Published online 1 February 2019

Nucleic Acids Research, 2019, Vol. 47, No. 5 ***2177–2189***
doi: 10.1093/nar/gkz046

Systematic investigation of sequence requirements for DNA i-motif formation

Petra Školáková[1], Daniel Renčiuk[1], Jan Palacký[1], Daniel Krafčík[2], Zuzana Dvořáková[1], Iva Kejnovská[1], Klára Bednářová[1] and Michaela Vorlíčková[1,*]

[1]Institute of Biophysics of the Czech Academy of Sciences, v.v.i., Královopolská 135, 612 65 Brno, Czech Republic and [2]CEITEC—Central European Institute of Technology, Masaryk University, Kamenice 735/5, 625 00 Brno, Czech Republic

Received June 20, 2018; Revised January 16, 2019; Editorial Decision January 18, 2019; Accepted January 18, 2019

ABSTRACT

The formation of intercalated motifs (iMs) — secondary DNA structures based on hemiprotonated C.C$^+$ pairs in suitable cytosine-rich DNA sequences, is reflected by typical changes in CD and UV absorption spectra. By means of spectroscopic methods, electrophoresis, chemical modifications and other procedures, we characterized iM formation and stability in sequences with different cytosine block lengths interrupted by various numbers and types of nucleotides. Particular attention was paid to the formation of iMs at pH conditions close to neutral. We identified the optimal conditions and minimal requirements for iM formation in DNA sequences, and addressed gaps and inaccurate data interpretations in existing studies to specify principles of iM formation and modes of their folding.

INTRODUCTION

Guanine-rich (G-rich) and cytosine-rich (C-rich) sequences in complementary strands of DNA are able to form non-standard tetra-stranded DNA arrangements, namely G-quadruplexes and i-motifs (iMs). These potential tetraplex-forming sequences are frequently located in functionally important regions of DNA such as telomeric or centromeric sequences and promoter regions of genes, e.g. (1,2). The search for relationships between tetraplex formation and its functional consequences is a current scientific challenge.

Although G-quadruplexes have been extensively studied for the last two decades (3,4), tetra-stranded structures formed by C-rich DNA regions i.e. i-motifs, have attracted relatively less attention. One of the reasons is the acidic pH needed for the protonation of N3 cytosine and the formation of the hemiprotonated cytosine.cytosine (C.C$^+$) pairs. The pairs join two parallel chains, which are zipped together in an antiparallel orientation (5,6). In contrast to G-quadruplexes, folding a sugar-phosphate DNA chain into the iM is unequivocally determined as only in a chair-type conformation regularly altering the up and down oriented C.C$^+$ pairs can be formed. Individual iMs can differ in folding direction (the first loop spans either a narrow or wide groove) (7) and in the intercalation topology (3′E or 5′E topology, according to which 3′or 5′ cytosine forms the outer pair of the iM core) (8–10). Transition into an intramolecular iM is highly reversible with the possibility of cyclic repetition (11). For this reason, the iM is suitable for use in molecular engineering. Common applications of the iM include the construction of both static (12,13) and dynamic arrangements (14). Highly cooperative formation of the iM is a prerequisite for the use of this structure as a pH sensor (15). pH sensors are also applicable *in vivo* (16).

Common ligands that can modify G-quadruplex formation are not very appropriate for iM stabilization. However, a small molecule of IMC48 stabilizes the iM of the BCL2 gene under certain physiological conditions (17) and similarly graphene quantum dots can selectively stabilize human telomeric iM DNA (18). Moreover, a specific cell environment can stabilize the iM, shifting its formation towards a less acidic pH. These conditions have been simulated using molecular crowding in PEG (200/8000) for d(CCT)$_n$ repetitive sequence (19,20) and in nanocavity water pools of reverse micelle for human telomeric repeat (21). It has also been shown that iM formation can be influenced by the presence of metals (22,23) and specific ligands (24,25). Factors affecting the stability of iM structures have been thoroughly discussed in a recent review (26).

The stability of the iM can be influenced not only by changing external conditions but also by primary structure modifications of DNA. For example, formation of the human telomeric iM can be shifted towards less acidic pH by introducing 5-methylcytidines in the iM core (27). Previous research on loop length revealed that the stability of the iM decreases with the increasing length of loops above three nucleotides, with this effect being the most pronounced in the middle loop (28,29). The most stable was reported (30) as being the iM with only one nucleotide in the loop. Sta-

*To whom correspondence should be addressed. Tel: +420 541 517 188; Email: mifi@ibp.cz

bility of the iM also depends on the type of nucleobases in the loops (26,31) and their precise sequence (32). The most stabilizing arrangement of the first and the third loops was reported to include three thymines or two thymines with a cytosine in the middle (33). Two thymines in the first and third loop can stabilize the iM due to thymine.thymine pairing (34).

It has already been reported (35,36) that, in spite of pK_A 5 of cytosine protonation, iM can be formed in pH regions close to neutral in suitable C-rich sequences. Recently, detailed studies of this effect have shown (30,37) that the elongation of cytosine tracts stabilizes the formation of iM at pH values even higher than neutral. Moreover, the recent paper of Waller's group has demonstrated that sequences which have the capacity to fold into the iM under neutral pH are frequently present in genomes and predominantly in their functionally important regions (37).

Despite recent cellular experiments (38,39) strongly supporting the biological importance of the iM, direct evidence of its existence *in vivo* had been still lacking. Two quite recent studies were published reporting the detection of iM formation in the nuclei of human cells. Using in-cell NMR spectroscopy, Dzatko *et al*. (40) showed that iMs, formed from fragments corresponding to naturally occurring C-rich sequences in the human genome, are stable and able to persist in the nuclei of living human cells. In agreement with this observation, Zeraati *et al*. (41) by means of fluorescently labelled antibody fragments, detected iM formation in the regulatory regions of the human genome in permeabilized cells.

The above-mentioned importance of the iM structure emphasized by its reported existence in cells and presumable biological functions requires further detailed research on conditions of its formation. In spite of numerous thorough studies, there are still questions that remain unanswered. In this paper, we broaden and clarify data on the minimal and optimal number of block cytosines required for iM formation, on the kinetics of iM formation, on optimal loop compositions of intramolecular iMs and on the influence of single nucleotide loops on iM molecularity. Using a variety of methods, we search for the basis of distinctly thermostable iM forms existing in $(C_nT_3)_3C_n$ with seven or more cytosines in blocks, under pH values close to physiological ones.

MATERIALS AND METHODS

Sample preparation

All the oligonucleotides used in this study (Supplementary Table S1) were synthesized and purified by Sigma-Aldrich. The lyophilized oligonucleotides were dissolved in 1 mM sodium phosphate, 0.3 mM EDTA, pH 8, to give a stock solution concentration of ∼10 mM in DNA nucleosides. The oligonucleotides were denatured before further use to remove possible higher structures.

UV absorption and CD spectroscopy

The precise oligonucleotide concentrations were determined based on their absorption measured at 90°C in 1 mM sodium phosphate, 0.3 mM EDTA, pH 8 on a UNICAM 5625 UV/Vis spectrometer using the calculated (42) molar extinction coefficients at 260 nm (Supplementary Table S1).

All experiments were performed at 23°C, in 0.5× Britton-Robinson buffer (20 mM H_3BO_3, 20 mM H_3PO_4, 20 mM CH_3COOH, 37 mM KOH (K-RB)) giving pH 8. The pH values were adjusted by adding 2 M HCl directly to the cells and measured using a Sentron Titan pH meter with Sentron Red-Line Probe electrode or Mettler Toledo pH meter with InLab micro pH electrode. Samples were left to equilibrate and all measurements were taken the following day without thermal annealing. Annealing (90°C for 5′ followed by slow cooling), which was also pursued for a control, led to lowering $\Delta\varepsilon$, ε and T_m values and induced formation of higher-order structures at acidic pH values (Supplementary Figure S1A, B).

CD spectra were measured using a Jasco J815 spectrometer (Japan) in 1cm (unless stated otherwise) path length Hellma cells in a Peltier holder. Strand concentrations of oligonucleotides were 2–4 μM (∼0.07 mM in nucleoside residues). Circular dichroism was expressed as $\Delta\varepsilon = \varepsilon_L - \varepsilon_R$ in units of $[M^{-1}.cm^{-1}]$ (the molarity being related to DNA strand concentration, unless stated otherwise). The pH titration curves were measured by changes of $\Delta\varepsilon$ at 287 nm (CD maximum for iM formation). The experimental data were fitted by three-parameter sigmoidal function using SigmaPlot 10 software (Systat Software Inc., USA). The pK_A values are inflection points of the fits. We determined the standard error of pK_A values for selected control oligonucleotides from three independent pH dependencies as 0.05 pH value.

Kinetics of iM formation was measured using a Chirascan Plus dichrograph equipped with a Stopped-Flow accessory (Applied Photophysics, UK) and a Xe/Hg lamp. The device was set to 297 nm, 2 nm bandwidth and 1 cm optical path length. Both circular dichroism and absorbance were recorded. For mixing reaction, the oligonucleotide was prepared in 1 mM sodium phosphate, 0.3 mM EDTA buffer, pH 8 and mixed at 1:1 volume ratio with the 1× K-RB buffer of certain pH (pH 5 and pH 7.1) to give desired pH after mixing. Final oligonucleotide concentration was set to 3 μM in strands. Each mixing reaction was repeated eight times and each repeat consisted of 10 000 experimental points (1–300 s). The average of selected experimental traces, expressed as $\Delta\varepsilon$, was fitted by a single-exponential rise to maximum according to the function $f(x) = y_0 + a*(1 - \exp(-b*x)) + c*(1 - \exp(-d*x))$ using SigmaPlot software.

UV absorption melting curves

Prior to melting, all samples were left overnight at 23°C to achieve equilibrium. Melting curves (5–85, 85–5 and again 5–85°C) were measured using a Varian Cary 4000 spectrophotometer (Australia) in 1 cm cells. The temperature was increased/decreased in 1°C steps and the samples were equilibrated for 2 min before each measurement to give an average temperature change around $0.25°C.min^{-1}$. Melting curves were monitored by absorbance at 297 nm and expressed as 0–1 normalized data. The data were normalized using a dual baseline-corrected melting curve according to Mergny (43). Accuracy of T_m determination was ±0.5°C.

