2386–2397 *Nucleic Acids Research, 2006, Vol. 34, No. 8*
doi:10.1093/nar/gkl098

Kinetics of double-chain reversals bridging contiguous quartets in tetramolecular quadruplexes

Jean-Louis Mergny*, Anne De Cian, Samir Amrane and Mateus Webba da Silva[1,2,*]

Laboratoire de Biophysique, Muséum National d'Histoire Naturelle USM503, INSERM U565, CNRS UMR 5153, 43 rue Cuvier, 75231 Paris cedex 05, France, [1]School of Biomedical Sciences, University of Ulster, Coleraine BT52 1SA, Northern Ireland, UK and [2]Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA

Received January 23, 2006; Revised February 22, 2006; Accepted March 7, 2006

ABSTRACT

Repetitive 5′GGXGG DNA segments abound in, or near, regulatory regions of the genome and may form unusual structures called G-quadruplexes. Using NMR spectroscopy, we demonstrate that a family of 5′GCGGXGGY sequences adopts a folding topology containing double-chain reversals. The topology is composed of two bistranded quadruplex monomeric units linked by formation of G:C:G:C tetrads. We provide a complete thermodynamic and kinetic analysis of 13 different sequences using absorbance spectroscopy and DSC, and compare their kinetics with a canonical tetrameric parallel-stranded quadruplex formed by TG_4T. We demonstrate large differences (up to 10^5-fold) in the association constants of these quadruplexes depending on primary sequence; the fastest samples exhibiting association rate equal or higher than the canonical TG_4T quadruplex. In contrast, all sequences studied here unfold at a lower temperature than this quadruplex. Some sequences have thermodynamic stability comparable to the canonical TG_4T tetramolecular quadruplex, but with faster association and dissociation. Sequence effects on the dissociation processes are discussed in light of structural data.

INTRODUCTION

Guanine-rich regions abound in the human genome; they have the propensity to fold into higher order DNA structures such as quadruplexes (1,2). These architectures result from the hydrophobic stacking of several quartets (3–5) and may have applications in areas ranging from supramolecular chemistry to medicinal chemistry [for a recent review see (6)]. Rules have been proposed to describe the properties of simple short segments that do not form loops such as $T_2G_4T_2$ (7). However, G-quadruplexes are in fact highly polymorphic (8,9). This polymorphic nature is in part due to formation of double-chain reversals loops, which span two or more contiguous tetrads resulting in parallel stranded architectures. Formation, stabilization, dynamic behavior as well as structure–function and structure–stability relationships of these quadruplexes are of current interest (10–15). Recent reports established that d-5′GCGG**T**GGAT (16) and d-5′GCGG**TT**GGAT (17) adopt a topology in which two bistranded quadruplex monomeric units are interlocked by formation of four-stranded G:C:G:C tetrads. A schematic representation of the four-stranded topology as well as core hydrogen bond alignments of the stem quadruplex are shown in Figure 1A–C.

In this article, we demonstrate that the same global folding pattern may apply to a rather large sequence family 5′GCGGXGGY (where X and Y may correspond to one or two bases). We first provide a structural analysis of the GCGG**A**GG**Y** motif, then analyze the contribution of various base substitutions on the kinetics and thermodynamics affecting the global architecture of these quadruplexes in view of structural data.

MATERIALS AND METHODS

Nomenclature, synthesis and purification of oligonucleotide sequences

Oligodeoxynucleotides have the general sequence 5′-GCGGXGGY-3′ where X = T, U, TT, UC, TC or A and Y = AT, AU, T or A. (U is a deoxyuracyl; all sugars are deoxyriboses). All oligonucleotides with X = T or U belong to the serie '1' and are labeled 1, 1B, 1UA and 1UB, depending on the nature of the 3′ base(s) (Table 1). All oligonucleotides with X = TT or UU belong to the serie '3' and are labelled **3**, 3A and 3AU. Oligonucleotides with a 'TC' core are labeled **4**,

*To whom correspondence should be addressed. Jean-Louis Mergny: Tel: +33 1 40 79 36 89; Fax: +33 1 40 79 37 05; Email: faucon@mnhn.fr
*Correspondance may also be addressed to M. Webba da Silva. Tel: +44 28 7032 4009; Fax: +44 28 7032 4375; Email: mm.webba-da-silva@ulster.ac.uk

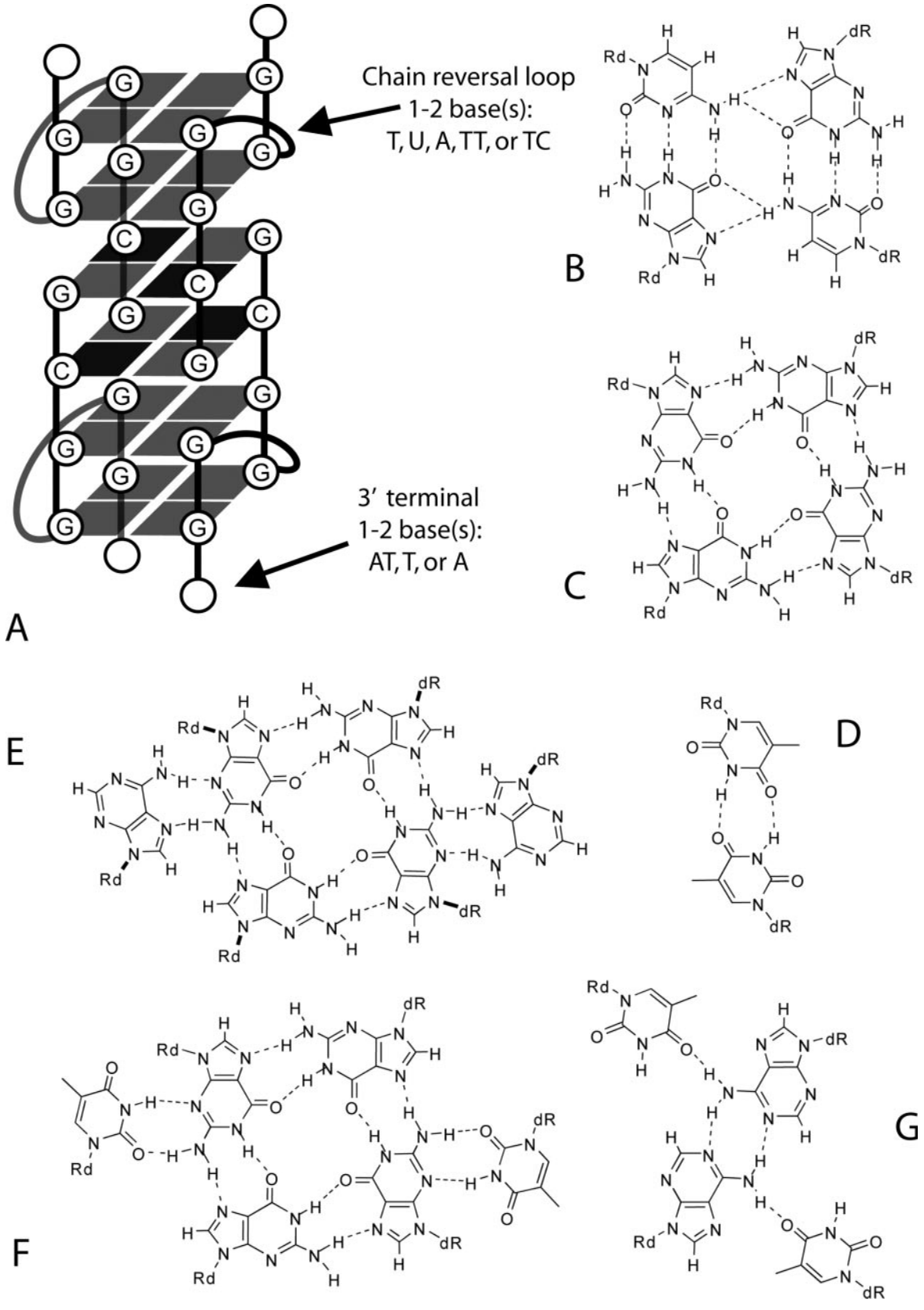


Figure 1. Folding topology and pairing schemes. (**A**) Scheme of the general folding topology of the four-stranded bimolecular quadruplexes. The backbone tracing of the individual strands is shown by thick lines and non-guanine bases are depicted in black. All bases of the stem quadruplex adopt *anti* alignments. Pairing alignments for the two-stranded G:G:G:G tetrads (**B**) and the four-stranded G1:C2:G1:C2 tetrad (**C**) are also shown. The T:T mismatch present in the sequences **1**, **3** and **2** (**D**), the two-stranded hydrogen bonding alignments for A5(G3:G6:G3:G6):A5 in **2**, (**E**), T6(G3:G7:G3:G7):T6 hexad (**F**) and T5:A9:A9:T5 tetrad (**G**) in **3** are depicted.

