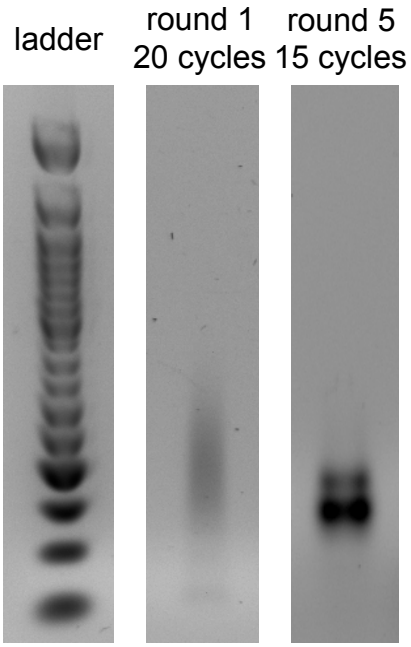
Electrophilic Activity-Based RNA Probes Reveal a Self-Alkylating RNA for RNA Labeling

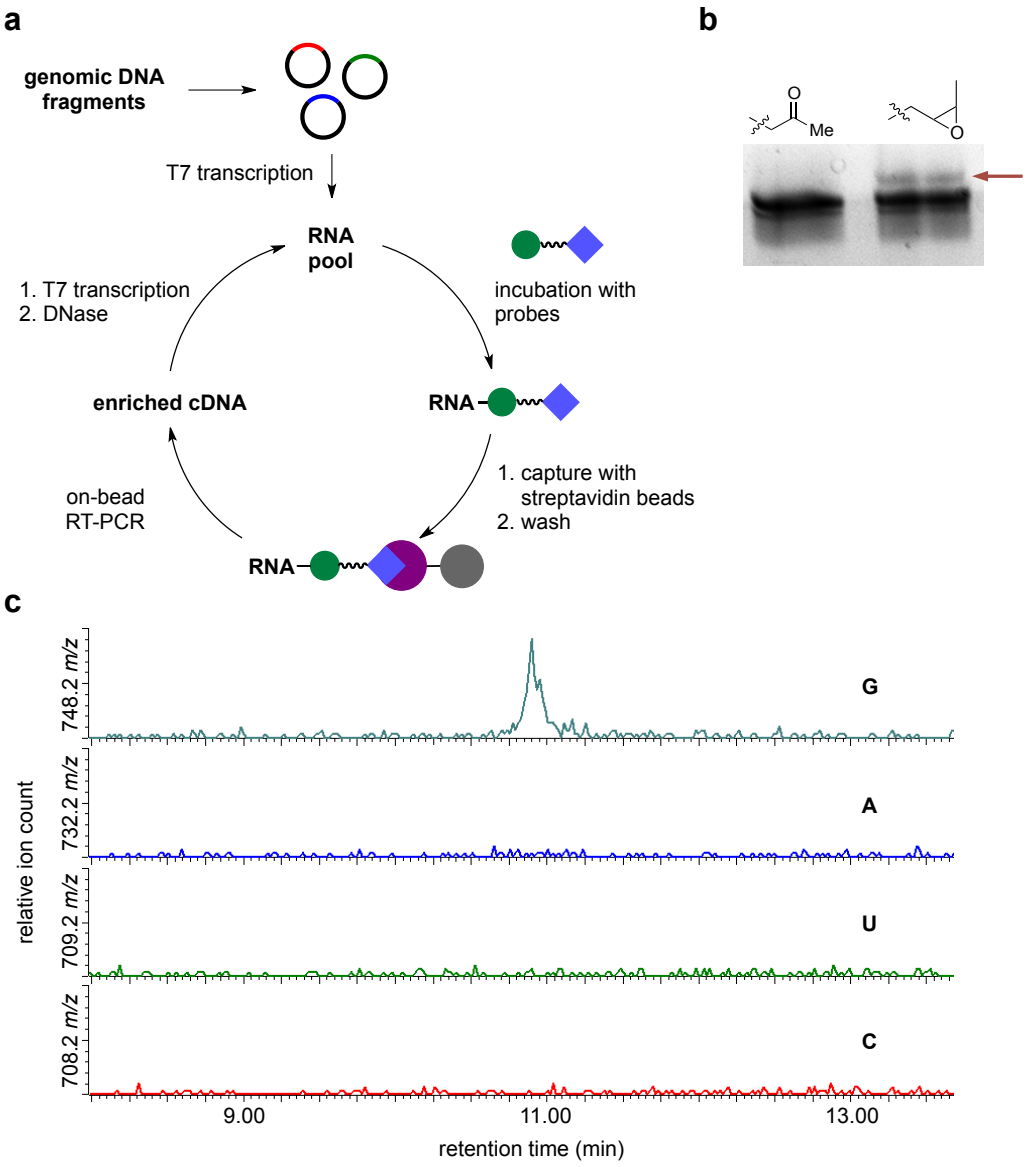
Richard I. McDonald, John P. Guilinger, Shankar Mukherji, Edward A. Curtis, Won I. Lee, David R. Liu

SUPPLEMENTARY INFORMATION

Supplementary Results



Supplementary Figure 1. Agarose gel electrophoresis of cDNA following 1 and 5 rounds of selection. The Round 1 material was amplified for 20 cycles and Round 5 was amplified for 15 cycles. The smallest band in the molecular weight ladder corresponds to 50 bp, and each higher band differs by 50 bp.



Supplementary Figure 2. *In vitro* selection leads to the identification of genome-encoded RNA species that react with epoxide 1. (**a**) *In vitro* selection cycle of genome-encoded RNAs capable of reacting with electrophilic, biotinylated probes. Electrophilic groups are represented as green circles, biotin groups are shown as blue diamonds, and streptavidin is represented in purple. (**b**) The post-round 6 RNA pool was incubated with each of the electrophilic probes and a streptavidin gel mobility shift assay was performed to identify the reactive electrophile(s). Lane 1 shows the RNA band following incubation with an unreactive ketone substrate. Lane 2 reveals that RNA species (arrow) within the RNA pool react with disubstituted epoxide probe **1**. The complete gel is shown in Supplementary Figure 11. (**c**) Following incubation with the epoxide probe, the RNA pool was digested into mononucleotides by nuclease P1 and analyzed by LC/MS. Guanosine nucleotides, but not A, U, or C nucleotides, modified with epoxide **1** were observed.

