University of
Strathclyde
Glasgow

Strathprints Institutional Repository

http://strathprints.strath.ac.uk/

Chemical Science

EDGE ARTICLE

Rationalising sequence selection by ligand assemblies in the DNA minor groove: the case for thiazotropsin A†

Hasan Y. Alniss,‡[a] Nahoum G. Anthony,[a] Abedawn I. Khalaf,[b] Simon P. Mackay,[a] Colin J. Suckling,[b] Roger D. Waigh,[a] Nial J. Wheate[a] and John A. Parkinson*[b]

DNA-sequence and structure dependence on the formation of minor groove complexes at 5′-XCYRGZ-3′, where Y = T and R = A, by the short lexitropsin thiazotropsin A are explored based on NMR spectroscopy, isothermal titration calorimetry (ITC), circular dichroism (CD) and qualitative molecular modelling. The structure and solution behaviour of the complexes are similar whether X = A, T, C or G and Z = T, A, I (inosine) or C, 5′-CCTAGI-3′ being thermodynamically the most favoured ($\Delta G = -11.1 \pm 0.1$ kcal mol^{-1}). Binding site selectivity observed by NMR for 5′-ACTAGT-3′ in the presence of 5′-TCTAGA-3′ when both accessible sequences are concatenated in a 15-mer DNA duplex construct is consistent with thermodynamic parameters ($|\Delta G|_{ACTAGT} > |\Delta G|_{TCTAGA}$) measured separately for the binding sites and with predictions from modelling studies. Steric bulk in the minor groove for Z = G causes unfavourable ligand–DNA interactions reflected in lower Gibbs free energy of binding ($\Delta G = -8.5 \pm 0.01$ kcal mol^{-1}). ITC and CD data establish that thiazotropsin A binds the ODNs with binding constants between 10^6 and 10^8 M^{-1} and reveal that binding is driven enthalpically through hydrogen bond formation and van der Waals interactions. The consequences of these findings are considered with respect to ligand self-association and the energetics responsible for driving DNA recognition by small molecules in the DNA minor groove.

Introduction

Selective recognition of short sequence nucleic acid targets or larger transcription factor binding sites by sequence reading, minor groove binding small molecules (MGBs) remains relatively difficult, despite efforts directed into this field over several decades. Tailoring this recognition requires a full understanding of the energetic and structural factors responsible if an informed approach to DNA targeted drug-design is to be effectively deployed. This matters if artificial interference with gene expression is to become a reality *via* selective gene transcription activation or repression.[1–5] Sequence recognition limited to between six and eight DNA base-pairs is sufficient to generate allosteric effects capable of substantially influencing gene transcription processes.[6] However, controlling binding strength and

sequence selectivity even over such short regions remains a profound challenge.

In recent years elegant solutions to DNA sequence targeting have emerged through the design of residue specific hairpin and cyclic polyamides containing *N*-methylpyrrole, *N*-methylimidazole[4–7] and other heterocyclic building blocks.[8] The long-term goal of controlling gene expression with small molecules is therefore on the horizon.[9,10]

Detailed rationalization of MGB molecular recognition characteristics can only be completely realized through combined structural, dynamic and energetic analyses. Commonplace biophysical methods, including isothermal titration calorimetry (ITC), capillary electrophoresis (CE) together with linear and circular dichroism (CD), have facilitated understanding in this field.[11] These techniques offer complementary data relevant to assessing the energetic factors responsible for driving drug-DNA binding processes.[11,12] NMR-spectroscopic and X-ray crystallographic approaches continue to provide intimate and complementary structural detail of molecular recognition and dynamic assembly processes.[13,14] By this armoury of techniques the factors governing DNA minor groove recognition by small molecules are becoming apparent.

We are exploring the relationship of ligand dimer pairing and assembly with super-specific DNA recognition in contrast to hairpin or cyclic molecules. The aim is to determine whether designed, low molecular weight complement ligands can

[a] Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 161 Cathedral Street, Glasgow, G4 0NR, United Kingdom

[b] WestCHEM, Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow, G1 1XL, United Kingdom. E-mail: john.parkinson@strath.ac.uk; Fax: +44 141 548 4822; Tel: +44 141 548 2820

‡ Current address: *College of Pharmacy, An-Najah National University, Nablus, Palestine.* E-mail: hasan.alniss@najah.e

assemble and bind to non-complementary DNA targets with similar binding constants to those of hairpin and cyclic MGBs.[15]

In this work we have used the model compound thiazotropsin A, **1**, (FoPyPy^{iPr}ThDp-(formyl)-(N-methylpyrrole)-(N-methylpyrrole)-(isopropylthiazole)-(dimethylaminopropyl)) as a recognition probe against a variety of analogous DNA sequences. By improving our understanding of the relationship between structure and energetics for this recognition event, we hope to influence our future design strategy. Thiazotropsin A has previously been shown to read the self-complementary DNA sequences 5′-NNACT^{5}A^{6}G^{7}T^{8}NN-3′ as a head-to-tail, side-by-side slipped dimer assembly. The DNA complex is stabilized by hydrogen bonds (H2 of **1** to T^{5}O2, H9 of **1** to A^{6}N3, H16 of **1** to G^{7}N3, thiazole N of **1** to G^{7}H22 and H26 of **1** to T^{8}O2).[16] ITC and molecular dynamics simulations[17] have allowed the thermodynamics of binding to be evaluated. CE studies not only confirmed the affinity of **1** for ACTAGT but also suggested that higher order DNA binding of such molecules is possible, as previously acknowledged in other work.[18] Footprinting indicated that alteration of the DNA reading frame flanking bases results in subtle effects on the DNA binding of **1**[19] but to date these results have lacked an integrated structure/energetic explanation.
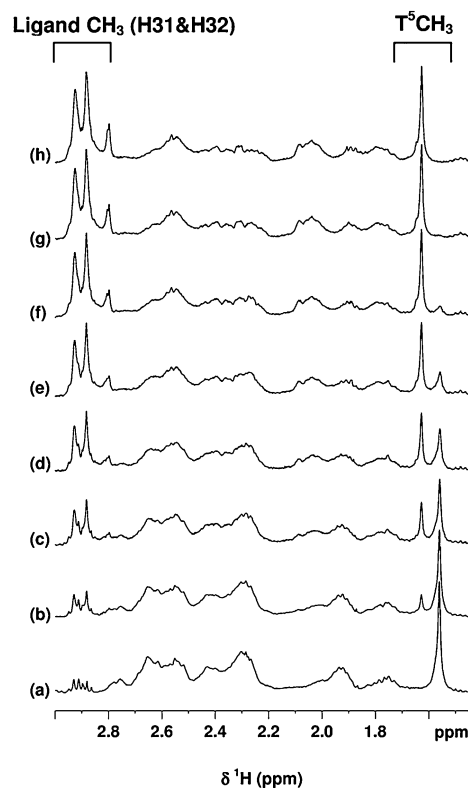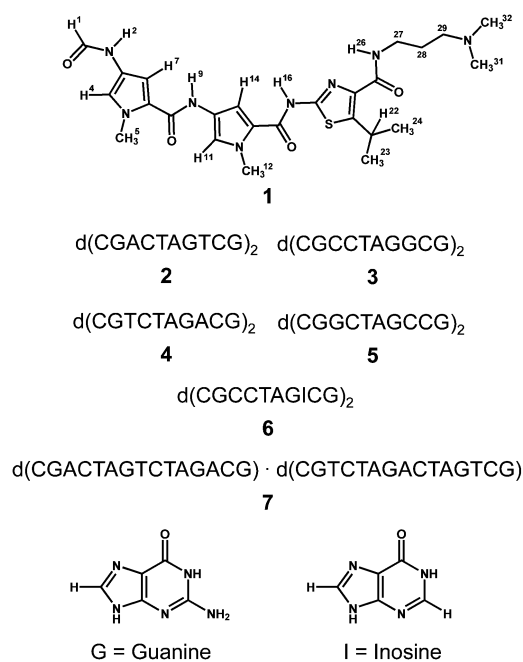
To address this matter, a complementary approach is presented here by systematically altering the flanking bases X and Z in the generic recognition sequence 5′-XCTAGZ-3′ to rationalize footprinting data and provide a comprehensive description of the thermodynamic and structural consequences of subtly altering the DNA sequence offered to **1**. A 15-mer construct incorporating two intersecting sequences represented under separate studies and a sequence in which inosine replaces guanine in one specific instance are included to allow the influence of steric hindrance and other effects to be evaluated. The findings are considered with reference to thermodynamic investigations of **1** bound to d(CGACTAGTCG)$_2$.[17,20] Through this work, we show that the contributions from enthalpy and entropy to the overall binding interactions are highly sequence-dependent with steric

effects dramatically influencing the choice of recognition target selected.

