Dramatic Effect of Furanose C2′ Substitution on Structure and Stability: Directing the Folding of the Human Telomeric Quadruplex with a Single Fluorine Atom

Nerea Martín-Pintado,[†] Maryam Yahyaee-Anzahaee,[‡] Glen F. Deleavey,[‡] Guillem Portella,[§] Modesto Orozco,[§] Masad J. Damha,*[,‡] and Carlos González*[,†]

[†]Instituto de Química Física "Rocasolano", CSIC, Serrano 119, 28006 Madrid, Spain

[‡]Department of Chemistry, McGill University, 801 Sherbrooke Street West, Montreal, Quebec H3A 0B8, Canada

[§]Joint IRB-BSC Program on Computational Biology, Institute for Research in Biomedicine, Barcelona Supercomputing Center, and Department of Biochemistry, University of Barcelona, Baldir I Reixac 10-12, 08028 Barcelona, Spain

ⓢ *Supporting Information*

ABSTRACT: Human telomeric DNA quadruplexes can adopt different conformations in solution. We have found that arabinose, 2′F-arabinose, and ribose substitutions stabilize the propeller parallel G-quadruplex form over competing conformers, allowing NMR structural determination of this particularly significant nucleic acid structure. 2′F-arabinose substitution provides the greatest stabilization as a result of electrostatic (F−CH---O4′) and pseudo-hydrogen-bond (F---H8) stabilizing interactions. In contrast, 2′F-rG substitution provokes a dramatic destabilization of the quadruplex structure due to unfavorable electrostatic repulsion between the phosphate and the 2′-F.

G-quadruplexes are nucleic acid structures occurring in human telomers and oncogene-promoter regions, and they have garnered considerable attention due to their involvement in telomere maintenance and gene regulation.[1−3] They have also emerged as novel building blocks for supramolecular assemblies,[4] DNA-based nanodevices,[5−9] and potential antiviral agents, and as a new class of molecular targets for anti-cancer drugs.[10,11] G-quadruplexes exhibit significant structural diversity.[3,12,13] High resolution structural studies on human telomeric DNA have mainly focused on short DNA sequences able to form individual quadruplexes. In many cases, these oligonucleotide sequences can adopt and interconvert between different conformations, obscuring structural analysis. Among the observed telomeric G-quadruplex topologies, it has been suggested that the simple folds of the parallel "propeller" and "hybrid" (3+1)[14,15] structural models provide more adequate scaffolds for the compact-stacking multi-quadruplex structure of human telomeric DNA than other antiparallel models.[12,13] Interestingly, human telomeric RNA (TERRA), which is known to play a crucial role in telomere maintenance in vivo,[16,17] always adopts propeller-like parallel structures.[18,19]

The tendency of TERRA and other RNA quadruplexes to form parallel structures is related to the strong preference of guanosine ribonucleosides to adopt the *anti* glycosidic bond conformation.[20] The *syn*/*anti* preference of dG vs rG has been used to modulate G-quadruplex folding, and different nucleoside modifications have been proposed following this strategy.[20−26] In particular, 2′-deoxy-2′-fluoro-arabinonucleosides (e.g., 2′F-araG) favor the *anti* orientation in order to avoid steric interaction between guanine and the top-face fluorine of the arabinose sugar.[22,27] The 2′F-arabinose modification is especially interesting since it can modulate the structure and stability of quadruplexes,[22,23] and unlike ribose, it does not disrupt the conformation of the furanose ring.[28−30] Thus, replacement of dG's adopting the *anti* glycosidic bond conformation with 2′F-araG's stabilizes the thrombin binding aptamer[22] and the "hybrid (3+1)" quadruplex[23] up to +3.4 °C per 2′F-araG modification.

