DNAzyme-Catalyzed Site-Specific N-Alkylation of DNA Oligonucleotide Nucleobases by Reductive Amination

Robert D. Boyd, Morgan M. Kennebeck, Aurora A. Miranda, Zehui Liu, and Scott K. Silverman*

Department of Chemistry, University of Illinois Urbana-Champaign,
600 South Mathews Avenue, Urbana, IL 61801, USA

Table of Contents

HPLC purification of 5′-benzaldehyde acetal oligonucleotide .... page S2

Nucleotide and primer sequences for selection .... page S2

Uncatalyzed splinted background reactions .... page S3

Sequences of individual DNAzymes .... page S4

Selections with 5-MA .... page S5

Yield data for DNAzymes .... page S7

Reaction site assignment data for non-site-specific DNAzymes.. page S8

Data for reducing agents other than $NaCNBH_3$ .... page S9

DNAzyme secondary structure predictions using mfold .... page S9

Metal ion requirements of the DNAzymes .... page S18

Mass spectrometry of DNAzyme products .... page S20

Additional data for DNAzymes with RNA substrates .... page S21

3′-Overhang sequence requirements of DNAzymes .... page S21

Reaction site assignments for DNAzymes that branch within the duplex binding region .... page S22

References for Supplementary Information .... page S22

HPLC purification of 5′-benzaldehyde acetal oligonucleotide

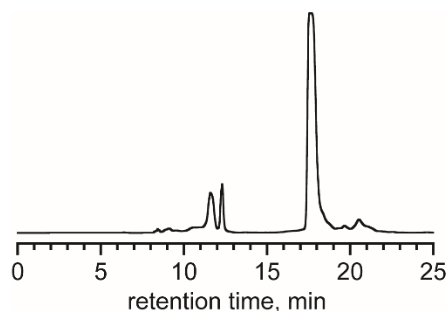


Figure S1. HPLC purification of the 18 nt 5′-benzaldehyde acetal oligonucleotide (retention time 17.6 min), prepared by solid-phase synthesis. After acetal deprotection with 80% aqueous acetic acid to form the 18 nt 5′-benzaldehyde oligonucleotide. MALDI-TOF mass spectrometry: m/z $[M+H]^+$ calcd. 5817.7, found 5818.3, Δ +0.01%.

Nucleotide and primer sequences for selection

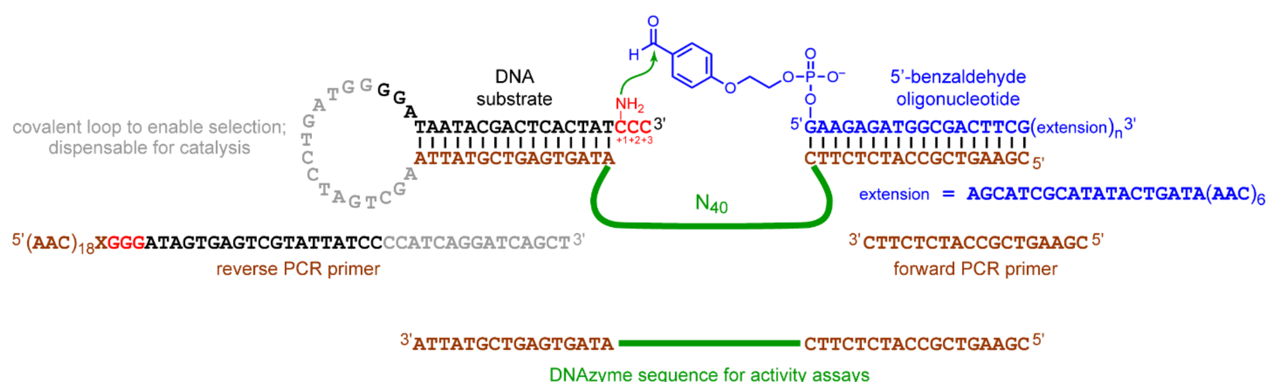


Figure S2. Nucleotide details of the in vitro selection experiments with the 5′-benzaldehyde oligonucleotide as reductive amination reaction partner. The DNA pool, which includes both the N_{40} region and the DNA substrate, is 111 nt. The DNA substrate is shown with CCC 3′-overhang, and the reverse primer has GGG to match; the appropriate changes in the reverse primer were made for GGG and AAA 3′-overhangs. In the reverse primer, X = hexa(ethylene glycol), HEG, spacer to stop Taq polymerase. The short (n=0; 18 nt) version of the 5′-benzaldehyde oligonucleotide was used in even-numbered selection rounds, and the long (n=1; 55 nt) version in odd-numbered rounds, to avoid enrichment of noncatalytic DNA sequences that have anomalous PAGE migration positions. The 55-mer was prepared from the 18-mer by splint ligation with the 37 nt extension oligonucleotide, splint 5′-TATCAGTATATGCGATGCTCGA-AGTCGCCATCTCTTC-3′, and T4 DNA ligase. Although one of the three CCC 3′-overhang nucleotides (C+1) is shown as providing the amine nucleophile, for any individual DNAzyme this role could be played by any of these three C nucleotides, as well as other nucleotides of the DNA substrate. The product migration standard during each selection round was the uncatalyzed background product formed from the CCC N_{40} pool, the 5'-benzaldehyde oligonucleotide, and splint 5′-CGAAGTCGCCATCTCTTCATAGTGAGTCGTATTATCCCCATCAGGATCAGCT-3′, incubated in 150 mM NaOAc, pH 5.2, 40 mM Mg^{2+}, 20 mM Mn^{2+}, 1 mM Zn^{2+}, and 100 mM $NaCNBH_3$ at 37 °C for 16 h. The pH was 5.2 because of higher background yield (e.g., pH 5.2, 4.8% versus pH 6.0, 0.7%).

Uncatalyzed splinted background reactions

Before we began in vitro selection, in order to determine suitable incubation conditions for the enrichment step with all six combinations of trinucleotide 3′-overhang (CCC, GGG, AAA) and pH (6.0, 7.5), we performed splinted background assays (Figure 2B). The goal was to identify conditions in which the uncatalyzed background yield does not exceed ~5%, where 5% background would limit the enrichment factor per selection round to 20-fold. From each background yield at 16 h we found that this incubation time was suitable, as shown in Figure S3A and marked in Figure 3. Separately, for all six combinations of 3′-overhang and pH, we quantitatively determined k_{bkgd} by performing assays of the uncatalyzed splinted background reaction at 24 to 96 h (Figure S3B). The k_{bkgd} values shown in the figure were used along with the k_{obs} values from Figure 6 to compute rate enhancements for individual DNAzymes.

