Structure of i-Motif/Duplex Junctions at Neutral pH

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

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.

© 2021 The Authors. Published by American Chemical Society.
Assignment of the NMR spectra of (Figure S4). Thanks to the good signal dispersion, complete DNA concentrations clearly show that the structure is coexist at neutral pH, as illustrated in Figure 1A. At pH 7, the formation of a DNA molecule in which B- and i-DNA regions exhibit a maximum at 284 nm that blue shifts upon temperatures are shown in Figure S2. At low temperature the denaturation of the i-DNA moiety, and the higher lower °C, whereas at lower pH only a single transition is observed at 59.4 °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 Tm corresponds to the denaturation of the i-DNA moiety, and the higher Tm 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 1H 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 1. (A) Scheme of IDJ1. (B) UV melting curves at pH 5 and 7 ([DNA] = 2 μM). (C) Imino protons region of NMR spectra (pH 7, T = 5 °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 °C, whereas at lower pH only a single transition is observed at 59.4 °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 Tm corresponds to the denaturation of the i-DNA moiety, and the higher Tm 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 1H 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 °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 5mC, dU, or 15N-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 (70SE).

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 ~30° and ~27°, respectively. The lack of disruption of the base-pair stacking at the junction,

https://doi.org/10.1021/jacs.1c04679
J. Am. Chem. Soc. 2021, 143, 12919–12923
and duplex moieties. UV denaturation curves indicate lower NMR spectra are consistent with formation of similar i-motif substituting the T:T mismatch was also explored (IDJ6). A behavior similar to the one observed in additional DNA constructs in which this T:T pair is either absent (IDJ4). Further exploration of the importance of this interfacial T:T pair to the structure and stability of the junction, we studied the 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. 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. 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
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 Laurens 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.; Hansel-Hertsch, R.; Fessl, T.; Fiala, R.; Loja, T.; Krafcik, D.; Mergny, J.-L.; Foldyna-Trantírek, L.; Trantírek, 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) (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. 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) Fujiy, T.; Sugimoto, N. Loop nucleotides impact the stability of intrastrand i-motif structures at neutral pH. Phys. Chem. Chem. Phys. 2015, 17, 16719–16722.
(4) 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. 2007, 372, 1381–1390.
(5) (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.
(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) Diaz-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) (a) Lim, K. W.; Jenjaroenpun, P.; Low, Z. J.; Khong, Z. J.; Ng, Y. S.; Kuznetsova, 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.; Istvánkóvá, E.; Višková, P.; Amrane, S.; Guédin, A.; Chen, J.; Lacroix, L.; Ju, H.; Trantírek, L.; Sahakyan, A. B.; Zhou, J.; Mergy, 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′-Deoxyctydine 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) Fujiy, T.; Sugimoto, N. Loop nucleotides impact the stability of intrastrand i-motif structures at neutral pH. Phys. Chem. Chem. Phys. 2015, 17, 16719–16722.
(13) (a) Escaja, N.; Pedroso, E.; Rico, M.; González, C. Dimeric solution structure of two cyclic octamers: four-stranded DNA structures stabilized by A: T: A: G and G: C: G: C tetrads. J. Am. Chem. Soc. 2000, 122, 12732–12742.
(14) (a) Chu, B.; Zhang, D.; Huang, W.; Paulkstelis, 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. Tetrahedral 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:T:G tetrad: Solution structure of the d-(TCCGTGGTTCA) 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.

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
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.; Šket, 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.