

DNAzyme-mediated catalysis with only guanosine and cytidine nucleotides

Kenny Schlosser¹ and Yingfu Li^{1,2,*}

¹Department of Biochemistry and Biomedical Sciences and ²Department of Chemistry, McMaster University, Hamilton, Ontario, Canada L8N 3Z5

Received September 7, 2008; Revised November 3, 2008; Accepted November 4, 2008

ABSTRACT

Single-stranded DNA molecules have the capacity to adopt catalytically active structures known as DNAzymes, although the fundamental limits of this ability have not been determined. Starting with a parent DNAzyme composed of all four types of standard nucleotides, we conducted a search of the surrounding sequence space to identify functional derivatives with catalytic cores composed of only three, and subsequently only two types of nucleotides. We provide the first report of a DNAzyme that contains only guanosine and cytidine deoxyribonucleotides in its catalytic domain, which consists of just 13 nucleotides. This DNAzyme catalyzes the Mn²⁺-dependent cleavage of an RNA phosphodiester bond ~5300-fold faster than the corresponding uncatalyzed reaction, but ~10 000-fold slower than the parent. The demonstration of a catalytic DNA molecule made from a binary nucleotide alphabet broadens our understanding of the fundamental limits of nucleic-acid-mediated catalysis.

INTRODUCTION

DNA is typically characterized by the iconic double helix structure that was first proposed by Watson and Crick in 1953 (1). The stability and complementary nature of this structure is central to its well-known role in the storage and transfer of genetic information. However, single-stranded DNA can adopt many other types of structures, including some that are able to catalyze chemical transformations, not unlike more traditional enzymes made from protein. This latent catalytic ability was first demonstrated in 1994, when DNA molecules that catalyze the cleavage of RNA phosphodiester bonds, were selected *in vitro* from a library of random synthetic DNA sequences (2). In subsequent years many additional examples of catalytic DNA molecules (commonly known as DNAzymes or

deoxyribozymes) have been reported. These man-made DNAzymes can catalyze just over a dozen different types of reactions, with rate enhancements as high as 10¹⁰-fold (3). Research in the DNAzyme field is motivated by both the conceptual and practical significance, of an alternative catalytic platform that is less costly and more stable than protein or RNA catalysts.

Our laboratory is interested in understanding the fundamental limits of DNA-mediated catalysis. All existing DNAzymes have been assembled from all four standard nucleotides (or modified versions thereof). However, it remains to be determined if all four types of nucleotides represent an absolute prerequisite for DNA catalysis. In contrast, other catalytic biopolymers made from protein and RNA have been shown to be active with a simplified set of their monomeric building blocks. For instance, Hilvert and colleagues have shown that as few as nine different types of amino acids are sufficient for a functional chorismate mutase enzyme (4). Similarly, Joyce and colleagues have demonstrated that as few as two different types of ribonucleotides can encode a functional ribozyme that catalyzes RNA ligation (5). So far, analogous achievements with DNA have not been reported, which may reflect the inherent difficulty of this objective. The 2'-OH group in RNA provides an additional hydrogen-bond donor or acceptor that can aid in the formation of more complex, and potentially catalytic structures. Therefore, the absence of this functional group should make DNA less likely to achieve catalysis with a smaller nucleotide alphabet.

Herein, we sought to determine whether DNA also has sufficient structural and functional plasticity to mediate catalysis with less than four types of nucleotides. We chose a DNAzyme known as '8-17' to serve as a model system for this study (6). This DNAzyme catalyzes the site-specific cleavage of RNA phosphodiester bonds. The 8-17 DNAzyme contains a well-defined catalytic core measuring 14-15 nucleotides (nt) in length, which is flanked by two variable-sequence arms that can be tailored to bind to any desired substrate sequence through standard Watson-Crick base-pairing. When bound to a

*To whom correspondence should be addressed. Tel: +1 905 525 9140. Ext. 22462; Fax: +1 905 522 9033; Email: liying@mcmaster.ca

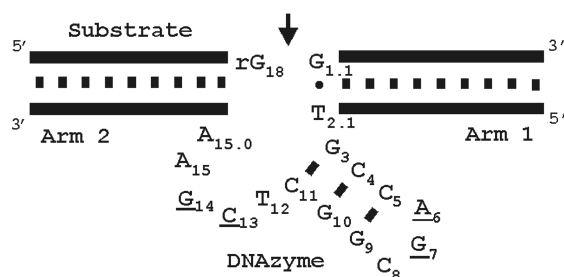


Figure 1. Sequence and secondary structure of an 8–17 variant. Only the sequence of the catalytic core domain is shown (represented by nucleotide positions 2.1 to 15.0), which is flanked by variable substrate-recognition arms (represented by lines) that can be changed to accommodate different substrate sequences. Underlined residues are highly conserved. Watson–Crick base-pairing is denoted by vertical or diagonal lines. The dinucleotide cleavage site is indicated by the arrow. rG: guanosine ribonucleotide.

cognate substrate, the 8–17 forms a three-way junction composed of the two peripheral substrate recognition arms and the central catalytic domain, which is typically characterized by a 3-bp stem-triloop and 4–5 nt single-stranded turn region (Figure 1).

The 8–17 DNAzyme contains all four types of nucleotides, but represented a good starting candidate because only three types of nucleotides at four positions in the catalytic core (i.e. positions A6, G7, C13 and G14) are highly conserved, as determined from hundreds of sequence variations that have been isolated from several *in vitro* selection experiments (6–12). The precise role of these highly conserved nucleotides has not been elucidated. Variation at the remaining positions in the catalytic core is tolerated, albeit with functional consequences for both the reaction rate and cleavage site selectivity (13). Our laboratory previously conducted a comprehensive mutational analysis on the catalytic core of an 8–17 variant, in which each nucleotide position was individually substituted with each of the remaining three types of nucleotides (13). We utilized this mutagenesis data to guide the semi-rational design of functional sequence derivatives composed of three, and subsequently just two, different types of nucleotides.

