ARTICLE

OBC
www.rsc.org/obc

# Tetrapeptides induce selective recognition for G-quadruplexes when conjugated to a DNA-binding platform

**Sylvain Ladame,[a] James A. Schouten,[a] John Stuart,[b] Jose Roldan,[b] Stephen Neidle[b] and Shankar Balasubramanian*[a]**

[a] *University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge, UK CB2 1EW. E-mail: sb10031@cam.ac.uk; Fax: +44 (0)1223-336913; Tel: +44 (0)1223-336347*

[b] *Cancer Research UK Biomolecular Structure Group, The School of Pharmacy, University of London, 29-39 Brunswick Square, London, UK WC1N 1AX*

***Received 28th June 2004, Accepted 1st September 2004***
***First published as an Advance Article on the web 20th September 2004***

3,6-Bis-peptide acridine and acridone conjugates have been designed and synthesised to selectively interact with G-quadruplex DNA. The ligand properties are peptide sequence dependant, the highest discrimination being obtained with the FRHR tetrapeptide (up to >50-fold specificity). Molecular modeling studies have helped us rationalise the data and suggest that human telomeric quadruplex DNA can readily accommodate tetrapeptides, and furthermore that FRHR contributes to stabilization of the complex by non-bonded interactions within the TTA loop pockets of the quadruplex. These studies indicate that targeting distinct features of a G-quadruplex with hybrid molecules is a promising strategy for discriminating between quadruplex and duplex DNA.

## Introduction

G-quadruplexes are four stranded DNA structures formed from stacked tetrads of hydrogen-bonded guanine bases. These stable, non-classical, secondary structures are formed *in vitro* at physiological salt concentration and pH, from DNA sequences containing stretches of adjacent guanine residues.[1] Quadruplexes can be intermolecular or intramolecular in nature and can exhibit structural polymorphism based on variability in both strand polarity and loop geometry. However, all quadruplexes share some fundamental differences in structure from classical double-stranded B-DNA helices, thus providing potential for selective recognition by ligands that can discriminate between duplex and quadruplex DNA. In general, quadruplexes contain at least two tetrads that are held together *via* π–π interactions. They also contain four sugar–phosphate grooves, together with specific loops surrounding the compact guanine-based structure.[2]

It has been established that the formation of G-quadruplexes is functionally relevant to a number of biological mechanisms.[3] In particular, stabilization of intramolecular telomeric DNA in a G-quadruplex conformation has been shown to inhibit its elongation by the enzyme telomerase.[4] As telomerase is selectively expressed in most tumour cells at substantially higher levels than in normal somatic cells, the enzyme is a potentially attractive target for selective anti-cancer therapy and drug development.[5] Consequently, chemists have taken up the challenge of synthesizing ligands that stabilize the quadruplex-folded conformation. Several classes of quadruplex-stabilizing ligands have been reported that selectively bind to and stabilize quadruplexes and thus inhibit telomerase activity.[6–7] In addition, an intramolecular-quadruplex forming sequence has also been identified in the promoter region of the *c-myc* oncogene, and stabilization of this quadruplex structure has been shown to down-regulate transcription of the gene.[8] It is a possibility that such a regulatory mechanism might also apply to other genes, especially given that sequences with potential to form a G-quadruplex have already been identified in the promoter regions of a number of other genes.[9] A deeper understanding of quadruplex recognition by synthetic molecules would thus provide a platform of knowledge that could contribute towards the targeting of a number of other functionally important quadruplexes in the future.

The quadruplex-stabilizing ligands reported include synthetic molecules[6–7] and also engineered proteins.[10] Most of the small molecule ligands bear common structural features such as a planar aromatic chromophore and a positive charge, somewhat similar to classic B-DNA intercalators. In particular, polycyclic aromatic chromophores (*e.g.* anthraquinones, acridines, perylenes, porphyrins) have been extensively studied for their ability to stabilize quadruplexes *via* strong π–π interactions with the top guanine tetrad. However, most of these flat heterocyclic derivatives have low specificity for quadruplex over duplex DNA. One approach towards the design of more potent and selective agents, is to append side-chains to these flat aromatic cores in order to provide the opportunity for additional interactions with the loops and grooves of quadruplex DNA. This approach has been shown to induce discrimination between quadruplexes and B-DNA from structure–activity studies on an acridine scaffold with two, or more preferably three, substituents at positions 3, 6 and 9.[11]

We report here the combination of selected peptides to classic DNA binding scaffolds that by themselves are known to bind to G-quadruplexes, but with very poor specificity. First generation 3,6-disubstituted acridines have been shown to bind both telomeric quadruplex and B-DNA with micromolar affinity. One example, the 3,6-bis-[3-pyrrolidinopropionamide] acridine (**BSU6039**, Fig. 1), has been recently co-crystallized with the dimeric intermolecular G-quadruplex formed from the Oxytricha nova telomeric DNA sequence d(GGGGTTTTGGGG).[12] This structure clearly demonstrated the key stacking of the acridine core on the terminal G-tetrad and also the participation of the amide groups at positions 3 and 6 in hydrogen bonds with loop residues. However, it also showed that the aminopropionamide side chains were too short to engage in significant groove interactions. This structure suggested that appending longer appropriate side chains to the 3-aminopropionamide linker could further enhance the specificity of these acridine derivatives by simultaneously targeting the top G-tetrad and two grooves of the quadruplex. Based on these results, we have generated a series of symmetrical 3,6-bispeptide–acridine (or acridone) conjugates (compounds **3**, **4** Fig. 2). The rationale behind the molecular design is to exploit the π–π interactions between the heterocyclic core and the G-tetrad to orientate both peptides into two different grooves of the quadruplex akin to the two pyrrolidine rings of compound

