Predictive modelling of topology and loop variations in dimeric DNA quadruplex structures

Pascale Hazel, Gary N Parkinson and Stephen Neidle*

CRUK Biomolecular Structure Group, The School of Pharmacy, University of London, 29-39 Brunswick Square, London WC1N 1AX, UK

ABSTRACT

We have used a combination of simulated annealing (SA), molecular dynamics (MD) and locally enhanced sampling (LES) methods in order to predict the favourable topologies and loop conformations of dimeric DNA quadruplexes with T2 or T3 loops. This follows on from our previous MD simulation studies on the influence of loop lengths on the topology of intramolecular quadruplex structures [P. Hazel *et al.* (2004) *J. Am. Chem. Soc.*, 126, 16 405–16 415], which provided results consistent with biophysical data. The recent crystal structures of d(G4T3G4)2 and d(G4BrUT2G4) (P. Hazel *et al.* (2006) *J. Am. Chem. Soc.*, in press) and the NMR-determined topology of d(TG4T2G4T)2 [A.T. Phan *et al.* (2004) *J. Mol. Biol.*, 338, 93–102] have been used in the present study for comparison with simulation results. These together with MM-PBSA free-energy calculations indicate that lateral T3 loops are favoured over diagonal loops, in accordance with the experimental structures; however, distinct loop conformations have been predicted to be favoured compared to those found experimentally. Several lateral and diagonal loop conformations have been found to be similar in energy. The simulations suggest an explanation for the distinct patterns of observed dimer topology for sequences with T3 and T2 loops, which depend on the loop lengths, rather than only on G-quartet stability.

INTRODUCTION

The increasing number of DNA G-quadruplex structures determined by X-ray and NMR methods is revealing the high degree of structural plurality available to quadruplex-forming sequences. Parallel or antiparallel dimeric and monomeric quadruplexes have been shown to form, with loops being either parallel (1–4), lateral (5–8) or diagonal (9–15). Structures with a mixture of loop types and parallel/antiparallel strands are also common (16–19). Potential G-quadruplex-forming G-rich sequences are found in the telomeric regions of chromosomes, in which the extreme 3′ end is single stranded. G-quadruplex structures formed from various numbers of T_2AG_3 (1,2), T_2G_4 (6,18) and T_4G_4 (9,10,20) telomeric repeats from several organisms have been the focus of studies using crystallography and NMR, as well as several biophysical techniques. In addition there is increasing evidence that the formation of G-quadruplexes is not limited to telomeric regions, but may occur in G-rich regions throughout the genome, either naturally or induced through the binding of small molecules. This is exemplified by a stable G-quadruplex formed by a G-rich sequence found in the nuclease hypersensitivity element (NHE) III1 in a promoter of the *c-myc* oncogene; this G-quadruplex has been suggested to be involved in transcriptional regulation of the *c-myc* gene (21,22). Many potential quadruplex-forming G-rich sequences can be found in the human and other genomes. They are much more varied than the tandem telomeric repeat sequences above, and contain differing G and loop residue length/sequence combinations (23–27). The development of drugs to specifically target non-telomeric G-quadruplexes will be aided by knowledge of the diverse structures formed by different G-quadruplex sequences (27–29).

In order to understand the factors governing the sequence-dependent folding of G-quadruplexes, we have examined the effect of loop length on intramolecular quadruplex structure (30). We were able to show, using a combined molecular dynamics (MD) and biophysical methods approach, that loop length in certain situations is a determinant of G-quadruplex structure, and specifically that sequences with three single-nucleotide loops restrict intramolecular quadruplexes to the formation of parallel structures, whereas one single-nucleotide loop together with two longer loops can potentially form both parallel and antiparallel structures. Parallel quadruplex structures have now been found in several sequences containing single-nucleotide loops (25,26),

*To whom correspondence should be addressed. Tel: 44 207 753 5969; Fax: 44 207 753 5970; Email: Stephen.neidle@pharmacy.ac.uk

including those derived from the *c-myc* promoter G-rich sequence (3,4).

MD simulation methods are commonly used to investigate G-quadruplex structures (31–34). The folding pathway of the tetrameric G-quadruplex $d(G_4)_4$ has been studied using MD simulations, and possible folding intermediates were suggested (33). Free energy calculations using the MM-PBSA method showed that these intermediates were less stable than the final folded structure. Detailed studies of the T_4 loop conformation were also carried out, however these have been less successful (34). These were unable to predict the experimentally observed diagonal loop conformation, and simulations of the experimental loop conformation were, moreover, unstable. A concern when studying the T_4 loop in a quadruplex structure is the presence, as observed in the crystal structure (9), of a loop-bound K^+ ion, which is not stable during MD simulations (34). This ion was not found in an NMR study (11), although a more recent study (35) using thallium ions (with a very similar ionic radius to potassium), suggests that an ion may be present in the loop. The solution structure (11) in K^+ adopts a somewhat different loop geometry. However this conformation was not predicted. MD simulations are known to suffer from limitations when these type of ionic effects are involved (31).

We have recently solved (P. Hazel *et al.* (2006) *J. Am. Chem. Soc.*, in press) the crystal structures of dimeric quadruplexes formed from the sequences $d(G_4T_3G_4)_2$ and $d(G_4^{Br}UT_2G_4)$, for comparison with the well-established structure of $d(G_4T_4G_4)_2$ (9). This last sequence forms an antiparallel dimeric quadruplex structure, with two diagonal T_4 loops, both in solution (10–12) and in the crystal (9,36). In contrast, the crystal structures of the sequences $d(G_4T_3G_4)_2$ and $d(G_4^{Br}UT_2G_4)_2$ show that they form dimeric quadruplexes with lateral loops, either on opposite sides of the G-quartets (head-to-tail dimer), or on adjacent sides (a head-to-head dimer quadruplex structure was found for the $d(G_4^{Br}UT_2G_4)$ sequence). In parallel with these crystallographic studies, we have attempted to predict the structure adopted by the sequences containing T_3 loops, using MD simulations. The crystal structures have also enabled the accuracy of the predicted structures to be assessed. Our crystallization trials with the $d(G_4T_2G_4)$ sequence have not yielded any crystals to date; however, simulation results on the quadruplexes potentially

available to this sequence have been compared with the NMR structure of the closely related $d(TG_4T_2G_4T)_2$ quadruplex (6).

We report here on the use of a combination of *in vacuo* simulated annealing (SA) and explicit solvent MD simulations with locally enhanced sampling (LES) (37,38) to generate favourable T_2 and T_3 loop conformations for dimeric quadruplexes. MM-PBSA free-energy calculations (39,40) have been used to compare the resulting different loop structures. This post-processing method allows absolute free energies to be estimated from snapshots obtained during MD simulations.

