Published online 6 July 2009

Nucleic Acids Research, 2009, Vol. 37, No. 16 ***5559–5567***
doi:10.1093/nar/gkp563

Stability of intramolecular quadruplexes: sequence effects in the central loop

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

[1]INSERM, U565, Acides nucléiques: dynamique, ciblage et fonctions biologiques, 43 rue Cuvier, CP26, Paris Cedex 05, F-75231 and [2]Muséum National d'Histoire Naturelle (MNHN) USM503, CNRS, UMR7196, Département de 'Régulations, développement et diversité moléculaire', Laboratoire des Régulations et dynamique des génomes, 43 rue Cuvier, CP26, Paris Cedex 5, F-75231, France

Received April 2, 2009; Revised June 16, 2009; Accepted June 17, 2009

ABSTRACT

Hundreds of thousands of putative quadruplex sequences have been found in the human genome. It is important to understand the rules that govern the stability of these intramolecular structures. In this report, we analysed sequence effects in a 3-base-long central loop, keeping the rest of the quadruplex unchanged. A first series of 36 different sequences were compared; they correspond to the general formula GGGTTTGGGHNHGGGTTTGGG. One clear rule emerged from the comparison of all sequence motifs: the presence of an adenine at the first position of the loop was significantly detrimental to stability. In contrast, adenines have no detrimental effect when present at the second or third position of the loop. Cytosines may either have a stabilizing or destabilizing effect depending on their position. In general, the correlation between the T_m or ΔG° in sodium and potassium was weak. To determine if these sequence effects could be generalized to different quadruplexes, specific loops were tested in different sequence contexts. Analysis of 26 extra sequences confirmed the general destabilizing effect of adenine as the first base of the loop(s). Finally, analysis of some of the sequences by microcalorimetry (DSC) confirmed the differences found between the sequence motifs.

INTRODUCTION

G-quadruplexes are unusual nucleic acids structures which result from the hydrophobic stacking of several quartets; each quartet being a planar association of four guanines held together by eight hydrogen bonds (1–5). Four (or more) tracts of two or more guanines are required to form an intramolecular structure. A cation (typically Na^+ or K^+) is located between two quartets and forms cation–dipole interactions with eight guanines. Quadruplex-prone regions abound in the human genome (6–8) and the sequence repertoire of sequences compatible with quadruplex formation is far more diverse than initially imagined (9). Intramolecular G-quadruplexes may form at telomeres, oncogene promoter sequences and other biologically relevant regions (10,11). Therefore, it is important to understand the rules that govern the formation of these intramolecular structures and to determine their stabilities. Previous works support the important role played by the nature and length of the loops in quadruplex stability (12–22) and reports suggest a strong influence of loop length on quadruplex stability. However, the stability cannot simply be deduced from total loop length, as elegantly demonstrated by Kumar and Maiti (21).

To understand the contribution of loop sequence, we chose to study model sequences involving three medium loops composed of three nucleotides. This leads to a general sequence of the type GGGTTTGGG**HNH**GGG TTTGGG and gives a total of 36 ($3 \times 4 \times 3$) different sequences, which are presented in Table 1. All sequences were able to form a quadruplex. A detailed quantitative and exhaustive analysis of all sequences allowed to obtain conclusions that we also validated in different sequence contexts.

MATERIALS AND METHODS

Nomenclature, synthesis and purification of oligonucleotide sequences

Oligonucleotides were synthesized by Eurogentec (Seraing, Belgium) at the 40-, 200- or 1000-nmol scale. Several sequences were re-synthesized; inter-lot reproducibility was excellent. Concentrations were estimated using

*To whom correspondence should be addressed. Tel: +33 1 40 79 36 89; Fax: +33 1 40 79 37 05l; Email: mergny@mnhn.fr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

extinction coefficients provided by the manufacturer and calculated with a nearest-neighbour model (23) as described before. Sequences are given in the 5′ to 3′ direction. H corresponds to A, C or T while N = A, C, T or G. We chose to exclude loop sequences that create an ambiguity in G tract definition, such as GNN or NNG, as there would be two different possibilities to select three consecutive guanines in a block of 4 G.

Absorbance measurements. Melting experiments were conducted as previously described (19,24,25) by recording the absorbance at 240 and 295 nm (26,27). Most sequences were tested at least twice at 5 μM strand concentration by two independent experimentators. As the T_m variations studied here are relatively modest, all samples were tested using strictly identical preparation protocols and buffers to maximize reproducibility. Despite all these precautions, the mean difference between two independent T_m determinations was 0.5°C. Hence, differences between two T_ms below 0.5°C were not considered significant. All transitions were reversible, indicating that the denaturation curves correspond to a true equilibrium process (28,29). The intramolecular formation of the G-quadruplexes was evaluated by varying concentration in the 5–241 μM range for subset of sequences.

Thermodynamic and statistical analysis. One can extract from the raw absorbance data the fraction of folded oligonucleotide as a function of temperature, assuming linear baselines. One may then perform a van't Hoff analysis of the melting curves using previously described procedures (30,31). However, most ΔH° values found were very similar and this analysis generally assumes that $\Delta C_p = 0$. This conventional assumption of a near-zero value for ΔC_p has been invalidated in a number of cases (32), including for quadruplexes (also see the 'DSC analysis' section).

Rather than extrapolating ΔG° values to the physiologically relevant temperature (37°C), we chose reference temperatures that are close to the average T_m values (50°C in sodium, 62°C in potassium). Hence, ΔG° values may be directly determined from the folded fraction θ at this temperature, [$\Delta G^\circ = -RT \ln (K)$] without any extrapolation or hypothesis on the temperature independency of the ΔH°. As the transition is predominantly intramolecular, $K = \theta/(1-\theta)$, in which θ (between 0 and 1) is deduced from the absorbance versus temperature profiles, with linear baselines assumption. These $\Delta G^\circ_{50°C(Na+)}$ or $\Delta G^\circ_{62°C(K+)}$ are of course less biologically relevant than an extrapolated $\Delta G^\circ_{37°C}$ but they are determined with a much higher confidence and fewer hypotheses. Comparison of T_m values (linear fits, Student's *t*-tests) were performed with Kaleidagraph 3.6 software. Error bars correspond to standard deviation values between independent experiments or sequences.

TDS and CD spectra. Thermal difference spectra (TDS) were obtained by difference between the absorbance spectra from unfolded and folded oligonucleotides that were recorded much above and below their T_m (33).

Circular dichroism (CD) spectra were recorded on a JASCO-810 spectropolarimeter as previously described (19).

