Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 12 (2004) 5735–5744
Bioorganic & Medicinal Chemistry

$d(G_3T_4G_4)$ forms unusual dimeric G-quadruplex structure with the same general fold in the presence of K^+, Na^+ or NH_4^+ ions

Primož Šket, Martin Črnugelj and Janez Plavec*

NMR Center, National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia

Received 25 June 2004; accepted 3 August 2004
Available online 9 September 2004

Abstract—We have recently communicated that DNA oligonucleotide $d(G_3T_4G_4)$ forms a dimeric G-quadruplex in the presence of K^+ ions [*J. Am. Chem. Soc.* **2003**, *125*, 7866–7871]. The high-resolution NMR structure of $d(G_3T_4G_4)_2$ G-quadruplex exhibits G-quadruplex core consisting of three stacked G-quartets. The two overhanging G3 and G11 residues are located at the opposite sides of the end G-quartets and are not involved in G-quartet formation. $d(G_3T_4G_4)_2$ G-quadruplex represents the first bimolecular G-quadruplex where end G-quartets are spanned by diagonal (T4–T7) as well as edge-type loops (T15–T18). Three of the G-rich strands are parallel while one is anti-parallel. The G12–G22 strand demonstrates a sharp reversal in strand direction between residues G19 and G20 that is accommodated with the leap over the middle G-quartet. The reversal in strand direction is achieved without any extra intervening residues. Here we furthermore examined the influence of different monovalent cations on the folding of $d(G_3T_4G_4)$. The resolved imino and aromatic proton resonances as well as (sequential) NOE connectivity patterns showed only minor differences in key intra- and interquartet NOE intensities in the presence of K^+, Na^+ and NH_4^+ ions, which were consistent with subtle structural differences while retaining the same folding topology of $d(G_3T_4G_4)_2$ G-quadruplex.

1. Introduction

It is well established that G-rich sequences can adopt multistranded DNA secondary structures called G-quadruplexes that are build of stacked G-quartets.[2,3] G-quadruplex DNA structures are a subject of great interest since their formation has been suggested to play an important role in variety of important biological processes as well as due to their potential therapeutic applications.[4–8] Recent studies have provided evidence of the G-quadruplex formation in vivo.[9,10] One of the most abundant sites with DNA sequences capable of forming G-quadruplex structures are telomeres, protein-DNA complexes located at the ends of chromosomes.[11–13] Telomeric DNA consists of multiple copies of repeating units of G-rich DNA (e.g., the repeating sequence is TTAGGG in mammals and TTTTGGGG in the protozoa *Oxytricha nova*), which are typically hundreds to thousands of residues long. Almost all of the telomeric DNA is double helical with a single stranded G-rich overhang protruding at the extreme 3′ terminus and having the propensity to form G-quadruplex structures.[14,15] In majority of normal cells telomeric DNA shortens upon each cell division. The critical shortening of telomeric DNA is thought to trigger the onset of cellular senescence in most normal cells. Cancer cells, however, exhibit an unlimited growth due to a mechanism that maintains telomere length, usually by the action of the enzyme telomerase.[16,17] Telomerase's optimal substrate is a nonfolded single stranded DNA, whereas folded G-quadruplex structures inhibit telomerase. The possibility to control cell ageing and senescence by selective inhibition through structural changes has made G-quadruplexes interesting targets for anti-cancer drug design.[18–22,7,23,8,24–27]

The folding topologies of G-quadruplex structures critically depend on the number of G-rich repeats, sequence details and the nature as well as concentration of metal ions present in solution.[28–30] For example, $d(G_4T_4G_4)$ (Fig. 1c) forms a symmetric G-quadruplex containing four stacked G-quartets with two T_4 segments looped diagonally across the two outer faces of stacked G-quartets.[31–34] On the other hand, $d(G_4T_4G_3)$ (Fig. 1b), a sequence with 3′ terminal dG residue missing from $d(G_4T_4G_4)$, forms G-quadruplex that consists of three G-quartets and two additional dG residues, which are not involved in G-quartet formation.[35] Here we describe

Keywords: G-quadruplex; NMR; Monovalent cations.
* Corresponding author. Tel.: +386 147 60 353; fax: +386 147 60 300; e-mail: janez.plavec@ki.si

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.08.009

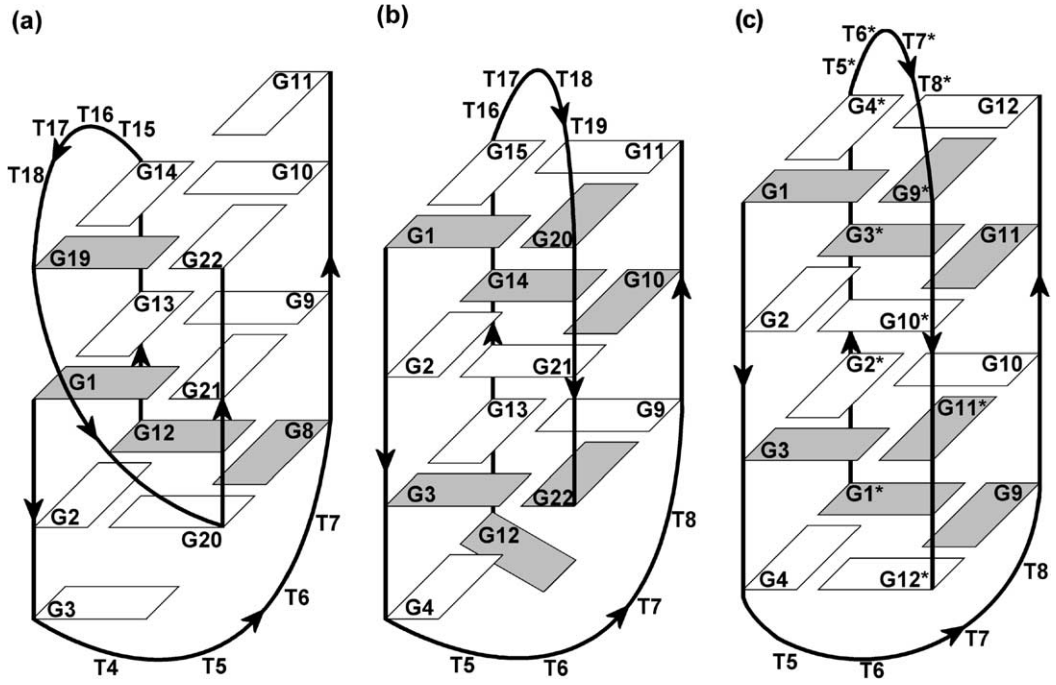

Figure 1. Folding topology of $d(G_3T_4G_4)_2$ (a), $d(G_4T_4G_3)_2$ (b) and $d(G_4T_4G_4)_2$ (c) G-quadruplex. Strand directions are indicated by arrows. The guanine bases are shown as rectangles, where shaded rectangles represent *syn* nucleobases. Thymine residues are omitted for clarity.

our studies on the folding of $d(G_3T_4G_4)$, a sequence with 5′ terminal dG residue missing from a well-studied $d(G_4T_4G_4)$ sequence. $d(G_3T_4G_4)$ forms G-quadruplex with unprecedented topology consisting of three G-quartet planes, overhanging G3 and G11 residues and diagonal as well as an edge-type loops (Fig. 1a).[1] In this work we extend our recent study[1] and describe the high-resolution NMR structure of the $d(G_3T_4G_4)_2$ G-quadruplex in detail as well as scrutinize structural changes in the presence of K^+, Na^+ and NH_4^+ ions.

