DNA Structures

DOI: 10.1002/ange.201302995

Structural Basis of DNA Quadruplex–Duplex Junction Formation**

*Kah Wai Lim and Anh Tuân Phan**

Aside from the double helix (duplex),[1] DNA can adopt alternative conformations, including the four-way Holliday junction[2] and the G-quadruplex.[3] The G-quadruplex is a four-stranded helical assembly made up of multiple stacked G-tetrads,[4] with a diverse range of possible topologies. G-quadruplex formation has been implicated in cellular processes,[5,6] including recombination[7] and replication.[8] Furthermore, G-quadruplex motifs have been incorporated in the design of DNA nanoarchitectures[9] and nanodevices.[10] An understanding on the compatibility between duplex and quadruplex DNA would open new avenues for drug design efforts targeting genomic G-rich sequences,[3,11,12] and could facilitate the integration of such motifs in DNA structural engineering and nanotechnology.[13]

In canonical B-DNA,[1] right-handed antiparallel polynucleotide chains are held together by Watson–Crick base pairs (Figure 1a). The relative alignment of the two phosphate backbones defines the major and minor grooves, with a strand separation (henceforth defined as the distance between matching pairs of phosphate groups) of approximately 18 Å. The G-quadruplex,[3] in contrast, is stabilized by Hoogsteen hydrogen bonds[14] and stacking interactions among the tetrad-forming guanine bases (Figure 1b), with coordinating cations running through the middle channel. The diversity of G-quadruplex topologies arises from different relative orientations of the four strands constituting the G-tetrad core (Figure 1c) and various possibilities of connecting them with linkers (known as loops). Each type of core has distinct groove dimensions (a combination of wide, medium, or narrow grooves); a wide groove (flanked by antiparallel strands) has a strand separation of approximately 19 Å, a medium groove (flanked by parallel strands) has a strand separation of approximately 16 Å, while a narrow groove (flanked by antiparallel strands, arranged in the opposite strand directionalities with respect to the wide groove) has a strand separation of approximately 12 Å.

To address the compatibility between duplex and quadruplex DNA, we explored the incorporation of a duplex hairpin (5′-CGCGAA*GCA*TTCGCG-3′, loop[15] in italics; Table S1 and Figure S1 in the Supporting Information) across the various geometries of a quadruplex core. Conceptually, with a duplex strand joint with one strand of a tetrad core, there are at least three possibilities of connection for the second strand that could lead to coaxial orientation of the helices (Figure 1d). Additionally, orthogonal connectivity of a duplex and a quadruplex (Figure 1e) could be possible. A stable quadruplex–duplex junction could take form, if duplex attachment for that particular arrangement is geometrically

