G-Quadruplexes

International Edition: DOI: 10.1002/anie.201505348
German Edition: DOI: 10.1002/ange.201505348

The Effect of DNA Sequence Directionality on G-Quadruplex Folding

*Maja Marušič and Janez Plavec**

Abstract: *Sequence inversion in G-rich DNA from 5′→3′ to 3′→5′ exerts a substantial effect on the number of structures formed, while the type of G-quadruplex fold is in fact determined by the presence of K^+ or Na^+ ions. The melting temperatures of G-quadruplexes adopted by oligonucleotides with sequences in the 5′→3′ direction are higher than those of their 3′→5′ counterparts with both KCl and NaCl. CD, UV, and NMR spectroscopy demonstrates the importance of primary sequence for the structural diversity of G-quadruplexes. The changes introduced by mere sequence reversal of the G-rich DNA segment have a substantial impact on the polymorphic nature of the resulting G-quadruplexes and their potential physiological roles. The insights resulting from this study should enable extension of the empirical rules for the prediction of G-quadruplex topology.*

G-Quadruplexes are four-stranded DNA structures that have become increasingly popular within the last 20 years, during which time alternative DNA structures have been shown to represent an additional level of control in DNA replication and transcription processes.[1–3] A large, still growing repertoire of possible G-quadruplex topologies differing in the orientation of their DNA strands and the types of connecting loops that a G-rich oligonucleotide can adopt has made the prediction of G-quadruplex topology very challenging. We have initiated a study to explore sequence directionality as a factor that is inherently connected with the folding of G-rich oligonucleotides with a clear question: will an oligonucleotide with a given primary sequence fold into a different G-quadruplex upon reversal of the sequence directionality? It has been previously shown that changing or modifying as much as one nucleotide can completely abolish G-quadruplex formation,[4] and one would intuitively expect that reversing the sequence direction from a 5′→3′ to 3′→5′ would cause dramatic changes in topology and/or the number of formed structures. It is noteworthy that DNA melting-temperature prediction algorithms for double-stranded stem regions do not differentiate between pairs of constructs with reversed 5′→3′ sequences (Table S1 in the Supporting Information). It has been shown, however, that double- and triple-stranded DNAs exhibit conformational polymorphism of the backbone and base pairs that in turn lead to different stacking interactions and thermal stability of the DNA structure(s).[5–8] Predicted helical parameters exhibit different values for pairs of oligonucleotides with reversed sequence (Figure S1 in the Supporting Information). If sequence inversion influences the structure of more common double-stranded DNA (e.g., different behavior of AT and TA base-pairs),[9–11] then its effect on more complex G-quadruplexes would be expected to be substantial.

In exploring the effects of sequence directionality on the folding of G-rich DNA oligonucleotides, we employed CD, UV, and NMR spectroscopy, which can together provide detailed insight into stability, molecularity, and topology changes that might appear with the reversal of a primary sequence from 5′→3′ to 3′→5′. The 27 nucleotide (nt) G-rich oligonucleotide s27, which has the sequence d[$G_3TAG_3CAG_4ACACAG_3TAG_3$], is found in the non-coding strand of the long control region of genome of human papillomavirus (HPV) type 52.[12] Since s27 contains five G-rich tracts and has as such high potential to form G-quadruplex structures, it was cut into two shorter oligonucleotides named s23 and s22, which comprise the first to the fourth and the second to the fifth G-tract, respectively (Figure 1 A). The sequence-reversed oligonucleotides r27, r23, and r22 were also prepared.

All six oligonucleotides were titrated with KCl and NaCl (Figures 1 B and 2 show spectra at 50 mM concentration) In

