Pergamon

Bioorganic & Medicinal Chemistry 10 (2002) 241–252

BIOORGANIC & MEDICINAL CHEMISTRY

Recognition of Nine Base Pair Sequences in the Minor Groove of DNA at Subpicomolar Concentrations by a Novel Microgonotropen

Alexander L. Satz and Thomas C. Bruice*

Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106, USA

Received 13 July 2001; accepted 13 July 2001

Abstract—The dsDNA interactions of the novel microgonotropen **L1** have been characterized via spectrofluorometric titrations and thermal melting studies. A microgonotropen consists of a DNA minor groove binding moiety attached to a basic side chain capable of reaching out of the minor groove and grasping the acidic DNA phosphodiester backbone. **L1** was synthesized employing solid-phase chemistry. **L1** is shown to distinguish nine base pair A/T rich binding sites from sites possessing fewer than nine contiguous A/T base pairs. Further, **L1** binds its preferred dsDNA sequences at subpicomolar concentrations. The equilibrium constant for complexation (K_1) of a nine base pair A/T rich dsDNA binding site by **L1** is roughly 10^{13} M^{-1}. Single base pair A/T → G/C substitutions within the nine base pair A/T rich binding site of **L1** decreases the equilibrium constant for DNA binding by 1–2 orders of magnitude. The three proplyamine side chains of **L1** enhance the agents free energy of binding by more than 5 kcal. Molecular modeling suggests that **L1** adopts a 'spiral-like' conformation which fits almost a full turn of the DNA helix.

Introduction

Studies indicate that minor groove binding agents may influence the regulation of gene expression by inhibiting the binding of regulatory proteins to their double-stranded DNA (dsDNA) binding sites.[1,2] Interest in controlling the expression of specific genes has spurred efforts toward the development of agents with greater sequence selectivity. It is expected that those agents capable of recognizing longer dsDNA sequences will exhibit the greatest specificity, however those few agents targeted to longer dsDNA sequences have generally lacked specificity.[3–12] We recently reported a novel minor groove binding agent, **L3**, capable of distinguishing nine bp A/T rich binding sites from sites possessing fewer than nine contiguous A/T base pairs (Chart 1).[13] In contrast, the 'classical' minor groove binding agents Hoechst 33258 (**L9**) and distamycin (**L7**) cannot distinguish between these A/T rich binding sites.[1,14] **L3** is proposed to adopt a 'spiral' shape which allows it to conform to almost a full turn of the dsDNA helix.[13]

Our interest in the control of gene expression via inhibition of transcription factor (TF) binding has also lead to the development of microgonotropens.[15,16] MGT-6a (**L6**) is an example of a first generation MGT which was based upon a tripyrrole polyamide minor groove binding moiety. To the central pyrrole unit of these MGTs is attached a basic polyamine chain capable of associating with the acidic phosphodiester backbone of dsDNA. The tren polyamine chain of **L6** is known to reach into the major groove of dsDNA and grasp the DNA backbone.[17] The polyamine tails of the MGTs endow them with a superior ability to inhibit the binding of transcription factors (TFs) to their dsDNA binding sites in cell free assays.[2,18,19] Second generation MGTs, or fluorescent MGTs, were based upon a bisbenzimidazole minor groove binding moiety which emits a significant fluorescence signal upon complexation of A/T rich dsDNA.[20,21] Thus, interesting aspects of ligand–dsDNA interactions can be readily investigated employing fluorescence spectroscopy. Again, the polyamine chains of these agents endow them with a superior ability to inhibit TF binding in cell free assays.[22]

We sought to design a DNA binding agent which would specifically recognize longer DNA sequences while also possessing the superior TF inhibitory activity of the

*Corresponding author. Tel.: +1-805-893-2044; fax: +1-805-893-2229; e-mail: tcbruice@bioorganic.ucsb.edu

0968-0896/02/$ - see front matter
PII: S0968-0896(01)00272-3

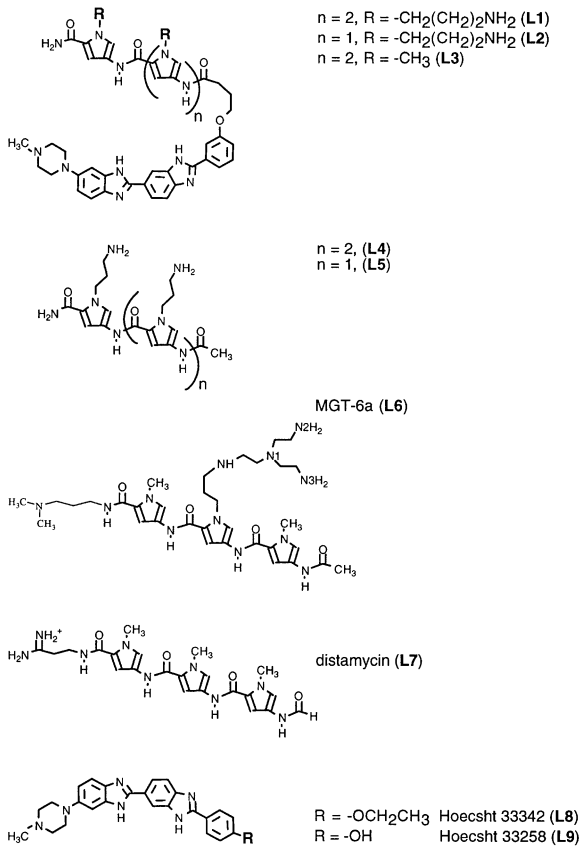


Chart 1.

MGTs. Thus, we replaced the pyrrole *N*-methyl substituents of **L3** with propylamine chains to yield a novel microgonotropen (**L1**). White et al. has recently shown this novel MGT to be a more potent inhibitor of TF complex formation (in cell free assays) than **L6**, **L7**, or **L8**. **L1** was also shown to inhibit gene expression in whole cells, making it the first MGT to possess whole cell activity.[23] The first and second generation MGTs show little activity in whole cells.[24] Thus, it appears that **L1** possesses the ability to traverse the cell membrane and localize in nuclear DNA.

oligomeric duplexes

(**1**) 5′-CGCAAAAACGCACC-3′

(**2**) 5′-CGCAAAAAAAAACGC-3′

(**3**) 5′-CGCAAATTCGCACC-3′

(**4**) 5′-CGCTATAACGCACC-3′

(**5**) 5′-GGACGTCGAATTGCAGTCGC-3′

(**6**) 5′-GGACGTCGTTGCAGTCGTCG-3′

(**7**) 5′-GCGGTATAAAATTCGACG-3′

Chart 2.

We report the binding characteristics of **L1**, and its tripyrrole polyamide precursor **L4**. Also, we have investigated **L2**, a novel analogue of **L1** possessing one less pyrrole unit and its dipyrrole polyamide precursor **L5**. We have investigated these compounds via spectrofluorometric titrations and thermal melting studies employing seventeen different oligomeric duplexes. The oligomeric duplexes **1**–**4** were employed to determine the effect of binding site size and sequence on equilibrium constants for dsDNA complexation by ligands (Chart 2). The duplexes **5** and **6** were employed to investigate the sequence selectivities of **L4** and **L6**. The oligomeric duplexes **7**–**17** were employed to investigate the effect of single and double base substitutions of the type A/T → G/C within the nine bp A/T rich binding site of **7** on binding by **L1** (Table 5). **7** contains the TATA box, which in eukaryotes consists of the consensus sequence 5′-TATAAAA-3′, and is recognized by the TBP (TATA binding protein).[25]

Materials and Methods

Materials

Purified DNA oligomers were purchased from the Biomolecular Resource Center, University of California at San Francisco. **L8**, **L9**, and 0.05 wt% 3-(trimethylsilyl)-propionic-2,2,3,3-d_4 acid, sodium salt in D_2O were purchased from Aldrich Chemical Company and used without further purification. **L7** was purchased from Sigma. Solvents and most reagents including triisopropylsilane (TIS), dimethylformamide (DMF), dicyclohexylcarbodiimide (DCC), diisopropylethylamine (DIPEA), acetonitrile (ACN), dimethyl sulfoxide (DMSO), and hydroxybenzotriazole (HOBt) were purchased from Aldrich Chemical Company. Rink amide MBHA resin and benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBOP) were bought from Novabiochem. The synthesis of compounds **L3**,[13] **18**,[18] and **22**[20] have been previously published.

DNA binding experiments

All DNA binding experiments employed 10 mM potassium phosphate pH 7.0 buffer treated with Chelex and filtered (0.45 μm). Solutions of differing ionic strengths (μ) were prepared by the addition of NaCl to buffer solutions prior to treatment with Chelex and filtering. Oligomeric duplexes were annealed and their molar extinction coefficients determined as previously discussed.[13] Solutions of known ligand concentrations were prepared via NMR peak integration where the ligand samples used contained a known quantity of the internal standard 3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt.[13,20] Thermal melting curves were acquired on a Cary 100 Bio UV–vis spectrophotometer equipped with a temperature programmable cell block. Data points were taken every 1 °C with a temperature ramp of 0.5 °C min^{-1}. Thermal melting temperatures (t_m) were determined as described by Marky and Breslauer.[26] ΔH° values for duplex formation were

determined by van't Hoff plots (ΔH° values are provided in Tables 1 and 2). Gibbs free energies for ligand binding (ΔG°) and corresponding apparent equilibrium constants (K_1) for DNA binding were calculated from t_m^0 and t_m values employing eq (1).[27] K_1 values are provided in Tables 3 and 5.

$$\frac{1}{Rn}\left(\frac{1}{t_m^0}-\frac{1}{t_m}\right)=\frac{-\ln a}{\Delta H}+\frac{\Delta G}{\Delta H R t_m} \quad (1)$$

t_m^0 and t_m are the melting points of the duplex without and with the presence of ligand, respectively, and are provided in Tables 1 and 2. a is the concentration of ligand free in solution when the temperature of the solution is equal to the t_m of the complex (a was estimated to be half the total ligand concentration). Calculated ΔG° and K_1 values are at the complex's t_m and at ionic strengths provided in the text, and are otherwise at standard state conditions. Although K_1 values in Tables 3 and 4 are not extrapolated to a common temperature, they are still very helpful when trying to interpret the observed t_m values.

