

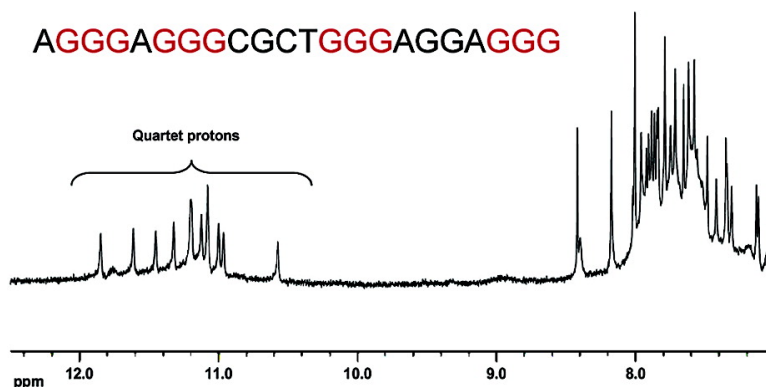
Article

Putative DNA Quadruplex Formation within the Human *c-kit* Oncogene

Sarah Rankin, Anthony P. Reszka, Julian Huppert, Mire Zloh, Gary N. Parkinson,
 Alan K. Todd, Sylvain Ladame, Shankar Balasubramanian, and Stephen Neidle

J. Am. Chem. Soc., **2005**, 127 (30), 10584-10589 • DOI: 10.1021/ja050823u • Publication Date (Web): 08 July 2005

Downloaded from <http://pubs.acs.org> on March 25, 2009



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 30 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

[View the Full Text HTML](#)

Putative DNA Quadruplex Formation within the Human *c-kit* Oncogene

Sarah Rankin,[†] Anthony P. Reszka,[†] Julian Huppert,[‡] Mire Zloh,[†]
Gary N. Parkinson,[†] Alan K. Todd,[†] Sylvain Ladame,[‡]
Shankar Balasubramanian,[‡] and Stephen Neidle^{*†}

Contribution from the Cancer Research UK Biomolecular Structure Group, The School of Pharmacy, University of London, 29-39 Brunswick Square, London WC1N 1AX, UK, and University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

Received February 8, 2005; E-mail: stephen.neidle@ulsop.ac.uk

Abstract: The DNA sequence, d(AGGGAGGGCGCTGGGAGGAGGG), occurs within the promoter region of the *c-kit* oncogene. We show here, using a combination of NMR, circular dichroism, and melting temperature measurements, that this sequence forms a four-stranded quadruplex structure under physiological conditions. Variations in the sequences that intervene between the guanine tracts have been examined, and surprisingly, none of these modified sequences forms a quadruplex arrangement under these conditions. This suggests that the occurrence of quadruplex-forming sequences within the human and other genomes is less than was hitherto expected. The *c-kit* quadruplex may be a new target for therapeutic intervention in cancers where there is elevated expression of the *c-kit* gene.

Introduction

G-quartets are planar structures which are formed from the hydrogen-bonding association of guanine bases in nucleic acids via their Hoogsteen and Watson–Crick faces.^{1,2} Stacks of several consecutive G-quartets in guanine-rich nucleic acid sequences can, in turn, form four-stranded inter- or intramolecular structures, termed quadruplexes. Telomeric DNA, with its tandem repeats of guanine-rich sequences, comprises the best-studied source of quadruplex-forming nucleic acids, with X-ray and NMR structural data as well as extensive biophysical data being available for G-quadruplexes formed from short sequences of human telomeric DNA^{3–5} and from *Oxytricha*^{6,7} and *Tetrahymena*^{8,9} telomere sequences. Telomeres in cancer cells are currently being studied as a target for anti-cancer therapies, with inhibition of the action of the reverse transcriptase enzyme, telomerase, being achieved by promoting the 3' single-stranded end of telomeric DNA to fold into G-quadruplex structures with the aid of quadruplex-selective ligands.^{10–14} Quadruplex structures abolish the ability of the 3'-end to act as a substrate for telomerase.

There has been increasing interest in whether nontelomeric regions of G-rich DNA in the human genome are inherently capable of forming stable G-quadruplex structures, in which runs of guanine bases are separated by short (1–6 bases) loop sequences. Putative quadruplexes have been identified in a number of nontelomeric genes and genomic sequences, including that of the oncogene *c-myc*,^{15–19} the fragile X syndrome²⁰ and other triplet repeat nucleotide sequences,²¹ and the promoter sequence²² of the *Ki-ras* oncogene. Quadruplex formation has also been reported in insulin-like growth factor II mRNA,²³ and CD and NMR experiments support G-quartet formation within

[†] University of London.

[‡] University of Cambridge.

