Solution structure of a DNA quadruplex containing ALS and FTD related GGGGCC repeat stabilized by 8-bromodeoxyguanosine substitution

Jasna Brčić¹ and Janez Plavec¹,²,³,*

¹Slovenian NMR Center, National Institute of Chemistry, Hajdrihova 19, SI-1000, Ljubljana, Slovenia, ²Faculty of Chemistry and Chemical Technology, University of Ljubljana, Ljubljana, Slovenia and ³EN-FIST Center of Excellence, Ljubljana, Slovenia

Received May 12, 2015; Revised July 29, 2015; Accepted July 30, 2015

ABSTRACT

A prolonged expansion of GGGGCC repeat within non-coding region of C9orf72 gene has been identified as the most common cause of familial amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), which are devastating neurodegenerative disorders. Formation of unusual secondary structures within expanded GGGGCC repeat, including DNA and RNA G-quadruplexes and R-loops was proposed to drive ALS and FTD pathogenesis. Initial NMR investigation on DNA oligonucleotides with four repeat units as the shortest model with the ability to form an unimolecular G-quadruplex indicated their folding into multiple G-quadruplex structures in the presence of K⁺ ions. Single dG to 8Br-dG substitution at position 21 in oligonucleotide d[(G4C2)3G4] and careful optimization of folding conditions enabled formation of mostly a single G-quadruplex species, which enabled determination of a high-resolution structure with NMR. G-quadruplex structure adopted by d[(G4C2)3GGBrGG] is composed of four G-quartets, which are connected by three edge-wise C-C loops. All four strands adopt antiparallel orientation to one another and have alternating syn-anti progression of glycosidic conformation of guanine residues. One of the cytosines in every loop is stacked upon the G-quartet contributing to a very compact and stable structure.

INTRODUCTION

Guanine-rich nucleic acids can form in the presence of cations such as K⁺ or Na⁺ non-canonical four-stranded structures called G-quadruplexes, which are composed of stacked layers of G-quartets, formed by four guanine residues connected by Hoogsteen-type hydrogen bonds. The distribution of G-rich sequences in the human genome is not random and their overrepresentation in the 5′-ends of human genes (1) suggests that G-quadruplex formation may influence gene expression (2,3). Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are related neurodegenerative diseases that share a common genetic and pathological background (4). Expansion in the hexanucleotide GGGGCC repeat tract within the first intron of C9orf72 was identified as the most frequent cause of familial ALS and FTD (5,6). While unaffected individuals most commonly carry 2–8 repeats (7), the number of repeats in patients was estimated to range between 700 and 1600 (5). Disease mechanism is currently still being investigated and the following three non-mutually exclusive hypotheses (8,9) have been proposed: (i) loss of function of C9orf72 gene (5,6,10), (ii) gain-of-function of C9orf72 RNA transcripts (5,11,12) and (iii) toxic dipeptide accumulation (13). The underlying mechanism of non-coding repeat expansion disorders most commonly involves an RNA dependent gain-of-function mechanism independently of the proteins encoded by their sequences (14). RNA gain-of-function mechanism involves unusual secondary structure formation in RNA transcripts leading to conformation-dependent sequestration of functionally important cellular proteins, a mechanism that was most thoroughly studied in myotonic dystrophy (15). G-rich GGGGCC repeat has a potential for G-quadruplex formation, and indeed it has been demonstrated that RNA (16–18) and DNA (17,19) oligonucleotides form G-quadruplex structures in vitro. Recently, a disease mechanism has been proposed in which structural polymorphism within expanded GGGGCC repeat, including G-quadruplex structures at both DNA and RNA level and transcriptionally induced RNA-DNA hybrids (R-loops) are implicated in the development of ALS and FTD (17). Formation of DNA G-quadruplexes and R-loops may cause pausing and abortion in transcription leading to a loss of full-length products and accumulation of abortive RNA transcripts. RNA G-quadruplex formation in RNA transcripts in turn leads to conformation-dependent interaction
with nucleolar proteins, causing nucleolar stress thereby increasing the sensitivity of the cells for chronic neurodegenerative disease. Nucleolin, an essential nucleolar protein, has been found to bind RNA (GGGGCC)₄ oligonucleotide and experimental data have suggested that it recognizes specifically folded G-quadruplex adopted by this oligonucleotide (17). Some of the transcribed RNA with GGGGCC repeats may leak out of the nucleus and undergo repeat-associated non-ATG dependent translation in the cytoplasm that results in formation of aggregative polydipeptides (20).

G-quadruplexes may adopt diverse architectures that vary in terms of molecularity, strand orientation, number of G-quartets and structure of loops connecting guanine residues involved in G-quartets. G-rich oligonucleotides are notorious for their polymorphism. In addition, several G-quadruplexes can coexist in solution, which represents a great challenge for high-resolution structure determination. Small changes in oligonucleotide sequence can have a profound effect on the folding topology (21,22) and mutations or chemical modifications of oligonucleotides are frequently employed to stabilize a particular fold (23–25). Furthermore, several G-quadruplex forming oligonucleotides have been shown to be very sensitive to experimental conditions used for folding, including type of cations and their concentration, pH and temperature (26–28).

Stimulated by the emerging experimental data on molecular details of ALS and FTD suggesting that G-quadruplex formation by repetitive GGGGCC repeats may be an important component in occurrence of devastating biological consequences, we screened d(GGGGCC)₄ oligonucleotide and several modified sequences as the shortest model for unmolecular G-quadruplex. Interestingly, occurrence of G-C Watson-Crick base pairs was not a common NMR spectral feature neither at low or high K⁺ ion concentrations. Screening revealed formation of multiple G-quadruplex structures in the presence of K⁺ ions for most oligonucleotides with four hexanucleotide repeats that were modified or expanded at 5'- and 3'-ends. Oligonucleotide d[(G₄C₃)₃G₄] labeled wt22 was shown to fold into two predominant structures with an antiparallel topology with characteristic syn glycosidic conformation of some guanine residues. Since 8-bromoguanine residue preferentially adopts a syn glycosidic conformation (29), substituting guanine with its 8-bromo analogue at a desired position in a sequence can lead to stabilization of anticipated structure (25,30,31). Single dG to 8Br-dG substitution in 22 residue oligonucleotide sl21 and careful optimization of folding conditions enabled us to obtain a sample of sufficient quality to determine a high-resolution structure with NMR. Notably, sl21 exhibits NMR resonances at chemical shifts of the species observed for its wild type analogue indicating equivalent structural characteristics. To the best of our knowledge, G-quadruplex adopted by sl21 represents the first high-resolution structure of DNA oligonucleotide related to ALS and FTD. Novel G-quadruplex structure expands the repertoire of known G-quadruplex folding topologies and provides a potential target for structure based design of a molecule directed toward ALS and FTD related GGGGCC hexanucleotide repeats.

