

“Understanding the effect of the nature of the nucleobase in the loops on the stability of the i-motif structure” Benabou, S., Garavís, M., Lyonnais, S., Eritja, R., González, C., Gargallo, R. *PhysChemChemPhys*, 18(11), 7997-8004 (2016). doi: 10.1039/c5cp07428b

Understanding the effect of the nature of the nucleobase in the loops on the stability of the i-motif structure

Sanae Benabou^a, Miguel Garavís^b, Sébastien Lyonnais^c, Ramon Eritja^d, Carlos González^b, Raimundo Gargallo^{a,*}

a. Department of Analytical Chemistry, University of Barcelona, Diagonal 645, E-08028 Barcelona, Spain

b. Institute of Physical Chemistry “Rocasolano”, CSIC, Serrano 119, E-28006 Madrid, Spain

c. Institute of Molecular Biology of Barcelona (IBMB-CSIC), Baldori i Reixac 4-8, 08028 Barcelona, Spain

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

Abstract. The nature and the length of loops connecting cytosine tracts in i-motif structures may affect their stability. In this work, the influence of the nature of the nucleobases located in two of the loops is studied using spectroscopic and separation techniques and multivariate data analysis. The insertion of bases other than thymine induces an additional acid-base equilibrium with $pK_a \sim 4.5$. The presence of two guanine bases placed opposite to each other decreases the thermal stability of the structure. In contrast, thymine and cytosine bases in these positions stabilize the structure.

Introduction

The i-motif structure is the only known DNA structure that consists of parallel-stranded duplexes held together by intercalated base pairs (Figure 1). The formation of the C·C⁺ base pair needs the protonation of one of the cytosines at N3; the pK_a value of which is around 4.5, depending on temperature and ionic strength. For this reason, stable i-motif structures are usually observed at a slightly acid medium¹. The formation *in vitro* of such structures in DNA sequences corresponding to centromeres, telomeres and to the promoter regions of several oncogenes, such as *c-kit*, *c-myc* or *bcl-2*, has been demonstrated.²⁻⁴ The study of these structures is not only interesting from a biophysical and biomedical point of view⁴, but also for their potential application in nanotechnology⁵. The i-motifs have been employed in the design of *in vivo* sensing systems⁶, artificial DNA-based nanomachines⁷ and logic gates⁸, among other applications. Recently, it has been demonstrated that the rational manipulation of i-motif composition can be utilized in designing pH sensors and the capability to tune both the response range and response sensitivity⁹.

The number and nature of the bases located in the loops may affect the stability of the i-motif structures. It has been suggested that i-motif structures may be classified into two groups

depending on the length of loops¹⁰: i-motifs showing short loop regions among cytosine tracts (class 1) and i-motifs showing longer loop regions (class 2). Class 2 i-motif structures are more stable than those in class 1, although recent results have cast some doubts on the generality of this effect¹¹. Using sequences that only contain thymine and cytosine bases it has been reported that shorter loops exhibit the highest stability. In this sense, flexibility related to long loop regions could be detrimental to the stability of the i-motif structure and long loops are only stabilizing if they can form additional intramolecular interactions that limit flexibility. However, it should be noted that this additional flexibility could also bring favourable entropic contribution to the folding free energy. The nature of the loops also influences the structure and stability of the i-motif. For example, it has been observed that the presence of adenine bases in the loops produces conformational changes involving disruption of the i-motif core^{12, 13}. In contrast, the mutation of the TAA repeats for TTT in the intramolecular i-motif structure formed by the cytosine-rich strand of the human telomere increases its stability¹⁴. The influence of bases adjacent to the cytosine-tracts, as well as the effect of the loop length, has been addressed in a recent study¹⁵. There, it was concluded that: (i) guanine or thymine in the loop positions adjacent to cytosine tracts in lateral loops stabilizes the i-motif; (ii) the same number of bases in lateral loops is optimal for stabilization; (iii) an additional C.C⁺ base pair also stabilizes the i-motif. To our knowledge, this is the first study focused on the influence of the nature of internal bases located at the lateral loops on the thermal and acid-base stabilities of i-motif structures. In this work we will use the sequence 5'-TT CCC TXT CCC TTT CCC TXT CCC TT-3', where X are T, A, C or G (S1, ESI) as a model to study these interactions.

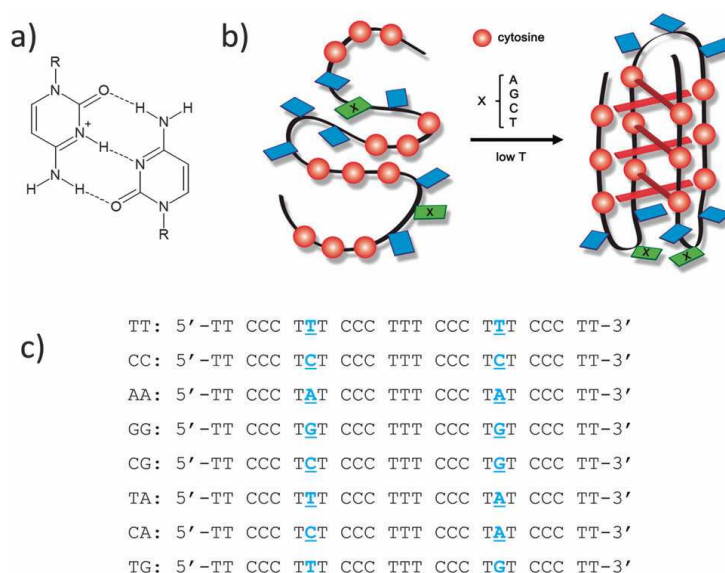


Figure 1. (a) Cytosine-protonated cytosine base pair; (b) Hypothetical scheme of the intramolecular i-motif structures adopted by the sequences studied in this work; X=A, C, G or T bases; (c) Sequences studied in this work.

Materials and methods

Apparatus

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, 3 mL volume) were used. NMR spectra were acquired either using a Bruker Avance spectrometer operating at 600 MHz (sequences CC, AA, GG, and CG) or using a Bruker Digital Avance 800 MHz (sequences TT, TA, CA and TG). Water suppression was achieved by the inclusion of a WATERGATE¹⁶ module in the pulse sequence prior to acquisition. Absorbance spectra were recorded on an

Agilent 8453 diode array spectrophotometer. The temperature was controlled by means of an 89090A Agilent Peltier device. Hellma quartz cells (1 or 10 mm path length, and 350, 1500 or 3000 μL volume) were used. The chromatographic system consisted of an Agilent 1100 Series HPLC instrument equipped with a G1311A quaternary pump, a G1379A degasser, a G1392A autosampler, a G1315B photodiode-array detector furnished with a 13 mL flow cell, and an Agilent Chemstation for data acquisition and analysis (Rev. A 10.02), all from Agilent Technologies (Waldbronn, Germany). A BioSep-SEC-S 3000 column (300 x 7.8mm, particle size 5 μm and pore size 290 \AA) from Phenomenex (Torrance, CA, USA) was used for the chromatographic separation at room temperature.

Reagents

The DNA sequences (Fig. 1) were obtained from Sigma Aldrich (St. Louis, Missouri, USA). Two control sequences were also synthesized for PAGE analysis (control1: 5'-TT CCC TTT AAA TTT CCC TTT CCC TT-3', and control 2: 5'-TT CCC TTT AAA TTT AAA TTT CCC TT-3'). DNA strand concentration was determined by absorbance measurements (260 nm) at 90 $^{\circ}\text{C}$ using the extinction coefficients calculated using the nearest neighbor method as implemented on the OligoCalc webpage.¹⁷ Before any experiment, DNA solutions were first heated to 90 $^{\circ}\text{C}$ for 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.

Procedures

Acid–base titrations were monitored in-line taking advantage of the stirrer incorporated in the Agilent cell holder. Experimental conditions were 25 $^{\circ}\text{C}$ and 150mMKCl. Titrations were carried out by adjusting the pH of solutions containing the oligonucleotides. pH was measured using an Orion SA 720 pH/ISE meter and a micro-combination pH electrode (Thermo). Absorbance spectra were recorded in a pH stepwise fashion. Thermal stability studies were monitored using an Agilent- 8453 spectrophotometer equipped with a Peltier unit. The DNA solution was transferred to a covered 10 mm-path-length cell and absorption spectra were recorded at 1 $^{\circ}\text{C}$ intervals with a hold time of 3 min at each temperature, which yielded an average heating rate of approximately 0.3 $^{\circ}\text{Cmin}^{-1}$. 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 began.

PAGE analysis

The oligonucleotides were suspended in water and diluted at a concentration of 5 mM in a solution buffered at pH (5.0) (50 mM sodium acetate pH (5.0), 10 mM potassium acetate) or at pH (7.8) (50 mM Tris acetate pH (7.8), 10 mM potassium acetate). The solutions were next boiled for 5 minutes at 95 $^{\circ}\text{C}$ and cooled down to room temperature. Glycerol was added (2.5% v/v) and 5 mL of the samples were loaded on 10 x 10.5 cm – 12% nondenaturing polyacrylamide gels (19/1 acrylamide/bisacrylamide, Sigma) either buffered with 50 mM Trisodium citrate adjusted to pH (5.0) with citric acid or with 0.5x Tris pH (8.0) borate EDTA (TBE). Electrophoresis was run at 4 $^{\circ}\text{C}$ using a mini VE apparatus (Hoeffer) for 4 h at 11 V cm^{-1} . 0.1 M tri-sodium citrate pH (5.0) was used as a running buffer for the acidic gel and 0.5x TBE for the other. The gel buffered at pH (5.0) was then incubated for 15 min in 1x TBE, and rinsed with water and both gels were stained with SybrGold (molecular probes) according to the manufacturer's instructions and digitalized using a Typhoon 8600 system (molecular dynamics).

Size-exclusion chromatography

The mobile phase was 300 mM KCl and 20mM buffer (phosphate or acetate) adjusted to the desired pH value. The flow rate was set to 1.0 mL min⁻¹. The injection volume was 15 µL. Absorbance spectra were recorded between 200 and 500 nm. Cytosine was used as a marker of the permeation limit (retention time, t_R equal to 12.5 min). T₁₅, T₂₀, T₂₅ and T₃₀ sequences were used as standards to construct the t_R vs. log(MW) calibration plot for unfolded sequences. Standards were injected twice to assess the reproducibility of the t_R values. At all pH values studied, the relative difference between t_R values for a given standard was lower than 0.5%.

Analysis of melting data

Absorbance data as a function of temperature were analyzed as described elsewhere.¹⁸ The physico-chemical model is related to the thermodynamics of DNA unfolding. Hence, for the one-step unfolding of intramolecular structures such as those studied here, the corresponding equilibrium constant may be written as:

$$K_{\text{unfolding}} = [\text{DNA}_{\text{unfolded}}]/[\text{DNA}_{\text{folded}}] \quad (1)$$

For thermal stability studies, 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:

$$\ln K_{\text{unfolding}} = -\Delta H_{\text{vH}}/RT + \Delta S_{\text{vH}}/R \quad (2)$$

It is assumed that ΔH_{vH} and ΔS_{vH} will not change throughout the range of temperatures studied here.

