JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

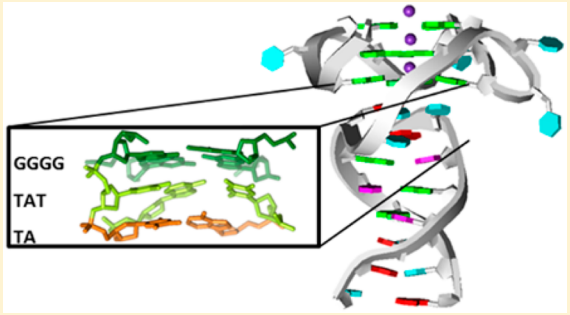
# Structural Insights into the Quadruplex–Duplex 3′ Interface Formed from a Telomeric Repeat: A Potential Molecular Target

Irene Russo Krauss,[‡] Sneha Ramaswamy,[†] Stephen Neidle,[†] Shozeb Haider,[†] and Gary N. Parkinson*[,†]

[†]UCL School of Pharmacy, University College London, London WC1N 1AX, United Kingdom
[‡]Department of Chemical Sciences, University of Naples Federico II, I-80126 Napoli, Italy

Supporting Information

**ABSTRACT:** We report here on an X-ray crystallographic and molecular modeling investigation into the complex 3′ interface formed between putative parallel stranded G-quadruplexes and a duplex DNA sequence constructed from the human telomeric repeat sequence TTAGGG. Our crystallographic approach provides a detailed snapshot of a telomeric 3′ quadruplex–duplex junction: a junction that appears to have the potential to form a unique molecular target for small molecule binding and interference with telomere-related functions. This unique target is particularly relevant as current high-affinity compounds that bind putative G-quadruplex forming sequences only rarely have a high degree of selectivity for a particular quadruplex. Here DNA junctions were assembled using different putative quadruplex-forming scaffolds linked at the 3′ end to a telomeric duplex sequence and annealed to a complementary strand. We successfully generated a series of G-quadruplex–duplex containing crystals, both alone and in the presence of ligands. The structures demonstrate the formation of a parallel folded G-quadruplex and a B-form duplex DNA stacked coaxially. Most strikingly, structural data reveals the consistent formation of a TAT triad platform between the two motifs. This triad allows for a continuous stack of bases to link the quadruplex motif with the duplex region. For these crystal structures formed in the absence of ligands, the TAT triad interface occludes ligand binding at the 3′ quadruplex–duplex interface, in agreement with in silico docking predictions. However, with the rearrangement of a single nucleotide, a stable pocket can be produced, thus providing an opportunity for the binding of selective molecules at the interface.




## INTRODUCTION

The ability of repeating G-rich sequences to form stable, four stranded arrangements in the presence of cations has been recognized since fiber-diffraction studies in 1962.[1] The potential for G-rich sequences, as identified in many key regions of the human genome, to form higher order structures suggests that G-quadruplexes are potential therapeutic targets. One such important class of target resides at the ends of human chromosomes, which are characterized by a long, 3′ single-stranded, G-rich hexanucleic tandem repeat d(TTAGGG) overhang that extends from the self-complementary duplex DNA of the main region of telomeric DNA. Characterization of this overhang has confirmed the in vitro formation of G-quadruplex motifs, which have also been observed in cells.[2,3] G-quadruplexes formed from these telomeric sequences are stable units in vitro, folding into a diverse range of topologies in dilute solution,[4−6] although in the crystalline state[7] and in concentrated solution the parallel form appears to be dominant. High-affinity ligands can induce particular topologies, and some such as the tetrasubstituted naphthalene diimides prefer the parallel form.[8,9]

Intrinsically G-quadruplexes terminate with both 3′ and 5′ G-quartets, providing large planar aromatic surfaces well-suited for binding to high affinity large polyaromatic ligands via $\pi-\pi$ stacking, a chemotype common to many G-quadruplex-selective molecules. Terminal G-quartets are a common feature that ultimately limits specificity to telomeric G-quadruplexes. This feature extends to many G-rich nontelomeric quadruplexes, including the nuclease hypersensitive region of the proto-oncogene c-*MYC*;[10,11] other proto-oncogenes promoters (for example: c-*KIT*, *BCL*-2, VEGF, H-*RAS*, N-*RAS*, K-*RAS*) and many other gene promoters, such as the chicken $\beta$-globin gene and human ubiquitin-ligase RFP2.[12,13]

Furthermore, genome-wide surveys based on quadruplex folding rules have identified a very large numbers of Putative Quadruplex Sequences (PQS) in the human genome, although it is likely that the number actually formed in vivo will be small, and dependent on cell type and status in the cell cycle.[12−14] So, although the G-quartet provides a suitable target for high affinity ligand binding, and some selectivity between the 3′ and 5′ G-quartets surfaces,[15] achieving complete specificity between G-quadruplexes motifs remains a challenge. Thus, there is a case to progress from targeting single distinct quadruplexes toward the identification and characterization of the interfaces formed adjacent to these individual G-quadruplexes, with the

**Received:** October 7, 2015
**Published:** January 5, 2016



DOI: 10.1021/jacs.5b10492
J. Am. Chem. Soc. 2016, 138, 1226−1233

goal of developing ligands that could selectively interact to stabilize these more complex motifs. Telomeres may provide a suitable opportunity to develop new ligands, as within human telomeres one can envisage two simple environments where G-quadruplexes might interact with adjacent structural motifs: a quadruplex-quadruplex junction,[8] and a duplex-quadruplex junction, analogous to that observed in spinach RNA.[16,17] Previous drug discovery efforts have focused on targeting isolated G-quadruplexes, while the duplex-quadruplex junction has not been explored.

