JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Specific Binding of Modified RGG Domain in TLS/FUS to G-Quadruplex RNA: Tyrosines in RGG Domain Recognize 2′-OH of the Riboses of Loops in G-Quadruplex

Kentaro Takahama and Takanori Oyoshi*

Faculty of Science, Department of Chemistry, Shizuoka University, 836 Ohya Suruga, Shizuoka, 422-8529, Japan

Supporting Information

ABSTRACT: Telomeric repeat-containing RNA (TERRA), which contains tandem arrays of short RNA repeats, r(UUAGGG), is an integral component of the telomere and contributes to telomeric heterochromatin formation and telomere-length regulation. TERRA forms a G-quadruplex, but the biologic significance of its G-quadruplex formation is unknown. Compounds that selectively bind to G-quadruplex RNA are useful for understanding G-quadruplex TERRA. Here we report that an engineered RGG domain translocated in liposarcoma (TLS) specifically binds to G-quadruplex TERRA. The Arg-Gly-Gly repeat (RGG) TLS binds to G-quadruplex human telomere DNA and TERRA simultaneously, but we show that substitution of Tyr for Phe in the RGG domain of TLS (TLSRGG3Y) converts its binding specificity solely toward G-quadruplex TERRA. TLSRGG3Y binds to dG tetrads with abasic RNA loops, but fails to bind to rG tetrads without loops or dG tetrads with abasic DNA loops. These findings suggest that TLSRGG3Y binds to loops within the G-quadruplexes of TERRA by recognizing the 2′-OH of the riboses. To our knowledge, TLSRGG3Y is the first known molecule that specifically recognizes the 2′-OH of the riboses of loops in the G-quadruplex. TLSRGG3Y will be useful for investigating the role of the G-quadruplex form of TERRA without affecting G-quadruplex telomere DNA functions.

Mammalian telomeres are specialized structures that cap chromosome termini to prevent chromosome loss and degradation, and are complexed with several proteins and telomeric repeat-containing RNA (TERRA).[1−3] Mammalian telomere DNA comprises tandem arrays of d(TTAGGG) repeats and forms a G-quadruplex.[4] TERRA, which contains tandem arrays of r(UUAGGG) repeats, forms the parallel propeller type of G-quadruplex.[5,6] TERRA performs various cellular functions, such as regulating telomere length,[7] forming telomeric heterochromatin,[8,9] and protecting the telomere.[2,3] The G-quadruplex-dependent functions of TERRA in telomere maintenance in human cells, however, are not clear. The functions of G-quadruplex DNA in gene promoters, telomeres, and genomes have been elucidated using G-quadruplex DNA-binding molecules.[10−12] Small molecules targeting the G-quadruplex in the 5′-untranslated regions of mRNA that regulate translation initiation could modulate translational activity by stabilizing or destabilizing the G-quadruplex structures.[13] G-quadruplex RNA-binding molecules will be useful for elucidating TERRA functions, but little is known about the molecules that bind to G-quadruplex TERRA without binding to G-quadruplex telomere DNA.[14−16] The negatively charged adduct derived from pyridostatin preferentially stabilizes the G-quadruplex of TERRA compared with the G-quadruplex of human telomere DNA (Htelo).[17] Quantitative values for the G-quadruplex TERRA binding affinity and the recognition mechanism of G-quadruplex TERRA binding, however, remain unclear.

