Article

pH-Dependent Capping Interactions Induce Large-Scale Structural Transitions in i-Motifs

Israel Serrano-Chacón, Bartomeu Mir, Lorenzo Cupellini, Francesco Colizzi, Modesto Orozco,* Núria Escaja,* and Carlos González*

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: We study here a DNA oligonucleotide having the ability to form two different i-motif structures whose relative stability depends on pH and temperature. The major species at neutral pH is stabilized by two C:C$^+$ base pairs capped by two minor groove G:C:G:C tetrads. The high pH and thermal stability of this structure are mainly due to the favorable effect of the minor groove tetrads on their adjacent positively charged C:C$^+$ base pairs. At pH 5, we observe a more elongated i-motif structure consisting of four C:C$^+$ base pairs capped by two G:T:G:T tetrads. Molecular dynamics calculations show that the conformational transition between the two structures is driven by the protonation state of key cytosines. In spite of large conformational differences, the transition between the acidic and neutral structures can occur without unfolding of the i-motif. These results represent the first case of a conformational switch between two different i-motif structures and illustrate the dramatic pH-dependent plasticity of this fascinating DNA motif.
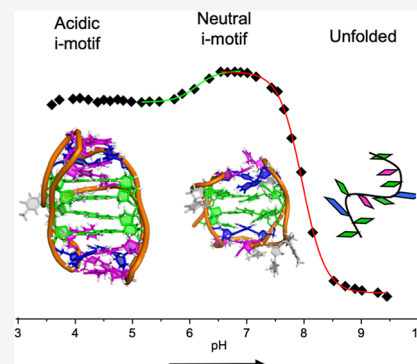
INTRODUCTION

Nucleic acids are dynamic and polymorphic molecules which can adopt a myriad of conformations in response to changes in the environment. Studying their conformational transitions is crucial to understand nucleic acids' biological activity and their potential applications in bio- and nanotechnology. In particular, conformational transitions involving i-motif structures are of special relevance due to their unique pH dependence. The i-motif is a four-stranded intercalated structure stabilized by the formation of hemiprotonated C:C$^+$ base pairs between parallel-oriented strands.[1−4] Since cytosine protonation is required for their formation, i-motif structures are usually observed at acidic pH. In spite of its preference for acidic environments, recent studies have shown that the i-motif is not an exotic structure generated only under special laboratory conditions but a conformation that can be formed in the cell. The growing evidence of their formation in vivo[5,6] and the increasing number of sequences being able to fold into stable i-motifs at neutral pH[7−11] is arousing great attention for this structure. Moreover, it has been shown that i-motif-forming sequences are common in the genome,[10,12] and numerous studies have described their potential role in processes like gene transcription,[13−15] DNA synthesis,[16] centromere[17] and telomere[18] maintenance, and so forth.

Besides, the strong dependency of i-motif stability on the pH makes it suitable for designing pH sensors[19,20] or other dynamic nanodevices.[21−23] In most cases, the conformational transitions involved in these potential devices are either i-motif-folding/unfolding transitions[19,20] or hybridization with the complementary strand.[24] However, the potential use of conformational switches between different i-motif structures has not been explored yet. Most studies on i-motif dynamics have focused on the folding/unfolding processes.[25−29] Although intrinsic conformational dynamics of folded i-motif structures has been studied by computational methods,[30,31] not many experimental studies have been performed, and most of them focused on the dynamics of loop regions.[18,32] To the best of our knowledge, transitions between different i-motif structures have not yet been studied experimentally in atomic detail.

A significant amount of work has been focused on understanding the basis of i-motif stabilization. Factors affecting its stability include cytosine tract length,[8] connecting loops,[11] and capping interactions at the sides of the C:C$^+$ stack.[33,34] Among the different capping interactions, minor groove tetrads (MGTs) are of particular interest since the interaction between MGTs and the terminal C:C$^+$ base pairs provides significant thermal and pH stabilization.[10,34] MGTs are the result of the association of two base pairs through their minor groove side and have been observed with different arrangements of Watson−Crick base pairs or mismatches (e.g., A:T:A:T,[35] A:G:A:G,[36] G:C:G:C,[35,37,38] G:C:G:T,[10] or G:T:G:T[39,40]). MGTs can be "direct" or "slipped", depending
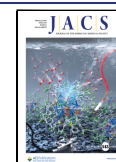
Table 1. Melting Temperatures (T_m) and pH$_T$ Values

name	sequence[a]	MGT	pH$_T$	T_m (°C)pH 7.0	T_m (°C)pH 6.0	T_m (°C)pH 5.0
LL4	d(L-T$_4$-L)	G:C:G:T	7.9	28.4	44.3	50.3
MM4	d(M-T$_4$-M)	G:T:G:T	6.7	16.6	29.5	38.0
NN4	d(N-T$_4$-N)	G:C:G:C	8.0	27.8	43.4[b]	53.3

[a]**L**: d(TCGTTCCGT); **M**: d(TCGTTTCGT); and **N**: d(CCGTTCCGT). [b]Apparent T_m corresponding to the melting of two coexisting species.

on the relative position of the two base pairs involved.[41] In the i-motif context, all the MGTs reported to date have been slipped, forming two H(N2)(G)−N3(G) hydrogen bonds between adjacent guanines.[10,34,39,40]

The MGT/C:C$^+$ interaction provides a dramatic stabilization to such an extent that i-motif forms even in sequences with very few cytosines, as those found in some repetitive sequences with no C-tracts.[10] These repetitive sequences fold into peculiar globular i-motif structures stabilized by two hemiprotonated C:C$^+$ base pairs, flanked by two MGTs. Interestingly, these structures can form in tandem and are prevalent in the human genome, being especially abundant in regulatory regions.[10]

The different base pair compositions of the MGTs may affect the conformational stability of these minimal i-motif structures. In particular, i-motifs stabilized by capping tetrads-containing G:C base pairs are very peculiar since they contain both neutral and protonated cytosines under the same experimental conditions. The simultaneous presence of C:C$^+$ and G:C base pairs raises the question on the impact of pH on the stability of these structures since the stabilizing effect of G:C-containing tetrads may be disrupted at acidic conditions.

By combining nuclear magnetic resonance (NMR) spectroscopy and theoretical calculations, we determined here two i-motif structures stabilized by MGTs at two different pH conditions and explored the structural determinants of their stability. Quite surprisingly, we found a dramatic but reversible conformational change triggered by pH changes, which happens without the unfolding of the structure. This unprecedented conformational transition widens the range of pH-dependent changes associated with i-motifs and can lead to a new paradigm for pH sensors. Furthermore, it opens the range of potential i-motif structures that can be stable under physiological conditions.

■ RESULTS

Repetitive sequences containing two repeats of d-(TCGTTCCGT) (**L**), d(TCGTTTCGT) (**M**), and d-(CCGTTCCGT) (**N**) have been studied. Previous studies indicate that these sequences fold into similar i-motif structures regardless of the number of residues connecting the repeats, although connectors of three or four nucleotides are optimal for their stability. For this reason, we focus on sequences containing two repeats connected by four thymines (Table 1).

