

## A G-Rich Sequence within the *c-kit* Oncogene Promoter Forms a Parallel G-Quadruplex Having Asymmetric G-Tetrad Dynamics

Shang-Te Danny Hsu,<sup>†</sup> Peter Varnai,<sup>‡</sup> Anthony Bugaut,<sup>†</sup> Anthony P. Reszka,<sup>§</sup>  
Stephen Neidle,<sup>\*,§</sup> and Shankar Balasubramanian<sup>\*,†,⊥</sup>

University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom, School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SP, United Kingdom, Department of Chemistry and Biochemistry, University of Sussex, Falmer Brighton BN1 9QJ, United Kingdom, and The Cancer Research UK, Biomolecular Structure Group, The School of Pharmacy, University of London, 29-39 Brunswick Square, London WC1N 1AX, United Kingdom

Received May 22, 2009; E-mail: stephen.neidle@pharmacy.ac.uk; sb10031@cam.ac.uk

**Abstract:** Guanine-rich DNA sequences with the ability to form quadruplex structures are enriched in the promoter regions of protein-coding genes, particularly those of proto-oncogenes. G-quadruplexes are structurally polymorphic and their folding topologies can depend on the sample conditions. We report here on a structural study using solution state NMR spectroscopy of a second G-quadruplex-forming motif (*c-kit2*) that has been recently identified in the promoter region of the *c-kit* oncogene. In the presence of potassium ions, *c-kit2* exists as an ensemble of structures that share the same parallel-stranded propeller-type conformations. Subtle differences in structural dynamics have been identified using hydrogen–deuterium exchange experiments by NMR spectroscopy, suggesting the coexistence of at least two structurally similar but dynamically distinct substates, which undergo slow interconversion on the NMR timescale.

### Introduction

DNA and RNA sequences containing repeated short guanine-rich tracts can form higher-order structures termed G-quadruplexes in which the underlying motif is four in-plane guanine bases hydrogen-bonded together to form a G-tetrad.<sup>1</sup> G-quadruplex arrangements can be tetramolecular, bimolecular, or intramolecular. A large number of quadruplex structures are possible by virtue of the variation in strand polarities and also the sequence and topologies of the loops (for bimolecular or intramolecular quadruplexes). Intramolecular G-quadruplexes typically comprise several G-tetrads stacked one on top of each other with intervening looped-out sequences.<sup>2,3</sup> The human telomeric intramolecular DNA G-quadruplex is the best studied quadruplex system and was first characterized in Na<sup>+</sup> solution by NMR to reveal an antiparallel topology with one diagonal and two lateral loops.<sup>4</sup> The crystal structure of the same system in K<sup>+</sup> showed an all-parallel topology with three chain-reversal loops.<sup>5</sup> More recently, studies have shown that the telomeric sequence can form a mixed (3 + 1) structure comprising a single

chain-reversal loop and two lateral loops.<sup>6</sup> Such studies have shown that G-quadruplex topologies can be sensitive to the experimental conditions. Quadruplex nucleic acid sequences have been located in human and other genomes,<sup>7,8</sup> with noteworthy overrepresentation in promoter sequences of genes involved in cellular proliferation.<sup>9–11</sup> As yet, the topology and molecular structures have been characterized for only a few oncogene promoter sequences that form G-quadruplex structures.<sup>12–14</sup> In contrast with the human telomeric G-quadruplex which has three trinucleotide loops, promoter G-quadruplexes appear to generally have one or more short loops that are predisposed to forming chain-reversal loops and promoting parallel quadruplex topology.<sup>15,16</sup> The *c-myc* oncogene has been studied in detail,<sup>17,18</sup> as the paradigm for promoter G-quadruplexes, with solution structures being reported for several G-quadruplexes using nuclear magnetic resonance (NMR) spectroscopy,<sup>19,20</sup> from within the NHE III<sub>1</sub> element of the *c-myc*

<sup>†</sup> University Chemical Laboratory, University of Cambridge.

<sup>‡</sup> University of Sussex.

<sup>§</sup> University of London.

<sup>⊥</sup> School of Clinical Medicine, University of Cambridge.

- (1) Davis, J. T. *Angew. Chem., Int. Ed. Engl.* **2004**, *43*, 668–98.
- (2) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402–15.
- (3) Patel, D. J.; Phan, A. T.; Kuryavyi, V. *Nucleic Acids Res.* **2007**, *35*, 7429–55.
- (4) Wang, Y.; Patel, D. J. *J. Mol. Biol.* **1993**, *234*, 1171–83.
- (5) Parkinson, G. N.; Lee, M. P.; Neidle, S. *Nature* **2002**, *417*, 876–80.

(6) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963–70.

(7) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908–16.

(8) Todd, A. K.; Johnston, M.; Neidle, S. *Nucleic Acids Res.* **2005**, *33*, 2901–7.

(9) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2007**, *35*, 406–13.

(10) Verma, A.; Halder, K.; Halder, R.; Yadav, V. K.; Rawal, P.; Thakur, R. K.; Mohd, F.; Sharma, A.; Chowdhury, S. *J. Med. Chem.* **2008**, *51*, 5641–9.

(11) Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2008**, *36*, 2700–4.

(12) Neidle, S. *Curr. Opin. Struct. Biol.* **2009**, *19*, 239–50.

(13) Qin, Y.; Hurley, L. H. *Biochimie* **2008**, *90*, 1149–71.

(14) Dai, J.; Chen, D.; Jones, R. A.; Hurley, L. H.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 5133–44.

promoter. It has been suggested that selective targeting of a promoter oncogene quadruplex by a small molecule may stabilize its tertiary structure, and that this may be of therapeutic utility.<sup>17,21</sup> Down-regulation of *c-myc* expression has been demonstrated using the quadruplex-binding porphyrin molecule TMPyP4<sup>18</sup> and the quindoline derivative SYUIQ-5,<sup>22</sup> and the supercoiling implications of G-quadruplex formation within a duplex DNA environment have been recently explored.<sup>23</sup> The G-quadruplex promoter targeting approach offers an alternative to the direct drug targeting of oncogenic kinase protein active sites, which has often resulted in severe clinical problems of resistance to these drugs, usually via active-site mutations.

The proto-oncogene *c-kit* codes for a 145–160 kDa tyrosine kinase receptor, which regulates several important signal transduction cascades that control cell growth and proliferation.<sup>24–26</sup> Functionally significant mutations and overexpression of the *c-kit* protein play a role in oncogenic transformation of certain cell types and are found in several highly malignant human cancers, notably gastrointestinal stromal tumors (GIST), in which a majority of GIST cells contain activating mutations in *c-kit*.<sup>27</sup> The drug imatinib has recently been licensed for clinical use in the treatment of GIST; its mode of action involves binding in the ATP pocket of the *c-kit* kinase domain and maintaining it in an inactive conformation.<sup>28–30</sup> This directly results in significant tumor regression, but resistance to the drug eventually occurs in almost all patients.<sup>31,32</sup> Newer kinase inhibitors have been developed, such as the compound sunitinib,<sup>33,34</sup> however, the heterogeneity of kinase mutations and of resistance mech-

anisms suggests that the clinical challenges of resistance still remain to be overcome.<sup>35</sup> We have used small molecules to target G-quadruplex DNA structures in the *c-kit* gene,<sup>36</sup> and have recently shown that cell growth can be halted in a panel of clinically derived GIST cell lines accompanied by a >90% decrease in *c-kit* expression.<sup>37</sup> Detailed structural descriptions of the G-quadruplexes in the *c-kit* gene will therefore be of great interest for structure-based small molecule developments.

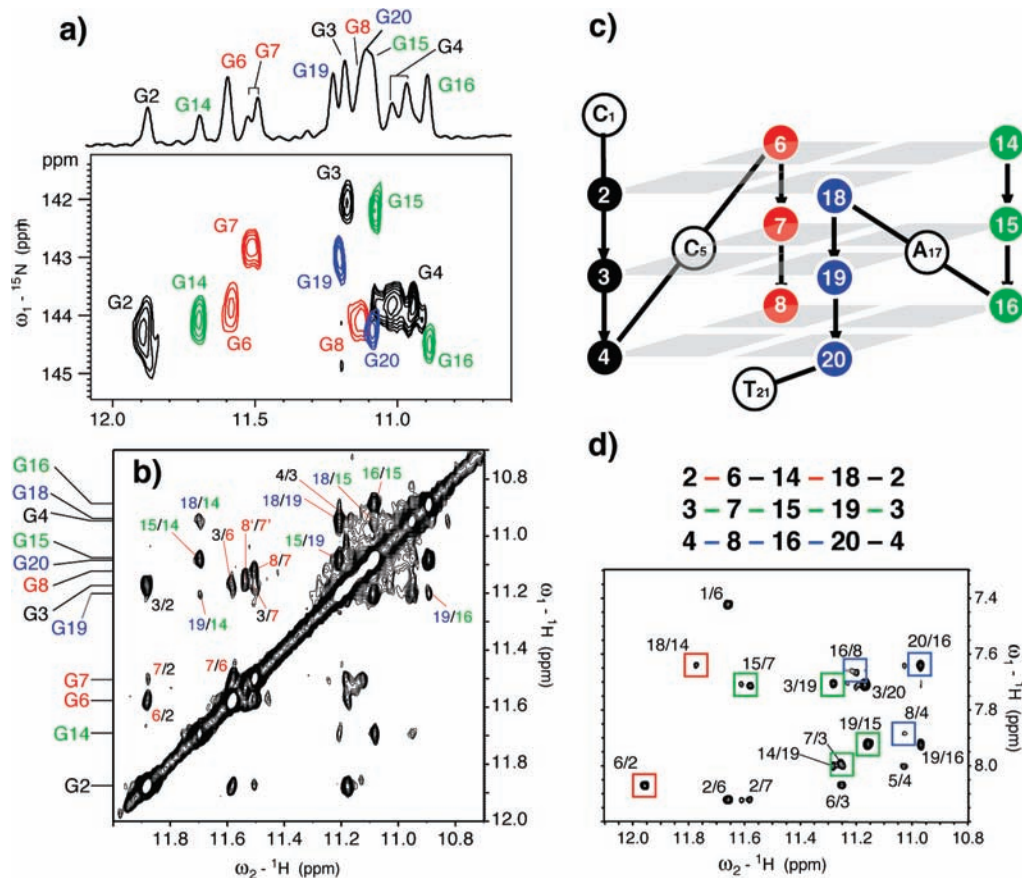
We have previously identified two sequences (*c-kit1* and *c-kit2*) that occur in a nuclease hypersensitive region of the promoter<sup>38</sup> of the *c-kit* oncogene, upstream of the transcription start site, and have shown them as being capable of forming G-quadruplex structures.<sup>39,40</sup> The three-dimensional (3D) structure of the *c-kit1* G-quadruplex has been determined by solution state NMR spectroscopy and shown to form a unique topology.<sup>41,42</sup> We report here on the topology and structure of the *c-kit2* quadruplex, whose sequence we have previously shown to promote the formation of nonduplex structure within the context of a natural DNA duplex sequence.<sup>43</sup> The *c-kit2* G-quadruplex adopts an all-parallel propeller-like topology with clear evidence of structural polymorphism, manifested by the presence of typically two sets of resonances in the imino proton NMR spectra, corresponding to regions of the tetrads that are asymmetrically distributed. The same asymmetry has been characterized in terms of local dynamics through hydrogen–deuterium exchange (HDX) experiments based on which the presence of at least two dynamically distinct (but structurally very similar) populations are identified.

