Published in final edited form as:

Methods Mol Biol. 2011 ; 735: 77–96. doi:10.1007/978-1-61779-092-8_8.

G-Quadruplex Structures and G-Quadruplex-Interactive Compounds

Raveendra I. Mathad and Danzhou Yang

Abstract

G-quadruplexes are noncanonical secondary structures formed in DNA sequences containing consecutive runs of guanines. DNA G-quadruplexes have recently emerged as attractive cancer therapeutic targets. It has been shown that the 3' G-rich single-stranded overhangs of human telomeres can form G-quadruplex structures. G-quadruplex-interactive compounds have been shown to inhibit telomerase access as well as telomere capping. Nuclear magnetic resonance (NMR) spectroscopy has been shown to be a powerful method in determining the G-quadruplex structures under physiologically relevant conditions. We present the NMR methodology used in our research group for structure determination of G-quadruplexes in solution and their interactions with small molecule compounds. An example of a G-quadruplex structure formed in the human telomere sequence recently solved in our laboratory is used as an example.

Keywords

Human telomeres; G-quadruplex structures; Structure polymorphism; G-quadruplex-interactive compounds; Telomerase inhibitors; Anticancer drug targets

1. Introduction

1.1. G-Quadruplex Structures

G-quadruplexes are noncanonical secondary structures formed in DNA sequences containing consecutive runs of guanines and have recently emerged as attractive cancer therapeutic targets (1). A G-quadruplex is a four-stranded DNA secondary structure that contains stacked G-tetrad planes of four guanines connected by a network of Hoogsteen hydrogen bonding (Fig. 1a). G-quadruplexes can be formed with one, two, or four G-rich strands (Fig. 1b). G-strands in a G-quadruplex structure can be parallel or antiparallel, and G-tetrad guanines can adopt *anti* or *syn* conformations around the glycosidic bonds depending on the orientation of DNA strands. G-tetrad guanines from parallel G-strands adopt the same glycosidic conformation, while those from antiparallel G-strands adopt the opposite glycosidic conformation (Fig. 1c). G-quadruplex formation requires monovalent cations, in particular K^+ and Na^+ , to coordinate with the eight electronegative O6 atoms of the adjacent stacked G-tetrads (2) (Fig. 1a). In general, K^+ is more preferred than Na^+ by G-quadruplex as K^+ has a better coordination with eight guanine O6s and a lower dehydration

¹¹Dialysis should be performed against at least 1,000 volumes of exchange solvent to obtain efficient change of solution conditions.

energy (3). The K^+ form is considered to be more biologically relevant due to the higher intracellular concentration of K^+ (~140 mM) than that of Na^+ (5–15 mM).

Intramolecular G-quadruplexes are of intensive current research interest due to their potential formation in telomeres and oncogene promoter regions (4–6). Intramolecular G-quadruplex structures form quickly in solution and are found to be DNA sequence specific, exhibiting great conformational diversity, such as folding topologies, loop conformations, and capping structures (1). High-field NMR spectroscopy has been shown to be a powerful method in determining the G-quadruplex structures under physiologically relevant conditions.

1.2. G-Quadruplexes Formed in Human Telomeres

The first biologically relevant G-quadruplex formation was observed in telomeric DNA (7). Human telomeres consist of tandem repeats of the hexanucleotide $d(TTAGGG)_n$ 5–25 kb in length, which terminate in a single-stranded 3'-overhang of 35–600 bases (8, 9). Telomeric DNA is extensively associated with various telosome proteins, and the structure and stability of telomeres are closely related with cancer (4), aging (10) and genetic stability (11). Vertebrate telomeric DNA repeats are highly conserved, which has been suggested to be related to their ability to form the DNA G-quadruplex (4). The most direct evidence of the *in vivo* existence of G-quadruplexes was established by using specific antibodies against parallel and antiparallel G-quadruplexes formed in telomeric DNA of the ciliate *Stylonychia lemnae* (12). Using the same antibody in ciliates, it was shown *in vivo* that the telomere end-binding proteins TEBP α and TEBP β are required to control G-quadruplex formation and that TEBP β phosphorylation is needed to resolve G-quadruplex structures during replication (13). A number of other proteins have been found to specifically interact with telomeric G-quadruplex structures (14). In addition, G-quadruplex formation was detected *in vivo* at human chromosomal ends by using the radiolabeled G-quadruplex ligand 360A (15) and the fluorescent G-quadruplex ligand BMVC (16).

In normal somatic cells, each cell replication results in a 50- to 200-base loss of the telomere. After a critical shortening of the telomeric DNA is reached, the cell undergoes apoptosis (17). However, telomeres of cancer cells do not shorten on replication, due to the activation of a reverse transcriptase, telomerase, that extends the telomeric sequence at the chromosome ends (18). Telomerase is activated in 80–85% of human cancer cells (19) and has been suggested to play a key role in maintaining the malignant phenotype (20). Formation of intramolecular G-quadruplex in human telomeric DNA induced by K^+ ion has been shown to inhibit the telomerase activity (21).

