Probing Essential Nucleobase Functional Groups in Aptamers and Deoxyribozymes by Nucleotide Analogue Interference Mapping of DNA

Falk Wachowius and Claudia Höbartner*

Research Group Nucleic Acid Chemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

S *Supporting Information*

ABSTRACT: Nucleotide analogue interference mapping of DNA (dNAIM) is here introduced as a new nonenzymatic interference-based approach that enables high-throughput identification of essential nucleobase functional groups in DNA aptamers and in the catalytic core of deoxyribozymes. Nucleobase-modified ribonucleotides are statistically incorporated into DNA by solid-phase synthesis, employing the $2'$-OH group as a chemical tag for analysis of interference effects. This method is exemplified on an AMP-binding DNA aptamer and was further used to identify indispensable nucleobase functional groups for DNA-catalyzed RNA-ligation by the Mg^{2+}-dependent deoxyribozymes 7S11 and 9DB1. dNAIM should prove broadly useful for facile structural probing of functional DNA for which active and inactive variants can be separated based on catalytic or ligand-binding activities.

Exploring the molecular details of nucleic acid catalysis and ligand binding by aptamers requires the identification of nucleotide functional groups that are essential for their specific activity. The catalytic mechanisms employed by natural and artificial ribozymes have been studied in detail.[1−3] For deoxyribozymes, mechanistic information is still very limited, and no 3-D structure of a DNA catalyst in an active conformation is known.[4] Although individual mutation, modification, or deletion variants of deoxyribozymes with short catalytic sequences have been studied,[5−7] this approach becomes experimentally impractical for longer DNA sequences. Therefore, we introduce a combinatorial analysis method that can examine many nucleotide positions in a high-throughput manner. The general applicability is demonstrated for different functional DNAs, including a DNA aptamer and two deoxyribozymes.

For ribozymes and other functional RNAs, efficient methods are available to comprehensively interrogate the function of every RNA nucleotide, but comparable methods are not readily available for studying functional DNA. For example, selective $2'$-hydroxyl acylation analyzed by primer extension (SHAPE) can reveal information on the structural flexibility and the local environment of every nucleotide in folded RNA.[8] Nucleotide analogue interference mapping (NAIM) can identify nucleotide functional groups required for RNA activity.[9] NAIM relies on the incorporation of phosphorothioate-tagged nucleotide analogues into an RNA of interest by in vitro transcription using T7 RNA polymerase. Interference effects are detected by iodine cleavage of the phosphate backbone at phosphorothioate-modified sites.[9] While

SHAPE is tailored to RNA, NAIM could in principle also be applied to DNA.[10] Mapping of essential nucleobase functional groups in DNA by the NAIM approach would depend on enzymatic incorporation of modified dNTPs by DNA polymerases that tolerate nucleobase modifications. However, such a template-dependent, polymerase-assisted approach is not suitable to address functional group modifications that interfere with Watson—Crick (W.C.) base-pairing.