To expand upon these findings and to investigate the structural basis of the effect of C2′ modifications in G-quadruplexes, we have studied the impact of several substitutions, 2′F-araG (**AFtel**), 2′F-rG (**RFtel**), araG (**Atel**), and rG (**Rtel**), in the human telomeric quadruplex (Figure 1). In this study, we have focused on the two-repeat telomeric sequence, d(TAGGGTTAGGGT)(**Dtel**), which offers a convenient model for high-resolution studies by either NMR[31] or crystallography.[14] In solution this sequence exists as a mixture of two interconverting forms: a parallel propeller-like fold and an antiparallel fold.[31] We have now found that a single dG-to-2′F-araG modification at the ninth position provides the propeller-like conformation exclusively, allowing for detailed structural analysis. Our rationale for substitution at G9 d(TAGGGTTA**G**GGT) stems from the observation that in antiparallel quadruplexes the guanine at the 3′-side of the first TTA loop tends to adopt a *syn* glycosidic bond conformation,[13] which is disfavored in 2′F-arabinonucleosides. The nucleotide structures and oligonucleotide sequences studied are shown in Table 1 and Figure 1.

G-quadruplex formation and melting were monitored by circular dichroism (CD) and NMR spectroscopy in K+-containing solution (Figure 1 and Figure S1 in the Supporting Information (SI)). The spectra of the unmodified sequence, **Dtel**, were characteristic of a mixture of parallel and antiparallel
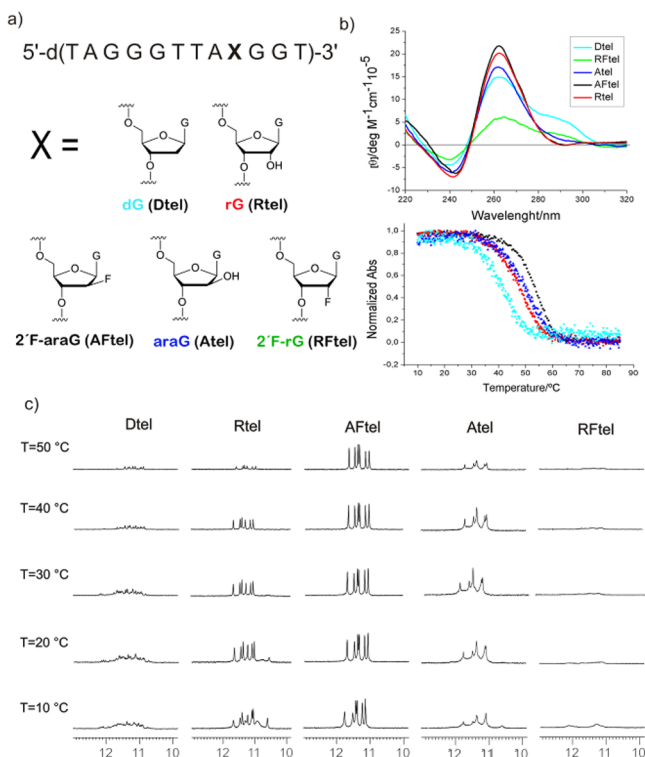
Figure 1. (a) Oligonucleotide sequence and schemes of the native (dG, rG) and chemically modified nucleotides (2′F-araG, araG, and 2′F-rG). (b) CD spectra and melting profiles. (c) Imino region of the ^{1}H NMR spectra at different temperatures of the five singly modified oligonucleotides. Same experimental conditions as in Figure S1.

Table 1. T_m Values for DNA Telomeric Sequence (Dtel) and for Singly Modified Sequences[a]

Name	Seq: (5′-3′)	T_m (°C)	ΔT_m (°C)
Dtel	TAGGGTTAGGGT	41	0
Rtel	TAGGGTTA(rG)GGT	47	+6
Atel	TAGGGTTA(araG)GGT	50	+9
AFtel	TAGGGTTA(2′F-araG)GGT	53	+12
RFtel	TAGGGTTA(2′F-rG)GGT	-	-

[a]Buffer conditions: 15 mM potassium phosphate, 5 mM KCl, pH 7.0; strand concentration 75 μM.

topologies (Figure 1b,c). By contrast, **AFtel**, **Atel**, and **Rtel** fold mainly into a single G-quadruplex form, as indicated by the number of proton resonances; furthermore, the six sharp imino signals observed (10−12 ppm) and the data presented below are consistent with a symmetric, bimolecular G-quadruplex complex (Figures 1 and S1). In marked constrast, the imino region of **RFtel** shows very broad and weak signals, suggesting the presence of multiple species of very low stability (Figures 1 and S1).

