

## SURVEY AND SUMMARY

# Loop flexibility in human telomeric quadruplex small-molecule complexes

Gavin W. Collie, Nancy H. Campbell and Stephen Neidle\*

UCL School of Pharmacy, University College London, 29-39 Brunswick Square, London WC1N 1AX, UK

Received March 20, 2015; Revised April 18, 2015; Accepted April 20, 2015

### ABSTRACT

**Quadruplex nucleic acids can be formed at the ends of eukaryotic chromosomes. Their formation and stabilisation by appropriate small molecules can be used as a means of inhibiting the telomere maintenance functions of telomerase in human cancer cells. The crystal structures have been determined for a number of complexes between these small molecules and human telomeric DNA and RNA quadruplexes. The detailed structural characteristics of these complexes have been surveyed here and the variations in conformation for the TTA and UUA loops have been explored. Loop conformations have been classified in terms of a number of discrete types and their distribution among the crystal structures. Sugar conformation and backbone angles have also been examined and trends highlighted. One particular loop class has been found to be most prevalent. Implications for in particular, rational drug design, are discussed.**

### BACKGROUND—QUADRUPLEX FORM AND FUNCTION

Sequences containing short repetitive G-tracts occur in a number of genomic contexts. They are often able to form higher-order structures, termed quadruplexes (1–3), especially under the influence of appropriate small molecules, proteins or negative supercoiling (4). Historically, the first such sequence category to be so characterised has been in eukaryotic telomeres, which comprise tandem repeats of a G-tract-containing DNA sequence together with associated telomeric proteins. The hexanucleotide repeat in human telomeric DNA is 5'-TTAGGG (5). Almost all of a telomeric DNA sequence (length varies from ~5 to ~10 kb) is in duplex form, with the exception of the ~150–200 nucleotides at the 3'-terminus which are single-stranded (6).

This single-stranded DNA (the 'overhang') can be folded to form either bimolecular (two-repeat) or unimolecular (four-repeat) quadruplexes, once the associated single-stranded binding proteins (principally several copies of hPOT1), have been competed away (7,8). The concept that quadruplex-binding small molecules can stabilise telomeric quadruplexes (9), has been widely used to discover small molecules with potential anti-cancer activity via inhibition of the action of the telomerase enzyme complex (10–14). Telomerase is a key hallmark of cancer and is highly over-expressed in many cancer cell types whereas it is only expressed at low levels in normal somatic cells, suggesting that it is a plausible anti-cancer target (15,16). The reverse transcriptase action of telomerase catalyses the synthesis of 5'-TTAGGG repeats onto the overhang, to counteract the loss of repeats during replication. This overhang is required to be in a single-stranded form to hybridise with the RNA domain of the telomerase complex and for the catalytic cycle to proceed. Small-molecule induction of the overhang into a four-stranded quadruplex structure, which can be augmented by a quadruplex-binding small molecule (9), results in the overhang becoming inaccessible to the telomerase RNA template, thus effectively inhibiting telomerase catalytic function (17).

### SMALL MOLECULE BINDING TO QUADRUPLEX DNA AND RNA

The demonstration of this concept, that telomerase inhibition can be achieved via quadruplex stabilisation with anthraquinone derivatives (9,18) and subsequently with many other small molecules (10–13) has led to the widespread development of this approach to selectively targeting human cancers. Information on the large and structurally diverse superfamily of quadruplex-binding small molecules is available from the G4LDB database [www.g4ldb.org](http://www.g4ldb.org) (19).

Several compounds including the acridine derivatives BRACO-19 (20), RHPS4 (21), the cyclic natural product telomestatin (22) and its analogues, and some tetra-

\*To whom correspondence should be addressed. Tel: +44 207 753 5969; Fax: +44 207 753 5970; Email: [s.neidle@ucl.ac.uk](mailto:s.neidle@ucl.ac.uk)

substituted naphthalene diimide (ND) derivatives (23), have been shown to selectively affect telomere maintenance in cancer cells. Mechanistic studies have demonstrated that the cellular responses to these agents, which occur considerably more rapidly than would be expected on the basis of a classic telomerase inhibition and telomere attrition model, involve the activation of a DNA damage response, presumably by sensing the presence of a quadruplex nucleic acid signal which is not associated with and protected by, telomeric proteins (20,21,24–30). This response is an important factor in the rapidity and selectivity of the growth-inhibitory action of these quadruplex-binding ligands in tumour cells and tumour xenografts (19,20,24).

The frequency of occurrence of quadruplex-forming sequences other than at telomeres in the human genome has been evaluated by informatics approaches (31–33). These studies have shown that such sequences are especially prevalent in promoter sequences (34) and in 5'-UTRs (35–37). Their over-representation in a number of oncogene promoter sequences (33) has led to the hypothesis that promoter quadruplexes can be targets for therapeutic intervention using small molecules which stabilise a particular quadruplex structure within a promoter sequence (38,39). This in principle results in inhibition of the transcription of that particular gene. The concept has been evaluated with small molecules targeting in particular the *c-MYC* and *c-KIT* promoter quadruplexes (see for example 10–13,39), although the challenge remains of devising small molecules capable of selectively targeting a particular quadruplex in the absence of off-target effects on other quadruplexes. Importantly, the concept of quadruplexes as targets in human cancer has been validated by their direct visualisation in human cells and tissues (40–42).

Extensive structural data on native human telomeric quadruplexes (HTQs) are now available from crystallographic and nuclear magnetic resonance (NMR) studies (43–53), whereas there is currently structural information available on only a small number of non-telomeric quadruplexes, of which the *c-KIT*, *c-MYC* and *BCL-2* promoter quadruplexes are the best-studied (54–64). Detailed NMR structures have been determined for three small-molecule complexes with *c-MYC* quadruplexes: bound to the porphyrin ligand TMPyP4 (PDB ID 2A5R) (65), bound to the bis-quinolinium compound Phen-DC(3) (PDB ID 2MGN) (66) and with a drug-like quindoline compound (PDB ID 2L7V) (67).

The available structural data on human telomeric quadruplex (HTQ) complexes with small molecules are dominated by results from X-ray crystallographic studies, with 16 structures deposited to date in the PDB (68–76). Of the few NMR studies on human telomeric quadruplex-small molecule complexes that have been reported, only one is of an intramolecular complex, bound to a derivative of the drug telomestatin (PDB ID 2MB3) (77). The present study focusses on the crystal structures and in particular, analyses aspects of their conformational variability in order to establish unifying trends and features. There are currently no crystal structures available for small-molecule complexes with promoter or 5'-UTR quadruplexes and only a very few NMR structures (involving promoter quadruplexes), but some of the underlying principles outlined below will also

apply to these complexes. One of the small-molecule complexes in the database is with a bimolecular human telomeric RNA (TERRA) quadruplex (71), and this is complemented by the availability of a crystal structure of the native RNA quadruplex (78).

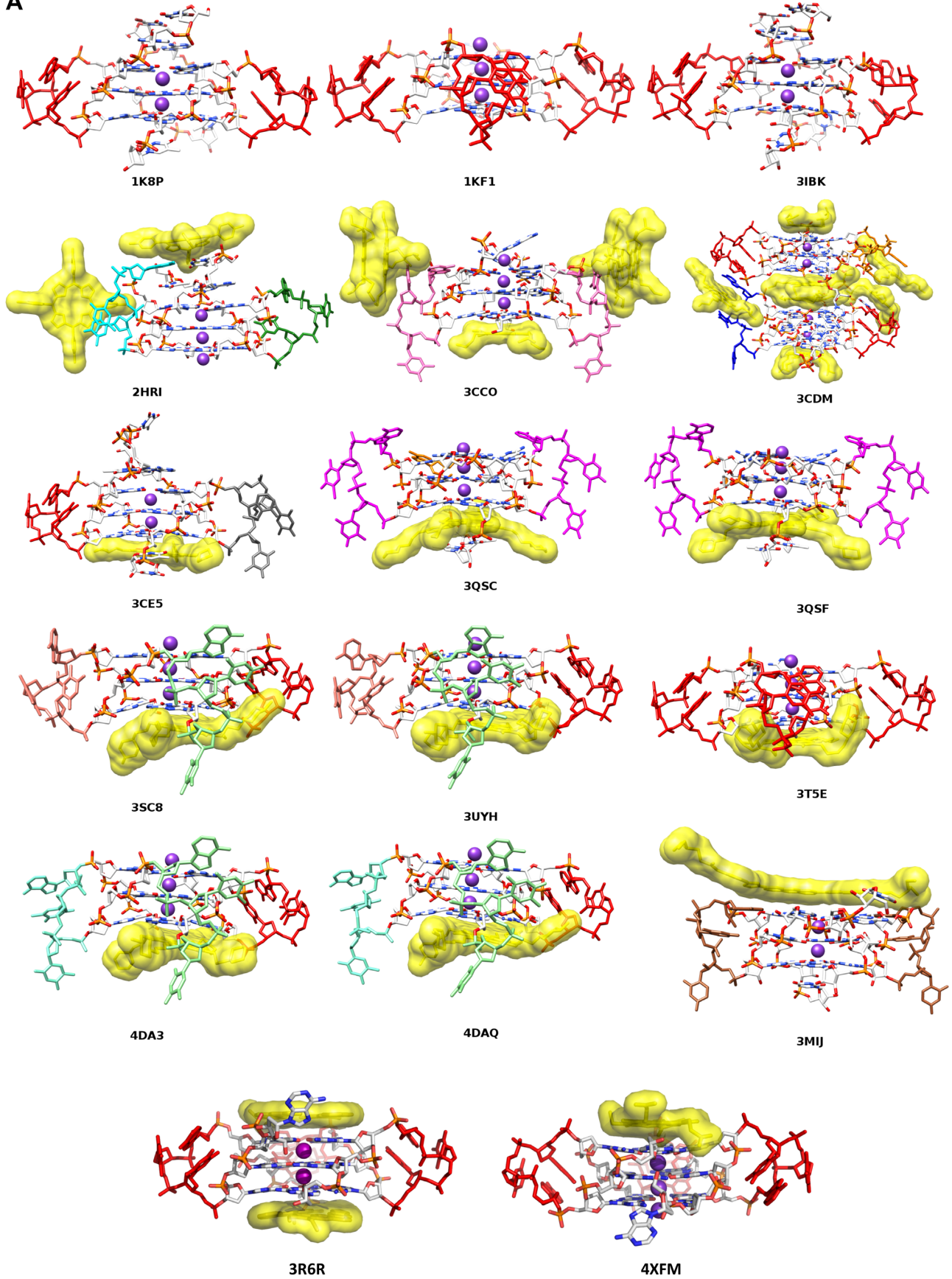
The overwhelming majority of quadruplex-binding small molecules reported to date possess planar groupings and their mode of binding predominantly involves 3' or 5' end-stacking onto terminal G-quartets (Figure 1). Intercalative binding has not been observed experimentally for these ligands and is widely considered not to be consistent with biophysical data on quadruplex-small molecule complexes (11–13). A small number of quadruplex-binding ligands, mostly based on a polyamide motif, have been proposed to bind in quadruplex grooves rather than on G-quartet surfaces. There is no detailed experimental structural information on their interaction with human telomeric bimolecular or unimolecular quadruplexes and so groove-binding structures will not be considered any further here.

