

“Porphyrin binding mechanisms is altered by protonation at the loops in G-quadruplex DNA formed near the transcriptional activation site of the human *c-kit* gene” Manaye, S., Eritja, R., Aviñó, A., Jaumot, J., Gargallo, R. *Biochim. Biophys Acta*, 1820, 1987-1996 (2012). Doi: 10.1016/j.bbagen.2012.09.006

Porphyrin binding mechanism is altered by protonation at the loops in G-quadruplex DNA formed near the transcriptional activation site of the human *c-kit* gene

Sintayehu Manaye ^a, Ramon Eritja ^b, Anna Aviñó ^b, Joaquim Jaumot ^a, Raimundo Gargallo ^{a,*}

^aSolution Equilibria and Chemometrics Group (Associate Unit UB-CSIC), Department of Analytical Chemistry, University of Barcelona, Diagonal 645, E-08028 Barcelona, Spain

Institute for Research in Biomedicine, IQAC-CSIC, CIBER-BBN, Baldiri i Reixac 15, E-08028 Barcelona, Spain

*Corresponding author. Tel.: +34 934039274; fax: +34 934021233. E-mail address: raimon_gargallo@ub.edu (R. Gargallo).

Abstract.

Background: G-quadruplex DNA structures are hypothesized to be involved in the regulation of gene expression and telomere homeostasis. The development of small molecules that modulate the stability of G-quadruplex structures has a potential therapeutic interest in cancer treatment and prevention of aging.

Methods: Molecular absorption and circular dichroism spectra were used to monitor thermal denaturation, acid base titration and mole ratio experiments. The resulting data were analyzed by multivariate data analysis methods. Surface plasmon resonance was also used to probe the kinetics and affinity of the DNA–drug interactions.

Results: We investigated the interaction between a G-quadruplex-forming sequence in the human *c-kit* proto-oncogene and the water soluble porphyrin TMPyP4. The role of cytosine and adenine residues at the loops of G-quadruplex was studied by substitution of these residues by thymidines.

Conclusions: Here, we show the existence of two binding modes between TMPyP4 and the considered G-quadruplex. The stronger binding mode (formation constant around 10^7) involves end-stacking, while the weaker binding mode (formation constant around 10^6) is probably due to external loop binding. Evidence for the release of TMPyP4 upon protonation of bases at the loops has been observed.

General significance: The results may be used for the design of porphyrin-based anti-cancer molecules with a higher affinity to G-quadruplex structures which may have anticancer properties.

Keywords: G-quadruplex, *c-kit*, Ligand, Conformational analysis, Multivariate analysis, Stacking interactions

1. Introduction

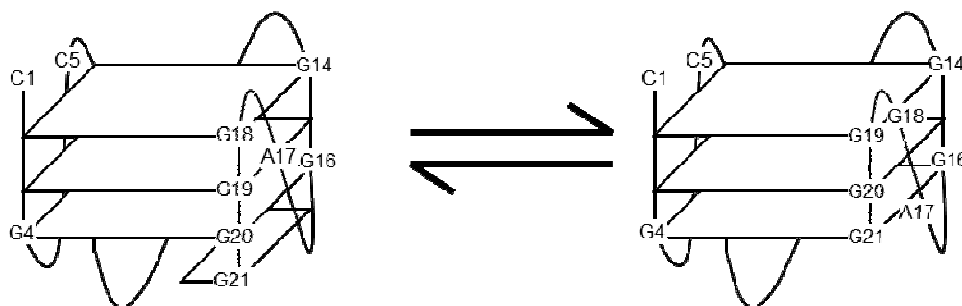
There is a large and growing interest in development and characterization of novel DNA binding drugs. Recently, this interest has focused on compounds that interact with nucleic acid sequences that adopt conformations other than B-DNA, *in vivo* [1]. Non-canonical higher order DNA conformations, such as G-quadruplex, might be formed transiently in small stretches of DNA sequences. G-quadruplex DNA is a very attractive target for highly selective, structure-specific therapeutic strategies, due to its significant structural difference from the double helix [2]. Nevertheless, there is little direct proof for the *in vivo* existence of G-quadruplex structures [3].

Potentially, G-quadruplex structures can be formed in hundreds of thousands of DNA sequences in the human genome [4]. G-rich sequences are particularly prominent in telomeres [5] and the promoter regions of several oncogenes such as *bcl2* [6], *k-ras* [7], *c-myc* [8] and *c-kit* [9,10]. *c-kit* is a human proto-oncogene coding for a 145–160 kDa membranebound glycoprotein [11]. Constitutive activation of the tyrosine kinase, the KIT receptor, is a central pathogenic event in most gastrointestinal stromal tumors.

The human *c-kit* oncogenic promoter contains two stretches of guanine-rich tracks, designated as c-kit1 (or *ckit81*) and c-kit2 (or *ckit21*). The c-kit2 sequence has been shown to form a parallel G-quadruplex in physiological conditions [9,12]. The multimeric nature of their folding still remains unclear. Recently, NMR-based studies showed that this sequence may adopt two distinct all-parallel-stranded conformations in a slow exchange, one of which forms a monomeric G-quadruplex (form-I) in 20 mM K⁺-containing solution and the other a novel dimeric G-quadruplex (form-II) in 100 mM K⁺-containing solution [13]. In both cases, the strand concentration of the samples was maintained between 0.2 and 4.0 mM. The c-kit2 promoter monomeric form-I G-quadruplex adopted an all-parallel topology with two single-nucleotide reversal loops (C5 and A17, respectively) and one long reversal loop (C9G10C11G12A13). The 3D structure showed these loops lying practically externally to the G-quadruplex core, where they are prone to interact with ligands such as porphyrins or protons. This intramolecular G-quadruplex structure had four 3'-end guanine bases. The guanine located at the 5'-end of this G-tract could assume either a loop or tetrad position leading to an intrinsic loop isomerism. Two energetically favorable loop isomers that are in dynamic equilibrium were proposed (Scheme 1). These are designated as a blunt-end conformer (1:5:2), in which only one loop is consisted of a single nucleotide, and a dangling-end conformer (1:5:1), in which two loops consisted of single nucleotides. In contrast, the c-kit2 promoter dimeric form-II G-quadruplex adopts an unprecedented all-parallel-stranded topology, in which individual c-kit2 promoter strands span a pair of three-G-tetrad-layer-containing, all-parallel-stranded G-quadruplexes aligned in a 3' to 5'-end orientation, with a stacking continuity between G-quadruplexes mediated by a sandwiched A·A non-canonical base pair.

Among small molecules, porphyrins have been shown to bind to and stabilize different types of G-quadruplexes and, in some cases, to facilitate G-quadruplex formation [7,14–19]. The cationic porphyrin mesotetrakis-(*N*-methyl-4-pyridyl)-porphyrin (TMPyP4) is the most extensively studied drug

[1,5,20]. Evaluation of the binding affinity and specificity of drugs for G-quadruplex DNA rely mainly on low resolution structural techniques such as UV-visible absorption spectroscopy and circular dichroism spectroscopy [21]. However, the measured signal contains a contribution from all the species present, which have different specificities and relative affinities to their ligands. It is presently unclear how conformational isomerism of G-quadruplex influences ligand binding. In this study, we employed a powerful method of multivariate curve resolution to extract previously unknown information about the effect of protonation in the loops of G-quadruplex structure on its affinity to porphyrin ligand TMPyP4. To our knowledge, the interaction between the intramolecular G-quadruplex formed by the *c-kit2* sequence and the porphyrin ligand TMPyP4 has not been established. Apart from the paper by Gunaratnam et al. [19], no data have been reported on the number of binding sites, the strength and the influence of pH on porphyrin binding. In the present study we also used biophysical tools to assess the interaction between TMPyP4 and the G-quadruplex-forming sequence in *c-kit*.



Scheme 1. Schematic representation of two loop isomers of *c-kit* G-quadruplex DNA. Dangling-end conformer (left) has one flanking nucleotide and one nucleotide in the third loop and blunt-end conformer (right) has two nucleotides in the third loop and no flanking nucleotide [28]. The white square represents a G-tetrad.

2. Materials and methods

2.1. Reagents

TMPyP4 was an analytical grade reagent purchased from Sigma Aldrich and was used without further purification. Stock solutions were prepared in MilliQ® water, stored at $-20\text{ }^{\circ}\text{C}$ and diluted to working concentrations in MilliQ water immediately before use. The composition of quadruplex stabilizing buffer (QSB) was 7 mM KH_2PO_4 , 10 mM Na_2HPO_4 , 147 mM KCl pH 7.2. Oligonucleotide 5'-d [CG3CG3CGCGAG3AG4]-3', designated as *ckit2*, was a G-quadruplex forming element in the promoter region of the human *c-kit* protooncogene. Oligonucleotide 5'-d [TG3TG3TGTGTG3TG4]-3', designated as *ckit2T*, was a mutated sequence in which adenine and cytosine bases of the wild type *ckit2* were systematically replaced with thymine (Table 1). Two additional sequences *ckit2T18* and *ckit2T21* were designed to eliminate the dangling-end to blunt-end equilibrium. These oligonucleotides, and the 3' biotin-labeled *ckit2* and *ckit2T*, were prepared as described elsewhere [22]. Oligonucleotide concentration was determined by UV absorbance measurements at 260 nm using calculated extinction coefficients approximated by the nearest-neighbor method.

2.2. Instrumentation

Absorbance and CD spectra were measured using an Agilent HP8453 photo diode array spectrophotometer and a Jasco J-810 spectropolarimeter, respectively. Both were equipped with a stirrer and Peltier units mounted in the spindle of the thermoelectric cuvette holders. Biosensor SPR experiments were performed with a four serial channel T100 (BIAcore, Inc.) optical biosensor system and streptavidin-coated sensor chips (Sensor chip SA; BIAcore, Inc.).

2.3. Procedures

Melting curves were collected on the J-810 instrument. Oligonucleotides were annealed and degassed by raising the temperature to 90 °C for 10 min and then cooling to 25 °C prior to the melting experiment. Samples were then heated at a linear temperature ramp of 0.3 °C/min with a data collection beginning at 25 °C and ending at 90 °C. The spectral bandwidth was 1 nm and the hold time was 1 min. A blank dataset taken from the buffer solution alone was also recorded and used for blank subtraction. Melting temperature (T_m) values are the average values of at least a pair of T_m values recorded during repeated melting experiments.

Acid–base titrations were performed for several mixtures of ckit2 and TMPyP4, and ckit2T and TMPyP4. The pH values were determined with an Orion SA 720 pH/ISE meter and a microcombination pH electrode (Thermo). The pH of the solutions containing each of the above mixtures was adjusted by adding small volumes of HCl or NaOH stock solutions. Following this, the absorbance spectra were recorded using an Agilent spectrophotometer pH stepwise.

To determine the equilibrium constants for the binding interactions, titration experiments were conducted using UV–vis, CD and SPR. In the UV–vis procedure, first the spectrum of pure TMPyP4 (or pure DNA) in QSB was recorded. This sample was then titrated by increasing the volumes of the DNA solution (or TMPyP4) in such a way that the resulting $C_{\text{TMPyP4}}:C_{\text{DNA}}$ (or $C_{\text{DNA}}:C_{\text{TMPyP4}}$) ratios were gradually lowered. Each titration was done at time intervals of 10 min with a magnetic stirrer on. Spectroscopic signals from the resulting mixture were recorded stepwise. Finally, the spectrum of a pure DNA (or TMPyP4) solution was recorded. In the CD procedure, first the CD spectrum of the pure DNA solution (ckit2 or ckit2T) in QSB was recorded. TMPyP4 was added to progressively increase the concentration of the ligand. The CD spectrum of the mixture was recorded stepwise. Finally, the CD spectrum of the pure porphyrin drug was recorded. For the SPR measurements, the streptavidin sensor chip surface was prepared for DNA immobilization with a regeneration buffer (1 M NaCl in 50 mM NaOH). It was then extensively washed with a running buffer. The running buffer was sterile filtered and degassed HBS-EP buffer (0.01 M HEPES (pH 7.4), 3 mM EDTA, and 0.005% surfactant Tween 20 with 0.1 M NaCl) fortified with 150 mM of KCl at pH 7.2. The 21mer 3'-biotinylated ckit2 and ckit2T in running buffer were immobilized on flow cells 2 and 4 respectively. The 3' biotinylated ckit2 with a poly(A) linker was immobilized on flow cell 3 to check the potential influence of the linker on binding. Flow cell 1 was coupled with biocytin and left as a blank control to characterize any non-specific binding. After immobilization, the structures were equilibrated by passing a running buffer over the surface for 1 h at 60 $\mu\text{l}/\text{min}$ flow rate to

attain base line stability. Plastic vials were randomly positioned in the rack to minimize systematic errors. Following this, dissociation from the surface was monitored for 60 s in a running buffer. The concentration ranges for TMPyP4 were 0, 0.01–0.08, 0.1 and 0.2 μM . Two samples (0, 0.06 μM) were replicated. Suitable blank control injections (zero concentrations) with a running buffer were performed.