Results

Complex formation monitored by NMR

Complex formation with decamer oligodeoxyribonucleotides (ODNs) containing central 5′-CCTAG*G*-3′, 5′-*T*CTAG*A*-3′, 5′-*G*CTAG*C*-3′, and 5′-CCTAG*I*-3′ (I = inosine) sequences (**3–6**, respectively), as well as with 15-mer **7** was monitored through titration by 1D ^{1}H NMR spectroscopy and compared with similar results obtained with ODN **2**.[16b] With the exception of **3** ("CCTAGG"), the result of each titration (shown for **5**, Fig. 1, by way of example) was complex formation at a ligand:DNA duplex ratio of 2 : 1. No evidence was found for formation of other complexes. The ligand was in slow chemical exchange on the NMR timescale (simultaneous appearance of ^{1}H NMR resonances from free and ligand-bound DNA at <2 equiv. **1** per DNA duplex). The replacement of one set of imino-proton resonances with another without any increase in the number of imino proton resonances proved that binding to each self-complementary ODN was by two ligand molecules as antiparallel, side-by-side pairs. Retention of duplex symmetry was the case for all but the non-self complementary ODN duplex **7**,

1

d(CGACTAGTCG)$_2$ d(CGCCTAGGCG)$_2$
2   **3**

d(CGTCTAGACG)$_2$ d(CGGCTAGCCG)$_2$
4   **5**

d(CGCCTAGICG)$_2$
6

d(CGACTAGTCTAGACG) · d(CGTCTAGACTAGTCG)
7

G = Guanine   I = Inosine

Fig. 1 The aliphatic region of the 400 MHz 1D ^{1}H NMR spectrum of **1** titrated into a sample of **5**. The ^{1}H NMR resonance of the T^{5}CH$_3$ group is visible at δ ^{1}H = 1.56 ppm (free DNA) and at 1.63 ppm (ligand:DNA complex). Resonances at δ ^{1}H = 2.88 and 2.93 ppm are from CH$_3$ protons H31 and H32 of **1**. (a) Free DNA; (b) with 0.3 equiv. **1**; (c) with 0.6 equiv. **1**; (d) with 0.9 equiv. **1**; (e) with 1.2 equiv. **1**; (f) with 1.5 equiv. **1**; (g) with 1.8 equiv. of **1**; (h) with 2 equiv. **1**.

which, following full data assignment, was also shown to bind only 2 equiv. of **1**.

Comparison of 1D ^{1}H NMR spectra for all 2 : 1 ligand:DNA complexes formed (Fig. 2) readily revealed similarities and differences between the data. Methyl proton resonances for **1** were generally identifiable. When compared with the data for **1** bound to CG*A*CTAG*T*CG, **2** (Fig. 2a), sharp singlet ^{1}H NMR resonances were observed for ligand Py and Dp *N*-methyl groups (δ ^{1}H = 3.6–4.2 and 2.8–3.0 ppm regions of the ^{1}H NMR data respectively) in complexes between **1** and ODNs **4–7**. Methyl doublet signals arising from ^{iPr}Th in **1** were typically observed at δ ^{1}H < 1.5 ppm as clearly identifiable resonances. For CG*CC*TAG*GCG*, **3**, CG*G*CTAG*C*CG, **5** and CG*CC*TAG*I*CG, **6**, each ODN possessed a single T per strand and for **4** (CG*T*CTAG*A*CG), **5** and **6**, a resonance is observed in the δ^1H ≈ 1.6 ppm region of the 1D ^{1}H NMR spectrum of each complex, corresponding to the methyl proton resonance of DNA residue T^5 in a C_{2v} symmetric environment.
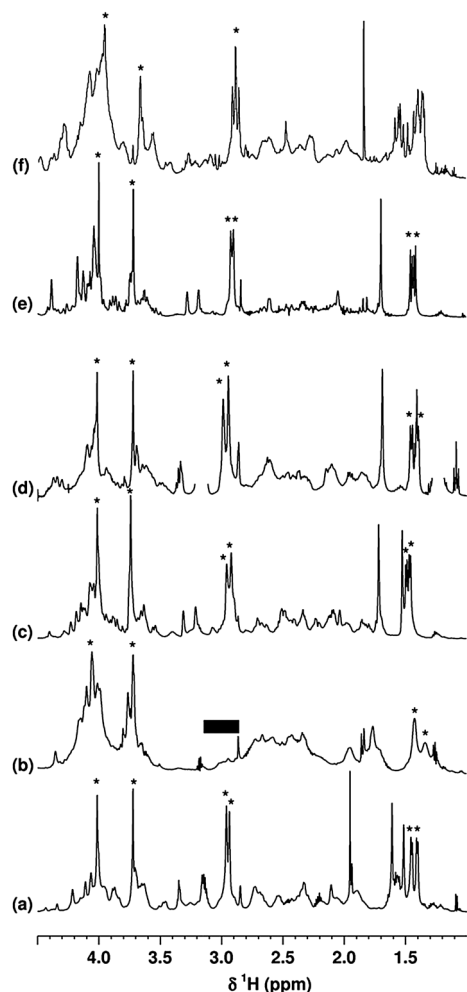
Fig. 2 1D ^{1}H NMR data in the aliphatic resonance region for the binding of 2 equiv. **1** to (a) **2**, CG*A*CTAG*T*CG; (b) **3**, CG*CC*TAG*G*CG; (c) **4**, CG*T*CTAG*A*CG; (d) **5**, CG*G*CTAG*C*CG; (e) **6**, CG*CC*TAG*I*CG and (f) **7**, CG*A*CTAG*T*CTAG*A*CG·CG*T*CTAG*A*CTAC*T*CG. Ligand methyl resonances are indicated by * in all cases. The solid bar in (b) indicates "missing" resonances for ligand methyl protons H31 and H32 of **1**.
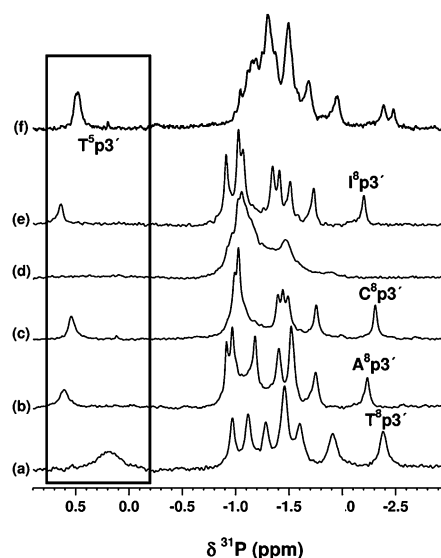
Fig. 3 1D ^{31}P-{^{1}H} NMR data for complexes of 2 equiv. **1** with (a) **2**, CG*A*CTAG*T*CG, (b) **4**, CG*T*CTAG*A*CG, (c) **5**, CGG*C*TAG*C*CG, (d) **3**, CG*CC*TAG*G*CG, (e) **6**, CG*CC*TAG*I*CG and (f) **7**, CG*A*CTAG*T*CTA-GACG·CG*T*CTAG*A*CTAC*T*CG acquired at 9.4 T. Selected ^{31}P resonance assignments indicated for each complex are based on 2D [^{31}P, ^{1}H] correlation data.

Contrastingly, CG*CC*TAG*G*CG, **3**, showed marked differences in behaviour compared with all of the other ODNs on binding **1**. Whilst the Py *N*-methyl resonances were visible (Fig. 2b), the quality of the NMR data was poor and degraded further with each additional aliquot of ligand. Broadened spectra are typical of more rapid ligand exchange (see also Figs S4 and S5 in the Supporting Information†). One-dimensional ^{31}P-{^{1}H} NMR data compared for each of the complexes confirmed these findings (Fig. 3). Substantial contrast existed between the data for the complex of **3**, CG*CC*TAG*G*CG (Fig. 3d, poor signal dispersion) and those of the remaining ODNs **2** and **4–7**. The resonance at high chemical shift (δ^{31}P *ca.* 0.5 ppm), consistent across all sequences except for **3**, was previously assigned to the phosphorous atom 3′ to T^5 for the complex between **1** and **2** and ascribed to a locked B$_{II}$ conformation of the phosphodiester backbone at this location. Its broadened nature compared with the other ^{31}P NMR resonances in the same NMR spectrum (Fig. 3a) and with similar data for complexes between **1** and **4–7** is attributed to differences in the dynamic behaviour of the DNA backbone at this position for different sequences.

NMR data assignment

^{1}H, ^{31}P and ^{13}C NMR signal assignments were determined for all ODNs and each complex using established strategies and data comparisons.[16b] Signal assignments for ODNs **4–7** and 2D [^{1}H, ^{1}H] NOESY NMR data typical of each 2 : 1 complex are reported in the Supporting Information (Tables S1–S4 and Fig. S1†, respectively).

Natural abundance 2D [^{1}H, ^{13}C] HSQC NMR data assisted in pinpointing sugar ring H4′ proton resonances that are subjected to the shielding influence of aromatic peptide *N*-methyl Py rings lodged in the DNA minor groove[16b,21] (Supporting Information

Fig. S2†) and used as a simple device for screening minor-groove binding, as well as assisting in the identification of ^{1}H NMR resonances that were far removed from their nominal free-DNA resonance position.