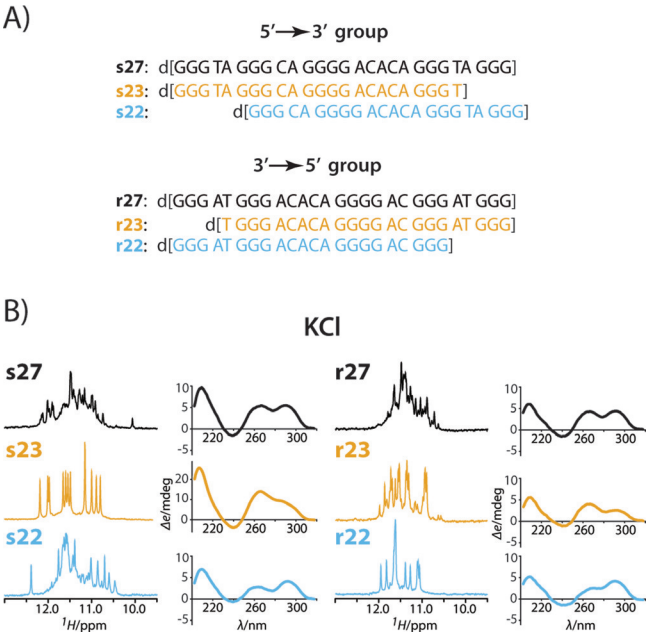

Figure 1. Sequences and spectroscopic characterization. A) Oligonucleotides with sequence in the 5′→3′ and 3′→5′ directions. B) Corresponding NMR and CD spectra recorded at 50 mM KCl.

[*] M. Marušič, Prof. Dr. J. Plavec
Slovenian NMR Center, National Institute of Chemistry
Hajdrihova 19, 1000 Ljubljana (Slovenia)
E-mail: Janez.plavec@ki.si
Prof. Dr. J. Plavec
EN-FIST Center of Excellence, 1000 Ljubljana (Slovenia)
and
Faculty of Chemistry and Chemical Technology
University of Ljubljana, 1000 Ljubljana (Slovenia)

Supporting information for this article (including experimental details) is available on the WWW under http://dx.doi.org/10.1002/anie.201505348.

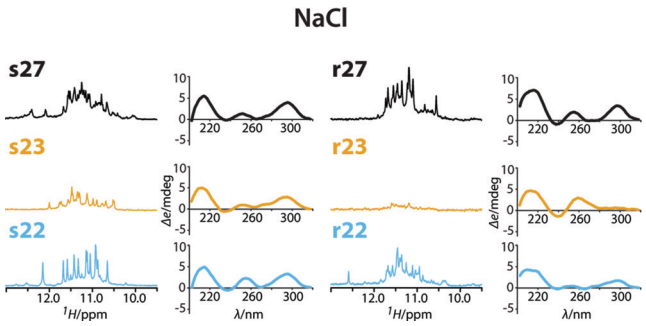

Figure 2. NMR and CD spectra recorded at 50 mM NaCl.

the presence of KCl, s27 and r27 show a high number of signals in the imino region of the proton NMR spectra from $\delta = 10.5$–12.5 ppm, which is characteristic of Hoogsteen-type hydrogen-bonded protons of guanine residues involved in G-quartets. Among the shorter oligonucleotides, s22 and r23 retained the highly polymorphic behavior of their parent oligonucleotides, while s23 and r22 each folded into a single structure with three G-quartets, which could be inferred from 12 partially resolved signals in the imino regions of the proton NMR spectra (Figure 1 B).

The topology of the G-quadruplex adopted by s23 in the presence of K^+ ions is of a peculiar (3+1) kind (Figure 3), as determined with the help of ^{15}N-labelled oligonucleotides and the imino–imino region of the 2D NOESY spectrum (Figure S2). The DNA chain formally starts in the central G-quartet and progresses in a circular manner from residue G1 to G11 via two consecutive edgewise loops (G3–T4–A5 and C9–A10). G11 completes the G1–G2 column of the G-quadruplex core. Consequently, there is no covalent linkage between the two G-quartets involving G11 and G1. The DNA chain progresses from G11 to G12 with a no-residue V-shaped loop that spans three G-quartet planes. The last loop connecting the G12–G14 and G20–G22 G-tracts is a 5-nt ACACA propeller-type loop. The DNA chain ends with T23 stacked on top of the G8–G11–G14–G22 quartet. Interestingly, the oligonucleotide chl1, which has the sequence d[$G_3TG_4A_2G_4TG_3T$], adopts the same G-quadruplex core as s23 (see Figure S3).[13] This suggests that the peculiar (3+1) fold adopted by s23 and chl1 is more robust and displays a wider range of loop lengths than was originally anticipated.

