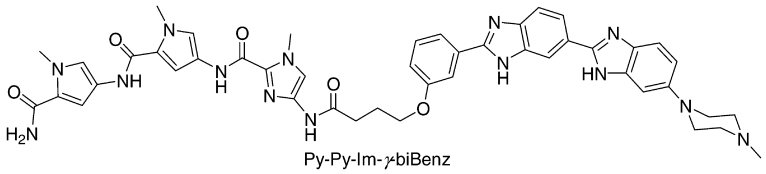
Article

# Sequence Selective Recognition in the Minor Groove of dsDNA by Pyrrole, Imidazole-Substituted Bis-benzimidazole Conjugates

Putta Mallikarjuna Reddy, Peter T. Jindra, Alexander L. Satz, and Thomas C. Bruice

*J. Am. Chem. Soc.*, **2003**, 125 (26), 7843-7848• DOI: 10.1021/ja035116k • Publication Date (Web): 10 June 2003






## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 06/10/2003

# Sequence Selective Recognition in the Minor Groove of dsDNA by Pyrrole, Imidazole-Substituted Bis-benzimidazole Conjugates

Putta Mallikarjuna Reddy, Peter T. Jindra, Alexander L. Satz, and Thomas C. Bruice*

*Contribution from the Department of Chemistry and Biochemistry, University of California at Santa Barbara, Santa Barbara, California 93106*

Received March 12, 2003; E-mail: tcbruice@bioorganic.ucsb.edu

***Abstract:*** A series of pyrrole, imidazole-substituted bis-benzimidazole conjugates, Py-Py-Im-$\gamma$-biBenz, Py-Py-$\gamma$-biBenz, Py-Im-$\gamma$-biBenz, and Im-Py-$\gamma$-biBenz (**1**−**4**), were prepared in an attempt to target dsDNA sequences possessing both A/T and G/C bps. The dsDNA interactions and sequence specificity of the conjugates have been characterized via spectrofluorometric titrations and thermal melting studies. All conjugates form 1:1 complexes with dsDNA at subnanomolar concentrations. The Im moiety selectively recognizes a G/C bp embedded in the A/T-rich binding site. This represents the first clear example of sequence selective recognition in a 1:1 motif.[1] The equilibrium association constant ($K_1$) for complexation of a specific nine-bp dsDNA site, 5′-gcggTATGAAATTcgacg-3′, by conjugate **1** is $\sim 2.6 \times 10^9$ $M^{-1}$. Displacement of the G/C position or G/C→A/T substitution within the nine-bp site decreases the $K_1$ by ∼8-fold, whereas two continuous G/C bps decrease the $K_1$ by ∼50-fold magnitude. The $K_1$ values for seven-bp dsDNA, 5′-gcggtaTGAAATTcgacg-3′ and 5′-gcggtaCAAAATTcgacg-3′, binding sites by conjugates Py-Im-$\gamma$-biBenz (**3**) and Im-Py-$\gamma$-biBenz (**4**) are $\sim 2.3 \times 10^9$ and $\sim 1.2 \times 10^9$ $M^{-1}$, respectively. However, the conjugates with no Im moiety, Py-Py-$\gamma$-biBenz (**2**) and Py-Py-Py-$\gamma$-biBenz (**5** and **6**), are specific for seven- to nine-bp A/T-rich sites and single A/T→G/C bp substitution within the binding site decreases the $K_1$ values by 1−2 orders of magnitude.

## Introduction

The incorrect expression or biochemical function of specific genes, whose protein products are essential for the regulation of cell growth, causes cancer.[2] By designing a drug capable of recognizing specific sequence in DNA, it may be possible to specifically inhibit a specific gene's expression and thereby control the growth of tumor cells. Studies indicate that low molecular weight minor groove binding agents may bind to specific sequences of DNA and influence gene expression by inhibiting the binding of transcription factors.[3−9] Interest in the control of specific gene's expression has spurred efforts in the development of new minor groove binding agents. It is expected that those agents capable of recognizing longer DNA sequences will exhibit the greatest specificity; however those agents targeted for longer sequences have generally lacked binding specificity.[10−20]

In search for potential minor groove binders (MGBs) targeting longer sequences, we have recently reported novel tripyrrole-Hoechst (*m*-hydroxy Hoechst derivative) conjugates (**5**, **6**) that can specifically recognize longer A/T base pair (bp) sequences in the minor groove of double strand (ds) DNA.[21,22] The tripyrrole-Hoechst conjugate **5** binds to nine-bp long A/T-rich sites at subnanomolar concentrations and complexes its targets at near diffusion-controlled rates.[21] The tripyrrole-Hoechst

(1) Kissinger, K.; Krowicki, K.; Dabrowiak, J. C.; Lown, J. W. *Biochemistry* **1987**, *26*, 5590.
(2) Latchman, D. S. *Gene regulation: a eukaryotic perspective*, 3rd ed.; Stanley Thornes: Cheltenham, 1998.
(3) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215.
(4) Reddy, B. S. P.; Sondhi, S. M.; Lown, J. W. *Pharmacol. Ther.* **1999**, *84*, 1.
(5) Bailly, C.; Chaires, J. B. *Bioconjugate Chem.* **1998**, *9*, 513.
(6) Neidle, S. *Nat. Prod. Rep.* **2001**, *18*, 291.
(7) Chiang, S.-Y.; Bruice, T. C.; Azizkhan, J. C.; Gawron, L.; Beerman, T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2811.
(8) He, G. X.; Browne, K. A.; Blasko, A.; Bruice, T. C. *J. Am. Chem. Soc.* **1994**, *116*, 3716.
(9) Browne, K. A.; He, G. X.; Bruice, T. C. *J. Am. Chem. Soc.* **1993**, *115*, 7072.
(10) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534.
(11) Kissinger, K. L.; Dabrowiak, J. C.; Lown, J. W. *Chem. Res. Toxicol.* **1990**, *3*, 162.
(12) Chen, Y. H.; Yang, Y. W.; Lown, J. W. *J. Biomol. Struct. Dyn.* **1996**, *14*, 341.
(13) Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 7636.
(14) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160.
(15) Singh, M. P.; Plouvier, B.; Hill, G. C.; Gueck, J.; Pon, R. T.; Lown, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 7006.
(16) Beerman, T. A.; Woynarowski, J. M.; Sigmund, R. D.; Gawron, L. S.; Rao, K. E.; Lown, J. W. *Biochim. Biophys. Acta* **1991**, *1090*, 52.
(17) Filipowsky, M. E.; Kopka, M. L.; Brazilzison, M.; Lown, J. W.; Dickerson, R. E. *Biochemistry* **1996**, *35*, 15397.
(18) Lown, J. W.; Krowicki, K.; Balzarini, J.; Newman, R. A.; Declercq, E. *J. Med. Chem.* **1989**, *32*, 2368.
(19) Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981.
(20) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B. *Nat. Struct. Biol.* **1998**, *5*, 104.
(21) Satz, A. L.; Bruice, T. C. *J. Am. Chem. Soc.* **2001**, *123*, 2469.
(22) Satz, A. L.; Bruice, T. C. *Bioorg. Med. Chem.* **2002**, *10*, 241.

