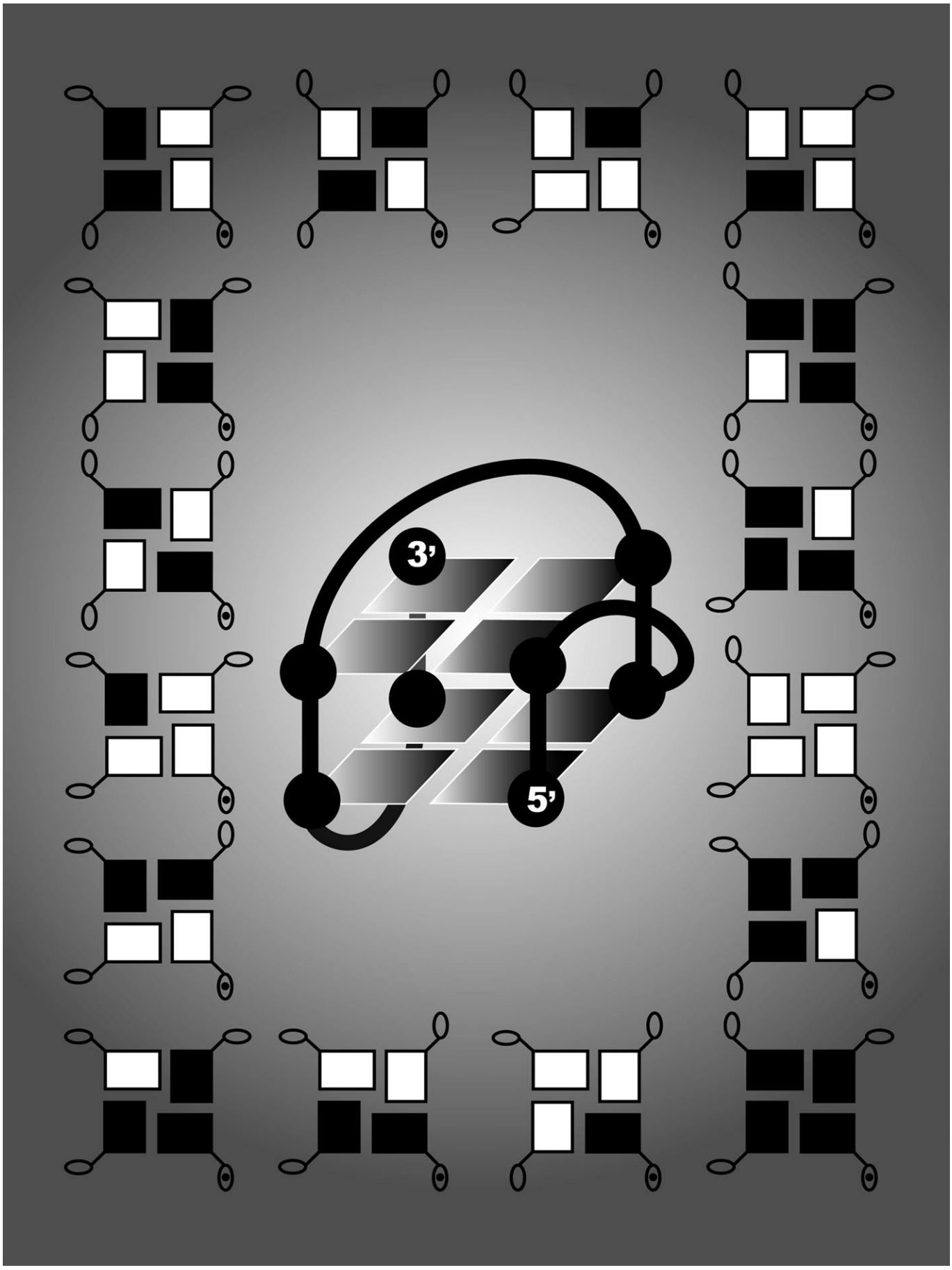
DOI: 10.1002/chem.200701255



# Geometric Formalism for DNA Quadruplex Folding

**Mateus Webba da Silva*[a]**

**Abstract:** Understanding the control of self-assembly and stereochemical properties of DNA higher order architectural folds is of fundamental importance in biology as well as biochemical technological applications. Guanine-rich DNA sequences can form tetrahelical architectures termed quadruplexes. A formalism is presented describing the interdependency of a set of structural descriptors as a geometric basis for folding of unimolecular quadruplex topologies. It represents a standard for interpretation of structural characteristics of quadruplexes, and is comprehensive in explicitly harmonizing the results of published literature with a unified language. The formalism is a fundamental step towards prediction of unimolecular quadruplex folding topologies from primary sequence.