We recently identified that the Arg-Gly-Gly repeat (RGG) domain of the C-terminal region (TLSRGG3) of translocated in liposarcoma (TLS) protein, also termed FUS, binds simultaneously to G-quadruplex Htelo and TERRA with structural specificity, but not to double-stranded telomere DNA, single-stranded DNA, or RNA (Figure 1A).[18] Ewing's Sarcoma (EWS) is related to TLS as a family of proteins containing RGG domains and binds both G-quadruplex Htelo and TERRA-like TLS (Figure 1A).[19,20] To understand the recognition mechanism of TLSRGG3, we analyzed the G-

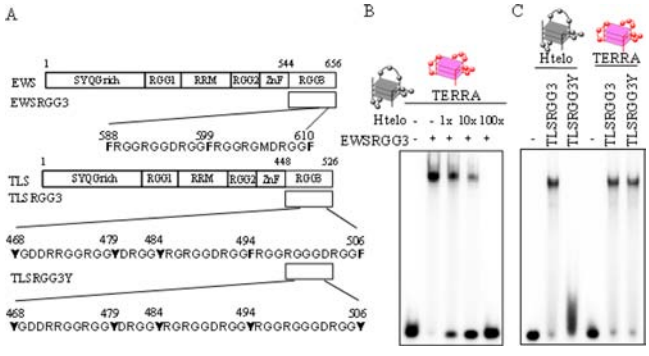

Figure 1. Identification of responsible residues at TLSRGG3 for TERRA binding selectivity. (A) Schematic illustration of amino acids 588−610 within EWSRGG3 and amino acids 468−506 within TLSRGG3 and TLSRGG3Y. Aromatic amino acids in RGG regions are shown in boldface type. (B) Binding competition assay, assaying binding of EWSRGG3 to ^{32}P-labeled TERRA in the presence of unlabeled Htelo at the indicated molar ratios of unlabeled/labeled nucleic acids. The structures of TERRA and Htelo are indicated. (C) EMSA was performed with TLSRGG3 or TLSRGG3Y with ^{32}P-labeled Htelo or TERRA. Red and gray in the cartoon shown in (B) and (C), respectively, represent RNA and DNA.

Received: August 29, 2013
Published: November 19, 2013

   dx.doi.org/10.1021/ja4086929 | J. Am. Chem. Soc. 2013, 135, 18016−18019

quadruplex DNA and RNA binding abilities of the RGG domain of the C-terminal region of EWS (EWSRGG3), and compared the binding abilities of TLSRGG3 and EWSRGG3. We examined ^{32}P-labeled TERRA binding of EWSRGG3 in the presence of cold Htelo using an electrophoretic mobility shift assay (EMSA; Table S1, Figure 1B). All G-quadruplex structures used in this paper were confirmed by circular dichroism (CD) spectroscopy (Table S1, Figure S1),[19,20] and each structure used as a probe is indicated above each lane. TERRA binding of EWSRGG3 was reduced by the addition of excessive Htelo competitor. Similarly, Htelo binding of EWSRGG3 was reduced by the addition of excessive TERRA competitor (Figure S2). These binding properties are in sharp contrast to those of TLSRGG3, which forms a ternary complex with G-quadruplex Htelo and TERRA.

We previously reported that substitution of three Phe with Ala in EWSRGG3 prevents its G-quadruplex binding.[20] This finding indicates that Phe is important for specific G-quadruplex binding. TLSRGG3 contains three Tyr in addition to two Phe as aromatic amino acids. To evaluate the role of Tyr and Phe in TLSRGG3 in G-quadruplex DNA and RNA recognition, we performed an EMSA to examine the impact of the simultaneous substitution of two Phe by Tyr within RGG3 on TLS (TLSRGG3Y) binding to TERRA or Htelo (Figure 1A and C). Tyr substitution dramatically reduced Htelo binding but did not affect TERRA binding. To test whether TLSRGG3Y specifically binds to a G-quadruplex form of TERRA, we analyzed the binding between TLSRGG3Y and a mutated TERRA r[UUAGGG(UUAGUG)$_2$UUAGGG] (mut TERRA), which fails to form the G-quadruplex.[19] TLSRGG3Y bound to the TERRA folded in a G-quadruplex, but not to mut TERRA (Figure S3A). TLSRGG3Y-TERRA binding was not affected by the addition of excessive Htelo or mut TERRA, whereas TLSRGG3Y-TERRA binding was affected by the addition of excessive TERRA (Figure S3B–D). To confirm the G-quadruplex TERRA in the presence of TLSRGG3Y, we performed a CD spectroscopic analysis (Figure S4), which revealed that the addition of 1.2 equiv of TLSRGG3Y did not alter the positive band at 265 nm as a parallel structure. These findings indicate that TLSRGG3Y binds specifically to G-quadruplex TERRA.