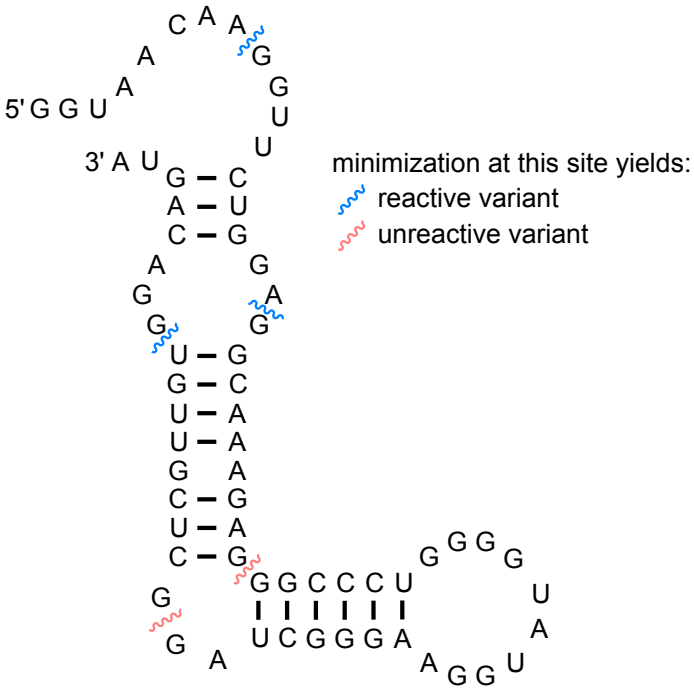
a

TAA CAA GGT TCT **a**GA GGC AAA GAG GGC CCT **a- -** GTA TG**a** AAG GGC TAG GCT C**a**T TGT GGA CAG **c**A

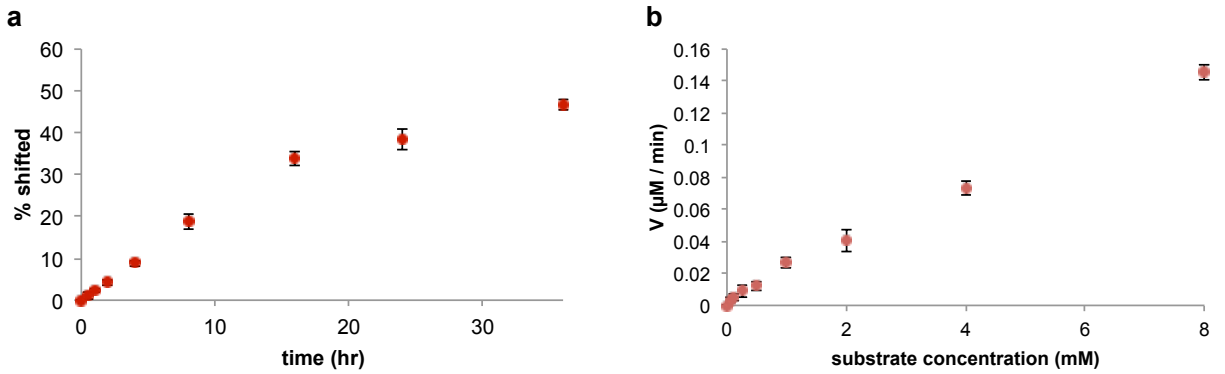
b

TAA CAA GGT TCT **g**GA GGC AAA GAG GGC CCT **ggg** GTA TG**g** AAG GGC TAG GCT C**g**T TGT GGA CAG **t**A

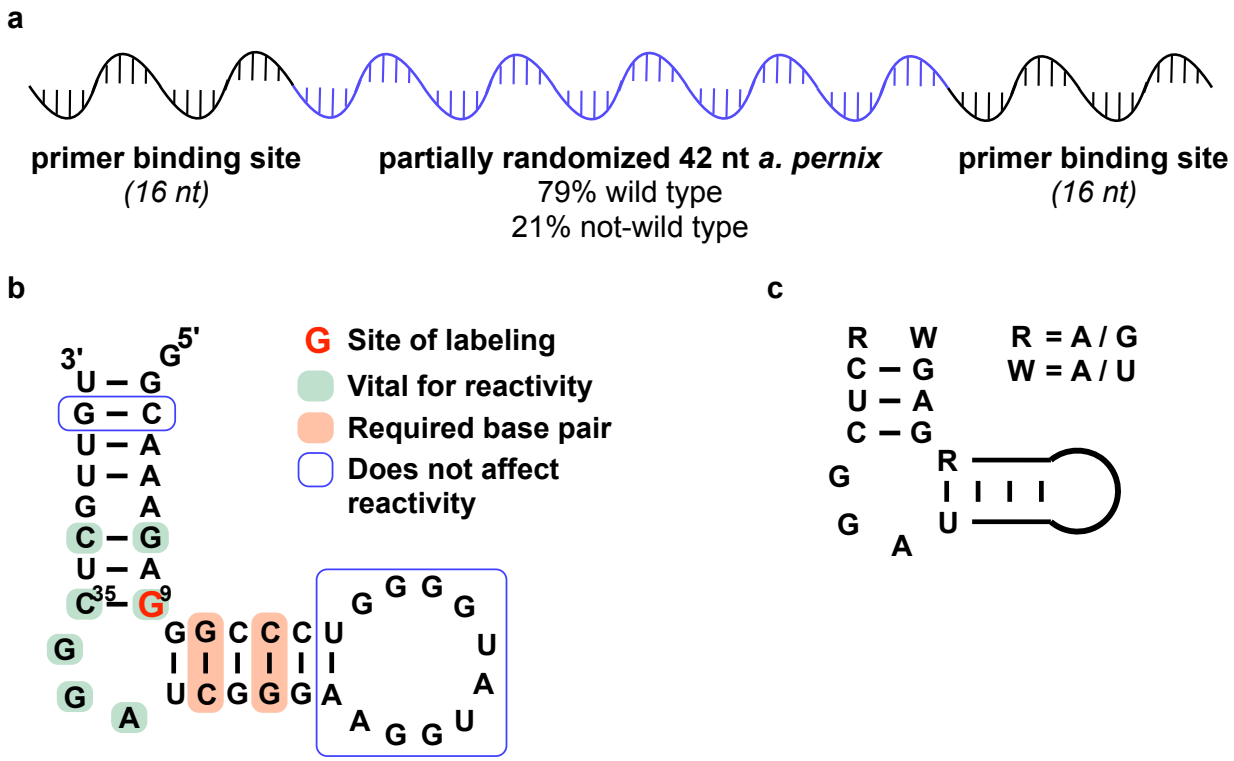
Supplementary Figure 3. *A. pernix*-based epoxide-opening catalytic RNA sequences. (a) The *A. pernix* sequence emerging from round 6 of the selection that was tested for reactivity with epoxide probe **1**. (b) The corresponding *A. pernix* reference genome sequence.



