Structural Insights into the Quadruplex-Duplex 3' Interface formed from a Telomeric Repeat: A Potential Molecular Target.

Irene Russo Krauss[2], Sneha Ramaswamy[1], Stephen Neidle[1], Shozeb Haider[1], Gary N Parkinson[1]*.

[1]UCL School of Pharmacy, University College London, London WC1N 1AX, United Kingdom. [2]Department of Chemical Sciences, University of Naples Federico II, Napoli, Italy.

ABSTRACT: We report here on a X-ray crystallographic and molecular modelling 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. In this study DNA junctions were assembled using different putative quadruplex-forming scaffolds that were 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, the structural data reveal the consistent formation of a TAT triad platform, at the transition region between the two motifs. The triad allows a continuous stack of bases to be formed which links 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 (Gellert M 1962). 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 (Biffi, Tannahill et al. 2013) (Henderson, Wu et al. 2014). G-quadruplexes formed from these telomeric sequences are stable units in vitro, folding into a diverse range of topologies in dilute solution (Wang and Patel 1993) (Ambrus, Chen et al. 2006) (Luu, Phan et al. 2006), although in the crystalline

state (Parkinson, Lee et al. 2002) 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 (Campbell, Parkinson et al. 2008) (Parkinson, Cuenca et al. 2008).

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 π-π 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 non-telomeric quadruplexes, including the nuclease hypersensitive region of the proto-oncogene c-*MYC* (Simonsson, Pecinka et al. 1998); (Siddiqui-Jain, Grand et al. 2002), other proto-oncogenes promoters (c-kit, bcl-2, VEGF, H-ras, N-ras, K-ras) and many other gene promoters, such as the chicken β-globin gene and human ubiqui-

tin-ligase RFP2(Todd, Johnston et al. 2005); (Huppert and Balasubramanian 2005).

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 (Huppert and Balasubramanian 2005); (Todd, Johnston et al. 2005) (Chambers, Marsico et al. 2015). 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 (Le, Di Antonio et al. 2015), achieving complete specificity between G-quadruplexes motifs remains a challenge. Thus, there is case to progress from targeting single distinct quadruplexes towards the identification and characterization of the interfaces formed adjacent to these individual G-quadruplexes, with the goal of developing ligands that could interact selectively 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 (Campbell, Parkinson et al. 2008), and a duplex-quadruplex junction, analogous to spinach RNA(Huang, Suslov et al. 2014) (Warner, Chen et al. 2014), and the stabilization of the quadruplex duplex interface with the binding of a fluorophore. Previous drug discovery efforts have in effect targeted 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 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 (Guedin, Gros et al. 2010) and preferentially folded into a parallel stranded G-quadruplex (Rachwal, Brown et al. 2007). 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 four-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, co-crystallisations in the presence of high-affinity quadruplex binding ligands and soaking experiments on preformed crystals were also conducted. Using X-ray crystallography and molecular modelling we report four new structures within two crystal forms. The binding of ligands at the interface between stacked duplex DNA stabilizes crystal packing interactions. Modelling 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.
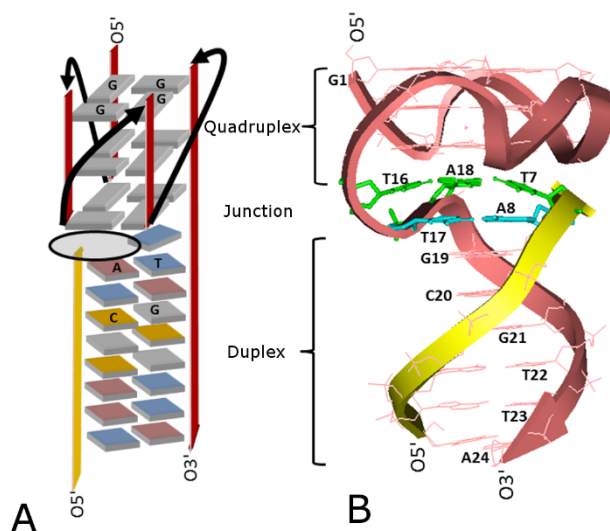

