DOI: 10.1002/anie.200902454

Quadruplex Self-Assembly

Design of a G-Quadruplex Topology through Glycosidic Bond Angles**

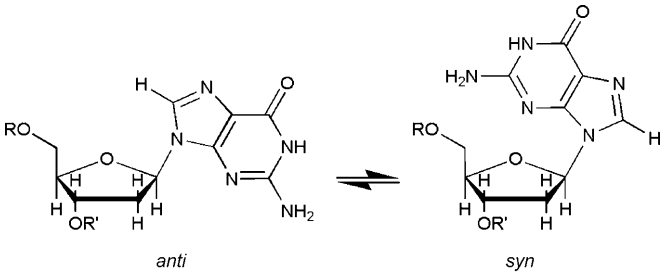
Mateus Webba da Silva, Marko Trajkovski, Yuta Sannohe, Nason Ma'ani Hessari, Hiroshi Sugiyama, and Janez Plavec**

To realize a programmed build up of DNA objects, devices, and materials, a systematization of the principles that form the basis of control for the assembly process is necessary. Thus far the discovery of four-stranded DNA architectures denominated G quadruplexes was either serendipitous or through coincidental emergence. Herein we utilize a rational approach for the design of G quadruplexes based on the two-state disposition of the glycosidic bond angle.

The duplex topology of DNA has a well-defined geometry, its self-assembly is highly predictable for forming programmable intra- and intermolecular interactions, the sequential context of its components is highly variable, and it can be conveniently chemically modified. As a result of its structural variability, general high-temperature stability, and the feasibility of controlling its dynamic behavior, DNA G quadruplexes can, in principle, be used for the same applications as double-helical DNA, and constitute suitable building blocks as scaffolds for functional materials and biotechnology. However, in contrast to double-helical DNA the rules for self-assembly of G quadruplexes are yet to be understood. Herein we demonstrate that by applying a formalism based on the two preferred glycosidic bond angle (χ) dispositions of guanosines in the quadruplex stem,[1] it is feasible, in principle, to derive a novel quadruplex topology by controlling the fold through the length of its loops.

DNA quadruplex structures are likely to be formed in vivo, having roles in key biological processes such as the maintenance of telomeres, regulation of gene transcription, DNA recombination, and packaging of the retroviral genome. G quadruplexes result from the stacking of (G:G:G:G) tetrads to form a quadruplex stem held together by cations. The stacking of tetrads is interdependent with the glycosidic torsion angles (χ) adopted by guanosines in the quadruplex stem—also known as glycosidic bond angle (GBA; Scheme 1). The individual G-rich tracts within a quadruplex stem are linked by "loops" which define the topology of the architecture. By defining a frame of reference for unimolecular quadruplex folds, it is possible to utilize the conformation of the glycosidic bond angle to derive the interdependency of loop type, groove–width combination, and tetrad combination which define the 26 possible folding topologies of unimolecular quadruplexes of three loops.[1] Loops can 1) bridge a quadruplex stem diagonally across a tetrad (diagonal loops), 2) bridge stacking tetrads in a medium groove (propeller loops), and 3) bridge both lateral-narrow and lateral-wide grooves (lateral loops).

Interdependencies of these parameters allow the description of the topologies. In this context, for example, parallel quadruplexes are composed of guanosines in a single tetrad GBA combination: (*anti*G:*anti*G:*anti*G:*anti*G). Alternatively, a theoretical all-parallel quadruplex composed of all-*syn*



Scheme 1. Chemical structures of *anti* and *syn* conformers of 2′-deoxyguanosine.

