Selectivity of small molecule ligands for parallel and anti-parallel DNA G-quadruplex structures

Thomas P. Garner,[a] Huw E. L. Williams,[a] Katarzyna I. Gluszyk,[b] Stephen Roe,[b] Neil J. Oldham,[b] Malcolm F. G. Stevens,[c] John E. Moses*[b] and Mark S. Searle*[a]

Received 28th May 2009, Accepted 30th June 2009
First published as an Advance Article on the web 14th August 2009
DOI: 10.1039/b910505k

We report CD, ESI-MS and molecular modelling studies of ligand binding interactions with DNA quadruplex structures derived from the human telomeric repeat sequence (*h*-Tel) and the proto-oncogenic *c*-kit promoter sequence. These sequences form anti-parallel (both 2 + 2 and 3 + 1) and parallel conformations, respectively, and demonstrate distinctively different degrees of structural plasticity in binding ligands. With *h*-Tel, we show that an extended heteroaromatic 1,4-triazole (**TRZ**), designed to exploit π-stacking interactions and groove-specific contacts, shows some selectivity for parallel folds, however, the polycyclic fluorinated acridinium cation (**RHPS4**), which is a similarly potent telomerase inhibitor, shows selectivity for anti-parallel conformations implicating favourable interactions with lateral and diagonal loops. In contrast, the unique *c*-kit parallel-stranded quadruplex shows none of the structural plasticity of *h*-Tel with either ligand. We show by quantitative ESI-MS analysis that both sequences are able to bind a ligand on either end of the quadruplex. In the case of *h*-Tel the two sites have similar affinities, however, in the case of the *c*-kit quadruplex the affinities of the two sites are different and ligand-dependent. We demonstrate that two different small molecule architectures result in significant differences in selectivity for parallel and anti-parallel quadruplex structures that may guide quadruplex targeted drug-design.

Introduction

Guanine-rich nucleic acid sequences are prevalent within the human genome.[1–3] The observation that they are able to form four-stranded G-quadruplex structures in the presence of physiological concentrations of monovalent cations (Na^+ and K^+) raises the possibility that they serve some key biological function linked to the regulation of gene expression.[4–7] Evidence for quadruplex formation within oncogenic promoter sequences and within human telomeric DNA has presented the opportunity for new therapeutic approaches in which small molecule anti-cancer drugs artificially down-regulate gene expression by stabilising quadruplex structures.[8–11] A large number of quadruplex-interactive ligands have now been described, many with encouraging *in vivo* activity.[12,13] However, the diversity of quadruplex structural topologies,[14] linked to the length of the loops connecting the G-rich tracts,[15,16] adds further complexity to attempts to target specific gene promoter sequences in the absence of detailed structural information. The observation that small molecules interact at the ends of the quadruplex with the terminal G-tetrads, rather than intercalating between the planes and displacing the monovalent cations,[17–21] brings the loops into play as a key structural feature for drug-quadruplex recognition.

We have focused on two well-characterised quadruplex forming sequences, namely the human telomeric repeat sequence (*h*-Tel) and the proto-oncogenic *c*-kit promoter sequence. Parallel and distinct anti-parallel (2 + 2 and 3 + 1) structures have now been reported for the *h*-Tel sequence, demonstrating conformational flexibility and sensitivity to monovalent cations (Na^+ and K^+) and to the nature of the terminal nucleotides.[22–30] In contrast, the *c*-kit promoter sequence forms a unique parallel architecture (Fig. 1a).[31] We have investigated by CD spectroscopy the interaction and selectivity of two quadruplex-binding ligands with *h*-Tel and *c*-kit, and demonstrate very different degrees of structural plasticity on the part of the quadruplex sequences, and binding selectivity on the part of the ligands which correlates with their unique structural characteristics (Fig. 1c). The first (**TRZ**) is a heteroaromatic 1,4-triazole assembled using the 'click' chemistry approach and has a novel extended scaffold designed to exploit π-stacking interactions and groove-specific contacts.[32,33] **TRZ** demonstrates potent telomerase inhibition in modified TRAP assays, while drug combination studies with *cis*-platin show interesting synergistic effects in MCF7 and A549 cell lines.[34] The polycyclic fluorinated acridinium cation (**RHPS4**) is similarly a potent telomerase inhibitor[35,36] which interacts largely through π-stacking interactions.[17,18] We demonstrate that two different small molecule architectures result in significant differences in selectivity for parallel and anti-parallel quadruplex structures.

Results

CD analysis of *h*-Tel and *c*-kit

Far UV CD spectra were collected on 4 μM samples of *h*-Tel and *c*-kit after annealing in the presence of 100 mM K^+ at pH 7.0 (Fig. 2a). The spectrum of *h*-Tel is dominated by a strong positive

[a]*Centre for Biomolecular Sciences, School of Chemistry, University Park, Nottingham, NG7 2RD, UK. E-mail: mark.searle@nottingham.ac.uk*
[b]*School of Chemistry, University Park, Nottingham, NG7 2RD, UK. E-mail: john.moses@nottingham.ac.uk*
[c]*School of Pharmacy, University Park, Nottingham, NG7 2RD, UK*