To determine the binding affinity and stoichiometry of G-quadruplex and TLSRGG3Y or TLSRGG3, we examined G-quadruplex binding of TLSRGG3Y and TLSRGG3 by isothermal titration calorimetry (ITC).[21,22] Figure 2 shows the results of the calorimetric titrations of TLSRGG3Y (A) or TLSRGG3 (C) with TERRA-containing solution and binding isotherms resulting from the integration of raw calorimetric data correction for the heat of TERRA dilution. The calorimetric titrations and binding isotherms of TLSRGG3 with Htelo are shown in Figure S5A. The binding isotherms of TLSRGG3Y or TLSRGG3 with TERRA or TLSRGG3 with Htelo revealed that all of the bindings are exothermic with a similar calculated K_d [10 ± 1 nM (TERRA/TLSRGG3Y), 11 ± 1 nM (TERRA/TLSRGG3), and 10 ± 1 nM (Htelo/TLSRGG3)] (Figure 2B,D, Figure S5B). The obtained K_d's for TLSRGG3 binding to TERRA and that to Htelo were in good agreement with the K_d's of 15 ± 2 nM (TERRA/TLSRGG3) and 10 ± 2 nM (Htelo/TLSRGG3) obtained from the EMSA, respectively.[18] Binding stoichiometries were determined by ITC as one TLSRGG3Y per two TERRA, one TLSRGG3 per one TERRA, and one TLSRGG3 per one Htelo. Therefore, in addition to reducing Htelo binding, Tyr

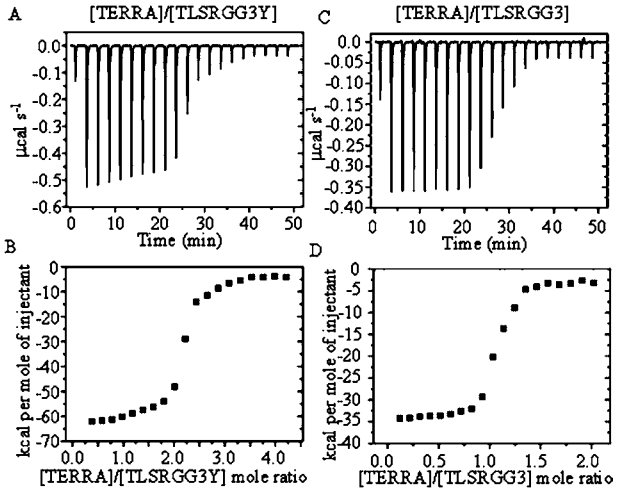

Figure 2. Binding affinities and stoichiometries of TLSRGG3 or TLSRGG3Y and TERRA measured by isothermal titration calorimetry (ITC). Raw calorimetric data for titration of the TLSRGG3Y solution (A) or TLSRGG3 (C) with serial injections of TERRA-containing solution. Binding isotherms (B, D) resulting from the integration of raw calorimetric data corrected for the dilution heat of TERRA.

substitution increased the TERRA-binding capacity of TLSRGG3. Besides the region containing the original three Tyr in TLSRGG3, the region containing two Tyr in TLSRGG3Y may be able to bind one molecule of TERRA.