Native electrophoresis

Non-denaturing polyacrylamide gel electrophoreses were performed in a thermostatted submersible apparatus (SE-600; Hoefer Scientific, San Francisco, CA, USA) with buffer circulation. Gels (16%, 29:1 acrylamide:bisacrylamide), were run for 19 h at 30 V (2 V/cm) and 23°C in 0.5× K-RB, pH 5.0 (pH 6.5 or pH 7.5). 2 μg of the respective oligonucleotides in the same solvent were loaded on the gel. As a marker, we used GeneRuler Ultra Low Range DNA ladder (bands refer to the number of bases).

The gels were stained with Stains-All (Sigma-Aldrich) and digitalized using Personal Densitometer SI 375-A (Molecular Dynamics, Sunnyvale, CA, USA).

Bromine probing of unpaired cytosines

All samples were 5′-labeled by ^{32}P-γ-ATP using T4-polynucleotide kinase. Labeled samples were diluted into 0.5× K-RB buffer of desired pH. Reactions of unpaired cytosines with bromine were carried out in accordance with published procedure (44) modified according to Guo *et al.* (45). In brief, the samples (60 μl) were reacted with molecular Br_2 (generated by mixing an equal molar amount of KBr with $KHSO_5$). The reactions were terminated after 10 min for pH 5 or 20 min for pH 8 by adding 20 μl of stop buffer containing 1.2 M sodium acetate buffer (pH 5.3), 7.5 mM HEPES and 20 μg glycogen. Unreacted Br_2 was removed by ethanol precipitation and the pellets were rinsed with 70% ethanol. The modified oligonucleotides were backbone cleaved at sites of modification by heating (30 min) in 1 M piperidine at 90°C.

After extensive lyophilization, the products were denatured with loading dye and separated on a denaturing (7 M urea) 20% (29:1 acrylamide:bisacrylamide) sequencing gel. Electrophoreses were run about 90 min at 2.6 kV (45 W) on a S2 Model (BRL, Life Technologies) sequencing instrument. The gels were visualized using Phosphor Storage Screen (Amersham Biosciences) and digitalized by a Typhoon FLA9000 device. The densitograms were calculated using Image Quant 5 software.

Data on other methods used in the experiments shown in Supplements are given along with figure legends.

RESULTS AND DISCUSSION

iM formation is sensitively reflected in the spectra of the electronic circular dichroism and in UV absorption spectra (Figure 1A). As shown for $(C_3T_3)_3C_3$ (abbreviated C_3T_3, Supplementary Table S1, Figure 1A), single-stranded unordered C-rich DNA fragments yield a CD spectrum with a positive maximum at 275 nm. With decreasing pH below the neutral value, this positive maximum increases and shifts toward the red and, simultaneously, the CD region around 260 nm deepens, resulting in the formation of a negative band. CD spectrum with a high positive maximum at 287 nm and a negative one at 265 nm is characteristic of the iM (Figure 1A). In contrast to distinctly different spectra of G-quadruplexes (46), which can adopt various structures, CD spectra of iM sequences are principally of the same type due to their single folding topology. iM formation is a cooperative two-state process as revealed by the sigmoidal shape of the dependence of the $\Delta\varepsilon_{287}$ as a function of pH (25), and by the presence of isodichroic points at 240 and 278 nm in the CD spectra. Cytosine protonation is reflected by an increase in the UV absorption spectrum around 300 nm, whereas the formation of the ordered iM is associated with a decrease of the absorption maximum (at 267.5 nm) and its bathochromic shift (Figure 1A, insert). Cytosine protonation that does not lead to the formation of any ordered structure, as shown by CD for $(CT_3)_3C$ (abbreviated CT_3, Supplementary Table S1, Figure 1B), starts at pH values more close to the pK_A of cytosine and leads to an overall increase in absorption spectrum at long wavelengths, including the maximum ∼267 nm (Figure 1B, insert). Thus the absorption maximum at 267 nm increases, while it decreases when the ordered iM structure is formed (Figure 1C).

Table 1. T_m/T_{ren} at pH 5 and pK_A values of studied sequences

Oligonucleotide	Abbreviation	pK_A	T_m/T_{ren} [°C]	ΔT_m [°C]
$(C_2T_3)_3C_2$	C_2T_3	5.99	36.5	
$C_3T_3C_2T_3C_3T_3C_2$	$C_3T_3C_2$		52.0	15.5
$(C_3T_3)_3C_3$	C_3T_3	6.76	61.0/60.6	9.0
$C_4T_3C_3T_3C_4T_3C_3$	$C_4T_3C_3$		68.0	7.0
$(C_4T_3)_3C_4$	C_4T_3	6.95	71.7/70.2	3.7
$C_5T_3C_4T_3C_5T_3C_4$	$C_5T_3C_4$		75.5	3.8
$(C_5T_3)_3C_5$	C_5T_3	7.25	79.6/77.2	4.1
$(C_6T_3)_3C_6$	C_6T_3	7.48	83.3/81.5	3.7

iM formation of C_nT_3

The study commenced with the sequences contained in Table 1, which differ in the number of cytosines in blocks interrupted by three thymines.

Figure 2A shows the pH-induced formation of iM of sequences C_nT_3 indicated by changes in $\Delta\varepsilon_{287}$. C_3T_3 transforms to iM with the midpoint (pK_A) around pH 6.8. Oligonucleotides with longer C blocks adopt iM at pH more close to the neutral value, and sequences C_5T_3 and C_6T_3 are transformed to iM even above pH 7 (Figure 2A). The transition midpoint does not further shift for $n > 6$. The transition curves become more steep with increasing n, i.e. more cooperative (steepness of the curve is related to the degree of cooperativity (25)). Δε values of the final C_nT_3 iMs increase with Cn length, and, simultaneously, the structures become more thermostable (Table 1). An increase in T_m with increasing number of cytosines in blocks is quite uniform for sequences starting from $C_4T_3C_3$ (expected 7 $C.C^+$ pairs) to C_6T_3 (12 pairs) with ∼3.7°C per additional $C.C^+$ pair (Table 1). For C_3T_3 and shorter sequences, the drop in T_m increases with decreasing number of cytosines; the relative effect of loops and ends starts to prevail and the shortest sequence able to form iM at reasonable pH value is C_2T_3.

Kinetics of iMs formation of C_nT_3. All the iMs of C_nT_3 are intramolecular (Supplementary Figure S1B). Interestingly, while the iM of C_3T_3 appears with fast kinetics, the complete formation of iMs of C_nT_3 with $n > 3$ is slow. A stopped flow CD measurement shown in Figure 3 demonstrates that the iM of C_3T_3 is completely formed within the

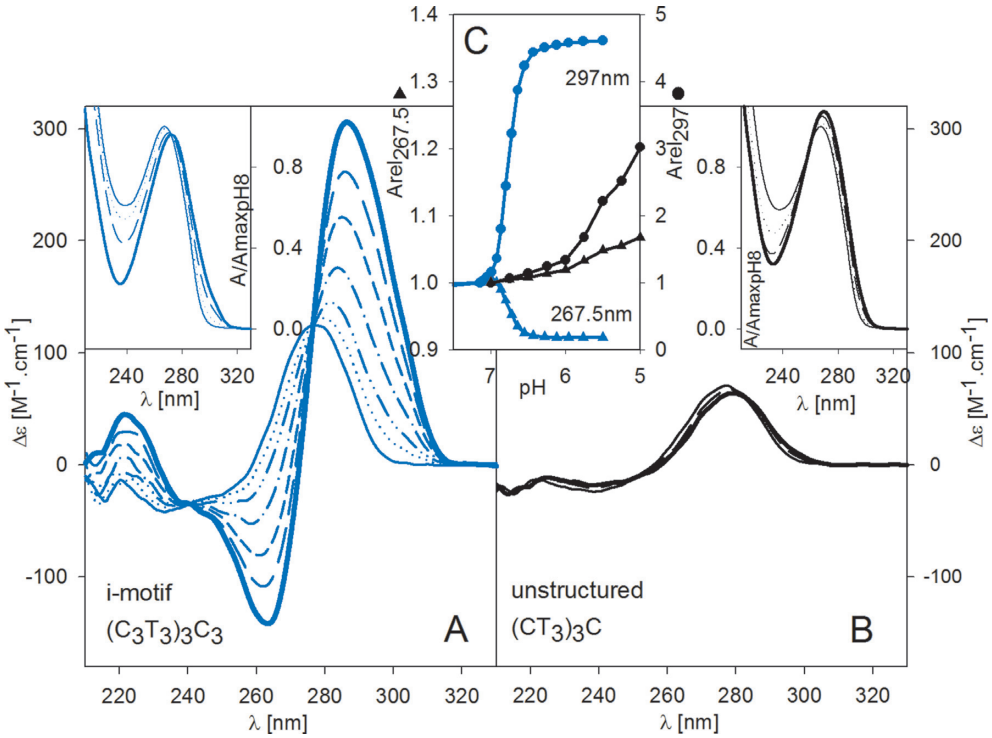

Figure 1. CD and normalized UV absorption spectra of (**A**) $(C_3T_3)_3C_3$ (blue) and (**B**) $(CT_3)_3C$ (black) measured at pH decreasing from pH 8 (thin solid line) to pH 5 (thick solid line). (**C**) Relative absorption (A_{pHi}/A_{pH8}) at 267.5 (triangles, left y-axis) and 297 nm (circles, right y-axis) of $(C_3T_3)_3C_3$ (blue) and $(CT_3)_3C$ (black) plotted as a function of pH. All measurements were carried out in 0.5× K-RB buffer at 23°C.

