deficient in hydrolysis. *Eur. J. Biochem.*, **271**, 4932–4938.
15. Ohuchi, S.J., Ikawa, Y., Shiraishi, H. and Inoue, T. (2002) Modular engineering of a group I intron ribozyme. *Nucleic Acids Res.*, **30**, 3473–3480.
16. Couture, S., Ellington, A.D., Gerber, A.S., Cherry, J.M., Doudna, J.A., Green, R., Hanna, M., Pace, U., Rajagopal, J. and Szostak, J.W. (1990) Mutational analysis of conserved nucleotides in a self-splicing group I intron. *J. Mol. Biol.*, **215**, 345–358.
17. Oe, Y., Ikawa, Y., Shiraishi, H. and Inoue, T. (2000) Analysis of the P7 region within the catalytic core of the Tetrahymena ribozyme by employing *in vitro* selection. *Nucleic Acids Symp. Ser.*, **44**, 197–198.
18. Oe, Y., Ikawa, Y., Shiraishi, H. and Inoue, T. (2001) Conserved base-pairings between C266-A268 and U307-G309 in the P7 of the Tetrahymena ribozyme is nonessential for the *in vitro* self-splicing reaction. *Biochem. Biophys. Res. Commun.*, **284**, 948–954.
19. Ayre, B.G., Köhler, U., Turgeon, R. and Haseloff, J. (2002) Optimization of trans-splicing ribozyme efficiency and specificity by *in vivo* genetic selection. *Nucleic Acids Res.*, **30**, e141.
20. Köhler, U., Ayre, B.G., Goodman, H.M. and Haseloff, J. (1999) Trans-splicing ribozymes for targeted gene delivery. *J. Mol. Biol.*, **285**, 1935–1950.
21. Pan, J. and Woodson, S.A. (1998) Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *J. Mol. Biol.*, **280**, 597–609.
22. Zarrinkar, P.P. and Williamson, J.R. (1994) Kinetic intermediates in RNA folding. *Science*, **265**, 918–924.
23. van derHorst, G., Christian, A. and Inoue, T. (1991) Reconstitution of a group I intron self-splicing reaction with an activator RNA. *Proc. Natl Acad. Sci. USA*, **88**, 184–188.
24. Treiber, D.K., Rook, M.S., Zarrinkar, P.P. and Williamson, J.R. (1998) Kinetic intermediates trapped by native interactions in RNA folding. *Science*, **279**, 1943–1946.
25. Beaudry, A.A. and Joyce, G.F. (1990) Minimum secondary structure requirements for catalytic activity of a self-splicing group I intron. *Biochemistry*, **29**, 6534–6539.
26. Williams, K.P., Imahori, H., Fujimoto, D.N. and Inoue, T. (1994) Selection of novel forms of a functional domain within the Tetrahymena ribozyme. *Nucleic Acids Res.*, **22**, 2003–2009.