Published online 30 January 2020 *Nucleic Acids Research, 2020, Vol. 48, No. 5* **2749–2761**
doi: 10.1093/nar/gkaa058

GC ends control topology of DNA G-quadruplexes and their cation-dependent assembly

Daša Pavc[1,2], Baifan Wang[1], Lea Spindler[3,4], Irena Drevenšek-Olenik[4,5], Janez Plavec[1,2,6] and Primož Šket[1,*]

[1]Slovenian NMR Center, National Institute of Chemistry, 1000 Ljubljana, Slovenia, [2]University of Ljubljana, Faculty of Chemistry and Chemical Technology, 1000 Ljubljana, Slovenia, [3]University of Maribor, Faculty of Mechanical Engineering, 2000 Maribor, Slovenia, [4]Department of Complex Matter, Jozef Stefan Institute, 1000 Ljubljana, Slovenia, [5]University of Ljubljana, Faculty of Mathematics and Physics, 1000 Ljubljana, Slovenia and [6]EN-FIST Center of Excellence, 1000 Ljubljana, Slovenia

Received October 11, 2019; Revised January 17, 2020; Editorial Decision January 20, 2020; Accepted January 21, 2020

ABSTRACT

***GCn* and *GCnCG*, where *n* = ($G_2AG_4AG_2$), fold into well-defined, dimeric G-quadruplexes with unprecedented folding topologies in the presence of Na^+ ions as revealed by nuclear magnetic resonance spectroscopy. Both G-quadruplexes exhibit unique combination of structural elements among which are two G-quartets, A(GGGG)A hexad and GCGC-quartet. Detailed structural characterization uncovered the crucial role of 5′-GC ends in formation of *GCn* and *GCnCG* G-quadruplexes. Folding in the presence of $^{15}NH_4^+$ and K^+ ions leads to 3′–3′ stacking of terminal G-quartets of *GCn* G-quadruplexes, while 3′-GC overhangs in *GCnCG* prevent dimerization. Results of the present study expand repertoire of possible G-quadruplex structures. This knowledge will be useful in DNA sequence design for nanotechnological applications that may require specific folding topology and multimerization properties.**

INTRODUCTION

Guanine-rich DNA sequences have propensity to fold into non-canonical, four stranded structures with G-quadruplexes being most well-known. Their main building block is a G-quartet formed by four guanine residues in planar arrangement held together by eight Hoogsteen-type hydrogen bonds. Cations, coordinated between G-quartets reduce repulsion of negatively charged O6 guanine atoms and are therefore crucial for G-quadruplex formation (1,2). Folding topologies are susceptible to changes in environment like pH, temperature and molecular crowding conditions as well as nature and concentration of cations, which all together add to wide repertoire of structural polymorphs. Among factors contributing to vast structural diversity of G-quadruplex structures are also different possibilities of base pairing alignments in addition to classical Hoogsteen-type hydrogen bonding between guanines in G-quartets. G-rich DNA sequences containing cytosine residues can be stabilized through formation of mixed GCGC-quartets in a major groove (3–6), minor groove (7,8) and slipped arrangement (9,10) or prefer formation of other tetrahelical structures such as AGCGA-quadruplexes (11). When G-tracts are separated by adenine residue, A(GGGG) pentads (12,13), A(GGGG)A hexads (14–16), G(A)G(A)G(A)G heptads (17,18) and mixed GAGA-quartets (11) could be formed.

It is believed that G-quadruplexes have important role in regulation of biological processes since G-rich sequences are over represented in human genome such as telomeres, promoter regions and even in genes connected with neurodegenerative diseases (19–21). In addition, G-quadruplexes also gained great attention in the field of nanotechnology. Their self-assembling ability, programmable control of their shape and size and unique optical and electrochemical properties make them attractive candidates for nanotechnological applications such as nano-electronics (22–25), nanosensors (26,27) and nanodevices (28). G-rich DNA oligonucleotides are able to form long, continuous nanostructures termed G-wires (29). In a recent review by Professors Mergny and Sen, G-wire was defined as an extended DNA nanostructure in one-dimension, formed by the self-assembly of one or more individual DNA oligonucleotides by way of G-quadruplex formation (30). One of the possibilities to assemble G-wires is through multimerization of individual G-quadruplex subunits. G-quadruplexes can form multimers via stacking or interlocking. Typically parallel G-quadruplexes with 'blunt-ends' can stack through π–π interactions of terminal G-quartets (31). Another possibility of stacking is through expanded π-

*To whom correspondence should be addressed. Tel: +386 1 4760223; Fax: +386 1 4760300; Email: primoz.sket@ki.si

© The Author(s) 2020. 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 Non-Commercial 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

systems such as hexads, heptads and octads, which facilitate association (14,17,32,33). Depending on which side of G-quadruplex stacking occurs, can be further classified as 5′–5′ (head-to-head) (34–38), 3′-3′ (tail-to-tail) (39,40) and 5′-3′ (head-to-tail) stacking (40,41). Potassium ions are known to more efficiently promote stacking in comparison to ammonium or sodium ions (42). Interlocks can be formed via: (i) extra G-quartet(s) formed by slipped G-rich strands from different G-quadruplexes or (ii) extra quartets formed by sticky ends. Similarly as stacking, interlocking can be classified as 3′–3′ (43), 5′–3′ and the most commonly observed 5′–5′ (6,12,13,15,16,40,44–46). A lot of efforts have been put into prediction and programming of multimerization of G-quadruplexes (31,41,47–49). One of the promising approaches for programmed self-assembly is via complementary GC ends, which could form linkages between two successive G-quadruplexes via inter-quadruplex GCGC-quartet formation (6,40,43,50,51). 5′-GC ends commonly form 5′-5′ interlocks (6,40,43,51), while the effect of 3′-GC ends has been reported to be more diverse (40,43,50).

In the presence of Na^+ ions, oligonucleotides *d*(GC$G_2TG_4TG_2$) and *d*(GC$G_2TG_4TG_2$CG) fold into dimeric, parallel G-quadruplexes composed of four G-quartets connected with thymine residues in propeller loops (40). Both G-quadruplexes exhibit two GC overhangs at their 5′-ends, which enable dimerization via inter-quadruplex GCGC-quartets. 5′-5′ interlocked G-quadruplexes adopted by *d*(GC$G_2TG_4TG_2$) undergo further multimerization via 3′–3′ stacking of terminal G-quartets resulting in four assembled G-quadruplexes. In contrast, the 3′-GC ends of a G-quadruplex formed by *d*(GC$G_2TG_4TG_2$CG) prevent multimerization.

Herein, we utilize nuclear magnetic resonance (NMR) experiments complemented with native PAGE, UV and circular dichroism (CD) spectroscopy in order to study DNA oligonucleotides *GCn* and *GCnCG,* where $n = G_2AG_4AG_2$. Adenine instead of thymine as a nucleobase separating G-tracts within *GCn* and *GCnCG* was expected to form additional hydrogen bonds with guanine residues, which can result in formation of A(GGGG) pentad or A(GGGG)A hexad and might thereby lead to additional stabilization of G-quadruplexes. Alternatively, adenines within G-rich oligonucleotides, might induce folding via highly stabilizing GNA type of loop leading to formation of different structures (52). Moreover, adenine residues might facilitate multimerization of G-quadruplexes when compared to other nucleobases (41,50). Additionally, mixed GAGA-quartets could be formed, however with specific requirements for grooves' dimensions that are critical for their ability to stack on a nearby G-quartet (11). Since both *GCn* and *GCnCG* exhibit GC ends, formation of intra-quadruplex GCGC-quartets rather than free overhangs is expected to control dimerization (multimerization) of individual G-quadruplexes. Furthermore, we deemed it essential to explore effect of various monovalent cations since GCGC-quartets stacked on G-quartets have been shown to exhibit specific (non)selectivity with respect to cation and water localization (53). At the outline, spectral features suggested alternative folding topologies and thus insights into the role of GC ends at the 5′- and 3′-ends of DNA sequence on G-quadruplex formation.

MATERIALS AND METHODS

Sample preparation

DNA oligonucleotides *GCn d*(GC$G_2AG_4AG_2$) and *GC nCG d*(GC$G_2AG_4AG_2$CG) were synthesized in DMT-off mode using DNA/RNA synthesizer H-8 (K&A Laborgeraete GbR) and standard phosphoramidite chemistry. *GCn* and *GCnCG* containing 8% ^{13}C and ^{15}N site-specifically labelled guanine and adenine residues were synthesized using the same procedure. Deprotection was done at 55°C over night with the use of aqueous ammonia which was later removed under low pressure. Samples were heated at 95°C for 5 min in the presence of LiCl and left to cool at room temperature. Desalting was done on Amicon ultrafilter at pH 10, which was adjusted with the use of LiOH solution. Concentrations of prepared samples were determined by measuring absorption at 260 nm on Varian CARY-100 BIO UV-VIS spectrometer. The molar extinction coefficients were calculated using the nearest neighbor method and were 125 300 M^{-1} cm^{-1} and 142 000 M^{-1} cm^{-1} for *GCn* and *GCnCG*, respectively. NMR samples were dissolved in 90% H_2O and 10% of D_2O in 100 or 175 mM salt concentration (NaCl, $^{15}NH_4Cl$ or KCl) and 10 mM NaPi or KPi buffer (pH 6.8). Samples containing 100% D_2O were prepared by lyophilization of previously prepared NMR samples and subsequently dissolved in 100% D_2O.

NMR spectroscopy

All NMR experiments were recorded on Agilent Technologies DD2 600 MHz and VNMRS 800 MHz NMR spectrometers at 25°C unless stated otherwise. For suppression of the water signal, the double-pulsed field gradient spin echo (DPFGSE) pulse sequence was used. The translation diffusion coefficients were obtained with the use of pulse field gradient stimulated echo (PFG-STE) pulse sequence. Identification of guanine H1 protons in partially (8%) ^{13}C and ^{15}N site-specific labelled samples was acquired with 1D ^{15}N-edited heteronuclear single quantum correlation (HSQC) experiment. Aromatic protons of adenines and guanines were identified with the use of 2D ^{13}C-edited HSQC experiment. Non-exchangeable proton resonances were assigned using the 2D Nuclear Overhauser Effect SpectroscopY (NOESY) with mixing times (τ_m) of 80, 150 and 250 ms recorded on NMR samples in 100% D_2O. 2D Total Correlation Spectroscopy (TOCSY) with τ_m of 80 ms, 2D Double Quantum Filtered Correlation SpectroscopY (DQF-COSY) and 2D ^1H-^{31}P COSY were used for cross-checking assignment of 2D NOESY spectra. Exchangeable proton resonances were assigned using 2D NOESY experiments with τ_m of 150 and 250 ms acquired on samples in 90% H_2O, 10% D_2O.

