

# Stabilization of i-motif structures by 2'- $\beta$ -fluorination of DNA

Hala Abou Assi<sup>1</sup>, Robert W. Harkness, V<sup>1</sup>, Nerea Martin-Pintado<sup>2</sup>, Christopher J. Wilds<sup>1,3</sup>, Ramón Campos-Olivas<sup>4</sup>, Anthony K. Mittermaier<sup>1</sup>, Carlos González<sup>2,\*</sup> and Masad J. Damha<sup>1,\*</sup>

<sup>1</sup>Department of Chemistry, McGill University, Montreal, QC H3A 0B8, Canada, <sup>2</sup>Instituto de Química Física 'Rocasolano', CSIC, 28006 Madrid, Spain, <sup>3</sup>Department of Chemistry and Biochemistry, Concordia University, Montreal, QC H4B 1R6, Canada and <sup>4</sup>Spanish National Cancer Research Center (CNIO), 28029 Madrid, Spain

Received February 25, 2016; Revised April 30, 2016; Accepted May 2, 2016

## ABSTRACT

**i-Motifs are four-stranded DNA structures consisting of two parallel DNA duplexes held together by hemi-protonated and intercalated cytosine base pairs (C:CH<sup>+</sup>). They have attracted considerable research interest for their potential role in gene regulation and their use as pH responsive switches and building blocks in macromolecular assemblies. At neutral and basic pH values, the cytosine bases deprotonate and the structure unfolds into single strands. To avoid this limitation and expand the range of environmental conditions supporting i-motif folding, we replaced the sugar in DNA by 2-deoxy-2-fluoroarabinose. We demonstrate that such a modification significantly stabilizes i-motif formation over a wide pH range, including pH 7. Nuclear magnetic resonance experiments reveal that 2-deoxy-2-fluoroarabinose adopts a C2'-endo conformation, instead of the C3'-endo conformation usually found in unmodified i-motifs. Nevertheless, this substitution does not alter the overall i-motif structure. This conformational change, together with the changes in charge distribution in the sugar caused by the electronegative fluorine atoms, leads to a number of favorable sequential and inter-strand electrostatic interactions. The availability of folded i-motifs at neutral pH will aid investigations into the biological function of i-motifs *in vitro*, and will expand i-motif applications in nanotechnology.**

## INTRODUCTION

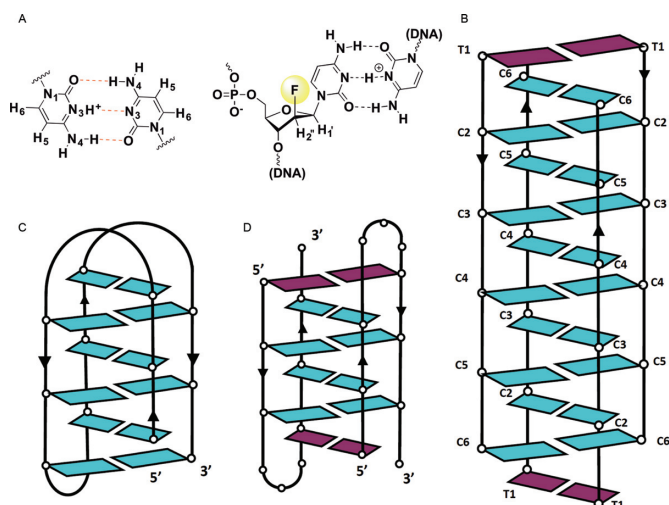
Since the discovery of the DNA double helix, the understanding of DNA structure has expanded considerably, with

a plethora of polymorphs now known (1). Two structures have attracted considerable interest due to their possible involvement in telomere maintenance and gene regulation: (i) the *G-quadruplex* (*G4*) (2,3), which consists of stacked guanine tetrads stabilized by monovalent cations (K<sup>+</sup>/Na<sup>+</sup>); and (ii) the *i-motif* (4), a cytosine-rich structure consisting of two hemi-protonated (C:CH<sup>+</sup>) parallel duplexes intercalated in an antiparallel orientation (5,6).

There is increasing evidence for the occurrence of the i-motif in regulatory regions of the human genome such as centromeres (7,8), telomeres (9,10) and oncogene promoter regions (11,12). These structures are also attracting much interest in nanotechnology, for example in the control of macromolecular structure assembly (13) and as pH responsive switches (14,15).

The i-motif is a compact DNA structure characterized by short (3.1 Å) base-pairing distances, a 12–16° helical twist between adjacent C:CH<sup>+</sup> base pairs (Figure 1A), and close sugar–sugar contacts stabilized by CH...O interactions (16). i-Motif structures can have different intercalation topologies known as 3'E and 5'E. In the 3'E topology, the terminal C:CH<sup>+</sup> base pair is at the 3'-end while the 5'E has the terminal C:CH<sup>+</sup> base pair at the 5'-end (9). Repulsive interactions exist between the charged C-imino protons and between the phosphate groups across the crowded narrow groove. *In vitro*, unmodified i-motif structures are stable across a narrow pH range (~3.5–5.5), with maximum stability occurring at pH equal to the p*K*<sub>a</sub> of the cytosine N3. At pHs above the cytosine N3 p*K*<sub>a</sub>, the stabilizing C:CH<sup>+</sup> base pairing is lost and the i-motif structure is denatured. However, it is likely that *in vivo* factors such as negative superhelicity (17), cellular proteins (18) and molecular crowding conditions (19) stabilize i-motif DNA structures under neutral pH. To help compensate for the absence of these factors *in vitro*, small molecule ligands and certain metal ions (Cu<sup>+2</sup>; Ag<sup>+</sup>) have been introduced to promote i-motif folding at neutral pH (20). Of the sugar and phosphate modi-

\*To whom correspondence should be addressed. Tel: +1 514 398 7552; Fax: +1 514 398 3797; Email: masad.damha@mcgill.ca  
Correspondence may also be addressed to Carlos González. Tel: +34 917 45 9533; Fax: +34 915 64 2431; Email: cgonzalez@iqfr.csic.es



**Figure 1.** (A) C:CH<sup>+</sup> base pair (left) and 2′F-araC:CH<sup>+</sup> base pair (right). (B) Schematic representation of the tetrameric i-motif formed by dTC-CCCC (H-1). (C) Schematic representation of human telomeric repeat (HT-0). (D) Schematic representation of the centromeric A-box (HC-0). Color code: dC are shown in cyan and capping dT are shown in magenta. Residues in the loops are not shown.

fications so far reported, none has provided sufficient stability at neutral pH to allow for biochemical studies *in vitro* (21–25).

The present study was prompted by observations that arabinose (2′OH) (26), but not ribose (2′OH (27) or 2′F (28)), is well tolerated within the i-motif structure. We hypothesized that 2′F-arabinose (2′F-araC) would enhance interstrand CH...O interactions within the narrow grooves (Figure 1A). Sequential FC-H2′...O4′ interactions have been previously observed in duplexes (29) and G-quadruplex structures (30). Furthermore, such a modification could potentially stabilize the i-motif by augmenting ion-dipole and stacking interactions of C:CH<sup>+</sup> base pairs (16). To explore the effect of 2′-β-fluorination on the structure and stability of different families of i-motif structures, 2′F-araC modifications were introduced in various i-motif forming sequences. We show that replacing 2-deoxyribose sugar with 2-deoxy-2-fluoroarabinose enables formation of stable i-motif structures over a wide pH range with thermal stabilities of ~30° at neutral pH.

## MATERIALS AND METHODS

### Oligonucleotide synthesis and purification

Oligonucleotide synthesis was performed on an ABI 3400 DNA synthesizer from Applied Biosystems at 1 μmol scale on Unylinker (Chemgenes) CPG solid support. Thymidine (dT), deoxycytidine (N-acetyl) (dC) and deoxyadenosine (N-Bz) (dA) phosphoramidites were used at 0.1 M concentration in acetonitrile, and coupled for 110 s. 2′F-araC was used at 0.13 M concentration and coupled for 600 s. After completion of the synthesis, CPG was transferred to a 1.5 ml screw-cap eppendorf. A total of 1000 μl aqueous ammonium hydroxide was added and the eppendorf was placed on a shaker at room temperature for 48 h. The deprotection solution was centrifuged and decanted from the

CPG. Samples were vented for 30 min, chilled in dry ice and evaporated to dryness. Hexamer sequences were purified by anion exchange HPLC on a Waters 1525 instrument using a Protein-Pak DEAE 5PW column (21.5 mm × 15 cm). The buffer system consisted of water (solution A) and 1 M aqueous lithium perchlorate (solution B), at a flow rate of 4 ml/min. The gradient was 0–40% solution B over 50 min at 60°C. Under these conditions the desired peaks eluted at roughly 28 min. The centromeric (HC) and telomeric (HT) sequences were purified by anion exchange HPLC on an Agilent 1200 Series Instrument using a Protein-Pak DEAE 5PW column (7.5 × 75 mm) at a flow rate of 1 ml/min. The gradient was 0–24% solution B over 30 min at 60°C. Under these conditions, the desired HC and HT peaks eluted at roughly 21 and 24 min, respectively. Samples were desalted on NAP-25 desalting columns according to manufacturer protocol. The extinction coefficient of the 2′F-araC was approximated to that of the unmodified sequence. Masses were verified by ESI-MS.

