High-throughput activity profiling of RNA-cleaving DNA catalysts by deoxyribozyme sequencing (DZ-seq)

Maksim V. Sednev*,‡, Anam Liaqat‡ and Claudia Höbartner*

Institute of Organic Chemistry, University of Würzburg, Am Hubland, 97074 Würzburg, Germany

ABSTRACT: RNA-cleaving deoxyribozymes have found broad application as useful tools for RNA biochemistry. However, tedious in-vitro selection procedures combined with laborious characterization of individual candidate catalysts hinder the discovery of novel catalytic motifs. Here, we present a new high-throughput sequencing method, DZ-seq, which directly measures activity and localizes cleavage sites of thousands of deoxyribozymes. DZ-seq exploits A-tailing followed by reverse transcription with an oligo-dT primer to capture the cleavage status and sequences of both deoxyribozyme and RNA substrate. We validated DZ-seq by conventional analytical methods and demonstrated its utility by discovery of novel deoxyribozymes that allow for cleaving challenging RNA targets or the analysis of RNA modification states.

DNA enzymes or deoxyribozymes are single-stranded DNA sequences that form three-dimensional actives sites for the catalysis of chemical reactions. Although deoxyribozymes have not been yet found in nature, in vitro selection procedures allowed the discovery of artificial deoxyribozymes with a vast repertoire of catalytic activities.1-4 A typical *in vitro* selection starts from a combinatorial library of up to ~1015 unique sequences and consists of repetitive rounds of activity screening, selection and amplification.5 Next-generation sequencing (NGS) techniques have been increasingly replacing the low-throughput Sanger method in the analysis of in-vitro selection libraries of aptamers⁶⁻¹¹ and (deoxy)ribozymes. 12-16 The candidate catalysts are usually selected from the NGS data using the relative sequence abundance or sequence enrichment as proxies for the catalytic activity. However, due to PCR and other biases, 17-18 the candidate catalysts have to be validated by conventional low-throughput biochemical assays. Novel NGSbased strategies provided direct activity measurement for large mutant libraries of RNA-cleaving ribozymes, 19-20 DNA-cleaving deoxyribozymes21 and in vitro selection libraries of aminoacylating ribozymes,22 but RNA-cleaving deoxyribozymes still lack a comparable technique.

In our efforts to develop tools for site-specific detection of epigenetic modifications in RNA, we performed *in vitro* selections to identify RNA-cleaving deoxyribozymes that differentiate between modified and unmodified RNA.²³⁻²⁵ Recently, we reported DNAzymes that distinguish between unmodified cytosine (C), 3-methylcytidine (m³C), N⁴-methylcytidine (m⁴C) and 5-methylcytidine (m⁵C).²⁵ Using Sanger sequencing and analysis of sequence enrichment in the NGS data with subsequent characterization of candidate catalysts by gel shift-based kinetic assays, we found few DNA enzymes with the desired properties. Screening of further deoxyribozymes was time-consuming and laborious as every candidate sequence had to be individually assayed for its cleavage activity.

To facilitate the discovery of further RNA-cleaving deoxyribozymes, we implemented deoxyribozyme sequencing (DZ-seq), an NGS-based strategy that allows massively parallel quantitation of RNA cleavage activity directly from sequencing data. In the first step of DZ-seq, the in-vitro selection library with the attached 5'-sequencing adapter was subjected to self-cleavage (Figure 1). Next, the 2',3'-cyclic phosphate residues resulting from the intramolecular attack of an RNA 2'-hydroxyl group on the adjacent phosphodiester linkage,26 were removed by T4 PNK, and both cleaved and uncleaved species were extended with poly(A) tails using E. coli poly(A) polymerase. The resulting poly(A) tails were then used as primer binding sites for the anchored oligo(dT) primer (D4, Table S1) which introduced the second sequencing adapter to the resulting cDNA during reverse transcription. After this step, the cDNA contained information about sequences of deoxyribozyme and RNA substrate, cleavage status of RNA and position of the cleavage site. Finally, the cDNA was used as a template in a PCR with two primers (D5 and D6) which introduced sequencing barcodes to produce sequencingready fragments.