1.3. G-Quadruplex Structures in Human Telomeres

Intramolecular G-quadruplexes formed by single-stranded DNA in human telomeres are considered to be attractive anticancer drug targets due to their association with telomerase function and telomere end-capping (see Refs. 4, 6 for reviews) (Fig. 1d). Understanding of the human telomeric G-quadruplex structure under physiologically relevant K^+ solution has been the subject of intense investigation. A 22-nt human telomeric DNA 5'-AGGG(TTAGGG)₃ (wtTel22) (Fig. 2a) has been shown by NMR to form a basket-type

three-tetrad intramolecular G-quadruplex in Na⁺ solution (22); the same 22-nt telomeric sequence has been shown to form a parallel-stranded three-tetrad G-quadruplex in the crystalline state in the presence of K⁺ (23). However, wtTel22 does not appear to form a single G-quadruplex structure in K⁺ solution (Fig. 2b). More recent work from our lab (24–26) and others' (27–30) has shown that the major G-quadruplexes formed in human telomeres in K⁺ solution are the hybrid-type intramolecular structures, distinct from the Na⁺ solution structure or the K⁺ crystal structure (see ref. 31 for a review). Two hybrid-type structures appear to coexist in K⁺ solution, with the hybrid-2 form being the major one for the extended four-G-tract human telomeric DNA (25) (Fig. 1e). The kinetics of the interconversion between the two forms appears to be rather slow, as shown by NMR experiments (25). Both hybrid-type structures contain three parallel G-strands and one antiparallel G-strand, five *syn* guanines, and asymmetrical G-arrangements (Fig. 2c). They differ in loop arrangements, strand orientations, G-tetrad arrangements, and capping structures which may provide specific ligand binding sites (Fig. 1e) (25, 26). The hybrid-form structures provide an efficient scaffold for a compact-stacking structure of multimers in the 3' single-stranded overhang of the human telomeric DNA (Fig. 1d). While short loop sizes of 1 and 2 nt are usually favored for the double-chain-reversal loop conformation (32, 33), the presence of a 3-nt double-chain-reversal TTA loop in the hybrid-type telomeric G-quadruplexes allows the 5' and 3' ends to point in opposite directions (Figs. 1d and 2c). However, this may contribute to the structural flexibility and polymorphism of telomeric G-quadruplexes. Nature may have chosen the human telomeric sequence with the low energy difference between various forms (see our recent reviews for more information (6, 31)).

1.4. G-Quadruplex-Interactive Ligands

Recognition of the biological significance of DNA G-quadruplexes has intensified research and development of G-quadruplex-interactive compounds. Targeting of DNA G-quadruplex secondary structures represents an entirely new approach for cancer therapeutics, with the first report of targeting G-quadruplexes for inhibiting telomerase activity in 1997 (34). Formation of diverse G-quadruplex structures offers an opportunity to design small molecules/ligands that can bind different G-quadruplexes selectively. Structure-based drug design has been playing an important role in the development of G-quadruplex-interactive compounds. In fact, G-quadruplex inhibitors themselves have contributed immensely to understanding G-quadruplexes as a therapeutic target. Most G-quadruplex compounds are the stacking ligands, which consist of aromatic systems that can stack onto the terminal G-tetrads. Additionally, stacking ligands may contain positively charged side-chain substituents that interact with phosphate groups in G-quadruplex grooves. More recently, unfused aromatic ring systems that can adopt flexible conformations have been shown to bind into G-quadruplex grooves (35, 36). A number of small molecules that can specifically bind and stabilize telomeric G-quadruplexes have been developed, and some have become prospective anticancer agents that display relatively low cytotoxicity. For example, BRACO19 is a trisubstituted acridine derivative, rationally designed by considering the unique structural features of DNA G-quadruplexes (37). Telomestatin, one of the most potent G-quadruplex stabilizing compounds, has been shown to be a highly potent inhibitor of telomerase (38). TMPyP4, another G-quadruplex-interactive compound by structure-based rational drug design (39), has been shown to inhibit human telomerase in HeLa cell

extracts. A number of other G-quadruplex-interactive compounds, including BMVC, RHPS4, 360A, and 12459 have been shown to inhibit telomerase activity with various levels of potency. It is noteworthy that Quarfloxacin, a G-quadruplex-targeting compound originated from rational drug design and developed by Cylene Pharmaceuticals (San Diego, CA), shows excellent *in vivo* activity in various solid tumors and is currently in Phase II clinical trials (see recent review (1, 6) for more information).

In this chapter, we focus on practical steps involved in the NMR structural determination of G-quadruplexes formed in human telomeric DNA in K^+ solution. In addition, NMR analysis of drug interactions is also presented.

2. Materials

2.1. Reagents

1. Phosphoramidites, resins, and other reagents (Glen Research, Sterling, VA) (see Note 1).
2. ^{15}N Guanine phosphoramidites (Cambridge Isotopes Laboratories, Andover, MA).
3. D_2O and $DMSO-d_6$ solvents (Sigma-Aldrich, St. Louis, MO).
4. Phosphate buffer: 25 mM phosphate, 95 mM potassium, pH 7.0 (see Note 2).

2.2. NMR Sample Preparation

1. Lyophilized DNA samples are dissolved in aqueous solution containing 25 mM phosphate, 95 mM potassium, pH 7.0 (see Note 3).
2. For H_2O samples, DNA powder is dissolved in 0.6 ml of 90/10% H_2O/D_2O . For D_2O samples, DNA is dissolved in 0.6 ml of 99% D_2O .
3. The final NMR samples contain 0.1–3 mM DNA with a final K^+ concentration of 95 mM (see Note 4).
4. DNA oligonucleotides are annealed by heating to 95°C for 15 min, then slowly cooling to room temperature.
5. Drug stock solutions (10–40 mM) can be prepared in H_2O or $DMSO$, depending on their solubility (see Note 5).

3. Methods

NMR spectroscopy is widely employed for structural studies of G-quadruplexes under physiologically relevant conditions. Significantly, unlike long thread-like double-helical DNA structures, e.g., B-DNA, that lack global interaction information, global interaction information is well observed in G-quadruplex structures because of extensive DNA folding.

¹Phosphoramidites are stored at room temperature until dissolution in acetonitrile.

²The buffer for NMR sample preparation is stored at room temperature.

³We used potassium salt as it is considered to be the more physiologically relevant ion. Na^+ can also be used if needed.