The (3+1) topology of the G-quadruplex adopted by r22 in the presence of K^+ ions has been reported previously and

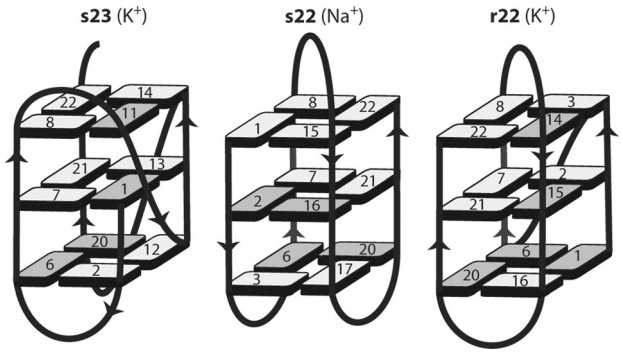

Figure 3. Topologies of s23, s22, and r22. *Syn* residues are shown in grey.

incorporates all three most common types of loops (Figure 3).[14]

The CD spectra for all six oligonucleotides in the presence of K^+ display the characteristic distribution of maxima at 260 and 290 nm and a minimum at 240 nm that is typical of the (3+1) topology (Figure 1 B). This suggests that while reversing the sequence influences the molecularity and topology of the oligonucleotides, they all belong to the (3+1) family of G-quadruplex structures. The relative intensities of the two maxima at 260 and 290 nm are retained pairwise for the original and reversed sequences. The CD spectra for s23 and r23 display a higher intensity maximum at 260 nm compared to the maximum at 290 nm, while in the CD spectra of s22 and r22, the intensity of the maximum at 260 nm is lower compared to that of the maximum at 290 nm. It can be concluded that the characteristics of the G-quadruplex core of s23 are retained in fold of r23, even though r23 folds into more structures, and likewise for s22 and r22.

The folding of s27 and r27 in the presence of NaCl demonstrated the formation of more than one structure (Figure 2). In the case of s23 and even more so for r23, the formation of G-quadruplex structures was hindered. r22 exhibits polymorphic behavior, while s22 forms mostly one G-quadruplex structure with three G-quartet planes. Additional signals of low intensity for s22 indicate the presence of another G-quadruplex structure, the population of which is approximately 30%. The topology of s22 in the presence of Na^+ ions is that of a basket-type G-quadruplex (Figure 3 and Figure S4) with two, six, and two nucleotides along edgewise, diagonal, and edgewise loops, respectively. Unequivocal assignment of the G6–G8 and G20–G22 strands determined their parallel orientation and *syn–anti–anti* distribution of glycosidic torsion angles along the strand. The distribution in the G1–G3 and G15–G17 strands that progress in the opposite direction to G6–G8 and G20–G22 was surprisingly found to be *anti–syn–anti* (see the Supporting Information and Figure S4). Such a distribution of glycosidic torsion angles in the G-quadruplex core differs from the prevailing distribution in antiparallel G-quadruplexes (alternating *syn–anti*). It is therefore clear that the typical CD spectrum of the antiparallel G-quadruplex, with a minimum at 260 nm and maxima at 240 and 290 nm, would not correspond to the CD spectrum of s22. In fact, CD spectrum of s22 displays spectral characteristics of the same (*syn-syn*/*anti-anti*) stacking steps, which results in the appearance of a positive peak at 255 nm and a negative peak at 235 nm (Figure 2). The atypical CD spectra of the other polymorphic oligonucleotides (s27, s23, r27, and r22) in NaCl fit very well with the CD spectrum of s22. It can therefore be concluded that the fold described for s22 is conserved in the other oligonucleotides in the presence of Na^+ ions. The only exception is r23, which displays a CD spectrum typical of a parallel G-quadruplex, with a minimum at 240 and a maximum at 260 nm.

Clear differences between the CD profiles in K^+- and Na^+-rich environments have demonstrated that—at least for our set of oligonucleotides—the topology of the G-quadruplex core is in fact determined by the nature of the stabilizing cation and not by the sequence directionality itself. Indeed, preference for an antiparallel fold in NaCl and parallel/(3+1)

Table 1: Melting temperatures of G-quadruplexes as measured by UV absorption.

	s27	s23	s22	r27	r23	r22
T_m (K^+)/°C	58	52	60	54	45	59
T_m (Na^+)/°C	48	39	49	41	36	43

topologies in KCl was previously noted[15–17] and becomes quite apparent after re-examination of the literature.[18,19]

The melting temperature (T_m) values for the six oligonucleotides were between 45 and 60 °C in the presence of K^+ and between 36 and 48 °C in the presence of Na^+ (Table 1). No hysteresis in the melting curves was observed. Interestingly, melting temperatures in the 5′→3′ group are higher than in 3′→5′ group in both KCl and NaCl. The highest thermal stability within the 5′→3′ group could be assigned to the shortest oligonucleotide s22, followed by the longest s27, while the least stable was s23. The same trend in thermal stability was observed in the 3′→5′ group. This curious differential thermal stability is reflected in the folding behavior of the oligonucleotides (Figure 4). The concentration of K^+ needed for complete folding of the G-quadruplex is lower than the concentration of Na^+, and this is true for both the 5′→3′ and 3′→5′ groups. The former folds at lower concentrations of cations regardless of the cation type. The length of the oligonucleotide is related to the cation concentration needed for folding in the 3′→5′ group. The shortest oligonucleotide r22 exhibits the highest cation affinity, followed by the longest r27, while the lowest cation affinity was observed for r23.

