ñSolution equilibria of the i-motif-forming region upstream of the B-cell lymphoma-2 P1 promoterò Khan, N., Aviñó, A., Tauler, R., González, C., Eritja, R., Gargallo, R. Biochimie, 89(12), 1562-1572 (2007). doi: 10.1016/j.biochi.2007.07.026

Solution Equilibria of the *i*-motif-forming Region Upstream of the B-Cell Lymphoma-2 P1 Promoter

*Nasiruddin Khan[1], Anna Aviñó[2], Romà Tauler[3], Carlos González[4], Ramon Eritja[2], Raimundo Gargallo[1]**

1. Department of Analytical Chemistry, University of Barcelona, Diagonal 647, E-08028 Barcelona, Spain

2. Institut de Biologia Molecular de Barcelona, CSIC, Jordi Girona 18-26, E-08034 Barcelona, Spain

3 Department of Environmental Chemistry, IIQAB-CSIC, Jordi Girona 18-26, E-08034 Barcelona, Spain

4 Instituto de Quimica-Fisica Rocasolano, CSIC, Serrano 119, E-28006 Madrid, Spain

* Corresponding author

fax: (34) 934021233

e-mail: raimon_gargallo@ub.edu

Abstract

The 5ô-end of the P1 promoter of the B-cell lymphoma-2 (bcl-2) gene contains a highly guanine-cytosine-rich region, which has a role in the regulation of bcl-2 transcription. Whereas the guanine-rich region has been the focus of recent studies, little attention has been paid to the cytosine-rich strand. Here we examine the structural transitions of the cytosine-rich sequence by means of acid-base, mole-ratio and melting experiments monitored by molecular absorption, circular dichroism, and NMR spectroscopies. Two intramolecular i-motif structures have been detected in the pH range 2 - 7, with maximal formation at pH 4 and 6, respectively. At pH 7.6 the majority species has been associated with a hairpin involving Watson-Crick base pairs. Upon addition of the quadruplex-interacting ligand TmPyP4, both bcl-2c structures at pH 6.1 and 7.6 yield identical interaction species with stoichiometries 1:1 and 1:2 (DNA : ligand) and logarithms of formation constant 12.4 ±0.2 and 11.7 ± 0.1, respectively. The initial i-motif structure at pH 6.1 is lost upon interaction with TmPyP4.

Keywords: Spectroscopic methods, Thermodynamics, Nucleic acids, Drugs, Chemometrics

1. Introduction

Bcl-2 (B-cell Lymphoma 2) is a potent oncoprotein that plays an essential role in cell survival and functions as an inhibitor of cell apoptosis [1]. The human bcl-2 gene contains a guanine-cytosine-rich region upstream of the P1 promoter that is implicated in the regulation of gene expression. In this region, one of the strands of the DNA double helix is guanine-rich, and the other one is cytosine-rich. Recently, the guanine-rich strand has been studied in an attempt to determine its structure and stability [2-4]. Guanine-rich DNA strands can form complex structures *in vivo* known as G-quadruplexes. In addition, some authors have studied the interaction of this guanine-rich strand with small organic ligands.

Although the guanine-rich strand has been widely studied, the cytosine-rich strand has received little attention. Cytosine-rich DNA strands can form a cytosine^{+}·cytosine tetraplex, known as an i-motif, at low or even neutral pH [5-8]. This structure consists of four strands arranged in two parallel duplexes ñzippedò together in antiparallel orientation (Figure 1a). The i-motif is the only known nucleic acid structure involving systematic base intercalation. Mainly through NMR studies, it has been established that the same cytosine-rich strands of oligodeoxynucleotides can form intercalated structural entities of the i-motif that differ essentially in their intercalation and loop topology.

Recent interest has been shown in the study of i-motif structures due to their potential application in nanotechnology [9-12] and their possible roles in gene transcription [13]. Although low pH favours the formation of i-motif structures and they have not been reported *in vivo*, the sequences with the potential to form i-motifs are frequent in the human genome. To date, i-motif structures have been postulated in both centromeric and telomeric regions of human chromosomes, as well as in the promoter region of several genes [14].

In addition to the lack of studies of the cytosine-rich strand in the region upstream the bcl-2 P1 promoter, few authors described an analytical approach to biophysical problems of this kind. For example, it is difficult to find studies of the acid-base properties of these DNA strands. In the case of the cytosine-rich strand in the bcl-2 gene, knowledge of such properties would be of value since pH

strongly influences the conformational equilibria and the formation of the i-motif structures.

Here we describe the acid-base properties and the interaction of an i-motif structure with the ligand 5,10,15,20-tetrakis(*N*-methyl-4-pyridyl)-21*H*,23*H*-porphyrin (TmPyP4, Figure 1b). The interaction of this ligand with G-quadruplex structures has been studied extensively. Reports of i-motif interaction with small organic ligands are scarce [15-16], perhaps because of the belief that such structures can form only at relatively low pH. However, recent studies indicate that they may also be present at neutral pH [17-18], giving a new insight into their possible biological role.

2. Experimental

2.1. Reagents

24-bases long oligonucleotide with sequence 5ôCCC GCC CCC TTC CTC CCG CGC CCG-3ô (bcl-2c) corresponding to the middle region of the bcl-2 NHE region was prepared on 1 µmol scale using standard 2-cyanoethyl phosphoramidites (Cruachem Ltd.). Syntheses were performed using an automatic DNA synthesizer (Applied Biosystems Mod. 392). Sequences were deprotected using standard protocols (concentrated ammonia, 55 ^{o}C, and overnight). After deprotection, oligonucleotides were purified using oligonucleotide purification cartridges following the protocol suggested by the manufacturer. Finally, purified oligonucleotides were desalted using Sephadex G-25 columns. The length and homogeneity of oligonucleotides was checked by 8M urea PAGE stained with stains-all.

DNA strand concentration was determined by UV absorbance measurements at 260 nm and 80 ^{o}C using a molar absorptivity of 186,000 M^{-1} cm^{-1}. This value was obtained using the nearest-neighbour method [19]. TmPyP4 was purchased from Sigma-Aldrich and used without further purification. A 190 µM stock solution of TmPyP4 was prepared in water, stored at -20 ^{o}C and diluted to working concentration immediately before use. A value of 226,000 (at 424 nm) for the extinction coefficient was used in the calculations. KCl, $MgCl_2$Å6H_2O, KH_2PO_4, K_2HPO_4, HAcO, HCl and NaOH (a.r.) were purchased to Panreac (Spain). MilliQ ® water was used in all experiments.

2.2. Instrumental

Absorbance spectra were recorded on an Agilent HP8453 diode array spectrophotometer. Temperature was controlled by an 89090A Agilent peltier device. Hellma quartz cuvettes (path length of 1.0 cm, 1500 µl or 3000 µl volume) were used. CD spectra were recorded on a Jasco J-810 spectropolarimeter equipped with Julabo F-25/HD temperature control unit. Hellma quartz cylindrical cuvette (path length of 1.0 cm, 650 µl volume) was used. pH was measured with an Orion SA 720 pH/ISE meter and micro combination pH electrode (Thermo). NMR spectra were acquired in a Bruker Advance spectrometer operating at 800 MHz and equipped with a crioprobe. Water suppression was achieved by the inclusion of a WATERGATE [20] module in the pulse sequence prior to acquisition.

2.3. Procedure

Acid-base titrations were monitored in-line (taking advantage of the stirrer incorporated at the cell holder of the Agilent instrument) or at-line (in the case of the CD instrument). Other experimental conditions were 25 ºC and 150 mM ionic strength (147 mM K^+, as KCl, and 1 mM Mg^{2+}, as $MgCl_2$Å6H_2O). Titrations were carried out by adjusting the pH of solutions containing the oligonucleotide, ligand, or mixtures of both by addition of small volumes of HCl or NaOH stock solutions. Absorbance or CD spectra were recorded pH stepwise.

Melting experiments were done in the temperature range from 15 ºC to 80 ºC with a linear temperature ramp of 0.5 ºC/min. Absorbance and CD spectra were recorded every 1 ºC and 3 ºC, respectively. Buffer solutions were 10 mM acetate (for experiments in which pH ranged from 4.0 to 5.6) or phosphate (for experiments which pH ranged from 6.1 to 6.8), 1 mM Mg^{2+} and adjusted to 150 mM ionic strength with KCl. Each sample was allowed to equilibrate at the initial temperature for 30 minutes before the melting experiment was started.

Mole ratio experiments were done by means of progressive additions of small volumes of TmPyP4 stock solution to an oligonucleotide solution (initial volume of 2300 µl). The additions were done at time intervals of 10 minutes. Spectroscopic monitoring of these experiments was done in-line. Other

experimental conditions were 25 ºC, pH 6.1, 6.6 or 7.8, 150 mM ionic strength. Buffer conditions were the same as those described for melting experiments.

2.4. Analysis of multivariate spectroscopic data

Spectra recorded in acid-base titrations, melting experiments or mole ratio studies were arranged in a table or data matrix **D**. The dimensions of this matrix were *n* rows x *m* columns, where *n* were the spectra recorded at successive pH, temperature or c_{TmPyP4} : c_{bcl-2c} ratio values (depending on whether data from an acid-base titration, a melting or a mole ratio experiment were being analyzed), and *m* was the number of wavelengths measured. Two multivariate data analysis methods were applied to analyze matrix **D**: the hard modelling-based EQUISPEC program [21] and the soft modelling-based MCR-ALS method [22]. Both approaches calculate concentration profiles and pure spectra for all spectroscopically active species present in the system from the decomposition of the experimental data matrix **D** according to equation:

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

where **C** and $\mathbf{S}^{\mathbf{T}}$ are respectively, the data matrices of the concentration profiles and of the pure spectra for each one of the spectroscopically active species or conformations present in the experiment. **E** is the data matrix of residuals not explained by the proposed species or conformations in **C** and $\mathbf{S}^{\mathbf{T}}$.

Decomposition of data matrix **D** according to equation 1 with EQUISPEC requires the fulfilment of a previously proposed simple chemical model. This model is defined by: (1) the stoichiometries of all species involved in the equilibria considered, and (2) approximate values of the equilibrium constants. EQUISPEC assumes that the value of the equilibrium constants do not vary upon the advance of the reaction considered. Logarithms of equilibrium constants calculated with EQUISPEC are given as their weighted means with twice their standard errors (units of the least significant digit).

Hard modelling-based programs, like EQUISPEC, are especially adequate for the study of chemical equilibria involving monomers (or TmPyP4) or short DNA sequences which do not show secondary

effects related to polymeric structures, like polyelectrolytic effects, polyfunctional effects or conformational changes. In contrast, for large DNA sequences or when analyzing data from melting experiments, the equilibrium constants vary upon advance of the reaction or conformational change considered. In these cases, application of hard modelling-based methods is challenging since it is difficult to construct a simple species model describing the spectral behaviour observed. In these cases, the multivariate data can be analysed by soft modelling-based methods, which do not require compliance of any species model. Multivariate Curve Resolution using Alternating Least Squares (MCR-ALS) is a soft modelling-based method that is widely applied in the study of acid-base and conformational transitions of DNAs and proteins [23-25]. As an example, the analysis of spectra recorded during a melting experiment with MCR-ALS has been included in the Supplementary Material.

