

SURVEY AND SUMMARY

Quadruplex DNA: sequence, topology and structure

Sarah Burge, Gary N. Parkinson, Pascale Hazel, Alan K. Todd and Stephen Neidle*

Cancer Research UK Biomolecular Structure Group, The School of Pharmacy, University of London,
29-39 Brunswick Square, London WC1N 1AX, UK

Received July 12, 2006; Revised August 25, 2006; Accepted August 27, 2006

ABSTRACT

G-quadruplexes are higher-order DNA and RNA structures formed from G-rich sequences that are built around tetrads of hydrogen-bonded guanine bases. Potential quadruplex sequences have been identified in G-rich eukaryotic telomeres, and more recently in non-telomeric genomic DNA, e.g. in nuclease-hypersensitive promoter regions. The natural role and biological validation of these structures is starting to be explored, and there is particular interest in them as targets for therapeutic intervention. This survey focuses on the folding and structural features on quadruplexes formed from telomeric and non-telomeric DNA sequences, and examines fundamental aspects of topology and the emerging relationships with sequence. Emphasis is placed on information from the high-resolution methods of X-ray crystallography and NMR, and their scope and current limitations are discussed. Such information, together with biological insights, will be important for the discovery of drugs targeting quadruplexes from particular genes.

INTRODUCTION

The knowledge that guanine-rich nucleic acids can self-associate has a long history, pre-dating the double helix itself by almost 50 years. For much of that time, the gels formed by such sequences were more of nuisance value than scientific worth. The molecular basis for the association was subsequently determined by fibre diffraction (1–3) and biophysical (4) studies using the concept (5,6) that the Hoogsteen hydrogen-bonded guanine (G)-tetrad (also termed a G-quartet) is the basic structural motif (Figure 1a). The synthetic polynucleotides poly(dG) and poly(G) were determined in these studies to form four-stranded helical structures (Figure 1b) with the G-tetrads stacked on one another, analogous to Watson–Crick base pairs in duplex DNA. These structures remained largely laboratory curiosities until it was found

that short G-rich sequences at the ends of telomeric DNA in eukaryotic chromosomes can associate together in physiological ionic conditions to form discrete four-stranded structures (variously termed quadruplexes, tetraplexes or G4 structures) that incorporate the fundamental structural feature of having at least two contiguous G-tetrads stacked one on another (7,8).

Formation of these quadruplex structures at telomere ends is possible since the terminal nucleotides at the 3' ends of all telomeric DNAs are single-stranded (9,10), albeit in association with single-strand-binding proteins, such as hPOT1 in *Homo sapiens* (11,12), where the single-strand overhang is ca. 100–200 nt long. Telomeric DNA sequences (13) comprise G-rich tandem repeats (Table 1), i.e. are not pure G sequences, and have short non-G tracts regularly interspersing the G ones. A few prokaryotic species, such as *Streptomyces* also have linear chromosomes, with repetitive DNA at the ends, but with distinct sequences that can form inverted repeat structures (14). A second category of quadruplexes involve oligonucleotide aptamers comprising quadruplex-forming sequences, which have the ability to selectively act as inhibitors of signal transduction or transcription via binding to particular targets, such as Stat3 (15) or nucleolin (16) in cancer cells. Few 3D structures of quadruplexes formed from aptamer sequences have been fully characterized; that of the thrombin-binding sequence d(GGTTGGT-GTGGTTGG) is a notable exception (17). The third category comprises potential quadruplexes that may be formed from appropriate G-rich sequences that are present within a wide range of genes (and very extensively in non-coding regions of many genomes). Now that extensive sequence data are available on a large number of eukaryotic and prokaryotic genomes, it is apparent that such sequences are highly prevalent (18–21), and an increasing number of quadruplexes arising from them have been reported. This survey will focus on some of the underlying principles and emerging issues concerning (i) sequence (primary structure), (ii) the diverse patterns of folding, i.e. quadruplex topology (secondary structure) and (iii) more detailed structural information (tertiary structure) on both telomeric and non-telomeric quadruplexes, especially those from the high-resolution methods of crystallography, molecular simulation and NMR.

*To whom correspondence should be addressed. Tel: +44 207 753 5969; Fax: +44 207 753 5970; Email: stephen.neidle@pharmacy.ac.uk

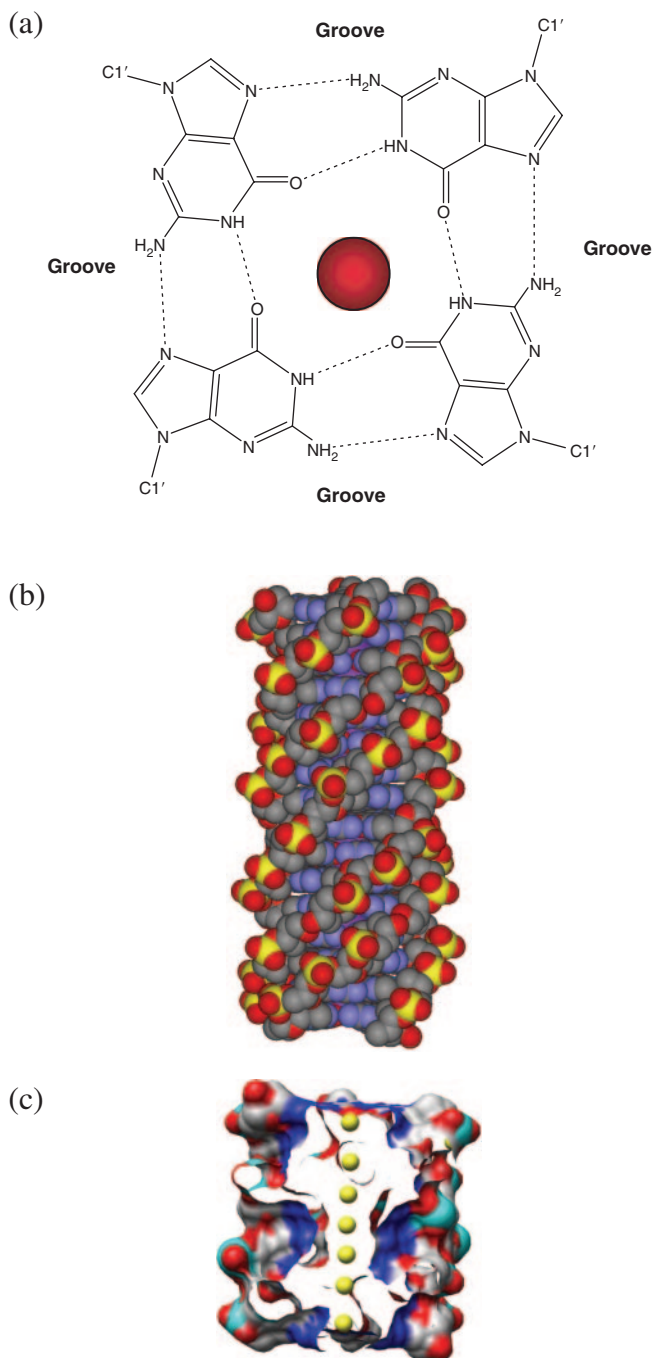


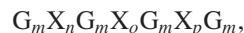
Figure 1. (a) The arrangement of guanine bases in the G-quartet, shown together with a centrally placed metal ion. Hydrogen bonds are shown as dotted lines, and the positions of the grooves are indicated. (b) The poly(dG) four-fold, right-handed helix. (c) Surface view representation of a quadruplex structure comprising eight G-quartets, with the central channel exposed to show an array of metal ions (coloured yellow).

GENERAL FEATURES OF QUADRUPLEX TOPOLOGY AND STRUCTURE

Quadruplexes can be formed from one, two or four separate strands of DNA (or RNA) and can display a wide variety of topologies, which are in part a consequence of various possible combinations of strand direction, as well as variations in

loop size and sequence. They can be defined in general terms as structures formed by a core of at least two stacked G-tetrads, which are held together by loops arising from the intervening mixed-sequence nucleotides that are not usually involved in the tetrads themselves. The combination of the number of stacked G-tetrads, the polarity of the strands and the location and length of the loops would be expected to lead to a plurality of G-quadruplex structures, as indeed is found experimentally.

Potential **unimolecular** (i.e. intramolecular) G-quadruplex-forming sequences can be described as follows:



where m is the number of G residues in each short G-tract, which are usually directly involved in G-tetrad interactions. X_n , X_o and X_p can be any combination of residues, including G, forming the loops. This notation also implies that the G-tracts can be of unequal length, and if one of the short G tracts is longer than the others, some of the G residues will be located in the loop regions. The assumption that all G tracts within a quadruplex sequence are identical is true for vertebrate telomeric sequences, but is not always the case for non-telomeric genomic sequences, or even for all telomeric sequences in some lower eukaryotics (see Table 1). In principle **bimolecular** (dimeric) and **tetramolecular** (tetrameric) quadruplexes can each be formed from the association of non-equal sequences, although very few quadruplexes with such features have yet been studied in detail. Thus, almost all bimolecular quadruplexes reported to date are formed by the association of two identical sequences $X_n G_m X_o G_m X_p$, where n and p may or may not be zero. Tetramolecular quadruplexes may be formed by four $X_n G_m X_o$ or $G_m X_n G_m$ strands associating together.