## Results

***c-kit2* Adopts a Parallel G-Quadruplex Topology with a Long Loop.** We have previously employed solution state NMR spectroscopy to show that the 21-nucleotide DNA sequence from *c-kit2*, 5'-CG<sub>3</sub>CG<sub>3</sub>CGCGAG<sub>3</sub>AG<sub>4</sub>-3', contains highly protected imino protons that we attributed as evidence for G-quadruplex structure formation.<sup>40</sup> The large number (i.e., >12) of observed imino proton resonances in the wild-type (wt) sequence however, suggests that the *c-kit2* quadruplex is structurally heterogeneous due to the four consecutive guanines at the 3'-end, each of which could in principle participate in G-tetrad formation. Indeed, a single nucleotide replacement at the last position, designated as G21T, gave rise to a much more simplified imino proton NMR spectrum, indicative of fewer

- (15) Hazel, P.; Parkinson, G. N.; Neidle, S. *J. Am. Chem. Soc.* **2006**, *128*, 5480–7.
- (16) Bugaut, A.; Balasubramanian, S. *Biochemistry* **2008**, *47*, 689–97.
- (17) Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.* **1998**, *26*, 1167–72.
- (18) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593–8.
- (19) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Am. Chem. Soc.* **2004**, *126*, 8710–6.
- (20) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. *Biochemistry* **2005**, *44*, 2048–58.
- (21) Hurley, L. H.; Von Hoff, D. D.; Siddiqui-Jain, A.; Yang, D. *Semin. Oncol.* **2006**, *33*, 498–512.
- (22) Liu, J. N.; Deng, R.; Guo, J. F.; Zhou, J. M.; Feng, G. K.; Huang, Z. S.; Gu, L. Q.; Zeng, Y. X.; Zhu, X. F. *Leukemia* **2007**, *21*, 1300–2.
- (23) Sun, D.; Hurley, L. H. *J. Med. Chem.* **2009**, *52*, 2863–74.
- (24) Yarden, Y.; Kuang, W. J.; Yang-Feng, T.; Coussens, L.; Munemitsu, S.; Dull, T. J.; Chen, E.; Schlessinger, J.; Francke, U.; Ullrich, A. *EMBO J.* **1987**, *6*, 3341–51.
- (25) Roskoski, R., Jr. *Biochem. Biophys. Res. Commun.* **2005**, *338*, 1307–15.
- (26) Yang, J.; Du, X.; Lazar, A. J.; Pollock, R.; Hunt, K.; Chen, K.; Hao, X.; Trent, J.; Zhang, W. *Cancer* **2008**, *113*, 1532–43.
- (27) Hoeben, A.; Schoffski, P.; Debiec-Rychter, M. *Br. J. Cancer* **2008**, *98*, 684–8.
- (28) Tuveson, D. A.; Willis, N. A.; Jacks, T.; Griffin, J. D.; Singer, S.; Fletcher, C. D.; Fletcher, J. A.; Demetri, G. D. *Oncogene* **2001**, *20*, 5054–8.
- (29) Zhu, M. J.; Ou, W. B.; Fletcher, C. D.; Cohen, P. S.; Demetri, G. D.; Fletcher, J. A. *Oncogene* **2007**, *26*, 6386–95.
- (30) Mol, C. D.; Dougan, D. R.; Schneider, T. R.; Skene, R. J.; Kraus, M. L.; Scheibe, D. N.; Snell, G. P.; Zou, H.; Sang, B. C.; Wilson, K. P. *J. Biol. Chem.* **2004**, *279*, 31655–63.
- (31) Heinrich, M. C.; Corless, C. L.; Demetri, G. D.; Blanke, C. D.; von Mehren, M.; Joensuu, H.; McGreevey, L. S.; Chen, C. J.; Van den Abbeele, A. D.; Druker, B. J.; Kiese, B.; Eisenberg, B.; Roberts, P. J.; Singer, S.; Fletcher, C. D.; Silberman, S.; Dimitrijevic, S.; Fletcher, J. A. *J. Clin. Oncol.* **2003**, *21*, 4342–9.
- (32) Heinrich, M. C.; Corless, C. L.; Blanke, C. D.; Demetri, G. D.; Joensuu, H.; Roberts, P. J.; Eisenberg, B. L.; von Mehren, M.; Fletcher, C. D.; Sandau, K.; McDougall, K.; Ou, W. B.; Chen, C. J.; Fletcher, J. A. *J. Clin. Oncol.* **2006**, *24*, 4764–74.
- (33) von Mehren, M. *Clin. Colorectal Cancer* **2006**, *6*, S30–34.

- (34) Le Tourneau, C.; Raymond, E.; Faivre, S. *Ther. Clin. Risk Manage.* **2007**, *3*, 341–8.
- (35) Liegl, B.; Kepten, I.; Le, C.; Zhu, M.; Demetri, G. D.; Heinrich, M. C.; Fletcher, C. D.; Corless, C. L.; Fletcher, J. A. *J. Pathol.* **2008**, *216*, 64–74.
- (36) Bejugam, M.; Sewitz, S.; Shirude, P. S.; Rodriguez, R.; Shahid, R.; Balasubramanian, S. *J. Am. Chem. Soc.* **2007**, *129*, 12926–7.
- (37) Gunaratnam, M.; Swank, S.; Haider, S. M.; Galesa, K.; Reszka, A. P.; Beltran, M.; Cuenca, F.; Fletcher, J. A.; Neidle, S. *J. Med. Chem.* **2009**, *52*, 3774–83.
- (38) Yamamoto, K.; Tojo, A.; Aoki, N.; Shibuya, M. *Jpn. J. Cancer Res.* **1993**, *84*, 1136–44.
- (39) Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.* **2005**, *127*, 10584–9.
- (40) Fernando, H.; Reszka, A. P.; Huppert, J.; Ladame, S.; Rankin, S.; Venkitaraman, A. R.; Neidle, S.; Balasubramanian, S. *Biochemistry* **2006**, *45*, 7854–60.
- (41) Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle, S.; Patel, D. J. *J. Am. Chem. Soc.* **2007**, *129*, 4386–92.
- (42) Todd, A. K.; Haider, S. M.; Parkinson, G. N.; Neidle, S. *Nucleic Acids Res.* **2007**, *35*, 5799–808.
- (43) Shirude, P. S.; Okumus, B.; Ying, L.; Ha, T.; Balasubramanian, S. *J. Am. Chem. Soc.* **2007**, *129*, 7484–5.



**Figure 1.** Imino proton resonance assignments and G-quadruplex topology determination aided by site-specific  $^{15}\text{N}$  labeling. (a) Overlaid 2D  $^1\text{H}$ – $^{15}\text{N}$  correlation spectra of site-specifically  $^{15}\text{N}$  labeled G21T oligonucleotides, recorded at 700 MHz and at 298K. Nucleotides that correspond to S1, S2, S3 and S4 are colored black, red, green and blue, respectively. The site-specific labeling enables unambiguous assignments of the 1D imino proton spectrum of G21T as shown on top of this figure. The presence of structural polymorphism is clearly visible for G4 and G7 in the 1D spectrum, and for G2 and G4 in the 2D spectrum. (b) Imino proton region of the 2D  $^1\text{H}$  NOESY spectrum of G21T. The assignments of the resolved cross-peaks are made through the imino proton resonance assignments obtained in panel (a) which are consistent with a G-quadruplex in a parallel conformation as shown in panel (c). (d) Interstrand NOEs between imino H1 (horizontal axis) and aromatic H8 (vertical axis) of unimolecular guanine bases within the same layer of G-tetrads (boxed in red, green and blue for the top, middle and bottom G-tetrads, respectively). Notably, G7H1 exhibits a minor conformation resulting in clear doubling. The expected connectivities between individual guanines are shown on top of this panel as dashes connecting the corresponding sequence numbers. Those that are observed in the NOESY spectrum are colored, accordingly. There are three cross-peaks (2/6, 2/7 and 14/19) that do not correspond to the unimolecular all-parallel topology and must arise due to the presence of an alternative conformation (see main text for details). The mixing time of this spectrum was 400 ms to yield better signals through spin-diffusion.

G-quadruplex structures; the total number of observed imino proton resonances remained in excess of the 12 that would have been expected from three G-tetrads (four imino protons per tetrad) of a single structure. The observed structural heterogeneity appeared to be kinetically driven and was insensitive to the amount of  $\text{K}^+$  ions present (from 10 to 100 mM) and DNA concentrations (from 100  $\mu\text{M}$  to 1 mM). After a prolonged incubation time, the imino proton NMR spectrum of the wt *c-kit2* sample eventually shows 12 expected imino proton resonances, as does the G21T sample, while the time required for the latter to reach apparent homogeneity is much shorter (see Figure S1 in Supporting Information, SI). Despite the structural heterogeneity, the unimolecular conformations in wt *c-kit2* and G21T are very similar as shown by the superimposable crosspeaks in the nuclear Overhauser effect (NOE) spectra (Figure S2). There were minor differences in the experimental conditions used here compared to our previous study,<sup>40</sup> which include the DNA concentration (100–200  $\mu\text{M}$  instead of 1 mM), buffer concentration (10 mM instead of 20 mM) and temperature (25 °C instead of 43 °C). The slight differences in the wt *c-kit2*

NMR spectrum of the previously observed,<sup>40</sup> and that in the current study, may be a consequence of different sample conditions.