MATERIALS AND METHODS

The *Oxytricha nova* $d(G_4T_4G_4)_2$ crystal structure (9) was used as a template for diagonal loop quadruplexes (PDB code 1JPQ). Lateral loop templates were generated from a previous, though incorrect, $d(G_4T_4G_4)_2$ crystal structure (PDB code 1D59) (41), as this had the desired topology, and the $d(GCGGT_3GCGG)$ NMR structure (PDB code 1A6H) (42). The latter has two mixed G-C quartets in the centre of the quadruplex. The C residues were replaced with G, and the quartet stem equilibrated for 2 ns to relieve strain. Both lateral loop templates (1D59 and 1A6H) have alternating *syn-anti* glycosidic angles around the G-quartets, however the 1A6H template has *syn-syn-anti-anti* conformations down each G-strand, compared to *syn-anti-syn-anti* ones for the 1D59 template. The 1D59 template structure was used both unaltered, and subsequent to a 2 ns MD equilibration, as the crystal structure itself is in a high-energy conformation (33). A model-built antiparallel G-quadruplex stem was also included, containing two wide and two narrow grooves, and alternating *syn-anti* glycosidic angles down the G-strands. Parallel structures were generated from the $d(TAG_3T_2AG_3T)_2$ crystal structure, (PDB code 1K8P) (1) in which a fourth G-quartet was added to the stem after removal of the loops. The resulting structure was minimized to relieve strain in the backbone. The $d(G_4T_3G_4)_2$ crystal structures (PDB id codes 2AVH and 2AVJ: Hazel *et al.*, manuscript submitted) were also used as templates for the simulations of quadruplexes containing T_2 loops. Schematic diagrams of the parallel, lateral and diagonal loop dimeric quadruplexes are shown in Figure 1.
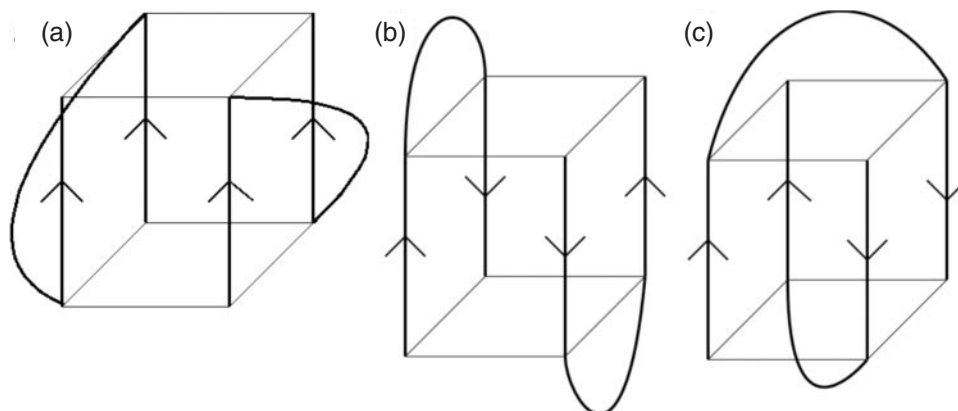
Figure 1. Schematic diagrams of the (**a**) parallel, (**b**) lateral and (**c**) diagonal loop dimeric G-quadruplexes which were simulated in this work.

All the initial model building and structural modifications were carried out with the Insight II suite of programs (Molecular Simulations Inc., San Diego, CA). The T_4 loops in the 1JPQ and 1D59 template structures were replaced with T_3 or T_2 loops, and the backbones were minimized to relieve strain. The 1A6H template already contains a T_3 loop, which was either kept as a starting structure, or modified to a T_2 one. Loop conformational space was searched with SA procedures using the discover module of Insight II. During all SA runs the residues involved in the G-quartets were kept fixed, and only the loop residues were allowed to move. The SA runs were carried out in implicit solvent, using a distance-dependent dielectric ($\epsilon = 4r$) to mimic the solvent. The initial loop conformation was minimized, then during each cycle the loop was first heated to 1000 K over 2 ps, simulated at 1000 K for 2 ps, cooled to 300 K for 1 ps and finally minimized. The next loop conformation was generated from heating of the latest minimized conformation.

The structures obtained from the SA runs were clustered into conformational families, according to root mean square (r.m.s.) deviation calculations between all structure pairs. Pairwise r.m.s. deviations between all the structures were calculated, then clustered according to the method used by the NMRCLUST program (43). NMRCLUST was designed to cluster NMR structures, and is therefore not able to handle the large number of structures generated here in each SA run. A Python script was written, which uses the same clustering methodology, but with an unlimited number of input structures. Moreover, it was written to directly read in the output files from the Insight SA runs, and output Insight format archive files for each cluster. This script can be downloaded from http://www.pascalehazel.org/cluster. Coordinates of the most frequently occurring loop conformations generated during SA runs are available as Supplementary Data. Bond and angle energetic contributions for the initial model loops were calculated using the Anal module of Amber 7 (44).

Selected structures from the clusters were subjected to more lengthy MD simulations in explicit solvent using the Amber 7 program. Three K^+ ions were placed, one between each G-quartet stack, equidistant from the eight G O6 atoms, when these were not present in the experimental template. Further solution K^+ ions were added to neutralize the system, which was then solvated in a pre-equilibrated box of TIP3P water. The box size depended on the system, but always extended at least 10 Å from the solute in every direction. The equilibration procedure consisted of 10 steps, beginning with 1000 steps of minimization and 25 ps of dynamics of the solvent only. The whole system was then minimized for 1000 steps, followed by 3 ps of dynamics with a restraint of 25 kcal.mol^{-1} on the DNA. The DNA restraint was lowered by 5 kcal.mol^{-1} during each of the next five 1000-step minimizations. Finally, the system was heated slowly to 300 K over 20 ps, with no further restraints. MD simulations were carried out at 300 K, using a 2 fs time step, with SHAKE applied to constrain the bonds containing hydrogen. The PME method was used to deal with long range electrostatic interactions, and Lennard–Jones interactions were cut off at 10 Å. Similar protocols were found to be reliable in previously reported simulations of G-quadruplexes (32,33).

LES simulations were carried out on a subset of loop conformations. After an equilibration period of between 500 ps and 1 ns of dynamics in explicit solvent, five copies of each loop were generated using the Addles module of Amber. Both the LES (loops) and non-LES regions (G-quartets) were maintained at 300 K, using separate water baths. LES simulations were carried out in explicit solvent.

The MM-PBSA method was used to calculate approximate free energies. Snapshots were collected every 20 ps for energetic analysis. The electrostatic contribution to the solvation free energy was calculated using the Delphi II program (BIOSYM., San Diego, CA). Dielectric constants of 1.0 and 80.0 were assigned to solute and solvent, respectively. A grid spacing of 0.5 Å was chosen, with the longest linear dimension of the molecule occupying 80% of this grid. The Amber parm99 charge set and BONDI radii were used (45). All MM-PBSA calculations explicitly included the three K^+ ions within the quadruplex channel. The K^+ radius was determined to be 2.025 Å, by adjusting it until ($\Delta G_{polar} + \Delta G_{nonpolar}$) was equal to the experimental $\Delta G_{solvation}$ of -80.6 kcal.mol^{-1}. All other energy terms were calculated with the programs distributed with Amber. The solute entropic contribution was estimated with the NMODE program, using snapshots collected every 200 ps. Each snapshot was minimized in the gas phase, using a distance-dependent dielectric of $\epsilon = 4r$, before calculation of the vibrational mode frequencies. The minimizations caused some distortion to the structures, however this did not have a significant effect on the entropies calculated. After LES simulations, the final loop copies were averaged, and non-LES dynamics were carried out for at least 1 ns. The MM-PBSA energies calculated can then be compared to pre-LES energies.

RESULTS

SA

The SA runs generated large numbers of structures, and only the most frequently repeated conformations were considered. Results from the clustering of lateral T_3 loop conformations over both the wide and narrow grooves are shown in Figures 2 and 3, respectively. L_w indicates a lateral loop over the wide quadruplex groove, as opposed to the narrow groove (L_n). PDB files of the most frequently obtained lateral and diagonal loop conformations are available as Supplementary Data. A large number of loop structures generated were very solvent exposed, with no stabilizing interactions between loop and G residues, possibly due to the implicit solvent approximation used during the simulations. However, these structures were generally structurally very different from one another, and did not appear in the final clusters created. Clusters containing fewer than 5% of the total number of conformations were not considered. Loop conformations containing stabilizing interactions, such as stacking or hydrogen bonding, were found to occur with greater frequencies than loop conformations with no such stabilizing interactions.

