RNA-Cleaving Deoxyribozymes Differentiate Methylated Cytidine Isomers in RNA

Anam Liaqat*, Maksim V. Sednev+, Carina Stiller, and Claudia Hübner*

In memory of Professor Siegfried Hünig

Abstract: Deoxyribozymes are emerging as modification-specific endonucleases for the analysis of epigenetic RNA modifications. Here, we report RNA-cleaving deoxyribozymes that differentially respond to the presence of natural methylated cytidines, 3-methylcytidine (m3C), N4-methylcytidine (m4C), and 5-methylcytidine (m5C), respectively. Using in vitro selection, we found several DNA catalysts, which are selectively activated by only one of the three cytidine isomers, and display 10- to 30-fold accelerated cleavage of their target m3C-, m4C- or m5C-modified RNA. An additional deoxyribozyme is strongly inhibited by any of the three methylcytidines, but effectively cleaves unmethylated RNA. The m3C-detecting deoxyribozymes are programmable for the interrogation of natural RNAs of interest, as demonstrated for human mitochondrial tRNAs containing known m3C and m5C sites. The results underline the potential of synthetic functional DNA to shape highly selective active sites.

Posttranscriptional nucleotide modifications are indispensable for the functional diversity of cellular RNA. More than 170 types of modified nucleotides are known,[1] but detection and quantification remain challenging. Transcriptome-wide maps are available for some modifications,[2] but the predicted and quantification remain challenging. Transcriptome-wide maps are available for some modifications,[2] but the predicted and quantification remain challenging. Transcriptome-wide maps are available for some modifications,[2] but the predicted and quantification remain challenging. Transcriptome-wide maps are available for some modifications,[2] but the predicted and quantification remain challenging. Transcriptome-wide maps are available for some modifications.[2] The finding of m6A- and i6A-sensitive deoxyribozymes inspired the search for DNA catalysts that target other RNA modifications. Deoxyribozymes that are able to differentiate the three natural monomethylated cytidine isomers 3-methylcytidine (m3C), N4-methylcytidine (m4C) and 5-methylcytidine (m5C), and distinguish them from unmethylated cytidine, are of particular interest. These methylated cytidines exhibit different structural properties depending on the location of the methyl group (Figure 1a). The methyl group in m3C disrupts Watson–Crick base pairing and confers a positive charge to the nucleobase.[7] As a result, m3C may promote the formation of alternative secondary structures. Similarly to m4A, the methyl group in m4C is attached to the exocyclic amino group, and depending on its conformation, m4C can either retain (in anti-conformation) or disrupt Watson–Crick base pairing (in syn-conformation).[8] The anti-conformation is enforced in a base-paired structure but destabilized by steric repulsion (Figure 1a). In contrast to m3C, the methyl group in m5C does not disturb base pairing directly but may

Figure 1. a) Structures of three natural methylated cytidines and their structural consequences. m3C blocks the Watson–Crick (WC) edge and confers a positive charge to the nucleobase. m4C retains WC base pairing in anti-conformation which is destabilized by steric clash between 5-H and N4-Me. In syn-conformation, m4C hinders WC base pairing. m5C enhances hydrophobicity in the major groove and improves base stacking. b) Design of in vitro selection libraries. Splint for ligation and primer binding sites are indicated. Detailed selection is shown in Figure S1.

[1] A. Liaqat,*,[1] Dr. M. V. Sednev,[1] C. Stiller, Prof. Dr. C. Hübner* Institute of Organic Chemistry, University of Würzburg Am Hubland, 97074 Würzburg (Germany) E-mail: claudia.hoebartner@uni-wuerzburg.de

[2] These authors contributed equally to this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202106517.

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
interact with neighboring nucleotides via the C–H-edge. The m^3C modification is known to enhance base stacking.\textsuperscript{39} Targeting these methylated cytidine isomers also addresses fundamental questions on scope and plasticity of DNA catalysis.