The presence and the nature of metal ions play a determining role in the folding and stability of G-quadruplex structures.[36–42,35,13,43,44] For example, recent X-ray crystallographic study has shown that DNA oligonucleotide consisting of four repeats of human telomere sequence, $d(AG_3(TTAG_3)_3)$, forms in the presence of K^+ ions a G-quadruplex structure with all parallel strands and double-chain reversal loops.[13] In contrast the same sequence has been shown by NMR spectroscopy to fold into fundamentally different G-quadruplex topology with anti-parallel strands in the presence of Na^+ ions.[45] On the other hand, $d(G_4T_4G_4)$ and $d(G_3T_4G_3)$ sequences form analogous diagonally looped G-quadruplex structures, which exhibit only minor structural differences in the presence of Na^+, K^+ and NH_4^+ ions.[36,37] The interaction of monovalent cations with G-quadruplexes is linked to their ionic radii and dehydration energies. The smaller Na^+ ion can fit in the centre of G-quartet plane whereas the larger K^+ and NH_4^+ ions reside between the two G-quartet planes.[46] The size of ions, their charge densities and coordination properties have been also shown to have a dramatic effect on ion exchange rates.[47] Our recent study[35] showed that $d(G_4T_4G_3)$ sequence adopts a single G-quadruplex structure only in the presence of Na^+ ions, while multiple structures form in the presence of K^+ or NH_4^+ ions.[35] In the present study we examined the influence of three monovalent cations with different ionic radii and coordination geometries on the folding of $d(G_3T_4G_4)$. Na^+, K^+ or NH_4^+ ions take up and preferentially reside at different binding sites within the architecture of G-quadruplex, which is manifested, in distinctive chemical (de)shielding of the nearby imino and aromatic resonances while the topology of $d(G_3T_4G_4)_2$ structure is equivalent in the presence of all three counterions.

2. Results and discussion

2.1. NMR study of $d(G_3T_4G_4)_2$ G-quadruplex in the presence of K^+ ions

The formation of the $d(G_3T_4G_4)_2$ G-quadruplex was initially followed by monitoring ^1H NMR resonances as a function of increasing potassium ion concentration. Twelve narrow and well-resolved imino resonances were observed in the range from 11.18 to 11.96 ppm, which is characteristic of guanine residues involved in Hoogsteen base pairing and substantiated the formation of three G-quartets (Fig. 2a).

The formation of a single asymmetric bimolecular G-quadruplex in the presence of 15 mM KCl was proposed on the basis of 12 imino and 22 aromatic resonances (Fig. 2a). Well-resolved resonances facilitated assignment and subsequent NMR restrained structural characterization. The NMR resonance assignments of G-quadruplex structure adopted by $d(G_3T_4G_4)$ oligonucleotide were performed using a combination of NOESY, COSY and TOCSY NMR experiments. Aromatic–anomeric region of 2D NOESY spectra of

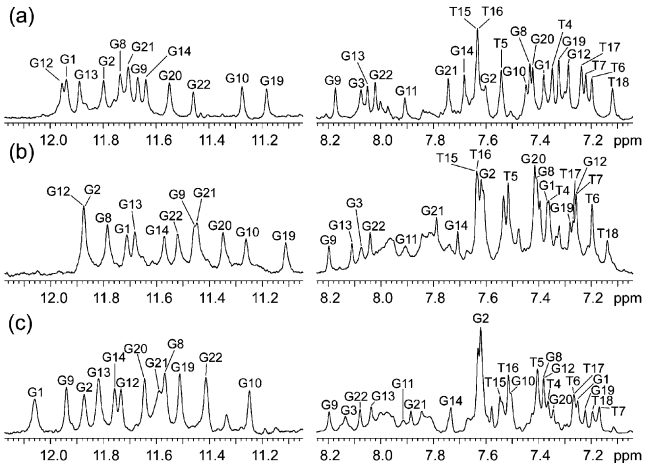

Figure 2. Imino and aromatic regions of the ^{1}H NMR spectra of $d(G_3T_4G_4)_2$ at 25 °C in 10% $^{2}H_2O$ in the presence of 15 mM KCl (a), 10 mM NH_4Cl (b) and 15 mM NaCl (c). The sample concentration was 4.0 mM in strand for (a), 3.8 mM for (b) and 3.6 mM for (c).

$d(G_3T_4G_4)_2$ revealed four strong intraresidual H8–H1′ cross-peaks that clearly identified four guanine bases (G1, G8, G12 and G19) in *syn* conformation and offered initial evidence for the formation of a fold-back structure.[11,48–50,35] An attempt has been made to perform a sequential assignment of the two strands of $d(G_3T_4G_4)_2$ G-quadruplex. The interruption of 5′–3′ sequential NOE connectivities at 5′*anti*–3′*syn* steps[50] and a small number of sequential connectivities amongst thymine residues in the loops prevented spectral assignment on the basis of NMR spectra of the original sequence alone. We then prepared six inosine and four uridine analogues with the single residue substitutions in the original sequence. NMR spectra of the uridine derivatives were essentially identical to those observed for the original sequence, which indicated the formation of the structure with the same topology. The uridine substitution was unambiguously identified through strong NOE correlation of H5 and H6 protons. In addition, uridine for thymine substitution resulted in disappearance of methyl resonance. On the other hand, replacement of inosine for guanine resulted in poorly resolved spectra in comparison to the original sequence. Nevertheless the same characteristic NOE connectivities confirmed the formation of G-quadruplex structure with the same topology. Large downfield shift of the inosine imino proton and a smaller shift of H8 proton assisted their assignment to the guanine base that has been substituted. Unique NOE contacts between H2 of inosine and H8 protons of the neighbouring guanine bases were used to establish the bases within individual G-quartets. While the above systematic substitutions enabled us to assign all thymine residues there were still some ambiguities in the assignment of guanines due to the chemical shift changes observed in inosine analogues. In order to complete the assignment we resorted to the residue specific ^{15}N labelled (15%) samples of $d(G_3T_4G_4)_2$. Partial ^{15}N isotopic labelling enabled us to unequivocally assign specific imino resonance to a particular guanine in a sequence with the use of one-bond ^{15}N–^{1}H correlation established by HSQC experiment. Furthermore, partially ^{15}N labelled guanine helped us to establish correlations via two bonds in ^{15}N–^{1}H HMBC experiments, which gave the residue specific assignment of H8 resonances.[51] Using this approach we were finally able to identify the following 5′–3′ sequential connectivities: G1*syn*–G2*anti*–G3*anti*–T4*anti*, T5*anti*–T6*anti*–T7*anti*, G8*syn*–G9*anti*, G10*anti*–G11*anti*, G12*syn*–G13*anti*–G14*anti* and G21*anti*–G22*anti*.

