"Influence of pH, temperature and the cationic porphyrin TMPyP4 on the stability of the i-motif formed by the 5'-(C3TA2)4-3' sequence of the human telomere". Fernández, S., Eritja, R., Aviñó, A., Jaumot, J. Gargallo, R. Int. J. Biol. Macromol., 49(4), 729-736, (2011).

PMID: 21777611, doi: 10.1016/j.ijbiomac.2011.07.004

Influence of pH, temperature and the cationic porphyrin TMPyP4 on the stability of the *i*-motif formed by the 5'-$(C_3TA_2)_4$-3' sequence of the human telomere

Sergio Fernández[1], Ramon Eritja[2], Anna Aviñó[2], Joaquim Jaumot[1], Raimundo Gargallo[1*]

1. Solution equilibria and Chemometrics group (Associate Unit UB-CSIC), Department of Analytical Chemistry, University of Barcelona, Diagonal 647, E-08028 Barcelona, Spain

2. Institute for Research in Biomedicine, IQAC-CSIC, CIBER-BBN Networking Centre on Bioengineering, Biomaterials and Nanomedicine, Baldiri Reixac 10, E-08028 Barcelona, Spain

* Corresponding author

Tel: (34)-934039274

fax: (34)-934021233

E-mail address: raimon_gargallo@ub.edu

Abstract

The influence of pH, temperature and the cationic porphyrin 5, 10, 15, 20-tetra(N-methyl-4-pyridyl)porphin (TMPyP4) on the conformational equilibria of the cytosine-rich strand of the human telomeric sequence 5'-$(C_3TA_2)_4$-3', as well as those of the related sequence 5'-$(C_3TT_2)_4$-3', has been studied by means of molecular absorption and circular dichroism spectroscopies. Data recorded along these experiments have been analyzed by means of multivariate data analysis methods.

Acid-base titrations of 5'-$(C_3TA_2)_4$-3' sequence throughout the pH range 3 - 7 and melting experiments showed the formation of up to two different intramolecular i-motif structures with a pH-transition midpoint around 4.6. Both structures show lower stability than the *i*-motif structure formed by 5'-$(C_3TT_2)_4$-3'. The results obtained have shown that the substitution of thymine by adenine at the loops destabilizes the *i*-motif structure.

The study of the interaction equilibria of i-motif structures formed by 5'-$(C_3TA_2)_4$-3' has revealed the formation of 1:1 DNA:TMPyP4 complex with a stability constant equal to $10^{5.9}$ M^{-1}. A similar study done with the sequence 5'-$(C_3TT_2)_4$-3' has shown the formation of 1:1 and 1:2 complexes, which points out to a role of the loop on the interaction with this ligand.

Keywords: *i*-motif, human telomere, Chemometrics, loops, TMPyP4 binding

Introduction

Cytosine-rich strands are known to form complex structures called *i*-motif. This structure is composed by two duplexes intercalated in an antiparallel way, i.e., is the only DNA structure where base pairs are intercalated. In *i*-motif, pairs are composed mainly by cytosine base pairs maintained thanks to the protonation of one of the cytosine bases in the pair. Due to this protonation requirement, *i*-motif structures have a marginal presence in neutral solutions at 37^{o}C. The *i*-motif structure may include four identical C-rich strands, two hairpins each carrying two cytidine stretches or a folded strand carrying four cytidine stretches [1]. Owing to its specific self-recognition and susceptibility to pH variation, *i*-motif structures have been utilized as building components for fabricating molecular devices [2, 3].

In vivo, cytosine-rich regions coexist with the complementary guanine-rich region. It has been shown that guanine-rich sequences can also form a complex structure called G-quadruplex, which is stabilized by the formation of hydrogen bonds among four planar guanines [4]. In spite of the high stability of the Watson-Crick duplex, G-quadruplex seems to be stable at physiological pH when appropriate conditions are established. Accordingly, the complementary cytosine-rich region could be involved in the formation of *i*-motif structures in similar conditions. In this context, *i*-motif structures have been proposed in several cytosine-rich sequences corresponding to the oncogenes *RET [5]*, *c-myc [6, 7]*, *bcl-2* [8], *Rb* [9] or *c-jun* [10]. In addition, human telomere DNA is composed of multiple repeats of 5'-TTAGGG on one strand and 5'-CCCTAA on the other. The guanine-rich strand is longer and has a single-strand overhang of approximately 100-150 base pairs, allowing the formation of a G-quadruplex secondary structure. Except for this short 3'-end guanine-rich overhang, all chromosomal DNAs potentially capable of forming G-quadruplexes are masked by their Watson-Crick complementary cytosine-rich strand DNA. Formation of a G-quadruplex structure within genomic DNA should therefore be coupled with the self-organization of the complementary cytosine-rich strand [11-14].

The potential biological importance of the intramolecular *i*-motif seems to be evidenced by its involvement in human telomeric and centromeric DNA structures and RNA intercalated structures, and by the discovery of several proteins that bind specifically to cytosine to DNA sequences containing four cytosine-stretches with at least three cytidines [15-18]. Because of this, along with human telomeric G-quadruplex DNA, *i*-motif has also been postulated as an attractive drug target for cancer treatment and for modulation of gene transcription [19].

Several works have been published dealing with the solution equilibria of *i*-motif formed by several sequences related to telomeric DNA. Hence, the i-motif formed by the arrangement of four molecules of the shortest sequence 5'-(C_3TA_2)-3' has been studied by means of NMR, PAGE and CD [11, 20]. The bimolecular

folding of the two repeat telomere sequence (5'-($C_3TA_2)_2$-3') was first studied by Ahmed et al. [21]. Other structural research has been focused on longer sequences such as 5'-CCC(TAACCC)$_3$-3'[13], 5'-(C_3TA_2)CCC-3'[22], 5'-(C_3TA_2)CCCT-3'[23, 24]. The 24-bases long sequence 5'-($C_3TA_2)_4$-3' has also been focus of attention [21, 25-27].

To our knowledge, however, no attempt has been made to study from an analytical point of view the solution equilibria of the 24-bases long sequence and of the related sequence 5'-(CCCTTT)$_4$-3' [28]. In addition, there is only a single work dealing with the interaction of a sequence based on the human telomere with the model ligand 5, 10, 15, 20-tetra(N-methyl-4-pyridyl)porphin (TMPyP4) [11]. The present work deals with both objectives and, with this purpose, acid-base titrations and melting experiments have been carried out to establish the pH- and temperature-range of stability for each of the structures proposed here. In addition, mole-ratio experiments have been carried out to study the interaction of the model ligand TMPyP4 with the different structures. All experiments were monitored by means of spectroscopic techniques and the recorded data were analyzed by means of appropriate multivariate data analysis methods [29, 30]. The results described here show that the sequence 5'-(CCCTAA)$_4$-3' can form two different *i*-motifs depending on pH changes. Secondly, the presence of adenine bases destabilizes the *i*-motif structure in relation to that formed by the sequence 5'-(CCCTTT)$_4$-3'. Finally, the study of the interaction of TMPyP4 with both sequences by means of multivariate methods confirms a weak interaction, mainly electrostatic in nature.

Material and methods

Reagents

DNA sequences C3TT2 (5'-(CCC TTT)$_4$-3'), C3TA2 (5'-(CCC TAA)$_4$ -3') and T24 (5'-T$_{24}$ -3') were prepared as described elsewhere [8]. DNA strand concentration was determined by absorbance measurements (260 nm) using calculated extinction coefficients and the nearest-neighbor method [31]. Before any experiment, DNA solutions were first heated at 90^{o}C for 5 - 10 minutes and then allowed to reach room temperature. KCl, KH_2PO_4, K_2HPO_4, $NaCH_3COO$, HCl and NaOH (a.r.) were purchased from Panreac (Spain). MilliQ ® water was used in all experiments. The cationic porphyrin 5,10,15,20-tetra(N-methyl-4-pyridyl)porphin (TMPyP4) was purchased from Porphyrin Systems (Germany).

Procedures

Absorbance spectra were recorded on an Agilent HP8453 diode array spectrophotometer. The temperature was controlled *via* an 89090A Agilent Peltier device. Hellma quartz cells (1 or 10 mm path length, and 350, 1500 or 3000 μl volume) were used. CD spectra were recorded on a Jasco J-810 spectropolarimeter equipped with a Julabo F-25/HD temperature control unit. Hellma quartz cells (10 mm path length, 3000 μl volume) were used. pH measurements were determined with an Orion SA 720 pH/ISE meter and micro-combination pH electrode (Thermo).

Acid-base titrations were monitored either in-line (taking advantage of the stirrer incorporated in the Agilent cell holder) or at-line (in the case of the CD instrument). Experimental conditions were as follows: 25^{o}C and 150 mM KCl. Titrations were carried out by adjusting the pH of solutions containing the oligonucleotides. CD and/or absorbance spectra were recorded in a pH stepwise fashion.

Melting experiments were monitored with an Agilent-8453 spectrophotometer equipped with the Agilent temperature-controlling Peltier unit. The DNA solution was transferred to a covered 10-mm-path-length cell and UV/VIS absorption spectra were recorded at 1^{o}C intervals with a hold time of 3 min at each temperature value, which yields an average heating rate of ~ 0.3 ^{o}C min^{-1}. Buffer solutions were 20 mM phosphate or acetate, and 150 mM KCl. Each sample was allowed to equilibrate at the initial temperature for 30 minutes before the melting experiment was begun.

Mole ratio experiments were carried out either by addition of small volumes of a DNA stock solution to a TMPyP4 solution or vice versa. In the first case, experiments were monitored with molecular absorption spectroscopy, whereas in the second case they were monitored with circular dichroism spectroscopy.

Experimental conditions were as follows: 25°C, pH 7.0, 5.2 and 4.0, and 150 mM KCl. Buffer solutions (20 mM phosphate or acetate) were the same as those described for the melting experiments.

Data analysis

Spectra recorded along acid-base titrations, melting experiments or mole ratio studies were arranged in a table or data matrix, which size is *m* rows (spectra recorded) and *n* columns (wavelengths measured).

The goal of the data analysis is the calculation of the distribution diagrams and pure (individual) spectra for all the *nc* spectroscopically-active species considered throughout an experiment. The distribution diagram provides information about the stoichiometry and stability of the considered species (in case of acid-base and mole-ratio experiments), and about thermodynamics in case of the melting processes. In addition, the shape and intensity of the pure spectra may provide qualitative information about the structure of the considered species.

With this goal in mind, the data matrix **D** is decomposed according to the Beer-Lambert-Bouer's law in matrix form:

$$\mathbf{D} = \mathbf{C}\,\mathbf{S}^{\mathbf{T}} + \mathbf{E} \qquad (1)$$

Now, **C** is the matrix (*m* x *nc*) containing the distribution diagram, $\mathbf{S}^{\mathbf{T}}$ is the matrix (*nc* 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**, $\mathbf{S}^{\mathbf{T}}$, and **E** may be done basically in two different ways, depending whether a physico-chemical model is initially proposed (hard-modeling approach) or not (soft-modeling approach).

For hard-modelling approaches, the proposed model depends on the nature of the studied process. Hence, for acid-base experiments the model will include 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, like the following:

$$DNA + pH^{+} \leftrightarrow DNA{\cdot}H_p \qquad Beta_{1p} = [DNA{\cdot}H_p] \,/\, [DNA]\,[H^{+}]^{p} \qquad (2)$$

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

For mole-ratio experiments, the physico-chemical model is similar to the previous one. As example:

$$DNA + qL \leftrightarrow DNA \cdot L_q \qquad Beta_{1q} = [DNA \cdot L_q] / [DNA]\,[L]^q \qquad (3)$$

For melting experiments, the physico-chemical model is related to the thermodynamics of DNA unfolding. Hence, for the unfolding of intramolecular structures such as those studied in this work, the chemical equation and the corresponding equilibrium constant may be written as:

$$DNA_{\,folded} + heat \leftrightarrow DNA_{\,unfolded} \qquad K_{unfolding} = [DNA_{\,unfolded}] / [DNA_{\,folded}] \qquad (4)$$

For melting experiments, the concentration of the folded and unfolded forms is temperature-dependent. Accordingly, the equilibrium constant depends on temperature according to the van't Hoff equation:

$$\ln K_{\,unfolding} = - \Delta H_{vH} / RT + \Delta S_{vH} / R \qquad (5)$$

Here it is assumed that ΔH_{vH} and ΔS_{vH} will not change throughout the range of temperatures studied in this work.