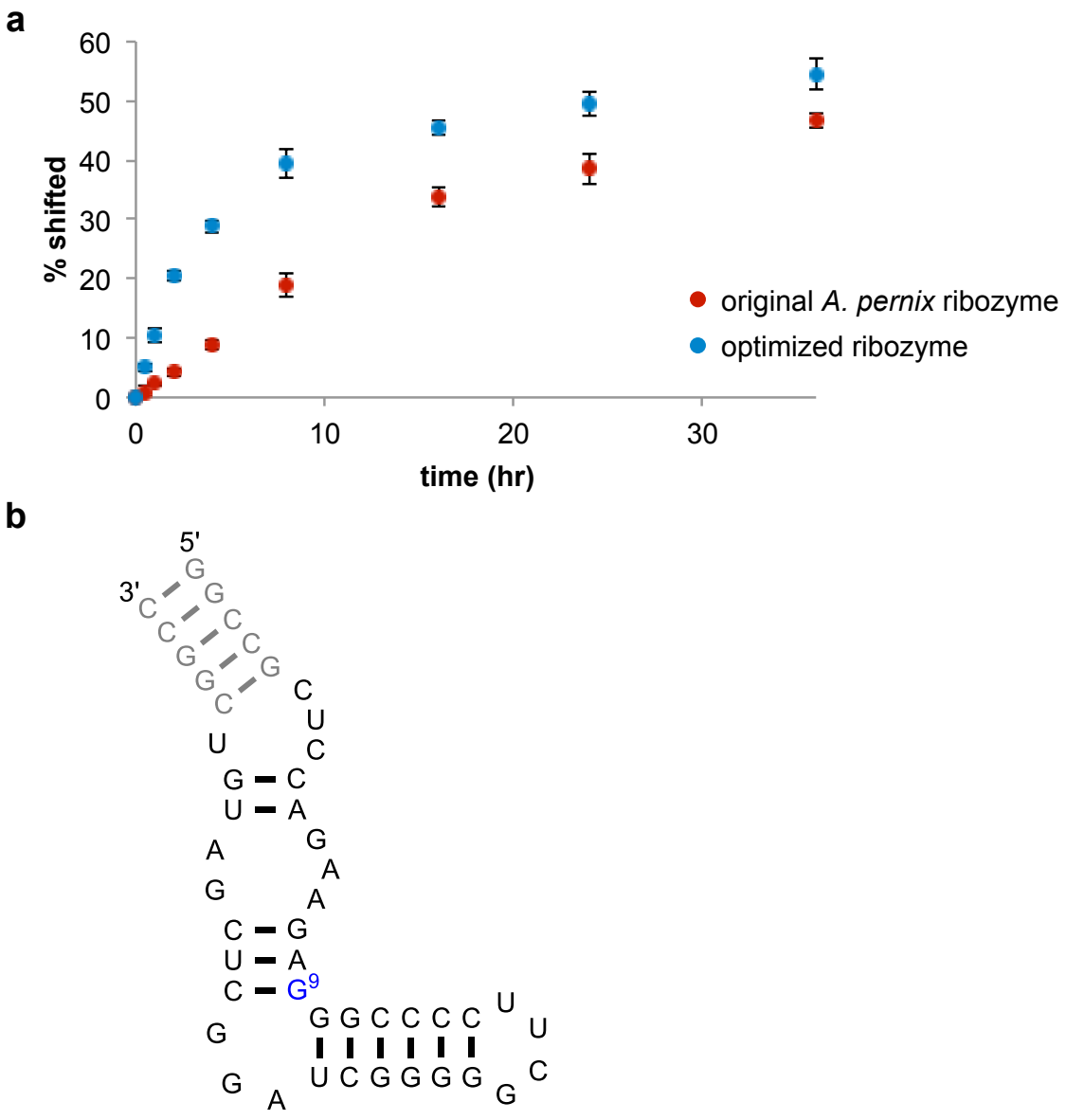
Supplementary Figure 4. Minimization of the *A. pernix* ribozyme. Progressive truncations from the 5'- and 3'-termini revealed a 42-nt transcript that effectively catalyzes epoxide ring opening.



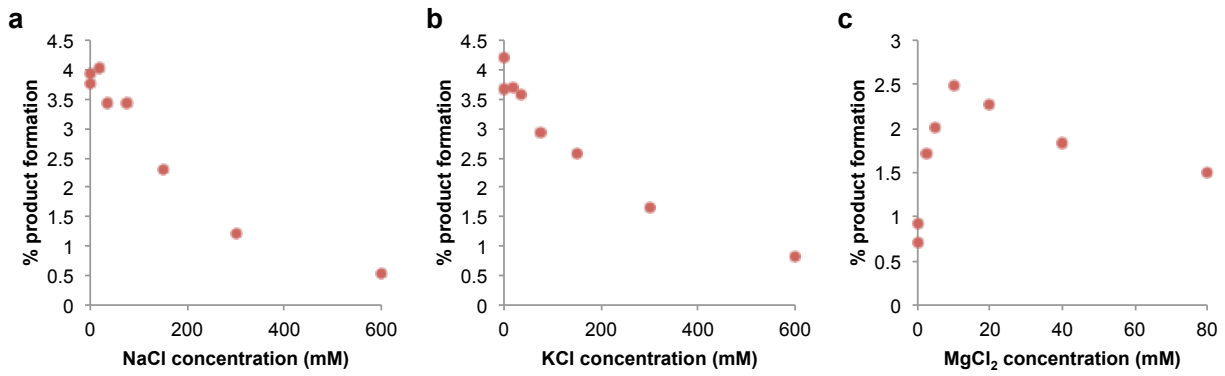
Supplementary Figure 5. Kinetic characterization of the minimized *A. pernix* ribozyme. (a) Product formation for the reaction between the ribozyme (1 μM) and biotin-epoxide 1 (1.3 mM). (b) Michaelis-Menten curves to determine k_{cat} and K_m (5 hr reaction time, 1 μM RNA). Formation of product was calculated as the mean of three independent experiments from streptavidin gel mobility shift assays. Michaelis-Menten parameter calculations were corrected to account for 47% of the RNA folding into a reactive conformation. Data represents mean values ± s.d. for three replicates.



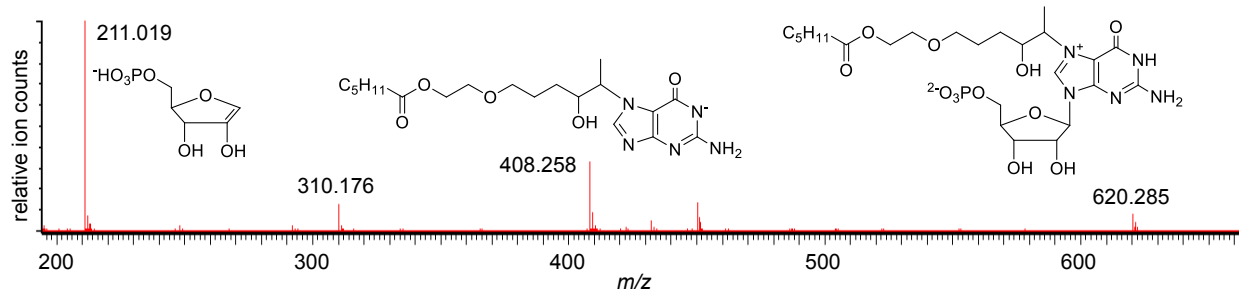
Supplementary Figure 6. Sequence requirements of the *A. pernix* ribozyme. (a) The partially randomized *A. pernix* library used for reselection. (b) Structure-activity relationships inferred from site-directed mutation studies. (c) Based on the results from reselection of the partially randomized *A. pernix* library and site-directed mutation studies, we developed a reactive minimal motif for bioinformatic searching. The motif requires three unspecified base pairs followed by a loop ranging from 4-25-nt in length.