conjugate was very specific for its A/T-rich binding site. A single A/T→G/C bp substitution led to substantial drops in the binding affinity. Substitution of the *N*-methyl groups of **5** with an *N*-linked alkylamine chain, which is protonated at neutral pH, provides the microgonotropen **6**. Most proteins bind to dsDNA via the major groove. The protonated and positively charged *N*-propylamines of the conjugate **6** reach out of the minor groove and electrostatically interact with the DNA phosphodiester linkages. This greatly increases the binding affinity of the MGB and inhibits protein−dsDNA interactions at picomolar concentrations.[22] Additionally, these MGBs were found to inhibit protein expression in whole cell assays. These compounds are capable of traversing the NIH 3T3 cell membrane and binding to its nuclear DNA target.[23,24] These findings are significant because they represent the first MGBs with the most important characteristics of cell penetration and inhibition of protein expression. However, the current scaffold is limited to targeting only longer A/T-rich target sites. To achieve the goal of specifically inhibiting the expression of a misregulated protein, minor groove binding agents must be developed with the ability to recognize G/C bps within their otherwise A/T-rich binding sites. A first step in that direction calls for the synthesis of imidazole-containing molecules, which have been previously shown to tolerate and possibly prefer a G/C bp.[3−6] For example, polyamides composed of *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) have been reported to bind in antiparallel 2:1 or "hairpin motifs" in the dsDNA minor groove with predictable sequence selectivity and high binding affinities.[3−6]

Although these *N*-methylimidazole-possessing MGBs have shown great promise, they still lack characteristics necessary for practical use as a drug. Immediate problems include a lack of sequence specificity, high molecular weight, and 2:1 stoichiometries of binding. While "short" Im-polyamides are sequence specific,[1] elongated Im-polyamides targeted to longer dsDNA sequences are not.[19,20] The 2:1 binding stoichiometries of the Im-polyamides are also a drawback. Drugs rarely (or perhaps never) complex their physiological binding sites with stoichiometries other than 1:1. Hairpin polyamides also suffer from a lack of specificity when elongated to target longer dsDNA binding sites.[19,20] Additionally, the hairpin polyamides are large molecules, with molecular weights twice that of their analogous 2:1 binding polyamides. As a general rule, greater molecular weight correlates with poor physicochemical properties.[25,26]

Herein we report the study of pyrrole, imidazole-substituted bis-benzimidazole (biBenz) conjugates (Figure 1), with $\gamma$-hydroxybutyric acid linker ($\gamma$), capable of recognizing G/C bp in the nine-bp specific sequences while forming 1:1 ligand:dsDNA complexes. The fluorescence emission of these conjugates increases greatly when bound in the minor groove of dsDNA. Hence, spectrofluorometric titrations were employed to determine the ligand:dsDNA stoichiometries and binding affinities. The sequence selectivity of each conjugate was determined by investigating the binding affinities to the 18-bp dsDNA by gradually changing the G/C position in the nine-bp minor groove site.