**Keywords:** DNA structures · glycosides · guanine

## Introduction

Inter- and intramolecular interactions modulate architectural motifs to control the structure of biomolecules. An understanding of this interplay is the basis for discovering the nature of biologically relevant structural motifs, and understanding the principles for the design of technological materials based on biopolymers. It is now understood that the primary sequence of DNA encodes for non-duplex topologies. Specifically, guanine-rich DNA sequences can form structurally diverse architectures, termed quadruplexes, that are associated with the telomeres, the control of gene expression,[1] and other biological functions.[2,3] Concurrently, there is an intense focus on technological applications of these architectures.[4] In the medical field DNA quadruplexes are of current interest due to their implication in cancer, diabetes, ageing, central nervous system, and cardiovascular diseases and are thus targets of drug design.[5–7] These guanine-rich sequences of DNA are highly polymorphic, adopting a variety of quadruplex topologies depending on primary sequence and environmental conditions (see e.g., reference [8]). The precise relationship between DNA sequence and folding into quadruplexes is fundamentally dependent on a hitherto undefined structural basis. Here we describe a formalism detailing the geometric structural requirements for topologies of unimolecular DNA quadruplexes.

[a] Dr. M. Webba da Silva
School of Biomedical Sciences, University of Ulster
W2064, CMB, Cromore Road, BT52 1SA (UK)
Fax: (+44) 28-7032-4375
E-mail: mm.webba-da-silva@ulster.ac.uk

## The Quadruplex Stem

In quadruplexes, four guanine bases align in a pseudo-plane through hydrogen-bond alignments involving the Watson–Crick edge of a guanine and the Hoogsteen edge of its partner; resulting in a (G:G:G:G) tetrad, Figure 1 (top). The quadruplex stem is composed of stacked tetrads with phosphodiester backbones delimiting cavities denominated grooves. The tetrads are held together by cations and interactions of π orbitals of stacked aromatic bases. To systematically describe quadruplex topologies in order to elucidate the rules for folding unimolecular quadruplexes, we classified tetrads according to the glycosidic bond angle (GBA) of the intervening bases. These can assume either an *anti* or a *syn* disposition (Figure 1, bottom).

By defining a frame of reference (Scheme 1), we can classify tetrads according to the GBA of the intervening bases. Thus, there are 16 possible GBA positions for (G:G:G:G) tetrads, Figure 2. A direct consequence of identifying the GBA is the definition of grooves spanning contiguous backbones. The type of groove is defined by the GBA of contiguous bases in the tetrad.[9,10] There are three possible groove types for a tetrad: narrow (n), medium (m), and wide (w; Figure 2 Ia). Please, note that these definitions are derived from the disposition of the two GBAs bridging the groove rather than its actual size. Narrow, as well as wide grooves, are the result of hydrogen-bond-aligned bases with different GBA. If contiguous bases in a tetrad have the same GBA,

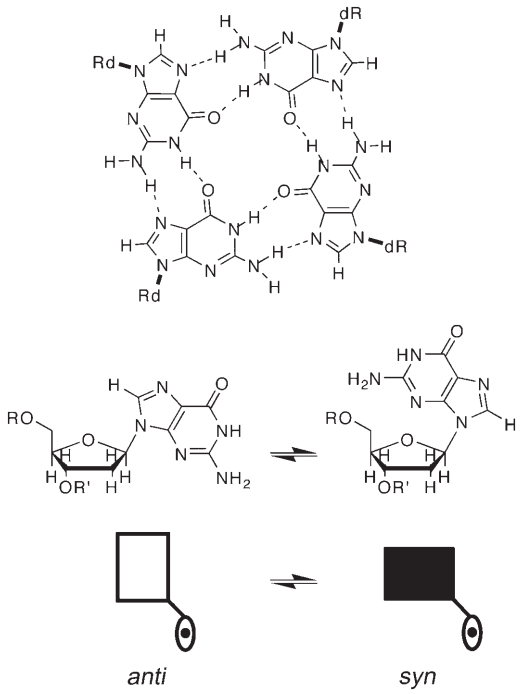
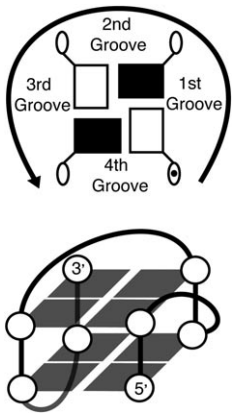



Figure 1. Top: The hydrogen bonding alignment of the (G:G:G:G) tetrad. Bottom: Chemical structures of *anti*- and *syn*-glycosidic bond angles. The two dispositions describing the orientations of the glycosidic bond angle are shown below.