### UV melting studies

UV thermal denaturation data were obtained on a Varian CARY 100 UV-visible spectrophotometer equipped with a Peltier temperature controller. The concentration of oligonucleotides used was 4.6 μM for the hexamer sequences and 4 μM for the centromeric and telomeric sequences. Samples were dissolved in appropriate buffer as indicated in the text. Concentrations were determined after quantitating the samples by UV absorbance at λ = 260 nm. Samples were heated to 90°C for 15 min, then cooled slowly to room temperature and stored in the fridge (5°C) at least 18 h before the measurements were performed. Denaturation curves were acquired at 265 nm at a rate of 0.5°C/min. Experiments at pH 5.0 and pH 7.0 with 0.5°C/min were performed at least in triplicates and experiments at 0.2°C/min were ran once for comparison. Samples were kept under a nitrogen sheath at temperatures below 12°C. The dissociation temperatures were calculated as the midpoint of the transition ( $T_{1/2}$ ) using the first derivative of the experimental data.

### Circular dichroism

Circular dichroism (CD) studies were performed at 5°C on a JASCO J-810 spectropolarimeter using a 1 mm path length cuvette. Temperature was maintained using the Peltier unit within the instrument. Spectra were recorded from 350–230 nm at a scan rate of 100 nm min<sup>-1</sup> and a response time of 2.0 s with three acquisitions recorded for each spectrum. The spectra were corrected by subtraction of the buffer scan. Data were smoothed using the means-movement function within the JASCO graphing software. Oligonucleotide solutions for CD measurements were prepared with 10 mM sodium phosphate buffer (pH 5.0 and pH 7.0) in a similar manner to that used for UV melting. The concentration of oligonucleotides used was 30 μM for the hexamer sequences, 100 μM for the centromeric sequences and 50 μM for the telomeric sequences.

## Gel electrophoresis

Native gel electrophoresis was performed utilizing 24% polyacrylamide at pH 5.0 using TAE (Tris:Acetate:EDTA) buffer. Running buffer was  $1 \times$  TAE. Samples were equilibrated in 10 mM sodium phosphate (pH 5.0) as described in the UV melting section. The gels were run at 200 V inside a refrigerator (5°C) for 12 h or at 280 V for 6 h. Gels were visualized by treatment with Stains-All dye (sigma E-9379) or UV shadowing. The oligonucleotide controls were dT<sub>12</sub> and dT<sub>24</sub> strands.

## NMR experiments

Samples for nuclear magnetic resonance (NMR) experiments were suspended in 300  $\mu$ l of either D<sub>2</sub>O or H<sub>2</sub>O/D<sub>2</sub>O 9:1 in 10 mM sodium phosphate buffer. NMR spectra were acquired on Bruker Avance spectrometers operating at 600, 700 or 800 MHz, and processed with Topspin software. TOCSY spectra were recorded with standard MLEV17 spinlock sequence and with 80 ms mixing time. NOESY spectra in H<sub>2</sub>O were acquired with 50 and 150 ms mixing times. For 2D experiments in H<sub>2</sub>O, water suppression was achieved by including a WATERGATE (31) module in the pulse sequence prior to acquisition. Two-dimensional experiments were carried out at temperatures ranging from 5 to 25°C. <sup>19</sup>F resonances were assigned from <sup>1</sup>H-<sup>19</sup>F HETCOR and <sup>19</sup>F detected HOESY spectra (32). The spectral analysis program Sparky (33) was used for semiautomatic assignment of the NOESY cross-peaks.

## Experimental NMR constraints

Distance constraints were obtained from a qualitative estimation of Nuclear Overhauser Effect (NOE) intensities. In addition to these experimentally derived constraints, hydrogen bond and planarity constraints for the base pairs were used in the initial DYANA calculations. Target values for distances and angles related to hydrogen bonds were set to values obtained from crystallographic data in related structures (34). Due to the relatively broad line-widths of the sugar proton signals, J-coupling constants were not accurately measured, but only roughly estimated from DQF-COSY cross-peaks. Loose values were set for the sugar dihedral angles  $\delta$ ,  $\nu_1$  and  $\nu_2$  to constrain deoxyribose conformation to north or south domain. No backbone angle constraints were employed. Distance constraints with their corresponding error bounds were incorporated into the AMBER potential energy by defining a flat-well potential term.

## Structure determination

Structures were calculated with the program DYANA 1.4 (35) and further refined with the SANDER module of the molecular dynamics package AMBER 7.0 (36). The resulting DYANA structures were used as starting points for the AMBER refinement, consisting of 1 ns trajectories in which explicit solvent molecules were included and using the Particle Mesh Ewald method to evaluate long-range electrostatic interactions. Non-experimental constraints used in the initial DYANA calculations were removed in the AMBER refinement. The specific protocols for these calculations have

been described elsewhere (37). The AMBER-98 force field (38) was used to describe the DNA, and the TIP3P model was used to simulate water molecules. Analysis of the representative structures as well as the MD trajectories was carried out with the programs Curves V5.1 (39) and MOLMOL (40).

## Differential scanning calorimetry

Differential scanning calorimetry (DSC) was performed using a NanoDSC-III (TA Instruments, USA). All samples were 150  $\mu$ M in 10 mM sodium phosphate (pH 5.0). Sample data were collected in triplicate by scanning from 5–80°C at a scan rate of 0.5°C per minute. Fitting procedures are explained in detail in the Supplementary Data. Errors were calculated according to the variance-covariance method (41).

## RESULTS

### The effect of 2'F-araC substitutions on thermal and pH stability of i-motifs

Sequences of the oligonucleotides prepared for this study are provided in Table 1. The first series (named H) is derived from dTCCCCC, a hexanucleotide that associates into a tetrameric i-motif structure (4) (Figure 1B), whereas the HT and HC series are derived from human telomeric and centromeric DNA sequences, respectively. Their control unmodified sequences; HT-0 and HC-0 form monomeric and dimeric i-motif structures respectively (Figure 1C and D).

### Effects of pH and temperature on i-motif structure

Circular dichroism provides a convenient way to detect the formation of i-motifs since their spectra exhibit characteristic negative and positive bands at  $\sim$ 265 and  $\sim$ 285 nm, respectively (42). As previously reported, the model sequence dTCCCCC (H-1) exhibits a positive band that decreases in magnitude and becomes blue shifted as the pH is raised above 5.5 (Figure 2A). By contrast, some of the modified hexanucleotides (e.g., H-3, H-4, H-5 and H-6) exhibit this characteristic i-motif CD signature at pH 6.5 (Figure 2B and Supplementary Figure S1). Similarly, the modified centromeric and telomeric sequences maintained the i-motif CD signature at neutral pH (Figure 2C and D). pH titrations were performed for the modified and unmodified i-motif structures to determine the pH at which 50% of the oligo is folded into an i-motif structure (pH<sub>1/2</sub>). The titration profiles were fit to a standard titration model assuming a single protonation event in order to extract populations of folded (protonated) and unfolded (deprotonated) states (Supplementary Figure S2) as a function of pH. The fraction protonated profiles indicate that the modified i-motif structures are markedly stabilized at pH 7.0 relative to the corresponding unmodified references. For example, H-4 and H-6 are roughly 40 and 50% folded at pH 7.0 respectively, whereas H-1 is 10% folded. The pH stabilizations translate to the centromeric and telomeric sequences where the fraction protonated profiles for the modified HC-3 and HT-4 sequences show folded populations of  $\sim$ 60 and 90% at pH 7.0 respectively, compared to the unmodified HC-0 and