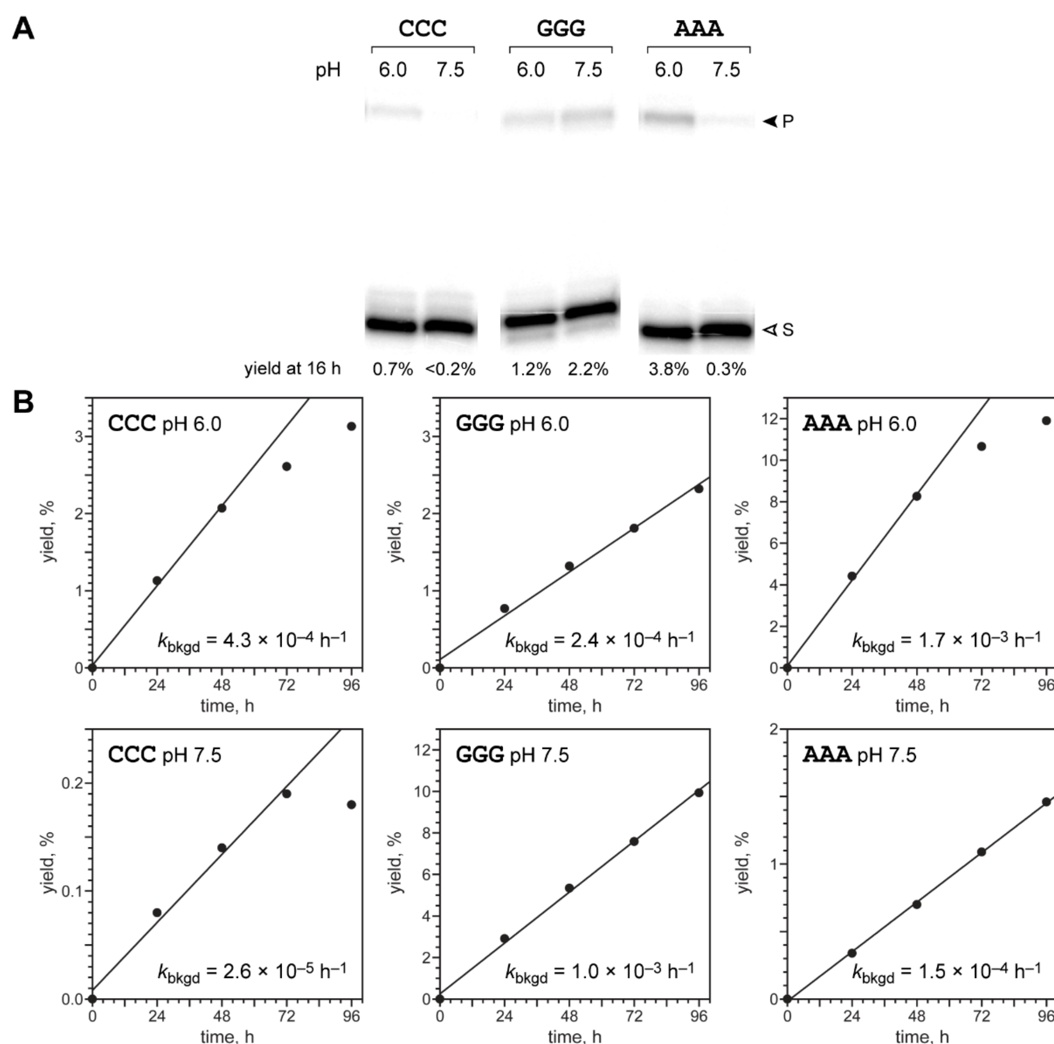


Figure S3. Assays of the uncatalyzed splinted background reactions. (A) Assays at 16 h for each combination of trinucleotide 3′-overhang (CCC, GGG, AAA) and pH (6.0, 7.5), to identify suitable incubation times for the enrichment step of selection. The observed yield for each combination is shown below each lane. S = substrate, P = background product. (B) Single-turnover assays at 24 to 96 h, to quantitatively determine the k_{bkgd} values shown with each plot.

<u>Sequences of individual DNAzymes</u>

```
           1        10         20         30         40
           |         |          |          |          |
*6LA230    GTATCGTCGG TGGAATTTTG GTTGGCGCAA CGTCTCGCCA 40 (2)
 6LA213    ...CG..G.. G.TGCCA.CT TA..AGA.GC ..GGG-.TTC 39 (2)

           1        10         20         30         40
           |         |          |          |          |
*11JP201   TGGTAGGAGC AAGAGAGGTA ACTACCTATT GGTCTGACCA 40 (12)
 11JP205   C.A.GACTAG TG..A....T .T.T.TC..C .C.GG.CGAC 40 (2)
 11JP209   .TACT..GTG GTAG.TACG. TGGTGTGGAG .AAAGTC.AT 40 (2)

           1        10         20         30         40
           |         |          |          |          |
*8LB203    GGAAGTAGTG AGGTCGGTAC AGATAGGGGA ATTTGCCTCA 40 (11)
 8LB223    AAGGA.T.A. CATCAC.A.G ..CGTT...T TA...G CGG 40 (1)

           1        10         20         30         40
           |         |          |          |          |
*8LD205    GTACAGCGGA CAGTTGATTA TCAAGCTGGA ACCGTAGCGG 40 (1)
 8LD204    .GC.G.GAA. AC.AG.GGA. C.GGATAAC. C.G.G.CGC. 40 (1)
 8LD207    .C........ ..C..TC.AC A.C.AG.... ....A..... 40 (1)
 8LD212    .GGGGCTATC ..CCA.CCC. ..GGCG.CTG GGG.CG.G.A 40 (1)

           1        10         20         30         40
           |         |          |          |          |
*8LF206    GCGTGGATCC GAACCCCAGG GCTTAGGTTA CGAGGTCTCG 40 (2)
*8LF239    CG.CACGGA. CGC..AG.TA CT.C.ACACT G..TTAAGAT 40 (1)
 8LF201    CA..A.CA.T .GTAGGGG.C CT.CTATGC. .CTCT.G.GA 40 (1)
 8LF202    ..AC.CGGTG TG..G.AC.C ..CCGA..CC .TCA..GGAC 40 (8)
 8LF203    AG.G.TTAG. .CGG.AGCC. AAGACAT.AG GTGTTCGCGT 40 (1)
 8LF208    CG..C..G.A CG-.ATT..C .G.ATA.CG. ACG...TAGC 39 (11)
 8LF212    CTCC...CGA ACGTA.GC.T .ACG..AA.. G.CACGT.T. 40 (1)
 8LF216    CG.A.....A AG-GT.GTAC ATC.C.CGAC G.T.TG.AAA 39 (1)
 8LF219    CG.G.CGAGT CCGG.AGGC. .TA.....AC TAG.AC.G.. 40 (1)
 8LF225    CGTCA.GAGG CTCAA.AGAA TGGGTAAACT GCG...TAGT 40 (1)
 8LF226    .G.G.TCAG. ACGGGAGTC. CAGCTC..CC GAGT.CGCA. 40 (1)
 8LF227    ......GC.. ....TGTGC. CAGGCAAAGG GACC.AGA.. 40 (1)

           1        10         20         30         40
           |         |          |          |          |
 6LG201    CTCGCGGCCG GGATAGTGCC GTCTAAGGCG AGTATGATAG 40 (2)

           1        10         20         30         40
           |         |          |          |          |
*8LH201    CCGGTCTACG GTTCCCGAAA GTAGGCGCCA GCCCAGCTGG 40 (1)
 8LH223    A..CATCGGC .A....AG.T .G.ACG.AGG A.GACC.C.T 40 (5)

           1        10         20         30         40
           |         |          |          |          |
*6LJ229    GAGGTGGATC GGTAACAAGG ACTGGGGAAA AAGAACGAGG 40 (1)
 6LJ201    TCACGC.GGG T..G.GC.CA .TG.T.A.G. TGTTGAACTA 40 (2)
 6LJ211    .CCAGATTGT T..GT.CCAT G..A.ATCTG GGACGGCGCA 40 (10)
 6LJ214    CGC.GCTC.A CA.GT.C.CA TAAA.ATGGG .G..CTTGCT 40 (1)
 6LJ217    AG.C.TACGT CAGGGGTTCT GAGC.CTTTG ....TGCGAT 40 (1)
 6LJ227    .GCCC..T.A A.CGGG.CA. .GGACC.TCG TG.GTAAG.A 40 (1)
 6LJ232    .T...C..CA A.AGTAT.TA CGA..AT.GG CG.CGG..AA 40 (1)
 6LJ234    CTAACCA.C. A.GTTA.C.. GAGC..AGGG TTTCT..GT. 40 (1)

           1        10         20         30         40
           |         |          |          |          |
 6LA202    GGGGGTACCT TCTACGATAG TGGATGTCGG CCATGTAGAC 40 (2)
 6LA235    .ACC..CAAA A..CG.G.C. ..AGG.CATT GGGGAG..CT 40 (1)
 11JP207   CATA.C.TTG GGGGG..G.T GTA.GCAT.. AGGG.CGCGT 40 (1)
 8LB201    ATA..ACTAG GGGTTCCA.T CTATGA..TT .TCG.C...G 40 (12)
 8LB204    CTCATG.GGG .A.CGCCAG. ..A.CTGGT. ..C.T.CA.G 40 (3)
 8LB218    C.CA.GGGA. .GATG.GAC. G..CA...T. AAGCTCCTCT 40 (2)
```

Figure S4. Sequences of the DNAzymes identified in this study. See section on "Procedures for in vitro selection and cloning" for details of the selection process, including incubation conditions. Only the initially random (N_{40}) sequences are shown. All DNAzymes were assayed as 5′-CGAAGTCGCCATCTCTTC-N_{40}-ATAGTGAGTCGTATTA-3′ (see Figure S2), with exceptions as noted in the accompanying supporting text. The last family includes all six branch-forming DNAzymes that modify a nucleotide within the portion of the DNA oligonucleotide substrate that binds to the DNAzyme by Watson-Crick base pairs. In each alignment, a dot denotes conservation, i.e., the same nucleotide as in the uppermost sequence within the family; a dash denotes a gap. On the far right is shown the sequence length and (in parentheses) the number of times that the particular sequence was found during cloning. Three sequences are 39 nt, each presumably due to a Taq polymerase deletion in an unknown selection round. In general, no common or conserved motifs were apparent either within or between sequence families. See Table S4 and Figure S8 for predicted secondary structures of the eight representative DNAzymes of Figure 6, which are each denoted here with an asterisk in front of their name.

