BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

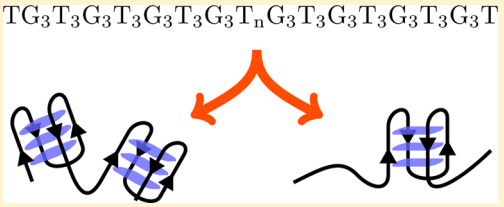
pubs.acs.org/biochemistry

Stability and Structure of Long Intramolecular G-Quadruplexes

Linda Payet*,§ and Julian L. Huppert

Cavendish Laboratory, University of Cambridge, J. J. Thomson Avenue, Cambridge CB3 OHE, U.K.

ABSTRACT: G-quadruplexes are formed from guanine-rich sequences of DNA and RNA. They consist of stacks of square arrangements of guanines called G-quartets. Increasing evidence suggests that these structures are involved in cellular processes such as transcription or translation. Knowing their structure and their stability in vitro should help to predict their formation in vivo and to understand their biological functions. Many studies have been performed on isolated G-quadruplexes, but little attention has been given to their interactions. Here, we present non-denaturing gel electrophoresis, UV melting, and circular dichroism data obtained for long sequences of DNA which are capable of forming two simultaneous G-quadruplexes, namely, d($TG_3T_3G_3T_3G_3T_3G_3T_nG_3T_3G_3T_3G_3T_3G_3T$), with n varying from one to seven. These sequences can form up to two separate G-quadruplexes. We also study mutated versions of these sequences designed to form one G-quadruplex at specific positions on the strand. Comparing results from the original sequences and their mutated versions, we show that for the former different folded states coexist: either with six stacked G-quartets or only three, in various combinations. Which ones are favored depends on n. Moreover, for n greater than three, the thermodynamic stability stays constant, contrary to an expected decrease in stability if the six G-quartets were stacked together in a single structure. This result agrees with a beads-on-a-string folding model for long sequences of G-quadruplexes, where two adjacent G-quadruplexes fold independently.



Guanine-rich sequences of DNA or RNA of type $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$ (N denotes any nucleotide, 3+ means at least 3, and 1−7 means between 1 and 7) can fold into intramolecular G-quadruplexes (*GQ*). These consist of π-stacked G-quartets as seen in Figure 1, formed by Hoogsteen hydrogen bonding and stabilized by monovalent cations.[1] In this article, we will generally consider G-quadruplexes formed by three G-quartets, since they are the most commonly studied. Starting with such a sequence (or a chain of such sequences), one or several G-quadruplexes might form under given physical conditions, and they might have different topologies,[2] interact with each other in different ways, and form various structures. Our aim is to understand these structures and how stable they are.

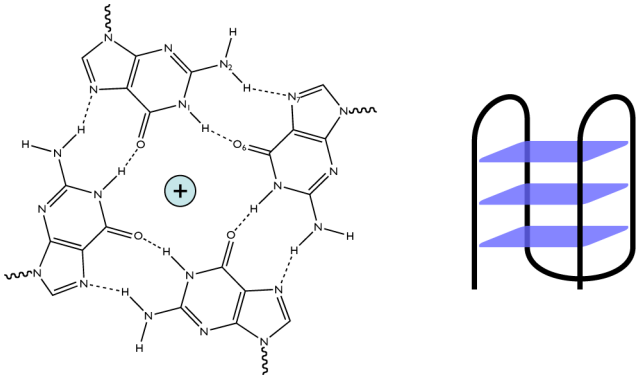


Figure 1. Left: A G-quartet formed by Hoogsteen hydrogen bondings and stabilized by a monovalent cation. Right: An example of a single G-quadruplex resulting from the stacking of three G-quartets.

By genome-wide bioinformatic searches, we have shown that there are around 376 000 such sequences in the human genome[3,4] and that around 40% of all human gene promoters contain at least one of them.[5] Intramolecular G-quadruplexes may fold in vitro with high stability under nearly physiological conditions (e.g., 100 mM KCl, pH 7.4).[6,7] The stability of the G-quadruplexes varies in particular with the length of the loops linking the G-quartets; the shorter the loops, the higher the melting temperature,[6−8] which can be greater than 90 °C, and the stability of some *GQ* is comparable to that of a double helix.[9] Evidence of *GQ* folding in vivo has been found for the telomeric repeat $d(TTTTGGGG)_n$ in the macronuclei of *Stylonychia lemnae*.[10] In addition, a number of proteins involved in biological activities have been shown to bind specifically to *GQ*.[11−13] All of these facts together motivate active research on their potential role in cellular processes (see the review written by Lipps et al.[14]).

The most studied intramolecular *GQ* is the long single-stranded human telomeric repeat d(TTAGGG) found at the end of the chromosomes. During each cell division, telomeres shorten by 50−200 base pairs, gradually inducing cell death. In contrast, the enzyme telomerase promotes the extension of telomeres and the lifetime of cells. In particular, 85% of cancer cells reveal an upregulation of telomerase.[15] It has been shown that this telomerase overactivity can be inhibited by folding the telomere overhang in G-quadruplex structures.[16] Also, a certain number of *GQ*-forming motifs in human gene promoters have been found to affect gene transcription, for example, in c-MYC,

