Encoding canonical DNA quadruplex structure

Scarlett A. Dvorkin, Andreas I. Karsisiotis, Mateus Webba da Silva*

The main challenge in DNA quadruplex design is to encode a three-dimensional structure into the primary sequence, despite its multiple, repetitive guanine segments. We identify and detail structural elements describing all 14 feasible canonical quadruplex scaffolds and demonstrate their use in control of design. This work outlines a new roadmap for implementation of targeted design of quadruplexes for material, biotechnological, and therapeutic applications.

INTRODUCTION

In the design of objects composed of nucleic acids, it is crucial to avoid the presence of alternative molecular recognition motifs, since these may result in nontargeted architectures. At the same time, quadruplex DNA is characterized by the presence of repetitive segments of guanines, which have the potential for a great variety of alternative hydrogen bond alignments. The prediction of quadruplex architectural folds encoded in guanine-rich DNA sequence is an important challenge relevant for understanding its significant regulatory roles (1), with implications on the evaluation of new therapeutics (2). The current general consensus is that predicting or controlling quadruplex folding is a mostly intractable problem. For example, biologically significant DNA sequences are often intrinsically polymorphic in vitro and respond to pH, cations, or crowding conditions. Nevertheless, there is substantial interest in resolving this issue due to the potential of these structures as functional materials (3, 4), in templated organization of materials (5), as nanowires (6), in catalysis (7), and as therapeutics (8). For example, a quadruplex-based nanomotor relying on a conformational switch between canonical quadruplex and duplex quadruplex resulted in approximately 5-nm displacement (3). Potential application of quadruplexes to nanoelectronics has been illustrated by measurements of currents greater than 100 pA in single-molecule quadruplex wires over 100 nm long (6). Direct imaging of quadruplex formation and their association with proteins have been enabled by constructing a quadruplex into a DNA origami scaffold (4). These examples illustrate the significant utility of the controlled design of quadruplexes. However, the design of quadruplexes is a complex problem also influenced by attributes of the self-assembly environment.

A major step in addressing this problem is to develop the ability to define the structural characteristics of quadruplexes. Canonical quadruplexes are composed of a single strand containing at least four tracts of two or more guanines that form a stem. These are linked by three loops, d(G_{n}L_{x}G_{n}L_{y}G_{n}L_{z}G_{n}), where n is the number of sequential guanines that form the stem, and L is the number (x, y, or z) of residues linking guanines in the stem. In the stem, a guanine from each of the four guanine tracts (G-tracts) engages in hydrogen bonds along its Hoogsteen and Watson-Crick edges to form a tetrad (also known as a quartet). In the stacked tetrads, each of the 2′-deoxyguanosines adopts one of two conformational states that relate base to sugar through the glycosidic bond: the glycosidic torsion angle \( \gamma \). These two nonoverlapping ranges are described as either syn (\(-90° \leq \gamma \leq +90°\)) or anti (\(+90° \leq \gamma \leq +180°\)) conformation.
**RESULTS**

**Propeller loops bridge parallel-stranded *synG-synG-antiG* grooves**

To evaluate the feasibility of propeller loop formation spanning grooves composed of parallel-stranded *synG-synG-antiG*, we designed the 3(d+pd) topology and determined its solution structure (Fig. 2) using nuclear magnetic resonance (NMR) spectroscopy. For experimental details, see Supplementary Materials. The sequence d(G₄T₃G₄T₃G₄) in 20 mM sodium adopts the target 3(d+pd) topology (see Fig. 2, C to F). The right-handed type 2 stem is composed of an equal number of *synG* and *antiG*. The propeller loop spans a groove composed of parallel-stranded *synG-synG-antiG*, demonstrating that this structural motif is feasible in quadruplex structure.

