J|A|C|S
ARTICLES
Published on Web 11/23/2004

Loop-Length-Dependent Folding of G-Quadruplexes

Pascale Hazel,[†] Julian Huppert,[‡] Shankar Balasubramanian,[‡] and Stephen Neidle*[,†]

Contribution from the Cancer Research UK Biomolecular Structure Group, The School of Pharmacy, University of London, 29-39 Brunswick Square, London WC1N 1AX, U.K., and University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

Received August 11, 2004; E-mail: stephen.neidle@ulsop.ac.uk

Abstract: Guanine-rich DNA sequences can form a large number of structurally diverse quadruplexes. These vary in terms of strand polarity, loop composition, and conformation. We have derived guidelines for understanding the influence of loop length on the structure adopted by intramolecular quadruplex-forming sequences, using a combination of experimental (using CD and UV melting data) and molecular modeling and simulation techniques. We find that a parallel-stranded intramolecular quadruplex structure is the only possible fold when three single residue loops are present. When single thymine loops are present in combination with longer length loops, or when all loops are longer than two residues, both parallel- and antiparallel-folded structures are able to form. Multiple conformations of each structure are likely to coexist in solution, as they were calculated to have very similar free energies.

Introduction

Guanine bases in DNA (and RNA) can associate by means of Watson−Crick and Hoogsteen hydrogen bonds to form planar G-quartets. Within guanine-rich nucleic acid sequences, the formation of several consecutive G-quartets can in turn form inter- or intramolecular four-stranded structures, termed G-quadruplexes, in which the G-quartets form a platform connected by the intervening sequences.[1−4]

Telomeres, at the ends of all eukaryotic chromosomes, comprise such G-rich regions, with tandem repeats of simple sequences. The d(T_2AG_3) telomeric repeat of human telomeres has been shown to form a G-quadruplex both in solution and in the crystalline state.[5,6] Other telomeric sequences, including d(T_4G_4) from *Oxytricha* and d(T_2G_4) from *Tetrahymena* also form stable G-quadruplexes.[7−9] The extreme 5′ ends of human telomeres consist of 100−200 bases of single-stranded DNA, which could potentially fold into G-quadruplexes in vivo. This feature is being exploited in the design and development of small molecules that stabilize quadruplex formation and thereby inhibit the action of the cancer-cell-selective reverse transcriptase enzyme telomerase. There is also growing evidence for the existence of proteins that bind G-quadruplexes in vivo.[10−13]

Potential quadruplex-forming sequences have been identified in other parts of the human genome. G-rich strands of the human insulin gene can form G-quadruplex structures, as shown by Catasti et al. using NMR and CD spectroscopy,[14] as can sequences from immunoglobulin switch regions.[15] A well-studied sequence is the nuclease hypersensitivity element III1 (NHE III1) upstream of the P1 and P2 promoters in the *c-myc* gene, which has been shown to form quadruplex structures.[16,17] Recent NMR and CD data suggest that this sequence may prefer to form parallel-stranded conformations rather than the antiparallel-stranded quadruplex conformations previously reported.[18,19] A mixed parallel/antiparallel structure has also been suggested when this quadruplex interacts with a stabilizing ligand.[19] G-rich RNA can also fold into quadruplex structures, for example, the insulin-like growth factor II (IGF II) mRNA.[20] Also the Fragile X mental retardation protein FMRP has been found to target an mRNA quadruplex.[21]

The polarity of strands with respect to each other can vary, with an all-parallel orientation being observed for quadruplexes

[†] University of London.
[‡] University of Cambridge.

(1) Sundquist, W. I.; Klug, A. *Nature* **1989**, *342*, 825−829.
(2) Henderson, E.; Hardin, C. C.; Walk, S. K.; Tinoco, I., Jr.; Blackburn, E. H. *Cell* **1987**, *51*, 899−908.
(3) Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. *Cell* **1989**, *59*, 871−880.
(4) Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1962**, *48*, 2013−2018.
(5) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263−282.
(6) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876−880.
(7) Haider, S.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2002**, *320*, 189−200.
(8) Schultze, P.; Smith, F. W.; Feigon, J. *Structure* **1994**, *2*, 221−233.
(9) Wang, Y.; Patel, D. J. *Structure* **1994**, *2*, 1141−1156.
(10) Horvath, M. P.; Schultz, S. C. *J. Mol. Biol.* **2001**, *310*, 367−377.
(11) Frantz, J. D.; Gilbert, W. *J. Biol. Chem.* **1995**, *270*, 9413−9419.
(12) Bashkirov, V., I.; Scherthan, H.; Solinger, J. A.; Buerstedde, J.-M.; Hindley, A. D. *J. Cell Biol.* **1997**, *136*, 761−773.
(13) Arimondo, P. B.; Riou, J.-F.; Mergny, J.-L.; Tazi, J.; Sun, J. S.; Garestier, T.; Hélène, C. *Nucleic Acids Res.* **2000**, *28*, 4832−4838.
(14) Catasti, P.; Chen, X.; Moyzis, R. K.; Bradbury, E. M.; Gupta, G. *J. Mol. Biol.* **1996**, *264*, 534−545.
(15) Sen, D.; Gilbert, W. *Nature* **1988**, *334*, 364−366.
(16) Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.* **1998**, *26*, 1167−1172.
(17) Rangan, A.; Fedoroff, O. Y.; Hurley, L. H. *J. Biol. Chem.* **2001**, *276*, 4640−4646.
(18) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Am. Chem. Soc.* **2004**, *126*, 8710−8716.
(19) Seenisamy, J.; Rezler, E. M.; Powell, T. J.; Tye, D.; Gokhale, V.; Joshi, C. S.; Siddiqui-Jain, A.; Hurley, L. H. *J. Am. Chem. Soc.* **2004**, *126*, 8702−8709.
(20) Christiansen, J.; Kofod, M.; Nielsen, F. C. *Nucleic Acids Res.* **1994**, *22*, 5709−5716.
(21) Darnell, J. C.; Warren, S. T.; Darnell, R. B. *Ment. Retard. Dev. Disab. Res. Rev.* **2004**, *10*, 49−52.

10.1021/ja045154j CCC: $27.50

formed from four separate strands, and antiparallel orientations for bimolecular quadruplexes. The sequences that intervene between the guanines that form G-quartets can form loops, typified by the T_4 loops in the d[$(G_4T_4G_4)_2$] bimolecular quadruplex and the T_2A loops in quadruplexes formed by repeats of the human telomeric sequence d(T_2AG_3). Loops play a key role in determining the nature of the folding and stability of G-quadruplexes, and thus different loop lengths and sequences can either stabilize or destabilize quadruplexes.[22] This has been attributed to intraloop interactions such as hydrogen bonding and stacking.[23] Recent studies involving quadruplexes with nonnucleosidic linkers instead of loops showed preferential formation of parallel quadruplexes, independently of the linker length.[23,24] However, no general rules governing the relationships between the loops and the nature of the quadruplex folding have been proposed. In large part this is due to the relative paucity of detailed structures for G-quadruplexes, with just 11 crystal structures having been reported,[6,7,10,25−30] and a similar number of high-resolution NMR structures.[5,8,9,31−36] There are as yet no rules about the specific folded structure adopted by quadruplexes of a given sequence, a long-term goal of such studies.

We report here on a combined experimental and molecular simulation study, as a first step in systematically probing the dependency of G-quadruplex folding on loop length. The sequences studied in this work are based upon the human telomeric repeat d[$AG_3(T_2AG_3)_3$]. Two very distinct structural polymorphs of this sequence have been separately identified, a Na^+ form by NMR[5] and a K^+ form by X-ray crystallography.[6] Recent data suggest that both these conformations are likely to coexist in K^+ solution.[37,38] These quadruplexes have also been probed using photochemical methods, which can detect the presence of diagonal loops in antiparallel structures.[39] There is also NMR evidence that the related shorter sequence d($TAG_3T_2AG_3T$) forms interconverting parallel and antiparallel conformations in solution, with relative stabilities of each form varying with temperature.[37] Other sequences have also shown this behavior, such as d($TG_4T_2G_4T$).[31] Structures selected in this study have all-thymine residue loops, as these are commonly found in telomeric G-quadruplex loops (for example, from *Oxytricha* and yeast). Thymine residues within the loops can form stacking interactions, but can also hydrogen bond with other residues, and are small enough, compared to adenine or guanine, to have several residues in very close proximity within the loops. Adenine residues are also found in telomeric quadruplex loops, but their occurrence is generally limited, as in the human TTA loop sequence. Cytosine residues are much less common in loops, probably due to the possible competition with Watson−Crick hydrogen bonding to guanine residues. Guanine residues are ambiguous in that they can form part of the loop or G-quartet regions, and were never included in the loops in this study.

