Comparison of DNA Complex Formation Behaviour for Two Closely Related Lexitropsin Analogues

by **John A. Parkinson***[a]), **Abedawn I. Khalaf**[a]), **Nahoum G. Anthony**[b]), **Simon P. MacKay**[b]), **Colin J. Suckling**[a]), and **Roger D. Waigh**[b])

[a]) WestCHEM, Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow G1 1XL, UK (phone: +44-141-5482820; fax: +44-141-5484822; e-mail: john.parkinson@strath.ac.uk)

[b]) Strathclyde Institute for Pharmaceutical and Biomedical Sciences, University of Strathclyde, 27 Taylor Street, Glasgow G4 0NR, UK

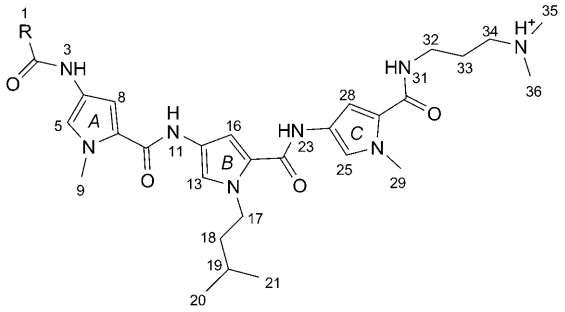
Two closely related lexitropsin analogues that differ only in the form of the 'headgroup' functionality (CHO (for **1**) *vs.* Ac (for **2**)) have been studied in their DNA-binding capacity for the sequence d(GCATATATGC) using ^{1}H-NMR spectroscopy. DNA-Complex formation for the CHO derivative was apparent from the observation of new NMR signals on titration of DNA with ligand. Detailed investigation and assignment of the data for a ligand/DNA-duplex ratio of 2:1 clearly delineated the structure as one associated with the 'minor groove' class of DNA complexes. The structure of the complex was determined on the basis of the acquired NMR data. Features characteristic of typical 2:1 minor-groove complexes were apparent. In a similar experimental approach, the Ac analogue ligand–DNA binding response was investigated. Despite the close similarity in chemical structure to the CHO case, the Ac analogue was found to produce NMR data of a much poorer quality. This was attributed to more rapid on/off chemical exchange equilibrium between ligand and DNA. From close analysis and comparison of the NMR data for the 'Ac' and 'CHO' headgroup ligand–DNA complexes, it was possible to ascertain that the same type of complex formed in each case but with different relative binding constants. Consideration of the nature and form of these complexes has been made with reference to a previously determined structure from our laboratory for the related lexitropsin analogue thiazotropsin A.

1. Introduction. – Solution-phase NMR studies of non-covalent minor-groove complex formation between DNA recognition sequences and minor-groove-binding small molecules (MGBs) have proved particularly fruitful for determining in detail the structural demand for such complexation [1]. Structure variations of typical MGBs based on short aromatic peptide molecules (lexitropsins) similar to the naturally occurring distamycin and netropsin antibiotics have become common and myriad in their diversity. Attempts to diversify the MGB theme with ligands possessing the capacity to act at the transcription level within cellular systems continue to meet with increasing success, as the sequence-reading capability of such molecules improves [2]. Nevertheless, despite the vast quantity of work directed towards this field, fundamental questions still remain regarding the electronic, thermodynamic, hydrophobic, H-bonding and structural demand required to enable the best possible sequence recognition and strongest binding capability for small MGBs. Understanding the role played by each of these factors is an essential prerequisite for developing a route to the design of much more effective MGBs both for pharmaceutical and analytical purposes.

Our teamwork has focussed on the development of small-molecule MGBs that, as far as possible, attempt to satisfy the 'rule of five' at least by reducing the overall number of H-bond donors and acceptors where a more general view would suggest that a literal take on the rule of five does not apply to anti-infective compounds [3]. By building enhanced hydrophobicity into each structure, our aim has been to develop a new class of MGBs with the potential for extended DNA sequence recognition. By dint of structural demand, MGBs that bind in the classical 2 : 1 side-by-side, head-to-tail self-assembled form within the DNA minor groove must, under symmetry considerations, recognise self-complementary DNA sequences. Evolving this concept to a fuller, more general rule for non self-complementary DNA sequence recognition within the DNA duplex structure requires hetero-dimerization of nonidentical ligands that could assemble to form the relevant self-asssembled structures capable of reading non-self-complementary DNA sequences. This contrasts with the approach used by *Dervan* and co-workers, which has focussed on the creation of much longer lexitropsin analogues incorporating the fold-back hairpin concept. Whilst such structures can be tailored for recognition of non-self-complementary DNA duplex sequence recognition, their relatively large size, which does not conform to the 'rule of five', may limit their suitability and utility as potential pharmaceuticals. Our intended approach to sequence reading, whilst potentially more demanding in design terms, has the scope to provide new advances in 'combination' therapy and could be helpful in the development of new DNA-sequence-reading tools in the context of DNA analysis and detection. To begin to more fully explore this avenue depends on an initial understanding of what the requirement for ligand self-assembly is on the formation of DNA minor-groove complexes.

To this end, we are carrying out solution-phase studies of ligand–DNA complex formation using the proprietary Strathclyde MGB compound library (StrathMGBs) in an attempt to begin to more fully understand the structural, energetic, and electronic demands for DNA minor-groove sequence recognition. By making small variations in the structures of our ligands, we hoped it would be possible to understand in more detail the subtle factors that influence the extent of ligand binding within the DNA minor groove. In this article, we report parallel solution-phase NMR studies of the DNA-complex formation between two closely related MGB analogues. A major structural feature of these compounds was the presence of an *N*-isopentyl (C_5) side chain attached to the central pyrrole of a three-pyrrole ring system. The hydrophobic nature of such a chain was anticipated to contribute significantly to the compound's ability to bind to DNA from two perspectives; firstly, from the formation of additional hydrophobic interactions with the sugar-phosphate DNA backbone, and, secondly, from potentially enhanced interaction of the ligands with one another when bound to DNA, assuming a favourable 2 : 1 ligand/DNA-duplex binding mode. In this article, these factors are considered together with the wider results of our observations, which are described in terms of data quality, complex formation and the influence of ligand structure on the nature of the DNA complex formed.

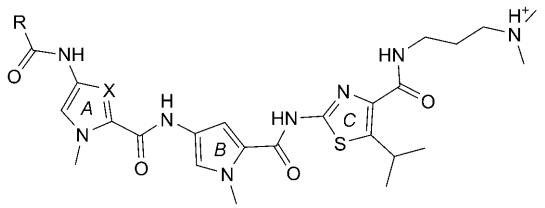
2. Results. – 2.1. *Titration of* **1** *and* **2** *with* **3**. The NMR responses observed when **1** was titrated against **3** are shown in *Figs. 1–3* for the Me/aliphatic, H1′/aromatic, and imino/amide NH resonance regions, respectively. The growth of new Me signals due to



1 R = H
2 R = Me

d(CGATATATCG)$_2$

3



Thiazotropsin A (**4**) R = H, X = CH
Thiazotropsin B (**5**) R = Me, X = N

the presence of ligand was observed at δ(H) values of 1.079 (Me20/21), 2.865 (Me35/36), 3.667 (Me29), and 3.899 (Me9) ppm (see chemical structure **1** for numbering scheme). Obvious new ^{1}H-NMR resonances ascribed to **1** were observed in the amide/amino N*H* resonance region, the main signals appearing at δ(H) values of 9.051 (H37) and 9.700 (H23) ppm. Comparison of the integrals of the signals ascribed to H37 and H23 with those for the DNA imino H-atom NMR responses confirmed that 2 equiv. of **1** were associated with the bound form of DNA. ^{1}H-NMR Responses from the DNA showed the complex to be in slow exchange on the NMR timescale based on the presence of signals from both free and bound forms of **3** at ligand concentrations of less than 2 full equiv. per DNA duplex. No evidence existed for the presence of a 1:1 ligand/DNA complex. The data were, therefore, consistent with a slow exchange complex in which the ligand was bound in the classical head-to-tail mode, thereby reflecting the retention of symmetry within the NMR data for the self-complementary DNA duplex. The data also identified a clear titration end-point at which all free DNA had been consumed in complex formation with the ligand.

By way of contrast, the ^{1}H-NMR responses for the titration of **2** with **3** are shown in *Figs. 4–6*. The results from this titration contrasted markedly with those of **1** titrated against **3**. On addition of 0.5 equiv. of **2** (titration point *b* of *Figs. 4–6*), characteristic new signals were apparent for some of the ligand Me signals (at δ(H) values of 1.057 (Me20/21) and 2.892 (Me35/36) ppm), but other Me responses were not immediately apparent. Additionally, the observed responses were very broad in comparison with

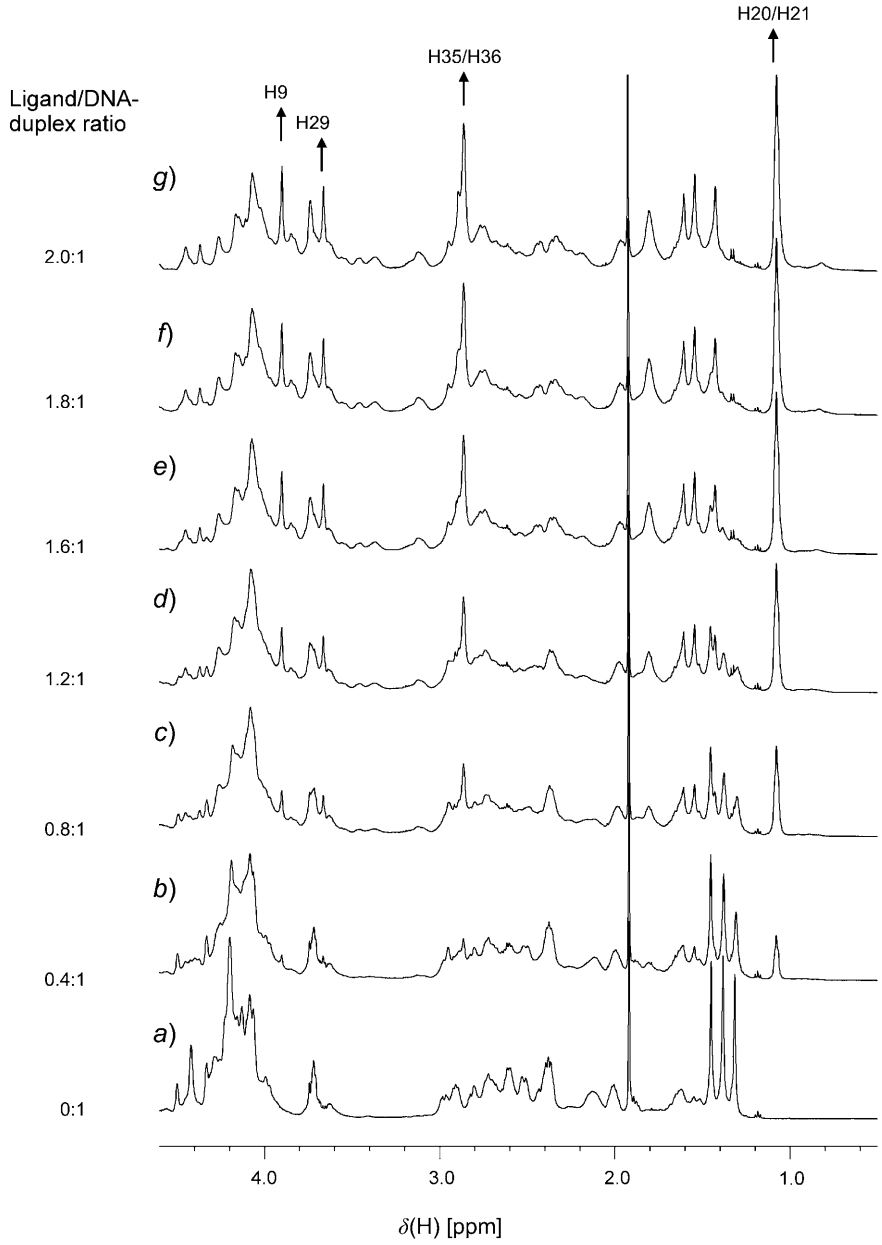


Fig. 1. *Aliphatic H-atom resonance region of the 600-MHz 1D ^{1}H-NMR spectrum of 2.52 mM d(CGATATATCG)$_2$* (**3**) *in aqueous solution at a sample temperature of 298 K and neutral pH in the absence* (*a*) *and presence* (*b*–*g*) *of increasing quantities of* **1** *in ligand/DNA-duplex ratios:* b) *0.4 : 1,* c) *0.8 : 1,* d) *1.2 : 1,* e) *1.6 : 1,* f) *1.8 : 1, and* g) *2.0 : 1.* Growth of ligand Me resonances are indicated by upward arrows labelled with the relevant atom assignment.

