

Polymorphism and resolution of oncogene promoter quadruplex-forming sequences†‡

M. Clarke Miller, Huy T. Le, William L. Dean, Patrick A. Holt, Jonathan B. Chaires and John O. Trent*

Received 3rd June 2011, Accepted 8th September 2011

DOI: 10.1039/c1ob05891f

We report the separation of several quadruplex species formed by ten promoter sequences by Size Exclusion Chromatography (SEC). Modification at the 5' or 3' ends or in loop regions of quadruplex forming sequences has become the standard technique for dealing with quadruplex polymorphism. However, conformations produced employing this method or by other means of artificially shifting the equilibrium may not represent the species that are present *in vivo*. This method enables an unperturbed view of the structural polymorphism inherent to quadruplex formation. Separation *via* SEC facilitates studies on quadruplex structure and biophysical properties without the need for sequence modification.

G-quadruplexes formed from a particular DNA sequence in solution are usually polymorphic.^{1,2} A common technique for simplifying the ensemble of species to enable structural analysis is to make sequence modifications.^{3,4} The presumption is that such alterations merely perturb the finely balanced equilibrium, which is in fact untested, and therefore structures produced in this manner may not represent the species that are present initially or *in vivo*. We have used Size Exclusion Chromatography (SEC) to analyze and separate the ensemble of quadruplex species formed by several oncogene promoter sequences. Separation *via* SEC facilitates studies on quadruplex structure without sequence modification or structural perturbation while potentially accessing a greater quadruplex folding space.

Guanine-rich DNA with at least four runs of two guanines can form three-dimensional structures called G-quadruplexes in the presence of certain monovalent cations such as Na⁺, K⁺, or NH₄⁺. G-quadruplexes are made up of stacks of two or more square planer arrays of four guanines (a G-quartet) comprised

of Hoogsten hydrogen bonds with coordination of a cation to the O6 of the guanines.⁵ Although most often associated with telomeres, potential quadruplex-forming sequences have been found throughout the genome⁶ including in the promoter regions of many proto-oncogenes such as c-myc,⁷ c-kit,^{8,9} bcl-2,¹⁰ VEGF,¹¹ and HIF-1 α ,¹² and in the 5'-untranslated regions of mRNA of proto-oncogenes.¹³ Quadruplex DNA is of interest because of possible roles in biological regulation,¹⁴ bio/nanotechnology applications,¹⁵ and as potential therapeutic drug targets^{15,16} or as therapeutic agents^{17,18} themselves.

Polymorphism is inherent in quadruplex formation for most quadruplex-forming sequences. When factors such as strand orientation, loop type and arrangement, glycosyl torsion angles, intramolecular *versus* intermolecular formation, ion type and concentration, DNA concentration, the presence of organic solvents and various biological molecules, and annealing profile are considered, even the relatively simple human telomeric sequence d(GGGTTA)₄ can theoretically fold into more than 200 intramolecular conformations.^{1,19,20} This intrinsic polymorphism is an issue for those who wish to investigate quadruplex structure or the thermodynamics of quadruplex formation. Structural and biophysical studies of quadruplex DNA have been hindered by our inability to isolate and study the individual quadruplex species that may occur in solution. Low resolution methods such as circular dichroism, UV-vis spectrophotometry, analytical ultracentrifugation, gel electrophoresis, or calorimetry may not be able to differentiate among conformations with similar physical properties and high resolution techniques such as NMR spectroscopy are limited for examining these complex mixtures.²

Polymorphism has been generally dealt with by altering the sequence of interest to produce an enriched configuration for study. This usually requires either adding or deleting bases at the 5' or 3' ends of the sequence or lengthening, shortening, or eliminating guanines from putative loop regions to ensure that a single fold is enriched.^{3,4} Substitution of 8-aminoguanine promotes formation of tetramolecular parallel quadruplexes such as those formed by TG₄T,²¹ incorporation of 8-methylguanine or 8-bromoguanine is known to produce quadruplex structures with a *syn* glycosidic configuration,²²⁻²⁴ while use of O⁶-methylguanine, inosine, or 6-thioguanine has been shown to destabilize quadruplex formation.²⁵⁻²⁸ Modification of the sugar phosphate backbone by insertion of 5'-5' or 3'-3' polarity inversion has also been shown to have a remarkable effect on quadruplex formation and