We applied DZ-seq to analyze endpoint activity (10 mM Mg²+, 6 h, 37 °C) of in-vitro selection pools AK, AL, AM and AN which have been trained previously to cleave NC-, Nm³C-, Nm³C- and Nm⁵C-dinucleotide containing all-RNA substrates, respectively (N = A, C, G, U).²5 We prepared in total 16 sequencing libraries by combining eight deoxyribozyme libraries from rounds 7 and 18 of each in-vitro selection with four RNA substrates (R1–R4, Table S2) in various combinations. Eight sequencing libraries contained DNA pools ligated to the unmodified NC-substrate R1. Six libraries were prepared by combination of DNA pools from AL, AM and AN selections with their corresponding cognate RNA substrates R2–R4, respectively. Further two sequencing libraries contained DNA pools from the selection AK and Nm⁵C-RNA R4.

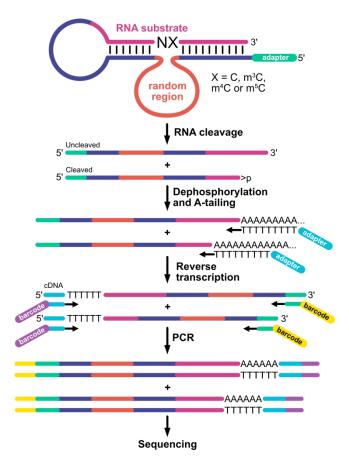


Figure 1. Outline of DZ-seq. ">p" denotes 2',3'-cyclic phosphate residue.

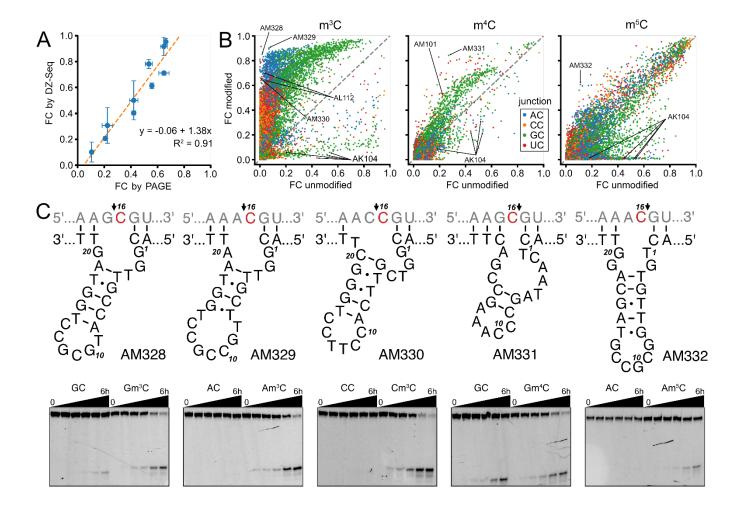
The sequencing-ready DNA fragments were subjected to NGS and a total number of ~220 million reads with ~10–20 million reads per library were obtained (Table S3). The raw reads were trimmed of binding arm and connecting loop sequences and dereplicated to obtain unique sequences and their counts. For every unique combination of deoxyribozyme and RNA substrate we determined the total numbers of cleaved and intact reads and the position of the cleavage site. For about 40 000 deoxyribozymes at least one such combination had at least 100 reads. This data was then used for calculation of cleaving activity which was expressed as fraction of cleaved reads (FC).

We validated the FC values obtained from DZ-seq by conventional PAGE-based kinetic assays. To this end, we chose ten deoxyribozymes AM301, AM319-AM327 (D4-D13) with varying predicted activities and tested them individually in cis-cleavage reactions, i.e. when the RNA substrate was covalently attached to the deoxyribozyme. DZ-seq correctly identified the position of the cleavage site for all ten cleavage reactions, but systematically overestimated FC values in comparison to PAGE (Figure 2A and

Table S₄). The overestimation can be attributed to additional cleavage events which occur during the dephosphorylation and A-tailing steps of DZ-seq as the corresponding reaction buffers contain Mg^{2+} ions and thus promote deoxyribozyme-catalyzed RNA cleavage. Nevertheless, the FC values obtained from DZ-seq displayed a good linear correlation with the PAGE assay (R² = 0.91, Figure 2A). Therefore, the obtained DZ-seq dataset provides a quantitative insight into catalytic activity of tens of thousands of RNA-cleaving deoxyribozymes.

