The Tail of the Telomere

Luigi Petraccone[1,2], **John O. Trent**[1], and **Jonathan B. Chaires**[1,*]

[1]*James Graham Brown Cancer Center, University of Louisville, 529 S. Jackson St., Louisville, KY 40202*

[2]*Dept. Chimica "P. Corradini", University of Naples "Federico II", 80122 Naples, Italy*

Abstract

The structure of a higher-order G-quadruplex structure for human telomeric DNA is presented. The structure was determined by a novel integrated approach in which molecular dynamics simulations were used to produce a stable structure, from which specific experimentally accessible properties were predicted. These properties were tested by sedimentation velocity and steady-state fluorescence measurements. The structure that emerges is a dimeric structure with two quadruplex units, each with a different structure. The interface between the quadruplex units is stabilized by specific stacking interactions of loop nucleotides. The interface is a unique structure and a unique target for drug design.

Telomeres are structures at the ends of chromosomes that protect chromosomal DNA from degradation, fusion, or recombination[1,2]. Human telomeric DNA consists of long repeats of the sequence 5'-TTAGGG. Several kilobases of that sequence are paired with a complementary strand to form duplex DNA, but approximately 200 nucleotides of the sequence remain unpaired and forms a single-stranded overhang. The telomere serves as a biological clock, and shortens with age. In immortalized cells and in cancer cells, however, telomerase is activated to maintain the length of the telomere[3,4]. Specific proteins interact with the overhang and regulate telomerase activity. Repeats of the 5'-TTAGGG sequence can spontaneously form intramolecular quadruplex structures in solution, with a variety of folded morphologies. There is strong evidence that quadruplex structures exist *in vivo*[5–7], and recent evidence suggests that quadruplexes form in telomeric DNA at specific times in the cell cycle[7,8].

Understanding telomere biology is hampered by the lack of knowledge of the structure of the DNA in the full length overhang. The structure of oligonucleotides containing approximately four repeats of the 5'-TTAGGG sequence have been well characterized under a variety of solution conditions[9–11]. Little is known, however, about the structure of longer sequences of the telomere repeat found in the overhang. No high-resolution structural data are available for these longer sequences, and such sequences may indeed be difficult to study by the usual NMR or crystallographic methods. The structure and stability of the full length overhang must be known in order to fully understand the interactions of the DNA with telomerase and other telomeric proteins and to understand the changes in telomere structure throughout the cell cycle.

We report here the initial results from a novel strategy that integrates molecular dynamics simulations with experimental validation to obtain realistic structural models for higher-order telomeric DNA. High-resolution quadruplex structures obtained for the four-repeat sequence were used to construct multimeric structures containing several quadruplex units. We focus here on dimeric structures containing two folded quadruplex units. These constructs were refined by molecular dynamics simulations to obtain stable structures which were then used

j.chaires@louisville.edu.

to predict experimentally accessible hydrodynamic properties using the HYDROPRO algorithm[12,13] and testable fluorescence properties of specific adenine residues. The experimental results eliminated several possible structures and narrowed the choice to one most probable structure in solution. A similar attempt to build multimeric telomeric DNA structures was recently reported[14], but that study was less comprehensive in its exploration of possible structures, and lacked any experimental validation.

