PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

PCCP

RSCPublishing

COMMUNICATION

Loop nucleotides impact stability of intrastrand i-motif structures at neutral pH

Cite this: DOI: 10.1039/x0xx00000x

Taiga Fujii[a] and Naoki Sugimoto[a,b]*

Received 00th January 2015,
Accepted 00th January 2015

DOI: 10.1039/x0xx00000x

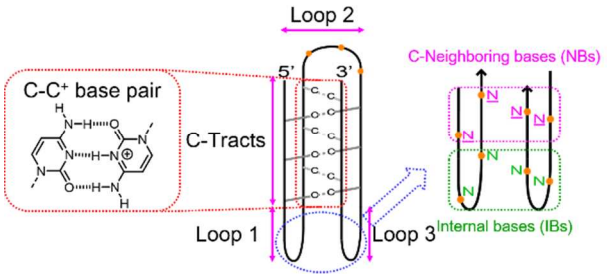
www.rsc.org/

The stability of i-motif structures at neutral pH is of interest due to the potential of these structures to impact gene expression. A systematic investigation of loop sequence and length revealed that certain loop nucleobases stabilize i-motif quadruplexes.

DNA and RNA sequences with the potential to form quadruplex structures are found in the human genome and transcriptome and may control transcription and translation.[1] There are two categories of quadruplexes: the G-quadruplex and the i-motif. The G-quadruplex is a four-stranded structure stabilized by stacked guanine tetrads.[2] Potential G-quadruplex forming sequences exist in promoter regions of various oncogenes such as *c-MYC*,[3] *BCL-2*,[4] *RET*,[5] *VEGF*,[6] *PDGF-A*,[7] *HIF-1α*,[8] and *hTERT*.[9] Experiments have shown that the formation of G-quadruplex structures in these promoter regions inhibit transcription, and slippage, halt, or arrest of transcription can be induced by their formation in open reading frames of genes.[1]

The i-motif is stabilized by hemiprotonated cytosine-cytosine ($C\text{-}C^+$) base pairs (Scheme 1)[10]. DNAs capable of i-motif formation are generally paired with a complementary strand with G-quadruplex forming potential.[11] Although G-quadruplexes are stable under physiological conditions, most i-motif structures are stable only at acidic pH (4.5-6.0) due to requirement for protonation of the cytosine building blocks. However, negative supercoil stress and molecular crowding conditions, which mimic the human cell environment, facilitate formation of i-motif structures at the near neutral pH.[3,12]

The factors that affect i-motif stability are pH, the number of $C\text{-}C^+$ base pairs, and the loop sequences and lengths. Sequences that form i-motif structures typically have four runs of two to four cytosines (Cs).[2] In logic gates and pH responsive elements, the number of the $C\text{-}C^+$ base pairs is the most significant factor.[13] Based on past investigations of the effects of loop length and sequence, i-motif forming sequences have been classified into two types.[14] Class I structures have short loop regions and include those formed within *VEGF* (transitional mid-point pH from a helix to coil: pH 5.8,)[6], *RET* (pH 6.4)[5], *Rb* genes (pH 5.9),[15] and in human telomere sequences (pH 6.8)[16]. Class II structures have longer loop regions and are found in *c-MYC* (pH 6.6),[3] *BCL-2* (pH 6.6),[4] *PDGF-A* (pH >7.0),[17] *HIF-1α* (pH 7.2),[17] and *hTERT* (pH >7.0)[17].



Scheme 1 Chemical structure of the $C\text{-}C^+$ base pair and schematic illustration of an intrastrand i-motif. There are three loop regions. We refer to nucleotides in the first and third loops that neighbor the $C\text{-}C^+$ base pairs as neighboring bases (NBs; purple box) and the other nucleotides as internal bases (IBs; green box).

