Aliphatic/Aromatic Amino Acid Pairings for Polyamide Recognition in the Minor Groove of DNA

James M. Turner, Susanne E. Swalley, Eldon E. Baird, and Peter B. Dervan*

Contribution from the Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125

Received January 13, 1998

Abstract: Selective placement of an aliphatic β-alanine (β) residue paired side-by-side with either a pyrrole (Py) or imidazole (Im) aromatic amino acid is found to compensate for sequence composition effects for recognition of the minor groove of DNA by hairpin pyrrole−imidazole polyamides. A series of polyamides were prepared which contain pyrrole and imidazole aromatic amino acids, as well as γ-aminobutyric acid (γ) "turn" and β-alanine "spring" aliphatic amino acid residues. The binding affinities and specificities of these polyamides are regulated by the placement of paired β/β, Py/β, and Im/β residues. Quantitative footprint titrations demonstrate that replacing two Py/Py pairings in a 12-ring hairpin (6-γ-6) with two Py/β pairings affords 10-fold enhanced affinity and similar sequence specificity for an 8-bp target sequence. The 6-γ-6 hairpin ImPyImPyPyPy-γ-ImPyPyPyPyPy-β-Dp, which contains six consecutive amino acid pairings, is unable to discriminate a single-base-pair mismatch site 5′-TGT**T**AACA-3′ from a 5′-TGTGAACA-3′ match site. The hairpin polyamide Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-β-Dp binds to the 8-bp match sequence 5′-TGTGAACA-3′ with an equilibrium association constant of $K_a = 2.4 \times 10^{10}$ M^{-1} and ≥48-fold specificity versus the 5′-TGT**T**AACA-3′ single-base-pair mismatch site. Modeling indicates that the β-alanine residue relaxes ligand curvature, providing for optimal hydrogen bond formation between the floor of the minor groove and both Im residues within the Im-β-Im polyamide subunit. This observation provided the basis for design of a hairpin polyamide, Im-β-ImPy-γ-Im-β-ImPy-β-Dp, which incorporates Im/β pairings to recognize a "problematic" 5′-GCGC-3′ sequence at subnanomolar concentrations. These results identify Im/β and β/Im pairings that respectively discriminate G·C and C·G from A·T/T·A as well as Py/β and β/Py pairings that discriminate A·T/T·A from G·C/C·G. These aliphatic/aromatic amino acid pairings will facilitate the design of hairpin polyamides which recognize both a larger binding site size as well as a more diverse sequence repertoire.

Polyamides containing *N*-methylpyrrole and *N*-methylimidazole amino acids are synthetic ligands that have an affinity and specificity for DNA comparable to those of naturally occurring DNA binding proteins.[1] DNA recognition depends on side-by-side aromatic amino acid pairings in the minor groove. Antiparallel pairing of imidazole (Im) opposite pyrrole (Py) recognizes a G·C base pair, while a Py/Im combination recognizes C·G.[2] A Py/Py pair is degenerate and recognizes either an A·T or T·A base pair.[2,3] Eight-ring pyrrole−imidazole polyamides have been shown to be cell permeable and to inhibit transcription of designated genes in cell culture.[4] This provides impetus to develop an ensemble of motifs which recognize a broad binding site size and sequence repertoire.

Given the sequence-dependent microstructure of the DNA helix,[5] it is surprising that a simple recognition code can be developed at all.[6] In both published and unpublished work, over 100 pyrrole−imidazole polyamides have been synthesized which recognize predetermined sequences. However, within this group, certain difficult Py-Im/DNA base-pair sequences have emerged. Sequence-dependent DNA structure features such as intrinsic minor groove width, minor groove flexibility, and inherent DNA curvature may reduce polyamide binding at certain sites.[5] However, it may also be possible to identify polyamide structural elements which will restore affinity at difficult sequences by providing an optimal fit between the hydrogen bond donors and acceptors displayed on the edges of both the Watson−Crick base pairs and the crescent-shaped polyamide dimers.

Hairpin Polyamide. Efforts have been made to increase DNA-binding affinity and sequence specificity by covalently linking polyamide heterodimers and homodimers.[1,7] A head-

(1) (a) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559. (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953. (c) Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 7636.

(2) (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385. (d) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (e) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.

(3) (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingham, M. *Nat. Struct. Biol.* **1994**, *1*, 169. (d) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532.

(4) Gottesfield, J. M.; Nealy, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202.

(5) (a) Wu, H.; Crothers, D. M. *Nature* **1984**, *308*, 509. (b) Steitz, T. A. *Annu. Rev. Biophys.* **1990**, *23*, 205. (c) Goodsell, D. S.; Kopka, M. L.; Cascio, D.; Dickerson, R. E. *Proc. Natl. Sci. U.S.A.* **1993**, *90*, 2930. (d) Paolella, D. N.; Palmer, R.; Schepartz, A. *Science* **1994**, *264*, 1130. (e) Kahn, J. D.; Yun, E.; Crothers, D. M. *Nature* **1994**, *368* 163. (f) Geierstanger, B. H.; Wemmer, D. E. *Annu. Rev. Biochem.* **1995**, *24*, 463. (g) Hansen, M. R.; Hurley, L. H. *Acc. Chem. Res.* **1996**, *29*, 249.

(6) (a) Dervan, P. B. *Science* **1986**, *232*, 464. (b) Dervan, P. B. In *Nucleic Acids and Molecular Biology*; Springer-Verlag: Heidelberg, 1988; Vol. 2, 49−64. (c) Thuong, N. T.; Helene, C. *Angew. Chem.* **1993**, *32*, 666. (d) Nielsen, P. E. *Chem. Eur. J.* **1997**, *3*, 505.

S0002-7863(98)00147-4 CCC: $15.00
Published on Web 06/12/1998

to-tail linked polyamide with γ-aminobutyric acid (γ) serving as a turn-specific internal guide residue binds to predetermined target sites with 100-fold enhanced affinity relative to unlinked subunits.[1,7] In contrast to unlinked dimers which can adopt a variety of "slipped" binding modes, *the hairpin structure locks the relative positions of the individual subunits* and allows greater control of amino acid ring pairings. Three-, four-, and five-ring polyamides covalently coupled to form six-, eight-, and 10-ring hairpin structures bind specifically to 5-, 6-, and 7-bp target sequences, respectively.[1,7] Recognition of 7 bps by a 10-ring hairpin represents an upper limit of five contiguous ring pairings which will match the curvature of the DNA helix without energetic penalty.[1c,8,9] For complexes of fully overlapped 2:1 polyamide dimers in the minor groove, a flexible β-alanine (β) spring was found to form A,T-specific β/β pairings which were necessary for recognition of longer binding sites.[10] It remained to be determined (i) if paired β-alanine residues could be accommodated in the hairpin structure to restore register of the hairpin with the DNA helix for recognition of larger binding sites and (ii) if the antiparallel polyamide dimer could accommodate aromatic amino acids paired side-by-side with aliphatic β-amino acids to give β/Py and β/Im ring pairings (Figures 1 and 2).