Received: November 27, 2011
Revised: February 20, 2012
Published: March 14, 2012

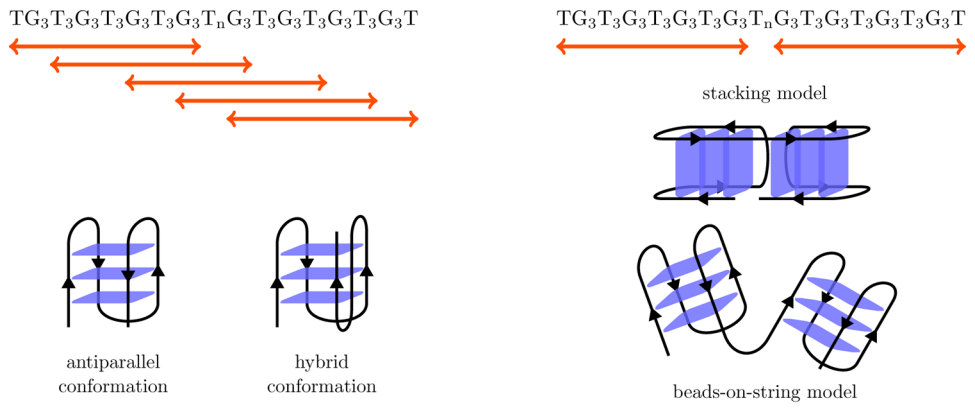


Figure 2. (A) Representation of the different possible folding positions of a single G-quadruplex on a sequence which could form two simultaneaous G-quadruplexes. (B) Representation of models of double G-quadruplex interaction: stacking and bead-on-string models. The orange arrows illustrate the possible positions of G-quadruplexes. The directions of strands along quartets with respect to each other imply different conformations such as antiparallel (two strands have the same direction) and hybrid conformations (three strands have the same direction).

Table 1. Oligonucleotides Used in This Study with *n* Varying between 1 and 7

name		oligonucleotide sequence
GQ	1PQS	TGGGTTTGGGTTTGGGTTTGGGT
2GQ (generic name)	$2GQ_n$	$TGGGTTTGGGTTTGGGTTTGGGT_nGGGTTTGGGTTTGGGTTTGGGT$
1GQ (generic name)	GQ_nT	$TTTTTTTTTTTTTGGGTTTGGGT_nGGGTTTGGGTTTTTTTTTTTTT$
	GQT_{25}	$TGGGTTTGGGTTTGGGTTTGGGTTTTTTTTTTTTTTTTTTTTTTTTT$
	$T_{25}GQ$	$TTTTTTTTTTTTTTTTTTTTTTTTTGGGTTTGGGTTTGGGTTTGGGT$
	T_nA	$TGGGTTTGGGTTTGGGTTTGGGT_nGGGTTTGGGTTTGGGTTTGAGT$
	AT_nA	$TGAGTTTGGGTTTGGGTTTGGGT_nGGGTTTGGGTTTGGGTTTGAGT$
	AAT_nAA	$TGAGTTTGAGTTTGGGTTTGGGT_nGGGTTTGGGTTTGAGTTTGAGT$
	$AAAT_nA$	$TGAGTTTGAGTTTGAGTTTGGGT_nGGGTTTGGGTTTGGGTTTGAGT$
	$AAAAT_n$	$TGAGTTTGAGTTTGAGTTTGAGT_nGGGTTTGGGTTTGGGTTTGGGT$

VEGF, HIF-1α, Ret, c-Kit, KRAS, and Bcl-2 promoters.[17,18] The study of G-quadruplex stability and structure is therefore important to understand their interaction with proteins and their role in cell processes and to help rational drug design.[19,20] G-quadruplexes may also be synthesized to produce aptamers such as AS1411, which induce breast tumor cell death,[21,22] or to construct nanodevices.[23,24]

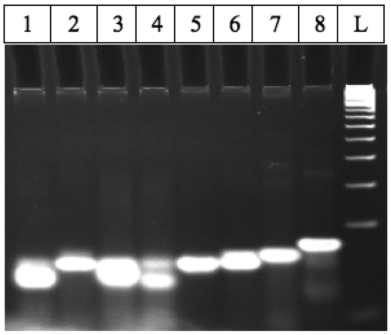
Our ultimate aim is to be able to predict computationally the structure and stability of putative G-quadruplexes in a given sequence. Toward this goal, we recently developed a computational prediction program based on Bayesian methods. The computer implementation was trained on data sets obtained by biophysical experiments,[25] involving short sequences allowing the formation of only one G-quadruplex. It predicts the melting temperature of G-quadruplexes within a 5 °C error margin. To extend its application to longer strands, it is necessary to have an experimental data set of thermodynamics where multiple G-quadruplex sequences could form.

Most studies of intramolecular G-quadruplex folding involve short sequences forming at most one *GQ*. There are a few looking at longer sequences,[26−35] mainly concerned with the human RNA and DNA telomere. It is however known that the human genome presents a number of other long sequences which could form several putative G-quadruplexes,[2,36−38] and it is currently unclear what structures theses sequences may form and with what stability. This motivates the study of model sequences forming possibly more than one *GQ*, and it raises specific questions. So, considering the sequence in Figure 2, we can ask: Do we have one (Figure 2A) or two G-quadruplexes (Figure 2B)? If we have only one, where is it on the strand (there are five potential positions)? If we have multiple *GQ*, how do all the formed G-quartets interact together? For structures with multiple *GQ*, two main models have been proposed: the beads-on-string model,[27,28,34,35] in which *GQ* are compared to beads moving independently of each other, and the stacking model,[26,30−32] in which *GQ* stack on each other in a higher-order structure. Two of many possible structural arrangements are sketched in Figure 2B. Martadinata et al. actually suggest that the two types of structures (beads-on-a-string and stacking) might coexist in long RNA or DNA telomeres, with their formation depending on various interacting partners in the cell.[32]

In this study, we investigate the stability and structure of oligonucleotides that could potentially form two simultaneous G-quadruplexes, that is, that have eight runs of d(GGG). We combine non-denaturing gel electrophoresis, UV melting, and circular dichroism (CD) techniques.

