9264–9275 *Nucleic Acids Research, 2018, Vol. 46, No. 18* *Published online 3 September 2018*
doi: 10.1093/nar/gky757

Loop permutation affects the topology and stability of G-quadruplexes

Mingpan Cheng [1,2], Yu Cheng[1,2], Jingya Hao[1,2], Guoqing Jia[1], Jun Zhou [3], Jean-Louis Mergny [3,4,5,*] and Can Li[1,*]

[1]State Key Laboratory of Catalysis, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China, [2]University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China, [3]State Key Laboratory of Analytical Chemistry for Life Science, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, China, [4]ARNA Laboratory, Inserm U1212, CNRS UMR5320, IECB, Université de Bordeaux, Pessac 33607, France and [5]Institute of Biophysics of the CAS, v.v.i., Královopolská 135, 612 65 Brno, Czech Republic

Received June 06, 2018; Revised August 08, 2018; Editorial Decision August 09, 2018; Accepted August 21, 2018

ABSTRACT

G-quadruplexes are unusual DNA and RNA secondary structures ubiquitous in a variety of organisms including vertebrates, plants, viruses and bacteria. The folding topology and stability of intramolecular G-quadruplexes are determined to a large extent by their loops. Loop permutation is defined as swapping two or three of these regions so that intramolecular G-quadruplexes only differ in the sequential order of their loops. Over the past two decades, both length and base composition of loops have been studied extensively, but a systematic study on the effect of loop permutation has been missing. In the present work, 99 sequences from 21 groups with different loop permutations were tested. To our surprise, both conformation and thermal stability are greatly dependent on loop permutation. Loop permutation actually matters as much as loop length and base composition on G-quadruplex folding, with effects on T_m as high as 17°C. Sequences containing a longer central loop have a high propensity to adopt a stable non-parallel topology. Conversely, sequences containing a short central loop tend to form a parallel topology of lower stability. In addition, over half of interrogated sequences were found in the genomes of diverse organisms, implicating their potential regulatory roles in the genome or as therapeutic targets. This study illustrates the structural roles of loops in G-quadruplex folding and should help to establish rules to predict the folding pattern and stability of G-quadruplexes.

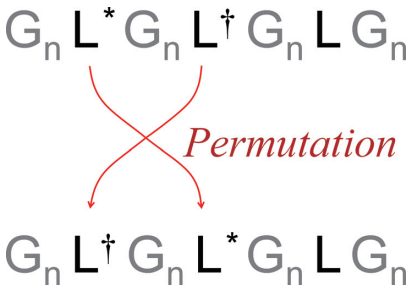
INTRODUCTION

Single-stranded guanine-rich DNA and RNA sequences can fold into stable G-quadruplex structures (1–3). Many G-quadruplex forming sequences fit the consensus $G_{\geq 2}L_{\geq 1}G_{\geq 2}L_{\geq 1}G_{\geq 2}L_{\geq 1}G_{\geq 2}$, in which adjacent G-tracts are spaced by a linker sequence composed of one or more nucleotide and called a loop ($L_{\geq 1}$). Four tracts of two or more guanines ($G_{\geq 2}$), or for a more conservative definition, three or more guanines ($G_{\geq 3}$) generally constitute the G-quadruplex core. G-quadruplex formation is achieved by π–π stacking of two or more planar G-quartets in the presence of cations such as potassium.

Potential DNA G-quadruplex sequences are distributed ubiquitously in diverse organisms including human (4–6), vertebrates (7), plants (8), viruses (9,10) and bacteria (11). More and more studies suggest that this unusual secondary DNA structure is important in telomere biology (12,13) and gene regulation (13–15). In addition, a number of small-molecule ligands with a high specificity for G-quadruplexes have been synthesized for their therapeutic potential or as structural probes (16–19). Conformational polymorphism and high stability endow G-quadruplex structures with diverse functional applications, such as drugs (20–22), DNAzymes (23–25), bio-sensors (26), aptamers (27–29), nano-devices (30,31) and fluorescent-protein mimics (32). Therefore, it is essential to uncover the rules that determine the folding pattern and stability of these intramolecular structures.

Loop parameters such as length (33–39) and base composition (40–44) play an important role in the folding conformations and stability of G-quadruplexes, and consequently affect their biological functions, molecular recognition and applications. Sugiyama *et al.* even claimed that loops contribute more to the structural stability than the stacking of G-quartets (45). Phan *et al.* elegantly demonstrated that

*To whom correspondence should be addressed. Tel: +86 0411 84379070; Fax: +86 0411 84694447; Email: canli@dicp.ac.cn
Correspondence may also be addressed to Jean-Louis Mergny. Tel: +33 540003022; Fax: +33 540003004; Email: jean-louis.mergny@inserm.fr



Scheme 1. Loop permutation is defined as the swap between two sequences (labelled L^* and $L^\dagger$) of an intramolecular G-quadruplex keeping length and overall base composition constant. These sequences belong to the same group.

the G-quadruplex within the CEB1 minisatellite containing short length and pyrimidine bases in loops adopts a parallel topology with a highly thermal stability, resulting in a severe genomic instability *in vivo* (46). Richter *et al*. (47) and Oyoshi *et al.* (48) found that nucleolin and Ewing's sarcoma proteins preferentially bind the G-quadruplex with longer loop length. Maiti *et al.* revealed that longer loop length shifts the equilibrium from G-quadruplex to duplex in the presence of its complementary cytosine-rich sequence (49). Additionally, loop regions are attractive binding sites for targeting ligands (50,51). Neidle *et al.* specially reviewed functions of loops in the interactions of human telomeric G-quadruplex with ligands (52). Moreover, Zhou *et al.* highlighted the role of adenine repeats from central loop in activity enhancement of G-quadruplex-based DNAzyme (53). To date, the effects of loop length and base composition on altering the structural stability and functions have been extensively studied. However, as the sequences space allowed for loops variation is huge, a number of questions remain regarding the dependency of G-quadruplex folding on loop sequence.

Loop permutation is defined as the exchange of two linker sequences, as shown in Scheme 1. The resulting G-quadruplex-forming sequences share identical G-tracts, loop length and base composition but only differ in the sequential order of their three loops. By definition, all sequences resulting from one or more loop permutations are considered to be in the same group. Previous works disclosed an impressive role of loop permutation in controlling the folding patterns of G-quadruplex (54). That study analyzed a very limited number of samples (six sequences from two groups) derived from human telomeric motif. Another study concluded that the effect of loop permutation on stability was hardly significant (<4°C) (55). This scarcity of data incited us to check whether this factor is a general role on topology and stability of G-quadruplexes.

Here, 99 sequences from 21 groups with different loop permutations were interrogated. Differences in melting temperature (T_m) caused by changing the sequential order of loops can reach 17°C. Both conformation and thermal stability are greatly dependent on loop permutation. Sequences containing the longest loop in central position have a high propensity to form highly stable non-parallel topologies. This study complements the structural roles of loops in G-quadruplex folding, as well as contributes to establish rules to predict the folding pattern and stability of intramolecular G-quadruplex *in vitro* through oligonucleotide sequences.

MATERIALS AND METHODS

DNA sample preparation

DNA oligonucleotides were purchased from Sangon Biotech, Co., Ltd. and purified by ultra-PAGE. Sequences are given in Table 1. Samples were prepared in distilled and deionized H_2O (18.2 MΩ, Milli-Q A10). Concentration of DNA sample was determined from the absorbance at 260 nm by using the extinction coefficients. Unless otherwise stated, DNA samples were heated in 20 mM KPi buffer (pH 7.0) supplemented with 80 mM KCl (total potassium concentration is 112.3 mM) for 3 min at 95°C and annealed slowly over the course of 2 h to room temperature, and then incubated at 4°C before spectroscopy measurements.

Circular dichroism (CD) spectroscopy and CD-melting experiments

Circular dichroism (CD) spectra were recorded on a dual beam DSM 1000 CD spectrophotometer (Olis, Inc.) equipped with a Peltier temperature controller, using quartz cells of 10-mm path length. DNA samples were prepared at 5 μM concentration in 20 mM KPi buffer (pH 7.0) supplemented with 80 mM KCl. Unless otherwise stated, each measurement was the average of 10 scans recorded from 230 to 320 nm at 20°C. The scanning rate was automatically selected by the Olis software as function of the signal intensity to optimize data collection. CD data were normalized to molar dichroic absorption ($\Delta\varepsilon$) on the basis of DNA strand concentration by using the following Equation (1):

$$\Delta\varepsilon = \theta / (32980 \times c \times l), \quad (1)$$

in which θ is the ellipticity in millidegree, c is the concentration of nucleotide in mol/l, and l is the path length in cm.

