Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 10389–10392, October 1995
Chemistry

Cyclic polyamides for recognition in the minor groove of DNA

JUNHYEONG CHO, MICHELLE E. PARKS, AND PETER B. DERVAN*

Beckman Institute, California Institute of Technology, Pasadena, CA 91125

Contributed by Peter B. Dervan, July 21, 1995

ABSTRACT **Small molecules that specifically bind with high affinity to any designated DNA sequence in the human genome would be useful tools in molecular biology and potentially in human medicine. Simple rules have been developed to rationally alter the sequence specificity of minor groove-binding polyamides containing *N*-methylimidazole and *N*-methylpyrrole amino acids. Crescent-shaped polyamides bind as antiparallel dimers with each polyamide making specific contacts with each strand on the floor of the minor groove. Cyclic polyamides have now been synthesized that bind designated DNA sequences at subnanomolar concentrations.**

Footprinting, affinity cleavage, NMR, and x-ray studies have established that polyamides containing *N*-methylimidazole (Im) and *N*-methylpyrrole (Py) amino acids can be combined in antiparallel side-by-side dimeric complexes with the minor groove of DNA (1–16). The DNA sequence specificity of these small molecules can be controlled by the linear sequence of pyrrole and imidazole amino acids (1–13). An imidazole ring on one ligand complemented by a pyrrolecarboxamide ring on the neighboring ligand recognizes a G·C base pair, whereas a pyrrolecarboxamide/imidazole combination targets a C·G base pair (2–4). A pyrrole/pyrrole pair is degenerate for A·T or T·A base pairs (1–4).

Despite this design breakthrough in molecular recognition of DNA, the binding affinities of this class of polyamide dimers are modest (4). For example, a three-ring polyamide dimer in complex with a 5-bp DNA site has a binding affinity typically in the range of $K = 2 \times 10^5\ M^{-1}$ (pH 7.0, 22°C) (4). In an effort to improve the energetics, antiparallel dimers were connected by a central γ-aminobutyric acid (γ) residue to create a single molecule that could bind in the minor groove by folding to a hairpin shape (11). Second-generation "hairpin" polyamides of sequence composition Im-Py-Py-γ-Py-Py-Py-Dp **1** (where Dp is 3-dimethylaminopropyl side chain) were shown to have improved equilibrium association constants of $8 \times 10^7\ M^{-1}$ for designated 5-bp target sites 5′-TGTTA-3′, an increase of 400 over the unlinked dimers (Fig. 1). In a formal sense, γ-linked pyrrole–imidazole polyamides could exist in at least two conformations, hairpin or extended, resulting in different DNA-binding motifs and, hence, possible different DNA specificities. Closing the ends of the hairpin to a circle would restrict conformational space for the DNA-binding molecule and presumably further increase the overall energetics. We report here the synthesis of cyclo-(Im-Py-Py-γ-Py-Py-Py-γ) **2** and analysis of the binding affinity and specificity of this class of cyclic DNA-binding small molecules (Fig. 1).

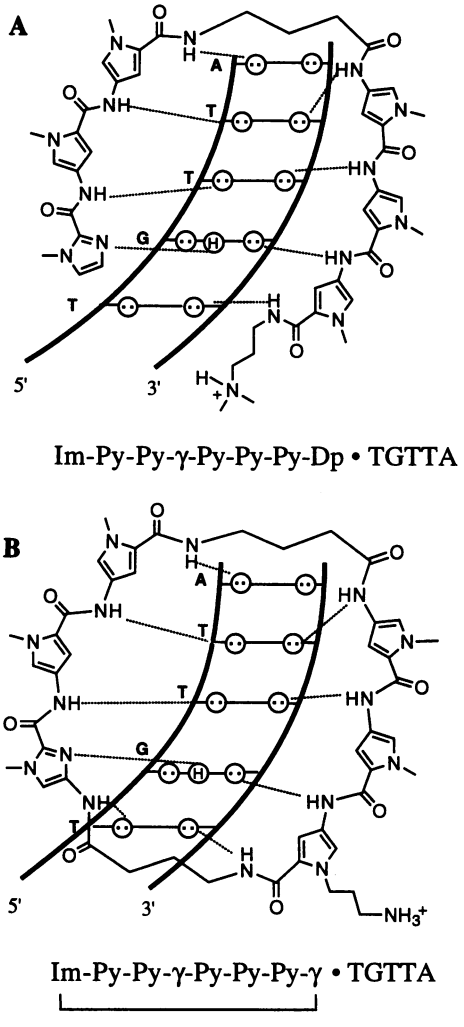

FIG. 1. Im-Py-Py-γ-Py-Py-Py-Dp **1** (*A*) and the cyclic polyamide cyclo-(Im-Py-Py-γ-Py-Py-Py-γ) **2** (*B*) with a 5′-TGTTA-3′ sequence. ⊙, Lone pairs of N3 of purines and O2 of pyrimidines; ⊛, 2-amino group of guanine; ····, putative hydrogen bonds. Dp, 3-dimethylaminopropyl side chain.