Here, we demonstrate a nonenzymatic approach for NAIM of DNA (dNAIM) that is based on solid-phase synthesis of nucleotide analogue libraries (Figure 1a). In contrast to the phosphorothioate tag in NAIM analysis, we use the $2'$-OH group of ribonucleotides as a chemical tag to encode nucleotide analogues in combinatorial DNA libraries. This is achieved by synthesizing the functional DNA with appropriate phosphoramidite mixtures containing a parent deoxyribonucleotide (dN) and a single nucleobase-modified ribonucleotide (rM) of the parent nucleobase (e.g., dG mixed with a rG analogue). The dN/rM ratio is adjusted to statistically yield one modified nucleotide per DNA molecule. Separation of active and inactive library members is based on the function of the DNA under investigation. Typically, ligand-binding DNA aptamers are separated by affinity chromatography using immobilized ligand (Figure 1b), and deoxyribozyme libraries are separated by denaturing polyacrylamide gel electrophoresis (PAGE) after performing the DNA-catalyzed reaction with the substrate covalently attached to the DNA (Figure 1c). Alkaline hydrolysis of the separated fractions leads to cleavage of the nucleic acid backbone only at modified positions. The interference patterns are then analyzed by PAGE. Missing hydrolysis bands in the active fraction identify positions where nucleobase modifications interfere with DNA function. Interference effects are quantified by taking the ratio of band intensities in the active and unreacted fractions (see Supporting Information). Positions at which modifications are tolerated (interference effect <2) are green, moderate (values between 2 and 6) are orange, and detrimental to activity (strong interference >6) are red. In principle, the $2'$-OH tag by itself can also interfere with DNA activity. This "tag interference effect" can be observed by decreased band intensities in the libraries containing parent ribonucleotides (rN). Although nucleotide analogue information is difficult to analyze in such cases, interference by the $2'$-OH group can also be informative, as it identifies positions that are most sensitive to conformational changes in the DNA backbone. The nucleotide analogues (rM) used in this study are depicted in
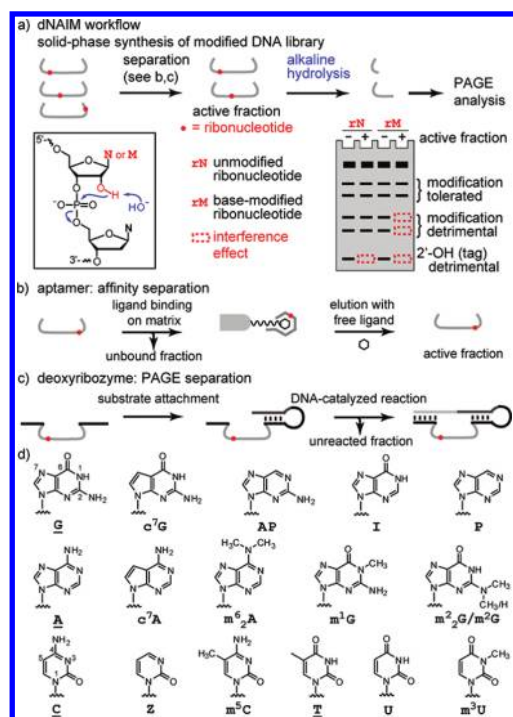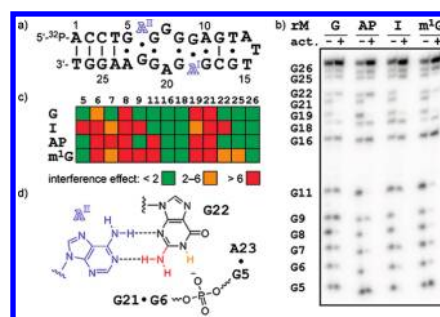
Figure 1. (a) dNAIM: ribonucleotides are statistically incorporated into DNA and active fractions separated. Alkaline hydrolysis leads to backbone cleavage only at modified positions (box). PAGE analysis identifies positions at which modifications are detrimental to catalysis. Separation techniques involve affinity chromatography for aptamers (b) or PAGE for deoxyribozymes (c). Nucleotides and analogues (d): G, guanosine; c[7]G, 7-deazaguanosine; AP, 2-aminopurine; I, inosine; P, purine; A, adenosine; c[7]A, 7-deazaadenosine; m[6]$_2$A, N^6,N^6-dimethyladenosine; m[1]G, 1-methylguanosine; m[2]$_2$G, N^2,N^2-dimethylguanosine; m[2]G, N^2-methylguanosine; C, cytidine; Z, zebularine; m[5]C, 5-methylcytidine; T, thymidine; U, uridine; m[3]U, 3-methyluridine. Unmodified nucleotides are underlined.

Figure 2. dNAIM with three guanosine analogues applied to the ATP/AMP-DNA aptamer.[12,14] (a) The 27mer DNA aptamer binds two molecules AMP (blue outlined) via G·A mismatches.[13] (b) dNAIM gel: act. + lanes, 5'-^{32}P-labeled active aptamer fraction, isolated by specific elution from ATP-agarose with 3 mM ATP, 5 mM MgCl$_2$, pH 7.6; act. − lanes, unseparated library (full gel, Figure S2). (c) Color-coded summary of interference effects. (d) Molecular recognition of AMP[II] (NMR structure PDB ID 1AW4). Essential G22 functional groups are highlighted.