Conformations are usually identified by CD spectra (56) and distinguished by values of r (Supplementary Figure S1; Equation 2):

$$r = \frac{CD_{265}}{|CD_{265}| + CD_{290}}, \quad (2)$$

where CD_{265} and CD_{290} are the molar dichroic absorption of peaks ~265 and 290 nm, respectively. $1 > r \geq 0.5$, $0.5 > r > 0$ and $r < 0$ correspond to predominantly parallel, hybrid and antiparallel topologies, respectively (56).

CD-melting experiment was performed on a Chirascan circular spectropolarimeter (Applied Photophysics). DNAs (5 μM) were tested in 20 mM KPi buffer (pH 7.0) containing 80 mM KCl. The melting profile was obtained at 265 nm with a temperature gradient of 1.0°C/min while thermally equilibrating at each step for 30 s prior to recording the absorbance. Melting temperature (T_m) was determined by the calculation of algebraic mean of pre- and post-melting baselines, assuming two-state melting process.

Table 1. Conformation (r) and stability (T_m) of sequences with different loop permutations

Name[a]	Loop permutation[b]	Sequence (5′→3′)	r[c]	T_m / °C[d]
124 Group				
124	A1	GGG T GGG TT GGG TTTT GGG	0.95	70.5
142	A2	GGG T GGG TTTT GGG TT GGG	0.94	72.8
214	A3	GGG TT GGG T GGG TTTT GGG	0.96	71.9
241	A4	GGG TT GGG TTTT GGG T GGG	0.91	71.0
412	A5	GGG TTTT GGG T GGG TT GGG	0.95	72.3
421	A6	GGG TTTT GGG TT GGG T GGG	0.94	72.8
133 Group				
133	A1	GGG T GGG TTT GGG TTT GGG	0.60	61.6
313	A2	GGG TTT GGG T GGG TTT GGG	0.90	60.9
331	A3	GGG TTT GGG TTT GGG T GGG	0.69	65.5
223 Group				
223	A1	GGG TT GGG TT GGG TTT GGG	0.77	59.7
232	A2	GGG TT GGG TTT GGG TT GGG	0.55	62.7
322	A3	GGG TTT GGG TT GGG TT GGG	0.59	61.4
125 Group				
125	A1	GGG T GGG TT GGG TTTTT GGG	0.94	68.2
152	A2	GGG T GGG TTTTT GGG TT GGG	0.90	67.4
215	A3	GGG TT GGG T GGG TTTTT GGG	0.94	68.2
251	A4	GGG TT GGG TTTTT GGG T GGG	0.91	68.5
512	A5	GGG TTTTT GGG T GGG TT GGG	0.94	67.8
521	A6	GGG TTTTT GGG TT GGG T GGG	0.95	69.1
134 Group				
134	A1	GGG T GGG TTT GGG TTTT GGG	0.51	65.3
143	A2	GGG T GGG TTTT GGG TTT GGG	0.21	70.6
314	A3	GGG TTT GGG T GGG TTTT GGG	0.91	65.2
341	A4	GGG TTT GGG TTTT GGG T GGG	0.32	65.7
413	A5	GGG TTTT GGG T GGG TTT GGG	0.80	64.7
431	A6	GGG TTTT GGG TTT GGG T GGG	0.81	65.0
224 Group				
224	A1	GGG TT GGG TT GGG TTTT GGG	0.95	64.5
242	A2	GGG TT GGG TTTT GGG TT GGG	0.44	64.7
422	A3	GGG TTTT GGG TT GGG TT GGG	0.92	64.7
233 Group				
233	A1	GGG TT GGG TTT GGG TTT GGG	0.40	67.8
323	A2	GGG TTT GGG TT GGG TTT GGG	0.46	63.6
332	A3	GGG TTT GGG TTT GGG TT GGG	0.41	67.8
126 Group				
126	A1	GGG T GGG TT GGG TTTTTT GGG	0.94	58.5
162	A2	GGG T GGG TTTTTT GGG TT GGG	0.88	56.5
216	A3	GGG TT GGG T GGG TTTTTT GGG	0.97	58.0
261	A4	GGG TT GGG TTTTTT GGG T GGG	0.87	58.1
612	A5	GGG TTTTTT GGG T GGG TT GGG	0.98	60.0
621	A6	GGG TTTTTT GGG TT GGG T GGG	0.97	58.3
135 Group				
135	A1	GGG T GGG TTT GGG TTTTT GGG	0.44	61.7
153	A2	GGG T GGG TTTTT GGG TTT GGG	0.28	70.2
315	A3	GGG TTT GGG T GGG TTTTT GGG	0.89	61.4
351	A4	GGG TTT GGG TTTTT GGG T GGG	0.36	62.9
513	A5	GGG TTTTT GGG T GGG TTT GGG	0.90	61.9
531	A6	GGG TTTTT GGG TTT GGG T GGG	0.88	61.1
144 Group				
144	A1	GGG T GGG TTTT GGG TTTT GGG	0.42	65.8
414	A2	GGG TTTT GGG T GGG TTTT GGG	0.64	60.4
441	A3	GGG TTTT GGG TTTT GGG T GGG	0.43	61.1
225 Group				
225	A1	GGG TT GGG TT GGG TTTTT GGG	0.92	59.3
252	A2	GGG TT GGG TTTTT GGG TT GGG	0.40	63.3
522	A3	GGG TTTTT GGG TT GGG TT GGG	0.90	60.5
234 Group				
234	A1	GGG TT GGG TTT GGG TTTT GGG	0.45	63.4
243	A2	GGG TT GGG TTTT GGG TTT GGG	0.34	71.3
324	A3	GGG TTT GGG TT GGG TTTT GGG	0.20	59.5
342	A4	GGG TTT GGG TTTT GGG TT GGG	0.27	66.4
423	A5	GGG TTTT GGG TT GGG TTT GGG	0.20	58.5
432	A6	GGG TTTT GGG TTT GGG TT GGG	0.47	64.2
127 Group				
127	A1	GGG T GGG TT GGG TTTTTTT GGG	0.96	56.2
172	A2	GGG T GGG TTTTTTT GGG TT GGG	0.92	54.1
217	A3	GGG TT GGG T GGG TTTTTTT GGG	0.97	57.1
271	A4	GGG TT GGG TTTTTTT GGG T GGG	0.89	56.7
712	A5	GGG TTTTTTT GGG T GGG TT GGG	0.97	57.1

Table 1. Continued

Name[a]	Loop permutation[b]	Sequence (5′→3′)	r[c]	T_m / °C[d]
721	A6	GGG TTTTTTT GGG TT GGG T GGG	0.97	55.7
136 Group				
136	A1	GGG T GGG TTT GGG TTTTTT GGG	0.59	61.0
163	A2	GGG T GGG TTTTTT GGG TTT GGG	0.41	68.3
316	A3	GGG TTT GGG T GGG TTTTTT GGG	0.92	58.1
361	A4	GGG TTT GGG TTTTTT GGG T GGG	0.55	61.8
613	A5	GGG TTTTTT GGG T GGG TTT GGG	0.89	59.0
631	A6	GGG TTTTTT GGG TTT GGG T GGG	0.70	58.4
145 Group				
145	A1	GGG T GGG TTTT GGG TTTTT GGG	0.43	61.9
154	A2	GGG T GGG TTTTT GGG TTTT GGG	0.44	65.1
415	A3	GGG TTTT GGG T GGG TTTTT GGG	0.70	56.5
451	A4	GGG TTTT GGG TTTTT GGG T GGG	0.46	58.5
514	A5	GGG TTTTT GGG T GGG TTTT GGG	0.63	56.5
541	A6	GGG TTTTT GGG TTTT GGG T GGG	0.28	58.3
226 Group				
226	A1	GGG TT GGG TT GGG TTTTTT GGG	0.85	55.2
262	A2	GGG TT GGG TTTTTT GGG TT GGG	0.39	58.8
622	A3	GGG TTTTTT GGG TT GGG TT GGG	0.92	56.3
235 Group				
235	A1	GGG TT GGG TTT GGG TTTTT GGG	0.41	62.0
253	A2	GGG TT GGG TTTTT GGG TTT GGG	0.05	70.0
325	A3	GGG TTT GGG TT GGG TTTTT GGG	0.26	56.5
352	A4	GGG TTT GGG TTTTT GGG TT GGG	0.38	63.7
523	A5	GGG TTTTT GGG TT GGG TTT GGG	0.27	55.5
532	A6	GGG TTTTT GGG TTT GGG TT GGG	0.38	60.4
244 Group				
244	A1	GGG TT GGG TTTT GGG TTTT GGG	0.26	64.5
424	A2	GGG TTTT GGG TT GGG TTTT GGG	−0.14	60.0
442	A3	GGG TTTT GGG TTTT GGG TT GGG	−0.25	62.5
334 Group				
334	A1	GGG TTT GGG TTT GGG TTTT GGG	0.28	62.9
343	A2	GGG TTT GGG TTTT GGG TTT GGG	0.33	68.2
433	A3	GGG TTTT GGG TTT GGG TTT GGG	0.38	63.2

[a]Name of sequence is the three consecutive numbers, referring to lengths of the three loops in the 5′ to 3′ direction. Group names are written in italics.
[b]A1–A6 and A1–A3 of each group are the arrangement of three loops, used for statistical analysis (see below).
[c]Conformations are identified by CD spectra and distinguished by values of r (see Equation 2). Standard deviation of r was <0.05 as determined by two or more independent measurements.
[d]T_m is determined by UV-melting experiment. Standard deviation of T_m of two independent measurements was <1.0°C.

