Published online 6 January 2016

Nucleic Acids Research, 2016, Vol. 44, No. 5 ***2409–2416***
doi: 10.1093/nar/gkv1534

Structural characterization of a dimer of RNA duplexes composed of 8-bromoguanosine modified CGG trinucleotide repeats: a novel architecture of RNA quadruplexes

Dorota Gudanis[1], Lukasz Popenda[2], Kamil Szpotkowski[1], Ryszard Kierzek[1] and Zofia Gdaniec[1,*]

[1]Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61-704 Poznan, Noskowskiego 12/14, Poland and [2]NanoBioMedical Centre, Adam Mickiewicz University, 61-614 Poznan, Umultowska 85, Poland

Received October 14, 2015; Revised December 23, 2015; Accepted December 23, 2015

ABSTRACT

Fragile X syndrome and fragile X-associated tremor/ataxia syndrome (FXTAS) are neurodegenerative disorders caused by the pathogenic expansion of CGG triplet repeats in the FMR1 gene. FXTAS is likely to be caused by a 'toxic' gain-of-function of the FMR1 mRNA. We provide evidence for the existence of a novel quadruplex architecture comprising CGG repeats. The 8-bromoguanosine (^{Br}G)-modified molecule GC^{Br}GGCGGC forms a duplex in solution and self-associates via the major groove to form a four-stranded, antiparallel (GC^{Br}GGCGGC)$_4$ RNA quadruplex with ^{Br}G3:G6:^{Br}G3:G6 tetrads sandwiched between mixed G:C:G:C tetrads. Self-association of Watson–Crick duplexes to form a four-stranded structure has previously been predicted; however, no experimental evidence was provided. This novel four-stranded RNA structure was characterized using a variety of experimental methods, such as native gel electrophoresis, NMR spectroscopy, small-angle X-ray scattering and electrospray ionization mass spectrometry.

INTRODUCTION

Trinucleotide repeats represent the most abundant type of simple sequence repeats in eukaryotic genoms and in transcripts. Expanded tracts of repeated triplet sequences are implicated in the pathogenesis of human neurological diseases, which are known under a common name of triplet repeat expansion diseases (TREDs) (1). In several TREDs, stable RNA structures formed by triplet repeats present in untranslated regions of the mRNA are responsible for pathogenesis. The CGG triplet repeats, which have been found within the 5′-UTR of the FMR1 gene are polymorphic in length in a normal population and may undergo pathogenic expansions responsible for two diseases known as fragile X-associated tremor/ataxia syndrome and fragile X syndrome (2,3).

We have recently shown that the structures of RNA molecules composed of CGG repeats depend on both the number of repeats and the identity of the cation used (4). Additionally, we demonstrated that two repeats CGG sequence, G(CGG)$_2$C, folded exclusively into G-quadruplex in K^+ solution; however, a G-quadruplex/duplex equilibrium was observed in the presence of Na^+ ions. This equilibrium was strongly dependent on RNA concentration. In ^{1}H NMR spectra, signals corresponding to the duplex form notably increased and the duplex emerged as the dominant form when RNA concentration was decreased. In the duplex, the presence of a broad resonance at 10.6 ppm in ^{1}H NMR spectrum recorded in Na^+-containing solution was characteristic of the formation of the dynamic G:G base pairs, where the G's swap positions between *syn*/*anti* conformation. In the crystal structure of G(CGG)$_2$C, 18 distinct A-form duplexes arranged in end-to-end manner, in semi-infinite columns, were found in the unit cell (5). In these duplexes G:C Watson–Crick base-pairs were separated by G:G mismatches of G(carbonyl):G(N1) - G(N7):G(amino) type. Our earlier studies on the thermodynamic stability of RNA structures formed by CNG trinucleotide repeats revealed that stability of the G(CGG)$_2$C duplex increased by 3.7 kcal/mol ($\Delta\Delta G^\circ_{37}$) when 8-bromoguanosine was inserted at the third position (6). This finding was attributed to the stabilizing effect of 8-bromoguanosine on G:G mismatch, because 8-bromoguanosine which adopts exclusively *syn* conformation excludes the unfavorable G(*anti*):G(*anti*) interactions. The formation of ^{Br}G(*syn*):G(*anti*) base pairs was later confirmed by X-ray study (PDB code 3R1E) (5). In the asym-

*To whom correspondence should be addressed. Tel: +48 61 8528503; Fax: +48 61 8520532; Email: zgdan@ibch.poznan.pl

metric unit of crystal of $(GC^{Br}GGCGGC)_2$ only one duplex was found with well-ordered $^{Br}G(syn)$:G(*anti*) pairs, in accordance with the observed enhanced thermodynamic stability for the 8-bromoguanosine modified molecule.

In addition to the above-discussed structural data providing model or high resolution structures, it was shown using several biochemical and biophysical methods that longer RNA molecules comprising seventeen or twenty CGG repeats fold into hairpins (7–9). The presence of AGG interruptions that occur between about every nine or ten CGGs in normal length tracts can influence their structures. It has been demonstrated that incorporation of even a single AGG interruption within CGG tracts resulted in the formation of branched hairpin structures (7–9). On the other hand, for $r(CGG)_9AGG(CGG)_{12}AGG(CGG)_{97}$ an evidence was presented that besides a number of different hairpin structures this molecule can also fold into intramolecular quadruplex (10). The ability of midrange premutation tracts of $(CGG)_{30–99}$, to form quadruplexes was also reported (11;12). Using quadruplex destabilizing proteins, CBF-A and hnRNP A2, the efficiency of translation of the reporter firefly luciferase gene with upstream positioned $(CGG)_{30–99}$ repeats was studied. The observed enhancement of $(CGG)_{30–99}$ mRNA translation by hnRNP A2 and CBF-A was explained by the quadruplex destabilizing activity of these proteins.

In the present study, we have focused on the structural preferences of 8-bromoguanosine (^{Br}G) modified RNA molecule, $GC^{Br}GGCGGC$, in the presence of Na^+ ion.

