Evaluation of the effect of polymorphism on G-quadruplex-ligand interaction by means of spectroscopic and chromatographic techniques

S. Benito[1], A. Ferrer[1], S. Benabou[1], A. Aviñó[2], R. Eritja[2], R. Gargallo[1]*

[1] Department of Chemical Engineering and Analytical Chemistry. University of Barcelona. Martí i Franquès, 1-11, E-08028 Barcelona, Spain

[2] Institute for Advanced Chemistry of Catalonia (IQAC-CSIC), CIBER-BBN, Jordi Girona 18-26, E-08034 Barcelona, Spain.

1. Abstract

Guanine-rich sequences may fold into highly ordered structures known as G-quadruplexes. Apart from the monomeric G-quadruplex, these sequences may form multimeric structures that are not usually considered when studying interaction with ligands. This work studies the interaction of a ligand, crystal violet, with three guanine-rich DNA sequences with the capacity to form multimeric structures. These sequences correspond to short stretches found near the promoter regions of *c-kit* and SMARCA4 genes. Instrumental techniques (circular dichroism, molecular fluorescence, size-exclusion chromatography and electrospray ionization mass spectrometry) and multivariate data analysis were used for this purpose. The polymorphism of G-quadruplexes was characterized prior to the interaction studies. The ligand was shown to interact preferentially with the monomeric G-quadruplex; the binding stoichiometry was 1:1 and the binding constant was in the order of 10^5 M^{-1} for all three sequences. The results highlight the importance of DNA treatment prior to interaction studies.

Keywords: G-quadruplex, crystal violet, binding equilibria, polymorphism, multivariate analysis.

2. Introduction

Guanine-rich sequences of nucleic acids may fold into complex structures known as G-quadruplexes. The building blocks of such structures are known as G-tetrads, which are formed by the planar arrangement of four guanine bases stabilized by hydrogen bonding (Figure 1a). The stability of the G-quadruplex depends heavily on the nucleic acid sequence (composition and length of both, guanine tracts and loops) and on external factors such as the nature of the counterion and the concentration, temperature and presence of appropriate ligands [1]. The biological importance of this structure lies in its presence in the promoter regions of several oncogenes and at the extreme ends of telomeres, which are related to cancer and aging [2]. The *in vivo* existence of these structures was demonstrated fairly recently [3]. In addition, the fact that their folding is strongly influenced by the presence of certain counterions, such as potassium, makes it a suitable tool for the development of analytical methods [4, 5] and for nanotechnology applications [6, 7].

In general, a 20-30-nucleotide long DNA sequence may fold intramolecularly into a monomeric G-quadruplex structure. This may show parallel, anti-parallel or 3+1 hybrid topologies (Figure 1b) [1]. In addition, G-quadruplexes have been shown to form multimeric forms such as dimers and tetramers, depending on the experimental conditions [8-11]. In the presence of 150 mM potassium, sodium or ammonium cations, parallel topologies are favored in G-quadruplex structures that contain short loops. Moreover, these parallel G-quadruplexes may form stable multimeric structures such as dimers or even trimers, even at low strand concentrations. The correlation between multimer formation and parallel G-quadruplex formation strongly suggests that multimeric assemblies are parallel. On the other hand, the G-quadruplex structures tend to be more intramolecular and anti-parallel as loop length increases [12].

The *in vivo* presence of G-quadruplexes is related to their role in certain biological processes such as DNA replication. In this regard, interaction with ligands is a subject that requires extensive research to modulate the stability, and consequently, the functionality of these structures [13-15]. However, to the best of our knowledge, the potential formation of multimeric species other than the monomer is rarely considered in studies on G-quadruplexes and interaction with ligands. Thus, the main objective of this work was to study the interaction of a ligand with several guanine-rich sequences prone to form multimeric structures in order to ascertain whether the presence of multimeric structures should be considered in such research.

Crystal violet (CV), the ligand used in this work, is a fluorescent triphenylmethane dye (Figure 1c) that has been shown to bind selectively to the G-quadruplex motif over duplex or single-stranded DNA [16]. Research has been carried out on the interaction of CV with the anti-parallel G-quadruplex formed by repeated subunits of the *Oxytricha* telomere [17]. Based on these studies, a binding mode was proposed in which the dye stacks onto the two external tetrads of the G-quadruplex structure. Many sensors have been developed based on the interaction of CV with DNAs in order to detect a plethora of analytes, such as potassium ions [18, 19], lead ions [20], copper ions [21] and ATP [22], and the activity of proteins, such as 3'→ 5' exonuclease [23].

The guanine-rich DNA sequence models used in this work are shown in Figure 1d. A guanine-rich sequence near the promoter region of the *c-kit* gene has previously been studied (ckit21) [24-27]. This sequence shows three tracts of three guanines and one tract of four guanines near the 3' end. Because of this asymmetric distribution of guanines, there is an equilibrium between two conformers that hinders the structural study of this sequence. To overcome this problem, appropriate G-T mutations in the fourth tract were proposed (ckit21T21 and ckit21T18), and the resulting G-quadruplex structures were shown to be more homogeneous than those formed from the ckit21 sequence. Whereas ckit21T21 formed a parallel topology, ckit21T18 formed mainly parallel G-quadruplex with some antiparallel G-contribution [24]. Later studies on ckit21T21 using NMR spectroscopy led to the proposal that this sequence adopts either a dimeric or a monomeric G-quadruplex structure, at high or low potassium concentrations, respectively [25, 26]. SMG03 corresponds to a guanine-rich sequence located between bases -71 and -28 upstream of the promoter region of the SMARCA4 gene, and has not yet been studied. Lastly, T21 was selected as a linear (i.e., non-compact) structure model.

Instrumental techniques such as circular dichroism (CD), molecular absorption and molecular fluorescence spectroscopies, size-exclusion chromatography (SEC) and mass spectrometry, as well as multivariate analysis methods, were used in this work.

3. Material and methods

3.1. Reagents

The DNA sequences were synthesized on an Applied Biosystems 3400 DNA synthesizer using the 200 nmol scale synthesis cycle. Standard phosphoramidites were used. Ammonia deprotection was performed overnight at 55°C. The resulting products were purified using a Glen-Pak Purification Cartridge (Glen Research, Sterling, VA, USA). DNA strand

concentration was determined by absorbance measurements (260 nm) at 90°C using the extinction coefficients calculated using the nearest-neighbor method, as implemented on the OligoCalc website [28]. KCl, KH_2PO_4, K_2HPO_4, $NaCH_3COO$, HCl and NaOH (a.r.) were purchased from Panreac (Barcelona, Spain). MilliQ® water was used in all experiments.

3.2. Instruments and apparatus

Absorbance spectra were recorded on an Agilent 8453 diode array spectrophotometer (Agilent Technologies; Waldbronn, Germany). Temperature was controlled by means of an Agilent 89090A Peltier device (Agilent Technologies). CD spectra were recorded on a Jasco J-810 spectropolarimeter equipped with a JULABO F-25-HD temperature control unit (Seelbach, Alemania). Molecular fluorescence spectra were measured with an AMINCO-Bowman AB2 spectrofluorometer. The temperature was controlled by means of a water bath. Measurements were taken at 570 nm excitation wavelength, with 900 V sensitivity, and 4 nm excitation and emission slides. In all spectroscopic studies, Hellma quartz cells (10 mm path length, and 350, 1500 or 3000 µl volume) were used.

For SEC, the chromatographic system consisted of an Agilent 1100 Series HPLC instrument equipped with a G1311A quaternary pump, a G1379A degasser, a G1392A autosampler, a G1315B photodiode-array detector with a 13-µL flow cell, and an Agilent ChemStation for data acquisition and analysis (Rev. A 10.02), all from Agilent Technologies. Two chromatographic columns were used for separation at room temperature: a BioSep-SEC-S3000 column (300 × 7.8 mm, particle size 5 µm and pore size 290 Å) from Phenomenex (Torrance, CA, USA) and an Acclaim SEC-300 column (300 × 7.8 mm, particle size 5 µm and pore size 300 Å) from Thermo Scientific (Waltham, MA, USA).

Electrospray ionization mass spectrometry (ESI-MS) spectra were recorded with an Agilent G1969A LC/MSD TOF Mass Spectrometer (Agilent Technologies, Santa Clara, CA, USA). The capillary voltage was set to 3.5 kV, the voltage range to 200 or 300 V, the gas temperature to 325°C, the pressure of nebulizer N_2 gas to 15 psi, and the N_2 flow rate for drying was set to 7.0 l·min^{-1}.

3.3. Procedures

The DNA solutions for measurements were prepared by dilution of DNA stock solutions in the appropriate medium. In accordance with the procedure usually applied when studying the conformational equilibria of DNA, all samples for measurement were subjected to a preliminary treatment that consisted of heating the samples at 90°C for 10 minutes

and slowly cooling them overnight to room temperature (from now on, referred to as the thermal treatment). As described in the Results section, another treatment based on the addition of LiOH 1 M to pH 11.0 and the subsequent addition of HCl 1 M to pH 7 was also carried out (from now on, referred to as the acid-base treatment).

The melting experiments were monitored using the Agilent 8453 spectrophotometer or the J-810 spectropolarimeter equipped with a Peltier unit. The DNA solution was transferred to a covered cell and spectra were recorded at 1°C intervals with a hold time of 3 min at each temperature, which yielded an average heating rate of approximately 0.3°C·min^{-1}. The medium consisted of 20 mM phosphate buffer (pH 7.0) and 50 mM KCl. Each sample was allowed to equilibrate at the initial temperature for 30 minutes before the melting experiment began.

For binding studies monitored by CD spectroscopy, small aliquots of stock CV solutions were added to a diluted DNA solution. In the case of molecular fluorescence-monitored titrations, the procedure consisted of the addition of small aliquots of DNA stock solution to CV diluted solutions. In all cases, the experimental conditions were 25°C, 20 mM phosphate buffer and 50 mM KCl. Spectra were recorded 5 minutes after each addition of titrant.

For SEC analysis, the mobile phase was 300 mM KCl and 20 mM phosphate buffer (pH 7.0). The flow rate was set to 1.0 mL min^{-1}. The injection volume was 15 μL. Absorbance spectra were recorded between 200 and 500 nm. T_{15}, T_{20}, T_{25}, T_{20} and T_{45} sequences were used as standards to construct the plot of logarithm of the molecular weight *vs.* retention time (t_R). Standards were injected twice to assess the reproducibility of the t_R values, and the relative difference between t_R values for a given standard was lower than 0.5%. SEC profiles were normalized to equal length (Euclidean normalization) to eliminate potential variations in the DNA concentration of samples that could hinder the comparison of chromatograms. Normalization was carried out using equation 2 [29]:
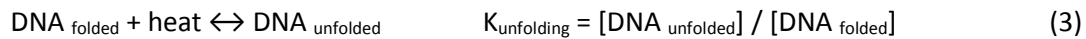
$$Normalized\ chromatogram = \frac{raw\ chromatogram}{\sqrt{\sum_1^n d_i^2}} \qquad (2)$$

For ESI-MS measurements, the sample (10 μl volume, 150 mM NH$_4$AcO, 20% CH$_3$OH) was introduced into the source by means of an Agilent 1100 HPLC pump. The flow rate was set to 200 μl·min^{-1} (H$_2$O:CH$_3$CN, 1:1).

3.4. Data analysis

3.4.1. Univariate analysis of melting data

For melting experiments, the absorbance or ellipticity data measured at one single wavelength (univariate data) as a function of temperature were analyzed, as described elsewhere [30]. For the unfolding of intramolecular structures such as those studied here, the chemical equation and corresponding equilibrium constant may be written as:

$$\text{DNA}_{\text{folded}} + \text{heat} \leftrightarrow \text{DNA}_{\text{unfolded}} \qquad K_{\text{unfolding}} = [\text{DNA}_{\text{unfolded}}] / [\text{DNA}_{\text{folded}}] \qquad (3)$$

The concentration of the folded and unfolded forms is temperature dependent. Accordingly, the equilibrium constant depends on temperature, in line with the Van't Hoff equation:

$$\ln K_{\text{unfolding}} = -\Delta H_{vH} / RT + \Delta S_{vH} / R \qquad (4)$$

It is assumed that ΔH_{vH} and ΔS_{vH} do not change throughout the temperature range studied here.

3.4.2. Multivariate analysis

While univariate analysis relies on process monitoring (a melting or mole-ratio experiment) using just one wavelength, multivariate analysis uses a large set of wavelengths. Hence, CD or absorbance spectra recorded stepwise along a process were arranged in a table or data matrix D, with m rows (spectra recorded) and n columns (wavelengths measured). The goals of the multivariate analysis were as follows: (i) determination of the number of nc spectroscopically active species present along the process, (ii) calculation of the distribution profiles for each one of these nc species, and (iii) calculation of their pure (or individual) spectra. The distribution profiles provide information about the thermodynamics of the melting processes, in addition to the stoichiometry and stability of the species considered (in the case of mole-ratio experiments). Moreover, the shape and intensity of the pure spectra may provide qualitative information about the structure of the species. Using multivariate analysis instead of univariate analysis would facilitate the detection of potential intermediates along the process.

With this goal in mind, data matrix D was decomposed according to the Beer-Lambert law in matrix form:

$$D = C S^T + E \qquad (5)$$

Where C is the matrix (m x nc) containing the distribution profiles, S^T is the matrix (nc x n) containing the pure spectra, and E is the matrix of residual data (m x n) not explained by the proposed decomposition.

The mathematical decomposition of **D** into matrices **C**, **S**^{T} and **E** may be conducted in two different ways, depending on whether a physico-chemical model is initially proposed (the so-called hard-modeling approach) or not (the soft-modeling approach) [31]. In this work, only the hard-modeling approach was used. Logically, the proposed physico-chemical model depended on the nature of the process under study. Hence, for the melting experiments, the model included a set of chemical equations, as previously used in univariate analysis [32]. For the binding titrations, the model included a set of chemical equations to describe the formation of the different DNA-ligand species, together with approximate values for the stability constants, such as the following:

$$DNA + n\ ligand \leftrightarrow DNA\cdot ligand_p \qquad\qquad Beta_{1n} = [DNA\cdot ligand_n] / [DNA]\ [ligand]^n \qquad\qquad (6)$$
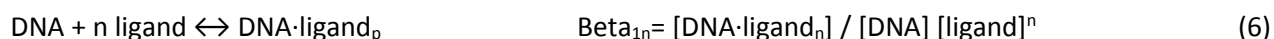
Whichever physico-chemical model is applied, the distribution profiles in **C** fulfilled the proposed model. Along the calculation, the values for the equilibrium constants in (3), (4) or (6) and the shape of the pure spectra in **S**^{T} were refined iteratively to explain the data in **D**, according to equation (5), whereas residuals in **E** were minimized. In this work, the hard modeling of the multivariate data recorded from the mole-ratio experiments was carried out using the EQUISPEC program [33].

