ACS OMEGA

Open Access

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

Article

Entangled World of DNA Quadruplex Folds

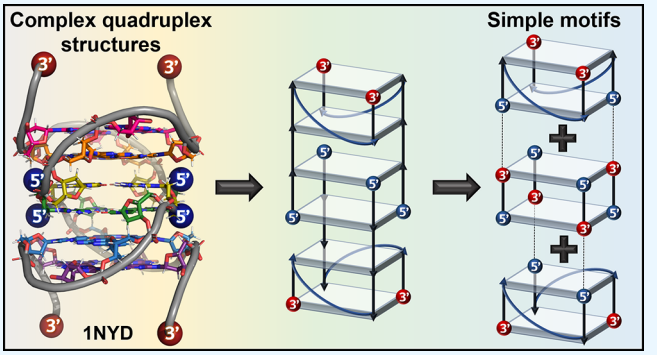
Sruthi Sundaresan, Patil Pranita Uttamrao, Purnima Kovuri, and Thenmalarchelvi Rathinavelan*

Cite This: *ACS Omega* 2024, 9, 38696−38709

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: DNA quadruplexes participate in many biological functions. It takes up a variety of folds based on the sequence and environment. Here, a meticulous analysis of experimentally determined 437 quadruplex structures (433 PDBs) deposited in the PDB is carried out. The analysis reveals the modular representation of the quadruplex folds. Forty-eight unique quadruplex motifs (whose diversity arises out of the propeller, bulge, diagonal, and lateral loops that connect the quartets) are identified, leading to simple to complex inter/intramolecular quadruplex folds. The two-layered structural motifs are further classified into 33 continuous and 15 discontinuous motifs. While the continuous motifs can directly be extended to a quadruplex fold, the discontinuous motif requires an additional loop(s) to complete a fold, as illustrated here with examples. Similarly, higher-order quadruplex folds can also be represented by continuous or discontinuous motifs or their combinations. Such a modular representation of the quadruplex folds may assist in custom engineering of quadruplexes, designing motif-based drugs, and the prediction of the quadruplex structure. Furthermore, it could facilitate understanding of the role of quadruplexes in biological functions and diseases.



INTRODUCTION

Besides the right-handed double helical structure,[1,2] nucleic acids can fold into a variety of secondary structures like hairpin, triplex, quadruplex, i-motif, *etc.*[3−9] Multiple pieces of evidence unfold the instrumental role of quadruplex in regulating the biological processes across all the domains of life.[10−22] This class of thermodynamically stable alternative structure encompasses four guanine (Figure 1A) rich strands[23] (thus the name quadruplex or tetraplex), stabilized by planar G-tetrads (Figure 1B,C) that stack on each other (Figure 1D,E). The inward orientation of the carbonyl groups of the guanines in the G-quartet creates an electronegative repulsion. Thus, there is a requirement for mono- or divalent cations to stabilize the G-quartets.[24,25] Although the quartet structure was originally found to be formed only by four guanines, recent evidences show the possibility of non-G-quartets.[26,27]

Sequences rich in Gs (which are prone to form thermodynamically favorable quadruplex structures) are found to be prevalent in the human genome[18,19,28,29] and have important regulatory roles that are influenced by their genomic location.[11] For instance, sequences prone to form quadruplexes are enriched (nearly 50%) in the human gene promoters,[29] suggestive of their role in the genetic transactions,[23,30] and also play a vital role in governing gene expression and genomic stability.[18,19,28,29] Notably, the quadruplex structures formed in the promoter regions of oncogenes like KRAS,[31] c-Myc,[32] *etc.* negatively regulate them. These structures are also present in the telomeres of eukaryotes[33] and are shown to inhibit telomere extension.[34] Quadruplex structures have been identified in the recombination hotspots of the human genome, insisting on their role in recombination.[35] The prevalence of quadruplex-forming sequences in the human transcriptome further facilitates the formation of RNA quadruplex structures.[27] RNA quadruplexes take part not only in mRNA cap-dependent[36] and independent[37] translation but also in alternative splicing,[38] mitochondrial transcription termination,[39] mRNA localization,[40] and maturation of miRNAs.[17,41] Besides independently forming quadruplex structures, RNA and DNA can together form an RNA-DNA hybrid quadruplex. Such a hybrid quadruplex structure formed between the template DNA and nascent mRNA strand is shown to play a role in transcription termination.[37] Recently, there have been a number of convincing studies on the role of quadruplex structures in misregulating the RNA biology of repeat expansion-associated neurodegenerative disorders and causing neuroglial pathogenies through nonconventional repeat-associated non-AUG (RAN) translation.[42−45] G-rich sequences that are prone to form quadruplex structures are also found in the regulatory regions of microbes, such as bacteria, fungi, and viruses. For example, a number of putative quadruplex-forming

Received: May 14, 2024
Revised: July 28, 2024
Accepted: August 21, 2024
Published: September 5, 2024

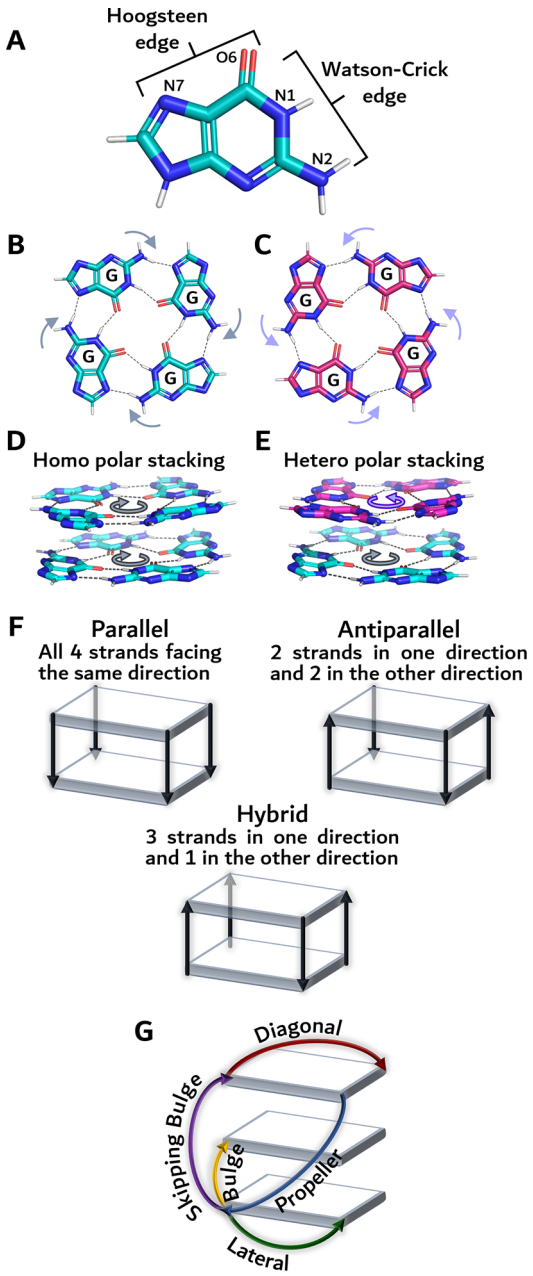


Figure 1. Nomenclature used in this study to describe the quadruplex folds. (A) Guanine base marked with Watson-Crick (donor) and Hoogsteen (acceptor) edges. (B) G-quartet with the donor to acceptor edge in a clockwise direction (indicated by arrows). The dotted line indicates the hydrogen bonding pattern. (C) G-quartet with the donor to acceptor edge in an anticlockwise direction (indicated by arrows). (D) Homo polar stacking illustrated by considering 2 G-quartets in a clockwise direction as an example. (E) Hetero polar stacking illustrated by considering one G-quartet in an anticlockwise direction (top, colored cyan, indicated by an arrow) and another in a clockwise direction (bottom, colored pink, indicated by an arrow). (F) Schematic illustration of parallel, antiparallel, and hybrid quadruplex folds. (G) Loops observed in quadruplex folds.

sequences are found in *Plasmodium falciparum*.[10] Quadruplex structures have also been shown to have a role in viral replication, translation modulation, and immune evasion.[11] Additionally, they play a role in modulating microbial transcription through radioresistance, antigenic variation, recombination, and latency.[11]

Due to its participation in the aforementioned diverse biological functions, quadruplex structures act as disease biomarkers[46] and potential therapeutic targets: anticancer,[34,47] antimicrobial,[11,48] anticoagulant,[49] and antineurodegenerative.[47,50] Furthermore, G-quadruplex-forming aptamers play a role in drug delivery[51] and diagnoses.[52] There are several proven examples of the role of quadruplexes in nanotechnological applications, such as biosensors,[53,54] nanomotors,[55] nanomaterials,[56] origami scaffolds (to capture direct interaction between quadruplexes and proteins),[57] and nanowires in nanoelectronics.[58]

Since quadruplex structures are intrinsically polymorphic in nature, they are greatly influenced by the sequence and environmental conditions such as pH, cations, *etc*. This makes it a universal challenge to find the thermodynamically stable quadruplex fold from its sequence information.[59,60] Several research groups around the world are working on getting the atomistic details of these structures using X-ray crystallography, NMR, and cryo-EM.[61−63] However, only a few studies have summarized in detail the atomistic structures of quadruplexes.[64−67] Recently, this lab has reported ten unique folding topologies of RNA quadruplexes[27] after analyzing in detail the structures deposited in the protein databank (PDB).[62] Considering the diverse therapeutic potential and nanotechnology applications of DNA quadruplexes, a systematic analysis of DNA quadruplex structures deposited in the PDB has been carried out here. The information provided here may enlighten the topological diversity of DNA quadruplex folds in understanding not only their biological significance and therapeutics but also the custom design of DNA quadruplex scaffolds for nanotechnological applications.

MATERIALS AND METHODS

DNA Quadruplex Structures. In order to understand the architecture of various DNA quadruplex folds, the Cartesian coordinates of quadruplex structures were downloaded from PDB,[62] NAKB (previously known as NDB),[68] and ONQUADRO[61] using the keyword search "quadruplex", "tetraplex", "tetrad", "quartet", or "G4-helices". After removal of the redundant structures, 437 DNA quadruplexes (433 PDBs as of July 15, 2024) were finally considered for analysis. This includes 11 left-handed structures. Note that the intercalating quadruplex structures (PDB IDs: 1V3P, 2DZ7) (viz., octaplex) are not considered for the analysis as it is beyond the scope of this article.

Deriving Two-Layered Structural Motifs from the DNA Quadruplex Structures. Since the objective of this study is to show that the quadruplex structures or folds are modular in nature in such a way that they can be represented in terms of small repetitive 3D structural units, a thorough analysis of 437 quadruplex PDB structures was carried out in this context. The 3D structural units, henceforth termed quadruplex motifs, were derived from the quadruplex folds by breaking the folds into two-layered structural blocks comprising two quartets.