Quadruplex structures may be classified according to their strand polarities and the location of the loops that link the guanine strand(s) for quadruplexes formed either from a single-strand or from two strands. Adjacent linked parallel strands require a connecting loop to link the bottom G-tetrad with the top G-tetrad, leading to **propeller** type loops (these are sometimes termed strand-reversal loops but we prefer the simpler term since this describes the appearance of this loop and does not introduce any potential confusion about strand direction). This feature has been found both in crystal structures (22) and in solution (23) for quadruplexes formed from human telomeric DNA sequences (see below), and more recently in a number of non-telomeric quadruplexes. Quadruplexes are designated as anti-parallel when at least one of the four strands is anti-parallel to the others. This type of topology is found in the majority of bimolecular and in many unimolecular quadruplex structures determined to date. Two further types of loops have been observed in these structures, in addition to parallel loops. **Lateral** (sometimes termed edge-wise) loops join adjacent G-strands, as observed in the structures of both two asymmetric quadruplexes observed in solution by NMR for the d(TG₄T₂G₄T) sequence (24) and in the bimolecular quadruplex structure formed by the sequence d(GGGCT₄GGGC) (25). Two of these loops can be located either on the same or opposite faces of a quadruplex, corresponding to head-to-head or head-to-tail, respectively when in bimolecular quadruplexes (Figures 2a and 4). Strand

Table 1. Some known telomeric DNA sequences

| Group | Organism | Telomeric repeat |
|------------------------|---|--|
| Vertebrates | Human, mouse, <i>Xenopus</i> | TTAGGG |
| Filamentous fungi | <i>Neurospora crassa</i> | TTAGGG |
| Slime moulds | <i>Physarum</i> , <i>Didymium</i> , <i>Dictyostelium</i> | TTAGGG AG(1–8) |
| Kinetoplastid protozoa | <i>Trypanosoma</i> , <i>Crithidia</i> | TTAGGG |
| Ciliate protozoa | <i>Tetrahymena</i> , <i>Glaucoma</i> <i>Paramecium Oxytricha</i> <i>Stylonychia</i> , <i>Euplotes</i> | TTGGGG TTGGG(T/G) TTTTGGGG |
| Apicomplexan protozoa | <i>Plasmodium</i> | TTAGGG(T/C) |
| Higher plants | <i>Arabidopsis thaliana</i> | TTTAGGG |
| Green algae | <i>Chlamydomonas</i> | TTTTAGGG |
| Insects | <i>Bombyx mori</i> | TTAGG |
| Roundworms | <i>Ascaris lumbricoides</i> | TTAGGC |
| Fission yeasts | <i>Schizosaccharomyces pombe</i> | TTAC(A)(C)G(1–8) |
| Budding yeasts | <i>Saccharomyces cerevisiae</i> | TGTGGGTGTGGTG (from RNA template) or G(2,3)(TG)(1–6)T (consensus) |
| | <i>Candida glabrata</i> | GGGGTCTGGGTGCTG |
| | <i>Candida albicans</i> | GGGTACGGATGTCTAACTTCTT |
| | <i>Candida tropicalis</i> | GGTGTA[C/A]GGATGTCACGATCATT |
| | <i>Candida maltosa</i> | GGGTACGGATGCAGACTCGCTT |
| | <i>Candida guilliermondii</i> | GGTGTAC |
| | <i>Candida pseudotropicalis</i> | GGGTACGGATTTGATTAGTTATGT |
| | <i>Kluyveromyces lactis</i> | GGGTACGGATTTGATTAGGTATGT |

polarities can vary, as in the example of the two distinct bimolecular quadruplexes formed by $d(G_4T_3G_4)$, with one being a head-to-tail lateral loop dimer in which all adjacent strands are anti-parallel, and the other is a head-to-head hairpin quadruplex with one adjacent strand parallel and the other is anti-parallel (26). The second type of anti-parallel loop, the **diagonal** loop joins opposite G-strands, as observed in the structure formed by the *Oxytricha nova* telomeric sequence $d(G_4T_4G_4)$ (27–31). In this instance the directionalities of adjacent strands must alternate between parallel and anti-parallel, and are arranged around a core of four stacked G-tetrads.

All parallel quadruplexes have all guanine glycosidic angles in an *anti* conformation. Anti-parallel quadruplexes have both *syn* and *anti* guanines, arranged in a way that is particular for a given topology and set of strand orientations, since different topologies have the four strands in differing positions relative to each other. All quadruplex structures have four grooves, defined as the cavities bounded by the phosphodiester backbones. Groove dimensions are variable, and depend on overall topology and the nature of the loops. Grooves in quadruplexes with only lateral or diagonal loops are structurally simple, and the walls of these grooves are bounded by monotonic sugar phosphodiester groups. In contrast, grooves that incorporate propeller loops have more complex structural features that reflect the insertion of the variable-sequence loops into the grooves (see Figure 5).

The formation and stability of G-quadruplexes is monovalent cation-dependent. This has been ascribed to the strong negative electrostatic potential created by the guanine O6 oxygen atoms, which form a central channel of the G-tetrad stack (4,32–34), with the cations located within this channel (Figure 1c). The precise location of the cations between the tetrads is dependent on the nature of the ion, with Na^+ ions within the channel being observed in a range of geometries; in some structures, a Na^+ ion is in plane with a G-tetrad whereas in others it is between two successive G-tetrads.

K^+ ions are always equidistant between each tetrad plane, and form the eight oxygen atoms in a symmetric tetragonal bipyramidal configuration. Other ions can substitute for these two. Thallium (1^+), with an ionic radius close to that of the K^+ ion, can substitute for it. The NMR structure (27) of the Tl^+ -containing bimolecular quadruplex formed from the *O.nova* sequence $d(G_4T_4G_4)$, shows identical quadruplex topology to that in the K^+ ion form found in the crystalline state (28), which is itself identical with the NMR structures of the well-characterized Na^+ form (29–31). On the other hand, there are a number of well-established examples where the change from Na^+ to K^+ induces profound structural alteration, implying high conformational flexibility for these particular quadruplexes. It is equally clear that some quadruplexes, such as the bimolecular $d(G_4T_4G_4)$ quadruplex (28–31) and the parallel-stranded structure formed by four $d(TGGGGT)$ molecules (35,36), have very stable and unique topologies. A series of very long time-scale molecular dynamics simulations (0.5–1 μ s) have shown that these structures retain their integrity not only in simulated solution but also in the gas phase, provided the cations are present (37).

Methods for quadruplex topology and structure determination

A number of quadruplex studies have employed the methods of biophysical chemistry, notably circular dichroism (CD), to assign topology. The main attraction of CD spectroscopy is its potential to discriminate between quadruplex topologies having differences in parallel and anti-parallel strand orientation, arising from different arrangements of *anti/syn* glycosidic angles. CD therefore can be a useful and rapid method for establishing an overall fold. It requires very little sample (μ M concentrations are sufficient) and is suited to examining a wide range of solution conditions and their influence on quadruplex formation. The method is, however, more sensitive than ultraviolet (UV) melting experiments

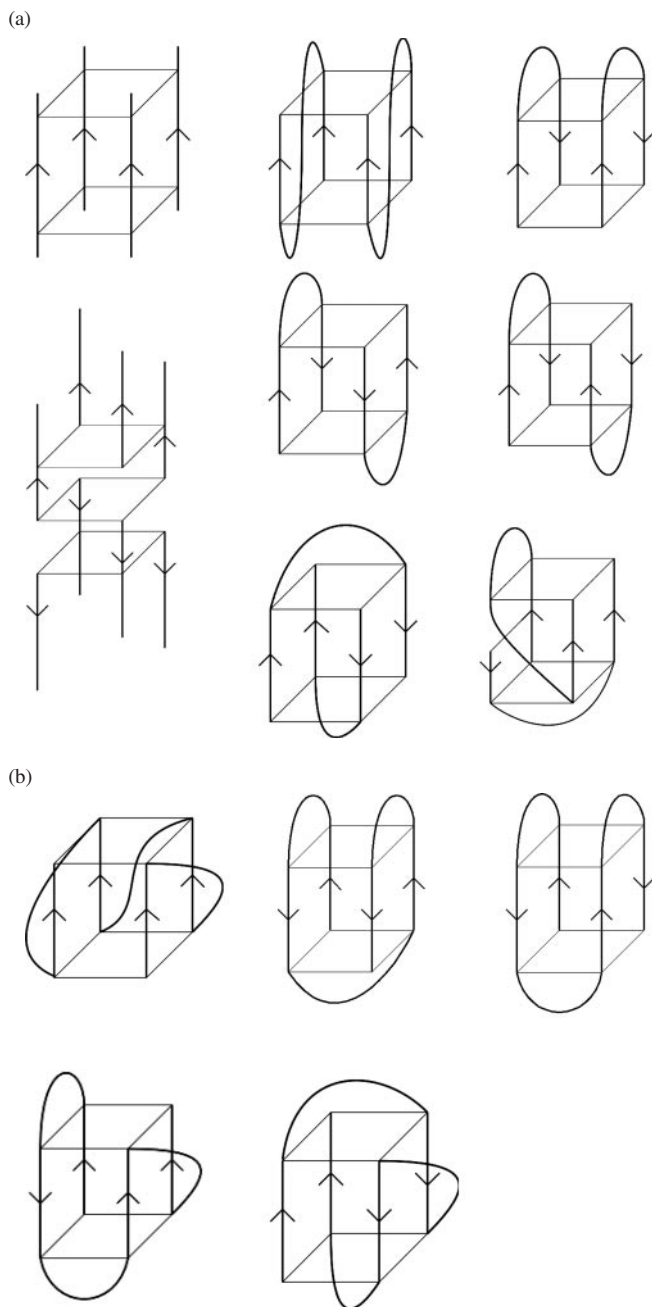


Figure 2. (a) Some possible topologies for simple tetramolecular (on the left-hand side) and bimolecular quadruplexes. Strand polarities are shown by arrows. (b) Some possible topologies for simple unimolecular quadruplexes.

to the buffer composition. Phosphate, acetate, sulphate and carbonate buffers should be avoided due to their strong absorbance at wavelengths commonly used for CD experiments. Many quadruplex-forming sequences have been studied using this technique and the majority of spectra conform to one of two characteristic spectral forms. Classic parallel and anti-parallel quadruplexes show similarly shaped traces but with maxima at distinct wavelengths. For quadruplexes assigned to be parallel-stranded, a maximum is present at ~ 260 nm and a minimum at ~ 240 nm; the maximum and minimum for an anti-parallel quadruplex are typically at

around 290 and 260 nm, respectively. These assignments have predominantly been used to examine telomeric and telomere-like sequences (i.e. sequences with regular repeating loop regions). As more complex quadruplex-forming sequences are examined, the reliability of assigning topology based on the comparison of a spectrum with the CD signature of known parallel or anti-parallel telomeric quadruplexes cannot always be assumed since (i) topologies may not conform to those observed with telomeric quadruplexes, (ii) multiple species cannot readily be identified in CD spectra and (iii) non-telomeric loop sequences may perturb the CD spectra in unforeseen ways.

X-ray crystallography and high-field NMR spectroscopy offer in principle the possibility of both topological assignment and more detailed atomic-level structure determination. However sometimes even with these methods, caveats are required. Successful structure determination by NMR methods relies on the sequence forming a kinetically stable species in solution; the presence of multiple species limits the structural information obtained. This may be overcome by the use of mutated or modified sequences based on the original G-rich sequence, but which only form a single species in solution environments. It is common practice to screen up to several tens of mutated sequences and other variants from wild-type until one is found that produces a well-resolved NMR spectrum showing a single species amenable to analysis. Favoured mutations are of thymine by uracil or 5-bromo-uracil. Variations in both 5' and 3' flanking sequence are also commonly explored. Crystallography similarly uses site mutations and/or sequence scanning, to find sequences that will crystallize. It is also necessary to use bases with heavy-atom substitutions (as in 5-bromo-thymine) for phasing purposes when confronted with structures that cannot be solved by molecular replacement. The various structures formed in solution by variants of the human telomeric two-repeat sequence (23) show that such mutations and changes cannot always be relied upon to preserve a particular topology and will inevitably alter the equilibrium between different ones, sometimes by forming additional stabilizing interactions. Thus generalizations from any one NMR or crystal structure need to be made with care, and need to take due regard of the role played by the modified/additional nucleotides, especially in the absence of independent data or more than one corroborating structure.

