Influence of pH and a porphyrin ligand on the stability of a G-quadruplex structure within a duplex segment near the promoter region of the SMARCA4 gene

Alba Navarro[1], Sanae Benabou[1,3], Ramon Eritja[2], Raimundo Gargallo[1]*

[1] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Spain

[2] Institute for Advanced Chemistry of Catalonia (IQAC), CSIC, Networking Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Barcelona, Spain

[3] Université de Bordeaux, CNRS, Inserm, Laboratoire Acides Nucléiques: Régulations Naturelle et Artificielle (ARNA, U1212, UMR5320), IECB, 2 rue Robert Escarpit, 33600, Pessac, France

* Corresponding author. Correspondence should be sent to Raimundo Gargallo, Dept. of Chemical Engineering and Analytical Chemistry, Marti i Franquès 1-11, E-08028 Barcelona, Spain. Phone: +34 934039116, e-mail: raimon_gargallo@ub.edu

1. Abstract

In a previous work, the formation of G-quadruplex structures in a 44-nucleotides long sequence found near the promoter region of the SMARCA4 gene was reported. The central 25 nucleotides were able to fold into an antiparallel G-quadruplex structure, the stability of which was pH-dependent. In the present work, the effect of the presence of lateral nucleotides and the complementary cytosine-rich strand on the stability of this G-quadruplex has been characterized. Moreover, the role of the model ligand TMPyP4 has been studied. Spectroscopic and separation techniques, as well as multivariate data analysis methods, have been used with these purposes. The results have shown that stability of the G-quadruplex as a function of pH or temperature is greatly reduced in the presence of the lateral nucleotides. The influence of the complementary strand does not prevent the formation of the G-quadruplex. Moreover, attempts to modulate the equilibria by an external ligand led us to determine the influence of the TMPyP4 porphyrin on these complex equilibria. This study could eventually help to understand the regulation of SMARCA4 expression.

Keywords: G-quadruplex; i-motif; SMARCA4; multivariate analysis; TMPyP4

2. Introduction

Guanine and cytosine-rich nucleic acid sequences may fold into characteristic structures known as G-quadruplex and i-motif, respectively. For many years, these structures were only matter of *in vitro* studies with the objective of deciphering the rules behind their stability in front of pH, temperature, ionic strength, cations or ligands. However, as their existence *in vivo* was shown not much time ago [1][2][3], it has been suggested that these structures may have an important role in some biological processes, like the regulation of gene expression [4]. In this sense, the interaction with proteins has been also reported [5][6][7][8].

Whereas the stability of G-quadruplex structures has been shown to depend strongly on the nature and concentration of appropriate cations, like potassium, the stability of i-motif structures is largely pH-dependent. However, recent works have described the existence of G-quadruplex-forming sequences that show also pH-dependent polymorphism [9][10]. *In vivo*, the stability of G-quadruplex or i-motif structures is also strongly dependent on the presence of the complementary strand. Hence, most of the bibliography dealing with this point shows that, at neutral pH and a temperature within 25 to 37°C, the formation of the duplex DNA stabilized by Watson-Crick bonds hinders the formation of both, G-quadruplex and i-motif structures.

In this context, we focused previously our attention in two 44-nucleotides long guanine (SMG01) and cytosine-rich (SMC01) sequences located near the promoter region of the SMARCA4 gen (Figure 1a). This gene is involved in some diseases, like small cell carcinoma of hypercalcemic type (SCOOHT), a type of ovarian cancer [11]. The study of SMG01 revealed the formation of G-quadruplex structures by the central 21-nucleotide long sequence (coded as SMG03) [9]. Surprisingly, it was found that the topology of these G-quadruplex structures was pH-dependent, with an apparent pH-transition midpoint around 7.4 (150 mM KCl, 25°C). Hence, at pH values lower than 7.4, the SMG03 sequence folds into a homogenous antiparallel G-quadruplex structure, whereas a mixture of structures was observed at pH values higher than 7.4. Based on the study of mutated sequences it was proposed that two cytosine bases were involved in a pH-dependent equilibrium that dramatically reduced the conformational heterogeneity of the SMG03 sequence and strongly increased the thermal stability of the resulting structure (Figure 1b). In the context of the whole SMG01 sequence, it was observed that the antiparallel G-quadruplex structure was still formed, presumably by the central SMG03 sequence, despite the presence of lateral nucleotides.

Concomitantly to the study of SMG03 and SMG01 sequences, the study of the SMC01 sequence, which is complementary to SMG01, revealed the formation of i-motif structures at acidic pH values [12]. Among all the possible structures, that formed by the 21-nucleotides long central sequence (SMC03) was observed to be the most stable in terms of pH or temperature, even though this stability (in terms of T_m values) was clearly lower than that observed for other i-motif formed by other cytosine-rich sequences.

In the present work we focus our study on the pH and temperature-driven competition between G-quadruplex and i-motif structures formed within SMG01 and SMC01 sequences, and the intermolecular Watson-Crick duplex. Spectroscopic and separation methods, as well as multivariate data analysis techniques, have been used with this purpose detecting the formation of an antiparallel G-quadruplex structure at pH 4-6 in both single-stranded SMG01

and duplex SMG01:SMC01. Moreover, the interaction of these sequences with a ligand, TMPyP4 (Figure 1c), is studied showing the interaction of this ligand with SMG01 sequence. In addition, this ligand also interacts with duplex SMG01:SMC01 but in this case TMPyP4 prefers to stabilize the duplex structure than the quadruplex.

3. Material and methods

3.1. Reagents

DNA oligonucleotides 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 Glen-Pak Purification Cartridge (Glen Research, VA, USA). The integrity of DNA sequences was checked by means of Mass Spectrometry. DNA strand concentration was determined by absorbance measurements (260 nm) at 90ºC using the extinction coefficients calculated using the nearest-neighbor method as implemented on the OligoCalc webpage [13]. Before any experiment, DNA solutions were first heated to 95ºC for 10 minutes and then allowed to reach room temperature overnight. KCl, $NaCH_3COO$, CH_3COOH, KH_2PO_4, K_2HPO_4, HCl and LiOH were purchased from Panreac Química S.L.U. (Castellar del Vallès, Spain). MilliQ® water was used in all experiments. The mesotetrakis-(N-methyl-4-pyridyl)-porphyrin (TMPyP4) was purchased from Porphyrin Systems (Lübeck, Germany).

3.2. Instruments and procedures

Absorbance spectra were recorded on an Agilent 8453 diode array spectrophotometer. The temperature was controlled by means of an 89090A Agilent Peltier device. Hellma quartz cells (10 mm path length, and 1500 or 3000 µl volume) were used. Circular dichroism (CD) spectra were recorded on a Jasco J-810 spectropolarimeter equipped with a Peltier accessory for temperature control. This instrument enabled the simultaneous acquisition of both, CD and molecular absorption spectra. Hellma quartz cells (10 mm path length, 3000 µl volume) were used. Molar ellipticity ($deg \cdot cm^2 \cdot mol^{-1}$) has been calculated according to: $[\Theta] = \Theta/C \cdot l$, where Θ is the measured ellipticity (mdeg), C is the analytical concentration ($mol \cdot L^{-1}$), and l is the optical path (cm).

Acid-base titrations were monitored by using the Jasco J-810 instrument. The experimental conditions were 20ºC and 150 mM KCl. Titrations were carried out by adjusting the pH of solutions containing the oligonucleotides by addition of 1 M LiOH or HCl solutions. pH was measured using an Orion SA 720 pH/ISE meter and a micro-combination pH electrode (Thermo Scientific, USA). Absorbance and CD spectra were recorded in a pH stepwise fashion.

Melting experiments were monitored either using the Agilent-8453 or the Jasco J-810 instruments. The DNA solution was transferred to a covered 10-mm-path-length cell and spectra were recorded at 2ºC intervals with a hold time of 3 minutes at each temperature, which yielded an average heating rate of approximately $0.6ºC \cdot min^{-1}$. Buffer solutions were 20 mM phosphate (pH 7.1 and 6.0) or acetate (pH 5.0 and 4.2) and 150 mM KCl.

For SE-HPLC, the chromatographic system consisted of a Waters 2695 HPLC instrument equipped with a quaternary pump, a degasser, an autosampler, a photodiode-array detector with a 13-µL flow cell, and software for data acquisition and analysis. The chromatographic column used for separation at room temperature was PSS Suprema

Analytical Lineal S 100-100.000 Da (PSS Polymer Standards Service GmbH, Mainz, Germany). The composition of the mobile phase was 300 mM KCl and 20 mM phosphate or acetate. The flow was set to 0.8 mL·min^{-1}. The injection volume was 15 µL. T_{15}, T_{20}, T_{25}, T_{30} and T_{45} sequences were used as standards to construct the plot of logarithm of the retention time (t_R) *vs.* molecular weight. Some 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%. SE-HPLC chromatograms 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 1, where each raw chromatogram is divided by the corresponding Euclidean norm. The variable d_i indicates the value of absorbance at time i, whereas n is here the total number of points at which absorbance was measured in each chromatogram.

$$Normalized\ chromatogram = \frac{raw\ chromatogram}{\sqrt{\sum_1^n d_i^2}} \qquad (1)$$

3.3. Data analysis

Two approaches (univariate and multivariate) were used to study the thermally induced unfolding, the acid-base equilibria or the interaction with TMPyP4 ligand. In the first case, the considered process was monitored at just one wavelength and the later analysis was limited to the interpretation of the resulting ellipticity or absorbance trace. In the second, case, the process was monitored in a whole range of wavelengths. In addition, some experiments were monitored simultaneously by using CD and molecular absorption spectroscopies. In all these cases, multivariate data analysis methods were then required to interpret the recorded data.

