Parallel DNA G-Quadruplex Induced and Stabilized by Curaxin CBL0137

Jing-Wei Kong(孔敬伟)[1,2], Shuo-Xing Dou(窦硕星)[1,2], Wei Li(李伟)[1,3*], Hui Li(李辉)[4*], and Peng-Ye Wang(王鹏业)[1,2,3*]

[1]*Laboratory of Soft Matter Physics and Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*
[2]*School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*
[3]*Songshan Lake Materials Laboratory, Dongguan 523808, China*
[4]*School of Systems Science and Institute of Nonequilibrium Systems, Beijing Normal University, Beijing 100875, China*

(Received 24 April 2023; accepted manuscript online 29 May 2023)

G-quadruplex (G4) is one of the higher-order DNA structures in guanine-rich sequences which are widely distributed across the genome. Due to their presence in oncogenic promoters and telomeres, G4 DNA structures become the novel targets in anticancer drug designs. Curaxin CBL0137, as an important candidate anticancer drug, can effectively inhibit the growth of multiple cancers. Although there is evidence that anticancer activity of curaxin is associated with its ability to bind DNA and to change the DNA topology, its therapeutic target and the underlying anti-cancer mechanism are still unclear. Here we show, for the first time, that curaxin CBL0137 induces G4 folding from anti-parallel to parallel structures, by single-molecule fluorescence resonance energy transfer technique. More importantly, we find that curaxin CBL0137 promotes G4 folding as well as stabilizes the folded G4 structures with long loops, giving a novel insight into effects of curaxin CBL0137 on DNA structures. Our work provides new ideas for the therapeutic mechanism of curaxin CBL0137 and for designs of new G4-targeting anticancer drugs.

DOI: 10.1088/0256-307X/40/7/078701

G-quadruplex (G4) is one of the non-canonical higher-order DNA structures with guanine-rich sequences, which is maintained by Hoogsteen-type base pairing.[1] G4-forming sequences locate throughout the human genome, such as telomeres and promoter regions of some genes (especially oncogene-associated promoters).[2–4] Since G4 plays critical roles in many biological processes, including gene transcription,[5] telomere maintenance,[6,7] DNA replication initiation, and replication fork progression,[8,9] it serves as an important therapeutic target for many diseases, particularly cancers.[10–12] Curaxin is a new class of non-genotoxic anticancer drugs with reported multi-targeting.[13] As the second generation of curaxin, CBL0137 [Fig. 1(a), top panel] has a promising development prospect on the basis of its water solubility and metabolic stability *in vivo*.[13,14] At present, CBL0137 is under phase-I clinical trials in patients with hematological malignancies, advanced extremity melanoma sarcoma and solid tumors.

Computer modeling suggested that the carbazole portion of curaxin CBL0137 is inserted into the DNA base pair, and the side strand interacts with the major and minor grooves.[15] Preliminary single-molecule magnetic tweezers studies have shown that CBL0137 inserts into base pairs of DNA, reducing the bending and twist rigidity of the DNA while strongly stabilizing the interactions between the DNA double strands.[16] By altering the structure and properties of DNA, CBL0137 could regulate the binding of key proteins on DNA such as CTCF and FACT and then prevent cancer development.[16,17]

Recently, the binding between CBL0137 and G4 was revealed by the nuclear magnetic resonance (NMR) method.[18] However, how CBL0137 regulates the structure and property of G4 remains unclear, limiting further applications of curaxin drugs in clinic. In contrast to ensemble-average methods, the single-molecule fluorescence resonance energy transfer (smFRET) technique may track the real-time dynamical behaviors of individual G4 molecules and provide precise information on G4 conformation changes and folding dynamics.[19–21] Therefore, characterizing the regulation of CBL0137 on G4 at single-molecule level should be potentially useful for understanding the anti-cancer mechanism of CBL0137 and clarifying the therapeutic targets of curaxin drug family.

