

Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human *c-kit2* promoter

Vitaly Kuryavyi¹, Anh Tuân Phan² and Dinshaw J. Patel^{1,*}

¹Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA and

²School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore

Received October 7, 2009; Revised May 28, 2010; Accepted June 1, 2010

ABSTRACT

Previous studies have demonstrated that nuclease hypersensitivity regions of several proto-oncogenic DNA promoters, situated upstream of transcription start sites, contain guanine-rich tracts that form intramolecular G-quadruplexes stabilized by stacked G•G•G•G tetrads in monovalent cation solution. The human *c-kit* oncogenic promoter, an important target in the treatment of gastrointestinal tumors, contains two such stretches of guanine-rich tracts, designated *c-kit1* and *c-kit2*. Our previous nuclear magnetic resonance (NMR)-based studies reported on the novel G-quadruplex scaffold of the *c-kit1* promoter in K⁺-containing solution, where we showed for the first time that even an isolated guanine was involved in G-tetrad formation. These NMR-based studies are now extended to the *c-kit2* promoter, which adopts two distinct all-parallel-stranded conformations in slow exchange, one of which forms a monomeric G-quadruplex (form-I) in 20 mM K⁺-containing solution and the other a novel dimeric G-quadruplex (form-II) in 100 mM K⁺-containing solution. The *c-kit2* promoter dimeric form-II G-quadruplex adopts an unprecedented all-parallel-stranded topology where individual *c-kit2* promoter strands span a pair of three-G-tetrad-layer-containing all-parallel-stranded G-quadruplexes aligned in a 3' to 5'-end orientation, with stacking continuity between G-quadruplexes mediated by a sandwiched A•A non-canonical pair. We propose that strand exchange during recombination events within guanine-rich segments, could potentially be mediated by a synapsis intermediate involving an intergenic parallel-stranded dimeric G-quadruplex.

INTRODUCTION

Guanine-rich DNA and RNA sequences can form G-quadruplexes stabilized by stacked G•G•G•G tetrads in monovalent cation-containing solution (1–4). The length and number of individual G-tracts and the length and sequence context of linker residues define the diverse topologies adopted by G-quadruplexes. These G-quadruplexes can serve as therapeutic targets (5) and considerable effort has gone toward identifying evidence in support of an *in vivo* functional role for G-quadruplexes (6–8).

Systematic bioinformatic sequence analysis of human genomes has identified a prevalence of guanine-rich tracts capable of G-quadruplex formation (9–13). Such enrichment in putative G-quadruplex-forming motifs is especially prevalent in promoter regions spanning 1 kb upstream of transcription start sites of genes (14) and since such sequences also correlate with nuclease hypersensitivity sites, current efforts have been directed toward understanding the role of promoter-mediated G-quadruplex formation in transcriptional regulation (15), as well as gene regulation (16).

The *c-kit* oncogene is an important target in the treatment of gastrointestinal tumors (17). Since the proto-oncogenic *c-kit* promoter encodes for a tyrosine kinase receptor, much effort has been directed towards identification of inhibitors of *c-kit* kinase activity. Even initially successful drugs such as Gleevec (imatinib) (18) have found limited efficacy, since new patterns of resistance mutations within the binding site have impacted on clinical effectiveness (19).

Two guanine-rich sequences containing four guanine tracts, designated *c-kit1* and *c-kit2*, have been identified within the nuclease hypersensitivity region of the promoter segment of the human *c-kit* gene, upstream of the transcription start site. Both *c-kit1* and *c-kit2* guanine-rich sequences exhibit nuclear magnetic resonance (NMR) imino proton chemical shifts characteristic

*To whom correspondence should be addressed. Tel: +1 212 639 7207; Fax: +1 212 717 3066; Email: pateld@mskcc.org

of G-quadruplex formation in K⁺-containing solution (20,21). 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 (22,23).

The four guanine-rich repeat *c-kit1* 22-mer sequence (20) is shown below:

A1-G2-G3-G4-A5-G6-G7-G8-C9-G10-C11-T12-G13-G14-G15-A16-G17-G18-A19-G20-G21-G22

Our group investigated the NMR-based solution structure of this *c-kit1* 22-mer sequence in K⁺-containing solution and determined that it formed an unanticipated all-parallel stranded G-quadruplex scaffold stabilized by three stacked G-tetrads (24). The guanine residues involved in G-tetrad formation are shown in bold in the above sequence and for the first time we demonstrated that an isolated guanine (G10) can be involved in G-tetrad core formation, despite the presence of four G-G-G tracts in the sequence. The G-quadruplex scaffold (Figure S1a and b, Supplementary Data) is composed of two single-residue double-chain-reversal loops, a two-residue loop, and a five-residue stem-loop, which contains base-pairing alignments. This novel G-quadruplex scaffold could serve as a specific platform for ligand-based drug design targeted to the *c-kit1* promoter (25).

Our NMR-based structural studies have now been extended to the four guanine-rich repeat *c-kit2* 21-mer promoter sequence (21) shown below:

C1-G2-G3-G4-C5-G6-G7-G8-C9-G10-C11-G12-A13-G14-G15-G16-A17-G18-G19-G20-G21

with guanines involved in G-tetrad formation (based on structural analysis outlined below) highlighted in bold. We show that minimal variants of the *c-kit2* sequence adopt two distinct all-parallel-stranded G-quadruplex conformations in slow exchange, one of which forms a monomeric (form-I) and the other a dimeric (form-II) G-quadruplex that forms a new fold.

MATERIALS AND METHODS

Sample preparation

The *c-kit2* promoter sequences for NMR studies were prepared using two protocols, one of which involved gel-filtration through Sephadex G-25 in a centrifuge (spin-down column) and the other involved equilibrium dialysis.

Spin-down columns were prepared by packing of G-25 Sephadex, previously swollen in water, in a 3-ml syringe, whose bottom was covered with glass wool. The suspended gel was packed by centrifugation of the syringe inserted into the centrifuge tube in a bucket rotor for 4 min at 1600 *g*. After packing, the column was washed five times with 200 μ l of the buffer. Subsequently, 200 μ l of the solution containing oligonucleotide was applied on the column top and recovered after centrifugation in the

ependorf vial inserted under the syringe in the centrifuge tube. (26,27).

Equilibrium dialysis was typically performed in 2 ml of the oligonucleotide solution placed inside the Spectrapor membrane tube (Mol. Wt. cutoff 1000 Da). The dialysis against 3.5 l of buffer involved the following sequence of changed buffers: water (1 h); 100 mM (or 20 mM) KCl (1 h); water (1 h); 100 mM (or 20 mM) KCl (1 h); water (1 h); 100 mM (or 20 mM) KCl + 5 mM potassium phosphate (overnight). Finally, the concentration of the dialysis buffer was lowered so as to account for the subsequent concentration of the sample for NMR experiments and the sample was dialyzed against this diluted buffer for 1 hr. After final dialysis, the sample was lyophilized.

Samples shown on Figures 1b and 5a–c were prepared by method of gel-filtration through Sephadex G-25 packed in spin-down columns. The remaining samples were prepared by equilibrium dialysis (24,28,29). The strand concentration of the samples varied from 0.2 to 4.0 mM (measured using millimolar extinction coefficients 202 and 204 for forms I and II, respectively) and the solution was either 20 mM KCl, 5 mM K-phosphate buffer, pH 6.8 (form-I) or 100 mM KCl, 5 mM K-phosphate buffer, pH 6.8 (form-II).

Annealing experiments were performed in the NMR spectrometer. The sample in the NMR tube was heated to 85°C and then gradually cooled down in 5°C decrements. NMR spectra were recorded at each temperature after 5 min of equilibration time.

Synthesis of oligonucleotides containing H8-deuterated guanines

Oligonucleotides containing H8-deuteration of specific guanine residues were synthesized on ABI-392 synthesizer using DMF-protected 8D-dG phosphoramidite (Glen Research). Protection groups were removed by incubation of the oligonucleotide in concentrated NH₄OH solution at room temperature for 17 h.

Gel electrophoresis

Electrophoresis experiment was performed with 10 \times 7 cm native gel containing 20% (Figure 1e) or 25% (Supplementary Figure S16c) acrylamide (Acrylamide: Bis-acrylamide = 37.5:1) in TBE buffer, pH 8.3 supplemented with 10 mM (Figure 1e) or 25 mM (Figure S16c) KCl. Each sample contains 5 μ g DNA at the concentration of 0.3–0.8 mM. Gel was viewed by ultraviolet (UV) shadowing (Figure 1e) or after staining with 0.1% toluidine blue (Figure S16c).

Circular dichroism spectra

Circular dichroism (CD) spectra were recorded on a Jasco-815 CD spectrophotometer using 1 cm quartz cuvette in a reaction volume of 600 μ l at 20°C. Scans between 220 and 320 nm were recorded at a rate of 200 nm/min, 1 nm pitch and 1 nm bandwidth. The DNA concentration was 5 μ M. The sample contained 70 mM KCl, 20 mM K-phosphate, pH 7.

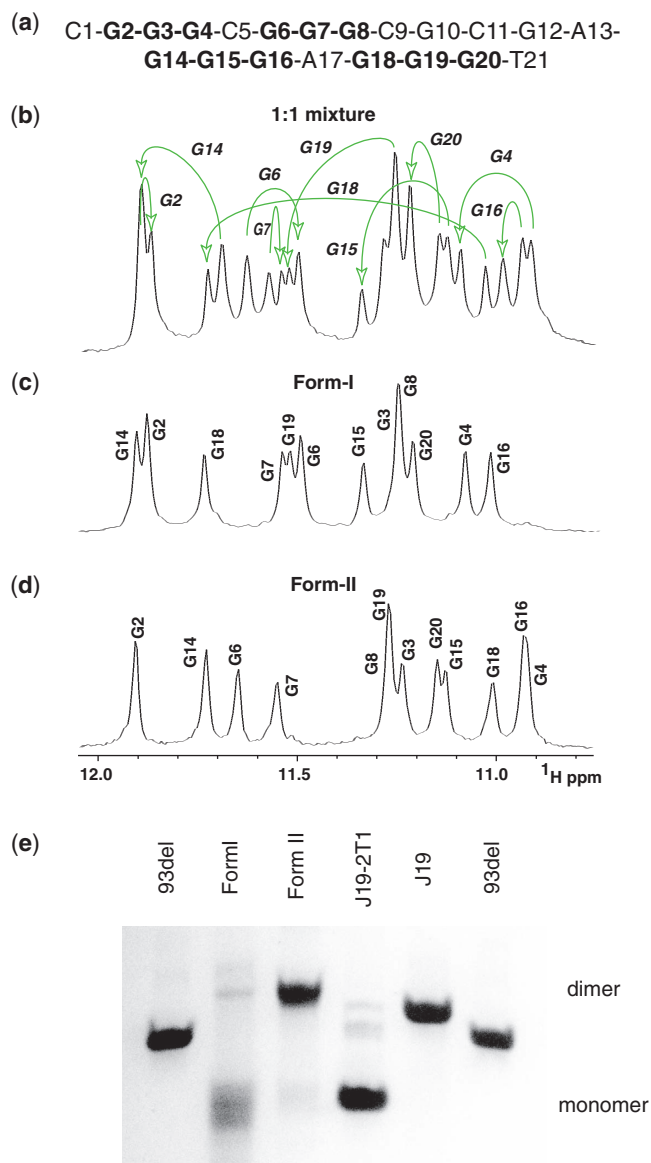


Figure 1. Imino proton NMR spectral patterns associated with form-I and form-II of the *c-kit2* promoter. (a) Chemical sequence of the *c-kit2* 21-mer sequence. The guanine residues involved in G-quadruplex formation are shown in bold. (b) Imino proton NMR spectrum of a 1:1 mixture of *c-kit2* T21 promoter form-I and form-II in slow equilibrium in 20 mM KCl, 5 mM phosphate, H₂O buffer, pH 6.8 at 43°C. (c) Imino proton NMR spectrum of the *c-kit2* T12/T21 promoter form-I in 20 mM KCl, 5 mM phosphate, H₂O buffer, pH 6.8, at 25°C, with assignments listed over the spectrum. The protocols for sample preparation of spectra in (b) to (d) are described in text (e) Non-denaturing polyacrylamide gel electrophoresis (PAGE) analysis of *c-kit2* T12/T21 (form-I) and *c-kit2* T21 (form-II) in K⁺-containing solution. Migration markers are provided: *93del*, an interlocked dimeric G-quadruplex (Phan *et al.*, 2005); *J19*: a stacked dimeric G-quadruplex (Phan, A.T., unpublished data); *J19 2T1*: a monomeric propeller-type G-quadruplex (Phan, A.T., unpublished data).

