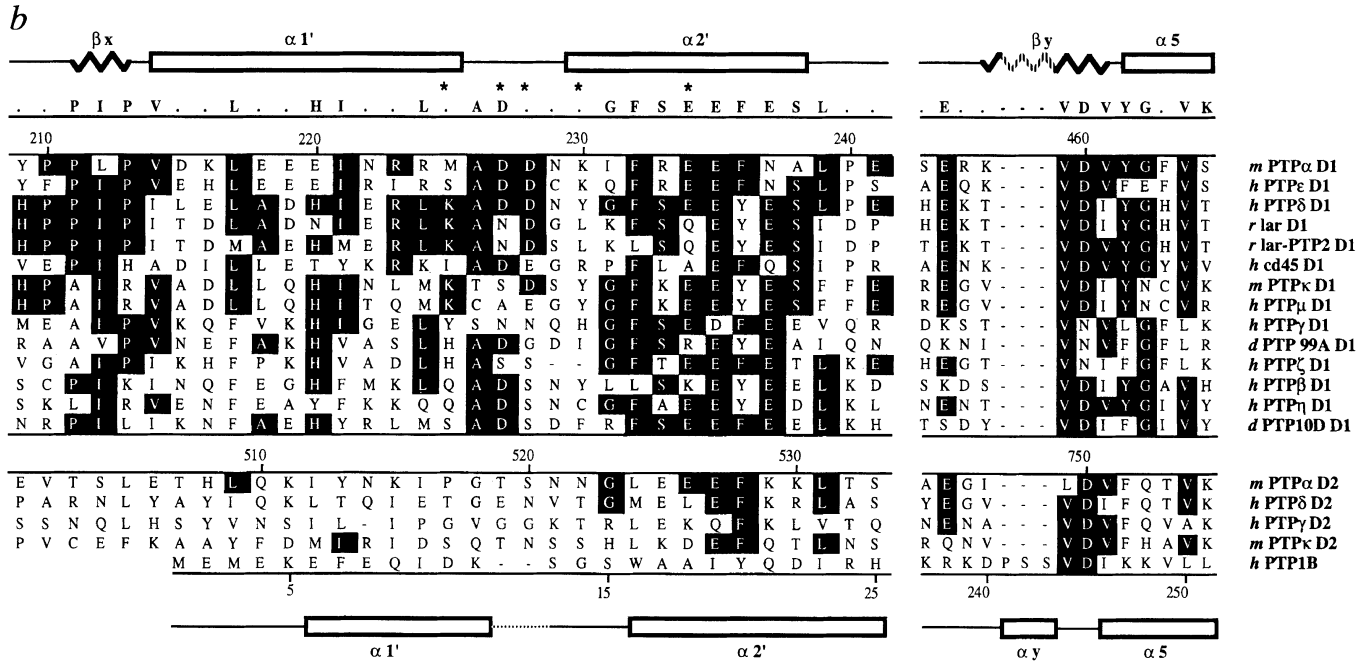


20. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* A**47,** 110–119 (1991).
21. Brunger, A. T. X-PLOR, Version 3.1. (Yale Univ. Press, New Haven, CT 1992).
22. Lamzin, V. S. & Wilson, K. S. *Acta crystallogr.* D**49,** 127–149 (1993).
23. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. *J. appl. Crystallogr.* **26,** 283–291 (1993).
24. Huang, C. C., Pettersen, E. F., Klein, T. E. & Langridge, R. *J. molec. Graphics* **9,** 230–236 (1991).
25. Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. *J. molec. Graphics* **6,** 13–27 (1988).
26. Evans, S. V. *J. molec. Graphics* **11,** 134–138 (1993).
27. Goldstein, B. J. *Phosphoprotein Phosphatases 1: Tyrosine Phosphatases* (Academic, London, 1995).
28. Sap, J., D'Eustachio, P., Givol, D. & Schlessinger, J. *Proc. natn. Acad. Sci. U.S.A.* **87,** 6112–6116 (1990).