MATERIALS AND METHODS

Oligonucleotides and reagents

Oligonucleotides were prepared by automated DNA synthesis using cyanoethylphosphoramidite chemistry (Integrated DNA Technologies; Mobix Central Facility, McMaster University). Oligonucleotides were purified by 10% preparative denaturing (8 M urea) polyacrylamide gel electrophoresis (PAGE) and their concentrations were determined by spectroscopic methods. [γ - ^{32}P]-ATP was purchased from Perkin–Elmer. T4 polynucleotide kinase (PNK) was purchased from Fermentas. All chemical reagents were purchased from Sigma.

Kinetic analyses

All rate constants were determined under single turnover conditions, in which a large excess of DNAzyme

($\sim 1.7 \mu\text{M}$) was used with a trace amount of 5'- ^{32}P -labelled substrate ($\sim 0.002 \mu\text{M}$). Substrate and DNAzyme were heated together at 90°C for 30 s, and allowed to cool at room temperature for ~ 10 min. An equal volume of $2\times$ reaction buffer (200 mM KCl, 800 mM NaCl, 15 mM MgCl_2 , 15 mM MnCl_2 , 100 mM HEPES pH 7.0 at 23°C) was added to the DNAzyme/substrate mixture to initiate the reaction. The reaction was terminated after a designated period of time by the addition of quenching buffer containing 60 mM EDTA, 7 M urea and loading dye solution. The cleavage products from a reaction timecourse were separated by denaturing 10% PAGE, and quantitated using a PhosphorImager and ImageQuant software. A graph of fraction cleaved versus time (t) was plotted for each timecourse, and the experimental data fit to a single exponential equation $Y = Y_{\text{max}}[1 - e^{-(k_{\text{obs}}^* t)}]$ using non-linear regression analysis in GraphPad Prism 4, from which the observed rate constant (k_{obs}) and maximum cleavage yield (Y_{max}) were determined. Kinetic parameters were determined from at least two independent experiments that typically differed by less than 30%. Because of the low level of cleavage observed between DNAzyme GC and S2, the initial rate constant for this reaction was determined from the negative slope of the line produced by a least squares fit to a plot of the natural logarithm of the uncleaved fraction versus time.

RESULTS

DNAzymes composed of three different types of nucleotides

Figure 1 shows the sequence and secondary structure of an 8–17 DNAzyme (denoted as AGCT) that contains all four types of nucleotides in a catalytic core defined by nucleotide positions 2.1 to 15.0. It should be noted that the sequence shown in Figure 1 is one of many possible variations of the canonical 8–17 sequence first reported by Santoro and Joyce (6). This specific sequence variant was chosen as the starting parent molecule because it exhibits one of the highest known cleavage rates among 8–17 variants [and is phenotypically equivalent to the sequence variant used in our previous mutational study (13), but contains a higher proportion of G and C residues]. We designed 8–17 sequence variants (Figure 2A) to represent each of the four different combinations of three nucleotides (denoted as DNAzymes AGC, GCT, AGT and ACT according to the composition of their catalytic cores). The sequences were designed by replacing each relevant position in the parent molecule, with one of the remaining three types of nucleotides expected to have the least detrimental effect on activity. The activity of each DNAzyme was tested under single turnover conditions *in trans*, against a chimeric substrate containing a 5'-riboguanosine-deoxyriboguanosine-3' (rG-G) dinucleotide junction as the scissile cleavage site within an otherwise all-DNA oligonucleotide (Figure 1). Initially, we used a substrate (denoted as S1) that contained all four types of nucleotides (and the complementary sequence in the substrate-binding arms of the DNAzyme), and only restricted the DNAzyme catalytic core to three different nucleotides. We did not want to obscure any potential