Gel electrophoresis. For some experiments, formation of G4-DNA was confirmed by non-denaturing PAGE. In that case, oligonucleotides were either 5′ labelled with T4 polynucleotide kinase or directly visualized by UV-shadow (see Supplementary Data). Prior to the incubation, the DNA samples were heated at 90°C for 5 min and slowly cooled (2 h) to room temperature. Samples were incubated at 50 nM or 4 μM strand concentration in Tris/HCl 10 mM pH 7.5 buffer with 100 mM NaCl or KCl. Ten percent sucrose was added just before loading. Oligothymidylate markers (dT_{15}, dT_{21} or dT_{30}) or double-stranded markers (Dx_9: 5′d-GCGTATCGG + 5′d-CCGA TACGC; Dx_{12}: 5′d-GCGTGACTTCGG + 5′d-CCGAA GTCACGC) were also loaded on the gel. One should note that the migration of the dT_n oligonucleotides does not necessarily correspond to single strands (34): these oligonucleotides were chosen here to provide an internal migration standard, not to identify intramolecular or higher-order structures.

Differential scanning calorimetry. Microcalorimetry experiments were carried out using a Nano DSC-II microcalorimeter as previously described (35,36). Buffer and oligo solutions were carefully degassed prior to their utilization and their thermal profiles were analysed in the 0 to 100°C temperature range at a scan rate of 1°C/min. A minimum of 12 scans (six heating/six cooling) was collected for each experiment (see Supplementary Data for examples of raw data). For all the scans, the oligo versus buffer scan was substracted by the previously performed buffer scan, which allowed us to obtain the best scan shapes. Subtraction of the constructed baseline and calculation of the thermodynamic parameters were carried out using the Cp-calc software (Applied Thermodynamics). The oligonucleotides were dissolved at concentrations ranging from 200 to 241 μM in 10 mM lithium cacodylate buffer at pH 7.2 containing 100 mM KCl or NaCl.

RESULTS

Evidence for quadruplex formation

Our objective is to compare the stability of close sequences under strictly identical conditions. All oligonucleotides (except one, see below) studied here form quadruplexes under both reference conditions (10 mM Lithium cacodylate pH 7.2 supplemented with 100 mM NaCl or KCl, abbreviated hereafter K^+ or Na^+ conditions). One can follow the typical evolution on temperature of the folded fraction determined at 295 nm in the presence of 100 mM KCl or NaCl (Figure 1A and B). One notices the presence of a cation-dependent conformational change associated with the temperature increase. As expected, stability of the quadruplex was always lower in NaCl and higher in KCl (Table 1). The TDS in K^+ and Na^+ buffers are shown in Figure 1C and D. These TDS spectra were obtained by difference between the absorbance spectra recorded above

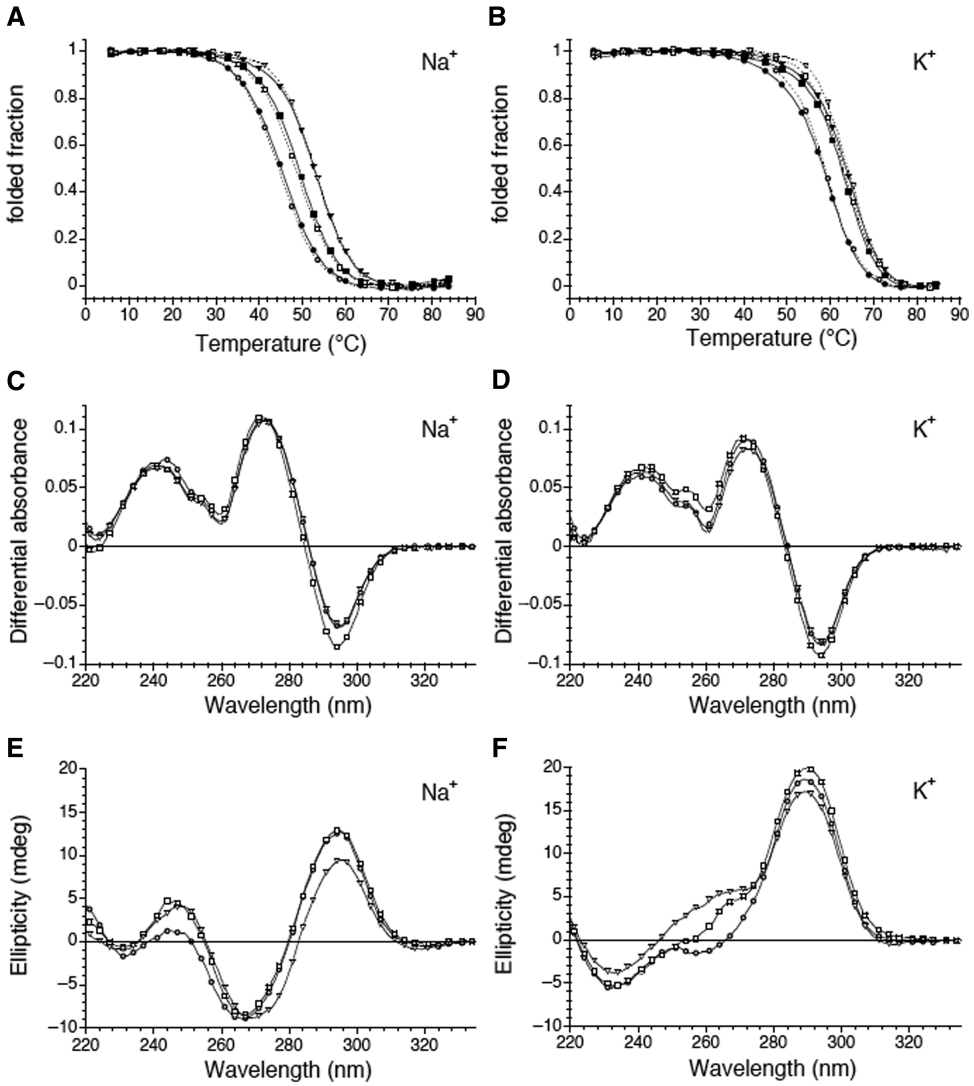

Figure 1. Spectroscopic data. (**A, B**) Folded fraction (derived from the absorbance at 295 nm) is plotted as a function of temperature for a selection of three oligonucleotides at two different concentrations: 5 μM (dotted lines; using 1-cm path length quartz cuvettes) and ≈225 μM (full lines; using 0.1-cm path length quartz cuvettes) for the ACT (circles), TCA (squares) and CGC (triangles) sequences (examples of raw melting profiles may be found in the Supplementary Data). Both buffers contained 10 mM lithium cacodylate at pH 7.2; (A) in 100 mM NaCl, (B) in 100 mM KCl. (**C, D**) Thermal difference spectra result from the difference between the absorbance recorded at 88 ± 2°C and at 4 ± 2°C for the same oligonucleotides (5 μM strand concentration) using 1-cm path length quartz cuvettes. They are normalized (TDS_{norm} = TDS/max(TDS) over the 220–335-nm wavelength range; (A) in 100 mM NaCl, (B) in 100 mM KCl. (**E, F**) Circular dichroism (CD) spectra (5 μM strand concentration) were recorded at 25°C using 1-cm path length quartz cuvettes; (A) in 100 mM NaCl, (B) in 100 mM KCl. Only a few experimental points are shown for clarity.

