Electronic Supplementary Material (ESI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2023

**Figure S1**

| A<br>Selection round | B<br>Incubation time (min) | C<br>[cTmp] (mM) | D<br>Number of PCR1 cycles |
|---|---|---|---|
| 1 | 180 | 50 | 15 |
| 2 | 180 | 50 | 15 |
| 3 | 180 | 50 | 15 |
| 4 | 180 | 50 | 15 |
| 5 | 180 | 50 | 7 |
| 6 | 60 | 20 | 7 |
| 7 | 20 | 10 | 6 |
| 8 | 20 | 1 | 8 |
| 9 | 20 | 0.3 | 8 |
| 10 | 20 | 0.1 | 8 |

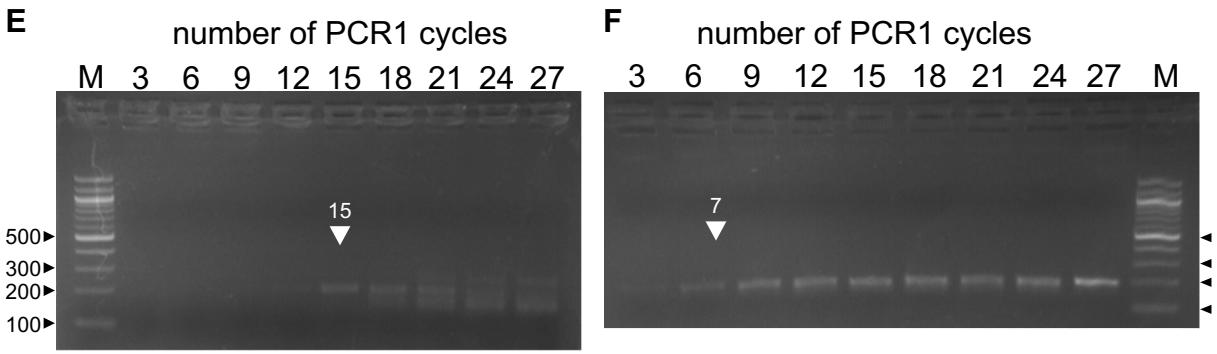


**Figure S1. Selection pressure during the rounds of the in vitro selection experiment. (A)** Number of the round of in vitro selection. **(B)** Incubation time of the RNA library with cTmp, $MgCl_2$, Tris/HCl pH 8.3. **(C)** Concentration of cTmp during its incubation with the RNA library. **(D)** Number of PCR1 cycles that were necessary to obtain a clear, but sub-saturated signal during the PCR step following reverse transcription of the selected RNA sequences. This number was used as an estimate for activity in the RNA library. Note that the incubation time, and the concentration of cTmp were reduced in round 6, after an increase in pool activity was suggested by the decreased number of PCR1 cycles in selection round 5. **(E)** 2% Agarose 1x TAE gel image of PCR1 products after selection round 1. The first lane with a size marker is labeled as 'M', with the size in base pairs for relevant bands given on the left. The other lanes are labeled with the number of PCR1 cycles. A white arrowhead indicates the chosen number of PCR1 cycles (see D). The PCR artifact that appeared starting PCR1 cycle 18 was not detectable in later selection rounds. **(F)** As (E) but for the PCR1 of selection round 5. Since the band at PCR1 cycle 6 was slightly weaker than the threshold used for other bands, the number 7 was used in (D). This semi-quantitative approach was sufficient for the purpose of monitoring the progress of the selection, as later confirmed by HTS analysis.

## Figure S2

### A

**C2**
GAGACCGACATGTCTGAATAACCTGCCAATTACAATTCCAGAGGCACCATTCAGTATATATGCGGACGGACCCCCAACAAGGTCCATGCTGCCTTCA
AGTAGCATTGTTGGATTTTGACACCCACGGTTACGTCGCCCGACGACAGGTCATAACCGTTGTCATTTTATACACGTAATCGCA

**C4**
GAGACCGACATGTCTACCAATGATGGGGATCTCCGAGGACAATGGGAGAGCGGCGAAAAGAGACGCTTTCCCCTGATCTAAGATGACATGCGTGTAA
AGTGATGGGAATCCCTGCTCGTAGCTAGAGTTAGCCGTAATGATAGAGCCATTATATTAAAGTAGTATACACGTAATCGCA