Fluorescence spectra were obtained on a Perkin-Elmer LS50B fluorimeter equipped with a constant temperature water bath set at 26 °C. Solutions were excited at 345 nm. Emissions were monitored at 475 nm except in the cases of **L8** and **L9** whose emissions were monitored at 450 nm. The generation of isothermal binding curves and their fitting to determine K_1 values has been previously discussed.[13] In short, a nanomolar concentration of dsDNA is titrated with a relatively

Table 1. Melting temperatures (Δt_m, °C) for ligand complexes of **1**–**6**[a]

Ligand		μ	**1**[c]	**2**	**3**	**4**	**5**	**6**
None	t_m^0	0.032	44	36	45	39	53	58
	$-\Delta H^\circ$ (kcal)		7373		70	57	96	93
	t_m^0	0.17	56	50	57			
	$-\Delta H^\circ$ (kcal)		100	102	111			
L1		0.032	~40[b]	~51[b]				
		0.17	14	~33[b]				
L2		0.03	22	~38[b]				
		0.17	9	23				
L3		0.032	9	24				
		0.17	0	16				
L4		0.032	22	36	22	21	16	9
		0.17	10	15	9			
L5		0.032	9				6	1
		0.17	2					
L6		0.032	22	31	21	23	15	8
		0.17	7	11	6			
L7		0.032	9	13	8	11	4	0
		0.17	4	6	2			
L8		0.032	10	16	12	9	7	0
		0.17	4	5	5			

[a]Thermal melting curves acquired in 10 mM potassium phosphate pH 7.0 buffer, 10 (μ = 0.032) or 150 mM NaCl (μ = 0.17). t_m^0 values are melting temperatures for 0.3 μM oligomeric duplex in the absence of ligand. Δt_m are differences in melting temperatures for oligomeric duplexes in the absence and presence of two equiv ligand. Standard deviation for Δt_m values are roughly ± 1 °C. ΔH° values for oligomeric duplexes have standard deviations of ± 10% and were determined by van't Hoff plots.
[b]Some melting curves went off scale such that their Δt_m values are approximate (see Figure 1 for an example).
[c]DNA sequences are provided in Chart 2 of the manuscript and their potential A/T rich binding sites are underlined.

Table 2. Melting temperatures (Δt_m, °C) for **L1** Complexes of **7**–**7**[a]

Oligomeric duplexes[b]	$-\Delta H$(kcal)	t_m^0(C)	Δt_m(°C)
5'-GCGGTATAAAATTCGACG-3' (**7**)	138	56	21
5'-GCGG**C**ATAAAATTCGACG-3' (**8**)	159	61	16
5'-GCGGT**G**TAAAATTCGACG-3' (**9**)	141	59	14
5'-GCGGTA**C**AAAATTCGACG-3' (**10**)	131	59	13
5'-GCGGTAT**G**AAATTCGACG-3' (**11**)	135	57	16
5'-GCGGTATA**G**AATTCGACG-3' (**12**)	138	55	15
5'-GCGGTATAA**G**ATTCGACG-3' (**13**)	132	57	15
5'-GCGGTATAAA**G**TTCGACG-3' (**14**)	143	57	14
5'-GCGGTATAAAA**C**TCGACG-3' (**15**)	149	56	17
5'-GCGGTATAAAAT**C**CGACG-3' (**16**)	151	57	17
5'-GCGGTATA**GG**AATTCGCG-3' (**17**)	128	56	13

[a]Thermal melting curves acquired in 10 mM potassium phosphate pH 7.0 buffer, 150 mM NaCl (μ = 0.17). t_m^0 values are melting temperatures for 0.15 μM oligomeric duplex in the absence of ligand. Δt_m are differences in melting temperatures for oligomeric duplexes in the absence and presence of two eq ligand. Standard deviation for Δt_m values are roughly ± 1 °C. ΔH° values for oligomeric duplexes have standard deviations of ± 10% and were determined by van't Hoff plots.
[b]The ligand binding site is underlined. Substitution sites are in bold type and larger font.

Table 3. Apparent equilibrium constants ($K_1 \times 10^{-9}$, M^{-1}) for complexation of oligomeric duplexes **1**–**6**. K_1 values were determined via melting curves of ligand–dsDNA complexes[a]

Ligand	μ	**1**[b]	**2**	**3**	**4**	**5**	**6**
L1	0.032	**1300**	**~65,000**[c]				
	0.17	1.6	~7600[c]				
L2	0.032	**5.9**	**1430**				
	0.17	0.19	120				
L3	0.032	**0.080**	**17**				
	0.17	0.0033	5.7				
L4	0.032	**5.9**	**780**	**4.1**	**1.1**	**3.3**	**0.14**
	0.17	0.29	3.7	0.29			
L5	0.032	**0.080**				**0.048**	**$\Delta t_m = 1$**[d]
	0.17	0.0083					
L6	0.032	**5.9**	**160**	**3.0**	**1.8**	**2.2**	**0.092**
	0.17	0.079	0.60	0.067			
L7	0.032	**0.080**	**0.40**	**0.050**	**0.075**	**0.020**	**$\Delta t_m = 0$**[d]
	0.17	0.021	0.059	0.0092			
L8	0.032	**0.11**	**1.1**	**0.18**	**0.043**	**0.074**	**$\Delta t_m = 0$**[d]
	0.17	0.021	0.037	0.041			

[a]Apparent K_1 values are derived from melting curves of ligand–dsDNA complexes and were calculated employing eq (1) as discussed in the text. Ligand/dsDNA melting curves acquired in 10 mM potassium phosphate pH 7.0 buffer with 10 mM NaCl (μ = 0.032) or 150 mM NaCl (μ = 0.17) are listed in bold and regular type, respectively. t_m values for oligomeric duplexes and their ligand complexes are provided in the supporting information. Standard deviations for K_1 values are less than ± 60%.
[b]DNA sequences are provided in Chart 2 and their potential A/T rich binding sites are underlined.
[c]Some melting curves went off scale such that their K_1 values are approximate (see the L1/2 complex melting curve in Figure 1).
[d]There was little difference in t_m values between the oligomeric duplex with and without the presence of ligand (Δt_m). Thus, K_1 values could not be calculated for these complexes.

Table 4. Equilibrium constants ($K_1 \times 10^{-9}$ M^{-1}, 26 °C) for complexation of oligomeric duplexes **1**–**7**. Values were determined via isothermal binding curves[a]

Ligand	μ[b]	EtOH (%)	**1**[c]	**2**	**3**	**4**	**7**
L1	**0.17**	**0**		∼ **5000**[d]			∼**10,000**[d]
	0.77	25	0.028	2.6			5.4
L3	**0.17**						**1.9**[g]
	0.77						0.024[g]
L4[e]	**0.17**	**0**	**65**		**360**	**1.6**	
	0.77	25	0.032				
L6[e]	**0.17**	**0**	**0.18**				
L7[e]	**0.17**	**0**	**0.15**				
L8	**0.17**	**0**	**0.25**	$K_1 = 0.9$[f] $K_2 = 0.004$[f]	**6.0**	**0.045**	
	0.77	25	0.0063				
L9	**0.17**	**0**		$K_1 = 0.050$[f] $K_2 = 0.008$[f]			

[a]K_1 values determined by nonlinear least squares fitting of isothermal binding curves at 26 °C employing eqs (3), (4), or (5). The stated equilibrium constants have a standard deviation of ± 50%.
[b]K_1 values acquired in 10 mM potassium phosphate pH 7.0 buffer with 150 mM NaCl (μ = 0.17) are listed in bold type. K_1 values acquired in a solution of 75% 10 mM potassium phosphate pH 7.0 buffer with 1 M NaCl and 25% ethanol (μ = 0.17) are listed in regular type.
[c]DNA sequences provided in Chart 2.
[d]K_1values estimated as discussed in text.
[e]K_1 values determined via spectrofluorometric competition assays with **L8**.
[f]Complex stoichiometries are 2:1.
[g]The **L3–7** complex has a stoichiometry of 2:1.[13] The K_1 and K_2 terms for the **L3–7** complex cannot be separated due to their similarity in magnitude. Approximate K_1 values in units M^{-1} are provided by taking the square root of the multiplied through K_1K_2 M^{-2} value.