As a consequence of the rise of the quadruplex targeting concept, a large number of diverse quadruplex-binding small molecules have been devised and evaluated. Most of these have been discovered by biophysical screens of compound libraries (11–13). Rather fewer studies have employed rational design and optimisation of ligands using structural data from X-ray crystallography and NMR studies. *In silico* docking methods, mostly using native quadruplex structures as starting-points, have been used in a few instances to screen large compound libraries, (79–81). Experimental structures are crucial starting-points for *in silico* studies so it is hoped that the analyses and correlations presented here will help provide sound starting-points for such future studies.

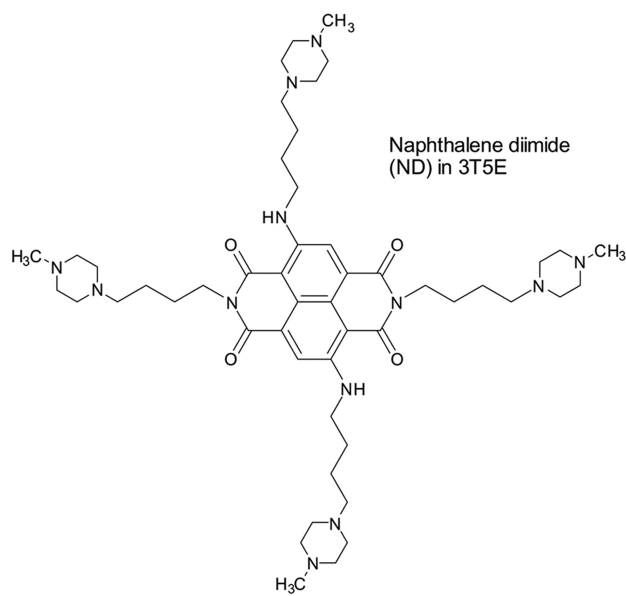
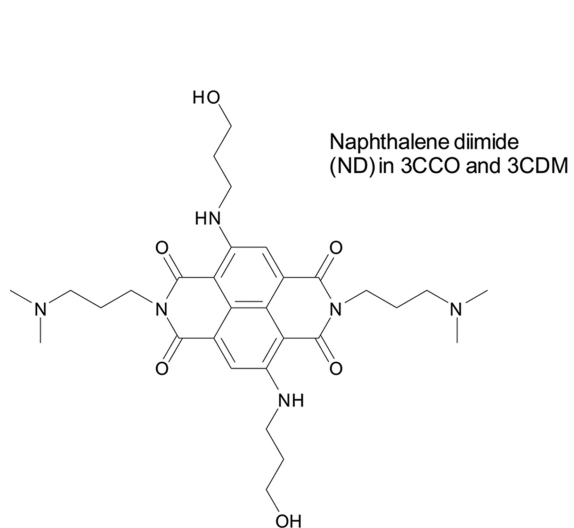
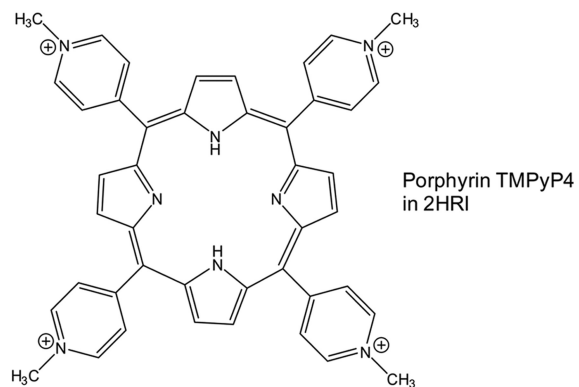
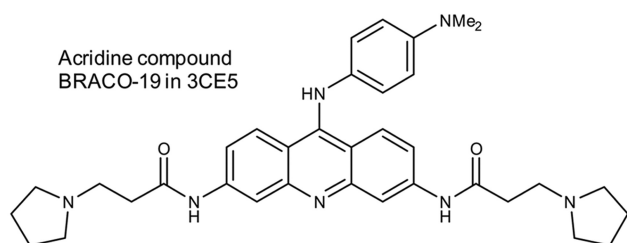
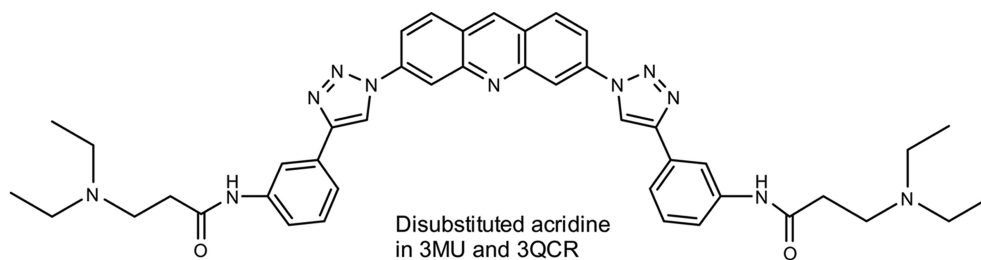
## THE FLEXIBILITY OF HUMAN TELOMERIC QUADRUPLEXES

It is now well-established that intramolecular quadruplexes comprising four repeats of the human telomeric sequence 5'-TTAGGG, can form a variety of topological isomers in solution, differing in the orientation of the backbone and the nature of the loops formed by the 'spacer' TTA triplet sequences (which can adopt lateral, diagonal or parallel arrangements dependent on the orientation of the attached strand). A number of these quadruplex topological isomers have been identified by NMR methods (43–50); the parallel form with all loops of the propeller type (also termed strand-reversal) is the sole form found in the X-ray crystallographic studies (52). The observation of a particular quadruplex topological isomer in solution is dependent on a number of factors, primarily quadruplex concentration, the presence or otherwise of a molecular crowding environment, the nature of the counter-ion and the flanking sequences at the 5' and 3' ends (see for example 51,82–85). There continues to be controversy as to which topological isomer is the most biologically relevant, although it is plausible that both parallel and anti-parallel ones can co-exist in the high local concentration of cellular conditions. Small molecule binding can induce the stabilisation of a particular form (for example the induction of the parallel topology by N-methylmesoporphyrin IX (86)); although there is to date very incomplete understanding of the molecular basis for

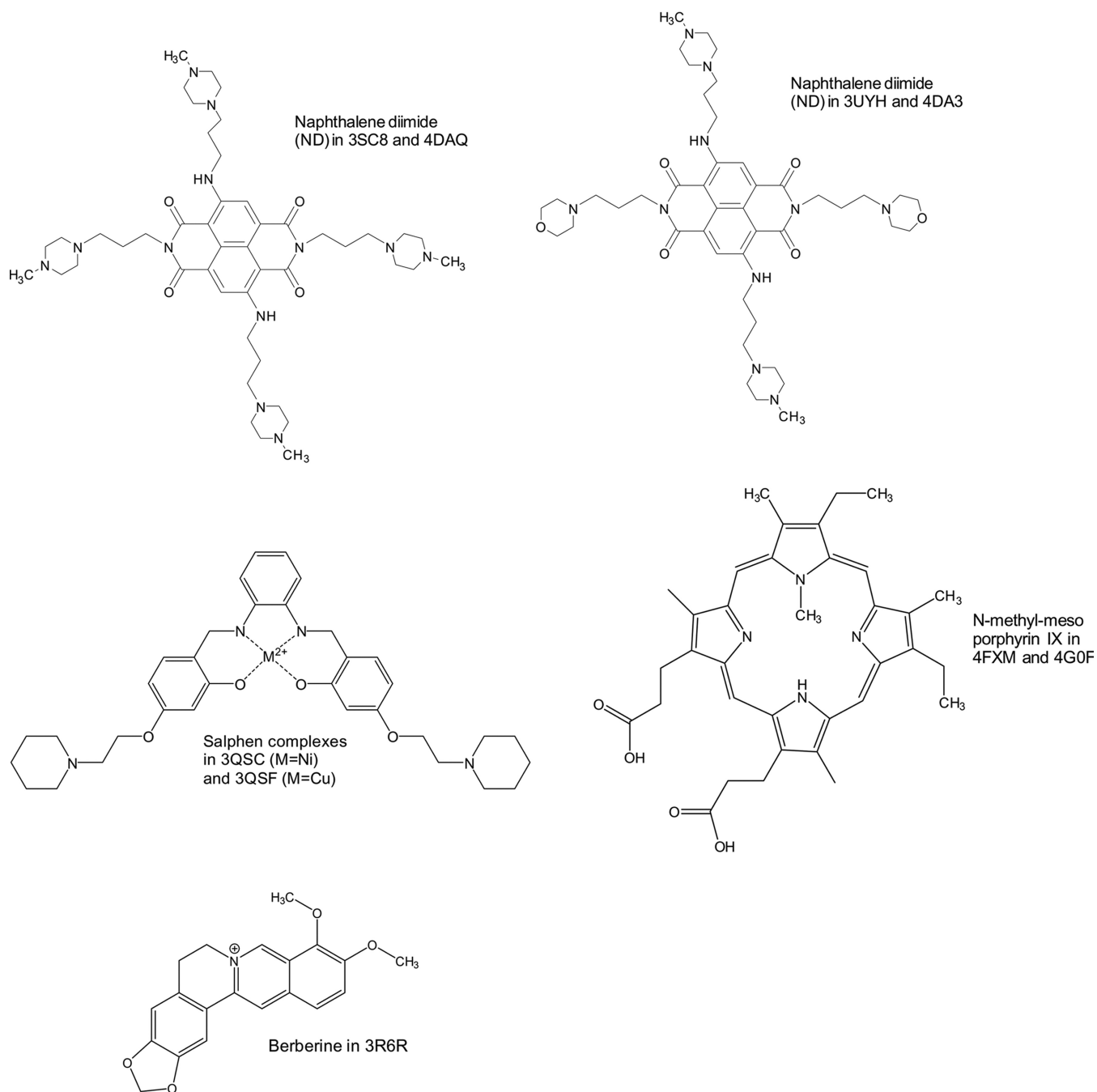
A



**B**







**Figure 1.** (A) View of the human telomeric quadruplex native and ligand complex crystal structures, with nucleic acids drawn in stick form and ligands in surface representation. Metal ions (potassiums) are shown as purple spheres. Individual loops are highlighted with sticks in single colours. The colour coding corresponds to the more detailed view of each loop type shown in Figure 2. (B) Structures of individual ligands, shown with the PDB IDs for the structures in which they occur.

the preference for a particular topology; accessible surface area of a terminal G-quartet has been suggested as being an important factor (87).

