MINIREVIEW

Human telomeric G-quadruplex: structures of DNA and RNA sequences

Anh Tuân Phan

School of Physical & Mathematical Sciences, Nanyang Technological University, Singapore

Keywords
DNA; G-quadruplex; G-quadruplex structure; G-quadruplex topology; G-tetrad; G-tetrad core; grooves in G-quadruplexes; human telomere; loops in G-quadruplexes; RNA

Correspondence
A. T. Phan, Division of Physics & Applied Physics, School of Physical & Mathematical Sciences, Nanyang Technological University, Singapore 637371
Fax: +65 6795 7981
Tel: +65 6514 1915
E-mail: phantuan@ntu.edu.sg

(Received 25 June 2009, revised 14 September 2009, accepted 6 October 2009)

doi:10.1111/j.1742-4658.2009.07464.x

Telomeres play an important role in cellular aging and cancer. Human telomeric DNA and RNA G-rich sequences are capable of forming a four-stranded structure, known as the G-quadruplex. Such a structure might be important for telomere biology and a good target for drug design. This minireview describes the structural diversity or conservation of DNA and RNA human telomeric G-quadruplexes, discusses structural views on targeting these G-quadruplexes and presents some future challenges for structural studies.

Introduction

Human telomeric DNA contains thousands of tandem repeats of the G-rich $(GGGTTA)_n$ sequence [1], with a 3′-end overhang of 100–200 nucleotides [2]. Telomeres can be transcribed by DNA-dependent RNA polymerase II, and telomeric-repeat-containing RNA molecules, ranging from 100 to 9000 nucleotides, have been detected in nuclear fractions [3–5]. Under physiological ionic conditions, human telomeric DNA and RNA G-rich sequences are capable of forming a four-stranded helical structure, known as the G-quadruplex [6–14], based on the stacking of multiple G•G•G•G tetrads (or G-tetrads) [15] (Fig. 1A). Such a structure might be important for telomere biology [5–14,16–19] and a good target for drug design [5–14,19–21]. G-quadruplex structures formed by various G-rich sequences have been reviewed previously [6–14]. This minireview focuses on a simple topological and structural description of different DNA and RNA G-quadruplexes formed by short and long human telomeric sequences. Structural views on targeting these G-quadruplexes by small molecules are also discussed. Finally, the minireview highlights some future challenges for structural studies of human telomeric G-quadruplexes. The interactions of proteins with G-quadruplexes have been studied [16–19], but the structural knowledge on these interactions is still limited and is not covered here. Accompanying minireviews in this series [22–24] discuss the thermodynamic and kinetic properties of human telomeric G-quadruplexes and the current status of their targeting by small molecules.

Basics of G-quadruplex topologies

Oligonucleotides containing G-stretches can form monomeric, dimeric or tetrameric G-quadruplexes by folding/assembling one, two or four separate strands

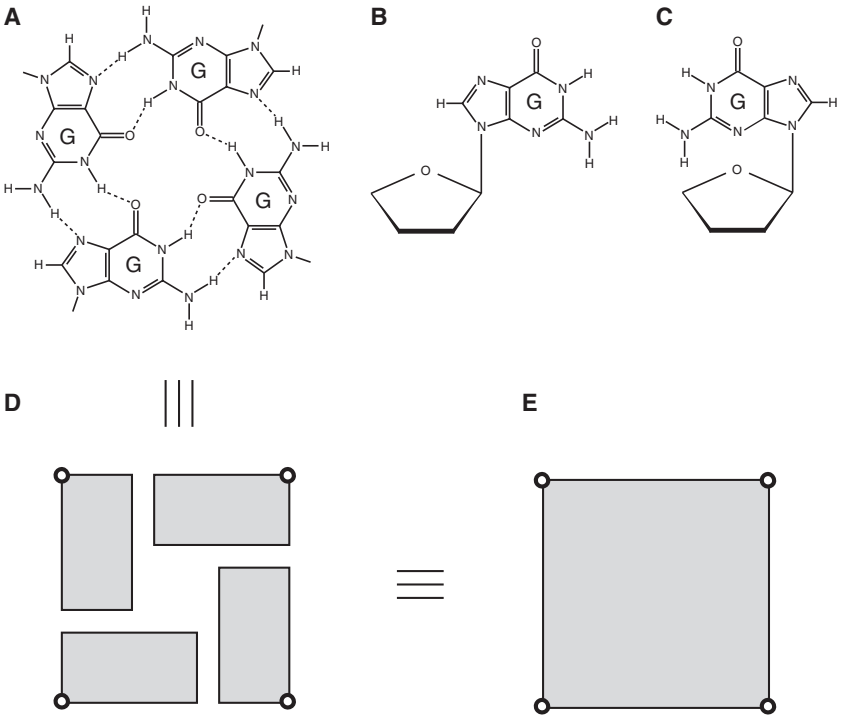

Fig. 1. (A) G-tetrad alignment. (B,C) Guanine in (B) *anti* and (C) *syn* glycosidic conformations. (D,E) Schematic presentation of a G-tetrad (D) used in Figs 3–6 with each guanine shown as a rectangular and (E) used in Fig. 2 with the G-tetrad shown as a square.

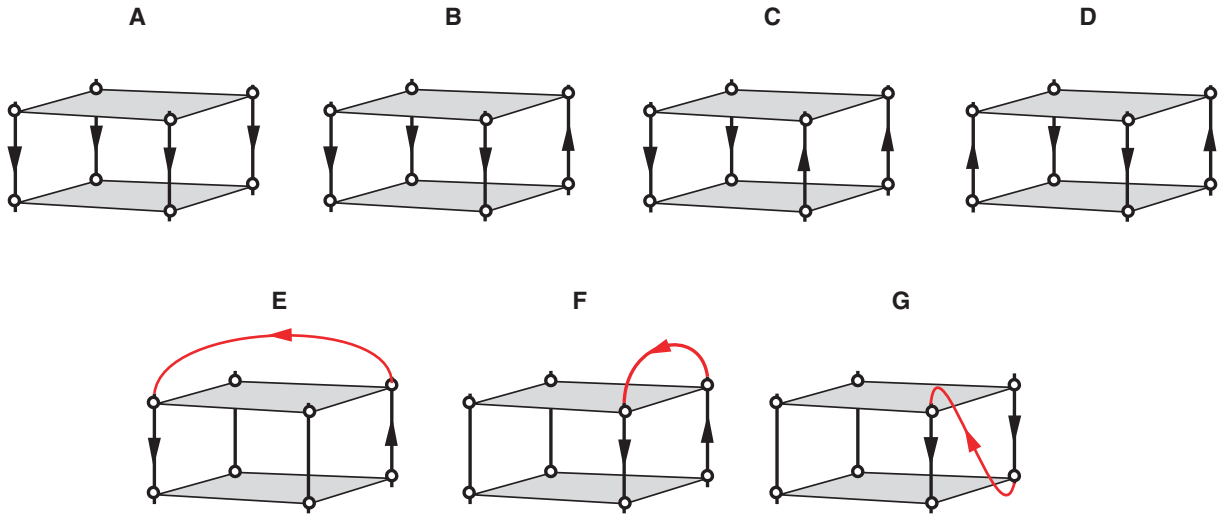

Fig. 2. (A–D) Four types of G-tetrad cores: (A) parallel G-tetrad core, (B) (3 + 1) G-tetrad core, (C) antiparallel G-tetrad core (up–up–down–down) and (D) antiparallel G-tetrad core (up–down–up–down). (E–G) Three types of loops (colored red): (E) diagonal loop, (F) edgewise loop and (G) double-chain-reversal loop. Arrows indicate the strand orientations, from 5′ to 3′ direction.

