Published in final edited form as:

J Am Chem Soc. 2007 April 11; 129(14): 4386–4392. doi:10.1021/ja068739h.

Structure of an Unprecedented G-Quadruplex Scaffold in the Human *c-kit* Promoter

Anh Tuân Phan[†,‡], **Vitaly Kuryavyi**[†], **Sarah Burge**[§,‖], **Stephen Neidle**[§], and **Dinshaw J. Patel**[†]

[†]Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021

[§]Cancer Research UK Biomolecular Structure Group, School of Pharmacy, University of London, London WC1N 1AX, United Kingdom

Anh Tuân Phan: phantuan@ntu.edu.sg; Stephen Neidle: stephen.neidle@pharmacy.ac.uk; Dinshaw J. Patel: pateld@mskcc.org

Abstract

The *c-kit* oncogene is an important target in the treatment of gastrointestinal tumors. A potential approach to inhibition of the expression of this gene involves selective stabilization of G-quadruplex structures that may be induced to form in the *c-kit* promoter region. Here we report on the structure of an unprecedented intramolecular G-quadruplex formed by a G-rich sequence in the *c-kit* promoter in K^+ solution. The structure represents a new folding topology with several unique features. Most strikingly, an isolated guanine is involved in G-tetrad core formation, despite the presence of four three-guanine tracts. There are four loops: two single-residue double-chain-reversal loops, a two-residue loop, and a five-residue stem-loop, which contain base-pairing alignments. This unique structural scaffold provides a highly specific platform for the future design of ligands specifically targeted to the promoter DNA of *c-kit*.

Introduction

The proto-oncogene *c-kit* encodes for a 145-160 kDa tyrosine kinase receptor,[1] which regulates key signal transduction cascades to control cell growth and proliferation. As the c-kit protein plays such a critical role in establishing normal cell growth, mutations in structurally important regions of the protein, or its overexpression, result in impaired function, which leads to oncogenic cellular transformations.[2] Gain-of-function mutations are found in several highly malignant human cancers.[3-5] Human gastrointestinal stromal tumors (GIST) have been associated with mutations around the two main autophosphorylation sites in the juxtamembrane region,[3] whereas kinase domain mutations are found in myeloid

Correspondence to: Anh Tuân Phan, phantuan@ntu.edu.sg; Stephen Neidle, stephen.neidle@pharmacy.ac.uk; Dinshaw J. Patel, pateld@mskcc.org.

[‡]Present address: Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore

[‖]Present address: MRC Centre for Protein Engineering, Cambridge, CB2 2QH, UK

Supporting Information Available: Table S1, a list of sitespecific low-enrichment ^{15}N-labeled sequences used for resonance assignments; Figures S1-S3, imino proton spectra of *c-kit87up* at 5 °C (Figure S1), in Na^+ and K^+ solution (Figure S2), and in real-time hydrogen exchange experiments (Figure S3). This material is available free of charge via the Internet at http://pubs.acs.org.

leukemias and human germ cell tumors.[4,5] These oncogenic mutations result in increased dimer formation by bringing together two c-kit kinases into close proximity and facilitating autophosphorylation.

The c-kit protein is currently the principal therapeutically important target in the treatment of GIST.[6] The small molecule Gleevec (imatinib), originally developed as an inhibitor of the c-abl kinase in chronic myelogenous leukaemia (CML), was subsequently found to be an effective inhibitor *in* V*itro* and *in* V*iVo* of c-kit kinase activity.[7] It is now widely used as a therapy for GIST, where it has made a major difference to survival rates.[6,8] However, as with CML, resistance to Gleevec occurs as a result of deactivating mutations in the active site, which diminish binding and clinical effectiveness of the drug.[9] A number of new c-kit kinase inhibitors are currently being developed and evaluated with the aim of circumventing this resistance,[10] although they themselves may well produce new patterns of resistance mutations.

An alternative approach to *c-kit* inhibition would involve selective gene regulation at the transcriptional level. Such an approach is currently being explored in the case of the *c-myc* oncogene,[11,12] which involves the induction of G-quadruplex DNA structures in a G-rich region of the *c-myc* promoter nuclease hypersensitivity element (NHE) III_1 that is responsible for up to 90% of *c-myc* transcription.[13] G-quadruplexes, which may have transient stability by themselves when embedded within the double-stranded DNA of a eukaryotic gene, may be stabilized further by a small-molecule ligand such as the porphyrin molecule TMPyP4.[12] The topology and structure of the *c-myc* DNA quadruplexes containing four[14] and five[15] G-tracts have been determined by nuclear magnetic resonance (NMR) spectroscopy, together with that of a TMPyP4 complex for the five-guanine-tract quadruplex.[15] The latter are parallel-stranded quadruplexes, with several strand-reversal loops and base-pair platforms.[15] They have structural complexity beyond that of quadruplexes formed from telomeric DNA repeats,[16-21] suggesting the possibility of identifying ligands that are selective for the *c-myc* quadruplexes.