In this work, by using the smFRET method, we, for the first time to our best knowledge, reveal that CBL0137 regulates the folding dynamics of G4 and increases the proportion of parallel conformation in the folded structures. The smFRET experimental system was constructed with an ssDNA sequence containing an acceptor fluorophore (Cy5) at the 3′ end of the G4 motif, and with its 5′ end hybridized with a complementary stem strand attached by a donor fluorophore (Cy3) at the third nucleotide from the 5′ end [Fig. 1(a), middle and lower panels]. All oligonucleotides for DNA substrates were purchased from Sangon Biotech (Shanghai, China). DNA used in smFRET measurements was annealed with a 1 : 2 mixture of the G4 strand and stem strand by incubating at 95 ℃ for 5 min and slowly cooling down to room temperature in about 7 h. A biotin-streptavidin bridge was used to fix the DNA to the PEG-passivated microscope coverslip. Curaxin CBL0137 was then flowed into the chamber in a buffer solution containing 25 mM Tris-HCl (pH 7.5), KCl and an oxygen scavenger system.

*Corresponding authors. Email: pywang@iphy.ac.cn; huili@bnu.edu.cn; weili007@iphy.ac.cn
© 2023 Chinese Physical Society and IOP Publishing Ltd

Single-molecule fluorescence resonance energy transfer (FRET) experiments were carried out under a total internal reflection fluorescence (TIRF) microscopy. Cy3 was excited by a 532 nm laser. The fluorescence signals from Cy3 and Cy5 were split by a dichroic mirror and collected by an EMCCD. The FRET efficiency was calculated using $E = I_A/(I_A + I_D)$, where I_D and I_A represent the fluorescence intensities of donor and acceptor, respectively. Single-molecule FRET histograms were generated by picking the initial 100 frames of each trace, and fitted by multi-peak Gaussian distributions. The fractions of different G4 conformations were determined by calculating areas of individual Gaussian peaks in the FRET histograms, as performed before.[19,20,22,23] In each experiment, about 300 smFRET traces were selected for the FRET analysis.

We first checked the G4 folding in the absence of drug CBL0137. It was observed that, in 100 mM KCl, the FRET histogram of G4 folding is well fitted by multi-peak Gaussian distributions with two major peaks at $E \approx 0.75$ and 0.9, corresponding to parallel and anti-parallel G4 conformations, respectively [Fig. 1(b), top panel], consistent with the previous studies.[20,24,25] About 80% of the G4 molecules fold into the anti-parallel conformation [Fig. 1(d)]. Next, to study the influence of curaxin CBL0137 on G4 folding, different concentrations of CBL0137 were then used. Interestingly, as the CBL0137 concentration increases, the FRET histograms show that the proportion of folding state at $E \approx 0.75$ increases while that at $E \approx 0.9$ decreases [Fig. 1(b)], which suggests that the G4 conformation changes from anti-parallel state to parallel state. The conformational transition processes could be seen clearly from the FRET traces, where the FRET signal drops from ∼0.9 to ∼0.75 after addition of CBL0137 [Fig. 1(c)]. To further confirm the observed phenomena, we then performed circular dichroism (CD) experiments.[26] Compared with the CD spectrum of G4 having two major positive peaks (partially overlapped) around 265 and 290 nm in the absence of CBL0137, the peak around 290 nm gradually disappears and that around 265 nm becomes more obvious with the increase of CBL0137 concentration [Fig. 1(e)]. Since it is known that the parallel G4 exhibits a positive peak at 260 nm and the anti-parallel one at 295 nm,[24] the CD results are in agreement with the smFRET experiments. These findings suggest that, in the presence of CBL0137, the G4 molecules tend to fold into the parallel conformation.