All EQUISPEC and MCR-ALS calculations were performed using MATLAB routines (version 6, The Mathworks Inc, Natick, MA). MCR-ALS codes and tutorials are freely available at the URL: www.ub.es/gesq/mcr/mcr.htm.

2.5. Calculation of thermodynamic data

From the MCR-ALS-resolved concentration profiles (matrix **C** in Equation 1) for each melting experiment, the free energy of i-motif folding was evaluated, where both enthalpy ΔH^0 and entropy ΔS^0 are assumed to be independent of temperature, using the Gibbs equation:

$$\Delta G^0 = -RT\ln(K_{eq}) = \Delta H^0 - T\Delta S^0 \qquad \text{(Equation 2)}$$

whereby:

$$\ln(K_{eq}) = -\frac{\Delta H^0}{R}\frac{1}{T} + \frac{\Delta S^0}{R} \qquad \text{(Equation 3)}$$

The folding equilibrium constant (K_{eq}) was calculated from the resolved concentration profiles in matrix **C** for the folded (i-motif) and unfolded conformations, and expressed as:

$$K_{eq} = \frac{[\text{folded conformation}]}{[\text{unfolded conformation}]} \qquad \text{(Equation 4)}$$

The plot of $\ln(K_{eq})$ against 1/T gives a straight line under the two-state intramolecular transition approximation, whose slope and intercept were used to evaluate the enthalpy and entropy of transition, respectively.

3. Results

3.1. Solution equilibria of the i-motif forming region

The bcl-2c sequence exhibits CD and absorbance spectra (Figure 2a) with similar features at slightly lower pH values than those previously reported in the study of i-motif structures [14, 18]. CD spectra in the pH range 4-6 show positive and negative bands around 280 nm and 250 nm, respectively. Absorbance spectra show a shift of the maximum from 266 (pH 7.2) to 277 nm (pH 2.0). Data analysis with MCR-ALS according to Equation 1 yields the results summarized in Figure 2b-c. Four spectroscopically active acid-base species were resolved with pH transition midpoints 2.7 ± 0.2, 5.2 ± 0.2 and 6.5 ± 0.1 (matrix $\boldsymbol{C}$), all of them are related to protonation of cytosine bases (deprotonation of guanine and thymine bases occurs at pH values higher than 9). Structural information about the resolved acid-base species can be obtained from the calculated concentration profiles and pure spectra (matrix $\mathbf{S}^{\mathbf{T}}$). The species predominantly present at neutral pH values has been related to a structure showing a certain degree of base stacking [26-27]. In this species, cytosine bases are in neutral form, i.e., deprotonated at the N3 position. The species predominantly present at pH 2 has been related bcl-2c showing fully protonated cytosines. This structure is a random coil as denoted by both CD and absorbance signatures. Two more species (labelled I and II) are also present in the pH range 2-7, which has been related to i-motif structures based on the position of CD bands and on the decrease in absorption maxima and red-shifts from 266 to 271 nm [18]. In these structures, hemiprotonated cytosine^{+}·cytosine pairs are formed, which stabilize the i-motif. The existence of the second i-motif structure in addition to that mainly formed at pH around 6 may be related to the different number of

cytosines present in the second track and/or to the three isolated cytosines not included in any of the four tracks. Therefore, species II would reflect the protonation of bases not involved in the previous formation of i-motif structure in species I. No other transition was observed, probably because of the lack of adenine bases (which show acidebase equilibria at pH near 3.5) in the loops, as in the case of the i-motif formed within the human c-myc promoter [17] and on the human centromeric satellite III [28].

NMR experiments provided additional structural information about the molecularity and nature of the species proposed. ^{1}H NMR experiments were recorded at different temperatures and oligonucleotide concentrations, and at pH ranging from 5 to 7. In all the experimental conditions, the NMR spectra exhibited broad signals, making the determination of the structure of bcl-2c impossible. However, structural information was obtained from the spectra. Thus the presence of imino resonances at 13.2 and 15.5 ppm and the dependence of their relative intensities upon acidification indicates the presence of an equilibrium involving two bcl-2c structures, one with cytosine^{+}Åcytosine base pairs, and the other one with Watson-Crick base pairs (Figure 3). Also, NOESY spectra recorded at pH 7.0 revealed cross-peaks between guanine imino and cytosine amino protons. These data, together with the melting behaviour of these signals (see below), suggest that the majority species at neutral pH is a hairpin with several Watson-Crick base pairs.

The proposed acid-base concentration profile was tested by melting experiments. As an example, Figure 4a shows the absorbance data recorded during the experiment at pH 6.1. Spectral features, like the sharp absorbance decrease at 295 nm (inset), are typical for the melting of an i-motif structure at pH higher than the pK_a of the isolated cytosine molecule (around 4.3-4.5) [29-30]. Analysis of matrix $\boldsymbol{D}$ with MCR-ALS according to Equation 1 yields the resolved concentration profile ($\boldsymbol{C}$) and pure spectra ($\boldsymbol{S}^{T}$) (Figure 4b, 4c). In this case, three components were necessary to explain the experimental data within the experimental error. Only two of them, however, are directly involved in the i-motif disruption. The first component, present at low temperatures, is the initial bcl-2c species, mainly as an i-motif structure according to Figure 2b. The second component is the hairpin produced from the melting of the i-motif, showing Watson-Crick base pairing and some degree of base stacking. The

estimated T_m value (transition midpoint between these two conformations) was 36 ± 1 ^oC, which agrees with previous values determined for similar cytosine-rich DNA sequences [29, 31] at similar pH and ionic strength values. Spectral signatures at 295 nm for the pure spectra of these two conformations agree with experimental observations. Finally, the third component, which is the predominant conformation at high temperatures, is a random coiled single strand. The presence of this wide transition has also been observed by other authors for similar sequences [31]. Temperature-dependent CD spectroscopy was used to confirm these findings. Decrease in molar ellipticity was observed at both the positive and negative bands characteristic of i-motif due to thermal denaturation (Supplementary material). At high temperatures the CD spectra are indicative of the random-coiled single strand.

A similar procedure was applied for the analysis of melting data at lower pH values (Table 1 and Supplementary Material for the description of the analysis of absorbance data recorded along the melting experiment at pH 5.6). The melting of the i-motif structure formed by the bcl-2c sequence was strongly pH-dependent, with an optimal pH around 4.3. In the pH range 6.6 to 4.5, T_m values were a linear function of pH. Previous studies have also found these relationships and shown that T_m was highest at pH around the pKa of cytosine, depending on experimental conditions such as ionic strength [29,31]. The shape of the transition at 295 nm was also pH dependent. At pH values above the pK_a of the isolated cytosine, the absorbance at 295 nm decreased upon disruption of the i-motif structure, whereas at pH values below pK_a, the absorbance increased. Around pH = pK_a, the absorbance at 295 nm was almost independent of temperature. These observations were already pointed out by Mergny et al. [29] for the 5ô-TCC TCC TTT CCT CCT-3ô sequence, and are in agreement with the requirement for protonation of half the cytosines in the bcl-2c sequence.

Melting transitions were monitored by UV at bcl-2c concentrations ranging from 1 to 8 μM, and by NMR at concentrations ranging from 20 to 600 μM. Throughout this concentration range T_m did not change significantly, indicating that the thermal denaturation of bcl-2c is a monomolecular process. From the concentration profiles resolved, enthalpy and entropy for the folding reaction were calculated (Table 1). In general, thermodynamic values agree with those previously described for similar

sequences [29]. Folding at 20 ^{o}C was accompanied by favourable free energy changes, resulting from favourable enthalpy and unfavourable entropy terms. As in the case of hyperchromicity, the slope of the transition, and thus the ΔH^0 of the reaction, was pH-dependent (Table 1). The formation of this structure was strongly enthalpy driven at pH around 5.6, as denoted by the maximum slope at this pH. Several authors have related this favourable enthalpy term with the electrostatic stabilization as a result of cytosine$^+$·cytosine base pairing and favourable interactions between the protonated bases and the sugar phosphates [32,33]. As previously measured, the formation of each cytosine$^+$·cytosine base pair leads to a decrease in enthalpy of 9-12 kcal mol^{-1} [29,32], thus suggesting that the number of cytosine$^+$·cytosine intercalated units could be 5-6, since the observed enthalpy change is around 50 - 60 kcal mol^{-1}. Unfavourable entropy changes arise largely from the loss of translational and rotational degrees of freedom by DNA monomers and counterions. According to the calculated ΔG^0 values, formation of the folded conformation seems to be slightly more favoured in the pH region around 4.5 compared to the formation of folded conformation at the other pH values. This agrees with the dependence of T_m on pH, and also with the resolved concentration profile for the i-motif II (Figure 2b) (see Section 4).

3.2. Interaction with TmPyP4

The interaction of the bcl-2c sequence with the quadruplex-interacting drug TmPyP4 was investigated. First, the acid-base properties of the isolated drug were characterized in the pH range 3.0 - 8.0 by means of a molecular absorption monitored acid-base titration. Data analysis (see Supplementary Material) yields a pKa = 5.3 ± 0.1, which probably corresponds to the protonation of pyrrolenine nitrogen atoms. The pure spectra corresponding to the protonated and deprotonated species are practically equal, the only noticeable difference being a slight red shift of the Soret band at 422 nm.

To obtain an initial picture of the interaction of bcl-2c and TmPyP4 according to pH, an acid-base titration of a mixture (1:1 C_{TmPyP4}:C_{bcl-2c} ratio) was done and the recorded spectral data were analyzed with MCR-ALS using two separate approaches. First, spectra recorded during the titration were analyzed according to Equation 1, in a similar way to the analysis of data recorded during the titration

of bcl-2c. This approach, the results of which are included in the Supplementary Material, did not allow complete resolution of the contribution of all acid-base and interaction species (if present). Therefore, a second approach involving the simultaneous analysis of spectral data recorded during the acid-base titrations of bcl-2c, of TmPyP4 and of the 1:1 mixture was tested. This approach has been shown to be useful for the analysis of such data sets [34]. The resolved concentration profiles and pure spectra (Figure 5) showed the contribution of all spectroscopically active species in the mixture. According to these plots, only one interaction species was present and resolved in the 1:1 mixture. At neutral pH values, the complex coexists with the proposed hairpin (continuous black line) and free neutral TmPyP4 (magenta). At pH below 6.7 (approximately) the formation of i-motif I (dashed black line) is observed, as in the titration of bcl-2c. At pH values lower than 6, protonation produces the disruption of the complex into protonated TmPyP4 (yellow) and a mixture of i-motif II (dotted black line) and random coiled single strand bcl-2c (dashed-dotted black line).

Figure 6 shows the NMR spectra of TmPyP4 : bcl-2c mixtures recorded at pH 6.0. The TmPyP4 signals exhibit significant line broadening, confirming the interaction of TmPyP4 with the oligonucleotide. Upon addition of TmPyP4, the intensity of the signal at 13.2 ppm decreased. The same effect was observed at pH 7, suggesting that the presence of TmPyP4 disrupts the structure of bcl-2c. These NMR spectra are consistent with TmPyP4 interacting with an unpaired single stranded structure of bcl-2c.

