Organic & Biomolecular Chemistry

Cite this: *Org. Biomol. Chem.*, 2011, **9**, 2266

www.rsc.org/obc

PAPER

Protein-inspired modified DNAzymes: dramatic effects of shortening side-chain length of 8-imidazolyl modified deoxyadenosines in selecting RNaseA mimicking DNAzymes†

Christopher J. Hipolito, Marcel Hollenstein, Curtis H. Lam and David M. Perrin*

Received 18th August 2010, Accepted 20th October 2010
DOI: 10.1039/c0ob00595a

The discovery of imidazole/amine-functionalized DNAzymes that efficiently cleave RNA independently of divalent metal cations (M^{2+}) and cofactors underscores the importance of expanding the catalytic repertoire with modified nucleosides. Considerable effort has gone into defining polymerase tolerances of various modified dNTPs for synthesizing and amplifying modified DNA. While long linkers are generally found to enhance incorporation and therefore increase sequence space, shorter linkers may reduce the entropic penalty paid for orienting catalytic functionality. Catalytic enhancement ultimately depends on both the functional group and appropriate linkage to the nucleobase. Whether a shorter linker provides enough catalytic enhancement to outweigh the cost of reduced polymerizability can only be determined by the outcome of the selection. Herein, we report the selection of DNAzyme 20–49 (Dz20–49), which depends on amine, guanidine, and imidazole-modified dNTPs. In contrast to previous selections where we used dA^{ime}TP (8-(4-imidazolyl)ethylamino-2′-dATP), here we used dA^{imm}TP (8-(4-imidazolyl)methylamino-2′-dATP), in which the linker arm is shortened by one methylene group. Although the most active clone, Dz20–49, was absolutely dependent on the incorporation of either dA^{imm}p or dA^{ime}p, it catalyzed cofactor independent self-cleavage with a rate constant of $3.1 \pm 0.3 \times 10^{-3}$ min^{-1}, a value not dissimilar from unmodified catalysts and strikingly inferior to modified catalysts selected with dA^{ime}TP. These results demonstrate that very subtle differences in modified nucleotide composition may dramatically effect DNAzyme selection.

Introduction

SELEX and related combinatorial methods of *in vitro* selection[1,2] have enabled the discovery of various nucleic acid catalysts including DNAzymes.[3–8] Nevertheless, in comparison to proteins, DNAzymes specifically, and nucleic acids generally, are functionality-poor. Although Mg^{2+} (and other divalent metal ions—generally M^{2+}) at concentrations much higher than physiological may relieve this functional deficiency, great effort has been directed to defining DNA polymerase tolerances for incorporating modified nucleoside triphosphates (dXTPs) bearing functional groups, some of which are commonly found at the active sites of protein enzymes and antibodies, *e.g.* imidazoles (His), amines (Lys), guanidines (Arg), aromatics (Phe, Trp), alkyls (Leu), thiols (Cys) *etc.*[9–33] Two early reports demonstrated the simultaneous use of two modified dNTPs[12,14] (*vide infra*) while two recent reports described selections with three modified dNTPs.[18,19] Finally, the potential of incorporating four modified nucleotides was described although to date there is no report of functional selection.[16]

Two critical conditions must be met in selecting with modified dXTPs: (1) the modified dXTP must be a substrate for a template-dependent polymerase and (2) the modified polymer must be a template for sequence specific recopying into DNA for PCR amplification. Efficient incorporation implies increased sequence space, which in turn is thought to result in selection of optimal activity. By the same token, poor incorporation is thought to limit sequence space, which in turn consequents selection of sub-optimal activity. Linker arm elements that promote efficient dXTP incorporation, including the ability to incorporate multiple modified dXTPs in a row, are: (a) position, (b) bond planarity (sp^1 or sp^2 *vs.* sp^3), and (c) length.[20,34] Although fewer studies have explored reamplification of modified DNA,[23] the same effects are likely to hold.

Yet even if all chemo-enzymatic conditions are optimized for incorporation and read-through, the *real* value of a modified dXTP can only be gauged by comparing the catalytic activity that evolves from modified and unmodified selections. Despite a large number of studies that address the polymerization of modified

Department of Chemistry, The University of British Columbia, 2036 Main Mall, Vancouver, BC, Canada. E-mail: dperrin@chem.ubc.ca

† Electronic supplementary information (ESI) available: *In vitro* selection protocols. Individual clone sequences. See DOI: 10.1039/c0ob00595a

dNTPs with an eye to eventual use in selection, there are very few examples of modified DNAzymes that have been selected. The contribution of modified dXTPs used in DNAzyme selections has been recently reviewed;[8] in many cases where a modified nucleic acid catalyst was selected,[30,35] an unmodified catalyst with very similar activity also was selected.[36,37] This raises questions as to the value of modified dXTPs.

Nevertheless, a uniquely challenging reaction in which modified dNTPs have dramatically increased rates compared to unmodified catalysts is M^{2+}/cofactor-independent RNA cleavage; unmodified DNA catalysts deliberately selected for M^{2+}-independent RNA cleavage manifest extremely modest rates.[38,39] This deficiency has been remediated by selecting DNAzymes with two modified dXTPs where, by analogy to RNaseA, imidazoles provide acid/base catalysis while cationic amines provide electrostatic stabilization.[12,14] Perrin and coworkers selected $Dz9_{25}$–11, which simultaneously utilized 8-(4-imidazolyl)ethylamino-2′-deoxyadenosine ($dA^{ime}TP$)‡ **3** and 5-aminoallyl-2′-deoxyuridine ($dU^{aa}TP$)§ **5**,[12] whereas Williams and coworkers selected a similar DNAzyme using modified dATP **8** and dUTP **9** (Fig. 1A).[14] A noted advantage of **8** and **9** is that they are easily polymerized compared to **3** and **5**. Yet in terms of catalytic activity, when *both* **8** and **9** were replaced with unmodified congeners, 7% self-cleavage was still observed in Williams' DNAzyme. By contrast, with a single exception (A_{24}), replacing even a single modified dU or dA in $Dz9_{25}$–11, with an unmodified congener resulted in >10-fold loss in activity. Mechanistic, kinetic, and affinity labelling studies further corroborated an RNaseA-like acid/base role for two essential imidazoles and an electrostatic role for at least one amine.[40] We hypothesized that the shorter linker arm length in **3** compared to **9** conformationally restricted the imidazole to play indispensable catalytic roles, therefore leading us to conclude that the modest incorporation of **3**, along with hindered amplification of duly modified strands, represented an acceptable trade-off that gave intended imidazole-dependent catalysts. Subsequently, Hollenstein *et al.* retained **3** while including both guanidiniumallyl-dUTP ($dU^{ga}TP$)¶ **6** and aminoallyl-dCTP ($dC^{aa}TP$) **7**,[17] (see Fig. 1), in the selection of vastly improved Mg^{2+}-independent self-cleaving DNAzymes from N20 and N40 random regions; respectively Dz9–86 ($k_{obs} = 0.13 \pm 0.03$ min^{-1}, 37 °C)[18] and Dz10–66 ($k_{obs} = 0.60 \pm 0.05$ min^{-1}, 37 °C, Fig. 1B).[19]