**The 5′ end of type 3 stem adopts *syn* conformation**

To evaluate whether a three-stacked type 3 quadruplex folds with a *synG* in the 5′ end of the stem, we designed a DNA sequence targeting the 3(−lₜₚ−lₚₚ) topology (Fig. 3A) and determined its solution structure using NMR spectroscopy. The DNA sequence d(G₄T₃G₄T₃G₄T₃G₄) (S064) folds into the 3(−lₜₚ−lₚₚ) topology in 80 mM NaCl and 20 mM NaH₂PO₄/Na₂HPO₄ (pH 6.8) (Fig. 3, C to E). Its right-handed type 2 stem includes a *synG* in its 5′ end. To evaluate whether formation of this motif is influenced by a base attached to the 5′ end of the stem, the DNA sequence d(TG₃T₃G₄T₃T₃G₄) (S093) was also found to adopt the same topology with a *synG* in the 5′ end of the stem.

**Encoding topology with conformationally locked 2′-deoxyguanosine derivatives**

To evaluate whether the formation of the 3(−lₜₚ−lₚₚ) topology in d(G₄T₃G₄T₃G₄T₃G₄) (S232) can be induced, we replaced the dG7 in this sequence with an rG d(G₄T₃G₄T₃G₄T₃G₄) (S209). The sequence S209 folds into a single species in 100 mM sodium solution at pH 6.8 (Fig. 3E) and adopts the desired 3(−lₜₚ−lₚₚ) topology.

**Loop length and number of stacking tetrads are interdependent in design**

To evaluate the hypothesized interdependency of loop length combinations and number of stacking tetrads in the programming of quadruplex topologies, we designed two-, three-, and four-stacked quadruplexes of the (lₜₚ−lₚₚ) topology with lateral loop variations as shown in Table 1 and structurally characterized them using solution NMR spectroscopy (see Supplementary Materials). Sodium solutions were selected since they favor the formation of type 3 stem (15).

The solution structures of DNA sequences d(G₄T₃G₄T₃G₄T₃G₄A₂T₄) (5J6U), d(G₄T₃G₄T₃G₄A₂T₄G₄) (2M6W), d(G₄T₃G₄T₃G₄T₃G₄) (2M6V), d(G₄T₃G₄T₃G₄T₃G₄) (5J4W), and d(G₄T₃G₄T₃G₄T₄G₄) (5J4P) were determined, and details are presented in Fig. 4. All sequences form right-handed quadruplexes with a *synG* at the 5′ end of the stem and alternation of glycosidic bonds along guanosine segments. The four-thymine diagonal loop adopts a very similar structural environment in all of them, with lateral loop residues stacking onto the stem. The third thymine of the loop stacks onto the first (5′) stem.

The glycosidic bond conformations to be adopted by guanosines throughout the stem of the targeted topology is required. This in turn enables segment length selection of residues in the primary sequence that will form loops in the final architecture. This approach has been previously applied in the design of a quadruplex architecture containing all three loop types (11). However, to render the approach more generally applicable, we addressed the remaining questions herein. Specifically, the topologies for all feasible canonical quadruplexes are described, and this knowledge is applied to the design of canonical quadruplex architectures using modified bases. We demonstrate that in the design of quadruplexes, the combination of lengths of loops is dependent not only on the arrangement of the groove widths but also on the number of stacking tetrads of the quadruplex stem.

