"Stabilization of telomeric i-motif structures by (2'*S*)-2'-deoxy-2'-methyl-cytidine residues" Aviñó, A., Dellafiore, M., Gargallo, R., González, C., Iribarren, A.M., Montserrat, J., Eritja, R. ChemBioChem, 18, 1123-1128 (2017); doi: 10.1002/cbic.201700112

Stabilization of telomeric i-motif structures by (2'*S*)-2'-deoxy-2'-*C*-methyl-cytidine residues

Anna Aviñó, * [a,b] María Dellafiore, [c] Raimundo Gargallo, [d,e] Carlos González, [f,e] Adolfo M. Iribarren[g], Javier Montserrat [h] and Ramon Eritja * [a,b]

[a]Institute for Advanced Chemistry of Catalonia (IQAC), Spanish Council for Scientific Research (CSIC), Jordi Girona 18-26, 08034 Barcelona, Spain; [b]Networking Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN); [c] NGEBI (CONICET), Vuelta de Obligado 2490-(1428), Buenos Aires, Argentina; [d]Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí i Franquès, 1-11, 08028 Barcelona, Spain; [e] BIOESTRAN, associated unit UB-CSIC; [f]Institute of Physical Chemistry Rocasolano (IQFR), Spanish Council for Scientific Research (CSIC), Serrano 119, 28006 Madrid, Spain; [g]Laboratorio de Biotransformaciones, Universidad Nacional de Quilmes, Roque Saenz Peña 352 (1876) Bernal, Prov. de Buenos Aires, Argentina; [h]Universidad Nacional de General Sarmiento J.M. Gutiérrez 1150, 1613 Los polvorines, Buenos Aires, Argentina.

Introduction

Nucleic acids can adopt secondary structures that play important roles in several cellular functions. Among them, i-motifs are present in oncogenic regions as well as in human telomeric DNA regions indicating their possible role during the regulation of oncogene expression at the transcription level.[1-3] Recently, i-motif structures were also found in the CCG triplet repeats,[4] in the G_2C_4 hexanucleotide repeat expansion mutation sequence,[5] and in centromeric sequences.[6] In addition, several proteins have been studied that specifically bind C-rich telomeric sequences.[7-10]

These i-motifs are intercalated tetraplex structures formed by the association of a hemiprotonated cytosine base pair ($C\text{-}CH^+$), and close sugar-sugar contacts stabilized by CH•••O interactions[11] in C-rich sequences. The $C\text{-}CH^+$ bond has been compared to the strong hydrogen bonds produced by equally hydrogen sharing observed in basic organic molecules with low hopping energy barriers.[12] Moreover, there are several analytical and biomedical applications of cytosine-rich

oligonucleotides[13] exploiting the advantages of i-motif pH-dependent formation especially in the nanotechnology field.[14,15]

Chemically modified nucleic acids are widely explored as therapeutic drugs, and diagnostic tools, in target validation for drug development and in other applications in the field of biomedicine and nanotechnology. In addition, chemically modified nucleic acids are good molecular models for the elucidation of nucleic acids structures, as well as excellent tools for physicochemical studies and for probing protein-nucleic acids interactions. To date, a large number of modifications have been developed and analyzed, providing important information for the design of new derivatives with tailor-made properties, including pairing specificity, nuclease resistance and cellular uptake.[16]

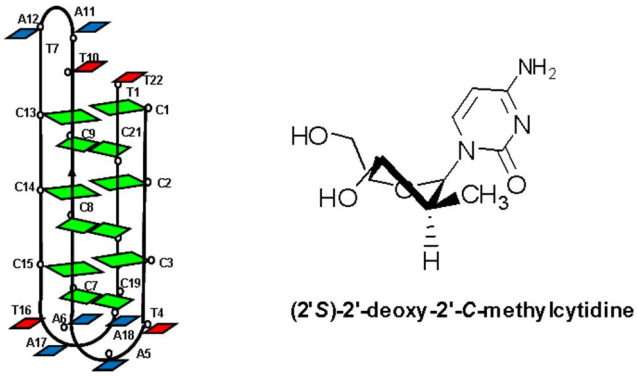
Several C-rich modified oligonucleotides have been developed. with the aim of increasing nuclease resistance and i-motif stability.[17] In most of cases, modifications are not well-tolerated in an i-motif because they might either destabilize the structure or prevent its formation.[17,18] Base modifications such as 5-methylcytosine and 5-hydroxymethylcytosine are permissive bases that can be adapted to the i-motif structures. They show a position-dependent stability and different behavior under experimental crowding conditions.[19-21] These bases are important epigenetic markers in the human genome.

The i-motif structure consists of two wide and two extremely narrow grooves. The stability of this structure depends on the repulsion between adjacent negatively charged phosphates backbones across these minor grooves. Various backbone substitutions have been studied in order to form intermolecular and intramolecular i-motifs.[17,18] Modifications such as Locked Nucleic Acid (LNA)[22,23], Unlocked Nucleic Acids (UNA),[24] acyclic threoninol[25] and peptide nucleic acids (PNA)[26] backbones have been introduced in biologically relevant i-motif structures.

The most common conformation of the sugars found in the i-motif structures is the 3'-endo. The ribose in RNA adopts a 3'-endo pucker but the i-motif is destabilized.[27] Steric interactions such as 2'-O-Me[27] or 2'-F[28] also produce a decrease in the stability. However, modifications with arabinose derivatives in which the 2'-OH group points toward the major groove are well tolerated within the i-motif structure.[29] Recently, 2'-F-araC have been studied showing stabilized i-motifs over a wide pH range without modifying the overall structure.[30]

To further explore the tolerance of 2'-substituyents in the arabino sugar configuration, we incorporated the (2'*S*)-2'-deoxy-2'-*C*-methyl-cytidine unit ($C_{Me}Up$)[31-33] (Scheme 1) in a well described C-rich fragment of the vertebrate telomere $(C_3TA_2)_3C_3T$. This i-motif structure folds intramolecularly with four cytidine stretches connected by three TAA loops and six C-CH^+ base pairs (Scheme 1).[34] A similar sequence has also been reported to show a polymorphism with two possible conformations.[35] The CMeUp derivative is characterized by its nuclease stability[36] and for its influence as conformationally restricted analogue on the activity of 10-23 DNAzymes. [36,37]

Scheme 1. Representation of C-rich fragment of the vertebrate telomere $(C_3TA_2)_3C_3T$ studied in this work and chemical structure of (2'*S*)-2'-deoxy-2'-*C*-methylcytidine unit ($C_{Me}Up$).



The introduction of carbon substituents at the 2'-position of deoxynucleotides deeply affects the sugar conformation state. In this sense, the 2'-deoxy-2'-*C*-methyl-nucleosides provided interesting properties since different preferred sugar conformations were observed depending on the absolute conformation at the 2'-carbon.[38-40] The (2'*S*)-2'-deoxy-2'-*C*-methyl-nucleosides mainly adopt the C3'-endo sugar puckering while the (2'*R*)-2'-deoxy-2'-*C*-methyl-nucleosides prefer the C2'-endo conformation. We rationalized that (2'*S*)-2'-deoxy-2'-*C*-methylcytidine ($C_{Me}Up$) with a methyl group in a pseudoequatorial orientation that stabilizes a North (RNA-like) puckering was an interesting unit to introduce in i-motif structures. Therefore, in this work we replaced dC by $C_{Me}Up$ in different positions of the C-core stretches of the C-rich fragment of the vertebrate telomere $(C_3TA_2)_3C_3T$. We selected single substitutions at 5'-terminal stretch ($VT1C_{Me}Up$), internal positions ($VT2C_{Me}Up$) or double substitution involved in the same internal C-CH^+ base pairs ($VT2,14C_{Me}Up$) (Table 1 and Figure S1).

Table 1. Sequences used in this study. **C**: 2'*S*-2'-deoxy-2'-*C*-methylcytidine unit ($C_{Me}Up$).

Name	Sequence
VTWT	CCCTAACCCTAACCCTAACCCT
VT1 $C_{Me}Up$	**C**CCTAACCCTAACCCTAACCCT
VT2 $C_{Me}Up$	C**C**CTAACCCTAACCCTAACCCT
VT2,14 $C_{Me}Up$	C**C**CTAACCCTAAC**C**CTAACCCT

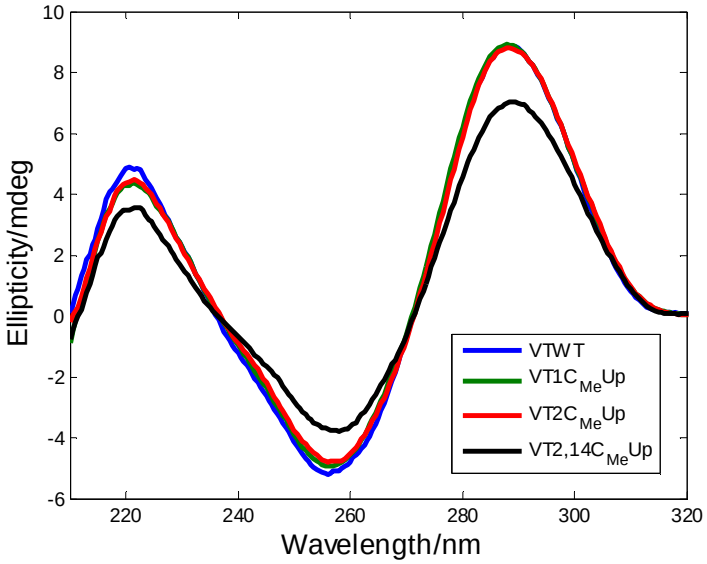
Results

CD spectra

First, the $C_{Me}Up$ i-motif formation was investigated by circular dichroism (CD). Figure 1 and Figure S2 shows the CD spectra recorded for all sequences at pH 5.5 and 15 ºC. We obtained the characteristic i-motif CD signature with positive and negative peaks at 287 and 262 nm, respectively, for all the modified oligonucleotides prepared in this study and at different pHs. In all cases, similar CD signature as unmodified WTVT

was observed, indicating that the modified nucleoside is well incorporated in the overall structure and at all pH studied.

Figure 1. CD spectra of the sequences measured at pH 5.5 and 15 ºC.



Acid-base titrations

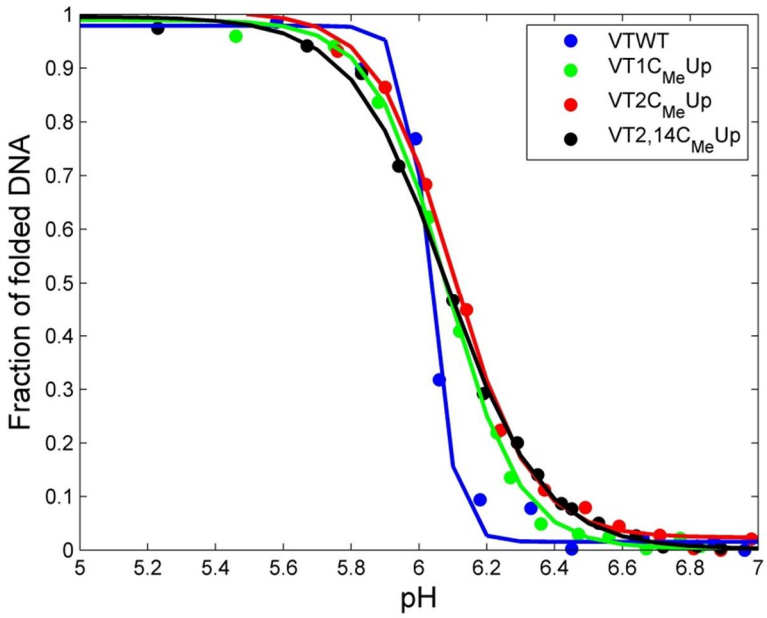
CD spectra of C_{Me}Up modified i-motif against pH variation were recorded. Acid-base titrations were carried out in the pH range of 3-7 and at 25 ºC (Figure 2 and Figure S3) show the DNA folded fraction vs. pH adjusted to a sigmoidal curve. The pH transition midpoint ($pH_{1/2}$) at which 50% of the oligonucleotide is folded into an i-motif structure was determined according to the procedure previously described. In all cases, the titration curves were analyzed assuming a single protonation event in order to extract populations of folded (protonated) or unfolded (deprotonated) states. This assumption was previously checked by means of multivariate analysis (data not shown). The calculated parameters from each curve are summarized in Table 2. In general, the modifications increase slightly the pH transition midpoint. This increase is almost within the range of experimental uncertainty. The C_{Me}Up derivative has a methyl group at the 2' position of the ribose. For this reason the introduction of the methyl group is not expected to have a strong impact in the pH transition midpoint. The proposed mutations have a more significant effect on the pH transitional range (R_T), which varies from 0.22 (for the wild-type) to 0.5 for the VT1C_{Me}Up and VT2C_{Me}Up, and even 0.64 for the double modification. In another words, the acid-base titration curve of the unmodified sequence is very sharp indicating a strong cooperativity in the deprotonation of the i-motif. Once the first protons are removed, it is easier to remove the rest of them leading to a quick destabilization of the i-motif. On the contrary the shape of the curves obtained for the CMeUp modified sequences are wider (less

cooperative). These differences may explain the stabilization of the i-motif at pHs higher than the midpoint (see below).

Table 2. pH transition midpoints and transitional range for the studied DNA sequences.

Name	$pH_{1/2}$	$pH_{10\%}$	$pH_{90\%}$	R_T
VTWT	6.03±0.02	5.92	6.14	0.22
VT1 C_{Me}Up	6.08±0.02	5.83	6.33	0.50
VT2 C_{Me}Up	6.10±0.02	5.85	6.39	0.54
VT2,14 C_{Me}Up	6.08±0.02	5.76	6.40	0.64

Figure 2. Dependence of the fraction of folded DNA with pH. The acid-base titrations were carried out at 25 ºC.



Thermal stability of i-motif

Next, we studied the thermal stability of the oligonucleotides using CD and molecular absorption spectroscopies. The thermal denaturation curves were obtained by measuring the UV absorbance of the oligonucleotide solutions at 295 nm at pHs 5.0-7.0 (Fig. S4). At this wavelength the denaturation of the i-motif affords a decrease in the absorbance from where the midpoint (T_m) can be determined as the minimum of the first derivative (Table 3 and Figure S4). The denaturation may also be followed at 260 nm obtaining similar results but in this case the denaturation process increases the absorbance as shown in Figures S5 and S6.

Table 3. Melting temperatures determined for the studied sequences in 20 mM phosphate buffer and 150 mM KCl at 295 nm. A two-state process has been considered. The uncertainty in the determination of T_m at pH 6.0, 5.5 and 5.0 is ±0.7 ºC.