Figure 1. Schematic diagrams of the quadruplex-duplex construct as designed (A) and as observed in crystal structures (B). The quadruplex and duplex motifs, the gap in panel A and the junction, a potential site of selective ligand-binding, in panel B are highlighted, along with strand A (pink) and strand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

EXPERIMENTAL

Sample preparation, crystallization and soaking

All sequences used in these studies were purchased from Eurofins, HPLC purified: a) a 24-mer, d(GGGTGGGTGGGTGGGTTAGCGTTA) where the first three TTA sequences were substituted with single thymines, b) a 25-mer, d(GGGTGGGTGGGTTGGGTTAGCGTTA) 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(GGGTTAGGGTTAGGGTTAGGGTTAGCGTTA), d) a 8-mer, d(TAACGCTA) complementary to the 3' end of the longer sequences (Table 1). The oligonucleotides were ini-

tially 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 (Gunaratnam, Green et al. 2009), BSU6037 and BG32 (Harrison, Gowan et al. 1999) (Guyen, Schultes et al. 2004) (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: 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 sulphate, 50 mM Tris/HCl pH 8.8 and 2 mM CuCl2. 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 (Chemistry. 2012 Nov 12;18(46):14752-9). These interactions would be paramount in stabilising 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 (J Am Chem Soc. 2008 May 28;130(21):6722-4; J Mol Biol. 2008 Sep 19;381(5):1145-

56). Based on these observations, we envisaged that a di-substituted ligand would bind to the junction and stabilise the 3' interface, by stacking on one half of the quartet (J Am Chem Soc. 2008 May 28;130(21):6722-4) 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 complexes 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 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 P4₂12 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 (McCoy, Grosse-Kunstleve et al. 2007) 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 (Emsley, Lohkamp et al. 2010) and REFMAC5 (Murshudov, Skubak et al. 2011) 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 Modelling

The quadruplex-duplex TTLOOP crystal structure was used as a starting point for modelling 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 Å³. 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 (Abagyan, Totrov et al. 1994). 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 (Parkinson, Lee et al. 2002) (Do, Lim et al. 2011) 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 anti-parallel 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 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 (Heddi and Phan 2011) (Petraccone, Malafronte et al. 2012). 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 towards the 3' end of the duplex region.
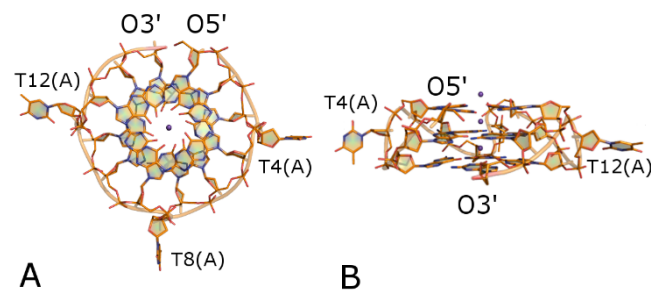

Figure 2. Top view (A) and lateral view (B) of the quadruplex region of TLOOP. Potassium ions are also shown as purple spheres. Loop residues are explicitly marked.

