**A molecular model for drug binding to tandem repeats of telomeric G-quadruplexes**

Shozeb M. Haider and Stephen Neidle1

CRUK Biomolecular Structure Group, The School of Pharmacy, University of London, 29–39 Brunswick Square, London, WC1N 1AX, U.K.

---

**Abstract**

The extreme 3'-ends of human telomeres consist of 150–250 nucleotides of single-stranded DNA sequence together with associated proteins. Small-molecule ligands can compete with these proteins and induce a conformational change in the DNA to a four-stranded quadruplex arrangement, which is also no longer a substrate for the telomerase enzyme. The modified telomere ends provide signals to the DNA-damage-response system and trigger senescence and apoptosis. Experimental structural data are available on such quadruplex complexes comprising up to four telomeric DNA repeats, but not on longer systems that are more directly relevant to the single-stranded overhang in human cells. The present paper reports on a molecular modelling study that uses Molecular Dynamics simulation methods to build dimer and tetramer quadruplex repeats. These incorporate ligand-binding sites and are models for overhang-ligand complexes.

---

**Background**

Telomeric DNA is the non-coding DNA at the ends of eukaryotic chromosomes that protects the genome from degradation, chromosomal end-to-end fusion and recombination [1]. The inability of DNA polymerase to fully replicate telomeric DNA results in progressive telomere shortening in proliferating somatic cells until a point of critical shortening is reached, when cells enter irreversible growth arrest. Most cancer cells prevent shortening of telomeric DNA by adding hexanucleotide repeats to 3'-ends, utilizing the reverse transcriptase activity of the telomerase enzyme complex [2]. Telomerase expression is a key marker of cellular immortalization and has been observed in 80–85% of cancer cells [3]. The extreme 150–250 nucleotides at the 3'-end of telomeric DNA are single-stranded, and appropriate small-molecule ligands can induce the formation of higher-order folded DNA structures [4–6]. These can inhibit telomerase activity, since elongation and catalysis require a single-stranded DNA substrate for effective hybridization by the RNA subunit of the telomerase complex, leading to telomere shortening and cell death in neoplastic cells. In order to form higher-order DNA, the small molecules compete with the single-strand-binding protein POT1 (protection of telomeres 1) and with the end-capping function of telomerase [7,8]. Their displacement rapidly leads a DNA-damage-response signal, which can be detected, for example by γ-H2AX (phosphorylated histone H2AX) [9]. This in turn activates apoptotic pathways. This approach is currently of interest as a selective therapeutic strategy in human cancer.

Human telomeric DNA comprises tandem repeats of the sequence d(TTAGGG). The formation of higher-order structures in the single-stranded telomeric DNA overhang is a consequence of the ability of a guanine base to form hydrogen-bonding interactions on both its Watson–Crick and Hoogsteen faces [10–12]. This enables guanines to readily self-associate to form a highly stable structural motif involving four guanines held together via eight Hoogsteen hydrogen bonds in a co-planar array, termed a G-quartet. Several G-quartets can stack on top of another, to form a four-stranded G-quadruplex arrangement, with the G-quartets held together by nucleotides from the sequences that occur between each G-tract. Quadruplexes can be formed by the folding of a single contiguous sequence containing multiple G-tracks (unimolecular) or by the association of two or four separate strands (bi- or tetra-molecular). They are stabilized by extensive π–π stacking non-bonded attractive interactions between each G-quartet together with the involvement of alkali metal ions (K⁺ and Na⁺). These are localized in the central channel of G-quadruplexes and co-ordinate to O6 atoms of the guanines in a bipyramidal prismatic arrangement.

G-quadruplexes can be highly polymorphic, adopting a variety of folds, especially when the loops are 3–4 nt. Several distinct structural topologies of G-quadruplexes have been identified using NMR and X-ray crystallography including those from human telomeric DNA [13–17] and Oxytricha nova telomeric DNA [18]. Strand polarity depends in large part upon the nature and length of the loop sequences intervening between the G-tracts. Crystallographic analyses of a human telomeric bimolecular quadruplex and a unimolecular quadruplex have revealed a topology with all strands parallel and propeller-like loops [13]. In contrast, several NMR studies on human telomeric unimolecular quadruplexes have shown distinct topologies involving mixed parallel and anti-parallel (3+1) backbone arrangements [14–17]. No structural
information is currently available on tandem repeats of quadruplexes (multimers) such as may be formed along the length of the single-stranded telomeric DNA overhang, although several models have been proposed [19,20].