Tetramolecular and bimolecular quadruplex structures

Tetramolecular G-quadruplexes comprises the simplest category of quadruplex nucleic acid (Figure 2a). Thus the crystallographic and NMR structures of $d(TG_4T)_4$ (35,36) and its RNA equivalent $(UG_4U)_4$ (38) show all the strands parallel to one another and the guanine glycosidic torsion angles are all in the *anti* conformation. However, even tetramolecular G-quadruplexes can form more complex structures, as shown by $d(GGGT)_8$, in which eight strands form an interlocked bimolecular quadruplex (39) with two symmetric parallel tetramolecular $d(GGGT)$ quadruplexes being linked by an external G-tetrad formed by slipped-out guanines from each quadruplex. The family of sequences $d(GCGGXGGY)$ form tetramolecular structures comprising two unusual bistranded quadruplex monomer units containing G:C:G:C tetrads (40).

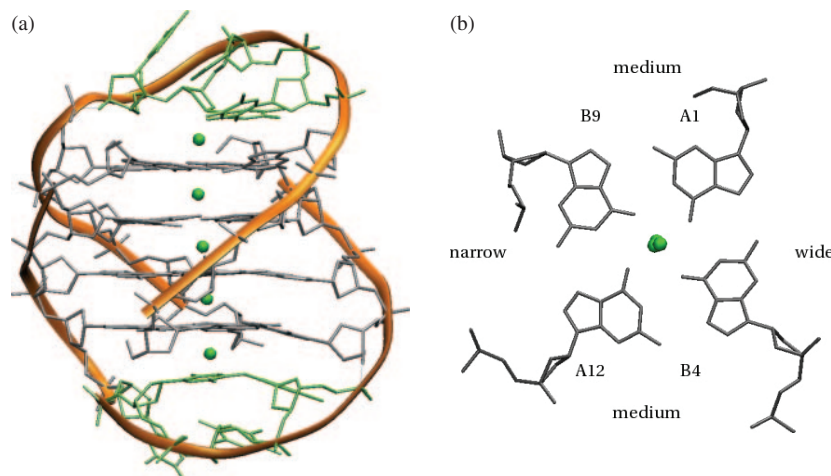


Figure 3. The crystal structure (28) of the bimolecular quadruplex formed by the *O.nova* telomeric sequence $d(G_4T_4G_4)$ (PDB entry 1JPQ). (a) Overall topology is indicated by the ribbon representation in orange. The details of the molecular structure are also shown. Potassium ions are shown as green spheres. (b) A projection down the central channel, indicating the relative widths of the four grooves

Association of two strands to produce bimolecular quadruplexes introduces increased topological variation (Figure 2a). The classic bimolecular quadruplex structure (Figure 3) is that formed by two strands of the *O.nova* sequence $d(G_4T_4G_4)$, with a diagonal T_4 loop at each end of the symmetric quadruplex (28–31). It is remarkable that even apparently conservative changes in this sequence have major topological consequences: thus $d(G_3T_4G_4)$, with one guanine at the 5' end less than in the wild-type sequence forms a bimolecular quadruplex having both lateral and diagonal loops (41). This is one of the few cases where a bimolecular quadruplex has an unequal number of parallel (three) and anti-parallel (one) strands; subsequent studies (42) showed that this topology is not dependent on the presence of NH_4^+ ions, but is retained in K^+ or Na^+ solution, as does a mixed di-cation form (43). The sequence isomer, now with one guanine less at the 3' end [i.e. $d(G_4T_4G_3)$], also forms an asymmetric bimolecular quadruplex, but with less dramatic differences compared to the *Oxytricha* parent structure. This structure has a core of three stacked G-tetrads, so the two guanines not included in this core are involved in one of the two diagonal loops (44). Reducing the number of guanines still further, to $d(G_3T_4G_3)$, results in a more conventional diagonal-looped quadruplex, but with asymmetry in guanine glycosidic angles (45,46). Decreasing the size of the thymine loops also results in topological change, as observed in the crystal structures (Figure 4) of the bimolecular quadruplexes formed by $d(G_4T_3G_4)$, with lateral loops being consistently favoured (26). The implication of this, that loops with three or less nucleotides dis-favour diagonal in preference to lateral loops, is borne out by the exclusive presence of lateral loops in both interconverting bimolecular quadruplexes formed by the $d(TG_4T_2G_4T)$ *Tetrahymena* sequence (24). These are closely similar to the head-to-head and head-to-tail lateral loop bimolecular quadruplexes of $d(G_4T_3G_4)$ (26). Interestingly, the 5' and 3' flanking thymine residues in this pair of sequences have no effect on quadruplex topology.

It is not possible at present to define a comprehensive set of rules that specifies the folding of bimolecular

G-quadruplexes, in the absence of much more structural and energetic information than is currently available, especially since in solution it is apparent that multiple structures sometimes exist in equilibrium. However, several significant contributing factors are apparent, notably loop length and sequence, and G-tract length (47,48). In general bimolecular quadruplex topology appears not to be markedly dependent on the nature of the cation, in striking contrast with unimolecular quadruplexes. Molecular dynamics simulations have been employed to model the stability of particular quadruplexes, such as that in the *Oxytricha* bimolecular topology (49–52). Simulations have suggested a set of preferences for thymine-containing loops (53), which are broadly in accord with the experimental observations from crystallographic and NMR studies, as outlined above, which show that T_3 loops have a marked preference for lateral loop conformations. This is not consistently indicated by the free-energy calculations, which may be a consequence of the inadequacies of current force fields to fully account for the electrostatics of quadruplexes, and of the likely small energy differences between differing loop conformations. On the other hand, shorter T_2 loops do restrict conformational flexibility of topological features. It also seems that differing numbers of guanines in the individual G-tracts results in quadruplexes with asymmetric topologies, which again, are not readily predictable at present.

Unimolecular quadruplexes

The same three loop types (propeller, lateral and diagonal) found in bimolecular quadruplexes also occur in unimolecular quadruplex structures (Figure 2b). For example the human telomeric sequence $d[AG_3(TTAGGG)_3]$ forms an anti-parallel arrangement in Na^+ solution (Figure 5a), with one diagonal and two lateral loops (54). In K^+ solution, this sequence appears to be able to access a number of distinct folds, as described further below; the crystal structure of this sequence (22) shows all strands in parallel orientations and therefore with the three TTA tracts forming three propeller loops. This all-parallel topology has been observed

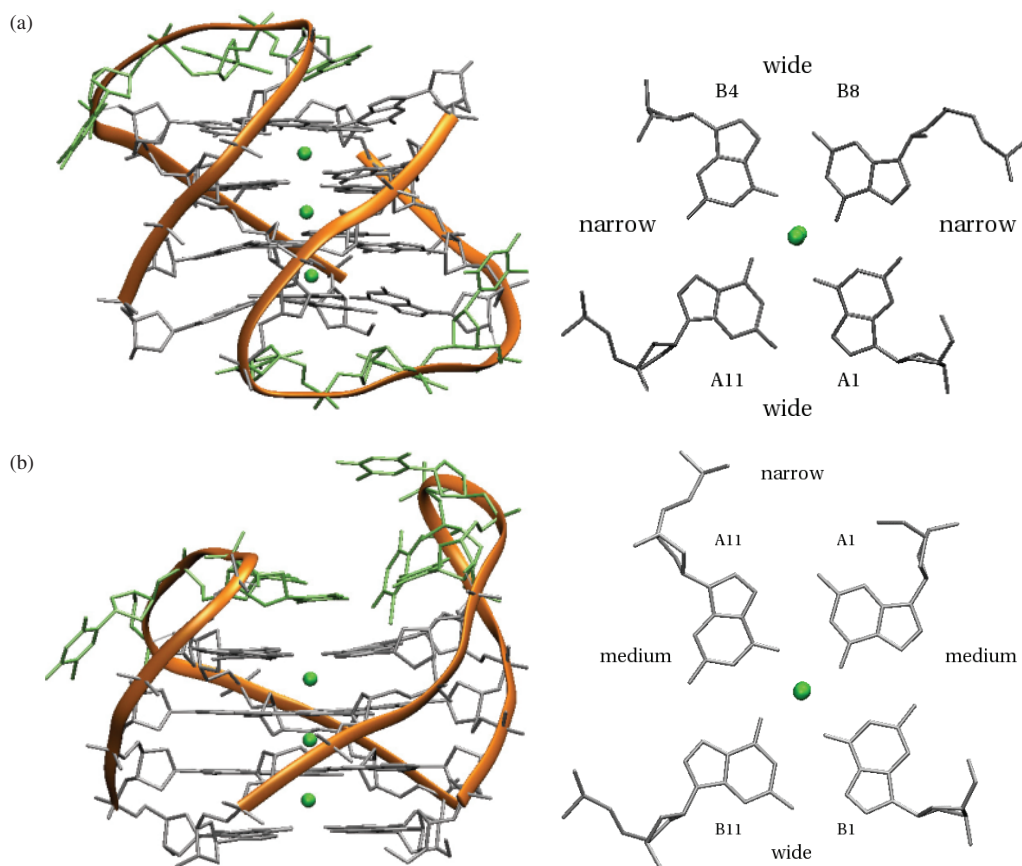


Figure 4. Crystal structure (26) of the two bimolecular quadruplexes found in the crystal structure of $d(G_4T_3G_4)$ (a) two views of the head-to-tail quadruplex (PDB entry 2AVH). (b) Two views of the head-to-head quadruplex (PDB entry 2AVJ).

for several other sequences in solution, e.g. for the aptamer sequence $d(G_4TG_3AG_2AG_3T)$, which is a potent inhibitor of HIV-1 integrase. This aptamer forms an interlocked quadruplex dimer, each with three single-nucleotide propeller loops (55). Propeller loops are also found in conjunction with lateral or diagonal loops, as in the $d(T_2G_4T_2G_4T_2G_4T_2G_4)$ and $d(G_2T_4G_2CAG_2GT_4G_2T)$ NMR structures (56,57). The size of the loop can affect unimolecular quadruplex stability (48); the *Oxytricha*-like unimolecular sequence, $d(G_4T_2G_4TGTG_4T_2G_4)$ has a more unfavourable ΔG^0 value than its bimolecular counterpart, although the former's melting temperature is considerably higher due to lower entropic contributions (58).