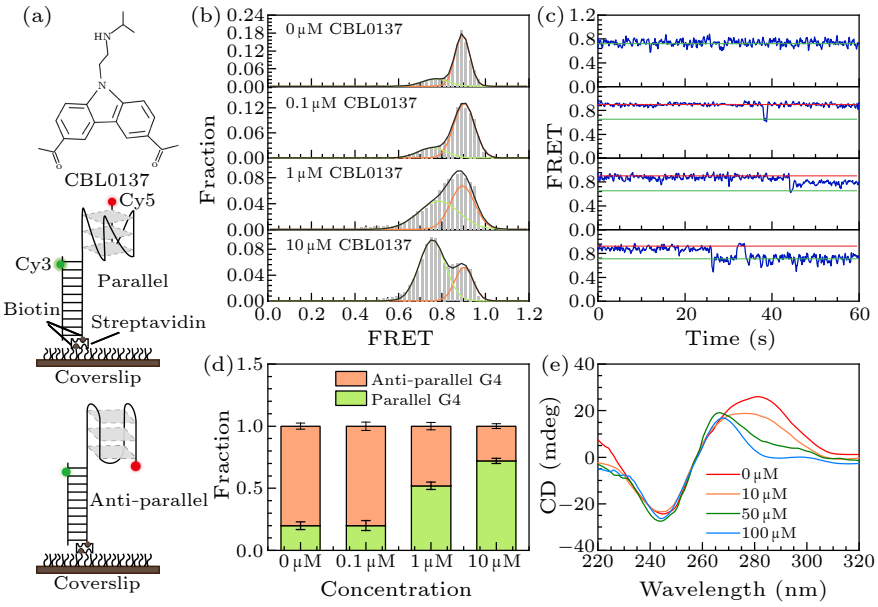

Fig. 1. CBL0137 induces G4 folding with parallel conformation. (a) Structure of CBL0137 (top panel) and schematic representation of the DNA construct with G4 folding in parallel or antiparallel conformations (middle and lower panels). The red dot represents Cy5 and the green dot represents Cy3. (b) FRET histograms for G4 in 100 mM KCl and different concentrations of CBL0137. The FRET histograms display two major peaks at $E \approx 0.75$ and 0.9, corresponding to parallel and anti-parallel G4 conformations, respectively. (c) Four representative smFRET traces with 1 μM CBL0137. The FRET efficiencies at 0.75 and 0.9 are marked by green and red lines. (d) Fractions of parallel or anti-parallel conformations of G4 in different concentrations of CBL0137, determined from the areas of individual Gaussian peaks in (b). Error bars represent the standard deviation. Sample size: $n = 297$, 295, 338, and 393 for 0, 0.1, 1, and 10 μM CBL0137, respectively. (e) CD spectra for G4 under different concentrations of CBL0137.

As the folding of G4 is affected by the concentration of KCl,[27,28] we then changed the concentration of KCl to examine whether the above-observed effect of CBL0137 on G4 folding will be altered. First, we reduced the concentration of KCl to 50 mM. In this case, the FRET histograms show three major peaks at $E \approx 0.45$, 0.75, and 0.9 [Fig. 2(a)]. The minor peak at $E \approx 0.45$ should be G3 or G-hairpin, which are intermediates in G4 folding and appeared due to the reduction in KCl concentration.[19,27] With the CBL0137 concentration increasing from 0 to 10 μM, the proportion of the parallel state at $E \approx 0.75$ increased from ∼35% to ∼60% [Fig. 2(c)]. Then we fur-

ther reduced the KCl concentration to 10 mM [Fig. 2(b)]. With the addition of CBL013, the proportion of the parallel state at $E \approx 0.75$ increased from $\sim$35% to $\sim$56% [Fig. 2(d)]. Moreover, a new peak at $E \approx 0.2$ appears in the FRET histograms, suggesting that a small proportion of ssDNA appears when CBL0137 is present. This may be due to the possible competition between the K^+ at such a low concentration and the CBL0137. We should note that the physiological concentration of K^+ in cells is reported to be around 140 mM.[29,30] Therefore, the concentration of K^+ at 10 mM is far below the physiological concentration, and we speculate that the induced ssDNA by CBL0137 would not be present in cells. After analyzing all the results, it is evident that CBL0137 has the ability to alter the folding behaviors of G4 in various concentrations of KCl. As a result, G4 prefers to fold in the parallel conformation.