NOE cross-peaks between imino and H8 protons of the neighbouring guanine residues shown in Figure 3 were consistent with the following G-quartets: G19→G22→G10→G14, G1→G21→G9→G13 and G12→G8→G20→G2 (arrows indicate sequential

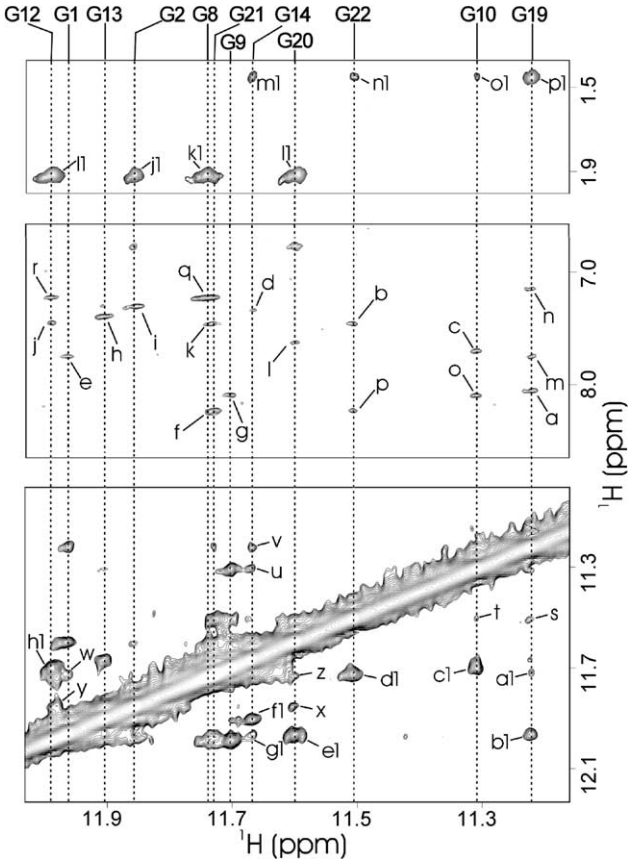

Figure 3. Details of the NOESY spectrum (τ_m = 300 ms) of $d(G_3T_4G_4)_2$ in 10% $^{2}H_2O$ at 5 °C and 15 mM KCl. NOE cross-peaks originating from imino protons of dG residues are labelled. The cross-peaks are assigned as follows: a, G19(NH)–G22(H8); b, G22(NH)–G10(H8); c, G10(NH)–G14(H8); d, G14(NH)–G19(H8); e, G1(NH)–G21(H8); f, G21(NH)–G9(H8); g, G9(NH)–G13(H8); h, G13(NH)–G1(H8); i, G2(NH)–G12(H8); j, G12(NH)–G8(H8); k, G8(NH)–G20(H8); l, G20(NH)–G2(H8); m, G19(NH)–G21(H8); n, G19(NH)–T18(H6); o, G10(NH)–G13(H8); p, G22(NH)–G9(H8); q, G8(NH)–T7(H6); r, G12(NH)–T7(H6); s, G19 (NH)–G22(NH); t, G10(NH)–G22(NH); u, G14(NH)–G10(NH); v, G14(NH)–G19(NH); w, G1(NH)–G21(NH); x, G20(NH)–G2(NH); y, G12(NH)–G2(NH); z, G20(NH)–G8(NH); a1, G19(NH)–G21(NH); b1, G19(NH)–G1(NH); c1, G10(NH)–G9(NH); d1, G22(NH)–G21(NH); e1, G20(NH)–G1(NH); f1, G14(NH)–G13(NH); g1, G14(NH)–G1(NH); h1, G12(NH)–G9(NH); i1, G12(NH)–T7(Me); j1, G2(NH)–T7(Me); k1, G8(NH)–T7(Me); l1, G20(NH)–T7(Me); m1, G14(NH)–T18(Me); n1, G22(NH)–T18(Me); o1, G10(NH)–T18(Me); p1, G19(NH)–T18(Me).

imino-H8 cross-peaks around each G-quartet, cross-peaks between the last and first residue of all three G-quartets are implied). The following intraquartet imino–imino NOE contacts (Fig. 3) were all consistent with the resonance assignment and topology model presented in Figure 1: G10–G14, G10–G22, G14–G19 and G19–G22 for the first G-quartet, G1–G21 and G9–G13 for the second G-quartet and G2–G12, G2–G20, G8–G20 for the third G-quartet. Interquartet imino–imino NOE cross-peaks between G1–G19, G1–G14, G1–G20, G9–G10, G9–G12, G13–G14, G19–G21 and G21–G22 supported the formation of three G-quartets and defined their relative positions within the structure (Fig. 3). Key NOE contacts of T18(Me)–G22(H8), G19(NH)–G22(H8), G19(NH)–G21(H8) and G19(NH)–G21(NH) (Fig. 3) were unexpected within the 3D fold of $d(G_3T_4G_4)_2$ G-quadruplex assuming topologies of related sequences and stipulated a search for a novel structural features within a dimeric G-quadruplex.

The solution structure of the $d(G_3T_4G_4)_2$ G-quadruplex in the presence of K^+ ions was calculated using 414 NOE restraints together with 66 sugar torsion angle restraints and 24 hydrogen-bond restraints (Table 1). Fifty simulated annealing calculations of 60 ps were performed with NMR restraints using a generalized Born implicit solvation model.[52–54] All 50 final structures showed similar structural features (root mean square deviation, rmsd below 1.3 Å) and comparable agreement with experimental NMR restraints (no violation of NOE distance restraints >0.5 Å). The eight lowest energy structures with small NMR restraints violation were selected and subsequently analyzed. The superposition of eight final NMR structures of $d(G_3T_4G_4)_2$ is shown in Figure 4a. The pairwise all atom rmsd between these eight structures was 0.76 ± 0.42 Å (Table 1).

Table 1. NMR restraints and structural statistics for $d(G_3T_4G_4)_2$

A. NMR restraints	
Distance restraints	
Nonexchangeable protons	335
Intraresidual	245
Sequential $(i, i+1)$	81
Long range $(i, i+j;\ j>1)$	9
Exchangeable protons	79
Sequential $(i, i+1)$	15
Long range $(i, i+j;\ j>1)$	64
Hydrogen-bond restraints	24
Sugar torsion angle restraints	66
B. Structure statistics	
NOE violations	
Number >0.3 Å	1.5 ± 0.5
Maximum violations (Å)	0.31 ± 0.01
Deviations from ideal covalent geometry	
Bond lengths (Å)	0.01 ± 0.00
Bond angles (deg)	2.95 ± 0.03
C. Pairwise all atom rmsd (Å) (eight structures from SA)	
All residues	0.76 ± 0.42
Without T4–T7 loop residues	0.50 ± 0.20
Without G3–T7 loop residues	0.45 ± 0.11
Without T15–T18 loop residues	0.76 ± 0.47