Few systematic studies have been reported of the effects of differing loop lengths and sequence on various unimolecular quadruplex folds and loop types. An analysis, restricted to loops with differing numbers of thymine residues, used molecular dynamics in conjunction with biophysical measurements (59), and has concluded that quadruplexes with three T_1 loops are constrained to only form parallel topologies, whereas quadruplexes with three T_2 loops can form both parallel and anti-parallel topologies (in this instance parallel structures are likely to be favoured). In addition, a single T_1 loop in a quadruplex is compatible with both parallel and anti-parallel arrangements, but the parallel type is more energetically favoured. Quadruplexes with a single T_2 to T_6 loop are stable with either parallel or anti-parallel topologies;

however, anti-parallel ones are likely to be slightly preferred. The conclusions regarding single-nucleotide loops are likely to be generally applicable to all four nucleotides A, T, G and C since loop size is the determinant of steric constraints on topology and energetics. However, the relative stabilities of loops with >1 nt are also dependent on relative nucleotide stacking energies within loops, as has been shown by a thermodynamic profiling study (60). Another factor has been highlighted by a study on the effects of ribonucleotide substitution for deoxynucleotide (61). Systematic substitutions in both loops and G-tracts have suggested that the greater tendency for ribonucleotides to be in an *anti* glycosidic conformation, resulting in a preference for parallel topologies in RNA quadruplexes.

Vertebrate telomeric quadruplexes

The large number of studies on the structure(s) adopted by repeats of the vertebrate telomeric sequence $d(TTAGGG)$ have, in large part, focused on the topology adopted by the folding of the single-stranded repeats at the 3' telomere end. The average length of this single-stranded overhang, of ca. 150 nt, corresponds to an assembly of 5–6 four-repeat unimolecular quadruplexes. Almost all considerations of the structural features of the 'human quadruplex' have focused on individual quadruplexes, especially the four-repeat unimolecular quadruplex(es) rather than the structure and

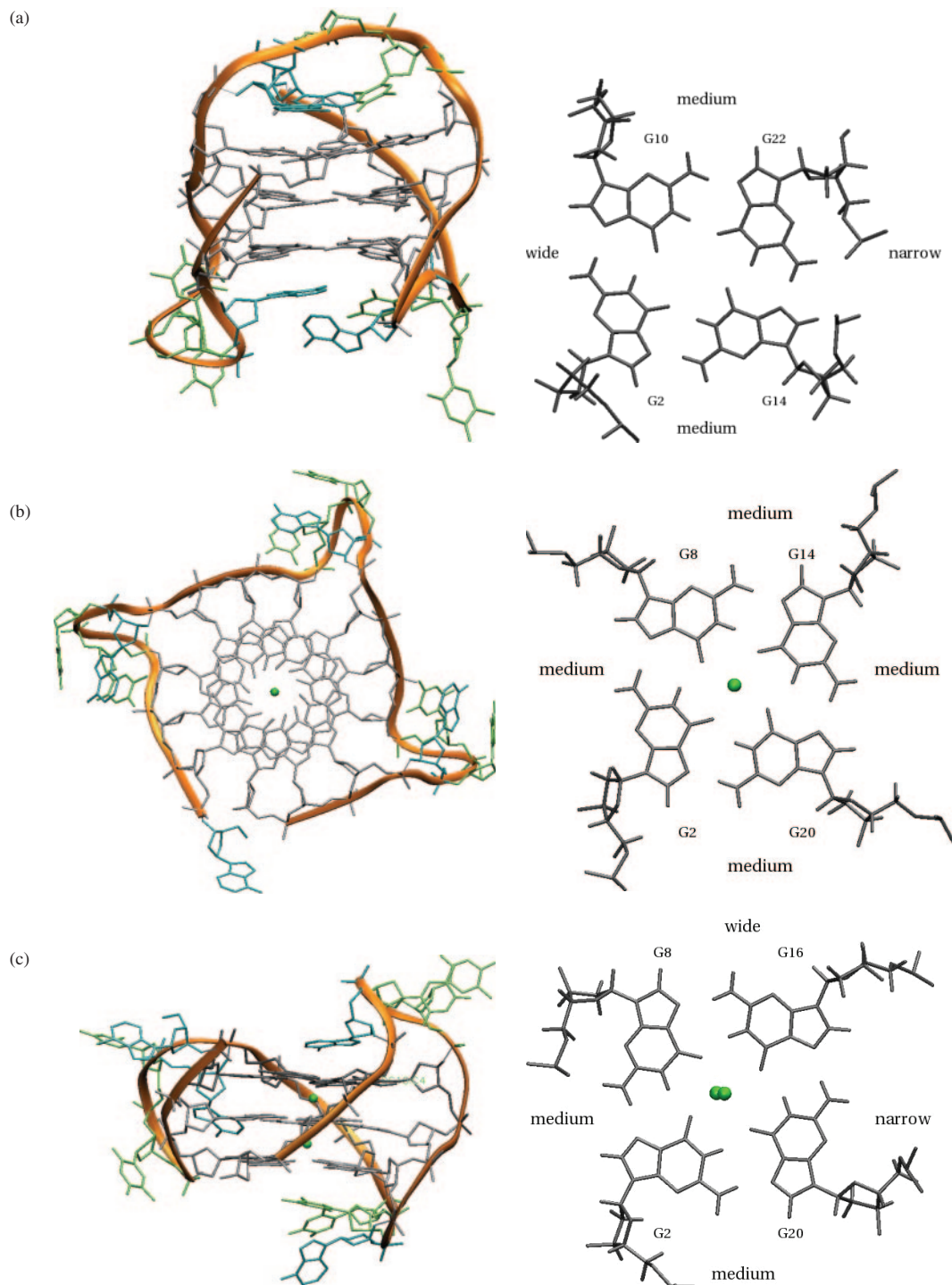


Figure 5. Structures of the human unimolecular telomeric quadruplex formed from the sequence d[AGGG(TTAGGG)]₃. In each case two views are shown (a) one of the deposited structures of the Na⁺ form, determined by NMR (PDB entry 143D) (54), with a diagonal and two lateral loops. (b) K⁺ form A, determined by crystallography (PDB entry 1KF1) (23), with three strand-reversal loops (c) K⁺ form B, showing the topology determined by NMR (75,96), with one strand-reversal and two lateral loops. Nucleotide loop conformations for the detailed atomic structure shown here have been obtained from a molecular dynamics simulation performed by Sarah Burge that has used this topology as a starting-point. The NMR-derived structure of one of the sequences determined experimentally (96), is also available as PDB entry id 2KGU.

dynamics of quadruplex assemblies, which are the more biologically relevant system (11,12). Apart from the 3' single-stranded overhang, all telomeric DNA is double-stranded [and associated with a number of telomeric proteins in the

'shelterin' complex (62)]. In the absence of proteins or small molecules, the equilibrium for vertebrate telomeric DNA has been found (63) to favour duplex over dissociation into quadruplex and i-form motifs (the four-stranded arrangements

formed by the complementary C-rich strand and organized around cytosine–cytosine base pairs).

There is good evidence from a range of biophysical techniques, that the four-repeat quadruplex formed by the sequence d(TTAGGG)₄ (and variants on it, notably d[AGGG(TTAGGG)₃]), adopt differing topologies in Na⁺ versus K⁺ solution (60,64–69). NMR analysis (54) of the species formed in Na⁺ conditions by the 22mer d[AGGG(TTAGGG)₃] has shown that the structure has an anti-parallel fold with two lateral and one diagonal loops, each loop comprising the TTA triad sequence (Figure 5a). Subsequent crystallographic analyses of this sequence and the related 12mer (i.e. two-repeat) sequence d(TAGGGTTAGGGT), in K⁺ solution (22), showed that they form a unimolecular (Figure 5b) and a bimolecular quadruplex, respectively in the crystal lattice. Both have the same topology with parallel orientations for all four strands, and propeller loops formed by the TTA sequence [single occurrences of these type of loops had been observed previously (56,57)]. This all-parallel arrangement, which is radically different from the Na⁺ structure, was subsequently observed in solution by NMR (23) for the closely-related sequence d(TAGGGUTAGGGT), although the same study also showed that the dominant form for another modified sequence, d(UAGGGT^{Br}UAGGGT), is that of an anti-parallel quadruplex with lateral loops. The propensity of telomeric quadruplexes for topological diversity is shown by the unusual asymmetric bimolecular quadruplex (69) formed by three telomeric repeats, with all three G-tracts of one strand associating with a single G-tract of another.

The unexpected nature of the quadruplex fold in the K⁺ crystal structures has led to a number of biophysical studies intent on identifying the nature of the species formed by d[AGGG(TTAGGG)₃] in solution [see e.g. Refs. (64–68,70–72)]. It is unsurprising that some studies suggest the co-existence of several forms (59,60,67), especially in view of the ability of quadruplexes with 3 nt loops to readily adopt topologically distinct structures upon small changes in environment or sequence, suggesting that the various forms are energetically-similar. This is in accord with both experimental and simulation studies, which show that there is only a small free-energy difference between the human telomeric parallel and anti-parallel quadruplexes with TTA loops (59,67). Thus a particular set of conditions or sequence will favour a particular fold or mixture of folds, analogous to the process of crystallization, which selects one or a few particular low-energy form(s) from solution, that are best able to pack effectively to give a well-ordered crystal. One key feature of the crystal structure's parallel fold is that the open nature of the G-tetrad surfaces of individual quadruplexes, due to the absence of lateral or diagonal loops, facilitates their stacking together into a very compact and stereochemically acceptable arrangement. This feature would also enable the assembly of successive quadruplexes, as would occur in biological telomeric DNA, and the binding of appropriate small-molecule drugs.