Closer examination of the non-palindromic loops shows that the 5′→3′ group possess TA and CA loops, while the 3′-5′ group possess AT and AC loops. An adenine at the first position in the loop was shown to have a substantial destabilizing effect,[15,18] which would explain why oligonucleotides with sequences in 5′→3′ direction display up to 7 °C

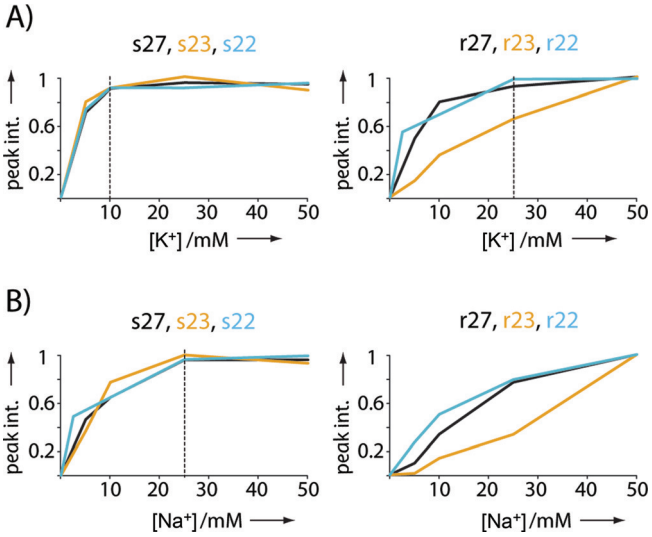

Figure 4. Folding curves in the presence of K^+ (A) and Na^+ (B). Vertical dashed lines represent concentration of cations at which folding was complete for most of the oligonucleotides. Peak int. = Integral of resolved proton NMR resonance normalized with respect to the value corresponding to the folded G-quadruplex.

higher T_m values in comparison to their 3′→5′ counterparts. The highest thermal stability for the shortest oligonucleotides s22 or r22 presumably indicates a preference for a longer central loop, but not the first and the third,[17,20] since s22 and r22 both have ACACA in the central position, with potential to adopt a diagonal orientation. For s27 and r27, the second most stable oligonucleotides, the ACACA loop retains the potential to form a long central loop. s23 and r23, however, have the longest ACACA loop as the third and the first loop, respectively, and exhibit the least stable G-quadruplexes of the group (see the Supporting Information for further discussion).

Our novel insights, together with recent literature data from studies that were done on a wide range of sequences, can be used to derive a collection of six semi-empirical rules that can offer an explanation for the folding preferences of the oligonucleotides in the current study and other known G-rich sequences. First, parallel or (3+1) topologies are preferred in KCl and antiparallel topologies in NaCl solutions.[15,18,19,16,21] Second, *anti-anti* and *syn-anti* steps contribute to higher stability in comparison to *anti-syn* and *syn-syn* steps.[22] Third, *syn* conformation at the 5′-end of oligonucleotide is preferred.[22,23] Fourth, an adenine residue is located preferentially at the last[24] but not the first[18] position of the loop. Fifth, long central and shorter first and third loops are preferred.[17,20] Sixth, the typical lengths for propeller, edgewise, and diagonal loops are one, three and four or more nucleotides, respectively.[25]

On the basis of the above generalized rules, the folding of r22 (d[$G_3ATG_3ACACAG_4ACG_3$]) into the (3+1) topology can be rationalized by its longer loops and the presence of K^+ ions. One guanine residue in the longer G_4 tract will participate in the third edgewise and not the second diagonal loop, since the second ACACA loop already has the length that is preferred for the diagonal loop. Simultaneously, this places the adenine residue in the preferred second position in the third GAC loop. The loop lengths of two, five, and three residues in the fold of r22 are close to ideal for propeller, diagonal, and edgewise loops, respectively. G1 adopts a *syn* conformation across the glycosidic bond as the first residue at the 5′ end.