Sequence	pH	T_m (ºC)	ΔT_m (ºC)
VTWT	6.5	≤12	--
VT1C_{Me}Up	6.5	~15	~3
VT2C_{Me}Up	6.5	~15	~3
VT2,14C_{Me}Up	6.5	18	~6
VTWT	6.0	23.9	--
VT1C_{Me}Up	6.0	24.5	0.6
VT2C_{Me}Up	6.0	27.0	3.1
VT2,14C_{Me}Up	6.0	28.5	4.6
VTWT	5.5	43.1	--
VT1C_{Me}Up	5.5	44.3	0.8
VT2C_{Me}Up	5.5	44.8	1.7
VT2,14C_{Me}Up	5.5	45.4	2.3
VTWT	5.0	54.8	--
VT1C_{Me}Up	5.0	55.5	0.7
VT2C_{Me}Up	5.0	55.4	0.6
VT2,14C_{Me}Up	5.0	54.6	-0.2

The determined T_m values showed that the C_{Me}Up-modified oligonucleotides slightly stabilize the i-motif structure, depending on the pH. shifting their T_m to higher values. At pH 5.0, the small changes on T_m values are not significant, indicating that the modification do not change the stability of the i-motif structure at this pH. The differences are more pronounced at pH 5.5, with an increase in thermal stability up to 2.3 ºC in the double modified variant and around 1.7 ºC for the single modification in VT2C_{Me}Up. At pH 6.0 there is a clear stabilization of the double modified variant (ΔT_m = 4.6 ºC) followed by VT2C_{Me}Up (ΔT_m = 3.1 ºC). The shifts in T_m values in relation to that of the wild-type sequence agree with the behavior of the pH transitional range RT mentioned above. Therefore, as this range becomes wider, the i-motif structure can be observed at higher pH values, even though as a minor species. Despite the analysis of the transitions at pH 6.5 is difficult, it is still observable for the double modified variant. Surprisingly, at pH 7, in which almost no i-motif is observed in the natural sequence, the denaturation curve measured at 295 nm of VT2,14C_{Me}Up variant (Figure S4) still shows a small hypochromic change that indicates the presence of the i-motif structure. These results suggest C_{Me}Up modification stabilizes human vertebrate i-motif structures even near neutral pH values.

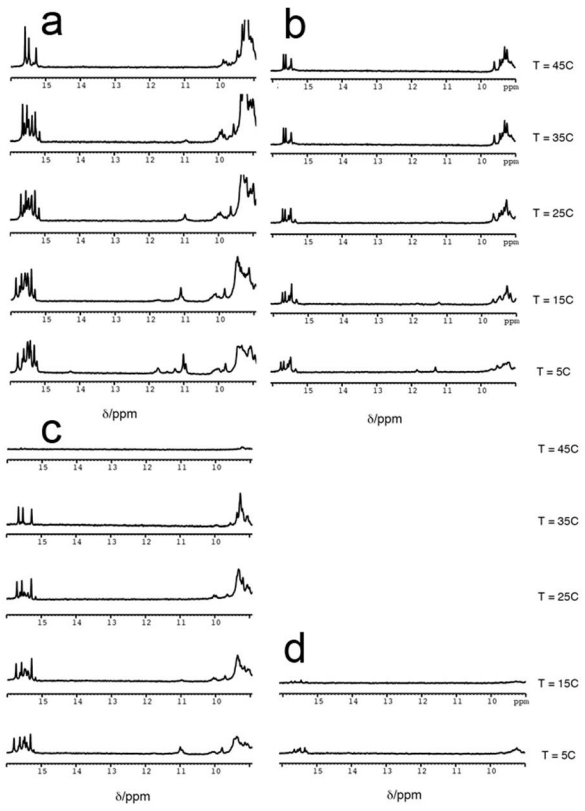
Similar results were obtained by following the denaturation of the i-motif sequences by circular dichroism (see supporting information). At pH 5.5 and 6.0 the higher stability of C_{Me}Up-modified oligonucleotides was confirmed being the double modified variant the most stable followed by VT2C_{Me}Up and VT1C_{Me}Up. Although there are some differences in Tm values obtained by CD- and UV-absorbance methods, the

relative order of stability was confirmed. At pH 6.5 the transitions are still observable but the first derivatives are too noisy and the estimated T_m's are not reliable. At pH 7.0 only the double modified variant shows changes in the CD spectra that may confirm the presence of i-motif structure at low temperatures.

NMR spectra

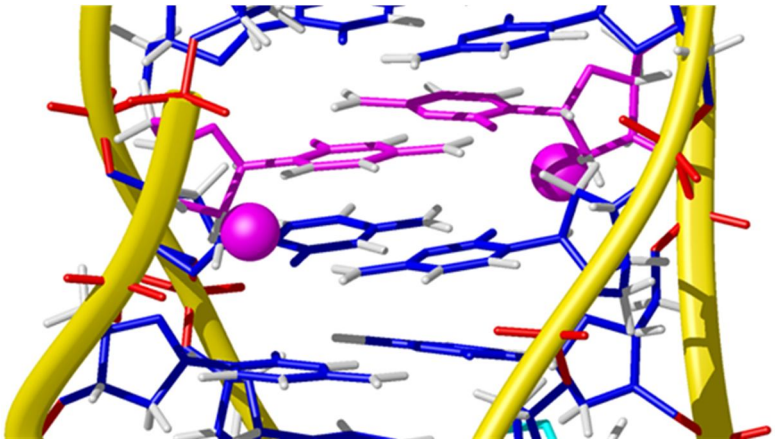
We further analyzed the structure of the modified i-motifs by NMR spectroscopy. ^{1}D NMR spectra show the characteristic cytosine imino protons in hemi-protonated C:CH^+ base pairs at around 15-16 ppm (Figure 3a,c). At neutral pH i-motif formation is significantly favored for VT2,14C_{Me}Up in comparison to the non-modified sequence, which does not exhibit the characteristic i-motif signals under these conditions. 2D NOESY spectra exhibit the characteristic cross-peak pattern of i-motif (Figure S7). Although the complete assignment of the NMR spectra is beyond the scope of this study, the 2'–CH_3 signals of the modified nucleotides could be identified (0.46 and 0.81 ppm). Their NOEs with the amino protons of hemiprotonated cytosines, shown in Figure S7 indicate that these methyl groups are oriented towards the i-motif major grooves, as expected (Fig. 3e). No steric clash with neighboring residues is expected in this orientation. To confirm this point, a molecular dynamics calculation was carried out with the AMBER program. After a 1-ns molecular dynamics run, the model structure of VT2,14C_{Me}Up remains stable (see Figure S8), with the two 2'-methyl groups well accommodated in the major grooves, and the sugar of the C_{Me}Up residues in North conformations (Figure 4).

Figure 3. Exchangeable proton region of the NMR spectra at different temperatures of VT2,14C_{Me}Up at pH 5 (a) and pH 7 (c), and the unmodified human telomeric sequence at pH 5 (b) and pH 7 (d).



These results confirm that the arabino sugar configuration is well tolerated within the i-motif structure, [29,30] and permits different substitutions at the 2' positions which are not allowed in the ribose configurations due to bad steric contacts throughout the i-motif minor groove. 2'-Fluorine substitutions provoke the most dramatic effect, since the fluorine electronegativity causes favourable changes in the charge distribution within the sugar. These changes in charge distribution are not expected in the case of 2'–CH_3 substitutions and, consequently, the stabilization effect is not so pronounced. However, (2'*S*)-2'-deoxy-2'-*C*-methyl-cytidine has a strong tendency to adopt a C3'-endo conformation, which is usually preferred in i-motifs. Preorganization of sugar in the proper conformation is most probably the reason of the observed stabilization of i-motif.

Figure 4. Model of VT2,14C_{Me}Up based on the structure of the telomeric sequence. (2'*S*)-2'-deoxy-2'-*C*-methyl-cytidine are shown in magenta with 2'-CH_3 groups displayed as spheres.



Discussion

Due to the potential applications of i-motif structures, there is an increasing interest in designing of modified cytosine derivatives capable to form stable i-motif particularly at near neutral pHs.[13-15] One of the first studied modifications was the backbone-modified derivatives known as peptide nucleic acids.[26] Some stabilization of the i-motif at neutral pH was observed, but these results were obtained in short tetrameric structures and are difficult to extrapolate to biologically relevant intramolecular i-motifs. As cytidine may be subjected to epigenetic modifications, the presence of 5-methyl-C and 5-hydroxymethyl-C in the C-rich telomere sequence has been studied, thus showing small changes in the stability of the modified i-motifs.[19-21] Interestingly, it has been observed that the inclusion of ribocytidines destabilizes i-motifs[27] but arabino-cytidines provoke the opposite effect.[29] It has also demonstrated that the presence of 2'-fluoro-araC produces very stable i-motifs even at neutral pH,[30]

mainly due to favorable electrostatic interactions provoked by the electrowithdrawing properties of the fluorine atom.

In this work we have studied the effect of the (2'*S*)-2'-deoxy-2'-*C*-methylcytidine derivatives on the stability of the i-motif, demonstrating that the substitution of a -H for a CH_3 group gives rise to more stable i-motifs and that these modified structures can also be observed at pH near neutrality. We replaced dC by CMeUp in different positions of the C-core stretches of the C-rich fragment of the vertebrate telomere (C3TA2)3C3T selecting single substitutions at 5'-terminal external position (VT1CMeUp), and single (VT2CMeUp) and double replacements (VT2,14CMeUp) in the next internal position. Thermal denaturation studies followed by UV- and CD-analyses showed that the most stable variant at mild acidic pH was the double modified sequence (VT2,14CMeUp) followed by the single modified sequence at internal position (VT2MeCup). The single modified sequence at external position (VT1CMeup) is the less stable but still it is more stable than the unmodified sequence at mild acidic pH. The most evident demonstration of the existence of i-motif at neutral pH for the double modified variant comes from the presence of the exchangeable protons at pH 7.0 in the NMR spectra of VT2,14CMeUp up to 35 ºC. Moreover, UV- and CD-spectroscopies confirm a distinct behavior of the double modified variant at pH 7.0 in accordance with the presence of an i-motif structure for VT2,14CMeUp at low temperatures.

A remarkable observation found is the low influence of the C_{Me}Up-modification on the pH midpoint transition values which oscillates between 6.03 and 6.10 (Table 2). On the contrary, other cytidine derivatives such as 5-methyl-C [20] and 2'-Fluoro-araC[30] induce larger variations in the i-motif $pH_{1/2}$ values that explain the changes in the stability. In our case the C_{Me}Up modification induces an alteration in the acid-base profile of i-motifs indicating that the protonation and deprotonation of cytidines are less cooperative. Therefore, the observed trend is that at pHs higher than the midpoint the stability of the i-motifs increases while at lower pHs it decreases.

It is interesting to compare these results with the stabilization provoked by the 2'-F-araC derivatives. Both studies confirm that i-motifs tolerate well 2'-substitutions in the arabino sugar configuration. However, 2'-F-araC derivatives stabilize i-motif at all pHs while (2'*S*)-2'-deoxy-2'-*C*-methyl-cytidine do not stabilize i-motif at pH 5.0. In contrast to the 2'-F substitution, the methyl group at 2' should not alter the charge distribution in the sugar and, therefore, the favorable electrostatic interactions involving 2'-F-arabinoses are absent. Consequently, the stabilization observed for (2'*S*)-2'-deoxy-2'-*C*-methyl-cytidine substitution is less pronounced. In this case, the preference of C_{Me}Up derivatives for the north-sugar conformation is probably the main factor in the stabilization. These results open the door to the design of other stabilizing *C*-derivatives.

Conclusions

In conclusion, we studied the effect of the replacement of the dC by C_{Me}Up in the ability to form intramolecular i-motifs. The relative stability of the modified structures have been analyzed by UV, CD and NMR spectroscopies, showing that C_{Me}Up residues induce a stabilization of the i-motif being the C_{Me}Up: C_{Me}Up pair more stable

than the C_{Me}Up: dC pair. This stabilization could be used to modulate the stability of i-motif structures at mild acidic to neutral pH values.

Experimental Section

Oligonucleotide synthesis

The oligonucleotides listed in Table 1 were prepared on a DNA/RNA Applied Biosystems 394 synthesizer by solid-phase 2-cyanoethylphosphoroamidite chemistry at a 0.2 µmol scale. Sequences containing C_{Me}Up were prepared using (2'*S*)-5′-*O*-(4,4′-dimethoxytrityl)-2'-deoxy-2'-*C*-methyl-N^4-benzoylcytidine-3′-O-(2-cyanoethyl-*N*,*N*-diisopropyl)-phosphoramidite that was prepared as reported previously. [32,33]

The oligonucleotides were synthesized in the DMT-ON mode. After the standard ammonia treatment (conc. aq. NH_3, 55 ºC, 6h), the oligonucleotides were purified using Glen-Pack™ DNA purification cartridges (Glen Research), analyzed by HPLC on a Semipreparative column: X-Bridge™ OST C_{18} (10x50 mm, 2.5 µm); 10 min linear gradient from 0 % to 30%, flow rate 2 mL•min^{-1}; solution A was 5% acetonitrile (ACN) in 0.1 M aqueous triethylammonium acetate (TEAA) and B 70% ACN in 0.1 M aqueous TEAA and confirmed by Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) spectra (Table 1, Figure S1).

UV Melting studies

UV thermal denaturation data[41] were obtained on a JASCO V-650 spectrophotometer equipped with a Peltier temperature control. Oligonucleotides were resuspended in 20 mM citrate/phosphate buffer pH 5.0 and 5.5, phosphate buffer pH 6.0, 6.5 and 7.0 containing 50 mM KCl to form solutions with a 5 µM concentration. These solutions were heated at 85 ºC and slowly cooled down to room temperature for annealing. The resulting solutions were stored at 4 ºC overnight. The thermal denaturing curves were obtained following the change of the absorption at 295 nm from 10 ºC to 80 ºC with a linear temperature ramp of 0.5 ºC•min-1. The dissociation melting temperatures were determined as the midpoint of the transition (T_m) using the first derivative of the experimental data. T_m values were determined as an average of three experiments.

CD spectroscopy

Oligonucleotides were resuspended in 20 mM citrate/phosphate buffer and 50 mM KCl at pH 5.0, 5.5 and in phosphate buffer at pH 6.0, 6.5 and pH 7.0 to form solutions with a concentration of 1.3 µM. These solutions were heated at 85ºC and slowly cooled down to room temperature for annealing and stored at 4ºC overnight. The thermal denaturing curves were obtained in a single experiment following the change of the CD absorption at 280 nm from 5ºC to 80ºC with a linear temperature ramp of 0.5 ºC•min-1 on a JASCO spectropolarimeter J-815V equipped with a Julabo F-25/HD temperature control unit. Spectra were registered at 15 ºC over a range of 210-350 nm with a scanning speed of 100 nm•min^{-1}, a response time of 4 s, a 0.5 nm

data pitch and a 1nm bandwidth. CD absorption at 280 nm data as a function of temperature were analysed as described elsewhere. [42]

Acid-base titrations

Acid-base titrations[25,43] of the sequences were monitored using the Jasco spectropolarimeter above mentioned. Experimental conditions were as follows: 25 ºC and 20 mM phosphate buffer with 50 mM KCl. Titrations were carried out by adjusting the pH value of solutions containing 2 µM of oligonucleotides in a Hellma quartz cell (10 mm path length, 3 mL). CD spectra were recorded in a pH-stepwise fashion. The ellipticity values at 280 nm were converted into "fraction of folded DNA" using the equation:

$$fraction\ of\ folded\ DNA = 1 - \frac{BS_{fold} - ellipticity_{pH}}{BS_{fold} - BS_{unfold}}$$

Where $ellipticity_{pH}$ is the ellipticity at 295 nm at a given pH, and "BS_{fold}" and "BS_{unfold}" correspond to the baseline values of the folded and unfolded species, respectively. These chosen baseline values correspond to pH values around 3.5 and 7.5, respectively. The transition midpoint of each transition ($pH_{1/2}$) was obtained by plotting the "fraction of folded DNA" vs. pH, and fitting to a sigmoidal using the Curve Fitting tool included in the Optimization Toolbox for Matlab (MathWorks, Natick, MA, USA). From the calculated values, the pH transitional range (RT, range between 10% and 90% of the unfolded structure) was also determined. The validity of this two-state transition process was validated prior to the calculations using multivariate analysis.