Assignments of ^{1}H NMR resonances of **1** bound to ODNs **4–7** (Supporting Information, Table S5†) show remarkable consistency. Without further detailed NOE analysis, the intrinsic similarity of these chemical shifts alone bears witness to the close similarity between the structures of each of the complexes formed in solution for **1** with ODNs **4–7**. ^{1}H NMR resonance assignments were identified for ODNs in complexes with **1** bound securely (Supporting Information, Tables S7–S10†).

Partial chemical shift assignments for proton-attached ^{13}C atoms (*via* 2D [^{1}H, ^{13}C] HSQC NMR data) were made for **1** in complex with CG*C*CTAG*I*CG, **6**, by direct comparison with the assigned proton NMR data for the complex (Supporting Information, Table S11†). These data are of benefit in identifying resonance assignments that are elusive in homonuclear proton NMR data.
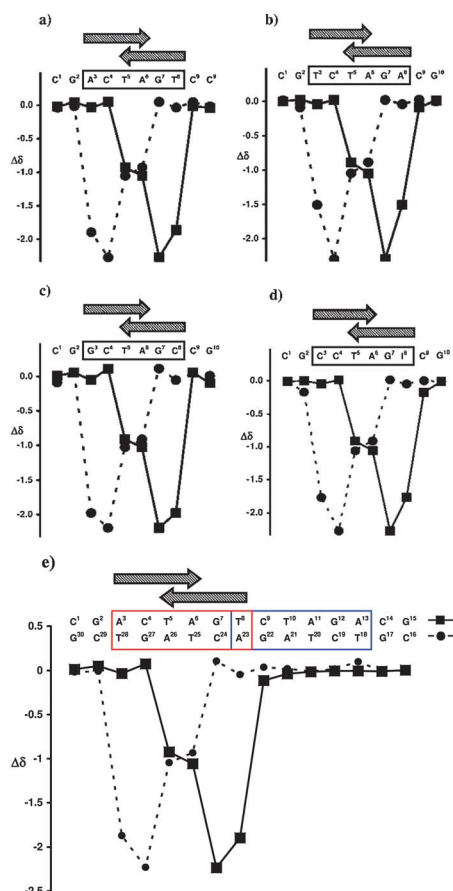
Fig. 4 A comparison of the ^{1}H NMR chemical shift differences (Δδ) measured for H4′ resonances as δ H4′ DNA [(free) – (bound)] for ligand-bound and ligand-free DNA duplexes: (a) **1** with **2**; (b) **1** with **4**; (c) **1** with **5**; (d) **1** with **6**; (e) **1** with **7** (solid line: C^1-G^{15}; dotted line: C^{16}-G^{30}). The ligand is indicated by arrows. Dashed lines show shift changes for opposing DNA strands (reverse sequence for (a–d); associating ligand represented by top arrow in each case). For (e), two potential binding sites are indicated by red and blue boxes.

Chemical shift differences (Δδ) between H4′ resonances of free DNA and those of DNA bound with two equivalents of **1** for **2**, CG*A*CTAG*T*CG, and ODNs **4–7** (Fig. 4) show excellent agreement and describe patterns that virtually superimpose. Juxtaposition of the ligand against each DNA sequence occurs at 5′-T^{5}A^{6}G^{7}A^8-3′ for **4**, 5′-T^{5}A^{6}G^{7}C^8-3′ for **5** and 5′-T^{5}A^{6}G^{7}I^8-3′ for **6**, creating the same footprint as that revealed for **1** binding to **2** (5′-T^{5}A^{6}G^{7}T^8-3′, Fig. 4a).

Remarkably, whilst **1** binds to both **2**, '*ACTAGT*', and **4**, '*TCTAGA*' under separate experimental conditions (Figs 4a and b), the binding of **1** to DNA duplex **7**, d (CG*ACTAGT*CTAGACG)·d(CG*TCTAGA*CTAGTCG), containing two possible recognition sites for **1**, but overlapping by one base pair, is specific only for *ACTAGT* in the presence of the co-joined *TCTAGA* binding site (Fig. 4e and Table S13 in the Supporting Information†). The design of this sequence was such that both *ACTAGT* and *TCTAGA* binding sites would be made simultaneously available to **1** for binding, but it contained an exclusion condition: steric hindrance, caused through the occupation of one binding site by **1**, was predicted to prevent simultaneous occupation of both sites, thereby enabling the subtle selection by **1** of one binding site over the other. In this way, sequence preference could be determined by a simple analysis of the chemical shift data.

The results are unambiguous (Fig. 4e): despite these sites being simultaneously available, a clear preference for *ACTAGT* over *TCTAGA* is shown by **1**. No evidence was found for occupation of the '*TCTAGA*' binding site under the experimental conditions used.

A 16-mer self-complementary DNA duplex longer by one base pair unit and containing two *adjacent* binding sites of sequence ACTAGT, namely d(CG*ACTAGTACTAGT*CG)$_2$, was able to fully accommodate 4 equivalents of **1** and resulted in NMR data consistent with perfect C_{2v} symmetry (See Fig. S10 of the Supporting Information†). These results are remarkable in view of the ITC and CD data and are discussed at length (*vide infra*).

As with the complex between **1** and **2**, CG*ACTAGT*CG, H4′ chemical shift values showed the greatest change upon ligand binding compared with all of the other DNA proton resonances. Similarities in Δδ for H5″ of A^6 in each instance (−1.107 ppm for **4**, CG*T*CTA^{6}G*A*CG, −1.179 ppm for **5**, CG*G*CTA^{6}G*C*CG and −1.164 ppm for **6**, CG*C*CTA^{6}G*I*CG) also compared favourably with a value of −1.167 ppm for **2** (CG*A*CTA^{6}G*T*CG) when ligand was bound. Phosphodiester backbone alteration at the T^{5}pA^6 step (correlated with unusual chemical shifts shown in the ^{31}P NMR data for each complex (Table 1) is consistent across all of the examples with the exception of the binding associated with ODN **3**, CG*C*CTAG*G*CG.

Modelling

Proton–proton NOEs assigned from the 2D [^{1}H, ^{1}H] NOESY NMR data acquired for each of the complexes of **1** with ODNs **4–7** (Supporting Information, Tables S14–S17†) were used for model building. Modelled complexes were based on PDB entry 1RMX by replacing the relevant residues of **2** with appropriate residues for each complex. These were used to check the assigned NOEs for plausibility. Ligand–DNA NOEs are illustrated schematically in Fig. 5, the density of NOE information largely

Table 1 A representative comparison of ^{31}P NMR chemical shift assignments. **4**, d(CGTCTAGACG)$_2$, in the absence (free) and presence (bound) of two equivalents of **1** per duplex

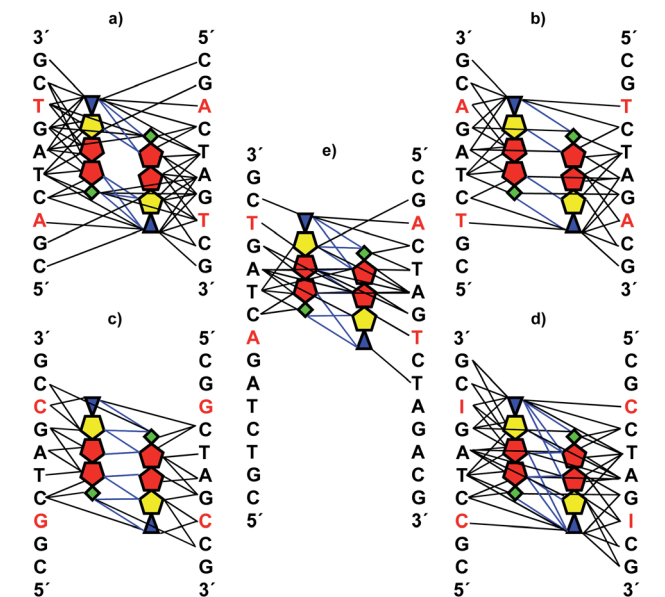| Base	δ ^{31}P (ppm)^a		$\Delta\delta$ ^{31}P^b
	Bound	Free	Difference
C1	−0.958	−0.954	−0.004
G2	−1.498	−1.374	−0.124
T3	−1.172	−1.063	−0.109
C4	−1.521	−1.290	−0.231
T5	**+0.609**	**−1.135**	**1.744**
A6	−1.377	−1.088	−0.289
G7	−1.732	−1.179	−0.553
A8	−2.209	−1.060	−1.149
C9	−0.907	−0.893	−0.014
G10	—	—	—

^{a} Assignments are for 3′ phosphates with respect to the base. ^{b} $\Delta\delta$ ^{31}P = (δ^{31}P$_{bound}$ − δ ^{31}P$_{free}$).

