ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

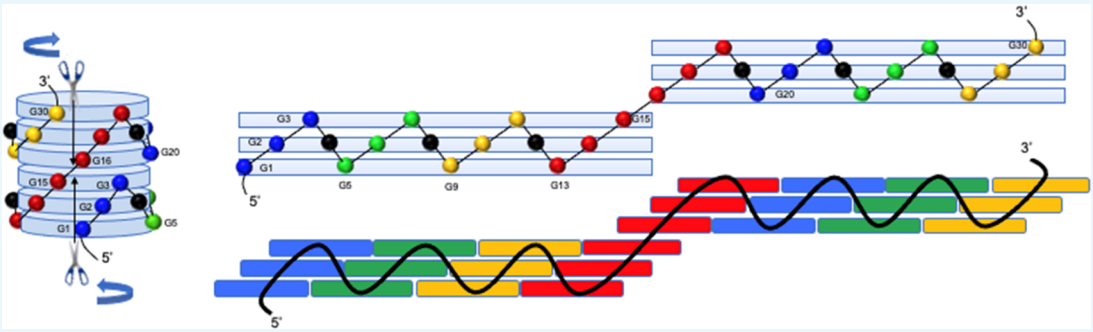
Structure of Tetrahelical DNA Homopolymers Supports Quadruplex World Hypothesis

Levan Lomidze,[§] Mengkun Yang,[§] David Khutsishvili, Nunu Metreveli, Karin Musier-Forsyth, and Besik Kankia*

Cite This: *ACS Omega* 2022, 7, 4311−4316

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: We previously reported a tetrahelical monomolecular architecture of DNA, tmDNA, which employs G-quartets and an all-parallel GGGTGGGTGGGTGGG (G3T) quadruplex as the repeating unit. Based on thermodynamic and kinetic studies, we proposed that covalently joined $(G3T)_n$ units formed an uninterrupted programmable homopolymer; however, structural evidence for the tmDNA architecture was lacking. Here, we used NMR spectroscopy of wild-type and single-inosine-substituted constructs to characterize both monomolecular $(G3T)_2$ and bimolecular quadruplex-Mg-coupled versions of tmDNA. The NMR results support an architecture consisting of uninterrupted stacked G-tetrads in both the monomolecular constructs and bimolecular assemblies. Taken together, these data support the formation of a stable programmable homopolymeric tmDNA architecture, which may have been a precursor to the modern-day Watson−Crick DNA duplex.

INTRODUCTION

The G3T (GGGTGGGTGGGTGGG) quadruplex is a truncated (i.e., lacks the 3′-terminal T) version of a DNA aptamer $(GGGT)_4$, selected for binding to HIV-1 integrase.[1] The quadruplex possesses unprecedented stability, which is attributed to the all-parallel alignment of GGG tracts and single-nucleotide propeller loops.[2,3] Earlier thermodynamic and kinetic studies suggested that the direct conjugation of the G3T sequences results in an uninterrupted monomolecular quadruplex polymer with all-parallel fold and increased stability.[4] Based on these macroscopic studies, a new architecture of tetrahelical monomolecular DNA (tmDNA) was suggested (Figure 1A).[4] The architecture plays a central role in the quadruplex world hypothesis, which postulates the *de novo* appearance of nucleic acid quadruplexes on primitive Earth.[5]

The tmDNA architecture can be visualized as a homopolymer consisting of n number of G3T segments, $(G3T)_n$. The 15-nt G3T unit folds into an all-parallel quadruplex with three G-tetrads and propeller T-loops and represents a structural domain of tmDNA. The terminal G_3-segments of adjacent G3T domains form a G_6-segment, which is responsible for vertical growth of the structure, while the zigzag movement of G3-segments and T-loops builds the quadruplex domains in a horizontal manner (Figure 1A). Whereas the tmDNA architecture was deduced from the results of thermodynamic and kinetic studies of $(G3T)_n$,[4,6] here, we used NMR spectroscopy to validate this architecture and gain additional insights using a bimolecular assembly of tmDNA.