Whenever a physico-chemical model is applied, the distribution diagram in **C** fulfills the proposed model. Accordingly, the proposed values for the equilibrium constants and the shape of the pure spectra in $\mathbf{S}^T$ are refined to explain satisfactorily data in **D**, whereas residuals in **E** are minimized.

The application of hard-modeling methods (i.e., those based on the fulfillment of a previously proposed model) has advantages and drawbacks. Hence, not only the calculated distribution diagrams and pure spectra are more robust in relation to the experimental noise, but also reliable values for the thermodynamic parameters (stability constants, ΔH_{vH}, ΔS_{vH}) may be calculated. In addition, mixtures of two or more species may be resolved with an acceptable incertitude when the stability constants have been calculated unambiguously. The main drawback, however, is the compulsory proposal of a model. Very often, it is difficult to ascertain the more appropriate model to explain a given process, and the chemical intuition must replace, hopefully, this incertitude. On the other hand, data in matrix **D** may include variance due to additional factors not related with the studied process (base line drift, impurities, …), which cannot be appropriately modeled.

The mathematical decomposition of **D** into matrices **C**, $\mathbf{S}^T$, and **E** may also done without applying a physico-chemical model, which is known as soft-modeling. In this case, the calculation of matrices **C** and $\mathbf{S}^T$ is based on the compliance of a series of constraints which reduce the initial large number of mathematical solutions to an (almost) unique physico-chemically meaningful solution. An example of constraint is that which impose that the elements in matrix **C** cannot be negative (as expected in a distribution diagram). The calculation of **C** and $\mathbf{S}^T$ is done throughout an alternate optimization process until a previously set degree of convergence is reached. The application of soft-modeling methods is proposed whenever a physico-chemical model cannot be easily proposed. Its main drawback is, probably, the inability to resolve complex mixtures without the

help of more complex experimental and data analysis setups (for instance, the simultaneous analysis of several data matrices corresponding to complementary experiments). Recently, several hybrid approaches which combine the advantages of hard- and soft-modeling methods have been proposed.

In this work, hard-modeling analysis of acid-base and mole-ratio experiments has been done with the EQUISPEC program [32]. Hard-modeling of melting experiments has been done with a modified version of the MCR-ALS procedure which includes the model proposed in equation (5) for unfolding of intramolecular structures. A detailed explanation of data analysis of melting experiments is included in Supporting Material. Soft-modeling analysis has been done with the graphical user interface of MCR-ALS [33, 34], freely available at the web address www.mcrals.info.

Results

C3TT2

The first oligonucleotide which solution equilibria have been studied in this work was the sequence C3TT2. It is expected the formation of an *i*-motif structure in slightly acid solution stabilized by a maximum of six $C^+ \cdot C$ base pairs [28]. Only thymine bases are present at the loops. Because of the high pK_a value of thymidine nucleoside (around 9.7 [35]) only acid-base equilibria involving cytosine will affect to the stability of the *i*-motif structure in the studied pH range.

First of all, spectroscopically-monitored acid-base titrations of C3TT2 solutions were performed. Molecular absorption spectra and circular dichroism (220 - 320 nm) were recorded stepwise throughout the pH range 3.1 – 7.6 (Supplementary Material). All spectra were arranged in a matrix **D** and later analyzed with the soft- and hard-modeling methods named MCR-ALS and Equispec, respectively. Detailed descriptions of these analyses were given previously [8, 29, 33]. Figure 1 shows the results obtained with Equispec, being the results obtained with MCR-ALS similar to these. Three spectroscopically-active acid-base species were enough to explain the measured spectroscopic data within the experimental incertitude. Increasing the number of species did not improve significantly the data fit (Supplementary Material).

According to the resolved distribution diagrams (matrix **C**) and pure spectra (matrix **S**), these three acid-base species were explained as follows. The major species in neutral solutions has been related to an partially stacked strand, probably left-handed [36], where cytosine bases are neutral (i.e., deprotonated at N3 [35]). Upon acidification, this conformation evolves to an *i*-motif structure which is the major species throughout the pH range 6.0 – 4.5, approximately. The presence of an *i*-motif structure is well characterized by a CD spectrum showing characteristic positive and negative bands centered at 285 and 265 nm, respectively (Figure 1c), being the intensity of the first one twice of the second. The acid-base model which best fits the experimental absorbance data is summarized in Table 1. The first transition, which implies the transition from the neutral species to the *i*-motif, is cooperative as the Hill's coefficient is equal to 3. The pH-transition midpoint is located at 6.7±0.1, which means that, at the experimental conditions of this experiment, there is a minor proportion of the *i*-motif at pH 7. At pH values lower than 4.5 the *i*-motif structure unfolds to yield an acid-base species where all cytosine bases are protonated at N3. The transition between these two species is also cooperative as the Hill's coefficient is also equal to 3. The location of the pH-transition midpoint (3.6±0.1), which is far from the pK_a of isolated cytosine (around 4.5 - 4.8), suggest a high stability for the *i*-motif structure.

Spectroscopically-monitored melting experiments of C3TT2 sequence were carried out at several pH values. Molecular absorption spectra were recorded stepwise in function of temperature and finally arranged in a data matrix **D**. Then, matrix **D** was analyzed to determine the number of transitions present throughout the temperature range studied and to evaluate thermodynamic parameters, such as ΔH and ΔS. As example, the experimental spectra recorded along the melting of C3TT2 at pH 6.6 and the steps done throughout the data analysis are shown and explained in detail in Supplementary Material. Figure 2 shows the distribution diagram (matrix **C**) and pure spectra (matrix **S**) resolved for the three proposed conformations. The species present at the beginning of the melting (Figure 2a) has been assigned to the initial mixture (where the intramolecular *i*-motif is the major conformation at this pH) which melts with a T_m around 23^{o}C to yield, probably, an partially stacked single strand (second species). Increasing the temperature produces a small variation of the absorbance spectra which is modeled by including a third species (the major at the end of the melting). The resolved spectra for the second and third species are rather similar, but no equal, which suggest that both species correspond to strands differing in the degree of base stacking. At pH 6.6 unfolding of C3TT2 is accompanied by ΔH=66 kcal·mol^{-1} and ΔS=222 cal·K^{-1}·mol^{-1}. Assuming that the breaking of a C$^+$·C base pairs needs around 11 ± 1 kcal·mol^{-1} [28] it can be deduced that the number of C$^+$·C base pairs broken throughout this transition was 6.0 ± 0.8, i.e., it can be deduced that six C$^+$·C base pairs were disrupted. The thermodynamic parameters of a similar sequence $(CCCTTT)_3CCC$ were previously determined from a melting experiment carried out at pH 6.0 (10mM cacodylate, 0.1 M NaCl) [28]. The authors proposed an intramolecular folding, characterized by T_m, ΔH and ΔS equal to 45^{o}C, -72 kcal·mol^{-1} and -226 cal·K^{-1}·mol^{-1}, respectively.

Table 3 summarizes the ΔH, ΔS, and the ΔG value at 37^{o}C values, as well as T_m values determined from the resolved distribution diagram for the transitions involving unfolding of the *i*-motif. The dependence of T_m values with pH (Figure 3) is that expected for *i*-motif structures, i.e., their stabilities are enhanced upon protonation. The low stability of the i-motif formed by C3TT2 at pH near 7 explains why at 25^{o}C practically all C3TT2 is unfolded, as reflected in the resolved acid-base diagram. The dependence of T_m values with DNA concentration has been studied at pH 6.1 and 4.7. In both cases, the determined T_m values remained unaltered (41^{o}C and 60^{o}C, respectively) for C_{DNA} ranging from 0.5 to 5 μM, suggesting a predominant intramolecular folding throughout this concentration range.

C3TA2

The study of the acid-base equilibria of C3TA2 provided different results to those observed previously for C3TT2. Hence, the model which best fitted the data comprised up to four acid-base species (i.e., three

transitions), one more than in the case of C3TT2 (Table 1 and Figure 4). The CD spectrum of the neutral species show negative and positive bands of similar intensity and centered at 250 and 275 nm, respectively. This spectrum is clearly different to that observed for C3TT2, and very similar to that of 5'-T_{24}-3' which is known to adopt a right handed helical structure (Supplementary Material) [35]. The first transition, which is related to the folding of *i*-motif from this neutral species, is cooperative (Hill's coefficient equal to 3). The third transition is less cooperative than the first one (Hill's coefficient equal to 2) and is clearly related to the unfolding of *i*-motif in acid solution. The second transition, however, is more difficult to explain. The CD spectra corresponding to the second and third species are similar (positive and negative bands centered at 285 and 258 nm, respectively), which confirms that both species are *i*-motif structures. On the other hand, the shape and intensity of the resolved absorbance spectra suggest that the difference between these two structures is due probably to the protonation of a few bases. It should be noted the slight shift of the negative CD band in relation to the same band in C3TT2, which points out to a slightly different *i*-motif conformation. Finally, the transition between these two acid-base species is strongly cooperative (Hill's coefficient equal to 3). Taken together all these evidences, the second transition has been attributed to a conformational change between two highly ordered structures (*i*-motif I and *i* -motif II).

Melting experiments showed that T_m values determined for C3TA2 were lower than those determined for C3TT2 throughout the pH range 4.5 – 6.5, approximately, revealing a destabilizing effect of adenine bases on the stability of *i*-motif (Table 2). The independence of T_m with concentration has been confirmed at pH 6.1, 4.5 and 4.0 throughout the DNA concentration range 0.5 - 5 μM, suggesting a predominant intramolecular folding at all these pH values.

Interaction with the cationic porphyrin TMPyP4

Finally, the interaction of C3TT2 and C3TA2 sequences with the model ligand TMPyP4 has been studied at several pH values. First, CD spectra of all five proposed acid-base species of C3TT2 and C3TA2 in absence and in presence of TMPyP4 were recorded (Supplementary Material). In all cases, the structure of the oligonucleotide did not show any significant modification upon binding, if any. In order to know the exact extent of ligand binding mole-ratio experiments monitored by molecular absorption spectroscopy at several pH values were carried out. Table 3 summarizes the results obtained. As example, Figure 5 shows the calculated distribution diagrams and pure spectra for the proposed species involving C3TA2 and C3TT2 sequences at pH 5.2.

At pH 7.0, both C3TT2 and C3TA2 adopt single-stranded helical structures which bind up to two TMPyP4 molecules, with a overall stability constant around 10^{12} M^{-1}. Mole-ratio experiments done with the control

sequence 5'-T_{24}-3', which also produces single-stranded helical structures, provide similar results (stability constant = $10^{12.4}$ M^{-1}).

At pH 5.2 and 4.0, the experimental CD spectra of C3TT2 and C3TA2 did not show any dramatic modification upon addition of drug, a fact which reflects the overall maintenance of the initial *i*-motif structure upon binding. In contrast, mole-ratio experiments provided slightly different mode of binding for both sequences (Table 3). Hence, whereas only one interaction complex (1:1) was enough to explain the experimental variation observed along the titration of TMPyP4 with C3TA2, two interaction complexes (1:1 and 1:2) were needed to achieve similar results for the C3TT2 sequence. Figure 5 (and b) shows the calculated distribution diagram and the pure molecular absorption spectra corresponding to each one of the three species involved (free TMPyP4, C3TA2 and 1:1 interaction complex). The shift of the Soret band, as well as the calculated value for the equilibrium constant ($10^{6.0}$ M^{-1}), is characteristics of a rather weak binding, probably electrostatics in nature. A similar conclusion can be drawn for the 1:1 complex present in the C3TT2:TMPyP4 system.