4. Results

In this study, three G-rich sequences representing three distinct intramolecular G-quadruplexes (parallel, anti-parallel and 3+1 hybrid) were selected (Figure 1) to analyze the interaction of CV with G-quadruplexes.

4.1. Characterization of G-quadruplex structures

Before starting to study the interaction of CV with the G-quadruplex structures formed by the chosen sequences, these were characterized in terms of topology, thermal stability and molecularity.

4.1.1. CD spectra of model G-rich sequences

Figure 2 shows the CD spectra of the chosen sequences at 25°C. Prior to measurement, DNA samples were treated in accordance with the procedure described in many studies that deal with G-quadruplex structures: an aliquot of a DNA stock solution is dissolved in the appropriate medium (20 mM phosphate buffer, pH 7.0 and 50 mM KCl) and later heated at 90°C for 10 minutes, then cooled slowly overnight. The CD spectrum of T21, which is expected to adopt a linear, partially stacked structure, is clearly different from the other spectra. The CD spectrum of ckit21T21 shows

negative and positive signals around 243 and 263 nm, respectively, which are typical of a parallel G-quadruplex topology [34]. In addition to the signals at 243 and 263 nm, the CD spectrum of the ckit21T18 sequence shows a shoulder around 295 nm, which is indicative of anti-parallel topology. Hence, the overall CD spectrum of ckit21T18 may correspond to a 3+1 hybrid G-quadruplex or to a mixture of two topologies (i.e., a major population of parallel G-quadruplex and a minor population of anti-parallel G-quadruplex) (Figure 1b). These differences between the two ckit21-derived sequences is consistent with the general observation that, when loop length increases, the preferred G-quadruplex topology tends to be anti-parallel [12]. Finally, the shape of the CD spectrum of SMG03 suggests an anti-parallel topology, with a clear positive signal around 295 nm [34].

4.1.2. Thermal stability

CD-monitored melting experiments were used to quantify the thermal stability of the folded structure formed by the chosen sequences. After the thermal treatment described above, the melting behavior of ckit21T21 presented a clear two-state transition (Figure 3a). Data analysis provided thermodynamic variable values and a T_m equal to $61 \pm 1°C$ (Table 1 and Figure S1). The shape and intensity of the CD spectrum at 90°C suggested that not all DNA sequences unfolded at this temperature, which may be indicative of the presence of multimeric structures other than monomeric G-quadruplex structures. These multimers would be more stable than the monomer and therefore would not unfold under these experimental conditions. This fact was further demonstrated as follows: after melting, the sample was treated with LiOH 1 M until a pH equal to 12.0 was measured (Figure 3b). Under these conditions, guanine bases lost the hydrogen at N1 (the pK_a of which is around 10), thereby causing disruption of the G-tetrad due to the breaking of hydrogen bonds between guanine bases (Figure 1a). The stepwise addition of an equivalent volume of HCl 1 M to return to a pH value equal to 7.0 led to the complete recovery of the CD spectrum prior to the addition of LiOH (Figure 3c). After this acid-base treatment, the formation of multimeric structures was expected to a minor extent due to the low temperature, high concentration and long reaction time. A second melting process was then carried out to confirm elimination of the multimeric structures (Figure 3d). As observed, the CD signal recorded at 90°C was much lower than that measured at the end of the first melting experiment, thus indicating almost complete unfolding. The thermodynamic parameters calculated for the second melting procedure were only slightly lower than those calculated for the first melting procedure (see Figure S1), which would suggest that both unfolding processes involved the same initial folded structure (i.e., the monomer).

The melting of a ckit21T18 sample, prepared with the thermal treatment, showed a slightly more complex unfolding process than that of ckit21T21 (Figure S2). Hence, the shape of the measured ellipticity *vs.* temperature trace at 263 nm suggests a three-state transition (Figure S3). As a result, it was not possible to analyze all data using a simple two-state process, and all of the spectra were analyzed using a multivariate analysis method [32]. The first transition, which was characterized by loss of the signal at 295 nm and the concomitant enhancement of the signal at 264 nm was related to the loss of the anti-parallel contribution. The second transition, which was characterized by a reduction in the signal at 264 nm was related to the unfolding of the parallel G-quadruplex structure. As with ckit21T21, the shape and intensity of the CD spectrum at 90°C may be indicative of the presence of stable multimeric structures. The addition of LiOH up to pH 12.0 led to the loss of the G-quadruplex folding. In contrast to ckit21T21, the addition of HCl down to pH 7.0 did not lead to the full recovery of the initial structure, which indicated that the kinetics of the folding/unfolding reaction was slow. In these conditions, the calculation of reliable thermodynamic parameters and T_m values was not possible and, because of that, only $T_{1/2}$ values, which gives rough information about the stability of the folded structure, were determined (Table 1). The kinetically favored structure formed by this acid-base treatment showed a higher anti-parallel structure contribution than that obtained through the thermal treatment. The next melting procedure led to the loss of the ordered structure in a two-state transition characterized by a $T_{1/2}$ value of around 75°C (Figure S3). As with ckit21T21, the shape and intensity of the CD spectrum at 90°C reflected the loss of the multimeric structures.

Finally, SMG03 sequence showed behavior that could be considered an intermediate between those of the ckit21T21 and ckit21T18 sequences (Figure S4). Hence, the melting of a SMG03 sample prepared using the thermal treatment showed a single transition with a low T_m value of around 30°C (Table 1, Figure S5). This was consistent with the generally lower stability of anti-parallel G-quadruplexes with respect to parallel G-quadruplexes, and the short guanine tracts present in this sequence (Figure 1c). On the other hand, the intensity and shape of the CD spectrum at 90°C suggested that most of the initial anti-parallel G-quadruplex structure was destroyed by heat. In other words, the presence of multimeric structures was low compared to the intramolecular structures. After the acid–base treatment, the initial anti-parallel topology was fully recovered after a few minutes. The reversibility of the process was confirmed by the similar melting profiles (Figure S5).

4.1.3. SEC and ESI-MS analysis: a more detailed picture

Size-exclusion chromatography (SEC) has shown to be a reliable tool for characterizing G-quadruplex polymorphism. In this work, this technique was used to characterize the potential polymorphism adopted by the chosen sequences. Figure 4 shows the chromatograms recorded for the sequences and for one of the standards (T21). The experimental concentration of salt used in the SEC analysis (300 mM KCl) was higher than that used in the CD measurements (50 mM) in order to prevent the ionic exclusion of the DNA from the pores in the silica-based stationary phase. Lower salt concentrations were found to cause the elution of all DNA samples at very low retention times (data not shown). This fact was related to the disappearance of the chromatographic process observed when all DNA sequences were excluded from the pores of the stationary phase. This exclusion was due to the electrostatic repulsion caused by the fact that both DNA and silica are negatively charged [35].

As with the CD measurements, the samples for SEC analysis were prepared using the thermal treatment. Thus, the chromatogram of ckit21T21 showed four main bands (Figure 4b). According to the calibration plot (Tables S6 and S7), these bands were assigned to a tetramer (8.0 min), a dimer (8.9 min), the unfolded sequence (9.6 min) and the monomeric G-quadruplex (10.0 min). The monomeric G-quadruplex structure eluted with a retention time greater than that of the unfolded sequence due to its compactness and, consequently, its smaller hydrodynamic volume. The formation of dimeric forms in the ckit21T21 sequence had already been suggested based on the NMR data [25]. As expected, the chromatogram of a second ckit21T21 sample, prepared through acid-base denaturation, showed a reduction in the peaks assigned to the multimeric structures and a slight enhancement of the peak assigned to the monomeric forms, which was consistent with the hypothesis previously proposed based on the CD melting data (Figure S8).

The chromatogram of the ckit21T18 sequence showed five bands. As in the previous case, the bands at 8.7, 9.5 and 9.8 minutes were assigned to the tetrameric, unfolded strand and a monomeric G-quadruplex structure, respectively. By contrast, the assignation of bands at 7.8 and 11.3 minutes was not straightforward. The band at 7.8 minutes corresponded to a tetramer or to another structure formed by a greater number of individual strands, such as the G-wire [36, 37]. Finally, the band at 11.3 minutes, which corresponds to a structure with a smaller hydrodynamic volume than the unfolded strand, was assigned to a highly compact monomeric structure. Again, the acid-base treatment caused the unfolding of multimeric structures, as revealed by the slight reduction in the peaks assigned to these forms (Figure S8).

The chromatogram of the SMG03 sequence was much simpler than those of the ckit21-derived sequences, since it showed just two bands. In this case, it seems that no linear structure was present, as none of the peaks eluted at around 9.5 minutes. The major band at 9.8 minutes was assigned to a folded structure, probably a monomeric G-quadruplex, while the minor band at 9.0 minutes was assigned to a multimeric structure, probably a dimer.

The presence of species other than monomers was confirmed by ESI-MS, a technique that has also shown to be a useful tool for this purpose [12, 38]. ESI-MS spectra of ckit21T21 and ckit21T18 and SMG03 in 150 mM NH_4AcO are shown in Figure S9. In this case, only the thermal treatment was used. Even though the experimental conditions at which DNA samples are monitored by ESI-MS are different from those used in solution experiments, these results pointed to the existence of dimers, in addition to monomers, in the case of ckit21T21 and ckit21T18. Accordingly, signals that would correspond to multimeric structures are missing from the spectra of SMG03.

4.2. Interaction of DNAs with CV

4.2.1. Mole-ratio experiments

Following characterization of the guanine-rich sequences, the interaction with the crystal violet (CV) ligand was studied. First, a set of CD-monitored mole-ratio experiments was carried out to obtain information about the effect of binding on the G-quadruplex topology and to determine the stoichiometry and binding constants (Figure 5). In these experiments, DNA was subjected to the thermal treatment. In general, adding the ligand caused induced CD signals to appear at 545 and 313 nm. As expected, the control sequence (T21) did not cause this induced CD signal to appear in the visible region, and only a minor variation was observed at 310 nm at high CV-DNA ratios. The CD signature of the ckit21T21 parallel topology was maintained after addition of the ligand (Figure 5a). On the contrary, the anti-parallel contribution observed in the CD spectrum of ckit21T18 was partially lost after the ligand was added (Figure 5b).

The results obtained from CD-monitored titrations were corroborated by means of binding studies analyzed using molecular fluorescence spectroscopy. For all sequences studied, the low fluorescence of the ligand was clearly enhanced upon interaction with the DNA (Figure 6). This finding was related to the stacking of the dye on the face of a G-tetrad, which may increase the rigidity of the dye structure and, hence, its intrinsic fluorescence [17]. The fluorescence enhancement was more pronounced in the case of the ckit21T18 sequence than in the case of ckit21T21. This finding is consistent with previous observations about the preference of CV for anti-parallel structures over parallel structures [16]. In the case of SMG03, the fluorescence enhancement was intermediate between the two

ckit21 variants. Data analysis made it possible to propose a 1:1 binding stoichiometry in all cases, with the logarithms of the binding constants equal to 4.5±0.3, 5.2±0.3 and 5.3±0.2 for ckit21T21, ckit21T18 and SMG03, respectively (Figure S10). The relatively low value of the binding constant, together with the minor spectral changes observed upon binding, pointed to a weak binding mode [39], probably related to the stacking of the dye on an external G-tetrad.

The effect of polymorphism on the titration of CV by ckitG1T21 was also studied with a DNA titrant solution previously treated with LiOH and HCl. In this case, the logarithm of the binding constants was determined as equal to 4.9±0.2, which was slightly higher than the value determined using the thermal treated DNA (4.5±0.3). This was attributed to the increase in monomeric forms after the acid-base treatment, which resulted in a higher proportion of DNA in relation to CV and, hence, an increase in the concentration of the DNA-CV interaction complex.

4.2.2. Melting experiments

CD-monitored melting experiments of DNA-CV (1:5) mixtures were conducted in order to gain an insight into the influence of CV on the thermal stability of the G-quadruplex. Figure 7 shows the spectral variations observed in all three cases. The increased stability of the G-quadruplex structure was monitored at 263 or 292 nm, and the fractions of folded DNA *vs.* T are plotted in Figure S12. In general, stabilization of the G-quadruplex structure upon CV binding was observed. In the case of ckit21T21, CV produced an increase in the T_m value of the DNA (65±2°C) with respect to the value measured in the absence of DNA (61±1C). It should be stressed, however, that the values of ΔH^0 and ΔH^0 were clearly higher in the presence of CV than in its absence, thus making the folding process more cooperative (Table 1 and Figure S12). Finally, the value of ΔG_{37oC} was more negative in the presence of CV, thereby making the folded DNA structure more stable at this temperature than in the absence of the ligand. The shape and intensity of the spectrum at 90°C reflect the persistence of G-quadruplex aggregates. It is important to highlight the differences between the changes in ellipticity at 545 (inset in Figure 7a) and 263 nm (Figure S11). While a sharp transition was observed at 263 nm, a rather smooth variation was seen at 545 nm. The sharp variation at 263 nm was obviously related to the unfolding of the monomer G-quadruplex structure. Therefore, the smooth variation at 545 nm must have been related to minor variations in the loop topology that slightly affected CV binding without disrupting the overall structure.

The effect of polymorphism on the interaction between ckitG1T21 and CV was also studied by means of a CD-monitored melting of a mixture (1:5), where the DNA was previously treated with LiOH and HCl. In this case, the T_m

value determined in the presence of the ligand was 63±1°C, slightly lower than in the case of the thermally treated DNA. This observation was attributed to the increase in monomeric forms after the basic treatment, which resulted in a higher proportion of DNA in relation to CV and, hence, a reduction in the T_m value observed for the DNA-CV mixture.

In the case of ckit21T18, the shape of the CD spectra at low temperatures and the trace recorded at 263 nm pointed to slight stabilization of the parallel topology induced by CV, as implied by the mole-ratio experiment. The presence of the ligand led to a slight increase in the $T_{1/2}$ value ($\Delta T_{1/2} \sim 5$°C). Finally, in the case of SMG03, the presence of CV induced clear stabilization of the G-quadruplex with a $\Delta T_M \sim 10$°C. In this case, the variation in the ellipticity measured at 292 and 545 nm was similar, thus suggesting a binding mechanism other than that observed for ckit21T18.

4.2.3. SEC and ESI-MS analysis

Figure 8 shows the chromatograms obtained for the mixtures of CV and the chosen sequences. In the case of ckit21T21, increasing concentrations of the ligand barely produced any changes in the band at 8.0 minutes, which was assigned to the tetrameric structure. A clear reduction in the band at 9.6 minutes, which was previously assigned to the unfolded strand, was observed. Concomitant increases were observed in the band at 8.9 and 10.0 minutes, which were assigned to the dimeric and monomeric G-quadruplex structures, respectively. These changes pointed to a shift in equilibrium from the unfolded strand to monomeric and dimeric G-quadruplexes upon interaction with the ligand. This result, which is consistent with the overall stabilization observed in the melting and mole-ratio experiments, provides a detailed picture of the interaction and points to a preferential interaction with dimeric and monomeric forms. The addition of the ligand, whose molecular weight is 372.5 g·mol^{-1}, to the DNA strand (6634.3 g·mol^{-1}) produced such small variations in the t_R values that they did not affect the above conclusions.