Table 1. Oligonucleotides

Name (5′→3′)	Sequence	k_{on}[a] (M^{-3} s^{-1})	$k_{on}^{(rel)}$[b]	E_{on}[c] (kcal.M^{-1})	$T_{1/2}$[d] (°C) (Na^+)	Gel[e]	ΔH°_{cal}(kcal M^{-1})[f]
TG_4T	TGGGGT	**1.8×10^8**	**1**	**−29 ± 2**	**57.4**	+	**−72**
1	GCGGTGGAT	1.3×10^9	7	−31 ± 3	43.0	+	−96 ± 5
1B	GCGGTGGT	5.8×10^7	0.3	nd	36.0	nd	−101 ± 5
1UA	GCGGUGGAT	1.1×10^9	6	−20 ± 2	48.9	+	−94 ± 5
1UB	GCGGTGGAU	1.1×10^9	6	−26 ± 2	41.5	+	nd
2	GCGGAGGAT	2.6×10^5	0.0014	+36 ± 7	37	+/−	−61 ± 12
3	GCGGTTGGAT	1.7×10^8	0.9	−29 ± 5	33.6	+	−103 ± 5
3A	GCGGTTGGA	1.3×10^8	0.7	+6 ± 6	31.6	+	nd
3U	GCGGUUGGAT	3.6×10^8	2	−26 ± 6	37.2	+	−96 ± 5
4	GCGGTCGGAT	2.8×10^5	0.0016	+42 ± 4	30.5	+/−	nd
4U	GCGGUCGGAT	1.1×10^6	0.006	+32 ± 7	31.8	+/−	nd
4AU	GCGGUCGGA	1.5×10^5	0.0008	nd	31.8	+/−	nd
4A	GCGGTCGGA	$<10^5$	$<10^{-4}$	nd	≈24	−	nd
4B	GCGGTCGGT	$<10^5$	$<10^{-4}$	nd	≈22	−	nd

nd, not done.
[a]Association rate constant at 8°C, pH 7, with 0.11 M Na^+, in M^{-3} s^{-1}. k_{on} is given with a ± 30% accuracy or better.
[b]Relative association rate constant at 8°C, as compared with TG_4T.
[c]Activation energy of association (±SD). As the dissociation temperature found for all oligonucleotides was lower than for TG_4T, it was difficult to measure k_{on} values above 20°C in a number of cases; hence E_{on} values derive from association rates on a relatively narrow T-range.
[d](Non-equilibrium) melting temperature of the preformed quadruplex, in (°C), in 0.11 M Na^+, determined with a temperature gradient of 0.45°C/min. $T_{1/2}$ is provided with a 0.5°C accuracy or better.
[e]Evidence for quadruplex formation on a non-denaturing gel. +, single retarded band when incubated in 0.11 M Na^+. +/−, smear (dissociation of the complex in the gel; migration at low temperature mandatory to observe the complex) or trace amounts. −, no retarded band, even after a 72 h incubation at 4 or 17°C at 500 μM strand concentration).
[f]Model-independent enthalpies determined by DSC.

4A, **4B**, **4U** and **4AU**. A single oligonucleotide (2) with a central '**A**' is tested here. Sequences were synthesized by Eurogentec (Seraing, Belgium) or as described previously (17). Concentrations of all oligodeoxynucleotides were estimated using published sequence-dependent extinction coefficients (18); values are provided in Supplementary Table S3.

Absorbance measurements

Absorbance was recorded as described previously (19). Most melting curves recorded by heating a preformed quadruplex do not correspond to equilibrium melting curves (hysteresis phenomenon) and the 'T_m' deduced from these experiments depends on the heating rate. In order to distinguish it from the true thermodynamic T_m, we will call this value $T_{1/2}$.

Differential scanning calorimetry

Microcalorimetry experiments were performed with a Nano DSC-II microcalorimeter (CSC) as described previously (20).

NMR spectroscopy

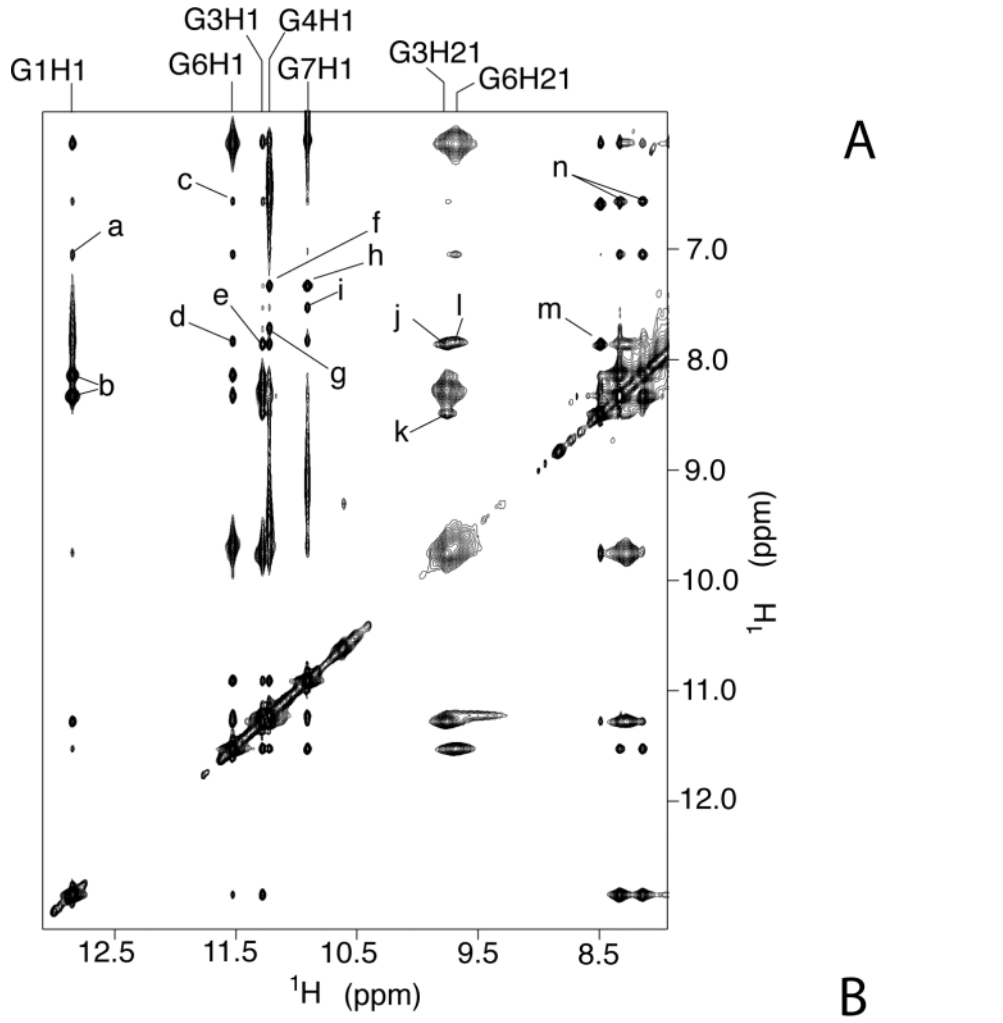
All experiments were performed on Varian Unity Inova NMR 500 and 600 MHz spectrometers. The datasets were processed using Varian software (VNMR) and the real matrices were transformed to the FELIX (Accelrys, Inc., San Diego, CA) format on a Silicon Graphics (Mountain View, CA) Octane workstation. Typically, assignments were based on homonuclear NOESY (21), DQF-COSY (22), TOCSY (23), and heteronuclear (^{1}H-^{13}C)HSQC (24), (^{1}H-^{15}N)HSQC (25) and (^{1}H,^{13}C) JR-HMBC (26) experiments. The final concentration of DNA in the sample was about 6 mM (∼1.5 mM per strand).

RESULTS

Topological architecture

Determination of the 3D solution structure of all sequences presented in Table 1 is beyond the scope of this study. Nevertheless, before comparing the kinetics and thermodynamics of these oligomers, we deemed necessary to establish that these sequences have the same global architecture. Quadruplex formation was thus analyzed by PAGE and revealed by an anomalously slow migrating band as compared with the migration pattern of the same 'single-stranded' oligomer (Supplementary Figure S1). Complete or near complete conversion into a lower mobility band was obtained with most sequences. However, in the case of **4** and its variants very high concentrations (mM range) were required to obtain quadruplex formation. Furthermore, the isothermal thermal difference spectra of all these structures (Supplementary Figure S2) were in agreement with the formation of quadruplexes (27,28). However, the actual shape of these spectra was not strictly identical to that of a canonical G-quadruplex (19), indicating also the formation of other base folding patterns (see below).