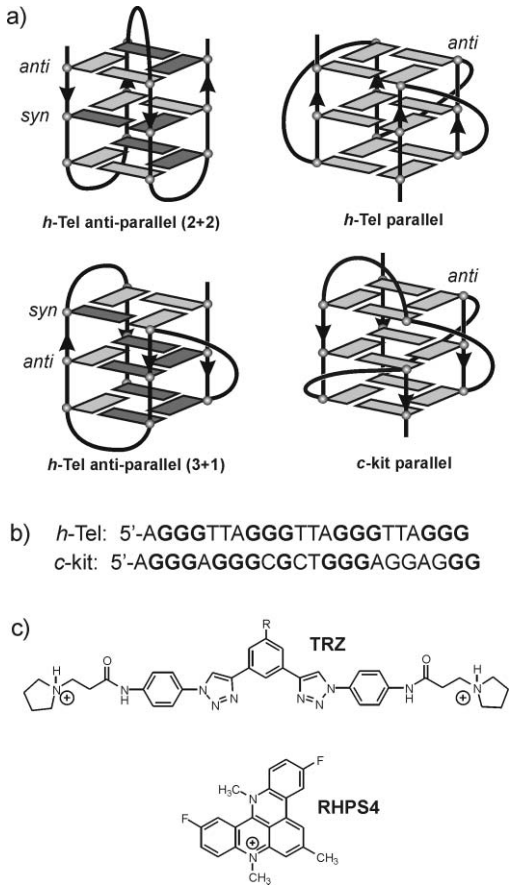

Fig. 1 (a) Schematic representation of the (2 + 2) and (3 + 1) anti-parallel and parallel quadruplex structures of *h*-Tel with relative *syn* and *anti* glycosidic conformations of the guanines. The unique parallel fold of the *c*-kit structure is also shown. (b) DNA sequences of *h*-Tel and c-kit used in these studies. Guanine residues shown in bold are those involved in formation of the G-tetrads. (c) Structures of **TRZ** and **RHPS4**: the former was synthesised as the free base and the latter as the methyl sulfate salt.

band at 290 nm, with a small minimum at 235 nm and a number of additional broad positive bands apparent at 250 nm and 275 nm which together suggest contributions from both the (2 + 2) and (3 + 1) anti-parallel folds (Fig. 1a) as previously discussed for closely related sequences in K^+ solution.[25,27] In contrast, the CD spectrum of *c*-kit gives a single maximum at 262 nm at the position expected for a parallel quadruplex structure (Fig. 2b),[37] in agreement with the structural model of *c*-kit determined by NMR.[31] There is no evidence for a significant population of any anti-parallel structure. The *h*-Tel and *c*-kit quadruplexes are of similar stability. The CD melting curve for *h*-Tel monitored at 290 nm gives a $T_m = 340$ K, while that of *c*-kit indicates a slightly lower stability ($T_m = 337$K; see Table 1).

Quadruplex–ligand interactions and binding selectivity

A superposition of the CD spectrum of *h*-Tel alone (4 μM in 100 mM KCl at 298 K) and in complex with 4 equivalents of **TRZ** and **RHPS4** reveals that the ligands perturb the conformational equilibrium between different quadruplex structures (Fig. 2a). **TRZ** has a significant effect on the relative population of parallel and anti-parallel structures, as judged from the intensity of the bands at 265 and 290 nm (Fig. 2a). We see an increase in the population (~60%) of the parallel conformer, with the emergence of a strong band at 262 nm, which suggests that this structure is thermodynamically preferred in the presence of an excess of ligand. Given that *h*-Tel alone favours anti-parallel structures, either (2 + 2) or (3 + 1) (Fig. 2a), ligand binding appears to show some selectivity for the parallel fold. We examined the near-UV CD spectrum up to 650 nm (data not shown) for possible evidence of an induced CD spectrum from the achiral ligand, as reported for groove-specific ligands such as the dicarbocyanine DODC and the bifuryl DB832.[38] We could detect no signal to support **TRZ** binding through a groove-only interaction, supporting a base stacking mode of binding at the end of the stack of G-tetrads. We characterised the binding interaction of **TRZ** with *h*-Tel using fluorescence, observing quenching of the ligand signal by *h*-Tel indicating strong binding ($K_A = 7.7 \pm 0.2 \times 10^5$ M^{-1} at 298 K).

The thermal unfolding of the parallel and anti-parallel structures formed by **TRZ** and *h*-Tel was monitored by CD at 265 and 290 nm, respectively, over the temperature range 293–368 K (Fig. 2c). We observe quite different melting profiles with different T_m values. The anti-parallel structure gives a much broader unfolding transition with a mid-point estimated to be 346 K and an enthalpy of unfolding at the T_m of 98 ± 1 kJ mol^{-1}. In contrast, the parallel structure, monitored at 265 nm, shows an initial increase in population with increasing temperature, as the less stable anti-parallel structure initially converts to the more stable parallel form. The latter has both a higher $T_m = 356$ K and a steeper transition ($\Delta H = 216 \pm 2$ kJ mol^{-1}). Consistent with the observation that the population of the parallel quadruplex structure of *h*-Tel increases in the presence of the ligand and is the thermodynamically preferred form, we also observe a 10 K difference in the transition mid-point of the complexes of the parallel *versus* anti-parallel folds. In both cases the T_m values show an overall stabilisation of *h*-Tel (see Table 1), with evidence for selectivity by **TRZ** for the parallel form.

Table 1 Thermodynamic stability data for *h*-Tel and *c*-kit quadruplexes with and without bound ligands

Quadruplex (Q)[a]	Q T_m (K)/(ΔH/kJ mol^{-1})[d]	Q + **TRZ** T_m (K)/(ΔH/kJ mol^{-1})[d]	$\Delta T_m(K)$**TRZ**	Q+**RHPS4** T_m (K)/(ΔH/kJ mol^{-1})[d]	$\Delta T_m(K)$**RHPS4**
h-Tel (290 nm) (*anti-parallel*)	340.4 (180 ± 2)	345.9 (98 ± 1)	5.4	361 (145 ± 10)	20.6
h-Tel (265 nm)[b] (*parallel*)	344.5 (228 ± 7)	355.6 (216 ± 2)	11.1	—	—
c-kit (265 nm)[c] (*parallel*)	337 (186 ± 2)	345 (130 ± 2)	8	358 (142 ± 17)	21

[a] All data were recorded on 4 μM samples of DNA in 100 mM KCl and 10 mM potassium phosphate buffer at pH 7.0. [b] The band at 265 nm is weak, but becomes significantly more populated in the presence of **TRZ** but is absent from the spectrum with **RHPS4**. [c] Only the band for the parallel quadruplex (at 265 nm) is detected. [d] Melting curves were fitted assuming a two-state model and a pseudo first-order transition with the ligand in excess; enthalpies were determined at the T_m.

