# A Pyrrole–Imidazole Polyamide Motif for Recognition of Eleven Base Pair Sequences in the Minor Groove of DNA

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

*Dedicated to Dieter Seebach on the occasion of his 60th birthday*

**Abstract:** A new upper limit of binding site size is defined for the 2:1 overlapped polyamide:DNA motif. Eight-ring polyamides composed of four-ring subunits containing pyrrole (Py) and imidazole (Im) amino acids linked by a central β-alanine (β) spacer ("4-β-4 ligands") were designed for recognition of eleven base pair sequences as antiparallel dimer (4-β-4)$_2$·DNA complexes in the minor groove. The DNA binding properties of three polyamides, ImPyPyPy-β-PyPyPyPy-β-Dp, ImImPyPy-β-PyPyPyPy-β-Dp, and ImImImPy-β-PyPyPyPy-β-Dp, were analyzed by footprinting experiments on DNA fragments containing the respective match sites 5′-AGTAATTTACT-3′, 5′-AGGTATTACCT-3′, and 5′-AGGGATTCCCT-3′ (Dp = dimethylaminopropylamide). Quantitative footprint titrations reveal that each polyamide binds its respective target site with subnanomolar affinity and 7-fold to over 30-fold specificity over double-base-pair mismatch sites. A 20-fold decrease in binding affinity is observed for placement of a side-by-side β–β pairing opposite G·C/C·G relative to placement opposite a A·T/T·A base pair. The use of side-by-side antiparallel β-alanine residues as an A·T/T·A-specific DNA binding element provides a new pairing rule for polyamide design. Expanding the DNA binding site size targeted by pyrrole–imidazole polyamides represents an important step in the development of cell-permeable synthetic ligands for the control of gene-specific regulation.

**Keywords**
DNA recognition · hydrogen bonds · ligand design · molecular recognition · sequence-specificity

## Introduction

Small molecules specifically targeted to any predetermined DNA sequence in the human genome would be potentially useful tools in molecular biology and human medicine. Polyamides containing *N*-methylpyrrole and *N*-methylimidazole amino acids are synthetic ligands that have an affinity and specificity for DNA comparable to naturally occurring DNA binding proteins.[1] DNA recognition depends on side-by-side 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] Pyrrole–imidazole polyamides have been shown to be cell-permeable and to inhibit the transcription of specific genes in cell culture.[4] Extension of the dimer motif to target longer sequences would allow recognition of statistically more unique sites in megabase-size DNA.[5] The optimal targeted binding site size required for effective biological regulation by polyamides has yet to be determined. This provides impetus to explore the binding site size limitations of polyamide dimers for DNA recognition.

Fully overlapped 2:1 polyamide dimers containing five contiguous rings and solely pyrrole and imidazole amino acids optimally recognize a maximum binding site size of 7 base pairs.[6] Insertion of a flexible β-alanine (β) linker allows the recognition of longer sites.[7] Six-ring polyamides based on three-ring subunits linked through C–N bonds by a central β-alanine bind to nine base pair sequences as "overlapped" (3-β-3)$_2$·DNA complexes.[7] The observation that four-ring polyamides bind as 2:1 complexes with 70-fold higher affinity relative to three-ring polyamides[6] suggests that an eight-ring (4-β-4)$_2$·DNA motif would be suitable for recognition of eleven base pair sequences (Figure 1), defining a new upper limit for minor groove binding polyamide dimers.

Three 4-β-4 polyamides, ImPyPyPy-β-PyPyPyPy-β-Dp (**1**), ImImPyPy-β-PyPyPyPy-β-Dp (**2**), and ImImImPy-β-PyPyPyPy-β-Dp (**3**) (Figure 2), were synthesized by solid-phase methods.[8] We report here the affinities, relative sequence specificity, and binding site size of the three 4-β-4 polyamides as determined by MPE·Fe$^{II}$[9] and DNase I[10] footprinting. Binding site size is more accurately determined by MPE·Fe$^{II}$ footprinting, while quantitative DNase I footprint titration is more suitable for measurement of apparent equilibrium association

[*] Prof. P. B. Dervan, S. E. Swalley, E. E. Baird
Arnold and Mabel Beckman Laboratories of Chemical Synthesis
California Institute of Technology, Pasadena, CA 91101 (USA)
Fax: Int. code +(626)568-8824
e-mail: dervan@cco.caltech.edu