Previously published results support the hypothesis that Class II i-motifs are more stable than Class I motifs due to the presence of stabilizing influences within the longer loop regions of the Class II motifs. However, two Class I motifs also have transitional mid-point pH values close to pH 7.0.[5,16] Recent reports have shown that the class II motifs with long T loops do not form stable i-motif structures.[18] Thus, the loop sequences rather than the loop length likely impact i-motif stability.

In this study, we systematically evaluated the effects of loop sequence and length on i-motif stability at neutral pH.

Intrastrand i-motifs are composed of cytidine tracts (C-tracts) and three loops. We refer to loop nucleobases that directly neighbor the C tracts as neighboring bases (NBs). We evaluated sequences of the form 5'-$(C)_l$-$(N)_m(T)_n(N)_m$-CCC-$(T)_3$-$(C)_l$-$(N)_m(T)_n(N)_m$-CCC-3' where N is any base and l = 3 or 4 nucleotides, m = 1-3, and n = 2-6 (Scheme 1, Fig. 1a and Scheme S1). We evaluated A, G, and T as NBs in loops 1 and 3 (loops numbered from the 5' end); cytosine residues were not placed at these positions to ensure that C-tracts did not shift in register. We refer to the bases located within the loops non-adjacent to the C runs as internal bases (IBs); these were thymine residues in all sequences evaluated. There were 4, 6, or 8 bases in the first and third loop regions, and, to limit conformational heterogeneity, the second loop contained three Ts, which stabilize the quadruplexes more than three As in second loop.[13,17] No nucleotides flanked the C-tracts at the oligonucleotide termini in any sequence.

a)

ODN1: 5'-CCC-TTTT-CCC-$(T)_3$-CCC-TTTT-CCC-3'
ODN2: 5'-CCC-GTTG-CCC-$(T)_3$-CCC-GTTG-CCC-3'
ODN3: 5'-CCC-GTTG-CCC-$(T)_3$-CCC-TTTT-CCC-3'
ODN4: 5'-CCC-TTTT-CCC-$(T)_3$-CCC-ATTA-CCC-3'
ODN5: 5'-CCC-ATTA-CCC-$(T)_3$-CCC-ATTA-CCC-3'
ODN6: 5'-CCC-GTTG-CCC-$(T)_3$-CCC-ATTA-CCC-3'
ODN7: 5'-CCCC-GTTG-CCC-$(T)_3$-CCCC-TTTT-CCC-3'
ODN8: 5'-CCCC-GTTI-CCC-$(T)_3$-CCCC-TTTT-CCC-3'
ODN9: 5'-CCC-GTTG-CCC-$(T)_3$-CCC-GTTTTG-CCC-3'
ODN10: 5'-CCC-TTTT-CCC-$(T)_3$-CCC-ATTTTA-CCC-3'
ODN11: 5'-CCC-GTTG-CCC-$(T)_3$-CCC-ATTTTA-CCC-3'

b)

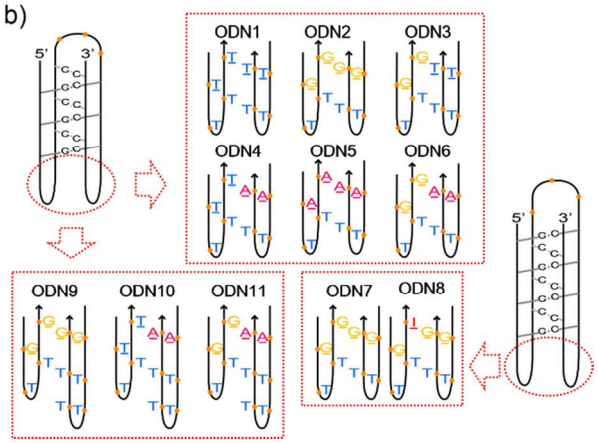


Fig. 1 (a) The eleven intrastrand i-motif-forming sequences that were used in this study. The NBs in the sequence are underlined. (b) Schematics of first and third loop regions of sequences analyzed. The NBs are underlined in the illustration.