MATERIALS AND METHODS

Sample preparation

Oligonucleotides wt22 and sl21 were synthesized on K&A Laborgeraete GbR DNA/RNA Synthesizer H-8 using standard phosphoramidite chemistry in DMT-off mode. Oligonucleotides containing 8% ¹⁵N site-specifically labeled guanine residues and oligonucleotides containing 8% ¹³C site-specifically labeled guanine residues were synthesized in DMT-off mode. Oligonucleotide wt22 was deprotected in concentrated aqueous ammonia at 55°C overnight. Oligonucleotides containing modified 8Br-dG base were deprotected in concentrated aqueous ammonia at room temperature for 48 h in the dark. Samples synthesized in DMT-off mode were desalted on a Sephadex G25 column using FPLC. Samples prepared in DMT-on mode were purified using reverse-phase HPLC chromatography, followed by removal of DMT group with reaction in 80% AcOH for 30 min. Oligonucleotides were then transferred to a pure water phase by ethyl ether extraction and desalted on a Sephadex G25 column using FPLC. Desalted samples were lyophilized and dissolved in 850 μl H₂O. The pH of the samples was adjusted to around 7 using LiOH prior to heating them at 90°C for 5 min. After heating, KCl and phosphate buffer with pH 7.2 were added to the samples, followed by slowly cooling from 90°C to 25°C over 16 h. Samples were again lyophilized and redissolved in 350 μl 90% H₂O and 10% ²H₂O solution. The sample in ²H₂O was prepared by repeated lyophilization and final dissolution in 350 μl ²H₂O. Concentration of the samples was determined by UV absorption at 260 nm using UV/VIS Spectrophotometer Varian CARY-100 BIO UV–VIS. Extinction coefficient of 204400 M⁻¹ cm⁻¹ was determined for wt22 by nearest neighbor method. Extinction coefficients of oligonucleotides with modified bases were determined using base composition method and were 204300 M⁻¹ cm⁻¹ for oligonucleotides with dG to 8Br-dG substitution and 202950 M⁻¹ cm⁻¹ for oligonucleotides with both dC to 5Me-dC and dG to 8Br-dG substitutions.

Circular dichroism spectroscopy

A circular dichroism (CD) spectrum was recorded on an Applied Photophysics Chirascan CD spectrometer at 25°C using a 1.0 mm path length quartz cell. The wavelength range was from 200 nm to 320 nm. CD sample was prepared at 10 μM oligonucleotide concentration in 120 mM KCl and 20 mM phosphate buffer. A blank sample containing only 120 mM KCl and 20 mM potassium phosphate buffer with pH 7.2 was used for baseline correction.

UV melting

UV melting experiments were performed on a Varian CARY-100 BIO UV–VIS spectrophotometer using a 2 mm (44 μM sample) and 1 cm (for lower concentrations) path-length cells. Samples were heated at 0.1°C min⁻¹ from 25°C to 95°C and absorbance at 295 nm was measured at 0.5°C steps. Mineral oil was used to prevent evaporation at high temperatures. Tm was determined from the first derivative of A295 versus temperature plot.
Native PAGE

Native gel electrophoresis was run in a temperature-controlled vertical electrophoretic apparatus at 5°C in TBA (pH 8.1) buffer and 120 mM KCl. Samples contained approximately 5 μg of DNA in a solution with 120 mM KCl and 20 mM phosphate buffer with pH 7.2. Ficoll was added to the samples prior to loading. Polyacrylamide gel concentration was 20%. GeneRuler Ultra Low Range DNA ladder with 10–300 base pairs (Thermo Scientific) was used to indicate approximate DNA size. DNA was visualized by Stains-all (Sigma-Aldrich) staining.

NMR experiments and restraints

All NMR experiments were performed on Agilent-Varian NMR Systems 600 MHz and 800 MHz spectrometers at 0°C. Double-pulsed field gradient spin echo (DPFGSE) pulse sequence was used to suppress the water signal. The translation diffusion coefficient was determined by performing double-pulsed field gradient stimulated echo pulse sequence using 20 different gradient strengths (1.3–31.00 G cm⁻¹). Identification of guanine imino protons in 8% 15N site-specifically labeled samples was achieved by performing 1D 15N-filtered HSQC experiments. Guanine H8 protons were identified by performing 1D 13C-filtered HSQC experiments on 8% 15N, 13C site-specifically labeled samples. Standard homonuclear 2D NMR experiments including 2D DQF-COSY, 2D TOCSY (20, 40 and 80 ms mixing time) and NOESY (80, 150 and 250 ms mixing time) recorded in 100% 2H₂O were used to assign non-exchangeable protons. Exchangeable proton resonances were assigned using 2D NOESY (80, 100, 150, 200, 300 ms) recorded in 90% H₂O, 10% 2H₂O. Most NOE (Nuclear Overhauser Effect) distance restraints for non-exchangeable protons were obtained from a 2D NOESY spectrum recorded at 0°C in 100% 2H₂O with a mixing time of 80 ms and the rest from 2D NOESY recorded at 0°C in 90% H₂O, 10% 2H₂O with a mixing time of 100 ms. The average volume of H5–H6 cross-peak of C11 was used as the distance reference of 2.45 Å. Cross-peaks were classified as strong (1.8–3.6 Å), medium (2.6–5.0 Å) and weak (3.5–6.5 Å). NOE distance restraints for exchangeable protons were obtained from a 2D NOESY spectrum recorded at 0°C in 90% H₂O, 10% 2H₂O with mixing times of 80 ms and 300 ms. Cross-peaks of medium and weak intensity in 2D NOESY spectrum with a mixing time of 80 ms were classified as strong (1.8–3.6 Å) and medium (2.6–5.0 Å), respectively. Cross-peaks that appeared in 2D NOESY spectrum with a mixing time of 300 ms were classified as weak (3.5–6.5 Å). Torsion angle χ around glycosidic bond was restrained to a range between 25 and 95° for syn and between 200 and 280° for anti residues. The endocyclic torsion angle ψ2 was used to restrain the sugar puckering to N-type sugar conformation for C5 and C17 and S-type conformation for all of the other residues, except for G22 which was left unrestrained.