DOI: 10.1039/b409698c

**Table 1** Dissociation constants (*Kd*) determined by SPR[a]

| Compound (peptide) | *Kd*/μM *Htelo* | *Kd*/μM *ds DNA* | *Kd*(*ds*)/*Kd*(*Htelo*) | *n* (drug/*Htelo*) |
|---|---|---|---|---|
| **1** | 9.51 ± 0.82 | 25.0 ± 3.20 | 2.6 | 0.9 |
| **3a** (RKKV) | 1.96 ± 0.10 | 21.0 ± 0.85 | 10.7 | 1.5 |
| **3b** (KRSR) | 0.81 ± 0.08 | 9.80 ± 0.43 | 12.1 | 1.2 |
| **3c** (FRHR) | 8.56 ± 0.15 | >390[b] | >46[b] | 1.1 |
| **2** | 3.60 ± 0.70 | 48.8 ± 2.3 | 13.6 | 1.2 |
| **4a** (RKKV) | 0.44 ± 0.10 | 3.70 ± 0.20 | 8.4 | 1.4 |
| **4b** (KRSR) | 0.78 ± 0.16 | 14.5 ± 0.75 | 18.6 | 1.0 |
| **4c** (FRHR) | 0.91 ± 0.07 | >48[b] | >53[b] | 1.0 |

[a] Experiments were carried out in 50 mM potassium phosphate, pH = 7.4 containing 0.1 M KCl. Errors are based on the standard deviation from three independent measurements. [b] Data estimated by calculating the minimum Kd value that would give a response of at least 2 RU (detection limit of the BIAcore) at the highest concentration of **3c** and **4c** for which zero binding was detected (25 μM for **3c** and 3 μM for **4c**).

**BSU6039**. The linker length has been carefully designed, from the X-ray co-crystal structure obtained with **BSU6039**, such that the peptide side-chains are orientated to have potential for selective interactions with grooves and loops of the quadruplex. The tetrapeptides employed were selected from a combinatorial library[7] and the corresponding symmetrical 3,6-bispeptide acridine or acridone conjugates were synthesized and studied. A number of the peptide conjugates showed enhanced specificity for G-quadruplexes as compared to their peptide-free analogues. In particular, the FRHR tetrapeptide induced significant discrimination for quadruplex over duplex DNA when conjugated to an acridine scaffold. The selectivity requirements for these amino acids have been explored and a structural hypothesis has been provided with the assistance of molecular modelling.

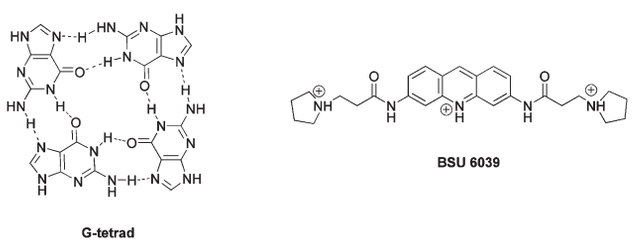


**Fig. 1** A G-tetrad (left) and the previously reported BSU 6039 bis-substituted acridine (right) are represented.

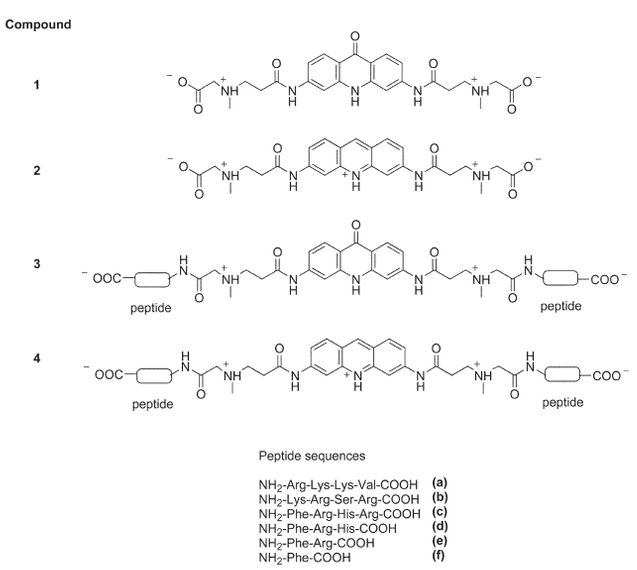


**Fig. 2** Chemical structures of acridine and acridone peptide conjugates and the peptide sequences are shown.

## Results and discussion

### Acridine/acridone peptide conjugate synthesis

Three tetrapeptides, FRHR, KRSR and RKKV were combinatorially selected from >25 000 possible sequences, as previously described,[7] based on their ability to selectively recognize G-quadruplexes as compared to duplexes (the quadruplex specificity being 5-fold for FRHR and 2-fold for KRSR and RKKV).

3,6-Bis-peptide acridine/acridone conjugates (Fig. 2) were synthesized on-bead *via* a bis-amide coupling between a 3,6-bis-sarcosine acridone/acridine and two peptides immobilized on resin based on a slight modification of our published protocol.[13] 3,6-Bis-tetrapeptide acridine/acridone conjugates (compounds **3a**–**c** and **4a**–**c**) were all obtained in high yields (>80%) and high purity (>95%) after HPLC purification. Similar yields were obtained for the bis-coupling of shorter peptides (tripeptides **3**–**4d** and dipeptides **3**–**4e**). Interestingly, no heterocycle peptide conjugate corresponding to the coupling of only one peptide onto the acridine or acridone was ever detectable by HPLC during all these syntheses, the main products being the expected bis-peptide acridone/acridine and traces of unreacted peptide.

### BIAcore surface plasmon resonance experiments

In order to investigate the binding affinity of the symmetrical 3,6-bispeptide acridine/acridone conjugates for their quadruplex target and also their level of specificity for quadruplex *versus* duplex DNA, we employed surface plasmon resonance (SPR). SPR has proven to be a powerful and valuable technology for the monitoring of DNA-small molecule interactions in real time.[14] We have previously used SPR to study the binding properties of a series of hemicyanine peptide conjugates to double-stranded and quadruplex DNA.[7] Others have used the same technology for investigating the binding of acridine derivatives to similar targets.[11] In the present study, the two DNA targets of interest were the human telomeric G-quadruplex (*Htelo*) formed from the oligonucleotide $(GTTAGG)_5$ and a double-stranded control DNA (*ds-DNA*) comprising the oligonucleotide $(GTTAGG)_5$ hybridized with its complementary sequence. A single-stranded control oligonucleotide ($(AGTTAG)_5$ *Atelo*) derived from the telomeric quadruplex sequence, but carrying an important G to A mutation that prevents quadruplex formation, was also used as a control. Each DNA target was loaded in a separate flow cell on the same sensor chip, keeping one lane DNA-free to detect possible non-specific interactions of the ligands with the dextran matrix. SPR was employed to determine both dissociation constants and stoichiometry of binding. For Kd determination, the response at equilibrium (Req) was plotted against the concentration of analyte, thus generating a hyperbolic binding curve which was fitted using a steady state algorithm (Fig. 3, SPR data for compound **4b**). Since the RU values are directly proportional to the amount of ligand bound to the immobilized DNA target,[15] the number of drug molecules bound to the DNA was also determined. The SPR results for the 3,6-bistetrapeptide acridone/acridine conjugates are summarized in Table 1.