To obtain further insight into TmPyP4 binding to the bcl-2c sequence, absorbance spectra were recorded upon titration of the bcl-2c sequence with the ligand at several pH values (7.8, 6.6 and 6.1) and temperature 25 ^{o}C (Figure 7). Multivariate analysis of titration data with EQUISPEC led us to propose a chemical model defined by a set of species and stability constants, together with the resolution of pure spectrum for each species. The best results for experiments carried out at pH 6.1 and 6.6 were obtained when one interaction species were considered with stoichiometry 1:2 (DNA:TmPyP4) and with binding constant $\log\beta = 12.4 \pm 0.2$. Fig. 7 shows the results obtained at pH 6.1, where i-motif I is the majority species in absence of TmPyP4. The resolved concentration profiles clearly indicate that, at the experimental conditions of the acid-base titration (1.1:1 C_{TmPyP4}:$C_{bcl\text{-}2c}$

ratio), almost 50% of bcl-2c is interacting with TmPyP4. The resolved pure CD and absorbance spectra reflect the observed experimental trends. Hence, the complex shows a induced negative CD band at 445 nm whereas the intensity and position of the band at 280 nm clearly indicates the disruption of the initial structure of DNA. Similar results were obtained from two independent mole ratio experiments carried out at pH 7.6, where hairpin is the initial majority species in absence of TmPyP4 (Supplementary Material). In this case, however, the value of the logarithm of the formation constant is slightly lower (11.7 ± 0.1) than at pH 6.1, where i-motif I is the initial majority species in absence of TmPyP4.

Melting experiments carried out at pH 6.1 and 7.6 for mixtures resulting from mole-ratio studies revealed high stability high stability of the bcl-2c:TmPyP4 complex. The Soret band shifted from 432 to 424 nm, reflecting the release of TmPyP4 from the complex (data not shown). At the end, some TmPyP4 is at the complex as the maximum did not achieve the characteristic position of free TmPyP4 (422 nm). Multivariate analysis of the whole spectral data set yields a distribution profile which explains the experimental findings (Fig. 8). Around 50% of bcl-2c still interacted with TmPyP4 at 80 ^{o}C. Similar results were obtained for the melting at pH 7.8..

4. Discussion

The solution equilibria involving i-motif structures have received little attention because such structures have not been found at biological pH. Recent studies, however, have shown that i-motif structures could occur at neutral pH. In these conditions, the formation of i-motif structures could parallel to the formation of G-quadruplex structures, which received greater attention. The results described herein show that the sequence 5ôCCC GCC CCC TTC CTC CCG CGC CCG-3ô (bcl-2c) forms two i-motif structures at pH between 2 and 7. Only about 5% of the sequence is present as an i-motif structure at pH 7. From the biological point of view, however, this small amount could be relevant

The acid-base concentration profile resolved for the bcl-2c sequence reflects high cooperativity in the

formation of the i-motif. This behaviour could be related to a ózipperô mechanism, in which the protonation of some cytosines, and the subsequent formation of cytosine$^+$·cytosine pairs, promotes the formation of neighbour pairs, leading to the formation of an intramolecular i-motif structure in a narrow pH range. In contrast, protonation of the base pair to yield isolated cytosine$^+$ bases is strongly difficult, as can be observed in the small slope of the concentration profile at pH below 5.6 - 6.3. The resolved concentration profile for the protonation of bcl2-c sequence agrees with results reported for similar sequences. Hence, the shape and transition midpoints shown in Figure 2 are consistent with those determined from NMR data for the sequence d(CCCTA$_2$5mCCCTA$_2$CCCUA$_2$CCCT) [34]. Simonsson et al. determined transition midpoints for the i-motif to neutral single strand transition around pH 6.6 ± 0.1 (22 ^{o}C) [18], depending on the length of the DNAs studied, whereas a pK_a = 6.4 ± 0.1 (20 ^{o}C) was determined for the disruption of the i-motif in the nuclease hypersensitive element of human *c-myc* promoter [32].

The resolved acid-base concentration profile resolved reflects high cooperativity in the formation of the i-motif I in the pH range 6-7. This behaviour could be related to a ózipperô mechanism, in which the protonation of some cytosines, and the subsequent formation of cytosine$^+$·cytosine pairs, promotes the formation of neighbour pairs, leading to the formation of an intramolecular i-motif structure in a narrow pH range. In contrast, protonation of the base pair to yield isolated cytosine$^+$ bases is more difficult, as can be observed in the small slope of the concentration profiles at pH below 6. The resolved concentration profile for the protonation of bcl2-c sequence in the pH range 5.5-7 agrees with results reported for similar sequences. Hence, the shape and transition midpoints shown in Fig. 2 are consistent with those determined from NMR data for the sequence d(CCCTA$_2$5mCCCTA$_2$CCCUA$_2$CCCT) [35]. Simonsson et al. determined transition midpoints for the i-motif to neutral single strand transition around pH 6.6 ± 0.1 (22 ºC) [18], depending on the length of the DNAs studied, whereas a pK_a = 6.4 ± 0.1 (20 ºC) was determined for the disruption of the i-motif in the nuclease hypersensitive element of human *c-myc* promoter [32].

Several authors have already described the presence of more than one species related to i-motif structures [17,18,28]. Simonsson et al. [17] studied several sequences of different length

corresponding to the *c-myc* oncogene and proposed the existence of distinct i-motif structures depending on the number and arrangement of cytosines. Following this study, Kumar et al. studied the 31-mer d(CCC CAC CTT CCC CAC CCT CCC CAC CCT CCC C) from the promoter site of *c-myc* oncogene and determined the presence of two i-motif structures with transitions around pH 5.9 and 6.8 [18]. Again, the existence of these two structures was explained by the different protonation on cytosines, resulting in two conformations showing a different core of cytosine$^+$ ·cytosine pairs. Other authors, however, have explained the presence of more than one i-motif structure because of differences in intercalation topologies, related to adenine protonation in the loops [28]. In our case, where no adenines were present, the presence of a second i-motif can only be explained by the protonation of cytosines not initially involved in the i-motif core. According to thermodynamic data, protonation of these cytosines does not affect the core of 5-6 cytosine$^+$ ·cytosine pairs. However, it could dramatically affect the topology on the loops.

The pH-dependence of the thermal stability of i-motif structures was described by Mergny et al. [29]. The results obtained in the present study agree with those reported by these authors for similar cytosine-rich DNA sequences. There is a reasonably good agreement between the resolved concentration profile and the pH-dependence of T_m (or the calculated ΔG^0). Whereas the resolved concentration profile shows the maximum degree of formation for the i-motif II at pH around 4.0, its relative stability seems to be higher at pH around 4.3. At this pH, practically half the cytosines on the single-strand are protonated. This should be the optimal situation for the i-motif formation, as no net protonation or deprotonation is required.

From the results shown in this work, the i-motif structure at pH 6.1 would be lost upon interaction with TmPyP4. This finding would agree with the preferential binding of TmPyP4 to melted or partially melted regions of DNA [36]. On the other hand, the proposed model apparently disagrees with that of Fedoroff et al. [15], where TmPyP4 binds to the sites symmetrical with respect to the helical axis of the i-motif formed by d(AACCC)$_4$. Porphyrin interaction with fold-over molecules (as the bcl-2c sequence) may differ from that of the tetrameric i-motif due to the influence of the structure of the connecting loops [15]. Very recently, Buscaglia et al. also proposed that 2 mol of TmPyP4 are

bound per mol of oligonucleotide (for an i-motif forming 23 base deoxyoligonucleotide derived from the *c-myc* P1 promoter sequence [16]). Finally, from the results obtained in this work, it seems that there is not a significant difference in binding constant and stoichiometry when TmPyP4 binds to i-motif I (logβ; = 12.4 ± 0.2 at pH 6.1) vs. hairpin (logβ ¼ 11.7 ± 0.1 at pH 7.6).

5. Acknowledgments

We acknowledge two grants from the Spanish *Ministerio de Educación y Ciencia* (projects BFU2004-02048/BMC and CTQ2006-15052-C02-01/BQU). We also thank Veronica Hernando for her contribution in carrying out some experiments.

6. Bibliography

[1] A. Chanan-Khan, Bcl-2 antisense therapy in B-cell malignancies, Blood Rev. 19 (2005) 213 - 221.

[2] J. Dai, C. Ding, R.A. Jones, L.H. Hurley, D. Yang, NMR solution structure of the majority G-quadruplex structure formed in the human BCL2 promoter region, Nucleic Acids Res. 34 (2006) 5133 - 5144.

[3] J. Dai, T.S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R.A. Jones, D. Yang, An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human Bcl-2 promoter region in solution, J. Am. Chem. Soc. 128 (2006) 1096 - 1098.

[4] T.S. Dexheimer, D. Sun, L. H. Hurley, Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the bcl-2P1 Promoter., J. Am. Chem. Soc. 128 (2006) 5404 - 5415.

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

[6] S. Ahmed, A. Kintanar, E. Henderson, Human telomeric C-strand tetraplexes, Nat. Struct. Biol. 1 (1994) 83 - 88.

[7] J. L. Leroy, M. Gueron, J.L. Mergny, C. Hélène, Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif, Nucleic Acids Res. 22 (1994) 1600 - 1606.

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

[9] F. Seela, S. Budow, pH-dependent assembly of DNA-gold nanoparticles based on the i-motif: a switchable device with the potential of a nanomachine, Helv. Chim. Acta 89 (2006) 1978 - 1985.

[10] J. Sharma, R. Chlabra, H. Yan, Y. Liu Y, pH-driven conformational switch of ñi-motifò DNA for the reversible assembly of gold nanoparticles, Chem. Comm. 5 (2007) 477 - 479.

[11] X. Li, Y. Peng, J. Ren, X. Qu, Carboxyl-modified single-walled carbon nanotubes selectively induce human telomeric i-motif formation, Proc. Natl. Acad. Sci USA 103 (2006) 19658 - 19663.

[12] H. B. Ghodke, R. Krishnan, K. Vignesh, G. V. P. Kumar, C. Narayana, Y. Krishnan, The I-tetraplex building block: rational design and controlled fabrication of robust 1D DNA scaffolds through non-Watson-Crick interactions. Agew Chem Int Ed. 46 (2007) 2646 - 2649.

[13] L. Hurley, T. Bialis, T.S. Dexheimer, D. Sun, M. Gleason-Guzman, Abstracts of the 19th Rocky Mountain Regional Meeting of the ACS, 2006, RM-251.

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

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

[16] R. Buscaglia, M.W. Freyer, M. Blynn, J. Ramos, D. Cashman, L. Hurley, E.A. Lewis, Calorimetric and spectroscopic evidence for the binding of TmPyP4 to a 23mer c-MYC i-motif mutant, Abstracts 58th Southeast Regional Meeting of the ACS, 2006, November 1-4.

[17] P. Kumar, A. Verma, S. Maiti, R. Gargallo, S. Chowdhury, Tetraplex DNA transitions within the human c-*myc* promoter detected by Multivariate Curve Resolution of Fluorescence Resonance Energy Transfer, Biochemistry 44 (2005) 16426 - 16434.