Structure calculations

Structure calculations were performed with AMBER 14 software using parmbsc0 force field (32) with parmχ OL4 (33) and parme/OL1 (34) modifications. In order to avoid discrepancy of bond angles in the final family of structures with respect to the standard values (35,36), the force field parameters for the following five bond angles had to be modified: N9-C1′-O4′, N1-C1′-O4′, C3′-O3′-P, N3-C2-O2 and N1-C2-O2. First, the respective force constants have been increased by up to an order of magnitude. In addition, the equilibrium bond angle values have been changed to match the standard values reported by Gelbin et al (35) and Clooney et al (36) Calculation was started from an initial extended structure of oligonucleotide d(G1-C2)3GG8G8GG, created with LEAP module of the AMBER 14 program. A total of 1000 structures were calculated in 130 ps of NMR restrained simulated annealing (SA) simulations using the generalized Born implicit model to account for solvent effects. The cut-off for non-bonded interactions was 20 Å and the SHAKE algorithm for hydrogen atoms was used with the tolerance of 0.0005 Å. SA calculations were initiated with random velocities. In the first 50 ps of SA, the temperature was raised from 300 K to 1000 K. Molecules were held at constant temperature of 1000 K for 20 ps and then cooled to 300 K in the next 30 ps, after which the temperature was scaled down to 0 K in the last 30 ps. Force constants were 20 kcal mol⁻¹ Å⁻² for hydrogen bond restraints, 20 kcal mol⁻¹ Å⁻² for NOE distances, 200 kcal mol⁻¹ rad⁻² for torsion angle ψ2, 20 kcal mol⁻¹ rad⁻² for torsion angle χ and 20 kcal mol⁻¹ Å⁻² for G-quartet base planarity restraints. Planarity restraints for G-quartets were excluded in the last 30 ps of SA. A family of 10 structures was selected based on the smallest energy and subjected to energy minimization with a maximum of 10 000 steps. As a control, NMR restraints were omitted in the minimization steps. Figures were visualized and prepared with UCSF Chimera software (37).

RESULTS

Stabilization of a predominant structure

For initial screening purposes, we designed several different mutational variants of d(GGGGCC)₄ oligonucleotide with CC or TT residues at either 5′, 3′ or both ends. Observation of a broad envelope of peaks in the imino region of Hoogsteen hydrogen-bonded structures in 1H NMR spectra indicated the formation of multiple G-quadruplex structures in the presence of K⁺ ions. The 22 residue oligonucleotide wt22 with sequence d(G1-C2)₃G₁₄ showed the best spectral characteristics and was chosen for further evaluation (Supplementary Figure S1). In addition to multiple weak signals in the imino and aromatic regions of 1D 1H NMR spectrum of wt22 indicating structural polymorphism, we observed two sets of sharp and intense proton signals, suggesting the coexistence of two predominant G-quadruplex species. In the 1D 1H NMR spectrum of wt22 we also observed signals around 6.13 ppm corresponding to Watson-Crick G-C base pairs. Since the oligonucleotide is GC-rich, we expect the structure to contain G-C-G-C quartets (38–41). However, diluting the DNA sample concentration down to 0.3 mM per strand caused these signals to disappear, indicating they are a part of a separate, probably dimeric structure (Supplementary Figure S2). The presence of syn guanine residues as evident from NOE spectra together with CD profile suggested antiparallel orientation of the strands.
in both predominant species. Stimulated by these encouraging data, we examined folding of oligonucleotides with a single dG to 8Br-dG substitution in sequence of wt22 at positions that were occupied by residues in syn conformation in the presumed topologies (Supplementary Figure S3). Sequence sl21, d[[(G12C1)3GGBrGG] with single dG to 8Br-dG substitution at position 21 exhibited favorable spectral characteristics with only two sets of sharp signals in 1H NMR spectrum. Notably, the signals in the 1H NMR spectrum of sl21 were very similar to two sets of more intense signals in wt22 (Figure 1, see Supplementary Figure S4 for extended chemical shift range), suggesting that both topologies adopted by the parent oligonucleotide were conserved in sl21. In order to possibly favor a single structure thus simplifying the process of high-resolution structure determination, we tested folding of sl21 at different solution conditions. We observed that the ratio between the two structures was not influenced by concentration of cation, but was however very sensitive to pH of the solution. Interestingly, the intensity of one set of signals decreased with increasing pH (Supplementary Figure S5). This characteristic was exploited to obtain a sample with 70% major and 30% minor species, suitable for structure determination. We determined the structure of the major G-quadruplex species adopted by sl21 under physiologically relevant pH 7.2 and 120 mM KCl.

Resonance assignment

Structural characterization of sl21 started with unambiguous assignment of guanine imino and aromatic resonances. In the imino region of 1D 1H NMR spectrum of sl21 in K+ ion containing solution, 16 major signals were observed in the region between δ 11.2 and 11.9 ppm. The presence of 16 imino resonances indicates that all of the 16 guanine residues are involved in G-quartets and therefore the structure contains four G-quartet planes. Unambiguous assignment of guanine imino resonances was achieved by recording 1D 15N-filtered HSQC spectra on 8% 15N residue specifically labeled oligonucleotides (Figure 2). Imino resonance of BrG21 that could not be isotopically labeled was identified unequivocally through its cross-peaks in NOESY spectra recorded at longer mixing times.

Aromatic H8 protons of guanine residues have been unambiguously assigned by recording 1D 13C filtered HSQC spectra of partially 8% 15N, 13C residue specifically labeled oligonucleotides (Supplementary Figure S6) which has been complemented by the sequential walk in NOESY spectra. Seven distinct, strong cross-peaks in the aromatic-anomeric region of NOESY spectrum with 80 ms mixing time in-
Figure 3. The aromatic H6/H8-sugar H2'/H2" region of NOESY spectrum (τm = 200 ms) of sl21 (A) and wt22 (B). Assignments are shown next to H6/H8-sugar H2'/H2" cross-peaks. The lines that connect aromatic H6/H8-sugar H2' and aromatic H6/H8-sugar H2" cross-peaks are depicted in orange and cyan, respectively. Letter x indicates C11H2'-C12H6 cross-peak visible at higher vertical scale or longer mixing time.