---

**Fig. 1. Structural descriptors of canonical quadruplexes.** (A) The 2(+lₚd−lₚp) fold- ing topology and hydrogen bond alignments for its top tetrad are shown. Magenta denotes syn glycosidic bond angles, and cyan denotes anti. The gray circle indicates the 5′ end of the stem. Strand directionalities are indicated by (−) when counter-clockwise and by (+) when clockwise. Propeller loops are indicated by the symbol “p,” diagonal loops by “d,” and lateral loops by “l.” (B) Schematic representations of named high-resolution solution structures publicly available in the Protein Data Bank (PDB) (42) are shown, with corresponding PDB ID and the respective schematic of the quadruplex topology. The topology of the two-stacked thyminobining aptamer “TBA” (PDB ID: 148D) is known as a chair-type quadruplex. It can be described as a quadruplex adopting the 2(+lₚd−lₚp) topology. The two two- stacked basket-type architectures of human telomeric repeats (2KF8) and Giardia telomeric repeats (2KOW) are denoted 2(−lₚd+lₚd) and 2(+lₚd−lₚp), respectively. The three-stacked form-1 and form-2 topologies of human telomeric sequences are described by 3(−p−lₚd−lₚd) for (2KOW) and 3(−lₚd−lₚd−p) for (2JPZ), respectively. Finally, the (3+1) scaffold of 2LOD can be named 3(−p+d+lₚp).

---

Dvorkin et al., Sci. Adv. 2018; 4 : eaat3007     31 August 2018
An attempt to fold loop combination 2M6V into 3(−l_w+d_l_n) proved unsuccessful. Instead, it adopted 2(−l_w+d_l_n), with a four-residue −l_w and a two-residue +l_n (see Fig. 4). In place of the designed two-residue −l_w, the 2(−l_w+d_l_n) has a four-residue loop that incorporates guanosines of the first and second G-rich segments. Thus, the energetically favored architecture still contains the expected diagonal loop formed by four residues, but it also includes a longer (four-residue) loop for −l_w. While the guanosine of the first G-rich segment forms a synG:T mismatch, the guanosine of the second G-rich segment stacks perfectly onto its preceding antiG of the stem. This stabilization motif is also observed for two-residue (2M6W), three-residue (5J6U and 5J4W), and four-residue (2KF8 and 2KKA in Table 1) −l_w loops that proceed from a synG to an antiG of the stem. The G6:T mismatch appears to be an essential stabilizing factor, since it is absent in the analogous DNA sequences that do not fold (S025).

The −l_w loop in two- and four- but not three-stacked (−l_w+d_l_n) can be stabilized by four residues. However, 3(−l_w+d_l_n) can be stabilized by a three-residue −l_w loop. This may be due to the fact that while in 2(−l_w+d_l_n) and 4(−l_w+d_l_n) the −l_w loop progresses from an antiG to a synG of the stem, the reverse is true for 3(−l_w+d_l_n). For a −l_w loop, a three-residue synG to antiG strand progression may be considered mechanically equivalent to a four-residue antiG to synG strand progression.

**DISCUSSION**

The challenge in design of quadruplexes is to determine the optimal sequence that reliably encodes for a given three-dimensional (3D)
Nucleobase substitutions can be used to induce quadruplex folding; however, not all site substitutions will result in well-folded species (31), suggesting that strategies for their use require further development. Guanine to 8-bromoguanine substitution has been shown to induce the folding of synG at the substituted site (21, 24, 32, 33), and riboguanosine substitution leads to the folding of antiG (34, 35). Our unsubstituted sequence d(G3T3G3T4G3T2G3) failed to adopt the expected 3(−lwd+ln) topology (Table 1), but substitution of a structure. Therefore, knowledge of structural motifs of the desired architecture is fundamental to this process. Twenty-six theoretical quadruplexes have been previously proposed (9, 11, 14); however, a fragment-based molecular mechanics approach applied to the derivation of three-stacked quadruplexes predicted that only 14 of these are mechanically feasible (17). Here, we use known and hypothesized glycosyl bond conformations to design 3D architectures to (i) provide experimental verification of all feasible glycosyl bond conformations in these 14 quadruplex topologies and (ii) demonstrate interdependency between loop-length combination and the number of stacking tetrads of quadruplexes.