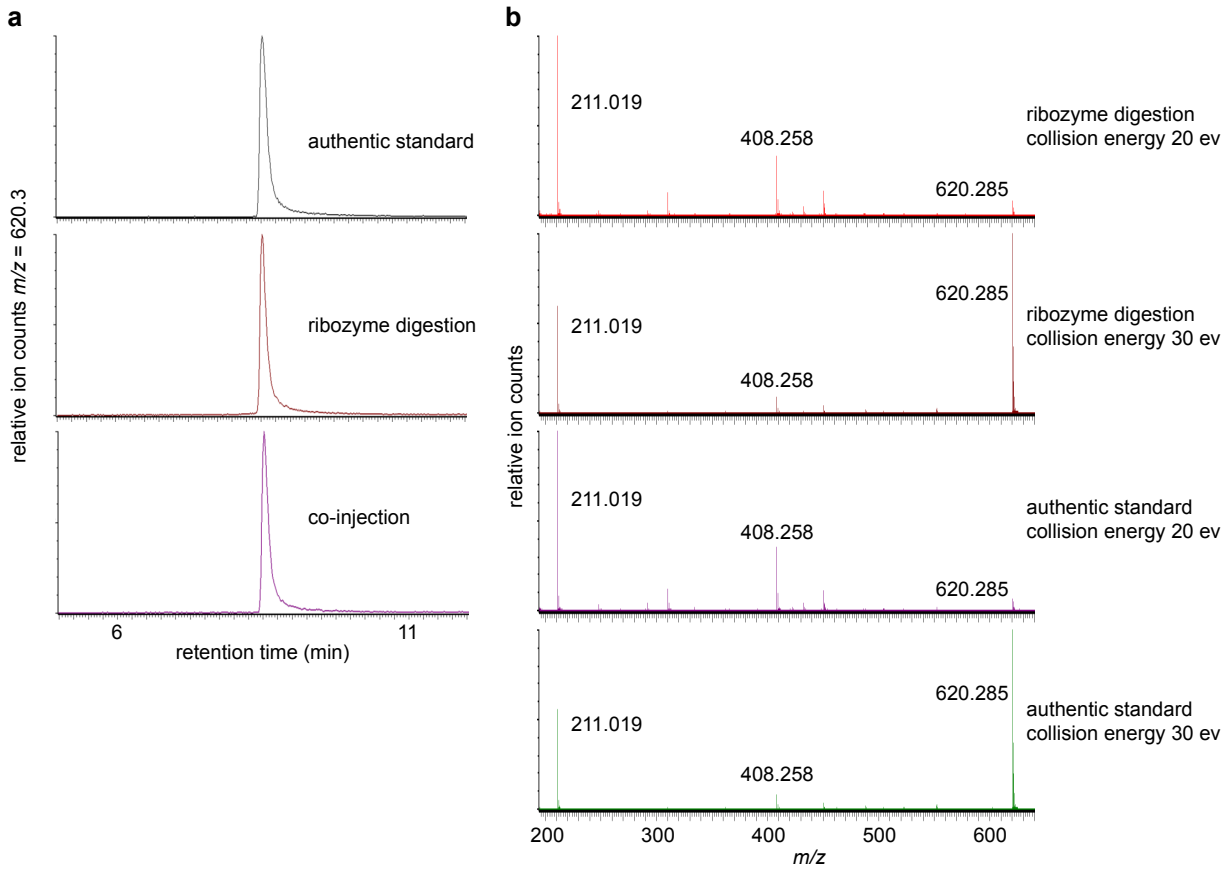
Supplementary Figure 7. Sequence, structure, and reactivity of the optimized ribozyme. (a) Product formation is shown for the reaction between the ribozyme (1 μM) and biotin-epoxide 1 (1.3 mM). Data represents mean values ± s.d. for three replicates. (b) Sequence and secondary structure model of the optimized epoxide-opening catalytic RNA. The grey-colored stem nucleotides are derived from the tRNA scaffold.



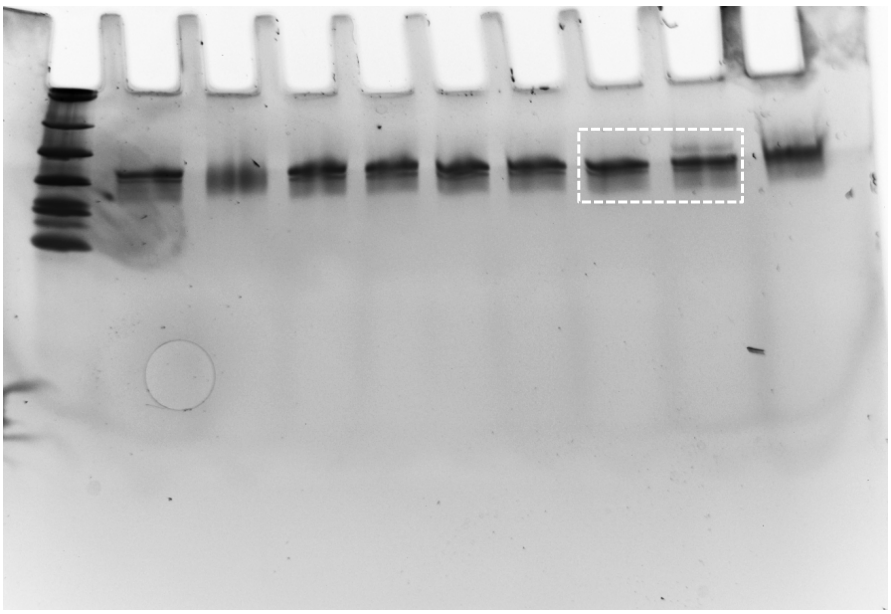
Supplementary Figure 8. The effect of monovalent and divalent salts on reaction efficiency. Ribozyme-epoxide reactivity as a function of (a) NaCl concentration, (b) KCl concentration, and (c) $MgCl_2$ concentration.



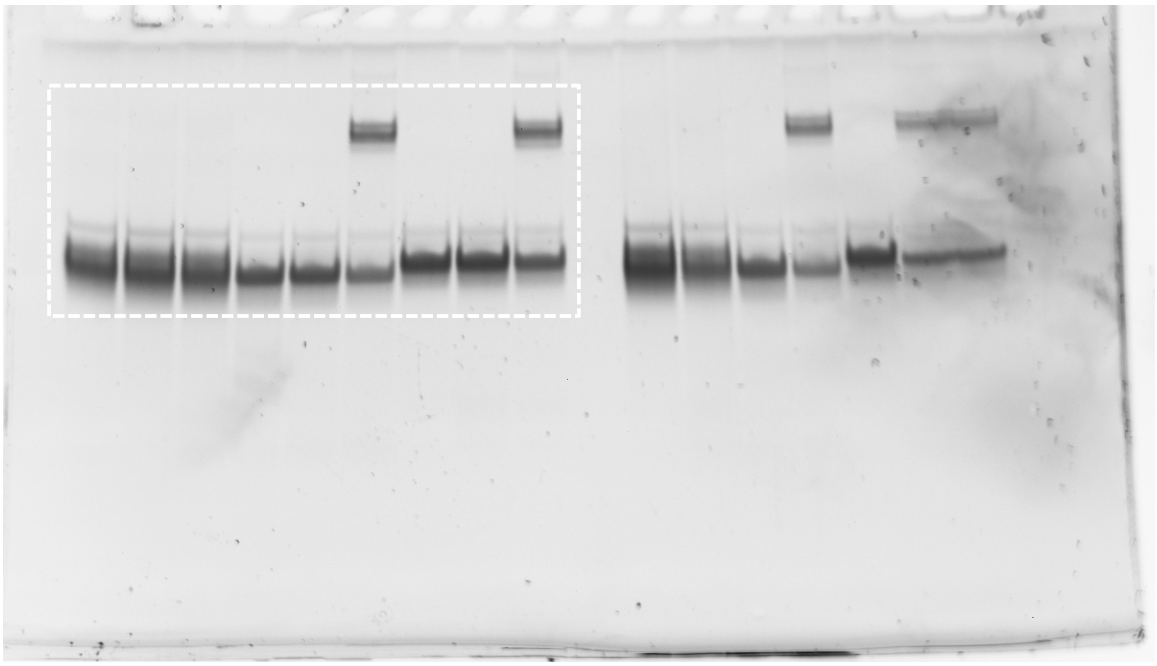
Supplementary Figure 9. MS characterization of the epoxide-ribozyme reaction. (a) Negative-ion mode MS/MS of the epoxide-GMP product from nuclease P1 digestion of the ribozyme (*m/z* = 620.285) reveals ions corresponding to epoxide-purine (*m/z* = 408.258) and the unmodified ribose (*m/z* = 211.019). An additional fragment resulting from hydrolysis of the ester group in the probe (*m/z* = 310.176) is also observed. The proposed structures show bond formation between N7 of guanosine and the less-hindered carbon of the epoxide, although the data are consistent with either or both epoxide carbon atoms undergoing bond formation.