Table 5. Apparent equilibrium constants ($K_1 \times 10^{-9}$ M^{-1}) for complexation of oligomeric duplexes **7**–**17** by **L1**. K_1 values were determined via melting curves of **L1**–dsDNA complexes[a]

Oligomeric duplexes[b]	K_1
5-GCGGTATAAAATTCGACG-3 (**7**)	970
5-GCGG**C**ATAAAATTCGACG-3 (**8**)	180
5-GCGGT**G**TAAAATTCGACG-3 (**9**)	18
5-GCGGTA**C**AAAATTCGACG-3 (**10**)	5.6
5-GCGGTAT**G**AAATTCGACG-3 (**11**)	43
5-GCGGTATA**G**AATTCGACG-3 (**12**)	33
5-GCGGTATAA**G**ATTCGACG-3 (**13**)	20
5-GCGGTATAAA**G**TTCGACG-3 (**14**)	23
5-GCGGTATAAAA**C**TCGACG-3 (**15**)	230
5-GCGGTATAAAAT**C**CGACG-3 (**16**)	250
5-GCGGTATA**GG**AATTCGCG-3 (**17**)	5.4

[a]Apparent K_1 values are derived from melting curves of ligand–dsDNA complexes which were acquired in 10 mM potassium phosphate pH 7.0 buffer with 150 mM NaCl (μ = 0.17). K_1 values were calculated from melting curves employing Eq. (1) as discussed in the text. t_m values for oligomeric duplexes and their ligand complexes are provided in the supporting information. Standard deviations for K_1 values are less than ±60%.
[b]A/T rich ligand binding sites are underlined. Substitution sites are in bold type and larger font.

concentrated solution of ligand dissolved in dimethyl sulfoxide. Formation of the ligand/dsDNA complex is monitored via the solutions fluorescence emission. If the total concentration of dsDNA employed ($[DNA]_T$) is sufficiently small compared to the ligand's K_1 value, $[DNA]_T = 1/K_1$, then a meaningful isothermal binding curve may be generated by plotting fluorescence signal versus the concentration of ligand free in solution ($[L]_f$) (Fig. 3). The $[L]_f$ term is calculated employing eq (2) where n is the stoichiometry of the ligand/dsDNA complex. For 1:1 and 2:1 stoichiometries, eqs (3) and (4) are employed to fit the isothermal binding curve, respectively. $[L]_{Bound}$ is the concentration of ligand bound to dsDNA. $\sum\Phi_f$ is the total fluorescence signal of the solution upon saturation of ligand binding sites. For 2:1 complexes, the equilibrium constants are given as the product of K_1K_2 if separation of the individual binding constants is not possible (due to similarity in magnitude).

$$F = \sum \Phi_f \frac{[L]_{Bound}}{n[DNA]_T} \quad (2)$$

$$F = \sum \Phi_f \left(\frac{K_1[L]_f}{1 + K_1[L]_f} \right) \quad (3)$$

$$F = \sum \Phi_f \left(\frac{0.5K_1[L]_f + K_1K_2[L]_f^2}{1 + K_1[L]_f + K_1K_2[L]_f^2} \right) \quad (4)$$

Competition assays were employed in an effort to determine K_1 values for **L4**, **L6**, **L7**, and in some cases **L1**. 5 nM dsDNA, in the presence of a large excess of **L8** (100–2000 nM) was titrated with ligand and decreases in the solutions fluorescence signal measured as **L8**-dsDNA complexes were replaced with ligand–dsDNA complexes possessing lower quantum yields of fluorescence. Plots were made of the negative change in fluorescence signal ($-\Delta F$) at 450 nm versus $[L]_f$, where $[L]_f$ was calculated according to eq (2) assuming $n = 1$ (see Fig. 4). The isothermal binding curves were fit employing eq (5).

$$-\Delta F = \sum \Phi_f \left(\frac{K_1[L]_f}{1 + K_1[L]_f + K_{Ht}[Ht]_f} \right) \quad (5)$$

The addition of **L6** causes a slight enhancement in the fluorescence emission of **L8** free in solution. Thus, the 'background' signal acquired from titration of **L8** (without the presence of dsDNA) with **L6** was subtracted from the agent's spectrofluorometric titration assays.

Model building

The molecular modeling program SYBYL was used to construct plausible ligand–dsDNA complexes. For the **L1**–dsDNA complex the agent was docked into the A/T rich minor groove of a model B-DNA helix. The dsDNA complexes of **L4** and **L6** were derived from X-ray crystal structure coordinates.[28]

Organic synthesis: general

^1H and ^{13}C NMR spectra were obtained on a Varian Unity Inova 400 spectrometer at 400 and 100 MHz, respectively. TLC was carried out on silica gel (Kieselger 60 F254) glass backed commercial plates and

visualized by UV light. Fast atom bombardment mass spectra, HRMS and LRMS, were obtained on a VG analytical, VG-70E double focusing mass spectrometer, with an Ion Tech Xenon Gun FAB source, and an OPUS/SIOS data interface and acquisition system. High-pressure liquid chromatography was accomplished using a Hewlett-Packard Series 1050 HPLC equipped with a diode array detector. For preparative separations an Alltech Macrosphere 300A, C8, silica, 7 μm, 250 × 10 mm reverse-phase column was used. For analytical separations an Alltech Macrosphere 300A, C18, silica, 7 μm, 250 × 4.6 mm reverse-phase column was used.

Solid-phase synthesis (SPS): general

SPS synthesis was accomplished using MBHA resin and standard manual solid-phase FMOC techniques.[29] Coupling reactions for **21** were accomplished using 1.6–2 equiv of **21**, 2 equiv HOBt, and 4 equiv of DIPEA in anhyd DMF and were run for 24 h. High coupling yields (70–100%) were measured by absorption at 290 nm of deprotected FMOC after resin was treated with a 20% piperidine/DMF solution. After each coupling, unreacted terminal amines were capped with a DMF solution of acetic anhydride and triethylamine. Coupling reactions for **22** (Scheme 1) were accomplished employing 2.6–3 equiv of **22**, 5.2–6 equiv PyBOP, 2 equiv HOBt, and 6 equiv DIPEA and were run for 24 h. Resin cleavage was accomplished in 2–4 h using a 95% TFA, 2.5% water, and 2.5% TIS solution. All final products synthesized via SPS were purified by HPLC chromatography (silica, reverse phase) with an increasing gradient of acetonitrile in 0.1% aq TFA solution. Product purity was checked by analytical HPLC analysis. Product was lyopholized and then reconstituted in a minimal amount of methanol. Product precipitated out of solution by the addition of diethyl ether followed by bubbling of the colloidal solution with HCl (g). Product was then collected via centrifugation as the HCl salt, reconstituted in H_2O, and lyopholized to dryness.

Pentafluorophenyl 1-[3-[*N*-(*tert*-butyloxycarbonyl)amino]propyl]-4-nitro-2-pyrrolecarboxylate (20). To 6 mL solution (3:2:1, 1 M $NaOH_{(aq)}$/dioxane/etoh) was added **18** (580 mg, 1.7 mmol). The solution was heated to 60 °C and stirred for 30 min. The disappearance of starting material **18** was monitored by TLC (silica, DCM). The solution was acidified with dilute HCl until a white precipitate appeared (pH ~ 2). Compound **19** was extracted out of the aq solution with ethyl acetate. The combined fractions of organic solvent were washed with water, dried over Na_2SO_4, and evaporated off to give a thick bright yellow paste/solution. The triethylamine salt of **19** has been previously characterized.[18] Without further purification, **19** was dissolved in 10 mL DMF to which was added DCC (421 mg, 2.0 mmol) and pentafluorophenol (626 mg, 3.4 mmol). The solution was stirred overnight at room temperature. Solvent was removed under vacuum and the remaining paste re-dissolved in diethyl ether after which reacted DCC (dicyclohexylurea) was removed by filtration. The diethyl ether was evaporated and the resulting crude product purified by flash chromatography (silica, DCM) to give **20** (767 mg, 92%, white powder). ^{1}H NMR ($CDCl_3$) δ 1.44 (s, 9H, *t*Boc), 2.02 (m, 2H, Ar–C–CH_2-C–), 3.19 (m, 2H, –CH_2–N–*t*Boc), 4.42 (t, J = 7.1 Hz, Ar–CH_2–), 4.75 (bm, 2H, –NH–*t*Boc), 7.77 (s, 1H, ArH), 7.94 (s, 1H, ArH); ^{13}C NMR ($CDCl_3$) δ 28.44 (-*t*Boc), 31.83 (Ar–C–*C*–C–), 37.49 (Ar–C–C–*C*–), 48.33 (Ar–*C*–C–C–), following 4 signals were detected in the aromatic region 116.31 + 118.95 + 129.43 + 136.19, 155.79 (–N–*C*(=0)–), 156.48 (–Ar–*C*(C=0)–). HRMS (ESI) 502.1017, M + Na^+ (502.1014 calcd for $C_{19}H_{18}F_5N_3O_6^+$Na).