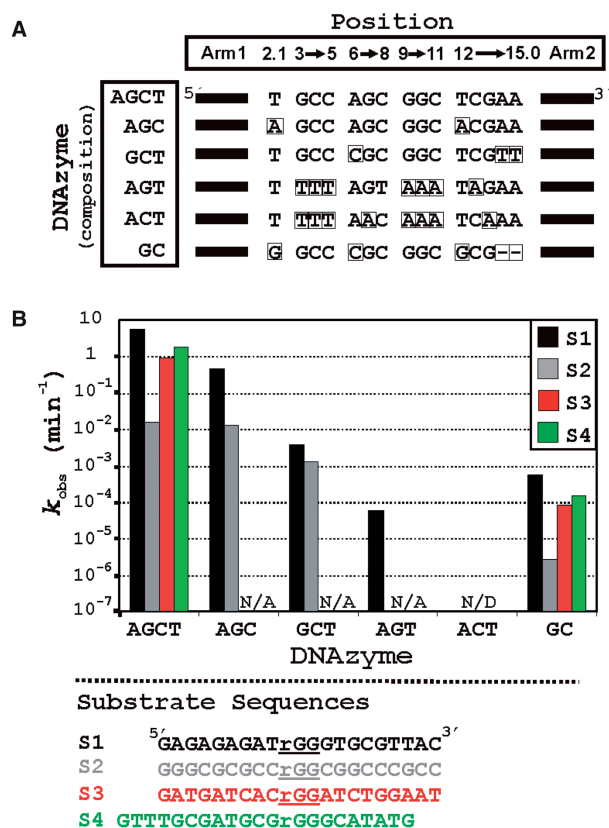


Figure 2. Comparison of activity between DNAzymes made from four, three or two different types of nucleotides. (A) Sequence derivatives of DNAzyme AGCT composed of three or two different types of nucleotides are shown. Mutations relative to the DNAzyme AGCT are highlighted by boxes. For clarity, only the sequence of the catalytic core is shown. (B) Cleavage rate constants for various DNAzymes against different substrate sequences, S1–S4. The dinucleotide cleavage site is underlined. DNAzymes AGC and GCT were not tested against substrates S3–S4 (N/A, not applicable). DNAzyme AGT was not tested against substrates S2–S4 (N/A, not applicable). DNAzyme ACT exhibited no detectable (N/D) activity against S1, and was not tested against substrates S2–S4. Reactions were conducted under single-turnover conditions, in the presence of 400 mM KCl, 100 mM NaCl, 7.5 mM MgCl₂, 7.5 mM MnCl₂ and 50 mM HEPES pH 7.0 at 23°C. Rate constants (k_{obs}) represent the average of at least two independent trials, which typically differed by <30%. Reaction endpoints (Y_{max}) ranged from ~80% to 90%, with the following exceptions; GC versus S2 = 4%, GC versus S4 = 49%. The background rate constant for cleavage under the reaction conditions is $\sim 10^{-7} \text{ min}^{-1}$. DNAzymes targeting substrates S1–S3 contained 9 nt + 9 nt binding arms, while 7 nt (arm 1) + 8 nt (arm 2) binding arms were used for DNAzymes targeting substrate S4.

activity due to misfolding, which we suspected would become problematic as the nucleotide alphabet was decreased. Activity assays were performed in the presence of 7.5 mM MgCl₂, 7.5 mM MnCl₂, 100 mM KCl, 400 mM NaCl and 50 mM HEPES pH 7.0 at 23°C. Under these conditions the rate constant of the uncatalyzed reaction is $\sim 10^{-7} \text{ min}^{-1}$ (14).

DNAzymes AGC, TGC and AGT exhibited detectable activity ranging from ~ 600 to 5×10^6 -fold over the uncatalyzed reaction (Figure 2B). The hierarchy of activity followed the order AGCT (5.6 min^{-1}) > AGC ($4.6 \times 10^{-1} \text{ min}^{-1}$) > GCT ($3.8 \times 10^{-3} \text{ min}^{-1}$) > AGT ($5.8 \times 10^{-5} \text{ min}^{-1}$) > ACT (ND, not detectable).

These results are consistent with expectations, given the importance of nucleotides A6, G7, C13 and G14 for 8–17-mediated catalytic activity (13,15). For the same reason, further attempts to design a DNAzyme composed of A, C and T nucleotides were abandoned. The maximum extent of substrate cleavage exhibited by each of the active DNAzymes was $\sim 90\%$.

Although the 8–17 DNAzyme engages its cognate substrate in a well-defined manner, we nevertheless wanted to eliminate the possibility that nucleotides in either the substrate, or substrate-recognition arms of the DNAzyme, may be essential for the observed activity. Therefore, a second substrate (denoted as S2) was designed to contain only G and C residues, and tested with DNAzymes AGC and TGC (whose binding arms were appropriately modified to maintain base-pairing with substrate S2). The rate constant for cleavage of substrate S2 by DNAzymes AGC and GCT was 1.3×10^{-2} and $1.3 \times 10^{-3} \text{ min}^{-1}$, respectively. Compared to the cleavage of S1, these values represent a decrease of ~ 38 - and ~ 3 -fold for DNAzymes AGC and GCT, respectively. DNAzyme AGCT was also tested against substrate S2, and exhibited a large decrease in activity (~ 358 -fold). The cleavage reaction mediated by these DNAzymes went to ~ 85 – 90% completion.

A DNAzyme composed of guanosine and cytosine residues

We next attempted to design DNAzymes composed of just guanosine and cytosine nucleotides. No attempt was made to engineer a DNAzyme containing only adenosine and thymidine, given the importance of nucleotides G7, C13 and G14 to 8–17-mediated catalysis as stated above (13,15). Initial activity assays were once again conducted with the S1 substrate. Various substitution mutants were designed and tested, however appreciable activity was only observed when these substitutions were combined with the deletion of residues 15 and 15.0 (Figure 2A). The resulting 13 nt sequence (denoted as GC) composed of 5.3 $\times 10^{-4} \text{ min}^{-1}$, and cleaves S1 nearly to completion ($\sim 90\%$) in 72 h. It should be noted that while the parent DNAzyme AGCT (or a related 8–17 variant) is capable of multiple turnover kinetics (16), DNAzyme GC is too slow to permit a practical evaluation of its multiple turnover parameters.

When tested against substrate S2, DNAzyme GC exhibited detectable but severely reduced activity ($\sim 2.7 \times 10^{-6} \text{ min}^{-1}$, or $\sim 4\%$ cleavage in 10 days). This low level of activity was not surprising given that a decrease in activity had also been observed with DNAzymes AGCT, AGC and GCT. Furthermore, a complete binary system (in which both DNAzyme and substrate are composed only of G and C residues) is expected to be highly susceptible to both intra and inter-molecular misfolding. In an effort to achieve a more active binary system, several different sequence variations of substrate S2, variations in substrate length and annealing conditions, and an intramolecular cis format were tested, but were unsuccessful.