2.2. G-quadruplex core of $d(G_3T_4G_4)_2$ fold

The solution structure of $d(G_3T_4G_4)_2$ G-quadruplex consists of three G-quartets, which are stacked in a right-handed quadruple helix (Fig. 4). Sugar rings of all residues adopt predominantly South-type conformation with puckering between C2′-*endo* and C2′-*endo*/C3′-*exo* canonical forms. The succession of the glycosidic torsion angles of guanine residues is 5′-*syn–anti–anti*-(T_4 loop)-*syn–anti–anti–anti*-3′ along both strands of $d(G_3T_4G_4)_2$. The details of 3D structures of both strands are however different. The first three guanine residues of G1–G11 strand follow 5′-*syn–anti–anti* pattern. G3 residue, which is part of the diagonal loop, is not involved in G-quartet formation and is oriented almost perpendicularly with respect to the neighbouring G-quartet and exhibits substantial degree of conformational freedom (Table 1). G8–G10 residues, which are involved in G-quartets, also follow *syn–anti–anti* pattern, whereas G11 is stacked on the end G-quartet. On the other hand, guanine residues of G12–G22 strand are all involved in the formation of G-quartets. In summary, we note that five out of twelve guanine residues involved in the G-quartet formation originate from G1–G11 strand, whereas seven originate from the second strand (i.e., G12–G22). The two G-tracts in G1–G11 strand are anti-parallel. The G12–G22 strand exhibits a sharp reversal in the strand direction between residues G19 and G20 that places its G-tracts in parallel orientation. As a consequence, there are three parallel and one anti-parallel G-runs in the overall fold of $d(G_3T_4G_4)_2$ G-quadruplex. It is noteworthy that residue G20, which is involved in a reversal of strand direction, exhibits unusual conformation across the glycosidic bond, which is close to *high-anti* region ($\chi \approx -41° \pm 1°$). The alternation of the glycosidic torsion angles within the individual G-quartet is *syn–anti–anti–anti* for G19–G22–G10–G14 and G1–G21–G9–G13 G-quartets and *syn–syn–anti–anti* for G12–G8–G20–G2 G-quartet. The hydrogen bonds in both adjacent G-quartets with *syn–anti–anti–anti* glycosidic torsion angle orientations follow anti-clockwise donor–acceptor directionality. On the other hand, G12–G8–G20–G2 G-quartet exhibits clockwise orientation of donors and acceptors in its hydrogen bonds. The stacking patterns between G-quartets of G19–G22–G10–G14 and G1–G21–G9–G13 are shown in Figure 5a. Five-membered rings of G19–G22–G10–G14 G-quartet are stacked through the overlap of six-membered rings of G1–G21–G9–G13 G-quartet (Fig. 5a). There is only a partial overlap between five-membered rings of G1–G21–G9–G13 and G2–G20–G8–G12 G-quartets that is indicative of overwinding of this part of the G-quadruplex structure (Fig. 5b).

2.3. Sharp reversal in the direction of G12–G22 strand

Perusal of NMR structures presented in Figure 4 shows three parallel and one anti-parallel G-rich runs in the $d(G_3T_4G_4)_2$ G-quadruplex. The unprecedented orienta-

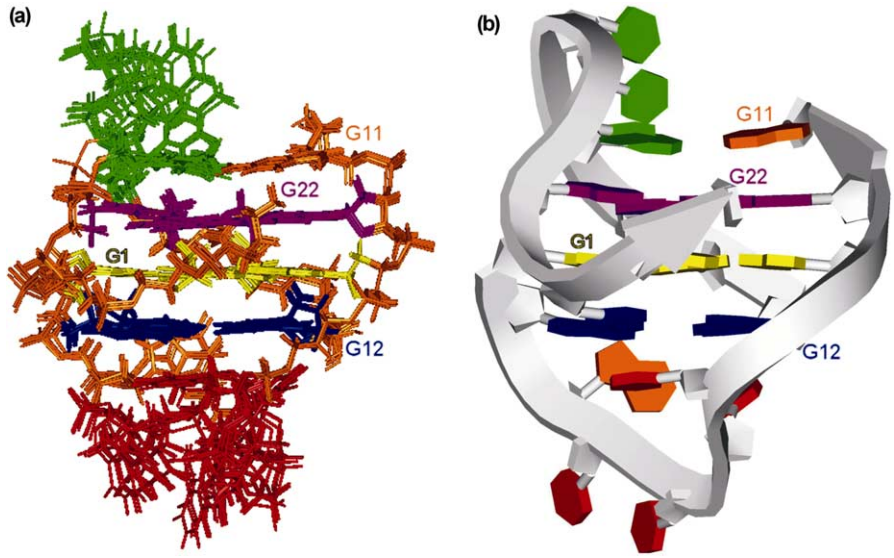

Figure 4. NMR solution structure of the K^+-form of $d(G_3T_4G_4)_2$ G-quadruplex. (a) Superposition of eight structures with the lowest energies and the lowest NMR restraints violations. (b) Ribbon presentation of the lowest energy structure of $d(G_3T_4G_4)_2$ G-quadruplex. End residues of each strand are labelled.

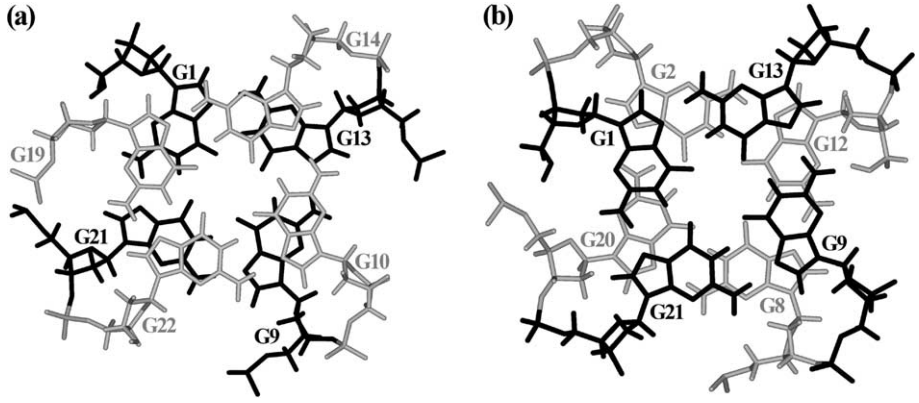

Figure 5. Base–base stacking patterns within the 3D structure of K^+-form of $d(G_3T_4G_4)_2$ G-quadruplex. (a) Stacking of G19–G22–G10–G14 quartet (in grey) over G1–G21–G9–G13 quartet (in black). (b) Stacking of G1–G21–G9–G13 quartet (in black) over G2–G20–G8–G12 quartet (in grey).

tion of the strands is accommodated by the sharp reversal in strand direction between residues G19 and G20 that results in both G-tracts of G12–G22 strand being parallel. Interestingly reversal in strand direction is accompanied by the leap over the middle G-quartet (Fig. 6) and is achieved without any extra intervening residues, which is a feature that has not been observed before in G-quadruplex structures. No unusual torsion angle conformations have been observed due to the sharp reversal in strand direction except for the conformation across the glycosidic bond for G20 residue, which is close to *high-anti* region ($\chi \approx -41° \pm 1°$). The subsequent G21 residue adopts normal *anti* ($\chi \approx -138° \pm 0.5°$) glycosidic bond conformation.