The addition of CV to the ckit21T18 sequence produced changes in the SEC chromatogram that differed slightly to those described above for ckit21T21. As in that case, the band at 7.8 minutes, which was related to the multimeric species, barely changed upon addition of the ligand. Moreover, the bands at 8.7 and 9,5 minutes, which were related to the dimeric G-quadruplex and unfolded strand, respectively, behaved in a similar way to that described for ckit21T21; while the former showed an increase, the latter showed a reduction upon addition of the ligand. A different behavior was observed in the case of the bands at 9.8 and 11.3 minutes, which were related to monomeric G-quadruplex structures; while the former decreased in intensity, the latter increased concomitantly. This behavior could be related to a shift in equilibrium caused by addition of the ligand.

Finally, the addition of CV to SMG03 also produced slight changes in the chromatogram that were similar to those observed for ckit21T18. While the intensity of the band at 9.0 minutes, which was assigned to a folded species, increased slightly, that of the band at 9.8 minutes, which was assigned to a dimeric form, reduced. Moreover, a new band appeared at 10.5 minutes.

The proposed 1:1 stoichiometry and preferential interaction with monomeric forms were also tested by means of ESI-MS (Figure S12). The addition of CV to DNA (1:4 DNA-ligand ratio) resulted in the disappearance of signals related to dimeric structures and the appearance of signals related to the complex. The proposed stoichiometry was 1:1, in accordance with the results obtained from fluorescence-monitored titrations.

5. Discussion

The main objective of this work was to study the interaction of a ligand with several guanine-rich sequences prone to forming multimeric structures in order to determine whether the presence of multimeric structures should be considered in such binding studies. To our knowledge, this aspect of DNA solution chemistry is rarely considered in the determination of stoichiometries and the calculation of stability constants in DNA-ligand interactions.

Using several instrumental techniques and methods, the selected sequences were shown to form G-quadruplex structures that differed in both topology (parallel, anti-parallel or hybrid) and polymorphism (monomer, dimer, etc.). In this regard, the present work complements prior studies [25], in which the formation of dimeric G-quadruplex structures by T12 and T21 mutations of the *ckit21* wild sequence (5'-C GGG C GGG CGC<u>G</u>A GGG A GGG<u>G</u>-3') has been postulated based on NMR data. It should be stressed that the interpretation of NMR spectra has been a matter of discussion among researchers who defend the formation of either monomeric [26] or dimeric [25] G-quadruplex structures by that sequence. In this regard, the present work highlights the importance of using instrumental techniques to provide complementary information to that obtained through NMR spectroscopy, as shown recently, in the elucidation of the two coexisting monomeric G-quadruplexes, parallel and hybrid, in the promoter region of the EGFT oncogene [40].

This work has demonstrated the preferential formation of multimeric structures by the ckit21T21 sequence in relation to the ckit21T18 or SMG03 sequences. This fact is consistent with the previously observed preferential multimer formation by parallel G-quadruplex [12], and suggests a mechanism for multimer formation based on the stacking of

monomeric parallel G-quadruplexes. Anti-parallel G-quadruplexes, like that formed by the SMG03 sequence, show lateral and/or diagonal loops that may hinder monomer stacking and thus prevent the formation of such multimeric structures. The ckit21T18 sequence, which has been shown to form a 3+1 hybrid structure, displays an intermediate behavior that lies somewhere between these two extreme scenarios. Previous studies proposed that guanine-rich sequences showing a 5'-GC end may form multimeric structures that are partially stabilized by G·C·G·C tetrads [41, 42]. Moreover, guanine-rich sequences showing a 5'-CG end may form multimeric structures that are stabilized by the G(:C):G(:C):G(:C):G(:C) octad, a particular arrangement of a G-tetrad interacting with 5'-ending cytosines [43, 44]. The experimental data reported in the present work point to the formation of such complex G-quadruplex structures. However, a solid conclusion can be drawn only from a more in-depth, structural study, which was considered outside the initial scope of this work.

Both topology and polymorphism may be strongly influenced by the treatment carried out on the DNA prior to binding studies [11]. Therefore, the effect of the treatment was studied in depth. It was observed that the most efficient way to break multimers lies in the deprotonation of the functional groups involved in the formation of the G-quartet. This observation complements previous studies, in which samples prepared with thermal treatment showed multimer formation to a lesser extent than dialyzed samples [42, 45]. Again, complementary methods and techniques such as spectroscopically monitored melting, SEC and ESI-MS have shown to be a useful strategy for arriving at this conclusion.

In a second phase, interaction of the selected sequences with CV revealed that this ligand interacts to a slightly greater extent with anti-parallel than with parallel or even hybrid structures. This was concluded based on enhanced fluorescence observed upon binding and on the higher values of the calculated binding constant and thermodynamic parameters associated with the unfolding processes of the DNA-CV complex in relation to free DNA. This observation complements previous studies by Kong *et al.* [16, 17] that examined the interaction of CV with several DNA structures (single strand, duplex and several G-quadruplex topologies). These works did not consider the polymorphism of the selected telomeric DNA sequences. On the other hand, the stoichiometry proposed for the DNA-CV interaction (1:1 in the present work) differed from that proposed for the interaction of CV with the anti-parallel G-quadruplex formed by repeated subunits of the *Oxytricha* telomere (1:2, DNA-CV). The reason for this could be the different sequence studied or the different data analysis method. The binding constant for each binding site, however, was similar (around 10^5 M^{-1}, in both cases), which clearly points to a weak binding mechanism. Interestingly, by using G-quadruplex structures that differed in the spatial arrangement of loops connecting G-tetrads, it was concluded that enhancement of CV

fluorescence upon interaction with DNA was dependent on the nature of the loops connecting guanine-rich tracts [17]. The interaction of other ligands with the wild-type ckit21T21 sequence was also studied. Among these, diarylethynyl amides were shown to interact with a 1:1 stoichiometry, like CV, and to present a stronger affinity (K_a equal to $10^{6.2}$ M^{-1}). Interestingly, these ligands favor the formation of a parallel topology [46, 47], as observed in the present work for the interaction of CV with parallel or hybrid topologies. Again, the potential polymorphism of DNA sequences was not considered in these studies.

Many works on the interaction of ligands with G-quadruplexes have been published. However, very few take the polymorphism of the selected sequences into account when addressing the binding results. One of these studies examined the interaction of the porphyrin TMPyP4 with a guanine-rich sequence from the promoter region of *c-myc* gene by means of SEC [8]. This sequence showed a high degree of polymorphism and the results were explained in terms of the preferential binding of TMPyP4 to multimer structures over monomers, thus indicating stabilization of the junctions of multimers by porphyrin end-stacking. In this regard, the results of the present work (melting experiments, binding titrations and SEC) suggest that crystal violet interacts more strongly with monomers than with dimers, a conclusion that differs slightly from the conclusion drawn in that study. The reason for this discrepancy could be due to both the DNA sequence and the chosen ligand.

6. Conclusions

In this work, the interaction of the crystal violet (CV) ligand with parallel, hybrid and anti-parallel G-quadruplex structures was studied. The polymorphism of the three G-quadruplex-forming sequences was characterized by means of several instrumental techniques prior to the binding studies. Overall, the results showed that CV stabilizes all G-quadruplex structures with 1:1 stoichiometries and relatively low binding constants at 25°C. However, separation techniques were used for a closer examination and revealed the preferential binding of CV with the monomeric G-quadruplex over the multimeric structures. This fact should be considered when studying the binding of ligands to polymorphic DNA sequences, such as G-quadruplex-forming sequences.

7. Acknowledgments

8. References

[1] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, Quadruplex DNA: sequence, topology and structure, Nucleic Acids Research 34 (2006) 5402-5415.

[2] S. Neidle, Quadruplex Nucleic Acids as Novel Therapeutic Targets, Journal of Medicinal Chemistry 59 (2016) 5987-6011.

[3] G. Biffi, M. Di Antonio, D. Tannahill, S. Balasubramanian, Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells, Nat Chem 6 (2014) 75-80.

[4] H.-Z. He, D.S.-H. Chan, C.-H. Leung, D.-L. Ma, G-quadruplexes for luminescent sensing and logic gates, Nucleic Acids Research 41 (2013) 4345-4359.

[5] B. Juskowiak, Analytical potential of the quadruplex DNA-based FRET probes, Analytica Chimica Acta 568 (2006) 171-180.

[6] V.C. Diculescu, A.-M. Chiorcea-Paquim, A.M. Oliveira-Brett, Applications of a DNA-electrochemical biosensor, TRAC Trends in Analytical Chemistry 79 (2016) 23-36.

[7] Y. Fu, X. Wang, J. Zhang, W. Li, Nanomaterials and nanoclusters based on DNA modulation, Current Opinion in Biotechnology 28 (2014) 33-38.

[8] H.T. Le, M.C. Miller, R. Buscaglia, W.L. Dean, P.A. Holt, J.B. Chaires, J.O. Trent, Not all G-quadruplexes are created equally: an investigation of the structural polymorphism of the c-Myc G-quadruplex-forming sequence and its interaction with the porphyrin TMPyP4, Organic & Biomolecular Chemistry 10 (2012) 9393-9404.

[9] P. Tóthová, P. Krafčíková, V. Víglaský, Formation of Highly Ordered Multimers in G-Quadruplexes, Biochemistry 53 (2014) 7013-7027.

[10] S. Kolesnikova, M. Hubálek, L. Bednárová, J. Cvacka, E.A. Curtis, Multimerization rules for G-quadruplexes, Nucleic Acids Research 45 (2017) 8684-8696.

[11] M.M. Dailey, M.C. Miller, P.J. Bates, A.N. Lane, J.O. Trent, Resolution and characterization of the structural polymorphism of a single quadruplex-forming sequence, Nucleic Acids Research 38 (2010) 4877-4888.

[12] N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E.S. Baker, M.T. Bowers, E. De Pauw, V. Gabelica, G-quadruplex DNA assemblies: Loop length, cation identity, and multimer formation, Journal of the American Chemical Society 130 (2008) 10208-10216.

[13] M. Bejugam, S. Sewitz, P.S. Shirude, R. Rodriguez, R. Shahid, Balasubramanian.S, Trisubstituted Isoalloxazines as a New Class of G-Quadruplex Binding Ligands: Small Molecule Regulation of c-kit Oncogene Expression, J. Am. Chem. Soc. 129 (2007) 12926-12927.

[14] J. Bidzinska, G. Cimino-Reale, N. Zaffaroni, M. Folini, G-Quadruplex Structures in the Human Genome as Novel Therapeutic Targets, Molecules 18 (2013).

[15] B. Pagano, S. Cosconati, V. Gabelica, L. Petraccone, S. De Tito, L. Marinelli, V. La Pietra, F.S. di Leva, I. Lauri, R. Trotta, E. Novellino, C. Giancola, A. Randazzo, State-of-the-Art Methodologies for the Discovery and Characterization of DNA G-Quadruplex Binders, Current Pharmaceutical Design 18 (2012) 1880-1899.

[16]   D.-M. Kong, Y.-E. Ma, J. Wu, H.-X. Shen, Discrimination of G-Quadruplexes from Duplex and Single-Stranded DNAs with Fluorescence and Energy-Transfer Fluorescence Spectra of Crystal Violet, Chemistry – A European Journal 15 (2009) 901-909.

[17]   D.-M. Kong, Y.-E. Ma, J.-H. Guo, W. Yang, H.-X. Shen, Fluorescent Sensor for Monitoring Structural Changes of G-Quadruplexes and Detection of Potassium Ion, Analytical Chemistry 81 (2009) 2678-2684.

[18]   D.-M. Kong, J.-H. Guo, W. Yang, Y.-E. Ma, H.-X. Shen, Crystal violet-G-quadruplex complexes as fluorescent sensors for homogeneous detection of potassium ion, Biosensors and Bioelectronics 25 (2009) 88-93.

[19]   B. Zheng, S. Cheng, H. Dong, H. Liang, J. Liu, M.H.-W. Lam, Label Free Determination of Potassium Ions Using Crystal Violet and Thrombin-Binding Aptamer, Analytical Letters 47 (2014) 1726-1736.

[20]   L. Zhang, B. Han, T. Li, E. Wang, Label-free DNAzyme-based fluorescing molecular switch for sensitive and selective detection of lead ions, Chemical Communications 47 (2011) 3099-3101.

[21]   Q. Shen, L. Zhou, Y. Yuan, Y. Huang, B. Xiang, C. Chen, Z. Nie, S. Yao, Intra-molecular G-quadruplex structure generated by DNA-templated click chemistry: "Turn-on" fluorescent probe for copper ions, Biosensors and Bioelectronics 55 (2014) 187-194.

[22]   H.-Z. He, V. Pui-Yan Ma, K.-H. Leung, D. Shiu-Hin Chan, H. Yang, Z. Cheng, C.-H. Leung, D.-L. Ma, A label-free G-quadruplex-based switch-on fluorescence assay for the selective detection of ATP, Analyst 137 (2012) 1538-1540.

[23]   C.-H. Leung, D.S.-H. Chan, B.Y.-W. Man, C.-J. Wang, W. Lam, Y.-C. Cheng, W.-F. Fong, W.-L.W. Hsiao, D.-L. Ma, Simple and Convenient G-Quadruplex-Based Turn-On Fluorescence Assay for 3' -> 5' Exonuclease Activity, Analytical Chemistry 83 (2011) 463-466.

[24]   H. Fernando, A.P. Reszka, J. Huppert, S. Ladame, S. Rankin, A.R. Venkitaraman, S. Neidle, S. Balasubramanian, A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene, Biochemistry 45 (2006) 7854-7860.

[25]   V. Kuryavyi, A.T.n. Phan, D.J. Patel, Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter, Nucleic Acids Research 38 (2010) 6757-6773.

[26]   S.-T.D. Hsu, P. Varnai, A. Bugaut, A.P. Reszka, S. Neidle, S. Balasubramanian, A G-Rich Sequence within the c-kit Oncogene Promoter Forms a Parallel G-Quadruplex Having Asymmetric G-Tetrad Dynamics, Journal of the American Chemical Society 131 (2009) 13399 - 13409.

[27]   S. Manaye, R. Eritja, A. Aviño, J. Jaumot, R. Gargallo, Porphyrin binding mechanism is altered by protonation at the loops in G-quadruplex DNA formed near the transcriptional activation site of the human c-kit gene, Biochimica et Biophysica Acta (BBA) - General Subjects 1820 (2009) 1987-1996.

