JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Structure of a Stable G-Hairpin

Martin Gajarský,[†,▽] Martina Lenarčič Živković,[‡,▽] Petr Stadlbauer,[§] Bruno Pagano,[⊥] Radovan Fiala,[†] Jussara Amato,[⊥] L'ubomír Tomáška,[∥] Jiří Šponer,[†,§] Janez Plavec,*[,‡,⊗,#] and Lukáš Trantírek*[,†]

[†]Central European Institute of Technology, Masaryk University, Kamenice 753/5, 62500 Brno, Czech Republic

[‡]Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia

[§]Institute of Biophysics, v.v.i., Academy of Sciences of the Czech Republic, Kralovopolska 135, 61265 Brno, Czech Republic

[⊥]Department of Pharmacy, University of Naples "Federico II", Via D. Montesano 49, I-80131 Naples, Italy

[∥]Department of Genetics, Faculty of Natural Sciences, Comenius University, Mlynska dolina B-1, Ilkovicova 6, 84215 Bratislava, Slovakia

[⊗]EN-FIST Centre of Excellence, Trg Osvobodilne fronte 13, SI-1001 Ljubljana, Slovenia

[#]Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, SI-1000 Ljubljana, Slovenia

Supporting Information

ABSTRACT: In this study, we report the first atomic resolution structure of a stable G-hairpin formed by a natively occurring DNA sequence. An 11-nt long G-rich DNA oligonucleotide, 5′-d(GTGTGGGTGTG)-3′, corresponding to the most abundant sequence motif in irregular telomeric DNA from *Saccharomyces cerevisiae* (yeast), is demonstrated to adopt a novel type of mixed parallel/antiparallel fold-back DNA structure, which is stabilized by dynamic G:G base pairs that transit between N1-carbonyl symmetric and N1-carbonyl, N7-amino base-pairing arrangements. Although the studied sequence first appears to possess a low capacity for base pairing, it forms a thermodynamically stable structure with a rather complex topology that includes a chain reversal arrangement of the backbone in the center of the continuous G-tract and 3′-to-5′ stacking of the terminal residues. The structure reveals previously unknown principles of the folding of G-rich oligonucleotides that could be applied to the prediction of natural and/or the design of artificial recognition DNA elements. The structure also demonstrates that the folding landscapes of short DNA single strands is much more complex than previously assumed.

It is well-known that DNA is an abundant biopolymer in all living organisms, where it functions in encoding, transmitting, and expressing genetic information. Based on sequence composition and environmental conditions, DNA can adopt diverse conformations, which predefine its physiological function. Aside from double helical B-DNA, DNA is capable of adopting alternative conformations, including parallel/antiparallel duplexes, various hairpins, triplexes and tetraplexes, and branched architectures. These so-called *non-B* DNA structures have been demonstrated to participate in cellular regulation, and their formation has also been associated with a number of human diseases.[1] In the past two decades, *non-B* DNA motifs have become validated targets for new anticancer drugs, and many leading compounds that target these motifs have entered (pre)-clinical trials.[2] In addition to the biological significance of *non-B* DNA, its unique nanoscale geometry, biocompatibility, biodegradability, and molecular recognition capacity have made *non-B* DNA motifs promising candidates for the construction of novel functional nanomaterials and nanodevices.[3]

Undoubtedly, most of the attention in *non-B* DNA-forming sequences currently is paid to G-rich DNA, particularly that displaying the potential to form G-quadruplexes.[1−4] Recent studies demonstrated the existence of G-quadruplexes *in vivo*[5] and their essential roles in the regulation of a number of cellular processes.[1] At the same time, synthetic G-quadruplexes became established elements in nanotechnology and found their widespread use as aptamers for therapeutic/diagnostic purposes.[3] Lately, the interest in the exploration of unique features of the conformational space of G-rich DNA has further intensified in reaction to the discovery of a stable G-triplex.[6] Until recently, the G-triplex has been considered as a (hypothetical) intermediate in the G-quadruplex folding.[7−11] Observation of stable G-triplex formation under physiological conditions[6] not only highlighted an unexpected richness in the conformational space of G-rich DNA but also suggested that other formerly hypothesized structures, including that of a G-hairpin,[8,11−13] a fold-back structure stabilized by G:G base pairs, might exist. Although the notion of the existence of a G-hairpin structure has recently been supported by the observation of the G-hairpin formation under enforced conditions of DNA origami[14] and theoretical MD simulations,[13] the formation of a stable G-hairpin from a natively occurring DNA sequence has never been experimentally observed. Here, we show that an 11-nt long sequence, 5′-d(GTGTGGGTGTG)-3′ (hereby referred to as SC11), the most abundant sequence motif from irregular telomeric DNA of *Saccharomyces cerevisiae*, which is also a binding motif for Cdc13, a single-stranded telomeric DNA binding protein in budding yeast,[15] adopts a stable G-hairpin structure.