Eight polyamides were synthesized containing either solely ring amino acids, a side-by-side pairing of two β-alanine residues (β/β), or a side-by-side pairing of a pyrrole and a β-alanine residue (Py/β).[11] We report here the DNA-binding affinity and sequence selectivity of the eight hairpin polyamides, ImPyPyPyPyPy-γ-ImPyPyPyPyPy-β-Dp (**1**), ImPyPyPy-β-Py-γ-Im-β-PyPyPyPy-β-Dp (**2**), ImPyPy-β-PyPy-γ-ImPy-β-Py-PyPy-β-Dp (**3**), ImPy-β-PyPyPy-γ-ImPyPy-β-PyPy-β-Dp (**4**), Im-β-PyPyPyPy-γ-ImPyPyPy-β-Py-β-Dp (**5**), Im-β-PyPyPyPy-γ-Im-β-PyPyPyPy-β-Dp (**6**), ImPy-β-PyPyPy-γ-ImPy-β-Py-PyPy-β-Dp (**7**), and ImPy-β-β-PyPy-γ-ImPy-β-β-PyPy-β-Dp (**8**) for an 8-bp 5′-TGTTAACA-3′ target sequence and a 5′-TGT**G**AACA-3′ single-base-pair mismatch sequence. Polyamides **2**−**5** vary the position of a single (β/β) pairing, while polyamides **6** and **7** have two (Py/β) pairings. As a control, polyamide **8** was synthesized with two consecutive (β/β) pairings (Figure 4). Additional polyamides ImPyImPyPyPy-γ-ImPyPyPyPyPy-β-Dp (**9**) and Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-β-Dp (**10**) were designed to target the 8-bp sequence 5′-TGTGAACA-3′. A polyamide, Im-β-ImPy-γ-Im-β-ImPy-β-Dp (**12**), which incorporates Im/β and β/Im pairings, was also synthesized, and its DNA binding affinity and specificity were determined for a 5′-TGCGCA-3′ site. Two separate techniques were used to characterize the DNA-binding properties of the designed polyamides: methidiumpropyl-EDTA·Fe(II) (MPE·Fe(II)) footprinting[12] and DNase I footprinting.[13] Information about precise binding site size is gained from MPE·Fe(II) footprinting, while quantitative DNase I footprint titrations allow determination of equilibrium association constants (K_a) of the polyamides for respective match and mismatch binding sites.

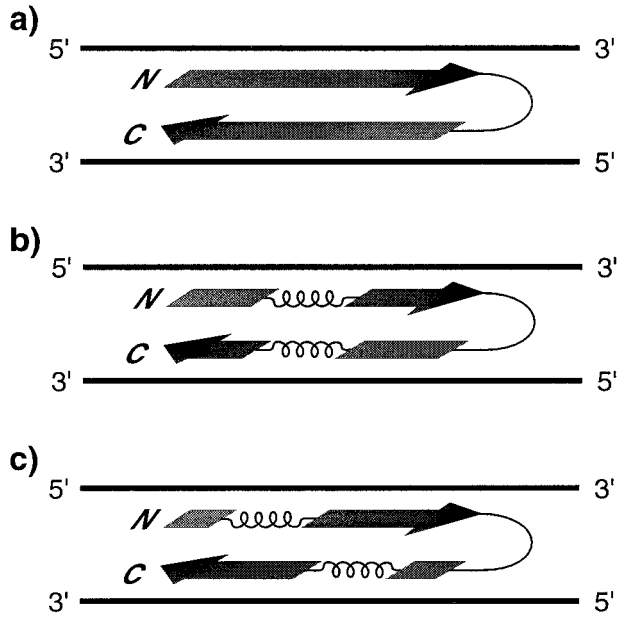


Figure 1. Ribbon model for the use of aliphatic amino acids in hairpin imidazole−pyrrole polyamides. Arrows represent aromatic imidazole and pyrrole amino acids, while springs represent β-alanine residues. (a) Six consecutive aromatic amino acid pairings. (b) One β/β pairing. (c) Two aromatic/β pairings.

Results and Discussion

Placement of Py/β and β/β Pairings. A series of eight hairpin polyamides were synthesized using Boc-chemistry machine-assisted protocols (Figure 3).[11] MPE·Fe(II) footprinting on a ^{32}P end-labeled 254-bp DNA restriction fragment (25 mM Tris-acetate, 10 mM NaCl, 100 mM calf thymus DNA, pH 7.0, 22 °C) reveals that each polyamide is binding to the 5′-TGTTAACA-3′ match site (Figure 4 and Supporting Information).[12] Footprinting patterns reveal asymmetrically 3′-shifted protection of the 8-bp sites, consistent with formation of a 1:1 hairpin polyamide−DNA complex in the minor groove. Polyamide **1** at 10 μM concentration protects both the 5′-TGTTAACA-3′ match site and the single-base-pair mismatch 5′-TGT***G***AACA-3′ site. Polyamides **4** and **7** each at 10 μM protect their cognate 5′-TGTTAACA-3′ match site; however, no protection is observed at the 5′-TGT***G***AACA-3′ single-base-pair mismatch site. Quantitative DNase I footprint titration experiments[13] (10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0, 22 °C) were performed to determine the equilibrium association constants of the polyamides for the two bound sites (Table 1).[13] The 5′-TGTTAACA-3′ match site was bound by the polyamides with decreasing affinity: ImPy-β-PyPyPy-γ-ImPyPy-β-PyPy-β-Dp (**4**, $K_a = 1.2\ (\pm 0.1) \times 10^{11}$ M^{-1}) > Im-β-PyPyPyPy-γ-Im-β-PyPyPyPy-β-Dp (**6**, $K_a = 4.5\ (\pm 2.7) \times 10^{10}$ M^{-1}) > ImPy-β-PyPyPy-γ-ImPy-β-PyPyPy-β-

(7) (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983. (b) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147. (c) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6153. (d) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1996**, *3*, 369. (e) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198. (f) Pilch, D. S.; Pokar, N. A.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306. (g) de Claire, R. P. L.; Geierstanger B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909. (h) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756. (i) White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569.