[*] Dr. M. Webba da Silva
School of Biomedical Sciences, University of Ulster
and Biomedical Sciences Research Institute, University of Ulster
Cromore Road, Coleraine BT52 1SA (UK)
Fax: (+44) 28-7032-4375
E-mail: mm.webba-da-silva@ulster.ac.uk
Prof. J. Plavec
Slovenian NMR Center, National Institute of Chemistry
Hajdrihova 19, 1000 Ljubljana (Slovenia)
and Faculty of Chemistry and Chemical Technology
University of Ljubljana, Aškerčeva 5, 1000 Ljubljana (Slovenia)
Fax: (+44) 28-7032-4375
E-mail: janez.plavec@ski.si
M. Trajkovski
Slovenian NMR Center, National Institute of Chemistry (Slovenia)
Y. Sannohe
Department of Chemistry, Graduate School of Sciences
Kyoto University (Japan)
N. Ma'ani Hessari
School of Biomedical Sciences, University of Ulster (UK)
Prof. H. Sugiyama
Department of Chemistry, Graduate School of Sciences
and Institute for Integrated Cell-Material Sciences
Kyoto University (Japan)

[**] This work was supported by the Royal Society (UK), the Slovenian Research Agency (P1-0242 and J1-0986), a Grant-in-Aid for Priority Research from the Ministry of Education, Culture, Sports, Science, and Technology, Japan, and the European Community (EU-NMR, contract no. RII3-026145). The authors also acknowledge support from the Global COE Program "Integrated Materials Science" (NoB-09) and COST MP0802.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200902454.

tetrads can also be described. For each of the remaining 24 theoretical folding possibilities the quadruplex stem is composed of two possible tetrad combinations. With knowledge of the order in which these two stacking possibilities appear in the quadruplex stem the topologies can be fully described. From a survey of high-resolution structures deposited in both the Protein Data Bank (PDB) and the Nucleic Acid Database (NDB) (January 2009) it is apparent that uni- and multimolecular quadruplex architectures composed of lateral and diagonal loops, or both, show alternation in the GBA of the guanosines in the quadruplex stem, thereby confirming earlier assertions derived from molecular mechanics calculations.[2] For two- and four-stacked tetrads in the quadruplex stem, the 5′-end of the quadruplex stem is *syn*G, and for three-stacked it is *anti*G. However, this is only valid for topologies composed of lateral loops, diagonal loops, or both.

In the PDB/NDB collections there are 11 structures of antiparallel quadruplexes containing propeller loops. In all of these the 5′-end guanosine of the quadruplex stem starts in a *syn* conformation. Moreover, in the three-stacked tetrads the first guanosine-rich segment has the sequence *syn*G-*anti*G-*anti*G. We aligned this segment for each of the 20 looping sequences that result in antiparallel quadruplexes with propeller loops (Figure 1). By stacking in this manner the two possible GBA tetrad combinations for each of the 20 looping combinations, the number of *syn*G units in the topology is minimized. Furthermore, in general, by following a propeller loop the guanine-rich segment is *syn*G-*anti*G-*anti*G, and by following a diagonal or lateral loop the guanine-rich segment is *syn*G-*syn*G-*anti*G. The exceptions coincide with the need to make allowances for a regular groove.

To provide evidence for these generalizations, we investigated the self-assembly of a representative topology. This topology is optimally illustrated both by controlling the assembly through loop length alone and choosing a novel topology that contains all three types of loops. The topology 10a shown in Figure 1 is described schematically in Figure 2 with its tetrad GBA combinations. In principle, with knowledge of one of the two tetrad GBA combinations in the quadruplex stem, it is possible to choose the length of loops. However, the tandem effects of loop length and nature of nucleosides in the sequence make it difficult to derive strict correlations between loop backbone lengths and the different types of loop defining the topology. We therefore utilized thymines as a suitable compromise for defining the loop length since they are less prone to stacking or to being involved in hydrogen-bond alignments with guanines. Thus we chose: 1) a single thymine for the anticlockwise propeller loop, 2) four thymines for the diagonal loop, and 3) three thymines for the clockwise lateral loop. The resultant oligonucleotide sequence is d[$G_3TG_3T_4G_3T_3G_3$]. Our choice of optimal loop lengths was based on their frequency in high-resolution structures deposited in the PDB and NDB as of January 2009.

