Nucleic Acids Research, 2024, **52**, 8702–8716
https://doi.org/10.1093/nar/gkae639
Advance access publication date: 25 July 2024
Chemical Biology and Nucleic Acid Chemistry

Site-specific N-alkylation of DNA oligonucleotide nucleobases by DNAzyme-catalyzed reductive amination

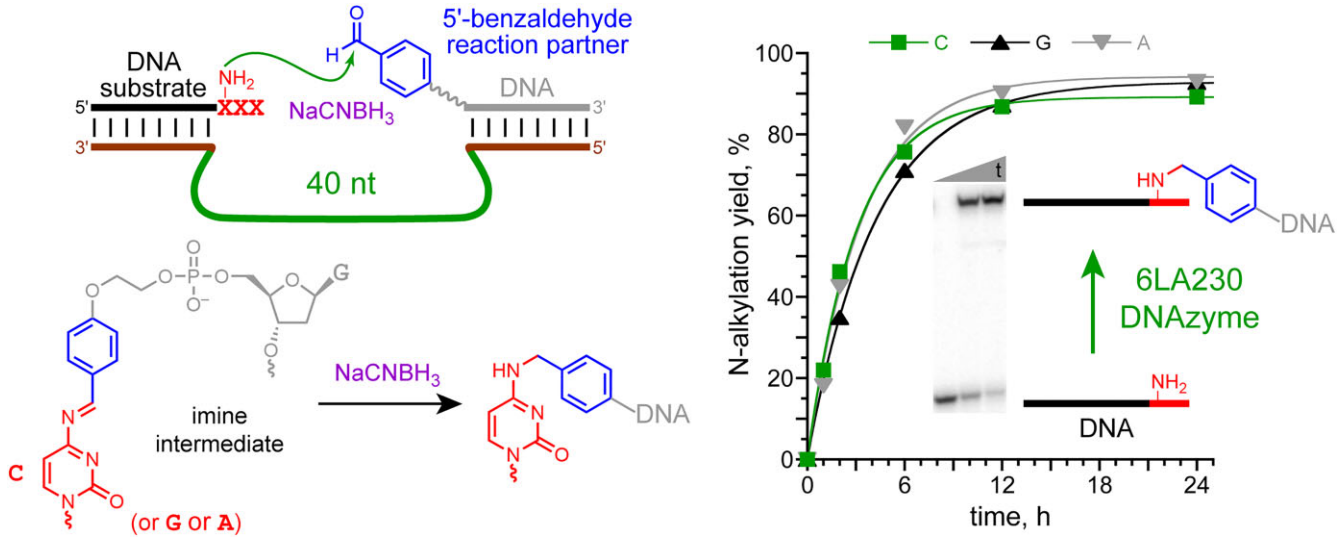
Robert D. Boyd, Morgan M. Kennebeck, Aurora A. Miranda, Zehui Liu and Scott K. Silverman*

Department of Chemistry, University of Illinois Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801, USA
*To whom correspondence should be addressed. Tel: +1 217 244 4489; Email: sks@illinois.edu

Abstract

DNA and RNA nucleobase modifications are biologically relevant and valuable in fundamental biochemical and biophysical investigations of nucleic acids. However, directly introducing site-specific nucleobase modifications into long unprotected oligonucleotides is a substantial challenge. In this study, we used *in vitro* selection to identify DNAzymes that site-specifically N-alkylate the exocyclic nucleobase amines of particular cytidine, guanosine, and adenosine (C, G and A) nucleotides in DNA substrates, by reductive amination using a 5′-benzaldehyde oligonucleotide as the reaction partner. The new DNAzymes each require one or more of Mg^{2+}, Mn^{2+}, and Zn^{2+} as metal ion cofactors and have k_{obs} from 0.04 to 0.3 h^{-1}, with rate enhancement as high as $\sim 10^4$ above the splinted background reaction. Several of the new DNAzymes are catalytically active when an RNA substrate is provided in place of DNA. Similarly, several new DNAzymes function when a small-molecule benzaldehyde compound replaces the 5′-benzaldehyde oligonucleotide. These findings expand the scope of DNAzyme catalysis to include nucleobase N-alkylation by reductive amination. Further development of this new class of DNAzymes is anticipated to facilitate practical covalent modification and labeling of DNA and RNA substrates.

Graphical abstract



Introduction

Modifications of DNA and RNA oligonucleotides play important roles in a myriad of biochemical, biophysical, and biological investigations. On the nucleobases, natural modifications that range from simple methyl (1–3) and acetyl (4–6) groups to much larger appended structures can influence both structure and function of the nucleic acid (7–10). Artificial modifications such as fluorophores, biotin, azides, alkynes, and others are commonly introduced to enable many kinds of studies (11–15). However, methods to incorporate natural and artificial modifications into DNA and RNA can be limited in various ways. A frequent bottom-up approach involves solid-phase oligonucleotide synthesis using commercially available or individually prepared nonstandard

Received: June 6, 2024. **Revised:** July 3, 2024. **Editorial Decision:** July 4, 2024. **Accepted:** July 8, 2024

nucleoside phosphoramidites (16–19), followed by ligation of two or more fragments to assemble longer oligonucleotides (20–22). A related chemoenzymatic method is based on nucleoside 3′,5′-bisphosphates and ligation (23–25). While useful, such approaches can be lengthy and laborious. In contrast, top-down chemical approaches can sometimes modify one nucleotide in a long oligonucleotide substrate without requiring any ligation steps (26–28). A distinct but relatively unexplored top-down approach is to develop new enzymatic modification reactions of intact oligonucleotides (29–32). Such reactions will ideally have both broad sequence generality and site specificity for directly modifying one freely chosen nucleotide within a long and unprotected oligonucleotide substrate.

In this study, we used *in vitro* selection (33–37) to identify new DNAzymes (deoxyribozymes, DNA enzymes) (38–46) for direct, top-down DNA nucleobase N-alkylation by reductive amination. Previously we reported DNAzymes for reductive amination of an aliphatic amino group (47). We also serendipitously identified DNAzymes that catalyze reductive amination to modify the N^2-amino group of the 5′-terminal guanosine nucleotide of an RNA substrate, by reaction with the periodate-oxidized dialdehyde 3′-terminus of a second oligonucleotide strand (48). Here we sought a more generalizable *in vitro* selection approach to identify DNAzymes for nucleobase N-alkylation by reductive amination. We targeted the exocyclic amino groups of the nucleobases of cytidine, guanosine, and adenosine (C, G and A), using a 5′-benzaldehyde oligonucleotide as the reductive amination reaction partner (Figure 1). Our results revealed many new DNAzymes with DNA nucleobase N-alkylation activity, including in some cases the ability to modify RNA rather than DNA substrates. In several instances, the DNAzymes can use small-molecule aldehydes in place of the 5′-benzaldehyde oligonucleotide.

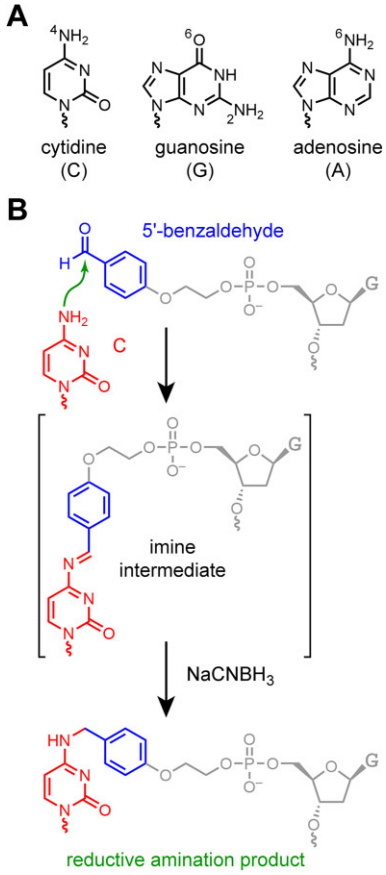


Figure 1. Nucleobase amino groups and their N-alkylation by reductive amination. (**A**) Nucleobases of C, G and A, with the respective N^4-, N^2- and N^6-amino groups marked. (**B**) N-Alkylation by reductive amination of N^4-C with a 5′-benzaldehyde oligonucleotide. The imine intermediate may be protonated as the iminium ion.

Materials and methods

Oligonucleotide preparative procedures

DNA oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA) or prepared by solid-phase synthesis on an ABI 394 instrument using reagents from Glen Research (Sterling, VA). Oligonucleotides were purified by 7 M urea denaturing 20% or 8% PAGE with running buffer 1× TBE (89 mM each Tris and boric acid and 2 mM EDTA, pH 8.3), extracted from the polyacrylamide with TEN buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 300 mM NaCl), and precipitated with ethanol.

The Glen Research 5′-benzaldehyde modifier (5′-aldehyde-modifier C2, cat. no. 10-1933) was coupled on the instrument according to the manufacturer's instructions, where the aldehyde group is protected as its 2,2-diethyl-1,3-propanediol acetal. The 18 nt 5′-benzaldehyde acetal oligonucleotide was purified by HPLC, using a Shimadzu Prominence instrument with a Phenomenex Gemini-NX C_{18} column (5 μm, 10 × 250 mm), solvent A (20 mM triethylammonium acetate in 50% acetonitrile/50% water, pH 7.0), solvent B (20 mM triethylammonium acetate in water, pH 7.0), gradient of 10% A/90% B at 0 min to 70% A/30% B at 30 min, and flow rate of 3.5 ml/min, with retention time 17.6 min (see Supplementary Figure S1 for HPLC trace). The 55 nt 5′-benzaldehyde acetal oligonucleotide was prepared from the 18 nt 5′-benzaldehyde acetal oligonucleotide by splint ligation using T4 DNA ligase and purified by 20% PAGE. The 18 nt and 55 nt 5′-benzaldehyde acetal oligonucleotides were each deprotected from the acetal to the aldehyde by treatment with 80% aqueous acetic acid (1 h, room temperature) and evaporated by SpeedVac. We chose to deprotect the acetal at the end of the synthetic procedure so that a single HPLC-purified 18 nt 5′-benzaldehyde acetal oligonucleotide could be used to prepare both the 18 nt and 55 nt 5′-benzaldehyde oligonucleotides.

Procedures for *in vitro* selection and cloning

The key enrichment step of each selection round using the 5′-benzaldehyde oligonucleotide is shown schematically in Figure 2A, with full nucleotide and primer sequences in Supplementary Figure S2. The 5′-benzaldehyde oligonucleotide was either 18 nt as depicted in Figure 2A or a 55 nt version with a long 3′-extension. The 18 nt version was used in even-numbered selection rounds, and the 55 nt version was used in odd-numbered selection rounds. This alternation was performed to suppress enrichment of noncatalytic DNA sequences that have anomalous PAGE migration positions.

