Recognition of 16 Base Pairs in the Minor Groove of DNA by a Pyrrole−Imidazole Polyamide Dimer

John W. Trauger, Eldon E. Baird, and Peter B. Dervan*

Arnold and Mabel Beckman Laboratories of Chemical Synthesis
Division of Chemistry and Chemical Engineering
California Institute of Technology,
Pasadena, California 91125

Received January 5, 1998

Cell-permeable small molecules that bind predetermined DNA sequences with affinities and specificities comparable to those of natural DNA-binding proteins have the potential to regulate the expression of specific genes. Recently, an eight-ring hairpin pyrrole−imidazole (Py−Im) polyamide which binds six base pairs of DNA was shown to inhibit transcription of a specific gene in cell culture.[1] Polyamides recognizing longer DNA sequences should provide more specific biological activity. To specify a single site within the 3 billion base pair human genome, ligands which specifically recognize 15−16 base pairs are necessary.[2] For this reason, recognition of 16 base pairs represents a milestone in the development of chemical approaches to DNA recognition.[2,3] We examine here the affinity and specificity of a Py−Im polyamide dimer which targets 16 contiguous base pairs in the minor groove of DNA.[4]

As the length of a polyamide dimer having the general sequence $ImPy_{2-6}$ increases beyond five rings (corresponding to a seven base pair binding site), the DNA-binding affinity ceases to increase with polyamide length.[5] A structural basis for this observation is provided by the recently determined X-ray crystal structure structure of a four-ring homodimer in complex with DNA, which reveals a perfect match of polyamide rise-per-residue with the pitch of the DNA duplex, but overwound ligand curvature.[6] The curvature mismatch explains the observation that flexible β-alanine residues reset an optimum fit of polyamide dimers with the DNA helix at long binding sites.[7]

We chose as our binding site the 16 base pair sequence 5′-ATAAGCAGCTGCTTTT-3′ present in the regulatory region of the HIV-1 genome.[8] Consideration of the previously published polyamide ring pairing rules,[9,10] the A,T specificity of β/β pairs,[7] and the "slipped" dimer motif[7a,11] suggested that the eight-ring polyamide ImPy-β-ImPy-β-ImPy-β-PyPy-β-Dp (**1**) would bind the target sequence as a cooperative antiparallel dimer (Figure 1). The polyamide was synthesized using solid-phase methods[12] and purified by HPLC, and its identity and purity confirmed by ^{1}H NMR, analytical HPLC, and MALDI-TOF MS. A quantitative DNase I footprinting experiment[13] carried out on a 245 base pair 3′-^{32}P-end-labeled restriction fragment revealed that the polyamide specifically binds it target site at subnanomolar concentrations (apparent monomeric association constant, $K_a \geq 3.5 \times 10^{10}$ M^{-1}) (Figure 2a).[14,15] The binding data were well-fit by a cooperative binding isotherm, consistent with formation of a cooperative 2:1 complex.[7a,14] To provide further evidence that **1** binds as an extended dimer, an affinity cleavage experiment[2a] was carried out with the polyamide-EDTA·Fe(II) conjugate **1-E** (Figure 2b). Cleavage was observed at each end of the match sequence, consistent with a dimeric, antiparallel binding mode. With regard to sequence specificity, there is a proximal two base pair mismatch site, 5′-**c**AGATGCTGCAT**A**Ta-3′, to the 5′ side of the ^{32}P-labeled strand which is bound with at least 35-fold lower affinity than the match site. However, other mismatch sites on the restriction fragment are bound with 10−20-fold lower affinity, revealing limitations of this first effort at 16 base pair recognition. Undoubtedly there is ample room for further optimization of sequence specificity.

The results reported here have two implications. First, the high binding affinity and the affinity cleavage pattern observed for the 16 base pair polyamide·DNA complex suggests that eight pairs of amide residues form a fully overlapped core which properly positions the six Im−Py pairs for recognition of six G,C base pairs and two β/β pairs for recognition of two A,T base pairs. Polyamides composed of two-ring subunits connected by β-alanine appear to be isohelical with B-DNA and allow placement of imidazole residues at any ring position, thus providing a generalizable motif for recognition of predetermined DNA sequences. Importantly, these results demonstrate that polyamides of similar size to those shown to permeate cells (i.e., MW ∼ 1200) can now bind 16 base pairs of DNA at subnanomolar concentrations, paving the way for investigation of the optimal polyamide binding site size required for specific biological activity. However, the specificity is suboptimal and this is likely a minimum

(1) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202−205.