UV-melting experiment and UV absorbance spectroscopy

Ultraviolet (UV)-melting and UV absorbance measurements were performed on a Shimadzu 2450 spectrophotometer equipped with a Peltier temperature control accessory in a sealed quartz cell of 10-mm path length. DNAs (5 μM) were generally tested in 20 mM KPi buffer (pH 7.0) containing 80 mM KCl. The melting profiles were obtained at 295 nm with a temperature gradient of 0.5°C/min (57). Melting temperature (T_m) was determined by the method described in CD section.

Differential scanning calorimetry (DSC)

Differential scanning calorimetry (DSC) measurements were carried out using a TA Nano DSC. DNA was prepared as 200 μM in 20 mM KPi buffer (pH 7.0) supplemented with 80 mM KCl. Scans were performed at 0.5°C/min in the 30–95°C temperature range. Reversibility for each DNA sample was confirmed. The DNA sample versus buffer scan was subtracted by the previously performed buffer for all the scans. T_m was obtained using TwoStateScaled model to fit the heat capacity curve by NanoAnalyze software.

1D imino NMR

^{1}H NMR experiments were performed on a 700 MHz Bruker Avance III HD spectrometer equipped with a cryogenic QCI probe at 20°C. DNA concentration was 200 μM. The WATERGATE pulse program was used in recording ^{1}H spectra.

Non-denaturing gel electrophoresis

DNA samples were incubated in 20 mM KPi buffer (pH 7.0) with 80 mM KCl. Samples were run on a 20% native polyacrylamide gel (7 cm × 10 cm), made up in 1× Tris-Borate-EDTA (TBE) buffer containing 20 mM KCl (also used for the running buffer) for 4.0 h at 100 V. Gels were stained with Stains-all (Sigma-Aldrich, 95%) and then destained under sunlight.

RESULTS

Sequences with different loop permutations

To understand the effects of loop permutation on G-quadruplex folding, 19 groups including 87 model se-

quences containing four tracts of three guanines were initially investigated (Table 1). Taking into account that flanking bases may affect the topology of intramolecular structures, no flanking bases were added to G-quadruplex cores in all sequences. To focus on loop permutation, loops for all sequences are composed of only thymine unless otherwise specified. These sequences are represented as 5′-GGG L_a GGG L_b GGG L_c GGG-3′. Letters a, b and c are the nucleotide numbers of first, second and third loops in the 5′ to 3′ direction, respectively. Total loop length (a + b + c) varied between 7 and 10 nts. 12 of these 87 sequences analyzed here were previously studied by Guédin *et al.* (39); the conformations and relative stabilities are in agreement with their findings. Each sequence is defined by three successive digits a, b and c. Each group is named after the first sequence in this group and it is written in italics. Sequences are divided into three types based on lengths of their three loops (Table 2).For convenience, letter x refers to one of a, b and c; letter y represents one of the left two; letter z denotes the last one.

- Sequences with $x \neq y \neq z$ in a group have six different arrangements of loop lengths: A1 ($a < b < c$), A2 ($a < c < b$), A3 ($b < a < c$), A4 ($c < a < b$), A5 ($b < c < a$) and A6 ($c < b < a$). For example, in the *125* group (three loops of 1, 2 and 5 nts), A1 corresponds to 125, A2 to 152, A3 to 215, A4 to 251, A5 to 512 and A6 to 521. A2 and A4 correspond to the sequences having a long central loop, while A3 and A5 to the motif with a short central loop.
- Sequences with $x = y < z$ in a group have three different arrangements of loop lengths: A1 ($a = b < c$), A2 ($a = c < b$) and A3 ($b = c < a$). For example, in the *223* group (three loops of 2, 2 and 3 nts), A1 corresponds to 223, A2 to 232 and A3 to 322. For these groups, A2 corresponds to the situation where the central loop is the longest.
- Sequences with $x = y > z$ in a group have three different arrangements of loop lengths: A1 ($b = c > a$), A2 ($a = c > b$) and A3 ($b = c > a$). For example, in the *133* group (three loops of 1, 3 and 3 nts), A1 corresponds to 133, A2 to 313 and A3 to 331. For these groups, A2 corresponds to the situation where the central loop is shorter than the two others.

There are 10, 5 and 4 groups of sequences with loop lengths patterns of $x \neq y \neq z$, $x = y < z$ and $x = y > z$, respectively (Table 2). All following experiments were carried out in near-physiological solution conditions (80 mM KCl and 20 mM KPi buffer, pH 7.0).

Roles of loop permutation on conformation

CD spectra were recorded to identify G-quadruplex topologies (Figure 1 and Supplementary Figure S2). Intensities of peaks ∼265 and 290 nm are presented in Supplementary Table S1. The corresponding *r* values are given in Table 1. Based on these *r* values, 24 sequences from six groups (*124*, *133*, *223*, *125*, *126* and *127*) fold into parallel topologies. Eighteen sequences from four groups (*233*, *234*, *235* and *334*) adopt hybrid topologies. Other sequences from the same groups adopt two disparate conformations. The sequences from eight groups (*134*, *224*, *135*, *144*, *225*, *136*, *145* and *226*) fold into either parallel or hybrid topologies. For examples, in *134* and *135* groups, the predominant conformations of sequences 134, 314, 413, 431, 315, 513 and 531 are parallel topologies, while 143, 341, 135, 153 and 351 sequences adopt hybrid topologies (Figure 1B and Supplementary Figure S2F). A homogeneous intramolecular structure of hybrid topology formed by 143 was determined using 2D NMR and photochemical methods by Webba da Silva *et al.* (58). The *244* group is a more intriguing case, as not only hybrid (244 sequence) but also antiparallel (424 and 442 sequences) topologies are formed (Figure 1D). This illustrates that sequences from the same group, thus having the exactly same loop length and nucleotide base composition may differ in topology. These examples prove that cases of G-quadruplex conformations dictated by loop permutation are more widespread than expected.

Figure 2 summarizes the CD results by *r* values (see the definition at Supplementary Figure S1, calculated in Supplementary Table S1) for each sequence (belonging to A1–A6 for groups with $x \neq y \neq z$ or A1–A3 for groups with $x = y < z$ and $x = y > z$). Through counting the occurrence frequencies of parallel, hybrid and antiparallel conformations for each loop permutation, a salient phenomenon of loop permutation dependent conformation is disclosed (Figure 3A). In groups with $x \neq y \neq z$, sequences with A3 and A5 arrangements have the highest tendency (verified for 8 of the 10 groups) to fold into parallel topologies, while the remaining two sequences adopt hybrid topologies. Sequences with A2 and A4 arrangements often (6/10) form hybrid and less frequently (4/10) fold into parallel topologies. In groups with $x = y < z$, four in five sequences with A2 arrangement adopt hybrid topologies, while the remaining one (232 sequence) folds into a parallel topology (Figure 3B). In comparison, most sequences (4/5) with A1 and A3 arrangements form parallel topologies and only 334 and 433 sequences fold into hybrid topologies. In groups with $x = y > z$, sequences with A2 arrangement have the highest (2/4) and the lowest possibility (1/4) to adopt parallel and hybrid topologies, respectively (Figure 3C). On the contrary, sequences with A1 and A3 arrangements have the highest possibility (4/4) to fold into hybrid topologies. Generally speaking, sequences for which the central loop is the longest have a high propensity to form hybrid topologies and are less likely to adopt parallel topologies as compared to other arrangements. In contrast, sequences for which the central loop is the shortest have a high propensity to form parallel topologies and a lower propensity to fold into hybrid topologies.