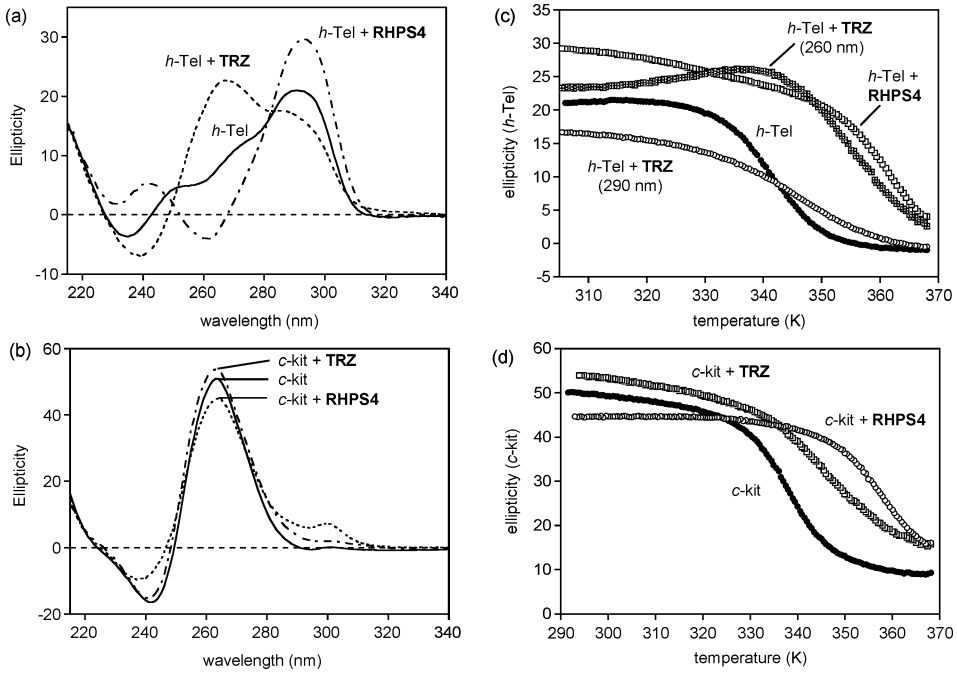

Fig. 2 (a) CD spectrum of *h*-Tel (4 μM DNA in 100 mM KCl and 10 mM potassium phosphate buffered to pH 7.0 at 298K) and the complexes with 4 equivalents of **TRZ** and **RHPS4**. The solid line representing *h*-Tel shows evidence for predominantly an anti-parallel fold (maximum at 290 nm) with possible minor conformers evident from bands at 255 and 275 nm. **TRZ** leads to a significant increase in the population of the parallel structure (positive band at 265 nm), while **RHPS4** pushes the equilibrium strongly in favour of the anti-parallel quadruplex (strong maximum at 290 nm). When overlayed, the three spectra show an isodichroic point at 282 nm. (b) Corresponding CD spectra for *c*-kit. The bound ligands produce only small perturbations to the absorption envelope. A small band at 300 nm in the spectrum with **RHPS4** is not readily assigned. (c) CD melting curves for the *h*-Tel and *c*-kit complexes with 4 equivalents of **TRZ** and **RHPS4**. Melting curves for *h*-Tel (black dots) and the *h*-Tel + **RHPS4** complex (open squares) monitored at 290 nm, and for the parallel and anti-parallel complexes with **TRZ** monitored at 260 nm and 290 nm, respectively. (d) Melting curves for *c*-kit under identical conditions, showing that both **TRZ** and **RHPS4** stabilise the unfolding transition despite small apparent structural changes in the CD spectra. Thermodynamic data is presented in Table 1.

Titrations of *h*-Tel with **RHPS4** under otherwise identical conditions produce quite different effects on the CD spectrum (Fig. 2a). The band at 290 nm strengthens considerably with the ellipticity becoming negative at 260 nm. The weaker positive bands observed for *h*-Tel at 250 nm and 265–270 nm in the absence of ligand are no longer apparent and the spectrum with **RHPS4** bound shows the characteristic features of a predominantly anti-parallel (2 + 2) basket-like structure reminiscent of that observed in Na^+ solution.[25,27] We monitored the melting transition at a single wavelength (290 nm) and observed a significant enhancement of quadruplex stability of 21 K compared to the unliganded *h*-Tel (Fig. 2c). In contrast to **TRZ**, the interaction of **RHPS4** shows selectivity for the anti-parallel (2 + 2) basket-like structure as well as significantly larger effects on quadruplex stability (Table 1), consistent with SPR and fluorescence binding studies[35,36] which show a telomere binding affinity ($K_A \sim 1 \times 10^7$ M^{-1}), which is an order of magnitude larger than for **TRZ**.