This diversity in native HTQ topologies has not been reflected in the folds observed in the crystal structures of small-molecule complexes with HTQs. All have been found to have a parallel topology, regardless of whether they are bimolecular or unimolecular complexes (Table 1). They crystallise in ten different crystallographic space groups,

strongly suggesting that crystal packing forces are not the determinant of the single observed (parallel) topology. The present survey has examined in particular the principal nucleic acid conformational variable regions in these structures, the TTA loops. The structural cores of all quadruplexes comprise two or more stacked G-quartets. In the case of HTQ crystal structures the core comprises three stacked G-quartets, which has been observed to be exceptionally stable. Structural alignment of the three stacked G-quartets

**Table 1.** Native DNA and RNA human telomeric quadruplexes and small-molecule complex crystal structures showing PDB codes, basic crystallographic data and TTA/UUA loop types

PDB ID	Sequence	Description of structure	Resolution (Å)	Space group	R-free	R-factor	No. of crystallographically unique loops	Loop type	Ref.
1K8P	d[U <sub>Br</sub> AG <sub>3</sub> U <sub>Br</sub> TAG <sub>3</sub> T]	Native DNA	2.40	P3 <sub>1</sub> 2 1	0.280	0.193	2 <sup>a</sup>	Type-1	(52)
1KF1	d[AG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	Native DNA	2.10	P6	0.263	0.231	3	Type-1 (x3)	(52)
2HRI	d[TAG <sub>3</sub> T <sub>2</sub> AG <sub>3</sub> ]	Porphyrin TMPyP4 complex	2.09	C222 <sub>1</sub>	0.257	0.208	2	Type-2	(68)
3CE5	d[TAG <sub>3</sub> T <sub>2</sub> AG <sub>3</sub> T]	Acridine complex	2.50	I4	0.213	0.184	2	Type-3	
3CCO	d[TAG <sub>3</sub> T <sub>2</sub> AG <sub>3</sub> T]	ND <sup>b</sup> complex	2.20	P6 <sub>4</sub> 22	0.303	0.257	1	Type-1	(70)
3CDM	d[TAG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	ND complex	2.10	P2 <sub>1</sub>	0.295	0.234	6	Type-7	
3IBK	r(U <sub>Br</sub> AG <sub>3</sub> U <sub>2</sub> AG <sub>3</sub> U)	Native RNA	2.20	P3 <sub>1</sub> 21	0.231	0.216	2 <sup>c</sup>	Type-4	(69)
3MIJ	r(UAG <sub>3</sub> U <sub>2</sub> AG <sub>3</sub> U)	RNA acridine complex	2.60	P23	0.248	0.236	1	Type-1 (x4)	(69)
3QCR	d[TAG <sub>3</sub> T <sub>2</sub> AG <sub>3</sub> T]	Acridine complex.	3.20	P6 <sub>5</sub> 22	0.378	0.346	1 <sup>d</sup>	Type-5	
3SC8	d[AG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	ND complex	2.30	P3 <sub>1</sub> 21	0.288	0.260	3	Type-6	(71)
3T5E	d[AG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	ND complex	2.10	P6	0.282	0.245	3	Type-1	(72)
3QSC	d[AG <sub>3</sub> T <sub>2</sub> AG <sub>3</sub> T <sub>2</sub> ]	Salphen complex	2.40	C222	0.234	0.214	1	Type-1 (x3)	(72)
3QSF	d[AG <sub>3</sub> T <sub>2</sub> AG <sub>3</sub> T <sub>2</sub> ]	Salphen complex	2.40	C222	0.320	0.240	1	Type-9	(75)
3UYH	d[AG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	ND complex	1.95	P3 <sub>1</sub> 21	0.275	0.231	3	Type-10	(76)
4DA3	d[G <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	ND complex	2.40	P3 <sub>1</sub> 21	0.282	0.242	3	Type-11	
4DAQ	d[G <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	ND complex	2.75	P3 <sub>1</sub> 21	0.275	0.212	3	Type-1	(76)
3R6R	d[AG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	Berberine complex	2.30	P6	0.252	0.217	3	Type-10	
4G0F	d[AG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	Mesoporphyrin complex	2.15	P6	0.262	0.239	3	Type-11	(73)
4FXM	d[AG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	Mesoporphyrin complex	1.65	P2 <sub>1</sub> 2 <sub>1</sub> 2	0.262	0.224	3	Type-1 (x3)	(74)
Total no. of experimentally determined TTA or UUA loops: 43									
Total no. of distinct TTA loop types: 12									
Total no. of type-1 TTA loops occurrences: 26									
Total number of crystal structures: 19									
Total no. of distinct space groups: 10									
Total no. of HTQ-ligand complexes: 16									

<sup>a+c</sup>Chain B loop not resolved in the electron density; <sup>b</sup>ND, Naphthalene diimide; <sup>d</sup>Loop poorly resolved in the electron density.

of all HTQ crystal structures analysed in this work using the native HTQ structure (PDB ID 1KF1) as a reference gives an average r.m.s.d of just 0.86 Å (see Supplementary Table S1 and Figure S1). This stability of the HTQ G-quartets has also been confirmed by molecular dynamics simulations (88–91). Accordingly, the geometry of the G-quartet core has not been surveyed here.

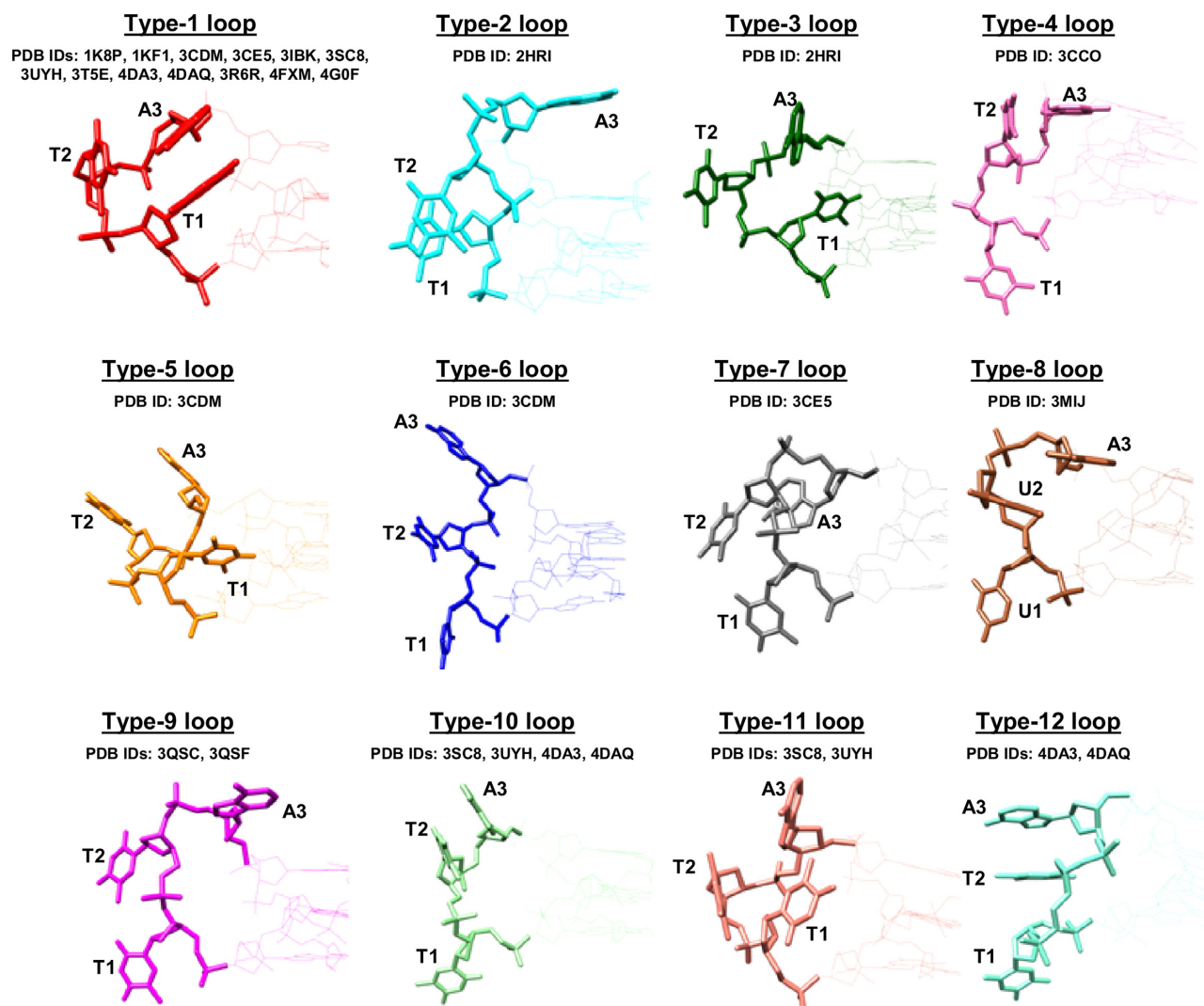
### Methodology used in the survey

Atomic coordinates for the telomeric quadruplex structures were extracted from the Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)), with geometric and conformational analyses performed using the 3DNA program (92) at [www.x3dna.org](http://www.x3dna.org) and visualised with the programs CHIMERA (93) at [www.cgl.ucsf.edu/chimera](http://www.cgl.ucsf.edu/chimera) and PyMOL (94) at [www.pymol.org](http://www.pymol.org). Each loop was analysed individually and no averaging of conformational features was done. Disordered and poorly resolved loops were excluded from the analyses. Sugar puckers were assigned directly from pseudorotation parameters calculated by 3DNA from crystallographic coordinates.

