JOC The Journal of Organic Chemistry

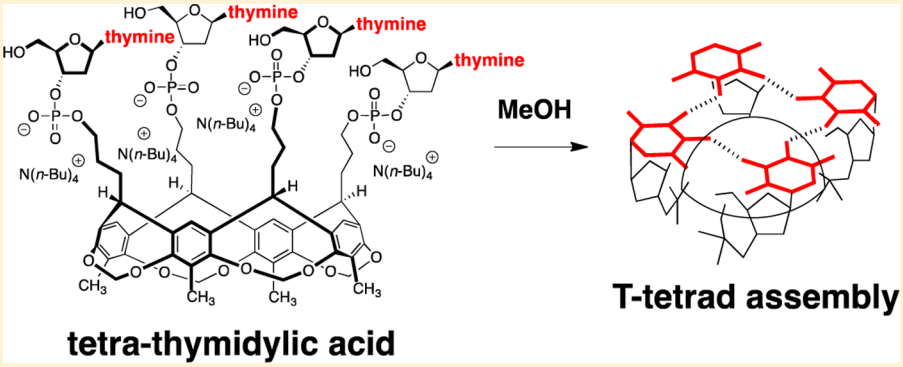
Note

pubs.acs.org/joc

A Thymine Tetrad Assembly Templated from Thymidylic Acid

Grant A. L. Bare and John C. Sherman*

Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, BC, Canada, V6T 1Z1



ABSTRACT: A template tetra-coupled with thymidylic acid through a phosphate linkage was characterized in methanol for emergent properties of nucleobase tetrad formation. Intramolecular hydrogen bonded base pairing in the absence of a cation was indicated for the thymidylic acid species supporting a monomeric template-assembled structure. Thus, an initial report of a stabilized individual thymine tetrad assembly is presented here. Consistent with previous investigations, a deoxyguanylic acid variant templated an analogous methanolic monomeric G-tetrad in comparison to the thymine species.

Nucleobase tetrad formations of guanine bases also called G-tetrads have been extensively studied due to their discovery as a fundamental nucleic acid structure in Nature.[1,2] G-tetrads are characterized by interactions between four guanine bases in a coplanar Hoogsteen hydrogen bonded base pairing arrangement normally supported by metal cation coordination to the guanine O6 position. Tetrad stacking forms the basis of four stranded G-quadruplex structures thermodynamically stable under conditions of the cell, and such structures have been implicated for a transcriptional regulatory function and a chromosome stabilizing role in the human genome; recent visualization of G-quadruplex DNA in human cells lends increasing support to this.[3] Observations of tetrad structures of canonical nucleobases other than guanine are rare in contrast, and potential biological functions of these tetrads have not yet been revealed at the present time. In the case of adenine, an A-tetrad structure has been proposed in K^+ solution at a terminus of a truncated human telomeric G-quadruplex, $d(AG_3T)_4$, and at the center of a G-quadruplex containing a c-MYC oncogene repeat sequence, $d(TGGAGGC)_4$.[4,5] The four-stranded RNA G-quadruplex, $r(UGGGGU)_4$, derived from the *Escherichia coli* 5S RNA has revealed a U-tetrad at its 3′ end in solution in the presence of K^+ and in the solid-state in the presence of Sr^{2+}.[6,7] Recent observation of thermodynamic stabilization imparted on a G-quadruplex through the presence of a 3′ U-tetrad for the telomeric RNA G-quadruplex, $r(UAGGGU)_4$, has questioned the role of the G-quadruplex scaffold as a stabilizing element for nonguanine tetrad structures.[8] Of the few reports of T-tetrads in existence, terminating T-tetrads have been observed in the DNA G-quadruplex, $d(TGGGGT)_4$, for Tl^+/Na^+ X-ray crystal structures and for NH_4^+ solution species.[9,10] In common with these systems outlined above is for the nonguanine tetrad to be organizationally dependent on a cation-stabilized G-quadruplex structure.