NMR spectroscopy

NMR experiments were performed on 600 MHz Varian NMR spectrometers with data recorded at 25°C.

Guanine base resonances were assigned unambiguously by using site-specific low-enrichment labeling and through-bond correlation at natural abundance (28,29). Assignment for some residues were verified and confirmed on independently synthesized samples with specific substitutions. Spectral assignments were also assisted and supported by COSY, TOCSY, ¹³C-HSQC and NOESY spectra. Interproton distances involving exchangeable protons were categorized as strong (1.2–3.8 Å), medium (2.0–6.0 Å) or weak (3.5–6.5) based on cross-peak intensities recorded in NOESY spectra (50 and 300 ms mixing time) in H₂O solution. Interproton distances involving nonexchangeable protons were measured from nuclear Overhauser enhancement (NOE) buildups using nuclear Overhauser enhancement spectroscopy (NOESY) experiments recorded at mixing times of 50, 100, 200 and 300 ms for form-I and 100, 150, 200 and 300 ms for form-II *c-kit2* G-quadruplexes in ²H₂O solution.

Structure calculations

The structures of the *c-kit2* G-quadruplexes were calculated using the X-PLOR (30) and XPLOR-NIH (v.2.11-2) (31) programs as described previously (32), with protocols differing in order to account for non-crystallographic symmetry of the dimeric form-II G-quadruplex. The initial folds guided by NMR restraints listed in Tables 1 and 2 were obtained using torsion dynamics with R^{-6} distance averaging for monomeric form-I and sum-averaging (with ambiguous restraints) for dimeric form-II G-quadruplexes. The distance restraints for dimeric form-II G-quadruplex obtained from build-up measurements were augmented by single-mixing time (300 ms) distances from the NOESY spectrum of the I4 analog of *c-kit2* promoter sequence (which showed better spectral resolution of resonances associated with G3 and I4 residues). The structures were further refined by Cartesian dynamics and, finally, using relaxation matrix refinement.

The initial fold consisted of an extended DNA strand (two strands in the case of form-II) with randomized chain torsion angles of constituent nucleotides, whose angles and bonds were set up in accordance with the most updated measurements (33,34). Folding of the dimeric form-II G-quadruplex from two extended strands resulted in substantial overpopulation by high-energetic left-handed forms, with the stacking order inverted relative to the chemical order of bases. The initial folding of the dimeric form-II G-quadruplex was therefore achieved in two steps. In the first step, only restraints for the 5'-end G-quadruplex were activated. Three independently obtained 5'-end associated right-handed G-quadruplex molecules were used in the second step, where restraints for both 5'-end and 3'-end G-quadruplexes were activated. At initial stages of dimeric form-II G-quadruplex computations, with lesser amount of restraints used, some 3'- to 3'-end oriented G-quadruplex associations were obtained (with five-residue linker in extended conformation). The NMR spectra were examined for the possible formation of cross-peaks indicating presence of dimeric 3'- to 3'-end

Table 1. Statistics of NMR restraint-guided computations of *c-kit2* promoter monomeric form-I G- quadruplex

A. NMR restraints		
Distance restraints ^a	Non-exchangeable	Exchangeable
Intra-residue distance restraints	158	0
Sequential (<i>i</i> , <i>i</i> +1) distance restraints	61	10
Long-range (<i>i</i> , <i>i</i> ≥ <i>i</i> +2) distance restraints	7	29
Other restraints		
Hydrogen bonding restraints (H-N, H-O, and heavy atoms)		52
Torsion angle restraints ^a		54
Intensity restraints		
Non-exchangeable protons (each of four mixing times)		223
B. Structure statistics of 12 molecules following intensity refinement		
NOE violations		
Number (>0.2 Å)		0.25 ± 0.45
r.m.s.d. of violations		0.02 ± 0.00
Deviations from the ideal covalent geometry		
Bond lengths (Å)		0.01 ± 0.00
Bond angles (deg.)		0.86 ± 0.02
Impropers (deg)		0.42 ± 0.03
NMR R-factor (<i>R</i> _{1/6})		0.03 ± 0.01
Pairwise all heavy atom r.m.s.d. values (12 refined structures)		
All heavy atoms in G-tetrads		0.43 ± 0.09
All heavy atoms except C9-T12		0.62 ± 0.12
All heavy atoms		1.49 ± 0.51

^aAll residues were restrained to χ values in the 240 (± 70)° range, characteristic of *anti* glycosidic torsion values. The ϵ of the residues C1-G20 was restrained to the stereochemically allowed range 225 (± 75)°. The γ torsion angle of the residues 2-4, 6-8, 14-16 and 18-21 was restrained to the values of 60 (± 35)° identified experimentally.

Table 2. Statistics of NMR restraint-guided computations of *c-kit2* promoter dimeric form-II G-quadruplex

A. NMR restraints		
Distance restraints ^a	Non-exchangeable	Exchangeable
Intra-residue distance restraints	189	0
Sequential (<i>i</i> , <i>i</i> +1) distance restraints	87	11
Long-range (<i>i</i> , <i>i</i> ≥ <i>i</i> +2) distance restraints	6	42
Other restraints		
Hydrogen bonding restraints (H-N, H-O, and heavy atoms)		104
Torsion angle restraints ^a		155
Intensity restraints		
Non-exchangeable protons (each of four mixing times)		145
B. Structure statistics of 10 molecules following intensity refinement		
NOE violations		
Number (>0.2 Å)		0.40 ± 0.70
r.m.s.d. of violations		0.03 ± 0.00
Deviations from the ideal covalent geometry		
Bond lengths (Å)		0.06 ± 0.00
Bond angles (deg.)		0.79 ± 0.06
Impropers (deg)		0.45 ± 0.03
NMR R-factor (<i>R</i> _{1/6})		0.03 ± 0.00
Pairwise all heavy atom r.m.s.d. values (10 refined structures)		
All heavy atoms in G-tetrads		0.57 ± 0.17
All heavy atoms except C5, A17		0.79 ± 0.26
All heavy atoms		1.23 ± 0.29

^aAll residues were restrained to χ values in the 240 (± 70)° range, characteristic of *anti* glycosidic torsion values. The ϵ of the residues C1-G20 was restrained to the stereochemically allowed range 225 (± 75)°. The γ torsion angle of the residues 1-4, 7-8, 14-16 and 19-21 was restrained to the values of 60 (± 35)°, the sugar pucker of the residues 2-4, 6-8, 14-16 and 18-20 was restrained in C2'-endo domain, identified experimentally.

G-quadruplex association. We do not observe NOE cross-peaks between the methyl group of T21 and imino-protons of G4 and G8, which, together with other set of observed and assigned cross-peaks, rule out formation of a dimeric 3'- to 3'-end (tail-to-tail) G-quadruplex,

in favor of a dimeric 3'- to 5'-end (tail-to-head) G-quadruplex.

Torsion dynamics. In the heating stage, the regularized extended DNA chain was subjected to 60 ps of

torsion-angle molecular dynamics at 40 000 K using a hybrid energy function composed of geometric and NOE terms. The van der Waals (vdW) component of the geometric term was set to 0.1, thus facilitating torsional bond rotations, while the NOE term included NOE-derived distances with the scaling factor of 150. The structures were then slowly cooled from 40 000 K to 1000 K over period of 60 ps during which the vdW term was linearly increased from 0.1 to 1. At the third stage, the molecules were slowly cooled from 1000 K to 300 K for 6 ps of Cartesian molecular dynamics (35). The 26 best structures of form-I with no 0.5 Å distance violations and 20 best structures of form-II, with as much as five 0.5 Å distance violations allowed, and minimal energies were selected for further refinement.

Distance-restrained molecular dynamics. Cartesian molecular dynamics was initiated at 300 K and the temperature was gradually increased to 1000 K during 7 ps. The system was equilibrated for 0.5 ps, while the force constants for the distance restraints were kept at $1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. Subsequently, the force constants were linearly scaled up to 150 during 17.5 ps. The system was then slowly cooled to 300 K in 14 ps and equilibrated at 300 K for 12 ps. The coordinates saved every 0.5 ps during the last 4.0 ps were averaged. The resulting average structure was subjected to minimization until the gradient of energy was $<0.1 \text{ kcal mol}^{-1}$. The soft planarity restraints imposed on the G-tetrads with the weight $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and base pairs with the weight $2 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ before the heating process were removed at the beginning of equilibration stage. The electrostatic term was excluded from the energy function to increase the weight of covalent geometry terms during minimization process. The dihedral and hydrogen-bonding restraints for G-tetrad formation were maintained throughout the computations.

Relaxation matrix intensity refinement. To account for spin diffusion effects, 12 best (form-I) and 10 best (form-II) distance-refined structures were next subjected to the energy minimization with back-calculation of the NOESY spectra with X-PLOR (36). The relaxation matrix was set up for the nonexchangeable protons, with the exchangeable imino and amino protons replaced by deuterons. NOE intensity volumes from 223 (form-I) and 145 (form-II) nonexchangeable cross-peaks for each of four mixing times (50, 100, 200 and 300 ms for form-I and 100, 150, 200 and 300 ms for form-II) were used as restraints, with uniform upper and lower bounds of $\pm 30\%$. Dynamics was started at 5 K, and the system was heated up to 300 K in 0.6 ps. During the subsequent relaxation, the force constant for NOE intensities was gradually increased from 0 to $300 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ with simultaneous decrease of the distance force constant of nonexchangeable protons from 50 to $30 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. The force constant for exchangeable protons and hydrogen bonds was kept at $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. After equilibration at 300 K for 3.0 ps the resulting structure was subjected to minimization until the gradient of energy was $<0.1 \text{ kcal mol}^{-1}$. The NMR *R*-factor ($R_{1/6}$)

improved from initial value of 6.0% to 3.0% with simultaneous improvement of structural convergence. Two different rotational conformers obtained for form-II at distance refinement step converged to single form after intensity refinement. One additional round of distance refinement with subsequent intensity refinement was undertaken to generate 10 best conformers of the dimeric form-II G-quadruplex.

RESULTS

Sample preparation

Our studies have focused on the *c-kit2* sequence analog where T21 replaced G21 (designated *c-kit2* T21 promoter, Figure 1a), because this analog, unlike the parent sequence, gave better resolved imino proton NMR spectra in K^+ -containing solution, as reported previously (21).

We used either a spin-down column or equilibrium dialysis methods for purification and buffer exchange of the *c-kit2* samples used for NMR studies (see 'Materials and methods' section for protocol details). An NMR sample (0.21 mM in strands) prepared using the spin-down approach was dissolved in 20 mM KCl, 5 mM phosphate, H_2O , next heated to 90°C , and subsequently cooled over a 2-h period. The imino proton NMR spectrum recorded at 43°C after the sample stood overnight at this temperature is shown in Figure 1b, and consisted of doubling of the imino proton resonances, indicative of roughly equimolar amounts of two conformations in slow exchange on the NMR timescale. One of the conformers corresponded to form-I, while the other is designated form-II.

We subsequently found that G12 to T12 substitution of *c-kit2* T21 promoter (designated *c-kit2* T12/T21 promoter) stabilized form-I in 20 mM KCl solution. An NMR sample of the *c-kit2* T12/T21 promoter (0.54 mM in strands), prepared using the equilibrium dialysis approach, was dissolved in 20 mM KCl, 5 mM K-phosphate, H_2O , pH 6.8. Its imino proton NMR spectrum recorded at 25°C is shown in Figure 1c, and consists of a single conformer exhibiting 12 partially resolved imino protons between 11.0 and 12.0 ppm, thereby allowing us to collect two-dimensional NMR spectra of form-I in the absence of form-II.

An NMR sample of the *c-kit2* T21 promoter (0.24 mM in strands), prepared using the equilibrium dialysis approach, was dissolved in 100 mM KCl, 5 mM K-phosphate, H_2O , pH 6.8. Its imino proton NMR spectrum recorded at 25°C is shown in Figure 1d, and consists of a single conformer exhibiting 12 partially resolved imino protons between 10.8 and 12.0 ppm corresponding to form-II. This spectrum recorded under higher KCl (100 mM) conditions, has allowed us to collect two-dimensional NMR spectra for form-II in the absence of form-I.

