Published online 27 July 2011 *Nucleic Acids Research, 2011, Vol. 39, No. 20* ***9023–9033***
doi:10.1093/nar/gkr612

c-MYC promoter G-quadruplex formed at the 5′-end of NHE III_1 element: insights into biological relevance and parallel-stranded G-quadruplex stability

Raveendra I. Mathad[1], Emmanuel Hatzakis[1], Jixun Dai[1] and Danzhou Yang[1,2,3,4,*]

[1]College of Pharmacy, The University of Arizona, 1703 East Mabel St, Tucson, AZ 85721, [2]BIO5 Institute, The University of Arizona, AZ 85721, [3]The Arizona Cancer Center, Tucson, AZ 85724 and [4]Department of Chemistry, The University of Arizona, Tucson, AZ 85721, USA

Received May 6, 2011; Revised July 8, 2011; Accepted July 12, 2011

ABSTRACT

We studied the structures and stabilities of G-quadruplexes formed in Myc1234, the region containing the four consecutive 5′ runs of guanines of c-MYC promoter NHE III_1, which have recently been shown to form in a supercoiled plasmid system in aqueous solution. We determined the NMR solution structure of the 1:2:1 parallel-stranded loop isomer, one of the two major loop isomers formed in Myc1234 in K^+ solution. This major loop isomer, although sharing the same folding structure, appears to be markedly less stable than the major loop isomer formed in the single-stranded c-MYC NHE III_1 oligonucleotide, the Myc2345 G-quadruplex. Our NMR structures indicated that the different thermostabilities of the two 1:2:1 parallel c-MYC G-quadruplexes are likely caused by the different base conformations of the single nucleotide loops. The observation of the formation of the Myc1234 G-quadruplex in the supercoiled plasmid thus points to the potential role of supercoiling in the G-quadruplex formation in promoter sequences. We also performed a systematic thermodynamic analysis of modified c-MYC NHE III_1 sequences, which provided quantitative measure of the contributions of various loop sequences to the thermostabilities of parallel-stranded G-quadruplexes. This information is important for understanding the equilibrium of promoter G-quadruplex loop isomers and for their drug targeting.

INTRODUCTION

Overexpression of the c-MYC proto-oncogene is linked to a wide variety of human cancers, including colon, breast, prostate, cervical and lung carcinomas, osteosarcomas, lymphomas and leukemias (1–9). In addition, elevated levels of c-MYC expression are often associated with poor therapeutic prognosis. c-MYC overexpression can be caused by different mechanisms, including gene amplification (10,11), translocation (12–14) and simple upregulation of transcription (1,15). The transcriptional regulation of c-MYC expression involves multiple promoters, with P1 and P2 being the predominant ones (16). A highly conserved NHE III_1 located 142–115 bp upstream from the P1 promoter has been shown to be required for 80–95% of c-MYC transcription, regardless of whether the P1 or P2 promoter is used (17,18). This NHE III_1 element can form transcriptionally active and silenced forms in the promoter (19); the formation of DNA G-quadruplex structures is critical for c-MYC transcriptional silencing (20–22), and compounds that stabilize the G-quadruplex can repress c-MYC gene expression (20,23). DNA G-quadruplexes are a family of secondary DNA structures that consist of stacked G-tetrads connected by Hoogsteen hydrogen bonds and stabilized by monovalent cations such as potassium and sodium. Intramolecular G-quadruplexes have been found in a number of G-rich regions with biological significance, such as human telomeres, oncogene promoters and 5′-UTR regions (24–26).

A special requirement for promoter sequences to form G-quadruplexes is that the DNA secondary structures must be generated in a region of duplex DNA. It has been recently shown that the transcriptionally induced negative superhelicity results in the dynamic equilibrium

*To whom correspondence should be addressed. Tel: +520 626 5969; Fax: +520 626 6988; Email: yang@pharmacy.arizona.edu

between duplex, single-stranded DNA and secondary DNA structures of the c-MYC NHE III_1 that likely controls c-MYC transcription (27). A recent report from the Levens lab at NCI (28,29) demonstrated *in vivo* that transcriptionally induced supercoiling in the c-MYC promoter is not immediately relieved by topoisomerase I and II and directs the melting of the FarUpStream Element (FUSE) 1.7 kb upstream of the P1 promoter, which binds the positive and negative regulating FBP and FIR proteins to control the rate of promoter firing through a feedback loop. The NHE III_1 element, which is the G-quadruplex forming region in the c-MYC promoter, is much closer to the source of induced negative superhelicity and thus is likely to be subjected to greater torsional stress than the FUSE. Transcriptional factors that bind to either the duplex (e.g. Sp1) or single-stranded (e.g. CNBP, hnRNP K) NHE III_1 elements cause transactivation, while the secondary DNA structures formed from the same element under negative superhelicity can silence transcription (27). NM23-H2 and nucleolin have been identified as proteins that facilitate the unwinding and folding of the G-quadruplex, respectively (30,31).

The G-rich strand of the c-MYC NHE III_1 is a 27-nt segment composed of five consecutive runs of guanines (Pu27, Figure 1A). DMS footprinting showed that the major G-quadruplex formed in the Pu27 oligonucleotide in K^+ solution is a quadruplex involving the II, III, IV, V runs of guanines, i.e. G7-G9, G11-G14, G16-G18, G20-G23, but not the first run of guanines, G2-G5 (Myc2345, Figure 1A). Mutational analysis in conjunction with a luciferase reporter system has also shown that the major G-quadruplex structure responsible for c-MYC transcriptional silencing in K^+ solution appears to involve the four consecutive 3′ runs of guanines (20,32). This structure adopts a parallel-stranded folding (32,33) and we have determined the molecular structure of the major isomer with the 1:2:1 loop size arrangement in K^+ solution using Pu22_T14/T23 (1:2:1 loop isomer, with the numbers describing the lengths of the three double chain-reversal loops) (Figure 1A) (34). Interestingly, a recent footprinting study using a supercoiled plasmid system in aqueous solution with the c-MYC NHE III_1 showed a G-quadruplex formed under superhelicity appears to involve the I, II, III, IV runs of guanines, i.e. G2–G5, G7–G9, G11–G14 and G16–G18 (Myc1234, Figure 1A) (35).

In this report, we studied the structures and stabilities of G-quadruplexes formed in Myc1234, the region containing the four consecutive 5′ runs of guanines of c-MYC NHE III_1. We found that two parallel-stranded major loop isomers, one with 1:2:1 and another with 1:1:2 loop size arrangements, formed in Myc1234 in K^+ solution with comparable stability. However, while they both contain the same total loop length, the two major loop isomers formed in Myc1234 appear to be markedly less stable than the major 1:2:1 loop isomer G-quadruplex formed in Myc2345 of NHE III_1 in K^+ solution. We determined the molecular structure of the 1:2:1 loop isomer formed in Myc1234 in K^+ solution. We also performed a systematic mutational analysis in combination with thermodynamic studies to understand the effect of loop