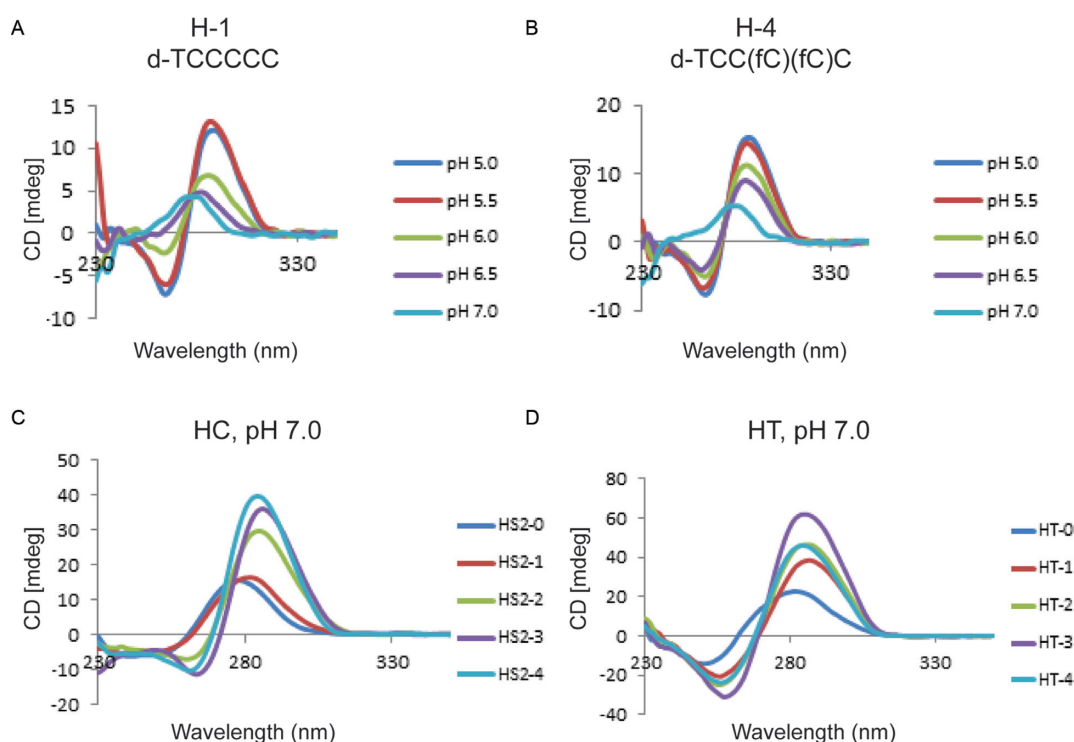


**Table 1.**  $T_{1/2}$  values for the (un)modified sequences<sup>a</sup>

Code	Sequence (5'-3')	$T_{1/2}$ pH 5.0	$\Delta T_{1/2}$ pH 5.0	$T_{1/2}$ pH 7.0	$\Delta T_{1/2}$ pH 7.0
H-1	dTCCCCC	51.5	-	-	-
H-2	dTCC(fC)CC	10.0, 53.6	+2.1	-	-
H-3	dTC(fC)(fC)CC	25.7, 58.7	+7.2	-	-
H-4	dTCC(fC)(fC)C	21.5, 57.0	+5.5	-	-
H-5 <sup>b</sup>	dTC(fC)(fC)(fC)C	32.0	-	-	-
H-6 <sup>b</sup>	dTf(CCCCC)	37.5	-	10.0	-
HC-0	dTCCTTTTCCA	19.0	-	9.0	-
HC-1	dT(fC)CTTTTCCA	23.0	+4.0	11.0	+2.0
HC-2	dT(fC)(fC)TTTTCCA	31.6	+12.6	24.1	+15.1
HC-3	dT(fC)CTTTT(fC)CA	34.1	+15.1	27.1	+18.1
HC-4	dT(fC)(fC)TTTT(fC)(fC)A	36.2	+17.2	28.2	+19.2
HT-0	d(CCCTAA) <sub>3</sub> C <sub>3</sub>	60.0	-	14.0	-
HT-1	d(C(fC)CTAA) <sub>3</sub> C(fC)C	67.5	+7.5	23.0	+9.0
HT-2	d(C(fC)(fC)TAA) <sub>3</sub> C(fC)(fC)	77.1	+17.1	32.6	+18.6
HT-3	d((fC)(fC)CTAA) <sub>3</sub> (fC)(fC)C	74.1	+14.1	26.6	+12.6
HT-4	d((fC)(fC)(fC)TAA) <sub>3</sub> (fC) <sub>3</sub>	80.2	+20.2	31.2	+17.2

<sup>a</sup>Oligonucleotide concentration is 4.6  $\mu\text{M}$  in single strands for hexamers and 4.0  $\mu\text{M}$  for HC and HT.  $T_{1/2}$  data were calculated from UV-visible spectroscopy thermal denaturation profiles where  $T_{1/2}$  corresponds to the midpoint of the dissociation transition obtained at 0.5°C/min. This value may be sensitive to scan rate (refer to Supplementary Table S2 for data at 0.2°C/min).  $\Delta T_{1/2}$  values were calculated for the main melting transitions relative to the respective unmodified strands. (-) indicates sequences for which no melting transition was detected.

<sup>b</sup>Sequences H-5 and H-6 exhibit lower  $T_{1/2}$  values compared to the control (H-1). This is due to the melting of a dimeric structure rather than an i-motif structure. Therefore,  $\Delta T_{1/2}$  cannot be calculated since the comparison with H-1 will be invalid.



**Figure 2.** (A and B) pH dependent CD spectra at 30  $\mu\text{M}$  and  $T = 5^\circ\text{C}$  for H-1 and H-4 respectively. (C) CD spectra for centromeric sequences at pH 7.0, 100  $\mu\text{M}$  and  $T = 5^\circ\text{C}$ . (D) CD spectra for telomeric sequences at pH 7.0, 50  $\mu\text{M}$  and  $T = 5^\circ\text{C}$ .

HT-0 that are roughly 10 and 30% folded (Supplementary Figure S2). This 6- and 3-fold increase in folded population is translated to an increase in thermal stability for HC-3 ( $\Delta T_{1/2} = +18.1$ ) and HT-4 ( $\Delta T_{1/2} = +17.2$ ). Here we use  $T_{1/2}$  to refer to the midpoint of the dissociation transition, owing to the observed hysteresis between forward and reverse scans for many of the studied sequences. The  $\Delta\text{pH}_{1/2}$  from the modifications are listed in Table 2. An opposite

trend is observed for the nucleosides: the  $\text{pK}_a$ s of 2'-F-araC and dC are 3.9 and 4.4, respectively. This indicates that the effective  $\text{pK}_a$  of 2'-F-araC residues is strongly affected by the three-dimensional structure of the i-motif. Most likely, this large  $\text{pK}_a$  shift is due to the hydrogen-bonding network and electrostatic interactions in the neighborhood of the protonation sites (43).

**Table 2.** Nucleoside p*K*<sub>a</sub> and i-motif p*H*<sub>1/2</sub> values<sup>a,b</sup>

Code	Nucleoside	p <i>K</i> <sub>a</sub>	Δp <i>K</i> <sub>a</sub>
dC	2'-deoxycytidine	4.4 ± 0.02	-
2'F-araC	2'-deoxy-2'-fluoroarabincytidine	3.9 ± 0.05	-0.5 ± 0.05
Code	Sequence	p <i>H</i> <sub>1/2</sub>	Δp <i>H</i> <sub>1/2</sub>
H-1	5'-dTCCCCC-3'	6.1 ± 0.2	-
H-4	5'-dTCC(fC)(fC)C-3'	6.7 ± 0.2	0.6 ± 0.3
H-6	5'-dTf(CCCCC)-3'	6.8 ± 0.3	0.7 ± 0.3
HC-0	5'-dTCCCTTTTCCA-3'	6.3 ± 0.2	-
HC-3	5'-dT(fC)CTTTT(fC)CA-3'	7.0 ± 0.2	0.7 ± 0.3
HT-0	5'-d(CCCTAA) <sub>3</sub> CCC-3'	6.8 ± 0.1	-
HT-4	5'-d((fC)(fC)(fC)TAA) <sub>3</sub> (fC)(fC)(fC)-3'	7.6 ± 0.3	0.8 ± 0.3

<sup>a</sup>The p*K*<sub>a</sub> values for dC and 2'F-araC were calculated from the plot of UV absorbance at 280 nm in 10 mM NaPi buffer as a function of pH. The p*H*<sub>1/2</sub> values for i-motif structures were calculated from the plot of molar ellipticity at 284 nm versus pH in 10 mM NaPi buffer at 5°C.

<sup>b</sup>CD versus pH data were fit to a standard titration model involving a single protonation event using  $CD_{obs} = CD_{high} + (CD_{low} - CD_{high}) / (1 + 10^{pH-pK_a})$ . Errors were calculated according to the variance-covariance method (41).

### <sup>1</sup>H-NMR and UV melting

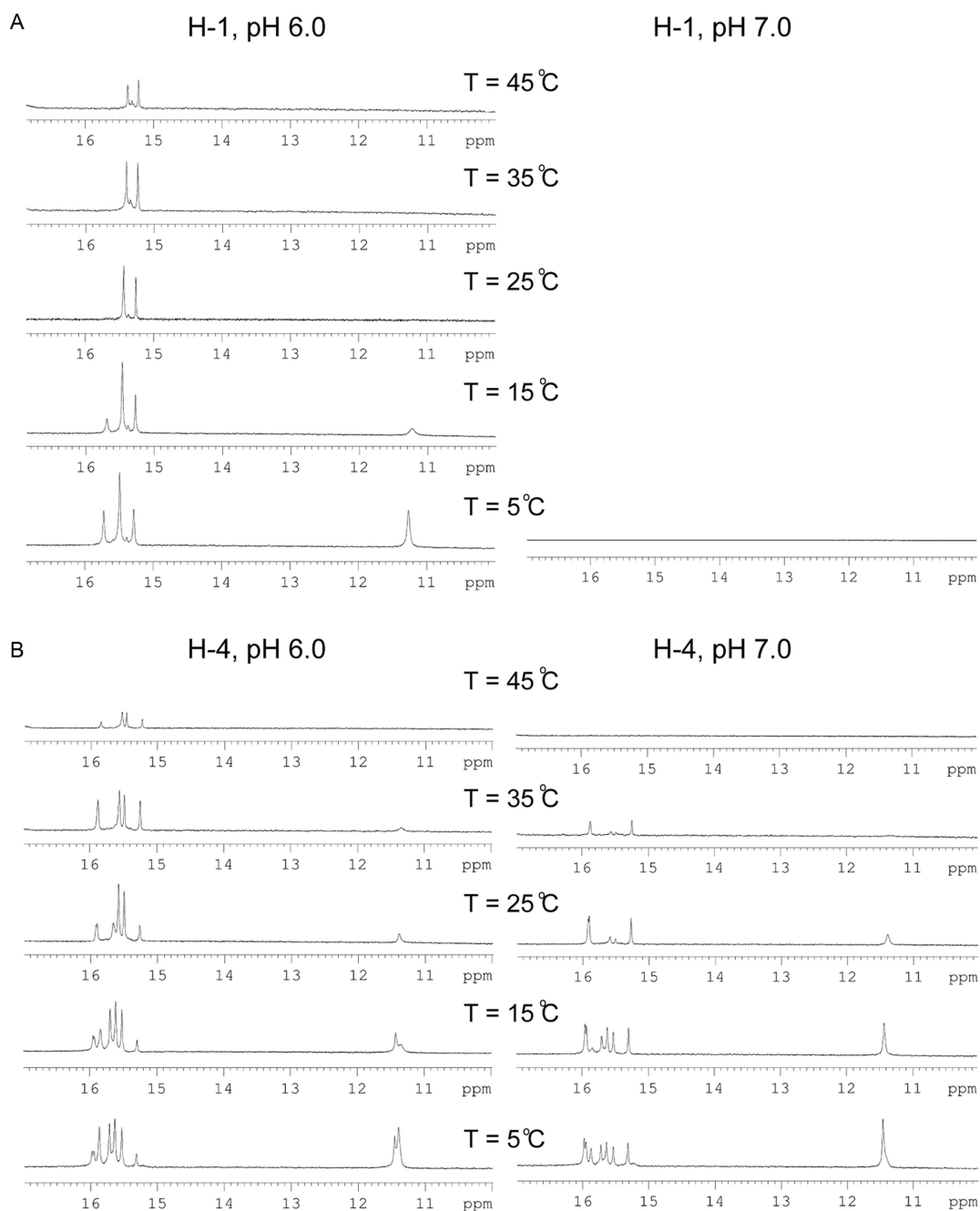
i-Motif formation can be also detected by <sup>1</sup>H-NMR. Signals at 15–16 ppm, characteristic of cytosine imino protons in hemi-protonated C:CH<sup>+</sup> base pairs, were observed for most of the modified sequences over a wide temperature and pH range (Supplementary Figures S3–5). For instance, the control i-motif (H-1) exhibits no imino signals at pH 7.0 (Figure 3A) whereas i-motif structures formed by H-3, H-4, H-5 and H-6 maintain C:CH<sup>+</sup> base pairing at neutral pH and at relatively high temperatures (35°C) (Figure 3B and Supplementary Figure S3). This effect is more pronounced in the case of the dimer centromeric sequences (Supplementary Figure S4) and monomer human telomeric sequences (Supplementary Figure S5), in which <sup>1</sup>H-NMR spectra exhibit imino signals at 15–16 ppm which are clearly visible at physiological temperatures and neutral pH.