A recent CD study (73) of the K⁺ form of the 22mer sequence has exploited the property of 8-bromo-guanosine to favour the *syn* glycosidic angle conformation, and has incorporated this modification at various positions in the sequence to determine topology from CD measurements. This is challenging since not all the CD spectra of individual

modified sequences show behaviour consistent with the proposed structures. It was concluded that d[AGGG(TTAGGG)₃] in solution is a mixture of two forms, one of which has a new topology for telomeric unimolecular quadruplexes, having anti-parallel/parallel strands with one propeller and two lateral loops. The unambiguous identification of all the species present in solution using CD alone may not be straightforward (74), so the topology of any other components have not yet been clarified. This fold (Figure 5c) has also been reported in two separate NMR analyses (75,96). Both have used sequences that have been slightly altered at the termini from telomeric regularity: d[TTG₃(TTAG₃)₃A] (96) and d[AAAG₃(TTAG₃)₃AA] (75), since NMR finds that the native 22mer as used in the crystal structure determination, forms a mixture of species in K⁺ solution, which is not amenable to structure analysis. The structure of the former has been reported in detail (96), and shows that the extra flanking residues are involved in Watson–Crick and reverse Watson–Crick base pairs that are stacked one on each end of the core of G-tetrads, and help to stabilize this particular topology. This explains why the fold has not been observed to date with the 22mer, which cannot form such base pairs. Thus what remains still to be determined by fine structure methods is the precise nature of all the species present in the K⁺-solution of d[AGGG(TTAGGG)₃].

UNIMOLECULAR NON-TELOMERIC QUADRUPLEXES

Sequence occurrences

The realization that potential quadruplex-forming sequences can occur in double-stranded non-telomeric regions of the human genome (and therefore in other eukaryotic and prokaryotic genomes), is not new, and they have been identified, e.g. in promoter and immunoglobulin switch regions and in recombination hot spots (76). There have been two recent systematic surveys of the complete human genome sequence, searching for potential unimolecular quadruplex-forming sequences (18,19). Both have used the same criteria for the definition of a potential quadruplex sequence and have agreed on the overall number of these sequences present, even though the statistical and analytical approaches used were quite different. These studies assumed that long-range and even medium length loops, although feasible are impractical to include because of the very large number of possibilities, which would be present. The criteria for a potential quadruplex sequences was therefore restricted to:



where N_{L1-3} are loops of unknown length, within the limits 1 < N_{L1-3} < 7 nt.

Potential quadruplex sequences are distributed throughout the human genome in exons, introns, in untranslated regions, in promoter sequences (sometimes though not invariably directly upstream of transcription start sites) and within gene desert regions. The majority of potential quadruplex sequences appear to be involved in more than one possible quadruplex, either as a result of being in a sequence with more than four consecutive runs of guanine (Table 2a) or because a lack of parity in the lengths of the loop sequence

Table 2. Examples of ambiguity in potential quadruplex sequences showing (a) uneven guanine runs creating a choice of loop sequence and (b) where more than four consecutive guanine tracts gives rise to more than one possible quadruplex fold for a sequence

| | |
|---|-----------------------------------|
| (a) Uneven guanine runs | |
| choice 1 | GGGCCTGGGCTGGGCTGGG |
| choice 2 | GGGCCTGGGCTGGGCTGGG |
| (b) More than four consecutive guanine tracts | |
| choice 1 | GGGCCTGGGCCTGGGCCTGGGCCTGGGCCTGGG |
| choice 2 | GGGCCTGGGCCTGGGCCTGGGCCTGGGCCTGGG |
| choice 3 | GGGCCTGGGCCTGGGCCTGGGCCTGGGCCTGGG |

Table 3. The top 20 most frequently occurring loop sequences (18)

| Rank | Sequence | Population |
|------|----------|------------|
| 1 | A | 193 756 |
| 2 | T | 121 406 |
| 3 | C | 44 020 |
| 4 | AA | 40 026 |
| 5 | CT | 32 472 |
| 6 | CA | 32 070 |
| 7 | G | 29 623 |
| 8 | AT | 19 957 |
| 9 | AGA | 19 144 |
| 10 | TT | 17 089 |
| 11 | TA | 12 641 |
| 12 | CC | 10 955 |
| 13 | AGT | 9896 |
| 14 | AGGA | 9463 |
| 15 | AGGT | 9434 |
| 16 | TGA | 9237 |
| 17 | AAA | 7839 |
| 18 | CCT | 7151 |
| 19 | TGT | 6619 |
| 20 | CCA | 6269 |

(Table 2b) means that some of the guanines in the G-runs have to be part of at least one of the loop sequences. Although it is possible that many of the potential topologies exist in dynamic equilibrium we cannot at present predict which are stable. This fold ambiguity will require much more extensive experimental data before generalized theoretical approaches to predicting folding can be reliable.

When every possible combination is considered, a survey of the Ensemble database (V20 NCBI assembly 34c) yields 5 713 900 possible potential quadruplex sequences in the human genome. However, this corresponds to a maximum of 375 157 distinct non-overlapping potential quadruplex sequences. Since a unimolecular quadruplex sequence can be broken down into four equal sized G-tracts and three distinct loop regions we can characterize different quadruplex sequences by the contents of their loop regions. Over a range of loop lengths of 1 to 7 bases there are 21 844 possible sequences of which 20 492 are actually found at least once in the human genome. The large differences in the number of times that these loops occur indicates that some sequences are over-represented whereas others are highly under-represented within the entire population of potential quadruplex-forming sequences (Table 3).

Unsurprisingly there is a tendency for a high proportion of quadruplex sequences to occur within promoter sequences given that these are G-rich regions. This is reflected in

the frequency of occurrence of each loop sequence. Small single-base loops are the most common and loops which are made up of guanines in the centre of a sequence are abundant e.g. AGGA and AGGT are very common loop sequences. In general, longer sequences occur less frequently within quadruplex loops. There are however notable exceptions. For example, CCTGTT and TAGCATT are highly over-represented among possible six and seven base pair loop sequences.

Unique sequence motifs have been identified in the human genome by examining how frequently a given loop sequence is found in the first, second or third loop position e.g. the sequence CCTGTT occurs in the human genome 1266 times as a first loop sequence, only 18 times as a second loop and just 9 times in the third loop position. There are several variants of this sequence with a similar pattern of loop distribution, with the common feature that they tend to have CCTGT within the sequence of the first loop. Although seemingly ubiquitous throughout the human genome, this sequence is also strongly represented in the Human Endogenous Retrovirus Database (77).

The findings that quadruplex sequences occur in the promoter regions of several cancer genes have stimulated a number of structural studies, which are outlined in a subsequent section. We list in Table 4 a selection of potential quadruplex sequences, mostly directly upstream of the transcription start site in a set of cancer-associated genes (from the Wellcome Trust Sanger Institute Cancer Genome Project web site (<http://www.sanger.ac.uk/genetics/CGP>)).

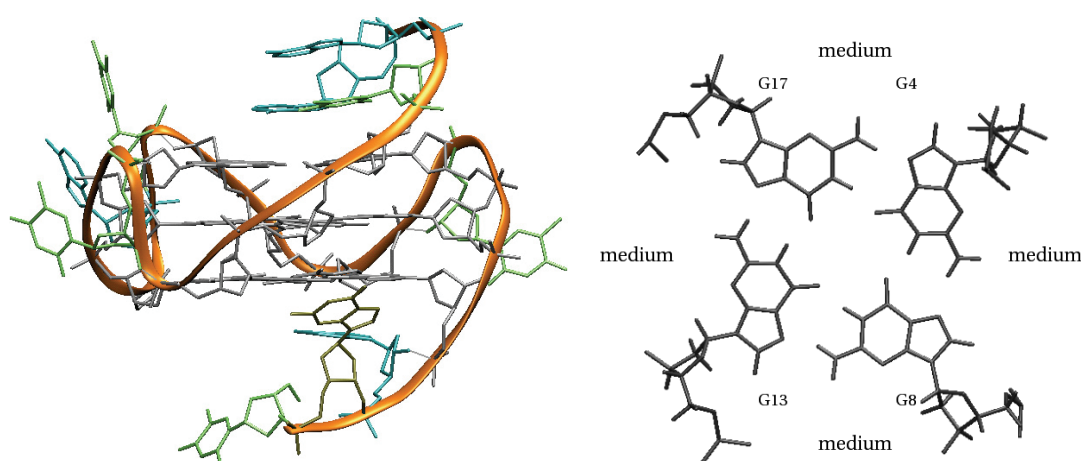
The concept that quadruplex formation may in particular provide a transcriptional regulatory signal has received support from the analysis of quadruplex occurrence in *Escherichia coli* and 17 other prokaryotic genomes (21), where G4 sequences are statistically significantly over-represented in promoter regions proximal to transcription start sites, and may be associated with global supercoiling-sensitive gene regulation. The occurrence of quadruplex sequences in alternatively spliced mammalian pre-mRNA sequences has been surveyed, and results are now available in the GRSDDB database (78). There is yet rather little experimental data on RNA quadruplexes; since some RNA quadruplexes are likely to have high stability these would be more likely to be present in non-translated mRNA sequences. It has been suggested that the fragile X mental retardation protein (FMRP) binds with high affinity to G-rich mRNA targets in yet undefined genes, which can form quadruplex arrangements: one with a possible parallel topology has been identified (79). A survey of all 16 654 genes in the human gene database has found that there is a significant correlation between quadruplex sequence occurrence and classes of gene (80), with proto-oncogenes having a high quadruplex-forming potential compared with a low potential for tumour-suppressor genes. It is suggested (80) that this reflects genomic structure being selected based on gene function.

Potential quadruplex-forming sequences also occur within regions of chromosomal translocations. One such well-characterized example is the breakpoint region on human chromosome 14 associated with the lymphoma-associated bcl-2 gene translocating to chromosome 18 (81). The region just downstream of the breakpoint is G-rich with runs of short G-tracts characteristic of a quadruplex-type sequence.

Table 4. Sequences in cancer-related genes that have been identified as forming quadruplex structures

| Gene | Sequence | Ref |
|---------------------------------|--------------------|--|
| <i>c-myc</i> | <i>Pu27</i> 88 | T T A T G G G G A G G G T G G G G A G G G T G G G G A A G G |
| | <i>Myc-2345</i> 88 | T G A G G G T G G G G A G G G T G G G G A A |
| | <i>Myc-1245</i> | T G G G G A G G G T T T T A G G G T G G G G A |
| | <i>Myc-22</i> 89 | T G A G G G T G G G T A G G G T G G G T A A |
| | <i>Pu241</i> 90 | T G A G G G T G G I G A G G G T G G G G A A G G |
| <i>c-kit</i> | <i>ckit1</i> 91 | C A G A G G G A G G G C G C T G G G A G G A G G G G C T G |
| | <i>ckit2</i> 92 | C C C C G G G C G G G C G C G A G G G G A G G G G A G G C |
| <i>VEGF</i> | 96 | C C C G G G G C G G G C C G G G G C G G G G T C C C G G C G G G G C G G A G |
| <i>HIF-1α</i> | 95 | G C G A G G G C G G G G A G A G G G G A G G G G C G C G |
| <i>bcl-2</i> | 93 | G T C G G G G C G A G G G C G G G G A A G G A G G G C G C G G G C G G G G A |
| <i>k-ras</i> | 97 | G G G A G G G A G G G A A G G A G G G A G G G A G G G A |
| <i>Rb</i> | 98 | C G G G G G T T T T G G G C G G C |

The likely G-tracts that are involved in G-quartets are marked in red.