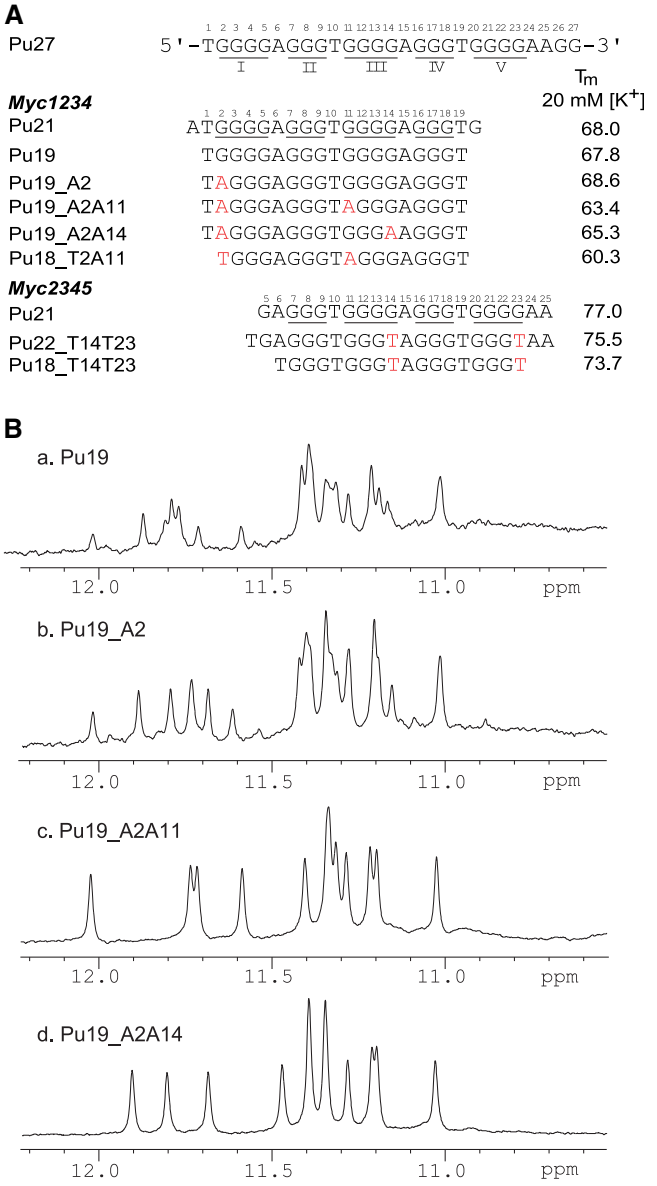

Figure 1. (**A**) The promoter sequences of the NHE III_1 element of the c-MYC gene and its modifications, and their melting temperatures in 20 mM K^+. The mutations are colored in red. Pu27 is the wild-type 27-mer G-rich sequence of the c-MYC NHE III_1. Myc1234 is the region containing the four consecutive 5′ runs of guanines of c-MYC NHE III_1. Within Myc1234, Pu21 and Pu19 are the 21-mer and 19-mer wild-type G-rich sequences of the c-MYC NHE III_1. Pu19_A2A11 is the modified sequence of Pu19, with G–A mutations at positions 2 and 11, which adopts a single parallel-stranded G-quadruplex structure with a 1:2:1 loop size arrangement in K^+ solution whose structure has been determined by NMR in this study. Myc2345 is the region containing the four consecutive 3′ runs of guanines of c-MYC NHE III_1. Within Myc2345, Pu21 is a 21-mer wild-type G-rich sequence of the c-MYC NHE III_1. Pu22_T14/T23 was the modified Myc2345 sequence used to determine the molecular structure of the major isomer with the 1:2:1 loop size arrangement in K^+ solution (34). (**B**) 1D ^{1}H-NMR profiles of Pu19 and its derived sequences. Pu19 (a) forms the same major G-quadruplexes as Pu19_A2 (b). Pu19_A2 adopts a mixture of two G-quadruplexes, which can be isolated by Pu19_A2A11 (1:2:1 loop size arrangement) (c) and P19_A2A14 (1:1:2 loop size arrangement) (d).

constitution and position on the overall stability of parallel-stranded G-quadruplexes. Parallel-stranded structural motifs with short loops are found to be prevalent in promoter sequences, and it has become evident that G-quadruplexes formed in gene promoters are often a mixture of multiple structures or loop isomers (36). While parallel-stranded G-quadruplexes appear to favor short loop lengths in K^+ solution (34,37–40), systematic information on the effect of loop sequences will be important to understand the equilibrium of promoter G-quadruplex loop isomers, which is critical for the development of small molecular compounds targeting promoter G-quadruplexes for gene regulation.

MATERIALS AND METHODS

Oligonucleotides

The DNA oligonucleotides were synthesized and purified as described previously (34,41,42). Samples in D_2O were prepared by repeated lyophilization and final dissolution in 99.96% D_2O. Samples in water were prepared in 10%/90% D_2O/H_2O solution. The final Nuclear Magnetic Resonance (NMR) samples contained 0.1–2.5 mM DNA in 25 mM K-phosphate buffer (pH 7.0) and 70 mM KCl or in the low-salt solution, 10 mM K-phosphate buffer.

Circular dichroism spectroscopic study

Circular dichroism (CD) spectroscopic study of the oligonucleotides was performed on a Jasco J-810 spectropolarimeter (Jasco Inc., Easton, MD, USA) equipped with a thermoelectrically controlled cell holder as described previously. The quartz cell of 1 mm optical path length was used. A blank sample containing only buffer was used for the baseline correction. CD spectroscopic measurements were the averages of three scans collected between 200 and 350 nm. The scanning speed of the CD instrument was 100 nm/min, and the response time was 1 s. For measuring the T_m value of an oligonucleotide, CD melting and annealing experiments were performed at 265 nm for three repeats, with a heating or cooling rate of 2°C/min, respectively.

NMR spectroscopic study

NMR experiments were performed on a Bruker DRX-600 MHz spectrometer as described previously (34,41,42). Six percent of 1,2,7-^{15}N, 2-^{13}C-labeled guanine phosphoramidite was used for site-specific labeled DNA synthesis. One-dimensional ^{15}N-filtered gradient HMQC experiments with site-specific labeling were performed to identify guanine imino and H8 protons. Homonuclear 2D NMR experiments, including NOESY, TOCSY and DQF-COSY, were collected at 5°C, 15°C and 25°C for Pu19_A2A11 sample in water and D_2O solution at pH 7, 95 mM K^+. Jump-return and watergate experiments were used to suppress the water signal in the spectra. The relaxation delays were set to 2–2.5 s in 2D experiments. The software Sparky (UCSF) was used in peak assignments and integrations. Non-exchangeable protons were estimated based on the Nuclear Overhauser Effect (NOE) cross-peak volumes at 50–300 ms mixing times, with the upper and lower boundaries assigned to ±20% of the estimated distances. The thymine base proton Me-H6 distance (2.99 Å) was used as a reference. For unresolved cross-peaks, the distance restraints were used with higher boundaries (±30%).

The structure of Pu19_A2A11 was calculated using X-PLOR (43). Metric matrix distance geometry (MMDG) and simulated annealing calculations were carried out in X-PLOR to embed and optimize 100 initial structures for the Pu19_A2A11 sequence, as described previously (41,42). The experimentally obtained distance restraints and G-tetrad hydrogen-bonding distance restraints were included during the calculation. A total of 343 distance restraints, of which 238 are from intra-residue NOEs, and 105 are from inter-residue NOEs, were incorporated into the NOE-restrained structure calculation. Hydrogen bond restraints were applied to the G-tetrads, using a quadratic energy function with a force constant of 100 kcal/mol/Å^2. A low-level planarity restraint (5 kcal/mol/Å^2) was applied on the G-tetrad in the simulated annealing step of the structure calculation. The planarity restraints were removed in the final molecular dynamics simulation with energy minimization. Dihedral angle restraints were used for the glycosidic torsion angle (χ) based on the experimentally determined 'anti' conformations. The 30 best molecules were selected based both on their minimal energy terms and number of NOE violations, and were further subjected to NOE-restrained molecular dynamics calculations at 300 K for 25 ps. The average structure from the trajectory saved during the last 2.0 ps of NOE-restrained molecular dynamics calculations was obtained and subjected to further energy minimization. The 10 best molecules were selected based both on their minimal energy terms and number of NOE violations.