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

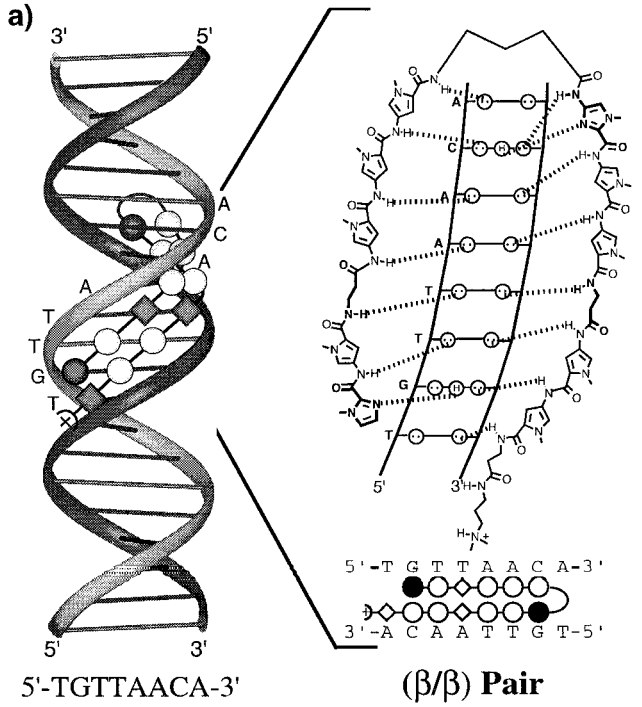
(9) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Biol.* **1988**, *5*, 104.

(10) (a) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan. P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160. (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1997**, *3*, 1600. (c) Trauger, J. W.; Baird, E. E.; Dervan. P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534.

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

(12) (a) Van Dyke, M. W.; Dervan, P. B. *Nucleic Acids Res.* **1983**, *11*, 5555. (b) Van Dyke, M. W.; Dervan, P. B. *Science* **1984**, *225*, 1122.

(13) (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8462. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.



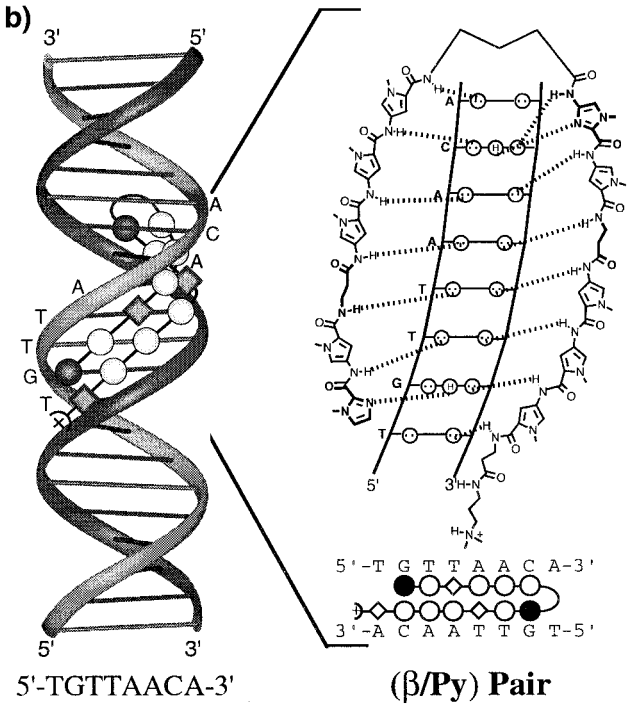


Figure 2. Binding model for the complexes formed between the DNA and (a) ImPy-β-PyPyPy-γ-ImPyPy-β-PyPy-β-Dp; and (b) ImPy-β-PyPyPy-γ-ImPy-β-PyPyPy-β-Dp. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. Ball and stick models are also shown. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent β-alanine residues.

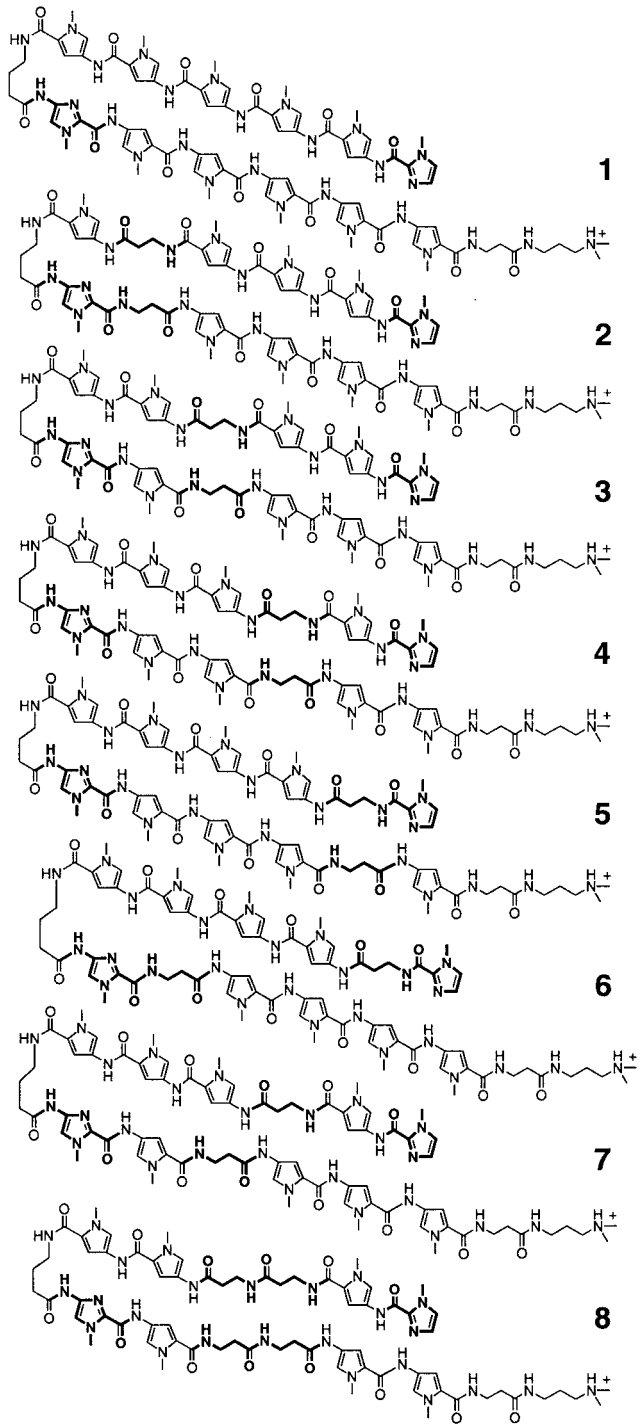


Figure 3. Structures of the hairpin polyamides **1**−**8** synthesized by solid-phase methods. Polyamides were synthesized using Boc-chemistry machine-assisted solid-phase protocols.[11] The trifluoroacetate salts of polyamides **2**−**8** are soluble in aqueous solutions at concentrations of ≤1 mM. Polyamide **1** is soluble at ≤0.1 mM.