UV melting experiments showed a significant increase in *T*<sub>1/2</sub> at acidic and neutral pH in all the monomeric (HT) and dimeric (HC) sequences studied herein (Table 1 and Figure 4). Although very similar spectral features (NMR and CD) were observed for the native (HT-0) and modified telomeric sequences (e.g. HT-2 and HT-4), the modified structures were significantly more stable at both pH 5.0 and 7.0 relative to the unmodified strands (Δ*T*<sub>1/2</sub> ≈ 17–20°C; Table 1). HC-4 and HT-2 exhibit *T*<sub>1/2</sub> values of 28.2 and 32.6°C at pH 7.0, respectively (Figure 4). It is interesting to note that Δ*T*<sub>1/2</sub> values obtained between modified and control oligonucleotides were not dependent on the temperature gradient used in the melting experiments (Table 1 and Supplementary Table S2). In addition, the dissociation/association profiles obtained at 0.5 and 0.2°C/min for every sequence studied were virtually identical, indicating that the measured processes are the same, despite the presence of hysteresis. The same stabilization effect was observed in the majority of the tetrameric sequences (Table 1 and Supplementary Table S1). Melting of H-1 was detected at pH 3.5–5.5, whereas melting of the modified sequences was observed over a broader pH range (3.0–7.0). However, the denaturation curves of the tetrameric structures are more complex, showing more than one transition, which suggests the occurrence of several species. The most

substituted tetrameric sequences exhibit lower *T*<sub>1/2</sub> values, in contrast to the general stabilization effect observed in all the other sequences. These apparently contradictory results prompted us to study the tetrameric sequences in more detail.

### Gel electrophoretic experiments

Native polyacrylamide gel electrophoresis (PAGE) experiments confirmed the formation of more than one species in the tetrameric i-motif structures (Supplementary Figures S6 and 7) at low temperature (5°C). For example, the denaturation curve of the singly modified oligomer (H-2, 4.6 μM; pH 5.0) exhibits a minor transition with a *T*<sub>1/2</sub> of 10.0°C and a major transition with a *T*<sub>1/2</sub> of 53.6°C that corresponds closely to that of the control sequence (*T*<sub>1/2</sub> of 51.5°C). This is also evident in a native gel, where H-2 (40 μM, Supplementary Figure S6) appeared as a mixture of closely moving bands that migrated with the same mobility as the tetrameric control structure H-1. Results obtained for H-2, H-3, H-4 and H-5 from concentration-dependent gel electrophoresis experiments are consistent with the formation of topologically similar i-motif structures at all concentrations studied (4.6–200 μM; Supplementary Figure S7). A major species was observed for the unmodified control sequence at 4.6–20 μM, with a minor component appearing on the gel at >40 μM strand concentration (Supplementary Figure S7) (4). Interestingly, the fully substituted sequence (H-6) exhibited a significantly lower *T*<sub>1/2</sub> value (37.5°C, Table 1), fast association-dissociation kinetics (Supplementary Figure S8) and aberrant gel electrophoretic mobility (Supplementary Figure S7), all consistent with the formation of a dimeric structure in this case. For instance, at pH 5.0 and 40 μM, H-6 appears as one fast moving band and an additional band exhibiting co-migration with the H-1 control (Supplementary Figure S7; lane 6). The slow moving band corresponds to a tetrameric structure and becomes the predominant species at higher strand concentration (100–200 μM; Supplementary Figure S7).



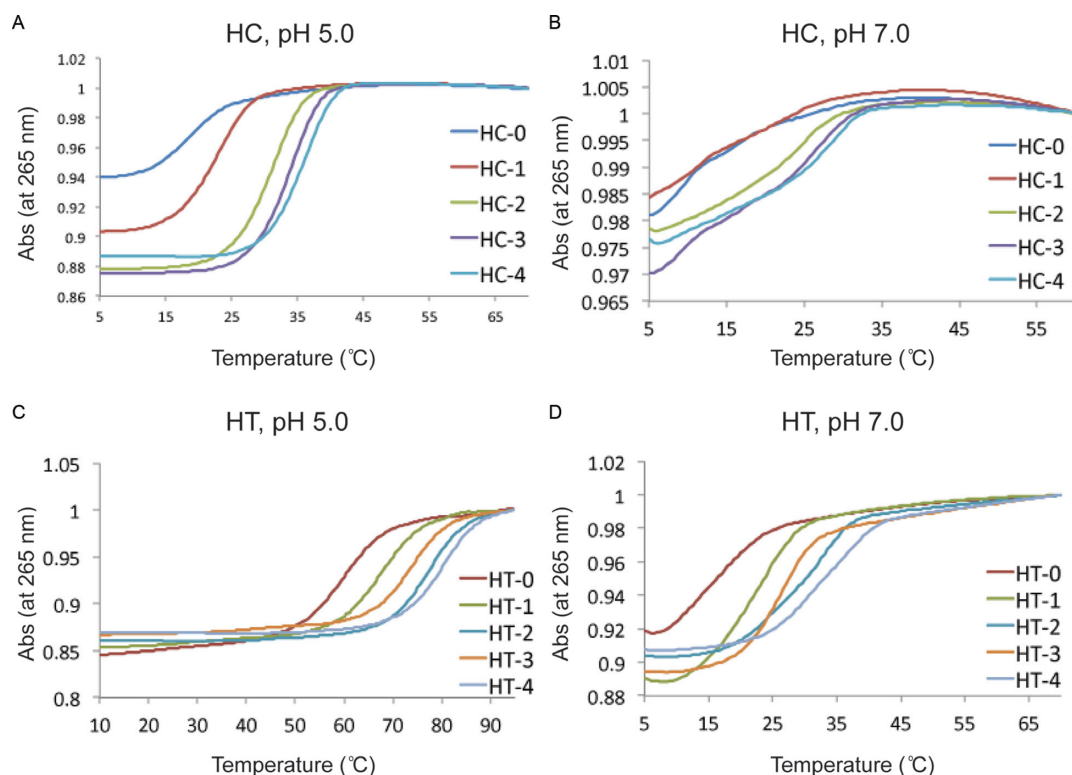
**Figure 3.** (A)  $^1\text{H}$  NMR melting experiments for H-1(5'-dTCCCCC-3', 2.0 mM). At pH 7.0 and  $T = 5^\circ\text{C}$ , no imino peaks were observed. (B)  $^1\text{H}$  NMR melting experiments for H-4(5'-dTCC(fC)(fC)C-3', 2.53 mM).

### $^{19}\text{F}$ -NMR

The wide dispersion of fluorine chemical shifts in the  $^{19}\text{F}$  NMR spectra facilitates the study of multiple species in equilibrium under various experimental conditions (44). In most of the tetrameric structures, the number of  $^{19}\text{F}$  signals is not consistent with a single conformation, in agreement with PAGE data.  $^{19}\text{F}$  resonances from different species, including the unfolded oligonucleotide, can be observed simultaneously in the whole range of temperatures explored. This indicates that the equilibria are slow on the NMR timescale at all temperatures, as observed previously in

other 2'-F-araC-modified oligonucleotides (45,29). With the exception of sequences H-5 and H-6,  $^{19}\text{F}$  spectra recorded at high (1-2 mM) and low (0.1 mM) concentrations are very similar (Supplementary Figure S9).