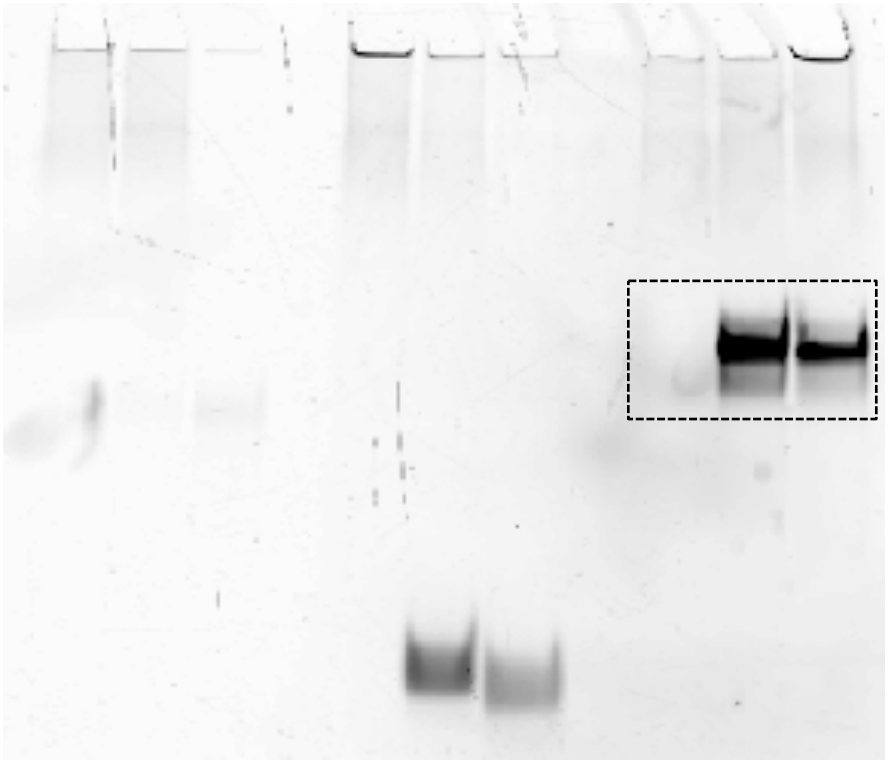
Supplementary Figure 10. Structural elucidation of the epoxide-guanosine product. (a) LC comparison of the GMP-epoxide authentic product and digested ribozyme product. (b) MS/MS fragmentation comparison of the authentic and digested ribozyme products.



Supplementary Figure 11. Complete gel from Supplementary Figure 2b. The dashed box highlights the portion of the gel shown in the figure.

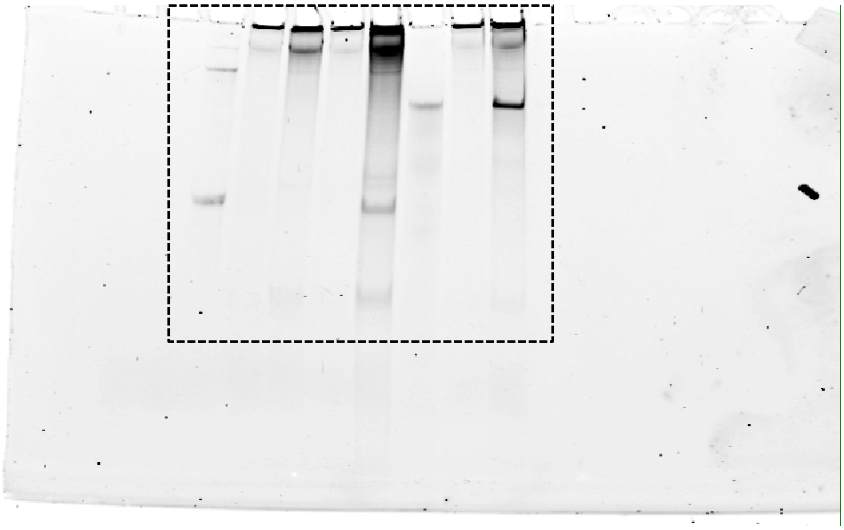


Supplementary Figure 12. Complete gel from Figure 2a. The dashed box highlights the portion of the gel shown in the figure.

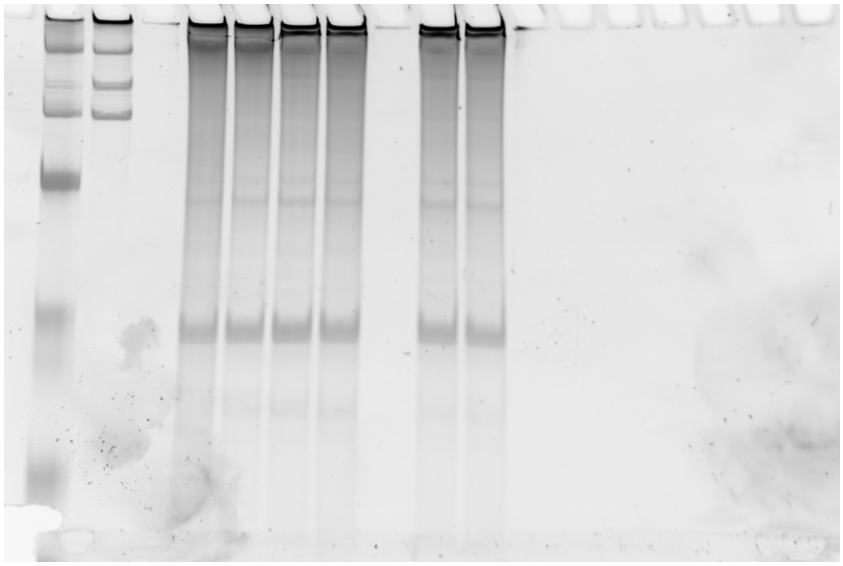


Supplementary Figure 13. Complete gel from Figure 3d. The dashed box highlights the portion of the gel shown in the figure.