To provide alternative evidence excluding the possibility that A or T nucleotides in either the substrate, or substrate-binding arms of the DNAzyme were essential for catalysis, we confirmed the activity of DNAzyme GC

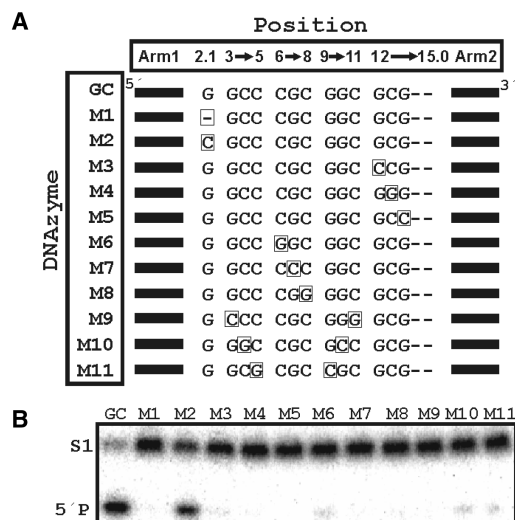


Figure 3. Mutational analysis of DNAzyme GC. (A) Mutations relative to DNAzyme GC are highlighted by boxes. (B) Phosphorimage scan showing cleavage activity of DNAzyme GC and related mutants M1–M11. A 5'-³²P-labelled substrate S1 was separately incubated with each DNAzyme for 72 h. The 5'-cleavage product (5'-P) was separated from uncleaved substrate by 10% denaturing PAGE.

against two completely new substrates (denoted as S3 and S4) containing all four types of nucleotides but scrambled in sequence (Figure 2B). This test also confirmed that the observed activity between DNAzyme GC and S1, was not due to accidental contamination with another DNAzyme. DNAzyme GC was indeed active against each substrate, albeit with a modest reduction in activity relative to S1 (~6.3-fold for S3 and ~3.7-fold for S4). However, when DNAzyme AGCT was subsequently tested against these substrates, it also exhibited approximately the same relative decrease in activity (~6.2-fold for S3 and ~3.1-fold for S4), suggesting that the peripherally located A and T residues of DNAzyme GC are not compensating for the absence of such residues in the catalytic core (Figure 2B).

Sequence requirements, mechanism, metal ion requirements and cleavage site versatility of DNAzyme GC

A focused mutational analysis of DNAzyme GC was conducted to gain some insight into its sequence requirements (Figure 3). Each nucleotide position was individually substituted with another type of nucleotide (i.e. either a G or C residue) to evaluate the effect on cleavage activity. Only a G2.1C substitution was tolerated (with a modest reduction in activity), while all other substitutions eliminated activity within a 72 h timeframe. Interestingly, compensatory substitutions to any position in the original stem region was not tolerated, which may suggest that this stem no longer forms as part of the active structure. However, it is also possible that these mutations may stabilize an alternative inactive conformation or simply perturb an important functional group.

The mechanism of DNAzyme GC was investigated by analyzing the cleavage products of substrate S4.

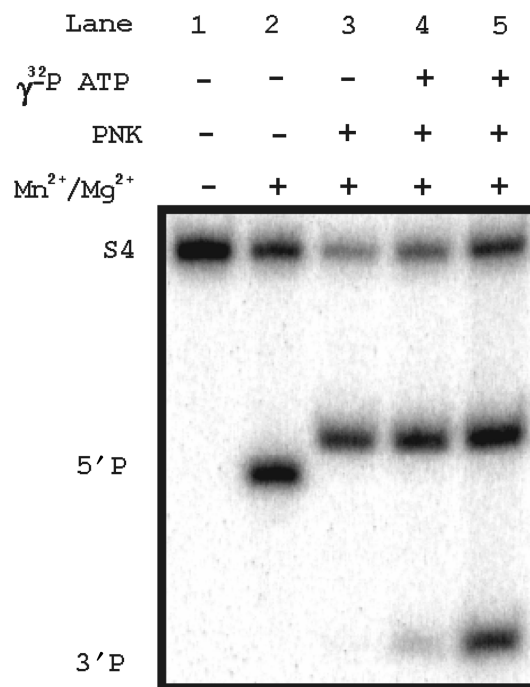


Figure 4. Analysis of cleavage products. A 5'-³²P-labelled substrate S4 was incubated with DNAzyme GC alone (lane 1) or in the presence of reaction buffer (lane 2). The reaction products were subsequently incubated in the presence of T4 PNK (lane 3) or T4 PNK and [³²P]-ATP (lanes 4 and 5). The products of lane 5 were incubated with 10-fold more [³²P]-ATP than lane 4. The 13 nt 5'-cleavage product (5'-P) and 8 nt 3'-cleavage product (3'-P) were separated from the uncleaved substrate by 20% denaturing PAGE. The reduced mobility of the 5'-cleavage product after treatment with PNK (lanes 3–5) is consistent with the removal of a 2',3'-cyclic phosphate. The appearance of the 3'-cleavage product after treatment with PNK and [³²P]-ATP (lanes 4–5), is consistent with the phosphorylation of a free 5'-hydroxyl group.

When incubated with T4 polynucleotide kinase (PNK) the electrophoretic mobility of the 5' cleavage product was reduced (Figure 4), consistent with the removal of a 2', 3' or 2',3'-cyclic phosphate. It should be noted that T4 PNK is known to have 2',3'-cyclic nucleotide 3'-phosphodiesterase and 2'- and 3'-phosphatase activities, in addition to its more commonly known kinase activity (17). The 3' cleavage product was also readily phosphorylated with [³²P]-ATP and T4 PNK, consistent with the presence of a free 5'-OH group (Figure 4). These cleavage products are indicative of a reaction mechanism involving nucleophilic attack of the 2'-hydroxyl group on the adjacent phosphorus center of the scissile phosphodiester bond. This is a common mechanism used by several RNA-cleaving DNAzymes including the parent (9), as well as some natural RNA-cleaving ribozymes (18).