⁴A low-concentration DNA sample is usually used for 1D experiments, while a high-concentration sample is needed for 2D NMR, particularly NOESY experiments. It needs to be confirmed that the same conformation of DNA is formed at all concentrations.

⁵For $DMSO$ stock solution, deuterated $DMSO-d_6$ is used to minimize the solvent peak in 1H -NMR spectra.

Thus, a G-quadruplex structure can be derived from extended single-stranded DNA conformation by NMR-based structure calculation, analogous to a globular protein with global folding that can be well defined by the NMR-restrained structure calculation.

3.1. Synthesis and Purification of DNA with and Without Labeling

1. The DNA oligonucleotides are synthesized at the 1 mmol scale on an Expedite™ 8909 Nucleic Acid System from Applied Biosystems (Foster City, CA) in DMT-on mode using solid-phase β -cyanoethylphosphoramidite chemistry (40) (see Note 6).
2. For labeled DNA oligonucleotide synthesis, ^{15}N -guanine phosphoramidite is used for site-specific ^{15}N -labeling of individual bases in DNA sequences at a 6% low-level enrichment. This low enrichment is sufficient to perform special NMR experiments such as JR-HMQC (see Note 7).
3. The oligonucleotides are cleaved from the solid support by treatment with concentrated ammonia for 15 h at 55°C (see Note 8).
4. DMT-on oligonucleotides are purified by reverse-phase high-performance liquid chromatography (RP-HPLC) on a C_{18} column (250 $\times$ 10.0 mm VARIAN Dynamax Microsorb 300-10) with a linear gradient of 5–45% acetonitrile over 45 min in 1% triethylammonium acetate (TEAA), pH 7.0, with a flow rate of 3 ml/min. DNA is monitored with UV at 254 nm. The DMT-on DNA is collected at 25 min and is then lyophilized.
5. DMT protection groups are removed from DNA by treating with 2 ml of 80% acetic acid for 2 h. The sample is then diluted with 1 ml of water and extracted with 10 ml of ether to remove DMT (see Note 9). The aqueous phase containing DNA is collected and concentrated on a Speedvac.
6. The concentrated DMT-off DNA samples are purified by reverse-phase HPLC on the C_{18} column with a linear gradient of 5–30% acetonitrile over 30 min in 1% TEAA, pH 7.0, with a flow rate of 3 ml/min. DMT-off DNA is collected at 10 min.
7. After HPLC purification, the DMT-off DNA samples are further processed by sequential dialysis in glass beakers with water, low salt (10 mM NaCl), high salt (150 mM NaCl), and water (see Notes 10–12). A weak base (10 mM NaOH) can be used to remove persistent secondary structures. The purified and dialyzed samples are lyophilized to a homogeneous powder (see Note 13).

⁶DNA can also be synthesized at a larger scale, e.g., at 15 mmol; however, the yield is anticipated to be lower for the larger scale synthesis.

⁷The level of enrichment can vary depending on resonances. An enrichment lower than 6% can be used for detecting resonances with high intensity, while a higher level of enrichment is needed for detecting weaker resonances.

⁸Cleavage of blocking groups from synthesized oligonucleotides with ammonium hydroxide can also be done at 62°C for 10 h.

⁹Careful extraction of acetic acid-treated oligonucleotides with fresh ether is necessary to obtain the high-quality final material. Presence of cloudiness in the extraction process is an indication of unsuccessful extraction. Addition of 1 ml of high-purity water increases the aqueous phase volume, leading to better phase separation and better recovery of DNA products.

¹⁰Dialysis in glass beakers is necessary to prevent the leaching of chemicals while stirring in plastic beakers, which produces contaminant peaks in NMR spectra.

¹²Oligonucleotides greater than 20-mer in lengths can be dialyzed in tubing with 3,000–3,500 MWCO; shorter oligonucleotides should be dialyzed in tubing with a MWCO of 1,000.

3.2. Determination of the G-quadruplex Formation by 1D ¹H-NMR

NMR spectroscopy has a unique strength in studying DNA G-quadruplex secondary structures. The imino peaks associated with guanines in G-tetrad formation give rise to characteristic chemical shifts around 10.5–12 ppm. This chemical shift region is completely separated from imino chemical shifts from any other DNA conformations, such as duplex DNA, single-stranded DNA, or other secondary DNA structures. Therefore, the imino regions of G-quadruplex guanines provide a direct and clear monitoring system for not only the formation of a G-quadruplex structure but also its drug binding interactions.

1. As a first step in the NMR structure determination process, it is important to determine the G-quadruplex formation, sample purity and line widths by ¹H-NMR spectra in solution. Formation of G-quadruplexes can be identified by chemical shifts of imino H1 protons in ¹H-NMR in H₂O solution. The appearance of imino peaks in the downfield region 10.5–12 ppm indicates their involvement in a Hoogsteen hydrogen-bonded network and is characteristic for the formation of G-quadruplex. The number of imino resonances corresponds to the number of G-tetrad-associated guanines present in the system (see Note 14).
2. For samples in aqueous solution, the ¹H-NMR experiments are performed with water suppression techniques such as Watergate or Jump-and-Return, and the relaxation delay is set to 3 s.

3.3. Sequence Design and Selection

It is important to note that the intramolecular G-quadruplex structure formed in a G-rich sequence can be complex. Not only do different sequences adopt different structures, but also a given sequence can fold into a variety of different conformations, as in the case of the human telomeric DNA sequence (see below). A DNA sequence that forms a major stable structure with good NMR spectral properties in solution, i.e., sharp line widths, is necessary for successful structure determination by NMR. NMR and sequence analysis are unique in being able to identify the major stable G-quadruplex structures under physiologically relevant ion conditions (pH 7, 100 mM KCl).