In Fig. 5, the schematic representations of glycosidic bond conformations for 14 feasible canonical quadruplex topologies are summarized. Only four of nine type 2 quadruplexes have been experimentally verified: 3(−p−lwd−ln) (18), 3(−lwd−ln−p) (19), 3(−p+lwd) (20), and 3(−lwd+p+ln) (21). In all four of these topologies, propeller loops bridge grooves of parallel-stranded synG-antiG-antiG, demonstrating that formation of the previously hypothesized (11), but as yet unobserved, topologies 3(−lwd−p) (30), 3(−p−p−lwd) (30), and 3(−p+lwd+p) is possible. The two remaining type 2 topologies, 3(−p+ln) and 3(d+pd), have a propeller loop bridging a medium groove of parallel-stranded synG-synG-antiG. The propeller loop spanning this groove has been confirmed in this study, thus verifying the hypothesized glycosyl bond conformation. Therefore, this structural motif permits formation of the yet to be observed 3(−p+ln). We propose here that type 2 quadruplexes are restricted to two- and three-stacked topologies, since formation of propeller loops for four-stacked canonical type 2 quadruplexes is improbable.

All four type 3 quadruplex topologies shown in Fig. 5 have been experimentally verified. For all two- and four-stacked, the 5′ end of the stem is synG: (+lwd−ln) (22), (−lwd+ln) (23, 24), (+lwd+ln) (25, 26), and (−lwd−ln−lwd) (27). In these architectures, more syn-anti steps are present than the less stabilizing (28) anti-syn steps. However, a single three-stacked structure exists with an antiG in the 5′ end of the stem: the human telomeric sequence d[AG2(T2AG3)3] (29), which adopts the topology 3(−lwd+ln) in 100 mM sodium solution at pH 6.8. Nonetheless, a synG conformation has been suggested to be more stabilizing for a quadruplex stem (30). Furthermore, in any three-stacked type 3 quadruplex, the number of syn-anti steps will be the same as that of anti-syn, regardless of the conformation adopted by the 5′-stem guanosine. All seven sequences designed to fold 3(−lwd+ln) topology were observed to adopt a synG at the 5′ end of the quadruplex stem, thus demonstrating it to be the most stable conformation. Other type 3 topologies may also adopt two alternative dispositions of glycosyl bond conformation. The preference for antiG at the 5′ end of the stem observed in the human telomeric sequence may outmatch the general preference for synG through the additional stacking of triads or other mismatch alignments onto the type 3 stem.

Individual loop length does not define successful design; it is their combination that does so. For example, although four residues in the first or one residue in the last loop each separately allow folding of (−lwd+ln), the combination of the two in a single sequence does not. The difference between lateral loops of 2- and 3(−lwd+ln) can be up to two residues, but only one residue for 4(−lwd+ln). Also, only 4(−lwd+ln) is able to fold with lateral loops of one or two residues.

In addition, we determined that the number of stacking tetrads of the targeted topology predicates the optimal loop-length combination. Combinations of two and three, or just three-residue segments for the lateral loops enabled reliable assembly of any n(−lwd+ln) topology. However, greater versatility is permitted for the design of 2(−lwd+ln) architectures. These can fold with a −lwd of four residues, in contrast to 3(−lwd+ln), and can accommodate a larger difference of residues between lateral loop lengths, in contrast to 4(−lwd+ln).

