A DNA G-quadruplex/i-motif hybrid

Betty Chu, Daoning Zhang and Paul J. Paukstelis*

Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, MD 20742, USA

ABSTRACT

DNA can form many structures beyond the canonical Watson–Crick double helix. It is now clear that noncanonical structures are present in genomic DNA and have biological functions. G-rich G-quadruplexes and C-rich i-motifs are the most well-characterized noncanonical DNA motifs that have been detected *in vivo* with either proscribed or postulated biological roles. Because of their independent sequence requirements, these structures have largely been considered distinct types of quadruplexes. Here, we describe the crystal structure of the DNA oligonucleotide, d(CCAGGCTGCAA), that self-associates to form a quadruplex structure containing two central antiparallel G-tetrads and six i-motif C–C$^+$ base pairs. Solution studies suggest a robust structural motif capable of assembling as a tetramer of individual strands or as a dimer when composed of tandem repeats. This hybrid structure highlights the growing structural diversity of DNA and suggests that biological systems may harbor many functionally important non-duplex structures.

INTRODUCTION

Non-Watson–Crick base pairing interactions in DNA can give rise to a variety of structural motifs beyond the canonical double helix. New types of DNA structural motifs continue to be reported (1–9), suggesting that our understanding of DNA's structural diversity has not been reached. The G-quadruplex and the i-motif are two noncanonical structures that have been studied extensively, and each is characterized by specific types of noncanonical interactions. G-quadruplexes (G4s) are formed from G-rich sequences and contain stacked guanosine tetrads, organized in a cyclic hydrogen bonding arrangement between the Hoogsteen and Watson–Crick faces of neighboring nucleobases (1,10). G4s can be formed through inter- or intramolecular interactions in a variety of topologies and are stabilized by central cations (11–13). The DNA i-motif is characterized by the formation of hemiprotonated C–C$^+$ parallel-stranded base pairs, which are organized to allow two duplexes to interca-

late in an antiparallel fashion to form a quadruplex structure (2,14). Both G4s and i-motifs can form as unimolecular, bimolecular or tetramolecular assemblies, leading to diverse folding topologies (15,16).

Though G4 and i-motif structures tend to form from sequences that contain contiguous stretches of G's or C's, respectively, structural characterization has revealed a relatively wide distribution of sequences capable of forming these and similar noncanonical motifs. A unimolecular G4 consensus motif, $G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}$, was initially used for G4 identification (17), leading to initial estimates of ∼300,000 possible G4-forming structures in the human genome (18). However, mounting structural evidence indicated that the sequences capable of forming G4s, and the G4 structures themselves, were more diverse than originally thought. Structural variations of G4 structures include motifs that incorporate non-G-tetrads (19), bulged residues (20), G-triads (21,22), G-tetrads as part of pentad assemblies (23) and hybrid G-quadruplex/duplexes (24,25). This sequence and structural diversity led to the doubling of the predicted G4-forming sequences in the human genome to >700,000 (26). Similarly, a unimolecular i-motif folding rule was formulated based on experimental evidence (27). This specified five cytosine residues for each of the four C-tracts, but allowed for greater variation in the length and sequence of the loop regions. Based on this, a preliminary search predicted >5000 i-motif-forming sequences in the human genome (27). However, isolated i-motif structures with shorter or longer C-tracts have been reported (28–30), and the characteristic C–C$^+$ base pair of i-motifs is prevalent in a variety of other noncanonical DNA structures (4,6,8,31,32), suggesting that they can serve as building blocks or structural units for other types of structures. Additionally, the structural topology of i-motifs is not limited to only C–C$^+$ base pairs. Even the earliest i-motif structures incorporated other noncanonical base pairs (2,33–36) or base triples (37,38) that stabilize the motif through stacking on the hemiprotonated cytosine base pairs (39). As a result, the number of sequences in the human genome with the potential to form i-motifs or related structures is likely much greater than previously predicted.

Both of these noncanonical structural motifs are present in cellular DNA, though their roles in biological processes are just beginning to be understood. G4s have been impli-

---

*To whom correspondence should be addressed. Tel: +1 301 405 9933; Fax: +1 301 314 2779; Email: paukstel@umd.edu

cated in a wide variety of normal cellular processes, including DNA replication and transcription, as well as a number of disease states (40). Telomeric G4 structures have been visualized using specific antibodies (41). The active formation of G4s (42,43), as well as their stabilization by small molecule ligands (42), in human cells have also been confirmed. With a predicted 50% of human genes containing G4s at or around promoter regions, DNA G4 structures are predicted to have widespread roles in gene expression (44). In particular, the significant enrichment of the G4 motif in a wide range of oncogene promoters suggests its functional importance in cancer (45). Examples of G4s modulating gene transcription have been found in the c-MYC (46), bcl-2 (47), and KRAS (48) oncogene promoters. Additionally, the stabilization of G4s by small molecule ligands at the hTERT (49) and PDGFR-β (50) oncogene promoters has been associated with downregulated activity. Nonetheless, the highly thermostable G4s can be detrimental to biological processes and lead to genome instabilities (40,51). DNA i-motifs have long been implicated in biological processes (27,45,52), but have now been observed *in vivo*. In-cell NMR identified characteristic i-motif signals in HeLa extracts with transfected i-motif DNAs, providing direct evidence that i-motif structures are stable in cellular environments (53). Furthermore, the antibody-mediated observation of i-motifs in the nuclei of human cells (54) and the discovery of i-motif binding proteins that regulate gene activity (55) demonstrate that i-motifs can have biological function. The sequence and structural diversity of G4s and i-motifs and their growing importance in cellular DNA transactions open the possibility of new variations of these motifs with distinct biological functions.

Here, we describe the crystal structure of a G-quadruplex/i-motif hybrid structure, formed by the oligonucleotide, d(CCAGGCTGCAA). Two distinct strands form a dimer through parallel-stranded interactions, while a symmetry-related dimer interacts in an antiparallel orientation to form the tetramer. The tetramer contains two central G-tetrads that are stabilized by a barium ion and are flanked on either side by a base triple, one unpaired guanosine and an i-motif of three C–C$^+$ base pairs. Solution studies indicate that the same hybrid quadruplex is formed from tandem sequence repeats, suggesting the potential for this type of structure to form from repetitive DNA elements. d(CCAGGCTGCAA) represents the first structural observation of a G-quadruplex/i-motif hybrid and further expands the wide-ranging structural diversity of DNA.

MATERIALS AND METHODS

Oligonucleotide synthesis and purification

The 11-mer, d(CCAGGCTGCAA), the derivative, d(CCAGGCU^{Br}GCAA) and d(CCAGGCTGC^{Br}AA), and the 22-mer, d(CCAGGCTGCAACCAGGCTGCAA), oligonucleotides were synthesized using standard phosphoramidite chemistry on an Expedite 8909 Nucleic Acid Synthesizer (PerSeptive Biosystems, Inc.), with reagents from Glen Research (Sterling, VA). The 11-nucleotide-long sequences were purified using the Glen-Pak cartridges according to the manufacturer's protocol. The 22-mer

oligonucleotide was purified using denaturing gel electrophoresis followed by electroelution.