Discussion

The cytosine-strand of the human telomere 5'-$(C_3TA_2)_4$-3' (C3TA2) contains four tracts of three cytosines which may yield the formation of an intramolecular *i*-motif structure in slightly acid solutions. Much effort has been done in the elucidation of the spatial structure of the i-motif formed by this sequence [21, 22, 24, 37], as well as those formed by truncated DNA sequences based on C3TA2. In our opinion, there is still a lack of analytical information regarding the acid-base equilibria of this particular sequence, as well as of the interaction with model ligands, such as the porphyrin TMPyP4.

In the present work, the acid-base equilibria of the C3TA2 sequence, as well of those of the related sequence C3TT2, have been studied. Melting experiments have been done to corroborate the results obtained in these studies. Whereas the C3TT2 sequence clearly showed the formation of an intramolecular *i*-motif throughout the pH range from 6.5 to 3.0, the acid-base behavior of C3TA2 in this pH range needed the proposal of two acid-base species, both related to *i*-motif structure. The differentiation between these two species has been possible because both multivariate data were recorded and appropriate data analysis methods were used. It would have been very difficult to discern the existence of these two species from the measurement of absorbance values at just one or two wavelengths. Two additional evidences may support our hypothesis of two acid-base species formed by C3TA2. First, the plot of T_m *vs.* pH for C3TA2 and C3TT2 sequences (Figure 3) reveals a different stability of both sequences in the pH range of existence of i-motif I (6.5 – 4.8, approximately) . A previous work dealing with the sequence 5'-TCCTCCTTTTCCTCCT-3' [28], lacking of adenine bases, showed a dependence of T_m with pH similar to that observed here for C3TT2. Secondly, the NMR study of a natural fragment of the human centromeric satellite III, 5'-CCATTCCATTCCTTTCC-3', showed that it forms two monomeric *i*-motif structures that differ in their intercalation topology and that are favored at pH values higher and lower than 4.6 [38]. The change in intercalation topology was suggested being related to adenine protonation. Interestingly, the pH transition midpoint coincides with that observed in our work (4.6, Figure 4). However, the pK_a of free adenosine is around 3.5 – 4.1 [35, 39] and, therefore, an additional mechanism (such as base pair binding) should be taken into account to explain an apparent pK_a around 4.6.

It is clear from the present study and from previous works that the presence of adenine bases at the potential loops affects the stability of *i*-motif structure, as already observed for several topologies of G-quadruplex structures [40]. The higher stability showed by C3TT2 in comparison to C3TA2 was already observed by Mergny *et al.* when studying the sequences 5'-$(C_3TA_2)_3$CCC-3' and 5'-$(C_3TT_2)_3$CCC-3' [28]. This higher stability could be due to the presence of T·T base pairs which expand the six C^+·C base pairs located at

the core of the *i*-motif. According to Figure 3, it seems that the formation of T·T base pairs may increase the T_m value of the *i*-motif structure more than the formation of T·A base pairs. Another possibility could be the formation of a three-base hydrogen-bonding network, in which two adenines and a thymine form four hydrogen bonds via a reverse Hoogsteen and an asymmetric adenine-adenine base pairing [41].

Recently, Kaushik *et al.* studied the 24-bases long 5'-$(C_3TA_2)_4$-3' sequence [25], as well as several truncated 9- and 12-bases long [27]. Their results showed that the 24-bases long sequence may form both intra and intermolecular *i*-motifs depending on the experimental conditions. At the high concentrations used in DSC experiments, intermolecular folding was preferred over intramolecular one at pH 4.8. In general, the T_m values determined in our work agree with those previously determined by Kaushik *et al*. for the same sequence at different experimental conditions (10 mM acetate or cacodylate, no KCl added). However, an independence of T_m with concentration has been observed here at several pH values throughout the DNA concentration range 0.5 - 5 μM, suggesting an intramolecular folding at all these pH values. Previous works dealing with very similar sequences also observed intramolecular folding in the presence of salts like KCl, NaCl or LiCl [21, 42]. Because of this, it seems that the presence of salts may influence the inter- or intramolecular nature of folding of these cytosine-rich DNA sequences.

The presence of additional protonated bases in the i-motif formed by C3TA2 could explain why its interaction with TMPyP4 differed from that for C3TT2. Whereas only one 1:1 spectroscopically active complex was proposed for the system C3TA2:TMPyP4, two 1:1 and 1:2 complexes were detected for the system C3TT2:TMPyP4. To our knowledge, there is only one previous study of the interaction of TMPyP4 with an i-motif whose sequence resembles that of the human telomere (5□-AACCCC-3□, which forms a tetraplex i-motif [11]). TMPyP4 was found to promote the formation of the i-motif structure, binding to the edge of the DNA core by a non-intercalative mechanism. Analysis of absorbance data revealed the anti - cooperativity of binding, 1:2 (DNA:drug) stoichiometry and a binding constant of 104.34M−1. This value is lower than those determined in this study for the two i-motifs of C3TA2 (around 106M−1, at pH 5.2 and 4.0). This suggests that porphyrin interactions with fold-over molecules may be different due to the influence of the structure of connecting loops.

Conclusions

The present work deals with the solution equilibria of the 5'-$(C_3TA_2)_4$-3' sequence located at the cytosine-rich strand of the human genome. Two *i*-motif structures have been observed with a pH transition midpoint around 4.6 at 25°C and 150 mM KCl. It has been suggested that they differ in the degree of protonation of adenine bases at the edge of the C^+·C core of the structure. The presence of adenine has been also shown to

destabilize the *i*-motif structure when compared with that formed by the 5'-$(C_3TT_2)_4$-3' sequence. Finally, the interaction equilibria of all *i*-motif structures formed by these two sequences with the model ligand TMPyP4 have been shown to be non-intercalative in nature.

Acknowledgements

We acknowledge funding from both the Spanish government (CTQ2009-11572 and CTQ2010-20541-C03-01) and Catalan government (2009 SGR 45 and 2009 SGR 238).

Figures and tables

Table 1. Parameters calculated for the acid-base transitions using the hard-modeling program Equispec.

Table 2. Thermodynamic parameters calculated from melting experiments. Experiments carried out in phosphate or acetate buffers. T_m values are given in ^{o}C. ΔH and ΔG values (at 37^{o}C) are given in kcal·mol^{-1}. ΔS is given in cal·K^{-1}·mol^{-1}.

Table 3. Proposed models for the interaction of TMPyP4 with the considered sequences. $\Delta\lambda$ (nm) refers to the spectral shift measured between the Soret band corresponding to free TMPyP4 and that corresponding to the calculated spectrum for the predominant interaction complex. The values for log beta(11) and log beta(12) refer to the logarithm of the overall stability constant for the 1:1 and 1:2 DNA:TMPyP4 interaction complexes, respectively.

Figure 1. Results obtained after analysis with Equispec of the spectra recorded along acid-base titrations of C3TT2 monitored by molecular absorption and circular dichroism spectroscopies. (a) Distribution diagram; (b) pure molecular absorption spectra; (c) pure CD spectra. Solid line: neutral species; dotted line: i-motif; dash-dotted line: random coil. C_{C3TT2} = 1.8 μM. Other experimental conditions as detailed in the text.

Figure 2. Results obtained after analysis with constrained MCR-ALS of absorbance data recorded along the melting experiment of C3TT2 at pH 6.6. (a) Resolved distribution diagram, (b) Resolved pure spectra. C_{C3TT2}: 5.3 μM. Other experimental conditions as detailed in the text.

Figure 3. T_m values determined against pH. Black diamond: C3TT2; open square: C3TA2.

Figure 4. Results obtained after analysis with Equispec of the spectra recorded along acid-base titrations of C3TA2 simultaneously monitored by molecular absorption and circular dichroism spectroscopies. (a) Distribution diagram; (b) pure molecular absorption spectra; (c) pure CD spectra. Solid line: neutral species; dotted line: i-motif I; dashed line: i-motif II; dash-dotted line: random coil. C_{C3TA2} = 3.7 μM. Other experimental conditions as detailed in the text.

Figure 5. Interaction of TMPyP4 with C3TA2 and C3TT2 at pH 5.2. Mole-ratio experiment monitored with molecular absorption spectroscopy. C3TA2: (a) Calculated distribution diagram. (b) Calculated pure molecular absorption spectra. Initial C_{TMPyP4} = 3.3 μM, C_{C3TA2} ranges from 0 to 3.1 μM. C3TT2: (c) Calculated distribution diagram. (d) Calculated pure molecular absorption spectra. Initial C_{TMPyP4} = 3.0 μM, C_{C3TT2} ranges from 0 to 3.5 μM. Solid line: *i*-motif, dotted line: TMPyP4, dashed line: 1:1 (DNA:ligand) interaction complex, dash-dotted line: 1:2 interaction complex. 25^{o}C, 20 mM acetate buffer (pH 5.2) and 150 mM ionic strength.

Table 1. Parameters calculated for the acid-base transitions using the hard-modeling program Equispec. Experiments done at 25^{o}C and 150 mM KCl. Standard deviation values are shown between parentheses.

Sequence	Transition from neutral species to i-motif structure		Transition from i-motif I to i-motif II		Transition from i-motif to random coil	
	p	log stability constant	*p*	log stability constant	*p*	log stability constant
C3TT2	3	20.2(0.1)	---	---	3	10.9(0.1)
C3TA2	3	17.8(0.1)	3	13.7(0.2)	2	5.6(0.2)

Table 2. Thermodynamic parameters calculated from melting experiments. T_m values are given in ^oC. ΔH and ΔG values (at 37^oC) are given in kcal·mol^{-1}. ΔS is given in cal·K^{-1}·mol^{-1}. Experiments carried out in 20 mM phosphate (pH 7.0) or acetate (pH 5.0) buffer, 150 mM KCl, C_{DNA} ranging from 0.5 to 5 μM.

	pH 4.0				pH 4.8				pH 5.3				pH 6.1				pH 6.6			
	T_m	ΔH	ΔS	ΔG	T_m	ΔH	ΔS	ΔG	T_m	ΔH	ΔS	ΔG	T_m	ΔH	ΔS	ΔG	T_m	ΔH	ΔS	ΔG
C3TT2	61	62.8	187.9	-4.5	60	63.4	190.2	-4.4	52	65.9	202.8	-3.0	41	63.2	201.3	-1.0	23	65.7	221.8	3.1
C3TA2	58	70.3	212.6	-4.4	60 [a]	54.2	162.9	-3.7	46	56.1	175.6	-1.6	32	61.7	202.0	0.9	27 [b]	66.7	222.0	2.1

[a] pH = 4.5; [b] pH = 6.3

Table 3. Proposed models for the interaction of TMPyP4 with the considered sequences. Δλ values refer to the spectral shifts measured between the Soret band corresponding to free TMPyP4 and those corresponding to the calculated spectra for the interaction complexes. The values for log beta(11) and log beta(12) refer to the logarithm of the overall stability constant for the 1:1 and 1:2 DNA:TMPyP4 interaction complexes, respectively. Experiments carried out at 25^oC, phosphate (pH 7.0) or acetate (pH 5.0) buffer, 150 mM KCl. The maximum of the Soret band for free TMPyP4 is 422 nm (25^oC, pH 7).

