Structure-based design of selective high-affinity telomeric quadruplex-binding ligands†

Caterina Maria Lombardo,[a] Iria Sánchez Martínez,[a] Shozeb Haider,[a] Valérie Gabelica,[b] Edwin De Pauw,[b] John E. Moses[c] and Stephen Neidle*[a]

A library of triazole-based telomeric quadruplex-selective ligands has been developed that mimic an established family of tri-substituted acridine-based ligands, using crystal structure data as a starting-point for computer-based design. Binding affinities, estimated by electrospray mass spectrometry, are in accord with the design concept.

Oligonucleotides and nucleic acids containing G-tracts can be organised as G-quadruplexes.[1] These are polymorphic tertiary structures, characterised by a hydrophobic core of G-quartets and negatively-charged loops. Quadruplexes show exceptional stability over other conformations in the presence of Na^+ or K^+ ions. Putative quadruplex sequences have been identified in G-rich genomic sequences,[2] with over-representation in telomeres,[3] as well as in other genomic regions for example in promoter sequences of a number of proto-oncogenes,[4] such as c-myc[5a] and c-kit,[5b] in 5′ untranslated regions[6a] and in introns.[6b] A number of these putative quadruplexes are appealing targets for cancer therapeutics. For instance, inducing the single-stranded telomeric DNA overhang to fold into G-quadruplexes has been shown to inhibit telomerase activity[7a] and cancer cell growth.[7b] Such precise targeting of human telomeres is significant since in >80% of cancers telomerase is up-regulated and contributes to the malignant phenotype by maintaining cancer cell immortalization.[7c]

A considerable number of small organic molecules have been found to stabilise quadruplex DNA structures.[8] Many, though not all, are based on polycyclic heteroaromatic cores, with the acridine nucleus being especially well explored.[9] However, the selectivity of many of these molecules for G-quadruplexes over duplex DNA is frequently less than what would be therapeutically acceptable, and their polycyclic features can make their druggability a challenge.

The 3,6,9-trisubstituted acridine ligand BRACO-19 (Fig. 1) has high affinity for human telomeric quadruplex DNAs and is a potent inhibitor of the telomerase enzyme.[10] It has selective cytotoxic activity against a range of human tumour cell lines
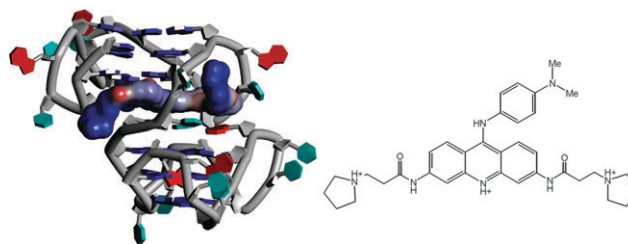

Fig. 1 (Left) Representation of a 45-mer telomeric DNA sequence, with two G-quadruplexes and a trisubstituted ligand (compound 15) in a low-energy docked position at the interface between the two quadruplexes. The ligand is shown as a solvent-accessible surface colored by charge. (Right) The BRACO-19 molecule.

and shows antitumour activity against xenograft models.[11] The BRACO-19 molecule was designed using qualitative molecular modeling, with the crystal structure of the native parallel human telomeric quadruplex as a template. It was rationalized that each of the three substituents emanating from the acridine core of BRACO-19 would be able to interact with a quadruplex groove.[10] This feature would, it was suggested, provide binding selectivity over duplex DNA, which has just two grooves. A more recent crystal structure of a BRACO-19 complex with a bimolecular quadruplex has confirmed the essential correctness of this hypothesis and has also provided a more detailed view of the interactions involved.[12] The structure has a parallel-stranded quadruplex arrangement, with the biological unit being two 5′ to 3′ stacked quadruplexes. Each bimolecular quadruplex in this structure contains three planar stacked G-quartets with a BRACO-19 molecule stacking directly onto the 3′ end G-quartet face.