3.3.1. Univariate analysis

In this work, univariate analysis was used to characterize melting experiments. As the sequences under study could produce several structures in equilibrium, the analysis was restricted to the determination of the melting temperature (T_m), the temperature at which half of the molecules are in ordered form, whereas the other half are unfolded. This was done by calculating the first derivative of the melting trace and by fitting a sigmoidal curve. The respective calculations were done by using the *savgol.m* function or the *cftool* toolbox, as implemented in Matlab®.

3.3.2. Multivariate analysis

In this case, spectra recorded during the considered process were arranged in a table or data matrix **D**, with *m* rows (spectra recorded) and *n* columns (wavelengths measured). The goal of data analysis was the calculation of distribution diagrams and pure (individual) spectra for all *ns* spectroscopically active species considered throughout the process. The distribution diagram provides information about the stoichiometry and stability of the species considered. In addition, the shape and intensity of the pure spectra may provide qualitative information about the structure of the species. With this goal in mind, data matrix **D** was decomposed according to Beer-Lambert-Bouer's law in matrix form:

$$\mathbf{D} = \mathbf{C}\,\mathbf{S} + \mathbf{E} \qquad (2)$$

C is the matrix (*m* x *ns*) containing the distribution diagram, **S** is the matrix (*ns* x *n*) containing the pure spectra, and **E** is the matrix of data (*m* x *n*) not explained by the proposed decomposition.

The mathematical decomposition of **D** into matrices **C**, **S**, and **E** may be conducted in two different ways, depending on whether a chemical model is initially proposed (hard-modeling approach [14]) or not (soft-modeling approach [15][16]). For hard-modeling approaches, the proposed model depends on the nature of the process under study. In this work, hard-modeling has been applied to the analysis of data recorded along acid-base and ligand titrations.

For acid-base experiments the model includes a set of chemical equations describing the formation of the different acid-base species from the neutral species, together with approximate values for the stability constants, such as the following:

$$DNA + nH^+ \leftrightarrow DNA{\cdot}H_n \qquad Beta_{1n} = [DNA{\cdot}H_n] / [DNA]\,[H^+]^n \qquad (3)$$

In this equation, the parameter *n* is related to the Hill coefficient and describes qualitatively the cooperativity of the equilibrium. Values of *n* greater than 1 indicate the existence of a cooperative process.

For ligand titrations the model includes a set of chemical equations describing the formation of the different species from DNA and TMPyP4, together with approximate values for the stability constants, such as the following:

$$DNA + q\ ligand \leftrightarrow DNA{\cdot}ligand_q \qquad Beta_{1p} = [DNA{\cdot}ligand_q] / [DNA]\,[ligand]^q \qquad (4)$$

For any of the models considered, a mathematical calculation is initiated to obtain the matrices **C** and **S** that, according to equation 2, better adjust the data in **D** within the experimental error. In other words, the proposed values for the equilibrium constants in equations 3 or 4 and the shape of the pure spectra in **S** are refined to explain satisfactorily data in **D**, whereas residuals in **E** are minimized. In this study, hard-modeling analysis of acid-base experiments and ligand titrations used the EQUISPEC program [14].

In some cases, the studied process was monitored by CD and molecular absorption spectroscopies because the instrument enabled the simultaneous acquisition of both spectra. Hence, the matrix **D** in Equation 2 was the result of merging data matrices $\mathbf{D_{CD}}$ and $\mathbf{D_{abs}}$. As a result, the calculated matrix **S** according to models 3 or 4 contained the pure CD and molecular absorption spectra for each one of the proposed species. This simultaneous analysis has been shown to reduce the mathematical ambiguities related with multivariate analysis.

The analysis of multivariate data recorded along melting experiments was done by using a soft-modeling approach [17]. This treatment allowed the verification of the two-state assumption for a thermally-induced unfolding without the requirement of a complex model of species.

4. Results and discussion

4.1. Study of SMG01 polymorphism by means of spectroscopic techniques

First, the potential formation of folded structures by the SMG01 sequence was checked by means of molecular absorption melting experiments. At pH 7.1, melting experiments of SMG01 showed high hyperchromism at 260 nm,

which could be related with the presence of initial folded structures showing Watson-Crick base pairs, whereas no hypochromicity at 295 nm was observed, which could be related to the unfolding of G-quadruplex or i-motif structures (Figure 2a). The Thermal Difference Spectrum (TDS) confirmed the absence of any signature at 295 nm [18]. From the trace at 260 nm, a T_m value equal to 55.1°C was determined using a two-state approach (Table 1). Analysis of first derivatives did not show any trend that could be related to the presence of a folding involving more than two states (Figure S1).

The CD-monitored melting of SMG01 at pH 7.1 revealed unfolding of the initial structure(s) when the temperature was increased from 20 to 90°C. The whole set of spectra was analyzed by means of multivariate data analysis methods to check the validity of the considered two-state process (Figure S1). As in the case of molecular absorption, it was observed that the unfolding was well explained with only two components. However, the well-defined shape and the relatively high intensity of the CD spectrum at 90°C suggested the presence of high-order structures not unfolded yet (see below) [19].

At pH 6.0, hypochromicity at 295 nm was observed, which could be related to the unfolding of a G-quadruplex structure (Figure 2b). The potential formation of stable i-motif structures can be ruled out because of the absence of enough cytosine bases within SMG01 sequence. The hypochromicity at 295 nm was enhanced at pH 5 and reduced at pH 4, whereas hyperchromicity at 260 nm was reduced from pH 7 to pH 4 (Table 1).

To obtain additional information about the influence of pH on the conformational equilibria of SMG01 an acid-base titration was done. Hence, SMG01 was titrated from pH 12 to pH 3, approximately, and the evolution of the acid-base species was monitored by means of CD and molecular absorption spectroscopies. The experiment consisted on the following procedure. First, the sequence was placed into an optical cell. Then, both CD and molecular absorption spectra, as well as pH, were measured. Stepwise additions of LiOH 1M allowed the measurement of spectra and pH from the initial pH (around 7) to pH 12, approximately. Then, successive additions of HCl 1 M from pH 12 to pH 3 allowed the measurement of the CD spectra in this pH range. No hysteresis was observed, which pointed out to fast equilibria. Figure 3a and b, respectively, show representative CD and molecular absorption spectra measured along the titration, whereas the complete data set is available as Figure S3.

The evolution of CD spectra along the titration reflects the presence of several equilibria in the studied pH range. As described previously [9], the interpretation of data of this kind by just plotting the variation of the signal (ellipticity or absorbance, see insets in Figure 3) at one single wavelength may result in a poor approach because minor contributions could be lost. Hence, the analysis of the whole data set (i.e., multivariate analysis) by means of appropriate mathematical methods may result a much better approach to solve it into the contributions of all the species present. In this case, the whole set of CD and molecular absorption spectra were simultaneously analyzed.

The multivariate analysis revealed that a model involving four spectroscopically active acid-base species (*ns=4*) explained well the whole data set. To check the validity of the proposed model of four species, an alternative model of only three acid-base species was tested (Figure S3). From the shape of the experimental *vs.* fitted signal, it was clear that the model of four species was appropriate to explain the experimental data.

The interpretation of the four species is as follows (Figure 3c-e). The species depicted in orange corresponds to the species present at pH 7 that, according to the resolved CD spectrum, is probably a hairpin. The existence of minor components cannot be ruled out at this point. Upon addition of LiOH, a new species (depicted in blue) appears. The pH-transition midpoint ($pH_{1/2}$) (11.8±0.1, Table 2) corresponds to the deprotonation of guanine and thymine bases. The pK_a values for free bases are around 10 [20] but this value may be shifted to higher values when bases are stabilized by hydrogen bonds in framework of a hairpin structure. Both, species yellow and magenta show CD spectra whose spectral features (strong positive ellipticity around 290 nm) match with those of antiparallel structures. Hence, it may be deduced that antiparallel G-quadruplex in SMG01 is formed at pH below 6, approximately. The determined $pH_{1/2}$ values (4.7 and 3.8, Table 2) are like those corresponding to the protonation of free cytosine and adenine bases (around 3.5- 4.5). For all three transitions, the n values in Equation 3 were equal to 1 (i.e., low cooperativity). This fact denotes smooth structural changes concomitant to the acid-base equilibria between species.

4.2. Study of SMG01 polymorphism by means of SE-HPLC

Size-Exclusion Chromatography (SE-HPLC) may be a useful tool to study the conformational complexity of nucleic acids [21][19]. In this work, SE-HPLC has been applied initially to monitor conformational changes in SMG01 with temperature and pH (Figure S4a).

At 20°C, the chromatographic profile showed the presence of two main bands centered at 9.86 and 10.45 minutes, respectively. Additionally, a shoulder may be observed around 9.3 minutes. According to the calibration plot using a set of Tx standards, the unfolded SMG01 sequence would elute at 10.22 minutes. Consequently, the band at 10.45 minutes would correspond to a folded monomer species, probably a hairpin, whereas the band at 9.86 minutes would correspond to a dimer (Figure S5). The presence of a shoulder around 9.2 minutes would be related to the formation of higher order aggregates. Upon heating, the complexity of the chromatogram is reduced greatly, and a major band around 10.30 minutes appears that has been assigned to the unfolded strand. The chromatogram recorded at 80°C still shows a shoulder around 9.7 minutes, which denotes the existence of multimeric species, probably dimers, at this temperature. This agrees with the previous observation of CD spectra with well-defined shape and relatively high intensity at high temperature (Figure S1). Not only temperature but also pH produced clear changes in the conformational landscape of SMG01. Hence, lowering the pH to 6.0 produced the enhancement of the bands related to multimers (Figure S4b). At pH 4.0, the band at 10.4 minutes, which is related to the folded monomer, seems to be the major component but the presence of strong adsorption of the analyte on the column prevents drawing further conclusions.