[18] T. Simonsson, M. Pribylova, M. Vorlickova, A nuclease hypersensitive element in the human c-*myc* promoter adopts several distinct *i*-tetraplex structures, Biochem. Biophys. Res. Com. 278 (2000) 158 - 166.

[19] Allawi,H., SantaLucia,J.,Jr., Thermodynamics and NMR of internal G·T mismatches in DNA, Biochemistry, 36 (1997) 10581 - 10594

[20] M. Piotto, V. Saudek, V. Sklenar, Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, J Biomol NMR 2 (1992) 661- 665.

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

[22] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, 3-way data-analysis and ambiguity in Multivariate Curve Resolution, J. Chemom. 9 (1995) 31 - 58.

[23] J. Jaumot, M. Vives, R. Gargallo, Application of multivariate resolution methods to the study of biochemical and biophysical processes, Anal. Biochem. 327 (2004) 1 - 13.

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

[25] M. Vives, R. Tauler, R. Eritja, R. Gargallo, Spectroscopic study of the interaction of actinomycin D with oligonucleotides carrying the central base sequences -XGCY- and -XGGCC-Y using multivariate methods, Anal. Bioanal. Chem. 387 (2007) 311 - 320.

[26] S. Freier, K.O. Hill, T.G. Dewey, L.A. Marky, K.J. Breslauer, D.H. Turner, Solvent effects on the kinetics and thermodynamics of stacking in poly(cytidylic) acid, Biochemistry 20 (1981) 1419 - 1426.

[27] V.A. Bloomfield, D.M. Crothers, I. Tinoco Jr, Nucleic Acids: Structures, Properties and Functions (2000) University Science Books, Sausalito CA, USA

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

[29] J.L. Mergny JL, L. Lacroix, H. Xiaogang, J.L. Leroy, C. Leroy, C. Hélène, Intramolecular folding of pyrimidine oligodeoxynucleotides into an I-DNA motif, J. Am. Chem. Soc. 117 (1995) 8887 - 8898.

[30] J.L. Mergny, L. Lacroix, Analysis of thermal melting curves, Oligonucleotides 13 (2003) 515 - 537.

[31] A.T. Phan, J.L. Mergny, Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix, Nucleic Acids Res. 30 (2002) 4618 - 4625.

[32] V. Mathur, A. Verma, S. Maiti, S. Chowdhury, Thermodynamics of i-tetraplex formation in the nuclease hypersensitive element of human *c-myc* promoter, Biochem. Biophys. Res. Com. 320 (2004) 1220 - 1227.

[33] J. Gallego, E.B. Golden, D.E. Stanley, B.R. Reid, The folding of centromeric DNA strands into intercalated structures: a physicochemical and computational study, J. Mol. Biol. 285 (1999) 1039 - 1052.

[34] J. Jaumot, A. Aviñó, R. Eritja, R. Tauler, R. Gargallo, Resolution of parallel and antiparallel oligonucleotide triple helices formation and melting processes by Multivariate Curve Resolution, J. Biomol. Struct. Dyn. 21 (2003) 267 - 278.

[35] A.T. Phan, J.L. Leroy, Intramolecular i-motif structures of telomeric DNA, J. Biomol. Struc. Dyn. Conversation 11, Issue #2 (2000) 245 - 251.

[36] R.J. Fiel, Porphyrin-nucleic acid interactions: a review, J. Biomol. Struct. Dyn. 6 (1989) 1259 - 1274.

Table 1. Thermodynamic data for the folding reaction at 20 ^{o}C as a function of pH. Values obtained from the Vanô Hoff representation of the calculated ln K_{eq} *vs.* 1/T. K_{eq} were calculated from MCR-ALS resolved profiles for each melting experiment.

Figure 1. Molecules studied in this work. (a) Schematic representation of a general i-motif structure. White rectangles represent cytosine bases. (b) Chemical structure of 5,10,15,20-tetrakis(*N*-methyl-pyridnium-4-yl)-21*H*,23*H*-porphyrin (TmPyP4).

Figure 2. Acid-base titration of bcl-2c sequence. (A) Experimental CD and absorbance data. (B) Resolved concentration profiles for the four optically active species present along the titration. The I and II labels denote the species related to i-motif structures. (C) Resolved pure CD spectra. (D) Resolved pure molecular absorption spectra. Continuous line, neutral species; dashed line, i-motif structure I; dotted line, i-motif structure II; dashededotted line, protonated species. Concentration of bcl-2c: 3.3 mM. Other experimental conditions as detailed in the text.

Figure 3. NMR spectra of bcl-2c. Exchangeable proton spectra of bcl-2c at different temperatures and pH = 5.1 (A), 6.1 (B) and 7.0 (C) (buffer conditions: 110 mM K^+,1 mM Mg^{2+}, 10 mM K^+ phosphate; bcl-2c concentration: 0.6 mM). The signal around 13.2 ppm is characteristic of guanine imino protons involved in Watson-Crick base pairs, whereas the signal around 15.6 ppm indicates the presence of imino H3C+ protons.

Figure 4. Melting of bcl-2c sequence at pH 6.1. (A) Experimental absorbance data. Inset: absorbance trace at 295 nm (B) Resolved concentration profiles for the three optically active species present along the melting. (C) Resolved pure spectra. Continuous line: initial bcl-2c sequence mainly as an i-motif structure, dotted line: hairpin at intermediate temperatures, dash-dotted line: random-coil at high temperatures. Concentration of bcl-2c: 8 μM. Other experimental conditions as detailed in the text.

Figure 5. Acid-base titration of a bcl-2c : TmPyP4 mixture (1:1 C_{TmPyP4}:C_{bcl-2c} ratio). (A) Resolved concentration profiles for the mixture after simultaneous analysis of experimental data recorded along the acid-base titrations of bcl-2c, TmPyP4 and the mixture with MCR-ALS. (B) Resolved pure spectra. Black lines

represent bcl-2c species as in Fig. 2. Yellow and magenta lines represent protonated and deprotonated TmPyP4, respectively. Blue line represent the 1:2 bcl-2c:TmPyP4 interaction complex. bcl-2c concentration, 3.0 μM; TmPyP4 concentration, 3.3 μM; temperature, 25 ºC.

Figure 6. NMR spectra of bcl-2c with TmPyP4. (A) bcl-2c alone. (B) 1:1 bcl2-c:TmPyP4, and (C) 1:2 bcl2-c:TmPyP4 (110 mM K^+, 1 mM Mg^{2+}, pH = 6.0, T = 5 ºC, bcl-2c concentration 20 μM).

Figure 7. Interaction of bcl-2c with TmPyP4 studied by a mole ratio experiment at pH 6.1. (A) Experimental absorbance data. (B) Resolved concentration profiles. (C) Resolved pure CD spectra. (D) Resolved molecular absorption pure spectra. Continuous line, bcl-2c sequence; dotted line, 1:2 bcl2-c:TmPyP4 complex; dash-dotted line, TmPyP4. Initial bcl-2c concentration, 4.1 μM; temperature, 25 ºC.

Figure 8. Melting of a TmPyP4:bcl-2c mixture at pH 6.1. MCR-ALS resolved concentration profiles for all optically active species present along the melting. Continuous line: i-motif, dotted line: 1:2 bcl2-c:TmPyP4 interaction species, dash-dotted line: TmPyP4, squares: random coil bcl-2c. Concentration of bcl-2c: 1.8 μM. Concentration of TmPyP4: 7.8 μM, pH 6.1. Other experimental conditions as detailed in the text.

Table 1. Thermodynamic data for the folding reaction at 20 °C as a function of pH. Values obtained from the vanô Hoff representation of the calculated ln K_{eq} *vs.* 1/T. K_{eq} were calculated from MCR-ALS resolved profiles for each melting experiment.

pH	Tm (°C) [a]	Hyperchromicity at 295 nm (%)	ΔG^0 at 20 °C (kcal mol^{-1}) [b]	ΔH^0 (kcal mol^{-1})	ΔS^0 (cal K^{-1} mol^{-1})	Temperature range [c]	r^{2} [d]
4.0	63	+14	–6.6	–52	–153	59-67 (9)	0.9874
4.5	68	-8	-7.0	-50	-148	64-72 (9)	0.9920
4.7	64	-19	-8.0	-61	-182	60-68 (9)	0.9957
5.6	48	-39	-5.5	-63	-196	44-52 (9)	0.9999
6.1	36	-41	-2.5	-49	-157	32-40 (9)	0.9994
6.8	28	-46	-1.1	-49	-164	23-31 (9)	0.9993

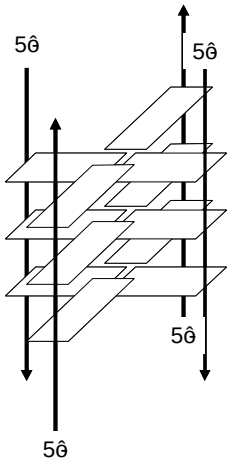
[a] Tm values are given with an uncertainty value of ± 1 °C.

[b] Estimated uncertainties values for thermodynamic data are 10 % (ΔG^0) and 5 % (ΔH^0 and ΔS^0).

[c] Temperature range used for the vanô Hoff representation and number of points used to calculate the regression line.

[d] Correlation coefficient

A)



B)

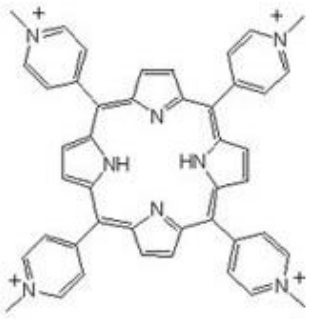


Figure 1.