Proton NMR is a powerful tool to determine high-resolution structures of short nucleic acids but is less practical for larger molecules due to the spectral overlap and costly site-specific labeling for peak assignments. Even the smallest version of tmDNA, $(G3T)_2$, contains 24 guanines involved in structurally similar G-tetrads; however, quadruplexes possess several unique features that reduce the spectral overlap, facilitating structural analyses using one-dimensional (1D) H NMR. First, quadruplexes are characterized by unique resonances in the

Received: October 27, 2021
Accepted: December 31, 2021
Published: January 26, 2022

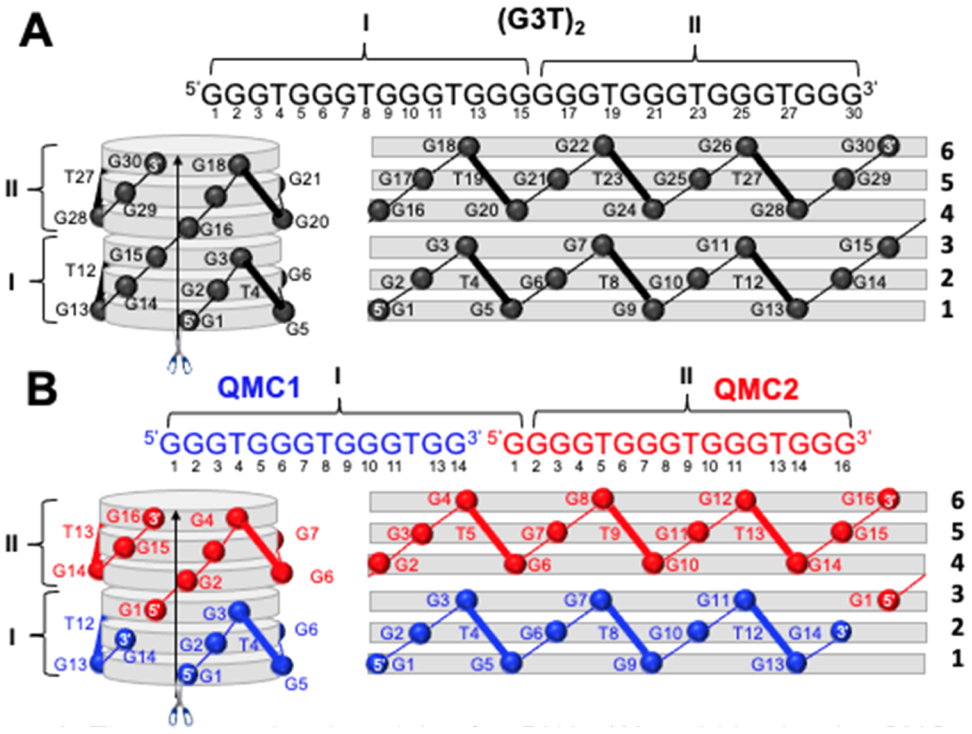


Figure 1. Three-dimensional models of tmDNA (A) and bimolecular QMC (B) and their two-dimensional maps. The spheres indicate G residues, and the thick diagonal bars correspond to T-loops. G-tetrads are depicted as gray disks. The Arabic and Roman numerals correspond to nucleotide and G3T domain. Bold Arabic numerals in the two-dimensional maps correspond to G-tetrad positions.

10−12 ppm region of the H NMR spectrum.[7] These resonances correspond solely to imino protons involved in G-tetrad formation; the same protons involved in the canonical GC base pairing demonstrate chemical shifts within the range of 13−14 ppm.[8] Second, the imino protons of internal G-tetrads exchange very slowly with the solvent and can be detected long after redissolving the samples in 100% D_2O. In contrast, the imino protons of terminal G-tetrads rapidly exchange with deuteron and their signals disappear.[9] For instance, the H NMR spectrum of a G3T monomer in water shows twelve sharp peaks corresponding to three G-tetrads, while in D_2O, the number of peaks is reduced to four, corresponding to the middle G-tetrad.[9] Third, even though inosine lacks the 2-amino group, which is important for hydrogen bond formation in G-quadruplexes, single guanine to inosine (G → I) substitutions are tolerated by G-tetrads (see Figure S1) and are accompanied by strong downfield chemical shifts.[2,10] Thus, G → I substitutions improve the NMR spectral quality by removing one peak from the overlapped region, allowing straightforward monitoring of the isolated peak.