1. The 22-nt human telomeric DNA 5'-AGGG(TTAGGG)₃ (wtTel22, Fig. 2a), which has been shown to form a basket-type G-quadruplex in Na⁺ solution (22) or a parallel-stranded intramolecular G-quadruplex in the crystalline state in the presence of K⁺ (23), was examined by 1D ¹H NMR (Fig. 2b top). The 1D ¹H NMR shows that wtTel22 does not form a single G-quadruplex structure in K⁺ solution.
2. We have screened a large number of variant four G-tract sequences containing the core wtTel22 with different flanking segments by ¹H NMR (24). A few sequences, including the 26-nt sequence Tel26 (Fig. 2a), were found to form a major hybrid-1 type intramolecular G-quadruplex in K⁺ solution. The 1D ¹H NMR spectrum of the

¹³Lyophilization is a critical step for DNA product quality. All materials should be hard-frozen prior to lyophilization and dried completely before removal from the lyophilizer at all stages. Lyophilized oligonucleotides are stored under desiccation at –20°C until dissolution.

¹⁴It is also important to determine the presence of multiple conformations. From the number of imino peaks arising from a particular sequence, it is possible to determine whether there are multiconformational species in solution and their relative populations.

Tel26 sequence in K^+ solution (Fig. 2b) shows a major intramolecular G-quadruplex structure with 12 resolved imino proton resonances between 10.5 and 12 ppm (see Note 15). The folding topology and molecular structure of the hybrid-1 structure was determined by NMR (PDB ID 2HY9) (24, 26) (Figs. 2c and 1e).

3. The wtTel26 sequence was found to form a major hybrid-2 type intramolecular G-quadruplex in K^+ solution. The folding topology and molecular structure of this hybrid-2 structure was determined by NMR (PDB ID 2JPZ) (25) (Figs. 2c and 1e).
4. CD (Circular Dichroism) can be used as a complementary method to determine the G-quadruplex formation. Antiparallel-stranded G-quadruplexes have a CD spectrum characterized by a positive ellipticity maximum at 295 nm and a negative minimum at 265 nm, while the parallel-stranded ones have a positive maximum at 264 nm and a negative minimum at 240 nm (31).

3.4. Assignment of Guanine Base Resonances Using Labeled DNA

Only sequences showing well-resolved NMR spectra are subjected to further structural investigation. Herein, we use the telomeric DNA sequence wtTel26 as an example for structure determination by NMR spectroscopy. NMR structure determination is composed of NMR resonance assignment and NMR-restrained structure calculation. NMR resonance assignment of G-quadruplexes includes guanine base proton assignment using site-specific ^{15}N -labeling and complete spectral assignment using 2D NMR methods. After the complete spectral assignment, the folding topology of a G-quadruplex is first determined, followed by three-dimensional structure determination using NMR-derived distance restraints in combination with MD calculations.

1. The 1D 1H NMR spectrum of the wtTel26 sequence in K^+ solution (Fig. 2b second top) shows a major intramolecular G-quadruplex structure with 12 resolved imino proton resonances between 10.5 and 12 ppm, indicating the formation of G-quadruplex with three G-tetrads (see Note 16).
2. The imino H1 and base aromatic H8 protons of 12 guanine residues can be unambiguously assigned by site-specific low enrichment (6%) using ^{15}N -labeled guanine nucleoside at each guanine position in the wtTel26 sequence. The guanine imino H1 and aromatic H8 proton resonances have one-bond coupling to N1 and two-bond coupling to N7, respectively (Fig. 1a). Both H1 and H8 protons of the site-specific labeled guanine are readily detected by 1D ^{15}N -filtered JR-HMQC experiments (Fig. 2d) (25).

¹⁵Tel26 was found to form two well-defined G-quadruplex conformations when freshly dissolved in K^+ solution, as indicated by two separate sets of relatively sharp guanine imino peaks (Fig. 2b bottom). One conformation (~40%) slowly converts to the other (~60%) overnight, and the complete conversion takes about a day (Fig. 2b). This observation led to the careful examination of the native 26-nt human telomeric sequence wtTel26, (TTAGGG)₄TT (Fig. 2a).

¹⁶However, this major conformation only accounts for ~70% of the total population, with about 30% population of minor conformations shown as weak and broader resonances. Thus, it is a more challenging process to get a complete resonance assignment for wtTel26 and a large number of different conditions are needed (25).

3. The assignment of guanine H8 protons can be further confirmed by long-range through-bond correlations with the already assigned imino H1 via $^{13}\text{C}5$, using 2D ^1H - ^{13}C HMBC experiment at natural abundance.

3.5. Assignment of Thymine/Cytosine Base Resonances

1. Thymine residues can be identified by a strong cross-peak between H6 and Me in COSY/TOCSY and NOESY spectra.
2. Cytosine residues can be identified by a strong cross-peak between H5 and H6 in COSY/TOCSY and NOESY spectra.
3. Base H6 protons of thymines or cytosines in a sequence can be unambiguously assigned by site-specific substitution for each thymine with deoxyuridine (dU) or for each cytosine with dT or dU one at a time (see Note 17).