Native TERRA in a cell is thought to contain a multi-G-quadruplex structure in vivo.[5,6,23] The calorimetric titrations and binding isotherms of TLSRGG3Y with r(UUAGGG)$_8$ or r(UUAGGG)$_{12}$ are shown in Figure S5C,D. The binding isotherm of TLSRGG3Y with r(UUAGGG)$_8$ and r-(UUAGGG)$_{12}$ forming a multi-G-quadruplex revealed exothermic binding with a calculated K_d of 8.8 ± 0.9 nM (r(UUAGGG)$_8$/TLSRGG3Y), and 9.0 ± 0.8 nM (r-(UUAGGG)$_{12}$/TLSRGG3Y), respectively (Figure S5D, F). Binding stoichiometries were determined to be 0.7 (r-(UUAGGG)$_8$/TLSRGG3Y) and 0.4 (r(UUAGGG)$_{12}$/TLSRGG3Y) by ITC, indicating that multiples TLSRGG3Y might be able to bind to native long TERRA.

The crystal structures of a complex of G-quadruplex DNA and G-quadruplex-binding compounds suggest that at least two different types of interactions contribute to G-quadruplex DNA binding of the G-quadruplex-binding compounds: G-quartet stacking and G-quartet-independent interactions, such as loop binding.[24] Because EWSRGG3 preferentially binds to G-quadruplex DNA with one longer loop,[20] we investigated the effect of the loop length of G-quadruplex RNA on TLSRGG3Y binding by EMSA (Figure 3). EMSA showed that if any of the loops bears rU3, TLSRGG3Y binding to the G-quadruplex was comparable to its binding to TERRA (rL333, rL131, rL113, rL311). Although TLSRGG3Y binds to those with a single longer loop (rL131, rL113, rL311), it seems to be lesser affinity than to those of rL333 and TLSRGG3Y. In addition, as the number of residues decreases, the binding affinity decreases (rL111, rL121, rL131). These findings suggest preferential binding of TLSRGG3Y to G-quadruplex RNA with a longer loop, independent of its position.

To examine whether TLSRGG3Y recognizes the loops and/or the G-tetrad of RNA, we examined TLSRGG3Y binding to G-quadruplexes comprising four d(GGG) repeats and three r(UUA) loops (rL9dG4), four r(GGG) repeats and three

dx.doi.org/10.1021/ja4086929 | J. Am. Chem. Soc. 2013, 135, 18016–18019

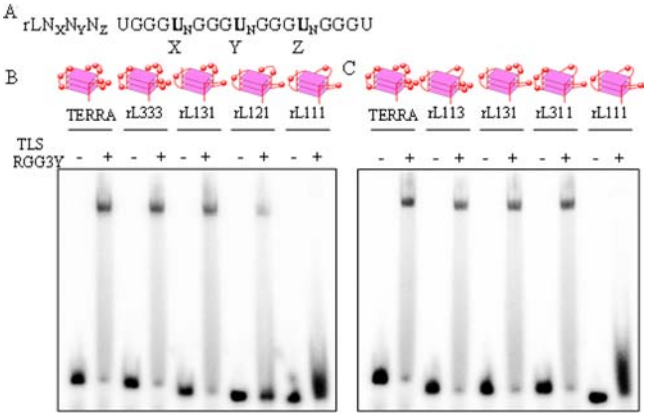

Figure 3. Effect of G-quadruplex RNA loop length on G-quadruplex-binding affinity of TLSRGG3Y. (A) The RNA sequences used are shown. Bold U shows loop regions: N shows number of loop residues. (B, C) EMSA was performed using TLSRGG3Y with ^{32}P-labeled (B) TERRA, rL333, rL131, rL121, or rL111 and (C) TERRA, rL113, rL131, rL311, or rL111.

d(TTA) loops (dL9rG4), or four r(GGG) repeats (rG4) by EMSA (Table S1, Figure 4). EMSA of TLSRGG3Y and these G-quadruplex structures showed that G-quadruplexes of rL9dG4 are favorable for binding, like TERRA, while G-quadruplexes of dL9rG4 and rG4 (Figure 4A) were unfavorable. These findings indicate preferential binding of TLSRGG3Y to G-quadruplexes with RNA loops, and not a guanine tetrad.

