J|A|C|S
ARTICLES
Published on Web 08/19/2004

Short Lexitropsin that Recognizes the DNA Minor Groove at 5′-ACTAGT-3′: Understanding the Role of Isopropyl-thiazole

Nahoum G. Anthony,[†] Blair F. Johnston,[†] Abedawn I. Khalaf,[‡] Simon P. MacKay,[†] John A. Parkinson,*[,‡] Colin J. Suckling,[‡] and Roger D. Waigh[†]

Contribution from the Department of Pure and Applied Chemistry, 295 Cathedral Street, Glasgow G1 1XL, and Department of Pharmaceutical Sciences, 27 Taylor Street, Glasgow G4 0NR, University of Strathclyde, Glasgow, UK

Received December 11, 2003; E-mail: john.parkinson@strath.ac.uk

Abstract: Isopropyl-thiazole ($^{i\text{Pr}}$Th) represents a new addition to the building blocks of nucleic acid minor groove-binding molecules. The DNA decamer duplex d(CGACTAGTCG)$_2$ is bound by a short lexitropsin of sequence formyl-PyPy$^{i\text{Pr}}$Th-Dp (where Py represents *N*-methyl pyrrole, $^{i\text{Pr}}$Th represents thiazole with an isopropyl group attached, and Dp represents dimethylaminopropyl). NMR data indicate ligand binding in the minor groove of DNA to the sequence 5′-ACT^{5}AG^{7}T-3′ at a 2:1 ratio of ligand to DNA duplex. Ligand binding, assisted by the enhanced hydrophobicity of the $^{i\text{Pr}}$Th group, occurs in a head-to-tail fashion, the formyl headgroups being located toward the 5′-ends of the DNA sequence. Sequence reading is augmented through hydrogen bond formation between the exocyclic amine protons of G^7 and the $^{i\text{Pr}}$Th nitrogen, which lies on the minor groove floor. The B_I/B_{II} DNA backbone equilibrium is altered at the T^5 3′-phosphate position to accommodate a B_{II} configuration. The ligands bind in a staggered mode with respect to one another creating a six base pair DNA reading frame. The introduction of a new DNA sequence-reading element into the recognition jigsaw, combined with an extended reading frame for a small lexitropsin with enhanced hydrophobicity, holds great promise in the development of new, potentially commercially viable drug lead candidates for gene targeting.

I. Introduction

The observation by Wemmer's group that distamycin could bind side-by-side in the minor groove of DNA with selectivity for A/T sequences[1] led to pioneering research in the laboratories of Dervan and Lown that resulted in the creation of highly sequence-selective DNA minor groove-binding molecules. The potential for such an approach in the treatment of diseases such as cancer hardly needs elaboration: specific, artificial interference with gene expression through the mediation of DNA site-specific transcriptional activation and repression[2,3] has far-reaching implications for the future of medicine.

Research into designing "side-by-side" minor groove binders has led to the development of a "code" for base recognition. Thus, antiparallel, side-by-side pairing of Imidazole (Im)/Pyrrole (Py) combinations distinguish G·C from C·G and both of these from A·T and T·A base pairs. Py/Py pairs bind A·T and T·A in preference to G·C and C·G, and discrimination of T·A from A·T is brought about by hydroxypyrrole(Hp)/Py and Py/Hp pairs.[4]

Specific DNA sequence recognition with the required nanomolar binding constant comes, however, at a price: that of molecular size. To date, molecules capable of reading a 10 base pair or longer DNA sequence have incorporated hairpin structure frameworks in the molecular weight range 900−1900 Da.[4,5] Although these have allowed convincing demonstrations of an ability to interrupt biochemical processes in vitro,[3,6] the "rule of five"[7] is certainly violated in at least one respect: poor drug absorption and permeation is more likely for molecules whose molecular weight is greater than 500. Targeting large binding sites is therefore limited by ligand size on cell permeability.[8] Although potential in vivo activity has been spoken of in respect of these types of molecules,[9] smaller molecules still have more likelihood of becoming commercial therapeutic drugs.

Most of the published work to date has reflected the assumption that selectivity and affinity of binding depends on a ligand's ability to hydrogen bond to the floor of the minor groove where the bases offer a sequence-dependent pattern of hydrogen bond donors and acceptors. Although this assumption

* Corresponding author: Phone: +44-141-548-2820. Fax: +44-141-548-4822.
[†] Department of Pharmaceutical Sciences.
[‡] Department of Pure and Applied Chemistry.

(1) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723−5727.
(2) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559−561.
(3) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202−205.
(4) Devan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215−2235.
(5) For reviews see: (a) Wemmer, D. E. *Biopolymers (Nucleic Acid Sci.)* **2001**, *52*, 197−211. (b) Neidle, S. *Nat. Prod. Rep.* **2001**, *18*, 291−309.
(6) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12890−12895.
(7) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Res.* **1997**, *23*, 3−25.
(8) Urbach, A. R.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4343−4348.
(9) (a) Lenzmeier, B. A.; Baird, E. E.; Dervan, P. B.; Nyborg J. K. *J. Mol. Biol.* **1999**, *291*, 731−744. (b) Bremer, R. E.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *Bioorg. Med. Chem.* **2000**, *8*, 1947−1955.

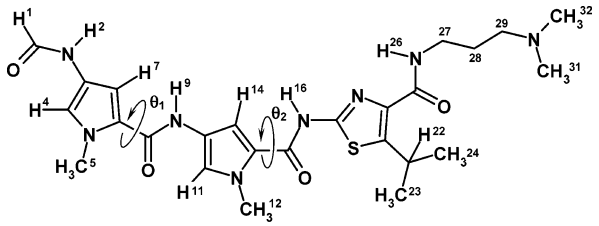
10.1021/ja030658n CCC: $27.50

has generated useful ideas and results, recent data suggest that the primary driving force is hydrophobicity, the main energy benefit of binding coming from the loss of structured water in the DNA minor groove and around the ligand.[10] Computer modeling of DNA with typical ligands such as distamycin shows that there is a shape mismatch: distamycin, although capable of adopting the required helical shape to fit the minor groove, is largely planar, whereas the minor groove has a very uneven surface. The surfaces of the minor groove reflect the nonplanar topology of the atoms that constitute the DNA backbone, and binding of ligands to this region appears to occur through an induced-fit mechanism.[11]

Our approach to the design of distamycin analogues has incorporated a strategy to raise the oil/water partition coefficient to aid penetration through lipid membranes while concurrently adapting the shape of the ligand to follow and mold to the contours of the minor groove. This partially empirical approach is required, because the minor groove is flexible and will adapt to fit a variety of ligand shapes, and is shown by the way that DNA adapts to either accommodate single ligands or side-by-side pairs of ligands in the minor groove. The mode of binding depends on the DNA base pair sequence but, as our current work is beginning to show, can also subtly and remarkably depend on what might otherwise be regarded as insignificant changes to the ligand. Alternative pairing schemes of new ligand monomers may expand the repertoire of targetable sequences using *N*-methyl pyrrole- and imidazole-based polyamides. The design of single binding, high-affinity ligands will open the way to the base pair code being read with smaller molecules.

In our work, we have chosen to introduce branched chain alkyl groups into our minor groove binding ligands. These include distamycin analogues containing an isopropyl group attached to the central pyrrole ring. Such compounds have shown higher affinity than the parent distamycin and with an altered pattern of binding selectivity.[12] These positive results have led to the synthesis of analogous compounds with branched alkyl side chains located at different positions. The use of thiazole instead of pyrrole was particularly attractive because the sulfur is large and will also have a major effect on partitioning into biological membranes. Lexitropsin **1**, used in these studies, has both a thiazole and an isopropyl side-chain, which renders the subunit nearest to the positive tail very bulky and hydrophobic. DNA footprinting with **1** using a 200 base pair DNA construct revealed only one binding site centered around the sequence 5′-ACTAGT-3′[12] and which had very high affinity. Comparative binding studies using capillary electrophoresis with the oligonucleotide 5′-CGACTAGTCG-3′ (**2**) confirmed the affinity, and competition studies confirmed the selectivity of binding. These studies also revealed the possibility that the ligand was binding sequence specifically to DNA in a higher ratio than had previously been observed. Molecular modeling studies showed a way in which such binding could be accommodated.

In an attempt to resolve this issue (important in terms of our overall aim of developing small molecules capable of reading



1

$C^1\ G^2\ A^3\ C^4\ T^5\ A^6\ G^7\ T^8\ C^9\ G^{10}$
$G^{20}C^{19}T^{18}G^{17}A^{16}T^{15}C^{14}A^{13}G^{12}C^{11}$

2

longer DNA sequences) and to understand in detail the molecular interactions that play a part in stabilizing the ligand−DNA complex, we have examined in detail, by solution-phase NMR spectroscopy, the structure and behavior of the complex formed between **1** (which we call thiazotropsin A) and the self-complementary DNA duplex **2**.

II. Experimental Section

A. Materials. *N*-[3-(Dimethylamino)propyl]-2-({[4-({[4-(formylamino)-1-methyl-1*H*-pyrrol-2-yl]carbonyl}amino)-1-methyl-1*H*-pyrrol-2-yl]carbonyl}amino)-5-isopropyl-1,3-thiazole-4-carboxamide (thiazotropsin A, **1**) was prepared as the TFA salt as described previously.[13] d(CGACTAGTCG)$_2$, **2**, was supplied by Cruachem, Ltd. (Glasgow, Scotland), as a desalted, lyophilized powder and was used without further purification. For NMR purposes, **2** was dissolved in 300 μL of a mixture of 90% H_2O/10% D_2O to a final concentration of 2.36 mM in DNA duplex prior to dilution effects caused through the addition of ligand solution. The sample was admitted to a water susceptibility-matched NMR tube (NMR tube type: BMS-005TB; Shigemi, Inc., Allison Park, PA).