Recently, two G-rich sequences have been identified in the promoter region of the human *c-kit* gene, and biophysical data, including NMR, have established that these sequences can form G-quadruplex structures.[22,23] In particular, the *c-kit87up* d(AGGGAGGGCGCTGGGAGGAGGG) sequence, 87 nucleotides upstream of the transcription start site of the human *c-kit* gene, forms a single G-quadruplex species.[22] Because this sequence contains four tracts of three consecutive guanines separated by linkers of one or four residues, it was initially assumed to form a conventional G-quadruplex structure involving these G-tracts in the G-tetrad core and the linkers in the loops. Surprisingly, modifications in the linker sequences were found to be detrimental to formation of a folded quadruplex structure,[22] suggesting that they were more intimately involved in the quadruplex fold than would be expected for a conventional parallel or antiparallel arrangement.

Here we report on the NMR-based solution structure of the G-quadruplex formed by this sequence in K^+ solution. The structure reveals a highly unusual and unprecedented G-quadruplex folding topology with several unique elements. This structure rationalizes the

data on previous sequence modifications,[22] as well as that on many systematic sequence modifications performed in the present work. The unique scaffold of the *c-kit87up* quadruplex may provide a platform for designing ligands that specifically target the G-quadruplex topology in the *c-kit* promoter.

Results

NMR Spectra

NMR spectra of *c-kit87up* in K^+ solution are of excellent quality with well-resolved resonances (Figure 1a). The number of peaks is consistent with formation of a single conformation. Spectral line-widths (2-4 Hz for sharpest peaks at 25 °C) suggest formation of a monomeric structure, in agreement with previous data,[22] which showed that the melting temperature of *c-kit87up* was independent of the oligonucleotide concentration.

We observe 12 sharp peaks of guanine imino protons at 25 °C, which likely belong to the G-tetrad core. This suggests formation of a G-quadruplex structure containing three G-tetrad layers. Three other imino protons (two for guanines and one for a thymine) are visible at lower temperatures (Figure S1, Supporting Information).

In contrast to the narrow and well-resolved NMR spectrum of *c-kit87up* in K^+ solution (Figure S2b, Supporting Information), its counterpart in Na^+ solution exhibits a broad and unresolved envelope probably associated with an aggregated species (Figure S2a, Supporting Information).

Spectral Assignments

We prepared samples (Table S1, Supporting Information) in which only one guanine was 2% ^{15}N-labeled at a time, using deliberately diluted ^{15}N-labeled phosphoramidite.[24] Guanine imino protons were unambiguously assigned by the site-specific low-enrichment approach[24] (Figure 1b). Unexpectedly, these unambiguous assignments revealed that G10, an isolated guanine within the C9-G10-C11-T12 “linker” sequence, was among 12 sharp imino protons, but not that of G20 from the last G-tract.

Most guanine H8 protons were assigned by natural-abundance through-bond correlations[25] to the already assigned imino protons via ^{13}C5 (Figure 1c, d). Guanine H8 proton assignments were completed (or confirmed) by multi-bond correlations to ^{15}N nuclei in site-specific 2% ^{15}N-labeled samples.[24] The spectral assignments for other resonances were completed by through-bond (COSY and TOCSY) and through-space (NOE-SY) correlations between protons.[26]

G-Quadruplex Folding Topology

On the basis of the characteristic NOEs between imino and H8 protons (Figure 2), we established an unprecedented G-quadruplex fold for *c-kit87up*, involving three G-tetrads: G2•G6•G10•G13, G3•G7•G21•G14, and G4•G8•G22•G15 (Figure 2d). The glycosidic conformations of all guanine are anti, as reflected by the observed H1′-H8 NOE intensities (not shown). These glycosidic conformations are consistent with the G-tetrad core containing all parallel G-tracts (Figure 2d). There are four loops in the structure. Two single-

residue linkers (A5 and C9) form two double-chain-reversal loops that bridge three G-tetrad layers. The third loop, C11-T12, connects two adjacent corners (G10 and G13) of the G-tetrad core (Figure 2d). The five-residue segment A16-G17-G18-A19-G20 forms the fourth loop that allows the terminal G21-G22 stretch to be inserted back to the G-tetrad core. This folding topology is also supported by proton exchange data, which showed that imino protons of the central G-tetrad (G3, G7, G21, and G14) are the most protected from exchange with water (Figure S3, Supporting Information).

Solution Structure of *c-kit87up* G-Quadruplex

The structure of the *c-kit87up* quadruplex (stereo pair, Figure 3a; representative structure, Figure 3b, c) was calculated on the basis of NMR restraints (Table 1). The G-tetrad core is well defined. We observe three base pairs formed within the top and bottom loops (Figure 3): A1•T12 (Figure 4a), A16•G20 (Figure 4b, c), and G17•A19 (Figure 4c, d). The stacking of the A1•T12 (Figure 4a) and A16•G20 (Figure 4b) on the top and the bottom of the G-tetrad core, respectively, is reflected by the observation of NOE cross-peaks between A1(H2) to the imino protons of guanines from the top tetrad (G2, G6, G10, G13) and those between G20(NH_2) to the imino protons of guanines from the bottom tetrad (G4, G8, G22, G15) (Figure 2b). Formation of the Watson-Crick A1•T12 base pair is supported by the observation of the imino proton of T12 at 5 °C (Figure S1, Supporting Information) and the NOE between this proton and the A1(H2) proton (not shown). A1 adopts a syn conformation, consistent with the observation of a strong intraresidue H8-H1′NOE cross-peak (not shown).

