10196–10207 *Nucleic Acids Research, 2014, Vol. 42, No. 15* *Published online 31 July 2014*
doi: 10.1093/nar/gku710

Distinctive structural motifs of RNA G-quadruplexes composed of AGG, CGG and UGG trinucleotide repeats

Magdalena Malgowska[1,†], Dorota Gudanis[1,†], Ryszard Kierzek[1], Eliza Wyszko[1], Valérie Gabelica[2,3,4] and Zofia Gdaniec[1,*]

[1]Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61–704 Poznan, Noskowskiego 12/14, Poland, [2]Laboratoire de Spectrométrie de Masse, Institut de Chimie, Bat. B6c, Université de Liège, B-4000 Liège, Belgium, [3]Inserm, U869 ARNA Laboratory, F-33000 Bordeaux, France and [4]University of Bordeaux, IECB, ARNA Laboratory, F-33600 Pessac, France

Received April 23, 2014; Revised July 07, 2014; Accepted July 22, 2014

ABSTRACT

Trinucleotide repeats are microsatellite sequences that are polymorphic in length. Their expansion in specific genes underlies a number of neurodegenerative disorders. Using ultraviolet-visible, circular dichroism, nuclear magnetic resonance (NMR) spectroscopies and electrospray ionization mass spectrometry, the structural preferences of RNA molecules composed of two and four repeats of AGG, CGG and UGG in the presence of K^+, Na^+ and NH_4^+ were analysed. $(AGG)_2A$, $(AGG)_4A$, $p(UGG)_2U$ and $p(UGG)_4U$ strongly prefer folding into G-quadruplexes, whereas CGG-containing sequences can adopt different types of structure depending on the cation and on the number of repeats. In particular, the two-repeat CGG sequence folds into a G-quadruplex in potassium buffer. We also found that each G-quadruplex fold is different: A:(G:G:G:G)A hexads were found for $(AGG)_2A$, whereas mixed G:C:G:C tetrads and U-tetrads were observed in the NMR spectra of $G(CGG)_2C$ and $p(UGG)_2U$, respectively. Finally, our NMR study highlights the influence of the strand sequence on the structure formed, and the influence of the intracellular environment on the folding. Importantly, we highlight that although potassium ions are prevalent in cells, the structures observed in the HeLa cell extract are not always the same as those prevailing in biophysical studies in the presence of K^+ ions.

INTRODUCTION

Guanine-rich DNA and RNA sequences can fold into inter- and intramolecular structures called G-quadruplexes. The fundamental unit of the G-quadruplex is a G-tetrad composed of a planar array of four guanine bases. G-quadruplex formation requires the presence of monovalent cations, in particular, K^+ or Na^+ (1–4). G-quadruplexes reveal remarkable thermodynamic and kinetic stability resulting from stacking interactions, hydrogen bonding, coordination of cations between the adjacent tetrads and hydration (5). The final topology of G-quadruplexes does not depend solely on their primary sequence, but is greatly affected by environmental conditions (6–9). Recently, many G-quadruplex structures have been reported to contain, apart from G-tetrads, other homo-tetrads (A-, T-, C- or U-tetrads) or mixed tetrads (G:C:G:C, G:T:G:T or A:T:A:T) (10). It has been demonstrated that even more than four nucleobases can be associated into a planar array of H-bonds, forming pentads, hexads, heptads or octads (10–14).

G-rich sequences occur in several biologically important DNA regions, such as telomeres, promoter regions of some human oncogenes and some variable number tandem repeats (15). Additionally, recent reports clearly indicate their existence *in vivo* (16,17). Also, a number of well-annotated transcripts of human genes have already been shown to contain G-rich sequences (18). Many more are still to be identified within the non-coding part of the transcriptome, which has been shown to cover over two-thirds of the human genome (19). Therefore, RNA G-quadruplexes are likely to be involved in various biological processes. For instance, it was shown that the G-quadruplex structures are implicated in specific protein binding and modulation of gene expression by regulating splicing, polyadenylation, translation and transcript turnover (20).

*To whom correspondence should be addressed. Tel: +48 61 8528503; Fax: +48 61 8520532; Email: zgdan@ibch.poznan.pl
†The authors wish it to be known that, in their opinion, the first two authors should be considered as Joint First Authors.

© The Author(s) 2014. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

The trinucleotide repeats (TNRs) belong to a subclass of microsatellite sequences also known as short tandem repeats. TNRs represent the most abundant type of simple sequence repeats in eukaryotic genomes and in transcripts (21,22). Some types of TNRs are widespread in exons while others are very rare, suggesting that their occurrence in exons is not random, but undergoes positive or negative selective pressure. These sequences have received special attention, primarily because some are known to undergo pathogenic expansions that cause triplet repeat expansion diseases (TREDs). More than 20 genetic disorders belong to this group, and most are neurodegenerative and neuromuscular disorders (23).

In several TREDs, stable RNA structures formed by triplet repeats present in untranslated regions of the responsible genes are implicated in pathogenesis (24). Of interest to us are RNA sequences composed solely of AGG, CGG and UGG repeats, with a potential to fold into a G-quadruplex structure. It was determined that AGG and CGG repeats are overrepresented in the human exonic sequences, whereas UGG repeats have been found to be underrepresented. TNRs occurring in different mRNA regions may have different functions. It was shown that AGG-rich sequences occur mainly in the open reading frames (ORF), that CGG repeats are most frequent in the 5′-UTR and that UGG TNRs appear mainly in the 3′-UTR (22). The CGG triplet repeats that have been found within the 5′-UTR of the *FMR1* gene are polymorphic in length in a normal population and may undergo pathogenic expansions being responsible for both fragile X-associated tremor/ataxia syndrome (25) and fragile X syndrome, which is the most common inherited form of the mental impairment (26).

An interesting insight into the structural diversity of RNA triplet repeats was published recently (24,27). It was demonstrated that oligoribonucleotides composed of 17 AGG or UGG repeats form G-quadruplexes, while CGG repeats have been found to form hairpins with a tendency to adopt several alternative alignments (27). Recently, a crystal structure of molecule composed of two CGG repeats has been determined (28), in which $G(CGG)_2C$ forms an A-type duplex with typical C:G Watson–Crick interactions and the non-canonical G·G pairs. A sequence comprising three GGC repeats was studied by nuclear magnetic resonance (NMR) spectroscopy and a duplex structure was proposed with G·G mismatches intercalated between the adjacent G:C pairs (29). In contrast, there is also evidence that CGG repeats can fold into a G-quadruplex (30–33), so the folding rules are yet unclear and may depend on the exact sequence and experimental conditions.

Tetramolecular and bimolecular G-quadruplex structures containing trinucleotide repeats have also interesting applications as biological therapeutics. An RNA aptamer containing four repeats of GGA has been isolated against a recombinant bovine prion protein and its structure was determined by NMR spectroscopy (34). It was found that two $(GGA)_4$ G-quadruplexes form a dimer stabilized through a stacking interaction between A:(G:G:G:G):A hexad planes of the two G-quadruplexes. Although experimental data available in the literature on RNA UGG repeats underline their tendency to form G-quadruplex structures (27,35), to the best of our knowledge there is yet no detailed structural information on RNA sequences containing UGG repeats.

In the present study, we have focused on the structural preferences of RNA molecules composed of two and four repeats of AGG, CGG and UGG in the presence of K^+, Na^+ and NH_4^+ ions. We employed ultraviolet (UV) molecular absorption, circular dichroism (CD), NMR spectroscopies and electrospray ionization mass spectrometry (ESI-MS) to obtain comprehensive information on the structural forms existing in different solution conditions. To maintain comparable conditions between different experimental methods, we have recorded the ^{1}H NMR spectra using the same solutions as for the CD and UV experiments. Only the ESI-MS could be carried out only in NH_4^+. Additionally, to complement our studies, we have tested the effect of intracellular environment on the folding of analysed molecules. In order to mimic the physiologically relevant conditions the ^{1}H NMR spectra were acquired in the HeLa cell extract.

MATERIALS AND METHODS

RNA sample preparation

All oligoribonucleotides: $(AGG)_2A$, $(AGG)_4A$, $G(CGG)_2C$, $G(CGG)_4C$, $p(UGG)_2U$ and $p(UGG)_4U$ were chemically synthesized on the Applied Biosystems DNA/RNA synthesizer using β-cyanoethyl phosphoramidite chemistry on solid support (36–38). The samples preparation procedure can be found in the Supplementary Data.

Our earlier experience with molecules composed of AGG, CGG and UGG repeats have shown that occasionally their spectroscopic features can be improved adding 5′-phosphate group. The role of 5′-phosphate has been discussed several times in the context of the G-quadruplex structures (39,40). In this study, this type of modification was used only for UGG repeats, because it improved the quality of the ^{1}H NMR spectra.

