Supplementary materials for:

Computational design and experimental verification of pseudoknotted ribozymes

Sabrine Najeh[1]*, Kasra Zandi[2]*, Nawwaf Kharma[3#] and Jonathan Perreault[1#]

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

2 Software Engineering & Computer Science Dept., Concordia University, 1455 boul. de Maisonneuve O., Montreal, QC, H3G 1M8, Canada

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

*authors contributed equally as indicated in RNA

#To whom correspondence should be addressed. Tel: +450 687 5010 (Ext 4411); Fax: +450 686 5301; Email: jonathan.perreault@iaf.inrs.ca; and Tel: +514 848-2424 (Ext 3117); Email: nawwaf.kharma@concordia.ca

Table of content

Supplementary Information

Table S1 Sequences of trans-acting ribozymes, their substrates and corresponding oligonucleotides

Ribozymes	Oligo Sequences	RNA sequences
Rz WT Mouse	TTTGGGTAGTTTCGCCGGCTCTCACCGTTGCTCATCAGGGGATTTATTCGGTACCCCTATAGTGAGTCGTATTA	ggGGUACCGAAUAAAUCCCCUGAUGAGCAACGGUGAGAGCCGGCGAAACUACCCAAA
Sb WT Mouse	TTATGGTACTATCCCGACTACCCTTGCCTATAGTGAGTCGTATTA	ggCAAGGGUAGUCGGGAUAGUACCAUAA
HHRz1	GAATATAACTTTCGCCTACGGTGTAACTGGCTCATCAGGACCCTTTTCGCTACGACCTATAGTGAGTCGTATTA	ggUCGUAGCGAAAAGGGUCCUGAUGAGCCAGUUACACCGUAGGCGAAAGUUAUAUUC
HHSb1	AGTATATATTGGGTCGAGTTATAATGCCTATAGTGAGTCGTATTA	ggCAUUAUAACUCGACCCAAUAUAUACU
HHRz2	GTTCTACAGTTTCGCTTTTATTCAGAGTTGCTCATCAGGGCGGAAGGGCTCCGTTCCTATAGTGAGTCGTATTA	ggAACGGAGCCCUUCCGCCCUGAUGAGCAACUCUGAAUAAAAGCGAAACUGUAGAAC
HHSb2	GTATAGAAAACCGCCGACCGTAGGTACCTATAGTGAGTCGTATTA	ggUACCUACGGUCGGCGGUUUUCUAUAC
HHRz3	TATGCTTGCTTTCGGTTTTTTTCTAGCTTCCTCATCAGGCCGCTTTTTGCTAGTTCCTATAGTGAGTCGTATTA	ggAACUAGCAAAAAGCGGCCUGAUGAGGAAGCUAGAAAAAAACCGAAAGCAAGCAUA
HHSb3	GTGTTTTTTTGCGGCGAGCAAGCGTACCTATAGTGAGTCGTATTA	ggUACGCUUGCUCGCCGCAAAAAAACAC
HHRz5	TTCACGAAGTTTCGCTTTTTTTGTCCCTCGCTCATCAGGGCCCTTTTTGGGACTTCCTATAGTGAGTCGTATTA	ggAAGUCCCAAAAAGGGCCCUGAUGAGCGAGGGACAAAAAAAGCGAAACUUCGUGAA
HHSb5	TTTTTTTCTCGGGCCGACTTCGCTCTCCTATAGTGAGTCGTATTA	ggAGAGCGAAGUCGGCCCGAGAAAAAAA
HHRz7	TGAACCAGGTTTCGTTTTTTTCGCGGGTCACTCATCAGGCCCCTCTCTCCCGCTTCCTATAGTGAGTCGTATTA	ggAAGCGGGAGAGAGGGGCCUGAUGAGUGACCCGCGAAAAAAACGAAACCUGGUUCA
HHSb7	CACTTCTCCTGGGGCGACCTGGCAGCCCTATAGTGAGTCGTATTA	ggGCUGCCAGGUCGCCCCAGGAGAAGUG
HHRz8	CATGGGCCCTTTCGCTTTTTTTCCACCTCGCTCATCAGGGCACTTTTTGGTGGTTCCTATAGTGAGTCGTATTA	ggAACCACCAAAAAGUGCCCUGAUGAGCGAGGUGGAAAAAAAGCGAAAGGGCCCAUG
HHSb8	CTTTTTTTTTGTGCCGAGGACCCGTTCCTATAGTGAGTCGTATTA	ggAACGGGUCCUCGGCACAAAAAAAAAG
HHRz9	TCTAGTGCCTTTCGTTTTTTTCCGGACTCACTCATCAGGCCGCTTCTTGTCCGCTCCTATAGTGAGTCGTATTA	ggAGCGGACAAGAAGCGGCCUGAUGAGUGAGUCCGGAAAAAAACGAAAGGCACUAGA
HHSb9	CGTTTTATTTGCGGCGAGGCACTCTACCTATAGTGAGTCGTATTA	ggUAGAGUGCCUCGCCGCAAAUAAAACG
HHRz10	GTTCGGACCTTTCGCTTTTTTTCCCAGGTCGCTCATCAGGGCCCTTTTTTCCTGGTTCCTATAGTGAGTCGTATTA	ggAACCAGGAAAAAGGGCCCUGAUGAGCGACCUGGGAAAAAAGCGAAAGGUCCGAAC
HHSb10	CGTTTTTTTTTGGGCCGAGGTCCGCTCCCTATAGTGAGTCGTATTA	ggGAGCGGACCUCGGCCCAAAAAAAACG