Department of Medicine, Clinical Translational Research Building, University of Louisville, 505 Hancock St., Louisville, KY 40202, USA. E-mail: john.trent@louisville.edu; Fax: +1-502-852-7979; Tel: +1-502-852-2194

† This work was supported by National Institutes of Health (CA113735-01 to J.O.T), National Institutes of Health Grant Number P20RR018733 from the NCRR, The JG Brown Foundation, and the Kentucky Challenge for Excellence. The authors would like to thank Dr Robert Gray for useful discussions and advice. The authors would also like to thank Dr Andrew N. Lane for assistance with NMR data collection and Robert Buscaglia for assistance with CD data collection.

‡ Electronic supplementary information (ESI) available: Methods, figures, and tables are available. See DOI: 10.1039/c1ob05891f

stability.²⁹ RNA and LNA force adoption of a *syn* glycosidic guanosine conformation.^{30–33}

Reduction of quadruplex polymorphism is possible using methods other than sequence modification. Solvent conditions can greatly affect the range of configurations formed. Choice of and concentration of monovalent cation, primarily K⁺ or Na⁺, or inclusion of divalent cations, such as Mn²⁺, Co²⁺, Ni²⁺, Mg²⁺, Pb²⁺, or Sr²⁺, can stabilize or destabilize G-quadruplex structures. Potassium concentration has been shown to be essential in selecting which quadruplex conformation is formed by the human telomere sequence,^{34,35} and inclusion of divalent cations has been shown to induce a transition from antiparallel to parallel G-quadruplex structure for the G₄T₄G₄ sequence.^{36,37} Finally, addition of organic or biological molecules such as PEG, acetonitrile, proteins, or polysaccharides can contribute greatly to quadruplex formation and stability.^{38,39}

In most cases a putative quadruplex forming sequence undergoes several modifications and/or truncations before producing a supposedly single species in solution. For example, the bcl-2 promoter parent thirty nine base sequence has six runs of three or more guanines and was truncated to a twenty three base sequence centred on the four central sets of guanine. This twenty three base sequence was chosen from a set of three possible sequences.^{10,40,41} The selection process included the use of CD data which has been shown to be unreliable for unambiguous determination of structural details of quadruplex configuration.⁴² The sequence was further modified by substitution of thymine for two guanines, thereby forcing a loop, to produce a final sequence for study somewhat removed from the original native sequence potential folding topology. This is not unique and is a common practice as further evidenced by the sequences used to produce the solution structures of c-kit and c-myc. Production of a c-kit sequence suitable for solution structure determination took a series of five sequence substitutions for the 2KJ2 structure and a series of twelve sequence substitutions for the 2O3M structure. In the case of c-myc, the parent twenty seven base sequence containing six consecutive runs of guanines, five of which are comprised of three or more guanines, was shown to form a complex mixture of quadruplexes by NMR.⁴³ The efforts to reduce and simplify the spectra for structural study produced three different structures from two different truncated sequences selected from different areas of the parent sequence and utilized two different thymine/guanine substitutions.^{7,43} Another difficulty is the reproducibility of reported data as these sequences can be exquisitely sensitive to solvation and annealing conditions. The

substitution of two thymines in the sequence used to produce the c-myc 1XAV solution structure was stated as required as the truncated sequence which produced a suitable 1D ¹H NMR spectrum for study by one research group⁴³ produced a complex spectra when examined by a second research group.⁷

While these methods can yield a less complex ensemble of species and have been used to produce many NMR and crystal structures, the relevance of these results is uncertain as they represent only a small number of the species possible in solution. If the array of quadruplex conformations is at or near equilibrium then sequence modification as a method for perturbing this equilibrium could lead to unreliable and unpredictable results. Therefore, the results from such modifications do not necessarily reflect the diversity of species actually present in solution for the parent sequence, and the modified sequences likely represent a further removal from conditions found *in vivo*.¹