Thermal Stability versus pH. UV-monitored thermal denaturation experiments were recorded at different pHs. As expected for structures belonging to the i-motif family, T_m values are higher at acidic pH (see Table 1). In all the conditions tested, **MM4** is less stable than **LL4** and **NN4**, in agreement with a higher stability of GC-containing tetrads versus pure G:T:G:T tetrads. However, **LL4** and **NN4** exhibit comparable T_m values at neutral pH, suggesting that G:C:G:C and G:C:G:T tetrads confer similar stability.

Changes in Circular Dichroism and NMR Spectra with pH. Circular dichroism (CD) and NMR spectra of **NN4** at

acidic pH differ from those obtained at neutral conditions, although they are in all cases consistent with i-motif folding (Figure 1a−d). Especially significant is the presence of imino signals around 15 ppm, clearly showing the formation of C:C$^+$ base pairs in the pH range from 5 to 7. At neutral conditions, fewer C:C$^+$ imino signals are observed, and several additional imino signals are detected between 11.5 and 13.5 ppm, most probably due to the formation of G:C base pairs. Interestingly, some of these G:C imino signals observed at neutral pH at 5 °C can also be observed at more acidic pH but only at higher temperatures. At acidic pH, signals around 10−11.5 ppm indicate the formation of G:T base pairs.

CD spectra are also pH-dependent. While at neutral pH, CD spectra exhibit a maximum band between 260 and 270 nm and a minimum band around 240 nm, at acidic conditions a second maximum is observed at around 295 nm. At basic pH, band intensities are reduced, and the CD spectrum contains a contribution from the statistical coil (see Figure 1d).

The midpoint of the pH denaturation transition (pH$_T$) was estimated by following the maximum at 270 nm of CD spectra versus pH (see Figures 1e and S1 and S2). In all cases, structures exhibit very high pH$_T$ values, being significantly higher for **LL4** and **NN4** than for **MM4**. This indicates that G:C-containing tetrads provide enhanced pH stability compared to G:T:G:T tetrads. The case of **NN4** is of particular relevance since the acid−base titration curve clearly exhibits two transitions with pH$_T$ values of 6.1 and 8.0 (Figure 1e).

CD and NMR data are consistent with the formation of two different structures for **NN4**; their relative populations are dependent on the pH and temperature but not on the presence of Na$^+$ or K$^+$ counterions (Figure S2). Observation of multiple exchangeable signals at high temperatures and intermediate pH (Figure 1b) indicates that the equilibrium between neutral and acidic species is slow on the NMR timescale.

NMR Assignment and Structural Calculation. To get insight into this peculiar behavior, we undertook the assignment of the NMR spectra of **NN4** at neutral and acidic conditions. Substitution of cytosines by 5-methyl-cytosines at certain positions was required for a complete assignment of the NMR spectra. Full details of the assignments are given in the Supporting Information (see Figures S3−S14).

As shown in Figure 2 (top), nuclear Overhauser enhancement spectroscopy (NOESY) spectra recorded at pH 7 show the coexistence of alternative i-motif species. One of the two imino signals from hemiprotonated C:C$^+$ base pairs (15.41 ppm) corresponds to two degenerated protons of the major species, whereas the signal at 15.66 ppm, which is less intense and shows fewer contacts, belongs to the minor species. We concluded that the hemiprotonated base pairs formed at this pH are C2:C20$^+$ and C7:C15$^+$, showing clear H41/H42C−H3$^+$C contacts. Amino protons of two cytosines (C6 and C19) exhibit cross-peaks with guanine imino protons (13.4 ppm), indicating the formation of two Watson−Crick G:C base pairs, which were assigned as G8:C19 and G21:C6. The two signals
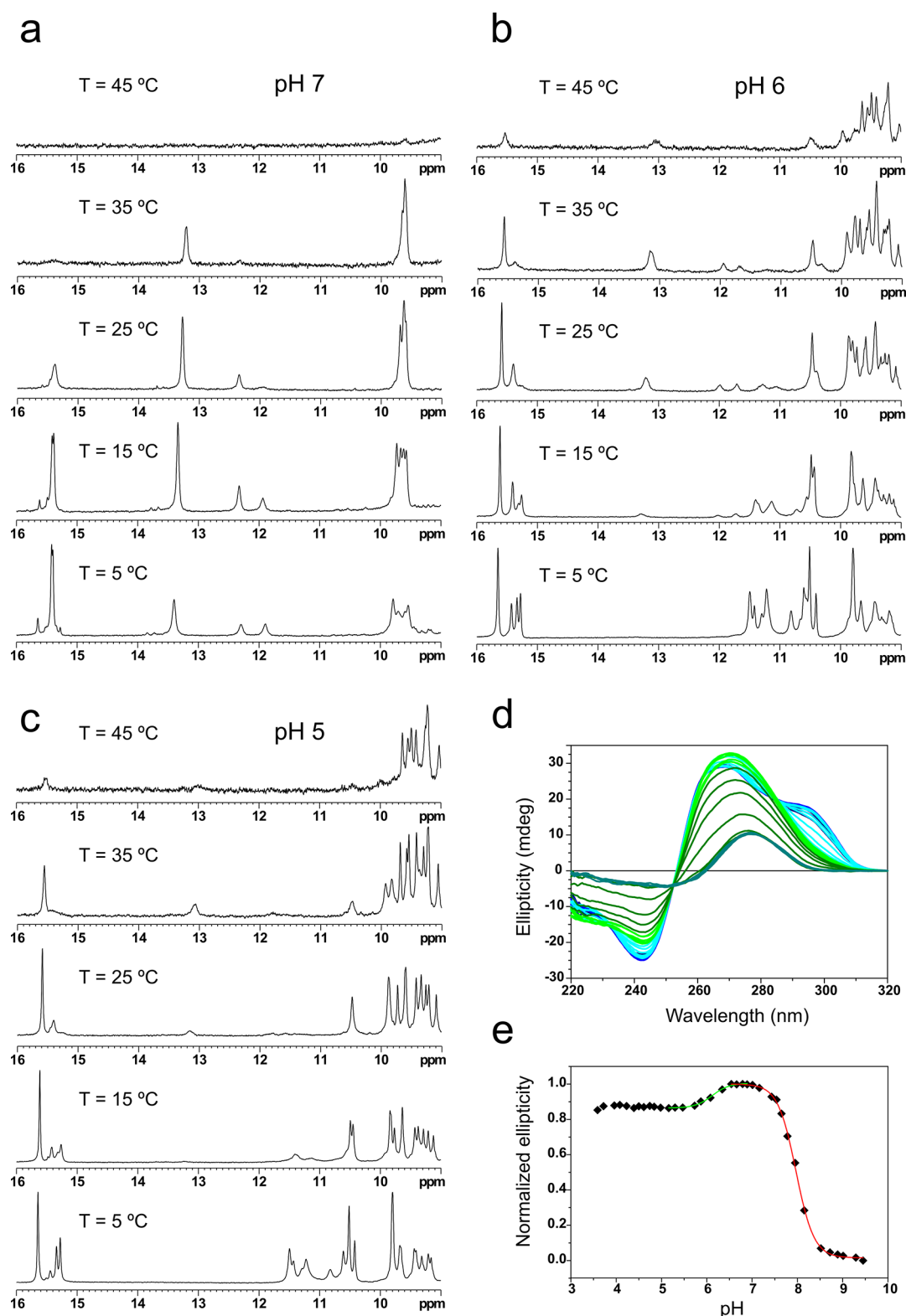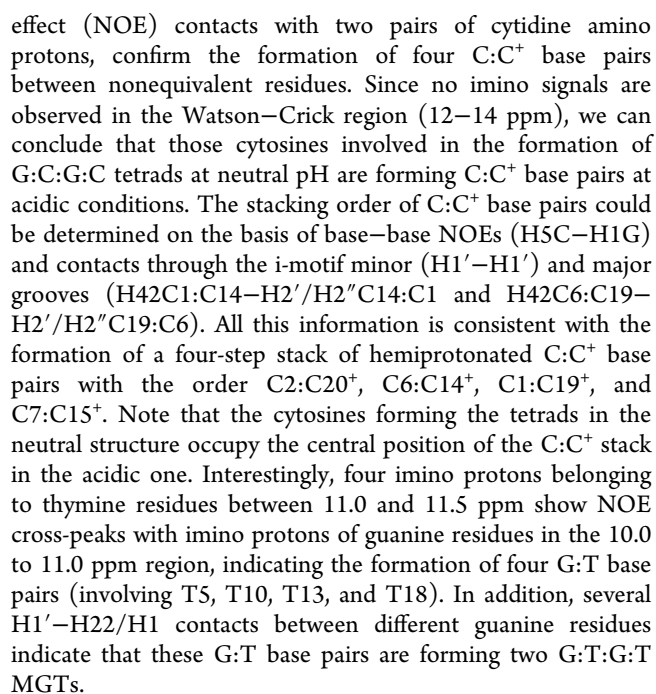
Figure 1. NMR spectra of **NN4** at different temperatures at pH 7 (a), 6 (b), and 5 (c). CD spectra at different pH values (from blue to green, acidic-to-basic pH) (d). Ellipticity versus pH profile (e); the two pH transitions are indicated with green (pH_T 6.1) and red (pH_T 8.0) lines. [DNA] = 500 μM (NMR) or 2 μM (CD).