3.6. Complete Spectral Assignment

1. After the site-specific assignment of base protons, the complete assignment of the base and sugar protons of a DNA sequence can be accomplished with 2D-NMR experiments (Fig. 3). Standard homonuclear 2D-NMR experiments are used to assign the nonexchangeable proton chemical shifts of the DNA, including DQF-COSY, TOCSY, and NOESY. The mixing times are set at 50, 100, 150, and 200 ms for NOESY and at 40 and 80 ms for TOCSY at various temperatures. The exchangeable proton chemical shifts are assigned for H_2O samples using 2D-NOESY experiments with WATERGATE or jump-return (NOE11) water suppression techniques. The relaxation delay is set to 2 s. The acquisition data points are set to $4,096 \times 512$. The 60° shifted sine bell functions are applied to both dimensions of NOESY and TOCSY spectra. The fifth-order polynomial functions are employed for the baseline corrections.
2. Assignment of base protons H8/H6 (see Subheadings 3.4 and 3.5) gives rise to the assignment of H1–H1 (Fig. 3a) and H1–H8 crosspeaks (Fig. 3b), which can be used for the determination of the folding topology of a G-quadruplex (see Subheading 3.8).
3. The assignment of guanine H1 and H8 protons can give rise to the assignment of the base proton H2/H8 of adenine residues that stack with the G-tetrads (Fig. 3b).
4. A standard DNA sequential assignment procedure is used for the assignment of nonexchangeable protons (Fig. 3c). The assignment of the aromatic H8 can lead to the direct assignment of H1' and H2'/H2'' and then can be extended to H3', H4', and H5'/H5''.
5. Sequential NOEs can be assigned for the neighboring residues in a DNA sequence. For example, $G_n\text{H}8\text{--}G_{n-1}$ H1'/H2'/H2'' NOEs allow the assignment of sequential guanines in the DNA (Fig. 3c).

¹⁷Site-specific substitution for adenines with inosines (dI) can also be used for the assignment of adenine base protons.

6. The glycosidic torsion angles (*syn/anti*) of G-tetrad guanines can be determined by the intensity of the intraresidue H8-H1' NOEs in NOESY spectra with different mixing times. Observation of a weak intraresidue H8-H1' NOE at a short mixing time indicates an *anti* conformation, while a strong H8-H1' NOE indicates a *syn* conformation (Fig. 3c). In addition, a characteristic downfield shift is observed for H2'/H2'' of guanines that are in *syn* conformation. In wtTel26, there are five *syn* tetrad guanines out of a total of 12 tetrad guanines (Fig. 2c), clearly shown in the NOESY spectrum (Fig. 3c).
7. Based on the complete spectral assignment, critical interresidue NOE interactions can be schematically drawn as in Fig. 3d. This schematic diagram of NOE interactions clearly defines the overall shape of the hybrid-2 G-quadruplex formed in wtTel26 in K⁺.

3.7. ³¹P Heteronuclear NMR

1. ³¹P heteronuclear NMR experiments can be used to refine NMR structure calculation, particularly for loop residues with unusual conformations. ³¹P-¹H correlation experiments provide useful information for torsion angles: ³¹P-H5'/H5'' coupling constants define β, while ³¹P-¹H couplings give information about ε torsions (33). All ³¹P NMR spectra are collected with an external standard of 85% H₃PO₄. The experiments include the 1D ¹H-decoupled ³¹P spectrum, 2D-heteronuclear ³¹P-¹H Correlation Spectroscopy (COSY), and Heteronuclear Single Quantum Correlation Spectroscopy (HSQC).

3.8. Determination of G-Quadruplex Folding Topology

The characteristic imino H1 and base H8 connectivity pattern lead to the direct determination of the folding topology of a G-quadruplex.

1. The G-tetrad alignments and folding topology of a sequence are determined by the interresidual H1 and H8 NOEs in exchangeable NOESY spectra (Fig. 3a and b). In a G-tetrad plane, the imino proton H1 of each guanine is in close spatial vicinity to the H1s of the two adjacent guanines, and to the base H8 of one of the adjacent guanines due to the Hoogsteen H-bonded network (Fig. 1a). The formation of three G-tetrads, namely (G4-G12-G16-G22, top), (G5-G11-G17-G23, middle), and (G6-G10-G18-G24, bottom), can be determined for wtTel26.
2. The overall G-quadruplex alignment can be further defined based on the inter-tetrad NOE connections from residues that are positioned far apart in the DNA sequence. The strong NOE interactions of G5H1/G12H1, G11H1/G16H1, and G17H1/G22H1 (Fig. 3a) connect the top two G-tetrads and define their reversed G-arrangements. The sequential inter-tetrad NOE interactions between H1 (Fig. 3a) and H8 proton indicate the same G-conformations of the bottom two G-tetrads. From these data, the folding topology of the G-quadruplex formed in wtTel26 can be unambiguously determined (Fig. 2c).

3.9. NMR-Restrained Structure Determination

1. For peak assignments and NOE volume integration (peak fitting function), the software Sparky (UCSF) is used (41). Distances between nonexchangeable protons are estimated based on the NOE cross-peak volumes at 50–300 ms mixing times, with the upper and lower boundaries assigned to $\pm 20\%$ of the estimated distances. Distances between exchangeable protons are assigned with looser boundaries (± 1 – 1.5 Å). The thymine base proton Me–H6 distance (2.99 Å) is used as a reference for wtTel26 (see Note 18). The distances involving the unresolved protons, e.g. methyl protons, are assigned using pseudo-atom notation to make use of the pseudo-atom correction automatically computed by X-PLOR (42). Only relatively isolated peaks are used for NOE-restrained structure calculation.
2. Metric matrix distance geometry (MMDG) calculations are carried out using an X-PLOR protocol to produce and optimize 100 initial structures. An arbitrary global conformation of G-quadruplex is used as a starting model, as a G-quadruplex contains enough global fold to derive a final NMR structure. Substructure embedding is performed to produce a family of 100 DG structures. The embedded DG structures are then subjected to simulated annealing.
3. All of the 100 conformations generated from the DGSA calculations are subjected to NOE-restrained Simulated Annealing refinement in X-PLOR with a distance-dependent dielectric constant. Hydrogen bonds in the G-tetrad planes are restrained with distances corresponding to ideal hydrogen bond geometry. Each individual hydrogen bond is restrained using two distance restraints (heavy atom–heavy atom and heavy atom–proton). The force constants are scaled at 30 and 100 kcal/mol/Å² for distance restraints and hydrogen bond constraints respectively. All possible intra- and interresidue NOE-distance restraints, together with backbone dihedral torsions and hydrogen-bond restraints for the G-tetrads, are incorporated into the structure calculation.
4. NOE-restrained simulated annealing refinement calculations are initiated at 300 K. The temperature is gradually increased to 1,000 K in 4 ps, equilibrated at 1,000 K for 20 ps, and then slowly cooled to 300 K in 10 ps. The 20 best conformations are selected based both on their minimal energy terms and number of NOE violations and are further subjected to NOE-restrained MD calculations. The resulting averaged structure is subjected to minimization. The ten best molecules are selected, based both on their minimal energy terms and on number of NOE violations, as representative solution structure of G-quadruplex (25).

