

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

Received 00th January 20xx,  
Accepted 00th January 20xx

Sanae Benabou<sup>a</sup>, Miguel Garavís<sup>b</sup>, Sébastien Lyonnais<sup>c</sup>, Ramon Eritja<sup>d</sup>, Carlos González<sup>b,e</sup>,  
Raimundo Gargallo<sup>a,e\*</sup>

DOI: 10.1039/x0xx00000x

www.rsc.org/

The nature and 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 of an intramolecular i-motif is studied using spectroscopy, separation techniques, and multivariate data analysis. Insertion of bases other than thymine induces an additional acid-base equilibrium with  $pK_a \sim 4.5$ . The presence of two guanine bases in the loops, 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<sup>+</sup> 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 medium<sup>1</sup>. 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.<sup>2-4</sup> The study of these structures is not only interesting from a biophysical and biomedical point of view<sup>4</sup>, but also for their potential application in nanotechnology<sup>5</sup>. The i-motifs have been employed in the design of *in vivo* sensing systems<sup>6</sup>, artificial DNA-based nanomachines<sup>7</sup> and logic gates<sup>8</sup>, 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 sensitivity<sup>9</sup>.

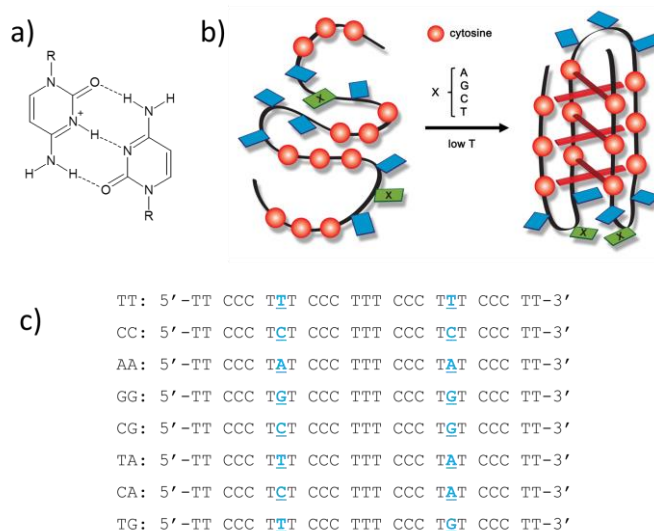


Figure 1. (a) Cytosine-protonated cytosine base pair; (b) Hypothetical scheme of the intramolecular i-motif structures adopted by the sequences studied in this work; X=A, C, G or T bases; (c) Sequences studied in this work

The number and nature of the bases located in the loops may affect the stability of i-motif structures. It has been suggested that these structures may be classified into two groups depending on the length of loops<sup>10</sup>: i-motifs showing short loop regions among cytosine tracts (class 1), and i-motifs showing longer loop regions (class 2). Class 2 structures are more stable than class 1, although recent results have casted some doubts on the generality of this effect<sup>11</sup>. Using sequences that only contain thymine and cytosine bases it has been reported that shorter loops exhibit highest stability. In this sense, flexibility related to long loop regions could be detrimental to the stability of the i-motif structure, and long loop regions 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 an additional favourable entropic contribution to the folding

<sup>a</sup> Department of Analytical Chemistry, University of Barcelona, Diagonal 645, E-08028 Barcelona, Spain

<sup>b</sup> Institute of Physical Chemistry "Rocasolano", CSIC, Serrano 119, E-28006 Madrid, Spain

<sup>c</sup> Institute of Molecular Biology of Barcelona (IBMB-CSIC), Baldori i Reixac 4-8, 08028 Barcelona, Spain

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

<sup>e</sup> BIOESTRAN, associated unit UB-CSIC

Electronic Supplementary Information (ESI) available: it includes experimental information and detailed figures. See DOI: 10.1039/x0xx00000x

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<sup>12, 13</sup>. In contrast, the mutation of the TAA repeats for TTT in intramolecular i-motif structure formed by the cytosine-rich strand of the human telomere increases its stability<sup>14</sup>. The influence of bases adjacent to the cytosine-tracts, as well as the effect of loop length, has been addressed in a recent work<sup>15</sup>. 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<sup>+</sup> base pair also stabilizes the i-motif. To our knowledge, the present work is the first study focused in 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 (Figure 1) as a model to study these interactions.