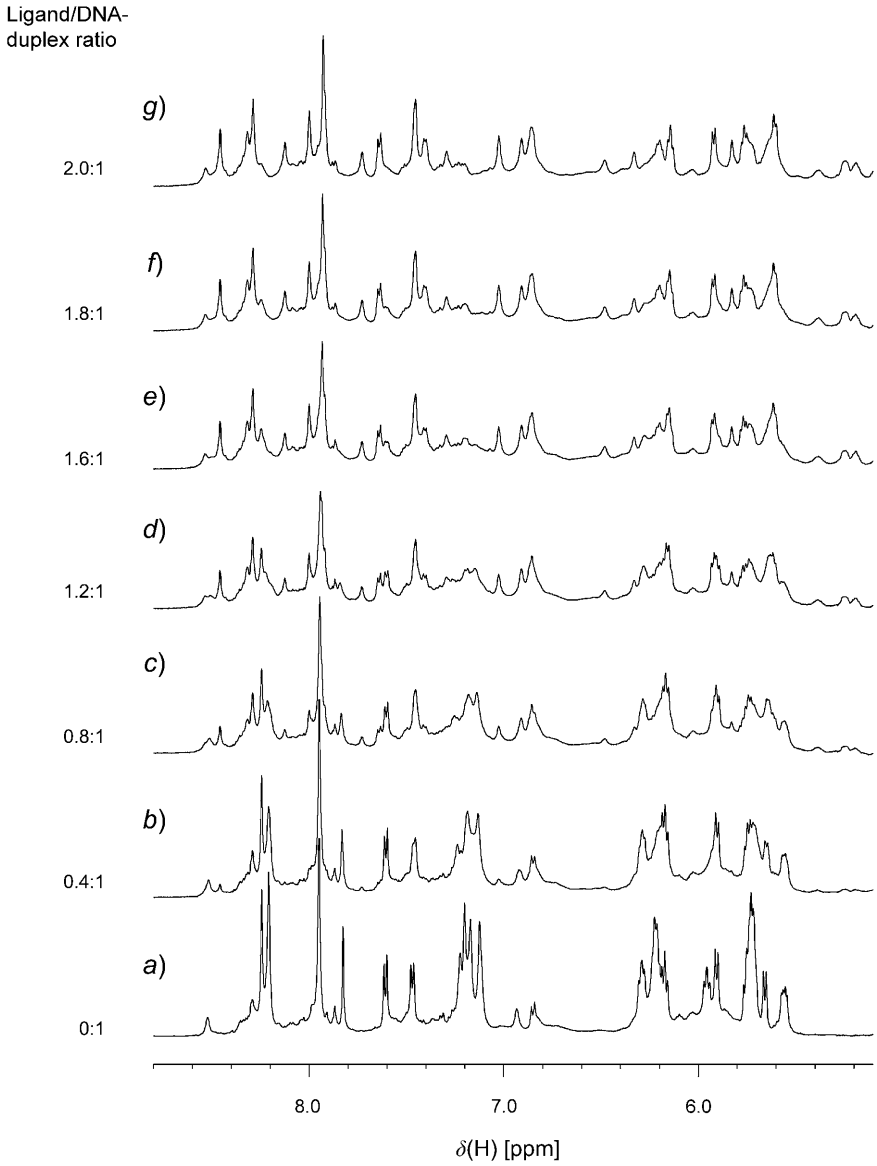


Fig. 2. *Aromatic and H1′ resonance region of the 600-MHz 1D ^{1}H-NMR spectrum of 2.52 mM d(CGATATATCG)$_2$* (**3**) *in aqueous solution at a sample temperature of 298 K and neutral pH in the absence* (*a*) *and presence* (*b*–*g*) *of increasing quantities of* **1** *in ligand/DNA-duplex ratios:* b) *0.4:1,* c) *0.8:1,* d) *1.2:1,* e) *1.6:1,* f) *1.8:1, and* g) *2.0:1*

those observed for the same groups in the complex of **1** with **3**. Whilst DNA signals from the free duplex reduced in intensity as expected (*e.g.*, the '*singlet*' thymine Me responses between δ(H) 1.2–1.6 ppm), the appearance of new signals for the complex

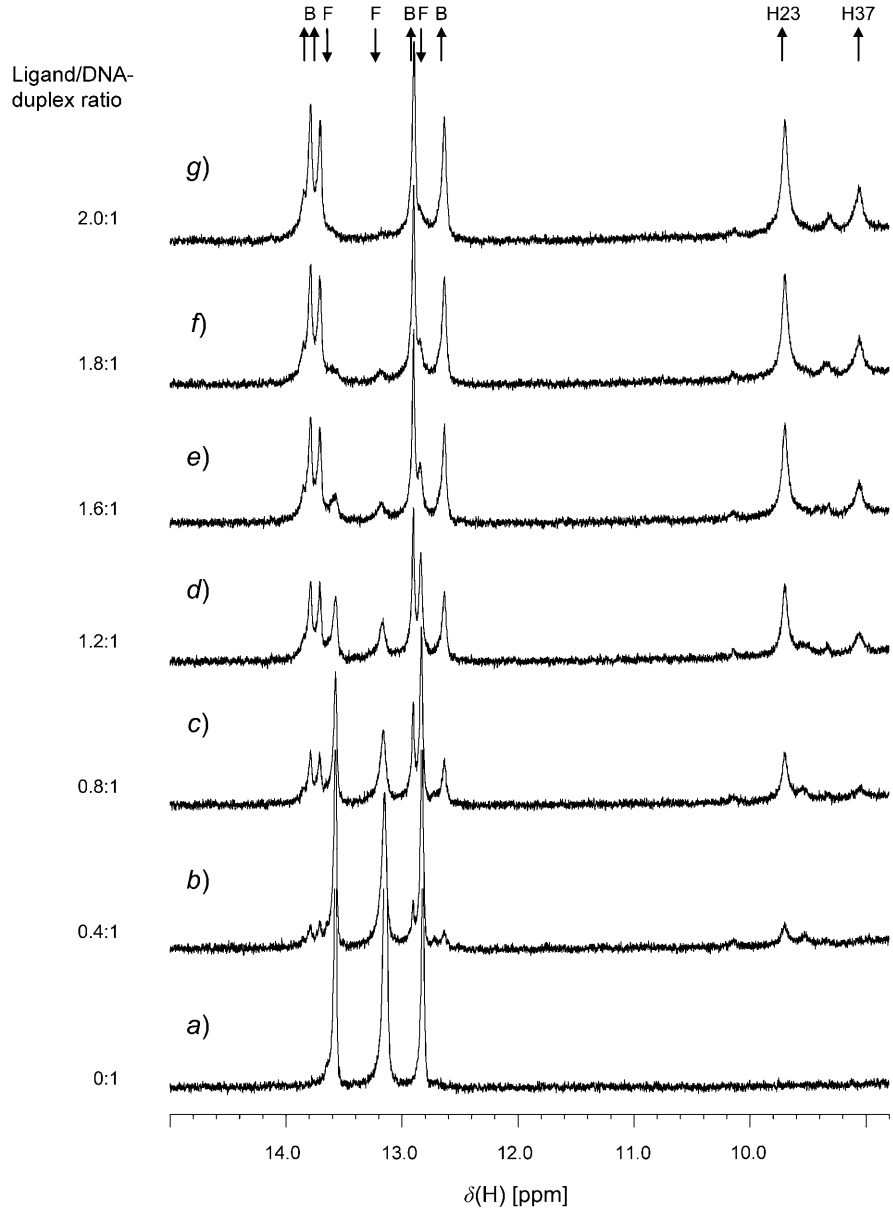


Fig. 3. *Imino and ligand peptide N*H *H-atom resonance region of the 600-MHz 1D* ^{1}*H-NMR spectrum of 2.52 m*M *d(CGATATATCG)*$_2$ (**3**) *in aqueous solution at a sample temperature of 298 K and neutral pH in the absence* (*a*) *and presence* (*b*–*g*) *of increasing quantities of* **1** *in ligand/DNA-duplex ratios:* b) *0.4 : 1,* c) *0.8 : 1,* d) *1.2 : 1,* e) *1.6 : 1,* f) *1.8 : 1, and* g) *2.0 : 1.* Loss of free DNA resonances (F) are indicated by downward arrows; growth of ligand-bound DNA resonance (B) are indicated by upward arrows. New resonances arising from ligand peptide N*H* resonances are indicated by upward arrows designated with the relevant atom assignment.

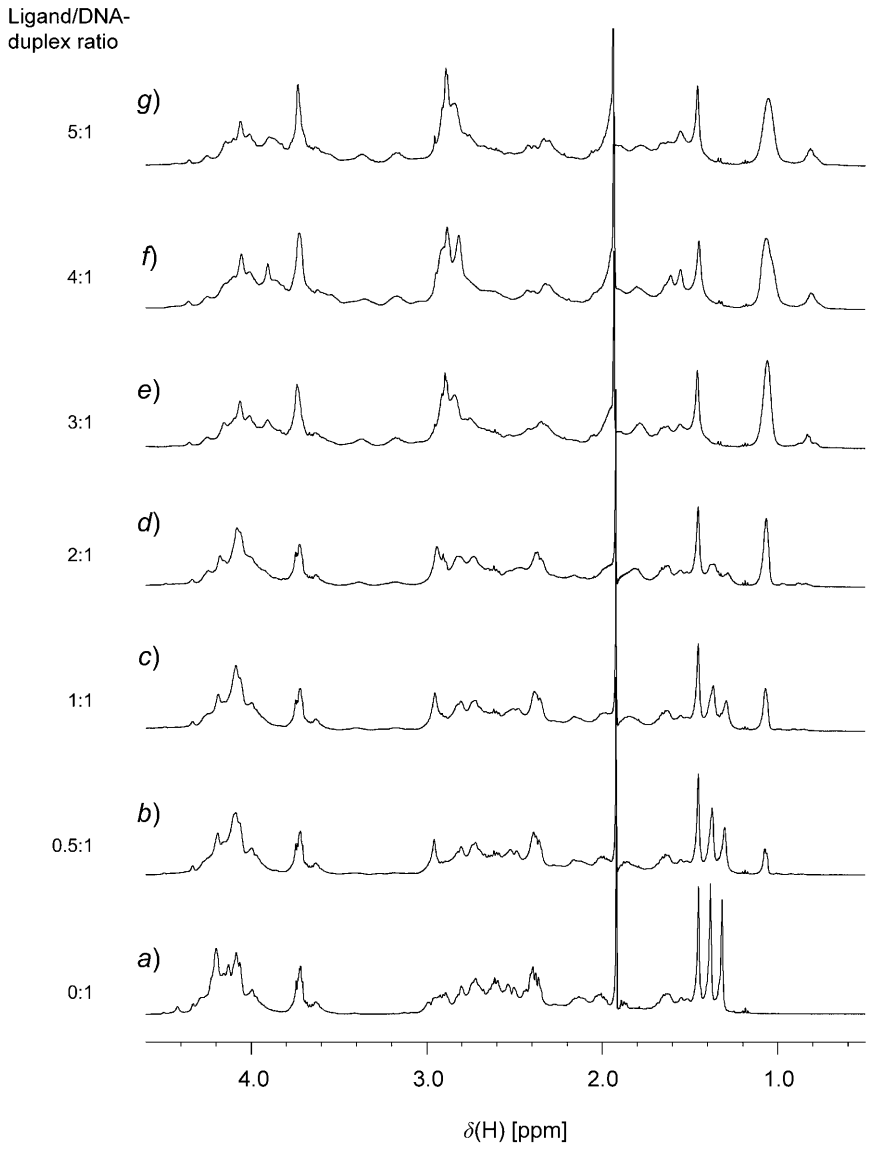


Fig. 4. *Aliphatic H-atom resonance region of the 600-MHz 1D ^{1}H-NMR spectrum of 2.52 mM d(CGATATATCG)$_2$* (**3**) *in aqueous solution at a sample temperature of 298 K and neutral pH in the absence* (*a*) *and presence* (*b*–*g*) *of increasing quantities of* **2** *in ligand/DNA-duplex ratios:* b) *0.5 : 1,* c) *1 : 1,* d) *2 : 1,* e) *3 : 1,* f) *4 : 1, and* g) *5 : 1*

were not observed, in contrast to the observations for **1** with **3**. This was especially apparent in the imino H-atom resonance region of the 1D ^{1}H-NMR spectrum (*Fig. 6*). Whilst some change in the appearance of the data was observed, the effect was not as dramatic as the former observations for **1** (compare *Figs. 3*,*b*, and *6*,*b*). No unique