(see below). A G-quadruplex contains a G-tetrad core (Fig. 2A–D), formed by the stacking of several tetrads and supported by four backbone strands (or columns). Linkers connecting these strands are called loops (Fig. 2E–G). G-quadruplex structures are polymorphic regarding the G-tetrad core and loops (Fig. 2). Cations, such as K^+ and Na^+, stabilize G-quadruplexes by coordinating the carbonyl groups of guanines at the center of the G-tetrad core, and the preferred G-quadruplex structures adopted by a G-rich sequence depend on the nature of cations.

The G-tetrad core can be classified with regard to two mutually related factors, the relative orientations of the strands and the glycosidic conformations [*anti*

(Fig. 1B) or *syn* (Fig. 1C)] of guanines, which in turn define specific patterns of groove dimensions. There are four different possibilities for the relative strand orientations in the G-tetrad core (Fig. 2A–D): (a) four strands are oriented in the same direction (designated a parallel-stranded core) (Fig. 2A); (b) three strands are oriented in one direction and the fourth in the opposite direction [designated a (3 + 1) core, also called a hybrid core in the literature] (Fig. 2B); (c) two neighboring strands are oriented in one direction and the two remaining strands in the opposite direction (designated an up–up–down–down core, also called an antiparallel-stranded core in the literature) (Fig. 2C); and (d) two strands across one diagonal are oriented in one direction and the two remaining strands across the other diagonal in the opposite direction (designated an up–down–up–down core, also called an antiparallel-stranded core in the literature) (Fig. 2D). The glycosidic conformations of guanines within a G-tetrad are geometrically associated with the relative strand orientations, being respectively: (a) *anti•anti•anti•anti* or *syn•syn•syn•syn*, (b) *syn•anti•anti•anti* or *anti•syn•syn•syn*, (c) *syn•syn•anti•anti* and (d) *syn•anti•syn•anti*. The hydrogen-bond directionality of a G-tetrad in the core can be clockwise or anticlockwise, and this is directly related to the glycosidic conformations of guanines for each type of strand orientations (e.g. Fig. 3). The stacking patterns between adjacent G-tetrads of the same hydrogen-bond directionality differ from those between adjacent G-tetrads of opposite hydrogen-bond directionalities.

There are three major loop types: (a) diagonal loop connecting two opposing antiparallel strands across the diagonal (Fig. 2E); (b) edgewise loop (also called lateral loop) connecting two adjacent antiparallel strands (Fig. 2F); and (c) double-chain-reversal loop (also called propeller loop or side loop) connecting two adjacent parallel strands (Fig. 2G). The latter shares some features with another loop type called V-shaped loop [14].

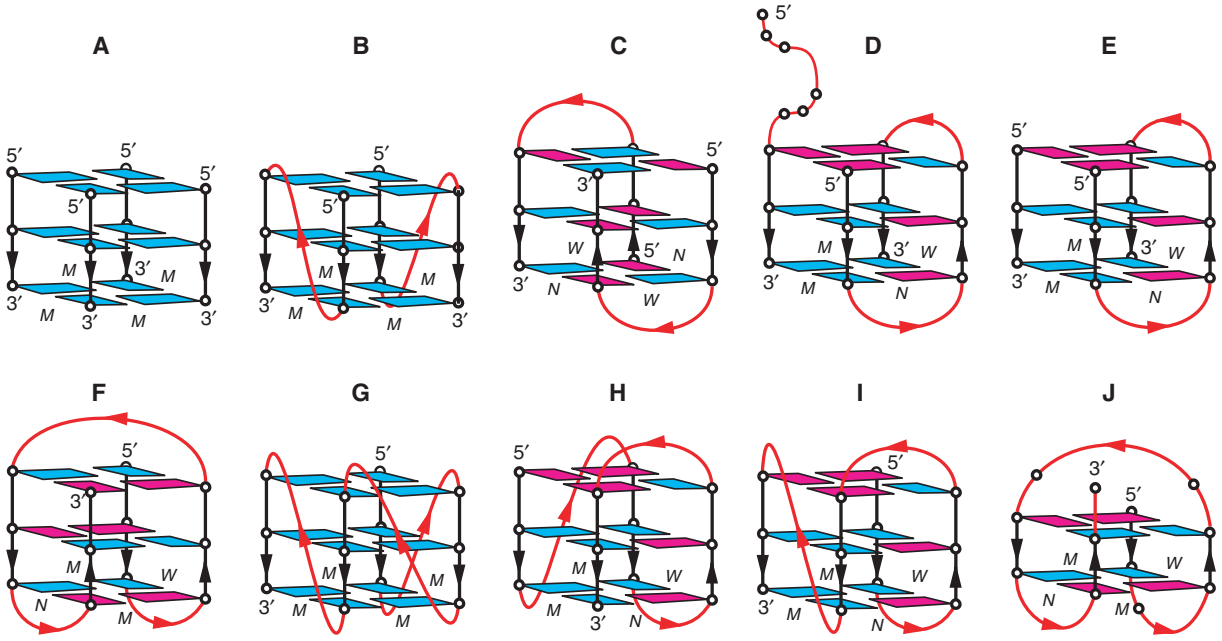

Fig. 3. Schematic structure of human telomeric G-quadruplexes. (A) Tetrameric parallel-stranded G-quadruplex observed for the single-repeat human telomeric sequences d(TTAGGG) and d(TTAGGGT) in K^+ solution [25]. (B) Dimeric parallel-stranded G-quadruplex observed for the two-repeat human telomeric sequence d(TAGGGTTAGGGT) in a K^+-containing crystal [27] and in K^+ solution [28]. (C) Dimeric antiparallel-stranded G-quadruplex observed for two-repeat human telomeric sequence d(TAGGGTTAGGGT) in K^+ solution [28]. (D) Asymmetric dimeric (3 + 1) G-quadruplex observed for the three-repeat human telomeric sequence d(GGGTTAGGGTTAGGGT) in Na^+ solution [30]. (E) Asymmetric dimeric (3 + 1) G-quadruplex association observed for the three-repeat human telomeric sequence d(GGGTTAGGGTTAGGGT) and the single-repeat human telomeric sequence d(TAGGGT) in Na^+ solution [30] and in K^+ solution (unpublished results). (F) Basket-type form observed for d[A$(GGGTTA)_3$GGG] in Na^+ solution [31]. (G) Propeller-type form observed for d[A$(GGGTTA)_3$GGG] in a K^+-containing crystal [27]. (H) (3 + 1) Form 1 observed for d[TA$(GGGTTA)_3$GGG] in K^+ solution [39–44,46]. (I) (3 + 1) Form 2 observed for d[TA$(GGGTTA)_3$GGGTT] in K^+ solution [41,45,46]. (J) Basket-type form observed for d[$(GGGTTA)_3$GGGT] in K^+ solution [47]. *anti* guanines are colored cyan; *syn* guanines are colored magenta; loops are colored red. *M*, *N* and *W* represent medium, narrow and wide grooves, respectively.