Scheme 1. On top the frame of reference for describing quadruplex topologies. The origin, the 5′-end, sits on a strand of the quadruplex stem progressing towards the viewer with polarity indicated by (●) on the sugar-pucker. Looping can be anticlockwise, as denoted by the arrow, or clockwise; (−) and (+), respectively. The description of the grooves follows anticlockwise from first to fourth. The bottom scheme depicts loops linking strands in a quadruplex stem. From 5′ to 3′ the first loop is a propeller, the second a diagonal, and the third a lateral loop.

they form a medium groove. According to the GBA of the bases in a tetrad, there are only eight possible groove width combinations. This is a structural limiting factor that serves a fundamental axiom. Only tetrads with the same groove width combinations may stack to form stable quadruplexes; that is, the tetrad combination Ia stacks only with itself and with Ib, IIa stacks only with itself and IIb, and so on. Thus, a quadruplex stem composed of any number of (G:G:G:G)

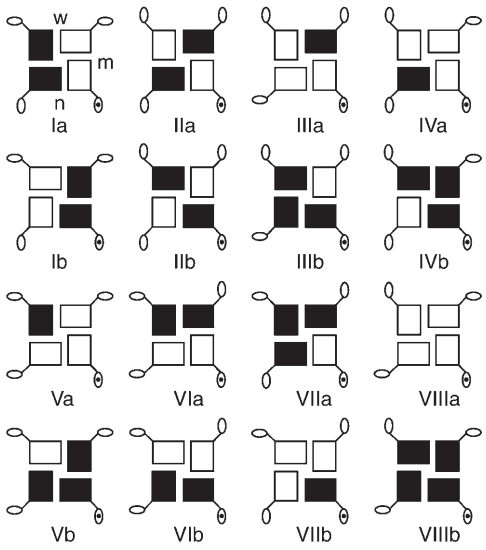



Figure 2. All possible combinations of GBA for (G:G:G:G) tetrads. In Ia, the definitions of medium (m), wide (w), and narrow grooves (n) according to the adjacent disposition of GBA are shown.

tetrads is defined by one, or a maximum by two GBA combinations of stacked tetrads of the same groove width disposition. Tetrads of the same, or different grove disposition, triads, or mismatches may be loosely stacked onto these.

## Relationship between Loops and GBA

Loops are here defined as the biopolymer chains linking the strands in a quadruplex stem. They may be involved in formation of further pseudo-planar architectures. A unimolecular quadruplex topology is defined by a minimum of three loops. There are three different types: diagonal, propeller (also known as double-chain reversal), and lateral (also known as edgewise)—Scheme 1. Their categorization is simple. Diagonal loops link bases of the same tetrad that do not share hydrogen bonds. In contrast, lateral loops link bases of the same tetrad that share hydrogen bonds. For both cases, the linked bases belong to antiparallel strands and result in different GBA.[10] Lateral loops may thus result in narrow or wide grooves, irrespective of starting on either GBA. Propeller loops link bases in the quadruplex stem that are not in the same tetrad but share a groove. Thus, the contiguous strands result parallel irrespective of the combination of GBA of the intervening bases. However, in the context of single tetrads, the grooves of propeller loops are composed of bases with the same GBA. This means that propeller loops invariably link medium grooves within a quadruplex stem.

## Looping Topologies

Loops spanning first–second–third grooves, in that order, progress anticlockwise; and conversely, loops spanning

third–second–first grooves show clockwise progression. Since a unimolecular quadruplex topology can be defined by three loops, and there are three loop types, we thus have 27 theoretical loop combinations. However, many of these are sterically not permissible. For example a topology that would include two sequential diagonals is impossible, and one that includes the diagonal-lateral-diagonal is highly improbable. The exclusion of these combinations results in 13 possible combinations starting from a clockwise loop, and 13 from an anticlockwise loop. Thus, there are 26 permissible looping combinations, which are described in Figure 3. Irrespective of the clockwise or anticlockwise progression of loops in a quadruplex, all topologies can be right handed. Eight topologies have loops progressing in a clockwise manner (topologies 1b–4b and 6b–9b), and the same number have an anticlockwise progression of loops- respectively (topologies 1a–4a and 6a–9a). There are four more pairs of topologies in which diagonal loops intervene to change the direction of the strand progression (5a/5b, 10a/10b, 11a/11b, and 13a/13b), and a pair of topologies with two diagonal loops (12a/12b). The pairing of suffixes "a" and "b" to a numeral does not imply that they are stereoisomers. In fact, only pairs 1a/1b, 4a/4b, 6a/6b, and 9a/9b depict pseudo-stereoisomeric topologies.