Electrospray mass spectrometry

ESI-MS experiments were carried out with a Synapt HDMS (Waters, Manchester, UK) with electrospray ionization in the negative ion mode (direct infusion at 4 μl/min). Sodium and potassium ions were not used in ESI-MS experiments because they condense around negatively charged nucleic acid backbones in solution and during solvent droplet evaporation, without a possibility of clean-up by evaporation once in the gas phase. To overcome this problem, we used NH_4^+ ions, which can stabilize G-quadruplexes by coordination in-between G-tetrads in a similar way as K^+ ions (41), but can evaporate as ammonia when bound to external phosphate sites (42–44). The stock solution was prepared from 500 μM RNA (single strand concentration) in 150 mM NH_4OAc. All the samples were annealed by heating at 90°C for 5 min and then slowly cooled down to the room temperature and stored at 4°C at least one day before the experiments. The final injection concentration of RNA (single strand) was 18 μM (prepared from 500 μM) in 150 mM NH_4OAc. About 1 h before

measurement methanol was added to the final concentration 10% v/v. A fixed concentration of dT_6 molecule (as an intensity reference) was also present in the sample. The part where ions are most activated is upon entering the IMS cell, because a bias voltage is required for the ions to enter the higher-pressure region of the cell. Therefore, the mass spectra were systematically recorded at different bias voltages, from 15 to 25 V, and the number of ammonium ions remaining bound to the RNA multistranded structures was monitored. Typically, ammonium ions persisting at up to 25 V bias voltage indicate very tight binding sites in-between G-tetrads. For more experimental details see the Supplementary Data.

UV thermal denaturation curves

Thermal denaturation curves were obtained at 295 nm with JASCO V-650 spectrophotometer using quartz optical cuvettes of 0.5 and 1 cm path lengths with the sample volumes of 150 and 300 μl, respectively. The samples were protected against evaporation by the silicone oil. Before measurements the cuvettes filled with samples were spun at 5000 revolutions per minute by 3 min in order to avoid air bubbles formation during measurements. The temperature range was 5–93°C, using the scan rate 0.5°C min^{-1}.

Thermal difference spectra (TDS)

The TDS were obtained by subtracting the UV spectra recorded above and below the melting temperature (90°C and 20°C, respectively). TDS were recorded in the range between 220 and 340 nm on JASCO V-650 spectrophotometer using quartz optical cuvettes of 0.5 and 1 cm path lengths with the sample volumes of 150 and 300 μl, respectively.

Circular dichroism

The JASCO J815 spectropolarimeter equipped with a Peltier temperature controller was used to collect CD spectra. A quartz cuvettes with a path length 0.5 and 1 cm were used with the sample volumes of 1300 and 2500 μl, respectively. Spectra were collected in the range between 220 and 340 nm at 25°C, from three scans and a buffer baseline was subtracted from each spectrum. CD was expressed as the difference in the molar absorption of the right- and left-handed circularly polarized light. $\Delta\epsilon$, in units of cm^2 $mmol^{-1}$, was normalized to the number of nucleoside residues in the RNA samples (45).

Nuclear magnetic resonance

The 1D ^1H NMR spectra as well as 2D homo- and hetero-nuclear spectra were collected on a 700 MHz Bruker AVANCE III spectrometer, equipped with a QCI CryoProbe. The 5 mm thin wall tubes were used with the final sample volume of 600 μl (90% H_2O/10% D_2O, v/v). The ^1H NMR spectra were collected with water suppression using excitation sculpting scheme with gradients (46), typically from 20 480 scans with 1.0 s relaxation delay. Experiments were performed at 25°C for AGG and UGG, whereas for $G(CGG)_2C$ and $G(CGG)_4C$ NMR spectra were recorded at 10°C in order to observe imino proton resonances from less thermodynamically stable forms, which were not observed at 25°C. The ^1H-^1H NOESY and ^1H-^{15}N HSQC spectra were collected for selected samples. Experimental details are given in the Supplementary Data. Spectra were processed and prepared with TopSpin 3.0 Bruker Software.

NMR studies in HeLa cell extract

HeLa cells were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS), 1× RPMI-1640 vitamin solution and 1× antibiotic antimycotic in 5% CO_2 for 48 h. In the next step, cells were trypsinized and subsequently washed three times with phosphate buffered saline to remove any remaining trypsin and medium. Cell extract was prepared by sonication of 1.6×10^7 cells in 1 ml deionized water with subsequent centrifugation at 20 000 × *g*, 5 min. For the NMR experiments 0.15–0.5 mM RNA samples were dissolved in 180 μl HeLa cell extract and 20 μl D_2O. The 3 mm thin wall sample tubes were used with the final sample volume of 200 μl. At those RNA concentrations, 256–2048 scans were usually sufficient to obtain a good signal-to-noise ratio. All spectra were recorded immediately after dissolving in HeLa extract and after 24 h. In all cases, the spectra remained unchanged, indicating that RNA samples were stable in HeLa cell extract.

RESULTS

$(AGG)_2A$ and $(AGG)_4A$ fold into G-quadruplex structures

The melting profiles of $(AGG)_2A$ and $(AGG)_4A$ obtained at 295 nm in the presence of 50 mM K^+, 150 mM NH_4^+ or 150 mM Na^+ were typical of G-quadruplexes, with a decrease of absorbance upon melting (Figure 1A, Supplementary Figure S1A and B) (47). The TDS of $(AGG)_2A$ and $(AGG)_4A$ shown in Supplementary Figure S2A and B, respectively, also display a negative band at 295 nm in all conditions and support G-quadruplex formation (48). All these profiles are similar, with the exception of the TDS of $(AGG)_2A$ obtained in the presence of sodium ions for which a slight shift of the band at ∼250 nm toward a higher wavelength was observed (Figure 1B).

The use of CD spectroscopy to determine the topology of DNA G-quadruplexes is well established (49) and was applied here. Note, however, that its use in studies of RNA G-quadruplexes can be problematic because CD spectra of the DNA and RNA parallel G-quadruplexes are very similar to those of A-form duplexes. The CD spectra recorded for $(AGG)_2A$ and $(AGG)_4A$ in the presence of K^+, Na^+ and NH_4^+ ions have positive bands at ∼265 nm and negative bands at 240 nm (Supplementary Figure S3A and B), a pattern characteristic of a parallel G-quadruplex. In a solution containing K^+ or NH_4^+ ions, an additional positive band of small intensity was detected around 305 nm. This band was previously observed in CD spectra of other RNA G-quadruplexes and was attributed to the formation of a dimer of G-quadruplexes stabilized by hexad–hexad stacking (50,51). The shape of the CD spectrum of $(AGG)_2A$ in solution containing Na^+ ions was different (Figure 1C). That observation, together with the observed

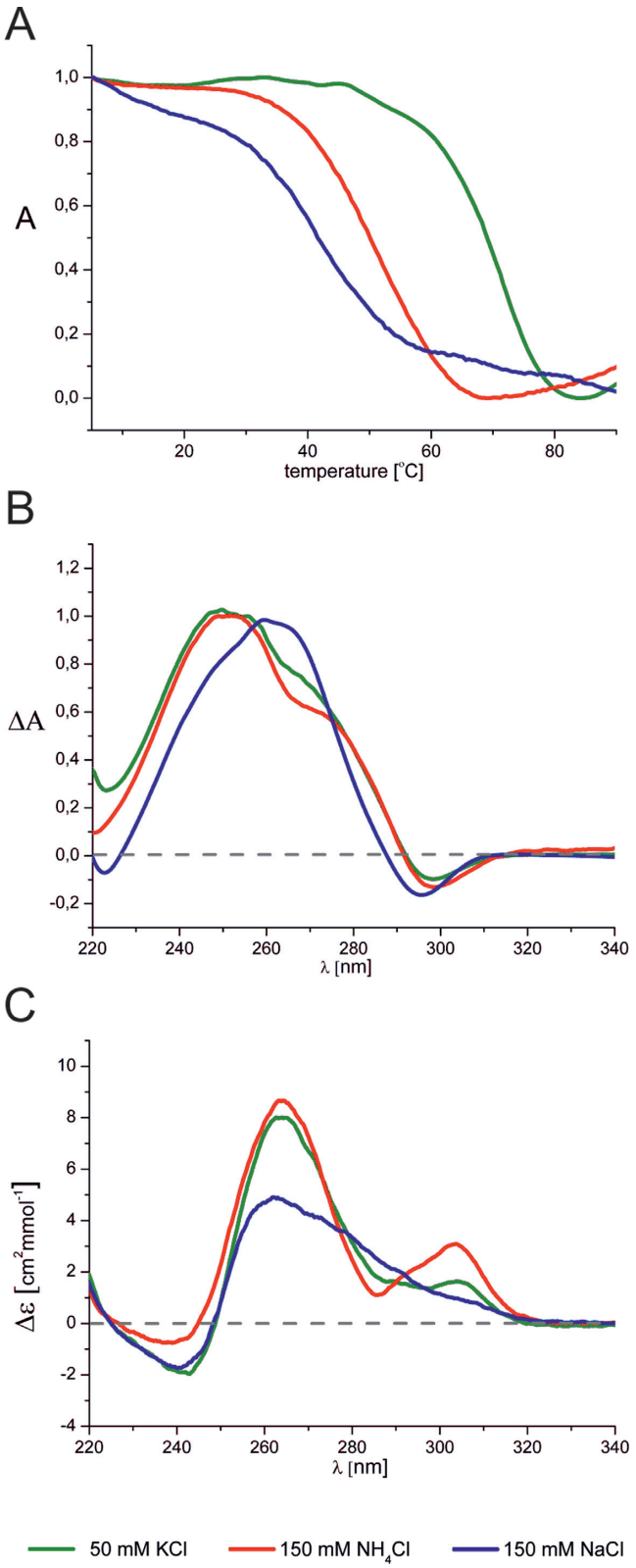

Figure 1. (A) Normalized UV melting profiles of $(AGG)_4A$ at 295 nm, (B) normalized thermal differential spectra of $(AGG)_2A$ and (C) CD spectra of $(AGG)_2A$ at 25°C. All the spectra were recorded in the presence of 50 mM KCl, 10 mM potassium phosphate and 0.1 mM EDTA, pH 6.8 (green), 150 mM NH_4Cl, 10 mM Tris/HCl, 0.1 mM EDTA, pH 6.8 (red) and 150 mM NaCl, 10 mM sodium phosphate, 0.1 mM EDTA, pH 6.8 (blue).

shift of the maximum TDS band, suggest the formation of a G-quadruplex with a different topology.