CD spectra of **AFtel**, **Atel**, and **Rtel** exhibit the distinctive features of a parallel quadruplex (a negative band at around 240, and a positive one at 265 nm), in contrast with CD spectra of the unmodified sequence (**Dtel**, Figure 1). The CD spectrum of **RFtel** is clearly different from the others, displaying low-intensity bands, suggesting that it retains very little quadruplex structure, if at all, under the experimental conditions.

Melting temperatures of all sequences were obtained from CD thermal denaturation experiments (Figure 1 and Table 1) and ^{19}F NMR spectroscopy for **AFtel** (Figure S2). Both arabinose modifications (**AFtel** and **Atel**) induce a strong stabilization, the effect being more pronounced for **AFtel** (ΔT_m = +12 and +9 °C, respectively, relative to **Dtel**). In fact, **AFtel** exhibits the highest melting temperature of the entire series, exhibiting a T_m (53 °C) which rivals that of the all-RNA parallel quadruplex under similar conditions [r(UAGGGUUAGGGU); T_m = 55 °C]. The rG substitution is also stabilizing (ΔT_m = +6 °C), in contrast to the 2′F-rG modification which is destabilizing (Figures 1 and S1). The overall trend observed was 2′F-araG > araG > rG > dG ≫ 2′F-rG (Table 1).

Detailed structural information could be extracted from two-dimensional NMR spectra. In all cases, the number of spin systems in the NMR spectra indicated the formation of a symmetric structure. Sequential connectivities for the guanine residues within the central core could be traced via NOESY experiments (see Figures S3 and S4 and Table S1 for chemical shift assignments). H8−H1′ NOE cross-peak intensity indicates that all guanines are in *anti* glycosidic conformation (Figure S4). As expected for a symmetrical G-quadruplex with three tetrads, six imino protons were observed in the 10.5−12.0 ppm region of **AFtel**, **Rtel**, and **Atel**. The exchangeable proton spectrum of **AFtel** is very similar to that of the parallel conformer in the mixture formed by d(TAGGGUTAGGGT); see Figure 3 in Phan et al.[31] Consequently, exchangeable proton assignment was carried out on the basis of this structure.[31] These assignments were fully consistent with the NOESY spectra recorded in H₂O (Figure 2). The G-tetrad alignment could be identified from the NOESY spectra on the
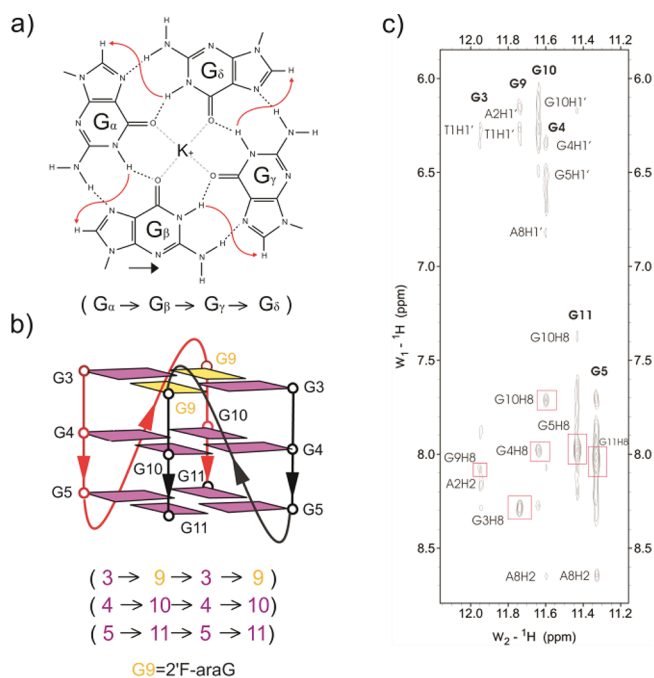