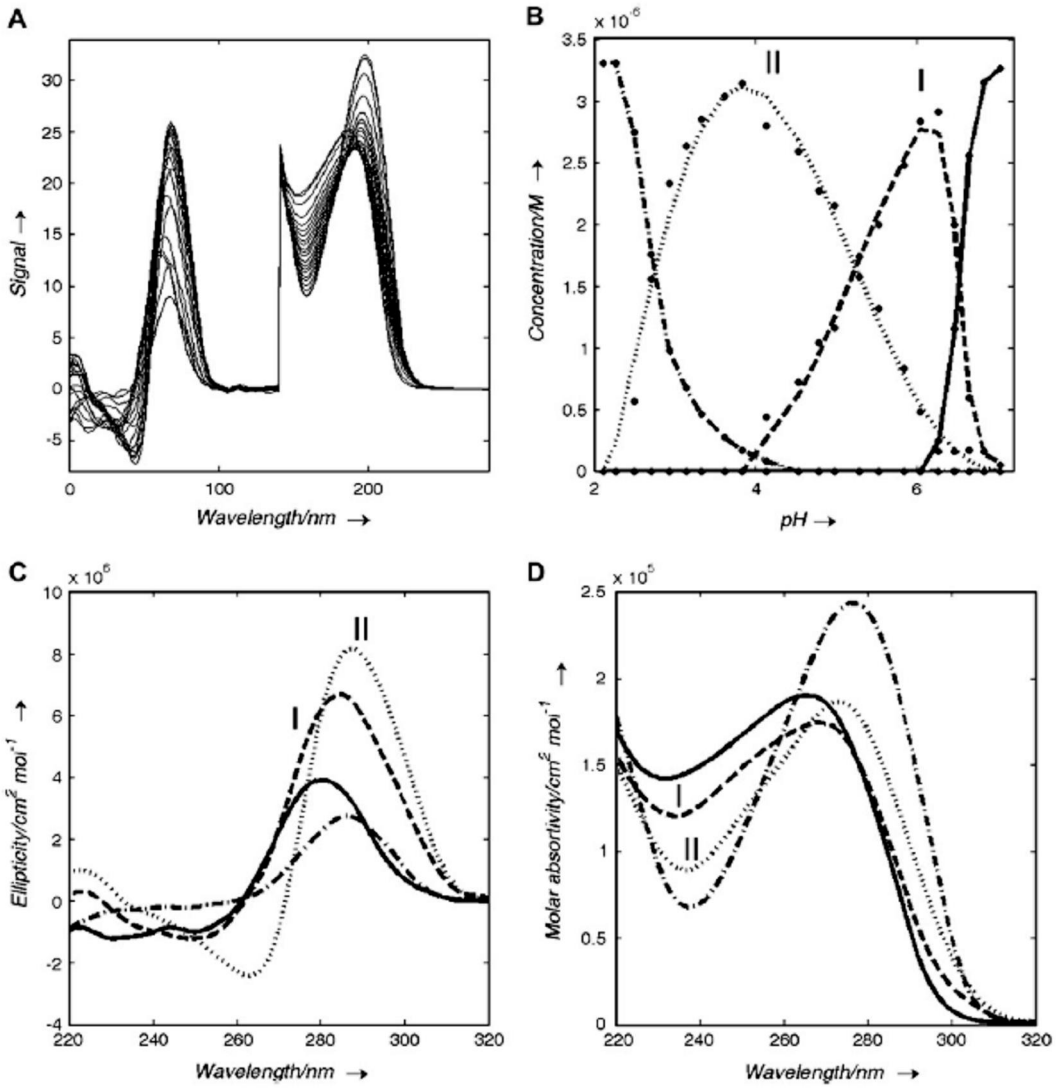

A
Signal →
Wavelength/nm →
B
x 10⁻⁶
Concentration/M →
pH →
II
I
C
x 10⁵
Ellipticity/cm² mol⁻¹ →
Wavelength/nm →
D
x 10⁵
Molar absorptivity/cm² mol⁻¹ →
Wavelength/nm →

Figure 2

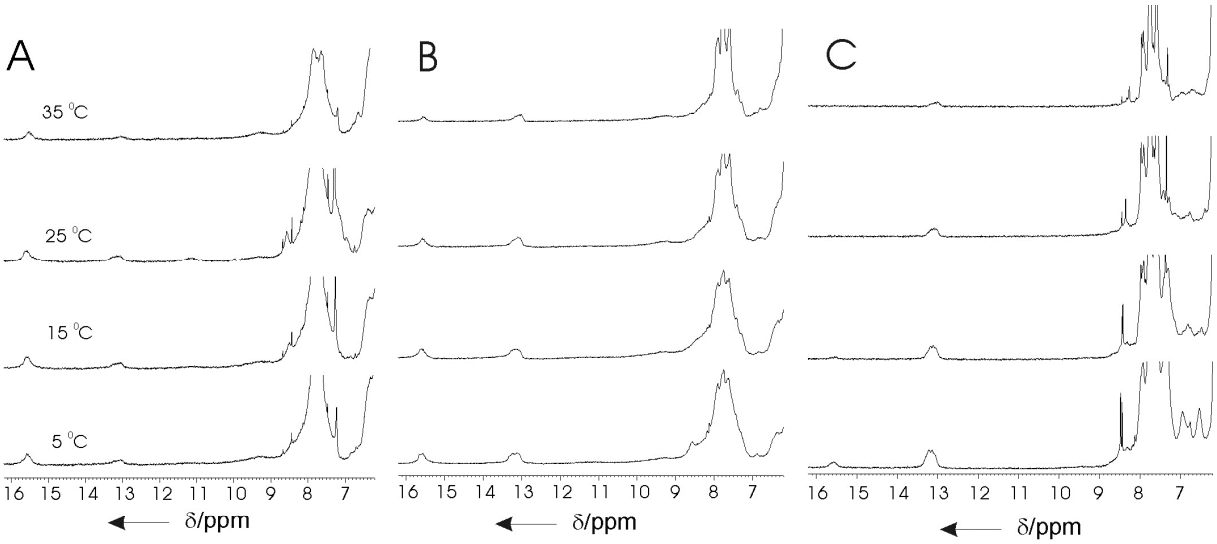

A
35 °C
25 °C
15 °C
5 °C
16 15 14 13 12 11 10 9 8 7
δ/ppm
B
16 15 14 13 12 11 10 9 8 7
δ/ppm
C
16 15 14 13 12 11 10 9 8 7
δ/ppm

Figure 3

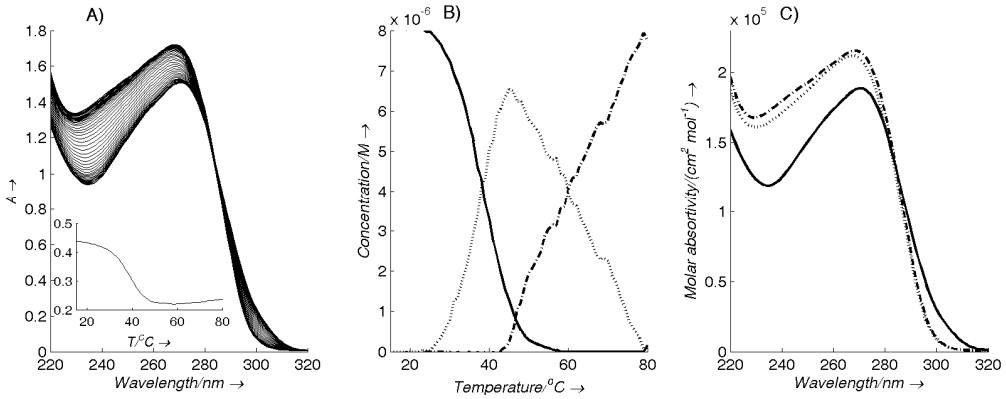


Figure 4

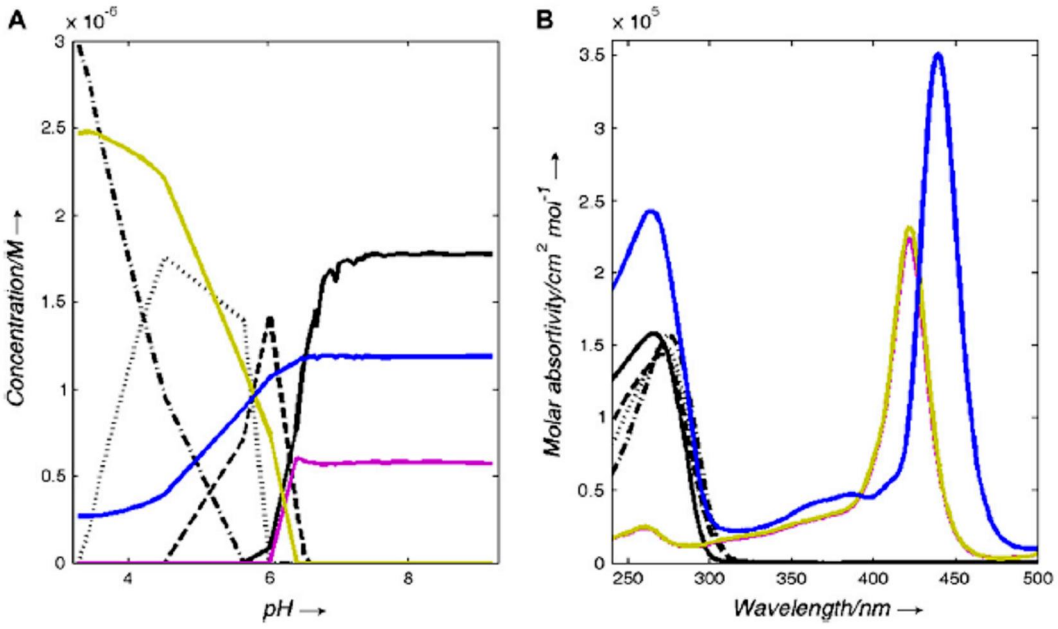


Figure 5.