In light of the success with Dz10–66, we sought to revisit the question of linker length with regards to the imidazole tether. Because **3** and **9** differ so greatly in both base and linker arm composition, and because **9** was not absolutely required for activity whereas **3** was, we sought to test the effect of linker arm length with 8-(4-imidazolyl)methylamino-2′-deoxyadenosine triphosphate ($dA^{imm}TP$)‖ **2**, an analog of **3** wherein the imidazole tether is shortened by one methylene. Previously, we showed that **2** is polymerized with virtually the same efficiency as **3**,[41] and *a priori* should provide similar sequence space coverage. Furthermore, because RNaseA requires only two properly placed imidazoles for acid/base catalysis, we hypothesized that imidazoles need not be abundantly represented for good activity, but rather, they must be conformationally constrained to properly orient the imidazoles for catalytic activity. To that end, shortening the linker length would hypothetically restrict conformational flexibility, and therefore potentially enhance the catalytic properties of a selected DNAzyme. In order to assess the effects of this discrete difference in linker length, we selected "analogs" of Dz10–66 using **2**, **6**, and **7**, whereby we held all other aspects of selection constant. Herein, we present the discovery of a new M^{2+}-independent RNA self-cleaving DNAzyme that was selected using **2**; while absolutely dependent on **2** it presented a significantly depressed rate compared to Dz10–66 and did not have significantly enhanced rates over unmodified DNAzymes. This work highlights how discrete changes in dNTP composition result in radically different selection outcomes.

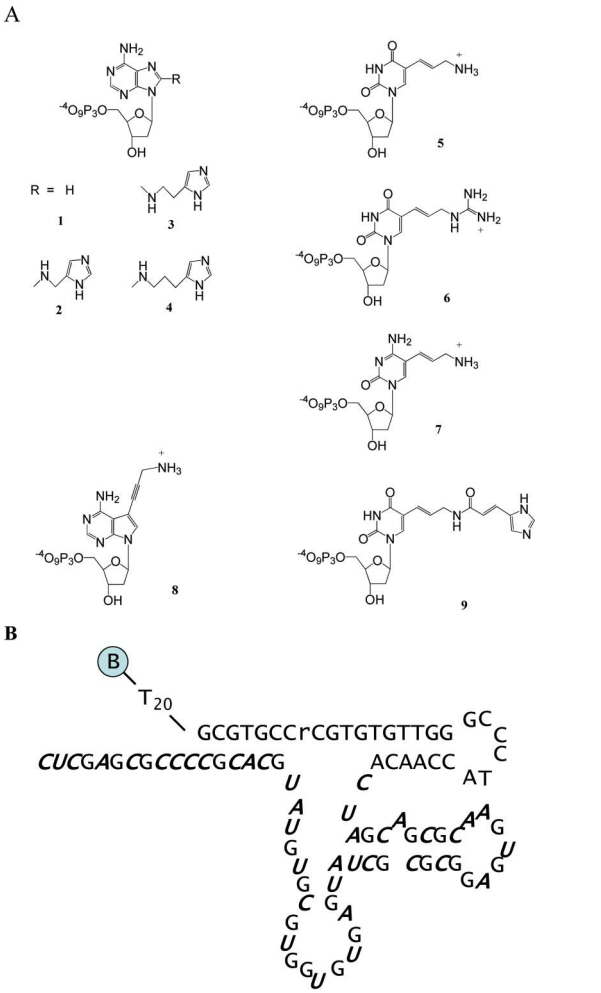

Fig. 1 A: Chemical structure of **1** (dATP), **2** ($dA^{imm}TP$), **3** ($dA^{ime}TP$), **4** ($dA^{imp}TP$), **5** ($dU^{aa}TP$), **6** ($dU^{ga}TP$), **7** ($dC^{aa}TP$). Chemical structures of amino-modified dATP **8** and imidazole-modified dUTP **9** used by Sidorov *et al.* B: Secondary structure of Dz10–66. Bold italic ***A***, ***C*** and ***U*** represent nucleosides resulting from the polymerization of **3**, **6** and **7**, respectively.

‡ The superscript ime refers to the (imidazolyl)aminoethyl group of the 8-modified deoxyadenosine.

§ The superscript aa refers to the aminoallyl of the 5-modified deoxyuridine or deoxycytidine.

¶ The superscript ga refers to the guanidiniumallyl group of the 5-modified deoxyuridine.

‖ The superscript imm refers to the (imidazolyl)methylamino linker group of the 8-modified deoxyadenosine.