Pentafluorophenyl 1-[3-[*N*-(*tert*-butyloxycarbonyl)amino]propyl]-4-[(9-fluorenylmethoxycarbonyl)amino]-2-pyrrolecarboxylate (21). Compound **20** (200 mg, 0.41 mmol) was dissolved in 30 mL 1:1 ethyl acetate/ethanol, 5 mL dioxane, and 2 mL saturated $NaHCO_3$. To this colloidal solution was added 25 mg of 10% palladium on carbon catalyst. The solution was stirred under an atmosphere of hydrogen gas (~1 atm) for 24 h. Reaction progress was monitored by TLC. The solution was filtered through Celite and the filtrate reduced to a volume of ~ 5–10 mL. To the reduced filtrate was then added 5 mL dioxane, 2 mL saturated $NaHCO_3$, and FMOC-Cl (127 mg, 0.49 mmol). The solution was stirred for 24 h. Product was extracted into diethyl ether, washed with water, and purified by flash chromatography (silica, 20:1 DCM/ethyl acetate) to give **21** (100 mg, 36%, white foam). ^{1}H NMR ($CDCl_3$) δ 1.43 (s, 9H, *t*Boc), 1.94 (m, 2H, Ar–C–CH2-C–), 3.1 (m, 2H,–CH_2–N–*t*Boc), 4.2–4.3 (m, 3H, Ar–CH_2– and Ph_2–CHR–), 4.51 (d, 2H,–CH_2OC(C=O)N–Py), 4.69 (bm, 2H,–NH–*t*Boc), 6.66 (s, 1H, ArH), 7.03 (s, 1H, ArH) 7.31 (m, 2H, FMOC ArH), 7.40 (t, 2H, FMOC ArH), 7.61 (d, 2H, FMOC ArH), 7.78 (d, 2H, FMOC ArH); ^{13}C NMR ($CDCl_3$) δ 28.56 (-*t*Boc), 32.02 (Ar–C–C–C–), 37.72 (Ar–C–C–C–), 46.96 + 47.33 (Ar–C–C–C–+–*C*OC(=O)N-Py), 67.28 (–*C*–COC(=O)N–Py), the following 6 signals were detected in the aromatic region 120.27 + 125.13 + 127.35 + 128.03 + 141.56 + 143.87, two signals belonging to carbonyl carbons were detected, a weak signal at 153.86 and a more intense signal at 156.25 ppm; HRMS (FAB) 671.203 (671.205 calcd for $C_{34}H_{30}F_5N_3O_6$).

L5. Synthesis employed MBHA rink amide resin (18 mg, 0.009 mmol loading sites). **L4** was collected as its HCl salt (0.0036 mmol, 40%, white powder). ^{1}H NMR (D_2O) δ 2.10–2.14 (two overlapping signals, 7 H, –C(C=O)–CH_3) + N_{py}–C–CH_2–C–), 2.92 (t, J = 7.02 Hz, 4 H, N_{py}–C–C–CH_2–), 4.37 (m, 4H, N_{py}–CH_2–C–C–), 6.79 (m, 1H, ArH), 6.86 (m, 1H, ArH), 7.22 (m, 1H, ArH), 7.27 (m, 1H, ArH). UV spectrum: λ_{max} = 305, λ_{min} = 262 nm; LRMS (ESI) 391 $(M + H)^+$.

L2. Synthesis employed MBHA rink amide resin (18 mg, 0.009 mmol loading sites). Product was collected as its HCl salt (0.0028 mmol, 31%, yellow powder). ^{1}H NMR (D_2O + DMSO–d_6) 1.80 + 1.92 (two sets of multiplets, 2H/multiplet, N_{py}–C–CH_2–C–N), 2.13 (m, 2H, Ph-O–C–CH_2–C–), 2.39 (bt, 2H, Ph–O–C–C–CH_2–), 2.67 + 2.80 (two sets of triplets, signal at 2.67 obscured by solvent, signal at 2.80 integrates for 2H, N_{py}–C–C–

CH_2–N), 3.00 (s, 3H, CH_3–NR_2), 3.1–3.3 (m, obscured by solvent, Ph–N(C–CH_2–)$_2$), 3.70 + 3.85 (two sets of broad signals, 6H, Ph–N(CH_2–)$_2$ + Ph–O–CH_2–), 3.98 + 4.18 (two sets of broad signals, 4H, N_{py}–CH_2–), signals detected between 6.5 and 8 ppm are due to ArH protons, 6.53 (d) + 6.77 (s) + 6.98 (m) + 7.04 (s) + 7.10 (s) + 7.26 (m) + 7.35–7.42 (m) + 7.56 (d) + 7.63 (s) + 7.75 (s); UV spectrum: $\lambda_{max} = 308$ with a distinctive shoulder at 350 nm, $\lambda_{min} = 280$ nm; LRMS (ESI) 841 $(M + H)^+$.

L4. Synthesis employed MBHA rink amide resin (21 mg, 0.010 mmol). Product was collected was as its HCl salt (0.0017 mmol, 17%, white product). **L5**, a byproduct of this reaction, was also purified and collected (0.0015 mmol, 15%, white powder). ^{1}H NMR (D_2O) δ 2.0 (two overlapping signals, 9 H,–C(C=O)–CH_3) + N_{py}–C–CH_2–C–), 2.78 (m, 6 H, N_{py}–C–C–CH_2–), 4.37 (bm, 6H, N_{py}–CH_2–C–C–), 6.92 (m, 2H, ArH), 7.04 (m, 1H, ArH), 7.30 (m, 1H, ArH), 7.38 (m, 1H, ArH), 7.41 (m, 1H, ArH). UV spectrum: $\lambda_{max} = 305$, $\lambda_{min} = 262$ nm; LRMS (ESI) 556 $(M + H)^+$.

L1. Synthesis employed MBHA rink amide resin (46 mg, 0.023 mmol). Product was collected as its HCl salt (0.0036 mmol, 16%, yellow powder). ^{1}H NMR (D_2O + DMSO-d_6) 1.95 + 2.08 (two sets of overlapping multiplets, 6H, N_{py}–C–CH_2–C–N + Ph–O–C–CH_2–C–), 2.46 (obscured by solvent, Ph–O–C–C–CH_2–), 2.72 (m, 6H, N_{py}–C–C–CH_2–N), 2.88 (s, 3H, CH_3–NR_2), 3.05 + 3.20 (two sets of multiplets, 2H/multiplet, Ph–N(C–CH_2–)$_2$), 3.55 + 3.89 (two sets of multiplets, 2H/multiplet, Ph–N(CH_2–)$_2$), 4.15 (obscured by HOD, Ar–O–CH_2–), 4.32 (broad multiplet, 6H, N_{py}–CH_2–), signals detected between 6.8 and 9 ppm are due to ArH protons, 6.86 (s) + 6.95 (s) + 7.05 (s) + 7.15 (bd) + 7.24 (s) + 7.28–7.38 (bm) + 7.39 (s) + 7.53 (t) + 7.71 (bd) + 7.80–7.98 (bm) + 8.07 (bd) + 8.49 (s); UV spectrum: $\lambda_{max} = 320$ with a slight shoulder at 355 nm, $\lambda_{min} = 278$ nm; LRMS (ESI) 1006 $(M + H)^+$.

Results

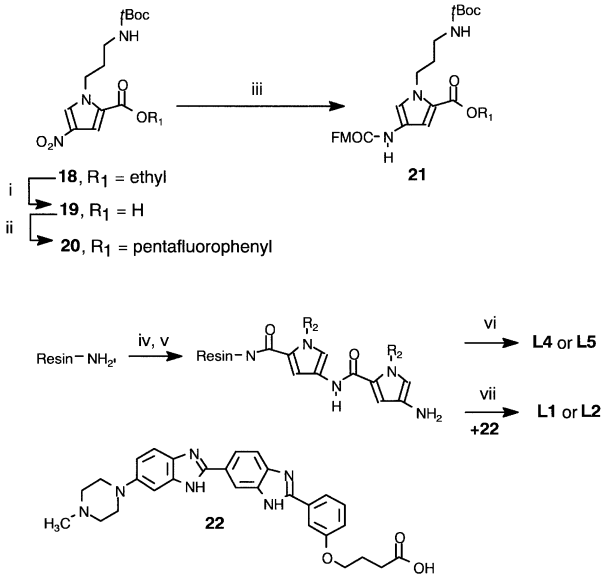
Synthesis

Synthesis of final products was accomplished in a stepwise manner from MBHA rink amide resin by employing FMOC chemistry and standard manual solid-phase synthetic techniques (Scheme 1).[29] It was previously found that *N*-methyl pyrrole carboxylic acid derivatives, analogous to **19**, do not reduce cleanly via catalytic hydrogenation.[13] However, *N*-methyl pyrrole pentafluorophenyl esters, analogous to **20**, do reduce cleanly without polymerization during catalytic hydrogenation or during carbamoylation with FMOC-Cl. Still, in the presence of HOBt catalyst these pyrrole pentafluorophenyl esters undergo coupling reactions via solid phase synthesis to form di- and tripyrrole polyamides. Thus, the activated monomer **21** was synthesized starting with the functionalized pyrrole **18**. The pyrrole **18** was hydrolyzed to give **19** and esterified to give **20**. Reduction of the nitro functional group of **20** followed by reaction of the free amine with FMOC-Cl gave the activated monomer **21**. Coupling yields between pyrrole units of **21** were observed to be between 70 and 100%. Pyrrole monomer coupling reactions were monitored by UV absorbance of the deprotected FMOC. The traditional Kaiser test is not compatible with the aromatic amine of the pyrrole ring.[30] Synthesis of **L2** and **L1** proceeded via coupling of **22** to the terminal amines of di- and tripyrrole polyamides, respectively. Capping of unreacted amines with acetic anhydride after each coupling reaction greatly simplified the purification of final products, particularly for **L1**. Coupling yields for the addition of **22** were estimated to be in the range of 30–40% by HPLC chromatography of resin cleaved product.