ACKNOWLEDGEMENTS. We thank S. Choe, R. Ranganathan and G. V. Louie for discussion; S. M. Redford and C. M. Starks for assistance during data collection at beamline 7-1 at the Stanford Synchrotron Radiation Laboratory (SSRL); M. Park for N-terminal sequencing; D. Wild, M. E. Bowman, W. Yang, F. D. Bushman, M. B. Harrington, and the staff at SSRL for technical help; and N. P. H. Møller and Novo Nordisk for peptides and financial support (A.M.B.). The work performed at SSRL was supported by the NIH and DOE. Work in the structural biology laboratory at the Salk Institute was supported in part from institutional funds and funds from the Lucille P. Markey Charitable Trust. J.d.H. is supported by a grant from the Dutch Cancer Society. T.H. is an American Cancer Society Professor.

CORRESPONDENCE and requests for materials should be addressed to J.P.N. (e-mail: noel@sbl.salk.edu). Coordinates for the RTPTαD1 dimer will be submitted to the Brookhaven Protein Data Bank.

Recognition of DNA by designed ligands at subnanomolar concentrations

John W. Trauger, Eldon E. Baird & Peter B. Dervan

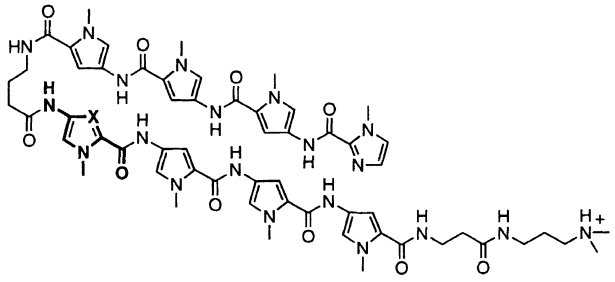
Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

SMALL molecules that specifically bind with high affinity to any predetermined DNA sequence in the human genome would be useful tools in molecular biology and potentially in human medicine. Simple rules have been developed to control rationally the sequence specificity of minor-groove-binding polyamides containing *N*-methylimidazole and *N*-methylpyrrole amino acids. Two eight-ring pyrrole–imidazole polyamides differing in sequence by a single amino acid bind specifically to respective six-base-pair target sites which differ in sequence by a single base pair. Binding is observed at subnanomolar concentrations of ligand. The replacement of a single nitrogen atom with a C-H regulates affinity and specificity by two orders of magnitude. The broad range of sequences that can be specifically targeted with pyrrole–imidazole polyamides, coupled with an efficient solid-phase synthesis methodology, identify a powerful class of small molecules for sequence-specific recognition of double-helical DNA.

For side-by-side complexes of pyrrole–imidazole polyamides in the minor groove of DNA, the DNA-binding sequence specificity depends on the sequence of side-by-side amino-acid pairings[1–3]. A pairing of imidazole (Im) opposite pyrrole (Py) targets a G·C base pair, whereas pyrrole opposite imidazole targets a C·G base-pair[1–3]. A pyrrole/pyrrole combination is degenerate and targets both T·A and A·T base pairs[1–5]. Specificity for G,C base pairs results from the formation of a hydrogen bond between the imidazole N3 and the exocyclic amino group of guanine[1–3]. The generality of these pairing rules has been demonstrated by targeting a wide variety of sequences and is supported directly by several NMR structure studies[1–10].

In parallel with the elucidation of the scope and limitations of the pairing rules already described, efforts have been made to increase the DNA-binding affinity and specificity of pyrrole–imidazole polyamides by covalently linking polyamide subunits[11–16]. The polyamide ImPyPy-γ-PyPyPy-dimethylaminopropylamide (Dp) , which contains a 'turn' amino acid γ-aminobutyric acid (γ) was shown to bind the five-base-pair target site 5′-TGTTA-3′ in a 'hairpin' conformation with an equilibrium association constant, $K_a = 8 \times 10^7\ M^{-1}$, an increase of 300-fold relative to unlinked three-ring polyamide dimers. A key issue was to determine whether low-molecular-weight (M_r 1,200) pyrrole–imidazole polyamides could be constructed that would bind DNA at subnanomolar concentrations without compromising sequence selectivity.

We report the DNA-binding affinities of two eight-ring hairpin polyamides, ImPyPyPy-γ-ImPyPyPy-β-Dp (**1**) and ImPyPyPy-γ-PyPyPyPy-β-Dp (**2**), which differ by a single amino acid, for two 6-base-pair (bp) target sites, 5′-AGTACT-3′ and 5′-AGTATT-3′, which differ by a single base pair. Based on the pairing rules for polyamide–DNA complexes, the sites 5′-AGTACA-3′ and 5′-AGTATT-3′ are for polyamide **1** 'match' and 'single-base-pair mismatch' sites, respectively, and for polyamide **2** 'single-base-pair mismatch' and 'match' sites, respectively (Figs 1 and 2).