We initially utilized enzymatic degradation and size-exclusion high-pressure liquid chromatography (HPLC) methods to establish that the oligonucleotide sequence

Figure 1. Schematic representations depicting the sequence of stacked tetrads for three-loop, antiparallel unimolecular quadruplexes containing propeller loops. Black rectangles depict *syn* and gray rectangles depict *anti* GBA for guanosines in the quadruplex stem. The blue circles represent the 5′ end which is the origin of the topologies. The numerals assigned to the depicted topologies follow reference [1]; those denoted as "a" start with an anticlockwise loop, and those denoted as "b" start with a clockwise loop.

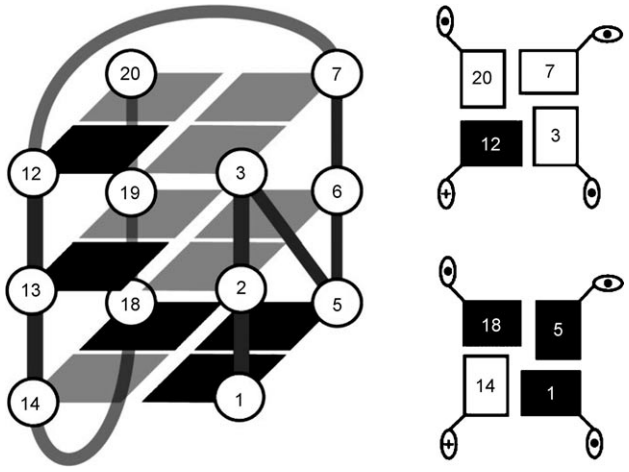


Figure 2. The folding topology and hydrogen-bond alignments adopted by the oligonucleotide sequence d[$G_3TG_3T_4G_3T_3G_3$] in KCl solutions. Black rectangles denote *syn* and gray rectangles represent *anti* GBA. Strand directionalities are indicated by "+" (away) and "●" (toward).

forms a unimolecular higher-order architecture in potassium solutions (see the Supporting Information).

The presence of a diagonal loop in the fold was initially established from the analysis of photoirradiated ^{I}U-substituted oligonucleotides. Photochemical 2′-deoxyribonolactone formation has been previously found, preferentially in the ^{I}U-containing diagonal loop of a G quadruplex.[3] Each of the eight thymine residues in the oligonucleotide sequence d[$G_3TG_3T_4G_3T_3G_3$] were substituted with ^{I}U to generate eight different oligonucleotides, ODNs 1–8 (Table 1).

Figure 3 shows the consumption of ODNs 1–8 after three minutes of irradiation by 302 nm UV light. ODNs 2–5, having ^{I}U in the diagonal loop, were highly photoreactive as compared with other oligonucleotides. A detailed product analysis of photoirradiated ODN 3, which was the most reactive ^{I}U-containing oligonucleotide, was carried out and appears in the Supporting Information. These results confirm the formation of 2′-deoxyribonolactone and indicate the presence of diagonal loop in the fold.