In particular, we study the influence of the number *n* of bases separating the two putative *GQ* in the sequence $TG_3T_3G_3T_3G_3T_3G_3T_nG_3T_3G_3T_3G_3T_3G_3T$ for *n* varying from 1 to 7. We compare the structure and stability of G-quadruplexes on the above sequences and on mutated sequences designed to allow the formation of only one *GQ* at various positions on the strand (see Figure 2A). It allows us to determine the position where G-quadruplexes actually form on our long sequences.

dx.doi.org/10.1021/bi201750g | *Biochemistry* 2012, 51, 3154−3161



Lane	1	2	3	4	5	6	7	8
Oligo	$2GQ_3$	$2GQ_7$	$2GQ_1$	$2GQ_2$	GQT_{25}	$T_{25}GQ$	GQ_1T	GQ_7T
Length	47-nt	51-nt	45-nt	46-nt	47-nt	47-nt	45-nt	51-nt

Figure 3. Gel electrophoresis in buffer 20 mM KCl, pH 7.4. Oligomers stabilized with 100 mM KCl. 20 bp ladder was used.

MATERIAL AND METHODS

Oligonucleotides were provided by Invitrogen. The sequences used in this work are presented in Table 1. Stock solutions of 100 μM were made with 10 mM Tris-HCl, pH 7.4. All experiments were performed with samples of around 4 μM ssDNA in a 10 mM Tris-HCl buffer at pH 7.4 with 100 mM KCl.

The stability of folding/unfolding of the different oligonucleotides was measured by UV melting experiments using a Varian Cary 300 spectophotometer. The UV absorbance profiles were recorded at 295 nm during two cycles of cooling/heating between 95 and 10 °C at a rate of 0.25 °C/min. Before measurement, the samples were annealed and degassed at 95 °C for 15 min.

The conformation of the oligonucleotides was investigated using CD in a Chirascan spectropolarimeter. The samples were annealed by heating at 95 °C and then cooled down by steps of 5 °C every 1500 s and stored at 5 °C. Measurements were usually performed at 10 °C, unless otherwise specified. For CD melting, samples were annealed and degassed at 95 °C for 15 min, and profiles were recorded with temperature varying from 95 to 10 °C at a rate of 0.25 °C/min.

Non-denaturing gel electrophoresis was performed on 20% polyacrylamide gel prepared at a final concentration 1× of TBE and 20 mM of KCl. Separation was performed at 80 V and at room temperature using a 1× TBE supplemented with 20 mM of KCl. The samples were annealed by heating at 95 °C and then cooled down by steps of 5 °C every 1500 s and stored at 5 °C until measurements. After electrophoresis, the gels were stained with Sybr Gold provided by Invitrogen.

RESULTS

Gel Electrophoresis. We first wanted to establish the molecularity of the structures formed from our $2GQ$ sequences, and in particular to check that they were all intramolecular under our experimental conditions. We also wished to find out whether there were several grossly different folded states for each sequence. We therefore started by performing native gel electrophoresis. This technique allowed us to evaluate the coexistence of different folded structures based on their different compactness. We also compared the $2GQ_n$ sequences with corresponding mutated sequences, each of which can only form one G-quadruplex (GQ_nT).

For all the sequences studied in this work, gel electrophoresis reveals that there are only intramolecular structures. For example, in Figure 3, the sequences are composed of 45−51 nucleotides, and they all migrate faster than the double-stranded 40 bp marker. Figure 3 also shows that for the $2GQ$ sequences, $2GQ_1$, $2GQ_2$, and $2GQ_3$, the bands are broad compared to $2GQ_7$ or compared to $1GQ$ sequences (GQ_1T, GQT_{25}, $T_{25}GQ$, and GQ_7T). This suggests that the $2GQ_1$, $2GQ_2$ and $2GQ_3$ form different species of different compactness. In contrast, the species formed by the sequence $2GQ_7$ have the same compactness. Finally, we observe that $2GQ_7$ diffuses faster than both GQ_7T and GQ_1T, although the molecular weight of the former is bigger than the molecular weight of the latter two. This difference in diffusion may have two explanations. First, Vorlíčková et al.[39] showed that oligo(dT) has an abnormally slow diffusion in gel, probably due to some interactions between the terminal thymines with the matrix. This may be the case for the sequences GQ_7T and GQ_1T. We believe also that the intramolecular structure formed by $2GQ_7$ is more compact than the single folded G-quadruplexes formed by GQ_7T and GQ_1T. This more compact structure with faster diffusion may correspond to a state where all the guanines are part of a G-quadruplex such as in Figure 2B. For $2GQ_1$, $2GQ_2$, and $2GQ_3$, it appears that there is a mixture (clearly resolved for $2GQ_2$) of different diffusive structures. This mixture may contain slow diffusive structures such as the single G-quadruplexes $1GQ$ and faster structures probably corresponding to a more compact structure such as in Figure 2B. In addition, we cannot exclude at this point that the single or multiple G-quadruplexes may fold in different topologies with different mobilities. The $2GQ_4$, $2GQ_5$, and $2GQ_6$ sequences behave the same way as $2GQ_7$, that is, show a single diffusion band with faster running than their mutated sequences (data not shown).

Thermodynamics. *Stability against Varying Number of Bases between Two Putative GQ*. The effect of the number of bases n (from 1 to 7) on the stability of the structures formed by $2GQ_n$ is studied by UV melting. We also study the short sequence 1PQS for comparison.

For all eight sequences, hypochromic melting transitions at 295 nm are observed characterizing the folding/unfolding of G-quadruplexes[40] such as shown in Figure 4 for $2GQ_3$.