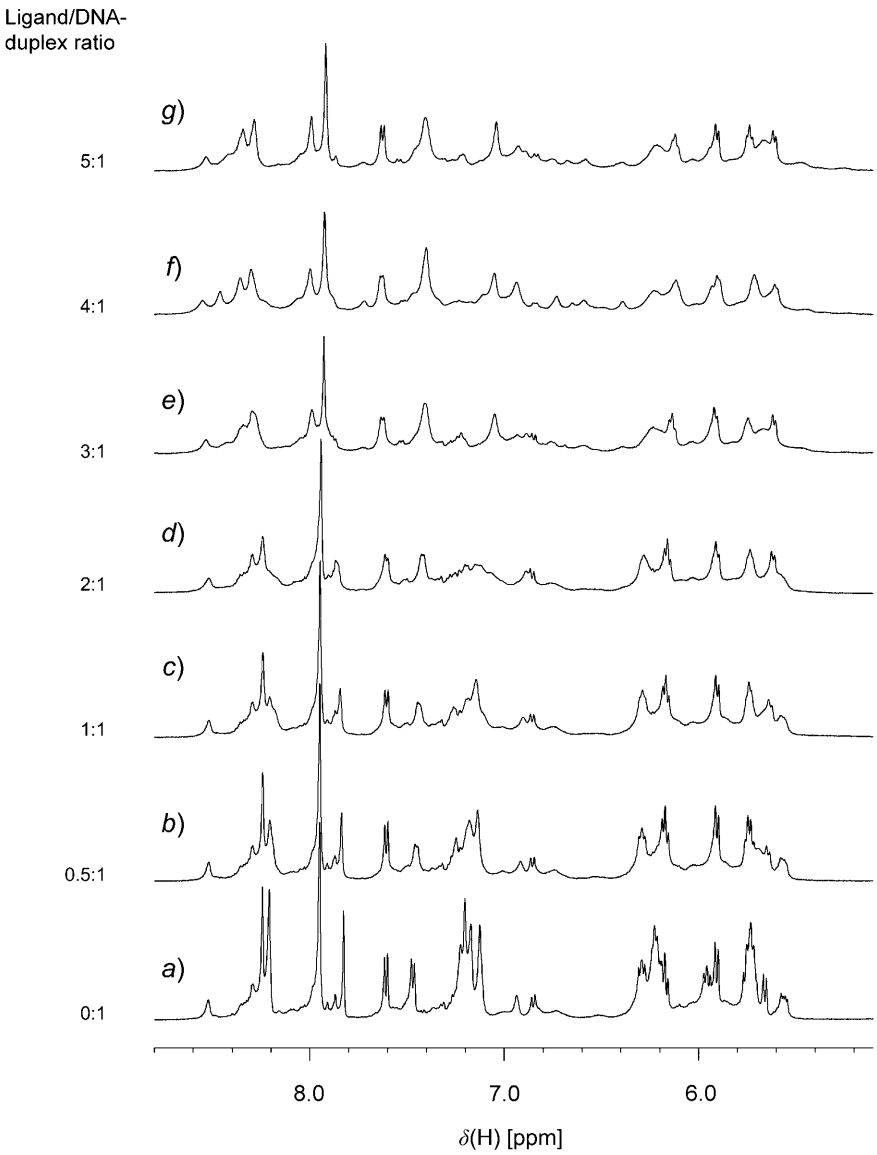


Fig. 5. *Aromatic and H1′ resonance region of the 600-MHz 1D* ^{1}*H-NMR spectrum of 2.52* mM *d(CGATATATCG)*$_2$ (**3**) *in aqueous solution at a sample temperature of 298 K and neutral pH in the absence* (*a*) *and presence* (*b*–*g*) *of increasing quantities of* **2** *in ligand/DNA-duplex ratios:* b) *0.5 : 1,* c) *1 : 1,* d) *2 : 1,* e) *3 : 1,* f) *4 : 1, and* g) *5 : 1*

signals were observed for the complexed form of the DNA. Additionally, the signals for the ligand amino/amide NH H-atoms were broad and relatively weak compared with expectations. With increasing quantities of **2** titrated against **3** up to an excess of 5

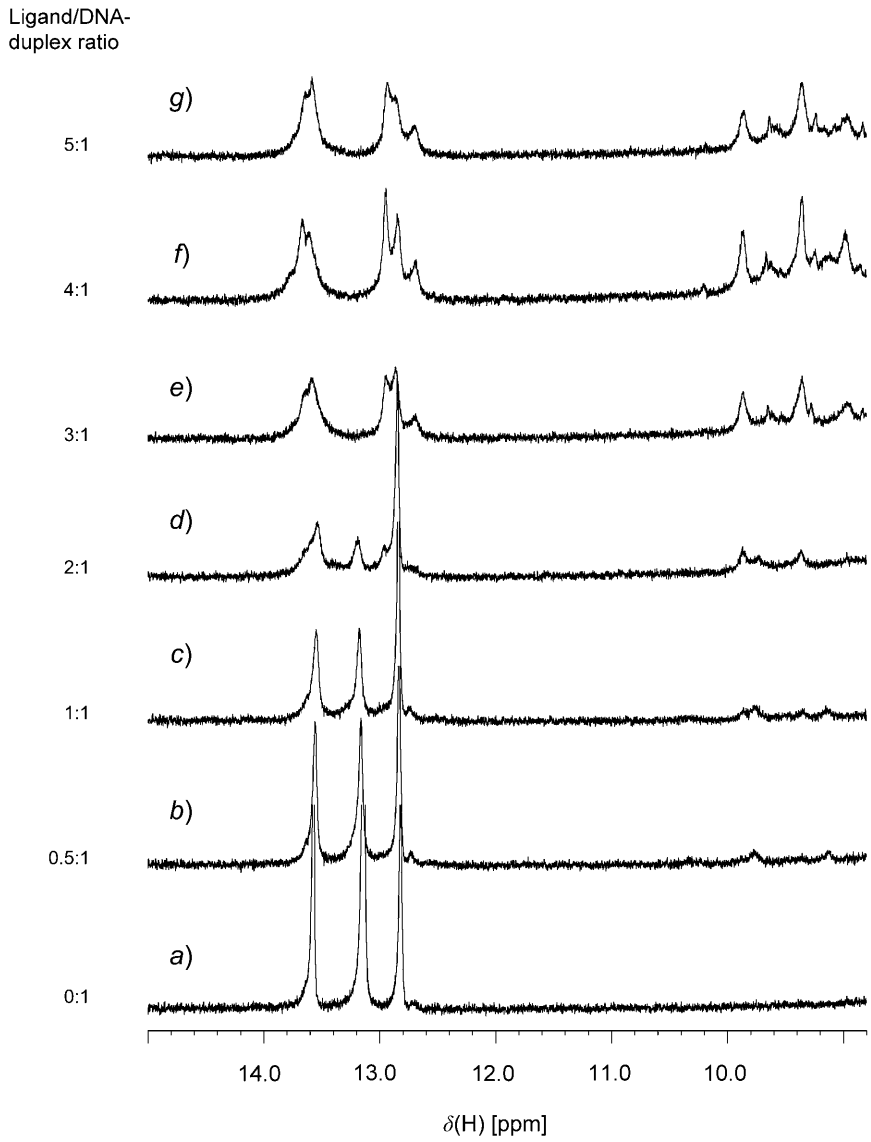


Fig. 6. *Imino and ligand peptide* N**H** *resonance region of the 600-MHz 1D* ^{1}*H-NMR spectrum of 2.52 m*M *d(CGATATATCG)*$_2$ (**3**) *in aqueous solution at a sample temperature of 298 K and neutral pH in the absence* (*a*) *and presence* (*b*–*g*) *of increasing quantities of* **2** *in ligand/DNA-duplex ratios:* b) *0.5 : 1,* c) *1 : 1,* d) *2 : 1,* e) *3 : 1,* f) *4 : 1, and* g) *5 : 1*

equiv. of ligand per DNA duplex, the 1D ^{1}H-NMR spectrum gradually changed with reduction of signal intensity and broadening for certain resonances, but retention of signal intensity, lineshape integrity and gradual drift of chemical shift for other signals. For example, A^{3}H2 signal at *ca.* 7.9 ppm appeared throughout the course of the titration

as a relatively sharp *singlet* signal, and the T^4Me signal at *ca.* 1.45 ppm appeared as a relatively sharp *singlet* compared with the dramatically broadened T^6Me and T^8Me resonances. Representative plots of the chemical-shift changes observed for both the A^3H2 and T^4Me resonances with increasing quantities of **2** (*Fig. 7*) are indicative of intermediate/fast exchange for the binding of **2** to **3** compared with the binding of **1** to **3**. This is remarkable given that the only difference between the two ligands was the substitution of a CHO (in the case of **1**) for an Ac (in the case of **2**) group, *i.e.*, substitution of a H-atom for a Me group at the 'head' end of the molecule.

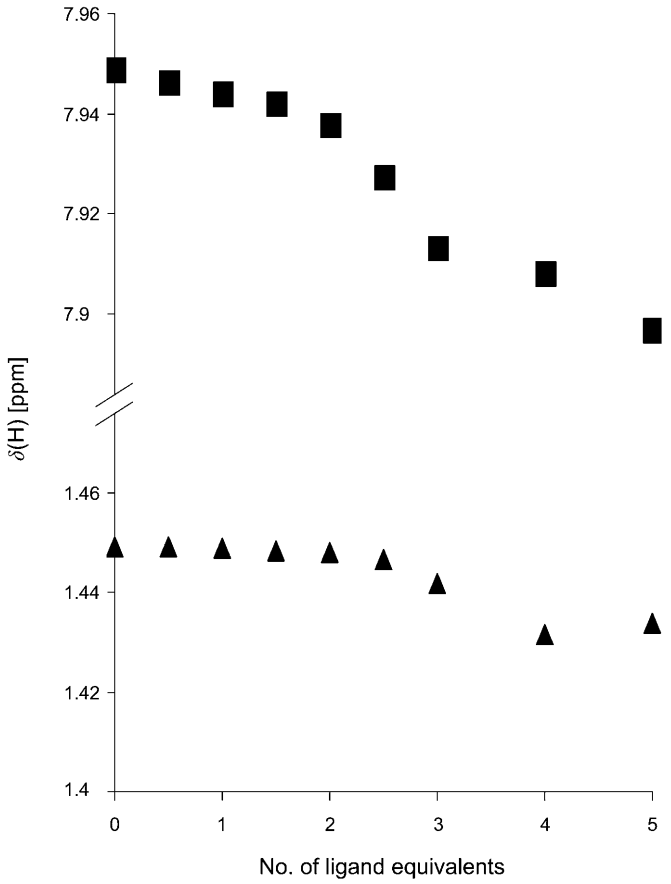


Fig. 7. *Plot of ^{1}H chemical shift* vs. *number of equiv. of* **2** *compared with 1 equiv. of DNA duplex* **3**: *A^3H2* (▲) *and T^4Me* (■)

2.2. *Assignment of ^{1}H-NMR Resonances in the Complex of* **1** *with DNA* **3**. In the slow exchange regime adopted by the complex between **1** and the DNA duplex **3**, it was possible to accumulate relatively good-quality homonuclear 2D ^{1}H,^{1}H-DQF-COSY, TOCSY, and NOESY NMR data in order to establish ^{1}H-NMR signal assignments for the molecules involved in the complex. This was with a view to determining the location and relative positions of ligand molecules with respect to one another and as a