We have employed molecular dynamics simulations to evaluate the stability of potential structures. Several G-quadruplex simulations have been previously reported, which have given stable trajectories over nanosecond time scales.[40−46] Free energy calculations using the MM-PBSA postprocessing method, in combination with the AMBER force field, have been used to study A- and B-DNA conformations,[47] RNA loop structures,[48] and protein folding.[49] Pathways along which four guanine strands might assemble into a quadruplex have been suggested on the basis of molecular dynamics and free energy calculations using the MM-PBSA method.[41] The inclusion of quadruplex channel ions was essential to obtain meaningful results in the free energy calculations.[41] The MM-PBSA method used in these studies only yields estimates of the free energy of each conformation. The more flexible systems studied here involving quadruplexes with loops of up to six residues would require excessively long simulations to sample all of their configurational space, and hence, any energy calculation must be approximate. Even in the absence of such long time scales, free energies should yield insights into the favored conformations of quadruplexes.

Materials and Methods

Oligonucleotides. The oligonucleotides were purchased from Invitrogen (Paisley, U.K.). They were synthesized on a 50 nmol scale and supplied lyophilized. Stock solutions (50 μM) of the oligonucleotides were made with Milli-Q water. The sequences of the oligonucleotides used in these studies are shown in Table 1.

Ultraviolet Melting. UV melting curves were collected using a Varian Cary 1E UV−vis spectrophotometer, measuring the spectral absorbance at 295 nm, which has been previously identified as a key absorbance region for quadruplex formation.[50] Samples were prepared at 2 μM in a buffer containing 100 mM KCl and 10 mM Tris−HCl, pH 7.4. These samples were then heat-annealed to 90 °C and allowed to slowly cool to 4 °C over a period of several hours. In a typical experiment 400 μL of a sample was degassed in a Speedvac for 3 min, transferred to a 1 cm path length quartz cuvette, and then covered with

(22) Smirnov, I.; Shafer, R. H. *Biochemistry* **2000**, *39*, 1462−1468.
(23) Keniry, M. A.; Owen, E. A.; Shafer, R. H. *Nucleic Acids Res.* **1997**, *25*, 4389−4392.
(24) Risitano, A.; Fox, K. R. *Nucleic Acids Res.* **2004**, *32*, 2598−2606.
(25) Caceres, C.; Wright, G.; Gouyette, C.; Parkinson, G. N.; Subirana, J. A. *Nucleic Acids Res.* **2004**, *32*, 1097−1102.
(26) Haider, S. M.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2003**, *326*, 117−125.
(27) Clark, G. R.; Pytel, P. D.; Squire, C. J.; Neidle, S. *J. Am. Chem. Soc.* **2003**, *125*, 4066−4067.
(28) Phillips, K.; Dauter, Z.; Murchie, A. I.; Lilley, D. M.; Luisi, B. *J. Mol. Biol.* **1997**, *273*, 171−182.
(29) Laughlan; G.; Murchie, A. I.; Norman, D. G.; Moore, M. H.; Moody, P. C.; Lilley, D. M.; Luisi, B. *Science* **1994**, *265*, 520−524.
(30) Padmanabhan, K.; Padmanabhan, K. P.; Ferrara, J. D.; Sadler, J. E.; Tulinsky, A. *J. Biol. Chem.* **1993**, *268*, 17651−17654.
(31) Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Mol. Biol.* **2004**, *338*, 93−102.
(32) Crnugelj, M.; Sket, P.; Plavec, J. *J. Am. Chem. Soc.* **2003**, *125*, 7866−7871.
(33) Schultze, P.; Hud, N., V.; Smith, F. W.; Feigon, J. *Nucleic Acids Res.* **1999**, *27*, 3018−3028.
(34) Strahan, G. D.; Keniry, M. A.; Shafer, R. H. *Biophys. J.* **1998**, *75*, 968−981.
(35) Smith, F. W.; Lau, F. W.; Feigon, J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10546−10550.
(36) Smith, F. W.; Feigon, J. *Nature* **1992**, *356*, 164−168.
(37) Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2003**, *125*, 15021−15027.
(38) Ying, L.; Green, J. J.; Li, H.; Klenerman, D.; Balasubramanian, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14629−14634.
(39) Xu, Y.; Sugiyama, H. *J. Am. Chem. Soc.* **2004**, *126*, 6274−6279.
(40) Spackova, N.; Berger, I.; Sponer, J. *J. Am. Chem. Soc.* **1999**, *121*, 5519−5534.
(41) Stefl, R.; Cheatham, T. E., III; Spackova, N.; Fadrna, E.; Berger, I.; Koca, J.; Sponer, J. *Biophys. J.* **2003**, *85*, 1787−1804.
(42) Malliavin, T. E.; Gau, J.; Snoussi, K.; Leroy, J. L. *Biophys. J.* **2003**, *84*, 3838−3847.
(43) Stefl, R.; Spackova, N.; Berger, I.; Koca, J.; Sponer, J. *Biophys. J.* **2001**, *80*, 455−468.
(44) Spackova, N.; Berger, I.; Sponer, J. *J. Am. Chem. Soc.* **2001**, *123*, 3295−3307.
(45) Han, H.; Langley, D. R.; Rangan, A.; Hurley *J. Am. Chem. Soc.* **2001**, *123*, 8902−8913.
(46) Chowdhury, S.; Bansal, M. *J. Phys. Chem. B* **2001**, *105*, 7572−7578.
(47) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.
(48) Simmerling, C.; Miller, J. L.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 7149−7155.
(49) Lee, M. R.; Duan, Y.; Kollman, P. A. *Proteins* **2000**, *39*, 309−316.
(50) Mergny, J.-L.; Phan, A.-T.; Lacroix, L. *FEBS Lett.* **1998**, *435*, 74−78.

Table 1. Sequences of Oligonucleotides Used in This Work, Together with UV Melting Temperatures (°C)[a]

Oligo name	Oligonucleotide sequence	T_m
G_3T	d(TGGGTGGGTGGGTGGGT)	-
G_3T_2	d(TGGGTTGGGTTGGGTTGGGT)	71
G_3T_3	d(TGGGTTTGGGTTTGGGTTTGGGT)	58
G_3T_4	d(TGGGTTTTGGGTTTTGGGTTTTGGGT)	48
G_3T_5	d(TGGGTTTTTGGGTTTTTGGGTTTTTGGGT)	34
G_3T_6	d(TGGGTTTTTTGGGTTTTTTGGGTTTTTTGGGT)	20
G_3T_7	d(TGGGTTTTTTTGGGTTTTTTTGGGTTTTTTTGGGT)	-
TTA-G_3TTA	d(AGGGTTAGGGTTAGGGTTAGGGT)	63
TTA-G_3T_1	d(AGGGTTAGGGTGGGTTAGGGT)	64
TTA-G_3T_2	d(AGGGTTAGGGTTGGGTTAGGGT)	66
TTA-G_3T_3	d(AGGGTTAGGGTTTGGGTTAGGGT)	68
TTA-G_3T_4	d(AGGGTTAGGGTTTTGGGTTAGGGT)	61
TTA-G_3T_5	d(AGGGTTAGGGTTTTTGGGTTAGGGT)	61
TTA-G_3T_6	d(AGGGTTAGGGTTTTTTGGGTTAGGGT)	59
TTA-G_3T_7	d(AGGGTTAGGGTTTTTTTGGGTTAGGGT)	56

[a] All samples were measured in 100 mM K^+ solutions. G_3T was also measured in 20 mM K^+ solution, with a T_m of >80 °C.

a layer of mineral oil (Sigma-Aldrich). It was then transferred to the spectrometer and equilibrated at 10 °C for 10 min. It was then heated to 90 °C and cooled to 10 °C twice consecutively at 0.1 °C/min, with data collection occurring every 0.5 min on both the annealing and melting steps. Blank samples containing only preannealed buffer were treated in the same manner and subtracted from the collected data.

Data Analysis. T_m values for the sequences were determined from van't Hoff analysis of the melting profiles following the method of Mergny.[50] This analysis assumes a simple two-state equilibrium between the folded and unfolded forms, with linear variations of absorbance with temperature for both species. All four transitions were studied for each sequence, with the T_m values differing by <0.5 °C. Minimal hysteresis was observed for these sequences at the heating/cooling rate employed; however, it was noticeable at faster rates (data not shown). This shows that the transition occurs sufficiently rapidly with respect to this rate for the system to be in thermodynamic equilibrium under the conditions of these experiments.