Table 1 Sequences of *c-kit* loop variants used in this study.

Name	5'		L1		L2		L3		3'
ckit2	C	GGG	C	GGG	CGCGA	GGG	A	GGG	G
ckit2T	T	GGG	T	GGG	TGTGT	GGG	T	GGG	G
ckit2T18	C	GGG	C	GGG	CGCGA	GGG	AT	GGG	
ckit2T21	C	GGG	C	GGG	CGCGA	GGG	A	GGG	T

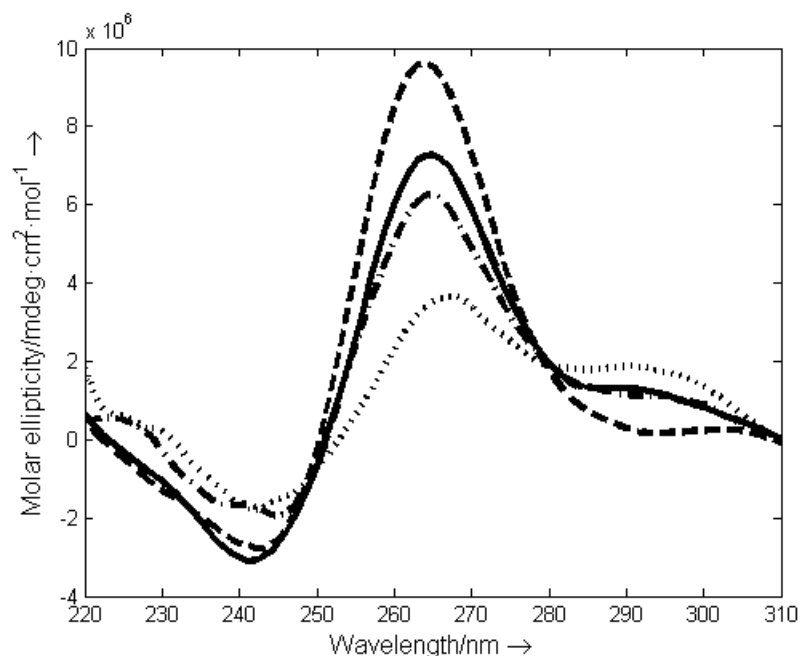


Fig. 1. CD spectra of ckit2 and mutants. CD spectrum of ckit2 (continuous black line), ckit2T (dashed red line), ckit2T18 (dotted blue line) and ckit2T21 (dash-dot green line). $T=25\text{ }^{\circ}\text{C}$, $\text{pH}=7.1$. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

2.4. Data analysis

In this study, the hard-modeling analyses of acid–base and mole-ratio experiments were carried out using the EQUISPEC program [23]. The soft-modeling analysis was undertaken with the graphic user interface of MCR-ALS [24], which is freely available at www.mcrals.info. As the details of the multivariate analysis have been explained extensively elsewhere [25–27], only a short description is given here as Supplementary material.

3. Results

Table 1 shows the DNA sequences studied here. In previous works, the solution equilibria of the ckit2 sequence were studied [9,26]. Spectroscopically-monitored melting experiments provided a value of T_m (74 ± 1 °C), which was shown to be independent of concentrations over the range 0.5 to 5 μ M. This suggests that an intramolecular folding has occurred. Acid–base titration experiments showed that the G-quadruplex structure was well maintained over the pH range from 3 to 7, approximately. In this pH range, only one macroscopic pH-induced transition at around pH 4.4 was observed. This was attributed to the protonation of adenine and cytosine bases located at the loops. Simultaneously, we studied the solution equilibria of the ckit2T sequence, in which all bases at the loops have been mutated to thymine. The T_m value determined at pH 7 was slightly higher than that for ckit2 (78 ± 1 °C). Acid–base titrations displayed no transition over the pH range of 3 to 7 because of the lack of protonable nucleobases.

3.1. Conformational equilibrium for ckit2 is shifted to the dangling-end conformer

Fig. 1 shows the CD spectra for ckit2 and for the mutated sequences ckit2T18 and ckit2T21. The mutation of a guanine by a thymine in the last two sequences reduces strongly the potential conformational heterogeneity in ckit2 by constraining loop isomerism. The CD spectrum of G-quadruplex is known to demonstrate the nature of the folding (parallel, antiparallel or mixed), depending on the shape and intensity of the bands at 260 and 295 nm. The CD spectrum of the ckit2T18 sequence, the third loop of which contains two bases, shows a relatively high intensity of the band at 295 nm. This indicates a significant contribution from an antiparallel conformation. In contrast, the spectrum of ckit2T21, the third loop of which contains only a base, shows a predominant contribution of the parallel conformation (intense band at 260 nm and small shoulder at 295 nm). The CD spectrum for the wild type ckit2 sequence is very similar to that displayed by ckit2T21, which points out to a similar spatial arrangement. This fact, as well as the previously shown superimposable cross peaks in the NOE spectra [28], indicates that the dangling-end to blunt-end equilibrium is shifted to the former one for the ckit2 sequence. Finally, the ckit2T sequence shows clearly a parallel structure, without any significant contribution at 295 nm.

3.2. TMPyP4 interacts with ckit2 and ckit2T at pH 7

The interaction of the ligand TMPyP4 with the G-quadruplex formed by ckit2 and ckit2T was studied by using molecular absorption and CD spectroscopies at pH 7 and 25 °C. Fig. 2a shows changes observed in the absorption band of the ligand upon the addition of the c-kit2 sequence.

After the addition of 3 equivalents of DNA, a bathochromic shift of up to 18 nm and 70% hypochromicity at 422 nm were observed. These features imply a relatively strong binding. Fig. 2b shows the CD spectra of ckit2 and of a mixture of TMPyP4 and ckit2. The shape of the CD spectrum in the UV region is related to a parallel G-quadruplex topology [9,29,30]. Upon an interaction with the ligand, the shape of the CD spectrum remains unaltered, and only small variations in CD intensity were observed at 240, 260 and 295 nm. A weak negative induced CD signal was observed around 440 nm, indicating slight changes in the chiral environment of the flat drug, probably due to stacking [31]. Since absorbance occurs at the induced CD wavelength, the signal to noise ratio is compromised.

A different spectral behavior was observed for the mutated sequence ckit2T. Molecular absorption spectra (Fig. 2c) clearly showed the interaction between the DNA and the ligand. A shift of up to 16 nm and 66% hypochromicity at 422 nm were observed after the addition of 3 equivalents of DNA. The results suggest that the binding of TMPyP4 to ckit2T is slightly weaker than ckit2. The most surprising fact is the absence of any induced CD band in the visible region of the spectrum (Fig. 2d), which suggests a different interaction mechanism.

3.3. *Melting experiments*

Thermal melting studies of mixtures of TMPyP4 with ckit2 and ckit2T were studied by CD and molecular absorption spectroscopy. Raw CD spectra are included as a Supplementary material (Fig. S1). The induced band at 440 nm disappeared when the temperature was raised, which suggests that the ligand is released from the G-quadruplex structure. Concomitantly, the ellipticity at 260 nm was dramatically reduced whereas the overall spectral shape remained unaltered. These observations were probably related with the fact that at 90 °C there was still a small portion of folded DNA ($T_m=74\pm 1$ °C). Fig. 3 shows the variation in CD intensity at 263 nm upon heating for the isolated strands and for the corresponding mixtures with TMPyP4. For both sequences, an interaction with TMPyP4 produced a shift in the measured CD signal at 263 nm, which indicates preferential binding with the folded structure over the unfolded strand. For ckit2 the shift was around 10 °C, whereas a larger shift for ckit2T can only be guessed. The shift reported in the case of ckit2 was clearly lower than that determined previously using FRET [19] (21 °C). The difference could be due to the slightly different experimental conditions used and the fact that the studied sequence was shorter (21 nucleotides against 26), and lacked terminal fluorescent labels.

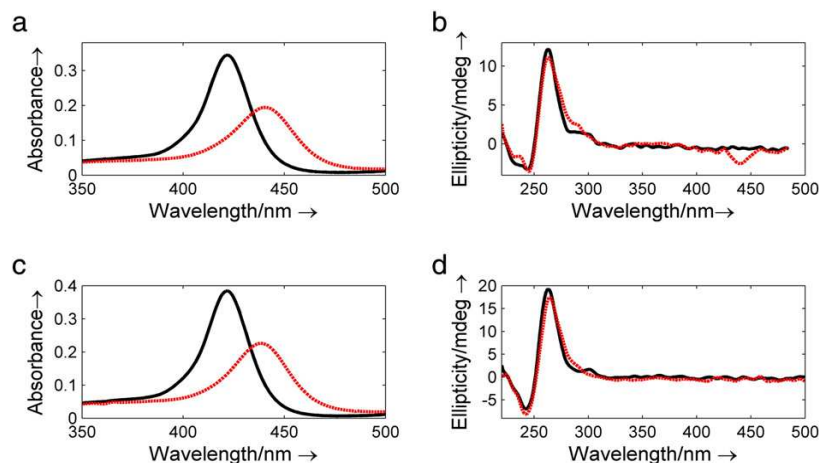


Fig. 2. Interaction of TMPyP4 with ckit2 and ckit2T. (a) Visible absorption spectra of a 1.6 μM solution of TMPyP4 (solid black line) and of a mixture of TMPyP4 (1.6 μM) and ckit2 (4.8 μM). (b) CD spectra of ckit2 (3 μM , solid black line) and of a mixture of ckit2 and TMPyP4 (3 and 9 μM , respectively). (c) Visible absorption spectra of a 1.7 μM solution of TMPyP4 (solid black line) and of a mixture of TMPyP4 (1.6 μM) and ckit2T (5.0 μM). (d) CD spectra of ckit2T (2.6 μM , solid black line) and of a mixture of ckit2T and TMPyP4 (2.5 and 7.4 μM , respectively). $T=25\text{ }^\circ\text{C}$, $\text{pH}=7.1$. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

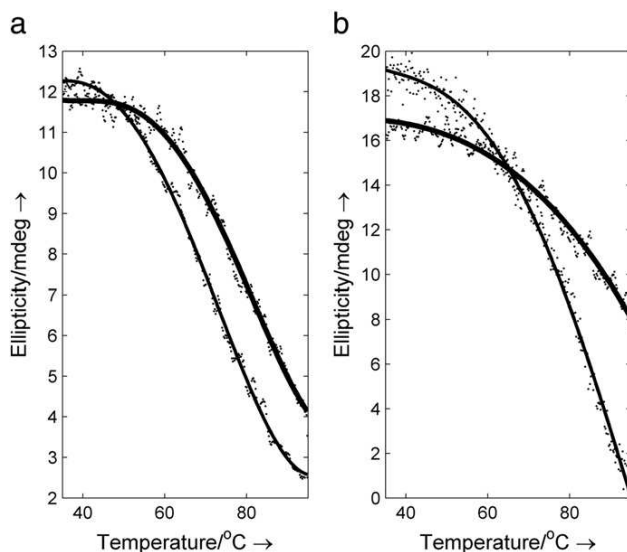


Fig. 3. CD-monitored melting experiments. (a) Variation of CD ellipticity at 263 nm upon heating for ckit2 (2.4 μM , thin line) and for a mixture of ckit2 and TMPyP4 (2.4 μM and 7.4 μM , thick line). (b) Variation of CD ellipticity at 263 nm for ckit2T (2.5 μM , thin line), and for a mixture of ckit2T and TMPyP4 (2.5 and 7.4 μM , thick line). $\text{pH}=7.1$.