Conformational Angle Preference. The sugar−phosphate backbone and glycosyl conformational angles [α (O3′-P-O5′-C5′), β (P-O5′-C5′-C4′), γ (O5′-C5′-C4′-C3′), δ (C5′-C4′-C3′-O3′), ε (C4′-C3′-O3′-P), ζ (C3′-O3′-P-O5′), and χ (O4′-C1′-N9/N1-C4′/C2′)] of each quartet-forming residues in the quadruplex folds (PDBs) were taken from NAKB. The conformational angle preference is reported based on the frequency of occurrence, which is individually normalized with respect to each fold, viz., separately for antiparallel, parallel, and

https://doi.org/10.1021/acsomega.4c04579
ACS Omega 2024, 9, 38696−38709

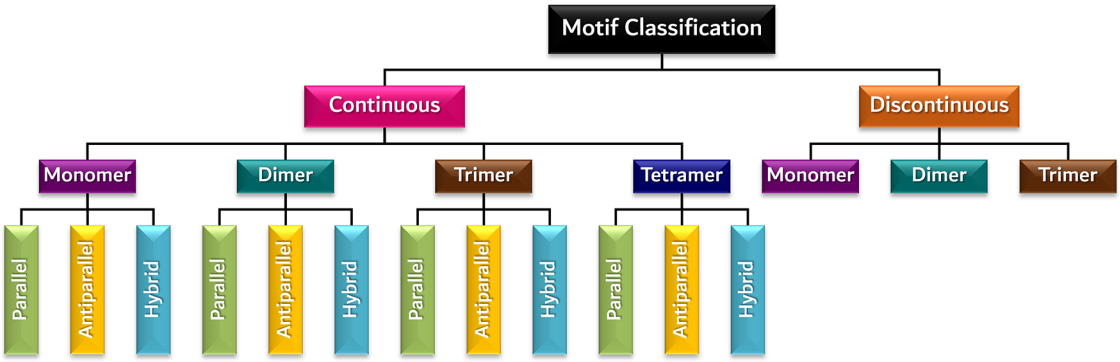


Figure 2. Quadruplex motif classification.

hybrid, which was then plotted as a bar plot using Microsoft Excel.

Plots. The contour density plots that show the conformational angle preferences of the quartet and loop residues were generated with the help of Gnuplot.[69]

RESULTS AND DISCUSSIONS

General Description of DNA Quadruplex Architecture. The nomenclatures used in the description of the quadruplex folds are shown in Figure 1. A guanine quartet (Figure 1B,C) is formed by the interaction of 4 guanines through hydrogen bonds involving the donors of the Watson-Crick edge and acceptors of the Hoogsteen edge (Figure 1A) either in a clockwise (Figure 1B) or in a anticlockwise (Figure 1C) direction. These quartets stack either in a homopolar [quartets of the same polarity (either clockwise (Figure 1D) or anticlockwise)] or in a heteropolar [quartets of different polarity (both clockwise and anticlockwise) (Figure 1E)] fashion. Figure 1F represents the three well-known broad classifications of quadruplexes based on the relative orientations of their four strands: parallel, antiparallel, and hybrid folds. In a parallel quadruplex fold, the directions of all four strands are the same. In the antiparallel fold, two strands are in one direction, and the other two strands are in the opposite direction. In the case of a hybrid fold, three strands are in one direction, and the other strand is in the opposite direction. However, this nomenclature is insufficient to understand the quadruplex folds, as the internal loops can further complicate the quadruplex folds. There are four loops observed in the quadruplex folds: diagonal, lateral, propeller, and bulge (Figure 1G).[70] The diagonal loop connects the guanines of the alternate strands in the terminal quartet, and the lateral loop connects the adjacent strand guanines of the terminal (same) quartet. The former leads to a change in the direction of the alternate strands, and the latter results in a change in the direction of the adjacent strands. Unlike the lateral and diagonal loops, the propeller and bulge loops connect two different quartets. The propeller loop connects the two adjacent strands in a parallel fashion. In most situations, the propeller loop connects two terminal quartets that are located at opposite ends of a quadruplex and turns around the strand polarity. Thus, the propeller loop is in an antiparallel direction with respect to the strands to which it connects. A propeller loop can also connect adjacent strands involving a terminal and a nonterminal quartet. Yet another special type of propeller loop is a V-shaped loop.[71] A V-shaped loop is observed when there is a discontinuity in the quadruplex strand. This facilitates the 3′ end of the disconnected quartet residue of one strand to directly connect to the 5′ end of the adjacent strand (with or without any intervening loop residues), resulting in a strand polarity reversal. The notable point is that in the case of a V-shaped loop, the two adjacent strand residues that form the quartet are in an antiparallel orientation. Unlike these three loops, the bulge loop connects the same strand, wherein it can be presented between 2 adjacent (bulge, Figure 1G) or nonadjacent (skipping bulge, Figure 1G) quartets without and with changing the direction of the strand, respectively.

Classification of Quadruplex Motifs. About 244 unique sequences from 437 DNA quadruplex structures (433 PDBs) reveal that there are 54 unique folds. Intriguingly, an in-depth analysis indicates that these quadruplex folds are modular in nature, as they can be represented in terms of small repetitive 3D structural units, henceforth referred to as quadruplex motifs. These motifs can be classified into two broad categories (Figure 2): continuous and discontinuous. In the former, all four residues of each adjacent quartet are connected continuously by the phosphodiester bonds of the backbone with or without bulge loops. In the latter, at least one of the four residues of each adjacent quartet is not connected by the phosphodiester bond. These two categories can further be classified into four classes (Figures 3 and 4): monomer (made with a single sequence) (Figure 3A and 4A), dimer (made up of 2 different sequences) (Figure 3B and 4B), trimer (made up of 3 different sequences) (Figure 3C and 4C), and tetramer (made up of 4 different sequences) (Figure 3D). The tetramer motifs are found only in the continuous category. These classes of the continuous category can further be classified into parallel, antiparallel, and hybrid motif types based on the strand orientation (Figure 1F). Although antiparallel quadruplex can further be subclassified in terms of groove widths,[72] this article broadly classifies the quadruplexes into antiparallel based on strand polarity to describe the overall quadruplex folding pattern (which is described by loop connectivity). There are 14, 9, 6, and 4 monomeric (Figure 3A), dimeric (Figure 3B), trimeric (Figure 3C), and tetrameric (Figure 3D) motifs, respectively, seen under the continuous motif category (Figure 3). Similarly, there are 6, 3, and 6 monomeric (Figure 4A), dimeric (Figure 4B), and trimeric (Figure 4C) motifs, respectively, seen under the discontinuous motif category (Figure 4). These motifs are two-layered quartets. Although the quartets of the two-layered quadruplex motifs can either be homo- (Figure 1D) or hetero- (Figure 1E) polar stacked, the diversity among the motifs defined here arises out of the strand orientation and the type of loops that connect the quartets. Henceforth, for the sake of description purpose, each motif is given a five-letter alphanumeric name as described below.

The first letter starts with the alphabet "Q", representing the quadruplex (Q) followed by the category (C (continuous) or D

https://doi.org/10.1021/acsomega.4c04579
ACS Omega 2024, 9, 38696−38709

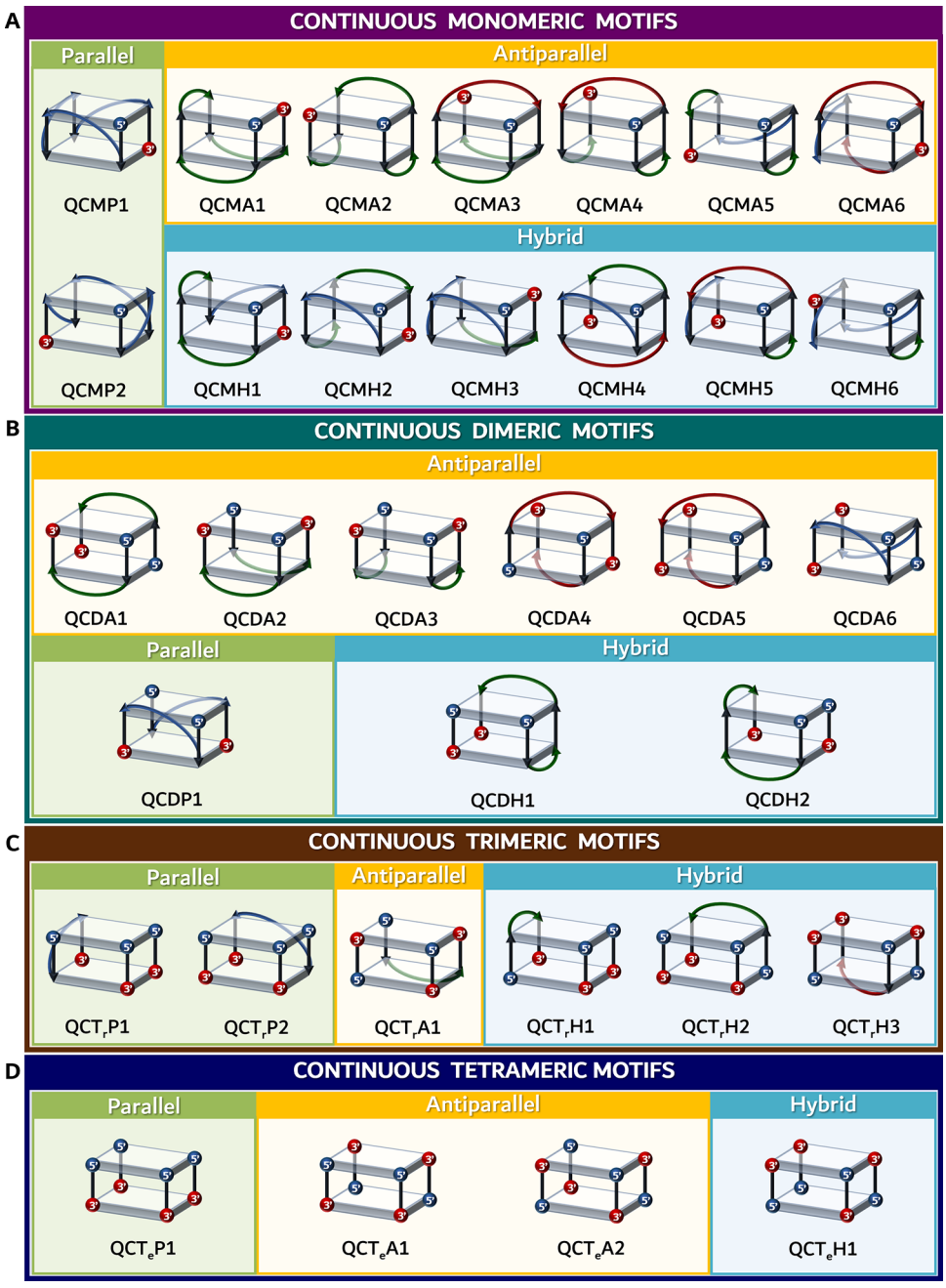


Figure 3. Schematic representation of the continuous quadruplex motifs. Note that the bulge loop is ignored in the representation as it would not affect the strand direction. Thus, for representation purposes, the bulge loop residues are considered to be zero. Note that to maintain uniformity, the 5′ end (blue-colored sphere) is always kept at the top right and front face of the quadruplex motif. In the multimeric dimer, trimer, and tetramer, the 5′ end of any one of the sequences is kept in the top right and front faces of the motif. The red-colored sphere represents the 3′ end.