B. Methods. 1. Complex Formation between Thiazotropsin A (1) and 2. Thiazotropsin A, **1** (2.3 mg), was dissolved in 50 μL of D_2O to provide a stock solution of ligand at a concentration of 152 mM. DNA **2** was dissolved in 300 μL of 90% H_2O/10% D_2O and added to a 25 μL aliquot of **1** (for handling purposes). This mixture (325 μL) was added to a further 10 μL of ligand stock solution. Mixing with a further 5 μL of ligand stock resulted in a final working solution of 340 μL, being 4.12 mM in **2** with 2 equiv of **1**. One-dimensional ^{1}H NMR spectroscopy confirmed the status of the sample at each stage of mixing. Ultimately, the complete disappearance of free DNA imino proton ^{1}H NMR resonances was noted with their replacement by a new set of signals. In this state the sample was used in the accumulation of complete NMR data sets for the purposes of structure determination.

2. NMR Spectroscopy. NMR data were acquired on Varian ^{UNITY}*INOVA* 800, Bruker Avance 600, Varian ^{UNITY}*INOVA* 600, and Bruker AMX 400 NMR spectrometers operating at 800.289, 599.89, 600.13, and 400.13 MHz for proton resonance, respectively. Standard geometry triple-resonance probe heads equipped for *z*-pulsed field gradients were used on each high-field NMR spectrometer. A 5 mm Bruker dual probehead (BBO, [X/^{1}H]) was used for all data acquisitions using the midfield AMX 400 instrument. Data acquisition was carried out in an identical manner for both **2** and for the complex between **1** and **2** unless otherwise stated. The probe temperature was 298 K in all instances except for work carried out at 400.13 MHz, where the probe temperature was 300 K.

One-dimensional ^{1}H NMR data were acquired using either presaturation or a double-pulsed field gradient spin echo (dpfgse)[14] approach

(10) Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chaires, J. B. *J. Mol. Biol.* **1997**, *271*, 244−257.
(11) Bostock-Smith, C. E.; Harris, S. A.; Laughton, C. A.; Searle, M. S. *Nucleic Acids Res.* **2001**, *29*, 693−702.
(12) Fox, K. R.; Khalaf, A. I.; MacKay, S. P.; McGroarty, I. S.; Parkinson, J. A.; Skellern, G. G.; Suckling, C. J.; Waigh, R. D. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1353−1356.
(13) Khalaf, A. I.; Suckling C. J.; Waigh, R. D. British Patent Application, PCT/GB02/05916, 2002.
(14) Hwang, T. L.; Shaka, A. J. *J. Magn. Reson. Ser. A* **1995**, *112*, 275−279.

to eliminate the solvent resonance. At 600 MHz, data were typically acquired using digital quadrature detection (DQD) with 128 transients over a frequency width of 15 kHz (25 ppm) centered 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 a ratio of 31:11.

At 600 MHz, two-dimensional (2D) NMR data sets were acquired as follows: 2D [^{1}H, ^{1}H] DQFCOSY NMR data (pulse program: cosydfphpr) were acquired with 16 transients for each of the 512 States-TPPI 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 for a total data accumulation time of 4.5 h. 2D [^{1}H,^{1}H] ROESY NMR data (pulse program: roesygpph19.2) were acquired with 16 transients for each of the 512 States-TPPI 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 1.4 s and a pulsed spin-lock time of 80 ms for a total accumulation time of 5 h. 2D [^{1}H, ^{1}H] TOCSY NMR data were acquired with 16 transients for each of the 512 States-TPPI t_1 increments over an ω_2 frequency width of 12 kHz (20.04 ppm) and an ω_1 frequency width of 5.48 kHz (9.14 ppm) into 8 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 for a total data accumulation time of 4.5 h. 2D [^{1}H, ^{1}H] NOESY NMR data for quantitation were acquired with eight transients for each of the 1024 States-TPPI t_1 increments over a frequency width of 12 kHz in both ω_2 and ω_1 (20.04 ppm) into 8 K complex data points (acquisition time 341 ms) with a recycle time of 7 s and a mixing time of 100 ms for a total accumulation time of 14 h.

One-dimensional ^{31}P−{^{1}H} NMR data were acquired at 161.977 MHz using 64 transients over a frequency width of 810 Hz (5 ppm) into 512 data points (acquisition time 316 ms) with a recycle time of 0.5 s and centered close to the center of the DNA phosphate resonance envelop. GARP composite pulse decoupling was used for ^{1}H decoupling during the acquisition time only. 2D [^{31}P, ^{1}H] correlations were acquired using an inverse INEPT 2D sequence with TPPI (pulse program hxineptp). Transients (80) were acquired for 80 t_1 increments into 128 data points over a ω_2 (^{31}P) frequency width of 486.4 Hz (3 ppm) and an ω_1 (^{1}H) frequency width of 1200 Hz. The evolution period 1/4J was varied for different experiments in order to select for differently sized couplings. J values were chosen in order to compromise between low optimization of couplings and loss of signal during the evolution period through T_2 relaxation. ^{31}P NMR data were referenced indirectly according to Maurer and Kalbitzer.[15]

All NMR data were processed on a Dell Precision 340 workstation running under Microsoft Windows 2000 using Xwin-nmr (version 3.0, Bruker Biospin, Karlsruhe, Germany) with appropriate processing parameters and imported into SPARKY (version 3.105)[16] for data analysis and reduction.

3. NMR Data Assignment Strategy. 2D [^{1}H, ^{1}H] NOESY NMR data were assigned for both free duplex **2** and for the complex between **1** and **2** by using established assignment strategies[17,18] for right-handed B-form DNA. 2D [^{1}H, ^{1}H] DQFCOSY NMR data were used to establish resonance assignments for specific protons of **1** in its complex with **2**. 2D [^{1}H, ^{1}H] ROESY NMR data were used to overcome ambiguities in the assignment of DNA H2′ and H2″ proton resonances.[19] ^{31}P resonance assignments were achieved on the basis of observed correlations between ^{31}P and H3′, H4′, and H5′/H5″ resonances.

4. Modeling Restraints. Quantitative interproton distance restraints were calculated from 100 ms NOESY data sets with an iterative relaxation matrix approach using the program MARDIGRAS.[20,21] 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 **2**, 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.

5. Structure Determination Strategy. Canonical B-DNA models were generated in Sybyl 6.3 using the Tripos 5.4 force field.[22] DNA atom charges were calculated using the Pullman method.[23] Backbone charges were reduced to incorporate counterion effects implicitly. A model of **1** was also generated in Sybyl 6.3. A qualitative model was created on the basis of 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 **2**. For both free DNA and for the complex, distance restraints were incorporated into the models. Periodic boundary conditions were applied to incorporate water explicitly employing a box with dimensions 38 × 38 × 48 Å containing 1904 water molecules. Periodic boundary condition simulations employed a RATTLE restraint on high-frequency bonds to enable a time step of 2 fs to be used. The initial structures were energy minimized using 500 steps of Powell minimization. Subsequent models were subjected to 100 ps rMD: models were heated from 10 to 300 K over the first 3.0 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: $E_{range\_c} = {}^1/_2k^r(d - d_{LL})^2$ for $d < d_{LL}$; $E_{range\_c} = 0$ for $d_{LL} < d < d_{UL}$; $E_{range\_c} = {}^1/_2k^r(d_{UL} - d)^2$ for $d_{UL} < d$, where the force constant for all restraints (k^r) was 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. Minimized low-energy structures from each run were averaged and compared with snapshots throughout the rMD trajectory. 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 of the structure. Similar comparative calculations were also made using the InsightII and Discover software packages (Accelrys. Inc., San Diego, CA) with the CFF91 force field[24] applied during simulations, which were performed using a 600 MHz R14000 dual processor Silicon Graphics Octane 2 workstation.

6. Structure Analysis. The program CURVES (version 5.2[25]) was applied to analyze the resulting structures by using the helicoidal parameters used to describe a nucleic acid duplex, as defined according to the EMBO workshop on DNA curvature and bending.[26]

III. Results

A. Titration of DNA with Ligand. The formation of the complex between **1** and **2** was carried out according to the experimental details described. One-dimensional (1D) ^{1}H NMR

(15) Maurer, T.; Kalbitzer, H. R. *J. Magn. Reson. Ser. B* **1996**, *113*, 177−178.
(16) Goddard, T. D.; Kneller, D. G. *SPARKY 3*; Univeristy of California, San Francisco, CA.
(17) Wijmenga, S. S.; Mooren, M. M. W.; Hilbers, C. W. In *NMR of Macromolecules, A Practical Approach*; Roberts, G. C. K., Ed.; Oxford University Press: New York, 1993; pp 217−288.
(18) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
(19) Tisné, C.; Hantz, E.; Hartmann, B.; Délèpierre, M. *J. Mol. Biol.* **1998**, *279*, 127−142.
(20) Borgias, B. A.; James, T. L. *Methods Enzymol.* **1989**, *176*, 169−183.
(21) Borgias, B. A.; James, T. L. *J. Magn. Reson.* **1990**, *87*, 475−487.
(22) Clark, M.; Cramer, R. D.; van Opdenbosch, N. *J. Comput. Chem.* **1989**, *10*, 982−1012.
(23) Berthod, H.; Pullman, A. *J. Chim. Phys.* **1965**, *62*, 942−946.
(24) (a) *Discover 2.9.5/94.0 User Guide*; Biosym Technologies: San Diego, CA, 1994. (b) Maple, J. R.; Thacher, T. S.; Dinur, U.; Hagler, A. T. *Chem. Des. Automation News* **1990**, *5*, 5−10.
(25) Lavery, R.; Sklenar, H. *Curves 5.2: Helical Analysis of Irregular Nucleic Acids*; Laboratorie de Biochimie Theorique, Institut de Biologie Physico-Chimique, CNRS, Paris, France, 1997.
(26) Dickerson, R. E. *J. Biomol. Struct. Dyn.* **1989**, *6*, 627−634.