Figure 1d. The W.C. base-pairing face and the major groove edge are modified by attachment of sterically demanding methyl groups or deletion of H-bonding donors or acceptors. Appropriate phosphoramidite building blocks of a subset of nucleotide analogues are commercially available; others were synthesized following published procedures.[11]

As a proof-of-principle, we exemplify the dNAIM approach on an AMP-binding DNA aptamer (Figure 2). This aptamer has been characterized in detail by biochemical methods,[12] and the solution structure of the 27-mer DNA in complex with two AMP molecules has been solved by NMR spectroscopy.[13] Knowledge of the molecular features of the DNA structure and DNA-ligand interactions allows us to predict interference effects for modified nucleotides. We synthesized four dNAIM libraries of the DNA aptamer, using rG, AP, I, and m[1]G, and separated active mutants by affinity selection on ATP-agarose.[14] The hydrolysis pattern of the 5'-^{32}P-labeled DNA aptamer libraries are depicted in Figure 2. As expected, 13 hydrolysis bands are present in the unselected library, indicating successful rN/rM incorporation by combinatorial solid-phase synthesis. In the hydrolysis pattern of the active (AMP-binding) fractions, several bands have reduced intensity or are entirely missing. At positions G8, G19, and G21, the 2'-OH group interferes with ligand binding, preventing the detailed analysis of nucleobase modifications.[15] Most other positions are highly informative. As an example, the interference pattern for G22 is

discussed (Figure 2d). The AMP ligand (blue) is recognized via a minor groove interaction with G22, critically depending on the presence of the N^2 amino group of G22. Accordingly, a G22I mutation is detrimental to ligand binding, resulting in the absence of this mutant in the active fraction (i.e., interference value >6). On the other hand, AP is tolerated, since O^6 and N^1 are not involved in critical structure-forming contacts. A moderate interference effect is observed for m[1]G at G22 indicating that a methyl group at N^1 is not favorably accommodated. NMR structure suggests a reason is the phosphate between G5 and G6 in close proximity to N^1 of G22. A similar pattern is observed for G9, which recognizes the second AMP ligand via a comparable minor groove interaction. Other strong interference effects can also be rationalized based on the NMR structure (see Figure S3). Thus, the dNAIM approach yields reliable information on nucleobase functional groups essential for ligand binding in DNA aptamers.

dNAIM was then applied to study two RNA-ligating deoxyribozymes. First, we analyzed the 7S11 deoxyribozyme,[16] which catalyzes the formation of 2',5'-branched RNA by forming a phosphodiester bond between the 2'OH group of an internal adenosine in one substrate and the 5'-end of a second RNA substrate (Figure 3a). The RNA substrates are bound via W.C. base-pairing and form a 3-helix-junction (3HJ) architecture with the 7S11 deoxyribozyme.[17] The DNA catalyst has two single-stranded loop regions which together contain 50% guanosines (9/18 nucleotides). Comprehensive mutation analysis identified guanosines at loop positions 8, 9, and 10 as critical nucleotides for 7S11 ligation activity.[17,18] However, it is not known which functional groups of these guanosines contribute to activity. Therefore, we synthesized 8 DNA libraries with guanosine analogues for dNAIM analysis of 7S11. The interference patterns are depicted in Figure 3b. Importantly, the 2'-OH tag is tolerated at all 9 guanosines. As expected, nucleotides G8, G9, and G10 show highly informative interference effects. For G8, all modifications lead to diminished activity, with some residual ligation product observed for I, P, and m[1]G. However, deletion of N^7 or O^6, as well as introduction of bulky methyl groups at N^2, leads to inactive DNA variants. This indicates that the Hoogsteen edge of G8 is involved in formation of an active conformation, which must be tightly packed such that methyl groups at the W.C. edge are not tolerated. The interference results for G9 clearly indicate
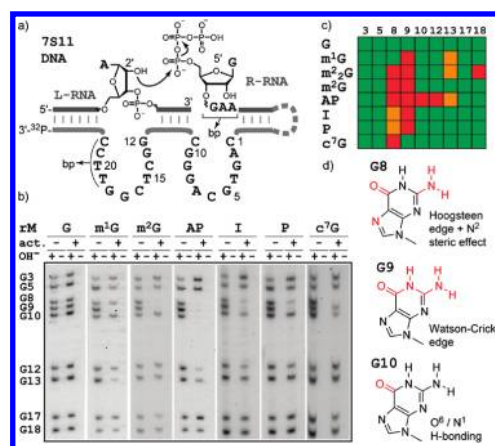
Figure 3. (a) 7S11-catalyzed formation of 2′,5′-branched RNA; bp, base-paired helix; broken line, covalent connection of RNA substrate and DNA during dNAIM. (b) dNAIM gels for guanosine analogues: act. + lanes, 3′-^{32}P-labeled active 7S11 fraction (connected to the ligation product) isolated after 3 h of reaction at pH 9.0, 40 mM $MgCl_2$, 37 °C; − lanes, unligated fraction. (c) Summary of interference results (colors as in Figure 2). (d) Crucial functional groups in the essential guanosine triad G8−10 are highlighted (Figure S4, full gels; Figure S5, quantitative interference values).