The formation of the junction between quadruplex and duplex regions is consistent with preliminary computational studies (Ramaswamy 2011). 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/Rfree 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 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). 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).
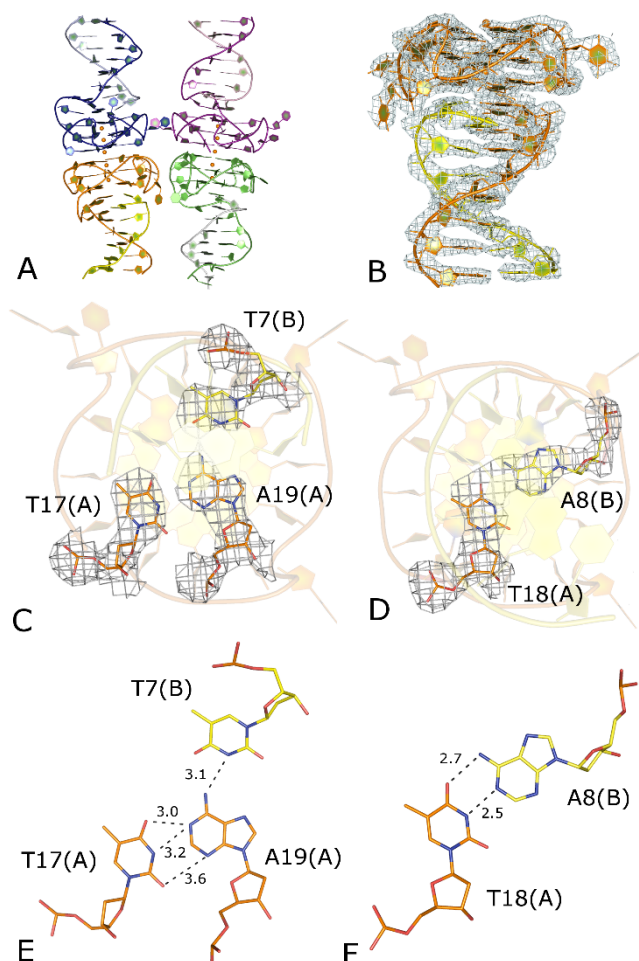

Figure 3. TTLOOP structure. A) Cartoon representation of the four TTLOOP molecules in the asymmetric unit. Different chains are marked in different colors. Potassium ions are shown as orange spheres. B) 2Fo-Fc electron density maps of the quadruplex and duplex regions of TTLOOP contoured at 2.0 and 1.5 σ level, respectively. C) Top view of the interface triad. 2Fo-Fc electron density map contoured at 1.9 σ level is also shown. D) Top view of the first duplex pair. 2Fo-Fc electron density map contoured at 1.9 σ 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.

Addition of ligands

In an attempt to identify ligands that could associate with the quadruplex-duplex interface other than peptides (Vasilyev, Polonskaia et al. 2015), 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. Based on our molecular modelling studies, several of these ligands were selected for co-crystallization experiments: BSU6037, a disubstituted acridine derivative analogous to ligand BSU6039 (PDB-ID 1L1H) (Haider, Parkinson et al. 2003); BRACO-19 (PDB-ID 3CE5) (Campbell, Parkinson et al. 2008), a trisubstituted acridine derivative; TMPyP4, a tetra-(N-methyl-4-pyridyl) porphyrin (PDB-ID 2HRI) (Parkinson, Ghosh et al. 2007); and FC4ND10(Cuenca 2008), a two dimethylamine, two hydroxyl tetra-substituted naphthalene diimide derivative, (PDB-ID 3T5E) (Collie, Promontorio et al. 2012). These ligands have also shown telomerase inhibitory activity in enzymatic assays (Campbell, Parkinson et al. 2008), (Gunaratnam, Greciano et al. 2007), (Martins, Gunaratnam et al. 2007). During co-crystallization 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 pre-formed 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 contacts the 3'-terminal ends of two symmetry-related TLOOP quadruplex-duplexes on the opposite side (Figure 4A).

