Thiazotropsin aggregation and its relationship to molecular recognition in the DNA minor groove

Marie-Virginie Salvia[a], *Fiona Addison*[a], *Hasan Y. Alniss*[b†], *Niklaas J. Buurma*[d], *Abedawn I. Khalaf*[a], *Simon P. Mackay*[b], *Nahoum G. Anthony*[b], *Colin J. Suckling*[a], *Maxim P. Evstigneev*[c], *Adrián Hernandez Santiago*[e], *Roger D. Waigh**[b] *and John A. Parkinson**[a]

[a] WestCHEM Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow G1 1XL, United Kingdom, [*]Corresponding author Dr. John A. Parkinson: Fax: +44 141 548 4822; Tel: +44 141548 2820; E-mail: john.parkinson@strath.ac.uk

[b] Strathclyde Institute of Pharmaceutical and Biomedical Sciences, University of Strathclyde, 161 Cathedral Street, Glasgow G4 0RE, United Kingdom.

[c] Department of Physics, Sevastopol National Technical University, Sevastopol 99053, Crimea, Ukraine.

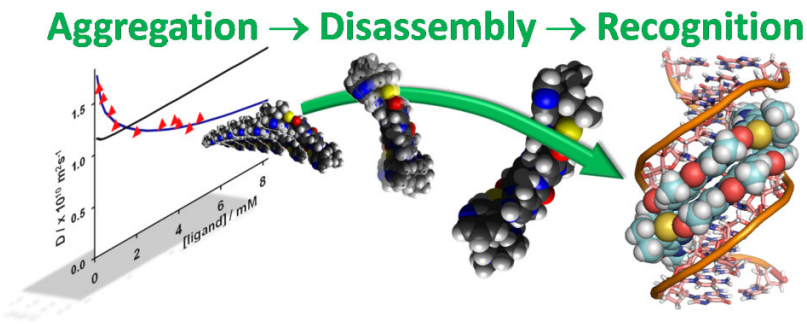
[d] School of Chemistry, University of Cardiff, Main Building, Park Place ,Cardiff CF10 3AT, United Kingdom.

[e] Department of Physics and Mathematics, Faculty of Chemistry, Autonomous University of Puebla, Puebla, Mexico CP 72570.

*Corresponding author: john.parkinson@strath.ac.uk. [†]Current address: College of Pharmacy, An-Najah National University, Nablus, Palestine.

ABSTRACT:

Aggregated states have been alluded to for many DNA minor groove binders but details of the molecule-on-molecule relationship have either been under-reported or ignored. Here we report our findings from ITC and NMR measurements carried out with AIK-18/51, a compound representative of the thiazotropsin class of DNA minor groove binders. The free aqueous form of AIK-18/51 is compared with that found in its complex with cognate DNA duplex d(CGACTAGTCG)$_2$. Molecular self-association of AIK-18/51 is consistent with anti-parallel, face-to-face dimer formation, the building block on which the molecule aggregates. This underlying structure is closely allied to the form found in the ligand's DNA complex. NMR chemical shift and diffusion measurements yield a self-association constant K_{ass} = (61 ± 19) × 10^3 M^{-1} for AIK-18/51 that fits with a stepwise self-assembly model and is consistent with ITC data. The deconstructed energetics of this assembly process are reported with respect to a design strategy for ligand/DNA recognition.



HIGHLIGHTS:

Aggregate assembly of a representative thiazotropsin MGB were studied by NMR and ITC.

Diffusion NMR measurements concur with numerical analysis of chemical shift data.

Molecular self-association occurs in a head-to-tail, face-to-face mode of assembly.

Cognate DNA recognition by the ligand goes *via* a closely allied, self-assembled state.

KEYWORDS: Ligand-assembly, aggregation, ITC, NMR, self-diffusion, molecular recognition, DNA minor groove.

1. Introduction

Molecular recognition between the DNA minor groove and classical small molecule DNA binders including Hoechst 33258 [1-234567] berenil [8,9] netropsin [10] distamycin A [11,12] and their derivatives (**Fig. 1**) typically occurs as either 1:1 or 2:1 ligand:DNA complexes [13-,14,15]. For aromatic peptide minor groove binders (MGBs) based on heterocyclic monomers such as *N*-methylpyrrole (Py) and *N*-methylimidazole (Im), binding in either 1:1 or 2:1 modes is governed by the DNA minor groove width, which is DNA sequence dependent [16,17]. Non-equivalent monomers are capable of binding to non-complementary DNA recognition sites when arranged anti-parallel to one another in a face-to-face manner [16,18-1920] and when joined either mid-molecule *via* polycarbon linker to create a sandwich dimer [21] or through intervening γ-aminobutyric acid residues to create hairpin [22,23] or cyclic [24-25,26] polyamide units. Hairpin and cyclic polyamides in particular display DNA binding association constants suitable for a programmed development of gene-targeting drugs [27,28], have cell penetration characteristics and promising biological potential [29,30]. Notwithstanding these advantages, the cyclic and hairpin MGBs are relatively large molecules supporting a case for the continued development of unlinked MGBs as smaller, simpler synthetic molecules. With scope arguably for more accessible commercial scale production together with other advantages associated with the smaller size of unlinked MGBs, such molecules remain attractive alternatives as potential therapeutic molecules, for example to treat infectious disease, or biotechnological tools.

A drawback to the unlinked strategy is that 2:1 ligand-DNA binding occurs with lower association constants for unlinked self-assembled dimers compared with cyclic or hairpin MGBs together with a potential entropic disadvantage. The symmetry of anti-parallel, face-to-face ligand assembly in the DNA minor groove allows recognition of self-complementary DNA sequences by such dimer assemblies. DNA sequences of this type often occur at gene control sites making such a process pertinent for applications in many fields [31]. Our MGBs of this type possess strong anti-bacterial, anti-viral, and

anti-parasitic properties and one group of compounds has entered preclinical development as antibacterial drugs [32]. Generalization of such a recognition process, as demonstrated previously for hetero-dimer assemblies, would allow tailored recognition to be targeted at specific, non-palindromic DNA binding sites akin to the behaviour of linked MGBs [16,18,28]. Increasing the strength of association between homo- or hetero-dimer partners might allow such molecules to be favourably compared with their hairpin or cyclic counterparts. Designing enhanced association behaviour into potential ligands would therefore be valuable but for control purposes requires an understanding of both the structural and energetic factors that govern ligand self-assembly outside the context of DNA-binding.

Our focus for such investigations lies with the thiazotropsins (**Fig. 1**). These constitute a class of biologically active molecule known to form tight binding complexes within the DNA minor groove [32-, 33, 34, 35]. The so-called 'Strathclyde Strategy' has evolved increasingly lipophilic character within its ligand design program resulting in the type of molecules shown. This feature not only assists with cell permeation but interestingly also provides scope for the assembly characteristics of pairs of ligands to be enhanced and tailored through additional hydrophobic interactions that can also occur between a DNA target and the MGB [36]. Our investigations have led us to consider the detailed structural and energetic characteristics of the unlinked homo-dimer MGBs belonging to this class of molecule in both the presence and absence of DNA, with a view to determining how the structure of ligand-on-ligand assembly compares between the two processes. Many studies have alluded to the propensity for DNA minor groove binding ligands to aggregate in aqueous solution with experimental evidence supporting face-to-face contact between near neighbours [37-3839]. The effects of aggregation may be alleviated by dilution which is suggested to favour monomers [40]. Our premise for the thiazotropsins is that a preferred arrangement of ligands may be the dimer rather than monomer and that aggregates may be constituted from assemblies of dimers. To our knowledge, no previous study has attempted to assess in detail the structural (including the relative orientation of neighboring ligands) and energetic relationship

between DNA binding ligands themselves in the absence of DNA *and* made a direct comparison with the ligand assembly structure when bound to the ligand's cognate DNA sequence. In this article we compare the overall energetics for MGB-DNA complex formation, for which a full theoretical consideration has been presented recently [41], with those of ligand-on-ligand assembly. Using NMR spectroscopy, isothermal titration calorimetry (ITC) and molecular modeling, we describe here a study of the self-assembly characteristics of AIK-18/51 (**Fig. 1**) representing one member of the thiazotropsin class of MGBs. The study is combined with an assessment of its complex-forming character with the DNA duplex d(CGACTAGTCG)$_2$. Our findings show that when enhanced lipophilicity is designed into MGBs, strong face-to-face association is apparent that is crucially of an anti-parallel nature in aqueous solutions for the free MGB in a manner directly akin to the structure found in the MGB-DNA complex. These observations suggest pre-association of ligand dimers in the form required for DNA binding. By such means, a lowered entropic penalty is expected for DNA binding compared with stepwise cooperative binding of monomer ligands, opening the way towards development of small MGBs with substantially enhanced DNA binding capability.