MATERIALS AND METHODS

RNA synthesis and purification

All oligoribonucleotides used were chemically synthesized on the BioAutomation MM12 synthesizer using β-cyanoethyl phosphoramidite chemistry on a solid support on the 1 μmol scale (13). The phosphoramidite of 8-bromoguanosine was synthesized according to the previously published procedure (14). After synthesis, the oligoribonucleotides were removed from the solid support and deprotected by treatment with concentrated ammonia in ethanol (2 ml, 3:1, v/v) at 25°C for 2 days. The soluble oligonucleotides were separated from the support and then the support was rinsed three times with water. The solutions of oligonucleotides were dried using CentriVap (Labconco). Subsequently, the silyl-protecting groups were removed by incubating at 50°C in a 9:1 TEA-3HF (triethylamine trihydrofluoride)/DMF solution for 3 h. Next, the samples were precipitated by an addition of 5 ml of 1-buthanol and stored at −20°C for 1 h. The precipitate was separated from the solution by spinning at 5000 rpm, 4°C by 10 min. The deprotected oligonucleotides were desalted using a Sep-Pak C18 cartridge (Waters) and purified on a large preparative Whatman TLC plate (20 cm x 20 cm) using a 55:35:10 1-propanol/ammonia/water mobile phase. The main product bands were visualized by UV shadowing, scraped from the plate, extracted with water (3 × 2 ml) and dried using CentriVap. Finaly, the oligonucleotides were desalted again with a Sep-Pak C18 cartridge. The purity of the oligonucleotides was verified by MALDI MS.

NMR sample preparation

The $GC^{Br}GGCGGC$ [Br-(CGG)] oligonucleotide was dissolved in a buffer containing 150 mM NaCl, 10 mM sodium phosphate, pH 6.8, 0.1 mM EDTA or 75 mM KCl, 10 mM potassium phosphate, pH 6.8, 0.1 mM EDTA and placed in a Shigemi tube or in the 3 mm thin wall tube. The oligonucleotide was annealed by heating to 90°C and then slowly cooled down to the room temperature. The final concentration of Br-(CGG) was 2.6 mM (unless otherwise indicated). A mixture of 90% H_2O and 10% D_2O was used for experiments undertaken to study exchangeable protons. For experiments carried out in D_2O, the oligonucleotide was dried from D_2O three times and redissolved in 99.996% D_2O.

NMR spectroscopy

NMR data were recorded either on a Bruker AVANCE III 600 MHz spectrometer or AVANCE III 700 MHz spectrometer equipped with a QCI CryoProbe. For spectra recorded in 90% H_2O /10% D_2O water signal was suppressed using the 3–9–19 WATERGATE pulse sequence or excitation sculpting with gradient pulse. For spectra recorded in D_2O residual water signal was suppressed using the low-power presaturation. The data were processed with TopSpin 3.0 (Bruker BioSpin Gmbh) software and analyzed with CARA (http://cara.nmr.ch) or FELIX (Felix NMR, Inc.) software. The assignment of exchangeable resonances was based on homonuclear ^1H-^1H NOESY and heteronuclear ^1H-^{15}N HSQC experiments. The ^1H-^{15}N HSQC spectrum was collected from 1680 scans at 15°C. On average, 2048 complex points and 128 FIDs were collected within the spectral width of 13227 Hz (^1H) and 7298 Hz (^{15}N). The ^1H-^1H NOESY spectrum in 90% H_2O/10% D_2O was collected from 360 scans with 150 ms mixing time at 5°C. 2048 complex points and 512 FIDs were collected within the spectral width of 14097 Hz. The ^1H-^1H NOESY spectra in D_2O were recorded with mixing times of 80, 150 and 400 ms using 64 scans. 2048 complex points and 512 FIDs were collected within the spectral width of 4800 Hz.

Native gel electrophoresis

2.8 nmol of GCCGCCGC, GCACGUGC, GCGGCGGC and $GC^{Br}GGCGGC$ were dissolved in the 12 μl of solution containing 150 mM sodium chloride, 10 mM sodium phosphate and 0.1 mM EDTA at pH 6.8. Each sample was treated with heat denaturation, renaturation and was stored at 25°C at least for 12 h. Then 3 μl of 30% glycerol was added before loading onto gel. Native gel electrophoresis (PAGE) was run on 15% polyacrylamide gel (acrylamide/bisacrylamide, 29:1) at 4°C, in 0.25X TBE buffer. The bands were visualized using *Stains-All* dye.

Electrospray mass spectrometry

The sample was prepared in 280 μM concentration of RNA (single strand) in 150 mM NH_4OAc. The sample was annealed by heating at 90°C for 5 min and then slowly cooled down to the room temperature and stored at 4°C. About 10 min before the measurement methanol was added to the final concentration 10% v/v.

The mass spectrum was obtained using a Bruker micrOTOF-Q mass spectrometer. The instrument was operated using an electrospray source in negative ion mode. The ionization source parameters were as follows: drying gas flow in the ion source was set up to 4 l/min and nebulizer gas pressure to 0.6 bar (both were N_2), the temperature of ion source was 30°C, voltage was 3.4 kV, collision energy was 10 eV. The equation: $m/z = (n \times M_{NA} + t \times 17 - z)/z$ was used to calculate the structural parameters of Br-(CGG) (n: the number of strands, M_{NA}: the molecular weight of neutral nucleic acid in daltons, t: the number of NH_4^+ ions, z: the charge of the ion).

Small angle X-ray scattering (SAXS)

SAXS experiments were performed at beamline P12 of Petra III storage ring. Twenty microliters of the sample and corresponding matching buffer were loaded into a 96-well plate. Data collection was performed at 15°C. Integration, scaling and buffer subtraction were accomplished using the program Primus (15). SAXS data were collected over the s range of 0.0088–4 nm^{-1}. Overlays of the merged data sets were used to detect concentration dependent scattering in the lowest s region. Indirect Fourier transform of the SAXS scattering curve was performed with Gnom (16). SAXS profile with the lowest noise and free of interparticle interference was subsequently used for *ab initio* modeling with the program DAMMIN. For each refined and merged SAXS curve, multiple independent DAMMIN runs were performed, superimposed and averaged with the program DAMAVER (17). Theoretical calculation of radii gyration and SAXS curves were performed using CRYSOL (18).