Crystallization and data collection

Sitting drops of d(CCAGGCTGCAA) were set up by mixing 1 µl of 500 µM DNA solution with 2 µl of crystallization solution (30% polyethylene glycol 400 (PEG400), 20 mM barium chloride, 10 mM spermidine, and 30 mM Bis–Tris at pH 8.5). Sitting drops of d(CCAGGCU^{Br}GCAA) were set up by mixing 1 µl of 500 µM DNA solution with 2 µl of crystallization solution (25% polyethylene glycol 400 (PEG400), 40 mM barium chloride, 10 mM spermidine, and 30 mM Bis–Tris at pH 8.5). These drops were equilibrated against 300 µl of 5% PEG400 in the well reservoir at 22°C for 15–20 h, followed by subsequent equilibration with 3–4 µl of glacial acetic acid added to the well reservoir. Crystals were observed 2 days after the addition of acid. Crystals were removed from the drops by nylon cryoloops and directly cryo-cooled in liquid nitrogen.

d(CCAGGCTGC^{Br}AA) was crystallized by mixing 3 µl of 500 µM DNA solution with 3 µl of crystallization solution (15% 2-methyl-2,4-pentanediol (MPD), 120 mM calcium chloride, 20 mM lithium chloride, 8 mM spermidine, and 30 mM sodium cacodylate at pH 5.5). Crystallization was performed at 22°C and in sitting drops, which were equilibrated against 300 µl of 20% MPD in the well reservoir. Crystals were observed 2 days after plating. Crystals were removed from the drops by nylon cryoloops, dipped in 30% MPD, and cryo-cooled in liquid nitrogen.

Diffraction data for d(CCAGGCTGCAA) and d(CCAGGCU^{Br}GCAA) were collected at the Advanced Photon Source (APS) 24-ID-C. Diffraction data for d(CCAGGCTGC^{Br}AA) were collected at APS 22-BM.

Structure determination

Data processing for d(CCAGGCTGCAA) and the U7-Br derivative, d(CCAGGCU^{Br}GCAA), was carried out in XDS (56) and Aimless (57,58). Diffraction data for the C9-Br derivative, d(CCAGGCTGC^{Br}AA), were indexed and integrated using iMosflm (59). In both derivative datasets, initial phases were determined by single-wavelength anomalous dispersion (SAD) phasing, using CRANK2 (60) and SHELX (61) in CCP4i2 (62). Two bromine sites were identified in each map, which enabled model building of two chains of each derivative in Coot (63). Subsequent refinement was carried out in Refmac (64,65). The refined U7-Br derivative structure was used as a molecular replacement search model in Phaser (66) for the native oligonucleotide. Further refinement was carried out in Refmac and additional model building was performed in Coot. The PDB-REDO (67) web server was used to conduct *k*-fold cross-validation of R_{free} values on all three structures and to generate the final models. Final refinement statistics are shown in Supplementary Table S1.

Nuclear magnetic resonance (NMR) spectroscopy

NMR data were acquired on a Bruker Avance III 600-MHz spectrometer equipped with a Cryo-TCI probe. The 11-mer

oligonucleotide was prepared at 500 μM in 30 mM sodium cacodylate buffer at pH 6.0 containing 40 mM BaCl$_2$ and 7% D$_2$O. The sample was lyophilized and dissolved in 100% D$_2$O for subsequent experiments. For both samples, a combination of 2D-NOESY and 2D-TOCSY experiments were performed at 10°C, in which the mixing time was set to 100 ms for the 2D-NOESY and 90 ms for the 2D-TOCSY. The oligonucleotide sequential assignment was conducted using the Computer Aided Resonance Assignment (CARA) program (68).

Circular dichroism (CD) spectroscopy

CD spectra were acquired using the Jasco J-810 spectropolarimeter fitted with a thermostated cell holder. Samples were prepared in 30 mM sodium cacodylate buffer at pH 6.0 or 7.4 containing varying concentrations of BaCl$_2$ or 100 mM monovalent (KCl or NaCl) cations. The 11-mer and 22-mer oligonucleotides were prepared at final DNA concentrations of 100 and 75 μM, respectively. Samples were equilibrated for 12–18 h at 4°C prior to the acquisition of the spectra. All spectra were collected at room temperature from 200 to 320 nm with a data pitch of 1.0 nm. For melting experiments, the sample was allowed to dwell for 7 min at the temperature set point.

Thermal denaturation

UV melting spectra were acquired using the Cary100 Bio UV–visible spectrophotometer equipped with a 12-cell sample changer and a Peltier heating/cooling system. The sample chamber was purged with N$_2$ throughout both melting and annealing data collection runs. Samples were prepared in 30 mM sodium cacodylate buffer at pH 6.0 supplemented with 40 mM BaCl$_2$ or 100 mM monovalent (KCl or NaCl) cations. The 11-mer and 22-mer oligonucleotides were prepared at final DNA concentrations of 14.4 and 7.25 μM, respectively. Samples were equilibrated for 15–20 h at 4°C prior to the acquisition of the spectra. Samples were transferred to self-masking quartz cuvettes with 1 cm path length for UV absorbance measurements. All spectra were collected at 260 nm. An initial fast heating ramp from 4°C to 95°C at 10°C/min was done. Data were collected every 1°C during a slow cooling ramp from 95°C to 4°C at 1°C/min and a subsequent slow heating ramp at the same temperature range and rate. Thermal melting analyses and curve fitting were conducted using MATLAB.

RESULTS AND DISCUSSION

Overview

As part of a screen to probe the structural diversity of DNA, we have crystallized many short DNA oligonucleotides, including d(CCAGGCTGCAA). Its structure was determined by single-wavelength anomalous dispersion using a 5-Br-deoxyuridine substitution at the T7 position. Initial phases from this derivative were used to create electron density maps for the higher resolution native structure (Supplementary Table S1). Refined native and derivative structures were virtually identical, with an RMSD of 0.377 Å for all DNA atoms of the asymmetric unit. The asymmetric unit
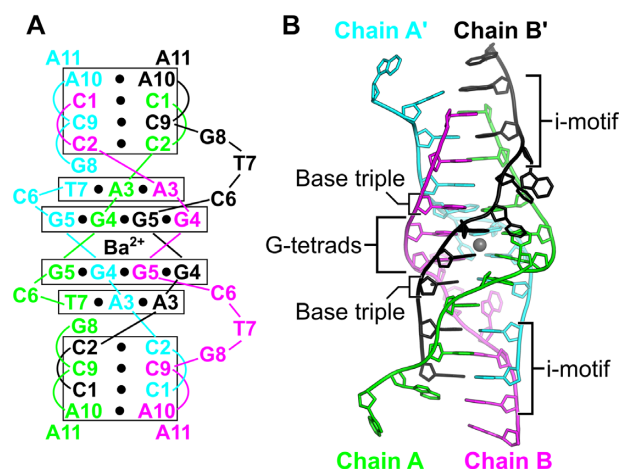

Figure 1. A G-quadruplex/i-motif hybrid structure formed from d(CCAGGCTGCAA). (**A**) Secondary structure of interactions formed between two symmetry-related dimers. Black circles represent hydrogen bonding interactions. Chains A and A' are in green and cyan, respectively. Chains B and B' are in magenta and black, respectively. (**B**) Cartoon representation of the hybrid quadruplex with labeled features. The gray sphere represents a barium ion.

contains two oligonucleotides (Chains A and B) that interact as a dimer. The two monomers show a large degree of structural similarity in the first five residues (RMSD, 0.757 Å for 84 atoms), with the largest deviation arising from the differing conformations of the A3 nucleobase (Supplementary Figure S1A). However, the latter half of the chain contains significant conformational differences in both the backbone and nucleobase atoms (Supplementary Figure S1B). Two dimers interact through crystal symmetry (symmetry molecules designated as Chains A' and B') to form a tetramer. This tetramer contains a number of distinct structural motifs, including a central G-quadruplex, a base triple interaction, a structurally variable spacer region, and a terminal i-motif (Figure 1).