To identify modification-specific catalysts, we performed gel-based in vitro selection experiments\textsuperscript{44} (arbitrarily named AL, AM and AN, Figure 1b) using an N\textsubscript{D} DNA library and three synthetic RNA substrates, each containing one of the three monomethylated cytidines (R2-R4, Table S1). Negative selection rounds were carried out with unmodified RNA (R1) to enhance the selectivity for the modified RNA. In this way, DNA enzymes that can cleave both unmodified and modified RNA were eliminated (Figure S1). In parallel, we sought DNA enzymes that preferentially cleave unmodified RNA and are inhibited by m^3C (AK selection). In this case, the positive selection was performed with R1 and the negative selection used an equimolar mixture of R2-R4. To evolve potentially universal DNA catalysts that respond to the cytidine modifications within all four possible dinucleotide sequence motifs NC, the RNA substrates contained a degenerate nucleotide upstream of the target cytidine.

After 18 rounds of selection, the enriched DNA pools were cloned, and 10 clones of each selection were sequenced. We identified 36 unique sequences (D6-D41, Table S2), were cloned, and 10 clones of each selection were sequenced. We identified 36 unique sequences (D6-D41, Table S2), and are inhibited by m^3C (AK selection). In this case, the positive selection was performed with R1 and the negative selection used an equimolar mixture of R2-R4. To evolve potentially universal DNA catalysts that respond to the cytidine modifications within all four possible dinucleotide sequence motifs NC, the RNA substrates contained a degenerate nucleotide upstream of the target cytidine.

After 18 rounds of selection, the enriched DNA pools were cloned, and 10 clones of each selection were sequenced. We identified 36 unique sequences (D6-D41, Table S2), suggesting a high diversity in the selection libraries. To obtain further insights into the library composition and to potentially identify further DNAzyme candidates, we applied next-generation sequencing (NGS) for a deeper analysis of AK, AL, AM and AN selection pools. In total, eight libraries were sequenced, including rounds 7 and 18 from each of the four in vitro selections. Deoxyribozymes specific for m^3C, m^4C or m^5C should be enriched in selection libraries AL, AM or AN, respectively, and depleted in selection AK. Conversely, the deoxyribozymes inhibited by a modification would be highly enriched in selection AK and depleted in selections AL, AM or AN. In this way, using log\textsubscript{2} fold change in abundance (log\textsubscript{2}F\textsubscript{CA}) between selection rounds 7 and 18 we identified nine additional candidate DNA sequences (D42-D50, Table S2).\textsuperscript{10}

Overall, 45 DNA sequences were then tested individually for their ability to cleave sixteen 3'-fluorescently labeled RNA substrates R5-R20 in independent cleavage reactions. The degenerate position was substituted with one of the four canonical ribonucleotides. Together R5-R20 represented every possible combination of four different junctions and four cytidine variants. After removing the 5'-overhang and 3'-loop region, 28 deoxyribozymes retained trans-cleaving activity (Table S3).

Using RNase T1 digestion and alkaline hydrolysis ladders, we determined the cleavage site for every active DNA enzyme. In accordance with the initial design, all tested deoxyribozymes cleaved the RNA substrate near the target cytidine (Table S3). The activity data partially supported our hypothesis that differences in enrichment in the selection libraries correlate with the specificity of deoxyribozymes. The correlation was good for DNAzymes found in the AK, AL and AM selections. However, poor correlation was observed for the selection AN, as the majority of the AN DNA enzyme candidates cleaved both m^3C-modified and unmodified RNA nearly to the same extent despite large differences in fold change abundance values (Table S3).

Interestingly, the inspection of predicted secondary structures revealed that none of the active deoxyribozymes contained the 8–17 motif, which had been repeatedly found in independent in vitro selections of RNA-cleaving deoxyribozymes.\textsuperscript{44,11} Its absence is especially surprising for AK selection, which used unmodified RNA as substrate. Probably, the 8–17 variants, that were initially present in the selection pools, were not particularly selective for either modified or unmodified RNA, and therefore were depleted during negative selection rounds. None of previously reported 8–17 variants was particularly efficient at cleaving N/C junctions.\textsuperscript{11}