NMR spectroscopy and molecular modeling of structures

Samples for NMR experiments were dissolved in either D_2O or 9:1 H_2O/D_2O in 10 mM sodium phosphate buffer. The sample concentration was 0.4 mM. The samples were annealed before the NMR experiments Experiments were carried out at different pH values, ranging from 5.0 to 7.0. The pH was adjusted by adding aliquots of concentrated solution of either DCl or NaOD. All NMR spectra were acquired on Bruker Advance spectrometers operating at 600 and 800 MHz, equipped with cryoprobes and processed with the TOPSPIN software. NOESY spectra were acquired with mixing times of 150 and 250 ms. TOCSY spectra were recorded with the standard MLEV-17 spin-lock sequence and a mixing time of 80 ms. In most of the experiments in H_2O, water suppression was achieved by including a WATERGATE module[44] in the pulse sequence prior to acquisition. Molecular structures were calculated with the SANDER module of the molecular dynamics package AMBER.

Acknowledgements

This work was supported by the Spanish Ministry of Economy and Competitiveness (grants CTQ2014-52588-R, CTQ2012-38616-C02-02, BFU2014-52864-R and CTQ2015-66254-C2-2-P), and the Generalitat de Catalunya (grant 2014 SGR 1106; 2014 SGR 187). CIBER-BBN is an initiative funded by the VI National R&D&I Plan 2008–2011, Iniciativa Ingenio 2010, Consolider Program, CIBER Actions and are financed by the Instituto de Salud Carlos III with assistance from the European

Regional Development Fund. MD was recipient of an EMBO short term fellowship (ASTF 515-2015).

Keywords: i-motif, (2'S)-2'-deoxy-2'-C-methyl-cytidine, telomere, Nuclear Magnetic Resonance, Circular Dichroism, thermal denaturation studies.

REFERENCES

[1] M. Gueron, J. L. Leroy, *Curr. Opin. Struct. Biol*., **2000**, *10*, 326-331.

[2] T. Simonsson, P. Pecinka, M. Kubista, *Nucleic Acids Res*., **1998**, *26*, 1167-1172.

[3] S. Kendrick, Y. Akiyama, S. M. Hecht, L. H. Hurley, *J. Am. Chem. Soc*., **2009**, *131*, 17667-17676.

[4] Y.-W. Chen, C.-R. Jhan, S. Neidle, M.-H. Hou, *Angew. Chem. Int. Ed*., **2014**, *53*, 10682-10686.

[5] A. Kovanda, M. Zalar, P. Sket, J. Plavec, B. Rogelj, *Sci. Rep*., **2015**, *5*, 17944.

[6] M. Garavís, M. Méndez-Lago, V. Gabelica, S. L. Whitehead, C. González, A. Villasante, *Sci. Rep*., **2015**, *5*, 13307.

[7] L. Lacroix, H. Lienard, E. Labourier, M. Djavaheri-Mergny, J. Lacoste, H. Leffers, J. Tazi, C. Helene, J. L. Mergny, *Nucleic Acids Res*., **2000**, *28*, 1564-1575.

[8] E. Marsich, A. Piccini, L. E. Xodo, G. Manzini, *Nucleic Acids Res*., **1996**, *24*, 4029-4033.

[9] E. Marsich, A. Bandiera, G. Tell, A. Scaloni, G. Manzini, *Eur. J. Biochem*., **2001**, *268*, 139-148.

[10] C. Sutherland, Y. Cui, H. Mao, L.H. Hurley *J. Am. Chem. Soc*., **2016**, *138*, 14138-14151.

[11] I. Berger, M. Egli, A. Rich, *Proc. Natl. Acad. Sci. U. S. A.,* **1996**, *93*, 12116-12121.

[12] A. L. Lieblein, M. Krämer, A. Dreuw, B. Fürtig, H. Schwalbe, *Angew. Chem. Int. Ed*., **2012**, *51*, 4067-4070.

[13] A. Dembska, *Anal. Chim. Acta*, **2016**, *930*,1-12.

[14] Y. Dong, Z. Yang, D. Liu, *Acc. Chem. Res*., **2014**, *47*, 1853-1860.

[15] J.J. Alba, A. Sadurní, R. Gargallo, *Critical Rev. Anal. Chem*., **2016**, *46*, 443-454.

[16] N.M. Bell, J. Micklefield, *ChemBioChem*, **2009**, *10*, 2691-2703.

[17] S. Benabou, A. Aviñó, R. Eritja, C. González, R. Gargallo, *RSC Adv*., **2014**, *4*, 26956-26980.

[18] H. A. Day, P. Pavlou, Z. A. E. Waller, *Bioorg. Med. Chem*., **2014**, *22*, 4407-4418.

[19] Y. P. Bhavsar-Jog, E. Van Dornshuld, T. A. Brooks, G. S. Tschumper, R. M. Wadkins, *Biochemistry*, **2014**, *53*, 1586-1594.

[20] B. Xu, G. Devi, F. Shao, *Org. Biomol. Chem*., **2015**, *13*, 5646-5651.

[21] B. Yang, M. T. Rodgers, *J. Am. Chem. Soc*., **2014**, *136*, 282-290.

[22] N. Kumar, M. Petersen, S. Maiti, *Chem. Commun*., **2009**, 1532-1534.

[23] N. Kumar, J. T. Nielsen, S. Maiti, M. Petersen, *Angew. Chem. Int. Ed*., **2007**, *46*, 9220-9222.

[24] A Pasternak, J. Wengel, *Bioorg. Med. Chem. Lett*., **2011**, *21*, 752-755.

[25] S. Pérez-Rentero, R. Gargallo, C. González, R. Eritja. *RSC Adv*., **2015**, *5*, 63278-63281.

[26] N. K. Sharma, K. N. Ganesh. *Chem. Commun*., **2005**, 4330-4332.

[27] D. Collin, K. Gehring, *J. Am. Chem. Soc*., **1998**, *120*, 4069-4072.

[28] C. P. Fenna, V. J. Wilkinson, J. R. P. Arnold, R. Cosstick, J. Fisher, *Chem. Commun*., **2008**, 3567-3569.

[29] S. Robidoux, M. J. Damha, *J. Biomol. Struct. Dyn*., **1997**, *15*, 529–535.

[30] H.Abou Assi, R.W. Harkness, N. Martin-Pintado, C.J. Wilds, R. Campos-Olivas, A.K. Mittermaier, C. González, M. J. Damha. *Nucleic Acids Res*., **2016**, *44*, 4998-5009.

[31] A.M. Iribarren, D.O. Cicero, P.J. Neuner, *Antisense Res. Dev*., **1994**, *4*, 95-98.

[32] N.S. Li, J.A. Piccirilli, *J. Org. Chem*., **2003**, *68*, 6799-6802.

[33] D.O. Cicero, M. Gallo, P.J Neuner, A.M. Iribarren, *Tetrahedron*, **2001**, *57*, 7613-7621.

[34] A. T. Phan, M. Guéron, J.-L. Leroy, *J. Mol. Biol*., **2000**, *299*, 123-144.

[35] A. L. Lieblein, J. Buck, K. Schlepckow, B. Fürtig, H. Schwalbe, *Angew. Chem. Int. Ed*., **2012**, *51*, 250-253.

[36] L. Robaldo, F. Izzo, M. Dellafiore, C. Proietti, P. V. Elizalde, J. M. Montserrat, A. M. Iribarren, *Bioorg. Med. Chem*. **2012**, *20*, 2581-2586.

[37] L. Robaldo, A. Berzal-Herranz, J. M. Montserrat, A. M. Iribarren, *ChemMedChem*, **2014**, *9*, 2172-2177.

[38] L.H. Koole, J.C. Wu, S. Neidle, J. Chattopadhyaya, *J. Am. Chem. Soc*., **1992**, *114*, 2687-2696.

[39] D.O. Cicero, A.M. Iribarren, R. Bazzo. *Appl. Magn. Reson*. **1994**, *7*, 95-105

[40] L.Robaldo, R. Pontiggia, S. DiLella, D.A. Estrín, J.W Engels, A.M. Iribarren, J.M. Montserrat *J. Phys. Chem. B* **2013**, *117*, 57-69.

[41] J.D. Puglisi, I. Tinoco, *Methods Enzymol*. **1989**, *180*, 304-325.

[42] S. Benabou, M. Garavis, S. Lyonnais, R. Eritja, C. González, R. Gargallo, *PhysChemChemPhys* **2016**, *18*, 7997-8004.

[43] G. Manzini, N. Yathindra, L.E. Xodo, *Nucleic Acids Res*. **1994**, *22*, 4634-4640.

[44] M. Piotto, V. Saudek, V. Sklenar, *J. Biomol. NMR*, **1992**, *2*, 661-665.

Supporting Information

Contents:

Name	M_{calc}	M_{found}
VTWT	6504.3	6502.2
VT1 C_{Me}Up	6518.3	6516.4
VT2 C_{Me}Up	6518.3	6515.2
VT2,14 C_{Me}Up	6532.3	6530.3