0947-6539/97/0310-1600 $ 17.50+.50/0

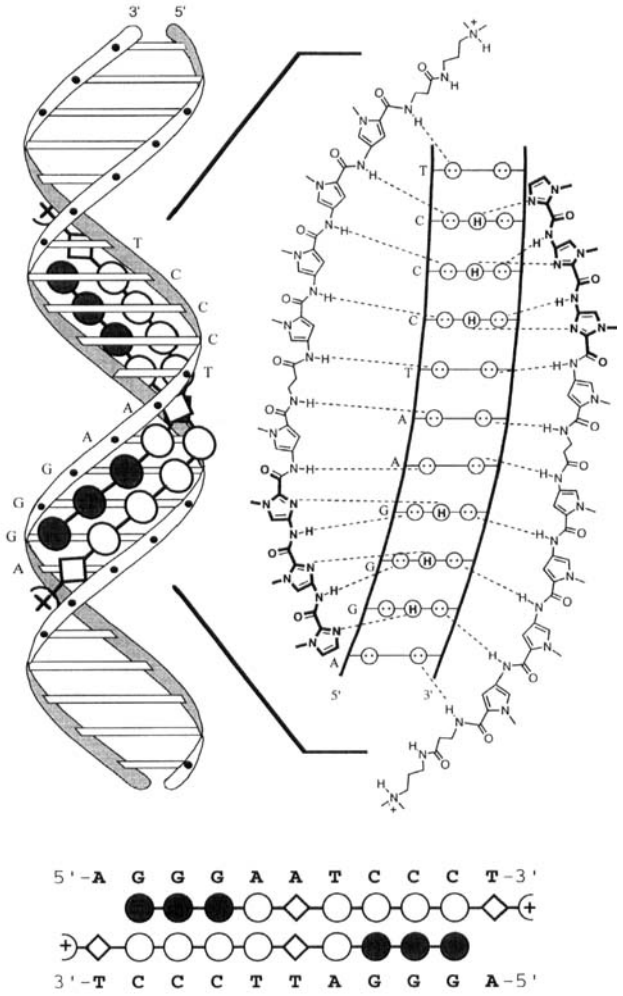


Figure 1. Models of the 2:1 polyamide:DNA complex formed with the eight-ring polyamide ImImImPy-β-PyPyPyPy-β-Dp (**3**) and the minor groove of double stranded B-form DNA. Left: Ribbon model for the 2:1 complex formed between the match sequence 5′-AGGGAATCCCT-3′ and **3**. *N*-methylimidazolecarboxamides are represented by filled circles, *N*-methylpyrrolecarboxamides by empty circles, and β-alanine amino acids unfilled triangles. Right: Hydrogen-bonding model for the 2:1 complex formed between the match sequence 5′-AGGGAATCCCT-3′ and **3**. 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. Bottom: Ball-and-stick representation of the dimerization of **3** with a 5′-AGGGAATCCCT-3′ site. The amino acids are represented as in the ribbon model.

constants ($K_a$) for the polyamide binding to match and mismatch sequences. Finally, the sequence preference of the central β–β pairing was examined by testing the polyamide ImImImPy-β-PyPyPyPy-β-Dp (**3**) against two potential 11-bp match sites, 5′-AGGGAATCCCT-3′ and 5′-AGGGAGTCCCT-3′, which differ by a single central base pair.

## Results

**Synthesis:** The polyamides **1**, **2**, and **3** were synthesized in 18 steps each from Boc-β-alanine–Pam resin (1 g resin/0.2 mmol $g^{-1}$ substitution) by using previously described solid-phase methods (Figure 3).[8] A sample of polyamide–resin (240 mg) was cleaved by a single step aminolysis reaction with ((dimethylamino)propyl)amine (55 °C, 18 h) and subsequently purified by reversed-phase HPLC chromatography. The trifluoroacetate salts of eight-ring 4-β-4 polyamides are stable at room temperature, and soluble in aqueous solution at concentrations ≤1 mM.

**Binding Site Size:** A plasmid was constructed, pSES123, which contains match sites according to the pairing rules for each of the three polyamides: 5′-atAGTAATTTACTgc-3′ (site 1), 5′-atAGGTATTACCTgc-3′ (site 2), and 5′-atAGGGATTCCCTgc-3′ (site 3) (Figure 4a). MPE·$Fe^{II}$ footprinting on the 3′- and 5′-$^{32}$P end-labeled 289 base pair restriction fragment from this plasmid (25 mM Tris-acetate, 10 mM NaCl, 100 μM calf thymus DNA, pH 7.0 and 22 °C) reveals that all three polyamides bind to their designated match sites (Figure 5). The polyamide ImPyPyPy-β-PyPyPyPy-β-Dp (**1**), which contains a single imidazole amino acid residue, binds to its match sequence 5′-atAGTAATTTACTgc-3′ (site 1) and weakly to the double mismatch sequence 5′-atAGGTATTACCTgc-3′ (site 2), but not to the quadruple mismatch sequence, 5′-atAGGGATTCCCTgc-3′ (site 3). Similarly, polyamide ImImPyPy-β-PyPyPyPy-β-Dp (**2**) reveals a strong footprint to its match sequence 5′-AGGTATTACCT-3′ (site 2) and only weak footprints on the double mismatch sequences, sites 1 and 3. Lastly, the polyamide with three contiguous Im residues ImImImPy-β-PyPyPyPy-β-Dp (**3**) recognizes its match sequence, 5′-AGGGATTCCCT-3′ (site 3), and its double mismatch sequence, site 2, but not site 1. The size