**C71**
GAGACCGACATGTCTTGGATGTGATCGCTATTGATTTACACAAAAATAAAAACTAAACCCCTGTTTATACGCGATACCACGCGAAAAGTTCAATTTT
ATCATGCGCGGTGCCGTATGAGCGGTACTGTACAGCCCGGTGAAACGATCTTGGCTACCACCTTTGAATATACACGTAATCGCA

**C20**
GAGACCGACATGTCTCCATGTTTGAAGGAAGCGGCCTAATCAAACAAAATACGGCACTTGTTTACCCGCATGGCGCCAACCTCTTAGATGTCGAGTC
ATCTTGCATTCCGCTGAAAGTTAAGGCCCCTAAGGGCGTTGACAATTTAACCATCAGGAGCGGTATATATACACGTAATCGCA

### B

**CLUSTAL O(1.2.4) multiple sequence alignment**

```
C2  GAGACCGACATGTCTGAATAACCTGCCAATTACAATTCCAGAGGC-------ACCATTCAGTATATATGCGGACG--GACCCCCAACAA
C20 GAGACCGACATGTCTCCATGTTTGAAGGAAGCGGCCTAATCAAAC-------AAAATACGGCACTTGTTTACCCGCATGGCGCCAACCT
C4  GAGACCGACATGTCTACCAATGATGGGGATCTCCGAGGACAATGGGAGAGCGGCGAAAAGAGACGCTTTCCCCTGATCTAAGATG---A
C71 GAGACCGACATGTCTTGGATGTGATCGCTATT-GATTTACACAAAAATAAAAACTAAACC---CCTGTTTATACGCGATACCACG---C
    ***************                                     *           *      *

C2  GGTCCATGCTGCCTTCAAGTAGCATTGTTGGATTTTGACACCCA---CGGTTACGTCGCCCGACGACAGGTCATAACCGTTGTCATTTTAT
C20 CTTAGATGTCGAGTCAT-CTTGC-----ATTCCGCTGAAAGTTAAGGCCCCTAAGGGCGTTGACAATTTAACCATCAGGAGCGGTATATAT
C4  CATGCGTGTAAAGTGAT-GGGAA-----TCCCTGCTCGTAGCTA---GAGTT---AGCCGTAATGATAGAGCCATTATATTAAAGTAGTAT
C71 GAAAAGTTCAATTTTAT-CATGC-GCGGTGCCGTATGAGCGGTA---CTGTACAGCCCGGTGAAACGATCTTGGCTACCACCTTTGAATAT
          *      *                     *       *                 *                          ***

C2  ACACGTAATCGCA         181
C20 ACACGTAATCGCA         180
C4  ACACGTAATCGCA         178
C71 ACACGTAATCGCA         181
    *************
```

**Figure S2: (A)** Four sequences that showed benefit of peptide *P4 as judged by the HTS data in figure 2B. Each sequence is the peak sequence of its corresponding cluster. **(B)** Alignment of these four sequences using Clustal Omega (https://www.ebi.ac.uk/Tools/services/web/). The 15 well-aligned sequences at the 5'-terminus and the 16 well-aligned sequences at the 3'-terminus are the conserved sequences at the ends of the previously randomized region of the library. Between these conserved sequences, no significant homology was found.