The CD analysis of the *c*-kit sequence provides evidence for far less structural plasticity in binding ligands than was evident for *h*-Tel. The **TRZ** ligand induced a small increase in intensity at 265 nm, however, overall the changes were not significant (Fig. 2b). The CD melting curves demonstrated an enhanced stability of 8 K over the unliganded structure (Fig. 2d), and fluorescence quenching experiments again show that **TRZ** binds with high affinity through base stacking interactions ($K_A = 1.2 \pm 0.13 \times 10^6$ M^{-1}). We repeated the experiments with **RHPS4** and observed similarly small changes in the spectrum (Fig. 2c). However, the large shift in the T_m value by 21 K (Fig. 2d), which is very similar to that observed with the anti-parallel structure of *h*-Tel, clearly demonstrated that **RHPS4** is binding strongly to the *c*-kit structure but without inducing changes in the population of conformers. The *c*-kit sequence forms a unique parallel quadruplex structure in solution, as determined by NMR.[31] An isolated guanine is recruited to form the G-tetrad core (Fig. 1a), despite the presence of four characteristic G_3 tracts which present the opportunity for formation of a number of possible alternative parallel or anti-parallel folds. However, these are not observed in solution by NMR, identifying the observed structure as uniquely low in energy. Moreover, the CD data indicate that there is little structural perturbation induced by ligand binding to the *c*-kit structure, and certainly no evidence for formation of anti-parallel structures as favoured by **RHPS4** in complex with *h*-Tel.

Ligand-induced quadruplex folding

The extent to which the apparent ligand selectivity for different quadruplex conformers is determined by the stabilising effects of the monovalent cation (K^+ in this instance) was investigated by titrating **TRZ** into 4 μM solutions of *h*-Tel and *c*-kit in the absence of K^+. Aqueous solutions of *h*-Tel and *c*-kit gave CD spectra devoid of substantial populations of quadruplexes (Fig. 3). The *h*-Tel sequence gives a positive ellipticity at 255 nm and a much weaker

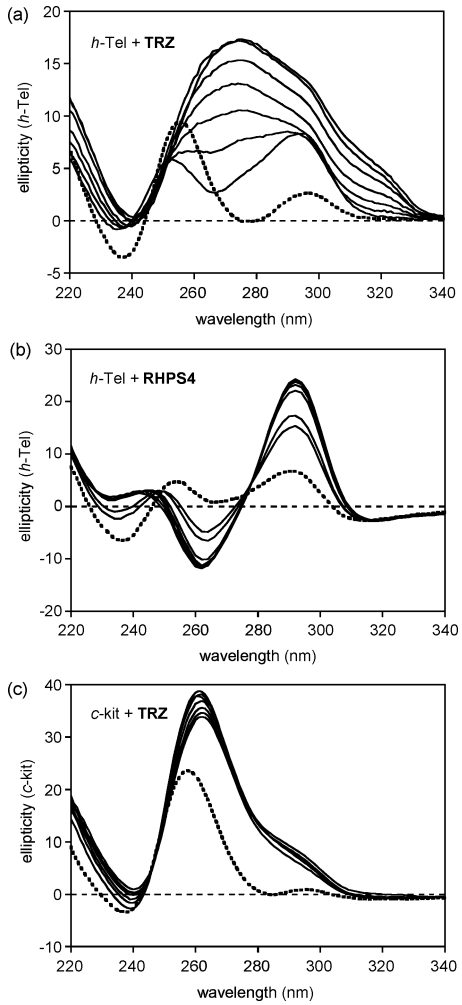

Fig. 3 Ligand-induced folding of *h*-Tel and *c*-kit in the absence of K^+ (4 μM DNA at 298K). (a) CD spectrum of *h*-Tel in the unfolded state (dotted line, maximum at 255 nm) and in the presence of up to 4 equivalents of **TRZ** resulting in an increase in ellipticity at 260 and 290 nm; (b) CD spectra showing anti-parallel structure (295 nm) induced in *h*-Tel by up to 4 equivalents of RHPS4 (dotted line showing *h*-Tel in the 'absence' of K^+ is taken from a different batch of material to that shown in (a); the slight difference in ellipticity at 295 nm suggests that some structure is already present in solution as a result of low levels of monovalent cations). (c) Similar experiment with *c*-kit (4 μM DNA, no K^+ at 298K), showing an increase in parallel quadruplex structure (CD maximum at 262 nm) in the presence of **TRZ** (up to 4 equivalents), although the structure is fully induced by one equivalent of ligand.

band at 295 nm, the former being characteristic of the unfolded state. Titration of *h*-Tel with **TRZ** initially leads to the appearance of a band at 290 nm with a shoulder at 275 nm indicative of the mixture of (2 + 2) and (3 + 1) anti-parallel structures observed for *h*-Tel in K^+ solution. However, beyond 1 equivalent of **TRZ** we observe a significant increase in intensity at 265–270 nm and at 290–295 nm indicative of both parallel and anti-parallel structures becoming highly populated. A much weaker absorbance at 320 nm is less readily assigned. The general features of the induced CD spectrum suggest populations of structures that are quantitatively similar to those observed upon addition of ligand to an annealed sample of *h*-Tel in 100 mM KCl. This also appears to be the case for **RHPS4** (Fig. 3b) which induces a strong positive band at 295 nm and negative band at 260 nm in the spectrum of *h*-Tel in the absence of significant amounts of K^+. The CD spectra of *h*-Tel bound to **RHPS4** either in the absence of K^+ or presence of 100 mM KCl both show the same strong preference for formation of anti-parallel quadruplex structure.

In the absence of K^+ ions and ligands, we observe a spectrum of the *c*-kit sequence with the unfolded state dominated by a band at 255 nm. In this case, addition of only 1 equivalent of **TRZ** induces a large increase in intensity at 260 nm characteristic of the parallel folded form of *c*-kit (Fig. 3c). Under these conditions a weaker band is induced at 290 nm which may indicate 10–15% of an anti-parallel form. However, the CD spectrum of the **TRZ** complex in the absence of K^+ is very similar to that observed for the annealed sample in 100 mM KCl. Thus, the structural preferences observed in the binding of **TRZ** to various parallel and anti-parallel conformations appear to be largely independent of whether the ligand–DNA complexes assemble from the pre-organised folded state in the presence of K^+, or are induced by the binding of **TRZ** to unfolded oligonucleotides. We assume that the native quadruplex structures are populated at very low levels in the absence of K^+ or ligands, and that the conformational equilibrium is pulled over to the bound state through a 'conformational selection' process, although the details of this mechanism remain to be established.

