

Structure of i-Motif/Duplex Junctions at Neutral pH

Israel Serrano-Chacón, Bartomeu Mir, Núria Escaja,* and Carlos González*



Cite This: *J. Am. Chem. Soc.* 2021, 143, 12919–12923



Read Online

ACCESS |



Metrics & More



Article Recommendations



Supporting Information

ABSTRACT: We report here the three-dimensional structure of an i-motif/duplex junction, determined by NMR methods at neutral pH. By including a minor groove tetrad at one side of the C:C⁺ stack of a monomeric i-motif, and a stem/loop hairpin at the other side, we have designed stable DNA constructs in which i-DNA and B-DNA regions coexist in a wide range of experimental conditions. This study demonstrates that i- and B-DNA are structurally compatible, giving rise to a distinctive fold with peculiar groove shapes. The effect of different residues at the i-motif/duplex interface has been explored. We also show that these constructs can be adapted to sequences of biological relevance, like that found in the promoter region of the KRAS oncogene.

The i-motif is a four-stranded intercalated structure stabilized by the formation of hemi-protonated C:C⁺ base-pairs.¹ Since its recent observation in human cells,² the i-motif (also named as i-DNA)³ is attracting extraordinary attention.⁴ Several studies in vitro and in vivo have found i-motifs in biologically relevant C-rich sequences involved in processes like gene transcription⁵ and DNA synthesis.⁶ As in the case of G-quadruplexes, bioinformatics searches indicate that i-motif-forming sequences are common in the genome.⁷ Since DNA is mainly a double helix in the cell, local G-quadruplex or i-motif formation entails the occurrence of interfaces (junctions) with canonical B-form DNA regions. In recent years, G-quadruplex/duplex junctions have been extensively studied⁸ and found to be an interesting target for selective recognition.⁹ However, very little is known about the structure of i-motif/duplex junctions (IDJs), although they are probably as common as those involving G-quadruplexes.¹⁰

One of the difficulties for studying IDJs is that different experimental conditions are usually required for i-motif and duplex formation. Although some C-rich sequences have been found to form i-motifs at neutral conditions, their stability is usually low. With only one recent exception, all structural studies on i-DNA were carried out at acidic pH.⁷

i-DNA stability follows complex rules still not fully understood. The effects of cytosine tract length and connecting loops have been widely studied in the past few years.¹¹ Capping interactions at the sides of the C:C⁺ stack also play important roles in i-motif stability.¹² In particular, minor groove tetrads (MGTs) have been found to be excellent capping elements, able to stabilize i-motifs at neutral conditions.⁷ MGTs result from the association of Watson–Crick or G:T base-pairs through their minor groove side. A:T:A:T,¹³ A:G:A:G,¹⁴ G:C:G:C,^{13,15} G:C:G:T, or G:T:G:T minor groove tetrads have been observed in different contexts, and the last three have been found to stabilize i-motifs.^{7,16} The first case of an i-motif stabilized by MGTs was observed in a centromeric sequence stabilized by a G:T:G:T tetrad.^{16b} These tetrads were then found in other dimeric i-motifs,^{16a} and more recently in the structure of the so-called mini i-motif.⁷ The

importance of these interactions is reflected by the fact that consensus sequences based on their capability to form favorable C:C⁺/MGT interactions have been found to be prevalent in the human genome, occurring preferentially near regulatory regions.

Following these previous findings, we propose to explore IDJs by designing constructs in which one end of the i-motif is stabilized by a minor groove G:T:G:T tetrad, whereas the other end forms an interface with sequences capable of forming a duplex. For practical reasons, we use a sequence forming a stem-loop hairpin structure, with the stem region being a suitable model of a B-form duplex. The sequence chosen for this region is known to form a stable hairpin with well-dispersed NMR spectra.¹⁷ A similar strategy has been previously used to build model interfaces between B-DNA and G-quadruplexes.^{8a} These hybrid DNA structures, containing i-motifs and stem/loop hairpins, have been proposed as part of molecular sensors¹⁸ and most probably occur in biologically relevant sequences in the genome, like those found in promoter regions of KRAS^{5a} or NMYC¹⁹ oncogenes, as well as in viral genomes.²⁰ However, despite their great interest, no detailed structural study in these systems has been published to date.