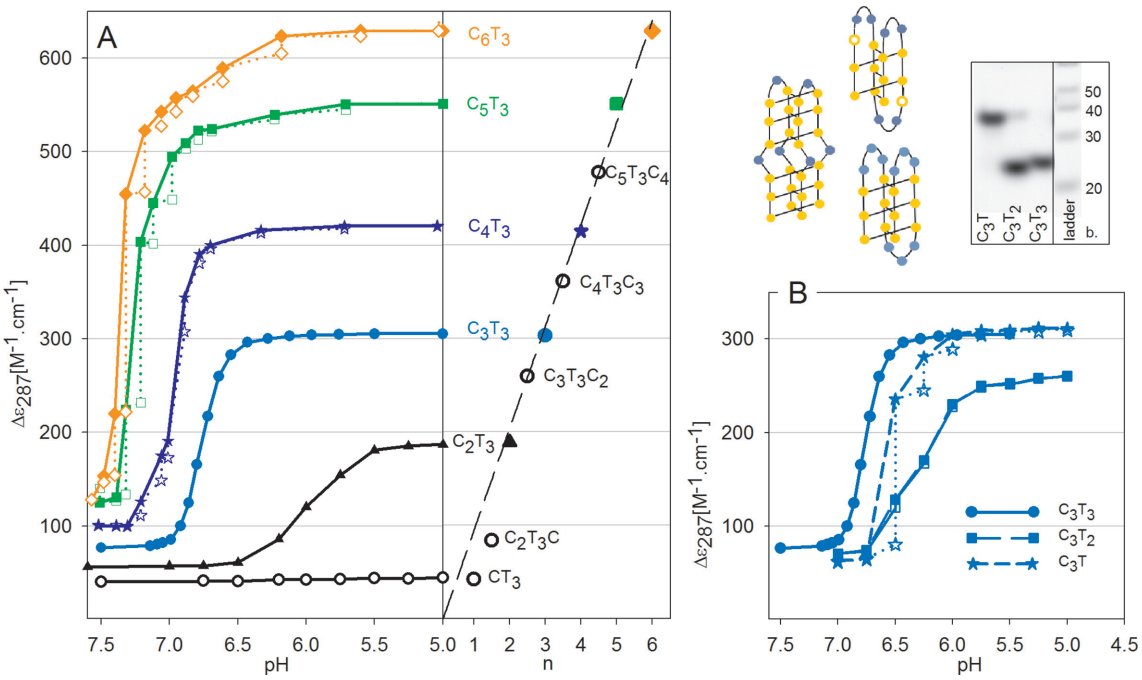

Figure 2. (**A**) (left) $\Delta\varepsilon_{287}$ of C_nT_3 for $n = 1$–6 monitoring the pH-induced iM formation: Open and full symbols connected by vertical dotted lines correspond to non-equilibrium (measured immediately) and equilibrium (measured next day) states, respectively. (A, right) $\Delta\varepsilon_{287}$ of the final iM structures of C_nT_3 at pH 5 plotted as a function of n; dashed line represents linear fit of the experimental points. (**B**) $\Delta\varepsilon_{287}$, plotted as a function of pH for sequences C_3T_3 (circles), C_3T_2 (squares), and C_3T (stars—open and full symbols correspond to non-equilibrium and equilibrium states, respectively); (top) Schematic drawings of C_3T_{1-3} and native PAGE performed at pH 5.

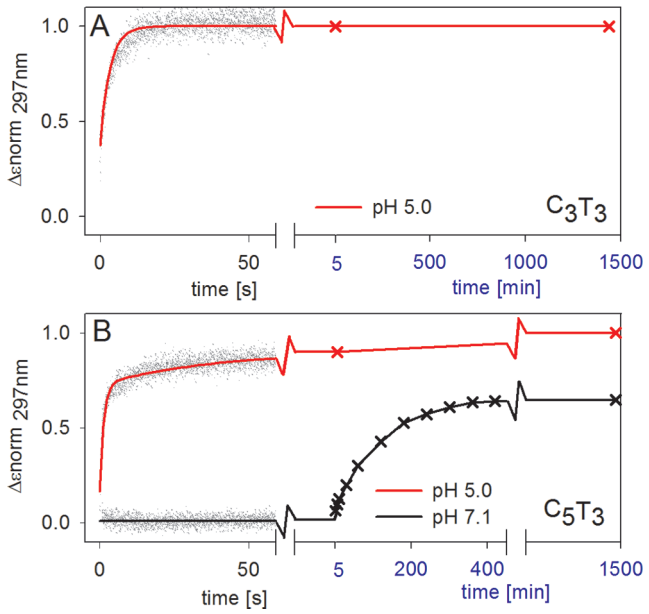

Figure 3. Kinetics of iM formation of (**A**) C_3T_3 and (**B**) C_5T_3 at pH 5 (red) and pH 7 (black) measured at room temperature by stopped-flow CD in seconds up to 5 minutes, and by standard CD spectroscopy in the course of minutes and the next day. Normalized Δε values are expressed as 0–1, where 0 corresponds to Δε measured at pH 8, and 1 corresponds to full iM form measured at pH 5.

first 10 s after adjusting pH to 5. The same $\Delta\varepsilon_{297}$ is observed at repeated CD measurements after 5 min and the next day. Changes in CD spectra of C_nT_3 with $n > 3$ are also fast in the first 10 seconds, but iMs are formed only partially (~70% in the case of C_5T_3, Figure 3, Supplementary Figure S2C). $\Delta\varepsilon_{297}$ further increases for the next CD measurement, and the final equilibrium values are only attained within hours, depending on pH (Figure 3B). With increasing pH, the time of reaching equilibrium increases. All experiments present in the paper were thus conducted the next day after adjusting pH to desired values. C_3T_3 adopts iM also slower at increased pH but faster than a repeated CD measurement can be performed, i.e. <5 min. We call this as 'fast kinetics' throughout the paper; in contrast, systematic changes measurable by standard CD are referred to as 'slow kinetics'.

Folding and unfolding curves of C_nT_3 sequences exhibit hysteresis, both with pH (Supplementary Figure S2A) and temperature changes (Supplementary Figure S2B), the extent of which depends on n. No hysteresis is observed with C_3T_3, the hysteresis is small with C_4T_3, but it becomes extensive with further increasing n and pH (Supplementary Figure S2A, B). Starting from C_4T_3 the iMs can probably adopt more conformations (as shown for long C_n blocks (37)), which may hinder the formation of the stable form and may be responsible for the observed slow kinetics and hysteresis.

CD reflects number of C.C$^+$ pairs forming the iM structure. CD amplitudes at 287 nm of the completely folded iMs linearly increase with the increasing number of cytosines in sequences C_nT_3, for n = 2–6 (Figure 2A, right). This dependence also contains the $\Delta\varepsilon_{287}$ values corresponding to iMs with odd number of C.C$^+$ pairs, represented by oligonucleotides $C_3T_3C_2$, $C_4T_3C_3$ and $C_5T_3C_4$ (5, 7 and 9 C.C$^+$ pairs, respectively). Every other extension of molecule by another C.C$^+$ pair results in a constant increase in $\Delta\varepsilon_{287}$. The linear dependence of Δε on C_nT_3 can be used for determination of the number of C.C$^+$ pairs in iMs of analogous sequences. C_2T_3C and CT_3 (Supplementary Table S1, Figure 2A, right) lie outside the linear dependence. The $\Delta\varepsilon_{287}$ of CT_3 does not change with pH (Figure 2A). The sequence C_2T_3 with four (two and two) C.C$^+$ pairs is thus the last one able to form iM at room temperature.

Interestingly, the alternating sequence $(CT)_{14}C$ does form iM (Supplementary Figure S3) with pK 5.92 and T_m 35°C. It can be expected that T in the CTCTC blocks are taken out of the iM core formed by three and three C.C$^+$ pairs, and TCT triplets form iM loops (sketch in Supplementary Figure S3). Shorter alternating (CT) sequence $(CT)_{10}C$ forms iM (similar to C_2T_3) at much lower pH.

Effect of loop length and composition on iM structure and stability

Effect of loop length. The next stage of our studies was a comparison of iM formation by $(C_3X_{1-3})_3C_3$ sequences in terms of number of C.C$^+$ pairs, pK_A values and thermodynamic stability with respect to iM molecularity (Table 2).

The 21-mer C_3T_3, described in the previous section, is used as a reference for other sequences. C_3T reaches approximately the same $\Delta\varepsilon_{287}$ value at acidic pH as C_3T_3 (Figure 2B). The transition, however, proceeds at slightly lower pH values and, mainy, with slow kinetics. This indicates more than one molecule participating in its iM structure. This is in accordance with native electrophoresis taken under the same conditions where C_3T was observed to run as a dimer (Figure 2B, Supplementary Figure S4). The iM of C_3T_2 reaches lower CD amplitudes and is formed with decreased cooperativity at more acidic pH values than the previous two oligonucleotides (Figure 2B, Table 2). The difference in $\Delta\varepsilon_{287}$ between C_3T_3 and C_3T_2 is the same as that between C_3T_3 (6 C.C$^+$ pairs) and $C_3T_3C_2$ (5 C.C$^+$ pairs). This indicates that some Cs are not involved in the iM of the C_3T_2 stem but are incorporated in loops.

The studied fragments were probed with nascent bromine to indicate which cytosines are involved in C.C$^+$ pairs, thus protected from bromination, and which are incorporated into loops (Supplementary Figure S5A). In the case of C_3T_3, all cytosines interact very slightly with bromine at pH 5, only the C3 nucleotide is somewhat more susceptible to interaction. In the case of C_3T_2, the C3 and C11 are accessible to bromination and, to a lesser extent, also C8 (Supplementary Figure S5A and sketch). This indicates that two thymines are not sufficient to form loops, especially the middle one. In case of C_3T iM, we observed a higher accessibility for bromine in cytosines C3, C7, C9 and C15 (Supplementary Figure S5A). As a result, the stem would contain only four C.C$^+$ pairs (Supplementary Figure S5). However, this result does not correlate with the CD results, indicating that all Cs are involved in the iM of C_3T. As revealed by PAGE (Figure 2B), C_3T forms a bimolecular iM. We found that the molecularity of C_3T is dependent upon DNA

Table 2. T_m/T_{ren} at pH 5 and pK_A values of sequences differing in the number and type of nucleotides in loops, m* indicates molecularity

Oligonucleotide	Abbreviation	pK_A	T_m /T_{ren} [°C]	m^*
$(C_3T)_3C_3$	C_3T	6.58	58.7/54.7	2
$(C_3T_2)_3C_3$	C_3T_2	6.26	54.9	1
$(C_3T_3)_3C_3$	C_3T_3	6.76	60.6	1
$(C_3A)_3C_3$	C_3A	6.10	42.6/40.8	2
$(C_3A_2)_3C_3$	C_3A_2	6.13	42.0/41.0	1
$(C_3A_3)_3C_3$	C_3A_3	6.23	45.4	1
$(C_3G)_3C_3$	C_3G	5.91	40.6/38.2	2
$(C_3G_2)_3C_3$	C_3G_2	5.37	53.3	1
$(C_3G_3)_3C_3$	C_3G_3	-	-	1, 2
C_{15}	C_{15}	7.05	71.6/70.3	2
C_{18}	C_{18}	7.13	77.6/75.5	1
C_{21}	C_{21}	7.25	83.1/81	1

concentration. While C_3T forms an intramolecular iM at 0.2μM DNA strand concentration used in the bromination experiments, it forms a bimolecular iM at 4 μM strand concentration used for spectroscopic methods. No free Cs were detected by bromine footprinting in a control experiment with DNA concentration used in CD (Supplementary Figure S5C).