(discontinuous)) followed by the class (M (monomer) or D (dimer) or T_r (trimer) or T_e (tetramer)) followed by motif type (P (parallel) or A (antiparallel) or H (hybrid)), which is finally followed by a number (assigned serially based on first come first serve basis) representing different motifs. Note that in the case of discontinuous motifs, the fourth alphabet is indicated as "X" since it is not feasible to identify the strand orientation due to the discontinuous nature of the motif. For instance, a continuous monomeric motif with parallel strand orientation is indicated as QCMP followed by a number (Figure 3A). However, a discontinuous monomeric motif is represented as QDMX followed by a number (Figure 4A).

Continuous Motifs. Among the 33 continuous quadruplex motifs, the antiparallel motifs (15) is predominantly seen, followed by hybrid (12) and parallel (6) motifs. There are two parallel (QCMP), six antiparallel (QCMA), and six hybrid (QCMH) motifs seen under the monomeric continuous (QCM) category. One can visualize from Figure 3 that the type and direction of the connecting loops (lateral, propeller, and diagonal) bring diversity to these motifs. For instance, although QCMP1 and QCMP2 motifs are connected through a propeller, they are different due to the direction of the propeller loop. Note that among the various hybrid motifs observed in the quadruplex PDB structures, the QCMH2 and QCMH1 motifs

https://doi.org/10.1021/acsomega.4c04579
ACS Omega 2024, 9, 38696−38709

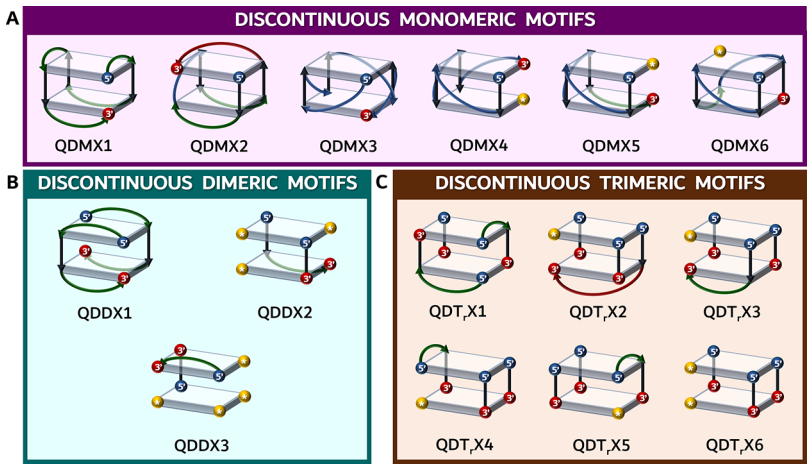


Figure 4. Schematic representation of the discontinuous quadruplex motifs. Note that the star enclosed by a golden-colored sphere represents the nucleotide monomer whose orientation can be of any direction (viz., 5′ or 3′). Such single nucleotides are not considered independent strands and, thus, are not used to define the quadruplex motif multimer.

correspond to the well-studied telomeric hybrid 1 and hybrid 2 structures, respectively.[73]

While the QCMP1 motif is connected by clockwise propeller loops, QCMP2 is connected by counterclockwise propeller loops; thus, they become mirror images of each other in terms of the loop (Supplementary Figure S1A). Note that the G-core maintains the same helicity in both cases. Similarly, QCMA1 and QCMA2 (Figure S1B), as well as QCMA3 and QCMA4 (Figure S1C), are mirror images. QCMA1 and QCMA3 are examples of how the type of loop leads to different motifs. Although both of them have two lateral loops, the presence of a third lateral loop in QCMA1 instead of a diagonal loop in QCMA3 leads to a different monomeric antiparallel quadruplex motif. In a similar fashion, dimeric, trimeric, and tetrameric motifs can be defined as shown in Figure 3B−D. Among the described multimeric motifs, QCDA2 and QCDA3, QCDH1 and QCDH2, QCT$_r$P1 and QCT$_r$P2, and QCT$_r$H1 and QCT$_r$H2 are the mirror image motif pairs. One can envisage that the remaining motifs can also have their mirror image based on the sequence and environment. One more well-known fact is that the parallel motifs are always connected by propeller loops, and the tetrameric motifs do not have any loops. Notably, the same quadruplex motifs (representing the type of connectivity and strand polarity) can also be found in quadruplex-duplex hybrids, as the length of the loop affects only the loop's secondary structural conformation and not the quadruplex motif pattern, viz., a duplex is generally formed when a greater number of nucleotides are in the loop. For instance, QCMA1 is the underlying motif behind the PDB IDs 6JKN and 8R6G, although the latter has a duplex as a part of a longer lateral loop.

Discontinuous Motifs. There are 15 discontinuous quadruplex motifs known as of now in which one or more adjacent quartet residues are not connected. Based on the number of strands, they can be classified into monomers, dimers, and trimers. There are 6, 3, and 6 monomeric, dimeric, and trimeric discontinuous motifs observed so far (Figure 4). Intriguingly, discontinuous quadruplex motifs can be derived from continuous quadruplex motifs. One such example is QCMP1 and QDMX4, wherein the removal of connectivity between the last two residues of QCMP1 leads to QDMX4. Thus, theoretically, each continuous motif can lead to one or more discontinuous motifs.

The different quadruplex motifs presented here may have different influences on interaction with the ligand molecules based on the presence or absence of different loops as well as their types. Furthermore, it has been discussed in a recent study that eight distinct G-quartets are possible based on the combination of *anti/syn* glycosyl conformations, which impact the quadruplex groove widths,[70,72] thereby affecting the ligand binding.[72] For instance, the glycosyl conformations of two adjacent bases of the same quartet can determine the groove width between the corresponding two strands. A *syn-anti* conformation (when quartet polarity is clockwise) results in a narrow groove, *syn-syn* or *anti-anti* leads to a medium groove, and *anti-syn* results in a wide groove, which may influence the ligand binding differently.[72] Since the complexity further increases when G-quartets are stacked on top of each other along with the inclusion of different types of loops and results in different motifs, it may further influence the ligand or protein binding with the quadruplex. Ligands generally bind with the quadruplexes through end-stacking, groove recognition, loop recognition, or intercalation.[74] Thus, the type of quadruplex topology affects the binding mode of ligands with the quadruplex.

An example of this is the interaction of the TMPyP4 ligand with the quadruplex. The planar TMPyP4 ligand binds to the parallel fold with a higher affinity than the antiparallel fold, as the diagonal loop in the latter hinders the ligand binding with the quadruplex through end-stacking mode.[75] Intriguingly, evidence shows that some ligands can switch the topology of the quadruplex to favor its interaction. One such example is the L2G2-2M2EG-6OTD ligand, which favors end-stacking and changes the antiparallel or hybrid to parallel fold to facilitate the binding with a telomeric quadruplex sequence.[76,77] Similar to small molecules, proteins interact with quadruplexes through end-stacking, groove-binding, and loop binding. For instance, the POT1 protein, a part of the telomere shelterin complex, shows selectivity for antiparallel over parallel folds, which suggests the probable groove-binding mode.[78,79] These examples indicate the selectivity of ligands and proteins toward a quadruplex fold.

Notably, the experimental conditions and structure calculation methods may influence the folding architecture of quadruplexes. Thus, the folding patterns discussed here are solely based on the quadruplex structures deposited in the PDB,

https://doi.org/10.1021/acsomega.4c04579
ACS Omega 2024, 9, 38696−38709

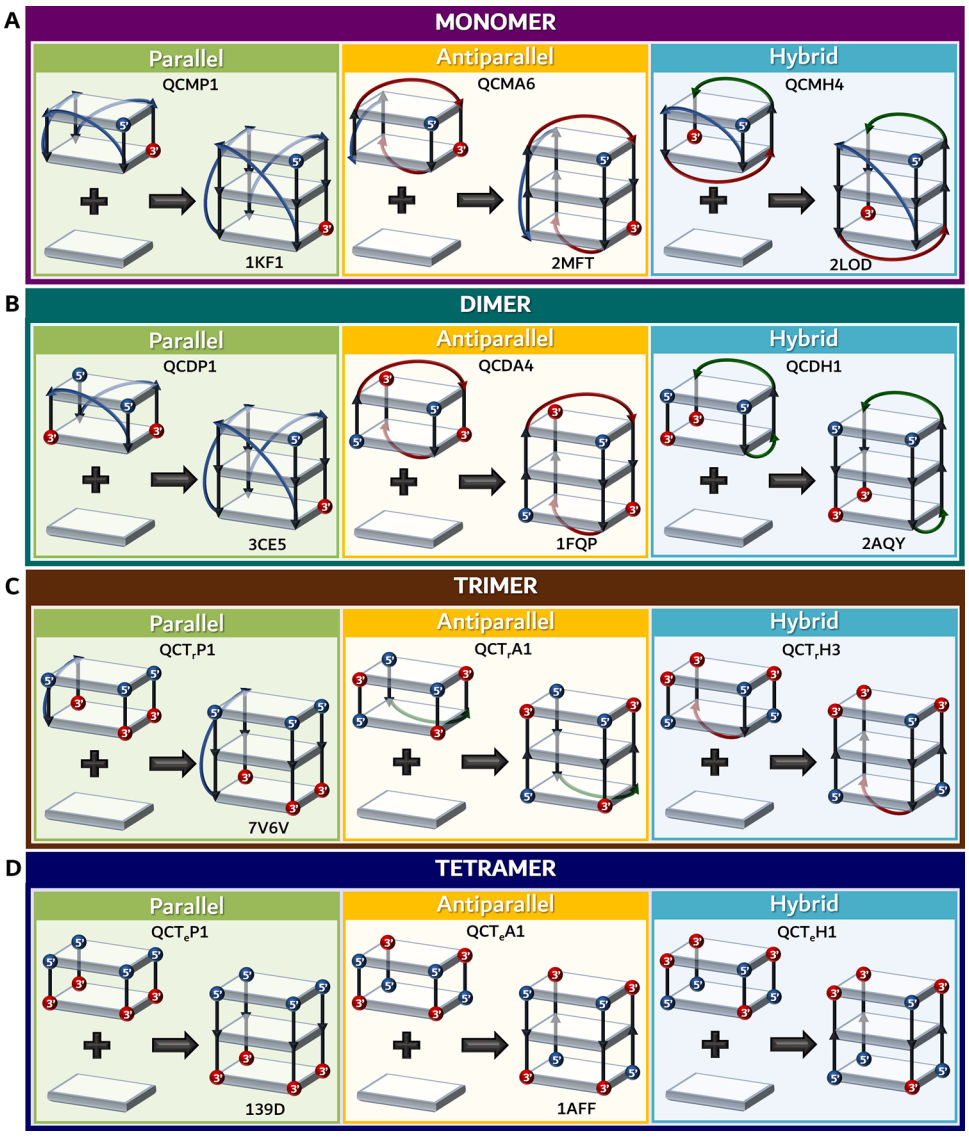


Figure 5. Examples illustrating the modular representation of (A) monomeric, (B) dimeric, (C) trimeric, and (D) tetrameric (left) parallel, (middle) antiparallel, and (right) hybrid quadruplex folds using the continuous quadruplex motifs described in Figure 3. Note that the name of the continuous monomeric motif is given alongside each panel. A representative PDB ID is given adjacent to each fold, wherever applicable.

which were solved predominantly by X-ray crystallography followed by NMR (solution-state), solid-state NMR, and cryo-electron microscopy techniques (Supplementary Table S1).