Several different models of dimeric quadruplexes with the $(TTAGGG)_8TT$ sequence were built by incorporating the available high-resolution structures of single quadruplexes reported in K^+ solution. There are two known high-resolution NMR structures for the human telomeric sequence in K^+ solution, hybrid 1 and hybrid 2[15–18]. The structures of the three G-quartets are similar in each; both consist of 2 lateral loops and one double chain reversal loop. The difference is that in hybrid 1 the double chain reversal loop is the first with respect to the 5'-end whereas in the hybrid 2 it is the last loop. There also exists a crystal structure of the human telomere sequence which has a unique symmetrical propeller-type structure with all parallel G-quartets and 3 double chain reversal loops[19]. This structure seems not to be the predominate form in solution[20]. Figure S1A (Supplementary Material) shows these structures. Using these monomeric motifs as building blocks we built five dimeric models that incorporated different combinations of structures, as shown in Figure S1B (Supplementary Material). These are named Hybrid-11 (formed by two hybrid 1 quadruplexes), Hybrid-22 (formed by two hybrid 2 quadruplexes), Hybrid-12 (formed by an hybrid 1 quadruplex at 5'-end and an hybrid 2 quadruplex at the 3'-end), Hybrid-21 (formed by an hybrid 2 quadruplex at 5'-end and an hybrid 1 quadruplex at the 3'-end) and all-propeller (formed by two propeller-type quadruplexes). In the building process, we either maximized or minimized the initial interactions between the two quadruplex units in each model to examine stacked or unstacked possibilities. We refined the initial structures using molecular dynamics. After an equilibration period of 250 ps, models were subjected to an unrestrained molecular dynamics trajectory of 10 ns duration (see Supplementary Material for details). In all of the simulations, the integrity of the individual quadruplexes was preserved throughout the simulation and the major conformational changes were in the relative motions of the quadruplex units. A detailed analysis of the trajectories will be reported elsewhere (in preparation). Here we focus on the main conformational features shown by the simulated models.

Of the five combinations, the optimal interactions and most stable structures were the stacked Hybrid-12 and the propeller models. The propeller model has the ideal geometry to form stable stacking between the two G-tetrad cores, and these stacking interactions were maintained over the whole simulation time. This indicates a very stable conformation, in agreement with a reported computational study on this particular dimer model[14]. In the Hybrid-12 model, there are strong interactions between the third loop of the hybrid 1 quadruplex unit and the second loop of the hybrid 2 quadruplex unit. In particular, the A-15 residue forms very stable stacking interactions with T-38 and T-25 residues for about 90% of the total simulation time (Figure 1A). Weaker interactions between the two quadruplex units were found in the initially stacked Hybrid-21 and Hybrid-11 models, and in the Hybrid-22 model all interfacial interactions were lost during the first nanosecond of the trajectory. These three less stable models also predicted fluorescence and circular dichroism properties that were inconsistent with experimentally observed behavior, as will be described in detail in a subsequent publication.

Figure 1B shows circular dichroism spectra of $(TTAGGG)_8TT$ together with the spectra of the $(TTAGGG)_4$ and $(TTAGGG)_4TT$ sequences reported to form hybrid-1 and hybrid-2, respectively. The CD spectrum of the dimer shows a peak at 290 nm, a major shoulder at 268 nm a minor shoulder at 250 nm and a small negative peak at 240 nm. The shape of this spectrum is consistent with the reported spectra of hybrid-type structures[21]. This observed spectrum of the dimer sequence in solution is inconsistent with a propeller structure, which is predicted to

have a spectrum with a maximum at 260 nm. Hybrid 1 and hybrid 2 quadruplexes have subtly different CD spectra (Figure 1).

The CD spectrum of the dimer closely matches the sum of the CD spectra of the hybrid 1 and hybrid 2 monomers. A more complete comparison of the calculated CD spectra for different models is shown in Figure S2 (Supplementary Material). No other model matches as well as the one shown in Figure 1.

The CD spectrum observed in Figure 1 is consistent with previously reported spectra obtained on long telomere repeat sequences[22–24], and all of these spectra are inconsistent with existence of a parallel-stranded structure in solution.

There are striking differences in loop topologies between the Hybrid-12 and propeller models. Figure 2 shows calculated relative solvent accessible surface areas for loop adenines for the two models. In the propeller model with all double-chain reversal loops, all adenines have similar accessibilities. In contrast, loop adenines show diverse accessibilities in the Hybrid-12 model. Differences in adenine loop residue accessibilities were probed using the fluorescent base analog 2-aminopurine. Single substitutions were made for each adenine within the dimer sequence Figure S3) and these were studied by steady-state fluorescence emission and collisional quenching experiments using acrylamide. The results are shown in Figure 3. It is readily apparent that the loop adenines are in diverse environments, and have widely different solvent accessibilities. These qualitative results are inconsistent with the predictions of the propeller model. Figure 2 shows the correlation between calculated solvent accessible surface areas of loop adenines and accessibility to the acrylamide quencher measured experimentally. There is good agreement between the predicted and actual behavior for the Hybrid-12 model, indicating that this structure can plausibly exist in solution. Such is not the case for the propeller model.