The identities of the nucleobases in loops 1 and 3, the bases that are adjacent to the C-tracts, had a significant effect on the thermal stability of intramolecular structures formed by the various i-motif sequences at pH 6.5. Each of the six sequences with loops of four nucleotides (Fig. 1b) had clear melting transitions when the absorbance at 295 nm was monitored as a function of temperature (Fig. S1), indicating that these oligonucleotides (ODNs) formed i-motif structures at pH 6.5. The mid-points of the melting transitions (T_ms) and thermodynamic parameters ($\Delta H°$, $\Delta S°$, and $\Delta G°_{25}$) were obtained by fitting of melting curves obtained at three different DNA concentrations (Table 1). Based on the values of T_m and $\Delta G°_{25}$, these i-motifs were classified into groups of high (**ODN1** and **ODN2**), medium (**ODN3** and **ODN4**), and low stabilities (**ODN6** and **ODN5**). In general, quadruplex stabilities, like duplex stabilities, result from contributions of stacking and hydrogen bonding interactions among the bases.[19] Stacking of purine bases is more favorable than that of pyrimidine bases. The quadruplex stabilities were inversely proportional to the number of the purine bases in NBs with the exception of **ODN2**, suggesting the hydrogen bonding among NBs had much stronger effect on the stability of the i-motif structures than did stacking of the NBs. In fact, $-\Delta\Delta G°_{25}$ values between **ODN6** and **ODN2** and between **ODN6** and **ODN1**, which is stabilizing energy due to the NB difference, are large at 1.83 and 1.88 kcal·mol^{-1}, respectively. Stacking interactions should be similar in **ODN2** and **ODN6**, but the structure formed by **ODN2** was significantly more stable, suggesting that there may be hydrogen bonds among the four Gs.

Table 1 Melting temperatures and thermodynamic parameters for the formation of i-motif quadruplexes with different NBs at pH 6.5[a]

Name	NB[b] [1/3]	$\Delta H°$ [kcal·mol^{-1}]	$T\Delta S°$[c] [kcal·mol^{-1}]	$\Delta G°_{25}$[c] [kcal·mol^{-1}]	$-\Delta\Delta G°_{25}$[c,d] [kcal·mol^{-1}]	T_m [°C]	ΔT_m[d] [°C]
ODN1	T/T	-82.1±1.6	-80.6±1.4	-1.56±0.38	+1.88	30.7±1.2	+7.2
ODN2	G/G	-73.2±0.8	-71.7±0.9	-1.51±0.28	+1.83	31.3±1.2	+7.8
ODN3	G/T	-74.3±2.9	-73.5±2.6	-0.75±0.32	+1.07	28.0±1.2	+4.5
ODN4	T/A	-72.8±1.1	-72.0±0.9	-0.73±0.30	+1.05	28.0±1.2	+4.5
ODN5	A/A	-66.6±0.8	-66.9±1.0	0.26±0.25	+0.06	23.8±1.0	+0.3
ODN6	G/A	-62.5±0.3	-62.8±0.4	0.32±0.17	-	23.5±0.8	-

[a] Buffer contained 10 mM Na_2HPO_4 (pH 6.5) and 0.1 mM Na_2EDTA. The UV melting and annealing curves were measured with three DNA concentrations: 1, 5, and 50 μM. The melting temperatures and thermodynamic parameters of quadruplexes were evaluated from average values obtained from curve fitting of data. [b] NB [1/3] represents the bases of NBs in the loop 1 and 3, respectively. [c] $T\Delta S°$ and $\Delta G°_{25}$ were calculated at 298 K (25 ºC). [d] $-\Delta\Delta G°_{25}$ and ΔT_m were calculated relative to values for **ODN6**.