Figure 2. (a) G-tetrad directionalities. (b) Parallel propeller G-quadruplex scheme with all guanines in *anti* conformation. 2′F-araG nucleotides are shown in yellow and dG residues in magenta. (c) **AFtel** imino region of the NOESY spectrum in H₂O (100 ms mixing time). The G-tetrad alignment was identified from the NOESY on the basis of the characteristic imino-H8 connectivity (indicated by red squares) around the G-tetrad.

basis of the characteristic imino-H8 connectivity, being G3·G9·G3·G9, G4·G10·G4·G10, and G5·G11·G5·G11 (G9 = 2′F-araG; Figure 2) with the clockwise directionality shown in Figure 2b. Two TTA loops connect the tetrads, with the terminal residues (T1-A2 and T12) stacked on top and bottom of the G-quadruplex.

The dG residues of **AFtel**, **Atel**, and **Rtel** all adopt the *South* sugar conformation as determined from DQF-COSY-derived *J*-coupling constants (Figure S6). The arabinose residues in **AFtel** and **Atel** adopt pure *South* sugar conformations (Figures S6 and S7), whereas rG (**Rtel**) favors the *North* conformation (Figure S8), as expected. NMR-derived distances and torsion angle constraints together with restrained molecular dynamics (MD) were used to calculate the three-dimensional structure of **AFtel** (see Methods in the SI and Table S2). The G-quadruplex core is parallel, symmetric, and very well defined (rmsd = 0.7 Å), with all guanines in the *anti* conformation (Figures 2 and 3). Continuous stacking of T1, A2, and G3 was supported by
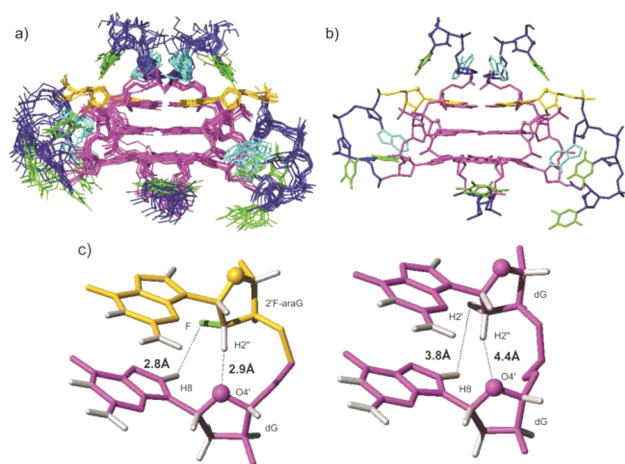

Figure 3. NMR structure of **AFtel** (PDB code: 2M1G). (a) Front view of the 10 superimposed refined structures. (b) Front view of the average structure. (c) Distances between F_i–H8$_{i+1}$ and H2″$_i$–O4′$_{i+1}$ in **AFtel** and H2′$_i$–H8$_{i+1}$ and H2″$_i$–O4′$_{i+1}$ in the crystallographic structure (PDB code: 1K8P). Guanines are shown in magenta for dG and yellow for 2′F-araG. Thymidines are shown in dark green, dA's in cyan, and fluorines in light green. Stereoviews are shown in Figure S11.

indicative NOEs, such as A2H2-G3H6 or T1CH3-G9H1 (Figure S9). The same applies to T12 at the bottom of the G-core, which stacks over the G5·G11·G5·G11 tetrad (Figure S9). NMR data indicate that **Atel** and **Rtel** adopt similar structures (see Figure S3). To better assess this point, state-of-the-art atomistic MD simulations were carried out on **Atel** and **Rtel** starting from the NMR structure of **AFtel**. In all three cases, the trajectories show very small deviations (see Figure S10 for details). In conclusion, **AFtel**, **Atel**, and **Rtel** adopt parallel, propeller-like, bimolecular quadruplex structures.