Thermostability of the C_3T_3 iM is 60.6°C and that is by 5.7°C higher than that of the iM of C_3T_2 (Table 2). A difference of 9°C was observed between melting temperatures of the iMs of $C_3T_3C_2$ and C_3T_3, differing by one C.C$^+$ pair (Tables 1 and 2). Thus, the decrease of C_3T_2 thermostability may be the cost of losing two cytosines from the iM core to be incorporated into the loops. The bimolecular iM of C_3T melts more cooperatively than the two other sequences (C_3T_2 and C_3T_3) and its thermostability, as compared with C_3T_3, is only slightly lower than that of C_3T_3, but yields a hysteresis of 4°C (Supplementary Figure S6). The results suggest that the optimal number of thymines in loops of an intramolecular iM is three. If only two Ts are present, one C is consumed by loops at the expense of losing one C.C$^+$ pair. Single Ts in all three loops do not permit intramolecular folding. The resulting bimolecular structure is almost as stable as an intramolecular one with three Ts (Table 2).The single T in the bimolecular iM is sufficient to form the central loop without disrupting the C.C$^+$ pairs.

The influence of single-nucleotide loop on the iM molecularity. Owing to the destabilizing influence of single Ts on the intramolecular iM structure, we observed the effect of replacing T_3 with a single T in individual loops (Supplementary Figure S7, Supplementary Table S2). T/T_3 substitution in only a single loop predominantly retains an intramolecular iM structure but traces of bimolecular iMs are present with T/T_3 in the middle loop. (Supplementary Figure S7A).

In the case of replacing two T_3 by single Ts in the first and second or in the second and third loop, bimolecular iMs are preferentially formed, while in the case of T,T_3,T in particular loops, the intramolecular structure is preserved (Supplementary Figure S7A). The distinct length of the first and the third loop is unfavorable for iM structure, which is consistent with (30,36,47). More detailed information on stability and $\Delta\varepsilon_{287}$ of T/T_3 substituted sequences is contained in Supplementary Table S2 (inside 7) and in Supplementary Figure S7B and its legend.

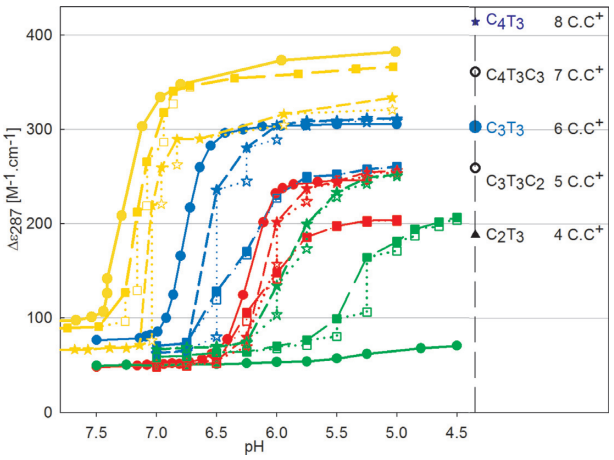

Figure 4. $\Delta\varepsilon_{287}$ reflecting iM formation of C_3X_n sequences for X = T (blue), A (red), G (green), and C (yellow); n = 1 (stars), 2 (squares) or 3 (circles), plotted as a function of pH. Open symbols correspond to non-equilibrium states. Right panel shows $\Delta\varepsilon_{287}$ of the final iM structures of C_nT_3 sequences at pH 5, taken from the right part of Figure 2A.

In contrast to only bimolecular iMs with single T in all three loops of C_3T, the intramolecular iM structure is preserved in the case of C_4T with four Cs in blocks (Supplementary Figure S7A). The length of the loops can be extended (to 2, 3, 2) by cytosines from the iM stem resulting in seven and six C.C$^+$ pairs in C_4T_2 and C_4T, respectively, as compared to eight pairs in C_4T_3 (Supplementary Figure S8, sketches). The $\Delta\varepsilon_{287}$ values of their iMs agree with this model (Supplementary Figure S8).

Effect of nucleotide types in loops. Next, we compared the results shown with C_3T_n sequences (Figure 2B) with the effect of other nucleotides in loops (Figure 4). pDifferent chromophores of loop nucleotides do not distinctly contribute to the CD effect reflecting iM formation (Supplementary Figure S9A).

21-mer, C_3A_3, (Table 2) folds with fast kinetics into an intramolecular iM (Supplementary Figure S4) with $\Delta\varepsilon_{287}$ value 246 being close to that of C_3T_2, the iM of which contains one C.C$^+$ pair less than C_3T_3 (Figure 4). Its formation takes place at pH values 0.5 lower than that corresponding to C_3T_3.

In analogy with C_3T_2, 18-mer, C_3A_2 (Table 2) folds predominantly into an intramolecular iM, though there is a

slight trace extending up to the bimolecular species on the electrophoresis (Supplementary Figure S4). The iM formation of C_3A_2 proceeds at a lower pH than that of C_3A_3, it is much less cooperative and the final structure reaches $\Delta\varepsilon_{287}$ value close to that of C_2T_3 (Figure 4) containing four C.C$^+$ pairs. The bromination of C_3A_2 identified C3, C6, C11 and C16 susceptible to interaction, which supports the presence of four C.C$^+$ pairs in C_3A_2 iM stem (Supplementary Figure S5B).

C_3A forms with a slow kinetics a bimolecular iM with $\Delta\varepsilon_{287}$ reaching the same value as C_3A_3. Its melting displays a hysteresis (Supplementary Figure S6). At the low DNA concentration used for bromination experiments however, it folds intramolecularly. The bromine footprint indicates C3, C5, C9, C13, and C15 (Supplementary Figure S5B). Thermostabilities of the iMs containing As decrease in the order $C_3A_3 > C_3A \geq C_3A_2$ (Table 2, Supplementary Figure S6). The difference in melting temperatures between C_3A_3 and C_3A_2 is 4.4°C, which is similar to that between C_3T_3 and C_3T_2 (loss of one C.C$^+$ pair). The CD results (Figure 4) imply, and the bromination experiments seem to support (Supplementary Figure S5B) that iMs of the set of sequences with the bulky A in the loops lose one C.C$^+$ pair as compared with their analogous sequences containing Ts. Supplementary Figure S9 shows in the example of several sequences containing T and A bases in loops that the mere presence of A is not the cause of the decrease in $\Delta\varepsilon_{287}$ in A versus T containing iMs. The same follows from Supplementary Figure S7B. Thermal stability of the set with A loops is still much lower than would correspond to the loss of one C.C pair (Table 2). A negative effect of bulky purines on iM stability was also reported by (25,31,33,34). Conversely, thermal stability of the sequences with Ts may be supported by T.T pairs stacked on the adjacent C.C$^+$ pairs (34).

Sequences containing a single or even two Gs in loops are also able to adopt the iM. Similarly to the previous 15-mers (C_3T and C_3A), C_3G (Table 2) forms with a slow kinetics a bimolecular iM, (Supplementary Figure S4) but less cooperatively and at an even lower pH (Figure 4, Table 2). The transition of C_3G starts close to pH 6 and the final iM reaches the same $\Delta\varepsilon_{287}$ value as C_3A. The iM of C_3G melts at ~40°C with a hysteresis of 3°C between melting and renaturation curves (Supplementary Figure S6). The iM of C_3G_2 arises with comparable kinetics as C_3G, but at a much more acidic pH value (Figure 4, Table 2). Nevertheless, $\Delta\varepsilon_{287}$ of the final iM structure of C_3G_2 achieves the same value as that of C_3A_2. C_3G_2 is mainly intramolecular, but its formation is hindered by an interfering bimolecular species, which can be seen by electrophoresis (Supplementary Figure S4). Both purines, adenine and guanine in loops influence the iM structure in a similar way, but its competition with a duplex in the case of Gs in the loops decreases the pH of the transition and slows down the kinetics of iM formation. The presence of G.C pairs may explain a distinct increase in the T_m of C_3G_2, i.e. by about 13°C as compared with the iM of C_3G (Table 2, Supplementary Figure S4). C_3G_3 adopts only duplex and hairpin structures, both in acid and in neutral pH (no iM is formed).

We have also studied analogous sequences with cytosines in loops i.e. C_{15}, C_{18} and C_{21}. The 21-mer adopts with a fast kinetics an intramolecular iM (Supplementary Figure S4), with the transition midpoint of 7.25 and $\Delta\varepsilon_{287}$ of the final structure even higher than that of C_3T_3 (Figure 4). The iM of C_{18} yields a similar value of $\Delta\varepsilon_{287}$, but its formation is slower. The increase in $\Delta\varepsilon$ means that three nucleotides are not needed for loops in the case of cytosines and that they form additional C.C$^+$ pairs. The dC_n oligonucleotides can adopt various iM arrangements. Various folding possibilities of long C_n stretches (C_{12}–C_{30}) into intramolecular iMs were studied by Burrows group (30,31). The authors demonstrated that the transition pH and T_m values of the C_n iM structures displayed a four-nucleotide ($4n - 1$) repeat pattern. The optimal folding for various C_n lengths was tested using model sequences with substituted non-C (T,U,A,G) deoxynucleotides in selected positions. The loop nucleotide influenced iM stability in the order C > T~U >> A~G, which is in line with our observations.