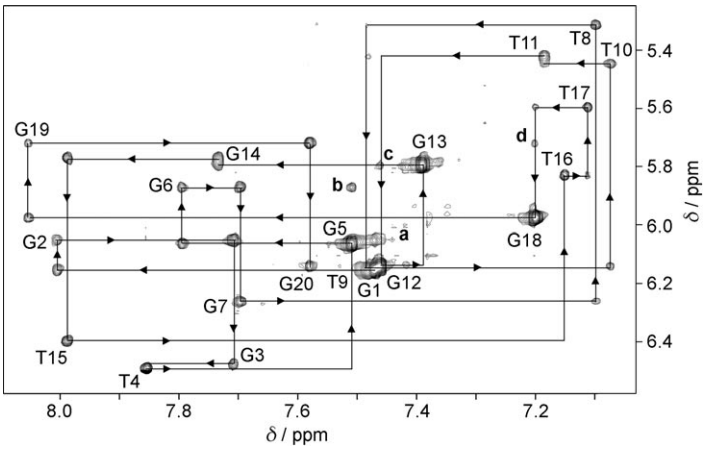


Figure 4. Expansions of NOESY spectrum (250 ms mixing time) of d[$G_3TG_3T_4G_3T_3G_3$] in $^{2}H_2O$ at 293 K. The lines trace the NOE connectivities between the base protons and their own and 5′-flanking sugar H1′ protons from G1 to G20 in the sequence. The regions identified as **a**, **b**, **c**, and **d** show the *syn*GH8(*i*)-*anti*GH1′(*i*+1) NOE contacts.

Table 1: The sequence of ^{I}U-substituted oligonucleotides.

Name	Sequence
ODN 1	5′-GGG^{I}UGGGTTTTGGGTTTGGG-3′
ODN 2	5′-GGGTGGG^{I}UTTTGGGTTTGGG-3′
ODN 3	5′-GGGTGGGT^{I}UTTGGGTTTGGG-3′
ODN 4	5′-GGGTGGGTT^{I}UTGGGTTTGGG-3′
ODN 5	5′-GGGTGGGTTT^{I}UGGGTTTGGG-3′
ODN 6	5′-GGGTGGGTTTTGGG^{I}UTTGGG-3′
ODN 7	5′-GGGTGGGTTTTGGGT^{I}UTGGG-3′
ODN 8	5′-GGGTGGGTTTTGGGTT^{I}UGGG-3′

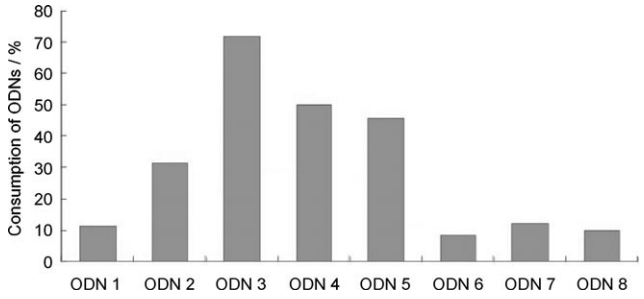


Figure 3. HPLC analysis of the degree to which each of the ODNs (1–8) were consumed when placed in a 2 mM sodium cacodylate buffer (pH 7.0) with 100 mM NaCl and irradiated with UV light for 3 min.

The sequence of GBA in the quadruplex stem, hydrogen-bond alignments of tetrads, their orientation and sequence within the quadruplex stem, as well as loop orientations, were established from solution NMR experiments (details can be found in the Supporting Information). In Figure 4, it is possible to trace the sequential NOE connectivities between the base and their own, and 5′-flanking sugar H1′ protons for G1-G4, G6-T8, T10-T11, G13-G20. Very strong H8–H1′ NOE intranucleotide connectivities and absent, or weak, N(*i*)–G(*i*+1) connectivities are observed for five guanosines (G1, G5, G12, G13, and G18). This data is an indication that their glycosidic torsion angles (χ) are in the *syn* conformation. Thus, the sequence of GBA in the quadruplex stem is proven to be as expected from design: (*syn-anti-anti*)-propeller-(*syn-*