Nine of the 40 DNAzymes, including five of the six that form highly branched products, were found to have one or more mutations or deletions in their binding arm that interacts with the DNA oligonucleotide substrate by Watson-Crick base pairs. These errors were possible because that segment of the DNAzyme was synthesized in each selection round by 3′-extension of the reverse primer (Figure S2) using Taq polymerase, which lacks proofreading ability. A summary of these DNAzymes is as follows. The relevant DNAzyme binding arm is N_{40}-ATAGTGAGTCGTATTA-3′, and its nucleotides are numbered as A–1, T–2, A–3, G–4, etc., where A–1 is underlined for reference, and the negative numbers are used to denote that we are counting in the substrate 5′-direction, moving away from the substrate 3′-overhang. Except in the three cases indicated otherwise below, the DNAzymes were not tested without the mutations or deletions that were found by sequencing. Regardless of the outcome of such tests, all assays in the remainder of the manuscript used these DNAzymes with their one or more mutations or deletions.

- 8LD204 T–5C (mutation found to be unimportant: T–5 88% and 86% at 17 and 48 h; cf. Table S2)
- 8LF206 T–2Δ (deletion found to be important: 48 h, T–2Δ 75%, T–2 31%)
- 8LF208 G–8Δ+T–9Δ
- 8LH223 T–5C
- 6LA202 T–5Δ
- 6LA235 G–6Δ+A–7Δ+G–8Δ+T–9Δ
- 11JP207 G–4Δ (deletion found to be important: 48 h, G–4Δ 90%, G–4 10%)
- 8LB204 T–2Δ
- 8LB218 A–3G

Selections with 5-MA

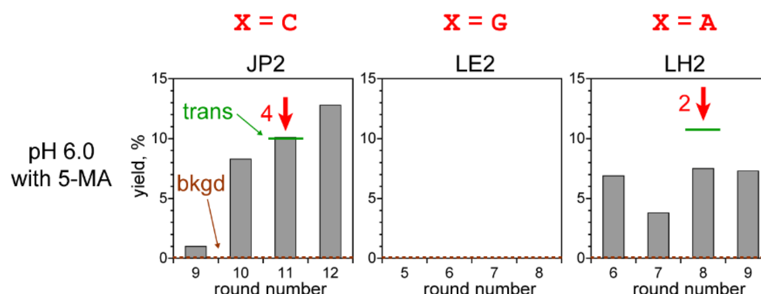


Figure S5. Progressions of the three additional in vitro selection experiments in which 1 mM 5-MA was included along with the 5′-benzaldehyde oligonucleotide reaction partner at pH 6.0. See Figure 3 for details. The splinted background yield for all three substrates was <0.3%. For the LH2 selection, the round 5 yield was <2%, and the round 8 in trans yield without 5-MA was 52% (11JP2 without 5-MA was not tested in trans). No activity was observed in any round of the LE2 selection.

DNAzyme	48 h yield –5-MA and –DMF, %	48 h yield –5-MA and +DMF, %	48 h yield +5-MA and +DMF, %
11JP201	79	82	6.7
11JP205	70	47	2.8
11JP207	91	38	12
11JP209	83	17	0.8
8LH201	97	93	13
8LH223	97	94	7.3

Table S1. Yields for DNAzymes identified by in vitro selection in the presence of 1 mM 5-MA, when assayed in the absence and presence of 1 mM 5-MA. The results reveal that all five of these DNAzymes do not need 5-MA for their catalytic activity, and indeed their activity is strongly suppressed by 5-MA. Each DNAzyme was assayed at 70 mM MES, pH 6.0, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, and 150 mM NaCl, without and with 1 mM 5-MA and 10% DMF (11JP2) or 5% DMF (8LH2), at 37 °C for 48 h.

Yield data for DNAzymes

DNAzyme	16 h yield, %	48 h yield, %
6LA213	43	69
6LA230 [a]	89	92
11JP201 [a]	47	76
11JP205	41	70
11JP209	59	83
8LB203 [a]	70	77
8LB223	*6.1*	*8.1*
6LG201	51	82
8LH201 [a]	85	92
8LH223	79	97
6LJ201	47	63
6LJ211	44	66
6LJ214	36	57
6LJ217	40	53
6LJ227	*19*	*29*
6LJ229 [a]	32	40
6LJ232	*17*	*22*
6LJ234	29	51

DNAzyme	16 h yield, %	48 h yield, %
8LD204	86	88
8LD205 [a]	91	95
8LD207	87	88
8LD212	53	83
8LF201	21	36
8LF202	55	77
8LF203	65	84
8LF206 [a]	40	67
8LF208	*12*	*21*
8LF212	45	78
8LF216	*6.6*	*7.3*
8LF219	65	89
8LF225	54	81
8LF226	86	93
8LF227	27	47
8LF239 [a]	75	84

Table S2. Yield data for the 34 DNAzymes. Left column: DNAzymes identified by selection with the CCC 3′-overhang (first seven rows) and AAA 3′-overhang (last eleven rows). Right column: DNAzymes identified by selection with the GGG 3′-overhang (16 rows). Italics denotes DNAzymes with yields <30%.

[a] Kinetic data for this DNAzyme is in Figure 6.

Reaction site assignment data for non-site-specific DNAzymes

DNAzyme	G+1	G+2	G+3
8LD212		20	80
8LF203		40	60
8LF219		68	32
8LF225	5	33	62
8LF226		77	23
8LF227	12	77	11

DNAzyme	A+1	A+2	A+3
6LG201	5	88	8
6LJ201	23	71	5

Table S3. Quantification of the multiple products formed by DNAzymes in Figure 4. Product yields in %. Blank entry means no observed product (<2%). Product yields may not sum to 100% due to rounding.

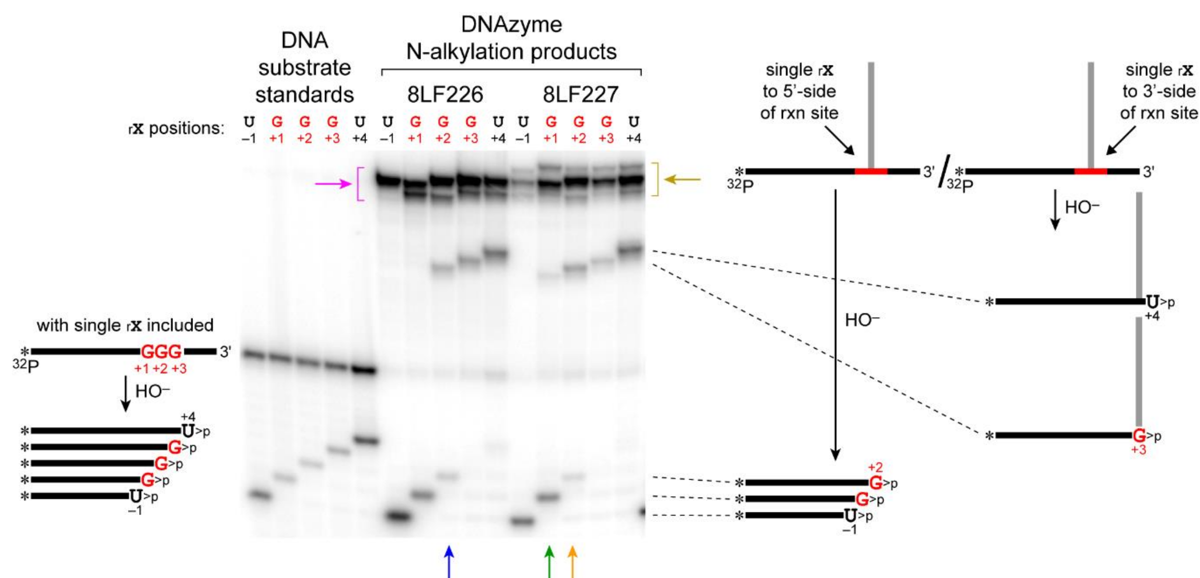


Figure S6. Assays analogous to those in Figure 5 for two DNAzymes that are not site-specific for a single nucleotide of the DNA substrate. The 8LF226 DNAzyme modifies both G+2 and G+3. The 8LF227 DNAzyme modifies all three of G+1, G+2, and G+3. On the left and right sides, the image is marked in the same way as Figure 5. For 8LF226, note the two resolved DNAzyme product bands at the top of the gel (pink arrow) and the corresponding two cleavage bands in the rG+2 lane (blue arrow). This set of bands is consistent only with a mixture of G+2 and G+3 N-alkylation, as revealed directly in Figure 4. For 8LF227, note the three resolved DNAzyme product bands at the top of the gel (yellow arrow), with the middle band predominant, and the corresponding pairs of cleavage bands in both the rG+1 and rG+2 lanes (green and orange arrows). These sets of bands are consistent only with a mixture of G+1, G+2 (favored), and G+3 N-alkylation, again as revealed directly in Figure 4.