Results and discussion

In vitro selection of Dz20–49

An initial population of modified DNA was created by the enzymatic copolymerization of modified dXTPs (**2**, **6**, and **7** in Fig. 1) along with dGTP and dGTPα^{32}P along a template containing 40 degenerate positions (N40). In contrast to the selection of Dz10–66, where self-cleavage activity appeared in generation 3 or 4, very slight (<5%) activity was detected only after 9 rounds of selection. While there was a slight increase in activity from generation 18 to generation 20 (see ESI†), the low overall yield (5% total cleavage after an hour) and the high number of rounds suggested that minimally active clones populated the majority of this gene pool. From round 20, thirty-three clones that contained a single insert of correct size were isolated and sequenced (many more plasmids were picked and initially mapped, hence the individual clone numbers go beyond 33); within this set, some library convergence was observed with nine clones containing either eight or nine modified dAs, while clone 35 (see ESI†) contained ten modified dAs, two of which were neighbours. Despite an abundance of modified dAs in certain clones, they had no detectable activity. In addition, the enzymatic synthesis of individual clones with **2**, **6** and **7** often resulted in significant amounts of truncated products. Of the isolated clones, only six had detectable activity and the most active clone, constituted as Dz20–49 (Fig. 2), was retained for more detailed characterization because of (i) a comparatively high self-cleavage rate, (ii) cleavage yields (approaching 80%), and (iii) the lowest amount of truncation products. The sequence and hypothetical 2D structure of Dz20–49 determined by mfold,[42,43] are shown in Fig. 2.

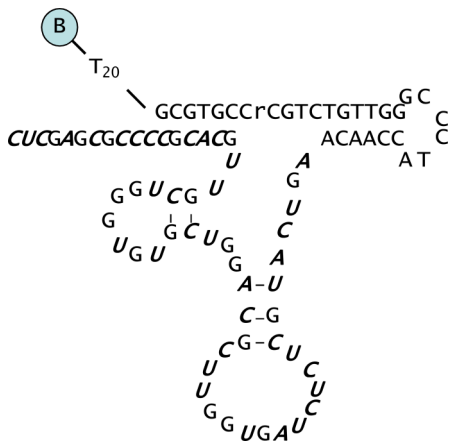

Fig. 2 Putative secondary structure of the most efficient DNAzyme, Dz20–49 based on mfold. Bold italic ***A***, ***U*** and ***C*** indicate the positions of modified nucleosides introduced by the polymerization of **2**, **6**, and **7** respectively. The encircled B indicates a biotin tether.

In contrast to Dz10–66 (Fig. 1B) whose putative catalytic region contained 43 nucleosides comprising 17 dGs, 8 amino-modified-dCs, 10 guanidinium-modified dUs, and eight dA^{ime}s, including two in a row, the catalytic loop of Dz20–49 contained 40 nucleosides with roughly the same percentage of dGs, dCs, a slightly greater percentage of guanidinium-dUs and a lower percentage of dA^{imm}s; there are only four dA^{imm}s, none of which are neighbours. Notably, other clones had a greater number of dA^{imm}s, including two clones in which the dA^{imm}s were neighbours, yet these species were not as active as Dz20–49.

Kinetics of self-cleavage

In order to investigate self-cleavage, Dz20–49 was prepared according to the same method as described for the selection. Although **2** (as with **3**) is a relatively poor substrate for most DNA polymerases,[41] enzymatic synthesis of Dz20–49 provided one major band with minimal truncation (Fig. 3A). The rate of self-cleavage was monophasic and proceeded with a rate constant, k_{obs} of $3.1 \pm 0.3 \times 10^{-3}$ min^{-1} (Fig. 3B). For two other assays, k_{obs} values were calculated to be $3.4 \pm 0.3 \times 10^{-3}$ min^{-1} and $3.9 \pm 0.3 \times 10^{-3}$ min^{-1}. Notably, up to 30% of the DNAzyme was inactive (P_∞ was calculated to be 0.75 ± 0.03, 0.68 ± 0.02, 0.69 ± 0.02 on three different occasions) and the reason for this was not pursued but may entail either misfolding or even misincorporation of nucleosides. In determining the rate constant of the active fraction, the contribution of these inactive DNAzymes was considered and is discussed in the ESI.† In addition, the data can also be fit to a double-exponential equation that comprises both a fast (~15%) and a slow cleaving fraction (~85%). Nevertheless, as the fast phase represents a minor species accurate determination of its rate constant would require a more thorough analysis due to the high associated error.[60] Notably the rate constant for the slow

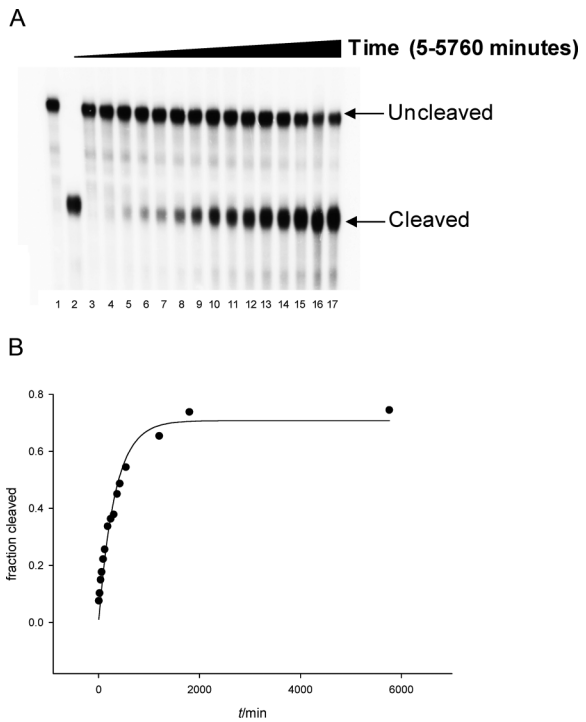

Fig. 3 Kinetic analysis of self-cleavage of Dz20–49. A Autoradiograph of a 7% denaturing polyacrylamide gel depicting self-cleavage over a period of 5760 minutes. B Graphical analysis of the kinetic data. Points are represented as the fraction of DNAzyme cleaved as a function of time. Lanes 1 and 2 show uncleaved full-length Dz20–49 and RNaseA-treated Dz20–49, respectively. Lanes 3–17 show the reaction at times points 5, 20, 40, 60, 90, 120, 180, 240, 300, 366, 420, 540, 1200, 1800, and 5760 minutes, respectively. Calculated k_{obs} for this particular experiment is $3.1 \pm 0.3 \times 10^{-3}$ min^{-1} ($R^2 = 0.95$).

phase did not differ appreciably from the value calculated using a single-exponential equation. The possibility that a minor fraction that is much more active than the major fraction that is minimally active, with the possibility of dynamic interconversion as discussed at length elsewhere[60] cannot be excluded in this case.