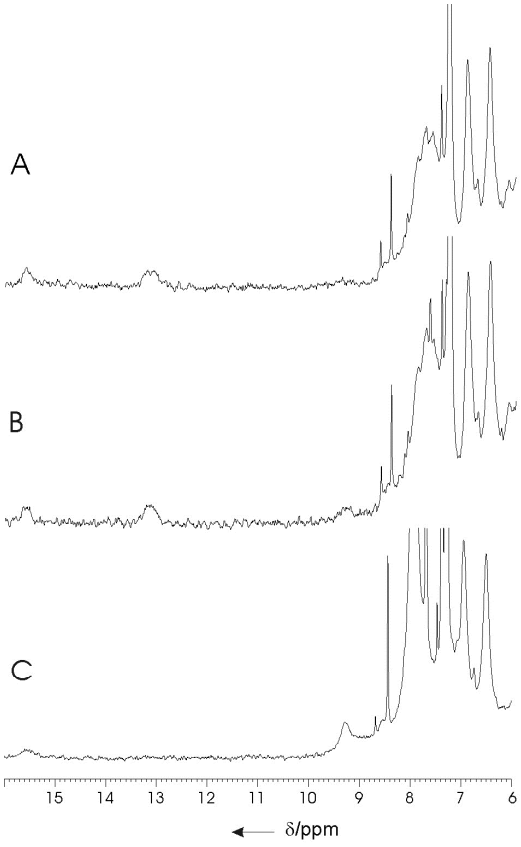


Figure 6

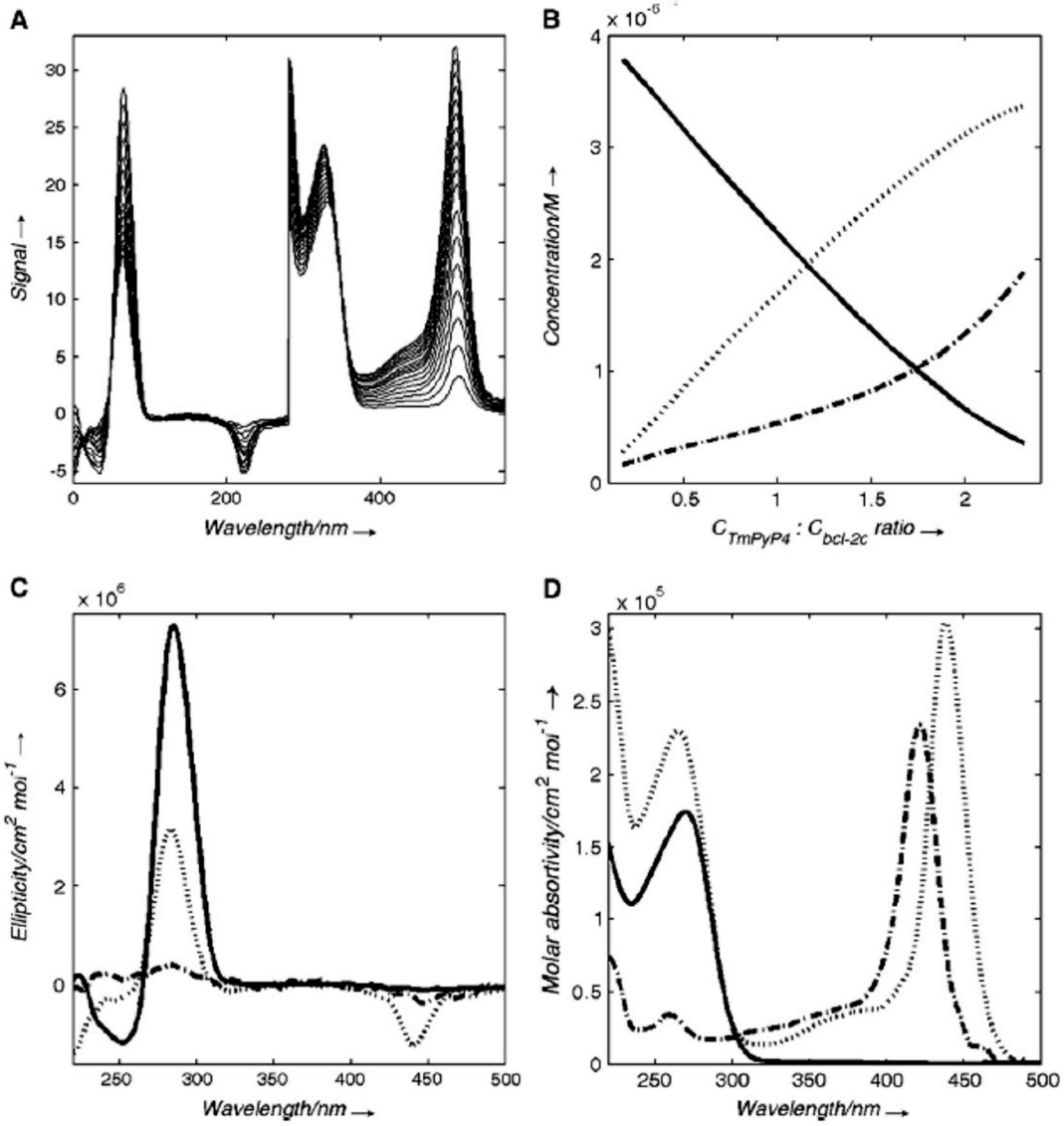

A
Signal →
Wavelength/nm →
B
x 10^-5
Concentration/M →
C_TmPyP4 : C_bcl-2c ratio →
C
x 10^6
Ellipticity/cm^2 mol^-1 →
Wavelength/nm →
D
x 10^5
Molar absorptivity/cm^2 mol^-1 →
Wavelength/nm →

Figure 7

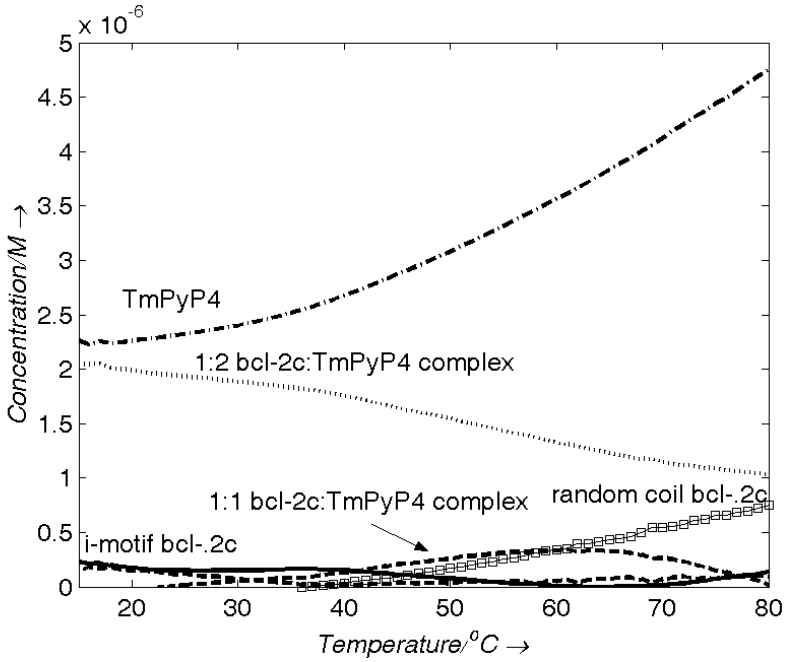

x 10^-6
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Concentration/M →
TmPyP4
1:2 bcl-2c:TmPyP4 complex
1:1 bcl-2c:TmPyP4 complex
random coil bcl-.2c
i-motif bcl-.2c
20
30
40
50
60
70
80
Temperature/°C →

Figure 8.

Supplementary Material

Solution equilibria of the *i*-motif Forming Region Upstream of the bcl-2 P1 Promoter

*Nasiruddin Khan[1], Anna Aviñó[2], Romà Tauler[3], Carlos González[4], Ramon Eritja[2], Raimundo Gargallo[1]**

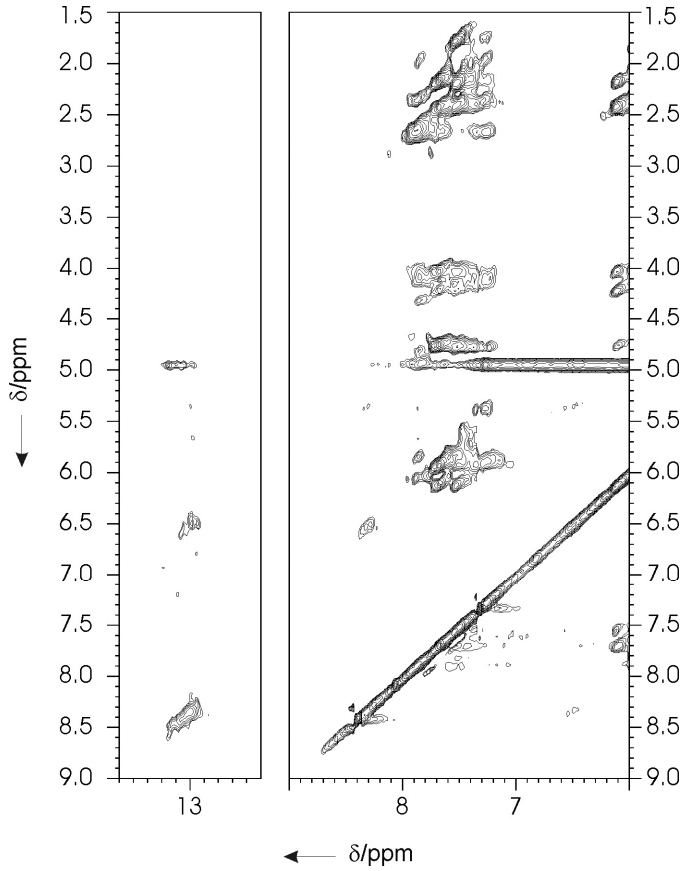
1. Department of Analytical Chemistry, University of Barcelona, Diagonal 647, E-08028 Barcelona, Spain

2. Institut de Biologia Molecular de Barcelona, CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain

3 Department of Environmental Chemistry, IIQAB-CSIC, Jordi Girona 18-26, E-08034 Barcelona, Spain

4 Instituto de Quimica-Fisica Rocasolano, CSIC, Serrano 119,. 28006 Madrid, Spain

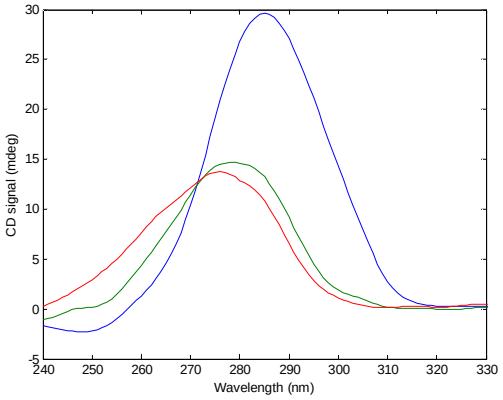
1. NOESY spectra



Regions of the NOESY spectrum (150ms mixing time) of bcl2-c in H_2O (0.6 mM oligonucleotide concentration, 110 mM KCl, 1 mM $MgCl_2$, T=5ºC, pH=7). The presence of several Watson-Crick base pairs can be established from the cross-peaks between guanine imino protons (between 13.0-13.3 ppm) and cytosine amino protons (around 8.3-8.5 ppm, and 6.4-6.6 ppm).

2. Melting of bcl-2c sequence at pH 6.6 monitored by CD

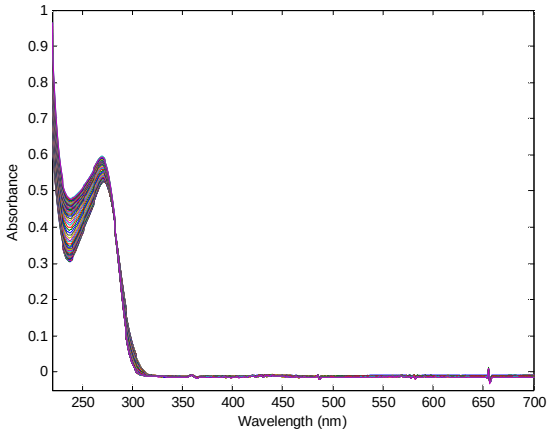
Figure S2. CD spectra of bcl-2c sequence at pH 6.6 recorded at 18 ^{o}C (blue), 39 ^{o}C (green) and 69 ^{o}C (red). $C_{bcl\text{-}2c}$ = 3 μM.



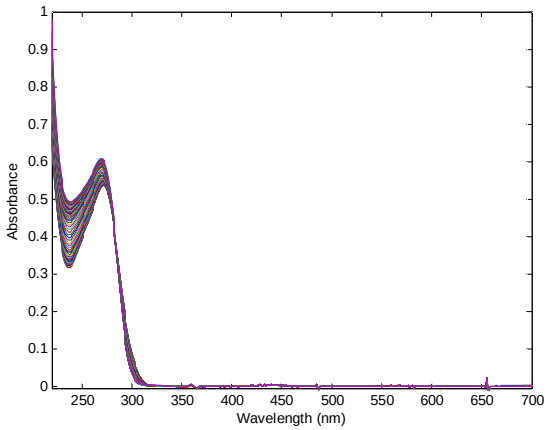
3. Analysis of absorbance data with Multivariate Curve Resolution

Here we include a brief description of the main steps done along the analysis of a dataset with Multivariate Curve Resolution. In this case, that corresponding to the melting of bcl-2c sequence at pH 5.6. Data analysis has been carried out by application of MCR-ALS procedure. For a detailed explanation of this procedure, including the use of the graphical interface of MCR-ALS, the reader is refereed to the tutorial of Jaumot et al. (Chem. Intel. Lab. Syst 2005, 76, 101 - 110).