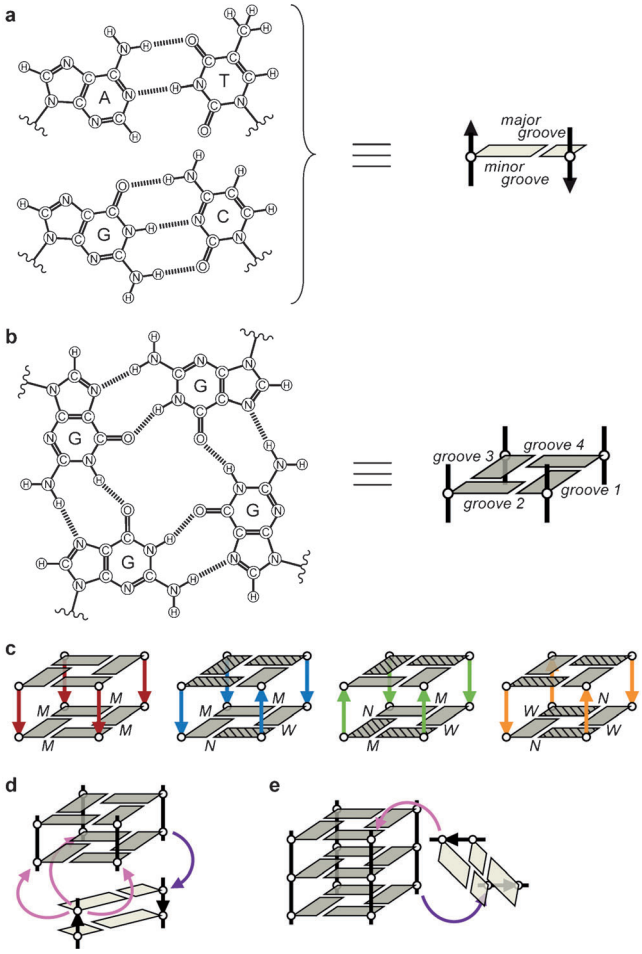

Figure 1. Geometry of duplex and quadruplex DNA and their juxtaposition. The base-pair and tetrad schematics in (a) and (b) apply throughout the text. a) Watson–Crick A·T and G·C base pairs, with the major and minor grooves marked in the schematic. b) The G-tetrad, with four separate grooves as indicated. c) Four G-quadruplex core topologies, each having a distinct arrangement of groove dimensions (*W*: wide, *M*: medium, and *N*: narrow). Guanine bases in *syn* conformation are shaded. d,e) Coaxial (d) and orthogonal (e) orientation of duplex and quadruplex helices at the quadruplex–duplex junction, shown here with a fixed (purple) and a variable (pink) attachment point for the two duplex strands.

[*] K. W. Lim, Prof. Dr. A. T. Phan
School of Physical and Mathematical Sciences
Nanyang Technological University
21 Nanyang Link, Singapore 637371 (Singapore)
E-mail: phantuan@ntu.edu.sg
K. W. Lim
School of Biological Sciences
Nanyang Technological University
60 Nanyang Drive, Singapore 637551 (Singapore)

[**] This research was supported by Singapore Ministry of Education and Nanyang Technological University grants to A. T. Phan.
Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201302995.

feasible and energetically favorable. Out of over twenty quadruplex–duplex constructs examined, we present here the high-resolution NMR structures of five representative constructs, each of which illustrates a disparate principle of connection between a duplex and a quadruplex. In each case, strategic placement of the hairpin and auxiliary structural elements culminated in a unique structure, and rigorous spectral assignment approaches were employed for structural characterization.

We first demonstrated that a duplex can be extended seamlessly across the wide groove of an antiparallel G-quadruplex[16] (construct I, "JWC": *j*unction across *w*ide groove with two *c*ontinuous strands; Figure 2a, Table S2 and Figure S2 in the Supporting Information). There is a gradual transition between the minor groove of the duplex and the wide groove of the quadruplex (strand separation expanding from 18 Å to 19 Å), and this compatibility was shown to apply universally across wide grooves of different G-quadruplexes (Figure S3 in the Supporting Information). Continual stacking of bases over the duplex and quadruplex segments contributes towards the overall stability of the structure.

A nick[17] can be introduced within the G-tract bordering the quadruplex–duplex junction such that the terminal G-tetrad now serves as a foothold for the loose strand,[18] which could be anchored at the 5′- or 3′-end (Figure S4 in the