The C9-G10-C11-T12 fragment forms an interesting configuration in the structure (Figure 5), where the middle G10 participates in the G-tetrad core.

Formation of the Watson-Crick-type A16•G20 base pair was supported by the observation of G20 imino proton at 12.8 ppm (Figure S1, Supporting Information) and the strong NOE between A16(H2) and G20(NH_2). This was independently confirmed in the sample *ck26* (Table 2), in which G20 was substituted by an inosine. NMR-restrained structure calculation revealed formation of the sheared G17•A19 pair (Figure 4c, d) and the single-residue turn G18 (Figure 4d). The positions of bases in the five-residue A16-G17-G18-A19-G20 loop were supported by several unusual chemical shifts: position of G18 over the G17•A19 base pair would explain the upfield shifts of its H1′and H5′/H5′p′rotons; upfield shifts of H5′a′nd H2″ (but not H5′and H2′) protons of A19 are consistent with the positions of these protons over the aromatic rings of both G18 and G20. The configuration of this unusual stem-loop resembles a hairpin loop observed previously,[27] which consists of a single residue turn closed by a sheared G•A pair, flanked by a Watson-Crick pair. Imino proton of the sheared G•A pair was also observed at 10.7 ppm,[27] as G17 in *c-kit87up*.

Analysis of Modified *c-kit87up* Sequences

We systematically probed the importance of different structural elements in the *c-kit87up* quadruplex fold by individual modifications of several residues (Table 2).

The single-residue loops A5 and C9 were substituted by T in the sequences *ck17* and *ck18* (Table 2), respectively, without altering the general fold, as suggested by their NMR spectra (Figure 6a). This was anticipated, as the structure of a single-residue double-chain-reversal loop bridging three G-tetrad layers has been shown to be independent of the nature of the base in it.[28]

Formation of the Watson-Crick A1•T12 base pair appears to be important for the structure, as the A1T mutation in the *ck16* sequence altered the general fold (Figure 6b). The *ck19* sequence containing the C11T mutation showed a doublings of imino proton peaks (Figure 6b). It is possible that for this sequence there are two conformations in which A1 can be paired with either T11 or T12.

Base pair alignments in the five-residue loop are also important for the fold since mutations that alter them (A16T, G17T, and A19T) would affect the general fold (Figure 6c). In this loop, however, the single-residue turn G18 could be substituted by a T in the *ck23* sequence, without altering the general fold (Figure 6a). This is not unexpected since this is the sole nucleotide in the loop not to be involved in tertiary interactions other than stacking onto the preceding guanine base. Substitution G20T (*ck24*) resulted in a similar general fold, except that the Watson-Crick A16•T20 may be formed instead of A16•G20. This possibility is structurally feasible. The G20I substitution (*ck26*) also resulted in a single conformation of the same general fold (Figure 6d), whereas the G17I substitution (*ck27*) maintained the general fold for the major conformation of only about 80%, indicating the significant role of the amino group of G17 in the G17•A19 pair (Figure 6d).

Finally, G10 should be important for the unique folding topology of *c-kit87up*, as NMR spectra (Figure 6b) suggest that the G10T mutation (*ck25*) favored an alternative G-quadruplex fold.

Discussion

We have solved the structure of an unprecedented intramolecular G-quadruplex formed by a promoter sequence from the *c-kit* oncogene in K^+ solution. Initial inspection of the sequence, with four equally sized G-tracts, suggested that the fold would be a straightforward parallel-type one, in accord with the biophysical data.[22] The present study has shown that this is not the case. The structure contains two single-residue double-chain-reversal loops spanning three G-tetrad layers, a robust type of loop reported previously[15,28] that could play an important role in the folds of many G-quadruplexes[29] in K^+ solution. This quadruplex is remarkable in requiring the active participation of 18 out of the 22 nucleotides in the structural organization. The sole exceptions are the single-nucleotide loops (A5 and C9) and turns (C11 and G18).

The structure also revealed several new topological elements. The most striking feature of the structure is the participation of an isolated non-G-tract guanine (G10) in formation of the G-tetrad core, despite the presence of the four G-tracts, each containing three consecutive guanines. This is counter-intuitive: generally, it was thought that G-tracts are most favorable for forming columns that support the G-tetrad core. This new folding principle, that G

residues in non-G-tract regions can participate in forming the structural core, should be kept in mind in future studies that attempt to predict G-quadruplex topologies from sequence data alone.