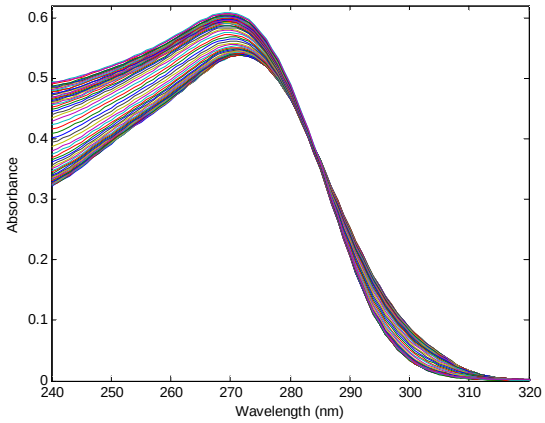
1. Raw dataset:



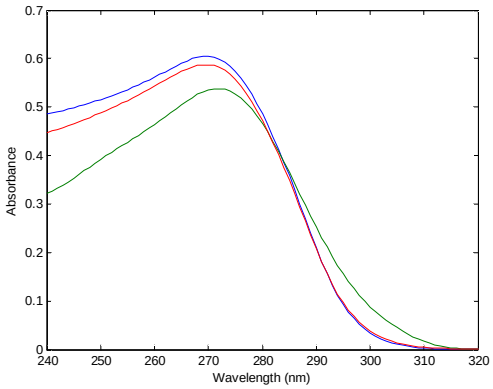
2. Removing base line drift. Sometimes, smooth is needed (as example, CD data).



3. Choosing the data interval to be analyzed. In this case, wavelengths below 240 nm have been removed because they contain a strong contribution of buffer absorption and are not directly related to any bcl-2c structural transition.

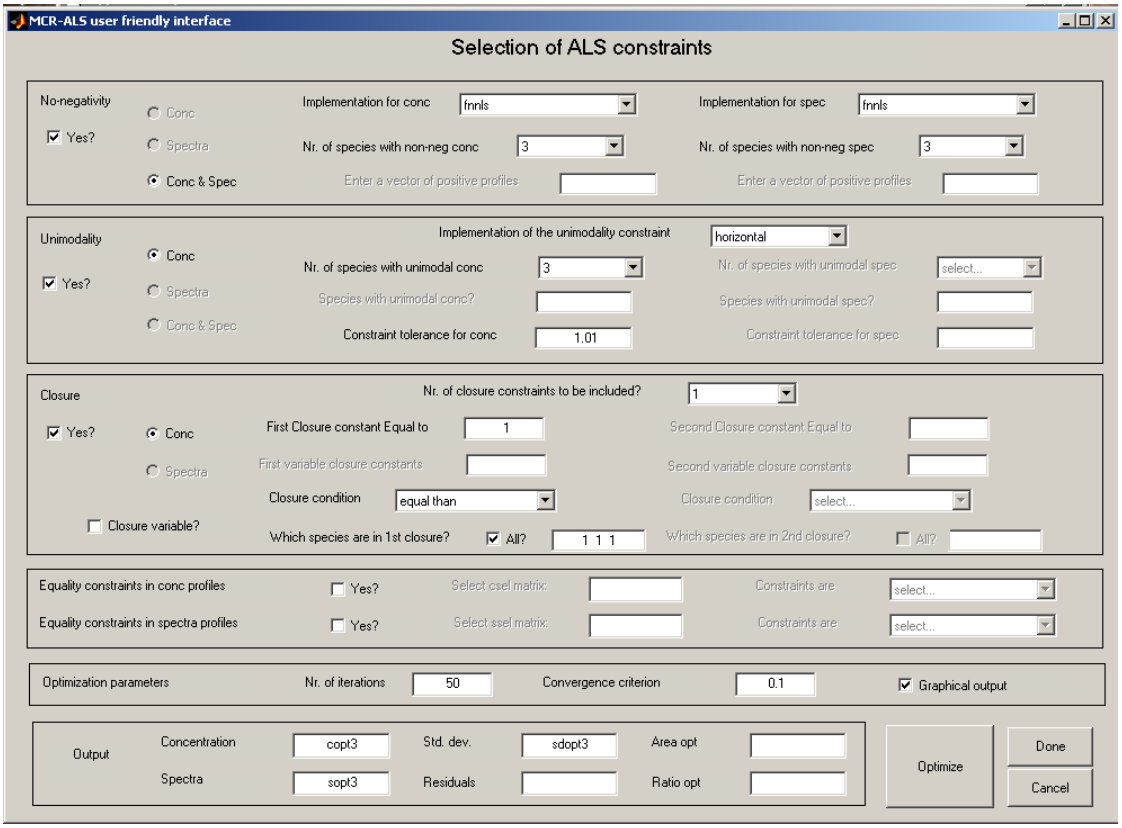


4. Obtaining an initial estimate of the pure spectra with SIMPLISMA. In this case, three components were considered. This number was selected after Singular Value Decomposition analysis of the data matrix plotted in 3). Alternatively, two and four were also tested. The best results (from a biophysical analytical point of view) were obtained when only three were considered.

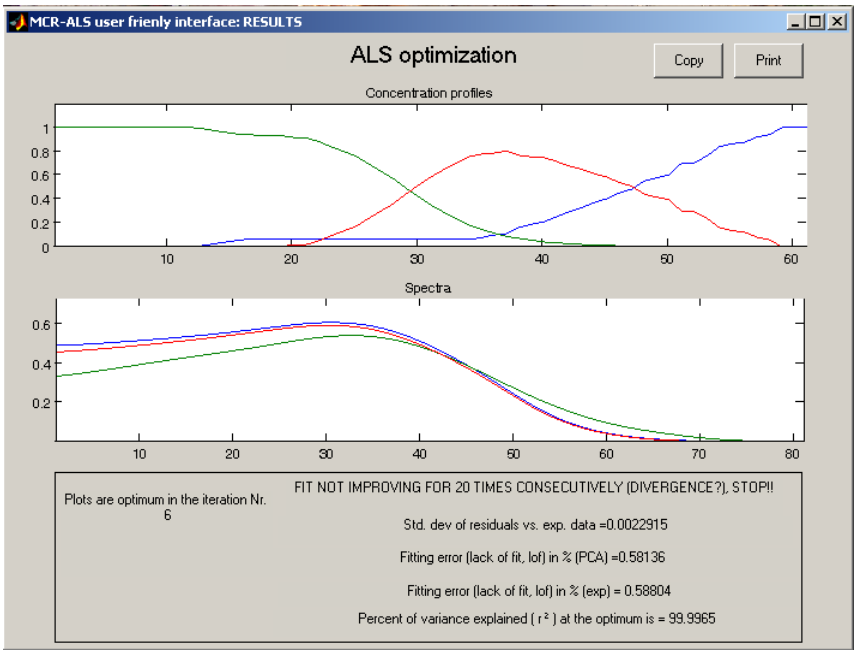


5. Optimization of the initial estimate of the pure spectra with ALS

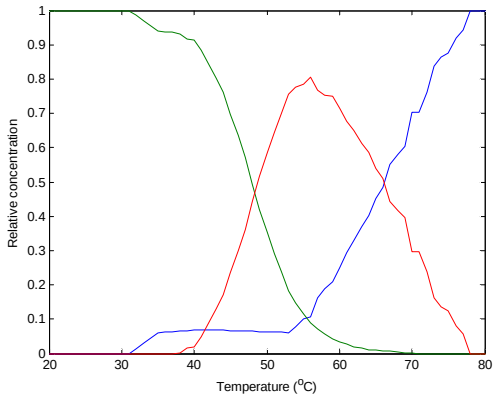
Next, a screen copy of the ALS graphical interface is shown. Application of several constraints is needed to optimize the initial estimates removing mathematical solutions non-sense from the biophysical point of view (as example, negative absorbances are set to zero). In this example, however, application of constraints is not critical as the initial estimates of the pure spectra are close to the real ones.

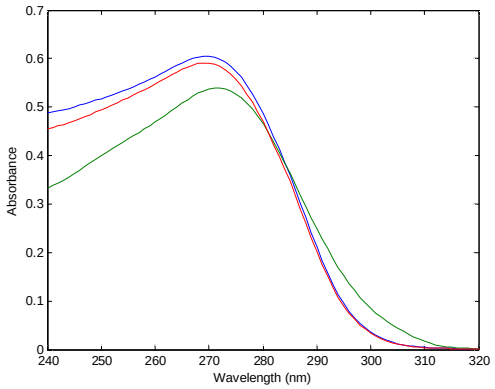


And the final MCR-ALS window showing the ALS resolved concentration profiles (against number of experimental spectrum) and pure spectra (against measurement channel, i.e., number of wavelength), together with some figures about the quality of ALS optimization:



Finally, the resolved concentration profiles are plotted against the measured temperature values and the resolved pure spectra are plotted against the measured wavelengths:

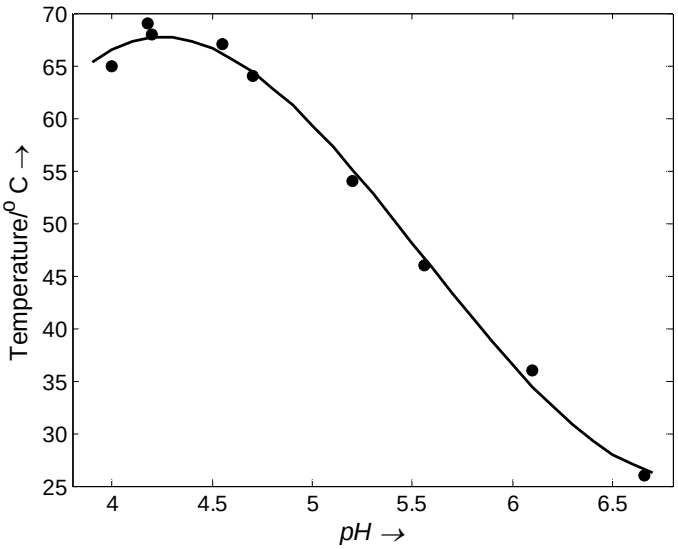




T_m is determined at the midpoint of the transition involving structured and unstructured species. Thermodynamic values can be obtained from the resolved concentration profiles, as described in the manuscript.

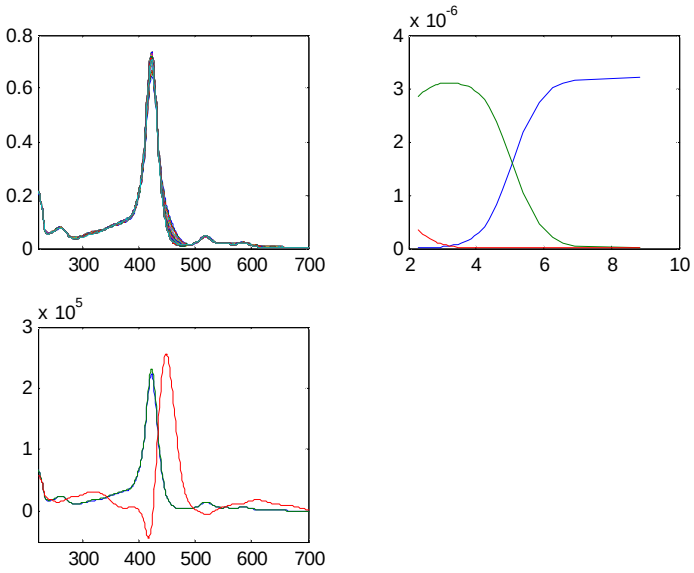
The analysis assumes that the variation of the experimental absorbance data upon heating is mainly due to the conformational changes, and not to the variation of the absorbance spectra with temperature. This assumption is supported by the straightforward determination of the optimal number of components by Singular Value Decomposition (SVD) [Haq et al., Eur. Biophys. J. 1997, 26, 419 - 426]. If the absorbance spectra of each component varied with temperature, SVD plot would not allow an unambiguous determination of the number of components.