Roles of loop permutation on thermal stability

G-quadruplex stabilities were evaluated by the melting temperatures (T_m), which were calculated from UV-melting curves (Supplementary Figure S3). T_m values of all sequences are given in Table 1. In addition, T_m values of *134* and *135* groups were further confirmed by DSC and CD-melting experiments (Supplementary Figures S4 and S5). T_m values versus different loop permutations (A1–A6 in groups with $x \neq y \neq z$ and A1–A3 in groups with $x = y < z$ and $x = y > z$) are plotted in Figure 4. The effects of loop permutation on stability are summarized in Figure 5. The differences between maximum and minimum T_m

Table 2. Sequence groups with different lengths of their three loops[a]

Total Loop length	x ≠ y ≠ z (A1–A6, 10 groups)		x = y < z (A1–A3, 5 groups)	x = y > z (A1–A3, 4 groups)
7	*124*		*223*	*133*
8	*125*	*134*	*224*	*233*
9	*126*	*135*	*225*	*144*
	234			
10	*127*	*136*	*226*	*244*
	145	*235*	*334*	

[a]Group names are written in italics.

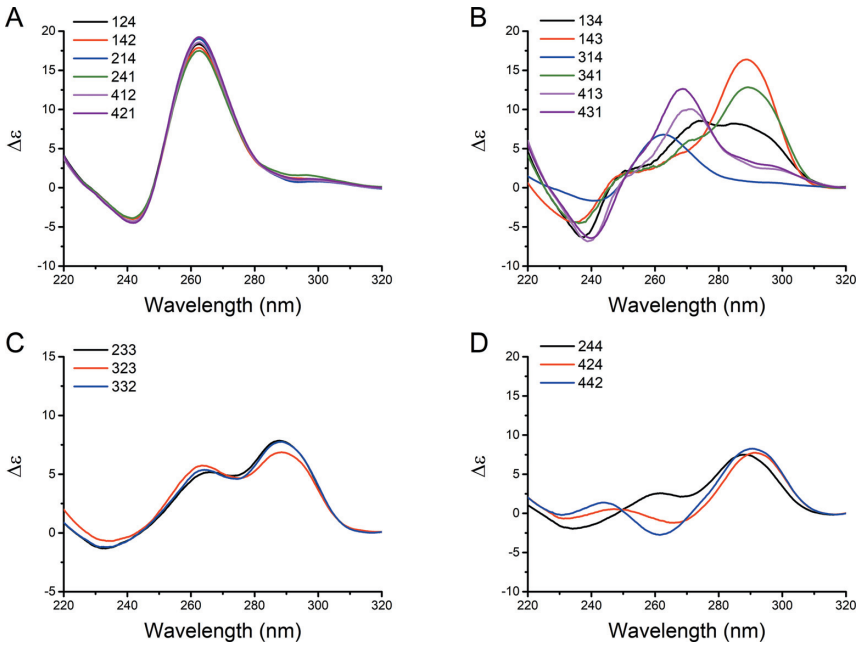


Figure 1. CD spectra of selected groups. (**A**) *124* group, all sequences adopt parallel topologies; (**B**) *134* group, sequences fold into either parallel or hybrid topologies; (**C**) *233* group, all sequences form hybrid topologies; (**D**) *244* group, sequences fold into either antiparallel or hybrid topologies; other groups are given in Supplementary Figure S2. CD experiments were carried out at 20°C by using 5 μM strand concentrations in 20 mM KPi buffer (pH 7.0) supplemented with 80 mM KCl.

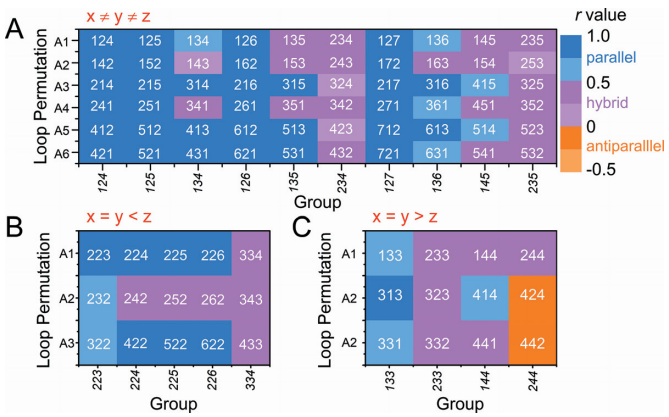


Figure 2. Distribution of *r* values of sequences with different loop permutations. (**A**) Groups with x ≠ y ≠ z; (**B**) Groups with x = y < z; (**C**) Groups with x = y > z. Parallel ($1 > r \geq 0.5$), hybrid ($0.5 > r > 0$) and antiparallel ($r < 0$) topologies are filled in blue, red and orange colors. CD spectroscopies were carried out using 5 μM strand concentrations in 20 mM KPi buffer (pH 7.0) with 80 mM KCl.

(ΔT_m) values within the same group vary from 0.2°C (*224* group) to 14.5°C (*235* group). This ΔT_m range is in line with ours and others' work (39,54,55,59,60). For eight groups (*134*, *135*, *144*, *234*, *136*, *145*, *235* and *334*), the ΔT_m values within a group are above 5°C. This highly significant difference in stability highlights the importance of loop permutation on stability, which may affect their biological relevance (15,61,62), such as the induction of a genomic instability by G-quadruplex structures (63,64).

Within each loop permutation, interesting trends can be found for the maximum and minimum T_m (Figure 6). In groups with x ≠ y ≠ z, A2 arrangement corresponds to the maximum T_m in 7 of the 10 groups (Figure 6A). In contrast, A5 corresponds to the least stable sequences in 4 of the 10 groups. For the groups with x = y < z, the trend is even clearer, as A2 and A1 always correspond to the most stable and least stable sequences, respectively (Figure 6B). In groups with x = y > z, all sequences with A2 arrangement exhibit the lowest T_m, while the most stable sequences correspond to A1 (in two of the four cases) or A3 (in the remaining two cases) (Figure 6C).

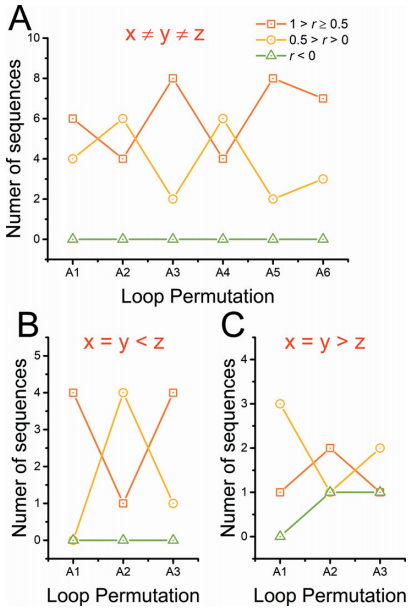


Figure 3. Conformations are dependent on the loop arrangements. (**A**) Groups with x ≠ y ≠ z; (**B**) Groups with x = y < z; (**C**) Groups with x = y > z. The number of sequences with parallel, hybrid and antiparallel conformations is counted for each loop permutation. A1–A6 and A1–A3 for each sequence are presented in Table 1.

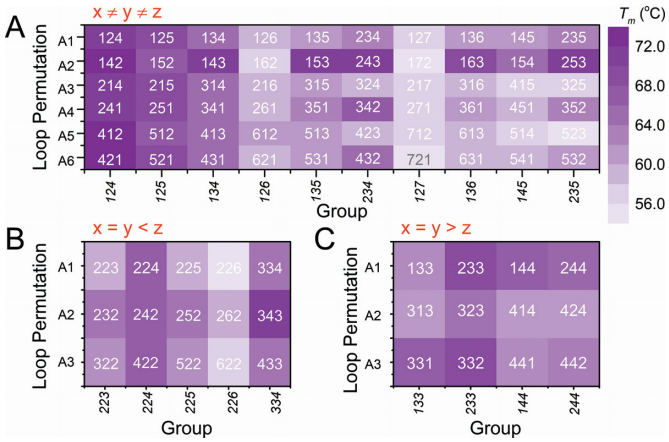


Figure 4. Distribution of T_m values of sequences with different loop permutations. (**A**) Groups with x ≠ y ≠ z; (**B**) Groups with x = y < z; (**C**) Groups with x = y > z. Darker color represents a higher temperature. T_m values were determined by UV-melting experiments, 5 μM strand concentration in 20 mM KPi buffer (pH 7.0) with 80 mM KCl.