Another notable feature of the structure is the snapback parallel-stranded G-tetrad core. The snapback arrangement was first observed in the G-quadruplex formed by a five-G-tract sequence from the *c-myc* promoter.[15] The comparison between the snapback G-quadruplexes from *c-kit* and *c-myc* promoters is shown in Figure 7. In both cases, the G-tetrad core is interrupted[30] and base pairings in the loops are important in stabilizing the snapback scaffold. The difference between *c-kit* and *c-myc* quadruplexes is that the latter involves insertion of a single syn guanine (Figure 7c, d), the former involves insertion of two anti guanines (Figure 7a, b). The snapback feature of this structure would also allow for continuation of the DNA sequence in both directions without significant steric hindrance.

Formation of G-quadruplex structures in the *c-myc* promoter has been shown to inhibit the transcription of the *c-myc* oncogene.[12] The same approach may also apply to the *c-kit* oncogene. Because the G-quadruplex structure formed by *c-kit87up* is quite unique, it appears to be a specific target for drug design. In particular, there is a pronounced cleft in the structure between the bottom G-tetrad and the stem-loop that has a complex topology, suggesting that it could be a recognition site for ligands.

Methods

Sample Preparation

The unlabeled and the site-specific low-enrichment (2% ^{15}N-labeled) oligonucleotides were synthesized and purified as described previously.[24] Some site-specific ^{15}N-labeled samples were independently resynthesized to definitively verify assignments. Unless otherwise stated, the strand concentration of the NMR samples was typically 0.5-5 mM; the solutions contained 70 mM of KCl and 20 mM of potassium phosphate (pH 7).

NMR Spectroscopy

Experiments were performed on 600 MHz Varian and 800 MHz Bruker NMR spectrometers at 25 °C, unless otherwise specified. Resonances were assigned unambiguously by using site-specific low-enrichment labeling[24] and through-bond correlations at natural abundance.[25,26] Assignments for some residues were verified and confirmed in different independently synthesized labeled samples. Spectral assignments were also assisted and supported by COSY, TOCSY, and NOESY spectra.[26] Interproton distances were measured by using NOESY experiments at different mixing times.

Structure Calculation

The structures of the *c-kit87up* quadruplex were calculated using the X-PLOR program.[31] NMR-restrained molecular dynamics computations were performed essentially as described previously.[20a] Folding of the five-residue A16-G17-G18-A19-G20 loop was facilitated by hydrogen bond restraints on the Watson-Crick A16•G20 base pair. On the basis of the observation of a cross-peak between I20(H2) and A19(H2) in the G20I modified sample

(*ck26*), the position of A19(H2) was additionally restrained with respect to G20(N2). The loop configurations from independent folding computations were inspected visually. Two groups of folds emerged, differing by the position of G18. Selection for the G18 configuration that satisfies the observed upfield chemical shifts of H1′and H5′/H5′’ resulted in the well-converged loop structure (Figure 3).

Data Deposition

The coordinates for the *c-kit87up* quadruplex have been deposited in the Protein Data Bank (accession code 2O3M).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by NIH Grant GM34504 to D.J.P. and CRUK Programme Grant C129/A4489 to S.N. D.J.P. is a member of the New York Structural Biology Center supported by NIH Grant GM66354.

References

(1). Yarden Y, Kuang WJ, Yang-Feng T, Coussens L, Munemitsu S, Dull TJ, Chen E, Schlessinger J, Francke U, Ullrich A. EMBO J. 1987; 6:3341–3351. [PubMed: 2448137]

(2). Roskoski R Jr. Biochem. Biophys. Res. Commun. 2005; 337:1–13. [PubMed: 16129412]

(3). Taniguchi M, Nishida T, Hirota S, Isozaki K, Ito T, Nomura T, Matsuda H, Kitamura Y. Cancer Res. 1999; 59:4297–4300. [PubMed: 10485475]

(4). Kitamura Y, Hirota S, Nishida T. Mutation Res. 2001; 477:165–171. [PubMed: 11376697]

(5). Tian Q, Frierson HF Jr. Krystal GW, Moskaluk CA. Amer. J. Pathol. 1999; 154:1643–1647. [PubMed: 10362788]

(6). Tarn C, Godwin AK. Curr. Treat. Options Oncol. 2005; 6:473–486. [PubMed: 16242052]

(7). Tuveson DA, Willis NA, Jacks T, Griffin JD, Singer S, Fletcher CDM, Fletcher JA, Demetri GD. Oncogene. 2001; 20:5054–5058. [PubMed: 11526490]

(8). von Mehren M, Watson JC. Hematol. Oncol. Clin. North Am. 2005; 19:547–564. [PubMed: 15939196]

(9). Schittenhelm MM, Shiraga S, Schroeder A, Corbin AS, Griffith D, Lee FY, Bokemeyer C, Deininger MW, Druker BJ, Heinrich MC. Cancer Res. 2006; 66:473–481. [PubMed: 16397263]

(10). Prenen H, Cools J, Mentens N, Folens C, Sciot R, Schöffski P, Van Oosterom A, Marynen P, Debiec-Rychter M. Clin. Cancer Res. 2006; 12:2622–2627. [PubMed: 16638875]

(11). Simonsson T, Pecinka P, Kubista M. Nucleic Acids Res. 1998; 26:1167–1172. [PubMed: 9469822]

(12). Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Proc. Natl. Acad. Sci. U.S.A. 2002; 99:11593–11598. [PubMed: 12195017]