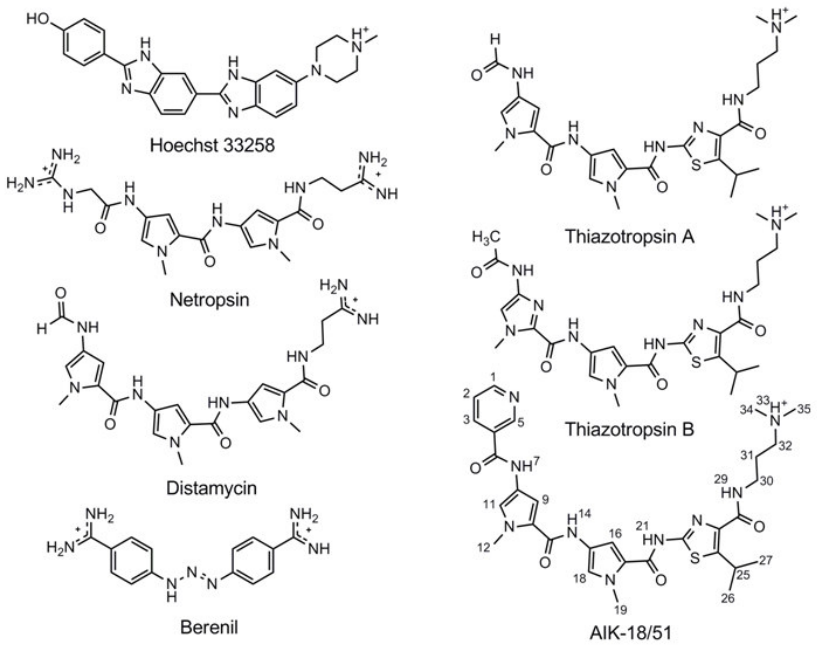


Fig. 1. The chemical structures of Hoechst 33258, Netropsin, Distamycin A, Berenil and examples of the thiazotropsin class of DNA minor groove binding ligand including Thiazotropsin A, Thiazotropsin B and AIK-18/51.

2. Materials and Methods

2.1 Materials

3-(5-isopropyl-2-(1-methyl-4-(1-methyl-4-(nicotinamido)-1*H*-pyrrole-2-carboxamido)-1*H*-pyrrole-2-carboxamido)-thiazole-4-carboxamido)-*N*,*N*-dimethylpropan-1-aminium TFA salt, AIK-18/51 (MW = 733.8 g.mol^{-1}) was synthesized at the University of Strathclyde according to previously reported procedures [32,42]. Piperazine-1,4-bis(2-ethanesulphonic acid), PIPES, 2-(carbamoylmethylamino)-ethanesulfonic acid, ACES, sodium chloride, NaCl and ethylenediaminetetraacetic acid, EDTA, were supplied by Sigma-Aldrich and used without further purification. DNA (sequence d(CGACTAGTCG)) was synthesized on a 15 μmol scale and supplied (1.14 micromoles) as a HPLC purified, dry lyophilized powder by AlphaDNA (Montreal, Quebec, Canada), which was used without further purification. Details of all buffer and sample preparations are supplied in the Supporting Information.

Briefly, ligand solutions were prepared for ITC at a concentration of 5 mM and for NMR studies, separate solutions were prepared for resonance assignment studies (1 mM,0.43 mg, 0.58 μmol, 550 μL 90% H_2O/ 10% D_2O) and for the purposes of self-association studies (5 mM, 2.02 mg, 2.75 μmol, 550 μL, 100 mM phosphate buffer solution, pH = 7.4 in 99.9% D_2O). Formation of a 2:1 ligand:DNA duplex complex was monitored by 1D ^{1}H NMR spectroscopy through titration experiments.

2.2 ITC Measurements

ITC measurements were carried out using a MicroCal VP-ITC instrument equipped with a sample cell volume of 1.4399 mL and running under Vp Viewer ITC (2000) software control. Ligand association titration experiments were carried out from a solution of AIK-18/51 at a concentration of 5

mM in PIPES buffer. Binding constant, K_{ass}, and enthalpy of association, ΔH_{ass} parameters were evaluated using the software application IC-ITC [43].

2.3 Diffusion Modeling

Self-assembly models were used to assess the nature of the extended mode of aggregation for AIK-18/51 based on NMR diffusion measurements. Theoretical models (Equations 1 and 2) were used for fitting, which link diffusion coefficient and concentration, where K represents the binding constant, K_{agg} (where $K_{agg} = 2 \times K_{dim}$, $mol^{-1}.dm^3$, evaluated *via* numerical analysis), [compound] is the concentration of material, k is Boltzmann's constant, T is temperature (298 K), η is viscocity (0.00111 Pa.s, equivalent to the viscosity of D_2O at 25 °C), and V is the monomer volume (adjusted for optimal curve fitting). The number of terms, n, refers to the summation runs over all aggregates from monomers to n-mers with a limiting value of $n = 50$.

$$D_{obs} = \sum_{n=1}^{n=\infty}\{x_n \cdot D_i\} \approx \sum_{n=1}^{n=terms}\left\{\frac{n \cdot K^{n-1} \cdot \left[\left\{\frac{\left(2 \cdot K \cdot [compound]_{total} + 1 - \sqrt{(4 \cdot K \cdot [compound]_{total} + 1)}\right)}{(2 \cdot K^2 \cdot [compound]_{total})}\right\}^n\right]}{[compound]_{total}} \cdot \frac{k \cdot T}{6 \cdot \pi^{0.666} \cdot \eta \cdot \{\frac{3}{4} \cdot n \cdot V_1\}^{0.333}}\right\} \quad \text{Eq. 1}$$

$$D_{obs} = \sum_{n=1}^{n=2}\{x_n \cdot D_i\} = \frac{k \cdot T}{6 \cdot \pi^{0.666} \cdot \eta \cdot \{\frac{3}{4} \cdot 1 \cdot V_1\}^{0.333}} \cdot \left(\frac{\left(\frac{\left(-1 + \sqrt{(1 + 8 \cdot K \cdot [compound]_{total})}\right)}{(4 \cdot K)}\right)}{[compound]_{total}}\right) + \frac{k \cdot T}{6 \cdot \pi^{0.666} \cdot \eta \cdot \{\frac{3}{4} \cdot 2 \cdot V_1\}^{0.333}} \cdot 2 \cdot \left(\frac{\left\{K \cdot \left(\frac{\left(-1 + \sqrt{(1 + 8 \cdot K \cdot [compound]_{total})}\right)}{(4 \cdot K)}\right)^2\right\}}{[compound]_{total}}\right) \quad \text{Eq. 2}$$

Monomer volumes were adjusted to obtain the best fit of the experimental data. Determination of the optimum volume, V, was carried out by minimizing the sum of the squares of the error, being the difference between theoretical and experimental values.

2.4 NMR Spectroscopy

NMR data were acquired on a Bruker AVANCE III 600 NMR spectrometer operating at a proton resonance frequency of 600.13 MHz (14.1 T magnetic field strength) using a TBI-z probehead and running under Topspin (version 2.1, Bruker Biospin, Karlsruhe). All data were accumulated under full automation using IconNMR. Ligand ^{1}H NMR signal assignments were made based on 2D [^{1}H, ^{1}H] NOESY, COSY and TOCSY NMR data. Complex formation with d(CGACTAGTCG)$_2$ was monitored by 1D ^{1}H NMR spectroscopy with data acquired using both excitation sculpting and presaturation for solvent suppression. Further details of the complete study of the DNA complex between AIK-18/51 and d(CGACTAGTCG)$_2$ is the subject of a separate article (in preparation).

Diffusion measurements were carried out using either a bipolar gradient routine (Bruker pulse program ledbpgppr2s) or a double stimulated echo routine (Bruker pulse program dstegppr3s) for which different numbers of transients, ns (typically ns = 32, 64, 128 or 256), were used depending on the concentration of the sample being studied in order to optimize the signal-to-noise ratio in each data set. The diffusion results were processed within the T1/T2 analysis module of Topspin and fitted according to Eq. 3

$$I = I_0 e^{-D(2\pi\gamma G\delta)^2(\Delta - x)\times 10^4} \qquad \text{Eq. 3}$$

where, γ is the gyromagnetic ratio (Hz/G, γ = 4257.7 Hz/G for proton), δ is the duration of the gradient (s), Δ is the time between spatially encoding gradients (s), I is the signal intensity, G is the amplitude of the pulsed field gradient (G/cm) and x is a pulse sequence dependent correction factor. Calibration of the spectrometer gradient strength for the purposes of diffusion measurements was carried out using a small amount of H_2O contained within a Shigemi NMR tube assembly. A literature value for the self-diffusion coefficient of H_2O at 298 K of 2.299×10^{-9} $m^2.s^{-1}$ measured by the diaphragm cell technique was used as the calibration value.

2.5 Molecular Modeling

Structure calculations were carried with a molecular mechanics procedure using X-PLOR (version 3.1) [44]. Atomic charges for the ligand were calculated using Gaussian03W [45] by the Mertz-Kollman method [46] with MP2/6-31G* basis set. Modeling of the aqueous environment was performed by water molecules in the form of TIP3P [47] placed in a box with sides of 54×30×30 Å (1663 molecules). The procedure of molecular dynamics (MD) simulations corresponded to that used previously in the energy analysis of drug self-association [48]. The total evolution time was equal to 1 ns. Coordinates for all atoms were saved every picosecond. The approach used in the energetic analysis of ligand dimerisation was broadly similar to the energy decomposition approach used previously for aromatic molecules [48,49] and the reader is directed to these articles for further information.