forerunner to calculating a detailed three-dimensional solution structure of the complex. The chemical-shift assignments for the H-atom resonances of **3** bound with 2 equiv. of **1** (*Table 1*) revealed significant differences compared with the chemical shifts recorded for the same H-atom resonances of the free oligonucleotide duplex. Most notable were the largest chemical-shift changes for which $|\Delta\delta| > 0.5$ ppm (where $\Delta\delta$ represents the difference in chemical shift between H-atom resonances of the free DNA and those of the DNA complex with **1**). The biggest changes in aromatic H-atom chemical shifts occurred for H2 H-atoms in residues A^5 ($\Delta\delta = +0.523$ ppm) and A^7 ($\Delta\delta = +1.032$ ppm), which lie on the floor of the DNA minor-groove. Such large deshielding effects are consistent with the nearby juxtaposition of the deshielding plane of an aromatic residue such as the *N*-alkylpyrrole rings. Precedent for such large shifts of A^nH2 H-atoms has been noted previously in our work on a related analogue of **1** and **2** [4], indicating close association of ligand with DNA. These data are useful in assisting with identification of the locality associated with the binding site. Hence, it can be concluded that the sequence of residues $A^5T^6A^7$ are closely associated with the bound ligand assembly. The floor of the DNA minor groove is also associated with imino H-atoms that are part of *Watson–Crick* base pairs. Large chemical-shift changes were also associated with those imino H-atom resonances of the same residues in close proximity to the proposed binding sequence. Hence, the $A^7 \cdot T^4$ base-pair imino H-atom underwent shielding ($\Delta\delta = -0.490$ ppm), and the $A^5 \cdot T^6$ base-pair imino H-atom underwent deshielding ($\Delta\delta = +0.650$ ppm). The $A^3 \cdot T^8$ base-pair imino H-atom experienced a smaller deshielding effect ($\Delta\delta = +0.130$ ppm) upon ligand binding. These data were consistent with anti-parallel self-assembled ligands binding in a centrally located position about the AT-rich region of the oligonucleotide duplex. Significant changes in the furanose sugar-ring H-atom chemical shifts of **3** on ligand binding with 2 equiv. of **1** were also noted (*Fig. 8*). The most significant changes ($|\Delta\delta| > 0.5$ ppm) in the H1′ chemical shifts were observed for residues A^5 ($\Delta\delta = -1.314$ ppm), T^6 ($\Delta\delta = -0.530$ ppm), A^7 ($\Delta\delta = -0.613$ ppm), and T^8 ($\Delta\delta = -0.584$ ppm), thus placing the pyrrole-ring sequence of one ligand against the A^5–T^8 sequence of one oligonucleotide strand. Additional significant changes were also observed for H4′ resonances (*e.g.*, A^7H4′ $\Delta\delta = -0.878$ ppm) and H5″ resonances (*e.g.*, A^7H5″ $\Delta\delta = -1.163$ ppm). Such a large change in H5″ chemical shift was previously observed in this research programme for a residue in the 6- rather than the 7-position of a decamer self-complementary oligonucleotide (*e.g.*, A^6 in CGACTA^{6}GTCG with thiazotropsin A, $\Delta\delta = -1.167$ ppm [4] compared with A^7 in CGATATA^{7}TCG with **1**, this work). The difference in H5″ chemical-shift change between the two cases upon ligand binding is within the experimental error of the NMR chemical-shift measurement and arises in both instances from shielding of the H5″-atom, caused by the close proximity of the magnetic shielding 'cone' of an *N*-alkylpyrrole ring. The difference in the position of the residue affected between the two cases is a subtle indicator of small structural differences in the binding of **1** with respect to its binding site in the DNA minor groove compared with the case for thiazotropsin A associated with its binding site [4]. This may result from differences in steric demand caused by the presence of a bulky *N*-isopentyl group attached to the central pyrrole unit of **1**, which does not occur with thiazotropsin A, a point that is further elaborated *vide infra* in relation to the detailed structure determined for the complex between **1** and **3**.

Table 1. *^{1}H-NMR Chemical Shift Assignments for the Complex between DNA Duplex d(CGATATATCG)$_2$* (**3**) *and Ligand* **1**[a])

Base	Chemical-shift assignment: δ(H) [ppm]														
	H8	H6	H2	H5	Me	H1′	H2′	H2″	H3′	H4′	H5′	H5″	GH1/TH3	H41	H42
C[1]		7.641[a])		5.917		5.762	1.952	2.431	4.728	4.075	3.738	3.738			
		(*0.026*)		(*0.011*)		(*0.015*)	(*0.068*)	(*0.044*)	(*0.028*)	(*0.014*)	(*0.016*)	(*0.016*)			
G[2]	8.002					5.729	2.772	2.898	5.039	4.367	4.261	4.261	12.90		
	(*0.047*)					(*0.190*)	(*0.049*)	(*0.088*)	(***0.709***)	(*0.037*)	(–)	(–)	(*0.060*)		
A[3]	8.290		7.921			6.200	2.681	2.747	5.073	4.453	4.073	4.156			
	(*0.033*)		(*0.107*)			(*– 0.093*)	(*– 0.001*)	(*– 0.228*)	(*0.028*)	(*– 0.046*)	(*– 0.130*)	(*– 0.137*)			
T[4]		7.464			1.430	5.249	2.439	2.558	4.923	4.154			12.64		
		(*0.324*)			(*– 0.014*)	(*– 0.447*)	(*0.327*)	(*0.047*)	(*0.032*)	(*– 0.057*)	–[b])	–	(*– 0.490*)		
A[5]	8.321		7.731			4.914	2.482	2.759							
	(*0.078*)		(***0.523***)			(***– 1.314***)	(*– 0.130*)	(*– 0.171*)	–	–	–	–			
T[6]		6.912			1.546	5.188	1.575	2.196	4.596				13.79		
		(*– 0.264*)			(*0.159*)	(***– 0.530***)	(***– 0.575***)	(*– 0.328*)	(*– 0.306*)	–	–	–	(***0.650***)		
A[7]	8.460		8.126			5.625	2.323	2.669	4.698	3.551	3.838	3.128			
	(*0.216*)		(***1.032***)			(***– 0.613***)	(*– 0.301*)	(*– 0.246*)	(*– 0.309*)	(***– 0.878***)	(*– 0.342*)	(***– 1.163***)			
T[8]		7.027			1.607	5.380	1.900	1.966	4.657				13.70		
		(*– 0.175*)			(*0.279*)	(***– 0.584***)	(*– 0.127*)	(*– 0.460*)	(*– 0.210*)	–	–	–	(*0.130*)		
C[9]		7.405		5.605		5.651	1.976	2.314	4.808	4.052				8.532	6.875
		(*– 0.076*)		(*– 0.059*)		(*– 0.074*)	(*– 0.052*)	(*– 0.062*)	(*– 0.041*)	(*– 0.089*)	–	–		(*0.005*)	(*– 0.078*)
G[10]	7.930					6.145	2.615	2.358	4.674	4.168	4.061	4.061			
	(*– 0.027*)					(*– 0.030*)	(*– 0.005*)	(*– 0.017*)	(*– 0.018*)	(*– 0.024*)	(–)	(–)			

[a]) Chemical-shift values are represented in a non-italicised font. Chemical-shift difference values are represented italicised in brackets below assignment entries within the table as $\Delta\delta = [\delta(\text{ligand-bound DNA}) - \delta(\text{ligand-free DNA})]$. [b]) Resonance unassigned. Colour coding: figures are colour coded blue or red for $|\Delta\delta| > 0.1$ ppm: red: base H-atoms; blue: furanose sugar H-atoms. Bold-typeface entries are for the most significant changes where $|\Delta\delta| > 0.5$ ppm.

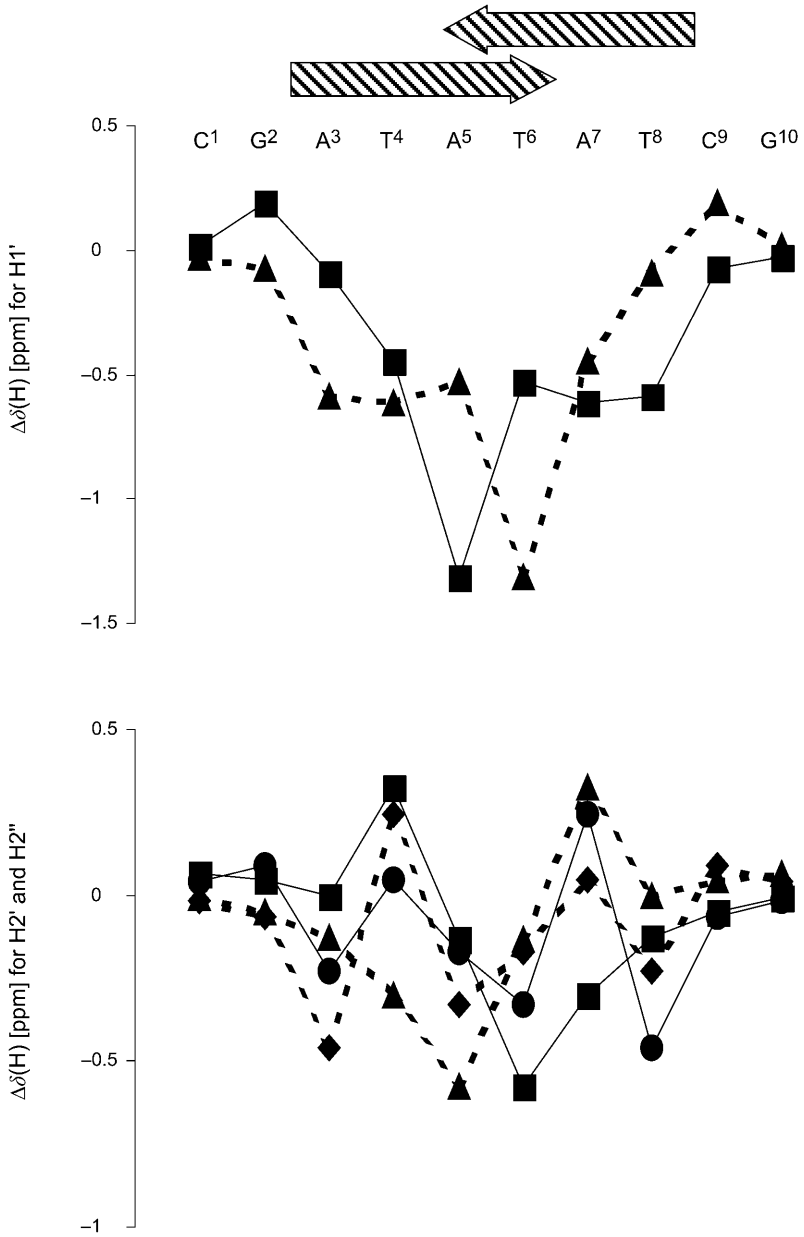


Fig. 8. *Chemical-shift differences for NMR resonances of ligand-bound and ligand-free DNA duplex.* Chemical-shift differences ($\Delta\delta$) are shown for H1′ (*top*) and H2′/H2″ H-atoms as δ(DNA) (free-bound). Shaded arrows above the plots indicate the location of the head-to-tail ligand pair assembly relative to the DNA sequence, indicating the overall footprint for ligand recognition in the minor groove. Dashed lines show the chemical-shift changes for the complementary DNA strand H-atom resonances. Key: bottom: ■ and ▲: H2′ chemical-shift changes; ◆ and ●: H2″ chemical-shift changes.

Assignment of the H-atom resonances for the ligand in its DNA-bound form (*Table 2*) show characteristic chemical shifts for pyrrole, *N*-methyl, peptide N*H*, and aliphatic H-atom resonances. A TOCSY cross-peak between a resonance at δ(H) 2.865 ppm and a resonance at δ(H) 9.051 ppm was initially designated as a correlation between the H-atoms NH31 and H32 of the ligand based on precedent [4]. However, closer inspection showed the signals at δ(H) 2.865 ppm to correspond to Me35/36 resonances implying that the signal at δ(H) 9.051 ppm was that of the quaternary amino H-atom H37, the chemical shift of the resonance presumably reflecting both H-bonding and a formal positive charge at the N-centre. Resonance connectivity between H37 and H34a/b confirmed this assignment, it being possible to subsequently assign the i-Pr H-atoms from further TOCSY correlations. The isopentyl H-atom resonances were assigned through TOCSY correlations that linked the obvious Me21/22 signal (δ(H) 1.079 ppm) with other resonances associated with H-atoms within the isopentyl unit. Narrow resonances at δ(H) 7.454 and 6.330 ppm, which correlated *via* TOCSY cross-peak but no NOE cross-peak, were associated with H13/H16 of the central *N*-alkylpyrrole ring as a result of NOE correlation between the resonance at δ(H) 7.454 ppm and that associated with H17a/b. Hence, the signal at δ(H) 7.454 ppm was assigned to H13, and the signal at δ(H) 6.330 ppm was assigned to H16. A large NOE between the resonance corresponding to H16 and a signal at δ(H) 9.7 ppm was ascribed to through-space proximity between H16 and H23. Based on former precedent for the highest chemical-shift NH resonance being associated with the peptide linkage between the second and third aromatic ring of such an aromatic peptide, it was concluded that the resonance at δ(H) 9.7 ppm was that arising from H23. Subsequently, NOE correlation assignment between H23 and H28 enabled the identities of the H-atoms associated with each remaining *N*-alkylpyrrole ring to be confirmed, thereby enabling virtually complete H-atom resonance assignment of the NMR spectrum of **1** in the DNA-bound form, and thus completing, as far as possible, the unique assignment of H-atom resonances for the 2 : 1 complex between **1** and **3**. Some H-atom resonances remained unassigned due to ambiguity and degenerate shifts, which resulted in some limitation in the subsequent definition of structure determined for the complex. Representative NOE data are shown for the H23 H-atom of **1** (*Fig. 9*) in which intra-ligand, inter-ligand and ligand–DNA close-proximity H–H contacts are all represented.