MATERIALS AND METHODS

General. ^1H NMR spectra were recorded at 300 MHz on a General Electric model 300 NMR in C^2HCl_3, dimethyl sulfoxide (DMSO)-d_6, or $C^2H_3O^2H$. Chemical shifts are reported in parts per million relative to tetramethylsilane or residual dimethyl sulfoxide-d_5. IR spectra were recorded on a Perkin–Elmer Fourier transform infrared spectrometer. High-resolution mass spectra were recorded using fast atom bombardment (FAB) techniques at the Mass Spectrometry Laboratory (University of California, Riverside). Preparatory HPLC was done on a Beckman 340 Instrument using a Waters DeltaPak 25 × 100 mm C_{18} column. Analytical HPLC was done on a Hewlett–Packard 1090 series II using a Vydac C_{18} reverse-phase column (0.46 × 25 cm, 5 mm, 201HS54). Flash column chromatography was done by using silica gel 60 (230–400 mesh, Merck). Thin-layer chromatography (TLC) was done on silica gel 60 F_{254}-precoated plates (Merck).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Im, *N*-methylimidazole; Py, *N*-methylpyrrole; γ, γ-aminobutyric acid; FAB, fast atom bombardment.

*To whom reprint requests should be addressed.

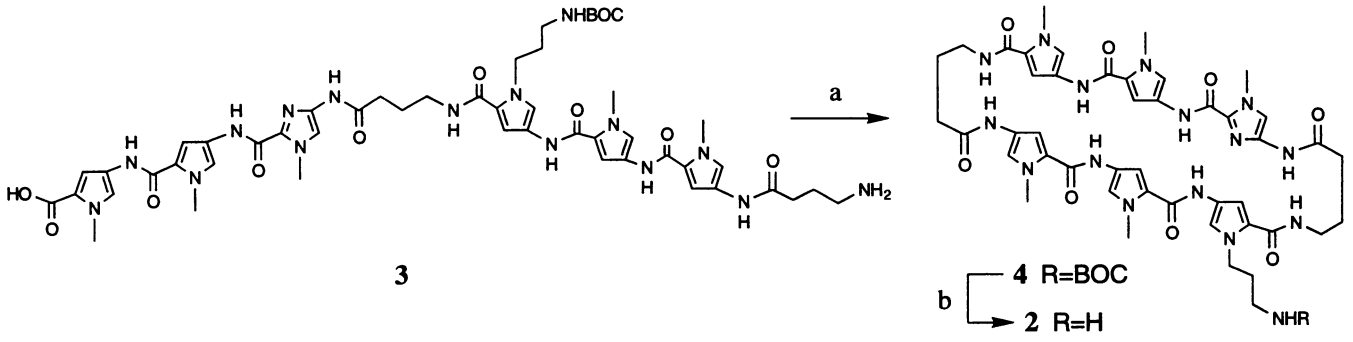

FIG. 2. Synthetic scheme for cyclization to cyclic polyamide **2**. (*a*) Diphenylphosphoryl azide/$NaHCO_3$/*N,N'*-dimethylformamide; (*b*) trifluoroacetic acid/thiophenol.

Chemicals for the syntheses were purchased from Aldrich unless otherwise specified. Dichloromethane and *N,N*-dimethylformamide were purchased as anhydrous solvents from Aldrich. *t*-Butoxycarbonyl (Boc)-γ and *N*-carbobenzoxy (Cbz)-γ were from Sigma. All enzymes were purchased from Boehringer Mannheim and used with the buffers supplied, except for *Bsr*BI, which was obtained from New England Biolabs. Deoxyadenosine 5'-[α-^{32}P]triphosphate and thymidine 5'-[α-^{32}P]triphosphate were from Amersham. Storage phosphor technology autoradiography was done by using a Molecular Dynamics model 400S PhosphorImager and IMAGEQUANT software.