- (1) Sundquist, W. I.; Klug, A. *Nature* **1989**, *342*, 825–829.
- (2) Henderson, E.; Hardin, C. C.; Walk, S. K.; Tinoco, I.; Blackburn, E. H. *Cell* **1987**, *51*, 899–908.
- (3) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876–880.
- (4) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263–282.
- (5) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2003**, *125*, 15021–15027.
- (6) Schultze, P.; Hud, N. V.; Smith, F. W.; Feigon, J. *Nucleic Acids Res.* **1999**, *27*, 3018–3028.
- (7) Haider, S.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2002**, *320*, 189–200.
- (8) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Mol. Biol.* **2004**, *338*, 93–102.
- (9) Phillips, K.; Dauter, Z.; Murchie, A. I. H.; Lilley, D. M.; Luisi, B. J. *Mol. Biol.* **1997**, *273*, 171–182.
- (10) Neidle, S.; Parkinson, G. N. *Nat. Rev. Drug Discovery* **2002**, *1*, 383–393.
- (11) Riou, J.-F.; Guittat, L.; Mailliet, P.; Laoui, A.; Renou, E.; Petitgenet, O.; Méglin-Chanet, F.; Hélène, C.; Mergny, J.-L. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2672–2677.
- (12) Read, M. A.; Harrison, R. J.; Romagnoli, B.; Tanious, F. A.; Gowan, S. H.; Reszka, A. P.; Wilson, W. D.; Kelland, L. R.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4844–4849.
- (13) Heald, R. A.; Modi, C.; Cookson, J. C.; Hutchinson, I.; Laughton, C. A.; Gowan, S. M.; Kelland, L. R.; Stevens, M. F. G. *J. Med. Chem.* **2002**, *45*, 590–597.
- (14) Shi, D. F.; Wheelhouse, R. T.; Sun, D.; Hurley, L. H. *J. Med. Chem.* **2001**, *44*, 4509–4523.
- (15) Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.* **1998**, *26*, 1167–1172.
- (16) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593–11598.
- (17) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Am. Chem. Soc.* **2004**, *126*, 8710–8716.
- (18) Seenisamy, J.; Rezler, E. M.; Powell, T. J.; Tye, D.; Gokhale, V.; Joshi, C. S.; Siddiqui-Jain, A.; Hurley, L. H. *J. Am. Chem. Soc.* **2004**, *126*, 8702–8709.
- (19) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. *Biochemistry* **2005**, *44*, 2048–2058.
- (20) Kettani, A.; Kumar, R. A.; Patel, D. J. *J. Mol. Biol.* **1995**, *254*, 638–656.
- (21) Matsugami, A.; Okuizumi, T.; Uesugi, S.; Katahira, M. *J. Biol. Chem.* **2003**, *278*, 28147–28153.
- (22) Cogo, S.; Quadrioglio, F.; Xodo, L. E. *Biochemistry* **2004**, *43*, 2512–2523.
- (23) Christiansen, J.; Kofod, M.; Nielsen, F. C. *Nucleic Acids Res.* **1994**, *22*, 5709–5716.

a sequence²⁴ from the human insulin gene. Particular emphasis to date has been on the *c-myc* gene, where parallel-stranded quadruplexes have been identified in the nuclease hypersensitivity element (NHE) III₁ upstream of the P1 and P2 promoters of the *c-myc* promoter region,^{17–19} and levels of gene expression have been altered by the addition of G-quadruplex stabilizing ligands.¹⁷ It has been proposed that ligand binding to the G-rich sequence results in a switch from a parallel to a mixed parallel/antiparallel quadruplex structure.¹⁸

We have initiated a systematic search for putative G4 sequences in the human genome and are examining not only their occurrence and frequency but also their ability to potentially form stable G4 structures under physiological conditions. We describe here one such sequence, from the human proto-oncogene *c-kit*, which encodes for a 145–160 kDa membrane-bound glycoprotein belonging to a family of growth factor receptors with tyrosine kinase activity.²⁵ The *c-kit* gene is expressed by and is critical for the development of mast cells, melanocytes, and hematopoietic stem cells²⁶ and is an attractive target in the treatment of gastrointestinal tumors (GIST), which have activating mutations of the gene.²⁷ The *c-kit* expression levels are maintained only in a number of other tumor types, such as prostate²⁸ and adenocarcinoma lung cancers.²⁹ Conversely, *c-kit* expression is diminished or absent in cutaneous melanoma³⁰ and breast cancers.³¹ We have identified a G4-type sequence, *c-kit87up*, 87 base pairs upstream of the transcription start site of the human *c-kit* gene, which we show here forms a single G-quadruplex species in K⁺-containing solution. This 22-nt sequence consists of four runs of three guanine bases, separated by three loop regions, consisting of a single A residue, a CGCT loop, and an AGGA loop, respectively.