In the following sections, we report on our NMR-based structure determination of form-I G-quadruplex formed by *c-kit2* T12/T21 promoter in 20 mM KCl-containing

solution and form-II G-quadruplex formed by *c-kit2* T21 promoter in 100 mM KCl-containing solution. Our structural studies establish that form-I is an all-parallel-stranded monomeric G-quadruplex, while form-II is a novel all-parallel-stranded dimeric G-quadruplex. The conclusions related to oligomerization state (monomer versus dimer) are verified by gel shift data outlined in Figure 1e, with the data presented and analyzed at the beginning of the 'Discussion' section.

NMR spectra of *c-kit2* form-I and form-II under annealing conditions

Imino proton NMR spectra of the *c-kit2* T12/T21 promoter (prepared using equilibrium dialysis, 1.12 mM in strands) in 20 mM KCl-containing solution have been recorded at 25°C (Figure S2a), as well as immediately following heating to 85°C and gradual annealing to 25°C (Figure S2b), and at this temperature after 7 days (Figure S2c) and 14 days (Figure S2d). The imino proton spectra are characteristic of the *c-kit2* form-I conformation both prior to (Figure S2a) and after annealing (Figure S2b), as well as over time at 25°C (Figures S2c and d).

Imino proton NMR spectra of the *c-kit2* T21 promoter (prepared using equilibrium dialysis, 0.065 mM in strands) in 20 mM KCl-containing solution have been recorded at 25°C (Figure S3a), as well as immediately following heating to 85°C and gradual annealing to 25°C (Figure S3b), and at this temperature after 1 day (Figure S3c), 4 days (Figure S3d) and 8 days (Figure S3e). The imino proton NMR spectra of the freshly prepared sample is a mixture of *c-kit2* forms-I and -II (Figure S3a), which converts to predominantly form-I after annealing (Figure S3b), and then gradually converts to form-II over time (Figure S3c–e; characteristic form-I resonances indicated by a pair of arrows).

Imino proton NMR spectra of the *c-kit2* T21 promoter (prepared using equilibrium dialysis, 0.065 mM in strands) in 100 mM KCl-containing solution have been recorded at 25°C (Figure S4a), as well as immediately following heating to 85°C and gradual annealing to 25°C (Figure S4b), and at this temperature after 16 h (Figure S4c), 46 h (Figure S4d) and 97 h (Figure S4e). The imino proton NMR spectra of the freshly prepared sample is characteristic of *c-kit2* form-II (Figure S4a), converts to a mixture of forms-I and -II after annealing (Figure 3b), and then gradually converts back to form-II over time (Figure S4c–e; characteristic form-I resonances indicated by a pair of arrows).

It should be noted that the imino proton NMR spectra of the *c-kit2* T21 promoter (prepared using equilibrium dialysis) in 100 mM KCl-containing solution at 25°C, with spectral characteristic of *c-kit2* form-II, remains unchanged as the strand concentration varies from 0.065 mM (Figure S5a), to 0.13 mM (Figure S5b), to 0.26 mM (Figure S5c), to 0.52 mM (Figure S5d). This implies that the *c-kit2* form-II conformer is independent of concentration over a 10-fold dilution range.

Proton NMR spectra of form-I fold of the *c-kit2* promoter

All NMR experiments on form-I were performed on samples prepared by equilibrium dialysis for purification and buffer exchange. The exchangeable proton spectrum of *c-kit2* T12/T21 promoter (0.54 mM in strands) corresponding to form-I in 20 mM KCl, 5 mM K-phosphate buffer, pH 6.8 at 25°C is plotted in Figure 2a. We observe 12 exchangeable imino protons between 11.0 and 12.0 ppm, in a spectral range characteristic of guanine imino protons involved N-H•••O hydrogen bond formation, a feature characteristic of G-tetrad formation (3). We pursued a strategy of assigning the guanine imino protons by incorporating 2% ¹⁵N-labeled-guanines one at a time into the *c-kit2* T12/T21 promoter sequence (Figure 2b) (29) and then correlating these established assignments with guanine H8 protons (Figure 2c) through long-range through-bond J-coupling experiments (data collected on a sample that was 1.7 mM in strands, Figure 2d) (28). The imino proton assignments of the *c-kit2* T12/T21 promoter are listed above the control NMR spectrum in Figure 2a, and are analyzed in the context of the four-guanine tracts G2-G3-G4, G6-G7-G8, G14-G15-G16 and G18-G19-G20 in the 21-mer sequence (Figure 1a). We observe four slowly exchanging guanine imino protons assigned to G3, G7, G15 and G19 in the spectrum of the *c-kit2* T12/T21 promoter recorded 1 h following transfer from 20 mM KCl, 5 mM phosphate, H₂O solution (after lyophilization) to its ²H₂O counterpart (Figure 2e).

The expanded NOESY contour plot (250 ms mixing time) correlating base and sugar H1' protons of *c-kit2* T12/T21 promoter (1.7 mM in strands) in 20 mM KCl, 5 mM K-phosphate buffer, H₂O, pH 6.8 at 25°C is shown in Figure 2f. The corresponding NOESY contour plot (250 ms mixing time) of the *c-kit2* T12/T21 promoter (1.1 mM in strands) recorded on a separately prepared sample in ²H₂O at 25°C is plotted in Figure S6 and exhibits better resolution. We can trace the NOEs between the base protons and their own and 5'-flanking sugar H1' protons and these connectivities are traced in Figure 2f (sample in H₂O) and Figure S6 (separate sample in ²H₂O), yielding the base and sugar proton assignments listed in Table S1 (Supplementary Data). The remaining sugar protons were assigned by through-bond COSY and TOCSY experiments and are also listed in Table S1. We do not observe strong NOEs between base and sugar H1' protons at short (50 ms) mixing times, indicative of *anti* torsion angles (37) for all guanines for form-I of the *c-kit2* T12/T21 promoter.

Fold of the *c-kit2* promoter form-I G-quadruplex

We have assigned guanines to the three G-tetrads by monitoring NOEs between guanine imino and H8 protons around individual G-tetrad planes (Figure 3a) in the NOESY spectrum (mixing time 250 ms) of *c-kit2* T12/T21 promoter (1.7 mM in strands) in 20 mM KCl, 5 mM K-phosphate, H₂O, pH 6.8, at 25°C (Figure 3b). Thus, the imino proton of G2 shows an NOE to the H8 proton of G6 (labeled 2/6 in red), the imino proton G6 shows an

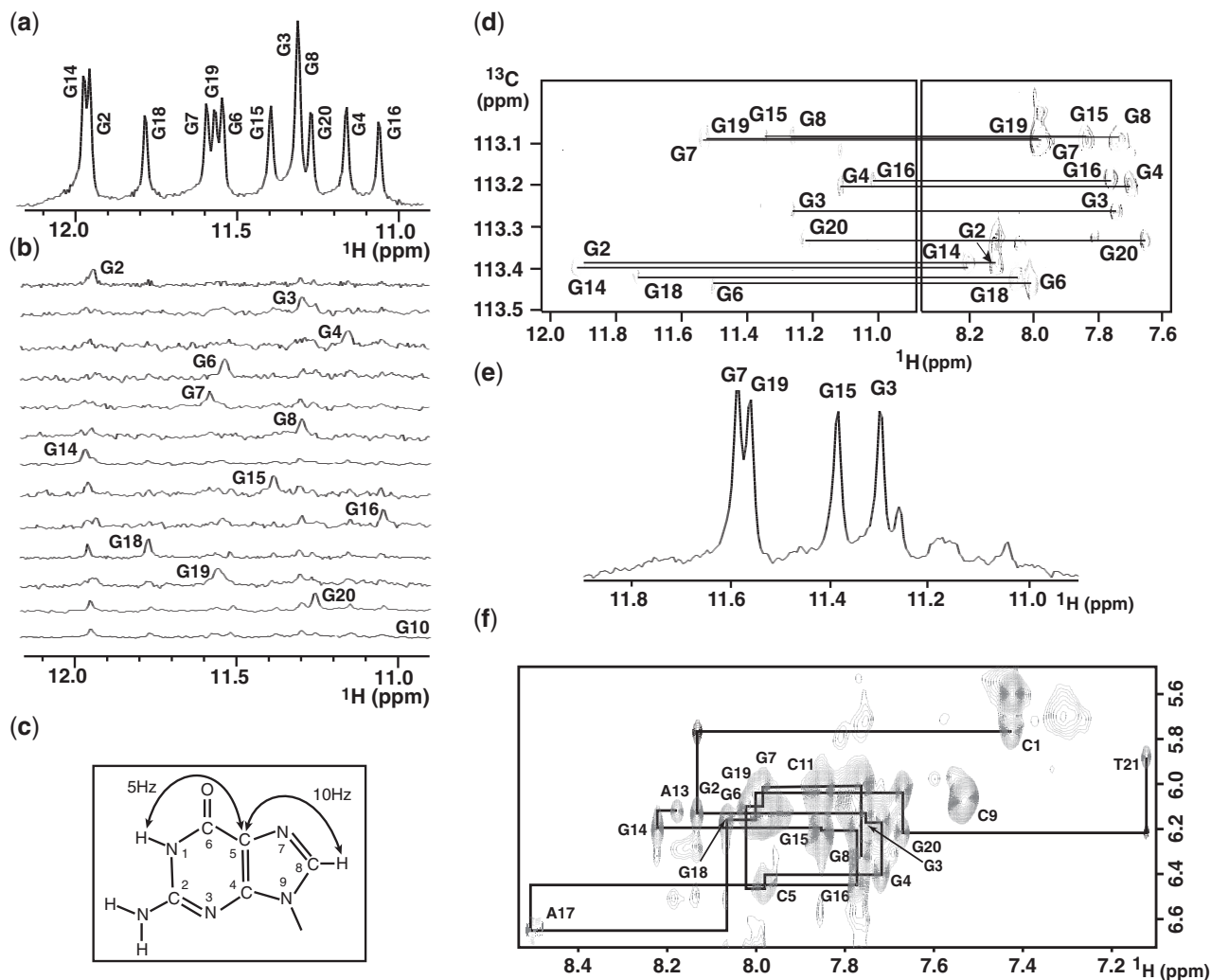


Figure 2. Exchangeable proton NMR spectra and assignments of form-I conformer of the *c-kit2* promoter. (a) Imino proton spectrum of *c-kit2* T12/T21 promoter form-I in 20 mM KCl, 5 mM phosphate, H₂O buffer, pH 6.8, at 25°C. Unambiguous imino proton assignments are listed over the spectrum. (b) Imino protons were assigned in ¹⁵N-filtered NMR spectra of samples containing 2% uniformly-¹⁵N,¹³C-labeled guanines at the indicated positions. (c) A schematic indicating long-range J couplings used to correlate imino and H8 protons within the guanine base. (d) Guanine H8 proton assignments of *c-kit2* promoter form-I based on through-bond correlations between imino and to be assigned H8 protons via ¹³C5 at natural abundance, using long-range J couplings shown in (c). (e) Imino proton spectrum of *c-kit2* promoter form-I recorded 1 h following transfer from H₂O (after lyophilization) to ²H₂O solution. Assignments of slowly exchanging imino protons are listed over the spectrum. (f) Expanded NOESY contour plot (250 ms mixing time) correlating base and sugar H1' non-exchangeable proton assignments of *c-kit2* T12/T21 promoter form-I in 20 mM KCl, 5 mM phosphate, ²H₂O buffer, pH 6.8, at 25°C. The line connectivities trace NOEs between base protons (H8 or H6) and their own and 5'-flanking sugar H1' protons. Intraresidue base to sugar H1' NOEs are labeled with residue numbers.

NOE to the H8 proton of G14, the imino proton of G14 shows an NOE to the H8 proton of G18, and the imino proton of G18 shows an NOE to the H8 proton of G2, thereby linking the guanines and their order around the G2•G6•G14•G18 G-tetrad plane (assignments in red) (Figure 3b–d). Related NOE tracings identify the G3•G7•G15•G19 (assignments in orange) and G4•G8•G16•G20 (assignments in blue) G-tetrads that generate the form-I G-quadruplex (Figure 3b–d). Since the imino protons of the G3•G7•G15•G19 G-tetrad exhibit the slowest exchange rates (Figure 2e), this central G-tetrad must be bracketed on either side by stacked G2•G6•G14•G18 and G4•G8•G16•G20 G-tetrads, resulting in the backbone tracing from G1 to T21 for the *c-kit2* monomeric form-I G-quadruplex as

shown in Figure 3d. All guanines adopt *anti* glycosidic torsion angles and all three loops are of the double-chain-reversal type (38). Loop 1 is made of a single nucleotide (C5), loop 2 is composed of 5 nt (C9-G10-C11-G12-A13), while loop 3 is also made up of a single nucleotide (A17).