We then established the molecularity of the G-quadruplexes as well as the number of the adducted ammonium ions preserved in the structure using negative ion mode ESI-MS. For parallel-stranded structures, ammonium ion preservation in the gas phase is easier than for antiparallel structures (42–44). As the structures formed here are parallel-stranded both in ammonium and in potassium, we are in a favourable case where the ammonium ion count by ESI-MS allows us to deduce the number of adjacent stacked tetrads. The ESI-MS spectrum of $(AGG)_2A$ shows two significant peaks (Supplementary Figure S4A). The peak at m/z 1151.86 corresponds to a monomer at charge state 2- without adducted ammonium ions, whereas the peak at m/z 1859.57 corresponds to the tetramolecular G-quadruplex structure at charge state 5- with three, four or five trapped ammonium ions. The distribution of ammonium ion adducts is presented in the insert of Supplementary Figure S4A. The ESI-MS spectrum of $(AGG)_4A$ is presented in Figure 2. Only one peak at m/z 1747.82 and charge state 5- was observed, corresponding to the bimolecular G-quadruplex structure stabilized by three ammonium ions, suggesting the contiguous stacking of four tetrads.

Another method that we used to assess G-quadruplex formation was NMR spectroscopy. We have used the same RNA solutions for NMR measurements as for CD and UV experiments. Consequently, the very low concentration of NMR samples (0.01 mM) entailed long acquisition times, typically about 14 h, in order to obtain a satisfactory signal-to-noise (S/N) ratio. Additionally, we have recorded ^{1}H NMR spectra at higher concentrations for each RNA sequence, in order to determine the effect of strand concentration on RNA structure. A comparison of ^{1}H NMR spectra of $(AGG)_2A$ recorded in solutions containing K^+, Na^+ or NH_4^+ ions is depicted in Figure 3A (a–c). A spectrum recorded in K^+ solution revealed the presence of two major and one minor forms. However, when the sample concentration was increased (1.14 mM) only one of these two major species was observed (Figure 3Ad). In NMR spectra of both concentrated and diluted samples, a characteristic signal appeared around 10 ppm. Signals in this region were previously observed in the NMR spectra of oligoribonucleotides containing an A:(G:G:G:G):A hexad motif (Figure 4A) and were assigned to guanosine amino protons involved in hydrogen bonding with adenosine (34,51,52). We recorded the ^{1}H-^{15}N HSQC spectrum for concentrated samples, in order to confirm that this high-field signal is due to amino proton (Supplementary Figure S5). The imino and amino protons can be easily distinguished based on the analysis of ^{15}N chemical shifts in ^{1}H-^{15}N HSQC spectra (53). Indeed, the chemical shift of the nitrogen atom corresponding to the proton at 10.14 ppm allowed us to unambiguously attribute this signal to the guanosine amino group. Finally, taking into account the results from the ESI-MS, we assumed that $(AGG)_2A$ folds into a dimer of bimolecular G-quadruplexes (Figure 4E), similar to that previously published (51).

In the presence of NH_4^+ ions the spectrum indicative of a single G-quadruplex form was acquired (Figure 3Ac). The

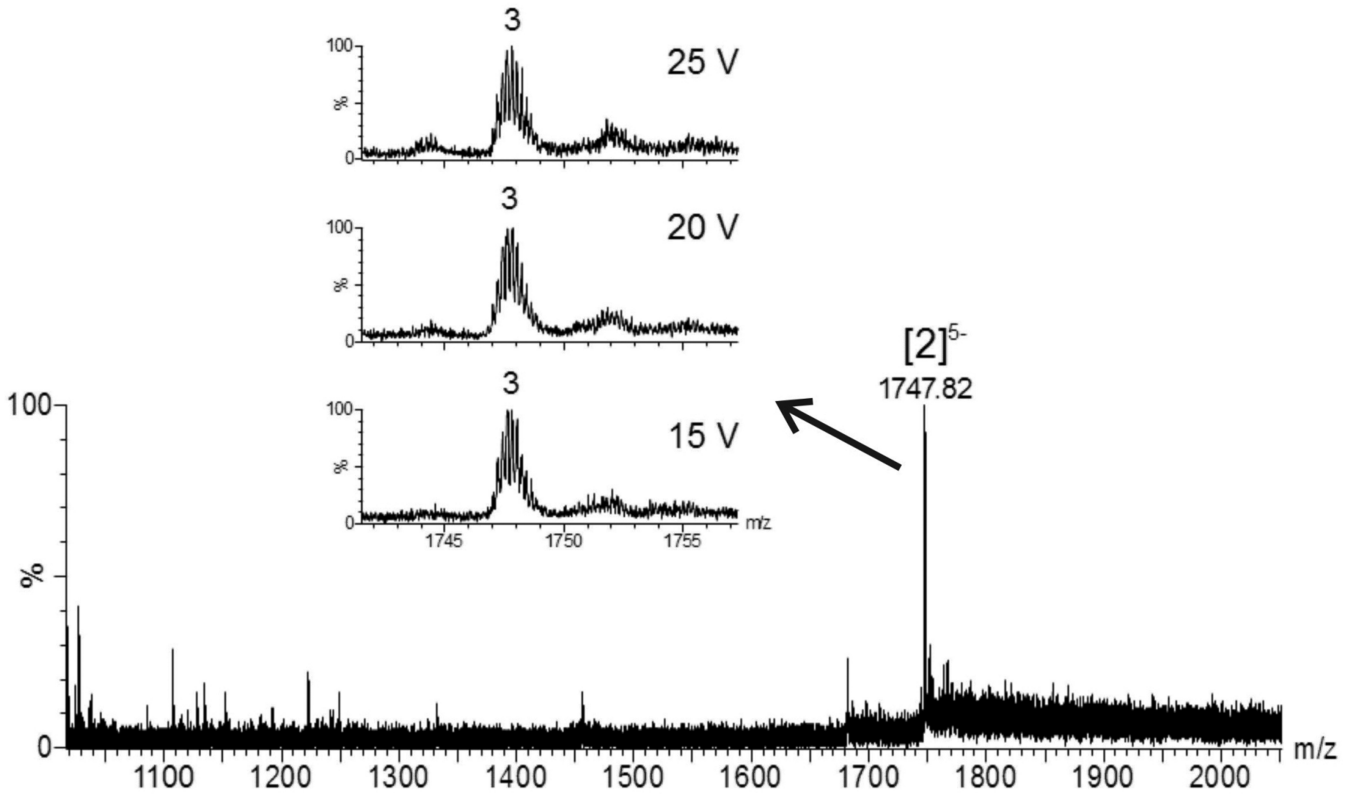

Figure 2. Electrospray mass spectrum of $(AGG)_4A$ in the presence 150 mM ammonium acetate. The final single strand RNA concentration was 18 μM. 10% v/v of methanol was added 1 h before measurement. The peak annotation $[n]^{z-}$ indicates the number of strands (n) and the total charge (z). The distribution of the number of ammonium ions preserved in the structure is shown in the inset. The distribution is presented at three bias voltages.