RESULTS

The 1:1:2 and 1:2:1 loop isomers are the two major loop isomers formed in Myc1234

The wild-type Pu19 of Myc1234 is composed of the four consecutive 5′-runs of guanines of mycPu27 of NHE III$_1$, with four (G2–G5), three (G7–G9), four (G11–G14) and three (G16–G18) guanines, respectively (Figure 1A). Utilizing a combination of mutational analysis and NMR, we found that G2 is not involved in the major G-quadruplex formation of Myc1234. As shown in Figure 1Ba and b, the same major G-quadruplexes were found to form in Pu19 and Pu19_A2, in which G2 was mutated to adenine, as evident by the imino protons of tetrad guanines of the two sequences. Therefore, the two major loop isomers formed in Pu19 are the 1:2:1 and 1:1:2 loop isomers in which G2/G11 or G2/G14 are not involved in tetrad formation, respectively. The two loop isomers can be isolated by the Pu19_A2A11 or Pu19_A2A14 sequences, respectively, in which G2/G11 or G2/G14 are mutated to adenines (Figure 1A). Adenine was substituted for guanine because both are purines. Pu19_A2A11 and Pu19_A2A14 each appeared

to form a well-defined G-quadruplex as shown by NMR (Figure 1Bc and d). The CD spectra of Pu19_A2A11 and Pu19_A2A14 showed a positive maximum ~265 nm and a negative minimum at 240 nm (Figure 2A) and indicated the formation of a parallel-stranded G-quadruplex structure (44). Interestingly, Pu19_A2 was shown by NMR to be a mixture of the two G-quadruplex loop isomers, with the 1:1:2 loop isomer accounting for ~55% and the 1:2:1 loop isomer accounting ~45% of the total population, indicating that the 1:1:2 loop isomer is slightly more stable than the 1:2:1 loop isomer (Figure 1B). This is consistent with the melting temperatures of Pu19_A2A11 (63.4°C) and Pu19_A2A14 (65.3°C) in 20 mM K^+, respectively (Figure 1A). The melting temperatures were determined using CD melting study at 265 nm. However, these melting temperatures are markedly lower than that of the major 1:2:1 loop isomer G-quadruplex formed in Myc2345, Pu22_T14T23 (Figure 1A) (34), which has a melting temperature of 75.5°C in 20 mM K^+, ~12°C higher than the 1:2:1 G-quadruplex formed by Pu19_A2A11. To understand the molecular level differences between the two 1:2:1 parallel structures formed in Myc1234 and Myc2345, we carried out complete NMR structure determination of the G-quadruplex formed by Pu19_A2A11.

Pu19_A2A11 forms a parallel-stranded G-quadruplex with 1:2:1 loop size arrangement in K^+

The CD melting temperature of Pu19_A2A11 is concentration-independent (Supplementary Table S1), suggesting the monomeric nature of the structure. ^{1}H-NMR of Pu19_A2A11 in 95 mM K^+ solution showed a sharp and well-resolved imino region (Figure 1Bc), which appears to be excellent for NMR structure determination. The presence of 12 imino protons between 10.5 and 12 ppm from the tetrad guanines indicates a G-quadruplex with three G-tetrads. Guanine imino protons are assigned to their positions in the sequence by site-specific 6% ^{15}N-labeled oligonucleotides. Using 1D ^{15}N-filtered HMQC experiments, imino H1 and aromatic H8 protons for each site-specific-labeled guanine can be detected and assigned (Figure 3A and B). Thymines in the sequence can be identified by strong H6-Me cross-peaks in DQF-COSY/TOCSY, and their positions were determined by the sequential assignment. The complete assignment of the protons of Pu19_A2A11 was achieved by sequential assignment (Figure 4) using 2D DQF-COSY, TOCSY and NOESY at different temperatures in both H_2O and D_2O. The chemical shifts of all proton resonances are listed in Table 1.

The folding topology of the Pu19_A2A11 G-quadruplex can be determined by the NOE connectivities of imino and aromatic H8 protons. In a G-tetrad plane with a Hoogsteen-type H-bond network, the imino proton NH1 of a guanine is in close spatial vicinity to the NH1s of the two adjacent guanines, and to the base H8 of one of the adjacent guanines (Figure 2B). For example, the NOEs of G12H1/G7H1, G12H1/G16H1, G3H1/G16H1 (Figure 4A), and G3H1/G7H8, G7H1/G12H8, G12H1/G16H8 and

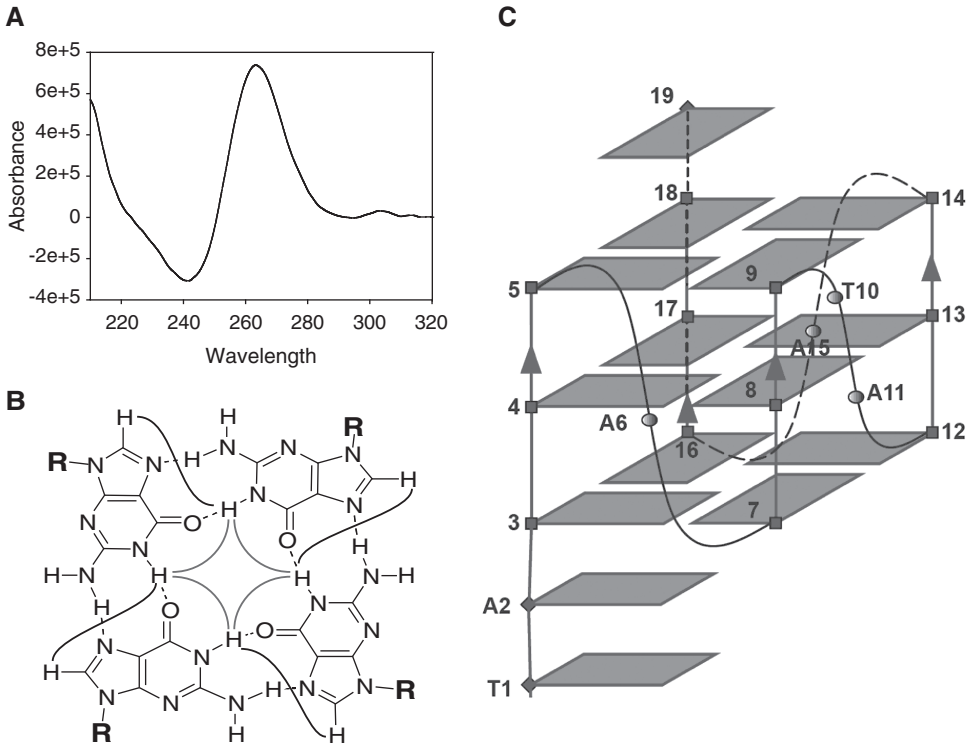

Figure 2. (**A**) A CD spectrum of Pu19_A2A11 in pH 7.0 H_2O containing 95 mM K^+ at 25°C with a DNA concentration of 10 μM. (**B**) The arrangement of guanine bases in a G-tetrad stabilized by a Hoogsteen hydrogen-bonded network. The H8-H1 and H1–H1 connectivities observable in NOESY spectra are shown. (**C**) The folding topology adopted by Pu19_A2A11 in K^+ solution.