The full range of factors influencing quadruplex structure and stability are not fully understood, in part due to the relative paucity of high-resolution X-ray and NMR quadruplex structures. The presence of alkali metal ions in the center of the quadruplex is important for the stability of quadruplexes,³² as are the number of guanine bases present in the sequence.³³ The length of the loop sequences also plays a vital role in the stabilization (or destabilization) of guanine quadruplex structures,³⁴ an effect that may be due to inter-residue hydrogen bonding and base stacking interactions between loop bases.³⁵ However, it is not yet clear whether or how the loop sequence itself affects quadruplex formation or stability. To investigate this issue further, we have studied the quadruplex-forming ability

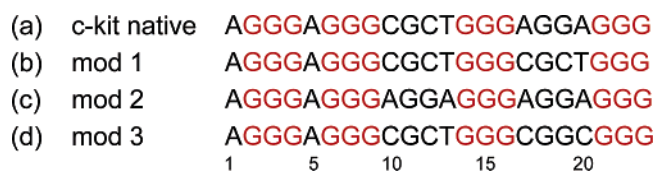


Figure 1. The *c-kit* native and modified sequences studied in this work. The regions that may form G-tetrad planes are highlighted in red with the loop regions shown in black.

of several loop variants of the *c-kit* native sequence. The first modification replaces the third loop region with the second one; the second modification replaces the second loop with the third loop, and the final modified sequence replaces the final AGGA loop with a CGGC loop. To maintain a level of similarity between the modified and native sequences, we left the first single base loop unchanged. Our intention with the first and second modification was to see if the position of loop sequences had an effect on the formation or stability of the guanine quadruplex. The third modification investigates the different effects on G-quadruplex formation of conservative purine and pyrimidine switching within the loop (Figure 1).

A range of biophysical techniques have been used in this work, NMR, UV melting profiles, and circular dichroism spectroscopy. In the absence of structural information from multidimensional NMR and X-ray crystallography, a multi-method approach may provide several complementary lines of evidence for quadruplex formation. NMR methods, in particular, have long been used to study quadruplexes in DNA, with several biologically important quadruplex structures and topologies being determined by 2D NMR.^{4,8,36–38} The imino protons in G-quadruplexes have been established to give rise to sharp and characteristic signals in 1D NMR spectra, and we have used these signals here to verify G-quadruplex DNA formation within the native and modified sequences. No topological information is immediately available from standard 1D NMR experiments, and therefore, circular dichroism techniques have been employed in order to attempt to assign the topologies of the quadruplexes. CD has been used to qualitatively assign quadruplex topology based on standard spectra for quadruplexes of known parallel and antiparallel topology.^{39–42} UV melting studies at 295 nm have also been used, a technique first proposed by Mergny et al.⁴³ for the study of G-quadruplex structures.

Materials and Methods

Oligonucleotide Sample Preparation. The native and modified sequences were synthesized on an Applied Biosystems DNA synthesizer using solid-phase β -cyanoethylphosphoramidite chemistry. The oligodeoxyribonucleotides were deprotected and cleaved from the solid resin support by immersing the resin in concentrated degassed ammonia solution at 60 °C overnight. The crude oligonucleotides were purified using HPLC on a BioCAD Workstation with an Oligo R3 column

- (24) Catasti, P.; Chen, X.; Moyzis, R. K.; Bradbury, E. M.; Gupta, G. *J. Mol. Biol.* **1996**, *264*, 534–545.
 (25) Yarden, Y.; Kuang, W. J.; Yangfeng, T.; Coussens, L.; Munemitsu, S.; Dull, T. J.; Chen, E.; Schlessinger, J.; Francke, U.; Ullrich, A. *EMBO J.* **1987**, *6*, 3341–3351.
 (26) Huizinga, J. D.; Thuneberg, L.; Kluppel, M.; Malysz, J.; Mikkelsen, H. B.; Bernstein, A. *Nature* **1995**, *373*, 347–349.
 (27) Sattler, M.; Salgia, R. *Leukaemia Res.* **2004**, *28*, S11–S20.
 (28) Simak, R.; Capodici, P.; Cohen, D. W.; Fair, W. R.; Scher, H.; Melamed, J.; Drobnyak, M.; Heston, W. D.; Stix, U.; Steiner, G.; Cordon-Cardo, C. *Histol. Histopath.* **2000**, *15*, 365–374.
 (29) Micke, P.; Hengstler, J. G.; Albrecht, H.; Faldum, A.; Bittinger, F.; Becker, K.; Wiewrodt, R.; Fischer, B.; Buhl, R. *Tumour Biol.* **2004**, *25*, 235–242.
 (30) Natali, P. G.; Nicotra, M. R.; Winkler, A. B.; Cavaliere, R.; Bigotti, A.; Ullrich, A. *Int. J. Cancer* **1992**, *52*, 197–201.
 (31) Natali, P. G.; Nicotra, M. R.; Sures, I.; Mottolose, M.; Botti, C.; Ullrich, A. *Int. J. Cancer* **1992**, *52*, 713–717.
 (32) Scaria, P. V.; Shire, S. J.; Shafer, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 10336–10340.
 (33) Fojtik, P.; Kejnovska, I.; Vorlickova, M. *Nucleic Acids Res.* **2004**, *32*, 298–306.
 (34) Smirnov, I.; Shafer, R. H. *Biochemistry* **2000**, *39*, 1462–1468.
 (35) Keniry, M. A.; Owen, E. A.; Shafer, R. H. *Nucleic Acids Res.* **1997**, *25*, 4389–4392.