To gain some additional insight into the reaction mechanism, the pH-dependent rate constant profiles of DNAzyme AGCT and GC were determined (Figure 5). Both DNAzymes exhibited a log-linear increase in rate constant with increasing pH (between pH 6.5 and 7.5), with a slope of ~1. Similar linear pH dependences have been reported previously for other 8–17 variants in the presence of various metal ions including Mg²⁺ and Ca²⁺ (16), Zn²⁺ (9), and Pb²⁺ (19). These results suggest

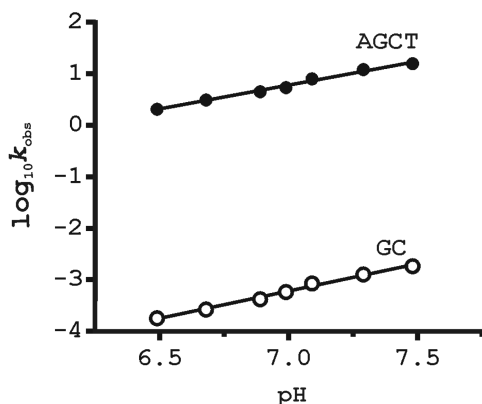


Figure 5. pH dependence of the rate constant of substrate S1 cleavage. k_{obs} values (min^{-1}) were determined under single-turnover conditions in the presence of 100 mM KCl, 400 mM NaCl, 7.5 mM MgCl_2 , 7.5 mM MnCl_2 and 50 mM MOPS pH 6.5–7.5 at 23°C. The data (represented by the average of at least two independent measurements) were fit log-linearly with slopes of 0.92 ± 0.04 ($R^2 = 0.99$) for DNAzyme AGCT (closed circles), and 1.06 ± 0.04 ($R^2 = 0.99$) for DNAzyme GC (open circles).

that the rate limiting step in the reaction catalyzed by DNAzyme GC remains the chemical cleavage step (as is the case with the parent DNAzyme AGCT), rather than a slow conformational change (20). This conclusion is also supported by the fact that burst kinetics were not observed in the reaction timecourse between DNAzyme GC and substrate S1 (Supplementary Figure S1).

The activity of DNAzyme GC was found to be largely dependent on Mn^{2+} ions, but also benefited from the presence of K^+ and Na^+ ions, as indicated in Figure 6. Divalent magnesium ions did not support wild-type activity, although a low level of activity was detectable over an extended period of time. This metal dependency is comparable to the parent DNAzyme, which also shows a preference for Mn^{2+} over Mg^{2+} ions (15,16).

The parent DNAzyme AGCT exhibits activity against all 16 possible combinations of dinucleotide cleavage junctions (13) (Figure 7A). The cleavage site versatility of DNAzyme GC was investigated using 16 versions of substrate S1 that differed only in the identity of the dinucleotide cleavage junction. In addition to the original 5'-GG dinucleotide junction, DNAzyme GC can cleave a 5'-AG junction approximately 2-fold faster with a rate constant of $1.2 \times 10^{-3} \text{ min}^{-1}$ ($Y_{\text{max}} = 93\%$). The remaining 14 dinucleotide junctions were not susceptible to cleavage within a 72 h timeframe (Figure 7B). The cleavage site specificity of DNAzyme GC was confirmed using an all-RNA version of substrate S1 (Supplementary Figure S2). DNAzyme GC cleaved this RNA substrate at a single location at the expected GG dinucleotide junction. However, the all-RNA substrate was cleaved ~11-fold slower than the chimeric substrate (data not shown). This discrepancy in the cleavage efficiency between chimeric and all-RNA substrates has been reported previously with other 8–17 and 10–23 DNAzymes (9,21). It has been suggested that the higher efficiency with which the chimeric substrate is cleaved, is a reflection of the higher amount of B-form like helix that can form between the DNAzyme and chimeric

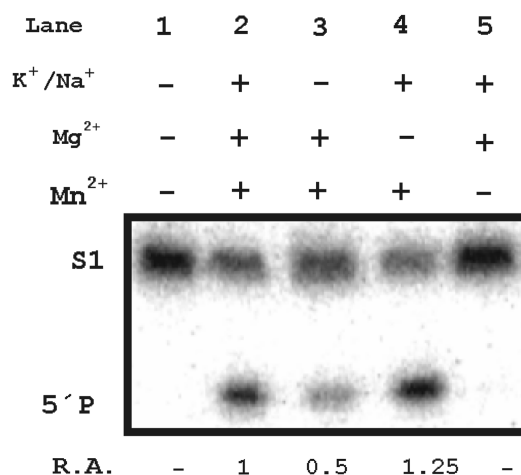


Figure 6. Analysis of metal ion requirements. A 5'- ^{32}P -labelled substrate S1 was incubated with DNAzyme GC for 24 h in the presence (+) or absence (-) of various metal ions from the original reaction buffer (400 mM KCl, 100 mM NaCl, 7.5 mM MgCl_2 , 7.5 mM MnCl_2 and 50 mM HEPES pH 7.0 at 23°C) as indicated. Note: the MnCl_2 concentration in lane 4 was supplemented to 15 mM, and the MgCl_2 concentration in lane 5 was supplemented to 15 mM. Reaction products were separated by 10% denaturing PAGE. R.A. denotes relative activity.

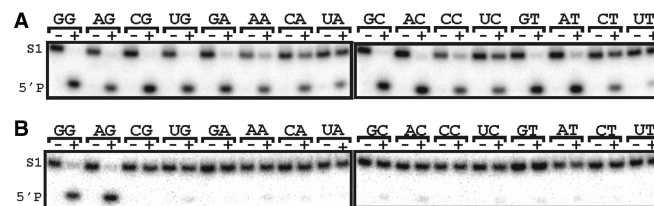


Figure 7. Cleavage site versatility of DNAzyme AGCT versus GC. (A) Sixteen versions of 5'- ^{32}P -labelled S1 (differing only in the identity of the dinucleotide cleavage junction as indicated above relevant lanes), were incubated in the presence (+) or absence (-) of DNAzyme AGCT for 72 h under standard reaction conditions. It should be noted that corresponding rate constants differ between dinucleotide junctions. (B) The same reactions described above were also conducted with DNAzyme GC. Dinucleotide junctions are written 5' to 3'. Reaction products were separated by 10% denaturing PAGE.

substrate (versus an all-RNA substrate), which may resemble the transition state structure more closely (21).