Procedure for selection step in round 1

Each selection experiment was initiated with 200 pmol of the N_{40} pool, which included the attached DNA substrate. A 20 μl sample containing 200 pmol of pool and 400 pmol of 5′-

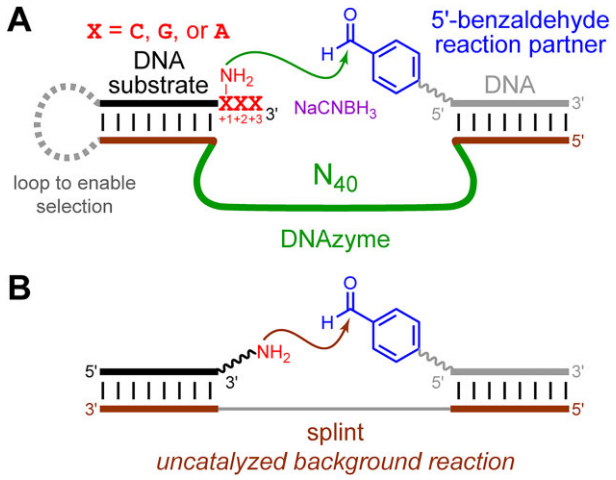


Figure 2. In vitro selection of DNAzymes for N-alkylation of DNA nucleobases by reductive amination. (**A**) Key selection step, which enables PAGE-shift enrichment of the DNAzyme population in catalytically active sequences. Any of the 'X' nucleotides may provide the amine nucleophile, enabling the particular DNAzyme sequence to survive the selection round. See Supplementary Figure S2 for nucleotide and primer details. Each DNAzyme that emerges from the selection process is assayed without the dashed loop at far left, such that the DNA substrate is bound intermolecularly (in trans) by the DNAzyme. The loop is required to enable selection, because attachment of the 5′-benzaldehyde reaction partner to the DNA substrate results in an upward PAGE shift only for catalytically active DNA sequences. Of the newly identified DNAzymes, about two-thirds (19 out of 29) catalyze the N-alkylation of a nucleobase amine on one specific nucleotide of the DNA substrate. Eight other individual DNAzymes can modify two or three consecutive nucleobase amines on the same DNA substrate, i.e. are not fully site-selective among the substrate nucleotide positions. (**B**) The uncatalyzed splinted background reaction. For this background reaction, any nucleobase amine on the DNA substrate can be the nucleophile.

benzaldehyde reaction partner was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95°C for 3 min and cooling on ice for 5 min. The selection reaction was initiated by bringing the sample to 40 μl total volume containing 70 mM MES, pH 6.0 or 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, 150 mM NaCl and 30 or 100 mM $NaCNBH_3$. When 1 mM 5-MA was included, 10% (JP2 selection) or 5% (LE2, LH2 selections) (v/v) DMF was also present for solubility. The $NaCNBH_3$ was stored as a solid in an inert atmosphere and added from a freshly prepared 1 M stock solution in water. The Mg^{2+} was added from a 20× stock solution containing 800 mM $MgCl_2$. The Mn^{2+} was added from a 20× stock solution containing 400 mM $MnCl_2$. The Zn^{2+} was added from a 20× stock solution containing 20 mM $ZnCl_2$, 40 mM HNO_3 and 400 mM MES, pH 6.0 or 400 mM HEPES, pH 7.5; this stock solution was freshly prepared from a 100× stock of 100 mM $ZnCl_2$ in 200 mM HNO_3. The metal ion stocks were added last to the final sample. The 5-MA was added from a 10× (JP2) or 20× (LE2, LH2) stock solution in DMF. The final DMF concentration (v/v) was decreased from 10% (JP2) to 5% (LE2, LH2) because the 11JP2 DNAzymes were found to be inhibited by DMF (Supplementary Table S1). The sample was incubated at 37°C for 16 h and separated by 8% PAGE.

Procedure for selection step in subsequent rounds

A 11 μl sample containing the pool and 50 pmol of 5′-benzaldehyde reaction partner was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95°C for 3 min and cooling on ice for 5 min. The selection reaction was initiated by bringing the sample to 20 μl total volume containing 70 mM MES, pH 6.0 or 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, 150 mM NaCl, 30 or 100 mM $NaCNBH_3$ and (when included) 1 mM 5-MA. The sample was incubated at 37°C for 16 h and separated by 8% PAGE.

Procedure for PCR

In each selection round, two PCR reactions were performed, 10-cycle PCR followed by 30-cycle PCR. First, a 100 μl sample was prepared containing the PAGE-purified selection product, 200 pmol of forward primer, 50 pmol of reverse primer, 20 nmol of each dNTP, 10 μl of 10× Taq polymerase buffer [1× = 20 mM Tris–HCl, pH 8.8, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$ and 0.1% Triton X-100], and 1 μl of Taq polymerase, expressed as reported (49). This sample was cycled 10 times according to the following PCR program: 94°C for 2 min, 10× (94°C for 30 s, 47°C for 30 s, 72°C for 30 s), 72°C for 5 min. Taq polymerase was removed by phenol/chloroform extraction. Second, a 50 μl sample was prepared containing 2 μl of the 10-cycle PCR product, 100 pmol of forward primer, 25 pmol of reverse primer, 10 nmol of each dNTP, 20 μCi of α-^{32}P-dCTP (3000 Ci/mmol), 5 μl of 10× Taq polymerase buffer, and 0.5 μl of Taq polymerase. This sample was cycled 30 times according to the following PCR program: 94°C for 2 min, 30× (94°C for 30 s, 47°C for 30 s, 72°C for 30 s), 72°C for 5 min. Samples were separated by 8% PAGE.

Cloning and screening of individual DNAzymes

The PCR primers used for cloning were 5′-CGAAGTCGCCATCTCTTC-3′ (forward primer; same as in selection) and 5′-TAATTAATTAATTACCCATCAGGATCAGCT-3′ (reverse primer). The 10-cycle PCR product from the appropriate selection round was diluted 10^3-fold. A 50 μl sample was prepared containing 1 μl of the diluted 10-cycle PCR product from the appropriate selection round, 100 pmol of forward cloning primer, 25 pmol of reverse cloning primer, 10 nmol of each dNTP, 5 μl of 10× Taq polymerase buffer, and 0.5 μl of Taq polymerase. This sample was cycled 30 times according to the following PCR program: 94°C for 2 min, 30× (94°C for 30 s, 47°C for 30 s, 72°C for 30 s), 72°C for 5 min. The sample was separated by 1% agarose gel and extracted using a GeneJET Gel Extraction Kit (Thermo Fisher). The extracted product was quantified by absorbance (A_{260}) and diluted to 5–10 ng/μl. A 4 μl portion of the diluted PCR product was inserted into the pCR2.1-TOPO vector using a TOPO TA cloning kit (Thermo Fisher). Individual *E. coli* colonies harboring plasmids with inserts were identified by blue-white screening and grown in LB/amp media. Miniprep DNA was prepared using a GeneJET Plasmid Miniprep Kit (Thermo Fisher) and screened for properly sized inserts by EcoRI digestion and agarose gel analysis. Before sequencing, single-turnover assays of individual DNAzyme clones were performed with PAGE-purified DNA strands prepared by PCR from the miniprep DNA.

Single-turnover DNAzyme or splinted background assay procedure

The general single-turnover assay procedure for each DNAzyme or background splint was as follows. The DNA or RNA substrate was 5′-^{32}P-radiolabeled using γ-^{32}P-ATP and T4 polynucleotide kinase. A 11 μl sample containing 0.5 pmol of 5′-^{32}P radiolabeled DNA or RNA substrate, 5 pmol of DNAzyme or splint, 10 pmol of 5′-benzaldehyde reaction partner, and 100 pmol of 'blocking oligonucleotide' [60-mer $(AAC)_{20}$, to suppress nonspecific binding to plastic tubes] was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95°C for 3 min and cooling on ice 5 min. The DNAzyme-catalyzed or splinted background reaction was initiated by bringing the sample to 20 μl total volume containing 70 mM MES, pH 6.0 or 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, 150 mM NaCl, and 100 mM $NaCNBH_3$. Refer to the selection procedure above for $NaCNBH_3$ and divalent metal ion stocks. The sample was incubated at 37°C. At each timepoint, a 2 μl aliquot was quenched with 7 μl of stop solution (80% formamide, 1× TBE [89 mM each Tris and boric acid and 2 mM EDTA, pH 8.3], 50 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol). Before PAGE for most DNAzyme assays, to each quenched sample was added 100 pmol of a 'decoy oligonucleotide', which was the 37-mer background splint that is complementary to both DNAzyme binding arms (Figure 2B). This decoy oligonucleotide was added to displace the DNAzyme from the product. In these cases when the decoy was omitted, gel bands were noticeably smeared, which inhibited proper quantification. Quenched samples were separated by 20% PAGE and quantified using a Phosphorimager. Values of k_{obs} were obtained by fitting the yield versus time data directly to first-order kinetics; i.e. yield $= Y\bullet(1 - e^{-kt})$, where $k = k_{obs}$ and Y is the final yield. For the background reactions (Supplementary Figure S3), the initial points were fit to a straight line, and k_{bkgd} was taken as the slope of the line.

Ribonucleotide substitution assay to assign nucleotide modification sites

For each DNAzyme, we 5′-^{32}P-radiolabeled each member of a set of five (for the assays of Figure 5) or six (for the assays of Supplementary Figure S11) DNA substrates, each with a single ribonucleotide at an internal position. Depending on the DNAzyme yield with the particular ribo-substituted substrate, 1–10 pmol of radiolabeled substrate was treated according to the single-turnover assay procedure for 48–72 h, except using 10–15 pmol of DNAzyme, 20–30 pmol of 5′-benzaldehyde reaction partner, and 100–150 pmol of blocking oligonucleotide, and quenching with 20 μl of stop solution. The DNAzyme product was separated by 12% PAGE (on 20% PAGE, the DNAzyme itself overlaps with the product) and precipitated with ethanol. For the assay, to each product was added 9 μl of 10 mM NaOH, pH 12.0, and 200 mM $MgCl_2$. The sample was incubated at 55°C for 1 h, quenched with 1 μl of 3 M NaOAc, pH 5.2, and 10 μl of stop solution, and separated by 20% PAGE. The ribo-substituted DNA substrate standards were generated in the same way, using 0.5 pmol of 5′-^{32}P-radiolabeled substrate.

Mass spectrometry of DNAzyme products

Each product for Supplementary Table S5 was prepared from a 16.5 μl sample containing 300 pmol of DNA substrate, 330 pmol of DNAzyme, and 360 pmol of 5′-benzaldehyde reaction partner. The sample was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl and 0.1 mM EDTA by heating at 95°C for 3 min and cooling on ice for 5 min. The DNAzyme-catalyzed reaction was initiated by bringing the sample to 30 μl total volume containing 70 mM MES, pH 6.0 or 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, 150 mM NaCl and 100 mM $NaCNBH_3$. Refer to the selection procedure above for $NaCNBH_3$ and divalent metal ion stocks.

The small-molecule benzaldehyde product for Figure 9B was prepared from a 10 μl sample containing 200 pmol of DNA substrate, 250 pmol of DNAzyme, and 300 pmol of the unmodified DNA oligonucleotide that was formerly connected to the benzaldehyde moiety. The sample was annealed in 5 mM HEPES, pH 7.5, 15 mM NaCl, and 0.1 mM EDTA by heating at 95°C for 3 min and cooling on ice for 5 min. The DNAzyme-catalyzed reaction was initiated by bringing the sample to 20 μl total volume containing 70 mM MES, pH 6.0 or 70 mM HEPES, pH 7.5, 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, 150 mM NaCl, 1 mM 4-(propargyloxy)benzaldehyde and 100 mM $NaCNBH_3$. The 4-(propargyloxy)benzaldehyde was added from a 20× stock solution in DMF.

The sample was incubated at 37°C for 48 h, separated by 20% PAGE, and desalted by Millipore C_{18} ZipTip. Data were acquired on a Bruker UltrafleXtreme MALDI-TOF mass spectrometer with matrix 3-hydroxypicolinic acid in positive ion mode at the UIUC School of Chemical Sciences Mass Spectrometry Laboratory. All m/z values are for $[M + H]^+$.