- (36) Cmugelj, M.; Sket, P.; Plavec, J. *J. Am. Chem. Soc.* **2003**, *125*, 7866–7871.
 (37) Smith, F. W.; Feigon, J. *Nature* **1992**, *356*, 164–168.
 (38) Smith, F. W.; Lau, F. W.; Feigon, J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10546–10550.
 (39) Balagurumoorthy, P.; Brahmachari, S. K.; Mohanty, D.; Bansal, M.; Sasisekharan, V. *Nucleic Acids Res.* **1992**, *20*, 4061–4067.
 (40) Berova, N.; Nakanishi, K.; Woody, R. W. *Circular Dichroism: Principles and Applications*; Wiley-VCH: New York, 2004.
 (41) Giraldo, R.; Suzuki, M.; Chapman, L.; Rhodes, D. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7658–7662.
 (42) Rujan, I. N.; Meloney, J. C.; Bolton, P. H. *Nucleic Acids Res.* **2005**, *33*, 2022–2031.
 (43) Mergny, J. L.; Phan, A. T.; Lacroix, L. *FEBS Lett.* **1998**, *435*, 74–78.

(PerSeptive Biosystems, Inc.). The sequences were then lyophilized and dialyzed against double-distilled water and lyophilized again to produce the final purified oligonucleotide. Concentration measurements of stock solutions were performed using samples dissolved in distilled water.

Nuclear Magnetic Resonance. NMR experiments were conducted on a Bruker Avance 500 MHz instrument. Experiments in H₂O used excitation suppression of the water signal. Samples were made up in 100 mM KCl solution with 20 mM potassium phosphate pH 7.0 buffer. Samples were annealed by heating to 85 °C followed by slow cooling over a period of 20 h; 540 μ L of sample and 60 μ L of D₂O were then mixed to give a strand concentration of around 2 mM with a 10% D₂O content. Samples were then placed in NMR tubes and flushed with argon gas. NMR experiments were conducted at 25 °C for the native and modified sequences. A limited temperature-dependence study of the behavior of the imino resonances for the native sequence was also undertaken.

Ultraviolet Melting Point Experiments. A significant hyperchromic shift at 295 nm has been observed upon G-quadruplex melting and is thought to be unique to G-quadruplex structures, so the method first reported by Mergny et al.⁴³ has been broadly followed. Solutions containing 20 μ M oligonucleotide, 10 mM sodium cacodylate pH 7.0 buffer, and 100 mM KCl were prepared; 1.5 mL of solution was transferred to a quartz cell with a path length of 1 cm, which was then sealed. No annealing was performed prior to melting curve experiments. UV melting profiles were conducted on a Varian Bio-300 UV/vis spectrophotometer with a Varian temperature controller. Samples were first heated to 90 °C with a temperature gradient of 10 °C/min and held at 90 °C for 2 min without data collection before being cooled to 5 °C with a temperature gradient of 0.5 °C/min, during which absorbance data were recorded. Samples were then returned to 90 °C with an identical temperature gradient, again with data collection. The data were then normalized and smoothed, with melting temperatures being calculated using the first derivative method.

Circular Dichroism. Circular dichroism experiments were performed on a Jasco J-810 spectropolarimeter using software supplied by the manufacturer. Samples were prepared at a 20 μ M concentration in 100 mM KCl solution. The samples were annealed as described above; 400 μ L of sample was placed in a quartz cuvette with a path length of 1 cm. Scans were performed over the range 220–320 nm at 25 °C with a constant flow of dried nitrogen applied to the sample. Samples were allowed to equilibrate at 25 °C for 10 min before runs. Each trace is the average result of five scans. A blank dataset taken from the salt solution alone was also recorded and subtracted from the results. The collected data were smoothed and zero-corrected at 320 nm.

Results

NMR Studies on the *c-kit* Native Sequence. G-quadruplex structures display characteristic resonances for the imino protons,^{5,8} whose presence may be used to confirm the presence of G-quadruplex structure in solution. The NMR spectrum of the *c-kit* native sequence is of exceptional quality with well resolved and sharp imino proton peaks (Figure 2a). Formation of a single quadruplex species is apparent from the number and intensity of the peaks, with the sharpest peaks having a line-width of around 3 Hz at 25 °C. This is consistent with intramolecular monomeric quadruplex formation. Between 10.5 and 12.0 ppm, eight peaks integrate to single protons, while the larger peaks at 11.08 and 11.20 ppm correspond to two protons each. These 12 protons correspond exactly to the 12 imino protons involved in inter-guanine hydrogen bonding within the G-quartet and are unequivocal evidence for three G-quartet planes in the structure.