Modular Representation of DNA Quadruplex Folds. *Example Illustrating the Representation of Simple Quadruplex Fold from the Motif.* Figure 5 shows a detailed representation of how the continuous parallel, antiparallel, and hybrid quadruplex motifs (Figure 3) can be used as scaffolds to represent simple quadruplex folds. For instance, QCMP1 is used for the modular representation of a monomeric parallel quadruplex fold (Figure 5A (Left)) by simply adding a G-quartet without affecting the overall quadruplex motif structure. Since QCMP1 has only two quartets, the third quartet can be added on either the top, bottom, or middle to form a fold. In the same way, more quartets can be added to increase the length of the quadruplex fold. However, when a non-G-quartet is present, it has to be placed in the appropriate place, unlike in the case of all G-quartets. In a similar fashion, the examples given in Figure 5A (middle) and A (right) illustrate the modular representation of monomeric antiparallel and monomeric hybrid folds using QCMA6 and QCMH4, respectively. One can also represent the dimeric, trimeric, and tetrameric continuous quadruplex folds using the appropriate motifs. The examples given in Figure 5B (Left), Figure 5B (Middle), and Figure 5B (Right) represent the dimeric fold by using parallel QCDP1, antiparallel QCDA4, and hybrid QCDH1 motifs, respectively. Figures 5C,D are examples of trimeric and tetrameric folds, respectively. A total of 12 different examples of deriving a simple quadruplex fold from the motifs are illustrated in Figure 5.

Example Illustrating the Representation of Complex Quadruplex Fold from the Motif. Unlike the cases described above, the representation of a complex quadruplex fold is not straightforward. Such a situation arises when at least one of the motifs is discontinuous. In order to represent a discontinuous quadruplex fold, the quartets common to two motifs that form the fold should be superimposed to obtain the complete fold. Unlike in the continuous fold, wherein the quartet can be added in the middle, top, or bottom, here, two quartets should be overlaid to maintain the directionality of the fold. Furthermore, based on the nature of the quadruplex fold, one or more

https://doi.org/10.1021/acsomega.4c04579
ACS Omega 2024, 9, 38696−38709

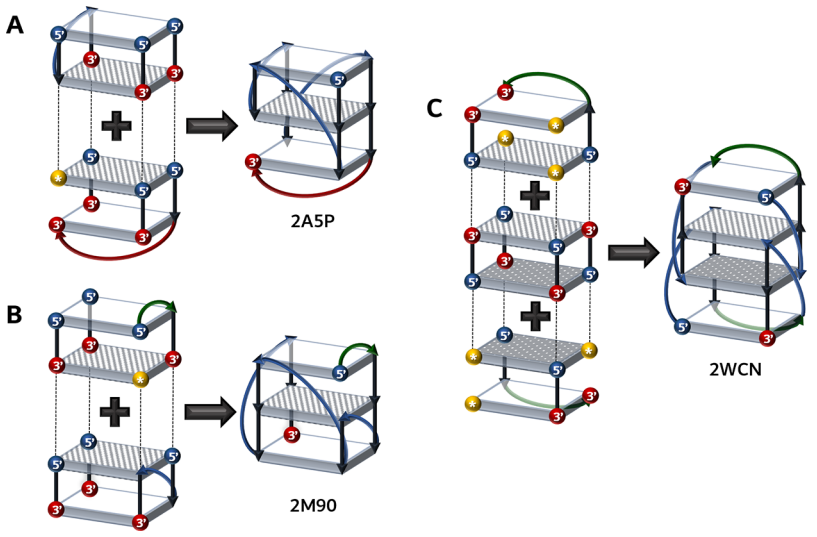


Figure 6. Examples illustrating the modular representation of monomeric (A and B) and dimeric discontinuous (C) quadruplex folds using the quadruplex motifs described in Figures 3 and 4. (A) and (B) Representing a quadruplex fold with the use of one discontinuous and one continuous motif. Note the formation of two different folds by just changing one of the motifs; viz., while the continuous motif is the same between the 2 cases, the discontinuous motif is different. (C) An example of forming a higher-order fold is shown by considering three motifs (one continuous motif amid two identical discontinuous motifs). Note the formation of four new propeller loops (colored blue) to complete the fold. (A−C) Dotted lines indicate the point of linkage between 2 different quadruplex motifs. The overlapping G-quadrates are indicated by striped (A−C) and checked (C) rectangular slabs. A representative PDB ID is given adjacent to each fold.

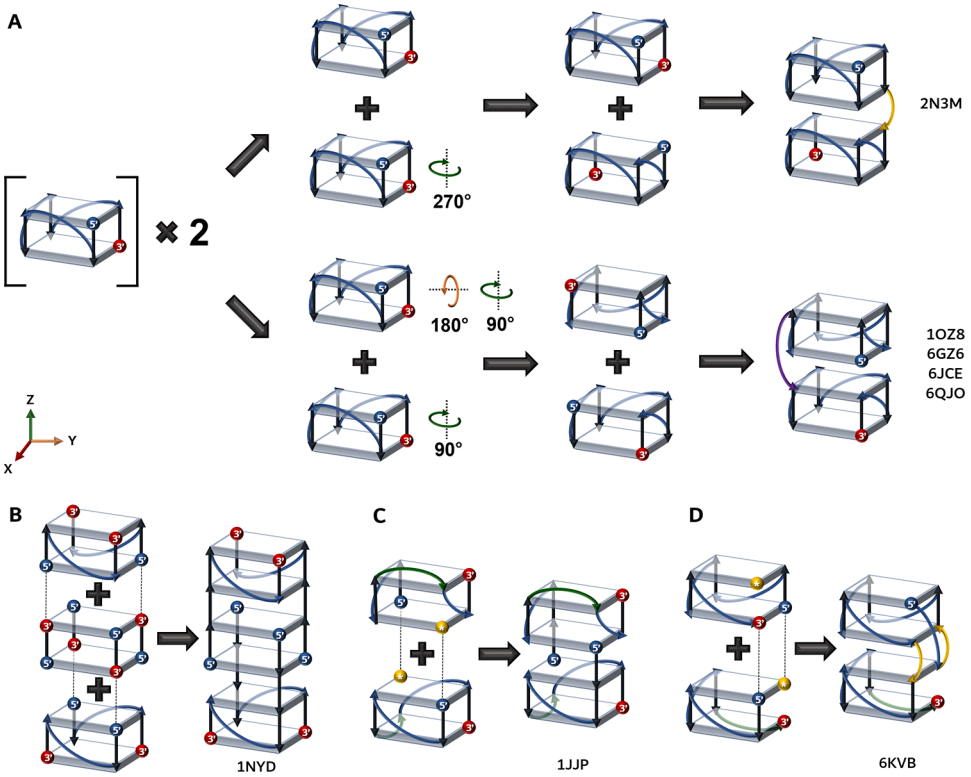


Figure 7. Schematic illustration of representing a higher-order (A and C) homo and (B and D) hetero quadruplex folds. (A) Two units of QCMP1 forming two different higher-order stacked quadruplex folds through a linker loop are shown. Note the possibility of having two different higher-order folds out of the same scaffold motif by just changing their relative orientations governed by the rotation around X, Y, and Z axes (indicated in circular arrows; X-, Y-, and Z-axes arrows are colored red, yellow, and green, respectively). Formation of a (B) hetero higher-order quadruplex fold using two different continuous motifs (QCDP1 and QCT$_e$A2), (C) homo higher-order quadruplex using two same discontinuous motifs (QDMX6), and (D) hetero higher-order quadruplex using two discontinuous motifs (QDMX4 and QDMX5). A representative PDB ID is given adjacent to each fold.

additional loops are required to complete the fold, which is dictated by the sequence. For instance, in Figure 6A, the quadruplex fold can be established by a continuous motif QCT$_r$P1 and discontinuous motif QDT$_r$X2, wherein these two motifs are superimposed with respect to the common quartet (indicated in gray colored stripes in Figure 6A,B). Note the two