UV spectroscopy

Samples for UV melting experiment were prepared by diluting NMR samples with blank solution, which contained 175 mM NaCl and 10 mM NaPi buffer (pH 6.8), to achieve final 0.01, 0.05, 0.15 or 0.2 mM concentration of *GCn* and *GCnCG* per strand. The melting experiment was carried out on Varian CARY-100 BIO UV-VIS spectrophotometer

with the Cary Win UV Thermal program using 10, 5, 2 and 1 mm path-length cells, respectively. The temperature was increased/decreased from 15 to 80 or 90°C with the rate of 0.1°C/min. Absorbance was measured at 295 nm. To prevent sample evaporation at higher temperatures the mineral oil and fixed cuvette caps were used. To avert condensation at lower temperatures the stream of nitrogen was applied. Melting temperatures were determined from the first derivative of A_{295} versus temperature plot.

CD spectroscopy

Circular dichroism (CD) spectra were recorded on an Applied Photophysics Chirascan CD spectrometer at 25°C from 200 to 320 nm. A blank sample containing only 100 mM NaCl and 10 mM NaPi buffer (pH 6.8) was used for baseline correction. Measurements were carried out in 0.1 mm path-length quartz cells for samples with 1.0 mM concentration per strand.

Native PAGE

Native gel electrophoresis was run in a temperature controlled vertical electrophoretic apparatus at 15°C in TBE buffer and 175 mM NaCl. Samples contained 0.1, 0.25 and 0.5 mM concentration of *GCn* or *GCnCG* per strand, 175 mM NaCl and 10 mM NaPi buffer (pH 6.8) in 90% H_2O and 10% D_2O solution. Ficoll was added to the samples prior to loading. Polyacrylamide gel concentration was 15% with 100 mM concentration of NaCl. Thermo Scientific GeneRuler Ultra Low Range DNA Ladder was used as standard. Electrophoresis was run at 50 mV for 16 h. DNA was visualized by Stains-all (Sigma-Aldrich) staining.

Structure calculations

The structures of *GCn* and *GCnCG* were calculated by the simulated annealing (SA) simulations based on NOE-derived distance restraints. Distance restraints for exchangeable and nonexchangeable protons were obtained from 2D NOESY spectra (τ_m 150 ms), recorded on samples with 10% and 100% D_2O, respectively. Spectra were recorded at 25°C, 175 mM NaCl and 10 mM NaPi buffer (pH 6.8). In case of *GCn*, volumes of C2 H2′-H2′ (1.9 Å) and C2 H5−H6 cross-peaks (2.45 Å) were used as references, for exchangeable and nonexchangeable protons, respectively. In case of *GCnCG*, average volume of C2 and C13 H5-H6 cross-peaks (2.45 Å) was used as reference for exchangeable and nonexchangeable protons. NOEs were classified as strong (1.8−3.6 Å), medium (2.5−5.0 Å) and weak (3.5−6.5 Å). SA simulations were performed using the CUDA version of pmemd module of AMBER 14 program suites (54,55) and Cornell *et al.* force field basic version parm99 (56) with the bsc0 (57), χOL4 (58), ε/ζOL1 (59) and βOL1 (60) refinements. The initial extended single-stranded DNA structure was obtained using the leap module of AMBER 14 program suites. A total of 200 structures were calculated in 80 ps of NMR restrained simulated annealing (SA) simulations using the generalized Born implicit solvation model (61,62). The cut-off for non-bonded interactions was 999 Å and the SHAKE algorithm (63) for hydrogen atoms was used with the 0.4 fs time steps. For each SA simulation, a random velocity was used. The SA simulation was as follows: in 0–2 ps, the temperature was raised from 300 K to 1000 K and held constant at 1000 K for 38 ps. Temperature was scaled down to 500 K in the next 24 ps and reduced to 100 K in the next 8 ps and was further reduced to 0 K in the last 8 ps. Restraints used in the calculation were NOE-derived distance restraints (force constant 20 kcal mol^{-1} $Å^{-2}$), hydrogen bond (force constant 40 kcal mol^{-1} $Å^{-2}$), torsion angle χ, ε (region 170°−290°) and sugar pucker phase angle restraints (force constant 200 kcal mol^{-1} rad^{-2}). Based on the intensity of respective H8−H1′ 2D NOESY cross-peaks, glycosidic torsion angles of G1 in *GCn* and *GCnCG* were restrained to the *syn*-region (30°−90°), while glycosidic torsion of other residues of *GCn* and *GCnCG* were restrained to the *anti*-region (200° - 280°). Torsion angle ε of each residue was constrained to the allowed range in nucleosides, nucleotides, deoxyoligonucleotides, and deoxypolynucleotides (64). The phase angle of pseudorotation of A10 residue in *GCnCG* and *GCn* were restrained in the range from −10° to +40°, corresponding to North-type conformation (64). The phase angle of pseudorotation of G4, G6, G7, G8, G11, G12, C13 residues in *GCnCG* and G1, C2, G3, G4, G7, G8, G9, G11 and G12 residues in *GCn* were restrained in the range from 140° to 185°, corresponding to South-type conformation (64). For the rest of residues in *GCn* and *GCnCG*, their phase angle of pseudorotation were not restrained. NOE-derived distance restraints were used after the first 3000 steps. All structures were minimized with a maximum of 10 000 steps of energy minimization and a family of 10 structures was selected based on the smallest restraints violations and lowest energy.

RESULTS

GCn and *GCnCG* fold into symmetric, dimeric G-quadruplexes

Folding of oligonucleotides *GCn* and *GCnCG* into well-defined G-quadruplex structures in aqueous solution is achieved in the presence of NaCl (Figure 1). 1D ^{1}H NMR spectrum of *GCn* reveals six sharp signals in the region between δ 11.2 and 11.7 ppm corresponding to imino protons of guanine residues involved in Hoogsteen-type hydrogen bonds. Observation of six sharp signals that correspond to 12 imino protons in three G-quartets indicates that *GCn* adopts a symmetric G-quadruplex fold. A downfield sharp signal at δ 13.23 ppm is characteristic for H1 proton of guanine residue involved in Watson−Crick (WC) base pair. Additional weak and broad signal is observed at δ 11.1 ppm. Its broad nature may be due to location of respective atom in flexible region of G-quadruplex structure or due exposure to exchange with bulk solvent. 1D ^{1}H NMR spectrum of *GCnCG* exhibits the same number of signals in the imino region corresponding to Hoogsteen and WC base pairing as observed for *GCn* despite additional 3′-GC end, which suggests formation of similar folds for both oligonucleotides. A great similarities between *GCn* and *GCnCG* are also reflected in aromatic regions of 1D ^{1}H NMR spectra with two additional signals for *GCnCG* at δ 7.40 and 7.48 ppm corresponding to G14 and C13 residues in the latter (*vide infra*).

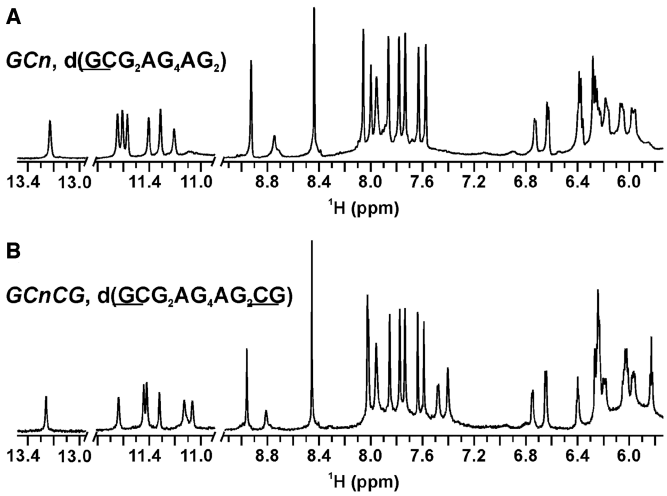

Figure 1. Expansion of selected regions of 1D ^{1}H NMR spectra of (**A**) *GCn* and (**B**) *GCnCG*. Spectra were recorded at 600 MHz, 25°C, 175 mM NaCl and 10 mM NaPi buffer (pH 6.8) in 90% H_2O/10% D_2O. Oligonucleotide concentrations were 1.0 mM.

Translational diffusion coefficients (D_t) at 25°C and 1.0 mM oligonucleotide concentrations are 1.6×10^{-10} and 1.5×10^{-10} m^2 s^{-1} for *GCn* and *GCnCG*, respectively. Observed values are in accordance with formation of dimeric G-quadruplexes formed by *GCn* and *GCnCG*. Similar size of both G-quadruplexes is nicely reflected in mobility on native polyacrylamide gel, where it can be clearly seen that *GCn* and *GCnCG* have the same gel mobility as approximately 12 base pairs (bp) long reference dsDNA (Supplementary Figure S1). Intermolecular nature of *GCn* and *GCnCG* G-quadruplexes is further supported by UV spectroscopy, which demonstrated concentration-dependence of temperature of mid-transition ($T_{1/2}$) in un/folding processes (Supplementary Figure S2A, B). G-quadruplexes display high thermal stability with $T_{1/2}$ of 64 and 68°C at 0.2 mM concentration per strand for *GCn* and *GCnCG*, respectively (Supplementary Figure S2A, B). From the experimentally determined slopes of van't Hoff plots, molecularities $n = 2$ were determined, which is in perfect agreement with dimeric nature of *GCn* and *GCnCG* G-quadruplexes (Supplementary Figure S2C, D).

Folding topology with unique combination of structural elements

Unambiguous assignment of imino and aromatic proton resonances of guanine and adenine residues was achieved by the use of 1D ^{15}N-edited and 2D ^{13}C-edited HSQC spectra acquired on ^{15}N, ^{13}C residue-specific partially (8%) labelled *GCn* and *GCnCG* (Supplementary Figures S3 and S4). Proton assignment was completed based on connectivities in 2D NOESY, TOCSY, ^{1}H-^{31}P COSY and DQF-COSY spectra (Figure 2, Supplementary Figures S5-S9). G1, G7, G8, G9, G11 and G12 residues are involved in formation of G-quartets, while G6 participates in WC base pair with C2. Broad signals at δ 11.1 ppm are assigned to G4H1 of *GCn* and *GCnCG*. G3H1 protons of *GCn* and *GCnCG* do not give observable signals, indicating that they are not involved in base pairing. No signal is observed for G14H1 proton of *GCnCG*. Interestingly, G1H8 protons of *GCn* and *GCnCG* resonate at δ 6.28 and 6.27 ppm, respectively, which is outside the range typical for guanine aromatic protons. On the other hand, A10H8 protons (δ 8.92 and 8.99 ppm for *GCn* and *GCnCG*, respectively) are observed downfield relative to the other aromatic protons.