Fig. 3. Alternative conformations of glycosyl bonds in a quadruplex stem are possible. (A) Schematic representations of the alternative sequence of glycosidic bond angles in the 3(−lwd+ln) topology for PDB ID 143D (left) and the designed 5J05, as well as corresponding groove-width combinations (B) in the stem. (C) A bundle of solution structures adopted by the DNA sequence 5J05 in 100 mM sodium solution at pH 6.8. (D) The capping of the diagonal onto the (G1:G9:G20:G14) tetrad, with the arrow indicating the position of synG1. (E) Detail of the intrastrand stacking of the interdigitated adenine stacking onto antiG18 of the stem and T6. (F) Design of the 3(−lwd+ln) topology by replacing the nucleoside dG7 of the DNA sequence S232 by an rG in S209, as shown in red in the schematics for the topology. In the expansion of the proton NMR spectrum in 0.1 M sodium solutions at 5°C the imino protons in the spectrum of S232 appear predominantly as “hump” at approximately 10.8 ppm, indicating that the sequence is mostly unfolded. Nineteen of 21 possible imino protons in the DNA sequence appear in the spectrum of S209, indicating formation of a three-stacked quadruplex fold. The topology has been structurally characterized (Supplementary Materials).
we chose the position at the end of the first loop. This exemplifies how identification of the glycosidic bond conformation of guanosines in the stem enables successful design. The structural details of canonical quadruplex topologies provided single antiG of 3(−lwd+ln) by the conformationally locked rG enabled successful folding of the targeted topology. Although a number of antiG positions could have been selected for this substitution, we chose the position at the end of the first loop. This exemplifies how identification of the glycosidic bond conformation of guanosines in the stem enables successful design.

The structural details of canonical quadruplex topologies provided single antiG of 3(−lwd+ln) by the conformationally locked rG enabled successful folding of the targeted topology. Although a number of antiG positions could have been selected for this substitution, we chose the position at the end of the first loop. This exemplifies how identification of the glycosidic bond conformation of guanosines in the stem enables successful design.

The structural details of canonical quadruplex topologies provided single antiG of 3(−lwd+ln) by the conformationally locked rG enabled successful folding of the targeted topology. Although a number of antiG positions could have been selected for this substitution, we chose the position at the end of the first loop. This exemplifies how identification of the glycosidic bond conformation of guanosines in the stem enables successful design.
each topology. This will inform both experimental and theoretical approaches, resulting in improved reliability and reproducibility in quadruplex design. Many principles of design are already established that make it possible to engineer the architectures described. For example, in canonical quadruplex architectures, positioning of loops (as first, second, and third) is fundamental in sequence design, that is, reversing the loop-length sequence in the combination of loops does not result in the same topology (36). The nature of residues able to realize the desired loop combinations is equally significant. Here, we used predominantly thymines, and less often adenines, while avoiding cytosines to prevent formation of G:C base pairs. Inclusion of purines in loops may be used to facilitate tetrad stacking and formation of hydrogen bond alignments that should also stack onto tetrads of the quadruplex stem. Successful formation of type 1 quadruplexes is favored in potassium, as well as type 2 and type 3 in sodium solutions (15). Potassium can also be used to fold type 2 and type 3 architectures.

Here, we have used four residues to generate diagonal loops. Molecular dynamics (MD) simulation studies (28, 37) and experimental studies (15) show that formation of propeller loops is favored by one or two residues. A combination of at least two such loop lengths is likely to be successful for type 1 topology. Loops of two and three residues favor (l, 0) over (l, 0), while four-residue loops favor (l, 0) (28). However, lack of loop selectivity for single propeller loops and clockwise lateral loops present unique challenges in design. To mitigate these problems, successful design can be further enabled by oligonucleoside modification strategies as demonstrated in this study. However, more general solutions can be sought from a greater understanding of the prefolding states and folding pathways. This will allow for the development of strategies to control thermodynamic parameters modulating the kinetic routes for self-assembly.

Here, we describe the feasible canonical quadruplex topologies and structural requirements for their design. These scaffolds can be used to inform the design of quadruplex architectures, as well as fine-tune a desired topology for specific applications.

Potential applications in therapeutics and nanotechnology of DNA quadruplexes are varied and diverse. Quadruplex-forming DNA sequences are highly prevalent in mammalian and bacterial genomes where they have established regulatory roles. Small molecules that can stabilize these architectures may be used as therapeutics and sensors (2). The structural diversity library based on the set of 14 topologies described here can be interrogated by small-molecule combinatorial libraries for the discovery of leads against the quadruplex topologies represented. Conversely, the quadruplex structural diversity library can also be used to “fish” for protein interaction partners in nuclear extracts to identify quadruplex binding proteins. Quadruplex topologies identified in this manner can then be further fine-tuned for the development of therapeutics, biomarkers, or other diagnostic purposes.