2.3. *Assignment of* ^{1}H-NMR *Resonances in the Complex of* **2** *with DNA* **3**. As shown in *Figs. 4–6*, onset of significant changes in the appearance of the NMR spectrum of the

Table 2. *^{1}H-NMR Chemical-Shift Assignments of* **1** *in Complex with DNA Duplex d(CGATATATCG)$_2$* (**3**)

δ(H) [ppm]												
H1	H3	H5	H8	Me9	H11	H13	H16	H17a	H17b	H18a/b	H19	Me20/21
–[a])	–	7.298	5.827	3.899	–	7.454	6.330	4.047	4.450	1.805	1.806	1.079
H23	H25	H28	Me29	H31	H32a	H32b	H33a	H33b	H34a	H34b	Me35/36	H37
9.700	6.855	6.485	3.667	8.059	2.781	3.368	1.507	1.776	2.956	3.102	2.865	9.051

[a]) Chemical shift could not be unambiguously assigned.

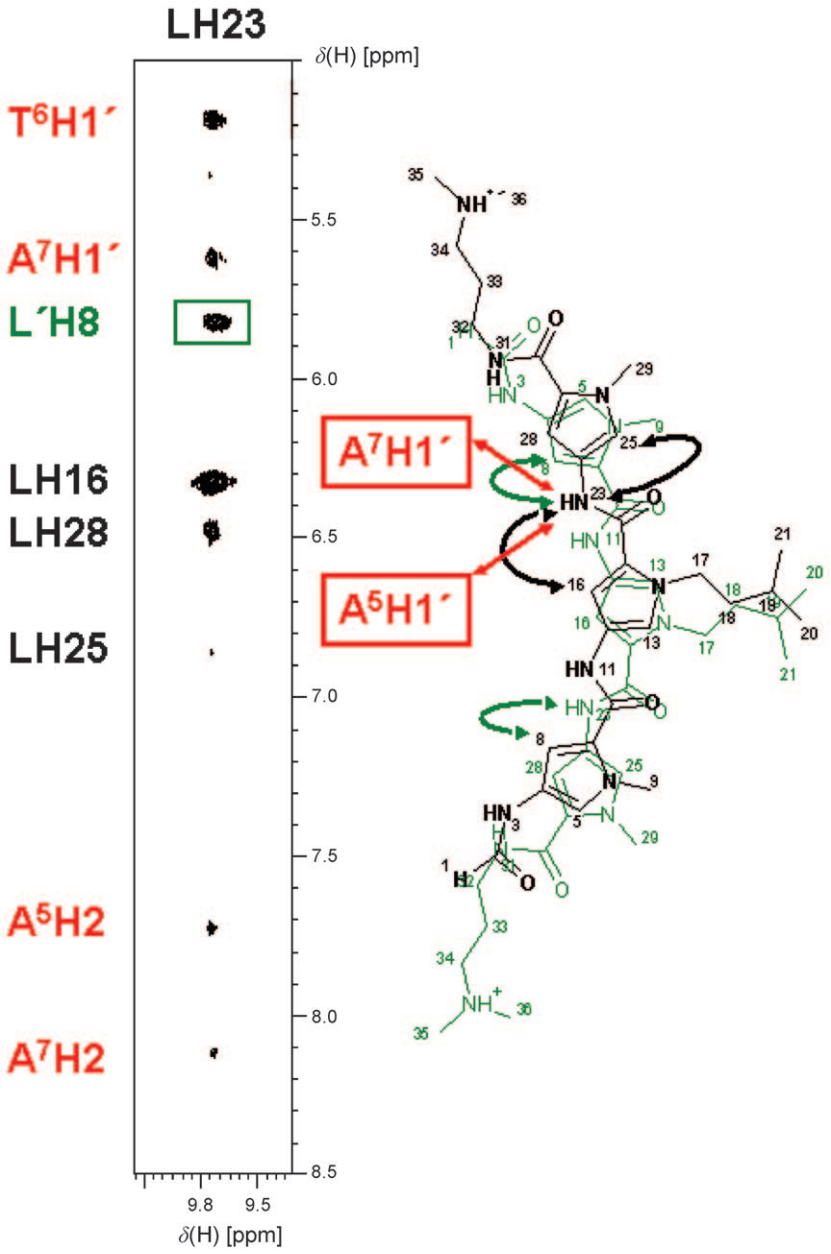


Fig. 9. *Assignment of the cross-peak region at the H23 chemical shift of the bound ligand in the 150-ms 600-MHz 2D NOESY NMR spectrum of the complex between* **1** *and* **3**, *together with a schematic representation of the ligand self-assembly showing* (with arrows) *the close proximity neighbourhoods represented by the data.* Black assignment and arrows: intra-ligand NOEs represented for the structure of **1**. Red assignment and arrows: ligand-DNA NOEs. Green assignment, boxes and arrows: inter-ligand NOEs derived from close proximity between two anti-parallel, side-by-side ligand molecules located adjacent to one another in a manner closely consistent with the NMR data.

DNA **3** on complexation with **2** did not appear until the ratio ligand/DNA was 2 : 1, and further changes in the data were apparent as the ratio of ligand to duplex was increased beyond 2 : 1 up to 5 equiv. of ligand. In the absence of the appearance of distinctly new NMR signals for 'bound' DNA on ligand addition, it was clear that the association of **2** with **3**, in terms of the strength of binding and the on/off chemical exchange rate, was markedly different as indicated by the NMR data. This was also reflected in the appearance of the 2D NOESY NMR spectrum of the complex of **2** with **3** compared with that of **1** with **3**. An overlay of the key fingerprint H1′ to aromatic cross-peak region and of the H2′/H2″ to aromatic cross-peak region of the 150-ms 2D ^{1}H,^{1}H 600-MHz NOESY NMR data for the two complexes is shown in *Fig. 10*. Evidence for very close similarity of the two complexes was apparent from the very tight overlay of many of the cross-peaks for the two complexes, indicating that the DNA had undergone extremely similar structural changes on complex formation with both **1** and **2**. On closer

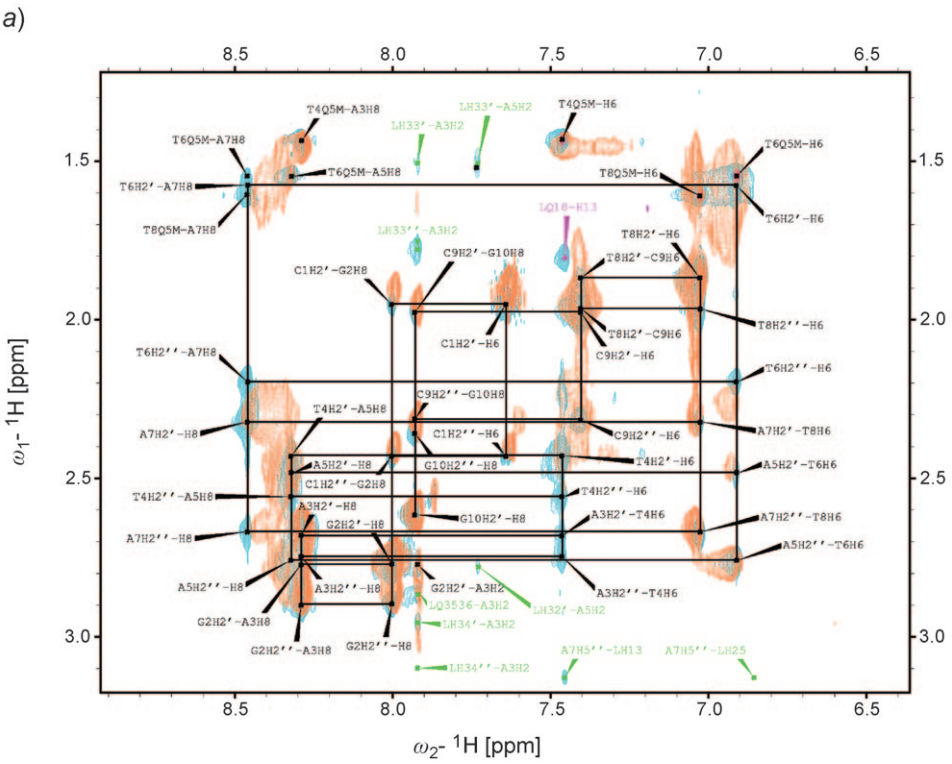


Fig. 10. *Overlay of 150-ms mixing time 600-MHz 2D ^{1}H,^{1}H-NOESY NMR data for the complex between* **1** *and* **3** (cyan-coloured cross-peaks), *and between* **2** *and* **3** (coral-coloured cross-peaks) *for H2′/H2″ to aromatic* (*a*) *and H1′ to aromatic* (*b*) *cross-peak regions.* Cross-peak labels are shown for the complex between **1** and **3** (black labels: intra-DNA cross-peaks; green labels: DNA-ligand NOEs; magenta labels: intra-ligand NOEs). Significant NOE absences are seen for the complex between **2** and **3** by the clear appearance of cyan-coloured cross-peaks without overlay by coral-coloured cross-peaks. Data for the complex between **2** and **3** were acquired at a ratio of 5 equiv. of **2** for each equiv. of DNA duplex **3**.

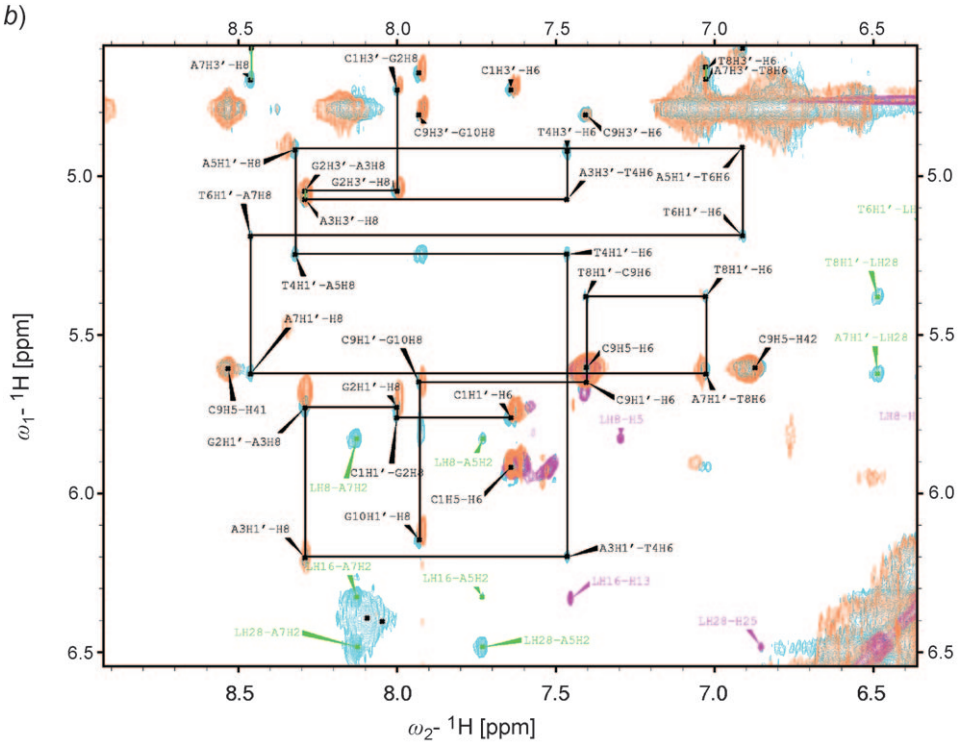


Fig. 10 (cont.)

inspection, however, significant loss of specific NOEs was apparent for the data from the complex between **2** and **3** compared with that from the complex between **1** and **3**. Missing H1′ to aromatic NOEs are detailed in *Table 3*. The data clearly indicated that, in the minor groove of the complex of **2** with **3** for the binding sequence region clearly delineated for the complex between **1** and **3**, an entire sequence of intra- and inter-residue NOEs appeared to be missing from their expected locations. As the chemical-shift position of the aromatic resonances and the assignment continuity in the complex of **2** with **3** were clearly observable in the H2′/H2″ to aromatic cross-peak region of the NOESY NMR data (*Fig. 10*), it was concluded from the data that NOE absences would arise from excessive broadening of one or both partner resonances that should have given rise to the expected NOE. The exchange characteristics or mobility of a ligand bound to the DNA that would give rise to the presence or absence of an expected resonance clearly impacted the overall appearance of the data in this case. This was all the more remarkable given the very slight difference in the structures of **1** and **2** at the 'head' of the molecule, which was not predicted to make such a major contribution to the mode of ligand binding. The fact that absences were apparent for the H1′ associated NOE cross-peaks and not for the H2′/H2″-related NOE cross-peaks was indicative of exchange broadening dynamics being associated with the minor rather than the major groove, supporting the assertion that **2** does bind to the DNA minor groove of duplex **3**

but in a more dynamic mode than the binding of **1** to the same structure. The nature of the DNA chemical-shift similarity for the two complexes was supportive of similar structures for the bound DNA duplex in both cases, and it could, therefore, be concluded that the ligand-exchange characteristics were responsible for imposing the resonance-broadening effects observed in the NMR data for **2** binding in association with **3**. As full complex formation was only achieved with an excess of **2** (as shown by the nature of the appearance of the 1D ^{1}H-NMR data sets shown in *Figs. 4–6*) it was clear that the 'looser' nature of the **2/3** complex arose from a different exchange equilibrium position in which excess ligand is required to force the complex to be the predominant species in solution. Whilst the key CHO/Ac difference between the two ligands was the cause of the difference observed in this behaviour, it was not possible to examine in detail the precise interactions that these groups had either with neighbouring ligand or with DNA. In the CHO case (*i.e.*, **1**), the resonance for H1 could not be unambiguously assigned due to extensive resonance overlap. In the Ac case (*i.e.*, **2**), the Me resonance was not assignable or definitely distinguishable from among the other resonance due to the broadened nature of the NMR data. Nevertheless, the effects transmitted to the rest of the data by these differences were enough to allow insight into the nature of the differences in complex formation to be established.