All peptide conjugates were found to bind tighter to the human telomeric quadruplex *Htelo* than their peptide-free analogues **1** and **2** (up to 12-fold for **3c** *versus* **1**). This suggests that peptide side-chains are probably involved in stabilizing interactions with the quadruplex structure. This enhancement in affinity was sequence-dependent and was more significant with RKKV and KRSR tetrapeptides than with FRHR, probably

**Table 2** Dissociation constants (*Kd*) determined by SPR[a]

| Compound (peptide) | *Kd*/μM *Htelo* | *Kd*/μM *ds DNA* | *Kd*(*ds*)/*Kd*(*Htelo*) | *n* (drug/*Htelo*) |
|---|---|---|---|---|
| **3c** (FRHR) | 8.56 ± 0.15 | >390[b] | >46[b] | 1.1 |
| **3d** (FRH) | 1.64 ± 0.04 | 18.3 ± 0.1 | 11.1 | 1.2 |
| **3e** (FR) | 4.06 ± 0.23 | 21.7 ± 3.9 | 5.3 | 0.8 |
| **3f** (F) | --[c] | --[c] | --[c] | --[c] |
| **4c** (FRHR) | 0.91 ± 0.07 | >48[b] | >53[b] | 1.0 |
| **4d** (FRH) | 5.24 ± 0.6 | >78[b] | >15[b] | 1.4 |
| **4e** (FR) | 3.26 ± 0.24 | 55.5 ± 3.2 | 17.0 | 1.4 |
| **4f** (F) | 4.38 ± 0.22 | 36.2 ± 1.0 | 8.3 | 1.0 |

[a] Experiments were carried out in 50 mM potassium phosphate, pH = 7.4 containing 0.1 M KCl. Errors are based on the standard deviation from three independent measurements. [b] Data estimated by calculating the minimum *Kd* value that would give a response of at least 2 RU (detection limit of the BIAcore) at the highest concentration of **3c**, **4c** and **4d** for which zero binding was detected (25 μM for **3c**, 6 μM for **4d** and 3 μM for **4c**). [c] DNA binding of **3f** could not be evaluated owing to its poor solubility in water.

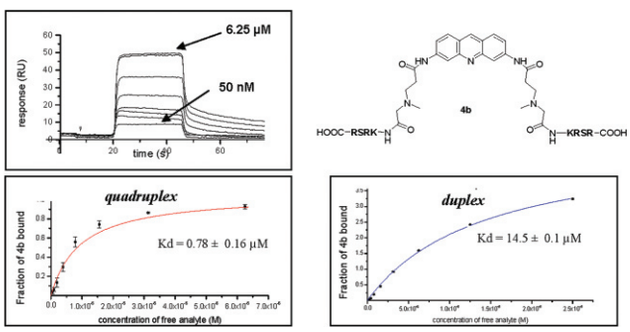


**Fig. 3** SPR sensorgram overlay for binding of compound **4b** to quadruplex target Htelo. Concentration range: 0.05, 0.10, 0.20, 0.39, 0.78, 1.56, 3.12, 6.25 μM from the bottom curve upwards (top left); Binding plots (fraction bound versus Cfree) used to determine the *Kd* value for **4b** with quadruplex (bottom left) and duplex (bottom right) DNAs.

as a result of the greater number of positively charged residues forming electrostatic interactions with the phosphate backbone of the TTA loops. There is also a general trend for acridine derivatives to bind slightly more tightly than the acridone analogues.

**Specificity for quadruplex over duplex DNA.** Some common features can be inferred for both series of acridine and acridone peptide conjugates. The symmetrical introduction of two peptides onto the 3,6-bis-sarcosine acridine/acridone **1** and **2** induced, for most tetrapeptide sequences, a significant improvement in the specificity for quadruplex over duplex DNA. Whereas compound **1** showed a weak quadruplex selectivity (Kd(*ds*)/Kd(*Htelo*) = 2.6), a discrimination of at least 10-fold for the G-quadruplex structure was obtained with all three 3,6-bispeptide acridones (**3a**–**c**). A similar but slightly weaker effect was observed with the three acridine analogues (**4a**–**c**), quadruplex selectivity increasing from 13.5 for the bis-sarcosine derivative **2** to >48 for the peptide conjugate **4c**. For both acridine and acridone series, the quadruplex selectivity was shown to be sequence dependent with >50-fold selectivity for the bis-FRHR acridine derivative **4c**. The increase in quadruplex specificity may be caused by peptide–quadruplex stabilizing interactions, but may also be due to peptide-induced destabilization of acridine–duplex DNA interactions. The latter effect is particularly noticeable for FRHR, since the FRHR acridone conjugate **3c** binds double-stranded DNA at least 15-fold weaker than its bis-sarcosine analogue **1**, showing that quadruplexes do accommodate FRHR more easily that DNA duplexes. The ability for FRHR, in particular, to enhance quadruplex recognition is consistent with our previous studies on hemicyanine–peptide conjugates,[7] where FRHR was also shown to induce high quadruplex *vs*. duplex discrimination. Whilst all heterocycle peptide conjugates bind *ds-DNA* with high stoichiometry (>3 : 1, data not shown) indicating the presence of several intercalating sites, they bind the telomeric G-quadruplex with a stoichiometry of near 1 : 1, suggestive of a much more specific binding mode. For duplex binding, although the stoichiometry was found to be higher than 3 : 1, the results fitted well with a model that assumes multiple sites of equal binding energy, suggesting that all duplex binding sites have comparable affinities. For both series of compounds, FRHR appeared to be the most promising tetrapeptide in terms of specificity for G-quadruplexes. Indeed, although FRHR conjugates **3c** and **4c** were the weakest quadruplex binders, they did not show any detectable binding to the duplex control at a concentration of 25 μM and 3 μM respectively.[16] In such cases, lower limits for Kd have been estimated (see Experimental section for details). Based on these estimates, it appeared that FRHR acridine and acridone conjugates show the most significant quadruplex selectivity (*ca.* 50 fold), slightly higher than that reported for the tri-substituted acridine **BRACO19**.[11]

**Influence of peptide length on quadruplex selectivity.** To investigate the capacity for shorter peptides to generate quadruplex specificity, the C-terminal amino acid of FRHR was systematically deleted (on both peptides at positions 3 and 6), generating six further acridine- and acridone-peptide conjugates with either FRH, FR or F peptide sequences. These compounds were evaluated by SPR to establish binding affinities for *Htelo* and *ds-DNA* (Table 2).

