7858–7868 *Nucleic Acids Research, 2010, Vol. 38, No. 21* *Published online 26 July 2010*
doi:10.1093/nar/gkq639

How long is too long? Effects of loop size on G-quadruplex stability

Aurore Guédin[1,2], Julien Gros[2], Patrizia Alberti[2] and Jean-Louis Mergny[1,2,*]

[1]INSERM, U869, European Institute of Chemistry and Biology, Bordeaux 2 University, 2 rue Robert Escarpit, Pessac F-33607 and [2]INSERM, U565, Acides nucléiques: dynamique, ciblage et fonctions biologiques, Muséum National d'Histoire Naturelle (MNHN) USM503, CNRS, UMR7196, Département de ''Régulations, développement et diversité moléculaire'', 43 rue Cuvier, CP26, Paris Cedex 5, F-75231, France

Received April 29, 2010; Revised July 2, 2010; Accepted July 3, 2010

ABSTRACT

We compared here 80 different sequences containing four tracts of three guanines with loops of variable length (between 1 and 15 bases for unmodified sequences, up to 30 for fluorescently labeled oligonucleotides). All sequences were capable of forming stable quadruplexes, with T_m above physiological temperature in most cases. Unsurprisingly, the melting temperature was systematically lower in sodium than in potassium but the difference between both ionic conditions varied between 1 and >39°C (average difference: 18.3°C). Depending on the sequence context, and especially for G4 sequences involving two very short loops, the third one may be very long without compromising the stability of the quadruplex. A strong inverse correlation between total loop length and T_m was found in K^+: each added base leads to a 2°C drop in T_m or ~0.3 kcal/mol loss in $\Delta G°$. The trend was less clear in Na^+, with a longer than expected optimal loop length (up to 5 nt). This study will therefore extend the sequence repertoire of quadruplex-prone sequences, arguing for a modification of the widely used consensus (maximal loop size of 7 bases).

INTRODUCTION

G-quadruplexes are a family of nucleic acid secondary structures stabilized by G-quartets that form in the presence of cations (1–5). These structures may be involved in key biological processes and recent articles highlight this notion (6–13). Four (or more) tracks of two or more guanines are required to form an intramolecular structure. Those are connected by three—sometimes four, see ref. (14)—loops of variable size and sequence. Previous works studying intramolecular quadruplexes formed with a central loop of 3–7 bases (15) or more (16) support the important role played by the nature and length of loops in quadruplex stability (15,17–25). To better understand this contribution, we chose to compare model sequences containing four tracts of three guanines. In this article, we investigated sequence variants, involving thymidine-only or thymidine-rich loops. This leads to a general sequence of the type GGG**L_1**GGG**L_2**GGG**L_3**GGG where L_1, L_2 and L_3 may either be T_2A or a loop composed solely of 1–30 thymines. Over 80 different sequences were tested; they are presented in Tables 1 and 2 and include 12 fluorescently labeled oligonucleotides for FRET analysis. This study complements our previous report in which we compared sequences of identical loop length (3 nt) but with variable base content (23).

MATERIALS AND METHODS

Nomenclature, synthesis and purification of oligonucleotide sequences

Oligonucleotides were synthesized by Eurogentec (Seraing, Belgium). Concentrations of all oligodeoxynucleotides were estimated using extinction coefficients provided by the manufacturer and calculated with a nearest neighbor model (26).

Sequences are given in the 5′- to 3′-direction and loops are numbered according to this convention:

- In the 5′-GGG**L_1**GGG**L_2**GGG**L_3**GGG-3′ sequence, L_1 refers to the 5′-side loop, L_3 to the 3′-side loop, while L_2 corresponds to the central loop.
- The three-digit oligonucleotide name **xyz** indicates loop lengths in L_1, L_2 and L_3, respectively. For example **353** corresponds to the GGG**TTT**GGG**TTT** **TT**GGG**TTT**GGG sequence. Bases in the loops are in bold/underlined characters.
- When loop length was >9 nt, letters were used: **B** = 12, **E** = 15, **K** = 21, **S** = 30. Hence, **1B1** corresponds

*To whom correspondence should be addressed. Tel: +33 5 4000 30 22; Fax: +33 5 4000 30 04; Email: jean-louis.mergny@inserm.fr