We have designed several duplex-quadruplex constructs for structural analysis using X-ray crystallography, with the aim of understanding the junction interface. All the constructs are formed by a strand containing four GGG runs and a 3′ ssDNA modified telomeric sequence 8 nt in length (strand A), and a complementary 8 bp strand B, which was designed to anneal and to form a 3′ duplex motif, generating a gap leaving an unpaired thymine above the G-quartet (Figure 1A). The three

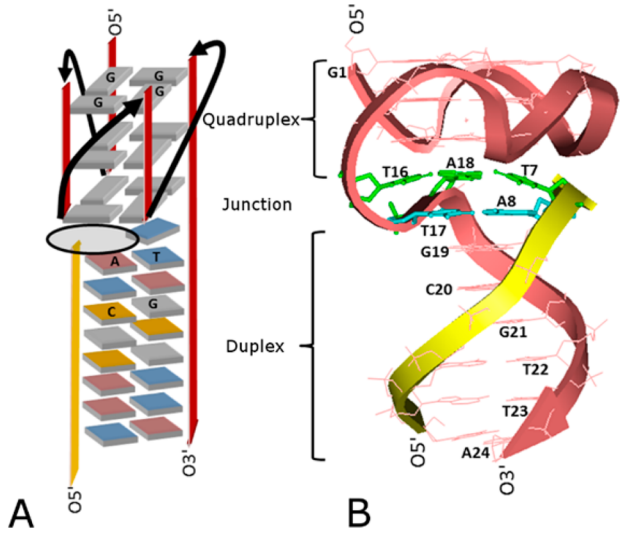



**Figure 1.** Schematic diagrams of the quadruplex–duplex construct as designed (A) and as observed in crystal structures (B). The quadruplex and duplex motifs, the gap in panel A and the junction, a potential site of selective ligand-binding, in panel B are highlighted, along with strand A (pink) and strand B (yellow). In panel B the dinucleotide (cyan) to trinucleotide (green) to G-quartet (pink) are also highlighted and TLOOP sequence is explicitly reported.

constructs differ for loops connecting the GGG runs of the quadruplex motif. The first construct to crystallize and diffract to high resolution contained single T nucleotide loops (TLOOP in Table 1). The core motif has been shown to be stable against thermal denaturation[18] and preferentially folds into a parallel stranded G-quadruplex.[19] The second construct is identical except it has an additional thymine added to the third chain-reversal loop (TT) to uniquely anchor the quadruplex within the crystal lattice breaking the quadruplex 4-fold symmetry: the TTLOOP (Table 1). The third construct contains TTA connecting loops consistent with the human telomeric sequence: the TELO (Table 1). All these three quadruplex-forming sequences are capable of adopting a unimolecular parallel topology with duplex DNA stacked above the quartets forming an interface comprising of a stable TAT triad platform (Figure 1B).

In order to investigate the potential of the G-quadruplex–duplex interface as a target for small molecule ligands, cocrystallizations in the presence of high-affinity quadruplex binding ligands and soaking experiments on preformed crystals were also conducted. Using X-ray crystallography and molecular modeling we report four new structures within two crystal forms. The binding of ligands at the interface between stacked duplex DNA stabilizes crystal packing interactions. Modeling data suggests that for ligands to bind at the G-quadruplex–duplex interface, a structural rearrangement around the TAT triad is required. The cavity formed by the rearrangement of one nucleotide is sufficient to accommodate a planar chromophore with suitable attached chemical groups that can be used to provide selectivity.

## EXPERIMENTAL SECTION

**Sample Preparation, Crystallization and Soaking.** All sequences used in these studies were purchased from Eurofins, HPLC purified: (a) a 24-mer, d(GGGTGGGTGGGTGGGTTA-GCGTTA) where the first three TTA sequences were substituted with single thymines, (b) a 25-mer, d(GGGTGGGTGGGTTGGGTT-AGCGTTA) where the first two TTA sequences were substituted with single thymines and the third one with two thymine residues, (c) a 30-mer containing the human telomeric sequence, d(GGGTTA-GGGTTAGGGTTAGGGTTAGCGTTA), (d) a 8-mer, d-(TAACGCTA) complementary to the 3′ end of the longer sequences (Table 1). The oligonucleotides were initially dissolved in water, to a final concentration of 2 mM for the longer strands and 8 mM for the 8-mer. In all cases, the quadruplex–duplex was assembled by means of a two-step annealing protocol. First, the longer oligonucleotides in the presence of buffer and salts (50 mM potassium chloride and 20 mM potassium cacodylate pH 6.5, diluted to final concentrations of 1.5 mM ssDNA), were heated at 90 °C and slowly allowed to cool to 55 °C, in order to refold the 5′ sequence into a quadruplex. Subsequently, the complementary strand was added (in 20 mM sodium cacodylate buffer at pH 6.5), bringing the final concentrations of DNA to 1 mM. The resulting mixture was slowly cooled to 20 °C overnight to induce duplex formation at the 3′ end. Three different quadruplex–duplex constructs were obtained: TELO, formed by the 30-mer and the complementary 8-mer, TLOOP, formed by the 24-mer and the complementary 8-mer, and TTLOOP, formed by the 25-mer and the complementary 8-mer (Table 1).

A series of crystallization screening trials were set up for the three quadruplex–duplex constructs, both alone and in the presence of ligands, using standard vapor diffusion techniques. Three high-affinity ligands with diverse side-chains were selected, AS1410,[20] BSU6037 and BG32[21,22] (see Supporting Information, Figure S1). Typically, crystallizations were carried out using small sample volumes (0.6 μL, of preformed complex of quadruplex–duplex DNA at concentrations around 0.33 mM, either alone or complexed with ligands at a concentration of 0.25 mM) and combined with the reservoir solutions at 1:1 ratios. Screening of many different conditions allowed us to identify conditions for the growth of crystals of all three constructs:

**Table 1. Sequences of the Oligonucleotides Used to Assemble the Quadruple Duplex Constructs**

| | Strand A | Strand B |
|---|---|---|
| TLOOP | GGGTGGGTGGGTGGGTTAGCGTTA | TAACGCTA |
| TTLOOP | GGGTGGGTGGGTTGGGTTAGCGTTA | TAACGCTA |
| TELO | GGGTTAGGGTTAGGGTTAGGGTTAGCGTTA | TAACGCTA |