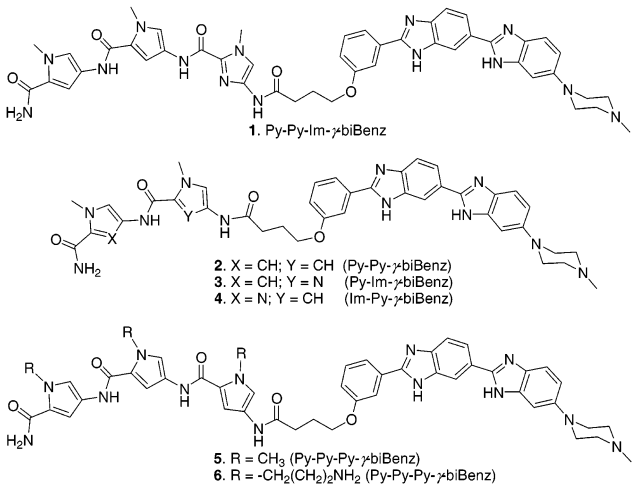



***Figure 1.*** Structures of pyrrole, imidazole-biBenz conjugates **1**−**6**.

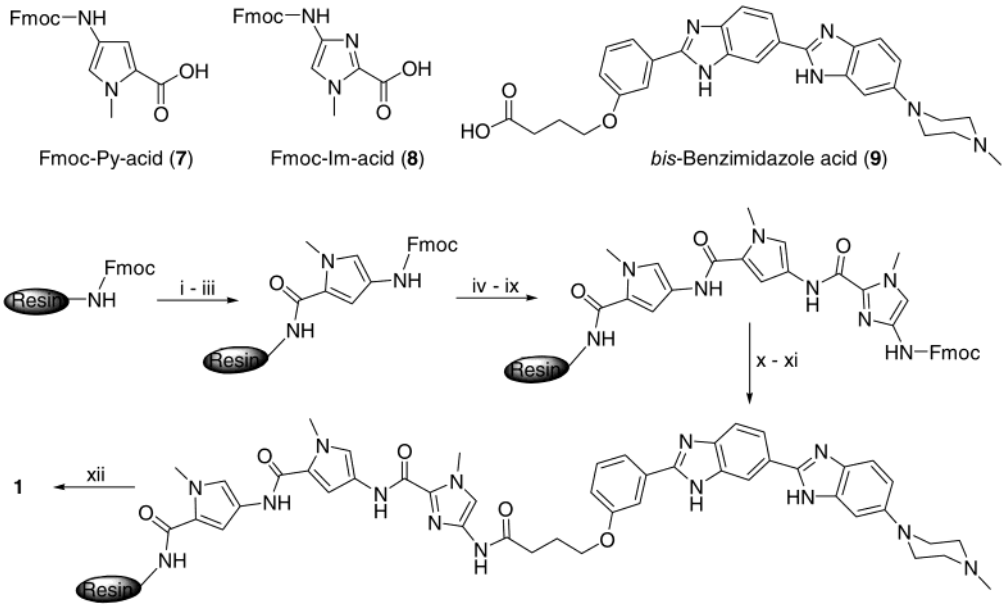
## Experimental Section

**Materials.** All solvents and reagents including Hoechst (Ht) 33258 were purchased from Aldrich and used without further purification. Rink amide MBHA resin and reagents used in the solid-phase synthesis were purchased from Novabiochem. Purified DNA oligomers were purchased from the Biomolecular Resource Center, University of California at San Francisco. All NMR spectra were recorded on a Varian 400 MHz instrument. ESI/TOF+ mass spectral analysis was performed on Micromass Q-Tof-2 quadrapole instrument. HPLC analysis was performed using a Hewlett-Packard Series 1050 instrument equipped with a diode array detector. Alltech Macrosphere 300A, C8, silica 7 $\mu$m, 250 mm × 10 mm (for preparative separations), and 250 mm × 4.6 mm (for analytical separations) reverse phase columns were used. The monomers, Fmoc-Py-acid (**7**),[27] Fmoc-Im-acid (**8**),[27] and substituted bis-benzimidazole acid (**9**),[28] were synthesized as described in the literature.

**Solid-Phase Synthesis of Py-Py-Im-$\gamma$-biBenz Conjugate (1).** The solid-phase synthesis of **1** was accomplished using rink amide MBHA resin (50 mg, 0.025 mmol loading sites) and standard manual solid-phase Fmoc techniques (Scheme 1). Coupling reactions were accomplished in 12 h using 3 equiv of **7** or **8**, 3 equiv of HOBt, 3 equiv of PyBOP, and 6 equiv of DIPEA in anhydrous DMF. After each coupling, unreacted terminal amine sites were capped with acetic anhydride/TEA solution in DMF. After each cycle, the terminal Fmoc was deprotected by treatment with a 20% piperidine/DMF solution. Coupling yields, in each cycle, between 90 and 100% were determined by the absorption of deprotected Fmoc at 290 nm. Final coupling with 2.5 equiv of **9** was accomplished in 24 h using 3 equiv of HOBt, 3 equiv of PyBOP, and 6 equiv of DIPEA in anhydrous DMF. Resin cleavage was achieved in 2 h using TFA containing 1% of TIS solution. The crude product was purified by reversed-phase HPLC using a C8 column with an increasing gradient of acetonitrile in 0.1% aqueous TFA solution. After purification, the product was reconstituted in a minimal amount of methanol and precipitated out of solution by the addition of ether. Product purity was checked by analytical RP-HPLC using the same solvent system. Yield was determined by NMR spectroscopy, using a known quantity of the internal standard 3-(trimethylsilyl)propionic-2,2,3,3-$d_4$ acid, sodium salt, to be 15 mg (17 $\mu$mol, 68% yield): $^1$H NMR (DMSO-$d_6$) $\delta$ 2.11 (m, 2H, $-CH_2CH_2-$

(23) White, C. M.; Satz, A. L.; Bruice, T. C.; Beerman, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10590.
(24) Satz, A. L.; Bruice, T. C. *Acc. Chem. Res.* **2002**, *35*, 86.
(25) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. *J. Med. Chem.* **2002**, *45*, 2615.
(26) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **1997**, *23*, 3.
(27) Wurtz, N. R.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Org. Lett.* **2001**, *3*, 1201.
(28) Satz, A. L.; Bruice, T. C. *Bioorg. Med. Chem.* **2000**, *8*, 1871.

***Scheme 1.*** Solid-Phase Synthesis of Conjugate **1** Using Rink Amide MBHA Resin[a]




[a] (i) *Deprotection*: 20% piperidine/DMF, 10 min; (ii) *Coupling*: Fmoc-Py-acid (**7**), PyBOP, HOBt, DIPEA, DMF, 12 h; (iii) *Capping*: acetic anhydride, TEA, DMF, 10 min; (iv) 20% piperidine/DMF, 10 min; (v) Fmoc-Py-acid, PyBOP, HOBt, DIPEA, DMF, 12 h; (vi) acetic anhydride, TEA, DMF, 10 min; (vii) 20% piperidine/DMF, 10 min; (viii) Fmoc-Im-acid (**8**), PyBOP, HOBt, DIPEA, DMF, 12 h; (ix) acetic anhydride, TEA, DMF, 10 min; (x) 20% piperidine/DMF, 10 min; (xi) bis-benzimidazole acid (**9**), PyBOP, HOBt, DIPEA, DMF, 24 h; (xii) *Cleavage*: TFA, TIS, 2 h. Conjugates **2**−**4** are also synthesized in a similar procedure.