The folding topology is based on studies on two of these sequences (**1** and **3**) (16,17). The exchangeable proton spectral regions for **2** and **4B** point to stable structures at 0°C. For these sequences the number of resonances in the imino region coincides with the number of guanines in the respective sequence, consistent with symmetry related multistranded architectures. The topology was primarily established through observation of analogous connectivities in the region connecting aromatic protons to H1′ in NOESY spectra (Supplementary Figures S3 and S4). Also, typical dipolar connectivities connecting the imino and amino exchangeable protons to the aromatic were consistent with formation of pseudo-planar elements. For example, in **2** the dipolar connectivities between the imino


A
G1H1
G3H1
G4H1
G6H1
G7H1
G3H21
G6H21
a
b
c
d
e
f
g
h
i
j
k
l
m
n
7.0
8.0
9.0
10.0
11.0
12.0
12.5
11.5
10.5
9.5
8.5
^{1}H (ppm)
^{1}H (ppm)
B
A5H2
A8H2
(a)
(b)
(c)
(d)
G1
T9
A8
G6
G7
G3
G4
C2
A5
5.4
5.6
5.8
6.0
6.2
6.4
6.6
8.2
7.8
7.4
7.0
6.6
^{1}H (ppm)
^{1}H (ppm)

of G6, G6H1 and G3H8 (peak d) and between G3H1 and G6H8 (peak e) protons are compatible with G3:G6:G3:G6 tetrad formation. The hydrogen bond alignment of the pseudo-planar architecture is further demonstrated through dipolar connectivities between $G6NH_2$ and G3H8 (peak l) and $G3NH_2$ to G6H8 (peak j). In addition, we observe an NOE between the $G3NH_2$ and A5H8 protons (peak k), consistent with recognition of the minor groove edge of G3 by the Watson–Crick edge of A5 resulting in the A5(G3:G6:G3:G6):A5 hexad alignment (Figure 1E). Accordingly, the amino protons of A5 resonate at 10.37 and 9.30 p.p.m. (Supplementary Figure S5A). These broadened amino protons are downfield shifted relative to their counterparts in Watson–Crick A:T pairs, which generally resonate between 7.5 and 8.0 p.p.m.

Analysis of through bond and through space correlations in D_2O buffer solution resulted in assignments for base and sugar protons of the **2** quadruplex. In a plot of a NOESY collected with mixing time of 250 ms we can trace the sequential connectivities between the base and its own and 5′-flanking sugar H1′ protons along individual strands (Figure 2B). For all oligonucleotides, weak intraresidue H8/H6-H1′ cross-peaks were observed indicating *anti* glycosidic torsion angles (χ).

In Figure 3 we depict the architecture formed by **2** as well as the tridimensional structures of **1**, and **3**. From the top view the overall topology of **2** appears regular cylindrical. However, **1** and **3** deviate from regular toroid due to the accommodation of the residues in the double chain reversal (bottom view). Whilst in **2** the one-residue looping base recognizes the sheared edge of a stem guanine, in **1** the thymine aligns with a quadruplex groove. The more salient backbone of **3** is due to accommodation of a two-residue loop.

Association of the isolated strands at low temperature: effects of concentration

Isothermal renaturation experiments were used to study the formation of the quadruplexes; representative examples are provided in Figure 4A and B for sample **1**. We first determined that the low temperature renaturation was minimal for the DNA sample at concentrations below 20 μM. In contrast, formation of the quadruplex at 3°C was nearly complete in 2 h at 80 μM (Figure 4B), demonstrating that concentration played a dramatic role in the kinetics of association. It was possible to fit the experimental profiles for the determination of the order n of the reaction. Data could be fitted with $3.3 \leq n \leq 3.9$; we defined $n = 4$ for all further studies. These fits (dotted lines) are in nearly-perfect agreement with the experimental points. Moreover, the k_{on} values determined from the curves at three different concentrations or at two different wavelengths (240 and 295 nm) were in excellent agreement [as seen in a dual wavelength parametric test (29); Supplementary Figure S2]. The association rate constant at 8°C was 1.3×10^9 $M^{-3}.s^{-1}$ for **1** (Table 1), i.e. seven times faster than TG_4T that we used as a reference (19).

Sequence and temperature effects on association

Next, we performed the same renaturation experiment at various temperatures. As shown in Figure 4C (for **3U**) and Supplementary Figure S6A (for **1**), an increase in temperature has a deleterious impact on the kinetics of association: at a given strand concentration, folding was slower at 12°C than at 3°C. The association rates are sumed up in Figure 4D and Supplementary Figures S6B–D (Arrhenius plots). At low temperature (<10°C), sample **1** folded 10^5 times faster than **4**. At 21°C, this difference was less pronounced, because of a different temperature dependency of k_{on}. The oligonucleotides could be subdivided into two distinct families:

(i) **1** and **3** (and their variants **1B**, **1UB**, **1UA** and **3U**) have a high association rate, especially at low temperature. Their association constant is either similar or higher (for **1**, **1UB** and **1UA**) than the k_{on} of TG_4T (Table 1). As most sequences have comparable negative activation energies of association (E_{on}), they have relatively similar temperature dependencies for k_{on}.

(ii) **2**, **3A**, **4** and **4U** have a low association rate, especially at low temperature. At 8°C, their association constant is roughly 1000-fold slower than the k_{on} of TG_4T (Table 1). The most striking difference with the first category is their opposite temperature dependencies: they have comparable positive activation energies of association (E_{on}). This puzzling behavior could be the result of the formation of competitive structure(s) (Figure 6C and Supplementary Data).

From the data shown in Table 1, one may rank the various sequences in decreasing order of k_{on}:

1 > **3** = TG_4T > **4** > **2**. A central T→U substitution generally leads to a faster association; it is clear that **3U** folds significantly faster than **3**, demonstrating that the central **TT→UU** substitution accelerates the association process. A similar phenomenum was found for **4U** as compared with **4**, demonstrating that the central **TC→UC** substitution also accelerates the association process.

Dissociation of the preformed quadruplexes: sequence effects

Starting from preformed quadruplexes (several days at 0°C and high strand concentration 100–1000 μM), one can simply follow the denaturation of this structure by recording the absorbance at 240 or 295 nm (19,20) (Figure 5A). The curve of the hypochromic effect produced upon heating

Figure 2. NOESY for 2, d(GCGGAGGAT). (**A**) Expanded JR-NOESY spectrum (200 ms) of 2, d(GCGGAGGAT). JR-NOESY spectrum (200 ms) in 100 mM NaCl, 2 mM sodium phosphate buffer at pH 6.7 in ^1H_2O at 0°C, exhibiting dipolar connectivities between imino and aromatic protons. Peaks a to m are assigned as follows: a, G1H1–C2H6; b, G1H1–$C2NH_2$; c, G6H1–G1H8; d, G6H1–G3H8; e, G3H1–G6H8; f, G4H1–A8H2; g, G4H1–G7H8; h, G7H1–A8H2; i, G7H1–G4H8; j, $G3NH_2$–G6H8; k, $G3NH_2$–A5H8; l, $G6NH_2$–G3H8; m, A5H8–G6H8; n, $C2NH_2$–G1H8. (**B**) NOESY for 2, d(GCGGAGGAT). An expanded NOESY (250 ms mixing time) spectrum recorded in ^2H_2O at 20°C correlating the aromatic base proton and sugar H1′ for **2**, d[(GCGGAGGAT)$_2$]$_2$. The lines trace the NOE connectivities between the base protons and their own and 5′-flanking sugar H1′ protons from G1 to T9 in the sequence. Both peaks (b), G6H8–G1H1′, and (d), G1H8–G6H1′, are crucial assignments leading to the identification of the architecture of the quadruplex formed by the sequence **2** d(GCGGAGGAT) in Na^+ salt solution. The corresponding cross peaks indicating the folding of tetrameric interlocked quadruplex appear in both **1** and **3**. Crosspeak (a) shows the strong NOE between the aromatic H8 and the very lowfield shifted H3′ of adenine A5 indicative of C3′-endo sugar puckering, and cross-peak (c) is sequential aromatic-aromatic.