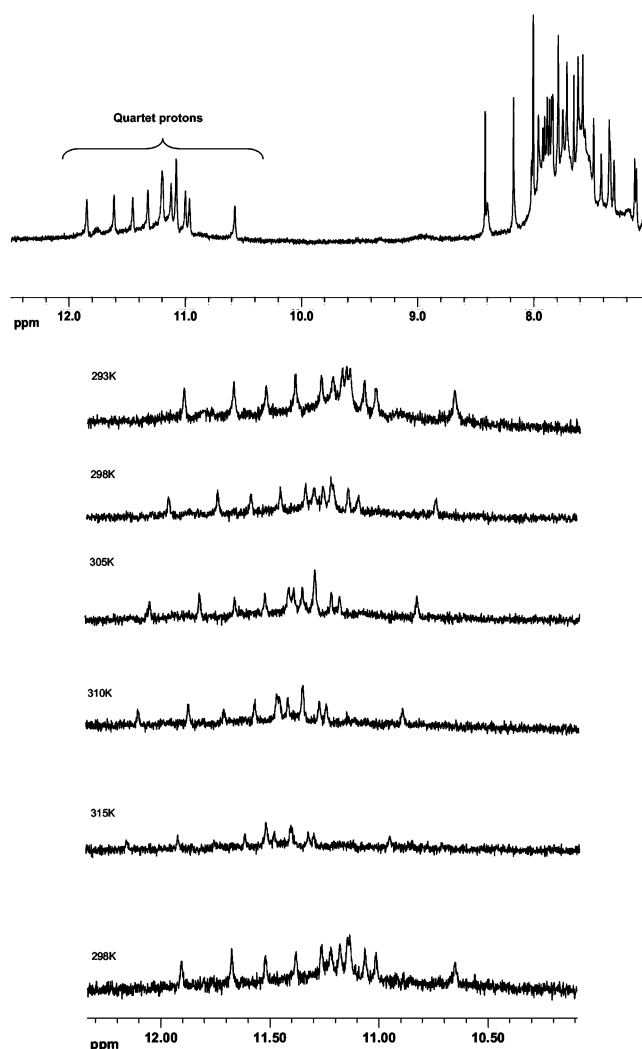


Figure 2. (a) One-dimensional 500 MHz proton spectrum of the *c-kit* native sequence. Experimental conditions: strand concentration, 2 mM; temperature, 25 °C, 100 mM KCl; potassium phosphate, 20 mM, pH 7.0. (b) One-dimensional 500 MHz proton spectra of the *c-kit* native sequence at various temperatures. Experimental conditions: strand concentration, 2 mM; 100 mM KCl; potassium phosphate, 20 mM, pH 7.0.

UV Melting Point Studies of the *c-kit* Native Sequence.

The melting curves obtained for the native sequence (Figure 3a) exhibit a single melting transition at 295 nm, with little difference between the heating and cooling curves. As with other G-quartet species,⁴³ a decrease in absorbance at 295 nm is observed upon heating. The lack of hysteresis and the sigmoidal shape of the curves are indicative of reversible quadruplex formation and relatively fast folding kinetics. Studies over a small range of strand concentrations (1–10 μ M) revealed no significant change in melting temperature (data not shown), pointing to a unimolecular transition, rather than an intermolecular one, during melting. We note that this result extends the protocol used by Mergny et al.⁴³ to nontelomeric quadruplexes. Variable-temperature NMR studies (Figure 2b) show, in the imino region, evidence of significant melting beyond around 40 °C, and we estimate the T_m to be ca. 50 °C, in excellent agreement with that from the UV study (Table 1). It should be noted that the concentrations used in these experiments are several orders of magnitude higher than those in the UV melting studies, so we cannot exclude the possibility of intermolecular complexes being formed.

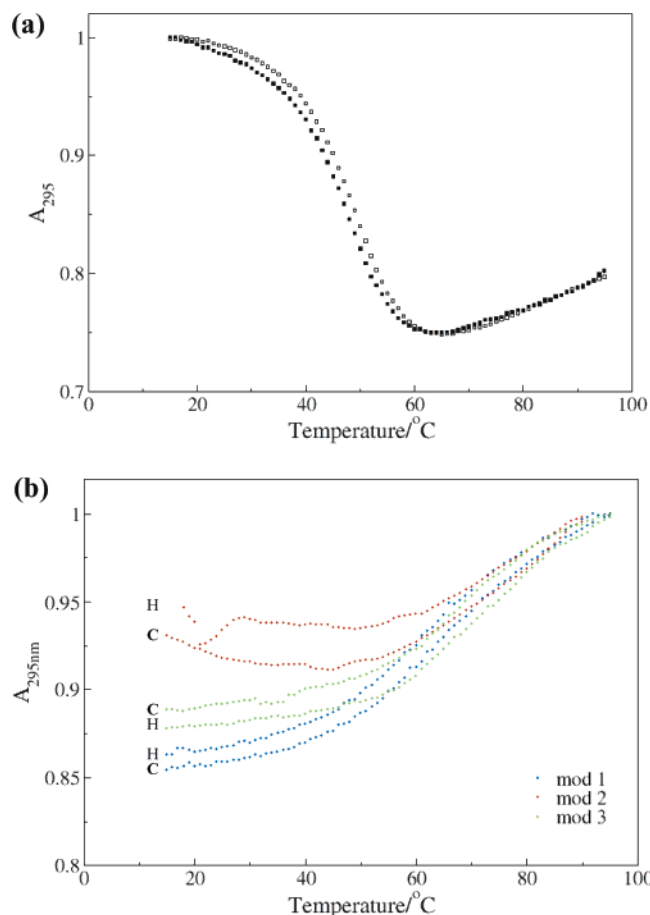


Figure 3. Normalized absorbance plots at 295 nm versus temperature, with strand concentration of 20 μ M, 10 mM sodium cacodylate pH 7.0 buffer, and 100 mM KCl. (a) For the *c-kit* native sequence: open circles represent cooling curves; filled circles represent heating curves. (b) For *c-kit* modified sequences: mod 1 (blue), mod 2 (red), and mod 3 (green). The heating and cooling curves are marked H and C, respectively.