Examination of a 1D ^{1}H NMR spectrum of SC11 in potassium buffer (100 mM KCl, 10 mM KPO_i, pH 7.0) has

Received: October 19, 2016
Published: February 20, 2017

DOI: 10.1021/jacs.6b10786
J. Am. Chem. Soc. 2017, 139, 3591−3594

shown seven imino proton resonances (Figure 1A). All imino proton resonances appear between δ 11.2 and 12.4 ppm, a

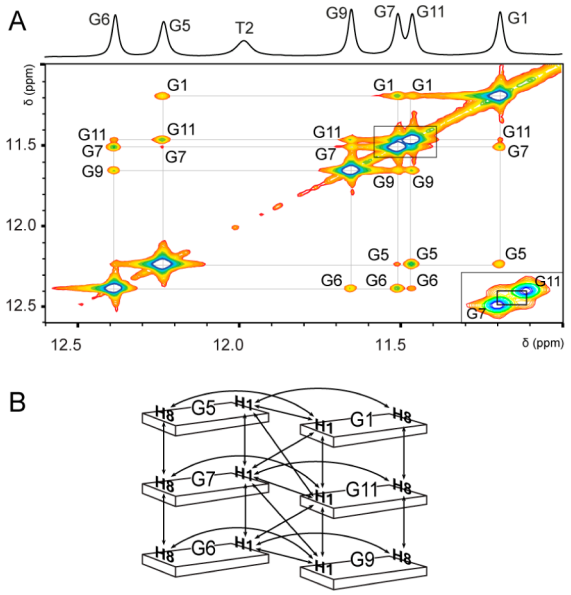

Figure 1. (A) Top: Imino region of the 1D ^{1}H NMR spectrum of SC11. Bottom: The imino−imino region of 2D NOESY spectrum (τ_m = 150 ms) in $^{1}H_2O/^{2}H_2O$ (90:10). (B) Schematic representation of NOE connectivities (indicated by arrows) between imino and aromatic protons for core guanine residues observed in the NOESY spectrum of SC11.

typical region for imino protons involved in hydrogen bonds between either G-to-G and/or G-to-T residues.[16] The native PAGE (Figure S1) and concentration-independent melting temperature (*vide infra*) showed that the SC11 forms a monomeric structure. The exchangeable protons were assigned with the help of the 1D ^{15}N-edited HSQC spectra acquired on residue-specific $^{15}N/^{13}C$-labeled oligonucleotides (Figure S2, S3) and 2D NOESY spectra acquired in $^{1}H_2O/^{2}H_2O$ (90:10) potassium-based buffer (for the assignment of proton chemical shifts, cf. Table S1). Inspection of the NOESY spectra revealed a rather intricate network of inter-residual NOE connectivities involving imino and aromatic protons from residues G1, G5, G6, G7, G9, and G11 (Figure 1B). Next to sequential H1−H1 and H8−H8 NOE connectivities (Figure 1A,B and 2B), connectivities defining base pairing, namely H1−H1 characteristic for N1-carbonyl symmetric G:G base pairs and H1−H8 NOE connectivities characteristic for both *H-WC* and *WC-H* types of N1-carbonyl, N7-amino G1:G5, G11:G7, and G9:G6 base pairs, were observed (for definition of the base pairing arrangements, see Figure S4).[16,17] The concurrent observation of NOE connectivities, which are inconsistent with single G:G base pairing and of only one set of resonances for given nucleotide in NMR spectra, is indicative of dynamic rearrangements of G:G base-pairs on time scale $\ll$1 μs (*vide infra*).[18]

Inspection of the NOESY spectra revealed two additional distinctive features of the SC11 structure, including a chain reversal arrangement of the backbone positioning residue G7 between residues G5 and G6 and a 3′-to-5′ stacking of terminal residues (G1 and G11). The position of G7 between G5 and G6 is evidenced by the observation of NOE cross-peaks between the sugar protons of G5 and the H8 proton of G7 and between the H8 proton of G5 and the H8 proton of G7 (Figure