## 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, 3000 ml volume) were used. NMR spectra were acquired either in a Bruker Advance spectrometer operating at 600 MHz (sequences CC, AA, GG, and CG) or in a Bruker Digital Avance 800 MHz (sequences TT, TA, CA and TG). Water suppression was achieved by the inclusion of a WATERGATE<sup>16</sup> 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 ml 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- $\mu$ L 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  $\times$  7.8 mm, particle size 5  $\mu$ m and pore size 290  $\text{\AA}$ ) from Phenomenex (Torrance, CA, USA) was used for the chromatographic separation at room temperature.

### Reagents

The DNA sequences (Figure 1) were obtained from SigmaAldrich (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°C using the extinction coefficients calculated using the nearest-neighbor method as implemented on the OligoCalc webpage

<sup>17</sup>. Before any experiment, DNA solutions were first heated to 90°C for 10 minutes and then allowed to reach room temperature. KCl, KH<sub>2</sub>PO<sub>4</sub>, K<sub>2</sub>HPO<sub>4</sub>, NaCH<sub>3</sub>COO, HCl and NaOH (a.r.) were purchased from Panreac (Spain). MilliQ<sup>®</sup> 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°C and 150 mM KCl. 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 the 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°C intervals with a hold time of 3 min at each temperature, which yielded an average heating rate of approximately 0.3°C min<sup>-1</sup>. 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  $\mu$ M in a solution buffered at pH(5.0) (50mM sodium acetate pH(5.0), 10 mM potassium acetate) or at pH(7.8) (50mM Tris acetate pH(7.8), 10mM potassium acetate). The solutions were next boiled for 5minutes at 95°C and cooled down to room temperature. Glycerol was added (2.5% v/v) and 5 $\mu$ L of the samples were loaded on 10x10.5 cm - 12% non-denaturing polyacrylamide gels (19/1 acrylamide / bisacrylamide, Sigma) either buffered with 50mM Tri-sodium 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°C with a miniVE apparatus (Hoeffer) for 4h. at 11V/cm. 0.1M Tri-sodium citrate pH(5.0) was used a running buffer for the acidic gel and 0.5x TBE for the other. The gel buffered at pH(5.0) was then incubated 15min. in 1X TBE, rinsed with water and both gels were stained with SybrGold (Molecular Probes) according to the manufacturer's instructions and digitalized with a Typhoon 8600 system (Molecular Dynamics).

### Size-Exclusion Chromatography

The mobile phase was 300 mM KCl and 20mM buffer (phosphate or acetate) adjusted to desired pH value. The flow rate was set to 1.0 mL min<sup>-1</sup>. The injection volume was 15  $\mu$ L. 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 12.5min). T<sub>15</sub>, T<sub>20</sub>, T<sub>25</sub> and T<sub>30</sub> 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<sup>18</sup>. 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  $\Delta H_{\text{vH}}$  and  $\Delta S_{\text{vH}}$  will not change throughout the range of temperatures studied here.

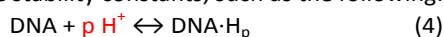
### Analysis of spectra recorded along acid-base titrations

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 for 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 intensity of the pure spectra may provide qualitative information about the structure of the species. With this goal in mind, data matrix **D** was decomposed according to Beer-Lambert-Bouer's law in matrix form:

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

where **C** is the matrix ( $m \times nc$ ) containing the distribution diagram, **S<sup>T</sup>** 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<sup>T</sup>**, 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)<sup>19-23</sup>. 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:



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

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<sup>T</sup>** 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<sup>24</sup>.

## 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 around 265 nm and a stronger positive band around 287 nm) was clear for all the considered sequences (Electronic Supplementary Information, 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) (Figure 2 and ESI). The imino proton resonances at ~15 ppm indicate the presence of protonated cytosines and the formation C-C<sup>+</sup> 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 (Figure 2).