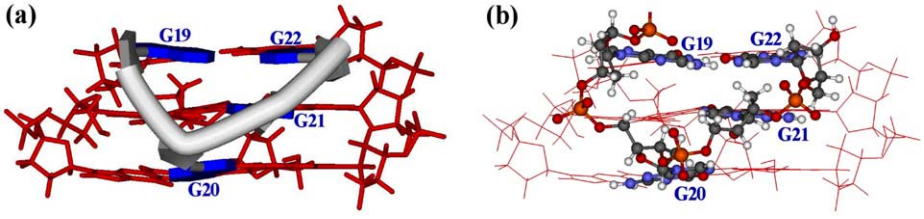

Figure 6. A view of the sharp reversal in strand direction between residues G19 and G20 and a leap over the middle G-quartet in the structure of $d(G_3T_4G_4)_2$ G-quadruplex. Ribbon (a) and ball and stick (b) views are shown. Residues G19–G22 are marked.

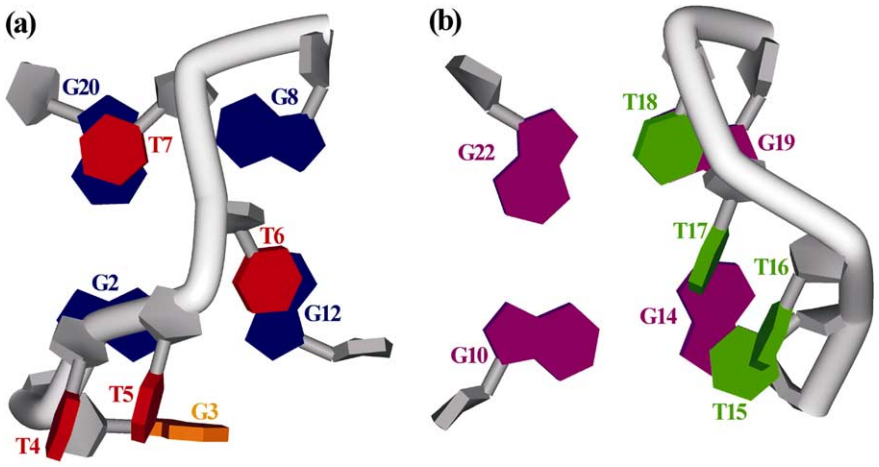

Figure 7. Bird's-eye view of the G3–T7 diagonal loop (a) and T15-T18 edge-type loop (b) with the neighbouring G-quartets.

2.4. Thymine loops

The examination of eight final structures obtained through NMR restrained simulated annealing calculations shows that $d(G_3T_4G_4)_2$ G-quadruplex contains two thymine loops that reside on the opposite sides of the G-quadruplex core. Thymines T4–T7 together with G3 residue form a loop that spans the diagonal of G12–G8–G20–G2 end G-quartet (Fig. 7a). The perusal of Figure 7a shows that T7 is stacked on G20 whereas T6 is stacked on G12. T4 and T5 are stacked onto each other. Residues T15–T18 constitute edge-type loop (Fig. 7b). T15 and T18 are stacked on G14 and G19 residues of the end G-quartet, respectively, while T16 and T17 stack onto each other. It is noteworthy that T4–T7 loop exhibits smaller number of NOE contacts compared to T15–T18 loop, which might correspond to a larger degree of conformational flexibility of the T4–T7 loop. Structural details of both loops influence ion exchange through the specific steric hindrance imposed on the nearby G-quartets. The comparison of Figure 7a and b suggests that the diagonal loop represents a larger steric barrier than the edge-type loop for bulk ions (or solvent molecules) to enter G-quadruplex core.

The comparison of NMR structure of $d(G_3T_4G_4)_2$ with NMR and X-ray structures of $d(G_4T_4G_4)_2$ G-quadruplex[31–34] shows that T4–T7 diagonal loop in the former adopts significantly different conformation compared to the diagonal loops of the latter. The first two thymines in $d(G_4T_4G_4)_2$ G-quadruplex structure form a continuous stack on their neighbouring G-quartet. T4 and T5 in $d(G_3T_4G_4)_2$, on the other hand, are stacked onto each other but not on the neighbouring G-quartet. The remaining two thymine residues also show different stacking pattern in $d(G_3T_4G_4)_2$ compared to $d(G_4T_4G_4)_2$ G-quadruplex. It is noteworthy that G3 residue in $d(G_3T_4G_4)_2$ also participates in the diagonal loop and is likely to be one of the reasons for aforementioned conformational differences between the diagonal loops of $d(G_3T_4G_4)_2$ and $d(G_4T_4G_4)_2$ quadruplex structures. The second loop of $d(G_3T_4G_4)_2$ spans the edge of the outer G-quartet, which is not found in the related $d(G_4T_4G_4)_2$ high-resolution NMR and X-ray structures.

2.5. Influence of different monovalent cations on the folding of the $d(G_3T_4G_4)_2$

As the folding of G-rich oligonucleotides into G-quadruplex structures is extremely sensitive to the nature of cation species present in solution and their concentration[48,42,36–40,35] we then examined the influence of different monovalent cations on the folding of $d(G_3T_4G_4)$ oligonucleotide. Aqueous solutions of KCl, NaCl and NH_4Cl were titrated into the samples of unfolded oligonucleotide. Twelve well-resolved imino resonances were observed in the region characteristic of G-quartet formation, which indicated the formation of three G-quartets in the presence of any of the three monovalent cations (Fig. 2). The assignment of imino resonances, however, revealed significant chemical shift differences as a function of cation species in solution. It is noteworthy that shielding effects for particular nuclei are not identical for all imino resonances. For example, G1(NH) is found most downfield in the presence of Na^+ ions and 0.35 ppm upfield in NH_4^+ ion containing solution. In contrast, G2(NH) is most downfield in the presence of NH_4^+ ions while it is shielded by 0.08 ppm in the presence of K^+ ions (Fig. 2). The comparative analysis of chemical shift changes for imino protons in the presence of K^+, NH_4^+ or Na^+ ions showed no correlation with *syn* or *anti* orientations of individual guanine residue or its position within a particular G-quartet and in a 3D fold of $d(G_3T_4G_4)_2$ G-quadruplex. Similarly, specific (de)shielding of aromatic protons for individual guanine and thymine residues in the presence of K^+, NH_4^+ or Na^+ ions could not be correlated with structural details of $d(G_3T_4G_4)_2$ G-quadruplex. The chemical shift changes observed for H6 protons of thymines are of comparable magnitude to the changes for H8 of guanines. The comparison of NOESY spectra of $d(G_3T_4G_4)_2$ G-quadruplexes folded in the presence of K^+ or Na^+ ions also reveals substantial chemical shift differences as evident from NOE cross-peaks in the aromatic–anomeric region (Fig. 8a and c). Distinctive