We report here the structure-based design, synthesis and preliminary assessment of a novel series of non-polycyclic trisubstituted ligands, whose affinity and selectivity for telomeric G-quadruplex DNA has been evaluated using electrospray mass spectrometry (ESI-MS).[13] The goal has been to design molecules (i) with potentially enhanced selectivity based on the quadruplex concept of selective groove binding, and (ii) that do not have a polycyclic heteroaromatic core, so potentially enhancing drug-like features. We have used the BRACO-19-quadruplex complex crystal structure[12] as a starting point for the structure-based design of non-polycyclic mimetics of BRACO-19. The single-stranded overhang of human telomeric DNA is 100–200 nucleotides in length, and in principle several quadruplex structures can be formed along its length. We have previously modeled such a higher-order arrangement using a simple linker between 5′ to 3′ ends in this crystal structure to form a continuous sequence, which does not involve any

[a] CRUK Biomolecular Structure Group, The School of Pharmacy, University of London, London WC1N 1AX, UK.
E-mail: stephen.neidle@pharmacy.ac.uk; Tel: +44 (0)20 7753 5969
[b] Physical Chemistry and Mass Spectrometry Laboratory, Department of Chemistry, University of Liege, B-4000 Liege, Belgium
[c] The School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD, UK
† Electronic supplementary information (ESI) available: Details of synthetic chemistry, molecular modelling, electrospray mass spectroscopy and results of ESI-MS competition experiments. See DOI: 10.1039/c0cc02917c

perturbation of the ligand binding site.[14] The resulting 45-mer bis-quadruplex ligand sandwich complex modeled structure has been used here as the template for initial qualitative modelling followed by *in silico* docking and binding energy calculations for plausible ligands.[15] We started with the hypothesis that such ligands would require three substituents as in BRACO-19 itself, and have explored their length and size. Ligand positions were explored computationally at the interface between the two quadruplexes, stacked on the 3′ and 5′ terminal G-quartet surfaces. Ligands were derived from a previously devised[16] series of bis-triazole quadruplex-binding ligands which contain two alkylamine side-chain arms linked to a benzene core through 1,4-triazoles. These had been prepared using a click chemistry approach,[17] which was also used here to add an additional triazole 'side arm' to the benzene core.

Initial qualitative modeling with the 45-mer suggested that addition of the third alkylamine side-chain (Fig. 1) could effectively mimic the key structural features of the BRACO-19 structure. This was followed by molecular dynamics simulations[15] to ascertain whether the resulting complex was conformationally stable. The resulting structure (Fig. 1) indicates that the arrangement is structurally sound, with the central and three attached phenyl rings constituting the core of the ligand, all being involved in π–π stacking with G-quartet guanine bases in the binding site. The three alkylamino arms are each positioned in a groove structure, analogous to disposition of the side-chains in the BRACO-19 complex crystal structure. Circular dichroism studies on the binding of compound 15 to a human telomeric quadruplex sequence show that the parallel form is induced (see the ESI†), giving confidence to the use of the parallel crystal structure in the modelling and simulation studies. Even so, docking studies were also performed on an alternative polymorph of the human telomeric quadruplex, one of the (3 + 1) hybrid anti-parallel structures,[18] but all low-energy arrangements involved stereochemically unacceptable buckling of ligand, DNA or both. These *in silico* trials were then abandoned.