reflecting both the quality of NMR data accumulated and the extent to which these were assigned for ligand docking purposes. A representative example of the decamer model complexes is shown for DNA duplex **6**, together with the complex modelled for 15-mer duplex **7** (Fig. 6). Data fitted the 2 : 1, head-to-tail, side-by-side binding of **1** within the widened minor groove of each DNA structure alongside the 5′-CTAGZ-3′ DNA sequence (Z = A, C or I). For the symmetrical complexes of **1** with ODNs
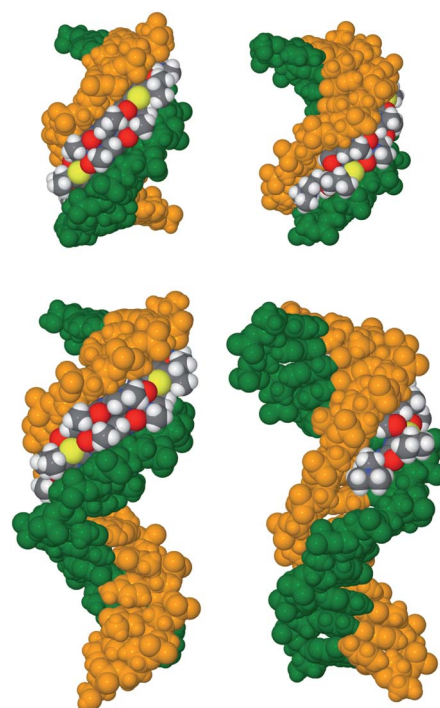
Fig. 6 Model structures based on NOE data. Top: complex with deca-mer duplex **6**, d(CGCCTAGICG)$_2$. Bottom: complex with 15-mer duplex **7**, d(CGACTAGTCTAGACG)·d(CGTCTAGACTAGTCG). Ligands are coloured according to atom type. Individual DNA strands are col-oured in green and orange. In each case the right-hand image results from an anti-clockwise rotation of the left-hand image by 90°.

4–6, identification of inter-ligand NOEs was only reliable between the "head" of one ligand and the "tail" of the adjacent ligand, whereas for the unsymmetrical complex between **1** and 15-mer ODN **7**, the structural asymmetry of the DNA allowed four inter-ligand NOEs to be identified between protons associ-ated with the central *N*-methyl Py rings of adjacent ligands.

Additional contacts were noted between isopropyl protons of ^{i}PrTh and the "outer edges" of the C^9 and G^{10} nucleotides of each complex (*e.g.* NOEs ligand(H24)–C^{9}H4′ and ligand(H24)–G^{10}H4′). The nature of the 'tail' is often overlooked in such studies, but it is clear from these results and from those discussed for the complex between **1** and ODN **3**, CG*C*CTAG*G*CG (*vide infra*) that the Dp 'tail' plays a key role in recognition and stabilization of the complex through contour surface matching and hydrophobic interaction.

Isothermal titration calorimetry (ITC)

Energetic parameters associated with ligand–DNA recognition for the interaction of **1** with the five related dodecamer ODNs (5′-GCG*ACTAGT*CGC-3′, **2′**, 5′-GCG*CCTAGG*CGC-3′, **3′**, 5′-GCG*TCTAGA*CGC-3′, **4′**, 5′-GCG*GCTAGC*CGC-3′, **5′** and 5′-GCG*CCTAGI*CGC-3′, **6′**) were evaluated using ITC. The results (Table 2) represent the mean ± standard error of dupli-cate experiments. In each case, the first 10 injections of ligand titrated the available ODN in the sample cell and generated an exothermic signal response (Fig. 7).

Fig. 5 A schematic representation of the inter-ligand (blue lines) and ligand–DNA NOEs for the complexes between **1** and DNA duplexes **4–7** (b–e) as compared with the pattern of ligand-related NOEs observed in the 2D [^{1}H, ^{1}H] NOESY NMR data for the complex of **1** with **2** (a). DNA base modifications are highlighted with red lettering. Individual lines represent clusters of NOEs between residues. Ligand colour coding: green diamond – formyl head-group (Fo); red pentagon – *N*-methylpyrrole unit (Py) linked *via* peptide bonds to adjacent residues; yellow pentagon – *C*-isopropylthiazole unit (^{i}PrTh); blue triangle –dimethyl aminopropyl tail-group (Dp).

Table 2 Thermodynamic parameters for the interaction of **1** with five DNA sequences based on ITC data

Sequence	ΔH kcal mol^{-1}	$T\Delta S$ kcal mol^{-1}K^{-1}	K M^{-1}/10^7	ΔG kcal mol^{-1}
ACTAGT (**2′**)	-12.8 ± 0.1	-2.7 ± 0.3	3.0 ± 1.1	-10.2 ± 0.2
TCTAGA (**4′**)	-10.4 ± 0.5	-0.3 ± 0.2	2.0 ± 0.8	-9.9 ± 0.2
GCTAGC (**5′**)	-12.0 ± 0.2	-2.5 ± 0.4	0.96 ± 0.09	-9.5 ± 0.1
CCTAGG (**3′**)	-5.5 ± 0.3	3.0 ± 0.3	0.17 ± 0.01	-8.5 ± 0.01
CCTAGI (**6′**)	-9.6 ± 0.4	1.5 ± 0.4	11.0 ± 0.5	-11.1 ± 0.1
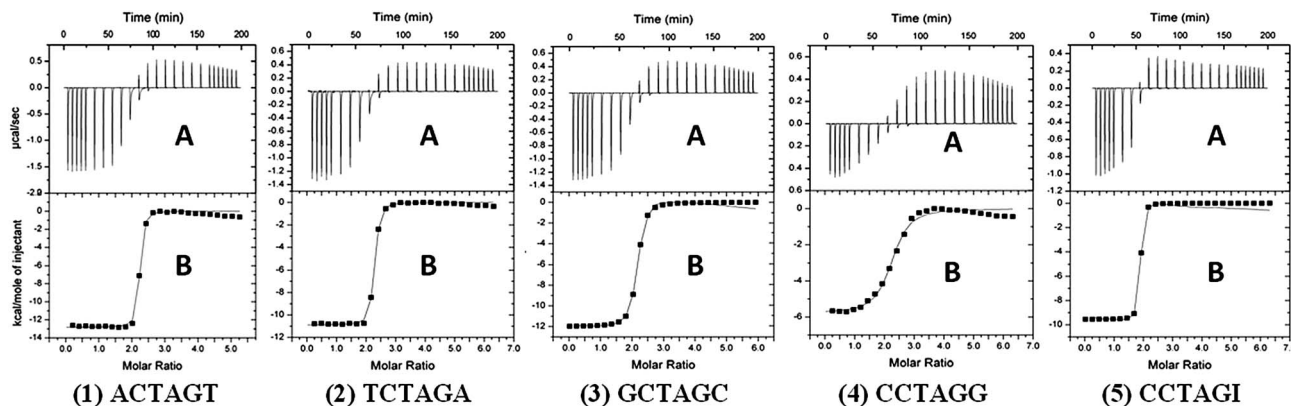
Fig. 7 The ITC titrations of **1** with five DNA sequences in PIPES buffer at 25 °C (pH 6.8). (A) Raw data for the titration of thiazotrosin A, **1**, into: (1) GCG**ACTAGT**CGC; (2) GCG**TCTAGA**CGC; (3) GCG**GCTAGC**CGC; (4) GCG**CCTAGG**CGC; (5) GCG**CCTAGI**CGC. (B) Enthalpogram retrieved from (A) corrected for the heat of dilution. The line represents the least-squares-fit to the single-site binding model.

Subsequent injections generated an endothermic response as a result of the heats of dilution of the ligand in the buffer. Dilution responses (Fig. S11 in the Supporting Information†) were all endothermic and their intensity decreased as more ligand was added, indicating the aggregation of **1** in buffered aqueous solution.

Enthalpograms generated from integration of the raw experimental binding data followed by subtraction of the heats of ligand dilution yielded binding enthalpy, ΔH, binding free energy, ΔG, entropy changes, ΔS and stoichiometry of binding, N.

Significant heat changes detected upon the addition of **1** to four of the ODNs characterized tight minor groove binding. Binding to the ACTAGT-containing sequence, **2′**, involved a single binding process with a binding constant of $K = 3.0 \times 10^7$ M^{-1} and a binding stoichiometry (ligand to DNA) of 2 : 1. The enthalpy (ΔH) and entropy ($T\Delta S$) of this interaction was negative (-12.8 kcal mol^{-1} and -2.7 kcal mol^{-1}, respectively), characterizing an exothermic binding event with entropic opposition.

The large negative enthalpy change results in a negative Gibbs free energy of binding ($\Delta G = -10.2$ kcal mol^{-1}). Titration of **1** with dodecamer ODNs containing the central sequences TCTAGA, **4′**, and GCTAGC, **5′**, revealed very similar binding characteristics compared with the titration of **1** against ACTAGT. Relatively weak binding was observed for CCTAGG ($\Delta G = -8.5$ kcal mol^{-1}) compared with other sequences. Use of the equivalent dodecamer in which guanine in 'CCTAG*G*' (**3′**) was replaced by inosine in 'CCTAG*I*' (**6′**) (removal of exocyclic 2-amino group protruding from the minor groove floor) was accompanied by a significant increase in binding affinity: $\Delta G = -8.5$ kcal mol^{-1} for **3′** compared with $\Delta G = -11.1$ kcal mol^{-1} for **6′** (Fig. 7 and Table 2).