**Figure 6.** NMR-derived topology and one of the deposited structures of the *c-myc* quadruplex (86) (PDB entry 1XAV).

Analogous G-tracts have been mapped in the breakpoint region of the SHANK3 gene (82).

Much of the interest in non-telomeric quadruplex sequences and their possible structures are due to their occurrence in genes associated with proliferation, especially in *c-myc* and a number of oncogenes (outlined below). The biological implications of these sequences is as yet at an early stage of study, although the possibility is currently being actively explored that they may be involved in the regulation of gene expression, and that this might be amenable to exploitation by small-molecule therapy (83). Evidence from the small number of non-telomeric quadruplex structures available to date suggests that there is large diversity both in topology and molecular structure between topoisomers, which may be exploited in the future for therapeutic gain (84,85).

Topology and structure

Quadruplex formation has been examined *in vitro* in a number of non-telomeric sequences (Table 4). The NHE III₁ G-rich sequence in the promoter region of the *c-myc* oncogene, which is responsible for 80–90% of its transcriptional activity, has been especially studied. The existence of a quadruplex within this promoter region was initially

proposed (20) based on the data from chemical probe and gel mobility measurements, and from fluorescence resonance energy transfer spectroscopy (86). Subsequent studies established a relationship between quadruplex stabilization within this sequence and suppression of *c-myc* transcriptional activation (83), with the porphyrin TMPyP4 acting to stabilize the quadruplex structure. NMR in solution (87–89), as well as that of a porphyrin TMPyP4 complex (89) has determined topology and detailed structures of several *c-myc* quadruplex sequences. Non-telomeric G-rich regions often contain more than four consecutive G-tracts (see above) which, as in the case of *c-myc*, results in the formation of multiple quadruplex species in the native *Pu27* region (this dynamic behaviour, which may involve shuffling between G-tracts, is distinct from the conformational rearrangements shown by the human telomeric 22mer, e.g.). Shorter sequences from within the *Pu27* region have been successfully analysed by NMR methods. *Myc-2345* and *Myc-1245* each contain four G-tracts and form very stable parallel-stranded unimolecular quadruplexes in solution (Figure 6), with G-tracts joined through propeller loops, and with all guanines involved in the G-tetrads having an *anti* conformation (87). These structures therefore share features with the human telomeric 22mer K⁺ crystal structure (22). In both *myc-1245* and *myc-2345*, structures the first and third loops are single-nucleotide propeller

loops; the second (central) loop in *myc-2345* is a GA propeller loop and that in the non-natural *myc-1245* consists of a large six-base TTTTAA loop. This loop is destabilizing compared to that in the shorter *myc-2345* sequence, as shown by the 16°C difference in their melting temperatures. Both of these *c-myc* quadruplexes have higher melting temperatures than the human telomeric quadruplexes in similar K⁺ conditions. One remarkable feature of the *Pu241* NMR structure, which has five G-tracts, is the fold-back of the 3' terminal G in the last G-tract, enabling participation in a G-tetrad and the establishment of a neighbouring G-A•G hydrogen-bonded triad that is positioned as a planar platform-like diagonal loop above this G-tetrad. The other face of the quadruplex has a stack of base pairs and bases arising from the 5' end sequence. The NMR solution structure of a complex of the *Pu241* quadruplex with the porphyrin ligand TMPyP4 (89) shows that it stacks on the other terminal G-tetrad, sandwiched against one of base pair platforms, with overall little perturbation from the ligand-free *Pu241* quadruplex structure.

NMR methods have been used to validate the existence of quadruplex species in two G-rich stretches of the promoter region of the *c-kit* kinase gene, an important therapeutic target in gastro-intestinal tumours (90,91). Unusually, NMR has observed only a single quadruplex species in K⁺ solution for one of these sequences (90), which is 87 bp upstream of the transcription start site, and therefore there will be no ambiguity as to the topology of the quadruplex species, once determined. This is probably due to both the four G-tracts constraining the number of quadruplexes, and the presence of topologically restraining single-nucleotide loops (Table 4). The second quadruplex sequence that has been identified in the promoter region of the *c-kit* gene (91) occupies a site required for core promoter activity. This sequence requires to be mutated in order to act as a single quadruplex species, probably with a parallel fold. Both these quadruplexes have high conservation across vertebrate species, suggestive of a functional role for them.

Chemical footprinting and CD methods have been used to characterize (92) a quadruplex species found in a nuclease-hypersensitive sequence within the vascular endothelial growth factor (VEGF) promoter region that is essential for basal promoter activity in human cancer cells. This quadruplex, which is induced to form from the duplex sequence by the ligands telomestatin or TMPyP4, has been assigned a parallel topology, which is consistent with the presence of two single-nucleotide loops. A parallel quadruplex has also been proposed (93) for a sequence in the hypoxia inducible factor 1 α (HIF-1 α) promoter region, based on footprinting and CD data; again the sequence (Table 4) has two likely single-nucleotide loops. The *bcl-2* oncogene has a major transcriptional promoter sequence ca. 1400 bp upstream of the transcription start site, that has been characterized as showing quadruplex characteristics (94,95). A potential loop sequence in this quadruplex region has the sequence AGGA in common with one of the *c-kit* sequences (90); this sequence was predicted (18) to have a high frequency of occurrence. An NMR study of the *bcl-2* quadruplex (94) shows that one of the topologies for this mixed parallel/anti-parallel unimolecular quadruplex has two lateral loops and one propeller loop, analogous to one of the 22mer vertebrate telomeric

quadruplex topologies (75,96), but with a reversed order of loops. Putative quadruplexes have been assigned within the *k-ras* (97) and neuroblastoma oncogenes (98).

It is striking that a high proportion of these quadruplex topological assignments suggest folds with all strands parallel [for which the vertebrate telomeric 22mer crystal structure (22) is the paradigm]. This is the inevitable consequence of the presence of at least two single-nucleotide loops, which energetically disbar these sequences from forming lateral or diagonal-looped anti-parallel structures. Although we know little as yet about the structural basis for loop sequence preferences, it is also reassuring that the sequences with high occurrences in the human genome overall are among those that have been observed (in the admittedly rather small sample base accumulated to date). This may suggest that these are the loop sequences with greatest stability. The NMR structure of the G-quadruplex from the *c-myc* promoter sequence with five G-tracts (89) shows that the presence of the fifth (very short) tract can produce unexpected and significant changes in topology compared to analogous sequences with four tracts, so topology and structure prediction for new quadruplexes based on the very small number of known quadruplex structures is probably premature at present.

CONCLUDING REMARKS

The quadruplexes studied to date by crystallography and NMR has revealed a diversity of topology and structure not shown by any other type of DNA. The symmetric telomeric quadruplexes comprise a small group of structures, although the topology and fine structure of consecutive unimolecular quadruplexes formed on the vertebrate single-stranded telomeric overhang remain to be established. The potentially much larger group of genomic quadruplexes have inherent sequence diversity (and usually asymmetry), and this will undoubtedly be reflected in high structural diversity once further structures beyond the *c-myc* quadruplex, are established. A note of caution is needed. Even more than with telomeric quadruplexes, double helical (or mRNA for quadruplexes in transcribed sequences) sequence context will ultimately need to be taken into account when describing genomic quadruplex form and function. The question of the extent to which quadruplex structural complexity can rival that of RNA folding (and possibly even have catalytic activity), remains to be fully answered, but again, sequence context may play a role. The folding of RNA itself into unimolecular quadruplex structures is largely unexplored, and it would be unsurprising if these RNA quadruplexes were to have greater complexity, given the propensity of complex RNAs for example to exploit the 2'-hydroxy group in folding. Other backbone chemistries, such as in peptide nucleic acids (PNA) and conformationally locked nucleic acids (LNA), can also form tetramolecular and bimolecular quadruplexes (99–102) that do not always conform to the topological patterns shown by DNA ones, and again it is to be expected that unimolecular quadruplexes with such backbones may also have topological novelty.

This review has not discussed the possible genomic quadruplex-related arrangements that may be formed by the expansion of triplet repeats, such as CCG, CTG, CAG or GAA, which are found in a range of genetic disorders repeats.

They share at least some features with the G-quadruplexes discussed here [see e.g. Refs. (103,104)], together with the added complexity of non-G base associations that can form motifs, such as base pentads and heptads.

Whether or not duplex genomic G-rich sequences may form quadruplex-based structures *in vivo* remains to be fully demonstrated, although supportive data are starting to emerge (105). Data from the *c-myc* sequences (84) when embedded in plasmids, which are then transfected in cells is strongly indicative of quadruplex formation in a more complex environment than the NMR tube; however, the difficulty of observing endogenous DNA quadruplex structures in live cells is a major challenge for the future. Resolution of the requirement for duplex unwinding in promoter sequences in order for the G-rich strand to transiently become single-stranded, may be achieved by strand scission; it has been recently reported that DNA topoisomerase II produces double-strand breaks in promoter sequences (106), which could be a suitable mechanism for achieving this. Even if the propensity for native sequences to form quadruplexes is low, it may be that their induction is achievable with ligands with selectivity not only for particular topologies but also for the detailed structural features of individual quadruplexes. The demonstration (107) that a *c-myc* quadruplex can be induced by the porphyrin TMPyP4 to form in preference to its duplex sequence, is an important step in this direction. Thus, a major focus for future quadruplex studies is as a target for selective therapeutic intervention. For example stabilization of a G-quadruplex structure in upstream regions important for maximum promoter activity, as in the case of the NHE III sequence in the *c-myc* gene (83), would result in down-regulation of gene expression. The large number of potential quadruplex sites in a genomes, implies that many target quadruplexes may have unique architectures, and therefore that selective stabilizing ligands can be devised. The number of known quadruplex structures is as yet very limited compared to the large number of quadruplex-forming sequences (18,19), strongly suggesting that we can look forward to the determination of a large number of diverse quadruplex topologies and structures.