Short human telomeric DNA sequences

Short human telomeric sequences often serve as models for high-resolution structural studies of the telomere. Various G-quadruplex structures have been observed for human telomeric DNA sequences containing one, two, three or four repeats under different experimental conditions [25–54]. The number of G-tracts is often taken as the number of repeats, when the studied sequences do not contain exact multiples of the TTAGGG repeat. For example, the sequence d[AGGG(TTAGGG)$_3$] is usually considered as a four-repeat human telomeric sequence [31]. Multiple G-quadruplex conformations can be observed for a given sequence, making structural elucidation difficult [28,41]. This conformational heterogeneity can be overcome by judicious choices of the flanking nucleotides and/or base-analogue substitutions [28,39–47].

In K^+ solution, the single-repeat human telomeric sequences d(TTAGGG) and d(TTAGGGT) form a tetrameric parallel-stranded G-quadruplex containing three G-tetrad layers in which all guanines adopt the *anti* glycosidic conformation [25] (Fig. 3A), indicating that this structure is preferred in the absence of looping constraints. There are four medium-size grooves in such a structure. For the d(TTAGGG) sequence, high concentrations of K^+ and/or DNA favor the 3′-end stacking of two such G-quadruplex blocks into a structure containing six G-tetrad layers [25,26] (Fig. 4A).

In a K^+-containing crystal, the two-repeat human telomeric sequence d(TAGGGTTAGGGT) forms a dimeric parallel-stranded propeller-type G-quadruplex [27] (Fig. 3B). In this structure, all guanines are *anti*, the four grooves are of medium size and the two loops

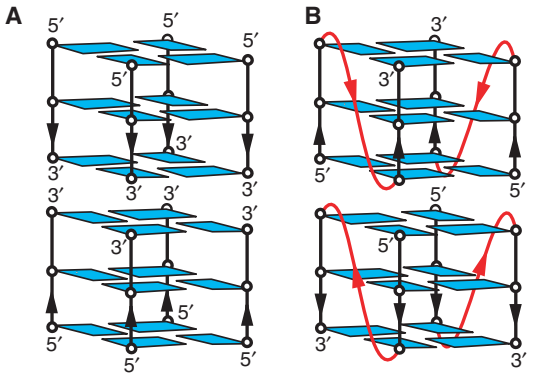

Fig. 4. Schematic structure of the stacking between two human telomeric G-quadruplex blocks, each involving three G-tetrads. (A) 3′-3′ stacking observed for the human telomeric DNA sequence d(TTAGGG) in K^+ solution [25,26] and (B) 5′-5′ stacking observed for human telomeric RNA sequence r(GGGUUAGGGU) in K^+ solution [66].

are double-chain-reversal (or propeller). In K^+ solution, the same sequence interconverts between parallel- and antiparallel-stranded dimeric G-quadruplexes [28] (Fig. 3B,C), whose folding and unfolding rates are distinct [28]. The parallel form is symmetric (Fig. 3B) and similar to the propeller-type structure observed in the crystalline state. The antiparallel form (up–down–up–down core) is asymmetric (Fig. 3C): glycosidic conformations of guanines along two consecutive G-tracts are 5′-*syn*-*anti*-*anti*-3′ and 5′-*syn*-*syn*-*anti*-3′ for one strand and 5′-*syn*-*syn*-*anti*-3′ and 5′-*syn*-*anti*-*anti*-3′ for the other strand of the dimer; glycosidic conformations of guanines around G-tetrads are *syn*•*anti*•*syn*•*anti*; there are two wide and two narrow grooves; the structure has two edgewise loops at the two ends of the G-tetrad core that span across the wide grooves. In Na^+ solution, CD spectra suggest that two-repeat human telomeric sequences adopt antiparallel-stranded G-quadruplex(es) [29].

The three-repeat human telomeric sequence d(GGGTTAGGGTTAGGGT) forms in Na^+ solution an asymmetric dimeric G-quadruplex, whose G-tetrad core involves all three G-tracts of one strand and only the 3′-end G-tract of the other strand [30] (Fig. 3D). The core of this structure, called the (3 + 1) core, has three strands oriented in one direction and one strand oriented in the opposite direction; glycosidic conformations of guanines along G-tracts are 5′-*syn*-*anti*-*anti*-3′ and 5′-*syn*-*syn*-*anti*-3′; there are two *syn*•*anti*•*anti*•*anti* G-tetrads and one *anti*•*syn*•*syn*•*syn* G-tetrad; there are one narrow, one wide and two medium grooves. The two edgewise grooves span the neighboring wide and narrow grooves, respectively. The three-repeat human telomeric sequence d(GGGTTAGGGTTAGGGT) can also associate with the single-repeat human telomeric sequence d(TAGGGT) in Na^+ solution [30] and K^+ solution (unpublished results) to form the same G-quadruplex topology (Fig. 3E).

Extensive research has been dedicated to the structures formed by sequences containing four human telomeric TTAGGG repeats, because this is considered the minimum length required for intramolecular G-quadruplex folding. Several G-quadruplex folding topologies have been proposed [27,31–51], with high-resolution structures reported for five intramolecular G-quadruplexes [27,31,42–47].

In Na^+ solution, the four-repeat human telomeric sequence d[AGGG(TTAGGG)$_3$] forms an antiparallel-stranded basket-type G-quadruplex [31] (Fig. 3F). The core (up–up–down–down type) of this structure consists of three *syn*•*syn*•*anti*•*anti* G-tetrads, which occur with 5′-*syn*-*anti*-*syn*-3′ or 5′-*anti*-*syn*-*anti*-3′ along the G-tracts; there are one narrow, one wide and two

medium grooves. Loops are consecutively edgewise–diagonal–edgewise.

In a K^+-containing crystal, the same sequence forms a propeller-type parallel-stranded G-quadruplex involving three G-tetrad layers [27] (Fig. 3G): all guanines are *anti*; loops are double-chain-reversal; four grooves are of medium size, three of which are occupied by loops.

In K^+ solution, multiple G-quadruplex conformations have been observed for four-repeat human telomeric sequences [28,29,32–54]. The d[TAGGG(TTAGGG)$_3$] and d[TAGGG(TTAGGG)$_3$TT] sequences form predominantly intramolecular (3 + 1) G-quadruplexes Form 1 [39–44,46] (Fig. 3H) and Form 2 [41,45,46] (Fig. 3I), respectively. These structures contain a three-G-tetrad (3 + 1) core with one double-chain-reversal and two edgewise loops, but they differ in the order of loop arraignments: in Form 1 the first linker adopts the double-chain-reversal loop configuration, whereas in Form 2 the third linker adopts the double-chain-reversal loop configuration. Form 1 and Form 2 are also the predominant conformations of the d[TTAGGG(TTAGGG)$_3$] [41,46] and d[TTAGGG(TTAGGG)$_3$TT] [41,45,46] sequences, respectively. The human telomeric sequence d[TAGGG(TTAGGG)$_3$T] adopts both Form 1 and Form 2 with comparable proportions in K^+ solution ([46] and unpublished results).