We have therefore chosen to focus on the structural characterization of the dominant population and have obtained the assignments of the resolved imino proton resonances in G21T through site-specific  $^{15}\text{N}$ -labelings (Figure 1a).<sup>44</sup> With an apparent concentration of the  $^{15}\text{N}$ -labeled guanines in each NMR sample of just 8  $\mu\text{M}$ , the use of 2D  $^{15}\text{N}$  band-selective optimized flip-angle short transient heteronuclear multiple quantum coherence (SOFAST-HMQC) spectroscopy<sup>45</sup> allows high resolution NMR data to be recorded within a few hours and effectively resolves all the ambiguities along the  $^1\text{H}$  dimension due to partial overlaps for G8, G15 and G20 by separating the overlapped cross-peaks along the  $^{15}\text{N}$  dimension (Figure 1a). Notably, the imino groups of G2, G4 and G7 exhibit clear satellite peaks that have distinct chemical shifts along the  $^1\text{H}$  dimension, and

(44) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2002**, *124*, 1160–1.

(45) Schanda, P.; Kupce, E.; Brutscher, B. *J. Biomol. NMR* **2005**, *33*, 199–211.

**Table 1.** Sequences of *c-kit2* Variants Used in This Study<sup>a</sup>

name	5'	S1	L1	S2	L2	S3	L3	S4	3'
wt	C	GGG	C	GGG	CGCGA	GGG	A	GGG	G
G21T	C	GGG	C	GGG	CGCGA	GGG	A	GGG	T
G21T/C9T	C	GGG	C	GGG	<b>TGCGA</b>	GGG	A	GGG	<b>T</b>
G21T/G10A	C	GGG	C	GGG	<b>CACGA</b>	GGG	A	GGG	<b>T</b>
G21T/C11T	C	GGG	C	GGG	<b>CGTGA</b>	GGG	A	GGG	<b>T</b>
G21T/A17T	C	GGG	C	GGG	CGCGA	GGG	<b>T</b>	GGG	<b>T</b>

<sup>a</sup> Changes in sequence are highlighted in bold; S and L stand for strands and loops, respectively.

for G2 in particular, the <sup>15</sup>N dimension. Together, these suggest the presence of a mixture of conformations in the NMR samples of G21T.

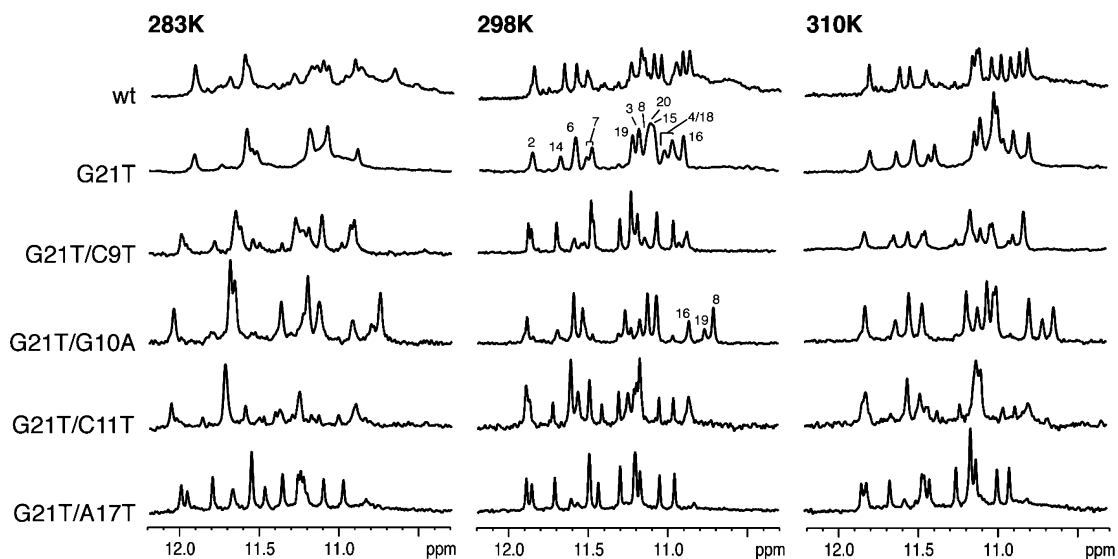
On the basis of the imino proton resonance assignments, we have assigned individual NOE correlations observed in the imino region of the 2D <sup>1</sup>H NOESY spectrum of G21T (Figure 1b). In addition, we have identified inter-strand NOEs (Figure 1d) that are consistent with a parallel G-quadruplex topology in which the four G strands are connected via two single-nucleotide propeller loops, loops 1 and 3 (L1 and L3), and a long loop 2 (L2) encompassing five nucleotides, CGCGA, through double chain-reversal (Figure 1c). Additionally, we have identified key NOE connectivities corresponding to interstrand correlations between imino proton H1 and aromatic proton H8 (Figure 1d). Of particular importance are the cross-peak of G3(H1)xG19(H8) between strands 1 and 4 (S1 and S4), and those of G7(H1)xG15(H8) and G8(H1)xG16(H8) between strands 2 and 3 (S2 and S3; boxed in Figure 1d), which essentially define the all-parallel topology together with the expected double chain-reversal configurations for L1 and L3. A much longer mixing time (400 ms) is required to obtain these key NOEs, in part due to the intrinsic dynamics of L2, and also the lack of covalent linkage to connect S1 and S4. Even with such a prolonged mixing time, the expected correlations of G14H8/G6H1 and G2H8/G18H1 from the top G-tetrad and that of G4H8/G20H1 from the bottom G-tetrad remain unidentified. Additionally, after assigning most of the resolved (more than 260) through-space <sup>1</sup>H–<sup>1</sup>H NOEs in the 2D NOESY spectra with various mixing times at different temperatures, we have been unable to identify a significant number of NOEs originated from L2, most likely due to conformational heterogeneity and/or motions that lead to NMR line broadening. The cross-peaks G6(H1)xG2(H8), G7(H1)xG2(H8) and G19(H1)xG14(H8) in the NOESY spectra (Figure 1d) do not correspond to our unimolecular all-parallel-quadruplex and can be attributed to an alternative conformation that is also present in the solution. In order to clarify NOE assignments due to the contributions of mixed populations, we introduced nucleotide replacements in other loops, in addition to the G21T replacement.

**There Is Dynamical Coupling between the G-Tetrads and the Long Loop 2.** Four additional *c-kit2* variants were synthesized (Table 1), each of which contains a single nucleotide replacement in addition to the G21T replacement, for NMR spectral comparisons. Comparison of the 1D <sup>1</sup>H spectra as well as the 2D <sup>1</sup>H NOESY and TOCSY spectra of all the *c-kit2* variants shows marginal spectral changes in terms of the relative chemical shift changes of the nonlabile protons, particularly for H6/H8 in the bases and H1', H2' and H2'' in the ribose moieties. The cytosine to thymine replacements in G21T/C9T and G21T/C11T enable immediate identification of the corresponding correlations between H5 and H6 in the spectra of G21T. Similarly, the assignments of the H2 and H8 protons of A17

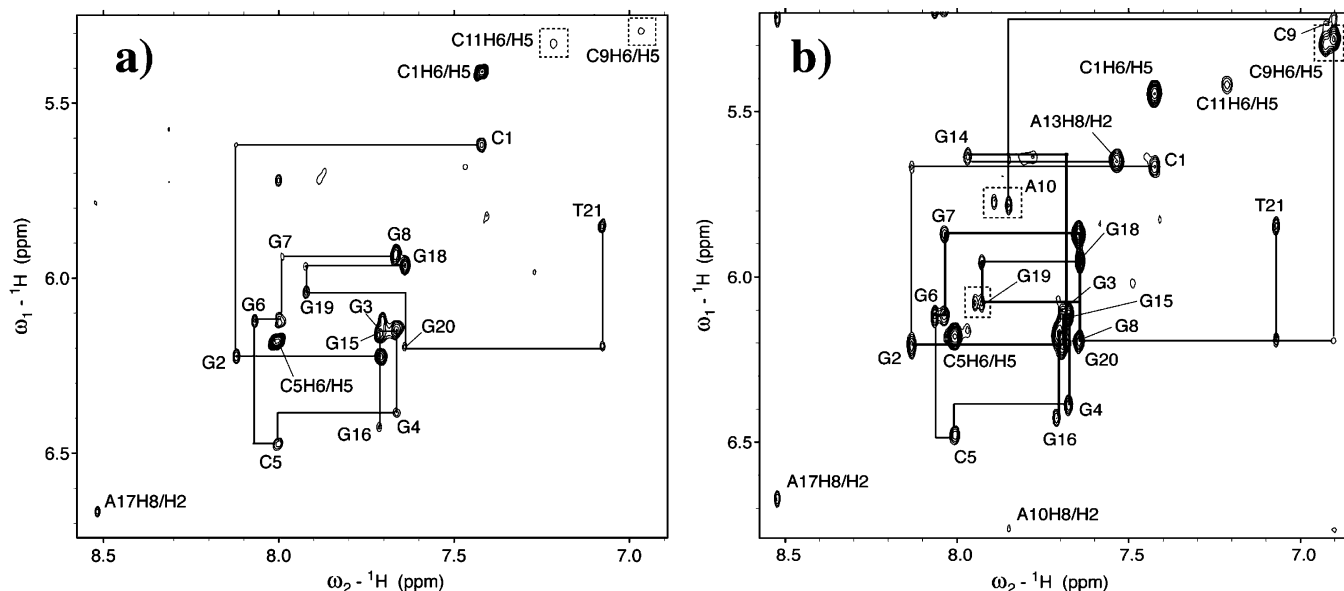
have been confirmed by the disappearance of the most downfield shifted <sup>1</sup>H resonance at 8.5 ppm, corresponding to the H8 proton of the adenine base in G21T/A17T (data not shown). Interestingly, the imino proton spectra of individual *c-kit2* variants exhibit marked differences both in chemical shift dispersions and in line widths, that is, dynamics (Figure 2). At 298 K, all loop variants exhibit resolved imino proton resonances, most of which could be assigned by comparing the 2D <sup>1</sup>H NOESY spectra with respect to that of G21T. Note that these spectra were recorded after the overnight annealing. The cross-peaks G6(H1)xG2(H8), G7(H1)xG2(H8) and G19(H1)xG14(H8), observed for G21T (Figure 1d), were absent in some of the loop variants (e.g. G21T/A17T, see below), suggesting that the second conformation was not significantly populated in these loop variants. In addition, 1D <sup>31</sup>P spectra of the loop variants show a similar peak distribution with a cluster centered at –1 ppm, while the relative intensities and the degree of chemical shift dispersions are broadly correlated with the observations seen in the 1D <sup>1</sup>H spectra, suggesting some degree of structural differences at the backbone level also (data not shown). The differences in the folding kinetics between different loop variants may cause the differences in the overall appearance of the corresponding spectra at the particular time points when the spectra were recorded.