As observed in UV melting and gel electrophoresis experiments, the fully substituted sequence (H-6) exhibits distinctive features.  $^{19}\text{F}$  NMR spectra for H-6 at high and low oligonucleotide concentration are completely different (Supplementary Figure S9e). Thus,  $^{19}\text{F}$  NMR spectra confirm the presence of species of different molecularity, as suggested by UV and gel electrophoretic mobility experi-



**Figure 4.** (A)  $T_{1/2}$  spectra for HC i-motifs: 4  $\mu\text{M}$  strand concentration in 10 mM NaPi buffer pH 5. (B)  $T_{1/2}$  spectra for HC i-motifs: 4  $\mu\text{M}$  strand concentration in 10 mM NaPi buffer pH 7.0. (C)  $T_{1/2}$  spectra for HT i-motifs: 4  $\mu\text{M}$  strand concentration in 10 mM NaPi buffer pH 5.0. (D)  $T_{1/2}$  spectra for HT i-motifs: 4  $\mu\text{M}$  strand concentration in 10 mM NaPi buffer pH 7.0.  $T_{1/2}$  were recorded at 265 nm and 0.5°C/min heating rate.

ments.  $^{19}\text{F}$  signals of the low concentration species are observed in the high concentration sample prepared under a fast annealing procedure. On the other hand, the high concentration spectrum is partially recovered after storing the low concentration sample for several months at  $T = 5^\circ\text{C}$  (Supplementary Figure S10). This indicates that the dimeric species is kinetically favored. Such strong kinetic bias explains the apparent destabilization observed in H-6, since the  $T_{1/2}$  determined by UV (37.5°C) corresponds to a kinetically trapped species detected by NMR at low concentrations. Interestingly,  $^1\text{H}$ -NMR spectra of H-6 recorded in conditions in which the dimeric species is predominant exhibit signals in the 15–16 ppm region, as expected for hemiprotonated cytosines (Supplementary Figure S11). Most probably, this structure is a parallel homoduplex stabilized by  $\text{C}:\text{CH}^+$  base pairs (46,47). Since there is no indication of the formation of this dimeric species in the unmodified oligonucleotide, we must consider that 2'F-araC substitutions may also have an effect on the stabilization of parallel duplexes, which in this case can be considered as intermediates in the formation of i-motif structures.

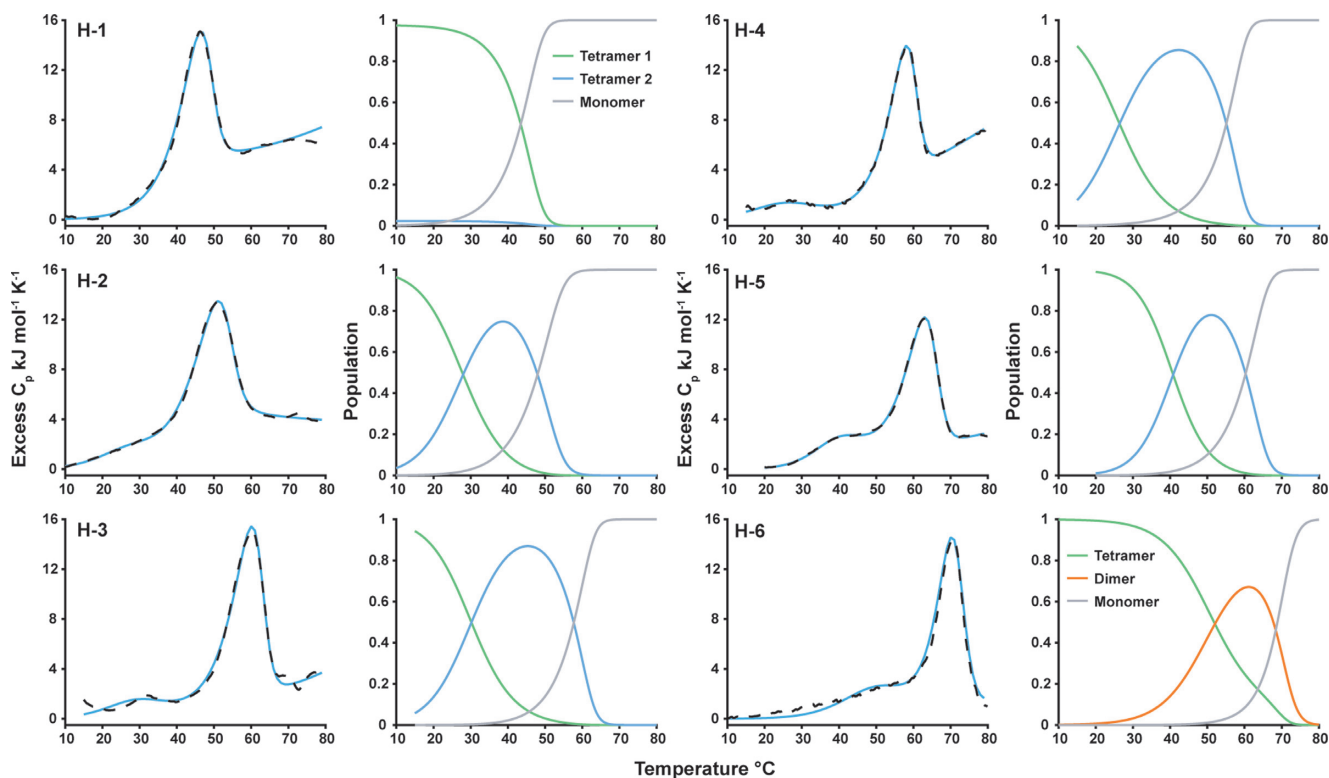
$^{19}\text{F}$  spectra suggest that a similar effect may also occur in H-5. Therefore, the only two exceptions in Table 1 can be explained by the formation of kinetically trapped dimeric structures. To extract more reliable thermodynamic data on the tetrameric i-motif formation, we made use of calorimetric techniques.

#### Thermodynamic analysis by differential scanning calorimetry

DSC is well-suited to detecting the presence of multiple structural populations. We performed DSC experiments on H1-H6, finding the main melting transition to occur at higher temperatures with increasing numbers of 2'F-araC modifications (Figure 5). With the  $\sim 33$ -fold higher concentrations relative to UV-Vis experiments, the association kinetics were much faster, reflected in the little to no hysteresis across the sample set (Supplementary Figure S12). Interestingly, an additional low temperature transition became more prominent as the number of modifications was increased (H-3 through H-6). The DSC profile for H-6 was in greatest contrast, with a wide low temperature shoulder and the highest main transition temperature of the entire set. Positive unfolding  $\Delta C_p$ s  $\approx 2$ –6  $\text{kJ mol}^{-1} \text{K}^{-1}$  were found for each structural ensemble.

We fit the DSC profiles assuming non-two-state folding behavior, finding in all cases that the extracted structural populations were in close agreement with gel and NMR data. H1-H5 were fit with a thermodynamic model assuming the presence of two tetrameric structures, while the model for H6 assumed a tetrameric and dimeric structure, in accordance with gel and NMR results. Thermodynamic parameters resulting from the DSC fits are displayed in Table 3. Because certain samples had slight hysteresis (Supplementary Figure S12), we use  $T_{1/2}$  here as the temperature at which respective structural populations are equal to the population of the monomer. For the samples with no





**Figure 5.** Excess heat capacity ( $C_p$ ) profiles for the control and 2'F-araC-modified tetrameric i-motif structures. Sequence codes are indicated in the top left corners of the excess  $C_p$  profiles. Experimental and fitted excess  $C_p$ s are shown as black dashed and blue solid lines, respectively. Structural populations from the fits to each excess  $C_p$  profile are shown in the panels immediately to the right. Tetramer, dimer and monomer populations are shown as green and blue, orange and gray lines respectively.

hysteresis, this value is the equilibrium  $T_m$ . Optimal data quality with minimal hysteresis was found with scan rates of 0.5°C/min, therefore we applied this rate across the H1-H6 dataset. The populations extracted from the fit for H-1 show a small fraction of a second tetrameric structure at low temperature. DSC populations for H2-H5 indicate the presence of minor tetramer structures at low temperature, which become the dominant populations toward the  $T_{1/2}$ . These unfolding intermediates may be different topologies or frayed structures, such as those observed in G-quadruplexes and homoduplexes (48,49). The NMR for H-4 (discussed below) suggests that the stable structure at pH 7.0 has opened terminal bases. This might be the case for the unfolding intermediates in H-2 through H-5. H-6 populations indicate the tetrameric structure is favored at low temperature and the dimeric intermediate structure is strongly populated toward the  $T_{1/2}$  to give the main transition. Importantly, population of intermediate states in H2-H6 is consistent with the NMR data. The NMR spectra contain peaks which shift with increasing temperature, suggesting the initial structures at low temperature are in fast chemical exchange on the NMR timescale (50,51).

In addition to identifying structural populations and quantifying their unfolding thermodynamics, the DSC fitting allows determination of  $T_{1/2}$ s for individual populations, as well as the global  $T_{1/2}$  at which 50% of the structural ensemble is denatured. To quantify the increase in thermal stability due to the modifications, we compared the

increase in  $T_{1/2}$  for each modified structure relative to the corresponding unmodified structure ( $\Delta T_{1/2} = T_{1/2}^{\text{modified}} - T_{1/2}^{\text{unmodified}}$ ) with the number of modifications. The  $\Delta T_{1/2}$ s as a function of number of modifications are strongly correlated ( $R = 0.98, 0.98, 0.98$ ), showing the  $T_{1/2}$ s for structures 1, 2 and the global  $T_{1/2}$  are increased by 7.5, 5.1 and 5.2°C respectively with each 2'F-araC modification (Supplementary Figure S13).