https://doi.org/10.1021/acsomega.4c04579
ACS Omega 2024, 9, 38696−38709

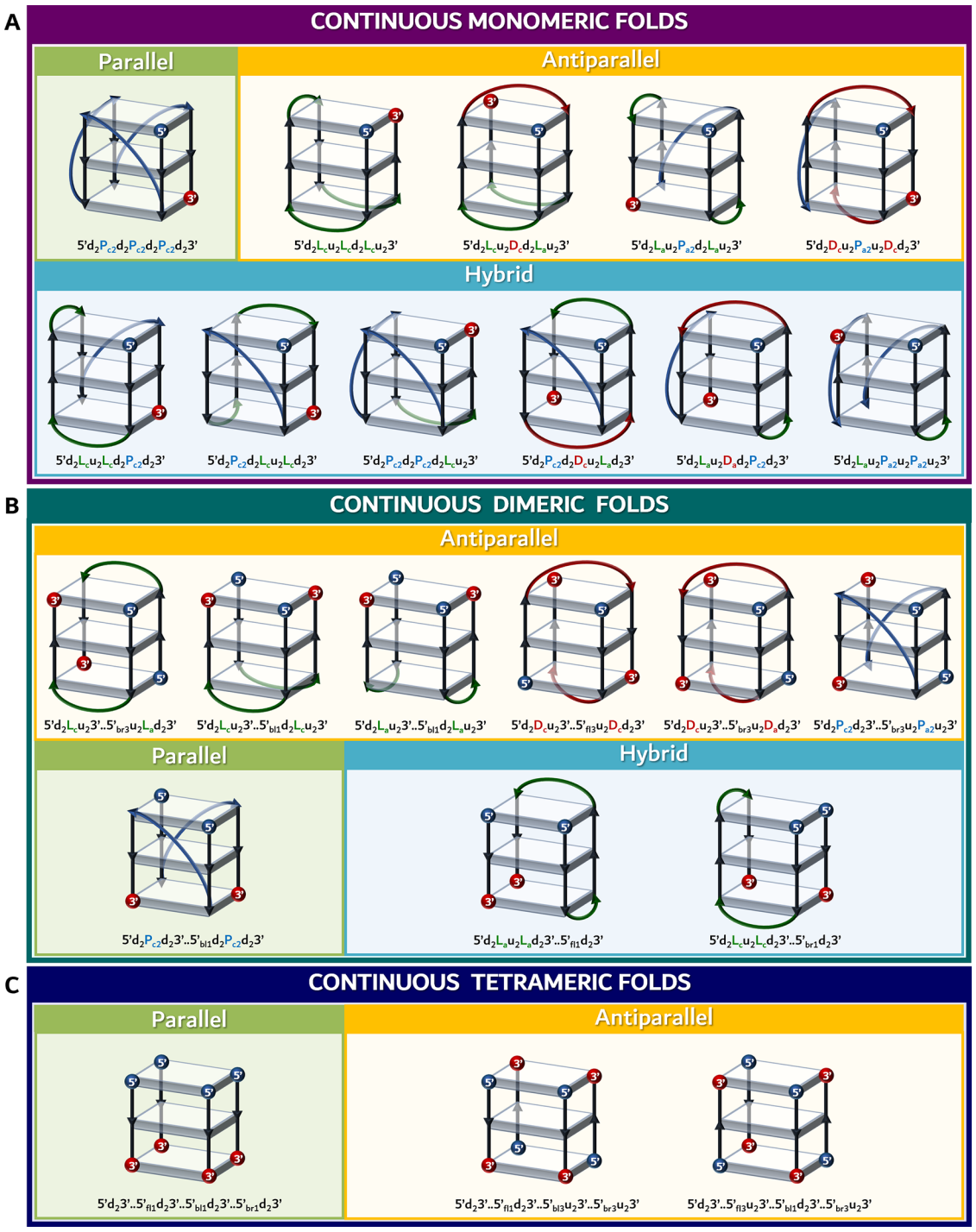


Figure 8. Schematic representation of unique continuous DNA quadruplex folds. For illustration purposes, each fold is represented by only three quartets; refer to Supplementary Table S1 for details about the PDB IDs, sequence, number of quartets, *etc.*, corresponding to each fold described. Note that the corresponding alphanumeric nomenclature is indicated adjacent to each fold.

new propeller loops in the final fold (colored blue) that make the fold monomeric. Interestingly, the addition of QCT_rP1 and QDT_rX2 motifs without the formation of two propeller loops makes a trimeric fold. It is noteworthy that placing the QDT_rX2 motif on the top of QCT_rP1, which is the reverse of the above, would result in a different fold (no such quadruplex fold structures are observed in the PDB as of now) depending on the sequence and environmental condition. Yet another example is shown in Figure 6B. An example of creating a quadruplex fold with more than two quadruplex motifs is shown in Figure 6C.

Example Illustrating the Representation of Higher-Order Quadruplex Structures Using the Motifs. A number of studies have shown that higher-order quadruplex folds exist in nature,[34,42,80−84] and there are 16 such quadruplex topologies seen in PDB (labeled as "Higher-order" in Supplementary Table S1). The analyses of the known quadruplex fold structures indicate two types of higher-order topologies: homo- and hetero. In the homo quadruplex folds, more than one repetition of a continuous or discontinuous motif occurs. However, the hetero quadruplex folds are formed by the combination of more

https://doi.org/10.1021/acsomega.4c04579
ACS Omega 2024, 9, 38696−38709

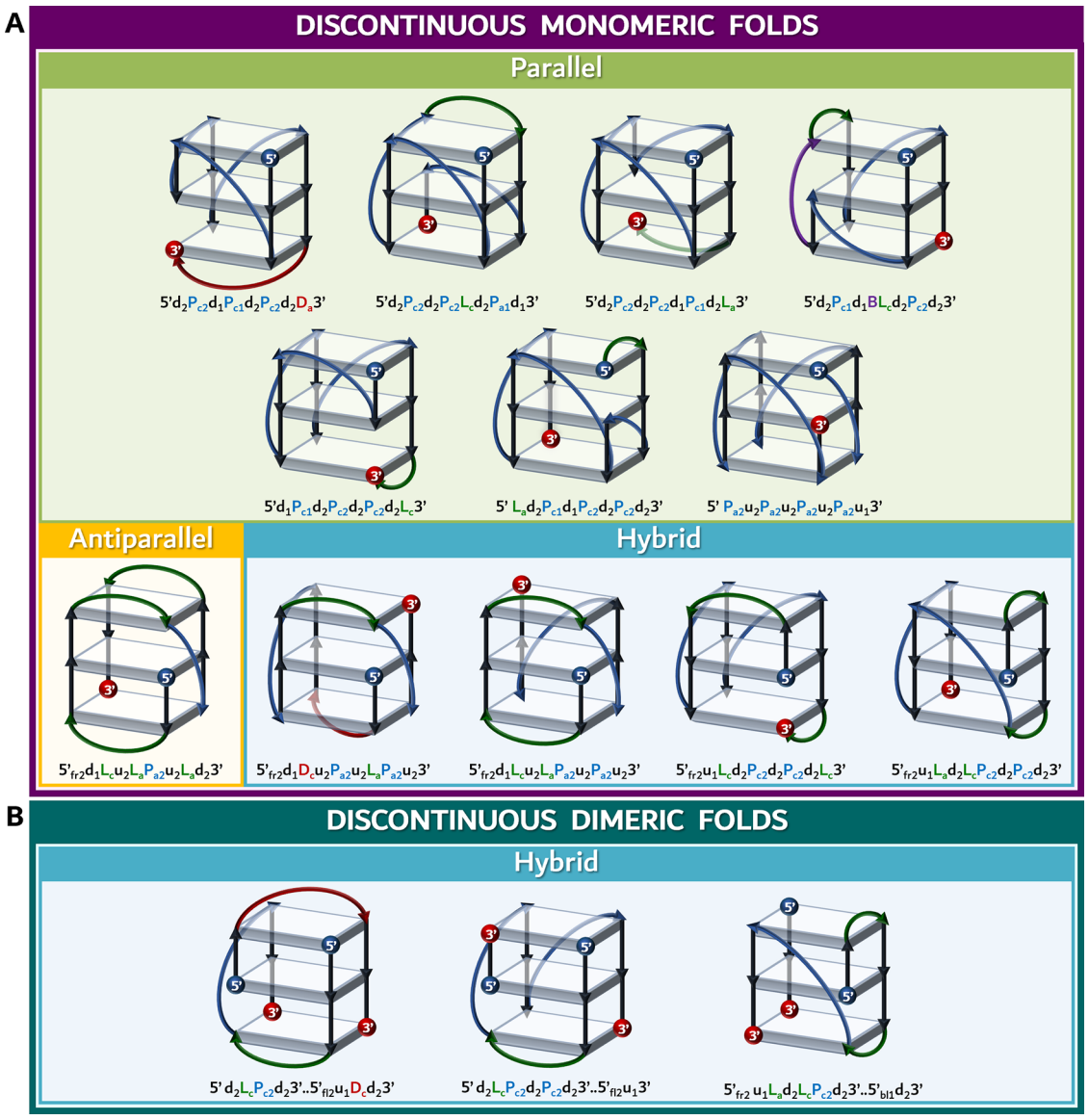


Figure 9. Schematic representation of unique discontinuous DNA quadruplex folds. Refer to Supplementary Table S1 for details about the PDB IDs, sequence, number of quartets, *etc.*, corresponding to each fold described. Note that the corresponding alphanumeric nomenclature is indicated adjacent to each fold, and the classification into parallel, antiparallel, or hybrid is based on the literature that reports the structure.

than one continuous and/or discontinuous motif. In the higher-order quadruplex, two or more of the same or different quadruplex motifs are connected through linker (l) loop(s), which can be considered a special case of a bulge loop that connects two motifs. One such example is given in Figure 7A, wherein two higher-order homo quadruplex folds are formed by two units of the same QCMP1 motif, which is dictated simply by the difference in the orientations between the two motifs in the Y- and Z-directions. While there are two thymines (T) in the linker loop for the fold given in Figure 7A (top), there is either a single 'T' or two 'T's or two adenines (A's) in the linker loop for the fold given in Figure 7A (bottom). This suggests that the relative orientation between two motifs (homo or hetero) may lead to many combinations of higher-order quadruplex folds depending on the rotation along the X- and/or - Y and/or Z axes. Among the higher-order structures analyzed, it was observed that a few structures have left- and right-handed motifs linked through a linker loop (PDB IDs: 6JCE, 6QJO, and 8TAA), and a few more structures have two left-handed motifs linked through a linker loop (PDB IDs: 2MS9, 4U5M, 6GZ6, 7D5D, 7D5E, 7D5F, and 7DFY). When the number of motifs increases, more complications are created in the higher-order arrangement of the quadruplex topology. In addition, the position of linker loop nucleotides in the sequence may lead to different arrangements among the motifs. An example of having more than two continuous motifs forming a hetero quadruplex topology is given in Figure 7B, wherein the end motifs of the quadruplex fold are continuous dimeric parallel, whereas the middle one is a tetramer antiparallel continuous motif. One can expect more complexities with many such combinations of parallel, antiparallel, and hybrid motifs. Homo higher-order quadruplex folds can be formed with two discontinuous motifs (Figure 7C). A hetero quadruplex fold can also be formed with discontinuous and continuous motifs, as given in Figure 7D.