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-quartets (Figure 4D). This architecture seems to occlude the binding of ligands at the interface between the two structural motifs.
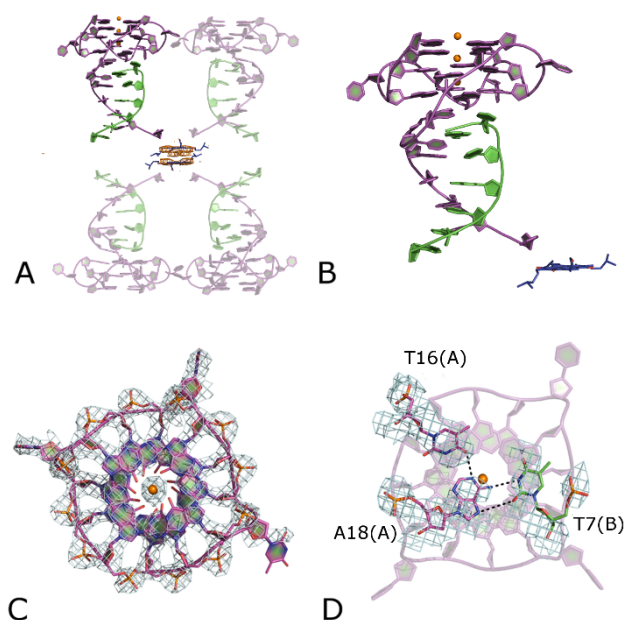

Figure 4. TLOOP- FC4ND10 structure with the quadruplex-duplex DNA represented as cartoon (longer strand A in magenta, complementary strand B in green), the ligand represented as stick in blue, the potassium ions as orange spheres. A) Packing interactions: ligand molecule is shown placed on the 4-fold 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 σ 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 σ level D) 2Fo-Fc electron density map of residues belonging to the interface triad contoured at 1.0 σ level.

Ligand docking into the quadruplex-duplex interface

Since no structural data is available on ligand binding at DNA quadruplex-duplex interfaces, we used molecular modelling to investigate the likely binding features of the hybrid junction. The loops of human telomeric quadruplexes are very flexible (Collie, Campbell et al. 2015), particularly the ones that connect segments together (Islam, Sgobba et al. 2013) (Haider, Parkinson et al. 2008) (Haider and Neidle 2009). 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 (Huang, Suslov et al. 2014). We constructed a pseudo ligand-binding site by rotating T17(A) towards 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 π-stacking interactions with G11, G16 and T18 (Figure 5). The central protonated nitrogen of the acridine 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 (Haider, Parkinson et al. 2003). 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.
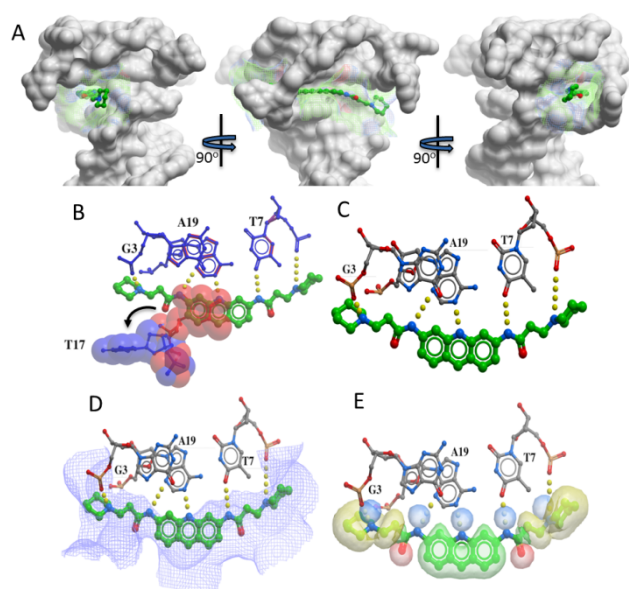

Figure 5. Structural features around the pseudo ligand-binding site. The ligand is colored in green; hydrogen bonds are represented as yellow dotted lines. A) The spatial position of the binding site at the quadruplex-duplex interface. B) The T17 base is flipped out (blue) of the TAT triad (red), generating 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.