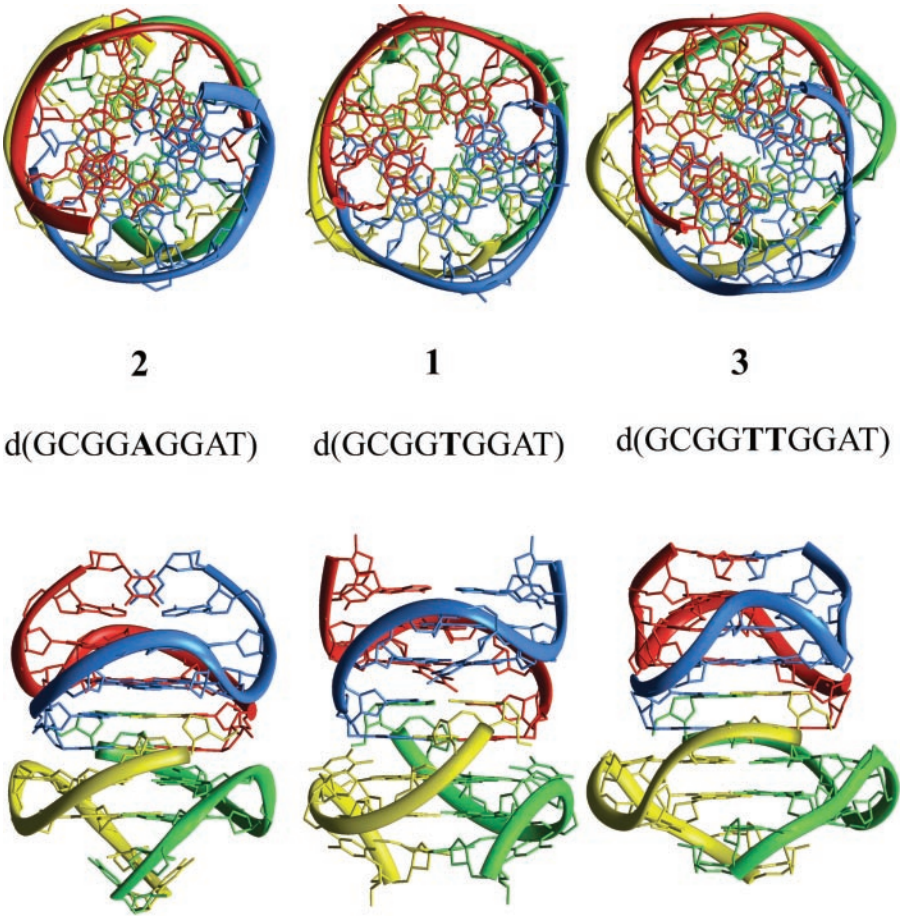


Figure 3. Views of the architectures formed by **2**, d(GCGGAGGAT); **1**, d(GCGGTGGAT); and **3**, d(GCGGTTGGAT). On top views from the top through the quadruplex stem axis depicting deformation of the DNA toroid. On the bottom corresponding views across the stem axis. The model for d(GCGGAGGAT) **2** was derived from substitution of T5 for A5 in the structure formed by **1**: d(GCGGTGGAT) in 0.1 M NaCl; PDB ID, 1NYD (16), followed by energy minimizations under Insight II (Accelrys Inc.). The Hoogsteen edge of A5 was brought to hydrogen bonding distance, keeping an anti glycosidic bond angle using the Biopolymer module of Insight II (Accelrys).

appears to be that of a typical cooperative melting transition. However, this melting transition does not represent an equilibrium denaturation process, for upon cooling, renaturation of the DNA quadruplex is very slow (Figure 5B). Furthermore, this apparent melting temperature does not depend on oligonucleotide concentration but strongly depends on the rate of heating (Figure 5C and Supplementary Figure S7; a 10–20°C difference in $T_{1/2}$ may be found between the fastest and the slowest gradient), again indicating that this profile does not correspond to an equilibrium curve but solely reflects the dissociation of the quadruplex. In this temperature range, what we see in practice is a simple one-way reaction from a folded quadruplex to a dissociated state: differences in $T_{1/2}$ reflect differences in thermal lability (19). We were therefore able to extract dissociation rate constant (k_{off}) values in the 20–50°C temperature range. These values were plotted for **1** and **3** against $1/T$ and compared with the values found for TG_4T (Figure 5D). A quasi linear relation between $\ln(k_{off})$ and $1/T$ was found, corresponding to activation energies of dissociation (E_{off}) of 62 ± 3 and 65 ± 4 kcal/mol for **1** and **3**, respectively, as compared with 43 ± 1 kcal/mol for TG_4T. At all temperatures above 20°C, **1** and **3** have higher dissociation rate constants than the canonical quadruplex: the quadruplex lifetimes at 37°C of **1** and **3** were 16 and 400 times shorter than the lifetime of the TG_4T tetramer, respectively. All other oligonucleotides are significantly less stable than the TG_4T quadruplex as shown by a lower $T_{1/2}$ (Table 1). Within each family, one may find interesting results concerning thymine to uracyl substitution. When a central thymine is replaced by an uracyl (T→U for **1**, TT→UU for **3** and TC→UC for **4**) a significant increase in $T_{1/2}$ is found (+5.9, 3.6 and 1.3°C, respectively). We found the following stability order TG_4T > **1** > **2** > **3** > **4**. This ranking is therefore different from the one found on association.

Ionic strength and pH effects

All profiles shown before were obtained in a pH 7.2 buffer containing Na^+ ions at 110 mM. We investigated the effects of changing buffer conditions on the **1** quadruplex GCGGTGGAT. k_{on} and $T_{1/2}$ were only weakly dependent on pH in the 6.0–7.8 range (Figure 4E for association, Figure 5E for thermal stability). Changing the pH from 7.8–6.0 led to a 1.2- and 2-fold increase in association for TG_4T and **1**, respectively. A similar pH change had no effect on the thermal stability of **1** ($T_{1/2}$ = 42.5 at pH 6.0 and 7.8).