Intensities of intra-nucleotide H8-H1′ cross-peaks in 2D NOESY (τ_m 80 ms) spectra revealed *syn* conformation along glycosidic bond for G1 and *anti* conformation for all other residues in *GCn* and *GCnCG* (Supplementary Figure S10). In addition to NOEs characteristic for *anti-anti* sequential steps, G1 displays G1H8−C2H1′ connectivity typical for *syn-anti* step of sequential walk (Figure 2). Interruption of sequential walk occurs only between G4 and A5, which implies unusual inter-nucleotide conformation in loop regions of G-quadruplexes adopted by *GCn* and *GCnCG*. Head-to-head stacking of bases characteristic for G1(*syn*)-C2(*anti*) step is nicely reflected in CD spectra of *GCn* and *GCnCG* with a shoulder at around 300 nm. The maxima at 270 and 265 nm in CD spectra of *GCn* and *GCnCG*, respectively as well as minima at 237 nm are characteristic for head-to-tail stacking of nucleobases, which is in agreement with *anti-anti* sequential steps (Supplementary Figure S11). CD spectrum of *GCnCG* exhibits additional shoulder at 285 nm, which is attributed to stacking interactions of 3′-GC ends with a nearby G-quartet.

Perusal of imino-aromatic regions of 2D NOESY spectra reveals formation of G1−G7−G1−G7, G8−G11−G8−G11 and G9−G12−G9−G12 quartets (labelled with green, violet and red, respectively in Figure 3A and Supplementary Figure S12A). Observation of only two imino-aromatic NOEs per G-quartet suggests their symmetric arrangement, where base pairs are formed by residues from different strands. G1(*syn*)−G7(*anti*)−G1(*syn*)−G7(*anti*) quartet exhibits variable arrangement of conformations along glycosidic bond, while all residues in the remaining two G-quartets adopt *anti* conformation. Mutual orientation of G-quartets is defined by inter-quartet imino-aromatic NOEs (labelled with black in Figure 3A and Supplementary Figure S12A). In addition, several inter-quartet imino-imino connectivities observed in 2D NOESY spectra (Supplementary Figure S12B) are in full agreement with the established G-quartets and their relative positions within G-quadruplexes (Figure 3B).

NOE correlations between hydrogen bonded amino G8H21 and G8H22 protons (δ 9.48 and 8.34 ppm for *GCn*, δ 9.33 and 8.36 ppm for *GCnCG*) with A10H8 suggest formation of sheared G8-A10 base pairs (Figure 3C and Supplementary Figure S12C). These correlations together with A10H8-G11H8 NOE interactions substantiate formation of A10−(G8−G11−G8−G11)−A10 hexads in *GCn* and *GCnCG* (Figure 3D). Positions of A10 residues are further defined by NOEs with G1 (G1H8−A10H2, G1H1′−A10H2, G1H1′−A10H8, G1H4′−A10H1′, G1H4′−A10H2, G1H5′−A10H2) and C2 (C2H4′−A10H2) residues in *GCn* and *GCnCG* G-quadruplexes. It is interesting to note that A10 residues adopt N-type sugar conformation as indicated by the absence of H1′-H2′ and the presence of an intense H3′−H4′

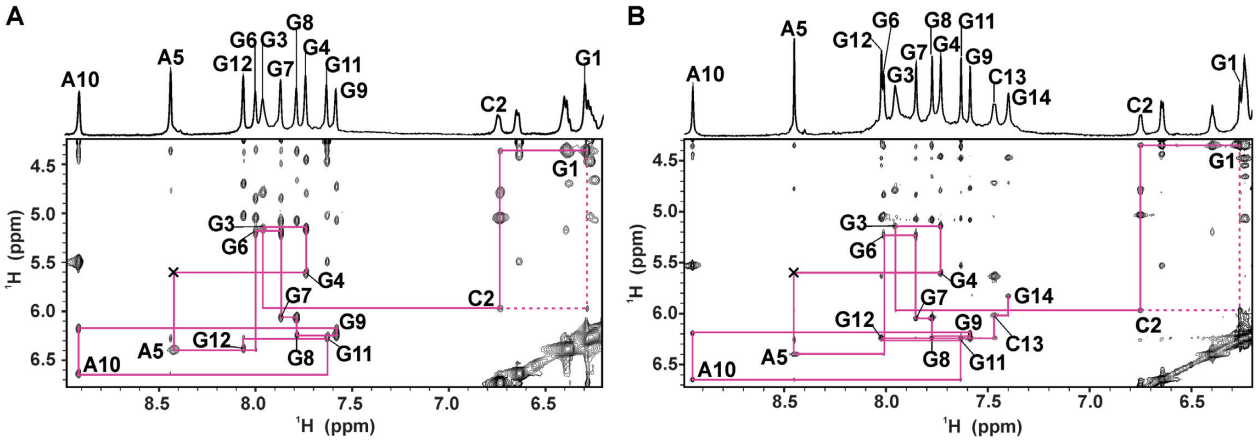

Figure 2. Aromatic-anomeric regions of 2D NOESY spectra (τ_m 150 ms) with marked sequential walks for (**A**) *GCn* and (**B**) *GCnCG*. The intra-nucleotide H8/H6$_{(n)}$−H1′$_{(n)}$ NOEs are labeled with residue numbers. Dashed lines show reverse inter-nucleotide H8$_{(i)}$−H1′$_{(i+1)}$ NOE interactions for G1(*syn*)−C2(*anti*) step. Interruptions of sequential walks between G4−A5 residues are denoted with crosses.

cross-peaks in 2D DQF-COSY spectra of *GCn* and *GCnCG* (Supplementary Figure S9).

G6−C2 base pair formation in WC geometry is supported by G6H1−C2H41 and G6H1−C2H42 NOEs. Downfield chemical shifts of C2H41 and C2H42 protons (δ 8.72 and 8.75 ppm for *GCn*, δ 8.81 and 8.96 ppm for *GCnCG*) confirm association of G6-C2 base pairs into G6−C2−G6−C2 quartet through their major groove edges (Figure 3E). Chemical shift of G4H1 protons (δ 11.1 ppm for *GCn* and *GCnCG*) suggests formation of symmetric G4−G4 N1-carbonyl base pairs, which is the only symmetric G-G base pair geometry with imino protons involved in hydrogen bonds (Figure 3F). 3′-GC ends in G-quadruplex adopted by *GCnCG* are less well defined in comparison to other parts of the structure, where C13 residues exhibit no stacking interactions, while G14 residues fold back as evident by G14H8−G12H8, G14H2′/H2′−G12H8 and G14H8−G11H8 cross-peaks in 2D NOESY spectra.

Diversity of structural elements and unique G-quadruplex folds adopted by *GCn* and *GCnCG* are nicely reflected in dispersion of ^{31}P NMR resonances. 1D ^{31}P spectra exhibit 11 and 13 resolved signals, which are dispersed between δ −1.4 and 0.5 ppm for *GCn* and *GCnCG*, respectively (Supplementary Figure S8). A10P and G11P resonate downfield relative to the other phosphorous atoms, which suggests unusual backbone conformation associated with formation of A(GGGG)A hexad.

Hexad is the most stable structural element

Local structural flexibility and unfolding were evaluated by analysis of a set of 1D ^{1}H NMR spectra acquired between 25 and 80°C on *GCn* (Supplementary Figure S13). As expected from its location in more solvent-exposed region of the structure, decrease in intensity of G4H1 signal is observed already at 30°C. Partial dissociation of G4−G4 N1-carbonyl base pair is reflected in changes in local environment as suggested by better dispersion of signals corresponding to C2H41, C2H42 and A5H8. At 40°C, complete dissociation of symmetric G4−G4 N1-carbonyl base pair occurs. Its destacking is reflected in downfield shift of neighboring G6H1 proton resonance. At the same temperature, intensities of G6H1, G9H1 and G12H1 proton signals start to decrease, which is well correlated with their location in the outer quartets of G-quadruplex. With opening of terminal G9−G12−G9−G12 quartet, G8H1 and A10H2 protons of the adjacent A10−(G8−G11−G8−G11)−A10 hexad became deshielded. Complete dissociation of G9−G12−G9−G12 and G6−C2−G6−C2 quartets occurs at 75°C, where signals corresponding to G9H1, G12H1, G6H1, C2H41 and C2H42 protons broaden to baseline. Interestingly, at 75°C signals belonging to residues within hexad are still detectable, indicating its high thermal stability. At 80°C signals corresponding to *GCn* G-quadruplex are no longer visible.

GCn and *GCnCG* form intertwined G-quadruplex structures

High-resolution G-quadruplex structures of *GCn* and *GCnCG* (Figure 4) were calculated using 748 and 804 NOE derived distance-restraints together with 149 and 135 torsion angle restraints, respectively in addition to 32 hydrogen bond restraints for both G-quadruplexes (Supplementary Table S1). Family of 10 refined structures was selected based on the lowest energy and the lowest number of restraint violations (Supplementary Figure S14). Overall pairwise heavy atom RMSDs for family of 10 refined structures are 1.01 ± 0.31 Å and 1.39 ± 0.36 Å for *GCn* and *GCnCG*, respectively. NMR based structural calculations resulted in a symmetric, dimeric, intertwined G-quadruplexes of *GCn* and *GCnCG*, which can be viewed as structures composed of two units – an antiparallel unit at the 5′- and a parallel unit at the 3′-end. The two units are connected with two phosphodiester linkages between G7 and G8 residues. Nucleobases in both G-quadruplexes exhibit extensive stacking interactions with partial overlap between their six- and five- membered rings (Figure 4A and Supplementary Figure S15).