No difference in stability was found between the quadruplexes with different lateral and diagonal T_3 loop types. However, both diagonal and lateral (over the wide quadruplex groove) T_2 loops were found to be strained in the structures obtained from the SA runs. Diagonal loops have to span an average of 19.5 Å in the G-quadruplex models considered, and lateral (wide groove) loops span an average of 15.5 Å
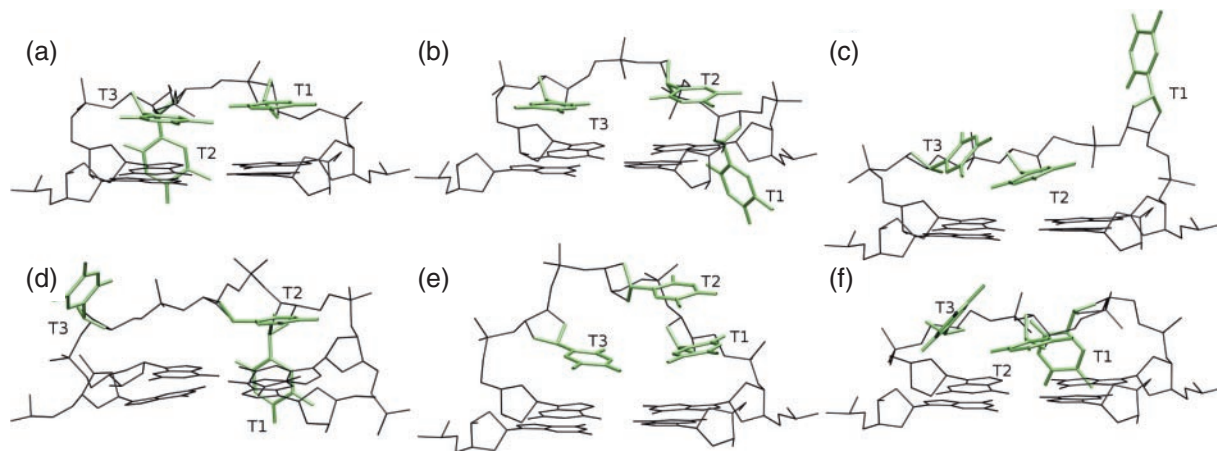
Figure 2. Structures with lateral T3 loops over the wide quadruplex groove, generated during SA runs. (**a**) T3-Lw-1, (**b**) T3-Lw-2, (**c**) T3-Lw-3, (**d**) T3-Lw-4, (**e**) T3-Lw-5 and (**f**) T3-Lw-6. The loop bases are shown in green and a G-quartet in black. Lw indicates a lateral loop over the wide quadruplex groove, as opposed to the narrow groove (Ln).
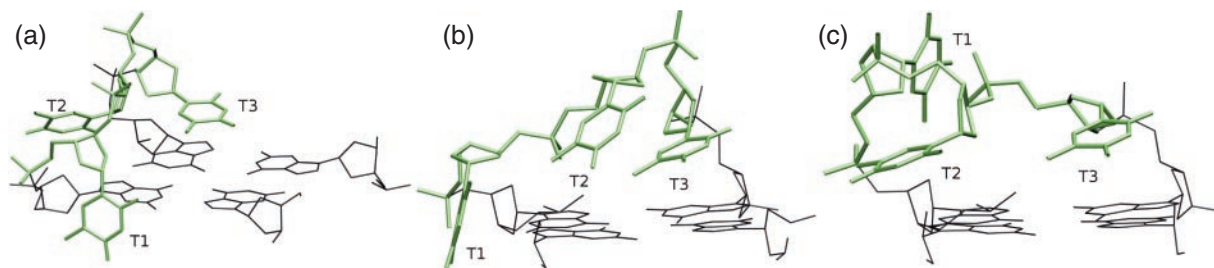
Figure 3. Structures with lateral T3 loops over the narrow quadruplex groove, generated using SA. (**a**) T3-Ln-1, (**b**) T3-Ln-2 and (**c**) T3-Ln-3.

(C4′ to C3′ distances across the groove). As only the loop residues were allowed to move during the conformational search, no effects of the short loop length could be observed on the G-quartets. However, the diagonal and lateral loops over the wider groove themselves appeared to be somewhat strained. This was apparent through the appearance of some increased bond lengths, e.g. P-O3′ or P-O5′ distances of 1.7 Å rather than around 1.6 Å, and was reflected in calculated backbone bond energy contributions. For example T_2 diagonal, lateral (wide groove) and lateral (narrow groove) loops had backbone bond energies of 11.1, 11.0 and 2.3 kcal.mol^{-1}, respectively. The T_2 short loops were much less conformationally flexible than the T_3 loops. Thus the lateral T_2 loop over the wide quadruplex groove formed only four different conformations, out of the 200 sampled. This limited number of possible conformations, together with their reduced flexibility suggests that their backbones are under strain. Lateral T_2 loops over the narrow quadruplex groove were much more flexible, and many possible conformations were obtained. These have to span an average C4′ to C3′ distance of 12.4 Å. In this case, the flexibility of the backbone suggests that the structures were not under strain. These SA results suggest that T_2 loops bridging distances of 15 Å and above are under strain.

Although SA methods in implicit solvent were useful to generate many possible loop conformations, the stability of the resulting conformations could not be assessed. Even if the number of SA runs was large enough to be able to use the number of times a structure appeared as an accurate indicator of stability, no comparisons could be made between the lateral and diagonal loop types. The total potential energies of each system can be calculated using Insight II; however, these are necessarily only approximate values. The calculations include solvent effects using a distance-dependent dielectric, which is only a crude approximation, and do not take entropy into account. Moreover, only the loop residues were allowed to move during the conformational search, meaning that the G-quartets could not respond to any pressure caused by strained loop conformations. In order to further assess the loop stability, fully solvated MD simulations were carried out on a subset of structures. Due to the computational cost of carrying out fully solvated MD simulations, only the most favourable structures from the SA runs were considered.

MD and LES simulations

MD simulations of T_3 and T_2 loop quadruplexes were carried out on a selected number of structures. The three most frequently occurring T_3 lateral loop conformations found for the alternating *syn-anti* model dimeric quadruplex, T_3-L_w-1, T_3-L_w-2 and T_3-L_n-2, were considered. In the T_3-L_w-1 conformation, the middle T residue is in the quadruplex groove, while the first and third T residues stack on the G-quartet plane (Figure 2a). A structure with two loops in the T_3-L_w-1 conformation at each end of the quadruplex showed no structural changes after 1 ns MD, apart from a slight movement of the middle T residues out of the quadruplex groove. A 4 ns LES simulation was carried out starting from the 1 ns equilibrated structure. The two loops in the same starting
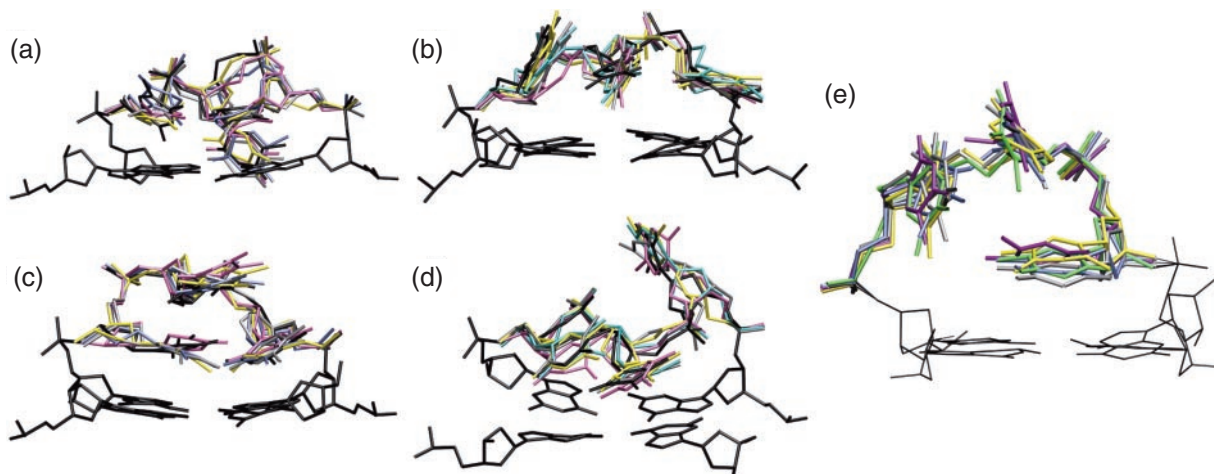
Figure 4. T_3 lateral loop final structures after LES simulations. (**a**) T_3-L_w-1 loop 1 and (**c**) 2 after 4 ns LES simulation, (**b**) T_3-L_w-2 loop 1 and (**d**) loop 2 after 2.7 ns LES simulation and (**e**) T_3-L_n-1 loop 1 after 4 ns LES simulation. The five LES copies are shown overlapped in each case. A G-quartet is also shown in black.