Apparent equilibrium constants for complexation of 1–6 were determined via melting curves of ligand–dsDNA complexes (Table 3)

Melting curves were acquired in 10 mM potassium phosphate pH 7.0 buffer at 10 ($\mu = 0.032$) or 150 ($\mu = 0.17$) mM NaCl concentrations. Representative melting curves for ligand–dsDNA complexes acquired at $\mu = 0.17$ are presented in Figure 1. Observed t_m values were converted to apparent equilibrium constants (K_1, M^{-1}) for DNA binding employing eq (1).[27]

L1 is shown to complex a nine bp A/T rich binding site at subpicomolar concentrations. The K_1 value for complexation of **2** by **L1** is 8×10^{12} M^{-1} at $\mu = 0.17$ (the oligomeric duplex **2** has the binding site -AAAAAAAAA-). Additionally, **L1** is shown to distinguish between the nine bp binding site of **2** and the shorter five bp A/T rich site of **1** (the oligomeric duplex **1** has the binding site -AAATT-). For **L1**, K_1 for complexation of **2** is 5000-fold greater than for **1**. In com-



Scheme 1. (i) NaOH, 60 °C, H_2O; (ii) pentafluorophenol, DCC; (iii) a. H_2(g) 1 atm, 10% Pd/C, b. FMOC-Cl, $NaHCO_3$; (iv) deprotection, 20% piperizine in DMF; (v) **21**, HOBt, DIPEA; (vi) a. capping with acetic anhydride, triethylamine, b. resin cleavage in 95% TFA; (vii) a. **22**, PyBOP, HOBt, DIPEA, b. resin cleavage.

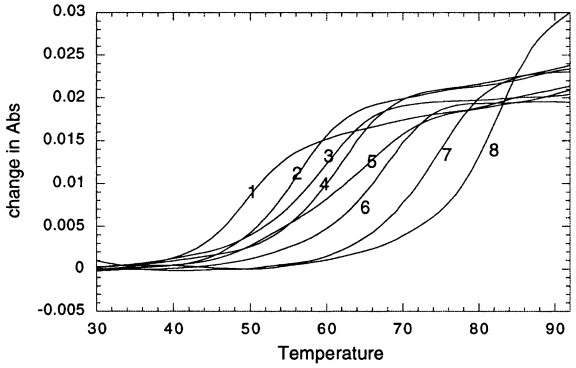


Figure 1. Thermal melting curves for oligomeric duplex **2** (0.3 μM) and its ligand complexes (0.6 μM ligand) (10 mM potassium phosphate pH 7.0 buffer, 150 mM NaCl (μ = 0.17)). Plots show change in absorption units versus °C. 1 is dsDNA with no ligand, 2 is dsDNA with **L8**, 3-**L7**, 4-**L6**, 5-**L4**, 6-**L3**, 7-**L2**, and 8-**L1**.

parison to classical minor groove binders, K_1 for complexation of **2** by **L1** is 5 orders of magnitude greater than for complexation by **L7** or **L8**. Also, as judged by K_1 values, neither **L7** or **L8** can distinguish between the 'small' five bp binding site of **1** and the 'large' nine bp site of **2**.

The propylamine chains of **L1** are shown to greatly enhance its K_1 values for DNA binding while not adversely affecting the sequence specificity of the agent. At μ = 0.17, K_1 for complexation of **2** by **L3** is roughly 3 orders of magnitude less than for complexation by **L1** (**L3** lacks the propylamine chains of **L1**). Also for **L3**, K_1 for complexation of **2** is 1700-fold greater than for **1**. Thus **L1** and **L3** possess the same magnitude of specificity for the longer A/T rich binding site of **2** relative to the 5 bp binding site of **1**.

L2 possesses one less pyrrole subunit than **L1**. The K_1 for complexation of **2** by **L2** is ~ 60-fold less than for complexation by **L1**. Also for **L2**, K_1 for complexation of **2** is 630-fold greater than for **1**, indicating that **L2** possesses less specificity than **L1**. The tripyrrole polyamide microgonotropen **L4** lacks the bisbenzimidazole moiety of **L1**. At μ = 0.17, K_1 for complexation of **2** by **L4** is 2100-fold less than for complexation by **L1** but still roughly 70-fold greater than for complexation by either of the classical minor groove binders **L7** or **L8**. **L4** shows only a slight preference for complexation of **2** relative to **1**, indicating that unlike **L1**, **L4** shows little preference for 'longer' A/T rich binding sites. However, **L4** is still specific for A/T rich binding sites relative to G/C rich dsDNA. For **L4** at μ = 0.032, K_1 for complexation of **5** is 24-fold greater than for **6**. The oligomeric duplex **5** contains an A/T rich binding site (-AATT-) whereas **6** does not contain an A/T rich binding site. **L5** possesses one less pyrrole subunit than **L4**. At μ = 0.17, K_1 for complexation of **1** by **L5** is equivalent to K_1 values for DNA binding by the classical minor groove binders **L7** and **L8**.

Decreasing the ionic strength from μ = 0.17 to μ = 0.032 increased the K_1 values for complexation of **2** by **L4** and **L6** by 210- and 270-fold, respectively. In contrast, the same change in ionic strength increases K_1 values for complexation of **2** by **L7** and **L8** to a lesser extent, 7- and 30-fold, respectively. Notably, at μ = 0.032 complexation of **2** by **L1** occurs at femtomolar concentrations (K_1 is 6×10^{13} M^{-1}). However, at this low ionic strength the K_1 for complexation of **2** by **L1** is only 50-fold greater than for complexation of **1**, compared to a 5000-fold difference when the ionic strength is μ = 0.17. Thus, the sequence specificity of **L1** is sensitive to changes in ionic strength.

Apparent equilibrium constants for complexation of 1–4 were determined via isothermal binding curves. Determination of K_1 values for **L1** was complicated by its very large magnitude. Even at nanomolar dsDNA concentrations, the lower limit for spectrofluorometric titrations, the titration of oligomeric duplexes [particularly **2** and **7** in 10 mM potassium phosphate buffer pH 7.0 with 150 mM NaCl (μ = 0.17) and 26 °C] with **L1** gave plots unusable for determination of K_1 (see Chart 2 for DNA sequences). This is because the plots showed a linear relationship between fluorescence intensity and ligand concentration up to the point of binding site saturation. However, the titrations did show that **L1** formed 1:1 complexes with both **2** and **7**.

A competition assay was conducted between **L8** and **L1** in an attempt to acquire a usable isothermal binding curve. We chose to employ **L8** in competition assays because it possesses a less intense fluorescence signal when free in solution than its more widely employed analogue **L9**.[31] Although both **L1** and **L8** fluorescence in the 400–500 nm range when bound to dsDNA, the quantum yield of fluorescence for **L8** is greater than that of **L1**. Thus, titration of dsDNA in the presence of a large excess of **L8** with **L1** leads to a decrease in emission signal. However, even in the presence of 400 equiv of **L8**, where the solution contained 5 nM **2**, 2000 nM **L8**, and μ = 0.17, the titration showed a linear relationship between fluorescence signal and **L1** concentration (Fig. 2). This linear relationship indicates that K_1 for complexation of **2** by **L1** is much greater than 400 times the K_1 for complexation of **2** by **L8**. Thus, K_1 for complexation of **2** must be much greater than 4×10^{11} M^{-1}. Attempts

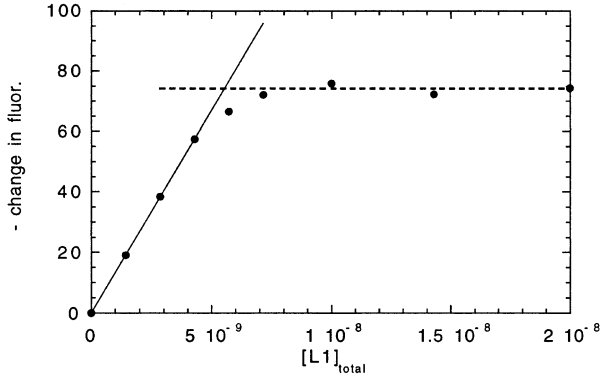


Figure 2. Titration of 5 nM **2** in the presence of 2000 nM **L8** with **L1** [10 mM potassium phosphate pH 7.0 buffer, 150 mM NaCl (μ = 0.17), 26 °C]. Plot shows the negative change in the fluorescence signal at 450 nm versus the concentration of **L1**. The two straight lines intersect at an **L1–2** stoichiometry of 1:1.

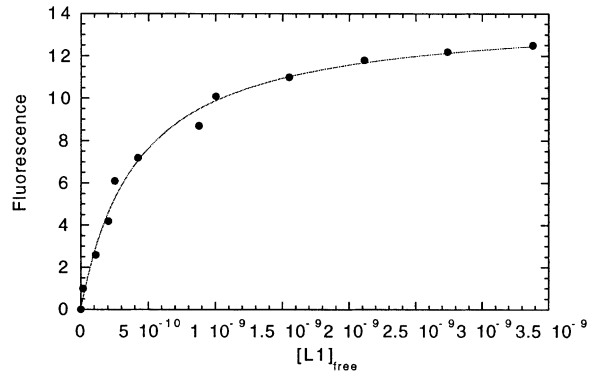


Figure 3. Titration of 2 nM **2** with **L1** in a solution of 75% 10 mM potassium phosphate pH 7.0 buffer, 1 M NaCl with 25% ethanol ($\mu = 0.77$), 26 °C. Fluorescence signal at 475 nm versus concentration of **L1** free in solution as calculated by eq (2). The data have been fit by a nonlinear least squares fitting routine employing eq (3) to determine the equilibrium constant for complex formation ($R^2 = 0.991$).