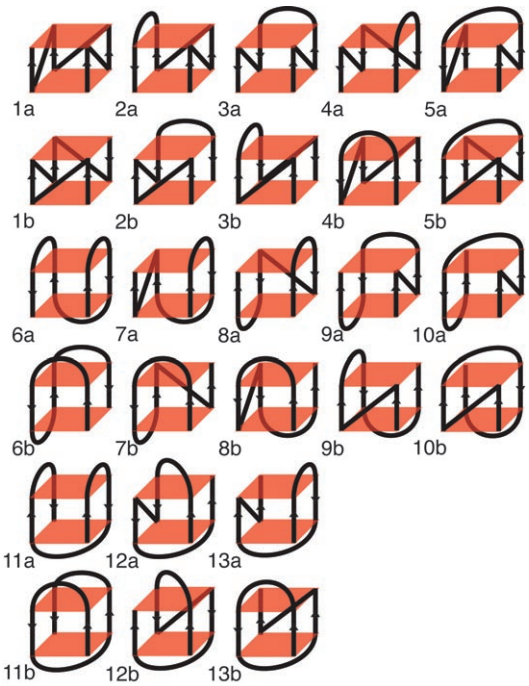



Figure 3. Representations of all looping topologies possible for three loop unimolecular quadruplex topologies. The topologies denoted by "a" start with anticlockwise progressing loops, and conversely the topologies denoted by "b" start with clockwise progressing loops.

## Disambiguation of Chain Progression

Irrespective of the right- or left-handedness of the groove propagation, the biopolymer chain will assume at every looping opportunity a clockwise or anticlockwise progression. By inspecting the possible tetrad combinations in Figure 2 it becomes apparent that: immediately from the 5′-end 1) no anticlockwise narrow grooves are possible, and 2) no clockwise wide grooves are possible. Thus, it is not only the propensity of the primary sequence to assume a particular fold that drives quadruplex folding. The stacking of tetrads with the same combinations of GBA is determinant in allowing for the assembly of the topology.

## Looping and Steric Considerations

The propensity for formation of a particular loop type is dependent on its primary sequence and its length.[11–16] At current understanding, there is no firm set of rules that establishes a unique propensity for folding quadruplexes from DNA sequences alone. However, there are some observations that merit consideration. Thus, a one-residue loop is characteristic of propeller loops between two[17] or three[18] stacked tetrads. One[19] or two[20] residue loops can also be characteristic of a lateral narrow loop, but would be too short a DNA chain to form a diagonal loop. The latter can be formed with at least three residues.[21] If a three-residue loop starts with a pyrimidine, both diagonal loops[22] and lateral loops with wide grooves[19] are preferred, but a three-layer propeller is also possible.[21] It is possible to fold diagonal loops starting with guanine moieties,[23] but there has not been any example of such a loop starting with adenines. Propeller loops are rather versatile. There has been one such loop exclusively composed of sugar-phosphate backbone.[24] Four-residue loops are not favoured in lateral narrow loops. This and greater number of residues in the biopolymer chain can be potentially accommodated by lateral wide,[25] diagonal,[26] and propeller loops. This is especially true when there is a possibility to form hydrogen-bond alignments for the looping bases.[27]

Further considerations to account for the propensity for loop formation are the dependency of primary sequence on flexibility, and dynamic (fluxional) motions. In general, the propensity for folding into loops with more than one residue is, at least in part, determined by the relative base-stacking energies of the sequence of residues in the loop. If tracks of two or more consecutive adenines make 50% or more of the number of residues in a potential loop, it is unlikely to fold due to the rigidity of an A-track effect.[28] Loop fluxional motions are brought about by two sequential loops bearing guanines at their 5′ and 3′-ends; that is, they are formed by segments bearing different numbers of sequential guanine moieties. This fluxional behaviour results in equilibria within the same fold.[29]

## Sequence of Stacking Tetrads

For each topology there are two alternatives per pair of stacked tetrads with the same groove width combination. For an $n$ number of stacked tetrads there would be $2^n$ possibilities for stacking. Thus, there is a need to address the sequence of stacked tetrads that are possible within a particular topology. We can use the GBA to further classify the stacking interactions in the quadruplex stem. Thus, within a strand of two stacked tetrads there are four possibilities of propagation of the GBA: *syn–anti*, *anti–syn*, *anti--anti*, or *syn--syn* (Figure 4). For each pair of intrastrand stacked

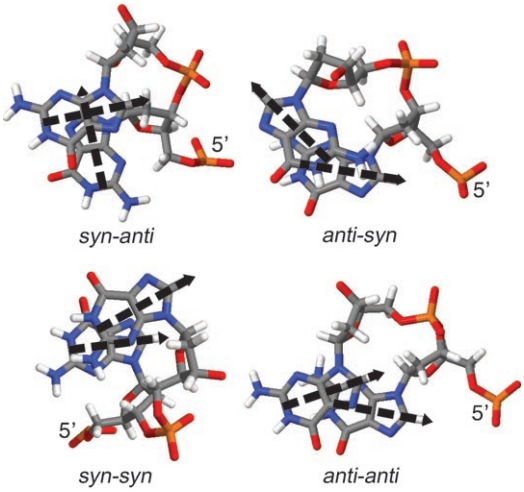



Figure 4. Intrastrand sequencial stacking of guanines in the quadruplex stem in the structures of the human telomeric sequences [$T_3$-($G_3AT_2$)$_3G_3A$] PDB 2GKU,[30] and d[$AG_3$($T_2AG_3$)$_3$] PDB 143D.[31] The 5′ to 3′ progression of the backbone is towards the viewer. The base-steps *syn–anti*,[30] *anti–syn*,[31] *syn–syn*,[30] and *anti–anti*[30] are depicted. The idealized angles between C8→H8 vectors of stacked guanines are −90°, +142°, +26°, and +26°; respectively.