To identify which part of the nucleotides on the loop is recognized by TLSRGG3Y, we investigated the role of the base and ribose on the loop in the G-quadruplex for TLSRGG3Y binding to G-quadruplexes comprising four d(GGG) repeats with several DNA/RNA loops (Figure 4B,C). The EMSA showed that G-quadruplexes comprising three RNA abasic loops (R9dG4) were favorable for binding, while G-quadruplexes with three DNA abasic loops (D9dG4) were unfavorable (Table S1, Figure 4B). G-quadruplexes comprising two DNA loops and a single RNA loop (rL3dG4), or a single RNA abasic loop (R3dG4), were also favorable for binding. Substitution of the RNA abasic loop in R3dG4 with a DNA loop abasic loop (D3dG4) eliminated binding favorability (Table S1, Figure 4C). These findings suggest that TLSRGG3Y distinguishes between one RNA loop and one DNA loop in the G-quadruplex. Moreover, we found that TERRA binding of TLSRGG3Y is reduced by the addition of excessive G-quadruplex with three RNA loops (R9dG4 and rL9dG4), but not by those with three DNA loops (dL9rG4 and D9dG4). Reduction was also observed by G-quadruplexes with a single RNA loop (rL3dG4 and R3dG4), but to a lesser extent (Figure S6A−G). Each K_d of G-quadruplexes with three RNA loops (R9dG4 and rL9dG4) with TLSRGG3Y was estimated to range from 10 to 100 nM, whereas those with a single RNA loop (rL3dG4 and R3dG4) were estimated to range from 100 to 1000 nM. These results indicate that TLSRGG3Y has a higher affinity for G-quadruplexes with three RNA loops than for those with one RNA loop. On the other hand, those with three DNA loops (dL9rG4 and D9dG4) were estimated to have a K_d greater than 1000 nM.

The fact that TLSRGG3Y recognizes abasic RNA loops suggests that 2′-OH in the loop is responsible for the recognition. To investigate the role of 2′-OH in the loop in