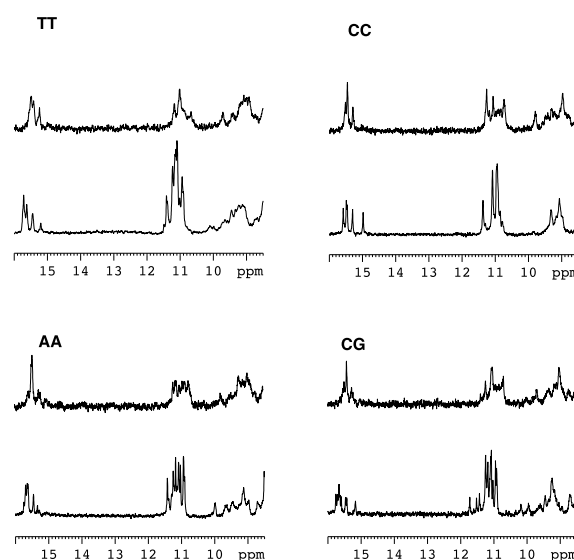


Figure 2. Exchangeable proton region of the <sup>1</sup>H NMR spectra of TT, CC, AA and CG recorded at pH 5.0 (top spectra in each panel) and pH 3.5 (bottom) (20 mM potassium phosphate, 100 mM KCl, 5°C, 90:1 H<sub>2</sub>O/D<sub>2</sub>O).

The imino signals are visible even at 45°C (Figure S2 in ESI), showing the relatively high thermo stability of the i-motifs formed. Analysis of NMR spectra revealed the formation of six C-C<sup>+</sup> base pairs (Figure 3a). Other NOEs characteristic of i-motif structures are clearly observable (i.e.: H1'-H2' cross-peaks in Figure 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 (Figure 2). Interestingly, all sequences exhibit sharp imino signals 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 (Figure 3c). Some differences are observed between NMR spectra acquired at pH 3.5 and 5.0, being such changes more pronounced in sequences different than TT (Figure 2). It is also worth noticing the significant differences in cytosine imino chemical shifts (illustrated in Figure S2.3) between oligonucleotides whose sequences differ only in loop nucleotides.

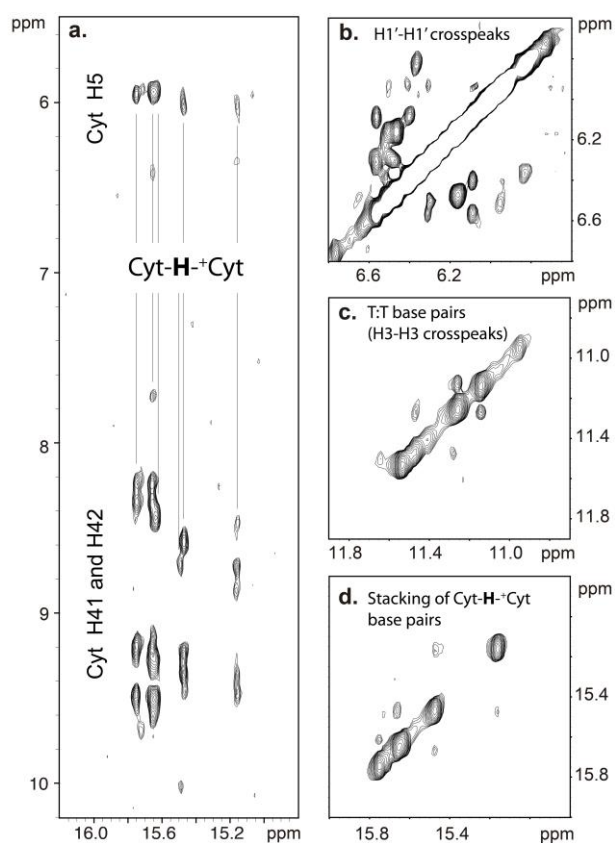


Figure 3. Regions of NOESY spectra of CC at pH 4.0 and  $T = 5^{\circ}\text{C}$ , 90:10  $\text{H}_2\text{O}/\text{D}_2\text{O}$  (mixing time: 150 ms). Buffer conditions were 20 mM potassium phosphate, 100 mM KCl, pH 4. a) Detail of the hemiprotonated cytosine imino region, showing the formation of six C-C<sup>+</sup> base pairs. b) Region of the NOESY spectra, showing H1'-H1' 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<sup>+</sup> 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) (Figure 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 intra-molecularly folded i-motif structure<sup>25</sup>. The same pattern of bands was observed for all the sequences considered in this study, which indicated a rather common structuration with equivalent stabilities in these conditions.