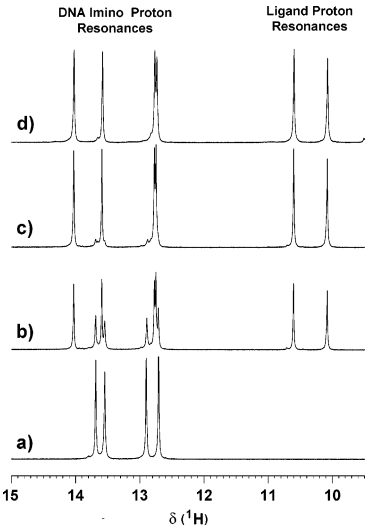


Figure 1. Imino proton resonance regions of ^{1}H NMR spectra acquired at 600 MHz using a dpfgse routine and showing the result of titrating a solution of **1** into a sample of **2**. The ^{1}H NMR resonances of the imino protons belonging to Watson−Crick base pairs were visible between 12.5 and 14.5 ppm. Resonances between 11.0 and 12.0 ppm were assigned to peptide N*H* protons in **1**. (a) Free DNA. (b) After addition of ca. 1.4 equiv of **1**. (c) After addition of slightly less than 2 equiv of **1**. (d) Exact 2:1 equivalence between ligand, **1**, and DNA, **2**. Ligand residency time was relatively long as shown by the presence of free and bound forms of DNA at a ligand/DNA duplex ratio of ∼1:1. Signal integration indicated that two molecules of **1** occupied the DNA minor groove (see Figure S4 in Supporting Information).

spectra were acquired on the sample at various ligand/DNA ratios. The results of the "titration" (Figure 1) focus on the region of the spectrum at high chemical shift (additional regions of the 1D ^{1}H NMR data are presented in Figures S1 and S2 in Supporting Information). The chemical shifts of four imino proton resonances for free DNA (expected for a symmetrical 10 base pair sequence with end base pair "fraying", Figure 1a) remained unaltered upon the addition of an aliquot of ligand. However, this region of the ^{1}H NMR spectrum increased in complexity with ligand addition (Figure 1b). Imino proton ^{1}H NMR signal assignments were initially readily made on the basis of saturation transfer effects (Table 1 and Figure S3 in Supporting Information), which were observed following a presaturation approach to solvent suppression and provided a "first impression" of signal assignments. At a later stage, detailed and rigorous NOE assignments confirmed these initial, speculative assignments. The ligand/DNA complex was associated with the slow chemical exchange regime NMR time scale: in the presence of free DNA, the ^{1}H NMR spectrum simultaneously showed the presence of imino proton resonances from both free and ligand-bound DNA. Further addition of ligand to the DNA sample simplified the NMR spectrum, which, in the imino proton resonance region, showed the presence of four imino proton resonances and two additional resonances from the ligand. The latter were anticipated to arise from N*H* protons associated with the ligands' peptide linkages (Figure 1d). Integration of ^{1}H NMR data (acquired using a dpfgse pulse sequence to suppress the solvent) showed that two molecules of **1** bind to **2** under the conditions used (Figure S4 in Supporting Information). The appearance of only four imino proton resonances under these conditions indicated the retention of symmetry in the DNA complex. From this evidence alone it was concluded that two ligands were binding to DNA in an antiparallel mode.

B. NMR Data Assignment. Complete solution structure analysis required ^{1}H and ^{31}P NMR data assignments to be completed on the basis of 2D [^{1}H, ^{1}H] DQFCOSY, TOCSY, ROESY, NOESY, and H−P correlation NMR data. This was carried out as far as possible for both free DNA, **2**, and for the complex between **1** and **2**. The signal assignments for **2** (Table 1) were made according to standard procedures for right-handed B-form DNA and were without unusual features. The fingerprint region of the 600 MHz, 100 ms, 2D [^{1}H, ^{1}H] NOESY NMR spectrum of **2** is shown with assignments in Figure S5 in Supporting Information. To complete the assignment of NOE cross-peaks in the NMR data from the complex between **1** and **2**, intraligand, interligand, intrastrand, interstrand, and ligand−DNA NOEs were distinguished from one another. This process is nontrivial, and initially various DQFCOSY correlations were used in order to distinguish key ligand resonances. Narrow DQFCOSY cross-peaks were observed between protons resonating at the following chemical shifts (Figure S6): 6.148 and 6.256 ppm (cross-peak a); 7.622 and 6.332 ppm (cross-peak b); 8.622 and 7.73 ppm (cross-peak c). These are unusual in a DNA setting and were consequently associated with **1**. Reference to the spectrum of **1** alone in DMSO-d_6 indicated the presence of narrow doublets at 8.13, 7.40 and 6.94 ppm (data not shown). From these data the signal at 8.13 ppm was assigned to the formyl proton (H1 in **1**). For the ligand/DNA complex, the DQFCOSY cross-peak at 8.622/7.73 ppm (c) was mirrored by a strong NOE in the NOESY NMR spectrum (c′). On the basis of the peak shape and line width compared with its partnering resonance at 7.73 ppm, the signal at 8.622 ppm was assigned to an N*H* signal (specifically H2 in **1**). Hence, the signal at 7.73 ppm was assigned to H1 in the complex. The shape and pattern of the remaining resonances at 6.148, 6.256, 6.332, and 7.622 ppm enabled these signals to be tentatively assigned to pyrrole "H2" and "H4" protons in **1**. Specific assignments of these resonances relied upon inspection of the relative sizes of NOEs associated with these resonances, together with NOE assignments of methyl resonances from the ligands. Specifically, a large NOE (Figure S6) was observed between the signals at 6.148 and 6.332 ppm (cross-peak d′), resonances that belonged to different spin systems. A large NOE (e′) was also observed between protons giving rise to the signals at 6.148 and 8.622 ppm, the latter having been provisionally assigned to H2 of **1**. The signals at 10.53 and 10.01 ppm were associated with the peptide N*H* protons: H9 and H16 of **1** (Figure 2). The signal at 10.01 ppm shared a very strong NOE with the signal at 6.148 ppm. These NOEs were explained by assigning the signal at 6.148 ppm to H7 of **1**. The NOE at 6.148/8.622 ppm was therefore due to close contact between H7 and H2 of **1**. The signal at 10.01 ppm was assigned to H9 of **1**, and thus the large NOE at 6.148/10.01 ppm was explained as the through-space contact between H9 and H7 of **1**. This initial stage of the assignment process proved to be crucial to the determination of the resonance assignments for the bound ligand, **1**. By a similar token, inspection of NOE intensities enabled the remaining pyrrole and peptide resonances for **1** bound to DNA to be assigned. Coupling observed between the signal for H26 (7.906 ppm) and ^{1}H NMR resonances at 3.927 and 3.129 ppm enabled H26 of **1** to be assigned. The latter two resonances were

Table 1. ^{1}H Chemical Shift Assignments for the DNA Duplex d(CGACTAGTCG)$_2$

	chemical shift assignment: δ ^{1}H (ppm)														
base	H8	H6	H2	H5	CH_3	H1′	H2′	H2″	H3′	H4′	H5′	H5″	GH1/ TH3	H41	H42
C1		7.511		5.788		5.61	1.75	2.26	4.58	3.96	3.62	3.62			
G2	7.862					5.40	2.62	2.70	4.91	4.23	3.99	3.87	12.80		
A3	8.131		7.792			6.17	2.65	2.81	4.96	4.40	4.15	4.09			
C4		7.155		5.100		5.68	1.83	2.35	4.58	4.10	3.97	3.97		7.88	6.53
T5		7.235			1.490	5.55	2.00	2.36	4.77	4.02	3.99	3.97	13.50		
A6	8.130		7.081			5.94	2.66	2.78	4.95	4.33	3.98	4.07			
G7	7.511					5.76	2.36	2.60	4.77	4.31	4.12	4.17	12.60		
T8		7.144			1.160	5.92	1.95	2.35	4.75	4.12	a	4.11	13.60		
C9		7.385		5.574		5.63	1.92	2.27	4.74	4.01	3.99	3.99		8.48	6.90
G10	7.839					6.06	2.52	2.28	4.59	4.08	3.99	3.99			

[a] Resonance unassigned.