3.4. Spectroscopically-determined binding constants

Spectroscopic mole-ratio direct and inverse titrations were performed. The recorded spectra were initially arranged in matrix D and later analyzed with the Equispec hard-modeling method. This procedure determines the binding constant and the pure spectra of DNA:ligand species

formed during titration according to a defined stoichiometry. As an example, Fig. 4 shows the results obtained after Equispec analysis of a set of molecular absorption spectra measured along the titration of a ckit2 sample with TMPyP4. The recorded spectra have been included as a Supplementary material (Fig. S2). In this case, the best fit was a model with a strong binding site and two equivalent weaker binding sites. The simplest 1:1 (DNA:ligand) model did not provide a good fit at higher ligand concentrations. The calculated binding constants are summarized in Table 2, and the calculated distribution diagram for this mole-ratio experiment is shown in Fig. 4a. In addition, the calculated pure spectra for each one of the species involved in the equilibria are shown in Fig. 4b. The interaction of TMPyP4 with ckit2 produced up to 68% hypochromicity, along with a red shift of up to 19 nm in the calculated spectrum for the 1:3 species. Mole-ratio titrations monitored with circular dichroism spectroscopy were also performed (Fig. S3). Upon addition of the ligand, the intensity of the CD band at 245 and 263 nm decreased significantly, and a small induced CD band appeared around 440 nm. In order to confirm that this signal was not an artifact due to the high absorption of TMPyP4 in this region, additional titrations were performed at lower TMPyP4 concentrations. The appearance of this signal even at low concentrations of TMPyP4 was confirmed. The spectra shown in Fig. S3 were deconvoluted using the set of stability constants calculated from the mole-ratio experiments monitored with molecular absorption spectroscopy. The calculated pure CD spectrum for each species is shown in Fig. 4c. Overall, the CD signature in the UV region denotes that the parallel structure was maintained for the two DNA:ligand interaction species. However, the 1:3 species showed a lower molecular ellipticity and the induced CD signal appeared in the visible region.

Moreover, the interaction of TMPyP4 with the mutated sequence ckit2T has been studied (Fig. S4 and S5). The interaction of TMPyP4 with ckit2T gave a slightly different result (Fig. 5 and Table 2). The model that best fits the data included a strong binding site and one weak binding site. The values of the equilibrium constants for the weaker binding sites were similar to those determined for ckit2. In contrast, the equilibrium constant for the stronger binding site was slightly lower than that in the case of ckit2, a fact which is correlated with the lower shift and hypochromicity values observed experimentally.

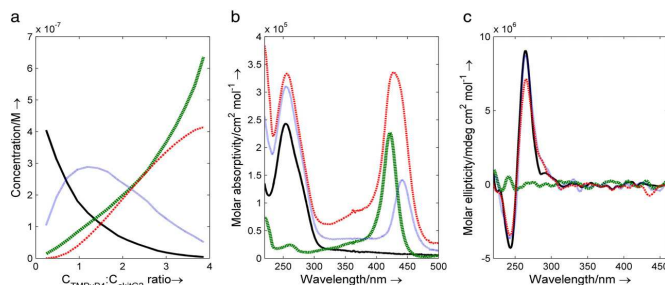


Fig. 4. Spectroscopically-determined binding constants for ckit2 sequence. (a) Distribution diagram calculated with Equispec. (b) Calculated pure molecular absorption spectra. (c) Calculated pure CD spectra. Solid black line: G-quadruplex ckit2; dotted blue line: 1:1 (DNA:ligand) complex; dash-dotted red line: 1:3 complex; dashed green line: TMPyP4. T=25 °C, pH=7.1. (For interpretation of the

references to color in this figure, the reader is referred to the web version of this article.)

3.5. Surface plasmon resonance-determined binding constants

SPR was used to obtain complementary information on the binding of ligands to the aforementioned G-quadruplex structures considered in this study. Fig. 6 shows the results obtained both in the steady-state and kinetic analysis. To a first approximation, the binding curves were fitted using the 1:1 (DNA:ligand) model. The equilibrium formation constants calculated for the interaction of TMPyP4 with ckit2 and ckit2T were $10^{7.1}$ and $10^{6.9}$, respectively (Supplementary material). These fits were improved when the two-site model was considered. Hence, the formation constants for ckit2 were $10^{8.0}$ and $10^{6.7}$, whereas the calculated formation constants for ckit2T were equal ($10^{8.0}$ and $10^{6.6}$).

The corresponding rate and equilibrium constants were also provided by fitting the sensorgrams to complex kinetic models. As for the steady-state studies, the best results were obtained when the two-site model was used. This suggests that TMPyP4 had a stronger binding site in the order of 10^8 on the G-quadruplex along with a weaker secondary binding interaction in the order of 10^6 (Table 3). Qualitatively, this result agrees with those previously obtained from spectroscopic measurements.

Table 2 Binding parameters obtained from UV-visible titration experiments.

Sequence	pH	% Hypochromicity at 422 nm	shift of the Soret band	Log $K_{a1}(M^{-1})$	n_1	Log $K_{a2}(M^{-1})$	n_2
ckit2	7.1	70	18	7.4(0.2)	1	6.7(0.3)	2
	3.0	69	18	7.6(0.1)	1	5.8(0.2)	1
ckit2T	7.1	66	16	7.1(0.1)	1	5.7(0.1)	1
	3.0	65	16	7.2(0.3)	1	6.0(0.4)	1

3.6. pH-induced binding changes

Mole-ratio experiments provide quite a reliable picture of the situation at a fixed pH value. To gain information about the influence of pH (and the consequent protonation of functional groups in the nitrogenated bases), acid-base titrations of the mixtures of ckit2 (or ckit2T) with TMPyP4 in the pH range of 7 to 1 were investigated spectroscopically. Firstly, the acid-base behavior of TMPyP4 was studied in the mentioned pH range (see Supplementary material, Fig. S7). A shift of the Soret band from 422 nm (pH 7) to 448 nm (at a pH of around 1.5) was observed. The experimental spectra were arranged in data matrix D and analyzed by means of Equispec. A simple acid-base equilibrium model was proposed and the calculated pK_a for the proposed equilibrium was 1.7 ± 0.1 at 25 °C and 150 mM ionic strength. However, the accurate determination of this pK_a is difficult because of the concomitant increase in ionic strength when the pH is lowered to values below 2, approximately. The calculated pK_a value agrees with other values reported previously [32]. The result was explained in terms of the protonation of nitrogen atoms located in the porphyrin moiety.

The UV-visible molecular absorption spectra recorded during the titration of a mixture of ckit2 and TMPyP4 (1:4) at 25 °C are shown in the

Supplementary material (Fig. S8). According to the calculated equilibrium constants at pH 7, at this ratio of concentrations almost all binding sites in ckitG2 are occupied by ligand molecules. The Soret band was initially located around 427 nm at pH 7. Based on the results obtained from the mole-ratio studies at pH 7, this indicates binding. The observed difference between 441 nm (the calculated wavelength for the fully complexed ligand) and 427 nm (the experimental value) was due to the presence of an excess of free, unbounded drug. At pH under 6, the Soret band shifts to shorter wavelengths, which is indicative of the release of free TMPyP4 from the complex. At pH 3, the band is centered at 422 nm, which is the position of free deprotonated TMPyP4. At even lower pH values, the band moves to longer wavelengths, according to the observed trend for protonated TMPyP4. Experimental spectra were arranged in data matrix D and analyzed by means of MCR-ALS, a soft-modeling approach. The number and complexity of the chemical equilibria in the mixture made it difficult to apply a hard-modeling approach like Equispec. The resolved distribution diagram and pure spectra, which are shown in Fig. 7a and b, reflect the experimentally observed facts.

Hence, the data were well explained when only three species were considered. The first species is related to the initial mixture of the 1:3 c-kit2: TMPyP4 complex and unbounded TMPyP4 at pH 7. The second species is the major one at pH 3. The midpoint of this transition is located around pH 4.4, which is also the location of the only transition induced by the pH observed for free c-kit2 [26]. This has been related to the protonation of bases at the loops. Accordingly, the major species at pH 3 have been related to a mixture of c-kit2 (where bases at the loops are protonated) and free deprotonated TMPyP4. The additional positive charge at the loops removes the ligand from the G-quadruplex. At even lower pH values, the protonation of TMPyP4 is observed (pH transition midpoint around 2) and the Soret band moves to 443 nm.

The role of the protonation of the cytosine bases at the loops was checked by carrying out a similar acid–base titration of a mixture of TMPyP4 and ckit2T. This sequence lacks cytosine and adenine bases at the loops, whereas the G-quadruplex structure is still well maintained [26]. The acid–base titration of the ckit2T: TMPyP4 (1:2.5) mixture did not show any spectral change over the pH range 2–7, and a shift from 428 to 443 nm was only observed at pH below 2, concomitant with the protonation of TMPyP4 (see the Supporting material, Fig. S9). The MCR-ALS analysis indicated that the experimental data could be explained satisfactorily with only one acid–base transition (Fig. 7c–d). In this case, the transition was located around 1.8 and was explained in terms of the protonation of TMPyP4.

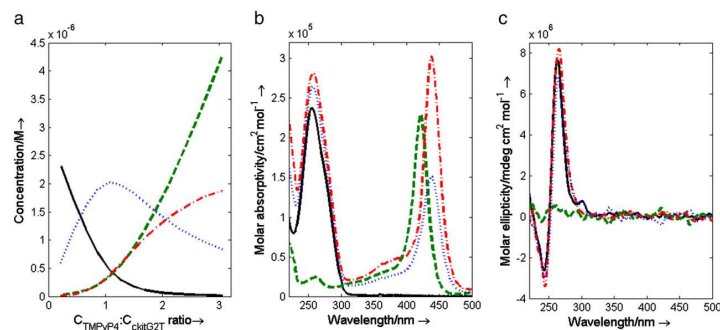


Fig. 5. Spectroscopically-determined binding constants for ckit2T sequence. (a) Distribution diagram calculated with Equispec. (b) Calculated pure molecular absorption spectra. (c) Calculated pure CD spectra. Solid black line: G-quadruplex ckit2T; dotted blue line: 1:1 (DNA:ligand) complex; dash-dotted red line: 1:2 complex; dashed green line: TMPyP4. T=25 °C, pH=7.1. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

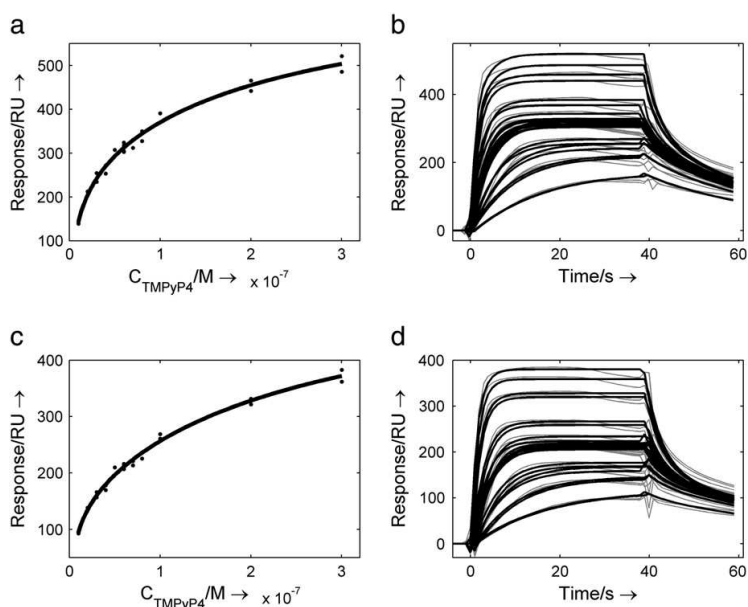


Fig. 6. SPR results for the interaction of TMPyP4 with ckit2 and ckit2T. (a) and (c) show the steady state affinity (1:2 model) plot for the interaction of TMPyP4 with ckit2 and ckit2T, respectively. (b) and (d) show the SPR sensorgrams (black) with fitting (grey) for the interaction of TMPyP4 with ckit2 and ckit2T, respectively, when the heterogeneous ligand model was considered. The concentration of TMPyP4 increased from $1 \cdot 10^{-8}$ M to $3 \cdot 10^{-7}$ (upper curve). The experiments were carried out in 10 mM HEPES buffer, 3 mM EDTA, 0.005% surfactant Tween 20, 150 mM KCl, pH 7.2, at 25 °C.