around 12 ppm were assigned to the imino protons of G3 and G16 based on a number of cross-peaks with the neighboring C:C⁺ base pairs (H5C2−H1G3). Although the chemical shifts of these imino signals might be indicative of G:C base pairs, no cross-peaks with cytosine amino protons could be detected.

Spectra of **NN4** at pH 5 exhibit more dispersed signals than those recorded at pH 7 (Figure 2, bottom). The observed minor signal at 15.66 ppm at pH 7 is more intense at pH 5, displaying a high number of cross-peaks. A comparison of signal intensities in this region suggests the overlapping of two

Figure 2. Exchangeable proton regions of NOESY spectra (150 ms mixing time) of **NN4** at pH 7 (top) and pH 5 (bottom), sodium phosphate buffer, $T = 5$ °C, and [oligonucleotide] = 500 μM.

different hemiprotonated imino signals. Two extra H3+ signals at 15.29 and 15.34 ppm, which also show nuclear Overhauser

effect (NOE) contacts with two pairs of cytidine amino protons, confirm the formation of four C:C+ base pairs between nonequivalent residues. Since no imino signals are observed in the Watson–Crick region (12–14 ppm), we can conclude that those cytosines involved in the formation of G:C:G:C tetrads at neutral pH are forming C:C+ base pairs at acidic conditions. The stacking order of C:C+ base pairs could be determined on the basis of base–base NOEs (H5C–H1G) and contacts through the i-motif minor (H1'–H1') and major grooves (H42C1:C14–H2'/H2"C14:C1 and H42C6:C19–H2'/H2"C19:C6). All this information is consistent with the formation of a four-step stack of hemiprotonated C:C+ base pairs with the order C2:C20+, C6:C14+, C1:C19+, and C7:C15+. Note that the cytosines forming the tetrads in the neutral structure occupy the central position of the C:C+ stack in the acidic one. Interestingly, four imino protons belonging to thymine residues between 11.0 and 11.5 ppm show NOE cross-peaks with imino protons of guanine residues in the 10.0 to 11.0 ppm region, indicating the formation of four G:T base pairs (involving T5, T10, T13, and T18). In addition, several H1'–H22/H1 contacts between different guanine residues indicate that these G:T base pairs are forming two G:T:G:T MGTs.

The three-dimensional structures were calculated on the basis of NMR data by following the protocols described in the Experimental Methods section. A total of 189 and 151 interproton distance constraints were used in the structural calculation of the acidic and neutral structures, respectively. Statistical analysis of the final structures including the number of distance and angular constraints are shown in Table S5. In both cases, the structures are overall well defined with total residual distance violations of 1.64 and 2.16 Å, respectively. Structural ensembles are shown in Figures S15 and S16. Overall, both structures are well defined with root-mean-square deviations (RMSDs) below 1 Å excluding some flexible residues in the loops.

Description of the Structures. The neutral structure of **NN4** consists of two C:C+ base pairs surrounded by two minor groove G:C:G:C tetrads (Figure 3, top). In each tetrad, one of the two GC base pairs (C1:G16 and C14:G3) is distorted (see Figures S16 and S19), probably due to interactions with the axial loop. The upfield shift of G3 and G16 imino signals and the lack of NOEs with the base-paired cytosine amino protons are consistent with these distortions. In the two lateral loops, the first thymine is stacked on top of the MGT and the second one is mainly disordered. The four thymines in the long axial loop are also disordered. Interactions contributing to the stability of this structure are most probably very similar as in the case of **LL4**. The additional stabilization expected for the extra hydrogen bond in the G:C:G:C tetrad versus G:C:G:T is not observed. This is most probably due to the high distortion in one of the G:C pairs forming the G:C:G:C tetrad. All deoxyriboses adopt a south conformation, and glycosidic angles are in anticonformation, with guanine residues in a high-anti conformation (Table S7).

The structure of the predominant species at acidic pH is formed by four C:C+ base pairs surrounded by two minor groove G:T:G:T tetrads, involving T5, T10, T13, and T18 (Figure 3, bottom). The two lateral loops consist of only one thymine that lies on top of the neighbor tetrad. The C:C+ stack, formed by C7:C15, C1:C19, C6:C14, and C2:C20, is of the 3'E type. In this structure, the axial loop consists of only