Results

In vitro selection experiments for DNA nucleobase N-alkylation by reductive amination

Our general *in vitro* selection design is in Figure 2A. We chose three related DNA substrates, each of which has an overhanging CCC, GGG, or AAA trinucleotide at its 3′-end, thus providing several nucleobase amines as potential nucleophiles. A repeating trinucleotide was used so that the resulting DNAzymes can be assessed for site selectivity among several competing nucleophiles that are chemically identical. The electrophilic reaction partner was a 5′-benzaldehyde oligonucleotide, chosen for its simple preparation and straightforward binding to the DNAzyme pool via Watson-Crick base pairs. Each of the three DNA substrates was used for *in vitro* selection at both pH 6.0 and 7.5, providing a set of six selection experiments. All selections included 40 mM Mg^{2+}, 20 mM Mn^{2+} and 1 mM Zn^{2+}, each of which has been a useful cofactor for DNAzyme catalysis (38–46), as well as 30 mM $NaCNBH_3$ as the reducing agent. Both the pH values and the divalent metal ions should enable a range of DNAzyme-catalyzed reductive amination mechanisms, which are not preprogrammed into the selection process. The N_{40} DNAzyme pool was connected to the DNA substrate, such that nucleobase N-alkylation with the 5′-benzaldehyde oligonucleotide results in a substantial upward PAGE shift and allows enrichment of the catalytically active DNA sequences.

In a series of iterated selection rounds, the pool population was enriched in the functional, catalytically active DNA sequences. In the 16 h incubation time for each combination of trinucleotide 3′-overhang and pH, the yield of the background

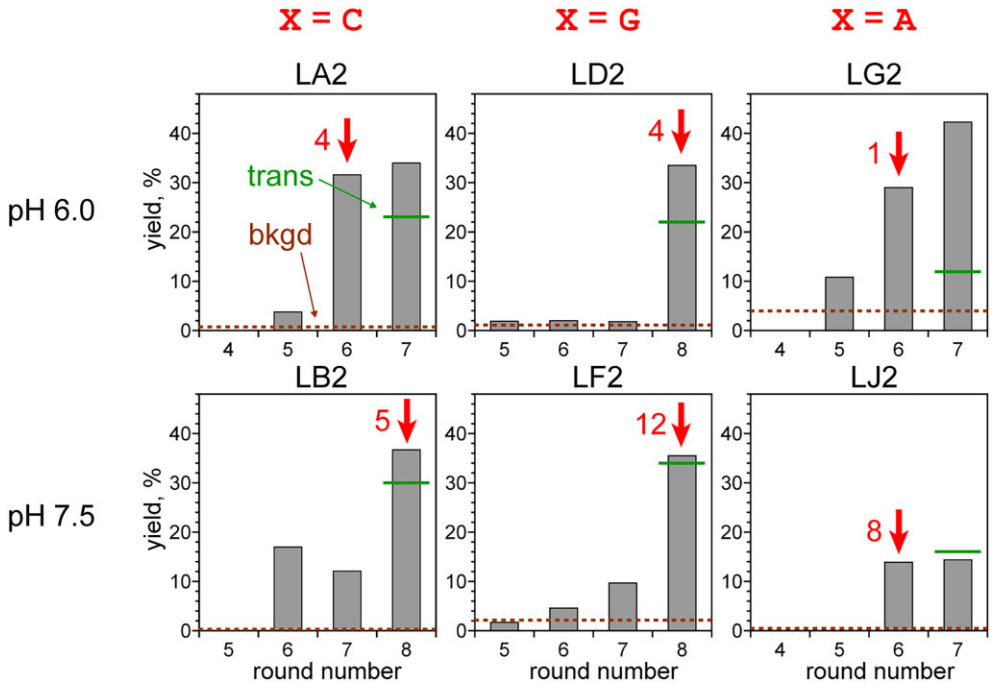


Figure 3. In vitro selection progressions. Each of the six selections has an arbitrary alphanumeric designation ranging from LA2 through LJ2, depending on the combination of 3′-overhang sequence and pH during the enrichment step of selection (Figure 2A). The incubation conditions were either 70 mM MES, pH 6.0, or 70 mM HEPES, pH 7.5, each including 40 mM $MgCl_2$, 20 mM $MnCl_2$, 1 mM $ZnCl_2$, 150 mM NaCl and 30 mM $NaCNBH_3$ at 37°C for 16 h. Each plot shows the yields for the final four selection rounds, and arrows mark the six cloned selection rounds. The number of individual DNAzymes identified from that selection experiment is shown next to each arrow, with sequences in Supplementary Figure S4. Two of the four 6LA2 DNAzymes and three of the five 8LB2 DNAzymes form branches within the duplex binding arm region (Figure 11). With each plot is shown the background yield under the same conditions (Supplementary Figure S3). The in trans assay yield (i.e. with grey loop of Figure 2A not intact) is shown for each cloned round, except the LA2, LG2, and LJ2 selections were each cloned from round 6 rather than round 7 in an effort to identify a more diverse range of DNAzyme sequences.

reaction (Figure 2B) was <5% (Supplementary Figure S3), which allows at least 20-fold enrichment in each selection round. Each selection was given an arbitrary alphanumeric designation as follows: CCC, pH 6.0 LA2 and pH 7.5 LB2; GGG, pH 6.0 LD2 and pH 7.5 LF2; AAA, pH 6.0 LG2 and pH 7.5 LJ2. When a suitable level of catalytic activity above background was reached (Figure 3), each selection pool was cloned to identify individual DNAzymes, whose sequences are collected in Supplementary Figure S4. In total, 29 new DNAzymes were found: four with the CCC substrate (two each at pH 6.0 and 7.5), 16 with GGG (four at pH 6.0, 12 at pH 7.5), and nine with AAA (one at pH 6.0; eight at pH 7.5). Each DNAzyme was named on the basis of its selection round number, selection designation, and clone number. For instance, the 6LA230 DNAzyme was identified from round 6 of the LA2 selection, clone 30.

Selection experiments with an alternative electrophile

In addition to providing each DNA oligonucleotide substrate with the 5′-benzaldehyde electrophile, in parallel we considered an alternative electrophile formed upon reaction of the benzaldehyde moiety with 5-methoxyanthranilic acid (5-MA, 2-amino-5-methoxybenzoic acid), where this compound is known to catalyze formation of oximes and hydrazones (50). The intention was that the 5-MA imine (or iminium) electrophile could replace the benzaldehyde itself in the reductive amination mechanism, offering an alternative pathway for the net reaction.

We performed three selections that included 1 mM 5-MA at pH 6.0, one selection with each of the CCC, GGG, and AAA 3′-overhangs. The selections with CCC and AAA led to modest pool activity, whereas the selection with GGG led to no activity above background (Supplementary Figure S5; alphanumeric selection designations: CCC JP2, GGG LE2, and AAA LH2). Upon cloning at respective rounds 11 and 8, the resulting five individual DNAzymes (three 11JP2 and two 8LH2; Supplementary Figure S4) were found to function much better in the absence of 5-MA (Supplementary Table S1). Therefore, for all subsequent characterizations, these five DNAzymes were included alongside the DNAzymes from the analogous two selections without 5-MA (LA2 and LG2), for an updated total of 29 + 5 = 34 DNAzymes. Of these 34 DNAzymes, 29 have >30% yield in 48 h (Supplementary Table S2). We speculate that the DNAzymes identified in the presence of 5-MA nevertheless form the nucleobase imine by catalyzing reaction of the nucleobase with benzaldehyde itself, rather than with the benzaldehyde 5-MA imine. Therefore we infer that 5-MA likely inhibits the reaction by unproductively forming the benzaldehyde 5-MA imine, and omitting 5-MA leads to better DNAzyme activity.

PAGE shift assay to assign particular nucleotide modification sites

We used a PAGE shift assay to determine the particular nucleotide within the DNA substrate that any individual DNAzyme modifies when catalyzing the reductive amination reaction. The basis of this assay is the well-known effect that

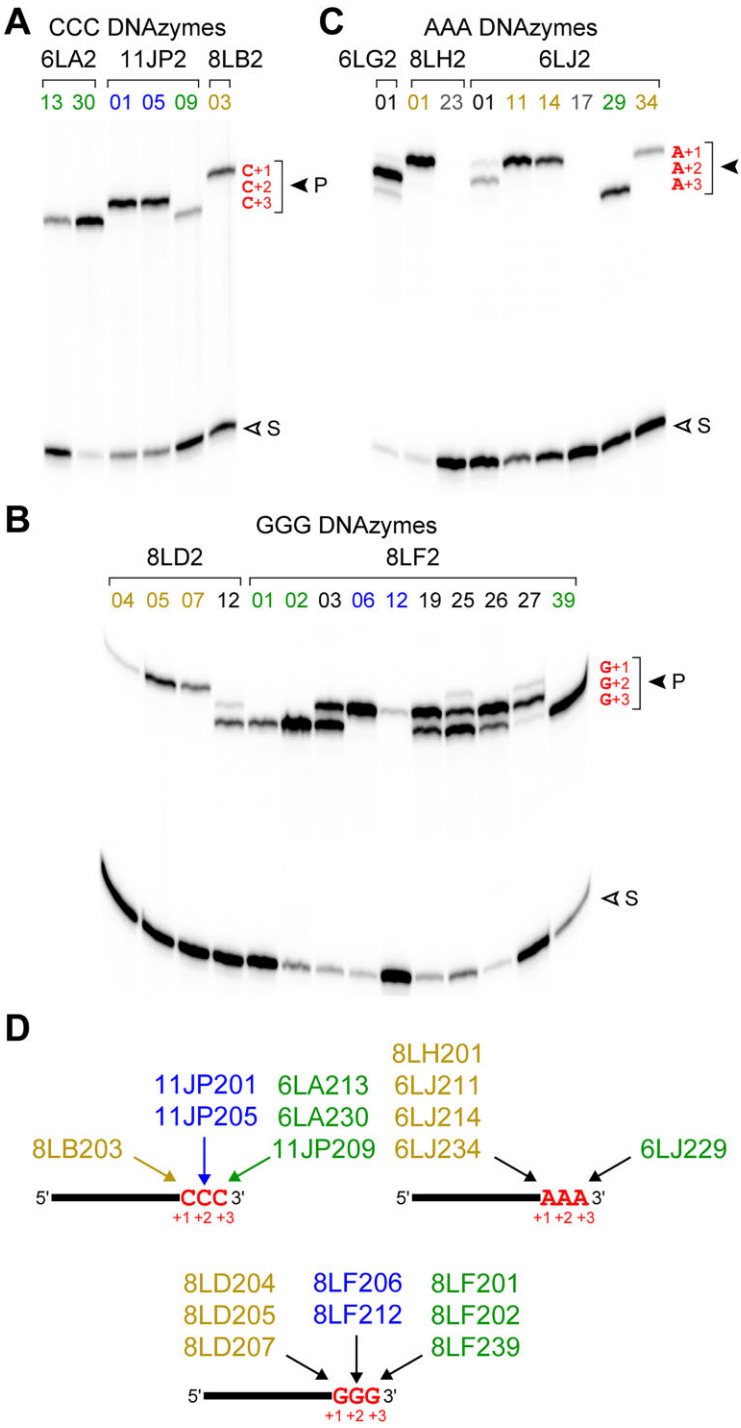


Figure 4. Using a PAGE-shift assay to assign the N-alkylation nucleotide sites on the DNA substrate for the new DNAzymes. All 29 DNAzymes with >30% yield in 48 h were assayed. (**A**) Six CCC DNAzymes. (**B**) Fourteen GGG DNAzymes. (**C**) Nine AAA DNAzymes. (**D**) Summary of the reaction sites on the DNA substrates used by the 19 new site-specific DNAzymes.

a branched oligonucleotide migrates more slowly on PAGE than does the isomeric linear oligonucleotide, and this migration effect is larger when the branch point is further from the oligonucleotide terminus. Using a DNA substrate that has an extended 3′-overhang, such as CCCTTTTTT ($CCCT_6$) for the substrate with the CCC 3′-overhang, all 29 new DNAzymes that have >30% yield in 48 h were assayed with regard to product migration position (Figure 4). Modification at the first (+1) nucleotide of the 3′-overhang results in a product that is slightly more branched, and thus slightly more slowly migrating, than for modification at the second (+2) nucleotide, and similarly for modification at the third (+3) nucleotide. The data reveal that the three isomeric products are all readily distinguishable on PAGE, thus allowing product assignment in terms of both modification site(s) and site specificity for nearly all of the new DNAzymes.