Data for reducing agents other than $NaCNBH_3$

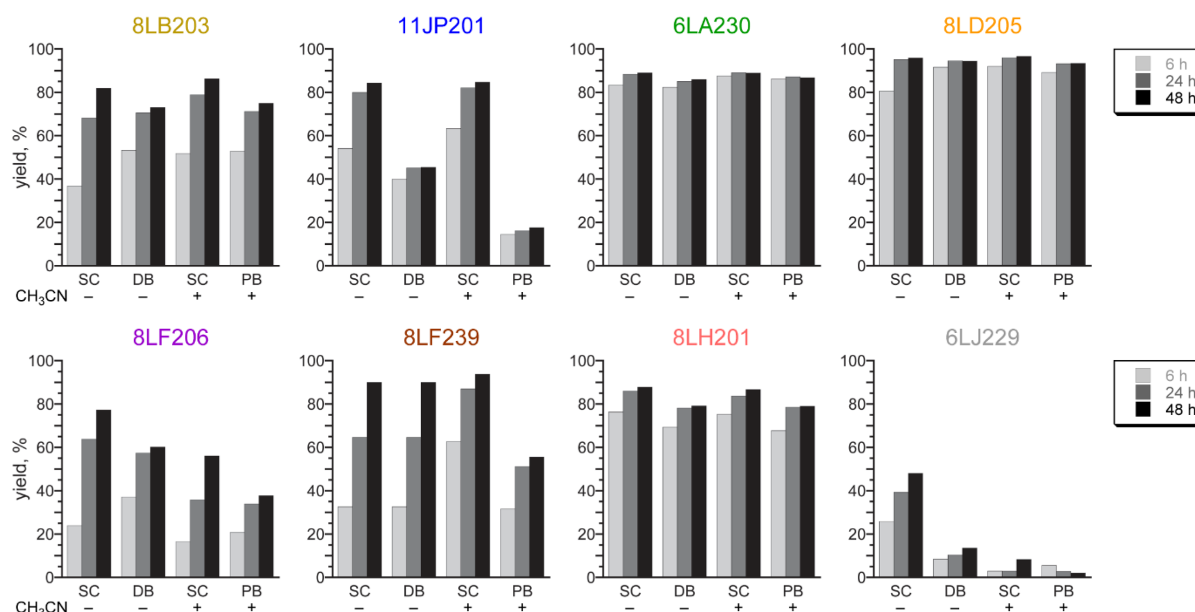


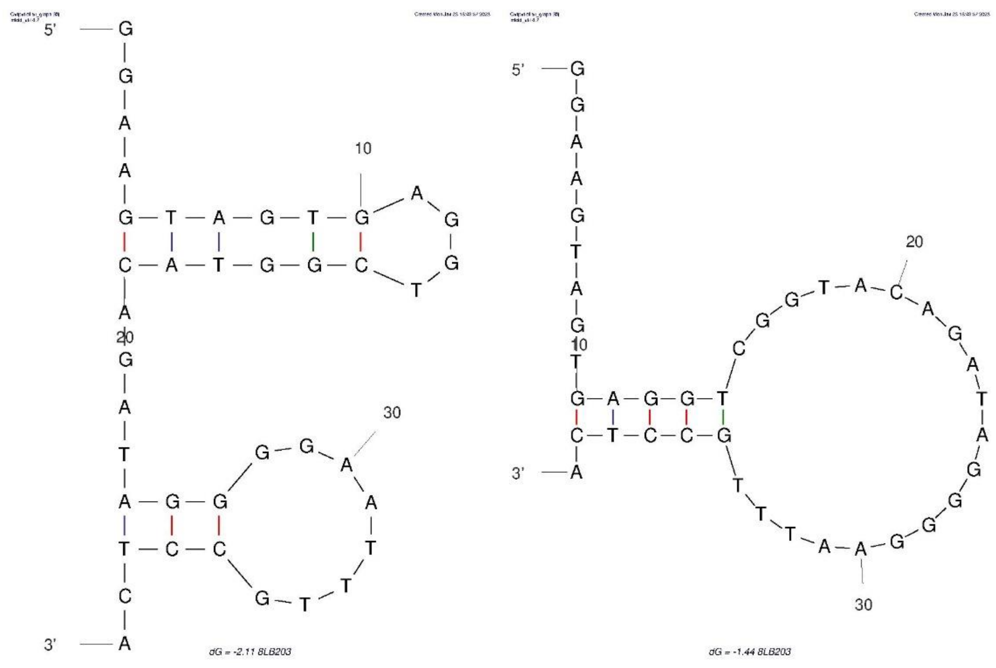
Figure S7. Data for the eight representative DNAzymes of Figure 6 with reducing agents other than $NaCNBH_3$. SC = $NaCNBH_3$, DB = dimethylamine-borane, PB = 2-picoline-borane, each at 100 mM. PB required 5% (v/v) CH_3CN for solubility.

DNAzyme secondary structure predictions using mfold

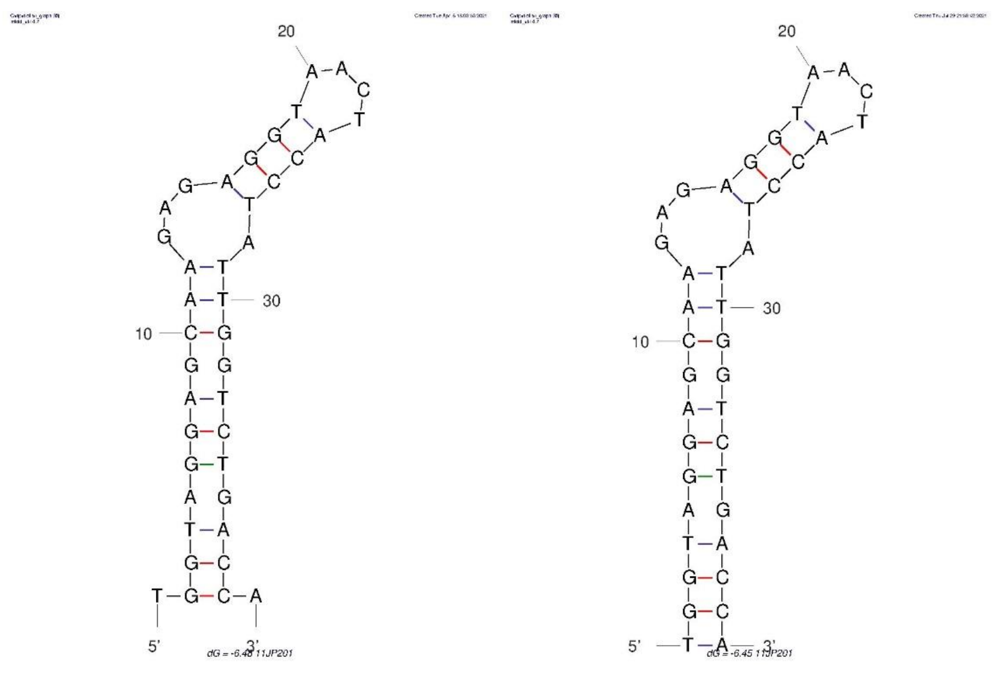
DNAzyme	number of structures	lowest ΔG, kcal/mol
8LB203	2	–2.1
11JP201	10	–6.5
6LA230	1	–4.1
8LD205	2	–4.2
8LF206	3	–3.2
8LF239	2	–3.6
8LH201	3	–6.3
6LJ229	5	+0.2

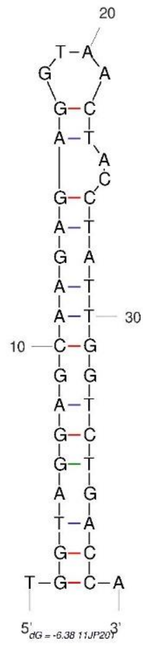
Table S4. Summary of mfold-predicted[1] secondary structures of the eight representative DNAzymes whose characterizations are shown in Figure 6. The default settings were used for the sequences of the initially random N_{40} regions with the DNA Folding Form at http://www.unafold.org/mfold/applications/dna-folding-form.php, adjusted to 150 mM Na^+ and 40 mM Mg^{2+}. The predicted secondary structures are shown in Figure S8.