(13). Cooney M, Czernuszewicz G, Postel EH, Flint SJ, Hogan ME. Science. 1988; 241:456–459. [PubMed: 3293213]

(14). Phan AT, Modi YS, Patel DJ. J. Am. Chem. Soc. 2004; 126:8710–8716. (a). [PubMed: 15250723] (b) Ambrus A, Chen D, Dai J, Jones RA, Yang D. Biochemistry. 2005; 44:2048–2058. [PubMed: 15697230]

(15). Phan AT, Kuryavyi V, Gaw HY, Patel DJ. Nat. Chem. Biol. 2005; 1:167–173. [PubMed: 16408022]

(16). Smith FW, Feigon J. Nature. 1992; 356:164–168. (a). [PubMed: 1545871] (b) Wang Y, Patel DJ. J. Mol. Biol. 1995; 251:76–94. [PubMed: 7643391] (c) Smith FW, Schultze P, Feigon J. Structure. 1995; 3:997–1008. [PubMed: 8590010]

(17). Wang Y, Patel DJ. Structure. 1993; 1:263–282. [PubMed: 8081740]

(18). Parkinson GN, Lee MPH, Neidle S. Nature. 2002; 417:876–880. [PubMed: 12050675]

(19). Wang Y, Patel DJ. Structure. 1994; 2:1141–1155. [PubMed: 7704525]

(20). Luu KN, Phan AT, Kuryavyi V, Lacroix L, Patel DJ. J. Am. Chem. Soc. 2006; 128:9963–9970. (a). [PubMed: 16866556] (b) Phan AT, Luu KN, Patel DJ. Nucleic Acids Res. 2006; 34:5715–5719. [PubMed: 17040899] (c) Ambrus A, Chen D, Dai J, Bialis T, Jones RA, Yang D. Nucleic Acids Res. 2006; 34:2723–2735. [PubMed: 16714449] (d) Xu Y, Noguchi Y, Sugiyama H. Bioorg. Med. Chem. 2006; 14:5584–5591. [PubMed: 16682210] (e) Zhang N, Phan AT, Patel DJ. J. Am. Chem. Soc. 2005; 127:17277–17285. [PubMed: 16332077]

(21). Phan AT, Patel DJ. J. Am. Chem. Soc. 2003; 125:15021–15027. (a). [PubMed: 14653736] (b) Phan AT, Modi YS, Patel DJ. J. Mol. Biol. 2004; 338:93–102. [PubMed: 15050825] (c) Haider S, Parkinson GN, Neidle S. J. Mol. Biol. 2002; 320:189–200. [PubMed: 12079378] (d) Parkinson GN, Ghosh R, Neidle S. Biochemistry. 2007; 46:2390–2397. [PubMed: 17274602]

(22). Rankin S, Reszka AP, Huppert J, Zloh M, Parkinson GN, Todd AK, Ladame S, Balasubramanian S, Neidle S. J. Am. Chem. Soc. 2005; 127:10584–10589. [PubMed: 16045346]

(23). Fernando H, Reszka AP, Huppert J, Ladame S, Rankin S, Venkitaraman AR, Neidle S, Balasubramanian S. Biochemistry. 2006; 45:7854–7860. [PubMed: 16784237]

(24). Phan AT, Patel DJ. J. Am. Chem. Soc. 2002; 124:1160–1161. [PubMed: 11841271]

(25). Phan AT. J. Biomol. NMR. 2000; 16:175–178. [PubMed: 10723997]

(26). Phan AT, Guéron M, Leroy JL. Methods Enzymol. 2001; 338:341–371. [PubMed: 11460557]

(27). Zhu L, Chou SH, Xu J, Reid BR. Nat. Struct. Biol. 1995; 2:1012–1017. [PubMed: 7583654]

(28). Phan AT, Kuryavyi V, Ma JB, Faure A, Andréola ML, Patel DJ. Proc. Natl. Acad. Sci. U.S.A. 2005; 102:634–639. [PubMed: 15637158]

(29). Todd AK, Johnston M, Neidle S. Nucleic Acids Res. 2005; 33:2901–2907. (a). [PubMed: 15914666] (b) Huppert JL, Balasubramanian S. Nucleic Acids Res. 2005; 33:2908–2916. [PubMed: 15914667] (c) Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Nucleic Acids Res. 2006; 34:5402–5415. [PubMed: 17012276]

(30). Phan, AT.; Kuryavyi, V.; Luu, KN.; Patel, DJ. Quadruplex Nucleic Acids. Neidle, S.; Balasubramanian, S., editors. Royal Society of Chemistry; U.K.: 2006. p. 81-99.

(31). Brünger, AT. X-PLOR: A system for X-ray crystallography and NMR. Yale University Press; New Haven, CT: 1992.