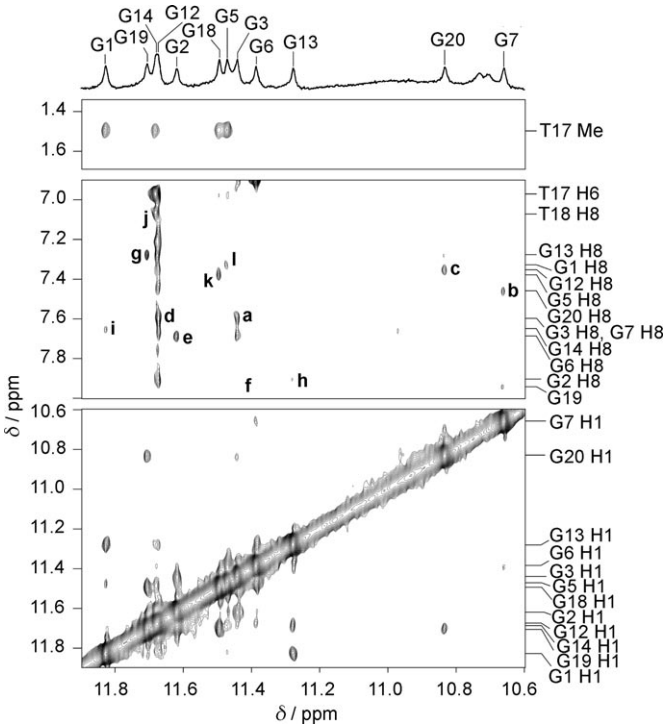


Figure 5. Imino–imino, imino–aromatic, and imino–methyl regions of NOESY spectrum (300 ms mixing time) at 278 K and pH 6.8 in $^{1}H_2O$. The sample concentration was 0.7 mM in strand. Hydrogen-bond directionalities within the three G tetrads are supported by the following NOE contacts: G3H1–G7H8 (**a**), G7H1–G20H8 (**b**), G20H1–G12H8 (**c**), G12H1–G3H8 (**d**), G2H1–G6H8 (**e**), G6H1–G19H8 (**f**), G19H1–G13H8 (**g**), G13H1–G2H8 (**h**), G1H1–G14H8 (**i**), G14H1–G18H8 (**j**), G18H1–G5H8 (**k**), and G5H1–G1H8 (**l**).

anti-anti)-diagonal-(*syn-syn-anti*)-lateral-(*syn-anti-anti*). The tetrad hydrogen-bond alignments were established from imino–imino and imino–aromatic NOESY cross-peaks as shown in Figure 5. They are consistent with the formation of a single G quadruplex containing three tetrads having the following directionalities: G3→G7→G20→G12, G2→G6→G19→G13 and G1→G14→G18→G5.

A number of inter-tetrad connectivities in the aromatic–aromatic, imino–aromatic, and imino–imino spectral regions define relative orientations of the three G tetrads with respect to each other. In addition, all amino protons of the guanines in the central (G2:G6:G19:G13) tetrad show cross-peaks to imino protons and to H8 inside the tetrad, and to the two outer tetrads. Loops in the architecture were also well-defined by NMR solution studies, and a description of NOE contacts is depicted in the low-resolution model shown in Figure 6.