OPh), 2.56 (t, $J$ = 7 Hz, 2H, −C$H_2$OPh), 2.90 (s, 3H, C$H_3$−N$R_2$), 3.24 (m, piperazine C$H_2$), 3.62 (m, piperazine C$H_2$), 3.80 (s, 3H, C$H_3$−NPy), 3.84 (s, 3H, C$H_3$−NPy), 3.96 (s, 3H, C$H_3$−NIm), 4.14 (t, $J$ = 6 Hz, 2H, −NHC(=O)C$H_2$−), signals detected between 7 and 9 ppm are due to Ar−$H$, Py−$H$, and Im−$H$, 6.85 (s, 1H), 7.02 (s, 1H), 7.13 (m, 2H), 7.21 (m, 2H), 7.27 (m, 2H), 7.51 (m, 2H), 7.69 (d, $J$ = 9 Hz, 1H), 7.82 (m, 2H), 7.88 (d, $J$ = 9 Hz, 1H), 8.05 (d, $J$ = 9 Hz, 1H), 8.46 (s, 1H), 9.15 (bs, 2H, N$H_2$−(C=O)−Py), 9.91 (s, 1H, −N$H$−(C=O)−Py), 9.97 (s, 1H, −N$H$−(C=O)−Im), 10.36 (s, 1H, −N$H$−(C=O)−$CH_2$−); ESI/TOF+ analysis exhibited the expected peaks at $m/z$ 877.50 (M + H) and 439.27 (M + 2H); calcd 877.40 (M + H) and 439.20 (M + 2H) for $C_{46}H_{48}N_{14}O_5$.

**Py-Py-$\gamma$-biBenz (2).** The solid-phase synthesis of **2** was accomplished using MBHA rink amide resin (45 mg, 0.0225 mmol loading sites). Yield was determined by NMR spectroscopy to be 13.5 mg (18 $\mu$mol, 80% yield): $^1$H NMR (DMSO-$d_6$) $\delta$ 2.10 (m, 2H, −C$H_2$-$CH_2$OPh), 2.57 (t, 2H, −C$H_2$OPh), 2.91 (s, 3H, C$H_3$−N$R_2$), 3.25 (m, piperazine C$H_2$), 3.81 (s, 3H, C$H_3$−NPy), 3.84 (s, 3H, C$H_3$−NPy), piperazine ring $CH_2$ signal hidden by HOD absorption, 4.15 (t, $J$ = 6 Hz, 2H, −NHC(=O)C$H_2$−), signals detected between 7 and 9 ppm are due to Ar−$H$, and Py−$H$, 6.81 (s, 1H), 7.02 (s, 1H), 7.11 (m, 2H), 7.17 (m, 2H), 7.51 (m, 4H), 7.69 (s, 1H), 7.88 (d, $J$ = 9 Hz, 1H), 8.29 (d, $J$ = 9 Hz, 1H), 9.84 (s, 1H, −N$H$−(C=O)−Py), 10.03 (s, 1H, −N$H$−(C=O)−$CH_2$−); ESI/TOF+ analysis exhibited the expected peaks at $m/z$ 754.35 (M + H); calcd 754.35 (M + H) for $C_{41}H_{43}N_{11}O_4$.

**Py-Im-$\gamma$-biBenz (3).** The solid-phase synthesis of **3** was accomplished using MBHA rink amide resin (45 mg, 0.0225 mmol loading sites). Yield was determined by NMR spectroscopy to be 12 mg (16 $\mu$mol, 73% yield): $^1$H NMR (DMSO-$d_6$) $\delta$ 2.08 (m, 2H, −C$H_2$-$CH_2$OPh), 2.56 (−C$H_2$OPh), 2.89 (s, 3H, C$H_3$−N$R_2$), 3.21 (m, piperazine C$H_2$), 3.54 (m, piperazine C$H_2$), 3.79 (s, C$H_3$−NPy), 3.94 (s, C$H_3$−NIm), 4.13 (t, $J$ = 6 Hz, 2H, −NH−C(=O)C$H_2$−), signals detected between 7 and 9 ppm are due to Ar−$H$, Py−$H$, and Im−$H$, 6.90 (s, 1H), 7.13 (m, 2H), 7.23 (m, 2H), 7.50 (m, 2H), 7.68 (d, $J$ = 9 Hz, 1H), 7.85 (m, 2H), 8.05 (d, $J$ = 9 Hz, 1H), 8.44 (s, 1H), 8.99 (bs, 2H, N$H_2$−(C=O)−Py), 9.83 (s, 1H, −N$H$−(C=O)−Im), 10.39 (s, 1H, −N$H$−(C=O)−$CH_2$−); ESI/TOF+ analysis exhibited the expected peaks at $m/z$ 755.40 (M + H) and 378.20 (M + 2H); calcd 755.35 (M + H) and 378.18 (M + 2H) for $C_{40}H_{42}N_{12}O_4$.

**Im-Py-$\gamma$-biBenz (4).** The solid-phase synthesis of **4** was accomplished using MBHA rink amide resin (45 mg, 0.0225 mmol loading sites). Yield was determined by NMR spectroscopy to be 4 mg (5.3 $\mu$mol, 24% yield): $^1$H NMR (DMSO-$d_6$) $\delta$ 2.08 (m, 2H, −C$H_2$-$CH_2$OPh), 2.51 (−C$H_2$OPh, hidden by DMSO peaks), 2.91 (s, 3H, C$H_3$−N$R_2$), 3.83 (s, C$H_3$−NPy), 3.92 (s, C$H_3$−NIm) piperazine ring $CH_2$ signals are hidden by HOD absorption, 4.15 (t, $J$ = 6 Hz, 2H, −NH−C(=O)C$H_2$−), signals detected between 7 and 9 ppm are due to Ar−$H$, Py−$H$, and Im−$H$, 6.93 (s, 1H), 6.96 (s, 1H), 7.09 (m, 2H), 7.15 (m, 2H), 7.22 (m, 2H), 7.52 (m, 2H), 7.81 (m, 2H), 8.05 (d, $J$ = 9 Hz, 1H), 8.43 (s, 1H), 8.95 (bs, 2H, N$H_2$(C=O)Im), 9.93 (s, 1H, −N$H$−(C=O)−Py), 10.34 (s, 1H, −N$H$−(C=O)−$CH_2$−); ESI/TOF+ analysis exhibited the expected peaks at $m/z$ 755.38 (M + H) and 378.21 (M + 2H); calcd 755.35 (M + H) and 378.18 (M + 2H) for $C_{40}H_{42}N_{12}O_4$.