The 5′-units with antiparallel topologies of *GCn* and *GCnCG* consist of two stacked quartets (G1−G7−G1−G7 and G6−C2−G6−C2) and two lateral loops formed by G3−G4−A5 residues. Loops are interconnected by symmetric G4−G4 N1-carbonyl and A5−A5 N1-amino

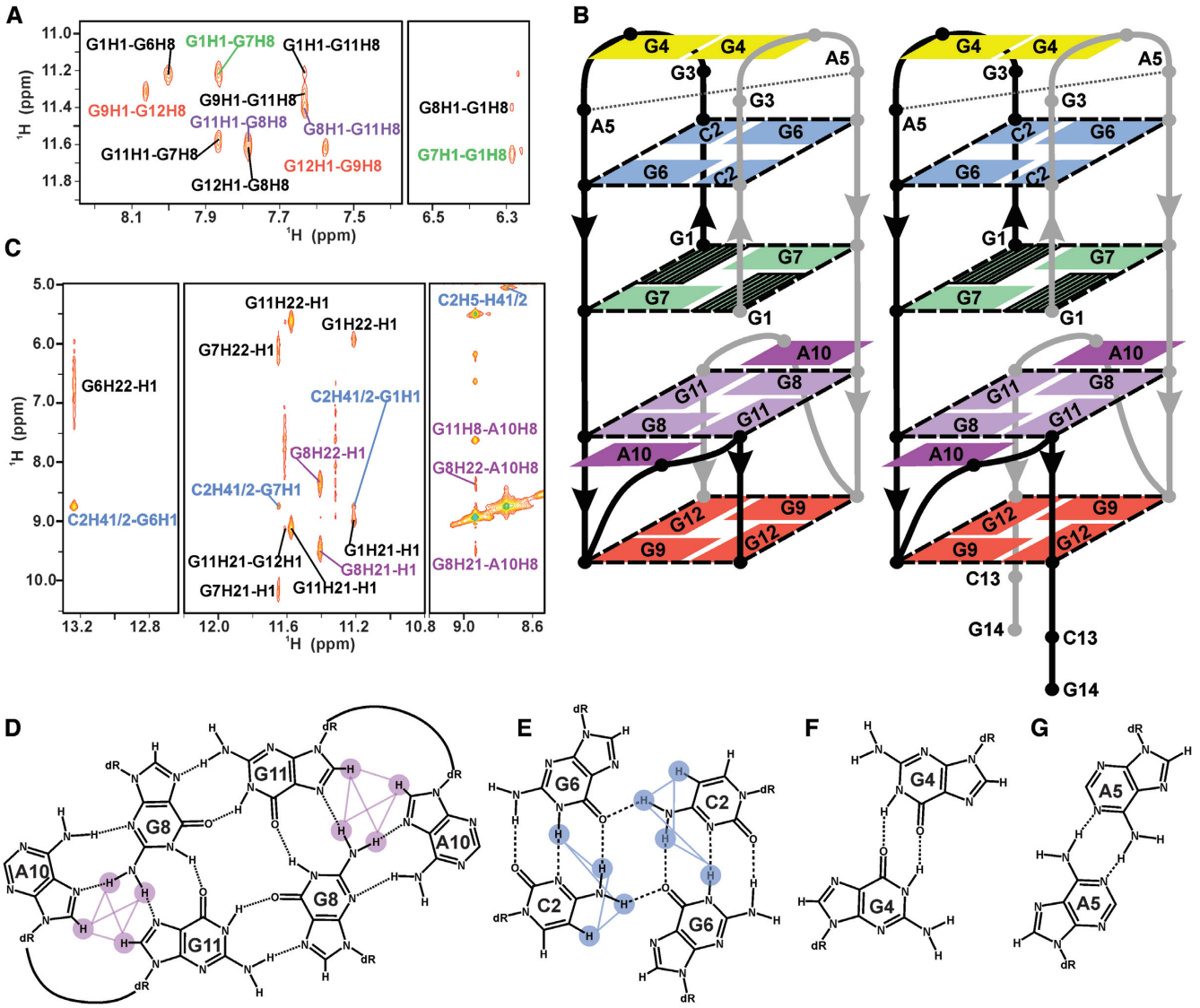

Figure 3. (**A**) Imino-aromatic region of 2D NOESY spectrum (τ_m 250 ms) of *GCn*. Cross-peaks coloured in green, violet and red indicate formation of G1–G7–G1–G7, G8–G11–G8–G11 and G9–G12–G9–G12 quartets, respectively. Inter-quartet NOE correlations are presented in black. (**B**) Folding topologies of *GCn* (left) and *GCnCG* (right) G-quadruplexes. Hatched rectangles indicate G1 in *syn* conformation, while all other residues adopt *anti* conformation. Dotted lines indicate symmetric A5-A5 N1-amino base pair revealed by MD simulations (*vide infra*). Dashed black lines indicate formation of GCGC- and G-quartets. (**C**) Imino-amino region of 2D NOESY spectrum (τ_m 150 ms) of *GCn*. NOE cross-peaks coloured in blue and violet support formation of major grove GCGC-quartet and A(GGGG)A hexad, respectively. Schematic representation of (**D**) A(GGGG)A hexad, (**E**) major groove GCGC-quartet, (**F**) symmetric GG N1-carbonyl base pair and (**G**) symmetric AA N1-amino base pair. Protons and observed NOE interactions crucial for formation of A(GGGG)A hexad and major groove GCGC-quartet are coloured in violet (C) or in blue (D), respectively.

base pairs (Figure 3F, G). Structural calculations suggest formation of A5–A5 base pair in *GCn* and *GCnCG*, which could be a consequence of favorable stacking interactions with lower G6–C2–G6–C2 quartet and upper symmetric G4–G4 N1-carbonyl base pair. The formation of A5–A5 base pair could not be verified experimentally due to dynamic nature or exposure to solvent. The 3′-units with parallel topologies consist of A10–(G8–G11–G8–G11)–A10 hexad and G9–G12–G9–G12 quartet in both *GCn* and *GCnCG* structures. Strands of the 3′-unit form single nucleotide (A10) double-chain reversal loops, which are part of A10–(G8–G11–G8–G11)–A10 hexad. The hexad is sandwiched between G9–G12–G9–G12 quartet and G1–G7–G1–G7 quartet (from the 5′-unit). In the case of *GCnCG*, the 3′-unit contains additional 3′-GC ends, which are rather flexible as seen in structural ensemble of 10 refined structures (Supplementary Figure S14). Consequently, G-quartets are less well defined in G-quadruplex of *GCnCG* (RMSD 0.62 ± 0.22 Å) compared to *GCn* (RMSD 0.40 ± 0.11 Å), which is reflected in distortion of planarity for G9–G12–G9–G12 and G8–G11–G8–G11 quartets (Supplementary Figure S14).

The antiparallel strands in the 5′-units of *GCn* and *GCnCG* form alternating narrow and wide grooves. The dimensions of narrow grooves formed by pairs of intra-strand

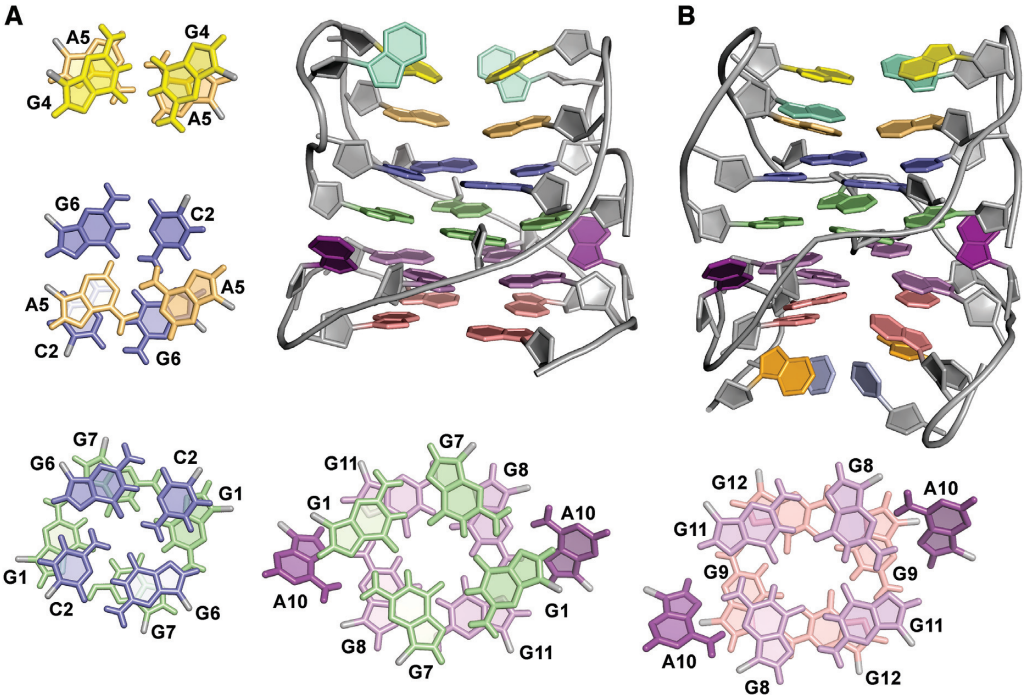

Figure 4. High-resolution structures of (**A**) *GCn* and (**B**) *GCnCG* G-quadruplexes. Stacking interactions between bases are shown for *GCn* in (A). Structural details and stacking (GCGC- and G-quartets, A(GGGG)A hexad, GG and AA base pairs) are presented with different colours for clarity.

G1–G7 and C2–G6 residues are extremely small with average width of 1.9 ± 0.9 Å and 3.1 ± 0.7 Å for *GCn* and *GCnCG*, respectively. Narrow grooves are bridged by G3–G4–A5 lateral loops. This trinucleotide arrangement is also known as GNA (G, guanine; N, any nucleotide; A, adenine) type of loop, which is known to form stable structures. Wide grooves formed by pairs of inter-strand G1–G7 and C2–G6 residues exhibit average width of 14.4 ± 0.5 Å and 14.6 ± 0.4 Å for *GCn* and *GCnCG*, respectively. The homo-purine G4–G4 and A5–A5 base pairs are formed diagonally on the wide grooves side in the 5′-units of the *GCn* and *GCnCG* G-quadruplexes. In contrast, the 3′-units with parallel topology of *GCn* and *GCnCG* exhibit four medium grooves of similar size with dimensions of 11.0 ± 0.3 Å for *GCn* and 11.4 ± 0.4 Å for *GCnCG*.

5′-GC ends and not GNA loops determine overall folds of *GCn* and *GCnCG* G-quadruplexes

Importance of 5′-GC in formation of *GCn* and *GCnCG* G-quadruplexes was evaluated by switching their respective positions in a sequence. Interestingly, d(CG$G_2AG_4AG_2$) forms completely different G-quadruplex structures with no WC GC base pairing (Supplementary Figure S16), which clearly demonstrates key role of WC GC pairing for stabilization and formation of intertwined G-quadruplexes.