and below the observed transition. They 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 (26,33,37). Furthermore, the CD spectra of these structures were in agreement with the formation of quadruplexes (Figure 1E and F) (38,39). Representative examples are shown in this figure. Spectra in sodium (Figure 1E) and in potassium (Figure 1F) are different, suggesting that the folding schemes for all these sequences are different in the two salts; nevertheless, CD spectra do not provide direct evidence for the folding topology of quadruplexes and this conclusion should be treated with caution. Finally, formation of a quadruplex was confirmed by the stability dependence of the structure on the nature of the monocation (Table 1).

For some sequences, concentration-independent melting temperatures confirmed that the folding process was intramolecular (Figure 1A and B). Non-denaturing gel electrophoresis allowed us to compare the behaviour in sodium and potassium. In the example provided in Figure 2B, (bottom) a single band is observed in potassium for all sequences. Its fast migration is in excellent agreement with the formation of intramolecular complexes, as judged by molecular size markers such as double-stranded sequences or oligothymidylate repeats (left lanes). Migration was not affected by the addition of unlabelled oligonucleotide, indicating that the predominant species was the same at 50 nM or 4 μM strand concentration, in full agreement with intramolecular structures. In sodium, one also observed a single major

Table 1. Sequence of the oligonucleotide used (part 1)

Name	Sequence 5′ => 3′	T_m[a] Na^+	T_m[a] K^+	ΔG°[b] Na^+ (50°C)	ΔG°[b] K^+ (62°C)
TTT	GGGTTTGGG**TTT**GGGTTTGGG	**49.3**	**64.1**	**+0.02**	**−0.37**
TAA	GGGTTTGGG**TAA**GGGTTTGGG	**49.9**	**65.2**	**−0.07**	**−0.45**
ATA	GGGTTTGGG**ATA**GGGTTTGGG	**46.3**	**62.0**	**+0.52**	**+0.07**
AAT	GGGTTTGGG**AAT**GGGTTTGGG	**44.8**	**61.0**	**+0.71**	**+0.24**
TTC	GGGTTTGGG**TTC**GGGTTTGGG	**53.9**	**64.2**	**−0.4**	**−0.4**
TCT	GGGTTTGGG**TCT**GGGTTTGGG	**47.9**	**63.2**	**+0.31**	**−0.23**
CTT	GGGTTTGGG**CTT**GGGTTTGGG	**53.4**	**63.6**	**−0.46**	**−0.27**
TCC	GGGTTTGGG**TCC**GGGTTTGGG	**51.8**	**62.2**	**−0.21**	**−0.11**
CTC	GGGTTTGGG**CTC**GGGTTTGGG	**53.6**	**64.0**	**−0.56**	**−0.48**
CCT	GGGTTTGGG**CCT**GGGTTTGGG	**49.5**	**61.6**	**+0.2**	**−0.08**
ATT	GGGTTTGGG**ATT**GGGTTTGGG	**45.8**	**61.2**	**+0.65**	**+0.1**
TAT	GGGTTTGGG**TAT**GGGTTTGGG	**49.9**	**64.2**	**+0.02**	**−0.31**
TTA	GGGTTTGGG**TTA**GGGTTTGGG	**51.3**	**64.7**	**−0.18**	**−0.51**
AAC	GGGTTTGGG**AAC**GGGTTTGGG	**49.3**	**60.5**	**+0.06**	**+0.28**
ACA	GGGTTTGGG**ACA**GGGTTTGGG	**45.1**	**60.3**	**+0.62**	**+0.28**
CAA	GGGTTTGGG**CAA**GGGTTTGGG	**50.1**	**63.0**	**−0.02**	**−0.27**
ACC	GGGTTTGGG**ACC**GGGTTTGGG	**48.6**	**58.1**	**+0.27**	**+0.6**
CAC	GGGTTTGGG**CAC**GGGTTTGGG	**53.6**	**65.0**	**−0.49**	**−0.5**
CCA	GGGTTTGGG**CCA**GGGTTTGGG	**48.4**	**63.0**	**+0.20**	**−0.24**
AAA	GGGTTTGGG**AAA**GGGTTTGGG	**44.8**	**60.1**	**+0.74**	**+0.32**
CCC	GGGTTTGGG**CCC**GGGTTTGGG	**49.8**	**61.8**	**+0.04**	**+0.04**
ATC	GGGTTTGGG**ATC**GGGTTTGGG	**49.7**	**60.0**	**+0.1**	**+0.3**
TAC	GGGTTTGGG**TAC**GGGTTTGGG	**53.3**	**63.7**	**−0.38**	**−0.29**
TCA	GGGTTTGGG**TCA**GGGTTTGGG	**50.0**	**63.3**	**+0.09**	**−0.21**
ACT	GGGTTTGGG**ACT**GGGTTTGGG	**45.9**	**59.2**	**+0.65**	**+0.5**
CAT	GGGTTTGGG**CAT**GGGTTTGGG	**50.8**	**63.5**	**−0.09**	**−0.32**
CTA	GGGTTTGGG**CTA**GGGTTTGGG	**50.5**	**63.6**	**−0.02**	**−0.34**
AGA	GGGTTTGGG**AGA**GGGTTTGGG	**45.5**	**60.2**	**+0.56**	**+0.26**
AGT	GGGTTTGGG**AGT**GGGTTTGGG	**47.3**	**61.2**	**+0.39**	**+0.1**
TGA	GGGTTTGGG**TGA**GGGTTTGGG	**50.4**	**64.4**	**0**	**−0.35**
TGT	GGGTTTGGG**TGT**GGGTTTGGG	**52.0**	**66.0**	**−0.31**	**−0.66**
TGC	GGGTTTGGG**TGC**GGGTTTGGG	**54.0**	**63.5**	**−0.58**	**−0.24**
CGC	GGGTTTGGG**CGC**GGGTTTGGG	**54.9**	**64.5**	**−0.55**	**−0.43**
CGT	GGGTTTGGG**CGT**GGGTTTGGG	**53.0**	**65.9**	**−0.5**	**−0.72**
AGC	GGGTTTGGG**AGC**GGGTTTGGG	**50.0**	**60.5**	**−0.03**	**0.35**
CGA	GGGTTTGGG**CGA**GGGTTTGGG	**49.8**	**64.3**	**+0.03**	**−0.35**