RESULTS

Our interest in establishing structural preferences of GC^{Br}GGCGGC [Br-(CGG)] in solution was derived from the observation of its unexpectedly slow migration in native gel electrophoresis. Figure 1 compares the migration of Br-(CGG) with G(CGG)$_2$C, G(CCG)$_2$C, and with a self-complementary duplex of similar sequence but with A:U Watson–Crick base-pairs instead of G:G mismatches (GCAGCUGC)$_2$. The duplex with all canonical base-pairs was used as size marker. These data indicate that in conditions of gel electrophoresis, G(CGG)$_2$C and GCAGCUGC form duplexes, G(CCG)$_2$C exists in a single stranded form and slow migration of Br-(CGG) strongly suggests the formation of higher order structure.

In order to get insight into the conformation of Br-(CGG) in solution we have used NMR spectroscopy. The imino region of ^{1}H NMR spectrum of the Br-(CGG) is shown in Figure 2. The presence of one set of imino proton resonances indicates the formation of one main conformation. The guanosine imino protons resonating in the range of 12.0–13.5 ppm are typical of Watson–Crick G:C base pairs and those between 10.2 and 11.0 ppm are characteristic of G:G mismatches. The assignment of the high-field resonances to guanosine imino protons was confirmed by the analysis of ^{15}N chemical shifts of attached nitrogen in ^{1}H-^{15}N HSQC spectrum recorded at 15°C (Supplementary Figure S1). The 2D NOESY spectrum acquired in H_2O/D_2O (90%/10%) at 5°C (Figure 3) was employed in the assignment of the exchangeable protons. The observation of a characteristic pattern of the cross-peaks between imino protons of guanosine and their respective cytidine partners allowed unambiguous identification of the guanosine imino and cytidine amino protons associated with G1:C8, G4:C5 and G7:C2 Watson–Crick base pairs (19). The resonance at 10.39 ppm was attributed to the imino proton of G6 based on the observation of NOEs to imino protons of G7 and G4 as well as the interaction of G6-NH1 with G7-H8 and G7-H1'. The imino proton at 10.92 ppm showed NOEs to C2-H6, the imino proton of G7, amino protons of C2 and, additionally, a strong NOE to amino protons at 6.46 ppm, characteristic of intraresidue imino-amino interactions. Accordingly, this remaining imino proton could be

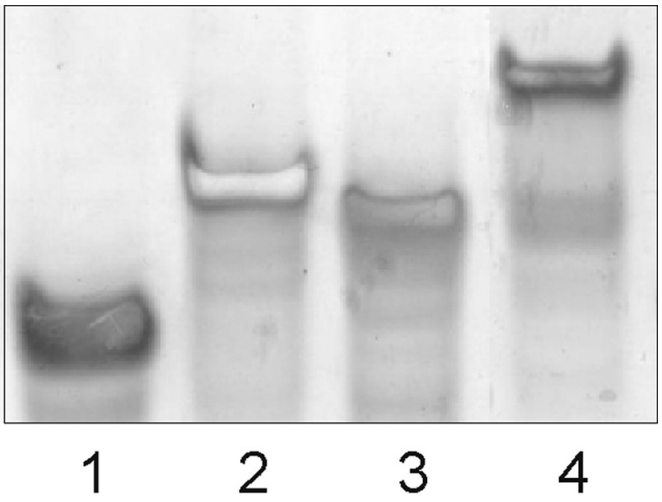


Figure 1. Comparison of gel electrophoretic mobility for the G(CCG)$_2$C (lane 1), GCACGUGC (lane 2), G(CGG)$_2$C (lane 3) and Br-(CGG) (lane 4).

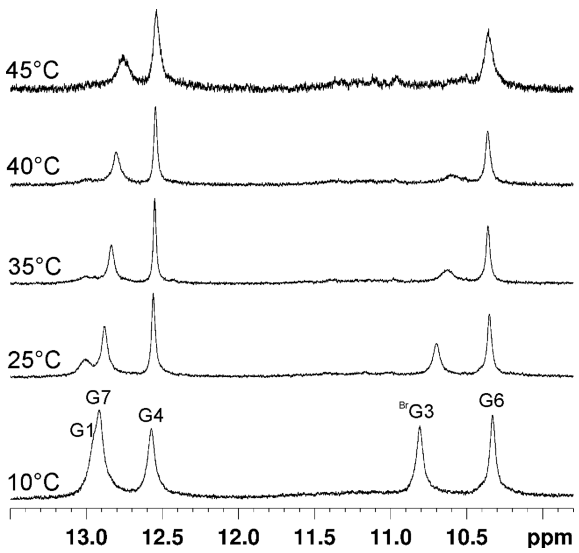
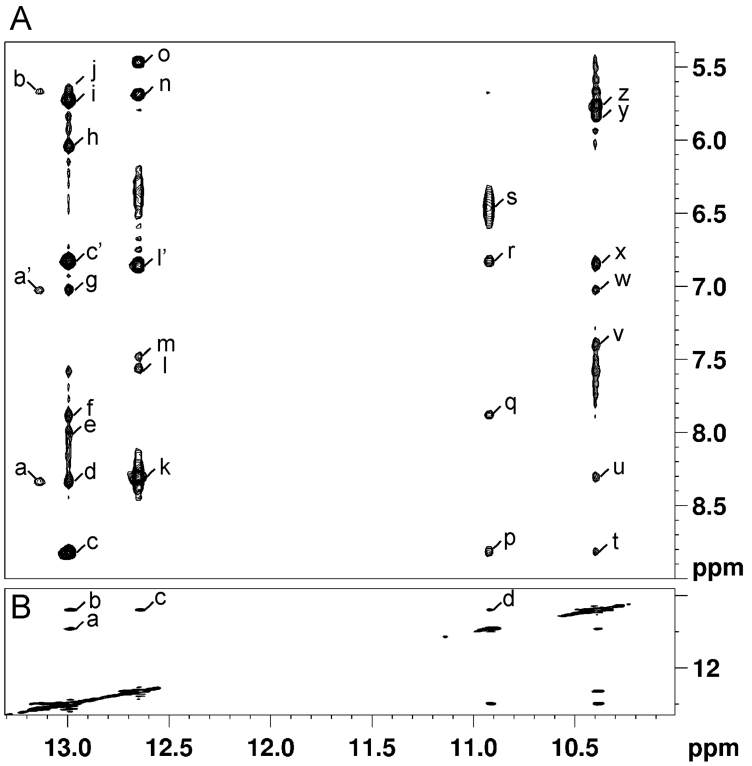