Table 2 Melting temperatures and thermodynamic parameters for the formation of i-motif quadruplexes with the different loop lengths at pH 6.5[a]

Name	Loop lengths[b]	$\Delta H°$ [kcal·mol^{-1}]	$T\Delta S°$[a] [kcal·mol^{-1}]	$\Delta G°_{25}$[a] [kcal·mol^{-1}]	$-\Delta\Delta G°_{25}$[a,c] [kcal·mol^{-1}]	T_m [°C]	ΔT_m[c] [°C]
ODN9	4/6	-72.2±4.2	-71.8±4.1	-0.38±0.07	-1.13	26.6±0.2	-4.7
ODN2	4/4	-73.2±0.8	-71.7±0.9	-1.51±0.28	-	31.3±1.2	-
ODN11	4/6	-67.1±4.7	-67.1±4.7	0.04±0.04	+0.28	25.0±0.2	+1.5
ODN6	4/4	-62.5±0.3	-62.8±0.4	0.32±0.17	-	23.5±0.8	-

[a] Experimental conditions and calculations were the same as those in Table 1. [b] The number of bases in loops 1 and 3, respectively. [c] $-\Delta\Delta G°_{25}$ and ΔT_m of **ODN9** and **ODN11** were calculated relative to those of **ODN2** and **ODN6**, respectively.

To evaluate the importance of hydrogen bonding connectivity, we compared the stability of **ODN7**, in which NBs are Gs, with that of **ODN8**, which has three Gs and one inosine (Fig. 1b). **ODN7** and **ODN8** have also seven C-C$^+$

base pairs to limit the superposition patterns of the C-C$^+$ base pairs while those of **ODN2** is six. Due to lack of an amino group in inosine, this base lacks some of the hydrogen bonding capability of G. The one-base difference between **ODN7** and **ODN8** did not significantly alter the i-motif stability suggesting that the four Gs do not form a G-quartet but that the two Gs base pair (Table S1 and Fig. S2). On the other hand, increment in the number of C-C$^+$ base pairs from **ODN2**, which had the hydrogen bonding of loop bases, largely increased the i-motif stability though the amount of the stabilization of **ODN7** was larger than that of the i-motif without hydrogen bonding in NBs such as **ODN12** (Fig. S2 and Table S1). Thus, these results demonstrate that hydrogen bonding between Gs in loops stabilizes the i-motif structures.

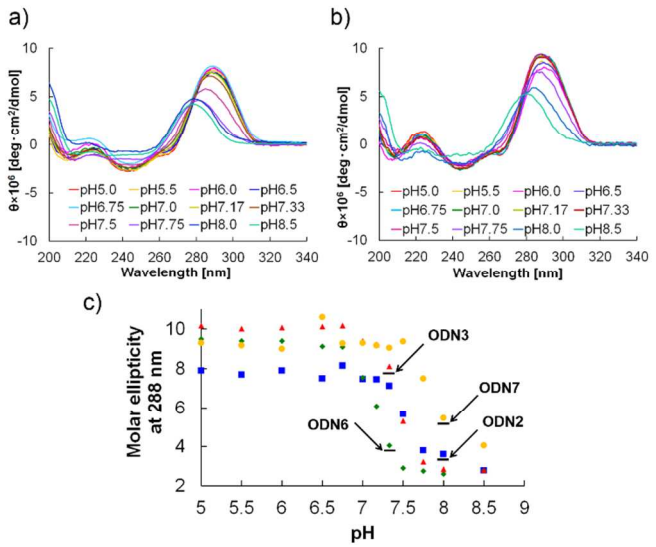


Fig. 2 (a) The CD spectra of 20 μM **ODN2** as a function of pH at 4 °C. (b) The CD spectra of **ODN7** as a function of pH at 4 °C. (c) Plots of pH vs. molar ellipticity at 288 nm of **ODN2** (blue ■), **ODN3** (red ▲), **ODN6** (green ♦), and **ODN7** (yellow ●).The experiments were carried out in the buffer containing 10 mM Na_2HPO_4 (pH 5.0-8.5) and 0.1 mM Na_2EDTA.