As a proof of concept, we explored a DNA construct based on the scheme shown in Figure 1A (IDJ1). Other related constructs (IDJ2 and IDJ3) are discussed in the Supporting Information (Supplementary Results and Figure S1). The NMR spectra of IDJ1 exhibit well-dispersed sharp signals (Figure 1C), allowing the identification of the imino resonances corresponding to C:C⁺ (15–16 ppm), Watson–Crick base-pairs (12.5–14.0 ppm), and G:T or T:T mismatches (10.5–12.0 ppm). UV melting curves were

Received: May 9, 2021

Published: August 9, 2021



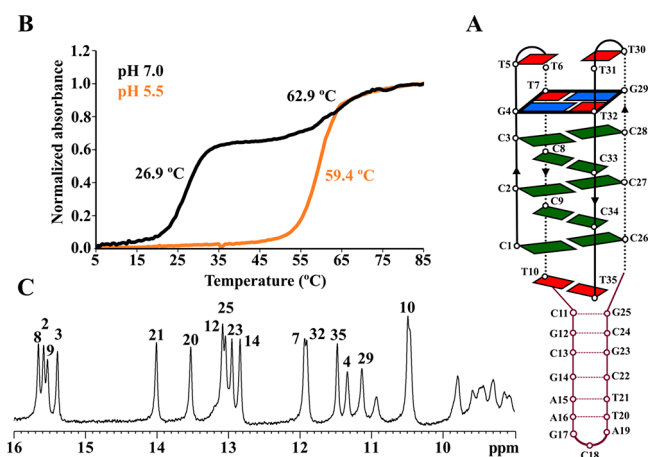


Figure 1. (A) Scheme of IDJ1. (B) UV melting curves at pH 5 and 7 ([DNA] = 2 μ M). (C) Imino protons region of NMR spectra (pH 7, T = 5 $^{\circ}$ C, [DNA] = 0.5 mM).

recorded at two different pH values (Figure 1B). At pH 7, the melting curve exhibits two clear transitions, at 26.9 and 62.6 $^{\circ}$ C, whereas at lower pH only a single transition is observed at 59.4 $^{\circ}$ C. CD spectra at neutral conditions and different temperatures are shown in Figure S2. At low temperature the spectra exhibit a maximum at 284 nm that blue shifts upon heating, and a minimum at 250 nm that red shifts at higher temperature. These experimental data are consistent with the formation of a DNA molecule in which B- and i-DNA regions coexist at neutral pH, as illustrated in Figure 1A. At pH 7, the lower T_m corresponds to the denaturation of the i-DNA moiety, and the higher T_m to the denaturation of the B-DNA part.

First, non-denaturing electrophoretic experiments were performed to rule out the formation of multimeric species (Figure S3). NMR and UV melting data recorded at different DNA concentrations clearly show that the structure is monomeric at conditions adequate for 2D spectra acquisition (Figure S4). Thanks to the good signal dispersion, complete assignment of the NMR spectra of IDJ1 could be carried out by standard 1 H NMR methods. The exchangeable protons regions of the NMR spectra are especially informative (Figure 2). Two A:T and four G:C Watson–Crick base-pairs could be easily identified, corresponding to residues in the hairpin stem.