Table 1. Sequence of the oligonucleotide used for UV melting

(Length) Name[a]	Sequence (5′→3′)								T_m Na^+ (°C)	T_m K^+ (°C)
		L_1		L_2		L_3				
T-var-T										
(16) 121	GGG	**T**	GGG	**TT**	GGG	**T**	GGG		(51)	84.0
(17) 131	GGG	**T**	GGG	**TTT**	GGG	**T**	GGG		45.5	81.8
(18) 141	GGG	**T**	GGG	**TTTT**	GGG	**T**	GGG		43.0	76.5
(19) 151	GGG	**T**	GGG	**TTTTT**	GGG	**T**	GGG		41.5	73.5
(20) 161	GGG	**T**	GGG	**TTTTTT**	GGG	**T**	GGG		40.3	70.5
(21) 171	GGG	**T**	GGG	**TTTTTTT**	GGG	**T**	GGG		37.0	68.5
(23) 191	GGG	**T**	GGG	**TTTTTTTTT**	GGG	**T**	GGG		32.0	67.0
(26) 1B1	GGG	**T**	GGG	**TTTTTTTTTTT**	GGG	**T**	GGG		26.1	63.0
(29) 1E1	GGG	**T**	GGG	**T_{15}**	GGG	**T**	GGG		24.8	62.5
TTT-var-TTT										
(19) 313	GGG	**TTT**	GGG	**T**	GGG	**TTT**	GGG		(49.5)	(65.5)
(20) 323	GGG	**TTT**	GGG	**TT**	GGG	**TTT**	GGG		(46.1)	(62)
(21) 333	GGG	**TTT**	GGG	**TTT**	GGG	**TTT**	GGG		51.0	65.3
(22) 343	GGG	**TTT**	GGG	**TTTT**	GGG	**TTT**	GGG		56.5	66.1
(23) 353	GGG	**TTT**	GGG	**TTTTT**	GGG	**TTT**	GGG		54.5	65.0
(23) 363	GGG	**TTT**	GGG	**TTTTTT**	GGG	**TTT**	GGG		54.5	61.8
(25) 373	GGG	**TTT**	GGG	**TTTTTTT**	GGG	**TTT**	GGG		nd	nd
(27) 393	GGG	**TTT**	GGG	**TTTTTTTTT**	GGG	**TTT**	GGG		45.5	53.9
(30) 3B3	GGG	**TTT**	GGG	**TTTTTTTTTTT**	GGG	**TTT**	GGG		39.3	49.0
(33) 3E3	GGG	**TTT**	GGG	**T_{15}**	GGG	**TTT**	GGG		36.5	45.0
TTA-var-TTA										
(19) H313	GGG	**TTA**	GGG	**T**	GGG	**TTA**	GGG		(45.0)	(62.0)
(20) H323	GGG	**TTA**	GGG	**TT**	GGG	**TTA**	GGG		(54.3)	(61.3)
(21) H333	GGG	**TTA**	GGG	**TTT**	GGG	**TTA**	GGG		56.6	65.3
(22) H343	GGG	**TTA**	GGG	**TTTT**	GGG	**TTA**	GGG		62.5	67.8
(23) H353	GGG	**TTA**	GGG	**TTTTT**	GGG	**TTA**	GGG		61.8	65.5
(24) H363	GGG	**TTA**	GGG	**TTTTTT**	GGG	**TTA**	GGG		60.8	62.3
(25) H373	GGG	**TTA**	GGG	**TTTTTTTT**	GGG	**TTA**	GGG		57.5	59.0
(27) H393	GGG	**TTA**	GGG	**TTTTTTTTTT**	GGG	**TTA**	GGG		52.8	54.4
(30) H3B3	GGG	**TTA**	GGG	**TTTTTTTTTTTT**	GGG	**TTA**	GGG		46.8	48.0
var-TTT-TTT										
(19) 133	GGG	**T**	GGG	**TTT**	GGG	**TTT**	GGG		45.8	67.2
(20) 233	GGG	**TT**	GGG	**TTT**	GGG	**TTT**	GGG		50.7	65.5
(21) 333	GGG	**TTT**	GGG	**TTT**	GGG	**TTT**	GGG		51.0	65.3
(22) 433	GGG	**TTTT**	GGG	**TTT**	GGG	**TTT**	GGG		44.4	62.0
(23) 533	GGG	**TTTTT**	GGG	**TTT**	GGG	**TTT**	GGG		40.7	58.5
(24) 633	GGG	**TTTTTT**	GGG	**TTT**	GGG	**TTT**	GGG		38.2	56.9
(25) 733	GGG	**TTTTTTT**	GGG	**TTT**	GGG	**TTT**	GGG		36.7	56.3
(27) 933	GGG	**TTTTTTTTT**	GGG	**TTT**	GGG	**TTT**	GGG		33.9	53.3
(30) B33	GGG	**TTTTTTTTTTT**	GGG	**TTT**	GGG	**TTT**	GGG		31.0	51.3
(33) E33	GGG	**T_{15}**	GGG	**TTT**	GGG	**TTT**	GGG		28.0	49.5
TTT-TTT-var										
(19) 331	GGG	**TTT**	GGG	**TTT**	GGG	**T**	GGG		46.2	(68.3)
(20) 332	GGG	**TTT**	GGG	**TTT**	GGG	**TT**	GGG		51.8	(67.3)
(21) 333	GGG	**TTT**	GGG	**TTT**	GGG	**TTT**	GGG		51.0	65.3
(22) 334	GGG	**TTT**	GGG	**TTT**	GGG	**TTTT**	GGG		44.5	62.3
(23) 335	GGG	**TTT**	GGG	**TTT**	GGG	**TTTTT**	GGG		39.5	59.9
(24) 336	GGG	**TTT**	GGG	**TTT**	GGG	**TTTTTT**	GGG		37.0	57.8
(25) 337	GGG	**TTT**	GGG	**TTT**	GGG	**TTTTTTT**	GGG		35.5	55.8
(27) 339	GGG	**TTT**	GGG	**TTT**	GGG	**TTTTTTTTT**	GGG		30.5	53.3
(30) 33B	GGG	**TTT**	GGG	**TTT**	GGG	**TTTTTTTTTTT**	GGG		27.8	51.0
(33) 33E	GGG	**TTT**	GGG	**TTT**	GGG	**T_{15}**	GGG		25.5	49.5
var-TTT-var										
(17) 131	GGG	**T**	GGG	**TTT**	GGG	**T**	GGG		45.5	81.8
(19) 232	GGG	**TT**	GGG	**TTT**	GGG	**TT**	GGG		53.8	(68.8)
(21) 333	GGG	**TTT**	GGG	**TTT**	GGG	**TTT**	GGG		51.0	65.3
(23) 434	GGG	**TTTT**	GGG	**TTT**	GGG	**TTTT**	GGG		40.8	59.4
(25) 535	GGG	**TTTTT**	GGG	**TTT**	GGG	**TTTTT**	GGG		34.3	53.7
(27) 636	GGG	**TTTTTT**	GGG	**TTT**	GGG	**TTTTTT**	GGG		31.5	51.0
(29) 737	GGG	**TTTTTTT**	GGG	**TTT**	GGG	**TTTTTTT**	GGG		26.0	46.9
(33) 939	GGG	**T_9**	GGG	**TTT**	GGG	**T_9**	GGG		18.5	41.5
(39) B3B	GGG	**T_{12}**	GGG	**TTT**	GGG	**T_{12}**	GGG		13.0	33.0[b]
(45) E3E	GGG	**T_{15}**	GGG	**TTT**	GGG	**T_{15}**	GGG		11.0	31.0[b]

(continued)

Table 1. Continued

(Length) Name[a]	Sequence (5′→3′)							T_m Na^+ (°C)	T_m K^+ (°C)
		L_1		L_2		L_3			
var-var-TTT									
(17) 113	GGG	**T**	GGG	**T**	GGG	**TTT**	GGG	46.2[b]	82.0
(19) 223	GGG	**TT**	GGG	**TT**	GGG	**TTT**	GGG	46.0	67.5
(21) 333	GGG	**TTT**	GGG	**TTT**	GGG	**TTT**	GGG	51.0	65.3
(23) 443	GGG	**TTTT**	GGG	**TTTT**	GGG	**TTT**	GGG	51.5	64.5
(25) 553	GGG	**TTTTT**	GGG	**TTTTT**	GGG	**TTT**	GGG	47.3	60.5
(27) 663	GGG	**TTTTTT**	GGG	**TTTTTT**	GGG	**TTT**	GGG	42.7	55.0
(29) 773	GGG	**TTTTTTT**	GGG	**TTTTTTT**	GGG	**TTT**	GGG	37.5	50.0
(33) 993	GGG	**T_9**	GGG	**T_9**	GGG	**TTT**	GGG	29.0	42.0
TTT-var-var									
(17) 311	GGG	**TTT**	GGG	**T**	GGG	**T**	GGG	(42.8)	(82.0)
(19) 322	GGG	**TTT**	GGG	**TT**	GGG	**TT**	GGG	46.0	(67.8)
(21) 333	GGG	**TTT**	GGG	**TTT**	GGG	**TTT**	GGG	51.0	65.3
(23) 344	GGG	**TTT**	GGG	**TTTT**	GGG	**TTTT**	GGG	50.5	60.8
(25) 355	GGG	**TTT**	GGG	**TTTTT**	GGG	**TTTTT**	GGG	43.5	56.3
(27) 366	GGG	**TTT**	GGG	**TTTTTT**	GGG	**TTTTTT**	GGG	38.0	49.9
(29) 377	GGG	**TTT**	GGG	**TTTTTTT**	GGG	**TTTTTTT**	GGG	32.8	45.0
(33) 399	GGG	**TTT**	GGG	**T_9**	GGG	**T_9**	GGG	22.0	38.3
var-var-var									
(18) 222	GGG	**TT**	GGG	**TT**	GGG	**TT**	GGG	45.5	(73.8)
(21) 333	GGG	**TTT**	GGG	**TTT**	GGG	**TTT**	GGG	51.0	65.3
(24) 444	GGG	**TTTT**	GGG	**TTTT**	GGG	**TTTT**	GGG	45.8	57.5
(27) 555	GGG	**TTTTT**	GGG	**TTTTT**	GGG	**TTTTT**	GGG	35.1	50.5
(30) 666	GGG	**TTTTTT**	GGG	**TTTTTT**	GGG	**TTTTTT**	GGG	27.6	42.3
(33) 777	GGG	**T_7**	GGG	**T_7**	GGG	**T_7**	GGG	18.0	35.9[b]
(39) 999	GGG	**T_9**	GGG	**T_9**	GGG	**T_9**	GGG	10.0	30.0[b]

[a]Length (in nucleotides) within parentheses is followed by the name of the sequence. See 'Materials and Methods' section for the definition of the convention used here. T_m values between parentheses correspond to situation in which significant formation of intermolecular complexes is either suspected or demonstrated. T_m within brackets: strong evidence for contributing intermolecular species. [b]slight hysteresis.