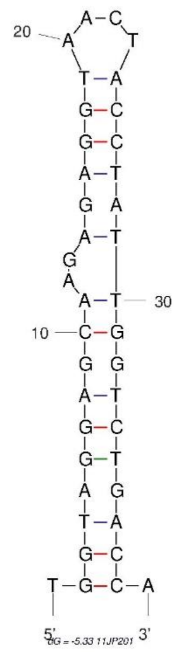
(A) 8LB203

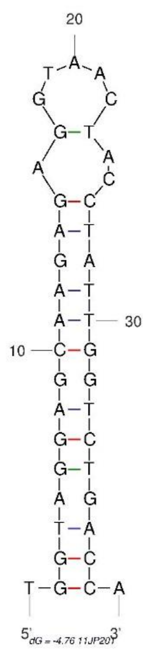


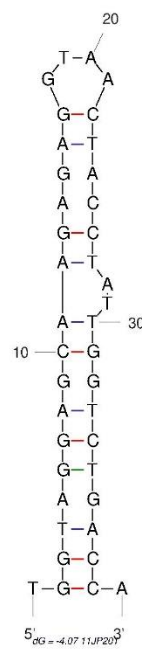
(B) 11JP201

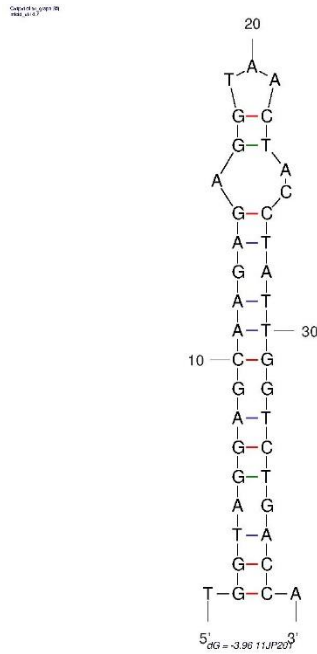


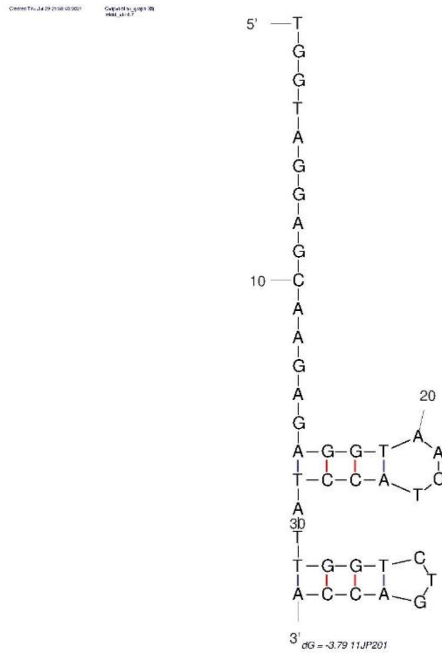

20
30
10
5'
3'
dG = -6.38 11JP201

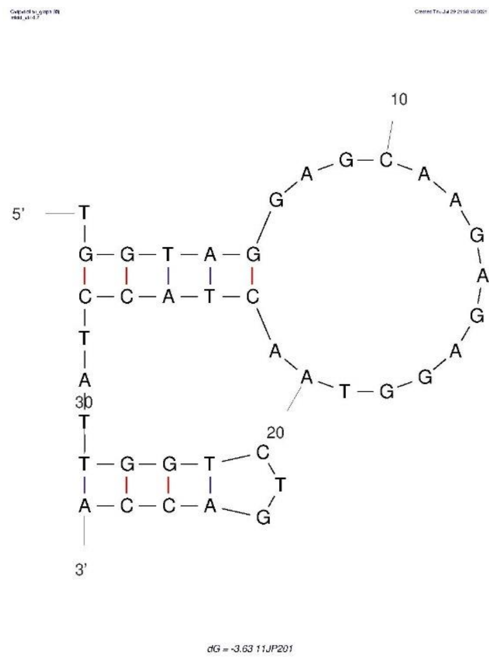

20
30
10
5'
3'
dG = -5.33 11JP201

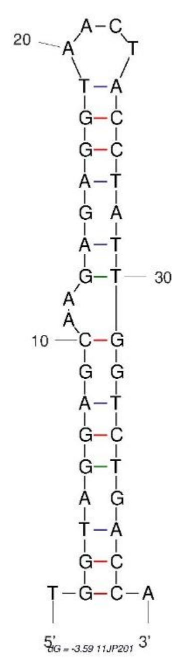

20
30
10
5'
3'
dG = -4.76 11JP201


20
30
10
5'
3'
dG = -4.07 11JP201

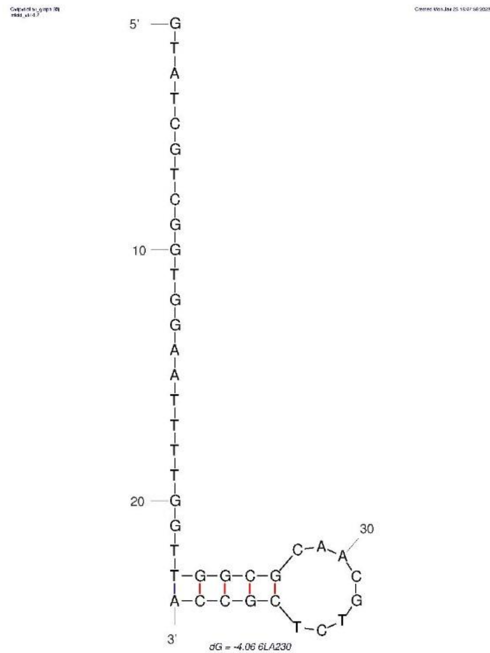

5'
3'
dG = -3.96 11JP201


5'
3'
dG = -3.79 11JP201

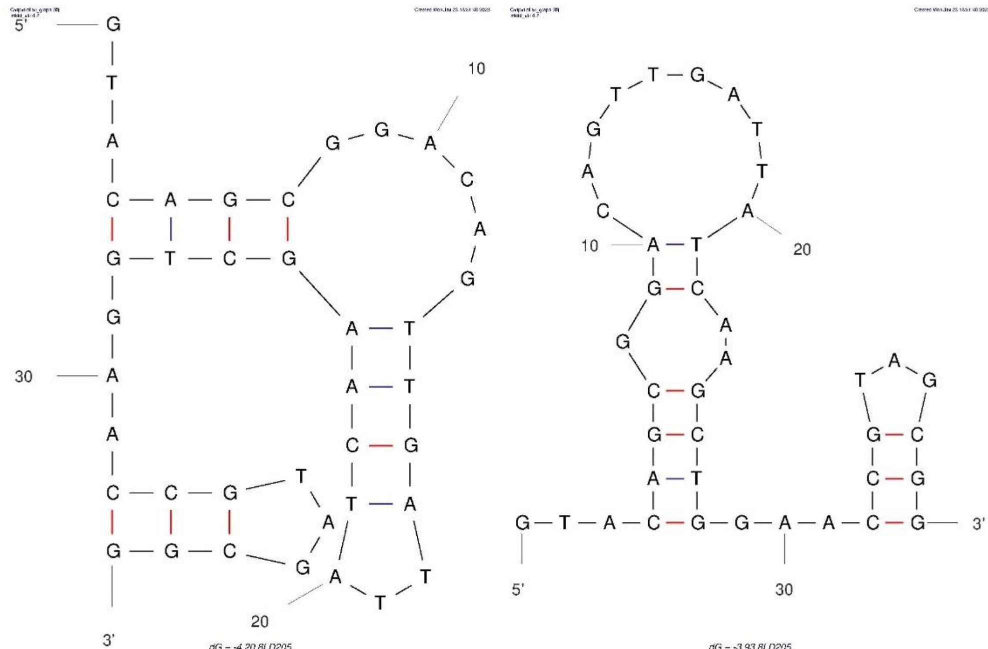

5'