Another way to investigate how the loop permutation affect stability is to calculate the average T_m for each class (Figure 7). In groups with x ≠ y ≠ z, average T_m of sequences with A2 and A4 arrangements is the highest and second highest, respectively. On the contrary, A3 and A5 arrangements correspond to the lowest and second lowest average T_m, respectively. For the groups with x = y < z, average T_m of sequences with A2 arrangement is higher than that with A1 and A3 arrangements. In groups with x = y > z, average T_m of sequences with A2 arrangement is lower than that with A1 and A3 arrangements. Overall, a consistent trend emerges: sequences containing a long central loop

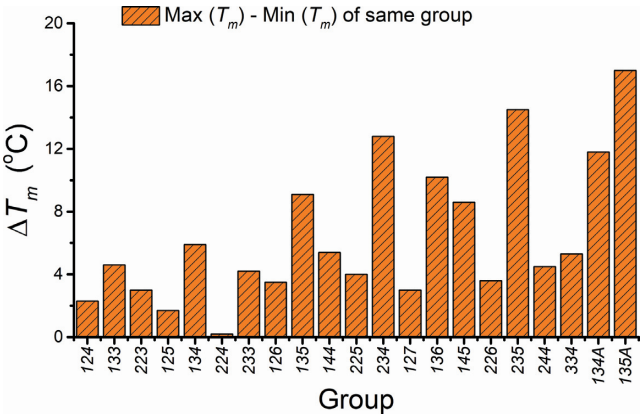


Figure 5. Effects of loop permutation on T_m for each group sequence. ΔT_m= Max (T_m) − Min (T_m), where Max (T_m) and Min (T_m) are the maximum and minimum T_m within the same group.

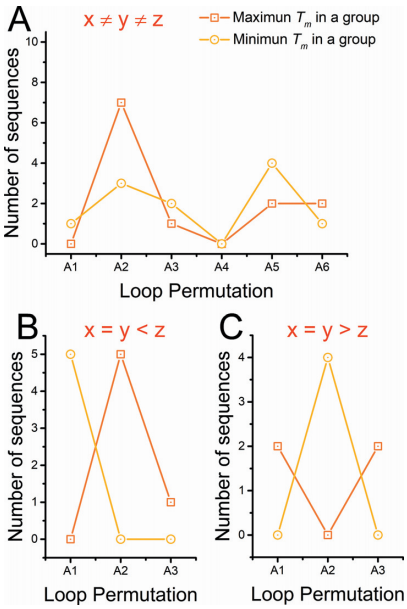


Figure 6. Thermal stabilities (T_m) are dependent on loop permutation. (**A**) Groups with x ≠ y ≠ z; (**B**) Groups with x = y < z; (**C**) Groups with x = y > z. The number of sequences with maximum and minimum T_m for the same group is counted for each loop permutation. A1–A6 and A1–A3 for each sequence are presented in Table 1.

are more likely to exhibit a high thermal stability, while sequences containing a short central loop tend to exhibit a lowest stability.

Molecularity

Twelve sequences from the *134* and *135* groups were selected to investigate the molecularity of G-quadruplexes. Thermal and isothermal difference spectra (TDS and IDS) with two positive peaks ∼240 and 275 nm and a negative peak ∼295 nm confirm the formation of G-quadruplex structures (Supplementary Figures S6 and S7) (57,65,66). Signals of imino protons upon ^{1}H NMR spectra between 9.5 and 14 ppm also demonstrate the G-quadruplex formation (Supplementary Figure S8). The formation of homogeneous in-

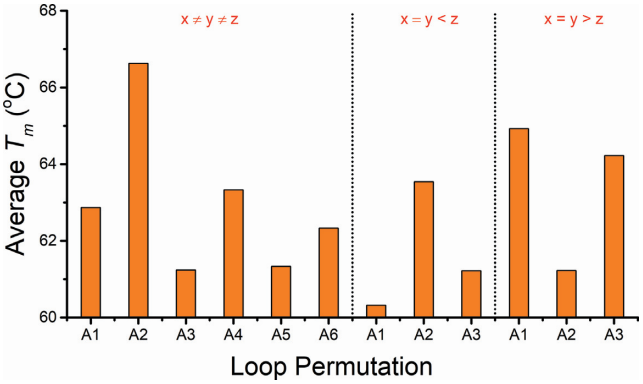


Figure 7. Thermal stabilities (T_m) are dependent on loop permutation. Averages T_m and corresponding standard deviation for each loop permutation are plotted as orange bar and black line respectively. A1–A6 and A1–A3 for each sequence are presented in Table 1.

tramolecular G-quadruplexes by 143 and 153 sequences was attested by the presence of 12 unambiguous imino-proton peaks (58). Broadening and partial overlapping of ^{1}H NMR peaking prevent us from concluding on molecularity of other sequences. Non-denaturing electrophoresis experiments (at 50 μM strand concentration) allow us to conclude that 134, 143 and 153 form intramolecular G-quadruplexes, while 341, 413, 431, 135, 351, 513 and 531 form species of higher molecularities, and 314 adopts both intermolecular and intramolecular structures (Supplementary Figure S9). We then investigated whether T_m value is dependent on DNA strand concentration (Supplementary Figures S10 and S11; 1–10 μM concentration range). T_m of 134, 143, 341, 135, 153, 351 and 531 are concentration-independent, indicating of intramolecular structures (Figure 8). The higher concentrations required for nuclear magnetic resonance (NMR) and electrophoresis may explain why more sequences were prone to form intermolecular structures with these methods. Previous work reported that sequences with two short loops of one or two nucleotides adopt a parallel topology and then often form intermolecular species (39,67). Figure 8 shows that G-quadruplexes with longer loops such as 314, 413, 431, 315 and 513 can also adopt intermolecular structures. Moreover, these results indicate that loop permutation also affects the G-quadruplex molecularity.

Loop permutation matters as much as loop length and base composition

Loop length is known to play an important role on the structural properties of G-quadruplexes. Supplementary Figure S12 plots values of r and T_m as function of total loop length, which vary between 7 and 10 nts in the present study. The impact of total loop length on topology (defined by r value) is limited as shown by a poor correlation (−0.33; Supplementary Figure S12A). The slight negative linear correlation coefficient indicates that conformations are more likely to be non-parallel for longer loops. As for stability, the T_m tends to decrease with loop length, as previously described (34,36,39,59), and this correlation is clearer than that for topology (−0.59; Supplementary Figure S12B). In summary, G-quadruplexes tend to fold into stable parallel structure when total loop is shorter (34,36,39,59). However, the relatively small values of linear coefficients (−0.33 and −0.59) indicate that loop length is far from being the only factor playing a role.

For this reason, we decided to investigate whether loop permutation could affect the stability of G-quadruplexes. We observed that these permutations strongly affect stability—T_m values vary by 13.1, 7.0, 14.8 and 15.9°C for total loop lengths of 7, 8, 9 and 10 nts, respectively (Supplementary Table S2). Paradoxically, formation of parallel is facilitated by the shortest loop being central, which is also detrimental to stability (Figures 3 and 6). These findings indicate that loop permutation plays a role on stability of the G-quadruplexes.

In addition, the loop base composition is also important for G-quadruplex folding. Using adenine in lieu of all thymine in the *134* and *135* groups produced the *134A* and *135A* groups containing all adenine loops (sequence information in Supplementary Table S3). CD spectroscopies and UV-melting experiments were carried out (Supplementary Figures S13 and S14), and the values of r and T_m are presented in Supplementary Table S3. Sequences 314, 413 and 431 fold into parallel topologies, while 314A and 413A adopt hybrid conformations, and 431A sequence even forms an antiparallel structure (Table 1 and Supplementary Table S3). Sequences 135 and 351 fold into hybrid topologies, and 513 and 531 adopt parallel topologies. In contrast, 135A and 351A fold into parallel topologies, while 513A and 531A adopt hybrid structures. Aside from G-quadruplex conformations, base composition has a significant impact on stability. In comparison with *134* and *135* groups, T_m values of *134A* and *135A* groups are generally much lower. The differences of T_m values between *134* and *134A* groups, *135* and *135A* groups are 12.9–23.0°C and 10.3–27.5°C, respectively. Similar results have been reported by Mergny and Fox (41,44,68). More importantly, the sequences in *134A* and *135A* groups with different loop arrangements have significant differences in both conformations and stabilities. Parallel, hybrid and antiparallel topologies can be folded by sequences in the *134A* group (Supplementary Figure S13A), while sequences in the *135A* group adopt either parallel or hybrid topologies (Supplementary Figure S13B). ΔT_m caused by loop permutation in *134A* and *135A* groups are 11.8 and 17.0°C (Supplementary Table S3 and Figure 5), which is comparable with difference resulted from the loop base compositions. This comparison indicates that loop permutation can robustly control the structural properties of G-quadruplex like the loop base composition.