3. Results and Discussion

Nuclear Magnetic Resonance (NMR) spectroscopy and isothermal titration calorimetry (ITC) studies of DNA binding by the thiazotropsin class of molecules show no observable evidence of single molecule binding in a stepwise, cooperative manner. The data recorded by us to date for the thiazotropsins suggest the occurrence of DNA binding by an assembled homo-ligand dimer unit [32-3334,50]. As detailed in the following, our findings support the notion that ligand assembly occurs before complex formation for this class of molecule, a feature that may be exploited as a general principle when considering the design of strongly self-associating DNA minor groove binders.

3.1. NMR data assignments

NMR data were initially acquired for AIK-18/51 with a view to determining a full ^{1}H NMR signal assignment. Data acquired on a 1 mM sample of the ligand in 90% H_2O/ 10% D_2O solution (**Fig. 2a**) afforded observation of labile (N<u>H</u>) proton resonances, necessary for sequential resonance assignment purposes. With the aid of 2D [^{1}H, ^{1}H] NOESY, COSY and TOCSY NMR data sets, a complete proton signal assignment of the molecule was achieved (Table 1).

Table 1. ^{1}H NMR chemical shift assignments for AIK-18/51 in both 90% H_2O/10% D_2O at C = 1 mM and in 99% D_2O phosphate buffer (pH = 7.4) at C = 5 mM.

Position	**Chemical shift $\delta^{1}H$ (ppm)**		**Signal Designation**
	1 mM aqueous buffer-free solution	**5 mM aqueous buffered solution (pH=7.4)**	
1	7.939	8.016	**F**
2	7.341	6.930	**G**
3	8.469	7.582	**D**
5	8.676	8.330	**C**
7	9.856	**-**	**-**
9	6.754	6.224	**K**
11	6.880	6.514	**I**
12	3.691	3.400	**M**
14	8.949	**-**	**-**
16	6.790	6.271	**J**
18	6.960	6.611	**H**
19	3.660	3.365	**N**
21[a]	-	-	**-**
25	3.818	3.523	**L**
26	1.219	1.087	**S**
27	1.219	1.087	**S**
29	8.282	**-**	**-**

30	3.232	3.002	**O/P**
31	1.955	1.836	**R**
32	3.151	2.962	**O/P**
34	2.900	2.821	**Q**
35	2.900	2.821	**Q**

[a] Chemical exchange results in the broadening of H21 beyond the limit of detection.

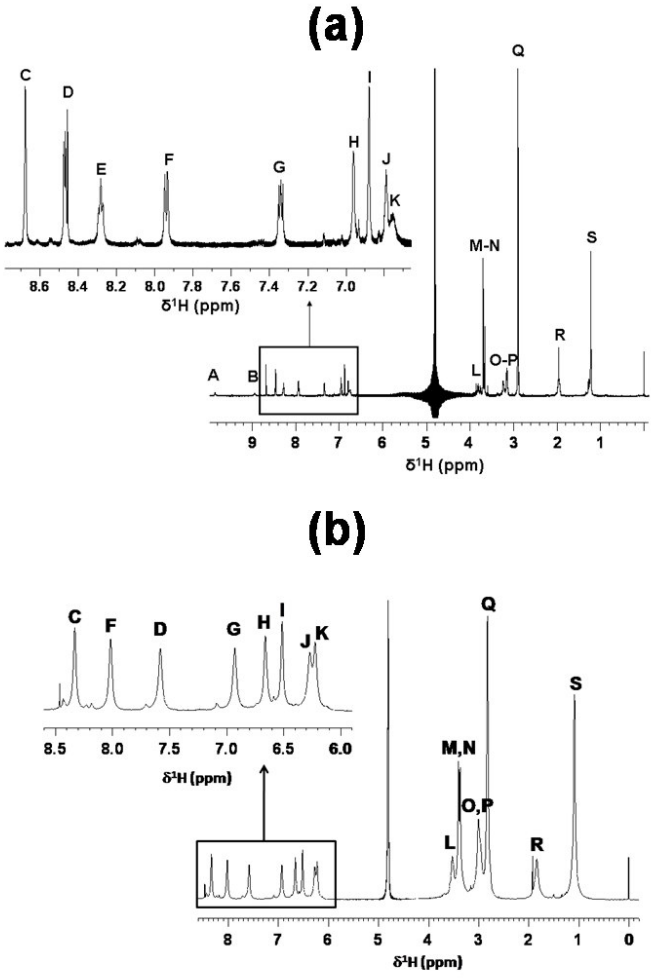


Fig. 2. 600 MHz 1D ^{1}H NMR spectra of AIK-18/51 in aqueous solution. **a)** Solution in 90% H_2O/10% D_2O at [C] = 1 mM. **b)** Solution in 99% D_2O phosphate buffer at [C] = 5 mM. Insets show expansions of the 'aromatic' region of the NMR spectra. See Table 1 for labelled signal assignments.

Subsequently, NMR data were also acquired and assigned (Table 1) for a 5 mM solution of the same ligand in 99% D_2O phosphate buffer solution at pH = 7.4. The signal assignments made for the 1 mM solution were used as a basis for establishing the chemical shift assignments for the 5 mM buffered solution in combination with the correlation data (**Fig. 2b**). Differences were observed for the ^{1}H NMR chemical shift values assigned for the signals under each set of solution conditions, associated with a concentration and/or pH effect, which suggested the presence in solution of self-assembled aggregates. This was corroborated by the following supporting evidence.

3.2. ^{1}H NMR chemical shift and NOE analysis

^{1}H NMR chemical shifts were measured as a function of solute concentration at 298 K (**Fig. 3a**) and also as a function of temperature at a solute concentration of 2 mM (**Fig. 3b**). The results were analyzed numerically to determine enthalpy of association (ΔH_{ass}) and self-association constant (K_{ass}) (Table 3).

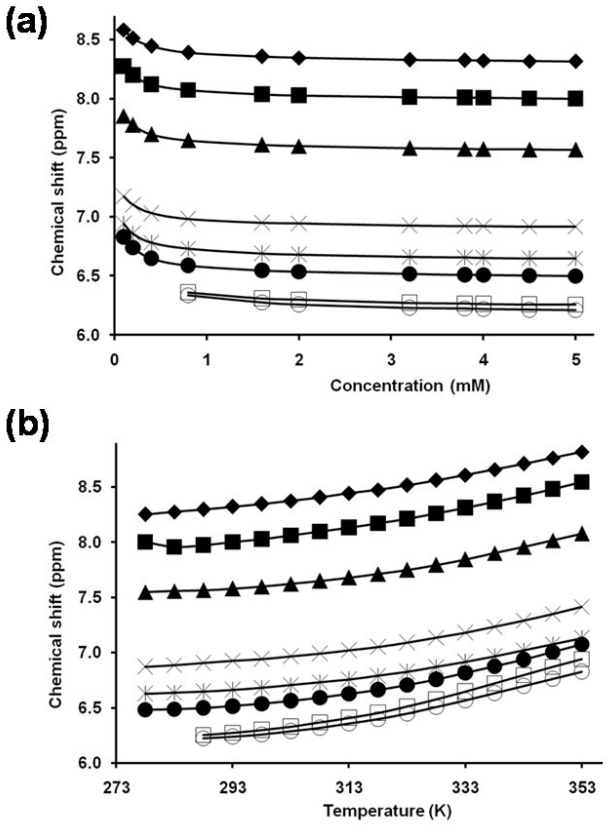


Fig. 3. Evolution of ^{1}H NMR chemical shifts **a)** as a function of concentration at T = 298 K and **b)** as a function of temperature for [C] = 2 mM for aromatic protons of AIK-18/51 in D_2O phosphate buffer. Key: ◆ - H**C**; ■ - H**F**;▲ - H**D**; × - H**G**; ∗ - H**H**; ● - H**I**; □ - H**J**; ○ - H**K**. Refer to Table 1 for signal assignments.

Increasing values of chemical shift measured as a function of reduced solute concentration and temperature increase supported the occurrence of aggregation. Deshielding arises as the equilibrium shifts away from aggregation towards disassembled ligands and is a classical response observed for 'face-to-face' H- or J-stacked aggregates bound by both electrostatic dipole-dipole and electronic π-π stabilizing interactions between ligands [39].

Supporting this assertion were the NOE data (**Fig. 4**) in which all NOEs were negative, a feature normally associated with large molecules displaying $\omega\tau_c > 1$ (where ω is the Larmor precession frequency (Hz) and τ_c is the molecular correlation time (rad.s^{-1})) but also characteristic of large aggregates of small molecules.