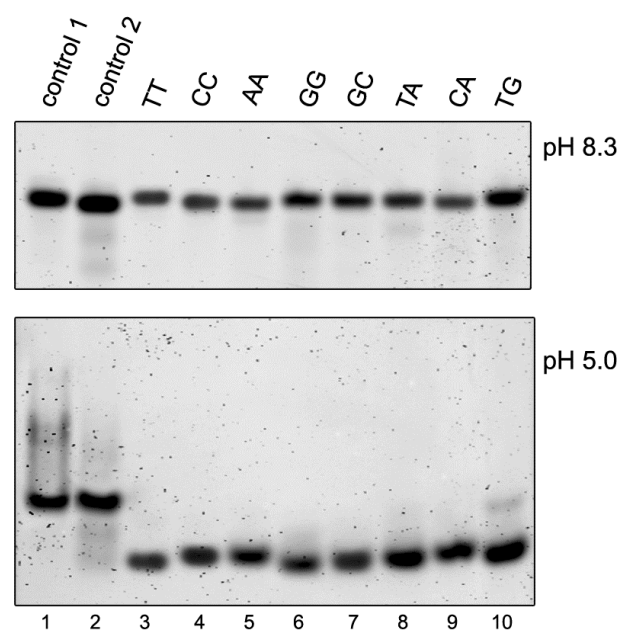


Figure 4. PAGE analysis. The migrations at pH 8.0 of the sequences incubated at pH 7.8 show equal mobility, indicative of unfolded oligonucleotides. The migrations at pH 5.0 of the sequences incubated at pH 5.0 show an acceleration of those sequences containing the four CCC tracts (lanes 3-10) as compared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

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), in 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 groups<sup>26</sup> (ESI). In fact, analyses done at pH 3.0 and 4.0 were unsuccessful because of the complete disappearance of chromatographic peaks.

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 range in the pH range 2-7 and 25°C. Figure 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.

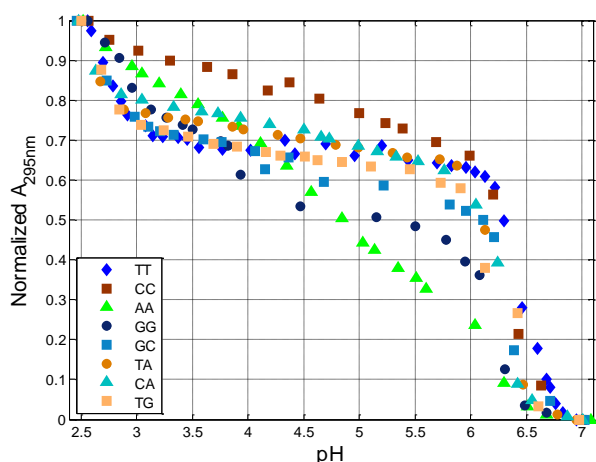


Figure 5. Absorbance values recorded during the acid-base titration of the sequences studied at 25°C.

In an attempt to go in depth in the 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 to its formation (Table 1). Figure 6 shows the calculated distribution diagrams and pure spectra for 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.

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 those of cytosine (~4.5) or adenine (~3.5). A detailed inspection of the traces (Figure 5) indicates 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 for AA sequence also reveals the presence of this intermediate transition (Figure 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 T<sub>25</sub> sequence, which only contains thymine bases ( $pK_a$  around 9.5), has been done (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 (Figure 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 i-motif in the smallest pH range (~3.4 pH units) (Table 1).