Circular dichroism

CD analysis of each ODN in PIPES buffer indicated a typical B-type DNA tertiary structure: a negative peak at 255 nm and a large positive peak at 285 nm consistent with other short oligonucleotides (Fig. 8). Ligand addition gave CD spectra consistent with widening of the DNA minor groove. The negative peak at 255 nm became positive with increasing ligand concentration for all five ODNs. New positive peaks were also observed at 316 and 355 nm, which increased in size with increasing ligand concentration, consistent with minor groove binding. Small differences between the CD responses of the five ODNs were apparent. Thus binding of **1** to TCTAGA and ACTAGT induced a bathochromic shift of the positive peak at 285 nm to 290 nm. For GCTAGC this shift was less pronounced and for CCTAGG; neither a change in molar ellipticity nor in the wavelength of the peak at 285 nm was seen. Plots of the molar ellipticity at 316 nm against ligand concentration yielded binding curves for all five ODNs (Fig. 8) and gave rise to a set of binding constants (Table 3). Binding was weakest to *G*CTAG*C* and *C*CTAG*G* and strongest to *C*CTAG*I*, which is in good agreement with the binding constants determined by ITC and with the NMR data, indicating that *C*CTAG*I* is the most favourable binding sequence for this ligand.

Discussion

Steric hindrance and the role of ligand self-assembly

The report by James *et al.*[19] details the sequence requirements for DNA minor groove binding of **1** and provides an important basis to this work. Debate over whether sequence recognition is driven by hydrophobic, topological and electrostatic factors or through
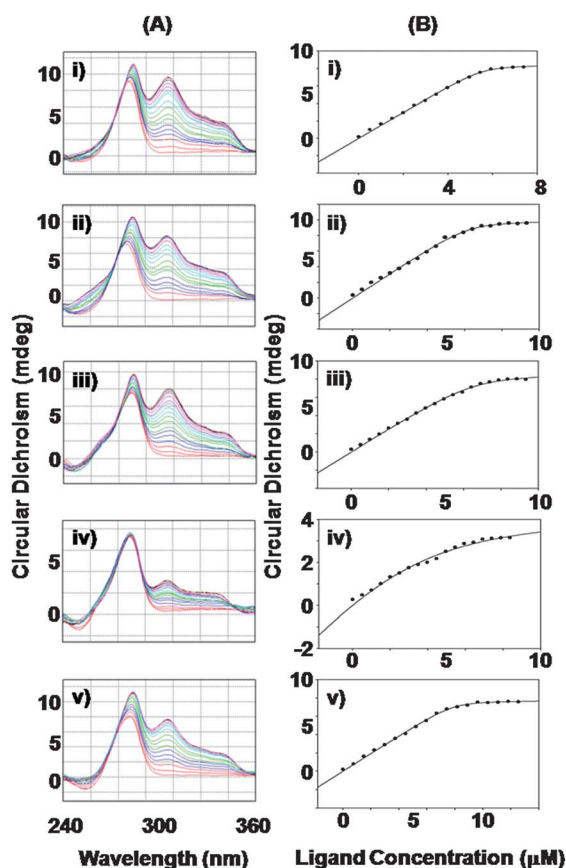
Fig. 8 CD spectra of **1** (A) titrated against five DNA sequences containing the central sequences ACTAGT (i), TCTAGA (ii), GCTAGC (iii), CCTAGG (iv), and CCTAGI (v). Increase in ellipticity (B) at 316 nm as a function of the quantity of **1** for (i–v).

hydrogen bonding continues, especially in view of studies such as those reported by Pang.[15] Part of the rationale behind our approach to designing small molecule DNA sequence readers encompasses the notion of enhanced lipophilicity. The observation by James *et al.* that **1** does not bind to 5′-XCYRGZ-3′ for X = C and Z = G tallies with the poor quality NMR data observed for **1** in complex with **3**, "CCTAGG", compared with the good quality data for **1** in complex with **5**, "GCTAGC" (Fig. 1). The difference exists through rotation of the flanking base pairs X·Z by 180° (Fig. S9 in the Supporting Information†). In the context 5′-XCYRGZ-3′, **1** tolerates X = C or G but does not tolerate Z = G. Removal of the exocyclic NH$_2$ from the 2 position of Z = G in the DNA minor groove by substitution of

G with inosine to give **6**, was predicted by modelling[16b,20] to alleviate this intolerance. The experimental ITC and NMR evidence bears this out.

Ligand self-assembly must be considered in this context. Independent studies of **1** alone in free solution have resulted in the observation of negative NOEs,[22] typical of larger molecules ($M_r \approx 2$ kDa or greater) or self-assembled aggregates of smaller molecules. The NOE data are consistent with head-to-tail dimer formation for **1** and related compounds. This has a bearing on how one ligand affects its partner if, as in the context of **1** binding to **3** (CCTAGG), the tail of one ligand is not accommodated within the DNA minor groove due to steric effects (Fig. 9).

The 'head' of an adjacent ligand may also lift out of the groove when the tail lifts out (or *vice versa*), significantly reducing the overall lipophilic contact between a pair of ligands and the walls of the DNA minor groove (Fig. 9b). Alternatively, the tail of one ligand may be forced out of the groove, but the head of the adjacent ligand remain firmly anchored to it (Fig. 9c). Loss in
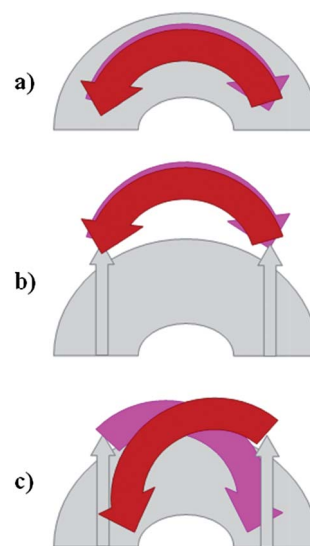
Fig. 9 A cartoon representing the effect on side-by-side ligand binding of a sterically demanding group lining the DNA minor groove floor: (a) unperturbed antiparallel side-by-side binding; (b) steric blocking group (grey arrow) present at either end of a palindromic DNA recognition sequence – the driving force for the ligand to remain associated with the groove < for ligands to remain associated with one another; (c) as for (b), but where the driving force for ligands to remain associated < the preference for ligand–DNA assembly. Grey: DNA minor groove; red/magenta arrows – antiparallel side-by-side DNA MGBs.

Table 3 A comparison of the ITC and CD results for the binding of **1** to five dodecamer DNA sequences

Sequence	ITC Results		CD Results	
	K M^{-1} (duplex)/10^7	ΔG kcal mol^{-1}	K M^{-1} (duplex)/10^7	ΔG kcal mol^{-1}
ACTAGT (**2′**)	3.0	−10.2	3.0	−10.2
TCTAGA (**4′**)	2.0	−9.9	2.6	−10.1
GCTAGC (**5′**)	0.96	−9.5	0.79	−9.4
CCTAGG (**3′**)	0.17	−8.5	0.11	−8.2
CCTAGI (**6′**)	10.7	−11.1	4.6	−10.5

lipophilic contact between the ligands in this way might result in energy losses, thereby perturbing overall DNA binding. An increase in the binding affinity for Z = I compared with Z = G in 5′-XCTAGZ-3′ is reported through a larger negative Gibbs free energy of binding. This may be the result of the combined effect of ligand-on-ligand alignment, as well as ligand-on-DNA association where the combination of hydrophobic/electrostatic interactions (entropic driving force) and hydrogen bonding (enthalpic driving force) balance one another out resulting in a tightly bound complex.

Exclusive binding test shows selective recognition

15-mer ODN **7** was designed to test the ability of **1** to select one or both potential binding sites within the DNA construct and to establish whether modelling studies correctly predicted the rank ordering of binding affinity for **1** based on energetics.[23] Simultaneous binding of two side-by-side ligand assemblies to two different binding sites was not expected in this case due to steric effects. Binding does occur in this way for *adjacent* binding sites as shown for 16-mer duplex d(CG***ACTAGT***ACTAGTCG)₂ in which both binding sites take up ligand simultaneously. At full occupation, 4 equivalents of **1** are bound and NMR data are consistent with a complex possessing perfect C_{2v} symmetry (Fig. S10, Supporting Information†). In contrast, 15-mer ODN, **7**, provides two binding conditions of specific interest. The first tests the ability of the ligand to bind as a single molecule (monomer) or not. The second tests the ability of the ligand to bind as a side-by-side pair exclusively to only one binding site (see Fig. S12 in the Supporting Information†for examples). In our hands, ligands remained associated side-by-side in an anti-parallel sense. Remarkably, the ligand pair preferentially and exclusively selects 5′-ACTAGT-3′ over 5′-TCTAGA-3′ in the context 5′-***ACTAGT***CTAGA-3′, despite the ability of ligand pairs to separately bind both sequences in a side-by-side fashion. One explanation for this behaviour concerns sequence context.