Circular Dichroism. Circular dichroism experiments were performed on a Jasco J-810 spectropolarimeter using inbuilt software. Samples were prepared at 4 μM in a buffer containing 100 mM KCl and 10 mM Tris−HCl, pH 7.4. These samples were then heat-annealed to 90 °C and allowed to slowly cool to 4 °C over a period of several hours. In a typical experiment 400 μL of a sample was transferred to a 1 cm path length quartz cuvette, and scans were performed over the range 220−320 nm. A wavelength of 220 nm is the lowest achievable, because Tris absorbs strongly below 220 nm, leading to a rise in the photomultiplier voltage and unreliable results. Each trace is the result of the average of five scans at 50 nm/min, with a 2 s response time, 1 nm pitch, and 1 nm bandwidth. The samples were left to equilibrate at 4 °C for 10 min before the scans were performed. A constant flow of dry nitrogen ensured there was no condensation. A blank sample containing only buffer was treated in the same manner and subtracted from the collected data. For graphing purposes, the data were smoothed slightly, with each data point being graphed as the average of those around it and a zero correction being applied at 320 nm.

Modeling and Molecular Dynamics. Three template quadruplex structures were obtained from the Protein Data Bank for use in the simulations. The parallel conformation was based on the crystal structure of the human telomeric repeat d[$AG_3(T_2AG_3)_3$], containing K^+ ions (PDB code 1KF1). This structure contains three T_2A strand-reversal loops. The NMR structure of the identical human sequence, but with Na^+ ions, has antiparallel backbone orientations and contains two lateral loops and a central diagonal loop. The first structure of the six in PDB entry 143D was used as an antiparallel template. Both parallel and antiparallel structures were used directly after modification of the central loop to the desired length, ranging from T_1 to T_6. For short loop lengths of one or two nucleotides, an alternative antiparallel template with a lateral central loop was derived from the NMR structure of $d(T_2G_4)_4$ (PDB code 186D). This sequence differs from the human sequence by one G to A mutation in each repeat. The 186D structure contains a mixture of parallel and antiparallel guanine strands, with two lateral loops and a third strand-reversal loop. The coordinates of the first PDB entry were also used in this case. The G-quartets were left unchanged, the first and third loops were modified to T_2A, and the central loop was modified as previously. A summary of the sequences simulated is shown in Table 2, and the three template structures used are shown in Figure 1. All starting and final averaged structures are available as Supporting Information.

Table 2. Molecular Dynamics Simulations[a]

sequence	parallel	antiparallel (143D)[b]	antiparallel (186D)[c]
TTA-G_3T	4		4
TTA-G_3T_2	4	2 (unstable)	4
TTA-G_3T_3	4	4	
TTA-G_3T_4	4	4	
TTA-G_3T_5	4	4	
TTA-G_3T_6	4	4	
G_3T	4		4 (unstable)
G_3T_2	4		4

[a] The values given in columns 2−4 are the simulation times (ns). [b] Refers to the antiparallel quadruplex built from the 143D PDB structure, with one lateral loop, one diagonal loop, and a final lateral loop. [c] Refers to the antiparallel quadruplex built from the 186D PDB structure, with two lateral loops and one strand-reversal loop. Sequences were the same as those in Table 1, except that the 3′ terminal T was removed, as it is not present in the experimentally derived structures.

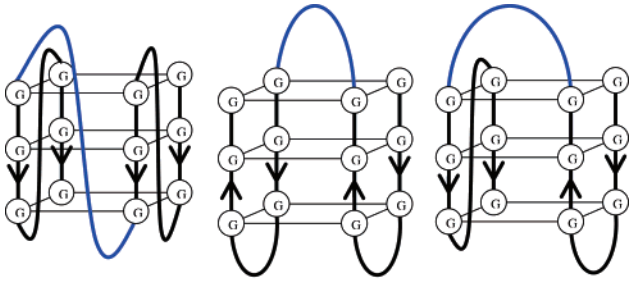

Figure 1. Template structures used in the simulations.

All structure manipulations were carried out within the InsightII suite of programs. Suitable loop conformations were generated using a simulated annealing protocol within the Discover3 program. The quartet stem was held fixed while the loops were heated to 1000 K. Molecular dynamics was carried out at this temperature for 2 ps, before slow cooling to 300 K over 1 ps, and minimizing for 10000 steps. The simulated annealing procedure was repeated on average 100 times for each quadruplex. The resulting structures were divided into clusters of similar conformations, and the lowest energy structure of the most populated cluster was typically used as a starting structure. As loop lengths were increased, an increasing number of clusters were obtained with very few structures in each. In those cases, initial structures were selected depending on their gas-phase energy as calculated by Insight as well as the presence of favorable nonbonded interactions within the loops. On average two starting structures were selected for each quadruplex, with only the most stable conformations being included in the discussion below. This procedure yielded a structure for the

parallel T_5 quadruplex with the loop residues facing the guanine quartets, with hydrogen-bonding potential between the loop and guanine residues. Structures for the T_4 and T_6 loops were generated from the T_5 loop by respective deletion/addition of a residue and simple in vacuo minimization with the Discover3 program.

Molecular dynamics simulations were carried out on all the selected structures using the AMBER7 package, with the AMBER99 force field. Two K^+ ions were placed between the quartet planes in the parallel quadruplexes, in accord with the X-ray structure. Experimental ion location information in the antiparallel NMR quadruplex structures is not available, so two K^+ ions were placed between the quartet planes, on the basis of the parallel quadruplex crystal structure. In the 143D antiparallel structures, a third K^+ ion was placed below the G-quartets, within the lateral loop region, as there was sufficient space to accommodate it. The leap module of AMBER was used to add further K^+ ions to neutralize the systems, which were then solvated in a box of TIP3P water molecules, extending 10 Å from the solute in each direction. The equilibration procedure consisted of 10 steps, beginning with a 1000 step minimization of the solvent. Dynamics of the solvent for 25 ps then enabled the water molecules to fill any gaps between the solvent and solute. The quadruplexes were minimized for 1000 steps, followed by 3 ps of dynamics with a constraint of 25.0 kcal mol^{-1}. A series of five minimizations were carried out, during which the constraint on the DNA was lowered each time by 5.0 kcal mol^{-1} until it reached 0.0 kcal mol^{-1}. Following minimization, the systems were heated to 300 K over 20 ps. MD simulations were carried out at 300 K with a 2 fs time step, and SHAKE was applied to constrain the bonds containing hydrogen. All stable quadruplexes were simulated for 4 ns. The particle mesh Ewald (PME) method of calculating long-range electrostatic interactions was employed, with a cutoff of 10 Å. The Carnal and Ptraj modules of AMBER were used to analyze the simulations.

Free Energy Calculations. Free energy calculations were carried out using the MM-PBSA postprocessing method in AMBER. Details of this method can be found in refs 47 and 51. Briefly, configurations were sampled from MD simulations in explicit solvent. The free energy was determined from molecular mechanics gas-phase energies, solvation free energies using continuum solvent methods, and entropic contributions. Additivity of all the free energy components was assumed. The gas-phase energies were calculated using the Anal module of AMBER. The Delphi program was used to calculate the electrostatic contribution to the solvation free energy. The K^+ ionic radius was adjusted until the calculated solvation energy ($\Delta G^{elect} + \Delta G^{nonpolar}$) was equal to the experimental ΔG^{solv}, of -80.6 kcal mol^{-1}. The K^+ radius was thus determined to be 2.025 Å. Dielectric constants of 1.0 and 80.0 were used for the solute and solvent, respectively. A grid spacing of 0.5 Å was chosen, with the longest linear dimension of the molecule occupying 80% of the grid. The nonpolar contribution to the solvation free energy ($\Delta G_{nonpolar} = \sigma a + b$) was determined using the MSMS program, where a is the surface area, $\sigma = 0.00542$, and $b = 0.92$ kcal mol^{-1}. Snapshots were collected every 20 ps after the equilibration period, giving 150 snapshots over 1−4 ns of each trajectory. As previously shown,[41] the inclusion of the channel K^+ ions is important in free energy calculations, and three K^+ ions were therefore included in all the calculations here. Where only two ions were present within the core of the quadruplex, the closest K^+ ion to the solute was included. The continuum model of electrostatic interactions implicitly averages over the solvent degrees of freedom, but the entropy of the solute is not included and must be calculated separately. This is more computationally demanding than the enthalpy calculations, and was therefore carried out only on a subset of structures taken every 200 ps. The Nmode module of AMBER was used to calculate the translational, rotational, and vibrational partition functions after minimization of the structures to a 10^{-5} root-mean-square gradient. A distance-dependent dielectric function was used to mimic the effect of the solvent during the minimizations (dielectric constant of $4r_{ij}$). The minimizations caused some structures to become slightly distorted, and these were not included in the entropy calculations (although this did not have an important effect on the entropies calculated).

Results

Stability of Quadruplexes with Oligo-dT Loops. A series of oligonucleotides consisting solely of four d(GGG) repeats separated by loops consisting of $d(T)_n$ were synthesized and examined (see Table 1, top half). These G_3T_n sequences were shown to form quadruplexes using the two techniques of UV melting and CD spectroscopy, which are both recognized methods for demonstrating quadruplex formation.