In order to evaluate importance of GNA trinucleotide loop in formation of *GCn* and *GCnCG* G-quadruplexes, we have substituted A5 for thymine within G3–G4–A5. Results showed that A5T substitution in *d*(GC$G_2TG_4AG_2$) leads to formation of a structure similar to *GCn* G-quadruplex (Supplementary Figure S16). Therefore, it appears that G3–G4–A5 loop element is not determining factor for structures shown in Figure 3B.

3′–3′ stacking of *GCn* G-quadruplexes induced by $^{15}NH_4^+$ and K^+ ions

GCn and *GCnCG* are folded into well-defined G-quadruplexes in the presence of $^{15}NH_4^+$ and K^+ ions (Figure 5). Signals for H1 protons of *GCn* folded in the presence of $^{15}NH_4^+$ (Figure 5B) and K^+ (Figure 5C) ions resonate upfield with respect to the same protons of *GCn* folded in the presence of Na^+ ions (Figure 5A). Upfield H1 proton signals suggest additional stacking interactions, which are also reflected in upfield shifts of some aromatic protons in 1D ^{1}H NMR spectra of *GCn* G-quadruplex in the presence of $^{15}NH_4^+$ and K^+ ions. Lower D_t values obtained at 25°C and 1.0 mM oligonucleotide concentration for *GCn* folded in the presence of $^{15}NH_4^+$ and K^+ ions (1.3 × 10^{-10} m^2 s^{-1}) in comparison to D_t value in the presence of Na^+ ions (D_t 1.6 × 10^{-10} m^2 s^{-1}) are in agreement with formation of twofold bigger G-quadruplex structures.

Most of H1 and H8 proton resonances of *GCn* G-quadruplex in the presence of $^{15}NH_4^+$ ions, were unambiguously assigned with the use of 1D ^{15}N-edited and 2D ^{13}C-edited HSQC spectra recorded on ^{15}N and ^{13}C residue-specific partially (8%) labeled oligonucleotides (Supplementary Figure S17). The remaining H1 and H8/H6 proton resonances were assigned based on connectivities in 2D NOESY spectra (Supplementary Figure S18). H8 protons of G1 and A10 residues display upfield (δ 5.93 ppm) and downfield (δ 8.92 ppm) shifts, respectively, relative to the other aromatic protons. Intra- and inter-nucleotide connectivities in sequential walk are observed from G1 to G4 and

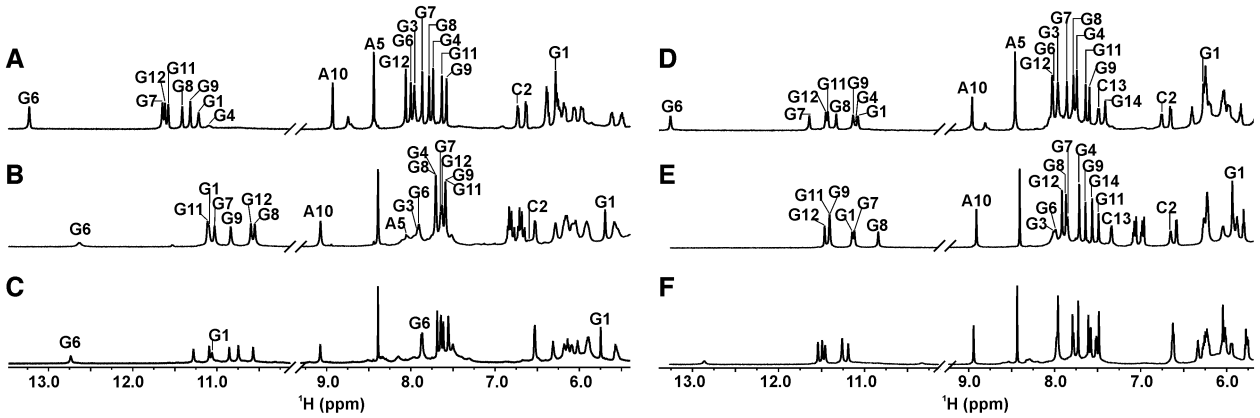

Figure 5. Selected regions of ^{1}H NMR spectra of *GCn* (**A**–**C**) and *GCnCG* (**D**–**F**) in the presence of 175 mM NaCl (**A**, **D**), $^{15}NH_4Cl$ (**B**, **E**) and KCl (**C**, **F**). The assignment of individual proton resonances is shown above individual spectrum. Spectra were recorded at 25°C, in 90% H_2O/10% D_2O and 1.0 mM oligonucleotide concentrations.

from G6 to G12 (Supplementary Figure S18A). Sequential walk is interrupted within G4–A5–G6 segment indicating their unusual inter-nucleotide conformation in the loop region. All residues in the *GCn* G-quadruplex folded in the presence of $^{15}NH_4^+$ ions adopt *anti* conformation along glycosidic bond except G1, which exhibits *syn* conformation. Analysis of imino-aromatic and imino-imino regions of 2D NOESY spectrum (Supplementary Figure S18B, C) reveals formation of G1–G7–G1–G7 and G9–G12–G9–G12 quartets and A10–(G8–G11–G8–G11)–A10 hexad (Supplementary Figure S18D), which is consistent with *GCn* G-quadruplex folded in the presence of Na^+ ions. The 5′-unit of *GCn* in the presence of $^{15}NH_4^+$ ions exhibits two WC GC base pairs as indicated by G6H1 signal at δ 12.72 ppm (Figure 6A). Lower intensity of G6H1 signal in the presence of $^{15}NH_4^+$ ions suggests that WC GC base pairs are not as well protected from the exchange with solvent or are more flexible than in *GCn* G-quadruplex folded in the presence of Na^+ ions. Further association of WC GC base pairs into GCGC-quartet is not observed in the presence of $^{15}NH_4^+$ ions since cross-peaks characteristic for its formation were not detected in 2D NOESY spectrum. Another difference observed in the 5′-unit of *GCn* G-quadruplex folded in the presence of $^{15}NH_4^+$ ions is the absence of symmetric G4–G4 N1-carbonyl base pair. In the 3′-unit stacking of terminal G-quartets is induced by the presence of $^{15}NH_4^+$ ions, which leads to dimerization of two dimeric *GCn* G-quadruplexes (Figure 6A). 3′–3′ stacking is supported by inter-residue NOE observed between G12H8-G12H1 protons (Supplementary Figure S18B). The 3′–3′ stacked *GCn* G-quadruplexes are able to form in the wide range of DNA concentration, as nicely demonstrated with a set of 1D ^{1}H NMR spectra, recorded with DNA concentration ranging from 3.0 to 0.09 mM per strand in the presence of 175 mM $^{15}NH_4^+$ ions (Supplementary Figure S19A). Interestingly, the 3′–3′ stacked *GCn* G-quadruplexes are preserved even at low DNA concentration (0.09 mM), which indicates their high stability.

In the presence of K^+ ions unambiguous assignment of imino and aromatic protons of G1 and G6 (Supplementary Figure S20) revealed similar chemical shifts as observed for the same protons in the presence of $^{15}NH_4^+$ ions (Figure 5B, C). In addition to the same D_t values obtained in the presence of both cations, comparable 1D ^{1}H and 2D NOESY NMR spectra confirmed similar *GCn* G-quadruplexes in the presence of $^{15}NH_4^+$ and K^+ ions (Figure 6A).

D_t values obtained at 25°C and 1.0 mM oligonucleotide *GCnCG* concentration folded in the presence of Na^+, $^{15}NH_4^+$ and K^+ ions (1.5×10^{-10} m^2s^{-1}) are the same, which suggest formation of dimeric G-quadruplexes. Therefore, 3′-GC ends in *GCnCG* irrespective of monovalent cation nature are not involved in inter-quadruplex GCGC-quartets but rather prevent formation of bigger G-quadruplex based structures. Contrary to differences observed for *GCn* G-quadruplexes formed in the presence of various monovalent cations, comparison of 1D ^{1}H (Figure 5D–F) and 2D NOESY NMR spectra of *GCnCG* G-quadruplexes folded in the presence of Na^+ (Figure 3B), $^{15}NH_4^+$ and K^+ ions reveals minor structural differences (Figure 6B). The 5′-unit of *GCnCG* G-quadruplexes in the presence of $^{15}NH_4^+$ and K^+ ions is characterized by less defined WC GC base pairs, which do not associate into GCGC-quartet. Additionally, no G4-G4 base pair was observed in the presence of $^{15}NH_4^+$ and K^+ ions. The molecularity of *GCnCG* G-quadruplex is preserved in DNA concentration ranging from 3.0 to 0.09 mM per strand in the presence of 175 mM $^{15}NH_4^+$ ions (Supplementary Figure S19B). Even at the highest DNA concentration no 3′-3′ interlocking was observed.

DISCUSSION

In the presence of Na^+ ions *GCn* and *GCnCG* form symmetric, dimeric G-quadruplexes, with unique combination of structural elements. *GCn* and *GCnCG* G-quadruplexes are composed of two G-quartets, A(GGGG)A hexad, major groove GCGC-quartet and symmetric GG N1-carbonyl and AA N1-amino base pairs. Only few high-resolution G-quadruplex structures containing major groove GCGC-quartets (3–6,46) and only one with A(GGGG)A hexad (14) in solution have been reported so far. Interestingly, 11 out of 12 residues in *GCn* are involved in hydrogen bonds. The only non-hydrogen bonded residues are