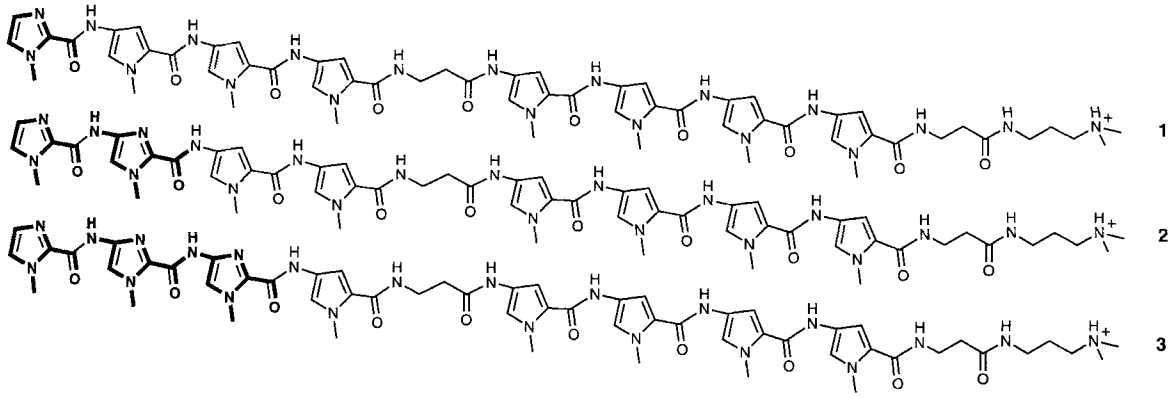


Figure 2. Structures of 4-β-4 polyamides **1**, **2**, and **3** [8].

