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

Aurore Guédin1,2, Julien Gros2, Patrizia Alberti2 and Jean-Louis Mergny1,2,*

1INSERM, U869, European Institute of Chemistry and Biology, Bordeaux 2 University, 2 rue Robert Escarpit, Pessac F-33607 and 2INSERM, 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).

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’-GGGL₁GGGL₂GGGL₃GGG-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 GGTTTGGTTT TIGGGTTG GGGG sequence. Bases in the loops are in bold/underline characters.
- When loop length was >9 nt, letters were used: \( B = 12, E = 15, K = 21, S = 30 \). Hence, \( 1B1 \) corresponds...
<table>
<thead>
<tr>
<th>(Length) Name</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>(T_m) Na(^{+}) (°C)</th>
<th>(T_m) K(^{+}) (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T-var-T</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(16) 121</td>
<td>T</td>
<td>ggg</td>
<td>ggg</td>
<td>T</td>
<td>(51)</td>
</tr>
<tr>
<td>(17) 131</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>45.5</td>
</tr>
<tr>
<td>(18) 141</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>43.0</td>
</tr>
<tr>
<td>(19) 151</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>41.5</td>
</tr>
<tr>
<td>(20) 161</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>40.3</td>
</tr>
<tr>
<td>(21) 171</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>37.0</td>
</tr>
<tr>
<td>(22) 181</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>32.0</td>
</tr>
<tr>
<td>(23) 191</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>26.1</td>
</tr>
<tr>
<td>(24) 201</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>24.8</td>
</tr>
<tr>
<td><strong>TTT-var-TTT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(19) 313</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>(49.5)</td>
</tr>
<tr>
<td>(20) 323</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>(46.1)</td>
</tr>
<tr>
<td>(21) 333</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>51.0</td>
</tr>
<tr>
<td>(22) 343</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>56.5</td>
</tr>
<tr>
<td>(23) 353</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>54.5</td>
</tr>
<tr>
<td>(24) 363</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>54.5</td>
</tr>
<tr>
<td>(25) 373</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>nd</td>
</tr>
<tr>
<td>(27) 393</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>45.5</td>
</tr>
<tr>
<td>(30) 3B3</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>39.3</td>
</tr>
<tr>
<td>(33) 3E3</td>
<td>T</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>36.5</td>
</tr>
<tr>
<td><strong>TTA-var-TTA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(19) H313</td>
<td>TTA</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>(45.0)</td>
</tr>
<tr>
<td>(20) H323</td>
<td>TTA</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>(54.3)</td>
</tr>
<tr>
<td>(21) H333</td>
<td>TTA</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>56.6</td>
</tr>
<tr>
<td>(22) H343</td>
<td>TTA</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>62.5</td>
</tr>
<tr>
<td>(23) H353</td>
<td>TTA</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>61.8</td>
</tr>
<tr>
<td>(24) H363</td>
<td>TTA</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>60.8</td>
</tr>
<tr>
<td>(25) H373</td>
<td>TTA</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>57.5</td>
</tr>
<tr>
<td>(27) H393</td>
<td>TTA</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>52.8</td>
</tr>
<tr>
<td>(30) H3B3</td>
<td>TTA</td>
<td>ggg</td>
<td>T</td>
<td>ggg</td>
<td>46.8</td>
</tr>
<tr>
<td><strong>var-TTT-TTT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(19) 133</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>45.8</td>
</tr>
<tr>
<td>(20) 233</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>50.7</td>
</tr>
<tr>
<td>(21) 333</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>51.0</td>
</tr>
<tr>
<td>(22) 433</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>44.4</td>
</tr>
<tr>
<td>(23) 533</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>40.7</td>
</tr>
<tr>
<td>(24) 633</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>38.2</td>
</tr>
<tr>
<td>(25) 733</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>36.7</td>
</tr>
<tr>
<td>(27) 933</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>33.9</td>
</tr>
<tr>
<td>(30) B33</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>31.0</td>
</tr>
<tr>
<td>(33) E33</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>28.0</td>
</tr>
<tr>
<td><strong>TTT-TTT-var</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(19) 331</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>46.2</td>
</tr>
<tr>
<td>(20) 332</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>51.8</td>
</tr>
<tr>
<td>(21) 333</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>51.0</td>
</tr>
<tr>
<td>(22) 334</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>44.5</td>
</tr>
<tr>
<td>(23) 335</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>39.5</td>
</tr>
<tr>
<td>(24) 336</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>37.0</td>
</tr>
<tr>
<td>(25) 337</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>35.5</td>
</tr>
<tr>
<td>(27) 339</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>tgg</td>
<td>30.5</td>
</tr>
<tr>
<td>(30) 3B3</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>27.8</td>
</tr>
<tr>
<td>(33) 3E3</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>25.5</td>
</tr>
<tr>
<td><strong>var-TTT-var</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(17) 131</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>45.5</td>
</tr>
<tr>
<td>(19) 232</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>53.8</td>
</tr>
<tr>
<td>(21) 333</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>51.0</td>
</tr>
<tr>
<td>(23) 434</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>ggg</td>
<td>40.8</td>
</tr>
<tr>
<td>(25) 535</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>tgg</td>
<td>34.3</td>
</tr>
<tr>
<td>(27) 636</td>
<td>T</td>
<td>ggg</td>
<td>TTT</td>
<td>tgg</td>
<td>31.5</td>
</tr>
<tr>
<td>(29) 737</td>
<td>TTTTTTT</td>
<td>ggg</td>
<td>TTT</td>
<td>tgg</td>
<td>26.0</td>
</tr>
<tr>
<td>(33) 339</td>
<td>T(_g)</td>
<td>ggg</td>
<td>TTT</td>
<td>tgg</td>
<td>18.5</td>
</tr>
<tr>
<td>(39) B3B</td>
<td>T(_{1,2})</td>
<td>ggg</td>
<td>TTT</td>
<td>tgg</td>
<td>13.0</td>
</tr>
<tr>
<td>(45) E3E</td>
<td>T(_{1,8})</td>
<td>ggg</td>
<td>TTT</td>
<td>tgg</td>
<td>11.0</td>
</tr>
</tbody>
</table>