To compare graphically the stability of the different sequences, we plot the normalized first derivative of absorbance as a function of temperature (Figure 5). While for n from 2 to 7 only one transition is observed, for n = 1 there are two transitions. This is consistent with what was observed by

dx.doi.org/10.1021/bi201750g | *Biochemistry* 2012, 51, 3154−3161

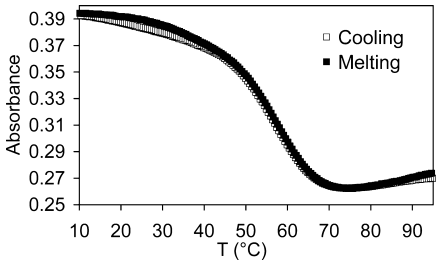


Figure 4. Melting/cooling curves for $2GQ_3$. No hysteresis observed around the transition.

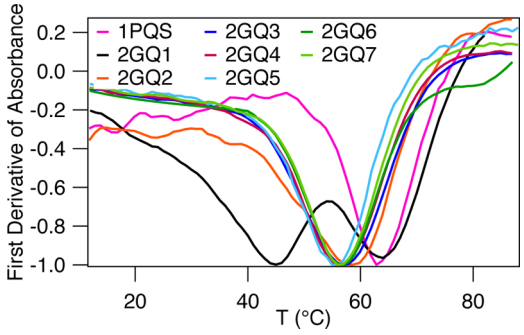


Figure 5. First derivative of absorbance (normalized by its maximum) as a function of temperature for 1PQS and $2GQ_n$ sequences with n varying from 1 to 7 in the presence of 100 mM KCl.

electrophoresis: we have different folded species for n = 1, and they have melting temperatures different enough to be distinguished. For n = 2 or 3, the different structures observed by electrophoresis probably have similar melting temperatures, making only one transition visible. These types of biphasic melting profiles[28,41−44] or single transition for mixtures of intramolecular structures[28,44] had previously been observed.

For almost all the sequences, at the cooling/heating rate of 0.25 °C/min, any hysteresis is negligible (e.g., for n = 3 in Figure 4). The n = 1 case (Figure 6) shows slight hysteresis, but

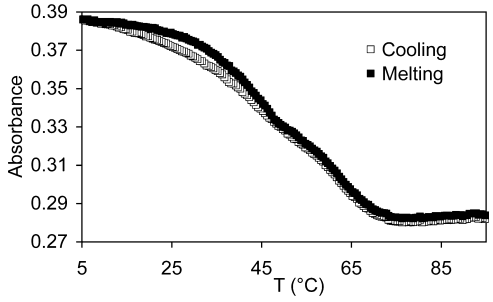


Figure 6. Melting/cooling curves for $2GQ_1$. Negligible hysteresis observed around the two transitions (T_m equals respectively 44 and 64 °C).

even there the folding and unfolding temperatures determined at the minima of the absorbance derivative for the first and second transition differ only by 1.5 and 0.3 °C, respectively.

Van't Hoff analysis of melting profiles assumes that there is a two-state transition with only one folded structure in solution.[40] We show by gel electrophoresis that this is not the case at least for n = 1, 2, or 3. We therefore fit the melting profiles either to a 2-state or to a 3-state model using the least-squares method, based on the number of observed transitions.

The following behavior has been suggested for G-quadruplexes:[44] When multiple folded states are present in solution, one may still observe a single transition if the states melt at similar temperatures. In this case, the temperature observed represents an intermediate value of the melting of the different species and the transition tends to be broader. This means that the energetic parameters ΔH, ΔS, obtained assuming a two-state equilibrium, are higher than for each isolated species. Thermodynamic parameters, and in particular the melting temperature T_m, still give an approximate understanding of the stability of the folded structures.

Table 2 lists the measured parameters for the sequences studied, and the results are also shown graphically as a function

Table 2. Thermodynamic Parameters for the Short Single G-Quadruplex (1PQS) and the Long Double G-Quadruplex ($2GQ_n$) Sequences[a]

	T_m (°C)	ΔH (kJ/mol)	ΔS (J/mol/°C)	ΔG (kJ/mol)
1PQS	66 ± 1	−240 ± 30	−720 ± 70	−20 ± 2
$2GQ_1$	64 ± 1	−185 ± 15	−550 ± 40	−15 ± 2
	44 ± 0.5	−150 ± 20	−480 ± 40	−4 ± 2
$2GQ_2$	61.5 ± 0.5	−180 ± 10	−530 ± 40	−13 ± 1
$2GQ_3$	58 ± 1	−170 ± 10	−520 ± 40	−11 ± 1
$2GQ_4$	58 ± 1	−190 ± 20	−590 ± 50	−12 ± 1
$2GQ_5$	57 ± 1.2	−180 ± 20	−550 ± 80	−11 ± 1
$2GQ_6$	58 ± 0.5	−200 ± 20	−590 ± 40	−12.5 ± 0.5
$2GQ_7$	57 ± 1	−200 ± 10	−600 ± 40	−12.0 ± 1.5

[a] ΔG was calculated at 37 °C.

of n in Figure 7. The data for the short sequence 1PQS are shown as a filled marker and the data for n = 1 corresponds to the transition at the highest melting temperature.

Interestingly, the melting temperature decreases and ΔG increases as n varies from 1 (for the transition at high temperature) to 3, while for n varying from 3 to 7 it is constant. Similarly, ΔH and ΔS slightly increase for n varying from 1 (for the transition at high temperature) to 3, showing higher values than for the sequences with n greater than 3; we expect that the presented energies of the long sequences with n less than or equal to 3 are underestimated, due to the presence of different folded structures, as explained above.