Figure 4. Imino-imino and imino-aromatic regions of NOESY spectra of sl21. (A) Assignments in imino-imino region of NOESY spectrum (τm = 200 ms) are shown next to the cross-peaks. Intraquartet cross-peaks correlating residues of adjacent strands are depicted in red (between outer and inner G-quartets) and brown (between inner G-quartets). Intraquartet sequential cross-peaks are depicted in pink. Intraquartet cross-peaks are in blue. Stars indicate signals of minor species. Assigned 1D 1H NMR spectrum is shown on top. Cross-peak marked with an X could be interpreted either as G4H1-G19H1, G4H1-G7H1, G16H1-G19H1 or G16H1-G7H1. (B) NOESY spectrum (τm = 300 ms) showing typical H1-H8 correlations between neighboring guanines in G-quartets. Assignments corresponding to individual G-quartets are shown in the same color. Assignments in yellow correspond to contacts between C12 and imino protons of nearby G-quartet. Cross-peak G2 H1-BrG21 H8 is missing because residue BrG21 is brominated. Spectra were recorded in 90% H2O, 10% 2H2O at 0°C, 120 mM KCl, 20 mM phosphate buffer with pH 7.2 and 0.5 mM oligonucleotide concentration per strand on 600 MHz spectrometer.

C6 and C18 had the same chemical shifts and identical cross-peak patterns in 2D spectra. Assignment of cytosine residues was achieved through residue specific dC to 5Me-dC substitution (Supplementary Figure S8 and Table S1 for chemical shifts). Analysis of DQF COSY and TOCSY spectra demonstrated that all residues except C5 and C17 exhibited large 3JH1'H2' coupling constants indicating their bias toward S-type sugar conformation.

Antiparallel topology with four G-quartets

In depth analysis of NMR and other spectroscopic data showed that the major form of sl21 adopts an antiparallel topology with four G-quartets, where every strand is antiparallel with respect to adjacent strands. Antiparallel orientation of the strands with alternating syn-anti glycosidic conformation along each strand is in complete agreement with CD profile with minimum at 260 nm and maxima at 245 nm and 295 nm (Figure 1). Unambiguously assigned imino and H8 proton resonances enabled us to determine the topology of sl21, based on characteristic cyclic imino-H8 connectivities (Figure 4). Adjacent G-quartets have opposite hydrogen-bond directionalities and variable
immediately after exposure to $^2$H$_2$O, while signals for outer G-quartets. Signals or disappearance of the signal. As expected, the most exposed imino protons are the first to exchange into deuterated water. In such an experiment, supported by the rates of H-D exchange when sl21 is transferred into deuterated water. In such an experiment, the most exposed imino protons are the first to exchange with bulk solvent. Although both inner G-quartets exhibit expected slow exchange in $^2$H$_2$O, the exchange rates are not the same. Signals for imino protons of G2→$^b$G21→G14→G9 disappear in a matter of days, while signals for G3→G8→G15→G20 are present even after 25 days of exposure to $^2$H$_2$O (Figure 5A). The difference in exchange rates is related to the structural features of sl21 G-quadruplex. The inner G3→G8→G15→G20 quartet is protected from exchange by an outer G-quartet that is covered by a capping structure consisting of two C-C loops stacked upon the outer G-quartet. On the other hand, G2→$^b$G21→G14→G9 quartet is stacked by an outer G-quartet that is covered by only one stacked C-C loop (Figure 5B).

In the imino-imino region of 2D NOESY spectra, the most intense NOE cross-peaks connecting G1→G10→G13→G22 with G2→$^b$G21→G14→G9 quartet and G3→G8→G15→G20 with G4→G19→G16→G7 quartet are between guanine residues located on adjacent strands (red in Figure 4). Cross-peaks G22H1-G21H1 and G21H1-G13H1 are not resolved because of the close proximity to the diagonal. In the structure, these NOE interactions correspond to distances below 3.5 Å and such relatively short distances are in agreement with the right hand twist of the structure. No cross-peaks connecting sequential residues between outer and inner G-quartets were observed in NOESY spectra, which is in agreement with measured distances higher than 4.9 Å. In comparison, the most intense cross-peaks between the inner G2→$^b$G21→G14→G9 and G3→G8→G15→G20 quartets are observed between sequential residues (pink in Figure 4) and correspond to distances around 3.5 Å. Cross-peaks correlating adjacent residues in the two inner G-quartets are weaker (brown in Figure 4), with the corresponding NH-NH distances around 4.3 Å. Cross-peaks G3H1-G9H1 and G8H1-G14H1 were not resolved due to proximity to the diagonal. Spectral overlap, proximity to the diagonal and ambiguity due to symmetry allowed the assignment of only four NH-NH contacts between neighboring residues within a G-quartet (blue in Figure 4).

**Figure 5.** H-D exchange of imino protons upon transfer into $^2$H$_2$O and topology of sl21. (A) Imino proton spectra of sl21 recorded at different times after dissolving in $^2$H$_2$O. Assignments are indicated by residue number above the signal. Spectra were recorded at 25 °C, 120 mM KCl, 20 mM phosphate buffer with pH 7.2 and 0.1 mM oligonucleotide concentration per strand on 600 MHz spectrometer. (B) Schematic presentation of G-quadruplex adopted by sl21. *Anti* and *syn* guanines are designated with lighter and darker shades of the same color, respectively.

(syn:anti:syn:anti) or (anti:anti:syn:anti) arrangement of conformations around glycosidic bonds in guanine residues. G-quartets are linked by three edgewise loops that progress in a clockwise manner, forming a (+1, +1, +1) topology (42). G-quadruplex core consists of the following four G-quartets: G1→G10→G13→G22, G2→$^b$G21→G14→G9, G3→G8→G15→G20 and G4→G19→G16→G7, where an arrow represents hydrogen bond directionality. The proposed topology is supported by the rates of H-D exchange when sl21 is transferred into deuterated water. In such an experiment, the most exposed imino protons are the first to exchange with deuterium, which results in the decreased intensity or disappearance of the signal. As expected, the most exposed in sl21 are residues in outer G-quartets. Signals for outer G1→G10→G13→G22 quartet disappear immediately after exposure to $^2$H$_2$O, while signals for outer G4→G19→G16→G7 quartet are observed after at least 15 min in $^2$H$_2$O, with signals for residues G7 and G19 present even after 4 h in $^2$H$_2$O, suggesting they are especially well protected from exchange with bulk solvent. Although both inner G-quartets exhibit expected slow exchange in $^2$H$_2$O, the exchange rates are not the same. Signals for imino protons of G2→$^b$G21→G14→G9 disappear in a matter of days, while signals for G3→G8→G15→G20 are present even after 25 days of exposure to $^2$H$_2$O (Figure 5A). The difference in exchange rates is related to the structural features of sl21 G-quadruplex. The inner G3→G8→G15→G20 quartet is protected from exchange by an outer G-quartet that is covered by a capping structure consisting of two C-C loops stacked upon the outer G-quartet. On the other hand, G2→$^b$G21→G14→G9 quartet is stacked by an outer G-quartet that is covered by only one stacked C-C loop (Figure 5B).