With respect to the degenerate position in the RNA substrate that was introduced to evolve for sequence generality, only deoxyribozyme AL112 distinguished a cytidine modification almost equally well in all four NC sequence contexts. On the other hand, many DNA catalysts were good at differentiating methylated from unmethylated RNA in one or two sequence contexts. In fact, we identified a corresponding deoxyribozyme for almost every combination of modification and junction (Table S3). For the UC junction in particular, we found that AL112, AM101, AN05 and AK104 (Figure 2a) deoxyribozymes together achieved efficient differentiation for all three monomethylated cytidines. Among these, AL112, AM101 and AN05 preferentially cleaved Um^3C-, Um^4C- and Um^5C-containing RNAs, respectively (fraction cleaved (FC) for the corresponding modified RNA FC\textsubscript{Um^3C} ≥ 70\%, and unmodified RNA FC\textsubscript{Um^3C} ≤ 30\% after 6 h, see Figure 2b and Table S3), while AK104 efficiently cleaved only the UC substrate and was inhibited by all the modified nucleotides (FC\textsubscript{Um^3C} = 85\%, FC\textsubscript{Um^4C} ≈ 0 after 6 h). Therefore, this set of deoxyribozymes was further characterized.

First, we confirmed the cleavage sites determined by the gel-based assay (Figure S3) by mass spectrometry. Consistent with the general mechanism proposed for DNA-catalyzed RNA cleavage,\textsuperscript{12} the cleavage products for all four selected catalysts contained 2',3'-cyclic phosphate and 5'-OH termini (Figure S4).

Deoxyribozyme AL112 preferentially cleaved m^3C-modified RNA in all four NC sequence contexts with the highest cleavage yield for the UC junction (FC\textsubscript{Um^3C} = 70\%, Figure S5). In the UC context, AL112 showed a ca. 30-fold higher k\textsubscript{cat} value for m^3C-RNA compared to unmodified RNA (Figure 2b). The cleavage site was situated directly upstream of the modified nucleotide m^3C. Analysis of the predicted secondary structure showed that presence of m^3C might be required for disruption of the dG1-rC16 base pair that otherwise inhibited RNA cleavage for the NC- and Nm^3C- and Nm^5C-RNA substrates (Figure 2b, Table S3).

Deoxyribozyme AM101 was specific for m^3C and cleaved Um^3C-RNA ca. 10 times faster than UC-RNA. The cleavage site of AM101 was one nucleotide downstream of the modification site. Analysis of the secondary structure revealed the presence of a dG18-rU15 wobble base pair, which might be responsible for selectivity of AM101 for the UC junction (Figure 2a). However, in the CC sequence...
The deoxyribozyme AN05 for cleavage of mA-C-modified RNA was found in Sanger sequencing data of the AN selection. Analysis of NGS data showed that it was also present in the AK selection, where it was strongly depleted from round 7 to round 18 (Table S3). In the UC context, AN05 cleaved only mA-C-modified and unmodified substrates (FCmA-C = 90%, FCUC = 33%), and was completely inhibited by mA-C and mA-C. The kinetic discrimination ability of AN05 was comparable to AM101: AN05 cleaved UCmC-RNA ca. 10 times faster than the corresponding unmodified substrate (Figure 2b, S8, and Table S3). The cleavage site of AN05 was located one nucleotide upstream of the target cytosine. The predicted secondary structure contained the dG1-rC16 base pair that might stabilize an active site which is responsible for preference of AN05 to mA-C and C (Figure 2a). In case of mA-C and mA-C, this base pair might be destabilized or disrupted and thus lead to inhibition of cleavage.

The recognition of mA-C by AN05 raised the question about the oxidized variants 5-hydroxymethylcytosin (hm5C) and 5-formylcytosin (f5C), previously reported in RNA as products of TET2 and ALKBH1 enzymes. AN05 cleaved hm5C-containing RNA slightly slower but to a similar extent as mA-C RNA, while f5C was comparable to unmodified RNA (Figure S9). Surprisingly, this result is in line with bisulfite sequencing, which is unable to distinguish mA-C from hm5C, and f5C from C. Given the functional importance of hm5C and f5C in natural RNA, future in vitro selection experiments may be directed to evolve deoxyribozymes specific for the individual oxidized mA-C variants.