Solution structures of all-parallel monomeric quadruplexes, formed by G3T and similar sequences, are well characterized.[2,10] It was shown that the $(GGGT)_4$ aptamer forms stable dimers through 5′−5′ stacking interactions.[2] However, the addition of a 5′-TT dinucleotide efficiently disrupts the dimerization process and transforms the dimers into monomers. An NMR study of the monomeric $TT(GGGT)_4$ demonstrated 12 well-separated peaks in the 11−12 ppm range.[9] Here, we carried out a 1D NMR study of the monomolecular wild-type $(G3T)_2$ quadruplex, as well as a variant with a single G → I substitution. We also studied a related quadruplex-Mg-coupling (QMC) system[11] formed via a bimolecular interaction involving G3T variants capable of tmDNA formation in the presence of Mg^{2+} ions (Figure 1B). Thus, the QMC system represents a bimolecular analogue of $(G3T)_2$ (Figure 1). This study strongly supports the formation of the proposed tmDNA architecture using both monomolecular and bimolecular assemblies and provides structural support for the quadruplex world hypothesis.

RESULTS AND DISCUSSION

Effects of G3 → I Substitutions on Stability and the Secondary Structure. It was previously reported that inosines, which are G analogues that lack the exocyclic amino group, can form I-tetrads, and G-tetrads can tolerate single G → I substitutions.[12−16] These substitutions not only improve spectral characteristics in NMR studies but also allow unambiguous monitoring of the selected peaks.[2,10] However, these substitutions can also be accompanied by structural rearrangements and changes in quadruplex topology.[15,16] Therefore, we first studied the impact of G → I substitutions on the secondary structure and thermal stability of G3T and $(G3T)_2$ quadruplexes.

The G → I substitutions at three different positions of the G3T monomer (G1, G2, and G3) revealed similar destabilization effects, ∼13 °C per substitution (see Figure S2). However, the substitutions do not induce any measurable changes in the CD profile of G3T (Figure 2).[2,10] Thus, single G → I substitutions in guanosines that participate in the formation of each of the three G-tetrads (see Figure 1A) significantly destabilize the G3T quadruplex without changing its secondary structure or topology. Next, we tested the effect of G → I substitutions on the folding topology and stability of $(G3T)_2$. For this, we designed a construct with a G20 → I substitution in the first tetrad of the second G3T quadruplex, TT-$(G3T)_2$-I-T. Like the G3T-I constructs, TT-$(G3T)_2$-I-T showed a CD spectrum characteristic of an all-parallel topology

https://doi.org/10.1021/acsomega.1c06026
ACS Omega 2022, 7, 4311−4316

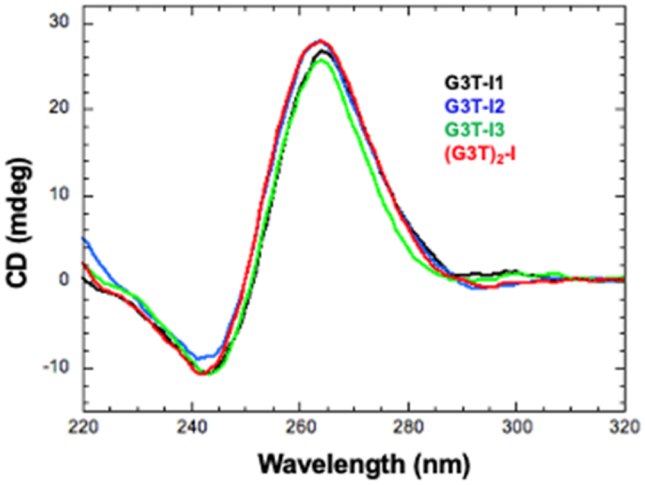


Figure 2. CD profiles of inosine-containing constructs at 20 °C in 0.1 mM KCl, 10 mM Tris, pH 7.