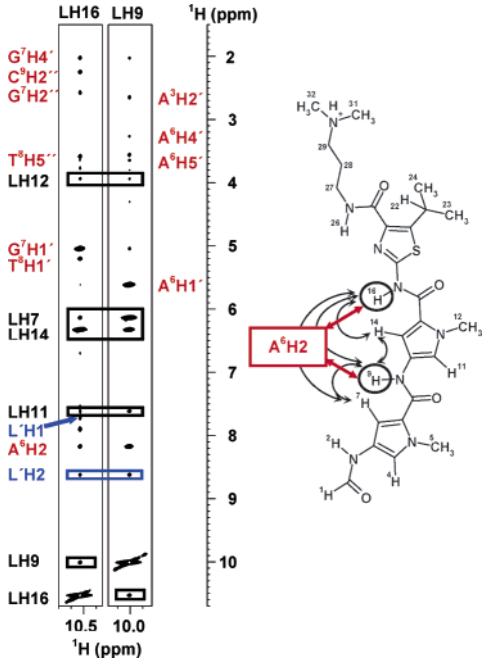


Figure 2. Strip plots of data taken from the 100 ms 2D [^{1}H, ^{1}H] NOESY NMR spectrum acquired on the 2:1 complex between **1** and **2** at 600 MHz. Data are shown at the ligand resonance chemical shifts of H16 and H9. Data labeling scheme: DNA resonance assignments, red labels; ligand resonance assignments, black labels; interligand NOEs, blue labels. A schematic indication is shown of how some of the NOEs relate to the structure of the complex. The relationship between representative pairs of protons is shown schematically, adjacent to the NMR data.

subsequently assigned to H27′ and H27″, respectively, on the basis of NOE measurements. The remaining resonances from the DMAP chain were assigned by recourse to DQFCOSY and TOCSY NMR data.

Methyl group ^{1}H NMR resonances from **1** were visible in three regions of the ^{1}H NMR spectrum, namely, 3.6−4.0, 2.86−2.89, and 1.33−1.38 ppm (Figure S2 in Supporting Information). The isopropyl methyl resonances were assigned from the DQFCOSY NMR spectrum, in which a strong correlation occurred between the methyl resonances at 1.337/1.383 ppm and a multiplet at 3.936 ppm (H22 of **1**). The pyrrole methyl resonances were assigned through strong NOEs to their associated ring protons. The remaining DMAP methyl ^{1}H NMR resonances were assigned through correlation to H29′ and H29″ via NOEs. It was therefore possible to make a complete assignment of the ^{1}H NMR resonances of **1** bound to DNA, full details of which are noted in Table 2. In the process of making the ligand assignments, ^{1}H NMR resonance assignments were made for the oligonucleotide to which **1** was bound (Table 3 and Figure S7 in Supporting Information). Large chemical shift changes were anticipated for resonances of H1′, H4′, H5′, and H5″ on the assumption that **1** would bind in the minor groove. These protons line the "walls" of the DNA minor groove of B-form DNA. The presence of pyrrole rings in the locality of H1′, H4′, H5′, and H5″ protons results in strong local magnetic shielding effects, which can provide useful secondary structural restraints.[27] By plotting the difference ($\Delta\delta$) between the chemical shift of protons in the free DNA and the chemical shifts of the same protons in the ligand-bound DNA, it was possible to show the location of the ligand against the DNA sequence. A plot of $\Delta\delta$ for the H4′ resonances of **2** (from both free and ligand bound forms, Figure 3) indicated that the *N*-methyl pyrrole and isopropylthiazole rings of the ligand were lying in the minor groove juxtaposed against the sugar rings of the sequence 5′-$T^5A^6G^7T^8$-3′. The overall footprint for the side-by-side binding of two ligands is also indicated (a full catalog of all ^{1}H NMR chemical shift changes is provided in Table S1 in Supporting Information). Although the H4′ chemical shifts showed the biggest change when the ligand was bound, H1′ protons also showed significant shift changes (Figure S8 in Supporting Information). It was also noted that the H5″ ^{1}H chemical shift for A^6 experienced an unusually large change (−1.17 ppm) when the ligand was bound. This was believed to be related not only to the influence of nearby ligand (specifically via van der Waals contact of the plane of one of the ligand pyrrole rings) but also to changes observed in the DNA backbone conformation at the T^5pA^6 step. This was suspected from changes in the ^{31}P NMR data of the DNA when **1** was bound (Figure 4). Resonance assignments based on 2D [^{31}P, ^{1}H] COSY NMR data, summarized for both free and ligand-bound DNA in Table 4, enabled the broad ^{31}P NMR resonance at +0.18 ppm in the complex to be assigned to the T^5 3′-phosphate, compared with −1.37 ppm for the same phosphate in free **2**. Broadening of this resonance indicated possible conformational averaging in the complex and the shift change implied an alteration in the geometry of the DNA backbone at this position in the sequence.

C. Structure Determination. Full assignment of proton−proton NOEs for **2** and for the 2:1 complex between **1** and **2** enabled comprehensive structure calculations to be carried out in both cases. NOE assignments are presented for the complex in Table S2 in Supporting Information. Statistical details regarding the structural restraints used for both sets of calcula-

(27) Parkinson, J. A.; Barber, J.; Douglas, K. T.; Rosamund, J.; Sharples, D. *Biochemistry* **1990**, *29*, 10181−10190.

Table 2. ^{1}H NMR Chemical Shift Assignments of Thiazotropsin A (**1**) in Complex with d(CGACTAGTCG)$_2$

chemical shift assignment: δ ^{1}H (ppm)										
H(1)	H(2)	H(4)	CH_3(5)	H(7)	H(9)	H(11)	CH_3(12)	H(14)	H(16)	H(22)
7.728	8.625	6.258	3.653	6.141	10.010	7.617	3.948	6.326	10.530	3.936
chemical shift assignment: δ ^{1}H (ppm)										
CH_3(23)	CH_3(24)	H(26)	H(27′)	H(27″)	H(28′)	H(28″)	H(29′)	H(29″)	CH_3(31)	CH_3(32)
1.337	1.383	7.906	3.257	3.129	2.108	1.771	2.948	3.187	2.867	2.891

Table 3. ^{1}H Chemical Shift Assignments for DNA d(CGACTAGTCG)$_2$ in the Presence of 2 Equiv of Thiazotropsin A (**1**) Per Duplex

	chemical shift assignment: δ ^{1}H (ppm)															
base	H8	H6	H2	H5	CH_3	H1′	H2′	H2″	H3′	H4′	H5′	H5″	GH1/TH3	H41	H42	H21	H22
C1		7.543		5.774		5.583	1.824	2.299	4.605	3.940	a	a					
G2	7.893					5.664	2.666	2.832	4.944	4.274	4.012	3.892	12.67				
A3	8.159		7.655			6.070	2.663	2.624	4.965	4.365	4.144	4.144					
C4		7.381		5.083		5.960	1.855	2.347	4.885	4.148	4.075	4.075		8.552	5.979		
T5		7.386			1.544	4.310	1.966	2.026	4.508	3.093	3.81	3.552	13.96				
A6	8.176		8.003			5.620	2.459	2.659	a	3.277	3.571	2.903					
G7	7.549					5.051	1.994	2.585	4.600	2.039	3.695	3.654	12.70			7.187	6.453
T8		6.705			1.449	5.211	1.516	1.825	4.384	2.259	3.782	3.591	13.51				
C9		7.196		5.505		5.680	1.896	2.264	4.739	3.996	3.881	3.399		8.559	6.785		
G10	7.768					5.976	2.483	2.226	4.564	4.042	a	a					

[a] Resonance unassigned.

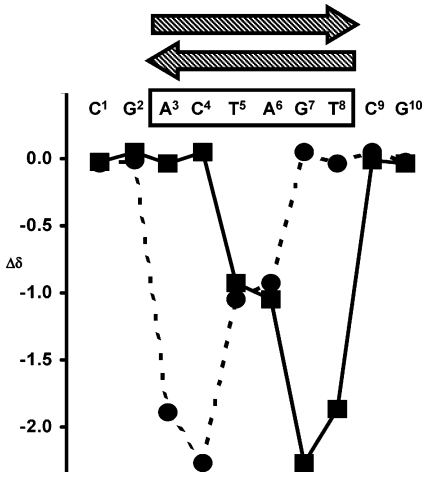


Figure 3. Chemical shift differences for NMR resonances of ligand-bound and ligand-free DNA duplex. Chemical shift differences ($\Delta\delta$) were measured for H4′ resonances as δ H4′ DNA [free − bound]. The location of the ligand relative to the DNA sequence is indicated by shaded arrows. The dashed line shows the shift changes for the opposing DNA strand, thereby indicating the overall footprint for side-by-side ligand DNA recognition.

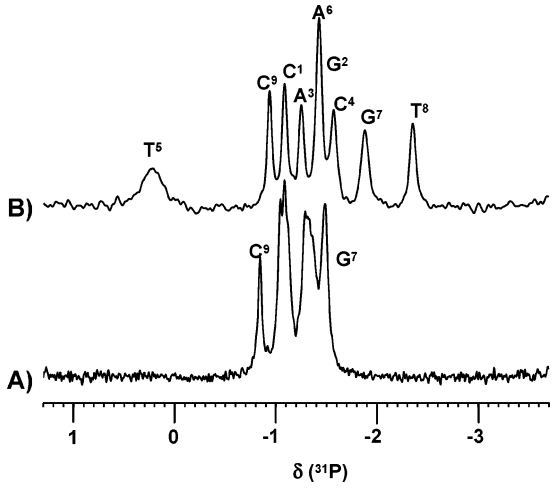


Figure 4. ^{31}P−{^{1}H} NMR spectra of **2** (A) and of the complex between **2** and **1** (B) acquired at 9.4 T. The effect of binding **1** to the DNA duplex was clear from the dispersion of signals that occurred for the complex, indicative of DNA backbone alteration.

tions are provided (Table 5). The restraints were carefully checked for consistency. For the complex, in particular, the initial set of 996 restraints were carefully analyzed against a crude model of the complex. Visualization of interproton distance restraints using this model made it possible to correct misassignments. Of those restraints originally incorporated, 14 were removed after inspection due to implausibility.