Analysis of spectra recorded along acid–base titrations

Spectra recorded during acid–base experiments were arranged in a table or data matrix **D**, with m rows (spectra recorded) and n columns (wavelengths measured). The goal of data analysis was the calculation of distribution diagrams and pure (individual) spectra of all nc spectroscopically active species considered throughout an experiment. The distribution diagram provides information about the stoichiometry and stability of the species considered (in the case of acid–base and mole-ratio experiments), as well as the thermodynamics of the melting processes. In addition, the shape and the intensity of the pure spectra may provide qualitative information about the structure of the species. With this goal in mind, data matrix **D** was decomposed according to Beer–Lambert–Bouyer's law in the matrix form:

$$\mathbf{D} = \mathbf{CS}^T + \mathbf{E} \quad (3)$$

where **C** is the matrix ($m \times nc$) containing the distribution diagram, **S**^T is the matrix ($nc \times n$) containing the pure spectra, and **E** is the matrix of data ($m \times n$) not explained by the proposed decomposition.

The mathematical decomposition of **D** into matrices **C**, **S**^T, and **E** may be conducted in two different ways, depending on whether a physico-chemical model is initially proposed (hard-modelling approach) or not (soft-modelling approach).^{19–23} For hard-modelling approaches, the proposed model depends on the nature of the process under study. 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, such as the following:



$$\text{Beta}_{1,p} = [\text{DNA}\bullet\text{H}_p]/[\text{DNA}][\text{H}^+]^p \quad (5)$$

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

Whenever a physico-chemical model is applied, the distribution diagram in **C** complies with the proposed model. Accordingly, the proposed values for the equilibrium constants and the shape of the pure spectra in **S^T** are refined to explain satisfactorily data in **D**, whereas residuals in **E** are minimized. In this study, hard-modelling analysis of acid-base experiments used the EQUISPEC program.²⁴

Results and discussion

CD spectra

The formation of i-motif structures was firstly assessed by means of circular dichroism (CD) experiments carried out at several pH values and 25 °C. The overall spectral signature characteristic of the i-motif structure (a negative band at around 265 nm and a stronger positive band at around 287 nm) was clear for all the considered sequences (ESI⁺). The CD spectra show practically no variation with pH for several sequences, such as TT. On the other hand, the CD spectra of other sequences, such as AA or GG, show clear changes with pH. Other sequences, such as CG, TA or CA, among others, show an intermediate behavior. These observations point to local modifications of the i-motif structure due to the nature of internal bases at the loops.

NMR spectra

I-motif formation was then demonstrated by nuclear magnetic resonance (NMR) (Fig. 2 and ESI⁺). The imino proton resonances at 15 ppm indicate the presence of protonated cytosines and the formation of C-C⁺ base-pairs. The number of signals and their chemical shifts are very similar in spectra acquired at different temperatures, ranging from 5 °C to 45 °C (ESI⁺), indicating that the structures remain unaltered, in this temperature range. In general, spectra recorded at pH 3.5 exhibit sharper and more dispersed signals than those acquired at pH 5.0 (Fig. 2).

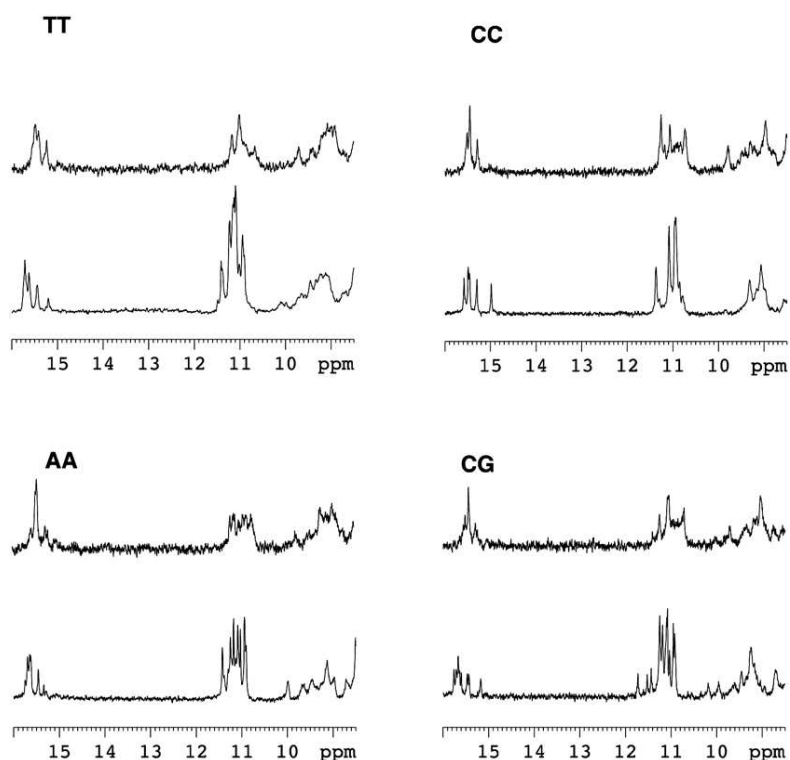


Fig. 2 Exchangeable proton region of the ^1H NMR spectra of TT, CC, AA and CG recorded at pH 5.0 (top spectra in each panel) and pH 3.5 (bottom) (20 mM potassium phosphate, 100 mM KCl, 5 $^\circ\text{C}$, 90 : 10 $\text{H}_2\text{O}/\text{D}_2\text{O}$).

The imino signals are visible even at 45 $^\circ\text{C}$ (Fig. S2 in ESI[†]), showing the relatively high thermostability of the i-motifs formed. Analysis of NMR spectra revealed the formation of six $\text{C}\cdot\text{C}^+$ base pairs (Fig. 3a). Other NOEs characteristic of i-motif structures are clearly observable (i.e.: $\text{H1}'\text{--H2}'$ crosspeaks in Fig. 3b). The absence of signals between ~ 12.5 and ~ 14 ppm indicates that Watson–Crick base pairs are not formed in any of the sequences studied, including those containing complementary bases, TA (ESI[†]) and CG (Fig. 2). Interestingly, all sequences exhibit sharp imino signals at around 11 ppm. These signals arise from loop thymines or guanines, and they are observable at relatively high temperature, as are the imino signals of the hemiprotonated cytosines. This suggests that loop residues adopt a defined structure. In most cases NOESY cross-peaks between some of these imino signals indicate the formation of TT mismatches (Fig. 3c). Some differences are observed between NMR spectra acquired at pH 3.5 and 5.0, such changes being more pronounced in sequences different from TT (Fig. 2). The significant differences in cytosine imino chemical shifts (illustrated in Fig. S2.3, ESI[†]) between oligonucleotides whose sequences differ only in loop nucleotides are also worth noticing.

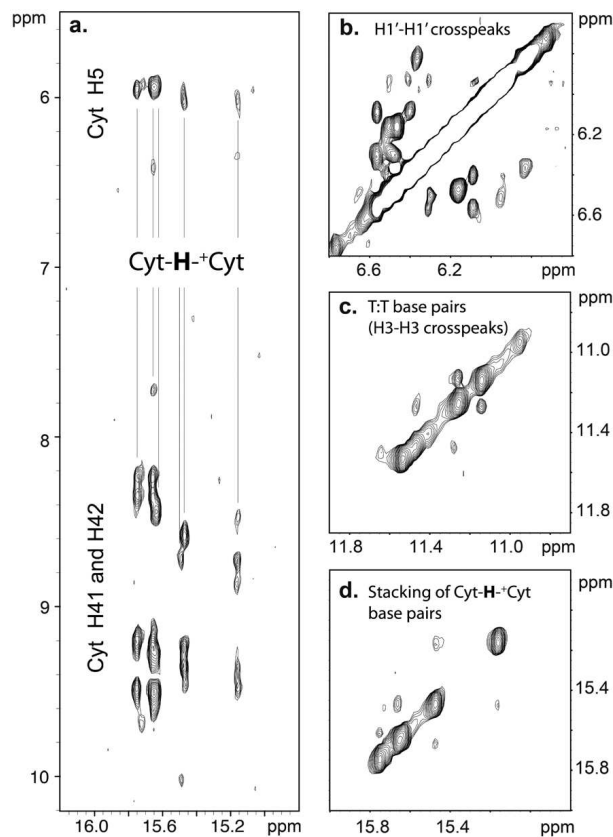


Fig. 3 Regions of NOESY spectra of CC at pH 4.0 and T = 5 1C, 90 : 10 H₂O/D₂O (mixing time: 150 ms). Buffer conditions were 20 mM potassium phosphate, 100 mM KCl, pH 4. (a) Detail of the hemiprotonated cytosine imino region, showing the formation of six C-C⁺ base pairs. (b) Region of the NOESY spectra, showing H10–H10 correlations characteristic of i-motif structures. (c) Detail of the imino region of the NOESY spectra, showing cross-peaks between thymine imino protons; and (d) imino–imino cross-peaks between adjacent C-C⁺ base pairs.

PAGE and SEC analysis

The overall conformational heterogeneity of the sequences studied at several pH values was studied by non-denaturing polyacrylamide gel electrophoresis (PAGE) (Fig. 4). The migrations at basic pH (8.3) of several sequences prone to form i-motif structures were similar to those of control sequences that cannot form this structure, indicative of unfolded oligonucleotides. On the other hand, the migrations at pH 5.0 showed an acceleration of the sequences containing the four CCC tracts as compared with the two control unfolded oligonucleotides. This acceleration is indicative of a more compact structure migrating faster, thus of the intramolecularly folded i-motif structure.²⁵ The same pattern of bands was observed for all the sequences considered in this study, which indicated a rather common structuration with equivalent stabilities under these conditions.

Finally, the size-exclusion chromatography (SEC) chromatograms showed that all sequences eluted as a single folded structure at pH 5.3, and 300 mM KCl (ESI⁺), under non-denaturing conditions. On the other hand, elution at pH 7.1 indicated that all sequences were present as an unfolded strand. Chromatographic analyses done at pH 4.7 showed the presence of significant peak tails that were explained as a result of the strong adsorption of DNA sequences on the silica surface due to the protonation of silanol groups²⁶ (ESI⁺). In fact, analyses done at pH 3.0 and 4.0 were unsuccessful because of the complete disappearance of chromatographic peaks.

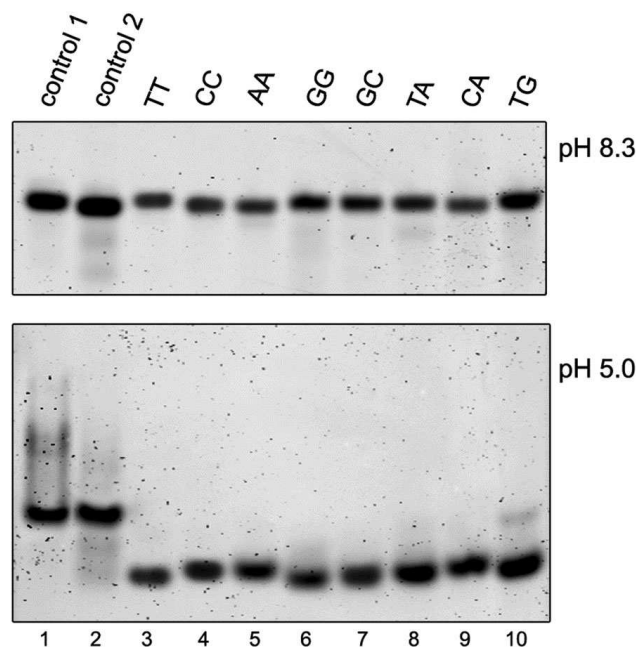


Fig. 4 PAGE analysis. The migrations at pH 8.0 of the sequences incubated at pH 7.8 show equal mobility, indicative of unfolded oligonucleotides. The migrations at pH 5.0 of the sequences incubated at pH 5.0 show an acceleration of those sequences containing the four CCC tracts (lanes 3–10) as compared with the two control oligonucleotides (lanes 1–2).