DNA	pH 4.0		pH 5.2		pH 7.0	
	Δλ in nm (complex)	Stability constant	Δλ (nm)	Stability constant	Δλ (nm)	Stability constant
C3TT2	13 (11)	log beta(11) = 6.0 ± 0.1	11 (11)	log beta(11) = 6.0 ± 0.2	11 (12)	log beta(12) = 12.0 ± 0.1
	7 (12)	log beta(12) = 11.9 ± 0.1	8 (12)	log beta(12) = 12.1 ± 0.2		
C3TA2	10 (11)	log beta(11) = 5.9 ± 0.1	10 (11)	log beta(11) = 6.0 ± 0.2	13 (12)	log beta(12) = 11.8 ± 0.1

Figure 1. Results obtained after analysis with Equispec of the spectra recorded along acid-base titrations of C3TT2 monitored by molecular absorption and circular dichroism spectroscopies. (a) Distribution diagram; (b) pure molecular absorption spectra; (c) pure CD spectra. Solid line: neutral species; dotted line: *i*-motif; dash-dotted line: random coil. Experimental conditions: 25°C, 150 mM KCl, C_{C3TT2} = 1.8 µM.

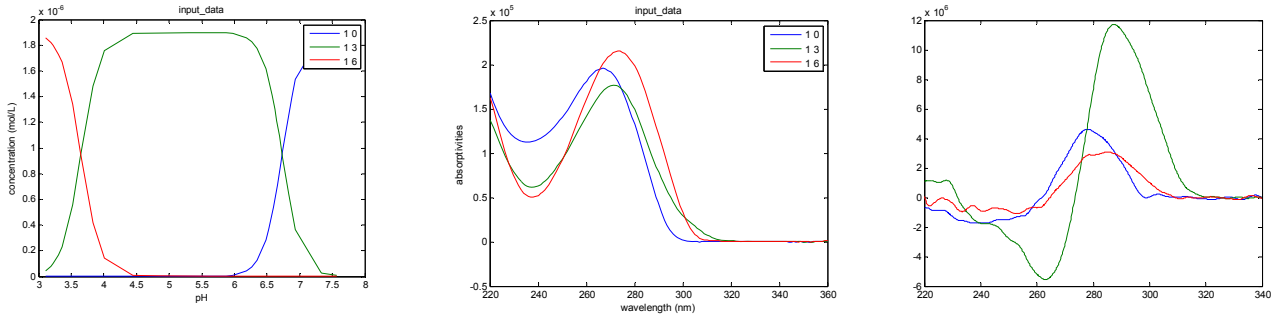


Figure 2. Results obtained after analysis with constrained MCR-ALS of absorbance data recorded along the melting experiment of C3TT2 at pH 6.6. (a) Resolved distribution diagram, (b) Resolved pure spectra. Experimental conditions: 20 mM phosphate buffer, 150 mM KCl,C_{C3TT2}= 5.3 µM.

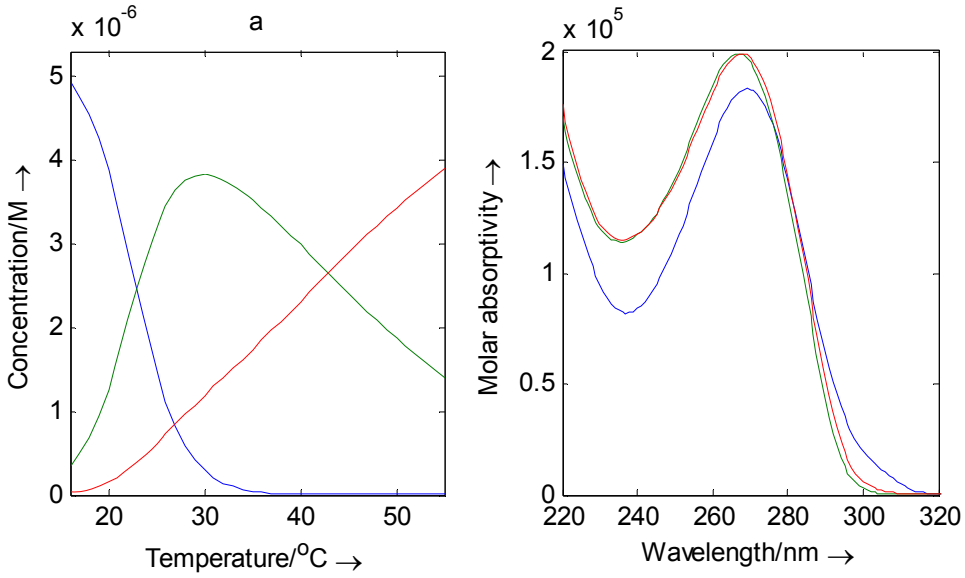


Figure 3. T_m values determined against pH. Black diamond: C3TT2; open square: C3TA2. Experimental conditions: 20 mM phosphate or acetate buffer, 150 mM KCl.

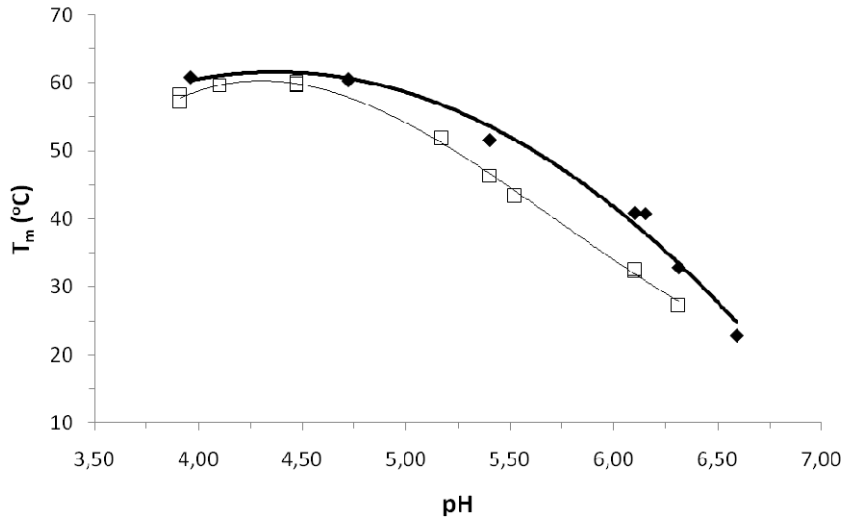


Figure 4. Results obtained after analysis with Equispec of the spectra recorded along acid-base titrations of C3TA2 simultaneously monitored by molecular absorption and circular dichroism spectroscopies. (a) Distribution diagram; (b) pure molecular absorption spectra; (c) pure CD spectra. Solid line: neutral species; dotted line: i-motif I; dashed line: i-motif II; dash-dotted line: random coil. Experimental conditions: 25°C, 150 mM KCl, C_{C3TA2} = 3.7 µM.

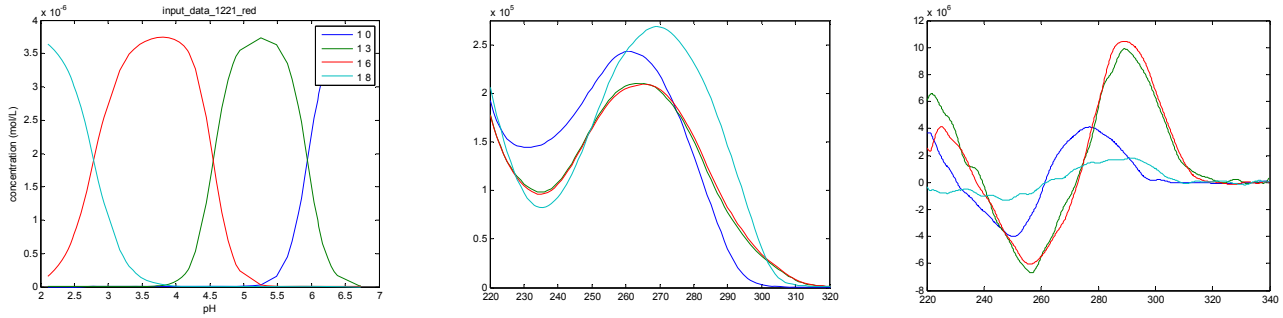
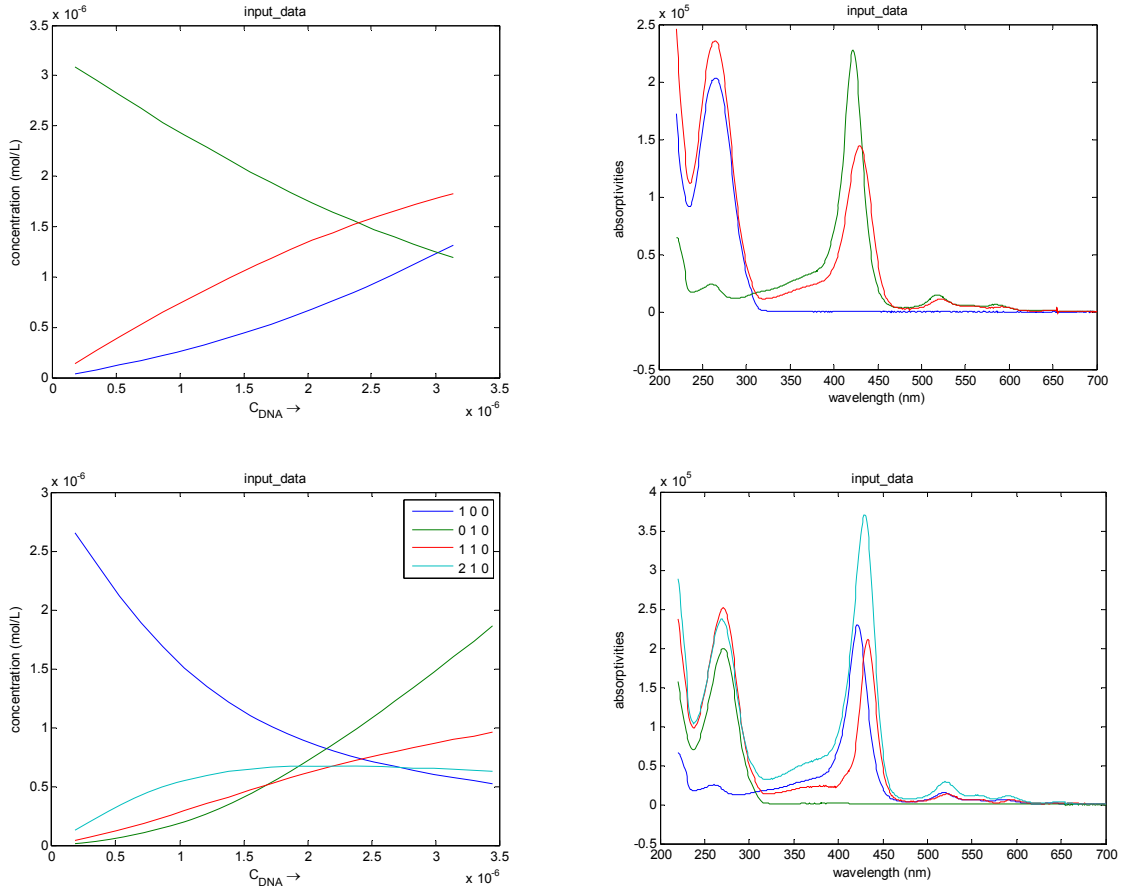


Figure 5. Interaction of TMPyP4 with C3TA2 and C3TT2 at pH 5.2. Mole-ratio experiment monitored with molecular absorption spectroscopy. C3TA2: (a) Calculated distribution diagram. (b) Calculated pure molecular absorption spectra. Initial C_{TMPyP4} = 3.3 μM, C_{C3TA2} ranges from 0 to 3.1 μM. C3TT2: (c) Calculated distribution diagram. (d) Calculated pure molecular absorption spectra. Initial C_{TMPyP4} = 3.0 μM, C_{C3TT2} ranges from 0 to 3.5 μM. Solid line: *i*-motif, dotted line: TMPyP4, dashed line: 1:1 (DNA:ligand) interaction complex, dash-dotted line: 1:2 interaction complex. 25°C, 20 mM acetate buffer (pH 5.2) and 150 mM ionic strength.