There are only restricted structural data available to date on quadruplex–ligand complexes. Two categories of topologies have been found: (i) the bimolecular diagonal loop crossover topology from O. nova telomeric DNA in complexes with di-substituted acridines [18,21]; and (ii) the parallel-stranded propeller-loop topology from human telomeric DNA, observed in three bimolecular and one unimolecular complexes, with three very different types of ligand [22–24]. This suggests that a parallel topology is an appropriate starting point for models of higher-order quadruplex multimers with bound ligand, which may also provide insight into the structural requirements of the DNA–quadruplex multimers with bound ligand, which may also provide insight into the structural requirements of the DNA–ligand interaction.

The drug molecule is asymmetrically sandwiched between the two quadruplexes, with one acridine face stacked on to a 5′-end. The other face is stacked on to one half of a G-quartet at the 3′-end. The drug molecule is asymmetrically sandwiched between the two quadruplexes, with one acridine face stacked on to a 5′-end. The other face is stacked on to one half of a G-quartet at the 3′-end. The 5′–3′ continuity of the two quadruplexes in the biological unit enables the ready construction of a higher-order model based on the scaffold observed in the crystal structure by means of molecular modelling. This has enabled a unimolecular model of the complex to be built, while retaining all the features present in the original crystal structure, using explicitly solvated all-atom MD (Molecular Dynamics) simulation methods to obtain low-energy structures and relative free energies estimated using MM-PBSA (Molecular Mechanics and the Poisson–Boltzmann surface area approximation).

The multimer model-building procedure
The crystal structure of the human telomeric DNA bimolecular quadruplex complexed with BRACO-19 [24] was taken from the Protein Data Bank (PDB code 3CE5) and used as a scaffold for further molecular modelling. Since the basic unit is a bimolecular quadruplex, there are only two TTA loops per individual unit. The modelling methodology is analogous to that described previously [19].

The first step was to generate a unimolecular model for both of the units. The crystal structure of human telomeric DNA d[AG₃(T₂AG₃)₃] (PDB code 1KF1) was also used and the terminal adenine at the 5′ end of it was removed (using the INSIGHT package) to generate a 21-mer. This was then superimposed on the individual units in the 3CE5 structure. The overall RMSD (root mean square deviation) of superimposition of the G-quartets was 0.48 and 0.41 Å for the top and bottom units respectively. The central G-quartets from 3CE5 were deleted and the new G-quartets from 1KF1 were superimposed on the old positions inserted into the units. An additional loop from the 1KF1 structure was also added to the unit. This resulted in one continuous unit consisting of 21 nucleotides. Both units in 3CE5 were replaced in this way. The two units were then joined by a TTA loop that was extracted from the 1KF1 crystal structure, to form a continuous 45-mer, i.e. a dimer of two unimolecular quadruplexes (Figure 1).

MD simulations
The final 45-mer was subjected to several rounds of molecular-mechanics energy minimization to relieve any steric clashes in the structure. K⁺ ions are required to stabilize quadruplexes and are located between the G-quartets in a vertical alignment along the axis in the electronegative channel within the core. The cations were retained from the BRACO-19 complex crystal structure. One feature of this structure is that the two quadruplex units are offset with respect to each other and are inclined by approx. 30° in two directions [24]. As a result of this offset, the ion channel is discontinuous. This is, however, compensated for by the positively charged nitrogen atom at the centre of the acridine group in the ligand which is positioned on top of the channel (Figure 1b). The system was subjected to 10 ns of explicitly solvated MD using the Amber ff99sb and parmbsc0 force-fields for nucleic acids, as outlined previously [19]. Free energies were estimated employing the MM-PBSA method.