In K^+ solution, the human telomeric sequence d[GGG(TTAGGG)$_3$T] predominantly forms an intramolecular basket-type G-quadruplex involving only two G-tetrads, designated Form 3 [47] (Fig. 3J): the antiparallel-stranded core is of the up–up–down–down type; the two G-tetrads are *syn*•*syn*•*anti*•*anti* G-tetrads; there are one narrow, one wide and two medium grooves; loops are consecutively edgewise–diagonal–edgewise. Several other four-repeat human telomeric sequences, which start with a G (e.g. d[GGG(TTAGGG)$_3$], d[GGG(TTAGGG)$_3$TT] and d[GGG(TTAGGG)$_3$TTA]), also adopt Form 3 in K^+ solution [47]. Despite the presence of only two G-tetrad layers, Form 3 adopted by d[GGG(TTAGGG)$_3$T] is more stable than Form 1 and Form 2 adopted by d[TAGGG(TTAGGG)$_3$] and d[TAGGG(TTAGGG)$_3$TT] in K^+ solution, respectively. With extensive base pairing and stacking in the loops, Form 3 is made up totally of four to six base pair/triple/tetrad layers, which might explain the high stability of this structure. The folding principle of Form 3 indicates that the overall G-quadruplex topology of a G-rich sequence is defined not only by maximizing the number of G-tetrads, but also by maximizing all possible base pairing and stacking in the loops.

G-quadruplex structures determined for human telomeric DNA sequences show different configurations for TTA loops in various contexts involving three G-tetrads [27,31,42–47,55–60], as well as GTTA and GTTAG loops in Form 3 involving two G-tetrads [47]. Base pairing and stacking are generally observed in these loops [27,31,42–47,55–60]. It has been suggested that these loops are dynamic and may be good targets for specific ligand recognitions [46,47,55–59].

Different patterns of the G-tetrad hydrogen-bond directionalities are observed for the structures described above (Fig. 3). For example, the hydrogen-bond directionality alternates clockwise–anticlockwise for adjacent G-tetrads in the Na^+ solution basket-type G-quadruplex (Fig. 3F), whereas it remains the same for all G-tetrads in the parallel-stranded G-quadruplexes (Fig. 3A,B,G).

Other G-quadruplex folds have been proposed for DNA sequences containing human telomeric TTAGGG repeats under different experimental conditions [39,48–51]. It has been reported that molecular crowding conditions can favor parallel-stranded G-quadruplex conformation(s) [54].

Human telomeres, which encompass thousands of canonical TTAGGG repeats, can be interspersed with some sequence-variant repeats [61,62]. In particular, short contiguous arrays of variant CTAGGG repeats in the human telomere (variation is underlined) are unstable in the male germline and somatic cells [63]. In K^+ solution, DNA sequences containing four human telomeric variant CTAGGG repeats (e.g. d[AGGG(CTAGGG)$_3$]) form a new antiparallel intramolecular G-quadruplex involving two G-tetrads and a G•C•G•C tetrad (Fig. 5) [64].

Short human telomeric RNA sequences

The two-repeat human telomeric RNA sequence r(UAGGGUUAGGGU) forms, in both Na^+ solution [65] and K^+ solution [66], a propeller-type parallel-stranded dimeric G-quadruplex, the same folding topology observed for the DNA counterpart in a K^+-containing crystal (Fig. 3B). However, unlike the propeller-type DNA G-quadruplex [27] in which DNA residues prefer the C2′-*endo* sugar puckering conformation, the high-definition structure of the propeller-type RNA G-quadruplex in K^+ solution [66] shows both C2′-*endo* and C3′-*endo* conformations (residues in the loops adopt C2′-*endo* conformation; residues in the central G-tetrad adopt C3′-*endo* conformation; residues in the external G-tetrad can adopt both C2′-*endo* and C3′-*endo* conformations).

In K^+ solution, the human telomeric RNA sequence r(GGGUUAGGGU) forms a structure involving

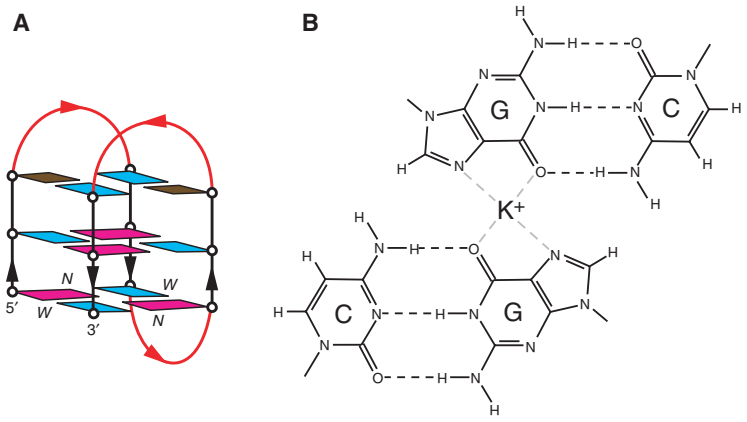

Fig. 5. Schematic structure of (A) the chair-type form G-quadruplex formed by variant human telomeric sequence d[A(GGGCTA)$_3$GGG] in K^+ solution, which contains two G-tetrads and (B) a G•C•G•C tetrad [64]. *anti* guanines are colored cyan; *syn* guanines are colored magenta; cytosines are colored brown; loops are colored red. *M*, *N* and *W* represent medium, narrow and wide grooves, respectively.

5′-end stacking of two propeller-type three-layer G-quadruplex blocks [66] (Fig. 4B). The lack of two residues UA at the 5′-end might favor this stacking structure [66].

Data suggest that the lack of U at the 3′-end of the human telomeric RNA sequence r(GGGUUAGGG) might favor further stacking of G-quadruplexes at this end to form a higher order structure [66]. CD spectra suggest that the four-repeat human telomeric RNA sequences also form parallel-stranded structure in Na^+ and K^+ solution [65,66]. The conservation of the G-quadruplex folding topology for human telomeric RNA sequences in Na^+ and K^+ solution [65,66] contrasts to the situation for human telomeric DNA counterparts in which multiple conformations are observed [25–54].