Figure 2. Temperature dependence of the imino region of the ^{1}H NMR spectrum of Br-(CGG) (2.6 mM). Spectra were recorded in 90% H_2O/10% D_2O (v/v) in the presence of 150 mM NaCl, 10 mM sodium phosphate and 0.1 mM EDTA, pH 6.8.



A	**a**	G1:NH1 - C8:NH4$_1$	**g**	G7:NH1 - C8:NH4$_2$	**m**	G4:NH1- G6:H8	**s**	G3:NH1-G3:NH2$_{1,2}$
	a'	G1:NH1 - C8:NH4$_2$	**c'**	G7:NH1 - C2:NH4$_2$	**l'**	G4:NH1- C5:NH4$_2$	**t**	G6:NH1- C2:NH4$_1$
	b	G1:NH1 - C2:H1'	**h**	G7:NH1 - G3:H1'	**n**	G4:NH1- G6:H1'	**u**	G6:NH1- C5:NH4$_1$
	c	G7:NH1 - C2:NH4$_1$	**i**	G7:NH1 - C8:H1'	**o**	G4:NH1- C5:H1'	**v**	G6:NH1-G7:H8
	d	G7:NH1 - C8:NH4$_1$	**j**	G7:NH1 - C2:H1'	**p**	G3:NH1- C2:NH4$_1$	**w**	G6:NH1-G4:H8
	e	G7:NH1 - G1:H8	**k**	G4:NH1 - C5:NH4$_1$	**q**	G3:NH1- C2:H6	**x**	G6:NH1- C5:NH4$_2$
	f	G7:NH1 - C2:H6	**l**	G4:NH1 - C5:H6	**r**	G3:NH1- C2:NH4$_2$	**y**	G6:NH1- G4:H1'
	z	G6:NH1 - G7:H1'						
B	**a**	G7:NH1-G3:NH1	**b**	G7:NH1-G6:NH1	**c**	G4:NH1-G6:NH1	**d**	G3:NH1-G6:NH1

Figure 3. Imino-imino and imino-amino regions of 2D NOESY spectrum (250 ms mixing time) of 1.3 mM Br-(CGG) in the presence of 150 mM NaCl, 10 mM sodium phosphate and 0.1 mM EDTA, pH 6.8 at 5°C. The cross peaks a to z (**A**) and a to d (**B**) are assigned as above.

assigned to ^{Br}G3. The presence of these two relatively sharp high-field imino resonances suggested the formation of stable ^{Br}G3:G6 mismatch and was confirmed by the observation of the imino-imino interactions involving ^{Br}G3-NH1 and G6-NH1 with adjacent G4:C5 and G7:C2 base pairs (peaks: a, b, c, Figure 3B).

The non-exchangeable protons have been assigned in 150 mM NaCl containing D_2O solution at 25°C following well established procedures (19). A complete list of resonance assignments is given in Supplementary Table S2 of the Supporting Material. The region of 2D NOESY spectrum corresponding to the interactions between base H6/H8 protons and H1'/H5 protons is shown in Supplementary Figure S2. Continuous set of sequential H6/H8-H1' NOE connectivities can be traced with a break occurring between C2 and ^{Br}G3 step due to the lack of proton at C8 position of the ^{Br}G3. Structural studies on nucleosides and oligonucleotides have shown that 8-bromoguanosine preferentially adopts the *syn* conformation. The *syn* conformation of ^{Br}G3 residue was confirmed by the analysis of ^{13}C chemical shift of C1' sugar carbon. The characteristic feature of the *syn* conformation of guanosine residues is a downfield shift of C1' relative to guanosine residues adopting *anti* conformation (20). Indeed, about 3 ppm downfield chemical shift of ^{Br}G3-C1' was observed relative to other residues in ^{1}H-^{13}C HSQC spectrum (Supporting Material, Supplementary Figure S3).

In ^{31}P NMR spectrum, most of the Br-(CGG) resonances is dispersed in a narrow region between −3.5 and −4.5 ppm (Supplementary Figure S4) typical of the regular helix con-

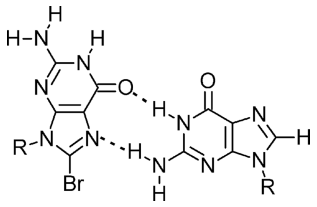


Figure 4. Non-canonical ^{Br}G(*syn*):G(*anti*) base pair. Hydrogen bonds are drawn with dashed lines.

formation. Only the ^{31}P signal corresponding to C2 - ^{Br}G3 step is shifted downfield to −2.95 ppm indicating local distortion of the phosphodiester backbone. Contrary to the results obtained from the native gel electrophoresis which suggested the formation of higher order structure of Br-(CGG), most of the NMR data were consistent with the formation of one conformation, most probably a duplex with ^{Br}G3:G6 mismatch.