guanines, we consider the projection of their C8→H8 vectors into a common plane as described by the view down the stem axis. The angle between the two vectors thus describes the intrastrand base-step twist angle (STA). For a *syn–anti* intrastrand stacked step the projection of both vectors results in an approximately −90° angle relative to each other. The converse *anti–syn* step results approximately in a +142° angle. In contrast, both *anti–anti*, and *syn–syn* steps result in an angle cantered at +26°.[32] For diagonal and lateral loops the observed tendency is for propagation of alternating GBA within the stem. That is explained by tendency for achieving the typical helical twist of +26°; that is a *syn–anti* STA of −90° combined with an *anti--syn* STA of +142° results in the median helical twist of +26° per step. By observing mono-, di- (e.g., see reference [23]), and tetramolecular[33,34] quadruplexes, we conclude that the sum of the STA of two sequential base-steps in each of the four strands tends to achieve the typical helical twist (+26°) in order to be stable. The set of quadruplex architectures determined thus far may not be a sufficiently representative sample for deriving the ranges possible. However, this descriptor will already limit the next tetrad that stacks to achieve a permissible twist angle.

## Notation and Structural Descriptors

We first consider a frame of reference with an origin at its 5′-end in the quadruplex stem, and the first strand progressing towards the viewer. Irrespective of the clockwise (+) or anticlockwise (−) progression of the bases, we ensure appropriate descriptors for right-handed quadruplexes by designating only one possible scheme position for each of *anti* and *syn* GBA (Figure 1, bottom). This representation guarantees all possible orientations resultant from hydrogen-bonding alignments. The placement of loops in a quadruplex is thus described sequentially by type and clockwise or anticlockwise progression from 5′ to 3′ end as they appear in the primary sequence and denoted in parentheses. For example, the descriptor for a quadruplex with a propeller, followed by a diagonal, and then a lateral loop appearing as the primary sequence all progressing in an anticlockwise manner is denoted −(pdl). In contrast, the occurrence of grooves in a quadruplex is described sequentially strictly in an anticlockwise manner from the 5′-end (Scheme 1). This is irrespective of the clockwise or anticlockwise progression of the primary sequence in the fold. As for looping architectures, the descriptor for groove width is denoted in parentheses. For example, the groove description for a quadruplex containing medium, wide, narrow, and again medium grooves appearing in sequence in a anticlockwise manner is denoted (mwnm). The descriptor for strand direction (polarity) of the strands in the quadruplex stem is relative to the first strand defined in the frame of reference. As for the descriptor for groove width disposition, strands are compared to the first strictly in an anticlockwise manner irrespective of the progression of the primary sequence. Thus, the looping architecture described in Figure 3, structure 6b, has a first antiparallel strand as compared to the strand of the frame of reference, followed by a parallel, and finally an antiparallel strand; thus the descriptor takes the form (apa).

## Description of Topologies

The topologies theoretically possible for unimolecular quadruplexes of three loops are described in Table 1. The looping architecture is described, with the respective tetrad combination, groove description, direction of loop progression and strand directionalities. The sequence of stacking tetrads in the topologies is not described, but may be derived if the number of stacked tetrads is known. Of the 26 theoretically possible, only seven of the topologies have been verified experimentally thus far. The remaining topologies are potentially amenable to experimental verification by structure determination utilizing X-ray crystallography or NMR spec-

Table 1. Architectural elements of unimolecular three loop DNA quadruplex topologies.