Finally, the predicted and observed distributions of sedimentation coefficients for the Hybrid-12 model are shown in Figure S4 (Supplementary Material). The calculated and observed sedimentation coefficients differ by less than 0.1S, fully consistent with the Hybrid-12 structure. A more complete discussion of the sedimentation properties of all of the constructed models is in preparation.

Our simulations and experiments show that the most plausible structure for telomeric DNA containing eight 5'-TTAGGG repeats is one in which two folded quadruplex units exist, but each with a different folding motif. The two units are not fully independent, and are stabilized by specific interactions at the interface. The interface region forms a unique structure, one that may present a unique target for ligands. We find that while a propeller dimeric structure is computed to be stable, its predicted properties are inconsistent with what is experimentally observed. It is unlikely to be a major species in solution.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Rhodes D, Giraldo R. Curr Opin Struct Biol 1995;5:311–322. [PubMed: 7583629]

2. McEachern MJ, Krauskopf A, Blackburn EH. Annu Rev Genet 2000;34:331–358. [PubMed: 11092831]

3. Cheung AL, Deng W. Front Biosci 2008;13:2075–2090. [PubMed: 17981693]

4. Londono-Vallejo JA. Biochimie 2008;90:73–82. [PubMed: 17728038]

5. Chang CC, Kuo IC, Ling IF, Chen CT, Chen HC, Lou PJ, Lin JJ, Chang TC. Anal Chem 2004;76:4490–4494. [PubMed: 15283592]

6. Granotier C, Pennarun G, Riou L, Hoffschir F, Gauthier LR, De Cian A, Gomez D, Mandine E, Riou JF, Mergny JL, Mailliet P, Dutrillaux B, Boussin FD. Nucleic Acids Res 2005;33:4182–4190. [PubMed: 16052031]

7. Paeschke K, Simonsson T, Postberg J, Rhodes D, Lipps HJ. Nat Struct Mol Biol 2005;12:847–854. [PubMed: 16142245]

8. Paeschke K, Juranek S, Simonsson T, Hempel A, Rhodes D, Lipps HJ. Nat Struct Mol Biol 2008;15:598–604. [PubMed: 18488043]

9. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Nucleic Acids Res 2006;34:5402–5415. [PubMed: 17012276]

10. Lane AN, Chaires JB, Gray RD, Trent JO. Nucleic Acids Res. 2008

11. Patel DJ, Phan AT, Kuryavyi V. Nucleic Acids Res 2007;35:7429–7455. [PubMed: 17913750]

12. Garcia De La Torre J, Huertas ML, Carrasco B. Biophys J 2000;78:719–730. [PubMed: 10653785]

13. Fernandes MX, Ortega A, Lopez Martinez MC, Garcia de la Torre J. Nucleic Acids Res 2002;30:1782–1788. [PubMed: 11937632]