4. Plot of Tm values (from melting experiments of bcl-2c sequence) *vs* pH



5. Acid-base titration of TmPyP4 monitored by molecular absorption

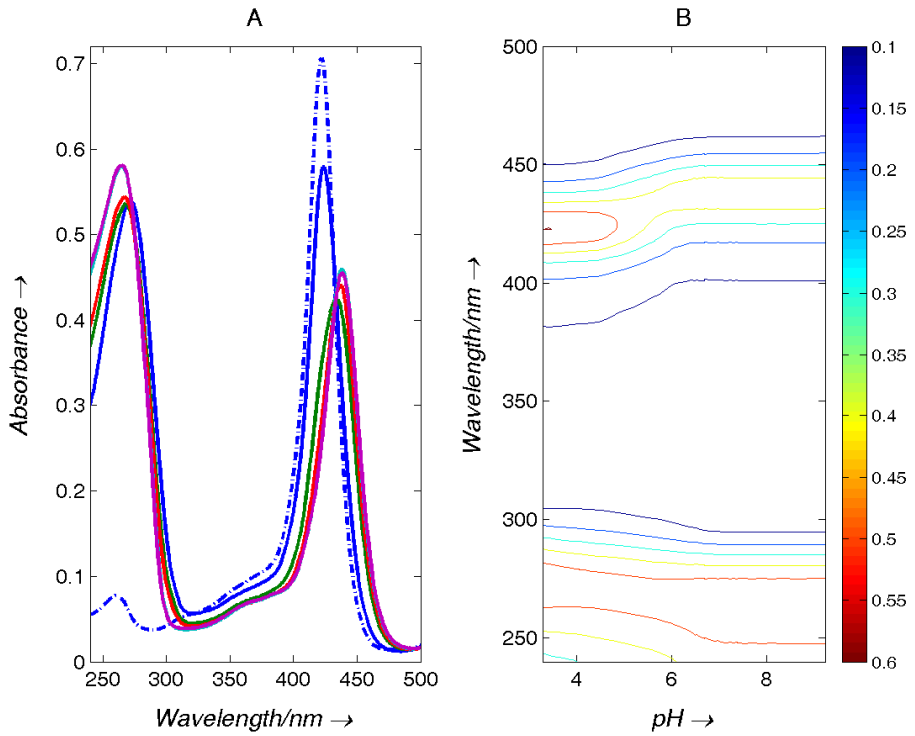
Absorbance data have been collected from 220 to 700 nm and pH ranging from 2.2 to 8.8. The experimental data set has been analyzed with the program EQUISPEC (Dyson et al., Anal. Chim. Acta 1997, 353, 381 - 393) which allow the calculation of equilibrium constants from multivariate data for systems involving monomers. Next figure show the experimental data and the resolved concentration profiles and pure spectra for each one of the three acid-base species considered:



Two pKa values were determined: 5.3 ± 0.1 and 1.4 ± 0.2. While the first pKa value has been correctly calculated, the second pKa value has been badly determined because of the small formation of the fully protonated species (red). Accordingly, the pure spectrum for this species has also badly calculated (it shows negative values).

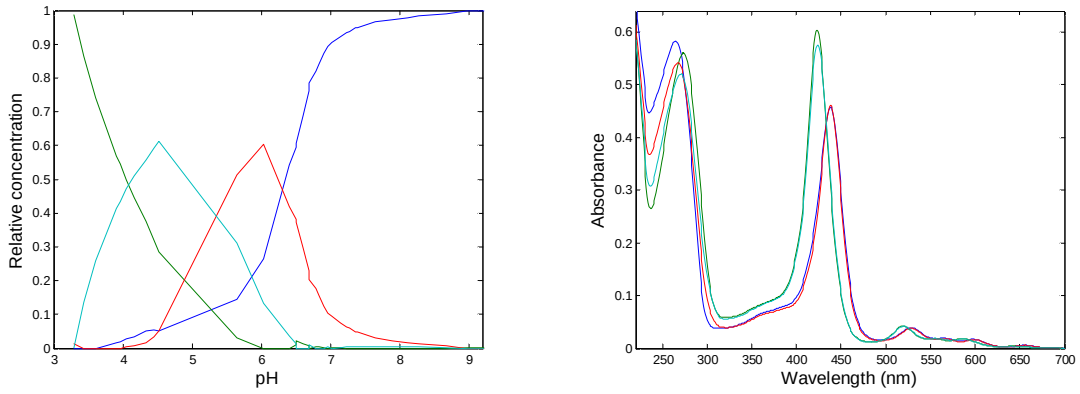
6. Acid-base titration of a 1:1 bcl-2c:TmPyP4 mixture monitored by molecular absorption

An acid-base titration of a mixture of bcl-2c and TmPyP4 was monitored by molecular absorption spectroscopy. Next, several experimental spectra recorded at pH 4.05 (blue), 5.65 (green), 6.03 (red), 7.02 (cyan) and 9.22 (magenta) are shown, together with the spectra of TmPyP4 (dotted line). Also, a contour plot showing the shift of the absorbance maxima upon pH change. Experimental conditions: bcl-2c concentration: 3.0 μM, TmPyP4 concentration: 3.3 μM, temperature: 25 ^{o}C.



According to the observed spectral features, interaction between bcl-2c and TmPyP4 occurs in the whole studied pH range, being the most dramatic changes on the visible region at pH higher than 6.

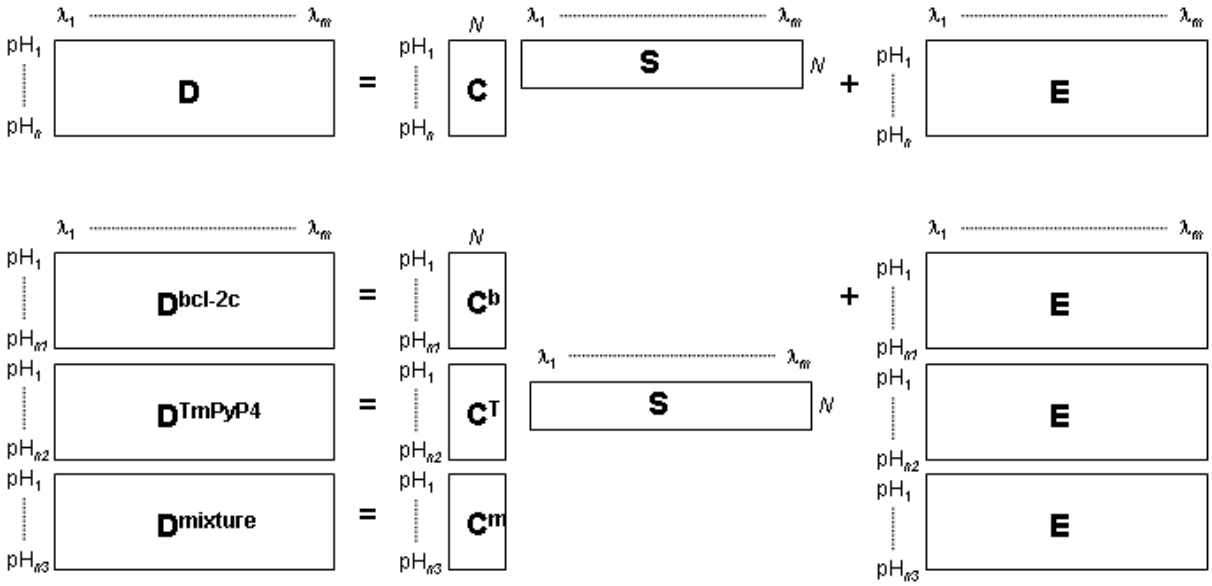
The experimental data recorded along the titration was analyzed in two different ways. First, the data matrix was analyzed according to the procedure described in the manuscript. The results obtained are summarized in the following figures:



The four resolved components can not be directly related to pure acid-base species or interaction complexes. For instance, the ôblueô species, which is the majority species at neutral pH values, is probably related to a mixture of hairpin bcl-2c and the interaction complex (as denoted by the shift of the visible band of TmPyP4 at 438 nm). Similar conclusion can be drawn for the other three species, where acid-base species of bcl-2c (including i-motif) and of the TmPyP4 are mixed.

In order to fully resolve the ambiguities still present in the results obtained in the previous analysis, another analysis strategy was tested. Hence, the experimental data recorded along the titration of the mixture was simultaneously analyzed together with the data recorded along the titration of bcl-2c (Figure 2 in the paper) and the data recorded along the titration of TmPyP4 (see above). The mathematical bases of this simultaneous analysis have been extensively explained elsewhere (see previous papers in J. Biomol. Struc. Dyn. 2003, 21, 267; Nucleic Acids Res. 2002, 30, e92; and previous references to the development of MCR-ALS by R. Tauler). In this manuscript, the simultaneous analysis was done by building up a column-wise augmented data matrix. This is done by grouping the single matrices containing the spectra recorded at different experimental conditions (analytical concentrations of bcl-2c and TmPyP4) but using the same instrumental technique (molecular absorption spectroscopy), according to the following scheme.

In this case, the initial estimate of the pure spectra for the ALS optimization was building up by joining the resolved spectra for the four acid-base species of bcl-2c (Figure 2 in the paper), the two resolved spectra for the two acid-base species of TmPyP4 (see above) and the resolved spectra 1:2 interaction complex resolved in the mole-ratio at pH 6.1.



N: number of spectroscopically active species considered

Now, the resolved concentration profiles and pure spectra showed the contribution of all

spectroscopically active species in the mixture. It must be stressed that the results shown in next figures must be considered critically because the results obtained with MCR-ALS can fail when a high overlap exists. This is our case as in the pH region below 7 coexist around seven acid-base species with very similar spectra.

7. Mole-ratio study carried out at pH 7.6

The best results were obtained when one interaction species were considered with stoichiometry 1:2 (DNA:TmPyP4) and with binding constant log $\beta = 11.7 \pm 0.1$. Next figure shows the results obtained at pH 7.6, where hairpin is the majority species in absence of TmPyP4: (A) Experimental CD and absorbance data. (B) Resolved concentration profiles. (C) Resolved pure CD spectra. (D) Resolved molecular absorption pure spectra. Continuous line: bcl-2c sequence, dotted line: 1:2 bcl2-c:TmPyP4 complex, dash-dotted line: TmPyP4. Initial bcl-2c concentration: 4.1 μM, temperature: 25 ^{o}C.