[a]T_m in °C, with a ± 0.5°C precision; average of two to four independent values; determined from the analysis of UV melting profiles at 295 nm and/or 240 nm.
[b]ΔG° in kcal/mol, determined at 50°C in sodium, 62°C in potassium from the analysis of UV melting profiles at 295 nm. Negative values mean that the oligonucleotide is predominantly folded at this temperature.

band, for which concentration-independent migration was also in agreement with intramolecular quadruplexes (Figure 2A). Nevertheless, although intramolecular complexes predominated in all cases, minor bands corresponding to higher-order structures (probably dimers) were also observed in Na^+ for some of the sequences with an ANN central loop (i.e. the sequences that form the least stable intramolecular quadruplexes—see below—are more prone to form species of higher molecularity). Interestingly, significant differences in migration between sequences were found in sodium, while migration was nearly insensitive to sequence in potassium (compare top and bottom gels). Analysis of all 36 mers in sodium was also performed at a higher strand concentration (30 μM) by UV-shadow analysis (Figure S2); it confirmed the results with radiolabelled oligomers.

Analysis of stability

The T_ms and ΔG° of the 36 different oligonucleotides are summarized in Table 1. For all sequences, T_m is higher in potassium than in sodium, as for nearly all quadruplexes published so far. In this series, the T_m difference is moderate as the average Na^+–K^+ T_m difference was 12.7°C. The formation of quadruplex structures, whatever is their type, is clearly enthalpy driven (Table 2). Despite a negative contribution of entropy to stability, all quadruplex structures studied here are stable at physiological temperature, especially when potassium is present. One should also note that the correlation between T_ms (Figure 3A) obtained in sodium and potassium is weak (correlation coefficient = 0.69; data not shown). This indicates that sequence dependent effects are somewhat different in sodium and potassium, and that results obtained under a given cation are weakly predictive of what can be expected under another cation. Differences in T_m accurately reflect differences in ΔG° (Figure 3B). A fair correlation may be found between these two parameters. A 1°C difference in T_m roughly translates into a 0.13 (in Na^+) to 0.15 (in K^+) kcal/mol difference in ΔG°.

DSC analysis

Three sequences were re-synthesized at a large scale (1 μM) in order to perform DSC experiments at 200–241-μM strand concentration. Representative examples

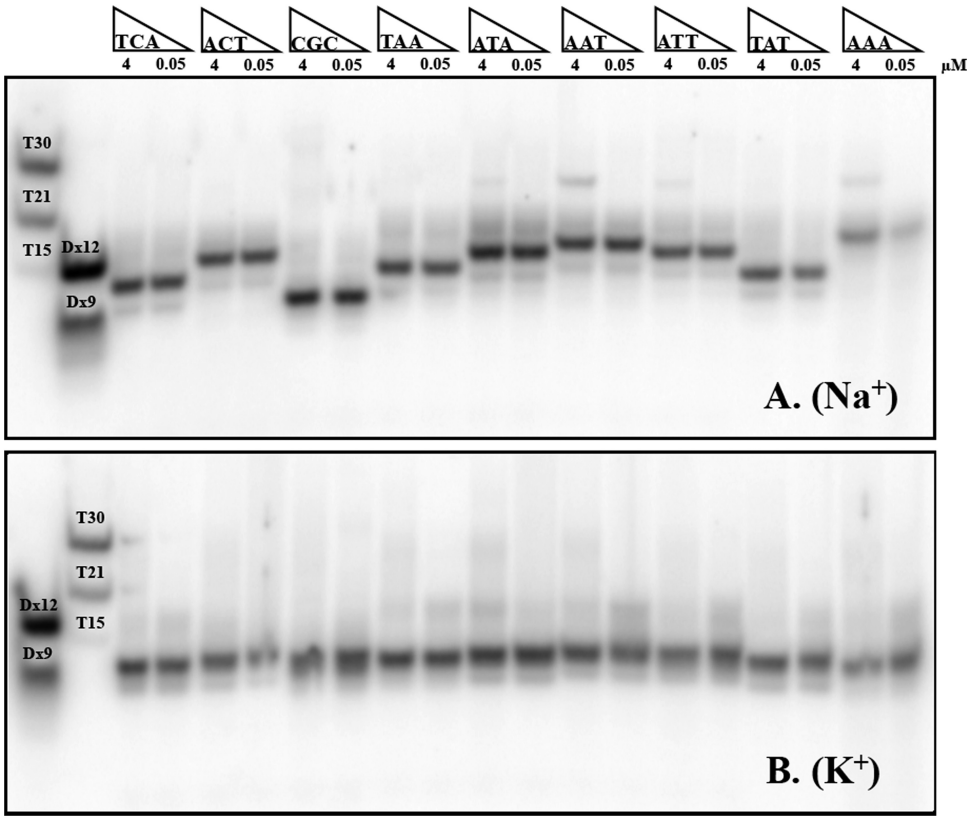

Figure 2. Oligonucleotide migration on a non-denaturing gel. PAGE profiles for a selection of nine oligonucleotides at two different concentrations: 0.05 μM (radio-labelled only) and 4 μM. Samples were prepared in a 100 mM NaCl (**A**) or KCl (**B**) buffer and loaded on a non-denaturing 15% acrylamide gel supplemented with 20 mM of the corresponding salt and run at 20°C. Migration markers are double-stranded sequences (Dx12 and Dx 9, forming 12 and 9 bp) and 'single-stranded' dT_n oligomers (15, 21 or 30 nt long).