Initially, by the analogy to the X-ray structure, we presumed that the Br-(CGG) duplex accommodates two N1-carbonyl, N7-amino type of base pairs. Analysis of the geometry of this type of G:G base pair reveals that ^{Br}G3 which is in *syn* conformation engages its Hoogsteen edge in pairing with guanosine G6 and exposes a Watson–Crick face to the solvent (Figure 4). In this arrangement only the imino proton of guanosine G6, which favors *anti* conformation, is hydrogen-bonded. Thus, the observation of both imino protons of G:G mismatch at elevated temperatures (Figure 2) was highly unexpected and suggested that ^{Br}G3-H1 proton must be somehow protected from the exchange with solvent. This was further supported by the relatively small negative value of thermal coefficient of −7.5 ppb/1°C (more downfield chemical shift as a function of decreased temperature) determined for ^{Br}G3 imino proton, what suggested that it can be engaged in a hydrogen bonding interactions (21).

In order to explain this observation, and given the slow migration of Br-(CGG) in a PAGE gel, we hypothesized that in solution the duplex self-associates via its major groove forming a four-stranded structure composed of ^{Br}G3:G6:^{Br}G3:G6 tetrads sandwiched between mixed G:C:G:C tetrads. In the major groove of ^{Br}G3(*syn*):G6(*anti*) base pair there are two donor (Watson–Crick edge of ^{Br}G3) and two acceptor sites (Hoogsteen edge of G6) of hydrogen bonds (Supplementary Figure S5A). Thus, two ^{Br}G3:G6 base pairs in this geometry have the proper orientation needed to form additional hydrogen bonds that would form a ^{Br}G3:G6:^{Br}G3:G6-tetrad (Supplementary Figure S5B). The formation of such a G-tetrad, typical of the G-quadruplex structure, could explain the relatively slow exchange of the ^{Br}G3-NH1 proton at elevated temperatures. In the ^{Br}G3:G6:^{Br}G3:G6 tetrad, the ^{Br}G3-NH1 proton would be involved in a hydrogen bond with O6 of the G6 residue and protected from exchange. Moreover, self-association of Br-(CGG) duplex would lead to the face-to-face alignment of the major groove edges of the Watson–Crick G:C pairs to form the G:C:G:C tetrads. As a consequence, the resulting structure would be an antiparallel, tetramolecular quadruplex (Figure 5A) with two ^{Br}G3:G6:^{Br}G3:G6 tetrads and six mixed G:C:G:C tetrads (Figure 5B,C). In contrast with all tetramolecular quadruplexes described so far, Br-(CGG) quadruplex is built of one-layer G-tetrads sandwiched by mixed G:C:G:C tetrads. The arrangement of the tetrads is similar to that found in DNA quadruplex containing elements of the fragile-X syndrome repeat (20).

Next, we analyzed 2D NOESY spectrum recorded in 90% H_2O/10% D_2O (v/v) to identify resonances characteristic of ^{Br}G3:G6:^{Br}G3:G6 and mixed C:G:C:G tetrads (Figure 3, Supplementary Figure S6). However, we were able to find only a few cross-peaks, which corroborated the formation of the quadruplex. For example, a cross peak of small intensity was detected between G6-NH1 and ^{Br}G3-NH1 (Figure 3B). Nonetheless, we were not able to identify the NOE between imino proton of ^{Br}G3 and H8 of G6, which is an interaction characteristic of a G-tetrad, in any conditions of temperature or RNA concentration range between 1.1 mM and 2.6 mM. The mixed G:C:G:C tetrads that are aligned via the major groove edges of the Watson–Crick base pairs can be arranged in a direct or a slipped manner. In the direct alignment the interaction between the base pairs occurs through two bifurcated amino(G)-O2(C) hydrogen bonds and in the slipped alignment the G:C base pairs are sheared along their hydrogen bonding axis (Supplementary Figure S7). The expected NOE contacts vary depending on the type of tetrad, direct or slipped (Supplementary Table S1) (22). Because for the G1:C8:G1:C8 tetrad we found only cross-peaks between the H8 proton of G1 and the amino proton of C8 (Supplementary Figure S6, peaks: f, l), but no NOE was detected between the H8 protons of G1 and H5 protons of C8, this observation suggested the slipped arrangement of the G1:C8:G1:C8 tetrad (Figure 5C). No signals typical of mixed G:C:G:C tetrads were found for the G7:C2:G7:C2 and G4:C5:G4:C5 tetrads. This could reflect the dynamic nature of the structure with a transient network of hydrogen bonds between duplexes, which can be reversibly broken and re-formed.

Circular dichroism which is widely used to characterize the structures of DNA and RNA quadruplexes, did not provide us additional information to support the hypothesis of self-association of RNA duplexes. In CD spectra positive bands at 260 nm or 295 nm are related to quadruplexes comprising a minimum of two adjacent G-tetrads, having parallel or antiparallel phosphodiester backbone, respectively. However, the Br-(CGG) sequence forms a single G-tetrad, which fails to give CD signal characteristic of antiparallel quadruplexes (Supplementary Figure S8A) (4,23). For the same reason we did not observe inverted melting profile at 295 nm in UV spectrum, a feature distinctive of quadruplexes.

Instead, we used mass spectrometry to obtain further evidence for a quadruplex formation (24). At present, electrospray ionization mass spectrometry (ESI-MS) is considered as one of the most reliable techniques for determining the number of strands of nucleic acid structures. For G-quadruplexes there may be one, two or four strands involved. The ESI-MS spectrum of Br-(CGG) shown in Supplementary Figure S9 unambiguously reveals the presence of bimolecular and tetramolecular species. The major peak