Table 3. *Presence* (✓) *or Absence* (x) *of Expected Intra- and Interresidue Aromatic-H1′ Cross-Peaks for the Complexes between* **1** *and* **3**, *or between* **2** *and* **3**

Expected NOE	Complex **1/3**	Complex **2/3**
C^1(H6-H1′)	✓	✓
C^1H1′–G^2H8	✓	✓
G^2(H8–H1′)	✓	✓
G^2H1′–A^3H8	✓	✓
A^3(H8–H1′)	✓	✓
A^3 H1′–T^4H6	✓	x
T^4(H6–H1′)	✓	x
T^4H1′–A^5H8	✓	x
A^5(H8–H1′)	✓	x
A^5H1′–T^6H6	✓	x
T^6(H6–H1′)	✓	x
T^6H1′–A^7H8	✓	x
A^7(H8–H1′)	✓	x
A^7H1′–T^8H6	✓	✓
T^8(H6–H1′)	✓	✓
T^8H1′–C^9H6	✓	✓
C^9(H6–H1′)	✓	✓
C^9H1′–G^{10}H8	✓	✓
G^{10}(H8–H1′)	✓	✓

2.4. *Solution Structure of the Complex between* **1** *and* **3**. A total of 170 NOEs were assigned for the complex of **1** with **3** (available as *Supplementary Material* upon request from the authors), which represented 340 distances based on symmetry considerations. Of the assigned NOEs, 27 represented ligand – DNA contacts, 19 represented intra- and

inter-ligand contacts, and five represented interstrand contacts. The remaining 119 NOEs arose from the DNA. Representation of the structure (*Fig. 11*) shows the overall nature of the complex formed between **1** and **3** in aqueous solution. For 50 structures selected from the end of the rMD production run, the all-atom r.m.s.d. was 0.58 Å with a standard deviation of 0.15 Å. Inter-digitation of alkyl groups on the outer edges of the opposing ligands allowed them to fit neatly and snugly against one another and against the 'walls' of the DNA minor groove formed by the sugar rings (*Fig. 11*,*a*). The combined ligand assembly occupied a widened minor groove. For the mean structure, 3DNA-calculated P–P distances across the minor groove in the central region showed only small variations between 11.7 and 13.0 Å, and sugar *pseudo* rotation angles generally lay in the S-form. The *N*-isopentyl groups, whilst being relatively compact, lay slightly proud of the minor groove. Being hydrophobic in nature, these elements of the ligands, by lying in close proximity to one another on opposing strands, may be expected to associate through hydrophobic interaction with one another. This is discussed in due course with regard to the reference frame of thiazotropsin A bound to

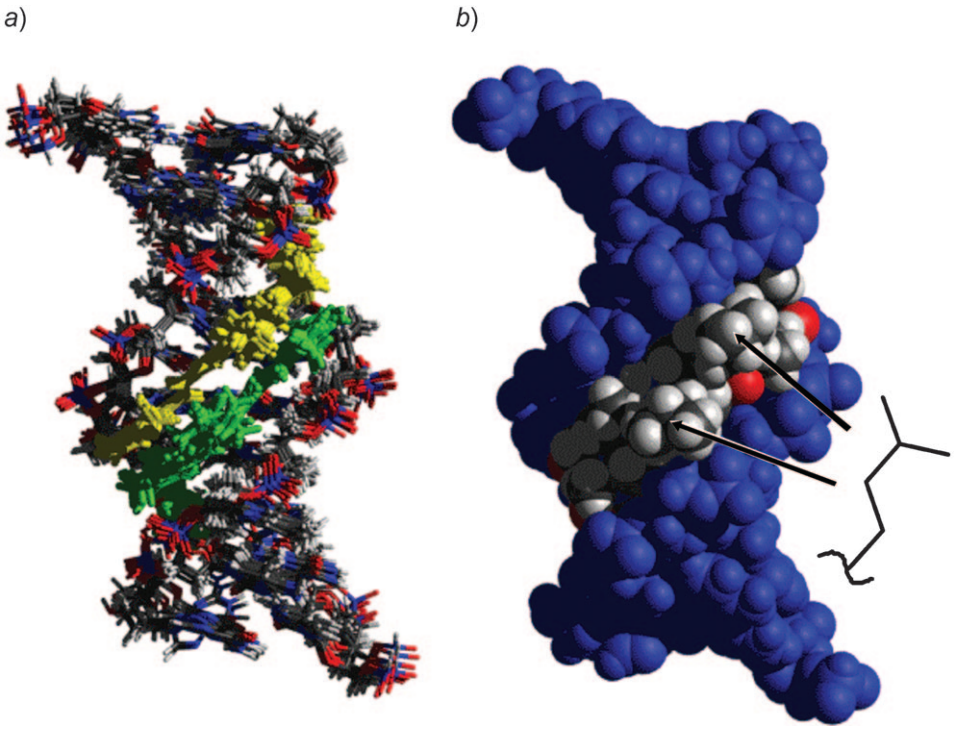


Fig. 11. *Representations of the structure of the complex between* **1** *and* **3**. *a*) Stick representation overlay of nine minimized structures from the end stages of the rMD production run together with the minimized mean structure from 50 minimized snapshots. Side-by-side ligands are coloured separately (*yellow* and *green*), and DNA is coloured according to atom type. *b*) Mean structure represented by *Van der Waals* surfaces for both DNA (*blue*) and ligands (coloured accorded to atom type) showing location of the *N*-isopentyl groups.

its preferred recognition sequence. DNA-Structural characteristics represented by the mean structure for the central region of the complex are detailed in the *Supplementary Material*. Although these features cannot be rigorously interpreted, it was interesting to note that some symmetry in their values existed. For instance, base-pair shear (S_x) and buckle (κ) values varied in a manner that was consistent with a symmetrical head-to-tail 2:1 ligand/DNA complex. In the same vein, inter-base-pair parameters including shift (D_x), slide (D_y), rise (D_z), roll (ρ) and twist (Ω) showed similar trends.

Although, on the NMR timescale, the NMR data represented a single symmetrical structure, the calculated mean structure allowed for some speculation regarding the orientation of bonds and the existence of potential H-bonding networks. For instance, two possible conformers of the CHO headgroup peptide bond were represented within the calculated mean structure. One ligand was present with the peptide bond in a conformation such that C=O O-atom and NH H-atom were *trans* to one another, whereas the opposing ligand showed the conformation about this bond to be *cis*. Due to the lack of NMR resonance assignment for signals arising from H1 and H3 of **1** in the current study, it was not possible to apply distance range restraints to these atoms due to lack of NOE information. Similarly, the conformation about the peptide bond associated with the 3-(dimethylamino)propyl tail was represented in the structure by one ligand displaying *cis* geometry and the opposing ligand displaying *trans* geometry about the C30–N31 peptide bond. In this case, the assignment of N*H*31 was made, but intermolecular NOEs between DNA and ligand were not observed. Without restraint, therefore, the orientation about this bond was subject to variability as shown by the results of the rMD calculation. These features of the structure in relation to the NMR data are discussed further below.

3. Discussion. – 3.1. *Nature of the Ligands.* Ligands **1** and **2** represent early synthetic elaboration by us of the lexitropsin theme by increasing the hydrophobicity of small ligands with the long-term view of developing effective pharmaceuticals based on DNA minor-groove recognition. This field has met with varying degrees of success over the years in terms of DNA sequence reading, quality of complex formation and resulting biological activity. Therefore, understanding the detailed influence of structural, electronic, topological, and functional features of a ligand's DNA recognition capability has been and still remains a requirement for reaching the goal of designing effective DNA-directed sequence-reading small molecules. Our early synthetic elaborations relied upon the attachment of more bulky alkyl substituents to what would be expected to represent the solvent-exposed edge of the groove binder when located in a DNA duplex minor groove. Increased association of ligands to one another through hydrophobic interactions together with increased ligand–DNA association through contact between the DNA minor groove 'walls' and the ligand's hydrophobic features were anticipated to result in better overall ligand–DNA binding. Use of different headgroups including CHO and Ac functional groups as in this case was not anticipated to result in significant differences in the extent and quality of ligand–DNA complex formation. Being relatively straightforward to create, it was in the nature of a synthetic programme of research to manufacture a library of molecules capable of being tested empirically for biological activity and used in comparative structural studies. Both the CHO and Ac ligands (in **1** and **2**, resp.) were anticipated to bind to an

AT-rich region of DNA based on precedent for similar ligands in other work. For ease of NMR-data assignment, an alternating purine/pyrimidine DNA sequence was selected as the testing recognition target for both ligands. Whilst the recognition sequence was not, therefore, optimised by other means, we expected it to yield sufficient information for comparing the DNA-binding characteristics of the two ligands.

3.2. *NMR-Based Observations.* Our comparative NMR study clearly revealed a difference in the nature of the way in which ligands **1** and **2** associated with the selected recognition sequence. As shown in the figures representing complex formation in the two cases (*Figs. 1–3* and *Figs. 4–6*, resp.), significant qualitative differences were immediately apparent from the ^{1}H-NMR data. A sharp titration end-point was observed in the binding of **1** to **3** at a ligand/DNA-duplex ratio of 2:1, when the entire free DNA was consumed as a ligand–DNA complex (complete loss of free DNA ^{1}H-NMR signals and replacement with ^{1}H-NMR signals from ligand-bound DNA). The same was not true of **2** binding to the chosen DNA sequence. Gradual increase of the relative quantity of ligand to DNA duplex, up to an excess of 5 equiv., resulted ultimately in the on/off free/bound equilibrium being pushed towards the bound limit of complex formation as evidenced by the significant change observed in the ^{1}H-NMR data towards the 5-equiv. titration point. Due to the looser nature of this complex (as evidenced by a much broader, lower intensity set of NMR responses), the quality of subsequent NMR data sets acquired on it were significantly lower than those acquired for the tighter complex formed between **1** and **3**. In our experience, ^{1}H-NMR resonance lineshape and physical clarity of a ligand/DNA titration solution can invariably be used as very good qualitative determinants of the ligand fit to the DNA recognition sequence. This study was no exception in this regard. Good quality, well-resolved and assignable NMR data were acquired for the complex between **1** and **3** from which it was possible to clearly assign the vast majority of NMR responses to specific H-atoms within the structure. However, based on this assignment, it was also possible to retrospectively examine similar data for the complex of **2** with **3**, when it became apparent that only certain NMR resonances were unobservable within the 2D data sets. As shown in the overlay comparisons for the two complexes (*Fig. 10*), large portions of the intra-DNA 2D NOESY NMR data were reproduced in the two complexes in a virtually identical manner under conditions where complex formation predominated (ligand/DNA duplex ratio of 2:1 in the case of **1** and 5:1 in the case of **2**). Chemical shift, in particular, is a sensitive qualitative probe of differences between molecular structures. The remarkable similarity between these data is, therefore, good evidence for concluding that the DNA structure responded in a similar way to the presence of each ligand. Hence, whilst no structure calculations were feasible based on the NMR data for **2** with **3**, it may still be concluded that complex formation occurred in a 2:1 head-to-tail manner, albeit in a rather 'looser' form. In essence, the presence of the additional bulky Me group in the case of **2** was enough to cause disruption between the way in which the ligands themselves self-associate in the minor groove. This is considered further in this discussion in terms of the structure determined for the complex between **1** and **3**. Finally, in consideration here of the retrospective analysis of the NMR data acquired with the Ac-bearing ligand **2**, it was possible to track gradual shift changes for several resonances that were identifiable from a comparison with the

data for the DNA complex of the CHO-bearing ligand (*Fig. 7*). This type of resonance behaviour is typical of a more rapid on/off chemical equilibrium relative to the NMR timescale. Distinct resonances for bound and free forms of the molecule present in excess (in this case, DNA) are not observed but instead an average chemical shift results, which is dependant upon the relative populations of the free and bound forms of the molecule present in solution at a given ligand/DNA ratio. The fact that specific aromatic-to-H1′ NOE cross-peaks were absent from the NMR data for the DNA complex of **2** together with associated intermolecular NOEs between H-atoms on the inner edge of the ligand and DNA H1′-atoms was a clear indicator that more significant mobility was occurring in the DNA minor groove for the complex of **2** compared with that of **1**, a point which is further elaborated upon below.