The results obtained until now indicate that the nature of bases in the considered sequences does not have a dramatic influence on the folding and molecularity of the i-motif structures formed.

Acid–base titrations

Now, to explore the effect of these bases on the pH-induced formation of i-motif structures spectrophotometrically monitored acid–base titrations were carried out in the pH range of 2–7 and at 25 °C. Fig. 5 shows the normalized absorbance values at 295 nm, a wavelength indicative of i-motif formation. Starting at pH 7, the absorbance for the sequence TT shows two leaps occurring in narrow pH ranges with pH-transition midpoints at ~ 6.5 and ~ 2.7, respectively. Other sequences, such as TA, CA or TG, show similar behaviour. In contrast, the sequence AA shows a continuous transition from pH 7 to pH 2.5, denoting the presence of at least one additional protonation process in this pH range. Finally, other sequences, such as GG, CC or GC, show an intermediate behaviour.

In an attempt to gain in-depth knowledge of the acid–base equilibria of these sequences, the recorded spectra were arranged in a **D** matrix and analysed by a multivariate data analysis method. This allowed the determination of the number of acid–base species and, for each one of these species, its pure spectrum and its individual concentration profile. The pure absorption spectrum provides qualitative information about the protonation state of bases in this acid–base species. On the other hand, the individual concentration profile provides information about the pH range of existence of this species, as well as the cooperativity associated with its formation (Table 1). Fig. 6 shows the calculated distribution diagrams and pure spectra of sequences TT and AA. The titration of TT comprises two acid–base equilibria, i.e., it may be explained by the linear combination of three acid–base species. Both equilibria are characterized by a high degree of cooperativity. The equilibrium observed at pH ~6.5 involves the i-motif and the unfolded strand where cytosines are fully deprotonated. As the

acid–base titrations were carried out at 25 °C, and the melting temperature of i-motif structures in this pH range is relatively low (~30 °C, see below), a mixture of species is expected. On the other hand, the equilibrium at pH ~ 2.7 involves the i-motif and the unfolded strand where cytosines are fully protonated.

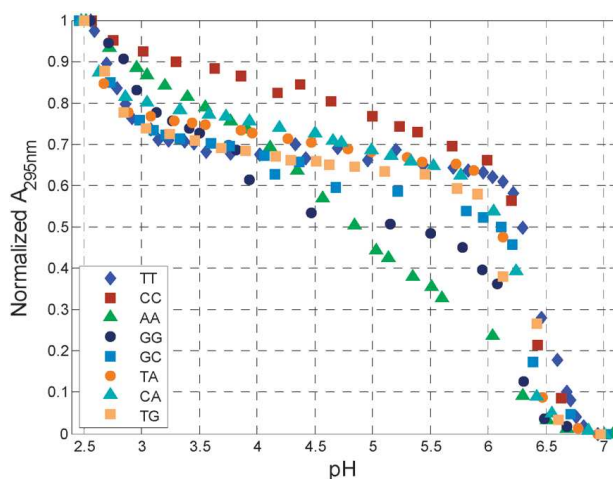


Fig. 5 Absorbance values recorded during the acid–base titration of the sequences studied at 25 °C.

Table 1. Calculated parameters from the multivariate analysis of spectra recorded along spectrophotometrically monitored acid-base titrations at 25°C.

Sequence	pH-Transition midpoint (p parameter) ^a	pH range predominance of the i-motif structure ^b
TT	6.47 ± 0.04 (3) 2.7 ± 0.1 (3)	3.8
CC	6.35 ± 0.03 (3) 4.5 ± 0.1 (1) 2.4 ± 0.1 (3)	3.9
AA	6.15 ± 0.04 (3) 4.59 ± 0.05 (1) 2.77 ± 0.07 (2)	3.4
GG	6.18 ± 0.04 (3) 3.02 ± 0.08 (1) 2.48 ± 0.1 (3)	3.4
GC	6.32 ± 0.07 (2) 4.0 ± 0.3 (1) 2.7 ± 0.3 (2)	3.6
TA	6.25 ± 0.03 (3) 4.4 ± 0.1 (1) 2.4 ± 0.1 (3)	3.8
CA	6.27 ± 0.03 (3) 4.5 ± 0.1 (1) 2.4 ± 0.2 (3)	3.9
TG	6.37 ± 0.06 (3) 3.9 ± 0.2 (1) 2.6 ± 0.3 (3)	3.8

^aAs commented before, this parameter describes the cooperativity of the transition.
^bCalculated from the difference between the first and the last pH-transition midpoints.

The introduction of bases other than thymine in the loops provokes the appearance of an additional protonation process with an apparent pK_a value close to that of cytosine (~ 4.5) or adenine (~ 3.5). A detailed inspection of the traces (Fig. 5) indicates that this intermediate protonation process is much more pronounced in AA than in other sequences, such as TA, CA or TG, where the effect is very small. The calculated distribution diagram and pure spectra of the AA sequence also reveal the presence of this intermediate transition (Fig. 6). The existence of this additional transition was validated by fitting the experimental data to a two pH-transition model (ESI⁺). The data not explained by the model of three components indicated that this intermediate transition actually exists. Also, a spectroscopically monitored titration of the T₂₅ sequence, which only contains thymine bases (pK_a around 9.5), has been carried out (ESI⁺). As expected, no transition was observed in the pH range from 7.4 to 2.6. Interestingly, the nature of the bases in the loops also affects the equilibrium involving the i-motif and the strand with deprotonated cytosines (Fig. 5 and ESI⁺). The TT and TG sequences show the highest pH-transition midpoints (6.5 and 6.4, respectively), whereas AA, GG and TA show the lowest values (~ 6.2). As a result of the effects in the two pH-transitions, the sequences with purines in both lateral loops (AA and GG) fold into the i-motif in the smallest pH range (~ 3.4 pH units) (Table 1).

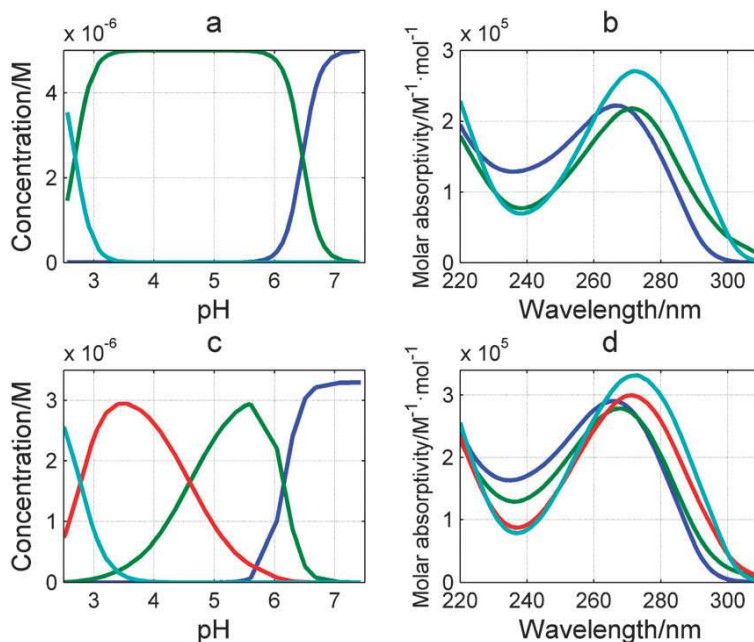


Fig. 6 Calculated distribution diagrams (matrix C) and pure spectra (matrix S) from multivariate analysis of data recorded along the acid–base titrations of TT (a and b, respectively) and AA (c and d, respectively) sequences at 25 °C.

To gain deeper knowledge about the effect of temperature on the acid–base equilibria, acid–base titrations of TT, AA, and CA sequences were also carried out at 37 °C (ESI⁺). The analysis by means of multivariate methods showed that the number of transitions and the cooperativity were similar to those observed at 25 °C. However, the pH-transition midpoint for the first transition (the one that involves the unfolded strand) was shifted to lower pH values, whereas the other transitions remained practically unaltered. For TT, the first transition was

shifted from pH 6.47 (25 1C) to 6.2 (37 °C), whereas the second transition took place at pH 2.7 for both temperatures. For AA, the first transition was shifted from pH 6.15 to 5.86, whereas the second and third transitions remained practically unaltered (pH 4.5 vs. 4.6, and pH 2.4 vs. 2.5). This fact was related to the low thermal stability of i-motif structures at pH values near neutral (see below).

Melting studies

Next, the stability of the i-motif structures formed by the considered sequences against temperature changes was evaluated. Therefore, spectrophotometrically monitored melting experiments were carried out at different pH values. The recorded melting curves exhibit a single melting transition at 295 nm, with little difference between the heating and cooling curves (ESI[†]). The general lack of hysteresis and the sigmoidal shape of the curves are indicative of reversible i-motif formation and relatively fast folding kinetics.²⁷ Under these conditions, reliable thermodynamic parameters from these experiments have been calculated.²⁸

It is possible to detect the presence of intermediates by using multivariate data analysis methods. For example, Fig. 7 shows the results of data analysis corresponding to the melting data of the CC sequence at pH 5.6. The presence of a process involving more than two states should be revealed by applying these procedures. From visual inspection of the trace at 295 nm (inset in Fig. 7a) a T_m value slightly lower than 50 °C may be determined. Table 2 summarizes the thermodynamic parameters of this folding, which have been calculated by using the usual methods based on a two-state process.

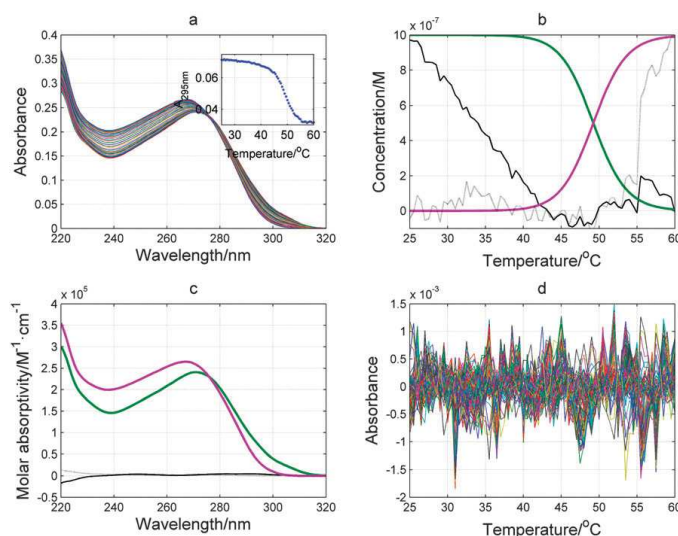


Fig. 7 Spectra recorded along the melting of the CC sequence at pH 5.6 (a). Inset: Absorbance trace at 295 nm. Calculated distribution diagram (matrix **C**) (b) and pure spectra (matrix **S**) (c). Absorbance data in matrix **E** not explained by the proposed model (d). The green and magenta components correspond to the folded and unfolded strands, respectively. The black components explain baseline drifts not related to the unfolding event.