HHRz11	**TCAATCCCGTTTCGCTTTTTTTGAGACTCGCTCATCAGGGCCTTTTTTGTCTCTCCCTATAGTGAGTCGTATTA**	**ggGAGAGACAAAAAAGGCCCUGAUGAGCGAGUCUCAAAAAAAGCGAAACGGGAUUGA**
HHSb11	**TCTTTCCTCCAGGCCGACGGGACATACCTATAGTGAGTCGTATTA**	**ggUAUGUCCCGUCGGCCUGGAGGAAAGA**
HHRz12	**GTCATGGAGTTTCGCGTTTTTTGGTCCGTGCTCATCAGGGCCCTTTTTGGACCTTCCTATAGTGAGTCGTATTA**	**ggAAGGUCCAAAAAGGGCCCUGAUGAGCACGGACCAAAAAAACGCGAAACUCCAUGAC**
HHSb12	**GGTTTTTTGTGGGCCGACTCCACTCTCCTATAGTGAGTCGTATTA**	**ggAGAGUGGAGUCGGCCCACAAAAAACC**
HHRz13	**CTCGTGCTGTTTCGCTTTTTTTGAGGGTCGCTCATCAGGAACCTTTTTCCCTCTTCCTATAGTGAGTCGTATTA**	**ggAAGAGGGAAAAAGGUUCCUGAUGAGCGACCCUCAAAAAAAGCGAAACAGCACGAG**
HHSb13	**TTCTTTTTTTGGTTCGACAGCACTTTCCTATAGTGAGTCGTATTA**	**ggAAAGUGCUGUCGAACCAAAAAAAGAA**
HHRz13 mut	**CTCGTGCTGTTTCGCTTTTTTTGGGGTTCGCTCATCAGGAACCTTTTTCTCTTTTCCTATAGTGAGTCGTATTA**	**ggAAAAGAGAAAAAGGUUCCUGAUGAGCGAACCCCAAAAAAAGCGAAACAGCACGAG**
HHRz13 mut com	**CTCGTGCTGTTTCGCTTTTTTTTAGGATTCGCTCATCAGGAACCTTTTTTATCCTTTCCTATAGTGAGTCGTATTA**	**ggAAAGGAUAAAAAGGUUCCUGAUGAGCGAAUCCUAAAAAAAAGCGAAACAGCACGAG**
RzWT YL	**GCCCCCACTGTTGATGTTTCGACGCTAGTCATGGGTTTGTTCTCATCAGGGCAGCCAGTCCCCCCCTATAGTGAGTCGTATTA**	**ggGGGGGACUGGCUGCCCUGAUGAGAACAAACCCAUGACUAGCGUCGAAACAUCAACAGUGGGGGC**
Sb WT YL	**TTATGACTAGTGGCTGCCGACACCAACACCTATAGTGAGTCGTATTA**	**ggUGUUGGUGUCGGCAGCCACUAGUCAUAA**
Rz YL1	**ATGCTCCGTTTCTCCTTCATCAAATCTCTTTGGATCATCAGGGAAGACGACACTATCCTATAGTGAGTCGTATTA**	**ggAUAGUGUCGUCUUCCCUGAUGAUCCAAAGAGAUUUGAUGAAGGAGAAACGGAGCAU**
Sb YL1	**TCTTTGATAACGTCTTCCGACGAAGCCCCTATAGTGAGTCGTATTA**	**ggGGCUUCGUCGGAAGACGUUAUCAAAGA**
RzYL2	**ATGCCTCGTTTCTCCTTCATCAAATTTTTTTGGATCATCAGGGAAGACGGGTATATCCTATAGTGAGTCGTATTA**	**ggAUAUACCCGUCUUCCCUGAUGAUCCAAAAAAAUUUGAUGAAGGAGAAACGAGGCAU**
SbYL2	**TCTTTGATAACGTCTTCCGACGAAGCCCCTATAGTGAGTCGTATTA**	**ggGGCUUCGUCGGAAGACGUUAUCAAAGA**

For the hammerhead ribozyme designs HHRz_1 to HHRz_14, *trans*-acting ribozyme assays were performed by opening the middle of the loop of stem III. Related information is also available in **Table S1**. Note that two “G” bases are added in 5´ (to allow efficient transcription), this is also true for the substrate of the *trans*-acting version.

Table S1 (continued) Oligonucleotide sequence for cis-acting ribozymes

Name	Sequence
T7	**TAATACGACTCACTATAGG**

WT	TTATGGTACTATCCCGACTACCCTTGTTTGGGTAGTTTCGCCGGCTCTCA CCGTTGCTCATCAGGGGATTTATTCGGTACCCCTATAGTGAGTCGTATTA
HHRz 1	AGTATATATTGGGTCGAGTTATAATGGAATATAACTTTCGCCTACGGTGT AACTGGCTCATCAGGACCCTTTTCGCTACGACCTATAGTGAGTCGTATTA
HHRz 2	GTATAGAAAACCGCCGACCGTAGGTAGTTCTACAGTTTCGCTTTTATTCA GAGTTGCTCATCAGGGCGGAAGGGCTCCGTTCCTATAGTGAGTCGTATTA
HHRz 3	GTGTTTTTTTGCGGCGAGCAAGCGTATATGCTTGCTTTCGGTTTTTTTCT AGCTTCCTCATCAGGCCGCTTTTTGCTAGTTCCTATAGTGAGTCGTATTA
HHRz 4	TTTTTTTTTTGGGTCGACCCATCTTAATCGACGGGTTTCGCTTTTTTTGT CGCTTGCTCATCAGGACCCTTTTTGCGACGGCCTATAGTGAGTCGTATTA
HHRz 5	TTTTTTTCTCGGGCCGACTTCGCTCTTTCACGAAGTTTCGCTTTTTTTGT CCCTCGCTCATCAGGGCCCTTTTTGGGACTTCCTATAGTGAGTCGTATTA
HHRz 6	AATAGGTAGAGGTCCGACCGAGAAAAAGATCTCGGTTTCGCTCGTTTTCG GCCTAGCTCATCAGGGACCGGGGGGGCCGTCCCTATAGTGAGTCGTATTA
HHRz 7	CACTTCTCCTGGGGCGACCTGGCAGCTGAACCAGGTTTCGTTTTTTTCGC GGGTCACTCATCAGGCCCCTCTCTCCCGCTTCCTATAGTGAGTCGTATTA
HHRz 8	CTTTTTTTTTTGTGCCGAGGACCCGTTCATGGGCCCTTTCGCTTTTTTTCC ACCTCGCTCATCAGGGCACTTTTTGGTGGTTCCTATAGTGAGTCGTATTA
HHRz 9	CGTTTTATTTGCGGCGAGGCACTCTATCTAGTGCCTTTCGTTTTTTTCCG GACTCACTCATCAGGCCGCTTCTTGTCCGCTCCTATAGTGAGTCGTATTA
HHRz 10	CGTTTTTTTTGGGCCGAGGTCCGCTCGTTCGGACCTTTCGCTTTTTTCCC AGGTCGCTCATCAGGGCCCTTTTTCCTGGTTCCTATAGTGAGTCGTATTA
HHRz 11	TCTTTCCTCCAGGCCGACGGGACATATCAATCCCGTTTCGCTTTTTTTGA GACTCGCTCATCAGGGCCTTTTTTGTCTCTCCCTATAGTGAGTCGTATTA
HHRz 12	GGTTTTTTGTGGGCCGACTCCACTCTGTCATGGAGTTTCGCGTTTTTTGG TCCGTGCTCATCAGGGCCCTTTTTGGACCTTCCTATAGTGAGTCGTATTA
HHRz 13	TTCTTTTTTTGGTTCGACAGCACTTTCTCGTGCTGTTTCGCTTTTTTTGA GGGTCGCTCATCAGGAACCTTTTTCCCTCTTCCTATAGTGAGTCGTATTA
HHRz 14	GGTGTTGTTTGGGTCGACCGGCCTTATCTGACCGGTTTCGCGTTTTTTTA GCCGTGCTCATCAGGACCCTTTTTGGCCATTCCTATAGTGAGTCGTATTA
HHRz mut	GTGTTTTTTTGCGGCGAGCAAGCGTATATGCTTGCGGTCGGTTTTTTTCT AGCTTCCTCATCCCTCCGCTTTTTGCTAGTTCCTATAGTGAGTCGTATTA
WT YL	TAATACGACTCACTATAGGGGGGGACTGGCTGCCCTGATGAGAACAAACCCAT GACTAGCGTCGAAACATCAACAGTGGGGGCTGTTGGTGTCGGCAGCCACTAGT CATAA
R WT YL	TTATGACTAGTGGCTGCCGACAC
Mut YL	TAATACGACTCACTATAGGGGGGGACTGGCTGCCCCAGTGAGAACAAACCCAT GACTAGCGTCGAAACATCAACAGTGGGGGCTGTTGGTGTCGGCAGCCACTAGT CATAA