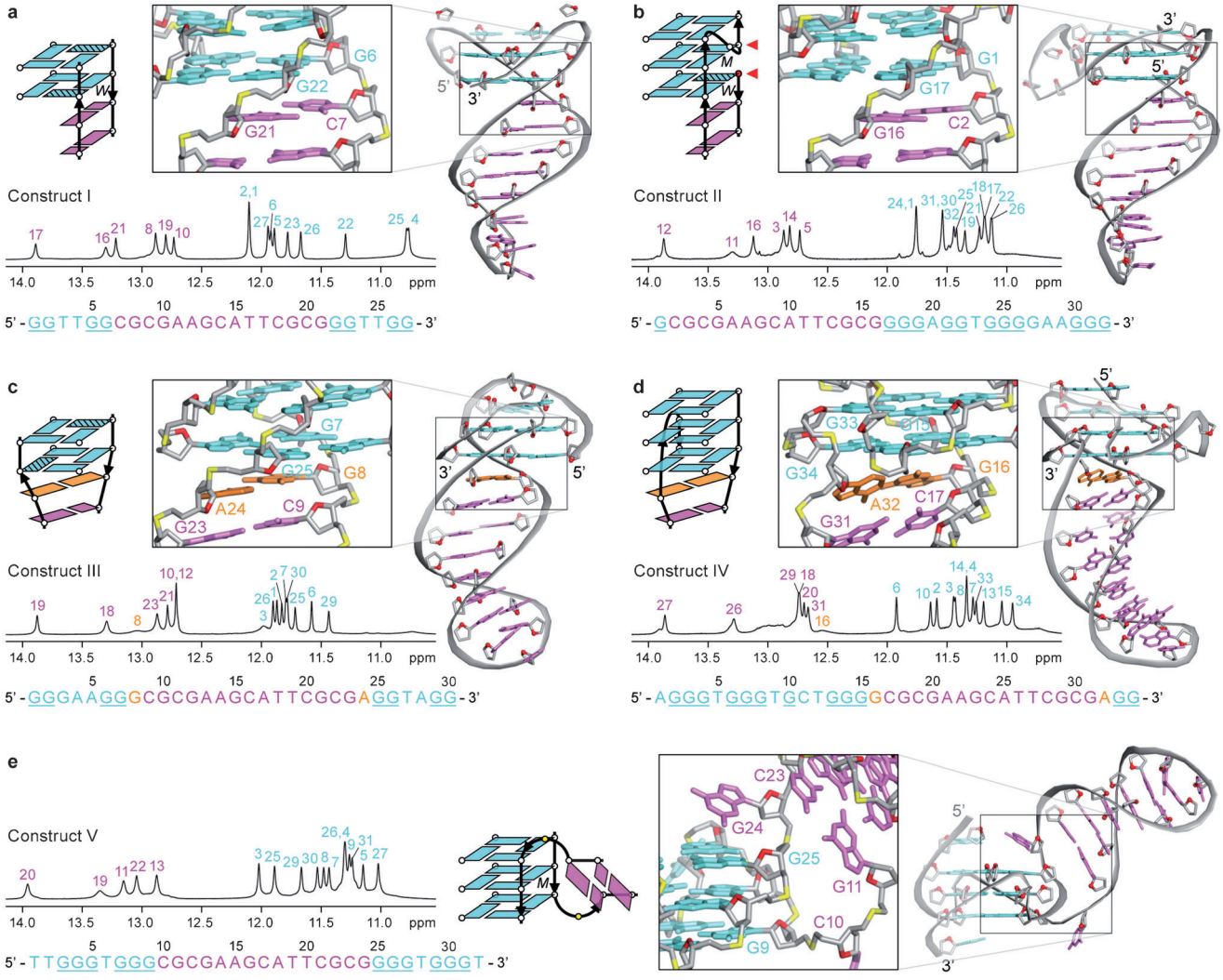

Figure 2. NMR structures of quadruplex–duplex constructs. Different duplex attachment strategies leading to the coaxial (a–d) and orthogonal (e) orientation of duplex and quadruplex helices at the quadruplex–duplex junctions. Ribbon view of a representative structure is shown in each panel together with the sequence (tetrad-forming guanines underlined) and the 1D imino proton NMR spectrum (assignments listed over the peaks) of the construct. Structural details at the junctions are enlarged, accompanied by schematics illustrating the principles of connection. Bases from the duplex and quadruplex segments are colored in magenta and cyan, respectively; G·A Watson–Crick mismatch base pairs: orange; backbone and sugar: silver; O4′ atoms: red; phosphorus atoms: yellow. a) A duplex can be extended continuously from the wide groove (PDB code: 2M8Z). b) Introduction of a nick onto a G-tract can accommodate a duplex across the medium-made-wide groove (PDB code: 2M90). c) An adaptor G·A base pair can be utilized to bridge a duplex across the diagonal corners of a tetrad (PDB code: 2M91). d) The snapback approach can be applied to sidestep the unfavorable connection of a duplex across the narrow groove, instead bridging the duplex across the same edge of the tetrad core through the diagonal corners (PDB code: 2M92). e) Sideway connectivity of a duplex, with the disruption of a base pair (highlighted as yellow circles in the schematic) at the junction (PDB code: 2M93).

Supporting Information). The resulting strand discontinuity permits the local accommodation of a wide groove across the medium groove edge of a quadruplex (construct II, "JW_MD": *j*unction across *m*edium-made-*w*ide groove with a *d*iscontinuous strand; Figure 2b, Table S3 and Figure S5 in the Supporting Information), effectively widening the strand separation from 16 Å to 18–19 Å at the anchor point. This approach removes the need to maintain the strict antiparallel strand orientation of the two G columns mediating the duplex.