Solution structure of the monomeric form-I G-quadruplex

Initial distance-restrained and subsequent intensity-restrained molecular dynamics calculations (see 'Materials and methods' section) of the solution structure of the *c-kit2* T12/T21 form-I monomeric G-quadruplex were guided by exchangeable and non-exchangeable proton restraints, with the numbers listed by category in

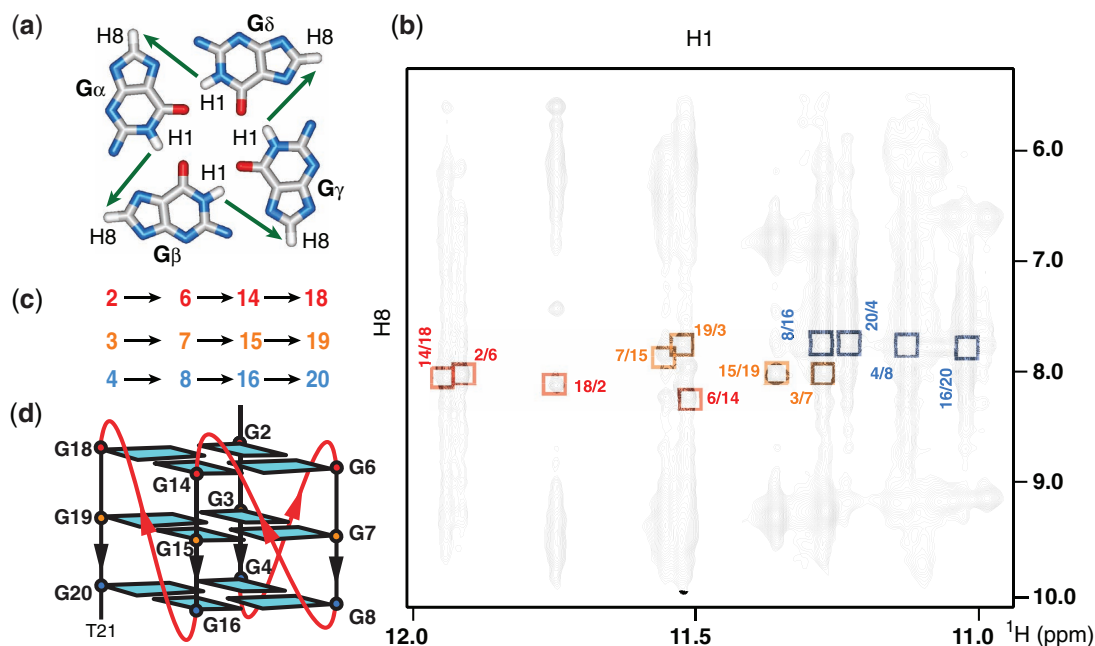


Figure 3. Assignment of individual G-tetrads within the *c-kit2* promoter form-I G-quadruplex. (a) Characteristic guanine imino-H8 NOE connectivity patterns around a G-tetrad (labeled $G\alpha\bullet G\beta\bullet G\gamma\bullet G\delta$) as indicated by arrows. (b) Expanded NOESY contour plot (mixing time = 250 ms) showing imino to H8 connectivities of *c-kit2* T12/T21 promoter form-I in 20 mM KCl, 5 mM phosphate, H_2O , pH 6.8, at 25°C. Cross-peaks identifying connectivities within individual G-tetrads are color-coded, with each cross-peak framed and labeled with the residue number of imino proton in the first position and that of the H8 proton in the second position. (c) Guanine imino-H8 connectivities observed for $G2\bullet G6\bullet G14\bullet G18$ (red), $G3\bullet G7\bullet G15\bullet G19$ (orange) and $G4\bullet G8\bullet G16\bullet G20$ (blue) G-tetrads. (d) Schematic representation of the *c-kit2* promoter monomeric form-I G-quadruplex fold. *Anti* bases are indicated in cyan and loop connectivities shown by red lines.

Table 1. The ensemble of 12 refined superpositioned structures is shown in stereo in Figure 4a, with a representative refined structure in the same orientation shown in a ribbon representation in Figure 4b and a surface representation in Figure 4c. The ensemble of refined structures is well converged, exhibiting pairwise rmsd values in the 0.43-range for the stacked G-tetrad core (Table 1).

The first loop is of the double-chain-reversal type and is bridged by a single C5 nucleotide, which unexpectedly stacks with C9 from the second double-chain-reversal loop (Figure 4d, stereo view of refined structures in Figure S7a), indicative of unanticipated interaction between residues in adjacent double-chain-reversal loops. In this regard, the observation of NOEs between the sugar protons of C5 and C9, namely $C5(H2',2'')\text{-}C9(H5)$, validates this proposed alignment. The second loop is also of the double-chain-reversal type and is composed of five nucleotide-segment C9-G10-C11-G12-A13 (Figure 4e), and despite its length, is reasonably well ordered (stereo view of refined structures in Figure S7b). The partial overlap of the bases of the non-adjacent G10 and T12 in this 5-nt double-chain-reversal loop in the *c-kit2* monomeric form-I G-quadruplex (Figure S8a) can be compared with the partial overlap of the bases of the non-adjacent T5 and A7 in the four-repeat human telomere G-quadruplex for the all-parallel-stranded form in the crystal (Figure S8b) (39) and for the related (3+1) form in solution (Figure S8c) (40–42). The third loop is of the double-chain-reversal type and is bridged by a single well-defined A17 nucleotide (Figure S9a and stereo view of refined structures in Figure S7c). Finally, a reversed

$C1\bullet A13$ non-canonical pair stacks over the $G14\bullet G18$ segment of the top $G2\bullet G6\bullet G14\bullet G18$ tetrad (Figure S9b), while terminal residue T21 stacks on G20 of the bottom $G4\bullet G8\bullet G16\bullet G20$ tetrad (Figure S9c).

Proton NMR spectra of form-II fold of the *c-kit2* promoter

The majority of NMR experiments on form-II were performed on samples prepared by spin-down column for purification and buffer exchange. The only exceptions involved the use of equilibrium dialysis for experiments with H8-deuterated G labels, I for G substitutions and 5mC for C substitutions. The exchangeable proton spectrum of *c-kit2* T21 promoter (0.3 mM in strands) corresponding to form-II in 100 mM KCl, 5 mM K-phosphate buffer, pH 6.8 at 25°C is plotted in Figure 5a. We observe 12 partially resolved exchangeable imino protons between 10.7 and 12.0 ppm, in a spectral range characteristic of guanine imino protons involved in G-tetrad formation. The imino proton NMR spectrum of the *c-kit2* T21 promoter in 100 mM KCl prepared by the spin-down approach (Figure 5a) differs somewhat from its counterpart prepared by the equilibrium dialysis approach (Figure 1d), but both exhibit four resolved imino protons between 11.5 and 12.0 ppm, while exhibiting small chemical shift differences among imino proton resonances between 10.9 and 11.3 ppm.

We pursued a strategy of assigning the guanine imino protons by incorporating 2% ^{15}N -labeled-guanines either in pairs or one at a time into the *c-kit2* T21 promoter sequence (Figure 5b) and then correlating these

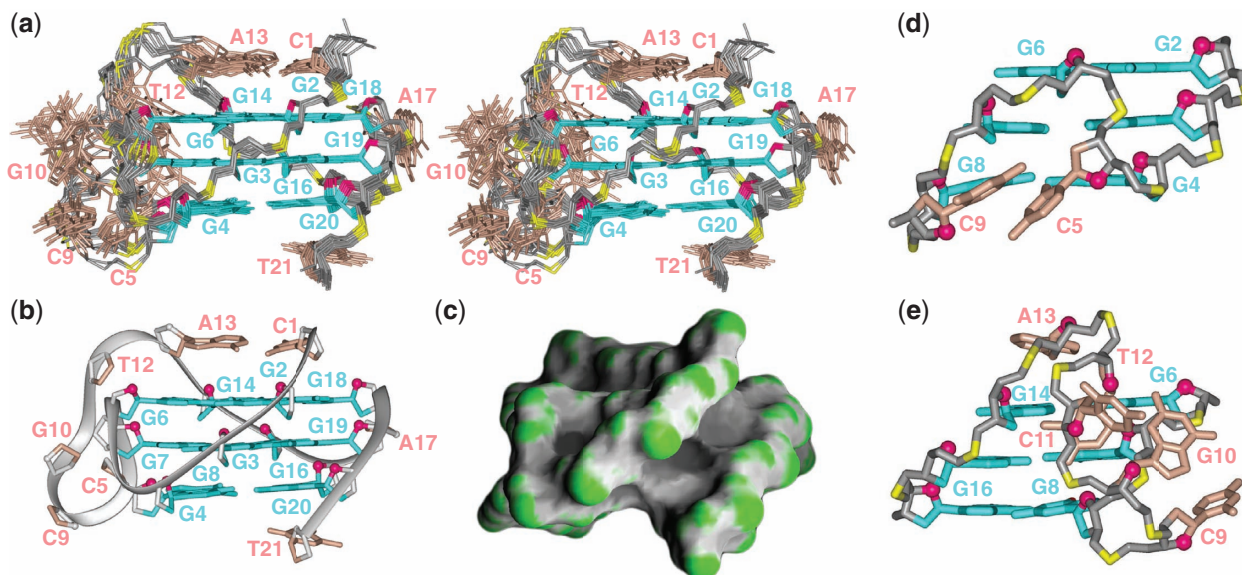


Figure 4. Solution structure of the *c-kit2* promoter monomeric form-I G-quadruplex in K^+ solution. (a) Stereo view of 12 superpositioned refined structures of the *c-kit2* T12/T21 promoter monomeric form-I G-quadruplex. Guanine bases in the G-tetrad core are colored cyan (*anti*). Bases in connecting loops are in biscuit color, with the backbone in gray, sugar ring oxygens in red and phosphorus atoms in yellow. (b) Ribbon and (c) surface representations of a representative refined structure. (d) Cross-groove stacking between the C5 and C9 bases (biscuit) in the C5-containing groove. (e) Base alignment within the five-residue C9-G10-C11-T12-A13-containing loop.

established assignments with guanine H8 protons (data collected on a sample that was 3.9 mM in strands, Figure 5c) through long-range through-bond J-coupling experiments (Figure 2c). The imino proton assignments of the *c-kit2* T21 promoter are listed above the control NMR spectrum in Figure 5a, and are analyzed in the context of the four-guanine tracts G2-G3-G4, G6-G7-G8, G14-G15-G16 and G18-G19-G20 in the 21-mer sequence (Figure 1a). We observe four slowly exchanging guanine imino protons assigned to G3, G7, G15 and G19 in the spectrum of the *c-kit2* T21 promoter recorded 1 h following transfer from 100 mM KCl-containing H_2O solution (after lyophilization) to its 2H_2O counterpart (Figure 5d).

The expanded NOESY contour plot (300 ms mixing time) correlating base and sugar H1' protons of *c-kit2* T21 promoter (0.5 mM in strands) in 100 mM KCl, 5 mM K-phosphate, 2H_2O , pH 6.8 at 25°C is shown in Figure 6a. Due to spectral overlap, we have also prepared samples containing single guanines deuterated at the H8 position, so as to independently confirm specific guanine H8 assignments in crowded regions of the NOESY contour plot. Such assignments can be readily made following comparison of the expanded NOESY contour plot for undeuterated *c-kit2* T21 promoter (Figure 6b), with its counterparts following deuteration of the H8 of G12 (Figure 6c), and the H8 of G14 (Figure 6d). Missing cross-peaks as a consequence of H8 deuteration are labeled with an 'x' in Figure 6c and d. We can trace the NOEs between the base protons and their own and 5'-flanking sugar H1' protons and these connectivities are traced in Figure 6a.

Inosine for guanine substituted derivative oligonucleotides (at positions G4 and G16) were also investigated (specific expanded regions of NOESY contour plots

shown in Figure 6e and f), with for example, I4 for G4 substitution, resulting in resolution of certain spectral overlaps in the NOESY contour plot in Figure S10 and were used for assignment purposes. Thus, we observe NOEs between H2 of I4 and H8 of G8 (peak a, Figure 6e) and G7 (peak b, Figure 6e), and between H2 of I16 and H8 of G20 (peak a, Figure 6f) and G19 (peak b, Figure 6f), as well as between the imino of I16 and H8 of G20 (peak c, Figure 6f) and G19 (peak d, Figure 6f).

Assignment of cytosines was achieved by incorporation of 5-methyl-cytosines at positions C11 (Figure 6g) and C5 (Figure 6h), which resulted in upfield shifts of the respective H6 protons (cross-peaks labeled in green for C11 and in red for C5 in Figure 6g and h).