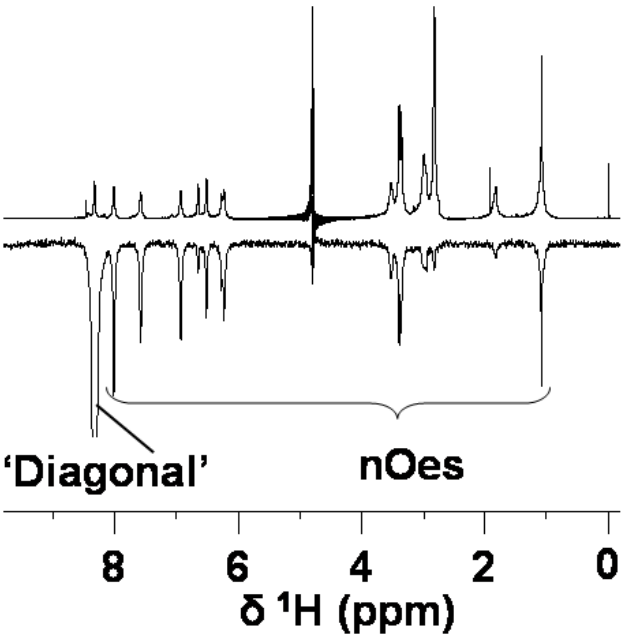


Fig. 4. NOE evidence of self-association for AIK-18/51. 600 MHz 1D ^{1}H NMR spectrum (TOP) and ω_2 slice at δ^1H = 8.33 ppm through the 240 ms mixing time 600 MHz 2D [^{1}H, ^{1}H] NOESY NMR spectrum (BOTTOM) for a solute concentration of 5 mM in 100 mM D_2O phosphate buffer, pH = 7.4, T = 298 K in which the negative NOE responses, anticipated more usually for macromolecules, result from small molecule aggregation.

Assignment of the NOE data revealed the presence of cross-peaks which could only be explained by the occurrence of intermolecular association as summarized (Table 2) and shown for selected NOE cross-peaks (**Fig. 5**). The only model consistent with the assigned NOEs and the chemical shift data was one in which molecules of AIK-18/51 were arranged anti-parallel to one another in a face-to-face fashion.

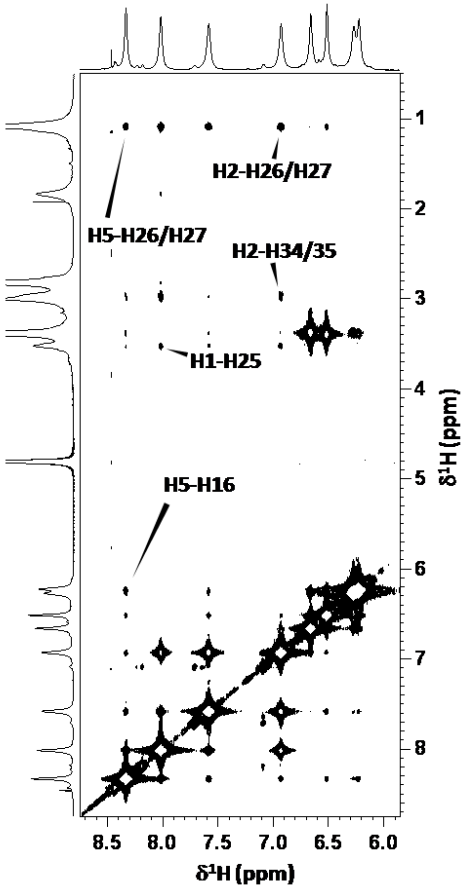


Fig. 5. Portion of the 600 MHz 2D [^{1}H, ^{1}H] NOESY NMR spectrum of AIK-18/51 at [C] = 5 mM, T = 298 K in D_2O phosphate buffer (pH = 7.4) acquired at a mixing time, τ_m = 40 ms and showing the NOE cross-peak region between aromatic proton resonances and between aromatic and aliphatic proton resonances. Selected NOE assignments are highlighted for those cross-peaks that can only be explained by reference to a head-to-tail arrangement of associated molecules.

Table 2. Observed NOEs between protons in AIK-18/51 (see structure for numbering scheme) together with distances calculated from an isolated spin pair approximation (ISPA) approach and based on NOE build-up curves.

Correlation	**Calculated Distance (Å)**
H5-H26/H27	4.2 ± 0.2
H2-H26/H27	4.0 ± 0.2
H11-H26/H27	4.7 ± 0.3
H5-H31	5.9 ± 0.3
H1-H31	5.0 ± 0.2
H1-H34/H35	4.9 ± 0.2
H2-H34/H35	5.6 ± 0.3
H5-H25	5.5 ± 0.3
H1-H25	4.6 ± 0.2
H3-H25	4.8 ± 0.2

H2-H25	4.5 ± 0.2
H5-H16	5.6 ± 0.3

3.3. NMR self-diffusion data analysis

Self-diffusion values, measured by NMR spectroscopy as a function of solute concentration, were also supportive of association into dimers followed by higher order association. These data were assessed against theoretical models to further refine the description of the self-assembly phenomenon. Ten concentrations of ligand in the accessible concentration range (for NMR study) of 5 - 0.2 mM inclusive, were used for which diffusion coefficients were measured from 7.22×10^{-11} m^2s^{-1} (5 mM) to 15.8×10^{-11} m^2s^{-1} (0.2 mM, **Fig. 6**). Evaluated against different diffusion models [51] and following adjustment of the molecular radius variable, with K_{agg} or K_{dim} restricted to values measured *via* ITC, a good fit was obtained to the stepwise (isodesmic) self-association model.

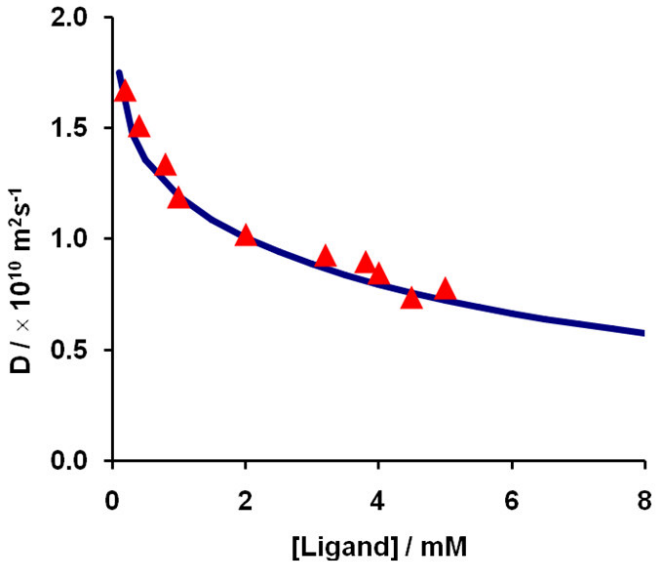


Fig. 6. Self-diffusion coefficient measured as a function of concentration of solute. Curve: stepwise self-assembly model; triangles: experimental self-diffusion data points for AIK-18/51 as a function of concentration. All experimental data were determined for AIK-18/51 in D_2O phosphate buffer by NMR and calibrated against the self-diffusion coefficient of H_2O at 298 K.

3.4. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) data measured as a function of temperature were used to independently assess the same aggregation phenomenon. Ligand aggregation in buffered aqueous solution is immediately evident from the occurrence of an endothermic response upon sample dilution (25 injections of 10 μL aliquots containing 5 mM AIK-18/51 in PIPES buffer (10 mM) into a cell of volume 1.4399 cm^3 filled with the same PIPES buffer, **Fig. 7**).

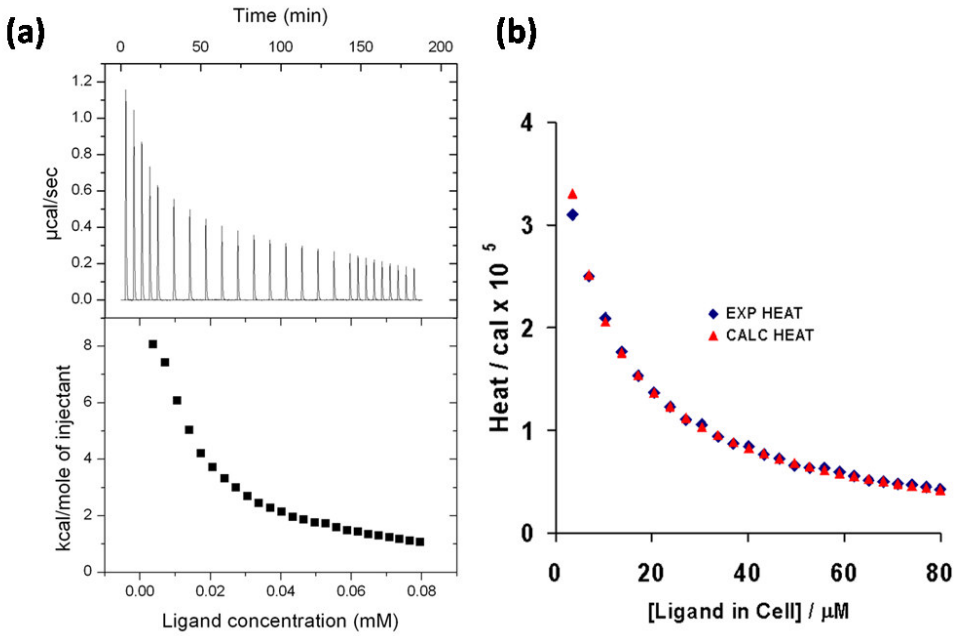


Fig. 7. Evidence of self-association for AIK-18/51 by ITC. **a)** Raw data traces showing heat effects per sample injection with time [TOP] and scaled per mole of injectant as a function of ligand concentration. **b)** Experimental and fitted ITC data showing characteristic heat response on sample dilution for AIK-18/51 as an aggregated molecular system.