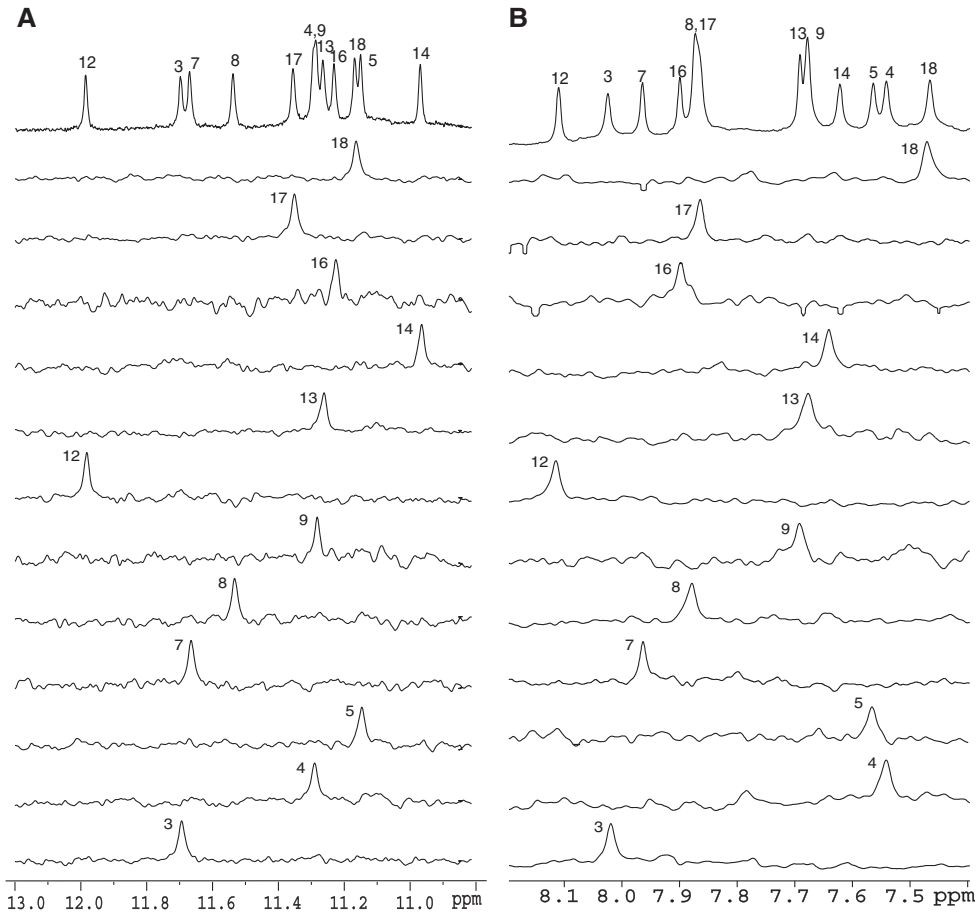

Figure 3. The assignment of (**A**) imino H1 and (**B**) aromatic H8 protons of Pu19-A2A11 as detected with ^{1}D-HMQC NMR experiments using site-specifically labeled oligonucleotides in aqueous solution containing 25 mM potassium phosphate, 70 mM potassium chloride, pH 7.0 at 25°C. The corresponding imino and aromatic protons are assigned in the reference spectrum at the top.

G16H1/G3H8 (Figure 4B) defined the tetrad plane of G3–G7–G12–G16. The other two tetrad planes, G4–G8–G13–G17 and G5–G9–G14–G18, were determined in a similar way. Based on the determination of three G-tetrad planes, Pu19_A2A11 forms a parallel-stranded intramolecular G-quadruplex of a 1:2:1 loop size arrangement with three double chain-reversal loops of A6, T10A11 and A15, respectively (Figure 2C). All the tetrad guanines appeared to adopt *anti* conformation based on the intensity of intra-residue H8-H1′ cross-peaks (Figure 4D). Along each G-strand, sequential H1–H1 connectivities of adjacent guanines were clearly observed, such as G3H1/G4H1 and G4H1/G5H1 (Figure 4A), as well as G3H8/G4H8 (Figure 4C). The sequential NOE connectivities along each G-strand are also clearly observed for G_n-H8 and G_{n-1}-H1′/H2′/H2″/H3′, typical for right-handed DNA backbone conformation (Figure 4D). Inter-tetrad NOE connectivities of non-sequential guanines, such as G3H8/G17H1 and G4H8/G18H1, G16H8/G13H1 and G17H8/G14H1, G12H8/G8H1 and G13H8/G9H1 and G7H8/G4H1 and G8H8/G5H1, were clearly observed (Figure 4B), supporting both the parallel-stranded conformation and the right-handed twist of the G-quadruplex. As observed previously (34), the H2′/H2″ protons of the tetrad guanines preceding the double chain-reversal loops were downfield shifted as they were more solvent exposed, with H2′ protons more downfield shifted than H2″ protons (Table 1).

Molecular structure of Pu19_A2A11 derived by NOE-restrained structure calculation

Many inter-residue NOEs are observed in 2D-NOESY of Pu19_A2A11 in K^+. Critical inter-residue NOE interactions of Pu19_A2A11 are schematically summarized in Figure 5, which clearly define the overall structure of the G-quadruplex structure formed by Pu19_A2A11. Solution structures of the Pu19_A2A11 G-quadruplex were calculated using a NOE-restrained distance geometry (DGSA) and molecular dynamics (RMD) approach (Figure 6, PDB ID 2lby), starting from an arbitrary extended single-stranded DNA. A total of 343 NOE distance restraints, of which 105 are from inter-residue NOE interactions, were incorporated into the NOE-restrained structure calculation (Table 2). The structure statistics are also listed in Table 2. The stereo view of the 10 lowest energy structures are shown in Figure 6A. Pu19_A2A11 forms a well-defined three G-tetrad

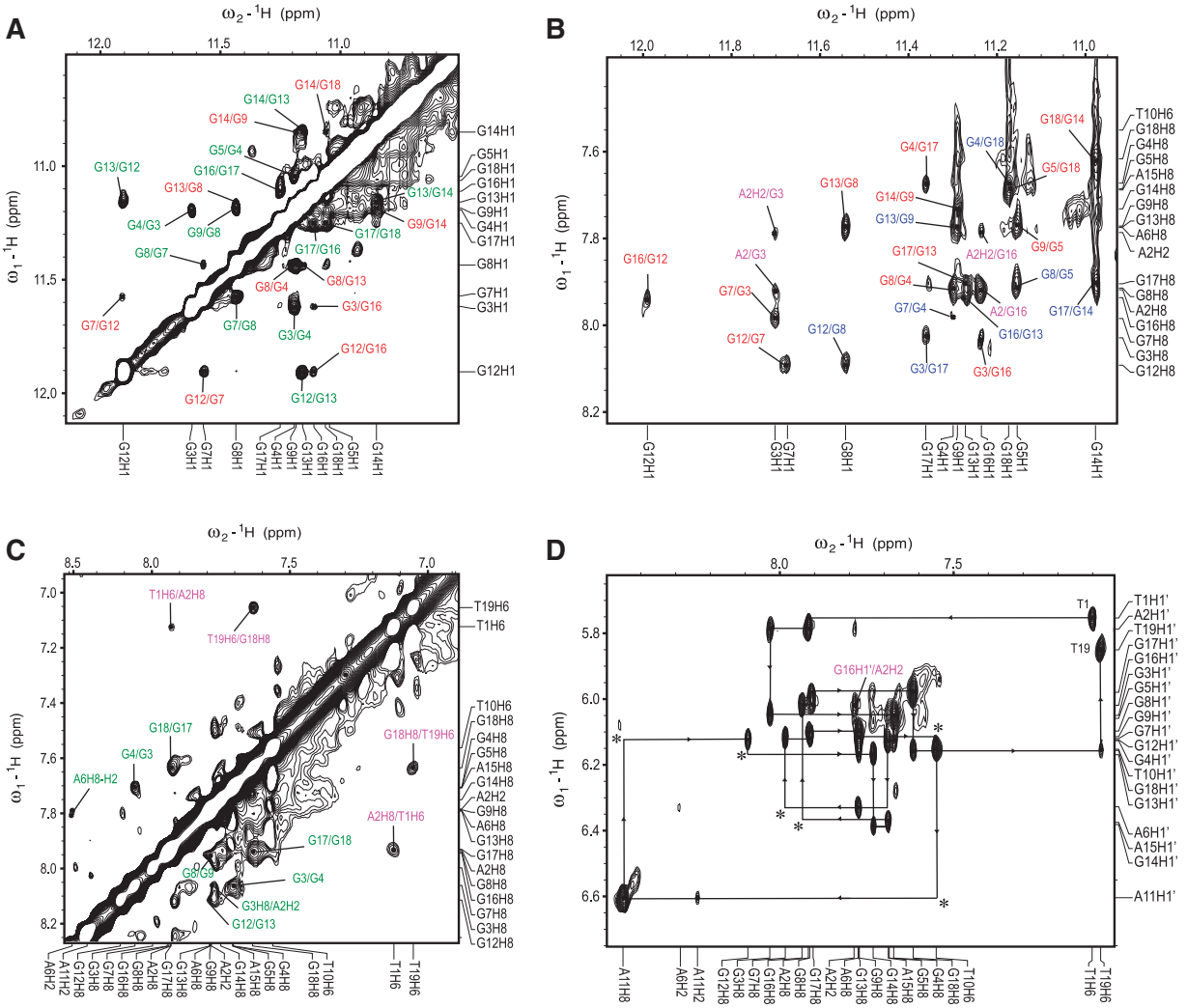

Figure 4. Expanded regions of H1–H1 (15°C) (**A**) and H1–H8 (25°C) (**B**) of the exchangeable 2D-NOESY spectra of Pu19_A2A11 with assignments. Expanded regions of H8/H8 (15°C) (**C**) and H8/H6-H1′ (25°C) (**D**) of the non-exchangeable 2D NOESY spectra of Pu19_A2A11 with assignments. The sequential assignment pathway is shown in (D), with the missing connections marked with asterisks. Intra-tetrad NOEs are in red, sequential NOEs in green, inter-tetrad NOEs in blue. NOEs of the flanking sequences are in magenta. Conditions: 25 mM potassium phosphate, 70 mM potassium chloride, pH 7.0.