**Spectrofluorometric Titrations.** Fluorescent titrations allowed the determination of ligand:dsDNA stoichiometries and binding constants. Oligomeric duplexes were prepared as described earlier,[21] and fluorescence spectra were obtained on a Perkin-Elmer LS50B fluorophotometer equipped with a constant-temperature water bath set at 25 °C. Solutions in 10 mM potassium phosphate buffer pH 7.0 containing 150 mM NaCl ($\mu$ = 0.17) were excited at 345 nm, and emissions were monitored at 445 or 475 nm. Spectrofluorometric titrations were obtained by titrating a constant concentration (5 nM) of dsDNA with a relatively concentrated solution of ligand. The two straight lines were generated by linear least-squares fitting of data points early and late in the titration (Figure 2). The point of intersection of the two lines over the $x$-axis provides the stoichiometry of the ligand:dsDNA complex.[28]

Following determination of ligand:dsDNA complex stoichiometry, equilibrium association constants ($K_1$) for ligand:dsDNA complexation were investigated by generating isothermal binding curves[21,22,28] (Figure 3). $K_1$ values for 1:1 ligand:dsDNA complexes were calculated by fitting isothermal binding curves using eqs 1−3. Equation 1 was employed to fit plots of fluorescence versus concentration of unbound ligand, $[L]_f$, where $[L]_f$ is calculated by eq 3 and eq 3 is derived from eq 2.

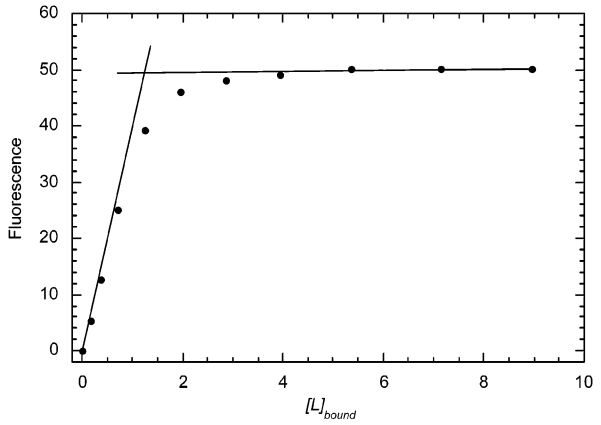



***Figure 2.*** Titration of 5 nM dsDNA **14** with conjugate **1**. Plot shows relative fluorescence intensity at 475 nm in arbitrary units vs ligand **1**:dsDNA ratio. The two straight lines were generated by linear least-squares fitting of data points early and late in the titration. The point of intersection of the two lines over the *x*-axis provides the stoichiometry of the ligand:dsDNA complex.

The derivation and use of eqs 1−3 have been previously discussed by our laboratory.[9,28]

$$F = \sum \Phi f\left(\frac{K_1[\mathrm{L}]_\mathrm{f}}{1 + K_1[\mathrm{L}]_\mathrm{f}}\right) \quad (1)$$

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

$$[\mathrm{L}]_\mathrm{f} = [\mathrm{L}]_\mathrm{T} - \frac{n[\mathrm{DNA}]_\mathrm{T}F}{\sum \Phi_\mathrm{f}} \quad (3)$$

In eq 1, $\sum\Phi_\mathrm{f}$ is the total fluorescence intensity upon saturation of dsDNA binding sites with ligand, $K_1$ is the equilibrium association constant for the first binding event, and $[\mathrm{L}]_\mathrm{f}$ is the concentration of ligand free in solution. In eq 2, $[\mathrm{L}]_\mathrm{Bound}$ is the concentration of ligand bound to dsDNA, $n$ is the stoichiometry of binding, and $[\mathrm{DNA}]_\mathrm{T}$ is the total concentration of duplex DNA in the sample. In eq 3 $[\mathrm{L}]_\mathrm{T}$ is the total ligand concentration added to the constant concentration of dsDNA.

***T*m Studies.** The thermal stability of ligand:dsDNA complexes were investigated by thermal denaturation experiments. UV/vis spectra were acquired on a Cary 100 Bio UV−vis spectrophotometer equipped with a temperature programmable cellblock. All $T_\mathrm{m}$ experiments were carried out using 10 mM potassium phosphate buffer, pH 7.0, containing 150 mM NaCl ($\mu = 0.17$). For each oligomeric duplex (3 $\mu$M), a series of $T_m$ curves were acquired (dsDNA + no ligand; dsDNA + Ht33258; dsDNA + **1** or **2** or **3** or **4**) using 1 equiv of ligand. Data points were taken, between 5 and 95 °C, for every 1 °C with a temperature ramp of 0.2 °C min$^{-1}$, and $T_\mathrm{m}$ temperatures were calculated by first-derivative analysis.

## Results

The pyrrole, imidazole-substituted bis-benzimidazole conjugates, **1**−**4**, were synthesized manually on rink amide MBHA resin by employing Fmoc chemistry using a series of PyBOP/HOBt-mediated coupling reactions as described in Scheme 1. The monomers Fmoc-Py-acid (**7**),[27] Fmoc-Im-acid (**8**),[27] and bis-benzimidazole acid (**9**)[28] were synthesized as described in the literature and were used as unactivated acids instead of the HOBt and HOAt derivatives.[29,30] The in situ activation and coupling reactions of the monomers on solid phase were achieved in the presence of PyBOP/HOBt in higher yields than obtained by HOBt or HOAt derivatives. However, this method did not overcome the poor coupling yields (~25%) of pyrrole onto an imidazole unit.[29]

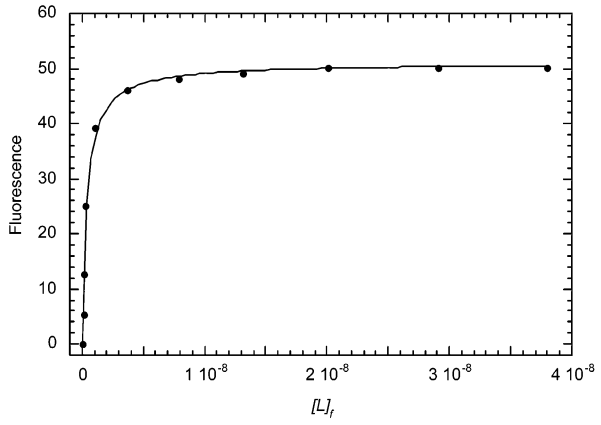



***Figure 3.*** Isothermal binding curve generated by spectrofluorometric titration of 5 nM dsDNA **14** with conjugate **1**. Plot shows relative fluorescence intensity in arbitrary units vs concentration of free ligand **1** as calculated by eq 3, and data points were fit using eq 1.