4.3. Study of the 1:1 SMG01:SMC01 mixture

From the results obtained in the acid-base titration of SMG01 it seems that an antiparallel G-quadruplex is also formed within the SMG01 sequence in the pH range 4-6, approximately. However, the presence of the complementary cytosine-rich strand could modulate the formation of this structure. To study the influence of this strand (SMC01), several experiments were carried out. Previously, the influence of pH and temperature on the folded structures formed by SMC01 were studied [12]. Briefly, it was shown that this sequence may fold into two similar i-motif structures at pH lower than 6.5, being their maximal formation between pH 4 and 5.5, approximately (Figure S6).

4.3.1. Acid-base titrations of 1:1 SMG01:SMC01 mixture

As in the case of SMG01, a 1:1 mixture of SMG01 and SMC01 was titrated from pH 12 to pH 3, approximately, and the evolution of the acid-base species was monitored by means of molecular absorption and CD spectroscopies (Figure 4 and Figures S7-S8). Contrarily to the acid-base titration of SMG01 hysteresis was observed, as spectra recorded when pH was decreased from 6 to 3 were not fully recovered when pH was raised in the opposite direction. This fact pointed out to relatively slow folding equilibria in the time scale of the measurements (10-15 minutes).

At pH 7, the CD spectrum shows the characteristics of a B-DNA duplex. At pH values below 6, the magnitude of the CD signal at 295 nm increases gradually, which denotes the formation of antiparallel structures. The variation of ellipticity at 265 nm (Figure 4b, symbols) clearly denotes the existence of several acid-base species. The further analysis of the whole data set (CD and molecular absorption data) revealed that a model involving four spectroscopically active acid-base species explained it well (Figure 4b, lines).

The interpretation of the four species is as follows. The species depicted in orange corresponds to the duplex stabilized by Watson-Crick base pairs. Again, the existence of minor components cannot be dismissed at this point. Upon addition of LiOH, a new species (depicted in blue) appears, which has been assigned to the mixture of isolated SMG01 and SMC01 strands. The pH-transition midpoint ($pH_{1/2}$) (12.1±0.1, Table 2) corresponds to the deprotonation of guanine and thymine bases. Contrarily to that observed for this transition in the case of SMG01, the value of n was equal to 2, which denoted a certain degree of cooperativity in this transition. This fact may be easily explained as a result of the unfolding of the 44-nucleotides long duplex. On the other hand, species colored in yellow and magenta show CD spectra whose spectral features (strong positive ellipticity around 290 nm) match with those of antiparallel G-quadruplex and/or i-motif. Hence, it may be hypothesized that either one or both structures, antiparallel G-quadruplex in SMG01 and i-motif in SMC01, are formed at pH below 6, approximately. The determined $pH_{1/2}$ values (4.8 and 4.0) are like those corresponding to the protonation of free cytosine and adenine bases (around 3.5- 4.5). The shape of the CD spectrum for the more acidic species (magenta) should be considered critically as the latest spectrum measured at pH 2.8 still shows a mixture of species. In this condition, the pure spectrum of this species cannot be determined accurately. Nevertheless, this spectrum is clearly different from that measured at pH 12, where the strands are completely unfolded. Therefore, it was deduced that species magenta still showed a certain degree of antiparallel and parallel folding within it.

4.3.2. CD-monitored melting experiments

In a previous work, the thermally-induced unfolding of SMC01 was studied in a short pH range (4.7 – 5.5) [12]. In the present work, this range was enlarged from 7 to 4 (Table 1 and Figure S9) to enable the comparison with results for SMG01 and for the 1:1 SMG01:SMC01 mixture. At pH 7.1, any well-defined transition that could be related to the unfolding of an ordered structure was observed. On the contrary, clear transitions accompanied by hypochromicity at 295 nm were observed at pH values higher than 4.2, which were assigned to the unfolding of i-motif structures. Below the pK_a of cytosine (4.5, approximately [reference]), unfolding was accompanied by hyperchromicity at 295 nm, as previously described by Mergny *et al.* [22]. Finally, the maximal stability of the folded structure in terms of T_m values was also reached near pH 4.5, which is the typical dependence of T_m with pH already described for i-motif structures.

Melting experiments of the mixture at different pH values were then carried out. Figure 5 (and Figure S10) includes CD spectra recorded at pH 7.1 and 4.2, as well as the normalized absorbance traces at 260 and 295 nm.

Melting at pH 7.1 (T_m = 79ºC) was related to the unfolding of the duplex structure stabilized by Watson-Crick bonds. Upon acidification, the CD signal at 285 nm increased slightly, being its maximal intensity at pH 4.2. From the absorbance melting curves at 260 and 295 nm, T_m values were determined (Table 1). The values at pH 7.1 and 6.0 were similar at both wavelengths, which indicated that the major structures within this pH range were similar, in agreement with the calculated distribution diagram (Figure 3c). However, at pH 5.0, the T_m value was reduced to 67ºC, when monitored at 260 nm, whereas the value was 47ºC when monitored at 295 nm. This indicated the presence of antiparallel component, also in agreement with the calculated distribution diagram. Finally, at pH 4.2, the trace at 260 nm was almost lost, whereas a clear transition was observed at 280 nm, which was indicative of antiparallel structures, with a T_m value equal to 47ºC.

4.3.3. SE-HPLC measurements

Finally, SE-HPLC was used to characterize the molecularity of the species formed by the mixture along the pH range 4.2-7.1. First, the molecularity of the complementary SMC01 strand was studied (Figure S4c). At pH 7.1, the chromatographic profile of SMC01 showed a single band centered at 10.32 minutes, which is similar to the retention time for the unfolded strand according to the calibration plot of T_x standards (10.26 minutes). This rather simple chromatogram agrees with the lack of unfolding transition described above. Moreover, it contrasts with the complex chromatogram recorded for SMG01 at the same pH. At lower pH values, the chromatographic peak was shifted to greater retention times, reflecting the folding of the sequence. Interestingly, a small contribution of a dimeric species was also observed at pH 4.2 and pH 5.0.

At pH 7.1, the chromatographic profile for the 1:1 SMG01:SMC01 mixture showed a single band centered at 9.7 minutes (Figure S4d). As expected from melting studies, the correlation of this retention time with the calibration plot of T_x standards confirmed that this species was the duplex formed by both sequences. Interestingly, any shift of the retention time was observed when lowering the pH from 7.1 to 4.2, which proved that the species remained mostly as a duplex within this pH range. Therefore, the antiparallel structures detected in acid-base and melting studies of the mixture are formed in the framework of the duplex structure, without disruption of this one into separated strands.

4.4. The cationic porphyrin TMPyP4 stabilizes the G-quadruplex structure within SMG01

The stability of G-quadruplex structures may be tuned by interaction with selective ligands. Among these, porphyrins have been shown to bind to and stabilize different types of G-quadruplexes and, in some cases, to facilitate G-quadruplex formation [23]. The cationic porphyrin mesotetrakis-(N-methyl-4-pyridyl)-porphyrin (TMPyP4) is one of the most extensively studied drugs and it is considered as a model ligand for these studies.

Figure 6a shows the spectral variations observed upon addition of TMPyP4 to the folded structure formed by SMG01 at pH 7. The intensity of negative and positive bands of the initial CD spectrum of SMG01 was reduced along the titration. Concomitantly, a shoulder appeared around 288 nm, which could reflect the stabilization of the antiparallel G-quadruplex formed by the central guanine-rich sequence. In addition, it should be noted the appearance of an

induced CD (ICD) band around 445 nm. This ICD signal has been related to the intercalation or stacking of the ligand on DNA bases [24].

Multivariate analysis according to equation 4 allowed the determination of the number of species present in the system, their stoichiometries, as well as the associated equilibrium constants (Figure 6b and 6c, and Table 3). Apart from the two single species (SMG01 and TMPyP4), two different interaction complexes were proposed with stoichiometries 1:1 and 1:4 (DNA:ligand). Whereas the pure spectra for both interaction complexes showed a shoulder around 290 nm, only the spectrum calculated for the complex 1:4 showed a greater extent of the ICD band at 445 nm.

For comparison, Table 3 also includes the determined stoichiometries and binding constants for the central sequence SMG03. Both SMG01 and SMG03 form 1:1 complexes. However, whereas SMG01 forms a second complex where DNA binds up to four ligand molecules, the second complex by SMG03 only is able to accommodate an additional ligand. This may be because SMG01 is a longer sequence than SMG03, which shows more complex polymorphism, as denoted by SE-HPLC analysis (Figure S4 and reference [9]). The calculated stoichiometries and binding constants for the interaction of TMPyP4 with the antiparallel G-quadruplex formed by SMG03 agreed with previously reported values for antiparallel telomeric G-quadruplex [25].

Finally, the interaction with the 1:1 SMG01:SMC01 mixture was studied (Figure S11). CD spectra measured along the titration also showed the appearance of a strong ICD band at 445 nm that was related to intercalation between base pairs, rather than stacking on the terminal bases. At the experimental conditions tested, it was observed that the duplex structure was able to bind up to 8 ligand molecules. This observation agrees with previous reports where up to 7 molecules were found to bind duplex structure [25].

For SMG01 sequence, melting experiments of the 1:6 DNA:ligand mixture showed a clear variation of T_m values associated to unfolding both at 260 and 290 nm. At 260 nm, T_m increased slightly from 56.1 (SMG01) to 60.7°C (in presence of TMPyP4) at pH 7.1, when monitored at 260 nm (Figure 7a). This increase suggests a certain stabilization of the folded hairpin in presence of the drug. However, when monitored at 290 nm, T_m in presence of drug was 72.3°C, which suggested a strong stabilization of the antiparallel structure. This behavior is similar to that observed for the melting of 1:6 SMG03:TMPyP4 mixture, the T_m of which (65.0°C) was clearly higher than that of the G-quadruplex (37°C, Table 1).