D. Solution Structure of Free DNA Duplex, 2. To directly assess the effect of ligand binding on the structure of the DNA duplex, the solution structure of **2** was determined. The NMR data were consistent with a right-handed B-form DNA duplex, in which the Watson−Crick base pairing was intact (as indicated by the presence of imino proton ^{1}H NMR resonances in the NMR data). The starting model for the solution structure calculations was therefore constructed as a canonical B-form DNA duplex structure. An average of 25 NOEs per residue were incorporated into the starting structure as distance restraints with a force constant of 1 kcal mol^{-1} Å^{-2}. Representations of the resulting calculated structure are shown in Figure S10 in Supporting Information. The family of representative structures show mean global backbone rmsd = 0.6 ± 0.1 Å and mean global heavy atom rmsd = 0.6 ± 0.1 Å. The overall nature of the DNA did not depart markedly from that of canonical B-form DNA. The average minor groove width decreased from 5.9 Å in canonical B-DNA to 3.6 Å in the calculated structure, which may reflect the higher A/T base pair content of the central recognition footprint region of the free DNA (66%). A/T-bearing regions of DNA generally encompass narrower minor grooves with more negative electrostatic potentials compared with C/G-bearing regions.[28] The structure displayed an overall axis

(28) Pullman, B. *Adv. Drug. Res.* **1989**, *18*, 1−113.

Table 4. Comparison of ^{31}P Chemical Shift Assignments for d(CGACTAGTCG)$_2$ in the Absence (Free) and Presence (Bound) of 2 Equiv of Ligand Per Duplex

	chemical shift assignment:[a] δ ^{31}P (ppm)		$\Delta\delta$ ^{31}P($\delta_{bound} - \delta_{free}$)
base	bound	free	(ppm)
C1	−1.08	−1.04	−0.04
G2	−1.43	−1.09	−0.34
A3	−1.29	−1.29	0.00
C4	−1.57	−1.04	−0.53
T5	+0.18	−1.37	1.55
A6	−1.42	−1.13	−0.29
G7	−1.88	−1.49	−0.39
T8	−2.36	−1.33	−1.03
C9	−0.93	−0.84	−0.09
G10	-	−1.48	

[a] Assignments are for the phosphates that are 3′ with respect to the base.

curvature of 20°. *x*- and *y*-displacement of base pairs was negligible (average dx = −0.8 Å, average dy = 0.8 Å). Some variation in the inclination (η) and tip (θ) of base pairs was noted (average $\eta = -9°$; average $\theta = -5°$). Average values of base pair shear (Sx), stretch (Sy), stagger (Sz), propeller twist (ω), and opening (σ) were similar to those of canonical B-DNA, although buckle (κ) in the complex did vary along the sequence with an average value of 5°. Average global interbase pair parameters (shift, Dx, slide, Dy, rise, Dz, and tilt, τ) were similar to those of canonical B-DNA, but roll (ρ) had an average value of 8° with a standard deviation of 14° compared with −0.02° with a standard deviation of 0.1° for canonical B-DNA. The largest contribution to the global axis curvature came from the T^{5}pA^6 step. Sugar puckers fell into either C1′-exo or C2′-endo configurations. The pseudorotation phase angle in the central eight-base pair region averaged at 136°, close to the 140−180° range anticipated for the sugar "*S*" conformation of B-DNA. We did not carry out a detailed analysis of ^{31}P chemical shifts or measure ^{1}H−^{31}P coupling constants to define backbone torsion angles. In general, it is recognized that rapid equilibrium exists between B_I ($\epsilon = 180°$ and $\zeta = -60°$) and B_{II} ($\epsilon = -60°$ and $\zeta = 180°$) states. We chose not to investigate the backbone structure in detail, but it was clear that all of the ^{31}P resonances for **2** were clustered together. No backbone restraints were applied to the structure calculation, but in the final analysis it was noted that ϵ (defined by C4′−C3′−O3′−P) and ζ (defined by C3′−O3′−P−O5′) backbone torsion angles (which are noted to have large influences on ^{31}P chemical shift and H3′−^{31}P coupling constants) were generally found to be consistent with expected values.

E. Solution Structure of the 2:1 Ligand/DNA Complex. As noted, 1D ^{1}H NMR data (Figure 1 and Figure S3 in Supporting Information) showed that complex formation between **1** and **2** was symmetrical. For an inherently unsymmetrical ligand, this only occurs if two ligand molecules bind in a head-to-tail fashion to one DNA duplex either in an end-to-end or in a side-by-side fashion. Detailed evidence for the head-to-tail, side-by-side nature of the ligands' relationship to one another was clear from NOE data: the protons at the "tail" of one ligand showed NOEs to the "head" of the partnering ligand and vice versa. The quality of the NMR data indicated an absence of any aggregation: ^{1}H and ^{31}P NMR line shapes were the same for all but a few instances for both free DNA and for the 2:1 ligand/DNA complex. Chemical shift changes for DNA proton resonances firmly placed **1** in the minor groove

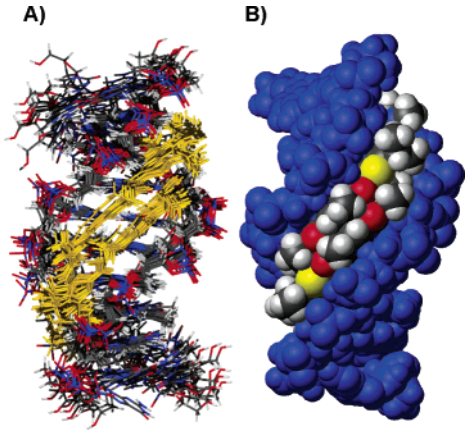


Figure 5. Representation of the solution structure of the complex between **1** and **2** determined in aqueous solution by NMR spectroscopy. (A) Overlay of a family of 10 lowest energy structures taken from different parts of the molecular dynamics trajectory; the ligand is represented in gold. (B) CPK representation of the average structure looking into the minor groove: ligand, atoms colored by atom type; DNA, all atoms shown in blue.

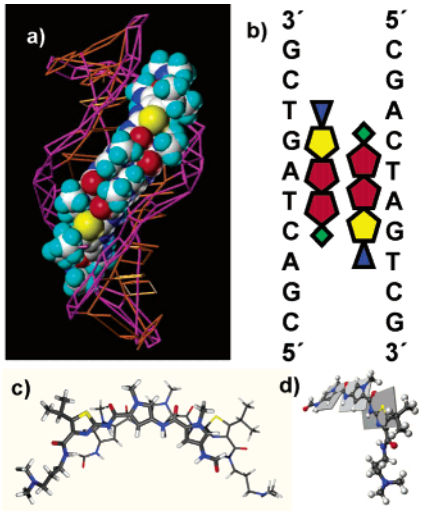


Figure 6. Cartoon and schematic representation of the complex between **1** and **2** showing the location of **1** with respect to the DNA sequence. (A) Cartoon showing a space-filling representation of the ligand molecules against a background of the CURVES framework calculated for the DNA structure of the complex. (B) Schematic indicating the ligand alignment relative to the DNA sequence. Color coding: green diamond = formyl "head"; red pentagon = *N*-methylpyrrole; yellow pentagon = isopropyl-thiazole; blue triangle = DMAP "tail". (C) Relationship between associated ligands in the complex of **1** with **2**. Thicker lines are shown for one ligand compared with its partner. (D) One ligand from the 2:1 complex of **1** with **2** showing how the ligand adapts to the curvature of the DNA minor groove by altering the interplane torsion angles θ_1 and θ_2.

of the duplex as anticipated (Figure 3, Figure S8, and Table S1 in Supporting Information), and ligand/DNA NOEs supported this assertion. A representation of the calculated solution structure of the complex is shown in Figure 5 and schematically in Figure 6a. The family of structures shown (Figure 5a) shows the 10 lowest energy structures from snapshots of a 100 ps molecular dynamics trajectory, together with a minimized average of the 10 structures. Over the whole structure, the best backbone fit rmsd to the average was 1 Å with a standard deviation of 0.2 Å. Consideration only of the central eight-base pair region (i.e., removal of the "end effects") reduced the all-atom rmsd fit to the average structure to 0.9 Å with a standard deviation of 0.2 Å. Overall, the position of **1** relative to the

Table 5. Statistical Details Concerning Interproton Distance Restraints for Both Free d(CGACTAGTCG)$_2$ and Its 2:1 Complex with Thiazotropsin A (**1**)

	structure	
	free DNA	complex
NOE peak assignments	498	793
Mardigras input after conversion based on duplex symmetry	564	1090
interproton distances	522	996
interproton distances used after error checking	514	982[a]

[a] For a symmetrical complex, of which NOEs are categorized as 34 interligand, 78 intraligand, 246 DNA−ligand, 36 DNA interstrand, and 588 DNA intrastrand.