Table 2. Model-dependent versus model-independent thermodynamic parameters

Salt	Oligo	(Strand) μM	UV-melting			DSC: average excess enthalpy/entropy			DSC: deconvolution general model, 1 transition[d]		
			T_m[a] (°C)	$\Delta H°_{VH}$ (kcal mol^{-1})	$\Delta S°_{VH}$ (cal K^{-1} mol^{-1})	ΔH_{cal}[c] (kcal mol^{-1})	ΔS_{cal}[c] (cal K^{-1} mol^{-1})	T_t(°C)	ΔH[e] (kcal mol^{-1})	ΔCp[e]	T_m (°C)
Na^+	**ACT**	223	45	42.0 ± 1.2	132 ± 3	30.1 ± 2.2	94 ± 7	47.3 ± 0.7	35.5 ± 0.7	−0.70 ± 0.14	50.8 ± 0.9
		5	45	44.1	139						
	TCA	241	50	43.5 ± 1.8	135 ± 5	31.0 ± 2.2	95 ± 7	52.0 ± 0.8	39.0 ± 0.6	−0.53 ± 0.11	54.5 ± 0.7
		5	49	44.0	136						
	CGC	225	53.5	44.9 ± 1.2	138 ± 4	28.6 ± 2.3	87 ± 7	56.2 ± 0.7	40.0 ± 0.9	−0.66 ± 0.20	59.2 ± 1.0
		5	53	48.4	149						
K^+	**ACT**	215	59	(44)[b]	(133)[b]	36.4 ± 3.8	109 ± 11	62.1 ± 0.5	44.6 ± 1.0	−0.79 ± 0.21	64.9 ± 0.5
		5	59	(49)[b]	(148)[b]						
	TCA	200	63	(48)[b]	(142)[b]	39.6 ± 3.5	117 ± 10	66.7 ± 0.3	48.6 ± 0.8	−0.86 ± 0.19	69.0 ± 0.4
		5	63	(49)[b]	(146)[b]						
	CGC	233	63.5	(43)[b]	(127)[b]	40.1 ± 3.8	118 ± 11	67.6 ± 0.4	47.9 ± 1.1	−0.86 ± 0.29	69.9 ± 0.6
		5	63.5	(53)[b]	(156)[b]						

[a]These values (which correspond to a different series of experiments) are in fair agreement with those presented in Table 1.
[b]$\Delta H°_{VH}$ and $\Delta S°_{VH}$ in KCl are provided for illustration only, as lnK versus $1/T$ graphs significantly deviate from linearity (see Supplementary Figure S1D for an example). Hence, linear fitting of these graphs is inappropriate.
[c]$\Delta H_{cal} = \int Cp^{excess}(T)dT$ and $\Delta S_{cal} = \int \frac{Cp^{excess}(T)}{T} dT$
where Cp^{excess} is the excess heat capacity function. Average of six heating and six cooling profiles, respectively.
[d]The general transition model directly fits the molar heat capacity Cp (and not the excess heat capacity Cp^{excess}). It is used for transitions with $\Delta Cp \neq 0$. In this model, $\Delta Cp(T)$ is fitted with a second order polynome: $\Delta Cp(T) = a + bT + cT^2 = \Delta Cp(T_m) + b(T - T_m) + c(T^2 - T_m^2)$. Average of six heating and six cooling profiles, respectively.
[e]ΔH and ΔCp at $T = T_m$

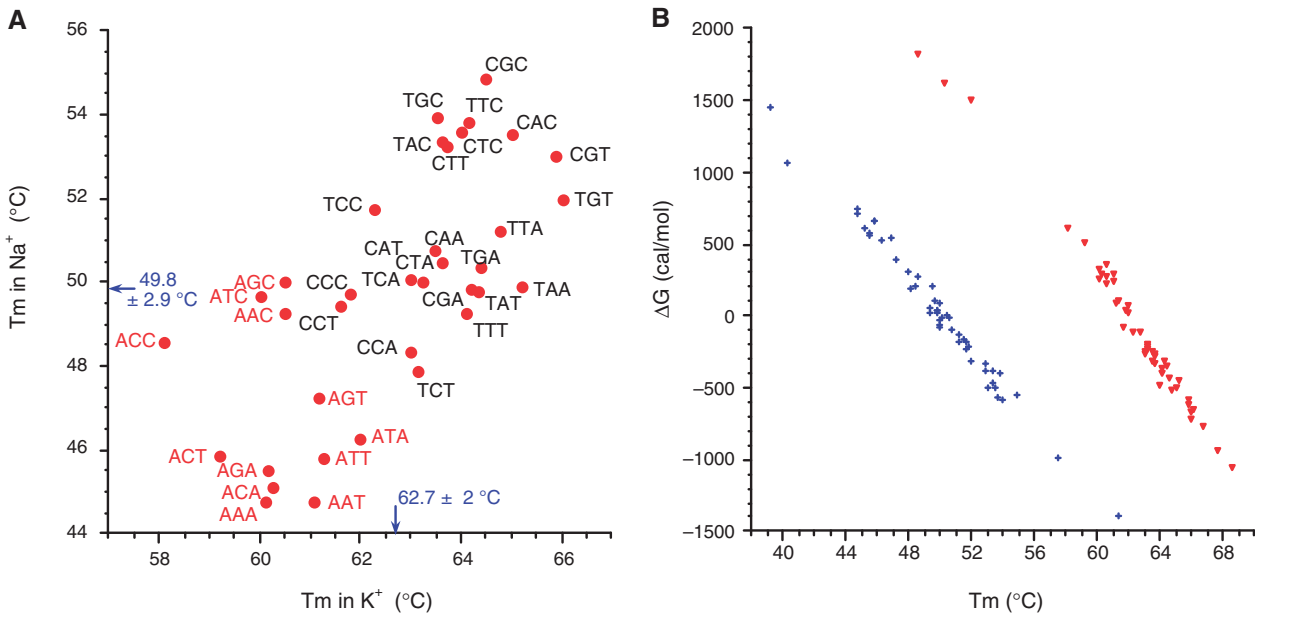

Figure 3. T_m and ΔG° in sodium and potassium. (**A**) T_m in 100 mM NaCl (with 10 mM lithium cacodylate at pH 7.2) is plotted versus T_m in 100 mM KCl (same buffer). All 36 sequences are shown. Average values in potassium (62.7°C) and sodium (49.8°C) are provided. (**B**) $\Delta G^\circ_{50°C}$ in 100 mM NaCl with 10 mM lithium cacodylate at pH 7.2 (blue crosses) or $\Delta G^\circ_{62°C}$ in 100 mM KCl (same buffer; red triangles) is plotted versus T_m, determined under identical conditions. (A linear fit between ΔG° and $1/T_m$ would be expected if all sequences had identical enthalpies).