### Structural basis of 2'F-araC stabilization

To get more insight into the origin of the stabilization due to fluoroarabinose modifications, an in-depth structural study was undertaken using NMR spectroscopy. Complete assignment of the NMR spectra of highly repetitive sequences such as those studied here is a difficult task which is further exacerbated in this case by the occurrence of multiple folded species (in the case of the tetrameric structures). However, the presence of 2'F-araCs in the sequence improves chemical shift dispersion and greatly facilitates 2D NMR analysis. In the case of H-4, we could obtain a complete sequential assignment of the major species at pH 5.0, consistent with a 3'E intercalation topology (Supplementary Figure S14). Strong H1'-H1' cross-peaks and amino-H2'/2'' NOEs clearly confirmed the formation of an i-motif structure. The stacking order could be determined by following H1'-H1' connectivities along the minor groove between C2-fC5, C2-C6, C6-T1, C3-fC4 and C3-fC5. This pattern was confirmed by the amino-H2'/H2'' contacts observed



**Table 3.** Thermodynamic parameters extracted from fits of the tetrameric i-motif DSC melting experiments

	$-\Delta H_1^a$	$-\Delta H_2^a$	$-\Delta S_1^b$	$-\Delta S_2^b$	$-\Delta G_1^a$	$-\Delta G_2^a$	$T_{1/2(1)}^c$	$T_{1/2(2)}^c$	Global $T_{1/2}^d$
<b>H-1</b>	572.6 ± 0.2	572.4 ± 3.8	1615 ± 35	1583 ± 18	61.5 ± 0.7	71.4 ± 0.7	23.9 ± 0.8	43.4 ± 0.7	43.5 ± 0.2
<b>H-2</b>	692.4 ± 4.4	562.3 ± 2.7	1960 ± 15	1528 ± 9	72.1 ± 0.4	78.7 ± 0.3	38.7 ± 0.6	48.0 ± 0.7	48.1 ± 0.2
<b>H-3</b>	771.8 ± 0.7	638.1 ± 0.8	2146 ± 5	1705 ± 2	92.6 ± 1.5	98.5 ± 0.9	45.7 ± 1.2	57.7 ± 0.8	57.8 ± 1.7
<b>H-4</b>	545.3 ± 0.4	422.9 ± 0.3	1457 ± 10	1049 ± 1	84.2 ± 0.7	90.9 ± 0.3	42.2 ± 0.9	55.1 ± 0.4	55.2 ± 0.3
<b>H-5</b>	663.0 ± 0.4	493.9 ± 0.4	1783 ± 54	1244 ± 37	98.7 ± 0.6	100.2 ± 0.4	50.6 ± 0.4	60.3 ± 0.5	60.4 ± 0.3
<b>H-6</b>	677.7 ± 6.1	205.5 ± 4.4	1743 ± 19	498 ± 16	126.0 ± 0.4	47.9 ± 0.9	63.5 ± 1.0	68.4 ± 1.2	68.9 ± 0.2

<sup>a</sup> $\Delta H$  and  $\Delta G$  are expressed in kJ mol<sup>-1</sup>.

<sup>b</sup> $\Delta S$  is expressed in J mol<sup>-1</sup> K<sup>-1</sup>. Thermodynamic parameters were calculated at the global  $T_{1/2}$  for the control sequence H-1.

<sup>c</sup> $T_{1/2s}$  (expressed in °C) for structures 1 and 2 were defined as the temperatures where the respective populations were equal to the population of the monomer.

<sup>d</sup>Global  $T_{1/2s}$  were defined as the temperatures at which the monomer is 50% populated. Numbered subscripts refer to thermodynamic parameters for melting tetramer 1 and tetramer 2 in sequences H-1 to H-5 and tetramer and dimer respectively in sequence H-6. Errors were calculated according to the variance-covariance method (41).

along the major groove between T1-C6, C2-fC5 and C3-fC4 (Supplementary Figure S14). The exchangeable proton spectra exhibit several signals between 15–16 ppm, indicative of C:CH<sup>+</sup> base pairs. These base pairs occur between magnetically equivalent cytosines since each of these signals exhibits cross-peaks with only two amino protons (8–10 ppm). NMR assignments are specified in Supplementary Table S3.

DQF-COSY experiments indicate that deoxyriboses adopt an N-type conformation (except C6), while 2'F-araCs adopt an S-type conformation (Supplementary Figure S15). The three dimensional structure of H-4 was calculated on the basis of 284 experimental distance constraints using restrained molecular dynamics methods, and following standard procedures previously described by our group (37). All residues are well defined, with an RMSD of 0.9 Å (Supplementary Table S4). The final AMBER energies and NOE terms are reasonably low in all the structures, with no distance constraint violation >0.3 Å. The coordinates of the final refined structures were deposited in the PDB (code: 2N89).

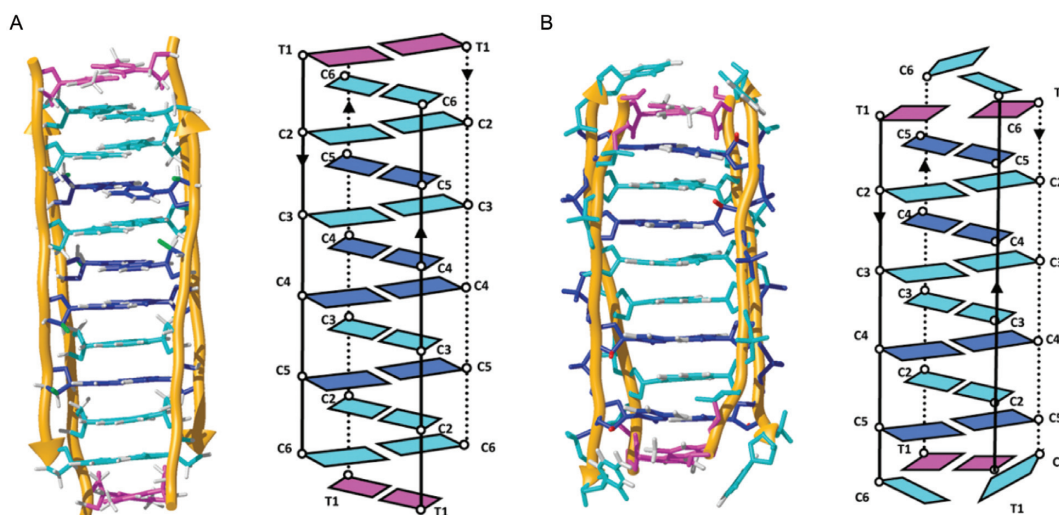
The structures show that 2'F-araC fits very well in the otherwise standard 3'E i-motif structure with the stacking order T1-C6-C2-fC5-C3-fC4 (Figure 6 and Supplementary Figure S16). The fluorine atoms in the S-type arabinoses point toward the major groove of the i-motif without affecting sugar–sugar contacts critical for i-motif stability (Supplementary Figure S16). Geometrical parameters are shown in Supplementary Table S5.

The major i-motif species at acidic pH is not the most stable one at neutral pH (Supplementary Figure S17). NMR data indicate that the terminal deoxycytidines (C6) are not protonated and are disordered at pH 7.0. Some key NOEs, such as those shown in Supplementary Figures S18 and 19, indicate that the intercalation order is T1-fC5-C2-fC4-C3 (Figure 6), which is different than that of the major conformer at acidic pH. Although the complete sequential assignment of the NMR spectra of the major conformer at neutral pH could not be carried out, a model structure consistent with the NOEs that could be assigned unambiguously was built and refined with molecular dynamics methods (Figure 6 and Supplementary Figure S19). Similar to the conformer observed at acidic pH, the neutral pH conformer maintains a 3'E topology with one fewer C:CH<sup>+</sup>,

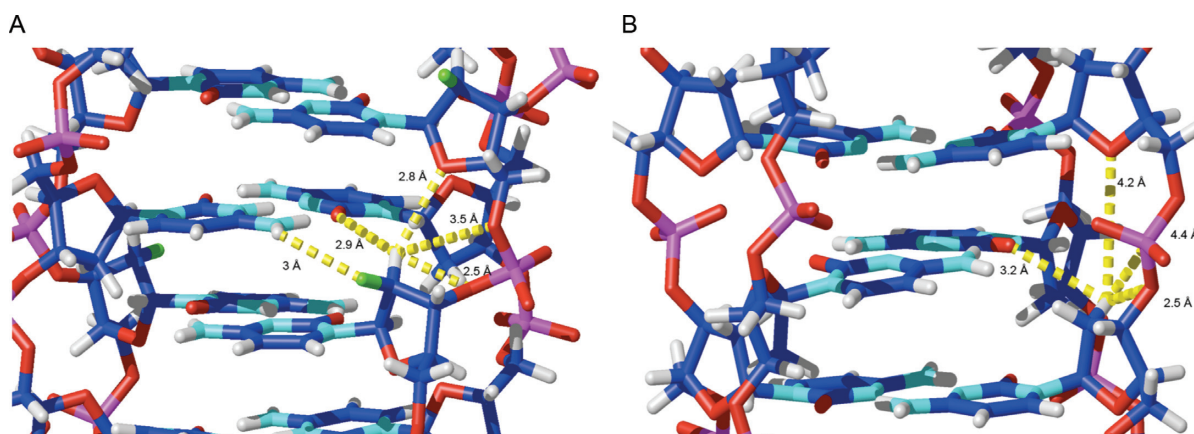
due to deprotonation of the terminal C:CH<sup>+</sup> base-pair. Thus, disruption of C6:C6<sup>+</sup> base pair upon raising the pH provokes a rearrangement of the intercalation order, preserving a 3'E topology. This reflects the well-known higher stability of 3'E vs 5'E i-motif topologies (52,53).