We next evaluated thermodynamic stabilities of the i-motif structures adopted by ODNs with different loop lengths (Fig. 1b and Scheme S1). Figures S3 and S4 show representative melting curves. An increase in the length of only loop 3 from four to six nucleotides changed the quadruplex stability in a manner that depended on the identities of the NB bases (Table 2 and S1). **ODN2** with four nucleotides in loop 3 had a T_m of 31.3 °C, whereas **ODN9** had a T_m of 26.6 °C. The same tendency was observed for **ODN1** and **ODN12** (see Table S1). These destabilizations likely derive from disruption of the ability of NBs to hydrogen bonds. On the other hand, the elongation of loop 3 in **ODN4** and **ODN6** that do not have NBs able to hydrogen bond slightly stabilized the i-motif structures (Table 2). In addition, further increase in the lengths of both loops 1 and 3 decreased in T_m values (Table S2) in agreement with previous reports.[18]

To explore the effect of pH on the conformations of the ODNs, the circular dichroism (CD) spectra were measured. All the sequences showed a positive Cotton effect at around 275-310 nm at pHs 6.5 (Figs. 2a, 2b, S5, S6, S7 and S8). This spectral feature is characteristic of the i-motif; therefore, the ODNs formed i-motif structures at pH values of 6.5 and lower.[6] The pH vs. molar ellipticity at 288 nm curves are shown in Figures 2c and S9. The transitional mid-point pHs of **ODN7**, **ODN2**, **ODN3** and **ODN6** were 7.9, 7.6, 7.4, and 7.1 respectively. The CD spectra evaluated as a function of temperature from 0-80 °C had isodichroic points, indicating that the melting curves in our experiment derived from two-state transitions from i-motif to coils (Fig. S5).[20] The mid-points of the pH vs. molar ellipticity curves were proportional to the T_m values of the i-motif structures. These data suggest interaction of loop nucleobases raised the pK_a of the N3 position of C.

It has been reported that long looping regions are necessary for i-motif formation at the near neutral pH.[13] The ODNs designed to have four, three, and four bases in first, second, and third loops, respectively, are classified with the Class I structures and were predicted to have low stabilities and low transitional mid-point pHs. However, we observed higher T_m values and higher transitional mid-point pHs than those previously reported for Class I i-motif structure formed by sequences from *VEGF* and *Rd* promoter regions.[6,15] This contradicts the hypothesis that the Class I i-motifs have low stability.[14] In our experiments elongation of the i-motif loop regions destabilized the structures, likely due to an increase in flexibility.[18]

Interactions between loop nucleobases in i-motif structures appear to significantly impact stability. We propose that the both sequence and length of the loop regions of the i-motif impact stability at near neutral pH in predictable ways: 1) Gs or Ts in the loop positions adjacent to C-rich regions in loops 1 and 3 stabilize the quadruplex structure due to inter-loop hydrogen bonding. 2) The same number of the bases in first and third loops are optimal for quadruplex stabilization. 3) An additional C-C$^+$ base pair also cooperatively stabilizes the quadruplex presumably by optimizing hydrogen bonding between the NBs. Data from our systematic study explains the relatively high transitional mid-point pH (6.6) of the i-motif in the *Bcl-2* promoter region, which has long loops. The *Bcl-2* i-motif sequence has seven C-C$^+$ base pairs in the C-tracts and eight and seven bases in first and third loops, respectively. The NBs in the first and third loops are Gs at three positions. The sequence satisfies the first and third conditions above for the i-motif stability. Although the previous studies revealed that the stacking interactions among bases in the second loop in the *Bcl-2* i-motif are important for the stability[4], our results suggest that stabilizing interaction in the first and third loops also make large contributions to i-motif stability.