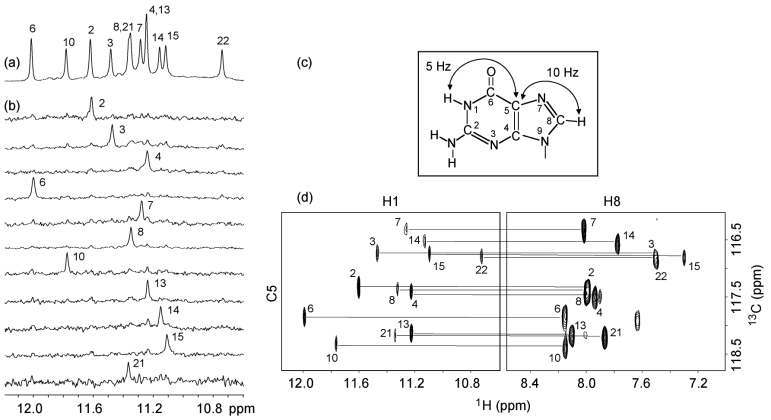

Figure 1.

Spectra and assignments of *c-kit87up*. (a) The 600 MHz imino proton spectrum of the *c-kit87up* sequence in K^+ solution at 25 °C. Unambiguous assignments are indicated. (b) Imino protons were assigned in ^{15}N-filtered spectra of samples, 2% ^{15}N-labeled at the indicated positions. (c) Long-range *J* couplings in a guanine used for imino-H8 correlation. (d) H8 proton assignments of *c-kit87up* by through-bond correlations between imino and H8 protons via ^{13}C5 at natural abundance, using long-range *J* couplings shown in (c).

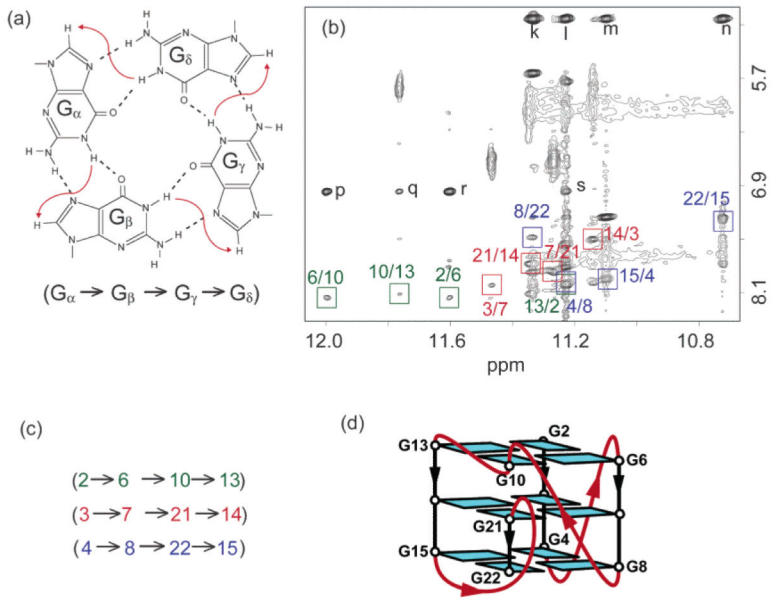

Figure 2.

G-tetrad alignments and folding topology of *c-kit87up*. (a) Schematic of a G-tetrad with characteristic guanine imino-H8 connectivity patterns indicated by arrows. (b) NOESY spectrum (mixing time, 200 ms). The characteristic guanine imino-H8 cross-peaks for G-tetrads are framed and labeled with the imino proton assignment in the first position and that of the H8 proton in the second position. Peaks k, l, m, and n show close distances between G20(NH_2) and the imino protons of guanines from the bottom tetrad; peaks p, q, r, and s show close distances between A1(H2) and the imino protons of guanines from the top tetrad. (c) Summary of characteristic guanine imino-H8 patterns observed for *c-kit87up*. (d) Topology of *c-kit87up* based on the G-tetrad alignments in (c).

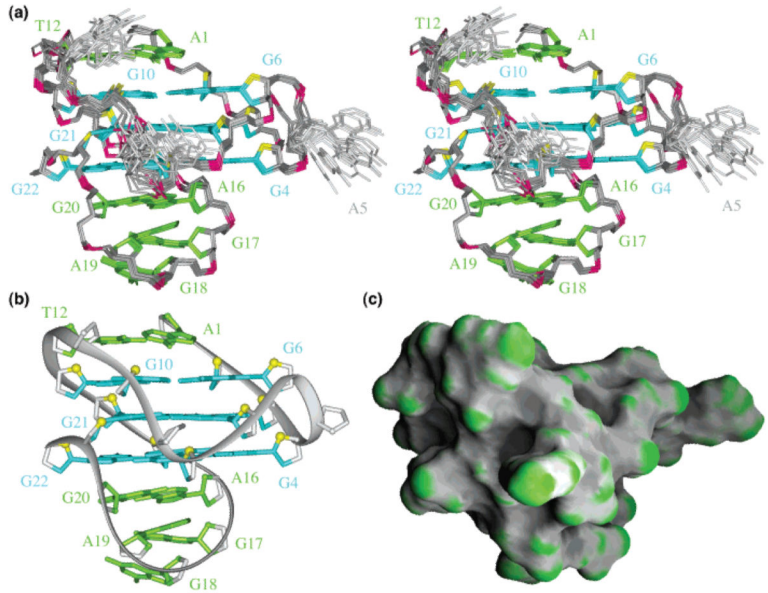

Figure 3.