In our study, only C_{15} behaves like the other sequences with non-C loops. It forms, in the same way as the other sequences with single nucleotide in loops, a bimolecular iM (Supplementary Figure S4) with $\Delta\varepsilon$ close to that of C_3T (Figure 4). All sequences with Cs in loops adopt iMs at higher pH values than analogous sequences with non-C loops and their stability is also distinctly higher (Table 2, Supplementary Figure S6) (31).

iM of long cytosine blocks in $(C_nT_3)_3C_n$ sequences

Similarly as shorter sequences, all the iMs of C_nT_3 with $n > 6$ are intramolecular for both pH 5 and 6.5 (Supplementary Figure S1). Higher associates were only observed after annealing iM structures stabilized at pH 5, but not at pH 6.5.

The dependence of the $\Delta\varepsilon_{287}$ of iMs formed at pH 5 by C_nT_3 on the number n of cytosines in blocks (Figure 2) remains linear up to the longest measured sequence $C_{10}T_3$ (Supplementary Figure S10A). The same is true for the dependence measured at pH 6.5. Only the sequences with the shortest C blocks reach lower $\Delta\varepsilon_{287}$ values than corresponds to the line (Figure 2), which is much more pronounced at pH 6.5 than at pH 5 (Supplementary Figure S10A). The CD data in the figures are related to molar strand concentration, which means that $\Delta\varepsilon_{287}$ values represent the sum of contributions of individual C.C$^+$ pairs in whole iM molecules responsible for the CD signal. To learn the contribution of individual C.C$^+$ pairs to CD signal with respect to the number of cytosines in blocks, the same amounts of DNA have to be compared. The dependence of $\Delta\varepsilon_{287}$ on the number of cytosines in blocks related to molar nucleoside concentration (Figure 5A) culminates for n around 5–6 and does not increase any more for longer C stretches. The long-wavelength UV absorption reflecting cytosine protonation changes in the same way as the molar CD and reaches a plateau for C runs longer than five (Figure 5B). Analogously, UV absorption in the absorption maximum around 260 nm decreases (Supplementary Figure S10D). These dependencies copy the course of T_m values of iMs as a function of the length of cytosine blocks at pH 5 (Figure 5C). The T_m dependence is, however, different at pH 6.5.

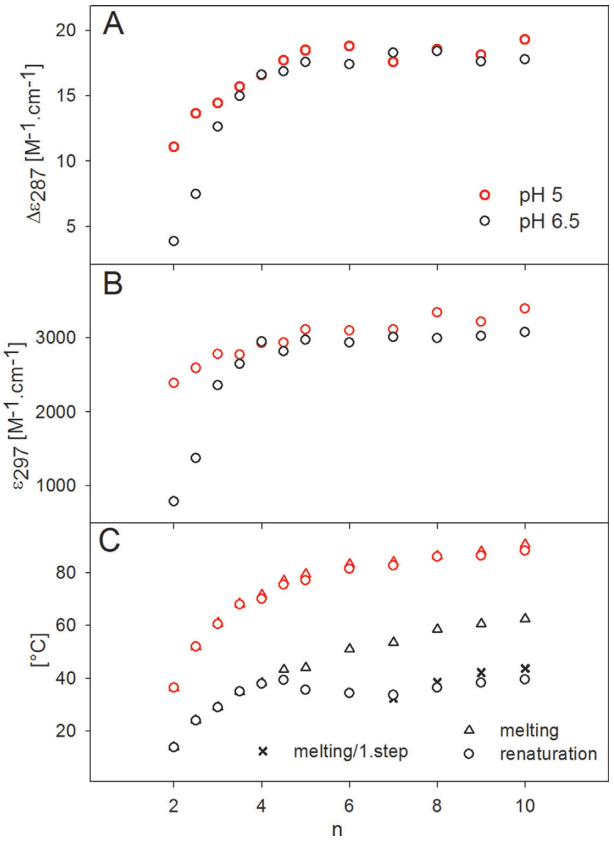

Figure 5. (**A**) $\Delta\varepsilon_{287}$ and (**B**) ε_{297} of C_nT_3 and $C_nT_3C_{n-1}$ sequences related to nucleoside concentration and (**C**) their T_m values plotted against the number (n) of block cytosines; (red) pH 5, (black) pH 6.5. In (C), triangles represent T_m of monophasic melting (pH 5) or T_m of the high-temperature phase of biphasic melting (pH 6.5), crosses represent T_m of the low-temperature phase of biphasic melting. Circles represent the midpoint (T_{ren}) of the renaturation phase.

Melting and renaturation proceed as single-step processes with only a small hysteresis at pH 5, whereas a distinct two-step melting with a strong hysteresis is observed in iMs at pH 6.5 for sequences with n>6 (Figure 5C). Thus, iMs of long C_nT_3 sequences fold into two types of structures, which melt at distinctly different temperatures. The two-step melting of C rich sequences at pH close to neutral has been already observed by Z. Waller's group (37). The authors suggested the iM and a foldback as a possible explanation of the two melting forms.

The two-step melting reflected by changes of ε_{297} is shown in Figure 6 for C_9T_3. The population of the two forms changes according to pH. The less stable form (form 1) prevails at neutral pH, and with decreasing pH gradually fades out. Nearly a single form, the more stable one (form 2), is present at pH 5. Analogous but inverted changes are displayed by ε_{260} (not shown).

Calorimetry, carried out at about ten times higher DNA concentration as compared to standard CD experiments, detected only a single form of C_9T_3 at pH 5, whereas two distinct structures are observed at pH 6.5 (Supplementary Figure S11). (Yet another weakly populated structure is detected by calorimetry at pH 6.5, the indication of which can be traced at low temperature on the melting curve in the insert of Figure 6B. The existence of another low-temperature transition can be clearly seen in Figure 6A at pH 7.2).

Renaturation proceeds in a single step. The hysteresis between melting and renaturation processes increases with pH increasing towards neutral conditions (Figure 6). (Interestingly, the single-step melting of C_nT_3 with $4 \leq n < 7$ - Supplementary Figure S2B provides the same type of hysteresis). After returning to low temperature, the original equilibrium between the two forms is attained again and the repeated melting curve copies the original one. The population of the two forms was the same for the C_9T_3 sample prepared one or three days before its measurement (Supplementary Figure S12). The return of the sequence C_9T_3 from its denatured state to a temperature within the region of the plateau between the two melting processes is not reversible: No increase in ε_{297} occurred after keeping the sequence at the respective temperature (four hours at 53°C for C_9T_3 at pH 6.5, or 24 h at 27°C for C_7T_3 at pH 7.2); the sequence remained denatured. Thus, form 2 does not occur without form 1 at pH values close to neutral.

The hyperchromicity of the 297 nm absorption band reflecting cytosine protonation is distinctly more extensive for pH close to neutral value than for pH 5 (Figure 6). To find the reason for this difference, the absolute ε_{297} values, instead of relative or normalized ones, have to be plotted. Long-wavelength (~297 nm) absorption of iMs at acidic and neutral pH values do not differ much for folded states, but the absorption values distinctly differ for denatured states: ε_{297} of the denatured state of C_9T_3 at pH 5 is nearly double that at pH 6.5 (Supplementary Figure S13A). CD spectra of the denatured state of C_9T_3 also differ at pH 5 and 6.5 (Supplementary Figure S13B). The maximum of the positive CD band of the denatured form remains close to 287 nm at pH 5 i.e. close to the maximum of the folded state, whereas it is shifted to 275 nm at pH 6.5, as shown for the unstructured form in Figure 1C. Thus, the denatured sequence may still contain protons at pH 5.

In Figure 7 the same dependences of ε_{297} on temperature are presented as in Figure 6, but at particular pHs for more C_nT_3 sequences. Depending on the length of the C blocks, the population of the two structural forms changes, while the total ε_{297} change during denaturation remains more or less constant at a given pH. With increasing the length of the C blocks, the population of the less stable form 1 increases. Simultaneously, form 1 becomes more populated with the increasing pH (Figure 7B). Thus, it seems that the number of available protons at particular pH values is sufficient for a limited number of Cs in the long blocks. The hysteresis of renaturation increases with increasing length of C_n blocks and increasing pH (Supplementary Figure S14).

What is form 1?

CD spectra of C_9T_3, measured in the course of melting (Figure 6B), reveal that forms 1 and 2 cannot be distinctly different. Their mixture at low temperature yields a CD spectrum characteristic of the iM structure and the spectra in the course of melting are linearly dependent (Supplementary Figure S15A) and intersect in isodichroic points (240.5 and 276.5 nm). The whole melting behaves as a two state process. This means that the CD does not discriminate between

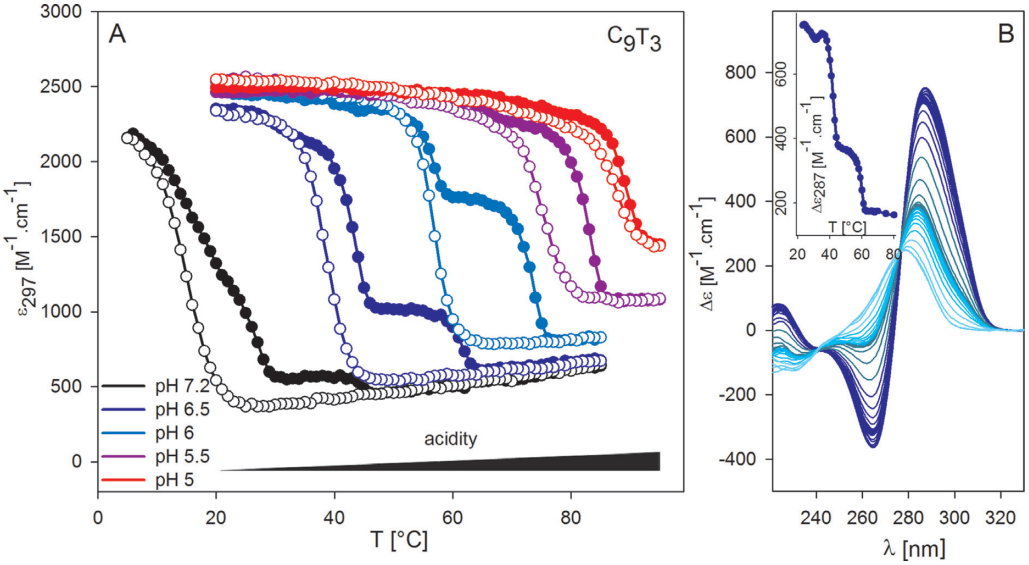

Figure 6. (**A**) Melting (full circles) and renaturation (empty circles) curves of C_9T_3 sequence, monitored by ε_{297}, at various pH given in the figure. (**B**) CD spectra of C_9T_3 recorded from 23°C (dark) to 80°C (light) with 1°C step at pH 6.5. Insert shows $\Delta\varepsilon_{287}$ as a function of temperature.