(Figure 2). The melting behavior of TT-$(G3T)_2$-I-T is slightly different from that of TT-$(G3T)_2$-T (Figure 3). In 0.1 mM K^+,

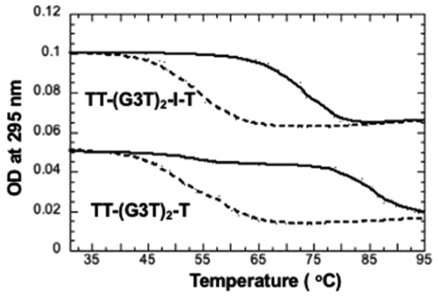


Figure 3. UV melting (solid) and cooling (dashed) curves of TT-$(G3T)_2$-I-T and TT-$(G3T)_2$-T conducted in 0.1 mM K^+, 10 mM Tris, pH 7. Curves are offset for clarity.

TT-$(G3T)_2$-T unfolds at ∼85 °C and refolds at ∼55 °C, resulting in a ∼30 °C hysteresis loop.[6] This hysteresis is attributed to domain assembly and interdomain stacking between the extended tetrahelix with six G-tetrads (Figure 1A).[6] TT-$(G3T)_2$-I-T also demonstrates a similar hysteresis loop; however, the unfolding transition takes place at significantly lower temperatures, ∼73 °C. This is because TT-$(G3T)_2$-I-T consists of two domains with unequal T_ms: G3T with a T_m of 55 °C and G3T-I with a T_m of 42 °C. During the heating cycle, unfolding starts from the less stable G3T-I domain, which shifts the melting curve to lower temperatures. During the cooling cycle, refolding starts from the more stable domain, G3T. As a result, refolding curves of TT-$(G3T)_2$-I-T and TT-$(G3T)_2$-T are almost superimposable (Figure 3). We concluded that G → I substitutions destabilize both TT-G3T-T and TT-$(G3T)_2$-T without affecting their folding topology or overall architecture.

NMR Spectra of $(G3T)_2$. As mentioned above, a high-resolution NMR structure of the monomolecular G3T quadruplex was solved using the sequence with thymidine attachments, TT-G3T-T, which prevents self-dimerization.[9] Therefore, (i) to avoid intermolecular interaction and (ii) to perform proper comparison with the literature data, our constructs used in NMR experiments contained the same attachments (see Materials and Methods).

The 1D proton NMR spectrum of TT-G3T-T revealed twelve peaks in the 11−11.8 ppm range (Figure S3), which was in excellent agreement with earlier studies performed in the absence of Mg^{2+} ions.[9] Figure 4A shows the 1D NMR spectrum of TT-$(G3T)_2$-T with strong signals in the region characteristic of guanine imino protons involved in G-tetrad formation. In comparison to TT-G3T-T, the spectrum of TT-$(G3T)_2$-T is shifted slightly upfield to the 10.6−11.7 ppm region. The spectrum also reveals a strong overlap in the 10.8−11 ppm range (Figure 4A). Both the upfield shift and peak overlap are consistent with the increased number of repeating units (G-tetrads) in TT-$(G3T)_2$-T, which increases structural symmetry and overall shielding of the imino protons. Similar effects were reported upon self-dimerization of G3T via the 5′−5′ stacking interface, $^{3'}$T-G3T$^{5'}$.$^{5'}$G3T-T$^{3'}$.[2]

Integration of the peaks in the H_2O spectrum shown in Figure 4A reveals 24 protons. Since the TT-$(G3T)_2$-T quadruplex is monomorphic (i.e., forming only one structure),[4,6] this indicates that all 24 Gs are engaged in G-tetrads. This result is consistent with the proposed tmDNA architecture containing G3T domains with three G-tetrads and propeller T-loops; any other folding pattern would require at least two-nt loops (i.e., TG),[17,18] resulting in at most four G-tetrads or 16 peaks.