Table 2. Sequence of the oligonucleotide used for FRET melting

Name	Sequence (5′→ 3′)							T_m Na^+ (°C)	T_m K^+ (°C)
		L_1		L_2		L_3			
f13lt	TGGG	**T**	GGG	**TTT**	GGG	**T**	GGGT	56.6	75.1
f16lt	TGGG	**T**	GGG	**TTTTT**	GGG	**T**	GGGT	50.5	68.0
f19lt	TGGG	**T**	GGG	**TTTTTTTTT**	GGG	**T**	GGGT	50.1	64.8
f1E1t	TGGG	**T**	GGG	**T_{15}**	GGG	**T**	GGGT	47.0	61.6
f1K1t	TGGG	**T**	GGG	**T_{21}**	GGG	**T**	GGGT	45.1	59.9
f1S1t	TGGG	**T**	GGG	**T_{30}**	GGG	**T**	GGGT	–[a]	–[a]
f333t	TGGG	**TTT**	GGG	**TTT**	GGG	**TTT**	GGGT	49.2	55.4
f363t	TGGG	**TTT**	GGG	**TTTTT**	GGG	**TTT**	GGGT	46.9	49.7
f393t	TGGG	**TTT**	GGG	**TTTTTTTTT**	GGG	**TTT**	GGGT	48.4	48.1
f3E3t	TGGG	**TTT**	GGG	**T_{15}**	GGG	**TTT**	GGGT	46.8	46.3
f3K3t	TGGG	**TTT**	GGG	**T_{21}**	GGG	**TTT**	GGGT	47.3	46.0
f3S3t	TGGG	**TTT**	GGG	**T_{30}**	GGG	**TTT**	GGGT	44.0	–[a]

Lower case letters (f and t) refer to 5′ FAM and 3′ TAMRA, respectively. Note that the presence of the fluorescent labels and different experimental conditions prevent a direct comparison with T_m values found by UV-absorbance experiments in Table 1.
[a]weird / low quality melting profile.

to the GGGTGGGTTTTTTTTTTTTGGGTGGG sequence. For the family of sequences in which the L_1 and L_3 loops are not solely composed of thymines but of the TTA motif (Table 1), the prefix letter **H** was used to distinguish them from the other sequences.

- For FRET melting experiments, lower case 'f' and 't' letters refer to FAM in 5′ (fluorescent donor for FRET) and tetramethylrhodamine in 3′ (TAMRA, acceptor for FRET), respectively.
- The different sequence families are defined in Table 1. For example T-var-T correspond to sequences in which the first and last loops are composed of a single thymine while the central loop may vary ('var').

Absorbance and circular dichroism measurements

Melting experiments were conducted as described earlier (23,27). Denaturation was followed by recording the absorbance at 240 or 295 nm (28,29). Our reference

conditions for this study were 10 mM Lithium cacodylate pH 7.2 supplemented with 100 mM NaCl or KCl, abbreviated hereafter K^+ or Na^+ conditions. The sequences were tested at 4 μM strand concentration. The intramolecular formation of the G-quadruplexes was evaluated by varying concentration in the 1–25 μM range or by using FRET melting experiments (see below) (30). Circular Dichroism (CD) spectra were recorded on a JASCO-810 spectropolarimeter using a 1-cm path length quartz cuvette as described earlier (27). TDS spectra were obtained by subtracting the absorbance spectra recorded below the observed transition from the absorbance at high temperature (31).

FRET melting

FRET melting experiments were conducted with double-labeled dyes (32) (Table 2) using a MX3000 realtime PCR machine (gain set at 1). Oligonucleotides were generally tested at 0.2 μM strand concentration in 10 mM lithium cacodylate buffer supplemented with 100 mM NaCl or 90 mM LiCl + 10 mM KCl, as for classical FRET melting experiments with ligands (33).

Gel electrophoresis

For some experiments, formation of G4-DNA was confirmed by non-denaturing PAGE (23,27). Prior to the incubation, the DNA samples were denatured at 90°C for 5 min and slowly cooled to room temperature (2 h). Oligonucleotides were then incubated at 30 μM strand concentration in 10 mM Tris–HCl pH 7.5 buffer with 100 mM NaCl or KCl. Ten percent sucrose was added just before loading. The gel was prepared at 12% acrylamide supplemented with 20 mM of the corresponding salt, run at 23°C and revealed by UV-Shadow. One should note that the migration of the dT_n oligonucleotides does not necessarily correspond to single-strands (34): these oligonucleotides were simply chosen here to provide an internal migration standard.

RESULTS

Evidence for quadruplex formation

We use a set of well established methods in order to demonstrate that the sequences we are studying are actually G-quadruplex structures.

Thermal Differential Spectra (TDS) exhibit the typical pattern of a G-quadruplex structure with two positive maxima at 240 and 275 nm and a negative minimum around 295 nm (28,31,35) (examples shown in Figures 1A–B). These spectra are qualitatively similar and may unambiguously be attributed to G4 structures. CD results are also in agreement with the formation of quadruplexes but different types are observed (36,37) (examples shown in Figures 1C–D). Within a family of sequences, the shortest representatives such as **131** (in KCl and NaCl) and **232** (in KCl only) tend to adopt a 'Type I' spectra, with a positive peak around 260 nm. Sequences of intermediate length exhibit a 'Type II' or 'Hybrid' spectra, clearly indicative of a quadruplex fold, but with a different topology. Finally, the longest sequences **B3B** and **E3E** have a relatively ambiguous CD signature which probably results from a significant contribution of two very long loops to the ellipticity. For these sequences, attribution of a quadruplex fold relies solely on the TDS spectra.

During absorbance UV-melting experiments, these sequences exhibit an inverted transition at 295 nm (Figure 1E–F). As for most quadruplex-forming sequences studied so far, stability was dependent on the nature of the monocation. As expected, stability of the quadruplex was lower in NaCl than in KCl (Table 1). This difference between sodium and potassium conditions is highly variable, with ΔT_m between +1.2 and +39.2°C (average +18.3°C, see below). Finally, non-denaturing gel electrophoresis indicated that most oligomers migrated in agreement with the formation of a predominant intramolecular structure (see below). These results indicate that all sequences studied here form quadruplexes.

Intramolecular versus intermolecular structures

For most of the sequences presented in this study, a concentration-independent melting temperature determined by UV-absorbance (strand concentration range 1–30 μM) confirmed that the folding process was intramolecular (Figure 2). We found evidence for intermolecular association in few cases. Among the examples shown in this figure, one should notice that oligonucleotide **121** exhibited a weak but significant dependency of the T_m on strand concentration (+5°C over a 10× or more concentration range). A few oligonucleotides were also analyzed by FRET melting (Table 2 and data not shown), with a different strand concentration range (from 50 to 1000 nM, signal saturation preventing us from analyzing data at higher strand concentrations). We observed a concentration independent T_m, arguing against the formation of a significant amount of intermolecular structures.