14. Haider S, Parkinson GN, Neidle S. Biophys J 2008;95:296–311. [PubMed: 18375510]

15. Ambrus A, Chen D, Dai J, Bialis T, Jones RA, Yang D. Nucleic Acids Res 2006;34:2723–2735. [PubMed: 16714449]

16. Phan AT, Luu KN, Patel DJ. Nucleic Acids Res 2006;34:5715–5719. [PubMed: 17040899]

17. Luu KN, Phan AT, Kuryavyi V, Lacroix L, Patel DJ. J Am Chem Soc 2006;128:9963–9970. [PubMed: 16866556]

18. Dai J, Carver M, Punchihewa C, Jones RA, Yang D. Nucleic Acids Res 2007;35:4927–4940. [PubMed: 17626043]

19. Parkinson GN, Lee MP, Neidle S. Nature 2002;417:876–880. [PubMed: 12050675]

20. Li J, Correia JJ, Wang L, Trent JO, Chaires JB. Nucleic Acids Research 2005;33:4649–4659. [PubMed: 16106044]

21. Dai J, Carver M, Yang D. Biochimie 2008;90:1172–1183. [PubMed: 18373984]

22. Pedroso IM, Duarte LF, Yanez G, Burkewitz K, Fletcher TM. Biopolymers 2007;87:74–84. [PubMed: 17549693]

23. Vorlickova M, Chladkova J, Kejnovska I, Fialova M, Kypr J. Nucleic Acids Res 2005;33:5851–5860. [PubMed: 16221978]

24. Yu HQ, Miyoshi D, Sugimoto N. J Am Chem Soc 2006;128:15461–15468. [PubMed: 17132013]
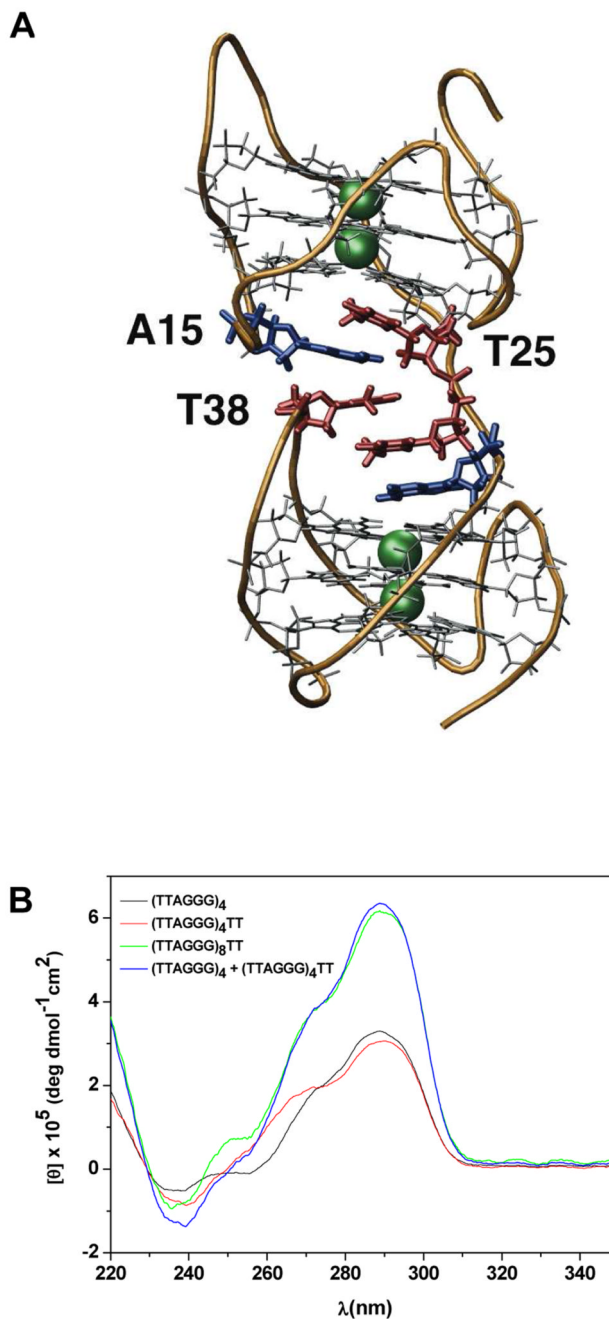
Figure 1.
A) Average structure from MD simulation of the Hybrid-12 model. dT and dA residues, in the junction region between the quadruplex units, are shown in red and blue, respectively. B) CD spectra of quadruplex structures. The $(TTAGGG)_8TT$ spectrum is shown in green, while the hybrid 1 and 2 monomer spectra are shown in black and red, respectively. The sum of the monomer spectra is shown by the blue line.
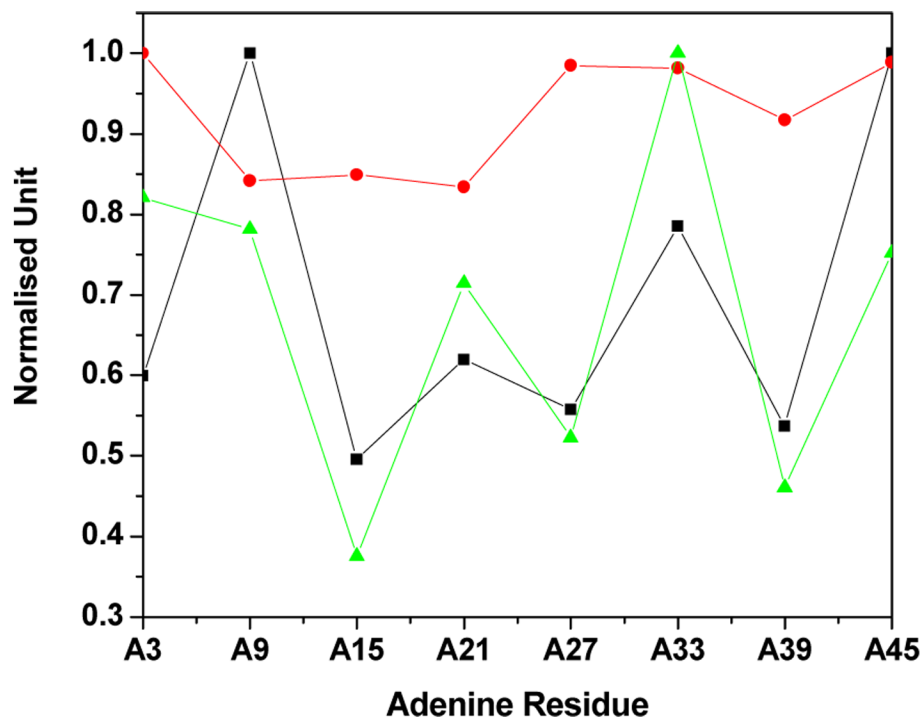
Figure 2.
Correlation of solvent accesible surface areas and expermental collisonal quenching results.
The behavior of the experimentally determined F_0/F value for the oligonucleotides containing
2-aminopurine (green line) is compared with the calculated solvent accesable surface area of
the dA residues over the MD trajectories of the Hybrid-12 (black line) and all-propeller (red
line) models. Each plot was normalized with respect to the highest SASA or F_0/F value.