Collectively we identified 19 DNAzymes that site-specifically (>95%) modify eight of the nine combinations

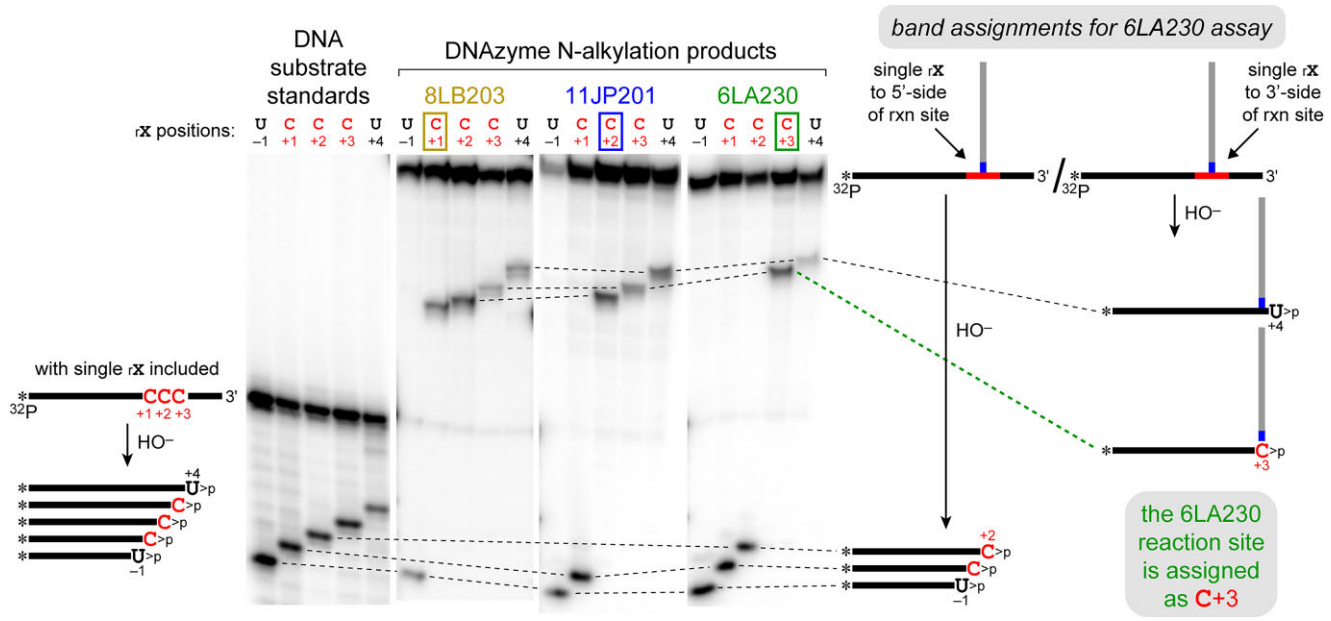


Figure 5. Using a ribonucleotide substitution assay to assign the N-alkylation nucleotide sites on the DNA substrate for the new DNAzymes. The representative PAGE image shows reaction site assignments for three DNAzymes (8LB203, 11JP201 and 6LA230) that modify the three respective 3′-overhang C nucleotides of this DNA substrate (C+1, C+2 and C+3). In the stylized depictions, * is the 5′-^{32}P radiolabel, and >p is the 2′,3′-cyclic phosphate at the 3′-end of the fragment formed upon base-promoted cleavage at the ribonucleotide position within the otherwise-DNA substrate. The assays included a 6 nt 3′-extension (T_6) on the DNA substrate, so that the upper cleavage bands are well-separated from the uncleaved N-alkylation products. All such assignments are consistent with those made from the PAGE-shift assay data in Figure 4.

of 3′-overhang (CCC, GGG, AAA) and nucleotide position (+1, +2, +3); only modification at A+2 was not found in this relatively small sample set. All six CCC DNAzymes modify a single C nucleotide. The 8LB203 DNAzyme modifies its DNA substrate site-specifically at the first of the three 3′-overhang CCC nucleotides, i.e. at C+1 of the DNA substrate. 11JP201 and 11JP205 modify C+2, and 6LA213, 6LA230, and 11JP209 modify C+3. In contrast, not all of the GGG and AAA DNAzymes were found to be site specific. Eight of the 14 GGG DNAzymes are site specific: 8LD204, 8LD205, and 8LD207 at G+1; 8LF206 and 8LF212 at G+2; and 8LF201, 8LF202, and 8LF239 at G+3. Of the remaining six GGG DNAzymes, 8LD212, 8LF203, 8LF219 and 8LF226 each modify both G+2 and G+3, and 8LF225 and 8LF227 modify all three of G+1, G+2 and G+3, with quantifications of product distributions in Supplementary Table S3. Finally, five of the nine AAA DNAzymes are site specific: 8LH201, 6LJ211, 6LJ214 and 6LJ234 at A+1, and 6LJ229 at A+3. 6LJ201 modifies both A+1 and A+2, and 6LG201 modifies all three of A+1, A+2 and A+3. Only the 8LH223 and 6LJ217 DNAzymes do not tolerate the T_6 3′-extension on the DNA substrate, and therefore their product PAGE positions could not be observed.

Ribonucleotide substitution assay to assign particular nucleotide modification sites

To corroborate the nucleotide modification site assignments made on the basis of PAGE shift assay data (Figure 4), we designed an alternative assay that is derived from the long-used 'partial alkaline hydrolysis' method for determining RNA branch points (51,52), except here we have DNA rather than RNA (Figure 5). For the DNAzyme's 5′-^{32}P-radiolabeled product, the reactive nucleotide site within the DNA substrate portion was determined by treating a series of individual, separately isolated, single-ribonucleotide-containing products with base (10 mM NaOH, pH 12, 200 mM Mg^{2+}, 55°C, 1 h) to generate a diagnostic cleavage pattern. When the ribonucleotide is placed to the 5′-side of the 5′-benzaldehyde oligonucleotide's attachment point on the substrate, base treatment of the product to cleave at the ribonucleotide position (which forms 2′,3′-cyclic phosphate and 5′-hydroxyl termini) also detaches the aldehyde oligonucleotide fragment, leading to a relatively small, faster-migrating product that has the ^{32}P radiolabel. In contrast, when the ribonucleotide is located to the 3′-side of the 5′-benzaldehyde oligonucleotide's attachment point, base treatment and ribonucleotide cleavage leaves the aldehyde oligonucleotide fragment with the ^{32}P-radiolabeled segment, resulting in a relatively large, slower-migrating product. The location of the discontinuity in the migration positions of the successive base-treatment products directly reveals the nucleotide modification site. For example, Figure 5 shows that the 6LA230 DNAzyme modifies its DNA substrate site-specifically at C+3. Similarly, 8LB203 and 11JP201 modify the DNA substrate at C+1 and C+2, respectively.

All 29 new DNAzymes that have >30% yield in 48 h were assayed by this procedure. Of the 19 out of 29 DNAzymes that were found by the PAGE-shift assay of Figure 4 to be site-specific, the ribo substitution experiment confirmed 14 of these outcomes, including the three representative examples in Figure 5. The other five DNAzyme products could not be assigned by this procedure, due to complications from one or more of the ribo substitutions. The eight DNAzymes (six for G modification, two for A modification) that modify more than one nucleotide according to the PAGE-shift assay were found to have the same outcomes by the ribo substitution procedure (Supplementary Figure S6). The final two DNAzymes again

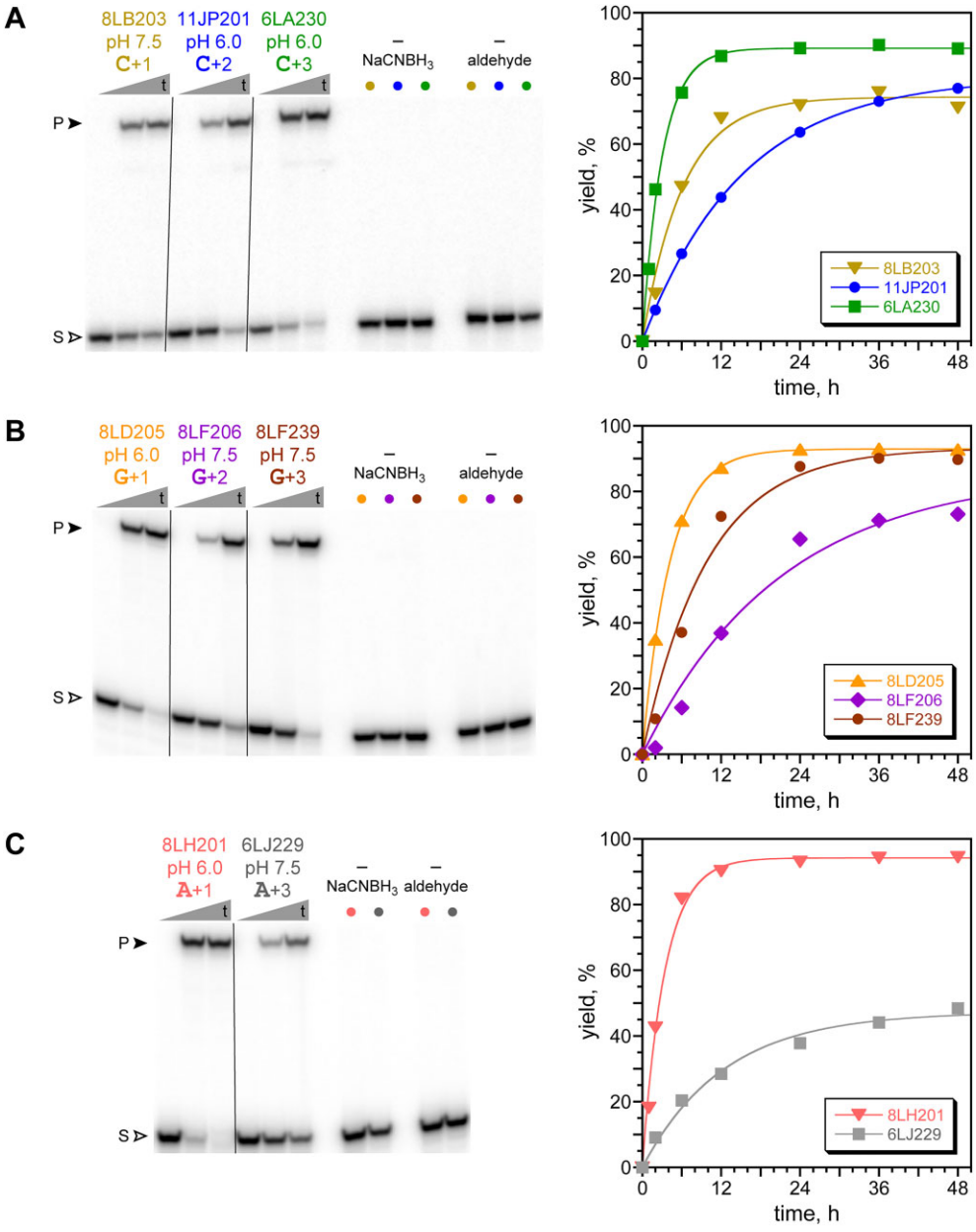


Figure 6. Assays of DNAzymes identified by *in vitro* selection for N-alkylation by reductive amination of DNA oligonucleotide nucleobase amines. Each panel shows PAGE and kinetic data for representative DNAzymes, where each DNAzyme is listed with the pH value used for its identification, and the indicated DNA substrate reaction site was determined from assays such as those in Figures 4 and 5. All DNAzymes were assayed in the incubation conditions used for their identification (exception: 6LJ229 with 1.5 rather than 1.0 mM Zn^{2+}), with 100 mM $NaCNBH_3$. Shown are representative timepoints ($t = 0.5$ min, 6 h, 48 h for DNAzymes; 48 h for negative controls, either without $NaCNBH_3$ or without 5′-benzaldehyde oligonucleotide reaction partner; S = substrate, P = product). DNAzymes were from the *in vitro* selections with DNA substrate that has (**A**) CCC 3′-overhang, (**B**) GGG 3′-overhang and (**C**) AAA 3′-overhang. k_{obs} values (h^{-1}, ± half of range, each $n = 2$): 8LB203, 0.23 ± 0.06; 11JP201, 0.068 ± 0.002; 6LA230, 0.33 ± 0.01; 8LD205, 0.24 ± 0.01; 8LF206, 0.044 ± 0.003; 8LF239, 0.12 ± 0.02; 8LH201, 0.29 ± 0.01; 6LJ229, 0.072 ± 0.009. k_{bkgd} values (h^{-1}, from Supplementary Figure S3): CCC pH 6.0, 4.3×10^{-4}; CCC pH 7.5, 2.6×10^{-5}; GGG pH 6.0, 2.4×10^{-4}; GGG pH 7.5, 1.0×10^{-3}; AAA pH 6.0, 1.7×10^{-3}; AAA pH 7.5, 1.5×10^{-4}.

could not be assayed because they do not tolerate the T_6 3′-extension on their DNA substrate.