DOI: 10.1021/jacs.5b10492
J. Am. Chem. Soc. 2016, 138, 1226–1233

**Table 2. Data-Collection Statistics**

| | TELO | TLOOP | TTLOOP | TLOOP-BSU6037 | TLOOP-FC4ND10 |
|---|---|---|---|---|---|
| Space group | *R*3 | $P42_12$ | *I*222 | $P42_12$ | $P42_12$ |
| Unit-cell parameters | | | | | |
| *a* (Å) | 115.85 | 59.83 | 71.61 | 59.68 | 59.84 |
| *b* (Å) | 115.85 | 59.83 | 101.23 | 59.68 | 59.84 |
| *c* (Å) | 65.32 | 72.31 | 145.45 | 72.65 | 72.73 |
| $\alpha$ (deg) | 90.00 | 90.00 | 90.00 | 90.00 | 90.00 |
| $\beta$ (deg) | 90.00 | 90.00 | 90.00 | 90.00 | 90.00 |
| $\gamma$ (deg) | 120.00 | 90.00 | 90.00 | 90.00 | 90.00 |
| Resolution limits (Å) | 30.00−2.84 (2.94−2.84) | 42.31−2.71 (2.78−2.71) | 45.56−2.79 (2.86−2.79) | 46.12−2.97 (3.15−2.97) | 36.57−3.23 (3.31−3.23) |
| No. of observations | 13355 | 44773 | 43807 | 18379 | 15684 |
| No. of unique reflections | 6862 | 3820 | 12767 | 2937 | 2356 |
| Completeness (%) | 88.8 (65.1) | 98.2(97.0) | 95.5 (98.6) | 93.6 (93.4) | 99.0 (94.6) |
| $I/\sigma(I)$ | 14 (2) | 12.3(3.2) | 12.9 (1.7) | 11.0 (3.2) | 8.5 (3.8) |
| Average multiplicity | 4.9 (4.3) | 6.6(6.8) | 3.4 (3.4) | 6.3 (6.7) | 6.7 (6.9) |
| $R_{merge}$ (%) | 12.8 (44.5) | 13.4(93) | 5.3 (67.9) | 10.5 (48.2) | 17.5 (44.1) |
| $V_M$ (Å$^3$ Da$^{-1}$) | 3.55 | 3.10 | 3.17 | 3.22 | 3.24 |
| No. of molecules in the asymmetric unit | 2 | 1 | 4 | 1 | 1 |
| Solvent content (%) | 76.6 | 73.2 | 73.8 | 74.2 | 74.4 |

**Table 3. Refinement Statistics**

| | TLOOP (5DWX) | TTLOOP (5DWW) | TLOOP-BSU6037 | TLOOP-FC4ND10 |
|---|---|---|---|---|
| Refinement results | | | | |
| Resolution limits (Å) | 42.31−2.71 | 45.56−2.79 | 46.12−2.97 | 36.58−3.23 |
| Number of reflections used in the refinement ($F > 0\sigma(F)$) | 3649 | 12137 | 2797 | 2242 |
| No. of reflections in working set | 3479 | 11508 | 2667 | 2136 |
| No. of reflections in test set | 170 | 629 | 130 | 106 |
| *R*-factor/$R_{free}$ | 0.297/0.329 | 0.208/0.245 | 0.282/0.319 | 0.259/0.338 |
| No. of oligonucleotide atoms | 485 | 3189 | 673 | 649 |
| No. of ions | 3 | 10 | 3 | 3 |
| No. of ligand atoms | – | – | – | 42 |
| RMSD from ideal values | | | | |
| Bond lengths (Å) | 0.01 | 0.01 | 0.01 | 0.01 |
| Bond angles (deg) | 1.70 | 1.55 | 1.58 | 1.62 |
| Average B-factors (Å$^2$) | | | | |
| Oligonucleotide | 66.56 | 93.46 | 59.64 | 94.54 |
| Ions | 23.96 | 32.38 | 20.57 | 5.93 |
| Ligand | – | – | – | 160.75 |

TELO, TLOOP and TTLOOP. In particular, well-formed crystals of all the three quadruplex−duplex constructs were grown at 12 °C from crystallization solutions containing 2 M lithium sulfate, 50 mM Tris/HCl pH 8.8 and 2 mM $CuCl_2$. Crystals of the TLOOP and TTLOOP were successfully grown without the presence of ligands while TELO crystals grew only in the presence of the compound BSU6037.

The 5′-end of the quadruplex would better reflect the context of the telomere where G-quadruplex is formed by the telomeric 3′-overhang, However, the main focus of this study was to explore the structure of the duplex-quadruplex junction and plausible ligand binding at the interface. Detailed analysis of the human telomeric quadruplex structures in the PDB revealed that there is a 5′−5′ stacking preference for parallel stranded quadruplex.[23] These interactions would be paramount in stabilizing the crystal lattice. It was assumed that in order to retain this stacking interaction, the quadruplex−duplex junction would have to be at the 3′ end. Additional details of crystal structures in the PDB, where ligands were bound to the human telomeric propeller-type quadruplex highlighted the preference of the ligands to bind to the 3′ surface.[8,9] On the basis of these observations, we envisaged that a disubstituted ligand would bind to the junction and stabilize the 3′ interface, by stacking on one-half of the quartet[8] and directly interacting with residues from the duplex should resolve the duplex-quadruplex junction.

In an attempt to improve the quality of preformed TLOOP crystals, several ligands were added by soaking. Soaking solutions at 10 mM ligand concentration also contained 25% glycerol as a cryo-protecting agent. Compounds known to bind strongly to G-quadruplexes were used: TMPyP4, tetra-substituted naphthalene diimides- and acridine-derived ligands. Most of these molecules are vividly colored, thus the soaking process can be followed by visual inspection of the crystals using a light microscope. In some cases, the soaking caused progressive damage to a crystal, detectable as cracking of crystal surfaces and reduction of diffraction quality. Thus, as soon as crystals changed color, they were flash-frozen to prevent dissolution.

**Data Collection and Structure Determination.** Crystals of free and complexed quadruplex−duplex DNAs were flash-frozen in liquid nitrogen. Cryoprotection with the addition of 25% glycerol was used in most of the flash-freezing experiments. Diffraction data were collected


DOI: 10.1021/jacs.5b10492
J. Am. Chem. Soc. 2016, 138, 1226−1233

at the Diamond Light Source synchrotron (beamline I04−1). All data sets were processed and scaled using the XDS, SCALA and XIA2 programs. TLOOP crystals belong to the tetragonal $P42_12$ space group, TELO crystals to the rhombohedral $R3$ space group and TTLOOP crystals to the orthorhombic $I222$ space group. All the crystals diffract X-rays up to about 3.0−2.7 Å resolution. Matthew's coefficient calculations indicated the presence of one TLOOP, two TELO and four TTLOOP quadruplex−duplex constructs in the asymmetric units, respectively. In all cases the solvent content was very high (>70%). Details of data collections are in Table 2.