Dp (**7**, K_a = 2.7 (±1.5) × 10^{10} M^{-1}) > ImPyPy-β-PyPy-γ-ImPy-β-PyPyPy-β-Dp (**3**, K_a = 1.7 (±0.4) × 10^{10} M^{-1}) > Im-β-PyPyPyPy-γ-ImPyPyPy-β-Py-β-Dp (**5**, K_a = 6.6 (±1.9) × 10^9 M^{-1}) > ImPyPyPyPyPy-γ-ImPyPyPyPyPy-β-Dp (**1**, K_a = 2.5 (±0.6) × 10^9 M^{-1}) > ImPyPyPy-β-Py-γ-Im-β-PyPyPyPy-β-Dp (**2**, K_a = 1.3 (±0.4) × 10^9 M^{-1}) > ImPy-β-β-PyPy-γ-ImPy-β-β-PyPy-β-Dp (**8**, K_a ≤ 1 × 10^8 M^{-1}). The 5′-TGT**G**AACA-3′ site was bound with decreasing affinity: Im-β-PyPyPyPy-γ-Im-β-PyPyPyPy-β-Dp (**6**, K_a = 7.7 (±4.5) × 10^9 M^{-1}) > ImPy-β-PyPyPy-γ-ImPy-β-PyPyPy-β-Dp (**7**, K_a = 5.7 (±4.1) × 10^9 M^{-1}) > ImPyPy-β-PyPy-γ-ImPy-β-PyPyPy-β-Dp (**3**, K_a = 2.7 (±1.5) × 10^9 M^{-1}) ≅ ImPy-β-PyPyPy-γ-ImPyPy-β-PyPy-β-Dp (**4**, K_a = 2.2 (±1.9) × 10^9 M^{-1}) >

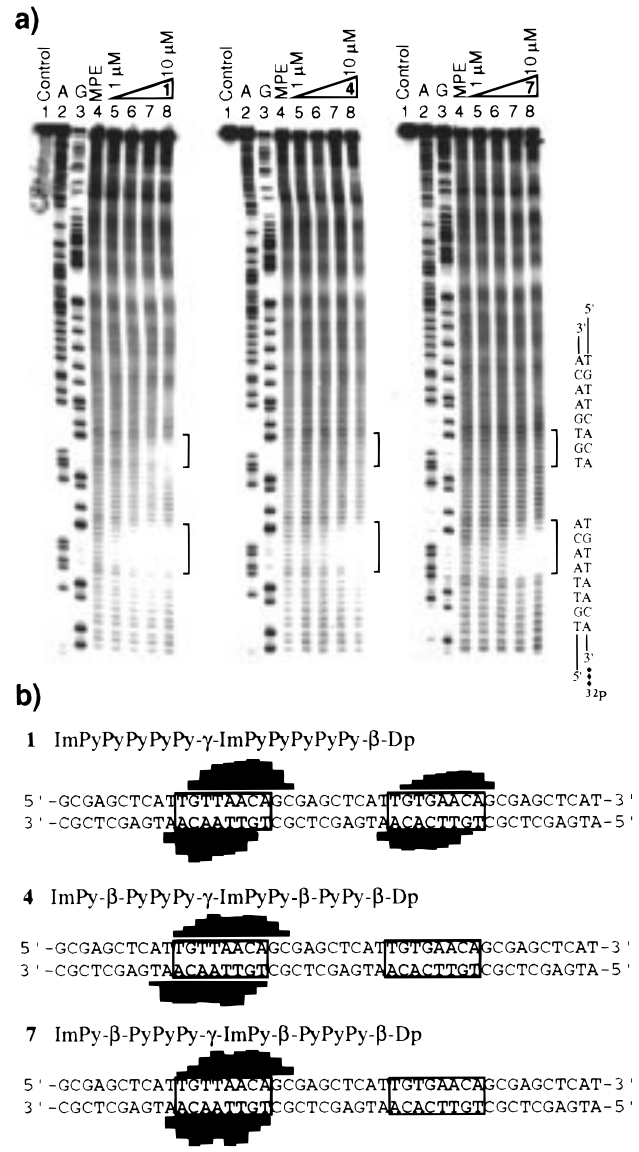
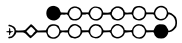
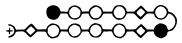
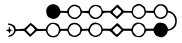
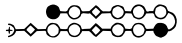
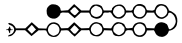
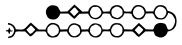
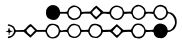
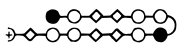


Figure 4. (a) MPE·Fe(II) footprinting on a 3′-^{32}P-labeled 254-bp *Eco*RI/*Pvu*II restriction fragment from plasmid pJK8. The 5′-TGTTAACA-3′ and 5′-TGTGAACA-3′ sites are shown on the right side of the autoradiograms: lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4, MPE·Fe(II) standard; lane 5, 1 μM **1**, **4**, or **7**; lane 6, 2 μM **1**, **4**, or **7**; lane 7, 5 μM **1**, **4**, or **7**; lane 8, 10 μM **1**, **4**, or **7**. All lanes contain 15-kcpm 3′-radiolabeled DNA and 25 mM Tris-acetate buffer (pH 7.0), 10 mM NaCl, and 100 μM/base pair of calf thymus DNA. (b) Results from MPE·Fe(II) footprinting of ImPyPyPyPyPy-γ-ImPyPyPyPyPy-β-Dp, ImPy-β-PyPyPy-γ-ImPyPy-β-PyPy-β-Dp, and ImPy-β-PyPyPy-γ-ImPy-β-PyPyPy-β-Dp. Boxes represent equilibrium binding sites determined by the published model. Only sites that were quantitated by DNase I footprint titrations are boxed. Bar heights are proportional to the relative protection from cleavage at each band.

ImPyPyPyPyPy-γ-ImPyPyPyPyPy-β-Dp (**1**, K_a = 3.9 ($\pm$1.7) $\times$ 10^8 M^{-1}) > Im-β-PyPyPyPy-γ-ImPyPyPy-β-Py-β-Dp (**5**, K_a = 2.5 ($\pm$0.4) $\times$ 10^8 M^{-1}) $\cong$ ImPyPyPy-β-Py-γ-Im-β-PyPy-PyPy-β-Dp (**2**, K_a = 2.0 ($\pm$0.1) $\times$ 10^8 M^{-1}) > ImPy-β-β-PyPy-γ-ImPy-β-β-PyPy-β-Dp (**8**, $K_a \leq 1 \times 10^8$ M^{-1}). Remarkably, the equilibrium association constant for the 5′-TGTTAACA-3′ match site varied 100-fold among polyamides **1**−**7**, indicating a sensitivity to the position relative to the γ-turn for placement of (β/β) or (Py/β) pairings. A structural basis for the β-alanine-mediated affinity enhancement awaits high-resolution X-ray and NMR studies which are in progress.

Table 1. Equilibrium Association Constants (M^{-1}) for Polyamides[a−c]

	Polyamide	5′-TGTTAACA-3′	5′-TGTGAACA-3′	Specificity[d]
1		2.5×10^9	3.9×10^8	6
2		1.3×10^9	2.0×10^8	7
3		$\mathbf{1.7 \times 10^{10}}$	2.7×10^9	6
4		$\mathbf{1.2 \times 10^{11}}$	2.2×10^9	**55**
5		$\mathbf{6.6 \times 10^9}$	2.5×10^8	**26**
6		$\mathbf{4.5 \times 10^{10}}$	7.7×10^9	6
7		$\mathbf{2.7 \times 10^{10}}$	5.7×10^9	5
8		$\leq 1 \times 10^8$	$\leq 1 \times 10^8$	1

[a] Values reported are the mean values obtained from three DNase I footprint titration experiments. [b] The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [c] Match site association constants and specificities higher than the parent hairpin are shown in bold type. [d] Specificity is calculated as K_a(match)/K_a(mismatch).