We report the separation profile of several quadruplex species formed by ten proto-oncogene promoter sequences by Size Exclusion Chromatography. The set of sequences was chosen because they have been previously studied and have either a deposited structure or a reported topology (Table 1). Each of the chosen sequences showed high polymorphism with variable degrees of separation (Fig. 1). These results agree with AUC data for each of the sequences (Supplementary Figure S1†) which indicates that each promoter sequence yields a mixture of conformations. It is clear that the modification of the parent sequence changes the quadruplex ensemble. This is consistent with previous reports for the human telomere³ and is dramatically shown for the two c-kit promoter studies that used significantly different sequences (Table 1) and obtained completely different structures.^{8,9} The SEC method shows that for these two sequences the distribution of species is indeed changed (Fig. 1A,B). More subtle sequence modifications, such as the inclusion of flanking bases as in the case of the HIF-1 α promoter (Table 1), also have significant perturbation of the quadruplex populations as seen by SEC (Fig. 1C,D). The remainder of the oncogene promoter sequences in Table 1 display significant polymorphism by SEC (Fig. 1E–J). In some cases, such as with the KRAS sequence (Fig. 1J), the chromatogram may also be indicative of higher order species such as G-wire. Fig. 1 shows that all of the oncogene promoter sequences have a wide distribution of species that are partially resolved by SEC. Many of the sequences show 6–8 peaks *via* SEC. Some of the individual peaks appear to be broadened. These broad peaks could be a combination of several conformations that elute at similar elution volumes. Broadened peak shapes may also signify

Table 1 Quadruplex-forming sequences from various proto-oncogene promoter sequences utilized in this study

Sequence name	Sequence	Topology or PDB ID	Reference
c-kit 2KJ2	CGGGCGGGCGCGAGGGAGGGT	2KJ2	8, 44
c-kit 2O3M	AGGGAGGGCGCTGGGAGGAGGG	2O3M	9
HIF-1 α with flanking bases	GCGCGGGAGGGGAGAGGGGGCGGGAGCGCG	all-parallel propeller	12
HIF-1 α without flanking bases	GGGGAGGGGAGAGGGGGCGGGA	all-parallel propeller	12
Retinoblastoma	CGGGGGGTTTTGGGCGGC	anti-parallel basket	45
c-myc	TGGGGAGGGTGGGGAGGGTGGGGGAAGG	1XAV	7
Her2	AGGAGAAGGAGGAGGTGGAGGAGGAGGGC	putative quadruplex	46
bcl-2	AGGGCGGGCGCGGGAGGAAGGGGGCGGGAGCGGGGC	2F8U	10
VEGF	GGGCGGGCCGGGGCGGGGTCCCGGCGGGGCGGGAG	all-parallel propeller	11
KRAS	GGGAAGAGGGAAGAGGGGGAGG	all-parallel propeller	47