Ligand binding stoichiometry revealed by mass spectrometry

To characterise the ligand binding stoichiometry we used electrospray ionisation-mass spectrometry (ESI-MS)[39,40] in the negative ion-mode at various quadruplex to ligand ratios in ammonium acetate buffer and could readily detect non-covalent complexes of **TRZ** and **RHPS4** with *h*-Tel and *c*-kit. Fig. 4a shows representative spectra of *h*-Tel and *c*-kit (10 μM) in the presence of 20 μM **TRZ**. In both cases the quadruplex produced $[M - 5H]^{5-}$ ions as the major species, with only traces of quadruply charged ions. Both Q:L and $Q:L_2$ complexes were visible for *h*-Tel, indicating the occupancy of both available sites at the ends of the *h*-Tel structure (Fig. 4a), however, $Q:L_2$ was not detected for *c*-kit. We analysed the fraction of free and bound quadruplex species present at different ligand concentrations and analysed the data assuming that in each case there are two independent sites of similar affinity. In the case of *h*-Tel (open symbols in Fig. 4b and c) the fraction of each bound species present fits well to this simple model (represented by the solid black lines) for both **TRZ** and **RHPS4**. However, with *c*-kit we see some deviations from this model. In the case of **TRZ**, we failed to observe the $Q:L_2$ complex in the accessible concentration range but detected a higher fraction of the singly bound species (Fig. 4b), indicating that one site on *c*-kit binds **TRZ** preferentially with higher affinity. This is consistent with the observation by CD (Fig. 3b) that only one equivalent of ligand is required to fully induce structure in *c*-kit in the absence of K^+ ions, while several equivalents are required for *h*-Tel. In contrast, titration studies with **RHPS4** and *c*-kit reveal a higher proportion of the $Q:L_2$ complex than expected from the two independent binding sites model, indicative of some degree of co-operativity between the two sites with binding at one location enhancing the affinity at the second site.

Molecular dynamics simulations of quadruplex–ligand complexes

Previous investigations have suggested a preference in X-ray structures for ligands to bind to the parallel form of *h*-Tel,[32,33] however,

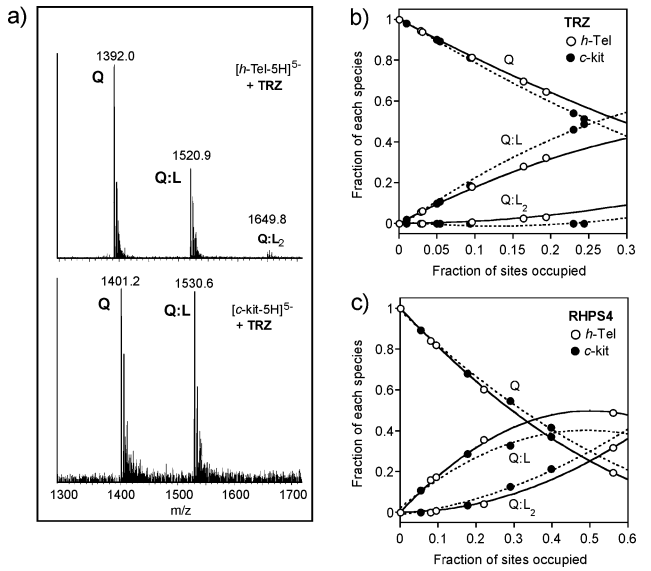

Fig. 4 ESI-MS analysis of the population of bound species in solution showing in (a) mass profiles of the *h*-Tel and *c*-kit quadruplexes (10 μM) in the presence of the **TRZ** ligand (20 μM) with the stoichiometry of the bound species indicated. The fraction of each species present in solution for (b) **TRZ** and (c) **RHPS4** in complex with *h*-Tel and *c*-kit *versus* the fraction of sites occupied is also shown. Solid black lines in each case represent populations derived from a simple 2:1 model involving two equivalent ligand binding sites of similar affinity. The fraction of each species present (open symbols for *h*-Tel complexes and black symbols for *c*-kit) were determined experimentally by MS as shown for **TRZ** (a) and **RHPS4** (b). The data for *h*-Tel in each case fit well to the simple 2:1 binding model, but for *c*-kit the $Q:L_2$ complex is not readily detected for **TRZ**, but for **RHPS4** is in higher abundance than predicted by the simple model. Dotted lines represent the polynomial fits to the experimental data to illustrate deviations from the simple model.

studies with telomestatin and diseleno sapphyrin demonstrate ligand-induced stabilisation of *h*-Tel in different conformational states.[41] CD analysis shows that anti-parallel quadruplexes predominate in aqueous buffer in 100 mM KCl in the absence of ligands.[25,27] **TRZ** perturbs the *h*-Tel equilibrium towards the parallel form and, when present in excess, parallel and anti-parallel complexes are both significantly populated. The ligand-induced shift in the equilibrium demonstrates some selectivity for the parallel structure which is otherwise not significantly populated in K^+ solution. In contrast, **RHPS4** strongly favours the (2 + 2) anti-parallel structure of *h*-Tel. To gain more insights into these conformational preferences we investigated the stability of various structures using solvent-explicit molecular dynamics simulations using AMBER8. Modelling of the complex of **TRZ** with the parallel quadruplex structure has already been reported and described a preference for stacking on the 3′-side of the structure, which leads to more favourable interactions with the TTA propeller loops forming the grooves.[32,33] Although the ESI-MS analysis of the complexes of *h*-Tel with **TRZ** (Fig. 4) suggests that the affinity of the two binding sites of $Q:L_2$ may be quite similar, we are unable to unambiguously assign this species to one or other quadruplex fold, indeed, a mixture of folds are evident from the CD analysis (Fig. 2a).