| Looping description[a] | Loop progression | Tetrad combination[b] | Groove description | Strand disposition | Ref. |
|---|---|---|---|---|---|
| 1a | −(ppp) | VIII | (mmmm) | (ppp) | [21,38,39,18] |
| 1b | +(ppp) | VIII | (mmmm) | (app) | |
| 2a | −(ppl) | IV | (mwmn) | (ppa) | |
| 2b | +(ppl) | III | (wnmm) | (app) | |
| 3a | −(plp) | I | (mwmn) | (paa) | |
| 3b | +(plp) | VI | (wmnm) | (aap) | |
| 4a | −(lpp) | VII | (wmmn) | (aaa) | |
| 4b | +(lpp) | VII | (wmmn) | (aaa) | |
| 5a | (−pd+p) | I | (wnmn) | (paa) | |
| 5b | (+pd−p) | VI | (wmnm) | (pap) | |
| 6a | −(lll) | II | (wnwn) | (apa) | |
| 6b | +(lll) | II | (wnwn) | (apa) | [20] |
| 7a | −(llp) | III | (wnmm) | (app) | [42,43] |
| 7b | +(llp) | IV | (mmwn) | (ppa) | |
| 8a | −(lpl) | VI | (wmnm) | (aap) | [46] |
| 8b | +(lpl) | I | (mwmn) | (paa) | |
| 9a | −(pll) | V | (mwnm) | (pap) | [30,41] |
| 9b | +(pll) | V | (mwnm) | (pap) | |
| 10a | (−pd+l) | IV | (mmwn) | (ppa) | |
| 10b | (+pd−l) | III | (wnmm) | (app) | |
| 11a | (−ld+l) | VI | (wmnm) | (aap) | [31,22] |
| 11b | (+ld−l) | I | (mwmn) | (paa) | |
| 12a | (d−pd) | I | (mwmn) | (paa) | |
| 12b | (d+pd) | VI | (wmnm) | (aap) | [45] |
| 13a | (−ld+p) | III | (wnmm) | (app) | |
| 13b | (+ld−p) | IV | (mmwn) | (ppa) | |

[a] The key numerals appear in Figure 3 for looping description. [b] The key numerals appear in Figure 2 for tetrad combination.

troscopy. We thus inspected all the structures deposited in the nucleic acids database[35] to date. The application of this formalism to representative topologies of the set of structures that have been experimentally verified utilizing these two techniques[35] is exemplified below.

The topology described by three sequential propeller loops all progressing in an anticlockwise manner, −(ppp), is frequently observed. How do we identify the stacking tetrad dispositions for this looping architecture? Since all loops are propeller they must link medium grooves, thus only the groove width arrangements of VIIIa and VIIIb in Figure 2 are possible. The relative propensity for purines for an *anti*-GBA[36] makes the tetrad disposition VIIIa the only one observed thus far. These topologies will have all-parallel strands. Correspondingly, strand directionality and groove width descriptors assume the form (ppp) and (mmmm), respectively. There are bimolecular[18] and tetramolecular[17,33,34,37] topologies made up solely of propeller loops. A unimolecular topology described by −(ppp) looping includes the X-ray crystal structure of the human telomeric sequence $d[AG_3(TTAG_3)_3]$,[21] the solution structure of a double-mutated form of a silencer element for c-MYC transcription $d[TG(AG_3TG_3T)_2AA]$,[38] the solution structure of an inosine-modified sequence representing a biologically relevant quadruplex element also of the human c-MYC promoter,[39] the $d[G_4TG_3(AGG)_2GT]$ aptamer inhibitor of HIV-1 integrase,[18] and other sequences.[40]

The topology described by −(pll) looping is exemplified in the solution structure of a mutated human telomeric sequence $d[A_3(G_3T_2A)_3G_3A_2]$[30] and base-modified human telomeric sequences.[30,41] The strand disposition is (pap), thus alternating parallel and antiparallel strands. It also has all possible types of grooves: (mwnm). Combinations of stacked Va/Vb tetrads are possible for this topology. However, depending on its number particular consideration has to be paid to the sequence of stacked tetrads to conform to a favourable STA.

The −(llp) looping topology is exemplified in the solution structure of a tetrahymena telomeric repeat $d[(T2G4)_4]$,[42] and a mutant of a guanine-rich region within the promoter of the BCL-2 gene.[43] In it, a wide groove bridges the first and the adjacent antiparallel strand, followed by a narrow groove with a change in strand polarity. The medium groove that follows implies no change in the polarity of the strand. The last groove is also between parallel adjacent strands and thus is bridged by a medium groove. In summary, the (app) strand disposition links the (wnmm) groove description. If, however, looping starts in a clockwise manner +(llp) groove description, tetrad combination, and strand disposition would be different. Even though, there has not been any structure determination for this looping architecture, the solution structure of an asymmetrical two-stranded molecule in which one of the strands has the sequential loop motif +(ll) has been determined.[44]

The (−ld+l) looping topology has been described for the solution structures of the $d[AG_3(T_2AG_3)_3]$ human telomeric sequence[31] and the $d[G_4(T_4G_4)_3]$ Oxytricha telomeric repeat.[22] It has antiparallel and parallel strands (aap) and a (wmnm) groove description. For this particular topology, a combination of VIa and VIb tetrad combinations is necessary for a permissible STA. The same tetrad combination is required for the (d+pd) looping disposition. The latter has been experimentally verified in the sequence $d[G_2T_4G_2CAG_3T_4G_2T]$.[45] In it a (aap) strand disposition links a (wnmm) groove description for the looping architecture 12b in Figure 3.