Similarly, for s22 (d[$G_3CAG_4ACACAG_3TAG_3$]), antiparallel topology would be expected in the presence of Na^+ ions. In this case, the G_4 tract donates its last guanine residue (G9) to the second loop, thereby preventing an adenine residue from occupying the first position of the loop. An antiparallel fold with loop lengths of two, six, and two residues in the fold of s22 can be achieved with a propeller–diagonal–propeller or edgewise–diagonal–edgewise combination of loop types. Several structures in the presence of Na^+ ions have shown a preference for edgewise over propeller-type loops,[26–29] presumably owing to the potential interaction of Na^+ cations in the outer G-quartets with the loop residues.

With the above six rules it is, however, quite difficult to rationalize the fold of s23 (d[$G_3TAG_3CAG_4ACACAG_3T$]) in the presence of K^+ ions. Nevertheless, topology of s23 clearly shows certain structural elements that have the potential to form but have escaped broad recognition owing to the low number of reported structures. The so-called GNA loop

(where N can be any nucleotide), for example, that can form in the G3–T4–A5 tract of s23, has been known for its considerable stability for 20 years.[30–33] The NAG_4 (C10–A11–G_4 in the case of s23) element comprising the third G-rich tract of the G-rich sequences can be recognized as crucial for the formation of the V-loop both in s23 and chl1[13] (Figure 3 and Figure S3). Moreover, several G-quadruplexes structures have indicated that oligonucleotides with uneven lengths of G-tracts exhibit an inclination for long G-tracts to be inserted into the G-quadruplex core as a whole.[13,34,35]

Oligonucleotides with reversed sequences do not represent a special case within biological systems. They are therefore interesting if found in different (parts of) genomes. It is, however, helpful to understand that these inverted sequence counterparts represent molecules with completely different folding preferences. Moreover, the current understanding of the role that G-quadruplexes might play in biological systems implies that there exists evolutionary pressure for the selection of G-rich sequences with the potential to form more stable G-quadruplex structures.[20,36,37] One would therefore expect to find higher number of cases of thermally more stable oligonucleotides in the 5′→3′ group than the in 3′→5′ group. However, since in vivo folding process might crucially depend on the order in which parts of G-rich sequence become available for G-quadruplex formation (during double-strand DNA melting or unwinding), the folding preferences of unrestrained oligonucleotides in solution might be very different. Additionally, the search for G-rich oligonucleotides that form only one structure may be exaggerated, since G-quadruplexes that adopt different structures might act as switches that can bind different protein partners.

Acknowledgements

Financial support from the Slovenian Research Agency (ARRS, grants P1-242 and J1-6733) is gratefully acknowledged.

Keywords: circular dichroism · DNA structures · folding · G-quadruplexes · NMR spectroscopy

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 11716–11719
Angew. Chem. **2015**, *127*, 11882–11885