By changing the sample temperature from 298 to 283 K and 310 K, followed by a brief equilibration period of ca. 10 min inside the spectrometer, the loop variants exhibit significant line-broadening for most *c-kit2* variants, except for G21T/A17T, suggesting conformational changes taking place on the millisecond (ms) timescale. Given that the very slow refolding kinetics of wt and G21T is on a timescale between days and months (Figure S1), the observed broadening could be a result of local rearrangements that have been slowed down at a lower temperature. At the elevated temperature of 310 K, wt *c-kit2* exhibits 12 well-resolved resonances (bearing in mind that this particular sample has been equilibrated extensively; Figure S1) with roughly equal peak intensities. A comparable improvement is observed for G21T/G10A. The spectral qualities of G21T/C9T and G21T/C11T, however, deteriorate somewhat compared to those obtained at 298 K. It was noteworthy that the 1D imino proton spectra of the loop variants, in particular G21T/A17T, recorded at 298K resemble the initial spectrum (recorded within 24 h) of G21T (Figures 2 and S1) more than that of an aged G21T (recorded after several weeks; Figure S1) based on which we assigned the individual imino proton resonances through <sup>15</sup>N enrichments (Figure 1a).

Among all the resolved imino proton resonances, G2 H1 has the most downfield chemical shifts for all *c-kit2* variants with much higher sensitivity toward temperature changes for G21T/G10A and slightly less so for G21T/C9T and G21T/C11T (Figure 2). The G2 H1 resonances of all loop variants but G21T/G10A also show clear doubling as observed in the initial spectra of wt and G21T (Figures 1 and 2). For G21T/G10A, the resonances of G8 H1 and G19 H1 show pronounced upfield shifts by ca. 0.5 ppm in comparison to those of G21T. The chemical shifts and line-shapes of these resonances are not sensitive to temperature changes, suggesting a well-defined conformation, involving S2 and S4, which is rather unexpected because the two imino protons are not involved in the same G-tetrad. Given the well-resolved imino proton spectrum of G21T/G10A, we decided to carry out further comparative spectral analyses for G21T and G21T/G10A. The common features consistently observed in all loop variants were used in



**Figure 2.** The 700 MHz imino proton spectra of *c-kit2* variants recorded at different temperatures. The resonances in the spectrum of G21T recorded at 298 K are labeled with the corresponding sequence numbers to aid visualization. Additional assignments are made for G21T/G10A to highlight the upfield shifts of G8 H1 and G19 H1. Note that the spectrum of G21T/A17T at 298K, which does not change over time, resembles the initial spectrum of G21T (Figure S1).



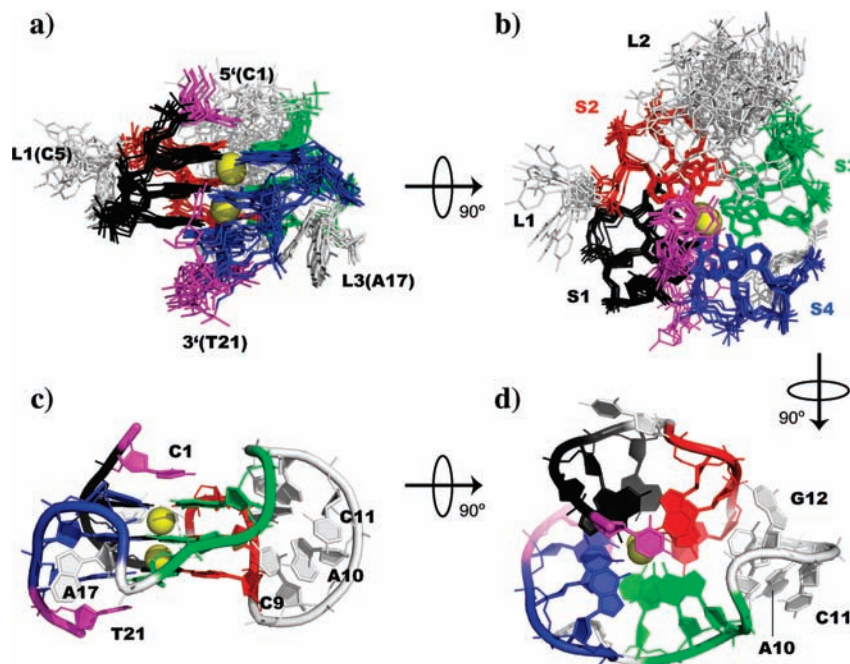
**Figure 3.** NOE assignments and spectral comparisons of *c-kit2* variants. Correlations between aromatic H6/H8 and anomeric proton H1' in the representative region of NOESY spectra of (a) G21T and (b) G21T/G10A were used to establish sequential connectivities between the spin systems. Intraresidue correlations are labeled with corresponding residue identities. For G21T, exceedingly weak intrabase H5/H6 correlations of C9 and C11 are indicated by dashed boxes to contrast with that of C5 and C1. The G10A replacement results in significant reduction in the dynamic range of observed peak intensities; at the same time, conformational polymorphism is clearly visible for a number of residues that give rise to two distinct cross-peaks (dashed boxes) with similar peak intensities. Both spectra were recorded at 700 MHz and at 298 K with the same acquisition parameters and similar sample concentrations (ca. 200  $\mu$ M).

our subsequent structural elucidation of the unimolecular all-parallel topology.

NMR-based structural characterization relies heavily on the resonance assignments of NOE correlations and the subsequent conversion of cross-peak intensities to distance restraints.<sup>46</sup> As mentioned above, the NOESY spectrum of G21T exhibits a large dynamic range, differences in the spin-relaxation properties due to dispersed internal motions, reflected in line widths, with many expected loop resonances missing, an observation that we have attributed to conformational variability and differences in

dynamics between different parts of the molecule. For example, the intrabase H5/H6 in cytosines has a fixed distance of 2.6 Å, which is expected to give rise to an intense NOEs. The intensities of the H5/H6 cross-peaks of C9 and C11, however, are about 1/10 of that of C5 and C11 in G21T and are much weaker than many of the intraresidue H8/H1' correlations (Figure 3a). The differences in the observed peak heights of the same H5/H6 cross-peaks, that is, dynamic range, are reduced in G21T/G10A, while the chemical shifts of most resonances exhibit marginal changes. Additionally, a number of residues, including C9, A10 and G19, exhibit clear peak doubling with similar peak intensities, suggesting the presence of another

(46) Wüthrich, K. *NMR of Protein and Nucleic Acids*; John Wiley & Sons, Inc.: New York, 1986.



**Figure 4.** Structural model of *c-kit2* G21T/G10A. (a and b) Two orthogonal views of an ensemble of 10 lowest total energy structures of G21T/G10A. These structures have been subject to explicit water refinement with counterions included. Residues in S1, S2, S3 and S4 are colored in black, red, green and blue, respectively. S1 and S4 are capped by C1 and T21 at the 5' and 3' ends, respectively, shown in magenta. Two coordinated potassium ions, which have been included in the explicit solvent refinement procedure, are shown in yellow spheres. (c and d) Cartoon representation of the representative structural model. The structure is orthogonal from the ones in panels (a and b) to highlight the base stacking in L2, i.e., C9–C11, and the compaction of L3, i.e., A17, as labeled.

similar but distinct conformation. Finally, the intraresidue correlations of G14 H8/H1 and A13 H8/H2 could now be identified through additional correlations with C11 in other regions of the NOESY spectrum (Figure 3b). Importantly, the relatively weak intraresidue H8/H1' peak intensities have led us to conclude that all the tetrad-forming guanines are in *anti* conformations<sup>47</sup> (G8 and G18 are somewhat ambiguous as the corresponding peak intensities are not as weak as other ones, which could again be attributed to the intrinsic dynamics; see below). Together with the H1–H1 (Figure 1) and H1–H1' (not shown) correlations, we have established the folding topology of *c-kit2* variants to be in a parallel topology with double chain reversals for all connecting loops and the nucleotides in all three G-tetrads in *anti*•*anti*•*anti*•*anti* conformations.<sup>2,3</sup>

**Solution Structures of the *c-kit2* Variants All Adopt a Parallel Propeller Topology.** We have generated models for the average structure of *c-kit2* variants, G21T and G21T/G10A, using the experimentally derived distance restraints. The molecules have all been assigned a parallel topology with three G-tetrads (Figure 4) and three chain-reversal (propeller) loops. L1 and L3 are single-nucleotide (C5 and A17, respectively), and L2 is a pentanucleotide loop (CGCGA in wt *c-kit2*). Considerable variability in the structure of the loops was seen, particularly the pentanucleotide loop, in part as a consequence of the lack of experimental distance restraints. The terminal bases, 5' C1 and 3' T21, stack on either side of the G-quadruplex core to cap the terminal tetrad planes: C1 against G2/G6 and T21 against G20. Although both L1 and L3 are single-nucleotide loops, the structure of C5 is much less defined, and the cytosine base adopts a wide range of orientations, consistent with the very intense H5/H6 cross-peak that indicates significant internal

mobility (Figure 3). A17, on the other hand, has a much more defined conformation and it is pointing toward the backbone of S4 (Figure 4a,c). The long L2 is very flexible as also evidenced by the presence of very few associated NOEs. However, there is a propensity to form an internal stacking arrangements within the loop. In G21T/G10A, the mutated A10 is often found to be stacked against C9 and C11 to form a more complex loop structure. Direct stacking between C9 and C11 has also been observed. The other two loop residues, G12 and A13, are very flexible, but G12 is able to stack on the top G-tetrad and hydrogen bond to C1. Altogether, the loop behaves as a short and highly flexible single strand of DNA with a tether at each end.