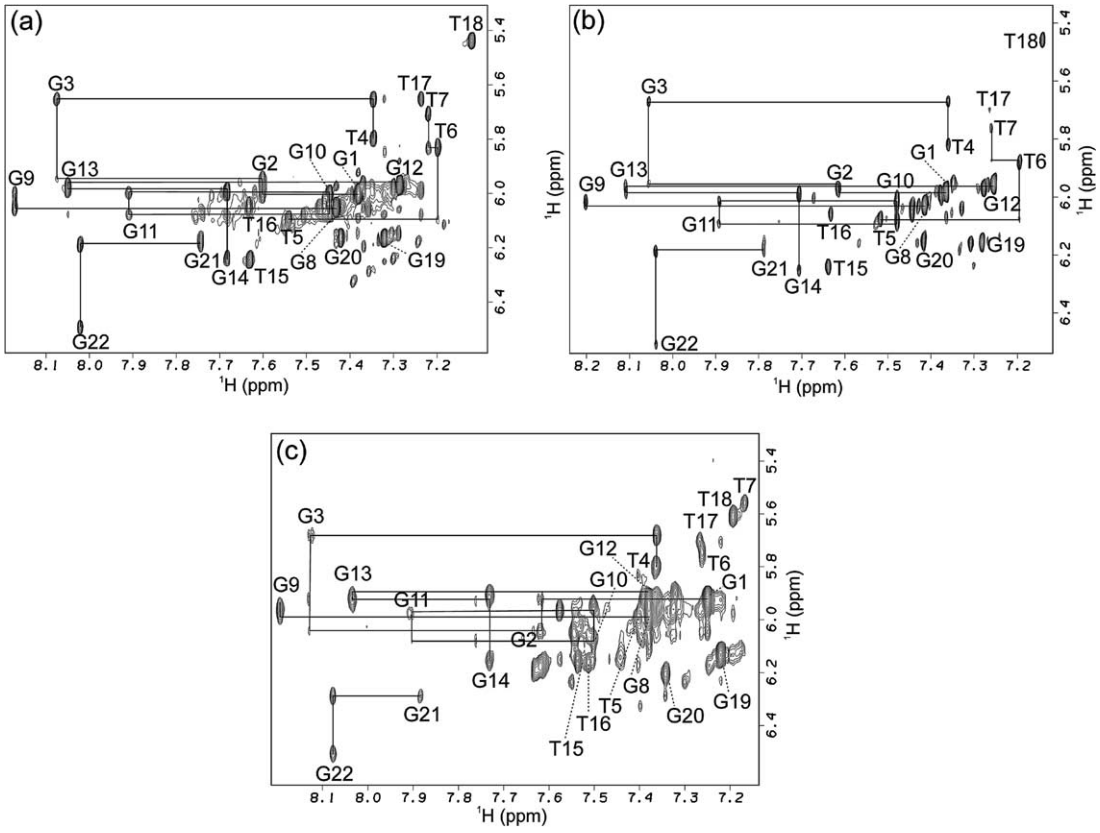

Figure 8. Aromatic–anomeric region of NOESY spectra of $d(G_3T_4G_4)_2$ at 25 °C and mixing time of 400 ms in the presence of 15 mM KCl (a), 15 mM $^{15}NH_4Cl$ (b) and 15 mM NaCl (c). Sequential aromatic-H1′ connectivities are labelled. The sample concentration was 4.0 mM (a), 3.8 mM (b) and 3.6 mM (c) in strand.

(de)shielding effects were attributed to different positions of cations along the central ion cavity of the G-quadruplex structure. It has been suggested that smaller Na^+ ions bind inside the G-quartet plane, while larger K^+ ions coordinate in the central position between two G-quartets.[46,37,34] In full support, there are almost negligible chemical shift differences between $d(G_3T_4G_4)_2$ G-quadruplexes formed in the presence of K^+ ions compared to NH_4^+ ions. The identical connectivities were observed in the aromatic–anomeric regions of respective NOESY spectra (Fig. 8a and b). This is consistent with similar ionic radii of K^+ and NH_4^+ ions, which consequently lead to their preferential binding to similar sites between the neighbouring G-quartets. In this way, the relative positions of stacked guanine rings and therefore their (de)shielding through ring currents is reflected in similar chemical shifts of nearby aromatic and sugar resonances.

The significant differences in chemical shifts of imino and aromatic ^{1}H resonances could possibly suggest different topologies of $d(G_3T_4G_4)_2$ G-quadruplexes formed in the presence of different monovalent cations. The thorough examination of NOE contacts in aromatic–anomeric region of NOESY spectra shown in Figure 8, however, revealed identical NOE sequential connectivity patterns and relative cross-peak volume intensities in the presence of K^+, Na^+ and NH_4^+ ions, which led us to the conclusion that $d(G_3T_4G_4)$ adopted the G-quadruplex structure with the same general fold in the presence of any of these three cations (compare Fig. 8a–c). This was additionally confirmed through detailed examination of other regions of NOESY spectra. In particular, NOE contacts in imino–aromatic region proved that each individual G-quartet consists of the same guanine bases regardless of ions present in solution (see Figs. 3 and S1). In the case of the NH_4^+ ions the assignment of hydrogen-bonding connectivities within the individual G-quartets was hampered due to the absence of some imino–H8 correlations. We then compared distances derived from several key interquartet and intraquartet H8–H8, NH–H8 and NH–NH NOE cross-peaks in the presence of K^+, Na^+ and NH_4^+ ions. The thorough comparative examination of NOEs clearly ascertained only minor differences in their intensities within the error limits of the experiment for K^+, Na^+ and NH_4^+ forms of G-quadruplex, which suggests that the three monovalent cations cause only subtle structural differences.

3. Conclusion

Guanine-rich regions of DNA display an extraordinary structural polymorphism. We have shown recently[35] that even a minor change at the 3′-end of the sequence strongly affects 3D structure, which in the case of $d(G_4T_4G_3)_2$ G-quadruplex exhibited a slipped strand in comparison to the related $d(G_4T_4G_4)_2$. To our surprise, the formal removal of 5′ terminal dG residue from

$d(G_4T_4G_4)$ sequence caused even more drastic structural perturbations that were found in $d(G_3T_4G_4)_2$ G-quadruplex. NMR study showed that $d(G_3T_4G_4)_2$ forms an asymmetric, dimeric G-quadruplex structure in the presence of K^+, Na^+ and NH_4^+ ions. The structure exhibits a number of unusual structural features. First, it consists of diagonal as well as edge-type loops. Second, G19–G22–G10–G14 and G1–G21–G9–G13 G-quartets exhibit G(*syn*)–G(*anti*)–G(*anti*)–G(*anti*) alternation of the glycosidic torsion angles and the same hydrogen-bonding donor–acceptor directionalities. Such exceptional *syn*/*anti* arrangement around G-quartet plane is rare.[55,56] Third, $d(G_3T_4G_4)_2$ G-quadruplex exhibits three parallel and one anti-parallel G-rich run. Fourth, the sharp reversal in strand direction between residues G19 and G20 accompanied with the leap over the middle G-quartet is another unique characteristic of the $d(G_3T_4G_4)_2$ G-quadruplex structure. It widens the palette of motifs that enable reversal of the strand direction within G-quadruplex structures. In the previously known double-chain reversal loops observed in $d(G_3AG_2T_3G_3AT)$ and $d(TAG_3TTAG_3T)_2$ G-quadruplexes strand direction was reversed by the means of TTA residues and resulted in the propeller-like structures with all four parallel G-strands.[56,13] In $d(G_3T_4G_4)_2$ G-quadruplex, however, reversal in strand direction is achieved without any intervening residues. It most closely resembles the so-called V-shaped loop in $d(G_3AG_2T_3G_3AT)$ G-quadruplex structure.[56] Fifth, $d(G_3T_4G_4)$ folds into a single G-quadruplex in the presence of K^+, Na^+ as well as NH_4^+ ions. The detailed examination of NOESY spectra revealed the same connectivity patterns in the presence of any of the three counterions, which is in agreement with the same folding topology. The comparison of distances derived from several key inter- and intraquartet H8–H8, NH–H8 and NH–NH NOE cross-peaks in K^+, Na^+ and NH_4^+ ion forms of G-quadruplex structures showed only minor differences. The significant differences in proton chemical shifts were interpreted in terms of distinct cation localization in binding sites along ion channel of G-quadruplex core. The stabilization of the same folding topology in the presence of potassium and sodium ions, that are omnipresent in our cells, contributes to our knowledge on the role of ions in the folding of G-rich sequences including telomere DNA repeats and biological roles of G-quadruplex structures.