The TLOOP and TTLOOP crystal structures were successfully solved by molecular replacement techniques using the PHASER program[24] and the G-core of the native telomeric quadruplex crystal structure (PDB-ID 1KF1) as a search model. Model building and refinement were performed using COOT[25] and REFMAC5[26] programs. Initial 2Fo−Fc maps showed clear electron density for G-quartets and potassium ions, as well as residual density for thymines of the loops. Refinement statistics are reported in Table 3. Attempts to solve the structure of the TELO-BSU6037 complex by molecular replacement methods and by soaking experiments with heavy metals such as platinum have failed to date.

The coordinates of the TLOOP and TTLOOP structures have been deposited in the Protein Data Bank with codes 5DWX and 5DWW, respectively.

**Molecular Modeling.** The quadruplex−duplex TTLOOP crystal structure was used as a starting point for modeling studies. T17 and A19 from the G-rich strand and T7 from the complementary strand constitute the TAT triad at the quadruplex−duplex interface. The pseudo-ligand-binding site at the interface was generated by rotating the phosphodiester backbone of T17, while maintaining strand polarity of the G-rich strand. The cavity generated has a volume of 421 Å$^3$. The backbone atoms were minimized employing 1000 steps of conjugate gradient energy minimization to relieve any structural distortions. The chemical structure of the BSU6037 ligand was built and docked using the ICM-Pro software package.[27] Grid maps were generated to encompass the pseudo-ligand-binding site at the interface. Docking was carried out using the automated docking module of the ICM software. The best docked structure was chosen based on the highest calculated binding energy.

## RESULTS AND DISCUSSION

Crystals for quadruplex−duplex constructs were obtained for TELO, which mimics the human telomeric sequence; TLOOP, which has three single-residue loops, and TTLOOP, where the third quadruplex loop comprises two thymines. The two quadruplex−duplex constructs solved (TLOOP and TTLOOP) confirm the expected self-assembly into a G-quadruplex motif connected to canonical B-form DNA, which is stacked coaxially to the 3′ G-quartet. The G-rich sequences adopt the expected unimolecular parallel G-quadruplex topology[7,28] comprised of three stacked quartets, as observed in the case of the human telomeric sequences. The last seven nucleotides at the terminal 3′ end base pair with the complementary B strand, in an antiparallel arrangement with five bases forming regular canonical B-form DNA (see Table S1). The TLOOP construct has one quadruplex−duplex in the crystallographic asymmetric unit, with positional disorder about the 4-fold axis of the quadruplex and poorer resolution at the 3′ end of the duplex DNA. The crystal structure of the TTLOOP construct comprises four well-defined and independent quadruplex−duplex structures in the asymmetric unit, with well-resolved duplex DNA. The TELO structure is predicted to have two quadruplex−duplexes and two ligands in the ASU based on cell volume.

**The Free TLOOP Structure.** The crystal structure clearly shows that the first 15 residues of the 24-mer (chain A) of TLOOP adopt a parallel-stranded intramolecular G-quadruplex fold, with three propeller loops formed by the three thymines, and three potassium ions interacting with the G-quartets (Figure 2), consistent with other G-quadruplex structures. This

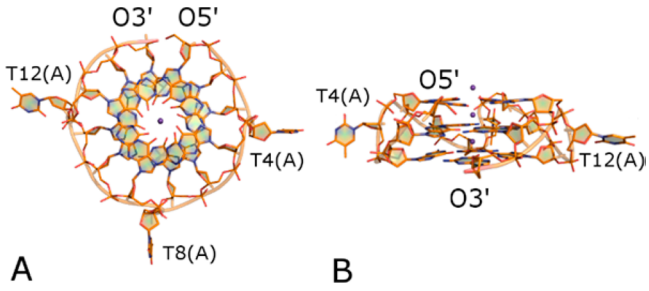



**Figure 2.** Top view (A) and lateral view (B) of the quadruplex region of TLOOP. Potassium ions are also shown as purple spheres. Loop residues are explicitly marked.

data confirms that the shortening of connecting loops, a single thymine in the place of a TTA segment, does not prevent the formation of the parallel G-quadruplex structure. On the contrary, the global fold of this region closely resembles that of the human telomeric quadruplex, as observed in the crystal structures and in crowded solution.[29,30] The quadruplex motif is extensively involved in crystal packing; in particular, a symmetry-related molecule is 5′−5′ stacked on the first G-quartet, with a potassium ion sandwiched between the two quadruplexes. The thymines of the loops also interact with symmetry mates, especially with T4 and T8 (Figure S2). The close packing of G-quadruplexes could explain the improved quality of electron density in this region as opposed to that observed toward the 3′ end of the duplex region.

The formation of the junction between quadruplex and duplex regions is consistent with preliminary computational studies.[31] The high-resolution crystallographic data reveals that T17 (the second nucleotide in the sequence after the quadruplex) bulges out to allow the pairing of T16 and A18, which stack on the 3′-quartet of the quadruplex. A residue from the complementary strand (T7) also interacts with this T−A base pair forming a TAT triad. Downstream of A18 the electron density maps are not continuous and possibly indicate disordering in the structure. We also observe electron density for a fourth propeller T-loop that could only be the result of static positional disorder of the assembled molecules in the crystal lattice. We surmise that the positional disorder would adversely affect the clarity of the duplex DNA and interface.

**The Free TTLOOP Structure.** In order to remove the contribution of static disorder observed in TLOOP crystals, we designed a new variant, TTLOOP, with an aim of anchoring the quadruplex to one orientation in the crystal lattice. We chose to insert an additional thymine in the third loop, since the first and the second loops (T4 and T8, respectively) in the TLOOP structure were extensively involved in packing interactions. We generated well-formed TTLOOP crystals that diffracted X-rays beyond 2.80 Å and the structure was solved by molecular replacement. The final structure was refined to $R/R_{free}$ values of 0.215/0.253. Four TTLOOP quadruplex−duplex constructs are present in the asymmetric unit (Figure 3A); they arrange themselves in two layers and interact as pairs through the same 5′−5′ stacking interaction found between TLOOP symmetry mates. Also, the interactions between the thymines of the loops observed in the TLOOP structure are mostly preserved. However, in the case of