For the acridone conjugates, peptide shortening actually resulted in a slight enhancement in affinity. This effect may be due to the greater flexibility of the shorter peptide chain in these conjugates to maximize attractive interactions within the loops. However, this was accompanied by a progressive loss of specificity, with *Kd*(*ds*)/*Kd*(*Htelo*) decreasing from 46 to 13.0 and then to 5.0 for FRHR, FRH and FR conjugates respectively. This demonstrates that, of the molecules studied, tetrapeptide derivatives generate the most specific recognition of G-quadruplexes with respect to *ds-DNA* although the last two residues do not appear to contribute significantly to quadruplex binding.

Within the acridine series, all truncated peptide conjugates **4d**–**f** bind to *Htelo* with comparable affinity (*Kd* ≈ 3–4 μM), which is about 4 times weaker than for FRHR derivative **4c**. More interestingly, these three compounds have similar affinities for the G-quadruplex to that of the bis-sarcosine analogue **2**. This suggests that FRH, FR and F do not directly add significant binding energy to the bis-sarcosine acridine **2**. However, although FRH, FR and F acridine conjugates have similar affinities for the G-quadruplex, the specificity for *Htelo versus ds-DNA* decreases when the length of the peptide is reduced. This suggests that all four residues play an important role in quadruplex specificity and are desirable to help prevent B-DNA binding (see below).

## Molecular modeling

Modeling studies were then carried out to investigate likely binding modes of the most promising heterocycle peptide conjugates and for providing structural understanding of the role of the peptide moieties. The parallel stranded G-quadruplex formed from four repeats of human telomeric

DNA, d[AGGG(TTAGGG)$_3$], is highly asymmetric as a consequence of the position of the three TTA loops. On one side of the G-quartet platform the phosphates are presented to the solvent while on the other the bases of the loops are facing outward. This arrangement makes one face of the human G-quadruplex rich in negative charge potential while the other side is more hydrophobic in nature. Since nearly all the ligands in this study had a net positive charge most of the docking procedures were naturally biased to docking to the polar face of the G-quadruplex. This was justified since, in cases where the docking was deliberately biased to the hydrophobic face, there was still a tendency for the ligands to bind to the opposite polar face.

**Binding of acridine/acridone central core.** The results of the docking trials show clearly that there is no unique orientation for either the acridone or acridine chromophore-based ligands on the G-quartet polar surface of the human G-quadruplex. There were no major differences between the binding modes for acridone and acridine-based ligands, although the positive charge of the acridine did result in slightly less mobility for the ligands. However, a number of common trends in binding behaviour were observed. The lowest energy conformations for all the ligands have the tricyclic cores aligned across the G-quartet and between the opposite clefts formed by the loops (Fig. 4). This enabled the peptide side chains (and occasionally even main-chain peptide carbonyl and amide atoms as well) to interact effectively with the negatively charged pockets formed in the grooves by the TTA loops. Even with the peptide side-chains bound within the TTA loops, there was still 2–3 Å of variability in the position of the tricyclic chromophores due to the flexibility of the linkers used.

**Binding of peptide side chains.** As a consequence of the ambiguity in chromophore positions, the low-energy structures did not fully converge during the minimizations. However, several preferred modes of interaction were observed for each amino acid coming off each side of the tricyclic core acridine or acridone. A number of specific interactions were observed between the first residue of the tetrapeptide and the human telomeric quadruplex. For RKKV and KRSR tetrapeptides, this position is occupied by a positively charged amino acid (an arginine and a lysine respectively) the side chain of which was usually tucked into a groove-binding pocket, coordinated with 3–5 hydrogen bonds to polar atoms in the DNA backbone. In a few cases these charged amino acids were also found bound to the phosphate backbone along the edge of the loops. For FRHR conjugates, the phenylalanine side-chain was usually located either stacked over the surface of the terminal G-quartet or over the phosphate backbone of the TTA loop. In both

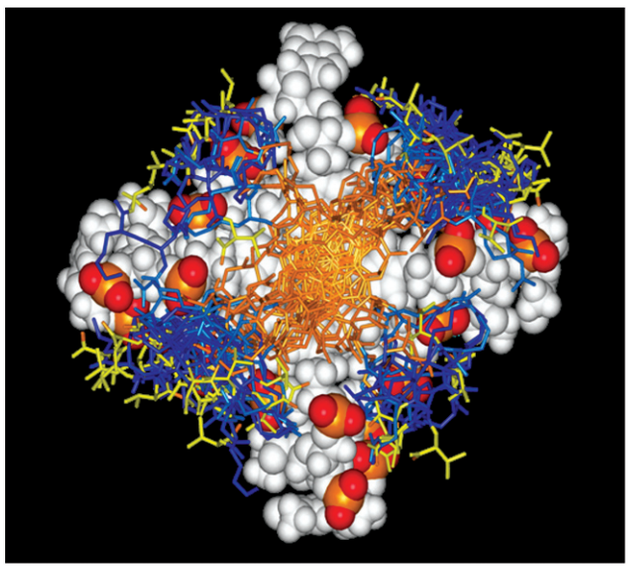

**Fig. 4** Results of the docking trial carried out with 3,6-bisRKKV acridone 3a showing global convergence. The tricyclic acridone cores (orange) are aligned across the G-quartet and between the opposite clefts formed by the loops. Arginine and lysine peptide side chains are shown in blue and valine in yellow. Only the phosphate atoms of the DNA are coloured for clarity.

orientations, but especially when stacked over the G-quartet platform (Fig. 5), this hydrophobic residue was found to have a stabilizing effect on the DNA–ligand complex.