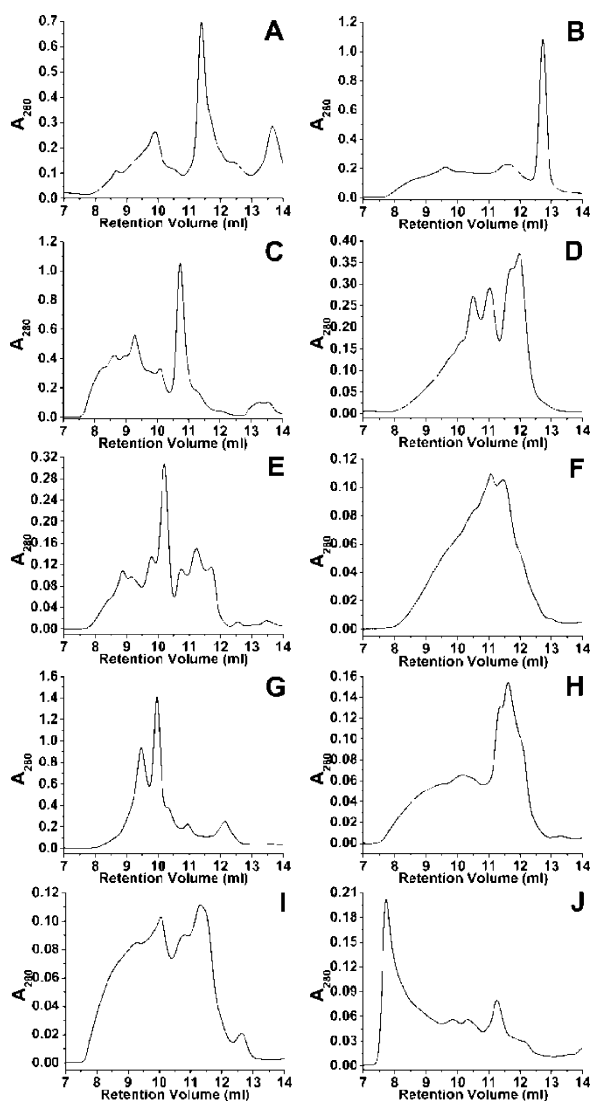


Fig. 1 SEC separations for the promoter sequences. A. c-kit 2KJ2, B. c-kit 2O3M, C. HIF-1 α , D. HIF-1 α without flanking sequences, E. Retinoblastoma, F. c-myc, G. Her2, H. bcl-2, I. VEGF, and J. KRAS.

rapid inter-conversion or re-equilibration of quadruplex species. If the distribution of conformations is near equilibrium then such inter-conversions are possible on the HPLC time scale. In such a case, the possibility of population perturbation by the act of separation must also be considered.

The c-kit 2KJ2 sequence was chosen for further study since this sequence has ample NMR data available.^{8,44} The sequence was annealed and separated *via* SEC and fractions corresponding to the major components were collected. Material from several HPLC runs were combined. The 1D ¹H NMR spectra of the combined fractions and of the un-separated material were recorded (see Supplementary Materials for Methods). The un-separated material produced an NMR spectrum indicative of a mixture of several quadruplex conformations. The spectrum displays many overlapping imino proton resonances with chemical shifts characteristic of quadruplex formation in K⁺ buffer^{48,49} (Fig. 2A). AUC for the parent mixture (Fig. 3A) supports the presence of multiple species in solution. The first purified fraction, corresponding to a retention volume of 9.90 ml, also displayed a

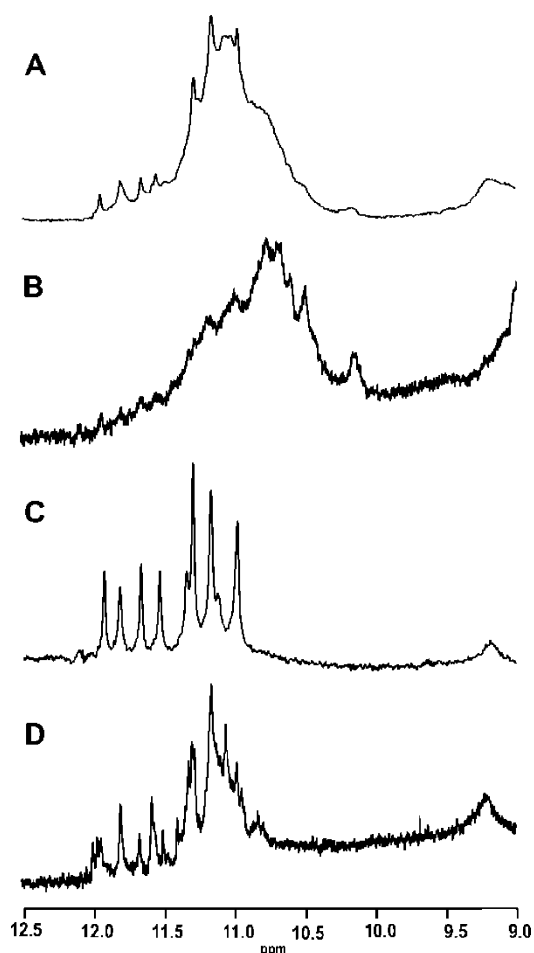


Fig. 2 NMR of the c-kit 2KJ2 sequence at 293 K. A. The spectrum of the parent c-kit sequence showing an overlapping set of GN1H resonances in the quadruplex imino/amino proton region. B. The NMR spectrum of the fraction 1. C. The NMR spectrum of fraction 2. D. The NMR spectrum of fraction 3.