Minimal synthetic model systems of unusual nucleobase associations independent of G-quadruplex are sought to assist in understanding conditions that may permit their discovery and characterization in biological systems or are sought in the design of new functional nanoscale assemblies. Previous studies in our lab demonstrated that triazole-linked uridine exhibits self-associative behavior when assembled with a template molecule in organic solvent.[11] It was hypothesized that a synthetic template-assembled system more closely representative of biomolecules in their native aqueous environment such as one composed of nucleotides in a polar protic solvent could show emergent properties of tetrad formation. We report here on the observation of thymidylic acid's template-assembly behavior by showing that tetra-thymidylic acid **1** readily assembles a single T-tetrad in methanol solvent (Figure 1).

Syntheses of methanol soluble tetra-thymidylic acid **1** and analogue templated tetra-nucleotides **2**–**4** chosen here for comparison study have been reported elsewhere.[12] Synthetic design incorporated (i) a rigidified template capable of preorganization of four nucleotides in close spatial proximity on one template face, (ii) deoxynucleotide DNA subunits, and (iii) solubilizing phosphodiester linkages paired with $(n\text{-Bu})_4N^+$ counterions. Owing to the widespread chemical phenomenon

Received: June 13, 2013
Published: July 22, 2013

dx.doi.org/10.1021/jo401281p | J. Org. Chem. 2013, 78, 8198–8202

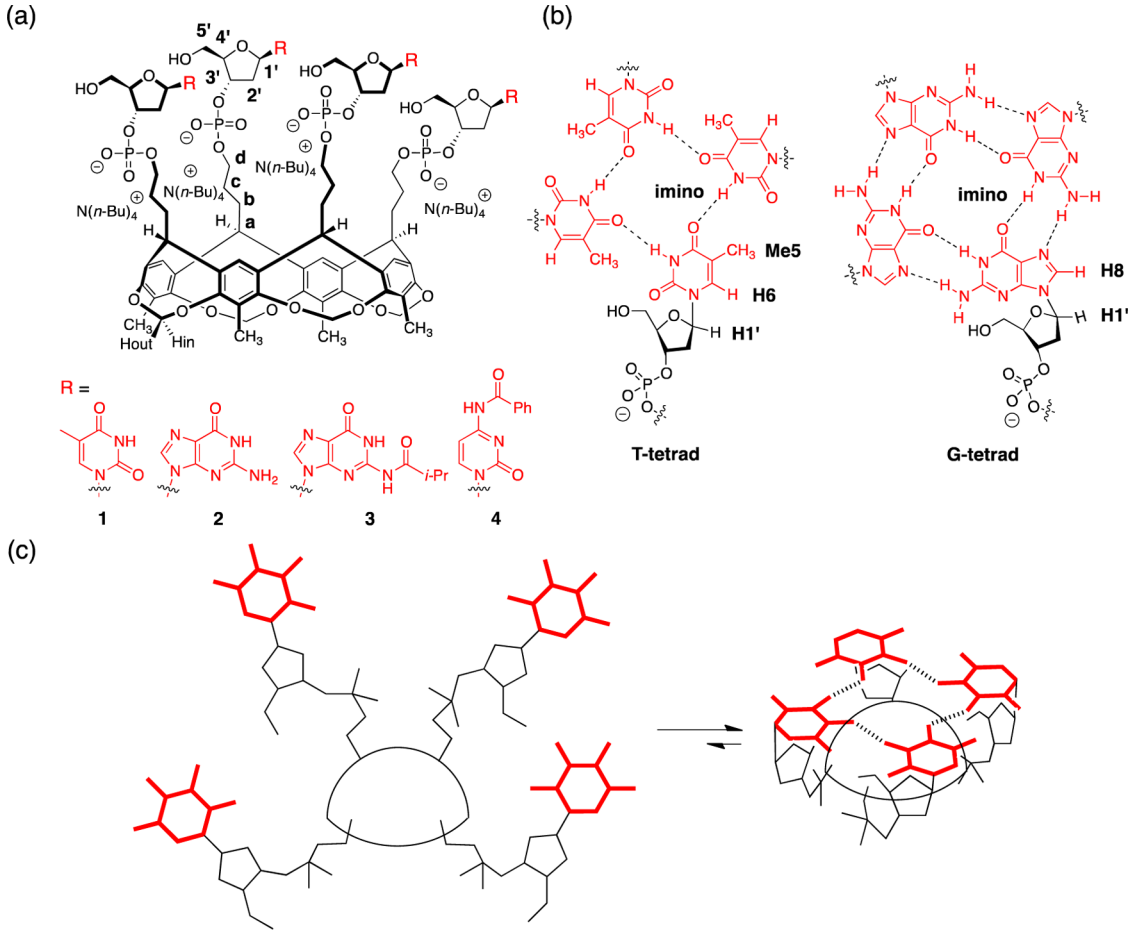


Figure 1. (a) Structures of template-assembled tetra-nucleotides. (b) A T-tetrad and G-tetrad assembly depicted with *syn* *N*-glycosidic bond conformation. (c) Equilibrium in methanol favoring a monomeric template-assembled T-tetrad for tetra-thymidylic acid **1** over an open non-hydrogen bonded conformation.

of guanine−guanine base pairing, tetra-deoxyguanylic acid **2** was anticipated to form G-tetrads in addition to compound **1**, provided that the phosphate-linker was sufficiently flexible for guanine bases to be oriented with correct spatial alignment.

Initial characterization by CD spectroscopy suggested tetrad assembly of tetra-nucleotides **1**−**2** (Figure 2). Tetra-thymidylic acid **1** exhibited CD with respective positive and negative bands at 269 and 218 nm. In comparison, the CD spectrum of the free thymidine nucleoside in methanol exhibited a shifted positive band at 274 nm along with a 46% and 83% reduction in CD signal at 274 and 218 nm, respectively. Despite the fact that the template linkage here forces parallel structures, tetra-deoxyguanylic acid **2** displayed a positive CD centered at 291 nm and a negative CD centered at 262 nm consistent with the observed CD relationship between stacked G-tetrads within antiparallel DNA G-quadruplexes.