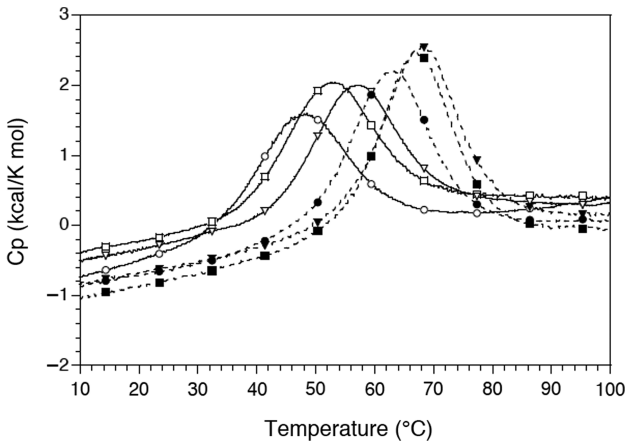

Figure 4. DSC analysis. *Cp* versus temperature plots for the ACT (circles), TCA (squares) and CGC (triangles) sequences at strand concentration of ≈225 µM. The experimental buffer contained 10 mM lithium cacodylate at pH 7.2 with either 100 mM NaCl (full lines) or 100 mM KCl (dotted lines).

of DSC scans (in sodium and potassium) are shown in Figure 4 and full scans are presented in Figure S5. A comparison of the thermodynamic parameters obtained from DSC experiments with a van't Hoff analysis derived from UV study can be found in Table 2. Calorimetry experiments confirmed the differences in stability between three sequences chosen for in depth analysis (ACT, TCA and CGC):

– a representative example of a low stability case (ACT)
– a representative example of an intermediate stability case (TCA)
– a representative example of a high stability case (CGC)

Figure 4 illustrates the DSC profiles for these oligonucleotides. The T_t temperatures (maximum of the *Cp* versus temperature plots) derived from DSC experiments were in fair agreement with the T_m deduced from UV-melting experiments (Table 2). The non-linear behaviour of ln(K) versus 1/T Arrhenius plots for UV-melting experiments in potassium (an example is provided in Supplementary Figure S1D) prevented us from accurately determining ΔH° and ΔS° in KCl. Hence, a comparison of ΔH°_{cal} and ΔH°_{VH} (and of ΔS°_{cal} and ΔS°_{VH}) could only be done in sodium. Model-independent ΔH°_{cal} values were less negative (by 28–36%) than ΔH°_{VH} for the three oligonucleotides. This less favourable enthalpy was (of course) compensated by a less negative entropy. Deconvolution of the DSC profiles with a general model for a single transition with a non-zero ΔCp provided intermediate thermodynamic values (Table 2).

Extension to other contexts

Some rules emerged from the wealth of data collected here. Nevertheless, one may wonder if these rules apply to this sequence context only, or if they may be generalized to various quadruplex types. Quadruplexes are highly polymorphic (40–42): rules that are valid for a given loop type (lateral, diagonal, propeller) do not necessarily apply to other types. To test this hypothesis, we chose two loops starting with an adenine (ACT and ATT) and compared them with TCA and CGC. These loop sequences were tested in six other quadruplex types (Table 3) with a variable number of quartets, different positions (i.e. when present in the first, central or last loops). Finally we analysed the effect of adenines in loops of different lengths (two or four bases). As shown in this table, the ACT and ATT sequences are less stable than the TCA and CGC motifs

Table 3. Loop sequence effects in various contexts

Name	Sequence 5′ => 3’	T_m[a] Na^+	T_m[a] K^+	$\Delta G°$[b] $Na^+_{50°C}$	$\Delta G°$[b] $K^+_{62°C}$
TT	GGGTTAGGG**TT**GGGTTAGGG	**46.9**	**63.9**	**+0.4**	**−0.3**
AT	GGGTTAGGG**AT**GGGTTAGGG	**43.8**	**59.5**	**+0.8**	**+0.4**
TA	GGGTTAGGG**TA**GGGTTAGGG	**55.2**	**66.2**	**−0.7**	**−0.7**
TTTT	GGGTTAGGG**TTTT**GGGTTAGGG	**58.0**	**68.2**	**−1.3**	**−1.1**
ATTT	GGGTTAGGG**ATTT**GGGTTAGGG	**57.5**	**68.2**	**−1.1**	**−1.3**
TTTA	GGGTTAGGG**TTTA**GGGTTAGGG	**64.0**	**71.5**	**−2.3**	**−2.1**
ACT	GGGTTAGGG**ACT**GGGTTAGGG	**52.9**	**62.7**	**−0.3**	**−0.1**
TCA	GGGTTAGGG**TCA**GGGTTAGGG	**57.5**	**66.1**	**−1.0**	**−0.7**
CGC	GGGTTAGGG**CGC**GGGTTAGGG	**61.4**	**68.5**	**−1.4**	**−1.0**
L3**ACT**	GGGTTTGGGTTTGGG**ACT**GGG	**46.8**	**60.5**	**+0.5**	**+0.2**
L3**ATT**	GGGTTTGGGTTTGGG**ATT**GGG	**47.5**	**61.9**	**+0.4**	**0**
L3**TCA**	GGGTTTGGGTTTGGG**TCA**GGG	**51.7**	**67.6**	**−0.2**	**−0.9**
L3**CGC**	GGGTTTGGGTTTGGG**CGC**GGG	**52.8**	65.8*	**−0.4**	**−0.6**
L1**ACT**	GGG**ACT**GGGTTTGGGTTTGGG	**45.5**	**61.9**	**+0.6**	**0**
L1**ATT**	GGG**ATT**GGGTTTGGGTTTGGG	**46.8**	**62.4**	**+0.5**	**−0.1**
L1**TCA**	GGG**TCA**GGGTTTGGGTTTGGG	**51.5**	**66.7**	**−0.2**	**−0.8**
L1**CGC**	GGG**CGC**GGGTTTGGGTTTGGG	51.1*	65.8*	**−0.1**	**−0.6**
X3**ACT**	GGG**ACT**GGG**ACT**GGG**ACT**GGG	**39.1**	**51.9**	**+1.5**	**+1.5**
X3**TCA**	GGG**TCA**GGG**TCA**GGG**TCA**GGG	**51.6**	**63.2**	**−0.2**	**−0.2**
X3**CGC**	GGG**CGC**GGG**CGC**GGG**CGC**GGG	–[c]	–[c]	–[c]	–[c]
17**ACT**	GGGTGGG**ACT**GGGTGGG	**40.2**	**>80**	**+1.1**	**<−3**
17**TCA**	GGGTGGG**TCA**GGGTGGG	**48.1**	**>80**	**+0.2**	**<−3**
17**CGC**	GGGTGGG**CGC**GGGTGGG	**50**	**>80**	**−0.1**	**<−3**
15**ACT**	GGTTGG**ACT**GGTTGG	**<12**	**35.4**	**>3**	**>3**
15**TCA**	GGTTGG**TCA**GGTTGG	**16**	**40.0**	**>3**	**>3**
15**TGT**[d]	GGTTGG**TGT**GGTTGG	**20**	**50.2**	**>3**	**+1.6**
15**CGC**	GGTTGG**CGC**GGTTGG	**26.7**	**48.6**	**>3**	**+1.8**