3.3. *Overall Structure of the Complex between* **1** *and* **3**. Virtually complete assignment of both ligand and DNA H-atom resonances for the CHO complex made it possible to determine a detailed structure of the complex formed based on assigned NOE data and rMD simulations. Whilst manual docking and manipulation of such complexes within a computer modelling framework can be helpful and sometimes essential for making sense of such NMR data, it is more acceptable to allow rMD simulation to refine the complex to satisfy as many restraints as possible and to fine-tune the relative position of the ligands both with respect to the DNA binding site and with respect to one another in a self-assembly context. Qualitative data (*e.g.*, $\Delta\delta$ information as shown in *Fig. 8*) are valuable for particularly helping to identify the effects of nearby ring currents arising from unsaturated aromatic rings, which directly assist with establishing the relative location of ligand with respect to the DNA sequence. Significant change ($\Delta\delta = -1.314$ ppm; *Table 1* and *Fig. 8*) in the A^5H1' chemical shift was a marker for the close proximity of a pyrrole ring from **1**. NOE Data and computer modelling firmly located the *N*-methylpyrrole ring adjacent to the CHO headgroup of **1** in close juxtaposition to A^5H1' (see *Figs. 8* and *12*,*a*). All NOEs along the length of each ligand inner edge between ligand and DNA were consistent with this location. Inter-ligand NOEs placed ligands in side-by-side opposition to one another in such a way that the terminal *N*-methylpyrrole ring of one ligand (adjacent to the tail) overlapped to a large extent with the *N*-methylpyrrole ring at the head of the neighbouring ligand (*Figs. 9* and *12*,*b*). Space-filling representation (*Fig. 11*,*b*) indicated the manner in which the *N*-alkyl groups nestled against one another, reinforcing the self-assembly character of the ligands through hydrophobic interaction.

3.4. *Conformation, Location and Binding of the Ligands.* As noted from our results, sampling of conformational space within the rMD run revealed the possibility of two conformers for the peptide bonds associated with either end of each ligand. In the absence of NOE restraints, either through lack of signal assignment or due to the apparent lack of NOE responses, the conformations about both peptide bonds were free to adopt the most appropriate fit. Due to the lack of experimental data representing these features of the structure, it was not possible to state categorically which was the preferred conformer at either end of each ligand. However, extended unrestrained molecular-dynamics (MD) simulations of a related complex with thiazotropsin A bound to its recognition sequence d(ACTAGT) have shown that lower-energy structures resulted from complexes in which the conformation about the CHO headgroup in particular was *trans* (*i.e.*, typical peptide-bond conformation),

which, in the case of thiazotropsin A, also allowed for the predicted formation of an additional stabilizing H-bond. From our experimental work, NOE restraints were derived and applied to the CHO headgroup NH H-atom, which showed a visible and clearly assignable H-atom NMR response [4]. In the current case (this work), a H-bond was predicted between N*H*3 of the ligand and A^5N3 of the DNA minor groove for the *trans* orientation of the CHO headgroup. Similarly, when the C30–N31 peptide bond was *trans*, N*H*31 of **1** was predicted to form a stabilizing H-bond with O2 of residue T^4. Details of these and other predicted H-bonds are available as *Supplementary Material*. Although the models represented both *cis* and *trans* possibilities for both of these peptide bonds, it would be most likely and energetically more favourable if both peptide bonds were *trans* in both cases of opposing ligands. Even in the presence of good-quality NOE data, it remains difficult to determine what the actual peptide bond conformations are. A more precise way of determining this information for N–H···N and N–H···O=C H-bonds would be by incorporation of ^{15}N into the ligands and by ^{13}C and ^{15}N incorporation into nucleic acids in order to observe through-bond ^{h2}J(N,N) and ^{h3}J(N,C′) couplings between peptide N*H* and the nucleic acid [5]. This would provide a measure of the relative strength and, at the same time, importance of ligand–nucleic acid H-bonds, addressing the matter of the extent to which H-bonding between ligands and nucleic acids in the DNA minor groove direct and enable complexation to occur. We are in the process of pursuing this line of enquiry. In the meantime, the study of **1** in complex with **3** predicts H-bonding to occur after the usual fashion for aromatic peptide minor-groove binders (see *Supplementary Material*). Of particular interest is the prediction of H-bonding between the H-atom associated with the N-atom of the 3-(dimethylamino)propyl moiety and two O-atoms in the sugar-phosphate DNA backbone. One of these H-bonds was predicted to be to the O-atom of the sugar ring associated with $G^{2/20}$, and the other was predicted to be to the O-atom at the 3′-position of the $C^{9/19}$ DNA residue. Interestingly, the model suggested that the terminal dimethylamino groups were relatively deeply buried in the minor groove, thus creating an anchor through which the ligand binding would be electrostatically and sterically stabilized.

3.5. *Comparison with Related Ligand–DNA Complexes.* The greater significance of these findings lies in the context of a more comprehensive study in which ligand-structure variations are being explored in a bid to understand structure–activity relationship (SAR). Our report concerning thiazotropsin A (**4**) in the binding context of ACTAGT [4] was followed up in a related article that described the DNA recognition capability of the very close analogue thiazotropsin B (**5**), which showed a DNA binding preference for (A/T)CGCG(T/A) [6]. While **5** differed from **4** in the key attribute of an imidazole N-atom in place of sp^2-CH in the *N*-methylpyrrole (ring *A*) adjacent to the headgroup (thus conferring its altered sequence-recognition properties), significance in the context of the current work lies in the CHO group of **4** being replaced by Ac in **5**. Our NMR investigations of **5** in complex with a (A/T)CGCG(A/T) self-complementary recognition sequence show excellent-quality NMR data with strong evidence for tight binding of **5** in a head-to-tail 2:1 ligand/DNA-duplex ratio, details of which will be reported in due course. Whilst the sequence contexts are different, we see little difference in the quality of data acquired for the DNA complexes of either compound. In both instances, the data quality is excellent with well-resolved

resonances throughout and complexes that exist in slow exchange on the NMR timescale. Comparison with the current study shows that subtle structural differences in the DNA recognition site are likely to also play a significant role in creating the right conditions for ligand binding (*e.g.*, minor-groove width is known to be wider in a GC-sequence rich context compared with an AT-sequence rich context). It is, therefore, necessary to consider each compound on it own merits. In this respect, the current study is helpful through its ability to show within a single DNA sequence context the effects on binding caused by a minor structural alteration. The steric interactions between the head of one ligand and the tail of another were satisfactory, when the steric bulk of the headgroup was small but became unsatisfactory with the introduction of a more sterically demanding Me group (*Fig. 12*, *c* and *d*). It will remain of interest to us to understand the precise differences between the structure determined in this work, and those of the DNA complexes formed by **4** and **5** with their respective recognition sequences, which display the same CHO/Ac headgroup difference. These findings will be reported in due course as part of our ongoing investigations into the structural and recognition consequences of our ligand-design strategy.

4. Conclusions. – Complex formation between an alternating AT-rich DNA duplex and two closely analogous lexitropsin molecules has been studied in detail by NMR spectroscopy. A minor functional group difference in the headgroup that distinguished the two ligands resulted in significant differences in the relative strength of complex formation as reflected in the differences in the quality of NMR data acquired for each complex system. Titration of DNA with the CHO headgroup ligand analogue resulted in clean formation of a complex undergoing slow exchange on the NMR timescale at a ligand/DNA-duplex ratio of 2 : 1. The complex formed was a classical head-to-tail side-by-side ligand DNA minor groove-bound system. It displayed all of the NMR-related characteristics for such a complex including significant chemical-shift changes and many ligand – DNA NOE responses, which were assigned as fully as possible. Ambiguity in some signal assignments meant that certain atoms could not be restrained in the resulting rMD simulations. Nevertheless, a structure for the complex between the CHO headgroup analogue, **1**, and the DNA **3** was obtained and was compared with the complex reported previously for a related ligand, thiazotropsin A. Despite initial cursory appearances that a 1 : 1 complex was formed, the Ac headgroup ligand analogue **2** was shown to actually form a similar type of complex with the same sequence of DNA but at a faster on/off exchange rate compared with the NMR timescale. Despite the inadequacy of the resulting NMR data for detailed structure determination, sufficient similarity in the NOE data of the ligand-bound DNA was enough to conclude that complex formation occurred in a similar way compared with the complex formed between the CHO headgroup ligand analogue, **1**, and DNA, albeit requiring a significant excess of ligand to push the exchange equilibrium over to the extent that ligand – DNA complex was predominant in solution (5 : 1 ligand/DNA-duplex ratio). Comparison of the general shape of the imino H-atom signal envelope of the 1D ^{1}H-NMR spectrum at the titration end-points were similar pointing to equivalent binding contexts. Consideration of the unexpected effect of what was originally anticipated to form a minor and insignificant alteration in structure with respect to ligand-binding strength was made in terms of steric demand and opposition between

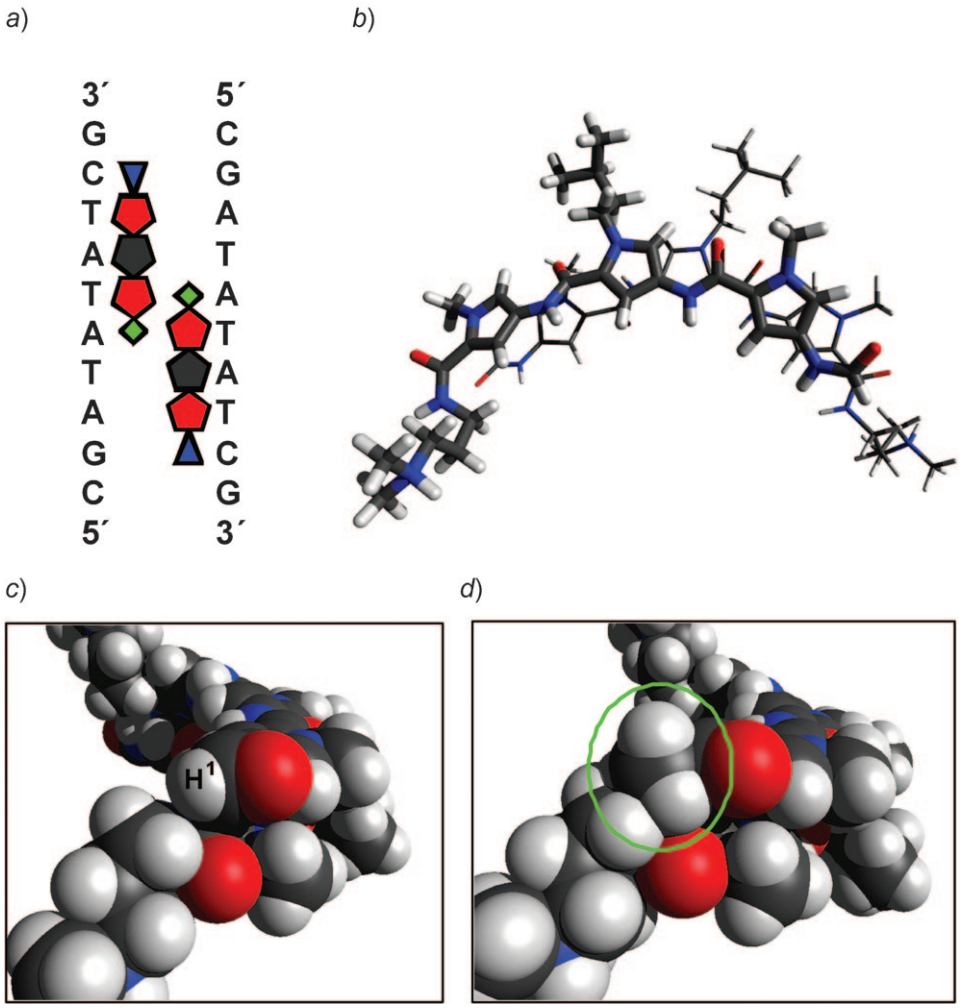


Fig. 12. *Schematic of complex and relative ligand proximity for the complex between* **1** *and* **3**. *a*) Schematic showing proximity and location of opposing ligands in the DNA minor-groove lined up against DNA sequence. *b*) Stick representation of opposing ligands from the mean structure of the complex. The uppermost ligand represented by thicker bonds is orientated with tail to the left and head to the right. The lower opposing ligand is represented by thinner bonds. *c*) Perspective view showing only space-filling representation of the two opposing self-assembled ligands in which atom H1 of **1** is shown highlighted. *d*) As for *c*) but with direct replacement of H1 by a Me group to represent **2** (circled), and showing the likely steric clash is responsible for causing disruption of complex formation in the case of **2**.

self-assembled ligands. Further evaluation of the significance of this will be possible on completion of a related study with an Ac headgroup bearing ligand, thiazotropsin B (**5**), which is known to bind in slow exchange on the NMR timescale to a recognition sequence of the form (A/T)CGCG(T/A), the detailed structural findings of which will be reported in due course. The 'staggered' self-assembly of such lexitropsins must be

considered very carefully when using rational design strategies to develop minor-groove binding, DNA recognizing, sequence reading molecules which fall into the 2:1 ligand/DNA-duplex binding regime.