[13] The circular dichroism signature of the major parallel and antiparallel forms of G-quadruplex DNA is linked to *N*-glycosidic bond conformer identity, and only the antiparallel form contains *syn* conformers in addition to *anti*. Tetra-N^2-isobutyryl-deoxyguanylic acid **3** with added steric bulk at its hydrogen bond donor exocyclic amino position was selected as a G-tetrad disrupting control compound, and as predicted, its CD spectrum did not exhibit signature positive and negative bands of G-quadruplex. Solvent effects were apparent from the evaluation by CD of tetra-nucleotides **1** and **2** for tetrad self-assembly in water. Characteristic peak CD magnitude was considerably diminished in aqueous buffer in accordance with free nucleoside and base-blocked methanolic controls demonstrating that methanolic structures of **1** and **2** are disrupted in water.

Further details of template-assembled solution behavior were investigated through ^{1}H NMR methods. Inspection of the ^{1}H NMR spectra for compounds **1**−**2** in deuterated methanol revealed one distinct set of resonances consistent with a 4-fold rotational symmetry requirement in a tetrad nucleobase assembly (Figure 3). Coupling constants, $^3J_{1'2'}$ = 7.0 Hz, between H1′ and H2′ protons extracted from H1′ proton triplet signals at 6.30 and 6.21 ppm for tetra-nucleotides **1** and **2**, respectively, matched those observed for the predominately S-type C2′-endo sugar pucker conformation for deoxyribose sugars in B-DNA.[14] Observable resonances in the ^{1}H NMR spectra for guanine imino protons in protic solvent are demonstrative of base pairing of the Hoogsteen type seen in G-quadruplexes, and these resonances are typically found in the hydrogen bonded region downfield of 10.5 ppm (Figure 4).[15] The thymine imino resonance for an NH_4^+-G-quadruplex scaffolded T-tetrad has been reported at approximately 11.8 ppm.[10] Tetra-thymidylic acid **1** and tetra-deoxyguanylic acid **2** accordingly gave rise to imino resonances at 11.1 and 11.2 ppm in methanol, respectively, confirming hydrogen bonded base pairing. However for a sample of **2**, good resolution of its guanine imino signal required cooling to −25 °C, whereas the thymine imino signal of **1** was well resolved at room temperature. This result may be in line with a temperature