5′-PyPu-ACTAGT-PyPu-3′
3′-PuPy-TGATCA-PuPy-5′

5′-PyPu-ACTAGT-PyPy-3′
3′-PuPy-TGATCA-PuPu-5′

5′-PuPu-TCTAGA-PyPu-3′
3′-PyPy-AGATCT-PuPy-5′

As shown above, dodecamer and decamer DNA constructs used in ITC and NMR studies, respectively, present different peripheral DNA sequences (top) to the ligand compared with those of the two possible 15-mer binding sites (middle and bottom; here Pu = purine and Py = pyrimidine). NOE data indicate hydrophobic contact taking place between the Dp tail and terminal residues in the decamer duplexes. These are clearly important for stable complex formation. It is entirely reasonable to suggest that such a hydrophobic interaction could be disrupted due to sequence variations peripheral to the binding footprint, which for the TCTAGA sequence are sufficient to prevent binding occurring at all. The implication is that the binding site extends over a larger DNA footprint. Further value therefore lies in testing the ability of **1** to select for these recognition sequences as a function of site ordering along the DNA

sequence and this and related studies form a current focus for our attention.

Such selectivity is also bound up with hydrogen bonding capability and the electrostatic potentials of electron donors and acceptors lining the edges of the DNA-bases along the DNA minor groove floor. For 5′-ACTAGT⁸-3′, hydrogen-bond formation is predicted between T⁸-O2 and N<u>H</u>26 of **1**. Former structure calculations[16b] place the distance between these two atoms at 1.8 Å. Rotation by 180° of the A·T base pairs flanking the central CTAG sequence to produce 5′-TCTAGA⁸-3′ would result in A⁸-N3 becoming the hydrogen bond partner of N<u>H</u>26 at a distance of 2.3 Å (Fig. S9 in the Supporting Information†). Although the negative electrostatic potential of A-N3 is greater than that of T-O2 in the context of a DNA double helix, the difference between the two values is relatively small.[24] Additional electronic factors including polarization and delocalisation may also be considered. These factors together cause subtle selection of 5′-ACTAGT-3′ over 5′-TCTAGA-3′ in the context of **1** binding to 15-mer duplex **7**.

Related *in silico* footprinting has been used to predict the difference in behaviour between the binding of **1** to **3** (CCTAG*G*) compared with the binding of **1** to **6** (CCTAG*I*).[20] The latter complex is favoured of the two, despite the only difference being the presence of NH_2 (in guanosine) in place of H (in inosine). Implicit molecular dynamics (MD) also distinguishes the binding of **1** to the allied recognition sequences 5′-ACTAGT-3′ and 5′-TCTAGA-3′, the former being favoured over the latter.[23] These results align with those reported here and confirm the viability of using *in silico* methods for rational ligand design in this context.

Thermodynamics of binding

Differences observed in the thermodynamics of **1** binding to ODNs **2′–6′** arise from variations in their ability to form non-covalent interactions with the ligand. Changes in the geometry of the hydrogen bond and van der Waals contacts between ligand and DNA bases would explain the higher ligand affinity shown for *ACTAGT vs. TCTAGA*. This conclusion agrees with the value of ΔH for **1** associating with *ACTAGT* ($\Delta H = -12.8$ kcal mol⁻¹) being more favourable than the association between **1** and *TCTAGA* ($\Delta H = -10.4$ kcal mol⁻¹) and suggests that stronger non-covalent interactions are formed in the former case. Unfavourable (negative) entropy changes observed for **1** binding to *ACTAGT*, *TCTAGA* and *GCTAGC* suggest an induced fit interaction by which entropically unfavourable conformational changes in either or both species allow better hydrogen bond contacts to form between ligand and DNA duplex.[25,26] Such phenomena together with allosteric effects are commonplace in the regulation of biological processes.[6,27] The association itself constrains the complex, generating an entropic penalty through losses in rotational and translational degrees of freedom.[28] These findings are consistent with NMR data, which indicate structural alterations in the DNA backbone as a result of minor groove ligand binding.[16b]

The amino groups of guanine residues protrude from the floor of the DNA minor groove and play an important role in assisting the binding event when hydrogen bonds are formed with the thiazole nitrogen. When a steric clash occurs between a guanine amino group and ligand (as for CCTAGG⁸ when ligand Dp tail

encounters G^8-NH$_2$, Figs S8 and S9 in the Supporting Information†) binding is opposed and affinity is reduced. The resulting substantial loss in enthalpy decreases binding affinity despite the favourable entropy contribution to ΔG. Speculation on the causes centre around gains in the conformational freedom of either ligand and/or DNA as a result of weaker non-bonded interactions formed in the complex. These are not enough to constrain or change the complex structure. A significant increase in the binding free energy was observed when G^8 in CCTAGG^8 was replaced by inosine in CCTAGI^8. This change in free energy was mainly enthalpic in origin ($\Delta H = -9.6$ kcal mol^{-1}) with a small entropic contribution ($-T\Delta S = -1.5$ kcal mol^{-1}). One could attribute this to release of more water and/or counter ions resulting from the ligand fitting more snugly within the groove in the absence of an exocyclic G-NH$_2$. Binding to the minor groove of GC-rich sequences is expected to release more water molecules and induce fewer conformational changes in the DNA structure.

For ligand–DNA interactions, favourable (positive) entropy arises from desolvation of the binding interface. Here the process is driven primarily by enthalpy, indicating that hydrogen bonding and van der Waals forces dominate the interaction. Such sequences have a widened minor groove capable of accommodating ligand without inducing large perturbations in DNA structure. This may explain why sequences such as CCTAGI have favourable entropy while others do not.

Binding *via* enthalpic drive with entropic opposition

ITC studies revealed an exothermic process when **1** was titrated with DNA dodecamers **2′–6′** combined with an endothermic dilution process, suggesting ligand aggregation prior to DNA binding. Favourable enthalpy of interaction is consistent with observations that exothermic processes occur for most DNA binding agents at room temperature.[29] The contribution that water and ions make to the formation of ligand–DNA complexes has been noted previously and occurs through several processes. Disruption of the solvent cage around a nucleic acid is entropically favourable, promoting binding affinity.[29,30] Bridging between a nucleic acid and the targeting ligand through hydrogen bond formation is enthalpically favourable and assists complex formation.[31–34] A favourable change in enthalpy and unfavourable or slightly favourable change in entropy for the complexes studied in this work could implicate water-assisted complex formation. Enthalpy changes reflect the strength of the non-covalent interactions between molecules relative to those existing with the solvent[35] and, as noted previously, knowledge of the changes in hydration state upon complex formation is needed for a complete picture of binding processes to be built up.[29] This requires volumetric measurements using densimetry, which is beyond the scope of the current work. Such a solvent-based entropic contribution to binding may not be evident if it is offset by greater opposing factors, such as the loss of conformational freedom, which is also entropic. Assessment of a large amount of calorimetric data on specific groove-binding and intercalating ligands concluded that minor groove binding was mainly entropically driven and was unrelated to structure.[29] Recent studies of HIV-1 protease inhibitors have found that a slight modification to the structure of a ligand can lead to completely different thermodynamic profiles.[36] The previous suggestion that minor groove-binding could be mainly entropically driven was based on a limited number of examples and was qualified by the need to critically test the hypothesis by expanding the database of thermodynamic data for minor groove binders. The studies detailed here for thiazotropsin A provide important new insights into this process and indeed extend the thermodynamic database, considerably adding to the significance of our findings in this work. Our results align with those reported for distamycin binding in the 2 : 1 mode, driven largely by favourable enthalpy contributions ($\Delta H = -15.7$ kcal mol^{-1}) and significantly opposed by entropic contributions ($T\Delta S = -7.8$ kcal mol^{-1}).[30] This also suggests that electrostatic effects are not a major contributor to the driving force of interaction, since ligand–DNA interactions, which have large contributions from hydrophobic and electrostatic forces, are largely driven by entropy due to the release of water and counter ions from the polyanion DNA duplex upon ligand binding.[30,37] The published literature to date has shown that minor groove recognition by small molecules can be enthalpically or entropically driven or both and that the thermodynamic signature of MGBs is highly dependent on ligand structure and the sequence of the binding site, which all of our studies with different sequences and related ligands confirm. Further detailed studies along similar lines to those reported here are required on related systems in order to widen our understanding of the role played by thermodynamics for alternative types of DNA minor groove binding ligands, such as those represented by **1**.