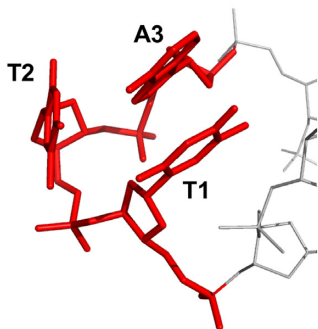
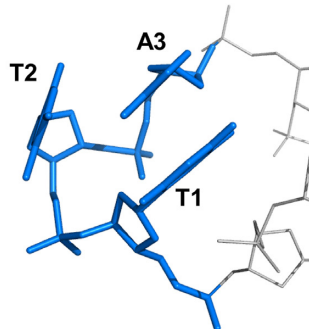
### STRUCTURAL ANALYSIS AND BACKBONE DIHEDRAL ANGLE COMPARISON OF HTQ TTA LOOPS

There are currently 19 X-ray crystal structures of native and small-molecule-complexed human telomeric DNA and RNA G-quadruplexes available in the Protein Data Bank. All of these contain d(TTA) (or r(UUA)) propeller loops (Figure 1a, Table 1). This data set comprises 17 DNA and two RNA (i.e. TERRA/telRNA) structures. The structures of individual ligands are shown in Figure 1b. These bimolecular and unimolecular quadruplex native and ligand-bound structures represent a total of 43 crystallographically-determined propeller-type loops (this number excludes three poorly-resolved loops: see Table 1). The range of resolution in these analyses is 1.65–3.2 Å, with an average of 2.30 Å. The loops have been categorised into 12 distinct groups (termed type-1, type-2... etc) on the basis of their overall three-dimensional shape (see Table 2 and Figure 2a, b).

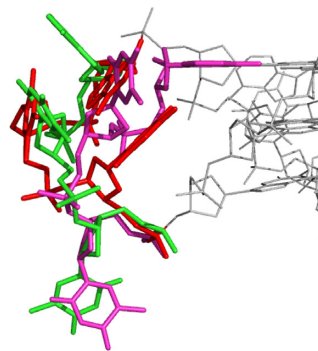
A



B

i) **Type-1 loop, sub-type 1****Type-1 loop, sub-type 2**

ii)



**Figure 2.** (A) Views of the various TTA loop types (coloured as in Figure 1), grouped into 12 distinct categories. (B) (i) The two sub-types of type-1 loops. Sub-type 1 (left) is from PDB structure 1KF1, sub-type 2 (right) is from PDB structure 3T5E. The structures shown here have been structurally aligned and are presented in the same orientation. (ii) An example of the structural diversity of non-type-1 loops: overlay of a type-4 loop (pink, from PDB structure 3CCO), a type-10 loop (green, from PDB structure 3SC8) and a type-1 loop (red, from PDB structure 1KF1). The non-type-1 loops are dramatically different to the 'native' type-1 TTA loop—as well as to one another. Note the retention of base-stacking interactions in the type-10 loop arrangement, which is a common feature of non-type-1 loops.

### Type-1 propeller loops

By far the most commonly observed loop type is the intercalated ‘TAT’ loop found in the native telomeric DNA structures (PDB IDs 1KF1 and 1K8P), termed type-1 loops. This loop arrangement always involves the adenine residue  $\pi$ - $\pi$  stacking above the first thymine, with the second (central) thymine adopting one of two similar orientations: either positioned perpendicular to the G-quartet planes or stacked (or nearly stacked) on the external face of the adenine (Figure 2b (i)). Type-1 loops have been found to occur in 13 crystal structures to date (including both native and ligand-complexes), with a total of 26 occurrences overall (therefore comprising over 60% of all crystallographically-observed TTA loops). This strongly suggests that the type-1 propeller loop is an energetically favourable and stable arrangement, which is independent of crystal packing mode—Table 1 also shows that these structures occur in a variety of crystallographic space groups (five distinct space groups in total). It is perhaps significant that, although ligand binding events often result in significant rearrangement of the type-1 loop geometry (described below), ten distinct quadruplex-ligand complexes exist in which the type-1 loop is preserved. This implies further that the type-1 loop arrangement is a structurally robust motif. Comparison of the backbone and glycosidic dihedral angles for the type-1 loops of each of the 13 structures which contain such a loop reveals a high degree of conservation in backbone geometry (Figure 3a). The only significant deviation can be seen for a 23-mer naphthalene diimide derivative complex structure (PDB ID 3CDM), which has changes in the  $\epsilon$  and  $\zeta$  angles for the loop adenine residue (Figure 3a).

Backbone dihedral comparison for the structures which contain multiple type-1 loops reveals a similar level of angle conservation. For example, the native 22-mer structure (PDB ID 1KF1) and the naphthalene diimide derivative-complexed 22-mer (PDB ID 3T5E, with identical sequence) both contain three type-1 loops, which display a significant level of intramolecular correlation when comparing backbone dihedrals (Figure 3b, c, respectively). A similar level of correlation in dihedral angles is also seen within the 22-mer complexes involving mesoporphyrin (PDB ID 4FXM) and berberine (PDB ID 3R6R) (Figure 3d, e, respectively).

### Non type-1 propeller loops

Of the 43 crystallographically observed TTA (or UUA) propeller loops, less than 40% (17 in total) adopt conformations other than the intercalated type-1 loop arrangement described above. These non-type-1 loops adopt a wide variety of dissimilar and diverse structures in the crystal state, which can be divided into 11 distinct groups (type-2 to type-12) based on the orientation of the three loop bases (Figure 2a and Tables 1 and 2). The structural diversity of the 11 non-type-1 loops indicates that TTA propeller loops possess a significant degree of structural flexibility, however, this does not necessarily signify that the crystallographically observed loop types are unstable. Indeed, the majority of these non-type-1 loop arrangements share the common feature of exploiting intra-loop base  $\pi$ -stacking interactions (Figure 2a and b(ii) and Table 2)—a feature which would

be expected to confer a degree of stability upon these loop structures.

Comparison of backbone and glycosidic dihedral angles of the 11 non-type-1 loops reveals a broad distribution in backbone geometry for the different loop arrangements (Figure 3f), which is to be expected, considering the dramatic structural variety of non-type-1 loops (Figure 2). However, there do appear to be several favourable groupings of torsion angles—such as a preference for *anti* glycosidic bond angles, particularly for the second loop thymine—in addition to some correlation with the type-1 backbone dihedral angle distribution. As with type-1 loops, backbone dihedral angle comparison of equivalent loop types between different structures reveals a high level of similarity, as seen when comparing the type-9 loops observed in the two bimolecular salphen-complexed quadruplexes 3QSF and 3QSC (Figure 3g).

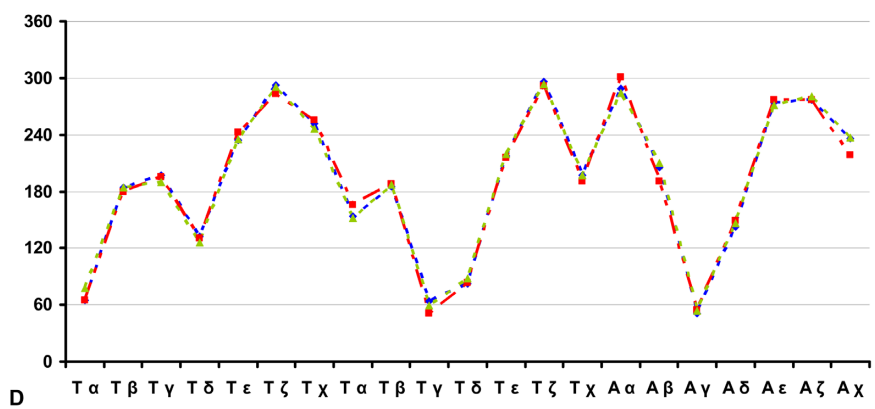
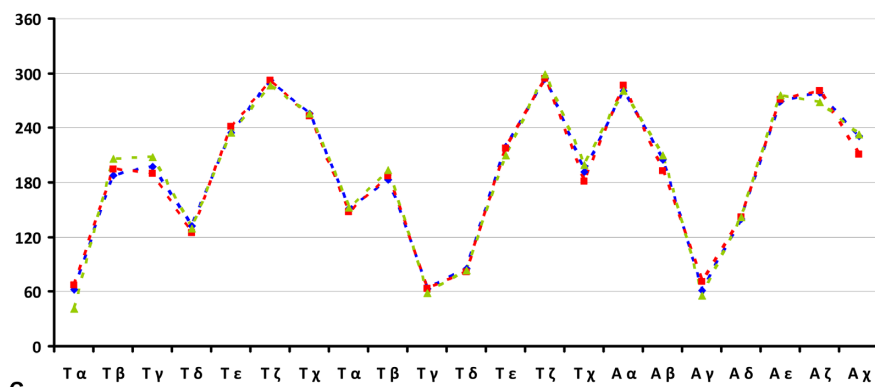
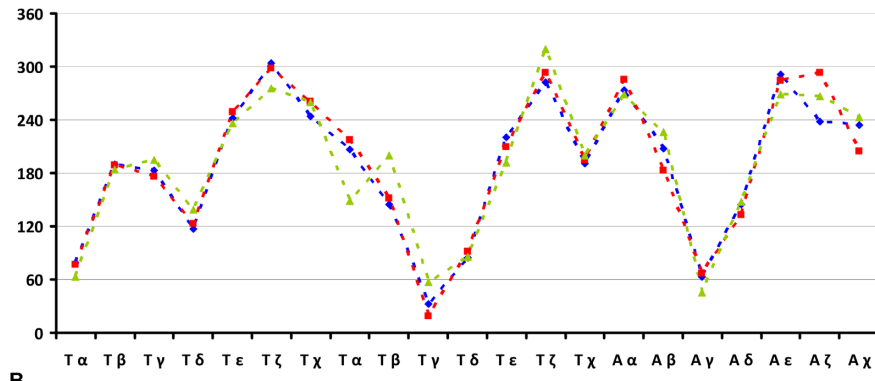
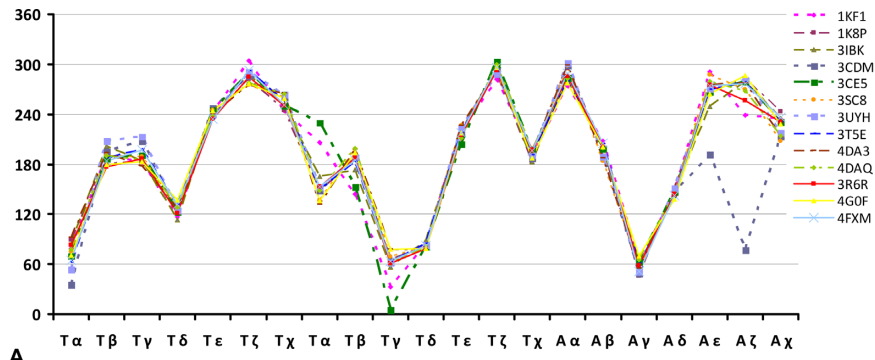
Although type-1 loops are indeed observed in a number of HTQ-ligand complexes (as stated above), all non-type-1 loops occur in ligand-bound HTQ structures, thus, non-type-1 loops are almost certainly a consequence of ligand-binding events. If one considers the type-1 loop to represent the ‘native’ TTA loop conformation, the observation of a further 11 distinct loop conformations—many of which are dramatically different to the type-1 arrangement (see for example Figure 2b(ii))—indicates clearly that TTA propeller loops of human telomeric G-quadruplexes possess a significant degree of structural polymorphism. This observation would be expected to have important consequences with respect to the rational design of small molecules targeted towards HTQs, however, the almost paradoxical observation of both type-1 and non-type-1 loops within a single ligand-bound HTQ structure (e.g. PDB IDs 3CDM, 3CE5, 3SC8, 3UYH, 4DA3, 4DAQ) precludes a clear conclusion: small molecule compounds are able to elicit both dramatic and negligible effects on TTA loop geometry.