[a]T_m in °C, with a ±0.5°C precision; determined from the analysis of UV melting profiles at 295 nm.
[b]$\Delta G°$ in kcal/mol, determined at 50°C in sodium, 62°C in potassium from the analysis of UV melting profiles at 295 nm. In some instance, values could not be accurately determined (stability of the quadruplex is either too low or too high at this temperature) and minimal/maximal values are provided.
[c]This GC-rich sequence does not form a quadruplex.
[d]Thrombin aptamer sequence.
*In this sequence context, CGC is slightly less stable than TCA (in contrast with all other samples).

in all cases (seven different sequence contexts in sodium versus six in potassium; no comparison could be made in one case due to high stability), demonstrating that the ‘No A as a first base’ rule is general. This rule also applies to loops consisting of two or four bases, as AT and ATTT loops were less stable than TA and TTTA, respectively. On the other hand, the CGC loop is often, but not always (8/11 cases) more stable than the TCA loop (Table 3).

DISCUSSION

In the present study, we analysed the effects of base substitutions in a 3-base-long loop for intramolecular quadruplexes. Although many oligomers adopt relatively similar conformations, the stabilities of these complexes may significantly vary. The comparison of an important number of sequences (a total of 54) in two ionic conditions allowed to draw general rules. The difference found between potassium and sodium is relatively modest for most of the sequences presented here. In comparison, this difference [$T_{m(K+)} - T_{m(Na+)}$] is in the same range (8°C) for the human telomeric motif (26) but can reach 36°C for single-base loops sequences (19). These general observations do not explain why some sequences are more sensitive than others to the nature of the cation.

We demonstrate here that overall base composition of the loops is a poor predictor of quadruplex stability. Whatever the parameter considered (total number of purines, or A, T or C in the loops), the correlation with thermal stability or $\Delta G°$ is poor, especially in potassium (Supplementary Figures S3 and S4). Nevertheless, some sequence effects may be evidenced: adenine is significantly disfavoured over a pyrimidine base (T or C) in a sodium buffer (correlation coefficient = 0.64). This is in agreement with previous observations (19,43). However, to our surprise, this destabilizing effect mainly applies to the ‘first’ position of the loop (Figure 5A) and was confirmed in loops consisting of 2, 3 or 4 bases. This is not observed when adenine is present at the central or last position.

Other less obvious but still significant effects may be evidenced as well. A cytosine is slightly detrimental (by 1–2°C) to quadruplex stability when present at the central position (data not shown). In contrast, the presence of a cytosine at the last position has little effect in potassium, and a significant stabilizing effect in sodium (Figure 5B). These observations allow us to propose consensus sequences for high- or low-stability 3-base-long loops:

‘Most’ stable (Na^+): *Y D **C***
‘Most’ stable (K^+): *Y D **H***
‘Least’ stable (Na^+): *A C **W***
‘Least’ stable (K^+): *A C **H***

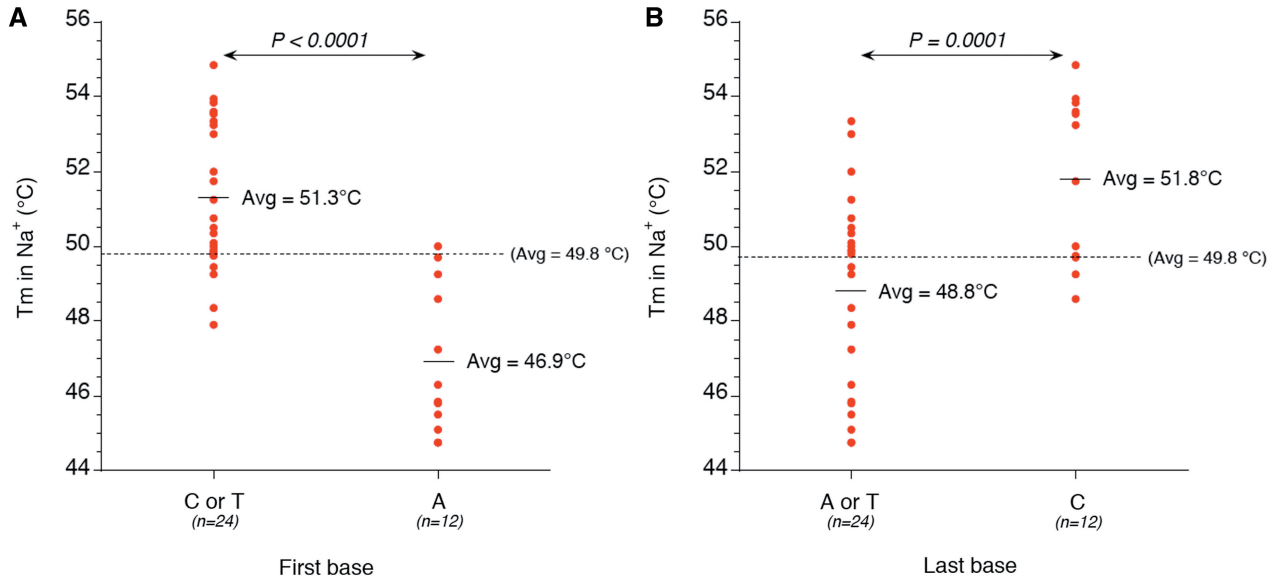

Figure 5. Position effects. Examples of paired comparisons: (**A**) Effect on T_m of the nature of the 'first' base in the loop (A versus C or T). (**B**) Effect on T_m of the nature of the 'last' base in the loop (C versus A or T). Student's paired *t*-test values are shown. Similar conclusions were reached when looking at ΔG° values (data not shown).

where $Y = C$ or T; $H = A$, C or T; $W = A$ or T and $D = A$, G or T. (Note that, due to ambiguity problem, we cannot easily determine the effect of a guanine as the first or last base of a loop.)