**NMR Hydrogen–Deuterium Exchange Experiment Reveals Hydrogen Bonding Asymmetry in the G-Tetrads.** To investigate the dynamical differences between the loop variants of *c-kit2*, we have carried out NMR hydrogen–deuterium exchange (HDX) experiments to extract the kinetics of solvent-mediated exchanges of the imino protons of *c-kit2* variants under equilibrium conditions. The rates of HDX ( $k_{\text{HDX}}$ ) reflect the degree of local fluctuations, that is, dynamics, of individual hydrogen bonds within the G-tetrads. Such information has been widely used to extract the free energy of unfolding at a residue-specific level for proteins.<sup>48,49</sup> For nucleic acids, the lifetimes of hydrogen bonded imino protons in Watson–Crick base-paired duplexes are on the timescale of 1–50 ms,<sup>50–52</sup> imino protons

(47) Smith, F. W.; Feigon, J. *Nature* **1992**, *356*, 164–8.

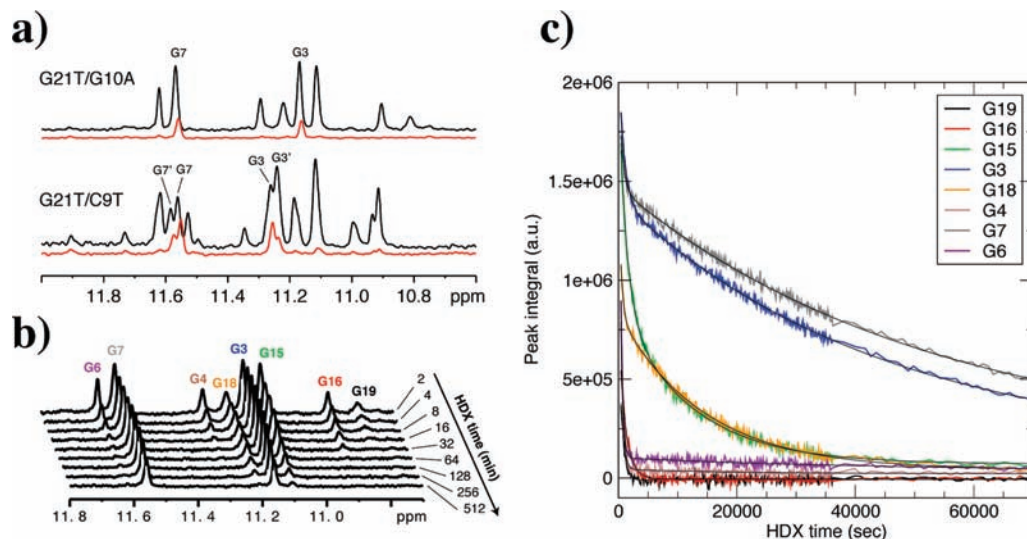
(48) Krishna, M. M. G.; Hoang, L.; Lin, Y.; Englander, S. W. *Methods* **2004**, *34*, 51–64.

(49) Englander, S. W.; Mayne, L.; Krishna, M. M. *Q. Rev. Biophys.* **2007**, *40*, 287–326.

(50) Varnai, P.; Canalia, M.; Leroy, J. L. *J. Am. Chem. Soc.* **2004**, *126*, 14659–67.

(51) Lee, J. H.; Pardi, A. *Nucleic Acids Res.* **2007**, *35*, 2965–74.

(52) Every, A. E.; Russu, I. M. *Biopolymers* **2007**, *87*, 165–73.



**Figure 5.** NMR HDX kinetics monitored by real-time 1D  $^1\text{H}$  NMR spectroscopy. (a) The 900 MHz 1D imino proton spectra of *c-kit2* variants recorded at time 0 and after 20 h of HDX at 298 K. The experimental dead time is about 4 min during which time the resonances of a number of imino protons in the top G-tetrad have been lost, due to rapid exchange with bulk solvent. (b) Stack plot of a series of 1D imino proton spectra of G21T/G10A for following the kinetics of HDX in real-time. Each spectrum along the time series is taken from a geometric sequence, ranging from 2 to 512 min. (c) Rate analysis of the *c-kit2* G21T HDX kinetics. The NMR signals of each imino proton (colored as in panel (b)) are plotted as a function of time and fitted to a double-exponential function with a baseline offset. The fitted functions are shown in thin black lines.

in higher order structures such as tRNA<sup>53</sup> and triplexes<sup>54</sup> exchange much more slowly, which are on the timescale of hours to days. For G-quadruplexes, the degree of protection for imino protons within the G-tetrads is much higher and the corresponding exchange rates are much slower. Phan et al. have reported<sup>19</sup> the rates of HDX for the four most protected imino protons of *myc-2345* and *myc-1245* to be in the range of 15–40 h, that is, an exchange rate  $k_{\text{HDX}}$  of ca.  $1 \times 10^{-5} \text{ s}^{-1}$ . Few studies on G-quadruplexes so far, however, have reported on the imino proton exchange kinetics; HDX can be employed to identify the most protected imino protons, usually the four in the central G-tetrads with similar degree of protection, by recording the corresponding imino proton spectra at a single time point after incubating the samples in  $\text{D}_2\text{O}$  for typically 4–24 h.<sup>6,47,55–59</sup>

Given the abundant internal dynamics of *c-kit2* observed in this work, we expected to see a broader distribution of HDX profiles of the 12 imino protons of individual *c-kit2* variants. Immediately after dissolving the lyophilized *c-kit2* variants into  $\text{D}_2\text{O}$ , between six and 10 imino proton resonances were identified in the first 1D  $^1\text{H}$  NMR spectrum with G2 H1 being the least protected, which is consistent with its higher temperature sensitivity as described above. Additionally, G3 H1 and G7 H1 are the most protected of all, with a significant level of residual signals over 20 h of HDX for all loop variants; those of G21T/A17T exchange at slightly faster rates, resulting in a complete exchange within 20 h. The use of a 900 MHz NMR spectrometer allows highly resolved imino proton resonances to be obtained, which is critical for signal integration as a

function of HDX time for subsequent data fitting. This has been particularly successful for G21T/G10A with all resonances exhibiting baseline separation (Figure 5). Additionally, the peak doublings of G3 H1 and G7 H1 remain visible in the spectrum of G21T/C9T after 20 h of HDX (Figure 5a). Note that the relative peak heights between the doublets are different at time 0 and after 20 h. We have assigned the identities of the fast exchanging ones as G3' and G7', and the slow exchanging ones as G3 and G7 hereafter (Figure 5a).

HDX under equilibrium conditions is expected to give rise to single-exponential decaying profiles for individual exchangeable groups. In the case of *c-kit2*, the HDX profiles for the slow-exchanging imino protons, particularly those of G3, G7, G15 and G19 in the central G-tetrad, are clearly double-exponential with a fast HDX rate, and a slower one that differs by at least 10-fold; the initial amplitudes of the two were comparable in magnitude (Table 2 and Experimental Section). The results suggest the presence of two distinct dynamic states that have very similar structures in NMR terms, that is, similar chemical shifts. In the case of G21/C9T, two distinct resonances have been identified and assigned to the imino protons of G3 and G3' (minor conformation of G3). As the two resonances were significantly overlapped, they were thus integrated together and subject to rate analysis, giving rise to two exchange rates,  $35 \times 10^{-6}$  and  $1035 \times 10^{-6} \text{ s}^{-1}$ , with similar initial amplitudes (Table 2). Assuming that the two populations independently exchange with bulk solvent, the resonance corresponding to the fast exchanging population will decay beyond detection after 20 h, which is in contrast to the experimental observation (Figure 5a). It is therefore plausible that an additional interconversion process exists between the slow and fast exchanging populations that leads to a significant fraction of the initial population, which is in the slow exchanging population, to switch to the fast exchanging conformation and vice versa. Hence, the apparent intensities of the two populations are comparable, as a result of the interconversion during the course of the HDX process.

Comparison of the observed  $k_{\text{HDX}}$  values of all *c-kit2* variants shows that, in general, G3 H1 and G7 H1 in the first two strands

- (53) Figueroa, N.; Keith, G.; Leroy, J. L.; Plateau, P.; Roy, S.; Gueron, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 4330–3.  
 (54) Coman, D.; Russu, I. M. *Nucleic Acids Res.* **2004**, *32*, 878–83.  
 (55) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2007**, *35*, 6517–25.  
 (56) Matsugami, A.; Xu, Y.; Noguchi, Y.; Sugiyama, H.; Katahira, M. *FEBS J.* **2007**, *274*, 3545–56.  
 (57) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2003**, *125*, 15021–7.  
 (58) Zhang, N.; Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2005**, *127*, 17277–85.  
 (59) Phan, A. T.; Kuryavyi, V.; Ma, J. B.; Faure, A.; Andreola, M. L.; Patel, D. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 634–9.