(continued)
to the \text{GGGTGGGT...} \text{GGGTGGG} sequence. For the family of sequences in which the L₁ and L₃ loops are not solely composed of thymines but of the TTA motif (Table 1), the prefix letter \text{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 dTm 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 ΔTm 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 Tm 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 Tm, 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 abundance, 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
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,
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 L1, L2 and L3.

(ii) The conclusions reached for L1 generally apply to L3. 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 L2 is variable (circles) is different from the two others, both in potassium and in sodium.

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).

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 (dT15, dT21 or dT30) or double-stranded markers (dx9:5'-d-GCGATACGG+5'-d-CCGATACGC dx12:5'-d-GCGTGACTTCC+5'-d-CCGAAGTCACGC). Oligonucleotide length (in nucleotide) is indicated above each band for comparison purposes.
(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 ($\Delta T_m$ 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...
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 deconvolute 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 $\Delta T_m$ 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 $T_m > 50^\circ C$ 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.

(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 $/C_14$ decrease in $T_m$ which may roughly be translated in a 0.3 kcal/mol loss in $/C_1$G/$/C_14$ (38). We did not observe a plateau around a total loop length of $\approx 7$ 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

Table 3. Effect of increasing loop length on $T_m$

<table>
<thead>
<tr>
<th>Loop Size</th>
<th>In Sodium $^a$</th>
<th>In Potassium $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>From 6 to 9 nt</td>
<td>5.2 ± 1.9</td>
<td>6.5 ± 2.2</td>
</tr>
<tr>
<td>From 6 to 12 nt</td>
<td>8.8 ± 3.1</td>
<td>11.0 ± 3.5</td>
</tr>
<tr>
<td>From 6 to 15 nt</td>
<td>11.6 ± 5.1</td>
<td>14.0 ± 3.2</td>
</tr>
</tbody>
</table>

$^a$destabilization, in $^\circ$ 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.

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 GGL1GGGL2GGGL3GG motif is currently underway (Alberti et 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 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 samples to be tested at different concentrations) one can easily underestimate 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