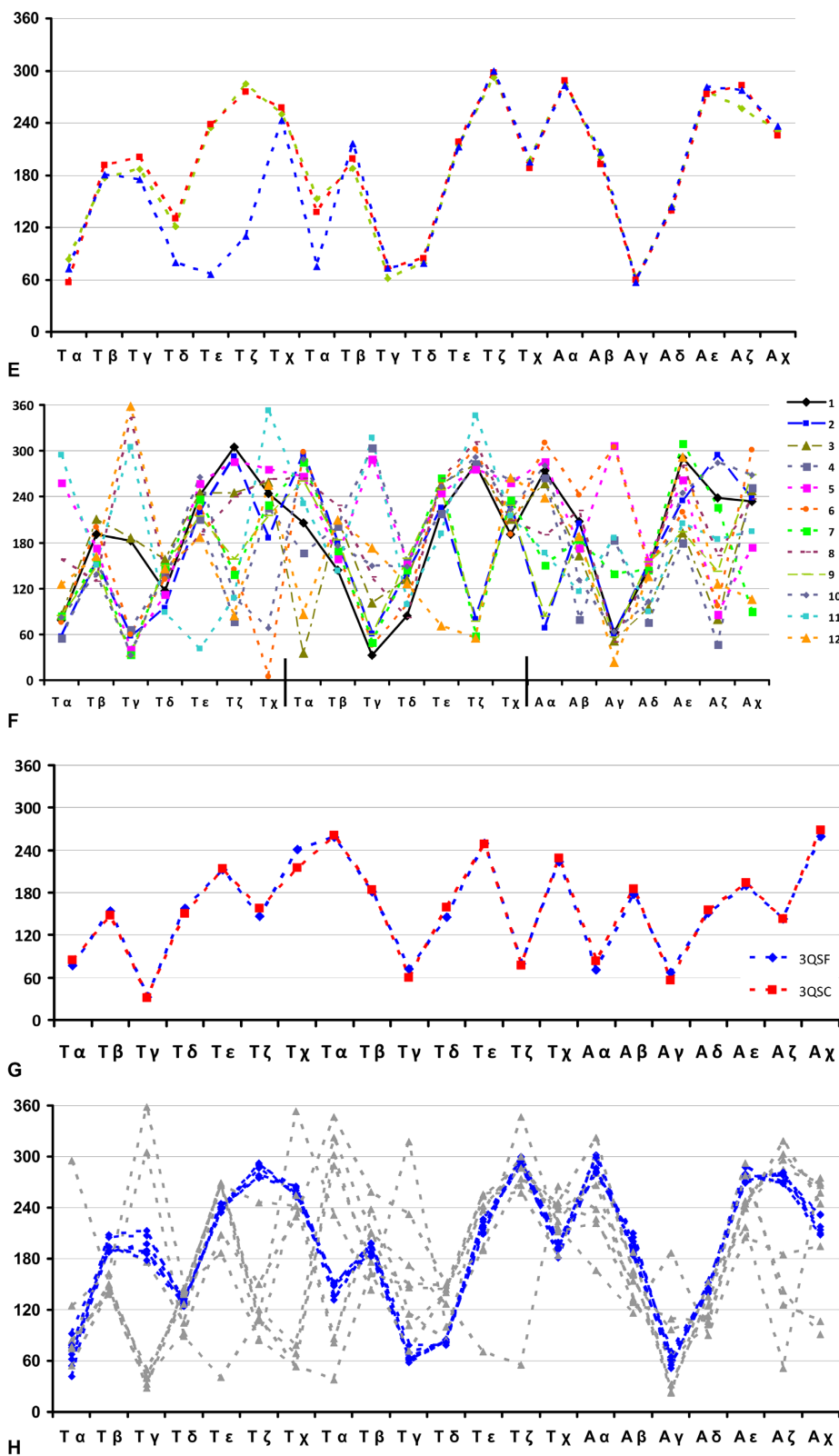
## SUGAR PUCKER ANALYSIS AND COMPARISON OF HTQ TTA LOOPS

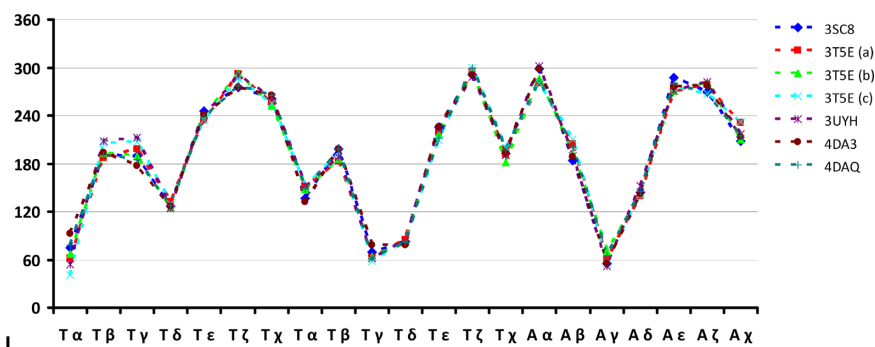
### Sugar pucker distribution of type-1 propeller loops

Comparison of the deoxyfuranose and furanose sugar puckers in the nucleotides of all type-1 propeller loops reveals an exceptionally high degree of similarity in pucker distribution. Of the 26 type-1 loops observed in crystal structures, 17 (65.4%) adopt a sugar pucker distribution of C1'-*exo*, C3'-*endo*, C2'-*endo* respectively for the three T-T-A nucleotides in a loop (Table 3). Interestingly, this consensus pattern of pucker arrangement—a similar pucker range to those found in DNA duplex crystal structures—is also observed for RNA G-quadruplex type-1 loops. Analysis of the frequency of pucker-types for each loop residue (i.e. T, T, A/U, U, A) reveals a high level of correspondence, with the first thymine sugar adopting a C1'-*exo* pucker in 69.2% of observed instances, the second thymine adopting a C3'-*endo* pucker conformation in 100% of the observed instances and the adenine sugar adopting a C2'-*endo* pucker in 88.5% of all type-1 loops (Table 4). The high level of correlation in the distribution of sugar pucker geometry and backbone dihedral angles of type-1 loops strongly implies that the folding









**Figure 3.** Plots of nucleotide conformational angles in the TTA loops. (A) *Backbone and glycosidic torsion angles.* These are shown for all structures containing at least one type-1 TTA (or UUA) loop. For structures which contain more than one type-1 loop only the first TTA loop is represented. (B) *Comparison of backbone and glycosidic torsions angles.* These are shown of the three type-1 propeller loops of the native human telomeric unimolecular quadruplex (PDB ID 1KF1) (each loop coloured separately; red, green or blue). (C) *Comparison of backbone and glycosidic torsions angles.* These are for the three type-1 propeller loops of a naphthalene diimide telomeric quadruplex complex (PDB ID 3T5E) (each loop coloured separately; red, green or blue). (D) *Comparison of the backbone and glycosidic torsions angles.* These are for the three type-1 propeller loops of the mesoporphyrin complex with a human telomeric quadruplex (PDB ID 4FXM) (each loop coloured separately; red, green or blue). (E) *Comparison of the backbone and glycosidic torsions angles.* These are for the three type-1 propeller loops in the berberine complex with a human telomeric quadruplex (PDB ID 3R6R) (each loop coloured separately; red, green or blue). (F) *Dihedral angle plots for each of the TTA residues of each of the 12 loop types* (each trace represents a distinct loop type). The loop 1 of 1KF1 (solid black line) can be used as a common reference. Dihedral angles are derived from loops from PDB structures 1KF1 (1), 2HRI (2 and 3), 3CCO (4), 3CDM (5 and 6), 3CE5 (7), 3MIJ (8), 3QSC (9), 3SC8 (10 and 11) and 4DA3 (12). (G) *Type-9 loop backbone and glycosidic dihedral angles.* These are from bimolecular human quadruplexes (PDB IDs 3QSF and 3QSC) complexed with salphen ligands. (H) *Comparison of backbone and glycosidic dihedral angles in the TTA loops of the naphthalene diimide (ND) complex structures.* Type-1 TTA loops are shown as blue dashed lines. Non-type-1 loops are shown as grey dashed lines. Structures represented are: PDB IDs 3SC8, 3T5E, 3UYH, 4DA3 and 4DAQ. (I) *Comparison of TTA loop backbone and glycosidic dihedrals of naphthalene diimide complex crystal structures, for type-1 loops only.*

**Table 2.** Qualitative TTA/UUA loop descriptors in terms of base stacking

Loop type	Stacked bases	Mutually perpendicular bases	Orientation of 3 <sup>rd</sup> base
1	T1, A3	T2, A3	T2 close to A3
2	T1, T2	-	A3 away from T1, T2
3	T2, A3	T1, A3	T1 close to A3
4	-	T2, A3	T1 away from T2, A3
5	T2, A3	-	T1 close to T2, A3
6	T2, A3	T1	T1 away from T2, A3
7	T2, A3	-	T1 close to T2
8	U2, A3	U1	U1 away from U2, A3
9	No stacking	-	All away from each other
10	T2, A3	-	T1 away from T2, A3
11	No stacking	T1, T2	A3 close to T1
12	T2, A3	-	T1 away from T2, A3

of TTA triplets into the type-1 arrangement is a highly precise, non-random process.