The effects of linker length on the activity of Dz20–49

Consistent with all other DNAzymes (*e.g.* $Dz9_{25}$–11, Dz9–86, Dz10–66) which are absolutely dependent on the incorporation of dA^{ime}, Dz20–49 was also absolutely dependent on the incorporation of dA^{imm}; enzymatic resynthesis with dATP led to a total loss of activity (see Fig. 4A). Since dA^{ime} (introduced *via* polymerization with **3**) was absolutely required for activity of Dz10–66, and because resynthesis of Dz10–66 with either **2** (shorter methyl linker) or **4** (longer propyl linker) resulted in total loss of activity, we wished to test the effects of lengthening the linker on Dz20–49. Therefore, Dz20–49 variants were prepared by enzymatic polymerization using either **3** or **4**, which contained increasingly longer linkers. Interestingly, when **3** (ethyl linker) replaced **2** (methyl linker) in the synthesis of Dz20–49, self-cleavage activity was almost identical to that selected with **2** (k_{obs} = 4.1 ± 0.5 × 10^{-3} min^{-1}; Fig. 4B). However when **4** (propyl linker) was used *in lieu* of **2**, virtually all activity was lost (Fig. 4C).

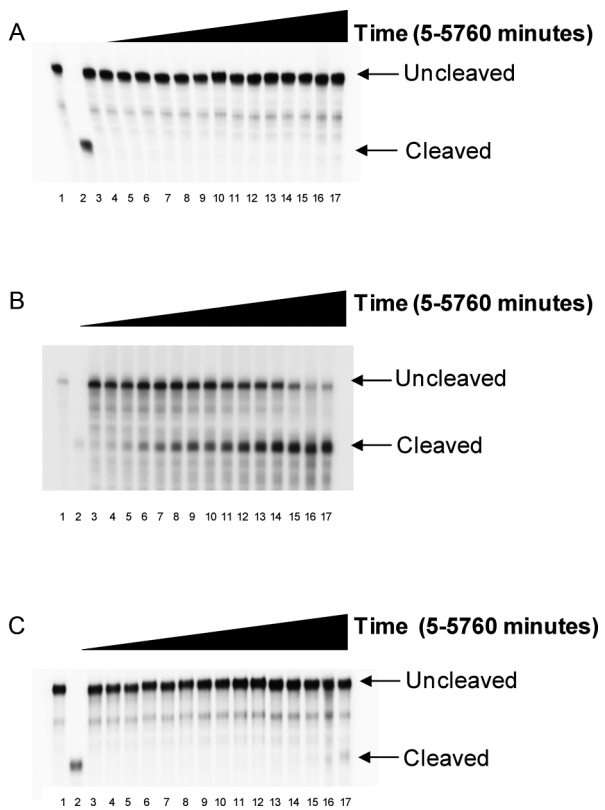

Fig. 4 Modified dA replacement studies. Time dependant self-cleavage of Dz20–49 analogues where **2** (dA^{imm}) is replaced with a) **1** (dA), b) **3** (dA^{ime}), c) **4** (dA^{imp}). For all three studies, Lanes 1 and 2 show uncleaved full-length Dz20–49and RNaseA-treated Dz20–49, respectively. Lanes 3–17 show the reaction at times points 5, 20, 40, 60, 90, 120, 180, 240, 300, 366, 420, 540, 1200, 1800, and 5760 minutes, respectively.

Effect of temperature and pH–rate profile

Since temperature variation can greatly affect the rate of M^{2+}-independent DNAzymes,[44] Dz20–49 was also assayed at 37 °C; at this temperature, the average self-cleavage rate constant dropped by about a half (2.0 ± 0.5 × 10^{-3} min^{-1}) compared to the rate constant at 24 °C. This observation lies in stark contrast to the temperature–rate profile of Dz10–66 for which the self-cleavage rate markedly increased with temperature from 24 °C to 37 °C. Instead, Dz20–49 mirrors the unmodified DNAzyme G3 selected by Geyer and Sen,[39] herein referred to as DzG3, where the k_{obs} sharply decreases above 30 °C. Similar effects were seen when Dz20–49 was prepared with **3** (data not shown).

The pH–rate profile for self-cleavage is bell-shaped (see Fig. 5). The rate increased, over the range 6.0–7.0, reaches a maximum at 7.0 and decreases over the range 7.0–9.0. Using eqn (2),[45] the pH-rate profile analysis yields pK_a values of 6.2 and 8.1, values consistent with a two-step acid–base mechanism. Nevertheless, the absence of log-linearity (see ESI†) suggests other effects such as a competitive folding step or multiple reaction channels that may not depend on concerted acid–base catalysis. Distinguishing these effects requires much more investigation and would change the focus of this work, which is to address the general effects of linker arm length on selection outcome.

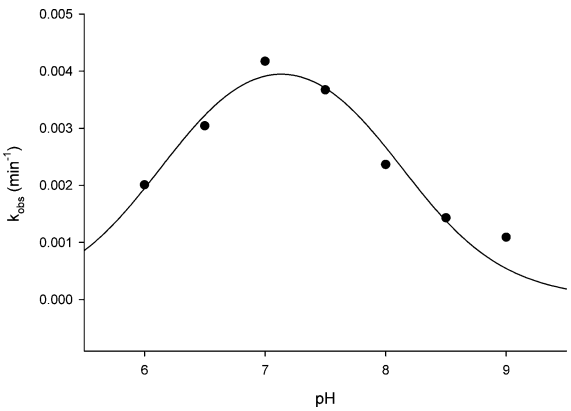

Fig. 5 pH–Rate profile. Rates of self cleavage were determined at pH 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0 in (50 mM buffer, 200 mM NaCl, 1 mM EDTA) at room temperature. Values for pK_a were determined using eqn (2) (see Materials and Methods) to be 6.2 and 8.1 and $k_{max} = 4.8 \times 10^{-3}$ min^{-1} (R^2 = 0.90).