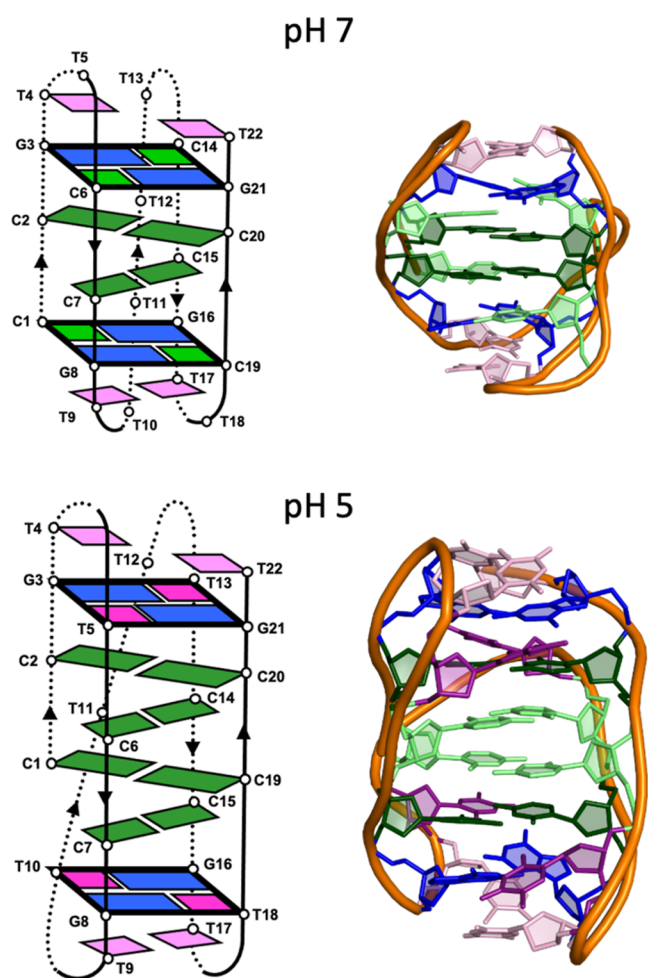Journal of the American Chemical Society
pubs.acs.org/JACS
Article

Figure 3. Schemes and average structures of **NN4** at neutral (top) and acidic pH (bottom). Color code: cytosines are in green, guanines are in blue, thymines involved in the tetrads are in magenta, and well-defined capping thymines are in pink. Cytosines involved in the G:C:G:C tetrad in the neutral structure are shown in lighter green. PDB codes: 8BV6 (top) and 8BQY (bottom).

Figure 4. (a) MGT; (b) interaction between a C:C$^+$ base pair and an MGT tetrad; and (c) calculated titration curves considering only the stack of two C:C$^+$ base pairs (blue) and the same stack plus the two MGTs (orange). The black crosses indicate the titration midpoints that define the pK_a.

two residues (T11 and T12), which are mainly disordered (see Figures S15 and S18).

Overall, the acidic structure is 5 Å more elongated than the neutral one (approximate dimensions are $23 \times 15 \times 8$ and $18 \times 15 \times 7$ Å for acidic and neutral structures, respectively). All sugars adopt a south conformation except for C1, C6, C14, and C19, which are in the general north domain. Glycosidic angles are all in anticonformation, with a tendency to adopt high-anti conformations (see Table S8).

MGTs Stabilize C:C$^+$ Pairs. To get more insight into the fundamental bases behind the favorable interaction between MGTs and C:C$^+$ base pairs, we performed theoretical titration experiments using Poisson−Boltzmann (PB) calculations with our CMIP program[42] (see the Experimental Section). Calculations were carried out considering the two cytosine pairs stacked with and without the capping tetrads. Results in Figure 4 clearly show a dramatic shift in pK_a, provoked by the tetrads. The first calculated pK_a matches the experimentally observed midpoint of the pH denaturation transition (pH$_T$) with remarkable precision. In this case, the second deprotonation event (predicted at ∼12.5) has no physical meaning since the structure is unfolded at that pH. We can conclude that the stabilization effect of MGTs in i-motifs is
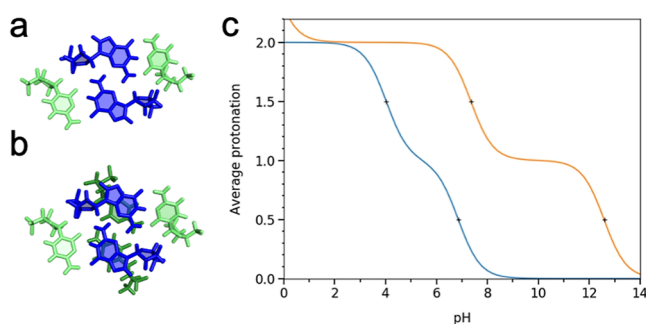
due to the ability of the stacked MGT to stabilize the adjacent positively charged cytosines.

Exploring the Conformational Change Pathway. To get further insight into the conformational transition between the two structures, we used classical molecular dynamics (MD) methods. As the timescale of conformational change is substantially longer than the timescales accessible to modern computer hardware, we used ratchet-and-pawl MD (rMD)[43] to enhance the sampling efficiency and trace the transition. Accordingly, the system is left to freely evolve along the "advance" coordinate according to thermal fluctuations, penalizing those movements that involve backward evolution along a given guess (hypothetical) reaction coordinate (information bias). In our case, the reaction coordinate is defined by the eRMSD metric taking a value of 0 when **NN4** adopts its neutral conformation (Figure 3, top). Fifty independent rMD trajectories were generated from the coordinates of the acidic structure (S in Figure 5c) but making cytosines C1, C6, C14, and C19 deprotonated as occurring at pH 7 (Figure 3, top).

Among the 50 rMD trial trajectories, 45 reached the target state within 20 ns (and the others were approaching), suggesting an easy transition that is confirmed by the free energy landscape along two intuitive coordinates: the number of stacked cytosines and the distance between the two MGTs (Figure 5a). This demonstrates that the E-state is more stable than the extended one, S. The transition occurs through a series of stable quasidegenerated intermediates which smoothly connect the S- and E-states (see Figure 5c). The first of these states corresponds to the flipping of C14 outside the stack, while the most populated intermediate (4) is a structure where only one of the deprotonated cytosines remains in the stack. Note that once the last cytosine flips, the system relaxes quickly toward the final state with a further convergence between the two tetrads. To test whether the conformational transition seen in the rMD trajectory is real and not forced by the bias, we performed a new set of 50 rMD simulations, starting from the same structure, but this time with the cytosines protonated as occurring at pH 5. These simulations, where all the hemiprotonated cytosine pairs are preserved, cannot reach the target structure (Figure 5b), strongly suggesting that there is no simulation artifact.