  0947-6539/97/0310-1601 $ 17.50+.50/0 

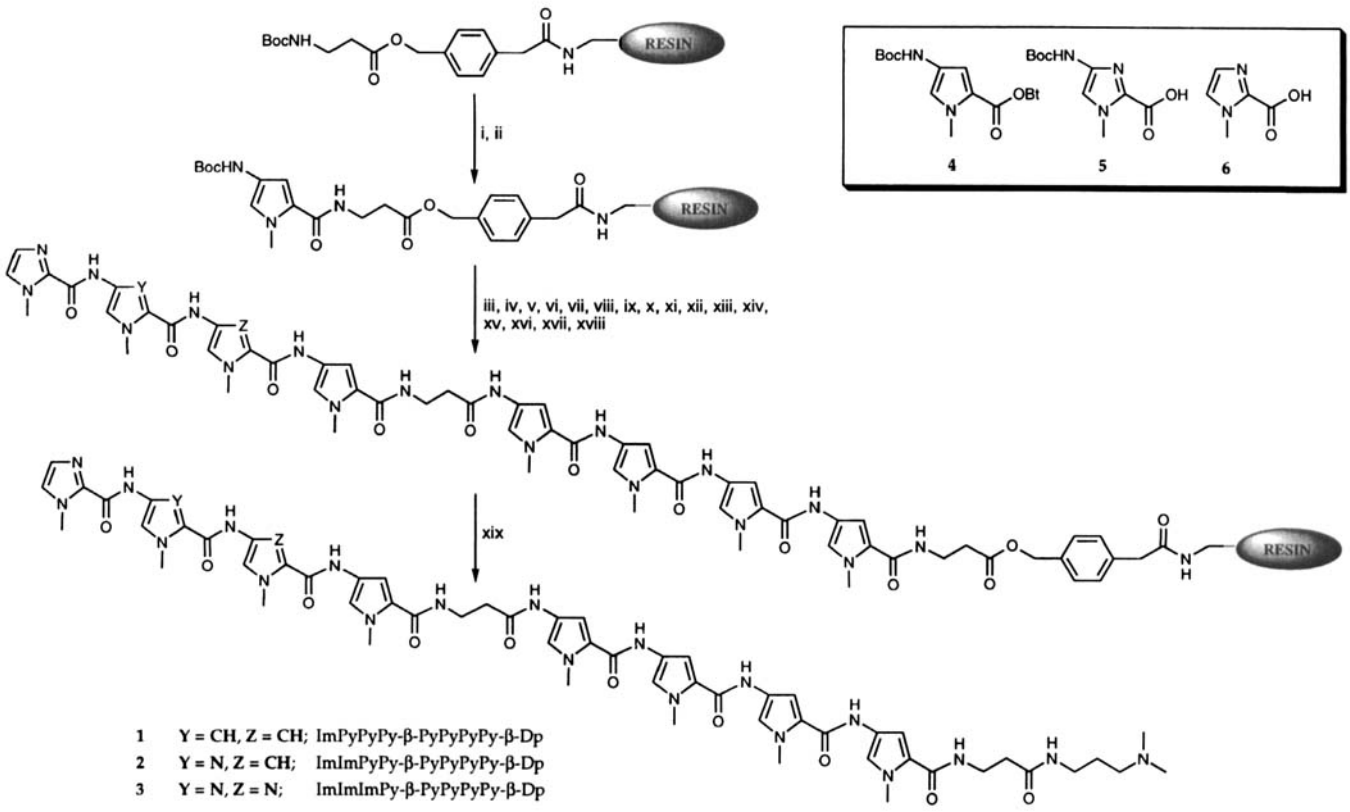


Figure 3. Box: Pyrrole and imidazole monomers for polyamide synthesis. Solid phase synthetic scheme for **1**, **2**, and **3** prepared from commercially available Boc-$\beta$-alanine-Pam resin (0.2 mmol g$^{-1}$): i) 80% TFA/DCM, 0.4 M PhSH; ii) BocPy-OBt, DIEA, DMF; iii) 80% TFA/DCM, 0.4 M PhSH; iv) BocPy-OBt, DIEA, DMF; v) 80% TFA/DCM, 0.4 M PhSH; vi) BocPy-OBt, DIEA, DMF; vii) 80% TFA/DCM, 0.4 M PhSH; viii) BocPy-OBt, DIEA, DMF; ix) 80% TFA/DCM, 0.4 M PhSH; x) Boc-$\beta$-alanine (HBTU, DIEA), DMF; xi) 80% TFA/DCM, 0.4 M PhSH; xii) BocPy-OBt, DIEA, DMF; xiii) 80% TFA/DCM, 0.4 M PhSH; xiv) BocPy-OBt, DIEA, DMF; (for **1** and **2**); BocIm-OBt (DCC/HOBt), DIEA, DMF (for **3**); xv) 80% TFA/DCM, 0.4 M PhSH; xvi) BocPy-OBt, DIEA, DMF (for **1**); BocIm-OBt (DCC/HOBt), DIEA, DMF (for **2** and **3**); xvii) 80% TFA/DCM, 0.4 M PhSH; xviii) imidazole-2-carboxylic acid (HBTU/DIEA); xix) *N,N*-((dimethylamino)propyl)amine, 55 °C.


**(a) pSES123**

5'-GATCGCGAGTGCTCAT**AGTAATTTACT**GCGAGTGCTCAT**AGGTATTACCT**GCGAGTGCTCAT**AGGGATTCCCT**GCGAGTGCTCAT -3'
3'- CGCTCACGAGTA**TCATTAAATGA**CGCTCACGAGTA**TCCATAATGGA**CGCTCACGAGTA**TCCCTAAGGGA**CGCTCACGAGTATCGA-5'

**(b) pSES4$\beta$4**

5'-GATCCGGTGTCATAA**AGGGAATCCCT**TACGCGGTGTAA**AGGTATTACCT**TACGCGGTGTAA**AGGGAGTCCCT**TACGCGGTAGTCTTAAGC -3'
3'- GCCACAGTATT**TCCCTTAGGGA**ATGCGCCACATT**TCCATAATGGA**ATGCGCCACATT**TCCCTCAGGGA**ATGCGCCATCAGAATTCGTCGA-5'

Figure 4. Sequence of the synthesized inserts from the plasmid a) pSES123, 289 bp; and b) pSES4$\beta$4, 294 bp. Top: illustration of the *Eco* RI/*Pvu* II restriction fragment with the *Bam* HI and *Hind* III insertion sites indicated. Only the sites that are shown in bold were analyzed by quantitative DNase I footprint titrations.

of the footprint cleavage protection for the polyamides is consistent with binding sites 11 bp in size.