DOI: 10.1021/jacs.5b10492
J. Am. Chem. Soc. 2016, 138, 1226−1233

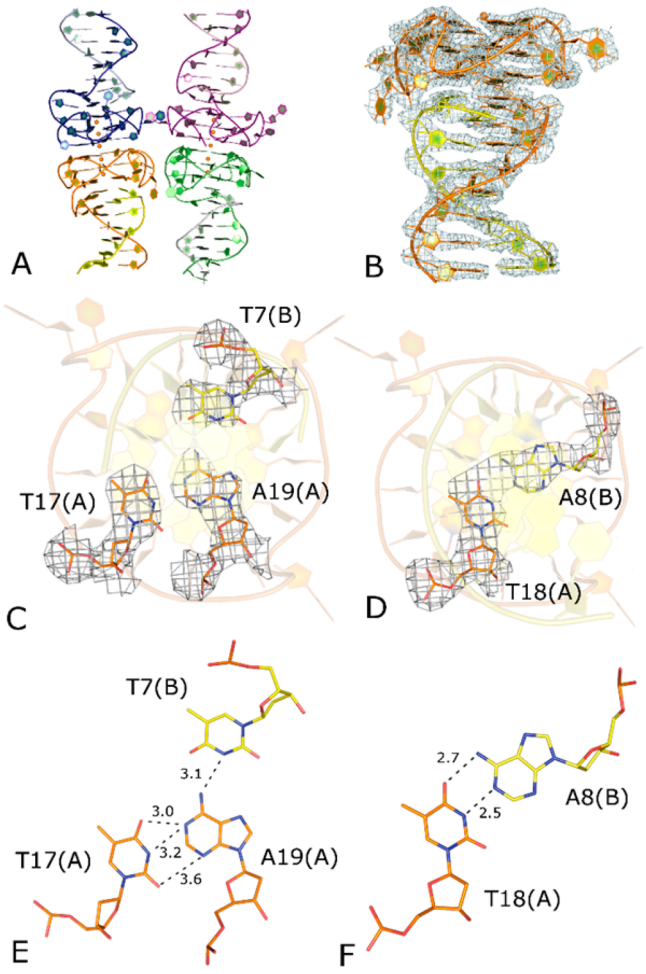



**Figure 3.** TTLOOP structure. (A) Cartoon representation of the four TTLOOP molecules in the asymmetric unit. Different chains are marked in different colors. Potassium ions are shown as orange spheres. (B) 2Fo−Fc electron density maps of the quadruplex and duplex regions of TTLOOP contoured at 2.0 and 1.5 $\sigma$ level, respectively. (C) Top view of the interface triad. 2Fo−Fc electron density map contoured at 1.9 $\sigma$ level is also shown. (D) Top view of the first duplex pair. 2Fo−Fc electron density map contoured at 1.9 $\sigma$ level is also shown. (E) H-bond interactions among residues of the triad. (F) H-bond interactions between residues of the first duplex pair. Distances are reported in angstrom.

TTLOOP, electron density maps are clear and continuous not only for the quadruplex moiety, but also for the duplex (Figure 3B), which allowed unambiguous rebuilding of all the quadruplex−duplexes, with the exception of one base pair in one of the four molecules.

The interface between quadruplex and duplex regions is enhanced over the TLOOP structure, providing four identical and well-defined interfaces consistent with that observed in the TLOOP structure. A TAT triad composed of T17, A19 of the A chain (corresponding to T16 and A18 in the TLOOP, respectively) and T7 of the B-strand bridges between the duplex and the quadruplex motifs (Figure 3C and E). Thus, T18(A) and A8(B), two residues that in the sequence follow the interface residues A19(A) and T7(B), form the first base pair of the duplex region (Figure 3D).

**Addition of Ligands.** In an attempt to identify ligands that could associate with the quadruplex−duplex interface other than peptides,[32] we chose a diverse range of G-quadruplex stabilizing ligands previously characterized by X-ray crystallography. These included acridines (BRACO-19 and its derivatives), naphthalene diimides, macrocyclic scaffolds, fluoroquinolone derivatives and porphyrins. On the basis of our molecular modeling studies, several of these ligands were selected for cocrystallization experiments: BSU6037, a disubstituted acridine derivative analogous to ligand BSU6039 (PDB-ID 1L1H);[33] BRACO-19 (PDB-ID 3CE5),[8] a trisubstituted acridine derivative; TMPyP4, a tetra-(*N*-methyl-4-pyridyl) porphyrin (PDB-ID 2HRI);[34] and FC4ND10,[35] a two dimethylamine, two hydroxyl tetra-substituted naphthalene diimide derivative, (PDB-ID 3T5E).[36] These ligands have also shown telomerase inhibitory activity in enzymatic assays.[8,37,38] During cocrystallization experiments, we were successful in generating yellow crystals of the TELO construct in the presence of BSU6037, but were not able to generate diffraction quality crystals for the other constructs.

We also pursued ligand soaking experiments with preformed crystals to determine if the complexes could be formed. The TLOOP had the most open lattice and several ligands were soaked into preformed TLOOP crystals. Ligands that did not significantly degrade crystal quality but appeared to bind to the construct were the naphthalene diimide FC4ND10, the porphyrin TMPyP4, a copper-porphyrin ligand, and the acridine BSU6037. Diffraction data were collected for all these complexes. Binding of TMPyP4 caused a large decrease in diffraction quality (from about 3.0 Å to 4.3 Å resolution) and a slight change of cell dimensions, whereas data collected on the FC4ND10 complex (3.23 Å resolution), on the copper derivative (3.15 Å resolution) and on the BSU6037 complex (2.91 Å resolution) appeared of good quality. Analysis of difference Fourier maps did not show the presence of BSU6037 or Cu-porphyrin bound to the quadruplex−duplex, but the quality of electron density maps for the duplex region of BSU6037 soaked crystals significantly improved, allowing the complete DNA molecule to be build (Figure S3). A residual density assignable to the ligand was observed for the FC4ND10 complex, as well as residual electron density in the duplex region.

**The TLOOP-FC4ND10 Complex Structure.** In the case of the TLOOP-FC4ND10 complex, a careful analysis of electron density maps allowed the manual model building of the quadruplex−duplex interface and provided some structural information about the duplex region and the position of the ligand. The final model was refined to R/Rfree values of 0.26/0.34. It should be stressed that the mobility of the duplex region is very high, as expected on the basis of the loose packing. This mobility results in high thermal factor values and poor quality of electron density maps. Static disorder can also be a cause of the low quality of electron density maps. Indeed, since the quadruplex region is highly symmetric whereas the duplex region is not, a different orientation of the quadruplex−duplex in the crystal would have a dramatic effect on maps of the duplex segment and not on those of the quadruplex one. Thus, even if it is a reliable model, the duplex region should be seen as a trace, since it lacks the accuracy of the quadruplex segment.