Solution structure of the *c-kit87up* quadruplex in K^+ solution. (a) Stereoview of 11 superpositioned refined structures. Guanine bases in the G-tetrad core are colored cyan; other bases, green; the backbone, gray; phosphorus atoms, red; O4′atoms, yellow. (b) Ribbon and (c) surface views of a representative refined structure.

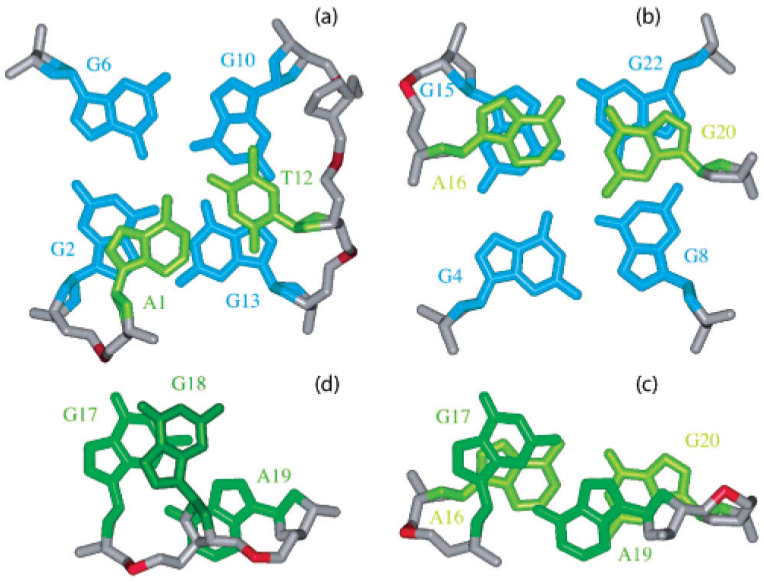

Figure 4.

Base pairs in loops. Stacking (a) between the Watson-Crick A1•T12 pair and the G2•G6•G10•G13 tetrad, (b) between the Watson-Crick A16•G20 pair and the G4•G8•G22•G15 tetrad, (c) between the Watson-Crick A16•G20 pair and the sheared G17•A19 pair, and (d) between the sheared G17•A19 pair and the single-base G18.

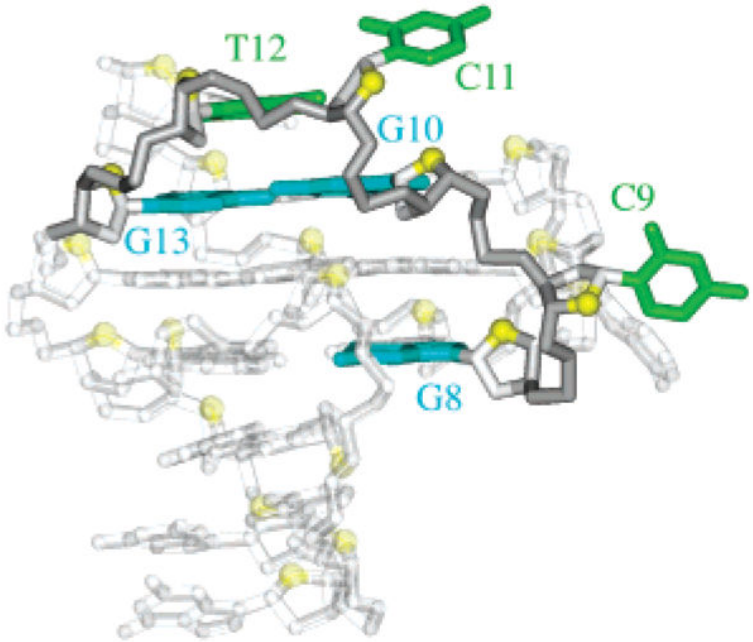

Figure 5.
Highlight of the C9-G10-C11-T12 fragment within the *c-kit87up* quaduplex.

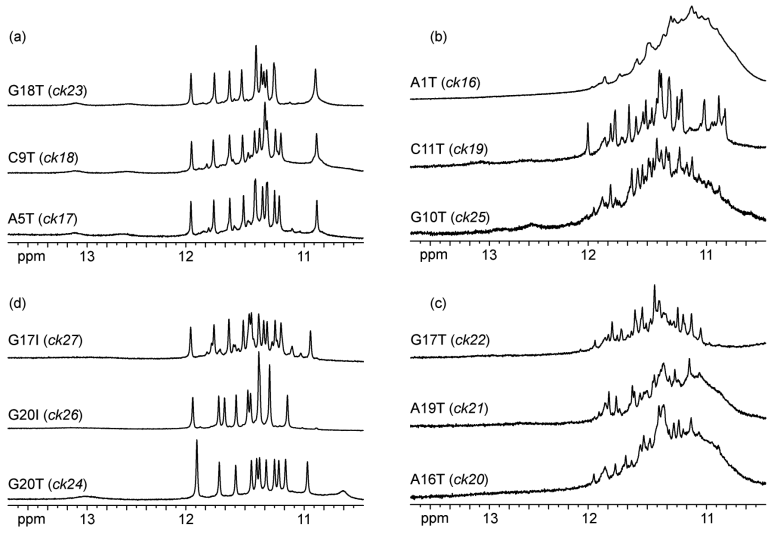

Figure 6.