The synthesis of a small library of the trisubstituted click ligands used a convergent click chemistry approach.[17] A series of azide building blocks with focussed structural diversity was initially synthesised, then clicked onto a central core, the commercially available trialkyne 1,3,5-triethynylbenzene (Scheme 1). The diversity of building blocks was designed in order to establish structure–activity relationships. They all have the common features of a phenyl ring (providing an aromatic surface) and a series of basic side-chains. These were built onto a starting 2-nitroaniline or 3-nitroaniline in a maximum of four steps. Briefly, the required nitroaniline was acylated, then one-pot substituted with diverse amines. The nitro group was readily reduced with H_2, then the azide was synthesised from the resulting aniline with one-pot diazotisation and azide substitution.[19] The 1,3,5-triethynylbenzene was then extended with one-pot click reaction formation of three triazole rings *via* Cu(I) catalysed Huisgen 1,4-dipolar cyclo-addition. Catalytic Cu(I) was formed *in situ* and a second catalyst, bathophenanthrolinedisulfonic acid disodium salt hydrate, was necessary to achieve complete trisubstitution. The reaction was complete after only 15 minutes of microwave irradiation; an excess of the required amine in the reaction
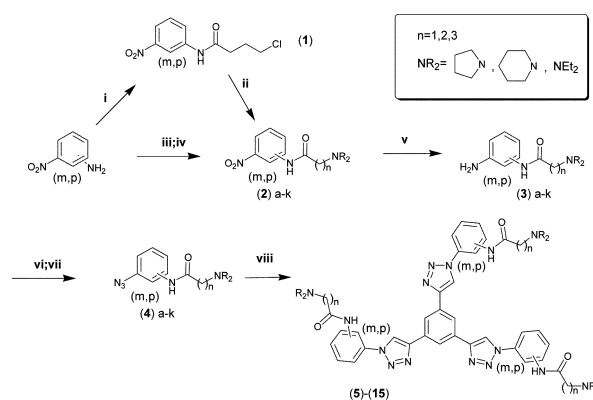
Scheme 1 Synthesis of ligands **5–15**: (i) 4-chlorobutyryl chloride (4 eq.), 4 °C to rt, overnight; (ii) pyrrolidine (5 eq.), 4 °C to rt, overnight; (iii) chloroacetyl chloride/3-chloropropionyl chloride, TEA (2 eq.), THF, 4 °C to rt, 2 h; (iv) piperidine/pyrrolidine/diethylamine (3 eq.), THF, 4 °C to rt, overnight; (v) anhydrous THF, H_2, Pd/C (10% m/m), rt, overnight; (vi) ^{t}BuONO (2.5 eq.), HCl (5.5 eq.), THF, 0 °C, 1.5 h; (vii) NaN_3 (3 eq.), H_2O, 0 °C to rt, overnight; (viii) 1,3,5-triethynylbenzene, compound (**4a–k**) (4 eq.), $CuSO_4$ (0.05 eq.), sodium ascorbate (0.2 eq.), bathophenanthrolinedisulfonic acid disodium salt hydrate (0.1 eq.), H_2O–^{t}BuOH, 110 °C, microwave, 15 min.

mixture was necessary to avoid elimination of the amine for side-chains with $n \geq 2$. The resulting compounds (**5–15**), all have an extended aromatic surface. Initial attempts to study DNA binding using techniques to measure the elevation in melting temperature as a result of ligand were unsuccessful since it appears that the compounds are unstable when UV irradiated at the elevated temperatures required in melting experiments. Their quadruplex and duplex DNA binding abilities were therefore assessed by electrospray mass spectrometry.

Binding of each ligand to the 22-mer human telomeric G-quadruplex sequence d[AGGG(TTAGGG)₃] (tel22) was examined at 1 : 1 and 2 : 1 drug : DNA ratios (Table 1, see also Fig. S1 and accompanying text, ESI†). All the K_d values reported refer to the 1 : 1 complex $(Q + L)^{5-}$. The *meta*-substituted compounds are consistently the strongest binders, with compounds 5, 8, 12 and 15 having lower K_d values than that of BRACO-19 itself.[20] All of these have short side chains ($n = 1, 2$) with either pyrrolidine or diethylamine basic groups. These compounds have consistently higher affinities than the corresponding *para* compounds, with those having pyrrolidine and diethylamine end-groups being superior to the piperidine derivatives 11 and 13. Compounds 5 ($n = 1$) and 15 ($n = 2$) are more active than the related compound (7) ($n = 3$), which suggests that shorter side chains result in superior quadruplex binding.

ESI-MS has also been used to assess quadruplex:duplex selectivity in a competition experiment between three sequences.[21] The telomeric 22-mer (tel22), the parallel intermolecular G-quadruplex sequence (dTG₄T)₄ and a duplex DNA, d(CGCGAATTCGCG)₂, each at 5 μM, were injected with the ligand at 10 μM. Almost all the compounds showed selectivity for G-quadruplex sequences and did not bind to the duplex (Fig. S5, ESI†).