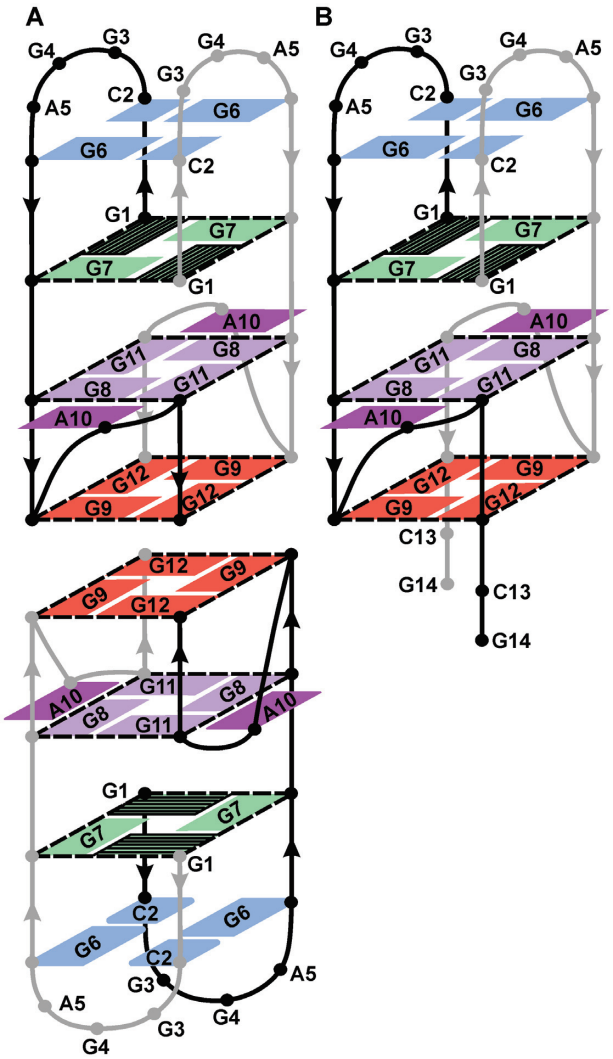

Figure 6. Folding topologies of (**A**) *GCn* and (**B**) *GCnCG* G-quadruplexes in the presence of $^{15}NH_4^+$ and K^+ ions. Hatched rectangles indicate *syn* conformation of G1 residues, while all other residues adopt *anti* conformation. Dashed black lines indicate formation of G-quartets. $^{15}NH_4^+$ and K^+ ions lead to structural changes in the 5′-units of *GCn* and *GCnCG*. In the case of *GCn*, $^{15}NH_4^+$ and K^+ ions induce dimerization through 3′-3′ stacking interactions, whereas the 3′-GC ends in *GCnCG* G-quadruplexes prevent formation of bigger structures.

G3. They are part of two lateral loops, which are interconnected by symmetric G4-G4 N1-carbonyl and A5-A5 N1-amino base pairs. In case of *GCnCG*, additional nucleobases at the 3′-end (C13 and G14) are not involved in hydrogen bonds. Both, *GCn* and *GCnCG* G-quadruplexes can be viewed as structures composed of two units with 5′-antiparallel and 3′-parallel topologies. G-quadruplex composed of 5′-antiparallel and 3′-parallel units was till now described only for crystal structure of oligonucleotide d($G_3CG_4AG_5A_2G_3A$) derived from B-raf gene in the presence of K^+ ions (65). In addition, this G-quadruplex contains distorted GCGC-quartet and similar arrangements of adenines as observed within A(GGGG)A hexad in *GCn* and *GCnCG* G-quadruplexes. Existence of *GCn* and *GCnCG* structures in solution undoubtedly demonstrates that G-quadruplexes composed of two distinct units with different strand orientations and various hydrogen bonding patters are not only a consequence of crystallization conditions in solid state. The two DNA strands intertwine in order to fold into G-quadruplexes, which is another interesting structural feature of *GCn* and *GCnCG* G-quadruplexes. Dimeric, intertwined G-quadruplexes were described before for oligonucleotides *d*($G_3CG_4AG_5A_2G_3A$), *d*($TAG_3CG_3AG_3AG_3A_2$) and d($CG_3CG_3CGCGAG_3AG_4$) (65–67).

Diverse structural elements of *GCn* and *GCnCG* G-quadruplexes are easily distinguished by NMR since involved residues display characteristic spectroscopic properties. A10H8 protons exhibit downfield shifts of δ 8.92 and 8.99 ppm for *GCn* and *GCnCG*, respectively. Such downfield shifts for adenine H8 protons were previously observed for adenine residues involved in a sheared G-A base pair within pentad (12) or heptad (18). Interestingly, not all H8 protons of adenines involved in formation of pentad, hexad or heptad display such downfield chemical shift (13,14,16–18). Another interesting feature of *GCn* and *GCnCG* G-quadruplexes are G1 residues, which display upfield shifts of their aromatic and anomeric protons (δ 6.28 and 4.36 ppm for *GCn*, δ 6.27 and 4.35 ppm for *GCnCG*, respectively). Such upfield shifts of aromatic and anomeric protons were observed for guanines stacked above G-(A-G) triad (12), which is in good agreement with location of G1 above A10-(G8–G11–G8–G11)–A10 hexad in *GCn* and *GCnCG* G-quadruplexes. Interestingly, only G1 residues in *GCn* and *GCnCG* G-quadruplexes exhibit *syn* conformation along the glycosidic bond. Furthermore, G1 residues occupy unusual position at the interface of two G-quadruplex units within the *GCn* and *GCnCG* G-quadruplexes. Although G1 residues are usually located in the outer G-quartets, their positions in the middle of G-quadruplex were observed before (12,68–70). Additionally, unusual upfield chemical shifts were observed for H4′ protons of G4 (δ 3.07 and 3.08 ppm for *GCn* and *GCnCG*, respectively) and G9 (δ 3.10 and 3.13 ppm for *GCn* and *GCnCG*, respectively). Such upfield shifts were previously observed for H4′ protons of N nucleotides in trinucleotide GNA segments when forming loops (52,71,72) or participating in mixed AG hexads (14) or heptads (17,18). This is in perfect agreement with our observations since G4H4′ are involved in G3–G4–A5 (lateral loops) and G9H4′ in G8–G9–A10 (A10 form single nucleotide double-chain reversal loops).

Both, *GCn* and *GCnCG* have 5′-GC ends, while *GCnCG* has additional 3′-GC end. 5′-GC ends exhibit propensity to participate in 5′–5′ interlocking of G-quadruplexes, which seems to be preferential way of multimerization via inter-quadruplex GCGC-quartet formation (6,16,40,46). No such interactions were observed for *GCn* and *GCnCG* G-quadruplexes. G1 and C2 (5′-GC ends) participate in G1–G7–G1–G7 and G6–C2–G6–C2 quartets in the fold-back motif of the 5′-unit with antiparallel topology of G-quadruplexes. Switching positions of G1 and C2 residues leads to formation of different structures with no WC GC base pairs and thereby confirms importance of 5′-GC ends in folding of *GCn* and *GCnCG* G-quadruplexes. Furthermore, 5′-5′ interlocking was observed for G-quadruplexes formed by *d*(GC$G_2TG_4TG_2$) and *d*(GC$G_2TG_4TG_2$CG) se-

quences, which are similar to *GCn* and *GCnCG* but with thymines separating G-tracts instead of adenines (40). It seems that adenine residues can affect ability for multimer formation. Both, *GCn* and *GCnCG* possess G_2A segments with tendency to form GNA loop (52,71,72). Surprisingly, A5 for thymine substitution still leads to formation of G3-G4-T5 lateral loop and therefore excludes importance of GNA loop in folding of *GCn* and *GCnCG* G-quadruplexes. Second adenines (A10) in front of the shortest G-tracts within oligonucleotides *GCn* and *GCnCG* form A(GGGG)A hexads, which is in perfect agreement with observation that adenine before the shortest G-tract will participate in a pentad or in our case hexad formation (45). The influence of adenine on multimerization ability was observed before (31,39,41,50). For short G-rich oligonucleotides *d*(GCG_2XG_2CG) and *d*(GCG_2XG_2), where X = A, T or TC, it was demonstrated that sequences with X = A most readily form long and organized structures in the presence of Na^+ and Na^+/Mg^{2+} ions on mica substrate (50). For d(GCG_2AG_2CG) the formation of G-wire was also observed in solution (43). Oligonucleotides d(GCG_2AG_2CG) and d(GCG_2AG_2) have similar nucleotide sequence as *GCn* and *GCnCG*, but lack one AG_4 segment. Our NMR study of *GCn* and *GCnCG* clearly shows that only dimeric G-quadruplexes were present in Na^+ ion containing solution. A5 in *GCn* and *GCnCG* are part of lateral loops and A10 residues participate in hexad hydrogen bond alignments. Therefore, the role of adenines in formation of longer structures is not straightforward. It should be noted that surface organization of G-quadruplexes could differ from their behavior in solution, due to molecular crowding effect reached during evaporation process on surface (50). Furthermore, in tetrameric G-quadruplexes formed by *d*(T_2AG_{3-5}), bulky adenine bases prevent formation of G-quadruplexes with a slipped strands, which are essential for assembly of the interlocking G-quadruplex dimer as observed for d(G_3T) (39,44). When varying Gs in the central quartet of unimolecular, parallel-stranded G-quadruplex formed by *d*($G_3TG_3A_2G_3TG_3A$), adenine was demonstrated to induce dimer as well as multimer formation, while variations with cytosines or thymines led to monomeric or dimeric structures, respectively (41). On the other hand, by varying nucleotide sequence in loops of monomeric G-quadruplexes composed of four G_2 or four G_3 tracts, adenine and cytosine residues display different multimerization mechanisms compared to thymines and mutations that mimic abasic sites, which shifted equilibrium towards higher-order structures (31).

Contradictory data were reported for the role that 3′-GC ends play in elongation of G-quadruplex based nanostructures. For example, d(GCG_2AG_2CG) is able to multimerize into G-wires on surface as well as in solution, where the mechanism of interlocking includes 3′-GC ends (43). For oligonucleotides d(GCG_2XG_2CG), where X = A, T or TC, it was shown that G-wires growth on surface through connectivities formed by GC ends was equally effective as growth through stacking of terminal G-quartets in sequences where 3′-GC was removed (50). On the other hand, for *d*(GC$G_2TG_4TG_2$CG) and *d*($G_2TG_4TG_2$CG) no interactions through 3′-GC ends were observed in solution (40). In the case, if free 3′-GC ends of *GCnCG* would enable inter-quadruplex connections, two major groove GCGC-quartets would link two *GCnCG* G-quadruplexes. Major groove GCGC-quartets are characterized by alternating narrow and wide grooves. The 3′-GC ends are part of the 3′-unit with parallel topology of *GCnCG* G-quadruplex, which exhibits medium grooves with average dimensions of 11.4 ± 0.4 Å. It may be assumed that different groove dimensions of 3′-unit with parallel topology and GCGC-quartet might prevent interconnections through 3′-GC ends.