**Figure S3**

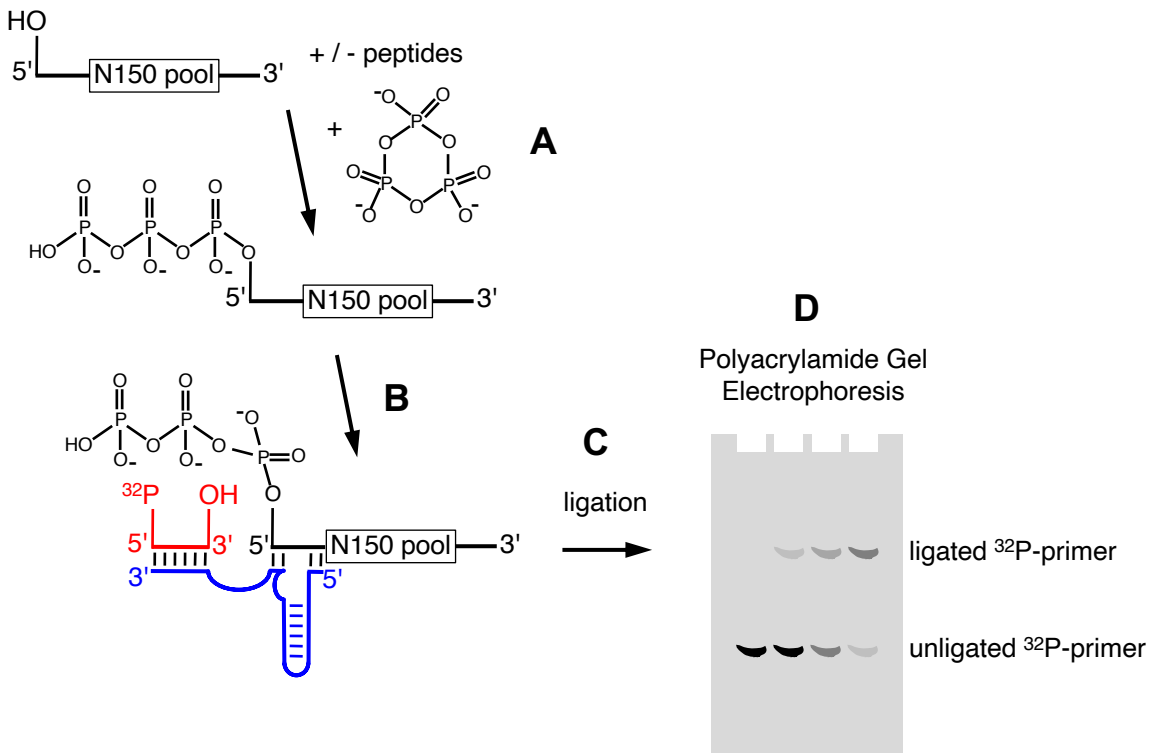


**Figure S3. Schematic for the biochemical assay used to measure the activity of selected ribozymes.** (A) Ribozymes were prepared with a 5'-hydroxyl group as for the selection procedure, and incubated in the presence of cTmp and peptides under the given conditions. (B) The reaction products were heat renatured with a ligase ribozyme and a radiolabeled 16-mer oligonucleotide. This oligonucleotide had the same sequence as the biotinylated oligonucleotide during the selection procedure but instead of the biotin group it was radiolabeled with $^{32}$P-phosphate. (C) After incubation of the ligase ribozyme / 16mer / triphosphorylation ribozyme complex, the ligase ribozyme was separated by a fully complementary DNA oligonucleotide, and the reaction products were loaded on a denaturing polyacrylamide gel. (D) Polyacrylamide gel electrophoreses separated the products into two bands, the unligated primer (16-mer) and the ribozyme / oligonucleotide ligation product (>100 nt). The gels were exposed to phosphorimager plates for quantitation.

**Figure S4**

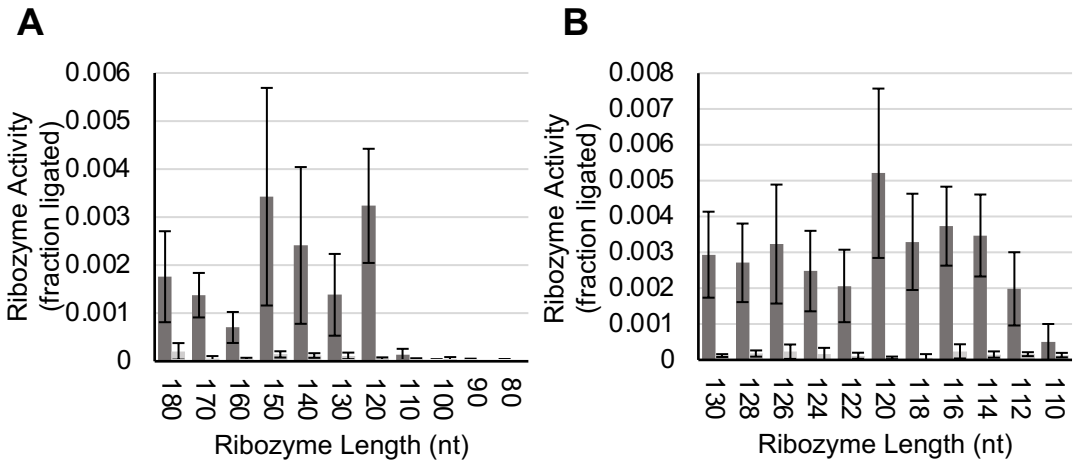


**Figure S4. Truncation of ribozymes 3'-termini from the cluster 20 ribozyme to the shortest length with maximal activity. (A)** In the first step, the ribozyme was truncated in ten-nucleotide steps, where activity was equally high with a length of 150 and 120 nucleotides. **(B)** In the second step, the ribozyme was truncated in two-nucleotide steps between 110 and 130 nucleotides. The length of 120 nucleotides was chosen as the minimal length with full activity. Columns in dark grey represent reactions in the presence of the ten peptides (each at 0.5 mM), columns in light grey represent reactions without peptides. Error bars represent standard deviations from at least three experiments.