1 X=N; ImPyPyPy-γ-**Im**PyPyPy-β-Dp
2 X=CH; ImPyPyPy-γ-**Py**PyPyPy-β-Dp

FIG. 1 Structures of polyamides ImPyPyPy-γ-ImPyPyPy-β-Dp (**1**) and ImPyPyPy-γ-PyPyPyPy-β-Dp (**2**). (Im = imidazole, Py = pyrrole, β = β-alanine, Dp = dimethylaminopropylamide). The identity and purity of the polyamides was verified by ^{1}H NMR, matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS), and analytical HPLC. MALDI-TOF MS: **1**, 1223.4 (1223.3 calculated); **2**, 1222.3 (1222.3 calculated).

Polyamides **1** and **2** were synthesized by solid-phase methods and purified by reversed-phase high-performance liquid chromatography (HPLC)[17]. Equilibrium association constants for complexes of **1** and **2** with match and mismatch six-base-pair binding sites on a 3′ ^{32}P-labelled 229-bp restriction fragment were determined by quantitative DNase I footprint titration experiments[18–20] (Fig. 3 and Table 1). Polyamide **1** binds its match site 5′-AGTACT-3′ at 0.03 nM concentration and its single-base-pair mismatch site 5′-AGTATT-3′ with nearly 100-fold lower affinity. Polyamide **2** binds its designated match site 5′-AGTATT-3′ at 0.3 nM concentration and its single-base-pair mismatch site 5′-AGTACT-3′ with nearly 10-fold lower affinity. The specificity of **1** and **2** for their respective match sites results from very small structural changes (Fig. 1). Replacing a single nitrogen atom in **1** with C-H (as in **2**) reduces the affinity of the polyamide·5′-AGTACT-3′ complex by ~75-fold, representing a free energy difference of ~2.5 kcal mol^{-1}. Similarly, replacing a C-H in **2** with N (as in **1**) reduces the affinity of the polyamide·5′-AGTATT-3′ complex ~10-fold, a loss in binding energy of ~1.3 kcal mol^{-1}.

Crystal structures of protein–DNA complexes reveal that nature chose a combinatorial approach for specific DNA recognition. Although there are several highly conserved structural