***t*-Butoxycarbonyl-Protected Cyclic Polyamide 4.** To a solution of polyamide **3** (150 mg, 0.18 mmol) in *N-N'*-dimethylformamide (70 ml) was added sequentially $NaHCO_3$ (78 mg, 0.93 mmol) and diphenylphosphoryl azide (154 mg, 0.56 mmol). The mixture was allowed to stir at room temperature for 3 days and then filtered. The filtrate was concentrated *in vacuo*, and the residue was purified by flash column chromatography [15% (vol/vol) methanol/CH_2Cl_2] to afford the cyclic polyamide **4** (90 mg, 58%) ^{1}H NMR [$CDCl_3$-$C^2H_3O^2H$ (1:1)] δ 7.30 (s, 1H), 7.26 (s, 1H), 7.18 (s, 1H), 7.02 (s, 2H), 6.88 (s, 1H), 6.70 (s, 1H), 6.68 (s, 1H), 6.62 (s, 1H), 6.58 (s, 1H), 6.43 (s, 1H), 4.35 (m, 2H), 3.92 (s, 6H), 3.90 (s, 3H), 3.89 (s, 3H), 3.82 (s, 3H), 3.49 (m, 4H), 3.08 (m, 2H), 2.42 (m, 4H), 1.93–2.15 (m, 6H), 1.45 (s, 9H); IR (KBr) 3422 (s), 1654 (s), 1545 (s), 1438 (s), 1406 (m), 1260 (m), 1208 (m), 1105 (w), 919 (w), 777 (w) cm^{-1}; FAB MS *m/e* observed 1047.4884 (M+H, 1047.4913 calculated for $C_{50}H_{62}N_{16}O_{10}$).

Cyclic Polyamide 2. To a mixture of trifluoroacetic acid and thiophenol (4 ml, 0.5 M thiophenol/trifluoroacetic acid) was added the *t*-butoxycarbonyl (Boc)-protected cyclic polyamide **4** (30 mg, 28.6 mmol), and the reaction mixture was allowed to stir at room temperature for 2 hr. Excess trifluoroacetic acid was removed *in vacuo*, and the residue was purified by flash column chromatography (5% NH_4OH/methanol) to afford cyclic polyamide **2** (23 mg, 87%). The product was purified by using reverse-phase HPLC on a preparatory Waters DeltaPak 25 × 100 mm C_{18} column with linear gradients of 60% acetonitrile plus 0.1% trifluoroacetic acid versus 0.1% aqueous trifluoroacetic acid. ^{1}H NMR (dimethyl sulfoxide-d_6) δ 10.00 (s, 1H), 9.99 (s, 1H), 9.96 (s, 2H), 9.95 (s, 2H), 9.87 (s, 1H), 7.47 (s, 2H), 7.44 (s, 1H), 7.39 (s, 1H), 7.34 (s, 1H), 7.16 (s, 1H), 6.94 (s, 1H), 6.90 (s, 1H), 6.88 (s, 1H), 6.71 (s, 1H), 6.65 (s, 1H), 4.29 (m, 2H), 3.94 (s, 3H), 3.84 (s, 3H), 3.83 (s, 3H), 3.82 (s, 3H), 3.79 (s, 3H), 3.22 (br s, 4H), 2.52 (m, 2H), 2.39 (m, 2H), 2.30 (t, 2H, J = 7.0 Hz), 1.78 (m, 6H); IR 3383 (s), 1637 (s), 1560 (s), 1438 (s), 1407 (s), 1265 (m), 1207 (w), 1122 (w), 1062 (w), 777 (w) cm^{-1}; UV (H_2O) λmax (ε) 240 (37,075 cm^{-1}·M^{-1}), 310 (48,096 cm^{-1}·M^{-1}) nm; FAB MS *m/e* observed 947.4333 (M+H, 947.4389 calcd. for $C_{45}H_{54}N_{16}O_8$).