***Table 1.*** Equilibrium Association Constants for Complexation ($K_1 \times 10^{-7}$ M$^{-1}$)[a]

| dsDNA | 1 | 2 | 3 | 4 | Ht |
|---|---|---|---|---|---|
| 5′-gcggTATAAAATTcgacg-3′ (**10**) | 35 | **310** | 90 | 28 | 72 |
| 5′-gcgg**C**ATAAAATTcgacg-3′ (**11**) | 20 | **290** | 55 | 30 | |
| 5′-gcggT**G**TAAAATTcgacg-3′ (**12**) | 21 | **250** | 30 | 35 | 48 |
| 5′-gcggTA**C**AAAATTcgacg-3′ (**13**) | 25 | 55 | 65 | **120** | |
| 5′-gcggTAT**G**AAATTcgacg-3′ (**14**) | **260** | 10 | **230** | 10 | 42 |
| 5′-gcggTATA**G**AATTcgacg-3′ (**15**) | 35 | 9 | 15 | 12 | |
| 5′-gcggTATAA**G**ATTcgacg-3′ (**16**) | 8 | 4 | 9 | 6 | 23 |
| 5′-gcggTATAAA**G**TTcgacg-3′ (**17**) | 12 | 22 | 25 | 18 | |
| 5′-gcggTATAAAA**C**TcgacgTcgacg-3′ (**18**) | 7 | 110 | 12 | 21 | 30 |
| 5′-gcggTATAAAAT**C**cgacg-3′ (**19**) | 15 | 120 | 15 | 24 | |
| 5′-gcggTATA**GG**AATTcgacg-3′(**20**) | 5 | 2 | 6 | 4 | 61 |

[a] The dsDNA binding site is typed in caps. Ht binds 2:1 in the binding site of all dsDNA, and the $K_1$ values are the square root of $K_1K_2$. The stated equilibrium constants have a standard deviation of ±60%.

The ligand:dsDNA complex stoichiometries and equilibrium association constants ($K_1$) for conjugates **1**−**4** were determined by fluorescence titrations. The fluorescence emission of ligand increases greatly upon being complexed by dsDNA. When excited at 345 nm, ligand:dsDNA complexes emit a broad fluorescence signal centered at 475 nm. This signal is red shifted by ~30 nm with respect to Ht33258:dsDNA complexes, which emit a broad fluorescence signal centered at 445 nm.[8,9,31−34] All of the oligomeric duplexes investigated (Table 1) were observed to form 1:1 ligand:dsDNA complexes with conjugates **1**−**4**, whereas Ht33258 forms a 2:1 $(\mathrm{Ht33258})_2$:dsDNA complex.[21] The $K_1$ values for Ht33258:dsDNA complexation listed in Table 1 are the square root of $K_1K_2$. Figure 2 shows a representative example of a plot used to determine **1**:dsDNA (**14**) complex stoichiometry.

The equilibrium association constants for complexation of 11 different dsDNAs by conjugates **1**−**4** were determined as

(29) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.
(30) Fattori, D. Kinzel, O.; Ingallinella. P.; Bianchi, E.; Pessi, A. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1143.
(31) Loontiens, F. G.; Regenfuss, P.; Zechel, A.; Dumortier, L.; Clegg, R. M. *Biochemistry* **1990**, *29*, 9029.
(32) Bostock-Smith, C. E.; Searle, M. S. *Nucl. Acids Res.* **1999**, *27*, 1619.
(33) Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chaires, J. B. *J. Mol. Biol.* **1997**, *271*, 244.
(34) Loontiens, F. G.; McLaughlin, L. W.; Diekmann, S.; Clegg, R. M. *Biochemistry* **1991**, *30*, 182.

***Table 2.*** Melting Temperatures ($T_m$) for Ligand:dsDNA Complexes (°C)[a]

| dsDNA | $T^0_m$ | $\Delta T_m$ 1 | 2 | 3 | 4 | Ht |
|---|---|---|---|---|---|---|
| 5′-gcggTATAAAATTcgacg-3′ (**10**) | 65 | 6 | **8** | 4 | 5 | 3 |
| 5′-gcgg**C**ATAAAATTcgacg-3′ (**11**) | 64 | 4 | **8** | 4 | 4 | 3 |
| 5′-gcgg**T**GTAAAATTcgacg-3′ (**12**) | 65 | 5 | **8** | 5 | 4 | 3 |
| 5′-gcggTA**C**AAAATTcgacg-3′ (**13**) | 64 | 4 | 5 | 4 | 7 | 3 |
| 5′-gcggTAT**G**AAATTcgacg-3′ (**14**) | 65 | **9** | 3 | **8** | 3 | 3 |
| 5′-gcggTATA**G**AATTcgacg-3′ (**15**) | 65 | 3 | 4 | 5 | 2 | 2 |
| 5′-gcggTATAA**G**ATTcgacg-3′ (**16**) | 64 | 4 | 2 | 2 | 1 | 2 |
| 5′-gcggTATAAA**G**TTcgacg-3′ (**17**) | 65 | 3 | 4 | 4 | 2 | 2 |
| 5′-gcggTATAAAA**C**TTcgacg-3′ (**18**) | 65 | 4 | 5 | 3 | 2 | 2 |
| 5′-gcggTATAAAAT**C**cgacg-3′ (**19**) | 65 | 4 | 3 | 5 | 1 | 2 |
| 5′-gcggTATA**GG**AATTcgacg-3′(**20**) | 67 | 2 | 0 | 0 | 0 | 2 |

[a] All $T_m$ studies were carried out in potassium phosphate buffer, pH 7.0, containing 150 mM NaCl, and $T_m$ values were determined by first derivative analysis. $T^0_m$ are melting temperatures of dsDNA in the absence of ligand. $\Delta T_m$ are differences in melting temperatures for dsDNA in the presence and absence of ligand. Standard deviations for $\Delta T_m$ values are ±1 °C. DNA duplex binding site is typed in caps.

shown in Table 1. Figure 3 shows a plot used for the determination of $K_1$ for **1**:dsDNA (**14**) complexation. Conjugate **1** recognizes the nine-bp site, and it is highly selective for the G/C containing nine-bp site 5′-gcggTATGAAATTcgacg-3′ (dsDNA **14**). Displacement of the G/C position or substitution of G/C with A/T bp in the nine-bps site, 5′-gcggTATGAAAT-Tcgacg-3′, causes a significant decrease in $K_1$ by ∼40-fold. Additionally, two continuous G/C bps in the center of the binding site (dsDNA **20**) cause further decrease in $K_1$ by ∼50-fold.[21] However, Ht33258 shows no such selectivity (Table 1). For example, $K_1$ values for the complexation of Ht33258 with dsDNA **10**−**20** are nearly equal, whereas $K_1$ values for **1** with dsDNA differ by greater than 50-fold.