In D_2O, the TT-$(G3T)_2$-T spectrum is reduced to 16 peaks as a result of H → D exchange at two of the G-tetrads. This is consistent with an uninterrupted tmDNA structure with an exchange only occurring at the terminal G-tetrads (Figure S4). The peaks that disappeared in D_2O include a quartet between 11.4−11.7 ppm and another quartet between 10.7−10.9 ppm (Figure 4A, red stars). Earlier NMR studies revealed that the peaks of TT-G3T-T are grouped according to G-tetrads (Figure S3): the most downfield and the most upfield quartets correspond to 5′- and 3′-end G-tetrads, respectively (see Figure S3).[9] Thus, we propose that the downfield peaks (11.4−11.7) correspond to the 5′-end G-tetrad and the peaks between 10.7−10.9 ppm correspond to the 3′-end G-tetrad. The existence of 16 peaks in D_2O further confirms that the TT-$(G3T)_2$-T structure is monomorphic. In the case of polymorphism (i.e., two distinct structures with three G-tetrads each), the NMR spectrum in D_2O would result in the disappearance of 16 peaks corresponding to the 4 terminal G-tetrads. Thus, the number of observed peaks, both in H_2O and D_2O, fully supports the folding of TT-$(G3T)_2$-T into an uninterrupted tmDNA architecture, as shown in Figure 1A.

NMR Spectra of QMC Complex. We next investigated the structure of a bimolecular QMC assembly based on the G3T quadruplex as shown in Figure 1B. The main difference between the QMC system and TT-$(G3T)_2$-T is that the former lacks a phosphodiester bond between G14 and G15. The 1D H NMR spectra of the QMC system and TT-$(G3T)_2$-T are therefore very similar, with a well-separated quartet of peaks in the downfield regions (11.4−11.7 ppm) and a significant peak overlap at 10.8−11 ppm (compare Figure 4A,B). All major features observed for TT-$(G3T)_2$-T hold true for QMC, which shows signals for 24 protons in H_2O, 16 protons in D_2O, and clustering of the disappeared peak quartets. The lack of one phosphodiester bond decreases the structural symmetry and reveals better separation of some of the peaks relative to the monomolecular system. For instance, three protons of TT-$(G3T)_2$-T that result in an overlapping chemical shift at 11.15 ppm are well separated in the QMC system (Figure 4B). Thus, the NMR study supports the