conformation behaved differently during the LES simulation. The first loop remained in the same conformation throughout the dynamics (Figure 4a). After about 2 ns, a structural rearrangement of the second loop occurred so that the second T residue, previously in the quadruplex groove, moved to stack on top of the two other T residues (Figure 4c). This happened gradually over a few 100 ps, and the second conformation was stable for the remainder of the LES run. The new loop conformation was similar to the T_3-L_w-5 structure in Figure 2e, sampled during SA. Lowering of the energy barriers enabled the third T residue to flip by 180° in both loops during LES simulations, allowing potential hydrogen bonds to form between the first and third T residues. However, the T residues were only within hydrogen-bonding distance in the second loop, and only after the conformational change to a T_3-L_w-5 type structure. During these simulations, each loop conformation was stable over nanosecond timescales, even when using the LES increased sampling method.

The SA runs generated several structures in which the first T residue was positioned in the quadruplex groove (T_3-L_w-2, T_3-L_w-4 and T_3-L_w-6 in Figure 2). A 1 ns MD simulation of the alternating *syn-anti* model quadruplex with two T_3-L_w-2 loop types suggested that the latter were less stable than the previously sampled T_3-L_w-1 and T_3-L_w-5 conformations. During the initial 1 ns equilibration phase, the loop residues were very flexible, especially the first and third T residues in the first loop. The first T residue moved gradually out of the quadruplex groove in the first few hundreds of picoseconds, and then remained exposed to the solvent (Figure 4b). The second loop in the same conformation was stable over the 1 ns MD simulation. During the subsequent 2.7 ns LES simulation, both loops rearranged to the conformations shown in Figure 4b and d. Stacking of the second and third T residues was conserved during the LES simulation, however the first T residue was much more flexible. The final structures obtained are similar to the T_3-L_w-3 conformation from the SA runs, in which the first T residue moves closer to stack with the other loop residues, thus minimizing solvent exposure.

Quadruplex structures with the native 1A6H T_3 loops were simulated for 4 ns after mutation of the C residues to G in the quadruplex stem. Multiple loop conformations are present in

the NMR structure, and the first PDB entry was chosen for MD. In this, the first T residue is in the quadruplex groove, the third T stacks on the G-quartets, and the second T is pointing into the solvent. Within the first 400 ps of dynamics, the second T residue of both loops formed stacking interactions with either the G-quartets, or the third T residue, and these were stable throughout the simulation. This is in accordance with the NMR structures, with some having both second and third T residues stacking with the G-quartets. The first T residue remained within the quadruplex groove in only one of the loops. In the second loop, this T residue moved into the solution, an arrangement corresponding to the second NMR structure in the 1A6H PDB entry. This simulation revealed that a loop with the first T residue in the quadruplex groove can be stable, although rearrangements to other conformations can also occur (in the 1A6H simulation as well as the T_3-L_w-2 simulation described above). The T_3 loops were flexible in the NMR structure, and interchange between the different experimental conformations was observed during the simulations. The final loop conformations obtained were similar to the T_3-L_w-2 and T_3-L_w-6 structures in Figure 2b and 2f. The loop flexibility within the NMR structure suggests that the conformational changes which occur during the simulations are due to real structural flexibility of the loops, rather than force field effects.

Lateral loops over the narrow quadruplex groove in the T_3-L_n-2 conformation were also simulated using MD. The loop conformation remained unchanged during a 1 ns equilibration, with two residues forming a stack over the G-quartets, and the first T residue slightly arranged within the quadruplex groove. The loops were more flexible over a 4 ns LES simulation, although no major rearrangements were observed (Figure 4e). The individual loop copy r.m.s. deviations were around 3.0 Å, after fitting of the G-quartets; however, the final structure was similar to the starting T_3-L_n-2 conformation. As with several other simulations, the first T residue, originally within the quadruplex groove, adopted a slightly elevated position, in order to form interactions with the other two loop residues.

A quadruplex with two diagonal T_3 loops was subjected to a 4 ns MD simulation. The two loops were found to behave

differently during the dynamics. One of the loops remained in the same conformation throughout the dynamics, with the first T residue bound deep within the quadruplex groove (Figure 5a). On the other hand, the first T residue of the second loop moved in and out of the quadruplex groove during the dynamics. The motions of this latter loop also caused a G-quartet to become distorted during the final few 100 ps of simulation, when a G residue left the quartet plane, and formed a hydrogen bond with a loop phosphate group oxygen atom of the loop backbone (Figure 5b). This shows the importance of carrying out lengthy simulations, as the loop conformation was stable for almost 4 ns before showing any signs of distortion. This also revealed that simulations can be sensitive to small changes in loop conformation, as the first loop in essentially the same conformation was stable during the whole simulation.

Most sampled loop conformations were stable over nanosecond timescales, however interchange between different loop conformations was observed. Similar loop conformations exhibited different behaviour within the same simulation,
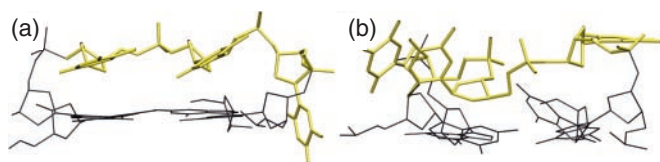
Figure 5. T_3 diagonal loop MD simulation. (**a**) Stable loop conformation averaged over the final 2 to 4 ns of MD and (**b**) final structure of the unstable loop, which formed within the last 200 ps of the simulation.

suggesting that several T_3 loop conformations are equally possible. Unstable loops did not generally affect the stability of the G-quadruplex stem, which always had lower r.m.s. deviations. These results are encouraging in that there was generally a good agreement between the conformations found in the SA runs and during the dynamics. Conformations adopted during the MD simulations were generally structurally close to structures previously generated in the SA runs. The diagonal loop simulation outlined some of the main difficulties when using theoretical models as starting structures in MD simulations. Small differences in loop conformation can lead to major structural changes over long timescales, and it can be difficult to establish whether the differences are due to the general loop conformation or to the particular starting structure which was used. Both simulated lateral T_3-L_w-3 loop conformations suggested that having the first T residue located within the quadruplex groove is not the most stable conformation, as rearrangements occurred. However, this conformation was stable during a simulation of the 1A6H experimental structure with T_3 loops. The results from MD simulations are therefore highly dependent on the initial starting structure.