A barium-stabilized G-quadruplex

The central G-quadruplex is composed of two symmetrically equivalent G-tetrads, each of which is formed through two G4 and two G5 residues (Figure 2). The two dimers are antiparallel with respect to each other, with G4-G5 dinucleotide steps along each strand, leading to heteropolar stacking between the two G-tetrads. The G-tetrads are arranged in the *abab* topology (15,69,70). Like other antiparallel G4s, the tetrad adopts *syn-anti-syn-anti* glycosidic angles with residue G4 in *syn* and G5 in *anti* for both chains. The observed base pair and base step geometries are comparable to other quadruplex structures containing only two G-tetrads with the same topology (71–73). This arrangement gives rise to two grooves of distinct widths (69). The G5–G5 phosphate distances across the narrow and wide grooves are 12.78 and 19.15 Å, respectively.

The eight guanosines coordinate directly with a central cation that is located between the two G-tetrad planes. Both the native and U7-Br oligonucleotides were crystallized in the presence of barium chloride, and a strong (11 σ) anomalous difference electron density peak between the two G-
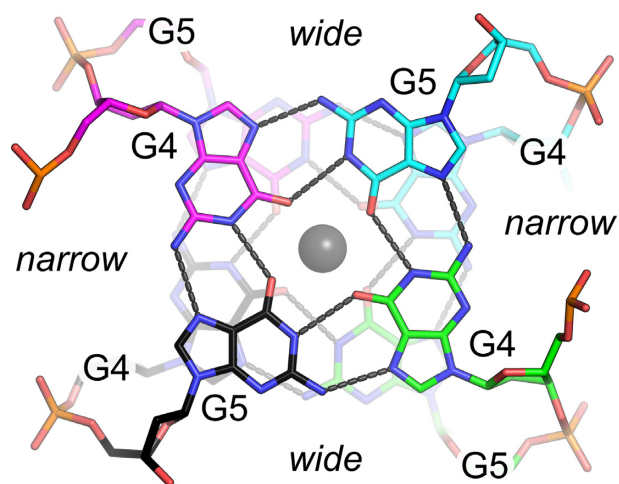
Figure 2. G-quadruplex. Top view of two stacked G-tetrads. Hydrogen bonds are indicated by gray dashes. The gray sphere represents a barium ion. The wide and narrow grooves are indicated.

tetrads was observed in both native and derivative structures. This peak is most consistent with Ba^{2+}, given the crystallization conditions and data collection energy (Supplementary Figure S2). The Ba^{2+} ion lies on or very near a crystallographic symmetry axis with a refined final occupancy of 0.50 and B-factors of 73.27 $Å^2$. The coordination distances between the cation and guanosine O6 positions range from 2.5 to 2.8 Å, with an average distance of 2.63 Å. This average is slightly shorter than the ~2.75 Å average coordination distance observed in previous examples of G-tetrads stabilized by Ba^{2+} (74,75). The apparent shorter metal-oxygen coordination may be the result of several factors, including difficulty in refining the cation residing near a special position. Alternatively, this more compact arrangement of guanosine residues could be a structural preference arising from the fewer base stacking interactions on either side of the two G-tetrads.

Reverse-Hoogsteen base triple

Flanking each side of the G-quadruplex is an A–A–T base triple. This noncanonical base triple involves both A3 residues from the dimer and T7 from Chain A' (Figure 3A). The A3–A3 base pair is formed through the Watson–Crick face of Chain A and the Hoogsteen face of Chain B, which adopts a *syn* glycosidic torsion angle to facilitate the N1–N6 and N6–N7 hydrogen bonds. The base triple is completed by interactions between A3 of Chain A and T7 from Chain A'. This is a reverse Hoogsteen base pair through the N6–O2 and N7–N3 hydrogen bonds. The *syn* glycosidic angle of A3 from Chain B allows the Watson-Crick face to make direct hydrogen bonding contacts with phosphate oxygens of T7 from Chain B' of the tetramer (Figure 3B). With both N1 and N6 of A3 in hydrogen bonding distance with the non-bridging phosphate oxygens, this arrangement suggests protonation of the N1 position to serve as a hydrogen bond donor. Similar to observations in RNA structures, the electrostatic stabilization between the localized positive charge following N1 protonation and the negatively charged phosphate would facilitate this pK_a perturbation (76).

Surprisingly, there is little direct nucleobase stacking between the G-tetrad and the A–A–T base triple. Rather, the adenosine and thymidine residues are largely positioned between the tetrad guanosines (Figure 3C, Supplementary Figure S3A). This is in contrast to the only other example of a base triple flanking one side of a G-quadruplex containing two G-tetrads of the same topology (77). In this case, the 22-nucleotide-long d[AGGG(CTAGGG)₃] contains a C–G–A base triple that forms significant stacking interactions with the G-tetrad (Supplementary Figure S3B). The large differences in stacking interactions between the triples and the tetrads suggest significant structural variability in these types of interactions based on intrinsic sequence differences and local structural constraints.

Variable spacer region

The most distinct structural differences between the two molecules of the asymmetric unit are in residues C6, T7, and G8 that collectively make up the spacer regions between the central G-tetrad and the peripheral i-motif. Interestingly, these residues have contrasting functional roles in the overall architecture of the tetramer. C6 of Chain A forms a single hydrogen bond with the G5 (N3–N2) from Chain B' and is tucked into the quartet's wide groove. C6 in Chain B does not form any base pairing interactions within the tetrameric structure. It is bulged from the tetramer core and serves primarily in mediating crystal contacts through base stacking interactions with the sugar of C6 from Chain A' and with the nucleobase of A11 from a symmetry-related dimer. As described above, T7 of Chain A is involved in base triple interactions. In contrast, T7 of Chain B is not involved in any base pairing interactions within the tetramer. Instead, this bulged residue base pairs with A11 from a symmetry-related molecule through standard Watson–Crick pairing interactions to stabilize crystal packing. The G8 residues of both molecules are unpaired. In Chain A, G8 is positioned within the nucleobase core, stacking with A3 of the base triple on one face and with the C2–C2⁺ base pair on the other (Supplementary Figure S4A). However, the G8 residue in Chain B is flipped out from the core, where it stacks with A11 of Chain A from an adjacent symmetry-related molecule to serve in crystal lattice packing contacts (Supplementary Figure S4B). This stacking is facilitated by A11 adopting a *syn* glycosidic angle, leading to partial stacking of both the pyrimidine and indole rings of the two purines.