Sequence Specificity of Py/β and β/Py. Polyamide **1** places a Py/Py pair opposite T·A in the fourth position and A·T in the fifth position in the sequence 5′-TGT*T*AACA-3′. The mismatch sequence 5′-TGT**GA**ACA-3′ (Py/Py opposite G·C in the fourth position and A·T in the fifth position) is discriminated by 6-fold. For polyamide **7**, each of these Py/Py pairs is replaced sequentially with a β/Py or Py/β pair. Similar sequence specificity is observed for polyamides **1** and **7**, indicating that the single Py/β and β/Py pairings discriminate A·T, T·A from G·C, C·G at least as effectively as the Py/Py pair.

Recognition of 5′-GWG-3′. On the basis of the pairing rules for polyamide−DNA complexes, the sites 5′-TGTTAACA-3′ and 5′-TGTGAACA-3′ are for polyamides **1**−**8** "match" and "single-base-pair mismatch" sites, respectively. The full-ring hairpin polyamide ImPyImPyPyPy-γ-ImPyPyPyPyPy-β-Dp (**9**) (Figure 5) recognizes the same 5′-TGTGAACA-3′ match site with reduced affinity (K_a = 5 $\times$ 10^9 M^{-1}) and no specificity versus the 5′-TGT**T**AACA-3′ single-base-pair mismatch site. This may be due to excessive curvature of the ligand preventing both imidazoles of the ImPyIm subunit from making favorable contacts with the DNA.[9] The substitution of Im for Py in Im-β-PyPyPyPy-γ-ImPyPyPy-β-Py-β-Dp to give Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-β-Dp (**10**), formally the single atomic substitution of N for C−H, regulates affinity and specificity by more than an order of magnitude. MPE·Fe(II) footprinting on a ^{32}P end-labeled 254-bp DNA restriction fragment (Figure 6) reveals asymmetrically 3′-shifted protection of 8-bp sites, consistent with formation of a 1:1 hairpin polyamide−DNA complex in the minor groove. Polyamide **5** protects its cognate 5′-TGTTAACA-3′ match site at 1 μM concentration, while no protection is observed at the cognate 5′-TGT**G**AACA-3′ mismatch site. Polyamide **10** protects its cognate 5′-TGTGAACA-3′ match site at 1 μM; however, no protection is observed at the 5′-TGT**T**AACA-3′ single-base-pair mismatch site. Quantitative DNaseI footprint titrations (Figure 7) were performed

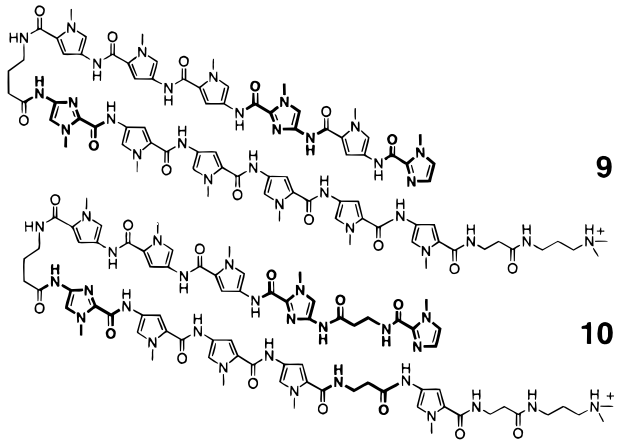


Figure 5. Structures of the hairpin polyamides ImPyImPyPyPy-γ-ImPyPyPyPyPy-β-Dp (**9**) and Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-β-Dp (**10**) synthesized by solid-phase methods. Polyamides were synthesized using Boc-chemistry machine-assisted solid-phase protocols.[11] The trifluoroacetate salt of polyamide **10** is soluble in aqueous solutions at concentrations of $\leq$1 mM. Polyamide **9** is soluble at $\leq$0.1 mM.

to determine the equilibrium association constants for polyamide **10** binding its 5′-TGTGAACA-3′ match site and 5′-TGT-**T**AACA-3′ mismatch site. Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-β-Dp (**10**) recognizes its 5′-TGTGAACA-3′ match site with $K_a = 2.4\ (\pm 0.2) \times 10^{10}\ M^{-1}$. The 5′-TGT**T**AACA-3′ mismatch sequence is bound with $\geq$48-fold ($K_a \leq 5 \times 10^8\ M^{-1}$) reduced affinity, respectively (Table 2).

The specificity and affinity of **10** suggests that all three imidazole residues make favorable DNA contacts. Previous studies have demonstrated that internal amino acids located in the third or fourth position relative to the subunit N-terminus do not form optimal hydrogen bonds.[1b] Modeling suggests that the increased affinity and specificity of hairpin **10** is due to a partial straightening of the ligand (Figure 8), allowing both imidazoles to adopt a curvature which better matches that of the DNA helix. These results identify Im-β-Im subunits as a general structure motif for the design of ligands that target certain "problematic" sequences which require an imidazole at the third position.

Im/β Pairings. In an effort to target "core" G·C-rich sequences, we had reported previously that the eight-ring hairpin polyamide ImPyImPy-γ-ImPyImPy-β-Dp (**11**) (Figure 9) binds to its 5′-TGCGCA-3′ match sequence with good specificity but with only a modest affinity ($K_a = 3.7\ (\pm 0.9) \times 10^7\ M^{-1}$) compared to other eight-ring hairpin polyamides which bind their target sites with subnanomolar affinity.[1] Because paired β-residues reset the imidazole residues of polyamide **10** for recognition of 5′-GWG-3′ sequences, it seemed plausible that Im/β and β/Im pairings could be used to optimize the positioning of the imidazole amino acids of **11** in a similar manner (Figure 10). To determine whether Im/β pairs would be G·C-specific recognition elements, the polyamide Im-β-ImPy-γ-Im-β-ImPy-β-Dp (**12**) was synthesized and its DNA binding properties compared with ImPyImPy-γ-ImPyImPy-β-Dp (**11**). MPE·Fe(II) footprinting[12] (25 mM Tris-acetate, 10 mM NaCl, 100 μM/base pair of calf thymus DNA, pH 7.0 and 22 °C) was performed on the 3′- and 5′-^{32}P end-labeled 263-bp restriction fragments from the plasmid pSES11 (Supporting Information). The assays reveal identical footprinting protection at the designed 5′-TGCGCA-3′ match site for polyamides **11** and **12** and are consistent with 6-bp recognition. Quantitative DNaseI footprint titrations were performed to determine the equilibrium association constants for polyamide **12** binding to three G,C se-