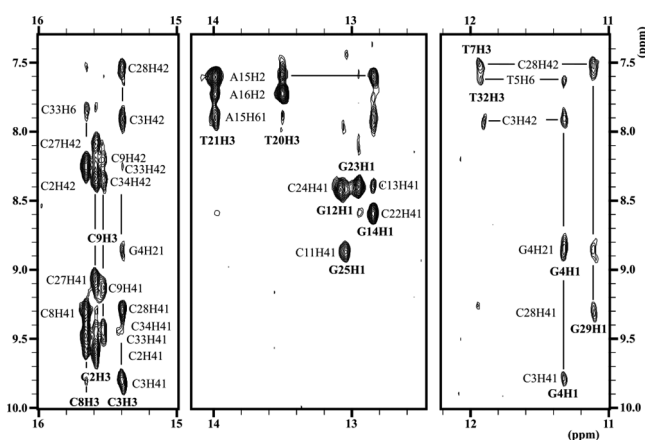


Figure 2. Regions of the NOESY spectra of IDJ1 (T = 5 $^{\circ}$ C, pH 7, [DNA] = 0.5 mM).

NOE sequential connections between these residues could be followed in the sugar/aromatic (Figure S5) and aromatic/aromatic regions. Two G:T base-pairs and one T:T base-pair were detected by their imino–imino and other cross-peaks. Their sequential assignments were performed by analyzing NMR spectra of constructs incorporating 13 C, dU, or 15 N-labeled guanines in key positions (see details in the Supplementary Results and Figures S6–S9). Four cytosine imino signals corresponding to C:C $^+$ base-pairs were found. Assignment of the cytosines involved in the i-motif moiety could be done by identifying first the terminal C:C $^+$ base-pair adjacent to the G:T:G:T tetrad and then following the characteristic sugar–sugar cross-peaks through the i-motif minor groove. Of particular relevance are the NOEs of T10:T35 with C11:G25 and with cytosines 1 and 26, which indicate the formation of C1:C26 $^+$, although the corresponding imino proton signal could not be observed. Overall, the spectra are fully consistent with the schematic representation shown in Figure 1. The chemical shifts are listed in Table S2.

The three-dimensional structure of IDJ1 was determined on the basis of 238 experimental distance constraints by using restrained molecular dynamics methods (see Supporting Information). Torsion angle constraints were also used for those sugars with a clear south conformation according to J-coupling data (Figure S10). Statistical analysis of distance constraints and the resulting structures are given in Table S3 and Figure S11. Final coordinates are deposited in the PDB (7O5E).

The resulting structure (Figure 3) consists of a stack of five hemi-protonated C:C $^+$ base-pairs, surrounded on one side by a minor-groove G:T:G:T tetrad. Two-residue loops connect the residues involved in this tetrad. The first thymine of each loop stacks on top of the tetrad, whereas the other one remains exposed to the solvent. A T:T mismatch is formed at the other end of the C:C $^+$ stack, and the structure continues with six Watson–Crick base-pairs without interrupting the base-pair stacking (Figure S12). This stem region adopts a B-form structure with all glycosidic angles in anti and sugars in south conformation. Geometrical parameters are shown in Table S4.

The overall molecule's surface is dominated by two main grooves (Figure 3), resulting from the connection of each of the i-DNA major grooves with the major and minor grooves of the B-DNA region (named as ImajorBmajor and ImajorBminor grooves in the following discussion). Interestingly, the two major grooves in the i-motif region are not identical, as in other i-motifs, as the one connecting with the B-DNA minor groove is narrower than the other (Figure S13). This is most probably a conformational adjustment to the presence of the adjacent B-DNA region. The resulting ImajorBmajor groove exhibits an approximate width of 11–12 Å , being slightly wider in the i-DNA region. On the other hand, the ImajorBminor groove exhibits a width of around 6 Å , also being slightly wider in the i-DNA moiety (Figures 3 and S13). Although the groove widths are similar in the i-motif and duplex regions, their depths are significantly different, being much shallower in the i-DNA.