UV Melting Studies on Oligo-dT Loop Sequences. It has been previously shown[50] that quadruplexes melt with a significant hyperchromic shift at 295 nm. The resulting melting curves can be analyzed either by finding a maximum in the first derivative, yielding the melting temperature, T_m, or more precisely using a van't Hoff analysis, fitting the melting curve to a two-state model. The latter can be used to yield full thermodynamic parameters as well as the T_m. The $d(T)_n$ loops studied here showed clear melting transitions at 295 nm, with the exception of G_3T and G_3T_7, where any transition was either at too low or too high a temperature to be observed. These sequences were rerun in low-salt (20 mM KCl) solutions, which should result in lower T_m values.[52] G_3T showed a melting transition over 80 °C, whereas G_3T_7 showed no transition. The other sequences had melting temperatures that decreased with the length of the loops, indicating decreased stability with respect to the unfolded structures. This may be explained by considering the increased entropic cost of constraining the longer loops. Melting temperature data are shown in Table 1, and sample melting curves are shown in Figure 2. Studies over a small concentration range showed no change in the UV melting curves, suggesting that they fit much better to a unimolecular transition than an oligomolecular one.

CD Spectroscopy on Oligo-dT Loop Sequences. Quadruplexes exhibit characteristic circular dichroism signals, depending on which of two broad classes of fold they belong to. Antiparallel folds exhibit a positive CD signal at around 295 nm, with a negative signal at around 260 nm. On the other hand, parallel folds have a positive signal at approximately 260 nm, and a negative signal near 240 nm.[53−55] Unstructured sequences exhibit neither of these characteristics. The CD data shown in Figure 3 appear to imply that the sequence G_3T is purely parallel, G_3T_3−G_3T_7 are antiparallel, and G_3T_2 exhibits substantial polymorphism, and appears to consist of a superposition of the two alternative states. The decrease in the peak at 295 nm with increasing loop length can be attributed to the fact that these long loops behave largely as single strands, and oligo-dT exhibits a small negative CD signal at 295 nm.

Stability of Quadruplexes with Two TTA Loops and a Central Oligo-dT Loop. To study sequences more similar to

(51) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(52) Jing, N.; Rando, R. F.; Pommier, Y.; Hogan, M. E. *Biochemistry* **1997**, *36*, 12498−12505.
(53) Balagurumoorthy, P.; Brahmachari, S. K.; Mohanty, D.; Bansal, M.; Sasisekharan, V. *Nucleic Acids Res.* **1992**, *20*, 4061−4067.
(54) Giraldo, R.; Suzuki, M.; Chapman, L.; Rhodes, D. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7658−7662.
(55) Berova, N.; Nakanishi, K.; Woody, R. W. *Circular dichroism: principles and applications*; Wiley-VCH: New York, Chichester, U.K., 2000.

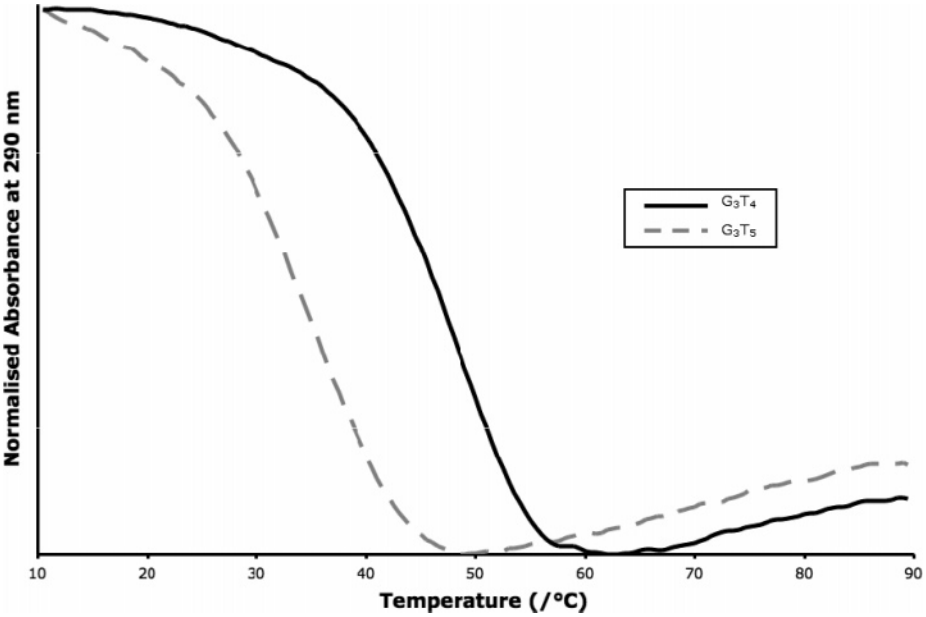

Figure 2. Sample UV melting curves for sequences G_3T_4 (solid line) and G_3T_5 (dashed line). These results are the average of a series of heating and cooling cycles between 10 and 90 °C.

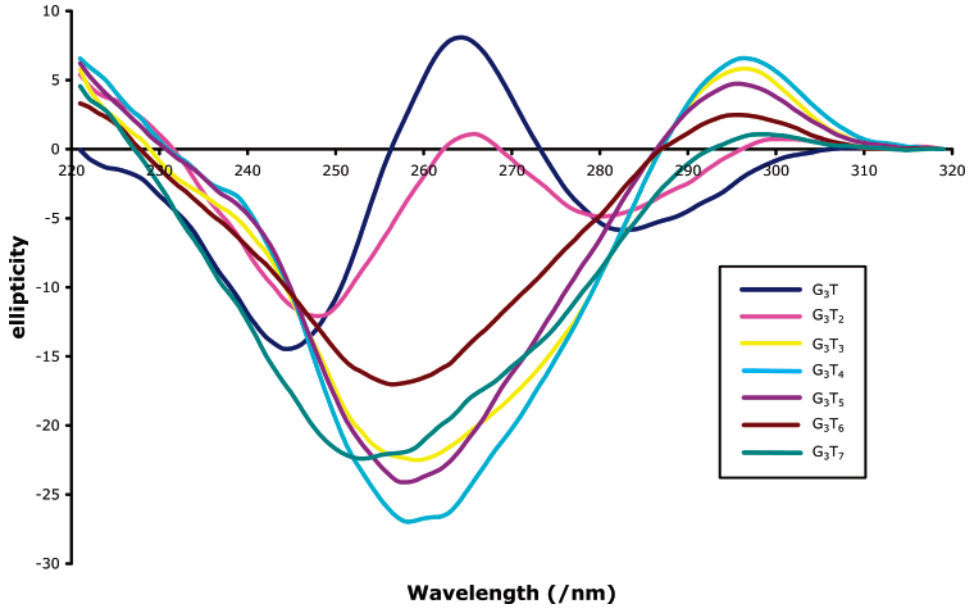

Figure 3. CD data for sequences with all loops being varied. Data were collected at 4 μM DNA concentration at 4 °C. Samples G_3T and G_3T_2 exhibit a peak at 265 nm characteristic of parallel folds (G_3T_2 possibly being polymorphic); the other sequences have a peak at 295 nm, characteristic of antiparallel folding.

the well-characterized human telomeric sequence of $(TTAGGG)_n$, facilitating accurate molecular modeling, a second series of DNA oligomers was studied using the same techniques. These are shown in Table 1 (bottom half). The parent sequence TTA-G_3TTA, which consists only of unmodified telomeric repeats, was also studied.

UV Melting Studies on Single-Loop Variants. This series of sequences showed relatively little difference in the melting temperatures, all showing a hyperchromic transition at 295 nm with T_m values that ranged from a maximum of 68 °C for TTA-G_3T_3, in which all the loops were the same length, to a minimum of 56 °C for TTA-G_3T_7. These results confirm that these sequences all form quadruplexes under these conditions.

CD Spectroscopy on Single-Loop Variants. These samples all showed CD spectra consistent with antiparallel folds, with the exception of TTA-G_3T, which had a polymorphic spectrum very similar to that for G_3T_2, suggesting that an extremely short loop can be accommodated in an antiparallel fold as long as there is only one such loop.

Molecular Dynamics Simulations. Relative stabilities of the simulated quadruplexes were compared using analysis of various criteria, such as rms deviations, planarity, and hydrogen-bonding integrity of the G-quartets. Loop structures were assessed using stacking and hydrogen-bonding abilities as well as rms deviations. The presence of ions within the central core of the quadruplexes was also related to stability. The average structure of each quadruplex over the last 2 ns of simulation is shown in Figures 5 and 6.