(2) (a) Dervan, P. B. *Science* **1986**, *232*, 464. (b) Dervan, P. B. In *The Robert A. Welch Foundation Conference on Chemical Research XXXI. Design of Enzymes and Enzyme Models*; Houston, TX, November 2−4, 1987; pp 93−109. (c) Dervan, P. B. In *Nucleic Acids and Molecular Biology*; Springer-Verlag: Heidelberg, Germany, 1988; Vol. 2, pp 49−64.

(3) (a) Moser, H. E.; Dervan, P. B. *Science* **1987**, 238, 645−650. (b) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J. L.; Thoung, N. T.; Lhomme, J.; Helene, C. *Nucleic Acids Res.* **1987**, *15*, 7749. (c) Strobel, S. A.; Doucettestamm, L. A.; Riba, L.; Housman, D. E.; Dervan, P. B. *Science* **1991**, *254*, 1639−1642. (d) Thuong, N. T.; Helene, C. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 666−690.

(4) For early examples of minor groove recognition of extended A,T tracts based on the 1:1 distamycin−DNA model, see: (a) Youngquist, R. S.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2565−2569. (b) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 7564−7566.

(5) Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981−6985.

(6) Keilkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104−109.

(7) β/β pairs specify for A,T base pairs: (a) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160−6166. (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1997**, *3*, 1600−1607.

(8) (a) Jones, K. A.; Peterlin, B. M. *Annu. Rev. Biochem.* **1994**, *63*, 717−743. (b) Frech, K.; Brack-Werner, R.; Werner, T. *Virology* **1996**, *224*, 256−267.

(9) Im/Py and Py/Im pairs specify for G·C and C·G, respectively: (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783−8794. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geirstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586−7590. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385−11389. (d) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474−4482. (e) White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569−578.

(10) Py/Py pairs specify for A,T base pairs: (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723−5727. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393−1399. (c) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Nat. Struct. Biol.* **1994**, *1*, 169−175. (d) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532−12537.

(11) (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Nat. Struct. Biol.* **1996**, *3*, 321−324. (b) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1996**, *3*, 369−377.

(12) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141−6146.

(13) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132−181.

(14) For the treatment of data on cooperative association of ligands, see: Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules*; W. H. Freeman: New York, NY, 1980; p 863.

(15) The method used here for determining association constants involves the assumption that $[L]_{tot} \approx [L]_{free}$, where $[L]_{free}$ is the concentration of polyamide free in solution (unbound). For very high association constants, this assumption becomes invalid, resulting in underestimated association constants. In the experiments described here, the DNA concentration is estimated to be ∼5 pM. As a consequence, apparent association constants greater than $(1-2) \times 10^{10}$ M^{-1} represent a lower limit of the true association constant.

S0002-7863(98)00037-7 CCC: $15.00 © 1998 American Chemical Society
Published on Web 03/28/1998