Supplementary Material

Influence of pH, temperature and the cationic porphyrin TMPyP4 on the stability of the *i*-motif formed by the 5'-$(C_3TA_2)_4$-3' sequence of the human telomere

Sergio Fernández[1], Ramon Eritja[2], Anna Aviñó[2], Joaquim Jaumot[1], Raimundo Gargallo[1*]

1. Solution equilibria and Chemometrics group (Associate Unit UB-CSIC), Department of Analytical Chemistry, University of Barcelona, Diagonal 647, E-08028 Barcelona, Spain

2. Institute for Research in Biomedicine, IQAC-CSIC, CIBER-BBN Networking Centre on Bioengineering, Biomaterials and Nanomedicine, Edifici Helix, Baldiri Reixac 15, E-08028 Barcelona, Spain

* Corresponding author

Tel: (34)-934039274; fax: (34)-934021233

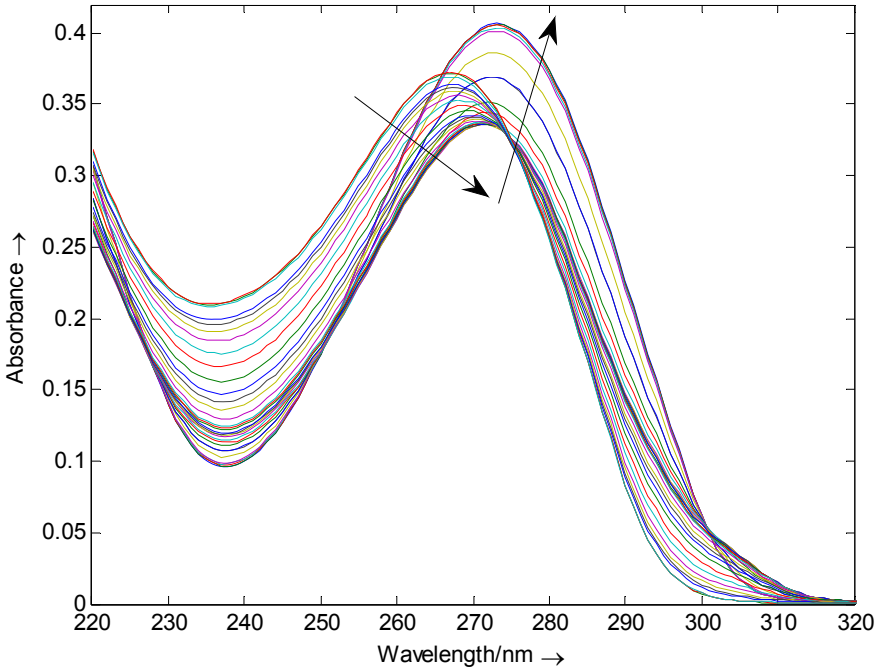
E-mail adress: raimon_gargallo@ub.edu

Contents:

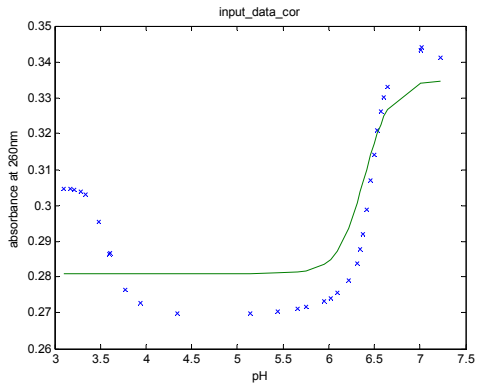
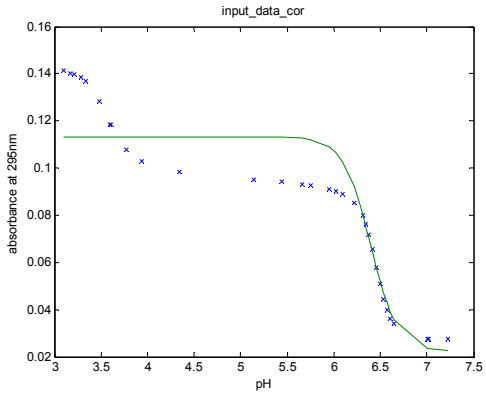
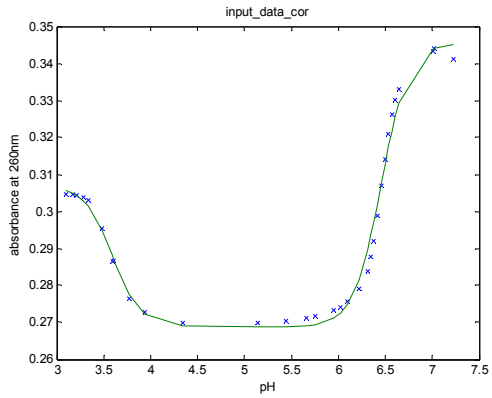
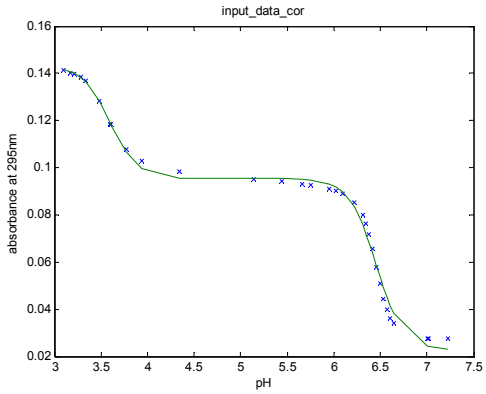
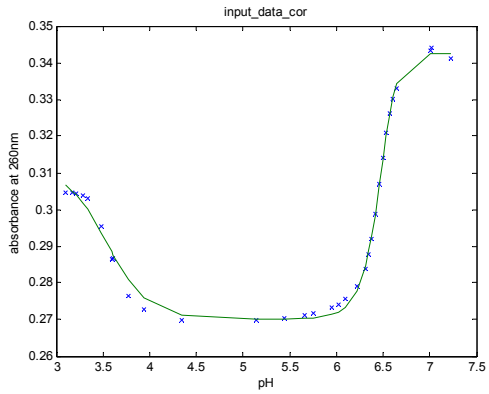
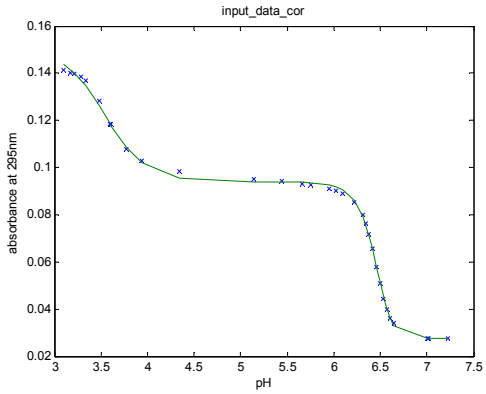
1. Acid-base titration of C3TT2 at 25°C.

1.1. Experimental spectra.

Arrows indicate the spectroscopic changes observed when the pH is raised from 7 to 3, approximately.



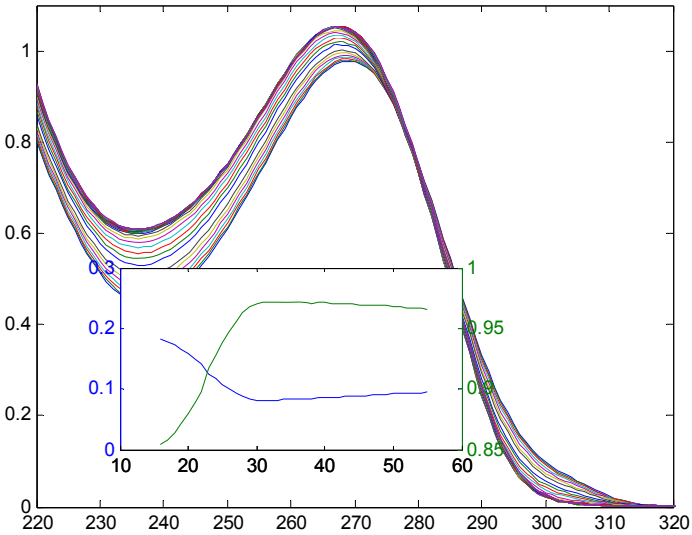
1.2. Fits obtained after analysis of absorbance data recorded along the acid-base titration of C3TT2 for different number of components (*nc*). ssq refers to the sum of squares of residuals.

nc	Fit at 260 nm	Fit at 295 nm	ssq
2			0.6890
3			0.014
4			0.006

2. Example of multivariate analysis of spectroscopic data recorded along a melting experiment

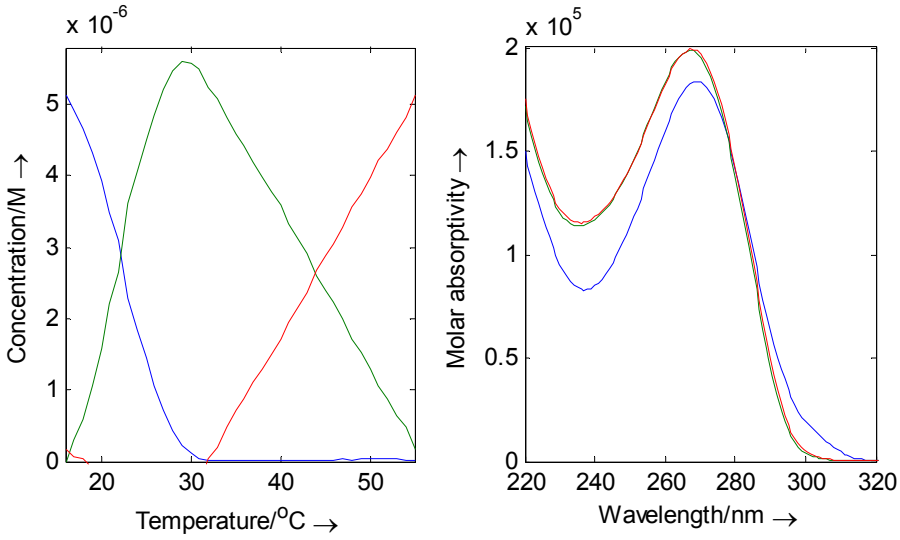
2.1. Experimental absorbance spectra

In this particular case, absorbance data were recorded stepwise throughout the temperature range 16 – 55°C, every 1°C. The traces at 295 (blue line) and 260 (green line) nm are studied (inset). The trace at 295 nm shows a clear hypochromism which may be related to the unfolding of *i*-motif structure (Mergny and Lacroix, *Oligonucleotides* **2003**, 13, 515). The monophasic transitions correspond to the presence of a single structural species in solution.



2.2. Analysis with MCR-ALS

First, spectroscopic data were analyzed by means of MCR-ALS. Several detailed examples of application of this procedure were given elsewhere (Khan *et al. Biochimie* **2007**, 89, 1562 and Bucek *et al. Chem. Eur. J.* **2009**, 15, 12663). In this case, three components (n=3) were considered. This number was selected after application of mathematical procedures such as Singular Value Decomposition or SIMPLISMA analysis of the data matrix plotted in 2.1. Alternatively, the analysis was repeated considering two or four components but the more reasonable results (from a biophysical point of view) were obtained when only three components were considered. Next figure shows the resolved distribution diagram and the corresponding pure spectra.