Analysis of the enthalpies of **1** binding to five dodecamer ODNs at ligand:ODN ratios (r) varying between 0–7 reveals that when $r \leq 2$ the enthalpy of binding to the dodecamer remains constant. Since **1** spans about six base pairs when bound to a DNA duplex,[38] the dodecamer potentially provides two consecutive binding sites (a feature confirmed in the case of 4 equivalents of **1** binding to d(CG*ACTAGT*A**CTAGT**CG)). NMR data conclusively confirms that **1** binds exclusively to one site as a dimer. In the case of monomers, a noticeable difference in the enthalpy of binding to the first and the second binding sites would be expected due to the different base pair sequences of the sites. ΔH remaining constant when $r \leq 2$ is highly suggestive of binding taking place exclusively in a dimeric 2 : 1 mode fully consistent with all of our observations.

Although different forces may contribute to ΔH, the results presented here suggest that hydrogen bonds and van der Waals interactions are the main molecular forces that contribute to ΔH for **1** associating with DNA when high affinity complexes are formed.

Experimental section

Materials

N-[3-(Dimethylamino)propyl]-2-({[4-({[4-(formylamino)-1-methyl-1H-pyrrol-2-yl]carbonyl}amino)-1-methyl-1H-pyrrol-2-yl]carbonyl}-amino)-5-isopropyl-1,3-thiazole-4-carboxamide (thiazotropsin A, **1**) was prepared as the TFA salt.[39] For NMR studies the self complementary ODNs d(CG*A*CTAG*T*CG)$_2$, **2**, d(CG*C*CTAG*G*CG)$_2$, **3**, d(CG*T*CTAGA*CG)$_2$, **4**, d(CG*G*CTAG*C*CG)$_2$, **5**, d(CG*C*CTA-G*I*CG)$_2$, **6** and 15-mer duplex d(CGACTAGTCTAGACG)·d(CGTCTAGACTAGTCG), **7**, were supplied by Alpha DNA

Ltd. (Montreal, Canada) as desalted, cartridge-purified, ethanol precipitated, lyophilised powders, which were used without further purification. For ITC and CD measurements, the self-complementary dodecameric ODNs d(GCG*A*CTAG*T*CGC)₂, **2**′, d(GCG*C*CTAG*G*CGC)₂, **3**′, d(GCG*T*CTAG*A*CGC)₂, **4**′, d(GCG*G*CTAG*C*CGC)₂, **5**′ and d(GCG*C*CTAG*I*CGC)₂, **6**′, were purchased from MWG-BIOTECH AG (Anzinger str. 7a, D-85560 Ebersberg, Germany) as HPLC-purified salt-free ODNs custom synthesized on a 1 μmol scale. Millipore-filtered water was used in the preparation of all solutions. PIPES (piperazine-1, 4-bis (2-ethanesulfonic acid), EDTA, and sodium chloride used to prepare the buffers were purchased from Sigma Aldrich (Poole, Dorset, UK).

NMR spectroscopy

NMR data were acquired on Varian Unity *INOVA* 600, Unity Plus 500 and Bruker AMX and DPX 400 NMR spectrometers operating at 599.89, 500.06 and 400.13 MHz for ^1H resonance, respectively, and were acquired at a probe temperature of 298 K, unless otherwise stated. A standard geometry triple resonance probehead equipped for z-pulsed field gradients was used for NMR work at 500 and 600 MHz; 5 mm dual (BBO, [X/^1H]) or QNP-z [^1H,^{13}C,^7Li,^{31}P] probeheads were used for data acquisitions carried out at a magnetic field strength of 9.4 T.

Isothermal titration calorimetry (ITC)

ITC titrations were performed on a Microcal VP-ITC system and analysed using Origin (version 7.0, OriginLab, Massachusetts). All samples were degassed in a Microcal Thermovac sample degasser for 20 min prior to use to decrease noise and achieve stable baselines. Experiments were performed at pH 6.8 in PIPES buffer (10 mM PIPES, 20 mM sodium chloride, and 1 mM EDTA) at 298 K. The instrument was equilibrated at the start of each experiment by means of a known electrical pulse until a stable baseline was obtained. The heats of ligand dilution were subtracted from the heat measured for the binding interactions and the corrected data were used to determine the binding enthalpies.

Circular dichroism (CD)

CD experiments were conducted using an Applied Photophysics Chirascan spectrophotometer at 25 °C in a 1 cm quartz cuvette. To the ODNs (5 μM, 1.0 mL) in pH 6.8 PIPES buffer (10 mM PIPES, 20 mM NaCl and 1 mM EDTA) ligand solutions were added (0.5 mM) in 1 μL increments to a total of 20 additions. At each titration point the molar ellipticity was measured between 240–380 nm using a bandwidth of 1 nm. Binding constants were calculated by non-linear least squares fitting of Engel's equation[40] for tight ligand binding to the CD data.

Molecular modelling

Qualitative modelling was carried out on the basis of the previously defined structure of **1** with **2** (PDB accession code 1RMX, rscb020892). Inter-proton distance restraints were calculated for ODN-1 interactions from 100 ms NOESY NMR data sets on the basis of an isolated spin pair approximation (ISPA). Models based on 1RMX were created within Sybyl (version 6.3, Tripos Inc.) running on a Silicon Graphics O2 R12000 workstation operating under IRIX 6.5. Energy minimised structures were compared with one another and the NOE data were checked for consistency in each case. Full and comprehensive details of all experimental parameters and methods may be found in the Supporting Information†§.

Conclusions

NMR spectroscopy, isothermal titration calorimetry and circular dichroism have been used to investigate the structural and energetic requirements for recognition of the DNA minor groove by thiazotropsin A, **1**, in a variety of different DNA contexts. Experimental findings are in agreement with theoretical predictions and show that the fine details of DNA sequence context are directly responsible for the subtle way in which ligands recognize their minor groove targets. In the context 5′-XCTAGZ-3′, **1** was shown to be capable of binding when Z = I if X = C, thereby implicating the exocyclic guanosine NH₂ in a steric clash that causes the dramatic drop in the quality of NMR data observed for the complex between **1** and d(CG*CC*TAG*G*CG)₂ compared with related DNA recognition sequences and reflected in the significant change observed in the thermodynamic parameters measured by ITC and CD. Despite the fact that the same ligand is able to bind separately to both 5′-*A*CTAG*T*-3′ and to 5′-*T*CTAG*A*-3′, the preference of ligand binding for the former recognition sequence when both are concatenated is in agreement with both modelling predictions and energetic analysis. Even though the flanking base pairs are only turned around by 180°, this subtly influences the binding potential of the recognition sequence, being ascribed to differences in hydrogen bond distances, electrostatic potentials and related characteristics that when factored together result in a preference for 5′-*A*CTAG*T*-3′ over 5′-*T*CTAG*A*-3′. Although scope exists for ligands to bind individually to different DNA sequences within the context d(CG*A*CTAG*TC*TAGA*C*G)·d(CG*T*CTAG*A*CTAG*T*CG), antiparallel side-by-side dimer association of ligand molecules to one binding site continues to be the preferred mode of binding.

The distinct thermodynamic signature of thiazotropsin A-DNA interactions adds important new information to the thermodynamic database allowing a more complete energetic and structural picture to be built for different types of DNA minor groove binding molecules. In this work, the large favourable enthalpy (negative) indicates formation of a large number of favourable hydrogen bond contacts and/or van der Waals interactions between the DNA and ligand. The unfavourable entropy (negative) suggests a conformational change in both or either of the molecules that produce a more restrained complex through an 'induced fit' process. This would appear to outweigh any solvent rearrangement, desolvation, release of counterions or hydrophobic drive that is characterized by favourable entropy measurements observed with some MGBs.[30] Establishing a link between the energetics of binding and structure is important when trying to understand biomolecular interactions and improving the binding affinity. In this study, it is evident that despite the NMR spectra of the complexes showing remarkable consistency, the thermodynamic

evaluations clearly show that small changes impact on the binding energetic at the molecular level. Such differences would not have been revealed by a structural study alone. However, the relationship is complex: improving the binding enthalpy does not necessarily lead to a higher binding affinity because of enthalpy–entropy compensation, leading to no net increase in affinity. A major cause of this compensation mechanism is the nature of non-covalent interactions; the enthalpic gain *via* hydrogen bond formation within a complex being offset by entropic losses since these new bonds limit movement within the complex.

Speculation has been made on the role that side-by-side association plays in the recognition of the DNA minor groove, but more information is required in order to understand the energetic penalties associated with minor groove binding in the context of ligand self-association/dissociation. Further work is being carried out in this context within our laboratory and we will report the results of our findings in due course.