In order to retain the structural features of the 3CE5 crystal structure, modifications were carried out to the loops. This was based on previous observations that the TTA loops are highly flexible and can adopt several conformations [19]. First, modifications were carried out to the first TTA loop (residues 4–6). The first thymine (Thy4) from the TTA loop flips back into the ligand-binding site and interacts with the ligand, making hydrogen bonds with the central nitrogen in the acridine chromophore and the amide nitrogen in one side chain. This arrangement has also been observed in the O. nova antiparallel bimolecular quadruplex crystal structures in complexes with di-substituted acridine ligands [18,21], where it functions to position the ligands in their binding site. As a result of this base flipping and thymine stacking in the binding site, the loop connecting to the top of the next guanine run is now reduced to two bases (TA). The second loop is a conventional loop from the crystal structure with no modification to its conformation (TTA, residues 10–12). The first thymine (Thy10) in the third loop is involved...
Figure 1 | Four views of the model for a complex between the drug BRACO-19 and two contiguous G-quadruplexes

(A) Backbone representation of the overall structure. Backbone atoms contributed by guanines are coloured blue, bases from the loops are coloured red, and the connecting loop is coloured green. The ligand (BRACO-19) is coloured yellow, and the ions present in the central electronegative channel within the structure are shown as green balls. (B) Stacking arrangement in the 45-mer modelled structure. Thy17 makes π-stacking interactions in the crystal structure, which is retained within the model and the free Thy22 base is modelled to stack on top of Thy17. The ions are represented as orange balls, highlighting the discontinuous ion channel ending in the positively charged nitrogen group in the central acridine chromophore of the BRACO-19 molecule. The ion channel extends on the other side beginning from the positively charged nitrogen group at the 9 position of the ligand. (C) Snapshot of the 45-mer structure at the end of a 10 ns MD run. The structure is remarkably stable while retaining most of the features from the crystal structure. (D) Stacking arrangement of the ligand on the TATA tetrad and G-quartet at the interface between the two component quadruplexes. The guanines are coloured blue, bases from the loops are coloured red and those from the connector loop (residues 22–24) are coloured green. The BRACO-19 ligand is coloured yellow, and the ions in the structure are coloured orange.

in the TATA tetrad. The second thymine (Thy17) loops back in to form a π-stack with the anilino substituent group in the ligand. This structural feature was also retained from the crystal structure. The backbone atoms of Thy18 and Ade19 in this loop connect to the top of the G-quartet. The next TTA loop (residues 22–24) or the connector loop connects the top quadruplex unit (residues 1–21) to the bottom unit (residues 25–45). The first thymine (Thy22) base in this connector loop forms π-stacking interactions with Thy17 (Figure 1b). The two following bases (Thy23 and Ade24) contribute towards the TATA tetrad, while their backbone atoms link the two quadruplex units together. There are no modifications in the next loop (TTA, residues 28–30). Adenine from the next loop (Ade36) is modified to stack and form the fourth base in the TATA tetrad (Thy16–Ade36–Thy23–Ade24).

Structural stability of the model
The conformational stability of the model can be assessed by measuring the RMSD over the course of an MD simulation.
Figure 2 | Structural drift observed during the MD simulation

Left: RMSD plot showing the stability of the model during the MD run. RMSD values calculated for all atoms (black) and backbone-only atoms (red) are plotted. Right: RMSD plot comparing the different segments in the 45-mer model. The G-quartets (blue) form the most stable segments in the model. The backbone-only atoms (black) and all atoms (red) from the loops are also illustrated. It is evident that the higher RMSD values observed for the model arise from the wobble exhibited by the free bases in the loops. Ang = Å.

Table 1 | RMSD values [for all-atoms and backbone (bb) calculations] and numbers of atoms in the simulated structures

<table>
<thead>
<tr>
<th>Model</th>
<th>Timescale (ns)</th>
<th>Environment</th>
<th>DNA (bb)</th>
<th>Waters</th>
<th>RMSD (all) (Å)</th>
<th>RMSD (bb) (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>45-mer (+BRACO-19)</td>
<td>10</td>
<td>DNA (bb)</td>
<td>1463 (269) (+90)</td>
<td>23256</td>
<td>3.20</td>
<td>2.49</td>
</tr>
<tr>
<td>45-mer* (+BSU6039)</td>
<td>15</td>
<td>DNA (bb)</td>
<td>1463 (269) (+70)</td>
<td>22961</td>
<td>3.68</td>
<td>2.90</td>
</tr>
</tbody>
</table>

*Data from [19].