Singular value decomposition (SVD) analysis suggested the presence of four spectroscopically active components in the data. These four components may be related directly to the unfolding process or may be just related to baseline drifts. In this case, a recently published approach based on hybrid modelling was applied.²⁹ Fig. 7b and c show the calculated distribution diagrams and pure spectra of each one of the four considered components. Component “green” is related to the initial i-motif structure, whereas the component

“magenta” corresponds to the unfolded strand. The calculated T_m , ΔH° and ΔS° values for this transition are 49.2 ± 0.3 °C, 95.2 ± 0.4 kcalmol⁻¹ and 295 ± 2 calK⁻¹ mol⁻¹, respectively, which agree with the values calculated from the classical univariate analysis. Components plotted in black explain baseline drifts at low and high temperatures, respectively. The residual variance (Fig. 7d) (matrix **E** in eqn (3)) shows both small values (around 0.001 absorbance units) and a random behavior, which means that the proposed model explains satisfactorily the experimental data. In summary, the unfolding of the CC sequence under these experimental and instrumental conditions maybe explained satisfactorily as a two-state process. This procedure has also been applied to other sequences at pH values where a mixture of acid–base species has been proposed (ESI⁺). In both cases, it has been possible to resolve the initial mixture of folded acid–base species and the calculation of thermodynamic parameters related to the unfolding process. The resolved spectra of the folded species, as well as the relative concentrations, are not exactly the same as expected from the acid–base distribution diagram. It is very difficult to match all data when the pure spectra of the components in the mixture are so similar.^{30,31} Moreover, the distribution diagram shows that these components are overlapped in the temperature direction, making more difficult the exact resolution of the system.

Table 2. Thermodynamic parameters calculated from thermal stability studies for the unfolding process. A two-state process was considered in the calculations. Uncertainty in T_m (°C) values was ± 0.7 °C. Uncertainties in ΔH (kcal/mol) and ΔS ((cal/K·mol) values were lower than ± 5 %. Uncertainty in ΔG (kcal/mol) values was lower than ± 10 %.

		pH 4.6					pH 5.1					pH 5.6					pH 6.2			
DNA	T_m	ΔH°	ΔS°	$\Delta G^\circ_{37^\circ\text{C}}$	T_m	ΔH°	ΔS°	$\Delta G^\circ_{37^\circ\text{C}}$	T_m	ΔH°	ΔS°	$\Delta G^\circ_{37^\circ\text{C}}$	T_m	ΔH°	ΔS	$\Delta G^\circ_{37^\circ\text{C}}$				
TT	62	83	248	6.1	55	89	271	4.9	47	96	298	3.6	34	91	295	-1.0				
CC	67	84	247	7.4	59	85	256	5.6	48	96	299	3.3	38	83	267	0.2				
AA	60	74	222	5.1	53	76	233	3.7	44	93	293	2.1	32	94	307	-1.2				
GG	59	79	238	5.2	55	83	253	4.5	45	93	292	2.4	31	92	304	-1.6				
GC	64	88	261	7.1	57	88	260	5.5	46	97	304	2.7	34	86	282	-1.5				
TA	65	74	219	6.1	56	87	264	5.1	46	89	279	2.5	33	92	301	-1.4				
CA	63	75	223	5.8	54	86	263	4.4	47	93	289	3.4	35	94	305	-0.6				
TG ₄	63	78	232	6.0	56	79	240	4.6	46	88	275	2.7	32	87	285	-1.4				

In terms of T_m values, CC forms the most stable structure throughout the pH range studied;¹⁵ whereas the GG and TG sequences form the least stable structures. The inclusion of additional cytosine bases in the loops slightly increases the stability of the i-motif, whereas the inclusion of purines clearly reduces their stability. The inclusion of a potential W·C base pair (GC or TA sequences) does not provoke a significant effect on the stability compared to the control sequence, TT. Analysis of the thermodynamic parameters shows that the values of ΔH° for the unfolding process at pH 6.2 range from ~ 83 (CC) to ~ 94 (AA or CA) kcal mol⁻¹. Considering that there are 6 C·C⁺ base pairs, and assuming that the disruption of each of them requires around 11 ± 1 kcal mol⁻¹ at pH ~ 6.0 ,³² it must be concluded that there are additional interactions contributing to the i-motif stability, like proton transfer equilibria or solvation. These contributions may be due, at least in part, to interactions involving loop residues. For example, T·T base pairs are observed in the NMR experiments. T·T based pairs have been found in many i-motif structures.³³ In addition to the two hydrogen bonds formed in each T·T base pair, these base pairs are isomorphous with the C·C⁺ ones, and fit well in the intercalative structure, contributing to its stability through favourable stacking interactions. In general, an increment of the ΔH° value was observed at lower pH values, with a maximal value around pH ~ 4.5 . As a general trend, $\Delta G^\circ_{37^\circ\text{C}}$ values were seen proportional to pH (ESI[†]). In terms of $\Delta G^\circ_{37^\circ\text{C}}$, the AA and GG sequences are the least stable structures in the pH range of 4.6–5.6. On the other hand, the CC sequence shows a higher stability throughout the whole pH range studied. Altogether, the results shown here suggest that the nature of the bases in the loops not only affects the stability of i-motif structures with respect to temperature, but also produces structural modifications that can be detected by studying the acid–base behaviour. In terms of T_m values, the CC, TT and CA sequences form the most stable i-motif structures at pH 5.6; whereas the AA and GG sequences form the least stable ones. Therefore, it may be concluded that the thermal stability of the i-motif structure decreases when two purine bases are placed opposite to each other. Concomitantly, the results obtained from the acid–base titrations show that the AA and GG sequences exhibit the lowest stability with respect to pH changes. These sequences fold into the i-motif structure in a narrower range of pH compared to the others (2.8–6.2 vs. 2.4–6.5). It has also recently been found for duplex DNA that the presence of mismatches produced significant local structural alterations, especially in the case of purine–purine mismatches,³⁴ due to their greater size.

Conclusions

Our results suggest that the thermal stability of intramolecular i-motifs and their pH dependence are clearly related to the composition of at least two of their loops. The presence of purine bases destabilizes the structure, whereas cytosine bases in these loops confer more stability at acidic pH. An additional acid–base transition has been observed at pH around 4.5 for all the considered sequences except for TT, which only contains cytosine and thymine bases. The pH range of the existence of i-motif structures is dependent both on the nature of the loops and on the temperature at which the measurement is carried out. Increasing the temperature reduces the pH at which takes place the transition involving the i-motif and the unfolded, deprotonated sequence. The different pH range in which these intramolecular i-motifs can fold may be useful for the development of pH responsive systems, such as selective delivery mechanisms based on molecular switches, biosensors and nanomachines.

Acknowledgements

Anna Sadurní (University of Barcelona) is acknowledged for carrying out some experiments. We acknowledge funding from the Spanish government (BFU2014-52864-R, CTQ2012-38616-CO2-02 and CTQ2014-52588-R). This research has been recognized by the regional Catalan authorities (2014 SGR 1106).

Notes and references

- 1 J. L. Leroy, M. Gueron, J. L. Mergny and C. Helene, *Nucleic Acids Res.*, 1994, 22, 1600–1606.
- 2 M. Garavís, N. Escaja, V. Gabelica, A. Villasante and C. González, *Chem. – Eur. J.*, 2015, 21, 9816–9824.
- 3 S. Benabou, A. Avino, R. Eritja, C. González and R. Gargallo, *RSC Adv.*, 2014, 4, 26956–26980.
- 4 J. Amato, N. Iaccarino, A. Randazzo, E. Novellino and B. Pagano, *ChemMedChem*, 2014, 9, 2026–2030.
- 5 Y. Dong, Z. Yang and D. Liu, *Acc. Chem. Res.*, 2014, 47, 1853–1860.
- 6 J. Huang, Y. He, X. Yang, K. Wang, L. Ying, K. Quan, Y. Yang and B. Yin, *Chem. Commun.*, 2014, 50, 15768–15771.
- 7 F. Xia, W. Guo, Y. Mao, X. Hou, J. Xue, H. Xia, L. Wang, Y. Song, H. Ji, Q. Ouyang, Y. Wang and L. Jiang, *J. Am. Chem. Soc.*, 2008, 130, 8345–8350.
- 8 Y. Yang, G. Liu, H. Liu, D. Li, C. Fan and D. Liu, *Nano Lett.*, 2010, 10, 1393–1397.
- 9 I. V. Nesterova and E. E. Nesterov, *J. Am. Chem. Soc.*, 2014, 136, 8843–8846.
- 10 T. A. Brooks, S. Kendrick and L. Hurley, *FEBS J.*, 2010, 277, 3459–3469.
- 11 S. P. Gurung, C. Schwarz, J. P. Hall, C. J. Cardin and J. A. Brazier, *Chem. Commun.*, 2015, 51, 5630–5632.
- 12 S. Nonin-Lecomte and J. L. Leroy, *J. Mol. Biol.*, 2001, 309, 491–506.
- 13 A. L. Lieblein, B. Fürtig and H. Schwalbe, *ChemBioChem*, 2013, 14, 1226–1230.
- 14 S. Fernández, R. Eritja, A. Aviñó, J. Jaumot and R. Gargallo, *Int. J. Biol. Macromol.*, 2011, 49, 729–736.
- 15 T. Fujii and N. Sugimoto, *Phys. Chem. Chem. Phys.*, 2015, 17, 16719–16722.
- 16 M. Piotta, V. Saudek and V. Sklenář, *J. Biomol. NMR*, 1992, 2, 661–665.
- 17 W. A. Kibbe, *Nucleic Acids Res.*, 2007, 35, W43–W46.
- 18 K. J. Breslauer, *Methods Enzymol.*, Academic Press, 1995, pp. 221–242.
- 19 J. Jaumot, V. Marchan, R. Gargallo, A. Grandas and R. Tauler, *Anal. Chem.*, 2004, 76, 7094–7101.
- 20 E. R. Henry, *Biophys. J.*, 1997, 72, 652–673.
- 21 I. Haq, B. Z. Chowdhry and J. B. Chaires, *Eur. Biophys. J.*, 1997, 26, 419–426.
- 22 R. D. Gray and J. B. Chaires, *Current Protocols in Nucleic Acid Chemistry*, John Wiley & Sons, Inc., 2011.
- 23 N. J. Greenfield, *Nat. Protoc.*, 2006, 1, 2527–2535.
- 24 R. Dyson, S. Kaderli, G. A. Lawrence, M. Maeder and A. D. Zuberbühler, *Anal. Chim. Acta*, 1997, 353, 381–393.
- 25 V. Mathur, A. Verma, S. Maiti and S. Chowdhury, *Biochem. Biophys. Res. Commun.*, 2004, 320, 1220–1227.
- 26 A. Mendez, E. Bosch, M. Rose and U. D. Neue, *J. Chromatogr. A*, 2003, 986, 33–44.
- 27 S. Rankin, A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian and S. Neidle, *J. Am. Chem. Soc.*, 2005, 127, 10584–10589.
- 28 J. L. Mergny and L. Lacroix, *Nucleic Acids Res.*, 1998, 26, 4797–4803.
- 29 R. Gargallo, *Anal. Biochem.*, 2014, 466, 4–15.
- 30 R. Tauler, A. Smilde and B. Kowalski, *J. Chemom.*, 1995, 9, 31–58.
- 31 R. Tauler, R. Gargallo, M. Vives and A. Izquierdo-Ridorsa, *Chemom. Intell. Lab. Syst.*, 1999, 46, 275–295.
- 32 J. L. Mergny, L. Lacroix, X. G. Han, J. L. Leroy and C. Helene, *J. Am. Chem. Soc.*, 1995, 117, 8887–8898.
- 33 M. Canalia and J. L. Leroy, *Nucleic Acids Res.*, 2005, 33, 5471–5481.
- 34 G. Rossetti, P. D. Dans, I. Gomez-Pinto, I. Ivani, C. González and M. Orozco, *Nucleic Acids Res.*, 2015, 43, 4309–4321.