HHRz05	**ACCCGAGCTCTTTACACTTTATGCTTCCGGCTCGTATGTTATATACCCGTCTT CCCTGATGATCCAAAAAAATTTGATGAAGGAGAAACGAGGCATGGCTTCGTCG GAAGACGTTATCAAAGACGTTATCAAAGAGTTCATGCGTTTCAAAGTTCGTAT GGAAGGTTCCGTTAACATGC**
HHRz 05 mut	**TTTACACTTTATGCTTCCGGCTCGTAATATACCCGTCTTCCCTTGCAATCCAA AAAAATCTGATGAAGGAGAAACGAGGCATGGCTTCGTCGGAAGACGTTATCAA AGA**
F HHRz05	**TAATACGACTCACTATAGGATATACCCGTCTTCCCTGATG**
R HHRz05	**GCATGTTAACGGAACCTTCCATACGAACTTTGAAACGC**
F HHRz05 mut	**TAATACGACTCACTATAGGATATACCCGTCTTCCCTTGC**
F HHRz00	**TAATACGACTCACTATAGGATAGTGTCGTCTTCCCTGATGA**
R HHRz00	**TCTTTGATAACGTCTTCCGAC**
F HHRz Mut	**TAATACGACTCACTATAGGATAGTGTCGTCTTCCCGGATAAT**

***glmS* WT**	**AGCGCCTGGACTTAAAGCCATTGCACTCCGGCTTTAAGTTGACGAG GGCAGGGTTTATCGAGACATCGGCGGGTGCCCTGCGGTCTTCCTGC GACCGTTAGAGGACTGTGAAAACCACAGGCGACTGTGGCATAGAGC AGTCCGGGCAGGAA**
F *glmS* WT	**TAATACGACTCACTATAGGGAAAAAAAAAAAAAAAAAAAG CGCCTGGACTTAAAGCCA**
R *glmS* WT	**TTCCTGCCCGGCTGCTCTATGC**
F *glmS* cleaved	**TAATACGACTCACTATAGGGCCTGGACTTAAAGCCATTGCACTCCGG**
F *glmS* mut	**TAATACGACTCACTATAGGGAAAAAAAAAAAAAAAAAAAAAC CTGGACTTAAAGCCA**
***glmS* 1**	**TGTGCGGTGGCAACCGTGTATTAGGCGTTTTCACGCCTATTTGAGG TCGCCTTTTGGGGCCCGCACTGCCCCCCCAGCTTCCGCCGATTTTT CGATTTGCTGGGCCCTCGTCCTCCATGTCAGCTTTTTTTTTCTGAC ATAGAGCTGGCGCTCGCTATAGTGAGTCGTATTA**
F *glmS* 1	**TAATACGACTCACTATAGGGAAAAAAAAAAAAAAAAAAAG CGCCAGCUCUAUGUCA**
R *glmS* 1	**TGTGCGGTGGCAACCGTGTATTAGG**
F *glmS* 6	**TAATACGACTCACTATAGGGAAAAAAAAAAAAAAAAAAAAGCGCCAGGTCGCG TCTATAAGGTTAAAAGTAGGCGCGACGACGAGGGCCAGCCAAATCGAGACATCG GCGGGAGGCTGG**
R *glmS* 6	**GAGTGCGTCGGTGGCTTTTATGCAGGAGTTTCCCCCTGTTTTGCACCACCTCTC TGGACCCACACCCGGCCCCCAGCCTCCCGCCGATGTCTCGATTTGGCTGGCCCT CGTCGTCGCGCC**