2.3. Analysis with constrained MCR-ALS

Now, the distribution diagram was constrained to fulfill the van't Hoff equation:

$$\ln K_{unfolding} = -\Delta H_{vH} / RT + \Delta S_{vH} / R$$

Where:

$$K_{unfolding} = [DNA_{unfolded}] / [DNA_{folded}]$$

Hence, for the unfolding of an intramolecular structure throughout a single transition, the concentration at equilibrium for the unfolded and folded conformations can be written as:

$$[DNA_{unfolded}] = \frac{C_{DNA}\, e^{\left(-\frac{\Delta H}{RT}+\frac{\Delta S}{R}\right)}}{1 + e^{\left(-\frac{\Delta H}{RT}+\frac{\Delta S}{R}\right)}}$$

$$[DNA_{folded}] = \frac{C_{DNA}}{1 + e^{\left(\frac{\Delta H}{RT}-\frac{\Delta S}{R}\right)}}$$

Where C_{DNA} is the analytical concentration of DNA:

$$C_{DNA} = [DNA_{unfolded}] + [DNA_{folded}]$$

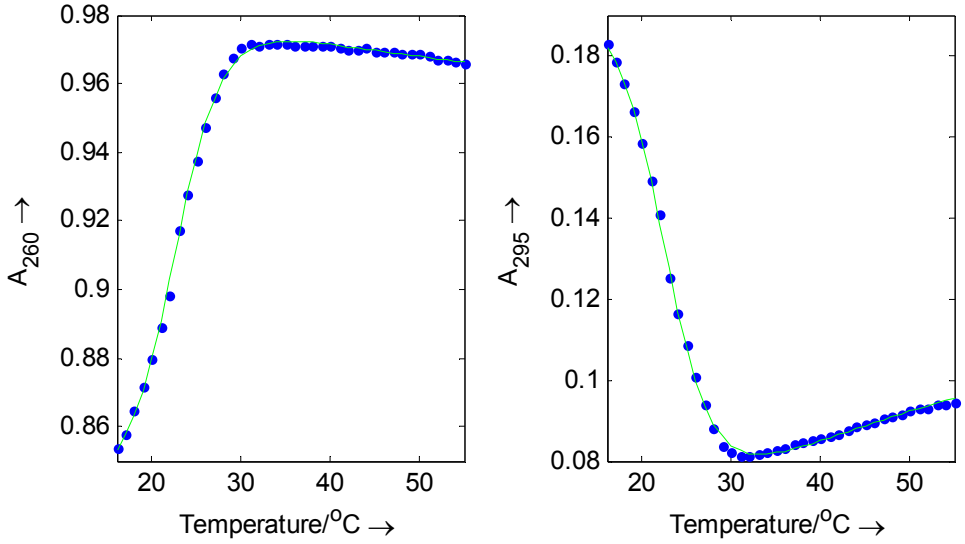
Hence, the equilibrium concentrations of both unfolded and folded DNA are constrained to fulfill the van't Hoff equation. By this way, most of the incertitude associated to the results obtained with soft-modeling methods can be removed. Moreover, estimates of thermodynamic parameters and of the melting temperature are obtained. The uncertainty associated with the thermodynamic values is ± 5-10 %,

depending on the level of noise of original data and of the degree of overlap among conformations. This uncertainty level was calculated from the analysis of previously simulated unfolding transitions.

Next table shows the fits obtained when one or two transitions (i.e., two or three components, *nc*) were considered. The best fit (given as sum of squares of residuals, ssqFit) was obtained when two transitions (i.e., three components) were considered. The intramolecular nature of the transition was simultaneously checked by doing melting experiments at increasing concentrations.

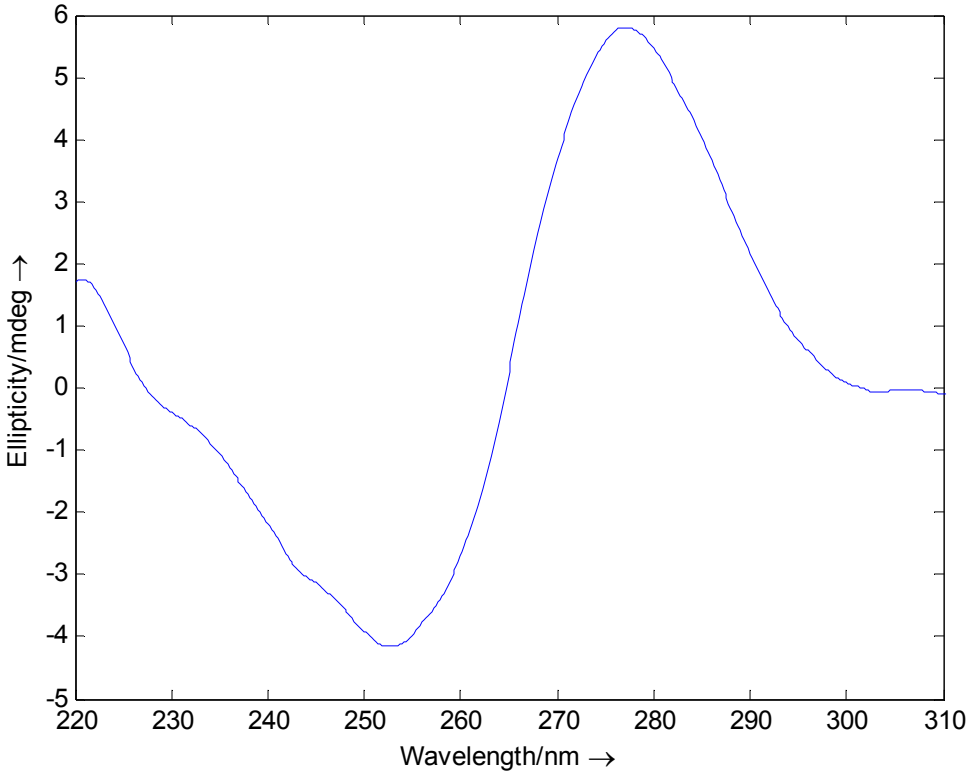
nc	Proposed model	ΔH ($kJ \cdot mol^{-1}$)	ΔS ($J \cdot K^{-1} \cdot mol^{-1}$)	T_m (^oC)	ΔH ($kJ \cdot mol^{-1}$)	ΔS ($J \cdot K^{-1} \cdot mol^{-1}$)	T_m (^oC)	ssqFit
2	Intramolecular structure → random coil	356.3	1205.8	22.3				0.0711
3	Intramolecular structure → intramolecular structure → random coil	273.7	924.2	23.0	72.7	230.0	---	0.0045

The resolved distribution diagram and the corresponding pure spectra have been included in the main text (Figure 2). Next figure shows the experimental (blue dots) and fitted (green line) absorbance data at 260 and 295 nm when the model of three components is considered. The goodness of the fit proves the proposed model of species. The presence of a higher number of conformations was rejected because the lack of fit obtained was below the experimental incertitude of the measurement.

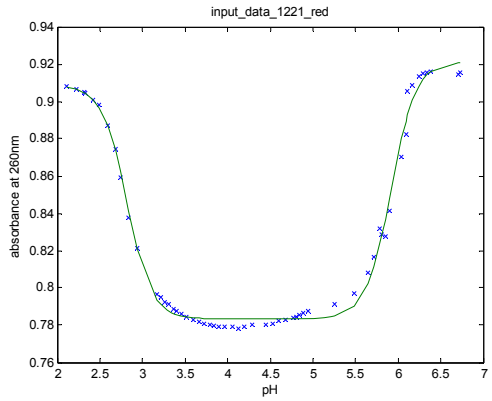
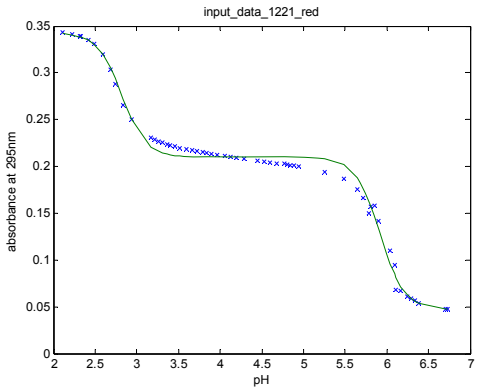
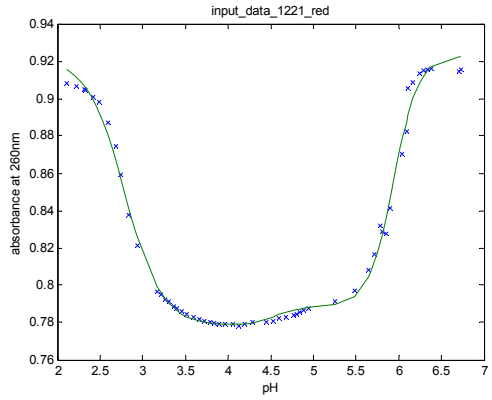
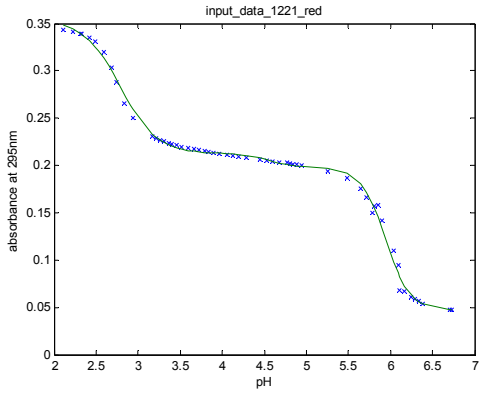


3. Acid-base titration of C3TA2 at 25°C.

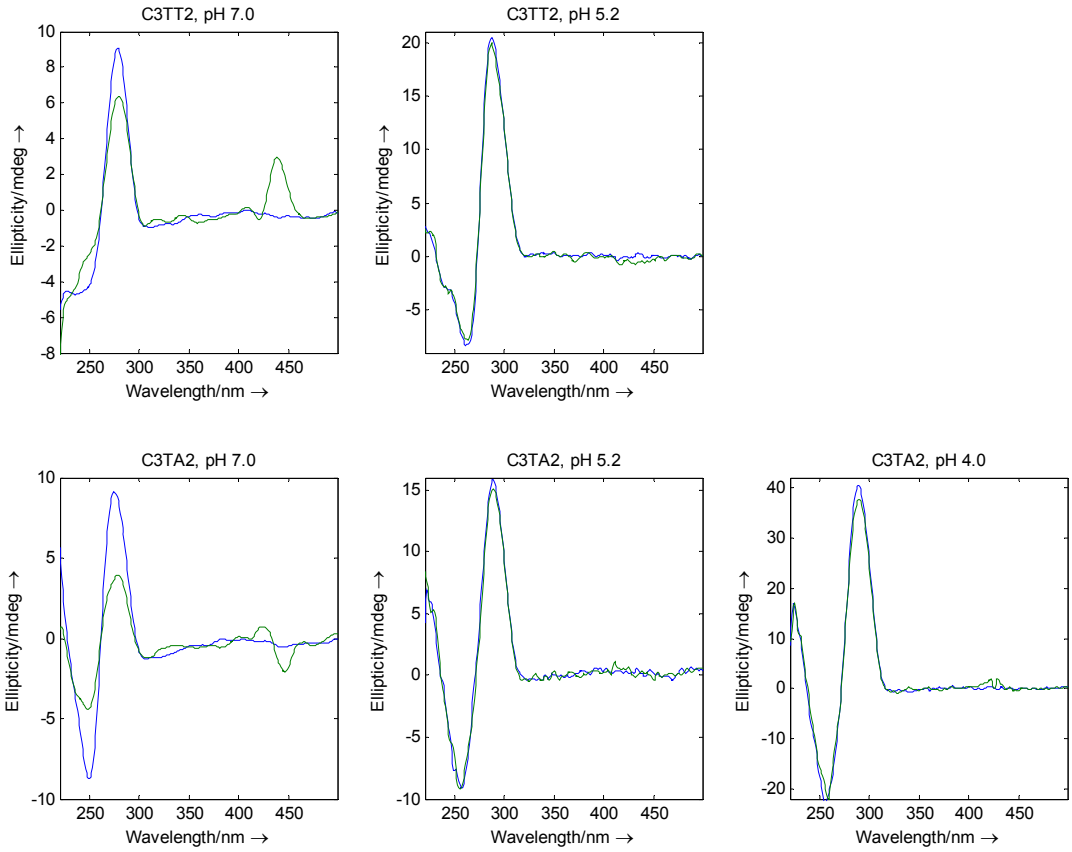
3.1. CD spectrum of 5'-T_{24}-3' in 150 mM KCl, 20 mM phosphate buffer, pH 7.0, 25°C, C_{T24} = 1.8 μM.