Long human telomeric DNA and RNA sequences

The next step toward understanding the structure of the 'real' telomeres is to address the question on the structure of long human telomeric DNA sequences [27,37,67,68]. The problem is the same for the long human telomeric RNA sequences [66,69]. Data [37,67,69] suggested that the structures of long human telomeric DNA and RNA sequences are based on multiple G-quadruplex blocks, each formed by a four-repeat

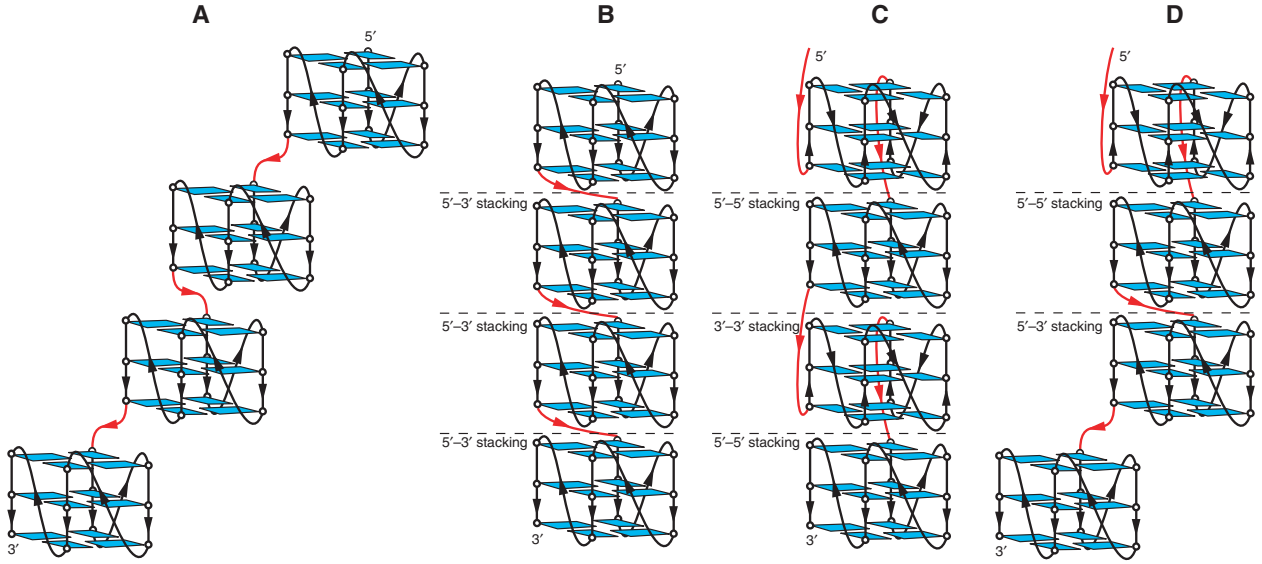

Fig. 6. Models for arrangements G-quadruplexes in long human telomeric DNA and RNA sequences. (A) 'Beads-on-a-string' [67], (B) 'same-direction stacking' [27], (C) 'alternate-direction stacking' [66] and (D) coexistence of all the three modes (A, B and C) for connection between G-quadruplex blocks. Linkers connecting consecutive G-quadruplex blocks are colored red.

segment (see above). Several models have been proposed regarding the arrangements of these G-quadruplex blocks [27,66,67]. In one model, G-quadruplex blocks are arranged like 'beads-on-a-string' [67], i.e. they can move relatively independently of each other and are constrained only by the connecting linkers (Fig. 6A).

Alternatively, G-quadruplex blocks can stack on each other to form a higher order structure. There may be three possible stacking modes between two parallel-stranded G-quadruplex blocks: (a) 5′-to-5′, in which the stacking interface is formed between the 5′-end of each block; (b) 3′-to-3′, in which the stacking interface is formed between the 3′-end of each block; and (c) 5′-to-3′, in which the stacking interface is formed between the 5′-end of one block and the 3′-end of the other. In the 'same-direction stacking' model proposed for long human telomeric DNA, successive propeller-type parallel-stranded G-quadruplex blocks, which are oriented in the same direction, stack 5′-to-3′ continuously (Fig. 6B) [27]. It has been suggested that a 200-nucleotide human telomeric DNA sequence, if folded into a stack of G-quadruplex, would form a rod of ~ 60 Å (compared with a 680 Å-long B-DNA helix) [27]. Successive (3 + 1) G-quadruplex blocks can also stack continuously according to this model [39,40,42]. In the 'alternate-direction stacking' model proposed for long human telomeric RNA (or DNA), the successive propeller-type G-quadruplex blocks stack according to 5′-to-5′ and 3′-to-3′ modes (Fig. 6C) [66]. In a model built for the 12-repeat human telomeric RNA r[UAGGG(UUAGGG)$_{11}$] sequence (Fig. 7), the linkers that connect two consecutive G-quadruplex blocks match well with the connecting distances, thereby resulting in these linkers being nicely packed in the grooves [66]. This type of linker arrangement can also connect G-quadruplex blocks of different folding topologies without generating knots. It is also possible that all these arrangements of G-quadruplexes coexist in the contexts of long telomeric DNA (or RNA). Figure 6D shows an example of the coexistence of three different connecting interfaces between consecutive G-quadruplex blocks.

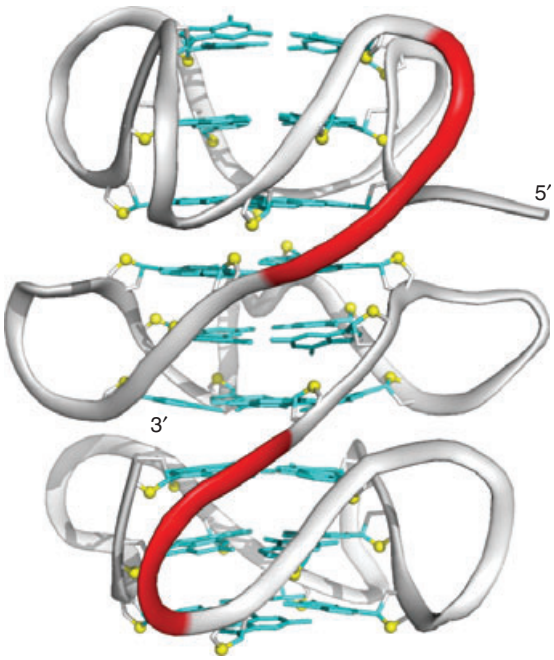

Fig. 7. A model for the high-order structure of the long human telomeric RNA sequence r[UAGGG(UUAGGG)$_{11}$]. Bases of guanines are colored cyan; O4′ of guanines yellow; UUA linkers connecting consecutive G-quadruplex blocks are colored red. Figure adapted from Martadinata & Phan [66].

A structural model for the eight-repeat human telomeric DNA sequence, built to satisfy various biophysical measurements, shows the stacking of two (3 + 1) G-quadruplex blocks (Form 1 [39–44] and Form 2) through bases in the loops [60].

Biochemical data on DNA sequences containing up to seven human telomeric repeats suggested that G-quadruplex preferentially forms at the 3′-end [70].

The dimeric (3 + 1) G-quadruplex assembly was proposed to be formed in the so-called T-loop [30], where the 3′-end overhang invades the preceding double-stranded part of the telomere [71]. This looping configuration of the telomere was illustrated in a stable lariat, in which the connection point was a (3 + 1) G-quadruplex [72].

Targeting human telomeric sequences by small molecules: structural views

The formation of G-quadruplexes by the telomeric G-rich DNA overhang has been shown to inhibit the activity of telomerase [73], an enzyme [74] required for the proliferation of most cancer cells [75]. Therefore, G-quadruplexes formed by human telomeric DNA are promising anticancer targets [20,21]. Human telomeric RNAs might also be potential drug targets based on their biological importance [5].