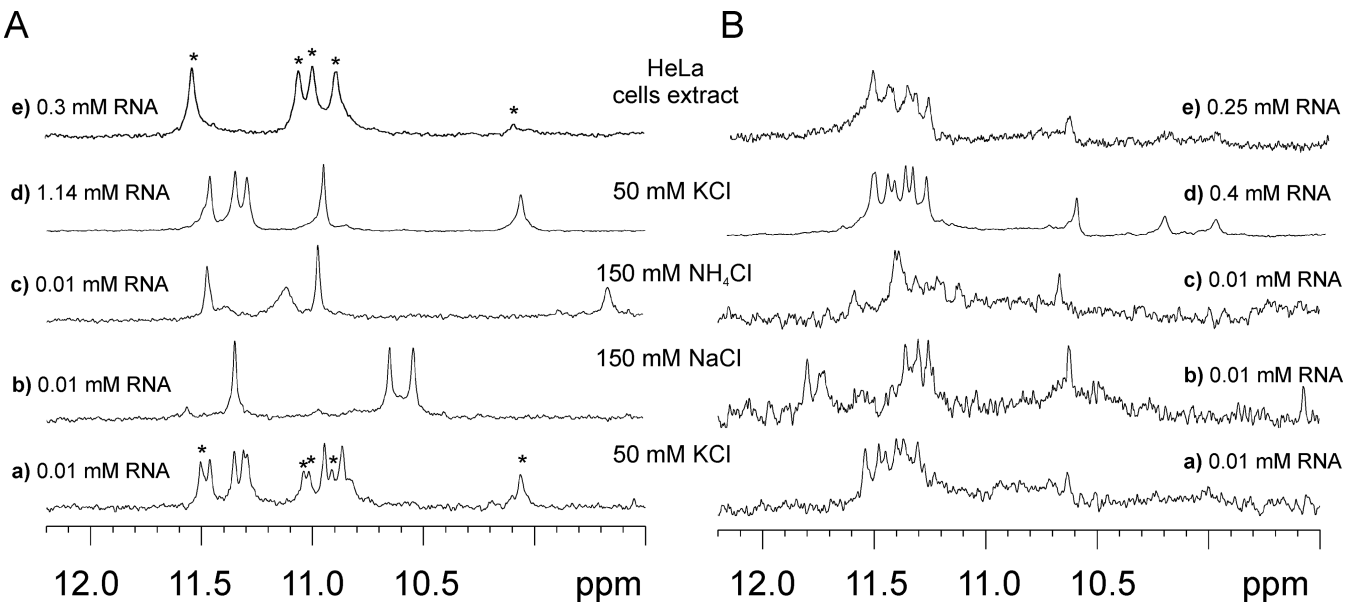

Figure 3. Imino region of the ^1H NMR spectra of (A) $(AGG)_2A$ and (B) $(AGG)_4A$ at 25°C in the presence of 50 mM KCl, 10 mM potassium phosphate and 0.1 mM EDTA, pH 6.8 (Aa, Ba, Ad, Bd), 150 mM NaCl, 10 mM sodium phosphate, 0.1 mM EDTA, pH 6.8 (Ab, Bb), 150 mM NH_4Cl, 10 mM Tris/HCl, 0.1 mM EDTA, pH 6.8 (Ac, Bc) and the HeLa cell extract (Ae, Be). The poor S/N ratio observed for $(AGG)_4A$ is due to polymorphism or/and the presence of higher order structures.

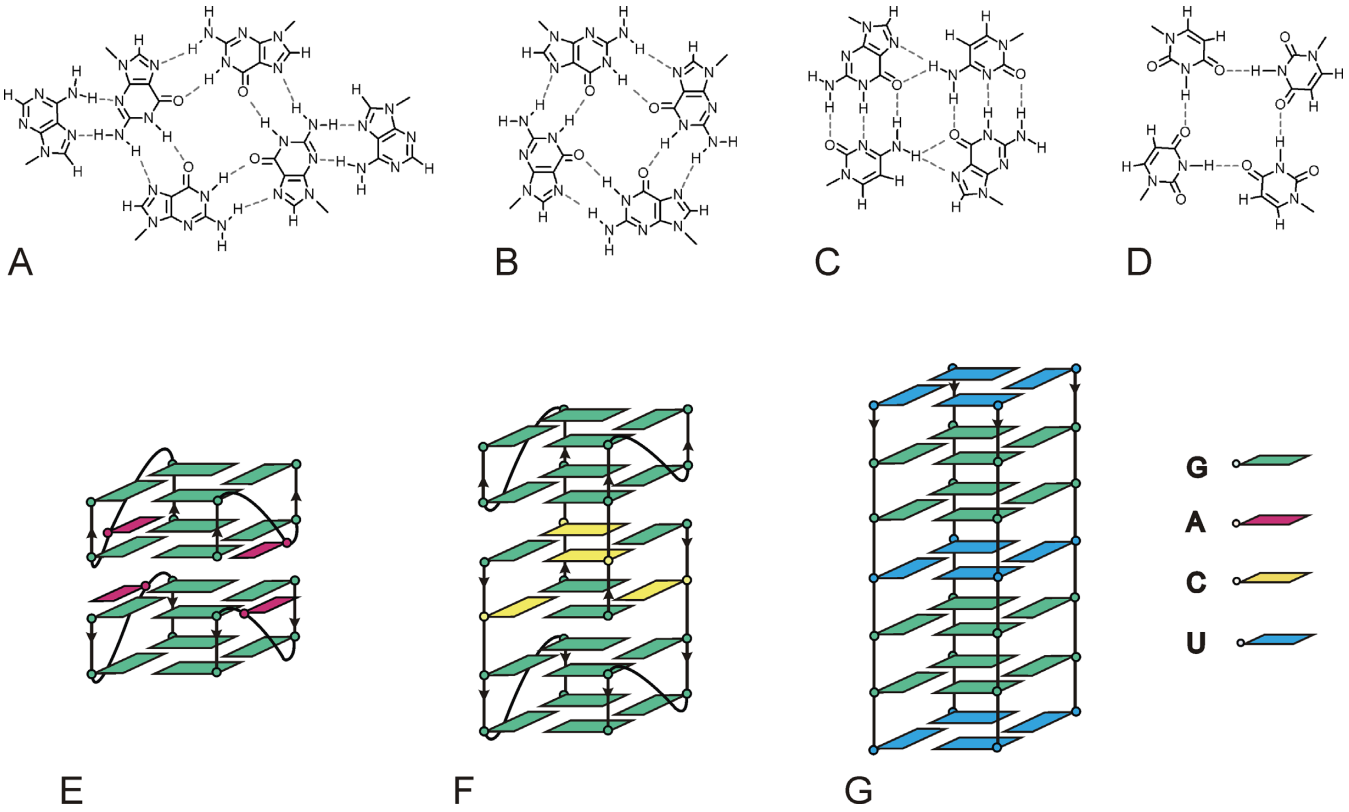

Figure 4. Schematic outline of (A) A:(G:G:G:G):A hexad, (B) G-tetrad, (C) G:C:G:C tetrad, (D) U-tetrad motifs and fold of (E) $(AGG)_2A$, (F) $G(CGG)_2C$ and (G) $p(UGG)_2U$.

high-field resonance characteristic of the hexad motif also occurred, suggesting that $(AGG)_2A$ adopts a similar fold to that obtained in the potassium buffer. In the NMR spectrum of $(AGG)_2A$ in a buffer containing Na^+ ions (Figure 3Ab), only three major imino resonances were visible, two of which were shifted high-field in respect to the spectra recorded in K^+ and NH_4^+ solutions. The absence of characteristic guanosine amino resonance around 10 ppm, together with the observed shift of a positive band in the TDS spectrum and the different shape of the CD spectrum, suggested that $(AGG)_2A$ folds differently in the presence of Na^+ ions. Next, we have performed a study in the HeLa cell extract, to find out how the intracellular environment could influence the $(AGG)_2A$ folding. Since the ^1H NMR spectrum of such an extract does not contain any signals in the region above 9.5 ppm, therefore, this model system is very convenient to monitor RNA structures. The ^1H NMR spectrum recorded in the HeLa cell extract is presented in Figure 3Ae. The number of proton resonances indicated the formation of a single G-quadruplex form. However, the spectrum was different from those recorded in other conditions. The chemical shifts of the observed imino protons were similar to that of the minor form (marked by *) present in the ^1H NMR spectrum recorded in the potassium buffer at a low concentration of RNA (Figure 3Aa,e).

The ^1H NMR spectra of diluted and concentrated $(AGG)_4A$ samples obtained in the presence of K^+ ions are similar (Figure 3Ba,d) and resemble that previously published for $(GGA)_4$ (34), suggesting the formation of the dimer of unimolecular subunits including the A:(G:G:G:G):A hexad (Figure 4E). It is difficult to interpret the NMR spectrum of $(AGG)_4A$ obtained in solution containing NH_4^+ ions (Figure 3Bc), because of the lack of spectral resolution and the extensive line broadening leading to a low signal-to-noise ratio. Inspection of Figure 3Bb implies that the major structure of $(AGG)_4A$ in buffer containing sodium ions can be different from that existing in potassium and ammonium buffers, which together with results obtained for $(AGG)_2A$ suggests that the presence of sodium ions can induce a different fold in the case of AGG-rich sequences. Finally, the spectra of $(AGG)_4A$ obtained in the HeLa cell extract and potassium buffer were virtually identical, regardless of the RNA concentration (Figure 3Be).

Structural preferences of $G(CGG)_2C$ and $G(CGG)_4C$

The same set of experimental methods was applied to the other sequences. Supplementary Figure S1C and D show the UV melting curves obtained at 295 nm for $G(CGG)_2C$ and $G(CGG)_4C$, respectively, in the presence of K^+, NH_4^+ and Na^+ ions. Here, we observed melting profiles typical of G-quadruplexes only in the presence of K^+ ions. In the melting curve of $G(CGG)_4C$, besides the decrease of absorbance visible at higher temperatures, an additional, low temperature transition with absorbance increase occurred in K^+ solution. This observation suggested that other types of structure could co-exist with the G-quadruplex. No fea-

tures characteristic of the G-quadruplexes were detected in the melting curves of $G(CGG)_2C$ and $G(CGG)_4C$ obtained in sodium and ammonium-containing buffers.

To get more insight into the type of structures which were formed in the sodium and ammonium buffers, we analysed the melting curves at 260 nm (Supplementary Figure S6). The concentration dependence of melting temperature (T_m) of $G(CGG)_2C$ was characteristic of melting profile of bimolecular form, such as a duplex (Supplementary Figure S7). The sigmoidal shape of the profiles at 260 nm obtained for $G(CGG)_4C$ in the ammonium and sodium buffers suggested that this molecule can fold into a duplex or hairpin structure. TDS of $G(CGG)_2C$ and $G(CGG)_4C$ are shown in Supplementary Figure S2C and D, respectively. A negative band at 295 nm was found only when potassium ions were present in solution. The shapes of the CD spectra of both molecules were similar, except for the spectrum of $G(CGG)_2C$ recorded in potassium buffer (Supplementary Figure S3C and D).