The −(lpl) loop progression has been described for a solution structure of the $d[(T_2AG_3)_4T_2]$ human telomeric sequence in potassium.[46] It has two parallel, and one antiparallel strands (aap), and a (wmnm) groove description. The stem of this particular topology in made up of tetrad combinations VIa and VIb.

A sequence of three lateral loops is known to adopt a clockwise disposition, +(lll) in the solution structure of the thrombin-binding aptamer $d[G_2T_2G_2TGTG_2T_2G_2]$.[20] The first lateral loop of two nucleotides thus results in a narrow groove. If a single-stranded quadruplex starts with a narrow groove, it can only be with a clockwise loop progression. For the current case, the progression of alternate GBA for the three loops can only result in the tetrad combination IIa/IIb. Therefore, the +(lll) looping disposition describes the 6b looping topology with (wnwn) grooves. Bimolecular right-handed quadruplexes folded exclusively through lateral loops have been experimentally determined in solution for $d[(G_3T_2CAG_2)_2]$.[47] Accordingly, the loops adopt a clockwise strand progression. A −(lll) looping disposition is also theoretically possible. However, the first loop has to be longer than two nucleotides to accommodate a wide groove. Since the tetrad combination is the same, both −(lll) as well as +(lll) looping dispositions result in (wnwn) grooves and are therefore true stereoisomeric topologies.

## Strand Interruptions in the Quadruplex Stem

Quadruplexes that have interrupted stem strands are made up of more than three loops.[39,48] For every number of interruptions, $I$, there are $(3+I)$ loops. Provided there is no change in the strand directionality of the interrupted strand, a finite number of loops are theoretically possible in a single quadruplex topology that depends on the number of stacked tetrads. This proviso is not a new concept, but derives from the requirement of stacking tetrads with the same groove width combination. An example is the structure of $d[(AG_3)_2CGCTG_3(AG_2)_2G]$ appearing in the c-Kit promoter determined by solution NMR spectroscopy.[47] The looping disposition is identified as (−p−p−p+p), and thus with all medium grooves as described in the tetrad combination VIIIa. Accordingly all strands are parallel. The three stacked tetrads allow for a single *anti*-guanine to mediate two propeller loops.

## Concluding Remarks

The progression from observation, to theory, to prediction is a means to unify and consolidate our understanding of folding of biopolymers. Here we utilize the previously described concept of groove widths derived from the glycosidic bond angle[9] to introduce three further concepts and analyze their interdependency. The formalism thus created is based on a comprehensive assessment of the publicly available set of three-dimensional structures determined thus far. It establishes the interdependency of glycosidic bond angle, strand polarity, groove width combination, and type of loop and thus provides the geometric structural basis for folding of unimolecular DNA quadruplexes. It is a fundamental step towards prediction of unimolecular quadruplex topologies from primary sequence, and of general use as an interpretative basis for the identification of the fold, as well as design of quadruplex architectures.

## Acknowledgement

The author gratefully acknowledges The Royal Society (UK) for their generous support.