Electronic Supplementary Information

Understanding the effect of loop mutations on the stability of the i-motif structure

Sanae Benabou¹, Miguel Garavís², Ramon Eritja³, Carlos González², Raimundo Gargallo^{1*}

1. Solution Equilibria and Chemometrics Group, Department of Analytical Chemistry, University of Barcelona, Diagonal 645, E-08028 Barcelona, Spain

2. Institute of Physical Chemistry "Rocasolano", CSIC, Serrano 119, E-28006 Madrid, Spain

3. Institute for Advanced Chemistry of Catalonia (IQAC-CSIC), CIBER-BBN Networking Centre on Bioengineering, Biomaterials and Nanomedicine, Jordi Girona 18-26, E-08034 Barcelona, Spain

Contents

S1. CD spectra

S2 NMR spectra

S3. SEC analysis in non-denaturing conditions

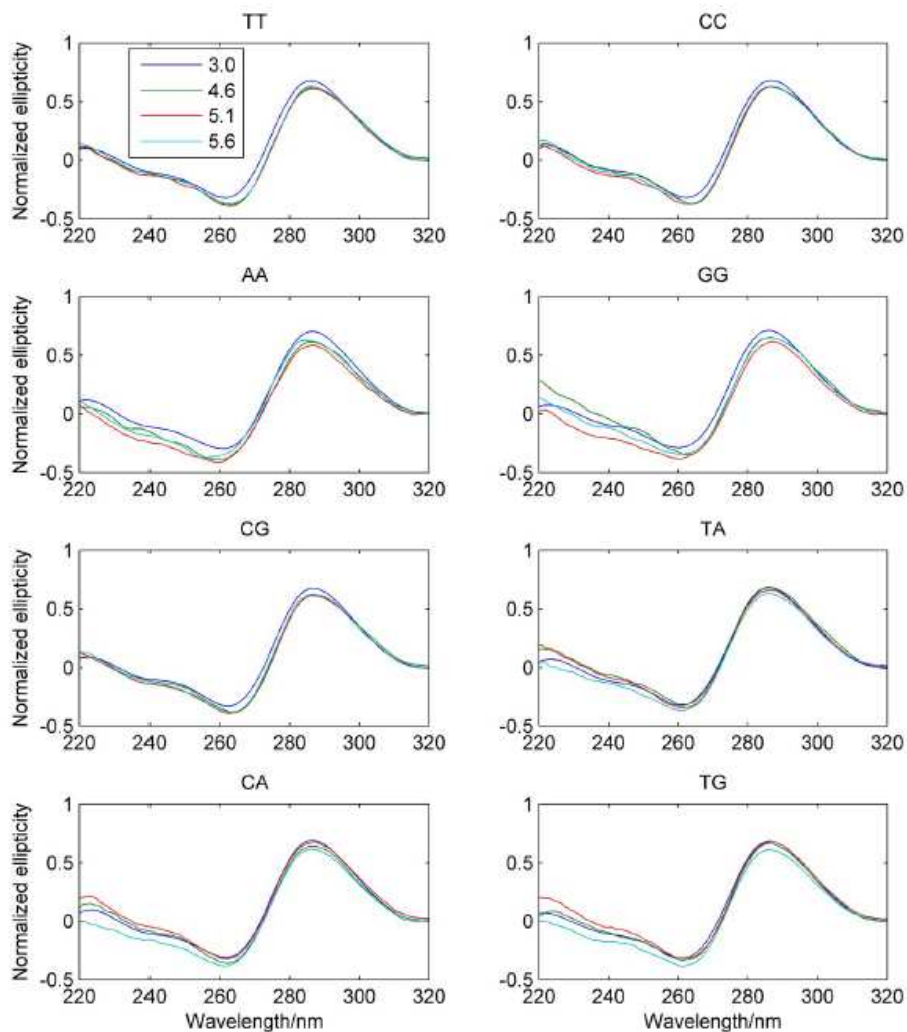
S4. Acid-base titrations

S5 Thermal stability studies

S1. CD Spectra

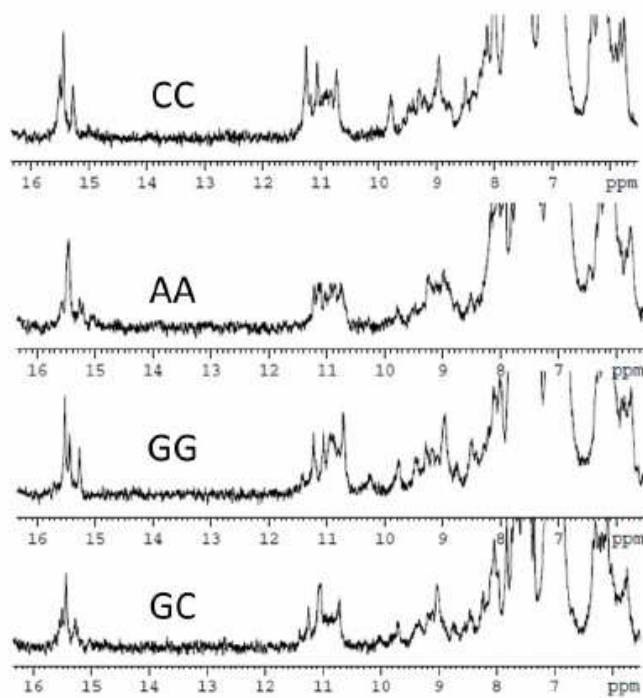
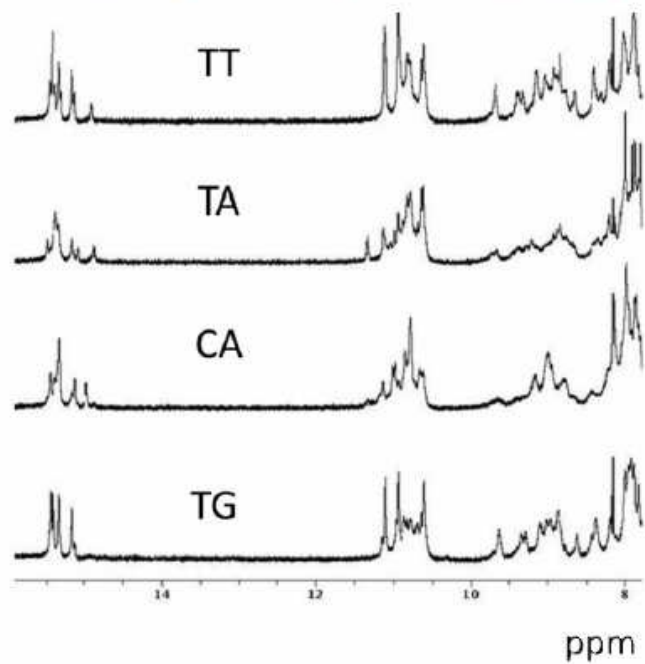
Normalized CD spectra measured at pH values 3.0, 4.6, 5.1, and 5.6 (20mM sodium acetate or phosphate buffer), 150mM KCl and 25°C. The normalization has been carried out to remove artifacts due to the potential differences in DNA concentration among different measurements. The normalization has been done as follows:

$$\text{Normalized CD spectrum} = \text{experimental CD spectrum} / (\text{maximum}(\text{ellipticity}) - \text{minimum}(\text{ellipticity})).$$