The structure of **AFtel** was then compared with the crystallographic structure of d(^{Br}UAGGG^{Br}UTAGGGT) (PDB ID: 1K8P) as a model of the DNA telomeric sequence.[14] The root-mean-squared deviations (rmsd's) between both structures are around 3.4 and 1.5 Å for all heavy atoms and the G-core, respectively. Superposition of the G-core residues in both structures indicates that the main differences reside around the 2′F-araG modification (Figure 3c). Specifically, the 2′F-araG sugar residues (position 9 in each subunit) are

displaced toward the minor groove and consequently toward their 3′-neighboring base (Figures 3 and S11 and Table S3). This effect provokes a substantial decrease in the distance between the 2′-fluorine and H8 of the adjacent (3′-dG10) base (Figure 3c). The 2′F–H8 distance for **AFtel** is around 2.8 Å, compared to 3.8 Å for the crystallographic structure (Figures 3 and S11). The close 2′F---H8–C8 contact[32] is indicative of a pseudo-hydrogen bond,[33] like those found in 2′F-ANA:RNA hybrid duplexes.[34] For **AFtel**, the 2′F---H8–C8 angle is around 140° (Table S4), which is 20° higher than that observed in the unmodified dG core. This value (140°) is similar to that observed in the 2′F-ANA:RNA hybrid (145°), and it is considered optimum for formation of X–H---F (X = N, O, or C) pseudo-hydrogen-bonds.[35,36] The **AFtel** structure revealed an additional close contact between H2″ and O4′ of the neighboring 3′-dG residue (2.9 Å in **AFtel** vs 4.4 Å in the crystal structure), with a C2″–H2″---O4′ angle of 150° (versus 120° for 1K8P) (see Figures 3 and S11 and Tables S5 and S6). Stabilizing C–H---O interactions have previously been described for nucleic acid[37] and protein–protein interactions.[38,39] Therefore, we hypothesize that the enhanced stability imparted by the 2′F-araG modification is due, at least in part, to C2′–F---H8 and FC2′–H2″---O4′ interactions, both of which are lacking in the native structure.

For **Atel**, the top-side 2′-OH of the araG nucleotide can easily be accommodated without incurring the steric penalty observed in duplex structures.[28,40–42] Furthermore, the ara 2′-OH is involved in favorable electrostatic interactions with the neighboring phosphates, as also seen for rG (**Rtel**) but not **RFtel** (see Figure S12a). This makes ara 2′-OH modifications particularly useful for applications where additional stabilization of a quadruplex and concomitant destabilization of a competing duplex structure are desired. The contrasting behavior of **Rtel** relative to **RFtel** is quite striking, given the small structural differences between these nucleotides and the fact that 2′-F-rN modifications are generally stabilizing.[43–45] Whereas MD simulations failed to detect spontaneous unfolding of the **RFtel** sequence, thermodynamic integration (TI) calculations (see Methods in the SI) indicated that the apparently innocuous change from 2′-OH to 2′-F generates a strong electrostatic repulsion between the fluorine group and oxygen atoms in the backbone (around 26 kJ/mol per substitution; see Figure S12).

Gas-phase quantum mechanical (QM) calculations on the dinucleotide models 2′F-rGpdG vs rGpdG revealed that the 3′-dG nucleoside unit has little influence (ΔE = +2.5 kJ/mol) to distinguish between the two sugars (see Methods in the SI and Figure S13). However, the interaction with the 3′-phosphate group seems crucial, since it stabilizes rGpdG over 2′F-rGpdG by as much as −49 kJ/mol. Further analysis of the six pairs of representative structures revealed that the ribo 2′-OH makes strong hydrogen bond contacts with the 3′-phosphate, whereas unfavorable electrostatic repulsion between phosphate and the 2′-F destabilizes 2′F-rGpdG (Figure S13). Clearly, the large energy difference is likely to be reduced when hydration free energy of the sugars is taken into account; that is, the 2′-OH group is expected to be better solvated than fluorine, and, accordingly, it will be more stabilized by the solvent in the unfolded state. One can estimate the difference in relative hydration free energy of ribose vs 2′F-deoxyribose in the unfolded state from the experimental hydration free energy difference between methanol and fluoromethane (20 kJ/mol[46]). Combining this number with the preferential

stabilization energy determined via QM, we expect the rG substitution to be more stable by ca. 24 kJ/mol (relative to 2′F-rG), a value that agrees quite well with MD/TI estimates (26 kJ/mol) that take into account classical enthalpic and entropic terms. This may provide an explanation for the surprising difference between 2′F-rG and rG modifications in modulating and stabilizing this particular G-quadruplex.