MD simulations of quadruplexes with T_2 loops were much more dependent upon the loop type, compared to the T_3 loop simulations. The diagonal T_2 loop quadruplex was very unstable during the simulation. After only 400 ps of dynamics, the G-quartet below the T_2 loop was severely distorted, as shown in Figure 6a. The strain caused by the short loop was apparent from the very beginning of the simulation, as the upper quartet G residues were forced closer together, making them tilt inside the quadruplex, and causing
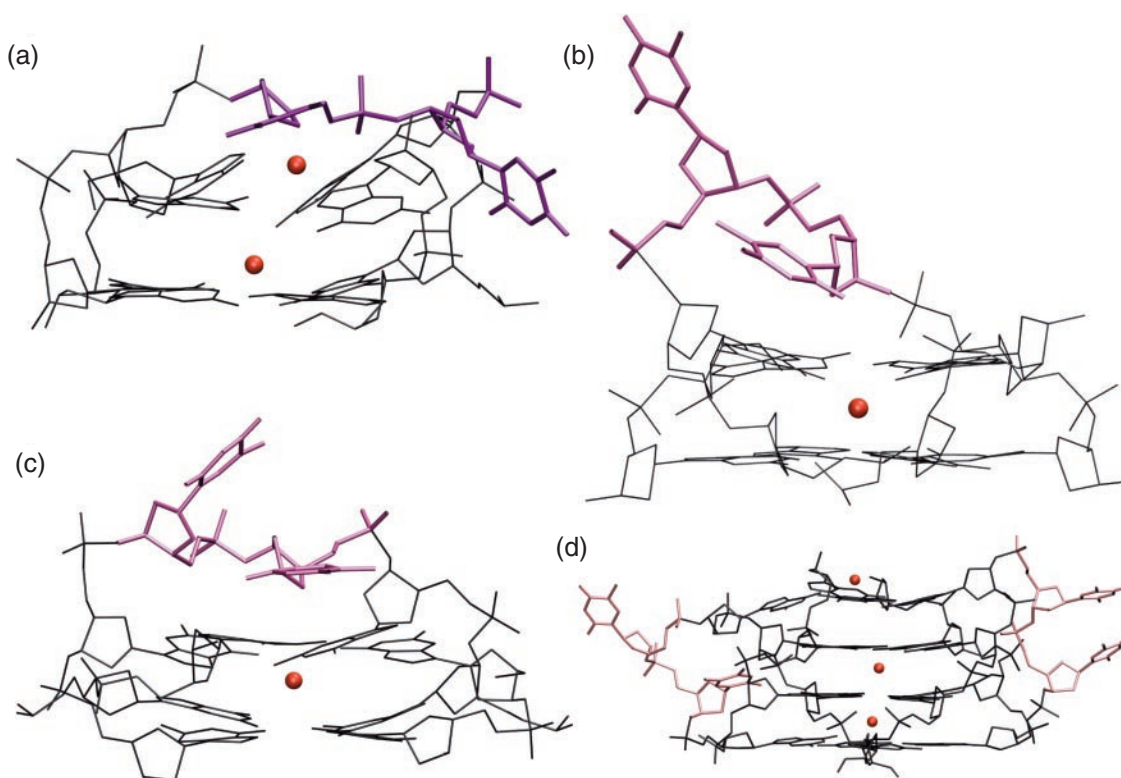
Figure 6. T_2 final loop conformations after 4 ns MD simulation (400 ps only for the diagonal loop). (**a**) Diagonal T_2 loop, (**b**) lateral T_2 loop over the narrow groove, (**c**) lateral T_2 loop over the wide groove and (**d**) parallel T_2 loops. The loops are shown in purple, and G-quartets in black, channel K^+ ions are shown as red spheres.

a K^+ ion to move out from the channel and into the loop region. These simulations suggest that dimeric quadruplexes with diagonal T_2 loops are unlikely to form in solution. The same strain effect was observed for the lateral T_2 loop over the wide quadruplex groove, although this was less pronounced than in the diagonal loop case. Figure 6c shows a selected lateral T_2 loop over the wider quadruplex groove after 4 ns of dynamics. Over the course of the simulation, the G-quartets became somewhat distorted, and not all Hoogsteen hydrogen bonds between the guanine bases were maintained. However, stacking between the G residues was mostly conserved. On the other hand, the T_2 lateral loop bridging a narrow groove caused no distortion of the G-quartets (Figure 6b), as was also suggested by the SA results. A parallel T_2 loop quadruplex was also stable over a 4 ns simulation (Figure 6d). A K^+ ion moved out of the quadruplex channel, but remained at the channel exit for the duration of the dynamics. A G-quartet was slightly out-of-plane, however, overall, the quadruplex structure was unaffected. Thus, MD simulations suggest that either parallel or lateral (narrow groove) loops are likely to be preferentially formed by short T_2 loops.

MM-PBSA calculations

Using SA and MD, stable conformations with both lateral and diagonal T_3 loops were found. Absolute free energy calculations using MM-PBSA were carried out in order to energetically rank the different structures, and the results are summarized in Table 1. The total free energy G_{total} was calculated for the structures simulated with MD in explicit solvent. G_{total} was further subdivided into contributions from each of the loops, following the method of Stefl *et al.* (33). G_{stem}, $G_{stem+loop1}$ and $G_{stem+loop2}$ were each calculated, and G_{loop1} and G_{loop2} were derived from $G_{stem+loop1} - G_{stem}$ and $G_{stem+loop2} - G_{stem}$, respectively. In this manner, the contribution of the loop-G-quartet interaction is included in the loop free energies. The data in Table 1 shows that all the loop conformations are close in energy. This was also observed during the dynamics as one loop type could rearrange to another, although each was stable for several nanoseconds. The LES simulations enabled more structural flexibility in the loops, and free energies after the LES simulations are generally lower than before the enhanced sampling. This suggests that during the LES simulations, more favourable loop conformations were formed. The T_3-L_w-5 conformation was the most favourable T_3 loop obtained, with a free energy of -461 kcal.mol^{-1}. Unfortunately, the G_{loop} values are not significantly different enough to be used to compare the stabilities of various loop conformations. Local fluctuations in loop geometry had significant effects on the calculated free energies at each step of the simulations, which tended to overshadow the free energy differences between structurally distinct loop types. Lateral T_3 loops over the narrow and wide groove of the quadruplex have similar energies, suggesting that both topologies are equally possible. Moreover, the free energies of diagonal and lateral loops were also similar, which does not allow differentiation between the two. The separation of the free energy into loop and quadruplex stem contributions is an approximation, which adds further uncertainty to the energies calculated. However, this approximation is necessary, as loop conformations adopted at each end of the quadruplexes during the simulations often differed, resulting in up to 10 kcal.mol^{-1} difference in free energy between the loops within a single quadruplex (Table 1). The total free energy of the G-quadruplex molecules does therefore not reflect the stability of individual loop conformations.

The solute entropic contribution was not included in the G values in Table 1, although the solvent entropy is implicitly taken into account in the solvation energy term. The entropic contribution is the least accurately calculated component of the free energy, and was therefore included separately. The entropy was calculated for the complete quadruplex structures, and was similar for the different quadruplexes considered. The entropic component is, however, important, as its inclusion can in principle alter the ranking of quadruplex energies.

MM-PBSA calculations were also carried out in order to differentiate between the stable T_2 lateral and parallel loop quadruplexes which were obtained. The decomposition of the free energy into loop and stem components, as used above, is inappropriate in this case, due to the completely different manner in which lateral and parallel loops interact with the G-quartets. Antiparallel quadruplex loops (lateral or diagonal) interact primarily with the G-quartet face. On the other hand, the parallel quadruplex loops interact with the groove region of the quadruplex. Only overall free energies of parallel and antiparallel quadruplexes can therefore be compared. The lateral (narrow groove) quadruplex was more favourable than the parallel quadruplex, with $G_{total} = -4237$ and $G_{total} = -4222$ kcal.mol^{-1}, respectively. This was still the case when the solute entropy was taken into account, as TS = 539 and TS = 543 kcal.mol^{-1}, respectively. For the particular loop conformations which were simulated, the antiparallel structure was more stable than the parallel structure. However, this does not imply that a more favourable parallel loop structure could not be found.