In the mixture, the behavior is reversed (Figure 7b). Hence, the high value of T_m at 260 nm (>90°C) indicated a strong stabilization of the duplex structure due to intercalation. On the contrary, the antiparallel component showed a lower T_m value (51.5°C).

5. Discussion

Several G-quadruplex-forming sequences have been identified in G-rich eukaryotic telomeres and recently in non-telomeric genomic DNA such as in nuclease-hypersensitive promoter regions [2]. The natural role and biological validation of these structures is starting to be examined, and there is interest in them as targets for therapeutics.

In promoters, G-quadruplex is considered to act as a molecular switcher that turns transcription process on/off in association with transcriptional proteins that may facilitate their folding [26]. This alternative way of regulation involving G-quadruplexes may have relation with oncogenic conversion, where a critical protein might be found with a G-quadruplex in the core or near promoter of oncogene, for example *c-kit* [27], *c-myc* and *KRAS* [28], *hTERT* [28], *bcl-2* [29], *PDGF-A* [28] or *VEGF-A* [30].

In this context, the present work had three main objectives. The first one was to determine the effect of the flanking nucleotides on the G-quadruplex structure formed by SMG03 sequence constituting the 44mer sequence that has been named as SMG01. From the results obtained, it was clear that this structure was less stable in front of pH than in the case of SMG03 (Tables 1 and 2). Hence, the $pH_{10\%}$, the pH at which 10% of the antiparallel structure was formed, was reduced from 8.2 in SMG03 to 5.7 in SMG01. Concomitantly, the thermal stability of the folded structure was also decreased at pH 7 and 6. Despite the fact that the stability of the G-quadruplex in SMG03 is strongly dependent on pH, the obtained results agree with previous observations on G-quadruplex structures whose stability is not as strongly pH-dependent. Hence, the destabilization of G-quadruplex structures with increasing flank length was already described [31]. Other effects have been reported. As example, Renciuk *et al*. proposed that increasing the number of flanking nucleotides may produce a shift in conformational equilibria towards an hybrid G-quadruplex [32]. Recently, Marchand *et al.* showed that flanking bases increase the K^+ binding cooperativity [33].

The second objective was to determine the effect of the complementary cytosine-rich strand on the stability of the G-quadruplex formed within SMG01 sequence. Despite the fact that this subject has been matter of interest of many studies previously [34][31][35][36][37], the present work provided a quantitative point of view that is usually not considered in such studies. At pH 7 and 37^{o}C, the duplex is the main structure. Only at pH values lower than 6, the antiparallel contributions appear, being their maximum formation around pH 4. It has been described that moderate acidic pH (around 6) could produce a destabilization of a duplex into G-quadruplex and i-motifs because of the driving force of the formation of this last structure [31][36][38]. Other authors, however, have proposed that the driving force behind dissociation of the duplex is the folding of the G-strand into the G-quadruplex [39][40]. The i-motif alone could not induce dissociation of the duplex even at pH 5.0, at which it is most stable. In this sense, It has been proposed that the formation of G-quadruplex and i-motif in the two complementary strands are mutually exclusive in a variety of DNA templates, which include human telomere and promoter fragments of hINS and hTERT genes [41]. In our case, the i-motif formed by SMC01 is not very stable (Table 1), and the disruption of the duplex was mainly due to the formation of the antiparallel G-quadruplex by the central SMG03 sequence.

Finally, the last objective was to evaluate the potential stabilizing effect of TMPyP4, which can be considered as a model ligand, on the G-quadruplex structure formed by SMG03 sequence at neutral pH. It should be taken into account, however, that its selectivity towards G-quadruplex structure is still a matter of debate [42][43]. It has been shown that the structure formed by SMG03 may bind up to two ligands whereas two more ligands may be accommodated within the longer SMG01 sequence. This uptake takes place in two steps, in agreement with previous studies of different sequences [44]. In both cases, the appearance of a strong ICD signal denoted the existence of intercalative and/or end-stacking interaction. The exact assignment is difficult, as the binding modes as well as the

binding stoichiometry in TMPyP4-G-quadruplex complexes continue to be controversial in recent years [45]. Hence, binding modes including intercalation between adjacent G- quartets [46], stacking onto external G-quartets [47], and external stacking on loop nucleotides without any direct interaction with G-quartets [48] have been discussed. In the framework of the duplex structure formed by SMG01:SMC01, addition of TMPyP4 hardly induces the formation of the G-quadruplex structure whereas it produced a strong stabilization of the duplex.

6. Conclusion

In this work, the stability of a particular guanine-rich sequences located near the promoter region of SMARCA4 gene was studied by different spectroscopic and separation techniques. The interaction with model ligand on the stability of the structures formed was examined. Broadly, G-quadruplex structures were identified as well as their solution equilibria. Interestingly, the stabilization of these G-quadruplex structures was increased in the presence of porphyrin TMPyP4. These finding can help in the understanding and modulation of SMARCA4 gene expression in a biological context.

7. Acknowledgments

Gerard Pastor (University of Barcelona, Spain) and Esther Miralles (*Centres Científics I Tecnològics*, University of Barcelona, Spain) are thanked for their help in carrying out some of the experiments. Funding from Spanish government (CTQ2017-84415-R and CTQ2015-66254-C2-2-P) and recognition from the Autonomous Catalan government (2017SGR114) are acknowledged.

8. References

[1] M. Zeraati, D.B. Langley, P. Schofield, A.L. Moye, R. Rouet, W.E. Hughes, T.M. Bryan, M.E. Dinger, D. Christ, I-motif DNA structures are formed in the nuclei of human cells, Nat. Chem. 10 (2018) 631–637. https://doi.org/10.1038/s41557-018-0046-3.

[2] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, Quantitative visualization of DNA G-quadruplex structures in human cells, Nat. Chem. 5 (2013) 182–186. https://doi.org/10.1038/nchem.1548.

[3] S. Dzatko, M. Krafcikova, R. Hänsel-Hertsch, T. Fessl, R. Fiala, T. Loja, D. Krafcik, J.L. Mergny, S. Foldynova-Trantirkova, L. Trantirek, Evaluation of the Stability of DNA i-Motifs in the Nuclei of Living Mammalian Cells, Angew. Chemie - Int. Ed. 57 (2018) 2165–2169. https://doi.org/10.1002/anie.201712284.

[4] D. Rhodes, H.J. Lipps, G-quadruplexes and their regulatory roles in biology., Nucleic Acids Res. 43 (2015) 8627–8637. https://doi.org/10.1093/nar/gkv862.

[5] H. Abou Assi, M. Garavís, C. González, M.J. Damha, i-Motif DNA: structural features and significance to cell biology, Nucleic Acids Res. (2018) 1–19. https://doi.org/10.1093/nar/gky735.

[6] J. Spiegel, S. Adhikari, S. Balasubramanian, The Structure and Function of DNA G-Quadruplexes, Trends Chem. (2019). https://doi.org/10.1016/J.TRECHM.2019.07.002.

[7] H. Masai, N. Kakusho, R. Fukatsu, Y. Ma, K. Iida, Y. Kanoh, K. Nagasawa, Molecular architecture of G-quadruplex structures generated on duplex Rif1-binding sequences, J. Biol. Chem. 293 (2018) 17033–17049. https://doi.org/10.1074/jbc.RA118.005240.

[8] G. Miglietta, S. Cogoi, E.B. Pedersen, L.E. Xodo, GC-elements controlling HRAS transcription form i-motif structures unfolded by heterogeneous ribonucleoprotein particle A1, Sci. Rep. 5 (2016) 18097. https://doi.org/10.1038/srep18097.

[9] S. Benabou, S. Mazzini, A. Aviñó, R. Eritja, R. Gargallo, A pH-dependent bolt involving cytosine bases located in the lateral loops of antiparallel G-quadruplex structures within the SMARCA4 gene promotor, Sci. Rep. 9 (2019) 15807. https://doi.org/10.1038/s41598-019-52311-5.

[10] J. Brčić, J. Plavec, NMR structure of a G-quadruplex formed by four d(G4C2) repeats: insights into structural polymorphism, Nucleic Acids Res. 46 (2018) 11605–11617. https://doi.org/10.1093/nar/gky886.

[11] L. Witkowski, J. Carrot-Zhang, S. Albrecht, S. Fahiminiya, N. Hamel, E. Tomiak, D. Grynspan, E. Saloustros, J. Nadaf, B. Rivera, C. Gilpin, E. Castellsagué, R. Silva-Smith, F. Plourde, M. Wu, A. Saskin, M. Arseneault, R.G. Karabakhtsian, E.A. Reilly, F.R. Ueland, A. Margiolaki, K. Pavlakis, S.M. Castellino, J. Lamovec, H.J. Mackay, L.M. Roth, T.M. Ulbright, T.A. Bender, V. Georgoulias, M. Longy, A. Berchuck, M. Tischkowitz, I. Nagel, R. Siebert, C.J.R. Stewart, J. Arseneau, W.G. McCluggage, B.A. Clarke, Y. Riazalhosseini, M. Hasselblatt, J. Majewski, W.D. Foulkes, Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type, Nat. Genet. 46 (2014) 438–443. https://doi.org/10.1038/ng.2931.

[12] S. Benabou, A. Aviñó, S. Lyonnais, C. González, R. Eritja, A. De Juan, R. Gargallo, i-motif structures in long cytosine-rich sequences found upstream of the promoter region of the SMARCA4 gene, Biochimie. 140 (2017) 20–33. https://doi.org/10.1016/j.biochi.2017.06.005.