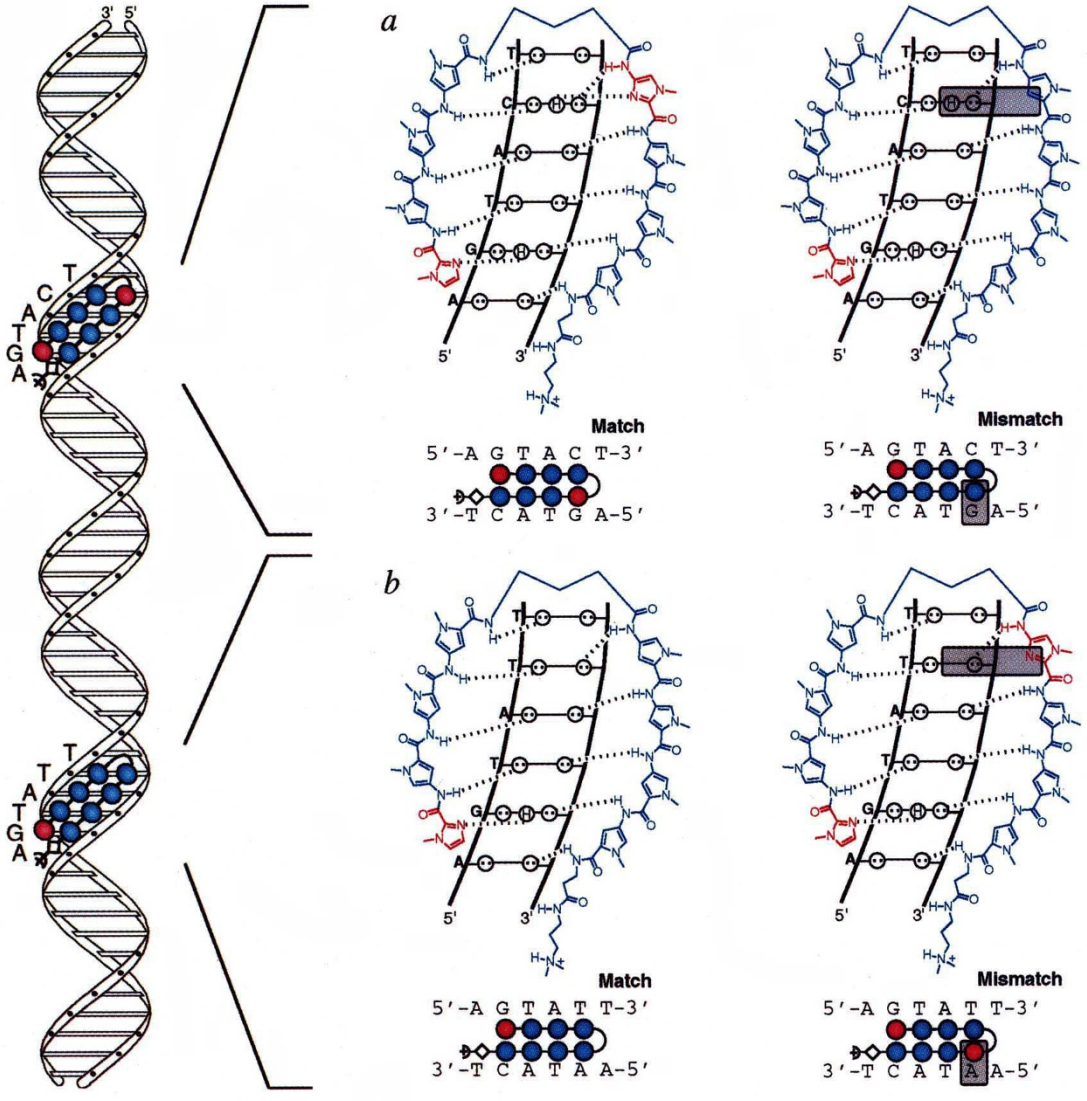


FIG. 2 Binding models for *a*, 5′-AGTACT-3′ in complex with polyamides **1** (match) and **2** (mismatch), and *b*, 5′-AGTATT-3′ in complex with polyamides **2** (match) and **1** (mismatch). Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines; circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are indicated by dashed lines. The red and blue circles represent imidazole and pyrrole rings, respectively; the curved line represents γ-aminobutyric acid, and the diamond, β-alanine. Single hydrogen-bond mismatches are highlighted.