Amplification of the modified templates

Although we had already established that **2** was polymerized with very similar efficiency as **3**,[41] here we examined the relative amplifiability of the Dz20–49 sequence as duly modified to seek hints as to why a selection using **2** produced such an inferior catalyst compared to Dz10–66 that was selected with **3**. Four biotinylated DNA strands based on the Dz20–49 sequence were prepared enzymatically, affixed to avidin, separated from their templates by a brief NaOH wash, removed by RNaseA treatment, purified by PAGE and subjected to PCR amplification to assess the effects of the linker length of 8-(4-imidazolyl)-modified dA upon amplification. One test template was synthesized using only natural dNTPs while three others were synthesized using dGTP, **6**, **7**, and one of **2**, **3**, or **4**. As expected, the template containing

natural dA produced the highest amount of amplicon. The PCRs that used modified templates containing the nucleosides of **3** or **4** produced about the same amount of amplicon, although in slightly lower amounts than amplicon of the unmodified template. The PCR of template DNA containing the nucleoside of **2** illustrates the most dramatic decrease in amplifiability. While Vent (*exo*-) was used in the selections of Dz10–66 and Dz20–49, both Vent (*exo*-) (Fig. 6) and Taq polymerase (see ESI†) were assayed to determine the amplifiability of the templates. Vent (*exo*-), which was used in this and previous selections, produced the cleanest product.

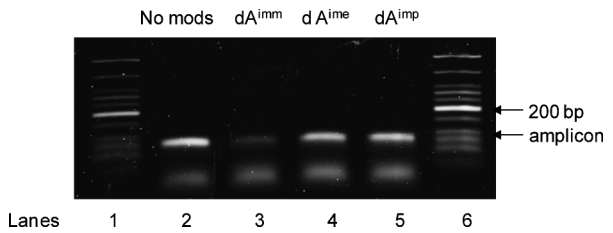

Fig. 6 PCR amplification of modified templates by Vent (*exo*-) (45 cycles). Lanes 1 and 6 contain NEB Low Molecular Weight Ladder. For a control, a completely unmodified ribophosphodiester bond cleaved version of Dz20–49 was used as the control template (Lane 2). For modified templates, Dz20–49 was constructed with dGTP, **6**, **7** and one of: **2** (Lanes 3), **3** (Lane 4) or **4** (Lane 5).

General discussion

In light of numerous reports detailing efforts to define polymerase tolerances for modified dNTPs for the eventual use in producing artificial DNAzymes, we were in a unique position to test the effects on selection outcome following the use of two very closely related dATP analogs with discrete linker arm differences. Because imidazole-modified DNAzymes show greatly enhanced activity over unmodified congeners in the context of M^{2+}-free RNA cleavage, we selected for this activity to shed some light on the effects of modifying the linker length of the imidazole functionality that is generally essential for catalytic activity. Based on the comparison between DNAzymes selected by both Williams and Perrin, it was hypothesized that shortening the length of the linker between the nucleobase and the side chain would restrict conformational freedom and thereby provide a more rigid and therefore more efficient catalyst.[28] The effect on the selection nevertheless became apparent as self-cleavage activity was slow to appear over 9 rounds of selection, especially when compared to the selections of DNAzymes such as Dz8–17, Dz10–23[46] and Dz10–66. Often, appearance of activity at an early stage of the selection is a good indication of the presence of highly active catalysts. Consequently, the slow increase in activity hinted at the catalytic inferiority that would result. Indeed the activity in later generations, even though sluggishly improving, never reached levels attained in the selection of Dz10–66. Furthermore, resynthesis of G20 with **3** did not improve cleavage suggesting that clones with additional modified dAs in G20 were also inactive and not simply inactive variants of Dz10–66 (data not shown).

While hindered incorporation of both **2** and **3** was discussed previously,[41] the efficiency of read-through of duly modified templates in PCR containing dC^{aa} and dU^{ga} along with various modified dAs had not yet been explored. Various PCR templates were synthesized and amplified to help assess this efficiency. The amplicons produced using Vent (*exo*-) clearly showed that the dA^{imm}-modified template, which also contained the modifications dC^{aa} and dU^{ga}, is poorly amplified compared to the control template which is devoid of any modifications (Fig. 6). To compare the amplifiability of various dA^{x}-modified templates that also contain dC^{aa} and dU^{ga}, Dz20–49 strands containing dA^{ime} and dA^{imp} were found to amplify better than the dA^{imm}-modified template. Taq, which we did not use in this or past selections for amplification of cleavage products, showed similar trends, but also revealed truncation artifacts, the origins of which were not pursued (data not shown).

Despite the fact that 30% of the clones contained 8 or more modified dAs, an abundance similar to Dz10–66, the majority of them exhibited little or no activity with the exception of Dz20–49, which exhibited a rate constant of $3.1 \pm 0.3 \times 10^{-3}$ min^{-1} that represents a ~200 fold drop in activity compared to Dz10–66. Both in this case, and in the selection of Dz10–66, there seemed to be little correlation between the percentage of modified dAs and activity. Nevertheless, it is difficult to explain why inactive clones with a greater percentage of modified dAs survived the selection despite pressures against both the polymerization of **2** and the amplification of duly modified strands.

Although the self-cleavage rate of Dz20–49 compares to that of the unmodified DzG3,[39] ablation of all imidazoles by incorporating dA leads to a total suppression of the catalytic activity. These findings clearly show that while the modified dA is required for activity, shortening the linker length does not enhance catalytic activity. As with several other modified DNAzymes whose activity is rivaled by unmodified counterparts,[35,36] here is another example of a modified catalyst which, while absolutely dependent on the incorporation of three modified nucleosides, does not surpass the unmodified DzG3 in terms of activity or thermal stability.

While the bell-shaped pH–rate profile suggests a two step acid–base mechanism characteristic of imidazole directed catalysis, the modified adenosine may also promote pH-dependent folding. Nevertheless, if an 8-imidazole-alkyl-amino-dA were required only for folding, then replacing **2** (methyl linker) with **4** (propyl linker) might have also afforded a catalytically active strand. Although this was not the case, we cannot fully exclude the possibility that proper positioning of the imidazole, linked to a methylene or ethyl group but *not* a propyl group is somehow essential for folding and not catalysis.