broad envelope of overlapping imino proton signals (Fig. 2B). This result could be indicative of either a complex mixture unresolvable by SEC or of a structure consisting of multiple strands of DNA such as a G-wire. The AUC data for this fraction supports the presence of a species that is at least a tetramer (Fig. 3B). In contrast, the 1D ¹H NMR data for the purified fraction 2, corresponding to a retention volume of 11.38, indicates only one quadruplex species in solution (Fig. 2C). AUC data for this fraction indicates that fraction 2 is at least a dimeric species of quadruplex (Fig. 3C). Finally, the third fraction collected yields a complex mixture of imino signals (Fig. 2D). AUC data provides support that fraction 3 is a mixture of components (Fig. 3D). The 1D ¹H NMR spectrum of the parent sequence and each fraction displays GN1H resonances between 12 to 10 ppm. The occurrence of these imino/amino resonances is indicative of Hoogsteen hydrogen bonding characteristic of the presence of a G-quadruplex.^{48,49}

Re-injection of each of the NMR samples shows that fraction 1 and fraction 2 matched the original elution profile, whereas fraction 3 showed significant re-equilibration (Supplementary Figure S2 \ddagger). UV-vis and CD data for each of the fractions is similar (Supplementary Figure S3 and S4 respectively \ddagger) and therefore

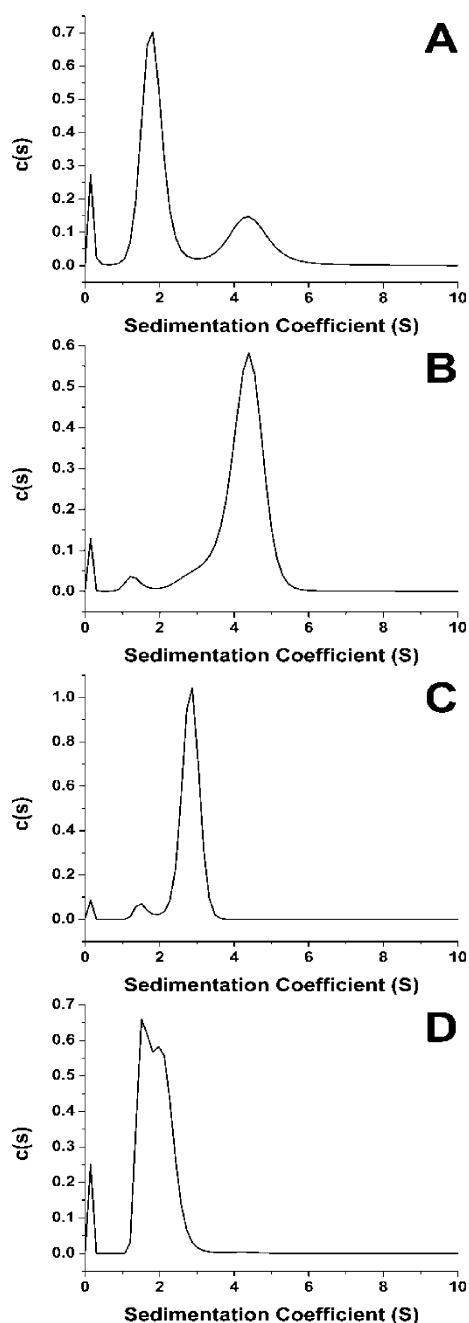


Fig. 3 AUC of the unseparated c-kit 2KJ2 sequence (A) and of each of the major fractions presented in Fig. 2(B–D).

little use for indicating different species. Interestingly, none of the purified species or un-separated mixture matches the previously reported NMR data for this sequence.^{8,44} In fact, the major species formed by this sequence under the conditions of this study is a dimer which is at odds with the previously reported data. Clearly formation of a particular quadruplex configuration is dependent not only on the sequence but also several other factors, including buffer conditions and annealing profile.