Preparation of Labeled DNA. The 135-bp 3' end-labeled *Eco*RI/*Bsr*BI restriction fragment from plasmid pMM5 was prepared and purified as follows. Plasmid DNA was linearized using *Eco*RI with simultaneous fill in by Klenow fragment, deoxyadenosine 5'-[α-^{32}P]triphosphate, and thymidine 5'-[α-^{32}P]triphosphate. The linearized plasmid DNA was digested with *Bsr*BI and the 135-bp *Eco*RI/*Bsr*BI restriction fragment was isolated by nondenaturing 5% PAGE. The gel bands were visualized by autoradiography, isolated, and filtered to remove the polyacrylamide. The resulting solution was further purified by phenol extraction followed by ethanol precipitation. The plasmid pJT3 *Afl* II/*Fsp* I 281-bp fragment was prepared analogously. Chemical sequencing reactions were done according to published protocols (17, 18). Standard protocols were used for all DNA manipulations (19).

Quantitative DNase I Footprint Titration. All reactions were done in a total volume of 40 μl with final concentrations of each species as indicated. The polyamide ligands, ranging in concentration from 10 pM to 100 nM, were added to solutions of radiolabeled restriction fragment (10,000 cpm), Tris·HCl (10 mM, pH 7.0), KCl (10 mM), $MgCl_2$ (10 mM), and $CaCl_2$ (5 mM) and incubated for 24 hr at 22°C. We explicitly note that no carrier calf thymus DNA is used in these reactions. Footprinting reactions were initiated by the addition of 4 μl of a stock solution of DNase I (0.03 unit/ml) containing 1 mM dithiothreitol and allowed to proceed for 7 min at 22°C. The reactions were stopped by addition of 3 M sodium acetate/50 mM EDTA and ethanol-precipitated. The reactions were resuspended in 100 mM Tris borate/EDTA/80% formamide loading buffer and electrophoresed on 8% polyacrylamide denaturing gels (5% crosslink, 7 M urea) at 2000 V for 1 hr.

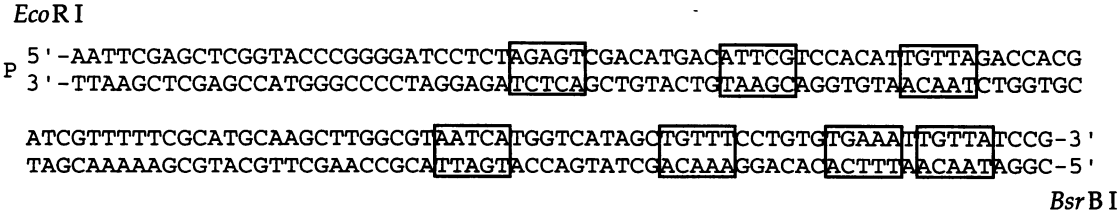

FIG. 3. Sequence of the 135-bp *Eco*RI/*Bsr*BI restriction fragment. Several 5-bp sites (boxed) were analyzed by quantitative footprint titration analysis. Note the additional binding sites on the upper half of the autoradiogram (Fig. 4) of sequence composition (A/T)G(A/T)$_3$ which are also occupied by the cyclic polyamide.

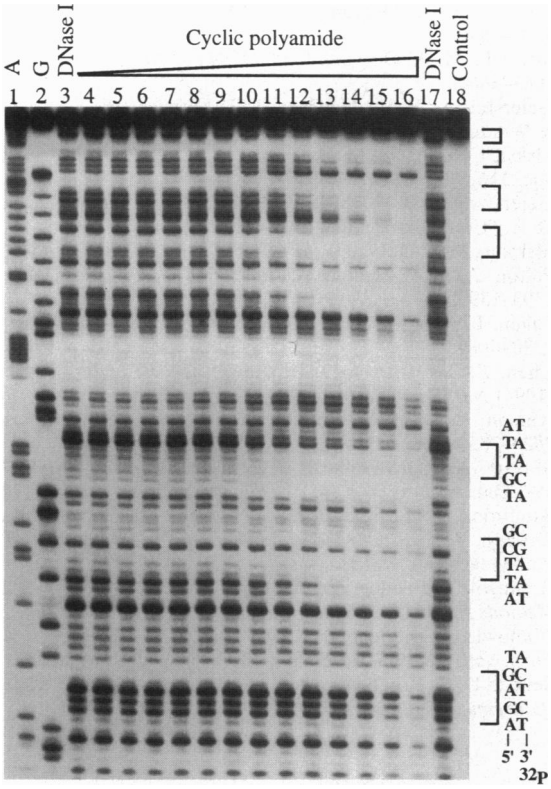

FIG. 4. Quantitative DNase I footprint titration experiment with the cyclic polyamide **2** on the 3′-^{32}P-labeled 135-bp *Eco*RI/*Bsr*BI restriction fragment from plasmid pMM5. Lanes: 1, adenine sequencing reaction; 2, guanine sequencing reaction; 3 and 17, DNase I standard; 4–16, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, and 100 nM cyclic polyamide, respectively; 18, intact DNA. The 5′-TGTTA-3′, 5′-ATTCG-3′, and 5′-AGAGT-3′ binding sites that were analyzed are shown at right. All reactions contain 10 kcpm of restriction fragment, 10 mM Tris·HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

The footprint titration gels were dried and quantitated using storage phosphor technology. Apparent first-order binding constants were determined as described (11).