To connect a duplex across the diagonal corners[18,19] of a tetrad (the phosphate groups would have to span a distance of >20 Å), an adaptor G·A Watson–Crick mismatch base pair[20] could be placed at the junction to ease the transition between the two segments with different dimensions (construct III, "JDC": *j*unction across *d*iagonal corners with two *c*ontinuous strands; Figure 2c, Table S4 and Figure S6 in the Supporting Information). There is an associated displacement of the helical axis of the duplex with respect to that of the quadruplex.

In contrast to the wide groove, continual stacking of a duplex over the narrow groove of a quadruplex would not be favorable; the two strands would have to shrink from approximately 18 Å apart to 12 Å at a single step. Instead, a snapback approach, previously observed in a promoter G-quadruplex[21] and an RNA quadruplex–duplex junction,[22] could be adapted to bridge a duplex across this edge of the tetrad core (construct IV, "JD_NS": *j*unction across *n*arrow groove-made-*d*iagonal corners through a *s*napback; Figure 2d, Table S5 and Figure S7 in the Supporting Information).

We also investigated the possibility of orthogonal orientation of duplex and quadruplex helices by integrating the hairpin as the middle propeller loop of an all-parallel-stranded G-quadruplex[23] (construct V, "JMC": *j*unction across *m*edium groove with two *c*ontinuous strands; Figure 2e, Table S6 and Figure S8 in the Supporting Information). A robust structure is formed with the breaking of a base pair at the quadruplex–duplex junction. The extrusion of the two bases is somewhat akin to the B–Z junction,[24] reflecting perhaps a general mechanism to mediate helical components having different handedness/axial orientation. In the B–Z junction, this permits the continual stacking of bases between two segments with different handedness (right-handed B-DNA onto left-handed Z-DNA). Here, local geometry at the junction necessitates the disruption of the base pair in order to achieve maximal base stacking separately at the duplex and quadruplex ends.

Note that even though the sequences were arbitrarily conceived, the five quadruplex–duplex constructs (Figure 2) were designed to cover the fundamental aspects of the connectivity between a duplex and a quadruplex (or the incorporation of a duplex in a quadruplex loop).[19,25,26] Junction variants can be rationalized as an adaptation or composite of the junctions presented. However, we should not discard the possibility that more exotic junction types may yet exist. Additionally, these continuous quadruplex–duplex junctions (in which there are no strand breaks at the points of connection between the duplex and the quadruplex) should be distinguished from discontinuous quadruplex–duplex junctions/interfaces (in which there is at least one strand break where the duplex and quadruplex abut), which could arise for instance at the beginning of the single-stranded overhang of telomeric DNA.[27] Nevertheless, the junction types presented could still provide valuable structural insights for the latter cases.

We further contemplated the conglomeration of multiple duplexes through a single G-quadruplex hub.[28] This would involve the grafting of each individual duplex stem onto the quadruplex core by utilizing a complementary set of attachment strategies appropriate at each of these junctions. We demonstrated this with the successful attachment of three hairpin stems onto a G-quadruplex scaffold, generating a G-junction (Figure 3, Table S7 and Figure S9 in the Supporting Information; basis of design detailed in Figure 3 captions). Overall, the structure comprises three orthogonally oriented duplex arms, brought together by the complement of twelve guanine residues constituting the tetrad core. Owing to the diverse G-quadruplex folding topologies available, a multitude of G-junction motifs could be conceived in principle.

The facile yet stable incorporation of duplex hairpin elements within the loops of G-quadruplexes (this work and unpublished data) calls for a consideration of G-quadruplexes with long loop lengths in the evaluation of potential G-quadruplex-forming sequences in the human genome.[11] On the other hand, an understanding on the compatibility between duplex and quadruplex DNA should help with the design of quadruplex-binding ligands.[3] In addition, alternative targeting strategies can be apprehended in light of this understanding, ranging from antigene targeting[29] to quadruplex–duplex junction binders.[22]

This study presents a paradigm to exact control over the folding topology of a G-quadruplex by the strategic placement of duplex components and auxiliary structural elements. The duplex stem drives the core topology towards the desired fold,[30] and eventual formation of a G-quadruplex dictates the orientation of the duplex arm. Such cooperativity could be exploited in the makeup of dynamic DNA assembly.[10] Previously, DNA aptamers and DNA enzymes containing G-rich elements were isolated from combinatorial pools of random-sequence DNA,[31] suggesting an organizational role of G-tetrad motifs in these oligonucleotides. A thorough reconciliation of duplex and quadruplex DNA could pave the way towards the rational design of such complexes[9] based on G-junction motifs.