An unstable model with incorrect geometries, simulated at 300 K, loses its structural integrity. The overall RMSD for the all-atom model (black) and backbone-only atoms (red) are illustrated in Figure 2 (left-hand panel). An initial jump in RMSD value is observed within the first 1 ns of the simulation. This corresponds to the relaxation of the starting model. The trajectory is eventually stabilized over the course of the simulation with relatively small fluctuations. The RMSD plot in Figure 2 (right-hand panel) highlights different segments in the model. The co-planar arrangement of guanines in a G-quartet (blue), held together by hydrogen bonds is the most stable segment of the model. There are very few differences in the RMSD values observed between an all-atom model and a backbone-only model for the G-quartets. The all-atom RMSD for the loops (red) is quite high; however, the backbone (black) RMSD of the loops are relatively stable. This clearly suggests that the higher RMSD is a result of the wobbling effect of the nucleotide bases. This correlates with our earlier study that suggested that the loops can rearrange to adopt multiple conformations [19]. Comparison has also been made with the model containing a ligand in its pseudo-intercalation site. The overall RMSD value is lower in the current model compared with the model containing the di-substituted acridine (Table 1). In the current model, thymine (Thy4) from the first loop flips back and enters the binding site, making strong interactions with BRACO-19. The average Thy4(O4)-BRA(N21) distance is 3.0 Å and Thy4(N3)-BRA(N7) is 3.4 Å. These hydrogen-bonding interactions are maintained throughout the course of the simulation, and serve to position and hold the ligand tightly in the binding site, allowing it only limited movement. As a result of this, the ligand stacks on one half of the G-quartet and makes π-stacking interactions with a guanine tetrad on one face and a TATA tetrad on the other (Figure 1d). Both of these stacking platforms are extremely stable with RMSDs of 1.1 and 1.3 Å respectively. The hydrogen-bonding pattern in the TATA tetrad is also maintained with donor–acceptor distances averaging 3.1 Å over the course of the simulation. The TTA linker loop (residues 22–24) is highly stable. This is because two residues from this
 linker loop take part in the TATA tetrad formation. The other thymine (Thy22) forms π-stacking interactions with Thy17.

The G-quartets

The G-quartets in the 45-mer are extremely stable, with their hydrogen-bonding networks preserved throughout the course of the simulation. There are minor differences in hydrogen-bonding distances. The average N1–O6 distance is 3.00 Å and N2–N7 distance is 2.97 Å. This has increased by 0.14 and 0.15 Å respectively compared with those in the crystal structures. The G-quartet-stacking geometry in terms of rise and twist is also retained in the model. The individual quadruplex units have a quasi-helical repeat of 12 G-quartets per turn, similar to that present in the crystal structure [24].

Conformations of the TTA loops

Previous studies on quadruplex multimers [19] confirmed that TTA loops are the most flexible part of a quadruplex structure. The TTA propeller loops in a parallel-stranded topology are arranged externally, directed away from the G-quartets. They have been previously shown to be highly mobile and can readily adopt several distinct conformations [19]. In the current model, some of the TTA loops contribute towards the structural features within the core of the model, including π-stacking and formation of the TATA tetrad. One thymine base, Thy1 in the first loop, is flipped back to interact with the ligand. This interaction is important in positioning the ligand in its binding site and is maintained throughout the course of the simulation. The backbone of the remaining two bases in the loop (Thy3 and Ade6) connects the stacked Thy1 to the guanines. The second loop has no structural modifications and exhibits dynamics similar to that observed in our previous quadruplex modelling studies [19], where stabilization of the loop coming from hydrogen-binding interactions of the central thymine with adjacent adenine and thymine. The first thymine in the third loop (Thy46) contributes to the TATA tetrad. The second thymine in this loop (Thy47) is arranged such that the base forms a π-stack with the anilino group in the side chain of BRACO-19, a feature observed in the crystal structure and is retained in the model. This structural arrangement is lost after approx. 2.2 ns, and the thymine base no longer makes stacking interactions with the anilino side chain of the ligand. Similarly to loop 1, the backbone atoms in thymine/adenine link the stacked thymine to the top of the next guanine run.