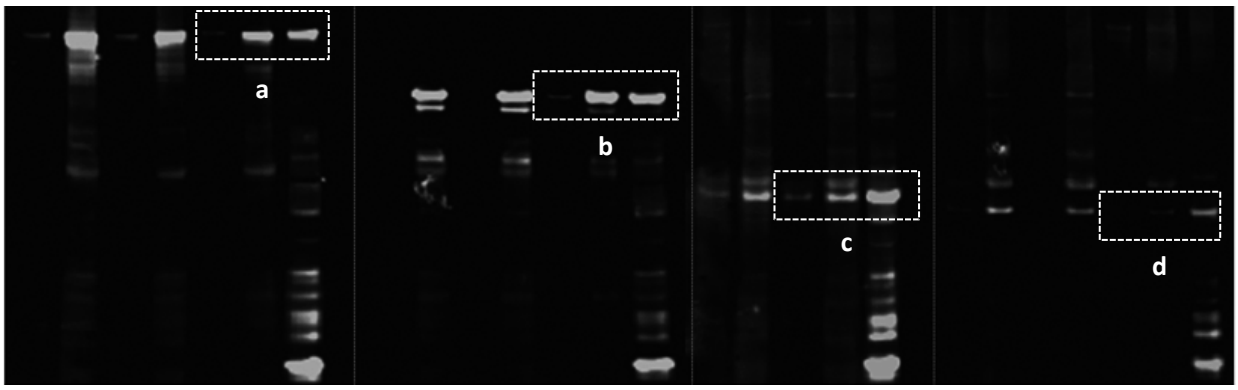
a



b



Supplementary Figure 14. Complete gel from Figure 4b. (a) The dashed box highlights the portion of the gel shown in the figure. (b) Following fluorescence imaging, the gel was stained with the general RNA stain SYBR Green II and imaged for total RNA content.



Supplementary Figure 15. Complete Western blot from Figure 4c. The portions of the Western blot corresponding to (a) Puf6, (b) Khd1, (c) She2, and (d) Gud1 portions of the figure are shown in the dashed boxes.

	5S rRNA	HPRT1	tubulin
inactive ribozyme	7.49	5.38	5.00
one copy of ribozyme	-0.06	4.66	4.56
three copies of ribozyme	-1.81	5.23	4.71

Supplementary Table 1. Quantification of RNA enrichment from total RNA. ΔC_T values for RNA enrichment of the 5S rRNA, HPRT1 control, and tubulin control, as determined by RT-qPCR. The ΔC_T values represent the difference between the experimental sample and a control sample lacking streptavidin-linked bead capture. See Online Methods and Supplementary Note for additional details.

Mutation	% Shifted
No changes	35
C^{3}-G	29
G^{42}-C	24
C^{3}-G^{42}	27
G^{7}-C	0
C^{38}-G	0
G^{7}-C^{38}	0
G^{9}-C	0
C^{35}-G	0
G^{9}-C^{35}	0
G^{10}-A	33
G^{10}-C	0
G^{10}-U	0
U^{31}-C	29
G^{11}-C	0
C^{30}-G	0
G^{11}-C^{30}	33
C^{13}-G	0
G^{28}-C	0
C^{13}-G^{28}	38
A^{32}-C	0
A^{32}-G	0
A^{32}-U	0
G^{33}-C	0
G^{33}-A	0
G^{33}-U	0
G^{34}-C	13
G^{34}-A	14
G^{34}-U	15
randomized 10 bp loop	34

Supplementary Table 2. Activity of A. pernix site-directed mutant RNAs. Percentage of shifted RNA was determined by streptavidin gel mobility shift assay with biotin-epoxide **1**.

Supplementary Note

In Vitro Selection of Nucleophilic, Genome-Encoded RNAs

The eight biotinylated probes were combined into one cocktail and incubated with a pool of genome-derived RNA fragments from nine organisms spanning all three kingdoms of life (*Arabidopsis thaliana, Aeropryum pernix, Haloarcula marismortui, Methanococcus jannaschii, Bacillus subtilis, Bacteroides fragilis, Escherichia coli, Gallus gallus domesticus, and Homo sapiens*). We recently discovered two families of naturally occurring GTP-binding RNAs by performing a selection on these RNA pools.[1,2] The libraries were constructed by random DNase I-catalyzed fragmentation of genomic DNA and isolation of fragments between 100 and 600 bp by gel electrophoresis. These DNA fragments were transcribed by T7 RNA polymerase to yield genome-derived RNA fragments flanked by primer binding sites suitable for *in vitro* selection. Genome-encoded RNA pools constructed in this manner, as opposed to RNA isolated from cells, avoid the underrepresentation of many RNA species due to large expression-level differences and, in the case of RNA from multicellular organisms, due to tissue-specific transcription.

The mixture of *in vitro*-transcribed genome-encoded RNA fragments from all nine organisms together with the eight biotinylated probes was incubated in buffer chosen to approximate physiological conditions (25 mM Na-HEPES, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$). Biotinylated RNA fragments were isolated using streptavidin-coated magnetic beads, then subjected to on-bead reverse transcription followed by PCR amplification (Supplementary Figure 2a). Following five rounds of selection, amplification, and transcription, we observed enrichment of a substantial portion of the RNA pool. Agarose gel electrophoresis of the corresponding cDNA revealed discrete bands consistent with convergence to discrete RNA species (Supplementary Figure 1). Collectively, these results suggest that unusually reactive RNA species exist within diverse genome-encoded RNA pools and can be isolated using small-molecule probes of tuned electrophilicity.

Sequence Requirements of the A. Pernix Catalytic RNA

Computational secondary structure prediction[3] suggests that the *A. pernix* catalytic RNA folds into a stem-bulge-stem-loop structure (Figure 2b). In order to probe this structural model and to identify the minimal sequence requirements for reactivity, we generated a partially randomized RNA pool derived from the minimized 42-nt *A. pernix* RNA and performed a reselection for epoxide reactivity (Supplementary Figure 6a). The RNA was incubated with the epoxide probe for 4 h each round and, following four rounds of selection, converged on a subset

of reactive constructs (see Online Methods). The enriched RNA pool was reverse transcribed and analyzed by high-throughput sequencing, resulting in the sequence logo shown in Figure 2c.[4] This analysis supported the predicted stem-bulge-stem model, revealed a highly conserved bulge region, and suggested that the 10-nt loop could be mutated without substantial loss of activity.

The resulting structural model was further probed by site-directed mutagenesis. Nucleotides predicted to form a base pair were mutated to determine whether catalytic RNA activity tolerated swapping of the presumed base-pair nucleotides. Constructs featuring compensatory mutations at G^{11}–C^{30} and C^{13}–G^{28} retained activity, whereas any mutation at G^{7}–C^{37} or G^{9}–C^{35} resulted in the complete loss of activity (Figure 2b). The requirement for the presence of the predicted 10-nt loop was investigated by testing the reactivity of an RNA pool featuring a randomized loop, which exhibited reactivity with the epoxide comparable to that of the canonical catalytic RNA. We also probed the predicted bulge region that was highly conserved during reselection by testing the reactivity of all four possible nucleotides at these positions. Mutation of G^{10} to A did not significantly affect RNA reactivity, while C and U were not tolerated. All mutations at A^{32} and G^{33} resulted in loss of RNA reactivity, while mutations at G^{34} significantly decreased activity. Collectively, the activities of these mutant RNAs are consistent with the reselection results and support the structural model in Figure 2b.