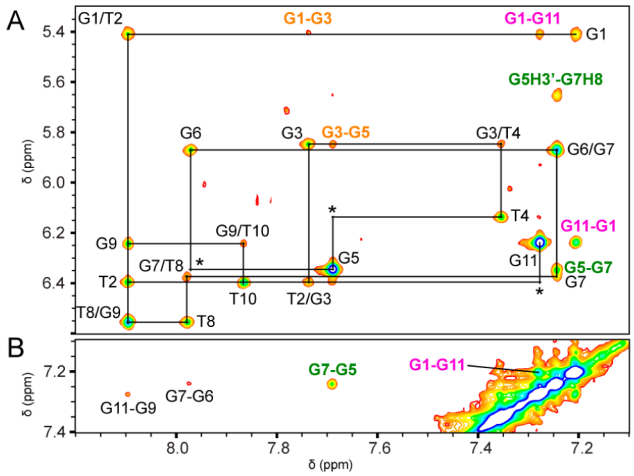

Figure 2. (A) Aromatic−anomeric region of the 2D NOESY spectrum (τ_m = 100 ms) in $^{2}H_2O$ with the indicated sequential-walk (H1′$_n$-H8/6$_{n+1}$). The missing peaks are designated with an asterisk. Although the T4/G5 and T10/G11 H1′$_n$-H8$_{n+1}$ NOE cross-peaks are missing due to consecutive *anti-syn* conformations along the glycosidic bond,[16] the absence of the G5/G6 cross-peak is a consequence of the chain reversal. (B) Aromatic−aromatic region of the same spectrum. Cross-peaks indicative of a chain reversal, 3′-to-5′ stacking, and stacking of G3 to G5:G1 base pair are shown in dark green, magenta, and orange, respectively.

2, dark green cross-peaks). The unusual G5-G7-G6 backbone arrangement is further corroborated by the absence of a sequential H1′−H8 NOE cross-peak between G5 and G6 (Figure 2A). G1-to-G11 stacking is supported by the observation of NOE cross-peaks between H1′ of G1 and H8 from G11, between H1′ of G11 and H8 of G1, and between G1 and G11 H8 protons (Figure 2, magenta cross-peaks). Furthermore, observation of NOE cross-peaks between H1′ of G1 and H8 of G3 and between H1′ of G3 and H8 of G5 suggest a stacking of residue G3 with G1 and G5 (Figure 2A, orange cross-peaks).

In total, 251 NOE-derived distance restraints, along with six hydrogen bonds, 22 torsion angles, and six planarity restraints, were used to calculate the high-resolution structure of the SC11 (for details, see Material and Methods). A conformational ensemble of the 10 lowest energy structures from a restrained MD simulation is shown in Figure 3A (for NMR restraints and structure statistics, see Table S2). The structural features of the refined model are schematically represented in Figure 3B and Figure S5. SC11 features a fold-back-like structure stabilized by a network of hydrogen bonds connecting guanine residues that form two N1-carbonyl symmetric base pairs by the G5:G1 and G7:G11 and a buckled G6:G9 base pair in an N1-carbonyl, N7-amino-like arrangement of the *WC-H* type (Figure 3C). The structure is classified as a mixed parallel/antiparallel hairpin, due to the chain reversal G5-G7-G6 backbone arrangement. The 3′-to-5′ stacking of terminal residues (G1 and G11) leads to the formation of a discontinuous G-tract from residues G1, G11, and G9. The loop residues also seem to contribute to the stability of the structure as follows: The O2 atom of T2 is positioned to interact with the amino proton at the sugar-edge of G1. T4 is positioned on the top of G3 in a stacking-like arrangement, whereas residue G3 stacks on the top of the G1 base paired with G5. The T8 and T10 residues, instead of flipping outward, collapse onto the terminal G6:G9 base pair,