The ESI-MS spectrum of $G(CGG)_2C$ shown in Supplementary Figure S4C revealed the presence of bimolecular and tetramolecular structures. The major peak (100%) at m/z 1288.89 and minor peak at m/z 1718.85 correspond to a bimolecular structure at charge states 4- and 3-, respectively. The absence of specifically bound ammonium ion adducts suggests the absence of stacked G-tetrads, and hence that this bimolecular structure may be stabilized by Watson–Crick base pairs instead of G-tetrads. Another minor peak at m/z 1724.54 corresponding to the tetramolecular structure at charge state 6- appeared, in this case with adducted ammonium ions. The ammonium ion distribution was monitored with the increase of the bias voltage and is shown in the insert of Supplementary Figure S4C, and the biased distribution toward two coordinated ammonium ions suggests that this tetramolecular structure, in contrast to the bimolecular one, contains stacked G-tetrads. In the ESI-MS spectrum of $G(CGG)_4C$ (Supplementary Figure S4D) the main peak observed at m/z 1141.49 and the minor peak at m/z 1522.42 correspond to a monomolecular form without adducted ammonium ions at charge state 5- and 3-, respectively. An additional, less intense peak corresponding to a bimolecular structure was observed at m/z 1844.32 (charge 5-). The distribution of ammonium adducts in this structure at low voltages is broad as a result of the mild desolvation conditions and makes it difficult to conclude whether or not different structures (G-quadruplex with labile ammonium ions (43) and non-G-quadruplex) coexist.

Figure 5Aa represents an ^{1}H NMR spectrum of $G(CGG)_2C$ in the presence of K^+ ions. In these conditions, four imino resonances appear in the region distinctive of G-quadruplex structures and one guanosine imino proton resonates at 12.44 ppm, a region characteristic of Watson–Crick G·C base pairs. It has been previously demonstrated that the presence of 5′-GC ends can lead to the association of two G-quadruplex moieties through G·C base pairs and the formation of two G:C:G:C-tetrads (54). Overall, the observation of the inverted UV melting profile at 295 nm and a negative band at this wavelength in the TDS spectrum, together with the analysis of the ESI-MS and ^{1}H NMR spectra indicated the formation of a highly symmetric tetramolecular G-quadruplex structure with G:G:G:G and mixed G:C:G:C tetrads (Figure 4B, C and F) in the K^+-containing solution.

The ^{1}H NMR spectrum of $G(CGG)_2C$ obtained in Na^+ solution revealed the presence of a mixture of conformers (Figure 5Ab). Although neither the UV melting analysis at 295 nm nor the TDS spectrum indicated the formation of a G-quadruplex under the Na^+ environment, the appearance of four sharp signals in the 11.01–11.22 ppm region of the ^{1}H NMR spectrum revealed that a minor population of the G-quadruplex form also exists in these conditions. We attributed resonances of the major form in the 12.5–13.2 ppm region to the G:C base pairs, and the broad signal at 10.61 ppm to a G·G mismatch (Figure 5Ab). Subsequently, we have studied the effect of oligoribonucleotide concentration on the G-quadruplex-duplex equilibrium in the sodium buffer. When the RNA concentration was increased, signals corresponding to the duplex diminished notably, and the G-quadruplex emerged as the dominant form (Figure 5Ad). The observed down-field shift of G:C imino resonance relative to that in the presence of K^+ could be related to the deshielding effect of the ion, or could reflect different metal binding sites within the G:C:G:C tetrad (55). In the ammonium buffer, only two sharp imino resonances at 12.63 and 12.71 ppm were visible and no signals corresponding to the G-quadruplex form could be determined (Figure 5Ac).

The dependence of the ^{1}H NMR spectra of $G(CGG)_4C$ on the solution conditions is presented in Figure 5B. Severe broadening of imino resonances was observed in a buffer containing potassium ions, indicating the coexistence of several species. We recorded ^{1}H NMR spectra as a function of the temperature in order to verify whether these species differ in thermal stability. Because the ^{1}H NMR spectra of diluted (0.01 mM) and concentrated (0.45 mM) samples were similar (Figure 5Ba,d), we performed this experiment on the most concentrated sample (Supplementary Figure S8). Imino resonances at about 12.7 ppm and a broad signal at 10.65 ppm diminished upon heating and vanished at 50°C. On the other hand, resonances around 11.4 ppm sharpened upon heating and remained very intense even at 50°C, a behaviour that is typical of G-quadruplexes. When the ^{1}H NMR spectrum of $G(CGG)_4C$ was recorded in the presence of Na^+ ions, only signals representing a hairpin (or duplex) form were detected (Figure 5Bb). No resonances corresponding to a G-quadruplex form were observed in buffer containing NH_4^+ ions either (Figure 5Bc). The ESI-MS spectrum of $G(CGG)_4C$ corresponds to the monomolecular form without adducted ammonium ions (Supplementary Figure S4D). This stoichiometry suggests that a hairpin, and not a duplex, is the preferred form of $G(CGG)_4C$.

In the NMR spectrum of $G(CGG)_2C$ recorded in the HeLa cell extract, only signals characteristic of Watson–Crick base pairs appeared, demonstrating that the duplex is the dominant structure (Figure 5Ae). In the NMR spectrum of $G(CGG)_4C$, the hairpin is the major form, even though a small population of G-quadruplex could also be detected (Figure 5Be).

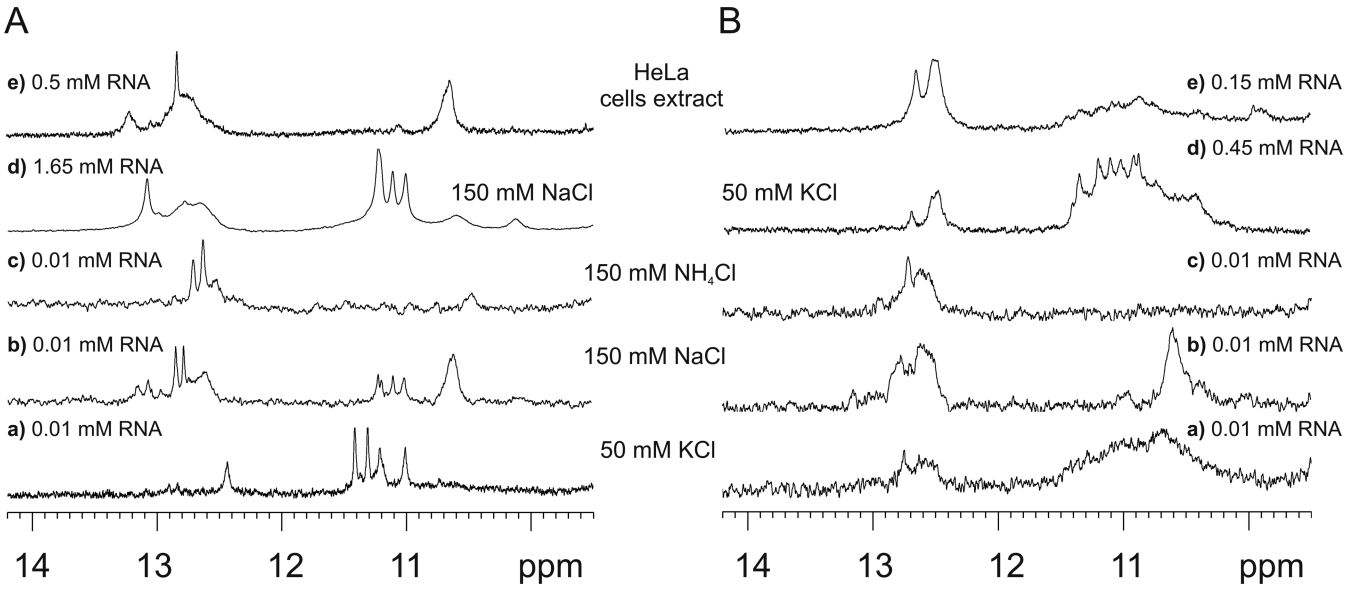

Figure 5. Imino regions of the ^{1}H NMR spectra of (A) G(CGG)$_2$C and (B) G(CGG)$_4$C at 10°C in the presence of (Aa, Ba, Ad, Bd) 50 mM KCl, 10 mM potassium phosphate and 0.1 mM EDTA, pH 6.8, (Ab, Bb) 150 mM NaCl, 10 mM sodium phosphate, 0.1 mM EDTA, pH 6.8, (Ac, Bc) 150 mM NH_4Cl, 10 mM Tris/HCl, 0.1 mM EDTA, pH 6.8 and (Ae, Be) HeLa cell extract. The poor S/N ratio observed for G(CGG)$_4$C is due to polymorphism or/and the presence of higher order structures.

G-quadruplex is the preferred form of p(UGG)$_2$U and p(UGG)$_4$U

Inverted UV melting profiles observed at 295 nm for p(UGG)$_2$U and p(UGG)$_4$U in the presence of K^+, Na^+ and NH_4^+ indicated the formation of G-quadruplexes (Supplementary Figure S1E and F). The tendency of these molecules to form G-quadruplexes was further confirmed by the presence of the negative band at 295 nm observed in the TDS (Supplementary Figure S2E and F). All CD profiles of p(UGG)$_2$U were similar (Supplementary Figure S3E), including the presence of a small negative band at ~290 nm. The CD shape of p(UGG)$_4$U obtained in ammonium buffer (Supplementary Figure S3F, red) differed from that recorded in potassium and sodium buffers (Supplementary Figure S3F, green and blue curves, respectively). In addition to the positive band at ~264 nm, a relatively intense positive band at ~290 nm and the absence of the negative band at 240 nm are clearly notable.