RESULTS AND DISCUSSION

A cyclic polyamide **2** containing eight amino acids was synthesized in 12 steps from readily available starting materials. The key cyclization reaction was accomplished in good yield (58%) using diphenylphosphoryl azide in the presence of $NaHCO_3$ followed by the deprotection with trifluoroacetic acid to give the cyclic product cyclo-(Im-Py-Py-γ-Py-Py-Py-γ) **2**, which was purified by reverse-phase HPLC (Fig. 2). The observed molecular mass of the $C_{45}H_{54}N_{16}O_8$ polyamide is 947.4333 (FAB MS) in good agreement with the calculated value.

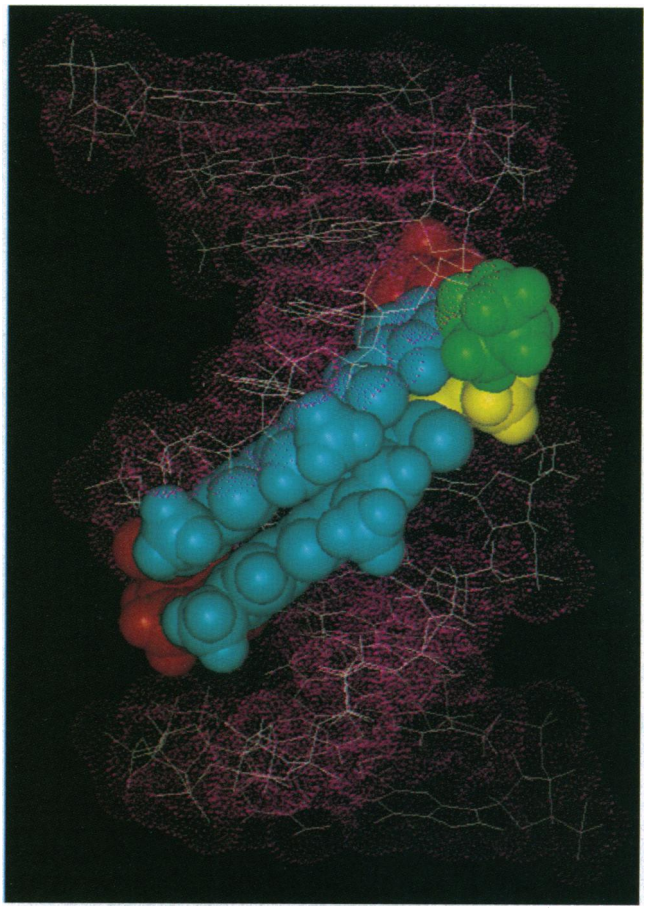

FIG. 5. Proposed model of the cyclic polyamide binding in the minor groove of DNA. Colors represent the following amino acids: cyan, Py; yellow, Im; magenta, γ; green, Dp side chain.

Quantitative DNase I footprint titration experiments [refs. 20–22; the quantitative footprint titration experiments and data analysis were done as described (11) with unlabeled carrier absent from all reactions] on a ^{32}P end-labeled 135-bp restriction fragment were done to obtain the binding affinity of the cyclic polyamide **2** for a designated match site, 5′-TGTTA-3′ (Fig. 3). The cyclic polyamide binds the 5′-TGTTA-3′ site very tightly with an equilibrium association constant of $K \geq 2.9 \times 10^9$ M^{-1} (Fig. 4). We attribute this increase of 40 in binding affinity over the corresponding hairpin polyamide **1** to a reduction in conformational entropy imparted via cyclization (Fig. 5). The binding affinities to single-base-pair mismatch sites, 5′-ATTCG-3′, 5′-AGAGT-3′, and 5′-AGACA-3′ are lower, 7.4×10^8, 1.7×10^8, and 6.2×10^8 M^{-1}, respectively (Table 1). The cycle appears to bind the match over single-base-pair mismatch site by a factor of 4–17, although these values may be lower limits due to the uncertainty in the very high equilibrium association constant value for the cyclic polyamide **2**–TGTTA complex. The hairpin binds