For all three tetrapeptide sequences RKKV, KRSR and FRHR, the amino acid in the second position is positively charged and was found to interact within the TTA pocket, coordinated in much the same way as the arginine or lysine of the first position. This second amino acid was usually found to be electrostatically bound along the phosphate backbone of the adjacent stacked G-quartet.

The conformations of the final two amino acids (KV, SR and HR) showed considerably more flexibility throughout the simulation procedures, although most conformations show strong electrostatic interactions between the positively charged residues and phosphate groups lining the surface of the loops and grooves. These interactions ensure that all conformations show the tetrapeptides folded so as to maximize these interactions.

**G-quadruplex *versus* duplex recognition.** To examine the differential binding of the FRHR-acridine and -acridone ligands to human telomeric quadruplex DNA, over double stranded DNA, a model for the latter was built where the

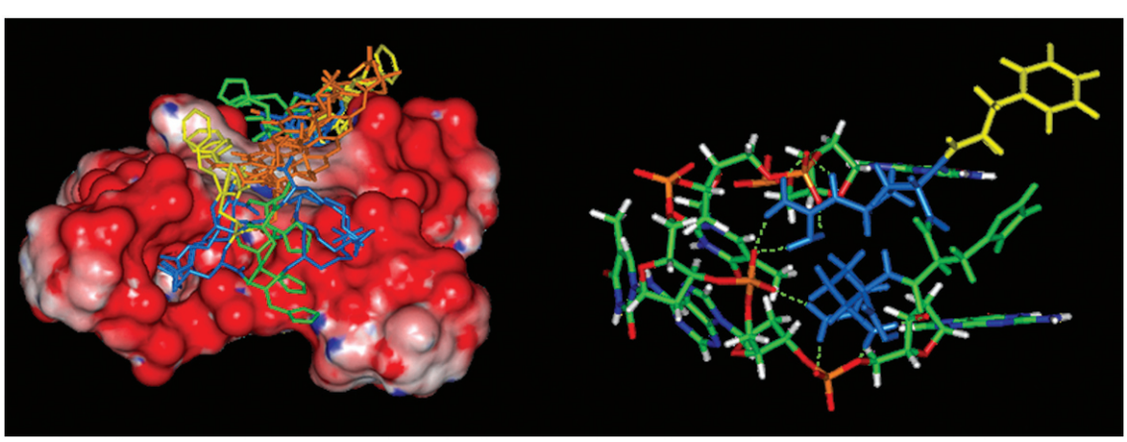

**Fig. 5** (Left) molecular model of the 3,6-bisFRHR acridone conjugate **3c**, showing an overlay of the conformations for one of the FRHR tetrapeptides in a groove of the parallel intramolecular human telomeric quadruplex. The quadruplex is coloured from red to blue over a range −5.0 to 5.0 kT e$^{-1}$ respectively. (right) Example of the H-bonding network of a low energy conformation of **3c**. Only the TTA DNA loop with flanking nucleotides and the peptide side chain of the acridone ligand is shown for clarity. The H-bonds between peptide atoms and the DNA are shown in green.

acridine was intercalated into a canonical B-DNA structure. An intercalated acridine mode is the most plausible mode of binding since that is the preferred mode for the acridine core, as shown by a number of crystallographic and modeling studies.[17] This mode shows a diminished level of interactions compared to quaduplex DNA. All models have the peptide groups in both major and minor grooves, in a spearing-type of arrangement. A simulated annealing study was carried out following a protocol similar to that followed for the human telomeric G-quadruplex. From this modelling study, we found that all low-energy conformations showed the arginine (2nd residue) side-chains lying in the DNA grooves. However the 3rd and 4th residues, especially in the minor groove as a consequence of its reduced width, cannot pack effectively into the groove, and so the number of electrostatic interactions with phosphate groups is reduced compared to their more effective packaging in the quadruplex loops (Fig. 6). Moreover, the intercalated B-DNA cannot accommodate the phenylalanine, or to a lesser extent, the valine side-chains that would be then, as a consequence, highly exposed to the solvent and these residues could not readily fold into any duplex groove. Although explicit solvent was not used in the modeling, there is no reason to believe that its exclusion could force these residues into the grooves. Comparison of the mean binding energy for the ten lowest-energy structures shows that FRHR-acridone favours the quadruplex over the duplex intercalated conformation; $-2453 \pm 35$ kcal mol$^{-1}$ *versus* $-2066 \pm 47$ kcal mol$^{-1}$ respectively.

We conclude that these ligand conformations are energetically significantly less favourable than when the same ligand is bound to the quadruplex, where the phenylalanine ring can strongly interact with the G-quartet (Fig. 6). The modelling shows that the various conformations of the RKKV and KRSR conjugate peptide side chains do have a somewhat improved fit into the duplex DNA grooves, since they are not exposing a hydrophobic residue to the solvent, as is the case with the phenylalanine residue in the FRHR conjugated acridone and acridine. This difference is reflected in the large difference in experimental Kd values between the FRHR conjugate and the other two when binding to duplex DNA (Table 1).

## Conclusion

General strategies for targeting B-DNA or RNA have included the design of hybrid molecules that combine a DNA intercalator with a sequence-specific DNA minor groove binder, such as netropsin, a distamycin analogue[18] or a peptide.[19] Herein, we have employed three selected tetrapeptides (FRHR, KRSR and RKKV), each with some intrinsic specificity for quadruplex over double-stranded DNA, and investigated their capacity to induce quadruplex over duplex DNA selectivity when appended to a heterocyclic DNA-binding scaffold. In conclusion, although acridine derivatives have been extensively studied for targeting DNA, high quadruplex specificity has previously only been achieved with 3,6,9-trisubstituted acridines and not with 3,6-disubstituted acridines.[11] Here, we have demonstrated that attaching tetrapeptides to positions 3- and 6- of an acridine or an acridone scaffold generates conjugates that bind tighter to G-quadruplexes than their individual components and which can also exhibit much higher quadruplex *versus* duplex DNA specificity. We show that these properties are peptide sequence dependent and that appropriate peptides such as FRHR can convert a 3,6-bis-substituted acridine/acridone B-DNA intercalator into a tight and quadruplex specific binder. Modeling studies suggest that human telomeric quadruplex DNA can readily accommodate tetrapeptides and that peptide side-chains can interact within the TTA loop pockets of the quadruplex. Structural diversity will now be introduced into the heterocyclic scaffold and X-ray crystallography experiments are in progress to elucidate the details of the molecular recognition by these molecules.