Metal Ion Requirements of the A. pernix Catalytic RNA

A common characteristic of functional RNAs is a dependence on metal cations.[5] The current selection was performed in the presence of Mg^{2+} and Na^{+}. RNA activity was not substantially affected by concentrations of Na^{+} or K^{+} up to 75 mM (Supplementary Figure 8). In contrast, RNA activity was dependent on Mg^{2+}, with optimal activity at 10 mM Mg^{2+} (Supplementary Figure 8). Testing other divalent cations instead of Mg^{2+} revealed that the RNA retains activity in the presence of Ca^{2+}, Mn^{2+}, or Co^{2+}, albeit at slightly reduced rates, but not in the presence of Ni^{2+} or Zn^{2+}.

Characterization of Alkylation Regiochemistry

LC/MS analysis revealed that RNA modification by **1** occurs selectively on guanosine (see Online Methods and Supplementary Figure 2c). To identify which guanosine was modified, the wild-type catalytic RNA and a panel of single-base mutants were incubated with the epoxide probe and then partially digested with RNase T1, an endonuclease that cleaves 3' of guanosine residues. The resulting nucleotide fragments were characterized by LC/MS. This approach

revealed G^9 as the site of alkylation by **1** (see Online Methods). LC/MS/MS fragmentation of the ion corresponding to the epoxide modified guanosine mononucleotide resulted in fragments consistent with glycosidic bond fragmentation to yield ribose ($[M\text{-}H]^-$ *m/z* = 211.019) and a species corresponding to the product of a reaction between the epoxide and a guanine base ($[M\text{-}H]^-$ *m/z* = 408.258) (Supplementary Figure 9).

To elucidate the structure of the guanine epoxide adduct, we synthesized an authentic chemical standard by heating the epoxide probe and 5'-monophosphate guanosine in acetic acid and determined the structure of the major reaction product by NMR spectroscopy to be the adduct arising from bond formation between N7 of guanosine and the epoxide probe (Supplementary Figure 9 and Online Methods). Comparison of this authentic synthetic product with the alkylated guanosine generated by P1 nuclease digestion of the catalytic RNA after incubation with the epoxide probe revealed identical LC retention times and MS/MS fragmentation patterns at varying collision energies (Supplementary Figure 10). These results indicate that N7 is the site of guanosine modification, similar to previous reports of DNA reacting with activated epoxides[6,7,8] (see Online Methods for details).

Optimization of the Catalytic RNA

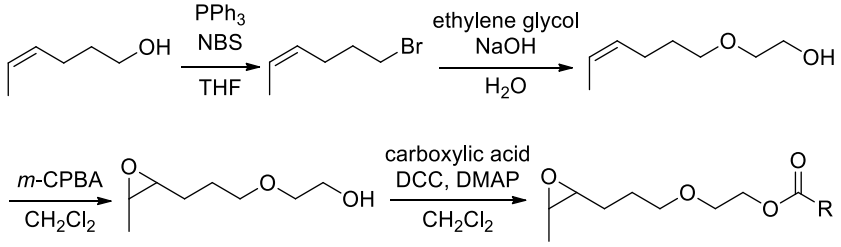
In order to improve the rate of the epoxide-RNA reaction to facilitate efficient covalent RNA modification, we used the partially randomized 42-nt *A. pernix* RNA library to perform high-stringency reselections with decreasing epoxide incubation times. After seven rounds of reselection, we characterized the resulting RNA pool by high-throughput sequencing. The ten most abundant library members were transcribed *in vitro* and their rates of epoxide self-labeling were assayed by gel mobility shift. Following selection of the most reactive candidates and additional minimization and engineering, we isolated a 41-nt variant containing four core mutations relative to the starting *A. pernix* catalytic RNA, a minimized 4-nt loop, and changes to the 5'- and 3'-ends (Supplementary Figure 7b). This optimized catalytic RNA exhibited a 5-fold faster rate of reaction with epoxide probe **1** (Supplementary Figure 7).

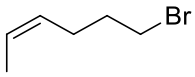
Quantification of RNA Enrichment from Total Cellular RNA

Following on-bead reverse transcription, we performed qPCR of the 5S rRNA and determined the difference in qPCR cycle threshold (ΔC_T) between the experimental sample and a control sample lacking streptavidin-linked bead capture for each of the three catalytic RNA fusion variants. To compare streptavidin capture of the 5S rRNA across the three samples, we calculated $\Delta\Delta C_T$ for each catalytic RNA sample compared to the inactive control, resulting in an

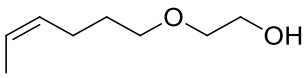
average $\Delta\Delta C_T$ of 6.96 for the 5S rRNA containing one copy of the catalytic RNA and an average $\Delta\Delta C_T$ of 9.08 for three copies of the catalytic RNA. These values correspond to a 125- and 541-fold enrichment, respectively, of the catalytic RNA-fused transcript over the inactive transcript.

Synthesis of Epoxide Probes





Triphenylphosphine (64 mmol; 1 equiv) was dissolved in dry THF (100 mL), followed by addition of cis-4-hexen-1-ol (64 mmol; 1 equiv). The solution was cooled in an ice bath, and then N-bromosuccinimide (67 mmol; 1.05 equiv) was added in portions. The solution was allowed to warm to room temperature and stirred overnight. After removing solvent at reduced pressure, the mixure was passed through a silica gel plug and flushed with hexanes (~500 mL). Hexanes was removed at reduced pressure to yield a colorless oil (7.72 g; 74% yield). Characterization data matched literature data.[9]

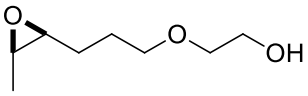


The alkyl bromide (28 mmol, 1 equiv) and ethylene glycol (224 mmol, 8 equiv) were combined in a round bottom flask. A solution of aqueous NaOH (112 mmol, 4 equiv in 6 mL H_2O) was then added slowly dropwise with vigorous stirring. The solution was submerged into a pre-heated 80 °C oil bath and stirred for 12 hours. After cooling to room temperature, the solution was diluted with H_2O and extracted three times with Et_2O. The combined organic layers were dried with $MgSO_4$ and the solvents removed at reduced pressure. The resulting oil was purified by silica gel chromatography (3:2 hexanes:EtOAc) to yield 2.18 g (54% yield) of a light-yellow oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ 1.59 (d, J=6.7 Hz, 3H), δ 1.64 (p, J=6.7 Hz, 2H), δ 2.10 (q, J=6.7 Hz, 2H), δ 2.24 (bs), δ 3.47 (t, J=6.7 Hz, 2H), δ 3.52 (t, J=4.7 Hz, 2H), δ 3.71 (t, J=4.7 Hz, 2H), δ 5.32-5.40 (m), δ 5.41-5.51 (m).

^{13}C NMR (100 MHz, $CDCl_3$) δ 129.75, δ 124.45, δ 71.78, δ 70.67, δ 61.81, δ 29.35, δ 23.32, δ 12.70.