Non-denaturing gel electrophoresis allowed us to evidence interesting different behaviors in sodium and potassium. We chose to reveal the gel using UV-shadow, a method that does not require any labeling of any kind and that relies solely on the absorbance of the nucleic acid in the far UV region (254 nm). This absorbance is very weakly dependent on the DNA conformation [10% hypo/hyperchromicity at most (31)]. Hence, band intensity accurately reflects species abundancy, in contrast with all other staining methods we tested (silver or fluorescence staining). The drawback of UV-shadow is its poor sensitivity, forcing us to use relatively high oligonucleotide concentration (30 μM; i.e. 7.5-fold higher than most UV melting experiments). In the examples provided in Figure 3 one may observe that a single band is observed in most sequences, with a migration in agreement with the formation of intramolecular complexes, as judged by molecular size markers such as double-stranded sequences (dx9 and dx12) or oligothymidylate repeats (right lanes). Notable exceptions are the shortest sequences for which a slower migration is observed indicative of a completely different

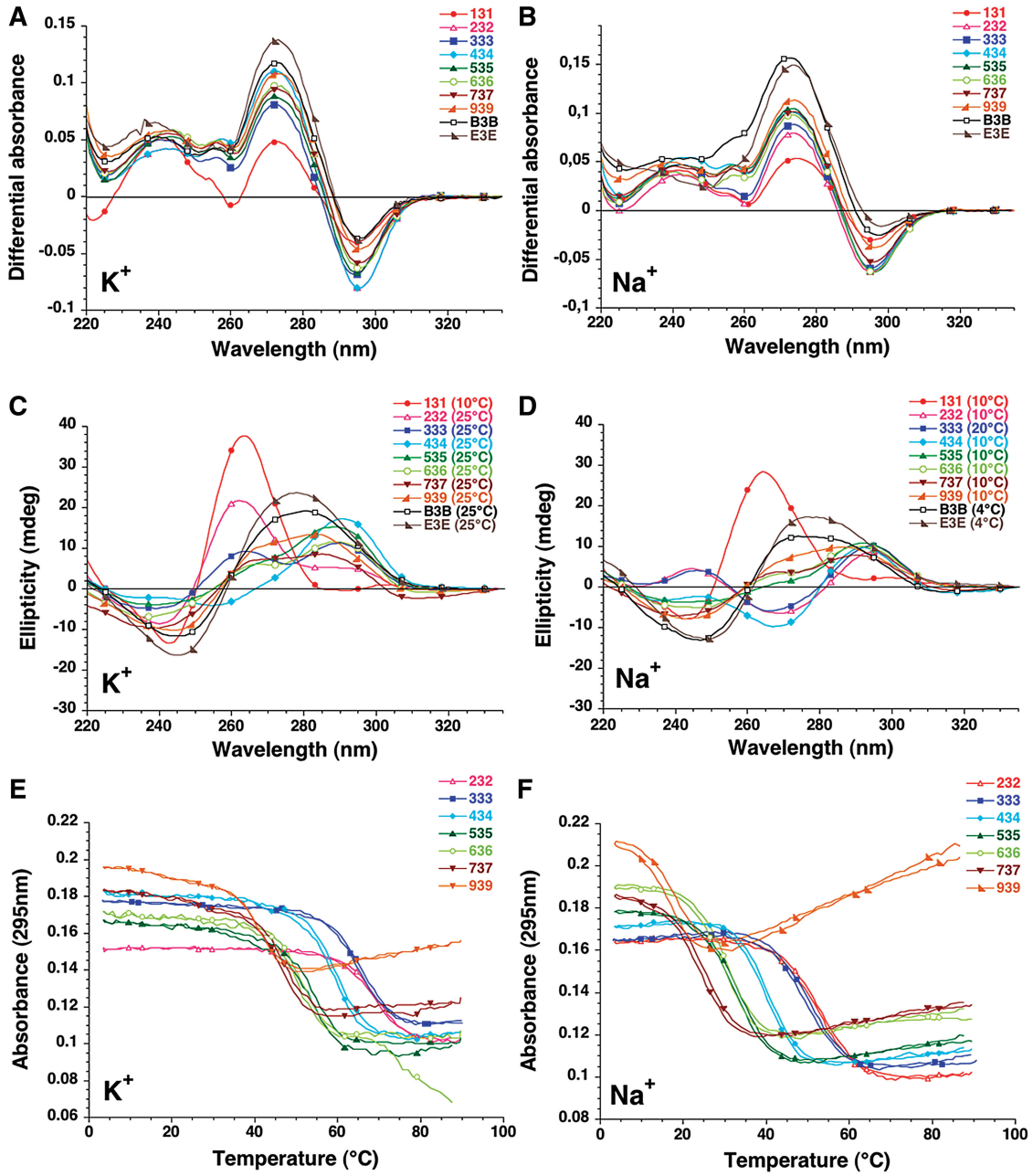

Figure 1. TDS, CD and UV melting experiments. Oligonucleotides from the var-TTT-var series are shown for illustration. Results obtained with the other series are shown as supplementary information (Supplementary Figures S1–S8). The sequences were tested at 4 μM strand concentration. Left panels in 100 mM KCl; Right panels in 100 mM NaCl. Both buffers contained 10 mM lithium cacodylate at pH 7.2. (**A–B**) Examples of Thermal difference spectra. Thermal difference spectra result from the difference between the absorbance recorded at 88 ± 2°C and at 04 ± 2°C. They were recorded over the 220–335 nm wavelength range. (**C–D**) Examples of Circular dichroism spectra. CD spectra were recorded at 10–25°C (in K^+) or 4–25°C in Na^+ (to maximize quadruplex formation in the case of relatively unstable quadruplexes) on a JASCO-810 spectropolarimeter using 1 cm path length quartz cuvettes. Oligonucleotides were prepared as a 4 μM and annealed by heating to 90°C for 2 min, followed by slow cooling. (**E–F**) Examples of UV melting profiles. Absorbance at 295 nm is plotted as a function of temperature for a selection of sequences.

topology or a higher molecularity. For some samples, one may observe several bands likely corresponding to different intramolecular species: in these cases, even though, the T_m still appears as concentration-independent.

Alternatively, a single band with a slower migration could be obtained with **121** in sodium (Figure 3A), or **313** and **323** in both ionic conditions (Figure 3B–D). This behavior could be indicative of an unusually slow intramolecular structure or, more likely, an intermolecular complex corresponding to dimers and tetramers of the oligonucleotide. Overall, as shown by the numbers in parentheses in Table 1, intermolecular complexes were often significantly present for sequences with two loops of 1–2 nt. These observations should be kept in mind when considering loop size effects on T_m! Finally some oligonucleotides gave several bands,

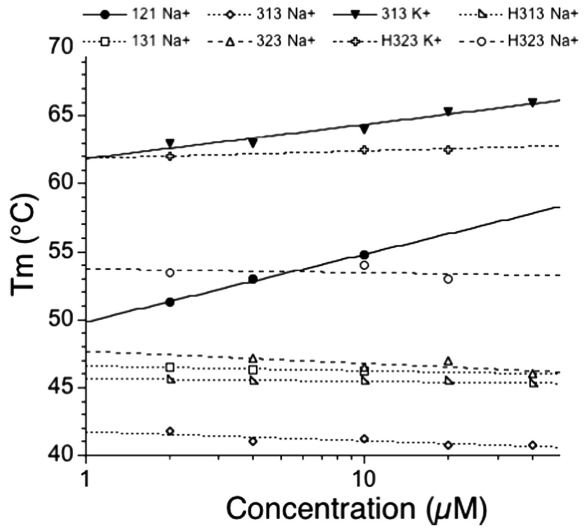

Figure 2. Effect of concentration on T_m. T_m (in °C) is shown versus strand concentration (log scale) for a few examples. For most sequences, the effect was negligible or modest, in agreement with intramolecular folding. Notable exceptions are **121** (in sodium) or **313** (in potassium) (solid lines; filled circles and triangles, respectively).

such as **1E1** (lane 15) in K^+ and Na^+, indicating that several conformations or molecularities could coexist.