These three residues from the parallel-stranded dimer have distinct functions within the structure. In Chain A, they form an integral part of the tetrameric structure, while the same residues in Chain B serve primarily as a bulged spacer that mediates crystal contacts. Because they have the same sequence, either strand could presumably take the role of the structural or bulged strand in solution. Though we cannot rule out the possibility of dynamic switching of these roles within the tetramer, there are several structural clues that suggest that this strand preference may arise at the time of assembly. The base triple interaction provides asymmetry between the parallel strands. This is seen in both the base pairing interactions with T7 and the *syn* A3 hydrogen bonding interactions with the phosphate from an antiparallel partner. These interactions bring the phosphate toward the
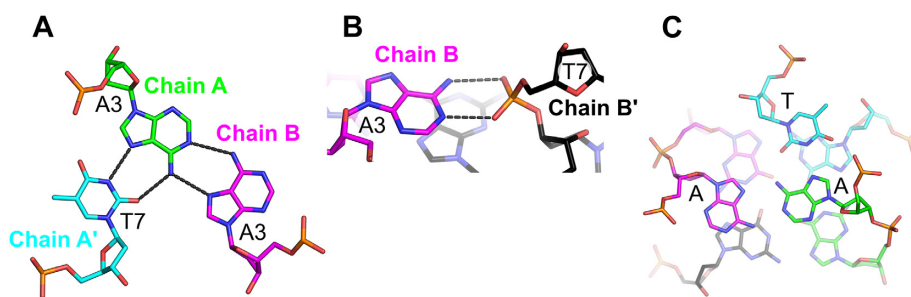
Figure 3. Base triple. (**A**) Residue A3 from Chain A (green) forms a reverse Hoogsteen base pair with T7 from Chain A' (cyan) through the A(N6)-T(O2) and A(N7)-T(N3) hydrogen bonds and interacts with A3 from Chain B (magenta) through the Watson-Crick/Hoogsteen faces. (**B**) Residue A3 from Chain B hydrogen bonds with the phosphate oxygens of T7 from Chain B'. Hydrogen bonds are indicated by gray dashes. (**C**) Top view of the A–A–T base triple above one G-tetrad.

stacked tetramer core and bias that partner strand toward bulging its nucleobases outward as found in the spacer. Additionally, the sequestration of the structural T7 in the base triple interaction would strongly bias the following nucleotide, G8, toward being stacked within the tetramer core.

i-motif and 3′-terminal nucleotides

The d(CCAGGCTGCAA) tetramer is capped at either end by i-motifs (Figure 1). The i-motif is comprised of three C–C$^+$ base pairs between C1, C2, and C9 residues of the dimers. The terminal C1–C1$^+$ base pair gives the i-motif a 5′-E topology (16). Residues C1 and C2 of both chains adopt C3′-*endo* sugar puckers, allowing the sugar-phosphate to stretch to a helical rise of 6.5 Å. This provides the necessary space to allow the C9–C9$^+$ base pair from the symmetry-related dimer to intercalate between them (Figure 4A). The geometries of the three hemiprotonated base pairs are similar, with the largest variation in the buckle and propeller angles (Supplementary Table S2), consistent with what has been observed in other i-motifs (33,78). Complete base pair and base step parameters are listed in Supplementary Table S2. Like the G-tetrads, the i-motif creates two grooves of dramatically different widths. The wide grooves are generated by the backbones of the parallel base paired strands and the narrow grooves are formed between one parallel-stranded dimer and the intercalated dimer.

Along with the C–C$^+$ interactions, a noncanonical A10-A10 base pair caps the i-motif. The capping of the 5′-E i-motif by noncanonical (i.e., A–A, T–T) base pairs has been observed in previous examples of i-motif structures (33,36,79). A large ζ angle between C9 and A10 in Chain A moves residue A10 away from the helical core, preventing direct stacking interactions between A10 and the intercalated C1 (Figure 4B). This creates a strong asymmetry with respect to the neighboring C1-C1$^+$ base pair. This asymmetry is likely induced by crystals contacts, most notably those made by the subsequent A11 nucleotides. These A11 residues are not involved in i-motif-like interactions, but form stabilizing contacts with the variable bulged region of another tetrameric assembly (see above).

NMR solution structure analysis

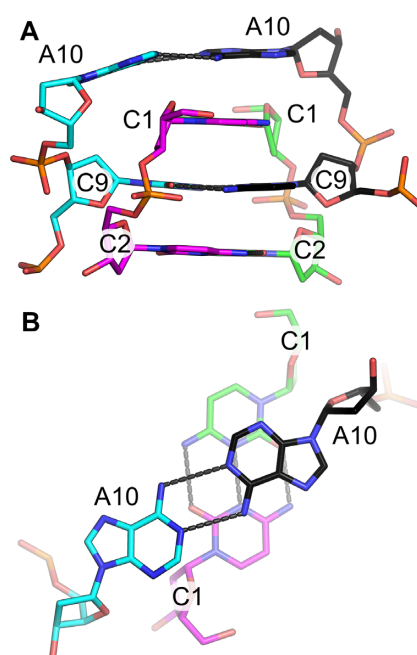We conducted 1D ^{1}H, 2D-NOESY and 2D-TOCSY experiments on the 11-mer oligonucleotide to directly assess the

Figure 4. Terminal i-motif. (**A**) Side view of three C-C$^+$ base pairs capped by an A10–A10 base pair. The C9–C9$^+$ base pair (cyan/black) is intercalated between the dimeric (green/magenta) C1–C1$^+$ and C2–C2$^+$ base pairs. (**B**) Top view of the A10–A10 base pair stacking above the C1–C1$^+$ base pair. Hydrogen bonds are indicated by gray dashes.

structure in solution. A detailed explanation of NMR assignments is provided in the Supplemental Methods. These spectra suffered from signal crowding and the appearance of multiple conformations made complete proton assignment difficult. The 1D ^{1}H profile is shown in Supplementary Figure S5. Sequential assignment allowed identification of the nucleobase, H1′, and H2′ protons in at least one conformation. Sugar-to-base connectivities were observed from C2 through G4 (Supplementary Figure S6) and from C6 through A11 (Supplementary Figure S7). Four imino proton signals were observed and assigned to T7H3, G4H1, G5H1 and C2H3 (Supplementary Figure S5/Supplementary Table S3). The C2H3 signal at 12.6 ppm is degenerate, indicating multiple conformations that are consistent with the crystal structure. These assignments allowed identification of several key structural features.

Three key structural features were confirmed by NMR analysis. First, the C2H3 signal demonstrates protonation at this position and evidence for a C–C$^+$ base pair of the i-motif. This was the only CH3 resonance observed. Typically, this proton is observed at chemical shift values near 15 ppm, though in this case there was a significant upfield shift to 12.6 ppm (Supplementary Figure S8). Cross-peaks to C2H41, C2H42, and C2H5 confirmed the assignment (Supplementary Figure S8). This large perturbation may be the result of cation-π interactions, with the localized positive charge at the C2H3 position stacking with the pyrimidine ring of G8 (Supplementary Figure S4A). The absence of C1H3 and C9H3 signals could be due to weak hydrogen bonding between the cytosines in solution, which has been previously reported in other structures consisting of multiple C–C$^+$ base pairs ([80]). NOEs confirmed the proximity of C2H3 and G8, as anticipated from the crystal structure (Supplementary Figure S8/Supplementary Table S3). Second, imino NOE cross-peaks confirmed the hydrogen bonding between T7 and A3 and additional NOEs between two independently assigned A3 residues indicated the formation of the A-A-T base triple (Supplementary Figure S9/Supplementary Table S3). Third, cross-peaks between the imino protons G4H1 and G5H1 suggest hydrogen bonding between the guanosine residues (Supplementary Figure S10/Supplementary Table S3), while resonances between guanosine H8 protons and neighboring guanosine imino and amino protons indicate their interaction through Watson-Crick and Hoogsteen faces. Importantly, these structural features were all internally consistent; NOEs were observed between the guanosines of the tetrads and multiple members of the base triple (Supplementary Figure S9 (resonances to A3), S10 (resonances to T7)), between the base triple and the unpaired G8 residue (Supplementary Figure S7), and between G8 and the C2–C2$^+$ base pair (Supplementary Figure S8). Though these NMR data do not allow independent structure determination, they are consistent with the three major base pairing motifs in the crystal structure.