The data presented here, exemplified by the chromatograms of the c-kit sequences, indicates that these modified sequences are capable of forming far more than the single species reported in the literature. This disparity has been observed previously with the c-

myc sequence.^{7,43} This discrepancy could be due to differences in annealing conditions producing an altered equilibrium of quadruplex conformations. The disparity in results could also be attributed to the ¹³C, ¹⁵N labelling method commonly used for quadruplex structural determination. At the low enrichment levels used for unambiguous assignment of residues, typically 3–6% or less,^{7,10,41,50,51} conformations with a low relative abundance would simply be lost in the baseline noise of the spectra. In other cases, such as in the determination of the topology of the Retinoblastoma sequence (Fig. 1E), researchers depended on low resolution methods to obtain their results.⁴⁵ The results in Fig. 1 and S1 clearly show multiple species in solution instead of a single structure.

When compared to the protein calibration curves (Supplementary Materials Figure S5†) results from the separation of the c-kit quadruplex sequence (Fig. 1A), at 6634.3 g mol⁻¹ with an estimated Stokes Radius (R_s) of approximately 16 Å, show that the species isolated for further study elutes at an apparent molecular weight of 28–30 kDa and R_s of approximately 26 Å. A facile answer to this observation would be the formation of a tetramer species, yet this material is shown to be a single, dimeric species *via* ¹H NMR (Fig. 2C) and AUC (Fig. 3C). The Retinoblastoma sequence (Fig. 1E) demonstrates a series of putative quadruplex species with apparent R_s ranging from 25 Å to over 40 Å. Comparison to the elution profile of several human telomere sequences illustrates very different behaviour (Supplementary Materials Figure S6 and Table S1†). Experimental values for the R_s of these sequences also somewhat disagree with the calculated values (Supplementary Materials Table S2), however, these sequences elute in a much more consistent manner than do the oncogene promoter sequences.

There are several possible explanations for these results. First, there is the basic difference between telomeric sequences and promoter derived sequences. Telomeric sequences form structures with only one of three types of loop, double chain reversal, chair, and diagonal, and no matter how the G-quartets are stacked the result is always a very similar compact structure. More irregular sequences, typical of promoter sequences, are capable of more diverse stacking, even formation of alternate G-quartets, and greater differences in loop configuration. However, it seems unlikely that these differences in configuration could result in the separation demonstrated here and yet be undetectable by other biophysical means. Second, proteins can undergo changes in stokes radius due to buffer dependent changes in configuration.⁵² This is unlikely for quadruplexes because of their exceptional stability. Third, protein samples can experience associative and repulsive interactions with column packing material.^{53,54} Some negatively charged species may experience exclusion by repulsive interaction with the column matrix. The results from the promoter sequences may be from differences in these interactions based on the differences in topology.

This approach facilitates the isolation of the actual species formed by a parent sequence without the need for extensive sequence modification with possible structural perturbation. An added benefit is that purification may yield more than one isolatable species providing the opportunity for simultaneous study of several structures. The separation technique could be used for a variety of applications including identification of protein/quadruplex interactions or drug/quadruplex binding studies from the parent mixtures. This method may not be

suitable for every quadruplex forming system. If the distribution of conformations at or near equilibrium for a specific quadruplex mixture then inter-conversions are possible on the HPLC time scale. As shown for 2KJ2 fraction 3, isolated fractions may also undergo re-equilibration and lack the stability required for further study. As such, each of these variables should be investigated thoroughly before incorporating SEC into procedures for the preparation of quadruplex DNA.