Molecular Dynamics 5 and 6 Simulations of Single-Loop Variants in Parallel Conformations. All the parallel quadru-

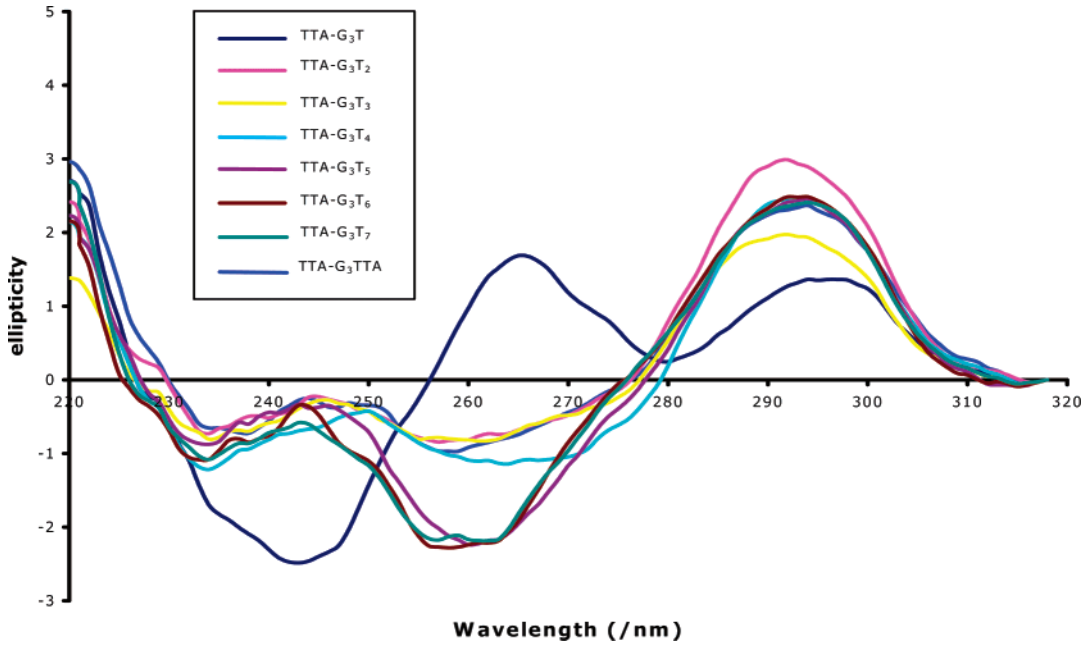

Figure 4. CD data for sequences with only the central loop being varied. Data were collected at a 4 μM DNA concentration at 4 °C. Sample TTA-G_3T exhibits a peak at 265 nm characteristic of parallel folds (though possibly being polymorphic); the other sequences have a peak at 295 nm, characteristic of antiparallel folding.

plex simulations gave stable trajectories for the G-quartets, with low rms deviations between 1.4 and 1.5 Å on average, as shown in Table 3. Different loop lengths did not influence G-quartet stability. The strand-reversal loops showed a much greater range of rms deviations than the G-quartets. The T_2A loop rms deviations were variable from simulation to simulation, ranging from 2.4 to 6.1 Å, despite being in identical starting conformations. These T_2A loops are stabilized by base-stacking interactions only, with no hydrogen bond interactions, and have high flexibility in solution. The parallel quadruplex X-ray structure loop conformation, in which the A residue stacks with the first T, and the second T faces the solvent, was often not retained during the simulations. The loops tended to adopt a less compact conformation, in which the second T residue reached further into the solvent. This T occasionally stacked with the A residue, although only for a few hundred picoseconds at most. At all other times, the A and first T loop residues stacked with each other. Only rarely during any of the simulations did no stacking take place between any two of the loop residues. The T_2A loop unraveling occurred as quickly as in the first 500 ps, in the T_6 simulation, or not at all, as in the T_5 simulation. In the T_3 loop simulation, the loops unraveled after about 1.5 ns, and then reformed at the end of the simulation to resemble the crystal structure again. The T_2A loop flexibility was not linked to the modified loop deviations, as shown by the uncorrelated rms deviations in Table 3. The same T_2A loop behavior was observed during simulations of the human sequence with three unmodified loops (results not shown).

The modified strand-reversal loops also displayed a range of rms deviations, depending on length, but also on the interactions of the loop residues with the G-quartets. Figure 7 shows the modified loop rms deviations. Increasing the loop length from T to T_3 increased the rms deviation of each loop, related to the increasing degrees of freedom (Table 3 and Figure 7). The T loop residue pointed toward the G-quartets, forming stable hydrogen bonds with two guanine residue N2 atoms. The T_2 loop residues were oriented toward the solvent, but maintained stacking interactions with each other during the whole 4 ns simulation. The T_3 loop behaved similarly to the T_2A loops during the first 3 ns of simulation, forming alternate stacking arrangements between two of the three residues. However, all stacking between the residues was lost during the final 1 ns of simulation, during which no interactions were present either between the loop residues or between the loop residues and G-quartets. The fact that no stable structure for this loop was achieved during the simulation is reflected in the higher rms deviation of 2.2 Å with respect to the average structure over the last 2 ns of simulation (Table 3). The T_3 loop rms deviation did not stabilize over the course of the simulation, as shown in Figure 7. This instability is not necessarily linked to the nature of the T_3 loop, as periods during which there were no interactions between the T_2A loop residues occurred frequently during the simulations. The greater length of the T_4 to T_6 loops enabled the residues to point inward toward the G-quartet, thus forming stable hydrogen-bonding interactions, resulting in a lower rms deviation for the T_4 loop. The thymine residues were also able to stack with each other, thus further stabilizing the loop. The T_5 and T_6 loops had higher rms deviations as some residues were more exposed to the solvent and unable to form stabilizing interactions with the G-quartet (Figures 5 and 7).

Molecular Dynamics Simulations of Single-Loop Variants in Antiparallel Conformations. The single-loop variants also formed stable antiparallel quadruplexes. The T and T_2 loops were too short to span the diagonal of the G-quartets in the 143D antiparallel structure, as shown by the distortion of the TTA-G_3T_2 143D antiparallel quadruplex after 2 ns of simulation (Figure 5). One of the G strands was displaced from the G-tetrad plane, and lost hydrogen-bonding interactions with the adjacent G residues. This affected all three G-tetrads. As a consequence, the K^+ ions moved from between to within the G-quartet planes. Both TTA-G_3T_1 and TTA-G_3T_2 were therefore simulated using the 186D template, yielding stable and undistorted structures

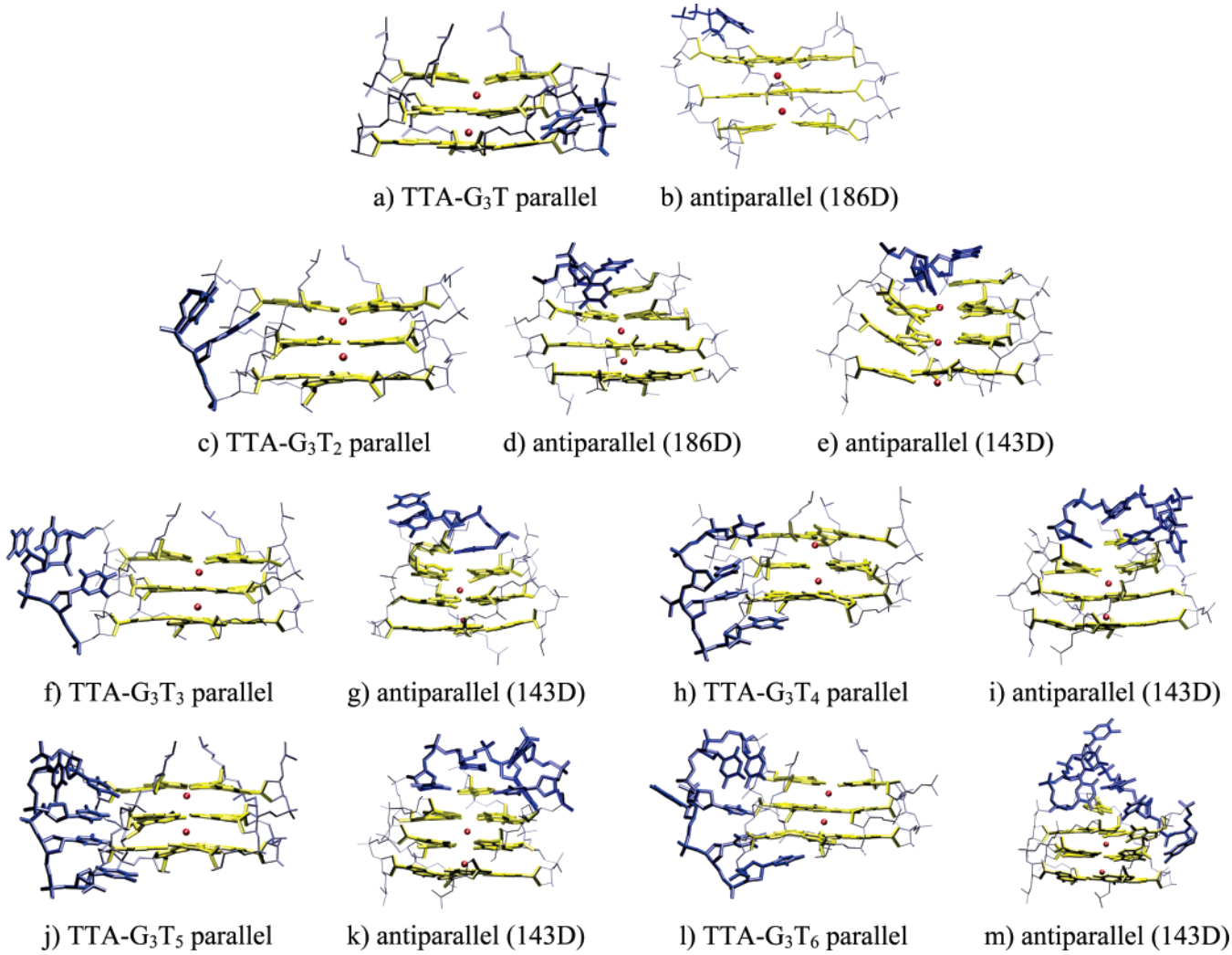

Figure 5. Structures of each single-loop variant sequence simulated. The structures were averaged over the last 2 ns of dynamics. For clarity only the G-quartets (yellow) and modified loop (blue) are shown. The K^+ ions within the central channel are shown in red. The 143D antiparallel structures also show the terminal A residue.