DISCUSSION

DNA intramolecular G-quadruplex formation has attracted much interest due to the possibility that this structure may form in telomeres, oncogene promoters and other biologically relevant regions of the genome. It is essential to understand the rules that govern the formation of these intramolecular structures and to determine their stabilities under near-physiological conditions. The impact of loop regions on G-quadruplex stability has been dissected from

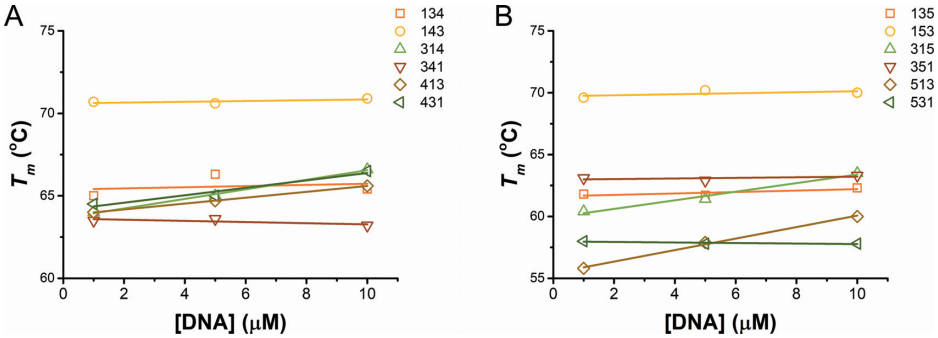


Figure 8. T_m values as a function of DNA concentration. Linear fitting of T_m versus DNA concentration of (**A**) *134* group and (**B**) *135* group were carried out. Strand concentrations were 1, 5 and 10 μM. Samples were incubated in 20 mM KPi buffer (pH 7.0) with 80 mM KCl.

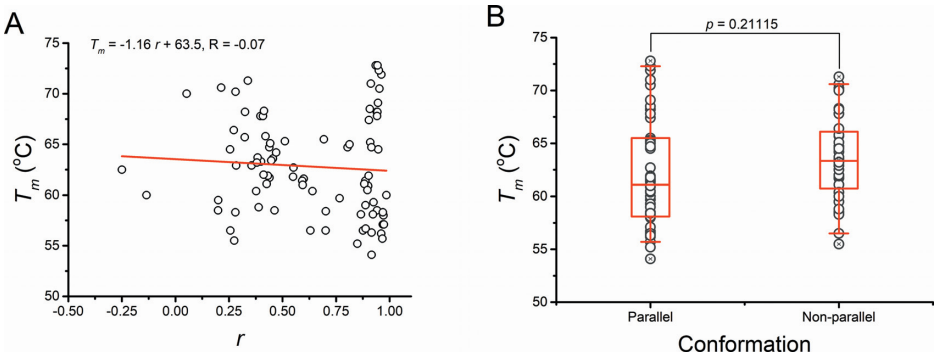


Figure 9. Lack of correlation between conformation (r) and stability (T_m). (**A**) T_m as a function of r. Fitting curves, equations and correlation coefficients obtained by linear regression method are in red colors. (**B**) Independent t-test of T_m between parallel ($1 > r \geq 0.5$) and non-parallel ($0.5 > r > 0$ and $r < 0$) topologies. *134A* and *135A* groups are not included. There is no significant relationship between conformation and stability according to both tests.

various perspectives (45,69–72) and this study complements that previous reported sequences of variable or identical loop length (but with variable base content) were compared. In the present study, we analyzed the effects of loop permutation on intramolecular quadruplex stability, keeping loop composition constant in most cases (thymine only). The work presented here offers a general view of this effect thanks to the number of sequences tested and the different loop combinations considered. Most quadruplexes studied here were stable at physiological temperature in the presence of potassium. To understand the contribution of loop permutation, we chose to study model sequences containing four tracts of three guanines; it will be interesting to determine if these results also apply to quadruplexes involving two or four quartets.

The sequences chosen for this study allowed us to minimize the contribution of hydrogen bonding and stacking interactions in the loops, such as (i) formation of base pairs or triples between loop regions and flanking residues, and then stacking on top of a terminal G-quartet (73–75), (ii) loop base itself stacking at external G-quartet (76,77) and (iii) formation of a stable secondary structures by internal loop bases, such as a double helix (78–80). Considering that sequences used here have no flanking sequence and are generally composed of only one type of base in loops, stable loop structures are unlikely to have a significant contribution. Interactions between loops and G-quartets or between internal bases of loop can affect the G-quadruplex folding. Further efforts are needed to understand the contribution of loop permutation on folding pattern and stability.

Stability is expressed here as a melting temperature; a ΔG° at 37°C would perhaps be a more biologically relevant feature. However, an accurate determination of Gibb's free energy is not always simple, especially when T_m are far above physiological temperature as enthalpies may be temperature-dependent ($\Delta C_p \neq 0$) and extrapolations are dangerous. Figure 9A shows that no correlation can be found between conformation and stability as shown by a near-zero linear correlation coefficient (−0.07). This observation is surprising, as both conformation and stability were dependent on loop sequences. Furthermore, a t-test analysis further demonstrates that T_m was not significantly different in parallel and non-parallel G-quadruplexes (Figure 9B). Some of these results were unexpected, as we anticipated that parallel structures would be more stable.

To determine whether these sequences could be biologically relevant, we performed a genome-wide search with BLAST (81). Most groups (18/21) and over half of the individual sequences (55/99) were found in the genome of diverse organisms; 12 of them were found at least once in the human genome (Supplementary Figure S15). These sequences have potential biological functions in gene regulation or could be used as therapeutic targets. In fact, given that a limited set of sequences are interrogated in this work, we can anticipate that more sequences with different loop permutation will be reported, especially taking the base varies (A/T/C) into account.

Our long-term goal is to establish rules to predict G4 stability based on primary sequence. Our concerned efforts should improve the reliability of these predictions, and several algorithms have been developed to predict the formation of stable G-quadruplexes under physiological conditions (82,83). This study constitutes a new step to understand the relationships between loop organization and thermal stabilities of intramolecular DNA quadruplexes. Our findings suggest that loop permutation is an indispensable factor in establishing rules to predict the folding pattern and stability of G-quadruplex *in vitro*. More sequences (not thymine only) will of course be needed to provide a complete picture, keeping in mind that screening all sequences is not experimentally feasible: assuming a total length of 10 nts, there are over 1 million (4^{10}) possible sequences for each group (226, 235, 244, 334…)! As the complexity increases geometrically with the length of each loop, one needs to select a subset of motifs, use degenerate sequences or combinatorial/SELEX approaches to identify quadruplexes of unusual stability.

CONCLUSION

In the present work, we compared the influences of loop permutation, length and base composition on G-quadruplex topology and stability, and found that the former one matters as much as the two others. Modulations on folding patterns and thermal stabilities vary remarkably based on the groups of sequences. The T_m difference resulting from loop permutation can be as high as 17.0°C. Sequences with a long central loop have a higher propensity to form hybrid topologies and show the highest stability. On the other hand, sequences containing a short central loop have a higher propensity to form parallel topologies and exhibit the lowest stability. These results indicate that loop permutation has a profound impact on structure and stability of G-quadruplexes, opening new perspectives in the prediction of G-quadruplex folding. Loop permutation could be biologically relevant as it exists in genomes, and it may affect the application of G-quadruplex *in vitro*.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We gratefully acknowledge Siyu Wang and Quan Shi from Dalian Institute of Chemical Physics (DICP) for their invaluable support in DSC measurements.

FUNDING

National Natural Science Foundation of China [21227801, 21503229]; Natural Science Foundation of Liaoning Province [2015020700]; Fundamental Research Funds for the Central Universities [020514380144]; SYMBIT Project, European Regional Development Fund [CZ.02.1.01/0.0/0.0/15_003/0000477]; Recruitment Program for Foreign Experts (1000plan) of China [WQ20163200397 to J.-L.M.]; Nanjing University [020514912216 to J.-L.M.]. Funding for open access charge: National Natural Science Foundation of China [21227801].
Conflict of interest statement. None declared.