To illustrate the distribution of cleaving activities and specificities of DNA enzymes, we plotted FC values for the m³C-, m⁴C- or m⁵C-modified RNA substrates against those for the unmodified RNA (Figure 2B). The top left region of the plot for the m³C modification is densely populated corresponding to numerous highly active DNA enzymes which are also highly specific for m3C. In contrast, the majority of data points in the plots for m4C and m5C is concentrated along the diagonal line corresponding to low specificity of the DNAzymes in the AM and AN selection pools for the modified substrate. This observation is consistent with the stronger influence of the positively charged and base pair-disrupting m3C modification on secondary structure of nucleic acids as compared to m4C and m5C.27 The high number of very active m3C-specific deoxyribozymes may indicate that m3C-containing RNA is an "easier" target for DNA-catalyzed cleavage than are m4C-, m5C-modified or unmodified RNA. Interestingly, few active deoxyribozymes that were inhibited by m3C survived selection pressure in the AL selection as several data points appear at the very right bottom of the corresponding plot.

DZ-seq interrogates the activity of RNA-cleaving deoxyribozymes in intramolecular reactions (in-cis) with the RNA substrates. For detection of RNA modifications an intermolecular arrangement of the DNA enzyme and the RNA substrate (in-trans) is more advantageous because it allows straightforward examination of RNA modifications without additional laborious experimental steps. Cis- and trans-activities do not necessarily correlate with each other as the tertiary interactions between the connecting loop and the catalytic core may be critical for proper folding into a functional catalytic structure.²⁸ Surprisingly, some of the deoxyribozymes used for the validation of the DZ-seq results showed even higher FC values when they were tested with the corresponding RNA substrates in-trans under pseudo-first order conditions (Table S4). However, most of the data points in the DZ-seq dataset agreed well with the in-trans selectivity profiles of the previously characterized deoxyribozymes (exemplary data points are highlighted in Figure 2B for deoxyribozymes AL112, AM101 and AK104).25 Hence, DZ-seq can be used for identification of deoxyribozymes that differentiate between modified and unmodified RNA substrates in intermolecular cleavage reactions.



³C, m⁴C or m⁵C). (C) Predicted secondary structures of deoxyribozymes AM₃₂8–AM₃₃2 (top) and their transcleavage activity with 3'-fluorescently labeled modified and unmodified RNA substrates analyzed by gel-based kinetic assays (bottom). Modified cytosine site is highlighted in red. Cleavage sites reported by DZ-seq are indicated by arrows. Cleaved fragments containing m³C migrate slower in PAGE, i.e. are shifted upwards due to the positive charge at the nucleobase.

To this end, we picked candidate sequences AM328-AM332 (D14-D18) which were predicted by DZ-seq to preferentially cleave Gm3C-, Am3C-, Cm3C-, Gm4C- and Am5Ccontaining RNAs, respectively (Figure 2C). Among them, Gm⁴C- and Am⁵C-contexts were particularly challenging for specific cleavage with the previously reported DNA enzymes.25 The new candidate DNA enzymes were individually tested in-trans in gel-based kinetic assays with their cognate and the corresponding unmodified RNA substrates (Table S₅). As expected, all five selected deoxyribozymes preferentially cleaved the modified RNA substrates, although in case of AM332 the FC value reported by PAGE (20%) was lower than that of DZ-seq (61%), what may be attributed to the influence of the connecting loop. Noteworthy, the DNA sequences here chosen for further analysis were low-abundant with no more than 5-300 counts per million (CPM) in the final selection rounds (Table S6), and thus most likely would have never been identified using conventional Sanger sequencing which typically yielded sequences with much higher abundance (5 000-30 000

CPM). Also, analysis of NGS data solely based on abundance and/or enrichment would have missed these candidates.

In fact, DZ-seq revealed a poor correlation between endpoint cleaving activity and abundance of deoxyribozymes in the final rounds of in-vitro selections AK, AL, AM and AN (Figure S2). A representative example is deoxyribozyme AM301 (Figure 3A) which displayed FC values that were among the highest found in the DZ-seq dataset. Specifically, AM301 cleaved CG junction in Am5C-, CC-, Cm4C-, Cm5C- Um4C- and Um5C-RNA substrates with very high yields (FCs of 80-98% after 6 h according to the PAGE assay, see Table S₅). Despite these favorable properties, AM301 was depleted in most of the selection pools and its abundance in the final rounds was very low (~90 CPM). This indicates that cleaving activity is not the only decisive factor for the outcome of the in-vitro selection, and that PCR amplification and/or ligation may distort the pool composition towards less active species.²⁹ In this respect, it would be compelling to analyze the similar in-vitro selections of the pre-NGS era³⁰⁻³¹ with DZ-seq for the presence

of depleted but nevertheless highly active (deoxy)ribozyme motifs.