Insight into the origins of the aggregation behaviour was gained by reference to numerical analysis procedures, modeling methods and the complementary NMR and ITC experimental approaches adopted. Numerical analysis, using both ITC [51,52] and variable temperature and concentration dependent proton chemical shift NMR data [53] were used to evaluate thermodynamic and equilibrium parameters associated with the ligand assembly (Table 3).

Table 3. Thermodynamic and Equilibrium parameters evaluated from ITC and NMR data for AIK-18/51 in aqueous buffered solutions using a stepwise self-association model.

Method	Temperature / °C	ΔH_{ass} / kcal.mol^{-1} [a]	K_{ass} / $10^{-3} \times M^{-1}$
ITC	25	-9.6 [-10.2 – -9.2][b]	49.8 [38 – 66]
		(-10.9 [-12.8 – -9.8])	(72.3 [47 – 125])
	35	-11.1 [-11.9 – -10.7]	16.2 [12.2 – 21.7]
	45	-17.7 [-26.5 – -13.6]	2.5 [1.3 – 3.8]
NMR	25	-15±1	61±19

[a] Buffers used: ITC – PIPES Buffer = 10 mM PIPES, 20 mM NaCl, 1 mM EDTA, pH = 6.8; (ACES Buffer = 10 mM ACES, 20 mM NaCl, 1mM EDTA, pH = 6.9); NMR – 100 mM phosphate buffer, pH = 7.4. [b] Numbers within square brackets are the lowest and highest values within the error margins.

ITC data for AIK-18/51 in PIPES buffer were measured at 25, 35 and 45 °C and fitted using IC-ITC [43,52]. From the variable temperature ITC data, the change in heat capacity $\Delta C_{p\ agg}$ for isodesmic self-aggregation was found to be -0.70 kJ mol^{-1} K^{-1} (see Supporting Information **Fig. S4**). Values for K_{ass} using a stepwise self-association model were broadly similar for NMR and ITC measurements at 25 °C. The concentration and temperature dependent NMR proton chemical shift data were also assessed against an indefinite non-cooperative (isodesmic) model of association (Eq. 4)

$$\delta = \delta_m + (\delta_d - \delta_m) \cdot \frac{2Kx_0 + 1 - \sqrt{4Kx_0 + 1}}{Kx_0} \qquad \text{Eq. 4}$$

where δ represents the experimentally measured ^{1}H NMR chemical shift, δ_m is the calculated limiting monomer ^{1}H chemical shift (i.e. at infinite dilution), δ_d is the calculated dimer ^{1}H chemical shift, K is the self-association constant and x_0 is the solute concentration. The self-association constant, K_{ass}, and

self-association enthalpy, ΔH_{ass}, determined for AIK-18/51 based on these NMR data were consistent with values for the self-association of compounds of similar molecular weight in aqueous solution [48].

3.5. Energetic Analysis

The factors responsible for the ordered nature of the self-assembled aggregate of AIK-18/51 were assessed by means of a detailed decomposition of the self-association energy using a molecular dynamics approach [48], results from which are detailed in Table 4. This process revealed that the main source of stabilization energy connected to dimer formation arises from intermolecular van der Waals forces (ΔG_{im} = -39.6 kcal.mol^{-1}), hydrophobic interactions (ΔG_{hyd} = -29.9 kcal.mol^{-1}), the creation of new vibrational degrees of freedom (ΔG_{vib} = -9.6 kcal.mol^{-1}) and electrostatic solvation (ΔG_{solv} = -6.2 kcal.mol^{-1}). Contributions that destabilize the dimer arise from van der Waals solvation (ΔG_{solv} = 34.1 kcal.mol^{-1}), loss of translational and rotational degrees of freedom (ΔG_{trot} = 20.5 kcal.mol^{-1}), loss of hydrogen bonds to water (ΔG_{HB} = 11.9 kcal.mol^{-1}) and electrostatic intermolecular interactions (ΔG_{im} = 10.1 kcal.mol^{-1}).

Table 4. Decomposition of the Self-Association Energy for AIK-18/51.

Interaction	**Energy / kcal.mol^{-1}**		
	ΔG_{solv} [a]	**ΔG_{im}** [b]	**Totals**
van der Waals (*vdW*)	34.1	-39.6	-5.5
Electrostatic (el)	-6.2	10.1	3.9
	ΔG_{trot} [c]	**ΔG_{vib}** [d]	

Entropic	20.5	-9.6	10.9
Hydrophobic (hyd)			-29.9
Hydrogen bonds			11.9
ΔG_{total}			**-8.7**
ΔG_{exp}			**-6.0**
$\lvert\Delta G_{exp} - \Delta G_{total}\rvert$			**2.7**

[a]Change in energy of interaction with the solvent (solvation energy). [b]Change in energy of intermolecular interaction of the MGB molecules in the dimer. [c]Change in energy due to loss of translational and rotational degrees of freedom. [d]Change in energy due to the formation of new vibrational degrees of freedom.

ΔG_{total} is within 2.7 kcal.mol^{-1} of ΔG_{exp} indicating that these results are reliable, considering the typical error associated with molecular dynamics simulations (typically several kcal.mol^{-1}) and the fact that ΔG_{total} is formed as the sum of large positive and negative numbers. The correlation of favourable and unfavourable interactions with individual structural features is therefore required, which may be used to build up complementary molecules by design. Selectivity with respect to DNA binding sequence has been well established in these MGBs and so the critical next stage is to create new MGB hetero-dimers taking account of the binding and destabilizing interactions described above. Empirical evidence to support this approach in terms of the avoidance of unfavourable interactions is to hand from our previous studies in which we showed that MGBs constructed from *N*-alkylpyrroles with bulky alkyl groups such as *iso*-propyl or 3-methylbutyl bound weakly or not detectably to DNA, Moreover such compounds were not active in antibacterial assays. Similar results were found with 1,5-dimethylpyrrole substructures and in compounds with *tert*-butyl head groups. All of these compounds are sterically

hindered from forming dimers that bind strongly to DNA which serves to emphasize the importance of the mutual fit of the components of a dimer [54].

3.6. Size and structure of aggregate components

Assessing the self-diffusion data in the context of the size and shape of monomer and dimer ligands allowed further conclusions to be drawn. To a first approximation the structures may be likened to cylinders. Molecular modeling makes it possible to evaluate their approximate length and diameter. Thus, the monomer has length, L = 23 Å and diameter, d = 5 Å; the dimer on the other hand has L = 23 Å and d = 10 Å. Using the standard expression for translational self-diffusion coefficient, $D_t = \frac{kT}{3\pi\eta L}\ln\frac{L}{d}$, the self-diffusion coefficients of monomer (mon) and dimer (dim) may be estimated: D_{mon}= 2.65 × 10^{-10} $m^2.s^{-1}$ and D_{dim}=1.45 × 10^{-10} $m^2.s^{-1}$. **Fig. 6** shows that the experimental D_t value asymptotically approaches a value of 2 × 10^{-10} $m^2.s^{-1}$ or higher as C→0. This qualitatively confirms the estimation of the value D_{mon}. If the real magnitude of D_{mon} is taken to be D_{mon}= 2 × 10^{-10} $m^2.s^{-1}$, it follows that D_{dim} should be close to D_{dim} = (1.45/2.65) × (2 × 10^{-10}) $m^2.s^{-1}$ = 1.09 × 10^{-10} $m^2.s^{-1}$. However, the experimental curve (**Fig. 6**) dips below the threshold value of 1 × 10^{-10} $m^2.s^{-1}$, thereby indicating formation of aggregates of higher order than dimers. The choice of the stepwise model for the NMR study is therefore vindicated.

Key to our results was the finding that aggregate assembly occurs for the ligand in free aqueous solution with an arrangement of monomers that mirrored the arrangement found for these molecules when bound to the DNA minor groove. Detailed assignment and evaluation of a series of 2D [^{1}H, ^{1}H] NOESY NMR data sets (mixing times in the range 30-240 ms) enabled the dimer geometry to be investigated in more detail. Intermolecular NOEs (nuclear Overhauser effect responses) were observed for molecules arranged anti-parallel to one another (i.e. head-to-tail, **Fig. 8**). The chemical shift and variable temperature NMR data also attested a face-to-face arrangement.