A desired ligand would recognize a G-quadruplex structure formed by human telomeric sequences with high affinity and specificity. Different G-quadruplex recognition modes are possible: (a) stacking on the ends of the G-tetrad core, (b) groove binding, (c) taking place of one or more strands in the core, (d) interacting with the backbone (core and loops), and (e) interacting with the loop bases. A ligand that uses several recognition modes may have an enhanced binding affinity and specificity.

Many of the reported G-quadruplex ligands [56–59,76–91] contain planar aromaric rings, which can interact with human telomeric G-quadruplex by stacking on the terminal G-tetrads [56–59,76,78–80,90,91]. To date, there is no conclusive evidence supporting the intercalation of a planar ligand between G-tetrad layers. In addition to the end-stacking binding mode of the aromatic rings, some ligands also contain other moieties that can recognize loops by stacking with loop bases or forming intermolecular hydrogen bonds [57–59,79] or recognize the backbone with electrostatic interactions [90,91]. The grooves in G-quadruplexes can also be recognized through hydrogen bonds [92] or hydrophobic interactions [93]. Alternatively, the G-rich human telomeric DNA (or RNA) strand can be trapped in a G-quadruplex structure with a linear guanine-containing molecule [30] based on a different backbone, such as PNA [94,95]. In the context of long human telomeric sequences, ligands can be designed to position between consecutive G-quadruplex blocks [96]. Finally, fluorescent ligands can be designed to probe the formation and the ligand-induced stabilization of telomeric G-quadruplexes in the cell [97,98].

Future challenges and prospects for structural studies

Despite a wealth of current knowledge about human telomeric G-quadruplexes, there remain many challenges associated with the structure and molecular recognition in the human telomeres. These include: (a) the structure and dynamics of all possible DNA, RNA and DNA/RNA hybrid G-quadruplexes formed by short and long human telomeric sequences; (b) the structural basis for molecular recognition of human telomeric G-quadruplexes by different small molecules and proteins; and (c) the detection of G-quadruplex structures and conformational transitions in the human telomeres in living cells.

Acknowledgements

This research was supported by Singapore Biomedical Research Council grant 07/1/22/19/542, Singapore Ministry of Education grants (ARC30/07 and RG62/07) and Nanyang Technological University start-up grants (SUG5/06 and RG138/06) to ATP.

References

1 Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, Meyne J, Ratliff RL & Wu JR (1988) A highly conserved repetitive DNA sequence, $(TTAGGG)_n$, present at the telomeres of human chromosomes. *Proc Natl Acad Sci USA* **85**, 6622–6626.
2 Makarov VL, Hirose Y & Langmore JP (1997) Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell* **88**, 657–666.
3 Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E & Lingner J (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**, 798–801.
4 Schoeftner S & Blasco MA (2008) Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol* **10**, 228–236.
5 Horard B & Gilson E (2008) Telomeric RNA enters the game. *Nat Cell Biol* **10**, 113–115.
6 Williamson JR (1994) G-quartet structures in telomeric DNA. *Annu Rev Biophys Biomol Struct* **23**, 703–730.
7 Gilbert DE & Feigon J (1999) Multistranded DNA structures. *Curr Opin Struct Biol* **9**, 305–314.
8 Simonsson T (2001) G-quadruplex DNA structures – variations on a theme. *Biol Chem* **382**, 621–628.
9 Neidle S & Parkinson GN (2003) The structure of telomeric DNA. *Curr Opin Struct Biol* **13**, 275–283.
10 Davis JT (2004) G-quartets 40 years later: from 5′-GMP to molecular biology and supramolecular chemistry. *Angew Chem Int Ed Engl* **43**, 668–698.
11 Phan AT, Kuryavyi V & Patel DJ (2006) DNA architecture: from G to Z. *Curr Opin Struct Biol* **16**, 288–298.
12 Burge S, Parkinson GN, Hazel P, Todd AK & Neidle S (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* **34**, 5402–5415.
13 Phan AT, Kuryavyi V, Luu KN & Patel DJ (2007) Structural diversity of G-quadruplex scaffolds. In *Quadruplex Nucleic Acids* (Neidle S & Balasubramanian S, eds), pp. 81–99. Royal Society of Chemistry, Cambridge.
14 Patel DJ, Phan AT & Kuryavyi V (2007) Human telomere, oncogenic promoter and 5′-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res* **35**, 7429–7455.
15 Gellert MN, Lipsett MN & Davies DR (1962) Helix formation by guanylic acid. *Proc Natl Acad Sci USA* **48**, 2013–2018.
16 Paeschke K, Simonsson T, Postberg J, Rhodes D & Lipps HJ (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures *in vivo*. *Nat Struct Mol Biol* **12**, 847–854.
17 Maizels N (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat Struct Mol Biol* **13**, 1055–1059.
18 Fry M (2007) Tetraplex DNA and its interacting proteins. *Front Biosci* **12**, 4336–4351.