Table 1. Melting Temperatures of the *c-kit* Native, Human Telomeric, and *c-myc* Quadruplexes

sequence	conditions	T_m /°C
<i>c-kit</i> native ^a	100 mM K ⁺	51
human telomeric repeat d[G ₃ (TTAG ₃) ₃] ^{a,c}	100 mM K ⁺	65
myc-2345 ^{b,d}	90 mM K ⁺	>80
myc-1245 ^{b,d}	90 mM K ⁺	75

^a Determined by UV/visible spectrometry. ^b Determined by NMR spectroscopy. ^c See ref 29. ^d See ref 17.

Circular Dichroism Studies on the *c-kit* Native Sequence.

G-quadruplexes exhibit characteristic CD spectra depending on the topology of the structure (parallel or antiparallel).^{39–42} Parallel folds exhibit a maximum positive signal at around 260 nm with a corresponding negative signal around 240 nm. Antiparallel quadruplex species show a characteristic positive signal at around 295 nm with a negative signal at around 260 nm. The spectrum for the native *c-kit* sequence is consistent with quadruplex formation, but does not suggest a clear assignment (Figure 4a). A maximum positive absorbance is seen at around 300 nm, which corresponds to that seen for antiparallel quadruplexes; however, the CD spectrum does not show a corresponding negative maximum at 260 nm. The negative maximum is at around 275 nm, and there is a gradual return to positive ellipticity values, rather than the sharp return associated

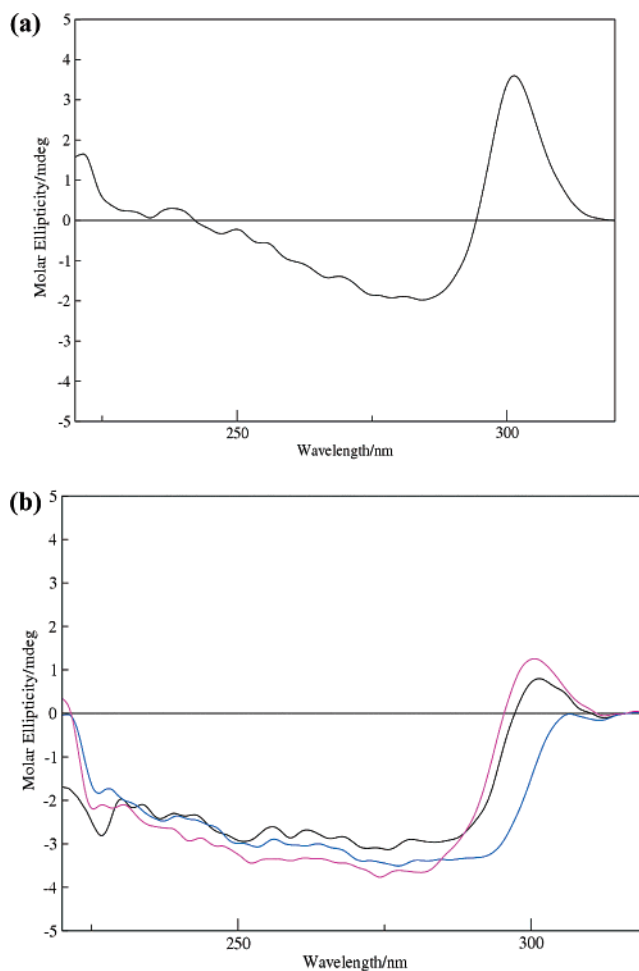


Figure 4. (a) CD spectrum of *c-kit* native sequence. (b) CD spectra of the modified sequences: mod 1 (black), mod 2 (blue), and mod 3 (magenta).

with previously reported CD spectra for antiparallel quadruplexes. The CD spectrum does not, therefore, correspond to published quadruplex CD spectra.

NMR Studies of the Modified Sequences. The imino and amino regions 1D proton spectra of the modified sequences are presented in Figure 5. Instead of the sharp imino protons present in the native sequence spectrum, broad and featureless envelopes are seen between 10 and 12 ppm for sequence mod 3, and very little signal is observed for mod 1. The broad envelope seen for the sequence mod 3 in the imino proton region indicates that G-quartet and secondary structure formation is minimal, and that no single quartet topology predominates among the strands forming G-quartets. The lack of any signal at all in the 10–12.5 ppm region for the sequence mod 1 is indicative of no G-quadruplex formation by this sequence. The mod 2 sequence shows a broad unresolved peak in the imino region. Again, this indicates a lack of significant quadruplex formation, though perhaps slightly more than mod 1.