Devising a Universal Nomenclature to Define the Folding Architecture of Quadruplexes. A total of 54 unique DNA quadruplex folds (Figures 8−10) are identified from the analysis of 433 PDBs. Among the 54-fold folds, 23, 15, and 16

https://doi.org/10.1021/acsomega.4c04579
ACS Omega 2024, 9, 38696−38709

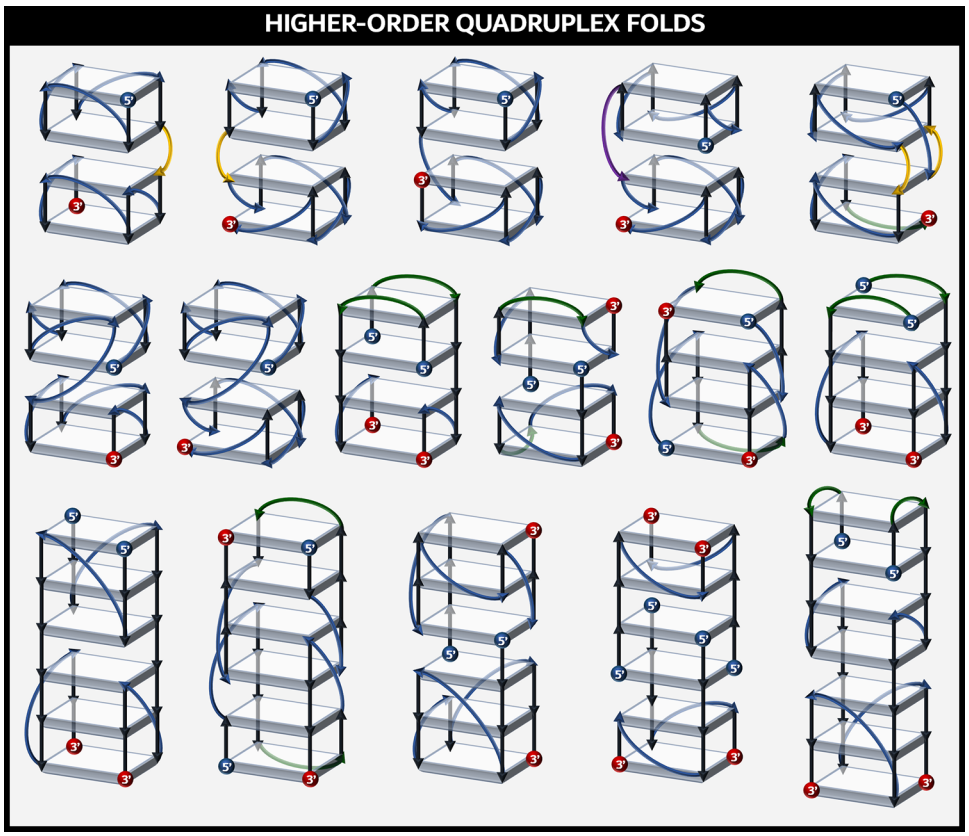


Figure 10. Schematic representation of unique higher-order DNA quadruplex folds. Refer to Supplementary Table S1 for the details about the PDB IDs, sequence, number of quartets, *etc.*, corresponding to each fold described.

are continuous, discontinuous, and higher-order folds. Although these folds are predominantly formed by G-quartets (Figure 1A), non-G-quartets like T-T-T-T (PDB ID: 7D31), C-C-C-C (PDB ID: 6A85), A-T-A-T (PDB ID: 5LS8), G-C-G-C (PDB ID: 7CV4), C-A-G-A (PDB ID: 6ZX7), and G-G-A-T (PDB ID: 5VHE) are also found. Out of the 244 sequences considered for the analysis, the number of sequences having continuous parallel folds (3-types as shown in Figure 8: $5'd_2P_{c2}d_2P_{c2}d_2P_{c2}d_23'$, $5'd_2P_{c2}d_23'..5'_{bl1}d_2P_{c2}d_23'$, and $5'd_23'..5'_{fl1}d_23'..5'_{bl1}d_23'..5'_{br1}d_23'$) is more compared to the antiparallel and hybrid folds; viz., 87 unique sequences among 152 structures prefer parallel folds. Thus, it is clear that different DNA sequences assume an identical fold. Interestingly, an identical sequence taking different folds is also observed, indicating the role of environmental conditions such as type of salt and salt concentration in dictating the quadruplex fold. For instance, "GGGTTAGGGTTAGGGTTAGGG" assumes both parallel fold (PDB ID: 1KF1) in the presence of 50 mM K+ in a crystal packing environment (X-ray crystallography)[85] and an anti-parallel fold (PDB ID: 143D) in the presence of 90 mM Na+ in a solution environment (solution NMR),[85] illustrating the impact of the environment on quadruplex folding. When this quadruplex-forming sequence has overhangs, it further assumes three more folds (PDB IDs: 2GKU, 2JPZ, and 2MBJ). A few more examples for the same sequence assuming different folds are PDB IDs: 2MCC, 2MCO, 3R6R, 3SC8, 3T5E, 3UYH, 4FXM, 4G0F, 6IP3, 6XCL, and 7KLP.

Previously, a naming scheme for canonical or monomeric quadruplexes has been proposed.[70] This nomenclature is helpful but limited to monomeric continuous monomeric quadruplexes. The complexity of the quadruplex structures (intermolecular, intramolecular, and discontinuous) necessitates a more elaborated nomenclature. To this end, an alphanumeric nomenclature has been devised here to represent the quadruplex folds (inter-, intramolecular, and discontinuous) and provide insight into the overall folding architecture of the DNA quadruplex backbone. Notably, incorporating the sequence information would lead to multiple names representing a single fold. Thus, the bases involved in the quartets are not included in the nomenclature. For instance, PDB IDs 2AVJ and 1A8N both are represented by QCDA1 motif and $5'd_3L_cu_33'..\ 5'_{br4}u_3L_ad_33'$ fold, although the former is formed by four G-quartets and the latter by two G-quartets and two G-C-G-C quartets.

The nomenclature starts with the 5′ end of the quadruplex strand. Next, the suffix '$_f$' or '$_b$' is added to the 5′ to indicate whether the 5′ end starts from the front or back side of the quadruplex fold, respectively. This is then followed by a suffix '$_l$' or '$_r$' to indicate the location of the 5′ end in the left or right side of the quadruplex, respectively. Finally, a numerical value is added in the suffix to represent the position of the quartet (with the numbering scheme going in ascending order from the top of the quadruplex fold). For instance, $5'_{bl3}$ refers to the 5′ end of the quadruplex strand that starts from the third quartet of the back face, located on the left-hand side of the quadruplex. By default, 5′ alone indicates the first ($_1$) quartet present on the right-hand side ($_r$) front face ($_f$) of the quadruplex fold. Further, the chain direction is indicated by a diagonal ($D_{c/a}$), lateral ($L_{c/a}$), bulge (B), or propeller loop ($P_{(c/a)n}$), or the number of quartet residues the chain has to travel down (d_n) or the number of quartet residues the chain has to travel up (u_n). While the suffix '$_{c/a}$' in D, L, and P refers to the clockwise or anticlockwise direction of the chain, '$_n$' in P, d, and u indicates the number of

https://doi.org/10.1021/acsomega.4c04579
ACS Omega 2024, 9, 38696−38709

quartet residues. Finally, the naming scheme ends with 3′ to denote the 3′ end of the quadruplex strand. This nomenclature can readily be used to refer to all of the continuous (Figure 8) and discontinuous (Figure 9) quadruplex folds. For instance, $5'd_2D_cu_2P_{a_2}u_2D_cd_23'$ refers to the monomeric continuous fold given in the right-most end of the first row of Figure 8A. In a similar fashion, each strand of the intermolecular quadruplex can be referred. Here, the different strands of the quadruplex fold are differentiated by '..'. By default, this nomenclature represents right-handed quadruplex folds, as they are predominantly seen. However, it can readily be migrated to the left-handed structures by adding 'L-' at the beginning of the left-handed structure since the alphanumerical nomenclature defines the folds based on the connectivity of the adjacent strands. For instance, the left-handed structure, 6FQ2, can be indicated as 'L-' in the beginning as follows $L\text{-}5'P_{c1}u_1P_{c1}u_1P_{c1}u_1P_{c1}3'$ (Figure S2A). Similarly, when both right- and left-handed motifs occur in the same fold, the motif corresponding to the left-handed motif can be distinguished by 'L-' (Figure S2B,C). Further, the same nomenclature can be extended for the higher-order quadruplex folds, wherein the intertwined quadruplex folds pose more challenges. Due to such complexity, the alphanumeric nomenclature for the higher-order quadruplexes is not mentioned here (Figure 10).

Among the 54 DNA folds discussed here, 6 of them are found to be common with RNA quadruplex folds.[27] However, some folds are specific for the RNA quadruplex, which are not discussed here.

Conformational Angle Preference for Different Quadruplex Folds. The conformational angle preference of the quadruplex folds is analyzed individually for G-quartet-forming residues and the loop residues. The conformational angles corresponding to each PDB ID (given in Supplementary Table S1) deposited in NAKB (previously known as NDB)[68] are downloaded and compiled in an Excel file. These conformational angles are then used for generating the density plots for the torsion angles with the help of Gnuplot.[69] The backbone conformational angles α (O3′-P-O5′-C5′), β (P-O5′-C5′-C4′), γ (O5′-C5′-C4′-C3′), δ (C5′-C4′-C3′-O3′), ε (C4′-C3′-O3′-P), ζ (C3′-O3′-P-O5′), and the glycosyl conformation angle χ (O4′-C1′-N9/N1-C4′/C2′) (Supplementary Figure S3A) of the quartet guanine residues exhibit a preference for *gauche-*, *trans*, *gauche+*, *trans*, *trans*, *gauche-*, and *anti* (Figure S3B−G), respectively, with a few exceptions. The *trans* conformation for 'δ' indicates the C2′-endo sugar puckering preference, and only a minor population of C3′-endo (*gauche+*) is seen. This is in contrast to RNA quadruplex, wherein both C2′-endo and C3-endo sugar puckers are found to be equally favorable.[27] In parallel quadruplex structures, the majority of the bases are observed to be in *anti* glycosyl conformation (>96%) (Figure S3H), facilitating the formation of a quartet with a clockwise polarity (Figure 1B) and naturally easing the hydrogen bonding formation between the quartet residues. However, a few parallel quadruplex structures (PDB IDs: 2CHJ, 2L88, 6ERL, 6JWD, 6JWE, and 8GP7) have an all-*syn* glycosyl conformation in one of the quartets, resulting in a quartet with an anticlockwise polarity (Figure 1C). In the antiparallel quadruplex structures, + *syn* and *anti* (2 *syn* and 2 *anti*) conformations are equally observed. Finally, in the hybrid quadruplex structures, a slightly higher population of *anti* is observed when compared to + *syn* conformation due to the fact that a pattern of 3 *syn* + 1 *anti* guanines is commonly observed in just one of the tetrads of hybrid quadruplex folds, whereas the remaining tetrads are of 3 *anti* + 1 *syn* (Figure S3H).[72]

The loop residue conformational angles α, β, γ, ζ, and χ predominantly prefer *gauche-*, *trans*, *gauche+*, *gauche-*, and *anti* conformations, respectively (Figure S4). Interestingly, ε takes a wide range of conformational angles in the range 180°−300° (Figure S4E). A significant population of *gauche+*, *trans*, and *gauche+*/*trans* is also seen for α, γ, and ζ, respectively, similar to the quartet guanine residues (Figure S3−S4). δ value further confirms the C2′-endo sugar pucker (Figure S4).