Our X-ray structure shows that the ligand does not interact with the quadruplex region, neither at the 5′ end, which is involved in packing contacts, nor at the 3′ end. FC4ND10 appears to be placed on the 4-fold symmetry axis at the end of the duplex segment and to form stacking interactions with a symmetry-related ligand molecule on one side, while it loosely


DOI: 10.1021/jacs.5b10492
J. Am. Chem. Soc. 2016, 138, 1226−1233

contacts the 3′-terminal ends of two symmetry-related TLOOP quadruplex−duplexes on the opposite side (Figure 4A).

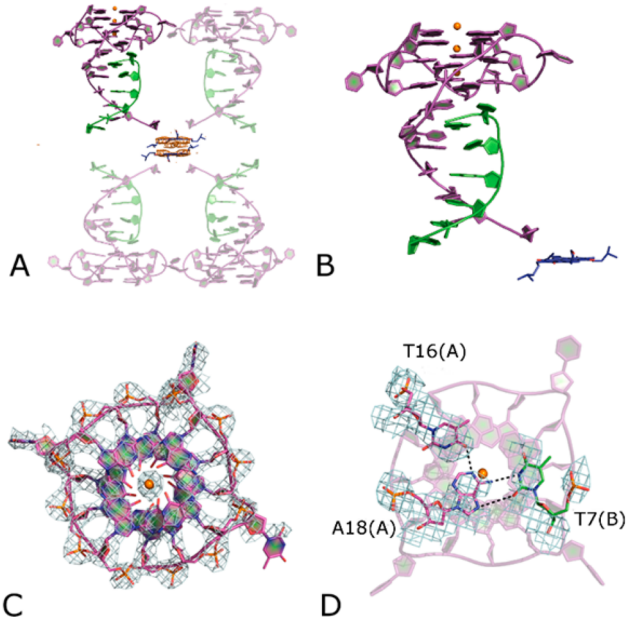



**Figure 4.** TLOOP-FC4ND10 structure with the quadruplex−duplex DNA represented as cartoon (longer strand A in magenta, complementary strand B in green), the ligand represented as stick in blue, the potassium ions as orange spheres. (A) Packing interactions: ligand molecule is shown placed on the 4-fold symmetry axis and interacts with the 3′-end of strand A. Symmetry related molecules are represented slightly faded. 2Fo−Fc electron density map of the ligand contoured at 1.5 $\sigma$ level is also shown in orange. (B) Close-up with one molecule in the ASU showing the potassium ions and ligand FC4ND10 stacked onto the duplex DNA. (C) 2Fo−Fc electron density map of the quadruplex region contoured at 1.5 $\sigma$ level. (D) 2Fo−Fc electron density map of residues belonging to the interface triad contoured at 1.0 $\sigma$ level.

The weak interaction between the naphthalene diimide and chain A of the duplex segment appears to be sufficient to cause a slight ordering of this region. Thermal factor values of the complementary strand (chain B) are even higher than those observed for the 3′ flanking end of the chain A. A model of this strand has been built taking in account electron density and base pairing (Figure 4B). Both the quadruplex region and the junction residues have rather clear electron density (Figure 4C) and lower thermal factors than duplex residues. As also observed in the case of free TTLOOP quadruplex−duplex, T16(A), A18(A) and T7(B) form a TAT triad stacking on the 3′ G-quartet (Figure 4D). This architecture seems to occlude the binding of ligands at the interface between the two structural motifs.

**Ligand Docking into the Quadruplex−Duplex Interface.** Since no structural data is available on ligand binding at DNA quadruplex−duplex interfaces, we used molecular modeling to investigate the likely binding features of the hybrid junction. The loops of human telomeric quadruplexes are very flexible,[39] particularly the ones that connect segments together.[40−42] For a ligand to bind at the interface, the TAT triad has to be disrupted, which would expose the G-quartet surface. Indeed disruption of a stable AUA triad above a G-quartet (observed in the apo state) and rearrangement has been observed with spinach RNA resulting from the binding of a fluorophore onto the G-quartet.[16] We constructed a pseudo-ligand-binding site by rotating T17(A) toward the solvent. The site is large enough to accommodate a planar chromophore, which can exploit different chemical features within the cavity. The disubstituted acridine, BSU6037, when docked in the pseudo-ligand-binding site sits on one-half of the quadruplex, making $\pi$-stacking interactions with G11, G16 and T18 (Figure 5). The central protonated nitrogen of the acridine

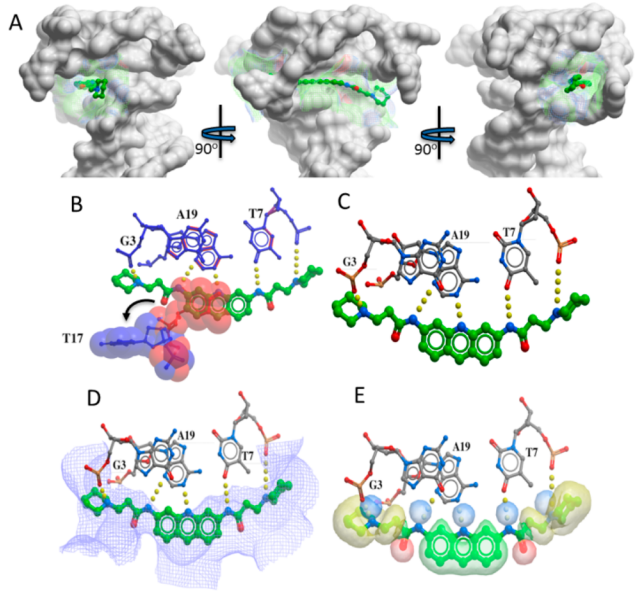



**Figure 5.** Structural features around the pseudo-ligand-binding site. The ligand is colored in green; hydrogen bonds are represented as yellow dotted lines. (A) The spatial position of the binding site at the quadruplex−duplex interface. (B) The T17 base is flipped out (blue) of the TAT triad (red), generating a cavity where planar chromophores can be accommodated. The position of the backbone atoms and strand polarity was retained. (C) The ligand (BSU6037) exploits several features within the binding site making five hydrogen bonds. (D) Negative occlusion space (blue) is available for the ligand in the pseudo-ligand-binding site. (E) Pharmacophore features in BSU6037 that complement chemical groups in the binding site, highlighting hydrogen bond donor (blue), hydrogen bond acceptor (red), hydrophobic (yellow) and aromatic (green) groups.