[28]   W.A. Kibbe, OligoCalc: An online oligonucleotide properties calculator, Nucleic Acids Research 35 (2007) W43-W46.

[29]   R.F. Lacey, Deconvolution of overlapping chromatographic peaks, Analytical Chemistry 58 (1986) 1404-1410.

[30]   K.J. Breslauer, Extracting thermodynamic data from equilibrium melting curves for oligonucleotide order-disorder transitions, Methods in Enzymology, vol. 259, Academic Press, 1995, pp. 221-242.

[31]   J. Jaumot, R. Tauler, R. Gargallo, Exploratory data analysis of DNA microarrays by multivariate curve resolution, Analytical Biochemistry 358 (2006) 76-89.

[32]   R. Gargallo, Hard/Soft hybrid modeling of temperature-induced unfolding processes involving G-quadruplex and i-motif nucleic acid structures, Analytical Biochemistry 466 (2014) 4-15.

[33]   R. Dyson, S. Kaderli, G.A. Lawrence, M. Maeder, A.D. Zuberbühler, Second order global analysis: the evaluation of series of spectrophotometric titrations for improved determination of equilibrium constants, Analytica Chimica Acta 353 (1997) 381-393.

[34]   G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, Quantitative visualization of DNA G-quadruplex structures in human cells, Nat Chem 5 (2013) 182-186.

[35]   G.B. Irvine, Size-exclusion high-performance liquid chromatography of peptides: a review, Analytica Chimica Acta 352 (1997) 387-397.

[36] T. Ilc, P. Sket, J. Plavec, M. Webba da Silva, I. Drevensek-Olenik, L. Spindler, Formation of G-Wires: The Role of G:C-Base Pairing and G-Quartet Stacking, The Journal of Physical Chemistry C 117 (2013) 23208-23215.

[37] T.C. Marsh, E. Henderson, G-Wires: Self-Assembly of a Telomeric Oligonucleotide, d(GGGGTTGGGG), into Large Superstructures, Biochemistry 33 (1994) 10718-10724.

[38] F. Rosu, V. Gabelica, C. Houssier, P. Colson, E. De Pauw, Triplex and quadruplex DNA structures studied by electrospray mass spectrometry, Rapid Communications in Mass Spectrometry 16 (2002) 1729-1736.

[39] E.W. White, F. Tanious, M.A. Ismail, A.P. Reszka, S. Neidle, D.W. Boykin, W.D. Wilson, Structure-specific recognition of quadruplex DNA by organic cations: influence of shape, substituents and charge, Biophys Chem 126 (2007) 140-153.

[40] M.L. Greco, A. Kotar, R. Rigo, C. Cristofari, J. Plavec, C. Sissi, Coexistence of two main folded G-quadruplexes within a single G-rich domain in the EGFR promoter, Nucleic Acids Research 45 (2017) 10132-10142.

[41] M. Webba da Silva, Association of DNA Quadruplexes through G:C:G:C Tetrads. Solution Structure of d(GCGGTGGAT), Biochemistry 42 (2003) 14356-14365.

[42] L. Spindler, M. Rigler, I. Drevensek-Olenik, N. Ma'ani Hessari, M. Webba da Silva, Effect of Base Sequence on G-Wire Formation in Solution, Journal of Nucleic Acids 2010 (2010).

[43] N. Borbone, J. Amato, G. Oliviero, V. D'Atri, V. Gabelica, E. De Pauw, G. Piccialli, L. Mayol, d(CGGTGGT) forms an octameric parallel G-quadruplex via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads, Nucleic Acids Research 39 (2011) 7848-7857.

[44] V. D'Atri, N. Borbone, J. Amato, V. Gabelica, S. D'Errico, G. Piccialli, L. Mayol, G. Oliviero, DNA-based nanostructures: The effect of the base sequence on octamer formation from d(XGGYGGT) tetramolecular G-quadruplexes, Biochimie 99 (2014) 119-128.

[45] M.C. Miller, H.T. Le, W.L. Dean, P.A. Holt, J.B. Chaires, J.O. Trent, Polymorphism and resolution of oncogene promoter quadruplex-forming sequences, Organic & Biomolecular Chemistry 9 (2011) 7633-7637.

[46] J. Dash, P.S. Shirude, S.-T.D. Hsu, S. Balasubramanian, Diarylethynyl Amides That Recognize the Parallel Conformation of Genomic Promoter DNA G-Quadruplexes, Journal of the American Chemical Society 130 (2008) 15950-15956.

[47] J. Dash, Z.A.E. Waller, G.D. Pantoş, S. Balasubramanian, Synthesis and Binding Studies of Novel Diethynyl-Pyridine Amides with Genomic Promoter DNA G-Quadruplexes, Chemistry – A European Journal 17 (2011) 4571-4581.

Table 1. Calculated thermodynamic variables for the folding process. Apart from the folding of ckit21T18, univariate analysis of the ellipticity data recorded at 263 or 292 nm was used and a two-state model was considered. In the case of ckit21T18, multivariate analysis was applied, as this melting could not be modeled as a simple two-state process (Figure S3).

System	T_m (°C)	ΔH^0 (kcal·mol^{-1})	ΔS^0 (cal·K^{-1}·mol^{-1})	$\Delta G_{37°C}$ (kcal·mol^{-1})
ckit21T21	61±1	-38±2	-112±3	-2.8±0.4
ckit21T21 + CV (1:5 ratio)	65±2	-64±3	-190±13	-5.3±0.2
ckit21T18, 1st	40±2[a,b]			
ckit21T18, 2nd	58±1 [a,b]			
ckit21T18 + CV (1:5 ratio)	65±1 [a,b]			
SMG03	~30	n.c.	n.c.	n.c.
SMG03 + CV (1:5 ratio)	39±2	-31±2	-99±5	-0.2±0.4

[a]: Calculated from the multivariate analysis (see Figure S3).

[b]: As hysteresis phenomenon is present and the reaction is not at thermodynamic equilibrium, these values actually correspond to $T_{1/2}$. The thermodynamic parameters could not be determined accurately.

n.c.: Not calculated because of an unreliable definition of one of the baselines.

Figure 1. Formation of G-quadruplex structures from the G-tetrad (a) and schematic representation of G-quadruplex topologies (b). The terms "parallel", "anti-parallel" and "3+1 hybrid" are used to describe the relative orientation of the four strands. The presence of potential loops linking DNA strands were omitted for the sake of clarity. Circles denote cations located between two adjacent G-tetrads. (c) Chemical structure of the acid-base form of crystal violet (CV) at pH 7.0. (d) DNA sequences studied in this work, together with the proposed topology at the experimental conditions used.