The above results suggest that a conformational transition from the acidic to the neutral structure can occur through flipping of four deprotonated cytosine bases and shortening of the distance between the MGTs. Remarkably, the hydrogen
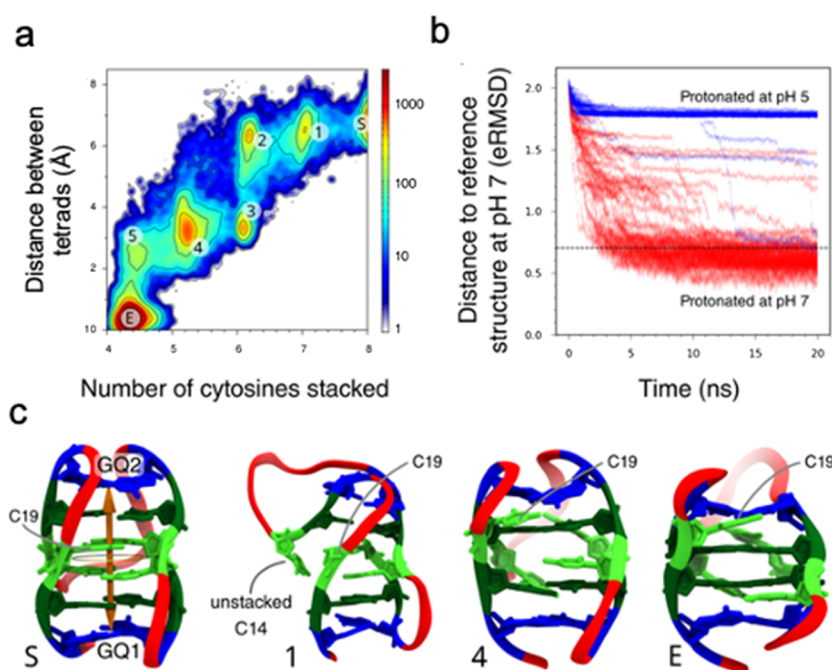
Figure 5. Enhanced MD results on the S (stretched, low pH) to E (exiguous, neutral pH) conformational transition. (a) Density plot of the 50 rMD trajectories (for the cytosines protonated as expected at pH 7) showing the distance between the two planes of guanines versus the number of cytosines within the stack. The color bar represents population density. The labels represent the starting (S) and end (E) states and different intermediates in the transition. (b) Distance (eRMSD) from the target structure (the neutral one) for all the rMD trajectories versus simulation time (red lines). The same data obtained from the acidic structure, in which no convergence to pH 7 structure is achieved, are also plotted. The dashed horizontal line is the eRMSD threshold value taken to be a successful transition. (c) Representation of the starting (acidic) structure, two intermediate states recognized in the density plot (a), and the final (neutral) state. Cytosines are represented in two green tones: light green for those cytosines that are neutral (C1, C6, C14, and C19) and dark green for those involved in the hemiprotonated base pairs in the neutral structure (C2, C7, C15, and C20); guanines are in blue and thymines are in red; thymine nucleobases are omitted for clarity. The double arrow represents the distance between guanine planes, whereas the ring represents the area in which cytosines are considered within the stack.

bonds between hemiprotonated cytosine pairs (C2:C20$^+$ and C7:C15$^+$) remain stable in all the rMD simulations. Illustrative examples of these rMD trajectories are shown in the Supporting Information movie. These results show that a conformational transition between the two structures can proceed without unfolding/refolding of the entire i-motif structure.

Conformational Changes in Longer Oligonucleotides. Previous studies have shown that sequences with an increasing number of **L** repeats in tandem can fold independently adopting bead-like superstructures. Consequently, we tested whether sequences based on more than two **N** repeats can adopt similar structures and if they undergo similar transitions. An oligonucleotide consisting of four **N** repeats was studied by NMR at different pH conditions. Despite having much broader lines, the main spectral features observed for **NN4** are reproduced (Figure S20). We conclude that this sequence forms i-motif domains in tandem and that each of these domains experiences the same conformational transition between acidic and neutral pH, as illustrated in Figure S21.

■ DISCUSSION

The effect of MGTs as stabilizing capping elements in i-motif structures is now well established. However, not all possible MGTs confer the same stabilization. As described here, G:C-containing tetrads are more stabilizing than pure G:T:G:T tetrads at neutral pH. This is most probably due to the additional hydrogen bond in Watson−Crick G:C versus G:T base pairs. In the case of **NN4**, the effect is mitigated because the G:C:G:C tetrads are partially distorted; one base pair is Watson−Crick, and the other is a wobble G:C with only two H-bonds.

MGTs induce a dramatic shift in the pK_a of the cytosines involved in the neighboring C:C$^+$ base pairs. Most probably, this effect is independent of the specific tetrads since all slipped MGTs have a similar geometry. In particular, the purines in the center of the tetrad are always located in the same position on top of the positively charged H3 proton, suggesting a strong PI−cation interaction.

At more acidic pH, cytosine protonation makes G:C:G:C tetrads unstable. However, G:C:G:C tetrads' disruption does not provoke a denaturation of the compact structure but a large-scale conformational change toward an elongated structure stabilized by minor groove G:T:G:T tetrads. Its stability arises from the formation of two additional C:C$^+$ base pairs, which compensate for the lower number of hydrogen bonds in G:T:G:T versus G:C:G:C tetrads. Due to the longer distance, the stabilizing effect of the tetrads on the central C:C$^+$ base pairs is probably weakened. This explains the lower pH stability of this structure, with a pH$_T$ between the acidic and neutral structures of 6.1 at 5 °C (Figure 1).

Interestingly, upon increasing the temperature, the neutral i-motif can be observed at a more acidic pH, as shown in the NMR spectra of Figure 1b,c. This observation reflects the lower cytosine protonation pK_a at higher temperatures and strongly suggests that the relative stability of the neutral and

acidic i-motif structures is determined by their cytosines' protonation states.

Not only the equilibria but also the transition between the two structures are entirely driven by the cytosines' protonation state, as shown by our ratchet-and-pawl MD calculations. The critical role of the cytosine protonation states in i-motif dynamics concurs with recent theoretical studies in other i-motif structures.[30] In our case, the conformational transition between the two structures involves important nucleobase rearrangements, with some residues showing dramatic rearrangements. For example, cytosines in the capping tetrads in the neutral structure occupy central positions in the i-motif core of the acidic structure. The fact that such a large conformational change may occur without the unfolding of the structure has implications for our understanding of the i-motif folding process since it indicates that, at least under certain conditions, additional C:C$^+$ base pairs can be inserted in the middle of a growing C:C$^+$ stack and not only in their terminal positions.

It is worth mentioning that related pH-induced conformational transitions have been observed in DNA crystals based on CGAA sequences which contain antiparallel- and parallel-stranded regions.[44] At the junctions between the two secondary structures, some cytosines form either C:C$^+$ base pairs at acidic pH or interhelical minor groove G:C:G:C tetrads at neutral conditions. Interestingly, the two variants can interconvert *in crystallo* in response to pH perturbations, giving rise to different 3D crystal lattices. These results suggest the possibility of designing dynamic biomaterial assemblies based on changes in cytosines' protonation states.

■ CONCLUSIONS

We have shown here that i-motifs may have complex dynamics. The formation of i-motif structures is not just an ON/OFF process induced by pH. Instead, some sequences can access different i-motif folds in a controlled way. Understanding these phenomena is important for applications of the i-motif in nano- and biotechnology and its potential role in the cell.

■ EXPERIMENTAL SECTION

Oligonucleotide Synthesis. Oligodeoxynucleotides were purchased at IDT with standard desalting purification. Samples were dissolved in the Na$^+$ form.

CD and UV Spectroscopy. CD spectra were recorded on a Jasco J-815 spectropolarimeter. UV spectra were recorded on a Jasco V-730 spectrophotometer. Both were fitted with Peltiers. Spectra were recorded in 25 mM sodium phosphate buffer at different pH values. Samples were initially heated at 90 °C for 5 min and slowly allowed to cool to room temperature and stored at 4 °C until use. UV melting curves were recorded at the wavelength of the larger positive band, ∼265 nm, with a heating rate of 0.5 °C·min^{-1}. Uncertainties in T_m values are estimated to be ±0.2 °C.