The reasons as to why a shorter linker results in such a dramatic loss in activity remain unclear. While the 8-NH may intramolecularly H-bond to the imidazole nitrogen of **2** in a 5-membered ring such that the imidazole cannot act as a base, a similar 6-membered ring H-bonded configuration can be drawn for **3**, which provided excellent catalysts. NMR characterization of nucleosides of **2** and **3** shows a high correlation between the chemical shifts of the protons and the carbons of the imidazoles of the two species, which suggests that both imidazoles are in similar chemical environments, at least in deuterated organic solvents, where such a hypothetical H-bonding interaction would be reinforced.[41,47] It is therefore likely that the electronic environments of the imidazoles of **2** and **3** will be similar in the aqueous buffer used for self-cleavage and thus unlikely that intramolecular H-bonding in **2** is diminishing the capacity of the imidazole for base catalysis.

With regards to thermal stability, the introduction of the guanidinium group substantially increased thermal stability to both Dz10–66 and Dz9–86 compared to $Dz9_{25}$–11, which lacked the guanidinium group. This is consistent with other findings on guanidinium-dU modified antisense and triplex forming oligonucleotides that have significantly higher melting temperatures.[48,49] Interestingly, Dz20–49 exhibited minimal thermostability despite having the same percentage of amino-modified dCs (17% *vs.* 18%) and a greater percentage of guanidinium-modified dUs (35% *vs.* 23%). We speculate that incorporation of **2** resulted in the selection of certain favoured sub-sequences that prevented selection for either high thermal activity or high catalytic activity.

Where a selection demands only catalytic activity, one would expect inactive sequences to be eliminated from the gene pool, rather than overpopulate it. Yet in this case, motifs that are barely active appear to be over-represented, including those that contain a greater percentage of modified dAs. This may occur if certain sequences are minimally active but preferentially amplified, such that they prevail because such a selection selects for both catalytic activity and amplifiability. Nevertheless, more work will be needed to fully address such effects. Although sequence dependent amplifiability may be more pronounced in modified selection, it may also effect unmodified catalysts;[50] it is well known that certain unmodified DNA sequences *e.g.* triplet repeats are poorly amplified in a PCR, and thus subtle amplification differences may also determine the outcome of unmodified selections, only to a lesser extent. Tyranny of the small motif, a phenomenon whereby certain small catalytic sequences are found embedded in larger libraries, may also reflect such preferences.

Experimental

Chemicals and reagents

Nucleoside triphosphates **2**, **3**, **4**, and **6** were synthesized according to literature reports[18,41,47,51] and converted to the triphosphates using the method of Ludwig and Eckstein.[52] Nucleoside triphosphate **7** was purchased from Trilink. dGTPα^{32}P was purchased from Perkin Elmer. Chemicals used were purchased from Sigma. Water used in all experiments was autoclaved following DEPC treatment.

Enzymes and proteins

Sequenase Version 2.0 was purchased from USB Corporation. Lambda exonuclease, Taq polymerase and Vent (*exo*-) were purchased from New England Biolabs. Single-Stranded Binding Protein was purchased from Epicentre. SUPERase-IN was purchased from Ambion. RNase A was purchased from Fermentas. pGEM-T Easy was purchased from Promega.

Oligonucleotides

All oligonucleotides were purchased from Integrated DNA Technologies. Biotinylated primer **2** was gel purified. All other oligonucleotides were extracted using phenol : chloroform : isoamyl alcohol 25 : 24 : 1 followed by precipitation with 3% $LiClO_4$ in acetone. The precipitate was centrifuged and the pellet was washed with ethanol (500 µl), dried and dissolved in water. The oligonucleotides were then passed through a G-25 column for desalting.

The following oligonucleotides were used (5′ to 3′):

GAGCTCGCGGGGCGTGCN$_{40}$CTGTTGGTAGGGCCCAACAGACG (**1**) biotin-T$_{20}$GCGTGCC**r**CGTCTGTTGGGCCC (**2**) phosphate-CGTCTGTTGGGCCCTACCA (**3**) GAGCTCGCGGGGCGTGC (**4**) phosphate-ACGACACAGAGCGTGCCCGTCTGTTGGGCCCTACCA (**5**) TTTTTTTTTTTTTTTTTTTTGAGCTCGCGGGGCGTGC (**6**) GAGCTCGCGGGGCGTGCAACGACCCACACGACCTGCGAACCACTAGAGAGCATGACTTGTGGTAGGGCCCAACAGACGGGCACGCTCGTGTTGT (**7**)

Buffers and cocktails

Relevant buffers and cocktails were made according to Hollenstein *et al.* without any modifications.[19] Detailed descriptions of the buffers and cocktails can be found in the ESI.†

Detection of radioactive DNA

Radioactive DNA resolved as bands on denaturing polyacrylamide gels were visualized by first exposing the gels to storage phosphor screens. Low activity gels were exposed overnight. Imaging of the screen was done using a GE Typhoon 9200 Phosphorimager. For data manipulation, the program Imagequant Version 5.2 was used to determine the number of counts associated with a particular band by encompassing the band and representing the counts as a function of image intensity.