With a stable core that can potentially serve as an immobile junction, the G-junction is well-poised for integration in DNA nanotechnology.[13] The orthogonal orientation of the three duplex arms in the G-junction motif presented (Figure 3) suggests an immediate route towards the assembly of three-dimensional DNA nanoarchitectures. Furthermore, G-rich functional elements (e.g. DNA aptamers,[31] enzymes,[31] sensors,[32] and nanowires[33]) should now be adaptable with an unprecedented level of control. The dependence of G-quadruplex formation on the presence of cations could also grant an additional layer of control over the assembly of DNA nanomaterials.

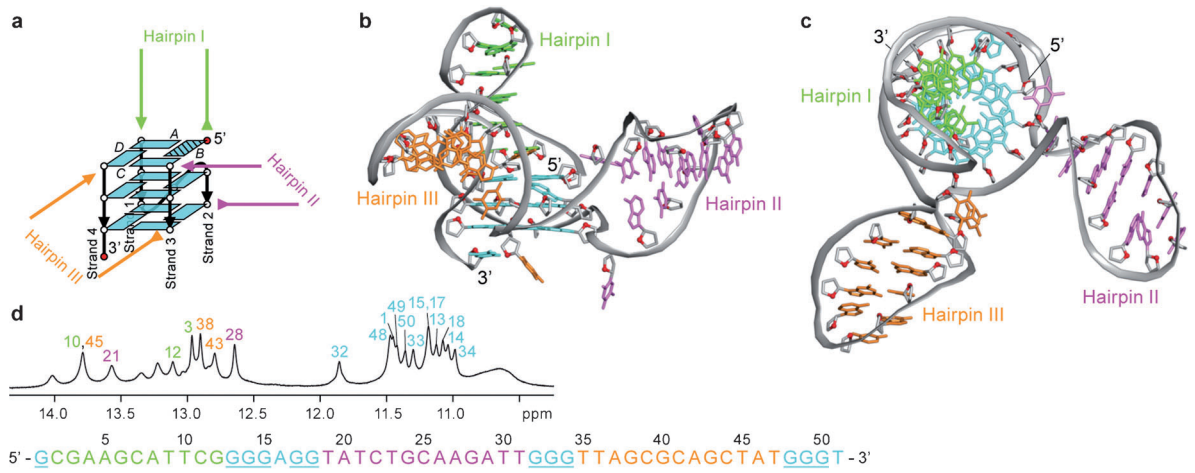

Figure 3. Rational design of a G-junction bridging three duplex stems. Starting from a G-tetrad core, duplex stems can be incrementally incorporated across its various edges by utilizing attachment strategies that are applicable at each of the resultant quadruplex–duplex junctions. a) Illustration of the duplex attachment strategies that led to the generation of the G-junction construct. Introduction of a nick on strand 2 allows the coaxial attachment of hairpin I (green) onto the G-tetrad core (cyan). Hairpin II (magenta) and hairpin III (orange) are connected across medium grooves (edge B and edge C, respectively) and project laterally from the core. The 5′- and 3′-terminals are shown as red circles. b, c) Side (b) and top-down (c) views of the G-junction model in ribbon representation. d) Sequence and 1D imino proton NMR spectrum (assignments listed over the peaks) of the G-junction construct.

The establishment of the quadruplex–duplex and G-junction constructs accomplished a comprehensive inquiry into the compatibility between duplex and quadruplex DNA, empowering us towards the structural and functional applications of such complexes. We believe the identification of genomic sequences with the capacity to form such motifs could have biological and therapeutic implications.