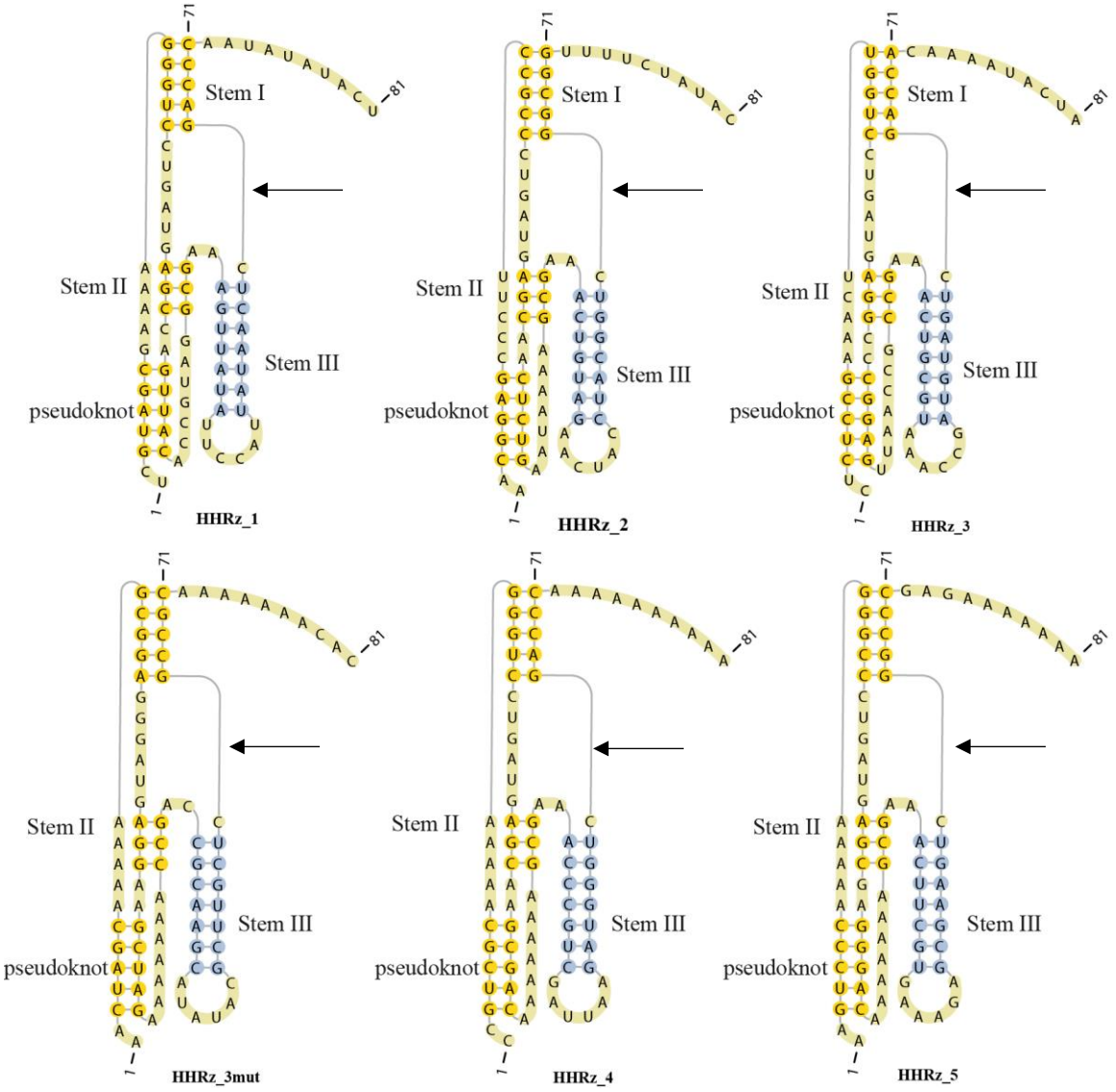
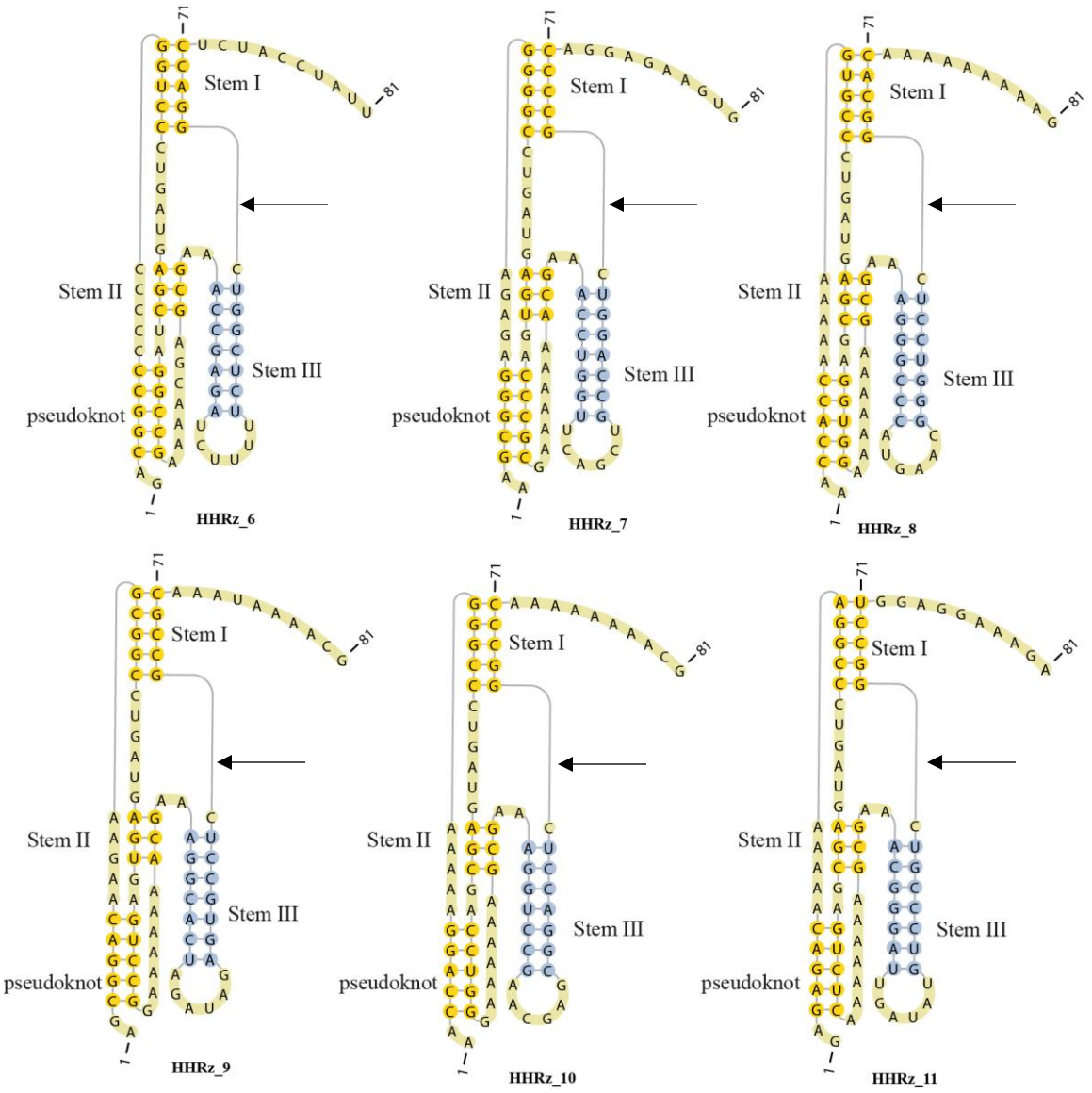