19 Oganesian L & Bryan TM (2007) Physiological relevance of telomeric G-quadruplex formation: a potential drug target. *BioEssays* **29**, 155–165.
20 Sun D, Thompson B, Cathers BE, Salazar M, Kerwin SM, Trent JO, Jenkins TC, Neidle S & Hurley LH (1997) Inhibition of human telomerase by a G-quadruplex-interactive compound. *J Med Chem* **40**, 2113–2116.
21 Mergny JL & Hélène C (1998) G-quadruplex DNA: a target for drug design. *Nat Med* **4**, 1366–1367.
22 Chaires JB (2009) Human telomeric G-quadruplex: thermodynamic and kinetic studies of telomeric quadruplex stability. *FEBS J* **277**, 1098–1106.
23 Neidle S (2009) Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J* **277**, 1118–1125.
24 Arora A, Kumar N, Agarwal T & Maiti S (2009) Human telomeric G-quadruplex: targeting with small molecules. *FEBS J* **277**, 1345.
25 Wang Y & Patel DJ (1992) Guanine residues in $d(T_2AG_3)$ and $d(T_2G_4)$ form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. *Biochemistry* **31**, 8112–8119.
26 Kato Y, Ohyama T, Mita H & Yamamoto Y (2005) Dynamics and thermodynamics of dimerization of parallel G-quadruplexed DNA formed from $d(TTAG_n)$ (n = 3-5). *J Am Chem Soc* **127**, 9980–9981.
27 Parkinson GN, Lee MPH & Neidle S (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876–880.
28 Phan AT & Patel DJ (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J Am Chem Soc* **125**, 15021–15027.
29 Rujan IN, Meleney JC & Bolton PH (2005) Vertebrate telomere repeat DNAs favor external loop propeller quadruplex structures in the presence of high concentrations of potassium. *Nucleic Acids Res* **33**, 2022–2031.
30 Zhang N, Phan AT & Patel DJ (2005) (3 + 1) assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. *J Am Chem Soc* **127**, 17277–17285.
31 Wang Y & Patel DJ (1993) Solution structure of the human telomeric repeat $d[AG_3(T_2AG_3)_3]$ G-tetraplex. *Structure* **1**, 263–282.
32 Redon S, Bombard S, Elizondo-Riojas MA & Chottard JC (2003) Platinum cross-linking of adenines and guanines on the quadruplex structures of the $AG_3(T2AG_3)_3$ and $(T_2AG_3)_4$ human telomere sequences in Na^+ and K^+ solutions. *Nucleic Acids Res* **31**, 1605–1613.
33 He Y, Neumann RD & Panyutin IG (2004) Intramolecular quadruplex conformation of human telomeric DNA assessed with ^{125}I-radioprobing. *Nucleic Acids Res* **32**, 5359–5367.
34 Risitano A & Fox KR (2005) Inosine substitutions demonstrate that intramolecular DNA quadruplexes adopt different conformations in the presence of sodium and potassium. *Bioorg Med Chem Lett* **15**, 2047–2050.
35 Rezler EM, Seenisamy J, Bashyam S, Kim MY, White E, Wilson WD & Hurley LH (2005) Telomestatin and diseleno sapphyrin bind selectively to two different forms of the human telomeric G-quadruplex structure. *J Am Chem Soc* **127**, 9439–9447.
36 Qi J & Shafer RH (2005) Covalent ligation studies on the human telomere quadruplex. *Nucleic Acids Res* **33**, 3185–3192.
37 Vorlickova M, Chladkova J, Kejnovska I, Fialova M & Kypr J (2005) Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic Acids Res* **33**, 5851–5860.
38 Li J, Correia JJ, Wang L, Trent JO & Chaires JB (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Res* **33**, 4649–4659.
39 Xu Y, Noguchi Y & Sugiyama H (2006) The new models of the human telomere $d[AGGG(TTAGGG)_3]$ in K^+ solution. *Bioorg Med Chem* **14**, 5584–5591.
40 Ambrus A, Chen D, Dai J, Bialis T, Jones RA & Yang D (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res* **34**, 2723–2735.
41 Phan AT, Luu KN & Patel DJ (2006) Different loop arrangements of intramolecular human telomeric (3 + 1) G-quadruplexes in K^+ solution. *Nucleic Acids Res* **34**, 5715–5719.
42 Luu KN, Phan AT, Kuryavyi V, Lacroix L & Patel DJ (2006) Structure of the human telomere in K^+ solution: an intramolecular (3 + 1) G quadruplex scaffold. *J Am Chem Soc* **128**, 9963–9970.
43 Dai J, Punchihewa C, Ambrus A, Chen D, Jones RA & Yang D (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res* **35**, 2240–2250.
44 Matsugami A, Xu Y, Noguchi Y, Sugiyama H & Katahira M (2007) Structure of a human telomeric DNA sequence stabilized by 8-bromoguanosine substitutions, as determined by NMR in a K^+ solution. *FEBS J* **274**, 3545–3556.
45 Dai J, Carver M, Punchihewa C, Jones RA & Yang D (2007) 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* **35**, 4927–4940.
46 Phan AT, Kuryavyi V, Luu KN & Patel DJ (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K^+ solution. *Nucleic Acids Res* **35**, 6517–6525.
47 Lim KW, Amrane S, Bouaziz S, Xu W, Mu Y, Patel DJ, Luu KN & Phan AT (2009) Structure of the human telomere in K^+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J Am Chem Soc* **131**, 4301–4309.
48 Pedroso IM, Duarte LF, Yanez G, Burkewitz K & Fletcher TM (2007) Sequence specificity of inter- and intramolecular G-quadruplex formation by human telomeric DNA. *Biopolymers* **87**, 74–84.
49 Gaynutdinov TI, Neumann RD & Panyutin IG (2008) Structural polymorphism of intramolecular quadruplex of human telomeric DNA: effect of cations, quadruplex-binding drugs and flanking sequences. *Nucleic Acids Res* **36**, 4079–4087.
50 Okamoto K, Sannohe Y, Mashimo T, Sugiyama H & Terazima M (2008) G-quadruplex structures of human telomere DNA examined by single molecule FRET and ^{Br}G-substitution. *Bioorg Med Chem* **16**, 6873–6879.
51 Amrane S, Ang RW, Tan ZM, Li C, Lim JK, Lim JM, Lim KW & Phan AT (2009) A novel chair-type G-quadruplex formed by a *Bombyx mori* telomeric sequence. *Nucleic Acids Res* **37**, 931–938.
52 Ying L, Green JJ, Li H, Klenerman D & Balasubramanian S (2003) Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc Natl Acad Sci USA* **100**, 14629–14634.
53 Lee JY, Okumus B, Kim DS & Ha T (2005) Extreme conformational diversity in human telomeric DNA. *Proc Natl Acad Sci USA* **102**, 18938–18943.
54 Xue Y, Kan ZY, Wang Q, Yao Y, Liu J, Hao YH & Tan Z (2007) Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K^+ solution under molecular crowding condition. *J Am Chem Soc* **129**, 11185–11191.
55 Haider S, Parkinson GN & Neidle S (2008) Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophys J* **95**, 296–311.
56 Parkinson GN, Ghosh R & Neidle S (2007) Structural basis for binding of porphyrin to human telomeres. *Biochemistry* **46**, 2390–2397.
57 Campbell NH, Parkinson GN, Reszka AP & Neidle S (2008) Structural basis of DNA quadruplex recognition by an acridine drug. *J Am Chem Soc* **130**, 6722–6724.
58 Parkinson GN, Cuenca F & Neidle S (2008) Topology conservation and loop flexibility in quadruplex–drug recognition: crystal structures of inter- and intramolecular telomeric DNA quadruplex–drug complexes. *J Mol Biol* **381**, 1145–1156.
59 Campbell NH, Patel M, Tofa AB, Ghosh R, Parkinson GN & Neidle S (2009) Selectivity in ligand recognition of G-quadruplex loops. *Biochemistry* **48**, 1675–1680.
60 Petraccone L, Trent JO & Chaires JB (2008) The tail of the telomere. *J Am Chem Soc* **130**, 16530–16532.
61 Allshire RC, Dempster M & Hastie ND (1989) Human telomeres contain at least 3 types of G-rich repeat distributed non-randomly. *Nucleic Acids Res* **17**, 4611–4627.
62 Baird DM, Jeffreys AJ & Royle NJ (1995) Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere. *EMBO J* **14**, 5433–5443.
63 Mendez-Bermudez A, Hills M, Pickett HA, Phan AT, Mergny JL, Riou JF & Royle NJ (2009) Human telomeres that contain $(CTAGGG)_n$ repeats show replication dependent instability in somatic cells and the male germline. *Nucleic Acids Res* **37**, 6225–6238.
64 Lim KW, Alberti P, Guedin A, Lacroix L, Riou JF, Royle NJ, Mergny JL & Phan AT (2009) Sequence variant $(CTAGGG)_n$ in the human telomere favors a G-quadruplex structure containing a G•C•G•C tetrad. *Nucleic Acids Res* **37**, 6239–6248.
65 Xu Y, Kaminaga K & Komiyama M (2008) G-quadruplex formation by human telomeric repeats containing RNA in Na^+ solution. *J Am Chem Soc* **130**, 11179–11184.
66 Martadinata H & Phan AT (2009) Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K^+ solution. *J Am Chem Soc* **131**, 2570–2578.
67 Yu HQ, Miyoshi D & Sugimoto N (2006) Characterization of structure and stability of long telomeric DNA G-quadruplexes. *J Am Chem Soc* **128**, 15461–15468.
68 Pomerantz AK, Moerner WE & Kool ET (2008) Visualization of long human telomere mimics by single-molecule fluorescence imaging. *J Phys Chem B* **112**, 13184–13187.
69 Randall A & Griffith JD (2009) Structure of long telomeric RNA transcripts: the G-rich RNA forms a compact repeating structure containing G-quartets. *J Biol Chem* **284**, 13980–13986.
70 Tang J, Kan ZY, Yao Y, Wang Q, Hao YH & Tan Z (2008) G-quadruplex preferentially forms at the very 3′-end of vertebrate telomeric DNA. *Nucleic Acids Res* **36**, 1200–1208.
71 Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H & de Lange T (1999) Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514.
72 Xu Y, Sato H, Sannohe Y, Shinohara KI & Sugiyama H (2008) Stable lariat formation based on a G-quadruplex scaffold. *J Am Chem Soc* **130**, 16470–16471.
73 Zahler AM, Williamson JR, Cech TR & Prescott DM (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350**, 718–720.