We have observed two types of peaks in ESI-MS spectra of p(UGG)$_2$U. The peak at m/z 1157.86 corresponding to a single strand at charge state 2-, and two peaks at m/z 1555.31 and 1870.34 corresponding to charge states 6- and 5- of a tetramolecular structure with four or five ammonium ion adducts (Supplementary Figure S4E). The evolution of the ammonium ion distribution upon increasing bias voltage for both peaks (shown in the insert of Supplementary Figure S4E) revealed that the structure stabilized by four ions were the most stable. In the ESI-MS spectrum of p(UGG)$_4$U (Supplementary Figure S4F) only one peak at m/z 1733.50 was observed, corresponding to a bimolecular structure at charge state 5-, stabilized by three ammonium ions.

The ^{1}H NMR spectra of p(UGG)$_2$U obtained in solutions containing K^+, Na^+ and NH_4^+ (Figure 6A) provided additional evidence for a G-quadruplex formation. In the ^{1}H NMR spectrum of p(UGG)$_2$U recorded in the presence of K^+ (Supplementary Figure S9), several resonances sharpened with the increase in temperature and were visible even at 80°C. Additionally, new resonances appeared at 3°C. To elucidate their origin and, in particular, to distinguish guanosine from uridine imino protons, we acquired a ^{1}H –^{15}N HSQC spectrum at 3°C. Four guanosine and three uridine imino protons were identified based on the chemical shifts of the attached nitrogen (Supplementary Figure S10). The appearance of uridine imino signals suggested the formation of U-tetrads (56). To verify whether the uridine residues constitute part of the G-quadruplex architecture, we assigned all guanosine and uridine imino protons based on the analysis of a 2D NOESY spectrum (data not shown). Each of the three observed uridine imino resonances exhibited a strong NOE to H5 and much weaker one to H6 protons (Supplementary Figure S11), a pattern that is distinctive of a U-tetrad motif (55). The 3′-end uridine imino proton at 11.08 ppm and that at 9.85 ppm, assigned to the internal uridine residue, sharpened with increasing temperature and were still observed at a high temperature, revealing the formation of two stable U-tetrads. Although there is no NMR structure known with an internal U-tetrad, a similar shift ~9.5 ppm was observed for imino protons of the thymidine tetrad accommodated in the centre of the DNA G-quadruplex (13,57,58). The 5′-end uridine imino proton at 10.74 ppm was observed only below 25°C (Supplementary Figure S9). These results, together with ESI-MS data (four very stable cation binding sites and one additional, more labile binding site in the tetramer), indicate that p(UGG)$_2$U folds into a symmetrical tetramolecular parallel G-quadruplex with four G-tetrads and three U-tetrads (Figure 4B, D and G).

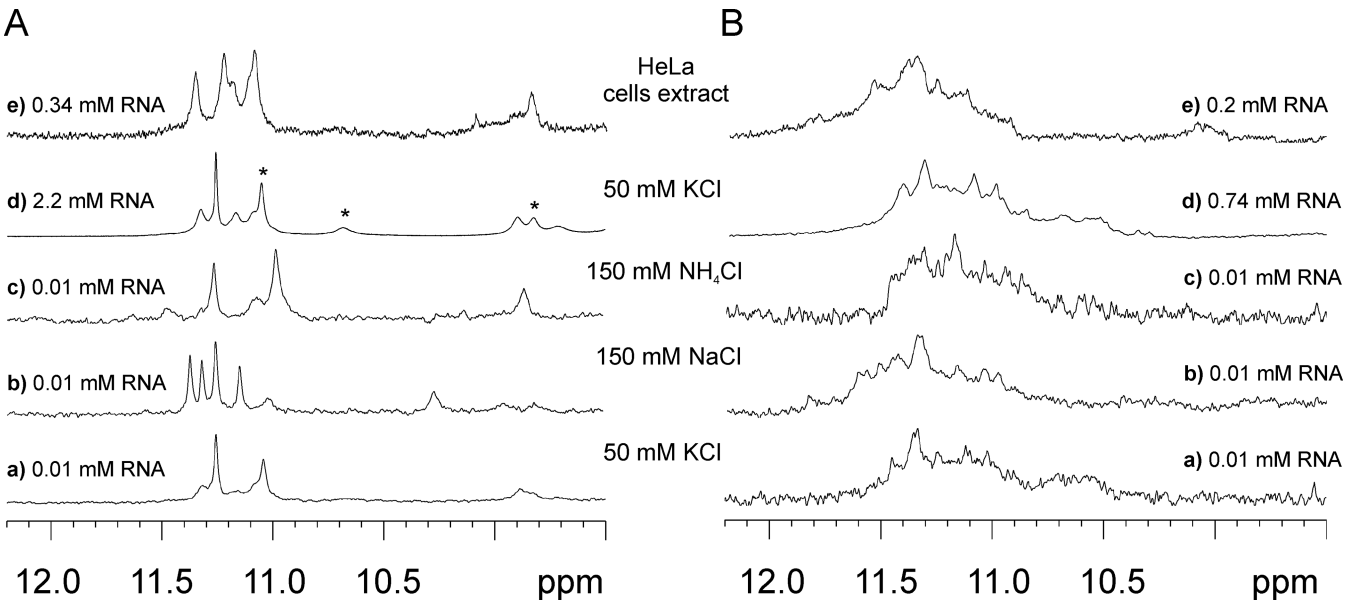

Figure 6. Imino region of the ^{1}H NMR spectra of (A) p(UGG)$_2$U and (B) p(UGG)$_4$U at 25°C in the presence of (Aa, Ba, Ad, Bd) 50 mM KCl, 10 mM potassium phosphate and 0.1 mM EDTA, pH 6.8, (Ab, Bb) 150 mM NaCl, 10 mM sodium phosphate, 0.1 mM EDTA, pH 6.8, (Ac, Bc) 150 mM NH_4Cl, 10 mM Tris/HCl, 0.1 mM EDTA, pH 6.8 and (Ae, Be) HeLa cell extract. Asterisks indicate imino proton resonances of uridines. The poor S/N ratio observed for p(UGG)$_4$U is due to polymorphism or/and the presence of higher order structures.

The imino proton spectrum obtained in ammonium buffer was similar to that recorded in the presence of K^+ (Figure 6Aa,c). The main change observed in the spectrum recorded in Na^+ solution (Figure 6Ab) was the sharpening of two resonances and the down-field shift of one of the uridine resonances (10.27 ppm). This could be explained by the different binding of sodium ions within U-tetrads, due to the smaller ionic radii of Na^+ relative to K^+ and NH_4^+. We have also observed sharp imino resonances in the NMR spectrum of p(UGG)$_2$U in the HeLa cell extract (Figure 6Ae). The similarity between the spectra obtained in the cell extract and in potassium buffer strongly suggest that also in cellular conditions, the p(UGG)$_2$U adopts a G-quadruplex structure with at least two stable U-tetrads.

As for the four-repeat sequence p(UGG)$_4$U, independently of the conditions used including cell extract, all ^{1}H NMR spectra were typical of G-quadruplexes. However, a strong tendency to form aggregates and multiple conformations was observed (Figure 6B). Due to the broadness of imino resonances we were not able, using the methods of NMR spectroscopy, to determine the role of uridine residues in a G-quadruplex fold of p(UGG)$_4$U, as we had succeeded in doing for p(UGG)$_2$U.

DISCUSSION

We examined here the structural differences between molecules composed of two or four AGG, CGG and UGG repeats. In general, all these molecules can fold into G-quadruplex structures, but for sequences comprising CGG repeats, the G-quadruplex structure competes with duplex or hairpin structures. We have identified G-quadruplex formation by combining different experimental techniques: UV, CD and NMR spectroscopies along with ESI-MS. The observation of an inverted profile of melting curve at 295 nm demonstrated, with a high degree of confidence, the existence of a G-quadruplex structure. The use of CD spectroscopy in the study of RNA G-quadruplexes has some limitations arising from the fact that spectra of RNA G-quadruplexes are very similar to those of A-form duplexes. Nevertheless, careful analysis of the shape of the CD curve may provide supporting arguments for the presence of specific structural motifs. For example, a small negative band at ~290 nm in the CD spectrum of p(UGG)$_2$U was attributed to the U-tetrad formation. A similar negative band had also been observed, but not discussed, in the CD spectra of other RNA sequences containing U-tetrads (55,59).

Using ESI-MS in NH_4^+, we determined that all three two-repeat sequences form tetramolecular G-quadruplexes, while (AGG)$_4$A and p(UGG)$_4$U fold into bimolecular G-quadruplexes. The number of ammoniuim ions remaining coordinated as detected by ESI-MS, however, differed depending on the sequence, suggesting different structural arrangements. The interpretation of the ESI-MS results obtained for G(CGG)$_4$C was ambiguous: the presence of peaks with no ammonium ions remaining bound can be attributed to non-G-quadruplex structures (as confirmed with the other technique) but ESI-MS alone would be inconclusive because the absence of bound ammonium ions could also be due particularly to high lability in the gas phase (44).