3.2. Fits obtained after analysis of absorbance data recorded along the acid-base titration of C3TA2 for different number of components (*nc*).

nc	Fit at 260 nm	Fit at 295 nm	ssq
3			0.207
4			0.109

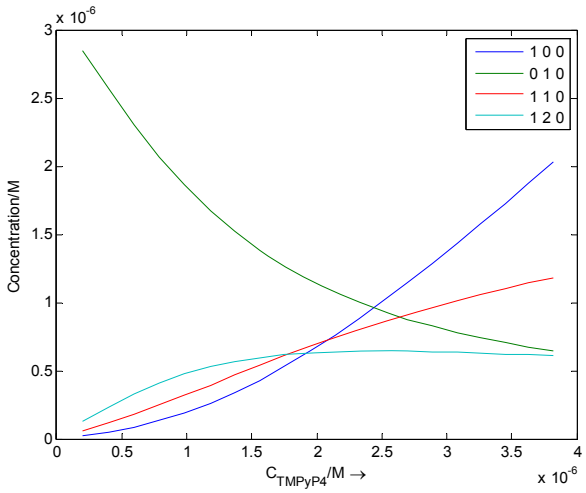
4. CD spectra of C3TT2 and C3TA2 in absence and in presence of TMPyP4 at several pH values. (a) C_{C3TT2} = 2.6 μM, ratio C_{TMPyP4}:C_{C3TT2} = 3.5; (b) C_{C3TT2} = 1.9 μM, ratio C_{TMPyP4}:C_{C3TT2} = 1.6; (c) C_{C3TA2} = 3.0 μM, ratio C_{TMPyP4}:C_{C3TA2} = 3.5; (d) C_{C3TA2} = 2.8 μM, ratio C_{TMPyP4}:C_{C3TA2} = 2.0; (3) C_{C3TA2} = 4.7 μM, ratio C_{TMPyP4}:C_{C3TA2} = 2.0. Other experimental conditions: 25°C, 10 mM acetate or phosphate buffer, 150 mM KCl.

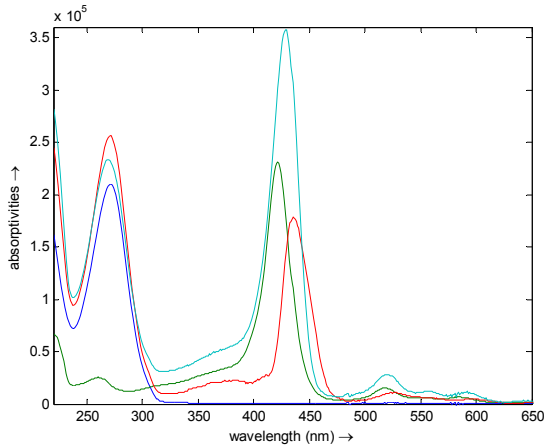


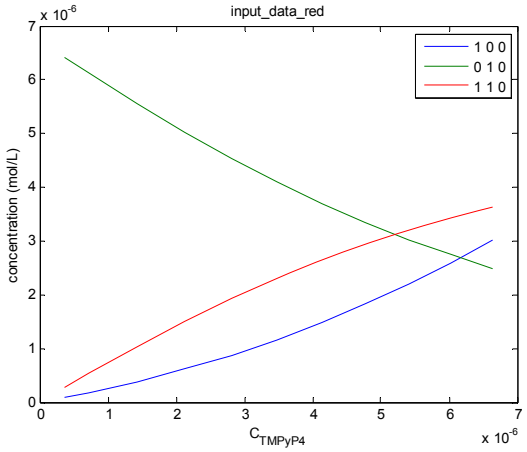
5. Mole-ratio experiments at pH 4.0

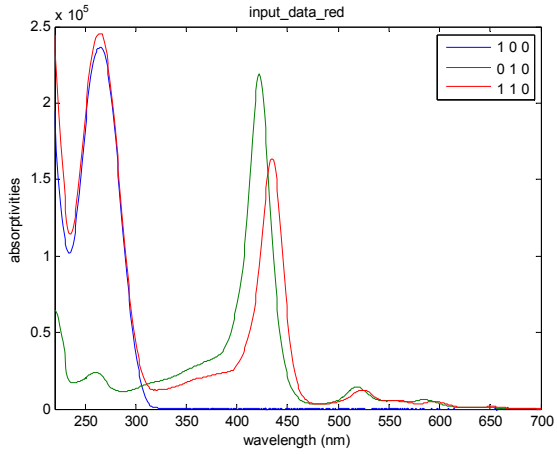
5.1 Interaction of TMPyP4 with the considered sequences studied by mole ratio experiments at pH 4.0.

Distribution diagrams (calculated according to the values of equilibrium constants shown in Table 3), and pure molecular absorption spectra for C3TT2 (1st row) and C3TA2 (2nd row).





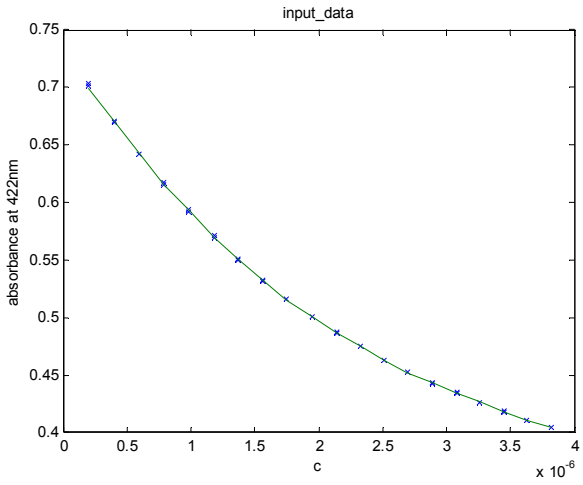




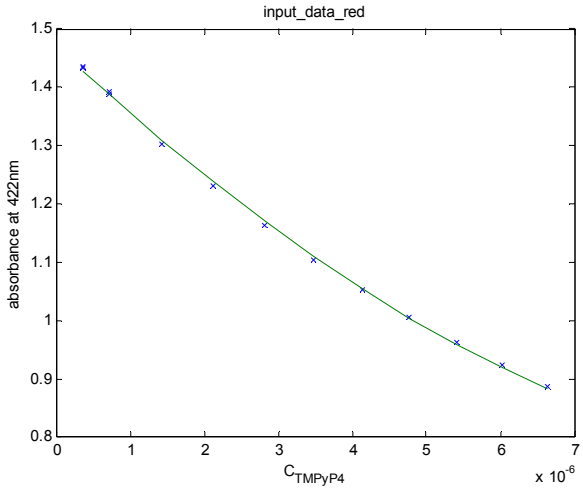
5.2 Fits at 422 nm for mole-ratio experiments at pH 4.0.

Symbols represent the experimental absorbance values measured at 422 nm. Continuous lines represent the fit obtained with the model shown in Table 3.

C3TT2



C3TA2

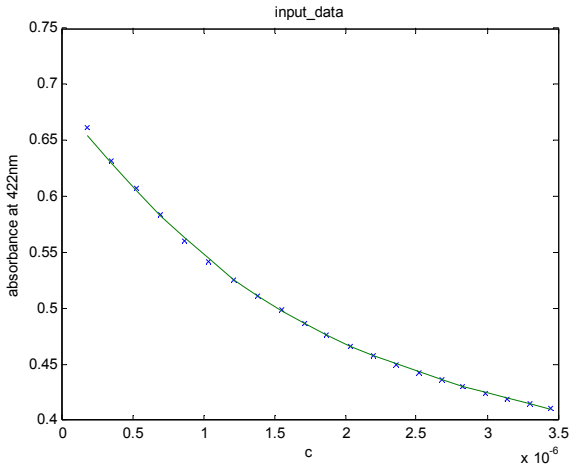


6. Mole-ratio experiments at pH 5.2

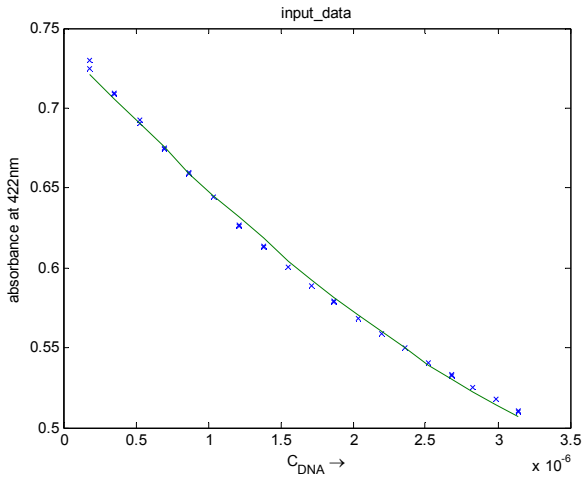
6.1 Fits at 422 nm for mole-ratio experiments at pH 5.0.

Symbols represent the experimental absorbance values measured at 422 nm. Continuous lines represent the fit obtained with the model shown in Table 3.

C3TT2



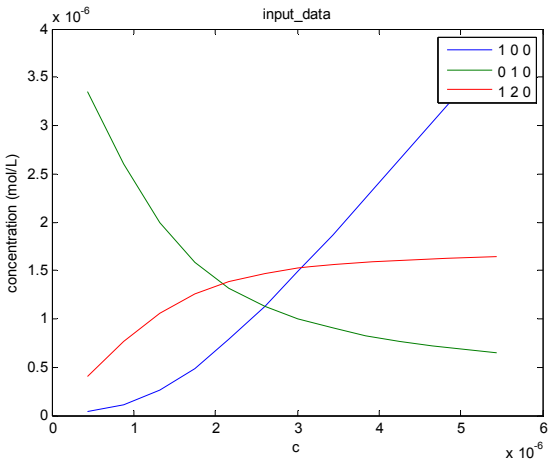
C3TA2

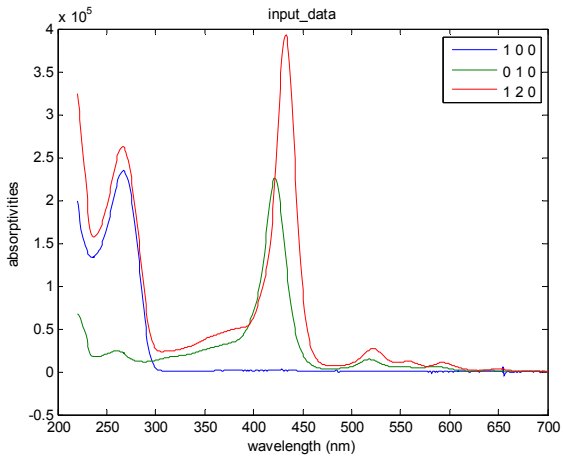


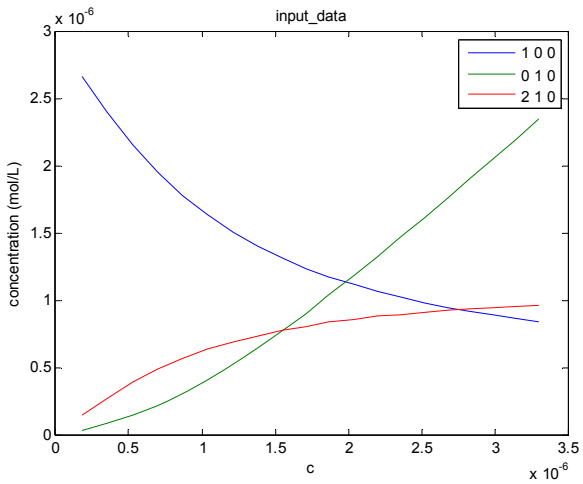
7. Mole-ratio experiments at pH 7.0

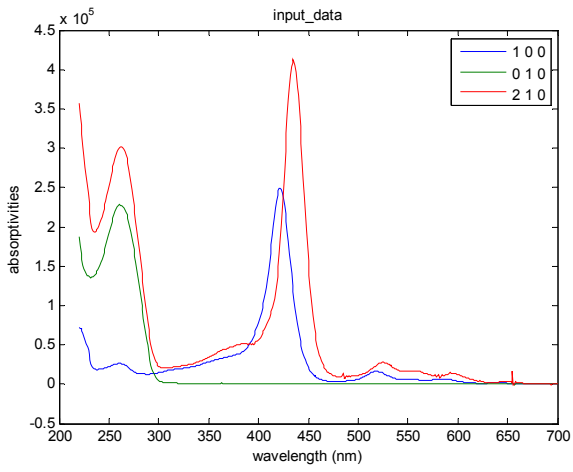
7.1 Interaction of TMPyP4 with the considered sequences studied by mole ratio experiments at pH 7.