Notes and references

- 1 A. N. Lane, J. B. Chaires, R. D. Gray and J. O. Trent, *Nucleic Acids Res.*, 2008, **36**, 5482–5515.
- 2 M. M. Dailey, M. C. Miller, P. J. Bates, A. N. Lane and J. O. Trent, *Nucleic Acids Res.*, 2010, **38**, 4877–4888.
- 3 J. Dai, M. Carver and D. Yang, *Biochimie*, 2008, **90**, 1172–1183.
- 4 D. Yang and K. Okamoto, *Future Med. Chem.*, 2010, **2**, 619–646.
- 5 J. R. Williamson, *Annu. Rev. Biophys. Biomol. Struct.*, 1994, **23**, 703–730.
- 6 J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2005, **33**, 2908–2916.
- 7 A. Ambrus, D. Chen, J. X. Dai, R. A. Jones and D. Z. Yang, *Biochemistry*, 2005, **44**, 2048–2058.
- 8 S. T. D. Hsu, P. Varnai, A. Bugaut, A. P. Reszka, S. Neidle and S. Balasubramanian, *J. Am. Chem. Soc.*, 2009, **131**, 13399–13409.
- 9 A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle and D. J. Patel, *J. Am. Chem. Soc.*, 2007, **129**, 4386–4392.
- 10 J. Dai, D. Chen, R. A. Jones, L. H. Hurley and D. Yang, *Nucleic Acids Res.*, 2006, **34**, 5133–5144.
- 11 D. Y. Sun, K. X. Guo, J. J. Rusche and L. H. Hurley, *Nucleic Acids Res.*, 2005, **33**, 6070–6080.
- 12 R. De Armond, S. Wood, D. Y. Sun, L. H. Hurley and S. W. Ebbinghaus, *Biochemistry*, 2005, **44**, 16341–16350.
- 13 D. J. Patel, A. T. Phan and V. Kuryavyi, *Nucleic Acids Res.*, 2007, **35**, 7429–7455.
- 14 N. Maizels, *Nat. Struct. Mol. Biol.*, 2006, **13**, 1055–1059.
- 15 J. L. Huppert, *Philos. Trans. R. Soc. London, Ser. A*, 2007, **365**, 2969–2984.
- 16 L. H. Hurley and A. Siddiqui-Jain, *Genet. Eng. News*, 2005, **25**, 26–26.
- 17 S. Cogoi, F. Quadrifoglio and L. E. Xodo, *Biochemistry*, 2004, **43**, 2512–2523.
- 18 C. R. Ireson and L. R. Kelland, *Mol. Cancer Ther.*, 2006, **5**, 2957–2962.
- 19 J. Li, J. J. Correia, L. Wang, J. O. Trent and J. B. Chaires, *Nucleic Acids Res.*, 2005, **33**, 4649–4659.
- 20 M. Webba da Silva, *Chem.–Eur. J.*, 2007, **13**, 9738–9745.
- 21 J. Gros, A. Avino, J. L. de la Osa, C. Gonzalez, L. Lacroix, A. Perez, M. Orozco, R. Eritja and J. L. Mergny, *Chem. Commun.*, 2008, 2926–2928.
- 22 A. Virgilio, V. Esposito, A. Randazzo, L. Mayol and A. Galeone, *Nucleic Acids Res.*, 2005, **33**, 6188–6195.
- 23 A. Virgilio, V. Esposito, A. Randazzo, L. Mayol and A. Galeone, *Bioorg. Med. Chem.*, 2005, **13**, 1037–1044.
- 24 V. Esposito, A. Randazzo, G. Piccialli, L. Petraccone, C. Giancola and L. Mayol, *Org. Biomol. Chem.*, 2004, **2**, 313–318.
- 25 C. S. Mekmaysy, L. Petraccone, N. C. Garbett, P. A. Ragazzon, R. Gray, J. O. Trent and J. B. Chaires, *J. Am. Chem. Soc.*, 2008, **130**, 6710–6711.
- 26 A. G. Petrovic and P. L. Polavarapu, *J. Phys. Chem. B*, 2008, **112**, 2255–2260.