[13] W.A. Kibbe, Oligo Calc: an online oligonucleotides properties calculator, Nucleic Acids Res. 35 (2007) W43–W46. https://doi.org/10.1093/nar/gkm234.

[14] R.M. Dyson, S. Kaderli, G. a Lawrance, M. Maeder, a D. Zunderbuhler, Second order global analysis: the evaluation of series of spectrophotometric titrations for improved determination of equilibrium constants, Anal. Chim. Acta. 353 (1997) 381–393. https://doi.org/Doi 10.1016/S0003-2670(97)87800-2.

[15] E. Casassas, R. Gargallo, A. Izquierdo-Ridorsa, R. Tauler, Application of a multivariate curve resolution procedure for the study of the acid-base and copper(II) complexation equilibria of polycytidylic acid, React. Funct. Polym. 27 (1995) 1–14. https://doi.org/10.1016/1381-5148(95)00025-B.

[16] R. Gargallo, R. Tauler, A. Izquierdo-Ridorsa, A. Diagonal, Acid–Base and Copper(II) Complexation Equilibria of Poly(inosinic)-Poly(cytidylic), Biopolymers. 42 (1997) 271–283. https://onlinelibrary.wiley.com/doi/10.1002/%28SICI%291097-0282%28199709%2942%3A3%3C271%3A%3AAID-BIP1%3E3.0.CO%3B2-J.

[17] J. Jaumot, N. Escaja, R. Gargallo, C. Gonzalez, E. Pedroso, R. Tauler, Multivariate curve resolution: a powerful tool for the analysis of conformational transitions in nucleic acids, Nucleic Acids Res. 30 (2002) 92e – 92. https://doi.org/10.1093/nar/gnf091.

[18] J.L. Mergny, J. Li, L. Lacroix, S. Amrane, J.B. Chaires, Thermal difference spectra: A specific signature for nucleic acid structures, Nucleic Acids Res. 33 (2005) 1–6. https://doi.org/10.1093/nar/gni134.

[19] S. Benito, A. Ferrer, S. Benabou, A. Aviñó, R. Eritja, R. Gargallo, Evaluation of the effect of polymorphism on G-quadruplex-ligand interaction by means of spectroscopic and chromatographic techniques, Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 196 (2018) 185–195. https://doi.org/10.1016/j.saa.2018.02.006.

[20] V.A. Bloomfield, D.M. Crothers, Ignacio Tinoco, Nucleic Acids. Structures, properties and functions, Sausalito, CA, USA, 1999.

[21] E. Largy, J.L. Mergny, Shape matters: Size-exclusion HPLC for the study of nucleic acid structural polymorphism, Nucleic Acids Res. 42 (2014) e149–e149. https://doi.org/10.1093/nar/gku751.

[22] J.L. Mergny, L. Lacroix, C. Hélène, X. Han, J.L. Leroy, Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an i-DNA Motif, J. Am. Chem. Soc. 117 (1995) 8887–8898. https://doi.org/10.1021/ja00140a001.

[23] S. Neidle, Quadruplex Nucleic Acids as Novel Therapeutic Targets, J. Med. Chem. 59 (2016) 5987–6011. https://doi.org/10.1021/acs.jmedchem.5b01835.

[24] H. Mita, T. Ohyama, Y. Tanaka, Y. Yamamoto, Formation of a complex of 5,10,15,20-tetrakis(N-methylpyridinium-4-yl)-21H,23H-porphyrin with G-quadruplex DNA., Biochemistry. 45 (2006) 6765–6772. https://doi.org/10.1021/bi052442z.

[25] A. Arora, S. Maiti, Effect of loop orientation on quadruplex-TMPyP4 interaction., J. Phys. Chem. B. 112 (2008) 8151–8159. https://doi.org/10.1021/jp711608y.

[26] C. Sutherland, Y. Cui, H. Mao, L.H. Hurley, A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III1, J. Am. Chem. Soc. 138 (2016) 14138–14151. https://doi.org/10.1021/jacs.6b09196.

[27] P. Bucek, J. Jaumot, A. Aviñó, R. Eritja, R. Gargallo, pH-modulated Watson-Crick duplex-quadruplex equilibria of guanine-rich and cytosine-rich DNA sequences 140 base pairs upstream of the c-kit transcription initiation site, Chem. - A Eur. J. 15 (2009) 12663–12671. https://doi.org/10.1002/chem.200901631.

[28] T.A. Brooks, S. Kendrick, L. Hurley, Making sense of G-quadruplex and i-motif functions in oncogene promoters, FEBS J. 277 (2010) 3459–3469. https://doi.org/10.1111/j.1742-4658.2010.07759.x.

[29] M. del Toro, P. Bucek, A. Aviñó, J. Jaumot, C. González, R. Eritja, R. Gargallo, Targeting the G-quadruplex-forming region near the P1 promoter in the human BCL-2 gene with the cationic porphyrin TMPyP4 and with the complementary C-rich strand, Biochimie. 91 (2009) 894–902. https://doi.org/10.1016/j.biochi.2009.04.012.

[30] K. Guo, V. Gokhale, L.H. Hurley, D. Sun, Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene, Nucleic Acids Res. 36 (2008) 4598–4608. https://doi.org/10.1093/nar/gkn380.

[31] A. Arora, D.R. Nair, S. Maiti, Effect of flanking bases on quadruplex stability and Watson-Crick duplex competition, FEBS J. 276 (2009) 3628–3640. https://doi.org/10.1111/j.1742-4658.2009.07082.x.

[32] D. Renciuk, I. Kejnovská, P. Školáková, K. Bednárová, J. Motlová, M. Vorlícková, D. Renčiuk, I. Kejnovská, P. Školáková, K. Bednářová, J. Motlová, M. Vorlíčková, Arrangements of human telomere DNA quadruplex in physiologically relevant K+ solutions, Nucleic Acids Res. 37 (2009) 6625–6634. https://doi.org/10.1093/nar/gkp701.

[33] A. Marchand, V. Gabelica, Folding and misfolding pathways of G-quadruplex, Nucleic Acids Res. 44 (2016) 10999–11012. https://doi.org/10.1093/nar/gkw970.

[34] K. Guo, A. Pourpak, K. Beetz-Rogers, V. Gokhale, D. Sun, L.H. Hurley, Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene, J. Am. Chem. Soc. 129 (2007) 10220–10228. https://doi.org/10.1021/ja072185g.

[35] M.L. Greco, M. Folini, C. Sissi, Double stranded promoter region of BRAF undergoes to structural rearrangement in nearly physiological conditions, FEBS Lett. 589 (2015) 2117–2123. https://doi.org/10.1016/j.febslet.2015.06.025.

[36] H. Abou Assi, R. El-Khoury, C. González, M.J. Damha, H. Abou Assi, R. El-Khoury, C. González, M.J. Damha, 2'-Fluoroarabinonucleic acid modification traps G-quadruplex and i-motif structures in human telomeric DNA, Nucleic Acids Res. 45 (2017) 11535–11546. https://doi.org/10.1093/nar/gkx838.

[37] J. Zhou, S. Amrane, D.N. Korkut, A. Bourdoncle, H. He, L. Ma, J. Mergny, Combination of i-Motif and G-Quadruplex Structures within the Same Strand : Formation and Application, Angew. Chemie - Int. Ed. 52 (2013) 7742–7746. https://doi.org/10.1002/anie.201301278.

[38] M. Trajkovski, J. Plavec, Chasing particularities of guanine- And cytosine-rich DNA strands, Molecules. 25 (2020) 434. https://doi.org/10.3390/molecules25030434.

[39] L. Liu, C. Ma, J.W. Wells, T. V. Chalikian, Conformational Preferences of DNA Strands from the Promoter Region of the c-MYC Oncogene, J. Phys. Chem. B. 124 (2020) 751–762. https://doi.org/10.1021/acs.jpcb.9b10518.

[40] T. Panczyk, P. Wolski, Molecular dynamics analysis of stabilities of the telomeric Watson-Crick duplex and the associated i-motif as a function of pH and temperature, Biophys. Chem. 237 (2018) 22–30. https://doi.org/10.1016/j.bpc.2018.03.006.

[41] Y. Cui, D. Kong, C. Ghimire, C. Xu, H. Mao, Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA, Biochemistry. 55 (2016) 2291–2299. https://doi.org/10.1021/acs.biochem.6b00016.

[42] L.N. Zhu, B. Wu, D.M. Kong, Specific recognition and stabilization of monomeric and multimeric G-quadruplexes by cationic porphyrin TMPipEOPP under molecular crowding conditions, Nucleic Acids Res. 41 (2013) 4324–4335. https://doi.org/10.1093/nar/gkt103.

[43] C. Romera, O. Bombarde, R. Bonnet, D. Gomez, P. Dumy, P. Calsou, J.-F.F. Gwan, J.-H.H. Lin, E. Defrancq, G.G. Pratviel, Improvement of porphyrins for G-quadruplex DNA targeting, Biochimie. 93 (2011) 1310–1317. https://doi.org/https://doi.org/10.1016/j.biochi.2011.06.008.

[44] L. Martino, B. Pagano, I. Fotticchia, S. Neidle, C. Giancola, Shedding light on the interaction between TMPyP4 and human telomeric quadruplexes, J. Phys. Chem. B. 113 (2009) 14779–14786. https://doi.org/10.1021/jp9066394.

[45] Y.X. Xiong, Z.S. Huang, J.H. Tan, Targeting G-quadruplex nucleic acids with heterocyclic alkaloids and their derivatives, Eur. J. Med. Chem. 97 (2015) 538–551. https://doi.org/10.1016/j.ejmech.2014.11.021.