The remaining sugar protons were assigned by through-bond COSY and TOCSY experiments, with proton chemical shifts listed in Table S2. We do not observe strong NOEs between base and sugar H1' protons at short (50 ms) mixing times, indicative of *anti* torsion angles (37) for all guanines for form-II of the *c-kit2* T21 promoter.

Fold of the *c-kit2* promoter form-II G-quadruplex

We have assigned guanines to specific G-tetrads by monitoring NOEs between guanine imino and H8 protons around individual G-tetrad planes (Figure 7a) in the NOESY spectrum (mixing time 250 ms) of the *c-kit2* T21 promoter (1.1 mM in strands) in 100 mM KCl, 5 mM K-phosphate, H_2O , pH 6.8, 25°C (Figure 7a). In an effort to overcome assignment problems associated with near degeneracies in cross-peak overlaps, we have also utilized samples selectively deuterated at the H8 positions of individual guanines in the *c-kit2* T21 promoter. Expanded NOESY contour plots are compared for the

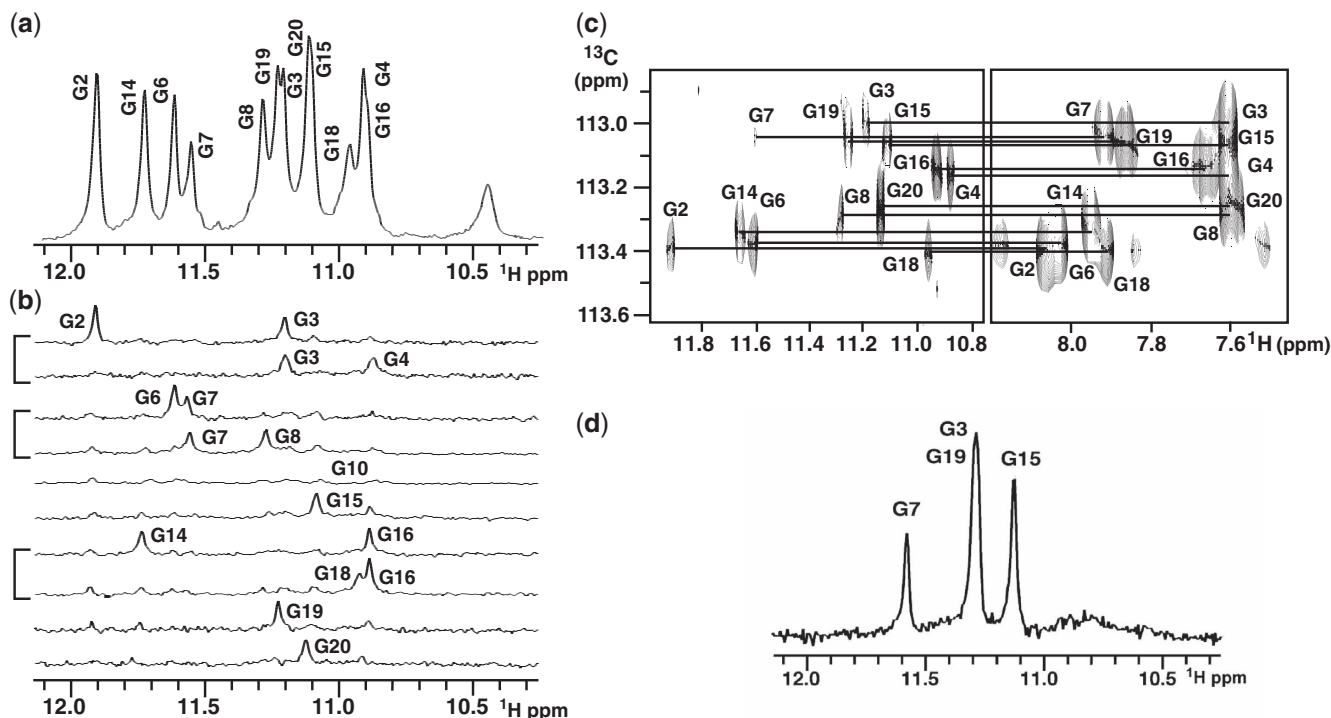


Figure 5. Exchangeable proton NMR spectra and assignments of form-II of the *c-kit2* promoter. (a) Imino proton spectrum of *c-kit2* T21 promoter form-II in 100 mM KCl, 5 mM phosphate, H₂O buffer, pH 6.8, at 25°C. Unambiguous imino proton assignments are listed over the spectrum. (b) Imino protons were assigned in ¹⁵N-filtered NMR spectra of samples containing 2% uniformly-¹⁵N,¹³C-labeled guanines at the indicated positions. Labels were incorporated either singly (G10, G15, G19, G20) or in pairs (G2 and G3, G3 and G4, G6 and G7, G7 and G8, G14 and G16, G16 and G18). (c) Guanine H8 proton assignments of *c-kit2* T21 promoter form-II based on through-bond correlations between assigned imino and to be assigned H8 protons via ¹³C5 at natural abundance, using long-range J couplings shown in Figure 1c. (d) Imino proton spectrum of *c-kit2* promoter form-II recorded 1 h following transfer from H₂O (after lyophilization) to ²H₂O solution. Assignments of slowly exchanging imino protons are listed over the spectrum.

undeuterated sample (Figure 7b), as well as deuteration of H8 of G3 (Figure 7c), H8 of G14 (Figure 7d), and H8 of G19 (Figure 7e). Missing cross-peaks as a consequence of H8 deuteration are labeled with an 'x' in Figure 7c and d.

We observe an unusual pattern of reciprocal NOEs as exemplified by cross-peaks between the imino proton of G2 and the H8 proton of G6, as well as between the imino proton of G6 and the H8 proton of G2 (boxed cross-peaks in red in Figure 7a), consistent with formation of a G2•G6•G2•G6 tetrad (red, Figure 7f). Other pairs of reciprocal NOEs (Figure 7a) identify G3•G7•G3•G7 (green, Figure 7f), G4•G8•G4•G8 (blue, Figure 7f), G14•G18•G14•G18 (cyan, Figure 7f), G15•G19•G15•G19 (magenta, Figure 7f) and G16•G20•G16•G20 (orange, Figure 7f) tetrads. Such a pattern of six G-tetrads can only be consistent with an all-parallel-stranded dimeric G-quadruplex shown schematically in Figure 7g for the *c-kit2* form-II G-quadruplex. Within each monomeric G-quadruplex component of the *c-kit2* dimeric form-II G-quadruplex, G3•G7•G3•G7 and G15•G19•G15•G19 constitute the central G-tetrads (Figure 7g), consistent with the observed slow exchange of the imino protons at positions G3, G7, G15 and G19 (Figure 5d).

Further, in NOESY spectra of inosine-substituted oligonucleotides (I4 for G4 and I16 for G16), we observe cross-peaks I4(H2)/G8(H8) and I4(H2)/G7(H8) (peaks a and b, respectively, Figure 6e), as well as

I16(H2)/G20(H8) and I16(H2)/G19(H8) (peaks a and b, respectively, Figure 6f), which independently corroborate features of the dimeric G-quadruplex core shown schematically on Figure 7g. All guanines adopt *anti* glycosidic torsion angles and the single-nucleotide first (C5) and third (A17) loops are of the double-chain-reversal type, with the central 5-nt C9-G10-C11-G12-A13 segment bridging the individual 5' (or upper) and 3' (or lower) G-quadruplex components in the context of the dimeric G-quadruplex.

Solution structure of the dimeric form-II G-quadruplex

Initial distance-restrained and subsequent intensity-restrained molecular dynamics calculations (see 'Materials and methods' section) of the solution structure of the *c-kit2* T21 dimeric form-II G-quadruplex were guided by exchangeable and nonexchangeable proton restraints, with the numbers listed by category in Table 2. The ensemble of 10 refined superpositioned structures is shown in stereo in Figure 8a, with a representative refined structure in the same orientation shown in a ribbon representation in Figure 8b and a surface representation in Figure 8c. The ensemble of refined structures is well converged, exhibiting pairwise rmsd values in the 0.57 range for the pair of stacked G-tetrad cores (Table 2).

The backbone of the C9-G10-C11-G12-A13 linker sequence, which connects the 5' (at position G8) and 3'

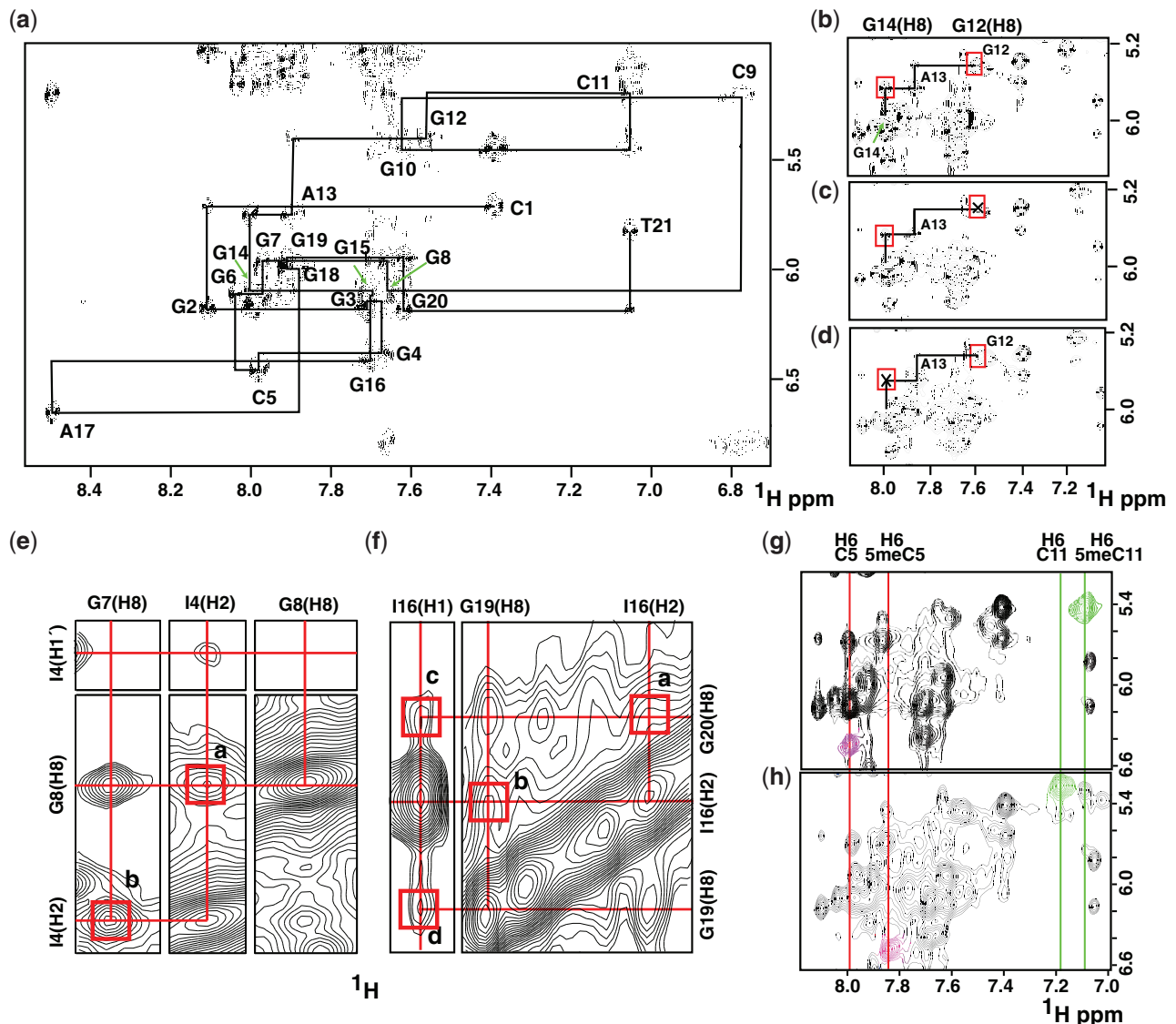


Figure 6. Non-exchangeable proton NMR spectra and assignments of form-II conformer of the *c-kit2* T21 promoter. (a) Expanded NOESY contour plot (300 ms mixing time) correlating base and sugar H1' non-exchangeable proton assignments of *c-kit2* T21 promoter form-II in 100 mM KCl, 5 mM phosphate, $^2\text{H}_2\text{O}$ buffer, pH 6.8, at 25°C. The line connectivities trace NOEs between base protons (H8 or H6) and their own and 5'-flanking sugar H1' protons. Intraresidue base to sugar H1' NOEs are labeled with residue numbers. (b–d) Expanded contour plots tracing the G12–A13–G14 base to H1' connectivities for undeuterated sample (panel b), for sample deuterated at H8 of G12 (c), and for sample deuterated at H8 of G14 (d). Missing cross-peaks are labeled by an 'x' symbol. (e, f) Expanded NOESY contour plots (300 ms mixing times) of I4-substituted (e) and I16-substituted (f) *c-kit2* T21 promoters in 100 mM KCl, 5 mM phosphate, $^2\text{H}_2\text{O}$ buffer, pH 6.8, at 25°C. Relevant cross-peaks (red box) are labeled a and b in (e) and labeled a–d in (f). (g, h) Expanded NOESY contour plots (300 ms mixing times) of 5meC11-substituted (g) and 5meC5-substituted (h) *c-kit2* T21 promoters in 100 mM KCl, 5 mM phosphate, $^2\text{H}_2\text{O}$ buffer, pH 6.8, at 25°C. Relevant cross-peaks are color-coded (labeled in red for H6 of C5/5meC5 and in green for H6 of C11/5meC11), with the cytosine H6 protons shifting upfield on methylation of adjacent cytosine ring 5 positions.