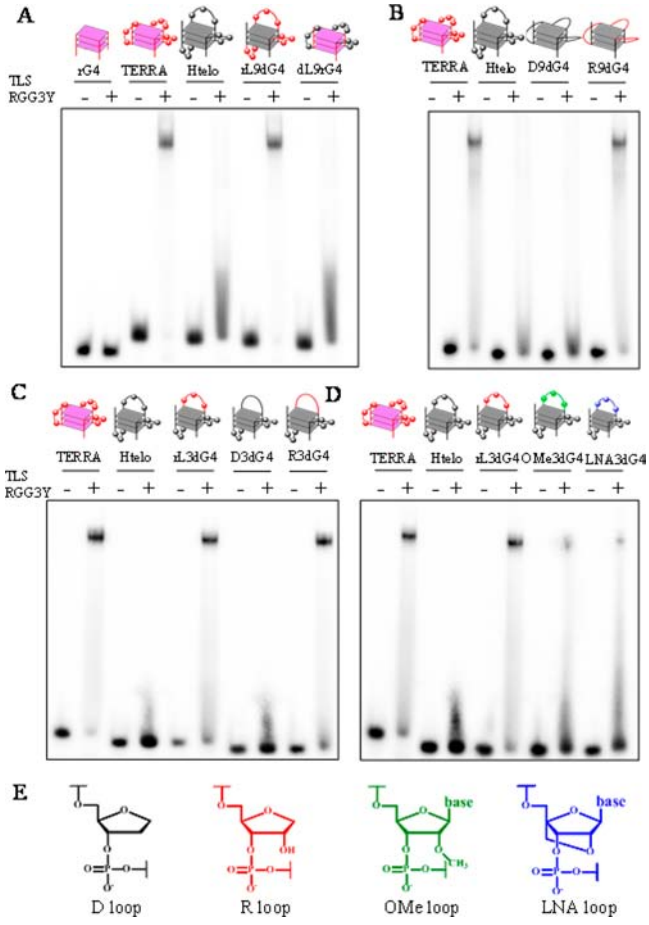

Figure 4. Effect of DNA or RNA loops on the G-quadruplex binding selectivity of TLSRGG3Y. (A−D) EMSA was performed with TLSRGG3Y and ^{32}P-labeled (A) rG4, TERRA, Htelo, rL9dG4, or dL9rG4; (B) TERRA, Htelo, D9dG4, and R9dG4; (C) TERRA, Htelo, rL3dG4, D3dG4, and R3dG4; and (D) TERRA, Htelo, rL3dG4, OMe3dG4, and LNA3dG4. (E) The nucleic acid structures of abasic DNA, abasic RNA, 2′-*O*-methylribonucleotide, and Locked Nucleic Acid are indicated. Red, gray, green, and blue in cartoon show, respectively, RNA, DNA, 2′-*O*-methylribonucleotides, and Locked Nucleic Acids.

TLSRGG3Y binding to G-quadruplex, we used an EMSA to examine TLSRGG3Y binding to G-quadruplexes with the loop containing an artificial 2′-modified nucleic acid loop, Locked Nucleic Acids (LNA3dG4), and 2′-*O*-methylribonucleotides (OMe3dG4), instead of an RNA loop of rL3dG4 (Table S1, Figure 4D).[25] Our findings indicate that LNA3dG4 and OMe3dG4 were weakly bound compared to rL3dG4. On the basis of the competitive assays, the K_d of LNA3dG4 and OMe3dG4 was estimated to be greater than 1000 nM (Figure S6H,I). These findings indicate the preferential recognition of the 2′-OH of the loop in the G-quadruplex by TLSRGG3Y. To our knowledge, this is the first report that G-quadruplex RNA-binding molecules recognize the 2′-OH of the riboses of loops.

In conclusion, TLSRGG3Y binds preferentially to G-quadruplex TERRA compared with telomere DNA. Moreover, TLSRGG3Y recognizes the ribose of the loop in the G-quadruplex RNA. Substitution of Phe with Tyr in human protection of telomeres 1 shifts the binding specificity of protection of telomeres 1 from DNA to RNA.[26] The hydrophilic environment created by Tyr may allow for ribonucleotide binding. The Tyr in the mutated POT1 may

dx.doi.org/10.1021/ja4086929 | J. Am. Chem. Soc. 2013, 135, 18016−18019

recognize 2′-OH of the riboses of loops in G-quadruplex like TLSRGG3Y. These findings might be useful toward the development of molecules that specifically bind to TERRA. Because the protein can be easily expressed in cells by transfection of its plasmid DNA, G-quadruplex RNA binding protein will be useful for investigating telomere maintenance by TERRA.

■ ASSOCIATED CONTENT

Supporting Information

Sequence of oligonucleotides used in EMSA and CD spectroscopy; CD spectra of rL333, rL311, rL131, rL113, rL121, rL111, rG4, rL9dG4, dL9rG4, R9dG4, rL3dG4, R3dG4, OMe3dG4, and LNA3dG4; competitive binding of EWSRGG3 to Htelo and TERRA; TERRA-binding specificities of TLSRGG3Y; CD spectra of TERRA in the presence of TLSRGG3Y; binding affinities and stoichiometries of TLSRGG3 or TLSRGG3Y and DNA or RNA measured by ITC; competitive binding of TLSRGG3Y to TERRA and various oligonucleotides. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

stohyos@ipc.shizuoka.ac.jp

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Ayumu Yamamoto at Shizuoka University for his valuable advice. This research was supported by a JGC-S Scholarship Foundation and a Grant-in-Aid for JSPS Fellows (No. 243925 to K.T.) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