## DISCUSSION

Our groups and others have extensively studied the stabilizing effect of fluorine substitutions at the sugar C2' position in different nucleic acid motifs (54–57). 2'F-Arabinonucleotides, when incorporated into DNA strands, increase binding affinity toward RNA through formation of non-canonical 2'F...purine (H8) hydrogen bonds (58,59). Similar interactions have been observed in pure 2'F-ANA:2'F-ANA duplexes (45). Moreover, the strong electronegativity of fluorine affects the sugar charge distribution (in particular in the H2' proton) and can induce the formation of FC-H...O hydrogen bonds between sequential sugars. This interaction is responsible for the enhanced stability of 2'F-araG substituted G-quadruplexes (30,60) and 2'F-ANA:RNA hybrid duplexes (29). The structural analysis of H-4 did not provide evidence for FC-H...O hydrogen bonds with the appropriate geometry. However, a number of favorable sequential and inter-strand contacts involving 2'F-araC residues were observed (Figure 7). Many of these close contacts are facilitated by the S-type conformation of the 2'F-arabinoses, and hence are absent in unmodified i-motifs in which the sugar conformations are N-type. In particular, sequential FC-H2'...O4' and inter-strand FC-H2'...O2 distances are much shorter in S-type 2'F-arabinoses. Since fluorine electronegativity provokes a positive charge polarization at the geminal H2' proton, its close contacts with electron-dense oxygen atoms cause favorable electrostatic interactions. Other favorable electrostatic interactions involve sequential F-C-H2'...X (where X = O3', O5') and inter-strand F...H<sub>2</sub>N. In addition to favorable electrostatic interactions, a long-range inductive effect may also strengthen hemi-protonated 2'F-araC:2'F-araC<sup>+</sup> base pairs as seen in 2'F-RNA duplexes (56,61). Stronger base pairing and long-range electrostatic interactions may also explain the large increase in pH<sub>1/2</sub> observed for 2'F-araC substituted i-motifs compared to the pK<sub>a</sub> of free 2'F-araC nucleoside.



**Figure 6.** (A) Average solution structure and schematic representation of the major 3'E species of H-4 (5'-dTCC(fC)(fC)C-3') at pH 5.0. (B) Schematic representation of the major 3'E species of H-4 at pH 7.0. Color code: dC: cyan, 2'F-araC: blue, dT: magenta; and fluorine atoms: green.



**Figure 7.** Details of sugar contacts in (A) 2'F-araC substituted i-motif H-4 and in (B) an unmodified i-motif H-1. Close distances corresponding to favorable electrostatic interactions in the substituted i-motif are shown in yellow (the structure in the left panel is slightly rotated along the Y-axis to better illustrate the relevant distances). The electronegative fluorine and the positively polarized H2' alleviate the unfavorable electrostatic interactions between O2, O3', O4' and O5', characteristic of i-motif structures. Coordinates of the unmodified i-motif are 1YBL.

The possibility of modulating the stability and pH dependence of i-motif structures through chemical modification has attracted much attention in recent years. Most substitutions tested thus far have led to destabilization, the exceptions being LNA and PNA for selected sequences (21,25). 2'F-araC is a rather conservative DNA-like modification which preserves the structure of natural i-motifs with only minor alterations. This makes 2'F-araC substituted sequences excellent mimics to study molecular recognition processes involving i-motifs. For example, pull-down cell-based assays that identify i-motif binding proteins (including potential i-motif-specific antibodies) will be significantly facilitated by carrying out these experiments under physiological conditions. The fluorine modification also provides an excellent handle to detect i-motif-ligand interactions via  $^{19}\text{F}$  NMR methods. Finally, the nuclease resistance properties conferred by 2'F-araC substitutions are important for *in vivo* applications of i-motif-based nanodevices (14).

## CONCLUSION

We have shown that i-motifs, including tetrameric sequences and those formed by centromeric and telomeric DNA sequences, can be significantly stabilized by replacing dC units with 2'F-araC residues. The effects observed are noteworthy, resulting in some instances in stabilization of the structure by over 10 kJ/mol per 2'F-araC incorporation. Replacement of DNA with 2'F-araC does not alter the overall i-motif structure, creating the opportunity to correlate the structural information of transient arrangements of G4 and i-motifs within duplex DNA with transcription factor or polymerase activity under physiological conditions. This would bring i-motifs into focus as drug targets through the screening of small molecule libraries that bind to i-motifs (62).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENT

This work is dedicated to the Memory of Alfredo Villasante, valuable collaborator and friend.

## FUNDING

Funding for open access charge: NSERC Discovery grant (to M.J.D., A.K.M.); CIHR DDTP Training Grant (to H.A., R.H.V.); MINECO [BFU2014-52864-R to C.G.]; CSIC-JAE contract (to N.M.P.).

*Conflict of interest statement.* None declared.

## REFERENCES

- Choi, J. and Majima, T. (2011) Conformational changes of non-B DNA. *Chem. Soc. Rev.*, **40**, 5893–5909.
- Sen, D. and Gilbert, W. (1988) Formation of parallel 4-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
- Moye, A.L., Porter, K.C., Cohen, S.B., Phan, T., Zyner, K.G., Sasaki, N., Lovrecz, G.O., Beck, J.L. and Bryan, T.M. (2015) Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.*, **6**, 1–12.
- Gehring, K., Leroy, J.L. and Gueron, M. (1993) A tetrameric DNA-structure with protonated cytosine. Cytosine base-pairs. *Nature*, **363**, 561–565.
- Benabou, S., Aviñó, A., Eritja, R., González, C. and Gargallo, R. (2014) Fundamental aspects of the nucleic acid i-motif structures. *RSC Adv.*, **4**, 26956–26980.
- Day, H.A., Pavlou, P. and Waller, Z.A. (2014) i-Motif DNA: structure, stability and targeting with ligands. *Bioorg. Med. Chem.*, **22**, 4407–4418.
- Garavís, M., Escaja, N., Gabelica, V., Villasante, A. and González, C. (2015) Centromeric alpha-satellite DNA adopts dimeric i-motif structures capped by AT Hoogsteen base pairs. *Chem. Eur. J.*, **21**, 9816–9824.
- Garavís, M., Méndez-Lago, M., Gabelica, V., Whitehead, S.L., González, C. and Villasante, A. (2015) The structure of an endogenous *Drosophila* centromere reveals the prevalence of tandemly repeated sequences able to form i-motifs. *Sci. Rep.*, **5**, 1–10.
- Phan, A.T., Gueron, M. and Leroy, J.L. (2000) The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. *J. Mol. Biol.*, **299**, 123–144.
- Kang, C.H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. and Rich, A. (1994) Crystal-structure of intercalated 4-stranded D(C3t) at 1.4 angstrom resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 11636–11640.
- Kang, H.J., Kendrick, S., Hecht, S.M. and Hurley, L.H. (2014) The transcriptional complex between the BCL2 i-motif and hnRNP LL is a molecular switch for control of gene expression that can be modulated by small molecules. *J. Am. Chem. Soc.*, **136**, 4172–4185.
- Kendrick, S., Kang, H.J., Alam, M.P., Madathil, M.M., Agrawal, P., Gokhale, V., Yang, D.Z., Hecht, S.M. and Hurley, L.H. (2014) The dynamic character of the BCL2 promoter i-motif provides a mechanism for modulation of gene expression by compounds that bind selectively to the alternative DNA hairpin structure. *J. Am. Chem. Soc.*, **136**, 4161–4171.
- Guittet, E., Renciuik, D. and Leroy, J.L. (2012) Junctions between i-motif tetramers in supramolecular structures. *Nucleic Acids Res.*, **40**, 5162–5170.
- Dong, Y.C., Yang, Z.Q. and Liu, D.S. (2014) DNA nanotechnology based on i-motif structures. *Acc. Chem. Res.*, **47**, 1853–1860.
- Li, T. and Famulok, M. (2013) I-motif-programmed functionalization of DNA nanocircles. *J. Am. Chem. Soc.*, **135**, 1593–1599.
- Berger, I., Egli, M. and Rich, A. (1996) Inter-strand CH...O hydrogen bonds stabilizing four-stranded intercalated molecules: stereoelectronic effects of O4' in cytosine-rich DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 12116–12121.
- Sun, D. and Hurley, L.H. (2009) The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J. Med. Chem.*, **52**, 2863–2874.
- Kendrick, S. and Hurley, L.H. (2010) The role of G-quadruplex/i-motif secondary structures as cis-acting regulatory elements. *Pure Appl. Chem.*, **82**, 1609–1621.
- Cui, J., Waltman, P., Le, V.H. and Lewis, E.A. (2013) The effect of molecular crowding on the stability of human c-MYC promoter sequence I-motif at neutral pH. *Molecules*, **18**, 12751–12767.
- Day, H.A., Huguin, C. and Waller, Z.A.E. (2013) Silver cations fold i-motif at neutral pH. *Chem. Commun.*, **49**, 7696–7698.
- Kumar, N., Nielsen, J.T., Maiti, S. and Petersen, M. (2007) i-motif formation with locked nucleic acid (LNA). *Angew. Chem.*, **46**, 9220–9222.
- Pasternak, A. and Wengel, J. (2011) Modulation of i-motif thermodynamic stability by the introduction of UNA (unlocked nucleic acid) monomers. *Bioorg. Med. Chem. Lett.*, **21**, 752–755.
- Pérez-Rentero, S., Gargallo, R., González, C. and Eritja, R. (2015) Modulation of the stability of i-motif structures using an acyclic threoninol cytidine derivative. *RSC Adv.*, **5**, 63278–63281.
- Kanaori, K., Sakamoto, S., Yoshida, H., Guga, P., Stec, W., Tajima, K. and Makino, K. (2004) Effect of phosphorothioate chirality on i-motif structure and stability. *Biochemistry*, **43**, 5672–5679.
- Krishnan-Ghosh, Y., Stephens, E. and Balasubramanian, S. (2005) PNA forms an i-motif. *Chem. Commun.*, 5278–5280.
- Robidoux, S. and Damha, M.J. (1997) D-2-deoxyribose and D-arabinose, but not D-ribose, stabilize the cytosine tetrad (i-DNA) structure. *J. Biomol. Struct. Dyn.*, **15**, 529–535.
- Collin, D. and Gehring, K. (1998) Stability of chimeric DNA/RNA cytosine tetrads: implications for i-motif formation by RNA. *J. Am. Chem. Soc.*, **120**, 4069–4072.
- Fenna, C.P., Wilkinson, V.J., Arnold, J.R., Cosstick, R. and Fisher, J. (2008) The effect of 2'-fluorine substitutions on DNA i-motif conformation and stability. *Chem. Commun.*, 3567–3569.
- Martin-Pintado, N., Deleavey, G.F., Portella, G., Campos-Olivas, R., Orozco, M., Damha, M.J. and González, C. (2013) Backbone FC-H...O hydrogen bonds in 2'-F-substituted nucleic acids. *Angew. Chem.*, **52**, 12065–12068.
- Martin-Pintado, N., Yahyaee-Anzahae, M., Deleavey, G.F., Portella, G., Orozco, M., Damha, M.J. and González, C. (2013) Dramatic effect of furanose C2' substitution on structure and stability: directing the folding of the human telomeric quadruplex with a single fluorine atom. *J. Am. Chem. Soc.*, **135**, 5344–5347.
- Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum Nmr-spectroscopy of aqueous-solutions. *J. Biomol. NMR*, **2**, 661–665.
- Yu, C. and Levy, G.C. (1984) Two-dimensional heteronuclear noe (Hoesy) experiments - investigation of dipolar interactions between heteronuclei and nearby protons. *J. Am. Chem. Soc.*, **106**, 6533–6537.
- Goddard, T.D. and Kneller, D.G. *SPARKY*. 3rd edn. University of California, San Francisco.
- Cai, L., Chen, L.Q., Raghavan, S., Ratliff, R., Moyzis, R. and Rich, A. (1998) Intercalated cytosine motif and novel adenine clusters in the crystal structure of the *Tetrahymena* telomere. *Nucleic Acids Res.*, **26**, 4696–4705.
- Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, **273**, 283–298.
- Case, D.A., Pearlman, D.A. and Caldwell, J.W.I. III (2007) *AMBER7*. University of California, San Francisco.
- Soliva, R., Monaco, V., Gómez-Pinto, I., Meeuwenoord, N.J., Van der Marel, G.A., Van Boom, J.H., González, C. and Orozco, M. (2001) Solution structure of a DNA duplex with a chiral alkyl phosphonate moiety. *Nucleic Acids Res.*, **29**, 2973–2985.
- Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
- Lavery, R. and Sklenar, H. (1990) CURVES, helical analysis of irregular nucleic acids. *Laboratory of Theoretical Biochemistry*. CNRS, Paris.
- Koradi, R., Billeter, M. and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 29–55.
- Tellinghuisen, J. (2001) Statistical error propagation. *J. Phys. Chem. A*, **105**, 3917–3921.