Several of the compounds have higher affinity for the tetra-molecular TG₄T quadruplex than for the telomeric quadruplex

Table 1 ESI-measured dissociation constants for compounds **5–15** with a human telomeric 22-mer quadruplex[a]. NR_2 groups are defined in Scheme 1

Compound	Substitution pattern	n	NR_2	$K_d/\mu M$
5	*meta-*	1	pyr	6.0 ± 0.6
6	*para-*	2	pyr	70 ± 50
7	*meta-*	3	pyr	35 ± 9
8	*meta-*	2	dieth	5.0 ± 1.0
9	*para-*	2	dieth	100 ± 19
10	*para-*	1	pyr	49 ± 26
11	*meta-*	1	pip	250 ± 130
12	*meta-*	1	dieth	3.0 ± 0.6
13	*para-*	1	pip	77 ± 12
14	*para-*	1	dieth	32 ± 5
15	*meta-*	2	pyr	4.9 ± 1.3
BRACO-19	n/a	n/a	n/a	7.9 ± 1.4

[a] Abbreviations: (pyr) pyrrolidino, (dieth) diethyl amino, (pip) piperidino. BRACO-19 has been used as a reference with data taken from ref. 20. Esds from two ligand concentrations (5 and 10 μM) and three voltage settings each.

or for duplex DNA. By examining the ESI-MS spectra (see Fig. S1 and accompanying text, ESI†), we noted that, in addition to binding to tel22 to form 1 : 1 complexes, all ligands with *para* substituents were causing partial dimerization of the $(dTG_4T)_4$ quadruplex. However, compounds **5**, **8**, **12** and **15** have high affinity for the telomeric sequence, even in the presence of $(dTG_4T)_4$. The correlation with the K_d values is shown in Fig. S3 (ESI†). These are also the compounds that have the highest affinity for the telomeric quadruplex. Calculated binding energies[15] ($\Delta E_{binding}$) for **5**, **7** and **15** are −8.64, −7.99 and −9.63 kcal mole^{-1}, respectively, which concurs with the ranking order in Table 1.

Ligand **11** did not bind significantly to either of the other sequences. None of the ligands bound to the duplex sequence. Mixtures of 5 μM DNA duplex + 40 μM ligand were run in an attempt to determine the K_d for ligand binding to duplex, in order to calculate the selectivity of the ligands for the telomeric quadruplex; no complex formation was detected with any of the ligands examined in these particular experiments (see Fig. S4, ESI†). From the noise level where the complex should be detected, we estimated the K_d that would lead to a detectable complex (S/N = 3 for the complex), and deduced a K_d value for duplex binding of > 3000 μM. This gives a quadruplex/duplex selectivity ratio > 1000 for ligand **12**, for example. Although K_d values for the disubstituted triazole compounds are not available, a K_a value[22] for the analogue of compound **6**, with pyrrolidino end-groups of 7.7×10^5 M^{-1}, suggests that the trisubstitution has resulted in an enhancement of affinity, at least for the *para*-substituted compounds.

In summary a group of ligands have been designed which demonstrate high affinity for an intramolecular human telomeric quadruplex that is comparable to, and in several instances, exceeds that of the established acridine ligand BRACO-19. The tris-triazole ligands have higher selectivity for quadruplex over duplex DNA compared to BRACO-19, with > 1000 fold difference in K_d values. Three compounds **8**, **12** and **15** have been selected for biological studies, which are currently underway.