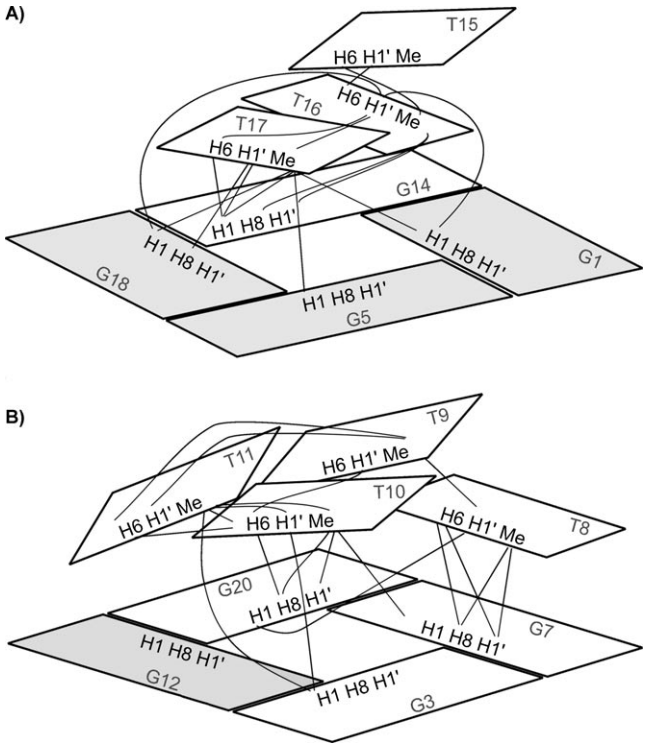


Figure 6. Schematic representation of low-resolution models for lateral loop consisting of residues T15–T17 (A) and for diagonal loop adopted by residues T8–T11 (B). Internucleotide NOE contacts between H1, H8/6, H1′, and Me protons are shown with solid lines. Panel (A) demonstrates spatial relations of T_3 loop with respect to the outer G1→G14→G18→G5 tetrad. For clarity, the following NOE contacts have been observed, but are not indicated in the figure: T17H1–G14H1 and T16H1–G1H1. In (B) stacking pattern of T_4-loop residues spanning the outer G3→G7→G20→G12 tetrad is depicted. Shaded residues are in *syn* conformation.

These studies indicate that we were successful in controlling the self-assembly of the designed topology. However, it does not prove the feasibility to control the self-assembly of all other topologies. We have therefore demonstrated that the GBA dispositions of guanosines of the tetrads in the quadruplex stem provide the basis for the design and control of folding of a quadruplex topology. We have shown that in principle it is feasible to control the folding through a choice of loops of appropriate length. However, not all topologies can be discriminated through their loop length alone. Several issues may impede successful design; for example, the nature of the residues in the loop and their interactions may influence the topology, or the self-assembly into multi-stranded architectures, or the nature and interactions of residues in either the 5′ and 3′ ends of the oligonucleotide chain, or both. More often than not, a sum of these factors in tandem determines the final topology. To minimize these effects it is currently feasible to control the fold of quadruplexes by forcing the GBA of selected guanosines of the quadruplex stem to adopt a desired conformation.[4–6]

Therefore, knowledge of the sequence of GBA that a quadruplex stem is able to adopt will, in principle, enable the rational design of quadruplex topologies. Moreover, understanding the role of loop length, loop composition, effect of metal ions on loops, and molecular recognition of these architectures is of general interest in realizing this goal.[1,7–9] To allow the general application of this formalism, two main premises have to be realized. Firstly, it is necessary to find out if, or which, of the remaining theoretical topologies are experimentally feasible. Secondly, the formalism should be extended to two-, four-, and possibly more-stacked tetrad systems. We are currently pursuing these studies.

Received: May 8, 2009
Revised: September 15, 2009
Published online: October 30, 2009

Keywords: glycosidic bonds · G quadruplexes · quadruplex folding · self-assembly · topology

[1] M. Webba da Silva, *Chem. Eur. J.* **2007**, *13*, 9738.
[2] D. Mohanty, M. Bansal, *Nucleic Acids Res.* **1993**, *21*, 1767.
[3] Y. Xu, H. Sugiyama, *Angew. Chem.* **2006**, *118*, 1380; *Angew. Chem. Int. Ed.* **2006**, *45*, 1354.
[4] A. Matsugami, Y. Xu, Y. Noguchi, H. Sugiyama, M. Katahira, *Febs J.* **2007**, *274*, 3545.
[5] Y. Xu, Y. Noguchi, H. Sugiyama, *Bioorg. Med. Chem.* **2006**, *14*, 5584.
[6] C. F. Tang, R. H. Shafer, *J. Am. Chem. Soc.* **2006**, *128*, 5966.
[7] A. N. Lane, J. B. Chaires, R. D. Gray, J. O. Trent, *Nucleic Acids Res.* **2008**, *36*, 5482.
[8] M. Cevec, J. Plavec, *Biochemistry* **2005**, *44*, 15238.
[9] P. Hazel, J. Huppert, S. Balasubramanian, S. Neidle, *J. Am. Chem. Soc.* **2004**, *126*, 16405.