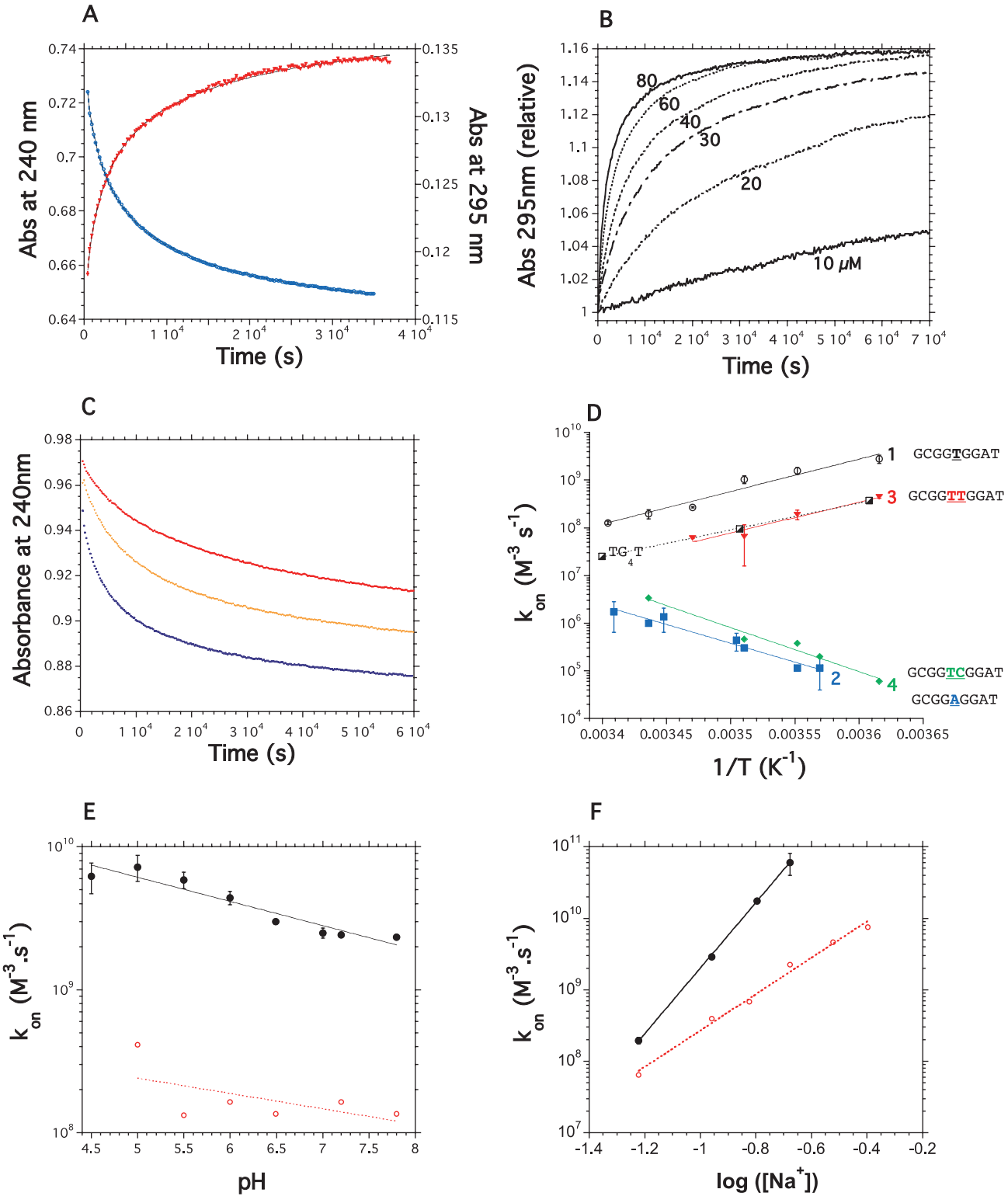


Figure 4. Absorbance data on association. (**A**) Experimental data at two wavelengths (240 and 295 nm) and fitted functions (solid lines). **1** oligonucleotide at 60 μM strand concentration in a 0.11 M Na^+, pH 7.2 buffer at 3°C. The fitted k_{on} values were 1.62×10^9 and 1.81×10^9 $M^{-3}.s^{-1}$ at 240 and 295 nm, respectively. (**B**) Effect of strand concentration (10–80 μM) for **1** at 3°C in 0.11 M Na^+ on the relative absorbance at 295 nm. Quadruplex formation leads to an increase in absorbance at this wavelength. (**C**) Effect of temperature on association for **3U** at 3°C (blue), 8°C (orange) and 12°C (red) in 0.11 M Na^+ on the absorbance at 240 nm. Quadruplex formation leads to a decrease in absorbance at this wavelength. (**D**) Arrhenius plots for the association of **1** (circles), **2** (triangles), **3** (squares), **4** (closed squares) and TG_4T (squares; dotted line) oligonucleotides. (**E**) Effect of pH on association. k_{on} is plotted versus pH for **1**, d(GCGGTGGAT, closed circles) and dTG_4T (red symbols). Solid and dotted lines, exponential fits (note the *Y*-axis log scale) for the data points. (**F**) Effect of ionic strength on association. k_{on} is plotted versus [Na^+] for **1**, d(GCGGTGGAT, closed circles) and dTG_4T (red symbols). These experiments were performed in a 10 mM lithium cacodylate pH 7.2 buffer with various concentrations of NaCl.

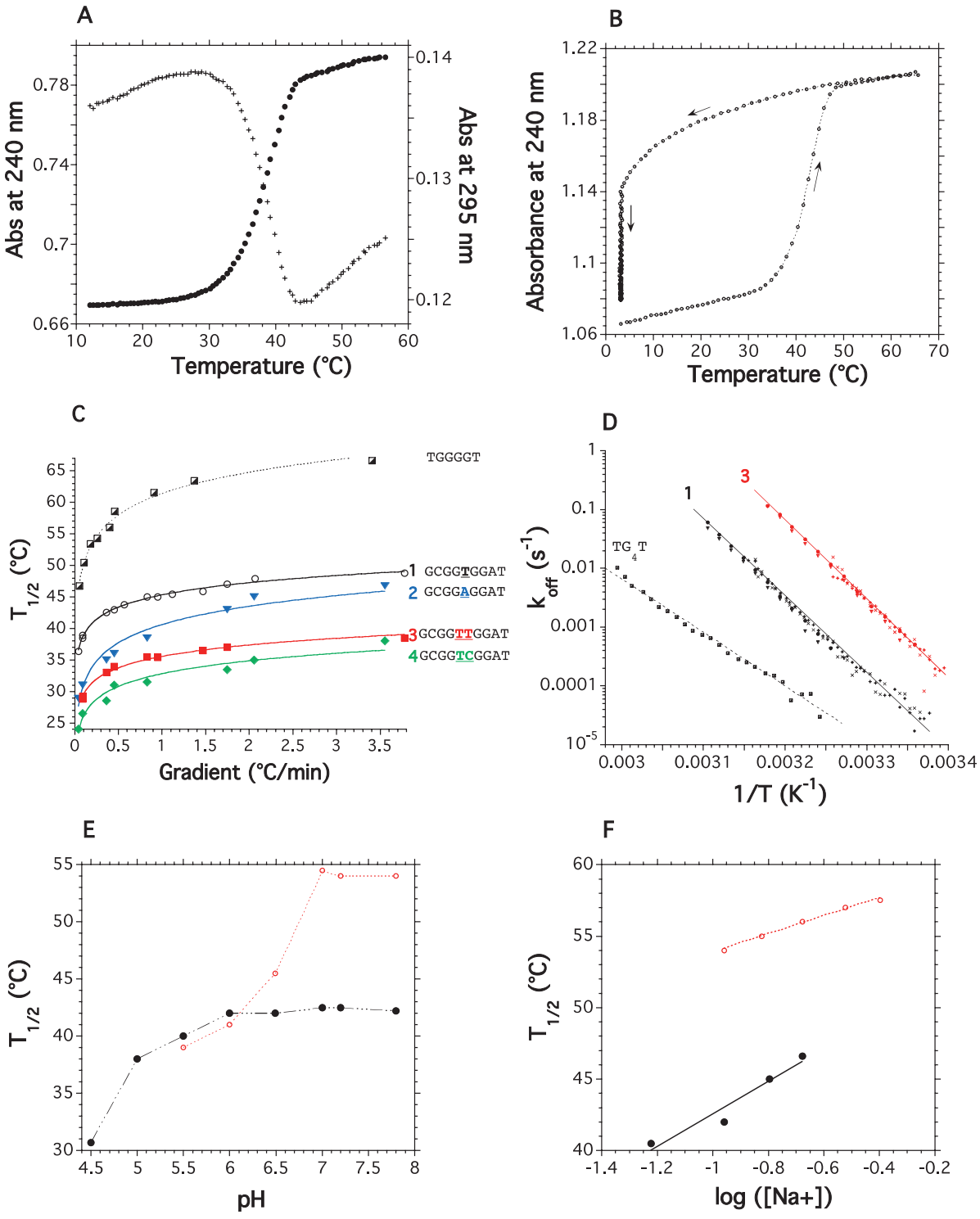


Figure 5. Absorbance data on thermal stability. (**A**) Melting profile of **1** oligonucleotide in 0.11 M Na^+. Absorbance at 295 nm (crosses) or 240 nm (circles) is recorded every 4 min with a thermal gradient of ~0.09°C/min in a pH 7.0, 10 mM sodium cacodylate buffer supplemented with 0.1 M NaCl. (**B**) Hysteresis in the denaturation/renaturation process of **1** in 0.11 M Na^+ at 20 μM strand concentration (temperature gradient 0.278°C/min). Arrows indicate directions of temperature changes. At this strand concentration, once melted, the tetramolecular quadruplex does not refold upon cooling, but a very slow renaturation is observed at 3°C after the cooling experiment. (**C**) $T_{1/2}$ of the **1** (circles), **2** (triangles) **3** (squares) and **4** (diamonds) quadruplexes as a function of the average temperature gradient. Note that the 'true' equilibrium T_m, which would be obtained with an infinitely slow gradient cannot be approached by this method, only an upper limit may be determined. (**D**) Arrhenius plots for the dissociation of **1** (black symbols), **3** (red symbols) and TG_4T (squares; dotted line) quadruplexes. Different symbols (squares, crosses and so on) correspond to independent melting profiles recorded at 240 or 295 nm using different temperature gradients. (**E**) Effect of pH on $T_{1/2}$ for **1**, d(GCGGTGGAT, closed circles) and dTG_4T (red symbols). Values determined for a heating temperature gradient of 0.48°C/min. (**F**) Effect of ionic strength on $T_{1/2}$ for **1**, d(GCGGTGGAT, circles) and dTG_4T (red symbols). Values determined for a heating temperature gradient of 0.48°C/min. These experiments were performed in a 10 mM lithium cacodylate, pH 7.2, buffer with various concentrations of NaCl.