In vitro selection

In our preceding investigation, DNAzyme GC was identified by conducting a limited, but focused sampling of sequence space, based around the structural scaffold of an existing DNAzyme. To gain insight into how small catalytic motifs like DNAzyme GC might be distributed over a broader unbiased region of sequence space, we conducted *in vitro* selection experiments. *In vitro* selection is a process by which DNAzymes (and other functional nucleic acids) can be identified from large libraries of random sequences, without systematically screening every sequence for activity (22,23). Instead, all sequences

within the library are simultaneously subjected to an iterative selective amplification process that progressively increases the ratio of DNAzyme sequences to non-functional sequences, until the DNAzymes are in significant enough excess to allow their detection by cloning and sequencing.

In keeping with our objective to probe the fundamental limits of DNA catalysis, we specifically sought to identify sequences that were comparable in size, or even smaller than DNAzyme GC. We therefore used two unbiased libraries composed of sequences containing only 15 or 20 random-sequence nucleotides (restricted to either guanosine or cytidine). For convenience these selection experiments are referred to as S_{15} and S_{20} , according to the size of their random-sequence domains. The library design and selection scheme are described in detail in Supplementary Figures S3 and S4, respectively. As part of the library design, we used a GC-rich substrate in the hope of identifying sequences that could effectively accommodate both catalysis and proper folding in a complete binary nucleotide system. However, it should be noted that this substrate was not susceptible to cleavage by DNAzyme GC, when they were tested together in a simple intermolecular cleavage assay (Figure S5). A probable explanation is the propensity for the substrate and DNAzyme to misfold into inactive conformations.

If DNAzymes were present in the original libraries, we estimated that it should take no more than around four rounds of selection to enrich any active sequences to a detectable level (depending on the actual enrichment factor per round). For instance, given an initial library size of 10^{14} molecules, there should be $\sim 10^8$ copies of each of the $\sim 10^6$ different sequence permutations represented in the S_{20} ($S = G$ or C) library. Therefore, even if the selection strategy provided only a modest enrichment factor of ~ 10 -fold per round, it should only take around four rounds to enrich the active sequences to a detectable number (i.e. $\sim 10^{12}$ molecules). In theory, even fewer rounds of selection should be required to detect active sequences in the S_{15} library. For comparison, we recently conducted a selection experiment using an N_{20} ($N = A, G, C$ or T) random library and the same selection strategy, which required only five rounds of selection to enrich the active sequences to a detectable level (8). These sequences were present in the initial library at a frequency of ~ 300 copies per sequence, suggesting an enrichment factor of ~ 100 -fold per round.

No cleavage signal was observed in the S_{20} selection experiment by round 5, which was therefore discontinued. However, a cleavage signal was observed by round 5 in the S_{15} experiment (later than expected), which was subsequently pursued for three additional rounds. The selection progress (i.e. percent cleavage per round) is shown in Figure S6. The terminal population was estimated to have an intramolecular cleavage rate constant of $\sim 7 \times 10^{-3} \text{ min}^{-1}$, which potentially represented a ~ 10 -fold improvement in activity relative to DNAzyme GC. However, subsequent analysis of 31 randomly selected clones revealed that the population had converged on one sequence class that contained a single A residue in the S_{15} random domain, and additional mutations in a

peripheral fixed-sequence region of the library (Figure S7), which were likely acquired during PCR. These mutations appeared to form the basis of an irregular 8–17 motif containing all four types of nucleotides (Figure S8), contrary to our objectives.

The fact that alternative binary catalytic motifs were not identified in these *in vitro* selection experiments, suggests that DNAzyme GC may represent a unique and optimal solution to RNA-cleavage within the sequence space afforded by the given libraries. However, this interpretation is made tentatively, because misfolding or even unpredictable population dynamics (24) are potential limitations of *in vitro* selection that could obscure the presence of some catalytic motifs.

DISCUSSION

Exploring the limits of DNA-mediated catalysis

Herein, we have probed the fundamental requirements for DNA catalysis, and shown that a simpler alphabet comprised of fewer than the standard four types of deoxyribonucleotides can support appreciable catalytic activity. Despite lacking the nucleophilic benefits of an extra 2'-OH group, DNA seems no less capable at catalysis than its RNA counterpart. Our results complement and extend the previous report of a ligase ribozyme lacking cytidine (25), by demonstrating that a combination of A, G and C, as well as G, C and T deoxyribonucleotides is sufficient for nucleic-acid-mediated catalysis. This suggests that RNA should also be able to rely exclusively on these combinations of three (ribo) nucleotides for catalysis. More importantly, we have shown that nucleic acid catalysis is possible with an even simpler alphabet comprised of only guanosine and cytidine deoxyribonucleotides, which complements the report of a ligase ribozyme composed of only uracil and 2,6-diaminopurine ribonucleotides (5). The results of this study are perhaps even more striking given the relatively small size of the relevant DNAzymes. One might expect that a catalyst made from an already functionally impoverished polymer like DNA (as compared to both protein and RNA), might require an extra long sequence of nucleotides to compensate for any reduction in the size of its nucleotide repertoire. However, we have demonstrated that a specific sequence composed of just seven guanosine and six cytidine nucleotides can form the catalytic core of a functional RNA-cleaving DNAzyme. In comparison, the presumed catalytic domain of the binary ribozyme reported by Joyce and colleagues measured ~ 66 nt in length (5).