**Table 2.** Analysis of NMR HDX Experiments of *c-kit2* Variants

		G21T		G21T/C9T		G21T/G10A		G21T/C11T		G21T/A17T	
		$k_{\text{HDX}}^s$ ( $10^{-6} \text{ s}^{-1}$ )	$k_{\text{HDX}}^f$ ( $10^{-6} \text{ s}^{-1}$ )	$k_{\text{HDX}}^s$ ( $10^{-6} \text{ s}^{-1}$ )	$k_{\text{HDX}}^f$ ( $10^{-6} \text{ s}^{-1}$ )	$k_{\text{HDX}}^s$ ( $10^{-6} \text{ s}^{-1}$ )	$k_{\text{HDX}}^f$ ( $10^{-6} \text{ s}^{-1}$ )	$k_{\text{HDX}}^s$ ( $10^{-6} \text{ s}^{-1}$ )	$k_{\text{HDX}}^f$ ( $10^{-6} \text{ s}^{-1}$ )	$k_{\text{HDX}}^s$ ( $10^{-6} \text{ s}^{-1}$ )	$k_{\text{HDX}}^f$ ( $10^{-6} \text{ s}^{-1}$ )
S1	G2										
	G3	22 <sup>a</sup> (76)	1504 <sup>c</sup> (24)	35 (57)	1062 (43)	23 (64)	1123 (36)	35 <sup>c</sup> (56)	1365 (44)	116 <sup>d</sup> (35)	1410 (65)
	G4			211 (46)	582 (54)	51 (9)	2655 (91)			130 (46)	962 (54)
S2	G6			2 (3)	2072 (97)	22 (2)	3017 (98)	40 (5)	2717 (95)	55 (29)	1811 (71)
	G7	470 (66)	2309 (34)	19 (46)	552 (54)	19 (70)	1664 (30)	20 (59)	1909 (41)	130 (45)	1908 (55)
	G8									3311 (100)	
S3	G14			10630 (100)						1834 (100)	
	G15	54 <sup>b</sup> (57)	881 (43)	86 (41)	750 (59)	83 (42)	767 (58)			58 (25)	1684 (75)
	G16	84 (63)	2353 (37)	1072 (100)		94 (100)		1784 (100)			
S4	G18					69 (53)	2224 (47)	1792 (100)		2095 (100)	
	G19	56 (77)	2762 (23)	130 (74)	409 (26)	177 (100)			o.l. <sup>c</sup>	o.l. <sup>d</sup>	
	G20	o.l.		2030 (100)				1520 (100)		1360 (100)	

<sup>a</sup> The relative amplitudes between the slow and fast exchanging populations, expressed as the percentages of the sum of the amplitudes corresponding to the fast and slow phases ( $A_f/(A_s+A_f)$ ,  $i = s$  or  $f$ ), are shown in parentheses. <sup>b</sup> The resonances of G15 H1 and G20 H1 are heavily overlapped (o.l.). <sup>c</sup> The resonances of G3 H1, G15 H1 and G19 H1 are partially overlapped. <sup>d</sup> The resonances of G19 H1 could not be unambiguously identified in the 1D spectrum, which is potentially overlapped with that of G3 H1. The fitting of a second slow phase is evoked only when the single-exponential fitting showed clear residual deviations from the experimental data to avoid overfitting. <sup>e</sup> The fitting of a second slow phase is evoked only when the single-exponential fitting showed clear residual deviations from the experimental data to avoid overfitting.

are the most protected imino protons and that, for G21T/C9T and G21T/G10A, G15 H1 and G19 H1 are exchanging ca. 3–5 times faster than the slowest exchanging imino protons of the same G-tetrad, G3•G7•G15•G19, indicating dynamical asymmetry within the core structure, which is in contrast with most of the previous NMR findings that the four imino protons in the central G-tetrads are equally protected during HDX.<sup>6,47,55–59</sup> The slowest exchanging imino group, G7 H1 for most *c-kit2* variants, have a half time,  $t_{1/2}$ , of ca. 10 h, which is at the lower bound of what has been reported for the *c-myc* G-quadruplex, which is in the range of 10–50 h,<sup>19</sup> suggesting that the central G-tetrad of *c-kit2* fluctuates more than that of *c-myc*. Additionally, G4 H1 and G6 H1 that flank the single nucleotide L1 also exhibit a higher degree of protection in comparison with those that flank L3, namely, G16 H1 and G18 H1.

## Discussion

There is good evidence from a large number of studies to suggest that a given G-rich nucleic acid sequence can adopt a variety of G-quadruplex folding topologies often accompanied by the ability to interconvert between these states over a broad range of time scales. Depending on the conditions used for structural investigations, one of several conformations can be preferentially populated.<sup>60,61</sup> We and others have used single molecule fluorescence spectroscopy to demonstrate that human telomeric DNA G-rich repeats can adopt multiple G-quadruplex conformations, both parallel and antiparallel, in dilute solution with temperature- and counterion-dependent folding and unfolding rates;<sup>62–64</sup> similar results have been derived by solution

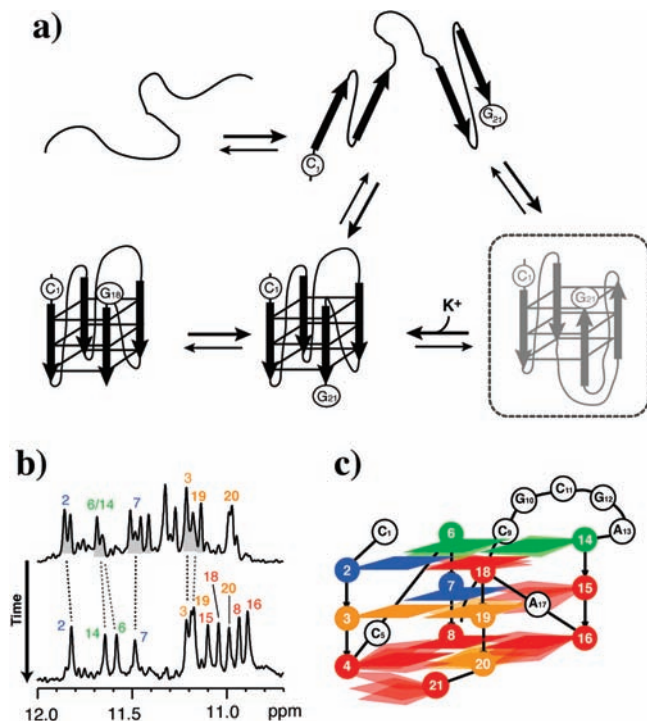
NMR spectroscopy.<sup>57,65,66</sup> In the case of *c-kit2*, we have also shown that the wt sequence exhibits two distinct sets of FRET, NMR and CD signals in the absence (or presence of low concentrations) of potassium ions, which we tentatively attributed to the coexistence of parallel and antiparallel conformations.<sup>43,67</sup> We have confirmed in the present study that *c-kit2* forms a parallel quadruplex in  $\text{K}^+$  solution with internal dynamics that are modulated by loop composition.<sup>16,68,69</sup>

Among a handful of intramolecular G-quadruplex structures that have been reported so far, the majority has been those of human telomeric DNA repeats with relatively long loops and hence much more polymorphic.<sup>4–6</sup> Meanwhile, intramolecular promoter G-quadruplex often have shorter loops.<sup>14,17,39,40,70</sup> On the basis of the sequence of *c-kit2* G21T, one could expect parallel arrangements for the first two (S1 and S2) and last two strands (S3 and S4), both of which are connected by a single nucleotide (G<sub>3</sub>NG<sub>3</sub>), which strongly favors double chain reversal.<sup>16,68,69</sup> G-quadruplexes in human *c-myc* promoter, *myc-2345*<sup>19</sup> and MYC22-G14T/G23T,<sup>20</sup> have been found to adopt an all-parallel topology with short loops (5′-G<sub>3</sub>NG<sub>3</sub>N<sub>2</sub>G<sub>3</sub>NG<sub>3</sub>-3′). Although L3 in wt *c-kit2* could in principle be two-nucleotides long when G19–G21 are involved in the G-tetrad formation, a single nucleotide L3 (A17 alone) is kinetically more favorable as has been demonstrated by a systematic survey of model oligonucleotides.<sup>71</sup> The relatively long and unstructured L2 loop provides the additional degree of freedom necessary for *c-kit2* in principle to adopt either an all-parallel, propeller-type conformation, or a mixed parallel/antiparallel conformation

- (60) Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. *Nucleic Acids Res.* **2008**, *36*, 5482–515.  
 (61) Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. *Nucleic Acids Res.* **2005**, *33*, 4649–59.  
 (62) Green, J. J.; Ying, L.; Klenerman, D.; Balasubramanian, S. *J. Am. Chem. Soc.* **2003**, *125*, 3763–7.  
 (63) Ying, L.; Green, J. J.; Li, H.; Klenerman, D.; Balasubramanian, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14629–34.  
 (64) Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18938–43.

- (65) Phan, A. T.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2006**, *34*, 5715–9.  
 (66) Kaushik, M.; Bansal, A.; Saxena, S.; Kukreti, S. *Biochemistry* **2007**, *46*, 7119–31.  
 (67) Dash, J.; Shirude, P. S.; Hsu, S. T.; Balasubramanian, S. *J. Am. Chem. Soc.* **2008**, *130*, 15950–6.  
 (68) Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.* **2004**, *126*, 16405–15.  
 (69) Rachwal, P. A.; Brown, T.; Fox, K. R. *FEBS Lett.* **2007**, *581*, 1657–60.  
 (70) Dai, J.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D. *J. Am. Chem. Soc.* **2006**, *128*, 1096–8.  
 (71) Mergny, J. L.; De Cian, A.; Amrane, S.; Webba da Silva, M. *Nucleic Acids Res.* **2006**, *34*, 2386–97.