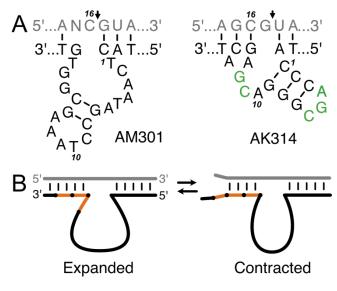


Figure 3. (A) Predicted secondary structures of deoxyribozymes AM301 and AK314. (B) Contraction of a deoxyribozyme's catalytic core due to strand slippage on tandem repeats. Repeating elements are shown as orange segments.

DZ-seq uncovered another low-abundant (< 18 CPM) and moderately active (FCs up to 60% after 6 h, Table S5) deoxyribozyme which had unusually short 13-nt catalytic core emerged from the originally 20-nt randomized region of the starting selection library (AK314, Figure 3A).25 AK314 possesses the structural features that are characteristic for the 8-17 family: a 3 bp stem flanked by an AGC triple in the apical loop and a CG-dinucleotide in the 3'-direction of the stem (shown green in Figure 3A).32 In fact, the sequence and the secondary structure of AK314 are identical up to a single base pair swap to those of the 8-17 variant discovered by Santoro et al.30 The unusually short length of AK314 raised the question about the mechanism that led to a such significant contraction of the in-vitro selection library. Screening of the sequencing dataset afforded a more abundant variant of AK314 with an additional TTCA fragment at the 3'-end of the catalytic core (AK314L, Table S1). Coincidentally, the 5'-end of the adjacent binding arm started with two further consecutive TTCA fragments. The resulting tandem repeat (TTCA)3 is prone to strand slippage33 and therefore might have induced contraction of AK314L to form AK314 during the PCR step of the in-vitro selection. Moreover, the same repeat may provide possibility for the formation of an alternative fold so that the TTCA fragment of the core is displaced into the binding arm and the actual catalytic core of AK314L is identical to that of AK314 (Figure

In conclusion, in the present study we reported DZ-seq, a high-throughput NGS-based method suitable for profiling activity of RNA-cleaving deoxyribozymes in in-vitro selection libraries. DZ-seq provided reliable information on cleavage activity and cleavage sites of thousands of deoxyribozymes. The novel method was successfully applied for

discovery of rare DNA catalytic motifs that were inaccessible by conventional methods. We expect DZ-seq can be applied for comprehensive examination of in-vitro selection or other combinatorial libraries and thus enable rapid discovery of catalysts with desired properties, e. g. deoxyribozymes specifically cleaving a modified RNA. Since DZ-seq should be indifferent towards the type of nucleic acid catalyst (i.e., deoxyribozyme or ribozyme), DZ-seq can be also applicable for discovery of novel RNA-cleaving motifs.

ASSOCIATED CONTENT

Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org.

Detailed methods, sequences, NGS data analysis, representative gel images, abundance plots, tables with abundance and activity data.

AUTHOR INFORMATION

Corresponding Authors

*maksim.sednev@uni-wuerzburg.de *claudia.hoebartner@uni-wuerzburg.de

ORCID

Maksim V. Sednev: <u>0000-0001-6746-1598</u> Claudia Höbartner: <u>0000-0002-4548-2299</u>

Author Contributions

[‡]M.V.S. and A.L. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

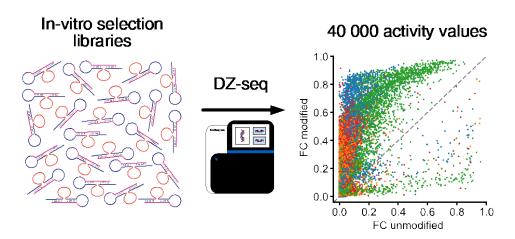
This work was supported by the DFG (SPP1784) and by the ERC (grant no. 682586). M.V.S. thanks the Graduate School of Life Sciences at the University of Würzburg for a Postdoc Plus fellowship. A.L. acknowledges funding by a PhD scholarship from the German Academic Exchange Service (DAAD). Illumina sequencing was performed at the Core Unit Systems Medicine and at the Biocenter of the University of Würzburg.