The G-quadruplex-duplex interface

A review of the few structures currently available (Lim, Khong et al. 2014) (Phan, Kuryavyi et al. 2011) (Lim and Phan 2013) in the PDB reveals the expected dominance of base-stacking at the interface, with all structures having the expected helical geometries associated with quadru-

plex 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 (Lim, Khong et al. 2014) 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 (Lim and Phan 2013) 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 non-related RNA G-quadruplex structure (PDB-ID 2LA5) (Phan, Kuryavyi et al. 2011) shows the RNA bound to a peptide, which has a similar transition 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). 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 S3).

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 S4).

CONCLUSIONS

These 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 the X-ray crystallography and molecular dynamic simulations (Ramaswamy 2011). 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 modelling 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, similarly to that already 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 tri- substituted acridine ligands, but not of the tetrasubstituted naphthaline diimide ligands, thus having a significant 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 thus may represent more biologically relevant models for ligand lead discovery and optimization than simple isolated quadruplexes.

ASSOCIATED CONTENT

Supporting Information. 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. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*gary.parkinson@ucl.ac.uk.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

REFERENCES

(1) Gellert M, L. M., Davies DR *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013.
(2) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nature chemistry* **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 research* **2014**, *42*, 860.
(4) Wang, Y.; Patel, D. J. *Structure (London, England : 1993)* **1993**, *1*, 263.
(5) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic acids research* **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.
(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. *Journal of molecular biology* **2008**, *381*, 1145.
(10) Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic acids research* **1998**, *26*, 1167.
(11) Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 11593.
(12) Todd, A. K.; Johnston, M.; Neidle, S. *Nucleic acids research* **2005**, *33*, 2901.
(13) Huppert, J. L.; Balasubramanian, S. *Nucleic acids research* **2005**, *33*, 2908.
(14) Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S. *Nature biotechnology* **2015**, *33*, 877.
(15) Le, D. D.; Di Antonio, M.; Chan, L. K.; Balasubramanian, S. *Chemical communications (Cambridge, England)* **2015**, *51*, 8048.
(16) Guedin, A.; Gros, J.; Alberti, P.; Mergny, J. L. *Nucleic acids research* **2010**, *38*, 7858.
(17) Rachwal, P. A.; Brown, T.; Fox, K. R. *FEBS Lett* **2007**, *581*, 1657.
(18) Gunaratnam, M.; Green, C.; Moreira, J. B.; Moorhouse, A. D.; Kelland, L. R.; Moses, J. E.; Neidle, S. *Biochem Pharmacol* **2009**, *78*, 115.
(19) Harrison, R. J.; Gowan, S. M.; Kelland, L. R.; Neidle, S. *Bioorganic & medicinal chemistry letters* **1999**, *9*, 2463.
(20) Guyen, B.; Schultes, C. M.; Hazel, P.; Mann, J.; Neidle, S. *Organic & biomolecular chemistry* **2004**, *2*, 981.
(21) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *Journal of applied crystallography* **2007**, *40*, 658.
(22) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta crystallographica. Section D, Biological crystallography* **2010**, *66*, 486.
(23) 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 crystallographica. Section D, Biological crystallography* **2011**, *67*, 355.
(24) Abagyan, R.; Totrov, M.; Kuznetsov, D. *Journal of Computational Chemistry* **1994**, *15*, 488.
(25) Do, N. Q.; Lim, K. W.; Teo, M. H.; Heddi, B.; Phan, A. T. *Nucleic acids research* **2011**, *39*, 9448.
(26) Heddi, B.; Phan, A. T. *J Am Chem Soc* **2011**, *133*, 9824.
(27) Petraccone, L.; Malafronte, A.; Amato, J.; Giancola, C. *J Phys Chem B* **2012**, *116*, 2294.
(28) Ramaswamy, S. *Unpublished master's thesis, The School of Pharmacy, London.* **2011**.

(29) Haider, S. M.; Parkinson, G. N.; Neidle, S. *Journal of molecular biology* **2003**, *326*, 117.

(30) Parkinson, G. N.; Ghosh, R.; Neidle, S. *Biochemistry* **2007**, *46*, 2390.