In conclusion, our data indicate that sequences and lengths in the first and third loop regions predictably contribute to i-motif stability at the neutral pH. The identities of nucleobases located adjacent to the C-rich regions are the main drivers of

the stability contributed by bases in the first and third loops. These bases appear to be capable of hydrogen bonding interactions as the i-motifs with two Gs or Ts at these positions have high thermodynamic stabilities and high transitional mid-point pHs. The nucleotides located at internal positions in the first and third loops are important for maintaining distances that allow hydrogen bonding networks to form in the loop regions. In fact, two base differences between first and third loops destabilized the quadruplex due to disarrangement of hydrogen bonding between NBs. Data from our systematic investigation can be utilized to predict whether particular genomic DNA regions are likely to adopt an i-motif structure of high stability and will guide design of DNA nano-materials.

Acknowledgements

This work was partly supported by the Grants-in-Aid for Scientific Research, by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) through the Program for the Strategic Research Foundation at Private Universities, Japan, (2014-2019), and Hirao Taro Foundation of the Konan University Association for Academic Research.

Notes and references

[a] Frontier Institute for Biomolecular Engineering Research (FIBER), Konan University, 8–9–1 Minatojima-minamimachi, Chuo-ku, Kobe 650–0047 (Japan); E-mail: sugimoto@konan-u.ac.jp
[b] Graduate School of Frontiers of Innovative Research in Science and Technology (FIRST), Konan University, 8–9–1 Minatojima-minamimachi, Chuo-ku, Kobe 650–0047 (Japan)

†Electronic Supplementary Information (ESI) available: [Experimental section, UV melting curves at pH 6.5, CD spectra at pH 5.0-8.5 and the plot of pH vs. molar ellipticity at 288 nm]. See DOI: 10.1039/c000000x/

1 H. Tateishi-Karimata, N. Isono, and N. Sugimoto, *PLoS ONE*, **2014**, *9*, e9058.; T. Endoh, Y. Kawasaki and N. Sugimoto, *Nucleic Acids Res.*, **2013**, *41*, 6222.; T. Endoh, Y. Kawasaki and N. Sugimoto, *Angew. Chem. Int. Ed.*, **2013**, *52*, 5522.

2 S. A. Ohnmacht, E. Varavipour, R. Nanjunda, I. Pazitna, G. D. Vita, M. Gunaratnam, A. Kumar, M. A. Ismail, D. W. Boykin, W. D. Wilson and S. Neidle, *Chem. Commun.*, **2014**, *50*, 960.; A. Marchand, R. Ferreira, H. Tateishi-Karimata, D. Miyoshi, N. Sugimoto and V. Gabelica, *J. Phys. Chem. B*, **2013**, *117*, 12391.; P. A. Rachwal, I. S. Findlow, J. M. Werner, T. Brown and K. R. Fox, *Nucleic Acid Res.*, **2007**, *35*, 4212.; P. A. Rachwal, T. Brown and K. R. Fox, *FEBS Lett.*, **2007**, *581*, 1657.; J. R. Williamson, M. K. Raghuraman, T. R. Cech, *Cell*, **1989**, *59*, 871.

3 J. M. Dettler, R.Buscaglia, J.J. Cui, D. Cashman, M Blynn and E. A. Lewis, *Biophys. J.*, **2010**, *99*, 561.; D. Sun and L. H. Hurley, *J. Med. Chem.*, **2009**, *52*, 2863.; T. Simonsson, M. Pribylova and M. Vorlickova, *Biochem. Biophys. Res. Commun.*, **2000**, *278*, 158.