42. Kanehara, H., Mizuguchi, M., Tajima, K., Kanaori, K. and Makino, K. (1997) Spectroscopic evidence for the formation of four-stranded solution structure of oligodeoxycytidine phosphorothioate. *Biochemistry*, **36**, 1790–1797.
43. Mata, G. and Luedtke, N.W. (2015) Fluorescent probe for proton-coupled DNA folding revealing slow exchange of i-motif and duplex structures. *J. Am. Chem. Soc.*, **137**, 699–707.
44. Kreutz, C., Kahlig, H., Konrat, R. and Micura, R. (2005) Ribose 2'-F labeling: A simple tool for the characterization of RNA secondary structure equilibria by F-19 NMR spectroscopy. *J. Am. Chem. Soc.*, **127**, 11558–11559.
45. Martín-Pintado, N., Yahyaee-Anzahaee, M., Campos-Olivas, R., Noronha, A.M., Wilds, C.J., Damha, M.J. and González, C. (2012) The solution structure of double helical arabino nucleic acids (ANA and 2'-F-ANA): effect of arabinoses in duplex-hairpin interconversion. *Nucleic Acids Res.*, **40**, 9329–9339.
46. Luo, J., Sarma, M.H., Yuan, R.D. and Sarma, R.H. (1992) Nmr-study of self-paired parallel duplex of D(Aaaaaccccc) in solution. *FEBS Lett.*, **306**, 223–228.
47. Geinguenaud, F., Liquier, J., Brevnov, M.G., Petrauskene, O.V., Alexeev, Y.I., Gromova, E.S. and Taillandier, E. (2000) Parallel self-associated structures formed by T,C-rich sequences at acidic pH. *Biochemistry*, **39**, 12650–12658.
48. Lim, K.W., Lacroix, L., Yue, D.J.E., Lim, J.K.C., Lim, J.M.W. and Phan, A.T. (2010) Coexistence of two distinct G-quadruplex conformations in the hTERT promoter. *J. Am. Chem. Soc.*, **132**, 12331–12342.
49. Patel, D.J., Kozlowski, S.A., Marky, L.A., Broka, C., Rice, J.A., Itakura, K. and Breslauer, K.J. (1982) Premelting and melting transitions in the D(Cgcgaattcgcg) self-complementary duplex in solution. *Biochemistry*, **21**, 428–436.
50. Rule, G.S. and Hitchens, T.K. (2006) *Fundamentals of Protein NMR Spectroscopy*. Springer, The Netherlands.
51. Wemmer, D.E., Chou, S.H., Hare, D.R. and Reid, B.R. (1985) Duplex-hairpin transitions in DNA - Nmr-Studies on Cgcgtatcgcg. *Nucleic Acids Res.*, **13**, 3755–3772.
52. Leroy, J.L., Snoussi, K. and Gueron, M. (2001) Investigation of the energetics of C-H...O hydrogen bonds in the DNA i-motif via the equilibrium between alternative intercalation topologies. *Magn. Reson. Chem.*, **39**, S171–S176.
53. Malliavin, T.E., Gau, J., Snoussi, K. and Leroy, J.L. (2003) Stability of the I-motif structure is related to the interactions between phosphodiester backbones. *Biophys. J.*, **84**, 3838–3847.
54. Lech, C.J., Li, Z., Heddi, B. and Phan, A.T. (2012) 2'-F-ANA-guanosine and 2'-F-guanosine as powerful tools for structural manipulation of G-quadruplexes. *Chem. Commun.*, **48**, 11425–11427.
55. Li, Z., Lech, C.J. and Phan, A.T. (2014) Sugar-modified G-quadruplexes: effects of LNA-, 2'-F-RNA- and 2'-F-ANA-guanosine chemistries on G-quadruplex structure and stability. *Nucleic Acids Res.*, **42**, 4068–4079.
56. Pallan, P.S., Greene, E.M., Jicman, P.A., Pandey, R.K., Manoharan, M., Rozners, E. and Egli, M. (2011) Unexpected origins of the enhanced pairing affinity of 2'-fluoro-modified RNA. *Nucleic Acids Res.*, **39**, 3482–3495.
57. Manoharan, M., Akinc, A., Pandey, R.K., Qin, J., Hadwiger, P., John, M., Mills, K., Charisse, K., Maier, M.A., Nechev, L. et al. (2011) Unique gene-silencing and structural properties of 2'-fluoro-modified siRNAs. *Angew. Chem. Int. Ed.*, **50**, 2284–2288.
58. Watts, J.K., Martín-Pintado, N., Gomez-Pinto, I., Schwartzenruber, J., Portella, G., Orozco, M., Gonzalez, C. and Damha, M.J. (2010) Differential stability of 2' F-ANA.RNA and ANA.RNA hybrid duplexes: roles of structure, pseudohydrogen bonding, hydration, ion uptake and flexibility. *Nucleic Acids Res.*, **38**, 2498–2511.
59. Anzahaee, M.Y., Watts, J.K., Alla, N.R., Nicholson, A.W. and Damha, M.J. (2011) Energetically important C-H...F-C pseudohydrogen bonding in water: evidence and application to rational design of oligonucleotides with high binding affinity. *J. Am. Chem. Soc.*, **133**, 728–731.
60. Peng, C.G. and Damha, M.J. (2007) G-quadruplex induced stabilization by 2'-deoxy-2'-fluoro-D-arabinonucleic acids (2'-F-ANA). *Nucleic Acids Res.*, **35**, 4977–4988.
61. Patra, A., Paolillo, M., Charisse, K., Manoharan, M., Rozners, E. and Egli, M. (2012) 2'-fluoro RNA shows increased Watson-Crick H-bonding strength and stacking relative to RNA: evidence from NMR and thermodynamic data. *Angew. Chem. Int. Ed. Engl.*, **51**, 11863–11866.
62. Stephenson, A.W.I., Partridge, A.C. and Filichev, V.V. (2011) Synthesis of beta-pyrrolic-modified porphyrins and their incorporation into DNA. *Chem. Eur. J.*, **17**, 6227–6238.