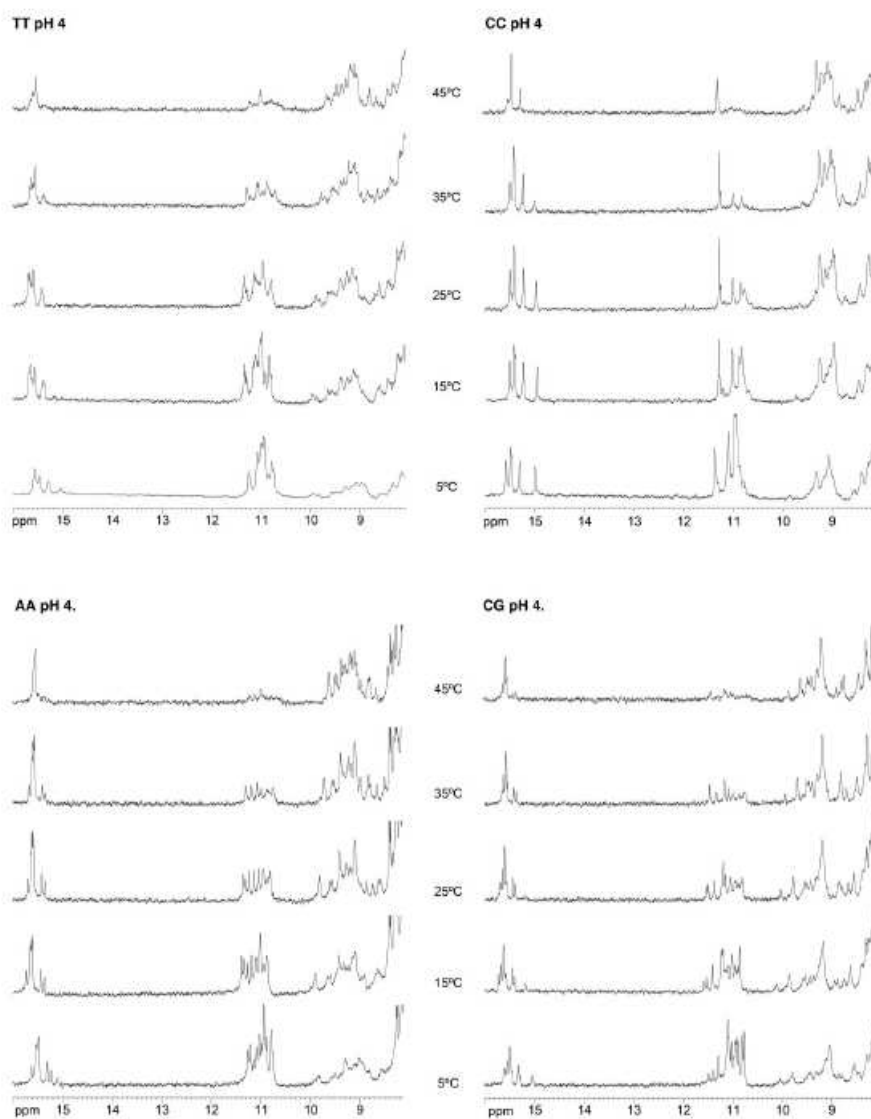


S2 NMR spectra

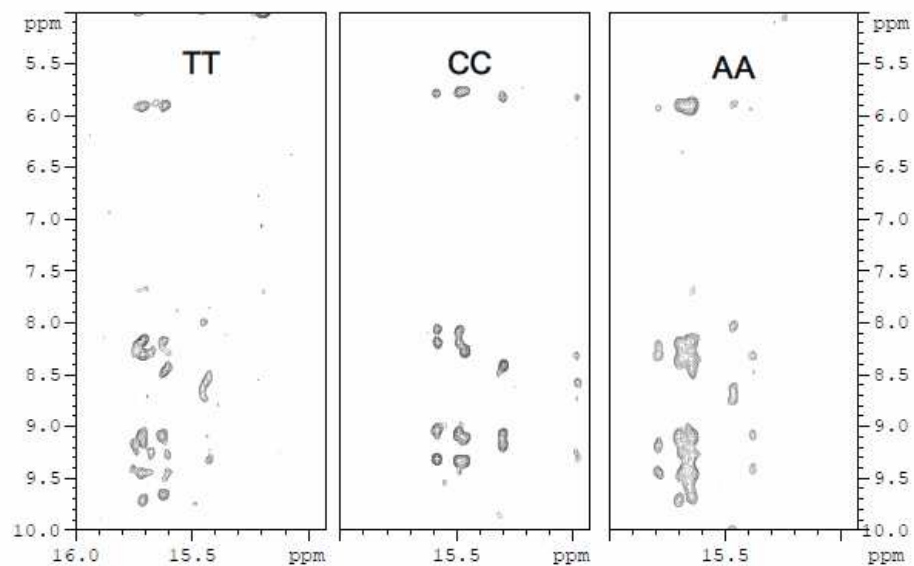
2.1. NMR spectra measured at pH 5.2, 20mM sodium phosphate buffer, 100mM KCl and 5°C. $C_{DNA} = 600 \mu M$



2.2. Exchangeable protons regions of the ^1H NMR spectra of TT (a), CC (b), AA (d) and CG (d) at different temperatures. Same buffer as in the previous figure.

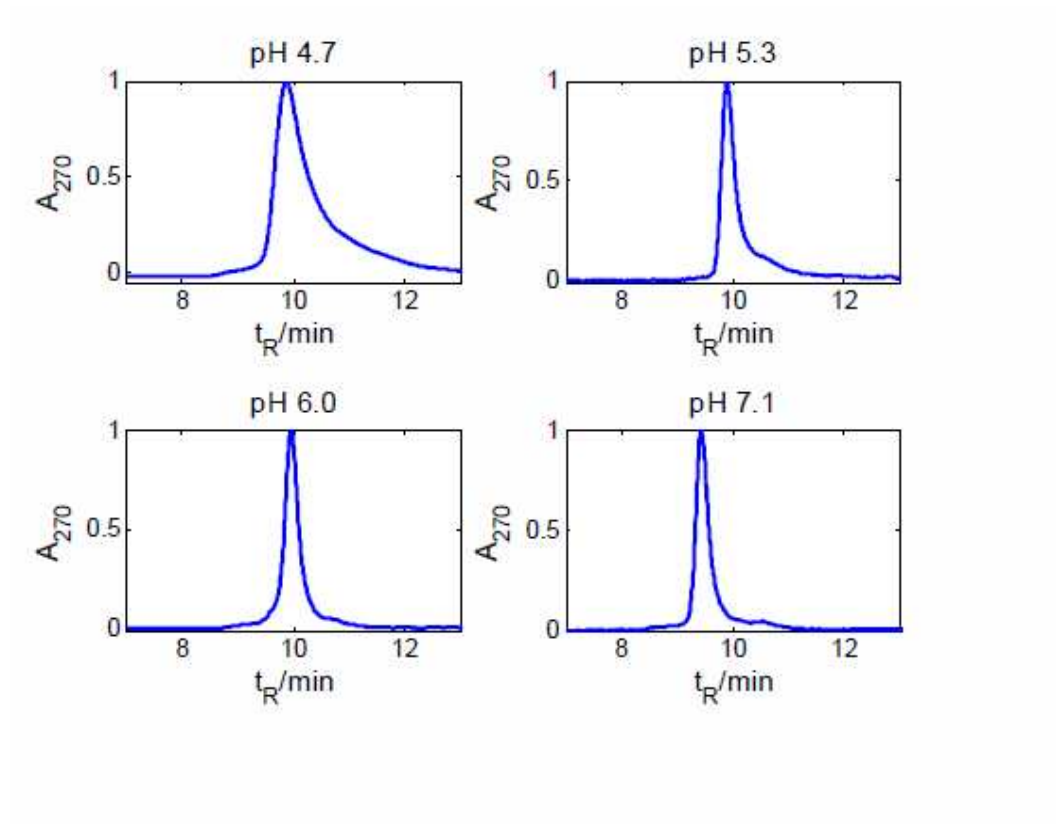


2.3. Cytosine imino/amino region of NOESY spectra of TT, CC and AA at pH 3.5 and T = 5°C, 90:10 H₂O/D₂O (mixing time: 150 ms). Buffer conditions were 20 mM potassium phosphate, 100 mM KCl. Differences in cytosine chemical shifts indicate that the loop sequence affects significantly the overall spectra.



3. SEC analysis in non-denaturing conditions

3.1 SEC chromatograms for GC sequence at several pH values

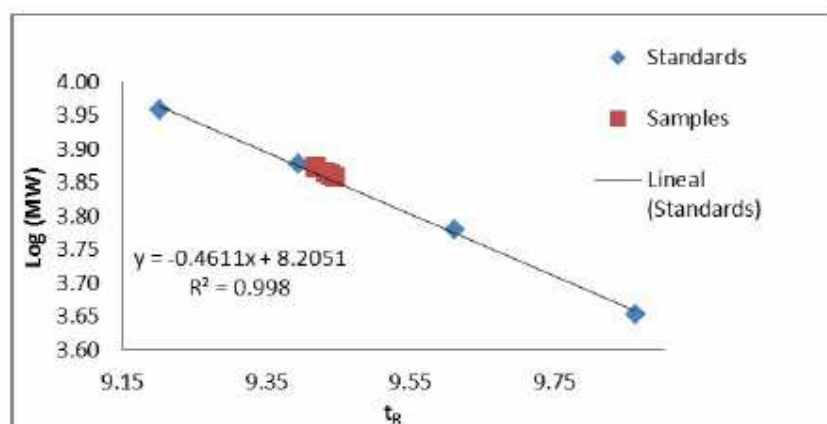


3.2. SEC data and calibration plot at pH 7.1 (phosphate buffer).

Standards **MW (g/mol)** **log (MW)** **Measured t_R (min)**

T15	4501.0	3.6533	9.86
T20	6022.0	3.7797	9.61
T25	7542.9	3.8775	9.39
T30	9063.9	3.9573	9.20

Samples	MW (g·mol⁻¹)	log(MW)	Measured t_R	Calculated t_R	Proposed structure
TT	7362.8	3.867	9.44	9.43	Unfolded sequence
CC	7332.8	3.865	9.42	9.43	Unfolded sequence
AA	7380.8	3.868	9.42	9.43	Unfolded sequence
GG	7412.8	3.870	9.44	9.42	Unfolded sequence
GC	7372.8	3.867	9.43	9.43	Unfolded sequence
TA	7372.0	3.867	9.44	9.43	Unfolded sequence
CA	7357.0	3.866	9.44	9.43	Unfolded sequence
TG	7388.0	3.868	9.45	9.44	Unfolded sequence

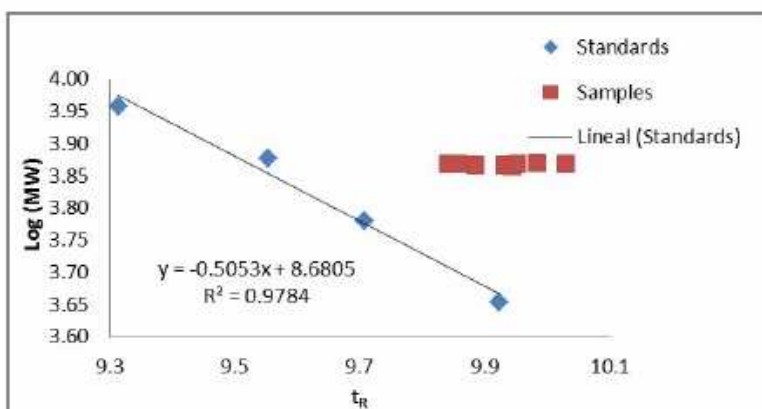


This graph represents the dependence of $\log(MW)$ with t_R for a series of four standard (T₁₅, T₂₀, T₂₅ and T₃₀). These standards do not show a folded structure in these experimental conditions. The variation of $\log(MW)$ with t_R shows a linear dependence. The samples eluted at t_R values equal (or very similar) to those calculated from the regression line, which means that these samples are not folded in this experimental conditions.

3.3. SEC data and calibration plot at pH 6.1 (phosphate buffer).

Standards	Measured t_R (min)		
	MW ($\text{g}\cdot\text{mol}^{-1}$)	log(MW)	
T15	4501.0	3.6533	9.92
T20	6022.0	3.7797	9.71
T25	7542.9	3.8775	9.55
T30	9063.9	3.9573	9.31

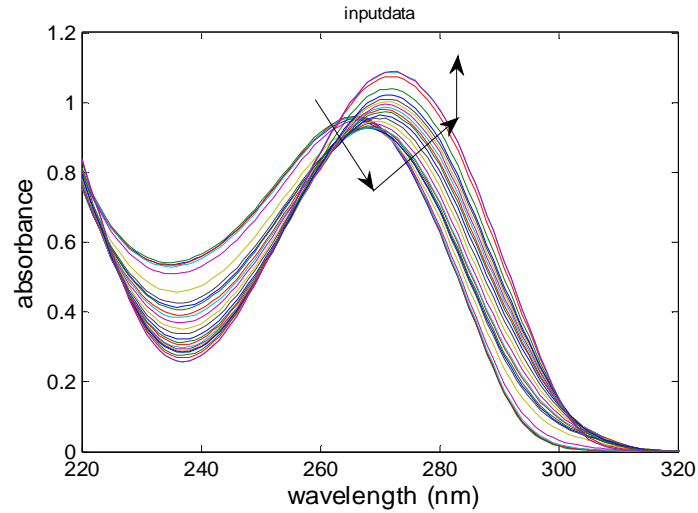
Samples	MW ($\text{g}\cdot\text{mol}^{-1}$)	log(MW)	Measured t_R	Calculated t_R	Proposed structure
TT	7362.8	3.867	9.93	9.53	Unfolded sequence
CC	7332.8	3.865	9.94	9.53	<i>i</i> -motif
AA	7380.8	3.868	10.03	9.52	<i>i</i> -motif
GG	7412.8	3.870	9.98	9.52	<i>i</i> -motif
GC	7372.8	3.867	9.96	9.52	<i>i</i> -motif
TA	7372.0	3.867	9.86	9.43	<i>i</i> -motif
CA	7357.0	3.866	9.89	9.43	<i>i</i> -motif
TG	7388.0	3.8678	9.84	9.42	<i>i</i> -motif