https://doi.org/10.1021/acsomega.1c06026
ACS Omega 2022, 7, 4311−4316

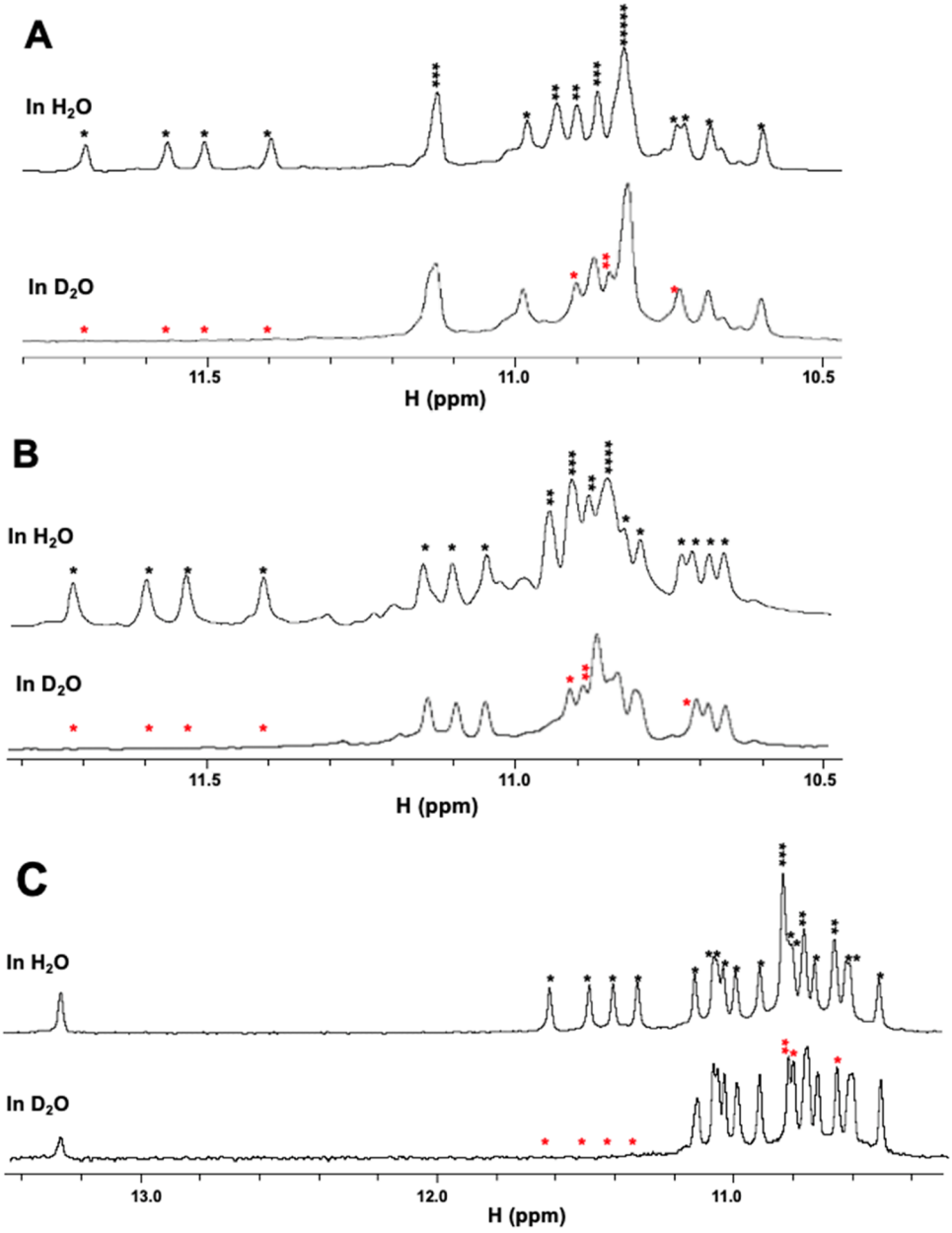


Figure 4. Imino proton spectra of TT-$(G3T)_2$-T (A), bimolecular quadruplex-Mg-coupling (QMC) system (B), and TT-$(G3T)_2$-I-T (C) in H_2O and D_2O. Red stars correspond to peaks that disappeared in D_2O. Buffer: 10 mM KCI, 2 mM $MgCI_2$, 10 mM Tris-HCJ, pH 7.0 at 15 °C.

formation of a QMC structure with six uninterrupted G-tetrads and supports the proposed tmDNA architecture.

NMR Spectra of Inosine-Containing TT-$(G3T)_2$-I-T. One of the main goals of studying constructs with G → I substitutions was to demonstrate the uninterrupted nature of the tmDNA architecture. For this experiment, we used the TT-$(G3T)_2$-I-T construct containing an inosine substitution at G20 in the 4th G-tetrad, which was involved in the interdomain interaction (see Figure S4). As expected, in H_2O, the H NMR spectrum of TT-$(G3T)_2$-I-T reveals an imino peak around 13.3 ppm corresponding to the inosine substitution (Figure 4C). The peak does not disappear in 100% D_2O, which supports the uninterrupted nature of the tmDNA architecture. Once again, all of the major features observed for TT-$(G3T)_2$-T and the QMC construct hold true for TT-$(G3T)_2$-I-T. In addition, the G → I substitution increases structural asymmetry and results in better separation of the peaks (see Figure S5).

CONCLUSIONS

This H NMR study of G_3- and G_6-tracts connected to each other through single-nucleotide propeller loops supports the formation of an uninterrupted tetrahelical monomolecular architecture with the all-parallel alignment. In addition, we demonstrate that matching QMC pairs forms a bimolecular complex with the same tetrahelical architecture. The tmDNA architecture provides key structural support for the quadruplex world hypothesis and may have been a precursor to the modern-day Watson−Crick DNA duplex.

https://doi.org/10.1021/acsomega.1c06026
ACS Omega 2022, 7, 4311−4316

MATERIALS AND METHODS

All DNA oligonucleotides were obtained from Integrated DNA Technologies. The concentration of the DNA oligonucleotides was determined by measuring UV absorption at 260 nm as described earlier.[19] Oligonucleotide names and sequences from 5′ to 3′ are