**Figure 5.** H-D exchange of imino protons upon transfer into $^2$H$_2$O and topology of sl21. (A) Imino proton spectra of sl21 recorded at different times after dissolving in $^2$H$_2$O. Assignments are indicated by residue number above the signal. Spectra were recorded at 25 °C, 120 mM KCl, 20 mM phosphate buffer with pH 7.2 and 0.1 mM oligonucleotide concentration per strand on 600 MHz spectrometer. (B) Schematic presentation of G-quadruplex adopted by sl21. *Anti* and *syn* guanines are designated with lighter and darker shades of the same color, respectively.

**NOE interactions of loop residues**

All three loops in sl21 G-quadruplex are defined and exhibit not only sequential but also several long-range NOE contacts (Supplementary Figure S9). In the C5-C6 and C17-C18 edgewise loops stacked on the G4→G19→G16→G7 quartet, observation of regular sequential NOE connectivities indicates both cytosine residues in the loops adopt a well-defined structure. Interestingly, clear NOE contacts are observed between C6H5/H1'/H2'/H2''/H3' and imino proton of G7. Moreover, unusually large number of NOE contacts are observed between protons of C6 and G4, including G4H8-C6H6 cross-peak of medium intensity. Taken together these NOE data indicate that C6 is stacked above G4 and G7, with its aromatic protons close to G4. Analogous pattern of NOE contacts is observed for C18 residue, indicating it is stacked above G16 and G19. In the C11-C12 loop, C12 is stacked on the G1→G10→G13→G22 quartet, as supported by many NOE contacts to guanine residues in the G-quartet, most notably C12H6 to imino protons of G1, G10 and G13. Observation of regular sequential NOE contacts for C11 indicate its structure is well-defined and its placement over the G-quartet is in agreement with unusual C11H1'→G10H1 NOE contact. However, C11 exhibits no long-range NOE contacts and could in principle be located in the wide groove between G10 and G13 with only small NOE violations. We examined NOESY spectrum of sl21 analogue with dC to 5Me-dC substitution at position 11. Weak, but clear NOE was observed between methyl protons of 5MeC11 and G1H1 confirming the position of C11 above the G-quartet.
Table 1. Structural statistics for G-quadruplex structure adopted by sl21

| NMR distance and torsion angle restraints          |          |
|--------------------------------------------------|----------|
| NOE-derived distance restraints                   |          |
| Total                                            | 337      |
| Intra-residue                                    | 140      |
| Inter-residue                                    | 197      |
| Sequential                                      | 120      |
| Long-range                                      | 77       |
| Hydrogen bond restraints                          | 32       |
| Torsion angle restraints                          | 43       |
| G-quartet planarity restraints                    | 48       |

**Structure statistics**

- Mean NOE restraint violation (Å): 0.11 ± 0.04
- Max. NOE restraint violation (Å): 0.23
- Max. torsion angle restraint violation (°): 0.00
- Deviations from idealized geometry
  - Bonds (Å): 0.012 ± 0.000
  - Angles (°): 2.45 ± 0.02
- Pairwise heavy atom RMSD (Å)
  - Overall: 0.67 ± 0.14
  - G-quartets: 0.63 ± 0.16
  - G-quartets and C5-C6: 0.62 ± 0.15
  - G-quartets and C11-C12: 0.65 ± 0.15
  - G-quartets and C17-C18: 0.67 ± 0.16

**Structure calculation**

Solution structure of sl21 major G-quadruplex species was calculated using 337 NOE derived distance restraints, together with 32 hydrogen-bond, 43 torsion angle and 48 planarity restraints (Table 1). On average 24.3 NOE distance restraints per residue were used (Supplementary Figure S10). A set of 1000 structures was calculated. Ten final structures were chosen based on the lowest energy and subjected to energy minimization (Figure 6). The G-quadruplex structure formed by sl21 is very well defined with overall pairwise heavy atom RMSD of 0.7 Å. The G-quartets are expectedly the best defined part of the structure. However, inclusion of C-C loops in the calculation of RMSD does not increase the value significantly (Table 1). Although unimolecular quadruplexes cannot be truly symmetrical, the structure does have a pseudo C2 symmetry axis running through the core of the structure, which is broken down by C11-C12 loop.

As a consequence, guanine and cytosine residues in part of the structure farther from the C11-C12 loop, namely C5-C6 and C17-C18 loops and G4-G19-G16-G7 were isochronous, and some NOE restraints involving them ambiguous. Ambiguous NOE restraints were not used in the initial stages of the calculation. After the first round of calculations, we examined the structure and measured the distances between protons that were ambiguously assigned. This analysis revealed that for some of the ambiguous NOE cross-peaks, only one of the possible assignments was possible. In this manner we obtained some additional distance restraints. However, their inclusion in a new cycle of calculation produced minor, if any structural change. NOE assignments that remained ambiguous after inspecting the initial calculated structure were not used as restraints in the later stages of calculation.

The antiparallel strands form alternating narrow and wide grooves. The dimension of wide groove between strands G1-G4 and G22-G19 is 9.8 Å, while the width of the opposite lying wide groove formed between strands G10-G7 and G13-G16 is 10.5 Å. The two narrow grooves are of similar dimensions (Supplementary Table S2). The loops are stacked upon the outer G-quartets making a very compact structure. The value of measured translational diffusion coefficient 1.85 ± 0.11 × 10⁻⁶ cm² s⁻¹ (Supplementary Figure S11) is in agreement with a compact unimolecular structure. Compact structure is supported by PAGE analysis (Supplementary Figure S12), where sl21 exhibited gel mobility comparable to unimolecular G-quadruplex with three G-quartet planes (43). Unimolecular structure is consistent with concentration-independent UV melting temperature experiments (Supplementary Figure S13). Furthermore, diluting the sample 10 and 20 times did not cause the appearance of any new signals in 1D ¹H NMR spectrum (Supplementary Figure S14).