CONCLUSIONS

A detailed analysis of experimentally derived quadruplex structures is carried out here to understand the architecture of the DNA quadruplexes. The modular nature of the quadruplex architecture is established here by defining quadruplex motifs. Utilization of these motifs as scaffolds for simple as well as higher-order quadruplex folds is also demonstrated. Such knowledge about the quadruplex topologies will be helpful in the programmed design of quadruplex folds, motif-specific ligand design, and understanding of the interaction with the proteins. One can further explore the sequence specificity of the quadruplex motifs and utilize them in the genomewide prediction of quadruplex motifs with the help of machine learning approaches. Further, a universal nomenclature was devised here to represent the quadruplex folds. Such an alphanumeric representation of quadruplex folds may be helpful in training machine learning models to predict and model the quadruplexes. Since meaningful and informative data labeling is required to preprocess the raw data to enable the supervised machine learning model to learn and make accurate predictions, the labeling scheme (nomenclature) provided here would be useful as labels to specify each fold.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c04579.

Table S1. List of PDB IDs (column 1), sequence (column 2), fold nomenclature (column 3), fold classification (viz., continuous, discontinuous or higher-order; column 4), fold type (viz., parallel, antiparallel or hybrid; column 5), number of quartets (column 6), residue numbers forming G-/non-G-quartets (column 7), details of loop residues (column 8), ligand (viz., ion, water, protein, and small molecules; column 9), experimental method (viz., solution NMR, X-ray diffraction; column 10), and DOI (column 11) (XLSX)

Figure S1. Illustration of the mirror image relationship between (A) QCMP1 and QCMP2, (B) QCMA1 and QCMA2, and (C) QCMA3 and QCMA4. Figure S2. Examples illustrating the usage of alphanumerical characters to describe the left-handed (indicated by a dotted box) quadruplex fold. Quadruplex folds having (A) only a left-handed motif, (B) 5′ right-handed motif flanked by a 3′-left-handed motif, and (C) 5′ left-handed motif flanked by a 3′-right-handed motif. Note that the PDB IDs corresponding to each quadruplex fold are indicated on the top. Figure S3. Contour density plot illustrating the conformational angle preference of quartet residues: (A) The schematic representation of a nucleotide (with guanine as a representative base) marked

https://doi.org/10.1021/acsomega.4c04579
ACS Omega 2024, 9, 38696−38709

with α, β, γ, δ, ε, ζ, and χ conformational angles and (B−G) 2D conformational angle plot. The density scale bar is shown at the bottom. (H) Bar plot illustrating the parallel, antiparallel, and hybrid quadruplex-wise χ conformational angle preference for the quartet guanines. Figure S4. Contour density plot illustrating the conformational angle preference of loop residues. (A−F) 2D conformational angle plots of α, β, γ, δ, ε, ζ, and χ conformational angles. The density scale bar is shown at the bottom (PDF)

AUTHOR INFORMATION

Corresponding Author

Thenmalarchelvi Rathinavelan − *Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Telangana 502284, India;* orcid.org/0000-0002-1142-0583; Email: tr@bt.iith.ac.in

Authors

Sruthi Sundaresan − *Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Telangana 502284, India*

Patil Pranita Uttamrao − *Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Telangana 502284, India*

Purnima Kovuri − *Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Telangana 502284, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c04579

Author Contributions

S.S. collected the PDB data and carried out quadruplex motif and fold analyses. P.P.U. collected the PDB data, participated in fold analysis, and carried out conformational angle analyses. P.K. carried out fold and loop residue analyses during the early stage of the project. S.S. and T.R. wrote the manuscript. T.R. designed and supervised the project.

Funding

We greatly acknowledge the support from BIRAC-SRISTI GYTI (PMU2019/007), BIRAC-SRISTI GYTI (PMU_2017_010), and SERB (CRG/2022/001825).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Sathyaseelan for suggestions on PDB data collection. The authors thank the Indian Institute of Technology Hyderabad for their computation resources. S.S. and P.P.U. thank MoE and CSIR, respectively, for the fellowship.

REFERENCES

(1) Watson, J. D.; Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature.* **1953**, *171*, 737−738.

(2) Rypniewski, W.; Adamiak, D. A.; Milecki, J.; Adamiak, R. W. Noncanonical G(syn)-G(anti) base pairs stabilized by sulphate anions in two X-ray structures of the (GUGGUCUGAUGAGGCC) RNA duplex. *RNA* **2008**, *14*, 1845−1851.

(3) Frank-Kamenetskii, M. D.; Mirkin, S. M. Triplex DNA structures. *Annu. Rev. Biochem.* **1995**, *64*, 65−95.

(4) Mirkin, S. M. Triple-Helical Nucleic Acids *J. Am. Chem. Soc.* 1997, *119*. DOI: 10.1007/978-1-4612-3972-7.

(5) Bacolla, A.; Wells, R. D. Non-B DNA conformations, genomic rearrangements, and human disease. *J. Biol. Chem.* **2004**, *279*, 47411−47414.

(6) Choi, J.; Majima, T. Conformational changes of non-B DNA. *Chem. Soc. Rev.* **2011**, *40*, 5893−5909.

(7) Ravichandran, S.; Subramani, V. K.; Kim, K. K. Z-DNA in the genome: from structure to disease. *Biophys Rev.* **2019**, *11*, 383−387.

(8) Neidle, S. The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin Struct Biol.* **2009**, *19*, 239−250.

(9) Sathyaseelan, C.; Vijayakumar, V.; Rathinavelan, T. CD-NuSS: A Web Server for the Automated Secondary Structural Characterization of the Nucleic Acids from Circular Dichroism Spectra Using Extreme Gradient Boosting Decision-Tree, Neural Network and Kohonen Algorithms. *J. Mol. Biol.* **2021**, *433*, No. 166629.

(10) Abiri, A.; Lavigne, M.; Rezaei, M.; Nikzad, S.; Zare, P.; Mergny, J. L.; Rahimi, H. R. Unlocking G-Quadruplexes as Antiviral Targets. *Pharmacol Rev.* **2021**, *73*, 897−923.

(11) Saranathan, N.; Vivekanandan, P. G-Quadruplexes: More Than Just a Kink in Microbial Genomes. *Trends Microbiol.* **2019**, *27*, 148−163.

(12) Cammas, A.; Millevoi, S. RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res.* **2017**, *45*, 1584−1595.

(13) Tateishi-Karimata, H.; Sugimoto, N. Roles of non-canonical structures of nucleic acids in cancer and neurodegenerative diseases. *Nucleic Acids Res.* **2021**, *49*, 7839−7855.

(14) Miglietta, G.; Russo, M.; Capranico, G. G-quadruplex-R-loop interactions and the mechanism of anticancer G-quadruplex binders. *Nucleic Acids Res.* **2020**, *48*, 11942−11957.

(15) Yu, H.; Qi, Y.; Yang, B.; Yang, X.; Ding, Y. G4Atlas: a comprehensive transcriptome-wide G-quadruplex database. *Nucleic Acids Res.* **2023**, *51*, D126−D34.

(16) Katsuda, Y.; Sato, S. I.; Inoue, M.; Tsugawa, H.; Kamura, T.; Kida, T.; et al. Small molecule-based detection of non-canonical RNA G-quadruplex structures that modulate protein translation. *Nucleic Acids Res.* **2022**, *50*, 8143−8153.

(17) Lyu, K.; Chow, E. Y.; Mou, X.; Chan, T. F.; Kwok, C. K. RNA G-quadruplexes (rG4s): genomics and biological functions. *Nucleic Acids Res.* **2021**, *49*, 5426−5450.

(18) Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 459−474.

(19) Spiegel, J.; Adhikari, S.; Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* **2020**, *2*, 123−136.

(20) Yadav, P.; Kim, N.; Kumari, M.; Verma, S.; Sharma, T. K.; Yadav, V.; Kumar, A. G-Quadruplex Structures in Bacteria: Biological Relevance and Potential as an Antimicrobial Target. *J. Bacteriol.* **2021**, *203*, No. e0057720.

(21) Griffin, B. D.; Bass, H. W. Review: Plant G-quadruplex (G4) motifs in DNA and RNA; abundant, intriguing sequences of unknown function. *Plant Sci.* **2018**, *269*, 143−147.

(22) Warner, E. F.; Bohalova, N.; Brazda, V.; Waller, Z. A. E.; Bidula, S. Analysis of putative quadruplex-forming sequences in fungal genomes: novel antifungal targets? *Microb Genom.* **2021**, *7*, 000570 DOI: 10.1099/mgen.0.000570.

(23) Lejault, P.; Mitteaux, J.; Sperti, F. R.; Monchaud, D. How to untie G-quadruplex knots and why? *Cell Chem. Biol.* **2021**, *28*, 436−455.

(24) Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **2016**, *4*, 38 DOI: 10.3389/fchem.2016.00038.

(25) Zaccaria, F.; van der Lubbe, S. C. C.; Nieuwland, C.; Hamlin, T. A.; Fonseca Guerra, C. How Divalent Cations Interact with the Internal Channel Site of Guanine Quadruplexes. *ChemPhysChem* **2021**, *22*, 2265 DOI: 10.1002/cphc.202100761.

(26) Escaja, N.; Mir, B.; Garavis, M.; Gonzalez, C. Non-G Base Tetrads. *Molecules* **2022**, *27*, 5287 DOI: 10.3390/molecules27165287.

(27) Uttamrao, P. P.; Sundaresan, S.; Rathinavelan, T. Structure and Folding Patterns of RNA G-Quadruplexes. In *RNA Structure and Function*; Springer, 2023.