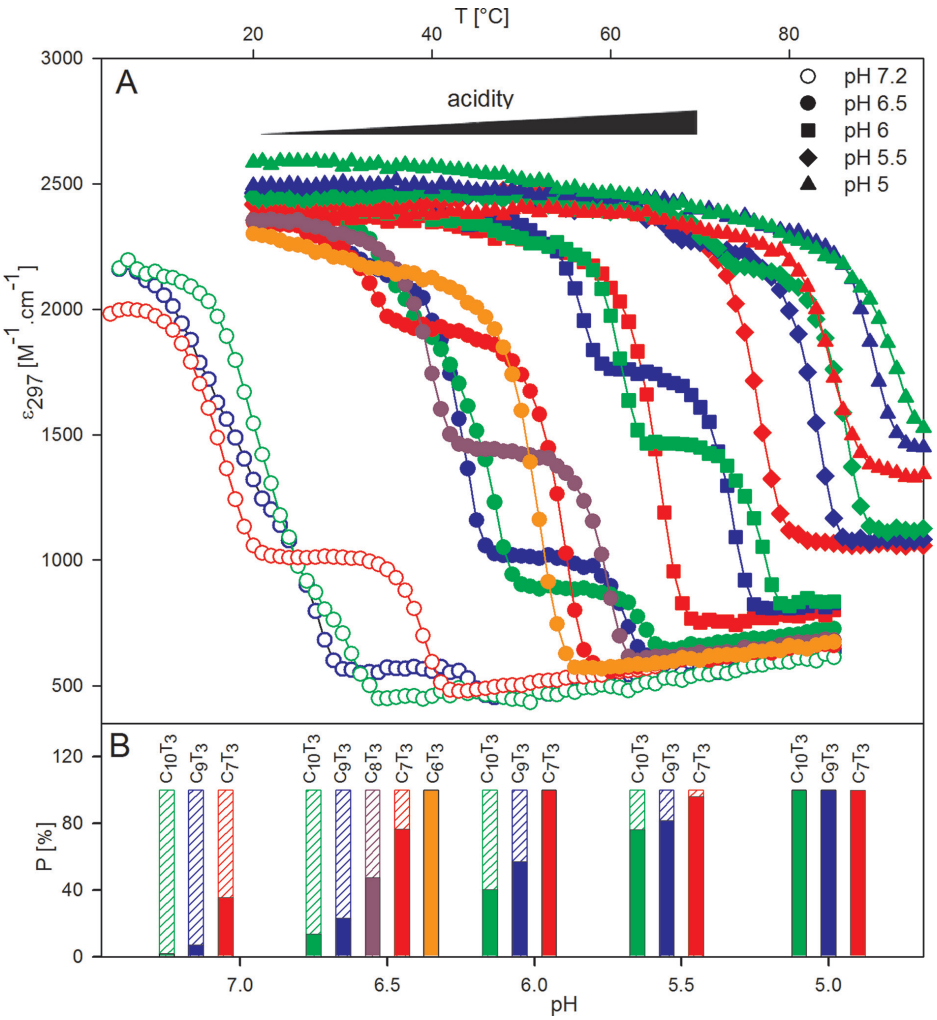

Figure 7. (**A**) Melting curves of C_nT_3 sequences for $n = 6$ (orange), 7 (red), 8 (violet), 9 (blue) or 10 (green), monitored by ε_{297} at various pH values. (**B**) Bar chart showing relative proportions of the less stable form 1 (pattern) and the more stable form 2 (full) of particular C_nT_3 at different pH values.

forms 1 and 2. This conclusion is supported by a SVD analysis, which detects only two spectrally distinguishable forms during C_9T_3 melting (Supplementary Figure S15B).

The same results follow from ^{1}H NMR measurements (Supplementary Figure S16). The spectrum of C_9T_3 at low temperature, thus of the mixture of the two forms, is characteristic of the iM structure. The signal at about 15.5 ppm belonging to H^+ between N3 of cytosine pairs, and the signal around 9 ppm of amino protons bound to oxygens decrease with increasing temperature in two steps (Supplementary Figure S16A), and in the same temperature regions as observed with CD. The population of the less stable form 1 slightly decreased for higher DNA concentration used for NMR (0.2mM DNA) as compared with CD (2μM DNA) (Supplementary Figure S16C) but the T_m values of both structures remained unchanged.

NMR signals in the region between 6 and 8 ppm (Supplementary Figure S16B) coming from H5 and H6 of cytosines are distinctly lower at 20 and 40°C than at 45°C, when form 1 became denatured. The spectra remain the same at temperatures from 45 to 52°C, i.e. in the plateau of CD dependence, and a distinct increase in signals takes place at 65°C, when form 2 also melted. The NMR results confirm the existence of two iM forms but, in the same way as CD results, do not distinguish between them.

What is thus the difference between the two iM forms? What does the structure of long C blocks look like?

'Superfolded' iM

It follows from the UV absorption and CD measurements and mainly from temperature dependences (Figure 5) that long C blocks are not advantageous for iM structure stability. A thermodynamic hindrance of C blocks longer than four to form a stem of intercalated $C.C^+$ pairs is noticeable from the electrophoretic results (Figure 8A) of sequences C_n and $C_nT_3C_n$: A 7-mer C_7 forms a (not very stable) tetramolecular iM, but C_8 isomerizes between the four molecular and bimolecular structures. C_9 and longer sequences up to C_{15} exclusively form bimolecular iMs, whereas still longer sequences C_{18}–C_{21} (Figure 8A, top) fold intramolecularly. Similarly, $C_4T_3C_4$ and $C_5T_3C_5$ form bimolecular iMs, but $C_6T_3C_6$ forms, in addition to bimolecular, an intramolecular iM and all other sequences with longer C blocks fold to iM only intramolecularly (Figure 8A, bottom). Analogous with the above results, we expect that the intramolecular iMs of $(C_7T_3)_3C_7$ and those with longer C blocks may turn back somewhere along the C stretch to superfold into shorter i-motifs.

If our expectation is correct, we should be able to observe in long C stretches some regions forming additional loops and thus be susceptible to interaction with bromine. In the case of C_9T_3 at pH 5, we can see in the middle of light areas of Cs forming $C.C^+$ pairs, dark places around Cs 5, 17, 29 and 41, which interacted with bromine (Figure 8B). The same holds for pH 6.5 but not for pH 8. The interacting Cs are less distinct at pH 6.5, which, however, says nothing about the population of the superfolded structures at the two pHs as the whole blocks are globally darker at pH 6.5 than pH 5. The iM is probably less compact at pH close to neutral, and thus more susceptible to interaction with bromine. It follows from these experiments that sequences C_nT_3 containing long C blocks prefer to fold into multiple iM structures with short blocks of probably three or four $C.C^+$ pairs.

A separation of the original iM with long C blocks into two (or more) short iMs could be accompanied with a change in the compactness of the two structures. In fact, the appearance of a slower band can be observed on the electrophoresis running at pH 6.5 for C_9T_3 and longer motifs, the population of which increases with the length of C blocks (Figure 9B, Supplementary Figure S1B). The slower bands can also be detected at pH 5, but they are in a minority (Supplementary Figure S1A). Similar two close-positioned electrophoretic bands were observed (48) with 3′E and 5′E iM forms (49), but this does not explain the two bands under discussion. The 3′E and 5′E iM forms do not differ in thermal stability.

Two iM forms differing in topology have been observed in a C rich fragment of the human centromeric satellite (10) and also in the fragment of the human telomere sequence $(C_3TAA)_4$ (25). In both cases, the second iM structure is formed at a pH lower than 4.6. The authors consider protonation of adenine to be responsible for the formation of the low-pH form, which does not thus explain the two distinct structures observed close to neutral pH.

In a recent paper, Rogers *et al.* observed an anomalous hysteresis (50) and a multiphasic pH-induced iM unfolding of the human telomere DNA sequence, of oligo C_{19} and of a fragment of the RAD 17 promoter (R17). The unusual hysteresis was observed to be dependent upon how fast the iM structure was formed. R17, consisting of two long C blocks interrupted by G, adopted, upon rapid folding (injection into pH 4.5), an iM arrangement, which unfolded in three phases. The authors explain the anomalous behavior by the presence of three differently folded iM structures: 5′E and 3′E iM structures, and by dimers of higher intermolecular products. Lieblein *et al.* also showed that a fast jump from pH 9 to 6 could give rise to 3′E and 5′E iM conformations, with the latter being more populated (49).

None of the above iMs can explain our distinctly different observations: Formation of the two iM structures does not depend on the way they are induced. The population of particular forms remains the same after repeated melting and annealing processes (Supplementary Figure S12). The structures are intramolecular, and no dimers or higher-order structures were observed even after annealing at pH 6.5 (Supplementary Figure S17). The two structures are induced at pH close to neutral values, while only a single structure is formed at pH 5. In contrast with the neutral iM forms, properties, namely molecularity, of the acidic iM form depend on the way of its induction (51), which is in line with our observations (Supplementary Figure S1A, B).

What is the difference between the iM forms 1 and 2?