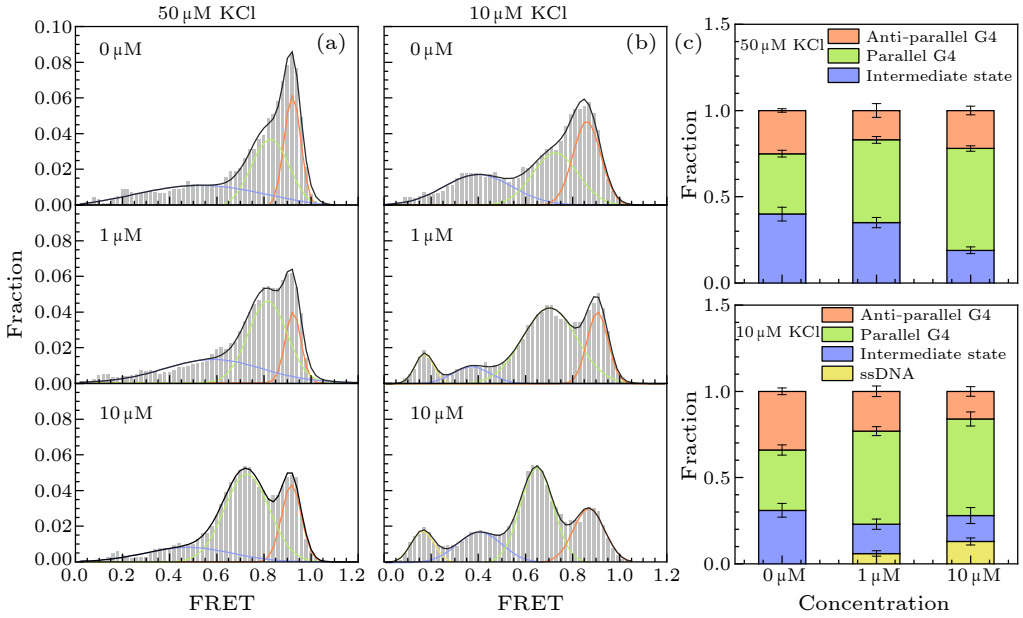

Fig. 2. The influence of CBL0137 on G4 folding is not altered by the concentration of KCl. [(a), (b)] FRET histograms for G4 in 50 and 10 mM KCl with different concentrations of CBL0137. Red, anti-parallel state; green, parallel state; blue, intermediate state; yellow, ssDNA. [(c), (d)] Fractions of different folding structures with different concentrations of CBL0137 in 50 and 10 mM KCl. Error bars represent the standard deviation. Sample sizes: n = 378, 316, and 328 (50 mM KCl) or 358, 325, and 354 (10 mM KCl) for 0, 1, and 10 μM CBL0137, respectively.

The phenomenon that CBL0137 induces G4 folding transition from anti-parallel to parallel conformations suggests that CBL0137 binds and interferes with the G4 folding. To further investigate how CBL0137 alters the topologies of G4, we designed a series of substrates with different loop lengths based on the human telomeric G4 (Table 1), where the name G4Ln indicates that the central TTA loop in the original G4 sequence has been replaced with n bases.

Table 1. Sequences of the oligonucleotides with different loop lengths.