### Sugar pucker distribution of non-type-1 propeller loops

As with backbone dihedral angle distribution, the sugar pucker of non-type-1 loops is also highly variable when comparing all 11 non-type-1 loops (Table 5). Although there is no common consensus pucker arrangement for the non-type-1 loops, there is a clear shift from the 'C1'-*exo*, C3'-*endo*, C2'-*endo*' consensus of type-1 loops to a C2'-*endo* conformation for *all* loop residues. For example, the first thymine of the TTA loop adopts a C2'-*endo* pucker in 52.9% of all non-type-1 propeller loops, compared to 23.1% in all type-1 loops (6/26 occurrences) (Table 6). This presumably reflects the energetic preference for C2'-*endo* puckered sugars in DNA, and suggests the less favourable C1'-*exo* and C3'-*endo* of the first two residues of the type-1 loops may be compensated for by the favourable energetics of base-base  $\pi$ -stacking interactions.

### ANALYSIS AND CONCLUSIONS

The solution NMR evidence for type-1 loops, which together with the diversity of crystal packing modes in the crystal structures, suggests that crystal packing forces are not contributing to the loop types. The diversity of space groups and resulting patterns of molecular packing in these crystals suggests that the intermolecular packing forces are weak and therefore different packing arrangements can readily occur. This is consistent with the assumption that what is observed in the crystalline state are low-energy conformations of these complexes. We suggest that here, the preferred conformations of the relatively flexible loops are always among those most observed in these crystal structures.

Although the type-1 loop is present in five distinct space groups, many of the non-type-1 loops are observed less frequently, and thus there is not currently sufficient data to make any definitive statements about the relationship between space group and loop type. It should be noted,

**Table 3.** Pucker distributions for TTA/UUAs of all type-1 loops

	1KF1-a	1KF1-b	1KF1-c	1K8P	3IBK	3CDM-a	3CDM-b	3CDM-c	3CDM-d	3CE5	3SC8	3UYH	3T5E-a
T	C1'- <i>exo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C4'- <i>exo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C2'- <i>endo</i>
T	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>
A	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C1'- <i>exo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>
	<b>3T5E-b</b>	<b>3T5E-c</b>	<b>4DA3</b>	<b>4DAQ</b>	<b>3R6R-a</b>	<b>3R6R-b</b>	<b>3R6R-c</b>	<b>4FXM-a</b>	<b>4FXM-b</b>	<b>4FXM-c</b>	<b>4G0F-a</b>	<b>4G0F-b</b>	<b>4G0F-c</b>
T	C1'- <i>exo</i>	C2'- <i>endo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C4'- <i>exo</i>	C2'- <i>endo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>
T	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>	C3'- <i>endo</i>
A	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C2'- <i>endo</i>

Yellow indicates a sugar which deviates from the consensus pucker conformation. Lower-case letters denote crystallographically unique type-1 loops within a single crystal structure.

**Table 4.** Percentage of each residue in type-1 loops which adopt sugar pucker consensus 'C1'-*exo*, C3'-*endo*, C2'-*endo*'

	Pucker type	% Observed
T	18/26 C1'- <i>exo</i>	69.2%
T	26/26 C3'- <i>endo</i>	100.0%
A	23/26 C2'- <i>endo</i>	88.5%

C3'-*endo* pucker for the second loop thymine is 100% conserved.

however, that the relationship between space group and (loop) conformation is not necessarily a one way process—a particular space group may affect the loop geometry, but equally, a loop type can influence the space group into which the molecule/DNA/RNA packs, i.e. loop-type may define the crystal packing, rather than the packing defining the loop-type.

Although there is no direct evidence that non-type-1 loops are produced by ligand binding, it is clear that ligand binding is involved in some way in the formation of non-type-1 loop arrangements. Since type-1 loops are also observed in the native DNA and RNA structures, we suggest that some ligands simply do not perturb this loop geometry, whereas others do.

The overall question of the relationship of these crystal structures to the structures and complexes present in solution can only be addressed once much more extensive data are obtained. However, there is certainly sufficient data for some tentative conclusions to be drawn. One important issue concerns the loops defined here—in particular, whether or not these structures are present in solution. Importantly, there is indeed evidence for the existence of type-1 loops in solution. The NMR structure of d[TAGGG(TTAGGG)]<sub>3</sub> in crowded conditions (PDB ID 2LD8) clearly shows stacked type-1 loops for all three TTA loops of all 10 structures present in the deposited ensemble (51). Thus, the type-1 loop—which is observed frequently crystallographically—appears to be highly stable and is very much present in solution. A structural alignment of the crystal and NMR structures, 1KF1 with 2LD8, (excluding the 5' flanking residues) gives an average RMSD of 1.981 Å (see Supplementary Table S2)—this is an excellent correlation between solution and solid state. The overall visual alignment of these two structures (1KF1 and 2LD8) may be observed by scrolling through the NMR ensemble and crystal structures and is very close. Figure 2 and Table 2 also show it is common for the loops to have at least one unstacked base, which is highly likely to be adopting multiple conformations in solution. The patterns of base-base

stacking seen in almost all the loop types, on the other hand, are more likely to have significant populations in solution.

An individual crystal structure provides quasi-static information about molecular structures—information embedded in temperature (B) factors in part reflects crystal lattice dynamics. Data on an ensemble of quadruplex crystal structures in which the principal variables are the bound small molecules, can, on the other hand, provide a detailed view of conformational trends in variable regions, i.e. the loops. The over-riding conclusion from this survey of available crystal structures is that these HTQ-ligand complexes have a significant degree of loop conformational variability (albeit with certain trends of conformational preference), which adds another layer of complexity to HTQ structures, and which needs to be taken into account when considering them as drug targets for *in silico* screening and molecular design studies. This feature of loops is likely to be common to all quadruplexes with loops, whether they contain single-nucleotide or much longer loops.

One key question posed by this survey is whether a particular loop conformational class is related to the nature of the bound ligand. The type-1 loop occurs at least once in each of all six naphthalene diimide complexes, in the berberine complex and in both mesoporphyrin complexes, whereas it does not occur in the salphen, or in the other porphyrin (TMPyP4) complex. Figure 3h, i shows that for the naphthalene diimide complexes, there is close correspondence between the detailed conformations of type-1 loops at the individual torsion angle level, but this is not the case for the other type-1 loops, where there is definite (albeit sometimes small) conformational variation between them. The type-1 loop also occurs in the structure of the DNA quadruplex acridine derivative (BRACO-19) complex, but not in the RNA one—one has to bear in mind that the acridines in these two structures are structurally very distinct from one another. The native RNA quadruplex shares with the two native DNA quadruplexes the preference for type-1 loops. The loop difference between the native and acridine-bound RNA quadruplexes is due mainly to the presence of the acridine, which exploits (and presumably stabilises) an extended binding surface created by the addition of loop adenine residues to the 5' G-quartet—a (re)arrangement very much dependent on hydrogen bonds involving the C2' hydroxyl groups of the ribose sugars. Overall the type-1 loop is the dominant one in the majority of these crystal structures.

Overall we conclude that loop types and conformations are most conserved within a closely-related group of small molecule complexes, as in the series of ND-quadruplex



**Table 5.** Sugar pucker distributions for each non-type-1 loop

	2RHI 2*	2RHI 3	3CCO 4	3CDM 5	3CDM 6	3CE5 7	3MIJ 8	3QSC 9	3QSF 9	3SC8 10	3SC8 11	3UYH 10	3UYH 11	4DA3 10	4DA3 12	4DAQ 10	4DAQ 12
T	C3'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C1'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C3'- <i>exo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C3'- <i>endo</i>	C2'- <i>endo</i>	O4'- <i>endo</i>	C2'- <i>endo</i>	C3'- <i>exo</i>	C1'- <i>exo</i>	C4'- <i>exo</i>
T	C1'- <i>exo</i>	C1'- <i>exo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C3'- <i>endo</i>	C2'- <i>endo</i>	C3'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C3'- <i>exo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C3'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>
A	C2'- <i>endo</i>	O4'- <i>endo</i>	C4'- <i>exo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C2'- <i>endo</i>	C1'- <i>exo</i>	C4'- <i>exo</i>	C1'- <i>exo</i>	C2'- <i>endo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>	C1'- <i>exo</i>

\*Numbers below PDB codes denote loop type.