**Figure 6.** Schematic representation of transitions between possible conformational arrangements of *c-kit2*. (a) The all parallel topology with two single nucleotide loops, L1 and L3, is the energetically most favorable conformation in the presence of potassium ions. Replacement of the last guanine in S4, G21T, reduces the complexity of the conformational space and therefore accelerates the process of conformational search for a low energy state in the associated free energy landscape. The gray dashed box highlights the presence of potassium ion-dependent conformation that is likely to be antiparallel (see Figure S3) whose detailed structure has hitherto not been characterized. (b) 1D imino proton spectra of wt *c-kit2* recorded immediately after annealing (top panel) and after several months (bottom panel). Tentative assignments for the top spectrum are made based on chemical shift similarities with the ones that have been unambiguously assigned in the bottom spectrum with peaks that exhibit doublets (or more for G7 H1) shaded in gray. The resonances that have correspondence between the two spectra are connected by dashed line. The assignments of G2 and G7 are colored in blue for their respective chemical shift differences between the two spectra which are small; those of G6 and G16 are colored in green for they exhibit larger but well-resolved differences; those of G3, G19 and G20 are colored in orange for they have corresponding resonances in the initial spectrum (top panel) with very high ambiguities due to a large number of unresolved resonances for G4, G8, G15, G16, G18 and G21 (colored in red). (c) Schematic of the initial parallel topology attained by wt *c-kit2* with varied degrees of structural heterogeneities. The residues that have reached lower energy conformations, with different degrees of peak doublings, including G2 and G7 are shown in blue as in panel (b); ambiguous assignments were made for G3, G6, G14, G19 and G20 in panel (a) and they are shown in green and orange as in panel (b); the remaining residues, including G21 which may participate in G-tetrad formation in the early stage of the folding process in slightly different conformations, are shown in red with multiple parallelograms to indicate the structural heterogeneities.

with a long lateral loop (Figure 6a). In the case of *myc-1245*,<sup>19</sup> which has a T<sub>4</sub> replacement for the third G-tract of the five consecutive G-tracts in the wt *c-myc* sequence, it was found to be in an all-parallel topology with a hexanucleotide loop connecting S2 and S3 (5'-G<sub>3</sub>NG<sub>3</sub>N<sub>6</sub>G<sub>3</sub>NG<sub>3</sub>-3'). In contrast, the G-quadruplex formed in human *bcl-2* promoter<sup>70</sup> was found to be in a mixed parallel/antiparallel topology with a largely disordered heptanucleotide loop connecting S2 and S3 (5'-G<sub>3</sub>NG<sub>3</sub>N<sub>7</sub>G<sub>3</sub>NG<sub>3</sub>-3').

The diversity of topologies that can be adopted by G-quadruplexes can be best appreciated by comparing the structure

of *c-kit2* and that of *c-kit1*, which are found proximally within the same promoter. Both have all strands parallel, but the former has three conventional strand-reversal loops, whereas the latter has two strand-reversal loops together with a dinucleotide loop connecting two adjacent corners and a pentanucleotide loop that folds back to allow a terminal guanine to participate in the tetrad core.<sup>36</sup> The pentanucleotide stem loop of *c-kit1* (of sequence AGGAG) is highly structured compared to the loop in *c-kit2*, since it is stabilized by stacked A·G base pairs. The relatively unconstrained all-parallel topology may be the origin of the dynamical asymmetry that has been observed for all variants of *c-kit2* delineated from the HDX data.

Having established the assignments of the imino proton resonances of all the loop variants of *c-kit2*, we have been able to assign those of the wt with 12 well-resolved resonances, denoted as the low energy state (lower panel in Figure 6b). Comparing the similarities of these resonances with respect to the initial spectrum of wt, a number of peaks in the initial spectrum could be tentatively assigned. G2, G3 and G7 have very similar chemical shifts in the initial spectrum to those in the low energy state. Additionally, G6 and/or G14 also exhibit the low energy state-like chemical shifts in the initial spectrum. Collectively, these residues are located in one corner of the three-layer G-tetrad core structure (Figure 6c). This could be rationalized by S4, particularly G20 and G21, tending to switch their G-tetrad registering in a latch sliding manner,<sup>72</sup> leading to fluctuations in the bottom G-tetrad as well as in the local arrangements of S3-L3-S4. In the low energy state, G8, G15, G16, G18, and G20 in the other half of the core structure exhibit well-resolved upfield shifted lines ( $\delta^1\text{H}$  10.7–11.2 ppm), while a collapsed broad line is observed in the initial spectrum as a result of conformational heterogeneity. Although it is assumed that a single nucleotide double chain reversal is kinetically the most favorable conformation for S3-L3-S4 within each topology,<sup>16,68,69</sup> L3 can also be a double nucleotide one in the wt sequence (consists of A17 and G18). The exceedingly slow folding kinetics observed for wt *c-kit2* may be the result of multiple parallel folding pathways in a rugged energy surface, rather than sequential folding events of obligatory intermediate formation with very high kinetic barriers.

Our NMR HDX experiments have provided intriguing evidence of the existence of dynamical asymmetry within the G-tetrads: the first half of the molecule, S1-L1-S2, is slightly better defined (less local fluctuations) than is the second half, S3-L3-S4. It is also apparent that there are at least two populations present in solution that are structurally very similar, yet distinguishable by NMR spectroscopy. The fact that two sets of NMR resonances, separated by very small chemical shift differences, have been identified for a number of G-tetrad forming residues (Figures 1, 2 and 5) suggests that the two conformations are in the same parallel conformation and that the exchange between the two states, if indeed present, is slow on the NMR timescale ( $> \text{s}^{-1}$ ). Indeed, previous NMR studies on *myc-2345* and *myc-1245* have shown that the imino proton chemical shifts (local structures) are not sensitive to a long loop extension in *myc-1245*, but the rates of HDX for the imino protons (dynamics) within the central G-tetrad of *myc-1245* are ca. 2 times faster compared to those of *myc-1245*. There was, however, no evidence of internal dynamics asymmetry nor conformational interconversion in this study. Here, the HDX

(72) Spackova, N. a.; Berger, I.; Sponer, J. *J. Am. Chem. Soc.* **1999**, *121*, 5519–5534.

data on G21T/C9T (Figure 5a) shows the possible interconversion between the two states as the doublet imino proton signals of G3 and G3', as well as those of G7 and G7', decay at a similar rate indicating that the pool of the fast exchanging imino protons is replenished by the slow exchanging ones during the apparently irreversible HDX process (Figure 5b). This is reminiscent of the interconversions between short-lived and long-lived human telomeric DNA G-quadruplexes observed by single molecule FRET measurement<sup>64</sup> with corresponding lifetimes in the order of 10 and 100 s, respectively. The relatively fast exchange process at equilibrium suggests that the conformations are similar in energy and are separated by relatively small kinetic energy barriers to allow redistribution of populations during the course of HDX on the timescale of  $10^3$ – $10^6$  s.

Genomic DNA sequences exist predominantly in a duplex form, but must necessarily be unwound into a single-stranded form to enable transcription. Thus, *in vivo*, duplex versus quadruplex stability may be further influenced by the local environment that will include factors such as superhelicity.<sup>23,73</sup> G-quadruplex forming sequences in the 5' promoter regions of human genome and formation of higher order structures may play an important role in transcription regulation, particularly those that are associated with oncogenesis.<sup>7,9,74</sup> Akin to protein folding, G-quadruplex formation is a highly cooperative process.<sup>60</sup> The folding pathways of G-quadruplexes, however, can consist of a large number of pathways with the potential to switch from one to another depending on the sample conditions. We have shown here that, where the presence of potassium ions and the G21T replacement favor a parallel conformation, *c-kit2* exists in at least two subtly distinct conformational states that are structurally and energetically similar. These are most likely substates populated in the global minimum, which would take a very long period of time for wt *c-kit2* to reach due to the complexity caused by the additional guanine at the 3' end (Figure 1). The HDX data on its own, however, cannot provide information about the folding pathways of G-quadruplex formation without the help of further kinetics information.<sup>49,75</sup> What it has demonstrated, however, is how sensitive the dynamics and stabilities are in response to the subtle structural differences manifested in small NMR chemical shift changes, for example, G21T/G10A (Figure 5a), and most likely indistinguishable by UV, CD and FRET measurements.

It remains to be seen as to how these low energy substates can be attained from a fully extended structure with the competition of a complementary strand. It has also been shown that a number of small molecules can target specific loop structures<sup>76</sup> with the potential to stabilize quadruplex structures. The L2 loop of *c-kit2* would be an ideal structural feature to target for recognition with the aim of selecting specific conformations through ligand binding.<sup>77</sup>

## Experimental Section

**Labeled Oligodeoxyribonucleotide Preparation.** The labeled oligodeoxyribonucleotide sequences were synthesized on an Applied Biosystems DNA synthesizer using solid-phase  $\beta$ -cyanoethylphosphoramidite chemistry.

(73) Kouzine, F.; Sanford, S.; Elisha-Feil, Z.; Levens, D. *Nat. Struct. Mol. Biol.* **2008**, *15*, 146–54.

(74) Huppert, J. L. *Philos. Trans. R. Soc., A* **2007**, *365*, 2969–84.

(75) Clarke, J.; Itzhaki, L. S.; Fersht, A. R. *Trends Biochem. Sci.* **1997**, *22*, 284–7.

(76) Campbell, N. H.; Patel, M.; Tofa, A. B.; Ghosh, R.; Parkinson, G. N.; Neidle, S. *Biochemistry* **2009**, *48*, 1675–80.

(77) Parkinson, G. N.; Cuenca, F.; Neidle, S. *J. Mol. Biol.* **2008**, *381*, 1145–56.

Bz-dA-CE phosphoramidite, iBu-dG-CE phosphoramidite, and dT-CE phosphoramidite were supplied by Sigma-Aldrich Co. Ltd., U.K., and Ac-dC-CE phosphoramidite by Link Technologies Ltd., U.K. ( $U$ - $^{13}C_{10}$ , 98%;  $U$ - $^{15}N_5$ , 98%) iBu-dG-CE phosphoramidite, manufactured by Cambridge Isotope Laboratories, Inc., was obtained from CK GAS Products Ltd., U.K. Dilution of the labeled iBu-dG-CE phosphoramidite (15 mg, 17.5  $\mu$ mol) was carried out by injecting a solution of the unlabeled G phosphoramidite in acetonitrile (0.0916 M, 2.1 mL) into the supplied vial, resulting in a 1.0 M solution of 8% [ $^{13}C$ ,  $^{15}N$ ] enriched phosphoramidite. To make more efficient use of the labeled phosphoramidite, a polymeric T resin (dT OligoPrep 250 PVA support - Link Technologies Ltd., U.K., specific loading 253  $\mu$ mol/g) was used rather than standard CPG support. Reusable synthesis columns were loaded with 12 mg (3  $\mu$ mol) of support and preswelled with acetonitrile for 1 h immediately before use. The standard 1  $\mu$ mol synthesis cycle was used for synthesis after being edited by prolonging phosphoramidite coupling times, and by increasing the volumes and reaction times of the deblocking, capping, and oxidizing reagents.