over 4 ns. The T_1 loop residue was able to stack on top of one G, and form a stabilizing hydrogen bond with the terminal A residue. One of the T_2 loop residues stacked on top of the terminal A residue, but in this case no hydrogen bonds could form. Both loops were stable, as shown by the low rms deviations in Table 3. The longer loop lengths formed stable quadruplexes in the 143D antiparallel conformation. The distorted 143D T_2 quadruplex retained three K^+ ions within its channel during the 2 ns of simulation (Figure 5). However, all the other antiparallel quadruplexes retained only two ions within the core channel, the K^+ ion initially positioned in the lateral loop region drifting into the solvent after a few hundred picoseconds at most.

In contrast to the parallel quadruplexes, the lateral T_2A loops were very stable during the simulations. Loop 1 had a high initial rms deviation, but quickly stabilized to a conformation common to all simulations (Table 3). The lateral T_2A loop residues formed a base triplet during the initial stages of each simulation. This triplet stacked with the G-quartets throughout the simulations. A greater range of rms deviations was observed for the modified diagonal loops. Loop lengths of T_3 to T_5 gave stable conformations after initial deviations from the starting structure, as shown by Table 3 and Figure 7. The T_3 loop formed thesame

Table 3. Rms Deviations (Å) for the Parallel and Antiparallel Quadruplexes of Single-Loop Variants[a]

	initial structure[b]				2 ns av structure[c]
	G-quartet	T_2A loop 1	T_2A loop 3	loop 2	loop 2
			Parallel		
TTA-G_3T	1.3	6.1	3.6	1.2	0.5
TTA-G_3T_2	1.4	3.2	3.3	2.3	1.2
TTA-G_3T_3	1.4	3.2	3.7	3.5	2.2
TTA-G_3T_4	1.4	3.3	3.8	2.3	0.9
TTA-G_3T_5	1.5	2.4	3.1	3.6	2.0
TTA-G_3T_6	1.4	3.3	3.6	5.6	1.8
			Antiparallel (143D)		
TTA-G_3T_2[d]	1.7	3.9	2.1	2.2	0.8
TTA-G_3T_3	1.4	4.1	2.0	2.7	0.9
TTA-G_3T_4	1.5	4.2	1.9	2.2	1.1
TTA-G_3T_5	1.5	3.7	1.8	2.3	0.9
TTA-G_3T_6	1.6	4.3	2.3	4.5	2.0
			Antiparallel (186D)		
TTA-G_3T	1.1	4.7	4.1	1.8	0.7
TTA-G_3T_2	1.2	1.5	4.2	1.5	1.2

[a] All loop rms deviations were calculated after the G-quartet regions only were overlapped. [b] The starting structure for the simulations was used as a reference for the rmsd calculations. [c] The average structure over the last 2 ns of simulation was used as a reference, and rmsd values were calculated over the last 2 ns only. [d] Values are over a 2 ns simulation.

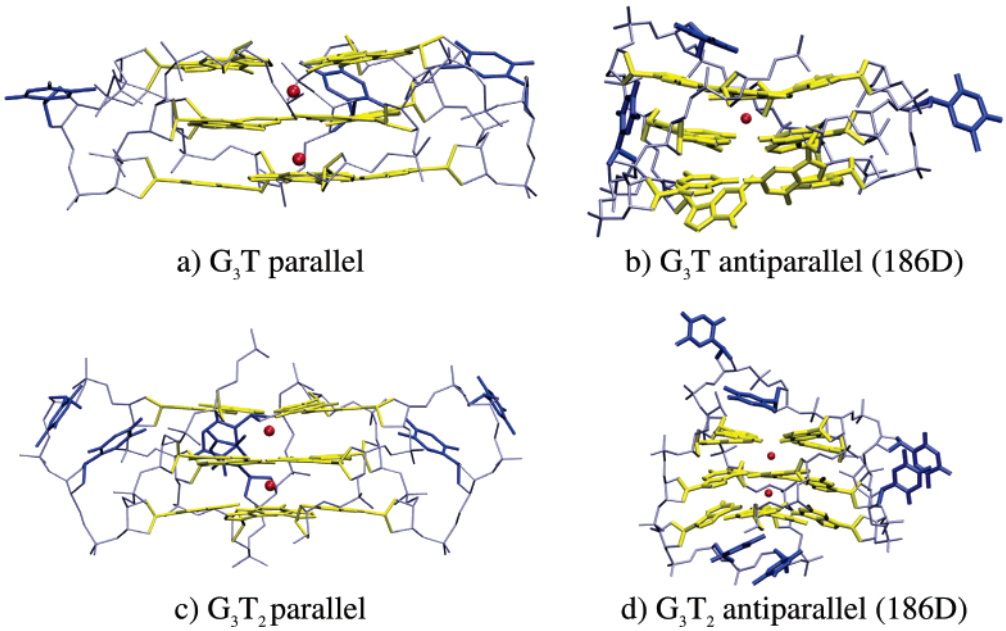

Figure 6. Structures of the multiple-loop variants. Structures were averaged over the last 2 ns of simulation. The color coding used is as in Figure 5.

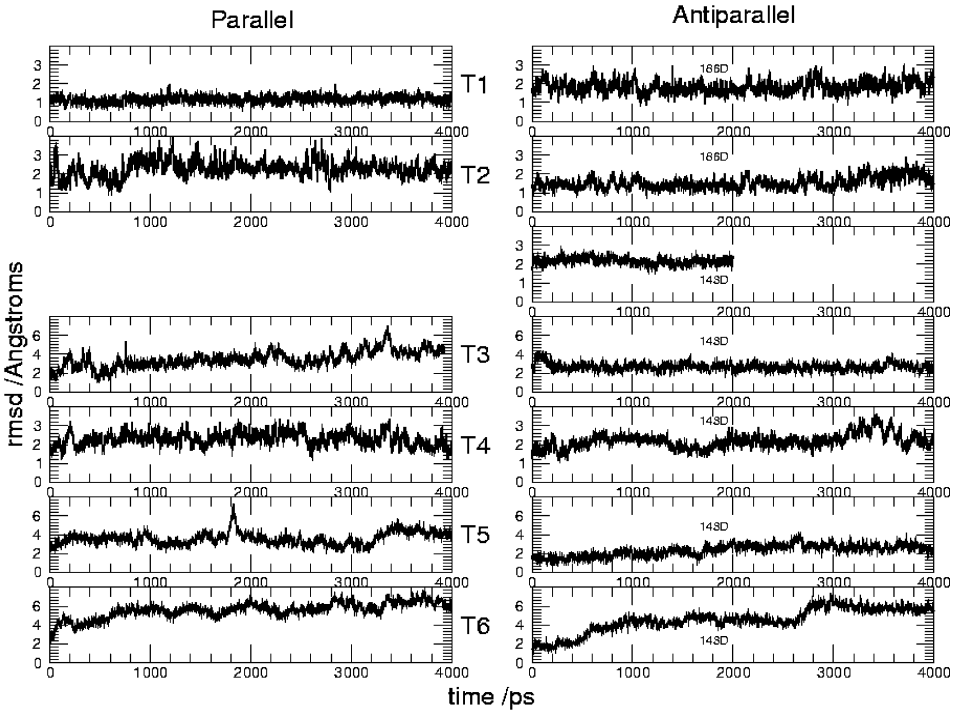

Figure 7. Modified loop rms deviations for the single-loop variants.

interactions as the T_2A diagonal loop in the NMR structure, with the first T being the furthest into solution, the second T stacking below on top of the terminal A, and the third T both stacking on the G-quartet and forming hydrogen bonds with the adenine (Figure 5). The first and last residues from the T_4 loop stacked on the G-quartets, with the second stacking on top of the triplet formed by the T and terminal A residues. The third T loop residue was stably located within the groove of the loop. The T_5 loop conformation is very similar, with one extra T stacking on top of the G-quartets. Only the T_6 loop had a higher flexibility. Stacking was still observed between the loop bases, but was more subject to fluctuations than in the shorter loop simulations.