parallel-stranded G-quadruplex structure with three double chain-reversal loops of A, TA and A. The three double chain-reversal loops are mostly exposed to solvent. The 5′-flanking sequence T1A2 adopts a well-defined conformation at the 5′-end, with A2 stacking very well with the 5′ tetrad (Figure 6B), as supported by NOE connectivities such as A2H8/G16H1 (strong-medium), A2H8/G3H1 (medium-weak), A2H2/G16H1 (medium-weak) and A2H2/G3H1 (medium-weak) (Figure 4B). The 3′-flanking sequence T19 appears to stack with the 3′ G-tetrad (Figure 6C), as supported by NOE connectivities such as T19Me/G14H1 and G18H8/T19Me (data not shown) and G18H8/T19H6 (Figure 4C). Regular right-handed backbone conformations were observed at both ends, as supported by sequential NOE connectivities at T1A2, A2G3 and G18T19 steps, such as G_n-H8/H6 and G_{n-1}-H1′/H2′/H2″/H3′.

Both Pu19_A2A11 (Myc1234) and Pu22_T14T23 (Myc2345) (Figure 1A) form parallel-stranded G-quadruplexes with a 1:2:1 loop arrangement. While different capping conformations were formed by different flanking sequences, the overall G-tetrad core structures of the two parallel G-quadruplexes are quite similar. However, a noticeable difference was found in the conformation of the 1-nt double chain-reversal loops of the two parallel G-quadruplexes. The structure formed by Pu19_A2A11 of Myc1234 contains two 1-nt double chain-reversal loops of adenine (A6 and A15), whereas the structure formed by Pu22_T14T23 of Myc2345 contains two 1-nt double chain-reversal loops of thymine (T10 and T19). Both the single nucleotide A loops of Myc1234 adopted a conformation in which the adenine base points away from the G-quadruplex groove almost perpendicularly (wing-up) (Figure 7A and B); whereas the two single nucleotide T loops of

Table 1. Proton chemical shifts for Pu19_A2A11 at 25°C[a]

	H6/H8	H2/Me H1/Me	H1′	H2′/H2″	H3′	H4′	H5′/H5″
T1	7.10	1.52	5.75	1.65/2.07	4.11	3.84	3.41
A2	7.91	7.78	5.78	2.35/2.37	4.40	4.04	3.83/4.02
G3	8.03	11.7	6.04	2.76/2.99	4.95	4.25	4.02
G4	7.67	11.3	6.12	2.55/2.86	4.94	4.49	4.26
G5	7.68	11.15	6.05	2.88/2.59	4.94	4.41	4.25
A6	7.77	8.28	6.32	2.58	4.93	4.40	3.64
G7	7.98	11.68	6.12	2.45/2.89	5.11	4.43	4.23/4.32
G8	7.09	11.54	6.09	2.36/2.62	5.00	4.03	3.60/3.82
G9	7.76	11.29	6.11	2.79/2.62	4.98	4.41	4.12/4.19
T10	7.55	1.84	6.15	2.12/2.37	4.62	3.79	3.64
A11	8.45	8.23	6.60	2.84/3.01	5.15	4.65	4.25
G12	8.09	11.99	6.12	2.60/2.92	4.97	4.39	4.08/4.19
G13	7.77	11.27	6.17	2.61/2.89	4.96	4.48	4.34
G14	7.73	10.97	6.38	2.60/2.50	5.04	4.33	4.24
A15	7.69	8.43	6.37	2.49/2.61	5.04	4.58	4.24
G16	7.93	11.23	6.01	2.34/2.77	5.03	4.42	4.22/4.31
G17	7.90	11.36	5.97	2.65	5.01	4.46	4.20
G18	7.61	11.17	6.15	2.52/2.71	4.82	4.45	4.17
T19	7.07	1.42	5.85	2.01	4.36	4.16	3.89/3.99

[a]The chemical shifts are measured in 95 mM K-phosphate, pH 7.0 at 25°C and referenced to DSS.

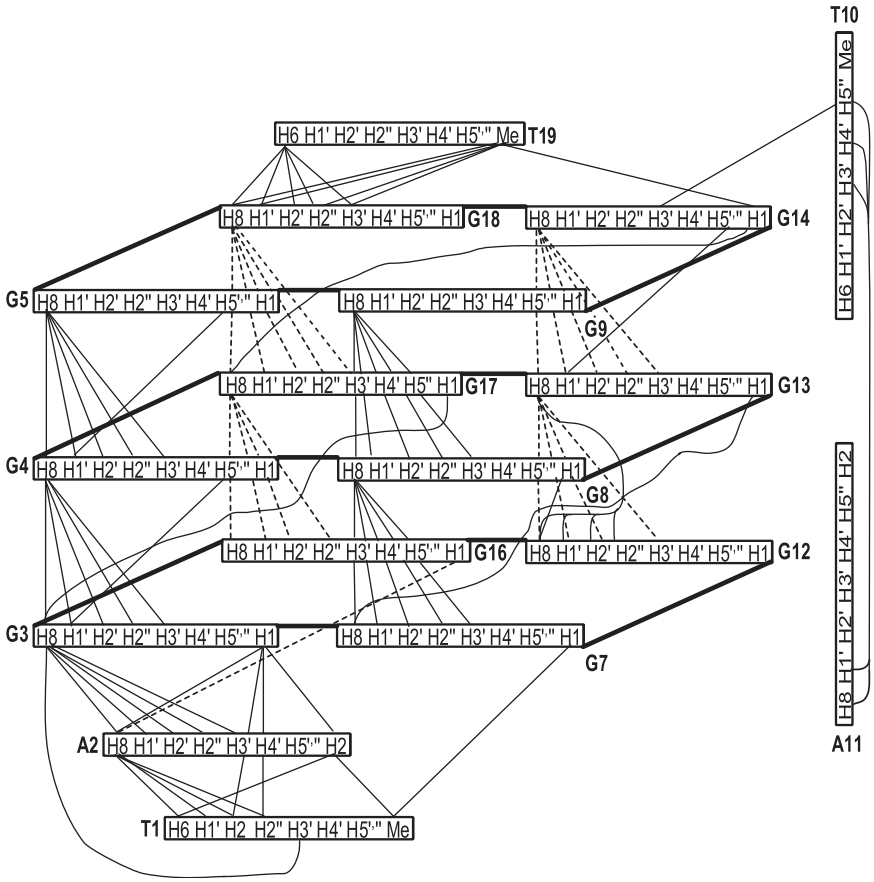

Figure 5. Schematic diagram showing NOE connectivities observed for Pu19_A2A11 G-quadruplex in the presence of K^+. Distance restraints are used for the structure calculations.

Myc2345 both adopted a conformation with the thymine base lying along the groove (wing-down) (Figure 7C and D) (34). The 1-nt adenine loop is more solvent exposed and thus less energetically favored. This difference in the single nucleotide loop conformation may explain the difference in thermodynamic stability of the two parallel-stranded 1:2:1 c-MYC G-quadruplexes, Pu19_A2A11 (Myc1234, T_m 63.4°C in 20 mM K^+) and Pu22_T14T23 (Myc2345, T_m 75.5°C in 20 mM K^+).