Melting Profiles of the Modified Sequences. The absorbance versus temperature melting profiles (Figure 3b) for the modified sequences do not display the characteristic quadruplex melting behavior⁴² with a general increase in the absorbance at 295 nm, in stark contrast with the decrease observed for authentic G-quadruplexes. The magnitude of the change in absorbance upon melting is typically smaller than that seen for the native sequence. In addition, significant hysteresis is observed for the

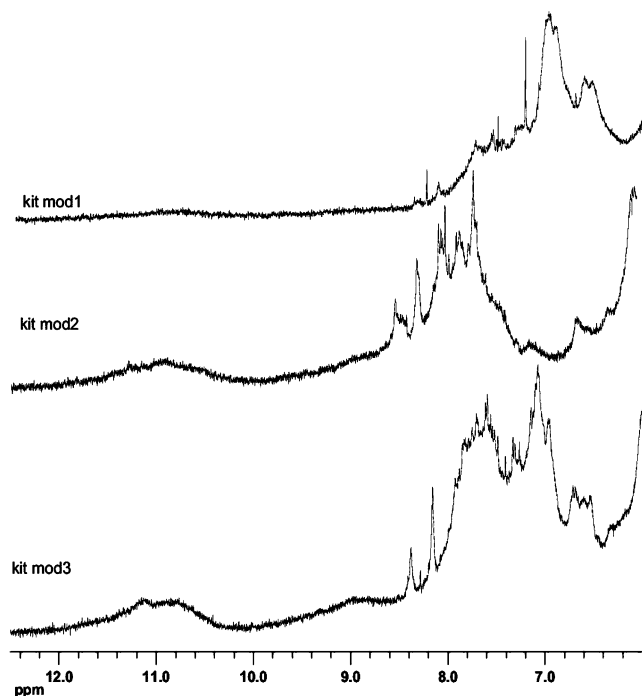


Figure 5. Imino and amino proton regions from 1D ^1H NMR spectra of the *c-kit* modified sequences.

mod 2 sequence during heating and cooling cycles. We were, therefore, unable to calculate melting temperatures for the modified sequences based on the UV melting technique. We conclude that these are not indicative of quadruplex formation under experimental conditions identical to that used for the native sequence.

Circular Dichroism of the Modified Sequences. The CD spectra for the modified sequences do not show data consistent with quadruplex formation. Similar data are reported for all three modified sequences (Figure 4b). A slight increase in positive ellipticity is seen around 300 nm; however, the negative ellipticity shows little change between 280 and 225 nm. The ellipticity values are low and, again, point to a lack of quadruplex formation within these sequences.

Discussion

A Novel G-Quartet in the *c-kit* Gene. We have identified a novel unimolecular quadruplex structure that forms in K^+ -containing solution, which is taken from a G-rich sequence in the human *c-kit* gene. Formation of this quadruplex species is reversible, as indicated by the nature of the heating and cooling curves produced during the UV melting experiments. The small line-widths of the imino proton peaks in the 1D NMR spectrum and the lack of hysteresis during melting point are interpreted as indicating the formation of a unimolecular quadruplex species, as shown by the 12 imino proton signals.¹⁷ We note that it is rare for an unmodified quadruplex sequence to exhibit a 1D NMR spectrum with such well-defined peaks in the imino proton region. Studies⁵ of the 12-mer human telomeric sequence, d(TAG₃TTAG₃T), show interconversion between two propeller quadruplex forms, while the imino proton region of the Pu27mer *c-myc* NHE III₁ quadruplex-forming sequences consists of a broad envelope which exhibits some fine structure.¹⁷ At present, we are unable to unequivocally assign the topology of this quadruplex species due to the inconclusive nature of the CD

results; however, the observation of a single quadruplex species, by NMR spectroscopy, suggests that one particular topological form is energetically favorable for the *c-kit* native sequence. The presence of a single-nucleotide loop in the native sequence is likely to be more consistent with a parallel-type arrangement.^{3,44} Equally, other possible conformations may be heavily disfavored in some manner; further structural evidence is required before this question can be fully answered. The high-quality 1D NMR spectra of the native *c-kit* sequence provide an excellent opportunity for further structural characterization, which is now underway in this laboratory.

The native sequence quadruplex melts at a temperature comparable to that of other telomeric and nontelomeric sequences (Table 1). Both the human telomeric and *c-myc* quadruplexes, which are suspected of forming more than one quadruplex species under similar conditions,^{5,17,18} melt at higher temperatures than the *c-kit* native sequence. However, its melting point is still well above physiological temperature and leaves room for further quadruplex stabilization by favorable ligand–quadruplex interactions.

Limitations of Circular Dichroism Data. The NMR spectrum for the *c-kit* native sequence is unambiguous in showing formation of a single, well-formed quadruplex species in K^+ solution. However, the CD spectrum (Figure 4a) does not conform to the characteristic shapes seen for either parallel or antiparallel quadruplex species. It is, therefore, not possible to unequivocally assign the topology of the *c-kit* native quadruplex from these data. It is thus unwise to rely solely on circular dichroism data for determining the topology of a new G-quadruplex since the CD spectrum of a particular quadruplex reflects the subtleties of that sequence and its fold. Recent studies have highlighted possible inconsistencies in CD data, especially where multiple quadruplex species are suspected of being formed.⁴⁵

Modified *c-kit* Sequences Do Not Exhibit Clear Quadruplex Formation. All the experimental data for the modified *c-kit* sequences examined here conclude that these are unable to form G-quadruplex species under experimental conditions identical to that of the native sequence, which are normal conditions for quadruplex formation. The NMR spectra lack clear imino proton signals, with broad envelopes being observed in the imino proton region for sequence mod 3, and sequence mod 1 shows very little signal in this region at all. Such poor quality signals from this imino proton region make it highly unlikely that G-quadruplex structures are forming. The UV melting point curves at 295 nm show distinct behavior to that expected from G-quartet species, and the CD spectra show no evidence of quadruplex formation.