G3T-I1; IGGTGGGTGGGTGGG
G3T-I2; GIGTGGGTGGGTGGG
G3T-I3; GGITGGGTGGGTGGG
TT-G3T-T; TTGGGTGGGTGGGTGGGT
T T - (G 3 T) $_2$ - T ; TTGGGTGGGTGGGTGGGGGGTGGGTGGGTGGGT
TT-(G3T)$_2$-I-T; TTGGGTGGGTGGGTGGGGGGT I GGTGGGTGGGT
QMC1; TTGGGTGGGTGGGTGG
QMC2; GGGGTGGGTGGGTGGGT

The first three single G- to I-substituted oligonucleotides shown above did not have 5′-terminal TT dinucleotides since they were only used in thermodynamic studies at 4 μM in the absence of divalent cations. Under these conditions, the quadruplexes do not reveal any significant dimerization typical for NMR studies performed at significantly higher concentrations.

NMR experiments were conducted on a 600 MHz Bruker spectrometer at 15 °C. DNA concentrations were between 0.2−1 mM per strand and oligonucleotides were dissolved in 10 mM KCl, 2 mM $MgCl_2$, 10 mM Tris-HCl, pH 7.0, and 10% D_2O. After recording the spectra in 10% D_2O, the samples were lyophilized, redissolved in 100% D_2O for 24 h, and reanalyzed by NMR.

UV unfolding/folding experiments were recorded at 295 nm as a function of temperature using a Varian UV−visible spectrophotometer (Cary 100 Bio). CD spectra were obtained with a Jasco-815 spectropolarimeter. The devices were equipped with thermoelectrically controlled cuvette holders. In a typical experiment, oligonucleotide stock solutions were mixed into 0.1 mM KCl, 10 mM Tris-HCl, pH 7.0 buffer in optical cuvettes. The solutions were incubated at 95 °C for a few minutes and annealed at room temperature for 10−15 min prior to ramping to the desired starting temperatures. The melting experiments were performed at a heating rate of 1 °C/min. All melting experiments of G3T, G3T-I1, G3T-I2, and G3T-I3 molecules revealed cooperative two-state transitions with the transition width around 20 °C. The melting temperature, T_m (± 0.5 °C), corresponds to the midpoint of the transition. CD and UV experiments were conducted at 4 μM concentration per G3T domain. The T_m values and CD profiles represent averages of at least three measurements. van't Hoff enthalpies, ΔH_{vH}, were also calculated using the following equation: $\Delta H_{vH} = 4\ R\ T_m^2\ \delta\alpha/\delta T$, where R is the gas constant and $\delta\alpha/\delta T$ is the slope of the normalized optical absorbance versus temperature curve at T_m.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.1c06026.

GGGI tetrad scheme; UV melting experiments; H NMR spectrum of TT-G3T-T; schematics of $(G3T)_2$; and imino proton spectra of TT-$(G3T)_2$-I-T (PDF)

AUTHOR INFORMATION

Corresponding Author

Besik Kankia − *Institute of Biophysics, Ilia State University, Tbilisi 0162, Republic of Georgia; Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States;* orcid.org/0000-0002-2091-485X; Phone: 614-688-8799; Email: kankia.1@osu.edu; Fax: 614-688-5402

Authors

Levan Lomidze − *Institute of Biophysics, Ilia State University, Tbilisi 0162, Republic of Georgia; Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States*

Mengkun Yang − *Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States;* orcid.org/0000-0003-1780-1210

David Khutsishvili − *Institute of Biophysics, Ilia State University, Tbilisi 0162, Republic of Georgia*

Nunu Metreveli − *Institute of Biophysics, Ilia State University, Tbilisi 0162, Republic of Georgia*

Karin Musier-Forsyth − *Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c06026

Author Contributions

[§]L.L. and M.Y. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Dr. Chunhua Yuan at the Campus Chemical Instrument Center for help in NMR experiments and many useful discussions. This work was funded by the National Institutes of Health grant RO1 AI153216 (to K.M.-F.) and the Shota Rustaveli National Science Foundation (grant FR17_140) (to B.K.).

REFERENCES

(1) Jing, N.; Hogan, M. E. Structure-activity of tetrad-forming oligonucleotides as a potent anti-HIV therapeutic drug. *J. Biol. Chem.* **1998**, *273*, 34992−34999.