■ REFERENCES

(1) de Lange, T. S. *Genes Dev.* **2005**, *19*, 2100−2110.
(2) Luke, B.; Lingner, J. *EMBO J.* **2009**, *28*, 2503−2510.
(3) Azzalin, C. M.; Reichenbach, P.; Khoriauli, L.; Giulotto, E.; Lingner, J. *Science* **2007**, *318*, 798−801.
(4) Luu, K. N.; Phan, A.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963−9970.
(5) Martadinate, H.; Phan, A. T. *J. Am. Chem. Soc.* **2009**, *131*, 2570−2578.
(6) Xu, Y.; Suzuki, Y.; Ito, K.; Komiyama, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 14579−14584.
(7) Schoeftner, S.; Blasco, M. A. *Nat. Cell Biol.* **2008**, *10*, 228−236.
(8) Deng, Z.; Norseen, J.; Wiedmer, A.; Riethman, H.; Lieberman, P. M. *Mol. Cell* **2009**, *35*, 403−413.
(9) Arnoult, N.; Van Beneden, A.; Decottignies, A. *Nat. Struct. Mol. Biol.* **2012**, *19*, 948−956.
(10) Blasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat. Rev. Drug Discovery* **2011**, *10*, 261−275.
(11) Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. *Nat. Chem. Biol.* **2012**, *8*, 301−310.
(12) Neidle, S. *FEBS J.* **2010**, *277*, 1118−1125.
(13) Bugaut, A.; Balasubramanian, S. *Nucleic Acids Res.* **2012**, *40*, 4727−4741.
(14) Collie, G. W.; Sparapani, S.; Parkinson, G. N.; Neidle, S. *J. Am. Chem. Soc.* **2011**, *133*, 2721−2728.
(15) Rzuczek, S. G.; Pilch, D. S.; Liu, A.; Liu, L.; LaVoie, E. J.; Rice, J. E. *J. Med. Chem.* **2010**, *53*, 3632−3544.
(16) Collie, G.; Reszka, A. P.; Haider, S. M.; Gabelica, V.; Parkinson, G. N.; Neidle, S. *Chem. Commun.* **2009**, *48*, 7482−7484.
(17) Di Antonio, M.; Biffi, G.; Mariani, A.; Raiber, E. A.; Rodriguez, R.; Balasubramanian, S. *Angew. Chem., Int. Ed.* **2012**, *51*, 11073−11078.
(18) Takahama, K.; Takada, A.; Tada, S.; Shimizu, M.; Sayama, K.; Kurokawa, R.; Oyoshi, T. *Chem. Biol.* **2013**, *20*, 341−350.
(19) Takahama, K.; Kino, K.; Arai, S.; Kurokawa, R.; Oyoshi, T. *FEBS J.* **2011**, *278*, 988−998.
(20) Takahama, K.; Sugimoto, C.; Arai, S.; Kurokawa, R.; Oyoshi, T. *Biochemistry* **2011**, *50*, 5369−5378.
(21) Haq, I.; Trent, J. O.; Chowdhry, B. Z.; Jenkins, T. C. *J. Am. Chem. Soc.* **1999**, *121*, 1768−1779.
(22) Pagano, B.; Martino, L.; Randozzo, A.; Giancola, C. *Biophys. J.* **2008**, *94*, 562−569.
(23) Martadinate, H.; Heddi, B.; Lim, K. W.; Phan, A. T. *Biochemistry* **2011**, *50*, 6455−6461.
(24) Collie, G. W.; Parkinson, G. N. *Chem. Soc. Rev.* **2011**, *40*, 5867−5892.
(25) Cuzic-Feltens, S.; Weber, M. H. W.; Hartman, R. K. *Nucleic Acids Res.* **2009**, *37*, 7638−7653.
(26) Nandakumar, J.; Pdell, E. R.; Cech, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 651−656.