**Figure S5**

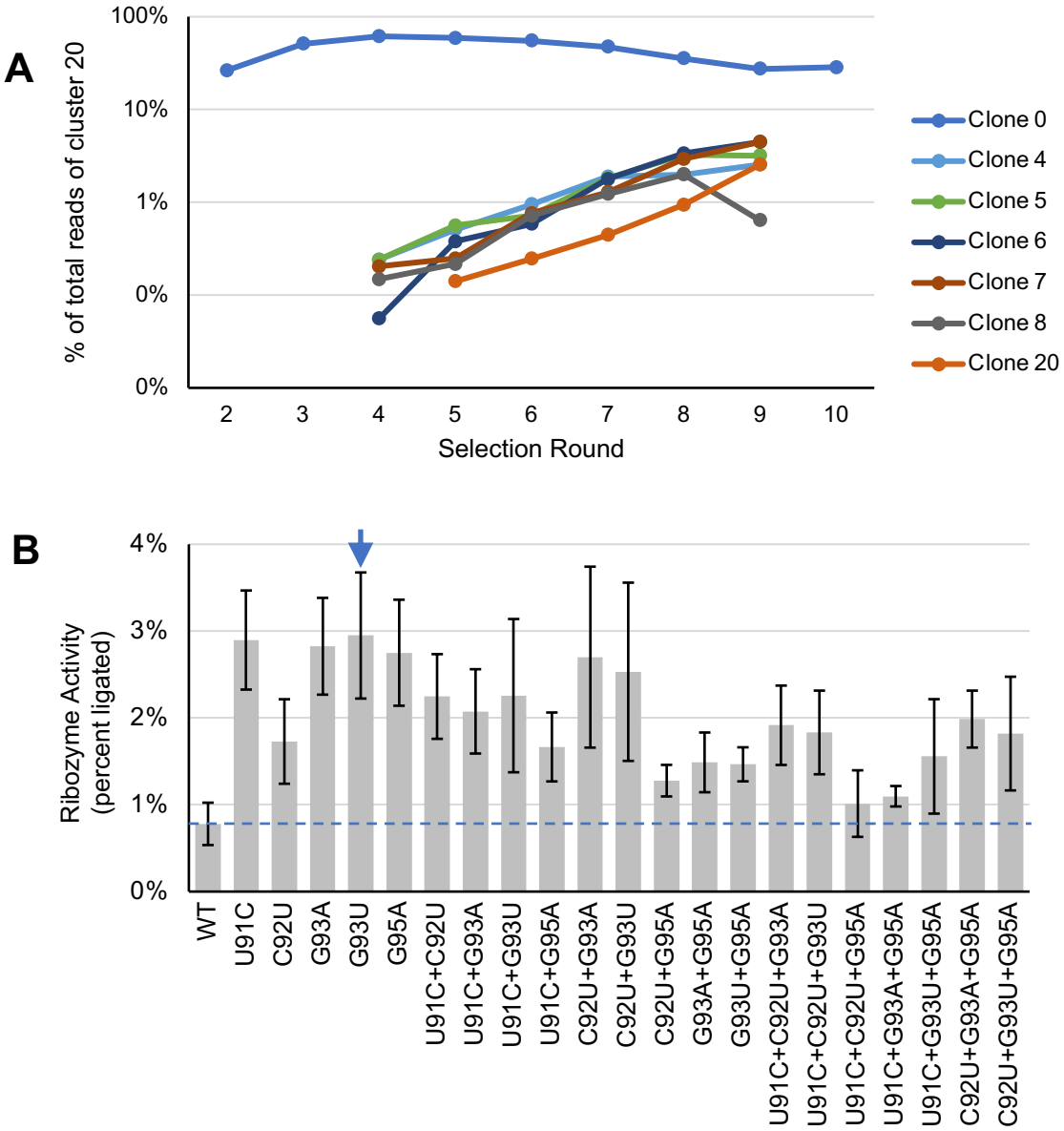


**Figure S5. Identification, and biochemical analysis of five beneficial mutations in ribozyme cluster 20. (A)** Frequency of five mutations as a function of the selection rounds.The peak sequence of the cluster (c0) is shown as comparison (blue) Since an increase in frequency is associated with high ribozyme activity, the five shown mutations were analyzed biochemically. Interestingly, all of these mutations were close to each other in the ribozyme sequence, at positions 91 to 95 (U91C, C92U, G93A, G93U, G95A). **(B)** Biochemical analysis of the five mutations. The activity of each ribozyme under selection conditions, with all ten peptides each at 0.5 mM concentration, is shown as a function of the ribozyme's mutations. The activity level of the unmutated ribozyme (WT) is shown as a dashed blue line. The ribozyme variant G93U showed 3.8-fold higher activity than the WT, and was chosen for future experiments (blue arrow). Error bars are standard deviations from triplicate experiments.