ACKNOWLEDGEMENTS

The authors are grateful for support from Cancer Research UK, the Association for International Cancer Research and the School of Pharmacy, and to Keith Fox and the anonymous reviewers for their constructive comments. Funding to pay the Open Access publication charges for this article was provided by JISC.

Conflict of interest statement. None declared.

REFERENCES

- Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) Helix formation by guanylic acid. *Proc. Natl Acad. Sci. USA*, **48**, 2013–2018.
- Zimmerman,S.B., Cohen,G.H. and Davies,D.R. (1975) X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid. *J. Mol. Biol.*, **92**, 181–192.
- Arnott,S., Chandrasekaran,R. and Marttila,C.M. (1974) Structures for polyinosinic acid and polyguanylic acid. *Biochem. J.*, **141**, 537–543.
- Howard,F.B., Frazier,J. and Miles,H.T. (1977) Stable and metastable forms of poly(G). *Biopolymers*, **16**, 791–809.
- Williamson,J.R. (1994) G-tetrad structures in telomeric DNA. *Ann. Rev. Biophys. Biomol. Struct.*, **23**, 703–730.
- Davies,J.T. (2004) G-tetrads 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Intl. Edit.*, **43**, 668–698.
- Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
- Sundquist,W.I. and Klug,A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825–829.
- Wright,W.E., Tesmer,V.M., Huffman,K.E., Levene,S.D. and Shay,J.W. (1997) Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev.*, **11**, 2801–2809.
- Sfeir,A.J., Chai,W., Shay,J.W. and Wright,W.E. (2005) Telomere-end processing the terminal nucleotides of human chromosomes. *Mol. Cell*, **18**, 131–138.
- Lei,M., Podell,E.M. and Cech,T.R. (2004) Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nature Struct. Mol. Biol.*, **11**, 1223–1229.
- Zaug,A.J., Podell,E.R. and Cech,T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension *in vitro*. *Proc. Natl Acad. Sci. USA*, **102**, 10864–10869.
- Cech,T.R. (2004) Beginning to understand the end of the chromosome. *Cell*, **116**, 273–279.
- Huang,C., Lin,Y., Huang,S. and Chen,C.W. (1998) The telomeres of *Streptomyces* chromosomes contain conserved palindromic sequences with potential to form complex secondary structures. *Mol. Microbiol.*, **28**, 905–916.
- Jing,N., Zhu,Q., Yuan,P., Li,Y., Mao,L. and Tweardy,D.J. (2006) Targeting signal transducer and activator of transcription 3 with G-tetrad oligonucleotides: a potential novel therapy for head and neck cancer. *Mol. Cancer Ther.*, **5**, 279–286.
- McMicken,H.W., Bates,P.J. and Chen,Y. (2003) Antiproliferative activity of G-tetrad-containing oligonucleotides generated by a novel single-stranded DNA expression system. *Cancer Gene Ther.*, **10**, 867–869.
- Kelly,J.A., Feigon,J. and Yeates,T.O. (1996) Reconciliation of the X-ray and NMR structures of the thrombin-binding aptamer d(GGTGGTGGTGGTGG). *J. Mol. Biol.*, **256**, 417–422.
- Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
- Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2901–2907.
- Simonsson,T., Pecinka,P. and Kubista,M. (1998) DNA tetraplex formation in the control region of *c-myc*. *Nucleic Acids Res.*, **26**, 1167–1172.
- Rawal,P., Kummarasetti,V.B., Ravindran,J., Kumar,N., Halder,K., Sharma,R., Mukerji,M., Das,S.K. and Chowdhury,S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res.*, **16**, 644–655.
- Parkinson,G.N., Lee,M.P.H. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
- Phan,A.T. and Patel,D.J. (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.
- Phan,A.T., Modi,Y.S. and Patel,D.J. (2004) Two-repeat *Tetrahymena* telomeric d(TGGGGTTGGGGT) sequence interconverts between asymmetric dimeric G-quadruplexes in solution. *J. Mol. Biol.*, **338**, 93–102.
- Kettani,A., Bouaziz,S., Gorin,A., Zhao,H., Jones,R.A. and Patel,D.J. (1998) Solution structure of a Na cation stabilized DNA quadruplex containing G.G.G.G and G.C.G.C tetrads formed by G-G-G-C repeats observed in adeno-associated viral DNA. *J. Mol. Biol.*, **282**, 619–636.
- Hazel,P., Parkinson,G.N. and Neidle,S. (2006) Topology variation and loop structure homology in crystal and simulated structures of a bimolecular DNA quadruplex. *J. Amer. Chem. Soc.*, **128**, 5480–5487.

27. Gill, M.L., Strobel, S.A. and Loria, J.P. (2005) ^{205}Tl NMR methods for the characterisation of monovalent cation binding to nucleic acids. *J. Amer. Chem. Soc.*, **127**, 16723–16732.
28. Haider, S., Parkinson, G.N. and Neidle, S. (2002) Crystal structure of the potassium form of an *Oxytricha nova* G-quadruplex. *J. Mol. Biol.*, **320**, 189–200.
29. Smith, F.W. and Feigon, J. (1992) Quadruplex structure of *Oxytricha* telomeric DNA oligonucleotides. *Nature*, **356**, 164–168.
30. Schultze, P., Hud, N.V., Smith, F.W. and Feigon, J. (1999) The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the *Oxytricha nova* telomere repeat oligonucleotide d(G₄T₄G₄). *Nucleic Acids Res.*, **27**, 3018–3028.
31. Schultze, P., Smith, F.W. and Feigon, J. (1994) Refined solution structure of the dimeric quadruplex formed from the *Oxytricha* telomeric oligonucleotide d(GGGGTTTTGGGG). *Structure*, **2**, 221–233.
32. Clay, E.H. and Gould, I.R. (2005) A combined QM and MM investigation into guanine quadruplexes. *J. Mol. Graphics Modelling*, **24**, 138–146.
33. van Mourik, T. and Dingley, A.J. (2005) Characterization of the monovalent ion position and hydrogen-bond network in guanine tetrads by DFT calculations of NMR parameters. *Chemistry*, **7**, 6064–6079.
34. Dingley, A.J., Peterson, R.D., Grzesiek, S. and Feigon, J. (2005) Characterization of the cation and temperature dependence of DNA quadruplex hydrogen bond properties using high-resolution NMR. *J. Amer. Chem. Soc.*, **127**, 14466–14472.
35. Aboul-ela, F., Murchie, A.I. and Lilley, D.M. (1992) NMR study of parallel-stranded tetraplex formation by the hexadeoxynucleotide d(TG₄T). *Nature*, **360**, 280–282.
36. Phillips, K., Dauter, Z., Murchie, A.I., Lilley, D.M. and Luisi, B. (1997) The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution. *J. Mol. Biol.*, **273**, 171–182.
37. Rueda, M., Luque, F.J. and Orozco, M. (2006) G-quadruplexes can maintain their structure in the gas phase. *J. Amer. Chem. Soc.*, **128**, 3608–3619.
38. Deng, J., Xiong, Y. and Sundaralingam, M. (2001) X-ray analysis of an RNA tetraplex (UGGGGU)₄ with divalent Sr²⁺ ions at subatomic resolution (0.61 Å). *Proc. Natl Acad. Sci. USA*, **98**, 13665–13670.
39. Krishnan-Ghosh, Y., Liu, D. and Balasubramanian, S. (2004) Formation of an interlocked quadruplex dimer by d(GGGT). *J. Amer. Chem. Soc.*, **126**, 11009–11016.
40. Mergny, J.-L., De Cian, A., Amrane, S. and Webba de Silva, M. (2006) Kinetics of double-chain reversals bridging contiguous tetrads in tetramolecular quadruplexes. *Nucleic Acids Res.*, **34**, 2386–2397.
41. Cmugelj, M., Sket, P. and Plavec, J. (2003) Small change in a G-rich sequence, a dramatic change in topology: new dimeric G-quadruplex folding motif with unique loop orientations. *J. Am. Chem. Soc.*, **125**, 7866–7871.
42. Sket, P., Cmugelj, M. and Plavec, J. (2004) d(G₃T₄G₄) forms unusual dimeric G-quadruplex structure with the same general fold in the presence of K⁺, Na⁺ or NH₄⁺ ions. *Bioorg. Med. Chem.*, **12**, 5735–5744.
43. Sket, P., Crnugelj, M. and Plavec, J. (2005) Identification of mixed di-cation forms of G-quadruplex in solution. *Nucleic Acids Res.*, **33**, 3691–3697.
44. Cmugelj, M., Hud, N.V. and Plavec, J. (2002) The solution structure of d(G₄T₄G₃)₂: a bimolecular G-quadruplex with a novel fold. *J. Mol. Biol.*, **320**, 911–924.
45. Smith, F.W., Lau, F.W. and Feigon, J. (1994) d(G₃T₄G₃) forms an asymmetric diagonally looped dimeric G-quadruplex with guanosine 5'-syn-syn-anti and 5'-syn-anti-anti N-glycosidic conformations. *Proc. Natl Acad. Sci. USA*, **91**, 10546–10550.
46. Strahan, G.D., Keniry, M.A. and Shafer, R.H. (1998) NMR structure refinement and dynamics of the K⁺-[d(G₃T₄G₃)]₂ quadruplex via particle mesh Ewald molecular dynamics simulations. *Biophys. J.*, **75**, 968–981.
47. Cevec, M. and Plavec, J. (2005) Role of loop residues and cations on the formation and stability of dimeric DNA G-quadruplexes. *Biochemistry*, **44**, 15238–15246.
48. Risitano, A. and Fox, K.R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **32**, 2598–2606.
49. Spackova, N., Berger, I. and Sponer, J. (1999) Nanosecond molecular dynamics simulations of parallel and antiparallel guanine quadruplex DNA molecules. *J. Amer. Chem. Soc.*, **121**, 5519–5534.
50. Spackova, N., Berger, I. and Sponer, J. (2001) Structural dynamics and cation interactions of DNA quadruplex molecules containing mixed guanine/cytosine tetrads revealed by large-scale MD simulations. *J. Amer. Chem. Soc.*, **123**, 3295–3307.
51. Stefl, R., Cheatham, T.E., III, Spackova, N., Fadrna, E., Berger, I., Koca, J. and Sponer, J. (2003) Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of substates. *Biophys. J.*, **85**, 1787–1804.
52. Fadrna, E., Spackova, N., Stefl, R., Koca, J., Cheatham, T.E., III and Sponer, J. (2004) Molecular dynamics simulations of guanine quadruplex loops: advances and force field limitations. *Biophys. J.*, **87**, 227–242.
53. Hazel, P., Parkinson, G.N. and Neidle, S. (2006) Predictive modelling of topology and loop variations in dimeric DNA quadruplex structures. *Nucleic Acids Res.*, **34**, 2117–2127.
54. Wang, Y. and Patel, D.J. (1993) Solution structure of the human telomeric repeat d[AG₃(T₂AG₃)₃] G-tetraplex. *Structure*, **1**, 263–282.
55. Phan, A.T., Kuryavii, V., Ma, J.B., Faure, A., Andreola, M.L. and Patel, D.J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. *Proc. Natl Acad. Sci. USA*, **102**, 634–639.
56. Wang, Y. and Patel, D.J. (1994) Solution structure of the Tetrahymena telomeric repeat d(T₂G₄)₄ G-tetraplex. *Structure*, **2**, 1141–1156.
57. Kuryavii, V., Majumdar, A., Shallop, A., Chernichenko, N., Skripkin, E., Jones, R. and Patel, D.J. (2001) A double chain reversal loop and two diagonal loops define the architecture of a unimolecular DNA quadruplex containing a pair of stacked G(syn)-G(syn)-G(anti)-G(anti) tetrads flanked by a G-(T-T) Triad and a T-T-T triple. *J. Mol. Biol.*, **310**, 181–194.
58. Petraccone, L., Erra, E., Esposito, V., Randazzo, A., Mayol, L., Nasti, L., Barone, G. and Giancola, C. (2004) Stability and structure of telomeric DNA sequences forming quadruplexes containing four G-tetrads with different topological arrangements. *Biochemistry*, **43**, 4877–4884.
59. Hazel, P., Huppert, J., Balasubramanian, S. and Neidle, S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Amer. Chem. Soc.*, **126**, 16405–16415.
60. Olsen, C.M., Gmeiner, W.H. and Marky, L.A. (2006) Unfolding of G-quadruplexes: energetic, and ion and water contributions of G-tetrad stacking. *J. Phys. Chem. B*, **110**, 6962–6969.
61. Tang, C.F. and Shafer, R.H. (2006) Engineering the quadruplex fold: nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes. *J. Amer. Chem. Soc.*, **128**, 5966–5973.
62. de Lange, T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.*, **19**, 2100–2120.
63. Phan, A.T. and Mergny, J.L. (2002) Human telomeric DNA: G-quadruplex, i-motif and Watson–Crick double helix. *Nucleic Acids Res.*, **30**, 4618–4625.
64. Marathias, V.M. and Bolton, P.H. (1999) Determinants of DNA quadruplex structural type: sequence and potassium binding. *Biochemistry*, **38**, 4355–4364.
65. Risitano, A. and Fox, K.R. (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.
66. Rujan, I.N., Meloney, J.C. and Bolton, P.H. (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.
67. Ying, L., Green, J.J., Li, H., Klenerman, D. and 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.
68. Lee, J.Y., Okumus, B., Kim, D.S. and Ha, T. (2005) Extreme conformational diversity in human telomeric DNA. *Proc. Natl Acad. Sci. USA*, **102**, 18934–18943.
69. Zhan, N., Phan, A.T. and Patel, D.J. (2005) (3+1) assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. *J. Amer. Chem. Soc.*, **127**, 17277–17285.