Deoxyribozyme AK104 efficiently cleaved unmodified RNA (FC/C21 = 85% after 6 h) and was inhibited by all cytosine modifications in AC, GC and UC sequence contexts. In this case, the cleavage site was situated directly upstream of the target cytosine. However, when the RNA substrates with the CC junction were used, AK104 cleaved the mA-C-containing RNA with a moderate yield (FC CmA-C ca. 30%) and was inactive with the unmodified substrate. The cleavage site for the CmA-C-RNA was shifted one nucleotide upstream (Figure S10, Table S3). This odd behavior can be explained by disruption of the dG1-rC16 base pair by mA-C, GmA-C- and UmC-RNAs and formation of an alternative GC base pair between dG1 and rC15 in case of CmA-C-RNA. The new base pair leads to repositioning of the AK104 active site and partial rescue of the cleavage activity. On the other hand, both rC15 and rC16 in the CC-substrate can form base pairs with dG2 and dG1, respectively, thereby disrupting the functional active site and abolish the catalytic activity of AK104. Thus, AK104 is a RNA-cleaving DNA enzyme for DC junctions (D = A,G,U), with high selectivity for unmodified RNA even under accelerated cleavage conditions at elevated Mg2+ concentrations. None of the modified RNAs was cleaved, and RNAs with rC16 changed to U, G, or A were also not cleaved (Figures S11, S12, S13).

Next, we examined the deoxyribozymes in the more challenging situation when all three RNA isomers (UmA-C-, UmG-C- and UmC-RNAs) are present at the same time. For this experiment, we labeled the three methylated UC-RNA substrates (R12, R16 and R20) at the 2'-OH of rA9 with fluorescent N6-(6-aminohexyl)-ATP derivatives, carrying...
6FAM, Cy3, and ATTO-647N, respectively. An equimolar mixture of the three labeled RNAs was then incubated with 10-fold excess of the AL112, AM101, or AN05 deoxyribozyme, respectively (Figure 3a). The cleavage reactions were analyzed by denaturing PAGE and imaged using a dedicated channel for each fluorescent label (Figure 3b). As expected, AL112 showed preferential cleavage of Um3C-RNA with negligible signals in the m4C- and m5C-channels. At the same time, AM101 and AN05 cleaved only their cognate substrates with no detectable cleavage products in other channels. The results demonstrated that the specificity of each DNAzyme for its target modification was retained, and this collection of deoxyribozymes can be used to examine the identity of methylated cytidines in UC motifs. In this regard, m4C-specific RNA cleavage is particularly important, as in contrast to m3C and m5C, there are no specific chemical reactions reported for the detection of m4C in RNA.

Next, we tested deoxyribozymes AL112, AN05, and AK104 on longer RNA substrates with complex secondary structures. To this end, we chose human mt-tRNAThr, which contains a known m3C in the anticodon loop (C32, Figure 4a), and mt-tRNA Glu with a known m5C at C49 (Figure 4c). The single-modified tRNAs were assembled from synthetic RNA fragments by splinted ligation with T4 RNA ligase, and the binding arms of deoxyribozymes were designed to cleave the tRNAs upstream of the corresponding modification sites. The results of the cleavage reactions for unmodified and modified tRNAs were compared (Figure 4b,d). AK104 cleaved both unmodified tRNAs with a high yield (> 90% after 18 h) and was nearly completely inhibited by m3C and by m5C. AL112 and AN05 cleaved m3C-modified mt-tRNAThr and m5C-modified mt-tRNA Glu, respectively, both somewhat less efficiently than their cognate substrates R12 and R20 (58% and 42% after 18 h), but the unmethylated tRNAs did not undergo any significant cleavage. Therefore, AL112 and AN05 can faithfully identify m3C- and m5C-modified tRNA, respectively.