Analysis of stability

Comparison of melting temperatures of the different oligonucleotides (Table 1) allowed us to reach the following conclusions:

(i) As observed earlier (16), and with the precautions mentioned above for very short sequences, T_m generally decreases when the length of a loop length is increased. This effect is striking both in potassium and in sodium (Figure 4) and valid for L_1 L_2 and L_3.

(ii) The conclusions reached for L_1 generally apply to L_3. Subtle differences may however be observed in Figure 4B and E (compare squares and diamonds). On the other hand, the profile found for the family in which L_2 is variable (circles) is different from the two others, both in potassium and in sodium.

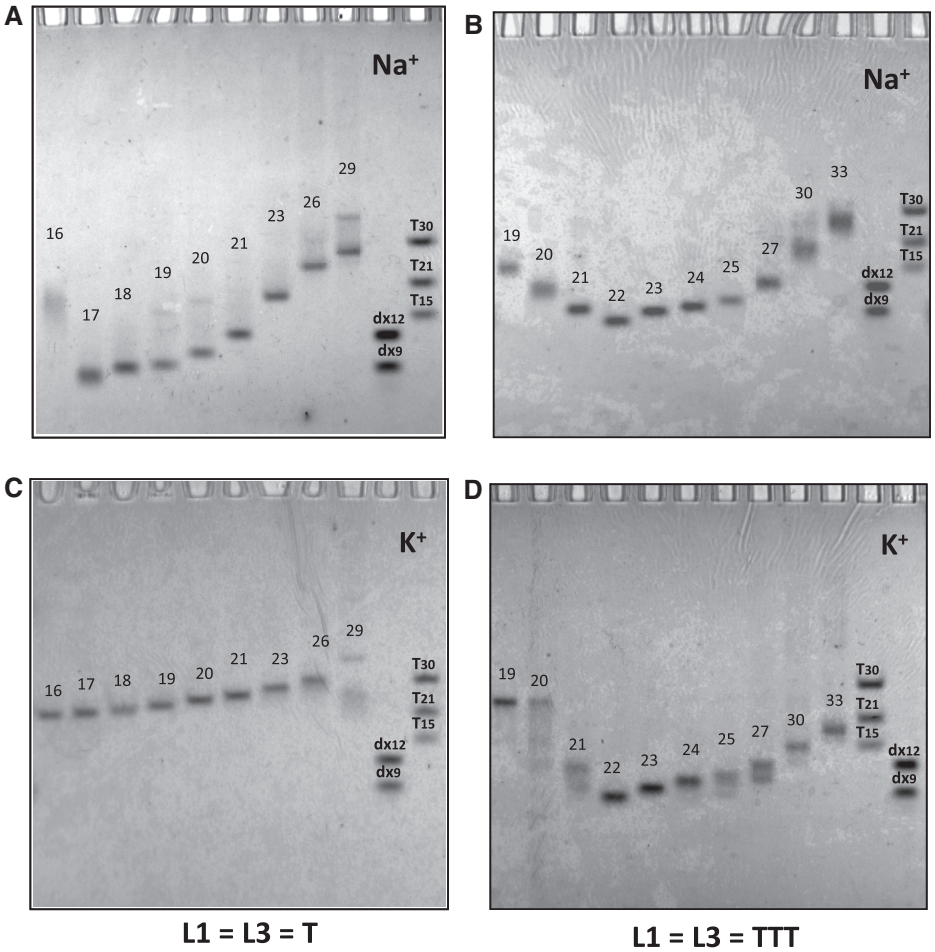

Figure 3. Behavior of the G-quadruplex forming oligonucleotides on a non-denaturing gel. Sequences were compared by non-denaturing electrophoresis and revealed by UV-shadow with 30 μM of oligonucleotide. Samples were incubated in 10 mM Tris–HCl pH 7.5 buffer with 100 mM NaCl (**A–B**) or 100 mM KCl (**C–D**). The gel was prepared at 12% acrylamide supplemented with 20 mM of the corresponding salt and run at 23°C. Migration markers are oligothymidylate markers (dT_{15}, dT_{21} or dT_{30}) or double-stranded markers (dx_9: 5′-d-GCGATACGG + 5′-d-CCGATACGC dx_{12}: 5′-d-GCGTGACTTCGG + 5′-d-CCGAAGTCACGC). Oligonucleotide length (in nucleotide) is indicated above each band for comparison purposes.

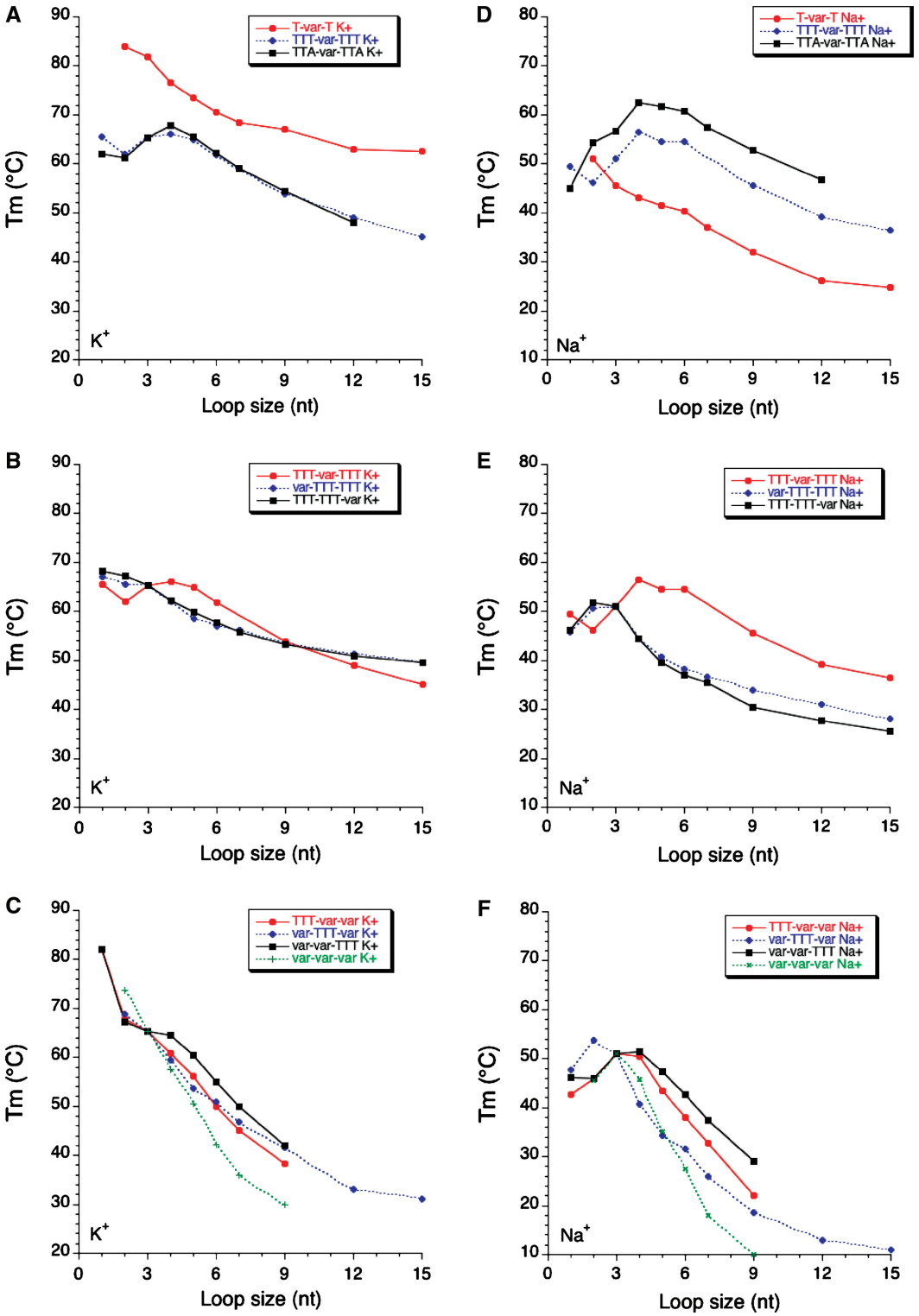

Figure 4. Sequence effects on T_m. T_m (deduced from UV-melting experiments) is shown as a function of loop size (in nucleotides) both in potassium (left panels) and sodium (right panels).