The presence of structural motifs, such as A:(G:G:G:G):A hexads, mixed G:C:G:C tetrads and U-tetrads, was confirmed mainly through the analysis of 1D and 2D NMR spectra. NMR spectroscopy helped not only in the identification of the various structural motifs, but also allowed us to observe the formation of various alternate structures in different environments. DNA and RNA sequences containing the GGAGG motif are

known to preferentially fold into a dimer of G-quadruplex subunits composed of the canonical G:G:G:G tetrad and A:(G:G:G:G):A hexad planes (34,50,51,60–62). The formation of the hexads increases the stacking contact between two subunits and facilitates their association. Both $(AGG)_2A$ and $(AGG)_4A$, contain such a motif and, indeed, adopt similar dimeric G-quadruplex structures involving A:(G:G:G:G):A hexads (Figure 4E). Furthermore, we have acquired the NMR spectra of all molecules in the HeLa cell extract to determine which *in vitro* observed structure exists in a native-like environment. Interestingly, although the potassium cations are predominant in the cellular environment, the NMR spectrum obtained in the K^+ solution was not always the same as that recorded in the cellular extract. Accordingly, for $G(CGG)_2C$ in K^+ solution a G-quadruplex was found to be the major form, but in the HeLa cell extract a duplex was formed almost exclusively.

The structural preferences of molecules composed of CGG repeats have been a matter of debate (28,31,32,63,64). Our study revealed that $G(CGG)_2C$ and $G(CGG)_4C$ are indeed polymorphic, and that their structures depend on both the length of sequence and type of cation. We have shown that $G(CGG)_2C$ fold exclusively into G-quadruplex in K^+ solution but a G-quadruplex/duplex equilibrium was observed in the presence of Na^+ ions. Under all conditions, $G(CGG)_4C$ showed a greater tendency to fold into a hairpin than into a G-quadruplex structure. Our data have shown that it is possible to switch between a duplex and G-quadruplex structures of $G(CGG)_2C$ by changing the type of monovalent cation or the concentration of the RNA strand. In a recent study on DNA G-wire formation, the dimerization of G-quadruplex structures through formation of G:C:G:C tetrads based on Watson−Crick G:C base pairing was tested (65). It was found that the presence of 5′GC-ends, but not 3′GC-ends, promotes self-assembly via Watson−Crick base pairing. Although for the RNA CGG TNRs the presence of G-quadruplex was postulated, no structure was proposed (30–33). We have demonstrated that $G(CGG)_2C$ folds into a symmetrical, tetramolecular parallel G-quadruplex structure with G:G:G:G and mixed G:C:G:C tetrads (Figure 4F). Both the presence of complementary 5′GC-ends in the sequence and G:C:G:C tetrads strongly suggest that $G(CGG)_2C$ adopts the same architecture as that determined by Webba da Silva for DNA (54).

We also established that $p(UGG)_2U$ folds into tetramolecular parallel G-quadruplex with two terminal and one internal uridine tetrads (Figure 4G). Until now, there have been no reports on RNA G-quadruplexes with a uridine tetrad intercalated into the G-quadruplex core. Analysis of the available X-ray structures of RNA G-quadruplexes indicates that internal uridine residues prefer to adopt the bulged-out conformation (66,67). Our data demonstrate for the first time that uridine residues can form U-tetrad inside G-tetrads. The stacking of a single bulge in or out of a G-quadruplex resembles the structures of DNA and RNA duplexes with a base being bulged out or stacked into a helix. In general, in crystal structures single bulges are more likely to be looped out, while in NMR structures the bases are usually found stacked into the helix (68–73). Following our finding that this holds true for short, tetramolecular G-quadruplexes, more thorough structural investigations on larger systems are needed to generalize these conclusions to longer sequences.

CONCLUSIONS

In summary, using UV molecular absorption, CD, NMR spectroscopies and ESI-MS, we have demonstrated that all oligoribonucleotides comprising two TNRs fold preferentially into G-quadruplexes in solution containing potassium ions. In their structure, these G-quadruplexes have distinct structural motifs characteristic of the type of repeats. Additionally, spectral features indicative of A:(G:G:G:G):A hexads have been found for $(AGG)_2A$, whereas signals characteristic of mixed G:C:G:C tetrads and U-tetrads have been observed in NMR spectra of $G(CGG)_2C$ and $p(UGG)_2U$, respectively. The knowledge of the structures formed by CGG repeats is crucial for understanding the pathogenesis of the family of fragile X-associated disorders, where the structure of RNA molecules play important role. Then, detailed understanding of the structural preferences of AGG and UGG trinucleotide tandem repeats may help in the recognition of the biological functions of these sequences.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Science Center [N N301 707040 to Z.G.; UMO-2013/08/A/ST5/00295, UMO-2011/03/B/NZ1/00576 and UMO-2011/03/B/ST5/011098 to R.K.]; Ministry of Science and Higher Education [N N301 255536 to Z.G.]; Ministry of Science and Higher Education; European Fund for Regional Development [UDA-POIG.02.01.00–30–182/09 to Z.G.]; EU COST action MP0802 [STSM n°11227 to M.M.]; Fonds de la Recherche Scientifique-FNRS [research associate position and FRFC grant 2.4528.11 to V.G.]. Source of open access funding: National Science Center [UMO-2013/08/A/ST5/00295].
Conflict of interest statement. None declared.

REFERENCES

1. Simonsson,T. (2001) G-quadruplex DNA structures—variations on a theme. *Biol. Chem.*, **382**, 621–628.
2. Davis,J.T. (2004) G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed. Engl.*, **43**, 668–698.
3. Phan,A.T., Kuryavyi,V. and Patel,D.J. (2006) DNA architecture: from G to Z. *Curr. Opin. Struct. Biol.*, **16**, 288–298.
4. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
5. 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.
6. Miyoshi,D., Nakao,A. and Sugimoto,N. (2002) Molecular crowding regulates the structural switch of the DNA G-quadruplex. *Biochemistry*, **41**, 15017–15024.
7. Ambrus,A., Chen,D., Dai,J., Bialis,T., Jones,R.a. and Yang,D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.*, **34**, 2723–2735.