chromophore lies in-plane with A19(A)-T7(B) base pair and makes hydrogen bonds with N1 atom of A19(A), while the amide nitrogen on one of the side chains makes hydrogen bonds with the carbonyl oxygen atom O4 of T7(B) and the other amide nitrogen with the N3 atom of A19(A) (Figure 5). This arrangement is closely analogous to that observed in the crystal structure of the complex of BSU6039 with the *Oxytricha* bimolecular quadruplex.[33] Two additional hydrogen bonds are formed between the protonated nitrogen atoms of the piperidine rings and the phosphate backbone. The quadruplex−duplex interface can accommodate this type of disubstituted acridine exploiting hydrogen bonding, and stacking interactions. However, tetrasubstituted napthaline diimide ligands, such as FC4ND10, cannot be so readily accommodated at the quadruplex−duplex interface due to size restrictions and steric clashes showing the potential for selectivity.

**The G-Quadruplex−Duplex Interface.** A review of the few structures currently available[43−45] in the PDB reveals the expected dominance of base-stacking at the interface, with all


DOI: 10.1021/jacs.5b10492
J. Am. Chem. Soc. 2016, 138, 1226−1233

structures having the expected helical geometries associated with quadruplex and duplex DNA topologies, the retention of the major and minor grooves, and a regular helical twist and standard rise, consistent with our structures reported herein (see Supporting Information, Table S1). The artificial quadruplex–duplex hybrid named Construct-I[43] consistently shows the planar G-quartet transitioning via a simple undistorted stacked C:G base-pair, continuing on into regular duplex DNA. Structure PDB-ID 2M8Z[45] is typical of the deposited structures, while PDB-ID 2M92 has an additional roll at the transition step. We observe a more complex transition that appears to maximize the base stacking at each step (Table S2). A nonrelated RNA G-quadruplex structure (PDB-ID 2LA5)[44] shows the RNA bound to a peptide, which has a transition similar to our TAT triad consisting of a transition from the terminal G- quartets, via a C:G:A triad into a C:G base pair-peptide complex and finally into the duplex region. Although displaying a more complex interface, the same geometric principles guiding nucleic acids are seen in this structure.

**Formation of the TAT Triad.** The consistency of the TAT triad formation at the G-quadruplex–duplex interface, as seen within the TTLOOP and TLOOP crystal forms and in all the molecules in the ASU, adds weight to the suggestion that the triad is an important structural feature of the quadruplex–duplex interface. All three residues (T17(A)-A19(A)-T7(B)) are well resolved and bridge the duplex and the quadruplex motifs (Figure 3C and 4C). This junction allows the formation of a regular duplex and the continuous stacking of bases from the quadruplex motif transitioning into the duplex. It would appear that the duplex DNA, comprising of strand B residues 1–6 paired to strand A residues 20–25, in the TTLOOP structure, forms a stable duplex with the CGC pairings providing rigidity. The collapse of this duplex end onto the G-quartet places residues A19(A)-T7(B) onto the quartet, allowing residue T17(A) to pair and complete the triad. The remaining unpaired nucleotide T18(A) folds back in, to pair with the free A8(B) and form the basis for further stacking (Figure 3D and F). The described folding pathway implies that the next adenine in the sequence A9(B) will project away from the duplex and not play a role in stabilizing the quadruplex–duplex interface (Figure S4).

All the structural features described for TTLOOP are also present in the TLOOP construct, and more clearly shown by the TLOOP soaked crystals. The ligand bound structure of the TLOOP-FC4ND10, shown in Figure 4A and 4B, with the ligand associated at the 3′ ends of the duplex DNA and not within the preformed quadruplex–duplex interface, and the TLOOP-BSU6037 soaked crystals as revealed by the well resolved and consistent electron density maps shown along on the entire DNA molecule (Figure S3).

## CONCLUSIONS

The structures reported here not only demonstrate the presence of typical DNA geometries, (parallel quadruplex and regular B-form duplex) within a hybrid quadruplex–duplex construct, but, more importantly, also reveal in detail the unique TAT triad interface. We show that the human telomeric DNA sequence forms a single stable interface, as revealed by X-ray crystallography and molecular dynamic simulations.[31] Ligand binding has been shown to readily occur at the ends of duplex DNA; none of the solved structures shows the binding of a ligand stacked on the G-quartets. However, using molecular modeling to explore the quadruplex–duplex interface, we have shown the potential for the formation of a selective ligand pocket. The pocket opening requires a slight structural rearrangement of the TAT triad, similar to that previously observed in the case of ligand-induced conformational changes of TTA loops of telomeric G-quadruplexes. Structural analysis of this potential pocket highlights that it is sufficient for the binding of both di- and trisubstituted acridine ligands, but not of the larger tetrasubstituted naphthalene diimide ligands, thus having a significantly higher selectivity for ligands than do isolated quadruplexes. It is notable that the tetrasubstituted naphthalerne diimides would not appear to function in cells via telomere maintenance mechanisms, in accord with these conclusions.

These structures may thus represent more biologically relevant models for ligand lead discovery and optimization than simple isolated quadruplexes.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b10492.

Additional tables presenting helical parameters for duplex region of TTLOOP and stacking interactions within TTLOOP structure. Additional figures for structures of ligands, lateral and top views of the symmetry contact for TLOOP, schematic representation of the TTLOOP interface, and electron density map of the TLOOP-BSU6037 construct. (PDF)

## AUTHOR INFORMATION


### Corresponding Author

*gary.parkinson@ucl.ac.uk


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Dr. Julia Viladoms for data collection of the TTLOOP construct during her participation in the DLS-CCP4 Data Collection & Analysis Workshop 2014, as well as for critical reading of the manuscript. The Career Excellence Fellowship in Computational Medicinal Chemistry, UCL School of Pharmacy, for supporting Dr. Shozeb Haider. Associazione Italiana di Crystallografia (AIC) is gratefully acknowledged for the Research Fellowship that allowed Dr. Irene Russo Krauss to spend two months in the group of Dr. Gary N. Parkinson.