[46] I. Lubitz, N. Borovok, A. Kotlyar, Interaction of monomolecular G4-DNA nanowires with TMPyP: Evidence for intercalation, Biochemistry. 46 (2007) 12925–12929. https://doi.org/10.1021/bi701301u.

[47] H. Han, D.R. Langley, A. Rangan, L.H. Hurley, Selective interactions of cationic porphyrins with G-quadruplex structures, J. Am. Chem. Soc. 123 (2001) 8902–8913. https://doi.org/10.1021/ja002179j.

[48] G.N. Parkinson, R. Ghosh, S. Neidle, Structural basis for binding of porphyrin to human telomeres, Biochemistry. 46 (2007) 2390–2397. https://doi.org/10.1021/bi062244n.

10. Tables

	pH 7.1		pH 6.0		pH 5.0		pH 4.2	
Sequence	295 nm	260 nm	295 nm	260 nm	295 nm	260 nm	295 nm	260 nm
SMG01	-- (+6.2%)	55.1 (+18.0%)	49.3 (-10.1%)	52.2 (+16.2%)	52.5 (-15.9%)	-- (+12.9%)	65 (-5.2%)	-- (+9.9%)
SMG03 (a)	37.0 (-22.9%)	-- (+19.0%)	48.7 (-38.8%)	-- (+3.2%)	59.7 (-34.2%)	-- (+8.5%)	not determined	not determined
SMC01	-- (+15.4%)	-- (+8.2%)	32.7 (-7.0%)	-- (+8.4%)	45.5 (-22.1%)	43.1 (+15.7%)	42.4 (+7.5%)	44.3 (+15.9%)
SMG01:SMC01	78.3 (+73.7%)	79.4 (+27.0%)	80.0 (+65.4%)	82.0 (+27.5%)	66.1 (-80.0%)	66.6 (+22.1%)	46.9 (+12.9%)	50.0 (+19.3%)

Table 1. Determined T_m values from absorbance melting traces. Hyperchromicity values are given between brackets. Experimental conditions were 150 mM KCl, 20 mM phosphate (pH 7.1 and 6.0) or acetate (pH 5.0 and 4.2) buffer, 2 μM DNA. The associated uncertainty in T_m values is ± 1°C. (a) Data from reference [9]. "--" indicates a smooth, not sigmoidal transition that did not show any well-defined T_m value.

Sequence	$pH_{1/2,\ 1}$	$pH_{1/2,\ 2}$	$pH_{1/2,\ 3}$	$pH_{10\%}$
SMG01	11.8±0.1 (1)	4.7±0.1 (1)	3.8±0.2 (1)	5.7
SMG01:SMC01	12.1±0.1 (2)	4.3±0.2 (1)	3.0±0.2 (1)	5.5
SMG03 (a)	11.1±0.1 (1)	7.1±0.2 (1)	4.7±0.4 (1)	8.2
SMG03:SMC03	12.2±0.1 (1)	4.8±0.1 (1)	4.0±0.2 (1)	5.8

Table 2. Summary of the pH-transition midpoints ($pH_{1/2}$), *n* values (among brackets) and $pH_{10\%}$ calculated from the multivariate analysis of spectroscopic data recorded along acid-base titrations of the considered DNA sequences. (a) Data from reference [9].

Sequence	Binding stoichiometry (DNA:ligand)	Logarithm of binding constant	T_m (°C)
SMG01	1:1	6.4 ± 0.5	60.7 (260 nm) 72.3 (290 nm)
	1:4	6.0 ± 0.5	
SMG03	1:1	6.7 ± 0.3	----- (260 nm) 65.0 (290 nm)
	1:2	6.0 ± 0.4	
SMG01:SMC01	1:8	6.1 ± 0.3	> 90 (260 nm) 51.5 (285 nm)

Table 3. Summary of the results obtained from binding experiments. CD and molecular absorption data were analyzed simultaneously. Experimental conditions of the titrations were 20°C, 150 mM KCl, 20 mM phosphate buffer, pH 7.1.

11. Figure captions

Figure 1. (a) DNA sequences studied in this work. (b) Proposed structure for the antiparallel G-quadruplex formed by SMG03 sequence at pH lower than 7.4 [9]. (c) Mesotetrakis-(N-methyl-4-pyridyl)-porphyrin (TMPyP4).

Figure 2. Melting of SMG01 monitored by molecular absorption spectroscopy. (a) Spectra measured in 150 mM KCl, 20 mM buffer, pH 7.1, 2 μM DNA. Inset: Thermal Difference Spectrum. (b) Normalized absorbance melting traces at 295 and 260 nm.

Figure 3. Acid-base titration of the SMG01 sequence at 20°C. (a) Selected CD spectra measured along the titration. Inset: ellipticity traces at 265 nm (blue symbols) and 290 nm (red symbols). (b) Selected molecular absorption spectra. Inset: absorbance traces at 295 nm. Experimental conditions were 3.5 μM DNA, 20°C and 150 mM KCl. (c) Calculated distribution diagram for a model of four acid-base species. (d) Calculated CD spectra for each one of these acid-base species. (e) Calculated molecular absorption spectra.

Figure 4. Acid-base titration of a 1:1 mixture of SMG01:SMC01 sequences at 20°C. (a) CD spectra recorded along the titration. (b) Experimental (blue symbols) and fitted (green line) ellipticity values at 265 nm. (c) Calculated distribution diagram of species. (d) Calculated pure CD spectra. Experimental conditions were: $C_{SMG01} = C_{SMC01} = 0.5$ μM, 20°C, 150 mM KCl.

Figure 5. CD-monitored melting experiments of SMG01:SMC01 mixtures at pH 7.1 (a) and 4.2 (b). Normalized absorbance melting curves for the SMG01:SMC01 mixture at several pH values monitored at 260 (c) and 295 (d) nm. In all cases, the experimental conditions were $C_{SMG01} = C_{SMC01} = 2$ μM, 150 mM KCl and 20 mM phosphate (pH 7.1 and 6.0) or acetate (5.0 and 4.2) buffer.

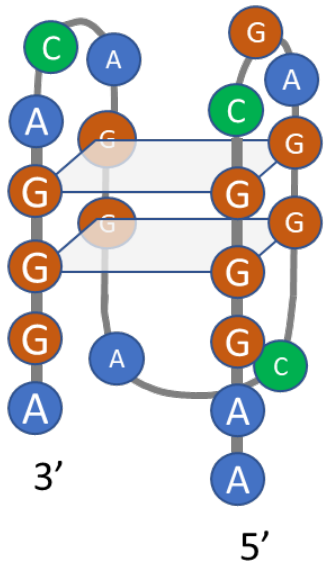
Figure 6. Interaction of SMG01 with TMPyP4. (a) CD spectra recorded along the titration of SMG01 with TMPyP4. C_{DNA} was 2 μM, 20°C, 150 mM KCl, and 20 mM phosphate buffer (pH 7.1). Insets refer to the TMPyP4:SMG01 ratio. (b) Calculated distribution diagram (up) and pure CD spectra (down) for the species considered.

Figure 7. Melting of the SMG01:TMPyP4 (2 μM:12 μM) mixture (a) and of the SMG01:SMC01:TMPyP4 (0.6 μM:0.6 μM:6 μM) (b). In both cases, CD spectra were recorded along the melting from 10 to 95°C. Insets show ellipticity traces measured at 260 (blue symbols) and 285 nm (red symbols). Experimental conditions were 150 mM KCl, 20 mM phosphate buffer, pH 7.1.

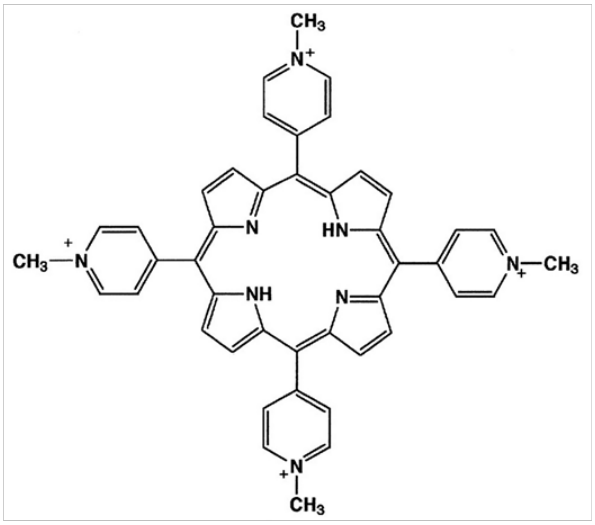
a

Sequence name	Sequence of bases (5’ -> 3’)	Length
SMG01	GGG GTT CAT GAC CAA GGG CGA GGC AGG ACA GGG ATA GCA AGG GA	44
SMC01	CCC CAA GTA CTG GTT CCC GCT CCG TCC TGT CCC TAT CGT TCC CT	44
SMG03	AA GGG **C**GA GGC AGG A**C**A GGG A	21
SMC03	TT CCC GCT CCG TCC TGT CCC T	21