Table 6. Comparison of Key Global Parameters for Three DNA Structures: Canonical B-DNA, the Calculated Ligand-Free B-DNA Structure, and the Calculated DNA Structure in the Presence of 2 Molar Equiv of Ligand Bound in the Minor Groove of the DNA d(CGACTAGTCG)$_2$

	minor groove[a]		
	width (Å)	depth (Å)	overall axis bend (deg)
canonical B-DNA[b]	6.0	5.0	0.0
ligand−Free DNA[c]	4.0	4.0	20.0
DNA with 2 equiv of **1**[c]	7.0	5.0	20.0

[a] Measured for the region -ACTAGT-. [b] Software-generated structure. [c] NMR structure.

DNA sequence and the relative position of ligands to one another in the minor groove varied little from structure to structure: the calculated structures confirmed the basic inter-ligand and ligand−DNA relationships.

The structure calculations resulted in average DNA minor groove parameters of 7.0 Å wide and 5.0 Å deep in the -ACTAGT- region. Comparison is made in Table 6 of the dimensions of the minor groove for both canonical B-DNA and the NMR solution structures calculated for the ligand-free and ligand-bound DNA structures. These figures are generally more a reflection of the computational methods used rather than the data, which poorly defines features such as curvature and groove width. Nevertheless, it is of some interest to examine these values and particularly local perturbations, which may reflect true changes in DNA structure. For instance, the depth of the minor groove increased in the region of the T^{5}pA^6 step at the center of the ligand-binding region, while the minor groove width concomitantly decreased in this part of the structure. The overall curvature was 20°, in close agreement with the value for the calculated ligand-free DNA structure. In general, the global shape of the DNA structure in the complex agreed well with that of the free DNA, most likely as a consequence of the procedures used to calculate the structures. The exception lies in the width of the minor groove caused by the presence of side-by-side binding of **1**. The formyl group (here designated the "head") was firmly orientated toward the 5′-end of the 5′-TAGT-3′ segment of DNA with the positively charged dimethylaminopropyl (DMAP) "tail" located toward the 3′-end, an orientation common with minor groove binders.

1. Hydrophobic Interactions and Solvent-Exposed Surfaces. Part of the proposed design for **1** was to increase the hydrophobic drive toward complex formation with the DNA minor groove. van der Waals interactions are indirectly apparent via NOE contacts that were observed between the DNA backbone and the isopropyl group. Hence (Figure S9) contacts

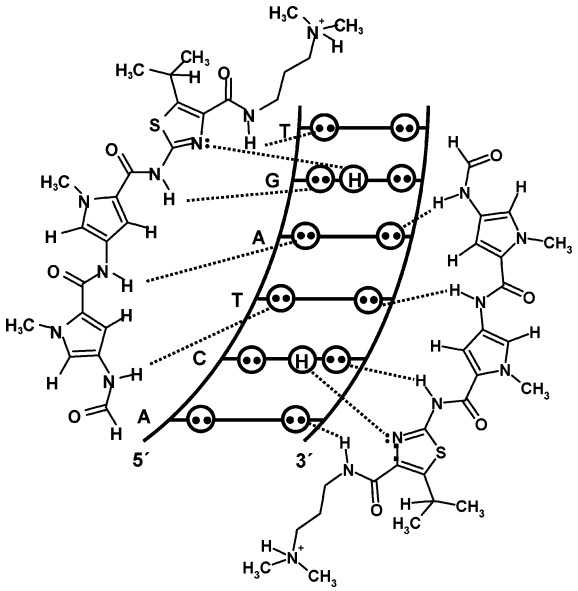


Figure 7. Deduced arrangement of hydrogen bonding between **1** and the DNA duplex d(CGACTAGTCG)$_2$, **2**.

were observed between the isopropyl methyl groups and C^{9}H5″ and T^{8}H4′ protons of the DNA backbone. Such close contact between the outer edge of the ligand and the DNA backbone is also apparent for the methyl protons of the *N*-methyl pyrrole groups (data not shown). On binding to DNA, the nominal change in solvent-accessible surface area experienced by each individual ligand molecule was on the order of 330 Å^2 (calculated on the basis of a 1.4 Å radius probe). Each ligand exposed only 180 Å^2 of surface area when bound to DNA compared with a total surface area of 514 Å^2 for each individual ligand in free solution. When comparing the solvent-exposed surface area of a methylpyrrole group and the isopropylthiazole group, the latter contributes an additional 40 Å^2 of hydrophobic surface area compared with the former. The expectation is that this feature ought to assist the drive toward complex formation and more pronounced exclusion of solvent once the complex has formed.

2. Overlapped vs Staggered Geometry of Ligands. Side-by-side antiparallel binding of polyaromatic amine minor groove binders can occur in either of two refined geometries, namely, fully overlapped or staggered.[29] Although in both binding modes the "aromatic" ring stacks on an amide bond, in the staggered geometry the first amide of each ligand hangs over the adjacent ligand. Binding of **1** occurs in the staggered mode (Figure 6c), thereby extending the reading frame while simultaneously increasing the separation of the charged tails, which reduces the electrostatic repulsion between the two ligands. The positively charged tails were therefore juxtaposed with A/T bases, thereby satisfying their drive to associate with a more negative electrostatic potential.

3. Labile Hydrogen Exchange Characteristics. The hydrogen-bonding scheme for the 2:1 ligand/DNA complex was deduced indirectly on the basis of the calculated solution structure (Figure 7 and Table 7). The labile protons of the ligand associated with the implied hydrogen bonds between the ligand and the DNA (i.e., H2, H9, H16, and H26 of **1**) appeared to be

(29) Kopka, M. L.; Goodsell, D. S.; Han, G. W.; Chiu, T. K.; Lown, J. W.; Dickerson, R. E. *Structure* **1997**, *5*, 1033−1046.

Table 7. Summary of the Implied Hydrogen Bonds in the 2:1 Complex Formed between Thiazotropsin A (**1**) and d(CGACTAGTCG)$_2$ Based on ^{1}H NMR Chemical Shift, Labile Proton Exchange Characteristics, and Solution Structure Information

ligand atom	DNA atom
H2	T^5O2
H9	A^6N3
H16	G^7N3
thiazole N21	G^7H22
H26	T^8O2

inaccessible to solvent (see, for example, Figure S3 in Supporting Information). Solvent accessibility of DNA and ligand protons is revealed by the presence of an exchange correlation between the solvent resonance and that of the labile proton. 2D [^{1}H, ^{1}H] NOESY NMR data acquired for the complex at a probe temperature of 298 K were devoid of exchange cross-peaks between the ligands' peptide N*H* protons and the solvent resonance, implying solvent inaccessibility (data not shown). Additionally, exchange cross-peaks were not observed between the solvent and the imino protons of the central four DNA base pairs of the complex, suggesting protection from solvent accessibility because of the presence of ligand in the minor groove. This is in contrast to the free DNA duplex, NMR data for which clearly showed the presence of exchange cross-peaks between the solvent resonance and all of the observable imino proton resonances (i.e., for the central eight base pairs). The presence of ligand in the minor groove therefore alters the behavior of the DNA, the minor groove being less solvent accessible due to its occupation by ligand molecules. The dynamics of base pair opening into the major groove are modified by the presence of ligand in the minor groove as indicated by the loss of exchange correlations between the solvent and the protons associated with the Watson−Crick base pairs in the vicinity of the ligand reading frame. Inspection of the calculated structure of the 2:1 ligand/DNA complex revealed the close proximity of N21 of the ligand thiazole ring to the non-hydrogen-bonded N2 exocyclic proton (H22) of G^7 (distance = 1.9 Å). Retrospective examination of the NMR data revealed a weak, broad, and unassigned cross-peak at the resonance position of ligand proton H26. This weak cross-peak was assigned to G^7 (H22−H21) where H22 and H21 are the non-hydrogen-bonded and hydrogen-bonded protons, respectively, of the exocyclic N2 amine of the G^7 base. Ordinarily, extreme broadening of the amino proton resonances of G and A bases in DNA occurs through rotation about the C−N bond.[30] The presence of the thiazole nitrogen implicated in the formation of a hydrogen bond with the H22 amine proton of G^7 would tend to arrest this rotational motion, thereby sharpening the resonances associated with the amino groups, enabling them to be observed by NMR, a feature that has been alluded to previously.[31]

4. DNA Backbone Conformation. As for the free DNA, quantitative $^3J(^1H,^{31}P)$ measurements were not carried out for the 2:1 ligand/DNA complex. However, comparison of the ^{31}P chemical shifts for free and ligand-bound DNA enabled general deductions to be made concerning the nature of the DNA backbone. As detailed in Figure 4, all of the ^{31}P NMR resonances occurred within a narrow range of chemical shifts for free DNA, which was not the case for ligand-bound DNA. Changes in ^{31}P chemical shifts that occurred on ligand binding (Table 4) are characteristic of changes in backbone conformation. Good correlation between ^{31}P chemical shift and conformation about torsion angles ϵ and ζ has been amply demonstrated previously.[32] It was not clear from our literature survey whether the observed changes in ^{31}P chemical shift are typical in the context of DNA minor groove ligand binding, although in at least one case ^{31}P NMR has been shown to be a sensitive measure of ligand binding.[33] In different contexts, it has been shown that variations in phosphate-ester torsion angles are the major contributor to variations in ^{31}P chemical shifts. A substantial chemical shift change to high ppm is indicative of conversion from the more stable B_I conformation to the less stable B_{II} conformation ($g^-t - \epsilon,\zeta = -60, 180$)[34] with $\Delta\delta$ (^{31}P) $\approx +1.5-1.6$ ppm.[35] In the current investigation, the T^5-3′-phosphate was the only resonance to show a positive chemical shift change when the ligand was bound. The order of magnitude of $\Delta\delta$ ^{31}P (+1.55 ppm) is in excellent agreement with a B_I to B_{II} conformational change although consideration must rightly be given to the possible influence of ring-current effects on the ^{31}P chemical shift of the T^5-3′-phosphate. This is particularly the case since the neighboring A^6H5″ proton resonance experienced a dramatic chemical shift change when **1** was bound ($\Delta\delta = -1.167$ ppm), an effect almost certainly caused through van der Waals interaction with a nearby pyrrole ring, a promoter of magnetic shielding. For the A^6H5″ proton to encounter a pyrrole ring in such a way as to generate the observed change in resonance frequency, the backbone conformation must alter and the required structural alteration results in the phosphate center moving further *away* from the ligand. Since the large change in ^{31}P chemical shift was the result of magnetic *deshielding*, we attributed the observed change in ^{31}P chemical shift to the influence of a large backbone conformational change. Torsion angle restraints along the DNA backbone were largely omitted from our structure calculations (although based on our findings a set of calculations were carried out in which torsion angles around the TpA steps were restrained). Despite this, the family of calculated structures for the 2:1 ligand/DNA complex reflected a substantial alteration of geometry at the T^5-3′-phosphate position consistent with a B_{II} conformation.