**Apparent Equilibrium Association Constants:** Quantitative DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$, pH 7.0 and 22 °C)[10] were performed in order to determine the apparent equilibrium association constants ($K_a$ from Equation (2) in the Experimental Section) of the polyamides for the three designed sites (Figures 6 and 7). ImPyPyPy-$\beta$-PyPyPyPy-$\beta$-Dp (**1**) recognizes its match site, 5'-atAGTAATTTACTgc-3', with an apparent equilibrium association constant of $K_a = 2.4 \times 10^{10}$ M$^{-1}$. The double mismatch 5'-atAGGTATTACCTgc-3' is bound with 10-fold lower affinity. ImImPyPy-$\beta$-PyPyPyPy-$\beta$-Dp (**2**) binds its match site 5'-atAGGTAATACCTgc-3' with the highest affinity of any of the three polyamides studied ($K_a \geq 5.4 \times 10^{10}$ M$^{-1}$). The double mismatches 5'-atAGGGATTCCCTgc-3' and 5'-atAGTAATTTACTgc-3' are bound with at least 27-fold and 30-fold lower affinity, respectively. ImImImPy-$\beta$-PyPyPyPy-$\beta$-Dp (**3**) binds its match site 5'-atAGGGATTCCCTgc-3' with $K_a = 7.4 \times 10^{9}$ M$^{-1}$. The double base pair mismatch 5'-atAGGTAATACCTgc-3' is bound with 7-fold lower affinity.