3.10. G-Quadruplex-Interactive Compounds

NMR spectroscopy is an indispensable technique to obtain high-resolution structural information about biologically important molecules and their interaction with ligands in solution. The appropriate sizes and the globular folding of DNA G-quadruplexes offer an excellent molecular system for NMR structural study. In addition, the well-separated

¹⁸When applicable, the cytosine base proton H5–H6 distance (2.45 Å) is normally used as a reference.

characteristic imino regions of G-quadruplex guanines provide a direct and unambiguous detection system for drug binding interactions. Furthermore, NMR allows a direct analysis of drug interactions with G-quadruplexes such as binding stoichiometry and binding kinetics, with direct monitoring of solution conditions such as temperatures and salt concentrations.

1. To study drug interactions of one particular G-quadruplex, the DNA sequence that forms a single major structure should be used (see Note 19). This sequence should have good NMR spectral properties, thus its drug interactions can be unambiguously characterized. For example, to study drug binding of the hybrid-1 telomeric G-quadruplex, the Tel26 sequence can be used; to study drug binding of the hybrid-2 telomeric G-quadruplex, the wtTel26 sequence can be used (Fig. 2a). The molecular structure of the G-quadruplex formed in the free DNA needs to be first determined and serves as the structural basis for studying its drug interactions.
2. 1D ^1H NMR titration experiments are used to monitor drug binding interactions. 1D NMR spectra of free drug and free DNA will be collected and used as references. Here we use wtTel26 as an example.
3. The purity of the drug is checked using ^1H -NMR or using analytical-HPLC. Impurities can cause additional peaks in ^1H -NMR spectra, which may overlap with the desired peaks or may even bind to wtTel26. The purity of the drug should be >95%.
4. For drug binding studies, concentrated drug stock solutions (20–40 mM) are prepared in H_2O or DMSO-d_6 (see Note 20).
5. The DNA (0.1–0.5 mM) solution containing 25 mM phosphate, 95 mM potassium, pH 7.0 is checked for G-quadruplex formation by ^1H -NMR.
6. The DNA–drug complex is prepared by step-by-step additions of a small volume of concentrated drug stock solution into DNA sample to achieve DNA–drug ratios of 1:0.5, 1:1, 1:2, and 1:3 (see Note 21). The complex sample is annealed if needed.
7. The change in chemical shifts of DNA–drug complex at various drug equivalences is monitored by 1D ^1H -NMR spectra (Fig. 4). The imino proton regions of tetrad-guanines are well separated from nonexchangeable protons and thus can be used in monitoring drug binding interactions and line width changes in titration profiles.
8. In general, the effects of a drug on a G-quadruplex are readily interpreted by NMR titration methods. A drug that binds specifically to a particular DNA sequence with sufficient affinity will produce an NMR spectrum with well-resolved peaks. In the case of a slow exchange binding regime, two sets of peaks of the free DNA and the bound DNA can both be observed at a lower drug equivalence, e.g., 0.5; and

¹⁹In some cases, the wild-type DNA sequence that forms multiple conformations is preferred for examining specific drug bindings. A drug that binds to one specific conformation may be able to shift the equilibrium of the multiple forms.

²⁰Highly concentrated drug stock solutions allow the use of small volume additions of drug solutions in the DNA–drug complex, so that the solvent effect is negligible.

²¹The DNA–drug ratio of 1:0.5 is important for drug binding studies: for a slow exchange binding regime, two sets of peaks of the free DNA and the bound DNA will be observed, while for a medium-to-fast exchange binding regime, only one set of peaks will be observed.

signals from the free DNA disappear once all the binding sites are occupied. In the case of a medium-to-fast exchange binding regime, only one set of peaks are observed for DNA and drug. In such cases, although the complex structure cannot be determined, some information on binding sites can be obtained. Compounds that do not bind tightly/specifically are readily discerned, as they do not lead to any shifts or cause spectral line broadening.

9. If a drug–DNA complex gives a well-resolved ^1H -NMR spectrum, the binding site(s) or mode(s) can be determined. For example, a stacking compound (with either end of the G-quadruplex) will affect the chemical shifts of the guanine imino protons from the end G-tetrad with which the drug stacks, while an intercalating drug will shift the guanine imino resonances of the two G-tetrads between which the drug intercalates. For the groove-binding drugs, the protons located in the groove region, such as sugar protons and G's H8, rather than the imino protons, will be shifted upon drug binding. The shifting of base protons' chemical shifts will also be different for different binding modes.
10. Once the complex system is optimized, the detailed structure determination can be carried out by following the steps in Subheading 3.

Acknowledgments

This research was supported by the National Institutes of Health funding (1S10RR16659 and CA122952). We thank Dr. Megan Carver for proofreading the manuscript.