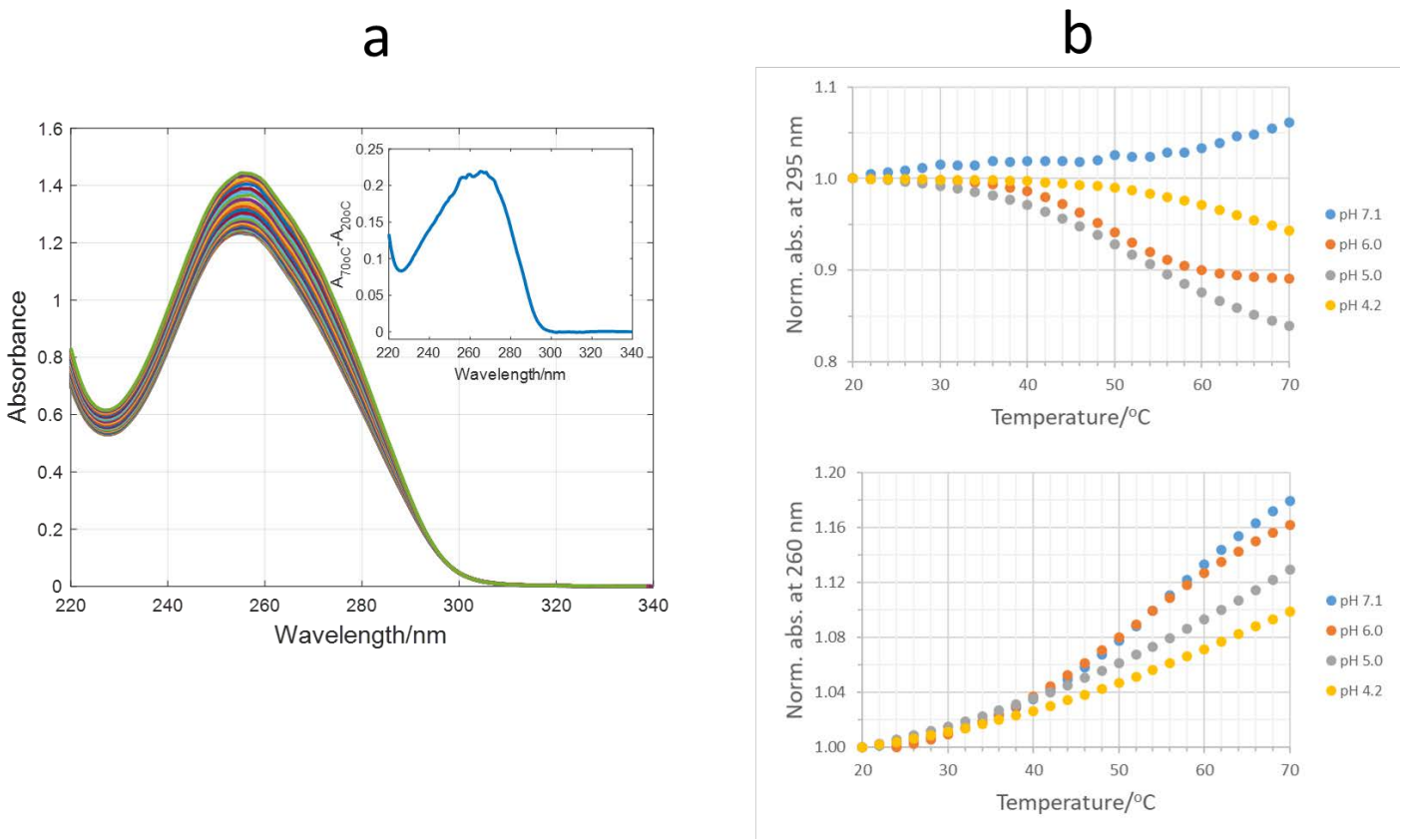
b



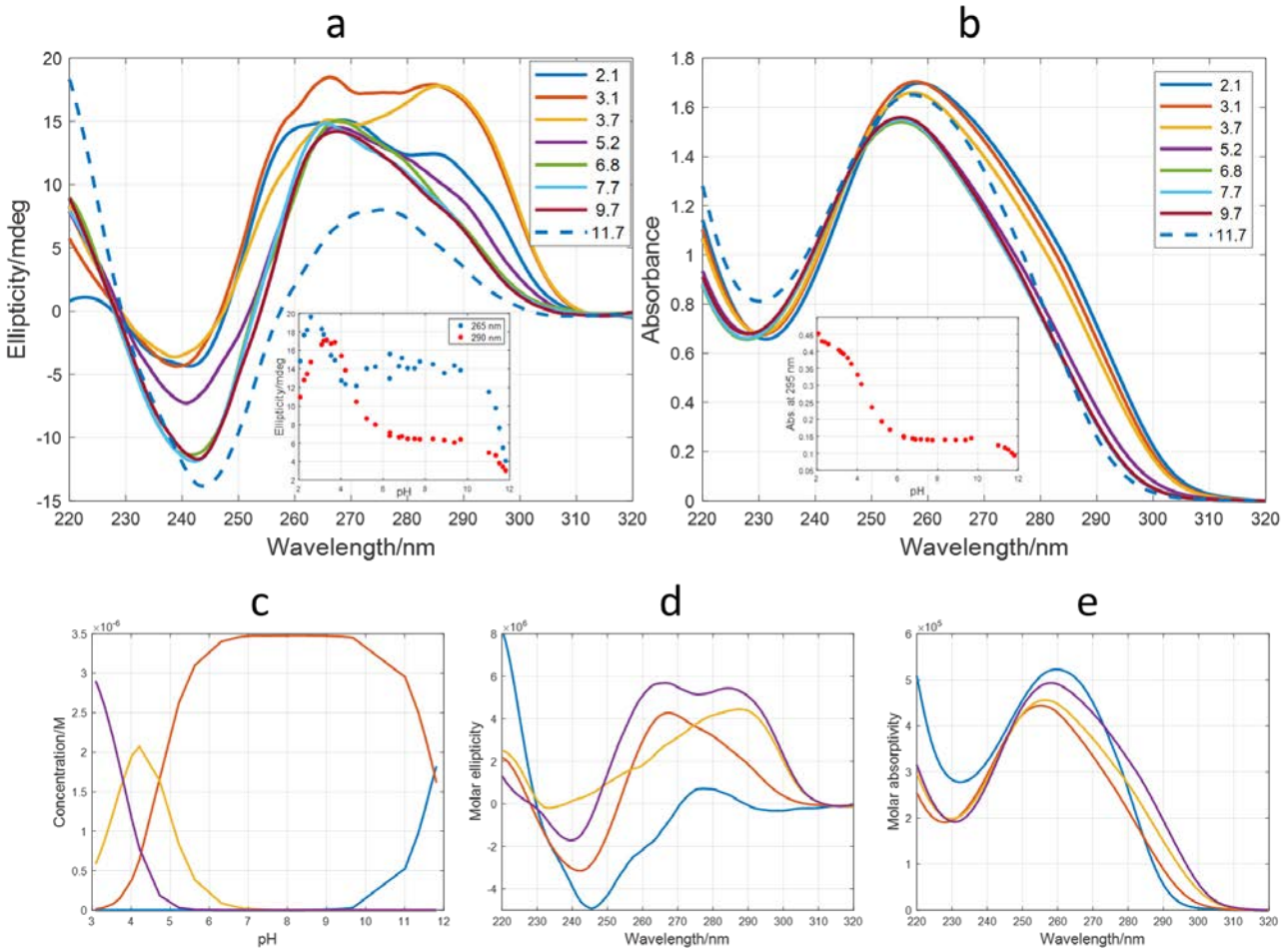
c



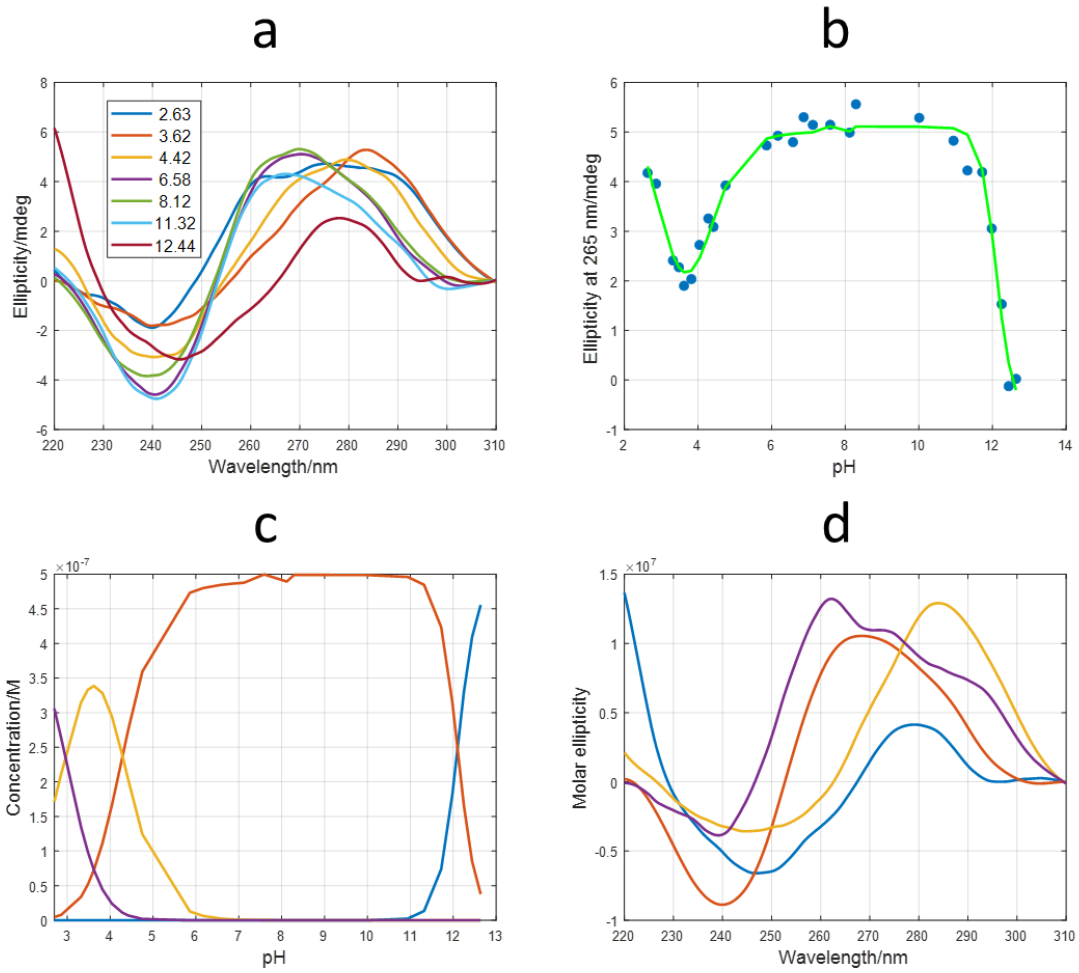
International Journal of Biological Macromolecules. Navarro et al. Figure 1.