Received: April 10, 2013
Published online: June 21, 2013

Keywords: DNA structures · duplex · helical junctions · NMR spectroscopy · quadruplex

[1] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.
[2] D. M. J. Lilley, *Q. Rev. Biophys.* **2000**, *33*, 109–159.
[3] *Quadruplex Nucleic Acids* (Eds.: S. Neidle, S. Balasubramanian), RSC Biomolecular Sciences, Cambridge, **2006**.
[4] M. Gellert, M. N. Lipsett, D. R. Davies, *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013–2018.
[5] N. Maizels, *Nat. Struct. Mol. Biol.* **2006**, *13*, 1055–1059.
[6] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, *Nat. Chem.* **2013**, *5*, 182–186.
[7] L. A. Cahoon, H. S. Seifert, *Science* **2009**, *325*, 764–767.
[8] K. Paeschke, J. A. Capra, V. A. Zakian, *Cell* **2011**, *145*, 678–691.
[9] U. Feldkamp, C. M. Niemeyer, *Angew. Chem.* **2006**, *118*, 1888–1910; *Angew. Chem. Int. Ed.* **2006**, *45*, 1856–1876.
[10] Y. Krishnan, F. C. Simmel, *Angew. Chem.* **2011**, *123*, 3180–3215; *Angew. Chem. Int. Ed.* **2011**, *50*, 3124–3156.
[11] J. L. Huppert, *Biochimie* **2008**, *90*, 1140–1148.
[12] S. Balasubramanian, L. H. Hurley, S. Neidle, *Nat. Rev. Drug Discovery* **2011**, *10*, 261–275.
[13] N. C. Seeman, *Annu. Rev. Biochem.* **2010**, *79*, 65–87.
[14] K. Hoogsteen, *Acta Crystallogr.* **1963**, *16*, 907–916.
[15] L. M. Zhu, S. H. Chou, J. D. Xu, B. R. Reid, *Nat. Struct. Biol.* **1995**, *2*, 1012–1017.
[16] P. Schultze, R. F. Macaya, J. Feigon, *J. Mol. Biol.* **1994**, *235*, 1532–1547.
[17] A. T. Phan, V. Kuryavyi, J. B. Ma, A. Faure, M. L. Andreola, D. J. Patel, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 634–639.
[18] A. T. Phan, V. Kuryavyi, H. Y. Gaw, D. J. Patel, *Nat. Chem. Biol.* **2005**, *1*, 167–173.
[19] G. D. Balkwill, T. P. Garner, H. E. L. Williams, M. S. Searle, *J. Mol. Biol.* **2009**, *385*, 1600–1615.
[20] N. B. Leontis, J. Stombaugh, E. Westhof, *Nucleic Acids Res.* **2002**, *30*, 3497–3531.
[21] A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle, D. J. Patel, *J. Am. Chem. Soc.* **2007**, *129*, 4386–4392.
[22] A. T. Phan, V. Kuryavyi, J. C. Darnell, A. Serganov, A. Majumdar, S. Ilin, T. Raslin, A. Polonskaia, C. Chen, D. Clain, R. B. Darnell, D. J. Patel, *Nat. Struct. Mol. Biol.* **2011**, *18*, 796–804.
[23] N. Q. Do, A. T. Phan, *Chem. Eur. J.* **2012**, *18*, 14752–14759.
[24] S. C. Ha, K. Lowenhaupt, A. Rich, Y. G. Kim, K. K. Kim, *Nature* **2005**, *437*, 1183–1186.
[25] A. Bourdoncle, A. E. Torres, C. Gosse, L. Lacroix, P. Vekhoff, T. Le Saux, L. Jullien, J. L. Mergny, *J. Am. Chem. Soc.* **2006**, *128*, 11094–11105.
[26] S. L. Palumbo, S. W. Ebbinghaus, L. H. Hurley, *J. Am. Chem. Soc.* **2009**, *131*, 10878–10891.
[27] J. S. Ren, X. G. Qu, J. O. Trent, J. B. Chaires, *Nucleic Acids Res.* **2002**, *30*, 2307–2315.
[28] E. A. Venczel, D. Sen, *J. Mol. Biol.* **1996**, *257*, 219–224.
[29] M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden, P. E. Nielsen, *Nature* **1993**, *365*, 566–568.
[30] J. Zhou, A. Bourdoncle, F. Rosu, V. Gabelica, J. L. Mergny, *Angew. Chem.* **2012**, *124*, 11164–11167; *Angew. Chem. Int. Ed.* **2012**, *51*, 11002–11005.
[31] R. R. Breaker, *Curr. Opin. Chem. Biol.* **1997**, *1*, 26–31.
[32] J. W. Liu, Z. H. Cao, Y. Lu, *Chem. Rev.* **2009**, *109*, 1948–1998.
[33] K. Dutta, T. Fujimoto, M. Inoue, D. Miyoshi, N. Sugimoto, *Chem. Commun.* **2010**, *46*, 7772–7774.