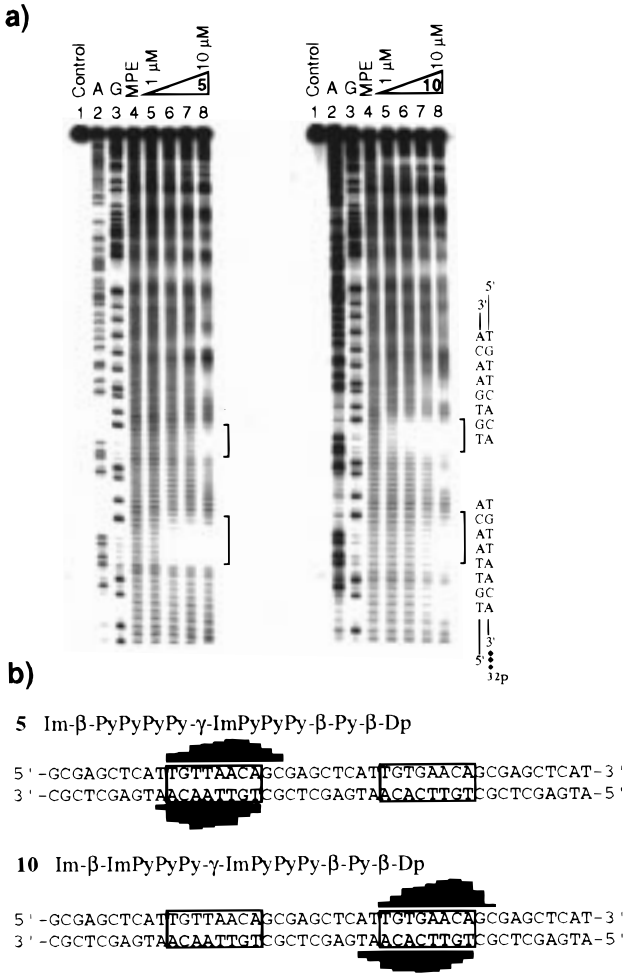


Figure 6. (a) MPE·Fe(II) footprinting on a 3′-^{32}P-labeled 254-bp *Eco*RI/*Pvu*II restriction fragment from plasmid pJK8. The 5′-TGT-TAACA-3′ and 5′-TGTGAACA-3′ sites are shown on the right side of the autoradiograms: lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4, MPE·Fe(II) standard; lane 5, 1 μM **5** or **10**; lane 6, 2 μM **5** or **10**; lane 7, 5 μM **5** or **10**; lane 8, 10 μM **5** or **10**. All lanes contain 15-kcpm 3′-radiolabeled DNA and 25 mM Tris-acetate buffer (pH 7.0), 10 mM NaCl, and 100 μM/base pair of calf thymus DNA. (b) Results from MPE·Fe(II) footprinting of Im-β-PyPyPyPy-γ-ImPyPyPy-β-Py-β-Dp and Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-β-Dp. Boxes represent equilibrium binding sites determined by the published model. Only sites that were quantitated by DNase I footprint titrations are boxed. Bar heights are proportional to the relative protection from cleavage at each band.

quences: 5′-TGCGCA-3′, 5′-TG**GC**CA-3′, and 5′-TG**G**G**G**A-3′. Im-β-ImPy-γ-Im-β-ImPy-β-Dp (**12**) recognizes its 5′-TGCGCA-3′ match site with $K_a = 3.7\ (\pm 1.5) \times 10^9\ M^{-1}$, a 100-fold increase in affinity over eight-ring polyamide **11**. (Table 3 and Supporting Information) Im/β pairings are found to be sequence specific for G·C base pairs relative to C·G, A·T, and T·A. For example, the mismatch sequences 5′-TG**GC**CA-3′ and 5′-TG**G**G**G**A-3′ are bound with 26-fold ($K_a = 1.4\ (\pm 1.0) \times 10^8\ M^{-1}$) and 34-fold ($K_a = 1.1\ (\pm 0.6) \times 10^8\ M^{-1}$) reduced affinity, respectively.

Generality. Substitution of β residues for ring residues does not always result in a gain in binding energy. For example, in unpublished results, it was determined that Im-β-PyPy-γ-Im-β-PyPy-β-Dp binds to a 5′-TGTACA-3′ match site with lower affinity than ImPyPyPy-γ-ImPyPyPy-β-Dp. Therefore, for 5′-TGTACA-3′ recognition, four-ring subunits containing exclusively ring amino acids are optimal, while for 5′-TGCGCA-3′

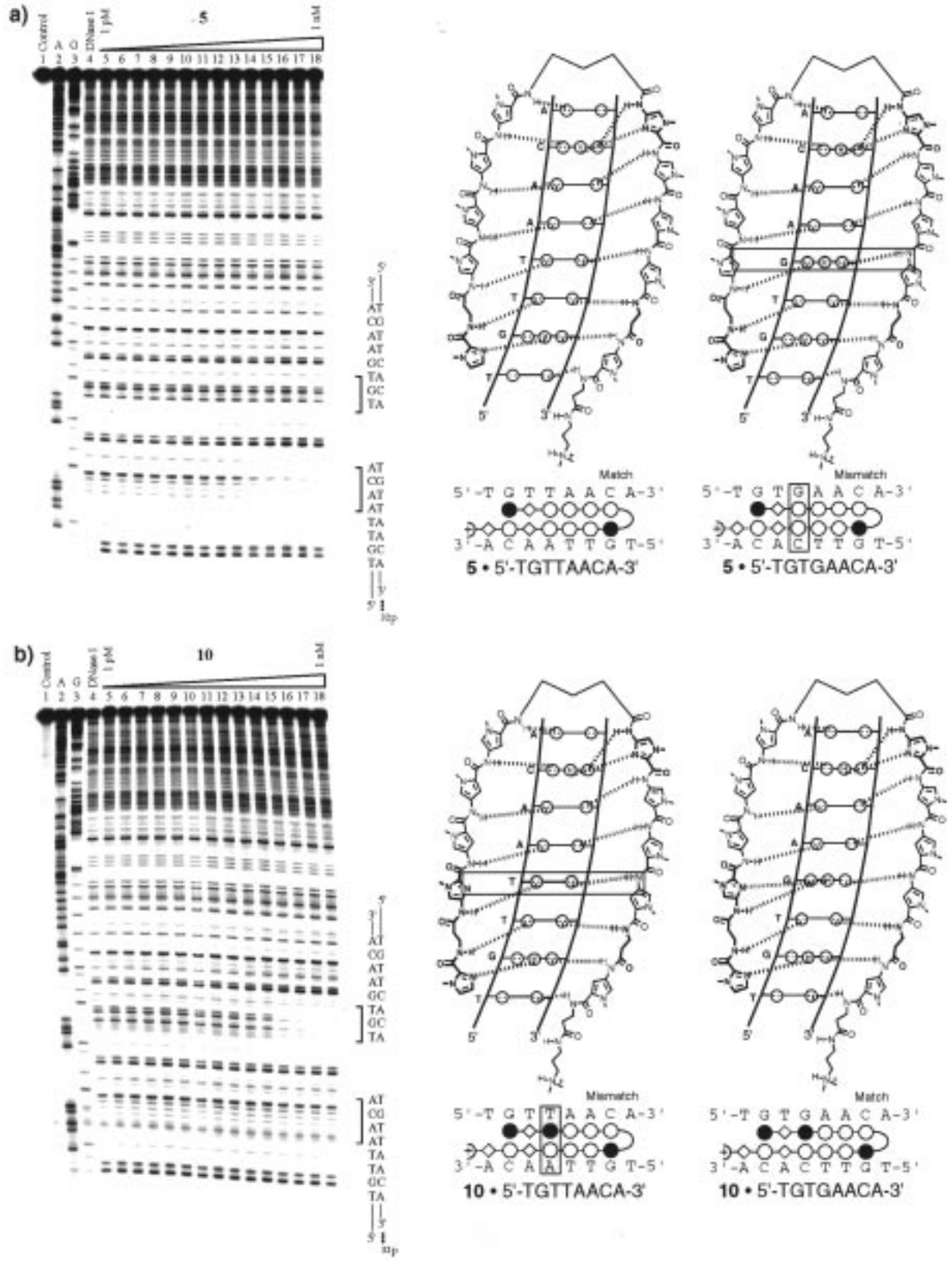
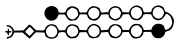
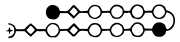
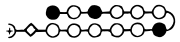
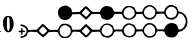