Table 3 Thermodynamic and kinetic parameters obtained from SPR. Results obtained from simultaneous analyses of two independent experiments. pH 7.1, 25 °C.

Sequence	Affinity (Two sites)		Kinetics (Heterogeneous ligand)					
	log Ka1	log Ka2	ka1 (1/Ms)	kd1 (1/s)	log Ka1 (M-1)	ka2 (1/Ms)	kd2 (1/s)	log Ka2 (M-1)
ckit2	8.0	6.7	$5.54 \cdot 10^{-6}$	0.0401	8.1	$1.15 \cdot 10^6$	0.3335	5.5
ckit2T	8.0	6.6	$5.59 \cdot 10^6$	0.0272	8.3	$1.17 \cdot 10^6$	0.3223	6.5

3.7. Mole-ratio studies at pH 3.5

Mole-ratio experiments were performed at pH 3.5, where most of the cytosine and adenine bases present at the loops of ckit2 were already protonated (Fig. S10). The experiments carried out for ckit2 showed the formation of two complexes with stoichiometries 1:1 and 1:2 (DNA:ligand), and with equilibrium constants equal to $10^{6.8}$ and $10^{6.3}$ i.e., two different binding sites were considered. The CD spectra measured for a mixture of ckit2 and drug showed the negative induced CD band around 440 nm observed at pH 7 (Fig. S11) disappeared.

In the case of ckit2T (Fig. S12), the results of mole-ratio experiments at pH 3.5 were very similar to those obtained previously at pH 7.1. The analysis of the data (Supplementary material) was successful when two complexes with stoichiometries of 1:1 and 1:2 (DNA:ligand) were used, with equilibrium constants equal to $10^{7.2}$ and $10^{6.0}$ i.e., two different binding sites were considered. The spectral shift, hypochromicity and binding constant values were similar to those determined at pH 7.1, which indicates that the binding mode was similar.

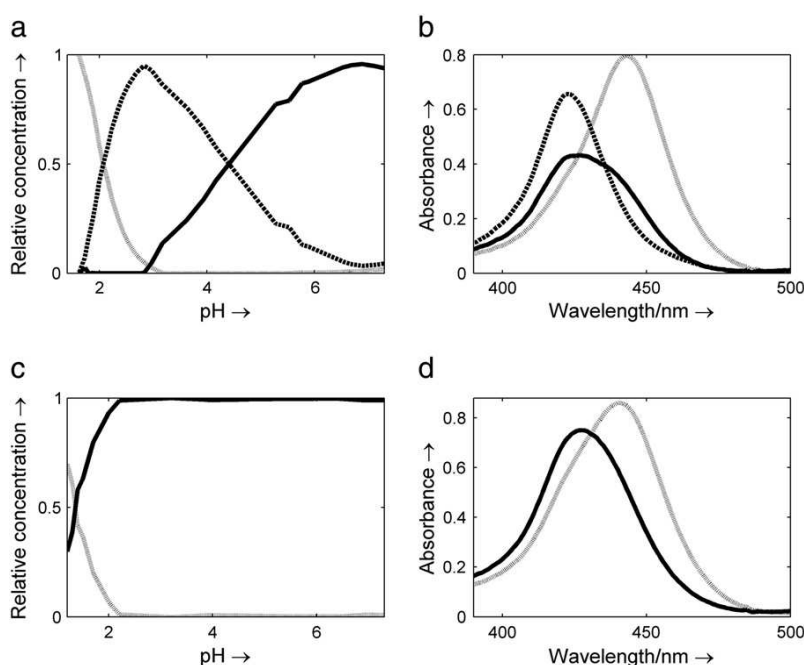


Fig. 7. Spectroscopically monitored acid–base titration of mixtures of TMPyP4 and ckit2, and TMPyP4 and ckit2T. Calculated distribution diagram (a) and pure molecular absorption spectra (b) for the mixture of TMPyP4 and ckit2. Continuous black line: mixture of the 1:3 complex and free unbounded TMPyP4. Dashed line: mixture of ckit2 (with protonated cytosines) and deprotonated TMPyP4. Dotted gray line: mixture of ckit2 (with protonated cytosines) and protonated TMPyP4. Calculated distribution diagram (c) and pure molecular absorption spectra (d) for the mixture of TMPyP4 and ckit2T. Continuous black line: mixture of the 1:2 complex and free unbounded TMPyP4. Dotted gray line: mixture of ckit2T and protonated TMPyP4.

4. Discussion

Spectroscopic techniques along with multivariate data analysis may become useful tools to generate plausible binding mechanisms. The present study examines one such case, by proposing a drug binding mechanism for a complex system involving G-quadruplex DNA. Multivariate data analysis was used to detect small absorbance and ellipticity changes manifested in response to subtle structural rearrangements during mole-ratio and acid-base titration experiments. The multivariate curve resolution technique added two important values to the spectroscopic data. First, it helped to extract new structural variants that would otherwise be indistinguishable in these spectroscopic techniques. Second, it helped to estimate thermodynamic parameters from the distribution diagrams. The combination of these tools was used to determine the affinity, specificity and binding mechanisms of the DNA drug interactions.

Based on NMR data, Hsu et al. proposed that the ckit2 sequence exists as an ensemble of at least two structures that share the same parallel-stranded propeller-type conformations, and which undergo slow interconversion on the NMR timescale (Scheme 1) [28]. In spite of the changes at the loops, the overall parallel structure of ckit2 seems to be conserved in the mutated sequence ckit2T, as the CD spectrum and T_m value are similar. The small difference observed (74 versus 78 °C) may be due to the substitution of adenine bases at the loops, whose destabilizing effect in G-quadruplex structures has been suggested previously [33].

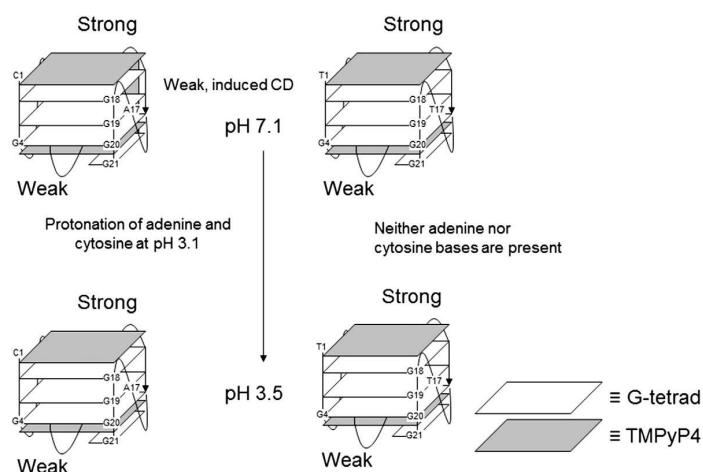
At pH 7.1, our results suggest the existence of three binding sites for ckit2 in the presence of a moderate excess of TMPyP4. One of these binding sites shows a stronger affinity (around 10^7) than the other two (around 106). This is in accordance with the previous models proposed by other authors. A 1:3 (DNA:ligand) stoichiometry was proposed for the interaction of TMPyP4 with the parallel intramolecular G-quadruplex formed by a sequence corresponding to the NHE III1 element of the c-myc oncogene [34], the structure of which is similar to that proposed for c-kit2 [35]. These authors proposed that the interaction of TMPyP4 with parallel quadruplexes takes place through three binding sites, one with a stronger binding (10^7) involving end-stacking over the end G-tetrad, and two binding sites with a weaker external binding (10^6). Other authors have proposed similar stoichiometries and binding modes. Wei et al. [36] proposed that three TMPyP4 molecules bind to the parallel four-stranded $(TG_4T)_4$ DNA with two independent binding sites of similar affinity to those determined in our study. Interestingly, the binding stoichiometry seems to be different for a similar porphyrin (TPrPyP4) in a Na^+ -containing medium [37].

Besides the existence of two binding sites with a different affinity, the number of TMPyP4 molecules bound to parallel G-quadruplex DNA depends on the sequence considered. Hence, Han et al. suggested that TMPyP4 binds to parallel G-quadruplex through external stacking at the ends with two molecules of ligand bound per quadruplex at saturation [15]. Parkinson et al. [38] reported the crystal structure of a DNA:TMPyP4 complex involving the bimolecular parallel G-quadruplex formed by a sequence corresponding to the human telomere. The particular sequence studied in this paper has several adenine and thymine bases at the loops. Consequently, TMPyP4 molecules did not stack directly onto the G-tetrad core but onto A·T base pairs and TT and TTA loops. The authors reported two main binding sites in the experimental conditions used in their study. Despite the fact that the

conformation is similar to that shown by ckit2 (parallel), the role of bases in the loops seems to be crucial.

It should be stressed that the methodology used in this study (spectroscopy hyphenated to multivariate analysis) could not deconvolute all three DNA:ligand complexes. The pure spectra and the formation constant for the 1:2 (ckit2:TMPyP4) complex could not be resolved. This fact is related to the equal spectral characteristics of 1:2 and 1:3 complexes, which hinder the mathematical resolution of the chemical system. It may be hypothesized that the similar spectral features of both complexes are due to the binding of TMPyP4 to very similar sites on the G-quadruplex. Another previously proposed possibility would be the existence of two sequential binding events, a first in which one molecule of TMPyP4 interacts with the quadruplex structures and a second in which more molecules bind to the structure [5].

At pH 7.1, the interaction of TMPyP4 with ckit2 led to the appearance of a weak negative induced CD band. A similar finding was reported for the interaction of TMPyP4 with a parallel G-quadruplex within the promoter region of the *k-ras* gene [7] or with a hybrid antiparallel/ parallel G-quadruplex within the promoter region of the *bcl-2* gene [27]. A comparison of the intensity of the induced CD signal with those previously reported for a hairpin formed by a cytosine-rich sequence of the *bcl-2* gene [25] or for duplex DNA [39] indicates that intercalation produces stronger induced CD signals than end-stacking. Multivariate analysis of spectroscopic data revealed that this induced CD band is characteristic of the 1:3 complex. In contrast, the 1:1 complex did not show this band. Similarly, the shift of the absorption band in the Soret region for the 1:3 complex is small compared to the 1:1 complex, a fact that is related to a weaker binding. At pH 7, the interaction of TMPyP4 with the mutated sequence ckit2T is slightly different to that observed for ckit2 as only one weak binding site was detected and no induced CD band appeared.



Scheme 2. Proposed mechanism for the interaction of TMPyP4 with ckit2 (left) and ckit2T (right).

Acid–base titrations showed that just one of the binding sites in ckit2 is dramatically modified by the protonation of bases such as cytosine and adenine, whose pKa values are around 4. As expected, decreasing the pH value from 7 to 3 did not affect the binding mechanism of ckit2T, as it lacks

cytosine and adenine bases. This site in ckit2 would be responsible for the negative induced CD band.