Details of interface region are shown in Figure 4. The T10:T35 base-pair, formed between parallel oriented strands, continues the intercalation pattern of the C:C $^+$ base-pairs in the i-DNA, while interacting with the neighboring G:C base-pair of the B-DNA (Figure 4). The twist angles between these consecutive base-pairs are $\sim 30^{\circ}$ and $\sim 27^{\circ}$, respectively. The lack of disruption of the base-pair stacking at the junction,

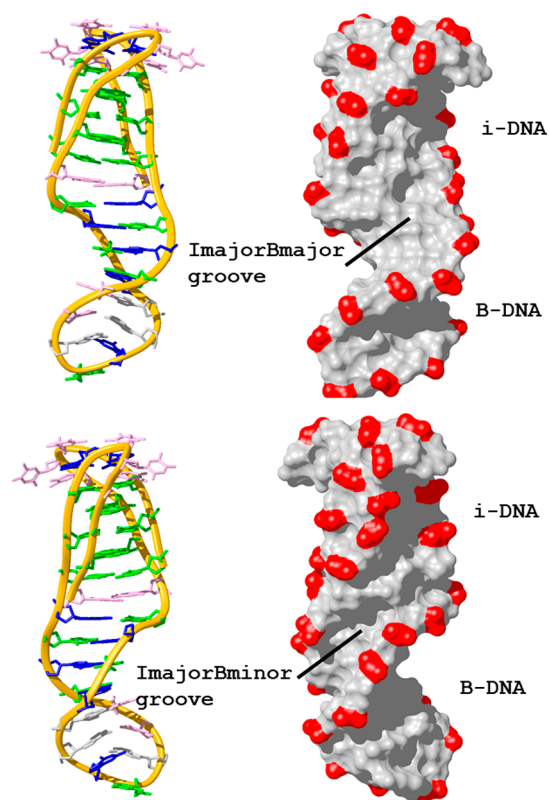


Figure 3. Two views of the IDJ1 structure showing the resulting characteristic grooves. Phosphate-sugar backbone is shown with a golden ribbon. Phosphates are colored in red in the surface representation. Base color code: cytosines in green, guanines in blue, thymines in pink, and adenines in gray (PDB 7O5E).

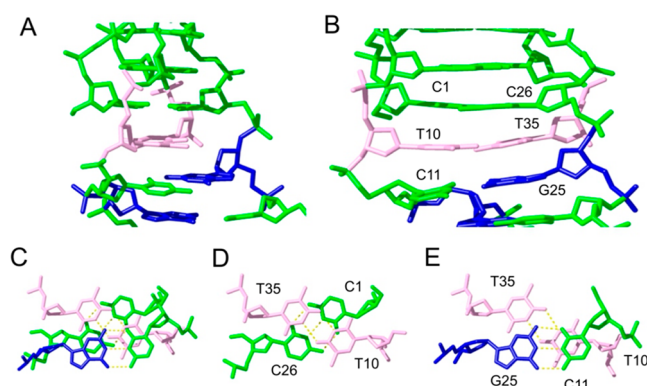


Figure 4. Side and top views of the IDJ1 interface. Same color code as in Figure 3. Hydrogen bonds in bottom panels are shown in yellow. Hydrogen atoms are not displayed.

together with the fine-tuning of the groove sizes, reflects the compatibility of these two families of DNA structures.

To further explore the importance of this interfacial T:T pair to the structure and stability of the junction, we studied additional DNA constructs in which this T:T pair is either absent (IDJ4) or in a different relative orientation with respect to the C:C⁺ stack (IDJ5). A junction with a C:C⁺ base-pair substituting the T:T mismatch was also explored (IDJ6) (Figures S14–S16). NMR and UV melting experiments show a behavior similar to the one observed in IDJ1. In all cases, the NMR spectra are consistent with formation of similar i-motif and duplex moieties. UV denaturation curves indicate lower

stabilities for the i-motifs in IDJ4 and IDJ5. In contrast, the i-motif in IDJ6 is more stable (Table S1). Analyses of their NOESY spectra suggest that these junctions, in particular IDJ5, are more flexible than IDJ1 at the site of the junction (Figures S17 and S18). The higher stability of IDJ6 vs IDJ4 and IDJ5 is probably due to the presence of an additional C:C⁺ base-pair and the formation of an i-motif moiety with the 3'E topology. Interestingly, the orientation of the interfacial base-pair, T:T in IDJ1 or C:C⁺ in IDJ6, may play an important role in the thermal stability of the IDJs.