We focused our modelling on the *h*-Tel complexes with **RHPS4** where differences were apparent that would appear to favour the formation of the (2 + 2) basket-like anti-parallel structure observed in Na^+ solution, over the parallel stranded structure (see Fig. 1).[22] We constructed models (illustrated in Fig. 5a–c) and used a combination of energy minimisation, molecular dynamics simulations and energy calculations to assess relative ligand binding affinities.[33] MMPBSA calculations suggest that binding through π-stacking on the terminal G-tetrad at either end of the (2 + 2) anti-parallel structure of *h*-Tel, in proximity to the diagonal TTA loop or the two lateral TTA loops, leads to highly favourable hydrophobic interactions that stabilise the complexes. Analogous contacts are evident in our earlier NMR structure of **RHPS4** binding to the 5-ApG and 5′-GpT steps in the intermolecular quadruplex $d(TTAGGGT)_4$,[17,18] and are observed in the X-ray structure of BRACO-19 bound to the diagonal loop of the $d(G_4T_4G_4)_2$ quadruplex dimer.[20] Very recent studies highlight the flexibility of these loops in accommodating bound ligands.[19] In contrast, the interaction of **RHPS4** with the parallel quadruplex structure (shown in Fig. 5a) appears to be energetically less favourable due to the greater solvent exposure of the bound ligand and absence of hydrophobic contacts with the loops.

Discussion

The development of small molecule ligands to interact selectively with G-quadruplex structures formed within telomeric DNA sequences or gene promoter regions has great potential for the discovery of anti-cancer agents that regulate gene expression. The structural topology of the target quadruplex is dictated by the relative orientation (parallel or anti-parallel) of the four

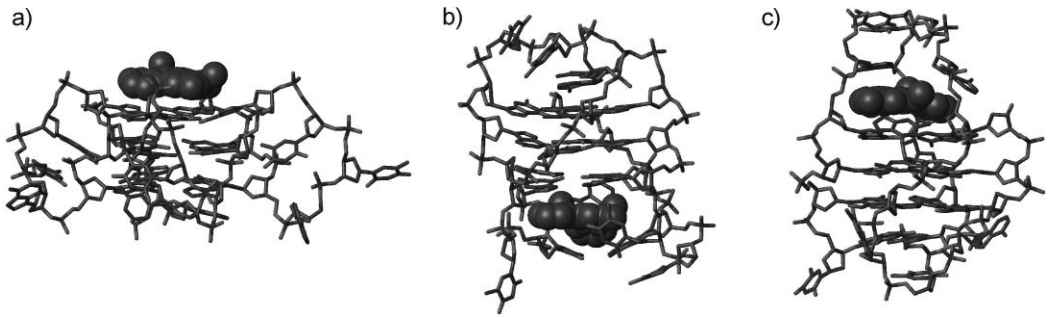

Fig. 5 Models of the quadruplex structures of *h*-Tel with **RHPS4**; high energy structure of the complex of the parallel form (a), and the low energy structures of the (2 + 2) anti-parallel form with **RHPS4** bound at either end, (b) interacting with the two lateral loops, and (c) interaction with the diagonal loop. Structures were refined from 3–5 ns of unrestrained molecular dynamics using an explicit solvent model.

strands, with the connecting loop sequences playing a pivotal role in dictating the thermodynamically preferred topology. As a consequence, the loops (diagonal or propeller) are brought into play as key structural features for drug recognition. We have demonstrated using CD spectroscopy and ESI-MS that recognition and stabilisation of specific G-quadruplex structures is possible using different small molecule scaffolds. Distinct structural differences between the parallel and anti-parallel folds of the human telomere (*h*-Tel) allow specific ligand contacts to 'switch' the conformational equilibrium, demonstrating both ligand binding selectivity and intrinsic plasticity of the *h*-Tel sequence. In contrast, the *c*-kit sequence populates a unique low energy parallel form in which 18 of the 22 nucleotides are involved in tertiary interactions (Fig. 1a). This is despite the fact that conformational heterogeneity could similarly arise from alternative parallel and anti-parallel folds with various loop lengths of between 1–4 nucleotides (Fig. 1b), but that these forms are energetically inaccessible even in the presence of the bound ligands. Thus, both **RHPS4** and **TRZ** bind to the unique fold of *c*-kit and further stabilise the structure but without perturbing the fold, masking the ligand binding selectivity evident in the case of *h*-Tel.

Both parallel and anti-parallel quadruplexes possess two potential ligand binding sites, one at either end of each structure. We demonstrate by ESI-MS that two drug molecules can bind to each quadruplex and, on the basis of the relative populations of the Q:L and Q:L_2 species present at different ligand concentrations, show that the two sites on the *h*-Tel structures have similar affinities. However, this is not the case for the highly asymmetric *c*-kit structure where different affinities for the two sites are apparent. Although structure-specific recognition of the grooves of quadruplexes has recently emerged as an attractive strategy for drug discovery,[38] the ends of quadruplexes, in particular the influence of the connecting loops, still remain to be fully exploited in nucleic acid recognition. Novel quadruplex folds with mixed parallel/anti-parallel strand alignments and novel loop sequences present attractive and potentially unique targets for drug design and the modulation of gene expression.[42]

Experimental

Oligonucleotides

The DNA oligonucleotides were purchased from Eurogentec SA (Seraing, Belgium) and synthesised on a 10 μmol scale, HPLC purified, desalted and supplied lyophilised. The DNA samples (4–6 μM) were prepared in a buffer comprising of potassium chloride (100 mM) and potassium phosphate (10 mM, monohydrogen and dihydrogen powders mixed to give pH 7) in Milli-Q water (>18 MΩ.cm). All samples were then heated up to 90 °C, held at this temperature for 10 minutes and then allowed to slowly cool to 20 °C over a period of 12 hours. To assess the concentration dependence, additional samples were prepared in the range 2.5–20 μM as described above.