Kinetic and other characterizations of the DNAzymes

For more detailed characterizations, we chose eight DNAzymes that each represent one of the eight marked nucleotide positions in Figure 4D. The k_{obs} values for these eight DNAzymes ranged from 0.04–0.3 h^{-1} with yields of 50–100% (all but one over 70%), and control experiments omitting either $NaCNBH_3$ or the 5′-benzaldehyde oligonucleotide reaction partner supported the assigned reductive amination reaction (Figure 6). Yield data for the other 26 out of 34 DNAzymes are in Supplementary Table S2; 21 out of 26 of these DNAzymes have yields of >30%, and 16 of these 21 yields are >60%. The splinted background rate constants k_{bkgd} were determined (Supplementary Figure S3; see Figure 6 caption for values), and rate enhancement values k_{obs}/k_{bkgd} were calculated. The highest rate enhancement of 8800

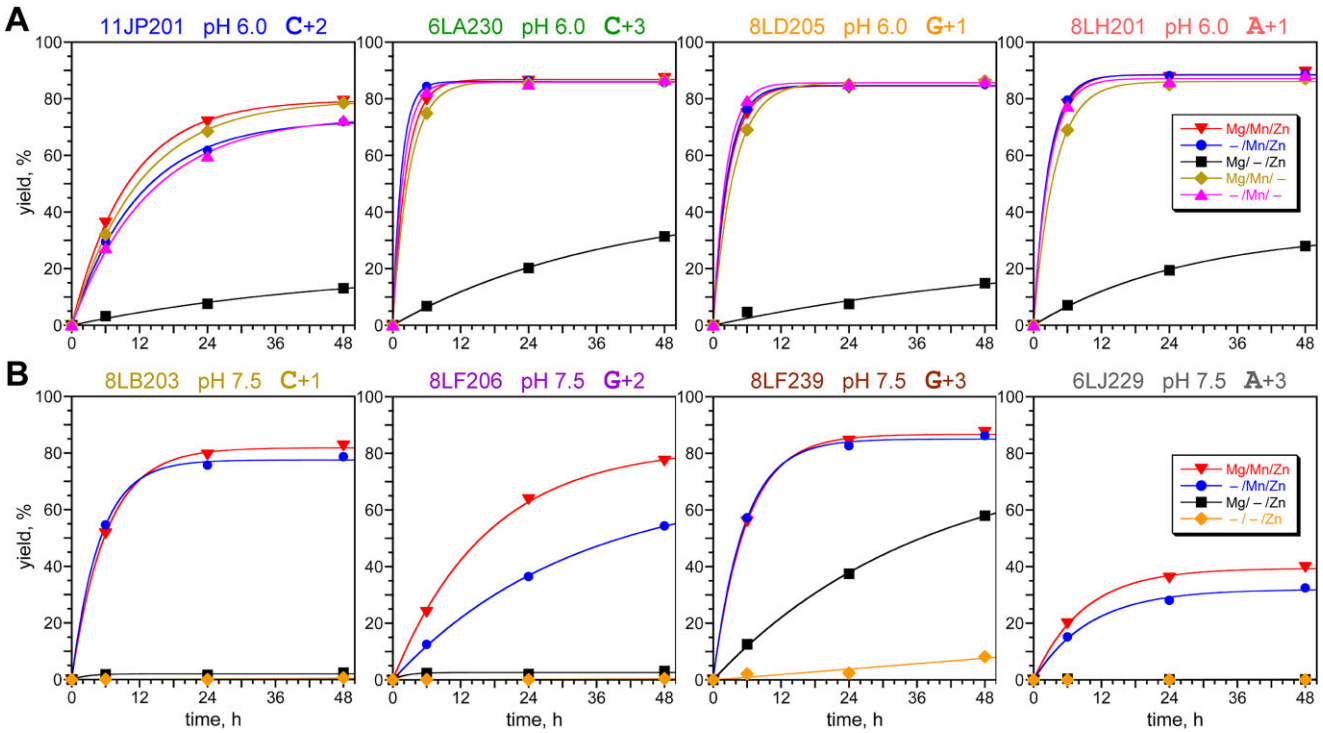


Figure 7. Divalent metal ion requirements of the eight representative DNAzymes from Figure 6. Each DNAzyme, as well as each of the 21 other DNAzymes with >30% yield (Supplementary Figure S9), was surveyed at the pH value of its identification (6.0 or 7.5 as marked) with all possible combinations of 40 mM Mg^{2+}, 20 mM Mn^{2+} and 1 mM Zn^{2+}. With one exception as noted, the 48 h yield with any metal ion combination not shown in the legend was <5%. (**A**) The four of eight representative DNAzymes identified at pH 6.0. Not plotted: 11JP201 has 19% yield at 48 h with Mg^{2+} alone. (**B**) The four of eight representative DNAzymes identified at pH 7.5.

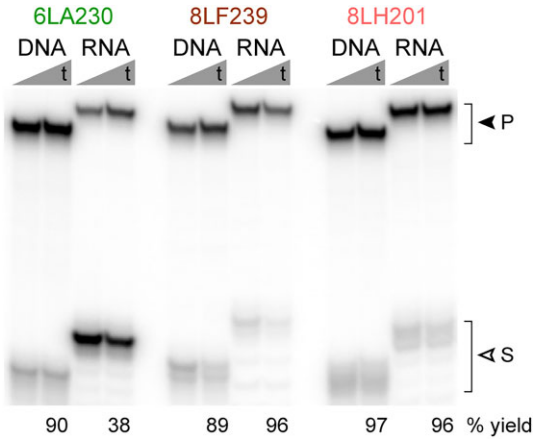


Figure 8. Many of the DNAzymes retain catalytic activity when the DNA substrate is replaced with RNA. $t = 16$ h, 48 h under the DNAzyme's selection conditions. See Supplementary Table S6 for yields for all 13 DNAzymes with DNA and RNA.

(i.e. $\sim 10^4$) was for 8LB203, owing in part to the low k_{bkgd} value found for this particular combination of CCC 3′-overhang and pH 7.5.

When $NaCNBH_3$ was replaced with the alternative reducing agent dimethylamine-borane or 2-picoline-borane (53,54), in nearly all cases we still observed a substantial amount of reductive amination product (Supplementary Figure S7). This outcome suggests that the DNAzymes catalyze the initial imine formation step of Figure 1, whereas the subsequent imine reduction step is likely uncatalyzed. Further experiments that are beyond the scope of this study are needed to explore the structures and mechanisms of the new DNAzymes. Mfold secondary structure predictions (55) for the eight representative DNAzymes are summarized in Supplementary Table S4 and shown in Supplementary Figure S8; these predictions have not been tested experimentally. High-resolution 3D DNAzyme structures are only recently emerging (56–59) and constitute a major challenge of their own.

Metal ion requirements of the DNAzymes

We assayed the eight representative DNAzymes of Figure 6 with all possible combinations of Mg^{2+}, Mn^{2+} and Zn^{2+} to deconvolute their metal ion requirements (Figure 7). Analogous assays for all 21 remaining DNAzymes that have >30% yield are in Supplementary Figure S9. Each of the 29 DNAzymes—12 identified at pH 6.0, 17 found at pH 7.5—had no activity (<0.5%) when tested in the absence of any divalent metal ions.

Nine of the 12 pH 6.0 DNAzymes, including all four in Figure 7A, achieve nearly the same yield with Mn^{2+} alone as they do with all three of Mg^{2+}, Mn^{2+} and Zn^{2+}, whereas they have very little activity with either Mg^{2+} or Zn^{2+} alone. Many of these DNAzymes also function, with lower yield, with Mg^{2+} and Zn^{2+} in the absence of Mn^{2+}. Two of the remaining three DNAzymes, 11JP209 and 8LH223, have substantial activity with all metal ion combinations (including Mg^{2+} or Mn^{2+} alone) except Zn^{2+} alone. Conversely, the final pH 6.0 DNAzyme, 8LD212, requires Zn^{2+} for its activity, including substantial yield with Zn^{2+} alone.

In contrast to the findings with the pH 6.0 DNAzymes, all 17 of the pH 7.5 DNAzymes require Zn^{2+} for their activity. Eleven of these DNAzymes, including three of four in Figure 7B, work well with all three of Mg^{2+}, Mn^{2+} and Zn^{2+}, or with Mn^{2+} and Zn^{2+} in the absence of Mg^{2+}, but no other metal combination. The other six of these DNAzymes, including the remaining example (8LF239) in Figure 7B, have high activity with all three of Mg^{2+}, Mn^{2+} and Zn^{2+}, or with either Mn^{2+}/Zn^{2+} or Mg^{2+}/Zn^{2+}. One of these six DNAzymes, 8LF226, also has considerable yield with Zn^{2+} alone.

Mass spectrometry analysis of DNAzyme products

The product from each of the eight representative DNAzymes of Figure 6 was prepared on the 300 pmol scale and assayed by MALDI-TOF mass spectrometry. In all cases, the observed product mass was consistent with the value calculated for reductive amination between the DNA substrate as amine nucleophile and the 5′-benzaldehyde oligonucleotide as electrophile (Supplementary Table S5).

Evaluating the DNAzymes with RNA instead of DNA substrates

DNAzymes that modify RNA rather than DNA will likely have considerable practical importance. Therefore, all 29 of the DNAzymes that have >30% yield were assayed with an RNA substrate of the same sequence in place of the original DNA substrate. This swap of RNA for DNA adds a 2′-hydroxyl group on each substrate nucleotide and also converts the substrate:DNAzyme Watson-Crick duplex (Figure 2A) from B-form DNA:DNA to A-form RNA:DNA. Any DNAzyme that still functions with the RNA substrate must be able to tolerate all of these structural changes. We found that seven DNAzymes (e.g. 8LF239 and 8LH201 in Figure 8; see Supplementary Table S6 for all yields), maintain >60% yield when presented with an RNA substrate. Six additional DNAzymes have substantial activity with RNA in place of DNA, although with lower yield (e.g. 6LA230 in Figure 8).