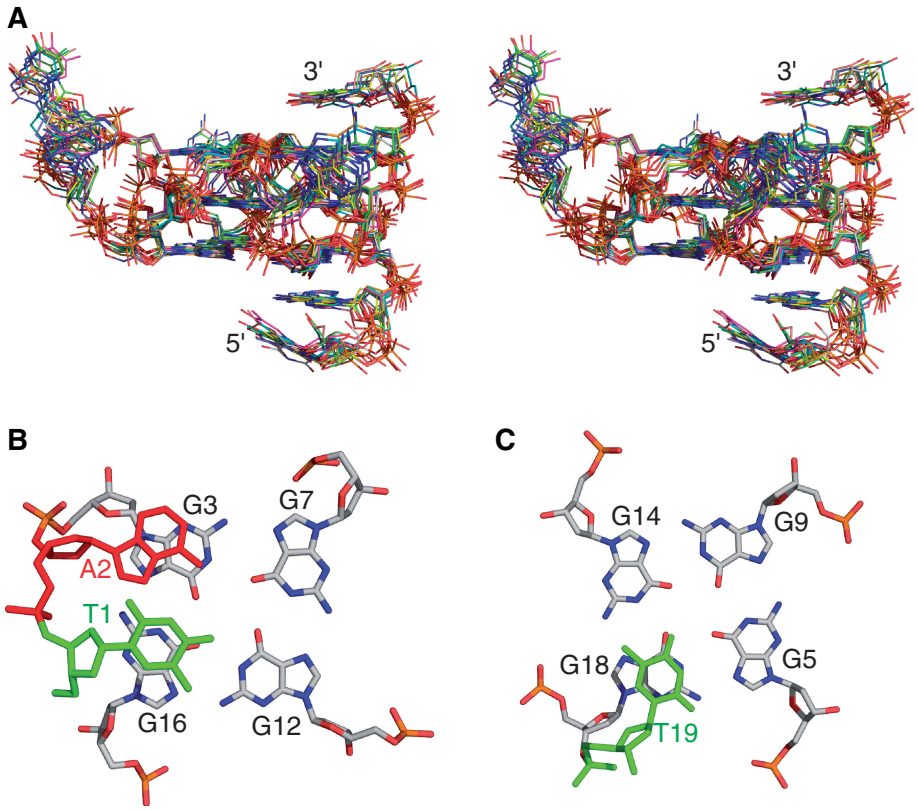

Figure 6. (**A**) Stereo view of the ensemble of the superimposed 10 lowest energy structures of Pu19_A2A11 obtained by NMR refinements. (**B**) The stacking conformation of the T1 and A2 at the 5′-end. (**C**) The capping structure of the T19 at the 3′-end stabilizing the G-tetrad core. The flanking thymines are shown in green, while adenine is in magenta.

Systematic mutational analysis in combination with CD melting studies of c-MYC G-quadruplexes with different loop sequences

While it is suggested that the more solvent-protruding single nucleotide adenine loop versus the single nucleotide thymine loop may be responsible for the reduced thermostability of Myc1234, we wanted to rule out the contribution of different flanking sequences. Therefore, we systematically designed a series of mutated c-MYC sequences with the same flanking sequences (Table 3) for thermodynamic analysis. CD melting experiments at 265 nm were used to determine the melting temperature (T_m) in pH 7 lithium phosphate buffer with 20 mM K^+ (40). All DNA oligonucleotides appeared to be completely folded at 25°C in 20 mM K^+, as indicated by their ^{1}H NMR and CD profiles, which were the same as those in 100 mM K^+.

We first examined modified Pu19_A2 sequences with various loop sequences (Table 3, top panel). Pu19_A2A11 forms a 1:2:1 parallel G-quadruplex with three loops of A, TA and A, and shows a T_m of 63.4°C. Pu19a forms the same 1:2:1 structure with a single-A loop replaced by a single-T loop, and shows a T_m of 70.5°C, indicating the ΔT_m between the single-T and single-A loop is ~7°C in 20 mM K^+. The loop positions do not appear to affect the thermostability, as Pu19b, which has the same loop sequences as Pu19_A2A11 but a different loop arrangement (1:1:2), shows a very similar T_m (62.9°C) to Pu19_A2A11; and Pu19c, which has the same loop sequences as Pu19a but different loop arrangement, also shows a very similar T_m (70.0°C) to Pu19a. Pu19_A2A14 forms the same 1:1:2 structure as Pu19c with the TA loop replaced by an AA loop, and shows a T_m of 65.3°C, indicating the ΔT_m between the TA and AA loops is ~4.6°C in 20 mM K^+.

We extended our analysis to compare with the G-quadruplex formed by Pu22 of Myc2345 (Figure 1A). To standardize the analysis, here we used a single flanking T at both the 5′- and 3′-ends for all the sequences (Table 3, middle and bottom panels). The ΔT_m between the 1:2:1 Myc1234 quadruplex formed by Pu18_T2A11 (T_m 60.3°C) and the 1:2:1 Myc2345 quadruplex formed by Pu18_T14T23 (T_m 73.7°C) is ~13.4°C in 20 mM K^+, which is quite significant. The ΔT_m of 13.4°C appears to be the summation of two single-T–single-A loop substitutions, which have been shown to each account for ~7°C ΔT_m. Again, the loop positions do not appear to affect the thermostability, as Pu18a (Table 3), which has the same loop sequences as Pu18_T14T23 but a different loop arrangement (2:1:1), shows a very similar T_m (73.4°C). The ΔT_m of ~7°C for a single-T–single-A loop substitution was also confirmed by the difference of T_ms between Pu18a and Pu18b, and between Pu18c and Pu18d (Table 3). We further examined parallel G-quadruplexes

with a single 1-nt and two 2-nt loops, and also found that the ΔT_m is ~7°C for a single-T–single-A loop substitution (between Pu19d and Pu19e, and between Pu19f and Pu19h) (Table 3). ΔT_m appears to be 3.5–4.5°C for a TT–TA loop substitution (between Pu18b and Pu18c, Pu19e and Pu19f, and Pu19d and Pu19h) and 8°C for a TT–AA loop substitution (between Pu18d and Pu18e), which may be considered to be the summation of TT–TA and TA–AA loop substitutions.

Table 2. Structural statistics for the Pu19_A2A11 G-quadruplex[a]

NMR distance and dihedral constraints	
Distance restraints	
Total NOE	352
Intra-residue	240
Inter-residue	112
Sequential ($\|i-j\| = 1$)	78
Non-sequential ($\|i-j\| > 1$)	34
Hydrogen bonds	50
Total dihedral angle restraints	28
Structure statistics	
NOE Violations (>0.2 Å)	1.7 ± 0.9
Deviations from idealized geometry	
Bond length (Å)	0.072 ± 0.001
Bond angle (°)	1.38 ± 0.01
Impropers (°)	0.98 ± 0.002
Average pairwise RMSD of heavy atoms (Å)	
G-tetrads	1.09 ± 0.10
With T1, A2	1.15 ± 0.10
With A6, A15	1.16 ± 0.13
With T10, A11	1.18 ± 0.10
With T19	1.15 ± 0.09
All residues	1.32 ± 0.11

[a]The ensemble of 10 structures is selected based both on the minimal energy terms and number of NOE violations.

DISCUSSION

While the major G-quadruplexes formed in Myc1234, the region containing the four consecutive 5′ runs of guanines of c-MYC NHE III_1, and Myc2345, the region containing the four consecutive 3′ runs of guanines of c-MYC NHE III_1, are both parallel-stranded G-quadruplexes with a total loop length of 4 (two 1-nt loops and one 2-nt loop), their thermostability is markedly different. The major G-quadruplex formed in Myc2345 (1:2:1) has a T_m ~14°C higher than that of the 1:2:1 quadruplex, one of the major loop isomers, formed in Myc1234. Our NMR-derived molecular structure of the 1:2:1 loop isomer formed in Myc1234 suggested that the different base conformations in the single nucleotide double chain-reversal loops are likely to be responsible for the different thermostability of the two 1:2:1 c-MYC G-quadruplexes. In particular, the adenine base in the single-A loop adopts a so-called 'wing-up' conformation in which the base moiety is extruding away from the groove, whereas the thymine base in the single-T loop adopts a 'wing-down' conformation in which the base