Figure 7. Quantitative DNase I footprint titration experiment with (a) Im-β-PyPyPyPy-γ-ImPyPyPy-β-Py-β-Dp (**5**) and (b) Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-β-Dp (**10**) on the 254-bp *Eco*RI/*Pvu*II restriction fragment from plasmid pJK8: lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4, DNase I standard; lanes 5−18, 1 pM, 2 pM, 5 pM, 10 pM, 15 pM, 25 pM, 40 pM, 65 pM, 100 pM, 150 pM, 250 pM, 400 pM, 650 pM, and 1 nM Im-β-PyPyPyPy-γ-ImPyPyPy-β-Py-β-Dp and Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-β-Dp. The 5′-TGTTAACA-3′ and 5′-TGTGAACA-3′ sites that were analyzed are shown on the right side of the autoradiogram. All reactions contain a 15-kcpm restriction fragment, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. Binding models are also shown. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent β-alanine residues.

Table 2. Equilibrium Association Constants (M^{-1}) for Polyamides[a–c]

	Polyamide	5'-TGTTAACA-3'	5'-TGTGAACA-3'	Specificity[d]
1		2.5×10^9	3.9×10^8	6
5		**6.6×10^9**	2.5×10^8	**26**
9		5×10^9	5×10^9	1
10		$\leq 5 \times 10^8$	**2.4×10^{10}**	**≥ 48**

[a] Values reported for **1**, **5**, and **10** are the mean values obtained from three DNase I footprint titration experiments. [b] The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [c] Match site association constants and specificities higher than parent hairpins are shown in bold type. [d] Specificity is calculated as K_a(match)/K_a(mismatch).

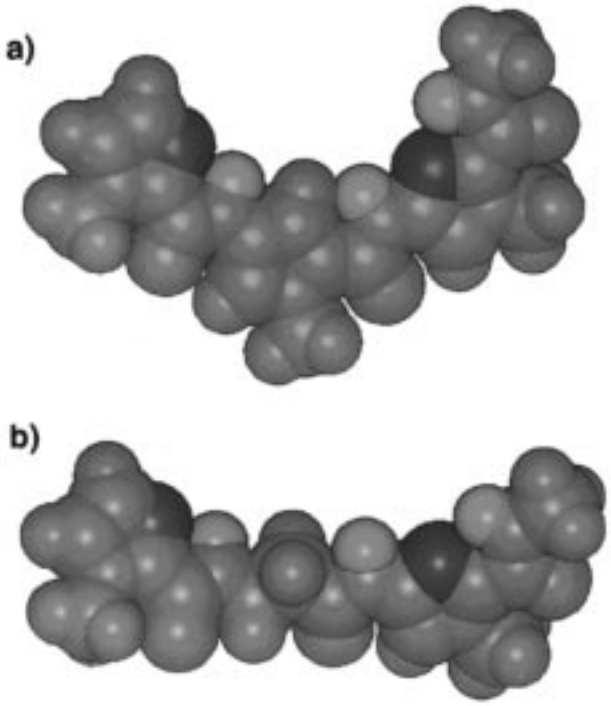
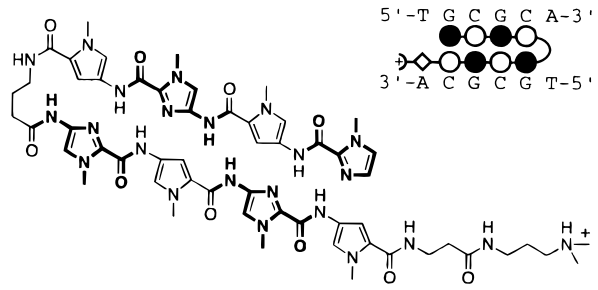


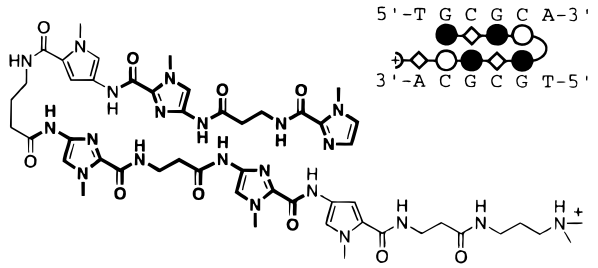
Figure 8. Model of the curvature adopted by (a) ImPyIm and (b) Im-β-Im subunits. The N3 of imidazole is shown in blue, and amide NH groups are shown in red. For simplicity, the C-terminus of both subunits is modeled as the methyl amide. Structures were minimized using MM2 calculations in Chem3D Pro v. 3.5 running on a Power Computing 180e computer. Minimized coordinates were then used to generate space-filling structures using Insight II v. 2.2 software on a Silicon Graphics workstation.

recognition, β-alanine "spring" amino acids are necessary to reset the register of the internal Im amino acids. These results indicate that the preferred polyamide template for sequence recognition is determined merely not only by site size but also by sequence composition.