DOI: 10.1021/jacs.6b10786
J. Am. Chem. Soc. 2017, 139, 3591−3594

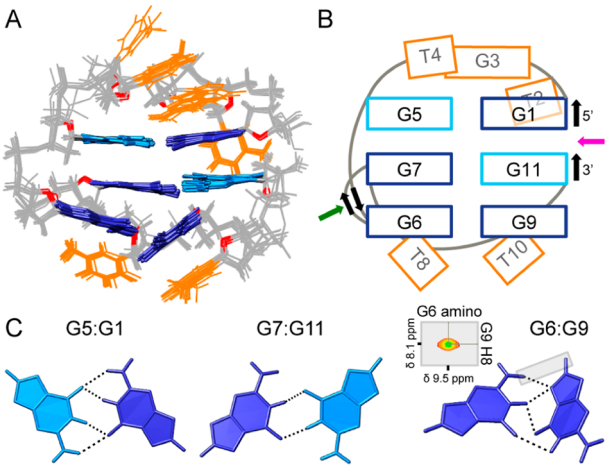

Figure 3. High-resolution structure of SC11 (PDB ID: 5M1W). (A) Ten lowest-energy structures. Loop residues are colored orange and O4′ atoms are colored red. (B) Schematic representation of SC11 folding topology. Chain reversal arrangement of the backbone and the 3′-to-5′ stacking of the terminal residues are indicated by dark green and magenta arrows, respectively. *Anti* and *syn* guanines that form G:G base pairs are colored dark and light blue, respectively. (C) Base pairing arrangement for G5:G1, G7:G11, and G6:G9 in the lowest energy structure. Hydrogen bonding patterns in individual base pairs are indicated by dashed lines. Above G6:G9 base pair: Amino–aromatic region of the 2D NOESY spectrum (τ_m = 100 ms) in $^1H_2O/^2H_2O$ (90:10). Strong cross-peak between amino proton from G6 and H8 proton from G9 supports the *WC-H* type arrangement for G6:G9 base pair.

allowing the O4′ of the thymines to hydrogen bond with the amino protons of G6 and G9, respectively.

To assess potential presence of submicrosecond dynamics, implicated by the appearance of NOESY spectra (*vide supra*), the selected NMR structures were subjected to unrestrained MD simulations in explicit solvent (cf. Material and Methods). Ten independent 2.5 μs long MD simulations were performed. In eight of them, the SC11 structure was stable over the whole simulation period (Table S4). All major structural features displayed by the NMR structure were preserved in the MD simulations, including the chain reversal and three G:G base pairs. Most importantly, inspection of the MD trajectories revealed the dynamic nature of the G5:G1 and G7:G11 base pairs, particularly the transitions between three base pairing arrangements, namely the N1-carbonyl symmetric and N1-carbonyl, N7-amino states (Figure S6, for occupancies of individual states cf. Table S5). Within the simulations, the G5:G1 base pair preferentially adopted the N1-carbonyl symmetric state, although both N1-carbonyl, N7-amino arrangements (i.e., *H-WC* and *WC-H*) were also observed. The G7:G11 base pair was observed to transit between two arrangements, N1-carbonyl symmetric and N1-carbonyl, N7-amino *WC-H* (Figure S6, Table S5). At the time-scale of the simulation (2.5 μs), almost no transitions were observed for the G6:G9 base pair, which remained in the N1-carbonyl, N7-amino *WC-H*-like state (Figure S6, Table S5). Inspection of MD trajectories revealed that time required for transitions between individual G:G hydrogen bonding registers was between 0.1 and 71.7 *ns* (with average transition time of ∼10 *ns*, cf. Table S5). Observation of almost no base pair transitions for the G6:G9 and no transitions to the *H-WC* arrangement for the G7:G11 base pair (presence of which was implicated by the NMR data) in the course of the MD simulations is likely to be a consequence of limited ability of MD to sample conformational space, which makes MD simulation outcome strongly dependent on geometry of starting (NMR) structure(s). Altogether, the MD simulation results support dynamic nature of G:G base-pairing in SC11 structure and suggest that the derived NMR structure should be considered as a representation of ensemble averaged experimental data.