Knowledge of the 14 feasible canonical topologies allows for selection of quadruplex type with structural characteristics appropriate for the loading of desired payloads. These topologies may have potential for use in future drugs for delivery to target locations through stimuli-responsive conformational change involving quadruplex structure and random single-stranded sequence. For example, superparamagnetic nanoparticles surface-functionalized with canonical quadruplexes carrying a payload for intracellular delivery can be thermally activated when subjected to an alternating magnetic field (38). Design of quadruplexes can be tailored to create temperature-dependent unfolding (payload delivery) dependent on a range of desired temperatures. The approach can also be used for the construction of devices for targeted delivery to solid tumors.

Quadruplex nanodevices are known to be able to sense cations, small molecules, and proteins [reviewed in (39)]. Reversible quadruplex folding can therefore also be controlled through photoregulation or presence of specific cations. For example, in the presence of hemin, quadruplexes are able to catalyze hydrogen peroxide–mediated oxidation of 2,2′-azinobis(3-ethylbenzothiazoline-6-sulfonic acid) to produce a color change which leads to chemiluminescence. Target recognition by the quadruplex is thus observed colorimetrically or through a fluorescence signal. The added ability to tune quadruplex structure in this context will be a powerful tool for a variety of
applications, which are expected to include the identification of pathogens and development of diagnostics and biomarkers.

The availability of feasible canonical topologies can also be used to inform the rational design of well-defined quadruplex nanowires (40). In principle, systematic design should allow for the modulation of electronic and photoelectronic properties of these materials due to their π-stacking system. This should ultimately enable the fine-tuning of properties that make these architectures good candidates for integrated multicomponent systems.

In summary, programming canonical quadruplex structure based on knowledge of the structural characteristics of the stem has been demonstrated. The complete description of all glycosyl bond angle conformations for canonical quadruplex topologies is presented for the first time. We also demonstrate the feasibility of alternative glycosyl bond conformations within a quadruplex topology. Optimal individual loop lengths have been shown to be dependent on the context of other loop types, as well as loop lengths, within the sequence. Finally, we have shown that the target number of stacking tetrads influences the combination of loop lengths required for successful folding. The improved ability presented here to reliably engineer high-fidelity, topology-specific quadruplex architectures enables the exploitation of this molecule’s unique potential.
Typically, (1H-31P)HSQC (heteronuclear single quantum correlation) NOESY experiments allowed for GH1 and GH21/GH22 assignments. JR-NOESY (60 and 200 ms) were used to derive distance restraints from the initial buildup rates of the NOE curves by the two-spin connectivities on H2/H2″ and H3/H5″(i) were acquired from the aromatic H8/H6(i)-H1′(i)-H8/H6(i+1) and the corresponding sequential connectivities in thymine residues, that is, 2.46 Å. Only two limiting mixing times average of the volume integral of the distance between H5-methyl protons from the initial buildup rates of the NOE curves by the two-spin connectivities involving H3(i-1)-P(i)-H4/i)-H5/H5″(i) were acquired to support or identify residue sequential connectivities. In selected cases, unambiguous assignment of imino H1 from the aromatic H8 guanines in the stem was also performed from natural abundance JR [1H,13C] HMBC (heteronuclear multiple-bond correlation) experiments. The chemical shift assignments are shown in tables S1 to S7.