TT**A3**GGGTT**A9**GGGTT**A15**GGGTT**A21**GGGTT**A27**GGGTT**A33**GGGTT**A39**GGGTT**A45**GGGTT
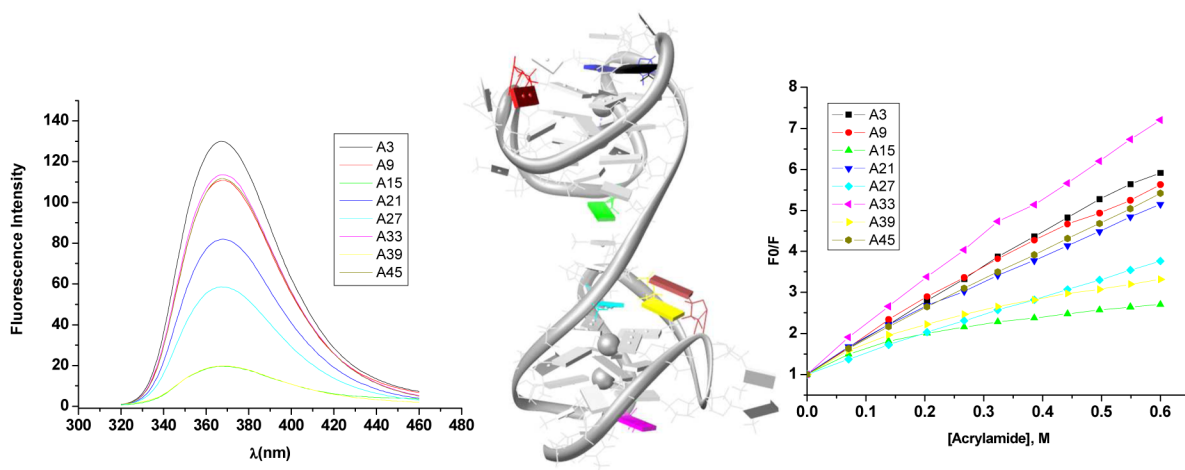

Figure 3.
The results of fluorescence intensity measurements (plots on the left) and fluorescence quenching studies using acrylamide (plots on the right) are shown for oligonucleotides containing 2-aminopurine. The structure of the Hybrid-12 model is also shown (in the middle). The colors indicate the position of the 2-aminopurine substitution as specified in the sequence written on the top. Each experiment was performed in 10 mM potassium phosphate, 100 mM KCl, 0.1 mM EDTA at pH 7.0.