(iii) Length dependent effects are somewhat different in sodium and potassium. This is reflected by the modest correlation (R = 0.69) found between T_m values in sodium and potassium (Figure 5A). Our study reveals that this T_m difference is highly variable (ΔT_m between 1 and 39°C!). One can observe on Figure 5B that this difference is often higher for the series with the shortest loops, but this effect might be partially the result of the presence of 'contaminating' intermolecular

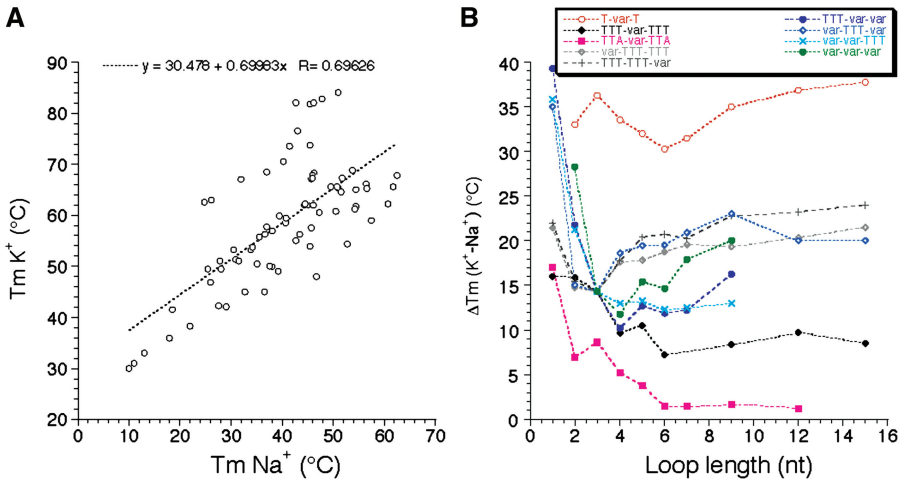

Figure 5. Sodium–Potassium difference. (**A**) Correlation between T_m values (deduced from UV-melting experiments) found in K^+ and Na^+. (**B**) ΔT_m (K^+ - Na^+) is shown as a function of loop size (in nucleotides) for each family of sequences.

complexes for those short sequences. For each series, the K^+–Na^+ difference tends to be relatively constant for loops of 7–15 nt, from 1–2°C (for TTA-var-TTA) to >30°C (for T-var-T).

(iv) While quadruplexes easily accommodate a single long loop, incorporating two or three loops of 9 nt or more becomes quickly detrimental, especially in sodium (Figure 4C and F).

(v) We tested a single family in which the loops are not entirely composed of thymines ($L_1 = L_3 =$ TTA). The comparison with ($L_1 = L_3 =$ TTT) may be made in Figure 4A and D (diamonds and square, respectively). The profiles are relatively similar but, in sodium, the adenines play a positive role on stability for all sequences with a central loop composed of at least 2 nt (up to 6–7°C).

DISCUSSION

In order to improve our understanding of quadruplex stability, we have chosen to deconvoluate sequence from loop length effects. Our previous work was dedicated to a systematic evaluation of all possible bases in 1- or 3-nt-long loops (23,38). In the present study, we analyzed the effects of loop length on intramolecular quadruplex stability, keeping loop composition constant (thymines only). Different studies have addressed this issue, but only for shorter loops (generally 7 bases or less) and in a peculiar sequence context (15). The work presented here offers a general view of this effect thanks to the number of sequence tested, the different motifs considered and the choice of two different cations. The comparison of an important number of sequences in two ionic conditions allow us several conclusions:

(i) Most quadruplex structures studied here are stable at physiological temperature, especially when potassium is present.

(ii) The difference found between T_m*s* in potassium and sodium is highly variable and thus deserves further scrutiny. A recurrent observation is that this ΔT_m is 'maximal' for sequences with two single-base loops (T-var-T) (38) (Figure 5B). This extreme difference is the result of at least two factors: (a) an unusually *high* T_m in potassium: the two short loops favor an all-parallel fold which is cation-dependent and (b) a unusually *low* T_m in sodium (compare Figures 4A and D, red circles). It is likely that the quadruplex topologies are different for these two conditions. In contrast, the difference is 'minimal' for the TTA-var-TTA family, as a result of an unusual high stability in sodium (Figure 4D, squares). As a consequence, it is difficult to predict the stability in sodium from the stability in potassium or vice versa as reflected by the poor correlation found between the results in these two ionic conditions (Figure 5A).

(iii) The most stabilizing loop length also depends on the nature of the cation: while very short loops are almost always favored in potassium (but may lead to intermolecular complexes), the situation is much less clear in sodium: for some families, the highest T_m is obtained for a loop of 5 nt (Figures 4D–E). In general, increasing the length of one loop leads to a decrease in stability (Figure 4). A destabilizing effect is *always* observed for loops of 6 nt or more. For the five different families here in which a single loop is affected, increasing its size from 6- to 9–15 nt leads to a T_m decrease summarized in Table 3. This effect is significant but relatively mild (−2°C or less per extra thymine) and tends to level off for long sequences.

(iv) This modest and regular effect prevents us from defining an upper limit for loop length compatible with quadruplex formation under physiological conditions. The commonly used upper limit

chosen for loop length in many bioinformatics studies (7 nt) is arbitrary and obviously too stringent, as $T_m > 50°C$ may be found with one loop of 12 nt or more, at least in potassium. We therefore tried to refine this upper limit. For technical reasons, to study extremely long sequences, we used a FRET melting assay as intermolecular complexes tend to form with very long oligomers: the FRET assay may be performed at a 10- to 100-fold lower strand concentration therefore limiting the contribution of intermolecular complexes. As shown in Figure 6, a relatively stable quadruplex may still be formed with a central loop of 21 or even 30 nt. Furthermore, the T_m tends to become length independent, demonstrating that one cannot even propose an upper limit for loop size *in vitro*. This observation has relevance for naturally occurring sequences with loops of variable size and contents such as the ones found in promoters or other genomic regions (38,39) and is in agreement with the lariat type quadruplex observed by Sugiyama and colleagues (40). Note, however that FRET dyes may affect folding and indeed stabilize some topologies, making a direct comparison with unmodified sequences difficult.

Table 3. Effect of increasing loop length on T_m

	In Sodium[a]	In Potassium[a]
From 6 to 9 nt	5.2 ± 1.9	6.5 ± 2.2
From 6 to12 nt	8.8 ± 3.1	11.0 ± 3.5
From 6 to 15 nt	11.6 ± 5.1	14.0 ± 3.2

[a]destabilization, in °C as compared to the T_m of the quadruplex with a TTTTTT loop. Results averaged for five different families of sequences.