70. Balagurumoorthy, P. and Brahmachari, S.K. (1994) Structure and stability of human telomeric sequence. *J. Biol. Chem.*, **269**, 21858–21869.
71. Li, J., Correia, J.J., Wang, L., Trent, J.O. and Chaires, J.B. (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.
72. Chang, C.C., Chu, J.F., Kao, F.J., Chiu, Y.C., Lou, P.J., Chen, H.C. and Chang, T.C. (2006) Verification of antiparallel G-quadruplex structure in human telomeres by using two-photon excitation fluorescence lifetime imaging microscopy of the 3,6-Bis(1-methyl-4-vinylpyridinium)carbazole diiodide molecule. *Anal. Chem.*, **78**, 2810–2815.
73. Xu, Y., Noguchi, Y. and Sugiyama, H. (2006) The new models of the human telomere d[AGGG(TTAGGG)₃] in K⁺ solution. *Bioorg. Med. Chem.*, **14**, 5584–5591.
74. Dapic, V., Abdomicovic, V., Marrington, R., Peberdy, J., Rodger, A., Trent, J.O. and Bates, P.J. (2003) Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. *Nucleic Acids Res.*, **31**, 2097–2107.
75. Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A. and 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.
76. Simonsson, T. (2001) G-quadruplex DNA structures--variations on a theme. *Biol. Chem.*, **382**, 621–628.
77. Paces, J., Pavlicek, A. and Paces, V. (2002) HERVd: database of human endogenous retroviruses. *Nucleic Acids Res.*, **30**, 205–206.
78. Kostadinov, R., Malhotra, N., Viotti, M., Shine, R., D'Antonio, L. and Bagga, P. (2006) GRSDb: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res.*, **34**, D119–D1124.
79. Zanotti, K.J., Lackey, P.E., Evans, G.L. and Mihaela-Rita, M. (2006) Thermodynamics of the fragile X mental retardation protein RGG box interactions with G tetrad forming RNA. *Biochemistry*, **45**, 8319–8330.
80. Eddy, J. and Maizels, N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.*, **34**, 3887–3896.
81. Raghavan, S.C., Swanson, P.C., Wu, X., Hsieh, C.-L. and Lieber, M.R. (2004) A non-B-DNA structure at the bcl-2 major breakpoint region is cleaved by the RAG complex. *Nature*, **428**, 88–93.
82. Bonaglia, M.C., Giorda, R., Mani, E., Aceti, G., Anderlid, B.M., Baroncini, A., Pramparo, T. and Zuffardi, O. (2005) Identification of a recurrent breakpoint within the SHANK3 gene in the 22q13.3 deletion syndrome. *J. Med. Genetics*, doi:10.1136/jmg.2005.038604.
83. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress *c-MYC* transcription. *Proc. Natl Acad. Sci. USA*, **99**, 11593–11598.
84. Liu, W., Sun, D. and Hurley, L.H. (2005) Binding of G-quadruplex-interactive agents to distinct G-quadruplexes induces different biological effects in MiaPaCa cells. *Nucleosides Nucleotides* *Nucleic Acids*, **24**, 1801–1815.
85. Shafer, R.H. and Smirnov, I. (2000) Biological aspects of DNA/RNA quadruplexes. *Biopolymers*, **56**, 209–227.
86. Simonsson, T. and Sjoback, R. (1999) DNA tetraplex formation studied with fluorescence resonance energy transfer. *J. Biol. Chem.*, **274**, 17379–17383.
87. Phan, A.T., Modi, Y.S. and Patel, D.J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human *c-myc* promoter. *J. Amer. Chem. Soc.*, **126**, 8710–8716.
88. Ambrus, A., Chen, D., Dai, J., Jones, R.A. and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human *c-myc* promoter. Implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048–2058.
89. Phan, A.T., Kuryavyi, V., Gaw, H.Y. and Patel, D.J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nature Chem. Biol.*, **1**, 167–173.
90. Rankin, S., Reszka, A.P., Huppert, J., Zloh, M., Parkinson, G.N., Todd, A.K., Ladame, S., Balasubramanian, S. and Neidle, S. (2005) Putative DNA quadruplex formation within the human *c-kit* oncogene. *J. Amer. Chem. Soc.*, **127**, 10584–10589.
91. Fernando, H., Reszka, A.P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A.R., Neidle, S. and Balasubramanian, S. (2006) A conserved quadruplex motif located in a transcription activation site of the human *c-kit* oncogene. *Biochemistry*, **45**, 7854–7860.
92. Sun, D., Guo, K., Rusche, J.J. and Hurley, L.H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.*, **33**, 6070–6080.
93. De Armond, R., Wood, S., Sun, D., Hurley, L.H. and Ebbinghaus, S.W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry*, **44**, 16341–16350.
94. Dai, J., Dexheimer, T.S., Chen, D., Carver, M., Ambrus, A., Jones, R.A. and Yang, D. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Amer. Chem. Soc.*, **128**, 1096–1098.
95. Dexheimer, T.S., Sun, D. and Hurley, L.H. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Amer. Chem. Soc.*, **128**, 5404–5415.
96. Luu, K.N., Phan, A.T., Kuryavyi, V., Lacroix, L. and Patel, D.J. (2006) Structure of the human telomere in K⁺ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J. Amer. Chem. Soc.*, **128**, 9963–9970.
97. Cogio, S. and Xodo, L.E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.*, **34**, 2536–2549.
98. Xu, Y. and Sugiyama, H. (2006) Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.*, **34**, 949–954.
99. Krishnan-Ghosh, Y., Stephens, E. and Balasubramanian, S. (2004) A PNA4 quadruplex. *J. Amer. Chem. Soc.*, **126**, 5944–5945.
100. Datta, B., Bier, M.E., Roy, S. and Armitage, B.A. (2005) Quadruplex formation by a guanine-rich PNA oligomer. *J. Amer. Chem. Soc.*, **127**, 4199–4207.
101. Randazzo, A., Esposito, V., Ohlenschlager, O., Ramachandran, R. and Mayola, L. (2004) NMR solution structure of a parallel LNA quadruplex. *Nucleic Acids Res.*, **32**, 3083–3092.
102. Nielsen, J.T., Arar, K. and Petersen, M. (2006) NMR solution structures of LNA (locked nucleic acid) modified quadruplexes. *Nucleic Acids Res.*, **34**, 2006–2014.
103. Matsugami, A., Okuizumi, T., Uesugi, S. and Katahira, M. (2003) Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(A):G(A):G(A):G heptad of GGA triplet repeat DNA. *J. Biol. Chem.*, **278**, 28147–28153.
104. Khateb, S., Weisman-Shomer, P., Hershco, L., Loeb, L.A. and Fry, M. (2004) Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)_n is mediated by homolog-conserved domains in three members of the hnRNP family. *Nucleic Acids Res.*, **32**, 4145–4154.
105. Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F. and Maizels, N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.*, **18**, 1618–1629.
106. Ju, B.G., Lunyak, V.V., Perissi, V., Garcia-Bassets, I., Rose, D.W., Glass, C.K. and Rosenfeld, M.G. (2006) A topoisomerase IIβ-mediated dsDNA break required for regulated transcription. *Science*, **312**, 1798–1802.
107. Rangan, A., Fedoroff, O.Y. and Hurley, L.H. (2001) Induction of duplex to G-quadruplex transition in the *c-myc* promoter region by a small molecule. *J. Biol. Chem.*, **276**, 4640–4646.