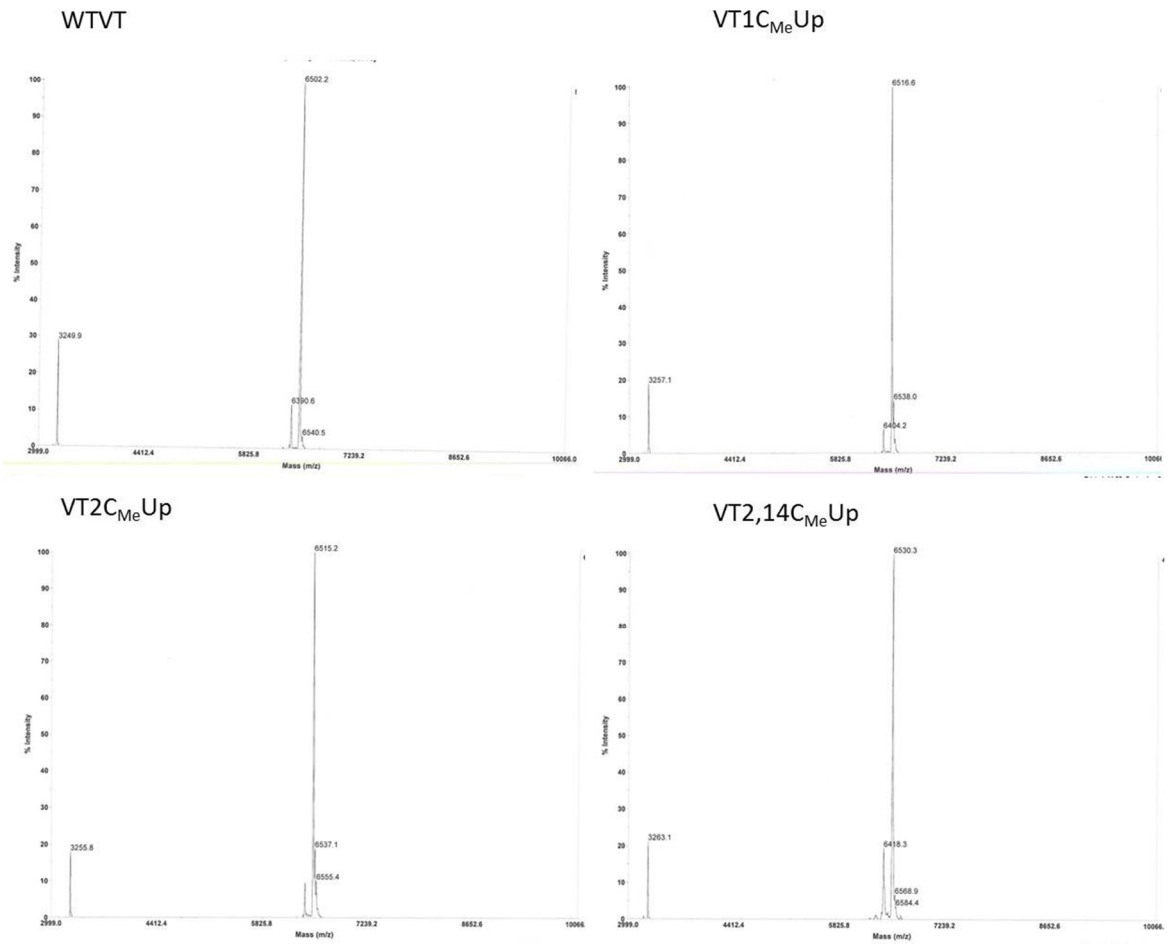


Figure S1. MALDI-TOF spectra of the oligonucleotides employed in this study

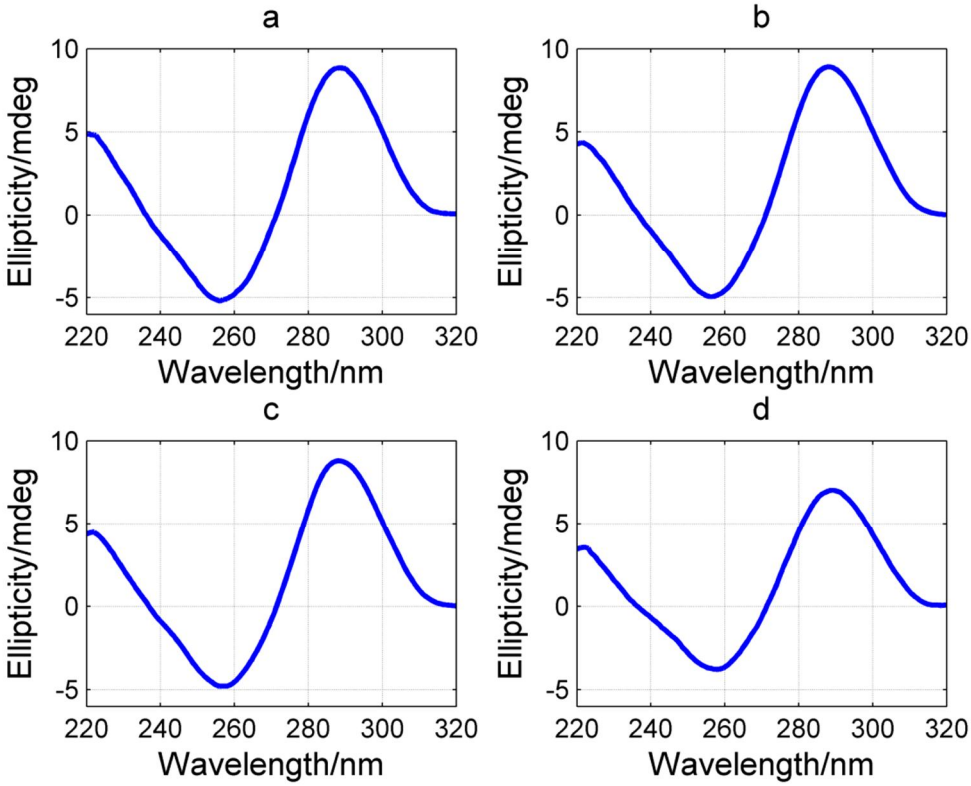
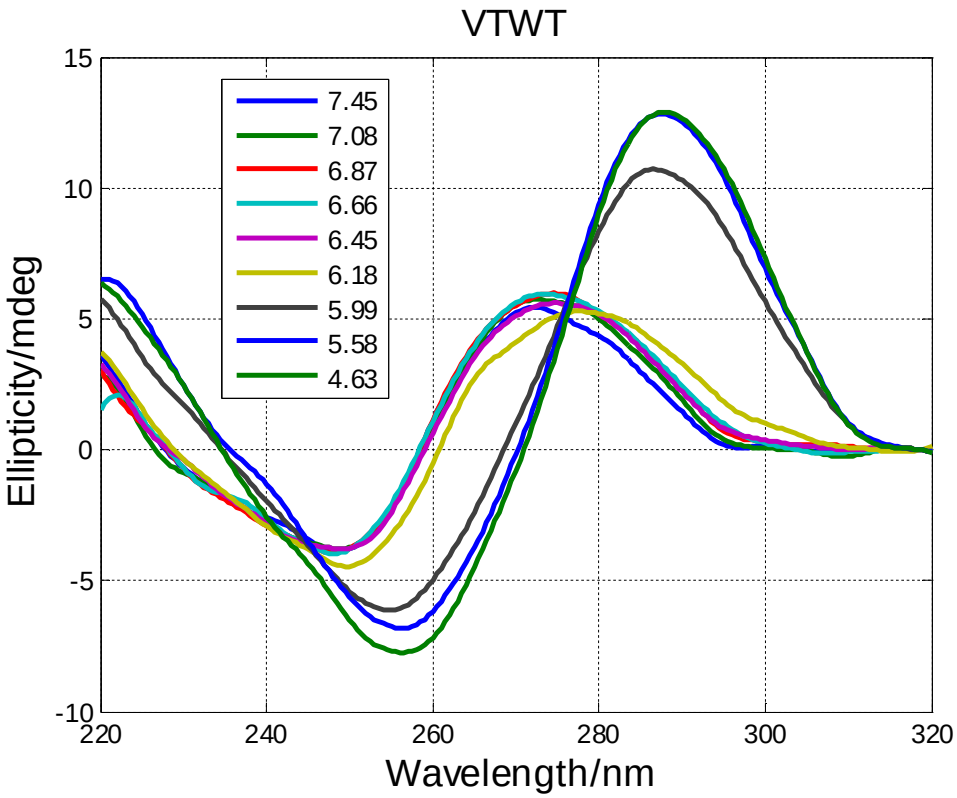
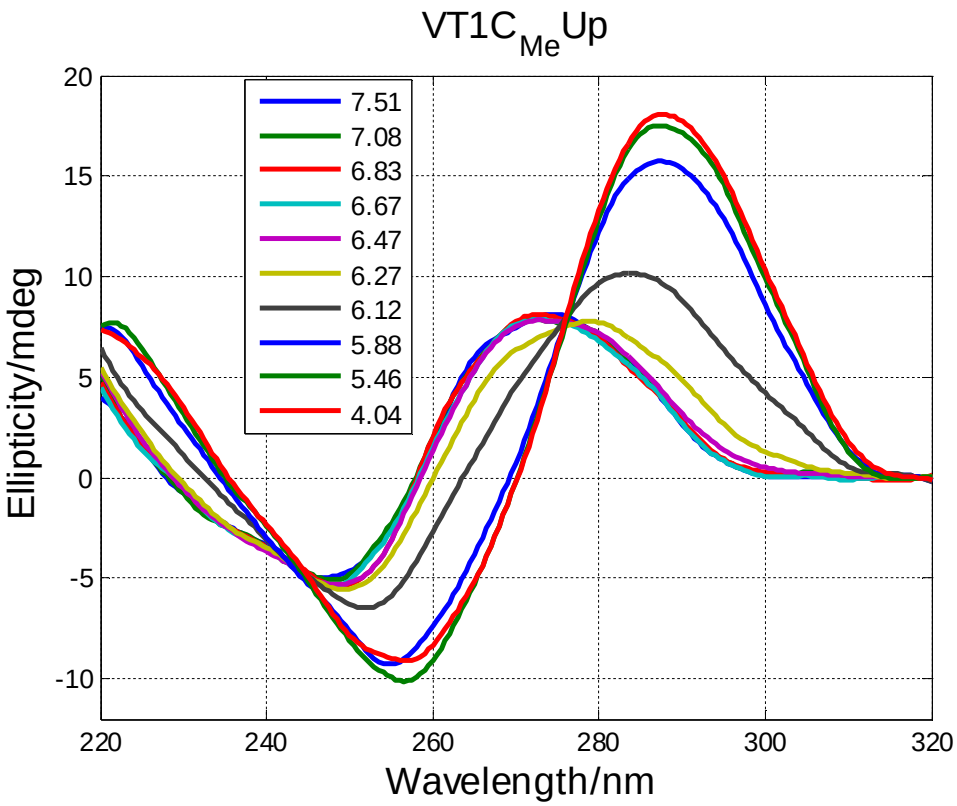


Figure S2. CD spectra of VTWT (a), VT1C_{Me}Up (b), VT2C_{Me}Up (c) and VT2,14C_{Me}Up (d) at pH 5.5 and 15ºC.


VTWT
7.45
7.08
6.87
6.66
6.45
6.18
5.99
5.58
4.63
Ellipticity/mdeg
Wavelength/nm
15
10
5
0
-5
-10
220
240
260
280
300
320


VT1C$_{Me}$Up
7.51
7.08
6.83
6.67
6.47
6.27
6.12
5.88
5.46
4.04
Ellipticity/mdeg
Wavelength/nm
20
15
10
5
0
-5
-10
220
240
260
280
300
320

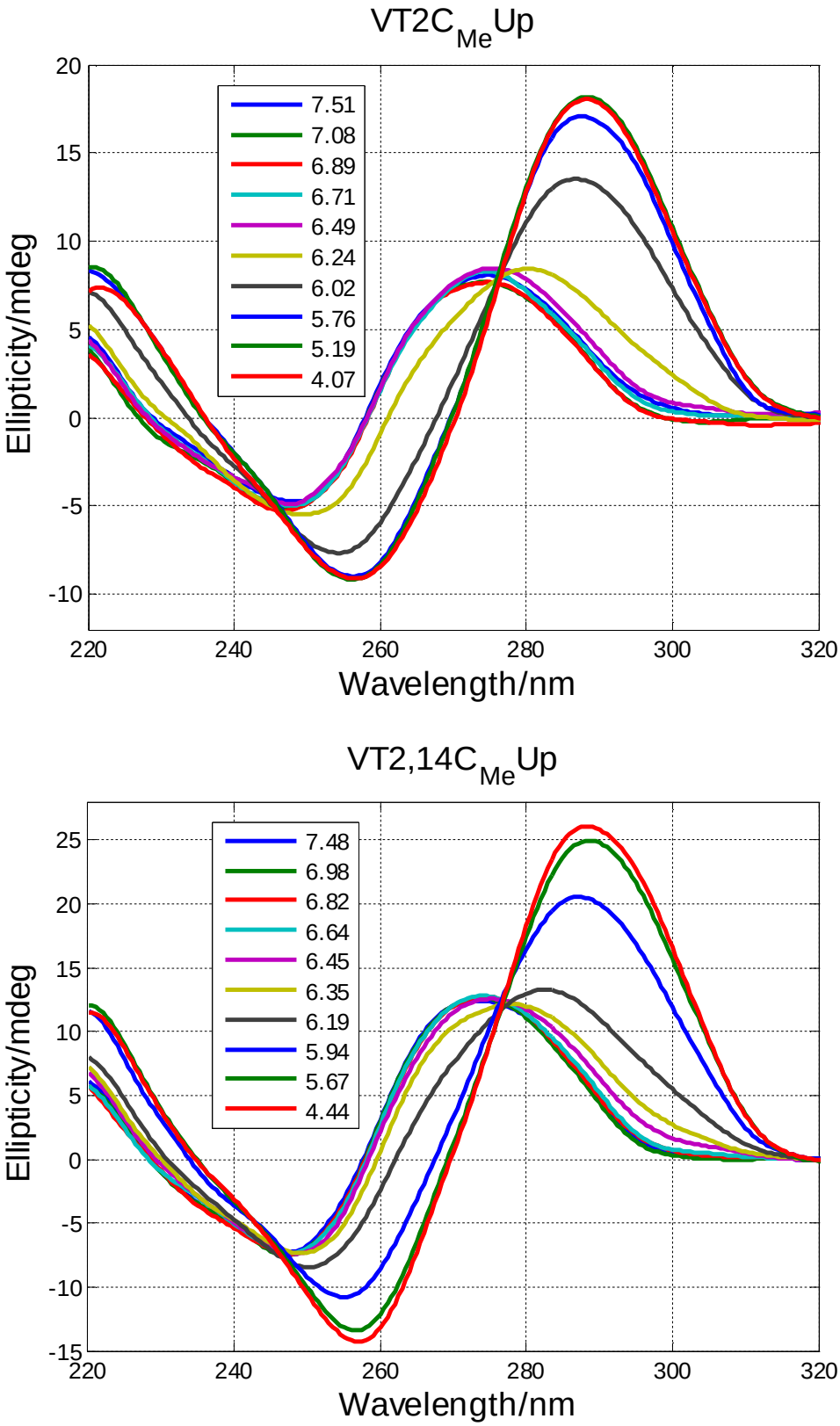
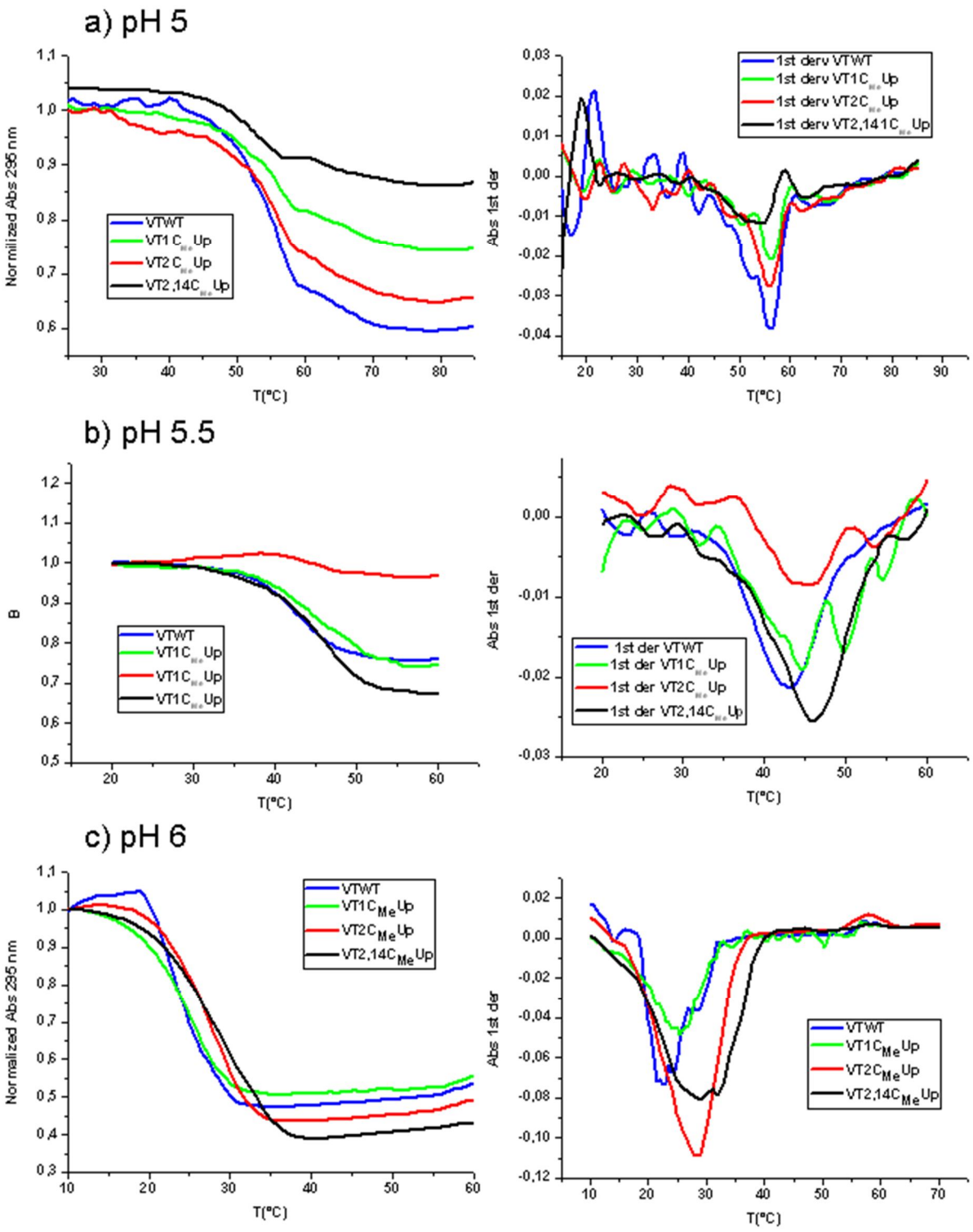


Figure S3. Acid-base titrations followed by CD spectroscopy.


a) pH 5
Normilized Abs 295 nm
T(°C)
VTWT
VT1 C$_{Me}$ Up
VT2 C$_{Me}$ Up
VT2,14C$_{Me}$ Up
Abs 1st der
1st derv VTWT
1st derv VT1C$_{Me}$ Up
1st derv VT2C$_{Me}$ Up
1st derv VT2,141C$_{Me}$ Up
b) pH 5.5
B
VTWT
VT1C$_{Me}$ Up
VT1C$_{Me}$ Up
VT1C$_{Me}$ Up
Abs 1st der
1st der VTWT
1st der VT1C$_{Me}$ Up
1st der VT2C$_{Me}$ Up
1st der VT2,14C$_{Me}$ Up
c) pH 6
Normalized Abs 295 nm
VTWT
VT1C$_{Me}$Up
VT2C$_{Me}$Up
VT2,14C$_{Me}$Up
Abs 1st der