8. Heddi,B. and Phan,A.T. (2011) Structure of human telomeric DNA in crowded solution. *J. Am. Chem. Soc.*, **133**, 9824–9833.
9. Marchand,A., Ferreira,R., Tateishi-Karimata,H., Miyoshi,D., Sugimoto,N. and Gabelica,V. (2013) Sequence and solvent effects on telomeric DNA bimolecular G-quadruplex folding kinetics. *J. Phys. Chem. B*, **117**, 12391–12401.
10. Tran,P.L.T., De Cian,A., Gros,J., Moriyama,R. and Mergny,J.-L. (2013) Tetramolecular quadruplex stability and assembly. *Top. Curr. Chem.*, **330**, 243–273.
11. Matsugami,A., Ouhashi,K., Kanagawa,M., Liu,H., Kanagawa,S., Uesugi,S. and Katahira,M. (2001) An intramolecular quadruplex of $(GGA)_4$ triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *J. Mol. Biol.*, **313**, 255–269.
12. Phan,A.T., Kuryavyi,V., Ma,J.-B., Faure,A., Andréola,M.-L. and Patel,D.J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 634–639.
13. Borbone,N., Amato,J., Oliviero,G., D'Atri,V., Gabelica,V., De Pauw,E., Piccialli,G. and Mayol,L. (2011) d(CGGTGGT) forms an octameric parallel G-quadruplex via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads. *Nucleic Acids Res.*, **39**, 7848–7857.
14. Islam,B., Sgobba,M., Laughton,C., Orozco,M., Sponer,J., Neidle,S. and Haider,S. (2013) Conformational dynamics of the human propeller telomeric DNA quadruplex on a microsecond time scale. *Nucleic Acids Res.*, **41**, 2723–2735.
15. Maizels,N. and Gray,L.T. (2013) The G4 genome. *PLoS Genet.*, **9**, e1003468.
16. Lipps,H.J. and Rhodes,D. (2009) G-quadruplex structures: *in vivo* evidence and function. *Trends Cell Biol.*, **19**, 414–422.
17. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
18. Huppert,J.L., Bugaut,A., Kumari,S. and Balasubramanian,S. (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.*, **36**, 6260–6268.
19. Djebali,S., Davis,C.a., Merkel,A., Dobin,A., Lassmann,T., Mortazavi,A., Tanzer,A., Lagarde,J., Lin,W., Schlesinger,F. *et al.* (2012) Landscape of transcription in human cells. *Nature*, **489**, 101–108.
20. Millevoi,S., Moine,H. and Vagner,S. (2012) G-quadruplexes in RNA biology. *Wiley Interdiscip. Rev. RNA*, **3**, 495–507.
21. Tóth,G., Gáspári,Z. and Jurka,J. (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.*, **10**, 967–981.
22. Kozlowski,P., de Mezer,M. and Krzyzosiak,W.J. (2010) Trinucleotide repeats in human genome and exome. *Nucleic Acids Res.*, **38**, 4027–4039.
23. Orr,H.T. and Zoghbi,H.Y. (2007) Trinucleotide repeat disorders. *Annu. Rev. Neurosci.*, **30**, 575–621.
24. 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.
25. 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**, 869–878.
26. Glass,I.A. (1991) X-linked mental retardation. *J. Med. Genet.*, **28**, 361–371.
27. Sobczak,K., Michlewski,G., de Mezer,M., Kierzek,E., Krol,J., Olejniczak,M., Kierzek,R. and Krzyzosiak,W.J. (2010) Structural diversity of triplet repeat RNAs. *J. Biol. Chem.*, **285**, 12755–12764.
28. 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.
29. Zumwalt,M., Ludwig,A., Hagerman,P.J. and Dieckmann,T. (2007) Secondary structure and dynamics of the r(CGG) repeat in the mRNA of the fragile X mental retardation 1 (*FMR1*) gene. *RNA Biol.*, **4**, 93–100.
30. 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.
31. Khateb,S., Weisman-Shomer,P., Hershco,I., Loeb,L.A. and Fry,M. (2004) Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)n is mediated by homolog-conserved domains in three members of the hnRNP family. *Nucleic Acids Res.*, **32**, 4145–4154.
32. 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.
33. 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.
34. Mashima,T., Matsugami,A., Nishikawa,F., Nishikawa,S. and Katahira,M. (2009) Unique quadruplex structure and interaction of an RNA aptamer against bovine prion protein. *Nucleic Acids Res.*, **37**, 6249–6258.
35. Joachimi,A., Benz,A. and Hartig,J.S. (2009) A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg. Med. Chem.*, **17**, 6811–6815.
36. Beaucage,S.L. and Caruthers,M.H. (1981) Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.*, **22**, 1859–1862.
37. Xia,T., SantaLucia,J., Burkard,M.E., Kierzek,R., Schroeder,S.J., Jiao,X., Cox,C. and Turner,D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**, 14719–14735.
38. Kierzek,E. and Kierzek,R. (2003) The synthesis of oligoribonucleotides containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines via post-synthetic modification of precursor oligomers. *Nucleic Acids Res.*, **31**, 4461–4471.
39. Uddin,M.K., Kato,Y., Takagi,Y., Mikuma,T. and Taira,K. (2004) Phosphorylation at 5' end of guanosine stretches inhibits dimerization of G-quadruplexes and formation of a G-quadruplex interferes with the enzymatic activities of DNA enzymes. *Nucleic Acids Res.*, **32**, 4618–4629.
40. Pedersen,E.B., Nielsen,J.T., Nielsen,C. and Filichev,V.V. (2011) Enhanced anti-HIV-1 activity of G-quadruplexes comprising locked nucleic acids and intercalating nucleic acids. *Nucleic Acids Res.*, **39**, 2470–2481.
41. Hud,N.V., Schultze,P. and Feigon,J. (1998) Ammonium ion as an NMR probe for monovalent cation coordination sites of DNA quadruplexes. **120**, 6403–6404.
42. Gros,J., Rosu,F., Amrane,S., De Cian,A., Gabelica,V., Lacroix,L. and Mergny,J.-L. (2007) Guanines are a quartet's best friend: impact of base substitutions on the kinetics and stability of tetramolecular quadruplexes. *Nucleic Acids Res.*, **35**, 3064–3075.
43. Collie,G.W., Parkinson,G.N., Neidle,S., Rosu,F., De Pauw,E. and Gabelica,V. (2010) Electrospray mass spectrometry of telomeric RNA (TERRA) reveals the formation of stable multimeric G-quadruplex structures. *J. Am. Chem. Soc.*, **132**, 9328–9334.
44. Balthasart,F., Plavec,J. and Gabelica,V. (2013) Ammonium ion binding to DNA G-quadruplexes: do electrospray mass spectra faithfully reflect the solution-phase species? *J. Am. Soc. Mass Spectrom.*, **24**, 1–8.
45. Sosnick,T.R., Fang,X. and Shelton,V.M. (2000) Application of circular dichroism to study RNA folding transitions. *Methods Enzymol.*, **317**, 393–409.
46. Hwang,T.-L. and Shaka,A.J. (1995) Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients. *J. Magn. Reson. Ser. A*, **112**, 275–279.
47. Mergny,J.-L. and Lacroix,L. (2009) UV melting of G-quadruplexes. *Curr. Protoc. Nucleic Acid Chem.*, **37**, 17.1–17.15.
48. Mergny,J.-L., Li,J., Lacroix,L., Amrane,S. and Chaires,J.B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138.
49. Randazzo,A., Spada,G.P. and Webba da Silva,M. (2013) Circular dichroism of quadruplex structures. *Top. Curr. Chem.*, **330**, 67–86.
50. Nishikawa,F., Murakami,K., Matsugami,A., Katahira,M. and Nishikawa,S. (2009) Structural studies of an RNA aptamer containing GGA repeats under ionic conditions using microchip electrophoresis, circular dichroism, and 1D-NMR. *Oligonucleotides*, **19**, 179–190.

51. Lipay,J.M. and Mihailescu,M.-R. (2009) NMR spectroscopy and kinetic studies of the quadruplex forming RNA r(UGGAGGU). *Mol. Biosyst.*, **5**, 1347–1355.
52. Liu,H., Kugimiya,A., Matsugami,A., Katahira,M. and Uesugi,S. (2002) Quadruplex structures of RNA 14-mer, r(GGAGGUUUUGGAGG) and DNA 14-mer, d(GGAGGTTTTGGAGG). *Nucleic Acids Res.*, **2**, 177–178.
53. Fürtig,B., Richter,C., Wöhnert,J. and Schwalbe,H. (2003) NMR spectroscopy of RNA. *Chembiochem*, **4**, 936–962.
54. Webba da Silva,M. (2003) Association of DNA quadruplexes through G:C:G:C tetrads. Solution structure of d(GCGGTGGAT). *Biochemistry*, **42**, 14356–14365.
55. Xu,Y., Ishizuka,T., Kimura,T. and Komiyama,M. (2010) A U-tetrad stabilizes human telomeric RNA G-quadruplex structure. *J. Am. Chem. Soc.*, **132**, 7231–7233.
56. Cheong,C. and Moore,P.B. (1992) Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structured. *Biochemistry*, **31**, 8406–8414.
57. Patel,P.K. and Hosur,R.V. (1999) NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisiae* telomere repeats. *Nucleic Acids Res.*, **27**, 2457–2464.
58. Oliviero,G., Amato,J., Borbone,N., Galeone,A., Varra,M., Piccialli,G. and Mayol,L. (2006) Synthesis and characterization of DNA quadruplexes containing T-tetrads formed by bunch-oligonucleotides. *Biopolymers*, **81**, 194–201.
59. Randazzo,A., Esposito,V., Ohlenschläger,O., Ramachandran,R. and Mayola,L. (2004) NMR solution structure of a parallel LNA quadruplex. *Nucleic Acids Res.*, **32**, 3083–3092.
60. Kettani,A., Gorin,A., Majumdar,A., Hermann,T., Skripkin,E., Zhao,H., Jones,R. and Patel,D.J. (2000) A dimeric DNA interface stabilized by stacked A · (G · G · G · G) · A hexads and coordinated monovalent cations. *J. Mol. Biol.*, **297**, 627–644.
61. Liu,H., Matsugami,A., Katahira,M. and Uesugi,S. (2002) A dimeric RNA quadruplex architecture comprised of two G:G(:A):G:G(:A) hexads, G:G:G:G tetrads and UUUU loops. *J. Mol. Biol.*, **322**, 955–970.
62. Mergny,J.-L., De Cian,A., Amrane,S. and Webba da Silva,M. (2006) Kinetics of double-chain reversals bridging contiguous quartets in tetramolecular quadruplexes. *Nucleic Acids Res.*, **34**, 2386–2397.
63. 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.
64. 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**, 10873–82.
65. Ilc,T., Šket,P., Plavec,J., Webba da Silva,M., Drevenšek-Olenik,I. and Spindler,L. (2013) Formation of G-wires: the role of G:C-base pairing and G-quartet stacking. *J. Phys. Chem. C*, **117**, 23208–23215.
66. Pan,B., Xiong,Y., Shi,K. and Sundaralingam,M. (2003) Crystal structure of a bulged RNA tetraplex at 1.1 Å resolution. *Structure*, **11**, 1423–1430.
67. Pan,B., Shi,K. and Sundaralingam,M. (2006) Base-tetrad swapping results in dimerization of RNA quadruplexes: implications for formation of the i-motif RNA octaplex. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 3130–3134.
68. Popenda,L., Adamiak,R.W. and Gdaniec,Z. (2008) Bulged adenosine influence on the RNA duplex conformation in solution. *Biochemistry*, **47**, 5059–5067.
69. Popenda,L., Bielecki,L., Gdaniec,Z. and Adamiak,R. (2009) Structure and dynamics of adenosine bulged RNA duplex reveals formation of the dinucleotide platform in the C:G-A triple. *Arkivoc*, **2009**, 130–144.
70. Hermann,T. and Patel,D. (2000) RNA bulges as architectural and recognition motifs. *Structure*, **8**, 47–54.
71. Sudarsanakumar,C., Xiong,Y. and Sundaralingam,M. (2000) Crystal structure of an adenine bulge in the RNA chain of a DNA·RNA hybrid, d(CTCCTCTTC)·r(gaagagagag). *J. Mol. Biol.*, **299**, 103–112.
72. Thiviyanathan,V., Guliaev,A.B., Leontis,N.B. and Gorenstein,D.G. (2000) Solution conformation of a bulged adenosine base in an RNA duplex by relaxation matrix refinement. *J. Mol. Biol.*, **300**, 1143–1154.
73. Xiong,Y., Deng,J., Sudarsanakumar,C. and Sundaralingam,M. (2001) Crystal structure of an RNA duplex r(gugucgcac)$_2$ with uridine bulges. *J. Mol. Biol.*, **313**, 573–582.