Sequence	Sequence (5'→ 3')	Proposed topology
ckit1T21	C GGG C GGG CGCGA GGG A GGG T	parallel
ckit1T18	C GGG C GGG CGCGA GGG AT GGG	Parallel / antiparallel
SMG03	AA GGG CGA GG CA GG ACA GGG A	anti-parallel
T21	TTT TTT TTT TTT TTT TTT TTT	Unfolded strand
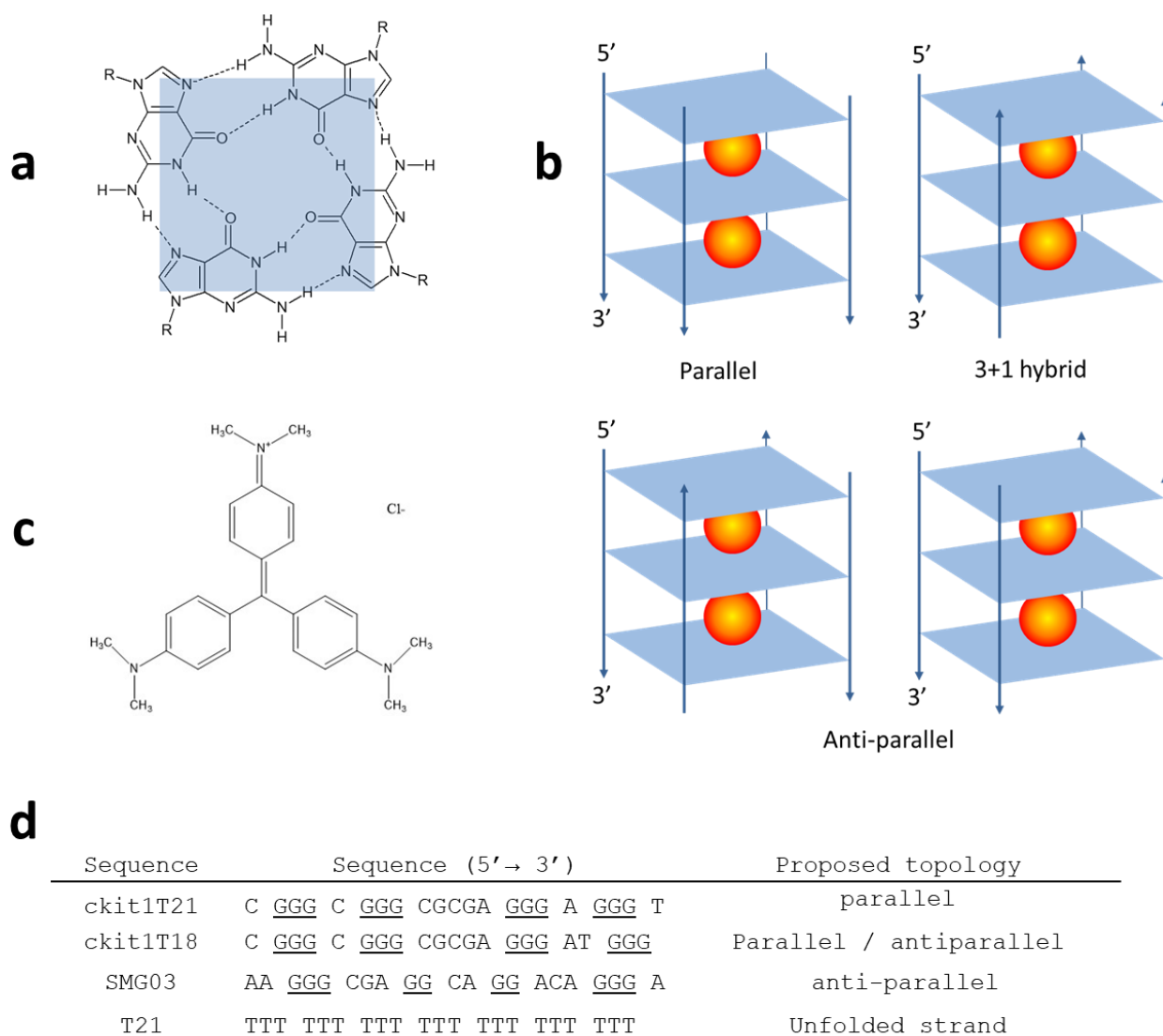
Figure 2. CD spectra of the studied sequences measured after heating at 90°C for 10 minutes and slow cooling overnight. C_{DNA} = 2 μM, 50mM KCl, 20mM phosphate buffer, pH 7.0, 25°C.
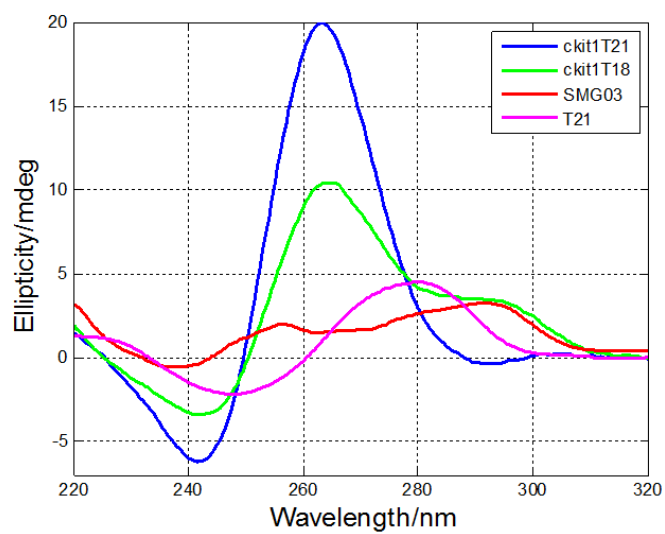
Figure 3. Denaturation and renaturation of ckit21T21. (a) CD spectra recorded throughout the process to melt a ckit21T21 2μM sample in 50 mM KCl, 20 mM phosphate buffer, pH 7.0 prepared from the stock solution, heated and cooled overnight. (b) CD spectrum measured after the addition of LiOH 1 M to the sample heated in (a) up to pH 12.0. (c) CD spectrum measured after the addition of HCl 1 M to the sample in (b) until pH was equal to 7.0. (d) CD spectra recorded throughout the process to melt the sample in (c).
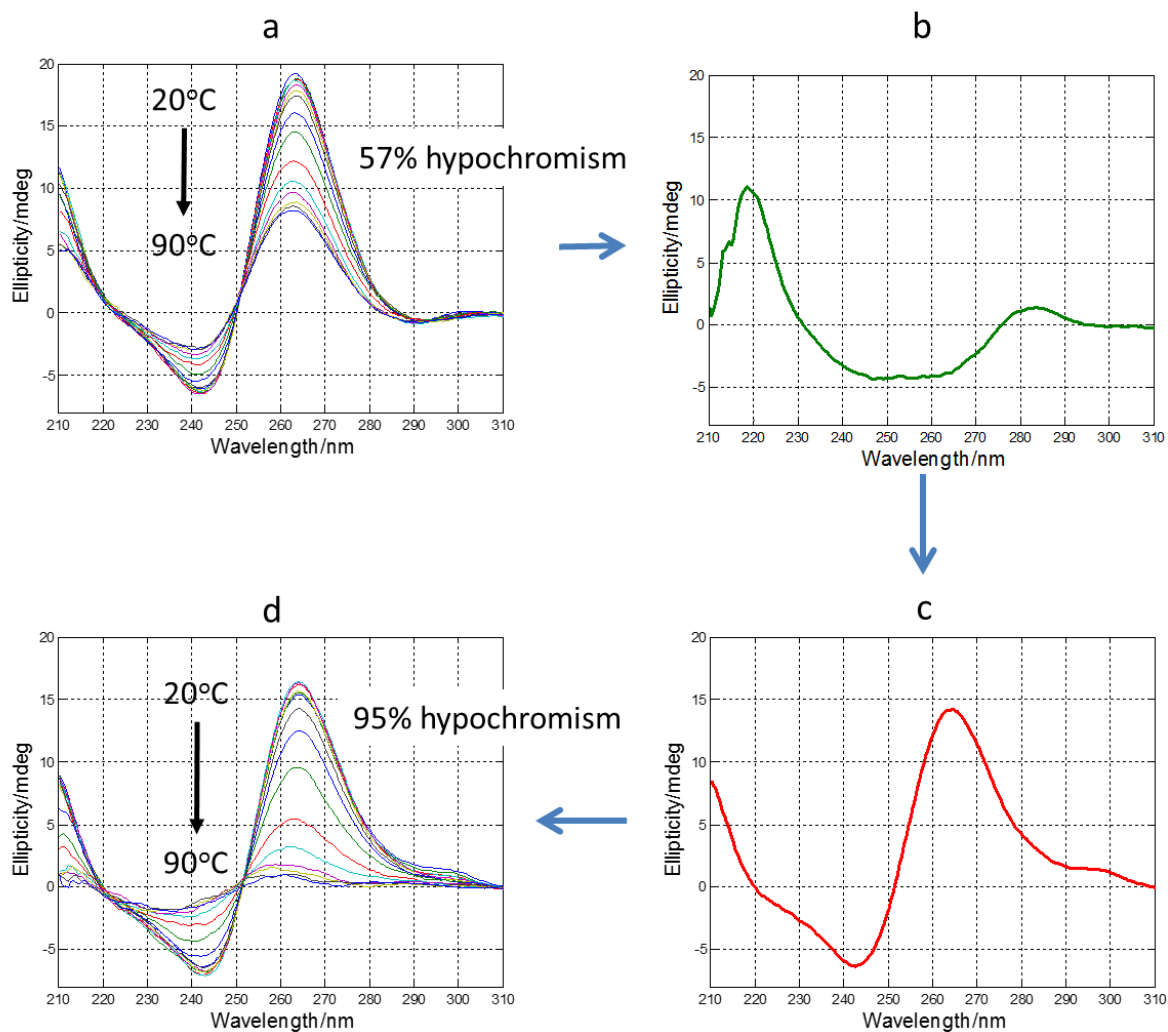
Figure 4. SEC profiles obtained for T21 (a), ckit21T21 (b), ckit21T18 (c), and SMG03 (d). Chromatograms have been normalized. C_{DNA} = 10 μM, 300 mM KCl, 20 mM phosphate buffer, pH 7.0, 25°C. A BioSep-SEC-S 3000 column was used.
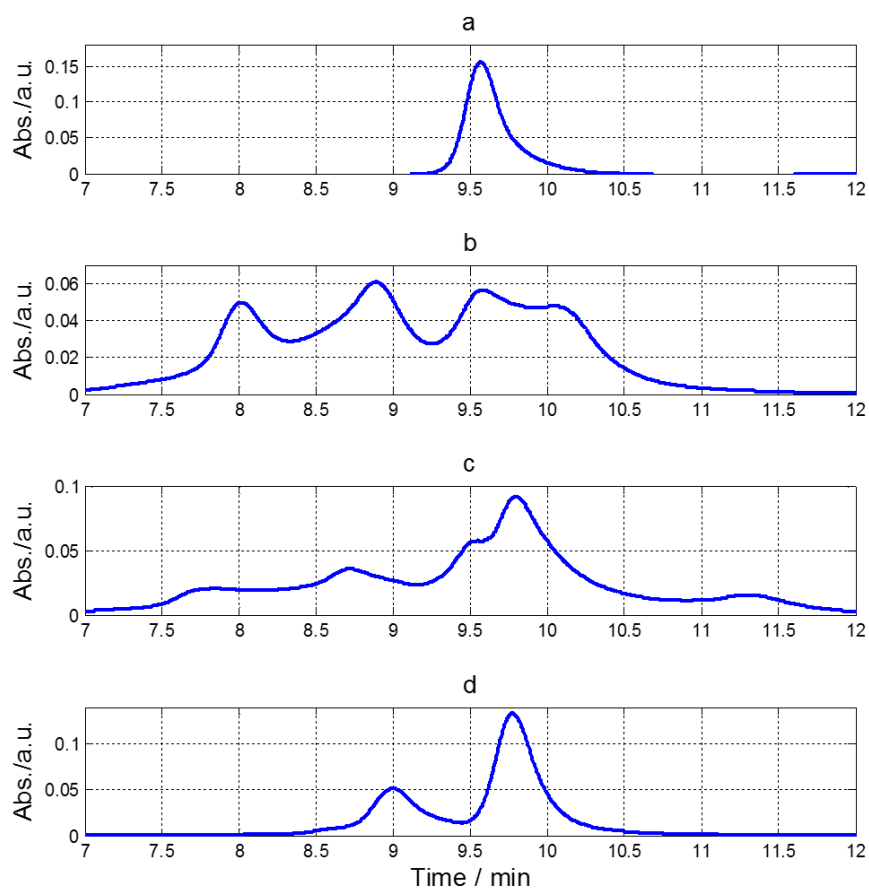
Figure 5. CD-monitored titration of DNA samples with CV. (a) ckit21T21, (b) ckit21T18, (c) SMG03, (d) T21. In all cases, C_{DNA} = 2μM, 25°C, 50 mM KCl, 20 mM phosphate buffer, pH 7.0. Insets describe the CV:DNA ratios at which spectra were measured.
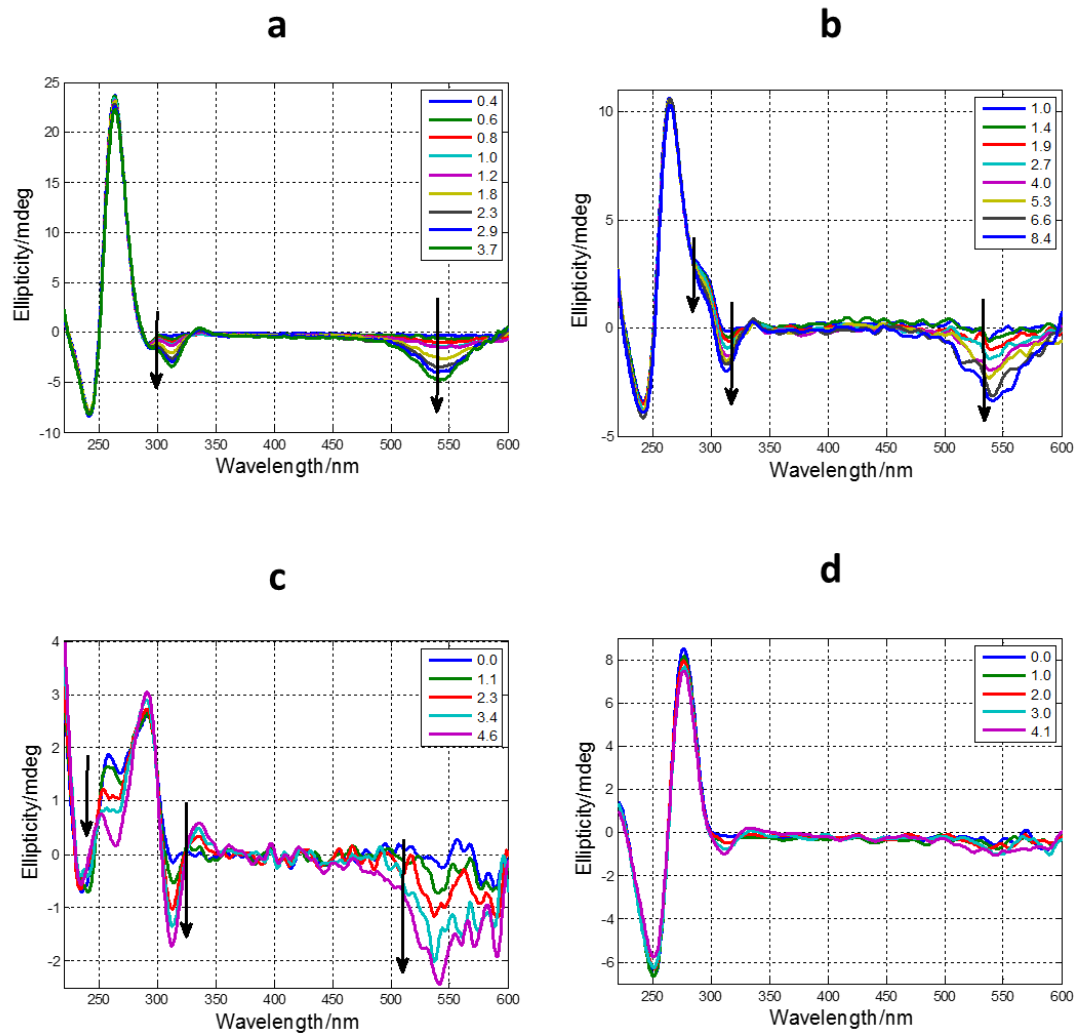
Figure 6. Selected spectra measured along the molecular fluorescence-monitored titrations of CV sample with DNAs. (a) ckit21T21, (b) ckit21T18, (c) SMG03, (d) T21. C_{CV} = 2μM, 25°C, 50 mM KCl, 20 mM phosphate buffer, pH 7.0. Insets describe the DNA:CV ratios at which spectra were measured.
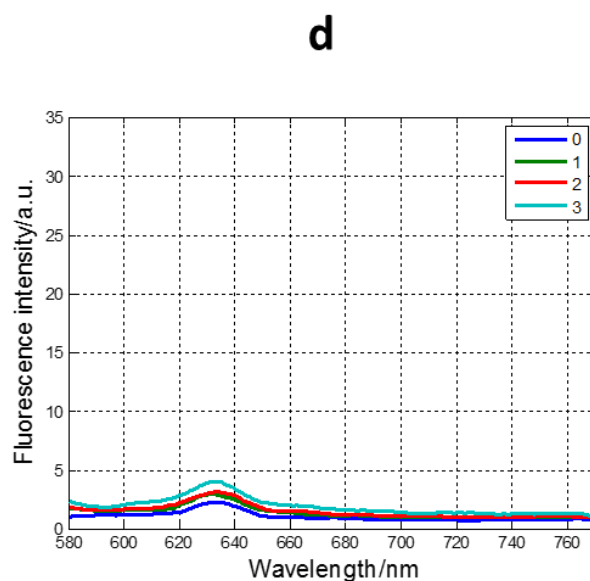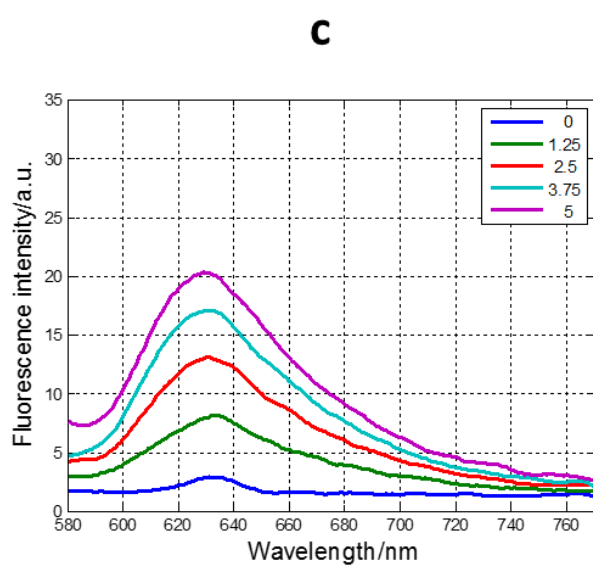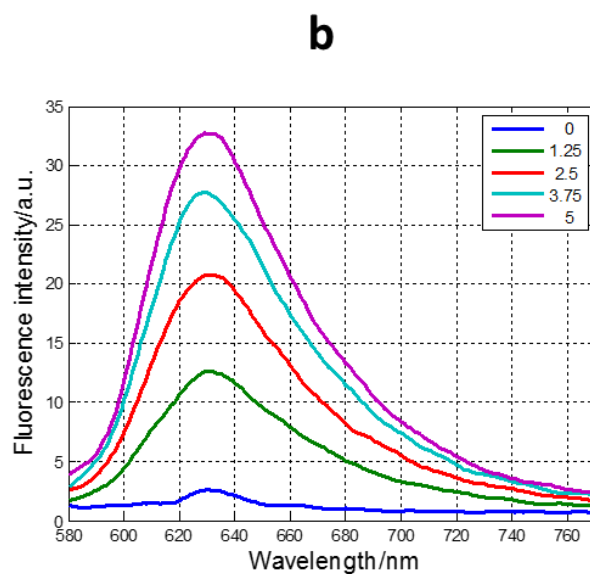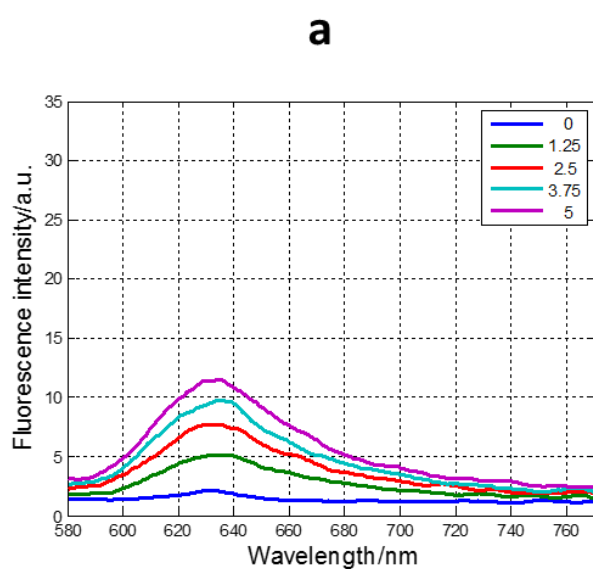
Figure 7. Melting of mixtures of CV with ckit21T21 (a), ckit21T18 (b) and SMG03 (c). Insets show the variation of ellipticity at 545 nm along the experiment. In all cases, C_{DNA} = 2 μM, C_{CV} = 10 μM, 50 mM KCl, 20 mM phosphate buffer, pH 7.0.
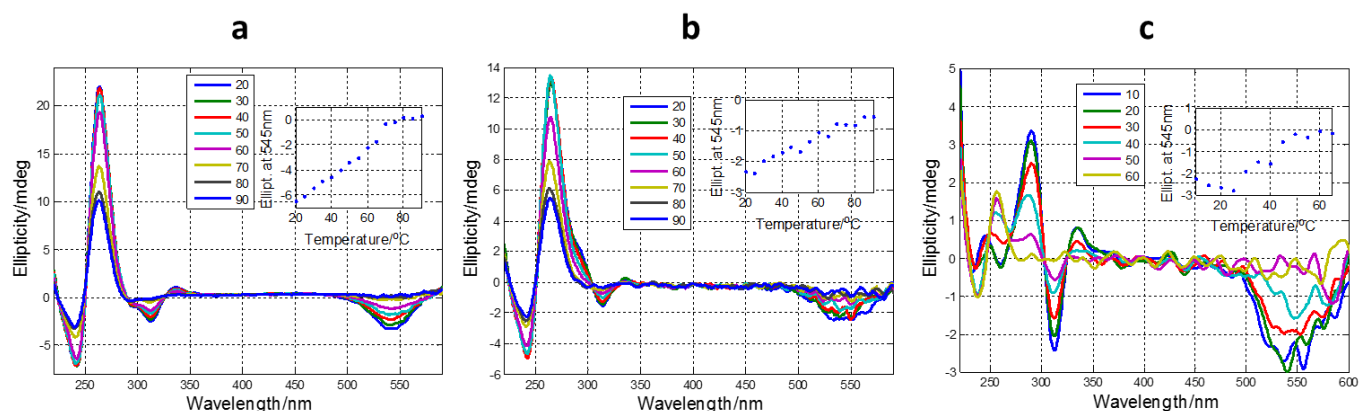
Figure 8. Normalized SEC profiles obtained for mixtures of ckit21T21 (a), ckit21T18 (b), and SMG03 (c) with CV. C_{DNA} = 10 μM, 300 mM KCl, 20 mM phosphate buffer, pH 7.0, 25°C. A BioSep-SEC-S 3000 column was used.
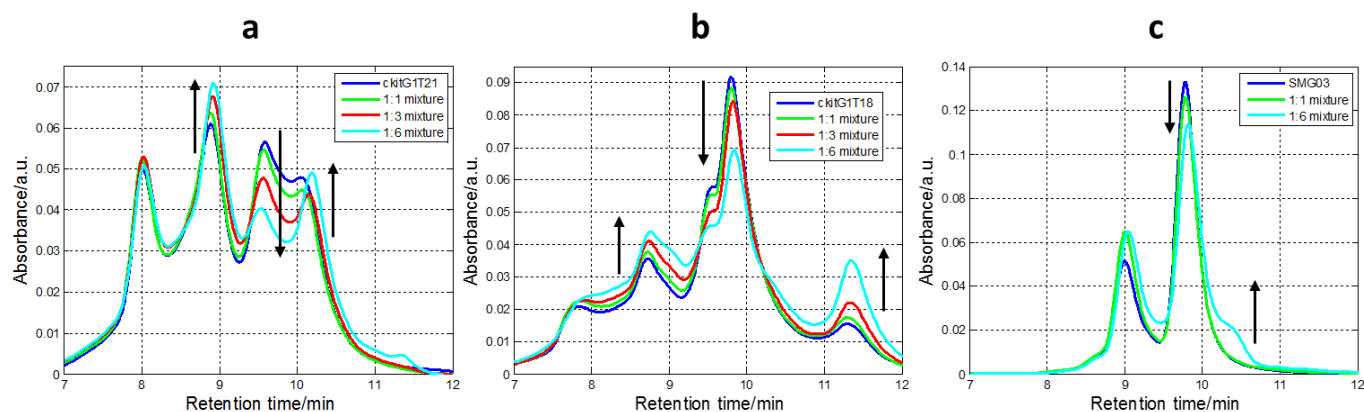
Electronic Supplementary Information