**Edgewise loops are stacked on the outer G-quartets**

In the edgewise loop comprised of C5 and C6, residue C6 is stacked above G7 and G4. Likewise, in the C17-C18 edge-wise loop, residue C18 is stacked above G19 and G16. Aromatic protons of C6 and C18 point away from the center of the G-quartet placing H5 protons of C6 and C18 right above the pyrimidine moiety of G4 and G16, respectively. Shielding due to ring-current effects correlates well with unusually shielded H5 protons of C6 and C18 (δ 4.34 ppm at 0°C, Supplementary Figure S7). The isochronous chemical shifts and identical NOE patterns are in agreement with C5-C6 and C17-C18 loops being highly symmetric (Figure 7). In the ¹H NMR spectra of sl21 analogues with a single 5Me-dC substitution (Supplementary Figure S8), it can be seen that spectrum of oligonucleotides with dC to 5Me-dC substitution at position 5 and 6 are very similar to oligonucleotides with substitution at positions 17 and 18, respectively. This observation is in complete agreement with symmetrical arrangement of C5-C6 and C17-C18 loops. While C5-C6 and C17-C18 both span narrow grooves, C11-C12 spans a wide groove which is reflected in its distinct structure. In the C5-C6 and C17-C18 loops, C5 and C17 are placed somewhat above C6 and C18, respectively. On the other hand, C11 and C12 are approximately in the same plane, parallel to the G-quartet. The C11-C12 edgewise loop is placed above the G1→G10→G13→G22 quartet, where C12 is stacked above G13 and G22 and its aromatic protons point toward the center of the G-quartet (Figure 7). Interestingly, this orientation brings sugar H4'/H5'/H5″ protons of C12 above the aromatic ring of G13, whose ring currents provide an explanation for this unusual upfield chemical shifts.

**DISCUSSION**

Oligonucleotides with four GGGGCC repeats represent the smallest possible model for unimolecular G-quadruplex formation of ALS and FTD related hexanucleotide repeat. A recent study indicated that d(GGGGCC)₄ folds into a polymorphic mixture of G-quadruplex structures and recognized antiparallel topology as the major form in K⁺ ions containing solution. G-quadruplex formation in the DNA strand within the region containing C9orf72 hexanucleotide repeat has been suggested to cause impairment...
One of the cytosines in every loop is stacked upon the G-quartet contributing to a very compact and stable structure. G-quadruplex formation in the DNA strand is likely among the first steps in biochemical pathway leading to C9orf72 linked ALS and FTD. Specific targeting of G-quadruplexes in the DNA strand of C9orf72 with an appropriate drug molecule could help to intervene with the disease at its root. Unique structural features as demonstrated by stacked edgewise loops in sl21 could serve as a recognizing element enabling selective recognition and binding of a ligand, providing a potential target for the development of a specific drug molecule.

Epigenetic alterations, such as methylation in the promoter region can cause gene silencing, which may also play a role in the C9orf72 expanded repeat disease mechanism (44). Studies of downstream consequences of methylation in C9orf72 repeat expansion suggest that promoter hypermethylation observed in a subset of patients that carry C9orf72 repeat expansion provides a protective role in disease state, as it causes a decrease in overall C9orf72 expression (45,46). One group hypothesized that in patients carrying the pathological expansion, an extra CpG-island is formed within the GGGGCC repeat itself. They showed that GGGGCC methylation actually occurs in 97% of expansion carriers (patients with over 50 repeats). Xi et al. have posed a question whether methylation prevents or enhances G-quadruplex formation (47). In the course of our study, we examined oligonucleotides with 5-methylated cytosine residues, the same type of methylation that occurs in CpG islands in vivo. These oligonucleotides folded in the same way as the parent oligonucleotide sl21, showing that methylation does not prevent G-quadruplex formation. Methylation within the GGGGCC repeat may have different consequences than methylation in the promoter region of C9orf72 gene, potentially facilitating G-quadruplex formation in the DNA strand and thereby increasing the production of abortive RNA transcripts. Methylation effects were extensively studied in fragile X syndrome, where large expansion of d(CGG)₆ triplet repeat leads to genetic
instability. Cytosine residues in triplet repeat are hyper-methylated (48) which is believed to worsen the disease condition (49,50). A study by Fry and Loeb on oligonucleotides with CGG showed facilitated G-quadruplex formation upon methylation of cytosine residues (51). High-resolution structure of dimeric G-quadruplex formed by d(CCGGT3GCGG) oligonucleotide suggested that favorable stacking between methyl group of 5Me-dC and guanine residues would occur upon methylation of cytosines (39). In accordance, it was previously proposed that methylation of cytosine residues stabilizes G-quadruplex structure by additional favorable stacking interactions (44). In G-quadruplex adopted by sl21, aromatic protons of C6 and C18 are positioned right above the aromatic rings of G4 and G16, respectively. Methylation of these cytosine residues would provide favorable stacking interaction and overall stabilization of the G-quadruplex structure.

The sl21 G-quadruplex structure adds to the topological diversity of G-quadruplexes, particularly expanding the handful of structures in the family of G-quadruplexes containing four G-quartets (52–58). To our knowledge, this is the first reported four G-quartet containing G-quadruplex with only edgewise loops and all strands antiparallel to one another. From examining a sequence alone, it is not easy to predict a folding topology of oligonucleotides. In contrast to the fold adopted by sl21, a oligonucleotide with similar TET4T sequence d(T2G4T) folds into a three G-quartet layered G-quadruplex with two edgewise and one propeller loop (59). Another closely related oligonucleotide d(G4C4T) containing GGGGC repeat from RET promoter, folds into a parallel G-quadruplex with three G-quartets and three propeller loops (60). A similar folding topology with antiparallel orientation of strands and three edgewise loops, but with only two G-quartets, was previously reported for thrombin binding aptamer (TBA) (61). Equivalent to the structure of the two C-C loops spanning the narrow grooves in sl21, the two T-T loops in TBA have one thymine residue in each loop stacked over neighboring G-quartet and positioned in a way to allow base-pairing between the thymine residues in the loops opposite to each other. In TBA, evidence suggests that hydrogen bond formation is between N3 hydrogen and carbonyl O2 atom (61,62). In sl21 G-quadruplex, the two cytosine residues C6 and C18 that are stacked on the neighboring G-quartet are also in alignment that allows formation of C6-C18 hydrogen bond. Specifically, inter-residue distances between O2 and H4 atoms that could form symmetric C-C carbonyl-amino bond are under 3 Å. Moreover, the inter-residue distances between N3 and H4 atoms are under 2 Å, making a potential symmetric C-C N3-amino bond even more probable. However, since we observed no NOE contacts involving N4 amino hydrogens of C6 or C18, we cannot tell if hydrogen-bond formation between C6 and C18 residues actually occurs.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**REFERENCES**