The sequence in the hairpin moiety was a convenient choice for NMR studies, but other sequences, including biologically relevant ones, can form analogous junctions. For example, formation of an i-motif/hairpin junction has been proposed in the promoter region of the KRAS oncogene.^{5a} In vivo studies suggest that an i-motif-forming region of this promoter containing a stem/loop hairpin can be involved in regulation of KRAS expression by direct interaction with the transcription factor hnRNP K. NMR spectra of the native sequence strongly suggest that i-motif and duplex regions coexist at nearly neutral pH.^{5a} However, the quality of the NMR spectra was very poor. We explored the construct shown in Figure S20 (IDJ7), in which the part of the sequence farthest from the i-motif/duplex interface has been substituted by a sequence that forms a minor-groove tetrad without perturbing the i-motif/duplex interface. The resulting molecule exhibits well-dispersed NMR spectra, as shown in Figure S20, clearly showing the formation of an i-motif/duplex junction at neutral conditions.

In conclusion, i-DNA and B-DNA can coexist at physiological conditions, giving rise to i-motif/duplex interfaces (or junctions). Such interfaces can be conveniently studied by taking advantage of the stabilizing effect of minor-groove tetrads in i-motifs. The three-dimensional structure of the i/B-DNA junction determined here reveals that these two DNA structures are perfectly compatible, with no large disruptions in base-pair stacks. T:T base-pairs, which are known to be stabilizing capping elements in i-motifs, are also very well suited for stabilization of i-motif/duplex junctions. The distinctive groove shapes of this structure suggest that i/B-DNA junctions may be a motif recognized by proteins in the cell, which might be targeted by selective ligands.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.1c04679>.

Detailed descriptions of the experimental procedures and NMR assignments; Tables S1–S5 with assignments, calculation statistics, and structural analysis; Figures S1–S20 showing UV melting curves, CD, electrophoretic experiments, NMR data, and details on the structural models and sequence analysis (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

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

Bartomeu Mir – Inorganic and Organic Chemistry Department, Organic Chemistry Section, and IBUB, University of Barcelona, 08028 Barcelona, Spain

Complete contact information is available at:

<https://pubs.acs.org/10.1021/jacs.1c04679>

Funding

This investigation was supported by research grants from the Spanish “Ministerio de Ciencia e Innovación” (BFU2017-89707-P, PID2020-116620GB-I00).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This manuscript is dedicated to the memory of Professor Enrique Pedroso, our dear colleague and friend. We acknowledge the “Manuel Rico” NMR laboratory (LMR), a node of the ICTS R-LRB. I.S.-C. acknowledges an FPI contract, and B.M. an UB-ADR fellowship. We gratefully acknowledge Dr. Douglas Laurents for careful reading of the manuscripts and his useful comments.