Evaluation of the effect of polymorphism on G-quadruplex-ligand interaction by means of spectroscopic and chromatographic techniques

S. Benito[1], A. Ferrer[1], S. Benabou[1], A. Aviñó[2], R. Eritja[2], R. Gargallo[1]*

[1] Department of Chemical Engineering and Analytical Chemistry. University of Barcelona. Martí i Franquès, 1-11, E-08028 Barcelona, Spain

[2] Institute for Advanced Chemistry of Catalonia (IQAC-CSIC), CIBER-BBN, Jordi Girona 18-26, E-08034 Barcelona, Spain.

Contents:

S1. Van't Hoff plots for the meltings of ckit21T21. (a) Melting of ckit21T21 sample after the thermal treatment. (b) Melting after the successive additions of LiOH and HCl. The thermodynamic parameters shown correspond to one of the three replicates done.

a
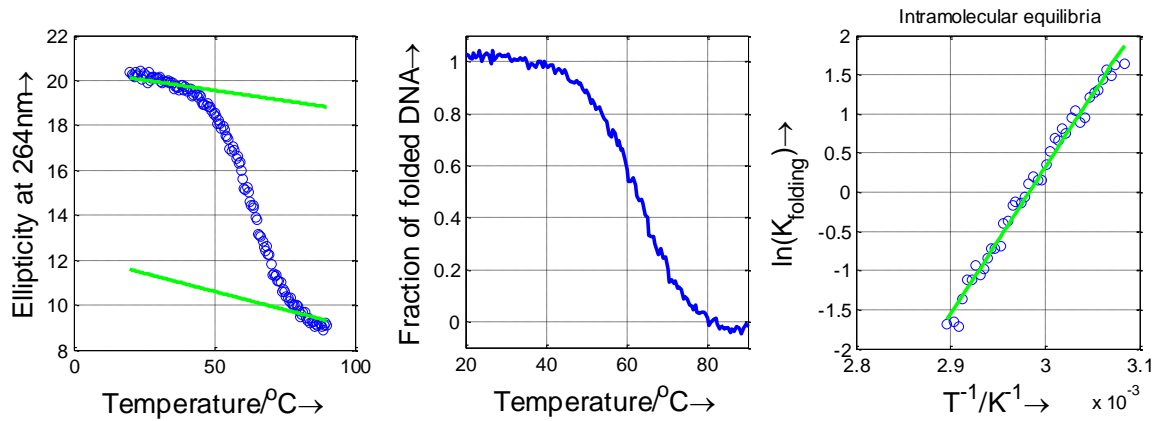
Calculated value for Tm (°C)= 61.6

ΔH (kcal·mol^{-1})= -37.1 +- 1.0 (95%)

ΔS (cal·K^{-1}·mol^{-1})= -110.0 +- 2.9 (95%)

ΔG_{37} (kcal·mol^{-1})= -3.0

b
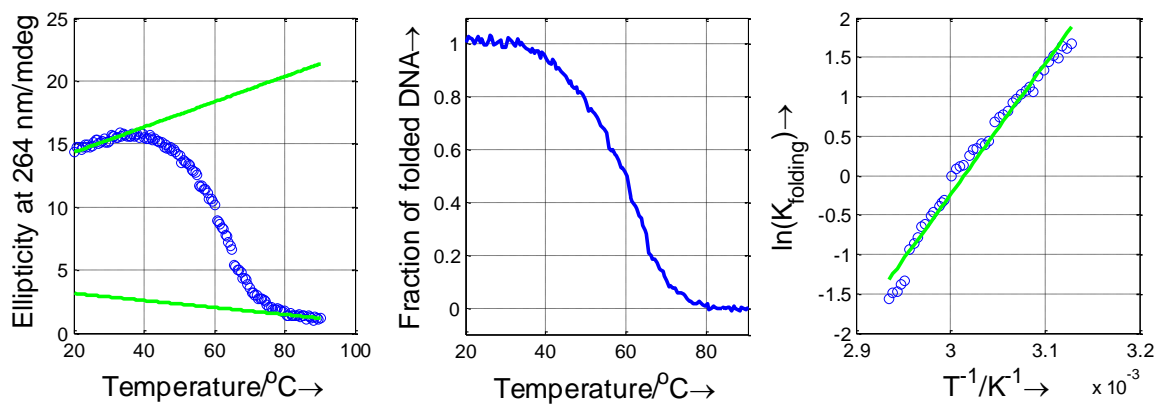
Calculated value for Tm (°C)= 58.8

ΔH (kcal·mol^{-1})= -34.0 +- 1.4 (95%)

ΔS (cal·K^{-1}·mol^{-1})= -102.4 +- 4.4(95%)

ΔG_{37} (kcal·mol^{-1})= -2.2

S2. Denaturation and renaturation of ckit21T18. (a) CD spectra recorded along the melting of a ckit21T18 0.8 μM sample in 50 mM KCl, 20 mM phosphate buffer, pH 7.0 prepared from the stock solution, heated and cooled overnight. (b) CD spectrum after addition of LiOH 1 M up to pH 12.0 to the sample heated in (a), (c) CD spectrum after the addition of HCl 1 M to the previous sample until pH was equal to 7.0. (d) CD spectra recorded along the melting of the sample at pH 7 after addition of LiOH and HCl.
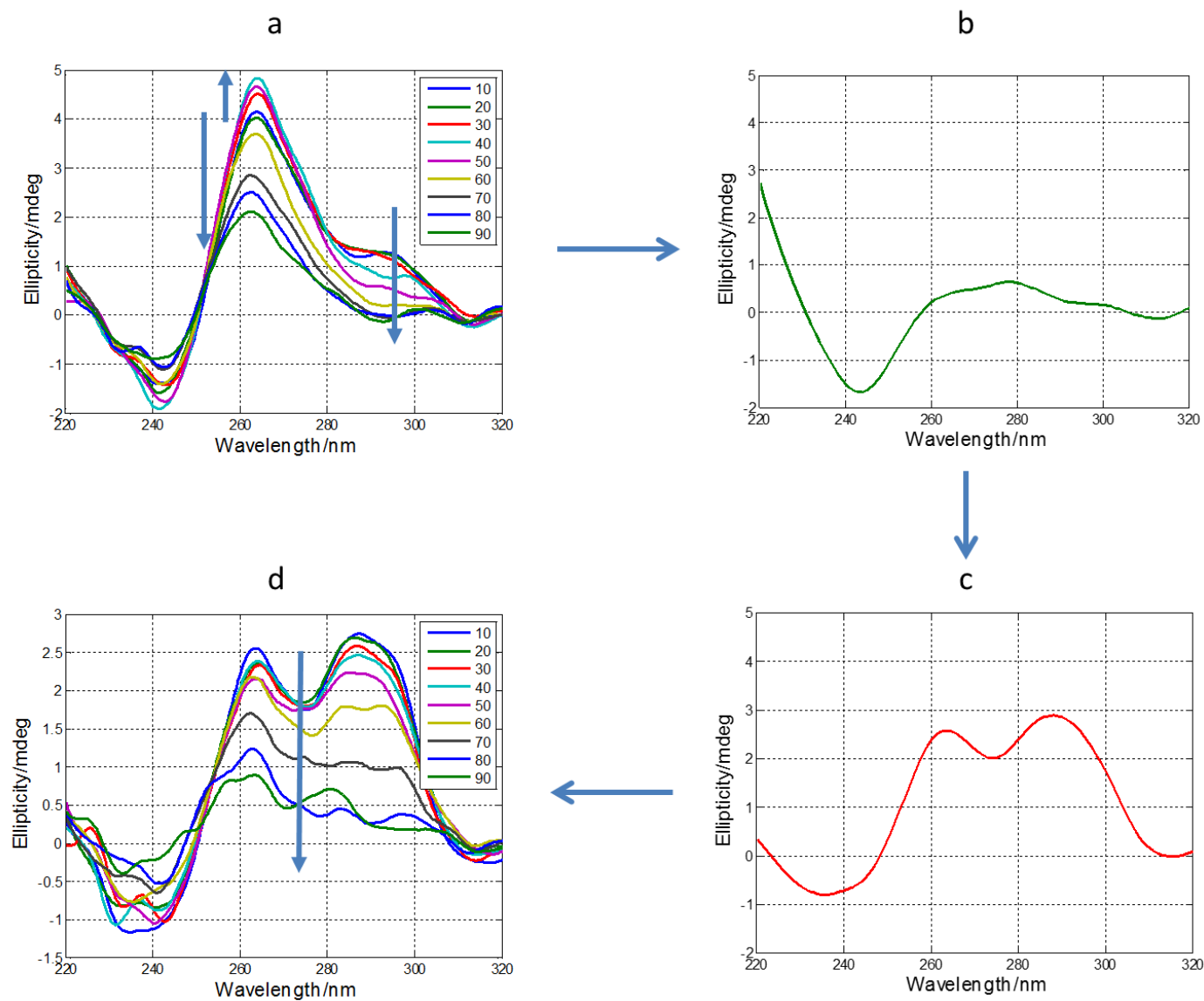
S3. Analysis of data recorded along the meltings of ckit21T18.

(A) Melting of ckit21T18 sample after the thermal treatment.

The mathematical analysis of CD spectra recorded along the melting of ckit21T18 after the thermal denaturation (Figure S2a) revealed the presence of three spectroscopically-active species, what results in two successive melting transitions. This may be also observed from the ellipticity data recorded at 263 nm:
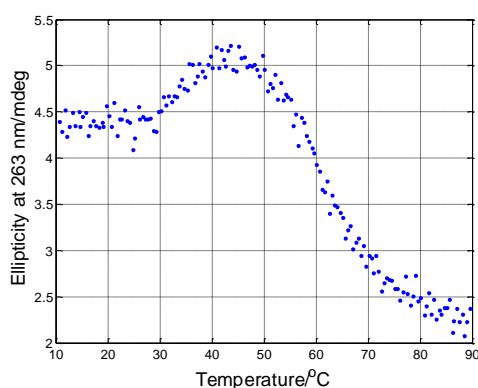
As a result, the whole data could not be analyzed considering a single two-state process. Accordingly, spectra were analyzed by means of a multivariate analysis method (Gargallo, 2014) that enables the calculation of thermodynamic parameters and T_m values. The distribution diagram and pure spectra for each one of the three resolved components are shown in next figure (panels a and b, respectively. Panels (c) and (d) show the quality of the fitting at 293 and 263 nm, respectively (blue symbols and green line represent the experimental and fitted data, respectively).
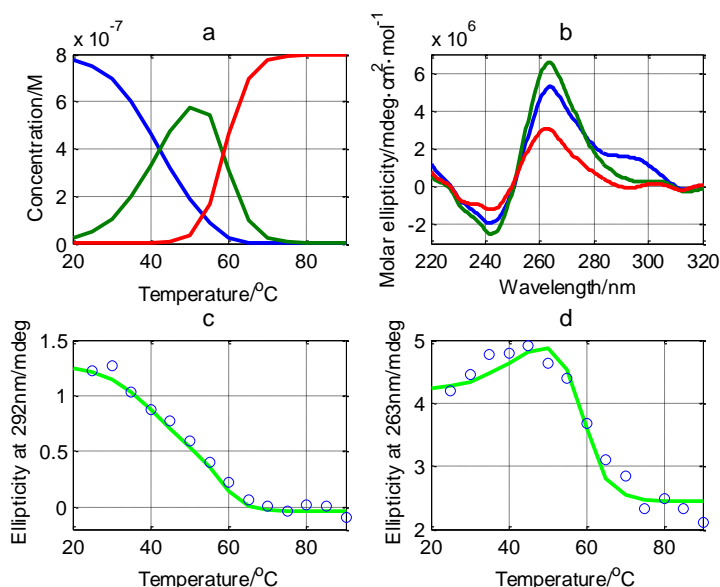
This reaction is not at thermodynamic equilibrium because of the slow kinetics. Therefore, T_m values actually correspond to $T_{1/2}$ and thermodynamic parameters only show a tendency.

(B) Melting of ckit21T18 sample after successive additions of LiOH and HCl.