Assaying DNAzymes with small-molecule aldehyde reaction partners

In our selection design, the aldehyde reaction partner has its aldehyde component tethered to an oligonucleotide (Figure 2A). This design has two key advantages. First, the oligonucleotide portion is bound to the DNAzyme by preprogrammed Watson-Crick base pairs. This means that the initially random (N_{40}) region does not need to interact with the aldehyde reaction partner entirely by noncanonical interactions, as would be required if the aldehyde component were not attached to an oligonucleotide. Second, the selection process is streamlined because each functional DNA sequence is directly PAGE-shift separable owing to the large added mass of the oligonucleotide. However, for practical application of the DNAzymes to modify a long DNA or RNA substrate, using an untethered small-molecule aldehyde reaction partner would be preferred. We therefore assayed all 29 DNAzymes that have >30% yield by testing them with the DNA substrate and various small-molecule benzaldehyde derivatives, where the covalent tether to the benzaldehyde was omitted. We found four DNAzymes that function detectably (Figure 9). These assays included a range of substituents on the small-molecule benzaldehydes, but no linear free energy relationships were evident.

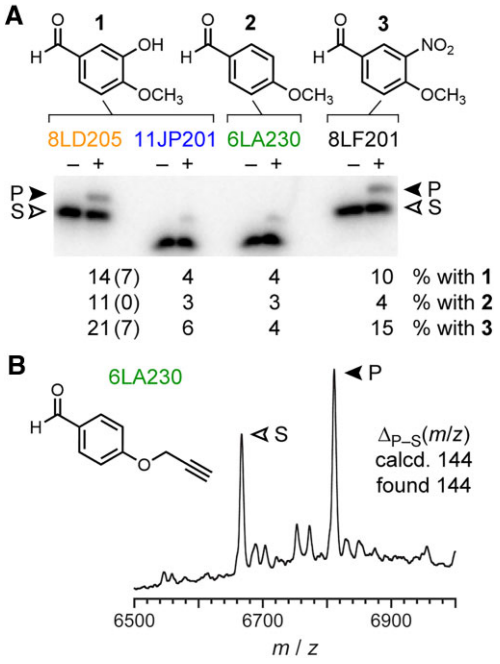


Figure 9. Several of the DNAzymes can use a small-molecule benzaldehyde compound as the reaction partner in place of the 5′-benzaldehyde oligonucleotide. (**A**) PAGE assay of four DNAzymes that work with small-molecule benzaldehydes (–/+ benzaldehyde; S = substrate, P = product; t = 48 h). To enable clear PAGE separation of S and P, the DNA substrate was shortened by eight nucleotides at its 5′-end (far away from the modification site), thus providing a 14 nt substrate. The calculated T_m (IDT OligoAnalyzer) is reduced from 57.7°C to 43.3°C, which may lead to lower yield than with the parent 22 nt substrate because of poorer binding to the DNAzyme. The benzaldehyde group of the 5′-benzaldehyde oligonucleotide used in selection is connected to the oligonucleotide via a para oxygen (Supplementary Figure S2). The assays included the unmodified DNA oligonucleotide that was originally attached to the benzaldehyde group. All of the DNAzymes were also assayed without this oligonucleotide. Only 8LD205 had any yield, observed with two of the three small-molecule benzaldehydes (% values in parentheses). (**B**) MALDI-TOF mass spectrometry of the 6LA230 product with 4-(propargyloxy)benzaldehyde after PAGE separation. Due to the small migration difference, some S was isolated along with P. m/z values $[M + H]^+$ S calcd. 6664.4, found 6667.0 (Δ = +0.04%). P calcd. 6808.6, found 6811.2 (Δ = +0.04%).

All of the above assays were conducted in the presence of the unmodified DNA oligonucleotide that was formerly connected to the benzaldehyde moiety, but now lacking the aldehyde group. This was done because the DNAzymes are likely to require a DNA duplex in this location; all of the DNAzymes were identified by selection with a duplex there (Figure 2A). Indeed, in the absence of this DNA oligonucleotide, three of the four DNAzymes were inactive with small-molecule benzaldehydes, although the remaining DNAzyme, 8LD205, had some activity.

Substrate sequence requirements of DNAzymes

For all eight representative DNAzymes of Figure 6, we performed a preliminary assessment of their substrate sequence requirements. We conceptually divided the DNA substrate into two portions: the 3′-overhang, and the longer segment that binds to the DNAzyme by Watson-Crick base pairs. Each trinucleotide 3′-overhang can have $3^3 = 9$ single mu-

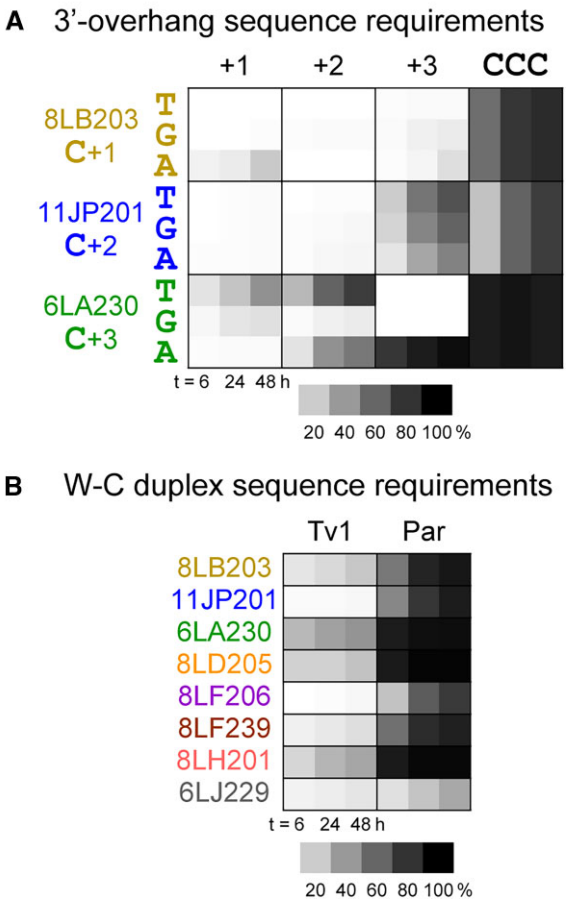


Figure 10. Preliminary assessment of DNAzyme substrate sequence requirements for the eight representative DNAzymes of Figure 6. For each assay, each horizontal set of three squares has data for $t = 6$, 24, 48 h under the DNAzyme's selection conditions. Percent yields are shown as the corresponding percent greyscale (0–100%). (**A**) Assays of the three CCC DNAzymes with systematic single-nucleotide mutations in the 3′-overhang portion of the DNA oligonucleotide substrate. Data with the parent CCC 3′-overhang is also shown. See Supplementary Figure S10 for analogous assays for the three GGG DNAzymes and two AAA DNAzymes. (**B**) Assays with systematic transversions (Tv1: A↔C, G↔T) in most of the Watson-Crick binding portion of the DNA substrate, along with compensatory changes in the DNAzyme binding arm. Reference data with the parent (Par) substrate is also shown.

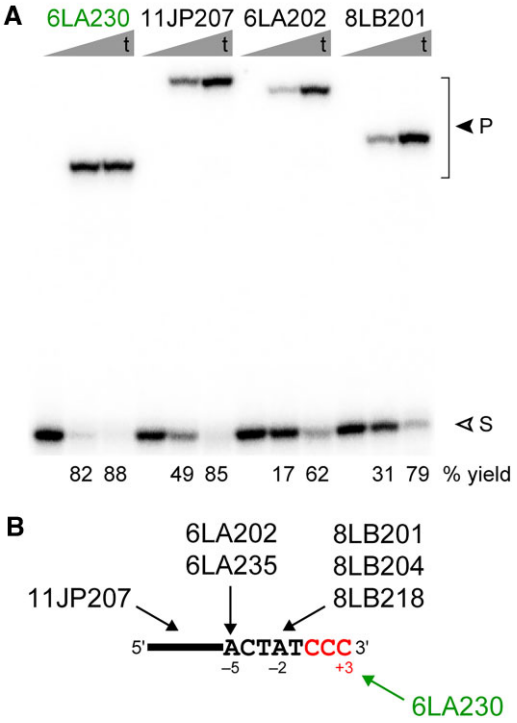


Figure 11. DNAzymes that N-alkylate a DNA substrate nucleotide within the duplex region formed by binding of the DNAzyme and substrate. (**A**) PAGE data for three of the six DNAzymes, under the DNAzyme's selection conditions with the 3′-CCC DNA substrate ($t = 30$ s, 6 h, 48 h). See Supplementary Figure S9 for yields for all six DNAzymes. (**B**) Nucleotide modification sites of these six DNAzymes, as described in the text.

tations. Experimentally, the eight DNAzymes responded in various ways to their corresponding nine mutations, resulting in a range of required sequence motifs. For the three CCC DNAzymes (Figure 10A), the respective sequence motifs for 8LB203, 11JP201 and 6LA230 can be summarized as mCC, CCN, and yhM, where the underlined position of the overhang is the modified nucleotide; M = A or C, N = any nucleotide, H = A, C or T, and Y = C or T; and lowercase denotes mutant yields that are substantially (more than twofold) lower than the parent CCC yield. Analogous assays for the three GGG DNAzymes and two AAA DNAzymes (Supplementary Figure S10) led to sequence motifs of 8LD205 GrG (R = G or A), 8LF206 GGG, 8LF239 GGG, 8LH201 MnA and 6LJ229 AAM. Taken together, these findings show that (i) most of the DNAzymes that alkylate C or A accept the other nucleotide of this pair at the modification position, consistent with the exocyclic amino group located on the upper face of both nucleobases; (ii) for each DNAzyme, the 3′-overhang nucleotide requirement encompasses more than just the nucleotide being modified; and (iii) the GGG DNAzymes are the least tolerant of 3′-overhang mutations, perhaps because only G has the exocyclic amino group on its lower face.

For the Watson-Crick duplex portion, we tested the eight representative DNAzymes with DNA substrates that have systematic transversion mutations (Tv1: A↔C, G↔T) at all positions except the two nucleotides closest to the 3′-overhang, which were kept unchanged. The compensatory Watson-Crick mutations were made in each tested DNAzyme's binding arm. Six out of eight DNAzymes showed clearly detectable activity with the heavily mutated substrate (Figure 10B); the two best instances are 6LA230 and 8LH201, with respective yields of 42% and 35% in 48 h.

DNAzymes that form branches within the duplex binding arm region

In addition to the above-described 34 DNAzymes, six other DNAzymes were found to modify a DNA substrate nucleotide that is located to the 5′-side of the 3′-overhang, within the duplex region where the DNAzyme binds to the DNA substrate (Figure 11; see DNAzyme sequences in Supplementary Figure S4). These DNAzymes emerged from three of the selections, all of which used the CCC 3′-overhang: two DNAzymes from the pH 6.0 selection (6LA202 and 6LA235), one DNAzyme from the pH 6.0 selection that included 5-MA (11JP207), and three DNAzymes from the pH 7.5 selection (8LB201, 8LB204, 8LB218). The products from these six DNAzymes migrate higher on PAGE than do the products from all 34 of the above-

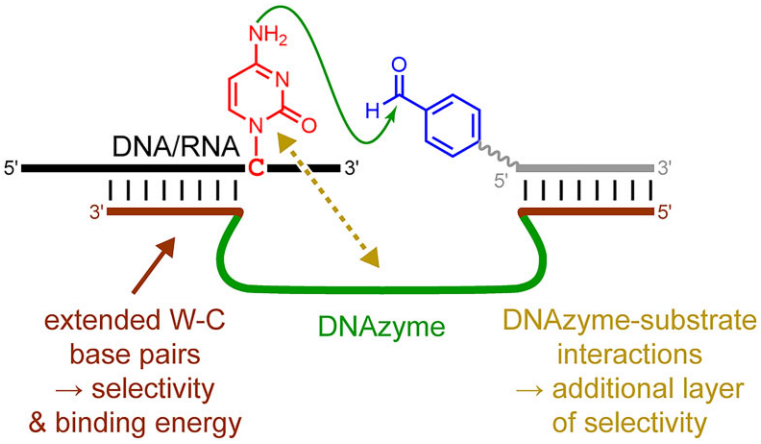


Figure 12. Model for the origin of site selectivity by nucleobase N-alkylation DNAzymes.

described new DNAzymes, revealing the relatively strongly branched nature of these products. The particular branch sites within the DNA substrate were assigned on the basis of assays like those in Figure 5, with data in Supplementary Figure S11. 11JP207 modifies an undetermined nucleotide to the 5′-side of the fifth nucleotide from the end of the duplex, A–5; both 6LA2 DNAzymes modify nucleotide A–5; and all three 8LB2 DNAzymes modify nucleotide A–2. The PAGE migration positions of the products in Figure 11 are consistent with these assignments, similar to the earlier results in Figure 4. This ability of DNAzymes to N-alkylate a nucleobase within the DNAzyme-substrate duplex region is reminiscent of our previously described DNAzymes that form branched RNA by using an internal RNA 2′-hydroxyl group as the nucleophile to attack the RNA 5′-triphosphate group of a reaction partner (52,60).