0947-6539/97/0310-1602 $ 17.50+.50/0 *Chem. Eur. J.* **1997**, *3*, No. 10

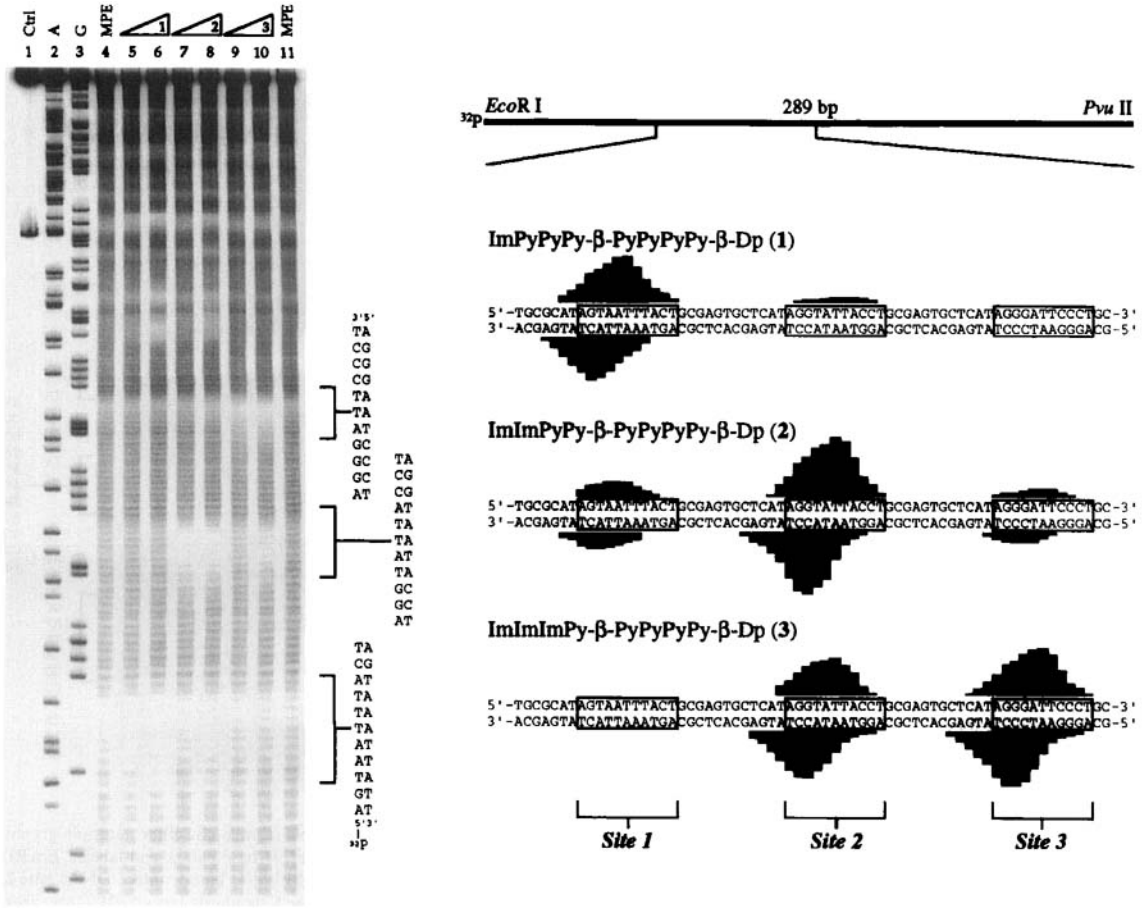


Figure 5. Left: MPE·Fe$^{II}$ footprinting experiments on the 5′-$^{32}$P-labeled 289 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pSES123. The 5′-AGTAATTTACT-3′, 5′-AGGTATTACCT-3, and 5′-AGGGATTCCCT-3′ sites are shown on the right side of the autoradiogram. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lanes 4 and 11, MPE·Fe$^{II}$ standard; lanes 5 and 6, 2 and 5 μM **1**; lanes 7 and 8, 2 and 5 μM **2**; lanes 9 and 10, 2 and 5 μM **3**. All lanes contain 15 kcpm 3′-radiolabeled DNA, 25 mM Tris-acetate buffer (pH 7.0), 10 mM NaCl, and 100 μM/base pair calf thymus DNA. Right: MPE·Fe$^{II}$ protection patterns for **1**, **2**, and **3** at 5 μM concentration. Top: Illustration of the 289 bp restriction fragment with the position of the sequence indicated. Bar heights are proportional to the relative protection from cleavage at each band. Boxes represent equilibrium binding sites determined by the published model, and only sites that were quantitated by DNase I footprint titrations are boxed. Shown with brackets are Site 1, 5′-AGTAATTTACT-3′; Site 2, 5′-AGGTATTACCT-3′; and Site 3, 5′-AGGGATTCCCT-3′.

**Sequence Specificity of a β–β Pairing:** A plasmid, pSES4β4, containing two 11-bp sites that differ by a single base pair in the sixth position, namely, 5′-aaAGGGAATCCCTta-3′ and 5′-aaAGGGAGTCCCTta-3, was constructed in order to explore the sequence specificty of a β/β pairing within the polyamide ImImImPy-β-PyPyPyPy-β-Dp (**3**). The relative binding affinity of **3** for the sites was determined on a 3′-$^{32}$P end-labeled 294 base pair restriction fragment derived from this plasmid (Figure 8). Quantitative DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$, pH 7.0 and 22 °C) reveal the apparent equilibrium association constants ($K_a$) of the polyamide for the two sites. Polyamide **3** binds the site with the A·T base pair in the sixth position with an equilibrium association constant of $K_a = 1.4 \times 10^{10}\,M^{-1}$. The sequence 5′-AGGGAGTCCCT-3′, which has a G·C base pair in the center position opposite the β–β pair, is bound with 20-fold lower affinity ($K_a = 6.9 \times 10^{8}\,M^{-1}$) (Figures 8–10).

## Discussion

MPE·Fe$^{II}$ footprinting reveals that the three 4-β-4 polyamides bind with highest affinity only to binding sites expected from the pairing rules (Figure 5). The asymmetric 3′-shifts in the footprint protection patterns are consistent with minor groove binding.[10] DNase I footprint titrations of polyamides **1**–**3** containing 1, 2, or 3 imidazole residues demonstrate the versatility of the dimeric 4-β-4 motif. Each polyamide binds its respective target site with apparent equilibrium association constants ranging from $K_a = 7 \times 10^{9}\,M^{-1}$ to $K_a \geq 5 \times 10^{10}\,M^{-1}$, and with DNA mismatch specificities ranging from 7-fold to >30-fold. Polyamides ImPyPyPy-β-PyPyPyPy-β-Dp (**1**) and ImImPyPy-β-PyPyPyPy-β-Dp (**2**) each recognize their respective target sites containing four or six G,C base pairs with $K_a \geq 2 \times 10^{10}\,M^{-1}$. ImImImPy-β-PyPyPyPy-β-Dp (**3**), which contains an ImImImPy sequence component, binds with 3-fold lower affinity compared to polyamide **1** (ImPyPyPy) and at least 7-fold lower than polyamide **2** (ImImPyPy). It could be that the imidazole ring located in the third position from the polyamides amino-terminus does not maintain precise ring–base pair register with the DNA and hence is not positioned optimally for hydrogen-bond formation.[11] Alternatively, Im-rich polyamides may not sit deeply in the minor groove of G,C-rich sites, diminishing energetically favorable van der Waals contacts with the wall of the DNA helix.[12]