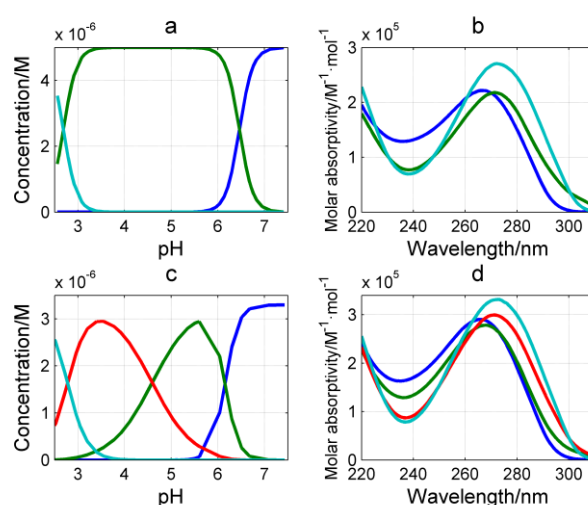


Figure 6. Calculated distribution diagrams (matrix **C**) and pure spectra (matrix **S**) from multivariate analysis of data recorded along the acid-base titrations of TT (a and b, respectively) and AA (c and d, respectively) sequences at 25°C.

To gain a 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 was 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°C) 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).

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) <sup>a</sup>	pH range predominance of the i-motif structure <sup>b</sup>
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) 3.92±0.08 (1) 2.8±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

<sup>a</sup> As commented before, this parameter describes the cooperativity of the transition. <sup>b</sup> Calculated from the difference between the first and last pH-transition midpoints.

### 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. 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<sup>27</sup>. In these conditions, reliable thermodynamic parameters from these experiments have been calculated<sup>28</sup>.

It is possible to detect the presence of intermediates by using multivariate data analysis methods. For example, Figure 7 shows the results of data analysis corresponding to the melting data of 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 Figure 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.

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<sup>29</sup>. Figure 7b and 7c show the calculated distribution diagrams and pure spectra for 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  $T_m$ ,  $\Delta H^0$  and  $\Delta S^0$  values for this transition are 49.2±0.3°C, 95.2±0.4 kcal·mol<sup>-1</sup> and 295±2 cal·K<sup>-1</sup>·mol<sup>-1</sup>, 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 (Figure 7d) (matrix E in equation 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 at 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 for 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<sup>30,31</sup>. Moreover, the distribution diagram shows that these components are overlapped in the temperature direction, making more difficult the exact resolution of the system.

In terms of  $T_m$  values, CC forms the most stable structure throughout the pH range studied<sup>15</sup>; 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 its stability. The inclusion of a potential W-C base pair (GC or TA sequences) does not provoke a significant effect in

the stability compared to the control sequence, TT. Analysis of the thermodynamic parameters shows that the values of  $\Delta H^0$  for the unfolding process at pH 6.2 range from  $\sim 83$  (CC) to  $\sim 94$  (AA or CA)  $\text{kcal}\cdot\text{mol}^{-1}$ . Considering that there are 6 C-C<sup>+</sup> base pairs, and assuming that the disruption of each of them requires around  $11 \pm 1 \text{ kcal}\cdot\text{mol}^{-1}$  at pH  $\sim 6.0$ <sup>32</sup>, 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<sup>33</sup>. In addition to the two hydrogen bonds formed in each T-T base pair, these base pairs are isomorphous with the C-C<sup>+</sup> ones, and fit well in the intercalative structure, contributing to its stability through favourable stacking interactions. In general, an increment of the  $\Delta H^0$  value was observed at lower pH values, with a maximal value around pH  $\sim 4.5$ . As a general trend,  $\Delta G_{370\text{C}}^0$  values were seen proportional to pH (ESI). In terms of  $\Delta G_{370\text{C}}^0$ , the AA and GG sequences are the least stable structures in the pH range 4.6 – 5.6. On the other hand, the CC sequence shows a higher stability throughout the whole pH range studied.