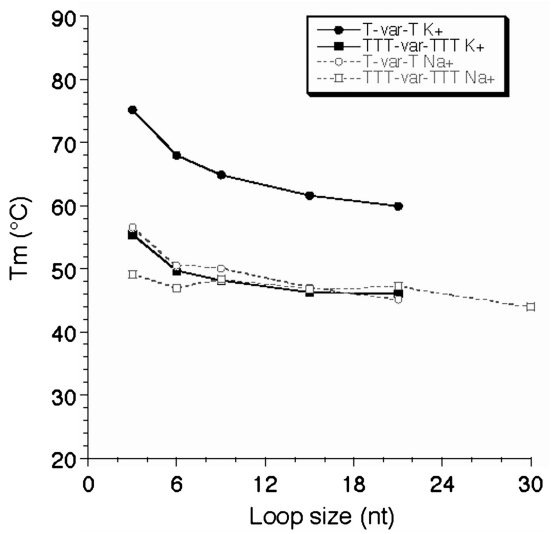

Figure 6. FRET melting experiments. T_m (deduced from FRET melting experiments) is shown as a function of loop size (in nucleotides) both in potassium (full lines) and sodium (dotted lines) for the T-var-T and TTT-var-TTT families.

(v) One may also try to correlate stability with total loop length (adding the number of nucleotides in loops L1, L2 and L3). Results are summarized in Figure 7. Interestingly, a clear negative correlation may be found in potassium while the trend is less defined in sodium, as shown by the corresponding linear correlation coefficients (0.93 and 0.69, respectively). In potassium, each extra nucleotide leads to a 2.04°C decrease in T_m which may roughly be translated in a 0.3 kcal/mol loss in $\Delta G°$ (38). We did not observe a plateau around a total loop length of ≈7 nt as found by Bugaut and Balasubramanian (17): T_m tends to reach a plateau but only for much longer loops (Figures 6 and 7A). The correlation found in Figure 7A is strikingly better than the one observed for naturally occurring sequences (24). In the later case, loop composition plays a predominant role and masks the length effects.

(vi) The effects found for shorter loops (5 nt or less) are more variable, especially in sodium. Depending on

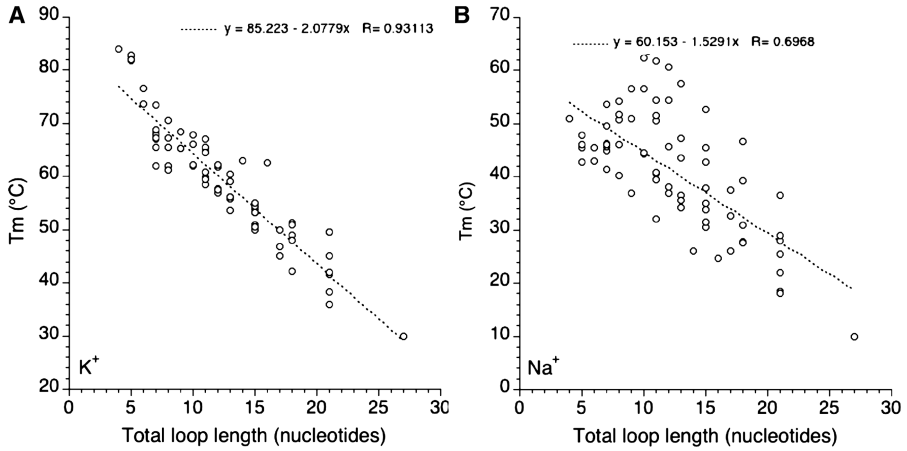

Figure 7. T_m as a function of total loop size. T_m (deduced from UV melting experiments) is plotted versus total ($L_1 + L_2 + L_3$) loop size in K^+ (**A**) and Na^+ (**B**).

the sequence context, the most stabilizing loop length may vary between 1 and 5 nt.

(vii) As expected short loops tend to allow for greater variation in loop type and consequently topology. Indeed, some of the extra species may correspond to different intramolecular topologies. Short loops also tend to favor 'contaminating' intermolecular complexes, which complicate a comparison with longer sequences. Such problem is often underappreciated but may be present even for well-known sequences (41). In these experiments, such higher order complexes may be revealed by an anomalous migration on non-denaturing gels. One way to circumvent—but not obliterate—this issue is to use a FRET melting assay and work at much lower concentrations.

(viii) The results collected here were obtained for intramolecular quadruplexes potentially involving three quartets. It will be interesting to determine if these results apply to G4 structures with two or four quartets. A systematic study of the $GGL_1GGL_2GGL_3GG$ motif is currently underway (Alberti e*t al*., manuscript in preparation). Data on such quadruplexes is scarce (42,43), despite strong biological relevance (thrombin aptamer, some telomeric motifs, etc.)

Our long term goal is to establish rules to predict the stability of intramolecular G-quadruplexes *in vitro* based on nucleotide sequences. An obvious application of this work will be to incorporate the numerical values found here in the algorithm developed by Huppert and colleagues (44). Other factors such as multiple quadruplexes (45,46) and molecular crowding (25,47,48) should also be taken into account. Altogether, our concerted efforts should improve the reliability of these predictions, especially for sequence spaces poorly covered before (very long loops). Nevertheless, one can argue that this data collection suffers from an inherent flaw: as quadruplexes adopt a variety of topologies (39,49–52) length and sequence effect are expected to be different for lateral, diagonal and chain reversal loops. Considering that several hundreds of melting profiles have been analyzed in this study (several replicates of 80 sequences done in sodium and potassium, some of them tested at different concentrations) one can easily understand that covering a wider sequence space is hardly feasible. Until recently, no rules were available to reliably predict what kind of quadruplex can be formed by a given sequence. However, recent works offer new possibilities: by inserting 8-bromo G (53,54) and ribo-G at key positions, one can control the topology of folding and hence loop type (55). Deconvolution of loop effects should be simpler in that case and lower the number of samples to be tested: efforts are now being made in that direction.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank L. Lacroix (Paris) P.L.T. Tran and A. Bourdoncle (Pessac) for helpful discussions.

FUNDING

Institut National de la Santé et de la Recherche Médicale (INSERM), the Fondation pour la Recherche Médicale (FRM); University of Bordeaux 2; *Agence Nationale de la Recherche* (G4-Toolbox ANR-09-BLAN-0355); Région Aquitaine grants. Funding for open access charge: INSERM.

Conflict of interest statement. None declared.