## REFERENCES


(1) Gellert, M.; L, M.; Davies, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48*, 2013.

(2) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182.

(3) Henderson, A.; Wu, Y.; Huang, Y. C.; Chavez, E. A.; Platt, J.; Johnson, F. B.; Brosh, R. M., Jr.; Sen, D.; Lansdorp, P. M. *Nucleic Acids Res.* **2014**, *42*, 860.

(4) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263.

(5) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723.

(6) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963.

(7) Parkinson, G. N.; Lee, M. P.; Neidle, S. *Nature* **2002**, *417*, 876.




DOI: 10.1021/jacs.5b10492
J. Am. Chem. Soc. 2016, 138, 1226–1233

(8) Campbell, N. H.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. *J. Am. Chem. Soc.* **2008**, *130*, 6722.
(9) Parkinson, G. N.; Cuenca, F.; Neidle, S. *J. Mol. Biol.* **2008**, *381*, 1145.
(10) Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.* **1998**, *26*, 1167.
(11) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11593.
(12) Todd, A. K.; Johnston, M.; Neidle, S. *Nucleic Acids Res.* **2005**, *33*, 2901.
(13) Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908.
(14) Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S. *Nat. Biotechnol.* **2015**, *33*, 877.
(15) Le, D. D.; Di Antonio, M.; Chan, L. K.; Balasubramanian, S. *Chem. Commun. (Cambridge, U. K.)* **2015**, *51*, 8048.
(16) Huang, H.; Suslov, N. B.; Li, N. S.; Shelke, S. A.; Evans, M. E.; Koldobskaya, Y.; Rice, P. A.; Piccirilli, J. A. *Nat. Chem. Biol.* **2014**, *10*, 686.
(17) Warner, K. D.; Chen, M. C.; Song, W.; Strack, R. L.; Thorn, A.; Jaffrey, S. R.; Ferre-D'Amare, A. R. *Nat. Struct. Mol. Biol.* **2014**, *21*, 658.
(18) Guedin, A.; Gros, J.; Alberti, P.; Mergny, J. L. *Nucleic Acids Res.* **2010**, *38*, 7858.
(19) Rachwal, P. A.; Brown, T.; Fox, K. R. *FEBS Lett.* **2007**, *581*, 1657.
(20) Gunaratnam, M.; Green, C.; Moreira, J. B.; Moorhouse, A. D.; Kelland, L. R.; Moses, J. E.; Neidle, S. *Biochem. Pharmacol.* **2009**, *78*, 115.
(21) Harrison, R. J.; Gowan, S. M.; Kelland, L. R.; Neidle, S. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2463.
(22) Guyen, B.; Schultes, C. M.; Hazel, P.; Mann, J.; Neidle, S. *Org. Biomol. Chem.* **2004**, *2*, 981.
(23) Do, N. Q.; Phan, A. T. *Chem. - Eur. J.* **2012**, *18*, 14752.
(24) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* **2007**, *40*, 658.
(25) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486.
(26) Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355.
(27) Abagyan, R.; Totrov, M.; Kuznetsov, D. *J. Comput. Chem.* **1994**, *15*, 488.
(28) Do, N. Q.; Lim, K. W.; Teo, M. H.; Heddi, B.; Phan, A. T. *Nucleic Acids Res.* **2011**, *39*, 9448.
(29) Heddi, B.; Phan, A. T. *J. Am. Chem. Soc.* **2011**, *133*, 9824.
(30) Petraccone, L.; Malafronte, A.; Amato, J.; Giancola, C. *J. Phys. Chem. B* **2012**, *116*, 2294.
(31) Ramaswamy, S. Master's thesis, The School of Pharmacy, London, 2011.
(32) Vasily ev, N.; Polonskaia, A.; Darnell, J. C.; Darnell, R. B.; Patel, D. J.; Serganov, A. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E5391.
(33) Haider, S. M.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2003**, *326*, 117.
(34) Parkinson, G. N.; Ghosh, R.; Neidle, S. *Biochemistry* **2007**, *46*, 2390.
(35) Cuenca, F. G.; Gunaratnam, M. O.; Haider, S.; Munnur, D.; Nanjunda, R.; Wilson, W. D.; Neidle, S. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1668−1673.
(36) Collie, G. W.; Promontorio, R.; Hampel, S. M.; Micco, M.; Neidle, S.; Parkinson, G. N. *J. Am. Chem. Soc.* **2012**, *134*, 2723.
(37) Gunaratnam, M.; Greciano, O.; Martins, C.; Reszka, A. P.; Schultes, C. M.; Morjani, H.; Riou, J. F.; Neidle, S. *Biochem. Pharmacol.* **2007**, *74*, 679.
(38) Martins, C.; Gunaratnam, M.; Stuart, J.; Makwana, V.; Greciano, O.; Reszka, A. P.; Kelland, L. R.; Neidle, S. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2293.
(39) Collie, G. W.; Campbell, N. H.; Neidle, S. *Nucleic Acids Res.* **2015**, *43*, 4785.
(40) Islam, B.; Sgobba, M.; Laughton, C.; Orozco, M.; Sponer, J.; Neidle, S.; Haider, S. *Nucleic Acids Res.* **2013**, *41*, 2723.
(41) Haider, S.; Parkinson, G. N.; Neidle, S. *Biophys. J.* **2008**, *95*, 296.
(42) Haider, S. M.; Neidle, S. *Biochem. Soc. Trans.* **2009**, *37*, 583.
(43) Lim, K. W.; Khong, Z. J.; Phan, A. T. *Biochemistry* **2014**, *53*, 247.
(44) Phan, A. T.; Kuryavyi, V.; Darnell, J. C.; Serganov, A.; Majumdar, A.; Ilin, S.; Raslin, T.; Polonskaia, A.; Chen, C.; Clain, D.; Darnell, R. B.; Patel, D. J. *Nat. Struct. Mol. Biol.* **2011**, *18*, 796.
(45) Lim, K. W.; Phan, A. T. *Angew. Chem., Int. Ed.* **2013**, *52*, 8566.


DOI: 10.1021/jacs.5b10492
*J. Am. Chem. Soc.* 2016, 138, 1226−1233