REFERENCES

1. Giancola,C. and Montesarchio,D. (2017) Not unusual, just different! Chemistry, biology and applications of G-quadruplex nucleic acids. *Biochim. Biophys. Acta*, **1861**, 1201–1204.
2. Neidle,S. and Balasubramanian,S. (2006) *Quadruplex Nucleic Acids*. Royal Society of Chemistry, Cambridge.
3. Chaires,J.B. and Graves,D. (2013) *Quadruplex Nucleic Acids*. Springer, Berlin.
4. Hansel-Hertsch,R., Di Antonio,M. and Balasubramanian,S. (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.*, **18**, 279–284.
5. Chambers,V.S., Marsico,G., Boutell,J.M., Di Antonio,M., Smith,G.P. and Balasubramanian,S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.*, **33**, 877–881.
6. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
7. Tang,J., Kan,Z.Y., Yao,Y., Wang,Q., Hao,Y.H. and Tan,Z. (2008) G-quadruplex preferentially forms at the very 3′ end of vertebrate telomeric DNA. *Nucleic Acids Res.*, **36**, 1200–1208.
8. Garg,R., Aggarwal,J. and Thakkar,B. (2016) Genome-wide discovery of G-quadruplex forming sequences and their functional relevance in plants. *Sci. Rep.*, **6**, 28211–28223.
9. Ruggiero,E. and Richter,S.N. (2018) G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Res.*, **47**, 3270–3283.
10. Marusic,M., Hosnjak,L., Krafcikova,P., Poljak,M., Viglasky,V. and Plavec,J. (2017) The effect of single nucleotide polymorphisms in G-rich regions of high-risk human papillomaviruses on structural diversity of DNA. *Biochim. Biophys. Acta*, **1861**, 1229–1236.
11. Di Salvo,M., Pinatel,E., Tala,A., Fondi,M., Peano,C. and Alifano,P. (2018) G4PromFinder: an algorithm for predicting transcription promoters in GC-rich bacterial genomes based on AT-rich elements and G-quadruplex motifs. *BMC Bioinform.*, **19**, 36–46.
12. Moye,A.L., Porter,K.C., Cohen,S.B., Phan,T., Zyner,K.G., Sasaki,N., Lovrecz,G.O., Beck,J.L. and Bryan,T.M. (2015) Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.*, **6**, 7643–7654.
13. Rigo,R., Palumbo,M. and Sissi,C. (2017) G-quadruplexes in human promoters: a challenge for therapeutic applications. *Biochim. Biophys. Acta*, **1861**, 1399–1413.
14. Hansel-Hertsch,R., Beraldi,D., Lensing,S.V., Marsico,G., Zyner,K., Parry,A., Di Antonio,M., Pike,J., Kimura,H., Narita,M. *et al.* (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.*, **48**, 1267–1272.
15. Rhodes,D. and Lipps,H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627–8637.
16. McLuckie,K.I., Di Antonio,M., Zecchini,H., Xian,J., Caldas,C., Krippendorff,B.F., Tannahill,D., Lowe,C. and Balasubramanian,S. (2013) G-quadruplex DNA as a molecular target for induced synthetic lethality in cancer cells. *J. Am. Chem. Soc.*, **135**, 9640–9643.
17. Piazza,A., Boule,J.B., Lopes,J., Mingo,K., Largy,E., Teulade-Fichou,M.P. and Nicolas,A. (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **38**, 4337–4348.
18. Zhao,A., Howson,S.E., Zhao,C., Ren,J., Scott,P., Wang,C. and Qu,X. (2017) Chiral metallohelices enantioselectively target hybrid human telomeric G-quadruplex DNA. *Nucleic Acids Res.*, **45**, 5026–5035.
19. Hu,M.H., Chen,S.B., Wang,B., Ou,T.M., Gu,L.Q., Tan,J.H. and Huang,Z.S. (2017) Specific targeting of telomeric multimeric G-quadruplexes by a new triaryl-substituted imidazole. *Nucleic Acids Res.*, **45**, 1606–1618.
20. Bates,P.J., Reyes-Reyes,E.M., Malik,M.T., Murphy,E.M., O'Toole,M.G. and Trent,J.O. (2017) G-quadruplex oligonucleotide AS1411 as a cancer-targeting agent: uses and mechanisms. *Biochim. Biophys. Acta*, **1861**, 1414–1428.

21. Do,N.Q., Chung,W.J., Truong,T.H.A., Heddi,B. and Phan,A.T. (2017) G-quadruplex structure of an anti-proliferative DNA sequence. *Nucleic Acids Res.*, **45**, 7487–7493.
22. Musumeci,D., Riccardi,C. and Montesarchio,D. (2015) G-quadruplex forming oligonucleotides as anti-HIV agents. *Molecules*, **20**, 17511–17532.
23. Li,Y., Jia,G., Wang,C., Cheng,M. and Li,C. (2015) Higher-order human telomeric G-quadruplex DNA metalloenzymes enhance enantioselectivity in the Diels-Alder reaction. *ChemBioChem*, **16**, 618–624.
24. Cheng,M., Li,Y., Zhou,J., Jia,G., Lu,S.M., Yang,Y. and Li,C. (2016) Enantioselective sulfoxidation reaction catalyzed by a G-quadruplex DNA metalloenzyme. *Chem. Commun.*, **52**, 9644–9647.
25. Dey,S., Ruhl,C.L. and Jaschke,A. (2017) Catalysis of michael additions by covalently modified G-quadruplex DNA. *Chemistry*, **23**, 12162–12170.
26. Lv,L., Guo,Z., Wang,J. and Wang,E. (2012) G-quadruplex as signal transducer for biorecognition events. *Curr. Pharm. Des.*, **18**, 2076–2095.
27. Connor,A.C., Frederick,K.A., Morgan,E.J. and McGown,L.B. (2006) Insulin capture by an insulin-linked polymorphic region G-quadruplex DNA oligonucleotide. *J. Am. Chem. Soc.*, **128**, 4986–4991.
28. Zhang,D., Han,J., Li,Y., Fan,L. and Li,X. (2016) Aptamer-based K^+ sensor: process of aptamer transforming into G-quadruplex. *J. Phys. Chem. B*, **120**, 6606–6611.
29. Macaya,R.F., Schultze,P., Smith,F.W., Roe,J.A. and Feigon,J. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *P. Natl. Acad. Sci. U.S.A.*, **90**, 3745–3749.
30. Yatsunyk,L.A., Mendoza,O. and Mergny,J.L. (2014) "Nano-oddities": unusual nucleic acid assemblies for DNA-based nanostructures and nanodevices. *Acc. Chem. Res.*, **47**, 1836–1844.
31. Zhou,J., Amrane,S., Korkut,D.N., Bourdoncle,A., He,H.Z., Ma,D.L. and Mergny,J.L. (2013) Combination of i-motif and G-quadruplex structures within the same strand: formation and application. *Angew. Chem. Int. Ed.*, **52**, 7742–7746.
32. Feng,G., Luo,C., Yi,H., Yuan,L., Lin,B., Luo,X., Hu,X., Wang,H., Lei,C., Nie,Z. *et al.* (2017) DNA mimics of red fluorescent proteins (RFP) based on G-quadruplex-confined synthetic RFP chromophores. *Nucleic Acids Res.*, **45**, 10380–10392.
33. Risitano,A. and Fox,K.R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **32**, 2598–2606.
34. Hazel,P., Huppert,J., Balasubramanian,S. and Neidle,S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.*, **126**, 16405–16415.
35. Kumar,N. and Maiti,S. (2008) A thermodynamic overview of naturally occurring intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **36**, 5610–5622.
36. 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.
37. Tran,P.L., Mergny,J.L. and Alberti,P. (2011) Stability of telomeric G-quadruplexes. *Nucleic Acids Res.*, **39**, 3282–3294.
38. Agrawal,P., Lin,C., Mathad,R.I., Carver,M. and Yang,D. (2014) The major G-quadruplex formed in the human BCL-2 proximal promoter adopts a parallel structure with a 13-nt loop in K^+ solution. *J. Am. Chem. Soc.*, **136**, 1750–1753.
39. Guédin,A., Gros,J., Alberti,P. and Mergny,J.L. (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.*, **38**, 7858–7868.
40. Tippana,R., Xiao,W. and Myong,S. (2014) G-quadruplex conformation and dynamics are determined by loop length and sequence. *Nucleic Acids Res.*, **42**, 8106–8114.
41. Rachwal,P.A., Brown,T. and Fox,K.R. (2007) Sequence effects of single base loops in intramolecular quadruplex DNA. *FEBS Lett.*, **581**, 1657–1660.
42. Sattin,G., Artese,A., Nadai,M., Costa,G., Parrotta,L., Alcaro,S., Palumbo,M. and Richter,S.N. (2013) Conformation and stability of intramolecular telomeric G-quadruplexes: sequence effects in the loops. *PLoS One*, **8**, e84113.
43. Li,Y.Y. and Macgregor,R.B. Jr (2016) A thermodynamic study of adenine and thymine substitutions in the loops of the oligodeoxyribonucleotide HTel. *J. Phys. Chem. B*, **120**, 8830–8836.
44. Guédin,A., De Cian,A., Gros,J., Lacroix,L. and Mergny,J.L. (2008) Sequence effects in single-base loops for quadruplexes. *Biochimie*, **90**, 686–696.
45. Ghimire,C., Park,S., Iida,K., Yangyuoru,P., Otomo,H., Yu,Z., Nagasawa,K., Sugiyama,H. and Mao,H. (2014) Direct quantification of loop interaction and pi-pi stacking for G-quadruplex stability at the submolecular level. *J. Am. Chem. Soc.*, **136**, 15537–15544.
46. Piazza,A., Adrian,M., Samazan,F., Heddi,B., Hamon,F., Serero,A., Lopes,J., Teulade-Fichou,M.P., Phan,A.T. and Nicolas,A. (2015) Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J.*, **34**, 1718–1734.
47. Lago,S., Tosoni,E., Nadai,M., Palumbo,M. and Richter,S.N. (2017) The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. *Biochim. Biophys. Acta*, **1861**, 1371–1381.
48. Takahama,K., Sugimoto,C., Arai,S., Kurokawa,R. and Oyoshi,T. (2011) Loop lengths of G-quadruplex structures affect the G-quadruplex DNA binding selectivity of the RGG motif in Ewing's sarcoma. *Biochemistry*, **50**, 5369–5378.
49. Kumar,N., Sahoo,B., Varun,K.A., Maiti,S. and Maiti,S. (2008) Effect of loop length variation on quadruplex-Watson Crick duplex competition. *Nucleic Acids Res.*, **36**, 4433–4442.
50. Guan,A.J., Zhang,E.X., Xiang,J.F., Li,Q., Yang,Q.F., Li,L., Tang,Y.L. and Wang,M.X. (2011) Effects of loops and nucleotides in G-quadruplexes on their interaction with an azacalixarene, methylazacalix[6]pyridine. *J. Phys. Chem. B*, **115**, 12584–12590.
51. Yu,H., Zhao,C., Chen,Y., Fu,M., Ren,J. and Qu,X. (2010) DNA loop sequence as the determinant for chiral supramolecular compound G-quadruplex selectivity. *J. Med. Chem.*, **53**, 492–498.
52. Collie,G.W., Campbell,N.H. and Neidle,S. (2015) Loop flexibility in human telomeric quadruplex small-molecule complexes. *Nucleic Acids Res.*, **43**, 4785–4799.
53. Chen,J., Guo,Y., Zhou,J. and Ju,H. (2017) The effect of adenine repeats on G-quadruplex/hemin peroxidase mimicking DNAzyme activity. *Chemistry*, **23**, 4210–4215.
54. Cheng,M., Zhou,J., Jia,G., Ai,X., Mergny,J.L. and Li,C. (2017) Relations between the loop transposition of DNA G-quadruplex and the catalytic function of DNAzyme. *Biochim. Biophys. Acta*, **1861**, 1913–1920.
55. 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.
56. Del Villar-Guerra,R., Trent,J.O. and Chaires,J.B. (2017) G-quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angew. Chem. Int. Ed.*, **57**, 7171–7175.
57. Mergny,J.L., Phan,A.T. and Lacroix,L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
58. Webba da Silva,M., Trajkovski,M., Sannohe,Y., Ma'ani Hessari,N., Sugiyama,H. and Plavec,J. (2009) Design of a G-quadruplex topology through glycosidic bond angles. *Angew. Chem. Int. Ed.*, **48**, 9167–9170.
59. Arora,A. and Maiti,S. (2009) Stability and molecular recognition of quadruplexes with different loop length in the absence and presence of molecular crowding agents. *J. Phys. Chem. B*, **113**, 8784–8792.
60. Babinsky,M., Fiala,R., Kejnovska,I., Bednarova,K., Marek,R., Sagi,J., Sklenar,V. and Vorlickova,M. (2014) Loss of loop adenines alters human telomere d[AG3(TTAG3)3] quadruplex folding. *Nucleic Acids Res.*, **42**, 14031–14041.
61. Cea,V., Cipolla,L. and Sabbioneda,S. (2015) Replication of Structured DNA and its implication in epigenetic stability. *Front. Genet.*, **6**, 209–215.
62. 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-myb promoter is a critical regulator of promoter activity. *Nucleic Acids Res.*, **36**, 1755–1769.
63. Piazza,A., Cui,X., Adrian,M., Samazan,F., Heddi,B., Phan,A.T. and Nicolas,A.G. (2017) Non-Canonical G-quadruplexes cause the hCEB1 minisatellite instability in Saccharomyces cerevisiae. *Elife*, **6**, e26884.