Figure S1 Secondary structures of ribozymes

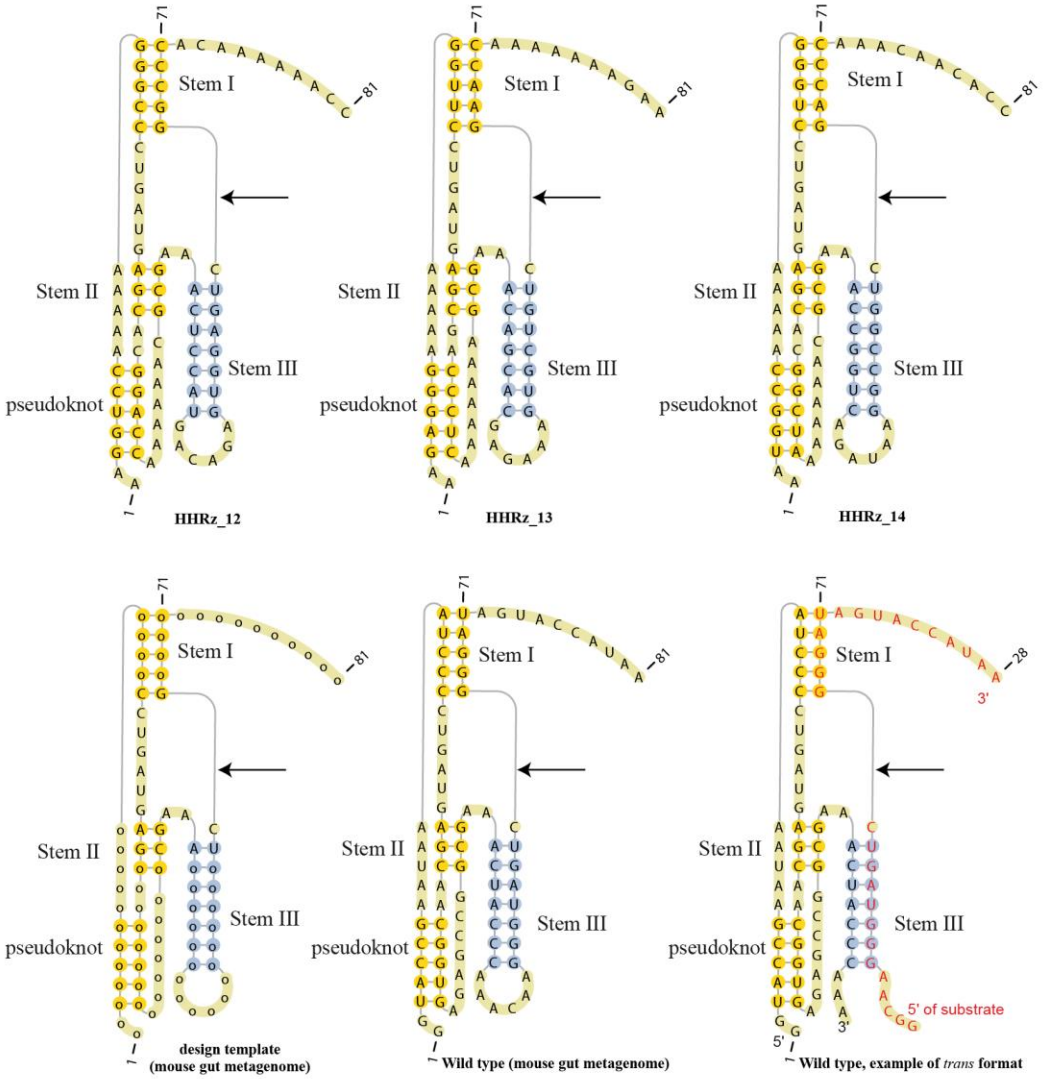
Stems are indicated as per the convention, in addition to the pseudoknot that contributes to bringing stems I and II closer. Beige nucleotides are single stranded, while yellow and blue are double stranded by design. Arrows indicate the cleavage site.

(figure continues on the following pages, a *trans*-acting HHRz example is presented in 2 pages).



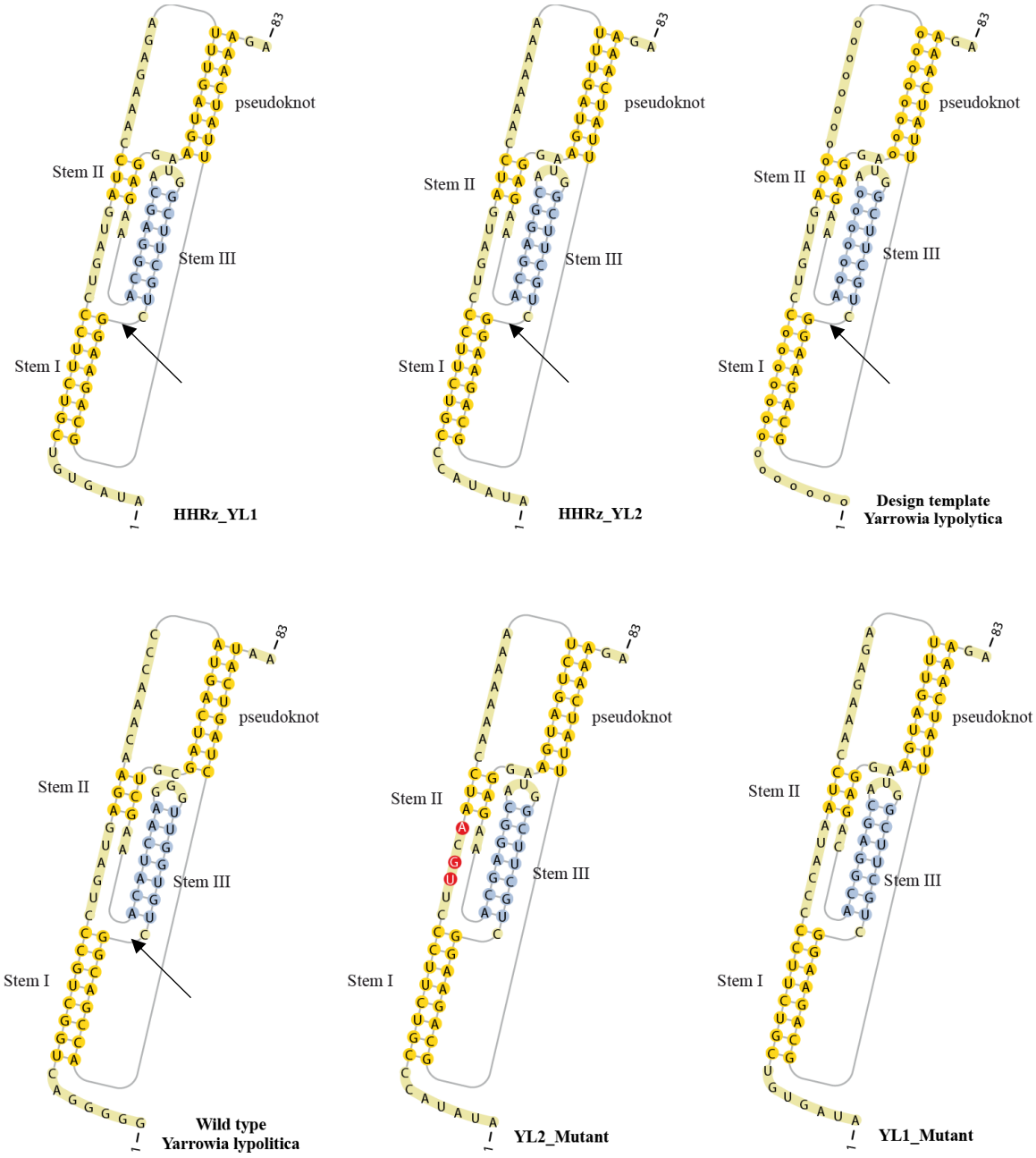
(continued) Figure S1. Secondary structures of ribozyme designs. Stems are indicated as per the convention, in addition to the pseudoknot that contributes to bringing stems I and II closer. Beige nucleotides are single stranded, while yellow and blue are double stranded by design. Arrows indicate the cleavage site.