CD spectroscopy

The CD experiments were recorded on a PiStar-180 spectrophotometer (Applied Photophysics Ltd, Surrey, UK) with temperature regulated using a RTE-300 circulating programmable water bath (Neslab Inc, Portsmouth, NH, USA) and a thermoelectric temperature controller (Melcor, Trenton, New Jersey, USA). Spectra were recorded using a 1 cm path length quartz cuvette with data points recorded at every 1 nm step with entrance/exit slit widths of 2 nm. Following the subtraction of the buffer baseline, the data were normalized to zero at 325 nm.

UV absorbance and melting studies

UV experiments were recorded on the PiStar-180 spectrophotometer or on a Cary 100 Bio UV-visible spectrophotometer on samples used for CD studies over the range 5 °C and 95 °C and wavelength range 325–215 nm. Samples were equilibrated at 5 °C for 20 minutes before being heated to 90 °C and then cooled back down to 5 °C at a rate of 0.2 °C/min, to obtain both the melting and annealing curves and at different sample concentrations in the range 2.5–20 μM. The T_m values were obtained from a van't Hoff analysis of the melting curve. The enthalpy and entropy changes were calculated assuming a two-state equilibrium between a folded and unfolded conformation using Kaleidagraph™ (version 3.5, Synergy Software, Reading, USA).

ESI-mass spectrometry

Electrospray ionisation-mass spectrometry was performed on a ThermoFisher LTQFT Ultra hybrid ion trap-Fourier transform ion cyclotron resonance mass spectrometry in the negative ion mode. Samples (10 μM DNA and various conc. of ligand from 2.5–20 μM) were sprayed from ammonium acetate buffer (25 mM, pH 7) at a rate of 5 μl min^{-1} using an electrospray capillary voltage of –4 kV, with a nebuliser gas flow rate of 15 (arbitrary units). The instrument was tuned for optimum signal intensity for the $[M - 5H]^{5-}$ ion of *h*-Tel and the tube lens adjusted down to –80 V to maintain the interaction between DNA and click ligand. Measurements were acquired in the FT mode at a nominal resolution of 100,000 (at m/z 400) from m/z 200–2000 and averaged over a 2 min period. Data were acquired and analysed using Thermo Xcaliber software.

Molecular modelling studies

All molecular modelling calculations were performed using a 12 CPU Opteron Clusters running RedHat Linux with MPI support for the sander module of AMBER8 or on workstations using 64-bit Fedora. The RHPS4 ligand was parameterised using the xLEaP and Antechamber modules of AMBER8 employing the General Amber Force Field (GAFF). Electrostatics were calculated using the bcc charge fitting function and no additional force constants were required to complete the parameterisation. RHPS4 was then neutralised using a chloride ion and subjected to an equilibration using the SANDER module of AMBER8. In total 1 ns of molecular dynamics calculations using explicit solvent (TIP3P) with truncated octahedron periodic boundary conditions were performed. Electrostatics were calculated using the particle mesh Ewald approach (PME). Final minimised geometries compared well to previous geometry optimisations using *ab initio* calculations.

The anti-parallel structure (PDB: 143D) and the parallel structure (PDB: 1KF1) were accessed *via* the PDB. All structures

were then parameterised within xLEaP using the AMBER99 forcefield. Explicit solvent modelling was carried out using MD simulations of 3–5 ns. RMSD values suggest that all structures reach a stable state by the first 500ps of modelling. RHPS4-quadruplex structures were initially modelled by positioning the ligand in contact with the external face of the G-tetrads guided by previous experimental evidence.[18] In the case of steric clashes with the loop regions of the anti-parallel sequences these loops were removed and reattached using long 7 Å bonds to create a binding pocket. Energy minimisation using a 100 kcal mol^{-1} coordinate restraining energy, followed by standard equilibration protocols, resulted in structures with good starting geometry. All structures were then subjected to 3–5 ns of MD simulation following the procedures detailed above.

MM PBSA methods allow calculation of a binding free energy from molecular modelling simulations. Models are taken from ligand, receptor and complex and energies calculated and summed to give approximated binding free energies. All structures were sampled from the last 2 ns of the molecular modelling runs. Energies are calculated for solvation/desolvation using the PBSA continuum model. This calculation assigns solute electrostatics using the AMBER partial charges and a dielectric constant of 1. Bulk solvent is then defined at the contact surface with a dielectric of 80. The change in energy for hydrophobic groups is also calculated based on a SASA dependant term. Entropy contributions are calculated from a normal mode analysis and gas phase interaction energies taken from minimised structures. The MM PBSA software was modified to include radii information for fluorine atoms.

Acknowledgements

This work was supported by the EPSRC of the UK, The Association for International Cancer Research, Cancer Research UK and the University of Nottingham.