  0947-6539/97/0310-1603 $ 17.50+.50/0 

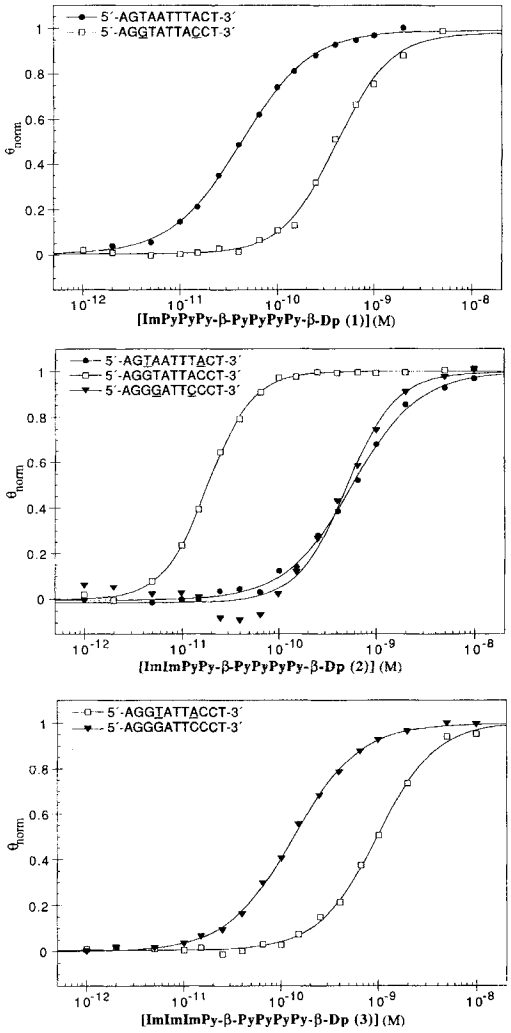


Figure 6. Data from the quantitative DNase I footprint titration experiments for polyamides **1**–**3** in complex with the designated sites. Top: polyamide **1**. Middle: polyamide **2**. Bottom: polyamide **3**. The $\theta_{norm}$ points were obtained by using photostimulable storage phosphor autoradiography and processed as described in the Experimental Section. The solid curves are the best-fit Langmuir binding titration isotherms obtained from nonlinear least-squares algorithm using Equation (2), with $n$ as an adjustable parameter.

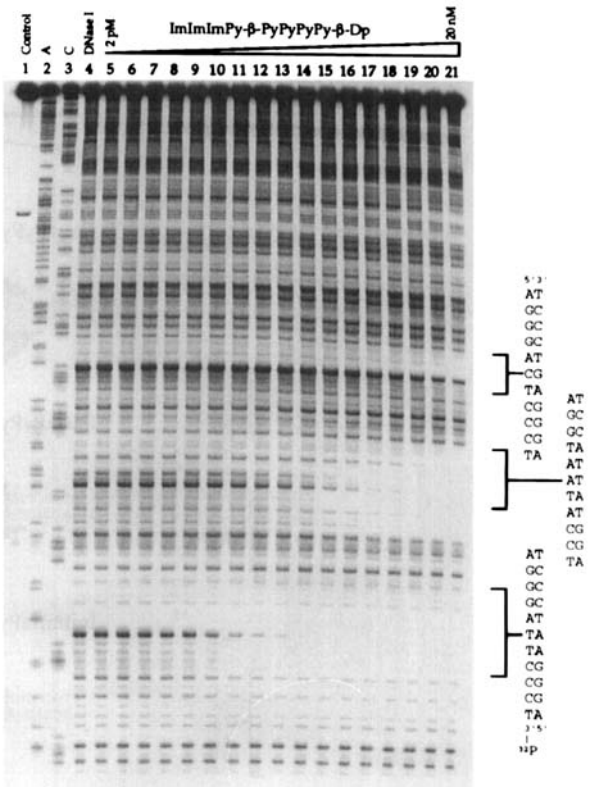


Figure 8. Image obtained by photostimulable storage phosphor autoradiography of a DNase I footprint titration of **3** on the 294 base pair 3′-end-labeled *Eco*R1/*Pvu*II restriction fragment from plasmid pSES4$\beta$4: lane 1, intact DNA; lane 2, A reaction; lane 3, C reaction; lane 4, DNase I standard; lanes 5–21, 2 pM, 5 pM, 10 pM, 20 pM, 40 pM, 65 pM, 100 pM, 150 pM, 250 pM, 400 pM, 650 pM, 1 nM, 1.5 nM, 2.5 nM, 5 nM, 10 nM, 20 nM polyamide. The 5′-AGGGAATCCCT-3′, 5′-AGGTATTACCT-3′, and 5′-AGGGAGTCCCT-3′ sites that were analyzed are shown on the right side of the autoradiogram. All reactions contain 15 kcpm restriction fragment, 10 mM Tris·HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$.