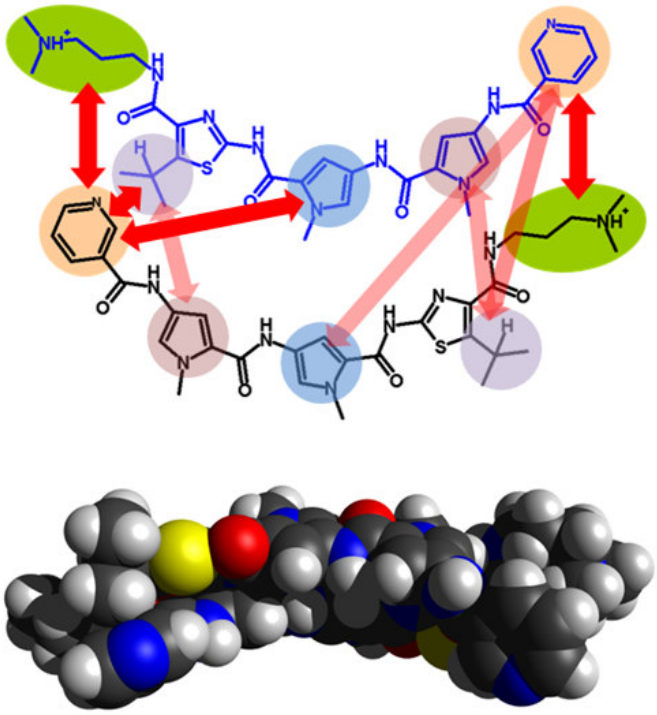


Fig. 8. Structure of the anti-parallel dimer assembly for AIK-18/51 based on 2D [1H, 1H] NOESY NMR data with schematic showing unambiguously assigned NOEs between substructures (double headed arrows, TOP) and a space filling representation (BOTTOM) from modelling of the dimer assembly.

This mode of assembly is not restricted to this molecule alone according to our ongoing investigations with analogues of AIK-18/51. We speculate that this mode of assembly is a generalized aqueous solution phase form of this class of molecule. Restrained molecular models and data from molecular dynamics simulations enabled realistic representations to be made of the anticipated homo-dimer assemblies (**Fig. 8**). Further aggregation to form larger assemblies might not necessarily be possible with the same strength of intermolecular interaction and could in fact be weaker but our data are unable to confirm this.

The data support the assertion that head-to-tail, face-to-face dimerization occurs and may also be the preferred mode of monomer association in the extended ligand aggregates. Such anti-parallel, face-to-face pairing is suggested and mediated substantially by hydrophobic and van der Waals interactions.

The head-to-tail dimer character provides maximum separation between electrostatic charges associated with the 3-dimethylaminopropyl 'tails' of each monomer. Hydrophobic interaction between the thiazole moiety of one molecule, especially its isopropyl group, and the pyridyl ring of its partner molecule (evidenced by appropriate NOEs, Table 2) drives the stabilization of head-to-tail dimer formation. Face-to-face dipole-dipole interactions between peptide linkages and heterocyclic rings may add to such stabilization [55]. Further detailed studies on related systems are required to validate this model, especially for the purposes of tailoring specific geometries of assembly for both like and unlike molecules. This will require additional insight into the local sources of relevant intermolecular forces.

3.7. Comparison with a ligand/DNA complex

These findings are all the more compelling when the minor groove DNA complexes of these molecules are considered in structural detail. Many studies of complex formation between MGBs and DNA duplexes, including our own with the thiazotropsins and related analogues [10], have shown DNA recognition to occur in a 2:1 ligand:duplex binding mode. What may have gone largely unnoticed, however, is that dimers are assembled anti-parallel to one another in the same manner as that described for the free ligand aggregates. AIK-18/51 is no exception in this respect. Titration of a solution of the DNA duplex d(CGACTAGTCG)$_2$, containing the recognition sequence ACTAGT, with a solution of AIK-18/51 and monitored by 1D ^{1}H NMR spectroscopy revealed formation of a stable complex with C_{2v} symmetry in slow exchange on the NMR timescale (**Fig. 9**).

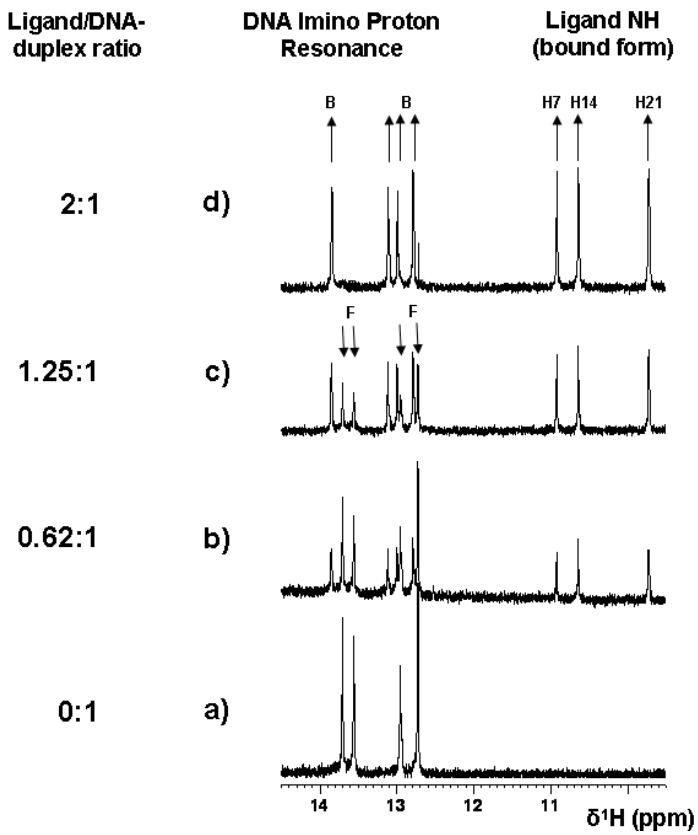


Fig. 9. DNA imino- and ligand peptide-NH resonance region of the 600 MHz 1D ^{1}H NMR spectrum of d(CGACTAGTCG)$_2$ in 90% H_2O/10% D_2O phosphate buffer (pH = 7.4) at a sample temperature of 298 K in the absence (a) and presence (b-d) of increasing quantities of AIK-18/51. Loss of free DNA resonances (F) are indicated by downward arrows; growth of ligand-bound DNA resonances (B) are indicated by upward arrows. New resonances arising from ligand peptide NH resonances are indicated by upward arrows and labeled H*n*.

The symmetry of the complex reported by these data is evidence enough to affirm 2:1 complex formation in which ligand molecules are disposed in an anti-parallel, face-to-face manner with respect to one another. Based on detailed analyses of the NMR data acquired on this complex (article in preparation) our structure calculations show that AIK-18/51 nestles comfortably within the DNA minor groove against the self-complementary ACTAGT binding sequence in the context d(CGACTAGTCG)$_2$ (**Fig. 10**).

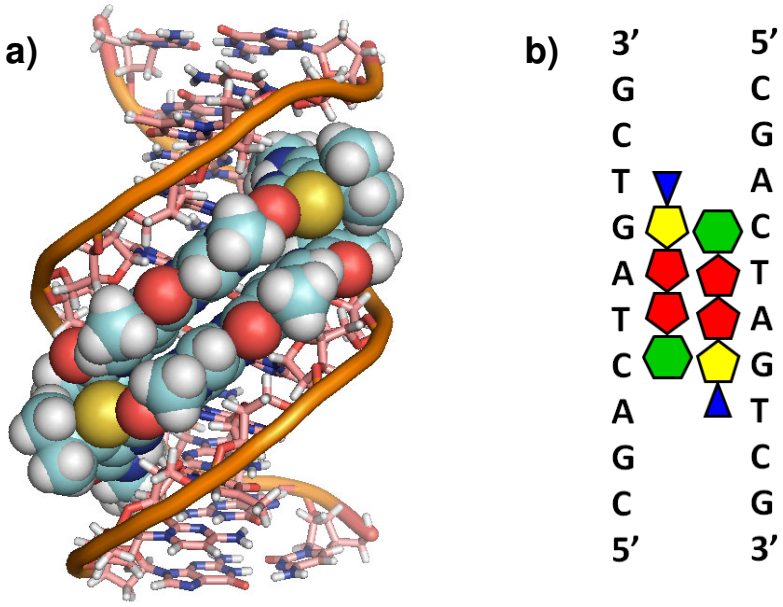


Fig. 10. Complex formation between the anti-parallel side-by-side binding of a homo-dimer assembly of AIK-18/51 bound in the minor groove of DNA against the sequence ACTAGT in the context d(CGACTAGTCG)$_2$ calculated from experimental NMR data. **a)** Calculated structure of the complex showing the homo-dimer assembly of AIK-18/51 (space filling, sulfurs are coloured gold; peptide linkage carbonyl oxygens are coloured salmon) bound to DNA within a widened minor groove setting (capped sticks together with a trace tube indicating the path of the DNA phosphodiester backbone). **b)** Schematic showing the relative position of bound monomers of AIK-18/51 with respect to one another and to the DNA sequence: triangle: dimethylaminopropyl (Dp) tail; light pentagon: isopropylthiazole unit (^{iPr}Th); dark pentagon: *N*-methylpyrrole unit (Py); hexagon: pyridyl head group (Pyr).