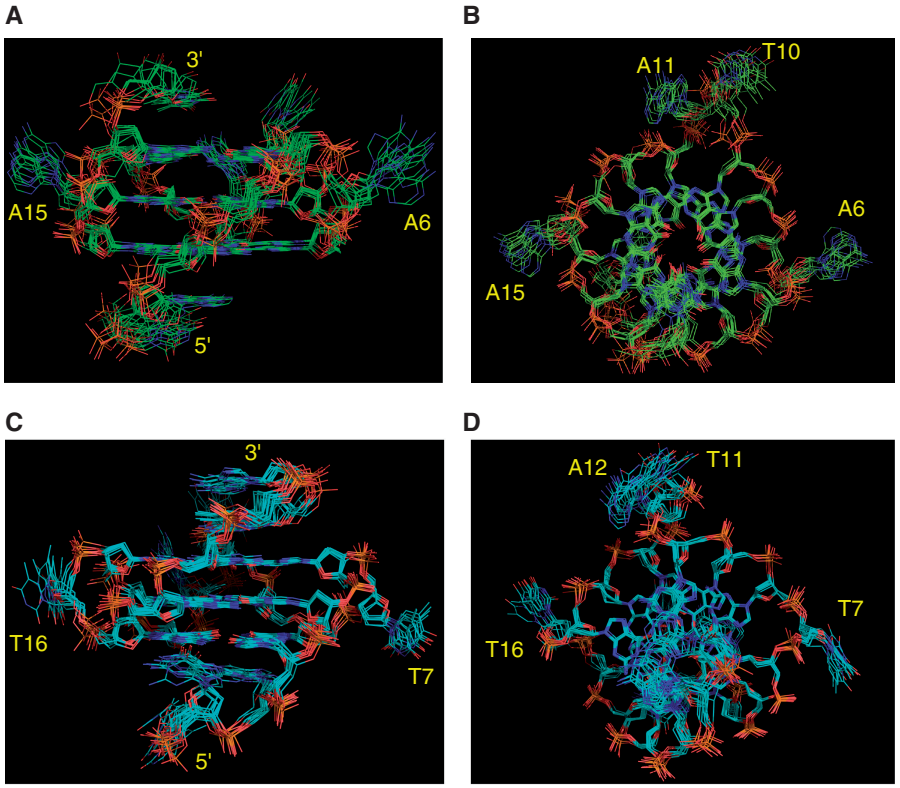

Figure 7. Comparison of the 1:2:1 parallel G-quadruplexes formed by Pu19_A2A11 of Myc1234 and by Pu22_T14T23 of Myc2345. NMR-derived overlaid 10 lowest energy structure bundles of Pu19_A2A11 of Myc1234 in side view (**A**) and top view (**B**), and of Pu22_T14T23 of Myc2345 in side view (**C**) and top view (**D**). The 1-nt adenine loops adopt a 'wing-up' conformation as shown in (A) and (B), whereas the 1-nt thymine loops adopt a 'wing-down' conformation as shown in (C) and (D).

Table 3. Melting temperatures of various modified NHE III_1 sequences

DNA/folding by loop-lengths	Sequence (loop residues underlined)	T_m (°C)
Pu19_A2A11/121	TAGGGAGGGTAGGGAGGGT	63.4
Pu19a/121	TAGGGAGGGTAGGGTGGGT	70.5
Pu19b/112	TAGGGAGGGAGGGTAGGGT	62.9
Pu19c/112	TAGGGAGGGTGGGTAGGGT	70.0
Pu19_A2A14/112	TAGGGAGGGTGGGAAGGGT	65.3
Pu18_T2A11/121	TGGGAGGGTAGGGAGGGT	60.3
Pu18_T14T23/121	TGGGTGGGTAGGGTGGGT	73.7
Pu18a/211	TGGGTAGGGTGGGTGGGT	73.4
Pu18b/211	TGGGTAGGGAGGGTGGGT	66.5
Pu18c/211	TGGGTTGGGAGGGTGGGT	70.2
Pu18d/211	TGGGTTGGGTGGGTGGGT	77.8
Pu18e/121	TGGGTGGGAAGGGTGGGT	69.5
Pu19d/122	TGGGAGGGTTGGGTTGGGT	63.1
Pu19e/212	TGGGTTGGGTGGGTTGGGT	69.9
Pu19f/122	TGGGTGGGTAGGGTTGGGT	65.4
Pu19h/212	TGGGTTGGGAGGGTAGGGT	58.4

moiety is lying along the groove (34) (Figure 7). The wing-up conformation of the single-A loop appears to be energetically less favored as it is more solvent exposed with less groove interactions, and appears to account for ~7°C decrease in T_m in 20 mM K^+ when compared with the wing-down conformation of the single-T loop.

Our systematic thermodynamic analysis of modified c-MYC NHE III_1 sequences provided quantitative measure of the contribution of various loop sequences for parallel G-quadruplexes in general, which have been shown to favor short loop lengths (36). Specifically, our results showed that, in 20 mM K^+, ΔT_m is about −7°C for a T–A substitution for a 1-nt loop, −3.5 to −4.5°C for a TT–TA substitution of the 2-nt loop, and about −8°C for a TT–AA substitution of the 2-nt loop (Table 3). If we consider the TT–AA loop substitution (8°C) to be the summation of two T–A substitutions of the 2-nt loop, i.e. TT to TA and TA–AA loop substitutions, the single T–A substitution in a 2-nt loop appears to account for about −4°C in ΔT_m in 20 mM K^+. Two major loop isomers appear to formed in Myc1234, with the 1:1:2 loop isomer accounting for ~55% and the 1:2:1 loop isomer accounting ~45% of the total population (Figure 1B). This is in accord with the slightly lower T_m (63.4°C) of the 1:2:1 loop isomer isolated by the Pu19_A2A11 sequence than that of the 1:1:2 loop isomers formed by the Pu19_A2A14 sequence (T_m of 65.3°C) (Figure 1A). The 1:2:1 loop isomer formed by Pu19_A2A11 has two 1-nt A loops and a TA loop, while the 1:1:2 loop isomer formed by Pu19_A2A14 has a 1-nt A loop, 1-nt T loop and a AA loop. Based on our thermodynamic analysis, ΔT_m between Pu19_A2A11 and Pu19_A2A14 is estimated to be 3°C in 20 mM K^+, i.e. [7°C (A–T substitution of 1-nt loop) − 4°C (T–A substitution of 2-nt loop)], consistent with the experimentally measured ΔT_m of the two sequences.

Interestingly, it appears that the different thermostability of the Myc2345 and Myc1234 G-quadruplexes is intrinsic to their sequences, as the four consecutive 3′ G-runs of the c-MYC NHE III_1 (Myc2345) are separated by T, A, T, respectively, whereas the four consecutive 5′ G-runs of c-MYC NHE III_1 (Myc1234) are separated by A, T, A, respectively. The two single-A loops in Myc1234 are likely to be responsible for the lower thermostability as compared with Myc2345 which contains two single-T loops. The observation of the formation of the Myc1234 G-quadruplex, which has lower stability, in the supercoiled plasmid footprinting (35) thus points to the potential role of supercoiling in the G-quadruplex formation in the promoter sequences. However, it needs to be noted that the level of superhelicity as well as the lifetimes of different structures would all be important factors for the dynamic equilibrium of multiple G-quadruplexes formed in a promoter element and should be taken into consideration for the development of drugs targeting promoter G-quadruplexes for gene regulation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Megan Carver for proofreading the article.

FUNDING

National Institutes of Health (RR16659 and CA122952 to D.Y.). Funding for open access charge: NIH CA122952.

Conflict of interest statement. None declared.