Table 1. Free energies (kcal.mol^{-1}) calculated using the MM-PBSA method for the T_3 loop quadruplexes

Loop conformation	G_{total}[a]	G_{stem}	G_{loop1}	G_{loop2}	$-TS_{total}$
T3-Lw-1 (before LES)	−4526 (4)	−3628 (4)	−449 (4)	−449 (4)	−585 (2)
T3-Lw-1 (after LES)	−4545 (4)	−3632 (4)	−451 (4)		−581 (1)
T3-Lw-5 (from T3-Lw-1 LES)			−461 (4)		
T3-Lw-3 (from 1A6H MD)[b]	−4522 (5)	−3620 (4)	−455 (4)		−582 (1)
T3-Lw-4 (from 1A6H MD)			−447 (4)		
T3-Lw-2 (before LES)	−4517 (5)	−3626 (4)	−442 (4)	−449 (4)	−580 (2)
T3-Lw-2 (after LES)	−4543 (5)	−3636 (4)	−456 (4)	−451 (4)	−586 (1)
T3-Ln-1 (before LES)	−4524 (4)	−3633 (4)	−446 (4)	−446 (4)	−582 (1)
T3-Ln-1 (after LES)	−4532 (4)	−3634 (4)	−451 (4)	−448 (4)	−582 (1)
T3-D-1	−4525 (5)	−3623 (4)	−452 (4)	−451 (4)	−579 (1)

Values were calculated before and after LES simulations (apart for the 1A6H template and diagonal loop simulations, for which LES simulations were not carried out). Standard errors of the mean are in parentheses. The entropic contribution was calculated using T = 300 K.
[a]G values do not include the solute entropy.
[b]Values calculated after a 2 ns equilibration.

DISCUSSION

### Comparison of T_3 experimental and predicted loops

Both X-ray (Hazel *et al.*, manuscript submitted) and NMR (42) structures of dimeric quadruplexes with T_3 loops show a clear preference for lateral over diagonal loop formation. The most frequently occurring experimental conformation, in which the first loop residue is in the quadruplex groove, and the other two residues are located above the G-quartets, was maintained during MD simulations of both the crystal structure and NMR lateral (wide groove) T_3 loops (Figure 7a and b, respectively). This preference was however difficult to reproduce in the simulations, as several loop conformations had similar energies (Table 1). There may well be some energetic contributions to these total energies that will favour one type of loop over the other. However these calculations cannot differentiate between these types. The contributions are included in the calculations of total free energies, but overall do not favour a particular structure. The most favourable T_3 loop conformation, predicted using MM-PBSA calculations, was a lateral loop; however, this conformation is not that observed experimentally, in which the first T residue is located in the quadruplex groove. Instead, the MM-PBSA calculations favour a conformation in which all three loop residues are stacked above the G-quartet planes (Figure 7c). This conformation was predicted to be more stable (by at least 6 kcal.mol^{-1}) than the experimental NMR conformation (1A6H simulation), as calculated by MM-PBSA (Table 1). Various simulations of the d($G_4T_3G_4$)$_2$ crystal structures have yielded loop free energies which were within a few kcal.mol^{-1} of the predicted T_3 loop conformation (Hazel *et al.*, manuscript submitted). The most favourable experimental T_3 loop conformation, over the wide quadruplex groove (shown in Figure 7a), is only 1 kcal.mol^{-1} more favourable than the predicted conformation; this is within the MM-PBSA calculation error.

There is experimental evidence suggesting that T_3 loops are flexible, and able to adopt several distinct conformations. Thus several loop conformations have been found by NMR methods to exist in solution, although the overall loop conformation is always conserved (with the first loop residue located within the quadruplex groove, and the second and third residues above the G-quartets) (42). Both lateral loops over the narrow and wide quadruplex grooves were found in the d($G_4^{Br}UT_2G_4$)$_2$ crystal structure, suggesting that these are indeed energetically similar, as suggested here by the MM-PBSA calculations in Table 1. However, the experimental results also suggest that lateral T_3 loops are favoured over diagonal loops. This was not reproduced in the simulations, as diagonal and lateral loops have similar computed energies.

### Comparison of T_2 predicted loops with NMR structure

It has been shown above that the conformations of T_3 loops in dimeric quadruplexes were difficult to predict with MD simulations because a number of conformers were found to be possible. This was not the case for dimeric quadruplexes containing T_2 loops, whose structures are constrained by the limited span of the 2 nt loop. In this case, the particular loop conformation is less important, as certain T_2 loop structures were found to be unstable due to strain in the loop backbone. The T_2 loops were found to be unable to form diagonal loops, and quadruplex structures with two lateral loops over the wide quadruplex groove were also distorted (Figure 6), using the d($G_4T_3G_4$) crystal structure (Hazel *et al.*, manuscript submitted) as a template, after replacement of the T_3 loops with T_2. The d($G_4^{Br}UT_2G_4$) crystal structure has two types of quadruplex dimer within the asymmetric unit of the structure (Hazel *et al.*, manuscript submitted), a head-to-tail homodimer with both loops across the wide grooves, and a head-to-head heterodimer with loops across the wide and narrow grooves. The d($TG_4T_2G_4T$) sequence forms two heterodimers, a head-to-tail and a head-to-head dimer both with loops across a wide and a narrow groove (6). Both these heterodimers have *syn-syn-anti-anti* G glycosidic angles around the G-quartets. Simulations were carried out in order to establish whether these dimerization difference could be the consequence of the differences in loop length, rather than any conformational differences in the G-quartets themselves. The d($G_4^{Br}UT_2G_4$) crystal structures suggest that both G-quartet core structures, with *syn-anti-syn-anti* (head-to-tail dimer) and *syn-syn-anti-anti* (head-to-head dimer) G glycosidic angles around the quartets, are equally stable.