Experimental Part

1. *Materials.* *3-{4-[4-(4-Formamido-1-methyl-1*H*-pyrrole-2-carboxamido]-1-isopentyl-1*H*-pyrrole-2-carboxamido]-1-methyl-1*H*-pyrrole-2-carboxamido}*-N,N-*dimethylpropan-1-aminium* (**1**) and *3-{4-[4-(4-acetamido-1-methyl-1*H*-pyrrole-2-carboxamido)-1-isopentyl-1*H*-pyrrole-2-carboxamido]-1-methyl-1*H*-pyrrole-2-carboxamido}*-N,N-*dimethylpropan-1-aminium* (**2**) were prepared as TFA salts as described in [7]. The self-complementary DNA sequence d(CGATATATCG)$_2$ (**3**) was supplied by *AlphaDNA* (Montreal, Canada) based on a 15-μmol synthesis followed by oligonucleotide purification cartridge (OPC) purification with additional desalting steps, and was used as supplied from a desalted, lyophilised powder. For NMR purposes, **3** was solubilized in 550 μl of a mixture of 90% H_2O/10% D_2O to a final concentration of 2.3 mM in DNA duplex prior to dilution effects caused through the addition of ligand soln.

2. *Method.* 2.1. *DNA-Complex Formation.* Complexes of **1** and **2** were prepared in a similar way. 1.5 mg of **1** were dissolved in 60 μl of 90% H_2O/10% D_2O to provide a stock soln. of ligand at a concentration of 37.5 mM. DNA **3** was dissolved in 550 μl of 90% H_2O/10% D_2O. Ligand **1** (50 μl of the ligand stock soln.) was added to **3** in 10-μl aliquots using a *Hamilton* syringe, and the soln. was gently mixed to disperse the immediate precipitate that formed on first contact between the ligand and DNA solns. Complex formation was monitored by acquisition of 1D ^{1}H-NMR spectra up to a total of 2 equiv. of **1** per equiv. of DNA duplex. Similarly **2** (2.5 mg) was dissolved in 100 μl of 90% H_2O/10% D_2O to provide a stock soln. of ligand at a concentration of 36.7 mM. Aliquot additions were monitored by 1D ^{1}H-NMR spectroscopy at each stage, and the ligand was added to the DNA soln. up to a total of 5 equiv. of **2** per equiv. of DNA duplex. The imino H-atom ^{1}H-NMR resonance regions of the NMR spectrum were monitored in each case to assess the extent of complex formation. At the end of each titration, the samples were each used in the accumulation of complete 2D ^{1}H-NMR data sets for the purposes of structure determination.

2.2. *NMR Spectroscopy.* NMR Data were acquired under VNMR on a *Varian UNITYINOVA* 600-NMR spectrometer operating at 599.89 MHz for H-atom resonance. A standard geometry triple-resonance probe-head equipped for *z*-pulsed field gradients was used. Data acquisition was carried out in an identical manner for both **3**, and for the complexes between **1** and **3** and between **2** and **3** unless otherwise stated. The probe temp. was maintained at 298 K in all instances. 1D ^{1}H-NMR Data were acquired using either presaturation or a double-pulsed-field-gradient-spin-echo (dpfgse) [8] approach to eliminate the solvent resonance. Data were typically acquired with 128 transients over a frequency width of 15 kHz (25 ppm) centred at 4.702 ppm into 32 K data points (acquisition time: 1.09 s) using a 90° hard pulse and a recovery delay of 2.0 s. For the dpfgse routine, rectangular soft pulses (bandwidth 125 Hz) were used for selective inversion at the solvent frequency together with sine shaped gradient pulses (1-ms duration) in the amplitude ratio 31:11. 2D-NMR Data sets were acquired as follows: 2D ^{1}H,^{1}H-DQF-COSY NMR data were acquired with eight transients for each of 256 states t_1 increments over a frequency width of 5.48 kHz in both ω_2 and ω_1 (9.14 ppm) into 4 K complex data points (acquisition time 374 ms) with a recycle time of 1.5 s. 2D ^{1}H,^{1}H-TOCSY NMR data were acquired with eight transients for each of 256 states t_1 increments over a frequency width of 5.4 kHz in both ω_2 and ω_1 (9.14 ppm) into 4 K complex data points (acquisition time 341 ms) with a recycle time of 1.5 s and a pulsed spin-lock time of 55 ms. 2D ^{1}H,^{1}H-NOESY NMR Data were acquired with 16 transients for each of 400 states t_1 increments over a frequency width of 10.5 kHz in both ω_2 and ω_1 into 8 K complex data points with a recycle time of 2 s, and a mixing time of 100 ms for a total accumulation time of 14 h. All NMR data were converted and processed on a *Dell Precision 340* workstation running under *Microsoft Windows 2000* using *TopSpin* (version 2.1, *Bruker Biospin*, D-Karlsruhe) with appropriate processing parameters and imported into SPARKY (version 3.105) [9] for data reduction and analysis. All NMR data were collected on samples admitted to *Wilmad 528-PP* 5-mm NMR tubes (*Goss Scientific*).

2.3. *NMR Data-Assignment Strategy.* 2D ^{1}H,^{1}H-NOESY NMR Data were assigned for both free duplex **3**, and for the complexes between **1** and **3** and between **2** and **3** by using established assignment strategies for right-handed B-form DNA [10][11]. 2D ^{1}H,^{1}H-DQF-COSY and TOCSY NMR data were used to establish resonance assignments for specific H-atoms of **1** and **2** in their complex with **3**.

2.4. *Modelling Restraints.* Quant. inter-H-atom distance restraints were calculated from 100-ms NOESY data with an iterative relaxation matrix approach using the program MARDIGRAS [12][13]. Averaged cross-peak volume integrals measured from 2D NOESY NMR data (*Sparky*, version 3.105) were used in relaxation matrix calculations starting from canonical B-form DNA models. For the complex between **1** and **3**, initial restraints were defined as fixed distances with force constants of 1 kcal mol^{-1} Å^{-2}. Following the construction of the starting model, these were redefined as range restraints with upper and lower limits set to $d \pm 0.5$ Å, where d was the calculated fixed distance.

2.5. *Structure-Determination Strategy.* Structure calculations were carried out on a *Silicon Graphics O2* workstation equipped with a *MIPS R12000* processor and operating under *IRIX* release version 6.5. Canonical B-DNA models were generated in *Sybyl 6.3* using the *Tripos 5.4* force field and predefined residue libraries within the *Sybyl Biopolymer* module [14]. A model of **1** was also generated within *Sybyl 6.3*. A qual. ligand assembly model was created based on NOE data in which a head-to-tail pair of ligand molecules was manipulated as one in order to construct a crude starting model of the complex between **1** and **3**. Ligand structures were manually docked with the DNA minor groove, and incorporated distance restraints derived from the NMR data were then added. Initial steric clashes were relieved by energy minimization using a conjugate gradient algorithm with an energy-convergence criterion of 1 kcal mol^{-1} Å^{-1}. Charges were calculated using the *Pullman* method [15], and charge neutrality was maintained by explicit incorporation of Na counterions placed 6 Å away from the DNA backbone P–O bisector. Periodic boundary conditions were applied to incorporate H_2O explicitly with a box of dimensions 38 Å × 38 Å × 48 Å. The solvated complex was energy-minimized, and the subsequent model was subjected to 107 ps of restrained MD: models were gradually heated from 0 to 300 K over the first 7.0 ps, followed by equilibration at 300 K for the remainder of each run. Calculations were carried out with distance-range restraints for which the energy term was defined by a flat-well potential with quadratic boundaries according to the following criteria: $E_{range\_c} = \frac{1}{2}k^r(d - d_{LL})^2$ for $d < d_{LL}$; $E_{range\_c} = 0$ for $d_{LL} < d < d_{UL}$; $E_{range\_c} = \frac{1}{2}k^r(d_{UL} - d)^2$ for $d_{UL} < d$, where the force constant for all restraints, $k^r = 1$ kcal mol^{-1} Å^{-2}, d_{LL} was defined as the lower distance limit, and d_{UL} was defined as the upper distance limit. Snapshots from the rMD run were taken every 100 fs. Representative structures (50) were taken from the end of the rMD production run. All NOE restraints were removed, and the resulting structures were energy-minimized using steepest decent, and conjugate gradient algorithms successively until a final derivative of 0.1 kcal mol^{-1} Å^{-2} was achieved. Structures were compared with one another to produce an overall average representation which was subsequently energy-minimized.

2.6. *Structure Analysis and Representation.* The program 3DNA (version 1.5) was used to analyse the resulting structures using the base-centred reference frame recommended at a workshop held at the *AIST-NIBHT Structural Biology Centre*, Tsukuba, Japan (1999) [16]. Structure representations were generated using MOLMOL (version 2K.1, ETH-Zurich, Switzerland [17]) and rendered using *POV-Ray* for *Windows* (version 3.6, *Persistence of Vision Raytracer Pty. Ltd.*).

We express our thanks to the *MRC* for an allocation of 600-MHz instrument time and the staff of the *National Institute for Medical Research* (NIMR), Mill Hill, London, for their invaluable assistance in the use of these facilities for NMR data accumulation purposes.

REFERENCES

[1] J. B. Chaires, C. Bailley, *Bioconjugate Chem.* **1998**, *9*, 513; M. Mrksich, W. S. Wade, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7589; S. Mazzini, M. C. Bellucci, R. Mondelli, *Bioorg. Med. Chem.* **2003**, *11*, 505; Q. Zang, T. J. Dwyer, V. Tsui, D. A. Case, J. Cho, P. B. Dervan, D. E. Wemmer, *J. Am. Chem. Soc.* **2004**, *126*, 7958; A. Blaskó, T. C. Bruice, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10018; E. Gavathiotis, G. J. Sharman, M. S. Searle, *Nucleic Acids Res.* **2000**, *28*, 728.

[2] P. B. Dervan, *Bioorg. Med. Chem.* **2001**, *9*, 2215.
[3] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Delivery Rev.* **1997**, *23*, 3.
[4] N. G. Anthony, B. F. Johnston, A. I. Khalaf, S. P. MacKay, C. J. Suckling, R. D. Waigh, J. A. Parkinson, *J. Am. Chem. Soc.* **2004**, *126*, 11338.
[5] I. Alkorta, J. Elguero, G. S. Denisov, *Magn. Reson. Chem.* **2008**, *46*, 599; A. J. Dingley, L. Nisius, F. Cordier, S. Grzesiek, *Nat. Protoc.* **2008**, *3*, 242.
[6] A. J. Hampshire, H. Khairallah, A. I. Khalaf, A. H. Ebrahimabadi, R. D. Waigh, C. J. Suckling, T. Brown, K. R. Fox, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3469.
[7] A. I. Khalaf, C. J. Suckling, R. D. Waigh, Br. Pat. Appl. PCT/GB02/05916, 2002.
[8] T. L. Hwang, A. J. Shaka, *J. Magn. Reson., Ser. A* **1995**, *112*, 275.
[9] T. D. Goddard, D. G. Kneller, SPARKY 3, University of California, San Fransisco.
[10] S. S. Wijmenga, M. M. W. Mooren, C. W. Hilbers in 'NMR of Macromolecules, A Practical Approach', Ed. G. C. K. Roberts, Oxford University Press, New York, 1993, pp. 217–288.
[11] K. Wüthrich, 'NMR of Proteins and Nucleic Acids', John Wiley & Sons, New York, 1986.
[12] B. A. Borgias, T. L. James, *Methods Enzymol.* **1989**, *176*, 169.
[13] B. A. Borgias, T. L. James, *J. Magn. Reson.* **1990**, *87*, 475.
[14] M. Clark, R. D. Cramer, N. van Opdenbosch, *J. Comput. Chem.* **1989**, *10*, 982.
[15] H. Berthod, A. Pullman, *Chim. Phys.* **1965**, *62*, 942.
[16] X.-J. Lu, Z. Shakked, W. K. Olson, *J. Mol. Biol.* **2000**, *300*, 819.
[17] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graphics* **1996**, *14*, 51.

Received October 27, 2008