64. Paeschke,K., Bochman,M.L., Garcia,P.D., Cejka,P., Friedman,K.L., Kowalczykowski,S.C. and Zakian,V.A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*, **497**, 458–462.
65. Mergny,J.L., Li,J., Lacroix,L., Amrane,S. and Chaires,J.B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138.
66. Largy,E., Marchand,A., Amrane,S., Gabelica,V. and Mergny,J.L. (2016) Quadruplex turncoats: cation-dependent folding and stability of quadruplex-DNA double switches. *J. Am. Chem. Soc.*, **138**, 2780–2792.
67. Do,N.Q., Lim,K.W., Teo,M.H., Heddi,B. and Phan,A.T. (2011) Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res.*, **39**, 9448–9457.
68. Guédin,A., Alberti,P. and Mergny,J.L. (2009) Stability of intramolecular quadruplexes: sequence effects in the central loop. *Nucleic Acids Res.*, **37**, 5559–5567.
69. Takahashi,S. and Sugimoto,N. (2017) Volumetric contributions of loop regions of G-quadruplex DNA to the formation of the tertiary structure. *Biophys. Chem.*, **231**, 146–154.
70. Wei,D., Husby,J. and Neidle,S. (2015) Flexibility and structural conservation in a c-KIT G-quadruplex. *Nucleic Acids Res.*, **43**, 629–644.
71. Fujimoto,T., Nakano,S., Sugimoto,N. and Miyoshi,D. (2013) Thermodynamics-hydration relationships within loops that affect G-quadruplexes under molecular crowding conditions. *J. Phys. Chem. B*, **117**, 963–972.
72. Olsen,C.M., Lee,H.T. and Marky,L.A. (2009) Unfolding thermodynamics of intramolecular G-quadruplexes: base sequence contributions of the loops. *J. Phys. Chem. B*, **113**, 2587–2595.
73. Zhang,Z., Dai,J., Veliath,E., Jones,R.A. and Yang,D. (2010) Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K^+ solution: insights into the interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Res.*, **38**, 1009–1021.
74. Lim,K.W., Amrane,S., Bouaziz,S., Xu,W., Mu,Y., Patel,D.J., Luu,K.N. and Phan,A.T. (2009) Structure of the human telomere in K^+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.*, **131**, 4301–4309.
75. Dai,J., Carver,M., Punchihewa,C., Jones,R.A. and Yang,D. (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.
76. Agrawal,P., Hatzakis,E., Guo,K., Carver,M. and Yang,D. (2013) Solution structure of the major G-quadruplex formed in the human VEGF promoter in K^+: insights into loop interactions of the parallel G-quadruplexes. *Nucleic Acids Res.*, **41**, 10584–10592.
77. Podbevsek,P., Sket,P. and Plavec,J. (2008) Stacking and not solely topology of T3 loops controls rigidity and ammonium ion movement within d(G4T3G4)2 G-quadruplex. *J. Am. Chem. Soc.*, **130**, 14287–14293.
78. Onel,B., Carver,M., Wu,G., Timonina,D., Kalarn,S., Larriva,M. and Yang,D. (2016) A new G-quadruplex with hairpin loop immediately upstream of the human BCL2 P1 promoter modulates transcription. *J. Am. Chem. Soc.*, **138**, 2563–2570.
79. Lim,K.W., Nguyen,T.Q. and Phan,A.T. (2014) Joining of multiple duplex stems at a single quadruplex loop. *J. Am. Chem. Soc.*, **136**, 17969–17973.
80. Lim,K.W. and Phan,A.T. (2013) Structural basis of DNA quadruplex-duplex junction formation. *Angew. Chem. Int. Ed.*, **52**, 8566–8569.
81. Boratyn,G.M., Camacho,C., Cooper,P.S., Coulouris,G., Fong,A., Ma,N., Madden,T.L., Matten,W.T., McGinnis,S.D., Merezhuk,Y. *et al.* (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.*, **41**, W29–W33.
82. Sahakyan,A.B., Chambers,V.S., Marsico,G., Santner,T., Di Antonio,M. and Balasubramanian,S. (2017) Machine learning model for sequence-driven DNA G-quadruplex formation. *Sci. Rep.*, **7**, 14535–14545.
83. Stegle,O., Payet,L., Mergny,J.L., MacKay,D.J. and Huppert,J.L. (2009) Predicting and understanding the stability of G-quadruplexes. *Bioinformatics*, **25**, i374–i382.