(2) Do, N. Q.; Lim, K. W.; Teo, M. H.; Heddi, B.; Phan, A. T. Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res.* **2011**, *39*, 9448−9457.

(3) Kelley, S.; Boroda, S.; Musier-Forsyth, K.; Kankia, B. I. HIV-integrase aptamer folds into a parallel quadruplex: a thermodynamic study. *Biophys. Chem.* **2011**, *155*, 82−88.

(4) Kankia, B. Tetrahelical monomolecular architecture of DNA: a new building block for nanotechnology. *J. Phys. Chem. B* **2014**, *118*, 6134−6140.

(5) Kankia, B. Quadruplex World. *Origins Life Evol. Biospheres* **2021**, *51*, 273−286.

(6) Kankia, B.; Gvarjaladze, D.; Rabe, A.; Lomidze, L.; Metreveli, N.; Musier-Forsyth, K. Stable Domain Assembly of a Monomolecular DNA Quadruplex: Implications for DNA-Based Nanoswitches. *Biophys. J.* **2016**, *110*, 2169−2175.

(7) Feigon, J.; Koshlap, K. M.; Smith, F. W. 1H NMR spectroscopy of DNA triplexes and quadruplexes. *Methods Enzymol.* **1995**, *261*, 225−255.

(8) Patel, D. J.; Tonelli, A. E. Assignment of the proton Nmr chemical shifts of the T-N3H and G-N1H proton resonances in isolated AT and GC Watson-Crick base pairs in double-stranded deoxy oligonucleotides in aqueous solution. *Biopolymers* **1974**, *13*, 1943−1964.
(9) Do, N. Q.; Phan, A. T. Monomer-dimer equilibrium for the 5′-5′ stacking of propeller-type parallel-stranded G-quadruplexes: NMR structural study. *Chemistry* **2012**, *18*, 14752−14759.
(10) Mukundan, V. T.; Do, N. Q.; Phan, A. T. HIV-1 integrase inhibitor T30177 forms a stacked dimeric G-quadruplex structure containing bulges. *Nucleic Acids Res.* **2011**, *39*, 8984−8991.
(11) Kankia, B. Quadruplex-and-Mg(2+) Connection (QMC) of DNA. *Sci. Rep.* **2015**, *5*, No. 12996.
(12) Petrovic, A. G.; Polavarapu, P. L. Quadruplex structure of polyriboinosinic acid: dependence on alkali metal ion concentration, pH and temperature. *J. Phys. Chem. B* **2008**, *112*, 2255−2260.
(13) Simard, C.; Savoie, R. A vibrational spectroscopic study of the self-association of polyinosinic acid and polyguanylic acid in aqueous solution. *Biopolymers* **1994**, *34*, 91−100.
(14) Thiele, D.; Guschlbauer, W. The structures of polyinosinic acid. *Biophysik* **1973**, *9*, 261−277.
(15) Risitano, A.; Fox, K. R. Inosine substitutions demonstrate that intramolecular DNA quadruplexes adopt different conformations in the presence of sodium and potassium. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2047−2050.
(16) Lim, K. W.; Lacroix, L.; Yue, D. J.; Lim, J. K.; Lim, J. M.; Phan, A. T. Coexistence of two distinct G-quadruplex conformations in the hTERT promoter. *J. Am. Chem. Soc.* **2010**, *132*, 12331−12342.
(17) Kankia, B. Monomolecular tetrahelix of polyguanine with a strictly defined folding pattern. *Sci. Rep.* **2018**, *8*, No. 10115.
(18) Smirnov, I.; Shafer, R. H. Effect of loop sequence and size on DNA aptamer stability. *Biochemistry* **2000**, *39*, 1462−1468.
(19) Kankia, B. I.; Marky, L. A. DNA, RNA, and DNA/RNA oligomer duplexes: a comparative study of their stability, heat, hydration and Mg(2+) binding properties. *J. Phys. Chem. B* **1999**, *103*, 8759−8767.

https://doi.org/10.1021/acsomega.1c06026
ACS Omega 2022, 7, 4311−4316