NMR Spectroscopy. Samples for NMR experiments were dissolved (in Na$^+$ form) in either D$_2$O or 9:1 H$_2$O/D$_2$O and 25 mM sodium phosphate buffer. The pH was adjusted by adding aliquots of concentrated HCl or NaOH. All NMR spectra were acquired on Bruker Neo Avance spectrometers operating at 600 and 800 MHz equipped with cryoprobes and processed with the TOPSPIN software. 2D NMR experiments for spectral assignment and acquirement of experimental constraints were recorded at $T = 5$ °C. A jump-and-return pulse sequence was employed to observe the rapidly exchanging protons in 1D H$_2$O experiments. NOESY spectra in D$_2$O and 9:1 H$_2$O/D$_2$O were acquired with mixing times of 150 and 250 ms. TOCSY (total correlation spectroscopy) spectra were recorded with the standard MLEV-17 spin-lock sequence and a

mixing time of 80 ms. In most of the experiments using H$_2$O, water suppression was achieved by including a WATERGATE module in the pulse sequence prior to acquisition. The spectral analysis program SPARKY was used for semiautomatic assignment of the NOESY cross-peaks and evaluation of the NOE intensities.

NMR Constraints. Qualitative distance constraints were obtained from NOE intensities. NOEs were classified as strong, medium, or weak, and distance constraints were set accordingly to 3, 4, or 5 Å. In addition to these experimentally derived constraints, hydrogen bond constraints for the base pairs were used. Target values for distances and angles related to hydrogen bonds were set to values obtained from crystallographic data in related structures. Force constants were 20 kcal/mol·Å^2 for experimental distance constraints and 30 kcal/mol·Å^2 for hydrogen bond distance constraints. Due to the relatively broad line widths of the sugar proton signals, J-coupling constants were not accurately measured but roughly estimated from H1′−H2′ and H1′−H2″ DQF-COSY (double quantum-filtered correlation spectroscopy) cross-peaks. When these cross-peaks were consistent with deoxyribose conformations in south or south/east, sugar dihedral angles were constrained to the following target intervals: $\nu0(−36.5, −6.5°)$; $\nu1(19.8, 49.8°)$; $\nu2(−49.8, −19.8°)$; $\nu3(6.5, 36.5°)$; and $\nu4(−15.0, 15.0°)$, with a force constant of 25 kcal/mol·rad. This is equivalent to loosely constraining the sugar pseudorotation phase angles (Ps) between 144 and 180°.

Structural Calculations. Structures were calculated with the program CYANA 3.0[45] and further refined with the SANDER module of the MD package AMBER 18.[46] The resulting CYANA structures were taken as starting points for the AMBER refinement, consisting of an annealing protocol in water, followed by trajectories of 500 ps each in which explicit solvent molecules were included and using the particle mesh Ewald method to evaluate long-range electrostatic interactions. Specific protocols for these calculations have been described elsewhere. The BSC1 force field[47] was used to describe the DNA, and the TIP3P model was used to simulate water molecules. Hemiprotonated C:C$^+$ base pairs were modeled by considering base pairs between neutral and protonated cytosine residues obtained from libraries included in the AMBER package. Analysis of the final structures was carried out with the programs MOLMOL[48] and X3DNA.[49] The conformation with the lowest AMBER total energy is taken as the best structure. Coordinates were deposited in the PDB data bank (codes 8BV6 and 8BQY for neutral and acidic structures, respectively).

MD Simulations. Classical MD simulations were initially used to probe the conformational change from the pH 5 structure to the pH 7 structure. As the timescale of conformational change is substantially longer than those accessible even to modern computer hardware, we resorted to an enhanced sampling approach, namely, ratchet-and-pawl MD (rMD).[43] All rMD simulations were carried out with GROMACS[50] patched with PLUMED[51,52] using the BSC1 force field.[47] In rMD, a hypothetical, plausible reaction coordinate is chosen, and the system is left free to evolve along such a reaction coordinate according to the room-temperature thermal fluctuations. Conversely, a bias potential is applied whenever the system evolves backward along the reaction coordinate. In this way, the evolution of the system is dictated only by spontaneous thermal motion, and any motions that drive the system far from the target structure are discouraged.

As a reaction coordinate, we chose the distance connecting the structure at pH 5 to the structure of the i-motif at pH 7. This distance is measured in terms of the eRMSD metric, which has been proposed for nucleic acids as a more effective alternative to the common RMSD metric in discriminating different configurations.[53] The eRMSD is able to finely distinguish different stacking and H-bonding patterns between nucleobases and is a vectorial analog of the contact maps commonly used in proteins.[54]

After thermalization and equilibration of the pH 5 structure (ensemble **S** in Figure 5c), 50 independent rMD trajectories were generated to investigate the intermediate states required to reach the NMR structure resolved at pH 7. Since deprotonation will be much faster than the conformational change, we deprotonated the cytosine

pairs C14:C6 and C1:C19 for these calculations. rMD simulations were carried out with an adiabatic force constant of 200 kcal/mol per eRMSD unit and a target value of zero (i.e., eRMSD from the structure at pH 7 = 0). The eRMSD was computed excluding the thymine bases, except for those forming tetrads in the pH 5 starting structure.

Theoretical pH Titration Calculations. Titration curves (Figure 4c) were calculated excluding the highly mobile thymine loop. The intrinsic pK_a (pK_{int}) of each cytosine was computed from electrostatic PB calculations of the electrostatic interaction with the nontitratable residues and of the desolvation energy. The pK_a of cytosine in water was assumed to be 4.4. The interaction term g_{ij} between cytosines C_i and C_j was calculated by computing the energies of the pairs $C_iH^+:C_j$, $C_iH^+:C_j$, $C_i:C_j$ H^+, and $C_i:C_j$. For PB calculations, the internal and external dielectric were set to 6 and 80, respectively.

The titration curve of each cytosine C_i can be computed by defining its average protonation $\langle x_i \rangle$ at each value of the pH. The latter can be calculated from the Boltzmann average

$$\langle x_i \rangle = \frac{\sum_{(X)} x_i^{(X)} e^{-G_X/kT}}{\sum_{(X)} e^{-G_X/kT}}$$

where X is a protonation state vector, whose elements x_i are either 1 or 0, depending on the protonation of site i. Each of these vectors represents a microstate with free energy G_X. Given the low number of cytosines, the protonation state vectors can be all enumerated, and the sums in the equation can be computed exactly. The Gibbs free energy G_X can be calculated at any pH as

$$G_X = \sum_i x_i \ln(10)\left(pK_{f,i} - pH\right) - \frac{1}{2}\sum_{ij} x_i x_j g_{ij}$$

where $pK_{int,i}$ is the aforementioned intrinsic pK_a of site i and g_{ij} is the interaction term between sites i and j.

The titration curve of the entire i-motif is obtained by summing all $\langle x_i \rangle$. The macroscopic pK_a values of the i-motif are defined here as the midpoint of the titration curves. Unless otherwise specified, we report the pK_a value for the second protonation, given that there are two C:C stacks in the structure at pH 7.

The effect of different residues on the macroscopic pK_a was calculated by computing the titration curve after either removing the whole residue from the calculation or setting its charges to zero. In addition, we also removed the charges of the nucleobases only.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.2c13043.