This question still remains open and we have no direct evidence to support our model: In the excess of protons at pH 5, not only the hemiprotonated $C.C^+$ pairs forming the iM stem contain protons but also the denatured forms remain protonated (Supplementary Figure S12). In contrast, in the lack of protons around neutral pH, the protons have

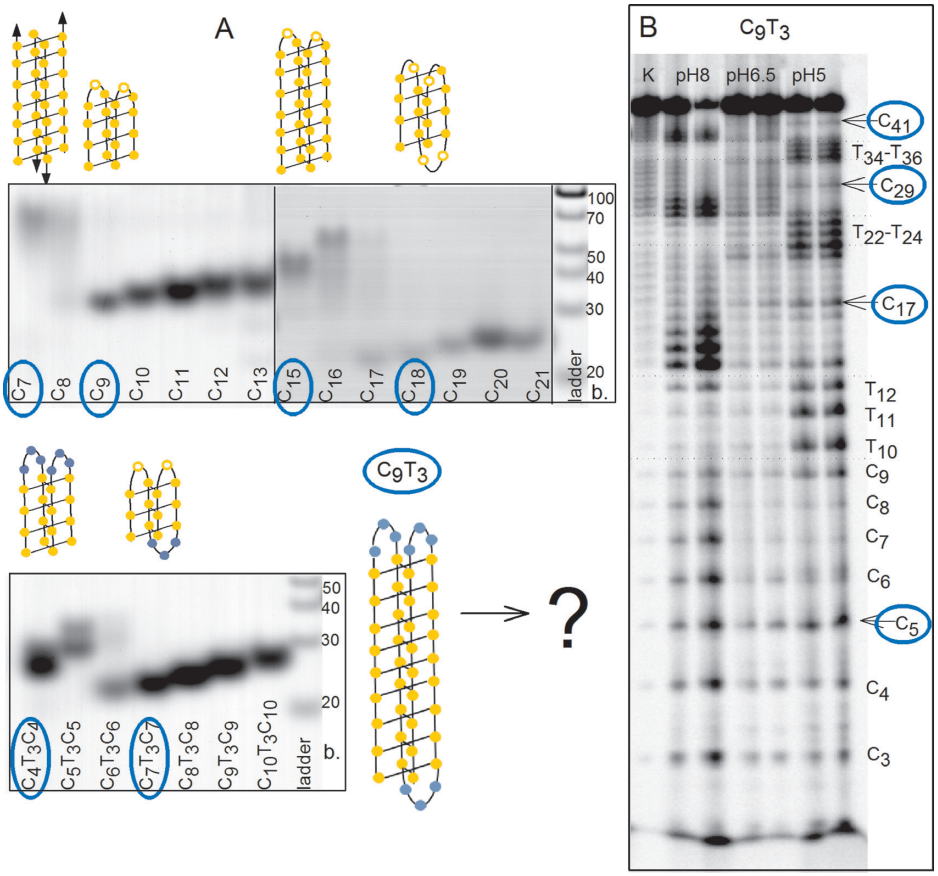

Figure 8. (**A**) Native electrophoreses of C_n (top) and $C_nT_3C_n$ (bottom) sequences run at pH 5. Sketches of iMs correspond to selected sequences differing in molecularity. (**B**) Nascent bromine probing of C_9T_3 sequence at three different pH, as indicated. The runs are in duplicate.

to be economized. They will be preferentially situated in areas where they can contribute to structure stability, i.e. in the iM stem, while they can be initially depleted in loops. If the loop cytosines in the superfolded part of iM structure of C_9T_3 do not carry protons, then $C_4TC_4T_3$ containing T instead of the middle C could melt similarly as the less stable form 1. Actually, $C_4TC_4T_3$ displays the same $\Delta\varepsilon$ values as C_9T_3 at pH 6.5 for both native and denatured states and its one step melting and renaturation curves are close to the renaturation curve of C_9T_3 (Figure 9A). The second part of C_9T_3 melting occurring at much higher temperatures does not take place with short iMs of $C_4TC_4T_3$. On the electrophoresis, the iM structure of $C_4TC_4T_3$ remains level with the slower band of C_9T_3 (Figure 9B). Both iMs thus adopt similarly compact structures formed by a couple of iMs with half the length size of C_9T_3 iM. At pH 5, T_m of the iM of $C_4TC_4T_3$ is lower than that of C_9T_3 and a hysteresis of its refolding into the superfolded iM structure is greater. We expect that the lack of protons supports bending of long C blocks to form short, less protonated and less stable structures. With decreasing pH the population of short, bent iMs decreases and that of iMs with long Cn stretches starts prevailing (Figures 6 and 7). We suggest that the difference between the iM forms 1 and 2 dwells in the different length and protonation of their structures.

CONCLUSION

We have identified the optimal conditions and minimal requirements for i-motif (iM) formation in C-rich DNA sequences. We show that the minimum number of block cytosines (n) required to form intramolecular iM of $(C_nT_3)_3C_n$ sequences is two, i.e. two couples of intercalated $C.C^+$ pairs. However, intramolecular iM can also occur if a suitable number of repetitions of the dinucleotide (CT) are present. Though iM stability increases with the length of C blocks, the optimal number of Cs for iM formation is a triad C_3. $(C_3T_3)_3C_3$ forms iM with fast kinetics and the pH- and temperature-induced iM folding is a reversible process without any hysteresis. In contrast, iMs of $(C_nT_3)_3C_n$ sequences where $n > 3$ form with slow kinetics lasting even hours depending on pH and C_n length.

We have shown that with an appropriate reference system CD spectroscopy can be used to determine the number of $C.C^+$ pairs in iM structures. We applied this and found that the optimal number of nucleotides in loops securing iM stability is three. In the case of only two non-C nucleotides, the iM is destabilized as it loses one $C.C^+$ pair in the stem to incorporate the Cs into the loop structure. Single nucleotides in loops influence the molecularity of C_3X_n iMs and their presence in all three loops (including C) prevents the formation of intramolecular iM folding. CD and bromine footprinting experiments have implied that bulky

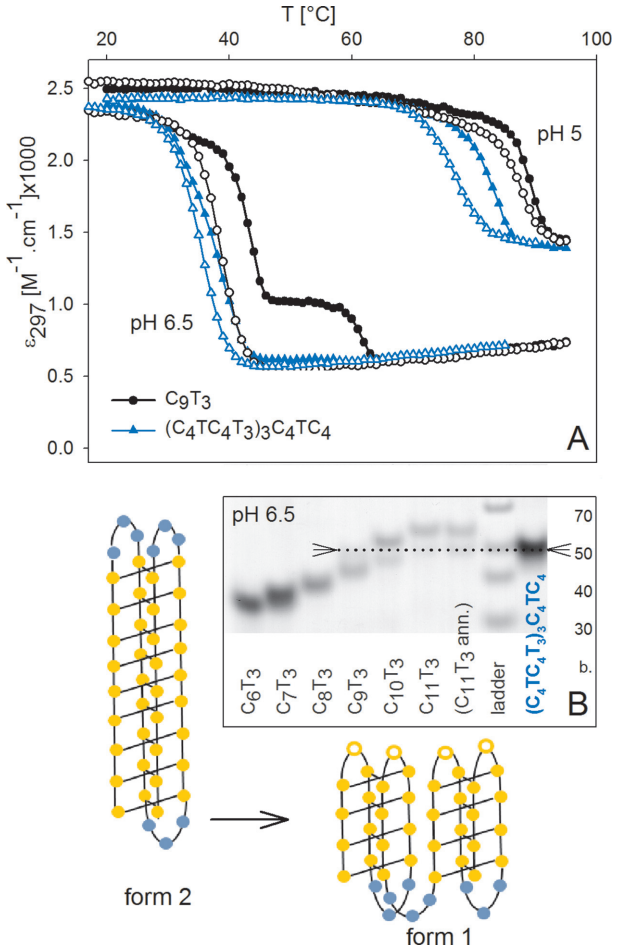

Figure 9. (**A**) Melting curves of $(C_9T_3)_3C_9$ (black) and $(C_4TC_4T_3)_3C_4TC_4$ (blue) at pH 6.5 and pH 5, measured by ε_{297}. (**B**) Native PAGE of C_nT_3 sequences for $n = 6$–11, and of $(C_4TC_4T_3)_3C_4TC_4$ performed at pH 6.5. Schematic drawings of the suggested iM structures of C_9T_3 are sketched.

purines in loops of iMs consume one extra $C.C^+$ pair from the stem as compared to thymines.

The CD signal ($\Delta\varepsilon_{287}$), absorption (ε_{297}) and stability of the iMs of $(C_nT_3)C_n$ increase with increasing C_n length and the midpoint of pH-induced iM formation shifts towards more alkaline pH up to a pH 7.5 for $n = 6$. For $n > 6$, the CD signal related to molar nucleotide concentration, absorption reflecting sequence protonation and the transition midpoint, stop increasing. Simultaneously, the iM denaturation profile shows two distinct phases at pH values close to neutral, indicating the presence of two different species. Their population strongly depends on the C-tract length and precise pH value, but not on the annealing procedure. Our ^{1}H NMR data and SVD analysis of CD spectra show that both species are iMs with the same spectral characteristics, thus indistinguishable with these methods. The thermally less stable form (form 1) prevails at neutral pH and with decreasing pH gradually fades out. Only a single form (form 2) is present at pH 5. Form 1 is present in greater quantities the higher the pH value and the longer the C blocks are, thus reflecting the decreasing ratio between the number of accessible protons and cytosines present.

Electrophoretic results and bromination experiments indicate that long C_n blocks in $(C_nT_3)_3C_n$ sequences turn back within the C runs and form two (or more) short iMs. We suggest that the difference between form 1 and 2 resides in the different lengths and protonation of their structures. We expect that the lack of protons is the driving force for iMs folding into short units, which do not need protons in newly created loops. The formation of short iMs instead of long ones may be analogous to splitting long repeated G-rich sequences into beads-on-a-string-like arrangements.

The results shown in this paper are of particular relevance in view of the long C-rich sequences found in important regions of the human genome.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

MV is indebted to J.-L. Mergny, L. Trantirek and P. Stadlbauer for fruitful discussions. The CIISB research infrastructure project LM2015043 funded by MEYS CR is gratefully acknowledged for an access to NMR instrumentation at the Josef Dadok National NMR Centre, CEITEC – Masaryk University.

FUNDING

Czech Science Foundation [17-12075S, 17-19170Y]; SYMBIT [CZ.02.1.01/0.0/0.0/15_003/0000477] financed by the ERDF. Funding for open access charge: SYMBIT [CZ.02.1.01/0.0/0.0/15 003/0000477].
Conflict of interest statement. None declared.