The stability of a structure at 37°C (directly related to ΔG°) is obviously more useful for biological purposes than a T_m. Unfortunately, its determination/extrapolation is more difficult for a number of reasons (to name a few: low reproducibility, baseline artefacts, non-zero ΔC_p°, large uncertainty in ΔH° determination and non-two-state melting behaviour). For these reasons, we propose to experimentally determine the folded fraction at a fixed temperature close to the average T_m for a series of sequences, which may immediately be translated into a ΔG°. We simply chose reference temperatures of 50 and 62°C close to the average T_m in sodium and potassium, respectively, so that one can determine θ and thus K and ΔG with maximal accuracy. Differences in stabilities ($\Delta\Delta G^\circ$) may then be determined with good confidence and reproducibility. Obviously, these $\Delta\Delta G^\circ$ are determined at a non-physiological temperature; nevertheless, provided that enthalpies of the various sequences are not too different, one may expect that the conclusions still hold at 37°C. Alternatively, Figure 3B demonstrates a fair correlation between T_m and ΔG°, suggesting that for this series of >50 sequences, a T_m difference of 1°C may be translated into a ΔG° difference of 0.13–0.15 kcal/mol. As the T_m values in sodium span a >10°C temperature range, sequence effects may affect the K_d by a factor of ≈10, showing that sequence effects play a very significant role which was underestimated before.

Our ultimate goal is to establish rules to predict the stability of intramolecular G-quadruplexes based on primary sequence. This approach is complementary to the recent work by Kumar and Maiti (21), who compared the stability of a number of biologically relevant sequences with various loop size and length (total loop length 5–18 bases). In this report, we decided to explore a limited sequence space (only 3-base-long loops are considered) but in a nearly exhaustive fashion, allowing us to evidence relatively subtle effects. Our results also demonstrate that loop composition affect quadruplex structure and thermodynamics, thus making it difficult to draw generalized correlations between loop length and thermodynamic stability (21). The destabilizing effect of adenine as the first base of the loop has not been reported before. Efforts are now being made to understand the structural basis of this destabilization. Therefore, thanks to the data accumulated by several groups, one may build a set of 'Santa Lucia-like' table for quadruplexes, in order to propose a predictive algorithm for G4 stability, which may eventually be incorporated in MFOLD. However, the situation will be more complex than for duplexes, as sequences effects cannot be simplified to nearest neighbour approximation. A first prediction algorithm provides a fair T_m estimate in a number of situations (44). Finally, it will be interesting to check at the genome-wide level if loops starting with an adenine are under or over-represented in G4 putative sequences. It is also striking that telomeric motifs from different species never correspond to 'loops' starting with an A (i.e. GGGTTA, GGG TTTA, GGGGTT, GGGGTTTT, etc.) as if the least stable quadruplex motif was selected against in telomeric motifs (Tran *et al.*, in preparation).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank L. Lacroix, J. Gros, L. L. Guittat, J.F. Riou and C. Trentesaux (Museum, Paris) for helpful discussions.

FUNDING

INSERM, CNRS and the Muséum National d'Histoire Naturelle. 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. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
7. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
8. Huppert,J.L. (2008) Hunting G-quadruplexes. *Biochimie*, **90**, 1140–1148.
9. 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.
10. 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.
11. 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.
12. 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.
13. Risitano,A. and Fox,K.R. (2003) Stability of intramolecular DNA quadruplexes: comparison with DNA duplexes. *Biochemistry*, **42**, 6507–6513.
14. 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.
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. Risitano,A. and Fox,K.R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **32**, 2598–2606.
17. 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.
18. 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.
19. 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.
20. Kumar,N., Sahoo,B., Varun,K.A.S., Maiti,S. and Maiti,S. (2008) Effect of loop length variation on quadruplex-Watson Crick duplex competition. *Nucleic Acids Res.*, **33**, 4433–4442.
21. Kumar,N. and Maiti,S. (2008) A thermodynamic overview of naturally occurring intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **36**, 5610–5622.
22. Vorlickova,M., Bednarova,K., Kejnovska,I. and Kypr,J. (2007) Intramolecular and intermolecular guanine quadruplexes of DNA in aqueous salt and ethanol solutions. *Biopolymers*, **86**, 1–10.
23. 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.
24. Mergny,J.L. and Lacroix,L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515–537.
25. Mergny,J.L. and Lacroix,L. (2009) UV melting of G-quadruplexes. *Curr. Protoc. Nucleic Acids Chem.*, **17**, 17.11.11–17.11.15.
26. Mergny,J.L., Phan,A.T. and Lacroix,L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
27. 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.
28. Rachwal,P.A., Findlow,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.
29. Gros,J., Rosu,F., Amrane,S., De Cian,A., Gabelica,V., Lacroix,L. and Mergny,J.L. (2007) Guanines are a quartet's best friend: impact of base substitutions on the kinetics and stability of tetramolecular quadruplexes. *Nucleic Acids Res.*, **35**, 3064–3075.
30. Breslauer,K.J. (1994) In Agrawal,S. (ed.), *Methods in Molecular Biology, Vol. 26, Protocols for Oligonucleotide Conjugates*. Humana Press Inc., Totowa, NJ, pp. 347–372.
31. Breslauer,K.J. (1995) In Johnson,M. L. and Ackers,G. K. (eds), *Energetics of Biological Macromolecules*. Vol. 259. Academic Press Inc., San Diego, CA, pp. 221–242.
32. Chalikian,T.V., Volker,J., Plum,G.E. and Breslauer,K.J. (1999) A more unified picture for the thermodynamics of nucleic acid duplex melting: a characterization by calorimetric and volumetric techniques. *Proc. Natl Acad. Sci. USA*, **96**, 7853–7858.
33. 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.
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. Amrane,S., Saccà,B., Mills,M., Chauhan,M., Klump,H.H. and Mergny,J.L. (2005) Length-dependent energetics of (CTG)n and (CAG)n trinucleotide repeats. *Nucleic Acids Res.*, **33**, 4065–4077.
36. Amrane,S. and Mergny,J.L. (2006) Length and pH-dependent energetics of (CCG)(n) and (CGG)(n) trinucleotide repeats. *Biochimie*, **88**, 1125–1134.
37. Petraccone,L., Erra,E., Randazzo,A. and Giancola,C. (2006) Energetic aspects of locked nucleic acids quadruplex association and dissociation. *Biopolymers*, **83**, 584–594.
38. 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.
39. 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.
40. Dai,J., Carver,M. and Yang,D. (2008) Polymorphism of human telomeric quadruplex structures. *Biochimie*, **90**, 1172–1183.
41. Shirude,P.S. and Balasubramanian,S. (2008) Single molecule conformational analysis of DNA G-quadruplexes. *Biochimie*, **90**, 1197–1206.
42. Neidle,S. and Parkinson,G.N. (2008) Quadruplex DNA crystal structures and drug design. *Biochimie*, **90**, 1184–1196.
43. 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.
44. Stegle,O., Payet,L., Mergny,J.L. and Huppert,J. (2009) Predicting and understanding the stability of G-quadruplexes. *Bioinformatics*, **25**, i374–i382.