that all three H-bonding groups at the W.C. edge are essential for catalytic activity, whereas the N^7 group is dispensable. Nucleotide G10 is the least demanding nucleotide in this triad, since apparently only the O^6 carbonyl functionality is required. These results demonstrate that dNAIM yields high-quality interference data that enable the efficient identification of critical functional groups in deoxyribozymes.

A comprehensive dNAIM analysis was then carried out for all nucleotides in the catalytic core of a second RNA-ligating deoxyribozyme. The 9DB1 DNA enzyme catalyzes the formation of a native 3′−5′ phosphodiester bond between two RNA fragments.[19] The RNA substrates are bound to the 9DB1 DNA via W.C. base-pairing and provide the 3′-OH of the L-RNA as nucleophile, and the 5′-triphosphate of the R-RNA as electrophile (Figure 4a). With our recently described approach for combinatorial mutation interference analysis (CoMA), we found that one of the two predicted stem-loops in the 40-nt core sequence was not required for activity and could be removed without severely affecting ligation performance.[18] Here, we employed the dNAIM approach to identify the functional groups essential for catalytic activity of the minimized deoxyribozyme.

The minimized DNA 9DB1* (Figure 4a) contains 31 nucleotides with a guanosine content of 39%. Using 15 different ribonucleotides, we synthesized 17 combinatorial 9DB1* DNA libraries to identify deoxyribozyme functional groups that define the catalytic center, form the active conformation, or that directly contribute to the RNA ligation activity of 9DB1*. All 31 positions in the catalytic core of 9DB1* were examined, yielding a total of 146 data points (Figure 4b,c).[20]

In the 5′-single-stranded region of the 9DB1* core (nucleotides 1−7), large interference effects are detected for nucleotide G1 with all analogues that modify the W.C. edge. At G2 and A6, all modifications are tolerated, which is consistent with mutation data.[18] Nucleotides A3, T4, and C5, however, are sensitive to the presence of the 2′-OH group, and therefore, interference effects of nucleobase modifications are difficult to quantify. Nevertheless, all tested analogues at A3, T4, and C5 result in inactive DNA derivatives (hydrolysis bands were extremely weak or
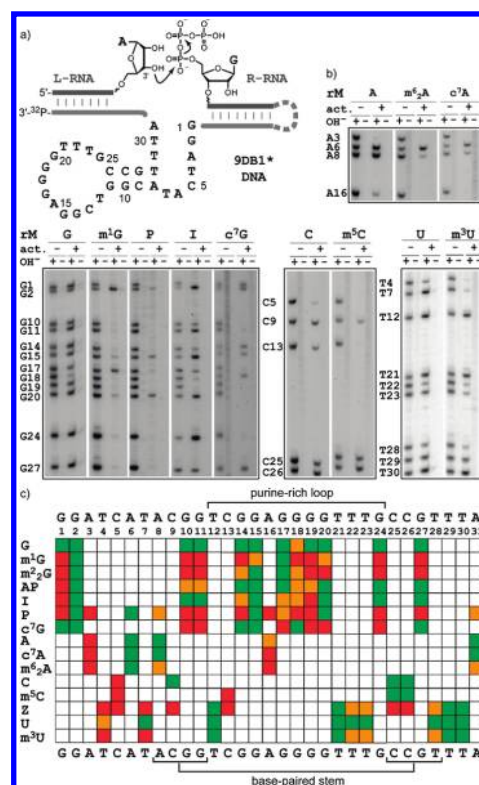
Figure 4. (a) 9DB1-catalyzed RNA ligation reaction and secondary structure of the minimized deoxyribozyme core (9DB1*); broken line, covalent connection of RNA substrate and DNA. (b) Examples of dNAIM gels for selected nucleotide analogues: act. + lanes, active 9DB1* fraction isolated after 3 h of reaction at pH 9.0, 40 mM $MgCl_2$, 37 °C; − lanes, unligated fraction. (c) Summary of interference results for all 146 data points (colors as in Figure 2; white squares: combinations not relevant for dNAIM). See Figures S6, S7, full gels; Figure S9, scheme of interference effects in core sequence; Figure S10, quantitative interference values.

below detection limit), which indicates that this region is important for catalytic activity.