The role of cations on G-quadruplex stabilization is well known as well as their effect on folding topologies (1,2,73). Structural rearrangement of G-quadruplexes can be triggered by change in nature of cations present in solution (74,75). For example, cation-dependent polymorphism of human telomeric sequence is likely one of the most well described (76). On the other hand, change of cations exhibits only minor effect on structure of G-quadruplexes formed by telomeric sequences derived from *Oxytricha nova* *d*($G_3T_4G_4$) and *d*($G_4T_4G_4$), which retain the same general fold in the presence of Na^+, $^{15}NH_4^+$ and K^+ ions (77,78). For *GCn* and *GCnCG* folded in the presence of $^{15}NH_4^+$ and K^+ ions, structural rearrangements of the 5′-units with antiparallel topologies were observed compared to the G-quadruplexes folded in the presence of Na^+ ions. Two G6-C2 WC base pairs are formed in the presence of $^{15}NH_4^+$ and K^+ ions instead of major groove G6–C2–G6–C2 quartet observed in the presence of Na^+ ions. Symmetric G4–G4 N1-carbonyl base pair are observed only in G-quadruplexes of *GCn* and *GCnCG* in the presence of Na^+ ions. Such structural transition of GCGC-quartet induced by change of Na^+ with $^{15}NH_4^+$ and K^+ ions was previously observed for *d*($G_3CT_4G_3C$) G-quadruplex (9,53). In the 3′-unit with parallel topology of *GCn* and *GCnCG* G-quadruplexes, G9–G12–G9–G12 quartet and A10–(G8–G11–G8–G11)–A10 hexad including 3′-GC ends in the case of *GCnCG* were preserved even in the presence of $^{15}NH_4^+$ and K^+ ions. Interestingly, in *GCn* G-quadruplexes $^{15}NH_4^+$ and K^+ ions promoted dimerization of two dimeric G-quadruplexes via 3′-3′ stacking. Dimerization of G-quadruplexes triggered by different cations was observed before. For example stacking through common U-quartet of *d*(UG_4T) tetramolecular G-quadruplex was induced by change of Na^+ ions to $^{15}NH_4^+$ and K^+ ions (38). Among monovalent cations, K^+ ions are known to most likely enhance multimerization of 'blunt-end' G-quadruplexes through stacking interactions as was demonstrated by ESI-MS (42).

In summary, we have shown that *GCn* and *GCnCG* form a novel six-layered dimeric G-quadruplex structures in solution of Na^+ ions. In the presence of larger cations, such as $^{15}NH_4^+$ and K^+, dimerization via 3′–3′ stacking interactions of terminal G-quartets was observed, which is precluded with 3′-GC ends. Our results might be useful for design of DNA oligonucleotides with ability to form G-quadruplexes with unique folding topologies as required for a specific nanotechnological application. In the designing process of G-rich oligonucleotides, introduction of adenine residues at critical position within DNA sequence might lead to additional hydrogen bonding or to formation of completely different structures. Furthermore, position of cytosines within the sequence is crucial due to their tendency to participate in formation of WC GC base pairs,

which may disrupt the anticipated structure. 5′-GC residues in *GCn* and *GCnCG* form G- and GCGC-quartets within G-quadruplexes and do not promote inter-quadruplex association. To the best of our knowledge, *GCn* and *GCnCG* G-quadruplexes exhibit unique combination of structural elements observed in solution for the first time.

DATA AVAILABILITY

Atomic coordinates and lists of chemical shifts for *GCn* and *GCnCG* have been deposited with 6SX6 and 6SYK accession numbers, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We gratefully acknowledge Professor Dr Jurij Lah at the University of Ljubljana, Faculty of Chemistry and Chemical Technology, for fruitful discussion regarding van't Hoff analysis.

FUNDING

Slovenian Research Agency [ARRS, grants P1-0242, P1-0192, J1-7108, J1-1704 and J7-9399]. Funding for open access charge: Slovenian Research Agency [P1-0242].
Conflict of interest statement. None declared.

REFERENCES

1. Bhattacharyya,D., Mirihana Arachchilage,G. and Basu,S. (2016) Metal cations in G-quadruplex folding and stability. *Front. Chem.*, **4**, 38.
2. Hud,N.V. and Plavec,J. (2006) In: Neidle,S and Balasubramanian,S (eds). *Quadruplex Nucleic Acids*. The Royal Society of Chemistry, Cambridge, UK, pp. 100–130.
3. Kettani,A., Kumar,A.R. and Patel,D.J. (1995) Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. *J. Mol. Biol.*, **254**, 638–656.
4. Kettani,A., Bouaziz,S., Gorin,A., Zhao,H., Jones,R.A. and Patel,D.J. (1998) Solution structure of a Na cation stabilized DNA quadruplex containing G·G·G·G and G·C·G·C tetrads formed by G-G-G-C repeats observed in adeno-associated viral DNA. *J. Mol. Biol.*, **282**, 619–636.
5. Zhang,N., Gorin,A., Majumdar,A., Kettani,A., Chernichenko,N., Skripkin,E. and Patel,D.J. (2001) Dimeric DNA quadruplex containing major groove-aligned A·T·A·T and G·C·G·C tetrads stabilized by inter-subunit Watson-Crick A·T and G·C pairs. *J. Mol. Biol.*, **312**, 1073–1088.
6. Webba da Silva,M. (2003) Association of DNA quadruplexes through G:C:G:C tetrads. Solution structure of d(GCGGTGGAT). *Biochemistry*, **42**, 14356–14365.
7. Escaja,N., Gómez-Pinto,I., Pedroso,E. and González,C. (2007) Four-stranded DNA structures can be stabilized by two different types of minor groove G:C:G:C tetrads. *J. Am. Chem. Soc.*, **129**, 2004–2014.
8. Viladoms,J., Escaja,N., Frieden,M., Gómez-Pinto,I., Pedroso,E. and González,C. (2009) Self-association of short DNA loops through minor groove C:G:G:C tetrads. *Nucleic Acids Res.*, **37**, 3264–3275.
9. Bouaziz,S., Kettani,A. and Patel,D.J. (1998) A K cation-induced conformational switch within a loop spanning segment of a DNA quadruplex containing G-G-G-C repeats. *J. Mol. Biol.*, **282**, 637–652.
10. 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.
11. Kocman,V. and Plavec,J. (2017) Tetrahelical structural family adopted by AGCGA-rich regulatory DNA regions. *Nat. Commun.*, **8**: 15355.
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. Zhang,N., Gorin,A., Majumdar,A., Kettani,A., Chernichenko,N., Skripkin,E. and Patel,D.J. (2001) V-shaped scaffold: a new architectural motif identified in an A·(G·G·G·G) pentad-containing dimeric DNA quadruplex involving stacked G(anti)·G(anti)·G(anti)·G(syn) tetrads. *J. Mol. Biol.*, **311**, 1063–1079.
14. 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.
15. Majumdar,A., Kettani,A., Skripkin,E. and Patel,D.J. (2001) Pulse sequences for detection of $NH_2\cdots N$ hydrogen bonds in sheared G·A mismatches via remote, non-exchangeable protons. *J. Biomol. NMR*, **19**, 103–113.
16. Mergny,J.-L., De Cian,A., Amrane,S. and Da Silva,M.W. (2006) Kinetics of double-chain reversals bridging contiguous quartets in tetramolecular quadruplexes. *Nucleic Acids Res.*, **34**, 2386–2397.
17. 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.
18. Matsugami,A., Okuizumi,T., Uesugi,S. and Katahira,M. (2003) Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. *J. Biol. Chem.*, **278**, 28147–28153.
19. Hänsel-Hertsch,R., Di Antonio,M. and Balasubramanian,S. (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.*, **18**, 279–284.
20. Rhodes,D. and Lipps,H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627–8637.
21. Simone,R., Fratta,P., Neidle,S., Parkinson,G.N. and Isaacs,A.M. (2015) G-quadruplexes: emerging roles in neurodegenerative diseases and the non-coding transcriptome. *FEBS Lett.*, **589**, 1653–1668.
22. Livshits,G.I., Stern,A., Rotem,D., Borovok,N., Eidelshtein,G., Migliore,A., Penzo,E., Wind,S.J., Di Felice,R. and Skourtis,S.S. (2014) Long-range charge transport in single G-quadruplex DNA molecules. *Nat. Nanotechnol.*, **9**, 1040–1046.
23. Sha,R., Xiang,L., Liu,C., Balaeff,A., Zhang,Y., Zhang,P., Li,Y., Beratan,D.N., Tao,N. and Seeman,N.C. (2018) Charge splitters and charge transport junctions based on guanine quadruplexes. *Nat. Nanotechnol.*, **13**, 316–321.
24. Liu,S.P., Weisbrod,S.H., Tang,Z., Marx,A., Scheer,E. and Erbe,A. (2010) Direct measurement of electrical transport through G-quadruplex DNA with mechanically controllable break junction electrodes. *Angew. Chem., Int. Ed.*, **49**, 3313–3316.
25. Calzolari,A., Di Felice,R. and Molinari,E. (2004) Electronic properties of guanine-based nanowires. *Solid State Commun.*, **131**, 557–564.
26. Gao,Z.F., Huang,Y.L., Ren,W., Luo,H.Q. and Li,N.B. (2016) Guanine nanowire based amplification strategy: enzyme-free biosensing of nucleic acids and proteins. *Biosens. Bioelectron.*, **78**, 351–357.
27. Miyoshi,D., Karimata,H., Wang,Z.-M., Koumoto,K. and Sugimoto,N. (2007) Artificial G-wire switch with 2, 2′-bipyridine units responsive to divalent metal ions. *J. Am. Chem. Soc.*, **129**, 5919–5925.
28. Usui,K., Okada,A., Sakashita,S., Shimooka,M., Tsuruoka,T., Nakano,S.-i., Miyoshi,D., Mashima,T., Katahira,M. and Hamada,Y. (2017) DNA G-wire formation using an artificial peptide is controlled by protease activity. *Molecules*, **22**, 1991.
29. Marsh,T.C. and Henderson,E. (1994) G-wires: self-assembly of a telomeric oligonucleotide, d(GGGGTTGGGG), into large superstructures. *Biochemistry*, **33**, 10718–10724.
30. Mergny,J.-L. and Sen,D. (2019) DNA quadruple helices in nanotechnology. *Chem. Rev.*, **119**, 6290–6325.
31. Varizhuk,A.M., Protopopova,A.D., Tsvetkov,V.B., Barinov,N.A., Podgorsky,V.V., Tankevich,M.V., Vlasenok,M.A., Severov,V.V., Smirnov,I.P. and Dubrovin,E.V. (2018) Polymorphism of G4