Table 1. Apparent first-order binding constants, M^{-1}

	Match site	Single-base mismatch sites		
Polyamide	5′-TGTTA-3′	5′-ATTCG-3′	5′-AGAGT-3′	5′-AGACA-3′*
Hairpin **1**	7.6×10^7 (0.8)	†	7.8×10^5 (2.0)	2.6×10^6 (0.9)
Cyclic **2**	2.9×10^9 (1.9)	7.4×10^8 (2.4)	1.7×10^8 (1.4)	6.2×10^8 (2.9)

Values are the mean values measured from at least three footprint titration experiments; the SD for each data set is indicated in parentheses. Assays were done at 22°C at pH 7.0 in 10 mM Tris·HCl/10 mM KCl/10 mM $MgCl_2$/5 mM $CaCl_2$.

*These data were determined on a different restriction fragment.

†Data were not determined because hairpin **1** binds neighboring 5′-TGACA-3′ single-base mismatch site.

more strongly to a match over single-base mismatch sites by a factor of 20. The cyclic polyamide, though higher in affinity, is less than or only equal to the hairpin in specificity. There may be an energetic price in specificity for closing the ends of the hairpin. Nevertheless, that wholly designed synthetic cyclic polyamides with a M_r of 950 bind designated 5-bp sequences at subnanomolar concentrations is an encouraging step forward. Further optimizing specificity remains a challenge.

We thank the Mass Spectrometry Laboratory (University of California, Riverside) for the mass spectral analysis. We are grateful to the National Institutes of Health (Grant 27681) for research support and a National Institutes of Health Research Service Award to M.E.P.

1. Pelton, J. G. & Wemmer, D. E. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5723–5727.
2. Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E. & Dervan, P. B. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7586–7590.
3. Wade, W. S., Mrksich, M. & Dervan, P. B. (1992) *J. Am. Chem. Soc.* **114,** 8783–8794.
4. Wade, W. S., Mrksich, M. & Dervan, P. B. (1993) *Biochemistry* **32,** 11385–11389.
5. Mrksich, M. & Dervan, P. B. (1993) *J. Am. Chem. Soc.* **115,** 2572–2576.
6. Geierstanger, B. H., Dwyer, T. J., Bathini, Y., Lown, J. W. & Wemmer, D. E. (1993) *J. Am. Chem. Soc.* **115,** 4474–4482.
7. Dwyer, T. J., Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1993) *J. Am. Chem. Soc.* **115,** 9900–9906.
8. Mrksich, M. & Dervan, P. B. (1993) *J. Am. Chem. Soc.* **115,** 9892–9899.
9. Mrksich, M. & Dervan, P. B. (1994) *J. Am. Chem. Soc.* **116,** 3663–3664.
10. Geierstanger, B. H., Jacobsen, J. P., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1994) *Biochemistry* **33,** 3055.
11. Mrksich, M., Parks, M. E. & Dervan, P. B. (1994) *J. Am. Chem. Soc.* **116,** 7983–7988.
12. Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1994) *Science* **266,** 646–650.
13. Mrksich, M. & Dervan, P. B. (1995) *J. Am. Chem. Soc.* **117,** 3325.
14. Pelton, J. G. & Wemmer, D. E. (1990) *J. Am. Chem. Soc.* **112,** 1393–1399.
15. Fagan, P. A. & Wemmer, D. E. (1992) *J. Am. Chem. Soc.* **114,** 1080–1081.
16. Chen, X., Ramakrishnan, B., Rao, S. T. & Sundaralingam, M. (1994) *Nat. Struct. Biol.* **1,** 169–175.
17. Iverson, B. L. & Dervan, P. B. (1987) *Nucleic Acids Res.* **15,** 7823–7830.
18. Maxam, A. M. & Gilbert, W. S. (1980) *Methods Enzymol.* **65,** 499–560.
19. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
20. Brenowitz, M., Senear, D. F., Shea, M. A. & Ackers, G. K. (1986) *Methods Enzymol.* **130,** 132–181.
21. Brenowitz, M., Senear, D. F., Shea, M. A. & Ackers, G. K. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8462–8466.
22. Senear, D. F., Brenowitz, M., Shea, M. A. & Ackers, G. K. (1986) *Biochemistry* **25,** 7344–7354.