The mathematical analysis of CD spectra recorded along the melting of ckit21T18 after the thermal denaturation (Figure S2d) revealed the presence of two spectroscopically-active species, what results in just one melting transition. This may be also observed from the ellipticity data recorded at 263 nm:
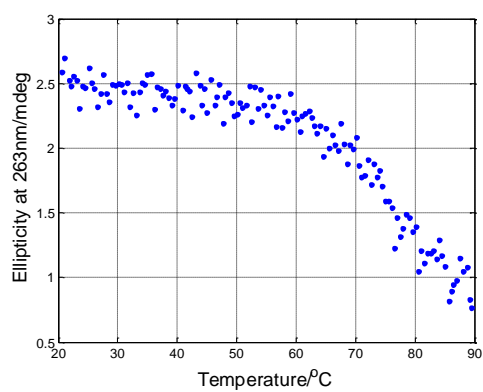
In this case, $T_{1/2}$ value is around 75°C.

S4. Denaturation and renaturation of SMG03. (a) CD spectra recorded along the melting of a SMG03 2µM sample in 50 mM KCl, 20 mM phosphate buffer, pH 7.0 prepared from the stock solution, heated at 90°C and cooled overnight. (b) CD spectrum after addition of LiOH 1 M up to pH 12.0 to the sample heated in (a), (c) CD spectrum after the addition of HCl 1 M to the previous sample until pH was equal to 7.0. (d) CD spectra recorded along the melting of the sample at pH 7 after addition of LiOH and HCl.
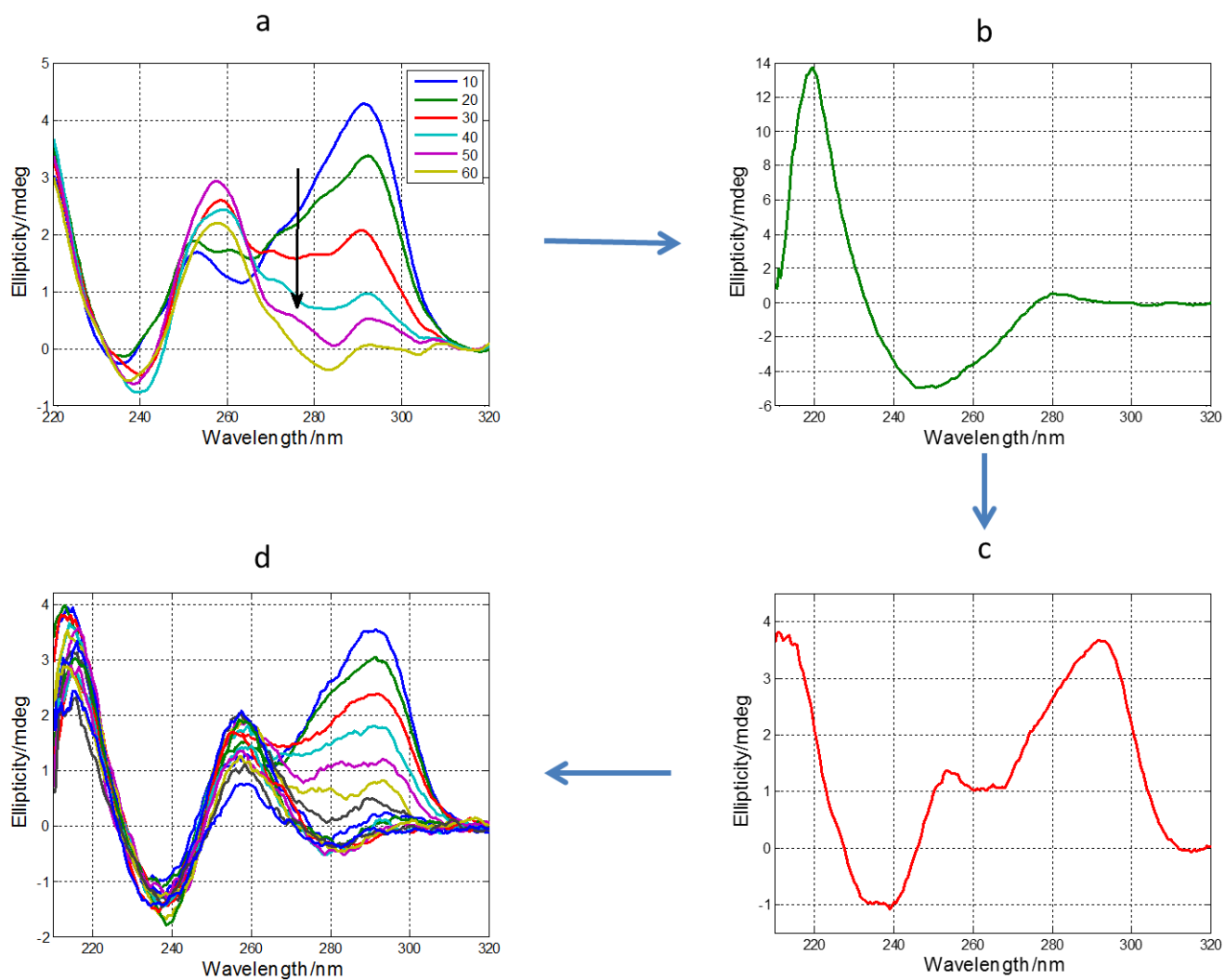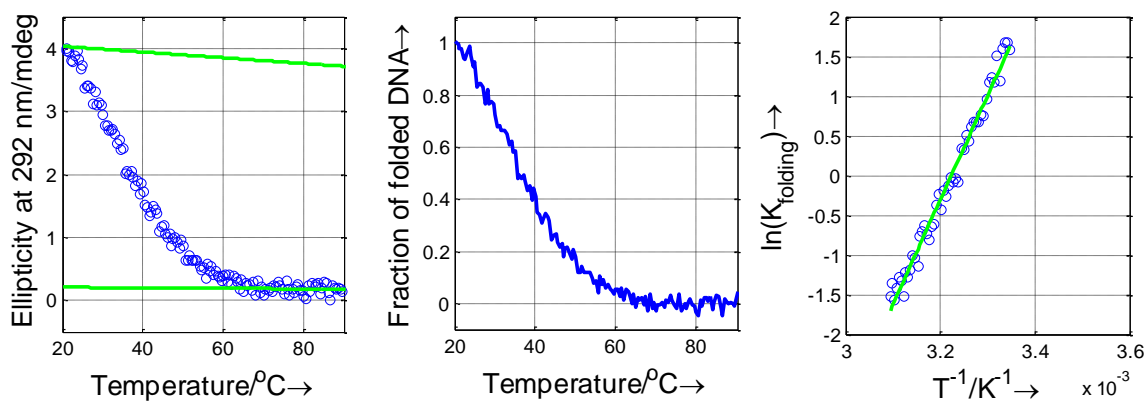
S5. Analysis of data recorded along the meltings of SMG03.

(A) Melting of SMG03 sample after the thermal pre-treatment. Experiments carried out in 50 mM KCl and 20 mM phosphate buffer, pH 7.
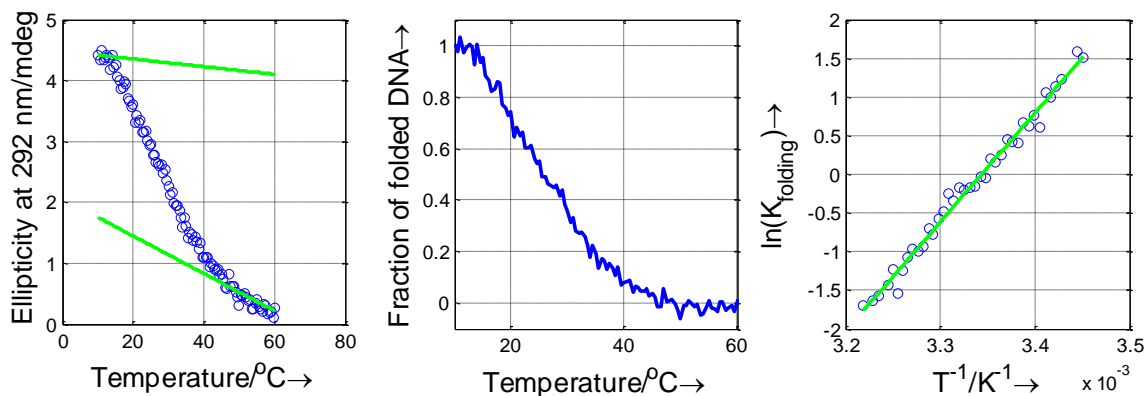

(B) Melting of SMG03 sample after successive additions of LiOH and HCl. Experiments carried out in 50 mM KCl and 20 mM phosphate buffer, pH 7.

In both cases, T_m value is around 30°C. However, the absence of a well-defined baseline at low temperatures makes the calculation of thermodynamic parameters inaccurate.

S6. SEC calibration plot at pH 7.0 (phosphate buffer).

Standard	MW $_{strand}$ (g·mol^{-1})	log (MW $_{strand}$)	Measured t_R (min)
			Biosep S-3000
T15	4501.0	3.65	9.84
T20	6022.0	3.78	9.57
T25	7542.9	3.88	9.37
T30	9063.9	3.96	9.18
T45	13626.9	4.13	8.69
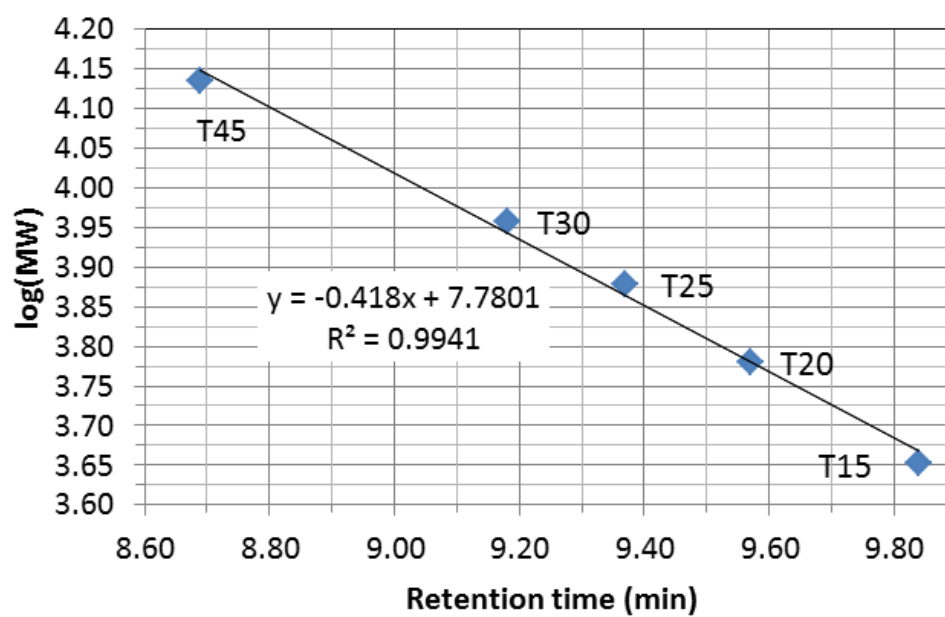
## S7. SEC data for samples at pH 7.0 (phosphate buffer).

| Sequence	sequence MW (g·mol^{-1})	log (MW)	Measured t_R (min)	Calculated MW (g·mol^{-1}) according to calibration plot	calculated MW / sequence MW	Proposed structure
ckit21T21	6634.3	3.82	8.02 [a]	26818.7	4.0	Tetrameric G-quadruplex
			8.89	11606.2	1.7	Dimeric G-quadruplex
			9.58	5973.1	0.9	Unfolded strand
			10.05 [a]	3799.2	0.6	Monomeric G-quadruplex
ckit21T18	6634.3	3.82	7.78 [a]	33789.5	5.1 [a]	Multimeric G-quadruplex
			8.72	13670.0	2.1	Dimeric G-quadruplex
			9.52	6328.3	1.0	Unfolded strand
			9.80	4833.0	0.7	Monomeric parallel G-quadruplex
			11.3 [a]	1108.0	0.2	Monomeric anti-parallel G-quadruplex
SMG03	6619.3	3.82	8.99	10541.0	1.6	Dimeric G-quadruplex
			9.77	4974.7	0.8	Monomeric G-quadruplex

[a] These values lie outside the calibration line. Therefore, molecular weight calculated from these retention times are approximate.