74 Greider CW & Blackburn EH (1985) Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**, 405–413.
75 Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PL, Coviello GM, Wright WE, Weinrich SL & Shay JW (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011–2015.
76 Searle MS & Balkwill GD (2006) DNA quadruplex–ligand recognition: structure and dynamics. In *Quadruplex Nucleic Acids* (Neidle S & Balasubramanian S, eds), pp. 131–153. Royal Society of Chemistry, Cambridge.
77 De Cian A, Lacroix L, Douarre C, Temime-Smaali N, Trentesaux C, Riou JF & Mergny JL (2008) Targeting telomeres and telomerase. *Biochimie* **90**, 131–155.
78 Clark GR, Pytel PD, Squire CJ & Neidle S (2003) Structure of the first parallel DNA quadruplex–drug complex. *J Am Chem Soc* **125**, 4066–4067.
79 Haider SM, Parkinson GN & Neidle S (2003) Structure of a G-quadruplex–ligand complex. *J Mol Biol* **326**, 117–125.
80 Cocco MJ, Hanakahi LA, Huber MD & Maizels N (2003) Specific interactions of distamycin with G-quadruplex DNA. *Nucleic Acids Res* **31**, 2944–2951.
81 De Cian A, Delemos E, Mergny JL, Teulade-Fichou MP & Monchaud D (2007) Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J Am Chem Soc* **129**, 1856–1857.
82 Barbieri CM, Srinivasan AR, Rzuczek SG, Rice JE, LaVoie EJ & Pilch DS (2007) Defining the mode, energetics and specificity with which a macrocyclic hexaoxazole binds to human telomeric G-quadruplex DNA. *Nucleic Acids Res* **35**, 3272–3286.
83 Bejugam M, Sewitz S, Shirude PS, Rodriguez R, Shahid R & Balasubramanian S (2007) Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: small molecule regulation of c-*kit* oncogene expression. *J Am Chem Soc* **129**, 12926–12927.
84 Shirude PS, Gillies ER, Ladame S, Godde F, Shin-Ya K, Huc I & Balasubramanian S (2007) Macrocyclic and helical oligoamides as a new class of G-quadruplex ligands. *J Am Chem Soc* **129**, 11890–11891.
85 Cetinkol OP, Engelhart AE, Nanjunda RK, Wilson WD & Hud NV (2008) Submicromolar, selective G-quadruplex ligands from one pot: thermodynamic and structural studies of human telomeric DNA binding by azacyanines. *Chembiochem* **9**, 1889–1892.
86 Kieltyka R, Englebienne P, Fakhoury J, Autexier C, Moitessier N & Sleiman HF (2008) A platinum supramolecular square as an effective G-quadruplex binder and telomerase inhibitor. *J Am Chem Soc* **130**, 10040–10041.
87 Kieltyka R, Fakhoury J, Moitessier N & Sleiman HF (2008) Platinum phenanthroimidazole complexes as G-quadruplex DNA selective binders. *Chemistry* **14**, 1145–1154.
88 Tera M, Ishizuka H, Takagi M, Suganuma M, Shin-ya K & Nagasawa K (2008) Macrocyclic hexaoxazoles as sequence- and mode-selective G-quadruplex binders. *Angew Chem Int Ed Engl* **47**, 5557–5560.
89 Tera M, Iida K, Ishizuka H, Takagi M, Suganuma M, Doi T, Shin-ya K & Nagasawa K (2009) Synthesis of a potent G-quadruplex-binding macrocyclic heptaoxazole. *Chembiochem* **10**, 431–435.
90 Phan AT, Kuryavyi V, Gaw HY & Patel DJ (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat Chem Biol* **1**, 167–173.
91 Dixon IM, Lopez F, Tejera AM, Estève JP, Blasco MA, Pratviel G & Meunier B (2007) A G-quadruplex ligand with 10000-fold selectivity over duplex DNA. *J Am Chem Soc* **129**, 1502–1503.
92 Kettani A, Gorin A, Majumdar A, Hermann T, Skripkin E, Zhao H, Jones R & Patel DJ (2000) A dimeric DNA interface stabilized by stacked A•(G•G•G•G)•A hexads and coordinated monovalent cations. *J Mol Biol* **297**, 627–644.
93 Martino L, Virno A, Pagano B, Virgilio A, Di Micco S, Galeone A, Giancola C, Bifulco G, Mayol L & Randazzo A (2007) Structural and thermodynamic studies of the interaction of distamycin A with the parallel quadruplex structure $[d(TGGGGT)]_4$. *J Am Chem Soc* **129**, 16048–16056.
94 Datta B, Schmitt C & Armitage BA (2003) Formation of a PNA2–DNA2 hybrid quadruplex. *J Am Chem Soc* **125**, 4111–4118.
95 Paul A, Sengupta P, Krishnan Y & Ladame S (2008) Combining G-quadruplex targeting motifs on a single peptide nucleic acid scaffold: a hybrid (3 + 1) PNA–DNA bimolecular quadruplex. *Chemistry* **14**, 8682–8689.
96 Xu Y, Yamazaki S, Osuga H & Sugiyama H (2006) The recognition of higher-order G-quadruplex by chiral cyclic-helicene molecules. *Nucleic Acids Symp Ser* **50**, 183–184.
97 Chang CC, Kuo IC, Ling IF, Chen CT, Chen HC, Lou PJ, Lin JJ & Chang TC (2004) Detection of quadruplex DNA structures in human telomeres by a fluorescent carbazole derivative. *Anal Chem* **76**, 4490–4494.
98 Yang P, De Cian A, Teulade-Fichou MP, Mergny JL & Monchaud D (2009) Engineering bisquinolinium/thiazole orange conjugates for fluorescent sensing of G-quadruplex DNA. *Angew Chem Int Ed Engl* **48**, 2188–2191.