Name	Sequence (5′ → 3′)
G4	GGGTTA GGGTTA GGGTTA GGGTTA
G4L1	GGGT GGGT GGGT GGGTTA
G4L2	GGGTT GGGTT GGGTT GGGTTA
G4L4	GGGTTTA GGGTTTA GGGTTTA GGGTTA
G4L5	GGGTTTTA GGGTTTTA GGGTTTTA GGGTTA

For the G4L1 and G4L2 with shorter loops, only the parallel structure of G4 was observed in 25 mM Tris–HCl (pH 7.5) with 100 mM KCl, as the same as previously reported.[31–33] Addition of CBL0137 did not change the G4 structures [Figs. 3(a) and 3(b)]. This can be easily understood because the short loop has already stabilized the G4 and the CBL0137 has no further influence on the folding of these two mutant G4. As for G4L4 and G4L5 with longer loops, the fraction of unfolded molecules increases as can be seen that from the FRET histograms with multiple peaks, which is caused by the increase of loop length.[31–33] However, when adding CBL0137 into the chamber, we observed that the proportion of parallel G4 increases while that of unfolded G4 molecules decreases. Specifically, for the G4L4 with 10 μM CBL0137, the proportion of parallel G4 increases from $\sim$20% to $\sim$30%, the proportion of unfolded G4 molecules (including ssDNA and intermediate state) decreases from $\sim$45% to $\sim$40%, and the proportion of antiparallel G4 remains essentially the same [Fig. 3(e)]. For the G4L5 with 10 μM CBL0137, the proportions of both parallel and antiparallel G4 increase to $\sim$35%, while that of unfolded G4 molecules decreases to $\sim$30% [Fig. 3(e)]. These results indicate that CBL0137 could promote the unfolded molecules to fold into parallel G4 and help stabilize the G4 folding. Moreover, we note that the proportions of anti-parallel structure for G4L4 and G4L5 with longer loops have not decreased after adding the CBL0137. According to the previous study using NMR and CD methods,[18] the CBL0137 binds to the exposed sides of the two external G-tetrads and stabilizes the G4 structure. Thus one possible explanation for the observed phenomenon is that, while a short

loop may partially block the binding of CBL0137 to G4 in the anti-parallel conformation [see Fig. 1(a)], a long loop no longer blocks the binding, and as a result, the anti-parallel G4 with longer loops can also well be stabilized by CBL0137.

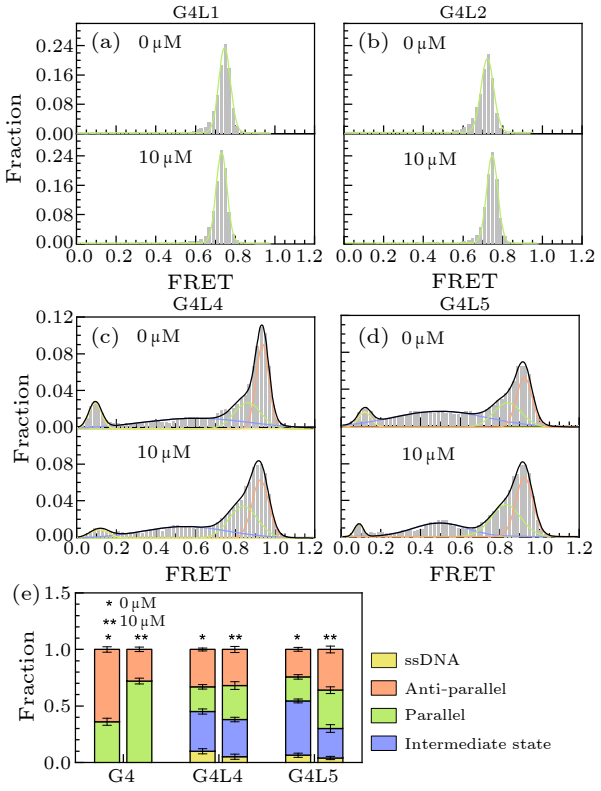

Fig. 3. Effect of CBL0137 on G4 with different loop lengths. (a)–(d) FRET histograms for G4L1 (a), G4L2 (b), G4L4 (c) and G4L5 (d) in 100 mM KCl without CBL0137 or with 10 μM CBL0137. Red line, anti-parallel state; green line, parallel state; blue, intermediate state; yellow, ssDNA. (e) Fractions of different folding structures for G4, G4L4, and G4L5 in 100 mM KCl without CBL0137 or with 10 μM CBL0137. Error bars represent the standard deviation. Sample sizes: $n = 297$ and 295 (G4), 428 and 340 (G4L4), 469 and 422 (G4L5), for 0 and 10 μM CBL0137, respectively.