S8. Effect of acid-base treatment on SEC chromatograms of ckit21T21 and ckit21T18. SEC chromatograms recorded for samples of ckit21T21 (a), and ckit21T18 (b). The samples injected after the standard thermal treatment are shown in blue, whereas those injected after the acid-base treatment are shown in green. An Acclaim SEC-300 column was used. Experiments carried out at 25°C, 300 mM KCl and 20 mM phosphate buffer, pH 7.
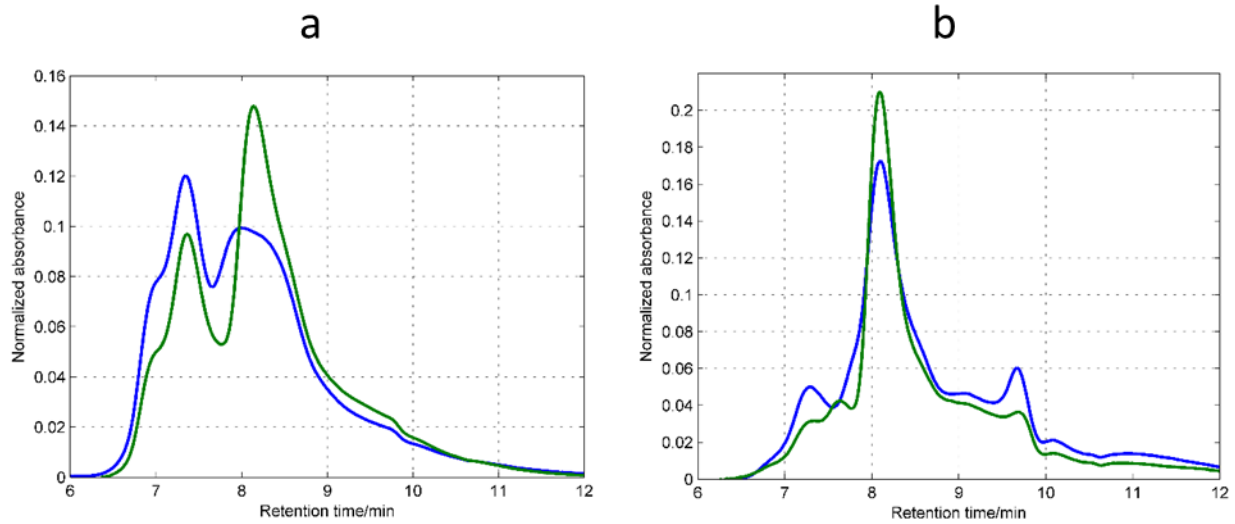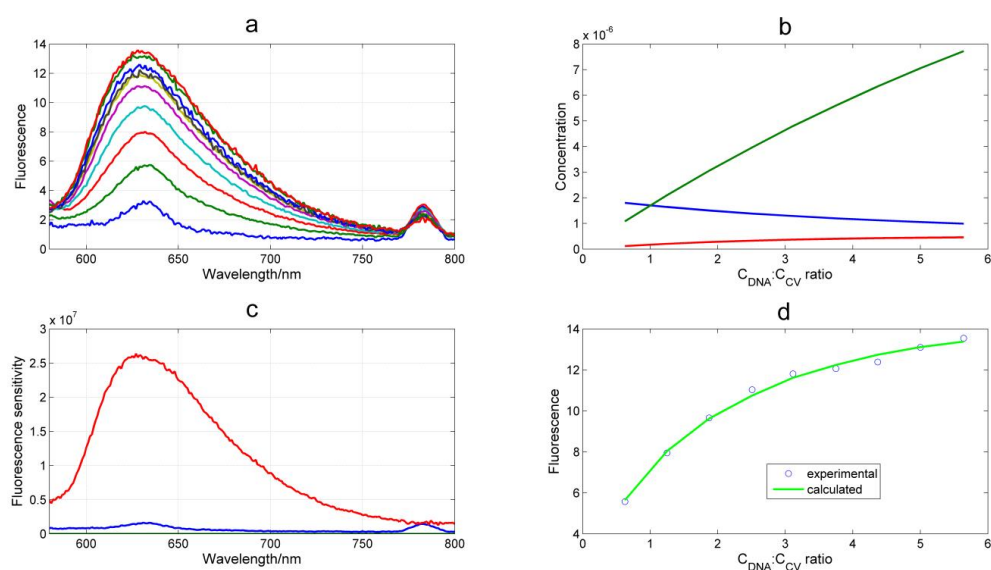
S9. ESI-MS spectra of ckit21T21, ckit21T18, and SMG03. M and D refer to monomer and dimer, respectively. Y-axis is in arbitrary units. Experiments carried out at 25°C, 150 mM NH₄AcO, pH 7.
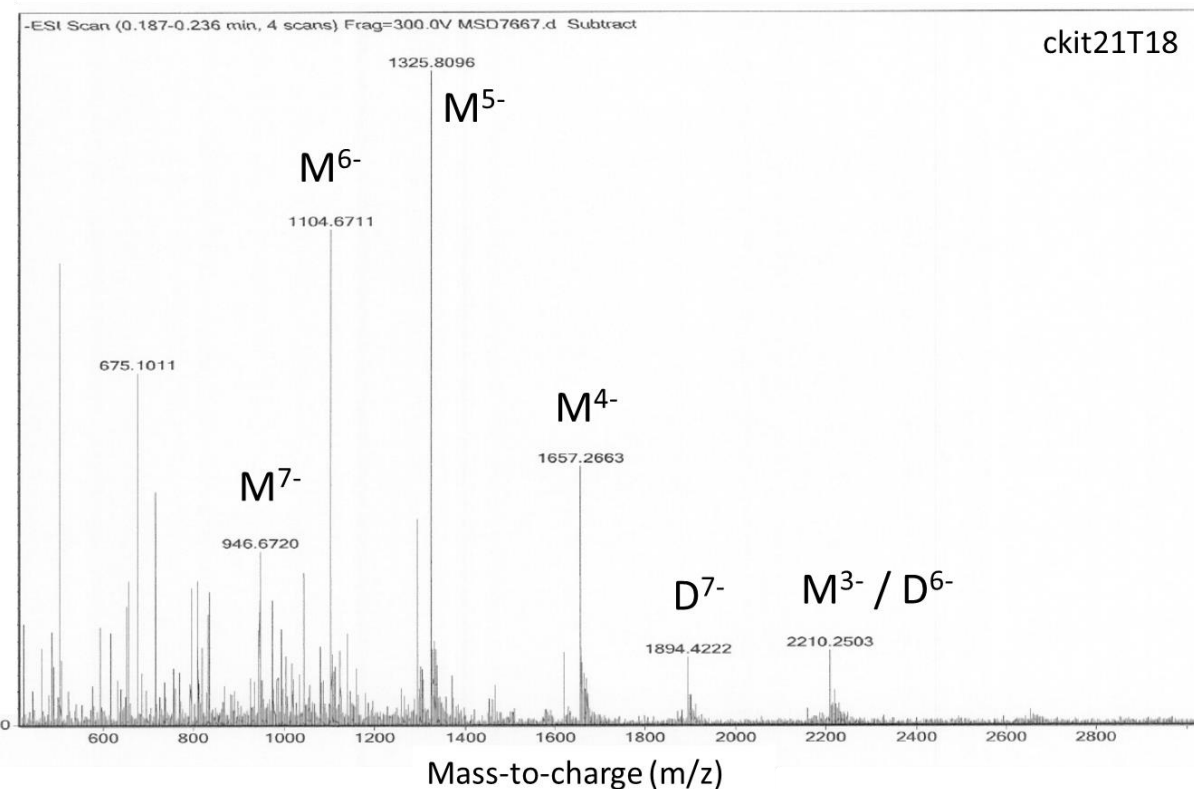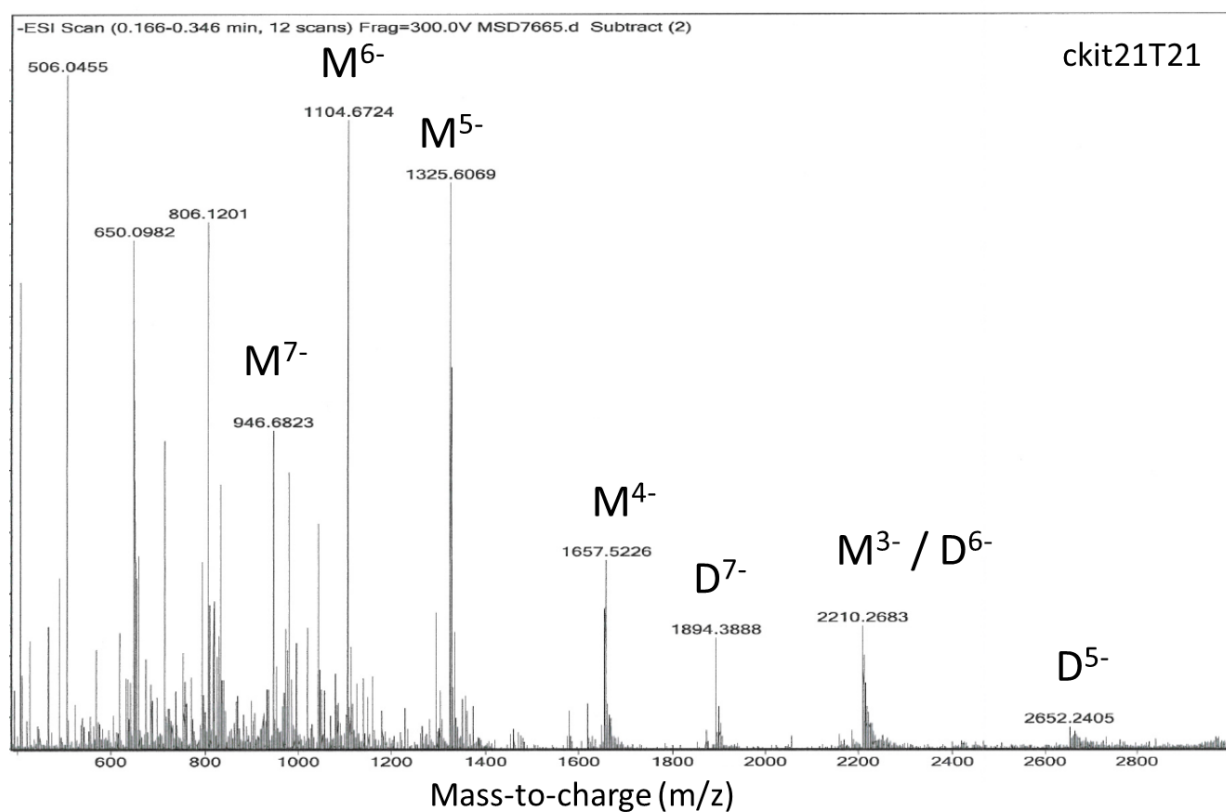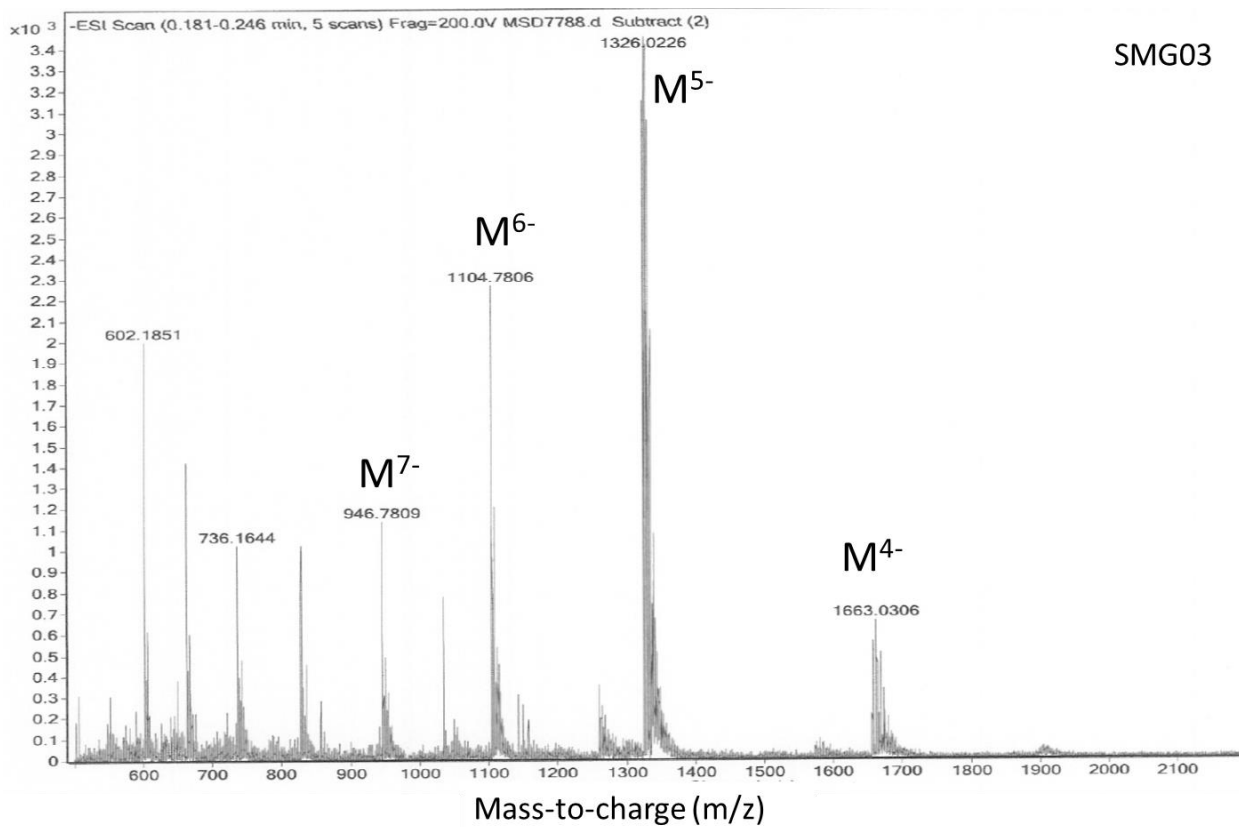
S10. Molecular fluorescence-monitored titration of CV with ckit21T21. (a) Experimental spectra (matrix **D**). (b) Calculated distribution diagram (matrix **C**) for each one of the three-species considered. This diagram has been calculated from the initial concentrations, proposed stoichiometry and fitted binding constants by using the Equispec program. (c) The corresponding pure spectra (matrix **S**). (d) Fit at 630 nm. In panels b and c, blue, green and red lines correspond to free CV, free DNA, and 1:1 interaction complex, respectively. In panel d, blue symbols and green lines correspond to experimental and fitted fluorescence values, respectively. Experiments were carried out at 25°C, 50 mM KCl and 20 mM phosphate buffer, pH 7.

S11. Melting curves of DNAs in absence and presence of CV. Plots of melting curves of ckit21T21 (a), ckit21T21 (b) and SMG03 (c) in absence (blue) and in the presence of CV (1:5 DNA:CV ratio, red). Experiments were carried out in 50 mM KCl and 20 mM phosphate buffer, pH 7. In panels a and b, ellipticity was measured at 263 nm, whereas in panel c, it was measured at 292 nm.
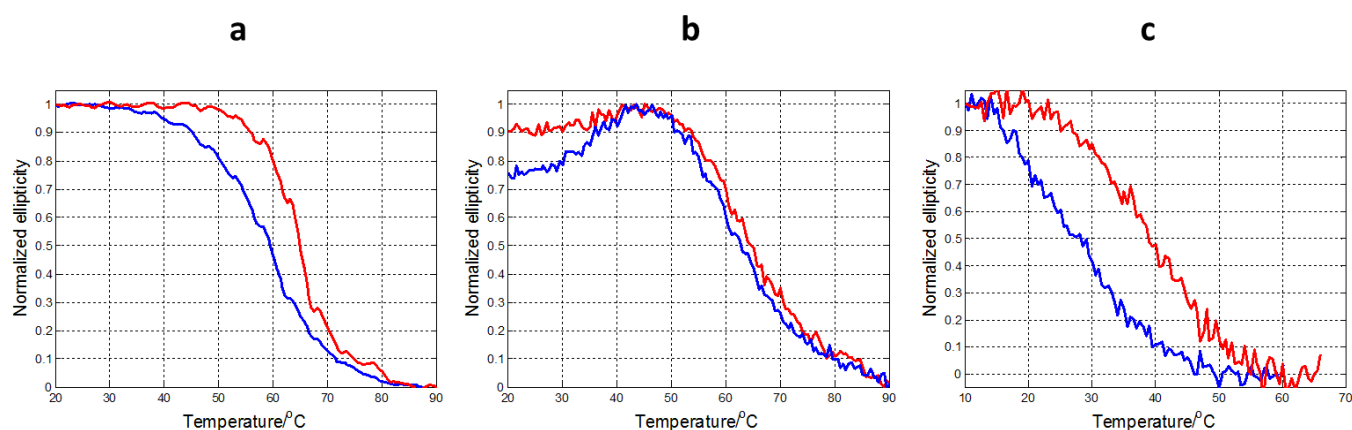
Figure S12. **ESI-MS spectra of ckit21T21, ckit21T18, and SMG03 in presence of CV (1:4 DNA:ligand ratio).** M and D refer to monomer and dimer, respectively. Y-axis is in arbitrary units. Experiments carried out at 25°C, 150 mM NH₄AcO, pH 7.