- 27 V. M. Marathias, M. J. Sawicki and P. H. Bolton, *Nucleic Acids Res.*, 1999, **27**, 2860–2867.
- 28 N. Spackova, E. Cubero, J. Sponer and M. Orozco, *J. Am. Chem. Soc.*, 2004, **126**, 14642–14650.
- 29 V. Esposito, A. Virgilio, A. Randazzo, A. Galeone and L. Mayol, *Chem. Commun.*, 2005, 3953–3955.
- 30 L. Bonifacio, F. C. Church and M. B. Jarstfer, *Int. J. Mol. Sci.*, 2008, **9**, 422–433.
- 31 N. Kumar and S. Maiti, *J. Phys. Chem. B*, 2007, **111**, 12328–12337.
- 32 C. F. Tang and R. H. Shafer, *J. Am. Chem. Soc.*, 2006, **128**, 5966–5973.
- 33 J. Qi and R. H. Shafer, *Biochemistry*, 2007, **46**, 7599–7606.
- 34 R. D. Gray, J. Li and J. B. Chaires, *J. Phys. Chem. B*, 2009, **113**, 2676–2683.
- 35 R. D. Gray, L. Petraccone, J. O. Trent and J. B. Chaires, *Biochemistry*, 2010, **49**, 179–194.
- 36 S. W. Blume, V. Guarcello, W. Zacharias and D. M. Miller, *Nucleic Acids Res.*, 1997, **25**, 617–625.
- 37 D. Miyoshi, A. Nakao and N. Sugimoto, *Nucleic Acids Res Suppl*, 2001, 259–260.
- 38 Y. Xue, Z. Y. Kan, Q. Wang, Y. Yao, J. Liu, Y. H. Hao and Z. Tan, *J. Am. Chem. Soc.*, 2007, **129**, 11185–11191.
- 39 M. C. Miller, R. Buscaglia, J. B. Chaires, A. N. Lane and J. O. Trent, *J. Am. Chem. Soc.*, 2010, **132**, 17105–17107.
- 40 T. S. Dexheimer, D. Sun and L. H. Hurley, *J. Am. Chem. Soc.*, 2006, **128**, 5404–5415.
- 41 J. Dai, T. S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R. A. Jones and D. Yang, *J. Am. Chem. Soc.*, 2006, **128**, 1096–1098.
- 42 S. Masiero, R. Trotta, S. Pieraccini, S. De Tito, R. Perone, A. Randazzo and G. P. Spada, *Org. Biomol. Chem.*, 2010, **8**, 2683–2692.
- 43 A. T. Phan, Y. S. Modi and D. J. Patel, *J. Am. Chem. Soc.*, 2004, **126**, 8710–8716.
- 44 V. Kuryavyi, A. T. Phan and D. J. Patel, *Nucleic Acids Res.*, 2010, **38**, 6757–6773.
- 45 Y. Xu and H. Sugiyama, *Nucleic Acids Res.*, 2006, **34**, 949–954.
- 46 A. J. Ziemba, Z. V. Zhilina, Y. Krotova-Khan, L. Stankova and S. W. Ebbinghaus, *Oligonucleotides*, 2005, **15**, 36–50.
- 47 S. Cogoi and L. E. Xodo, *Nucleic Acids Res.*, 2006, **34**, 2536–2549.
- 48 J. Feigon, K. M. Koshlap and F. W. Smith, *Methods Enzymol.*, 1995, **261**, 225–255.
- 49 H. Fernando, A. P. Reszka, J. Huppert, S. Ladame, S. Rankin, A. R. Venkitaraman, S. Neidle and S. Balasubramanian, *Biochemistry*, 2006, **45**, 7854–7860.
- 50 J. Dai, C. Punchedhewa, A. Ambrus, D. Chen, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 2440–2450.
- 51 A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2006, **34**, 2723–2735.
- 52 R. E. Martenson, *J. Biol. Chem.*, 1978, **253**, 8887–8893.
- 53 R. P. Frigon, J. K. Leypoldt, S. Uyeji and L. W. Henderson, *Anal. Chem.*, 1983, **55**, 1349–1354.
- 54 C. Y. Kuo and T. Provder, *ACS Symp. Ser.*, 1987, **352**, 2–28.