The oligodeoxyribonucleotides were cleaved from the support and deprotected by passing a 1:1 mixture of 0.88 ammonia and 40% aqueous methylamine back and forth through the synthesis columns for 10 min and then heating the solutions at 60 °C for 30 min in sealed microcentrifuge tubes. The ammonia and methylamine were removed using a centrifugal evaporator and the crude oligodeoxyribonucleotides dissolved in ammonium acetate buffer (100 mM, pH 10.0). Purifications were carried out by HPLC on a BioCAD Workstation with an Oligo R3 column (PerSeptive Biosystems, Inc.). The samples were then dialyzed against double-distilled water, yields determined by measuring the absorbance of the dialyzed oligodeoxyribonucleotide solutions at 260 nm, and lyophilized with a centrifugal evaporator to produce the final purified products.

**NMR Spectroscopy for Resonance Assignments.** All NMR samples have been subject to an annealing process by heating the NMR sample solutions at 95 °C for 5 min and slowly cooled to room temperature. The unlabeled wt and loop variants were synthesized and purified by double HPLC through a commercial source (IBA GmbH, Germany). All NMR samples, including the site-specifically  $^{13}C/^{15}N$  labeled samples, contain oligo concentrations of ca. 100  $\mu$ M, buffered in 10 mM potassium phosphate (pH 7.0), and 7%  $D_2O$  (v/v) as an NMR lock signal. All NMR data have been recorded at the  $^1H$  Larmor frequencies of 500, 700, and 900 MHz with NMR spectrometers equipped with cryogenic probes (Bruker, Germany). 1D imino proton spectra of all samples were recorded using the jump-and-return (JR) scheme to minimize saturation of the labile NMR signals as a result of water suppression.<sup>78</sup> 2D  $^1H$ - $^{15}N$  SOFAST-HMQC spectra<sup>45</sup> of 12 site-specifically  $^{15}N$ -labeled NMR samples of the G21T variant were recorded to obtain unambiguous resonance assignments: each NMR sample contains 8% [ $^{13}C$ ,  $^{15}N$ ] labeled guanine base introduced to one of the 12 G-tetrad forming guanine residues as described above; each NMR sample is designated as [8%- $^{13}C/^{15}N$ , G<sub>*i*</sub>]-G21T to indicate the labeling site *i* {*i* = 2, 3, 4, 6, 7, 8, 14, 15, 16, 18, 19, 20}. Additionally,  $^{13}C$ -edited 1D  $^1H$  spectra were recorded for all 12 [8%- $^{13}C/^{15}N$ , G<sub>*i*</sub>]-G21T samples to obtain H8 resonance assignments of individual guanines within the G-tetrad. Resonance assignments of the remaining protons have been determined through a combination of 2D  $^1H$  NOESY with various mixing times ranging from 50 to 400 ms, 2D  $^1H$  TOCSY with a mixing time of 60 ms, and 2D  $^1H$  DQF-COSY spectra.<sup>78</sup>

**Molecular Modeling.** Two sequences, corresponding to G21T (dCGGGCGGGCGAGGGAGGGT) and G21T/G10A (dCGGGCGGGCACGAGGGAGGGT) mutants of *c-kit2* were generated in

(78) Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Rance, M.; Skelton, N. J. *Protein NMR Spectroscopy*; 2nd ed.; Academic Press: San Diego, CA, 2007.

ideal extended conformation using the leap module of Amber10.<sup>79</sup> In addition, the same sequences were built into a parallel quadruplex topology using the crystal structure of the human telomeric quadruplex (PDB accession ID: 1KF1)<sup>5</sup> as a template. Both starting models were subjected to a restrained simulated annealing conformational search protocol to evaluate the convergence of the final ensembles using the simulation package Amber10.<sup>79</sup>

**NOE Restraints.** All assigned NOESY cross-peaks were classified to strong (1.8–2.8 Å), medium (1.8–3.6 Å), weak (1.8–5.5 Å) and very weak (1.8–6.5 Å) interproton distance restraints according to their intensities. Cross-peaks related to the thymine methyl group were interpreted as ambiguous restraints and the  $\langle r^{-6} \rangle^{-1/6}$  averaged values of the three equivalent distances was restrained. Beyond these bounds, a quadratic penalty potential was applied with a force constant of 10 kcal mol<sup>-1</sup> Å<sup>-2</sup>. A total of 236 distance restraints were used for G21T and 300 distance restraints were used for G21T/G10A. In addition, ideal hydrogen bond restraints were imposed on the guanine bases in the *anti•anti•anti•anti* tetrad core using two distances per hydrogen bond. Donor to acceptor distance was set to 2.8–3.0 Å and proton to acceptor distance to 1.8–2.0 Å with a force constant of 50 kcal mol<sup>-1</sup> Å<sup>-2</sup>. Bases corresponding to a G-tetrad were restrained to lie in the same plane with a force constant of 50 kcal mol<sup>-1</sup> rad<sup>-2</sup>. The glycosylic  $\chi$  angle (O4'–C1'–N9–C4) was restrained to an *anti* conformation (240 ± 70°) for all guanines participating in the G-quadruplex tetrad with a force constant of 50 kcal mol<sup>-1</sup> Å<sup>-2</sup>.

For restrained simulated annealing calculation, initial models were energy minimized without restraints and subjected to 200 simulated annealing cycles using the full set of restraints. A cycle started from the structure obtained in the previous cycle and included 10 ps heating from 300 to 1000 K followed by an equilibration of 10 ps at 1000 K without any restraints applied. The force constants for all restraints were then scaled gradually from zero to the final values during 20 ps at 1000 K followed by cooling the system to 300 K over 30 ps. Interactions within the system were calculated using the latest Amber force field for nucleic acids (ff99bsc0)<sup>80,81</sup> and Generalized Born solvation (*igb* = 2) with monovalent salt concentration corresponding to 0.1 M. The final structures were energy minimized using the same parameters and restraints described above. Ten best models for an average structure were selected based on the value of the overall potential energy and NMR restraint violations to undergo further refinement in explicit water and counterion environment.

For structure refinement with molecular dynamics in explicit solvent, each final model structure was placed in a truncated octahedral box of about 5000 water molecules and 20 K<sup>+</sup> counterions. Two K<sup>+</sup> ions were inserted in the core of the quadruplex with ideal coordination geometry between stacked guanine tetrads as seen in the crystal structure (PDB accession ID: 1KF1). Potassium ions were maintained at this position throughout using distance restraints (2.7–3.0 Å) to O6 of the corresponding guanines

with force constant of 10 kcal mol<sup>-1</sup> Å<sup>-2</sup>. The system was energy minimized and then heated to 300 K at constant volume during 50 ps while the quadruplex was kept under positional restraints with force constant of 25 kcal mol<sup>-1</sup> Å<sup>-2</sup>. The positional restraints were gradually removed over 300 ps at constant pressure (1 atm) and temperature (300 K) and a production phase was initiated for 2 ns with the full set of restraints applied, but the planarity and hydrogen bond restraints in the G-tetrad reduced to 5 kcal mol<sup>-1</sup> Å<sup>-2</sup>. The final structures were energy minimized and used for subsequent analysis. The coordinates of the ensembles of G21T and G21T/G10A have been deposited in the Protein Data Bank with the accession IDs of 2KJ2 and 2KJ0, respectively.

#### NMR Hydrogen–Deuterium Exchange (HDX) Experiments.

A total of 500 μL of each NMR sample in aqueous solution was flash frozen in liquid nitrogen and subsequently lyophilized overnight prior to the HDX experiments. A total of 500 μL of 99% D<sub>2</sub>O was added to dissolve the lyophilized sample immediately before each NMR measurement using a 900 MHz NMR spectrometer. A series of 1D <sup>1</sup>H NMR spectra were recorded until most imino proton signals have significantly attenuated after typically 14–20 h. HDX under NMR conditions follows a pseudo-first-order kinetics. The observed imino proton resonances were integrated and analyzed as a function of HDX time and fitted to the formula below,

$$I(t) = A_0 + A_s e^{(-k_{\text{HDX}}^s t)} + A_f e^{(-k_{\text{HDX}}^f t)}$$

where  $A_0$  is the baseline given by the residual H<sub>2</sub>O content,  $A_s$  and  $A_f$  are the amplitudes of two distinguishable decaying phases to the integrated NMR signals, respectively;  $k_{\text{HDX}}^s$  and  $k_{\text{HDX}}^f$  are the rate constants corresponding to the major and minor (slow and fast) contributions, respectively. Note that a single exponential decaying function with baseline is used when the two rates are similar,  $k_{\text{HDX}}^s \sim k_{\text{HDX}}^f$ , or when one amplitude is insignificant to the overall amplitude,  $A_s \gg A_f$  or  $A_s \ll A_f$ , to avoid overfitting. The data fittings were carried out using the built-in module within the program *Grace* (<http://plasma-gate.weizmann.ac.il/Grace/>).

**Acknowledgment.** We thank the BBSRC for project funding, Cancer Research UK for programme funding (to S.B. and S.N.). S.-T.D.H. is a recipient of a Human Frontier Long-term Fellowship (LT0798/2005) and is supported in part by the Postdoctoral Research Abroad Program of National Science Council of the Republic of China, Taiwan (NSC97-2917-1-564-102). We thank the staff and the use of the Biomolecular NMR Facility, Department of Chemistry, University of Cambridge. The 900 MHz NMR spectra were recorded at the SON NMR Large Scale Facility in Utrecht, The Netherlands (funded by the European Union, contract number RII3-026145).

**Supporting Information Available:** Additional 1D and 2D NOESY spectra of *c-kit2* variants recorded under different conditions. Imino proton chemical shift assignment of *c-kit2* loop variants. Complete authorship for ref 79. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA904007P

(79) Case, D. A., et al. *AMBER*; 10 ed.; University of California: San Francisco, CA, 2008.

(80) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–74.

(81) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–29.