Notes and references

1 D. J. Patel, A. T. Phan and V. Kuryavyi, *Nucleic Acids Res.*, 2007, **35**, 7429; S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res.*, 2006, **34**, 5402; S. Neidle, *Curr. Opin. Struct. Biol.*, 2009, **19**, 239.
2 J. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2005, **33**, 2908; A. Todd, M. Johnston and S. Neidle, *Nucleic Acids Res.*, 2005, **33**, 2901.
3 M. J. McEachern, A. Krauskopf and E. H. Blackburn, *Annu. Rev. Genet.*, 2000, **34**, 331.
4 J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2006, **35**, 406.
5 See for example: (*a*) A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 11593; Y. Qin, J. S. Fortin, D. Tye M. Gleason-Guzman, T. A. Brooks and L. H. Hurley, *Biochemistry*, 2010, **49**, 4208; (*b*) S. T. Hsu, P. Varnai, A. Bugaut, A. P. Reszka, S. Neidle and S. Balasubramanian, *J. Am. Chem. Soc.*, 2009, **131**, 13399; M. Gunaratnam, S. Swank, S. M. Haider, K. Galesa, A. P. Reszka, M. Beltran, F. Cuenca, J. A. Fletcher and S. Neidle, *J. Med. Chem.*, 2009, **52**, 3774.
6 (*a*) For example: S. Kumari, A. Bugaut, J. L. Huppert and S. Balasubramanian, *Nat. Chem. Biol.*, 2007, **3**, 218; M. Wieland and J. S. Hartig, *Chem. Biol.*, 2007, **14**, 757; A. Bugaut, R. Rodriguez, S. Kumari, S. T. Hsu and S. Balasubramanian, *Org. Biomol. Chem.*, 2010, **8**, 2771; (*b*) J. Eddy and N. Maizels, *Nucleic Acids Res.*, 2007, **36**, 1321.
7 (*a*) 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; (*b*) P. Phatak, J. C. Cookson, F. Dai, V. Smith, R. B. Gartenhaus, M. F. Stevens and A. M. Burger, *Br. J. Cancer*, 2007, **96**, 1223; (*c*) J. W. Shay and W. E. Wright, *Nat. Rev. Drug Discovery*, 2006, **5**, 577.
8 A. De Cian, L. Lacroix, C. Douarre, N. Temime-Smaali, C. Trentesaux, J. F. Riou and J. L. Mergny, *Biochimie*, 2008, **90**, 131; D. Monchaud and M.-P. Teulade-Fichou, *Org. Biomol. Chem.*, 2008, **6**, 627.
9 See for example: R. J. Harrison, S. M. Gowan, L. R. Kelland and S. Neidle, *Bioorg. Med. Chem. Lett.*, 1999, **9**, 2463; M.-K. Cheng, C. Modi, J. C. Cookson, I. Hutchinson, R. A. Heald, A. J. McCarroll, S. Missailidis, F. Tanious, W. D. Wilson, J.-L. Mergny, C. A. Laughton and M. F. G. Stevens, *J. Med. Chem.*, 2008, **51**, 963.
10 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. U. S. A.*, 2001, **98**, 4844.
11 A. M. Burger, F. Dai, C. M. Schultes, A. P. Reszka, M. J. Moore, J. A. Double and S. Neidle, *Cancer Res.*, 2005, **65**, 1489.
12 N. H. Campbell, G. N. Parkinson, A. P. Reszka and S. Neidle, *J. Am. Chem. Soc.*, 2008, **130**, 6722.
13 F. Rosu, E. De Pauw and V. Gabelica, *Biochimie*, 2008, **90**, 1074; V. Gabelica, *Methods Mol. Biol.*, 2010, **613**, 89.
14 S. M. Haider and S. Neidle, *Biochem. Soc. Trans.*, 2009, **37**, 583.
15 S. M. Haider and S. Neidle, *Methods Mol. Biol.*, 2010, **608**, 17. See also the ESI†.
16 A. D. Moorhouse, A. M. Santos, M. Gunaratnam, M. Moore, S. Neidle and J. E. Moses, *J. Am. Chem. Soc.*, 2006, **128**, 15972.
17 H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2001, **40**, 2004.
18 A. T. Phan, V. Kuryavyi, K. N. Luu and D. J. Patel, *Nucleic Acids Res.*, 2007, **35**, 6517; A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2006, **34**, 2723.
19 J. Das, S. N. Patil, R. Awasthi, C. P. Narasimhulu and S. Trehan, *Synthesis*, 2005, 1801.
20 G. Collie, A. P. Reszka, S. M. Haider, V. Gabelica, G. N. Parkinson and S. Neidle, *Chem. Commun.*, 2009, 7482.
21 A. Casagrande, A. Alvino, G. Bianco, G. Ortaggi and M. Franceschin, *J. Mass Spectrom.*, 2009, **44**, 530.
22 T. P. Garner, H. E. L. Williams, K. I. Gluszyk, S. Roe, N. J. Oldham, M. F. G. Stevens, J. E. Moses and M. S. Searle, *Org. Biomol. Chem.*, 2009, **7**, 4194.