Scheme 2 shows the proposed mechanism for the interaction of TMPyP4 with ckit2 and ckit2T under the experimental conditions. The existence of three binding sites in ckit2 is proposed. The strongest site corresponds to the end-stacking binding onto a G-tetrad. The location on the upper or bottom G-tetrad is still a matter of discussion. Whereas the 3'-end seems the most appropriate place for an intramolecular parallel structure [34], the 5'-end seems to be preferred for a bimolecular parallel structure [38]. Alternatively, two TMPyP4 molecules may stack onto both external G-tetrads, and one of the binding sites is stabilized additionally by means of an electrostatic interaction with bases of the larger loop [7]. This explanation agrees with our results and corresponds to the binding sites depicted in Scheme 2. One of the weak binding sites has been placed near a loop, and it has been shown that the protonation of bases affects this site. The location of the third binding site is quite ambiguous. Previous literature proposing the existence of two weak binding sites locates one of the weak binding sites near the external grooves. The location of the third site is near the bottom G-tetrad [34] or near the external groove [36]. However, in the case of ckit2, the role of the terminal guanine base, linked to the dangling-blunt-end conformer equilibria, cannot be discarded.

These proposed mechanisms should be further refined on the basis of additional structural pictures derived from NMR and X-ray studies. The extraction of a mechanistic conclusion is always associated with some uncertainties and is by no means definitive. Here a compelling mechanism is postulated that is open to rigorous theoretical and experimental tests.

Acknowledgments

This research was supported by the Spanish Ministerio de Ciencia e Innovación (grant numbers CTQ2009-11572 and CTQ2010-20541-C03-01), and the Generalitat de Catalunya (grant numbers 2009-SGR-45 and 2009-SGR-208).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.bbagen.2012.09.006>.

References

- [1] J. Ren, J.B. Chaires, Sequence and structural selectivity of nucleic acid binding ligands, *Biochemistry* 38 (1999) 16067.
- [2] N.W. Luedtke, Targeting g-quadruplex DNA with small molecules, *Chimia* 63 (2009) 134–139.
- [3] A.N. Lane, J.B. Chaires, R.D. Gray, J.O. Trent, Stability and kinetics of G-quadruplex structures, *Nucleic Acids Res.* 36 (2008) 5482–5515.
- [4] H.J. Lipps, D. Rhodes, G-quadruplex structures: in vivo evidence and function, *Trends Cell Biol.* 19 (2009) 414–422.
- [5] L. Martino, B. Pagano, I. Fotticchia, S. Neidle, C. Giancola, Shedding light on the interaction between TMPyP4 and human telomeric quadruplexes, *J. Phys. Chem. B* 113 (2009) 14779–14786.
- [6] J.X. Dai, T.S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R.A. Jones, D.Z. Yang, An intramolecular G-quadruplex structure with mixed

parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution, *J. Am. Chem. Soc.* 128 (2006) 1096–1098.

[7] S. Cogoi, L.E. Xodo, G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription, *Nucleic Acids Res.* 34 (2006) 2536–2549.

[8] A. Siddiqui-Jain, C.L. Grand, D.J. Bearss, L.H. Hurley, Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription, *Proc. Natl. Acad. Sci. U.S.A.* 99 (2002) 11593–11598.

[9] H. Fernando, A.P. Reszka, J. Huppert, S. Ladame, S. Rankin, A.R. Venkitaraman, S. Neidle, S. Balasubramanian, A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene, *Biochemistry* 45 (2006) 7854–7860.

[10] S. Rankin, A.P. Reszka, J. Huppert, M. Zloh, G.N. Parkinson, A.K. Todd, S. Ladame, S. Balasubramanian, S. Neidle, Putative DNA quadruplex formation within the human c-kit oncogene, *J. Am. Chem. Soc.* 127 (2005) 10584–10589.

[11] S. Hirota, K. Isozaki, Y. Moriyama, K. Hashimoto, T. Nishida, S. Ishiguro, K. Kawano, M. Hanada, A. Kurata, M. Takeda, G. Muhammad Tunio, Y. Matsuzawa, Y. Kanakura, Y. Shinomura, Y. Kitamura, Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors, *Science* 279 (1998) 577–580.

[12] P.S. Shirude, B. Okumus, L. Ying, T. Ha, S. Balasubramanian, Single-molecule conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene c-kit, *J. Am. Chem. Soc.* 129 (2007) 7484–7485.

[13] V. Kuryavyi, A.T.n. Phan, D.J. Patel, Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter, *Nucleic Acids Res.* 38 (2010) 6757–6773.

[14] I. Haq, J.O. Trent, B.Z. Chowdhry, T.C. Jenkins, Intercalative G-tetraplex stabilization of telomeric DNA by a cationic porphyrin1, *J. Am. Chem. Soc.* 121 (1999) 1768–1779.

[15] H.Y. Han, D.R. Langley, A. Rangan, L.H. Hurley, Selective interactions of cationic porphyrins with G-quadruplex structures, *J. Am. Chem. Soc.* 123 (2001) 8902–8913.

[16] M.-Y. Kim, M. Gleason-Guzman, E. Izbicka, D. Nishioka, L.H. Hurley, The different biological effects of telomestatin and TMPyP4 can be attributed to their selectivity for interaction with intramolecular or intermolecular G-quadruplex structures, *Cancer Res.* 63 (2003) 3247–3256.

[17] V.r. Gabelica, E. Shammel Baker, M.-P. Teulade-Fichou, E. De Pauw, M.T. Bowers, Stabilization and structure of telomeric and c-myc region intramolecular G-quadruplexes: the role of central cations and small planar ligands, *J. Am. Chem. Soc.* 129 (2007) 895–904.

[18] C. Romera, O. Bombarde, R. Bonnet, D. Gomez, P. Dumy, P. Calsou, J.-F. Gwan, J.-H. Lin, E. Defrancq, G. Pratviel, Improvement of porphyrins for G-quadruplex DNA targeting, *Biochimie* 93 (2011) 1310–1317.

[19] M. Gunaratnam, S. Swank, S.M. Haider, K. Galesa, A.P. Reszka, M. Beltran, F. Cuenca, J.A. Fletcher, S. Neidle, Targeting human gastrointestinal stromal tumor cells with a quadruplex-binding small molecule, *J. Med. Chem.* 52 (2009) 3774–3783.

[20] D.Monchaud, A. Granzham, N. Saettel, A. Guedin, J.-L.Mergny, M.-P. Teulade-Fichou, One ring to bind them all — part I: the efficiency of the

macrocyclic scaffold for G-quadruplex DNA recognition, *J. Nucleic Acids* 2010 (2010) 525682.

[21] J. Jaumot, R. Gargallo, Experimental methods for studying the interactions between G-quadruplex structures and ligands, *Curr. Pharm. Des.* 2012 (2012) 1900–1916.

[22] M. del Toro, R. Gargallo, R. Eritja, J. Jaumot, Study of the interaction between the G-quadruplex-forming thrombin-binding aptamer and the porphyrin 5,10,15,20-tetrakis-(N-methyl-4-pyridyl)-21,23H-porphyrin tetratosylate, *Anal. Biochem.* 379 (2008) 8–15.

[23] R.M. Dyson, S. Kaderli, G.A. Lawrance, M. Maeder, Second order global analysis: the evaluation of series of spectrophotometric titrations for improved determination of equilibrium constants, *Anal. Chim. Acta* 353 (1997) 381–393.

[24] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, *Chemom. Intell. Lab. Syst.* 76 (2005) 101–110.

[25] N. Khan, A. Avino, R. Tauler, C. Gonzalez, R. Eritja, R. Gargallo, Solution equilibria of the i-motif-forming region upstream of the B-cell lymphoma-2 P1 promoter, *Biochimie* 89 (2007) 1562–1572.

[26] P. Bucek, J. Jaumot, A. Avino, R. Eritja, R. Gargallo, pH-modulated Watson–Crick duplex–quadruplex equilibria of guanine-rich and cytosine-rich DNA sequences 140 base pairs upstream of the c-kit transcription initiation site, *Chem. Eur. J.* 15 (2009) 12663–12671.

[27] M. del Toro, P. Bucek, A. Aviñó, J. Jaumot, C. González, R. Eritja, R. Gargallo, Targeting the G-quadruplex-forming region near the P1 promoter in the human BCL-2 gene with the cationic porphyrin TMPyP4 and with the complementary C-rich strand, *Biochimie* 91 (2009) 894–902.

[28] S.-T.D. Hsu, P. Varnai, A. Bugaut, A.P. Reszka, S. Neidle, S. Balasubramanian, A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics, *J. Am. Chem. Soc.* 131 (2009) 13399–13409.

[29] J. Dash, Z.A.E. Waller, G.D. Pantoş, S. Balasubramanian, Synthesis and binding studies of novel diethynyl-pyridine amides with genomic promoter DNA G-quadruplexes, *Chem. Eur. J.* 17 (2011) 4571–4581.

[30] A.I. Karsiotis, N.M.a. Hessary, E. Novellino, G.P. Spada, A. Randazzo, M. Webba da Silva, Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism, *Angew. Chem. Int. Ed.* 50 (2011) 10645–10648.

[31] E.W. White, F. Tanious, M.A. Ismail, A.P. Reszka, S. Neidle, D.W. Boykin, W.D. Wilson, Structure-specific recognition of quadruplex DNA by organic cations: Influence of shape, substituents and charge, *Biophys. Chem.* 126 (2007) 140–153.

[32] V.A. Bloomfield, D.M. Crothers, I.J. Tinoco, Bases, nucleosides and nucleotides, in: *Nucleic Acids. Structures, Properties, and Functions*, University Science Books, Sausalito, CA, 2000, pp. 13–44.

[33] A. Arora, S. Maiti, Stability and molecular recognition of quadruplexes with different loop length in the absence and presence of molecular crowding agents, *J. Phys. Chem. B* 113 (2009) 8784–8792.

[34] A. Arora, S. Maiti, Effect of loop orientation on quadruplex–TMPyP4 interaction, *J. Phys. Chem. B* 112 (2008) 8151–8159.

[35] A. Ambrus, D. Chen, J. Dai, R.A. Jones, D. Yang, Solution structure of the biologically G-quadruplex element in the human c-MYC promoter.

Implications for G-quadruplex stabilization, *Biochemistry* 44 (2005) 2048–2058.

[36] C. Wei, L. Wang, G. Jia, J. Zhou, G. Han, C. Li, The binding mode of porphyrins with cation side arms to (TG₄T)₄ G-quadruplex: spectroscopic evidence, *Biophys. Chem.* 143 (2009) 79.

[37] C. Wei, G. Han, G. Jia, J. Zhou, C. Li, Study on the interaction of porphyrin with G-quadruplex DNAs, *Biophys. Chem.* 137 (2008) 19–23.

[38] G.N. Parkinson, R. Ghosh, S. Neidle, Structural basis for binding of porphyrin to human telomerase, *Biochemistry* 46 (2007) 2390–2397.

[39] N.V. Anantha, M. Azam, R.D. Sheardy, Porphyrin binding to quadruplexed T₄G₄, *Biochemistry* 37 (1998) 2709–2714.

Supporting Material

Porphyrin binding mechanism is altered by protonation at the loops in G-quadruplex DNA formed near the transcription activation site of the human *c-kit* gene

Sintayehu Manaye¹, Ramon Eritja², Anna Aviñó², Joaquim Jaumot¹, Raimundo Gargallo^{1*}

1. Solution equilibria and Chemometrics Group Associate Unit (UB-CSIC), Department of Analytical Chemistry, University of Barcelona, Diagonal 645, E-08028 Barcelona, Spain

2. Institute for Research in Biomedicine, IQAC-CSIC, CIBERN-BBN, Baldori i Reixac 15, E-08028 Barcelona, Spain

Contents:

1. Data analysis
2. CD-monitored melting experiments of TMPyP4:ckitG2 and TMPyP4:ckitG2T mixtures.
3. Mole-ratio experiment of TMPyP4 and ckitG2 monitored with molecular absorption spectroscopy.
4. Mole-ratio experiment of TMPyP4 and ckitG2 monitored with circular dichroism spectroscopy
5. Mole-ratio experiment of TMPyP4 and ckitG2T monitored with molecular absorption spectroscopy.
6. Mole-ratio experiment of TMPyP4 and ckitG2 monitored with circular dichroism spectroscopy
7. Supplementary SPR results for the interaction of TMPyP4 with ckit2 and ckit2T when a 1:1 model was considered.
8. Spectroscopically monitored acid-base titration of TMPyP4
9. Spectroscopically monitored acid-base titration of a mixture of TMPyP4 and ckit2
10. Spectroscopically monitored acid-base titration of a mixture of TMPyP4 and ckit2T
11. Mole-ratio experiment involving ckit2 and TMPyP4 at pH 3.1 and 25°C.
12. Mole-ratio experiment involving ckit2T and TMPyP4 at pH 3.1 and 25°C.