(figure continues on the following pages, a *trans*-acting HHRz example is presented next page).



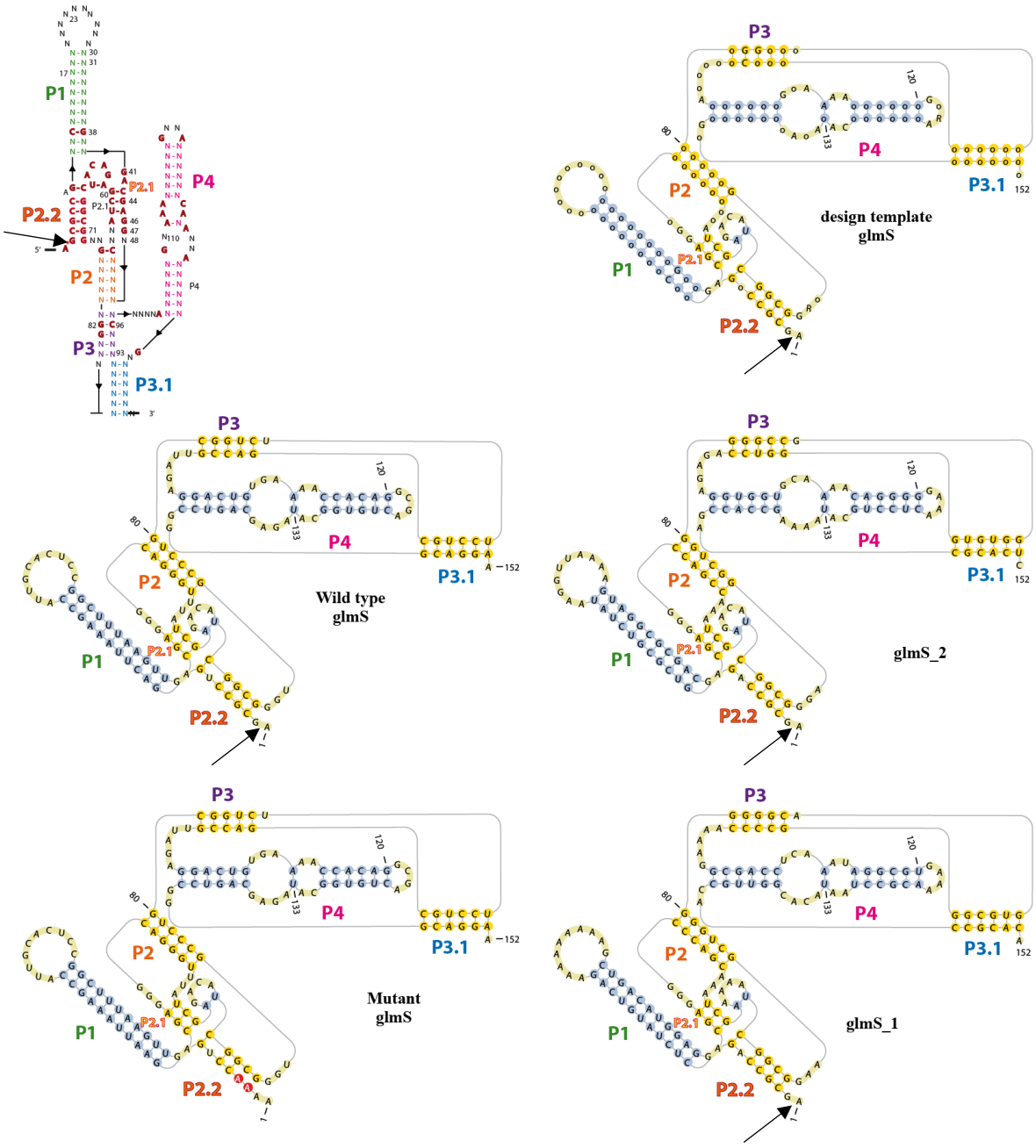
(continued) Figure S1. Secondary structures of ribozyme designs. Stems are indicated as per the convention, in addition to the pseudoknot that contributes to bringing stems I and II closer. Beige nucleotides are single stranded, while yellow and blue are double stranded by design. Arrows indicate the cleavage site. For the design template, open circles correspond to positions free to vary while letters correspond to nucleotide constraints. The last structure (bottom right) provides an example of the *trans*-acting configuration, where the substrate portion of the RNA is colored in red.

(figure continues on the following pages).



(continued) Figure S1. Secondary structures of ribozyme designs. Stems are indicated as per the convention, in addition to the pseudoknot that contributes to bringing stems I and II closer. Beige nucleotides are single stranded, while yellow and blue are double stranded by design. Arrows indicate the cleavage site. For the design template, open circles correspond to positions free to vary while letters correspond to nucleotide constraints. For the mutant, mutated positions are circled in red.

(figure continues on the following page).



(continued) Figure S1. Secondary structures of ribozyme designs. Stems are indicated as per the convention shown on the top left. Beige nucleotides are single stranded, while yellow and blue are double stranded by design. Arrows indicate the cleavage site. For the design template, open circles correspond to positions free to vary while letters correspond to nucleotide constraints. For the mutant, mutated positions are circled in red.