A d($G_4T_2G_4$) head-to-head dimer, with loops over the narrow and wide grooves was stable over a 4 ns simulation, and the loop residues were able to form stacking interactions with the G-quartets (Figure 8b). However, the head-to-tail dimer, which has both T_2 loops over wide quadruplex grooves, became distorted during the simulation (Figure 8a). It therefore appears that the strain caused by a T_2 loop bridging a wide groove can be compensated for by the second loop bridging a shorter distance. However, two T_2 loops bridging wide grooves are unstable. The stable head-to-head d($G_4T_2G_4$) dimer simulated has the same topology as one of the d($TG_4T_2G_4T$) solution NMR structures (6). The latter sequence also forms a head-to-tail dimer in solution, however this has a different topology to the d($G_4T_3G_4$) head-to-tail dimer. It is possible for the different head-to-tail topologies observed to be due to effects of the loop lengths, rather than G-quartet stem stability. The simulations carried out here tend to support this view. A head-to-tail dimer with
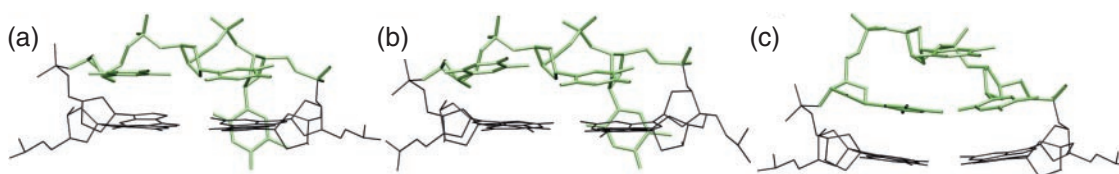
Figure 7. MD simulations of (**a**) d($G_4T_3G_4$)$_2$. X-ray structure T_3 loop conformation averaged over 2 to 4 ns (**b**) NMR structure T_3 loop conformation averaged over 2 to 4 ns and (**c**) most favourable predicted T_3-L$_w$-5 loop averaged over 1 to 2 ns MD.
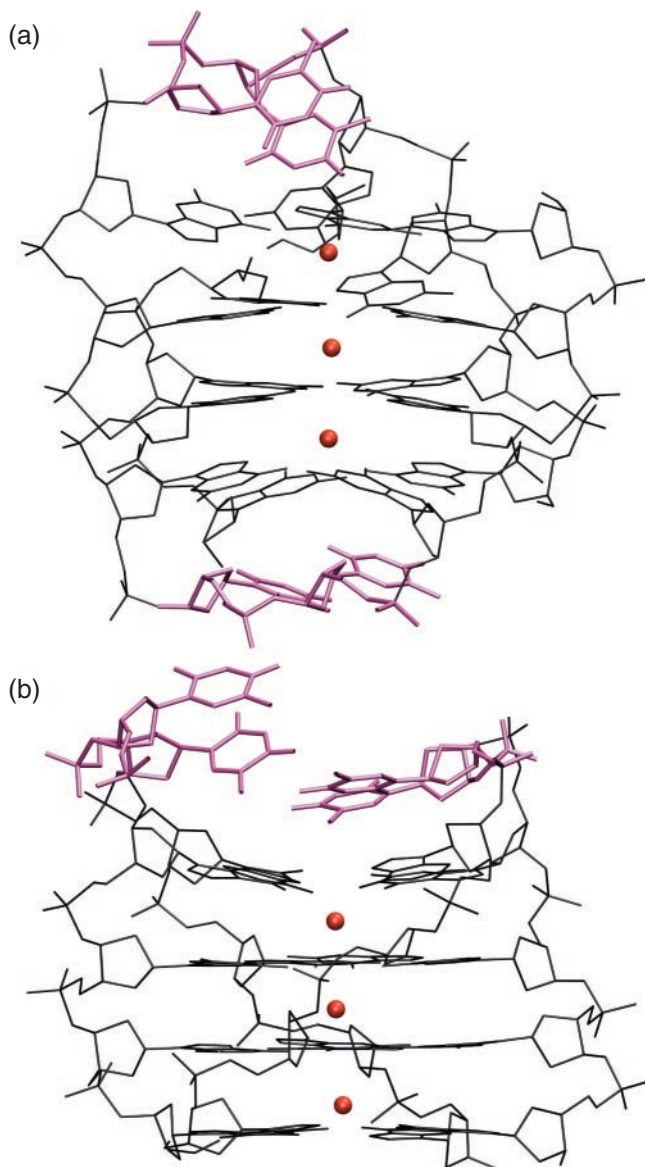
Figure 8. (**a**) Head-to-tail and (**b**) head-to-head quadruplex structures with two T_2 loops shown in purple. The complete quadruplexes are shown, with K^+ channel ions in red. These structures have been averaged over the final 2 ns of the 4 ns simulations.

both loops over the wide groove, as formed by the $d(G_4T_3G_4)$ sequence, cannot support two T_2 loops without the G-quartets becoming distorted. However, dimerization which enables T_2 loops to bridge a wide and a narrow groove can lead to stable structures. Strain caused by the loop length could therefore explain the different head-to-tail dimers formed by 3 nt compared to 2 nt loops.

One of the limitations of MD simulations of quadruplex structures is their inability to differentiate between K^+ and Na^+ ion binding within the channel. All simulations in this work were carried out with K^+ ions bound within the quadruplex channel. This is in contrast to the $d(TG_4T_2G_4T)$ NMR structure, which was determined in Na^+ ion containing solution. The different ions are, however, not expected to affect the results, as strain was the dominant factor in the

simulations, and this is unlikely to be affected by different channel ions.

CONCLUSIONS

We have shown here that dimeric DNA quadruplexes with 3 nt-loops can adopt a wide range of conformations. NMR and crystal structures of quadruplexes with T_3 loops indicate a preference for lateral over diagonal loops. This was not fully reproduced during the simulations. Even though the most favourable T_3 loop found was indeed a lateral loop, diagonal loop conformations were found which had similar free energies. Moreover, simulations were unable to unambiguously identify the favoured lateral T_3 loop conformation found both in the crystal and in solution. However, the small free energy differences between the T_3 loop types simulated are also in accord with experimental findings. Thus NMR studies of these structures have shown that the loop regions are flexible, while loops over both the wide and narrow quadruplex grooves were found in the $d(G_4^{Br}UT_2G_4)$ crystal structure. The free energy calculations could therefore indicate real energetic similarities between the different loop types. Loop flexibility was also suggested by the structural transitions which occurred in the loops during the simulations. However, it appears that approximate calculations, such as MM-PBSA are not able to reproduce accurately enough the small energy differences between various loop conformations, in order to reliably predict the most favourable conformations in solution.

Shorter T_2 loops in these quadruplexes do restrict the conformational flexibility of the quadruplexes. Simulations suggested that the structures adopted by sequences with T_2 loops in solution may be governed by loop length, rather than only by the stability of the G-quartet cores. These simulations, and the $d(G_4T_3G_4)$ X-ray structures solved by us, both suggest that G-quartets, which have alternating *syn-anti-syn-anti* or *syn-syn-anti-anti* G-glycosidic angles around the G-quartets are equally stable. The structures of dimeric quadruplexes formed by the $d(G_4T_3G_4)$ and $d(G_4T_2G_4)$ sequences are most likely different due to the differing lengths of their loops, which does not allow the dimerization of two strands with T_2 loops over the wide quadruplex groove.

These simulations emphasize the influence of loop length on the folding of G-quadruplexes, in accord with earlier NMR studies (46). This may be the dominant factor in the folding of certain sequences, especially with short loops comprising one or 2 nt. Preliminary crystal structure data on the $d(G_4T_2AG_4)$ quadruplex (P. Hazel *et al.* (2006) *J. Am. Chem. Soc.*, in press) suggest that this adopts a structure very similar to $d(G_4T_3G_4)$, with lateral loops having the same conformation. This further emphasizes the innate importance of loop length, and that the nucleotide composition of short loops may be of lesser importance.

This study, together with our previous one (30) on the effect of loop length on intramolecular quadruplex structure, shows that valuable structural insights can be gained from MD simulations. Although particular loop conformations are difficult to predict, general features of quadruplex topology, especially when caused by strained shorter loops, can be established with relatively short time-scale simulations. Thus diagonal and lateral (wide groove) T_2 loop simulations showed G-quartet distortion within the first few 100 ps of simulation. Unlike T_4

loops, in which ion binding within the loop region can cause difficulties in the simulations (32), crystal structures of quadruplexes with T_3 loops do not show any ion binding within the loops. This absence of ion within the loop region means that MD simulations are able to reproduce loop conformations when the experimental structures are simulated.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