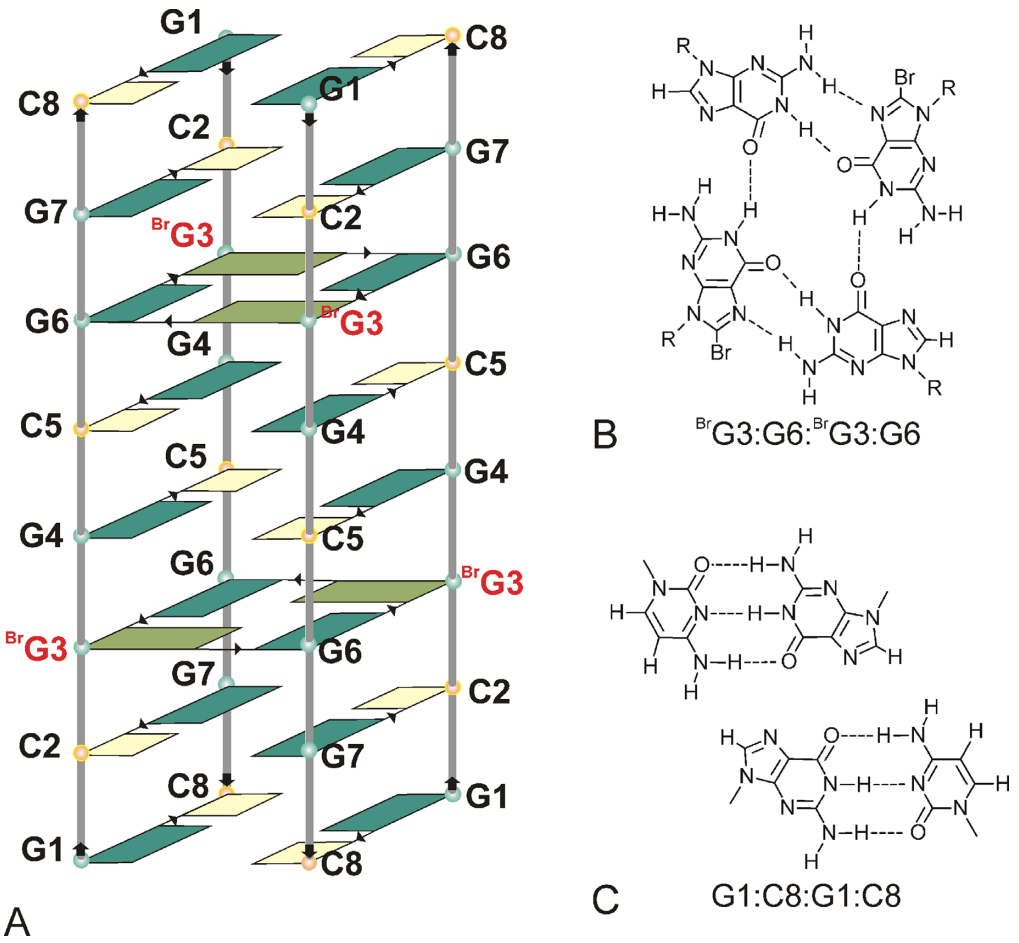


Figure 5. Schematic outline of the Br-(CGG) fold (**A**), ^{Br}G:G:^{Br}G:G-tetrad (**B**) and slipped G:C:G:C-tetrad (**C**).

at m/z 1328.10 corresponds to a bimolecular structure in charge state 4- with no ammonium ions. The minor peak at m/z 1782.68 and in charge states 3- is also consistent with a bimolecular structure but with two trapped ammonium ions. The minor peak which appears at m/z 2139.45 corresponds to the tetramolecular structure in charge state 5- in an adduct with three ammonium ions. The latter peak strongly supports the dimerization of duplexes. The optimization of experimental conditions was a major challenge in determining the stoichiometry of Br-(CGG) using ESI-MS. In our studies, due to the presence of weak or transient hydrogen bonds between duplexes, the use of the most gentle conditions possible was required in order to ensure preservation of the tetramolecular structure during ionization.

Once we got additional evidence supporting the dimerization of duplexes, we decided to determine the three dimensional structure of Br-(CGG) in solution. The structure was calculated by applying the previously published procedure (25,26) using a set of 736 NOE and 296 dihedral experimental constraints obtained from the analysis of NMR spectra (see Model calculation paragraph in Supplementary Material). The only constraints across duplexes were introduced based on the distance between imino protons within G1:C8:G1:C8 tetrad in accord with the NOE observations. The resulting structure (Figure 6) consists of four strands arranged in an antiparallel fashion. Self-association of two antiparallel RNA duplexes with ^{Br}G:G mismatches

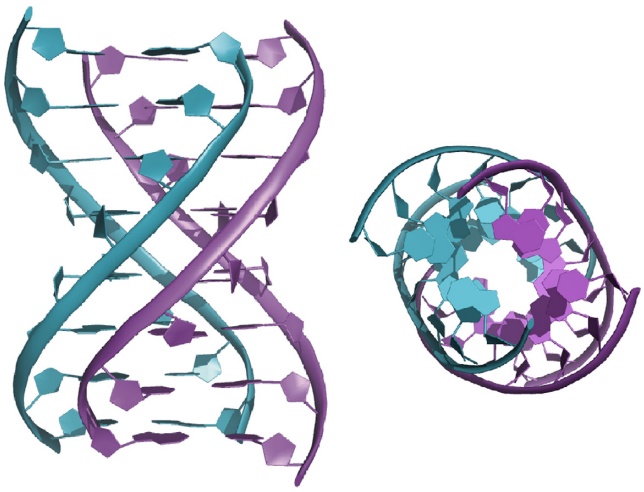


Figure 6. NMR structure of the Br-(CGG) quadruplex in Na^+ solution shown in side and top view. The individual duplexes are colored in magenta and cyan.

sandwiched between G:C Watson–Crick base pairs yields a quadruplex with two very wide and two very narrow grooves. The narrow grooves are made up of two strands in antiparallel orientation. The two backbone chains fit into each other remarkably well in a zig-zag fashion (Supplementary Figure S10). The average interchain P–P distances across the narrow groove is ≈5.6 Å and the average interstrand P–P distance across the wide groove is ≈18.5 Å. The formation of such a narrow groove is conformationally feasible and has been previously observed, for example in I-motif (27). Supplementary Figure S11 shows stacking interactions between bases. Except the ^{Br}G3:G6:^{Br}G3:G6 - G4:C5:G4:C5 step (Supplementary Figure S11C) the stacking interactions occur only between base pairs of the individual duplexes. This suggests that stabilization of this structure results from a balance of multiple factors including hydrogen bonding within the tetrads created from G:C and G:G base pairs, base stacking, hydration and interaction with Na^+ ions. Previous molecular dynamics simulations studies have shown that interactions between monovalent cations and mixed G:C:G:C tetrads are quite favorable although less attractive than interactions between ions and all-guanine tetrads (28).