REFERENCES

1. Spencer,C.A. and Groudine,M. (1991) Control of C-Myc regulation in normal and neoplastic-cells. *Adv. Cancer Res.*, **56**, 1–48.
2. Magrath,I. (1990) The pathogenesis of Burkitt's lymphoma. *Adv. Cancer Res.*, **55**, 133–270.
3. Kinzler,K.W. and Vogelstein,B. (1996) Lessons from hereditary colorectal cancer. *Cell*, **87**, 159–170.
4. Berns,E., Klijn,J.G.M., Vanputten,W.L.J., Vanstaveren,I.L., Portengen,H. and Foekens,J.A. (1992) C-Myc amplification is a better prognostic factor than Her2/Neu amplification in primary breast-cancer. *Cancer Res.*, **52**, 1107–1113.
5. Pertschuk,L.P., Feldman,J.G., Kim,D.S., Nayeri,K., Eisenberg,K.B., Carter,A.C., Thelmo,W.T., Rhong,Z.T., Benn,P. and Grossman,A. (1993) Steroid-hormone receptor immunohistochemistry and amplification of C-Myc protooncogene - relationship to disease-free survival in breast-cancer. *Cancer*, **71**, 162–171.
6. Nupponen,N.N., Kakkola,L., Koivisto,P. and Visakorpi,T. (1998) Genetic alterations in hormone-refractory recurrent prostate carcinomas. *Am. J. Pathol.*, **153**, 141–148.
7. Dang,C.V. (1999) c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol. Cell Biol.*, **19**, 1–11.
8. Nesbit,C.E., Tersak,J.M. and Prochownik,E.V. (1999) MYC oncogenes and human neoplastic disease. *Oncogene*, **18**, 3004–3016.
9. Schlagbauer-Wadl,H., Griffioen,M., van Elsas,A., Schrier,P.I., Pustelnik,T., Eichler,H.G., Wolff,K., Pehamberger,H. and Jansen,B. (1999) Influence of increased c-Myc expression on the

growth characteristics of human melanoma. *J. Investig. Dermatol.*, **112**, 332–336.
10. Freier,K., Joos,S., Flechtenmacher,C., Devens,F., Benner,A., Bosch,F.X., Lichter,P. and Hofele,C. (2003) Tissue microarray analysis reveals site-specific prevalence of oncogene amplifications in head and neck squamous cell carcinoma. *Cancer Res.*, **63**, 1179–1182.
11. Harris,C.P., Lu,X.Y., Narayan,G., Singh,B., Murty,V. and Rao,P.H. (2003) Comprehensive molecular cytogenetic characterization of cervical cancer cell lines. *Genes Chromosomes Cancer*, **36**, 233–241.
12. Dallafavera,R., Bregni,M., Erikson,J., Patterson,D., Gallo,R.C. and Croce,C.M. (1982) Human C-Myc Onc gene is located on the region of chromosome-8 that is translocated in Burkitt-lymphoma cells. *Proc. Natl Acad. Sci. USA*, **79**, 7824–7827.
13. Taub,R., Kirsch,I., Morton,C., Lenoir,G., Swan,D., Tronick,S., Aaronson,S. and Leder,P. (1982) Translocation of the C-Myc gene into the immunoglobulin heavy-chain locus in human Burkitt-lymphoma and murine plasmacytoma cells. *Proc. Natl Acad. Sci. USA*, **79**, 7837–7841.
14. Nowell,P., Finan,J., Favera,R.D., Gallo,R.C., Arrushdi,A., Romanczuk,H., Selden,J.R., Emanuel,B.S., Rovera,G. and Croce,C.M. (1983) Association of amplified oncogene C-Myc with an abnormally banded chromosome-8 in a human-leukemia cell-line. *Nature*, **306**, 494–497.
15. Marcu,K.B., Bossone,S.A. and Patel,A.J. (1992) Myc function and regulation. *Ann. Rev. Biochem.*, **61**, 809–860.
16. Marcu,K.B., Patel,A.J. and Yang,Y. (1997) Differential regulation of the c-MYC P1 and P2 promoters in the absence of functional tumor suppressors: implications for mechanisms of deregulated MYC transcription. *Curr. Top. Microbiol. Immunol: C-Myc in B-Cell Neoplasia*, **224**, 47–56.
17. Berberich,S.J. and Postel,E.H. (1995) Puf/Nm23-H2/Ndpk-B transactivates a human C-Myc promoter-cat gene via a functional nuclease hypersensitive element. *Oncogene*, **10**, 2343–2347.
18. Simonsson,T., Pecinka,P. and Kubista,M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.*, **26**, 1167–1172.
19. Collins,I., Weber,A. and Levens,D. (2001) Transcriptional consequences of topoisomerase inhibition. *Mol. Cell Biol.*, **21**, 8437–8451.
20. 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.
21. Hurley,L.H., Von Hoff,D.D., Siddiqui-Jain,A. and Yang,D.Z. (2006) Drug targeting of the c-MYC promoter to repress gene expression via a G-quadruplex silencer element. *Semin. Oncol.*, **33**, 498–512.
22. Yang,D.Z. and Hurley,L.H. (2006) Structure of the biologically relevant G-quadruplex in the c-MYC promoter. *Nucleosides Nucleotides Nucleic Acids*, **25**, 951–968.
23. Ou,T.M., Lu,Y.J., Zhang,C., Huang,Z.S., Wang,X.D., Tan,J.H., Chen,Y., Ma,D.L., Wong,K.Y., Tang,J.C.O. *et al.* (2007) Stabilization of G-quadruplex DNA and down-regulation of oncogene c-myc by quindoline derivatives. *J. Med. Chem.*, **50**, 1465–1474.
24. Yang,D.Z. and Okamoto,K. (2010) Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med. Chem.*, **2**, 619–646.
25. Punchihewa,C. and Yang,D.Z. (2009) Therapeutic targets and drugs-G-quadruplex inhibitors. In Hiyama,K. (ed.), *Telomeres and Telomerase in Cancer*. Springer, NJ, USA, pp. 251–280.
26. Kumari,S., Bugaut,A., Huppert,J.L. and Balasubramanian,S. (2007) An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.*, **3**, 218–221.
27. Brooks,T.A. and Hurley,L.H. (2009) The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nat. Rev. Cancer*, **9**, 849–861.
28. Lavelle,C. (2008) DNA torsional stress propagates through chromatin fiber and participates in transcriptional regulation. *Nat. Struct. Mol. Biol.*, **15**, 123–125.
29. Kouzine,F., Sanford,S., Elisha-Feil,Z. and Levens,D. (2008) The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat. Struct. Mol. Biol.*, **15**, 146–154.
30. Dexheimer,T.S., Carey,S.S., Zuohe,S., Gokhale,V.M., Hu,X., Murata,L.B., Maes,E.M., Weichsel,A., Sun,D., Meuillet,E.J. *et al.* (2009) NM23-H2 may play an indirect role in transcriptional activation of c-myc gene expression but does not cleave the nuclease hypersensitive element III1. *Mol. Cancer Ther*, **8**, 1363–1377.
31. Gonzalez,V., Guo,K., Hurley,L. and Sun,D. (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.*, **284**, 23622–23635.
32. Seenisamy,J., Rezler,E.M., Powell,T.J., Tye,D., Gokhale,V., Joshi,C.S., Siddiqui-Jain,A. and Hurley,L.H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.*, **126**, 8702–8709.
33. Phan,A.T., Modi,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.
34. Ambrus,A., Chen,D., Dai,J., Jones,R.A. and Yang,D.Z. (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.
35. Sun,D. and Hurley,L.H. (2009) The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J. Med. Chem.*, **52**, 2863–2874.
36. Qin,Y. and Hurley,L.H. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*, **90**, 1149–1171.
37. Dai,J.X., Dexheimer,T.S., Chen,D., Carver,M., Ambrus,A., Jones,R.A. and Yang,D.Z. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.*, **128**, 1096–1098.
38. Rachwal,P.A., Findlow,I.S., Werner,J.M., Brown,T. and Fox,K.R. (2007) Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Res.*, **35**, 4214–4222.
39. Bugaut,A. and Balasubramanian,S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry*, **47**, 689–697.
40. Hatzakis,E., Okamoto,K. and Yang,D.Z. (2010) Thermodynamic stability and folding kinetics of the major G-quadruplex and its loop isomers formed in the nuclease hypersensitive element in the human c-Myc promoter: effect of loops and flanking segments on the stability of parallel-stranded intramolecular G-quadruplexes. *Biochemistry*, **49**, 9152–9160.
41. Dai,J., Chen,D., Jones,R.A., Hurley,L.H. and Yang,D.Z. (2006) NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.*, **34**, 5133–5144.
42. Dai,J., Carver,M., Punchihewa,C., Jones,R.A. and Yang,D.Z. (2007) Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.*, **35**, 4927–4940.
43. Brünger,A.T. (1993) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
44. Masiero,S., Trotta,R., Pieraccini,S., De Tito,S., Perone,R., Randazzo,A. and Spada,G.P. (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Org. Biomol. Chem.*, **8**, 2683–2692.