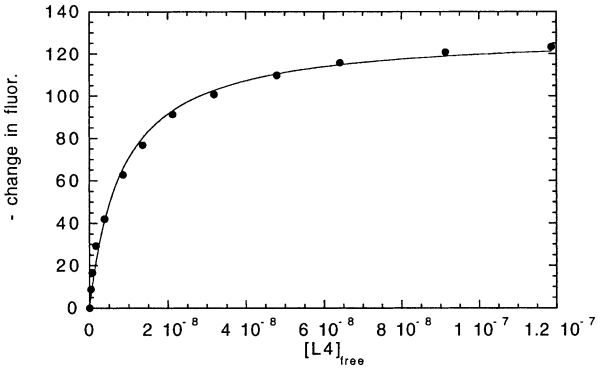


Figure 4. Titration of 5 nM **4** in the presence of 100 nM **L8** with **L4** ([0 mM potassium phosphate pH 7.0 buffer, 150 mM NaCl ($\mu = 0.17$), 26 °C]. Negative change in fluorescence signal at 450 nm versus concentration of **L4** free in solution as calculated by eq (2) with $n = 1$. The data have been fit by a non-linear least squares fitting routine employing eq (5) to determine the equilibrium association constant for complex formation ($R^2 = 0.993$).

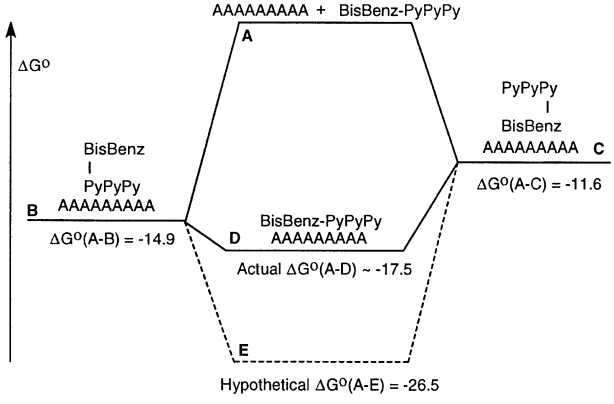


Figure 5. Schematic thermodynamic cycle showing the formation of the **L1/2** complex. ΔG° values (kcal) are calculated from equilibrium constants for DNA binding determined via spectrofluorometric titrations provided in Table 2 [10 mM potassium phosphate pH 7.0 buffer, 150 mM NaCl ($\mu = 0.17$), 26 °C]. (A) unbound **L1** (B) **2** bound by the tripyrrole moiety of **L1**. ΔG° estimated from K_1 for complexation of **1** by **L4**. (C) **2** bound by the bisbenzimidazole moiety of **L1**. ΔG estimated from K_1 for complexation of **1** by **L8**. (D) experimentally determined ΔG value for complexation of **2** by **L1**. (E) Hypothetical ΔG value for noncooperative complexation of **2** by **L1**. The difference in ΔG° between D and E is the magnitude of the ligand's negative cooperativity.

were made to conduct competition assays employing even larger excesses of **L8** in an attempt to generate a usable isothermal binding curve, however, **L8** slowly precipitates out of solution at greater concentrations.

High salt concentrations and the presence of ethanol are both known to weaken DNA–**L9** interactions.[32] A 'high salt–25% ethanol' solution where $\mu = 0.77$ (75% 10 mM potassium phosphate pH 7.0 buffer, 1 M NaCl and 25% ethanol) was found to lessen K_1 values for DNA binding by **L1** to magnitudes measurable via spectrofluorometric titration (Fig. 3 and Table 4). K_1 values at 26 °C are provided in Table 2. The extent to which the 'high salt–25% ethanol' solution ($\mu = 0.77$) decreases K_1 values for **L1** relative to 'normal' pH 7.0 buffer where $\mu = 0.17$ was estimated by determining K_1 values for **L4** in both solutions (the tripyrrole moiety of **L1**). Upon going from the 'normal' to the 'high salt–25% ethanol' solution, the K_1 for complexation of **1** by **L4** decreased by 2030-fold. Applying this value to **L1**, the K_1 for complexation of **2** by **L1** in 'normal' pH 7.0 buffer where $\mu = 0.17$ is estimated to be $\sim 5 \times 10^{12}$ M^{-1}, roughly 3 to 4 orders of magnitude greater than for complexation by **L8** and **L9**. Also, in 'normal' buffer the K_1 for complexation of **1** by **L4** was found to be ~260-fold greater than for complexation by **L6**, **L7**, or **L8** (see Figure 4 for a representative competition assay).

The two minor groove binding moieties of **L1** may complex dsDNA with positive or negative cooperativity with respect to one another. In an attempt to characterize the cooperativity between the two moieties, thermodynamic cycles such as shown in Figure 5 were constructed. Thermodynamic cycles at 26 °C were constructed for complexation of **2** in both 'normal' pH 7.0 buffer where $\mu = 0.17$ (cycle presented in Fig. 5) and the 'high salt–25% ethanol solution' where $\mu = 0.77$ (cycle not shown). Free energies of dsDNA binding for the tripyrrole moiety of **L1** were estimated via K_1 values for complexation of **1** by **L4** (Table 4). Similarly, free energies of binding for the bisbenzimidazole moiety were estimated via K_1 values for complexation of **1** by **L8**. Further, the ability of two minor groove binding molecules to simultaneously occupy the minor groove of **2** was investigated by the determination of K_1 and K_2 values for complexation of **2** by **L8** and **L9**. For both ligands, K_1 was roughly two orders of magnitude greater than K_2.

K_1 values for complexation of **3** (-AAATT-) and **4** (-TATAA-) by **L6** and **L8** are also provided in Table 2. The five bp A/T rich binding sites of **3** and **4** are representative of either side of the nine bp A/T rich binding site of **7** (for dsDNA sequences see Chart 2). **L6** and **L8** show the same type of selectivity, with the magnitude of their K_1 values for DNA binding in the order **3** > **1** > **4**. Both had K_1 values for complexation of **3** two orders of magnitude greater than for complexation of **4**. Interestingly though, the K_1 values derived at elevated temperatures via thermal melting curves (Table 3) suggest little or no selectivity by either ligand. It is reported that the stacking of A-tracts is disrupted by temperature (see the discussion section for a brief description of

A-tracts), and that A-tracts undergo a pre-melting transition around 30–37 °C to a structure little different from that of 'normal' B-DNA (see ref 33 and references therein). Thus, it would appear that the effect of 5-TpA-3 steps on minor groove binding cannot be detected via analysis of thermal melting curves when the DNA duplexes employed have t_m^0 values much greater than 30 °C. The ligand complexes of **3** and **4** melt at temperatures between 40 and 50 °C (see Table 1 for t_m values).

Apparent equilibrium constants for complexation of 7–17 by L1 were determined via melting curves of L1–dsDNA complexes (Table 5). The ability of **L1** to distinguish between a nine bp A/T rich binding site, as for **7**, and related binding sites containing a single A/T→G/C bp substitution, as for oligomeric duplexes **8**–**16**, was investigated. Apparent K_1 values are derived from **L1**–dsDNA complex melting curves employing eq (1) (see Table 5 for K_1 values, t_m values for **L1**–dsDNA complexes are provided in the supporting information). Apparent K_1 values decrease substantially, 20- to 200-fold, for single bp mismatch duplexes **9**–**14**. Substitutions at either end of the A/T rich binding site (**8**, **15**, and **16**) have little effect on K_1. The double bp substitution in **17** causes a 180-fold decrease in K_1.

Equilibrium constants for complexation of 7 by L1 and L3 were determined via isothermal binding curves (Fig. 3)

As mentioned previously, K_1 values for **L1**–dsDNA complexes where determined in a 'high salt–25% ethanol' solution where $\mu = 0.77$ (75% 10 mM potassium phosphate pH 7.0 buffer, 1 M NaCl and 25% ethanol) because K_1 values in a 'normal' phosphate pH 7.0 buffer where $\mu = 0.17$ were too large to measure via spectrofluorometric titrations. K_1 values for complexation of **7** by **L1** and **L3** are provided in Table 4. K_1 for complexation of **7** by **L1** in 'normal' pH 7.0 buffer where $\mu = 0.17$ is estimated to be 1×10^{13} M^{-1} in the same manner as described earlier for estimation of K_1 for complexation of **2** by **L1**. Direct comparison of equilibrium constants for complexation of **7** by **L1** and its analogue **L3** (**L3** lacks the propylamine chains of **L1**) is slightly complicated due to differences in the binding stoichiometries of the two agents. **L1** and **L3** form 1:1 and 2:1 complexes with **7**, respectively.[13] However, K_1 values for complexation of **7** by **L3** in units M^{-1} can be approximated by taking the square root of the multiplied through K_1K_2 M^{-2} values. Thus in 'normal' pH 7.0 buffer where $\mu = 0.17$, K_1, in units M^{-1}, for complexation of **7** by **L3** is determined to be ~5200-fold less than for complexation of **7** by **L1**. Thus, the three propylamine chains of **L1** increase the agent's free energy of binding by approximately 5.1 kcal.