In contrast, the values for n varying from 4 to 7 do not change particularly. This constant stability suggests that the two GQ fold independently and form a structure like beads-on-a-string. If there was positive cooperativity, then we would expect this separation to have an effect: if the two GQ were interacting as proposed in the stacking model, increasing n would induce a decrease of the stability, similarly to the way the loop length affects the stability of a single GQ.[6]

Moreover, the stability is lower for all the $2GQ$ sequences than for the 1PQS one. It confirms that there is no stacking between the single G-quadruplexes because such stacking would induce an increase of the thermodynamic stability compared to the single GQ; it again rules out cooperative formation. It is consistent with previous studies which showed that stability decreases when the sequence length increases by any number of repeats.[27,28]

Comparison with Mutated Sequences. The long sequences studied so far can form either one or two simultaneous GQ, and we wished to understand which position would be occupied if only one formed, especially for the sequences where n = 1−3 for which double GQ states coexist with single

dx.doi.org/10.1021/bi201750g | Biochemistry 2012, 51, 3154−3161

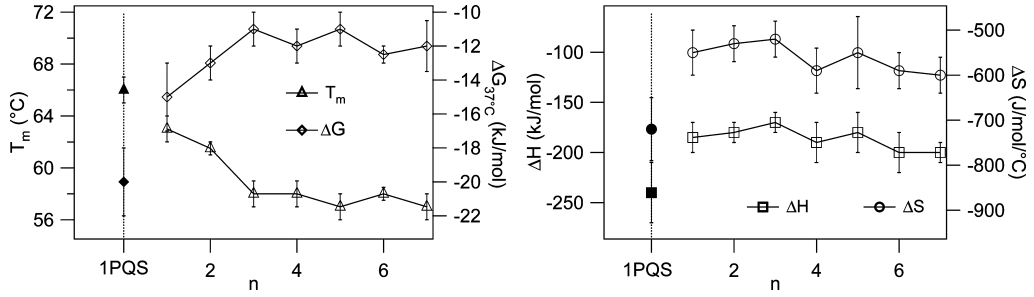


Figure 7. Thermodynamic parameters for the short single G-quadruplex 1PQS (filled marker) and the long double G-quadruplex $2GQ_n$ (opened marker) sequences. The data for n equal 1 corresponds to the transition at the highest melting temperature of $2GQ_1$. ΔG was calculated at 37 °C.

ones, as shown above. The objective was therefore to determine the stability of a single *GQ* at a particular position on a long $2GQ_n$ strand and to estimate if it may coexist (at that position) with the double *GQ* structure. We therefore designed new sequences by mutating slightly the original ones to prevent the formation of *GQ* at some positions, replacing the central guanine of one or more G-runs by adenine, which cannot form *GQ*-like structures and does not basepair with G. In each mutated sequence, we mutated G-runs at the ends, preserving central G-runs. See Table 1 for the notation of mutated sequences.

For all the sequences studied, we assume a two-state equilibrium, in agreement with gel electrophoresis which shows single-narrow bands and with the melting profile, which does not show any significant hysteresis and which implies a single transition (data not shown). The thermodynamic data are shown in Tables 3 and 4. We also determined

Table 3. Thermodynamic Parameters of Mutated Sequences of $2GQ_1$

sequence	T_m (°C)	ΔH (kJ/mol)	ΔS (J/mol/°C)	ΔG (kJ/mol)
TA	65.5 ± 1	−210 ± 20	−620 ± 70	−18 ± 1
ATA	64.4 ± 1	−200 ± 10	−600 ± 20	−16.4 ± 0.1
$2GQ_1$	64 ± 1	−185 ± 15	−550 ± 40	−15 ± 2
	44 ± 0.5	−150 ± 20	−480 ± 40	−4 ± 2
AATAA	59 ± 0.5	−190 ± 10	−580 ± 10	−12.7 ± 0.2
AAATA	60 ± 1	−200 ± 20	−600 ± 40	−13 ± 1
AAAAT	60 ± 1	−210 ± 30	−620 ± 80	−14 ± 1
T_1 mixture	61 ± 1	−180 ± 20	−550 ± 30	−13 ± 1

Table 4. Thermodynamic Parameters of Mutant Sequences of $2GQ_3$

sequence	T_m (°C)	ΔH (kJ/mol)	ΔS (J/mol/°C)	ΔG (kJ/mol)
$2GQ_3$	58 ± 1	−170 ± 10	−520 ± 40	−11 ± 1
AT_3A	54 ± 1	−200 ± 10	−610 ± 30	−10.5 ± 0.6
AAT_3AA	50 ± 1	−210 ± 10	−640 ± 20	−8.3 ± 0.3
$AAAAT_3$	59 ± 1	−210 ± 30	−640 ± 60	−14 ± 2
T_3 mixture	53.5 ± 0.5	−175 ± 15	−540 ± 20	−8.8 ± 0.5

the thermodynamics for mixed samples (AT_3A, AAT_3AA, and $AAAAT_3$) and (TA, ATA, AATAA, AAATA, and AAAAT) respectively called T_3 and T_1 mixtures. For these, we also analyze the melting curves as for a two-state equilibrium, as the transitions of individual sequence are very similar.

First, the melting temperature for $2GQ_1$ (corresponding to the transition at the highest melting temperature) and $2GQ_3$ are in the same range of value as their respective mutated sequences. Second, the value of ΔH and ΔS for the mixtures are lower (in absolute value) than for any isolated component for both n = 1 or 3, as expected for a transition involving different species in equilibrium (see above). This is also what we observed for $2GQ_1$ and $2GQ_3$ confirming that for these sequences the energies are overestimated and there are several folded states in equilibrium corresponding to single *GQ* folded at different positions on the strand and double *GQ*.