dx.doi.org/10.1021/jo401281p | *J. Org. Chem.* 2013, 78, 8198−8202

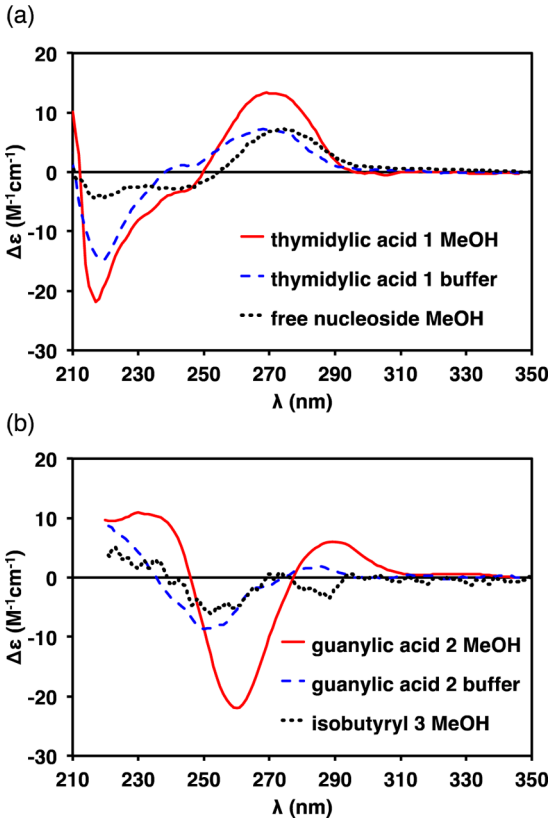


Figure 2. CD spectroscopy of 0.01 mM tetra-nucleotides **1**−**3** and 0.04 mM thymidine nucleoside in methanol or 10 mM Tris-Borate pH 8. Thymidine nucleoside CD was normalized to the nucleoside mass percent and stoichiometry of tetra-thymidylic acid **1**.

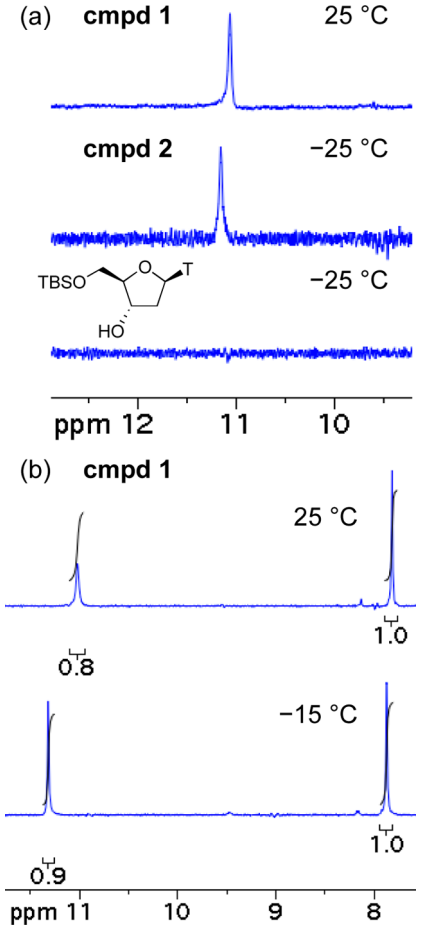


Figure 4. Imino region of ^{1}H NMR spectra at 400 MHz for 7.1 mg mL^{-1} solutions of tetra-thymidylic acid **1**, tetra-deoxyguanylic acid **2**, and 5′-TBS-5′-deoxythymidine in 9:1 CH_3OH/CH_3OD.