For the predicted base-paired stem (nucleotides 8−11 and 25−28), the interference effects of modifications addressing the W.C. edge are as expected for canonical base pairs. Consistent with a A−T base-pair, P and m^6_2A at position A8, as well as m^3U and Z at T28 result in diminished ligation activity. Similarly, the interference patterns of C9 and G27 are in agreement with a W.C. base-pair: Z is not tolerated at C9, and m^1G, m^2_2G, and P are not accepted at G27. For nucleotides G10 and G11, strong interference effects are found with m^1G and m^2_2G which interfere with base-pairing for steric reasons. Almost no effect is seen with I, which is consistent with the formation of I−C base-pairs, whereas AP and P do not provide appropriate functional group combinations for unperturbed base-pairing with C. Interestingly, additional strong interference effects are observed for modifications in the major groove of this stem. At positions G10 and G11, c^7G is detrimental, that is, deletion of N^7 abolishes catalytic activity. This suggests the presence of essential additional interactions at the Hoogsteen edge of G10 and G11, such as tertiary H-bonds with other nucleotide functional groups or coordination of divalent metal ions. Consistent with the W.C. interactions, C25 and C26 require the N^4 amino group, but tolerate the methyl group of m^5C. Together, these results classify the stem region as a hotspot for tertiary interactions in the
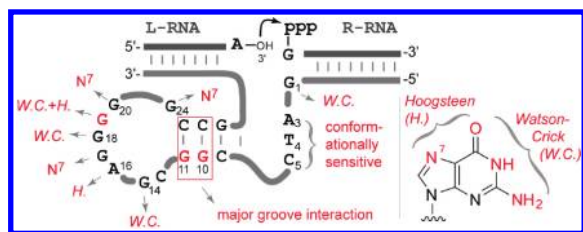
Figure 5. Summary of most important nucleotides in the 9DB1* deoxyribozyme core. Sensitive structural elements are highlighted, and essential interaction sites are abbreviated as *H.* = Hoogsteen edge and *W. C.* = Watson−Crick edge (see also inset).

active conformation of 9DB1* and explain the adverse effect of compensatory base-pair mutations in the stem.[18]

The most critical regions in the purine-rich loop (nucleotides 12−24) include the central five nucleotides, A16−G20, as well as C13, G14, and G24. C13 shows strong interference with Z and m^5C, indicating that it is involved in a tightly packed structural element. While G14 is sensitive to methylation at the W.C. edge, deletion of its N^7 is tolerated. At A16, N^7 and N^6 are absolutely required, indicating an important interaction at the Hoogsteen face. N^7 is also essential at nucleotide G17, but deletion of O^6 or N^2, as well as methylation at N^1 or N^2, is tolerated. Nucleotide G18 shows slightly diminished tolerance for the 2′-OH group, but deleting either O^6 or N^2, as well as double N^2 methylation, has almost no additional effects. Moreover, c^7G is tolerated at G18, but m^1G and P are detrimental. This pattern indicates sensitivity to local conformational changes around this central loop nucleotide. For G19, almost all tested modifications lead to inactive DNA; only with AP some residual activity is detected. This suggests that the W.C., the Hoogsteen, and likely also the minor groove edge of G19 participate in essential contacts. Similar to G17, G20 apparently only involves its N^7 in the formation of the active structure. At G24, N^7 is also required, but in contrast to G20, methylations and simultaneous deletion of all exocyclic functional groups are detrimental.

Overall, we observed many effects in the 9DB1* core, with only 6 out of 31 nucleotides tolerating all tested modifications. The results of the dNAIM analysis are summarized in Figure 5, highlighting the most important nucleotides and their essential functional groups for formation of a sophisticated interaction network required for catalysis.