Implications for the Minor Groove Binding Polyamides. The results presented here reveal that hairpin polyamides based on optimally spaced subunits provide a useful design for recognition of 8-bp binding sites in the hairpin motif. The high binding affinity of the paired (β/β) hairpin polyamides described here indicates that the *N*-methylpyrrolecarboxamide and *N*-methylimidazolecarboxamide amino acid residues make energetically favorable contact *across all eight base pairs* of the DNA helix. The specificity of polyamides which differ by a single atomic substitution, Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-



(**11**) ImPyImPy-γ-ImPyImPy-β-Dp



(**12**) Im-β-ImPy-γ-Im-β-ImPy-β-Dp

Figure 9. Structures of the hairpin polyamides ImPyImPy-γ-ImPy-ImPy-β-Dp (**11**) and Im-β-ImPy-γ-Im-β-ImPy-β-Dp (**12**) synthesized by solid-phase methods. Polyamides were synthesized using Boc-chemistry machine-assisted solid-phase protocols.[11] The trifluoroacetate salt of polyamides **11** and **12** are soluble in aqueous solutions at concentrations of ≤1 mM. Ball and stick models are also shown. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent β-alanine residues.

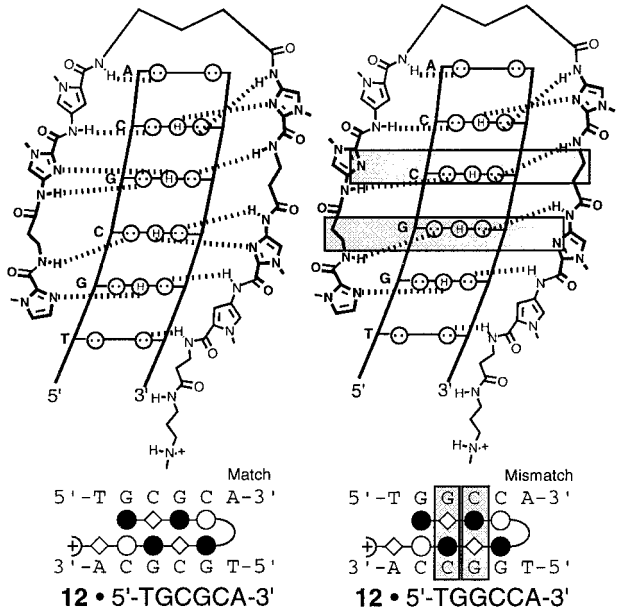
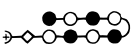
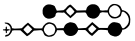


Figure 10. Binding model for the complexes formed between Im-β-ImPy-γ-Im-β-ImPy-β-Dp and the 5′-TGCGCA-3′ match and 5′-TG**GC**CA-3′ mismatch sites. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. Ball and stick models are also shown. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent β-alanine residues.

Table 3. Equilibrium Association Constants (M^{-1}) for Polyamides[a–c]

Polyamide	5'-TGCGCA-3'	5'-TGGCCA-3'	5'-TGGGGA-3'
10 	3.7×10^7	$< 10^7$	$< 10^7$
11 	3.7×10^9	1.4×10^8	1.1×10^8

[a] Values reported are the mean values obtained from a minimum of three DNase I footprint titration experiments. [b] The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [c] Equilibrium association constants for polyamide **10** are from ref 1b.

β-Dp and Im-β-PyPyPyPy-γ-ImPyPyPy-β-Py-β-Dp, for discrimination of sites which differ by a single base pair, identifies the Im-β-Im subunit as a structure motif for the design of ligands that target certain "problematic" sequences. The use of the Im-β-Im subunit in both the 6-bp (Im-β-ImPy-γ-Im-β-ImPy-β-Dp) and 8-bp (Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-β-Dp) hairpin templates indicates that β-alanine "springs" introduced as either aromatic/aliphatic or aliphatic/aliphatic pairs may offer a general approach to broaden the accessible sequence repertoire for hairpin polyamides. The reduced binding affinity of polyamide **8** suggests that consecutive side-by-side pairing of β-alanine creates unfavorable polyamide:DNA interactions. These results are consistent with the observation that motifs for recognition of longer sequences can use alternating sets of aromatic ring pairings separated by a single β/β pair.[10c] The sequence-dependent combination of rings and β-alanine required for optimal polyamide binding to longer sequences awaits further studies.

Polyamides which incorporate Py/β, β/Py, Im/β, and β/Im pairings provide the first example of aromatic/aliphatic amino acid pairings for minor groove recognition. Py/β and β/Py pairings are found to recognize A·T/T·A relative to G·C/C·G. The Im/β pairing is found to target G·C relative to C·G, A·T, and T·A, while the β/Im pair targets C·G relative to G·C, A·T, and T·A (Table 4). The results described here provide guidelines for placement of Im/β, β/Im, Py/β, β/Py, and β/β pairings within the hairpin template, setting the stage for investigation of the effects of polyamide molecular weight, affinity, and binding site size on targeted gene regulation in living cells.

Table 4. Aliphatic/Aromatic Pairing Code for Minor Groove Recognition[a]

Pair	G•C	C•G	T•A	A•T
β/β	-	-	+	+
β/Py	-	-	+	+
Py/β	-	-	+	+
Im/β	+	-	-	-
β/Im	-	+	-	-

[a] Favored (+), disfavored (−).

Experimental Section

Reagents and protocols for polyamide synthesis were as described.[1] Polyamides were purified by reversed-phase HPLC with a Waters DeltaPak 25 × 100 mm, 100-μm C18 column equipped with a guard, 0.1% (wt/v) TFA, 8.0 mL/min, 0.25% acetonitrile/min. Extinction coefficients were calculated based on $\epsilon = 8333$/ring at 304 nm.[7f]

MPE·Fe(II) and DNase I footprinting reactions were carried out as previously described.[1] Photostimulable storage phosphorimaging plates (Kodak Storage Phosphor Screen S0230 obtained from Molecular Dynamics) were pressed flat against gel samples and exposed in the dark at 22 °C for 12−16 h. A molecular dynamics 400S PhosphorImager was used to obtain all data from the storage screens. The data were analyzed by performing volume integrations of all bands using ImageQuant v. 3.2 software. All DNA manipulations were performed according to standard protocols.[14]

Preparation of 3′- and 5′-End-Labeled Restriction Fragments. The plasmids pJK8 and pSES9hp were 3′ or 5′ labeled as described.[1] Chemical sequencing reactions were performed according to published methods.[15]

Acknowledgment. We are grateful to the National Institutes of Health (GM-27681) for research support and a traineeship award (GM-08501) to S.E.S., J. Edward Richter for an undergraduate fellowship to J.M.T., and the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B.

Supporting Information Available: MPE·Fe(II) footprinting of polyamides **2**, **3**, **6**, **8**, **11**, and **12** and quantitative DNase I footprint titration experiment with **12** (6 pages, print/PDF). See any current masthead page for ordering information and Web access instructions.

JA980147E

(14) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.

(15) (a) Iverson, B. L.; Dervan, P. B. *Nucleic Acids Res.* **1987**, *15*, 7823. (b) Maxam, A. M.; Gilbert, W. S. *Methods Enzymol.* **1980**, *65*, 499.