In summary, we have studied the effect of curaxin CBL0137 on the folding of G4 by smFRET. Our experimental results have revealed that CBL0137 induces G4 to fold into a parallel structure, and the trend is not affected by KCl concentration. The increase in loop length results in an increase in the proportion of unfolded G4 molecules, while the addition of CBL0137 promotes and stabilizes the folding of G4. Compared with the ensemble-average and static results from NMR that CBL0137 interacts with G4 in both telomeres and oncogene promoter regions,[18] our study directly illustrates the folding dynamics of G4 under the influence of CBL0137, with single-molecule resolutions. It is known that G4 sequences are present in ∼50% of human gene promoters, especially in oncogene promoters[34,35] including c-MYC,[36] VEGF,[37] HIF-1α,[38] and c-KIT.[39] More importantly, the parallel G4 is shown to be more stable than anti-parallel ones from the previous NMR and CD measurements.[40] Therefore, considering that G4 formation at the promoter regions of genes could impair initiation of transcription by the RNA polymerase,[41] the stabilization of G4 especially in parallel conformation is of great significance for inhibiting the expression of oncogenes. The experimental results obtained here suggest that CBL0137 not only induces transition of G4 from anti-parallel to parallel conformations but also stabilizes the folding of parallel conformation. As a result, the binding of transcription factors on promoters with G4 sequences may be significantly impeded, thus effectively inhibiting the expression of oncogenes. The present study improves our understanding of the anti-cancer mechanism of CBL0137, and provides new insights into the therapeutic targets for drug designs.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (Grant Nos. 10225417, 21991133, 12122402, and 12074043), and the National Basic Research Program of China (Grant No. 2006CB601003).

References

[1] Simonsson T 2001 *Biol. Chem.* **382** 621
[2] Maizels N and Gray L T 2013 *Plos Genet.* **9** e1003468
[3] Varizhuk A, Isaakova E, and Pozmogova G 2019 *Bioessays* **41** 1900091
[4] Bochman M L, Paeschke K, and Zakian V A 2012 *Nat. Rev. Genet.* **13** 770
[5] Wu G H, Xing Z, Tran E J, and Yang D Z 2019 *Proc. Natl. Acad. Sci. USA* **116** 20453
[6] Wang Q, Liu J Q, Chen Z, Zheng K W, Chen C Y, Hao Y H, and Tan Z 2011 *Nucleic Acids Res.* **39** 6229
[7] Moye A L, Porter K C, Cohen S B, Phan T, Zyner K G, Sasaki N, Lovrecz G O, Beck J L, and Bryan T M 2015 *Nat. Commun.* **6** 7643
[8] Bryan T M 2019 *Molecules* **24** 3439
[9] Liu Y R, Wang P Y, Li W, and Xie P 2021 *Chin. Phys. Lett.* **38** 118701
[10] Spiegel J, Adhikari S, and Balasubramanian S 2020 *Trends Chem.* **2** 123
[11] Balasubramanian S and Neidle S 2009 *Curr. Opin. Chem. Biol.* **13** 345
[12] Li B, Ji C, Lu X M, Liu Y R, Li W, Dou S X, Li H, and Wang P Y 2018 *Chin. Phys. Lett.* **35** 118701
[13] Di Bussolo V and Minutolo F 2011 *ChemMedChem* **6** 2133
[14] Jin M Z, Xia B R, Xu Y, and Jin W L 2018 *Front. Oncology* **8** 598
[15] Safina A, Cheney P, Pal M, Brodsky L, Ivanov A, Kirsanov K, Lesovaya E, Naberezhnov D, Nesher E, Koman I, Wang D, Wang J M, Yakubovskaya M, Winkler D, and Gurova K 2017 *Nucleic Acids Res.* **45** 1925
[16] Lu K, Liu C F, Liu Y N, Luo A F, Chen J, Lei Z C, Kong J W, Xiao X, Zhang S M, Wang Y Z, Ma L, Dou S X, Wang P Y, Li M, Li G H, Li W, and Chen P 2021 *Biochemistry* **60** 494
[17] Kantidze O L, Luzhin A V, Nizovtseva E V, Safina A, Valieva M E, Golov A K, Velichko A K, Lyubitelev A V, Feofanov A V, Gurova K V, Studitsky V M, and Razin S V 2019 *Nat. Commun.* **10** 1441
[18] Dallavalle S, Mattio L M, Artali R, Musso L, Avino A, Fabrega C, Eritja R, Gargallo R, and Mazzini S 2021 *Int. J. Mol. Sci.* **22** 6476
[19] Lu X M, Li H, You J, Li W, Wang P Y, Li M, Dou S X, and Xi X G 2018 *J. Phys. Chem. B* **122** 9499
[20] Hou X M, Fu Y B, Wu W Q, Wang L, Teng F Y, Xie P, Wang P Y, and Xi X G 2017 *Nucleic Acids Res.* **45** 11401
[21] Pan B Y, Dou S X, Yang Y, Xu Y N, Bugnard E, Ding X Y, Zhang L Y, Wang P Y, Li M, and Xi X G 2010 *J. Biol. Chem.* **285** 15884
[22] Teng F Y, Wang T T, Guo H L, Xin B G, Sun B, Dou S X, Xi X G, and Hou X M 2020 *J. Biol. Chem.* **295** 17646
[23] Wu W Q, Zhang M L, and Song C P 2020 *J. Biol. Chem.* **295** 5461
[24] Tippana R, Xiao W K, and Myong S 2014 *Nucleic Acids Res.* **42** 8106