(at position G14) G-quadruplexes, adopts an inverse S-shaped fold (Figure 8d). A13 residues from partner strands form a *trans* A13•A13 non-canonical pair aligned through pairing of their Watson–Crick edges, with the 5' and 3' G-quadruplexes sandwiching this pair (Figure 8e and f).

The residues within the C9–G10–C11–G12–A13 linker sequence are well-defined, including stacked C11 and G12 bases, indicative of an ordered five-nucleotide linker segment (Figure 8d and stereo view in Figure S11; also stereo views of the refined structures in Figure S12a and b). The inter-residue NOE cross peaks that contribute

to the fold of the C9–G10–C11–G12–A13 linker sequence are listed in Table S3, with an interpretation of the key NOEs listed as footnotes to Table S3.

The relative alignment of the A13•A13 non-canonical pair sandwiched between the G4•G8•G4•G8 and G14•G18•G4•G18 tetrads (Figure 7g) is defined by a set of inter-residue NOE cross-peaks listed in Table S4, with an interpretation of the key NOEs listed as footnotes to Table S3.

Single residue loops C5 and A17 are positioned with respect to each other by approx 90° rotation around the helical axis (colored in purple in stereo views of the refined

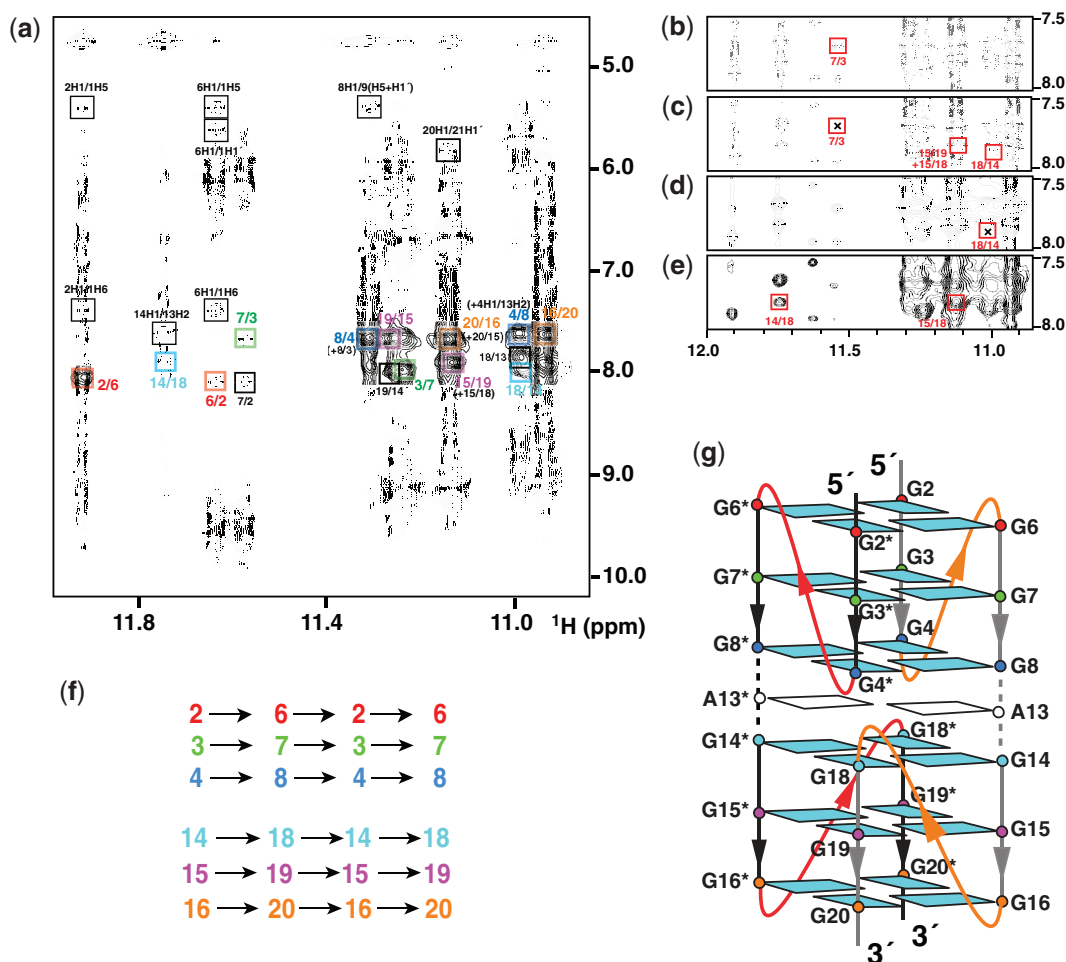


Figure 7. Assignment of individual G-tetrads within the *c-kit2* promoter form-II G-quadruplex. (a) Expanded NOESY contour plot (mixing time = 250 ms) showing imino to H8 connectivities of *c-kit2* T21 promoter form-II in 100 mM KCl, 5 mM phosphate, H₂O, pH 6.8, at 25°C. Cross-peaks identifying connectivities within individual G-tetrads are color-coded, with each cross-peak framed and labeled with the residue number of imino proton in the first position and that of the H8 proton in the second position. (b–e) Expanded NOESY contour plots showing NOE connectivities between guanidine imino and H8 protons following deuteration of specific guanidine H8 protons. Undeuterated sample is shown in (b). Sample deuterated at H8 of G3 is shown in (c). Sample deuterated at H8 of G14 is shown in panel d. Sample deuterated at H8 of G19 is shown in (e). Missing cross-peaks as a result of deuteration are labeled by an 'x'. (f) Guanine imino-H8 connectivities observed for G2•G6•G2•G6 (red), G3•G7•G3•G7 (green), G4•G8•G4•G8 (blue), G14•G18•G14•G18 (cyan), G15•G19•G15•G19 (magenta), and G16•G20•G16•G20 (orange) G-tetrads. (g) Schematic representation of the *c-kit2* T21 promoter dimeric form-II G-quadruplex fold. *Anti* bases are indicated in cyan and loop connectivities shown by red and orange lines.

structures in Figure S12a and b), such that residue A17 becomes a part of the cluster of C9-G10-C11-G12-A13 linker residues. The individual grooves of the 5' and 3' G-quadruplexes, that are separated by the A13•A13 non-canonical pair, align to form one continuous groove in the structure of the dimeric G-quadruplex (represented by a pair of dashed lines in Figure 8e).

Time dependence of imino proton NMR spectrum of the dimeric form-II G-quadruplex

We observe small chemical shift changes for a subset of imino protons in the *c-kit2* promoter dimeric form-II G-quadruplex spectrum on standing over period of many months in the refrigerator (Supplementary Figure S13a changes over time to Figure S13b). The signature of a *c-kit2* promoter dimeric form-II G-quadruplex is retained over time since we still observe four

well-dispersed imino protons (G2, G14, G6 and G21) between 11.5 and 12.0 ppm (Figure S12b). The small time-dependent chemical shift changes are observed for several guanine imino protons, amongst those include G4, G8, G14 and G18, the very guanines located at the interface between 5' and 3' quadruplexes (Figures S13a and S13b). These results imply the existence of an as-yet undefined time-dependent conformational stabilization at the interface between the two halves of the dimeric form-II G-quadruplex.

DISCUSSION

Our studies establish that the *c-kit2* promoter T12/T21 sequence folds to form an intramolecular monomeric all-parallel stranded form-I G-quadruplex in 20 mM KCl-containing buffer solution (Figures 3d and 4b),

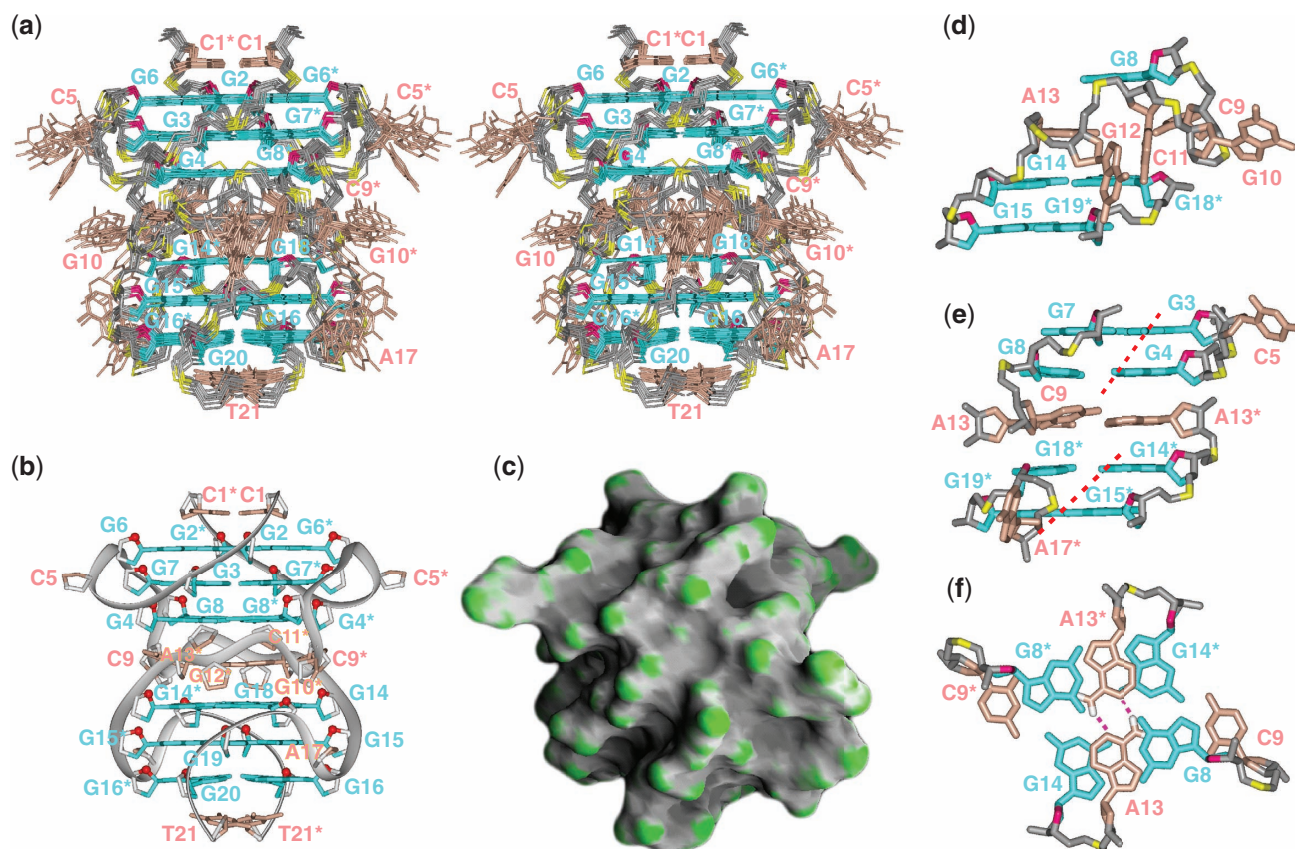


Figure 8. Solution structure of the *c-kit2* promoter dimeric form-II G-quadruplex in K^+ solution. (a) Stereo view of 10 superpositioned refined structures of the *c-kit2* T21 promoter dimeric form-II G-quadruplex. Guanine bases in the G-tetrad core are colored cyan (*anti*). Bases in connecting loops are in biscuit color, with the backbone in gray, sugar ring oxygens in red and phosphorus atoms in yellow. Residues from the second strand are marked with an asterisk (b) Ribbon and (c) surface representations of a representative refined structure. Residues from the second strand are marked with an asterisk. (d) Base alignment within the inverted S-shaped five-residue C9-G10-C11-T12-A13-containing loop. (e, f) *trans* A13•A13 non-canonical pair is sandwiched between 5' and 3'-G-quadruplexes. Also labeled are residues C5 and A17 involved in single residue double-chain-reversal loops in (e). The grooves associated with the 5' and 3' G-quadruplexes are labeled by dashed lines in (e). For clarity purposes, the A13•A13* non-canonical pair is shown flanked by only one of two G-tetrads in (f).

while the *c-kit2* promoter T21 sequence folds to form a dimeric all-parallel stranded form-II G-quadruplex in 100 mM KCl-containing buffer solution (Figures 7g and 8b). Further, our accumulated experimental observations are that the *c-kit2* promoter T21 and T12/T21 sequences switch over time from the form-I monomeric G-quadruplex to the form-II dimeric G-quadruplex in KCl-containing solution, establishing that the form-II dimeric G-quadruplex is the more thermodynamically stable conformer in solution (Figures S3 and S4).