(31) Cuenca, F. G., O. Gunaratnam, M. Haider, S. Munnur, D. Nanjunda, R., Wilson W. D., Neidle, S. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1668 1673.

(32) Collie, G. W.; Promontorio, R.; Hampel, S. M.; Micco, M.; Neidle, S.; Parkinson, G. N. *J Am Chem Soc* **2012**, *134*, 2723.

(33) Gunaratnam, M.; Greciano, O.; Martins, C.; Reszka, A. P.; Schultes, C. M.; Morjani, H.; Riou, J. F.; Neidle, S. *Biochem Pharmacol* **2007**, *74*, 679.

(34) Martins, C.; Gunaratnam, M.; Stuart, J.; Makwana, V.; Greciano, O.; Reszka, A. P.; Kelland, L. R.; Neidle, S. *Bioorganic & medicinal chemistry letters* **2007**, *17*, 2293.

(35) Collie, G. W.; Campbell, N. H.; Neidle, S. *Nucleic acids research* **2015**, *43*, 4785.

(36) Islam, B.; Sgobba, M.; Laughton, C.; Orozco, M.; Sponer, J.; Neidle, S.; Haider, S. *Nucleic acids research* **2013**, *41*, 2723.

(37) Haider, S.; Parkinson, G. N.; Neidle, S. *Biophysical journal* **2008**, *95*, 296.

(38) Haider, S. M.; Neidle, S. *Biochem Soc Trans* **2009**, *37*, 583.

(39) Lim, K. W.; Khong, Z. J.; Phan, A. T. *Biochemistry* **2014**, *53*, 247.

(40) 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.

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

Table 2. Data-collection statistics

	TELO	TLOOP	TTLOOP	TLOOP-BSU6037	TLOOP-FC4ND10
Space group	R3	$P4_2{}_12$	I222	$P4_2{}_12$	$P4_2{}_12$
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
α (°)	90.00	90.00	90.00	90.00	90.00
β (°)	90.00	90.00	90.00	90.00	90.00
γ (°)	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)
$R_{merge}{}^{\dagger}$ (%)	12.8 (44.5)	13.4(93)	5.3 (67.9)	10.5 (48.2)	17.5 (44.1)
V_M (Å³ 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/Rfree	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 (°)	1.70	1.55	1.58	1.62
Average B-factors (Å2)				
Oligonucleotide	66.56	93.46	59.64	94.54
Ions	23.96	32.38	20.57	5.93
Ligand	-	-	-	160.75

Insert Table of Contents artwork

Abagyan, R., M. Totrov and D. Kuznetsov (1994). "ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation." Journal of Computational Chemistry **15**(5): 488-506.

Ambrus, A., D. Chen, J. Dai, T. Bialis, R. A. Jones and D. Yang (2006). "Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution." Nucleic acids research **34**(9): 2723-2735.

Biffi, G., D. Tannahill, J. McCafferty and S. Balasubramanian (2013). "Quantitative visualization of DNA G-quadruplex structures in human cells." Nat Chem **5**(3): 182-186.

Campbell, N. H., G. N. Parkinson, A. P. Reszka and S. Neidle (2008). "Structural basis of DNA quadruplex recognition by an acridine drug." J Am Chem Soc **130**(21): 6722-6724.

Chambers, V. S., G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith and S. Balasubramanian (2015). "High-throughput sequencing of DNA G-quadruplex structures in the human genome." Nat Biotechnol **33**(8): 877-881.

Collie, G. W., N. H. Campbell and S. Neidle (2015). "Loop flexibility in human telomeric quadruplex small-molecule complexes." Nucleic Acids Res **43**(10): 4785-4799.

Collie, G. W., R. Promontorio, S. M. Hampel, M. Micco, S. Neidle and G. N. Parkinson (2012). "Structural basis for telomeric G-quadruplex targeting by naphthalene diimide ligands." J Am Chem Soc **134**(5): 2723-2731.