Sequence Dependence of Quadruplex Formation. After examining the sharp imino proton resonances present in the native sequence NMR spectrum, we expected at least one of the modified sequences to produce a quadruplex species after identical annealing. However, we were surprised by the dramatic destabilizing effect that the small sequence modifications had on the ability of the modified sequences to form or maintain stable G-quadruplexes. We conclude that quadruplex formation for this G-rich tract is strongly sequence-dependent, with

(44) Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.* **2004**, *126*, 16405–16415.

(45) Dapic, V.; Abdomerovic, V.; Marrington, R.; Peberdy, J.; Rodger, A.; Trent, J. O.; Bates, P. *J. Nucleic Acids Res.* **2003**, *31*, 2097–2107.

conservative changes in the loop regions having a major detrimental effect on quadruplex formation. By replacing one four-base loop with another (for example, in sequences mod 1 and mod 2), we were able to prevent quadruplex formation and, therefore, demonstrate that all loops contribute in some degree to quadruplex stability. We have also demonstrated that the sequence itself within the loop region contributes to the ability of a G-rich sequence to form a quadruplex; by altering two purine residues in a loop to pyrimidines (sequence mod 3), we were again able to prevent formation of a stable quadruplex species.

Importance of Interactions within the Loops. The *c-kit* native quadruplex must form particular stabilizing interactions involving the loop regions; without such favorable loop interactions, quadruplex formation would not be so dramatically affected by conservative changes to the loop sequence. However, we cannot rule out the possibility that our observations result from destabilizing interactions being introduced in the loops of the variants. This result serves to stress the importance of interactions within the loop regions and that these interactions, whether intraresidue within the loop or loop–quartet interactions, play an important part in the formation and stability of G-quadruplex structures, although we are as yet far from fully understanding them. The negative results for the modified sequences also suggest that the loop regions may contribute cooperatively in some measure to quadruplex formation and stability. A more extensive study of modified sequences needs to be undertaken before the exact role of each base within the loop region on quadruplex stability may be examined; this will undoubtedly be facilitated by further structural information on the *c-kit* native quadruplex. We are also unable to identify the exact mechanisms by which G-quadruplexes are destabilized by the modified loop sequences. We, therefore, conclude that the influence of the loop region interactions on quadruplex stability can be much more complex and involved than previously thought. This is particularly the case for nontelomeric sequences with their wide variety in loop sequence and lengths.

Biological Relevance. The quadruplex-forming ability of the *c-kit* native sequence may be an attractive target for ligand design, with the aim of altering levels of *c-kit* gene expression

using quadruplex-binding therapeutic agents. Successful targeting of the native sequence would rely on the accessibility and quadruplex-forming ability of this sequence *in vivo*; however, selective agents may be designed for the single quadruplex species that forms, in contrast to the *c-myc* gene which shows multiple quadruplex species formation under similar conditions. With the increasing number of quadruplexes being reported in the genome,^{46,47} ligand design must now be directed not only to differentiating between duplex and quadruplex DNA species but also toward designing selectivity between different quadruplex species. The loop regions of G-quadruplexes provide a natural and accessible structure element for exploitation in order to achieve this selectivity.

We have established that high guanine content in a given DNA sequence is not a sufficient determinant of quadruplex formation, but that the nature of the loop regions plays a crucial role in quadruplex stability, in terms of not only the loop sequence length but also the loop sequence itself. This dependence on loop regions is also shown by the *c-myc* G-rich sequences, for which the native sequence, Pu27mer, forms multiple quadruplex species in solution.¹⁷ G-rich sequences derived from the *c-kit* native sequence, however, form a single quadruplex species under identical conditions as judged by NMR. The rules governing quadruplex stability and formation are evidently more complex than was initially thought and can depend significantly upon interactions formed between both the loops and the loop–quartet interactions. A further consequence of the present study is that the total number of actual quadruplexes encoded within a genome *and* capable of forming a stable quadruplex structure is probably less than the total of putative quadruplex sequences that are present in the genome.^{46,47}

Acknowledgment. We are grateful for grant support from Cancer Research UK (to S.N. and S.B.), the Isaac Newton Trust, the BBSRC, and Trinity College, Cambridge (funding to J.L.H.), and the School of Pharmacy (research studentship to S.R.).

JA050823U

(46) Todd, A. K.; Johnston, M.; Neidle, S. *Nucleic Acids Res.* **2005**, *33*, 2901–2907.

(47) Huppert, J. H.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908–2916.