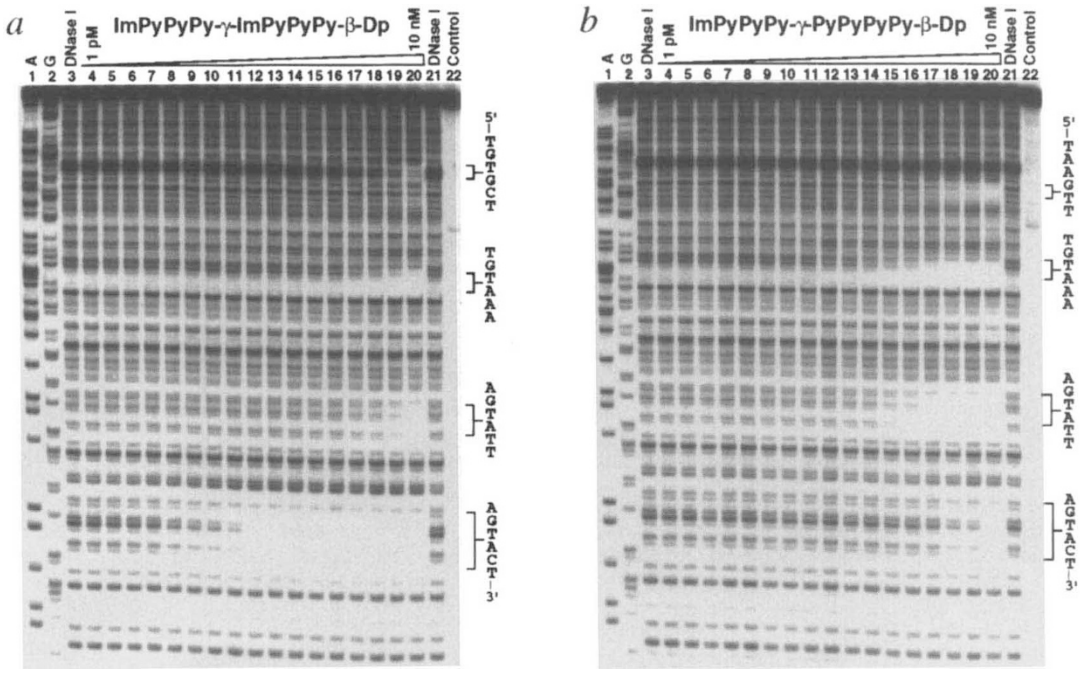


FIG. 3 Quantitative DNase I footprint titration experiments with polyamides **1** and **2** on the 3′ ^{32}P-labelled 229-bp pJT8 *Afl*II/*Fsp*I restriction fragment. Lanes 1 and 2: A and G sequencing lanes[29,30]; lanes 3 and 21; DNase I digestion products in the absence of polyamide; lanes 4–20: DNase I digestion products in the presence of 1, 2, 5, 10, 15, 25, 40 and 65 pM, and 0.1, 0.15, 0.25, 0.4, 0.65, 1, 2, 5 and 10 nM polyamide, respectively; lane 22: intact DNA. Polyamide binding sites for which association constants were determined are 5′-AGTACT-3′ and 5′-AGTATT-3′. Additional sites not analysed are 5′-TGTAAA-3′, 5′-TGTGCT-3′ and 5′-TAAGTT-3′. All reactions were done in a total volume of 400 μl. A polyamide stock solution or H_2O was added to an assay buffer containing radiolabelled restriction fragment, with final solution conditions of 10 mM Tris–HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0. Solutions were allowed to equilibrate for 12–15 h at 22 °C before initiation of footprinting reactions. Footprinting reactions, separation of cleavage products, and data analysis were carried out as described[14]. Plasmid pJT8 was prepared by hybridizing two 5′-phosphorylated complementary oligonucleotides, 5′-CCGGTTAGTATTTGG-ATGGGCCTGGTTAGTACTTGGATGGGAGACCGCCTGGGAATACCAGGTGTCGTATC-TTAAGAG-3′ and 5′-TCGACTCTTAAGATACGACACCTGGTATTCCCAGGCGGTC-TCCCATCCAAGTACTAACCAGGCCCATCCAAATACTAA-3′, and ligating the resulting duplex to the large pUC19 *Ava*I/*Sal*I restriction fragment.

TABLE 1 Equilibrium association constants (M^{-1})

Binding site	**1**	**2**
5′-ttAGTACTtg-3′	3.7×10^{10} (0.8)	5.0×10^{8} (0.5)
5′-ttAGTATTtg-3′	4.1×10^{8} (0.5)	3.5×10^{9} (0.8)

The association constants are the average values obtained from three DNase I footprint titration experiments. The standard deviation for each data set is indicated in parentheses. Assays were carried out in the presence of 10 mM Tris–HCl, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$ at pH 7.0 and 22 °C. The six-base-pair binding sites are in capital letters, with flanking sequences in lower-case.

modules that bind DNA, no single motif exists which generates an amino acid–base pair code for all DNA sequences[21]. One encouraging success is the development of zinc-finger libraries with different DNA-binding specificities based on a one-finger–three-nucleotide code[22–27]. The three-zinc-finger DNA-binding domain of the transcription factor Zif268 has a dissociation constant $K_d = 0.5–3.0$ nM for its 9-bp binding site and is 2–12-fold specific over single-base-pair mismatches[24–27]. Similarly, the three-zinc-finger transcription factor Sp1 has a $K_d = 0.5$ nM for its 9-bp recognition sequence and is 3–30-fold specific over single-base-pair mismatches[28]. Using a simple molecular shape and a two-letter aromatic amino-acid code, pyrrole–imidazole polyamides achieve affinities and specificities comparable to these DNA-binding proteins and, in addition, have the potential to be general for any desired DNA sequence. This non-biological approach to DNA recognition could provide an underpinning for the design of cell-permeable molecules for the control of gene-specific regulation *in vivo*. □

Received 22 March; accepted 25 June 1996.