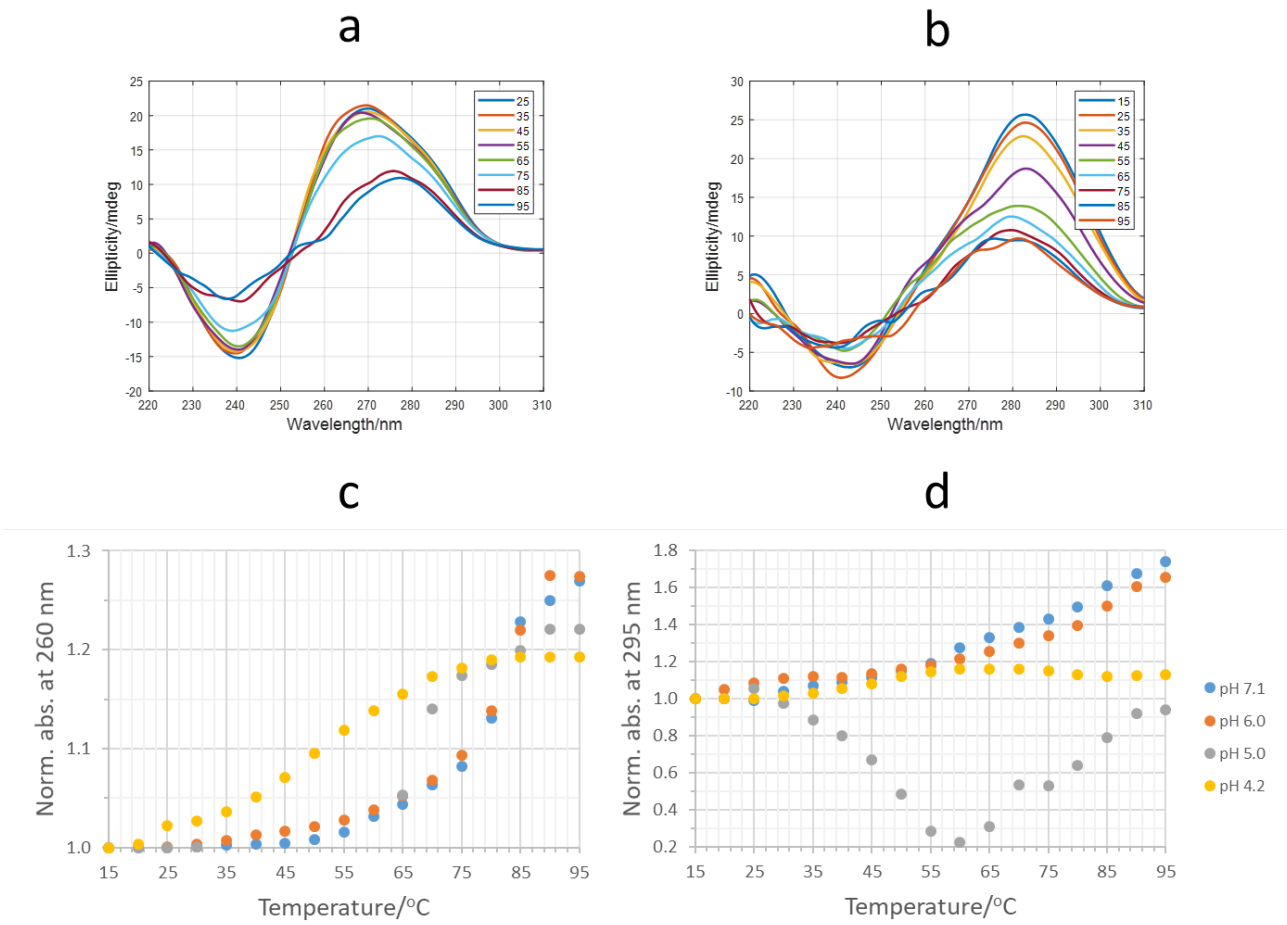
International Journal of Biological Macromolecules. Navarro et al. Figure 2.



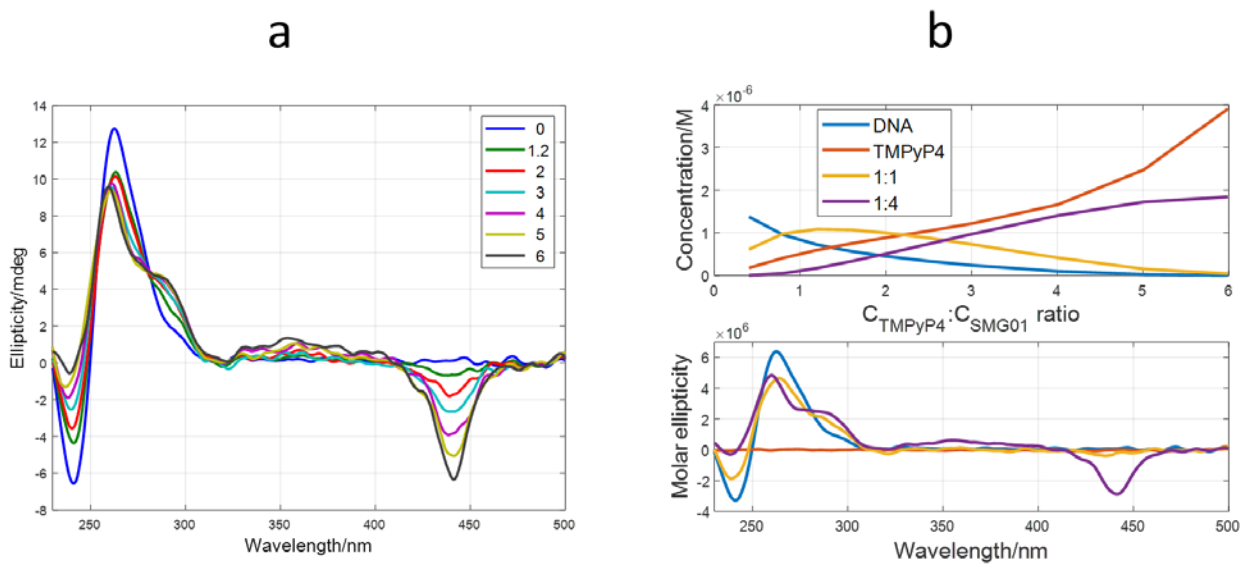
International Journal of Biological Macromolecules. Navarro et al. Figure 3.



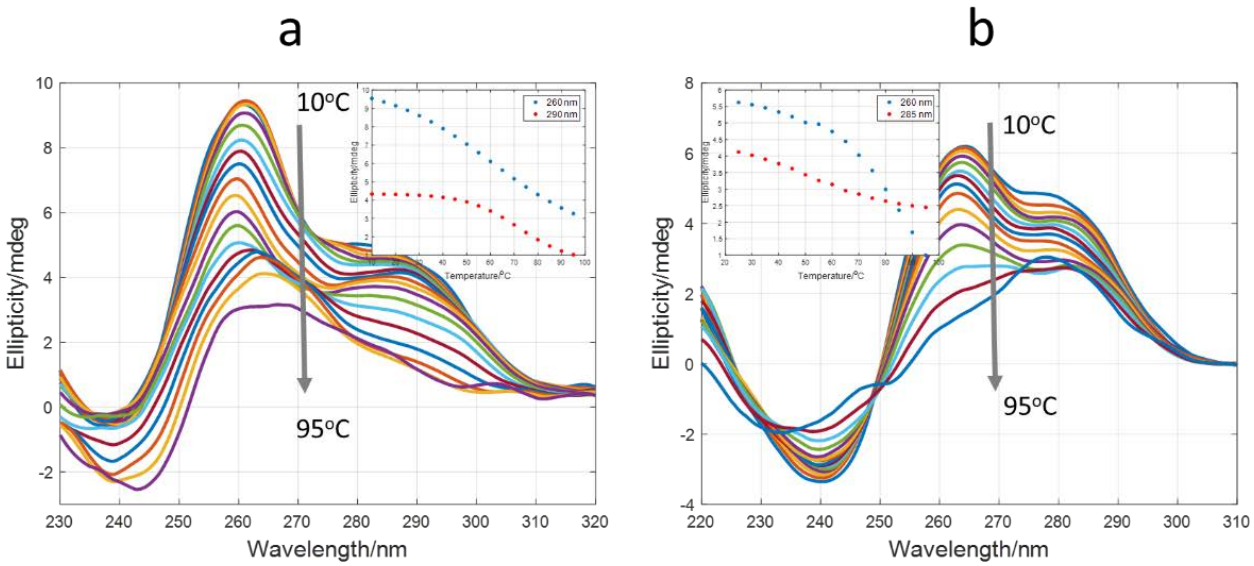
International Journal of Biological Macromolecules. Navarro et al. Figure 4.



International Journal of Biological Macromolecules. Navarro et al. Figure 5.



International Journal of Biological Macromolecules. Navarro et al. Figure 6.



International Journal of Biological Macromolecules. Navarro et al. Figure 7.