[1] C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H. J. Lipps, A. Plückthun, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 8572–8577.
[2] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, L. H. Hurley, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11593–11598.
[3] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, *Nat. Chem.* **2013**, *5*, 182–186.
[4] A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle, D. J. Patel, *J. Am. Chem. Soc.* **2007**, *129*, 4386–4392.
[5] T. Maehigashi, C. Hsiao, K. K. Woods, T. Moulaei, N. V. Hud, L. D. Williams, *Nucleic Acids Res.* **2012**, *40*, 3714–3722.
[6] I. Mamajanov, A. E. Engelhart, H. D. Bean, N. V. Hud, *Angew. Chem. Int. Ed.* **2010**, *49*, 6310–6314; *Angew. Chem.* **2010**, *122*, 6454–6458.
[7] D. A. Rusling, V. E. C. Powers, R. T. Ranasinghe, Y. Wang, S. D. Osborne, T. Brown, K. R. Fox, *Nucleic Acids Res.* **2005**, *33*, 3025–3032.
[8] N. T. Thuong, H. Claude, *Angew. Chem. Int. Ed.* **1993**, *32*, 666–690; *Angew. Chem.* **1993**, *105*, 697–723.
[9] R. D. Wells, J. E. Larson, R. C. Grant, B. E. Shortle, C. R. Cantor, *J. Mol. Biol.* **1970**, *54*, 465–497.
[10] K. J. Breslauer, R. Frank, H. Blöcker, L. A. Marky, *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 3746–3750.
[11] R. A. Friedman, B. Honig, *Biophys. J.* **1995**, *69*, 1528–1535.
[12] K. Tlučková, M. Marušič, P. Tóthová, L. Bauer, P. Šket, J. Plavec, V. Viglasky, *Biochemistry* **2013**, *52*, 7207–7216.
[13] V. Kuryavyi, D. J. Patel, *Structure* **2010**, *18*, 73–82.
[14] M. Marusic, P. Sket, L. Bauer, V. Viglasky, J. Plavec, *Nucleic Acids Res.* **2012**, *40*, 6946–6956.
[15] A. Guédin, A. De Cian, J. Gros, L. Lacroix, J.-L. Mergny, *Biochimie* **2008**, *90*, 686–696.
[16] N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E. S. Baker, M. T. Bowers, E. De Pauw, V. Gabelica, *J. Am. Chem. Soc.* **2008**, *130*, 10208–10216.
[17] P. A. Rachwal, I. S. Findlow, J. M. Werner, T. Brown, K. R. Fox, *Nucleic Acids Res.* **2007**, *35*, 4214–4222.
[18] A. Guédin, P. Alberti, J.-L. Mergny, *Nucleic Acids Res.* **2009**, *37*, 5559–5567.
[19] A. Guédin, J. Gros, P. Alberti, J.-L. Mergny, *Nucleic Acids Res.* **2010**, *38*, 7858–7868.
[20] J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2005**, *33*, 2908–2916.
[21] P. A. Rachwal, T. Brown, K. R. Fox, *Biochemistry* **2007**, *46*, 3036–3044.
[22] J. Šponer, A. Mládek, N. Špačková, X. Cang, T. E. Cheatham, S. Grimme, *J. Am. Chem. Soc.* **2013**, *135*, 9785–9796.
[23] X. Cang, J. Šponer, T. E. Cheatham, *Nucleic Acids Res.* **2011**, *39*, 4499–4512.
[24] X. Cang, J. Šponer, T. E. Cheatham, *J. Am. Chem. Soc.* **2011**, *133*, 14270–14279.
[25] M. W. da Silva, M. Trajkovski, Y. Sannohe, N. M. Hessari, H. Sugiyama, J. Plavec, *Angew. Chem. Int. Ed.* **2009**, *48*, 9167–9170; *Angew. Chem.* **2009**, *121*, 9331–9334.
[26] P. Balagurumoorthy, S. K. Brahmachari, D. Mohanty, M. Bansal, V. Sasisekharan, *Nucleic Acids Res.* **1992**, *20*, 4061–4067.
[27] Y. Wang, D. J. Patel, *Structure* **1993**, *1*, 263–282.
[28] P. Balagurumoorthy, S. K. Brahmachari, *J. Biol. Chem.* **1994**, *269*, 21858–21869.
[29] N. Zhang, A. T. Phan, D. J. Patel, *J. Am. Chem. Soc.* **2005**, *127*, 17277–17285.
[30] L. Zhu, S. H. Chou, J. Xu, B. R. Reid, *Nat. Struct. Biol.* **1995**, *2*, 1012–1017.
[31] I. Hirao, G. Kawai, S. Yoshizawa, Y. Nishimura, Y. Ishido, K. Watanabe, K. Miura, *Nucleic Acids Res.* **1994**, *22*, 576–582.
[32] A. Kettani, A. Gorin, A. Majumdar, T. Hermann, E. Skripkin, H. Zhao, R. Jones, D. J. Patel, *J. Mol. Biol.* **2000**, *297*, 627–644.
[33] A. Matsugami, K. Ouhashi, M. Kanagawa, H. Liu, S. Kanagawa, S. Uesugi, M. Katahira, *J. Mol. Biol.* **2001**, *313*, 255–269.
[34] M. Marušič, R. N. Veedu, J. Wengel, J. Plavec, *Nucleic Acids Res.* **2013**, *41*, 9524–9536.
[35] M. Crnugelj, P. Sket, J. Plavec, *J. Am. Chem. Soc.* **2003**, *125*, 7866–7871.
[36] A. Risitano, K. R. Fox, *Biochemistry* **2003**, *42*, 6507–6513.
[37] A. K. Todd, M. Johnston, S. Neidle, *Nucleic Acids Res.* **2005**, *33*, 2901–2907.

Received: June 11, 2015
Revised: July 20, 2015
Published online: August 12, 2015