In this study, DNAzymes composed of just three or two different nucleotides were identified by examining a limited amount of sequence space based around the structural scaffold of an existing DNAzyme. This strategy was practical due to the relatively small size of the parent 8–17 DNAzyme, and the availability of detailed mutational data. We expect that a more comprehensive and unbiased exploration of sequence space could potentially reveal additional examples of catalytic DNA motifs, including DNAzymes composed of those combinations of nucleotides that we chose not to pursue herein

(i.e. ACT and AT). Similarly, it may be possible to identify faster binary DNazymes by removing the size constraints imposed herein, and increasing the structural complexity of potential candidates.

Functional consequences for a simplified nucleotide alphabet

DNAzyme AGCT appears to use all four types of nucleotides productively to achieve a level of catalytic activity that cannot be sustained when one or more types are replaced. In general, this observation is not surprising since a larger nucleotide alphabet can provide more structural options to pursue different catalytic strategies, and to do so more optimally, by achieving a greater level of structural refinement necessary for the precise alignment and spatial arrangement of catalytic/reacting groups. Furthermore, structures based on a larger nucleotide alphabet are expected to be more clearly defined, and less susceptible to misfolding (26). Based on the compositional complexity of the 8–17 variants tested herein, we observed a hierarchy of activity that followed the order AGCT > AGC > GCT > GC > AGT > ACT (where DNAzyme ACT was completely inactive). These Variants of DNAzyme 8–17 composed exclusively of A and T residues were not explicitly tested, but assumed to be inactive given the indispensability of G and C residues reported in the literature (13,15), which is also reflected in the preceding order of activity. It should be noted that this hierarchy of activity is not expected to be an artifact of the limited (but focused) sampling of 8–17 variants tested herein, because it is fully consistent with the results of comprehensive mutational analyses (13,15) and *in vitro* selection experiments that have considered many more sequence permutations (7,11,27).

The precise role(s) played by each type of nucleotide in the catalytic core of the 8–17 DNAzyme has yet to be clearly defined, as no high resolution structure is currently available. Nevertheless, previous biochemical and mutational studies can provide some context to suggest how the removal of A and T residues may adversely affect the activity of 8–17. A recent cross-linking analysis of 8–17 has suggested that nucleotides T2.1, A6 and A15 (as well as G7, C8, C13 and G14) form close contact interactions with cleavage site residues G1.1 and G18 in the substrate (28). In particular, substitution of A6 (and G7) with non-standard nucleotides has suggested that these residues may be directly involved in a network of functionally important hydrogen bonds (15). Therefore, the replacement of A and T residues may remove important contacts or create steric or electrostatic clashes that impair the ability of 8–17 to fold into an optimal catalytic conformation. It is also possible that the removal of A and T residues may cause a structural rearrangement that disrupts the binding site for an important activating metal ion. For instance, the TCGAA unpaired region of 8–17 has been suggested as a possible binding location for a single activating metal ion (16). The log-linear pH dependence (slope ~1) observed in the pH-rate constant profile of DNAzyme GC (Figure 5), is indicative of a single deprotonation event during the rate limiting step, which can most likely be assigned to the 2'-OH group of

G18 at the cleavage site (29,30). A disrupted binding site could limit the ability of the metal ion (acting either as a Lewis acid or general base) to assist in the deprotonation of this 2'-OH group (30). Alternatively, the potential structure-stabilizing benefits of the metal ion could be diminished.

Implications for the evolution of nucleic acid function

This study also demonstrates that simpler nucleotide alphabets have the capacity to support catalysis for more than one type of chemical reaction, extending a list that previously consisted of only RNA ligation (5,25), to now include RNA cleavage as well. This finding provides additional support to the plausibility of a primordial world in which (ribo)nucleotides may have formed the first polymers responsible for both the genetic and catalytic requirements of life (31). From an evolutionary perspective, our results suggest how an increase in the complexity of the nucleotide alphabet over time would be selectively advantageous. This is exemplified by the ~10 000-fold increase in activity, and ~8-fold increase in the number of susceptible cleavage sites, observed between DNAzyme GC and DNAzyme AGCT.

Unlike ribozymes, DNazymes have so far not been found in nature. Nevertheless, the inherent catalytic properties of DNA beg a very simple, but important question: is this catalytic ability merely a coincidence? The potential sequence and structural simplicity of DNA-based catalysis demonstrated herein suggests how natural DNazymes might easily be overlooked among the billions of nucleotides that make up some genomes. Therefore, we are left to wonder if there may be some merit to the notion of an ancient 'DNA world' (31), in analogy to the well-known 'RNA world' hypothesis (32,33).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Canadian Institutes of Health Research (grant MOP37964). Y.L. is a Canada research chair. K.S. holds a Natural Sciences and Engineering Research Council Doctoral Canada Graduate Scholarship. Funding for open access charge: Canadian Institutes of Health Research.

Conflict of interest statement. None declared.

REFERENCES

1. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
2. Breaker, R.R. and Joyce, G.F. (1994) A DNA enzyme that cleaves RNA. *Chem. Biol.*, **1**, 223–229.
3. Baum, D.A. and Silverman, S.K. (2008) Deoxyribozymes: useful DNA catalysts in vitro and in vivo. *Cell Mol. Life Sci.*, **65**, 2156–2174.