3'
dG = -3.63 11JP201


5'
3'
dG = -3.59 11JP201

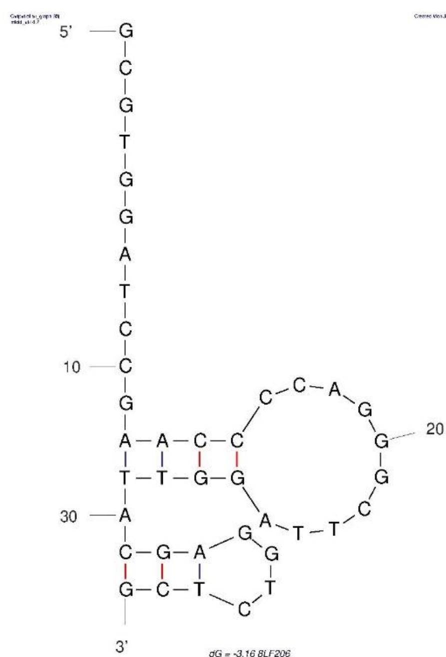
(C) 6LA230

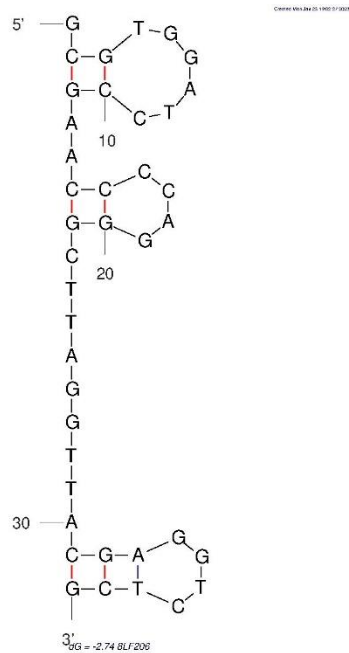


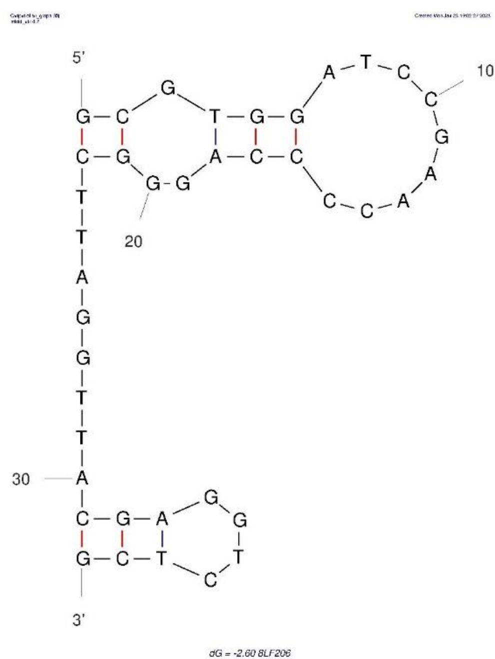
(D) 8LD205



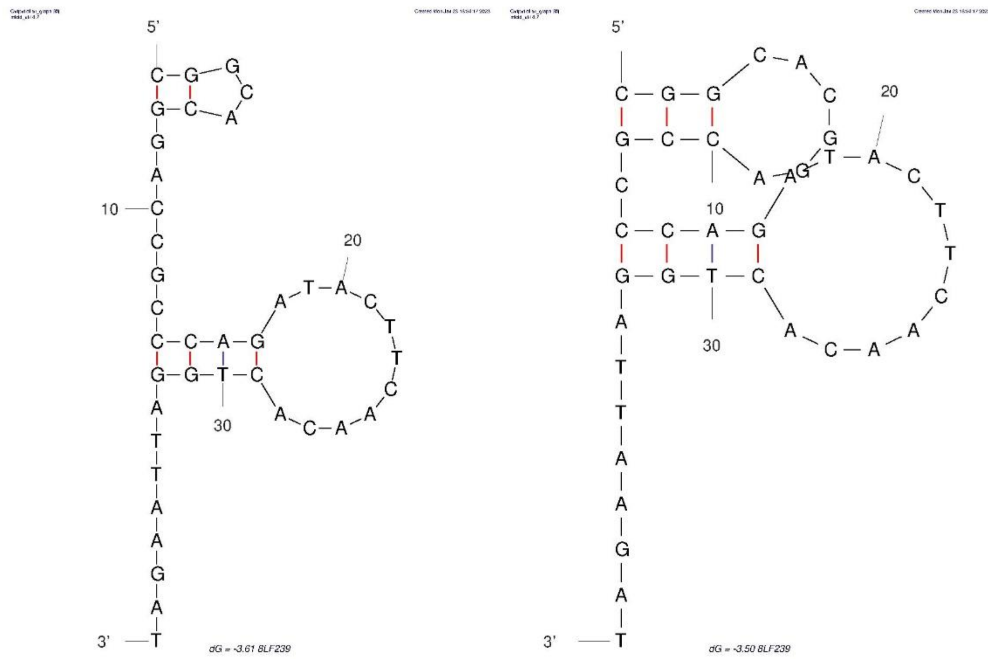
(E) 8LF206



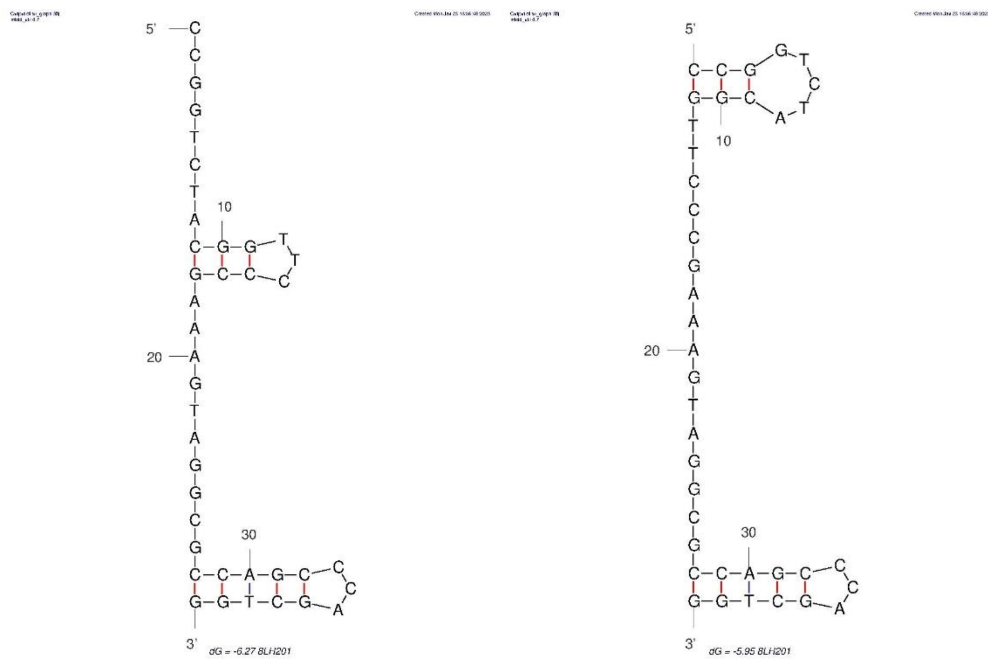


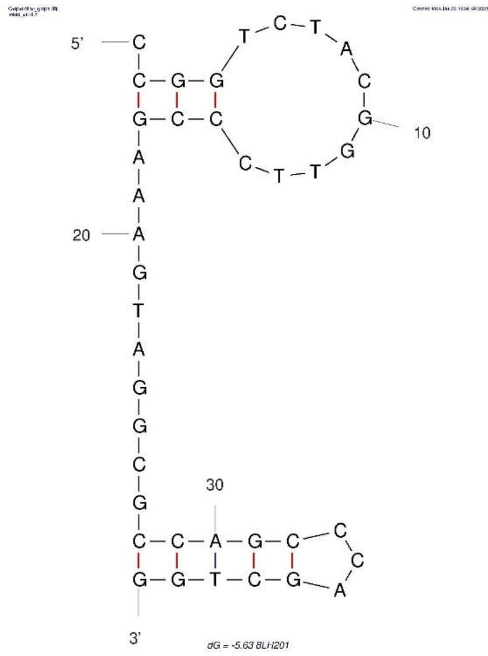


(F) 8LF239

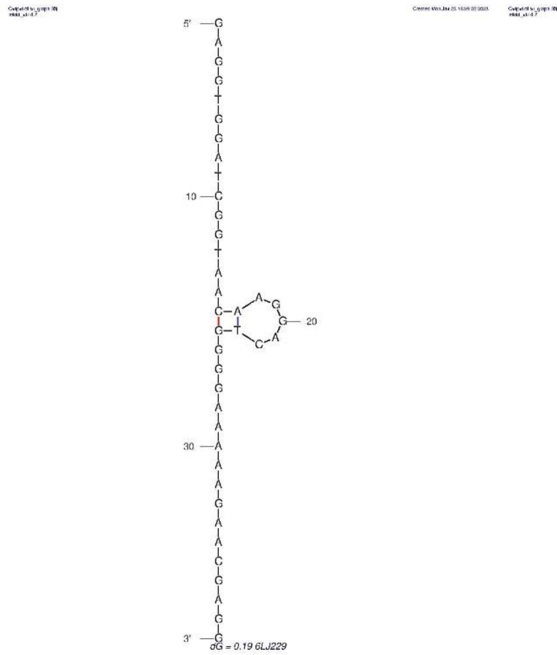


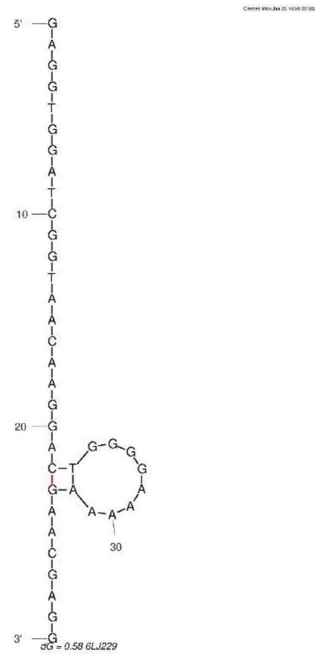
(G) 8LH201





(H) 6LJ229





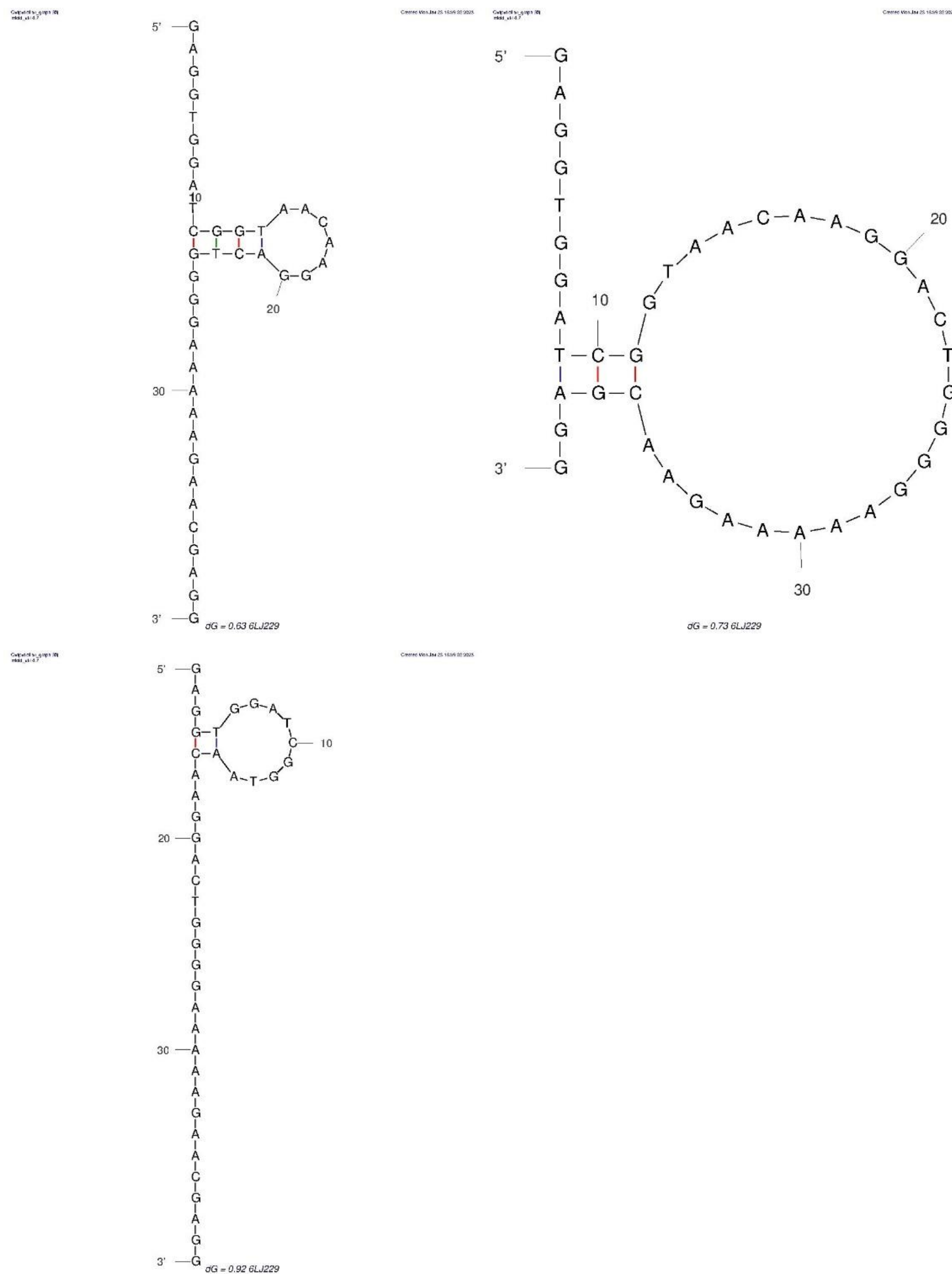
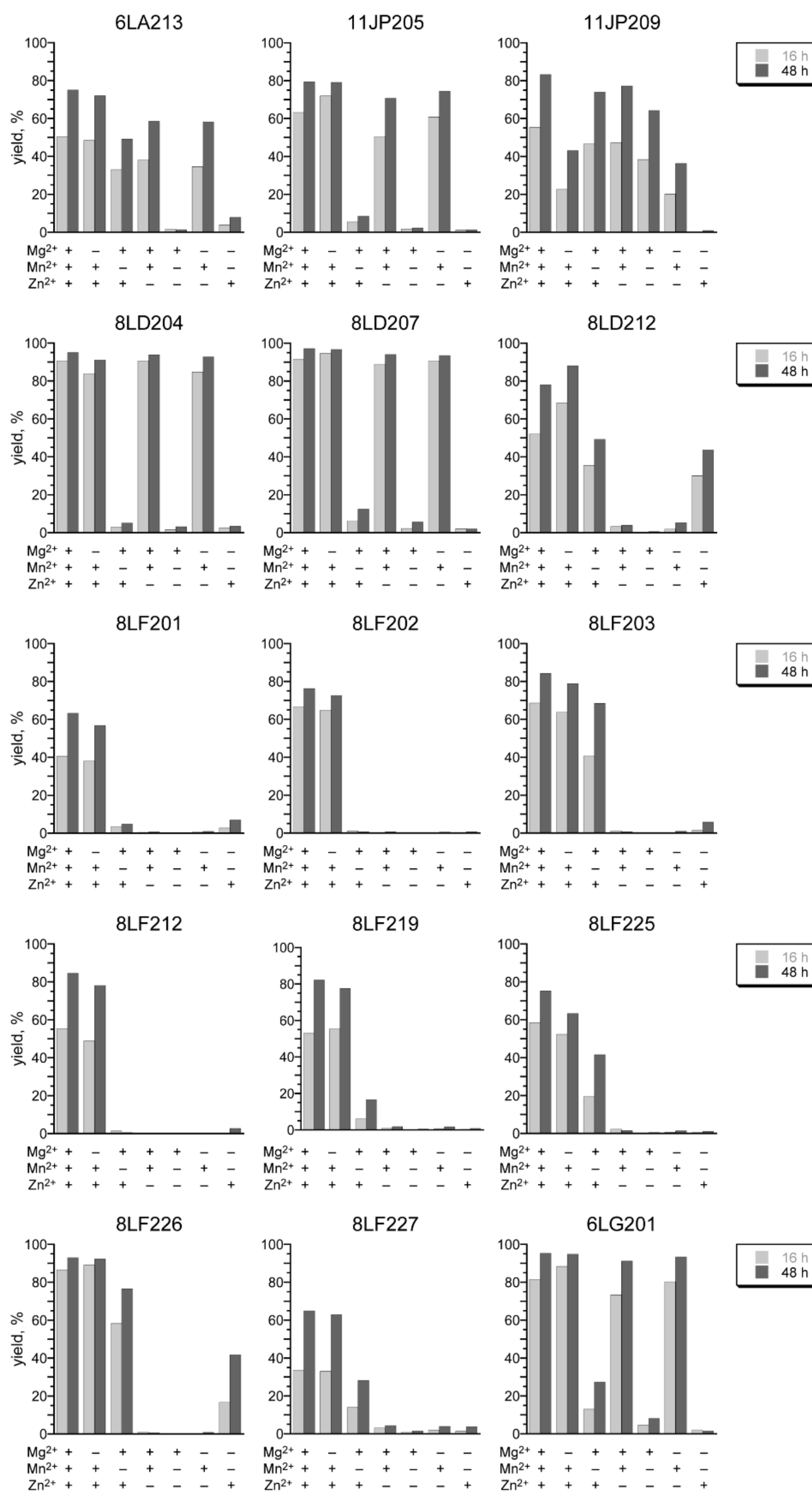


Figure S8. Mfold-predicted[1] secondary structures for the eight representative DNAzymes whose characterizations are shown in Figure 6. The default settings were used for the initially random N_{40} regions with the DNA Folding Form at http://www.unafold.org/mfold/applications/dna-folding-form.php, adjusted to 150 mM Na^{+} and 40 mM Mg^{2+}. Where multiple structures are shown for an individual DNAzyme, the lowest-energy structure (with most negative ΔG value) is shown first, followed by the remaining structure(s) in order of increasing energy. See Table S4 for full tabulation of number of structures and lowest ΔG value for each DNAzyme.