The final evidence supporting quadruplex formation was derived from SAXS analysis. The SAXS data collected for Br-(CGG) are presented in Supplementary Figure S12. A radius of gyration (Rg) of the four-stranded NMR structure and X-ray derived duplex (pdb code 3R1E) were compared in order to corroborate that the Br-(CGG) duplex dimerizes in solution. The calculated radii of gyration are 1.01 nm and 1.19 nm for duplex and quadruplex, respectively. The radius of gyration determined from experimental SAXS data is 1.27 nm and is in a good agreement with the radius of gyration calculated for the NMR structure (≈5% difference). A low resolution SAXS structure reproduces the shape of that obtained from NMR data very well. Superposition of both structures is presented in Supplementary Figure S13.

DISCUSSION

The self-association of duplexes was observed regardless of the type of cation, sodium or potassium (Supplementary Figures S8 and S14), RNA and salt concentration (Supplementary Figures S15 and S16). In the presence of potassium ions, however, the NMR spectrum revealed the formation of an additional, minor form. The signals of the minor form appeared in the region characteristic of G-tetrads and no signals typical of Watson–Crick G:C base-pairs were observed (Supplementary Figure S14). Moreover, in CD spectrum recorded in the presence of potassium ions no positive band at 295 nm (Supplementary Figure S8B) was detected. This suggested that the minor form corresponds to the formation of a quadruplex with all strands parallel to each other (Supplementary Figure S17).

In crystal, unlike the solution structure, Br-(CGG) forms A-type duplex with G:C Watson–Crick base-pairs separated by ^{Br}G(*syn*):G(*anti*) mismatches (5). To determine structural changes which occur upon the dimerization of Br-(CGG) duplexes we compared the helical parameters (29) of the X-ray duplex and the duplex derived from NMR quadruplex structure (Supplementary Table S3). The NMR derived duplex is longer which is reflected by the larger value of helical rise (Supplementary Figure S18A). The average value of the inclination (Supplementary Figure S18B) is smaller in NMR duplex. These two parameters are typical more for B-type duplexes than for A-type structure (Supplementary Table S4). Nevertheless, other parameters like helical twist, displacement and C3'-*endo* conformation of the sugars are characteristic of A-type structure. Although the NMR and X-ray structures of duplexes differ (Supplementary Figure S19), this structural transformation between the two forms seems to be easily accomplished. It is this flexibility that makes RNA such a versatile molecule.

Except for typical G-quadruplex structures which are formed of stacks of G-tetrads, nucleic acids can fold into other tetrahelical structures. For example, it has been shown recently that in solution the G-rich VK1 sequence found in the regulatory region of the PLEKHG3 gene in the 14th human chromosome adopts tetrahelical structure with no G-tetrads but with four G:C, four G:A and six N1- carbonyl symmetric G:G base pairs (30). The stability of this unique structure is achieved through a fine balance of hydrogen bonding and aromatic stacking interactions. The solution structure of Br-(CGG) quadruplex presented in this work is equally unusual as it results from the self-association of two antiparallel RNA duplexes that contain ^{Br}G:G mismatches sandwiched between G:C Watson–Crick base pairs. As previously demonstrated by us, an unmodified oligonucleotide, $G(CGG)_2C$, can also fold into a quadruplex with G:G:G:G and mixed G:C:G:C tetrads; however, its architecture is entirely different. It forms a parallel, highly symmetric, interlocked tetramolecular G-quadruplex structure (4). Hence, it is evident that the antiparallel fold is facilitated by the presence of an 8-bromoguanosine residue, which first promotes the formation of a stable ^{Br}G:G mismatch and subsequently induces the formation of ^{Br}G3:G6:^{Br}G3:G6 tetrads. Our data indicate that self-association of duplexes composed entirely from G:C base pairs should be possible; however, it would be unspecific because many alternative alignments of such duplexes are feasible. In case of Br-(CGG), the formation of ^{Br}G3:G6:^{Br}G3:G6 tetrads is crucial and ensures the specificity of interactions between duplexes. It is worth mentioning that self-association of Watson–Crick duplexes into four-stranded structure has been previously predicted; however, no experimental evidence was given to support this thesis (31).

CONCLUSIONS

In summary, using native gel electrophoresis, NMR spectroscopy, small-angle X-ray scattering and ESI-MS, we have demonstrated that Br-(CGG) fold preferentially into antiparallel RNA quadruplex in solution containing sodium ions. The present study expands the repertoire of structures adopted by molecules composed of CGG repeats. The knowledge of these structures may become crucial for understanding the pathogenesis of FXTAS, where the structure of a RNA molecule is believed to play a significant role.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We would like to thanks SAXS beamline P12 of the EMBL at the storage ring Petra-3.

FUNDING

Ministry of Science and Higher Education of the Republic of Poland by KNOW program, National Science Center [UMO-2014/13/B/ST5/04144 to Z.G., UMO-2013/08/A/ST5/00295 to R.K.]; National Centre for Research and Development [PBS1/A9/13/2012 to L.P.]. Source of open access funding: National Science Center [UMO-2014/13/B/ST5/04144] and [UMO-2013/08/A/ST5/00295.

Conflict of interest statement. None declared.