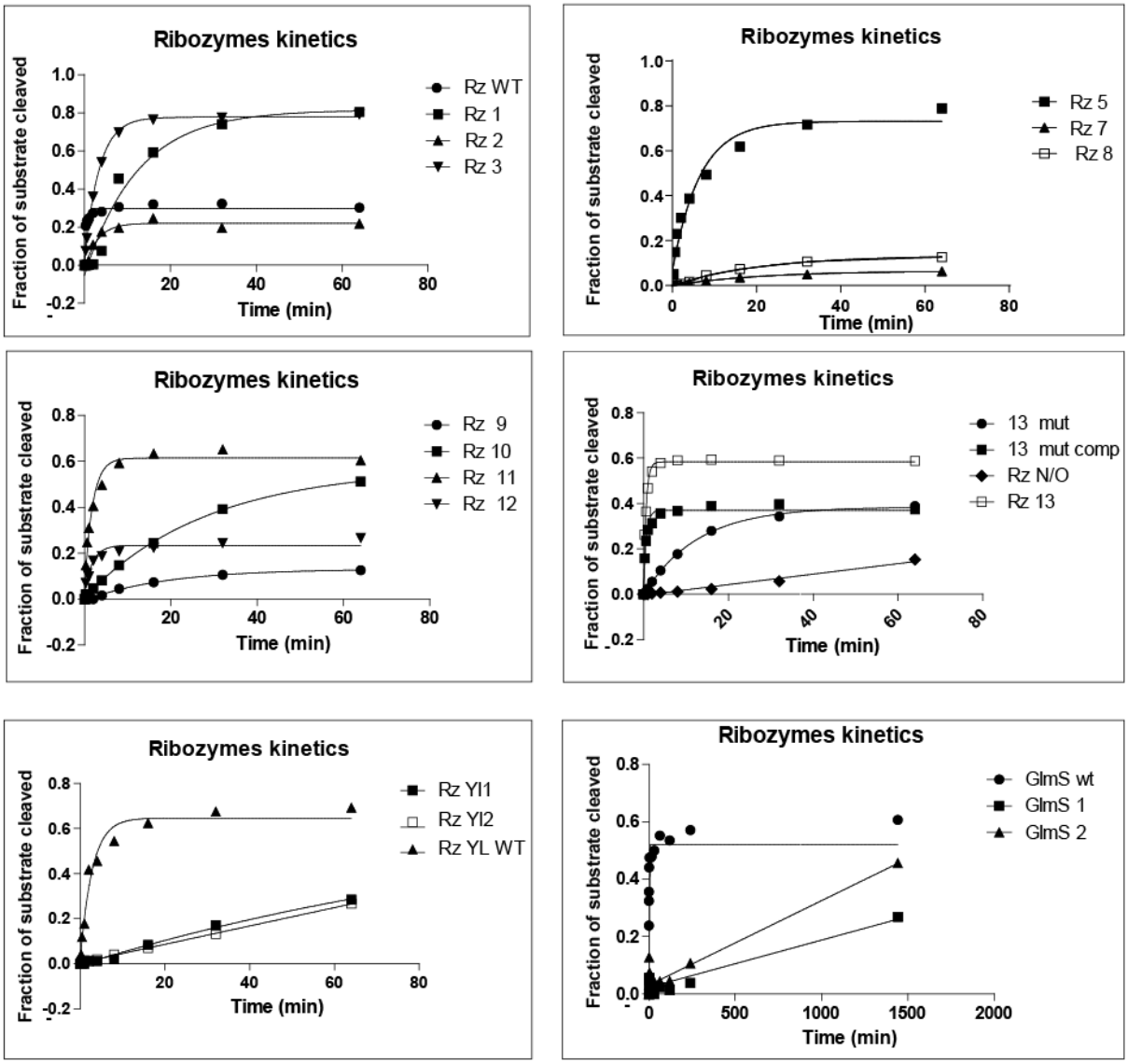


Figure S2 Ribozymes activity

Cleavage over time is plotted for all ribozymes assayed in order to determine the k_{obs} (presented in table S2). Note that this is for the trans-acting versions (except for *glmS*). It can be noted that the WT (from the HHRz found in the mouse gut metagenome) has a much lower % cleavage, with an apparent plateau around 30%. It is nevertheless the HHRz with the best cleavage rate (3 min^{-1}). The Rz13 presenting the best k_{obs} compared to all the other ribozymes shows a higher cleavage % than the WT with a plateau at almost 60%.

- **Supplementary details on sequence design by Enzymer**

The design goal of the algorithm is to interactively evolve the initial template by lowering the predicted normalized ensemble defect of the candidate sequence at each iteration. The algorithm was set to stop after 600 iterations or when the predicted normalized ensemble defect reaches 0.01. In the final sequence, the initial catalytic core remained unchanged. When we run Enzymer, when the run is finished, it returns a single final sequence. The sequences generated internally are of no value and Enzymer does not keep track of them.

Sequences were generated at different temperatures of 17, 27, 37, 47 and 57 (according to Dirks model). For each temperature about 10 sequences were generated, which were sorted from each temperature group based on their ensemble defect and then from each group two or more were picked for experiments (as described in main text), but the ensemble defects were all very close. For the YL and *glmS* they were all designed at 37. The final ensemble defects were very close.

Tables S2, S3, S4 in excel sheets.

Table S2. Details for k_{obs} determination

Detailed results of rate determination are provided in an excel sheet

Table S3. *glmS*1 and *glmS*2 alignment with natural *glmS* ribozyme sequences

Detailed results of *glmS*1 and *glmS*2 alignments with natural *glmS* ribozymes are provided in an excel sheet.

Table S4. Ensemble defect

Detailed calculated free energies and defects at different temperatures for the HHRzs based on the mouse gut model.