## Experimental

### Compounds and oligonucleotides

Peptides were synthesized on Polystyrene (PS) Wang resin (Novabiochem, initial loading: 0.9 mmol g$^{-1}$, 1% DVB) using standard Fmoc chemistry. The symmetrical 3,6-bispeptide acridone conjugates were synthesized in high yields (80–85%) from 3,6-bis carboxylic acid acridone **1** by a one-step bis-coupling on resin as previously reported.[13] A similar synthetic pathway was used for the preparation of the 3,6-bispeptide acridine conjugates. The heterocyclic precursors were synthesized by reacting sarcosine on the previously described 3,6-bis(3-chloropropionamido) acridone and acridine. All the compounds were purified by HPLC and characterized by high resolution mass spectrometry and amino acid analysis. Aqueous solutions of each compound were quantitated by amino acid analysis. Oligodeoxynucleotides probes were synthesized by Oswel Ltd. All concentrations were expressed in strand molarity with a nearest-neighbor approximation for the absorption concentrations of the unfolded species.

Three 5′-biotinylated oligonucleotides were used for SPR (Oswel, HPLC purified): the *Htelo* human telomeric quadruplex d(biotin-[(GTTAGG)$_5$]), the dsDNA double-stranded telomeric control d(biotin-[GACGTGTGGACTGTGA(GGTTAG)$_5$]) hybridized with its complementary sequence and the *Atelo* single-stranded DNA control d(biotin-[(AGTTAG)$_5$]).

Biotinylated (GTTAGG)$_5$ (*Htelo*) was annealed in 50 mM Tris (pH = 7.4) with 150 mM KCl by heating to 95 °C for 5 min and then slow cooling to room temperature as described previously.[10] After overnight incubation at 4 °C, the stock so-

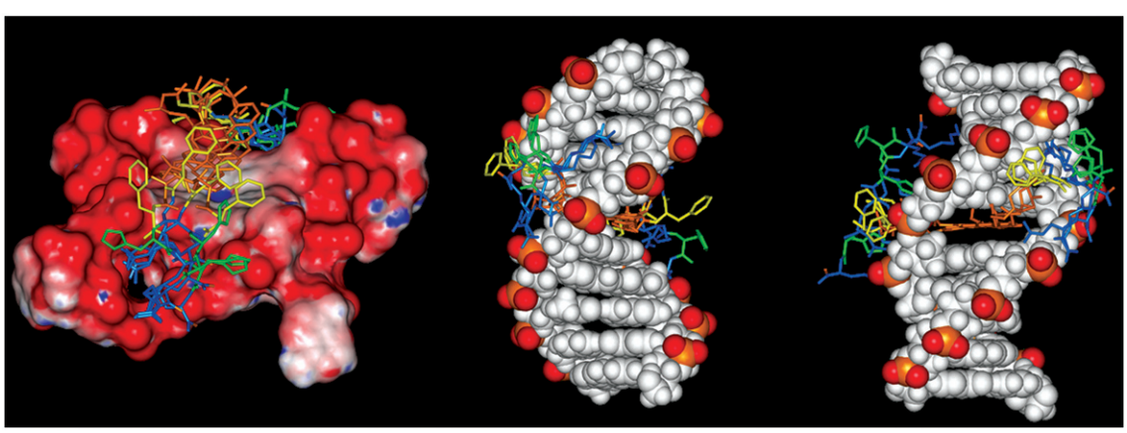

**Fig. 6** Low-energy conformations of the 3,6-bisFRHR acridine conjugate **4c** docked on the top-tetrad of a parallel intramolecular telomeric quadruplex coloured as in Fig. 5 (left) and (right) in a B-DNA intercalation site showing interaction of peptides in both minor and major grooves. Phe residues tend to be bound in a hydrophobic pocket when interacting with a quadruplex (left) while they are mostly unbound to B-DNA and are thus highly exposed to solvent (right).

lutions of quadruplex were stored at −20 °C. The folding of *Htelo* into a quadruplex structure was confirmed by circular dichroism (characteristic peak observed at 295 nm). Under similar conditions, the single-stranded telomeric control shows no signal at 295 nm indicative of an unstructured conformation. The biotinylated double-stranded telomeric control was obtained by annealing the biotinylated [GACGTGTGGACTGTGA] with an equimolar amount of [(CTAACC)5TCACAG TCCACACGTC] in 20 mM Tris pH 8.0, 100 mM NaCl by heating to 90 °C for 5 min and slow cooling to room temperature. The fill-in reaction was carried out at 37 °C for 1 h in 10 mM Tris pH 8.0, 10 mM $MgCl_2$, 10 mM DTT, 2 mM dNTPs, 1.25 μM annealed oligo and 1 μM Klenow polymerase. The DNA was then purified using a nucleotide purification kit (Qiagen) and eluted in phosphate buffer.

### Surface plasmon resonance

Surface plasmon resonance (SPR) measurements were performed with a four-channel BIAcore 2000 optical biosensor system (Biacore Inc.) and streptavidin-coated sensor chips (Biacore SA-chip). In a typical experiment, biotinylated *Htelo* (200 response unit RU), biotinylated *Atelo* (200 RU) and biotinylated *ds-DNA* (450 RU) were immobilized in flow cells 2, 3 and 4, leaving flow cell 1 as a blank. DNA binding experiments were carried out in phosphate buffer at a flow rate of 10 μl $min^{-1}$. Ligand solutions were prepared in freshly degassed phosphate buffer by serial dilutions from stock solution. These solutions were injected using the INJECT command (Biacore 2000 Control Software version 3.1) for 2 min followed by a 30 s 1 M KCl injection for chip regeneration. Each experiment was repeated in triplicate.