In addition, the structure and stability of SC11 were investigated by CD spectroscopy and differential scanning calorimetry (DSC). As expected, the CD spectrum of denatured SC11 at 90 °C showed no significant bands. Conversely, at 5 °C, it showed a positive band at 270 nm with a shoulder at approximately 300 nm and a weak negative band at approximately 245 nm (Figure S7A). CD melting and annealing profiles of SC11, recorded at 0.5 °C/min, showed a considerable hysteresis (Figure S7B), indicating that the thermal melting curve was not at thermodynamic equilibrium because of a slow rate of association and/or dissociation of SC11 structure. Under these conditions, the full thermodynamic characterization is impractical. Only the apparent melting temperature ($T_{1/2}$) at specific heating rate and the calorimetric enthalpy change relative to the G-hairpin dissociation process could be directly determined.[19] The DSC melting curves of the G-hairpin, recorded at two different SC11 concentrations (310 and 620 μM) using the 0.5 °C/min heating rate (Figure S7C), exhibited a peak with a maximum at the $T_{1/2}$ of 53.4 (±0.6) °C. The integration of the denaturation peaks gives an average enthalpy change value (ΔH^0) of 115 (±8) kJ/mol, which is fully compatible with the number and type of base pairs determined by NMR that contribute to the stability of this structure.[20] Interestingly, the $T_{1/2}$ acquired by DSC is in good agreement with that obtained by CD at the same heating rate and concentration of 2.5 μM ($T_{1/2}$ = 52.2 (±1.0) °C). The virtual independence of $T_{1/2}$ on SC11 concentration is consistent with native PAGE experiment showing that the SC11 structure results from unimolecular folding.[21] It is noteworthy that the presence of Na^+ ions seems to interfere with the formation of a well-defined structure from the SC11 sequence (including that of the G-hairpin) (Figure S8).

For a long time, G-hairpins have been hypothesized to exist as short-lived intermediates in the G-quadruplex folding process.[8,11−13,22] Our present data demonstrate that G-hairpins might also be unexpectedly stable, spontaneously formed structures. However, it is important to emphasize that the reported DNA G-hairpin structure is considerably different from the Hoogsteen stem–loop G-hairpins that were suggested as early intermediates in G-quadruplex folding. Along with a variety of right-handed G-quadruplex structures,[23] a recently discovered DNA G-triplex,[6] and a left-handed G-quadruplex,[24] the structure of the G-hairpin presented here highlights the extraordinary complexity of the conformational space of G-rich DNA sequences. The studied sequence first appears to possess a low capacity for base pairing. Despite this, it forms a thermodynamically stable structure with a rather complex topology. This indicates that the folding landscapes of short DNA single strands may be much more intricate than previously assumed. This finding may have substantial implications for the folding pathways of noncanonical DNA sequences. The structure reveals previously unknown principles of the folding of G-rich oligonucleotides that could be applied to the prediction of natural and/or the design of artificial

DOI: 10.1021/jacs.6b10786
J. Am. Chem. Soc. 2017, 139, 3591−3594

recognition DNA elements. This fold also extends the toolbox of motifs available for the construction of DNA assemblies.[25]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b10786.

Material and Methods; additional experimental data (Figures S1–S9 and Tables S1–S5). (PDF)

Starting structures used in MD simulations (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

*janez.plavec@ki.si

*lukas.trantirek@ceitec.muni.cz

ORCID

Lukáš Trantírek: 0000-0001-5948-4837

Author Contributions

[▽]These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This project was supported by grants from the Czech Science Foundation [13-28310S, 16-13721S], project CEITEC 2020 (LQ1601) with financial contribution from the MEYS CR and National Programme for Sustainability II, the EMBO Installation grant [IG2535], Marie-Curie grant [ECOPOD], the Slovenian Research Agency [ARRS, grant nos. P1-242 and J1-6733], the Slovak Research and Development Agency (APVV-15-0022), and by "Programma STAR" 2014 of University of Naples "Federico II" [no. 14-CSP3-C03-141]. The authors acknowledge the CERIC-ERIC Consortium and the CIISB research infrastructure (LM2015043 funded by MEYS CR) for access to experimental facilities and financial support.

■ REFERENCES

(1) (a) Bacolla, A.; Wells, R. D. *Mol. Carcinog.* **2009**, *48*, 273. (b) Maizels, N. *EMBO Rep.* **2015**, *16*, 910. (c) Rhodes, D.; Lipps, H. J. *Nucleic Acids Res.* **2015**, *43*, 8627.

(2) Ali, A.; Bhattacharya, S. *Bioorg. Med. Chem.* **2014**, *22*, 4506.

(3) (a) Tang, Y.; Ge, B.; Sen, D.; Yu, H. Z. *Chem. Soc. Rev.* **2014**, *43*, 518. (b) Yatsunyk, L. A.; Mendoza, O.; Mergny, J. L. *Acc. Chem. Res.* **2014**, *47*, 1836.

(4) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2006**, *35*, 406.

(5) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182.