Gel shift confirmation of oligomerization state of form-I and form-II G-quadruplexes

The oligomerization states of the *c-kit2* promoter as a function of K^+ cation concentration has been confirmed from gel shift studies. The 18-nt J192T1 sequence which forms a monomeric G-quadruplex (Phan, A.T. unpublished data) was used as a monomer control while the 16-nt 93del (43) and the 16-nt J19 (Phan, A.T. unpublished data) sequences, which form dimeric G-quadruplexes, were used as dimer controls in the gel shift studies (Figure 1e).

A sample of *c-kit2* T21/T12 promoter prepared in 20 mM KCl-containing solution exhibited an imino proton NMR spectrum consisting of a monomeric form-I G-quadruplex. This was also reflected in the gel shift pattern, which consists predominantly of the monomer G-quadruplex band (lane labeled form-I, Figure 1e). By contrast, a sample of *c-kit2* T21 promoter prepared in 100 mM KCl-containing solution exhibited imino NMR spectra consistent with formation of a dimeric form-II G-quadruplex. This is also reflected in the gel shift pattern, which consists predominantly of the dimeric G-quadruplex band (lane labeled form-II, Figure 1e).

All-parallel-stranded intramolecular G-quadruplexes

All parallel-stranded intramolecular G-quadruplexes were first reported in solution from NMR studies of the d(GGA)₂G sequence, which formed a dimeric G-quadruplex containing a stacked G•G•G•G tetrad and A•(G•G•G•G)•A hexad (44). NMR studies have also investigated the d(GGA)₄ sequence, which forms an all-parallel-stranded G-quadruplex formed by stacked

tetrads and heptads (45–47). Soon after, an all-parallel-stranded intramolecular G-quadruplex was reported for the crystal structure of the four-repeat TTAGGG human telomere from crystals grown in K^+ -containing solution (39).

The next demonstration of all-parallel stranded intramolecular G-quadruplex formation emerged from studies of the folding topology of the nuclease hypersensitivity element III of the human *c-myc* promoter containing six-guanine tracts (48,49). NMR-based studies of sequences containing combinations of four- of the six-guanine tracts established that the *c-myc-2345* (containing guanine tracts 2, 3, 4 and 5) (50,51) and *c-myc-1245* (50) both formed intramolecular all-parallel-stranded G-quadruplexes in K^+ -containing solution. Notably, two of the three double-chain-reversal loops in both G-quadruplexes involved a single nucleotide bridging three G-tetrad planes.

Elements of an all-parallel-stranded G-quadruplex topology containing three double-chain-reversal loops were also observed for the NMR-based structures of the five-guanine tract *c-myc-23456* sequence (52) and the four-guanine tract *c-kit1* sequence (24), both in K^+ -containing solution. Both structures adopt unique G-quadruplex folding topologies where one of the columns is restricted to a G-G rather than a G-G-G step, and both contain snap-back ($3'$ -G for *c-myc-23456* and $3'$ -G-G for *c-kit1*) elements completing G-tetrad formation (24,52).

CD studies have verified the structural conclusions of all parallel-stranded topologies for form-I monomeric and form-II dimeric G-quadruplexes adopted by the *c-kit2* promoter as a function of K^+ concentration. The CD spectrum for the form-I monomeric (*c-kit2* T12/T21 promoter in 20 mM KCl, green curve, Figure S14) and form-II dimeric (*c-kit2* T21 promoter in 100 mM KCl, red curve, Supplementary Figure S14) exhibit a negative peak at 240 nm and a positive peak at 260 nm, characteristic of parallel-stranded G-quadruplex formation (53).

Cross-talk between loops in monomeric form-I G-quadruplex

The *c-kit2* T12/T21 promoter monomeric form-I G-quadruplex adopts a standard all-parallel-stranded G-quadruplex topology involving three stacked G-tetrads, four G-G-G columns, and three double-chain reversal loops (Figure 3d). The single bases in both the first (C5) and third (A17) double-chain-reversal loops are well defined (stereo pairs of refined structures in Figure S7a and c), as is the five-nucleotide C9-G10-C11-T12-A13 linker associated with the second double-chain reversal loop (Figure 4e and stereo pair of refined structures in Figure S7b). This could reflect in part the unprecedented stacking between C5 of the first double-chain-reversal loop with C9 of the second double-chain-reversal loop (Figure 4d and stereo pair of refined structures in Figure S7a), an unique feature of cross-talk between adjacent loop residues reported here for the first time.

Novelty of the all parallel-stranded dimeric form-II G-quadruplex

There is no precedence for the topology of the *c-kit2* promoter dimeric form-II G-quadruplex presented in this paper, which is distinctly different from all other dimeric G-quadruplexes reported in the literature (43–46,54–56). The two monomeric units in the *c-kit2* form-II G-quadruplex adopt a parallel orientation, namely tail-to-head alignment (Figure 7g), in sharp contrast to all previously reported dimeric parallel-stranded G-quadruplexes, where the orientation is anti-parallel, namely head-to-head, with respect to the alignment of the two monomeric units (Figure S15a–e).

The novelty of this structure resides in its in-register self-assembly of two $5'$ -guanine tracts (G2-G4 and G6-G8) and two $3'$ -guanine tracts (G14-G16 and G18-G20) to form an all-parallel-stranded dimeric G-quadruplex scaffold (Figures 7g and 8b). The $5'$ and $3'$ G-quadruplexes sandwich a *trans* A13•A13 non-canonical pair (Figure 8e and f), with the five-residue C9-G10-C11-G12-A13 linker segment adopting an inverted S-shaped fold, stabilized by stacking interactions (Figure 8d and stereo pair in Figure S12a and b).

Inverted S-shaped fold of 5-nt linker segment in the dimeric form-II G-quadruplex

The inverted S-shaped fold of the C9-G10-C11-G12-A13 linker segment results in base-sugar stacking interactions as reflected in upfield shifts of the sugar H1' protons of C9 (5.20 ppm), G10 (5.47 ppm), C11 (5.19 ppm), G12 (5.46 ppm) and A13 (5.73 ppm) within the *c-kit2* promoter dimeric form-II G-quadruplex (Table S2 and Figure 6a). Several of these sugar H1' protons are stacked over guanine rings of the linker segment and/or terminal G-tetrads (sugar H1' protons are colored as blue balls in the stereo pair in Figure S11), thereby accounting for their upfield shifts. By contrast, such upfield shifts were not observed for the corresponding sugar H1' protons of the *c-kit2* promoter monomeric form-I G-quadruplex (Table S1 and Figure 2f).

A pair of phosphates (C9 and C11) are positioned in close proximity within the 5-nt linker segment (Figure 8d) connecting the $5'$ and $3'$ G-quadruplexes of the *c-kit2* promoter dimeric form-II G-quadruplex. It is conceivable that monovalent cations bridge these phosphates under higher KCl (100 mM)-containing buffer conditions and stabilize the inverted S-shaped fold of the C9-G10-C11-G12-A13 linker segment.

The 5-nt linker segment and A13•A13 non-canonical pair between the $5'$ - and $3'$ G-quadruplexes is likely to constitute a dynamic interface in the dimeric form-II G-quadruplex. This is reflected in broadening of aromatic proton-sugar H1' proton cross-peaks for residues C9-G10-C11-G12 (Figures 6a and S10), as well as in the absence of protection for protons of the G4•G8•G4•G8 and G14•G18•G14•G18 junctional G-tetrads during proton for deuterium exchange (Figure 5d).

Impact of flanking sequences on *c-kit2* dimeric form-II G-quadruplex formation

We have monitored the impact of addition of flanking sequences to either end of the *c-kit2* promoter sequence on dimeric form-II G-quadruplex formation. The imino proton spectra of the CC(*c-kit2* T21)AG and TT(*c-kit2* T21)TT sequences in 100 mM KCl-containing solution, where 2-nucleosides were added to the 5'- and 3'-ends, are plotted in Figure S16a and S16b, respectively). Each sequence forms a dimer under non-denaturing polyacrylamide gel electrophoresis conditions (panel 2 for CC(*c-kit2* T21)AG and panel 3 for TT(*c-kit2* T21)TT, Figure S16c). This implies that the addition of flanking sequences appear not to impact on *c-kit2* dimeric form-II G-quadruplex formation.

Implications for G-quadruplex-based synapsis and recombination

Genome-wide analysis of recombination prone regions (57) and a fine-scale map of recombination rates and hot spots across the human genome (58), have shown that they can occur within guanine-rich segments of DNA sequence space. We postulate that all-parallel-stranded monomeric G-quadruplexes embedded within duplex segments (Figure S17a) can undergo synapsis, following strand exchange, to form the all-parallel-stranded dimeric G-quadruplex shown in Figure S17b. The folding topologies of the monomeric and dimeric G-quadruplexes shown in Figure S17a and b are those reported in this paper for the monomeric form-I (Figure 3d) and dimeric form-II (Figure 7g) G-quadruplex folds of the *c-kit2* guanine-rich sequence. Following further strand exchange, the monomeric G-quadruplexes could be regenerated in a recombined manner (Figure S17c). Our postulate implies that two strands running in a parallel direction could undergo recombination after cleavage (designated by crosses at A•A non-canonical base pair, Figure S17b), rotating by 180° and subsequent rejoining mediated by an as-yet undefined nuclease-topoisomerase enzyme complex.

Such strand exchange processes as shown in Figure S17 build on an earlier postulated model of intergenic recombination involving parallel-stranded DNA alignments (59) and self-recognition of guanine-rich motifs in meiotic prophase involving parallel-G-quadruplexes (60).

Comparison with a recent structure of the *c-kit2* promoter G-quadruplex

During the writing of this paper, a related manuscript on the solution structure of the *c-kit2* promoter G-quadruplex was reported in the literature (61). This study also undertook NMR studies on the *c-kit2* T21 promoter sequence in K⁺-containing buffer (70 mM KCl, 20 mM potassium phosphate, pH 7), on a sample exhibiting imino proton NMR spectra characteristic of a mixture of at least two conformers, with the major conformer under study exhibiting an imino proton NMR spectral pattern (namely, imino protons of G2, G14, G6 and G7 between 11.5 and 12.0 ppm) similar to our *c-kit2*

T21 promoter form-II spectrum in 100 mM K⁺-containing buffer shown in Figures 1d and 5a. The authors of this paper interpreted their NMR data to conclude that the major conformer in their spectral mixture of conformers adopts a monomeric all-parallel-stranded G-quadruplex (61).

This interpretation (61) is at variance with our conclusion that *c-kit2* T21 promoter form-II adopts a dimeric all-parallel-stranded G-quadruplex (Figure 7g). Our study of the *c-kit2* T21 promoter form-II G-quadruplex has the advantage that the NMR spectrum in Figure 5a correspond to a single species and the analysis is not complicated by cross-peaks from other species. Further, we have identified a monomeric all-parallel-stranded G-quadruplex (Figure 3d), but this *c-kit2* T12/T21 promoter form-I G-quadruplex exhibits a very different imino proton NMR spectrum (recorded in 20 mM KCl-containing buffer, Figure 2a), from that of the *c-kit2* promoter T21 form-II G-quadruplex (recorded in 100 mM KCl-containing buffer, Figure 5a). Our study of the *c-kit2* promoter T12/T21 form-I G-quadruplex also has the advantage that the NMR spectrum in Figure 2a corresponds to a single species, thus facilitating assignment and structure determination.