**Table 6.** Occurrence of each pucker conformation in non-type-1 loops

	C1'- <i>endo</i>	C2'- <i>endo</i>	C3'- <i>endo</i>	C1'- <i>exo</i>	C2'- <i>exo</i>	C3'- <i>exo</i>	C4'- <i>exo</i>	O4'- <i>endo</i>	% C2'- <i>endo</i> for non-type-1 loops	% C2'- <i>endo</i> for type-1 loops
<b>T</b>	1	<b>9</b>	2	1	0	2	1	1	52.9% (9/17)	23.1% (6/26)
<b>T</b>	0	<b>10</b>	3	2	1	1	0	0	58.8% (10/17)	0.0% (0/26)
<b>A</b>	0	<b>8</b>	0	6	0	0	2	1	47.1% (8/17)	88.5% (23/26)

There is a pronounced shift towards C2'-*endo* puckering for all residues of the TTA loops for non-type-1 loops when compared to type-1 loops (highlighted in bold).

structures. In these, the differences between the structures are in the nature of the ND end-groups—all the ND compounds are tetra-substituted but with different cationic termini. There is some degree of structure conservation of type-1 loops beyond the ND complexes, especially when the ligands are not over-sized, as is the case for the acridine BRACO-19 and the berberine derivative. The implications for rational drug design of quadruplex-binding ligands are clear: *in silico* modelling based on existing crystal structures is most reliable when modelling closely similar analogues, but even then the variability in some loops is unpredictable and can best be determined by further experimental structures rather than for example, molecular dynamics with restricted simulation times. Taking one of the existing crystal structures (Table 1) as a starting-point will therefore be useful in providing qualitative guidance on optimising existing or related ligand hits, but extension for screening of structurally diverse libraries and scaffolds needs to be approached with caution.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

Cancer Research UK [C129/A4489]. Funding for open access charge: UCL.

Conflict of interest statement. None declared.

## REFERENCES

- Davies, J.T. (2004) G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Intl. Edit.*, **43**, 668–698.
- Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
- Bochman, M.L., Paeschke, K. and Zakian, V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.*, **13**, 770–780.
- Sun, D. and Hurley, L.H. (2009) The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J. Med. Chem.*, **52**, 2863–2874.
- Meyne, J., Ratliff, R.L. and Moyzis, R.K. (1989) Conservation of the human telomere sequence (TTAGGG)<sub>n</sub> among vertebrates. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 7049–7053.
- Wright, W.E., Tesmer, V.M., Huffman, K.E., Levene, S.D. and Shay, J.W. (1997) Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev.*, **11**, 2801–2809.
- Zaug, A.J., Podell, E.R. and Cech, T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 10864–10869.
- Nandakumar, J. and Cech, T.R. (2013) Finding the end: recruitment of telomerase to telomeres. *Nat. Rev. Mol. Cell Biol.*, **14**, 69–82.
- Sun, D., Thompson, B., Cathers, B.E., Salazar, M., Kerwin, S.M., Trent, J.O., Jenkins, T.C., Neidle, S. and Hurley, L.H. (1997) Inhibition of human telomerase by a G-quadruplex-interactive compound. *J. Med. Chem.*, **40**, 2113–2116.
- De Cian, A., Lacroix, L., Douarre, C., Temime-Smaali, N., Trentesaux, C., Riou, J.-F. and Mergny, J.-L. (2008) Targeting telomeres and telomerase. *Biochimie*, **90**, 131–155.
- Monchaud, D. and Teulade-Fichou, M.-P. (2008) A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.*, **6**, 627–636.
- Sissi, C. and Palumbo, M. (2014) Telomeric G-quadruplex architecture and interactions with potential drugs. *Curr. Pharm. Design*, **20**, 6489–6509.
- Zhang, S., Wu, Y. and Zhang, W. (2014) G-quadruplex structures and their interaction diversity with ligands. *ChemMedChem*, **9**, 899–911.
- Sekaran, V., Soares, J. and Jarstfer, M.B. (2014) Telomere maintenance as a target for drug discovery. *J. Med. Chem.*, **57**, 521–638.
- Hahn, W.C., Stewart, S.A., Brooks, M.W., York, S.G., Eaton, E., Kurachi, A., Beijersbergen, R.L., Knoll, J.H., Meyerson, M. and Weinberg, R.A. (1999) Inhibition of telomerase limits the growth of human cancer cells. *Nat. Med.*, **5**, 1164–1170.
- Hanahan, D. and Weinberg, R.A. (2011) The hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.
- Zahler, A.M., Williamson, J.R., Cech, T.R. and Prescott, D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350**, 718–750.
- Perry, P.J., Gowan, S.M., Reszka, A.P., Polucci, P., Jenkins, T.C., Kelland, L.R. and Neidle, S. (1998) 1,4- and 2,6-disubstituted amidoanthracene-9,10-dione derivatives as inhibitors of human telomerase. *J. Med. Chem.*, **41**, 3253–3260.
- Li, Q., Xiang, J.F., Yang, Q.F., Sun, H.X., Guan, A.J. and Tang, Y.L. (2013) G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res.*, **41**, D1115–D1123.
- Burger, A.M., Dai, F., Schultes, C.M., Reszka, A.P., Moore, M.J., Double, J.A. and Neidle, S. (2005) The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res.*, **65**, 1489–1496.

21. Phatak,P., Cookson,J.C., Dai,F., Smith,V., Gartenhaus,R.B., Stevens,M.F.G. and Burger,A.M. (2006) Telomere uncapping by G-quadruplex ligand RHPS4 inhibits clonogenic tumour cell growth in vitro and in vivo consistent with a cancer stem cell targeting mechanism. *Brit. J. Cancer*, **96**, 1223–1233.
22. Tauchi,T., Shin-ya,K., Sashido,G., Sumi,M., Okabe,S., Ohyashiki,J.H. and Ohyashiki,K. (2006) Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: in vitro and in vivo studies in acute leukemia. *Oncogene*, **25**, 5719–5725.
23. Cuenca,F., Greciano,O., Gunaratnam,M., Haider,S., Munnur,D., Nanjunda,R., Wilson,W. D. and Neidle,S. (2008) Tri- and tetra-substituted naphthalene diimides as potent G-quadruplex ligands. *Bioorg. Med. Chem. Lett.*, **18**, 1668–1673.
24. Gomez,D., O'Donohue,M. F., Wenner,T., Douarre,C., Macadré,J., Koebel,P., Giraud-Panis,M.-J., Kaplan,H., Kolkes,A., Shin-Ya,K. et al. (2006) The G-quadruplex ligand telomestatin inhibits POT1 binding to telomeric sequences in vitro and induces GFP-POT1 dissociation from telomeres in human cells. *Cancer Res.*, **66**, 7908–7912.
25. Salvati,E., Leonetti,C., Rizzo,A., Scarsella,M., Mottolose,M., Galati,R., Sperduti,I., Stevens,M.F.G., D'Incalci,M., Biasco,M. et al. (2007) Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *J. Clin. Investig.*, **117**, 3236–3247.
26. Gunaratnam,M., Greciano,O., Martins,C., Reszka,A.P., Schultes,C.M., Morjani,H., Riou,J.-F. and Neidle,S. (2007) Mechanism of acridine-based telomerase inhibition and telomere shortening. *Biochem. Pharmacol.*, **74**, 679–689.
27. Rodriguez,R., Müller,S., Yeoman,J. A., Trentesaux,C., Riou,J.-F. and Balasubramanian,S. (2008) A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Amer. Chem. Soc.*, **130**, 15758–15789.
28. Rizzo,A., Salvati,E., Porru,M., D'Angelo,C., Stevens,M.F., D'Incalci,M., Leonetti,C., Gilson,E., Zupi,G. and Biroccio,A. (2009) Stabilization of quadruplex DNA perturbs telomere replication leading to the activation of an ATR-dependent ATM signaling pathway. *Nucleic Acids Res.*, **37**, 5353–5364.
29. Müller,S., Sanders,D.A., Di Antonio,M., Matsis,S., Riou,J.-F., Rodriguez,R. and Balasubramanian,S. (2012) Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells. *Org. Biomol. Chem.*, **10**, 6537–6546.
30. Rodriguez,R., Miller,K.M., Forment,J.V., Bradshaw,C.R., Nikan,M., Britton,S., Oelschlaegel,T., Xhemalce,B., Balasubramanian,S. and Jackson,S.P. (2012) Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.*, **8**, 301–310.
31. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
32. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
33. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
34. Eddy,J. and Maizels,N. (2008) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.*, **36**, 1321–1333.
35. Huppert,J.L., Bugaut,A., Kumari,S. and Balasubramanian,S. (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.*, **36**, 6260–6268.
36. Halder,K., Wieland,M. and Hartig,J.S. (2009) Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes. *Nucleic Acids Res.*, **37**, 6811–6817.
37. Bugaut,A. and Balasubramanian,S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*, **40**, 4727–4741.
38. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress *c-MYC* transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 11593–11598.
39. Balasubramanian,S., Hurley,L.H. and Neidle,S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.*, **10**, 261–275.
40. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
41. Henderson,A., Wu,Y., Huang,Y.C., Chavez,E.A., Platt,J., Johnson,F.B., Brosh,R.M. Jr, Sen,D. and Lansdorp,P.M. (2013) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.*, **42**, 860–869.
42. Biffi,G., Tannahill,D., Miller,J., Howat,W. J. and Balasubramanian,S. (2014) Elevated levels of G-quadruplex formation in human stomach and liver cancer tissues. *PLoS One*, **9**, e102711.
43. Wang,Y. and Patel,D.J. (1993) Solution structure of the human telomeric repeat d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] G-tetraplex. *Structure*, **1**, 263–282.
44. Phan,A.T. and Patel,D.J. (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J. Amer. Chem. Soc.*, **125**, 15021–15027.
45. Ambrus,A., Chen,D., Dai,J., Bialis,T., Jones,R.A. and Yang,D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.*, **34**, 2723–2735.
46. Phan,A.T., Luu,K.N. and Patel,D.J. (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K<sup>+</sup> solution. *Nucleic Acids Res.*, **34**, 5715–5719.
47. Luu,K.N., Phan,A.T., Kuryavyi,V., Lacroix,L. and Patel,D.J. (2006) Structure of the human telomere in K<sup>+</sup> solution: an intramolecular (3+1) G-quadruplex scaffold. *J. Amer. Chem. Soc.*, **128**, 9963–9970.
48. Dai,J., Carver,M., Punchihewa,C., Jones,R.A. and Yang,D. (2007) Structure of the hybrid-2 type intramolecular human telomeric G-quadruplex in K<sup>+</sup> solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.*, **35**, 4927–4740.
49. Phan,A.T., Kuryavyi,V., Luu,K.N. and Patel,D.J. (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K<sup>+</sup> solution. *Nucleic Acids Res.*, **35**, 6517–6525.
50. Lim,K.W., Amrane,S., Bouaziz,S., Xu,W., Mu,Y., Patel,D.J., Luu,K.N. and Phan,A.T. (2009) Structure of the human telomere in K<sup>+</sup> solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J. Amer. Chem. Soc.*, **131**, 4301–4309.
51. Heddi,B. and Phan,A.T. (2013) Structure of human telomeric DNA in crowded solution. *J. Amer. Chem. Soc.*, **133**, 9824–9833.
52. Parkinson,G.N., Lee,M.P.H. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
53. Campbell,N., Collie,G.W. and Neidle,S. (2012) Crystallography of DNA and RNA G-quadruplex nucleic acids and their ligand complexes. *Curr. Protoc. Nucleic Acid Chem.*, Ch. 17, Unit 17.6.
54. Phan,A.T., Kuryavyi,V., Burge,S., Neidle,S. and Patel,D.J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human *c-kit* promoter. *J. Amer. Chem. Soc.*, **129**, 4386–4392.
55. Todd,A.K., Haider,S.M., Parkinson,G.N. and Neidle,S. (2007) Sequence occurrence and structural uniqueness of a G-quadruplex in the human *c-kit* promoter. *Nucleic Acids Res.*, **35**, 5799–5808.
56. Hsu,S.T., Varnai,P., Bugaut,A., Reszka,A.P., Neidle,S. and Balasubramanian,S. (2009) A G-rich sequence within the *c-kit* oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. *J. Amer. Chem. Soc.*, **131**, 13399–13409.
57. Kuryavyi,V., Phan,A.T. and Patel,D.J. (2010) Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human *c-kit2* promoter. *Nucleic Acids Res.*, **38**, 6757–6773.
58. Wei,D., Parkinson,G.N., Reszka,A.P. and Neidle,S. (2012) Crystal structure of a *c-kit* promoter quadruplex reveals the structural role of metal ions and water molecules in maintaining loop conformation. *Nucleic Acids Res.*, **40**, 4691–4700.
59. Wei,D., Husby,J. and Neidle,S. (2015) Flexibility and structural conservation in a *c-KIT* G-quadruplex. *Nucleic Acids Res.*, **43**, 629–644.
60. Dai,J., Chen,D., Jones,R.A., Hurley,L.H. and Yang,D. (2006) NMR solution structure of the major G-quadruplex structure formed in the human *BCL2* promoter region. *Nucleic Acids Res.*, **34**, 5133–5144.
61. Agrawal,P., Lin,C., Mathad,R.I., Carver,M. and Yang,D. (2014) The major G-quadruplex formed in the human *BCL-2* proximal promoter adopts a parallel structure with a 13-nt loop in K<sup>+</sup> solution. *J. Amer. Chem. Soc.*, **136**, 1750–1753.