5. DNA Structural Parameters. A CURVES[25] description of the average DNA structure of the ligand−DNA complex differed from that of the free DNA structure. Average global base pair-to-axis displacement parameters generally only showed minor deviations from the values calculated for canonical B-DNA- (dx = −0.9 Å, dy = 0.3 Å) although inclination (η) and tip (θ) showed higher average values ($\eta = 2°$, $\theta = -2°$). Interestingly the central A/T base pairs showed significant inclination (11 and 13°) compared with the other base pairs. Average buckle (κ) and propeller twisting (ω) of base pairs was also more pronounced than for the free-DNA structure (average $\kappa = -10°$ and average $\omega = -10°$), although large contributions

(30) Mueller, L.; Legault, P.; Pardi, A. *J. Am. Chem. Soc.* **1995**, *117*, 11043−11048.
(31) Keniry, M. A.; Shafer, R. H. *Methods Enzymol.* **1995**, *261*, 575−604.
(32) Gorenstein, D. G. *Methods Enzymol.* **1992**, *211*, 254−286.
(33) Seaman, F. C.; Hurley, L. H. *Phosphorus, Sulphur Silicon Relat. Elem.* **1999**, *146*, 297−300.
(34) (a) Chou, S. H.; Cheng, J. W.; Reid; B. R. *J. Mol. Biol.* **1992**, *288*, 138−155. (b) Legault, P.; Pardi, A. *J. Magn. Reson. Ser. B* **1994**, *103*, 82−86.
(35) Gorenstein, D. G. *Annu. Rev. Biophys. Bioeng.* **1981**, *10*, 355−386; Roongta, V. A.; Jones, C. R.; Gorenstein, D. G. *Biochemistry* **1990**, *29*, 5245−5258.

to these figures came from the terminal base pairs. Of the global inter-base pair parameters, roll (ρ) was significant at the T^5pA^6 step (−24°), which was also the largest contributor to the global axis curvature. Sugar pseudorotation phase angles varied, but most were consistent with the "*S*"-type sugar conformation. Even though the phosphodiester backbone had not been restrained, the structure indicated some alteration at the T^5pA^6 step, which was consistent with the ^{31}P chemical shift change that was observed for the phosphorus atom in this position.

IV. Discussion

Pioneering research carried out over the past decade has helped to define the characteristics of DNA minor groove binding ligands and their potential to act as therapeutic drugs. Milestones have included the development of tight binding, DNA sequence-specific molecules, new paradigms in minor groove ligand structure and functionality, and increased cell (including cell nucleus) permeability. However, to create a successful therapeutic drug, low toxicity with cell permeability and targeted DNA sequence recognition of 10−16 base pairs must be achieved. In contrast to the approach of increased ligand complexity, functionality, and molecular weight to achieve these ends, commercially viable options for drug candidates require simplicity and ease of manufacture to be combined with these DNA recognition and cell penetration characteristics. To this end, we have taken the rational approach of designing a new class of lexitropsin with hydrophobicity as one of its key elements. While the aim here was to work with small molecules, we have demonstrated that such compounds are capable of sequence specificity and strong binding to extended DNA sequences through targeted minor groove molecular recognition processes.

A. New Role for Thiazole. Previously, thiazole (Th) subunits have been incorporated into aromatic amino acid minor groove binders in such a way as to locate the sulfur on the minor groove floor.[36−38] The effect has been to confer A·T recognition preference over G·C, overcoming the degeneracy, by steric means, of polyamide A·T/T·A recognition.[38] In the present case, however, we have designed the thiazole subunit so that the sulfur supplements the hydrophobicity of the ligand by lying at the top of the minor groove on its outside edge rather than being buried deeply against the minor groove floor. In combination with the adjacent isopropyl group, the design was aimed at increasing the hydrophobicity rather than affecting steric interactions. At the same time, the hydrogen-bonding capacity was to be retained via the nitrogen center of the same subunit through coordination with the exocyclic N*H* of G residues on the groove floor. In this way it was hoped that enhanced affinity (through hydrophobicity) *and* selectivity (through additional hydrogen bonding interactions) could be achieved and that thiazole itself could be made specific for G recognition. The NMR data presented here indicate that both aims have been achieved. The alteration in behavior of the exocyclic NH_2 ^1H NMR resonances for G^7 on ligand binding indicates how the presence of the thiazole ring nitrogen atom on the floor of the minor groove is important for alignment of the ligand against the DNA sequence presented to it. It is therefore clear that in this context, *S*-alkyl thiazoles can be classified as a new recognition element for G bases in the minor groove of DNA.

B. Occupancy of the DNA Minor Groove by Two Ligand Molecules Creates a Highly Stabilized, Low-Energy Complex. The degree of induced-fit of a ligand with DNA is known to be a consequence of optimizing the van der Waals contacts between a ligand and the minor groove walls.[11] Groove flexibility is regarded as asymmetric, and a low energetic cost is incurred when groove widening occurs.[5] A comparison of the minor groove dimensions of the free and 2:1 ligand-bound form of the DNA can be made directly from the DNA structures determined using the NMR data (Table 6). From a comparison of the overall geometries of the ligand-free DNA and the ligand-bound DNA structures, it is clear that minor groove width is one of the few obvious differences. From the titration of DNA with ligand and experiments using the reverse approach (in which ligand was titrated with DNA), it was clear that the minor groove was occupied with 2 equiv of ligand. Our NMR data showed no evidence for an intermediate step in which the DNA minor groove became fully occupied with 1 equiv of ligand prior to the addition of a second equivalent of ligand. However, this process cannot be ruled out. Our modeling calculations indicate a highly stabilized, low-energy complex. If such an intermediate step does occur in this case, NMR spectroscopy is unlikely to be able to detect it, since the data provide a time average of the binding process, which for a 1:1 complex is likely to occur with a shorter residence time compared with the 2:1 complex. A study of the forces driving the binding process has not been attempted here, but we speculate that a positive ligand electrostatic potential combines with increased hydrophobicity to drive the binding process forward. The presence of 2 equiv of ligand in the minor groove is required to satisfy the minor groove shape (through induced-fit) and to maximize hydrophobic interactions between the DNA sugar residues and the ligand.

C. B_{II} Substate Appears to be Stabilized at the TpA Step. Of additional interest in the alteration of the DNA geometry on ligand binding was the apparent stabilization of a B_{II} substate in the phosphate backbone 3′ to T,[5] as demonstrated by the large chemical shift change (to higher ppm) for the 3′-phosphate resonance of the T^5 residue ($\Delta\delta = +1.55$ ppm). This is supported in our structures calculated using restrained molecular dynamics but without phosphodiester backbone restraints, in which base pair roll (ρ) at the T^5pA^6 step and B_{II}-type backbone ϵ and ζ torsion angles are consistent with the observed NMR data. Arrest of dynamic motion and alteration of the structure of the DNA backbone plays a crucial role in the mechanism by which minor groove binding drugs distort the binding sites of proteins in the major groove. B_I/B_{II} substate transitions are believed to act as additional conversation tools to transfer binding information from a ligand in the minor groove to a protein-binding site in the major groove.[39] By locking part of the backbone in the B_{II} state, as we believe is the case described here, the minor groove binding ligand can potentially offer a mode through which protein binding in the major groove may be mediated. Physical bulk is provided by the isopropyl group being adjacent to the exposed sulfur atom of the thiazole moiety,

(36) Rao, K. E.; Shea, R. G.; Yadagiri, B.; Lown, J. W. *Anti-Cancer Drug Des.* **1990**, *5*, 3−20.
(37) Sharma, S. K.; Tandon, M.; Lown, J. W. *J. Org. Chem.* **2000**, *65*, 1102−1107.
(38) Nguyen, D. H.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 7−17.
(39) Flader, W.; Wellenzohn, B.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. *J. Phys. Chem. B* **2001**, *105*, 10379−10387.