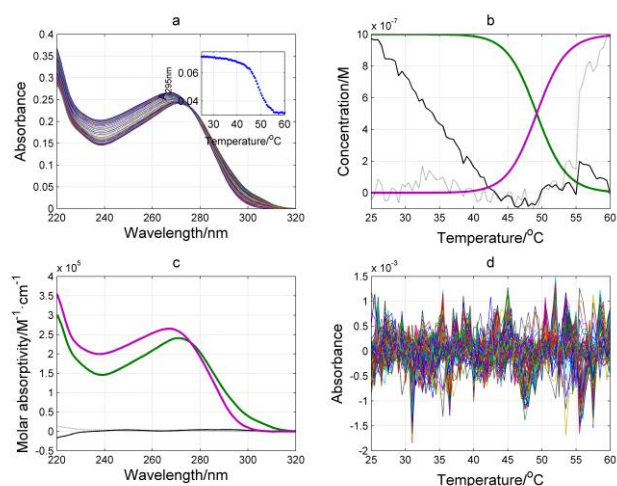


Figure 7. Spectra recorded along the melting of CC sequence at pH 5.6 (a). Inset: absorbance trace at 295 nm. Calculated distribution diagram (matrix C) (b) and pure spectra (matrix S) (c). Absorbance data in matrix E not explained by the proposed model (d). The green and magenta components correspond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

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<sup>34</sup>, due to their greater size.

## Conclusions

Our results suggest that the thermal stability of intramolecular i-motifs and its pH dependence is 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 existence of i-motif structures is dependent both, on the nature of the loops and on the temperature at which the measurement is done. 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.

Anna Sadurní (University of Barcelona) is thanked for carrying out some experiments. We acknowledge funding from the Spanish government (BFU2014-52864-R, CTQ2012-38616-C02-02 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, *Chemistry - A European Journal*, 2015, **21**, 9816-9824
3. S. Benabou, A. Avino, R. Eritja, C. Gonzalez and R. Gargallo, *RSC Advances*, 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.
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 Journal*, 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. Fernandez, R. Eritja, A. Aviño, J. Jaumot and R. Gargallo, *Int. J. Biol. Macromol.*, 2011, **49**, 729-736.
15. T. Fujii and N. Sugimoto, *PCCP*, 2015, **17**, 16719-16722.
16. M. Piotto, V. Saudek and V. Sklenár, *J. Biomol. NMR*, 1992, **2**, 661-665.
17. W. A. Kibbe, *Nucleic Acids Res.*, 2007, **35**, W43-W46.
18. K. J. Breslauer, in *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, *European Biophysics Journal*, 1997, **26**, 419-426.
22. R. D. Gray and J. B. Chaires, in *Current Protocols in Nucleic Acid Chemistry*, John Wiley & Sons, Inc., 2011.
23. N. J. Greenfield, *Nature Protocols*, 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, *Chemometrics and Intelligent Laboratory Systems*, 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. Gonzalez and M. Orozco, *Nucleic Acids Res.*, 2015, **43**, 4309-4321.



Table 2. Thermodynamic parameters calculated from thermal stability studies for the unfolding process. A two-state process was considered in the calculations. Uncertainty in  $T_m$  values was  $\pm 0.7^\circ\text{C}$ . Uncertainties in  $\Delta H$  and  $\Delta S$  values were lower than  $\pm 5\%$ . Uncertainty in  $\Delta G$  values was lower than  $\pm 10\%$ .

DNA	pH 4.6				pH 5.1				pH 5.6				pH 6.2			
	$T_m$ ( $^\circ\text{C}$ )	$\Delta H^\ddagger$ (kcal/mol)	$\Delta S^\ddagger$ (cal/K-mol)	$\Delta G_{37}^\ddagger$ (kcal/mol)	$T_m$ ( $^\circ\text{C}$ )	$\Delta H^\ddagger$ (kcal/mol)	$\Delta S^\ddagger$ (cal/K-mol)	$\Delta G_{37}^\ddagger$ (kcal/mol)	$T_m$ ( $^\circ\text{C}$ )	$\Delta H^\ddagger$ (kcal/mol)	$\Delta S^\ddagger$ (cal/K-mol)	$\Delta G_{37}^\ddagger$ (kcal/mol)	$T_m$ ( $^\circ\text{C}$ )	$\Delta H^\ddagger$ (kcal/mol)	$\Delta S^\ddagger$ (cal/K-mol)	$\Delta G_{37}^\ddagger$ (kcal/mol)
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	26	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	63	78	232	6.0	56	79	240	4.6	46	88	275	2.7	32	87	285	-1.4