In summary, we have shown that dNAIM is generally applicable to structured DNAs that have a selectable function, in particular to aptamers and deoxyribozymes. Since several functional groups at every nucleotide position are analyzed in parallel, this method provides an efficient and unbiased route to identify key functional elements for DNA activity. Solid-phase synthesis of modified DNA libraries for dNAIM enables investigation of modifications at the W.C. base-pairing edge that remain elusive to enzymatic interference methods. The data provided by dNAIM are indeed informative because they represent the active DNA conformation and can therefore guide meaningful structure−function analyses. Several candidate nucleotides were identified in the 7S11 and 9DB1 deoxyribozymes that warrant further investigation, for example, by using affinity labels,[21] cross-linking,[22] phosphorothioate interference, and metal ion rescue experiments.[23] In combination with site-specifically modified ligation substrates, the dNAIM libraries will provide additional powerful information in experiments similar to nucleotide analogue interference suppression (NAIS) studies of ribozymes.[9] This will allow us to address the question if DNA, despite the lack of the 2′-OH group, uses similar catalytic strategies as RNA. This is interesting from a conceptual viewpoint of understanding key aspects of DNA catalysis and may guide engineering of DNA enzymes to expand the scope of current applications.

■ ASSOCIATED CONTENT

Ⓢ **Supporting Information.** Experimental details, hydrolysis gels, and interference values. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
claudia.hoebartner@mpibpc.mpg.de

■ ACKNOWLEDGMENT

■ REFERENCES

(1) Cochrane, J. C.; Strobel, S. A. *Acc. Chem. Res.* **2008**, *41*, 1027.
(2) Lilley, D. M. *Biochem. Soc. Trans.* **2011**, *39*, 641.
(3) Fedor, M. J.; Williamson, J. R. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 399.
(4) Nowakowski, J.; Shim, P. J.; Prasad, G. S.; Stout, C. D.; Joyce, G. F. *Nat. Struct. Biol.* **1999**, *6*, 151.
(5) Peracchi, A.; Bonaccio, M.; Clerici, M. *J. Mol. Biol.* **2005**, *352*, 783.
(6) Zaborowska, Z.; Furste, J. P.; Erdmann, V. A.; Kurreck, J. *J. Biol. Chem.* **2002**, *277*, 40617.
(7) Wang, B.; Cao, L.; Chiuman, W.; Li, Y.; Xi, Z. *Biochemistry* **2010**, *49*, 7553.
(8) Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M. *J. Am. Chem. Soc.* **2005**, *127*, 4223.
(9) Suydam, I. T.; Strobel, S. A. *Methods Enzymol.* **2009**, *468*, 3.
(10) Gish, G.; Eckstein, F. *Science* **1988**, *240*, 1520.
(11) Höbartner, C.; Kreutz, C.; Flecker, E.; Ottenschläger, E.; Pils, W.; Grubmayr, K.; Micura, R. *Monatsh. Chem.* **2003**, *134*, 851.
(12) Huizenga, D. E.; Szostak, J. W. *Biochemistry* **1995**, *34*, 656.
(13) Lin, C. H.; Patel, D. J. *Chem. Biol.* **1997**, *4*, 817.
(14) Although selected for binding to ATP, the 27mer aptamer binds ATP, AMP, and adenosine with similar affinities; refs 12 and 13.
(15) The OH interference can be rationalized based on the known structure. Nucleotides 8, 19, and 21 are involved in critical three-base platforms in the ligand binding pocket. The tag effect indicates sensitivity to local conformational changes in the backbone that prevent proper orientation of the nucleobases for productive H-bonding.
(16) Coppins, R. L.; Silverman, S. K. *Nat. Struct. Mol. Biol.* **2004**, *11*, 270.
(17) Coppins, R. L.; Silverman, S. K. *J. Am. Chem. Soc.* **2005**, *127*, 2900.
(18) Wachowius, F.; Javadi-Zarnaghi, F.; Höbartner, C. *Angew. Chem., Int. Ed.* **2010**, *49*, 8504.
(19) Purtha, W. E.; Coppins, R. L.; Smalley, M. K.; Silverman, S. K. *J. Am. Chem. Soc.* **2005**, *127*, 13124.
(20) We probed 12 G's with seven analogues, 5 A's with four analogues, 5 C's and 9 T's with three analogues each. The binding arms were not modified. The 2′-OH tag was well tolerated at the majority of the 31 core nucleotides.
(21) Thomas, J. M.; Perrin, D. M. *J. Am. Chem. Soc.* **2008**, *130*, 15467.
(22) Sekhon, G. S.; Sen, D. *Biochemistry* **2010**, *49*, 9072.
(23) Frederiksen, J. K.; Piccirilli, J. A. *Methods* **2009**, *49*, 146.