Tandem repeats alter the oligomeric solution state

The crystal structure suggested that the flexibility at the A11 position could allow tandem sequence repeats to form a dimeric quadruplex, analogous to loops in bi- and unimolecular G4- and i-motif-forming sequences. We synthesized the 22-mer tandem repeat, d(CCAGGCTGCAACCAGGCTGCAA), and compared it to the 11-mer by circular dichroism (CD) and UV absorption spectroscopy. The results demonstrate that the assemblies are structurally similar, that the dimeric assembly is significantly more stable than the tetramer, and that both are preferentially stabilized by Ba^{2+}.

CD spectra of the 11-mer oligonucleotide titrated with Ba^{2+} (up to 100 mM) showed the appearance of a positive band at ∼240 nm, a strong negative band at ∼255 nm, and a weak negative band at ∼295 nm (Figure 5A). CD melting analysis suggested that these features were due to the formation of a specific structure, with nearly identical forward and reverse temperature dependence spectra (Supplementary Figure S11A). Interestingly, this was not pH-dependent, as the same profile was observed at pH 7.4 (Supplementary Figure S11B). This suggests that the hybrid G-quadruplex/i-motif structure can form at physiological pH, similar to other i-motifs ([27,81]). The 22-mer had a similar CD profile with more pronounced characteristic peaks, suggesting that the tandem repeat forms the same or similar structure as that of the 11-mer (Figure 5B). UV melting analysis showed a dramatic difference in melting temperature between these two assemblies: 41.7 ± 1.3°C for the tetrameric assembly and 73.7 ± 2.5°C for the dimeric assembly (Figure 5C). The fewer number of DNA strands in the dimeric quadruplex likely results in reduced end fraying, which can account for the apparent stability increase in melting experiments and the strength of CD signals.

Both the tetrameric and dimeric assemblies showed a preference for Ba^{2+} over monovalent cations. The spectra for both sequence lengths differed slightly in monovalent cations (K$^+$ or Na$^+$), with respect to Ba^{2+}, showing a shoulder at ∼240 nm and a negative band at ∼255 nm, but lacking the ∼295 nm negative band (Supplementary Figure S11C). These bands were largely absent for the 11-mer in buffer alone, suggesting additional cations were necessary for structure formation. Thermal denaturation experiments of the 11-mer showed no observable melting transition in conditions containing the buffer alone, whereas melting transitions were observed with additional K$^+$ or Na$^+$ (37.2 ± 3.9°C and 37.3 ± 4.1°C, respectively; Supplementary Figure S11D). The 22-mer showed comparable CD profiles between 100 mM monovalent conditions and buffer only, suggesting that the Na$^+$ cation from the cacodylate buffer was sufficient to induce some assembly. A melting transition at 56.7 ± 0.3°C was observed in buffer, while additional K$^+$ or Na$^+$ increased the T_m (62.9 ± 0.2°C and 65.3 ± 1.3°C, respectively; Supplementary Figure S11E). For both the 11-mer and 22-mer oligonucleotides, the observed T_m values in monovalent conditions were lower than that in Ba^{2+}, suggesting that the divalent cation plays a significant role in structural stability.

An alternative hybrid motif

Finally, we determined that d(CCAGGCTGCAA) can also assemble into an alternative hybrid structure. We crystallized a 5-Br-deoxycytidine substitution at the C9 position and determined its structure (Supplementary Table S1). The bromine substitution and different crystallization conditions resulted in an overall different structure, but with some similar features. The two molecules in the asymmetric unit interact with symmetry-related strands to form a hybrid quadruplex structure, in this case juxtaposing an i-motif at the 5′ end and a partial antiparallel duplex at the 3′ end (Figure 6).

In this structure, the i-motif C–C$^+$ base pairs are formed exclusively from residues C1 and C2. The four strands create a 5′-E topology, with the symmetry axis between the intercalated C2–C2$^+$ base pairs. This i-motif region is extended on either side by a homo base pairing region that includes symmetric A3–A3 (N6–N7) and G4–G4 (N1–O6) pairs. The base stacking interactions provided by these noncanonical base pairs stabilize the i-motif tertiary interactions. The brominated C9 residues are no longer involved in i-motif
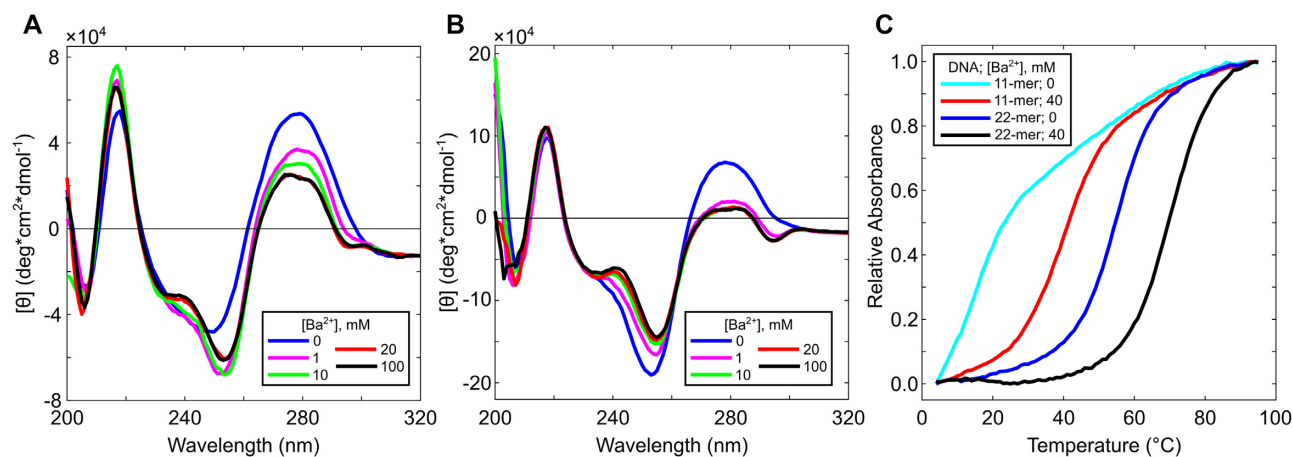
Figure 5. CD and Absorbance spectra. Ba^{2+} titration up to 100 mM at room temperature for (**A**) the 11-mer, d(CCAGGCTGCAA), and (**B**) the 22-mer, d(CCAGGCTGCAACCAGGCTGCAA). (**C**) Thermal denaturation curves from 4°C to 95°C of the 11-mer and 22-mer in sodium cacodylate buffer pH 6.0 or supplemented with 40 mM Ba^{2+}.
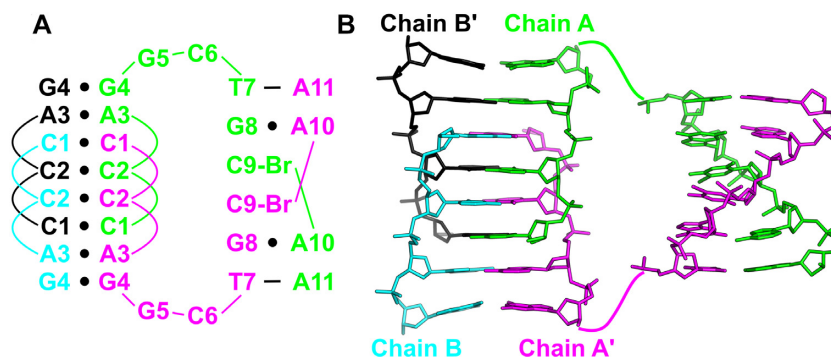