Distribution diagrams (calculated according to the values of equilibrium constants shown in Table 3) and pure molecular absorption spectra for C3TT2 (1st row) and C3TA2 (2nd row).





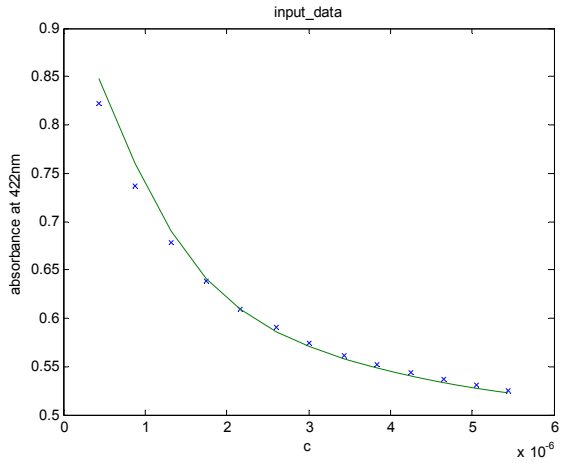




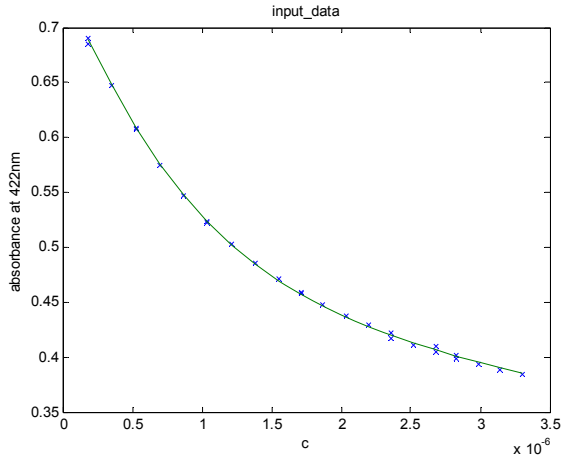
7.2 Fits at 422 nm for mole-ratio experiments at pH 7.0.

Symbols represent the experimental absorbance values. Continuous lines represent the fit obtained with the model shown in Table 3.

C3TT2



C3TA2



Bibliography

[1] Gueron M., Leroy J.L., The i-motif in nucleic acids, Current Opinion in Structural Biology 10 (2000) 326-331.

[2] Peng Y., Wang X., Xiao Y., Feng L., Zhao C., Ren J., Qu X., i-Motif Quadruplex DNA-Based Biosensor for Distinguishing Single- and Multiwalled Carbon Nanotubes, Journal of the American Chemical Society 131 (2009) 13813-13818.

[3] Sharma J., Chhabra R., Yan H., Liu Y., pH-driven conformational switch of "i-motif'' DNA for the reversible assembly of gold nanoparticles, Chemical Communications (2007) 477-479.

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

[5] Guo K., Pourpak A., Beetz-Rogers K., Gokhale V., Sun D., Hurley L.H., 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.

[6] Simonsson T., Pribylova M., Vorlickova M., A nuclease hypersensitive element in the human c-myc promoter adopts several distinct i-tetraplex structures, Biochemical and Biophysical Research Communications 278 (2000) 158-166.

[7] Kumar P., Verma A., Maiti S., Gargallo R., Chowdhury S., Tetraplex DNA transitions within the human c-myc promoter detected by multivariate curve resolution of fluorescence resonance energy transfer, Biochemistry 44 (2005) 16426-16434.

[8] Khan N., Avino A., Tauler R., Gonzalez C., Eritja R., Gargallo R., Solution equilibria of the i-motif-forming region upstream of the B-cell lymphoma-2 P1 promoter, Biochimie 89 (2007) 1562-1572.

[9] Xu Y., Sugiyama H., Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb), Nucleic Acids Research 34 (2006) 949-954.

[10] Saxena S., Bansal A., Kukreti S., Structural polymorphism exhibited by a homopurine·homopyrimidine sequence found at the right end of human c-jun protooncogene, Archives of Biochemistry and Biophysics 471 (2008) 95-108.

[11] Fedoroff O.Y., Rangan A., Chemeris V.V., Hurley L.H., Cationic porphyrins promote the formation of i-motif DNA and bind peripherally by a nonintercalative mechanism, Biochemistry 39 (2000) 15083-15090.

[12] Cimino-Reale G., Pascale E., Alvino E., Starace G., D'Ambrosio E., Long Telomeric C-rich 5â€²-Tails in Human Replicating Cells, Journal of Biological Chemistry 278 (2003) 2136-2140.

[13] Zhou J., Wei C., Jia G., Wang X., Feng Z., Li C., Human telomeric G-quadruplex formed from duplex under near physiological conditions: Spectroscopic evidence and kinetics, Biochimie 91 (2009) 1104-1111.

[14] Gehring K., Leroy J.-L., Gueron M., A tetrameric DNA structure with protonated cytosine-cytosine base pairs, Nature 363 (1993) 561-565.

[15] Cortes A., Huertas D., Fanti L., Pimpinelli S., Marsellach F.X., Pina B., Azorin F., DDP1, a single-stranded nucleic acid-binding protein of Drosophila, associates with pericentric heterochromatin and is functionally homologous to the yeast Scp160p, which is involved in the control of cell ploidy, EMBO J 18 (1999) 3820-3833.

[16] Lacroix L., Lienard H., Labourier E., Djavaheri-Mergny M., Lacoste J., Leffers H., Tazi J., Helene C., Mergny J.-L., Identification of two human nuclear proteins that recognise the cytosine-rich strand of human telomeres in vitro, Nucleic Acids Research 28 (2000) 1564-1575.

[17] Marsich E., Xodo L.E., Manzini G., Widespread presence in mammals and high binding specificity of a nuclear protein that recognises the single-stranded telomeric motif (CCCTAA)n, European Journal of Biochemistry 258 (1998) 93-99.

[18] Zeng Z.-X., Zhao Y., Hao Y.-H., Tan Z., Tetraplex formation of surface-immobilized human telomere sequence probed by surface plasmon resonance using single-stranded DNA binding protein, Journal of Molecular Recognition 18 (2005) 267-271.

[19] Huppert J.L., Quadruplexes in the Genome, in: Neidle S. and Balasubramanian S. (Eds.), Quadruplex Nucleic Acids, The Royal Society of Chemistry, Cambridge, 2006, pp. 208-223.

[20] Kanaori K., Shibayama N., Gohda K., Tajima K., Makino K., Multiple four-stranded conformations of human telomere sequence d(CCCTAA) in solution, Nucleic Acids Research 29 (2001) 831-840.

[21] Ahmed S., Kintanar A., Henderson E., Human Telomeric C-Strand Tetraplexes, Nature Structural Biology 1 (1994) 83-88.

[22] Leroy J.L., Gueron M., Mergny J.L., Helene C., Intramolecular Folding of a Fragment of the Cytosine-Rich Strand of Telomeric DNA into an I-Motif, Nucleic Acids Research 22 (1994) 1600-1606.

[23] Manzini G., Yathindra N., Xodo L.E., Evidence for Intramolecularly Folded I-DNA Structures in Biologically Relevant Ccc-Repeat Sequences, Nucleic Acids Research 22 (1994) 4634-4640.

[24] Phan A.T., Leroy J.L., Intramolecular i-motif structures of telomeric DNA, Journal of Biomolecular Structure & Dynamics (2000) 245-251.

[25] Kaushik M., Suehl N., Marky L.A., Calorimetric unfolding of the bimolecular and i-motif complexes of the human telomere complementary strand, d(C3TA2)4, Biophysical Chemistry 126 (2007) 154-164.

[26] Zhao Y., Zeng Z.-x., Kan Z.-y., Hao Y.-h., Tan Z., The Folding and Unfolding Kinetics of the i-Motif Structure Formed by the C-Rich Strand of Human Telomere DNA, ChemBioChem 6 (2005) 1957-1960.

[27] Kaushik M., Prasad M., Kaushik S., Singh A., Kukreti S., Structural transition from dimeric to tetrameric i-motif, caused by the presence of TAA at the 3′-end of human telomeric C-rich sequence, Biopolymers 93 (2009) 150-160.

[28] Mergny J.L., Lacroix L., Han X.G., Leroy J.L., Helene C., Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an I-DNA Motif, Journal of the American Chemical Society 117 (1995) 8887-8898.

[29] Bucek P., Jaumot J., Aviñó A., Eritja R., Gargallo R., 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, Chemistry – A European Journal 15 (2009) 12663-12671.

[30] Jaumot J., Eritja R., Tauler R., Gargallo R., Resolution of a structural competition involving dimeric G-quadruplex and its C-rich complementary strand, Nucleic Acids Research 34 (2006) 206-216.

[31] Allawi H.T., SantaLucia J., Thermodynamics and NMR of Internal G·T Mismatches in DNA, Biochemistry 36 (1997) 10581-10594.

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

[33] Jaumot J., Gargallo R., de Juan A., Tauler R., A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, Chemometrics and Intelligent Laboratory Systems 76 (2005) 101-110.

[34] Tauler R., Multivariate curve resolution applied to second order data, Chemometrics and Intelligent Laboratory Systems 30 (1995) 133-146.

[35] Bloomfield V.A., Crothers D.M., Tinoco I.J., Nucleic acids. Structures, properties, and functions, University Science Books, Sausalito, CA, 2000.

[36] Broido M.S., Kearns D.R., 1H NMR evidence for a Left-Handed Helical Structure of Poly(ribocytidylic acid) in Neutral Solution, Journal of the American Chemical Society 104 (1982) 5207-5216.

[37] Phan A.T., Gueron M., Leroy J.L., The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere, Journal of Molecular Biology 299 (2000) 123-144.

[38] Nonin-Lecomte S., Leroy J.L., Structure of a C-rich strand fragment of the human centromeric satellite III: A pH-dependent intercalation topology, Journal of Molecular Biology 309 (2001) 491-506.

[39] Dawson R.M.C., Elliott D.C., Elliott W.H., Jones K.M., Data for Biochemical Research, Clarendon Press, Oxford, 1989.

[40] Guedin A., Alberti P., Mergny J.-L., Stability of intramolecular quadruplexes: sequence effects in the central loop, Nucl. Acids Res. 37 (2009) 5559-5567.

[41] Weil J., Min T.P., Yang C., Wang S.R., Sutherland C., Sinha N., Kang C.H., Stabilization of the i-motif by intramolecular adenine-adenine-thymine base triple in the structure of d(ACCCT), Acta Crystallographica Section D-Biological Crystallography 55 (1999) 422-429.

[42] Phan A.T., Mergny J.L., Human telomeric DNA: G-quadruplex, i-motif and watson-crick double helix, Nucleic Acids Research 30 (2002) 4618-4625.