In summary, we have identified and characterized a series of RNA-cleaving deoxyribozymes that differentially respond to the presence of methylated cytidines in RNA. On the one hand, the methyl groups in m3C, m4C, and m5C can dramatically inhibit deoxyribozyme-catalyzed RNA cleavage (as in case of AK104). On the other hand, the presence of methylcytidines can lead to accelerated cleavage of modified RNA (as in case of AL112, AM101, and AN05). Owing to these kinetic differences, new deoxyribozymes can be used to examine the identity of methylcytidines and the degree of their methylation at selected target sites. The effect of the methyl group on catalytic activity might be related to steric, electrostatic (m3C) or hydrophobic interaction of the modified nucleotide with its neighbors and Mg2+ in the active site.
The detailed structural basis for the selective recognition of methylcytidines remains to be identified.

As the enriched libraries for selection of methylcytidine-sensitive deoxyribozymes remained highly diverse even after 18 rounds of selection, we expect that many DNA enzymes with potentially interesting properties remained undiscovered. To this end, massively parallel activity assays that are able to screen cleavage reactions of thousands of candidate deoxyribozymes are required. Given the recent success in being able to screen cleavage reactions of thousands of candidate deoxyribozymes remained highly diverse even after

Acknowledgements

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

Conflict of Interest

The authors declare no conflict of interest.

Keywords: deoxyribozymes · epitranscriptomics · in vitro selection · RNA modification · site-specific RNA cleavage