Study by NMR spectroscopy of both the free form of AIK-18/51 assembled in buffered aqueous solution *and* of the 2:1 ligand:DNA duplex complex of the same molecule bound within the DNA minor groove carries with it the advantage that a comparison can be drawn between the assembled structure of AIK-18/51 in both free and DNA bound forms. The similarities are especially notable. In both cases, the set of inter-ligand NOEs observed and assigned in the 2D [^{1}H, ^{1}H] NOESY NMR data are fully consistent with the same arrangement of monomers within each dimer assembly. This has important connotations when considering the energetics of binding. Global rearrangement of a dimer assembly prior to DNA binding would lead to unfavourable contributions to the free energy of binding. However, binding of a preformed structure requiring only subtle hand-in-glove changes to accommodate the ligand assembly within the DNA minor groove would be substantially favoured. The benefits of such pre-assembly are threefold. Firstly, the dynamics of the binding process are simplified compared with monomer binding processes in a one-at-a-time manner according to a more classical basis of cooperativity. Secondly, significant reduction in the entropic penalty incurred on DNA complex formation is implied. Thirdly, the design of smaller molecules as tools for DNA minor groove recognition will be facilitated if preassembly is well defined and catalogued. Pre-organized states suitable for directed, specific recognition of DNA in the form of composite ligands can therefore become a focus for molecular scale structural design. This concept goes beyond single molecule design towards composite assembly design for groups of well-defined molecular clusters. For success the rules of engagement need to be established, tested and embedded into classical chemical design and synthesis. Such information is not only useful in the current context but also extends to fields in which self-assembly plays a central role in materials work such as nanofabrication.

3.8. A role for designed assembly processes

The premise of our MGB design strategy has been based to date on the supposition that in a medicinal chemistry context, cell wall penetration is expected to be improved by engineering enhanced hydrophobicity into small MGBs [33]. This can be achieved without significantly increasing the size or

weight of the molecule. The challenge now lies in the *de novo* design of composite ligand assemblies as described. Engineering of complementary pairs of non-identical small MGBs (such as for hetero-dimer assemblies) based on precedent, would allow non self-complementary duplex DNA to be accessed through generalized programming in a manner akin to that described for hairpin and cyclic MGBs but instead using much smaller recognition templates. The hetero-assembly character of complement partners would have to take energetic preference over self-assembly homo-dimer formation for this to be successful, which, for monomers X and Y constituting such a composite assembly, would need to satisfy the equilibrium condition such that $K_{XY} >> K_{XX}$ or K_{YY}. Precedent for this occurring already exists and further extension of this work is therefore entirely feasible [18]. This forms the basis of the challenge for further developments in this field. Despite former studies, research work tailored specifically to the purpose of exploring this challenge has not been attempted in a systematic manner. Successful outcomes from such an approach would lead to new tools for DNA sequence analysis and detection and to the potential development of new biologically active composite drugs. Further developments along these lines will rely upon detailed knowledge of how electrostatic, van der Waals, topological and hydrophobic factors can be used to deliver hetero-assembled systems whilst simultaneously disfavouring homo-dimer formation. Decomposition of the relevant contributions to the energetics of this process, as shown by this article, can be combined with experimental characterization of the behaviour of such systems based on NMR measurements and ITC data. The prospects for generating a model suitable for describing the behaviour of extended forms of such assemblies in solution may be proposed, a work currently under consideration between our laboratories.

4. Conclusions

For the thiazotropsin class of molecules, represented in this work by AIK-18/51, monomer assembly occurs in free aqueous solution to form aggregates, data from which align with an assembly based on an anti-parallel, face-to-face geometry. NOE NMR data indicate that the geometry of the dimer in free solution very closely mirrors that of the same molecule when bound to the minor groove of

DNA in a 2:1 ligand:DNA complex with the sequence d(CGACTAGTCG)$_2$. Higher order aggregates of the ligand exist in solution as shown by ITC and diffusion NMR data, which fit to an indefinite non-cooperative self-assembly model. A molecular dynamics approach reveals that the main factors driving the assembly process are intermolecular van der Waals forces and hydrophobic interactions. Our data suggest that preassembly of dimer pairs of MGBs constituted as components of larger aggregates and represented here by AIK-18/51 could occur in advance of binding to a DNA recognition site. The finding suggests an entropic advantage for this type of dynamic assembly process. Exploiting this phenomenon in a directed manner may prove to be advantageous when considering generalized programmable DNA recognition schemes for non-covalently linked small molecule partners.

Acknowledgements

We thank the EPSRC for funding support (to N.J. B., grant EP/D001641/1).

References

[1] M.-K. Teng, N. Usman, C. A. Frederick and A. H.-J. Wang, The molecular structure od the complex of Hoechst-33258 and the DNA dodecamer d(CGCGAATTCGCG), Nucleic Acids Research 16 (1988) 2671-2690.

[2] J. A. Parkinson, J. Barber, K. T. Douglas, J. Rosamund and D. Sharples, Nuclear Magnetic Resonance probing of binding interactions used by minor groove binding, DNA-directed ligands – assignment of the binding site of Hoechst-33258 on the self-complementary oligonucleotide d(CGCGAATTCGCG), Journal of the Chemical Society Chemical Communications (1989) 1023-1025.

[3] J. A. Parkinson, J. Barber, K. T. Douglas, J. Rosamund and D. Sharples, Minor-groove recognition of the self-complementary duplex d(CGCGAATTCGCG)$_2$ by Hoechst-33258 – a high-field NMR study, Biochemistry 29 (1990) 10181-10190.

[4] J. A. Parkinson, J. A. Barber, B. A. Buckingham, K. T. Douglas and G. A. Morris, Hoechst-33258 and its complex with the oligonucleotide d(CGCGAATTCGCG)$_2$ – H-1 NMR assignments and dynamics, Magnetic Resonance in Chemistry 30 (1992) 1064-1069.

[5] M. S. Searle and K. J. Embrey, Sequence-specific interaction of Hoechst-33258 with the minor groove of an adenine-tract DNA duplex studied in solution by H-1-NMR spectroscopy, Nucleic Acids Research 18 (1990) 3753-3762.

[6] S. E. S. Ebrahimi, J. A. Parkinson, K. R. Fox, J. H. McKie, J. Barber and K. T. Douglas, Studies of the interaction of a meta-hydroxy analog of Hoechst-33258 with DNA by melting temperature, footprinting and high-resolution H-1-NMR spectroscopy, Journal of the Chemical Society Chemical Communications (1992) 1398-1399.

[7] J. A. Parkinson, S. E. S. Ebrahimi, J. H. McKie and K. T. Douglas, Molecular design of DNA-directed ligands with specific interactions – solution NMR studies of the interaction of a m-hydroxy analog of Hoechst 33258 with d(CGCGAATTCGCG)$_2$, Biochemistry 33 (1994) 8442-8452.

[8] D. G. Brown, M. R. Sanderson, J. V. Skelly, T. C. Jenkins, T. Brown, E. Garman, D. Stuart and S. Neidle, Crystal structure of a berenil dodecanucleotide complex – the role of water in sequence specific ligand binding, EMBO Journal 9 (1990) 1329-1334.

[9] M. Yoshida, D. L. Banville and R. H. Shafer, Structural analysis of d(GCAATTGC)$_2$ and its complex with berenil by nuclear magnetic resonance spectroscopy, Biochemistry 29 (1990) 6585-6592.

[10] M. Coll, J. Aymami, G. A. van der Marel, J. H. van Boom, A. Rich and A. H.-J. Wang, Molecular structure of the netropsin-d(CGCGATATCGCG) complex – DNA conformation in an alternating AT segment, Biochemistry 28 (1989) 310-320.

[11] J. G. Pelton and D. E. Wemmer, Structural characterization of a 2:1 distamycin A d(CGCAAATTGGC) complex by two-dimensional NMR, Proceedings of the National Academy of Sciences USA 86 (1989) 5723-5727.

[12] J. G. Pelton and D. E. Wemmer, Binding modes of distamycin-A with $d(CGCAAATTTGCG)_2$ determined by 2-dimensional NMR, Journal of the American Chemical Society 112 (1990) 1393-1399.

[13] P. B. Dervan, Molecular recognition of DNA by small molecules, Bioorganic and Medicinal Chemistry 9 (2001) 2215-2235.

[14] S. Neidle, DNA minor-groove recognition by small molecules, Natural Product Reports 18 (2001) 291-309.

[15] D. E. Wemmer, Ligands recognizing the minor groove of DNA: development and applications, Biopolymers 52 (1999) 197-211.

[16] B. H. Geierstanger, J. P. Jacobsen, M. Mrksich, P. B. Dervan and D. E. Wemmer, Structural and dynamic characterization of the heterodimeric and homodimeric complexes of distamycin and 1-methylimidazole-2-carboxamide-netropsin bound to the minor-groove of DNA, Biochemistry 33 (1994) 3055-3062.

[17] M. Munde, A. Kumar, R. Nhili, S. Depauw, M.-H. David-Cordonnier, M. A. Ismail, C. A. Stephens, A. A. Farahat, A. Batista-Parra, D. W. Boykin and W. D. Wilson, DNA minor groove induced dimerization of heterocyclic cations: compound structure, binding affinity and specificity for a TTAA site, Journal of Molecular Biology 402 (2010) 847-864.