dependent process whereby chemical exchange of imino protons with solvent molecules is dependent on breaking open the tetrad in an analogous manner to Watson−Crick base pair opening described in B-DNA (Figure 1c).[16] Kinetic stabilization of the T-tetrad of **1** in contrast to the G-tetrad of **2** may reflect less than ideal sterics and or methanol solvation by polar methanol molecules for the bulkier guanine bases in these systems. A methanol soluble free nucleoside 5′-TBS-5′-deoxythymidine did not display any imino resonance down to −25 °C in comparison. Integration for **1** of the exchangeable imino resonance with respect to the H6 resonance gave integral ratios of 0.8 and 0.9 at 25 and −15 °C, respectively, confirming that the template-assembled T-tetrad is the major equilibrium species in methanol solvent. Investigation of thermodynamic stability differences through monitoring of UV absorbance as a function of temperature did not produce distinct melting

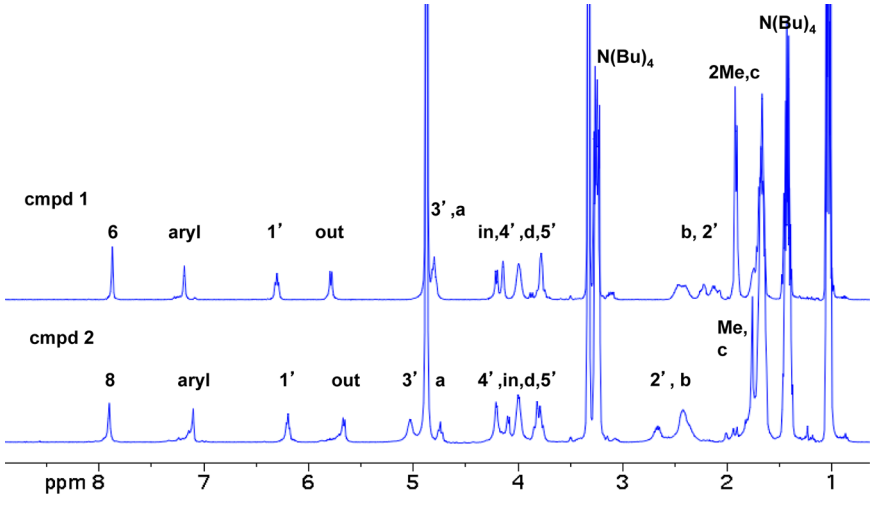


Figure 3. ^{1}H NMR spectra at 400 MHz for 7.1 mg mL^{-1} solutions of tetra-thymidylic acid **1** and tetra-deoxyguanylic acid **2** in CD_3OD at 25 °C.

dx.doi.org/10.1021/jo401281p | J. Org. Chem. 2013, 78, 8198−8202

temperatures typically observed for thermal denaturation of nucleic acid secondary structure for **1** or **2** below methanol's boiling point (data not shown).

Information about conformation and oligomeric association was obtained from 2D NMR experimentation. The sugar-base *N*-glycosidic bond conformation was confirmed as *syn* for both T-tetrad **1** and G-tetrad **2** through the observation of respective H6/H1′ ROE and H8/H1′ NOE cross-peaks (Figure 5).