In vitro selection

The *in vitro* selection was carried out based on the protocol developed by Hollenstein *et al.* with **2** used in place of **3**.[19] Twenty rounds of selection were performed. A detailed protocol of the selection, cloning, and screening can be found in the ESI.†

Kinetics of native Dz20–49 and modified dA^{X} replacements

The template for Dz20–49 was annealed to primer **2** and extended under the same conditions as the selection rounds with the exception of using dNTPs of varying compositions as noted. The modified nucleoside triphosphate **2** was replaced with **1**, **3**, or **4**. Ten picomoles were bound to the beads and washed as before. Beads were incubated in cleavage buffer (100 µl) at room temperature (24 °C) and portions (5 µL) were removed and quenched at the following time points 5, 20, 40, 60, 90, 120, 180, 240, 300, 366, 420, 540, 1200, 1800, and 5760 min. Rate constants for self-cleavage were calculated using eqn (1) at various conditions as noted below (*e.g.* pH, temperature).

$$P_t = P_\infty \times (1 - e^{-kt}) \quad (1)$$

pH–Rate profile

The self-cleavage rate constants of Dz20–49 at various pH values ranging from 6.0 to 9.0 were determined; cleavage buffers used were composed of 50 mM Tris-HCl, 1 mM EDTA and 200 mM NaCl at pH values of 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, and 9.0. Approximately 1 pmol of extension product was bound to streptavidin beads and washed. After washing, the beads were treated with 40 µl of one of the cleavage buffers and incubated at room temperature (24 °C). Aliquots (5 µL) were removed at

5, 60, 120, 330, 540, 1350 and 1800 min and added to quench buffer formamide loading buffer containing biotin to dissociate the immobilized DNA upon heating (15 μL). The data were used to determine observed cleavage rate constants which are fitted to eqn (2)[45] to calculate putative pK_a values.

$$k_{obs} = \frac{k_{max}}{(1 + 10^{(pK_a - pH)} + 10^{(pH - pK_a)} + 10^{(pK_a - pK_a)})} \quad (2)$$

Standardizing modified PCR templates by autoradiography

Four modified templates based on the DNAzyme were produce using primer extension using dNTPs or a combination of dNTPs and modified dNTPs (see ESI†). The radioactive templates were quantified according to literature.[19] Standardized templates were used in PCRs using either Vent (*exo*-) or Taq polymerase (NEB).

Conclusions

Over the past decade numerous reports have examined the incorporation of modified dNTPs with an eye to enhancing the catalytic potential of DNA. We and others have shown that: (1) excellent incorporation does not always lead to functionality-dependent enhancement of catalysis; (2) in the case of using an 8-imidazolyl-modified deoxyadenosine, a shorter linker does not lead to enhanced catalysis; (3) dXTP design should take into consideration efficient read-through as well as efficient incorporation if the dXTP is intended for *in vitro* selection. To complement dXTP design and facilitate both incorporation and read-through would be the evolution of polymerases that readily accept unnatural nucleotides.[53–59] Future efforts will be applied towards this goal.

Nevertheless, no report has definitively demonstrated which modified dNTPs will actually *enhance* catalysis and which will not. In light of the contrasting results for **3** and **9**, we must appreciate that incorporation, while necessary for functional selection, is clearly insufficient. Here we show that slight differences in dNTP structure, which are not evident from examining the propensity for dNTP incorporation, strongly govern the selection outcome. Indeed this discrete change in linker length produced a DNAzyme that is ~200-fold less active than Dz10–66. Moreover, *a priori* no antecedent report could have informed our decision to use **2** instead of **3** as both were incorporated with nearly identical efficiency. Indeed the choice of **3** in preceding selections was fortuitous as we might have easily chosen to use **2** instead. While some readers might prefer that we report only on catalytic enhancements, we contend it is essential to report this large drop in activity to enhance our current understanding of the nuances surrounding the use of modified nucleotides in combinatorial selection.

Acknowledgements

The authors would like to thank Dr Elena Polishchuk and UBC Biological Services for providing the laboratory environment for the cloning work. Funding for this work was provided by grants from CIHR Operating Funds and NSERC.