REFERENCES

- (1) Gehring, K.; Leroy, J. L.; Guéron, M. I-Motif. *Nature* **1993**, *363*, 561–565.
- (2) (a) 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., Int. Ed.* **2018**, *57*, 2165–2169. (b) 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.
- (3) Mergny, J. L.; Lacroix, L.; Hélène, C.; Han, X.; Leroy, J. L. Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an i-DNA Motif. *J. Am. Chem. Soc.* **1995**, *117*, 8887–8898.
- (4) (a) Assi, H. A.; Garavis, M.; González, C.; Damha, M. J. I-motif DNA: Structural features and significance to cell biology. *Nucleic Acids Res.* **2018**, *46*, 8038–8056. (b) 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. (c) 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) (a) 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. (b) 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. (c) 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.
- (6) 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.
- (7) 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.
- (8) (a) Lim, K. W.; Phan, A. T. Structural basis of DNA quadruplex-duplex junction formation. *Angew. Chem., Int. Ed.* **2013**, *52*, 8566–8569. (b) Russo Krauss, I.; Ramaswamy, S.; Neidle, S.; Haider, S.; Parkinson, G. N. Structural Insights into the Quadruplex-Duplex 3′ Interface Formed from a Telomeric Repeat: A Potential Molecular Target. *J. Am. Chem. Soc.* **2016**, *138*, 1226–1233.
- (9) Díaz-Casado, L.; Serrano-Chacón, I.; Montalvillo-Jiménez, L.; Corzana, F.; Bastida, A.; Santana, A. G.; González, C.; Asensio, J. L. De Novo Design of Selective Quadruplex–Duplex Junction Ligands and Structural Characterisation of Their Binding Mode: Targeting the G4 Hot-Spot. *Chem. - Eur. J.* **2021**, *27*, 6204–6212.
- (10) Lim, K. W.; Jenjaroenpun, P.; Low, Z. J.; Khong, Z. J.; Ng, Y. S.; Kuznetsov, V. A.; Phan, A. T. Duplex stem-loop-containing quadruplex motifs in the human genome: a combined genomic and structural study. *Nucleic Acids Res.* **2015**, *43*, 5630–5646.
- (11) (a) Cheng, M.; Qiu, D.; Tamon, L.; Ištávková, E.; Všíšková, P.; Amrane, S.; Guédin, A.; Chen, J.; Lacroix, L.; Ju, H.; Trantirek, 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. (b) 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.
- (12) (a) Benabou, S.; Garavis, 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. (b) Fujii, T.; Sugimoto, N. Loop nucleotides impact the stability of intrastrand i-motif structures at neutral pH. *Phys. Chem. Chem. Phys.* **2015**, *17*, 16719–16722. (c) Han, X.; Leroy, J. L.; Guéron, M. An intramolecular i-motif: The solution structure and base-pair opening kinetics of d-(5mCCT3CCT3ACCT3CC). *J. Mol. Biol.* **1998**, *278*, 949–965.
- (13) 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.
- (14) 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.
- (15) (a) 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. (b) Kocman, V.; Plavec, J. Tetrahelical structural family adopted by AGCGA-rich regulatory DNA regions. *Nat. Commun.* **2017**, *8*, 15355.
- (16) (a) Escaja, N.; Viladoms, J.; Garavis, 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. (b) 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.
- (17) Zhu, L.; Chou, S. H.; Xu, J.; Reid, B. R. Structure of a single-cytidine hairpin loop formed by the dna triplet gca. *Nat. Struct. Mol. Biol.* **1995**, *2*, 1012–1017.
- (18) (a) 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. (b) Shi, L.; Peng, P.; Zheng, J.; Wang, Q.; Tian, Z.; Wang, H.; Li, T. I-Motif/miniduplex hybrid structures bind benzothiazole dyes with unprecedented efficiencies: a generic light-up system for label-free DNA nanoassemblies and bioimaging. *Nucleic Acids Res.* **2020**, *48*, 1681–1690.
- (19) Benabou, S.; Ferreira, R.; Aviñó, A.; González, C.; Lyonnais, S.; Solà, M.; Eritja, R.; Jaumot, J.; Gargallo, R. Solution equilibria of

cytosine- and guanine-rich sequences near the promoter region of the n-myc gene that contain stable hairpins within lateral loops. *Biochim. Biophys. Acta, Gen. Subj.* **2014**, *1840*, 41–52.

(20) Ruggiero, E.; Lago, S.; Sket, P.; Nadai, M.; Frasson, I.; Plavec, J.; Richter, S. N. A dynamic i-motif with a duplex stem-loop in the long terminal repeat promoter of the HIV-1 proviral genome modulates viral transcription. *Nucleic Acids Res.* **2019**, *47*, 11057–11068.