Figure 6. An i-motif/duplex hybrid structure formed from the C9-Br derivative, d(CCAGGCTGC^{Br}AA). (**A**) Secondary structure representation of interactions formed between symmetry-related strands. Black dashes represent Watson–Crick base pairs. Black circles represent noncanonical base pairs. (**B**) Stick representation of the 5′ i-motif-consisting CCAG homo base stacking motif linked to the 3′ antiparallel duplex. The i-motif and duplex are rotated 70° about the x-axis with respect to each other. Chains A and A′ are in green and magenta, respectively. Chains B and B′ are in cyan and black, respectively.

formation, and instead form Watson–Crick base pairing interactions with G5 from symmetry-related molecules that promote crystal packing. Examination of the native structure suggests that the C9 bromine substitution would preclude the formation of the hybrid G4/i-motif structure due to significant steric clashes between the bromine in Chain B and the phosphodiester backbone.

The 3′ end of the structure is a short imperfect duplex with noncanonical features. Symmetry interactions between two identical strands (Chains A and A') form an antiparallel base pairing arrangement consisting of the brominated C9 residues at the interior, each of which is immediately flanked by a G8–A10 base pair formed through N2–N7 and N3–N6 hydrogen bonding, and finally capped by a T7–A11 Watson–Crick base pair (Supplementary Figure S12). This duplex interacts with a second duplex that is formed from the other unique molecule (Chains B and B') in the asymmetric unit. The two distinct duplexes are held together by the G5–C9-Br base pair described above and by the C–G–A base triple, which is converted from the G8–A10 base pair from a single duplex (Supplementary Figure S13).

Structural and biological implications

Though previous biophysical studies characterized oligonucleotides capable of forming a parallel G4/i-motif hybrid in solution (82), the results presented here provides the first structural snapshot of a hybrid G4/i-motif. This information provides the beginnings of a structural paradigm for how these two distinct quadruplex motifs can coexist. Most notably, this structure suggests a requirement for spacer elements to separate the two base pairing motifs. These spacer elements serve to bridge the large differences in interstrand backbone distances of the two motifs. This is necessitated by an exchange of the wide and narrow grooves between the individual motifs; the G-tetrad wide groove is continuous with the i-motif narrow groove and vice versa. The variable spacer regions that include the structurally integrated base triple and unpaired guanosine allow progressive changes of interstrand backbone distances to facilitate this transition.

Biologically, this structure hints at the potential complexity of noncanonical DNA structures that may be harbored within genomes. The demonstration that the DNA studied here forms a highly stable dimeric structure from tandem

repeats suggests that longer repetitive sequences may have the ability to form complex structures, perhaps containing existing known DNA motifs. Repetitive DNA makes up >50% of the human genome (83), with microsatellite (1–10 nt), minisatellite (10 to several hundred nt), and macrosatellite (up to thousands of nt) repeats, making up ∼3% (84). Satellite DNA is involved in a variety of biological functions and pathologies (85), and repeat sequences have been implicated as drivers of evolution through the formation of noncanonical structures that result in genomic instability (86). Though there are now some examples for how repetitive DNA can impact biological function, the structural basis for this is largely unknown. The discovery of new, stable noncanonical DNA structures suggests the possibility that these repeat sequences can form complex motifs that may not be predictable from existing sequence/structure relationships.

DATA AVAILABILITY

Atomic coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes 6TZQ, 6TZR and 6TZS.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
2. Gehring,K., Leroy,J.L. and Gueron,M. (1993) A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature*, **363**, 561–565.
3. Salisbury,S.A., Wilson,S.E., Powell,H.R., Kennard,O., Lubini,P., Sheldrick,G.M., Escaja,N., Alazzouzi,E., Grandas,A. and Pedroso,E. (1997) The bi-loop, a new general four-stranded DNA motif. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 5515–5518.
4. Robinson,H., van der Marel,G.A., van Boom,J.H. and Wang,A.H. (1992) Unusual DNA conformation at low pH revealed by NMR: parallel-stranded DNA duplex with homo base pairs. *Biochemistry*, **31**, 10510–10517.
5. Kettani,A., Bouaziz,S., Skripkin,E., Majumdar,A., Wang,W., Jones,R.A. and Patel,D.J. (1999) Interlocked mismatch-aligned arrowhead DNA motifs. *Structure*, **7**, 803–815.
6. Sunami,T., Kondo,J., Kobuna,T., Hirao,I., Watanabe,K., Miura,K. and Takenaka,A. (2002) Crystal structure of d(GCGAAAGCT) containing a parallel-stranded duplex with homo base pairs and an anti-parallel duplex with Watson-Crick base pairs. *Nucleic Acids Res.*, **30**, 5253–5260.
7. Paukstelis,P.J., Nowakowski,J., Birktoft,J.J. and Seeman,N.C. (2004) Crystal structure of a continuous three-dimensional DNA lattice. *Chem. Biol.*, **11**, 1119–1126.
8. Escaja,N., Viladoms,J., Garavis,M., Villasante,A., Pedroso,E. and Gonzalez,C. (2012) A minimal i-motif stabilized by minor groove G:T:G:T tetrads. *Nucleic Acids Res.*, **40**, 11737–11747.
9. Chu,B., Zhang,D., Hwang,W. and Paukstelis,P.J. (2018) Crystal structure of a tetrameric DNA Fold-Back Quadruplex. *J. Am. Chem. Soc.*, **140**, 16291–16298.
10. Williamson,J.R. (1993) Guanine quartets. *Curr. Opin. Struct. Biol.*, **3**, 357–362.
11. Davis,J.T. (2004) G-quartets 40 years later: from 5′-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed. Engl.*, **43**, 668–698.
12. Williamson,J.R. (1994) G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 703–730.
13. Hud,N.V. and Plavec,J. (2006) In: Neidle,S and Balasubramanian,S (eds). *Quadruplex Nucleic Acids*. The Royal Society of Chemistry, Cambridge, pp. 100–130.
14. Leroy,J.L. and Gueron,M. (1995) Solution structures of the i-motif tetramers of d(TCC), d(5methylCCT) and d(T5methylCC): novel NOE connections between amino protons and sugar protons. *Structure*, **3**, 101–120.
15. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
16. Gueron,M. and Leroy,J.L. (2000) The i-motif in nucleic acids. *Curr. Opin. Struct. Biol.*, **10**, 326–331.
17. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
18. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
19. Kocman,V. and Plavec,J. (2017) Tetrahelical structural family adopted by AGCGA-rich regulatory DNA regions. *Nat. Commun.*, **8**, 15355.
20. Mukundan,V.T. and Phan,A.T. (2013) Bulges in G-Quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.*, **135**, 5017–5028.
21. Heddi,B., Martin-Pintado,N., Serimbetov,Z., Kari,T.M. and Phan,A.T. (2016) G-quadruplexes with (4n-1) guanines in the G-tetrad core: formation of a G-triad.water complex and implication for small-molecule binding. *Nucleic Acids Res.*, **44**, 910–916.
22. Jiang,H.X., Cui,Y., Zhao,T., Fu,H.W., Koirala,D., Punnoose,J.A., Kong,D.M. and Mao,H. (2015) Divalent cations and molecular crowding buffers stabilize G-triplex at physiologically relevant temperatures. *Sci. Rep.*, **5**, 9255.
23. Zhang,N., Gorin,A., Majumdar,A., Kettani,A., Chernichenko,N., Skripkin,E. and Patel,D.J. (2001) V-shaped scaffold: a new architectural motif identified in an A x (G x G x G x G) pentad-containing dimeric DNA quadruplex involving stacked G(anti) G(anti) x G(anti) x G(syn) tetrads. *J. Mol. Biol.*, **311**, 1063–1079.
24. Lim,K.W. and Phan,A.T. (2013) Structural basis of DNA quadruplex-duplex junction formation. *Angew. Chem. Int. Ed. Engl.*, **52**, 8566–8569.