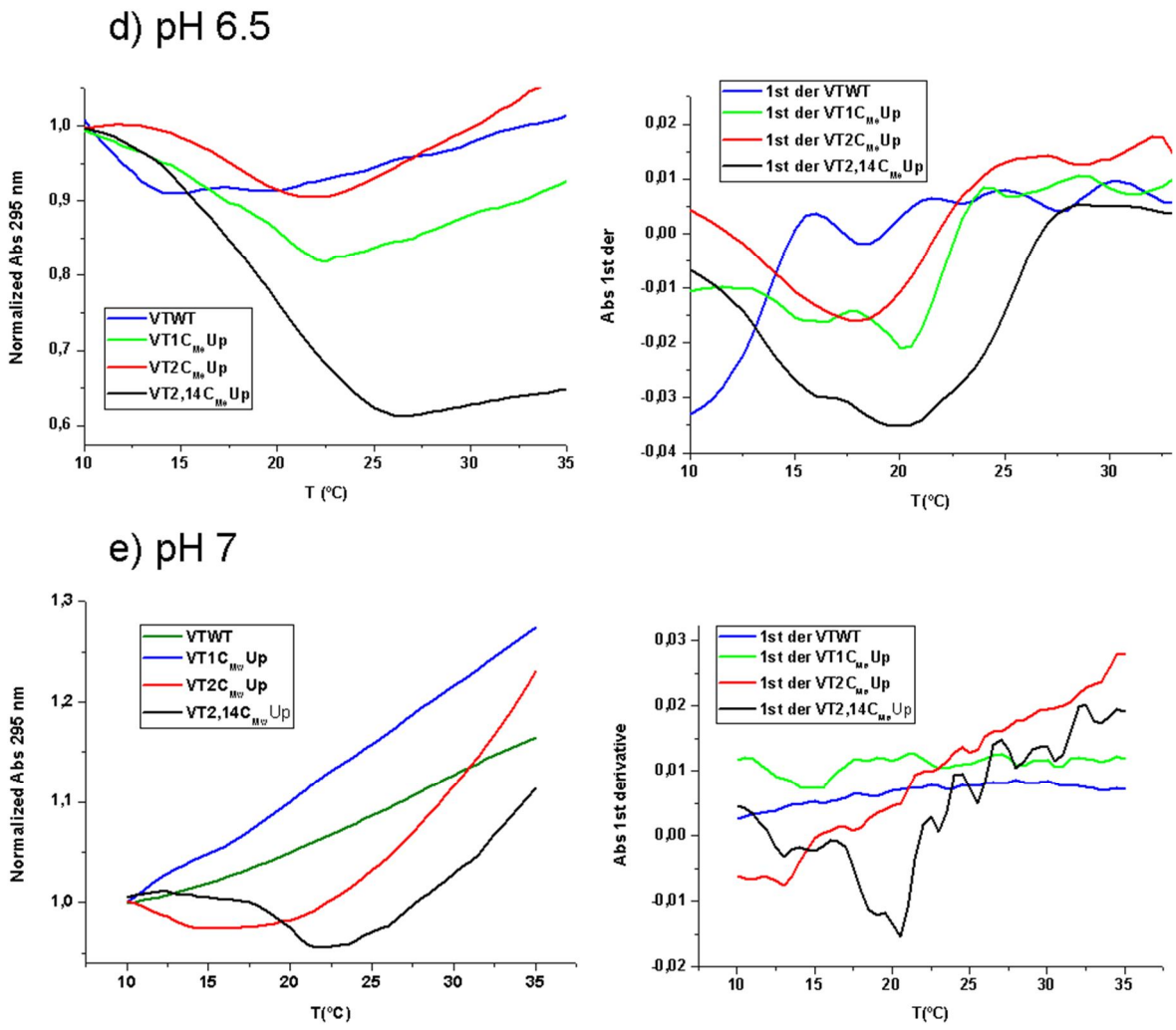


Figure S4. Thermal denaturation curves and the 1st derivative at pH 5.0 (a), pH 5.5 (b), pH 6.0 (c), pH 6.5 (d) and pH 7 (e) determined from UV measurements at 295 nm.

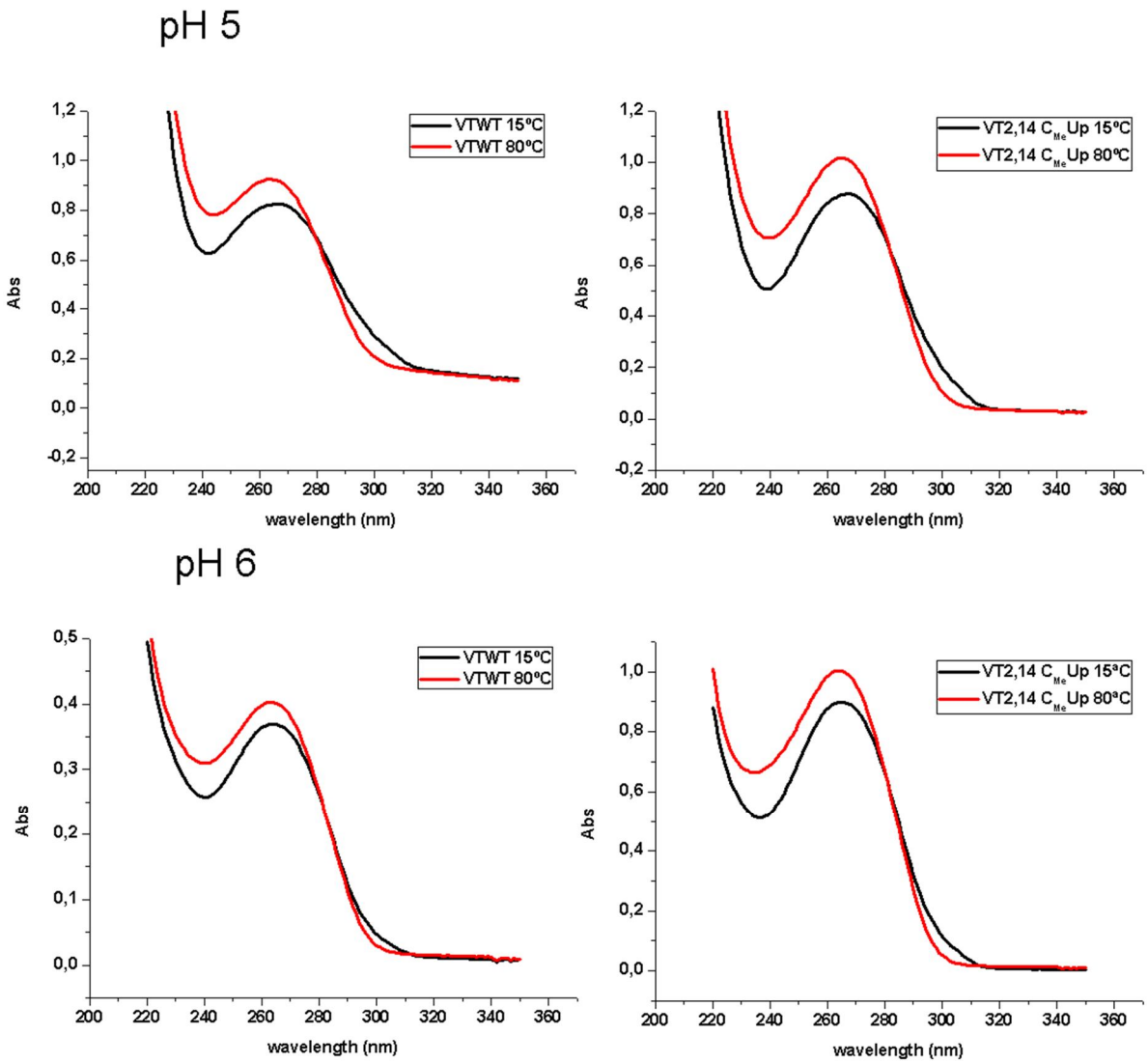


Figure S5. Changes in the UV spectra of VTWT (left) and VT2,14C$_{Me}$Up (right) caused by denaturation at pH 5 and 6. The UV spectra at 15 ºC are represented in red and the UV spectra at 80ºC in black. Upon heating the UV absorbance at 260 nm of the solutions increases but the UV absorbance at 295 nm decreases due to the differences in the absorption between neutral and protonated cytidines.

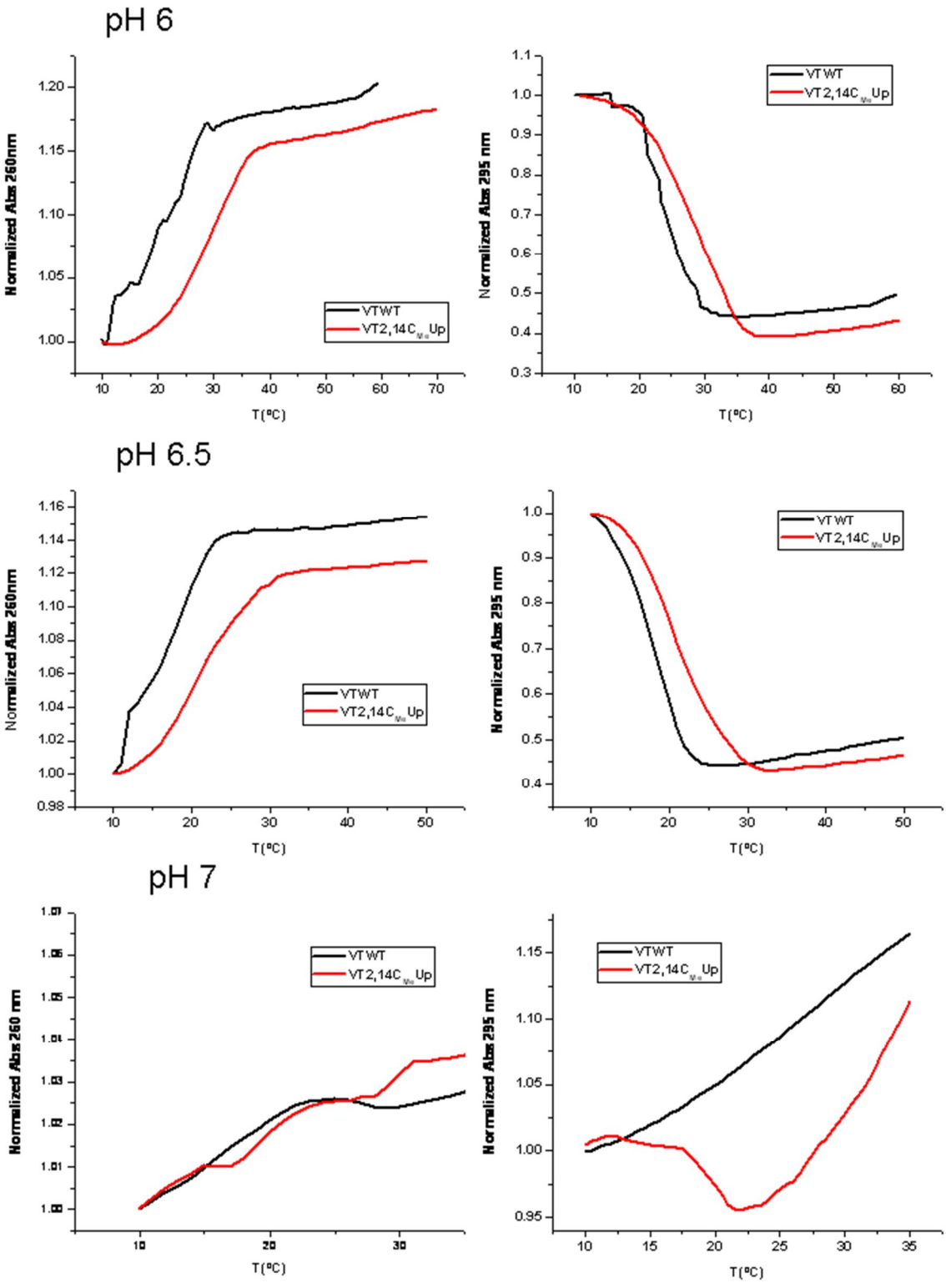


Figure S6. Comparison of UV melting curves followed at 260 nm and 295 nm of VTWT (black) and VT2,14C$_{Me}$Up (red). At pH 6.0 and 6.5 the double modified variant (VT2,14C$_{Me}$Up) is more stable. At pH 7.0 the transition cannot be observed at 260 nm because the hyperchromocity associated to denaturation is very low and cannot be distinguished from the increase of absorbance due to thermal heating. On the contrary at 295 nm it is possible to distinguish the hypochromicity associated to the transition

from the increase of absorbance due to thermal heating, suggesting the presence of the i-motif in the double modified variant.

Thermal denaturation experiments followed by CD
RESULTS

Sequence	pH*	T_m(°C)	ΔT_m(°C)	ΔH^0 (kcal/mol)	ΔS^0 (cal/K·mol)	ΔG^0_{37oC} (kcal/mol)
VTWT	7.0	--	--	--	--	--
VT1C_{Me}Up	7.0	--	--	--	--	--
VT2C_{Me}Up	7.0	--	--	--	--	--
VT2,14C_{Me}Up	7.0	~12	--	--	--	--
VTWT	6.5	--	--	--	--	--
VT1C_{Me}Up	6.5	~15	--	--	--	--
VT2C_{Me}Up	6.5	~20	--	--	--	--
VT2,14C_{Me}Up	6.5	~20	--	--	--	--
VTWT	6.0	26.7	-	-69±8	-231±26	+2.33
VT1C_{Me}Up	6.0	27.0	0.3	-55±7	-184±23	+1.84
VT2C_{Me}Up	6.0	28.1	1.4	-60±7	-200±22	+1.77
VT2,14C_{Me}Up	6.0	30.2	3.5	-47±3	-155±11	+1.05
VTWT	5.5	43.4	-	-70±8	-222±27	-1.42
VT1C_{Me}Up	5.5	43.5	0.1	-76±7	-240±23	-1.57
VT2C_{Me}Up	5.5	44.8	1.4	-67±6	-211±18	-1.65
VT2,14C_{Me}Up	5.5	45.5	2.1	-63±6	-197±20	-1.68

*1.3 µM concentration in 20 mM citrate/phosphate or phosphate buffer and 50 mM KCl

Analysis of thermal denaturation by CD

The thermal denaturing curves were obtained in a single experiment following the change of the CD absorption at 280 nm from 5°C to 80 °C with a linear temperature ramp of 0.5 °C•min^{-1}. In the experiments performed at pHs 5.5 and 6.0, the denaturation curves were complete allowing the extraction of thermodynamic data associated to the transition. In these cases, CD absorption at 280 nm data as a function of temperature were analysed as previously described.[42] 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 chemical equation and the corresponding equilibrium constant may be written as:

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

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

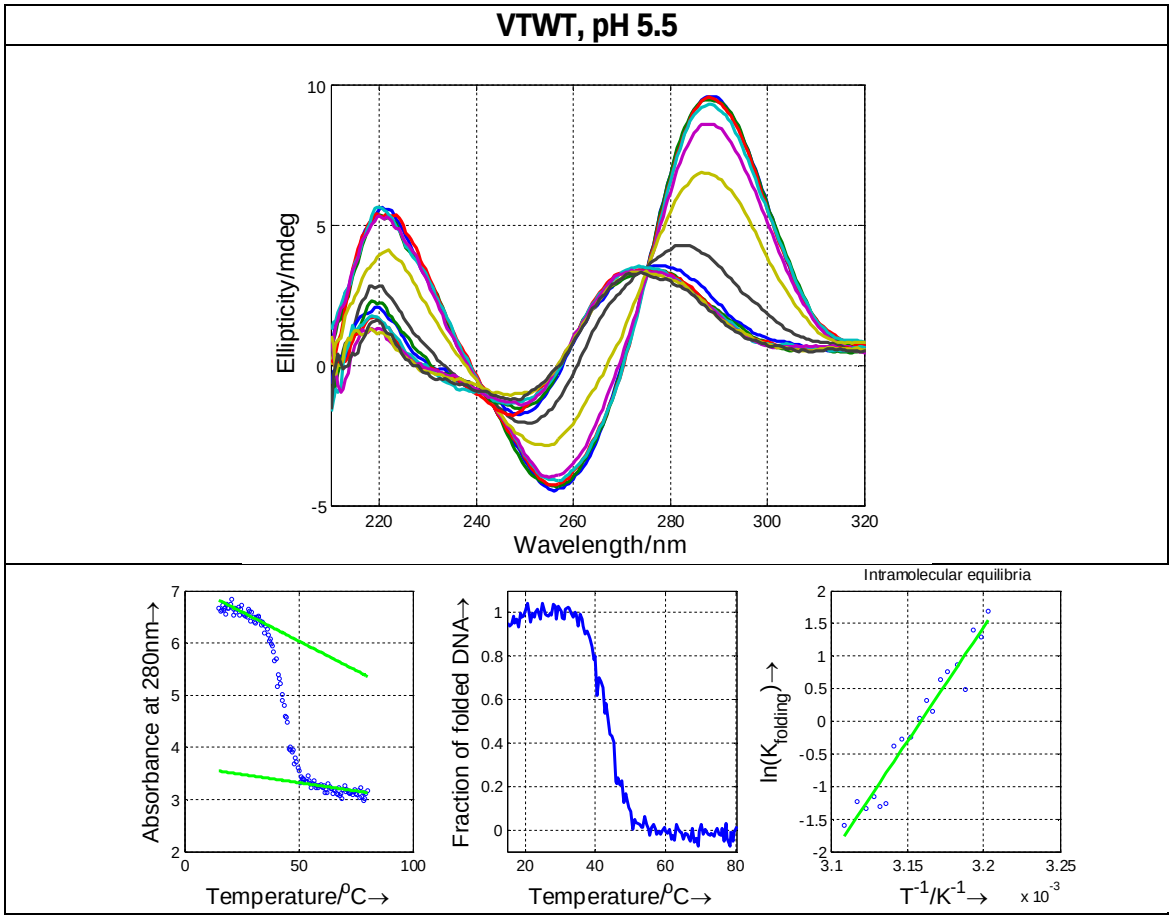
$\ln K_{unfolding} = - \Delta H_{vH} / RT + \Delta S_{vH} / R$

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

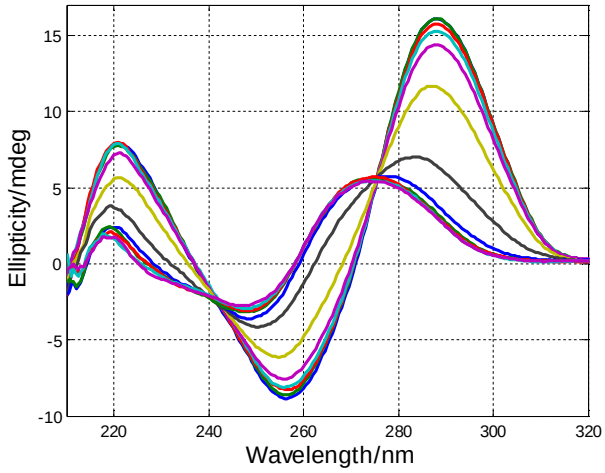
At pHs 6.5 and 7.0 partial denaturation curves were obtained and for this reason, T_m were estimated by analyzing the first derivative of the denaturation curve.

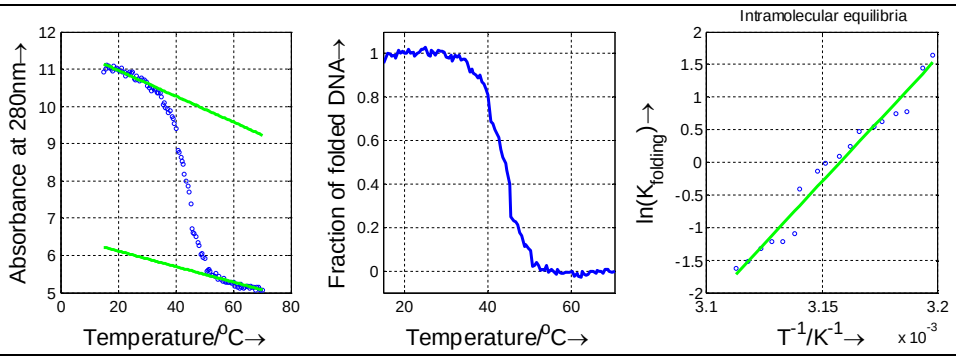
The changes of the CD spectra during denaturation as well as the denaturation curves are shown below.