The response at equilibrium ($R_{eq}$) was plotted against concentration of analyte to generate a hyperbolic binding curve. A single-stranded control oligonucleotide ($(AGTTAG)_5$ *Atelo*) derived from the telomeric quadruplex sequence but carrying a G→A mutation that prevents quadruplex formation[10] was also used as a suitable blank control since no significant binding was detected in the concentration range of these experiments. An upper limit of 25 μM was then chosen where no binding was detectable to either the single-stranded (*Atelo*) control or the blank cell. The final graphs were obtained by subtracting Atelo sensorgrams from the duplex or quadruplex sensorgrams. For all compounds whose *Kd* was found to be below 25 μM, dissociation constants were determined by fitting the binding curve (from at least six concentrations) using the steady state algorithm (Biaevaluation 3.1). For compounds **3c** and **4c**, no binding to double-stranded DNA was detectable at a concentration of 25 μM and 3 μM of analyte respectively. The lower limit for the *Kd* was then estimated from eqn. 1 by calculating the minimum *Kd* value that would give a response of at least 2 RU (detection limit of the BIAcore) at a concentration of 25 μM for **3c** (or 3 μM for **4c**). Applied to all the other derivatives, this estimation gave *Kd* lower limits approximately two times lower than the actual *Kd*, thus supporting this approximation and suggesting that the limits quoted for compounds **3c** and **4c** are possibly under-estimated.

$$Kd > (C\mathrm{RU}_{max} - C\mathrm{RU}_{min})/\mathrm{RU}_{min} \quad (1)$$

where C is the maximum concentration of analyte showing no detectable binding ($RU_{eq} < 2$) to the considered DNA target; $RU_{max}$ is the predicted maximal response per bound compound in the steady-state region, for a binding stoichiometry of 1 : 1; $RU_{min}$ is the weakest signal that is detectable from noise signal by SPR and has been fixed to 2.

### Docking procedure

The X-ray derived structure of the parallel-stranded G-quadruplex formed from the 22-mer human telomeric DNA sequence $d[AGGG(TTAGGG)_3]$ (PDB id 1KF1),[2] was used as the binding host to the various acridine/acridone peptide conjugates. The ligands were built using the Insight II package (Accelrys Inc) on an SGI workstation. Partial charges for the acridine and acridone central scaffold were derived by fitting the HF/6–31G* electrostatic potential, calculated with GAMESS *ab initio* software,[20] to the atomic centers with the RESP program.[21] Atomic parameters for the peptide side chains were taken directly from the Amber force field[21] and atom types for the core assigned automatically. The protonation state of the amino acid side chains were determined based on an assumed physiological pH of 7.0. The N10 position of the acridine was also protonated with a +1 charge based on the observation of its role as a hydrogen bond donor in a quadruplex–ligand crystal structure.[12]

Docking was performed with the Affinity Docking module of Insight II. A multi-phase docking protocol was used. First the ligand was randomly orientated 200 times, centered on the more polar face of the quadruplex. Van der Waals (VDW) radii were set to 10% of the full value, charges were not considered and non-bonded cut-offs were set to 8 Å. The system was minimized for 300 steps using the conjugate gradient method. The maximum allowable change for succeeding structures (energy tolerance) was set to 10 000 kcal $mol^{-1}$ and the energy range was set to 30 to 40 kcal $mol^{-1}$. The 75 lowest energy structures were used for the second phase of the modeling.

Simulated annealing was used for the second phase in order to further refine the initial placement for the 75 structures. During this phase the Van der Waals radii were adjusted to their full values, charges were included with a distance dependent dielectric of $4 \cdot r$ and the non-bonded cut-off set to 12 Å. The system was again minimized for 300 steps of conjugate gradient and then molecular dynamics were performed, starting at a temperature of 500 K and cooling the system to 300 K over 7.5 to 10 ps. The resulting structures were minimized for 800 steps of conjugate gradients and the 25 structures with the lowest total energy were used for further evaluation. Docked molecules were selected based on three criteria, total energy of the system, energy of the ligand and binding pocket (with all atoms less than 5.5 Å from ligand), and total number of intermolecular hydrogen bonds.

Several of the ligands were further refined by another round of simulated annealing. In this phase the best-fit lowest-energy structure from the first round of simulated annealing was used to define a binding pocket containing atoms 5 Å from the ligand while the bases of the G-quartet were tethered to their original position. Then 30 further rounds of simulated annealing were performed, starting at a temperature of 800 K and cooling the system to 200 K over 10 ps. This allowed for optimal conformational sampling of the peptide side-chains while maintaining the acridine/acridone orientation.

For the modelling of the FRHR-acridine to duplex DNA an initial model was built from an idealized model of acridine intercalated into B form DNA. This was used as a starting structure for 30 rounds of simulated annealing. The same parameters were used as for the further refinement stage where the initial temperature was 800 K and the system cooled to 200 K over 10 ps with a distance-dependent dielectric constant. During the simulation the B-DNA was held fixed except for the two base pair residues above and below the intercalation site. As before, this was sufficient to sample many side chain conformations of the ligand and maintain the overall integrity of the model.

## Acknowledgements

This research was supported by an European Community Marie Curie Fellowship and by grants from the BBSRC and Cancer Research UK.

## References

1 T. Simonsson, *J. Biol. Chem.*, 2001, **368**, 621–628; S. Neidle and G. N. Parkinson, *Curr. Opin. Struct. Biol.*, 2003, **13**, 275–283.