The next loop is the connector loop (residues 22–24) that joins the two individual quadruplex units together. Thy22, which stacks on Thy17, also loses its π-stacking arrangement at approx. 2 ns and does not contribute towards any interactions for the remainder of the simulation. The loss of π-stacking interactions between Thy17, Thy22 and the anilino group can be attributed to the flexibility of the anilino side chain at the 9 position in the ligand combined with the restraints on the bases set while constructing the model.

The next loop in the next quadruplex unit (TTA, residues 28–30) was added to the model without any modifications. However, during the course of the simulation, it adopts a conformation which is reminiscent of those seen in our previous work, where the loops extend outwards away from the quadruplex [19]. Similar dynamics behaviour is also observed for the two thymines (Thy34 and Thy35) in the next loop (TTA, residues 34–36), where they stack on top of one another during the simulation. Ade35 is stacked to complete the TATA tetrad. The last loop (TTA, residues 40–42) adopts a conformation similar to that observed in the original crystal structure.

Estimated energies can provide relative semi-quantitative indications of the stability of the model and its components. Models with lower energies are expected to be more stable than those with higher values. We have calculated free energies of the TATA tetrad, GGGG stacking tetrad and the loops to identify the individual contributions of these segments. The values are summarized in Table 2, and demonstrate that there are some differences in the stability of the loops, with the connecting loop having the highest energy of the seven loops in the structure. The energies correlate well with RMSD values.

Conclusions

Molecular modelling has shown that a stereochemically plausible and stable structure can be readily constructed for two parallel-topology G-quadruplexes with a drug-binding site at the interface between them. The process has been straightforwardly extended to form a structure with four quadruplexes (Figure 3). In this, the two central quadruplexes do not have a bound drug at their interface, so the channel of K+ ions is continuous. The pattern of external TTA loops is evident along the length of this gently writhing superhelix. It is tempting to suggest that these loops can act as recognition features and play a role in the quadruplex recognition of telomere-associated proteins such as poly(ADP-ribose) polymerase-1 [28] and in the DNA-damage response to the loss of telomere-capping proteins [9].

### Table 2 | RMSD values and relative energies (kcal/mol, where 1 kcal = 4.184 kJ) for the backbone atoms in the loops

<table>
<thead>
<tr>
<th>Segment (residues)</th>
<th>RMSD (Å)</th>
<th>Relative energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loop 1 (4–6)</td>
<td>2.12 (1.04)</td>
<td>–369.30</td>
</tr>
<tr>
<td>Loop 2 (10–12)</td>
<td>2.41 (1.67)</td>
<td>–359.51</td>
</tr>
<tr>
<td>Loop 3 (16–18)</td>
<td>2.28 (1.28)</td>
<td>–362.83</td>
</tr>
<tr>
<td>Connecting loop (22–24)</td>
<td>2.47 (1.36)</td>
<td>–358.49</td>
</tr>
<tr>
<td>Loop 4 (28–30)</td>
<td>2.17 (1.04)</td>
<td>–368.76</td>
</tr>
<tr>
<td>Loop 5 (34–36)</td>
<td>2.10 (1.04)</td>
<td>–368.91</td>
</tr>
<tr>
<td>Loop 6 (40–42)</td>
<td>1.84 (0.95)</td>
<td>–372.75</td>
</tr>
<tr>
<td>TATA tetrad (16,23,24,36)</td>
<td>1.33 (1.10)</td>
<td>–452.14</td>
</tr>
<tr>
<td>GGGG stacking quartet</td>
<td>1.10 (1.31)</td>
<td>–710.67</td>
</tr>
<tr>
<td>(3,9,15,21)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3 | A 93-mer model comprising four quadruplex units and two bound BRACO-19 molecules, constructed from two 45-mer units

Funding

This work was supported by Cancer Research UK [programme grant C129/A4489].

References


Received 2 March 2009 doi:10.1042/BST0370583

© The Authors Journal compilation © 2009 Biochemical Society