Five of these six DNAzymes have a metal ion dependence that is the same as one of the patterns found for the DNAzymes that modify the 3′-overhang (Supplementary Figure S9). 6LA202 and 6LA235 work well with Mn^{2+} alone; 11JP207 has activity with all combinations except Zn^{2+} alone; 8LB201 functions with all three ions as well as Mn^{2+}/Zn^{2+}; and 8LB204 needs all three metal ions or either Mn^{2+}/Zn^{2+} or Mg^{2+}/Zn^{2+}. The sixth DNAzyme, 8LB218, works best with either all three metal ions or Mg^{2+}/Zn^{2+}. Of these six DNAzymes, 8LB201 works well using an RNA substrate in place of DNA, and 8LB204 functions with RNA but with a several-fold decrease in the 48 h yield (Supplementary Figure S6). The ability of these two DNAzymes to modify RNA substrates is remarkable, given their modification sites within the duplex binding arm region and the helicity change of that duplex from B-form DNA:DNA to A-form RNA:RNA.

Discussion

In this study, we used *in vitro* selection to identify DNAzymes that N-alkylate nucleobases in DNA oligonucleotides by reductive amination (Figure 2 and Figure 3). Many of the new DNAzymes are >95% site selective for a particular nucleotide within the DNA substrate (Figure 4 and Figure 5), with k_{obs} up to 0.3 h^{-1}, rate enhancement up to ~10^4, and yield as high as 95% (Figure 6). A key advance is the identification of DNAzymes that achieve top-down, site-specific modification of DNA oligonucleotide nucleobases by N-alkylation. Our model for the origin of the DNAzyme site selectivity is in Figure 12. Extended Watson-Crick base pairs between the DNA or RNA oligonucleotide substrate and the DNAzyme binding arm localize the DNAzyme onto the substrate. This base pairing provides both the binding energy for the DNAzyme-substrate interactions and the selectivity basis for the DNAzyme to bind one sequence element among a potentially much larger number of similar elements within a long substrate. In parallel, the segment of the DNAzyme that emerged from the selection process (i.e. the original N_{40} region) can interact with the reactive nucleotide in the substrate 3′-overhang, and possibly the surrounding nonreactive nucleotides as well, providing an additional layer of selectivity for the reaction site.

The nucleobase N-alkylation reactions require the $NaCNBH_3$ reducing agent, likely in an uncatalyzed step after DNAzyme-catalyzed imine formation because other reducing agents can replace $NaCNBH_3$ (Supplementary Figure S7). A considerable variety of divalent metal ion requirements were found for the DNAzymes (Figure 7 and Supplementary Figure S9), and some key themes emerged. All of the selection experiments used a combination of Mg^{2+}, Mn^{2+} and Zn^{2+}, and the resulting DNAzymes unsurprisingly have maximal yield in the presence of all three metal ions together. For most of the DNAzymes identified at pH 6.0, Mn^{2+} alone is sufficient for their maximal yield, whereas nearly all of the pH 7.5 DNAzymes require Zn^{2+} and one of the other two metal ions. Understanding these interesting outcomes will require more detailed biochemical and structural characterizations. Many of the DNAzymes can use an RNA oligonucleotide substrate in place of DNA, despite the presence of 2′-hydroxyl groups and the helical difference (A-form rather than B-form) in the substrate:DNAzyme binding duplex (Figure 8). Four of the DNAzymes can function with a small-molecule benzaldehyde compound provided in place of the 5′-benzaldehyde oligonucleotide as the reductive amination reaction partner (Figure 9). One DNAzyme, 6LA230, can work separately with either an RNA substrate or a small-molecule benzaldehyde, although no activity was found for this DNAzyme when RNA was used simultaneously with small-molecule benzaldehydes.

The extended Watson-Crick base pairs of Figure 12 ideally have minimal sequence requirements, such that a spe-

cific segment of any long DNA or RNA substrate can be targeted simply by choosing the DNAzyme binding arm to have sequence complementarity. Our considerable prior work on sequence-general RNA ligase DNAzymes (61,62) and DNA-hydrolyzing DNAzymes (63) supports this expectation, as do at least two of the nucleobase N-alkylation DNAzymes from the current study (Figure 10B). The detailed DNAzyme-substrate interactions at and near the reactive nucleotide in the 3′-overhang are not preprogrammed into the selection process. The expectation, again supported by our prior work for other DNAzymes (52,60), is that these interactions can foster selectivity for a particular reactive nucleotide while allowing substantial generality at other 3′-overhang sequence positions. Our current findings indicate a range of sequence requirements at these nonreactive 3′-overhang nucleotides, including several nucleotides that have considerable generality (Figure 10A). Future selection and reselection (i.e. directed evolution) experiments can readily impose a pressure for generality at these nucleotides, by changing their identities in successive selection rounds. Extending the 3′-overhang past the first three nucleotides was tolerated well by nearly all of the DNAzymes, noting the high activity of most of the 29 DNAzymes in Figure 4, where a T_6 3′-extension was included. For comparison, many RNA ligase DNAzymes also tolerate long 3′-extensions (61,62).

One interesting practical question is whether the simplest possible aldehyde, formaldehyde, can be used by DNAzymes as the reductive amination reaction partner, thereby enabling N-methylation of oligonucleotide nucleobases (1–3). The commercial availability of several N^6-methyladenosine antibodies, partnered with the ability to perform *in vitro* selection that relies on antibody binding rather than PAGE shift for enrichment (64), suggests that DNAzymes for nucleobase N-methylation should be identifiable in this way. Independently, a ribozyme was identified for N^1-A methylation by direct methyl transfer from O^6-methylguanine (65), and several other ribozymes for various nucleobase alkylation reactions have been reported (66–72).

In future studies, we will expand our investigations of RNA rather than DNA substrates and small-molecule reaction partners, both by reselection of the newly found DNAzymes and by entirely new *in vitro* selections. Our ultimate goal is to identify sets of DNAzymes for site-specific nucleobase N-alkylation of DNA and RNA, using small-molecule reaction partners that are broadly applicable for introducing biochemical modifications and biophysical labels.

Data availability

The data underlying this article are available in the article and in its online Supplementary Material.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank Shukun Yang and Prakriti Das for helpful discussions and Clair Colee and Abhinav Masih for technical assistance.

Funding

Stephen J. Elledge Scholarship for Undergraduate Research [to M.M.K.]; Snyder Scholar fellowship [to A.A.M]; National Institutes of Health [R01GM149566 to S.K.S.]. The open access publication charge for this paper has been waived by Oxford University Press - *NAR* Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.

Conflict of interest statement

None declared.

References

1. Hu,L., Liu,S., Peng,Y., Ge,R., Su,R., Senevirathne,C., Harada,B.T., Dai,Q., Wei,J., Zhang,L., *et al.* (2022) m^6A RNA modifications are measured at single-base resolution across the mammalian transcriptome. *Nat. Biotechnol.*, **40**, 1210–1219.
2. Chung Kim Chung,K., Mahdavi-Amiri,Y., Korfmann,C. and Hili,R. (2022) PhOxi-Seq: single-nucleotide resolution sequencing of N^2-methylation at guanosine in RNA by photoredox catalysis. *J. Am. Chem. Soc.*, **144**, 5723–5727.
3. Franco,M.K. and Koutmou,K.S. (2022) Chemical modifications to mRNA nucleobases impact translation elongation and termination. *Biophys. Chem.*, **285**, 106780.
4. Arango,D., Sturgill,D., Alhusaini,N., Dillman,A.A., Sweet,T.J., Hanson,G., Hosogane,M., Sinclair,W.R., Nanan,K.K., Mandler,M.D., *et al.* (2018) Acetylation of cytidine in mRNA promotes translation efficiency. *Cell*, **175**, 1872–1886.
5. Arango,D., Sturgill,D., Yang,R., Kanai,T., Bauer,P., Roy,J., Wang,Z., Hosogane,M., Schiffers,S. and Oberdoerffer,S. (2022) Direct epitranscriptomic regulation of mammalian translation initiation through N4-acetylcytidine. *Mol. Cell*, **82**, 2797–2814.
6. Nance,K.D., Gamage,S.T., Alam,M.M., Yang,A., Levy,M.J., Link,C.N., Florens,L., Washburn,M.P., Gu,S., Oppenheim,J.J., *et al.* (2022) Cytidine acetylation yields a hypoinflammatory synthetic messenger RNA. *Cell Chem. Biol.*, **29**, 312–320.
7. Carell,T., Brandmayr,C., Hienzsch,A., Müller,M., Pearson,D., Reiter,V., Thoma,I., Thumbs,P. and Wagner,M. (2012) Structure and function of noncanonical nucleobases. *Angew. Chem. Int. Ed.*, **51**, 7110–7131.
8. Wiener,D. and Schwartz,S. (2021) The epitranscriptome beyond m^6A. *Nat. Rev. Genet.*, **22**, 119–131.
9. Suzuki,T. (2021) The expanding world of tRNA modifications and their disease relevance. *Nat. Rev. Mol. Cell Biol.*, **22**, 375–392.
10. Sun,H., Li,K., Liu,C. and Yi,C. (2023) Regulation and functions of non-m^6A mRNA modifications. *Nat. Rev. Mol. Cell Biol.*, **24**, 714–731.
11. Qin,P.Z. and Pyle,A.M. (1999) Site-specific labeling of RNA with fluorophores and other structural probes. *Methods*, **18**, 60–70.
12. Merkel,M., Peewasan,K., Arndt,S., Ploschik,D. and Wagenknecht,H.-A. (2015) Copper-free postsynthetic labeling of nucleic acids by means of bioorthogonal reactions. *ChemBioChem*, **16**, 1541–1553.
13. Ivancová,I., Leone,D.-L. and Hocek,M. (2019) Reactive modifications of DNA nucleobases for labelling, bioconjugations, and cross-linking. *Curr. Opin. Chem. Biol.*, **52**, 136–144.
14. Klöcker,N., Weissenboeck,F.P. and Rentmeister,A. (2020) Covalent labeling of nucleic acids. *Chem. Soc. Rev.*, **49**, 8749–8773.
15. Xiao,L., Habibian,M. and Kool,E.T. (2020) Site-selective RNA functionalization via DNA-induced structure. *J. Am. Chem. Soc.*, **142**, 16357–16363.
16. Flamme,M., McKenzie,L.K., Sarac,I. and Hollenstein,M. (2019) Chemical methods for the modification of RNA. *Methods*, **161**, 64–82.