References

1 A. D. Ellington and J. W. Szostak, *Nature*, 1990, **346**, 818–822.
2 C. Tuerk and L. Gold, *Science*, 1990, **249**, 505–510.
3 S. K. Silverman, *Org. Biomol. Chem.*, 2004, **2**, 2701–2707.
4 S. K. Silverman, *Nucleic Acids Res.*, 2005, **33**, 6151–6163.
5 S. K. Silverman, *Chem. Commun.*, 2008, (30), 3467–3485.
6 D. A. Baum and S. K. Silverman, *Cell. Mol. Life Sci.*, 2008, **65**, 2156–2174.
7 K. Schlosser and Y. Li, *Chem. Biol.*, 2009, **16**, 311–322.
8 S. Franzen, *Current Opinion in Molecular Therapeutics*, 2010, **12**, 223–232.
9 K. Ohsawa, T. Kasamatsu, J. I. Nagashima, K. Hanawa, M. Kuwahara, H. Ozaki and H. Sawai, *Anal. Sci.*, 2008, **24**, 167–172.
10 A. Shoji, M. Kuwahara, H. Ozaki and H. Sawai, *J. Am. Chem. Soc.*, 2007, **129**, 1456–1464.
11 H. A. Held and S. A. Benner, *Nucleic Acids Res.*, 2002, **30**, 3857–3869.
12 D. M. Perrin, T. Garestier and C. Hélène, *J. Am. Chem. Soc.*, 2001, **123**, 1556–1563.
13 K. Sakthivel and C. F. Barbas, *Angew. Chem., Int. Ed.*, 1998, **37**, 2872–2875.
14 A. V. Sidorov, J. A. Grasby and D. M. Williams, *Nucleic Acids Res.*, 2004, **32**, 1591–1601.
15 R. H. Symons, *Annu. Rev. Biochem.*, 1992, **61**, 641–671.
16 S. Jäger, G. Rasched, H. Kornreich-Leshem, M. Engeser, O. Thum and M. Famulok, *J. Am. Chem. Soc.*, 2005, **127**, 15071–15082.
17 M. Hollenstein, C. Hipolito, C. Lam, D. Dietrich and D. M. Perrin, *Angew. Chem., Int. Ed.*, 2008, **47**, 4346–4350.
18 M. Hollenstein, C. Hipolito, C. Lam and D. M. Perrin, *Nucleic Acids Res.*, 2009, **37**, 1638–1649.
19 M. Hollenstein, C. J. Hipolito, C. H. Lam and D. M. Perrin, *ChemBioChem*, 2009, **10**, 1988–1992.
20 S. E. Lee, A. Sidorov, T. Gourlain, N. Mignet, S. J. Thorpe, J. A. Brazier, M. J. Dickman, D. P. Hornby, J. A. Grasby and D. M. Williams, *Nucleic Acids Res.*, 2001, **29**, 1565–1573.
21 S. E. Lee, J. S. Vyle, D. M. Williams and J. A. Grasby, *Tetrahedron Lett.*, 2000, **41**, 267–270.
22 T. Gourlain, A. Sidorov, N. Mignet, S. J. Thorpe, S. E. Lee, J. A. Grasby and D. M. Williams, *Nucleic Acids Res.*, 2001, **29**, 1898–1905.
23 S. Jäger and M. Famulok, *Angew. Chem., Int. Ed.*, 2004, **43**, 3337–3340.
24 O. Thum, S. Jager and M. Famulok, *Angew. Chem., Int. Ed.*, 2001, **40**, 3990–3993.
25 M. Hocek and M. Fojta, *Org. Biomol. Chem.*, 2008, **6**, 2233–2241.
26 H. Cahova, R. Pohl, L. Bednarova, K. Novakova, J. Cvacka and M. Hocek, *Org. Biomol. Chem.*, 2008, **6**, 3657–3660.
27 P. Capek, H. Cahova, R. Pohl, M. Hocek, C. Gloeckner and A. Marx, *Chem.–Eur. J.*, 2007, **13**, 6196–6203.
28 J. D. Vaught, C. Bock, J. Carter, T. Fitzwater, M. Otis, D. Schneider, J. Rolando, S. Waugh, S. K. Wilcox and B. E. Eaton, *J. Am. Chem. Soc.*, 2010, **132**, 4141–4151.
29 T. Ohbayashi, M. Kuwahara, M. Hasegawa, T. Kasamatsu, T. Tamura and H. Sawai, *Org. Biomol. Chem.*, 2005, **3**, 2463–2468.
30 S. W. Santoro, G. F. Joyce, K. Sakthivel, S. Gramatikova and C. F. Barbas, *J. Am. Chem. Soc.*, 2000, **122**, 2433–2439.
31 M. M. Masud, M. Kuwahara, H. Ozaki and H. Sawai, *Bioorg. Med. Chem.*, 2004, **12**, 1111–1120.
32 H. Macickova-Cahova and M. Hocek, *Nucleic Acids Res.*, 2009, **37**, 7612–7622.
33 J. A. Latham, R. Johnson and J. J. Toole, *Nucleic Acids Res.*, 1994, **22**, 2817–2822.
34 M. Kuwahara, K. Hanawa, K. Ohsawa, R. Kitagata, H. Ozaki and H. Sawai, *Bioorg. Med. Chem.*, 2006, **14**, 2518–2526.
35 T. M. Tarasow, S. L. Tarasow and B. E. Eaton, *Nature*, 1997, **389**, 54–57.
36 B. Seelig and A. Jäschke, *Chem. Biol.*, 1999, **6**, 167–176.
37 B. Seelig, S. Keiper, F. Stuhlmann and A. Jäschke, *Angew. Chem., Int. Ed.*, 2000, **39**, 4576–4579.
38 D. Faulhammer and M. Famulok, *J. Mol. Biol.*, 1997, **269**, 188–202.
39 C. R. Geyer and D. Sen, *Chem. Biol.*, 1997, **4**, 579–593.
40 J. M. Thomas, J. K. Yoon and D. M. Perrin, *J. Am. Chem. Soc.*, 2009, **131**, 5648–5658.
41 C. Lam, C. Hipolito and D. M. Perrin, *Eur. J. Org. Chem.*, 2008, 4915–4923.
42 J. SantaLucia, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 1460–1465.
43 M. Zuker, *Nucleic Acids Res.*, 2003, **31**, 3406–3415.

44 R. Ting, J. M. Thomas, L. Lermer and D. M. Perrin, *Nucleic Acids Res.*, 2004, **32**, 6660–6672.
45 M. Roychowdhury-Saha and D. H. Burke, *RNA*, 2006, **12**, 1846–1852.
46 S. W. Santoro and G. F. Joyce, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 4262–4266.
47 L. Lermer, J. Hobbs and D. M. Perrin, *Nucleosides, Nucleotides Nucleic Acids*, 2002, **21**, 651–664.
48 T. P. Prakash, A. Puschl and M. Manoharan, *Nucleosides, Nucleotides Nucleic Acids*, 2007, **26**, 149–159.
49 V. Roig and U. Asseline, *J. Am. Chem. Soc.*, 2003, **125**, 4416–4417.
50 K. Schlosser, J. C. F. Lam and Y. F. Li, *Nucleic Acids Res.*, 2009, **37**, 3545–3557.
51 T. P. Prakash, R. Krishna Kumar and K. N. Ganesh, *Tetrahedron*, 1993, **19**, 4035–4050.
52 J. Ludwig and F. Eckstein, *J. Org. Chem.*, 1989, **54**, 631–635.
53 A. D. Keefe and S. T. Cload, *Curr. Opin. Chem. Biol.*, 2008, **12**, 448–456.
54 M. Fa, A. Radeghieri, A. A. Henry and F. E. Romesberg, *J. Am. Chem. Soc.*, 2004, **126**, 1748–1754.
55 A. M. Leconte, L. J. Chen and F. E. Romesberg, *J. Am. Chem. Soc.*, 2005, **127**, 12470–12471.
56 F. J. Ghadessy, N. Ramsay, F. Boudsocq, D. Loakes, A. Brown, S. Iwai, A. Vaisman, R. Woodgate and P. Holliger, *Nat. Biotechnol.*, 2004, **22**, 755–759.
57 D. Loakes, J. Gallego, V. B. Pinheiro, E. T. Kool and P. Holliger, *J. Am. Chem. Soc.*, 2009, **131**, 14827–14837.
58 D. Loakes and P. Holliger, *Chem. Commun.*, 2009, 4619–4631.
59 N. Ramsay, A. S. Jemth, A. Brown, N. Crampton, P. Dear and P. Holliger, *J. Am. Chem. Soc.*, 2010, **132**, 5096–5104.
60 R. Ting, J. M. Thomas and D. M. Perrin, *Can. J. Chem.*, 2007, **85**, 313–329.