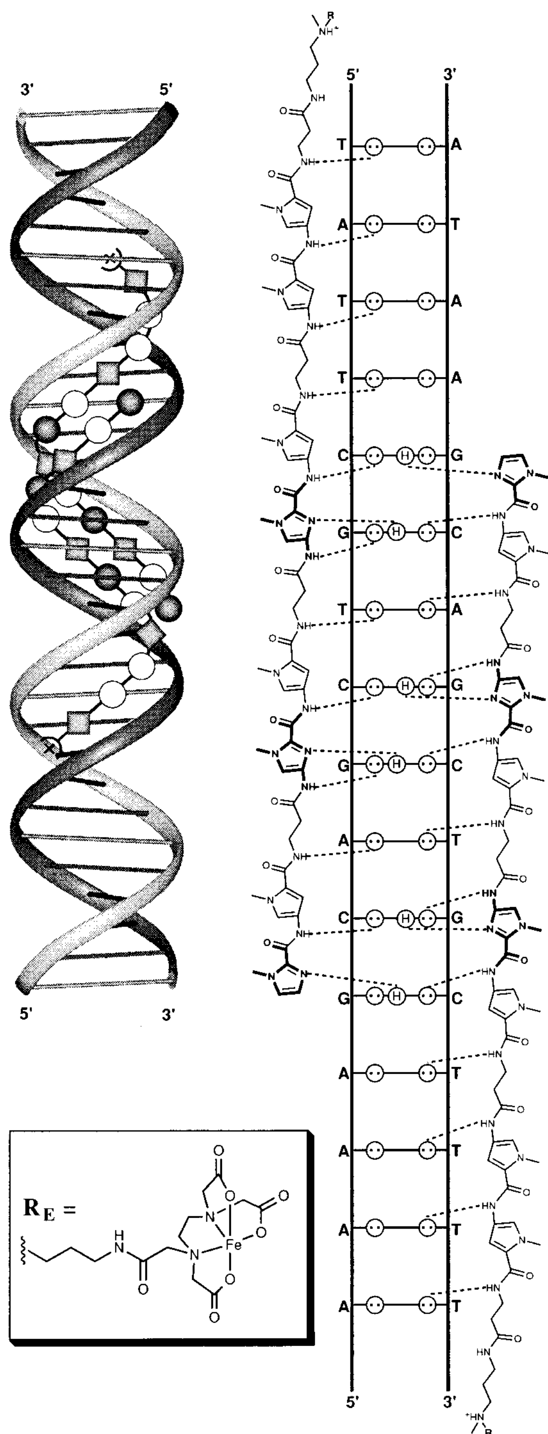

Figure 1. Model of the complex of ImPy-β-ImPy-β-ImPy-β-PyPy-β-Dp (**1**, R = H) or ImPy-β-ImPy-β-ImPy-β-PyPy-β-Dp-EDTA•Fe(II) (**1-E**, R = R_E) (Im = *N*-methylimidazole, Py = *N*-methylpyrrole, β = β-alanine, Dp = dimethylaminopropylamide) with 5′-ATAAGCAGCT-GCTTTT-3′. The shaded and open circles represent imidazole and pyrrole rings, respectively, and the diamonds represent β-alanine. Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines, and circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dashed lines.

step forward. These data provide a baseline for comparison with other cooperative binding motifs which combine minimum polyamide size (for cell permeation) with maximum DNA sequence size (for biological specificity) which will be reported in due course.[16]

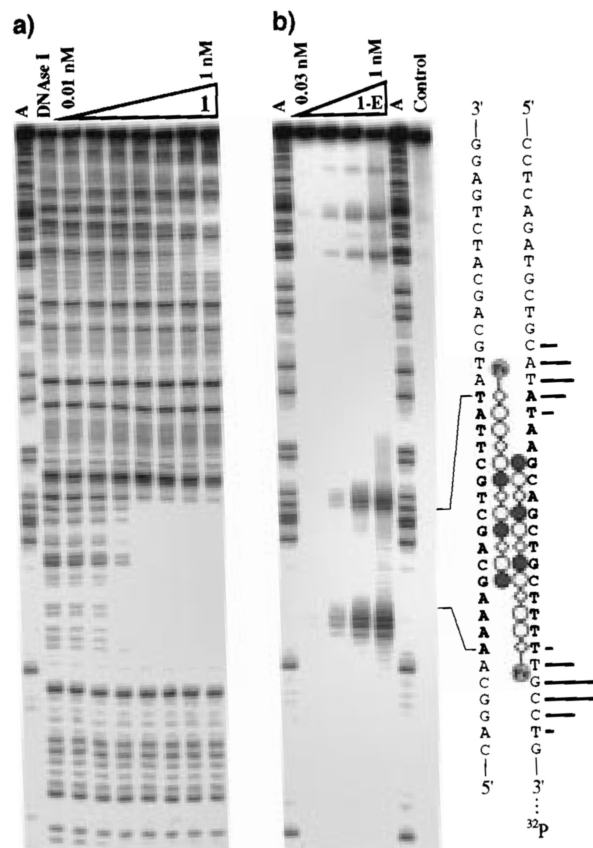

Figure 2. (a) Storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel used to separate the fragments generated by DNase I digestion in a quantitative footprint titration experiment with polyamide **1**: lane 1, A lane; lane 2, DNase I digestion products obtained in the absence of polyamide; lanes 3−9, DNase I digestion products obtained in the presence of 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, and 1 nM polyamide **1**, respectively. All reactions contain 3′-^{32}P-end-labeled *Eco*RI/*Hin*dIII restriction fragment from plasmid pJT-LTR (15 kcpm), 10 mM Tris•-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ (pH 7.0, 24 °C). (b) Autoradiogram of a gel used to separate the fragments generated by an affinity cleavage reaction using polyamide **1-E**: lanes 1 and 6, A sequencing lanes; lanes 2−5, cleavage products obtained in the presence of 0.03, 0.1, 0.3, and 1 nM **1-E**, respectively; lane 7, intact DNA. All reactions contain labeled restriction fragment (7 kcpm), 20 mM HEPES, 300 mM NaCl, 50 μg/mL glycogen, 1 μM Fe(II), and 5 mM DTT (pH 7.3, 24 °C). The sequence of the restriction fragment in the region of the 16 base pair target site and a model of the $(\mathbf{1\text{-}E})_2$•DNA complex are shown along the right side of the autoradiogram. Line heights are proportional to the observed cleavage intensity at the indicated base.

Acknowledgment. We are grateful to the National Institutes of Health (Grant GM-51747) for research support, to the National Science Foundation and the Ralph M. Parsons Foundation for predoctoral fellowships to J.W.T., and to the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B.

JA9800378

(16) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Angew. Chem., Int. Ed. Engl.*, in press.