1. Wade, W. S., Mrksich, M. & Dervan, P. B. *J. Am. chem. Soc.* **114,** 8783–8794 (1992).
2. Mrksich, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **89,** 7586–7590 (1992).
3. Wade, W. S., Mrksich, M. & Dervan, P. B. *Biochemistry* **32,** 11385–11389 (1993).
4. Pelton, J. G. & Wemmer, D. E. *Proc. natn. Acad. Sci. U.S.A.* **86,** 5723–5727 (1989).
5. Pelton, J. G. & Wemmer, D. E. *J. Am. chem. Soc.* **112,** 1393–1399 (1990).
6. Mrksich, M. & Dervan, P. B. *J. Am. chem. Soc.* **115,** 2572–2576 (1993).
7. Geierstanger, B. H., Jacobsen, J. P., Mrksich, M., Dervan, P. B. & Wemmer, D. E. *Biochemistry* **33,** 3055–3062 (1994).
8. Geierstanger, B. H., Dwyer, T. J., Bathini, Y., Lown, J. W. & Wemmer, D. E. *J. Am. chem. Soc.* **115,** 4474–4482 (1993).
9. Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. *Science* **266,** 646–650 (1994).
10. Mrksich, M. & Dervan, P. B. *J. Am. chem. Soc.* **117,** 3325–3332 (1995).
11. Mrksich, M. & Dervan, P. B. *J. Am. chem. Soc.* **115,** 9892–9899 (1993).
12. Dwyer, T. J., Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. *J. Am. chem. Soc.* **115,** 9900–9906 (1993).
13. Mrksich, M. & Dervan, P. B. *J. Am. chem. Soc.* **116,** 3663–3664 (1994).
14. Mrksich, M., Parks, M. E. & Dervan, P. B. *J. Am. chem. Soc.* **116,** 7983–7988 (1994).
15. Chen, Y. H. & Lown, J. W. *J. Am. chem. Soc.* **116,** 6995–7005 (1994).
16. Cho, J. Y., Parks, M. E. & Dervan, P. B. *Proc. natn. Acad. Sci. U.S.A.* **92,** 10389–10392 (1995).
17. Baird, E. E. & Dervan, P. B. *J. Am. chem. Soc.* **118,** 6141–6146 (1996).
18. Galas, D. & Schmitz, A. *Nucleic Acids Res.* **5,** 3157–3170 (1978).
19. Fox, K. R. & Waring, M. J. *Nucleic Acids Res.* **12,** 9271–9285 (1984).
20. Brenowitz, M., Senear, D. F., Shea, M. A. & Ackers, G. K. *Meth. Enzym.* **130,** 132–181 (1986).
21. Steitz, T. A. *Rev. Biophys.* **23,** 205–280 (1990).
22. Desjarlais, J. R. & Berg, J. M. *Proc. natn. Acad. Sci. U.S.A.* **89,** 7345–7349 (1992).
23. Desjarlais, J. R. & Berg, J. M. *Proc. natn. Acad. Sci. U.S.A.* **90,** 2256–2260 (1993).
24. Jamieson, A. C., Kim, S.-H. & Wells, J. A. *Biochemistry* **33,** 5689–5695 (1994).
25. Rebar, R. J. & Pabo, C. O. *Science* **263,** 671–673 (1994).
26. Choo, Y. & Klug, A. *Proc. natn. Acad. Sci. U.S.A.* **91,** 11163–11167 (1994).
27. Choo, Y. & Klug, A. *Proc. natn. Acad. Sci. U.S.A.* **91,** 11168–11172 (1994).
28. Letovsky, J. & Dynan, W. S. *Nucleic Acids Res.* **17,** 2639–2653 (1989).
29. Maxam, A. M. & Gilbert, W. S. *Meth. Enzym.* **65,** 499–560 (1980).
30. Iverson, B. L. & Dervan, P. B. *Nucleic Acids Res.* **15,** 7823–7830 (1987).

ACKNOWLEDGEMENTS. We are grateful to the NIH for research support, to the NSF for a predoctoral fellowship to J.W.T., and to the HHMI for a predoctoral fellowship to E.E.B.

CORRESPONDENCE and requests for materials should be addressed to P.B.D. (e-mail: dervan@starbase1.caltech.edu).