We remain confident about our conclusions given that the NMR studies reported in our paper were intentionally undertaken under single conformer conditions for both *c-kit2* promoter monomeric form-I and dimeric form-II G-quadruplexes, thereby avoiding potential complications over data interpretation from minor conformations.

ACCESSION NUMBER

The coordinates of the *c-kit2* T12/T21 promoter form-I (accession code: 2KYP) and *c-kit2* T21 promoter form-II (accession code: 2KYO) G-quadruplexes have been deposited in the protein data bank.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We thank Ngoc Quang Do from ATP's laboratory for the gel shift data in Figure 1e and CD spectra in Figure S14.

FUNDING

National Institutes of Health Grant GM34504 (to D.J.P.); the NY Structural Biology Center supported by NIH grant GM66354 (to D.J.P.); and the Singapore Ministry of Education (grants ARC30/07 and RG62/07 to A.T.P.). Funding for open access charge: National Institutes of Health Grant GM34504 (to D.J.P.).

Conflict of interest statement. None declared.

REFERENCES

- Davis, G.T. (2004) G-Quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Edn.*, **43**, 668–698.
- Burge, S.E., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
- Patel, D.J., Phan, A.T. and Kuryavyi, V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse high order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.*, **35**, 7429–7455.
- Neidle, S. (2009) The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.*, **19**, 239–250.
- Balasubramanian, S. and Neidle, S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.*, **13**, 345–353.
- Cahoon, L.A. and Seifert, H.S. (2009) An alternative DNA structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. *Science*, **325**, 764–767.
- Maizels, N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.*, **13**, 1055–1059.
- Lipps, H.J. and Rhodes, D. (2009) G-quadruplex structures: *in vivo* evidence and function. *Trends in Cell Biol.*, **19**, 412–422.
- Todd, A.K., Johnston, M. and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
- Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
- Rawal, P., Kummarasetti, V.B., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S.K. and Chowdhury, S. (2006) Genome wide prediction of G4 DNA as regulatory motifs: role in *E. coli* global regulation. *Genome Res.*, **16**, 644–655.
- Eddy, J. and Maizels, N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.*, **34**, 3887–3896.
- Eddy, J. and Maizels, N. (2009) Selection for the G4 DNA motif at the 5'-end of human genes. *Mol. Carcinogenesis*, **48**, 319–325.
- Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the genome. *Nucleic Acids Res.*, **35**, 406–413.
- Dexheimer, T.S., Fry, M. and Hurley, L.H. (2006) DNA quadruplexes and gene regulation. In Neidle, S. and Balasubramanian, S. (eds), *Quadruplex Nucleic Acids*. RSC Publishing, Cambridge, UK, pp. 180–207.
- Du, Z., Zhao, Y. and Li, N. (2009) Genome-wide colonization of gene regulatory elements by G4 DNA motifs. *Nucleic Acids Res.*, **37**, 6784–6798.
- Hoeben, A., Schoffski, P. and Debiec-Rychter, M. (2008) Clinical implications of mutational analysis in gastrointestinal stromal tumors. *Br. J. Cancer*, **98**, 684–688.
- Tuveson, D.A., Willis, N.A., Jacks, T., Griffin, J.D., Singer, S., Fletcher, C.D., Fletcher, J.A. and Demetri, G.D. (2001) STI571 inactivation of the gastrointestinal stromal tumor *c-KIT* oncoprotein: biological and clinical implications. *Oncogene*, **20**, 5048–5058.
- Schittenhelm, M.M., Shiraga, S., Schroeder, A., Corbin, A.S., Griffith, D., Lee, F.Y., Bokemeyer, C., Deininger, M.W., Druker, B.J. and Heinrich, M.C. (2008) Dasatinib (BMS-354825), a dual SRC/ABL kinase inhibitor, inhibits the kinase activity of wild-type, juxtamembrane and activation loop mutant *KIT* isoforms associated with human malignancies. *Cancer Res.*, **66**, 473–481.
- Rankin, S., Reszka, A.P., Huppert, J., Zloh, M., Parkinson, G.N., Todd, A.K., Ladame, S., Balasubramanian, S. and Neidle, S. (2005) Putative DNA quadruplex formation within the human *c-kit* oncogene. *J. Am. Chem. Soc.*, **127**, 10584–10589.
- Fernando, H., Reszka, A.P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A.R., Neidle, S. and Balasubramanian, S. (2006) A conserved quadruplex motif located in a transcription activation site of the human *c-kit* oncogene. *Biochemistry*, **45**, 7854–7860.
- Bejugam, M., Sewitz, S., Shrude, P.S., Rodriguez, R., Shahid, R. and Balasubramanian, S. (2007) Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: Small molecule regulation of *c-kit* oncogene expression. *J. Am. Chem. Soc.*, **129**, 12926–12927.
- Gunaratnam, M., Swank, S., Haider, S.M., Galesa, K., Reszka, A.P., Beltran, M., Cuenca, F., Fletcher, J.A. and Neidle, S. (2009) Targeting human gastrointestinal stromal tumor cells with a quadruplex-binding small molecule. *J. Med. Chem.*, **52**, 3774–2783.
- Phan, A.T., Kuryavyi, V., Burge, S., Neidle, S. and Patel, D.J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human *c-kit* promoter. *J. Am. Chem. Soc.*, **129**, 4386–4392.
- Todd, A.K., Haider, S.M., Parkinson, G.N. and Neidle, S. (2007) Sequence occurrence and structural uniqueness of a G-quadruplex in the human *c-kit* promoter. *Nucleic Acids Res.*, **35**, 5799–5808.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Rippe, K. (1991) Versuche zur Strukturaufklärung und physiologischen Bedeutung von parallelsträngiger DNA. *Ph.D. Thesis*, Georg-August Universität Göttingen.
- Phan, A.T. (2000) Long-range imino proton-¹³C J-couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled DNA. *J. Biomol. NMR*, **16**, 175–178.
- Phan, A.T. and Patel, D.J. (2002) A site-specific low-enrichment ¹⁵N, ¹³C isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids. *J. Am. Chem. Soc.*, **124**, 1160–1161.
- Brünger, A.T. (1992) *X-PLOR. A System for X-ray Crystallography and NMR*. Yale University Press, New Haven.
- Schwieters, C.D., Kuszewski, J.J., Tjandra, N. and Clore, G.M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, **160**, 65–73.
- Kuryavyi, V. and Patel, D.J. (2010) Solution structure of a unique G-quadruplex scaffold adopted by a guanosine-rich human intronic sequence. *Structure*, **18**, 73–82.
- Gelbin, A., Schneider, B., Clowney, L., Hsieh, S.-H., Olson, W.K. and Berman, H.M. (1996) Geometric parameters in nucleic acids: sugar and phosphate constituents. *J. Am. Chem. Soc.*, **118**, 519–528.
- Clowney, L., Jain, S.C., Srinivasan, A.R., Westbrook, J., Olson, W.K. and Berman, H.M. (1996) Geometric parameters in nucleic acids: nitrogenous bases. *J. Am. Chem. Soc.*, **118**, 509–518.
- Stein, E.G., Rice, L.M. and Brünger, A.T. (1997) Torsion angle molecular dynamics: a new, efficient tool for NMR structure calculation. *J. Magn. Reson.*, **124**, 154–164.
- Nilges, M., Habazettl, J., Brünger, A.T. and Holak, T.A. (1991) Relaxation matrix refinement of the solution structure of squash trypsin inhibitor. *J. Mol. Biol.*, **21**, 499–510.
- Patel, D.J., Kozłowski, S.A., Nordheim, A. and Rich, A. (1982) Right-handed and left-handed DNA: Studies of B-DNA and Z-DNA by using proton nuclear Overhauser effect and phosphorus NMR. *Proc. Natl Acad. Sci. USA*, **79**, 1413–1417.
- Wang, Y. and Patel, D.J. (1994) Solution structure of the *Tetrahymena* telomeric repeat d(T₂G₄)₄ G-quadruplex. *Structure*, **2**, 1141–1156.
- Parkinson, G.N., Lee, M.P. and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
- Luu, K.N., Phan, A.T., Kuryavyi, V., Lacroix, L. and Patel, D.J. (2006) Structure of the human telomere in K⁺ solution: an intramolecular (3+1) G-quadruplex scaffold. *J. Am. Chem. Soc.*, **128**, 9963–9970.
- Xu, Y., Noguchi, Y. and Sugiyama, H. (2006) The new models of the human telomere d[AGGG(TTAGGG)]₃ in K⁺ solution. *Biorg. Med. Chem.*, **14**, 5584–5591.
- Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A. and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.*, **34**, 2723–2725.
- Phan, A.T., Kuryavyi, V., Ma, J.B., Faure, A., Andreola, M.L. and Patel, D.J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. *Proc. Natl Acad. Sci. USA*, **102**, 634–639.
- Kettani, A., Gorin, A., Majumdar, A., Hermann, T., Skripkin, E., Zhao, A., Jones, R. and Patel, D.J. (2000) A dimeric DNA interface

- stabilized by stacked A•(G•G•G•G)•A hexads and coordinated monovalent cations. *J. Mol. Biol.*, **297**, 627–644.
45. Matsugami, A., Ouhashi, K., Kanagawa, M., Liu, H., Kanagawa, S., Uesugi, S. and Katahira, M. (2001) An intramolecular quadruplex of (GGA)₄ triplet repeat DNA with a G•G•G•G tetrad and a G(A)•G(A)•G(A)•G heptad, and its dimeric interaction. *J. Mol. Biol.*, **313**, 255–269.
 46. Matsugami, A., Okizumi, T., Uesugi, S. and Katahira, M. (2003) Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. *J. Biol. Chem.*, **278**, 28147–28153.
 47. Palumbo, S.L., Memmott, R.M., Uribe, D.J., Krotova-Khan, Y., Hurley, L.H. and Ebbinghaus, S.W. (2008) A novel G-quadruplex-forming GGA repeat region in the c-myc promoter is a critical regulator of promoter activity. *Nucleic Acids Res.*, **36**, 1755–1769.
 48. Simonsson, T., Pecinka, P. and Kubitsa, M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.*, **26**, 1167–1172.
 49. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl Acad. Sci. USA*, **99**, 11593–11598.
 50. Phan, A.T., Mody, Y.S. and Patel, D.J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.*, **126**, 8710–8716.
 51. Ambrus, A., Chen, D., Dai, J., Jones, R.A. and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048–2058.
 52. Phan, A.T., Kuryavyi, V., Gaw, H.W. and Patel, D.J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol.*, **1**, 163–173.
 53. Kypr, J., Kejnovska, I., Rencuk, D. and Vorlickova, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.
 54. Kuryavyi, V., Kettani, A., Wang, W., Jones, R. and Patel, D.J. (2000) A diamond-shaped zipper-like DNA architecture containing triads sandwiched between mismatches and tetrads. *J. Mol. Biol.*, **295**, 455–469.
 55. Zhang, N., Gorin, A., Majumdar, A., Kettani, A., Chernichenko, N., Skripkin, E. and Patel, D.J. (2001) V-shaped scaffold: a new architectural motif identified in an A•(G•G•G•G) pentad-containing dimeric G-quadruplex involving stacked G(anti)•G(anti)•G(anti)•G(syn) tetrads. *J. Mol. Biol.*, **311**, 1063–1079.
 56. Nielsen, J.T., Arar, K. and Petersen, M. (2009) Solution structure of a locked nucleic acid modified quadruplex: introducing the V4 folding topology. *Angew. Chem. Int. Edn.*, **48**, 3099–3103.
 57. Mani, P., Yadav, V.K., Das, S.K. and Chowdhury, S. (2009) Genome-wide analysis of recombination prone regions predict role of DNA structural motif in recombination. *PLoS ONE*, **4**, e4399.
 58. Myers, S., Bottolo, L., Freeman, C., McVean, G. and Donnelly, P. (2005) A fine-scale map of recombination rates and hotspots across the human genome. *Science*, **310**, 321–324.
 59. Kuriavyi, V.V. (1987) Parallel DNAs (English translation: Kuryavyi, V.V. (1988) Plenum Publishing Corporation, New York 1209–1217). *Molekulyarnaya Biologiya*, **21**, 1486–1495.
 60. Sen, D. and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
 61. Hsu, S.-T.D., Varnai, P., Bugaut, A., Reszka, A.P., Neidle, S. and Balasubramanian, S. (2009) A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. *J. Am. Chem. Soc.*, **131**, 13399–13409.