Construction of ligand–dsDNA models. A plausible structure for a **L1**–dsDNA complex was constructed and its binding site size measured to be 9 bp (Fig. 6). Models for **L4** and **L6** dsDNA complexes were also constructed (Fig. 7). In this case, the basic side chains of the ligands are shown reaching out of the minor groove. The longer side chain of **L6** is capable of reaching a greater distance beyond the phosphodiester backbone than the shorter propylamine side chains of **L4**. When the propylamine chains of **L4** are fully extended, the N–P distances between the propylamine nitrogens and the DNA phosphorous atoms are at most 6–7 Å. Also, the three propylamine chains of **L4** are spaced at intervals equivalent to the P–P distances of the DNA backbone. Alternatively, N–P distances for **L6** average ~11 Å (for the three nitrogen atoms N1, N2, and N3 as labeled in Chart 1). Also, relatively short distances between these three nitrogen atoms prohibits them from simultaneously associating with three different phosphate groups.

Discussion

Spectrofluorometric and thermal denaturation experiments were employed to investigate stoichiometries and equilibrium constants (K_1) for formation of ligand–dsDNA complexes. **L1** is observed to form 1:1 complexes with oligomeric duplexes **2** and **7** at subpicomolar concentrations (see Chart 2 for DNA sequences). Both **2** and **7** contain binding sites of nine contiguous A/T base pairs. Also, **L1** possesses the ability to distinguish between its preferred nine bp A/T rich binding sites and shorter sites with fewer than nine contiguous A/T bases pairs. For **L1**, the K_1 for complexation of **2** is 5000-fold greater than for complexation of **1** whereas the classical minor groove binders **L7** and **L8** demonstrate no significant preference for either duplex. The specificity of **L1** was further investigated employing the oligomeric duplexes **8**–**16**, each of which contains a single A/T → G/C bp substitution within the -TATAAAATT- binding site of **7** (Table 5). Single bp substitutions at either termini of the A/T rich binding site of **7** had little effect on the agents K_1 values. However, substitutions elsewhere (for duplexes **9**–**14**) led to significant 20- to 200-fold drops in K_1.

The propylamine chains of **L1** increase its free energy of binding (ΔG°) by about 5.1 kcal as calculated from K_1 values for complexation of **7** by **L1** and **L3** (Table 4). The overall effect of the propylamine chains is to strengthen **L1**–dsDNA interactions without a loss of sequence specificity. For example, K_1 values for both molecules are three orders of magnitude less for complexation of the 5 bp A/T rich site of **1** than the 9 bp site of **2** (see Table 3). Initially the development of minor groove binders capable of recognizing longer dsDNA binding sites was hindered because longer molecules failed to conform to the curvature of the minor groove.[8] This problem was at least partially overcome by connecting 'rigid' minor groove binding segments, such as polypyrroles or bisbenzimidazole units with flexible linkers such as for extended polyamides[3,5] or the polyamide-bisbenzimidazole conjugate **L3.** In the case of **L3**, model building demonstrated how the molecule could adapt a 'spiral-like' conformation which matched that of the DNA minor groove.[13] Figure 6 shows a plausible structure for an **L1**–dsDNA complex where it too is shown to adapt to the curvature of the minor groove.

ΔG° values, calculated from the K_1 values presented in Table 4, were employed to characterize the nature of binding cooperativity between the tripyrrole and bisbenzimidazole moieties of **L1**. As shown in Figure 6, the two moieties of **L1** are estimated to possess a negative cooperativity of 9 kcal. It is assumed that the free energy of binding for either moiety of **L1** is equivalent to that for complexation of **1** by either **L4** or **L8**. Thus, the ΔG° for complexation of **2** by the tripyrrole peptide or bisbenzimidazole moieties of **L1** are estimated to be -14.9 and -11.6 kcal, respectively, at 26 °C and an ionic strength of $\mu = 0.17$. Hypothetically, if no cooperativity existed between the two moieties of **L1**, an **L1**–**2** complex would possess a ΔG° of -26.5 kcal and a K_1 of 1×10^{19} M^{-1} (represented by a dotted line in Fig. 5). Instead, the complex is measured to have a ΔG° of 7.5 kcal and a K_1 of 5×10^{12} M^{-1}.

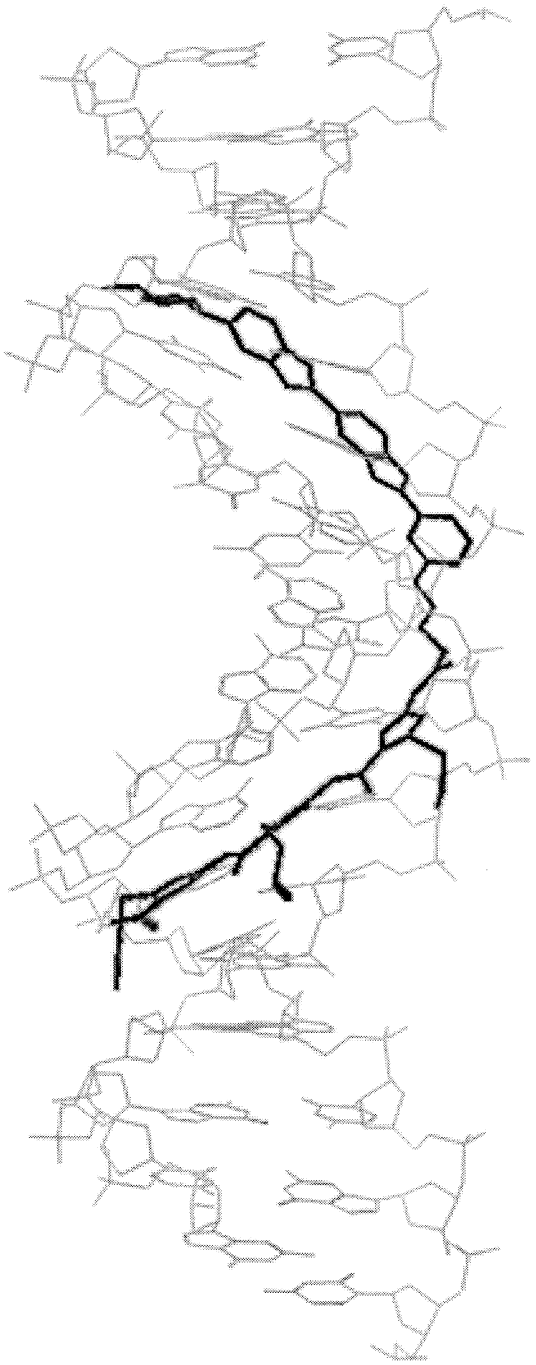


Figure 6. Computer generated model of proposed **L1**–dsDNA complex.

At first one might be quick to assume that the negative cooperativity demonstrated between the two moieties of **L1** is due to a poorly designed linker which prevents the agent's moieties from correctly registering with the DNA bases. After all, this is the problem most often ascribed to longer or extended minor groove binders.[8] However, that formation of the $(\mathbf{L8})_2$–**2** complex also occurs with negative cooperativity suggests otherwise. K_1 for complexation of **2** by **L8** is two orders of magnitude less than K_2 for complexation of **2** by a second molecule of **L8** (Table 2). Similar results are observed for **L9**. It seems that the nine bp binding site of **2** is simply not large enough for two **L8** molecules to complex in a linear end-to-end manner. Judging only by length, two molecules of **L8** should be capable of fitting within a nine bp site. Still, complexation of the first molecule may cause conformational changes in the duplex which adversely affects binding by the second. Another possibility is that the molecule's preferred binding site exists in the middle of the nine bp binding site of **2**. Complexation by two molecules then forces both to either end of the A/T rich sequence.

Our results suggest that an optimized molecule similar to **L1** could possess an equilibrium constant as large as 10^{19}–10^{20} if the source of negative cooperativity between the two moieties could be removed. The source of negative cooperativity appears to be inherent in the targeting of a nine base pair binding site — that is, the two minor groove binding moieties need to be spaced further apart and the molecule directed towards a longer A/T rich binding site.

We were interested in determining the effect of 5′-TpA-3′ dinucleotide steps on the minor groove binding moieties of **L1**. Particularly, we wished to determine if the binding orientation of **L1** in its complex with **7** could be predicted from the relative selectivities of its two minor groove binding moieties — that is, does **L1** exclusively bind the -TATA- region of **7** with its tripyrrole or bisbenzimidazole moiety. Towards this end, K_1 values were determined for complexation of **3** and **4** by **L4** and **L8**. Oligomeric duplexes **3** and **4** contain the A/T rich binding sites -TATAA- and -AAATT-, respectively, similar to either half of the nine bp site of **7**. It has been suggested that 5′-TpA-3′ dinucleotide steps produce a DNA structural alteration which discourages minor groove binding.[31,34,35] Dickerson and co-workers have reported that A/T rich regions of dsDNA usually adopt poly(dA)/poly(dT)-like structures (referred to as 'A-tracks').[33,36] A-tracks consist of successive adenine bases which stack in a rigid and unbent column. Emphasized is that the A-track structure can incorporate 5′-ApT-3′ steps, but is broken by 5′-TpA-3′ steps. For both **L4** and **L6**, K_1 values for complexation of **3** were roughly 100-fold greater than for complexation of **4** (Table 2). The specificity of the bisbenzimidazole and tripyrrole moieties of **L1** for an A-track appears to be

equivalent. Thus, neither moiety of **L1** likely binds either region of **7**'s binding site, exclusively.