Conjugates **2**−**4**, with different combinations of Py and Im moieties, were synthesized to study the sequence selectivity in the minor groove of dsDNA. Conjugate **2**, with no Im moiety, exhibited very little change in $K_1$ when the minor groove length was reduced from a nine- to a seven-bp A/T-rich binding site (dsDNA **10**−**12**). Further decrease of the minor groove length is accompanied by a dramatic change in the $K_1$ (Table 1) of conjugate **2**, indicating that it is selective for a seven-bp A/T-rich site. Substitution of one of the Py moieties with Im, conjugates **3** and **4**, allows recognition of G/C in the seven-bp site. Conjugates **3** and **4** are specific for 5′-gcggtaTGAAAT-Tcgacg-3′ and 5′-gcggtaCAAAATTcgacg-3′ sites, respectively. Similar to conjugate **1**, conjugates **3** and **4** exhibited significant change in $K_1$ when G/C is substituted by A/T or the G/C position is changed in the binding site (Table 1). The $K_1$ values for **2** are affected only by A/T to G/C substitution within the central region of the nine-bp A/T-rich binding site. Alternatively, **3** and **4** (both of which possess an Im moiety) bind most tightly to oligomers that possess a G/C bp located near the center of the nine-bp binding site. So, the two sequences (**13** and **14**) disfavored by conjugate **2** are the preferred binding sites for **3** and **4**. Conjugate **3** binds to dsDNA **14** with a $K_1$ of $2.3 \times 10^9$, and conjugate **4** binds to dsDNA **13** with a $K_1$ of $1.2 \times 10^9$.

As shown in Table 2, the thermal melting data strongly support the data acquired via the equilibrium assays. Conjugate **1** forms a significantly more stable complex with dsDNA **14** than Ht33258 (Figure 4). The effect of **1** on $\Delta T_m$ (the difference between $T_m$ values for oligomeric duplexes in the presence and absence of ligand) decreases greatly upon change in the G/C position or substitution of G/C with A/T bp in the binding site (Table 2). In agreement with $K_1$ data, $\Delta T_m$ values were affected more strongly when the two continuous G/C bps are present in the center of the binding site (dsDNA **20**) (Figure 5). Again, conjugates **3** and **4** most favor those sequences possessing internal G/C bps in otherwise A/T-rich binding sites (dsDNA **14** and **13**) by exhibiting a higher $\Delta T_m$ of 3−4 °C. Conjugate **2**, with no Im moiety, exhibited no change in $\Delta T_m$ when minor groove length was reduced from nine to seven bps (dsDNA **10**−**12**). Alternatively, **2** disfavors the internal G/C substitution, and Ht33258 exhibited no such selectivity for all the oligomers investigated (Table 2).

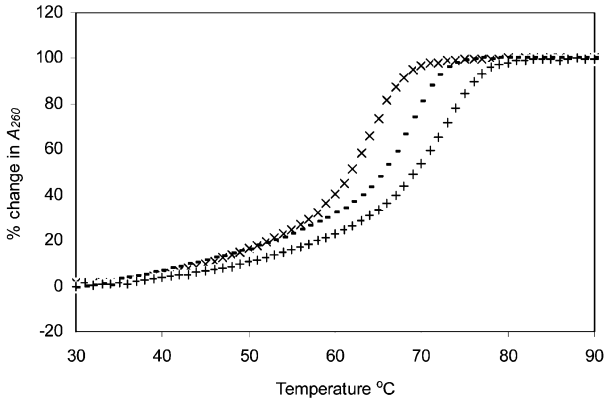



***Figure 4.*** Normalized $T_m$ curves for dsDNA **14** ($c = 3\ \mu$M) with 1 equiv of ligand: (×) dsDNA with no ligand, (−) dsDNA + Hoechst 33258, and (+) dsDNA + conjugate **1**.

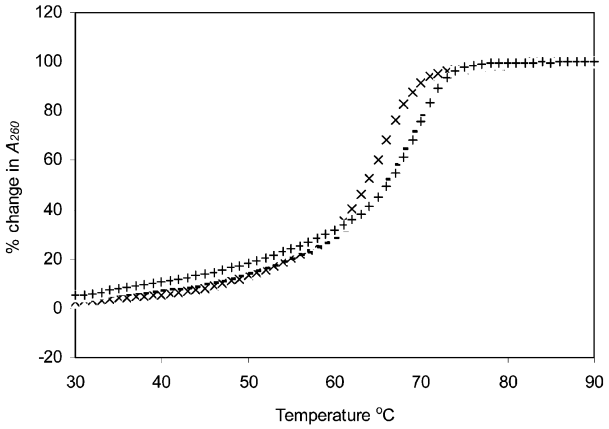



***Figure 5.*** Normalized $T_m$ curves for dsDNA **20** ($c = 3\ \mu$M) with 1 equiv of ligand: (×) dsDNA with no ligand, (−) dsDNA + Hoechst 33258, and (+) dsDNA + conjugate **1**.