Data analysis

Spectra recorded along acid-base titrations, melting experiments or mole ratio studies were analyzed by means of multivariate analysis methods. The procedures and the software used to analyze data have been extensively explained elsewhere (Bucek, 2009; del Toro, 2009; Jaumot et al, 2011), and only a short description will be given here. The goal of the data analysis was to calculate the distribution diagrams and pure (individual) spectra for all the spectroscopically-active species considered throughout the experiment. The distribution diagram provides information about the stoichiometry and stability of the species considered in case of acid-base and mole-ratio experiments. In addition, the shape and intensity of the pure spectra may provide qualitative information about the structure of the species.

With this goal in mind, measured spectra were arranged in a data matrix **D** and later decomposed according to Beer-Lambert-Bouyer's law in matrix form:

$$\mathbf{D} = \mathbf{C} \mathbf{S}^T + \mathbf{E} \quad (1)$$

where **C** is the matrix containing the distribution diagram, **S^T** is the matrix containing the pure spectra, and **E** is the matrix of data not explained by the proposed decomposition.

The mathematical decomposition of **D** according to equation 1 may be done basically in two different ways, depending on whether a physico-chemical model is initially proposed (hard-modelling approach) or not (soft-modelling approach).

For mole-ratio experiments, the physico-chemical model is similar to the previous one.

For example:



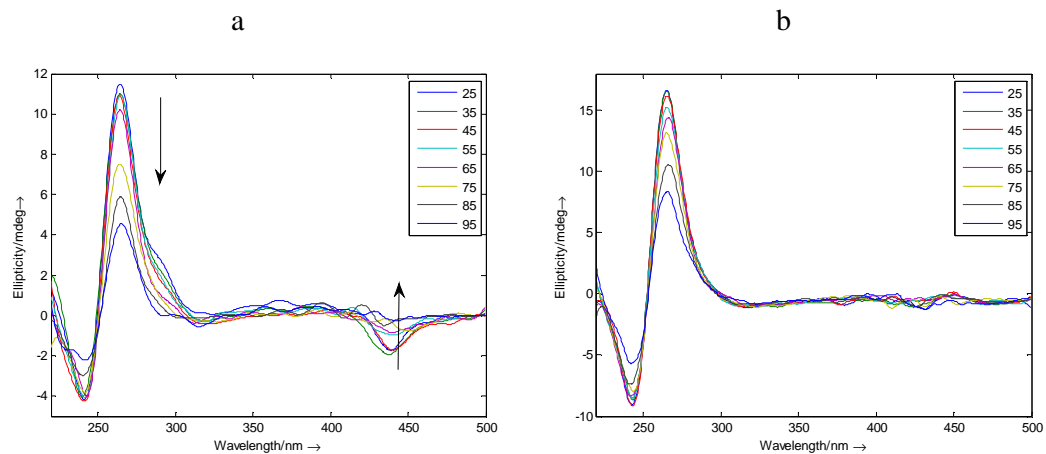
Whenever a physico-chemical model is applied, the distribution diagram complies with the proposed model. Accordingly, the proposed values for the equilibrium constants and the shape of the pure spectra are refined to explain satisfactorily data in **D**, whereas residuals in **E** are minimized.

The application of hard-modelling methods (i.e., those based on compliance with a previously proposed model) has advantages and drawbacks. Hence, not only the calculated distribution diagrams and pure spectra are more robust in relation to the experimental noise, but reliable values for the thermodynamic parameters (stability constants) may be calculated, too. In addition, mixtures of two or more species may be resolved with acceptable incertitude when the stability constants have been calculated unambiguously. The main drawback, however, is the compulsory proposal of a model. Very often, it is difficult to find the most appropriate model to explain a given process and the chemical intuition is needed to replace this incertitude. However, data in matrix **D** may include variance due to other factors unrelated to the process studied (base line drift, impurities...) and cannot be appropriately modelled.

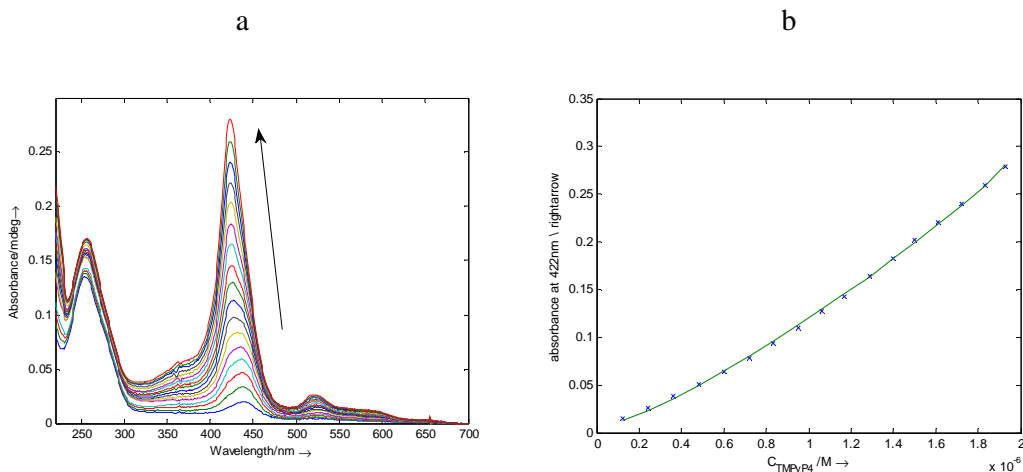
The mathematical decomposition of **D** into matrices **C**, **S^T**, and **E** may also be done without applying a physico-chemical model, which is known as soft-modelling. In this case, the calculation of matrices **C** and **S^T** is based on compliance with a series of constraints which reduce the initially large number of mathematical solutions to (almost) a single physico-chemically meaningful solution. **C** and **S^T** are calculated through an alternate optimization process until a previously set degree of convergence is reached. The application of soft-modelling methods is proposed whenever a physico-chemical model cannot be easily proposed. Its main drawback is, probably, the inability to resolve complex mixtures without the help of more complex experimental and data

analysis setups (for instance, the simultaneous analysis of several data matrices corresponding to complementary experiments). Recently, several hybrid approaches which combine the advantages of hard- and soft-modelling methods have been proposed.

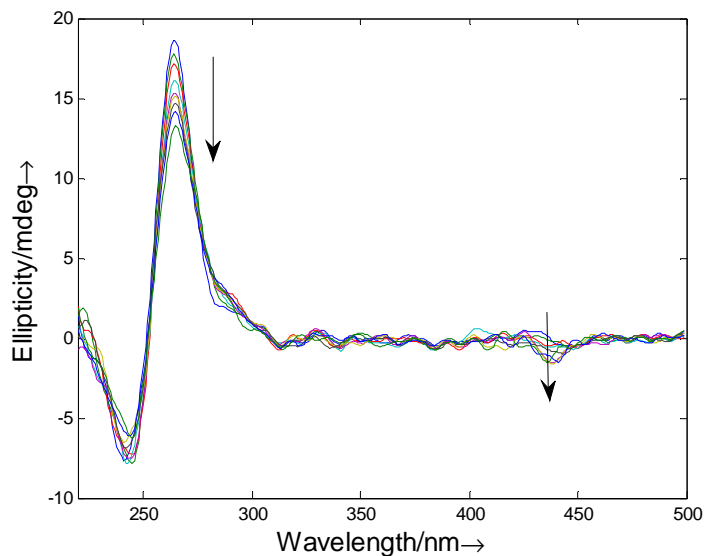
S1. CD-monitored melting experiments of TMPyP4:ckitG2 and TMPyP4:ckitG2T mixtures. (a) Spectra measured in function of temperature for a mixture of ckit2 (2.4 μM) and TMPyP4 (7.4 μM); (b) Spectra measured in function of temperature for a mixture of ckit2T (2.5 μM) and TMPyP4 (7.4 μM). pH=7.1. Arrows indicate the spectral variation upon heating.



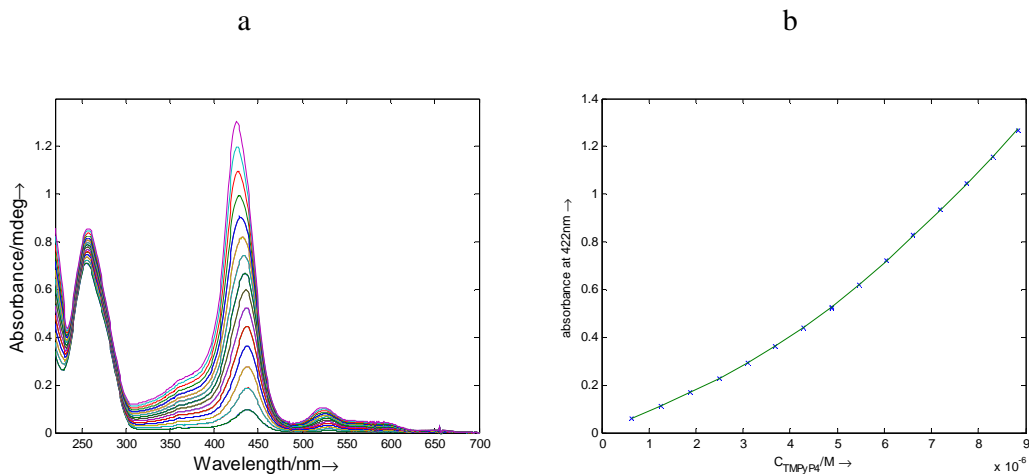
S2. Mole-ratio experiment of TMPyP4 and ckitG2 monitored with molecular absorption spectroscopy. (a) Experimental molecular absorption spectra arranged recorded along the titration of a ckitG2 sample with TMPyP4. The concentration of DNA was $0.5 \mu\text{M}$ and the TMPyP4 concentration ranged from 0.1 to $1.9 \mu\text{M}$. (b) (Experimental (blue symbols) versus calculated (green line) absorbances at 422 nm .



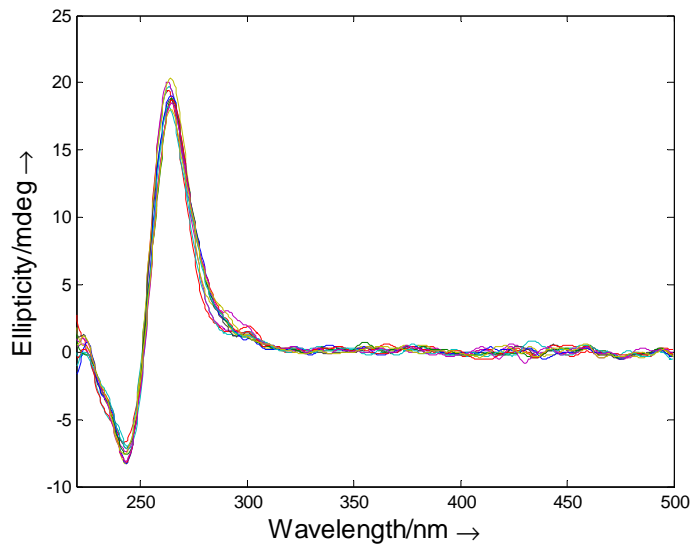
S3. Mole-ratio experiment of TMPyP4 and ckitG2 monitored with circular dichroism spectroscopy. Experimental circular dichroism spectra arranged recorded along the titration of a ckitG2 sample with TMPyP4. The concentration of DNA was $2.0 \mu\text{M}$, and the TMPyP4 concentration ranged from 0.0 to $6.3 \mu\text{M}$. Arrows indicate the spectral variation upon addition of ligand.