62. Ambrus, A., Chen, D., Dai, J., Jones, R.A. and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048–2058.
63. Mathad, R.I., Hatzakis, E., Dai, J. and Yang, D. (2011) c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. *Nucleic Acids Res.*, **39**, 9023–9033.
64. Wei, D., Todd, A.K., Zloh, M., Gunaratnam, M., Parkinson, G.N. and Neidle, S. (2013) Crystal structure of a promoter sequence in the B-raf gene reveals an intertwined dimer quadruplex. *J. Amer. Chem. Soc.*, **135**, 19319–19329.
65. Phan, A.T., Kuryavyy, V., Gaw, H.Y. and Patel, D.J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nature Chem. Biol.*, **1**, 167–173.
66. Dai, J., Carver, M., Hurley, L.H. and Yang, D. (2011) Solution structure of a 2:1 quindoline-c-MYC G-quadruplex: insights into G-quadruplex-interactive small molecule drug design. *J. Amer. Chem. Soc.*, **133**, 17673–17680.
67. Chung, W.J., Heddi, B., Hamon, F., Teulade-Fichou, M.-P. and Phan, A.T. (2014) Solution structure of a G-quadruplex bound to the bisquinolinium compound Phen-DC(3). *Angew. Chem. Int. Ed. Engl.*, **53**, 999–1002.
68. Parkinson, G.N., Ghosh, R. and Neidle, S. (2007) Structural basis for binding of porphyrin to human telomeres. *Biochemistry*, **46**, 2390–2397.
69. Parkinson, G.N., Cuenca, F. and Neidle, S. (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**, 1145–1156.
70. Campbell, N.H., Parkinson, G.N., Reszka, A.P. and Neidle, S. (2008) Structural basis of DNA quadruplex recognition by an acridine drug. *J. Amer. Chem. Soc.*, **130**, 6722–6724.
71. Collie, G.W., Sparapani, S., Parkinson, G.N. and Neidle, S. (2011) Structural basis of telomeric RNA quadruplex-acridine ligand recognition. *J. Amer. Chem. Soc.*, **133**, 2721–2728.
72. Collie, G.W., Promontorio, R., Hampel, S.M., Micco, M., Neidle, S. and Parkinson, G.N. (2012) Structural basis for telomeric G-quadruplex naphthalene diimide ligand targeting. *J. Amer. Chem. Soc.*, **134**, 2723–2731.
73. Bazzicalupi, C., Ferraroni, M., Bilia, A.R., Scheggi, F. and Gratteri, P. (2012) The crystal structure of human telomeric DNA complexed with berberine: an interesting case of stacked ligand to G-tetrad ratio higher than 1:1. *Nucleic Acids Res.*, **41**, 632–638.
74. Nicoludis, J.M., Miller, S.T., Jeffrey, P.D., Barrett, S.P., Rablen, P.R., Lawton, T.J. and Yatsunyk, L.A. (2012) Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. *J. Amer. Chem. Soc.*, **134**, 20446–20456.
75. Campbell, N.H., Karim, N.H., Parkinson, G.N., Gunaratnam, M., Petrucci, V., Todd, A.K., Vilar, R. and Neidle, S. (2012) Molecular basis of structure-activity relationships between salphen metal complexes and human telomeric DNA quadruplexes. *J. Med. Chem.*, **55**, 209–222.
76. Micco, M., Collie, G.W., Dale, A.G., Ohnmacht, S.A., Pazitna, I., Gunaratnam, M., Reszka, A.P. and Neidle, S. (2013) Structure-based design and evaluation of naphthalene diimide G-quadruplex ligands as telomere targeting agents in pancreatic cancer cells. *J. Med. Chem.*, **56**, 2959–2974.
77. Chung, W.J., Heddi, B., Tera, M., Iida, K., Nagasawa, K. and Phan, A.T. (2013) Solution structure of an intramolecular (3 + 1) human telomeric G-quadruplex bound to a telomestatin derivative. *J. Amer. Chem. Soc.*, **135**, 13495–13501.
78. Collie, G.W., Haider, S., Neidle, S. and Parkinson, G.N. (2010) A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.*, **38**, 5569–5580.
79. Cosconati, S., Marinelli, L., Trotta, R., Virno, A., Mayol, L., Novellino, E., Olson, A.J. and Randazzo, A. (2009) Tandem application of virtual screening and NMR experiments in the discovery of brand new DNA quadruplex groove binders. *J. Amer. Chem. Soc.*, **131**, 16336–16337.
80. Cosconati, S., Marinelli, L., Trotta, R., Virno, A., De Tito, S., Romagnoli, R., Pagano, B., Limongelli, V., Giancola, C., Baraldi, P.G. et al. (2010) Structural and conformational requisites in DNA quadruplex groove binding: another piece to the puzzle. *J. Amer. Chem. Soc.*, **132**, 6425–6433.
81. Alcaro, S., Musetti, C., Distinto, S., Casatti, M., Zagotto, G., Artese, A., Parrotta, L., Moraca, F., Costa, G., Ortuso, F. et al. (2013) Identification and characterization of new DNA G-quadruplex binders selected by a combination of ligand and structure-based virtual screening approaches. *J. Med. Chem.*, **56**, 843–855.
82. Renčičuk, D., Kejnovská, I., Školáková, P., Bednářová, K., Motlová, J. and Vorlíčková, M. (2009) Arrangements of human telomere DNA quadruplex in physiologically relevant K<sup>+</sup> solutions. *Nucleic Acids Res.*, **37**, 6625–6634.
83. Yu, H., Gu, X., Nakano, S., Miyoshi, D. and Sugimoto, N. (2012) Beads-on-a-string structure of long telomeric DNAs under molecular crowding conditions. *J. Amer. Chem. Soc.*, **134**, 20060–20069.
84. Petraccone, L., Malafronte, A., Amato, J. and Giancola, C. (2012) G-quadruplexes from human telomeric DNA: how many conformations in PEG containing solutions? *J. Phys. Chem. B*, **116**, 2294–2305.
85. Buscaglia, R., Miller, M.C., Dean, W.L., Gray, R.D., Lane, A.N., Trent, J.O. and Chaires, J.B. (2013) Polyethylene glycol binding alters human telomere G-quadruplex structure by conformational selection. *Nucleic Acids Res.*, **41**, 7934–7946.
86. Sabharwal, N.C., Savikhin, V., Turek-Herman, J.R., Nicoludis, J.M., Szalai, V.A. and Yatsunyk, L.A. (2014) N-methylmesoporphyrin IX fluorescence as a reporter of strand orientation in guanine quadruplexes. *FEBS J.*, **281**, 1726–1737.
87. Haider, S.M., Autiero, I. and Neidle, S. (2011) Surface area accessibility and the preferred topology of telomeric DNA quadruplex-ligand complexes. *Biochimie*, **93**, 1275–1279.
88. Cang, X., Šponer, J. and Cheatham, T.E. III (2011) Insight into G-DNA structural polymorphism and folding from sequence and loop connectivity through free energy analysis. *J. Amer. Chem. Soc.*, **133**, 14270–14279.
89. Cang, X., Šponer, J. and Cheatham, T.E., III (2011) Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes. *Nucleic Acids Res.*, **39**, 4499–4512.
90. Islam, B., Sgobba, M., Laughton, C., Orozco, M., Šponer, J., Neidle, S. and Haider, S. (2013) Conformational dynamics of the human propeller telomeric DNA quadruplex on a microsecond time scale. *Nucleic Acids Res.*, **41**, 2723–2735.
91. Rebič, M., Mocchi, F., Laaksonen, A. and Uličný, J. (2015) Multiscale simulations of human telomeric G-quadruplex DNA. *J. Phys. Chem. B*, **119**, 105–113.
92. Xiang, J.L. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
93. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comp. Chem.*, **25**, 1605–1612.
94. DeLano, W.L. (2008) The PyMOL Molecular Graphics System. Version 1.7.4. Schrodinger, LLC.