REFERENCES

1. Parkinson,G.N., Lee,M.P.H. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
2. Phan,A.T. and Patel,D.J. (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms an interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J. Am. Chem. Soc.*, **125**, 15021–15027.
3. Ambrus,A., Chen,D., Dai,J., Jones,R.A. and Yang,D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-*myc* promoter. Implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048–2058.
4. Phan,A.T., Modi,Y.S. and Patel,D.J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human *c-myc* promoter. *J. Am. Chem. Soc.*, **126**, 8710–8716.
5. Zhan,N., Phan,A.T. and Patel,D.J. (2005) (3+1) assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. *J. Am. Chem. Soc.*, **127**, 17277–17285.
6. Phan,A.T., Modi,Y.S. and Patel,D.J. (2004) Two-repeat *Tetrahymena* telomeric d(TGGGGTTGGGGT) sequence interconverts between asymmetric dimeric G-quadruplexes in solution. *J. Mol. Biol.*, **338**, 93–102.
7. Padmanabhan,K., Padmanabhan,K.P., Ferrara,J.D., Sadler,J.E. and Tulinsky,A. (1993) The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J. Biol. Chem.*, **268**, 17651–17654.
8. Macaya,R.F., Schultze,P., Smith,F.W., Roe,J.A. and Feigon,J. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl Acad. Sci. USA*, **90**, 3745–3749.
9. Haider,S., Parkinson,G.N. and Neidle,S. (2002) Crystal structure of the potassium form of an *Oxytricha nova* G-quadruplex. *J. Mol. Biol.*, **320**, 189–200.
10. Smith,F.W. and Feigon,J. (1992) Quadruplex structure of *Oxytricha* telomeric DNA oligonucleotides. *Nature*, **356**, 164–168.
11. Schultze,P., Hud,N.V., Smith,F.W. and Feigon,J. (1999) The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the *Oxytricha nova* telomere repeat oligonucleotide d($G_4T_4G_4$). *Nucleic Acids Res.*, **27**, 3018–3028.
12. Schultze,P., Smith,F.W. and Feigon,J. (1994) Refined solution structure of the dimeric quadruplex formed from the *Oxytricha* telomeric oligonucleotide d(GGGGTTTTGGGG). *Structure*, **2**, 221–233.
13. Crnugelj,M., Hud,N.V. and Plavec,J. (2002) The solution structure of d($G_4T_4G_4$)$_2$: a bimolecular G-quadruplex with a novel fold. *J. Mol. Biol.*, **320**, 911–924.
14. Strahan,G.D., Keniry,M.A. and Shafer,R.H. (1998) NMR structure refinement and dynamics of the K$^+$-[d($G_3T_4G_3$)]$_2$ quadruplex via particle mesh Ewald molecular dynamics simulations. *Biophys. J.*, **75**, 968–981.
15. Smith,F.W., Lau,F.W. and Feigon,J. (1994) d($G_3T_4G_3$) forms an asymmetric diagonally looped dimeric G-quadruplex with guanosine 5′-syn-syn-anti and 5′-syn-anti-anti N-glycosidic conformations. *Proc. Natl Acad. Sci. USA*, **91**, 10546–10550.
16. Cmugelj,M., Sket,P. and Plavec,J. (2003) Small change in a G-rich sequence, a dramatic change in topology: new dimeric G-quadruplex folding motif with unique loop orientations. *J. Am. Chem. Soc.*, **125**, 7866–7871.
17. Wang,Y. and Patel,D.J. (1993) Solution structure of the human telomeric repeat d(AG$_3$(T$_2$AG$_3$)$_3$) G-tetraplex. *Structure*, **1**, 263–282.
18. Wang,Y. and Patel,D.J. (1994) Solution structure of the *Tetrahymena* telomeric repeat d(T$_2$G$_4$)$_4$ G-tetraplex. *Structure*, **2**, 1141–1156.
19. Kuryavyi,V., Majumdar,A., Shallop,A., Chernichenko,N., Skripkin,E., Jones,R. and Patel,D.J. (2001) A double chain reversal loop and two diagonal loops define the architecture of a unimolecular DNA quadruplex containing a pair of stacked G(syn)·G(syn)·G(anti)·G(anti) tetrads flanked by a G·(T-T) triad and a T·T·T triple. *J. Mol. Biol.*, **310**, 181–194.
20. Wang,Y. and Patel,D.J. (1995) Solution structure of the *Oxytricha* telomeric repeat d[G$_4$(T$_4$G$_4$)$_3$] G-tetraplex. *J. Mol. Biol.*, **251**, 76–94.
21. Simonsson,T., Pecinka,P. and Kubista,M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.*, **26**, 1167–1172.
22. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-*MYC* transcription. *Proc. Natl Acad. Sci. USA*, **99**, 11593–11598.
23. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
24. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2901–2907.
25. De Armond,R., Wood,S., Sun,D., Hurley,L.H. and Ebbinghaus,S.W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry*, **44**, 16341–16350.
26. Sun,D., Guo,K., Rusche,J.J. and Hurley,L.H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.*, **33**, 6070–6080.
27. Rankin,S., Reszka,A.P., Huppert,J., Zloh,M., Parkinson,G.N., Todd,A.K., Ladame,S., Balasubramanian,S. and Neidle,S. (2005) Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.*, **127**, 10584–10589.
28. Phan,A.T. and Patel,D.J. (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J. Am. Chem. Soc.*, **125**, 15021–15027.
29. Rezler,E.M., Seenisamy,J., Bashyam,S., Kim,M.-Y., White,E., Wilson,W.D. and Hurley,L.H. (2005) Telomestatin and diseleno sapphyrin bind selectively to two different forms of the human telomeric G-quadruplex structure. *J. Am. Chem. Soc.*, **127**, 9439–9447.
30. Hazel,P., Huppert,J., Balasubramanian,S. and Neidle,S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.*, **126**, 16405–16415.
31. Spackova,N., Berger,I. and Sponer,J. (1999) Nanosecond molecular dynamics simulations of parallel and antiparallel guanine quadruplex DNA molecules. *J. Am. Chem. Soc.*, **121**, 5519–5534.
32. Spackova,N., Berger,I. and Sponer,J. (2001) Structural dynamics and cation interactions of DNA quadruplex molecules containing mixed guanine/cytosine quartets revealed by large-scale MD simulations. *J. Am. Chem. Soc.*, **123**, 3295–3307.
33. Stefl,R., Cheatham,T.E.III, Spackova,N., Fadrna,E., Berger,I., Koca,J. and Sponer,J. (2003) Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of substates. *Biophys. J.*, **85**, 1787–1804.

34. Fadrna,E., Spackova,N., Stefl,R., Koca,J., Cheatham,T.E.III. and Sponer,J. (2004) Molecular dynamics simulations of guanine quadruplex loops: advances and force field limitations. *Biophys. J*., **87**, 227–242.

35. Gill,M.L., Strobel,S.A. and Loria,J.P. (2005) [205]Tl methods for the characterisation of monovalent cation binding to nucleic acids. *J. Am. Chem. Soc*., **127**, 16723–16732.

36. Horvath,M.P. and Schultz,S.C. (2001) DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein-DNA complex. *J. Mol. Biol*., **310**, 367–377.

37. Elber,R. and Karplus,M. (1990) Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc*., **112**, 9161–9175.

38. Simmerling,C. and Elber,R. (1994) Hydrophobic 'collapse' in a cyclic hexapeptide: computer simulations of CHDLFC and CAAAAC in water. *J. Am. Chem. Soc*., **116**, 2534–2547.

39. Srinivasan,J., Cheatham,T.E.III., Cieplak,P., Kollman,P.A. and Case,D.A. (1998) Continuum solvent studies of the stability of DNA, RNA and phosphoramidate-DNA helices. *J. Am. Chem. Soc*., **120**, 9401–9409.

40. Jayaram,B., Sprous,D., Young,M.A. and Beveridge,D.L. (1998) Free energy analysis of the conformational preferences of A and B forms of DNA in solution. *J. Am. Chem. Soc*., **120**, 10629–10633.

41. Kang,C., Zhang,X., Ratliff,R., Moyzis,R. and Rich,A. (1992) Crystal structure of four-stranded *Oxytricha* telomeric DNA. *Nature*, **356**, 126–131.

42. Kettani,A., Kumar,R.A. and Patel,D.J. (1995) Solution structure of a DNA quadruplex containing the fragile X syndrome triple repeat. *J. Mol. Biol*., **254**, 638–656.

43. Kelley,L.A., Gardner,S.P. and Sutcliffe,M.J. (1996) An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng*., **9**, 1063–1065.

44. Case,D.A., Pearlman,D.A., Caldwell,J.W., Cheatham,T.E.III., Wang,J., Ross,W.S., Simmerling,C.L., Darden,T.A., Merz,K.M., Stanton,R.V. *et al*. (2002) *AMBER 7*. University of California, San Francisco.

45. Bondi,A. (1964) Van der Waals volumes and radii. *J. Phys. Chem*., **68**, 441–451.

46. Marathias,V.M. and Bolton,P.H. (1999) Determinants of DNA Quadruplex Structural Type: Sequence and Potassium Binding. *Biochemistry*, **38**, 4355–4364.