(28) Raguseo, F.; Chowdhury, S.; Minard, A.; Di Antonio, M. Chemical-biology approaches to probe DNA and RNA G-quadruplex structures in the genome. *Chem. Commun.* **2020**, *56*, 1317−1324.
(29) Rhodes, D.; Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **2015**, *43*, 8627−8637.
(30) Marsico, G.; Chambers, V. S.; Sahakyan, A. B.; McCauley, P.; Boutell, J. M.; Antonio, M. D.; Balasubramanian, S. Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.* **2019**, *47*, 3862−3874.
(31) Cogoi, S.; Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **2006**, *34*, 2536−2549.
(32) Stump, S.; Mou, T. C.; Sprang, S. R.; Natale, N. R.; Beall, H. D. Crystal structure of the major quadruplex formed in the promoter region of the human c-MYC oncogene. *PLoS One* **2018**, *13*, No. e0205584.
(33) Tran, P. L. T.; Mergny, J. L.; Alberti, P. Stability of telomeric G-quadruplexes. *Nucleic Acids Res.* **2011**, *39*, 3282−3294.
(34) Patel, D. J.; Phan, A. T.; Kuryavyi, V. Human telomere, oncogenic promoter and 5′-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* **2007**, *35*, 7429−7455.
(35) Mani, P.; Yadav, V. K.; Das, S. K.; Chowdhury, S. Genome-wide analyses of recombination prone regions predict role of DNA structural motif in recombination. *PLoS One* **2009**, *4*, No. e4399.
(36) Bugaut, A.; Balasubramanian, S. 5′-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* **2012**, *40*, 4727−4741.
(37) Fay, M. M.; Lyons, S. M.; Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.* **2017**, *429*, 2127−2147.
(38) Brown, V.; Jin, P.; Ceman, S.; Darnell, J. C.; O'Donnell, W. T.; Tenenbaum, S. A.; et al. Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **2001**, *107*, 477−487.
(39) Kumari, S.; Bugaut, A.; Balasubramanian, S. Position and stability are determining factors for translation repression by an RNA G-quadruplex-forming sequence within the 5′ UTR of the NRAS proto-oncogene. *Biochemistry.* **2008**, *47*, 12664−12669.
(40) Subramanian, M.; Rage, F.; Tabet, R.; Flatter, E.; Mandel, J. L.; Moine, H. G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep.* **2011**, *12*, 697−704.
(41) Tassinari, M.; Richter, S. N.; Gandellini, P. Biological relevance and therapeutic potential of G-quadruplex structures in the human noncoding transcriptome. *Nucleic Acids Res.* **2021**, *49*, 3617−3633.
(42) Ajjugal, Y.; Tomar, K.; Rao, D. K.; Rathinavelan, T. Spontaneous and frequent conformational dynamics induced by A···A mismatch in d(CAA).d(TAG) duplex. *Sci. Rep.* **2021**, *11*, No. 3689.
(43) Zhou, Z. D.; Jankovic, J.; Ashizawa, T.; Tan, E. K. Neurodegenerative diseases associated with non-coding CGG tandem repeat expansions. *Nat. Rev. Neurol.* **2022**, *18*, 145−157.
(44) Cheng, W.; Wang, S.; Mestre, A. A.; Fu, C.; Makarem, A.; Xian, F.; et al. C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2alpha phosphorylation. *Nat. Commun.* **2018**, *9*, No. 51.
(45) Endoh, T.; Takahashi, S.; Sugimoto, N. Endogenous G-quadruplex-forming RNAs inhibit the activity of SARS-CoV-2 RNA polymerase. *Chem. Commun.* **2023**, *59*, 872−875.
(46) Zhang, H.; Z, J.; Ye, Y. *Prediction and Validation of Circulating G-quadruplex as a Novel Biomarker in Colorectal Cancer*; Research Square, 2022.
(47) Xu, J.; Huang, H.; Zhou, X. G-Quadruplexes in Neurobiology and Virology: Functional Roles and Potential Therapeutic Approaches. *JACS Au* **2021**, *1*, 2146−2161.
(48) Mishra, S. K.; Jain, N.; Shankar, U.; Tawani, A.; Sharma, T. K.; Kumar, A. Characterization of highly conserved G-quadruplex motifs as potential drug targets in Streptococcus pneumoniae. *Sci. Rep.* **2019**, *9*, No. 1791.
(49) Tasset, D. M.; Kubik, M. F.; Steiner, W. Oligonucleotide inhibitors of human thrombin that bind distinct epitopes. *J. Mol. Biol.* **1997**, *272*, 688−698.
(50) Simone, R.; Fratta, P.; Neidle, S.; Parkinson, G. N.; Isaacs, A. M. G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. *FEBS Lett.* **2015**, *589*, 1653−1668.
(51) Shahsavar, K.; Hosseini, M.; Shokri, E.; Xu, G. New insight into G-quadruplexes; diagnosis application in cancer. *Anal. Biochem.* **2021**, *620*, No. 114149.
(52) Roxo, C.; Kotkowiak, W.; Pasternak, A. G-Quadruplex-Forming Aptamers-Characteristics, Applications, and Perspectives. *Molecules.* **2019**, *24*, 3781 DOI: 10.3390/molecules24203781.
(53) Neidle, S. Quadruplex Nucleic Acids as Novel Therapeutic Targets. *J. Med. Chem.* **2016**, *59*, 5987−6011.
(54) Xi, H.; Juhas, M.; Zhang, Y. G-quadruplex based biosensor: A potential tool for SARS-CoV-2 detection. *Biosens Bioelectron.* **2020**, *167*, No. 112494.
(55) Alberti, P.; Mergny, J. L. DNA duplex-quadruplex exchange as the basis for a nanomolecular machine. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 1569−1573.
(56) Koirala, D.; Dhakal, S.; Ashbridge, B.; Sannohe, Y.; Rodriguez, R.; Sugiyama, H.; et al. A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands. *Nat. Chem.* **2011**, *3*, 782−787.
(57) Sannohe, Y.; Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H. Visualization of dynamic conformational switching of the G-quadruplex in a DNA nanostructure. *J. Am. Chem. Soc.* **2010**, *132*, 16311−16313.
(58) Livshits, G. I.; Stern, A.; Rotem, D.; Borovok, N.; Eidelshtein, G.; Migliore, A.; et al. Long-range charge transport in single G-quadruplex DNA molecules. *Nat. Nanotechnol.* **2014**, *9*, 1040−1046.
(59) Jana, J.; Weisz, K. Thermodynamic Stability of G-Quadruplexes: Impact of Sequence and Environment. *Chembiochem.* **2021**, *22*, 2848−2856.
(60) Yuan, W. F.; Wan, L. Y.; Peng, H.; Zhong, Y. M.; Cai, W. L.; Zhang, Y. Q.; et al. The influencing factors and functions of DNA G-quadruplexes. *Cell Biochem. Funct.* **2020**, *38*, 524−532.
(61) Zok, T.; Kraszewska, N.; Miskiewicz, J.; Pielacinska, P.; Zurkowski, M.; Szachniuk, M. ONQUADRO: a database of experimentally determined quadruplex structures. *Nucleic Acids Res.* **2022**, *50*, D253−D258.
(62) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; et al. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(63) Coimbatore Narayanan, B.; Westbrook, J.; Ghosh, S.; Petrov, A. I.; Sweeney, B.; Zirbel, C. L.; et al. The Nucleic Acid Database: new features and capabilities. *Nucleic Acids Res.* **2014**, *42*, D114−D122.
(64) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006**, *34*, 5402−5415.
(65) Kwok, C. K.; Merrick, C. J. G-Quadruplexes: Prediction, Characterization, and Biological Application. *Trends Biotechnol.* **2017**, *35*, 997−1013.
(66) Karsisiotis, A. I.; O'Kane, C.; Webba da Silva, M. DNA quadruplex folding formalism–a tutorial on quadruplex topologies. *Methods.* **2013**, *64*, 28−35.
(67) Bochman, M. L.; Paeschke, K.; Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **2012**, *13*, 770−780.
(68) Lawson, C. L.; Berman, H. M.; Chen, L.; Vallat, B.; Zirbel, C. L. The Nucleic Acid Knowledgebase: a new portal for 3D structural information about nucleic acids. *Nucleic Acids Res.* **2024**, *52*, D245−D54.
(69) Williams, T.; Kea, C. Gnuplot 4.6: An Interactive Plotting Program; 2013.
(70) Dvorkin, S. A.; Karsisiotis, A. I.; Webba da Silva, M. Encoding canonical DNA quadruplex structure. *Sci. Adv.* **2018**, *4*, No. eaat3007.
(71) Maity, A.; Winnerdy, F. R.; Chang, W. D.; Chen, G.; Phan, A. T. Intra-locked G-quadruplex structures formed by irregular DNA G-rich motifs. *Nucleic Acids Res.* **2020**, *48*, 3315−3327.

https://doi.org/10.1021/acsomega.4c04579
ACS Omega 2024, 9, 38696−38709

(72) Farag, M.; Mouawad, L. Comprehensive analysis of intramolecular G-quadruplex structures: furthering the understanding of their formalism. *Nucleic Acids Res.* **2024**, *52*, 3522−3546, DOI: 10.1093/nar/gkae182.
(73) Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.* **2007**, *35*, 4927−4940.
(74) Largy, E.; Granzhan, A.; Hamon, F.; Verga, D.; Teulade-Fichou, M. P. Visualizing the quadruplex: from fluorescent ligands to light-up probes. *Top Curr. Chem.* **2013**, *330*, 111−177.
(75) Arora, A.; Maiti, S. Effect of loop orientation on quadruplex-TMPyP4 interaction. *J. Phys. Chem. B* **2008**, *112*, 8151−8159.
(76) Ma, Yue.; I, K.; N, K. *Macrocyclic G-Quadruplex Ligands of Telomestatin Analogs*; Springer Nature: Singapore, 2023.
(77) Ma, Y.; Tsushima, Y.; Sakuma, M.; Sasaki, S.; Iida, K.; Okabe, S.; et al. Development of G-quadruplex ligands for selective induction of a parallel-type topology. *Org. Biomol Chem.* **2018**, *16*, 7375−7382.
(78) Meier-Stephenson, V. G4-quadruplex-binding proteins: review and insights into selectivity. *Biophys Rev.* **2022**, *14*, 635−654.
(79) Ray, S.; Bandaria, J. N.; Qureshi, M. H.; Yildiz, A.; Balci, H. G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 2990−2995.
(80) Petraccone, L. Higher-order quadruplex structures. *Top Curr. Chem.* **2013**, *330*, 23−46.
(81) Zhao, J.; Zhai, Q. Recent advances in the development of ligands specifically targeting telomeric multimeric G-quadruplexes. *Bioorg Chem.* **2020**, *103*, No. 104229.
(82) Monsen, R. C. Higher-order G-quadruplexes in promoters are untapped drug targets. *Front Chem.* **2023**, *11*, No. 1211512.
(83) Bose, K.; Lech, C. J.; Heddi, B.; Phan, A. T. High-resolution AFM structure of DNA G-wires in aqueous solution. *Nat. Commun.* **2018**, *9*, No. 1959.
(84) Xu, Y.; Ishizuka, T.; Kurabayashi, K.; Komiyama, M. Consecutive formation of G-quadruplexes in human telomeric-overhang DNA: a protective capping structure for telomere ends. *Angew. Chem., Int. Ed.* **2009**, *48*, 7833−7836.
(85) Nishio, M.; Tsukakoshi, K.; Ikebukuro, K. G-quadruplex: Flexible conformational changes by cations, pH, crowding and its applications to biosensing. *Biosens. Bioelectron.* **2021**, *178*, No. 113030.

https://doi.org/10.1021/acsomega.4c04579
ACS Omega 2024, 9, 38696−38709