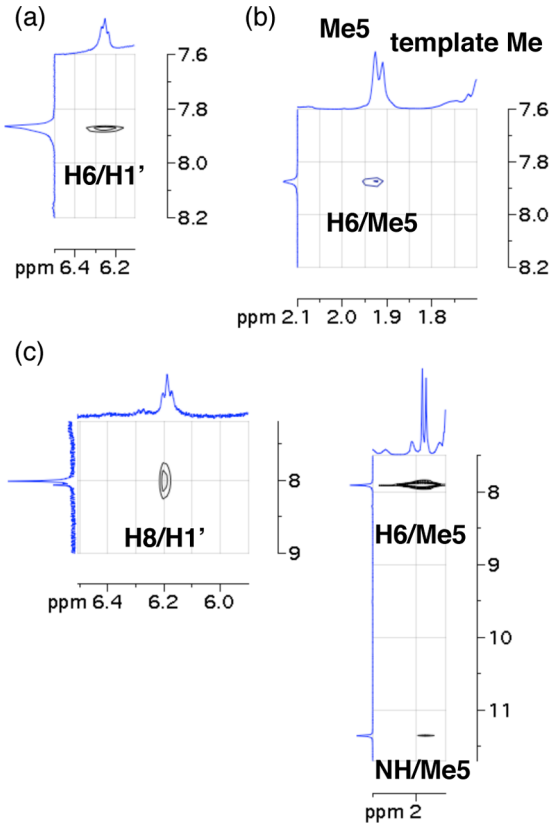


Figure 5. Key dipolar and scalar couplings observed for tetranucleotides **1**–**2** at 7.1 mg mL^{-1} concentrations on a 400 MHz spectrometer. (a) 2D NOESY spectrum of **2** in CD_3OD at −25 °C with 800 ms mixing time. (b) ^{1}H ^{1}H COSY spectrum of **1** in CD_3OD at 25 °C. (c) 2D ROESY spectrum of **1** in 9:1 CH_3OH/CD_3OD at −15 °C with WET suppression.

Furthermore, in the 2D ROESY spectrum of **1** at −15 °C, dipolar coupling between the NH and one of two methyl singlets in addition to expected H6/Me5 coupling was apparent suggesting the close spatial proximity of neighboring thymine residues in a tetrad arrangement through an NH-Me5 interaction.

Tetra-thymidylic acid **1** exhibits two methyl singlets at 1.93 and 1.91 ppm in CD_3OD at 25 °C arising from the thymine base Me5 and template Me groups. A ^{1}H ^{1}H COSY cross-peak with H6 centered on the downfield 1.93 ppm singlet allowed its assignment as the Me5. The Me5 of **1** shows a notable downfield chemical shift difference with respect to thymine methyl groups in B-DNA, which have been reported in the 1.39–1.55 ppm region.[17]

Routine measurement of diffusion coefficients using 2D DOSY NMR gave values within 99–97% of tetrad disrupting standard compounds in line with nonassociating monomeric species observed in lypophilic nucleobase systems.[11] Diffusion coefficients for 7.1 mg mL^{-1} solutions of G-tetrad **2** and tetra-N^2-isobutyryl-deoxyguanosine **3** were measured as 2.54×10^{-10} and 2.56×10^{-10} m^2 s^{-1}, respectively. For 7.1 mg mL^{-1} T-tetrad **1**, its diffusion coefficient was found to be 2.72×10^{-10} m^2 s^{-1} in comparison to 2.64×10^{-10} m^2 s^{-1} for hydrogen bond donor protected pyrimidine standard **4** at the identical concentration.

Although G-quadruplexes are observed to be dependent on metal cation localization to the G-tetrad centers via O6 lone pair coordination, we did not observe this dependence on tetrad formation for either putative T-tetrad or G-tetrad structures **1** or **2**. Cation-free stabilization is not an unexpected result for template-assembled synthetic structures.[11] We ruled out Na^+ contamination by showing that less than 10 mol % of sodium exists by inductively coupled plasma (ICP)-MS for phosphate-linked tetra-nucleotide preparations (see Experimental Section). The examination of the role of a cation in these systems is of interest. However, it was found that the introduction of salts gave rise to problematic line broadening effects in NMR samples attributed to aggregation and thus prevented their characterization.

We have shown here that the assembly of thymidylic acid into a hydrogen bonded T-tetrad array proceeds with the assistance of a template molecule without the need for a cation or G-quadruplex stabilization in a polar protic solvent. This effect is also seen for G-tetrads from templated deoxyguanylic acid. Synthetic modifications to template and linkage design may be required to stabilize template-assembled T- and G-tetrads in aqueous media. Further research into the self-assembling potential of the thymine base is necessary to conclude whether T-tetrads have a functional role in Nature.