**Figure S6**

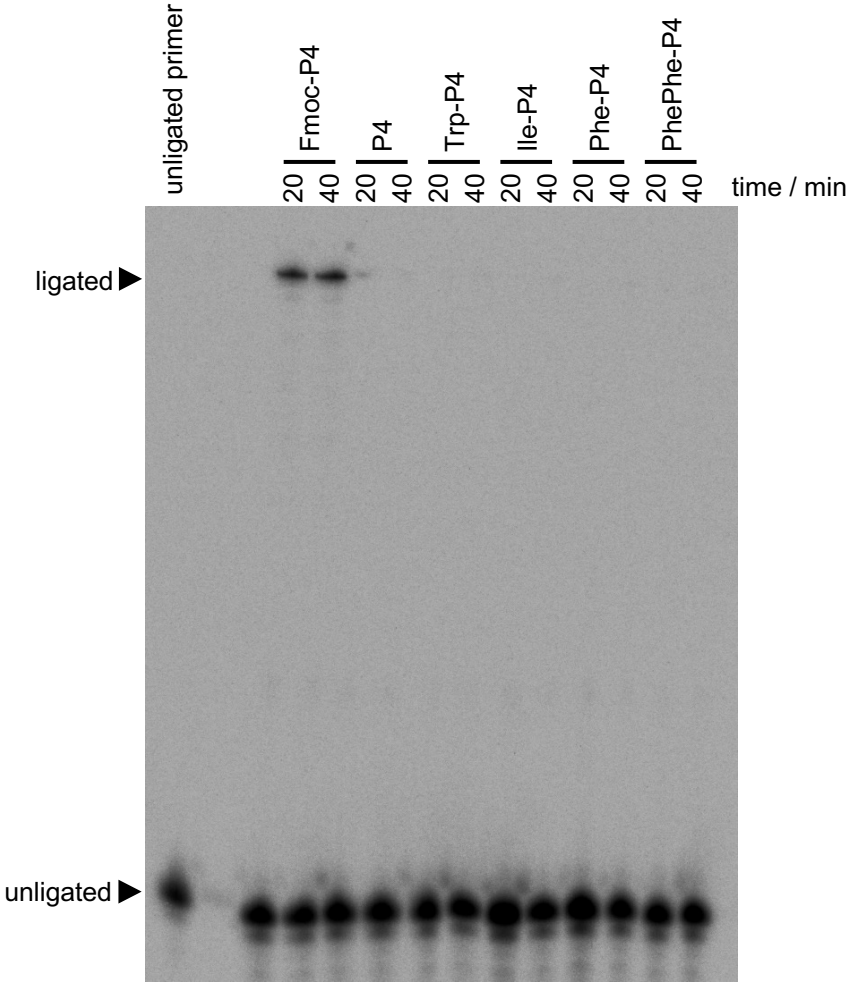


**Figure S6. Test of ribozyme 20 benefit by peptide 4 variants with hydrophobic amino acids added to the N-terminus.** Shown is a phosphorimager scan of a denaturing polyacrylamide gel that was used to separate the reaction products of the ligase ribozyme assay that measures self-triphosphorylation activity as shown in figure S2. The positions of unreacted and reacted oligonucleotides are indicated on the left. Six different reactions were analyzed, with six different variants of peptide 4, and for each variant, two self-triphosphorylation times (20 minutes and 40 minutes, as indicated) were analyzed. The first two variants are controls, with Fmoc-P4 (the Fmoc protection group at the peptide 4 N-terminus), and P4 (peptide 4 without N-terminal modification) confirming the results in figure S7. The remaining four peptide variants contained N-terminal extensions by one or two amino acids, such that the sequences were WAAEAAAKA (Trp-P4), IAAEAAAKA (Ile-P4), FAAEAAAKA (Phe-P4), and FFAAEAAAKA (PhePhe-P4).