Various concentrations of sodium were also tested. Increasing ionic strength played a significant role in the thermal dissociation of the **1** quadruplex (Figure 5F; the higher the salt concentration, the higher the $T_{1/2}$). Varying the sodium concentration had an even more dramatic effect on the association process (Figure 4F). Raising the NaCl concentration from 60 to 210 mM leads to a 6°C increase in $T_{1/2}$ and 300-fold increase in the association constant. For both parameters (k_{on} and $T_{1/2}$), one may note that the effect is more pronounced on GCGG**T**GGAT than on TG_4T: in the latter case, the same increase in NaCl concentration leads to a 35-fold faster association process.

Thermodynamics of the 1–4 quadruplexes

Qualitatively, from the data on association (**1** ≈ **1UA** > **3** > **4** > **2**) and dissociation (**1UA** > **1** > **2** > **3** > **4**), one may easily deduce that **1UA** forms the thermodynamically most stable complex, immediately followed by **1**. It is interesting to compare the thermodynamics of **1** with the canonical TG_4T tetramolecular parallel quadruplex. At 37°C, the quadruplex formed by **1** unfolds 16 times faster than the TG_4T tetramer. On the other hand, **1** folds approximately seven times faster than TG_4T. As a result, one can deduce that both quadruplexes have relatively similar ΔG° at 37°C, with a 2-fold difference in the association equilibrium constants $K_a = k_{on}/k_{off}$. However, one should note that this calculation is valid only at 37°C (these structures have different enthalpies), and that experimental points for association and dissociation of **1** and TG_4T are not determined in the same temperature range (for **1**, 2–21°C for association, 25–50°C for dissociation). Fortunately, a reasonable linear fit may be obtained for most association and dissociation rate constants (Figures 4D and 5D): their slopes allow the determination of the association activation energies (E_{on}, Table 1 and E_{off}) with some confidence.

It is more difficult to rank the other sequences, as contrasting effects were found on association and dissociation. **2** and **4** are thermodynamically less stable than **1** and TG_4T at all temperatures, but the difference in K_a is more pronounced at low temperatures. For example, at 8°C, there is a 10^4–10^5 difference in K_a between **4** and **1** (which results both from fastest dissociation and much slower association of the former). At 20°C, there is a ≈300-fold difference only, as a result of opposite effects of temperature on k_{on} for the two sequences.

In order to better understand the thermodynamics of the quadruplexes, we performed DSC experiments on several different sequences. This technique allows us to evaluate the energetics of G-quadruplex and compare the impact of base substitution on the enthalpy of formation. Examples of heating/cooling experiments are presented in Figure 6. Similar to the UV-melting experiments, a large hysteresis is obtained. This phenomenon is not unexpected: at a typical scan rate of 1°C/min, reformation of the quadruplex upon cooling occurs at a lower temperature. The profile for **4** (GCGG**TC**GGAT) is even more complex (Figure 6C); one may see two distinct transitions when heating a preformed sample (several days at 4°C). The area under the first melting curve allows a model-independent determination of ΔH° for these sequences. Assuming 97 and 91% quadruplex formation before the onset of the first DSC run for 1 and 3, respectively (based on the k_{on} values found in Table 1), the enthalpy values found for **1** and **3** quadruplexes were −96 and −103 kcal/mol quadruplex, respectively. It is interesting to compare these numbers with the model-dependent ΔH° values deduced from the thermokinetic analysis ($\Delta H^\circ = E_{on} - E_{off}$). The values found for **1** and **3** quadruplexes were −93 ± 6 and −94 ± 9 kcal/mol, respectively, in excellent agreement with the model-independent DSC values. The formation of **1** and **3** quadruplexes are therefore strongly enthalpy-driven, with an even more negative ΔH° than TG_4T (72 kcal/mol under identical conditions). This result is hardly surprising as **1** and **3** quadruplexes not only involve the formation of 4 G-quartets (as TG_4T) but also 2 (G:C:G:C) quartets. DSC measurements also allow us to rank the samples according to their enthalpy:

3 ≈ **1** ≈ **1B** > **3U** ≈ **1UA** ≫ TG_4T ≫ **2**.

DISCUSSION

We have presented evidence suggesting that the global folding pattern established for GCGG**T**GGAT (1) and GCGG**TT**GGAT (3) may indeed be extended to a large family of sequences GCGG**X**GG**Y**. We analyzed the contribution of various base substitutions within the double chain reversal (**X**) on the kinetics and thermodynamics of these quadruplexes.

Architecture

We have demonstrated that the sequences GCGG**A**GGAT and GCGG**TC**GGT/A fold into a tetramolecular quadruplex that results in an interlocked bimolecular topology as established for **1** (16) and **3** (17). The quadruplexes formed result from anti-parallel strands at their four-stranded core, and parallel strands in two symmetry-related bistranded regions. The parallel strandedness observed for the bistranded regions of the stem quadruplex results from double chain reversals. Indeed, all guanines in the stem quadruplex structure for **1**, **2**, **3** and **4** show *anti* glycosidic bond angles. In contrast to T5 in **1**, for each single strand of **2**, the Hoogsteen edge of A5 recognizes the sheared edge of G3 thus forming a single-stranded triad, G3(A5–G6). The involvement of G3 and G6 in formation of a tetrad leads to an overall bistranded hexad alignment: A5(G3:G6:G3:G6)A5. Double chain reversals containing this particular structural motif are now well established in quadruplexes (30–33). We have also shown that this topology can accommodate more than one base in the double chain reversal. Indeed, substitution of a single base (X = T or A) for two bases (TT or TC) in the GCGG**X**GGA**Y** segment does also result in four-stranded quadruplex structures.

Sequence and temperature effects on association, dissociation and energetics

We report a complete analysis of the kinetic properties of a relatively large family of sequences (13 different oligonucleotides) coupled with a structural study for some of its members. The kinetics were compared with the parallel-stranded tetrameric quadruplex formed by TG_4T which was used as a reference. Calorimetry measurements allowed us to determine a model independent enthalpy of quadruplex formation (which was strongly negative) and evaluate the impact of each