2 Y. Wang and D. J. Patel, *Structure*, 1993, **1**, 262–282; G. N. Parkinson, M. P. Lee and S. Neidle, *Nature*, 2002, **417**, 876–880; A. T. Phan and D. J. Patel, *J. Am. Chem. Soc.*, 2003, **125**, 15021–15027.
3 J. C. Darnell, K. B. Jensen, P. Jin, V. Brown, S. T. Warren and R. B. Darnell, *Cell*, 2001, **107**, 489–499; R. H. Shafer and I. Smirnov, *Biopolymers*, 2001, **56**, 209–227; H. Arthanari and P. H. Bolton, *Chem. Biol.*, 2001, **8**, 221–230; C. Marchand, P. Pourquier, G. S. Lac, N. Jing and Y. Pommier, *J. Biol. Chem.*, 2002, **277**, 8906–8911; E. M. Rezler, D. J. Bearss and L. H. Hurley, *Curr. Opin. Pharmacol.*, 2002, **2**, 415–423.
4 S. Neidle and G. Parkinson, *Nat. Rev. Drug Discovery*, 2002, **1**, 384–393.
5 T. R. Cech, *Angew. Chem., Int. Ed.*, 2000, **39**, 34–43.
6 For example see: D. Sun, B. Thompson, B. E. Cathers, M. Salazar, S. M. Kerwin, J. O. Trent, T. C. Jenkins, S. Neidle and L. H. Hurley, *J. Med. Chem.*, 1997, **40**, 2113–2116; P. J. Perry, S. M. Gowan, A. P. Reszka, P. Polucci, T. C. Jenkins, K. R. Kelland and S. Neidle, *J. Med. Chem.*, 1998, **41**, 3252–3260; H. Han, D. Langley, A. Rangan and L. H. Hurley, *J. Am. Chem. Soc.*, 2001, **123**, 8902–8913; A. Maraval, S. Franco, C. Vialas, C. Pratviel, M. A. Blasco and B. Meunier, *Org. Biomol. Chem.*, 2003, **1**, 921–927; M.-P. Teulade-Fichou, C. Carrasco, L. Guittat, C. Bailly, P. Alberti, J.-L. Mergny, A. David, J.-M. Lehn and W. D. Wilson, *J. Am. Chem. Soc.*, 2003, **125**, 4732–4740; R. J. Harrison, J. Cuesta, G. Chessari, M. Read, S. K. Basra, A. P. Reszka, J. Morrell, S. M. Gowan, C. M. Incles, F. A. Tanious, W. D. Wilson, L. R. Kelland and S. Neidle, *J. Med. Chem.*, 2003, **46**, 4463–4476.
7 J. A. Schouten, S. Ladame, S. J. Mason, M. A. Cooper and S. Balasubramanian, *J. Am. Chem. Soc.*, 2003, **125**, 5594–5595.
8 A. Rangan, O. Y. Fedoroff and L. H. Hurley, *J. Biol. Chem.*, 2001, **276**, 4640–4646; A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 11593–11598.
9 T. Simonsson, P. Pecinka and M. Kubista, *Nucleic Acids Res.*, 1998, **26**, 1167–1172.
10 M. Isalan, S. D. Patel, S. Balasubramanian and Y. Choo, *Biochemistry*, 2001, **40**, 830–836; C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H. Lipps and A. Plückthorn, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 8572–8577.
11 M. Read, R. J. Harrison, B. Romagnoli, F. A. Tanious, S. H. Gowan, A. P. Reszka, W. D. Wilson, L. R. Kelland and S. Neidle, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 4844–4849.
12 S. M. Haider, G. N. Parkinson and S. Neidle, *J. Mol. Biol.*, 2003, **326**, 117–125.
13 S. Ladame, R. J. Harrison, S. Neidle and S. Balasubramanian, *Org. Lett.*, 2002, **4**, 2519–2522.
14 L. Wang, C. Carrasco, A. Kumar, C. E. Stephens, C. Bailly, D. W. Boykin and W. D. Wilson, *Biochemistry*, 2001, **40**, 2511–2521; E. R. Lacy, N. M. Le, C. A. Price, M. Lee and W. D. Wilson, *J. Am. Chem. Soc.*, 2002, **124**, 2153–2163; C. Carrasco, M. Facompre, J. D. Chisholm, D. L. Van Vranken, W. D. Wilson and C. Bailly, *Nucleic Acids Res.*, 2002, **30**, 1774–1781.
15 T. M. Davis and W. D. Wilson, *Anal. Biochem.*, 2000, **284**, 348–353.
16 Nonspecific interactions between the ligand and the matrix of the sensor chip at higher concentrations of analyte, as observed by others, precluded the use of SPR for quoting an accurate *Kd* value for *ds-DNA*. For example, see: C. Carrasco, F. Rosu, V. Gabelica, C. Houssier, E. De Pauw, C. Garbay-Jaureguiberry, B. Roques, W. D. Wilson, J. B. Chaires, M. J. Waring and C. Bailly, *ChemBioChem*, 2002, **3**, 1235–1241.
17 H. S. Shieh, H. M. Berman, M. Dabrow and S. Neidle, *Nucleic Acids Res.*, 1980, **8**, 85–97; S. Neidle, H. M. Berman and H. S. Shieh, *Nature*, 1980, **288**, 129–133.
18 For example see: P. A. Boriack, D. W. Christianson, J. K. Wood and G. M. Whitesides, *J. Med. Chem.*, 1995, **38**, 2286–2291; D. J. Maly, I. C. Choong and J. A. Ellman, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 2411–2424; C. Bailly and J. B. Chaires, *Bioconjugate Chem.*, 1998, **9**, 513–538; C. Carrasco, P. Helissay, M. Haroun, B. Baldeyrou, A. Lansiaux, P. Colson, C. Houssier, S. Giorgi-Renault and C. Bailly, *ChemBioChem*, 2003, **4**, 50–61.
19 V. Guelev, J. Lee, J. Ward, S. Sorey, D. W. Hoffman and B. L. Iverson, *Chem. Biol.*, 2001, **8**, 415–425; V. Guelev, S. Sorey, D. W. Hoffman and B. L. Iverson, *J. Am. Chem. Soc.*, 2002, **124**, 2864–2865; Z. Zhang, P. Chaltin, A. Van Aerschot, J. Lacey, J. Rozenski, R. Busson and P. Herdenwijn, *Bioorg. Med. Chem.*, 2002, **10**, 3401–3413; C. B. Carlson, R. J. Spanggord and P. A. Beal, *ChemBioChem*, 2002, **124**, 8510–8511.
20 *GAMESS-UK*, a package of *ab initio* programs written by M. F. Guest, J. H. van Lenthe, J. Kendrick, K. Schoeffel, P. Sherwood, and R. J. Harrison, with contributions from R. D. Amos, R. J. Buenker, H. van Dam, M. Dupuis, N. C. Handy, I. H. Hillier, P. J. Knowles, V. Bonacic-Koutecky, W. von Niessen, A. P. Rendell, V. R. Saunders, A. Stone and D. Tozer. The package is derived from the original GAMESS code due to M. Dupuis, D. Spangler and J. Wendoloski, NRCC Software Catalog, vol. 1, program No. QG01 (*GAMESS*), 1980.
21 S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *J. Comput. Chem.*, 1986, **7**, 230–253; W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Goulg, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.