Interestingly, for all the mutated sequences, there is a single clear transition. The TA sequence (mutation of only one G at the extremities) is very close to the original $2GQ_1$: it contains the same putative *GQ* sequences but one and that missing one prevents the formation of two simultaneous *GQ*. Its melting temperature (65.5 °C) is very close to the highest melting temperature of $2GQ_1$ (64 °C). Furthermore, the sequence AAAAT is able to form a *GQ* in the spot present in $2GQ_1$ and missing in TA, and it also has a melting temperature around 60 °C. This clearly suggests that the lowest melting temperature transition (at 44 °C) involves a structure which can be formed by $2GQ_1$ and not by TA or AAAAT, that is, the double *GQ* structure. We believe that the thermodynamic parameters related to the highest melting temperature for the $2GQ_1$ sequence represents an equilibrium between the unfolded state and a mixture of all the different single *GQ* (Figure 2A) states. The lowest melting temperature should be associated with an equilibrium between the double *GQ* state and all the single *GQ* ones.

Notice that for *GQ* formed in the middle of the strand, the more consecutive G-runs, the greater the stability: ATA and AT_3A, with six consecutive G-runs, are respectively more stable than AATAA and AAT_3AA, with four. Therefore, we should expect that the equivalent structures will be even more stable in the original sequences than observed in the mutated forms containing less guanine. Also, in the mutated sequences, *GQ* forming in the middle of the strand are less stable than those forming at the extremities: AAT_3AA and AATAA are less stable than $AAAAT_3$ and AAAAT, respectively, despite having the same number of consecutive G-runs. This preference for folding at the extremities has also been observed with the human telomere repeat.[45] Finally, the data provide evidence that the effect of loop length for long sequences is the same as for short sequences;[7] structures involving at least one single base loop are more stable than sequences involving only triple base loops: ATA and AATAA are more stable than AT_3A and AAT_3AA, respectively.

Circular Dichroism. We performed CD experiments to confirm the formation of *GQ* in $2GQ_n$ and in order to better understand the different conformations that are compatible

dx.doi.org/10.1021/bi201750g | *Biochemistry* 2012, 51, 3154−3161

with the positions revealed by gel electrophoresis and UV melting experiments. For all the sequences, we observed CD spectra characteristic of *GQ* (Figure 8). As with the UV melting

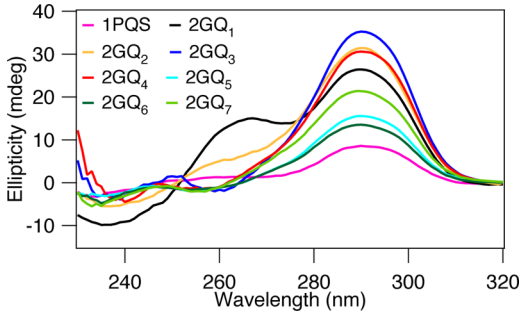


Figure 8. CD spectra of 1PQS and $2GQ_n$ sequences with n varying from 1 to 7 in the presence of 100 mM KCl at 10 °C.

profiles, the CD spectrum of the sequence with $n = 1$ differs strongly from the ones with $n \geq 2$.

For $n = 1$, the CD spectrum has positive peaks at 265 and 290 nm and a negative one around 240 nm. Given the sequence design, and what is known about the impact of loop lengths on conformations, this type of spectrum can be related to the structure of a *GQ* in the middle of the strand involving the single T loop.[6,7,46] The sample is therefore a mixture of parallel, antiparallel, and hybrid G-quadruplexes. In that case, single base loops link three G-quartets by adopting a double-chain-reversal conformation. With only one such short loop, the middle *GQ* forms a hybrid conformation with three parallel and one antiparallel strands (see Figure 2A). It was suggested that the presence of a single T loop may favor the longer loops to fold also in a double-chain-reversal conformation inducing then the formation of parallel structures.[7] As shown by UV measurements and gel electrophoresis, these parallel and hybrid conformations should coexist with single or double *GQ* which have only triple-base loops. These latter structures should contribute only to the 290 nm peak characteristic of antiparallel G-quadruplexes.

To confirm this hypothesis, we performed CD experiments on the mutated sequences for $n = 1$ (Figure 9). Recall that the

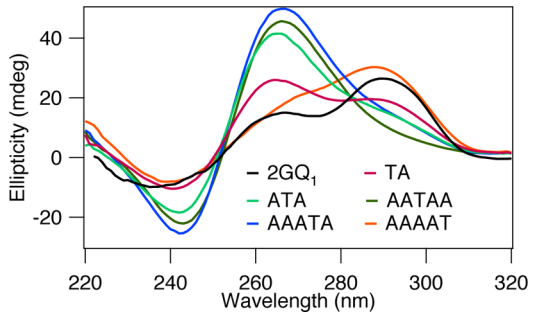


Figure 9. CD spectra of $2GQ_1$ and its mutated versions in the presence of 100 mM KCl at 10 °C.

mutated sequences are only able to form single *GQ*, at various positions on the strand. All the mutated sequences have spectra characteristic of *GQ* formation. As expected, the folded *GQ* involving a single T loop have CD spectra characteristic of mixture conformations (hybrid, parallel, and antiparallel) with both peaks around 265 and 290 nm. Furthermore, for sequences where only a G-quadruplex involving the single T loop can form (ATA, AATAA, AAATA), the peak around 290 nm is small compared to the one around 265 nm. In contrast, for TA which can fold in structures involving single T loops or structures having only T_3 loops, both peaks are quite similar and the overall spectrum resembles closely the $2GQ_1$ one. For AAAAT, the single T is not part of the *GQ* and, as expected, we only have the antiparallel conformation (see Figure 2A) represented by the peak at 290 nm (the peak around 270 is due to the unstructured tail). This demonstrates again that the $2GQ_1$ can fold in any configuration sketched in Figure 2A.