25. Karg,B., Mohr,S. and Weisz,K. (2019) Duplex-Guided Refolding into Novel G-Quadruplex (3+1) Hybrid Conformations. *Angew. Chem. Int. Ed. Engl.*, **58**, 11068–11071.

26. Chambers,V.S., Marsico,G., Boutell,J.M., Di Antonio,M., Smith,G.P. and Balasubramanian,S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.*, **33**, 877–881.

27. Wright,E.P., Huppert,J.L. and Waller,Z.A.E. (2017) Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res.*, **45**, 2951–2959.

28. Mergny,J.L., Lacroix,L., Han,X., Leroy,J.L. and Helene,C. (1995) Intramolecular folding of pyrimidine oligodeoxynucleotides into an i-DNA Motif. *J. Am. Chem. Soc.*, **117**, 8887–8898.

29. Benabou,S., Avino,A., Lyonnais,S., Gonzalez,C., Eritja,R., De Juan,A. and Gargallo,R. (2017) i-motif structures in long cytosine-rich sequences found upstream of the promoter region of the SMARCA4 gene. *Biochimie.*, **140**, 20–33.

30. Fujii,T. and Sugimoto,N. (2015) Loop nucleotides impact the stability of intrastrand i-motif structures at neutral pH. *Phys. Chem. Chem. Phys.*, **17**, 16719–16722.

31. Muser,S.E. and Paukstelis,P.J. (2012) Three-dimensional DNA crystals with pH-responsive noncanonical junctions. *J. Am. Chem. Soc.*, **134**, 12557–12564.

32. Tripathi,S., Zhang,D. and Paukstelis,P.J. (2015) An intercalation-locked parallel-stranded DNA tetraplex. *Nucleic Acids Res.*, **43**, 1937–1944.

33. Esmaili,N. and Leroy,J.L. (2005) i-motif solution structure and dynamics of the d(AACCCC) and d(CCCCAA) tetrahymena telomeric repeats. *Nucleic Acids Res.*, **33**, 213–224.

34. Assi,H.A., Harkness,R.W., Martin-Pintado,N., Wilds,C.J., Campos-Olivas,R., Mittermaier,A.K., Gonzalez,C. and Damha,M.J. (2016) Stabilization of i-motif structures by 2′-beta-fluorination of DNA. *Nucleic Acids Res.*, **44**, 4998–5009.

35. Kang,C., Berger,I., Lockshin,C., Ratliff,R., Moyzis,R. and Rich,A. (1995) Stable loop in the crystal structure of the intercalated four-stranded cytosine-rich metazoan telomere. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 3874–3878.

36. Berger,I., Kang,C., Fredian,A., Ratliff,R., Moyzis,R. and Rich,A. (1995) Extension of the four-stranded intercalated cytosine motif by adenine.adenine base pairing in the crystal structure of d(CCCAAT). *Nat. Struct. Biol.*, **2**, 416–429.

37. Weil,J., Min,T., Yang,C., Wang,S., Sutherland,C., Sinha,N. and Kang,C. (1999) Stabilization of the i-motif by intramolecular adenine-adenine-thymine base triple in the structure of d(ACCCT). *Acta. Crystallogr. D. Biol. Crystallogr.*, **55**, 422–429.

38. Mir,B., Soles,X., Gonzalez,C. and Escaja,N. (2017) The effect of the neutral cytidine protonated analogue pseudoisocytidine on the stability of i-motif structures. *Sci. Rep.*, **7**, 2772.

39. Day,H.A., Pavlou,P. and Waller,Z.A. (2014) i-Motif DNA: structure, stability and targeting with ligands. *Bioorg. Med. Chem.*, **22**, 4407–4418.

40. Rhodes,D. and Lipps,H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627–8637.

41. Lam,E.Y., Beraldi,D., Tannahill,D. and Balasubramanian,S. (2013) G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.*, **4**, 1796.

42. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.

43. Henderson,A., Wu,Y., Huang,Y.C., Chavez,E.A., Platt,J., Johnson,F.B., Brosh,R.M.J., Sen,D. and Lansdorp,P.M. (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.*, **42**, 860–869.

44. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.

45. Brooks,T.A., Kendrick,S. and Hurley,L. (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.*, **277**, 3459–3469.

46. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 11593–11598.

47. Dexheimer,T.S., Sun,D. and Hurley,L.H. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.*, **128**, 5404–5415.

48. Kaiser,C.E., Van Ert,N.A., Agrawal,P., Chawla,R., Yang,D. and Hurley,L.H. (2017) 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.*, **139**, 8522–8536.

49. Palumbo,S.L., Ebbinghaus,S.W. and Hurley,L.H. (2009) Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *J. Am. Chem. Soc.*, **131**, 10878–10891.

50. Onel,B., Carver,M., Agrawal,P., Hurley,L.H. and Yang,D. (2018) The 3′-end region of the human PDGFR-beta core promoter nuclease hypersensitive element forms a mixture of two unique end-insertion G-quadruplexes. *Biochim. Biophys. Acta. Gen. Subj.*, **1862**, 846–854.

51. Piazza,A., Serero,A., Boule,J.B., Legoix-Ne,P., Lopes,J. and Nicolas,A. (2012) Stimulation of gross chromosomal rearrangements by the human CEB1 and CEB25 minisatellites in Saccharomyces cerevisiae depends on G-quadruplexes or Cdc13. *PLoS Genet.*, **8**, e1003033.

52. Abou Assi,H., Garavis,M., Gonzalez,C. and Damha,M.J. (2018) i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.*, **46**, 8038–8056.