> Detailed descriptions of NMR assignments; assignment, calculation statistics, and structural analysis; CD and NMR data; and details on the structural models (PDF)
>
> Transition between the acidic and neutral structures (MP4)

AUTHOR INFORMATION

Corresponding Authors

Modesto Orozco — *Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain; Departament de Bioquímica i Biomedicina. Facultat de Biologia, Universitat de Barcelona, 08028 Barcelona, Spain;* ⓘ orcid.org/0000-0002-8608-3278; Email: modesto.orozco@irbbarcelona.org

Núria Escaja — *Inorganic and Organic Chemistry Department, Organic Chemistry Section, and IBUB, University of Barcelona, 08028 Barcelona, Spain; BIOESTRAN Associated Unit UB-CSIC, 08028 Barcelona, Spain;* Email: nescaja@ub.edu

Carlos González — *Instituto de Química Física "Rocasolano", CSIC, 28006 Madrid, Spain; BIOESTRAN Associated Unit UB-CSIC, 08028 Barcelona, Spain;* ⓘ orcid.org/0000-0001-8796-1282; Email: cgonzalez@iqfr.csic.es

Authors

Israel Serrano-Chacón — *Instituto de Química Física "Rocasolano", CSIC, 28006 Madrid, Spain; Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain*

Bartomeu Mir — *Instituto de Química Física "Rocasolano", CSIC, 28006 Madrid, Spain; Inorganic and Organic Chemistry Department, Organic Chemistry Section, and IBUB, University of Barcelona, 08028 Barcelona, Spain*

Lorenzo Cupellini — *Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain;* ⓘ orcid.org/0000-0003-0848-2908

Francesco Colizzi — *Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain; Present Address: Institute of Marine Sciences, ICM-CSIC, Barcelona, 08003, Spain;* ⓘ orcid.org/0000-0001-5601-1452

Complete contact information is available at:
https://pubs.acs.org/10.1021/jacs.2c13043

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval for the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

MGT   minor groove tetrad
rMD   ratchet-and-pawl molecular dynamics

■ REFERENCES

(1) Gehring, K.; Leroy, J. L.; Guéron, M. A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* 1993, 363, 561−565.

(2) Benabou, S.; Aviñó, A.; Eritja, R.; González, C.; Gargallo, R. Fundamental Aspects of the Nucleic Acid I-Motif Structures. *RSC Adv.* 2014, 4, 26956−26980.

(3) Abou Assi, H.; Garavís, M.; González, C.; Damha, M. J. I-Motif DNA: Structural Features and Significance to Cell Biology. *Nucleic Acids Res.* 2018, 46, 8038−8056.

(4) Day, H. A.; Pavlou, P.; Waller, Z. A. E. I-Motif DNA: Structure, Stability and Targeting with Ligands. *Bioorg. Med. Chem.* 2014, 22, 4407−4418.

(5) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. I-Motif DNA Structures Are Formed in the Nuclei of Human Cells. *Nat. Chem.* 2018, 10, 631−637.

(6) Dzatko, S.; Krafcikova, M.; Hänsel-Hertsch, R.; Fessl, T.; Fiala, R.; Loja, T.; Krafcik, D.; Mergny, J.-L. L.; Foldynova-Trantirkova, S.; Trantirek, L. Evaluation of the Stability of DNA i-Motifs in the Nuclei of Living Mammalian Cells. *Angew. Chem.* 2018, 57, 2165−2169.

(7) Brazier, J. A.; Shah, A.; Brown, G. D. I-Motif Formation in Gene Promoters: Unusually Stable Formation in Sequences Complementary to Known G-Quadruplexes. *Chem. Commun.* 2012, 48, 10739.

(8) Fleming, A. M.; Ding, Y.; Rogers, R. A.; Zhu, J.; Zhu, J.; Burton, A. D.; Carlisle, C. B.; Burrows, C. J. 4n-1 Is a "Sweet Spot" in DNA i-Motif Folding of 2′-Deoxycytidine Homopolymers. *J. Am. Chem. Soc.* 2017, 139, 4682−4689.

(9) Wright, E. P.; Huppert, J. L.; Waller, Z. A. E. Identification of Multiple Genomic DNA Sequences Which Form I-Motif Structures at Neutral pH. *Nucleic Acids Res.* 2017, 45, 2951−2959.

(10) Mir, B.; Serrano, I.; Buitrago, D.; Orozco, M.; Escaja, N.; González, C. Prevalent Sequences in the Human Genome Can Form Mini I-Motif Structures at Physiological pH. *J. Am. Chem. Soc.* 2017, 139, 13985−13988.

(11) Cheng, M.; Qiu, D.; Tamon, L.; Ištvánková, E.; Víšková, P.; Amrane, S.; Guédin, A.; Chen, J.; Lacroix, L.; Ju, H.; Trantírek, L.; Sahakyan, A. B.; Zhou, J.; Mergny, J.-L. Thermal and pH Stabilities of I-DNA: Confronting in Vitro Experiments with Models and In-Cell NMR Data. *Angew. Chem., Int. Ed.* 2021, 60, 10286−10294.

(12) Belmonte-Reche, E.; Morales, J. C. G4-IM Grinder: When Size and Frequency Matter. G-Quadruplex, i-Motif and Higher Order Structure Search and Analysis Tool. *NAR: Genomics Bioinf.* 2020, 2, lqz005.

(13) Kang, H. J.; Kendrick, S.; Hecht, S. M.; Hurley, L. H. 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.* 2014, 136, 4172−4185.

(14) Kendrick, S.; Kang, H. J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D.; Hecht, S. M.; Hurley, L. H. 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.* 2014, 136, 4161−4171.

(15) Kaiser, C. E.; Van Ert, N. A.; Agrawal, P.; Chawla, R.; Yang, D.; Hurley, L. H. Insight into the Complexity of the I-Motif and G-Quadruplex DNA Structures Formed in the KRAS Promoter and Subsequent Drug-Induced Gene Repression. *J. Am. Chem. Soc.* 2017, 139, 8522−8536.

(16) Takahashi, S.; Brazier, J. A.; Sugimoto, N. Topological Impact of Noncanonical DNA Structures on Klenow Fragment of DNA Polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 2017, 114, 9605−9610.

(17) Garavís, M.; Escaja, N.; Gabelica, V.; Villasante, A.; González, C. Centromeric Alpha-Satellite DNA Adopts Dimeric i-Motif Structures Capped by at Hoogsteen Base Pairs. *Chem.—Eur. J.* 2015, 21, 9816−9824.

(18) Phan, A. T.; Guéron, M.; Leroy, J. L. The Solution Structure and Internal Motions of a Fragment of the Cytidine-Rich Strand of the Human Telomere. *J. Mol. Biol.* 2000, 299, 123−144.

(19) Nesterova, I. V.; Nesterov, E. E. Rational Design of Highly Responsive pH Sensors Based on DNA I-Motif. *J. Am. Chem. Soc.* 2014, 136, 8843−8846.

(20) Dembska, A.; Bielecka, P.; Juskowiak, B. pH-Sensing Fluorescence Oligonucleotide Probes Based on an i-Motif Scaffold: A Review. *Anal. Methods* 2017, 9, 6092−6106.

(21) Dong, Y.; Yang, Z.; Liu, D. DNA Nanotechnology Based on I-Motif Structures. *Acc. Chem. Res.* 2014, 47, 1853−1860.