REFERENCES

1. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
2. Brooks,T.A., Kendrick,S. and Hurley,L. (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.*, **277**, 3459–3469.
3. Neidle,S. and Balasubramanian,S. (2006) *Quadruplex Nucleic Acids*. Royal Society of Chemistry, Cambridge.
4. Chaires,J.B. and Graves,D. (2013) In: Graves,D (ed). *Topics in Current Chemistry*. Springer-Verlag, Berlin Heidelberg, Vol. **330**.
5. Gehring,K., Leroy,J.L. and Gueron,M. (1993) A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature*, **363**, 561–565.
6. Gueron,M. and Leroy,J.L. (2000) The i-motif in nucleic acids. *Curr. Opin. Struct. Biol.*, **10**, 326–331.
7. Han,X., Leroy,J.L. and Gueron,M. (1998) An intramolecular i-motif: the solution structure and base-pair opening kinetics of d(5mCCT$_3$CCT$_3$ACCT$_3$CC). *J. Mol. Biol.*, **278**, 949–965.
8. Phan,A.T. and Leroy,J.L. (2000) Intramolecular i-motif structures of telomeric DNA. *J. Biomol. Struct. Dyn.*, **17**(Suppl. 1), 245–251.
9. 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.
10. Nonin-Lecomte,S. and Leroy,J.L. (2001) Structure of a C-rich strand fragment of the human centromeric satellite III: a pH-dependent intercalation topology. *J. Mol. Biol.*, **309**, 491–506.
11. Liu,D.S. and Balasubramanian,S. (2003) A proton-fuelled DNA nanomachine. *Angew. Chem.-Int. Edit.*, **42**, 5734–5736.

12. Ghodke,H.R., Krishnan,R., Vignesh,K., Kumar,G.V.P., Narayana,C. and Krishnan,Y. (2007) The I-Tetraplex building block: Rational design and controlled fabrication of robust 1D DNA scaffolds through non-Watson-Crick interactions. *Angew. Chem.-Int. Ed.*, **46**, 2646–2649.
13. Li,T. and Famulok,M. (2013) i-Motif-programmed functionalization of DNA nanocircles. *J. Am. Chem. Soc.*, **135**, 1593–1599.
14. Dong,Y.C., Yang,Z.Q. and Liu,D.S. (2014) DNA nanotechnology based on i-motif structures. *Acc. Chem. Res.*, **47**, 1853–1860.
15. Nesterova,I.V. and Nesterov,E.E. (2014) Rational design of highly responsive pH sensors based on DNA i-Motif. *J. Am. Chem. Soc.*, **136**, 8843–8846.
16. Modi,S., Swetha,M.G., Goswami,D., Gupta,G.D., Mayor,S. and Krishnan,Y. (2009) A DNA nanomachine that maps spatial and temporal pH changes inside living cells. *Nat. Nanotechnol.*, **4**, 325–330.
17. 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.
18. Chen,X., Zhou,X., Han,T., Wu,J., Zhang,J. and Guo,S. (2013) Stabilization and induction of oligonucleotide i-motif structure via graphene quantum dots. *ACS Nano*, **7**, 531–537.
19. Rajendran,A., Nakano,S. and Sugimoto,N. (2010) Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH. *Chem. Commun.*, **46**, 1299–1301.
20. Bhavsar-Jog,Y.P., Van Dornshuld,E., Brooks,T.A., Tschumper,G.S. and Wadkins,R.M. (2014) Epigenetic modification, dehydration, and molecular crowding effects on the thermodynamics of i-motif structure formation from C-rich DNA. *Biochemistry*, **53**, 1586–1594.
21. Pramanik,S., Nagatoishi,S. and Sugimoto,N. (2012) DNA tetraplex structure formation from human telomeric repeat motif (TTAGGG):(CCCTAA) in nanocavity water pools of reverse micelles. *Chem. Commun.*, **48**, 4815–4817.
22. Day,H.A., Huguin,C. and Waller,Z.A. (2013) Silver cations fold i-motif at neutral pH. *Chem. Commun. (Camb.)*, **49**, 7696–7698.
23. Saxena,S., Joshi,S., Shankaraswamy,J., Tyagi,S. and Kukreti,S. (2017) Magnesium and molecular crowding of the cosolutes stabilize the i-motif structure at physiological pH. *Biopolymers*, **107**, e23018.
24. 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.
25. Fernandez,S., Eritja,R., Avino,A., Jaumot,J. and Gargallo,R. (2011) Influence of pH, temperature and the cationic porphyrin TMPyP4 on the stability of the i-motif formed by the 5 '-(C(3)TA(2))(4)-3 ' sequence of the human telomere. *Int. J. Biol. Macromol.*, **49**, 729–736.
26. Abou Assi,H., Garavís,M., González,C. and Damha,M.J. (2018) i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.*, **46**, 8038–8056.
27. Lannes,L., Halder,S., Krishnan,Y. and Schwalbe,H. (2015) Tuning the pH response of i-motif DNA oligonucleotides. *ChemBioChem*, **16**, 1647–1656.
28. Gurung,S.P., Schwarz,C., Hall,J.P., Cardin,C.J. and Brazier,J.A. (2015) The importance of loop length on the stability of i-motif structures. *Chem. Commun. (Camb.)*, **51**, 5630–5632.
29. Reilly,S.M., Morgan,R.K., Brooks,T.A. and Wadkins,R.M. (2015) Effect of interior loop length on the thermal stability and pK(a) of i-motif DNA. *Biochemistry*, **54**, 1364–1370.
30. Fleming,A.M., Ding,Y., Rogers,R.A., Zhu,J., Zhu,J., Burton,A.D., Carlisle,C.B. and Burrows,C.J. (2017) 4n-1 is a "Sweet Spot" in DNA i-Motif folding of 2 '-Deoxycytidine homopolymers. *J. Am. Chem. Soc.*, **139**, 4682–4689.
31. Fleming,A.M., Stewart,K.M., Eyring,G.M., Ball,T.E. and Burrows,C.J. (2018) Unraveling the 4n-1 rule for DNA i-motif stability: base pairs vs. loop lengths. *Org. Biomol. Chem.*, **16**, 4537–4546.
32. McKim,M., Buxton,A., Johnson,C., Metz,A. and Sheardy,R.D. (2016) Loop sequence context influences the formation and stability of the i-motif for DNA oligomers of Sequence $(CCCXXX)_4$, where X = A and/or T, under slightly acidic conditions. *J. Phys. Chem. B*, **120**, 7652–7661.
33. Benabou,S., Garavis,M., Lyonnais,S., Eritja,R., Gonzalez,C. and Gargallo,R. (2016) Understanding the effect of the nature of the nucleobase in the loops on the stability of the i-motif structure. *Phys. Chem. Chem. Phys.*, **18**, 7997–8004.
34. Lieblein,A.L., Furtig,B. and Schwalbe,H. (2013) Optimizing the kinetics and thermodynamics of DNA i-motif folding. *ChemBioChem*, **14**, 1226–1230.
35. Manzini,G., Yathindra,N. and Xodo,L.E. (1994) Evidence for intramolecularly folded i-DNA structures in biologically relevant CCC-repeat sequences. *Nucleic Acids Res.*, **22**, 4634–4640.
36. Mergny,J.-L., Lacroix,L., Han,X., Leroy,J.-L. and Helene,C. (1995) Intramolecular folding of pyrimidine oligodeoxynucleotides into an i-DNA motif. *J. Am. Chem. Soc.*, **117**, 4797–4803.
37. Wright,E.P., Huppert,J.L. and Waller,Z.A.E. (2017) Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res.*, **45**, 2951–2959.
38. Cui,Y., Koirala,D., Kang,H., Dhakal,S., Yangyuoru,P., Hurley,L.H. and Mao,H. (2014) Molecular population dynamics of DNA structures in a bcl-2 promoter sequence is regulated by small molecules and the transcription factor hnRNP LL. *Nucleic Acids Res.*, **42**, 5755–5764.
39. 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.
40. Dzatko,S., Krafcikova,M., Hansel-Hertsch,R., Fessl,T., Fiala,R., Loja,T., Krafcik,D., Mergny,J.L., Foldynova-Trantirkova,S. and Trantirek,L. (2018) Evaluation of the stability of DNA i-Motifs in the nuclei of living mammalian cells. *Angew. Chem.-Int. Ed.*, **57**, 2165–2169.
41. Zeraati,M., Langley,D.B., Schofield,P., Moye,A.L., Rouet,R., Hughes,W.E., Bryan,T.M., Dinger,M.E. and Christ,D. (2018) I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.*, **10**, 631–637.
42. Gray,D.M., Hung,S.H. and Johnson,K.H. (1995) Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes. *Methods Enzymol.*, **246**, 19–34.
43. Mergny,J.L. and Lacroix,L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515–537.
44. Ross,S.A. and Burrows,C.J. (1996) Cytosine-specific chemical probing of DNA using bromide and monoperoxysulfate. *Nucleic Acids Res.*, **24**, 5062–5063.
45. Guo,K., Pourpak,A., Beetz-Rogers,K., Gokhale,V., Sun,D. and Hurley,L.H. (2007) Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. *J. Am. Chem. Soc.*, **129**, 10220–10228.
46. Del Villar-Guerra,R., Trent,J.O. and Chaires,J.B. (2018) G-Quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angew. Chem. (Int. Ed. Engl.)*, **57**, 7171–7175.
47. Fujii,T. and Sugimoto,N. (2015) Loop nucleotides impact the stability of intrastrand i-motif structures at neutral pH. *Phys. Chem. Chem. Phys.*, **17**, 16719–16722.
48. Dvorakova,Z., Renciuk,D., Kejnovska,I., Skolakova,P., Bednarova,K., Sagi,J. and Vorlickova,M. (2018) i-Motif of cytosine-rich human telomere DNA fragments containing natural base lesions. *Nucleic Acids Res.*, **46**, 1624–1634.
49. Lieblein,A.L., Buck,J., Schlepckow,K., Fürtig,B. and Schwalbe,H. (2012) Time-resolved NMR spectroscopic studies of DNA i-motif folding reveal kinetic partitioning. *Angew. Chem. Int. Ed. Engl.*, **51**, 250–253.
50. Rogers,R.A., Fleming,A.M. and Burrows,C.J. (2018) Unusual isothermal hysteresis in DNA i-Motif pH Transitions: a study of the RAD17 promoter sequence. *Biophys. J.*, **114**, 1804–1815.
51. Wu,S., Wang,X., Ye,X. and Zhang,G. (2013) pH-Induced conformational change and dimerization of DNA chains investigated by analytical ultracentrifugation. *J. Phys. Chem. B*, **117**, 11541–11547.