53. Dzatko,S., Krafcikova,M., Hansel-Hertsch,R., Fessl,T., Fiala,R., Loja,T., Krafcik,D., Mergny,J.L., Foldynova-Trantirkova,S. and Trantirek,L. (2018) Evaluation of the Stability of DNA i-Motifs in the nuclei of living mammalian cells. *Angew. Chem. Int. Ed. Engl.*, **57**, 2165–2169.

54. Zeraati,M., Langley,D.B., Schofield,P., Moye,A.L., Rouet,R., Hughes,W.E., Bryan,T.M., Dinger,M.E. and Christ,D. (2018) I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.*, **10**, 631–637.

55. Niu,K., Zhang,X., Deng,H., Wu,F., Ren,Y., Xiang,H., Zheng,S., Liu,L., Huang,L., Zeng,B. *et al.* (2018) BmILF and i-motif structure are involved in transcriptional regulation of BmPOUM2 in Bombyx mori. *Nucleic Acids Res.*, **46**, 1710–1723.

56. Kabsch,W. (2010) XDS. *Acta. Crystallogr. D. Biol. Crystallogr.*, **66**, 125–132.

57. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G.W., McCoy,A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta. Crystallogr. D. Biol. Crystallogr.*, **67**, 235–242.

58. Evans,P.R. and Murshudov,G.N. (2013) How good are my data and what is the resolution? *Acta. Crystallogr. D. Biol. Crystallogr.*, **69**, 1204–1214.

59. Leslie,A.G.W. and Powell,H.R. (2007) In: Read,RJ and Sussman,JL (eds). *Evolving Methods for Macromolecular Crystallography*. NATO Science Series, Springer, Dordrecht, Vol. 245, pp. 41–51.

60. Skubak,P. and Pannu,N.S. (2013) Automatic protein structure solution from weak X-ray data. *Nat. Commun.*, **4**, 2777.

61. Sheldrick,G.M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta. Crystallogr. D. Biol. Crystallogr.*, **66**, 479–485.

62. Potterton,L., Agirre,J., Ballard,C., Cowtan,K., Dodson,E., Evans,P.R., Jenkins,H.T., Keegan,R., Krissinel,E., Stevenson,K. *et al.* (2018) CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta. Crystallogr. D. Struct. Biol.*, **74**, 68–84.

63. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta. Crystallogr. D. Biol. Crystallogr.*, **66**, 486–501.

64. Murshudov,G.N., Skubak,P., Lebedev,A.A., Pannu,N.S., Steiner,R.A., Nicholls,R.A., Winn,M.D., Long,F. and Vagin,A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta. Crystallogr. D. Biol. Crystallogr.*, **67**, 355–367.

65. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta. Crystallogr. D. Biol. Crystallogr.*, **53**, 240–255.

66. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.

67. Joosten,R.P., Long,F., Murshudov,G.N. and Perrakis,A. (2014) The PDB_REDO server for macromolecular structure model optimization. *IUCrJ.*, **1**, 213–220.

68. Keller,R.L.J. (2004) *The Computer Aided Resonance Assignment Tutorial.* CANTINA Verlag, Goldau.

69. Esposito,V., Galeone,A., Mayol,L., Oliviero,G., Virgilio,A. and Randazzo,L. (2007) A topological classification of G-quadruplex structures. *Nucleos. Nucleot. Nucl.*, **26**, 1155–1159.

70. Cang,X., Sponer,J. and Cheatham,T.E. (2011) Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes. *Nucleic Acids Res.*, **39**, 4499–4512.

71. Marathias,V.M. and Bolton,P.H. (2000) Structures of the potassium-saturated, 2:1, and intermediate, 1:1, forms of a quadruplex DNA. *Nucleic Acids Res.*, **28**, 1969–1977.

72. Mao,X., Marky,L.A. and Gmeiner,W.H. (2004) NMR structure of the thrombin-binding DNA aptamer stabilized by Sr2+. *J. Biomol. Struct. Dyn.*, **22**, 25–33.

73. Amrane,S., Kerkour,A., Bedrat,A., Vialet,B., Andreola,M.L. and Mergny,J.L. (2014) Topology of a DNA G-quadruplex structure formed in the HIV-1 promoter: a potential target for anti-HIV drug development. *J. Am. Chem. Soc.*, **136**, 5249–5252.

74. Zhang,D., Huang,T., Lukeman,P.S. and Paukstelis,P.J. (2014) Crystal structure of a DNA/Ba2+ G-quadruplex containing a water-mediated C-tetrad. *Nucleic Acids Res.*, **42**, 13422–13429.

75. Pan,B., Xiong,Y., Shi,K., Deng,J. and Sundaralingam,M. (2003) Crystal structure of an RNA purine-rich tetraplex containing adenine tetrads: implications for specific binding in RNA tetraplexes. *Structure*, **11**, 815–823.

76. Tang,C.L., Alexov,E., Pyle,A.M. and Honig,B. (2007) Calculation of pKas in RNA: on the structural origins and functional roles of protonated nucleotides. *J. Mol. Biol.*, **366**, 1475–1496.

77. Lim,K.W., Alberti,P., Guedin,A., Lacroix,L., Riou,J.F., Royle,N.J., Mergny,J.L. and Phan,A.T. (2009) Sequence variant (CTAGGG)n in the human telomere favors a G-quadruplex structure containing a G.C.G.C tetrad. *Nucleic Acids Res.*, **37**, 6239–6248.

78. Malliavin,T.E., Snoussi,K. and Leroy,J.L. (2003) The NMR structure of [Xd(C2)4] investigated by molecular dynamics simulations. *Magn. Reson. Chem.*, **41**, 18–25.

79. Phan,A.T., Gueron,M. and Leroy,J.L. (2000) The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. *J. Mol. Biol.*, **299**, 123–144.

80. Robinson,H. and Wang,A.H. (1993) 5′-CGA sequence is a strong motif for homo base-paired parallel-stranded DNA duplex as revealed by NMR analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 5224–5228.

81. Kovanda,A., Zalar,M., Sket,P., Plavec,J. and Rogelj,B. (2015) Anti-sense DNA d(GGCCCC)n expansions in C9ORF72 form i-motifs and protonated hairpins. *Sci. Rep.*, **5**, 17944.

82. Zhou,J., Amrane,S., Korkut,D.N., Bourdoncle,A., He,H.Z., Ma,D.L. and Mergny,J.L. (2013) Combination of i-motif and G-quadruplex structures within the same strand: formation and application. *Angew. Chem. Int. Ed. Engl.*, **52**, 7742–7746.

83. Saha,S., Bridges,S., Magbanua,Z.V. and Peterson,D.G. (2008) Empirical comparison of ab initio repeat finding programs. *Nucleic Acids Res.*, **36**, 2284–2294.

84. Treangen,T.J. and Salzberg,S.L. (2012) Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.*, **13**, 36–46.

85. Bagshaw,A.T.M. (2017) Functional Mechanisms of Microsatellite DNA in Eukaryotic Genomes. *Genome Biol. Evol.*, **9**, 2428–2443.

86. Xie,K.T., Wang,G., Thompson,A.C., Wucherpfennig,J.I., Reimchen,T.E., MacColl,A.D.C., Schluter,D., Bell,M.A., Vasquez,K.M. and Kingsley,D.M. (2019) DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science*, **363**, 81–84.