In conclusion, arabinose and ribose substitutions, and in particular the 2′F-ara subsitution, can be used to stabilize the biologically relevant propeller parallel G-quadruplex form over competing conformers, allowing NMR structural determination of this particularly significant nucleic acid structure. 2′F-ara allowed the greatest stabilization (+12 °C) of the parallel quadruplex, as a result of electrostatic (F—CH---O4′) and pseudo-hydrogen-bonding (F---H8) stabilizing interactions. Sugar/phosphate interactions stabilize the ara and ribo substitutions, whereas 2′F-rG substitution provoked a dramatic destabilization (disruption) of the quadruplex structure. Thus, a single conservative oligonucleotide chemical modification can exert remarkable control over G-quadruplex formation, enabling significant structural insight into the nature of the propeller parallel G-quadruplex.

■ ASSOCIATED CONTENT

⑤ Supporting Information

Detailed descriptions of the experimental procedures; 7 tables with assignment, calculation statistics, and structural analysis; 13 figures showing NMR data and structural models. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
masad.damha@mcgill.ca; cgonzalez@iqfr.csic.es

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Sandell, L. L.; Zakian, V. A. *Cell* **1993**, *75*, 729.
(2) De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J.-F.; Mergny, J.-L. *Biochimie* **2008**, *90*, 131.
(3) Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. *Nat. Struct. Mol. Biol.* **2005**, *12*, 847.
(4) Nikan, M.; Sherman, J. C. *Angew. Chem., Int. Ed.* **2008**, *47*, 4900.
(5) So, H.-M.; Won, K.; Kim, Y. H.; Kim, B.-K.; Ryu, B. H.; Na, P. S.; Kim, H.; Lee, J.-O. *J. Am. Chem. Soc.* **2005**, *127*, 11906.
(6) Hsu, C.-L.; Chang, H.-T.; Chen, C.-T.; Wei, S.-C.; Shiang, Y.-C.; Huang, C.-C. *Chem.—Eur. J.* **2011**, *17*, 10994.
(7) Ge, B.; Huang, Y. C.; Sen, D.; Yu, H.-Z. *Angew. Chem., Int. Ed.* **2010**, *49*, 9965.
(8) Huang, Y. C.; Cheng, A. K. H.; Yu, H.-Z.; Sen, D. *Biochemistry* **2009**, *48*, 6794.
(9) Huang, Y. C.; Sen, D. *J. Am. Chem. Soc.* **2010**, *132*, 2663.
(10) Yang, D.; Okamoto, K. *Future Med. Chem.* **2010**, *2*, 619.
(11) Phan, A. T.; Kuryavyi, V.; Ma, J.-B.; Faure, A.; Andréola, M.-L.; Patel, D. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 634.
(12) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402.
(13) Phan, A. T. *FEBS J.* **2010**, *277*, 1107.
(14) Parkinson, G. N.; Lee, M. P.; Neidle, S. *Nature* **2002**, *417*, 876.
(15) Phan, A. T.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2006**, *34*, 5715.
(16) Azzalin, C. M.; Reichenbach, P.; Khoriauli, L.; Giulotto, E.; Lingner, J. *Science* **2007**, *318*, 798.
(17) Stefan, S.; Maria, A. B. *Nat. Cell Biol.* **2007**, *10*, 228.
(18) Martadinata, H.; Phan, A. T. *J. Am. Chem. Soc.* **2009**, *131*, 2570.