EXPERIMENTAL SECTION

2D ROESY NMR. A 2D ROESY NMR experiment for a 7.1 mg mL^{-1} solution of **1** in 9:1 CH_3OH/CD_3OD was performed on a 400 MHz spectrometer at −15 °C by incorporating water suppression enhanced through the T1 effects (WET) pulse sequence into a standard Bruker 2D ROESY program to simultaneously suppress both methanol signals. A spin-lock pulse of 800 ms was used.

2D DOSY Diffusion Coefficient Measurements. DOSY experiments were performed at the 7.1 mg mL^{-1} concentration in CD_3OD on a 400 MHz spectrometer at 25 °C. A standard Bruker program was used with big and little δ values optimized to 100 and 4 ms respectively. Gradient strength was increased over 16 steps from 2% to 95%. Curve fitting using the XWINNMR SimFit algorithm for the signal decay of the ribose sugar 1′ proton resonance as a function of gradient strength provided the diffusion coefficients reported here.

ICP-MS Sodium Analysis. A sample of **1** was dissolved in 1% trace element nitric acid and diluted accordingly. Standard addition after calibration with sodium-23 and indium-115 internal standards and subtraction from a control sample gave a sodium concentration of 6.1 mol % relative to the concentration of **1**.

AUTHOR INFORMATION

Corresponding Author

*E-mail: sherman@chem.ubc.ca.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support.

REFERENCES

(1) Sen, D.; Gilbert, W. *Nature* **1988**, *334*, 364–366.

(2) Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. *Cell* **1989**, *59*, 871−880.
(3) Biffi, G.; Tannahill, D.; John McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182−186 and references cited therein.
(4) Patel, P. K.; Koti, A. S. R.; Hosur, R. V. *Nucleic Acids Res.* **1999**, *27*, 3836−3843.
(5) Searle, M. S.; Williams, H. E. L.; Gallagher, C. T.; Grant, R. J.; Stevens, M. F. G. *Org. Biomol. Chem.* **2004**, *2*, 810−812.
(6) Cheong, C.; Moore, P. B. *Biochemistry* **1992**, *31*, 8406−8414.
(7) Deng, J.; Xiong, Y.; Sundaralingam, M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13665−13670.
(8) Xu, Y.; Ishizuka, T.; Kimura, T.; Komiyama, M. *J. Am. Chem. Soc.* **2010**, *132*, 7231−7233.
(9) Caceres, C.; Wright, G.; Gouyette, C.; Parkinson, G.; Subirana, J. A. *Nucleic Acids Res.* **2004**, *32*, 1097−1102.
(10) Sket, P.; Plavec, J. *J. Am. Chem. Soc.* **2010**, *132*, 12724−12732.
(11) Hui, B. W.-Q.; Sherman, J. C. *ChemBioChem* **2012**, *13*, 1865−1868 and references cited therein.
(12) Bare, G. A. L.; Sherman, J. C. *Tetrahedron Lett.* **2013**, *54*, 3207−3209.
(13) Masiero, S.; Trotta, R.; Pieraccini, S.; De Tito, S.; Perone, R.; Randazzo, A.; Spada, G. P. *Org. Biomol. Chem.* **2010**, *8*, 2683−2692.
(14) Feigon, J.; Leupin, W.; Denny, W. A.; Kearns, D. R. *Biochemistry* **1983**, *22*, 5930−5942.
(15) Henderson, E.; Hardin, C. C.; Walk, S. K.; Tinoco, I.; Blackburn, E. H. *Cell* **1987**, *51*, 899−908.
(16) Gueron, M.; Kochoyan, M.; Leroy, J. *Nature* **1987**, *328*, 89−92.
(17) Feigon, J.; Denny, W. A.; Leupin, W.; Kearns, D. R. *Biochemistry* **1983**, *22*, 5943−5951.