REFERENCES

- (1) Silverman, S. K., Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes. *Trends Biochem. Sci.* **2016**, *41* (7), 595–609.
- (2) Hollenstein, M., DNA Catalysis: The Chemical Repertoire of DNAzymes. *Molecules* **2015**, 20 (11), 20777–20804.
- (3) Schlosser, K.; Li, Y., Biologically Inspired Synthetic Enzymes Made from DNA. *Chem. Biol.* **2009**, *16* (3), 311–322.
- (4) Micura, R.; Höbartner, C., Fundamental studies of functional nucleic acids: aptamers, riboswitches, ribozymes and DNAzymes. *Chem. Soc. Rev.* **2020**, *49* (20), 7331–7353.
- (5) Jijakli, K.; Khraiwesh, B.; Fu, W.; Luo, L.; Alzahmi, A.; Koussa, J.; Chaiboonchoe, A.; Kirmizialtin, S.; Yen, L.; Salehi-Ashtiani, K., The in vitro selection world. *Methods* **2016**, *106*, *3*–13.
- (6) Rosenthal, M.; Pfeiffer, F.; Mayer, G., A Receptor-Guided Design Strategy for Ligand Identification. *Angew. Chem. Int. Ed.* **2019**, 58 (31), 10752–10755.
- (7) Schmitz, A.; Weber, A.; Bayin, M.; Breuers, S.; Fieberg, V.; Famulok, M.; Mayer, G., A SARS-coV-2 spike binding DNA aptamer that inhibits pseudovirus infection by an RBD-independent mechanism, *Angew. Chem. Int. Ed.* **2021**, *60* (18), 10279-10285.

- (8) Kacherovsky, N.; Yang, L. F.; Dang, H. V.; Cheng, E. L.; Cardle, I. I.; Walls, A. C.; McCallum, M.; Sellers, D. L.; DiMaio, F.; Salipante, S. J.; Corti, D.; Veesler, D.; Pun, S. H., Discovery and Characterization of Spike N-Terminal Domain-Binding Aptamers for Rapid SARS-CoV-2 Detection. *Angew. Chem. Int. Ed.* **2021**, *60* (39), 21211–21215.
- (9) Schütze, T.; Wilhelm, B.; Greiner, N.; Braun, H.; Peter, F.; Mörl, M.; Erdmann, V. A.; Lehrach, H.; Konthur, Z.; Menger, M.; Arndt, P. F.; Glökler, J., Probing the SELEX Process with Next-Generation Sequencing. *PLOS ONE* **2011**, *6* (12), e29604.
- (10) Blind, M.; Blank, M., Aptamer Selection Technology and Recent Advances. *Mol. Ther. Nucleic Acids* **2015**, *4*, e223
- (11) Liu, M.; Yin, Q.; Chang, Y.; Zhang, Q.; Brennan, J. D.; Li, Y., In Vitro Selection of Circular DNA Aptamers for Biosensing Applications. *Angew. Chem. Int. Ed.* **2019**, *58* (24), 8013–8017.
- (12) Streckerová, T.; Kurfürst, J.; Curtis, E. A., Single-round deoxyribozyme discovery. *Nucleic Acids Res.* **2021**, *49* (12), 6971–6981.
- (13) Zhang, C.; Li, Q.; Xu, T.; Li, W.; He, Y.; Gu, H., New DNA-hydrolyzing DNAs isolated from an ssDNA library carrying a terminal hybridization stem. *Nucleic Acids Res.* **2021**, *49* (11), 6364–6374.
- (14) Popović, M.; Ellingson, A. Q.; Chu, T. P.; Wei, C.; Pohorille, A.; Ditzler, M. A., In vitro selections with RNAs of variable length converge on a robust catalytic core. *Nucleic Acids Res.* **2021**, *49* (2), 674–683.
- (15) Inomata, R.; Zhao, J.; Miyagishi, M., Zn²⁺-dependent DNAzymes that cleave all combinations of ribonucleotides. *Commun. Biol.* **2021**, 4 (1), 221.
- (16) Ameta, S.; Winz, M.-L.; Previti, C.; Jäschke, A., Next-generation sequencing reveals how RNA catalysts evolve from random space. *Nucleic Acids Res.* **2014**, *42* (2), 1303–1310.
- (17) Tolle, F.; Wilke, J.; Wengel, J.; Mayer, G., By-Product Formation in Repetitive PCR Amplification of DNA Libraries during SELEX. *PLOS ONE* **2014**, *9* (12), e114693.
- (18) Komarova, N.; Kuznetsov, A., Inside the Black Box: What Makes SELEX Better? *Molecules* **2019**, *24* (19), 3598.
- (19) Kobori, S.; Nomura, Y.; Miu, A.; Yokobayashi, Y., Highthroughput assay and engineering of self-cleaving ribozymes by sequencing. *Nucleic Acids Res.* **2015**, *43* (13), e85.
- (20) Kobori, S.; Yokobayashi, Y., High-Throughput Mutational Analysis of a Twister Ribozyme. *Angew. Chem. Int. Ed.* **2016**, *55* (35), 10354–10357.