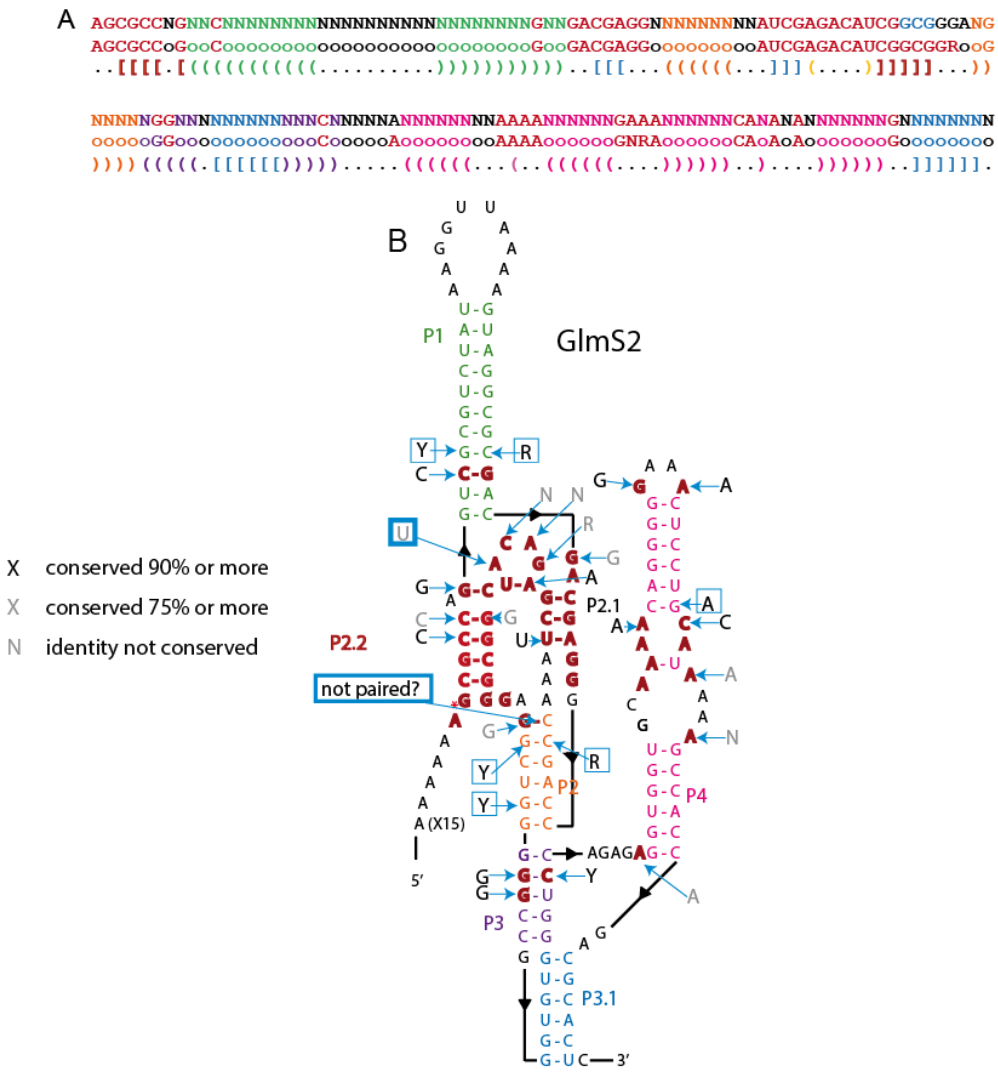


Figure S3 Discrepancies between the consensus from (McCown et al. 2011) and our model

Based on the structure and consensus from (Klein et al. 2006). The structure of *glmS2* is shown as in Figure 3 with the red nucleotides representing all the identities resulting from constraints from the model shown in A (used to generate *glmS2*). All the constrained positions for which the level of conservation is lower than 97% as the consensus (see color coding) of McCown et al. are shown. Positions with blue boxes high-light all the positions for which a discrepancy is observed for the base identity. Blue boxes with thicker lines highlight more notable discrepancies, notably: the U within the tetraloop of the single bp between P2.1 and P2.2 (which is nevertheless an A in the sequence used to determine the structure in Klein et al. 2006, as well as the top base pair of P2, which is absent in the consensus from McCown et al. 2011. In the latter case, the somewhat conserved G seems to be part of the junction between P2.2 and P2 instead of forming a base pair. It can also be noted in the structure determined from the crystal that this bp is not aligned as well as some other bp. Forming a more stable G-C bp might thus be detrimental, compared to a G-U bp, potentially only transient and captured within the snapshot of the structure determined from the crystal.

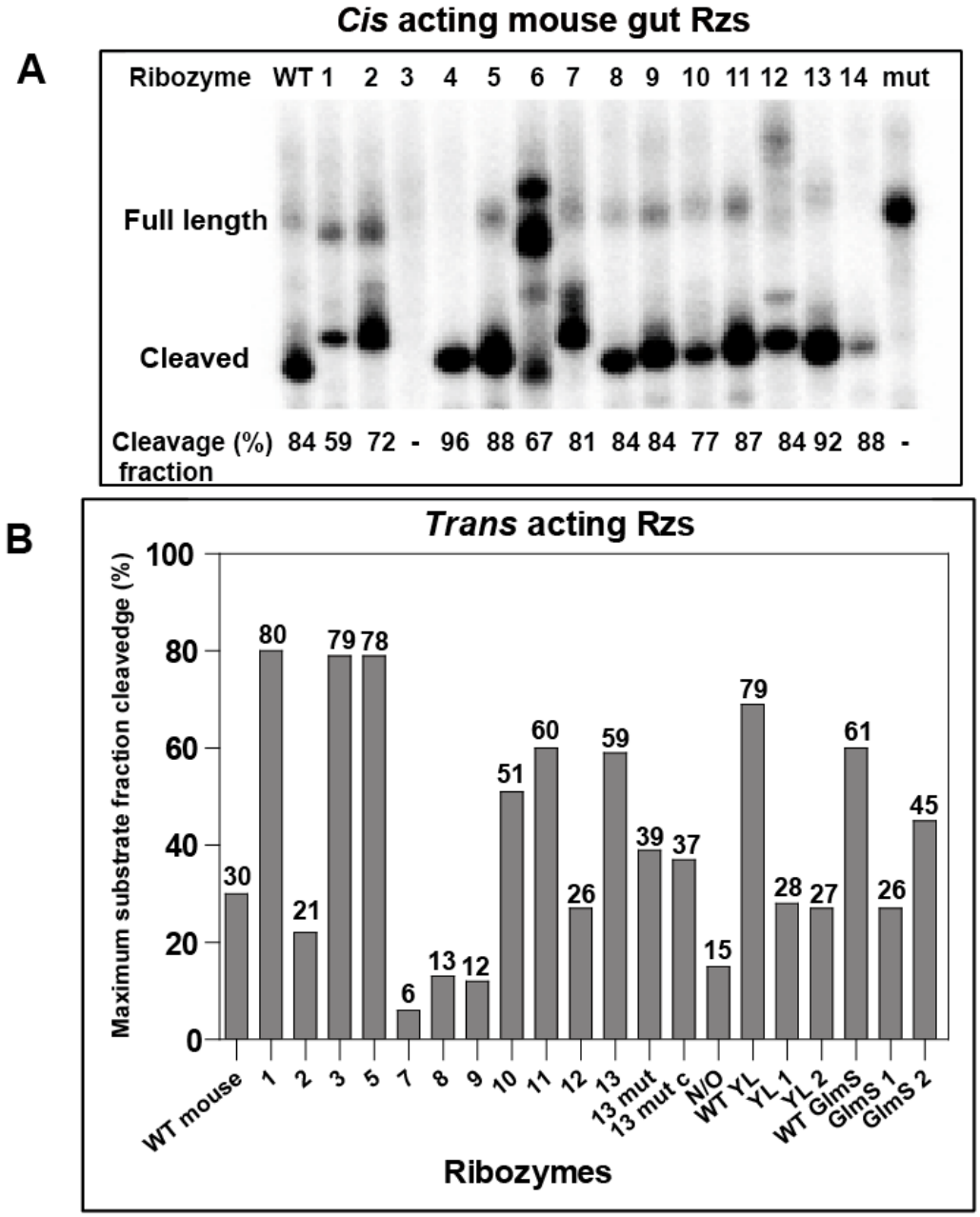


Figure S4 Cleavage extent of cis and trans acting ribozymes

(A) Cleavage activity during transcription, of all selected mouse gut-hammerhead ribozymes including Rzs 4, 6 and 14 that were eliminated from the kinetic tests because of the transcription failure of the *trans* versions. (B) Cleavage extent of all the tested ribozymes resulting from the last aliquot taken at 64 minutes during kinetics following. The used ribozymes were all *trans*-acting except for the *glmS* ribozymes. Note that Rz3 was not properly transcribed in the gel shown in A, this does not reflect its activity, the gel from Fig 1 includes a successful transcription of Rz3.