[18] B. H. Geierstanger, T. J. Dwyer, Y. Bathini, J. W. Lown and D. E. Wemmer, NMR characterization of a heterocomplex formed by distamycin and its analog 2-ImD with

d(CGCAAGTTGGC):d(GCCAACTTGGC): Preference for the 1:1:1 2-ImD:Dst:DNA complex over the 2:1 2-ImD:DNA and the 2:1 Dst:DNA complexes, Journal of the American Chemical Society 115 (1993) 4474-4482.

[19] M. Mrksich and P. B. Dervan, Antiparallel side-by-side heterodimer for sequence-specific recognition in the minor groove of DNA by a distamycin 1-methylimidazole-2-carboxamide-netropsin pair, Journal of the American Chemical Society 115 (1993) 2572-2576.

[20] H. Sugiyama, C. Y. Lian, M. Isomura, I. Saito and A. H.-J. Wang, Distamycin A modulates the sequence specificity of DNA alkylation by duocarmycin A, Proceedings of the National Academy of Sciences USA 93 (1996) 14405-14410.

[21] M. Mrksich and P. B. Dervan, Design of a covalent peptide heterodimer for sequence-specific recognition in the minor-groove of double-helical DNA, Journal of the American Chemical Society 116 (1994) 3663-3664.

[22] D. S. Pilch, N. Poklar, C. A. Gelfand, S. M. Law, K. J. Breslauer, E. E. Baird and P. B. Dervan, Binding of a hairpin polyamide in the minor groove of DNA: sequence-specific enthalpic discrimination, Proceedings of the National Academy of Sciences USA 93 (1996) 8306-8311.

[23] R. P. L. de Clairac, B. H. Geierstanger, M. Mrksich, P. B. Dervan and D. E. Wemmer, NMR characterization of hairpin polyamide complexes with the minor groove of DNA, Journal of the American Chemical Society 119 (1997) 7909-7916.

[24] J. Cho, M. E. Parks and P. B. Dervan, Cyclic polyamides for recognition in the minor-groove of DNA, Proceedings of the National Academy of Sciences USA 92 (1995) 10389-10392.

[25] Q. Zhang, T. J. Dwyer, V. Tsui, D. A. Case, J. Cho, P. B. Dervan and D. E. Wemmer, NMR structure of a cyclic polyamide-DNA complex, Journal of the American Chemical Society 126 (2004) 7958-7966.

[26] D. M. Chenoweth and P. B. Dervan, Allosteric modulation of DNA by small molecules, Proceedings of the National Academy of Sciences USA 106 (2009) 13175-13179.

[27] N. G. Nickols and P. B. Dervan, Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide, Proceedings of the National Academy of Sciences USA 104 (2007) 10418-10423.

[28] C. F. Hsu, J. W. Phillips, J. W. Trauger, M. E. Farkas, J. M. Belitsky, A. Heckel, B. Z. Olenyuk, J. W. Puckett, C. C. C. Wang and P. B. Dervan, Completion of a programmable DNA-binding small molecule library, Tetrahedron 63 (2007) 6146-6151.

[29] N. G. Nickols, C. S. Jacobs, M. E. Farkas and P. B. Dervan, Improved nuclear localization of DNA-binding polyamides, Nucleic Acids Research 35 (2007) 363-370.

[30] J. A. Raskatov, J. L. Meier, J. W. Puckett, F. Yang, P. Ramakrishnan and P. B. Dervan, Modulation of NF-κB-dependent gene transcription using programmable DNA minor groove binders, Proceedings of the National Academy of Sciences USA 109 (2012) 1023-1028.

[31]. J. H. Miller, The Operon (Cold Spring Harbour Laboratory 1980)

[32] N. G. Anthony, D. Breen, J. Clarke, G. Donoghue, A. J. Drummond, E. Ellis, C. Gemmell, J.-J. Hélèxbeux, I. S. Hunter, A. I. Khalaf, S. P. Mackay, J. A. Parkinson, C. J. Suckling and R. D. Waigh, Antimicrobial lexitropsins containing amide, amidine and alkene linking groups, Journal of Medicinal Chemistry 50 (2007) 6116-6125.

[33] K. R. Fox, A. I. Khalaf, S. P. Mackay, I. S. McGroarty, J. A. Parkinson, G. G. Skellern, C. J. Suckling and R. D. Waigh, DNA binding of a short lexitropsin, Bioorganic and Medicinal Chemistry Letters 14 (2004) 1353-1356.

[34] N. G. Anthony, B. F. Johnston, A. I. Khalaf, S. P. Mackay, C. J. Suckling, R. D. Waigh and J. A. Parkinson, Short lexitropsin that recognizes the DNA minor groove at 5′-ACTAGT-3′: Understanding the role of isopropyl-thiazole, Journal of the American Chemical Society 126 (2004) 11338-11349.

[35] N. G. Anthony, D. Breen, G. Donoghue, A. I. Khalaf, S. P. Mackay, J. A. Parkinson and C. J. Suckling, A new synthesis of alkene-containing minor-groove binders and essential hydrogen bonding in binding to DNA and in antibacterial activity, Organic and Biomolecular Chemistry 7 (2009) 1843-1850.

[36] C. J. Suckling, Molecular recognition and physicochemical properties in the discovery of selective antibacterial minor groove binders. Journal of Physical Organic Chemistry 21 (2008) 575-583.

[37] J. L. Seifert, R. E. Connor, S. A. Kushon, M. Wang and B. A. Armitage, Spontaneous assembly of helical cyanine dye aggregates on DNA nanotemplates, Journal of the American Chemical Society 121 (1999) 2987-2995.

[38] B. Nordén and F. Tjerneld, Optical studies on complexes between DNA and pseudoisocyanine, Biophysical Chemistry 6 (1976) 31-45.

[39] B. A. Armitage, Cyanine Dye-DNA interactions: intercalation, groove binding and aggregation, Topics in Current Chemistry 253 (2005) 55-76.

[40] J. B. Chaires and M. J. Waring, Drug-Nucleic Acid Interactions (Academic Press, 2001).

[41] V. V. Kostjukov, A. A. Hernandez Santiago, F. R. Rodriguez, S. Rosas Castilla, J. A. Parkinson and M. P. Evstigneev, Energetics of ligand binding to the DNA minor groove, Physical Chemistry Chemical Physics 14 (2012) 5588-5600.

[42] A. I. Khalaf, C. J. Suckling and R. D. Waigh, British Patent Application, PCT/GB02/05916, 2002. US Patent 7700765

[43] IC-ITC is available from the authors *via* http://www.cf.ac.uk/chemy/staffinfo/poc/software/ic_itc

[44] A. T. Brunger, X–PLOR (Yale University Press, 1992).

[45] Gaussian, Inc., 2004. Gaussian 03. Wallingford; CT.

[46] B. H. Besler, K. M. Mertz and P. A. Kollman, Atomic charges derived from semi-empirical methods, Journal of Computational Chemistry 11 (1990) 431-439.

[47] W. Jorgensen, J. Chandrasekhar, J. D. Madura, R. Impey and M. Klein, Comparison of simple potential functions for simulating liquid water, Journal of Chemical Physics 79 (1983) 926-935.

[48] V. V. Kostjukov, N. M. Khomytova, A. A. Hernandez Santiago, A.-M. Cervantes Tavera, J. Salas Alvarado and M. P. Evstigneev, Parsing of the free energy of aromatic-aromatic stacking interactions in solution, Journal of Chemical Thermodynamics 43 (2011) 1424-1434.

[49] V. V. Kostjukov, N. M. Khomytova and M. P. Evstigneev, Partition of thermodynamic energies of drug-DNA complexation, Biopolymers 91 (2009) 773–790.

[50] H. Y. Alniss, N. G. Anthony, A. I. Khalaf, S. P. MacKay, C. J. Suckling, R. D. Waigh, N. J. Wheate and J. A. Parkinson, Rationalizing sequence selection by ligand assemblies in the DNA minor groove: the case for thiazotropsin A, Chemical Science 3 (2012) 711-722.

[51] N. J. Buurma and I. Haq, Calorimetric and spectroscopic studies of Hoechest 33258: self-association and binding to non-cognate DNA, Journal of Molecular Biology 381 (2008) 607-621.

[52] N. J. Buurma and I. Haq, Advances in the analysis of isothermal titration calorimetry data for ligand-DNA interactions, Methods 42 (2007) 162-172.

[53] M. P. Evstigneev, V. P. Evstigneev and D. B. Davies, H-1 NMR determination of the self-association of an acridine homodimer and its complexation with ethidium bromide in aqueous solution, Journal of Molecular Structure 784 (2006) 162-168.

[54] A. I. Khalaf, R. D. Waigh, A. J. Drummond, B. Pringle, I. McGroarty, G. G. Skellern, and C. J. Suckling, Distamycin analogues with enhanced lipophilicity, Journal of Medicinal Chemistry 47 (2004) 2133-2156.

[55] C. J. Collar, M. Lee and W. D. Wilson, Setting anchor in the minor groove: in silico investigation into formamido N-methylpyrrole and N-methylimidazole polyamides bound by cognate DNA sequences, Journal of Chemical Information and Modeling 50 (2010) 1611-1622.