VTWT, pH 5.5

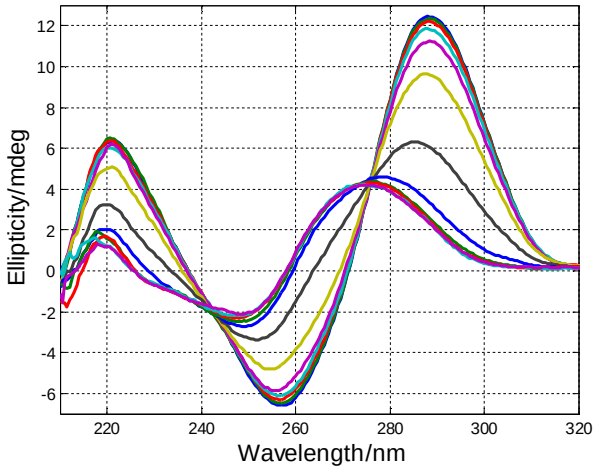


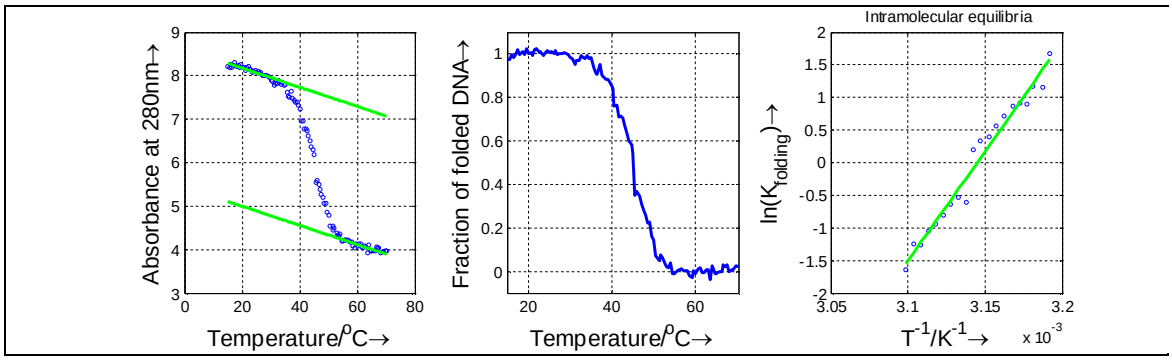
VT1C$_{Me}$Up, pH 5.5



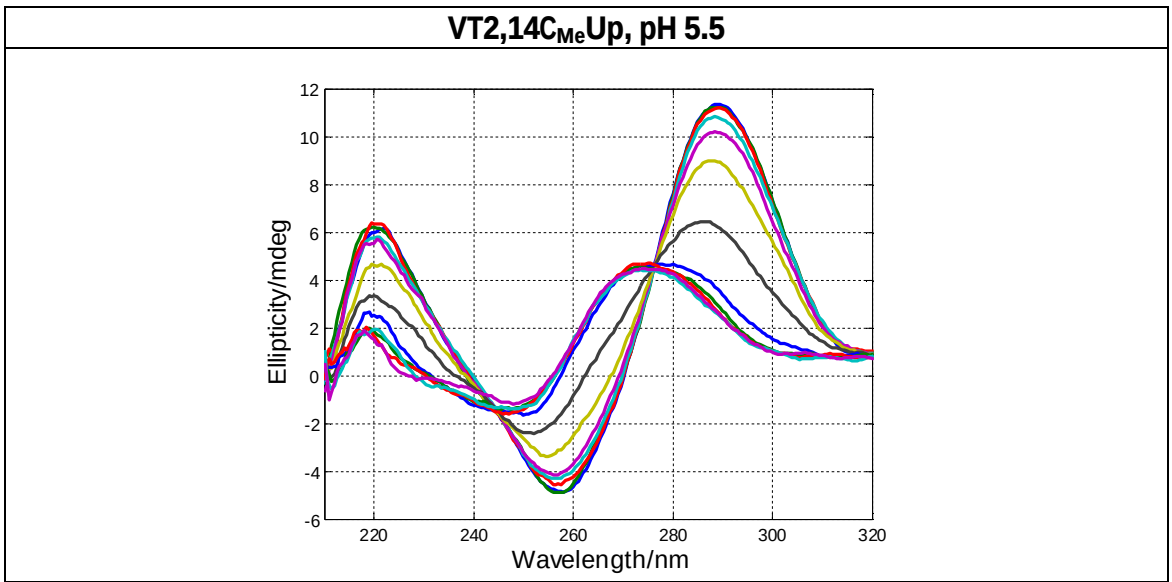


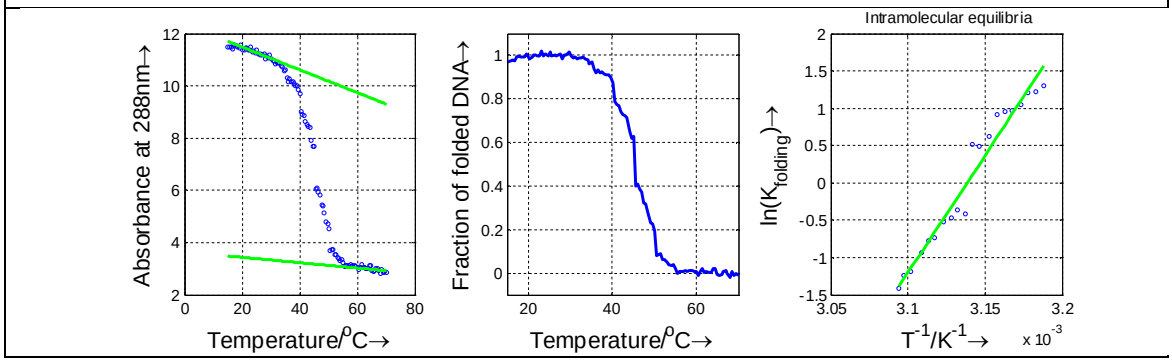
VT2C$_{Me}$Up, pH 5.5



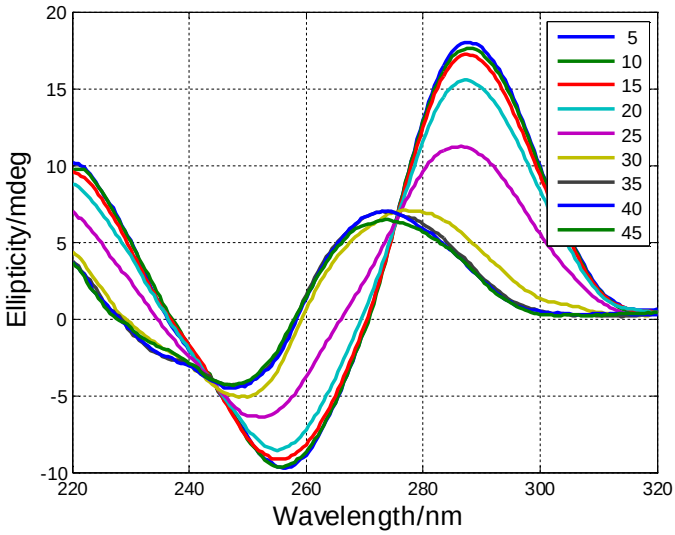


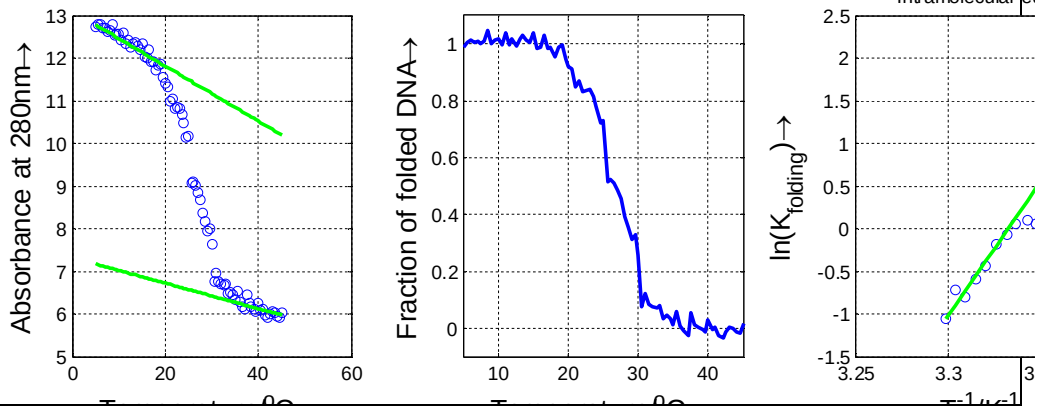
VT2,14C$_{Me}$Up, pH 5.5



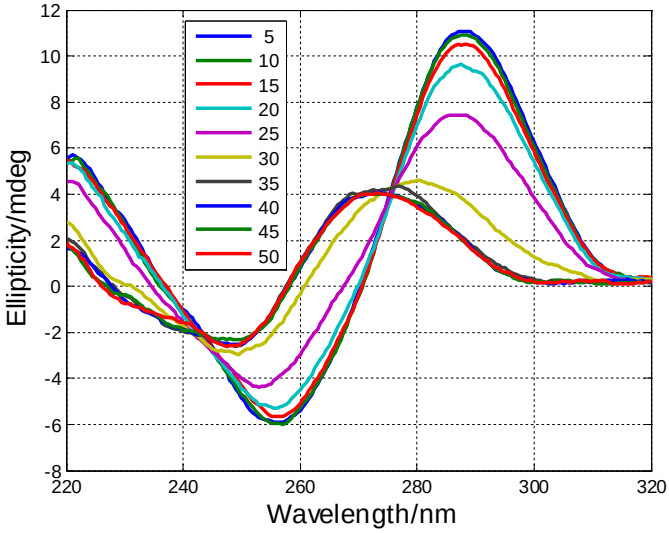


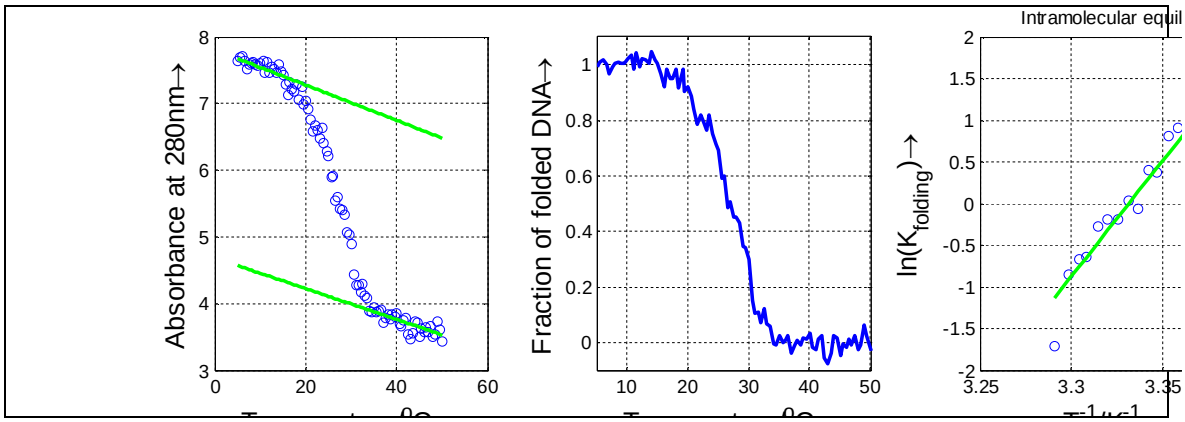
VTWT, pH 6.0

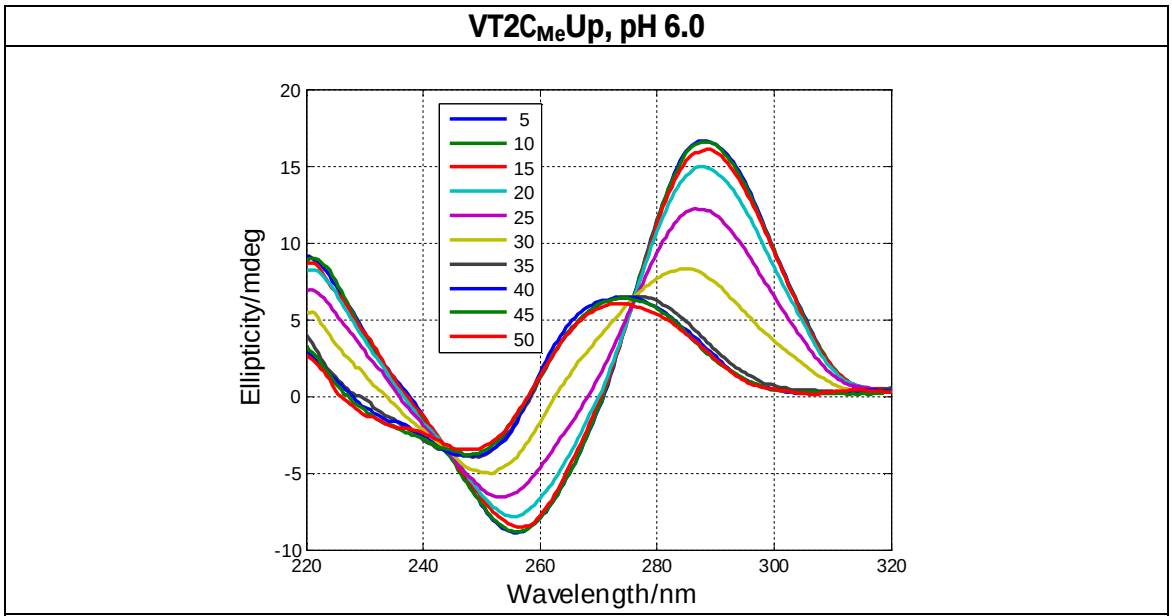


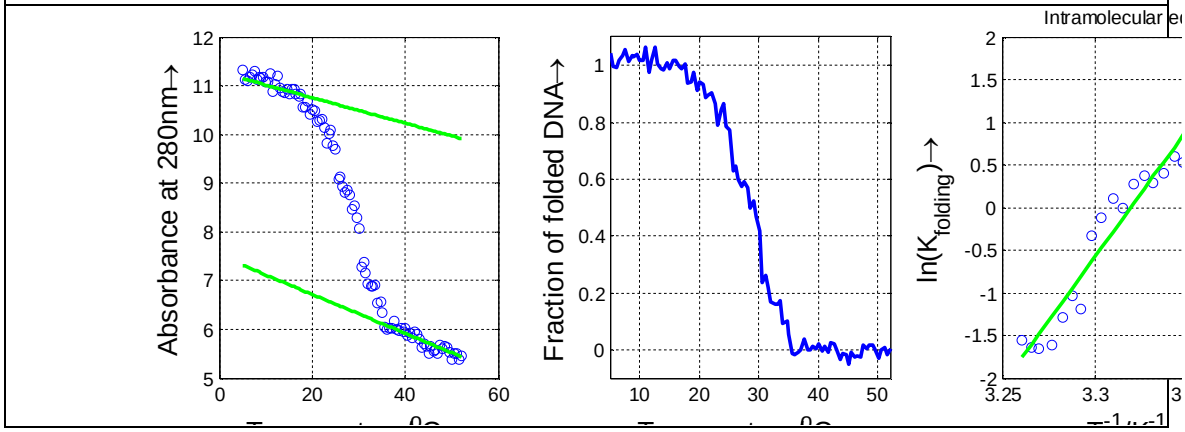


VT1C$_{Me}$Up, pH 6.0

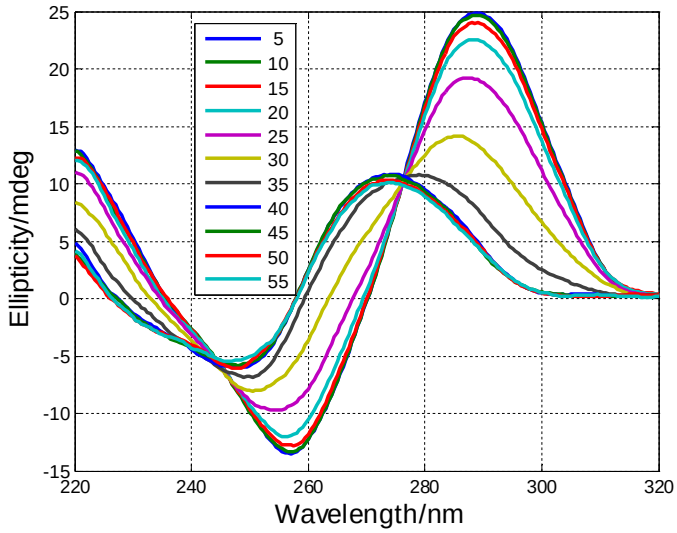


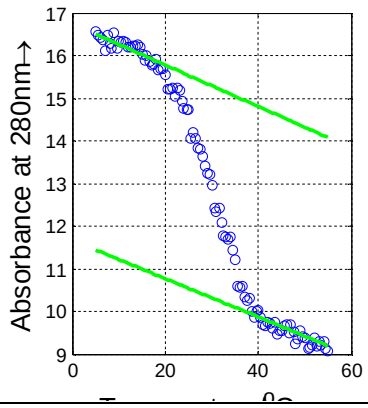

Intramolecular equil
Absorbance at 280nm→
Fraction of folded DNA→
ln(K_folding)→

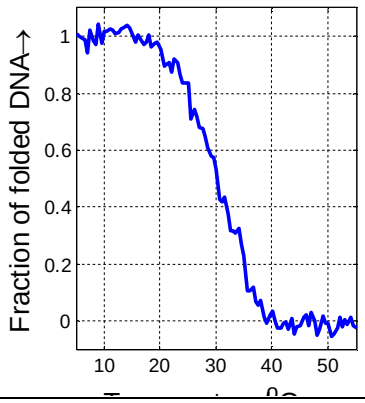

VT2C_MeUp, pH 6.0
Ellipticity/mdeg
Wavelength/nm
5
10
15
20
25
30
35
40
45
50

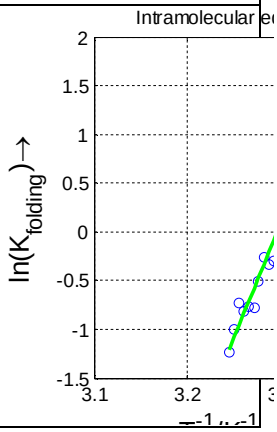

Intramolecular e
Absorbance at 280nm→
Fraction of folded DNA→
ln(K_folding)→

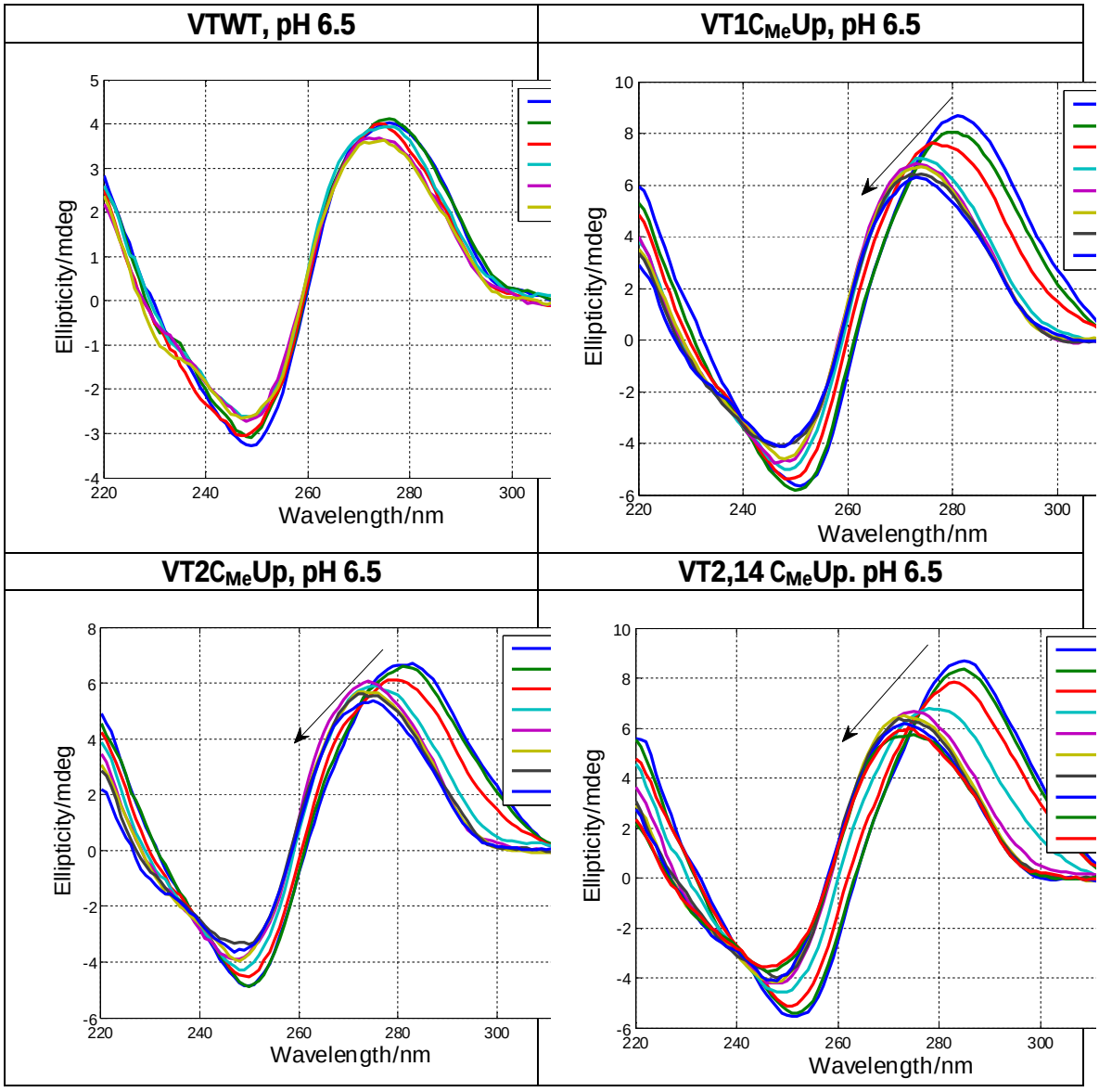
VT2,14C_{Me}Up, pH 6.0



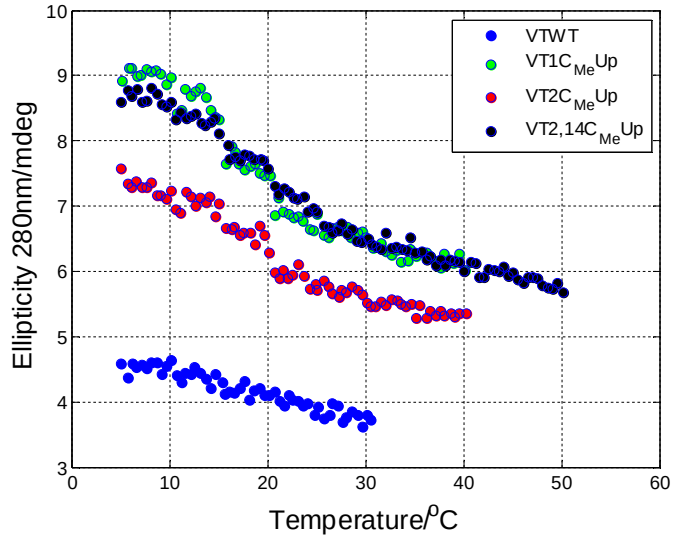


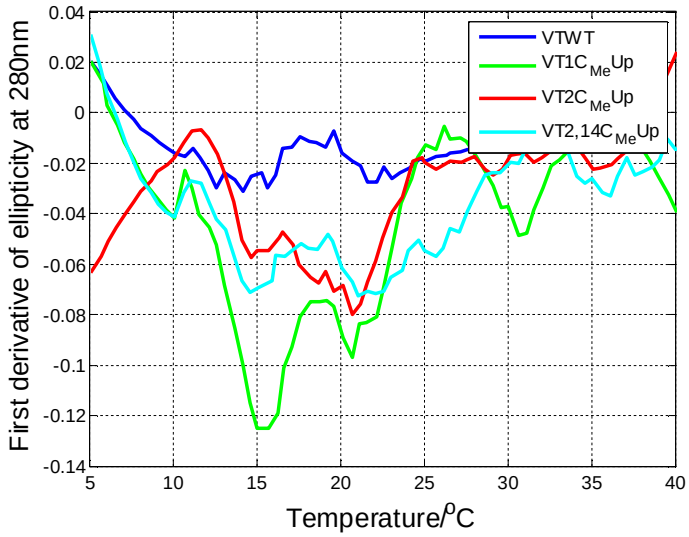




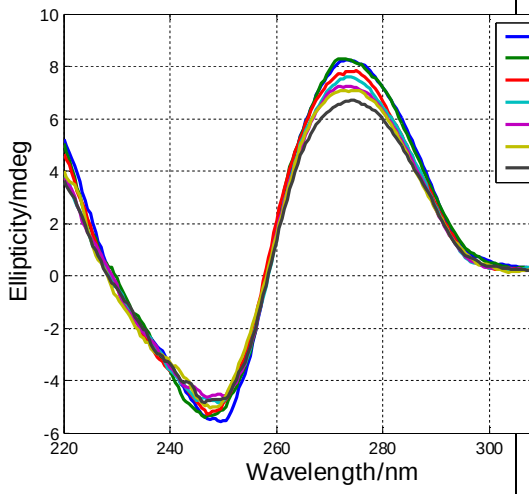


Thermal denaturation curves and first derivatives at pH 6.5 followed by CD at 280 nm