## Discussion

A series of pyrrole, imidazole-substituted bis-benzimidazole (*m*-hydroxy Hoechst derivative) conjugates have been synthesized and used to study the sequence selective recognition in the minor groove of dsDNA. The ligand:dsDNA stoichiometry and sequence specificity of the conjugates have been characterized via spectrofluorometric titrations and thermal melting studies. All conjugates are observed to form 1:1 complexes with all dsDNA used in this study at subnanomolar concentrations. The observed 1:1 ligand:dsDNA complex stoichiometries suggest the formation of complexes as illustrated in Figure 6. In contrast in $(Ht33258)_2$:dsDNA complex, each Ht33258 molecule

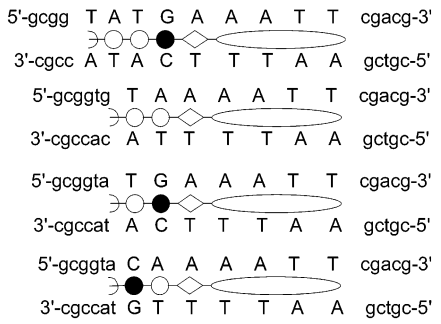



***Figure 6.*** Schematic illustration of binding of conjugates **1**−**4** with preferred dsDNA. Binding site typed in caps. Im and Py rings are represented as filled and open circles, respectively. $\gamma$-linker and biBenz are represented as diamond and elongated circle, respectively.

resides in a 4 dA/dT bp region such that two Ht33258 molecules are arranged end-to-end and do not overlap each other.[35,36]

The tripyrrole-biBenz conjugate (**5**) binds to a nine-bp A/T-rich site in 2:1 stoichiometry.[21] Substitution of one of the Py residues by an Im residue (conjugate **1**) results in a 1:1 stoichiometry of ligand:dsDNA complex (Figure 2). This is most interesting. This is probably due to the lack of cooperative binding between Py-Py-Im and biBenz moieties. Conjugate **1** is highly selective ($K_1 = \sim 2.6 \times 10^9$) for the nine-bp site 5′-gcggTATGAAATTcgacg-3′ (dsDNA **14**). The Im residues in polyamides that bind to dsDNA in a 1:1 stoichiometry generally do not distinguish G/C from A/T but bind to all four bps with high affinity.[37] However, the Im residue in conjugate **1** can distinguish G/C from A/T bp by ∼8-fold magnitude (Table 1). Though the dsDNA **16** possesses a binding site size similar to that of dsDNA **14**, $K_1$ is lower by ∼30-fold magnitude (Table 1). This is probably due to the preferential sequence selectivity by the biBenz moiety; that is, Hoechst is more selective for AATT than TATA sequence.[38,39]

Conjugates **2**−**4**, with different combinations of Py and Im moieties, were synthesized to study the sequence selectivity in the minor groove of dsDNA. The most selective sequence for each conjugate was determined by gradually changing the G/C position in the nine-bp binding site. On the basis of equilibrium association constants, conjugates **2**−**4** recognize only seven bps (Table 1) and the selective sequence for each conjugate is illustrated in Figure 6. Similar to **1**, the Im residue in conjugates **3** and **4** selectively targets G/C bp embedded in the A/T-rich binding site. Conjugates **3** and **4** are observed to form significantly stable complexes with 5′-gcggtaTGAAATTcgacg-3′ (dsDNA **14**) and 5′-gcggtaCAAAATTcgacg-3′ (dsDNA **13**) binding sites, respectively. Displacement of the G/C position or substitution of G/C with A/T bp exhibited significant change in $K_1$ (Table 1). Although the binding sites in dsDNAs **14** and **16** (for conjugate **3**) and **13** and **17** (for conjugate **4**) are quite similar, the $K_1$ values differ by several-fold (Table 1). This is also probably due to the preferential sequence selectivity of the biBenz moiety for the AATT over the TATA sequence.[38,39] Conjugate **2**, with no Im moiety, binds to seven contiguous A/T bps and also possesses the ability to distinguish between its preferred seven bp A/T-rich binding site and shorter sites with fewer than seven contiguous A/T bps (Table 1).

Thermal denaturation experiments were employed as an alternative method for the investigation of ligand:dsDNA complex stabilities and sequence selectivities. From the $T_m$ data shown in Table 2, it is clear that the thermal melting data strongly support the data acquired via the equilibrium assays.[40] The conjugates **1**−**4** form stable complexes with preferred sequences, as illustrated in Figure 6. Conjugate **1** forms stable complex ($\Delta T_m = 9$ °C) with the nine-bp site 5′-gcggTATGAAATTcgacg-3′ (dsDNA **14**). The $\Delta T_m$ values are greatly affected when G/C position is displaced or substituted with A/T bp. Similarly, conjugates **3** and **4** form stable complexes ($\Delta T_m = 7-8$ °C) with 5′-gcggtaTGAAATTcgacg-3′ (dsDNA **14**) and 5′-gcggtaCAAAATTcgacg-3′ (dsDNA **13**) sequences, respectively. Additionally, two continuous G/C bps in the binding site (dsDNA **20**) exhibited $\Delta T_m$ of 2 (for conjugate **1**) and 0 °C (for conjugates **2**−**4**). In contrast, Ht33258 $\Delta T_m$ values are fairly similar for all dsDNAs investigated (Table 2). Thus, from $K_1$ data and $\Delta T_m$ values, it is clear that an Im moiety in the pyrrole, imidazole-biBenz conjugates is able to recognize G/C and exhibit preference in binding over A/T bp in the A/T-rich binding site. Further, these are so far 1:1 ligand:dsDNA complexes.

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

JA035116K

(35) Neidle, S. *Biopolymers* **1997**, *44*, 105.
(36) Geierstanger, B. H.; Wemmer, D. E. *Annu. Rev. Biophys. Bimol. Struct.* **1995**, *24*, 463.
(37) Urbach, A. R.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4343.
(38) Abu-Daya, A.; Brown, P. M.; Fox, K. R. *Nucleic Acids Res.* **1995**, *23*, 3385.
(39) Satz, A. L.; White, C. M.; Beerman, T. A.; Bruice, T. C. *Biochemistry* **2001**, *40*, 6465.
(40) The $\Delta T_m$ is a complex function of the ligand binding constant, site size, and enthalpies of both ligand binding and dsDNA melting. For discussion see: Spink, C. H.; Wellman, S. E. *Methods Enzymol.* **2001**, *340*, 193.