[1] a) P. Boccaletto, M. A. Machnicka, E. Purta, P. Piażkowski, B. Bagiński, T. K. Wirecki, V. de Crécy-Lagard, R. Ross, P. A. Limbach, A. Kotter, M. Helm, J. M. Bujnicki, Nucleic Acids Res. 2018, 46, D303 – D307; b) P. J. McCown, A. Ruszkowska, C. Ni, Kunklar, K. Breger, J. P. Hulewicz, M. C. Wang, N. A. Springer, J. A. Brown, Wiley Interdiscip. Rev. RNA 2020, 11, e1595.
[2] a) J.-J. Xuan, W.-J. Sun, P.-H. Lin, K.-R. Zhou, S. Liu, L.-L. Zheng, J. Biol. Chem. 2019, 295, 1848 – 1856.
[3] a) L. Flemmich, S. Heil, S. Moreno, K. Breuer, R. Micura, Nat. Commun. 2021, 12, 3877; b) S. Mao, P. Haruchanreroanga, S. V. Ranganathan, F. Shen, T. J. Begley, J. Sheng, ACS Chem. Biol. 2021, 16, 76 – 85.
[4] a) R. Micura, W. Pils, C. Höbartner, K. Grubmayr, M. O. Ebert, B. Jaun, Nucleic Acids Res. 2001, 29, 3997 – 4005; b) S. Mao, B. Sekula, M. Ruszkowski, S. V. Ranganathan, P. Haruchanreroanga, Y. Wu, F. Shen, J. Sheng, Nucleic Acids Res. 2020, 48, 10087 – 10100.
[5] a) L. Trixl, A. Lusser, Wiley Interdiscip. Rev. RNA 2019, 10, e5150; b) E. M. Harcourt, A. M. Kietrys, E. T. Kool, Nature 2017, 541, 339 – 346; c) S. Wang, E. T. Kool, Biochemistry 1995, 34, 4125 – 4132.
[6] a) S. W. Santoro, G. F. Joyce, Proc. Natl. Acad. Sci. USA 1997, 94, 4262 – 4266; b) J. Li, W. Zheng, A. H. Kwon, Y. Lu, Nucleic Acids Res. 2000, 28, 481 – 488; c) R. P. G. Cruz, J. B. Withers, Y. Li, Chem. Biol. 2004, 11, 57 – 67; d) K. Schlosser, Y. Li, Biochemistry 2004, 43, 9895 – 9707; e) J. C. F. Lam, J. B. Withers, Y. Li, J. Mol. Biol. 2010, 400, 689 – 701.
[7] a) R. R. Breaker, G. F. Joyce, Chem. Biol. 1994, 1, 223 – 229; b) S. K. Silverman, Nucleic Acids Res. 2005, 33, 6151 – 6163; c) K. Schlosser, Y. Li, ChemBioChem 2010, 11, 866 – 879.
[8] a) C. He, J. Bozler, K. A. Janssen, J. E. Wilusz, B. A. Garcia, A. J. Schorn, R. Bonasio, Nat. Struct. Mol. Biol. 2021, 28, 62 – 70; b) S. Haag, K. E. Sloan, N. Ranjan, A. S. Warda, J. Kretschmer, C. Blessing, B. Hubner, J. Seikowskii, S. Dennerlein, P. Rehling, M. V. Rodnina, C. Höbartner, M. T. Bohnsack, EMBO J. 2016, 35, 2104 – 2119.
[9] a) C.-X. Song, C. Yi, C. He, Nat. Biotechnol. 2012, 30, 1107 – 1116; b) Y. Dai, B.-F. Yuan, Y.-Q. Feng, RSC Chem. Biol. 2021, https://doi.org/10.1039/D1CB00022E.
[10] RNA labeling was performed with FH14 ribozyme: M. Ghaem Maghami, C. P. M. Scheitl, C. Höbartner, J. Am. Chem. Soc. 2019, 141, 19546 – 19549.
[11] a) L. -H. Qu, J.-H. Yang, Nucleic Acids Res. 2014, 42, D327 – D334.
[12] a) V. de Crécy-Lagard, R. Ross, P. A. Limbach, A. Kotter, M. Helm, J. M. Bujnicki, Nucleic Acids Res. 2018, 46, D303 – D307; b) P. J. McCown, A. Ruszkowska, C. Ni, Kunklar, K. Breger, J. P. Hulewicz, M. C. Wang, N. A. Springer, J. A. Brown, Wiley Interdiscip. Rev. RNA 2020, 11, e1595.
[13] a) S. Kobori, Y. Yokobayashi, Angew. Chem. Int. Ed. 2016, 55, 10354 – 10357; Angew. Chem. 2016, 128, 10510 – 10513; b) V. Dhamodharan, S. Kobori, Y. Yokobayashi, ACS Chem. Biol. 2017, 12, 2940 – 2945; c) A. D. Pressman, Z. Liu, E. Janzen, C. Blanco, U. F. Muller, G. F. Joyce, R. Pascali, I. A. Chen, J. Am. Chem. Soc. 2019, 141, 6213 – 6223; d) J. O. L. Andreasson, A. Savinov, S. M. Block, W. J. Greenleaf, Nat. Commun. 2020, 11, 11664 – 11664.
[14] To facilitate access of the deoxyribozymes, disruptor oligonucleotides were included complementary to the 5’ or 3’ end of the acceptor stem, respectively (see DS5 and DS9 in the Supporting Information, Table S2).
[15] a) S. Kobori, Y. Yokobayashi, Angew. Chem. Int. Ed. 2016, 55, 10354 – 10357; Angew. Chem. 2016, 128, 10510 – 10513; b) V. Dhamodharan, S. Kobori, Y. Yokobayashi, ACS Chem. Biol. 2017, 12, 2940 – 2945; c) A. D. Pressman, Z. Liu, E. Janzen, C. Blanco, U. F. Muller, G. F. Joyce, R. Pascali, I. A. Chen, J. Am. Chem. Soc. 2019, 141, 6213 – 6223; d) J. O. L. Andreasson, A. Savinov, S. M. Block, W. J. Greenleaf, Nat. Commun. 2020, 11, 11664 – 11664; e) Y. Yokobayashi, Acc. Chem. Res. 2020, 53, 2903 – 2912; f) Y. Shen, A. Pressman, E. Janzen, I. A. Chen, Nucleic Acids Res. 2021, https://doi.org/10.1093/nar/gkab199.

Manuscript received: May 14, 2021
Revised manuscript received: June 20, 2021
Accepted manuscript online: June 29, 2021
Version of record online: July 26, 2021

19062 www.angewandte.org © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH Angew. Chem. Int. Ed. 2021, 60, 19058 – 19062