[25] Aznauryan M, Sondergaard S, Noer S L, Schiott B, and Birkedal V 2016 *Nucleic Acids Res.* **44** 11024
[26] You J, Li H, Lu X M, Li W, Wang P Y, Dou S X, and Xi X G 2017 *Biosci. Rep.* **37** Bsr20170771
[27] Lee J Y, Okumus B, Kim D S, and Ha T J 2005 *Proc. Natl. Acad. Sci. USA* **102** 18938
[28] Qiao H Z, Wu Y Y, Tu Y, and Ji C M 2021 *Chin. Phys. B* **30** 018702
[29] Stein W D 2002 *Int. Rev. Cytol.* **215** 231
[30] Lang F 2007 *J. Am. College Nutr.* **26** 613S
[31] Hazel P, Huppert J, Balasubramanian S, and Neidle S 2004 *J. Am. Chem. Soc.* **126** 16405
[32] Bugaut A and Balasubramanian S 2008 *Biochemistry* **47** 689
[33] Guedin A, Gros J, Alberti P, and Mergny J L 2010 *Nucleic Acids Res.* **38** 7858
[34] Huppert J L and Balasubramanian S 2007 *Nucleic Acids Res.* **35** 406
[35] Eddy J and Maizels N 2006 *Nucleic Acids Res.* **34** 3887
[36] You H J, Wu J Y, Shao F W, and Yan J 2015 *J. Am. Chem. Soc.* **137** 2424
[37] Agrawal P, Hatzakis E, Guo K X, Carver M, and Yang D Z 2013 *Nucleic Acids Res.* **41** 10584
[38] De Armond R, Wood S, Sun D Y, Hurley L H, and Ebbinghaus S W 2005 *Biochemistry* **44** 16341
[39] Hsu S T D, Varnai P, Bugaut A, Reszka A P, Neidle S, and Balasubramanian S 2009 *J. Am. Chem. Soc.* **131** 13399
[40] Chen Y and Yang D 2012 *Current Protocols in Nucleic Acid Chemistry* **17** 17.5
[41] Rhodes D and Lipps H J 2015 *Nucleic Acids Res.* **43** 8627