References

1 J. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2005, **33**, 2908–2916.
2 A. Todd, M. Johnston and S. Neidle, S., *Nucleic Acids Res.*, 2005, **33**, 2901–2907.
3 J. Huppert, *Chem. Soc. Rev.*, 2008, **37**, 1375–1384.
4 T. Fletcher, D. Sun, M. Salazar and L. H. Hurley, *Biochemistry*, 1998, **37**, 5536–5541.
5 J. Mergny, J. Riou, P. Mailliet, M. Teulade-Fichou and E. Gilson, *Nucleic Acids Res.*, 2002, **30**, 839–865.
6 R. De Armond, S. Wood, D. Sun, L. H. Hurley and S. W. Ebinghaus, *Biochemistry*, 2005, **44**, 16341–16350.
7 S. Kumari, A. Bugaut and S. Balasubramanian, *Biochemistry*, 2008, **47**, 12664–12669.
8 S. Neidle and G. Parkinson, *Nat. Rev. Drug Discovery*, 2002, **1**, 383–393.
9 M. Bejugam, S. Sewitz, P. S. Shirude, R. Rodriguez, R. Shahid and S. Balasubramanian, *J. Am. Chem. Soc.*, 2007, **129**, 12926–12927.
10 C. Grand, H. Han, R. Munoz and L. H. Hurley, *Molecular Cancer Therapeutics*, 2002, **1**, 565–573.
11 A. Siddiqui-Jain, C. L. Grand, D. J. Bears and L. H. Hurley, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 11593–11598.
12 M. S. Searle, G. D. Balkwill, in *Quadruplex Nucleic Acids* (Eds. S. Neidle, S. Balasubramanian) Royal Society of Chemistry Publishing, Cambridge, UK, 2007, pp 131–153.
13 A. De Cian, L. Lacroix, C. Douarre, N. Temime-Smaali, C. Trentesaux, J.-F. Rioux and J.-L. Mergny, *Biochimie*, 2008, **90**, 131–155.
14 A. T. Phan, V. Kuryavyi, K. Ngoc, D. J. Patel, in *Quadruplex Nucleic Acids* (Eds. S. Neidle, S. Balasubramanian) Royal Society of Chemistry Publishing, Cambridge, UK, 2007, pp 81–99.
15 P. A. Rachwal, I. S. Findlow, J. M. Werner, T. Brown and K. R. Fox, *Nucleic Acids Res.*, 2007, **35**, 4214–4222.
16 P. Hazel, J. Huppert, S. Balasubramanian and S. Neidle, *J. Am. Chem. Soc.*, 2004, **126**, 16405–16415.
17 E. Gavathiotis, R. A. Heald, M. F. G. Stevens and M. S. Searle, *Angew. Chem.*, 2001, **40**, 4749–4751.
18 E. Gavathiotis, R. A. Heald, M. F. G. Stevens and M. S. Searle, *J. Mol. Biol.*, 2003, **334**, 25–36.
19 G. N. Parkinson, F. Cuenca and S. Neidle, *J. Mol. Biol.*, 2008, **381**, 1145–1156.
20 S. M. Haider, G. N. Parkinson and S. Neidle, *J. Mol. Biol.*, 2003, **326**, 117–125.
21 A. T. Phan, V. Kuryavyi, H. Y. Gaw and D. J. Patel, *Nat. Chem. Biol.*, 2005, **1**, 167–173.
22 Y. Wang and D. J. Patel, *Structure*, 1993, **1**, 263–282.
23 A. Phan and D. J. Patel, D., *J. Am. Chem. Soc.*, 2003, **125**, 15021–15027.
24 K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix and D. J. Patel, *J. Am. Chem. Soc.*, 2003, **128**, 9963–9970.
25 A. T. Phan, K. N. Luu and D. J. Patel, *Nucleic Acids Res.*, 2006, **19**, 5715–5719.
26 A. Ambrus, D. Chen, J. Dai, B. Bialis, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2006, **9**, 2723–2735.
27 A. T. Phan, V. Kuryavyi, K. N. Luu and D. J. Patel, *Nucleic Acids Res.*, 2007, **19**, 6517–6525.
28 Y. Xue, Z.-Y. Kan, Q. Wang, Y. Yao, J. Liu, Y.-H. Hao and Z. Tan, *J. Am. Chem. Soc.*, 2007, **129**, 11185–11191.
29 K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu and A. T. Phan, *J. Am. Chem. Soc.*, 2009, **131**, 4301–4309.
30 G. N. Parkinson, M. Lee and S. Neidle, *Nature*, 2002, **417**, 876–880.
31 A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle and D. J. Patel, *J. Am. Chem. Soc.*, 2007, **129**, 4386–4392.
32 A. D. Moorhouse, A. M. Santos, M. Gunaratnam, M. Moore, S. Neidle and J. E. Moses, *J. Am. Chem. Soc.*, 2006, **128**, 15972–15973.
33 A. D. Moorehouse, S. Haider, M. Gunaratnam, D. Munnur, S. Neidle and J. E. Moses, *Mol. BioSyst.*, 2008, **4**, 629–642.
34 M. Gunaratnam, C. Green, J. B. Moreira, A. D. Moorehouse, L. R. Kelland, J. E. Moses and S. Neidle, *Biochem. Pharmacol.*, 2009, **78**, 115–122.
35 S. M. Gowan, R. Heald, M. F. G. Stevens and L. R. Kelland, *Mol. Pharmacol.*, 2001, **60**, 981–988.
36 R. Heald, C. Modi, J. C. Cookson, I. Hutchinson, C. A. Laughton, S. M. Gowan, L. R. Kell and M. F. G. Stevens, *J. Med. Chem.*, 2002, **45**, 590–597.
37 H. Fernando, A. P. Reszka, J. Huppert, S. Ladame, S. Rankin, A. R. Venkitaraman, S. Neidle and S. Balasubramanian, *Biochemistry*, 2006, **45**, 7854–7860.
38 E. W. White, F. Tanious, M. A. Ismail, A. P. Reszka, S. Neidle, D. W. Boykin and W. D. Wilson, *Biophys. Chem.*, 2007, **126**, 140–153.
39 W. M. David, J. Brodbelt, S. M. Kerwin and P. W. Thomas, *Anal. Chem.*, 2002, **74**, 2029–2033.
40 F. Rosu, E. De Pauw, L. Guittat, P. Alberti, L. Lacroix, P. Maillet, J. F. Riou and J. L. Mergny, *Biochemistry*, 2003, **42**, 10361–10371.
41 E. M. Rezler, J. Seenisamy, S. Bashyam, M.-Y. Kim, E. White, W. D. Wilson and L. H. Hurley, *J. Am. Chem. Soc.*, 2005, **127**, 9439–9447.
42 J. Dai, T. S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R. A. Jones and D. Yang, *J. Am. Chem. Soc.*, 2006, **128**, 1096–1098.