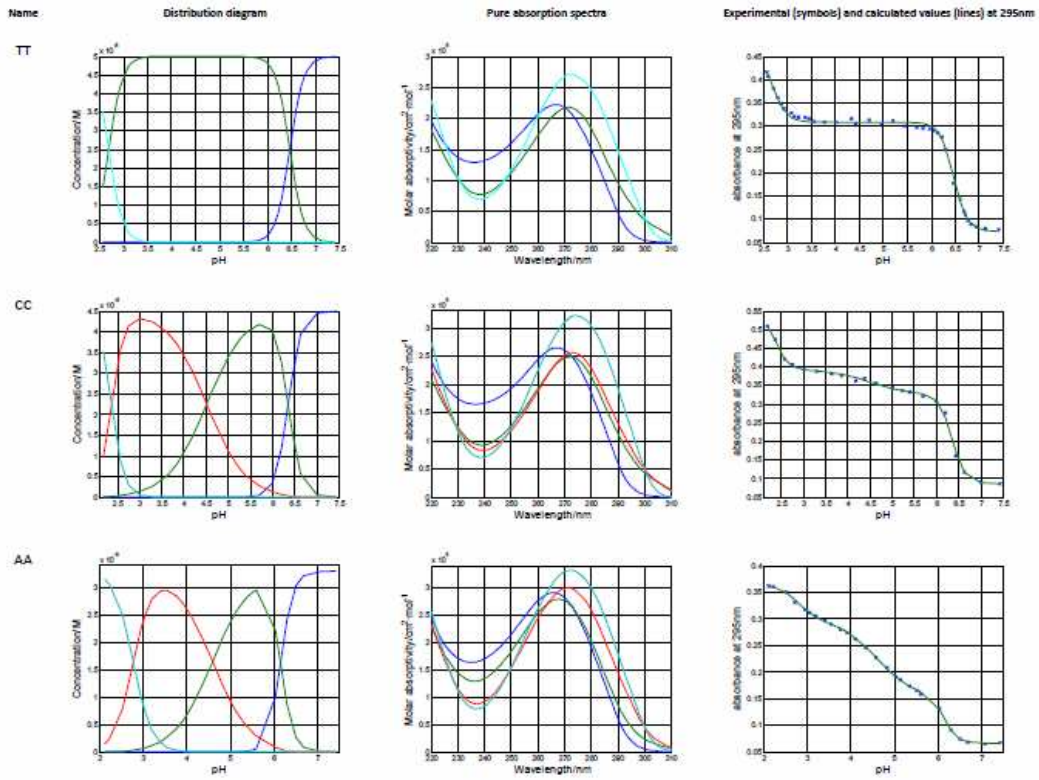
As in the previous graph (pH 7.1), the dependence of log(MW) with t_R for a series of four standards (T_{15} , T_{20} , T_{25} , and T_{30}) is shown. Again, these standards do not show a folded structure in these experimental conditions, and the variation of log(MW) with t_R shows a linear dependence. However, at pH 5.3, the samples eluted at t_R values greater than those expected for unfolded strands, which means that the hydrodynamic volume of samples is smaller than at pH 7.1. This fact is related to the folding into *i*-motif structures

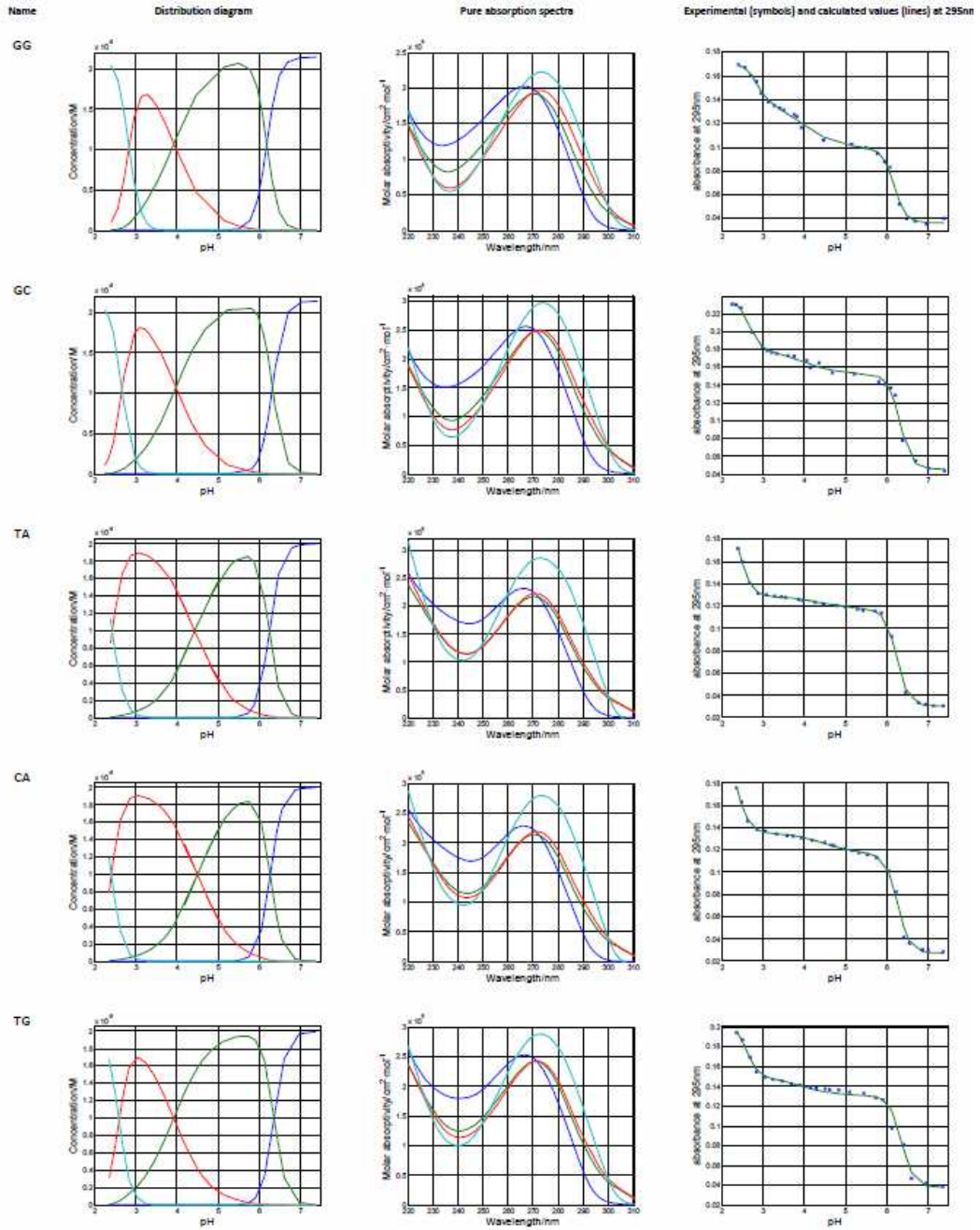
S4. Acid-base titrations.

4.1. Experimental spectra recorded throughout the titration of the AA sequence. Arrows indicate the absorbance changes when pH is lowered from pH 7 to 2.5 at 25°C.

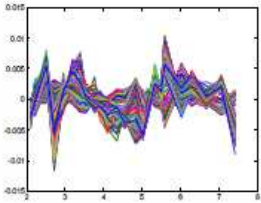
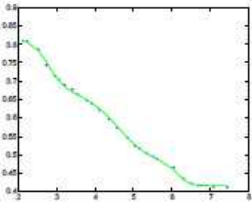
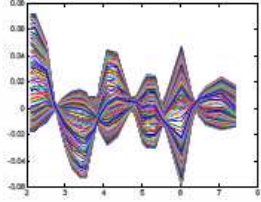
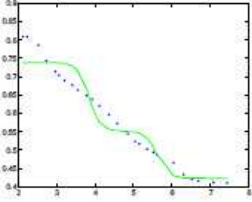
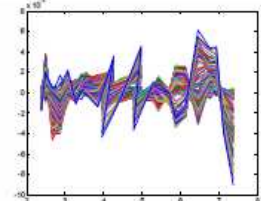
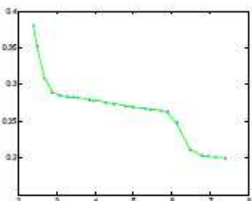
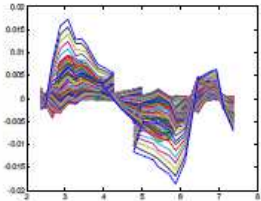
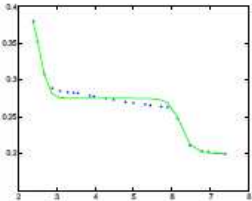


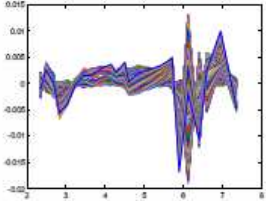
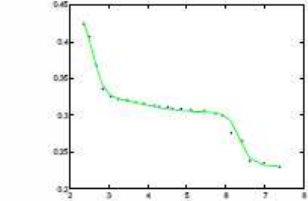
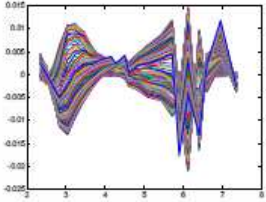
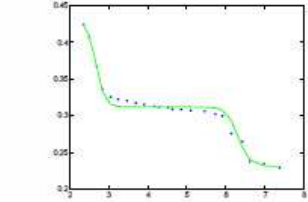
4.2. Calculated distribution diagrams and pure spectra from multivariate analysis of data recorded along acid-base titrations of the considered sequences at 25 °C.