(6) Limongelli, V.; De Tito, S.; Cerofolini, L.; Fragai, M.; Pagano, B.; Trotta, R.; Cosconati, S.; Marinelli, L.; Novellino, E.; Bertini, I.; Randazzo, A.; Luchinat, C.; Parrinello, M. *Angew. Chem., Int. Ed.* **2013**, *52*, 2269.

(7) Boncina, M.; Lah, J.; Prislan, I.; Vesnaver, G. *J. Am. Chem. Soc.* **2012**, *134*, 9657.

(8) (a) Gray, R. D.; Chaires, J. B. *Nucleic Acids Res.* **2008**, *36*, 4191. (b) Gray, R. D.; Trent, J. O.; Chaires, J. B. *J. Mol. Biol.* **2014**, *426*, 1629.

(9) Koirala, D.; Mashimo, T.; Sannohe, Y.; Yu, Z.; Mao, H.; Sugiyama, H. *Chem. Commun.* **2012**, *48*, 2006.

(10) Li, W.; Hou, X. M.; Wang, P. Y.; Xi, X. G.; Li, M. *J. Am. Chem. Soc.* **2013**, *135*, 6423.

(11) Zhang, A. Y. Q.; Balasubramanian, S. *J. Am. Chem. Soc.* **2012**, *134*, 19297.

(12) (a) Li, Y.; Liu, C.; Feng, X. J.; Xu, Y. Z.; Liu, B. F. *Anal. Chem.* **2014**, *86*, 4333. (b) Mashimo, T.; Yagi, H.; Sannohe, Y.; Rajendran, A.; Sugiyama, H. *J. Am. Chem. Soc.* **2010**, *132*, 14910.

(13) Stadlbauer, P.; Kuhrova, P.; Banas, P.; Koca, J.; Bussi, G.; Trantirek, L.; Otyepka, M.; Sponer, J. *Nucleic Acids Res.* **2015**, *43*, 9626.

(14) (a) Rajendran, A.; Endo, M.; Hidaka, K.; Sugiyama, H. *Angew. Chem., Int. Ed.* **2014**, *53*, 4107. (b) Rajendran, A.; Endo, M.; Hidaka, K.; Teulade-Fichou, M. P.; Mergny, J. L.; Sugiyama, H. *Chem. Commun.* **2015**, *51*, 9181.

(15) (a) Eldridge, A. M.; Halsey, W. A.; Wuttke, D. S. *Biochemistry* **2006**, *45*, 871. (b) Lloyd, N. R.; Dickey, T. H.; Hom, R. A.; Wuttke, D. S. *Biochemistry* **2016**, *55*, 5326. (c) Mitton-Fry, R. M.; Anderson, E. M.; Theobald, D. L.; Glustrom, L. W.; Wuttke, D. S. *J. Mol. Biol.* **2004**, *338*, 241.

(16) Adrian, M.; Heddi, B.; Phan, A. T. *Methods* **2012**, *57*, 11.

(17) Kocman, V.; Plavec, J. *Nat. Commun.* **2014**, *5*, 5831.

(18) Cavanagh, J, Fairbrother, W. J., Palmer, A. G., Rance, M, Skeltonv, M. J. *Protein NMR Spectroscopy*; Elsevier Academic Press: London, U.K., 2007; pp 333–403.

(19) Petraccone, L.; Pagano, B.; Esposito, V.; Randazzo, A.; Piccialli, G.; Barone, G.; Mattia, C. A.; Giancola, C. *J. Am. Chem. Soc.* **2005**, *127*, 16215.

(20) Antao, V. P.; Tinoco, I. *Nucleic Acids Res.* **1992**, *20*, 819.

(21) Mergny, J. L.; Lacroix, L. *Oligonucleotides* **2003**, *13*, 515.

(22) Ceru, S.; Sket, P.; Prislan, I.; Lah, J.; Plavec, J. *Angew. Chem., Int. Ed.* **2014**, *53*, 4881.

(23) Karsisiotis, A. I.; O'Kane, C.; Webba da Silva, M. *Methods* **2013**, *64*, 28.

(24) Chung, W. J.; Heddi, B.; Schmitt, E.; Lim, K. W.; Mechulam, Y.; Phan, A. T. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2729.

(25) Seeman, N. C. *Annu. Rev. Biochem.* **2010**, *79*, 65.

DOI: 10.1021/jacs.6b10786
J. Am. Chem. Soc. 2017, 139, 3591–3594