4. Walter, K.U., Vamvaca, K. and Hilvert, D. (2005) An active enzyme constructed from a 9-amino acid alphabet. *J. Biol. Chem.*, **280**, 37742–37746.
5. Reader, J.S. and Joyce, G.F. (2002) A ribozyme composed of only two different nucleotides. *Nature*, **420**, 841–844.
6. Santoro, S.W. and Joyce, G.F. (1997) A general purpose RNA-cleaving DNA enzyme. *Proc. Natl Acad. Sci. USA*, **94**, 4262–4266.
7. Schlosser, K. and Li, Y. (2004) Tracing sequence diversity change of RNA-cleaving deoxyribozymes under increasing selection pressure during in vitro selection. *Biochemistry*, **43**, 9695–9707.
8. Schlosser, K., Gu, J., Lam, J.C. and Li, Y. (2008) In vitro selection of small RNA-cleaving deoxyribozymes that cleave pyrimidine-pyrimidine junctions. *Nucleic Acids Res.*, **36**, 4768–4777.
9. Li, J., Zheng, W., Kwon, A.H. and Lu, Y. (2000) In vitro selection and characterization of a highly efficient Zn(II)-dependent RNA-cleaving deoxyribozyme. *Nucleic Acids Res.*, **28**, 481–488.
10. Faulhammer, D. and Famulok, M. (1996) The Ca²⁺ ion as a cofactor for a novel RNA-cleaving deoxyribozyme. *Angew Chem. Int. Ed. Engl.*, **35**, 2809–2813.
11. Cruz, R.P., Withers, J.B. and Li, Y. (2004) Dinucleotide junction cleavage versatility of 8-17 deoxyribozyme. *Chem. Biol.*, **11**, 57–67.
12. Peracchi, A. (2000) Preferential activation of the 8-17 deoxyribozyme by Ca(2+) ions. Evidence for the identity of 8-17 with the catalytic domain of the Mg5 deoxyribozyme. *J. Biol. Chem.*, **275**, 11693–11697.
13. Schlosser, K., Gu, J., Sule, L. and Li, Y. (2008) Sequence-function relationships provide new insight into the cleavage site selectivity of the 8-17 RNA-cleaving deoxyribozyme. *Nucleic Acids Res.*, **36**, 1472–1481.
14. Li, Y. and Breaker, R.R. (1999) Kinetics for specific base catalysis of RNA degradation by transesterification involving the 2'-hydroxyl group. *J. Am. Chem. Soc.*, **121**, 5364–5372.
15. Peracchi, A., Bonaccio, M. and Clerici, M. (2005) A mutational analysis of the 8-17 deoxyribozyme core. *J. Mol. Biol.*, **352**, 783–794.
16. Bonaccio, M., Credali, A. and Peracchi, A. (2004) Kinetic and thermodynamic characterization of the RNA-cleaving 8-17 deoxyribozyme. *Nucleic Acids Res.*, **32**, 916–925.
17. Kandadai, S.A., Chiuman, W. and Li, Y. (2006) Phosphoester-transfer mechanism of an RNA-cleaving acidic deoxyribozyme revealed by radioactivity tracking and enzymatic digestion. *Chem. Commun.*, 2359–2361.
18. Doudna, J.A. and Cech, T.R. (2002) The chemical repertoire of natural ribozymes. *Nature*, **418**, 222–228.
19. Brown, A.K., Li, J., Pavot, C.M. and Lu, Y. (2003) A lead-dependent DNAzyme with a two-step mechanism. *Biochemistry*, **42**, 7152–7161.
20. Stage-Zimmermann, T.K. and Uhlenbeck, O.C. (1998) Hammerhead ribozyme kinetics. *RNA*, **4**, 875–889.
21. Ota, N., Warashina, M., Hirano, K., Hatanaka, K. and Taira, K. (1998) Effects of helical structures formed by the binding arms of DNAzymes and their substrates on catalytic activity. *Nucleic Acids Res.*, **26**, 3385–3391.
22. Wilson, D.S. and Szostak, J.W. (1999) In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.*, **68**, 611–647.
23. Joyce, G.F. (2004) Directed evolution of nucleic acid enzymes. *Annu. Rev. Biochem.*, **73**, 791–836.
24. Schlosser, K. and Li, Y. (2005) Diverse evolutionary trajectories characterize a community of RNA-cleaving deoxyribozymes: a case study into the population dynamics of in vitro selection. *J. Mol. Evol.*, **61**, 192–206.
25. Rogers, J. and Joyce, G.F. (1999) A ribozyme that lacks cytidine. *Nature*, **402**, 323–325.
26. Szathmary, E. (2003) Why are there four letters in the genetic alphabet? *Nat. Rev. Genet.*, **4**, 995–1001.
27. Schlosser, K., Lam, J.C. and Li, Y. (2006) Characterization of long RNA-cleaving deoxyribozymes with short catalytic cores: the effect of excess sequence elements on the outcome of in vitro selection. *Nucleic Acids Res.*, **34**, 2445–2454.
28. Liu, Y. and Sen, D. (2008) A contact photo-cross-linking investigation of the active site of the 8-17 deoxyribozyme. *J. Mol. Biol.*, **381**, 845–859.
29. Emilsson, G.M., Nakamura, S., Roth, A. and Breaker, R.R. (2003) Ribozyme speed limits. *RNA*, **9**, 907–918.
30. Breaker, R.R., Emilsson, G.M., Lazarev, D., Nakamura, S., Puskarz, I.J., Roth, A. and Sudarsan, N. (2003) A common speed limit for RNA-cleaving ribozymes and deoxyribozymes. *RNA*, **9**, 949–957.
31. Dworkin, J.P., Lazcano, A. and Miller, S.L. (2003) The roads to and from the RNA world. *J. Theor. Biol.*, **222**, 127–134.
32. Gilbert, W. (1986) Origin of life: The RNA world. *Nature*, **319**, 618.
33. Muller, U.F. (2006) Re-creating an RNA world. *Cell Mol. Life Sci.*, **63**, 1278–1293.