1. Eddy, J. and Maizels, N. (2009) Selection for the G4 DNA motif at the 5′ end of human genes. *Mol. Carcinog.*, 48, 319–325.
2. Verma, A., Yadav, V.K., Basundra, R., Kumar, A. and Chowdhury, S. (2009) Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cells. *Nucleic Acids Res.*, 37, 4194–4204.
3. Wu, Y. and Brosh, R.M. Jr (2010) G-quadruplex nucleic acids and human disease. *FEBS J.*, 277, 3470–3488.
4. Lillo, P. and Hodges, J.R. (2009) Frontotemporal dementia and motor neuron disease: overlapping clinic-pathological disorders. *J. Clin. Neurosci.*, 16, 1131–1135.
5. DeJesus-Hernandez, M., Mackenzie, I.R., Boone, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J. et al. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*, 72, 245–256.
6. Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollison, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Molyk, L. et al. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*, 72, 257–268.
7. Rutherford, N.J., Heckman, M.G., Dejesus-Hernandez, M., Baker, M.C., Soto-Ortolaza, A.I., Rayaprolu, S., Stewart, H., Finger, E., Volkening, K., Seeley, W.W. et al. (2012) Length of normal alleles of C9ORF72 GGGGCC repeat do not influence disease phenotype. *Neurobiol. Aging*, 33, e5–e7.
8. Orr, H.T. (2013) Toxic RNA as a driver of disease in a common form of ALS and dementia. *Proc. Natl. Acad. Sci. U.S.A.*, 110, 7533–7534.
9. Vatovec, S., Kovanova, A. and Rogelj, B. (2014) Unconventional features of C9ORF72 expanded repeat in amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *Neurobiol. Aging*, 35, e1–e12.
10. Gijselinck, I., Van Langenhove, T., van der Zee, J., Sleepers, K., Philtjens, S., Kleinberger, G., Janssens, J., Betens, K., Van Cauwenbergh, C., Pereson, S. et al. (2012) A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. *Lancet Neurol.*, 11, 54–65.
11. Renoux, A.J. and Todd, P.K. (2012) Neurodegeneration the RNA way. *Prog. Neurobiol.*, 97, 173–189.
12. Xu, Z., Poidevin, M., Li, X., Li, Y., Shu, L., Nelson, D.L., Li, H., Hales, C.M., Gearing, M., Wingo, T.S. et al. (2013) Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc. Natl. Acad. Sci. U.S.A.*, 110, 7778–7783.
13. Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A. et al. (2011) Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U.S.A.*, 108, 260–265.
14. van Blitterswijk, M., Dejesus-Hernandez, M. and Rademakers, R. (2012) How do C9ORF72 repeat expansions cause amyotrophic lateral sclerosis and frontotemporal dementia: can we learn from other noncoding repeat expansion disorders? *Curr. Opin. Neurol.*, 25, 689–700.
15. Cooper, T.A., Wan, L. and Dreyfuss, G. (2009) RNA and disease. *Cell*, 136, 777–793.
16. Fratta, P., Mizielsinska, S., Nicoll, A.J., Zloh, M., Fisher, E.M., Parkinson, G. and Isaacs, A.M. (2012) C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci. Rep.*, 2, 1016.
17. Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A., Shaw, P.G., Kim, M.S., Maragakis, N.J., Oncono, J.C., Pandey, A., Satter, R. et al. (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, 507, 195–200.
18. Reddy, K., Zamiri, B., Stanley, S.Y., Macgregor, R.B. Jr and Pearson, C.E. (2013) The disease-associated r(GGGGCC)n repeat
from the C9orf72 gene forms tract-length dependent uni- and multimolecular RNA G-quadruplex structures. *J. Biol. Chem.*, 288, 9860–9866.

19. Skret,P., Polheven,J., Kovanda,A., Stakek,M., Zupunksi,V., Zalar,M., Plavec,J. and Rogelj,B. (2015) Characterization of DNA G-quadruplex species forming from C9orf72 G4C2-expanded repeats associated with amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *Neurobiol. Aging*, 36, 1091–1096.

20. Ash,P.E., Bieniek,K.F., Gendron,T.F., Caulfield,T., Lin,W.L., Dejesus-Hernandez,M., van Blitterswijk,M.M., Jansen-West,K., Paul,J.W. 3rd, Rademakers,R. et al. (2013) Unconventional translation of C9orf72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron*, 77, 639–646.

21. Cmrgel,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.

22. Sattin,G., Arteca,A., Nadai,M., Costa,G., Parrotta,L., Alcaro,S., Palumbo,M. and Richter,S.N. (2013) Conformation and stability of intramolecular telomeric G-quadruplexes: sequence effects in the loops. *PLoS One*, 8, e84113.

23. Luu,K.N., Phan,A.T., Kuryayi,V., Lacroix,L. and Patel,D.J. (2006) Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.*, 128, 9963–9970.

24. Dai,J., Carver,M., Punchihewa,C., Jones,R.A. and Yang,D. (2007) Structure of the hybrid 2-type intramolecular G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.*, 35, 4927–4940.

25. Lim,K.W., Amrane,S., Bouaziz,S., Xu,W., Mu,Y., Patel,D.J., Luu,K.N. and Phan,A.T. (2009) Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.*, 131, 4301–4309.

26. Wang,Y. and Patel,D.J. (1993) Solution structure of the human telomeric repeat [d(AG3[TAGA3])] G-tetraplex. *Structure*, 1, 263–282.

27. Parkinson,G.N., Lee,M.P. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, 417, 876–880.

28. Trajkovsky,M., da Silva,M.W. and Plavec,J. (2012) Unique structural features of interconvergent monomeric and dimeric G-quadruplexes adopted by a sequence from the intron of the N-myel gene. *J. Am. Chem. Soc.*, 134, 4132–4143.

29. Sarma,R.H., Lee,C.H., Evans,F.E., Yathindra,N. and Sundaralingam,M. (1974) Probing the interrelation between the glycosyl torsion, sugar pucker, and the backbone conformation in C(8) substituted adenine nucleotides by 1H and 1H-31P fast Fourier transform nuclear magnetic resonance methods and conformational energy calculations. *J. Am. Chem. Soc.*, 96, 7337–7348.