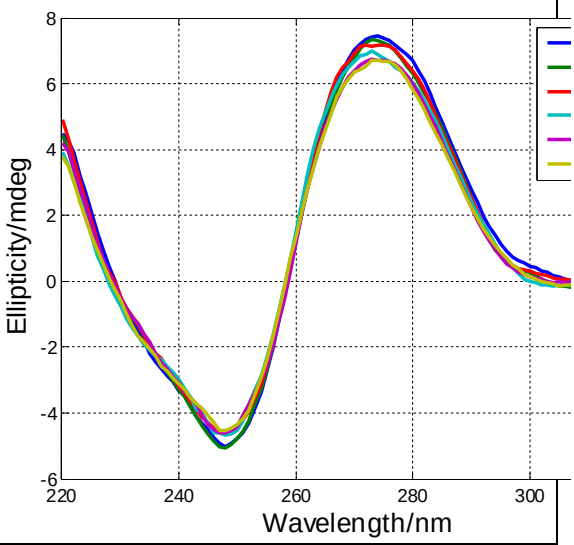

VTWT
VT1C Me Up
VT2C Me Up
VT2,14C Me Up
Ellipticity 280nm/mdeg
Temperature/°C


VTWT
VT1C Me Up
VT2C Me Up
VT2,14C Me Up
First derivative of ellipticity at 280nm
Temperature/°C

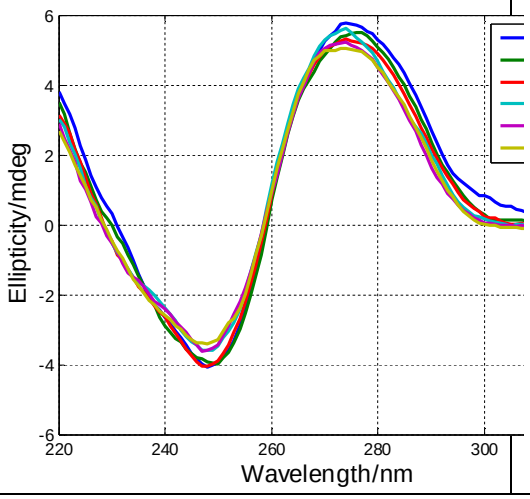
VTWT, pH 7.0



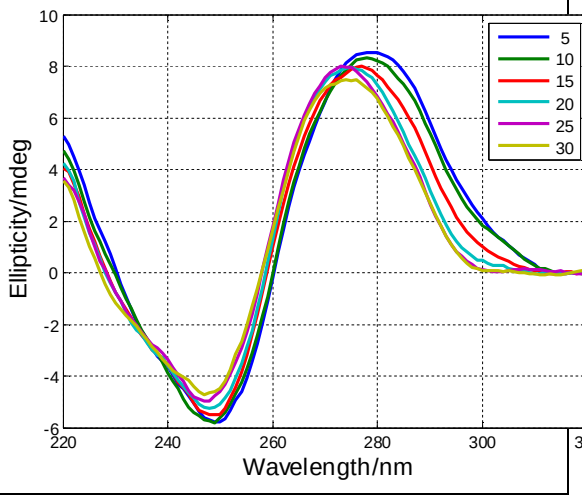
VT1C$_{Me}$Up, pH 7.0



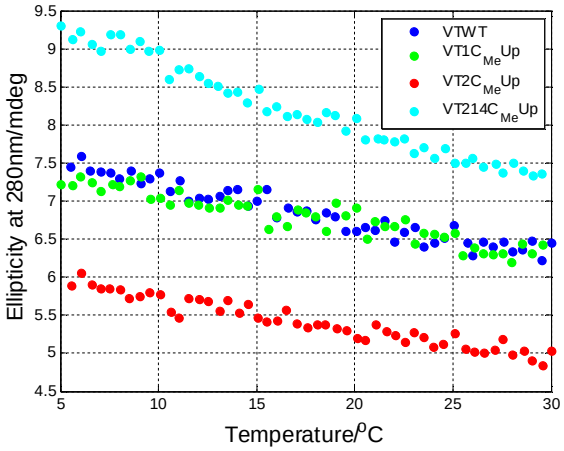
VT2C$_{Me}$Up, pH 7.0

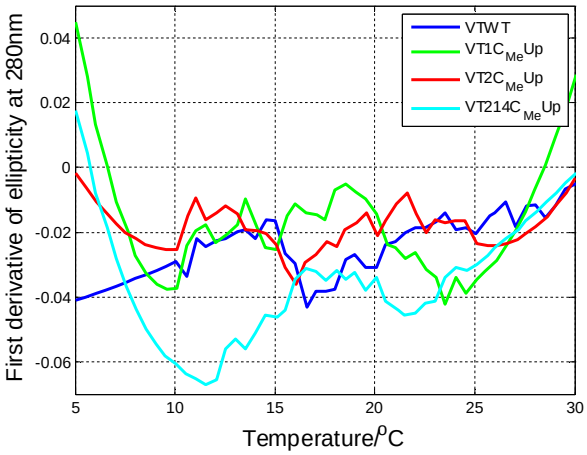


VT2,14 C$_{Me}$Up. pH 7.0



Thermal denaturation curves and first derivatives at pH 7.0 followed by CD at 280 nm





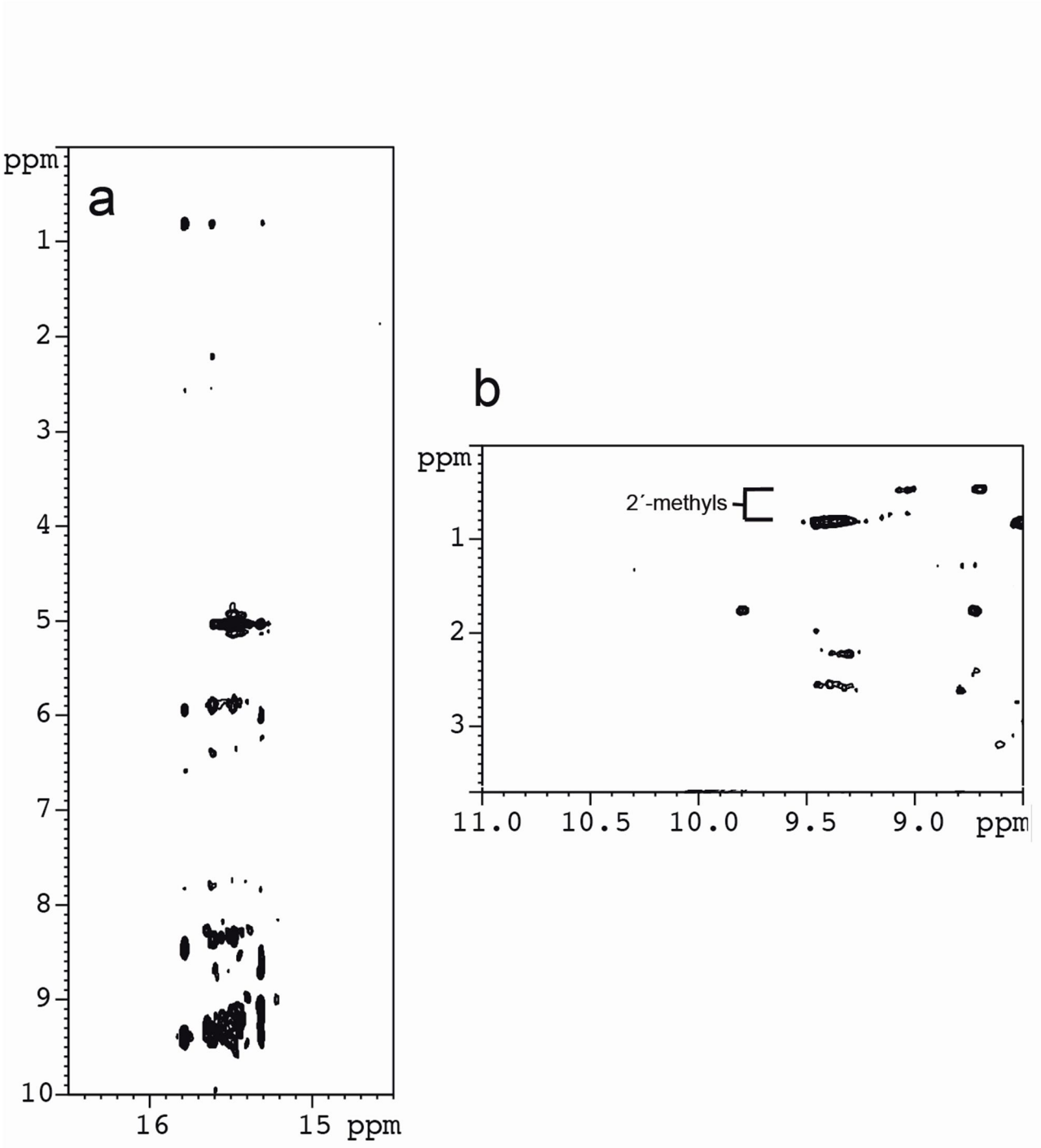


Figure S7. Two regions of the NOESY spectra of VT2,14C_{Me}Up (mixing tine 150 ms, buffer conditions 10 mM sodium phosphate, pH 6, T= 5ºC). Characteristic NOEs between exchangeable protons of hemiprotonated cytosines are shown in panel (a). NOEs between cytosine amino protons and 2'-methyl groups (0.46 and 0.81 ppm) are shown in panel (b).

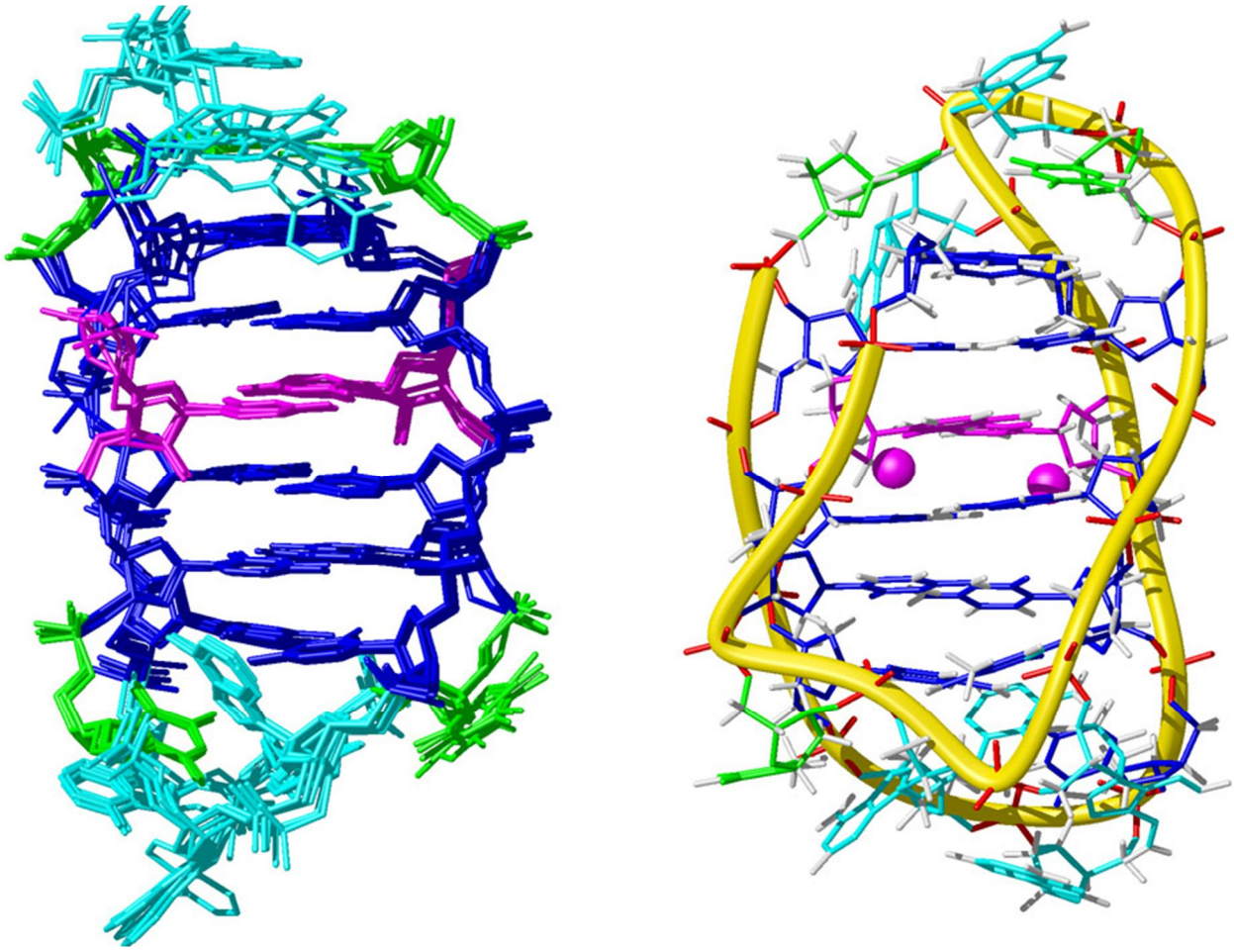


Figure S8. Model of VT2,14C_{Me}Up based on the structure of the C-rich strand of human telomeric sequence. A molecular dynamics calculation was run with the AMBER package, starting from the coordinates of the native sequence (pdb: 1ELN). The (2'*S*)-2'-deoxy-2'-*C*-methyl-cytidine residue, shown in magenta, was built with the program xleap, following standard procedures. Ten snapshots corresponding to a 1 ns molecular dynamics trajectory are shown in the left panel. The final model results from the average and minimization of the last 25 ps of this trajectory. 2'-CH_3 groups are displayed as spheres.