Our study represents a nice example where single nucleotide substitution/deletion leads to unexpected folding into a novel 3D structure that would be difficult to predict if one was to correlate substitution/deletion with chemical and enzymatic probing experiments.

4. Experimental

4.1. Sample preparation

Oligonucleotides were synthesised on an Expedite 8909 synthesizer using phosphoramidite chemistry following the manufacturer's protocol and deprotected with concentrated aqueous ammonia. DNA was purified on 1.0 m Sephadex G25 column. Fractions containing only full-length oligonucleotide were pooled, lyophilized, redissolved in 1 mL H_2O and extensively dialyzed against 10 mM LiCl. The DNA was then lyophilized and subsequently redissolved in 0.3 mL of 90% H_2O/10% ^2H_2O or 99.996% ^2H_2O. LiOH or HCl were added to adjust pH of the sample to around 5.5 for optimal observation of imino cross-peaks in G-quadruplexes. Sample concentrations were between 1 and 4 mM in strand (0.5–2.0 mM in G-quadruplex). For structure determination higher concentrations were used in order to achieve higher signal to noise ratios. KCl, NaCl or $^{15}NH_4Cl$ were titrated into the samples up to 20 mM concentrations.

4.2. NMR spectroscopy and data processing

NMR data were collected on a Varian Unity Inova 600 MHz NMR spectrometer. We recorded NOESY spectra in H_2O at 5 °C, and NOESY spectra, double-quantum filtered COSY (DQF-COSY), P.E.COSY and TOCSY spectra in ^2H_2O at 5 and 25 °C. NOESY spectra in H_2O (collected at mixing times of 80 and 300 ms) were acquired with solvent suppression using the WATERGATE pulse sequence, with 4096 complex points in the t_2 dimension, 400 increments in the t_1 dimension and a spectral width of 9.9 kHz. Recovery delay between experiments was 3 s. NOESY spectra in ^2H_2O were acquired at mixing times of 80, 150, 250 and 400 ms with 4096 complex points in the t_2 dimension and 400 increments in the t_1 dimension with a spectral width of 6.7 kHz. The phase-sensitive DQF-COSY spectra were collected with 8192 complex points in the t_2 dimension and 256 increments in the t_1 dimension. Homonuclear TOCSY spectra in ^2H_2O were collected with 4096 complex points in the t_2 dimension and 256 increments in the t_1 dimension with a spin lock time of 100 ms. P.E.COSY spectra, which were used to obtain $^3J_{H1'H2'}$ and $^3J_{H1'\,H2''}$ coupling constants, were collected with 8192 complex points in the t_2 dimension and 256 increments in the t_1 dimension with a spectral width of 6.3 kHz. Spectra were processed and analyzed using FELIX 2000 program (Accelrys) and VNMR 6.1B software (Varian).

4.3. NMR restraints derivation

Distance restraints for nonexchangeable protons were derived from NOESY spectra acquired at mixing times of 80, 150 and 250 ms. NOE cross-peaks were integrated using FELIX 2000 program (Accelrys). The H2′–H2″ NOE cross-peaks were equated with a distance of 1.9 Å and all the remaining intensities were converted to distances using the $I \sim r_{ij}^{-6}$ relation. The upper and lower bounds were assigned to ±30%. NOE cross-peaks involving exchangeable protons were derived from NOESY spectra in H_2O. These were classified as strong (strong intensity at 80 ms), medium (weak intensity at 80 ms) and weak (peak observed only at 300 ms) and were restrained to distances of 2.7 (±0.9), 3.8 (±1.2) and 5.0 (±1.5) Å, respectively. A total of 414 distance restraints, of which 79 involved exchangeable protons, were derived from NOESY spectra. In each of the three G-quartets an additional eight hydrogen-bond restraints were applied

to account for hydrogen bonding within individual G-quartet.

Homonuclear $^3J_{H1'H2'}$ and $^3J_{H1'H2''}$ coupling constants were derived from the P.E.COSY spectra and were interpreted in terms of a two-state North↔South pseudorotational equilibrium with the use of the computer program PSEUROT.[57] The PSEUROT program finds the best fit between experimental and calculated $^3J_{HH}$ coupling constants. During the iteration procedure we have systematically varied pseudorotation phase angles of North ($0° < P < 30°$) and South conformers ($140° < P < 180°$), as well as both puckering amplitudes ($35° < \Psi_m < 40°$), while optimizing populations of North- and South-type conformers. The sugar moieties of all guanine residues showed a strong bias of over 80% towards South-type conformations. The sugar moieties of thymine residues, which constitute the two loops, exhibited >70% preference for South-type conformers. The torsion angle restraints for ν_1, ν_2 and ν_3 were used in structure calculations in order to constrain the puckering of individual sugar moieties within $d(G_3T_4G_4)_2$ into the South region of conformational space.

4.4. Structure calculations

All calculations were performed using the AMBER 7[58] program with a Cornell et al. (1999) force field[59,60] using the described protocol.[1]

Coordinates of the eight structures of $d(G_3T_4G_4)_2$, which best satisfied NMR restraints and exhibited the lowest total energy, have been deposited in the RCSB Protein Data Bank (accession number 1U64).

Acknowledgements

We thank the Ministry of Education, Science and Sport of the Republic of Slovenia (Grant Nos J1-3309-0104 and P1-0242-0104) and European Commission (Contract No ICA1-CT-2000-70034) for their financial support.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2004.08.009. Figure of imino–aromatic region of a NOESY spectrum of $d(G_3T_4G_4)_2$ folded in the presence of Na^+ ions.