To investigate further the equilibrium for $n = 1$, we carried out CD melting experiments (Figure 10). At high temperatures,

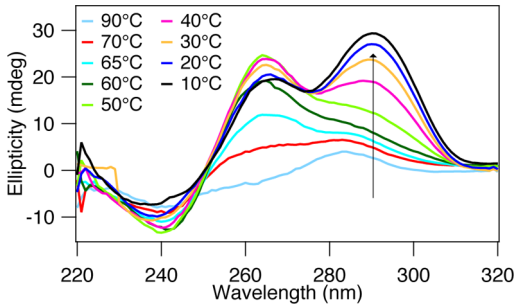


Figure 10. CD spectra of $2GQ_1$ during melting at 0.2 °C/min from 90 to 10 °C.

the peak at 265 nm predominates while at low temperatures the 290 nm one is predominant. It corresponds to a decrease of the proportion of the folded *GQ* involving the single T base in the middle of the strand at lower temperatures. These CD results agree with the $2GQ_1$ folding/unfolding scheme from UV melting data; that is, at high temperatures all the possible configurations presented in Figure 2A form and at low temperatures the single *GQ* structures, in particular, the ones in the middle of the strand, are replaced by double ones.

For $2GQ_2$, there is a very small peak at 265 nm and a high one at 290 nm. Hazel et al. showed that the T_2 loops on a G-quadruplex sequence can also be double-strand-reversal loops.[6] As for $n = 1$, extreme *GQ* (double or single) with T_3 loops coexist with the single middle *GQ* ones, they are in antiparallel conformation and contribute to the peak at 290 nm.

For $3 \leq n \leq 7$, by the previous methods, we know that the sequences $2GQ_n$ only form *GQ* with T_3 loops, thus antiparallel. As expected, they present no positive peak at 265 nm but a small negative one at 260 nm and a large positive peak at 290 nm. Therefore, for these sequences CD experiments cannot distinguish configurations.

DISCUSSION AND CONCLUSION

First of all, our results remind us that one should always be careful when analyzing UV melting experiments, even when a clear single transition is observed. The assumption of a two-state equilibrium should be cross-checked by other techniques, such as electrophoresis, and the analysis of the data should initially take into account the possible existence of more states.

We have shown that the sequences $2GQ_n$ can adopt a number of different configurations, many of them with roughly similar energies. They can fold into two *GQ* on the strand or only one, which can be at different positions along the sequence. This contrasts with the common assumption that as many G-quadruplexes as possible will form. For small n ($n \leq$

dx.doi.org/10.1021/bi201750g | *Biochemistry* 2012, 51, 3154–3161

3), we find that the proportion of single G-quadruplex strands is important, presumably because there is an entropic gain in having different structures; any configuration that can be envisaged is likely to form to some extent. In contrast, for larger n ($n > 3$), we have only one folded structure with two independent G-quadruplexes behaving like independent beads-on-a-string.

As for short sequences, shorter loops induce more stable structures, but a *GQ* with longer loops formed at the end of the strand can be as stable as a *GQ* with shorter loops in the middle of a strand. There are a number of sequences having between four and eight G-runs, and for which it is hard to tell where the G-quadruplex forms on the strand,[2,36−38] and our results should help in clarifying this issue.

This study is a step toward the prediction of the thermodynamics and structure of *GQ* on long sequences, where several *GQ* can form simultaneously or where one G-quadruplex structure can form at several positions. The position of the folding G-quadruplex on a sequence might be biologically important as it is expected for telomere,[45,47,48] so this line of research should be continued.

AUTHOR INFORMATION

Corresponding Author

*E-mail: lpayet@u-cergy.fr.

Present Address

[§]CNRS-UMR 8587, LAMBE, Université de Cergy. Telephone: (+33)134256649. Fax: (+33)134256694.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

J.L.H. is a Research Councils UK Academic Fellow in Computational Biology, on leave while he serves as the Member of Parliament for Cambridge.

REFERENCES

(1) Neidle, S., Balasubramanian, S., Eds. (2006) *Quadruplex Nucleic Acids*; Royal Society of Chemistry, Cambridge.

(2) 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.

(3) Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res. 33*, 2908−2916.

(4) Todd, A. K., Johnston, M., and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res. 33*, 2901−2907.

(5) Huppert, J. L., and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res. 35*, 406−413.

(6) Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc. 126*, 16405−16415.

(7) 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.

(8) Guédin, A., Gros, J., Alberti, P., and Mergny, J.-L. (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res. 38*, 7858−7868.

(9) Shirude, P., Okumus, B., Ying, L., Ha, T., and Balasubramanian, S. (2007) Single-molecule conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene C-kit. *J. Am. Chem. Soc. 129*, 7484−7485.

(10) 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. U.S.A. 98*, 8572−8577.

(11) Fry, M., and Loeb, L. A. (1999) Human Werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J. Biol. Chem. 274*, 12797−12802.

(12) Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., and Lipps, H. J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol. 12*, 847−854.

(13) Oganesian, L., and Bryan, T. M. (2007) Physiological relevance of telomeric G-quadruplex formation: a potential drug target. *Bioessays 29*, 155−165.

(14) Lipps, H. J., and Rhodes, D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol. 19*, 414−422.

(15) Kim, N., Piatyszek, M., Prowse, K., Harley, C., West, M., Ho, P., Coviello, G., Wright, W., Weinrich, S., and Shay, J. (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science 266*, 2011−2015.

(16) Neidle, S., and Parkinson, G. (2002) Telomere maintenance as a target for anticancer drug discovery. *Nat. Rev. Drug Discovery 1*, 383−393.

(17) 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.

(18) Balasubramanian, S., Hurley, L. H., and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug. Discovery 10*, 261−275.

(19) Wong, H. M., Payet, L., and Huppert, J. L. (2009) Function and targeting of G-quadruplexes. *Curr. Opin. Mol. Ther. 11*, 146−155.