HRMS: *m*/*z* (ESI) calculated $[M+H]^-$ = 143.1072, measured 143.1062 (Δ = 7.0 ppm)



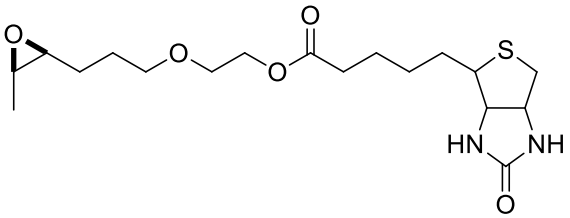
m-CPBA (4.6 mmol; 1.2 equiv) was dissolved in dry CH_2Cl_2 (8.5 mL) and cooled to 0 °C. A solution of the alkene (3.8 mmol; 1 equiv) in CH_2Cl_2 (1.5 mL) was added dropwise via syringe. The solution was removed from the ice bath and stirred at room temperature. Following disappearance of the starting material, the solution was washed 2 times with aq. 10% Na_2CO_3 and once with brine. The organic layer was then dried with $MgSO_4$ and the solvent removed at reduced pressure. Silica gel column chromatography (3:2 EtOAc:Hexanes) yielded the desired product as a colorless oil (206 mg, 34% yield).

^{1}H NMR (400 MHz, $CDCl_3$) δ

δ 1.26 (d, J= 5.5 Hz, 3H), δ 1.50-1.70 (m, 2H), δ 1.72-1.83 (2H), δ 2.35 (bs, 1H), δ 2.90-2.94 (m, 1H), δ 3.02-3.08 (m, 1H), δ 3.53-3.56 (m, 4H), δ 3.69-3.74 (m, 2H).

^{13}C NMR (100 MHz, $CDCl_3$) δ 71.84, δ 70.53, δ 61.74, δ 56.95, δ 52.69, δ 26.74, δ 24.45, δ 13.15.

HRMS: *m*/*z* (ESI) calculated $[M+H]^-$ = 159.1021, measured 159.1025 (Δ = 2.5 ppm)



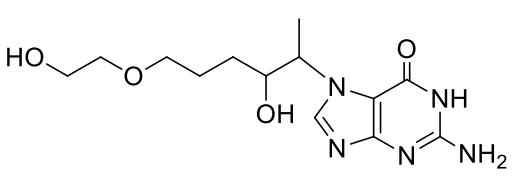
The epoxide (0.47 mmol; 1 equiv) was dissolved in CH_2Cl_2 (5 mL), followed by addition of D-biotin (0.52 mmol; 1.1 equiv). The solution was cooled to 0 °C, followed by addition of *N*,*N*'-dicyclohexylcarbodiimide (DCC; 0.52 mmol, 1.1 equiv) and 4-dimethylaminopyridine (DMAP; 0.02 mmol; 0.05 equiv). The solution was removed from the ice bath and stirred overnight at room temperature. Following removal of the solvent at reduced pressure, the reaction mixture was purified by silica gel column chromatography (7.5% MeOH in CH_2Cl_2) to yield epoxide-biotin **1** as a colorless solid (66 mg; 55 % yield).

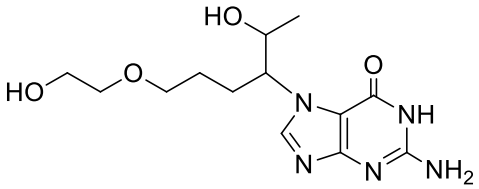
^{1}H NMR (400 MHz, $CDCl_3$) δ 1.25 (d, J= 5.6 Hz, 3H), δ 1.41-1.49 (m, 2H), δ 1.54-1.78 (m, 8H), δ 2.36 (t, J= 7.4 Hz, 2H), δ 2.70 (d, J= 12.8 Hz, 1H), δ 2.90-2.97 (m, 2H), δ

3.04-3.09 (m, 1H), δ 3.18-3.22 (m, 1H), δ 3.29-3.31 (m, 2H), δ 3.51-3.56 (m, 2H), δ 3.62-3.66 (m, 2H), δ 4.19-4.22 (m, 2H), δ 4.30 (dd, J= 7.9, 4.6 Hz, 1H), δ 4.48 (dd, J= 7.7, 4.6 Hz, 1H).

^{13}C NMR (100 MHz, $CDCl_3$) δ 201.0, δ 173.8, δ 70.2, δ 68.3, δ 63.1, δ 61.9, δ 60.2, δ 56.8, δ 55.5, δ 52.6, δ 39.6, δ 33.3, δ 28.2, δ 26.1, δ 24.5, δ 23.9, δ 11.9.

HRMS: *m*/*z* (ESI) calculated $[M+H]^+$ = 387.1979, measured 387.1961 (Δ = 4.6 ppm)





^{1}H NMR (600 MHz, D_2O) δ 0.92 (d, J=6.45 Hz, 3H), δ 1.11 (t, J=7.4 Hz, 3H), δ 1.19-1.27 (m, 1H), δ 1.33-1.41 (m, 1H), δ 1.75-1.84 (m, 2H), δ 3.01-3.07 (m, 3H), δ 3.44 (t, J=4.5 Hz, 2H), δ 3.48-3.53 (m, 2H), δ 3.62 (p, J = 6.3 Hz, 1H), δ 4.14 (t, J= 6.8 Hz, 2H), δ 7.79 (s, 1H).

^{13}C NMR (150 MHz, CD_3OD) δ 156.7, δ 155.6, δ 154.8, δ 142.5, δ 110.3, δ 86.2, δ 74.6, δ 70.9, δ 63.9, δ 63.5 δ 29.0, δ 28.8, δ 19.9.

^{1}H, ^{13}C, and 1D NOESY NMR spectroscopy were used to establish that the reaction occurred at N7 of guanine. Through space coupling with the hydrogen at C8 of guanine was observed with epoxide-substrate hydrogens are indicative of N7 labeling. In addition to the spectroscopic evidence, reaction at positions other than N7 would lead to modification on the Watson-Crick face and would inhibit reverse transcription. This would prohibit propagation of the RNA through the *in vitro* selection.

Supplementary References

1 Curtis, E.A. & Liu, D.R. Discovery of Widespread GTP–Binding Motifs in Genomic DNA and RNA. *Chem. Biol.* **20**, 521–532 (2013).

2 Curtis, E.A. & Liu, D.R. A Naturally Occurring, Noncanonical GTP Aptamer Made of Simple Tandem Repeats. *RNA Biol.* **11**, 1-10, (2014)

3 Zuker, M. Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

4 Machanick, P. & Bailey, T.L. MEME-ChIP: Motif Analysis of Large DNA Datasets. *Bioinformatics* **27**, 1696–1697 (2011).

5 Pyle, A.M. Metal Ions in the Structure and Function of RNA. *J. Biol. Inorg. Chem.* **7**, 679–690 (2002).

6 Boysen, G., Pachkowski, B.F., Nakamura, J. & Swenberg, J.A. The Formation and Biological Significance of N7-Guanine Adducts. *Mutat. Res.* **678**, 76–94 (2009).

7 Neagu, I., Koivisto, P., Neagu, C., Kostiainen, R., Stenby, K. & Peltonen, K. Butadiene Monoxide and Deoxyguanosine Alkylation Products at the N7-Position. *Carcinogenesis* **16**, 1809–1813 (1995).

8 Hansen, M.R. & Hurley, L.H. *Acc. Chem. Res.* Pluramycins. Old Drugs Having Modern Friends in Structural Biology. **29**, 249–258 (1996).

9 Burns, N. Z., Witten, M. R., Jacobsen, E. N. Dual Catalysis in Enantioselective Oxidopyrylium-Based [5+2] Cycloaddition. *J. Am. Chem. Soc.* **133**, 14578-14581 (2011).