Solution NMR assignments
Proton assignments were performed following well-established procedures and, in some cases, aided by inosine substitutions. Identification of intranucleotide anomeric signals was derived from DQF-COSY (double quantum-filtered correlation spectroscopy) and TOCSY (total correlation spectroscopy) experiments. G-quadruplex sequence-specific assignments were based on sequential nuclear Overhauser effect (NOE) connectivities of the type H8/H6(i)-H1′(i)-H8/H6(i+1) and the corresponding sequential connectivities on H2/H2′ and H3′ spin systems derived from nuclear Overhauser effect spectroscopy (NOESY) experiments. JR-NOESY experiments allowed for GH1 and GH21/GH22 assignments. Typically, (1H-31P)HSQC (heteronuclear single quantum correlation) experiments that allow for tracing intranucleotide spin system connectivities involving H3(i-1)-P(i)-H4/i)-H5/H5″(i) were acquired to support or identify residue sequential connectivities. In selected cases, unambiguous assignment of imino H1 from the aromatic H8 guanines in the stem was also performed from natural abundance JR [1H,13C] HMBC (heteronuclear multiple-bond correlation) experiments. The chemical shift assignments are shown in tables S1 to S7.

Restrains for structure calculations
Distance restraints were typically derived from NOESY experiments in "100% "H2O at three to five mixing times, and distances estimated from the initial buildup rates of the NOE curves by the two-spin approximation: \[ r_{ij} = r_{\text{ref}} \left( \frac{R_{ij}}{R_{\text{ref}}} \right)^{1/6}, \] where \( r_{ij} \) is the distance between protons \( i \) and \( j \), \( r_{\text{ref}} \) is a reference distance, and \( R_{ij} \) and \( R_{\text{ref}} \) are the initial buildup rates. Interproton distances were estimated using the average of the volume integral of the distance between H5-methyl in thymine residues, that is, 2.46 Å. Only two limiting mixing times (60 and 200 ms) were used to derive distance restraints from the exchangeable protons collected with jump-and-return NOESY spectra at 5°C in 90% H2O, 10% D2O. Distances were assumed to be 3.0 ± 0.8 Å for strong peaks observed in the 60-ms mixing time spectrum, 4.0 ± 1.2 Å for medium cross-peaks observed in the 200-ms mixing time spectrum, and 5.0 ± 1.8 Å for cross-peaks not observed in a 60-ms mixing time spectrum.

Structure calculations
Distance-restrained structure determinations were carried out using distance constraints from the NMR data. Calculations were performed using XPLOR-NIH (41) using the CHARMM force field and adapted for restrained MD (rMD) for nucleic acids. All calculations were executed in vacuo without explicit counterners. Typical-
Table S14. NMR restraints and structural statistics for the structures of 2M6V.

Table S13. NMR restraints and structural statistics for the structures of 5J4P.

Table S11. NMR restraints and structural statistics for the structures of 5J05.

Table S9. NMR restraints and structural statistics for the structures of 2M6W.

Table S3. Proton and phosphorous chemical shifts for structure of 5J6U.
Acknowledgments

Funding: This research was supported by a BBSRC grant (number BB/H005692) to M.W.d.S. and a Vice-Chancellor Research Studentship (University of Ulster) to S.A.D. We acknowledge use of the open-access NMR facilities: (i) at HWB-NMR, and the associated Wellcome Trust Biomedical Resources (083796/Z/07/Z); (ii) through EU FP7 EAST-NMR (228461) at the Slovenian NMR Center. Author contributions: M.W.d.S. conceived the project. M.W.d.S. and S.A.D. developed the concepts. M.W.d.S., S.A.D., and A.I.K. designed the primary sequences. A.I.K., S.A.D., and M.W.d.S. acquired the data. S.A.D., A.I.K., and M.W.d.S. analyzed the experimental data. S.A.D. and M.W.d.S. performed the structure calculations. S.A.D., M.W.d.S., and A.I.K. drafted the manuscript. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 12 February 2018
Accepted 24 July 2018
Published 31 August 2018
10.1126/sciadv.aat3007

Citation: S. A. Dvorkin, A. I. Karsisiotis, M. Webba da Silva, Encoding canonical DNA quadruplex structure. Sci. Adv. 4, eaat3007 (2018).