[1] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, S. Neidle, *Nucleic Acids Res.* **2006**, *34*, 5402.
[2] N. Maizels, *Nat. Struct. Mol. Biol.* **2006**, *13*, 1055.
[3] M. Fry, *Front. Biosci.* **2007**, *12*, 4336.
[4] J. T. Davis, *Angew. Chem.* **2004**, *116*, 684; *Angew. Chem. Int. Ed.* **2004**, *43*, 668.
[5] J. L. Mergny, C. Helene, *Nat. Med.* **1998**, *4*, 1366.
[6] L. H. Hurley, *Nat. Rev. Cancer* **2002**, *2*, 188.
[7] S. Neidle, G. Parkinson, *Nat. Rev. Drug Discovery* **2002**, *1*, 383.
[8] A. T. Phan, V. Kuryavyi, K. N. Luu, D. J. Patel, *Structural Diversity of G-Quadruplex Scaffolds*, Royal Society of Chemistry, Cambridge (UK), **2006**.
[9] F. W. Smith, J. Feigon, *Nature* **1992**, *356*, 164.
[10] D. Mohanty, M. Bansal, *Nucleic Acids Res.* **1993**, *21*, 1767.
[11] P. A. Rachwal, T. Brown, K. R. Fox, *Biochemistry* **2007**, *46*, 3036.
[12] A. Risitano, K. R. Fox, *Nucleic Acids Res.* **2004**, *32*, 2598.
[13] M. Cevec, J. Plavec, *Biochemistry* **2005**, *44*, 15238.
[14] P. Hazel, J. Huppert, S. Balasubramanian, S. Neidle, *J. Am. Chem. Soc.* **2004**, *126*, 16405.
[15] P. Hazel, G. N. Parkinson, S. Neidle, *Nucleic Acids Res.* **2006**, *34*, 2117.
[16] M. Zuker, *Nucleic Acids Res.* **2003**, *31*, 3406.
[17] A. Kettani, A. Gorin, A. Majumdar, T. Hermann, E. Skripkin, H. Zhao, R. Jones, D. J. Patel, *J. Mol. Biol.* **2000**, *297*, 627.
[18] A. T. Phan, V. Kuryavyi, J. B. Ma, A. Faure, M. L. Andreola, D. J. Patel, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 634.
[19] N. Zhang, A. Gorin, A. Majumdar, A. Kettani, N. Chernichenko, E. Skripkin, D. J. Patel, *J. Mol. Biol.* **2001**, *312*, 1073.
[20] P. Schultze, R. F. Macaya, J. Feigon, *J. Mol. Biol.* **1994**, *235*, 1532.
[21] G. N. Parkinson, M. P. Lee, S. Neidle, *Nature* **2002**, *417*, 876.
[22] Y. Wang, D. J. Patel, *J. Mol. Biol.* **1995**, *251*, 76.
[23] M. Crnugelj, N. V. Hud, J. Plavec, *J. Mol. Biol.* **2002**, *320*, 911.
[24] M. Cmugelj, P. Sket, J. Plavec, *J. Am. Chem. Soc.* **2003**, *125*, 7866.
[25] J. X. Dai, T. S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R. A. Jones, D. Z. Yang, *J. Am. Chem. Soc.* **2006**, *128*, 1096.
[26] P. Schultze, N. V. Hud, F. W. Smith, J. Feigon, *Nucleic Acids Res.* **1999**, *27*, 3018.
[27] V. Kuryavyi, A. Kettani, W. M. Wang, R. Jones, D. J. Patel, *J. Mol. Biol.* **2000**, *295*, 455.
[28] A. Barbic, D. P. Zimmer, D. M. Crothers, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2369.
[29] J. Seenisamy, E. M. Rezler, T. J. Powell, D. Tye, V. Gokhale, C. S. Joshi, A. Siddiqui-Jain, L. H. Hurley, *J. Am. Chem. Soc.* **2004**, *126*, 8702.
[30] K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix, D. J. Patel, *J. Am. Chem. Soc.* **2006**, *128*, 9963.
[31] Y. Wang, D. J. Patel, *Structure* **1993**, *1*, 263.
[32] G. D. Strahan, M. A. Keniry, R. H. Shafer, *Biophys. J.* **1998**, *75*, 968.
[33] M. Webba da Silva, *Biochemistry* **2003**, *42*, 14356.
[34] M. Webba da Silva, *Biochemistry* **2005**, *44*, 3754.
[35] H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S. H. Hsieh, A. R. Srinivasan, B. Schneider, *Biophys. J.* **1992**, *63*, 751.
[36] C. F. Tang, R. H. Shafer, *J. Am. Chem. Soc.* **2006**, *128*, 5966.
[37] J. L. Mergny, A. De Cian, S. Amrane, M. Webba da Silva, *Nucleic Acids Res.* **2006**, *34*, 2386.
[38] A. Ambrus, D. Chen, J. X. Dai, R. A. Jones, D. Z. Yang, *Biochemistry* **2005**, *44*, 2048.

[39] A. T. Phan, V. Kuryavyi, H. Y. Gaw, D. J. Patel, *Nat. Chem. Biol.* **2005**, *1*, 167.
[40] A. T. Phan, Y. S. Modi, D. J. Patel, *J. Am. Chem. Soc.* **2004**, *126*, 8710.
[41] A. Matsugami, Y. Xu, Y. Noguchi, H. Sugiyama, M. Katahira, *FEBS J.* **2007**, *274*, 3545.
[42] Y. Wang, D. J. Patel, *Structure* **1994**, *2*, 1141.
[43] J. X. Dai, D. Chen, R. A. Jones, L. H. Hurley, D. Z. Yang, *Nucleic Acids Res.* **2006**, *34*, 5133.
[44] N. Zhang, A. T. Phan, D. J. Patel, *J. Am. Chem. Soc.* **2005**, *127*, 17277.
[45] V. Kuryavyi, A. Majumdar, A. Shallop, N. Chernichenko, E. Skripkin, R. Jones, D. J. Patel, *J. Mol. Biol.* **2001**, *310*, 181.
[46] A. Kettani, G. Basu, A. Gorin, A. Majumdar, E. Skripkin, D. J. Patel, *J. Mol. Biol.* **2000**, *301*, 129.
[47] A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle, D. J. Patel, *J. Am. Chem. Soc.* **2007**, *129*, 4386.
[48] F. W. Smith, P. Schultze, J. Feigon, *Structure* **1995**, *3*, 997.

Published online: October 30, 2007