Molecular Dynamics Simulations of Multiple-Loop Variants. To further correlate the simulations with experimental data, multiple T_1 and T_2 loop variants were also simulated. Their average structures are shown in Figure 6. Rms deviations for the G-quartets and loops are shown in Table 4. The single- and multiple-loop variants differ in that TTA-G_3T is the only single-loop variant quadruplex that shows a characteristically parallel CD spectrum, whereas both the G_3T and G_3T_2 multiple-loop variants appear to have some parallel character (with a peak around 265 nm). MD simulations of parallel and antiparallel single- and multiple-loop variants showed that only the G_3T sequence precluded an antiparallel fold. The simulation of this structure resulted in a large distortion of the G-quartets, with

Table 4. Rms Deviations (Å) for the Parallel and Antiparallel Quadruplexes of Multiple-Loop Variants from the Initial Structure[a]

	G-quartet	loop 1	loop 2	loop 3
		Parallel		
G_3T	1.3	3.6	3.0	1.6
G_3T_2	1.2	2.9	3.2	3.5
		Antiparallel (186D)		
G_3T	1.8	3.7	2.0	4.4
G_3T_2	1.5	2.2	3.9	5.5

[a] All loop rms deviations were calculated after the G-quartet regions only were overlapped.

Table 5. Free Energy Differences (kcal mol^{-1}) between Antiparallel and Parallel Quadruplexes (Antiparallel − Parallel)[a]

	ΔH	$-T\Delta S$	ΔG
TTA-G_3T	−12(2)	+8(1)	−4(2)
TTA-G_3T_2 (143D)	+16(2)	+11(1)	+27(2)
TTA-G_3T_2 (186D)	−11(2)	+10(1)	−1(2)
TTA-G_3T_3	−14(3)	+8(1)	−6(3)
TTA-G_3T_4	−19(2)	+7(1)	−12(2)
TTA-G_3T_5	−12(2)	+7(1)	−5(2)
TTA-G_3T_6	−25(2)	+14(1)	−11(2)
G_3T	−8(2)	+6(1)	−2(2)
G_3T_2	−4(2)	+6(1)	+2(2)

[a] Standard errors of the mean are indicated in parentheses.

only one K^+ ion remaining within the core quadruplex channel as a result. It must also be borne in mind that the antiparallel quadruplex selected for the simulations was based on the 186D PDB structure, which contains a mixture of parallel and antiparallel strands. The distortion caused to an antiparallel quadruplex with three lateral T loops would probably be much greater. On the other hand, the TTA-G_3T antiparallel quadruplex was stable over the 4 ns simulation, confirming that a short lateral T loop is possible in an antiparallel quadruplex, as long as there is only one present. Figure 6d shows that the antiparallel G_3T_2 quadruplex lower G-quartet was slightly distorted (curved), although to a much lesser extent than antiparallel G_3T (Figure 6b). The hydrogen bonds between G residues in the G_3T_2 quadruplex were still maintained, and the channel K^+ ions were unaffected. This suggests that the CD spectrum of the G_3T_2 quadruplex could be due to a mixture of both conformations.

Free Energy Calculations. The free energy differences between parallel and antiparallel quadruplexes of the same loop lengths are shown in Table 5. The free energies were averaged over the last 3 ns of simulation. A time scale of 4 ns was generally found to be sufficient to obtain stable enthalpic contributions to the free energy. The qualitative values obtained show that antiparallel quadruplexes are generally favored, although the free energy differences are relatively small and often within error margins. The entropic contribution to the free energy differences was found to be nonnegligible as it is almost equal to the enthalpic contribution.

Discussion

Loop lengths play a major role in the stability of quadruplexes, as shown by the substantial decrease in melting points as loop lengths are increased for the multiple-loop variants. The G_3T sequence had the highest UV melting temperature, in accordance with results obtained on a closely related sequence by Jing et al.[52] This is in contrast with melting studies carried out on d($G_4T_4G_4$) and d($G_4T_3G_4$) by Balagurumoorthy,[53] in which the shorter loop length was found to destabilize a dimeric quadruplex. The destabilizing effect is much less apparent when only a single loop is varied, in which case all sequences have very similar melting points. This suggests that the TTA loops are able to compensate for one destabilizing longer loop in the quadruplex.

Comparison of Simulations and Experimental Results for T and T_2 Loop Structures. For quadruplexes with short loop lengths, the simulations correlated well with experimental CD data. The results suggest that short single T loops can be tolerated in antiparallel quadruplexes, as long as there is only one, balanced by other longer loops. The G_3T quadruplex is only able to fold in a parallel conformation, the antiparallel structure being distorted (Figure 6b), whereas the TTA-G_3T quadruplex can fold in both parallel and antiparallel conformations (with a lateral, rather than diagonal, T loop). This was shown both in the simulations and experimentally from the absence of a peak around 295 nm in the G_3T CD spectrum. On the other hand, the TTA-G_3T CD spectrum has peaks around both 260 and 295 nm, and simulations show that both parallel and antiparallel conformations are stable. The 186D antiparallel structure has both parallel and antiparallel loops, and would be expected to give peaks in the CD spectra at both 260 and 295 nm.[19] As this is a steric effect, due to the short span of a single-nucleotide backbone, we expect the above to be the case for any single-residue loop. A slightly longer loop length of two nucleotides enables both parallel and antiparallel conformations to form, whether one or all three loops are modified. The parallel TTA-G_3T_2 and G_3T_2 quadruplexes were both stable over the 4 ns simulations. Loop rms deviations were similar for both quadruplexes; however, the G-quartets were slightly more stable in the G_3T_2 structure. The antiparallel TTA-G_3T_2 and G_3T_2 quadruplexes were also stable during the simulations, as long as the short T_2 loops adopted a lateral, rather than diagonal, geometry. However, the G_3T_2 average structure showed some distortion of the G-quartets (Figure 6d), although hydrogen bonding and stacking of the guanine bases were maintained. This was reflected in a higher rms deviation of the G-quartets for the antiparallel G_3T_2 structure compared to the antiparallel TTA-G_3T_2 structure (1.5 and 1.2 Å, respectively). Although shortening the loops leads to an increased stability (lower rms deviations) of the parallel structures (for both T and T_2 loops), the opposite is true for antiparallel loops. G_3T_2 can form an antiparallel quadruplex, but the simulations show that it is probably more likely to form a parallel structure. These results predict that structures such as that formed by part of the *c-myc* promoter sequence, d($AG_3TG_3GAG_3TG_3$), with potentially two T loops and a GA loop, would be expected to favor a parallel conformation, in accord with the recently revised topology for this quadruplex.[18,19]

The results of the free energy calculations did not provide significant insight into the stability of the T and T_2 loop sequences. The parallel and antiparallel structures had very similar free energies, as shown in Table 5. More unambiguously, however, the 143D antiparallel TTA-G_3T_2 quadruplex was largely disfavored compared to the parallel ($\Delta G = +27$ kcal mol^{-1}) and 186D antiparallel ($\Delta G = +28$ kcal mol^{-1}) conformations. This was due to very high distortion of the G-quartets to accommodate the short diagonal T_2 loop. Such large free energy differences were not found for other short loop sequences. Both T and T_2 loops formed parallel and antiparallel quadruplexes

of very similar free energies, indicating that all the structures could potentially form in solution. Surprisingly, the antiparallel G_3T quadruplex was slightly more stable than the parallel structure, despite the distortion of one of the G-quartets, and only one K^+ remaining within the channel. This result does cast doubt about the ability of such free energy calculations to measure the relative stabilities of structurally diverse quadruplexes. The small ΔG values obtained could indicate that the structures are very close in energy, and are likely to coexist in solution.

Comparison of Simulations and Experimental Results for T_3 to T_6 Loop Structures. Structures with longer loop lengths of both the single- and multiple-loop variants gave characteristic antiparallel conformation CD traces (Figures 3 and 4). However, these structures are more difficult to compare on the basis of only molecular dynamics simulations, as both parallel and antiparallel conformations yielded stable trajectories. The loop residues in antiparallel structures did generally form more numerous and stable interactions between themselves and with the G-quartets. Due to the geometry of strand-reversal loops, interactions between the loop residues and the G-quartets are more difficult to form. Free energy calculations did show a preference for antiparallel over parallel conformations, consistent with the CD spectroscopic analysis, although the small free energy differences must be interpreted with caution (Table 5). The antiparallel quadruplexes were found to be between 1 and 12 kcal mol^{-1} more favorable than the parallel quadruplexes. Small values indicate that both conformations could coexist in solution, even in the longer loop length cases. This is not observed in CD spectra, where for loop lengths of three or more, no characteristic parallel quadruplex peaks are observed.