17. Glazier,D.A., Liao,J., Roberts,B.L., Li,X., Yang,K., Stevens,C.M. and Tang,W. (2020) Chemical synthesis and biological application of modified oligonucleotides. *Bioconjug. Chem.*, **31**, 1213–1233.
18. Bartee,D., Nance,K.D. and Meier,J.L. (2022) Site-specific synthesis of N^4-acetylcytidine in RNA reveals physiological duplex stabilization. *J. Am. Chem. Soc.*, **144**, 3487–3496.
19. Flemmich,L., Bereiter,R. and Micura,R. (2024) Chemical synthesis of modified RNA. *Angew. Chem. Int. Ed.*, **63**, e202403063.
20. Moore,M.J. and Sharp,P.A. (1992) Site-specific modification of pre-mRNA: the 2′-hydroxyl groups at the splice Site. *Science*, **256**, 992–997.
21. Depmeier,H., Hoffmann,E., Bornewasser,L. and Kath-Schorr,S. (2021) Strategies for covalent labeling of long RNAs. *ChemBioChem*, **22**, 2826–2847.
22. Hertler,J., Slama,K., Schober,B., Ozrendeci,Z., Marchand,V., Motorin,Y. and Helm,M. (2022) Synthesis of point-modified mRNA. *Nucleic Acids Res.*, **50**, e115.
23. Keyhani,S., Goldau,T., Blümler,A., Heckel,A. and Schwalbe,H. (2018) Chemo-enzymatic synthesis of position-specifically modified RNA for biophysical studies including light control and NMR spectroscopy. *Angew. Chem. Int. Ed.*, **57**, 12017–12021.
24. Blümler,A., Schwalbe,H. and Heckel,A. (2022) Solid-phase-supported chemoenzymatic synthesis of a light-activatable tRNA derivative. *Angew. Chem. Int. Ed.*, **61**, e202111613.
25. Sudakov,A., Knezic,B., Hengesbach,M., Fürtig,B., Stirnal,E. and Schwalbe,H. (2023) Site-specific labeling of RNAs with modified and ^{19}F-labeled nucleotides by chemo-enzymatic synthesis. *Chem. Eur. J.*, **29**, e202203368.
26. Sasaki,S., Onizuka,K. and Taniguchi,Y. (2011) The oligodeoxynucleotide probes for the site-specific modification of RNA. *Chem. Soc. Rev.*, **40**, 5698–5706.
27. Jahn,K., Olsen,E.M., Nielsen,M.M., Tørring,T., MohammadZadegan,R., Andersen,E.S., Gothelf,K.V. and Kjems,J. (2011) Site-specific chemical labeling of long RNA molecules. *Bioconjug. Chem.*, **22**, 95–100.
28. Lee,Y.-H., Yu,E. and Park,C.-M. (2021) Programmable site-selective labeling of oligonucleotides based on carbene catalysis. *Nat. Commun.*, **12**, 1681.
29. Deen,J., Vranken,C., Leen,V., Neely,R.K., Janssen,K.P.F. and Hofkens,J. (2017) Methyltransferase-directed labeling of biomolecules and its applications. *Angew. Chem. Int. Ed.*, **56**, 5182–5200.
30. Zhou,C.Y., Alexander,S.C. and Devaraj,N.K. (2017) Fluorescent turn-on probes for wash-free mRNA imaging via covalent site-specific enzymatic labeling. *Chem. Sci.*, **8**, 7169–7173.
31. Tomkuvienė,M., Mickutė,M., Vilkaitis,G. and Klimasauskas,S. (2019) Repurposing enzymatic transferase reactions for targeted labeling and analysis of DNA and RNA. *Curr. Opin. Biotechnol.*, **55**, 114–123.
32. Ghaem Maghami,M., Scheitl,C.P.M. and Höbartner,C. (2019) Direct *in vitro* selection of *trans*-acting ribozymes for posttranscriptional, site-specific, and covalent fluorescent labeling of RNA. *J. Am. Chem. Soc.*, **141**, 19546–19549.
33. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
34. Ellington,A.D. and Szostak,J.W. (1990) *In vitro* selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
35. Robertson,D.L. and Joyce,G.F. (1990) Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, **344**, 467–468.
36. Joyce,G.F. (2004) Directed evolution of nucleic acid enzymes. *Annu. Rev. Biochem.*, **73**, 791–836.
37. Joyce,G.F. (2007) Forty years of in vitro evolution. *Angew. Chem. Int. Ed.*, **46**, 6420–6436.
38. Schlosser,K. and Li,Y. (2009) Biologically inspired synthetic enzymes made from DNA. *Chem. Biol.*, **16**, 311–322.
39. Liu,J., Cao,Z. and Lu,Y. (2009) Functional nucleic acid sensors. *Chem. Rev.*, **109**, 1948–1998.
40. Hollenstein,M. (2015) DNA catalysis: the chemical repertoire of DNAzymes. *Molecules*, **20**, 20777–20804.
41. Silverman,S.K. (2016) Catalytic DNA: scope, applications, and biochemistry of deoxyribozymes. *Trends Biochem. Sci.*, **41**, 595–609.
42. Liu,M., Chang,D. and Li,Y. (2017) Discovery and biosensing applications of diverse RNA-cleaving DNAzymes. *Acc. Chem. Res.*, **50**, 2273–2283.
43. Morrison,D., Rothenbroker,M. and Li,Y. (2018) DNAzymes: selected for applications. *Small Methods*, **2**, 1700319.
44. Hollenstein,M. (2019) Nucleic acid enzymes based on functionalized nucleosides. *Curr. Opin. Chem. Biol.*, **52**, 93–101.
45. Cepeda-Plaza,M. and Peracchi,A. (2020) Insights into DNA catalysis from structural and functional studies of the 8-17 DNAzyme. *Org. Biomol. Chem.*, **18**, 1697–1709.
46. Ma,L. and Liu,J. (2020) Catalytic nucleic acids: biochemistry, chemical biology, biosensors, and nanotechnology. *iScience*, **23**, 100815.
47. Yang,S. and Silverman,S.K. (2023) Defining the substrate scope of DNAzyme catalysis for reductive amination with aliphatic amines. *Org. Biomol. Chem.*, **21**, 1910–1919.
48. Wong,O., Mulcrone,A.E. and Silverman,S.K. (2011) DNA-catalyzed reductive amination. *Angew. Chem. Int. Ed.*, **50**, 11679–11684.
49. Desai,U.J. and Pfaffle,P.K. (1995) Single-step purification of a thermostable DNA polymerase expressed in *Escherichia coli*. *BioTechniques*, **19**, 780–784.
50. Crisalli,P. and Kool,E.T. (2013) Water-soluble organocatalysts for hydrazone and oxime formation. *J. Org. Chem.*, **78**, 1184–1189.
51. Tuschl,T., Sharp,P.A. and Bartel,D.P. (1998) Selection *in vitro* of novel ribozymes from a partially randomized U2 and U6 snRNA library. *EMBO J.*, **17**, 2637–2650.
52. Wang,Y. and Silverman,S.K. (2003) Deoxyribozymes that synthesize branched and lariat RNA. *J. Am. Chem. Soc.*, **125**, 6880–6881.
53. Cabacungan,J.C., Ahmed,A.I. and Feeney,R.E. (1982) Amine boranes as alternative reducing agents for reductive alkylation of proteins. *Anal. Biochem.*, **124**, 272–278.
54. Ruhaak,L.R., Steenvoorden,E., Koeleman,C.A., Deelder,A.M. and Wuhrer,M. (2010) 2-Picoline-borane: a non-toxic reducing agent for oligosaccharide labeling by reductive amination. *Proteomics*, **10**, 2330–2336.
55. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
56. Ponce-Salvatierra,A., Wawrzyniak-Turek,K., Steuerwald,U., Höbartner,C. and Pena,V. (2016) Crystal structure of a DNA catalyst. *Nature*, **529**, 231–234.
57. Liu,H., Yu,X., Chen,Y., Zhang,J., Wu,B., Zheng,L., Haruehanroengra,P., Wang,R., Li,S., Lin,J., *et al.* (2017) Crystal structure of an RNA-cleaving DNAzyme. *Nat. Commun.*, **8**, 2006.
58. Borggräfe,J., Victor,J., Rosenbach,H., Viegas,A., Gertzen,C.G.W., Wuebben,C., Kovacs,H., Gopalswamy,M., Riesner,D., Steger,G., *et al.* (2022) Time-resolved structural analysis of an RNA-cleaving DNA catalyst. *Nature*, **601**, 144–149.
59. Wieruszewska,J., Pawłowicz,A., Połomska,E., Pasternak,K., Gdaniec,Z. and Andrałojć,W. (2024) The 8-17 DNAzyme can operate in a single active structure regardless of metal ion cofactor. *Nat. Commun.*, **15**, 4218.
60. Wang,Y. and Silverman,S.K. (2005) Efficient one-step synthesis of biologically related lariat RNAs by a deoxyribozyme. *Angew. Chem. Int. Ed.*, **44**, 5863–5866.
61. Purtha,W.E., Coppins,R.L., Smalley,M.K. and Silverman,S.K. (2005) General deoxyribozyme-catalyzed synthesis of native 3′–5′ RNA linkages. *J. Am. Chem. Soc.*, **127**, 13124–13125.
62. Pratico,E.D., Wang,Y. and Silverman,S.K. (2005) A deoxyribozyme that synthesizes 2′,5′-branched RNA with any branch-site nucleotide. *Nucleic Acids Res.*, **33**, 3503–3512.

63. Xiao,Y., Wehrmann,R.J., Ibrahim,N.A. and Silverman,S.K. (2012) Establishing broad generality of DNA catalysts for site-specific hydrolysis of single-stranded DNA. *Nucleic Acids Res.*, **40**, 1778–1786.
64. Walsh,S.M., Konecki,S.N. and Silverman,S.K. (2015) Identification of sequence-selective tyrosine kinase deoxyribozymes. *J. Mol. Evol.*, **81**, 218–224.
65. Scheitl,C.P.M., Ghaem Maghami,M., Lenz,A.-K. and Höbartner,C. (2020) Site-specific RNA methylation by a methyltransferase ribozyme. *Nature*, **587**, 663–667.
66. Wilson,C. and Szostak,J.W. (1995) *In vitro* evolution of a self-alkylating ribozyme. *Nature*, **374**, 777–782.
67. Sharma,A.K., Plant,J.J., Rangel,A.E., Meek,K.N., Anamisis,A.J., Hollien,J. and Heemstra,J.M. (2014) Fluorescent RNA labeling using self-alkylating ribozymes. *ACS Chem. Biol.*, **9**, 1680–1684.
68. Poudyal,R.R., Nguyen,P.D.M., Lokugamage,M.P., Callaway,M.K., Gavette,J.V., Krishnamurthy,R. and Burke,D.H. (2017) Nucleobase modification by an RNA enzyme. *Nucleic Acids Res.*, **45**, 1345–1354.
69. Flemmich,L., Heel,S., Moreno,S., Breuker,K. and Micura,R. (2021) A natural riboswitch scaffold with self-methylation activity. *Nat. Commun.*, **12**, 3877.
70. Jiang,H., Gao,Y., Zhang,L., Chen,D., Gan,J. and Murchie,A.I.H. (2021) The identification and characterization of a selected SAM-dependent methyltransferase ribozyme that is present in natural sequences. *Nat. Catal.*, **4**, 872–881.
71. Scheitl,C.P.M., Okuda,T., Adelmann,J. and Höbartner,C. (2023) Ribozyme-catalyzed late-stage functionalization and fluorogenic labeling of RNA. *Angew. Chem. Int. Ed.*, **62**, e202305463.
72. Okuda,T., Lenz,A.K., Seitz,F., Vogel,J. and Höbartner,C. (2023) A SAM analogue-utilizing ribozyme for site-specific RNA alkylation in living cells. *Nat. Chem.*, **15**, 1523–1531.

Received: June 6, 2024. **Revised:** July 3, 2024. **Editorial Decision:** July 4, 2024. **Accepted:** July 8, 2024