(20) Murat, P., Singh, Y., and Defrancq, E. (2011) Methods for investigating G-quadruplex DNA/ligand interactions. *Chem. Soc. Rev. 40*, 5293−5307.

(21) Teng, Y., Girvan, A. C., Casson, L. K., Pierce, W. M., Qian, M., Thomas, S. D., and Bates, P. J. (2007) AS1411 alters the localization of a complex containing protein arginine methyltransferase 5 and nucleolin. *Cancer Res. 67*, 10491−10500.

(22) Soundararajan, S., Chen, W., Spicer, E. K., Courtenay-Luck, N., and Fernandes, D. J. (2008) The nucleolin targeting aptamer AS1411 destabilizes Bcl-2 messenger RNA in human breast cancer cells. *Cancer Res. 68*, 2358−2365.

(23) Alberti, P., Bourdoncle, A., Sacca, B., Lacroix, L., and Mergny, J.-L. (2006) DNA nanomachines and nanostructures involving quadruplexes. *Org. Biomol. Chem. 4*, 3383−3391.

(24) Sen, D., Fahlman, R. P. DNA conformational switches as sensitive electronic sensors of analytes. U.S. Patent 7,943,301 B2, 2011.

(25) Stegle, O., Payet, L., Mergny, J.-L., MacKay, D. J. C., and Huppert, J. L. (2009) Predicting and understanding the stability of G-quadruplexes. *Bioinformatics 25*, i374−i382.

(26) Parkinson, G. N., Lee, M. P. H, and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature 417*, 876−880.

(27) Vorlíčková, M., Chládková, J., Kejnovská, I., Fialová, 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.

(28) 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.

(29) Pedroso, I. M., Duarte, L. F., Yanez, G., Burkewitz, K., and Fletcher, T. M. (2007) Sequence specificity of inter- and intramolecular G-quadruplex formation by human telomeric DNA. *Biopolymers 87*, 74−84.

(30) Petraccone, L., Trent, J. O., and Chaires, J. B. (2008) The tail of the telomere. *J. Am. Chem. Soc. 130*, 16530−16532.

(31) Haider, S., Parkinson, G. N., and Neidle, S. (2008) Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophys. J. 95*, 296−311.
(32) Martadinata, H., and Phan, A. T. (2009) Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K+ solution. *J. Am. Chem. Soc. 131*, 2570−2578.
(33) Collie, G. W., Parkinson, G. N., Neidle, S., Rosu, F., De Pauw, E., and Gabelica, V. (2010) Electrospray mass spectrometry of telomeric RNA (TERRA) reveals the formation of stable multimeric G-quadruplex structures. *J. Am. Chem. Soc. 132*, 9328−9334.
(34) Martadinata, H., Heddi, B., Lim, K. W., and Phan, A. T. (2011) Structure of long human telomeric RNA (TERRA): G-quadruplexes formed by four and eight UUAGGG repeats are stable building blocks. *Biochemistry 50*, 6455−6461.
(35) Bauer, L., Tlučková, K., Tóthová, P., and Viglaský, V. (2011) G-quadruplex motifs arranged in tandem occurring in telomeric repeats and the insulin-linked polymorphic region. *Biochemistry 50*, 7484−7492.
(36) Phan, A. T., Kuryavyi, V., Gaw, H. Y., and Patel, D. J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol. 1*, 167−173.
(37) Kumar, N., and Maiti, S. (2008) A thermodynamic overview of naturally occurring intramolecular DNA quadruplexes. *Nucleic Acids Res. 36*, 5610−5622.
(38) Todd, A. K., and Neidle, S. (2011) Mapping the sequences of potential guanine quadruplex motifs. *Nucleic Acids Res. 39*, 4917−4927.
(39) Kejnovská, I., Kypr, J., and Vorlíčková, M. (2007) Oligo(dT) is not a correct native PAGE marker for single-stranded DNA. *Biochem. Biophys. Res. Commun. 353*, 776−779.
(40) Mergny, J., and Lacroix, L. (2003) Analysis of thermal melting curves. *Oligonucleotides 13*, 515−537.
(41) Risitano, A., and Fox, K. R. (2003) Stability of intramolecular DNA quadruplexes: comparison with DNA duplexes. *Biochemistry 42*, 6507−6513.
(42) Brown, N. M., Rachwal, P. A., Brown, T., and Fox, K. R. (2005) Exceptionally slow kinetics of the intramolecular quadruplex formed by the Oxytricha telomeric repeat. *Org. Biomol. Chem. 3*, 4153−4157.
(43) Merkina, E. E., and Fox, K. R. (2005) Kinetic stability of intermolecular DNA quadruplexes. *Biophys. J. 89*, 365−373.
(44) Rachwal, P., Brown, T., and Fox, K. (2007) Sequence effects of single base loops in intramolecular quadruplex DNA. *FEBS Lett. 581*, 1657−1660.
(45) Tang, J., Kan, Z.-y., Yao, Y., Wang, Q., Hao, Y.-h., and Tan, Z. (2008) G-quadruplex preferentially forms at the very 3′ end of vertebrate telomeric DNA. *Nucleic Acids Res. 36*, 1200−1208.
(46) Dai, J., Dexheimer, T. S., Chen, D., Carver, M., Ambrus, A., Jones, R. A., and Yang, D. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc. 128*, 1096−1098.
(47) Henson, J. D., Neumann, A. A., Yeager, T. R., and Reddel, R. R. (2002) Alternative lengthening of telomeres in mammalian cells. *Nature 21*, 598−610.
(48) Rivera, M. A., and Blackburn, E. H. (2004) Processive utilization of the human telomerase template. *J. Biol. Chem. 279*, 53770−53781.