References

1. Yang D, Okamoto K. Structural Insights into G-quadruplexes: Towards New Anti-Cancer Drugs. *Fut Med Chem*. 2010 In press.
2. Sen D, Gilbert W. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*. 1990; 344:410–414. [PubMed: 2320109]
3. Hud, NV.; Plavec, J. The role of cations in determining quadruplex structure and stability. In: Neidle, S., editor. *Quadruplex Nucleic Acids*. Royal Society of Chemistry, RSC Publishing; Cambridge: 2006. p. 100-130.
4. Neidle S, Parkinson G. Telomere maintenance as a target for anticancer drug discovery. *Nature Reviews Drug Discovery*. 2002; 1:383–393. [PubMed: 12120414]
5. Qin Y, Hurley LH. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*. 2008; 90:1149–1171. [PubMed: 18355457]
6. Punchihewa, C.; Yang, DZ. Therapeutic Targets and Drugs-G-quadruplex inhibitors. In: Hiyama, K., editor. *Telomeres and Telomerase in Cancer*. Springer; NJ, USA: 2009. p. 251-280.
7. Henderson E, Hardin CC, Walk SK, Tinoco I Jr, Blackburn EH. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell*. 1987; 51:899–908. [PubMed: 3690664]
8. Moyazis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, Meyne J, Ratliff RL, Wu JR. A highly conserved repetitive DNA sequence (TTAGGG) $_n$, present at the telomere of human chromosomes. *Proc Natl Acad Sci USA*. 1988; 85:6622–6626. [PubMed: 3413114]
9. Wright WE, Tesmer VM, Huffman KE, Levene SD, Shay JW. Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev*. 1997; 11:2801–2809. [PubMed: 9353250]

10. Bodnar AG, Ouellette M, Frolkis M, Holt SE, Chiu CP, Morin GB, Harley CB, Shay JW, Lichtsteiner S, Wright WE. Extension of life-span by introduction of telomerase into normal human cells. *Science*. 1998; 279:349–352. [PubMed: 9454332]
11. Sun D, Lopez C, Von Hoff DD, Hurley LH. Modulation of the catalytic activity and processivity of human telomerase from HeLa cell. *Proc Am Assoc Cancer Res*. 1998; 39:569.
12. Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps HJ, Pluckthun A. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc Natl Acad Sci USA*. 2001; 98:8572–8577. [PubMed: 11438689]
13. Paeschke K, Simonsson T, Postberg J, Rhodes D, Lipps HJ. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat Struct Mol Biol*. 2005; 12:847–854. [PubMed: 16142245]
14. Oganessian L, Bryan TM. Physiological relevance of telomeric G-quadruplex formation: a potential drug target. *Bioessays*. 2007; 29:155–165. [PubMed: 17226803]
15. Granotier C, Pennarun G, Riou L, Hoffschir F, Gauthier LR, De Cian A, Gomez D, Mandine E, Riou JF, Mergny JL, Mailliet P, Dutrillaux B, Boussin FD. Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res*. 2005; 33:4182–4190. [PubMed: 16052031]
16. Chang CC, Kuo IC, Lin JJ, Lu YC, Chen CT, Back HT, Lou PJ, Chang TC. A novel carbazole derivative, BMVC: A potential antitumor agent and fluorescence marker of cancer cells. *Chemistry & Biodiversity*. 2004; 1:1377–1384. [PubMed: 17191915]
17. Harley CB, Futcher AB, Greider CW. Telomeres shorten during ageing of human fibroblasts. *Nature*. 1990; 345:458–460. [PubMed: 2342578]
18. Greider CW, Blackburn EH. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell*. 1985; 43:405–413. [PubMed: 3907856]
19. Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PLC, Coviello GM, Wright WE, Weinrich SL, Shay JW. Specific Association of Human Telomerase Activity with Immortal Cells and Cancer. *Science*. 1994; 266:2011–2015. [PubMed: 7605428]
20. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000; 100:57–70. [PubMed: 10647931]
21. Zahler AM, Williamson JR, Cech TR, Prescott DM. Inhibition of Telomerase by G-Quartet DNA Structures. *Nature*. 1991; 350:718–720. [PubMed: 2023635]
22. Wang Y, Patel DJ. Solution Structure of the Human Telomeric Repeat D[Ag(3)(T(2)Ag(3))3] G-Tetraplex. *Structure*. 1993; 1:263–282. [PubMed: 8081740]
23. Parkinson GN, Lee MPH, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*. 2002; 417:876–880. [PubMed: 12050675]
24. Ambrus A, Chen D, Dai JX, Bialis T, Jones RA, Yang DZ. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res*. 2006; 34:2723–2735. [PubMed: 16714449]
25. Dai JX, Carver M, Punchihewa C, Jones RA, Yang DZ. Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K⁺ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res*. 2007; 35:4927–4940. [PubMed: 17626043]
26. Dai JX, Punchihewa C, Ambrus A, Chen D, Jones RA, Yang DZ. Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res*. 2007; 35:2440–2450. [PubMed: 17395643]
27. Xu Y, Noguchi Y, Sugiyama H. The new models of the human telomere d[AGGG(TTAGGG)(3)] in K⁺ solution. *Bioorg Med Chem*. 2006; 14:5584–5591. [PubMed: 16682210]
28. Luu KN, Phan AT, Kuryavyi V, Lacroix L, Patel DJ. Structure of the human telomere in K⁺ solution: An intramolecular (3+1) G-quadruplex scaffold. *J Am Chem Soc*. 2006; 128:9963–9970. [PubMed: 16866556]
29. Phan AT, Kuryavyi V, Luu KN, Patel DJ. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K⁺ solution. *Nucleic Acids Res*. 2007; 35:6517–6525. [PubMed: 17895279]
30. Phan AT, Luu KN, Patel DJ. Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K⁺ solution. *Nucleic Acids Res*. 2006; 34:5715–5719. [PubMed: 17040899]