4 Y. Cui, D. Koirala, H. Kang, S. Dhakal, P. Yangyuoru, L. H. Hurley and H. Mao, *Nucleic Acid Res.*, **2014**, *42*, 5755.; H.-J. Kang, S. Kendrick, S. M. Hecht and L. H. Hurley, *J. Am. Chem. Soc.*, **2014**, *136*, 4172.; S. Kendrick, H.-J. Kang, M. P. Alam, M. M. Madathil, P. Agrawal, V. Gokhale, D. Yang, S. M. Hecht and L. H. Hurley, *J. Am. Chem. Soc.*, **2014**, *136*, 4161.; S. Kendrick, Y. Akiyama, S. M. Hecht and L. H. Hurley, *J. Am. Chem. Soc.*, **2009**, *131*, 17667.; T. S. Dexheimer, D. Sun and L. H. Hurley, *J. Am. Chem. Soc.*, **2006**, *128*, 5404.

5 K. Guo, A. Pourpak, K. Beetz-Rogers, V. Gokhale, D. Sun and L. H. Hurley, *J. Am. Chem. Soc.*, **2007**, *129*, 10220.

6 D. Sun, K. Guo and Y.J. Shin, *Nucleic Acid Res.*, **2011**, *39*, 1256.; K. Guo, V. Gokhale, L. H. Hurey and D. Sun, *Nucleic Acid Res.*, **2008**, *36*, 4598.

7 Y. Qin, E. M. Rezler, V. Gokhale, D. Sun and L. H. Hurey, *Nucleic Acid Res.*, **2007**, *35*, 7698.

8 R. D. Armond, S. Wood, D. Sun, L. H. Hurey and S. W. Ebbinghaus, *Biochemistry*, **2005**, *44*, 16341.

9 S. L. Palumbo, S. W. Ebbinghaus and L. H. Hurley, *J. Am. Chem. Soc.*, **2009**, *131*, 10878.

10 Y.-W. Chen, C.-R. Jhan, S. Neidle, and M.-H. Hou, *Angew. Chem. Int. Ed.*, **2014**, *53*, 10682. ; K. Gehring, J.-L. Leroy and M. Guéron, *Nature*, **1993**, *363*, 561

11 S. Benabou, A. Aviño, R Eritja, C. González and R. Gargallo, *RSC Adv.*, **2014**, *4*, 26956.

12 A. Rajendran, S. Nakano and N. Sugimoto, *Chem. Commun.*, **2010**, *46*, 1299.

13 I. V. Nesterova and E. E. Nesterov, *J. Am. Chem. Soc.*, **2014**, *136*, 8843.; T. Li and M. Famulok, *J. Am. Chem. Soc.*, **2013**, *135*, 1593.

14 T. A. Brooks, S. Kendrick and L. Hurley, *FEBS J.*, **2010**, *277*, 3459.

15 J. W. Park, Y. J. Seo and B. H. Kim, *Chem. Commun.*, **2014**, *50*, 52.; Y. Xu and H. Sugiyama, *Nucleic Acid Res.*, **2006**, *34*, 949.

16 A. L. Lieblein, B.Fürtig and H. Schwalbe, *ChemBioChem*, **2013**, *14*, 1226.; I.J. Lee, J. W. Yi and B. H. Kim, *Chem. Commun.*, **2009**, *45*, 5383.

17 J. A. Brazier, A. Shah and G. D. Brown, *Chem. Commun.*, **2012**, *48*, 10739.

18 S. P. Gurung, C. Schwarz, J. P. Hall, C. J. Cardin and J. A. Brazier, *Chem. Commun.*, **2015**, *51*, 5630.; S. M. Reilly, R. K Morgan, T. A. Brooks, and R. M. Wadkins, *Biochemistry*, **2015**, *54*, 1364.

19 D. H. Turner, N. Sugimoto, R. Kierzek and S. D. Dreiker, *J. Am. Chem. Soc.*, **1987**, *109*, .3783.

20 S. Dhakal, J. L. Lafontaine, Z. Yu, D. Koirala and H. Mao, *PLoS ONE*, **2012**, *7*, e39271.; S. Dhakal, D. J. Schonhoft, D. Koirala, Z. Yu, S. Basu and H. Mao, *J. Am. Chem. Soc.*, **2010**, *132*, 8991.