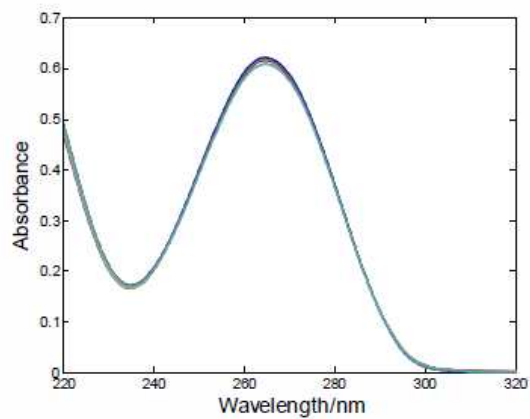


4.3. Validation of the existence of the additional transition for several acid-base titrations carried out at 25 °C.

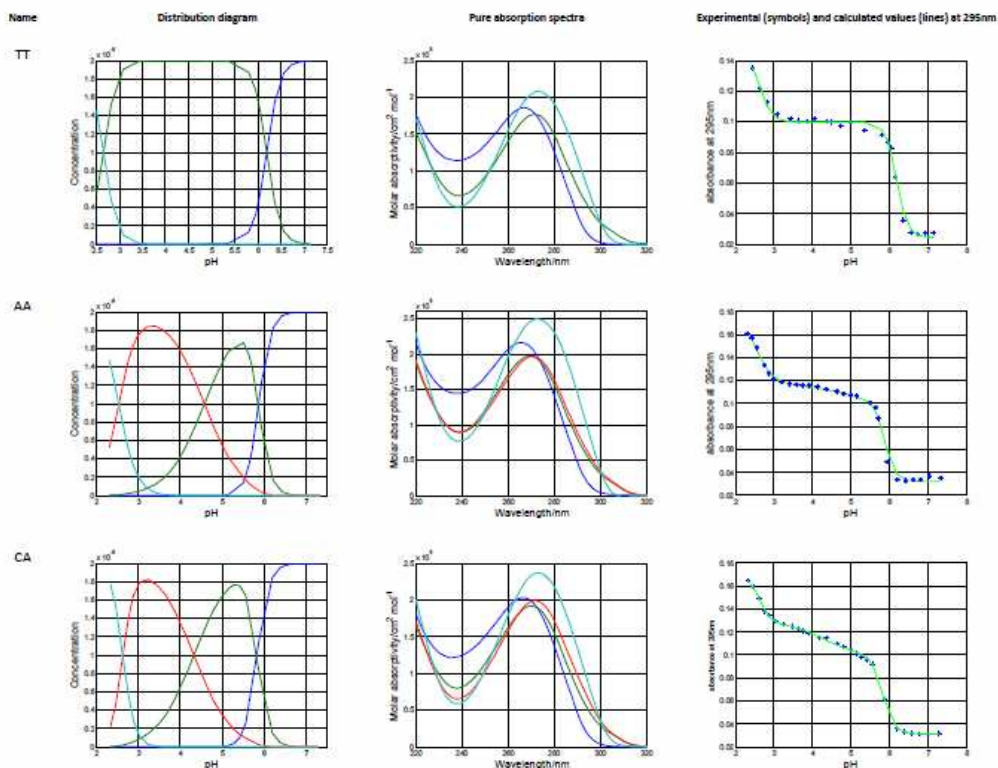
Sequence	Number of components (nc) considered in the model	Sum of squares in matrix E	Plot of absorbance data not explained by the proposed model (matrix E)	Fits at 295nm
				Symbols represent the experimental absorbance values. Lines represent the calculated absorbance values with the proposed model
AA	4	0.020		
	3	0.933		
TA	4	0.006		
	3	0.028		

Sequence	Number of components (nc) considered in the model	Sum of squares in matrix E	Plot of absorbance data not explained by the proposed model (matrix E)	Fits at 295nm
				Symbols represent the experimental absorbance values. Lines represent the calculated absorbance values with the proposed model
TG	4	0.035		
	3	0.063		

4.4. Absorbance spectra of T_{25} sequence measured from pH 7.4 to pH 2.6 at 25°C. The shape of the spectrum remains unaltered, and only the intensity shows a small variation due to dilution.



4.5. Calculated distribution diagrams and pure spectra from multivariate analysis of data recorded along acid-base titrations of the considered sequences at 37°C.



Calculated parameters from the multivariate analysis of spectra recorded along spectrophotometrically-monitored acid-base titrations at 37°C.

Name	pH-transition midpoint (p parameter ^a)	pH range predominance of the <i>i</i> -motif structure ^b
TT	6.2±0.04 (3)	3.5
	2.7±0.1 (3)	
AA	5.86±0.04 (3)	3.4
	4.6±0.2 (1)	
CA	2.5±0.3 (3)	3.2
	5.83±0.02 (3)	
	4.3±0.1 (1)	
	2.6±0.1 (3)	

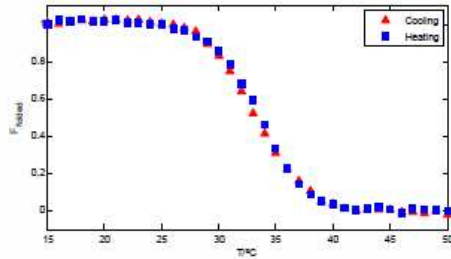
^a As commented before, this parameter describes the cooperativity of the transition.

^b Calculated from the difference between the first and last pH-transition midpoints.

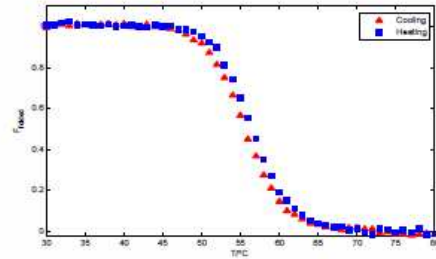
S5. Thermal stability studies

5.1. Normalized heating and cooling curves. Absorbance spectra were recorded at 1°C intervals with a hold time of 3 min at each temperature, which yielded an average heating rate of approximately 0.3°C min⁻¹. Buffer solutions were 20 mM phosphate or acetate, and 150 mM KCl.

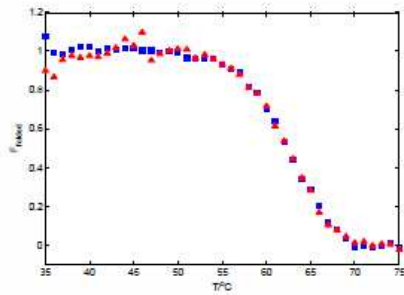
TT sequence at pH 6.2



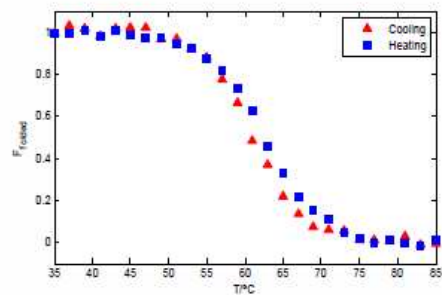
GC sequence at pH 5.1



TT sequence at pH 4.6



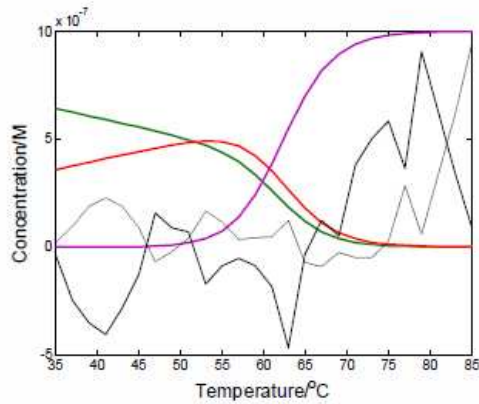
AA sequence at pH 4.6



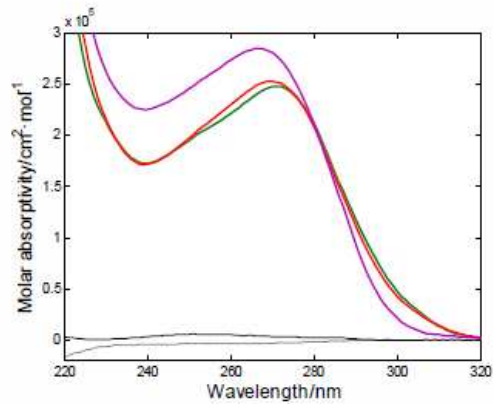
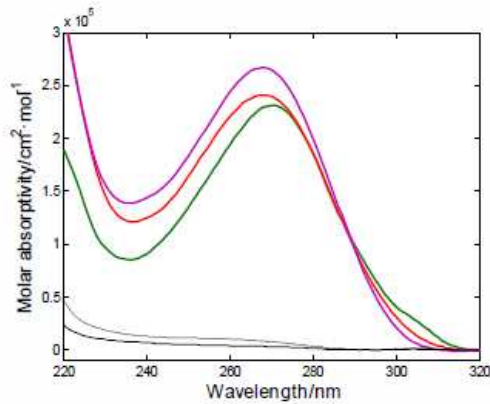
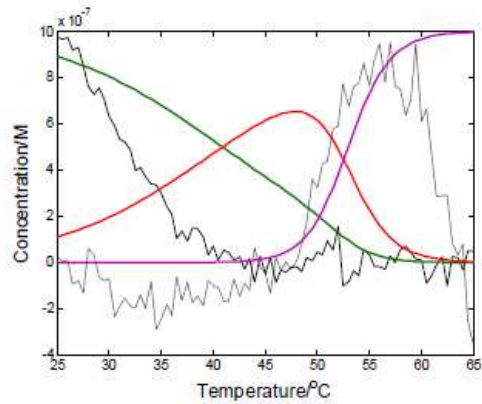
5.2. Multivariate analysis of melting data by means of hybrid modelling.

Green and red thick lines correspond to acid-base species in acid-base distribution diagrams. Magenta thick line represents the unfolded strand. Thin black lines correspond to mathematical components not related with the unfolding process, just explaining noise and baseline drifts.

Distribution diagram for the melting of AA sequence at pH 4.6



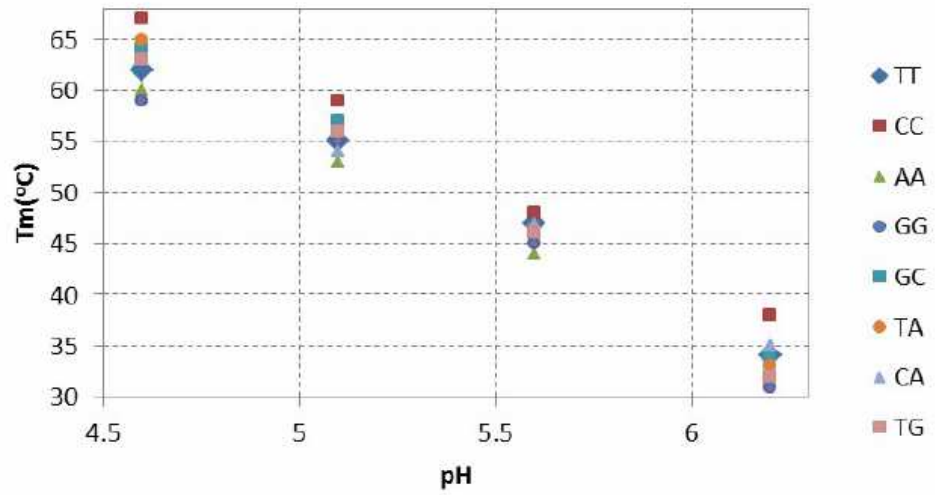
Distribution diagram for the melting of CA sequence at pH 5.1



Calculated parameters for the unfolding process:

System	First transition			Second transition		
	ΔH° (kcal·mol ⁻¹)	ΔS° (cal·K ⁻¹ ·mol ⁻¹)	T_m (°C)	ΔH° (kcal·mol ⁻¹)	ΔS° (cal·K ⁻¹ ·mol ⁻¹)	T_m (°C)
AA at pH 4.6	6.9±0.1	21±1	52±2	71.6±0.5	215±2	60.7±0.8
CA at pH 5.1	24.5±0.2	78±1	40.9±0.7	100.9±0.6	310±4	52.5±0.7

5.3. Melting temperatures corresponding to Table 1 (main text).



5.4. ΔG° dependence upon pH. Thermal stability studies were carried out at 20mM phosphate or acetate buffer and 150mM KCl.