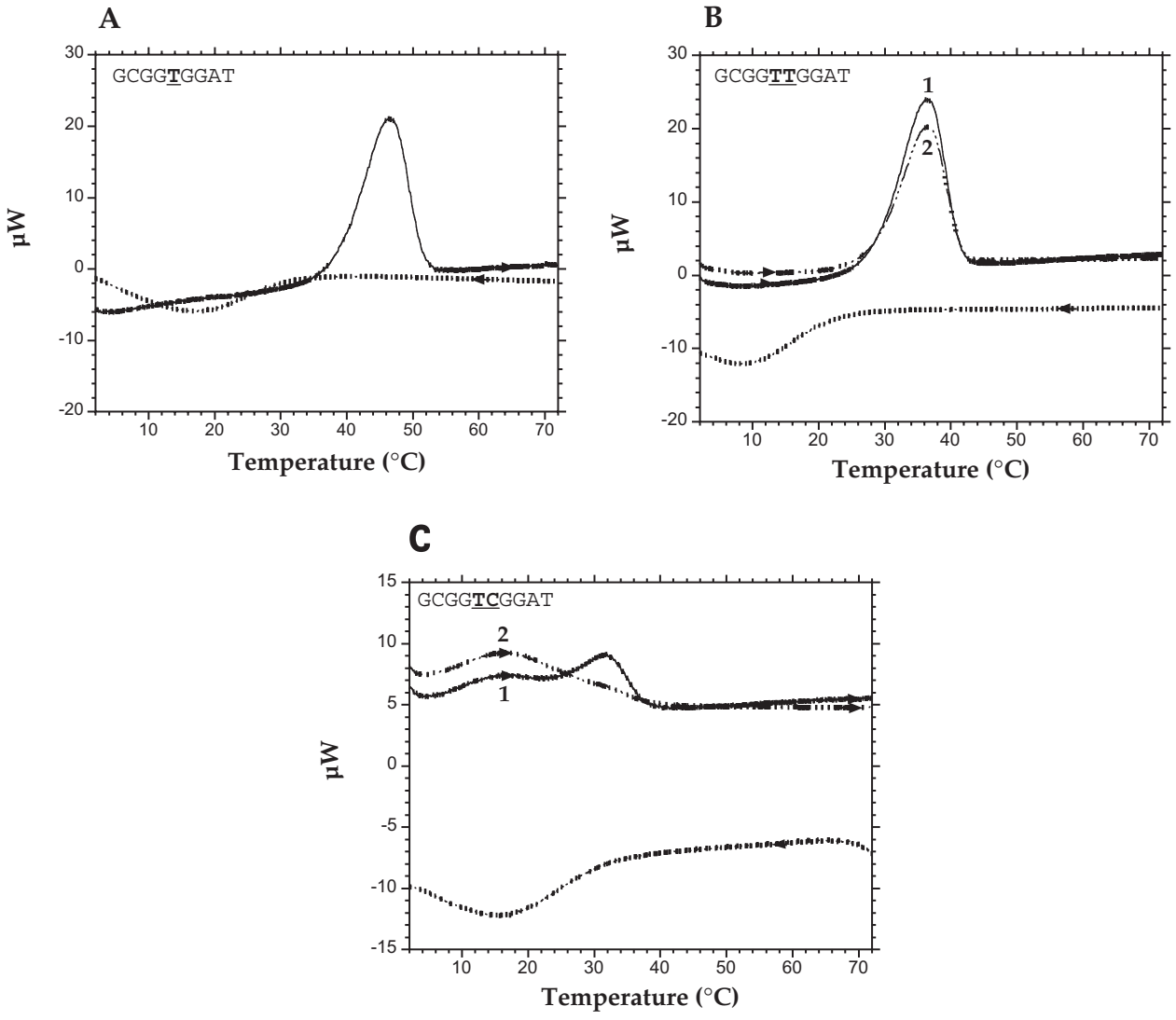


Figure 6. DSC analysis. DSC raw data of heating (solid line) and cooling (dotted line) (1°C/min) for (**A**) **1**, d(GCGG**T**GGAT) 357 µM, (**B**) **3**, d(GCGG**TT**GGAT) 382 µM, (**C**) **4**, d(GCGG**TC**GGAT) 317 µM. In (B and C), two heating profiles are presented (first heating, solid line; second heating, broken line). They illustrate partial, but not complete, reformation of the quadruplex after a first heating/cooling cycle [for (B)] and formation of a competitive structure (C).

modification on the energetics of the structure. Overall for sequences **1**, **3** and their variants, modifications did not strongly affect the enthalpy of formation. In contrast, a T→A substitution siginificantly destabilizes the structure (ΔH° drops to −60 kcal/mol; Table 1). It is important to note that the stabilities derived from the analysis of thermal denaturation curves alone do not take into account the very large differences in association rate constants (up to 10^5-fold) found between the different oligonucleotides. Although many oligomers adopt relatively similar conformations, the kinetics of these complexes may vary greatly.

The differences in the (non-equilibrium) melting temperatures ($T_{1/2}$) are a direct measure of the contribution of the stabilized structural elements within the structural environment of the double chain reversal. Thus, in general the **1** series is more thermally stable than **2**, and the latter more stable than the **3** series. The greater stability of **1** over **2** in a single residue loop may be explained from a combination of factors, such as backbone torsions of the double chain reversal and hydrophobic/hydrophilic interactions. The deviations from canonical backbone dihedral angle conformations were more numerous and severe in **2** than in **1** (17). Furthermore, optimum hydrophobic interactions of the methyl groups of T5 oriented towards the quadruplex stem (its solvent exposed Watson–Crick edge) contrast to hydrophilic Hoogsteen edge of A5 oriented towards the quadruplex stem coupled with a solvent exposed minor groove edge. Recognition of the Hoogsteen edge of a tetrad-forming G3 by A5 would contribute little to the stability of the local architecture, since hydrophobic interactions are usually of greater significance. Finally, the sequence **1B** has a lower apparent melting temperature ($T_{1/2}$) than **2**. In the latter there is contribution to stability by base stacking interactions of the A:A mismatch (hydrophobic interactions), which is absent in **1B**.

In extending to a two residue loop the thymines in **3** do not loop out. All thymines in this structure are well structured

and make significant hydrogen bond interactions with the stem. This is clearly supported by the observation of the exchangeable H3 proton for all thymines (pH 6.8, 0°C), as well as the observation of numerous NOE contacts of differing contribution to distance. Indeed, T5 and T6 are part of newly discovered planar architectures: a T5:A9:A9:T5 tetrad and a T6:(G3:G7:G3:G7):T6 hexad alignments. The observation of the T:A:A:T tetrad is documented by various NOE contacts of differing contribution to distance. For example, the adenine A9 shows cross-peaks A9H62–G8H1, T10H3–A9H62, T10H3–A9H61, T10H3–A9H2 clearly indicating stacking interactions with G8 and T10. In the first T5 does not form intrastrand base-stacking interactions. Thus it does not contribute significantly to the stability of the stem. Nevertheless, the appearance of T5H3 lower field shifted below 11 p.p.m. clearly indicates that it falls in a deshielding region of an isotropic aromatic field. This is consistent with the T5H3 proton positioning itself in a fairly narrow region. Compared with the backbone dihedral angles of **1** and **2**, in **3** there are no remarkable torsions. However, the methyl group of T6 is solvent exposed. This is likely to be the major contributor for its destabilization in comparison to **1** and **2**.

Corroborating structural conclusions with the kinetics of association is difficult, as differences in the association rate constants not only result from intrinsic propensity for folding of A, T, TT or TC base(s) within the GpG steps, but also from their propensity to favor competing architectures. It is relatively difficult to analyze these competing structures in detail, as they are relatively unstable (T_m below 20°C). This coupled equilibrium reduces to a variable extent the available free strand concentration for quadruplex formation, and therefore affects the apparent association rate constant of the quadruplex. Working around room temperature or above somewhat abrogates this contribution. A comparison of the k_{on} at 18–20°C indicates that the fastest-folding oligomers possess a central **T** base (as in **1**), whereas a **A**-loop slows the association and the two-base loops (**TC**, **TT**, **UC**) fall in between. The destabilizing effect of the purine (adenine) as compared with one pyrimidine (thymine) could actually result from at least two effects: (i) stabilization of a single-stranded or double-stranded structure that inhibits quadruplex formation, as for sequence **4** (Figure 6C) and (ii) a structural distorsion dominated by steric effects. This is further exacerbated in the folding of two-base loops. We did not observe a biphasic transition in the case of the sequence with a central adenine (**2**, data not shown) but DSC data suggests that this adenine *per se* has a strong detrimental effect on the enthalpy of the quadruplex (Table 1).

Biological relevance and implications for other sequences

Repetitive GG**X**GG segments show propensity to fold into quadruplexes bearing loops linking anti-parallel strands or double chain reversals resulting in parallel stranded quadruplexes (11,14–17,31,33–37). These sequence motifs are found in several important regions of the genome, especially in promoters (38), enhancers (39), locus control regions (40) and AGG islands present in fragile-X syndrome CGG repeats (41–43). Higher order structures with FMR1 sequence variants could be involved in the blockage of DNA processing in fragile X (44). Formation of a quadruplex may therefore play a role in the observed expansion in fragile X loci. Note, however, that all the results presented here were obtained in a sodium-containing buffer, in contrast with intracellular conditions, where potassium ions predominates. We are currently performing NMR experiments in a potassium-containing buffer. Preliminary results suggest that, at high DNA strand concentrations (~1 mM) **1**, **2** and **3** adopt a different topology we are currently investigating.

Our systematic approach gives insights into the propensity and stability of formation of double chain reversals for A, T, TT, and TC between two stacked tetrads. This study can be extended to three stacked tetrads if a suitable model system is developed and include a greater number and permutations of bases. The gained knowledge thus contributes to a better understanding of the propensity for transiently folding, dynamics and stability to be observed in double-chain reversals appearing in intramolecular quadruplexes folded in segments appearing in intronic regions, promoter regions, in, or near repressor and enhancer genetic sequences.

CONCLUSIONS

Moderately-rich guanine sequences (5 guanines for 8–10 base long oligomers) may form relatively stable tetramolecular quadruplexes, even if they involve 'blocks' of only two contiguous guanines. These quadruplexes are structurally diverse and involve a number of unusual pairing patterns such as hexads (Figure 1E–G), further extending the sequence repertoire of DNA quadruplexes. Some of these 'non-canonical' quadruplexes may actually form faster than the TG_4T reference sequence. We have demonstrated within the same topology that DNA sequence dramatically influences propensity for adopting double chain reversal loops. We have also probed modes of stabilization of bases within the double chain reversal loop, thus providing a structural basis for differential dissociation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank L. Lacroix and M. Rougée for helpful discussions. This work was supported by ARC (#3365) and EU FP6 'Mol Cancer Med' grants to J.L.M. S.A. is the recipient of a 'Fondation Jérôme Lejeune' fellowship. Funding to pay the Open Access publication charges for this article was provided by INSERM.

Conflict of interest statement. None declared.

REFERENCES

1. Williamson,J.R. (1994) G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 703–730.
2. Neidle,S. and Parkinson,G.N. (2003) The structure of telomeric DNA. *Curr. Opin. Struct. Biol.*, **13**, 275–283.

3. Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) Helix formation by guanylic acid. *Proc. Natl Acad. Sci. USA*, **48**, 2013–2018.
4. Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its applications for meiosis. *Nature*, **334**, 364–366.
5. Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) Monovalent cation induced structure of telomeric DNA: the G-quartet model. *Cell*, **59**, 871–880.
6. Davis,J.T. (2004) G-quartets 40 years later: from 5′-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed. Engl.*, **43**, 668–698.
7. Wyatt,J.R., Davis,P.W. and Freier,S.M. (1996) Kinetics of G-quartet-mediated tetramer formation. *Biochemistry*, **35**, 8002–8008.
8. Simonsson,T. (2001) G-quadruplex DNA structures—variations on a theme. *Biol. Chem.*, **382**, 621–628.
9. Keniry,M.A. (2001) Quadruplex structures in nucleic acids. *Biopolymers*, **56**, 123–146.
10. Phan,A.T., Kuryavyi,V., Ma,J.B., Faure,A., Andreola,M.L. and Patel,D.J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. *Proc. Natl Acad. Sci. USA*, **102**, 634–639.
11. Parkinson,G.N., Lee,M.P. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
12. Rangan,A., Fedoroff,O.Y. and Hurley,L.H. (2001) Induction of duplex to G-quadruplex transition in the c-myc promoter region by a small molecule. *J. Biol. Chem.*, **276**, 4640–4646.
13. Seenisamy,J., Rezler,E.M., Powell,T.J., Tye,D., Gokhale,V., Joshi,C.S., Siddiqui-Jain,A. and Hurley,L.H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.*, **126**, 8702–8709.
14. Phan,A.T., Modi,Y.S. and Patel,D.J. (2004) Propeller-type parallel-stranded g-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.*, **126**, 8710–8716.
15. Phan,A.T., Kuryavyi,V., Ma,J.B., Faure,A., Andreola,M.L. and Patel,D.J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: A potent inhibitor of HIV-1 integrase. *Proc. Natl Acad. Sci. USA*, **102**, 634–639.
16. Webba da Silva,M. (2003) Association of DNA quadruplexes through G:C:G:C tetrads. Solution structure of d(GCGGTGGAT). *Biochemistry*, **42**, 14356–14365.
17. Webba da Silva,M. (2005) Experimental demonstration of T(GGGG)T hexad and TAAT tetrad alignment within a DNA quadruplex system. *Biochemistry*, **44**, 3754–3764.
18. 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.
19. Mergny,J.L., de Cian,A., Ghelab,A., Saccà,B. and Lacroix,L. (2005) Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.*, **33**, 81–94.
20. 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.
21. Jeener,J., Meier,B.H., Bachmann,P. and Ernst,R.R. (1979) Investigation of exchange processes by 2-dimensional NMR-spectroscopy. *J. Chem. Phys.*, **71**, 4546–4553.
22. Piantini,U., Sorensen,O.W. and Ernst,R.R. (1982) Multiple quantum filters for elucidating NMR coupling networks. *J. Am. Chem. Soc.*, **104**, 6800–6801.
23. Braunschweiler,L. and Ernst,R.R. (1983) Coherence transfer by isotropic mixing—application to proton correlation spectroscopy. *J. Magn. Reson.*, **53**, 521–528.
24. Bodenhausen,G. and Ruben,D.J. (1980) Natural abundance N-15 NMR by Enhanced Heteronuclear Spectroscopy. *Chem. Phys. Lett.*, **69**, 185–188.
25. Varani,G., Abouela,F. and Allain,F.H.T. (1996) NMR investigation of RNA structure. *Prog. Nucl. Magn. Reson. Spectrosc*, **29**, 51–127.
26. Phan,A.T. (2000) Long-range imino proton-C-13 J-couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled DNA. *J. Biomol. NMR*, **16**, 175–178.
27. Mergny,J.L., Phan,A.T. and Lacroix,L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
28. 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.
29. Wallimann,P., Kennedy,R.J., Miller,J.S., Shalango,W. and Kenp,D.S. (2003) Dual wavelength parametric test of two-state models for circular dichroism spectra of helical polypeptides: anomalous dichroic properties of alanine-rich peptides. *J. Am. Chem. Soc.*, **125**, 1203–1220.
30. Kettani,A., Gorin,A., Majumdar,A., Hermann,T., Skripkin,E., Zhao,H., Jones,R. and Patel,D.J. (2000) A dimeric DNA interface stabilized by stacked A.(G.G.G.G).A hexads and coordinated monovalent cations. *J. Mol. Biol.*, **297**, 627–644.
31. Matsugami,A., Ouhashi,K., Kanagawa,M., Liu,H., Kanagawa,S., Uesugi,S. and Katahira,M. (2001) An intramolecular quadruplex of (GGA)(4) triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *J. Mol. Biol.*, **313**, 255–269.
32. Liu,H., Kugimiya,A., Sakurai,T., Katahira,M. and Uesugi,S. (2002) A comparison of the properties and the solution structure for RNA and DNA quadruplexes which contain two GGAGG sequences joined with a tetranucleotide linker. *Nucleosides Nucleotides Nucleic Acids*, **21**, 785–801.
33. Zhang,N., Gorin,A., Majumdar,A., Kettani,A., Chernichenko,N., Skripkin,E. and Patel,D.J. (2001) V-shaped scaffold: a new architectural motif identified in an A × (G × G × G × G) pentad-containing dimeric DNA quadruplex involving stacked G(anti) × G(anti) × G(anti) × G(syn) tetrads. *J. Mol. Biol.*, **311**, 1063–1079.
34. Wang,Y. and Patel,D.J. (1994) Solution structure of the Tetrahymena telomeric repeat d(T2G4)4 G-tetraplex. *Structure*, **2**, 1141–1156.
35. Matsugami,A., Okuizumi,T., Uesugi,S. and Katahira,M. (2003) Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G (:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. *J. Biol. Chem.*, **278**, 28147–28153.
36. Kuryavyi,V., Majumdar,A., Shallop,A., Chernichenko,N., Skripkin,E., Jones,R. and Patel,D.J. (2001) A double chain reversal loop and two diagonal loops define the architecture of a unimolecular DNA quadruplex containing a pair of stacked G(syn)-G(syn)-G(anti)-G(anti) tetrads flanked by a G-(T-T) Triad and a T-T-T triple. *J. Mol. Biol.*, **310**, 181–194.
37. Phan,A.T. and Patel,D.J. (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: Distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J. Am. Chem. Soc.*, **125**, 15021–15027.
38. Chen,J., Spector,M.S., Kunos,G. and Gao,B. (1997) Sp1-mediated transcriptional activation from the dominant promoter of the rat alpha-1B adrenergic receptor gene in DDT1MF-2 cells. *J. Biol. Chem.*, **272**, 23144–23150.
39. Masuda,E.S., Yamaguchi-Iwai,Y., Tsuboi,A., Hung,P., Arai,K.-I. and Arai,N. (1994) The transcription factor Sp1 is required for induction of the murine GM-CSF promoter in T cells. *Biochem. Biophys. Res. Commun.*, **205**, 1518–1525.
40. Pruzina,S., Antoniou,M., Hurst,J., Grosveld,F. and Philipsen,S. (1994) Transcriptional activation by hypersensitive site three of the human beta-globin locus control region in murine erythroleukemia cells. *Biochim. Biophys. Acta*, **1219**, 351–360.
41. Fu,Y.H., Kuhl,D.P.A., Pizzuti,A., Pieretti,M., Sutcliffe,J.S., Richards,S., Verkerk,A., Holden,J.J.A., Fenwick,R.G., Warren,S.T. *et al.* (1991) Variation of the CGG repeat at the fragile-X site results in genetic instability—resolution of the Sherman Paradox. *Cell*, **67**, 1047–1058.
42. Verkerk,A., Pieretti,M., Sutcliffe,J.S., Fu,Y.H., Kuhl,D.P.A., Pizzuti,A., Reiner,O., Richards,S., Victoria,M.F., Zhang,F.P. *et al.* (1991) Identification of a gene (Fmr-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in Fragile-X Syndrome. *Cell*, **65**, 905–914.
43. Eichler,E.E., Macpherson,J.N., Murray,A., Jacobs,P.A., Chakravarti,A. and Nelson,D.L. (1996) Haplotype and interspersion analysis of the FMR1 CGG repeat identifies two different mutational pathways for the origin of the fragile X syndrome. *Hum. Mol. Genet.*, **5**, 319–330.
44. Hansen,R.S., Canfield,T.K., Lamb,M.M., Gartler,S.M. and Laird,C.D. (1993) Association of Fragile-X Syndrome with delayed replication of the Fmr1 gene. *Cell*, **73**, 1403–1409.