Acknowledgements

Notes and references

§ Full description of ESI: Full experimental details; ^{1}H NMR resonance assignments for ODNs **4–7**: Tables S1–S4;^{1}H NMR resonance assignments for **1** bound to ODNs **2**, **4–7**: Table S5; revised ^{1}H NMR resonance assignments for the complex between **1** and **2**: Table S6; ^{1}H NMR assignments for complexes between **1** and **4–6**: Tables S7–S10; partial ^{13}C NMR chemical shift assignments for the complex between **1** and **6**: Table S11; chemical shift differences for ligand-bound and ligand-free DNA duplexes **2**, **4–7**: Tables S12–S16; inter-ligand and ligand-DNA inter-proton NOEs for complexes between **1** and **4–7**: Tables S17–20; finger-print region of the 600 MHz 100 ms 2D [^{1}H, ^{1}H] NOESY NMR spectrum of the complex between **1** and **6**: Fig. S1; part of the natural abundance 2D [^{1}H, ^{13}C] HSQC NMR spectrum of the complex between **1** and **6**: Fig. S2; 400 MHz 1D ^{1}H NMR data of the titration of **5** with **1**: Fig. S3; ^{1}H NMR data monitoring the titration of **3** with **1**: Fig. S4; comparison of a section of the 2D [^{1}H, ^{1}H] NOESY NMR data for **1** with **3** *vs.* **1** with **5**: Fig. S5; revised assignment of the fingerprint region of the 2D [^{1}H, ^{1}H] NOESY NMR spectrum of **1** in complex with **2**: Fig. S6; 2D [^{31}P, ^{1}H] COSY NMR data for the complex of **1** with **4**: Fig. S7; modelling images showing the steric clash and relief for G[8] and I[8] in complexes of **3** and **6** with **1**: Fig. S8; a comparison of models showing steric clash and the effects of base pair re-orientation in connection with the ligand Dp tail: Fig. S9; 1D ^{31}P-{^{1}H} NMR titration monitoring of **1** binding to self-complementary 16-mer d(CGA**CTAGT**A**CTAGT**CG)₂ to form a 4 : 1 (ligand:duplex) complex: Fig. S10; heats of dilution of **1** in PIPES buffer: Fig. S11; a schematic representation of potential binding sites for **1** on the 15-mer construct duplex: Fig. S12.

1 J. G. Pelton and D. E. Wemmer, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 5723.
2 P. B. Dervan, *Bioorg. Med. Chem.*, 2001, **9**, 2215.
3 J. W. Trauger, E. E. Baird and P. B. Dervan, *Nature*, 1996, **382**, 559.
4 J. M. Gottesfeld, L. Neely, J. W. Trauger, E. E. Baird and P. B. Dervan, *Nature*, 1997, **387**, 202.

5 For reviews see: (*a*) D. E. Wemmer, *Biopolymers*, 2001, **52**, 197; (*b*) S. Neidle, *Nat. Prod. Rep.*, 2001, **18**, 291.
6 D. M. Chenoweth and P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 13175.
7 D. E. Wemmer and P. B. Dervan, *Curr. Opin. Struct. Biol.*, 1997, **7**, 355.
8 M. A. Marques, R. M. Doss, A. R. Urbach and P. B. Dervan, *Helv. Chim. Acta*, 2002, **85**, 4485.
9 P. B. Dervan and B. S. Edelson, *Curr. Opin. Struct. Biol.*, 2003, **13**, 284.
10 C. J. Suckling, *J. Phys. Org. Chem.*, 2008, **21**, 575.
11 For a general reference on biophysical methods applied to drug–DNA interaction studies see Drug–DNA Interaction Protocols, *Methods in Molecular Biology*, ed. K. R. Fox, Humana Press Inc., Totowa, NJ, USA, 2010, vol. 613.
12 A. Rodger and B. Norden in *Circular Dichroism and Linear Dichroism*, Oxford Chemistry Masters, Oxford University Press, 1997, pp. 160.
13 J. Barber, H. F. Cross and J. A. Parkinson "High-Resolution NMR of DNA and Drug-DNA Interactions" in *Spectroscopic Methods and Analyses: NMR, Mass Spectrometry and Metalloprotein Techniques, Methods in Molecular Biology*, ed. C. Jones, B. Mulloy and A. H. Thomas, Humana Press Inc., Totowa, NJ. USA, 1993, vol. 17, pp. 87–113.
14 *Nucleic Acid Targeted Drug Design*, eds C. L. Propst and T. J, Perun, Marcel Dekker Inc., New York, USA, 1992.
15 Y.-P. Pang, *Bioorg. Med. Chem.*, 2004, **12**, 3063.
16 (*a*) N. G. Anthony, K. R. Fox, B. F. Johnston, A. I. Khalaf, S. P. MacKay, I. McGroarty, J. A. Parkinson, G. G. Skellern, C. J. Suckling and R. D. Waigh, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 1353; (*b*) N. G. Anthony, B. F. Johnston, A. I. Khalaf, S. P. MacKay, J. A. Parkinson, C. J. Suckling and R. D. Waigh, *J. Am. Chem. Soc.*, 2004, **126**, 11338.
17 W. Treesuwan, K. Wittayanarakul, N. G. Anthony, G. Huchet, H. Alniss, S. Hannongbua, A. I. Khalaf, C. J. Suckling, J. A. Parkinson and S. P. MacKay, *Phys. Chem. Chem. Phys.*, 2009, **11**, 10682.
18 A. Blaskó and T. C. Bruice, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 10018.
19 P. L. James, E. E. Merkina, A. I. Khalaf, C. J. Suckling, R. D. Waigh, T. Brown and K. R. Fox, *Nucleic Acids Res.*, 2004, **32**, 3410.
20 N. G. Anthony, G. Huchet, B. F. Johnston, J. A. Parkinson, C. J. Suckling, R. D. Waigh and S. P. Mackay, *J. Chem. Inf. Model.*, 2005, **45**, 1896.
21 J. A. Parkinson, J. Barber, K. T. Douglas, J. Rosamund and D. Sharples, *Biochemistry*, 1990, **29**, 10181.
22 We are in the process of studying in detail the influence of ligand self-assembly on DNA minor groove recognition and will subsequently report the details of our observations in this regard in a further article. For the purposes of this article, however, the observation of negative NOEs for thiazotropsin A in free aqueous solution at room temperature indicates that this molecule self-assembles as a result of hydrophobic and π–π interactions.
23 K. Wittayanarakul, N. G. Anthony, W. Treesuwan, S. Hannongbua, H. Alniss, A. I. Khalaf, C. J. Suckling, J. A. Parkinson and S. P. MacKay, *ACS Med. Chem. Lett.*, 2010, **1**, 376.
24 B. Pullman, D. Perahia and D. Cauchy, *Nucleic Acids Res.*, 1979, **6**, 3821.
25 P. D. Kwong, M. L. Doyle, D. J. Casper, C. Cicala, S. A. Leavitt, S. Majeed, T. D. Steenbeke, M. Venturi, I. Chaiken, M. Fung, H. Katinger, P. W. Parren, J. Robinson, D. Van Ryk, L. Wang, D. R. Burton, E. Freire, R. Wyatt, J. Sodroski, W. A. Hendrickson and J. Arthos, *Nature*, 2002, **420**, 678.
26 R. O'Brien and I. Haq Applications of Biocalorimetry: Binding, Stability and Enzyme Kinetics, *Biocalorimetry 2*, ed. J. E. Ladbury and M. Doyle, John Wiley and Sons, Ltd., 2004.
27 P. A. Borea, K. Varani, S. Gessi, P. Gilli and G. Gilli, *Biochem. Pharmacol.*, 1998, **55**, 1189.
28 R. S. Spolar and M. T. Record, Jr, *Science*, 1994, **263**, 777.
29 J. B. Chaires, *Arch. Biochem. Biophys.*, 2006, **453**, 26.
30 I. Haq, *Arch. Biochem. Biophys.*, 2002, **403**, 1.
31 S. Bergqvist, M. A. Williams, R. O'Brien and J. E. Ladbury, *J. Mol. Biol.*, 2004, **336**, 829.
32 M. W. Freyer, R. Buscaglia, A. Hollingsworth, J. Ramos, M. Blynn, R. Pratt, W. D. Wilson and E. A. Lewis, *Biophys. J.*, 2007, **92**, 2516.

33 M. W. Freyer, R. Buscaglia, B. Nguyen, W. D. Wilson and E. A. Lewis, *Anal. Biochem.*, 2006, **355**, 259.

34 J. W. Schwabe, *Curr. Opin. Struct. Biol.*, 1997, **7**, 126.

35 J. E. Ladbury, G. Klebe and E. Freire, *Nat. Rev. Drug Discovery*, 2010, **9**, 23.

36 V. Lafont, A. A. Armstrong, H. Ohtaka, Y. Kiso, L. Mario Amzel and E. Freire, *Chem. Biol. Drug Des.*, 2007, **69**, 413.

37 I. Haq, J. E. Ladbury, B. Z. Chowdhry, T. C. Jenkins and J. B. Chaires, *J. Mol. Biol.*, 1997, **271**, 244.

38 S. A. Harris, E. Gavathiotis, M. S. Searle, M. Orozco and C. A. Laughton, *J. Am. Chem. Soc.*, 2001, **123**, 12658.

39 A. I. Khalaf, C. J. Suckling, R. D. Waigh British Patent Application, 2002, PCT/GB02/05916.

40 G. Engel, *Anal. Biochem.*, 1974, **61**, 184.