- (21) Dhamodharan, V.; Kobori, S.; Yokobayashi, Y., Large Scale Mutational and Kinetic Analysis of a Self-Hydrolyzing Deoxyribozyme. *ACS Chem. Biol.* **2017**, *12* (12), 2940–2945.
- (22) Pressman, A. D.; Liu, Z.; Janzen, E.; Blanco, C.; Müller, U. F.; Joyce, G. F.; Pascal, R.; Chen, I. A., Mapping a Systematic Ribozyme Fitness Landscape Reveals a Frustrated Evolutionary Network for Self-Aminoacylating RNA. *J. Am. Chem. Soc.* 2019, *141* (15), 6213–6223.
- (23) Sednev, M. V.; Mykhailiuk, V.; Choudhury, P.; Halang, J.; Sloan, K. E.; Bohnsack, M. T.; Höbartner, C., N^6 -Methyladenosine-Sensitive RNA-Cleaving Deoxyribozymes. *Angew. Chem. Int. Ed.* **2018**, 57 (46), 15117–15121.
- (24) Liaqat, A.; Stiller, C.; Michel, M.; Sednev, M. V.; Höbartner, C., *N*⁶-Isopentenyladenosine in RNA Determines the Cleavage Site of Endonuclease Deoxyribozymes. *Angew. Chem. Int. Ed.* **2020**, 59 (42), 18627–18631.
- (25) Liaqat, A.; Sednev, M. V.; Stiller, C.; Höbartner, C., RNA-Cleaving Deoxyribozymes Differentiate Methylated Cytidine Isomers in RNA. *Angew. Chem. Int. Ed.* **2021**, *60* (35), 19058–19062.
- (26) Silverman, S. K., In vitro selection, characterization, and application of deoxyribozymes that cleave RNA. *Nucleic Acids Res.* **2005**, 33 (19), 6151–6163.
- (27) Mao, S.; Haruehanroengra, P.; Ranganathan, S. V.; Shen, F.; Begley, T. J.; Sheng, J., Base Pairing and Functional Insights into *N*³-Methylcytidine (m³C) in RNA. *ACS Chem. Biol.* **2021**, *16* (1), 76–85.
- (28) De la Peña, M.; Gago, S.; Flores, R., Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. *EMBO J.* **2003**, 22 (20), 5561–5570.
- (29) Takahashi, M.; Wu, X.; Ho, M.; Chomchan, P.; Rossi, J. J.; Burnett, J. C.; Zhou, J., High throughput sequencing analysis of RNA libraries reveals the influences of initial library and PCR methods on SELEX efficiency. *Sci. Rep.* **2016**, *6* (1), 33697.
- (30) Santoro, S. W.; Joyce, G. F., A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci.* **1997**, *94* (9), 4262–4266.
- (31) Pan, T.; Uhlenbeck, O. C., A small metalloribozyme with a two-step mechanism. *Nature* **1992**, 358 (6387), 560–563.
- (32) Schlosser, K.; Li, Y., A Versatile Endoribonuclease Mimic Made of DNA: Characteristics and Applications of the 8–17 RNA-Cleaving DNAzyme. *ChemBioChem* **2010**, *11* (7), 866–879.
- (33) Shinde, D.; Lai, Y.; Sun, F.; Arnheim, N., Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. *Nucleic Acids Res.* **2003**, *31* (3), 974–980.



TOC Graphic