associates: from stacks to wires via interlocks. *Nucleic Acids Res.*, **46**, 8978–8992.
32. 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.
33. D'Atri,V., Borbone,N., Amato,J., Gabelica,V., D'Errico,S., Piccialli,G., Mayol,L. and Oliviero,G. (2014) DNA-based nanostructures: The effect of the base sequence on octamer formation from d(XGGYGGT) tetramolecular G-quadruplexes. *Biochimie*, **99**, 119–128.
34. Mukundan,V.T., Do,N.Q. and Phan,A.T.n. (2011) HIV-1 integrase inhibitor T30177 forms a stacked dimeric G-quadruplex structure containing bulges. *Nucleic Acids Res.*, **39**, 8984–8991.
35. Do,N.Q., Lim,K.W., Teo,M.H., Heddi,B. and Phan,A.T. (2011) Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res.*, **39**, 9448–9457.
36. Do,N.Q. and Phan,A.T. (2012) Monomer-dimer equilibrium for the 5′-5′ stacking of propeller-type parallel-stranded G-quadruplexes: NMR structural study. *Chem.–Eur. J.*, **18**, 14752–14759.
37. Kuryavyi,V., Cahoon,L.A., Seifert,H.S. and Patel,D.J. (2012) RecA-binding *pilE* G4 sequence essential for pilin antigenic variation forms monomeric and 5′ end-stacked dimeric parallel G-quadruplexes. *Structure*, **20**, 2090–2102.
38. Šket,P. and Plavec,J. (2010) Tetramolecular DNA quadruplexes in solution: insights into structural diversity and cation movement. *J. Am. Chem. Soc.*, **132**, 12724–12732.
39. Kato,Y., Ohyama,T., Mita,H. and Yamamoto,Y. (2005) Dynamics and thermodynamics of dimerization of parallel G-quadruplexed DNA formed from d(TTAG$_n$) (n = 3− 5). *J. Am. Chem. Soc.*, **127**, 9980–9981.
40. 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.
41. Kolesnikova,S., Hubálek,M., Bednárová,L., Cvačka,J. and Curtis,E.A. (2017) Multimerization rules for G-quadruplexes. *Nucleic Acids Res.*, **45**, 8684–8696.
42. Smargiasso,N., Rosu,F., Hsia,W., Colson,P., Baker,E.S., Bowers,M.T., De Pauw,E. and Gabelica,V. (2008) G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *J. Am. Chem. Soc.*, **130**, 10208–10216.
43. Ma'ani Hessari,N., Spindler,L., Troha,T., Lam,W.-C., Drevenšek-Olenik,I. and Webba da Silva,M. (2014) Programmed self-assembly of a quadruplex DNA nanowire. *Chem.–Eur. J.*, **20**, 3626–3630.
44. Krishnan-Ghosh,Y., Liu,D. and Balasubramanian,S. (2004) Formation of an interlocked quadruplex dimer by d(GGGT). *J. Am. Chem. Soc.*, **126**, 11009–11016.
45. Phan,A.T. and Do,N.Q. (2012) Engineering of interlocked DNA G-quadruplexes as a robust scaffold. *Nucleic Acids Res.*, **41**, 2683–2688.
46. Webba da Silva,M. (2005) Experimental demonstration of T:(G:G:G:G):T hexad and T:A:A:T tetrad alignments within a DNA quadruplex stem. *Biochemistry*, **44**, 3754–3764.
47. Kankia,B. (2014) Tetrahelical monomolecular architecture of DNA: a new building block for nanotechnology. *J. Phys. Chem. B*, **118**, 6134–6140.
48. Oliviero,G., D'Errico,S., Pinto,B., Nici,F., Dardano,P., Rea,I., De Stefano,L., Mayol,L., Piccialli,G. and Borbone,N. (2017) Self-assembly of G-rich oligonucleotides incorporating a 3′-3′ inversion of polarity site: a new route towards G-wire DNA nanostructures. *ChemistryOpen*, **6**, 599–605.
49. Yatsunyk,L.A., Piétrement,O., Albrecht,D., Tran,P.L.T., Renciuk,D., Sugiyama,H., Arbona,J.-M., Aimé,J.-P. and Mergny,J.-L. (2013) Guided assembly of tetramolecular G-quadruplexes. *ACS nano*, **7**, 5701–5710.
50. Troha,T., Drevensek-Olenik,I., Webba da Silva,M. and Spindler,L. (2016) Surface-adsorbed long G-quadruplex nanowires formed by G:C linkages. *Langmuir*, **32**, 7056–7063.
51. Spindler,L., Rigler,M., Drevenšek-Olenik,I., Ma'ani Hessari,N. and Webba da Silva,M. (2010) Effect of base sequence on G-wire formation in solution. *J. Nucleic Acids*, **2010**, doi:10.4061/2010/431651.
52. Marušič,M. and Plavec,J. (2019) Towards understanding of polymorphism of the G-rich region of human Papillomavirus Type 52. *Molecules*, **24**, 1294.
53. Zavasnik,J., Podbevsek,P. and Plavec,J. (2011) Observation of water molecules within the bimolecular d($G_3CT_4G_3C$)$_2$ G-quadruplex. *Biochemistry*, **50**, 4155–4161.
54. Salomon-Ferrer,R., Goetz,A., Poole,D., Le Grand,S. and Walker,R. (2012) Routine microsecond molecular dynamics simuations with AMBER-part II: particle mesh Ewald. *J. Chem. Theory Comput.*, **9**, 3878–3888.
55. Case,D.A., Berryman,J.T., Betz,R.M., Cerutti,D.S., Cheatham,T.E., Darden,T.A., Duke,R.E., Giese,T.J., Gohlke,H. and Goetz,A.W. (2015) AMBER 2015. *UCSF*.
56. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
57. Pérez,A., Marchán,I., Svozil,D., Sponer,J., Cheatham,T.E., Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.*, **92**, 3817–3829.
58. Krepl,M., Zgarbová,M., Stadlbauer,P., Otyepka,M., Banáš,P., Koca,J., Cheatham,T.E. III, Jurecka,P. and Šponer,J.I. (2012) Reference simulations of noncanonical nucleic acids with different χ variants of the AMBER force field: quadruplex DNA, quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.*, **8**, 2506–2520.
59. Zgarbová,M., Luque,F.J., Šponer,J.i., Cheatham,T.E. III, Otyepka,M. and Jurecka,P. (2013) Toward improved description of DNA backbone: revisiting epsilon and zeta torsion force field parameters. *J. Chem. Theory Comput.*, **9**, 2339–2354.
60. Zgarbová,M., šponer,J.í., Otyepka,M., Cheatham,T.E. III, Galindo-Murillo,R. and Jurecka,P. (2015) Refinement of the sugar–phosphate backbone torsion beta for AMBER force fields improves the description of Z-and B-DNA. *J. Chem. Theory Comput.*, **11**, 5723–5736.
61. Onufriev,A., Bashford,D. and Case,D.A. (2000) Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B*, **104**, 3712–3720.
62. Onufriev,A., Bashford,D. and Case,D.A. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins: Struct., Funct., Bioinf.*, **55**, 383–394.
63. Ryckaert,J.-P., Ciccotti,G. and Berendsen,H.J. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.*, **23**, 327–341.
64. Neidle,S. (2010) *Principles of Nucleic Acid Structure*. Elsevier Science, Oxford, UK.
65. Wei,D., Todd,A.K., Zloh,M., Gunaratnam,M., Parkinson,G.N. and Neidle,S. (2013) Crystal structure of a promoter sequence in the B-raf gene reveals an intertwined dimer quadruplex. *J. Am. Chem. Soc.*, **135**, 19319–19329.
66. Trajkovski,M., Webba da Silva,M. and Plavec,J. (2012) Unique structural features of interconverting monomeric and dimeric G-quadruplexes adopted by a sequence from the intron of the N-myc gene. *J. Am. Chem. Soc.*, **134**, 4132–4141.
67. Kuryavyi,V., Phan,A.T. and Patel,D.J. (2010) Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter. *Nucleic Acids Res.*, **38**, 6757–6773.
68. Marušič,M. and Plavec,J. (2015) The effect of DNA sequence directionality on G-quadruplex folding. *Angew. Chem., Int. Ed.*, **54**, 11716–11719.
69. Kuryavyi,V. and Patel,D.J. (2010) Solution structure of a unique G-quadruplex scaffold adopted by a guanosine-rich human intronic sequence. *Structure*, **18**, 73–82.
70. Črnugelj,M., Šket,P. and Plavec,J. (2003) Small change in a G-rich sequence, a dramatic change in topology: new dimeric G-quadruplex folding motif with unique loop orientations. *J. Am. Chem. Soc.*, **125**, 7866–7871.
71. Yoshizawa,S., Kawai,G., Watanabe,K., Miura,K.-I. and Hirao,I. (1997) GNA trinucleotide loop sequences producing extraordinarily stable DNA minihairpins. *Biochemistry*, **36**, 4761–4767.

72. Balkwill,G.D., Williams,H.E. and Searle,M.S. (2007) Structure and folding dynamics of a DNA hairpin with a stabilising d(GNA) trinucleotide loop: influence of base pair mis-matches and point mutations on conformational equilibria. *Org. Biomol. Chem.*, **5**, 832–839.
73. Largy,E., Mergny,J.-L. and Gabelica,V. (2016), *The Alkali Metal Ions: Their Role for Life*. Springer, pp. 203–258.
74. Bouaziz,S., Kettani,A. and Patel,D.J. (1998) A K cation-induced conformational switch within a loop spanning segment of a DNA quadruplex containing GGGC repeats. *J. Mol. Biol.*, **282**, 637–652.
75. De Rache,A., Kejnovská,I., Vorlíčková,M. and Buess-Herman,C. (2012) Elongated thrombin binding aptamer: a G-quadruplex cation-sensitive conformational switch. *Chem.–Eur. J.*, **18**, 4392–4400.
76. Šket,P. and Plavec,J. (2015) In: Monchaud,D (ed). *Biological Relevance & Therapeutic Applications of DNA- & RNA-Quadruplexes*. Future Medicine Ltd, London, pp. 22–36.
77. Šket,P., Črnugelj,M. and Plavec,J. (2004) d($G_3T_4G_4$) forms unusual dimeric G-quadruplex structure with the same general fold in the presence of K^+, Na^+ or NH_4^+ ions. *Bioorg. Med. Chem.*, **12**, 5735–5744.
78. Schultze,P., Hud,N.V., Smith,F.W. and Feigon,J. (1999) The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the Oxytricha nova telomere repeat oligonucleotide d($G_4T_4G_4$). *Nucleic Acids Res.*, **27**, 3018–3028.