Cuenca, F. G., O. Gunaratnam, M. Haider, S. Munnur, D. Nanjunda, R., Wilson W. D., Neidle, S. (2008). "Tri- and tetra-substituted naphthalene diimides as potent G-quadruplex ligands." Bioorg. Med. Chem. Lett. **18**: 1668 1673.

Do, N. Q., K. W. Lim, M. H. Teo, B. Heddi and A. T. Phan (2011). "Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity." Nucleic acids research **39**(21): 9448-9457.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

Gellert M, L. M., Davies DR (1962). "Helix formation by guanylic acid." Proc. Natl. Acad. Sci. USA **48**: 2013–2018.

Guedin, A., J. Gros, P. Alberti and J. L. Mergny (2010). "How long is too long? Effects of loop size on G-quadruplex stability." Nucleic acids research **38**(21): 7858-7868.

Gunaratnam, M., O. Greciano, C. Martins, A. P. Reszka, C. M. Schultes, H. Morjani, J. F. Riou and S. Neidle (2007). "Mechanism of acridine-based telomerase inhibition and telomere shortening." Biochemical pharmacology **74**(5): 679-689.

Gunaratnam, M., C. Green, J. B. Moreira, A. D. Moorhouse, L. R. Kelland, J. E. Moses and S. Neidle (2009). "G-quadruplex compounds and cis-platin act synergistically to inhibit cancer cell growth in vitro and in vivo." Biochemical pharmacology **78**(2): 115-122.

Guyen, B., C. M. Schultes, P. Hazel, J. Mann and S. Neidle (2004). "Synthesis and evaluation of analogues of 10H-indolo[3,2-b]quinoline as G-quadruplex stabilising ligands and potential inhibitors of the enzyme telomerase." Org Biomol Chem **2**(7): 981-988.

Haider, S., G. N. Parkinson and S. Neidle (2008). "Molecular dynamics and principal components analysis of human telomeric quadruplex multimers." Biophys J **95**(1): 296-311.

Haider, S. M. and S. Neidle (2009). "A molecular model for drug binding to tandem repeats of telomeric G-quadruplexes." Biochemical Society transactions **37**(Pt 3): 583-588.

Haider, S. M., G. N. Parkinson and S. Neidle (2003). "Structure of a G-quadruplex-ligand complex." J Mol Biol **326**(1): 117-125.

Harrison, R. J., S. M. Gowan, L. R. Kelland and S. Neidle (1999). "Human telomerase inhibition by substituted acridine derivatives." Bioorg Med Chem Lett **9**(17): 2463-2468.

Heddi, B. and A. T. Phan (2011). "Structure of human telomeric DNA in crowded solution." Journal of the American Chemical Society **133**(25): 9824-9833.

Henderson, A., Y. Wu, Y. C. Huang, E. A. Chavez, J. Platt, F. B. Johnson, R. M. Brosh, Jr., D. Sen and P. M. Lansdorp (2014). "Detection of G-quadruplex DNA in mammalian cells." Nucleic Acids Res **42**(2): 860-869.

Huang, H., N. B. Suslov, N. S. Li, S. A. Shelke, M. E. Evans, Y. Koldobskaya, P. A. Rice and J. A. Piccirilli (2014). "A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore." Nat Chem Biol **10**(8): 686-691.

Huppert, J. L. and S. Balasubramanian (2005). "Prevalence of quadruplexes in the human genome." Nucleic acids research **33**(9): 2908-2916.

Islam, B., M. Sgobba, C. Laughton, M. Orozco, J. Sponer, S. Neidle and S. Haider (2013). "Conformational dynamics of the human propeller telomeric DNA quadruplex on a microsecond time scale." Nucleic Acids Res **41**: 2723-2735.

Le, D. D., M. Di Antonio, L. K. Chan and S. Balasubramanian (2015). "G-quadruplex ligands exhibit differential G-tetrad selectivity." Chem Commun (Camb) **51**(38): 8048-8050.