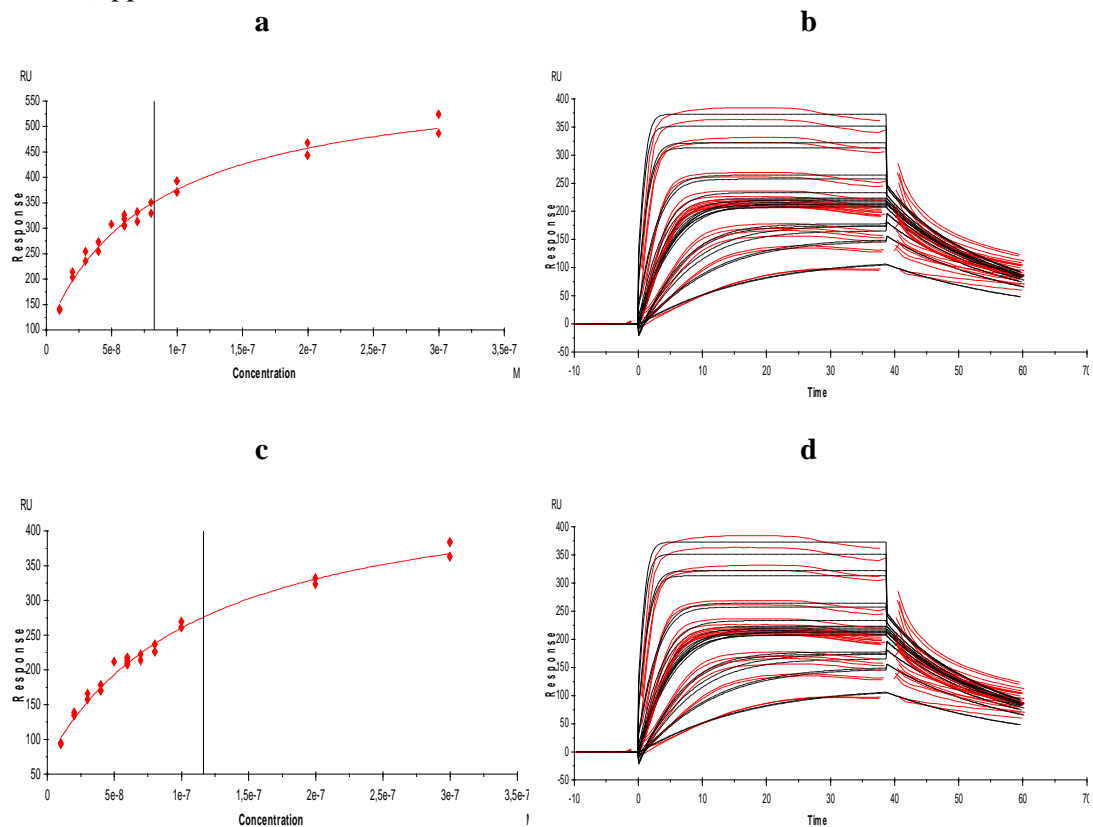
S4. Mole-ratio experiment of TMPyP4 and ckitG2T monitored with molecular absorption spectroscopy. (a) Experimental molecular absorption spectra arranged recorded along the titration of a ckitG2 sample with TMPyP4. The concentration of DNA was $2.9 \mu\text{M}$ and the TMPyP4 concentration ranged from 0.6 to $8.9 \mu\text{M}$. (b) (Experimental (blue symbols) versus calculated (green line) absorbances at 422 nm .



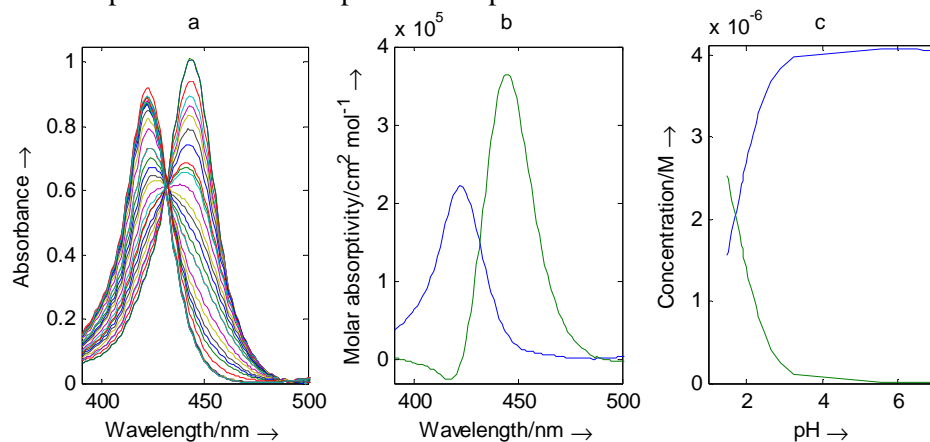
S5. Mole-ratio experiment of TMPyP4 and ckitG2T monitored with circular dichroism spectroscopy. Experimental circular dichroism spectra arranged recorded along the titrations of a ckitG2T sample with TMPyP4. The concentration of DNA was $2.5 \mu\text{M}$, and the TMPyP4 concentration ranged from 0.0 to $7.4 \mu\text{M}$.



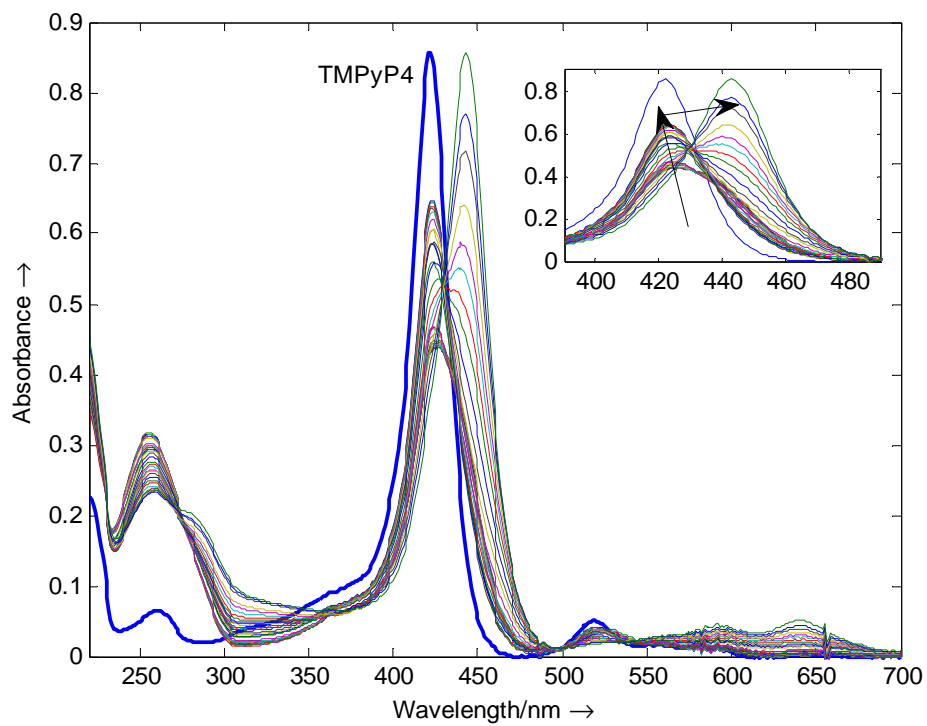
S6. Supplementary SPR results for the interaction of TMPyP4 with ckit2 and ckit2T when a 1:1 model was considered. (a) and (b) show the steady state affinity plot for the interaction of TMPyP4 with ckit2 and ckit2T, respectively. (c) and (d) show the SPR sensorgrams (black) with fitting (red) for the interaction of TMPyP4 with ckit2 and ckit2T, respectively. The concentration of TMPyP4 increased from $1 \cdot 10^{-8}$ M to $3 \cdot 10^{-7}$ (upper curve).



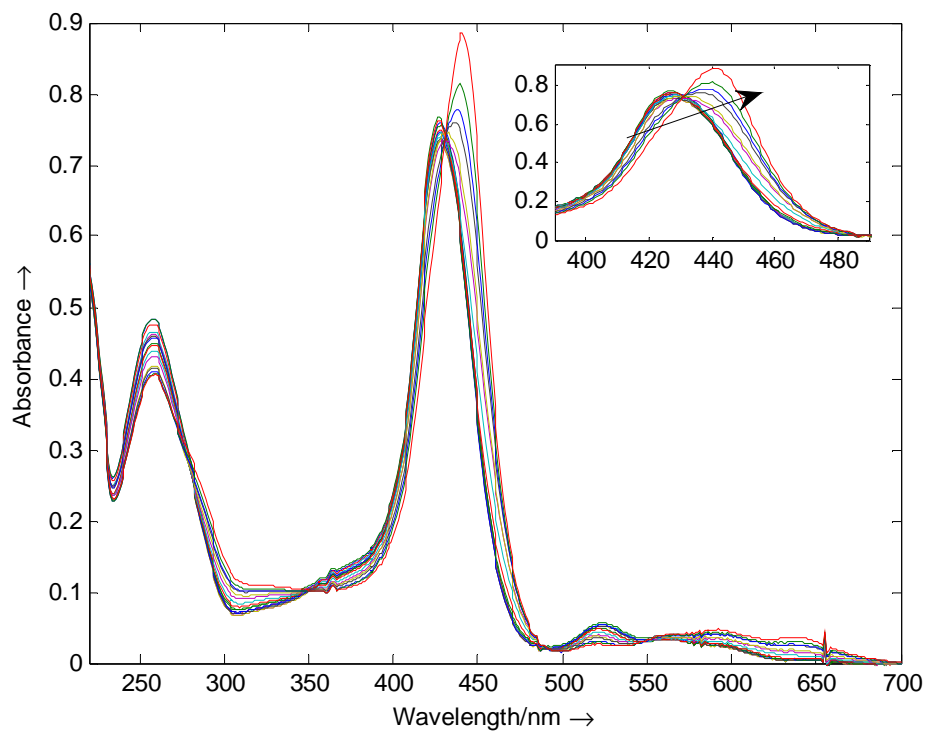
S7. Spectroscopically-monitored acid-base titration of TMPyP4. (a) UV-visible molecular absorption spectra measured throughout the pH range 7.5 – 1.5 at 25°C and 150mM ionic strength. (b) Calculated pure spectra with Equispec for each one of the two acid-base species considered. (c) Calculated distribution diagram. Blue line: deprotonated species. Green line: protonated species.



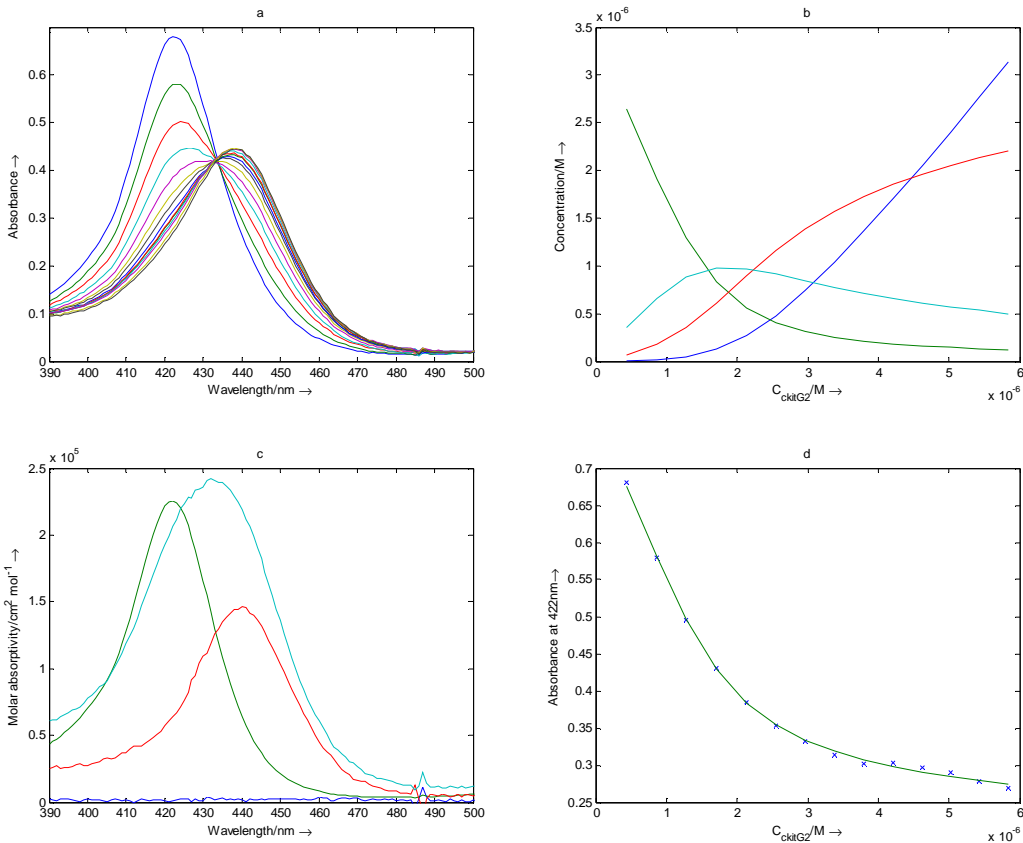
S8. Spectroscopically monitored acid-base titration of a mixture of TMPyP4 and ckit2. Experimental molecular absorption spectra recorded throughout the titration of a mixture of TMPyP4 4.0 microM and ckit2 1.0 microM at 25°C. The inset shows the spectral shifts observed in the Soret region.