which in the vicinity of the TpA step, could cause this localized alteration in backbone behavior. However, if steric bulk is a factor in altering the backbone behavior in this part of the structure, the effect is one that must be transmitted, since the $^{i\mathrm{Pr}}$Th is relatively remote: a second ligand lies between the $^{i\mathrm{Pr}}$Th moiety and the relevant phosphodiester group. In this case, it is entirely reasonable that the *N*-methyl pyrrole immediately juxtaposed with the T^5pA^6 step is forced into closer contact with the adjacent sugar units, which become locked into a close association with one another. The result is that the conformational flexibility normally experienced by the sugar/phosphodiester backbone is arrested at this point. "Locked out" of sampling B_I-type conformational space at this position, the DNA backbone adopts the next most stable B_{II}-type conformation. Clearly, it is of interest to pursue this theme of "freezing in" an altered B_I/B_{II} substate pattern in a form entirely different from that of normal DNA in order to provide additional tools for down regulating aberrant gene expression, especially since phosphate groups along the DNA backbone contribute the major part of DNA−protein recognition contacts in both nonspecific *and* highly specific DNA−ligand complexes.[40]

D. Extended Reading Frame is Created by a Staggered Relationship between Adjacent Ligands. From previous studies of *N*-methyl pyrrole-based DNA minor groove binding ligands, it has been noted that the *N*-methyl group supplies a possible role in aligning the side-by-side register of ring pairings.[41] Our studies suggest that a combination of the presence of thiazole sulfur and *C*-isopropyl functionalities contribute to this effect and aid in setting the pyrrole/thiazole rings deeply into the minor groove. In the case of the 5′-ACTAGT-3′ reading frame, thiazole nitrogens are expected to form hydrogen bonds with the guanosine exocyclic amino protons, thereby causing the ligands to align in a staggered relationship. The conditions for a staggered mode of binding are therefore provided, thereby extending the DNA reading frame to six base pairs.[42] We have found no evidence to date for the formation of an overlapped configuration of ligand molecules, which would recognize a shorter binding site. This may be possible in theory, but from our structural studies, steric bulk caused by the presence of both the thiazole sulfur atom and the adjacent isopropyl group gives the strong impression that a ligand overlap binding mode is unlikely. The staggered relationship between two ligand molecules in the bound state is illustrated in Figure 6c.

E. Induced-Fit Results in Elegant Matching of Hydrogen Bonding Elements. Overall, the geometry of each ligand, in terms of length and curvature, is optimal to position the component elements against the correct recognition groups of the DNA sequence. The torsion angles θ_1 and θ_2 define the relationship between ring planes within each ligand, and, adapting to the geometry of the minor groove, $\theta_1 = 22°$ and $\theta_2 = 27°$ as illustrated schematically in Figure 6d. As with previous reports of this type of binding, induced-fit occurs both for the DNA and for the ligand and optimizes hydrogen bonding between groups on the concave edge of the ligand and the minor groove floor, particularly the N*H* protons and DNA base N3 nitrogen or thymine O2 oxygen atoms (Table 6): H26 to T^8 O2 (1.91 Å); H16 to G^7 N3 (1.88 Å); H9 to A^6 N3 (1.98 Å); H2 to T^5 O2 (1.71 Å). This is particularly elegant in the case of additional hydrogen bond formed between the thiazole N and G^7 H22 of the exocyclic N2 amine, which is perfectly placed to assist with control of recognition and alignment with the correct reading frame. Thus, isopropylthiazole ($^{i\mathrm{Pr}}$Th) genuinely introduces a new sequence reading element.

F. Ligand Residency Time is Relatively Long. Studies of 1:1 complexes between unsymmetrical ligands and self-complementary DNA duplexes (where symmetry is destroyed as the result of tight binding of an unsymmetrical ligand in the DNA minor groove) have reported ligand interchange rates on the order of $1.2-2.5\ s^{-1}$.[11] Ligand interchange can sometimes be observed by the presence of NMR cross-peaks in the NOESY NMR data sets that are the result of chemical exchange processes. From our data, the extent of ligand interchange remains unclear, a fact partially associated with the study of symmetrical systems. In NOESY NMR data sets acquired on the system under discussion in this article, at the point of "half occupancy" (indicated in Figure 1b) no such cross-peaks were observed between ligand-free and ligand-bound forms of DNA. This information, together with the appearance of well-resolved signals for free and bound forms of DNA at the same titration point, is a reflection of a ligand binding to DNA with a relatively long residency time. This in turn is a reflection of the binding constant. Preliminary results using fluorescently tagged oligonucleotides show that the binding of **1** to ACTAGT is better by a factor of 2 than the binding of distamycin to the same DNA sequence (C_{50} [μM] = 11.2 ± 1.6 (ΔT_m = +8.0 °C at a ligand concentration of 10 μM) for **1** compared with 22.8 ± 2.1 (ΔT_m = +5.0 °C at a ligand concentration of 10 μM) for distamycin, where C_{50} is the ligand concentration that produces a half-maximal change in DNA melting temperature on ligand binding).[43] Further work in understanding the factors that influence the binding process, in particular the DNA sequence dependency of ligand binding, are currently in progress and will be reported in a subsequent article.

V. Conclusion

In this article we have shown clear evidence for a new class of lexitropsins that bind sequence specifically to the DNA minor groove in a 2:1 ratio with enhanced recognition for G DNA bases. Our detailed NMR studies show that the presence of the new $^{i\mathrm{Pr}}$Th element of the ligand introduces both hydrophobicity and enhanced sequence specificity through hydrogen bond formation between the $^{i\mathrm{Pr}}$Th nitrogen (on the floor of the minor groove) and the exocyclic amine protons of G residues. Both the shape of the molecule and its ability to undergo induced-fit recognition allows our ligand, thiazotropsin A, **1**, to target the specific DNA sequence 5′-ACTAGT-3′, molding itself to both the global shape and sequence of the DNA structure. A comparison of the ligand-free and ligand-bound forms of the DNA in solution indicates that the overall shape of the DNA is retained but that a widened groove accommodates two ligands

(40) (a) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. *J. Am. Chem. Soc.* **2001**, *123*, 5044−5049. (b) Kielkopf, C. L.; Ding, S.; Kuhn, P.; Rees, D. C. *J. Mol. Biol.* **2000**, *296*, 787−801. (c) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nature Struct. Biol.* **1998**, *5*, 104−109.
(41) Bremer, R. E.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *Bioorg. Med. Chem.* **2000**, *8*, 1947−1955.
(42) O'Hare, C. C.; Mack, D.; Tandon, M.; Sharma, S. K.; Lown, J. W.; Kopka, M. L.; Dickerson, R. E.; Hartley, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (*1*), 72−77.
(43) Fox, K. R. (University of Southampton, Southampton, UK) Personal communication.

side-by-side. The possible adoption of a B_{II} backbone conformation 3′ to T^5 is demonstrated by ^{31}P NMR data and thought to be the result of steric bulk at ^{iPr}Th combined with strong hydrophobic interactions between *N*-methyl pyrrole of a neighboring ligand and sugar rings bridged by the relevant phosphodiester group. Alteration of the DNA backbone in this way is of potential significance in terms of protein recognition and gene transcription. The staggered nature of the side-by-side minor groove ligand binding has enabled a six-base pair DNA footprint to be recognized in contrast to the normal expectation of four base pairs based on the size of the ligand. The importance of this in terms of gene targeting and drug-led candidate development cannot be understated. Small DNA-binding drug molecules that can target longer sequences of DNA will be those that become commercially viable, especially if they possess the hydrophobicity demonstrated in this new class of lexitropsins, an essential requirement for penetration of the cell membrane and nucleus while remaining intact and functional. The development of such gene-targeting, small molecules is of significant interest as their viability as drug candidates improves. Further studies on these developments are in progress in our laboratories, and the results of this work will be reported in due course. Overall, our work shows that considerable scope continues to exist in generating new types of DNA frame-reading ligands, of all the more significance with the completion of the human genome project and the requirement for targeted, feasible, and efficient drug therapies capable of tackling the sources of genetic disease.

Acknowledgment. This project was funded under the Glasgow/Strathclyde Synergy Fund and through a Strathclyde University Research Development Award (RDF 1160 to J.A.P.). We thank the University of Edinburgh, Scotland, UK, for the provision of 600 MHz instrument time and the MRC and the NMR staff of the National Institute for Medical Research, Mill Hill, London, UK, for the provision of 600 and 800 MHz NMR instrument time. We acknowledge the Wellcome Trust for the provision of computational facilities for the structure determinations detailed in this article. We thank Prof. Keith R. Fox of the University of Southampton for communicating the C_{50} values.

Supporting Information Available: Coordinate files for both **2** and the 2:1 complex between **1** and **2** are available under PDB accession codes 1RN9 (rcsb020900) and 1RMX (rcsb020892), respectively; table of chemical shift differences for ligand-bound and free DNA (Table S1); table of NOE assignments for ligand/DNA complex (Table S2); relevant 1D ^1H NMR subspectra (Figures S1−S4); fingerprint region of the NOESY NMR data acquired on free $d(CGACTAGTCG)_2$ (Figure S5); selected regions of DQFCOSY and NOESY NMR data of the ligand/DNA complex (Figure S6); fingerprint region of the NOESY NMR data acquired on the ligand/DNA complex (Figure S7); plot of chemical shift differences for H1′ protons in bound and free DNA (Figure S8); illustration of NOE contacts between the ligand isopropyl group and the DNA backbone in the complex (Figure S9); calculated structure of free $d(CGACTAGTCG)_2$ in solution (Figure S10); 2D [^{31}P, ^1H] NMR data for the complex (Figure S11). This material is available free of charge via the Internet at http://pubs.acs.org.

JA030658N