References and notes

1. Crnugelj, M.; Sket, P.; Plavec, J. *J. Am. Chem. Soc.* **2003**, *125*, 7866–7871.
2. Keniry, M. A. *Biopolymers* **2001**, *56*, 123–146.
3. Neidle, S.; Parkinson, G. N. *Curr. Opin. Struct. Biol.* **2003**, *13*, 275–283.
4. Arthanari, H.; Bolton, P. H. *Chem. Biol.* **2001**, *8*, 221–230.
5. Neidle, S.; Read, M. A. *Biopolymers* **2001**, *56*, 195–208.
6. Shafer, R. H.; Smirnov, I. *Biopolymers* **2001**, *56*, 209–227.
7. Rezler, E. M.; Bearss, D. J.; Hurley, L. H. *Curr. Opin. Pharmacol.* **2002**, *2*, 415–423.
8. Hurley, L. H. *Nat. Rev. Cancer* **2002**, *2*, 188–200.
9. Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Pluckthun, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8572–8577.
10. Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593–11598.
11. Williamson, J. R. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 703–730.
12. Bryan, T. M.; Cech, T. R. *Curr. Opin. Cell Biol.* **1999**, *11*, 318–324.
13. Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876–880.
14. Kerwin, S. M. *Curr. Pharm. Des.* **2000**, *6*, 441–471.
15. Simonsson, T. *Biol. Chem.* **2001**, *382*, 621–628.
16. Hahn, W. C.; Stewart, S. A.; Brooks, M. W.; York, S. G.; Eaton, E.; Kurachi, A.; Beijersbergen, R. L.; Knoll, J. H. M.; Meyerson, M.; Weinberg, R. A. *Nat. Med.* **1999**, *5*, 1164–1170.
17. Hanahan, D.; Weinberg, R. A. *Cell* **2000**, *100*, 57–70.
18. Kelland, L. R. *Anti-Cancer Drugs* **2000**, *11*, 503–513.
19. Han, H. Y.; Hurley, L. H. *Trends Pharmacol. Sci.* **2000**, *21*, 136–142.
20. Read, M.; Harrison, R. J.; Romagnoli, B.; Tanious, F. A.; Gowan, S. H.; Reszka, A. P.; Wilson, W. D.; Kelland, L. R.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4844–4849.
21. Shi, D. F.; Wheelhouse, R. T.; Sun, D. Y.; Hurley, L. H. *J. Med. Chem.* **2001**, *44*, 4509–4523.
22. Mergny, J. L.; Lacroix, L.; Teulade-Fichou, M. P.; Hounsou, C.; Guittat, L.; Hoarau, M.; Arimondo, P. B.; Vigneron, J. P.; Lehn, J. M.; Riou, J. F.; Garestier, T.; Helene, C. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3062–3067.
23. Gowan, S. M.; Harrison, J. R.; Patterson, L.; Valenti, M.; Read, M. A.; Neidle, S.; Kelland, L. R. *Mol. Pharmacol.* **2002**, *61*, 1154–1162.
24. Kim, M. Y.; Vankayalapati, H.; Kazuo, S.; Wierzba, K.; Hurley, L. H. *J. Am. Chem. Soc.* **2002**, *124*, 2098–2099.
25. Gavathiotis, E.; Heald, R. A.; Stevens, M. F. G.; Searle, M. S. *J. Mol. Biol.* **2003**, *334*, 25–36.
26. Tauchi, T.; Shin-ya, K.; Sashida, G.; Sumi, M.; Nakajima, A.; Shimamoto, T.; Ohyashiki, J. H.; Ohyashiki, K. *Oncogene* **2003**, *22*, 5338–5347.
27. Gomez, D.; Lemarteleur, T.; Lacroix, L.; Mailliet, P.; Mergny, J. L.; Riou, J. F. *Nucleic Acids Res.* **2004**, *32*, 371–379.
28. Seela, F.; Kroschel, R. *Org. Biomol. Chem.* **2003**, *1*, 3900–3908.
29. da Silva, M. W. *Biochemistry* **2003**, *42*, 14356–14365.
30. Risitano, A.; Fox, K. R. *Nucleic Acids Res.* **2004**, *32*, 2598–2606.
31. Schultze, P.; Smith, F. W.; Feigon, J. *Structure* **1994**, *2*, 221–233.
32. Smith, F. W.; Feigon, J. *Nature* **1992**, *356*, 164–168.
33. Smith, F. W.; Feigon, J. *Biochemistry* **1993**, *32*, 8682–8692.
34. Haider, S.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2002**, *320*, 189–200.
35. Crnugelj, M.; Hud, N. V.; Plavec, J. *J. Mol. Biol.* **2002**, *320*, 911–924.
36. Hud, N. V.; Smith, F. W.; Anet, F. A. L.; Feigon, J. *Biochemistry* **1996**, *35*, 15383–15390.
37. Schultze, P.; Hud, N. V.; Smith, F. W.; Feigon, J. *Nucleic Acids Res.* **1999**, *27*, 3018–3028.

38. Kettani, A.; Bouaziz, S.; Gorin, A.; Zhao, H.; Jones, R. A.; Patel, D. J. *J. Mol. Biol.* **1998**, *282*, 619–636.
39. Bouaziz, S.; Kettani, A.; Patel, D. J. *J. Mol. Biol.* **1998**, *282*, 637–652.
40. Miyoshi, D.; Nakao, A.; Toda, T.; Sugimoto, N. *FEBS Lett.* **2001**, *496*, 128–133.
41. Strahan, G. D.; Keniry, M. A.; Shafer, R. H. *Biophys. J.* **1998**, *75*, 968–981.
42. Keniry, M. A.; Strahan, G. D.; Owen, E. A.; Shafer, R. H. *Eur. J. Biochem.* **1995**, *233*, 631–643.
43. Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2003**, *125*, 15021–15027.
44. Searle, M. S.; Williams, H. E. L.; Gallagher, C. T.; Grant, R. J.; Stevens, M. F. G. *Org. Biomol. Chem.* **2004**, *2*, 810–812.
45. Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263–282.
46. Phillips, K.; Dauter, Z.; Murchie, A. I. H.; Lilley, D. M. J.; Luisi, B. *J. Mol. Biol.* **1997**, *273*, 171–182.
47. Sket, P.; Crnugelj, M.; Kozminski, W.; Plavec, J. *Org. Biomol. Chem.* **2004**, *2*, 1970–1973.
48. Strahan, G. D.; Shafer, R. H.; Keniry, M. A. *Nucleic Acids Res.* **1994**, *22*, 5447–5455.
49. Wang, Y.; Patel, D. J. *J. Mol. Biol.* **1995**, *251*, 76–94.
50. Feigon, J.; Koshlap, K. M.; Smith, F. W. *Methods Enzymol.* **1995**, *261*, 225–255.
51. Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2002**, *124*, 1160–1161.
52. Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489–2498.
53. Tsui, V.; Case, D. A. *Biopolymers* **2001**, *56*, 275–291.
54. Xia, B.; Tsui, V.; Case, D. A.; Dyson, H. J.; Wright, P. E. *J. Biomol. NMR* **2002**, *22*, 317–331.
55. Wang, Y.; Patel, D. J. *Structure* **1994**, *2*, 1141–1156.
56. Zhang, N.; Gorin, A.; Majumdar, A.; Kettani, A.; Chernichenko, N.; Skripkin, E.; Patel, D. J. *J. Mol. Biol.* **2001**, *311*, 1063–1079.
57. de Leeuw, F. A. A. M.; Altona, C. *J. Comput. Chem.* **1983**, *4*, 428–437.
58. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Chetham, T. E.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Mertz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 7; University of California: San Francisco, 2002.
59. Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845–862.
60. Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.