30. Matsugami,A., Xu,Y., Noguchi,Y., Sugiyama,H. and Katahira,M. (2007) Structure of a human telomeric DNA sequence stabilized by 8-bromoguanosine substitutions, as determined by NMR in K+ solution. *FEBS Lett.*, 574, 3545–3556.

31. Lim,K.W., Ng,Y.C., Martin-Pintado,N., Heddli,B. and Phan,A.T. (2013) Structure of the human telomere in Na+ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Res.*, 41, 10556–10562.

32. Perez,A., Marchan,I., Svozil,D., Sponer,J., Cheatham,T.E. 3rd, Zgarbova,M., Luque,F.J., Sponer,J., Cheatham,T.E. 3rd, Otyepka,M., Palumbo,M. and Richter,S.N. (2013) Solution structure of the dimeric quadruplex formed from the Oxytricha telomeric sequence. *J. Mol. Biol.*, 426, 221–233.

33. 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 GGGGCC G and GCCG C tetrads formed by G-G-G-C repeats observed in adeno-associated viral DNA. *J. Mol. Biol.*, 282, 619–636.

34. Kettani,A., Kumar,R.A. and Patel,D.J. (1995) Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. *J. Mol. Biol.*, 254, 638–656.

35. Zivasnik,J., Podbevsek,P. and Plavec,J. (2011) Observation of water molecules within the bimolecular d(G3C3T4)(G3C3)(G-quadruplex). *Biochemistry*, 50, 4155–4161.

36. Komav,C. and Patel,D.J. (2014) A tetrahedral DNA fold adopted by tandem repeats of alternating GGG and GCG tracts. *Nat. Commun.*, 5, 5831.

37. Wehsa da Silva,M. (2007) Geometric formalism for DNA quadruplexes. *Chem. Eur. J.*, 13, 9738–9745.

38. Marusic,M., Sket,P., Bauer,L., Viglasky,V. and Plavec,J. (2012) Solution-state structure of an intramolecular G-quadruplex with propeller, diagonal and edgewise loops. *Nucleic Acids Res.*, 40, 6946–6956.

39. Hardin,C.C., Corregan,M., Brown,B.A. 2nd and Frederick,L.N. (1993) Cytosine-cytosine+ base pairing stabilizes DNA quadruplexes and cytosine methylation greatly enhances the effect. *Biochemistry*, 32, 5870–5880.

40. Russ,J., Liu,E.Y., Wu,K., Neal,D., Suh,E., Irwin,D.J., McMillan,C.T., Harms,M.B., Cairns,N.J., Wood,E.M. et al. (2015) Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. *Acta Neuropathol.*, 129, 39–52.

41. Liu,E.Y., Russ,J., Wu,K., Neal,D., Suh,E., McNally,A.G., Irwin,D.J., Van Deerlin,V.M. and Lee,E.B. (2014) C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. *Acta Neuropathol.*, 128, 525–541.

42. Xi,Z., Zhang,M., Bruni,A.C., Maletta,R.G., Colao,R., Fratta,P., Polke,J., Sweeney,M., Suddenohwo,E., Nacmias,B. et al. (2015) The C9orf72 repeat expansion itself is methylated in ALS and FTLD patients. *Acta Neuropathol.*, 129, 715–727.

43. Oberle,J., Rousseau,F., Heitz,D., Kretz,C., Devys,D., Hanauer,A., Boue,J., Bertheas,M.F. and Mandel,J.L. (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science*, 252, 1097–1102.

44. Pieretti,M., Zhang,F.P., Fu,Y.H., Warren,S.T., Oostra,B.A., Caskey,C.T. and Nelson,D.L. (1991) Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell*, 66, 817–822.

45. Hansen,R.S., Canfield,T.K., Lamb,M.M., Gartler,S.M. and Laird,C.D. (1993) Association of fragile X syndrome with delayed replication of the FMR1 gene. *Cell*, 73, 1403–1409.

46. Fry,M. and Loeb,L.A. (1994) The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahedral structure. *Proc. Natl. Acad. Sci. U.S.A.*, 91, 4950–4954.

47. Kang,C., Zhang,X., Ratliff,R., Moynis,R. and Rich,A. (1992) Crystal structure of four-stranded Oxytricha telomeric DNA. *Nature*, 356, 126–131.

48. Abou-El,F., Marcher,A.I., Norman,D.G. and Lilley,D.M. (1994) Solution structure of a parallel-stranded G-quadruplex formed by d(TG4T) in the presence of sodium ions by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.*, 243, 458–471.

49. Laughlan,G., Marcher,A.I., Norman,D.G., Moore,M.H., Moody,P.C., Lilley,D.M. and Luiz,B. (1994) The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science*, 265, 520–524.

50. Wang,Y. and Patel,D.J. (1993) Solution structure of a parallel-stranded G-quadruplex DNA. *J. Mol. Biol.*, 234, 1171–1183.

51. Schulzle,P., Smith,F.W. and Feigon,J. (1994) Refined solution structure of the dimeric quadruplex formed from the Oxytricha telomeric oligonucleotide d(GGGGTGTTGGG). *Structure*, 2, 221–233.

52. Wang,Y. and Patel,D.J. (1995) Solution structure of the Oxytricha telomeric repeat d(G4[TAG4]) G-tetraplex. *J. Mol. Biol.*, 251, 76–94.
58. Nielsen, J.T., Arar, K. and Petersen, M. (2009) Solution structure of a locked nucleic acid modified quadruplex: introducing the V4 folding topology. *Angew. Chem. Int. Ed. Engl.*, **48**, 3099–3103.

59. Wang, Y. and Patel, D.J. (1994) Solution structure of the Tetrahymena telomeric repeat d(T2G4)4 G-tetraplex. *Structure*, **2**, 1141–1156.

60. Tong, X., Lan, W., Zhang, X., Wu, H., Liu, M. and Cao, C. (2011) Solution structure of all parallel G-quadruplex formed by the oncogene RET promoter sequence. *Nucleic Acids Res.*, **39**, 6753–6763.

61. Schultze, P., Macaya, R.F. and Feigon, J. (1994) Three-dimensional solution structure of the thrombin-binding DNA aptamer d(GGTTGGTGGTTG). *J. Mol. Biol.*, **235**, 1532–1547.

62. Macaya, R.F., Schultze, P., Smith, F.W., Roe, J.A. and Feigon, J. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 3745–3749.