The 600 MHz imino proton spectra of the modified *c-kit87up* sequences in K^+ solution at 25 °C.

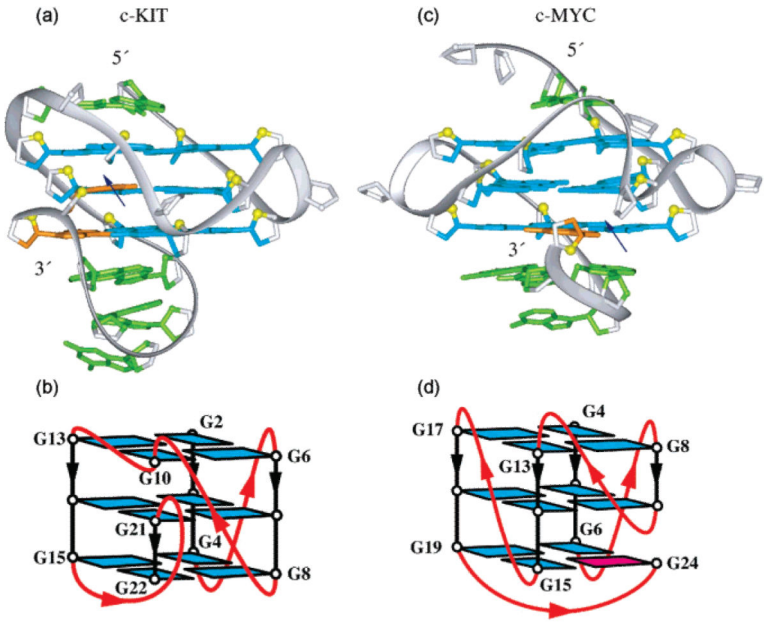

Figure 7.

Comparison between G-quadruplex structures observed in (a, b) *c-kit* (this work) and (c, d) *c-myc*[15] promoters. In (a) and (c): color coded as in Figure 3, except that snapback residues are colored orange; the blue arrows indicate break points within interrupted G-tetrad cores. In (b) and (d): loops are colored red; anti and syn guanines are colored cyan and magenta, respectively.

Table 1

NMR Restraints and Structure Statistics

A. NMR Restraints		
distance restraints	nonexchangeable	exchangeable
intra-residue distance restraints	200	0
sequential (i, i + 1) distance restraints	72	6
long-range (i, ≥ i + 2) distance restraints	13	38
other restraints		
hydrogen bond restraints		
(H–N, H–O, and heavy atoms)	60	
torsion angle restraints	54	
intensity restraints		
nonexchangeable protons	240	
B. Statistics for 11 Structures following Intensity Refinement		
NOE violations		
number (>0.2 Å)	0.182 ± 0.404	
maximum violation (Å)	0.284 ± 0.024	
rmsd of violations	0.018 ± 0.004	
deviations from the ideal covalent geometry		
bond lengths (Å)	0.004 ± 0.000	
bond angles (deg)	0.974 ± 0.018	
impropers (deg)	0.349 ± 0.008	
NMR R-factor ($R_{1/6}$)	0.017 ± 0.005	
pairwise all heavy atom rmsd values		
all heavy atoms except A5, C9, C11	0.50 ± 0.09	
all heavy atoms	0.83 ± 0.12	

Table 2

Investigated G-Rich Sequences in the *c-kit* Promoter[a]

	sequence									
name		**1**		**2**		**3**		**4**		
c-kit87up	A	GGG	A	GGG	C	G	CT	GGG	AGGAG	GG
ck16	**T**	GGG	A	GGG	C	G	CT	GGG	AGGAG	GG
ck17	A	GGG	**T**	GGG	C	G	CT	GGG	AGGAG	GG
ck18	A	GGG	A	GGG	**T**	G	CT	GGG	AGGAG	GG
ck19	A	GGG	A	GGG	C	G	**TT**	GGG	AGGAG	GG
ck20	A	GGG	A	GGG	C	G	CT	GGG	**T**GGAG	GG
ck21	A	GGG	A	GGG	C	G	CT	GGG	AGG**T**G	GG
ck22	A	GGG	A	GGG	C	G	CT	GGG	A**T**GAG	GG
ck23	A	GGG	A	GGG	C	G	CT	GGG	AG**T**AG	GG
ck24	A	GGG	A	GGG	C	G	CT	GGG	AGGA**T**	GG
ck25	A	GGG	A	GGG	C	**T**	CT	GGG	AGGAG	GG
ck26	A	GGG	A	GGG	C	G	CT	GGG	AGGA**I**	GG
ck27	A	GGG	A	GGG	C	G	CT	GGG	A**I**GAG	GG

[a]Modifications are in boldface. Loop numbers are listed in the column head.