REFERENCES

1. Orr,H.T. and Zoghbi,H.Y. (2007) Trinucleotide repeat disorders. *Annu. Rev. Neurosci.*, **30**, 575–621.
2. Glass,I.A. (1991) X-linked mental retardation. *J. Med. Genet.*, **28**, 361–371.
3. Jacquemont,S., Hagerman,R.J., Leehey,M., Grigsby,J., Zhang,L., Brunberg,J.A., Greco,C., Des Portes,V., Jardini,T., Levine,R. *et al.* (2003) Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am. J. Hum. Genet.*, **72**, 869–78.
4. Malgowska,M., Gudanis,D., Kierzek,R., Wyszko,E., Gabelica,V. and Gdaniec,Z. (2014) Distinctive structural motifs of RNA G-quadruplexes composed of AGG, CGG and UGG trinucleotide repeats. *Nucleic Acids Res.*, **42**, 10196–10207.
5. Kiliszek,A., Kierzek,R., Krzyzosiak,W.J. and Rypniewski,W. (2011) Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. *Nucleic Acids Res.*, **39**, 7308–7315.
6. Broda,M., Kierzek,E., Gdaniec,Z., Kulinski,T. and Kierzek,R. (2005) Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure. *Biochemistry*, **44**, 10873–10882.
7. Galka-Marciniak,P., Urbanek,M.O. and Krzyzosiak,W.J. (2012) Triplet repeats in transcripts: structural insights into RNA toxicity. *Biol. Chem.*, **393**, 1299–1315.
8. Krzyzosiak,W.J., Sobczak,K., Wojciechowska,M., Fiszer,A., Mykowska,A. and Kozlowski,P. (2012) Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res.*, **40**, 11–26.
9. Sobczak,K., de Mezer,M., Michlewski,G., Krol,J. and Krzyzosiak,W.J. (2003) RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res.*, **31**, 5469–5482.
10. Handa,V., Saha,T. and Usdin,K. (2003) The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by Dicer. *Nucleic Acids Res.*, **31**, 6243–6248.
11. Khateb,S., Weisman-Shomer,P., Hershco-Shani,I., Ludwig,A.L. and Fry,M. (2007) The tetraplex (CGG)n destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA. *Nucleic Acids Res.*, **35**, 5775–5788.
12. Ofer,N., Weisman-Shomer,P., Shklover,J. and Fry,M. (2009) The quadruplex r(CGG)n destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA. *Nucleic Acids Res.*, **37**, 2712–2722.
13. McBride,L.J. and Caruthers,M.H. (1983) An investigation of several deoxynucleosides phosphoramidites useful for synthesizing oligodeoxynucleotides. *Tetrahedron Lett.*, **24**, 245–249.
14. Proctor,D.J., Kierzek,E., Kierzek,R. and Bevilacqua,P.C. (2003) Restricting the conformational heterogeneity of RNA by specific incorporation of 8-bromoguanosine. *J. Am. Chem. Soc.* **125**, 2390–2391.
15. Konarev,P.V., Volkov,V.V., Sokolova,A.V., Koch,M.H.J. and Svergun,D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.*, **36**, 1277–1282.
16. Svergun,D.I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.*, **25**, 495–503.
17. Volkov,V.V. and Svergun,D.I. (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Cryst.*, **36**, 860–864.
18. Svergun,D.I., Barberato,C. and Koch,M.H.J. (1995) *CRYSOL* - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.*, **28**, 768–773.
19. Varani,G., Aboul-ela,F. and Allain,F.H.-T. (1996) NMR investigation of RNA structure. *Prog. Nucl. Magn. Reson. Spectrosc.*, **29**, 51–127.
20. 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.
21. Schroeder,K.T., Skalicky,J.J. and Greenbaum,N.L. (2005) NMR spectroscopy of RNA duplexes containing pseudouridine in supercooled water. *RNA*, **11**, 1012–1016.
22. Lim,K.W., Alberti,P., Guédin,A., Lacroix,L., Riou,J.-F., Royle,N.J., Mergny,J.-L. and Phan,A.T. (2009) Sequence variant (CTAGGG)n in the human telomere favors a G-quadruplex structure containing a G.C.G.C tetrad. *Nucleic Acids Res.*, **37**, 6239–6248.
23. Lane,A.N., Chaires,J.B., Gray,R.D. and Trent,J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, **36**, 5482–5515.
24. Gabelica,V. (2010) Determination of equilibrium association constants of ligand-DNA complexes by electrospray mass spectrometry. *Methods Mol. Biol.*, **613**, 89–101.
25. Brzezinska,J., Gdaniec,Z., Popenda,L. and Markiewicz,W.T. (2014) Polyaminooligonucleotide: NMR structure of duplex DNA containing a nucleoside with spermine residue, N-[4, 9, 13-triazatridecan-1-yl]-2′-deoxycytidine. *BBA-Gen Subjects*, **1840**, 1163–1170.
26. Popenda,L., Adamiak,R.W. and Gdaniec,Z. (2008) Bulged adenosine influence on the RNA duplex conformation in solution. *Biochemistry*, **47**, 5059–5067.
27. Cai,L., Chen,L., Raghavan,S., Rich,A., Ratliff,R. and Moyzis,R. (1998) Intercalated cytosine motif and novel adenine clusters in the crystal structure of the Tetrahymena telomere. *Nucleic Acids Res.*, **26**, 4696–4705.
28. Špačková,N., Berger,I. and Šponer,J. (2001) Structural Dynamics and Cation Interactions of DNA Quadruplex Molecules Containing Mixed Guanine/Cytosine Quartets Revealed by Large-Scale MD Simulations. *J. Am. Chem. Soc.*, **123**, 3295–3307.
29. Zheng,G., Lu,X.-J. and Olson,W.K. (2009) Web 3DNA - a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.*, **37**, W240–W246.
30. Kocman,V. and Plavec,J. (2014) A tetrahelical DNA fold adopted by tandem repeats of alternating GGG and GCG tracts. *Nat. Commun.*, **5**, 5831.
31. Neidle,S. and Balasubramanian,S. eds. (2006) *Quadruplex Nucleic Acids, Royal Society of Chemistry*, Cambridge.