<u>Metal ion requirements of the DNAzymes</u>

Of the 34 DNAzymes, all 29 that have >30% yield were evaluated with all combinations of Mg^{2+}, Mn^{2+}, and Zn^{2+} to deconvolute their metal ion requirements. In Figure S9 is data at 16 and 48 h for all 21 DNAzymes not shown in Figure 7, as well as the six additional DNAzymes that form branches within the duplex binding arm region, with all data shown for DNAzymes in the order of Figure S4.



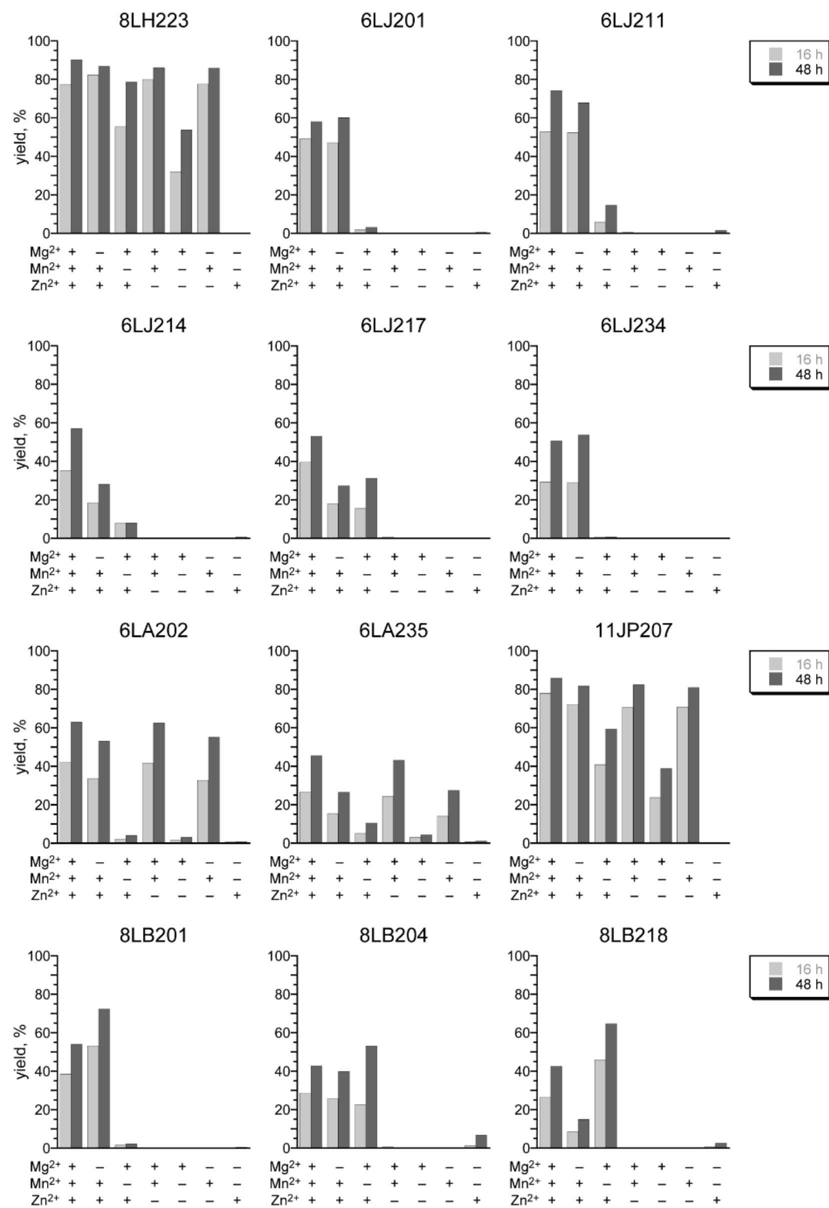


Figure S9. Metal ion requirements of the DNAzymes. Each bar plot shows the DNAzyme yield at 16 and 48 h with the indicated divalent metal ion or ions.

Mass spectrometry of DNAzyme products

DNAzyme	*m*/*z* calcd.	*m*/*z* found	Δ
8LB203	12465.2	12456.6	–0.07%
11JP201	12465.2	12463.3	–0.01%
6LA230	12465.2	12461.1	–0.03%
8LD205	12585.2	12582.2	–0.02%
8LF206	12585.2	12583.6	–0.01%
8LF239	12585.2	12586.0	+0.01%
8LH201	12537.2	12534.3	–0.02%
6LJ229	12537.2	12535.2	–0.02%

Table S5. MALDI-TOF mass spectrometry analysis of DNAzyme products. All *m*/*z* values are for $[M+H]^+$.

Additional data for DNAzymes with RNA substrates

DNAzyme	16 h yield with RNA, %	48 h yield with RNA, %	48 h yield with DNA, %
6LA230 [a]	19	38	90
8LD212 [b]	5.0	9.5	83
8LF202 [b]	6.8	9.5	77
8LF203 [b]	80	83	84
8LF219 [b]	53	80	89
8LF225 [b]	56	68	81
8LF226 [b]	97	99	93
8LF227 [b]	45	45	47
8LF239 [a]	93	96	89
6LG201 [b]	31	62	82
8LH201 [a]	95	96	97
8LH223 [b]	5.9	14	97
6LJ201 [b]	19	41	63
8LB201 [c]	56	89	79
8LB204 [c]	5.0	16	56

Table S6. Yields for the 15 DNAzymes that have catalytic activity with RNA in place of the DNA oligonucleotide substrate. [a] RNA assay shown in Figure 8; DNA yield data from Figure 6. [b] DNA yield data from Table S2. [c] DNA yield data from Figure 11.

3′-Overhang sequence requirements of DNAzymes

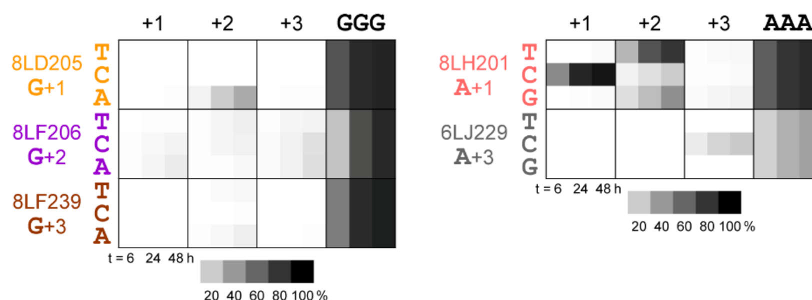


Figure S10. Assays of the GGG and AAA DNAzymes analogous to those in Figure 10.

Reaction site assignments for DNAzymes that branch within the duplex binding region

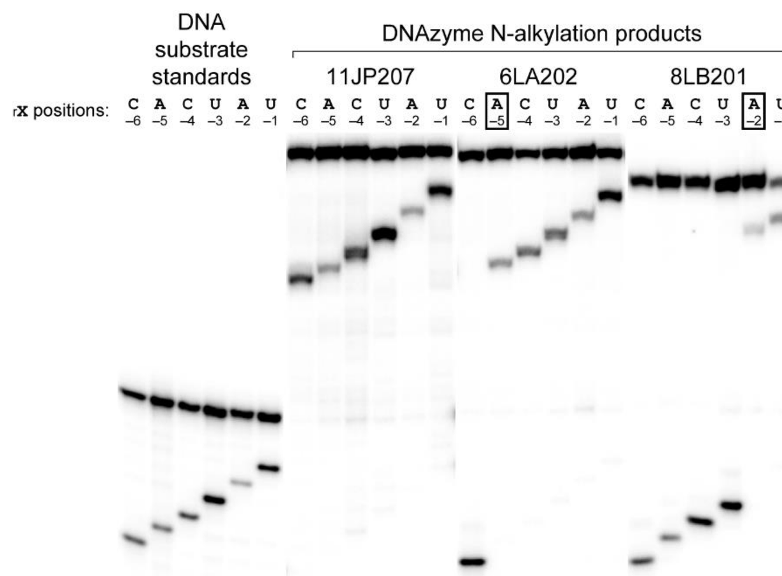


Figure S11. Reaction site assays analogous to those in Figure 5, for assigning the nucleotide branch sites of the DNAzymes that each modify a DNA substrate nucleotide to the 5′-side of the 3′-overhang, i.e., within the Watson-Crick binding duplex (Figure 11). All experiments used 3′-CCCT DNA substrates.

References for Supporting Information

(1) Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31*, 3406-3415.