(19) Collie, G. W.; Haider, S. M.; Neidle, S.; Parkinson, G. N. *Nucleic Acids Res.* **2010**, *38*, 5569.
(20) Tang, C. F.; Shafer, R. H. *J. Am. Chem. Soc.* **2006**, *128*, 5966.
(21) Karsisiotis, A.; Webba da Silva, M. *Molecules* **2012**, *17*, 13073.
(22) Peng, C. G.; Damha, M. J. *Nucleic Acids Res.* **2007**, *35*, 4977.
(23) Lech, C. J.; Li, Z.; Heddi, B.; Phan, A. T. *Chem. Commun.* **2012**, *48*, 11425.
(24) Webba da Silva, M.; Trajkovski, M.; Sannohe, Y.; Ma'ani Hessari, N.; Sugiyama, H.; Plavec, J. *Angew. Chem., Int. Ed.* **2009**, *48*, 9167.
(25) Dominick, P. K.; Jarstfer, M. B. *J. Am. Chem. Soc.* **2004**, *126*, 5050.
(26) Doluca, O.; Withers, J. M.; Filichev, V. V. *Chem. Rev.* **2013**, DOI: 10.1021/cr300225q.
(27) Sapse, A. M.; Snyder, G. *Cancer Invest.* **1985**, *3*, 115.
(28) Martin-Pintado, N.; Yahyaee-Anzahaee, M.; Campos-Olivas, R.; Noronha, A. M.; Wilds, C. J.; Damha, M. J.; González, C. *Nucleic Acids Res.* **2012**, *40*, 9329.
(29) Berger, I.; Tereshko, V.; Ikeda, H.; Marquez, V. E.; Egli, M. *Nucleic Acids Res.* **1998**, *26*, 2473.
(30) Watts, J. K.; Damha, M. J. *Can. J. Chem.* **2008**, *86*, 641.
(31) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2003**, *125*, 15021.
(32) Li, F.; Sarkhel, S.; Wilds, C. J.; Wawrzak, Z.; Prakash, T. P.; Manoharan, M.; Egli, M. *Biochemistry* **2006**, *45*, 4141.
(33) Anzahaee, M. Y.; Watts, J. K.; Alla, N. R.; Nicholson, A. W.; Damha, M. J. *J. Am. Chem. Soc.* **2010**, *133*, 728.
(34) Watts, J. K.; Martín-Pintado, N.; Gómez-Pinto, I.; Schwartzentruber, J.; Portella, G.; Orozco, M.; González, C.; Damha, M. J. *Nucleic Acids Res.* **2010**, *38*, 2498.
(35) Mehta, G.; Sen, S. *Eur. J. Org. Chem.* **2010**, *2010*, 3387.
(36) Dalvit, C.; Vulpetti, A. *ChemMedChem* **2012**, *7*, 262.
(37) Berger, I.; Egli, M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12116.
(38) Jiang, L.; Lai, L. *J. Biol. Chem.* **2002**, *277*, 37732.
(39) Horowitz, S.; Yesselman, J. D.; Al-Hashimi, H. M.; Trievel, R. C. *J. Biol. Chem.* **2011**, *286*, 18658.
(40) Noronha, A. M.; Wilds, C. J.; Lok, C. N.; Viazovkina, K.; Arion, D.; Parniak, M. A.; Damha, M. J. *Biochemistry* **2000**, *39*, 7050.
(41) Venkateswarlu, D.; Lind, K. E.; Mohan, V.; Manoharan, M.; Ferguson, D. M. *Nucleic Acids Res.* **1999**, *27*, 2189.
(42) Giannaris, P. A.; Damha, M. J. *Can. J. Chem.* **1994**, *72*, 909.
(43) Pallan, P. S.; Greene, E. M.; Jicman, P. A.; Pandey, R. K.; Manoharan, M.; Rozners, E.; Egli, M. *Nucleic Acids Res.* **2011**, *39*, 3482.
(44) Deleavey, G. F.; Watts, J. K.; Alain, T.; Robert, F.; Kalota, A.; Aishwarya, V.; Pelletier, J.; Gewirtz, A. M.; Sonenberg, N.; Damha, M. J. *Nucleic Acids Res.* **2010**, *38*, 4547.
(45) Patra, A.; Paolillo, M.; Charisse, K.; Manoharan, M.; Rozners, E.; Egli, M. *Angew. Chem., Int. Ed.* **2012**, *51*, 11863.
(46) Radzicka, A.; Wolfenden, R. *Biochemistry* **1988**, *27*, 1664.