REFERENCES

1. Neidle,S. and Balasubramanian,S. (2006) *Quadruplex Nucleic Acids*. RSC Biomolecular Sciences, Cambridge.
2. Neidle,S. and Parkinson,G.N. (2003) The structure of telomeric DNA. *Curr. Opin. Struct. Biol.*, **13**, 275–283.
3. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
4. Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) Helix formation by guanylic acid. *Proc. Natl Acad. Sci. USA*, **48**, 2013–2018.
5. Bates,P., Mergny,J.L. and Yang,D. (2007) Quartets in G-major. The first international meeting on quadruplex DNA. *EMBO Rep.*, **8**, 1003–1010.
6. Schaffitzel,C., Berger,I., Postberg,J., Hanes,J., Lipps,H.J. and Plückthun,A. (2001) In vitro generated antibodies specific for telomeric guanine quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl Acad. Sci. USA*, **98**, 8572–8577.
7. Duquette,M.L., Handa,P., Vincent,J.A., Taylor,A.F. and Maizels,N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.*, **18**, 1618–1629.
8. Paeschke,K., Simonsson,T., Postberg,J., Rhodes,D. and Lipps,H. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.*, **12**, 847–854.
9. Granotier,C., Pennarun,G., Riou,L., Hoffschir,F., Gauthier,L.R., De Cian,A., Gomez,D., Mandine,E., Riou,J.F., Mergny,J.L. *et al*. (2005) Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res.*, **33**, 4182–4190.
10. Kruisselbrink,E., Guryev,V., Brouwer,K., Pontier,D.B., Cuppen,E. and Tijsterman,M. (2008) Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective C. elegans. *Curr. Biol.*, **18**, 900–905.
11. Pontier,D.B., Kruisselbrink,E., Guryev,V. and Tijsterman,M. (2009) Isolation of deletion alleles by G4 DNA-induced mutagenesis. *Nat. Methods*, **6**, 655–657.
12. 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.
13. Cahoon,L.A. and Seifert,H.S. (2009) An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. *Science*, **325**, 764–767.
14. Phan,A.T., Kuryavyi,V., Burge,S., Neidle,S. and Patel,D.J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.*, **129**, 4386–4392.
15. Hazel,P., Huppert,J., Balasubramanian,S. and Neidle,S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.*, **126**, 16405–16415.
16. Bourdoncle,A., Estévez-Torres,A., Gosse,C., Lacroix,L., Vekhoff,P., Le Saux,T., Jullien,L. and Mergny,J.L. (2006) Quadruplex-based molecular beacons as tunable DNA probes. *J. Am. Chem. Soc.*, **128**, 11094–11105.

17. 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.
18. Risitano,A. and Fox,K.R. (2003) Stability of intramolecular DNA quadruplexes: comparison with DNA duplexes. *Biochemistry*, **42**, 6507–6513.
19. Risitano,A. and Fox,K.R. (2003) The stability of intramolecular DNA quadruplexes with extended loops forming inter- and intra-loop duplexes. *Org. Biomol. Chem.*, **1**, 1852–1855.
20. Risitano,A. and Fox,K.R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **32**, 2598–2606.
21. Guo,Q., Lu,M. and Kallenbach,N.R. (1993) Effect of thymine tract length on the structure and stability of model telomeric sequences. *Biochemistry*, **32**, 3596–3603.
22. Cevec,M. and Plavec,J. (2005) Role of loop residues and cations on the formation and stability of dimeric DNA G-quadruplexes. *Biochemistry*, **44**, 15238–15246.
23. 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.
24. Kumar,N. and Maiti,S. (2008) A thermodynamic overview of naturally occurring intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **36**, 5610–5622.
25. 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.
26. Cantor,C.R., Warshaw,M.M. and Shapiro,H. (1970) Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, **9**, 1059–1077.
27. Mergny,J.L. and Lacroix,L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515–537.
28. Mergny,J.L., Phan,A.T. and Lacroix,L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
29. Saccà,B., Lacroix,L. and Mergny,J.L. (2005) The effect of chemical modifications on the thermal stability of different G-quadruplexes-forming oligonucleotides. *Nucleic Acids Res.*, **33**, 1182–1192.
30. Amrane,S. and Mergny,J.L. (2006) Length and pH-dependent energetics of (CCG)(n) and (CGG)(n) trinucleotide repeats. *Biochimie*, **88**, 1125–1134.
31. 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.
32. Darby,R., Sollogoub,M., McKeen,C., Brown,L., Risitano,A., Brown,N., Barton,C., Brown,T. and Fox,K. (2002) High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. *Nucleic Acids Res.*, **30**, e39.
33. De Cian,A., Guittat,L., Kaiser,M., Saccà,B., Amrane,S., Bourdoncle,A., Alberti,P., Teulade-Fichou,M.P., Lacroix,L. and Mergny,J.L. (2007) Fluorescence-based melting assays for studying quadruplex ligands. *Methods*, **42**, 183–195.
34. Kejnovska,I., Kypr,J. and Vorlickova,M. (2007) Oligo(dT) is not a correct native PAGE marker for single-stranded DNA. *Biochem. Biophys. Res. Commun.*, **353**, 776–779.
35. Petraccone,L., Erra,E., Randazzo,A. and Giancola,C. (2006) Energetic aspects of locked nucleic acids quadruplex association and dissociation. *Biopolymers*, **83**, 584–594.
36. Gray,D.M., Wen,J.D., Gray,C.W., Repges,R., Repges,C., Raabe,G. and Fleischhauer,J. (2007) Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality*, **20**, 431–440.
37. Paramasivan,S., Rujan,I. and Bolton,P.H. (2007) Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. *Methods*, **43**, 324–331.
38. 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.
39. 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.
40. Xu,Y., Sato,H., Sannohe,Y., Shinohara,K.I. and Sugiyama,H. (2008) Stable lariat formation based on a G-quadruplex scaffold. *J. Am. Chem. Soc.*, **130**, 16470–16471.
41. Fialova,M., Kypr,J. and Vorlickova,M. (2006) The thrombin binding aptamer GGTTGGTGTGGTTGG forms a bimolecular guanine tetraplex. *Biochem. Biophys. Res. Commun.*, **344**, 50–54.
42. Smirnov,I. and Shafer,R.H. (2000) Effect of loop sequence and size on DNA aptamer stability. *Biochemistry*, **39**, 1462–1468.
43. Amrane,S., Ang,R.W., Tan,Z.M., Li,C., Lim,J.K., Lim,J.M., Lim,K.W. and Phan,A.T. (2009) A novel chair-type G-quadruplex formed by a Bombyx mori telomeric sequence. *Nucleic Acids Res.*, **37**, 931–938.
44. Stegle,O., Payet,L., Mergny,J.L. and Huppert,J. (2009) Predicting and understanding the stability of G-quadruplexes. *Bioinformatics*, **25**, i374–i382.
45. Yu,H.Q., Miyoshi,D. and Sugimoto,N. (2006) Characterization of structure and stability of long telomeric DNA G-quadruplexes. *J. Am. Chem. Soc.*, **128**, 15461–15468.
46. Vorlickova,M., Chladkova,J., Kejnovska,I., Fialova,M. and Kypr,J. (2005) Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic Acids Res.*, **33**, 5851–5860.
47. Miyoshi,D., Nakao,A. and Sugimoto,N. (2002) Molecular crowding regulates the structural switch of the DNA G-quadruplex. *Biochemistry*, **41**, 15017–15024.
48. Fujimoto,T., Miyoshi,D., Tateishi-Karimata,H. and Sugimoto,N. (2009) Thermal stability and hydration state of DNA G-quadruplex regulated by loop regions. *Nucleic Acids Symp. Ser.*, 237–238.
49. Patel,D.J., Phan,A.T. and Kuryavyi,V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.*, **35**, 7429–7455.
50. Dai,J., Carver,M. and Yang,D. (2008) Polymorphism of human telomeric quadruplex structures. *Biochimie*, **90**, 1172–1183.
51. Webba da Silva,M. (2007) Geometric formalism for DNA quadruplex folding. *Chem. Eur. J.*, **13**, 9738–9745.
52. Neidle,S. (2009) The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.*, **19**, 239–250.
53. Xu,Y., Noguchi,Y. and Sugiyama,H. (2006) The new models of the human telomere d[AGGG(TTAGGG)3] in K+ solution. *Bioorg. Med. Chem.*, **14**, 5584–5591.
54. Okamoto,K., Sannohe,Y., Mashimo,T., Sugiyama,H. and Terazima,M. (2008) G-quadruplex structures of human telomere DNA examined by single molecule FRET and BrG-substitution. *Bioorg. Med. Chem.*, **16**, 6873–6879.
55. 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. Engl.*, **48**, 9167–9170.