Lim, K. W., Z. J. Khong and A. T. Phan (2014). "Thermal stability of DNA quadruplex-duplex hybrids." Biochemistry **53**(1): 247-257.

Lim, K. W. and A. T. Phan (2013). "Structural basis of DNA quadruplex-duplex junction formation." Angew Chem Int Ed Engl **52**(33): 8566-8569.

Luu, K. N., A. T. Phan, V. Kuryavyi, L. Lacroix and D. J. Patel (2006). "Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold." Journal of the American Chemical Society **128**(30): 9963-9970.

Martins, C., M. Gunaratnam, J. Stuart, V. Makwana, O. Greciano, A. P. Reszka, L. R. Kelland and S. Neidle (2007). "Structure-based design of benzylamino-acridine compounds as G-quadruplex DNA telomere targeting agents." Bioorganic & medicinal chemistry letters **17**(8): 2293-2298.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.

Murshudov, G. N., P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin (2011). "REFMAC5 for the refinement of macromolecular crystal structures." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 355-367.

Parkinson, G. N., F. Cuenca and S. Neidle (2008). "Topology conservation and loop flexibility in quadruplex-drug recognition: crystal structures of inter- and intramolecular telomeric DNA quadruplex-drug complexes." J Mol Biol **381**(5): 1145-1156.

Parkinson, G. N., R. Ghosh and S. Neidle (2007). "Structural basis for binding of porphyrin to human telomeres." Biochemistry **46**(9): 2390-2397.

Parkinson, G. N., M. P. Lee and S. Neidle (2002). "Crystal structure of parallel quadruplexes from human telomeric DNA." Nature **417**(6891): 876-880.

Petraccone, L., A. Malafronte, J. Amato and C. Giancola (2012). "G-quadruplexes from human telomeric DNA: how many conformations in PEG containing solutions?" The journal of physical chemistry. B **116**(7): 2294-2305.

Phan, A. T., V. Kuryavyi, J. C. Darnell, A. Serganov, A. Majumdar, S. Ilin, T. Raslin, A. Polonskaia, C. Chen, D. Clain, R. B. Darnell and D. J. Patel (2011). "Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction." Nature structural & molecular biology **18**(7): 796-804.

Rachwal, P. A., T. Brown and K. R. Fox (2007). "Sequence effects of single base loops in intramolecular quadruplex DNA." FEBS letters **581**(8): 1657-1660.

Ramaswamy, S. (2011). "Structural Investigation of a Duplex-Quadruplex Junction." Unpublished master's thesis, The School of Pharmacy, London.

Siddiqui-Jain, A., C. L. Grand, D. J. Bearss and L. H. Hurley (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 U S A **99**(18): 11593-11598.

Simonsson, T., P. Pecinka and M. Kubista (1998). "DNA tetraplex formation in the control region of c-myc." Nucleic Acids Res **26**(5): 1167-1172.

Todd, A. K., M. Johnston and S. Neidle (2005). "Highly prevalent putative quadruplex sequence motifs in human DNA." Nucleic acids research **33**(9): 2901-2907.

Vasilyev, N., A. Polonskaia, J. C. Darnell, R. B. Darnell, D. J. Patel and A. Serganov (2015). "Crystal structure reveals specific recognition of a G-quadruplex RNA by a beta-turn in the RGG motif of FMRP." <u>Proc Natl Acad Sci U S A</u> **112**(39): E5391-5400.

Wang, Y. and D. J. Patel (1993). "Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex." <u>Structure</u> **1**(4): 263-282.

Warner, K. D., M. C. Chen, W. Song, R. L. Strack, A. Thorn, S. R. Jaffrey and A. R. Ferre-D'Amare (2014). "Structural basis for activity of highly efficient RNA mimics of green fluorescent protein." <u>Nat Struct Mol Biol</u> **21**(8): 658-663.