31. Dai JX, Carver M, Yang DZ. Polymorphism of human telomeric quadruplex structures. *Biochimie*. 2008; 90:1172–1183. [PubMed: 18373984]
32. Ambrus A, Chen D, Dai JX, Jones RA, Yang DZ. Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry*. 2005; 44:2048–2058. [PubMed: 15697230]
33. Dai JX, Chen D, Jones RA, Hurley LH, Yang DZ. NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res*. 2006; 34:5133–5144. [PubMed: 16998187]
34. Sun DY, Thompson B, Cathers BE, Salazar M, Kerwin SM, Trent JO, Jenkins TC, Neidle S, Hurley LH. Inhibition of human telomerase by a G-quadruplex-interactive compound. *J Med Chem*. 1997; 40:2113–2116. [PubMed: 9216827]
35. White EW, Tanious F, Ismail MA, Reszka AP, Neidle S, Boykin DW, Wilson WD. Structure-specific recognition of quadruplex DNA by organic cations: Influence of shape, substituents and charge. *Biophys Chem*. 2007; 126:140–153. [PubMed: 16831507]
36. Muller S, Pantos GD, Rodriguez R, Balasubramanian S. Controlled-folding of a small molecule modulates DNA G-quadruplex recognition. *Chem Commun*. 2009:80–82.
37. Read M, Cuesta J, Basra I, Harrison J, Reszka A, Gowan S, Kelland LR, Neidle S. Rational design approaches to increase the potency of G-quadruplex-mediated telomerase inhibitors. *Clin Cancer Res*. 2001; 7:713.
38. Shin-ya K, Wierzba K, Matsuo K, Ohtani T, Yamada Y, Furihata K, Hayakawa Y, Seto H. Telomestatin, a novel telomerase inhibitor from *Streptomyces anulatus*. *J Am Chem Soc*. 2001; 123:1262–1263. [PubMed: 11456694]
39. Wheelhouse RT, Han FX, Sun D, Hurley LH. The interaction of telomerase inhibitory porphyrines with G-quadruplex DNA. *Proc Am Assoc Cancer Res*. 1998; 39:430.
40. Dai JX, Punchihewa C, Mistry P, Ooi AT, Yang DZ. Novel DNA Bis-intercalation by MLN944, a potent clinical bisphenazine anticancer drug. *J Biol Chem*. 2004; 279:46096–46103. [PubMed: 15317822]
41. Goddard TD, Kneller DG. University of California, San Francisco. 2004
42. Brünger, AT. Version 3.1: A system for X-ray crystallography and NMR. Yale University press; New Haven, CT, USA: 1993.

Fig. 1.

(a) Schematic diagram of a G-tetrad containing a square planar alignment of four guanines connected by cyclic Hoogsteen hydrogen bonding between the N1, N2 and O6, N7 of guanine bases. Curved lines show the H1-H1 and H1-H8 connectivity pattern detectable in NOESY experiments. (b) Schematic diagrams of monomeric (intramolecular), dimeric, and tetrameric G-quadruplexes. (c) The guanines in a G-tetrad can adopt either syn or anti glycosidic conformation depending on the orientation of DNA strands. (d) A schematic model of human telomeres composed of compact-stacking multimers of DNA G-quadruplex secondary structure, and the proposed mechanism of drug-mediated inhibition of telomerase and interference of telomere capping by G-quadruplex-stabilizing compounds. (e) Two hybrid-type intramolecular G-quadruplex structures coexist in human telomeres in K^+ solution. The molecular structures are determined by NMR in pH 7.0, 95 mM K^+ solution.

Fig. 2.

(a) Human telomeric DNA sequences containing four G-tracts that were used for structure determination. The sequence numbering system is shown above wtTel26. The major conformations formed in each sequence are specified. (b) The imino regions of 1D ^1H -NMR spectra of wtTel22, wtTel26, and Tel26 (1 week and 1 h samples) in K^+ solution. The assignment of imino protons is shown for wtTel26 and Tel26. (c) Schematic drawing of the folding topologies of the hybrid-1 and hybrid-2 intramolecular telomeric G-quadruplexes in K^+ solution. Dark box=(syn) guanine, light box=(anti) guanine. (d) The imino proton region with the assignment of the 1D ^1H NMR spectrum of wtTel26 in K^+ solution (top), and imino proton assignments of wtTel26 using 1D ^{15}N -filtered experiments on site-specific labeled oligonucleotides. Conditions: 25 mM phosphate, 95 mM potassium, pH 7.0, 25°C, 0.5–0.6 mM DNA.

Fig. 3.

The expanded H1-H1' (a) and H1-H8/H2/H6 (b) regions of the exchangeable proton 2D-NOESY spectrum of wtTel26 in K⁺ solution. (c) The expanded H8/H6-H1' region of the nonexchangeable 2D-NOESY spectrum of wtTel26 in K⁺ solution. The sequential assignment pathway is shown. The H8-H1' NOEs of the guanines with *syn* conformation are labeled with residue names. Missing connectivities are labeled with *asterisks*. The characteristic G(i)H8/G(i+1)H1' NOEs for the *syn* G(i)s are labeled by *arrows*. Conditions: 15°C, 25 mM phosphate, 95 mM potassium, pH 7.0, 2.5 mM DNA. (d) Schematic diagram

of interresidue NOE connectivities of wtTel26 G-quadruplex formed in K⁺ solution. The guanines in *syn* conformation are represented using *gray circles*.

Fig. 4. The 1D ^1H NMR titration profiles of wtTel26 with a small molecule compound at different ratios. Conditions: 25°C, 0.2 mM DNA, 25 mM phosphate, 95 mM potassium, pH 7.0.