Substitution of the tripyrrole moiety of **L1** with a dipyrrole yields **L2** and a 30-fold decrease in K_1 for complexation of **2**. The specificity possessed by **L2** for the nine bp A/T rich site of **2** relative to the five bp site of **1** remains equivalent to that of **L1** (Table 3). **L4** is the tripyrrole polyamide moiety of **L1**. The K_1 for complexation of **2** by **L4** is 2100-fold less then for complexation by **L1**. Also, **L4** possess little specificity for longer A/T rich binding sites. Comparison of K_1 values derived via spectrofluorometric titrations show **L4** to possess a K_1 for complexation of **1** at least 260-fold greater than for the classical minor groove binders **L7** and **L8** and 360-fold greater than for the fellow microgonotropen **L6** (Table 4).

Molecular models are employed in an effort to explain why the K_1 for complexation of **1** by **L4** is 260-fold greater than for complexation by **L6**. In Figure 7, the basic side chains of both ligands are shown fully extended and reaching out from the minor groove. Still, the basic amines of **L4** are in close proximity to DNA phosphate groups on either side of the minor groove (N–P distances are at most 6–7 Å). Additionally, the nitrogen atoms of **L4** are spaced such that each propylamine chain may simultaneously form a salt bridge with a different phosphate group. In contrast, the polyamine side chain of **L6** is much longer with N–P distances averaging ~11 Å (for N1, N2, and N3 as labeled in Chart 1). Further, N1, N2, and N3 are not spaced at distances equal to the DNA phosphate groups. Interestingly though, **L6** is still as potent an inhibitor of TF binding as **L4**.[23] This may be because the long polyamine chain of **L6** wraps around the phosphodiester backbone and reaches into the major groove[17] where it may sterically hinder TF binding. In contrast, the shorter propylamine chains of **L4** are not long enough to reach that far.

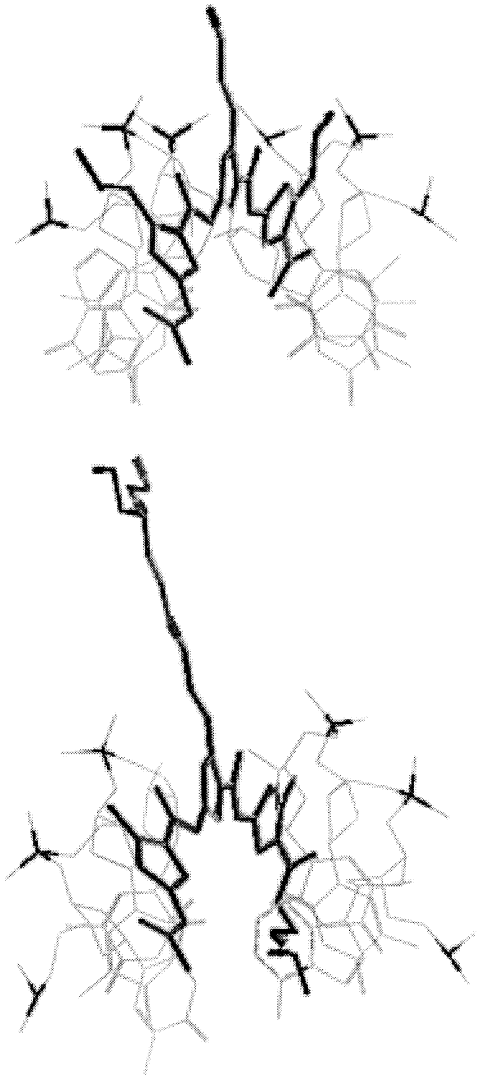


Figure 7. Computer generated models of **L4** (above) and **L6** (below) 1:1 complexes with dsDNA. **L4** and **L6** are shown bound within the minor groove. The positively charged chains of both agents are shown fully extended and reaching out of the minor groove. Only relevant DNA bases are shown.

Acknowledgements

This work was supported by a grant from the National Institute of Health (5R37DK09171-36).

References and Notes

1. Reddy, B. S. P.; Sondhi, S. M.; Lown, J. W. *Pharmacol. Ther*. **1999**, *84*, 1.
2. Chiang, S. Y.; Bruice, T. C.; Azizkhan, J. C.; Gawron, L.; Beerman, T. A. *Proc. Natl. Acad. Sci. U.S.A*. **1997**, *94*, 2811.
3. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc*. **1998**, *120*, 3534.
4. Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc*. **1996**, *118*, 6160.
5. Kissinger, K. L.; Dabrowiak, J. C.; Lown, J. W. *Chem. Res. Tox*. **1990**, *3*, 162.
6. Chen, Y. H.; Yang, Y. W.; Lown, J. W. *J. Biomol. Struct. Dyn*. **1996**, *14*, 341.
7. Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc*. **1997**, *119*, 7636.
8. Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A*. **1996**, *93*, 6981.
9. Singh, M. P.; Plouvier, B.; Hill, G. C.; Gueck, J.; Pon, R. T.; Lown, J. W. *J. Am. Chem. Soc*. **1994**, *116*, 7006.
10. Beerman, T. A.; Woynarowski, J. M.; Sigmund, R. D.; Gawron, L. S.; Rao, K. E.; Lown, J. W. *Biochim. Biophys. Acta* **1991**, *1090*, 52.
11. Filipowsky, M. E.; Kopka, M. L.; Brazilzison, M.; Lown, J. W.; Dickerson, R. E. *Biochemistry* **1996**, *35*, 15397.
12. Lown, J. W.; Krowicki, K.; Balzarini, J.; Newman, R. A.; Declercq, E. *J. Med. Chem*. **1989**, *32*, 2368.
13. Satz, A. L.; Bruice, T. C. *J. Am. Chem. Soc*. **2001**, *123*, 2469.
14. Zimmer, C.; Wähnert, U. *Prog. Biophys. Mol. Biol*. **1986**, *47*, 31.
15. He, G. X.; Browne, K. A.; Groppe, J. C.; Blasko, A.; Mei, H. Y.; Bruice, T. C. *J. Am. Chem. Soc*. **1993**, *115*, 7061.
16. He, G. X.; Browne, K. A.; Blasko, A.; Bruice, T. C. *J. Am. Chem. Soc*. **1994**, *116*, 3716.
17. Blasko, A.; Browne, K. A.; Bruice, T. C. *J. Am. Chem. Soc*. **1994**, *116*, 3726.

18. Xue, T. H.; Browne, K. A.; Bruice, T. C. *Bioconjugate Chem.* **1995**, *6*, 82.
19. Bruice, T. C.; Sengupta, D.; Blasko, A.; Chiang, S. Y.; Beerman, T. A. *Bioorg. Med. Chem.* **1997**, *5*, 685.
20. Satz, A. L.; Bruice, T. C. *Bioorg. Med. Chem.* **2000**, *8*, 1871.
21. Satz, A. L.; Bruice, T. C. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 3261.
22. White, C. M., Gawron, L., Satz, A. L., Bruice, T. C., Beerman, T. A. *Biochim. Biophys. Acta* submitted for publication.
23. White, C. M., Satz, A. L., Bruice, T. C., Beerman, T. A. *Proc. Natl. Acad. Sci. U.S.A.*, in press.
24. White, C. M., Beerman, T. A. unpublished results.
25. Latchman, D. S. *Gene Regulation: A Eukaryotic Perspective, 3rd ed.*; Stanley Thornes: Cheltenham, 1998.
26. Marky, L. A.; Breslauer, K. *J. Biopolymers* **1987**, *26*, 1601.
27. Crothers, D. M. *Biopolymers* **1971**, *10*, 2147.
28. Coll, M.; Frederick, C. A.; Wang, A. H.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8385.
29. Fields, G. B.; Noble, R. L. *Intl. J. Pep. Prot. Res.* **1990**, *35*, 161.
30. Sarin, V. K.; Kent, S. B. H.; Tam, J. P.; Merrifield, R. B. *Anal. Biochem.* **1981**, *117*, 147.
31. Satz, A. L.; White, C. M.; Beerman, T. A.; Bruice, T. C. *Biochemistry* **2001**, *40*, 6465.
32. Loontiens, F. G.; Regenfuss, P.; Zechel, A.; Dumortier, L.; Clegg, R. M. *Biochemistry* **1990**, *29*, 9029.
33. Dickerson, R. E.; Goodsell, D.; Kopka, M. L. *J. Mol. Biol.* **1996**, *256*, 108.
34. Ward, B.; Rehfuss, R.; Goodisman, J.; Dabrowiak, J. C. *Biochemistry* **1988**, *27*, 1198.
35. Chen, F. M.; Sha, F. *Biochemistry* **1998**, *37*, 11143.
36. Goodsell, D. S.; Kaczorgrzeskowiak, M.; Dickerson, R. E. *J. Mol. Biol.* **1994**, *239*, 79.