(22) Yatsunyk, L. A.; Mendoza, O.; Mergny, J.-L. "Nano-Oddities": Unusual Nucleic Acid Assemblies for DNA-Based Nanostructures and Nanodevices. *Acc. Chem. Res.* 2014, 47, 1836−1844.

(23) Alba, J. J.; Sadurní, A.; Gargallo, R. Nucleic Acid i- Motif Structures in Analytical Chemistry. *Crit. Rev. Anal. Chem.* 2016, 46, 443−454.

(24) Bielecka, P.; Dembska, A.; Juskowiak, B. Monitoring of PH Using an I-Motif-Forming Sequence Containing a Fluorescent Cytosine Analogue, TC. *Molecules* 2019, 24, 952.

(25) Lieblein, A. L.; Buck, J.; Schlepckow, K.; Fürtig, B.; Schwalbe, H. Time-Resolved NMR Spectroscopic Studies of DNA i-Motif Folding Reveal Kinetic Partitioning. *Angew. Chem., Int. Ed.* 2012, 51, 250−253.

(26) Choi, J.; Kim, S.; Tachikawa, T.; Fujitsuka, M.; Majima, T. PH-Induced Intramolecular Folding Dynamics of i-Motif DNA. *J. Am. Chem. Soc.* 2011, 133, 16146−16153.

(27) Smiatek, J.; Chen, C.; Liu, D.; Heuer, A. Stable Conformations of a Single Stranded Deprotonated DNA I-Motif. *J. Phys. Chem. B* 2011, 115, 13788−13795.

(28) Smiatek, J.; Heuer, A. Deprotonation Mechanism of a Single-Stranded DNA i-Motif. *RSC Adv.* 2014, 4, 17110−17113.

(29) Liu, L.; Kim, B. G.; Feroze, U.; Macgregor, R. B.; Chalikian, T. V. Probing the Ionic Atmosphere and Hydration of the C-MYC i -Motif. *J. Am. Chem. Soc.* 2018, 140, 2229−2238.

(30) Mondal, M.; Gao, Y. Q. Microscopic Insight into pH-Dependent Conformational Dynamics and Noncanonical Base Pairing in Telomeric i-Motif DNA. *J. Phys. Chem. Lett.* 2022, 13, 5109−5115.

(31) Mondal, M.; Bhattacharyya, D.; Gao, Y. Q. Structural Properties and Influence of Solvent on the Stability of Telomeric Four-Stranded i-Motif DNA. *Phys. Chem. Chem. Phys.* 2019, 21, 21549−21560.

(32) Lieblein, A. L.; Fürtig, B.; Schwalbe, H. Optimizing the Kinetics and Thermodynamics of DNA I-Motif Folding. *ChemBioChem* 2013, 14, 1226−1230.

(33) Benabou, S.; Garavís, M.; Lyonnais, S.; Eritja, R.; González, C.; Gargallo, R. Understanding the Effect of the Nature of the Nucleobase in the Loops on the Stability of the I-Motif Structure. *Phys. Chem. Chem. Phys.* 2016, 18, 7997−8004.

(34) Escaja, N.; Viladoms, J.; Garavís, M.; Villasante, A.; Pedroso, E.; González, C. A Minimal I-Motif Stabilized by Minor Groove G:T:G:T Tetrads. *Nucleic Acids Res.* 2012, 40, 11737−11747.

(35) Escaja, N.; Pedroso, E.; Rico, M.; González, C. Dimeric Solution Structure of Two Cyclic Octamers: Four-Stranded DNA Structures Stabilized by A:T:A:T and G:C:G:C Tetrads. *J. Am. Chem. Soc.* 2000, 122, 12732−12742.

(36) Chu, B.; Zhang, D.; Hwang, W.; Paukstelis, P. J. Crystal Structure of a Tetrameric DNA Fold-Back Quadruplex. *J. Am. Chem. Soc.* 2018, 140, 16291−16298.

(37) Escaja, N.; Gómez-Pinto, I.; Pedroso, E.; González, C. Four-Stranded DNA Structures Can Be Stabilized by Two Different Types of Minor Groove G:C:G:C Tetrads. *J. Am. Chem. Soc.* 2007, 129, 2004−2014.

(38) Kocman, V.; Plavec, J. A Tetrahelical DNA Fold Adopted by Tandem Repeats of Alternating GGG and GCG Tracts. *Nat. Commun.* 2014, 5, 5831.

(39) Gallego, J.; Chou, S. H.; Reid, B. R. Centromeric Pyrimidine Strands Fold into an Intercalated Motif by Forming a Double Hairpin

with a Novel T:G:G:T Tetrad: Solution Structure of the d(TCCCGTTTCCA) Dimer. *J. Mol. Biol.* **1997**, *273*, 840−856.

(40) Serrano-Chacón, I.; Mir, B.; Escaja, N.; González, C. Structure of I-Motif/Duplex Junctions at Neutral pH. *J. Am. Chem. Soc.* **2021**, *143*, 12919−12923.

(41) Escaja, N.; Mir, B.; Garavís, M.; González, C. Non-G Base Tetrads. *Molecules* **2022**, *27*, 5287.

(42) Gelpí, J. L.; Kalko, S.; Barril, X.; Cirera, X.; de La Cruz, J.; Luque, F. J.; Orozco, M. Classical Molecular Interaction Potentials: An Improved Set-up Procedure in Molecular Dynamics Simulations of Proteins. *Proteins* **2001**, *45*, 428−437.

(43) Rueda, M.; Cubero, E.; Laughton, C. A.; Orozco, M. Exploring the Counterion Atmosphere around DNA: What Can Be Learned from Molecular Dynamics Simulations? *Biophys. J.* **2004**, *87*, 800−811.

(44) Muser, S. E.; Paukstelis, P. J. Three-Dimensional DNA Crystals with pH-Responsive Noncanonical Junctions. *J. Am. Chem. Soc.* **2012**, *134*, 12557−12564.

(45) Güntert, P. Automated NMR Structure Calculation with CYANA. *Methods Mol. Biol.* **2004**, *278*, 353−378.

(46) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A.*AMBER 18*, 2018.

(47) Ivani, I.; Dans, P. D. P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L. J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A. S. A.; Case, D. A. D. A.; Orozco, M. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2015**, *13*, 55−58.

(48) Koradi, R.; Billeter, M.; Wüthrich, K. MOLMOL: A Program for Display and Analysis of Macromolecular Structures. *J. Mol. Graphics* **1996**, *14*, 51−55.

(49) Lu, X. J.; Olson, W. K. 3DNA: A Versatile, Integrated Software System for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic-Acid Structures. *Nat. Protoc.* **2008**, *3*, 1213−1227.

(50) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E.Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In *Solving Software Challenges for Exascale*; Markidis, S., Laure, E., Eds.; Springer International Publishing: Cham, 2015; pp 3−27.

(51) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.

(52) The PLUMED consortium; Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods***2019**, *16*, 670−673.

(53) Bottaro, S.; Di Palma, F.; Bussi, G. The Role of Nucleobase Interactions in RNA Structure and Dynamics. *Nucleic Acids Res.* **2014**, *42*, 13306−13314.

(54) Vendruscolo, M.; Najmanovich, R.; Domany, E. Protein Folding in Contact Map Space. *Phys. Rev. Lett.* **1999**, *82*, 656.