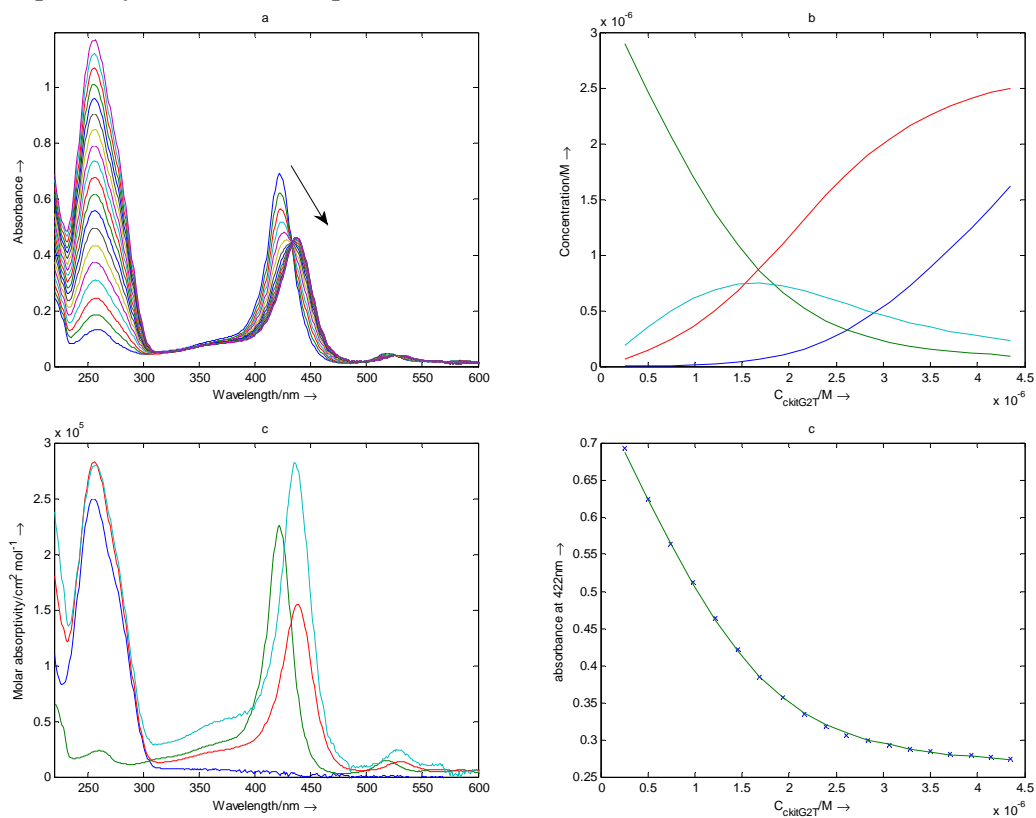
S9. Spectroscopically monitored acid-base titration of a mixture of TMPyP4 and ckit2T. Experimental molecular absorption spectra recorded throughout the titration of a mixture of TMPyP4 3.8 μM and ckit2 1.5 μM at 25°C. The inset shows the spectral shifts observed in the Soret region.



S10. Mole-ratio experiment involving ckit2 and TMPyP4 at pH 3.1 and 25°C. (a) Visible molecular absorption experimental spectra measured for the mixtures of TMPyP4 and ckitG2. $C_{\text{ckitG2}} = 0.4 - 6.6 \mu\text{M}$. $C_{\text{TMPyP4}} = 3.4 - 3.3 \mu\text{M}$. (b) Calculated distribution diagram with Equispec. (c) Calculated pure spectra for each one of the four species considered. (d) Experimental (blue symbols) versus calculated (green line) absorbances at 422 nm. Blue line: ckitG2. Green line: TMPyP4. Red line: 1:1 complex. Cyan line: 1:2 complex.



S11. Mole-ratio experiment involving ckit2T and TMPyP4 at pH 3.1 and 25°C. (a) UV-visible molecular absorption experimental spectra measured for the mixtures of TMPyP4 and ckitG2T. $C_{\text{ckitG2T}} = 0.3 - 4.5 \mu\text{M}$. $C_{\text{TMPyP4}} = 3.3 - 3.1 \mu\text{M}$. (b) Calculated distribution diagram with Equispec. (c) Calculated pure spectra for each one of the four species considered. (d) Experimental (blue symbols) versus calculated (green line) absorbances at 422 nm. Blue line: ckitG2T. Green line: TMPyP4. Red line: 1:1 complex. Cyan line: 1:2 complex.



References

- [1] J. Ren and J. B. Chaires, *Biochemistry* **1999**, *38*, 16067.
- [2] N. W. Luedtke, *Chimia* **2009**, *63*, 134-139.
- [3] A. N. Lane, J. B. Chaires, R. D. Gray and J. O. Trent, *Nucleic Acids Research* **2008**, *36*, 5482-5515.
- [4] H. J. Lipps and D. Rhodes, *Trends in cell biology* **2009**, *19*, 414-422.
- [5] L. Martino, B. Pagano, I. Fotticchia, S. Neidle and C. Giancola, *The Journal of Physical Chemistry B* **2009**, *113*, 14779-14786.
- [6] J. X. Dai, T. S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R. A. Jones and D. Z. Yang, *Journal of the American Chemical Society* **2006**, *128*, 1096-1098.
- [7] S. Cogoi and L. E. Xodo, *Nucleic Acids Research* **2006**, *34*, 2536-2549.
- [8] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 11593-11598.
- [9] H. Fernando, A. P. Reszka, J. Huppert, S. Ladame, S. Rankin, A. R. Venkitaraman, S. Neidle and S. Balasubramanian, *Biochemistry* **2006**, *45*, 7854-7860.
- [10] S. Rankin, A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian and S. Neidle, *Journal of the American Chemical Society* **2005**, *127*, 10584-10589.
- [11] S. Hirota, K. Isozaki, Y. Moriyama, K. Hashimoto, T. Nishida, S. Ishiguro, K. Kawano, M. Hanada, A. Kurata, M. Takeda, G. Muhammad Tunio, Y. Matsuzawa, Y. Kanakura, Y. Shinomura and Y. Kitamura, *Science* **1998**, *279*, 577-580.
- [12] P. S. Shirude, B. Okumus, L. Ying, T. Ha and S. Balasubramanian, *Journal of the American Chemical Society* **2007**, *129*, 7484-7485.
- [13] V. Kuryavyi, A. T. n. Phan and D. J. Patel, *Nucleic Acids Research* **2010**, *38*, 6757-6773.
- [14] I. Haq, J. O. Trent, B. Z. Chowdhry and T. C. Jenkins, *Journal of the American Chemical Society* **1999**, *121*, 1768-1779.
- [15] H. Y. Han, D. R. Langley, A. Rangan and L. H. Hurley, *Journal of the American Chemical Society* **2001**, *123*, 8902-8913.
- [16] M.-Y. Kim, M. Gleason-Guzman, E. Izbicka, D. Nishioka and L. H. Hurley, *Cancer Research* **2003**, *63*, 3247-3256.
- [17] V. r. Gabelica, E. Shammel Baker, M.-P. Teulade-Fichou, E. De Pauw and M. T. Bowers, *Journal of the American Chemical Society* **2007**, *129*, 895-904.
- [18] C. Romera, O. Bombarde, R. Bonnet, D. Gomez, P. Dumy, P. Calsou, J.-F. Gwan, J.-H. Lin, E. Defrancq and G. Pratviel, *Biochimie* **2011**, *93*, 1310-1317.
- [19] M. Gunaratnam, S. Swank, S. M. Haider, K. Galesa, A. P. Reszka, M. Beltran, F. Cuenca, J. A. Fletcher and S. Neidle, *Journal of Medicinal Chemistry* **2009**, *52*, 3774-3783.
- [20] D. Monchaud, A. Granzham, N. Saettel, A. Guedin, J.-L. Mergny and M.-P. Teulade-Fichou, *Journal of Nucleic Acids* **2010**, *2010*, 525682.
- [21] J. Jaumot and R. Gargallo, *Current Pharmaceutical Design* **2012**, *2012*, 1900-1916.
- [22] M. del Toro, R. Gargallo, R. Eritja and J. Jaumot, *Analytical Biochemistry* **2008**, *379*, 8 -15.
- [23] R. M. Dyson, S. Kaderli, G. A. Lawrance and M. Maeder, *Analytica Chimica Acta* **1997**, *353*, 381-393.

- [24] J. Jaumot, R. Gargallo, A. de Juan and R. Tauler, *Chemometrics and Intelligent Laboratory Systems* **2005**, *76*, 101-110.
- [25] N. Khan, A. Avino, R. Tauler, C. Gonzalez, R. Eritja and R. Gargallo, *Biochimie* **2007**, *89*, 1562-1572.
- [26] P. Bucek, J. Jaumot, A. Avino, R. Eritja and R. Gargallo, *Chemistry – A European Journal* **2009**, *15*, 12663-12671.
- [27] M. del Toro, P. Bucek, A. Aviñó, J. Jaumot, C. González, R. Eritja and R. Gargallo, *Biochimie* **2009**, *91*, 894 - 902.
- [28] S.-T. D. Hsu, P. Varnai, A. Bugaut, A. P. Reszka, S. Neidle and S. Balasubramanian, *Journal of the American Chemical Society* **2009**, *131*, 13399 - 13409.
- [29] J. Dash, Z. A. E. Waller, G. D. Pantoş and S. Balasubramanian, *Chemistry – A European Journal* **2011**, *17*, 4571-4581.
- [30] A. I. Karsisiotis, N. M. a. Hessary, E. Novellino, G. P. Spada, A. Randazzo and M. Webba da Silva, *Angewandte Chemie International Edition* **2011**, *50*, 10645-10648.
- [31] E. W. White, F. Tanious, M. A. Ismail, A. P. Reszka, S. Neidle, D. W. Boykin and W. D. Wilson, *Biophysical Chemistry* **2007**, *126*, 140-153.
- [32] A. Arora and S. Maiti, *The Journal of Physical Chemistry B* **2009**, *113*, 8784-8792.
- [33] A. Arora and S. Maiti, *The Journal of Physical Chemistry B* **2008**, *112*, 8151-8159.
- [34] A. Ambrus, D. Chen, J. Dai, R. A. Jones and D. Yang, *Biochemistry* **2005**, *44*, 2048-2058.
- [35] C. Wei, L. Wang, G. Jia, J. Zhou, G. Han and C. Li, *Biophysical Chemistry* **2009**, *143*, 79.
- [36] C. Wei, G. Han, G. Jia, J. Zhou and C. Li, *Biophysical Chemistry* **2008**, *137*, 19-23.
- [37] G. N. Parkinson, R. Ghosh and S. Neidle, *Biochemistry* **2007**, *46*, 2390-2397.
- [38] N. V. Anantha, M. Azam and R. D. Sheardy, *Biochemistry* **1998**, *37*, 2709-2714.