Importance of Interactions within the Loops. In general, the MM-PBSA calculations showed that the enthalpic contribution favors antiparallel quadruplex structures, whereas the entropic contribution favors parallel structures. The antiparallel structures formed a greater number of stable interactions within the loops and between the loops and G-quartets, which could explain the more favorable enthalpic contribution. In all three loops, the residues could stack with each other and with the G bases, and long-lived hydrogen bonds were formed between loop residues. Stacking interactions between loop residues in the parallel quadruplexes were generally observed, although the residues involved in these interactions varied during the simulations, and individual interactions were relatively short-lived. The multiple-loop variants had smaller ΔH values, indicating that there were fewer differences in loop interactions between the parallel and antiparallel conformations. Smaller loops are therefore not as enthalpically favored in antiparallel quadruplexes as longer loops are. On the other hand, entropic factors are more significant in parallel structures, in which loop mobility is greater, compared to antiparallel structures. Generally, the rms deviations of loops with respect to the average structure are larger for parallel than antiparallel quadruplexes (see Table 3 for loop 2 rms deviations).

Interactions within the loops probably play a major part in the preferential antiparallel folding of quadruplexes with loop lengths of three or more. In the 143D antiparallel quadruplexes, it was shown that the two T_2A lateral loops can form a triplet which stacks with the G-quartets, in addition to stacking of the diagonal loop residues. Fewer stabilizing interactions are possible within the parallel conformation. When the loop residues are removed and replaced by nonnucleosidic linkers, as shown recently,[24] the CD spectra of unimolecular quadruplexes exhibit characteristic parallel conformation peaks. However, the use of loops in the present work favors the formation of antiparallel structures. If several quadruplex structures do form in solution, the antiparallel quadruplex is probably the thermodynamically favored conformation, especially for longer loops. Fox also showed that a quadruplex sequence with the G_3T repeat was too stable for the melting point to be measured, and that a quadruplex with G_3T_2 repeats had a biphasic curve in K^+ solution.[24] This agrees with our simulations in which G_3T is only stable in a parallel conformation, whereas G_3T_2 can fold in both parallel and antiparallel conformations. It has been shown that both parallel and antiparallel conformations of a related sequence, d(TAGGGTTAGGGT), have different rates of folding and unfolding,[37] but both coexist in K^+ or Na^+ solution, and the free energy difference between the two forms is relatively small ΔG (<1 kcal mol^{-1}).[38] This strongly suggests that, even though the simulations have all been on K^+ forms, the results have applicability to species in Na^+ solution as well (a thorough exploration of an equivalent set of simulations using Na^+ instead of K^+ ions is outside the scope of the present study).

Parallel quadruplexes of the sequences in this study can all form, and are observed[24] when there are no loop residues. The interactions between loop residues and the G-quartets might be one of the factors that drive the equilibrium toward formation of antiparallel structures. This could explain the dominance of antiparallel conformations in the CD spectra, especially as the loop lengths are increased. We initially assumed that the loops are key elements in the stability of particular G-quadruplexes; comparing our present results with the behavior of non-residue-containing loops[24] shows that the loops and not the G-quartets are the determinant of the dominant fold of particular sequences. One caveat has to be that the size of the G-quartet core itself may alter folding preferences, as well as loop conformations, especially in the case of parallel quadruplexes.

Limitations of CD Spectra. The results obtained in this study highlight the dangers of overreliance solely on CD spectra to determine the structure of a particular quadruplex. For example, the unmodified d(G_3T_2A) sequence has a characteristic antiparallel CD spectrum (Figure 4), despite evidence that the parallel conformation is also present in K^+ solution.[37] The simulations show that both parallel and antiparallel structures can form for most loop lengths, and dismissing a conformation only on CD data is most likely incorrect. Spectra in which peaks around both 265 and 295 nm are obtained could indicate the coexistence of parallel and antiparallel species in solution, or the presence of one conformation with both parallel and antiparallel guanine strands (such as the 186D structure).

Limitations of the Computational Methods. The reliability of free energy calculations depends on several factors, including the extent of conformational sampling and the theoretical methods used in the calculations, as well as the quality of the parametrization, especially the assigned ionic radii of metal ions. This is true for both the simulated annealing runs and the MD simulations carried out. Due to the high flexibility of the quadruplex loops, especially in the parallel quadruplexes, and increasing loop lengths, much longer simulations might be required to sufficiently sample their configurational space.

Additive force fields may not be sufficient to describe structures such as quadruplexes in which ions play an integral part. A polarizable force field would be more appropriate to describe, in particular, the ion channel in these structures and its interactions with the guanine tetrads. Furthermore, the small free energy differences obtained indicate that both parallel and antiparallel conformations are energetically very similar, requiring very accurate methods to calculate ΔG. The free energy differences calculated should thus be viewed as qualitative ranking values, the entropic term being the least reliable contribution. Studies on protein folding have previously shown that folded and misfolded proteins have very similar entropic contributions,[49] although the latter are often ignored when the free energies of variously folded structures are compared. The values in Table 5 show that entropy cannot be neglected in the case of quadruplex folding, as the $-T\Delta S$ contributions are nearly equal to ΔH for different conformations of the same sequence. However, the fact that the errors were small for all the snapshots for each conformation indicates that different potential energy wells sampled during the trajectory had similar entropy values.

Due to the structural polymorphism of G-quadruplexes, it was beyond the scope of this study to simulate all possible antiparallel structures that can be adopted by these quadruplex-forming sequences. The two templates were chosen due to the availability of experimental structures for them, which did not then require changes to the G-quartets. Simulations tend to be very dependent on the starting structures used, and changes were therefore limited to the loops. For each loop length we endeavored to find a stable conformation, and not to discriminate between all the possible loop structures. The free energy calculations were carried out for a single antiparallel structure, although of course other potentially more favorable conformations cannot be ruled out.

Recently, the ability of current force fields to accurately represent G-quadruplex loops has been questioned.[56] The present work did not aim to find the "best" loop geometries for each of the quadruplexes studied, but only to show that a certain quadruplex could potentially form. Stable loop conformations were not obtained for several of the longer loop variants in this study, as shown in Figure 7. However, the simulations showed that the G-quartets were stable despite this. The distortions of certain short-loop G-quartets are expected to reflect genuine inabilities of these folds to form, rather than shortcomings of the force field.

Conclusions

We have shown that a combination of experimental and molecular modeling methods can yield complementary and overall consistent information about the structures of G-quadruplexes with varying loop lengths. All sequences studied, with loops ranging from T to T_7, were shown to form quadruplexes in K^+ solution. CD spectra suggest that parallel quadruplexes are only present when the loops are too short to accommodate an antiparallel conformation. G_3T was the only sequence studied which could only exist as a parallel quadruplex, due to steric constraints of the short-loop backbone. Since this is a steric factor, the same is very likely to be the case with single-residue loops of any nature. When only one T loop was present, both parallel and antiparallel quadruplexes were stable. Simulations showed that parallel conformations are stable for all of the sequences studied, and are likely to be very close in energy to antiparallel conformations. Assuming that parallel structures are present in solution only when characteristic peaks are observed in CD spectra could be misleading. The following initial rules for quadruplex folding can therefore be drawn up on the basis of the results obtained from these simulations and experimental data: (1) Structures with three T_1 loops can only form parallel structures. (2) Structures with three T_2 loops can form both parallel and antiparallel structures; however, parallel conformations are likely to be favored. (3) A single T_1 loop enables both parallel and antiparallel structures to be formed, but the parallel structure is more favorable. (4) Structures with a single T_2 to T_6 loop can also exist as parallel or antiparallel conformations; however, the antiparallel structure is likely to be preferred.

Melting studies showed that shorter loop lengths are energetically favored, with G_3T having the highest melting point. The free energy calculations are only relative values, and cannot currently be used to compare quadruplexes of different loop lengths to assess their stability relative to the number of loop residues. However, the simulations do show that longer loops have higher rms deviations, and fewer long-lived interactions.

Acknowledgment. We are grateful to Cancer Research UK (Program Grant C129/A4945 to S.N. and Project Grant C129/A3567 to S.B. and S.N.), the International Association for Cancer Research (studentship to P.H.), the Biotechnology and Biological Sciences Research Council (studentship to J.H.), and a number of colleagues for useful discussions.

Supporting Information Available: All starting and final averaged quadruplex structures described herein (34 Brookhaven Protein Databank files). This material is available free of charge via the Internet at http://pubs.acs.org.

JA045154J

(56) Fadrna, E.; Spackova, N.; Stefl, R.; Koca, J.; Cheatham, T. E., III; Sponer, J. *Biophys. J.* **2004**, *87*, 227−242.