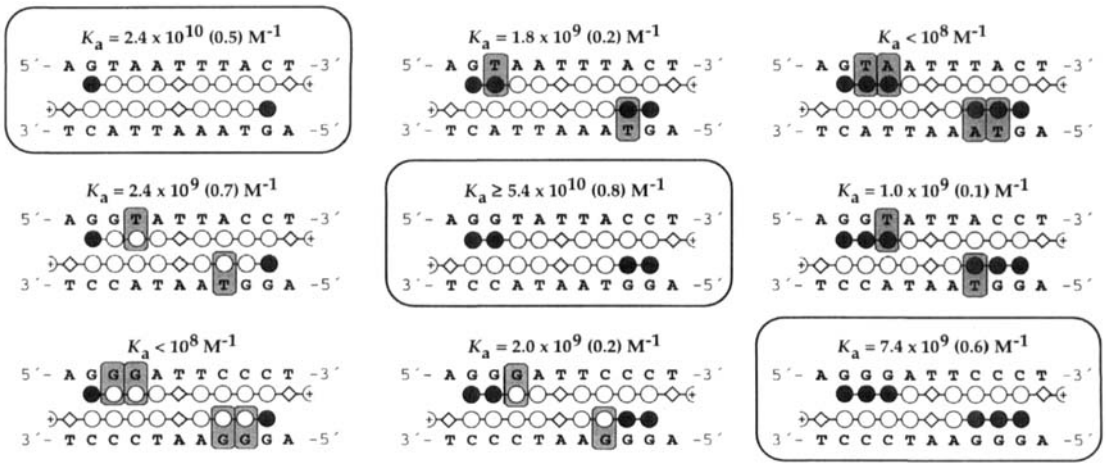


Figure 7. Ball-and-stick models of **1** (left), **2** (middle), and **3** (right) for each binding site, with the corresponding equilibrium association constants shown above each individual model. The binding sites shown are 5′-AGTAATTTACT-3′ (top), 5′-AGGTATTACCT-3′ (middle), and 5′-AGGGATTCCCT-3′ (bottom). Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent the $\beta$-alanine residue. Formally mismatched base pairs are boxed. Values reported are the mean values measured from four DNase I footprinting titration experiments, with the standard deviation for each data set indicated in parentheses. The assays were performed 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 $CaCl_2$.



0947-6539/97/0310-1604 $ 17.50+.50/0 *Chem. Eur. J.* **1997**, *3*, No. 10

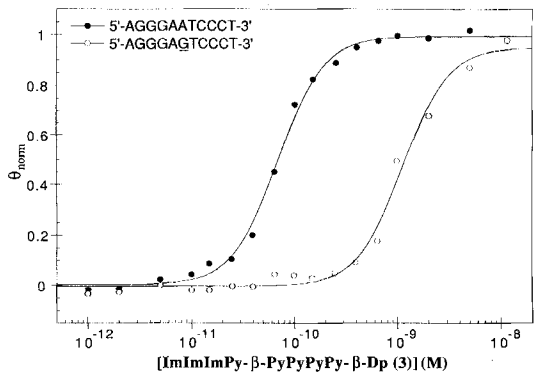


Figure 9. Data from the quantitative DNase I footprint titration experiments for **3** in complex with the designated sites (mismatched base underlined). The $\theta_{norm}$ points were obtained by means of photostimulable storage phosphor autoradiography and processed as described in the Experimental Section. The solid curves are the best-fit Langmuir binding titration isotherms obtained from nonlinear least-squares algorithm using Equation (2), $n = 2$.

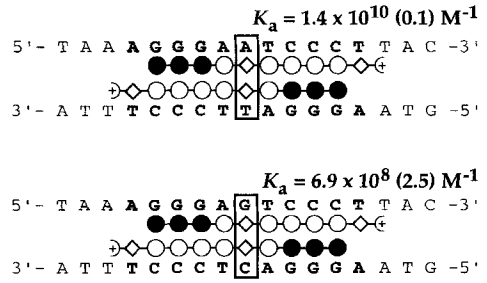


Figure 10. Ball-and-stick models comparing the binding of **3** to the match site 5′-AGGGAATCCCT-3′ (top) and to the $\beta$–$\beta$ single mismatch site 5′-AGGGAGTCCCT-3′ (bottom), with the corresponding equilibrium association constants shown above each individual model. Shaded and unshaded circles denote imidazole and pyrrole carboxamides, respectively, while diamonds represent the $\beta$-alanine residue. Values reported are the mean values measured from four DNase I footprinting titration experiments, with the standard deviation for each data set indicated in parentheses. The assays were performed 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 $CaCl_2$.

ImImImPy-$\beta$-PyPyPyPy-$\beta$-Dp (**3**) binds to 5′-AGGGATTCCCT-3′ with an apparent equilibrium association constant of $K_a = 1 \times 10^{10}\,M^{-1}$. For recognition of core 5′-GGG(A,T)-3′ sequences, four-ring ImImImPy subunits appear optimal.[11] C–N covalent head-to-tail linkage of polyamide subunits by a $\gamma$-aminobutyric acid ($\gamma$) linker to form "hairpin" polyamides increases affinity by 100-fold relative to the unlinked subunits.[13] As a control, the affinity of the eight-ring hairpin polyamide ImImImPy-$\gamma$-PyPyPyPy-$\beta$-Dp was measured on the 5′-AGGGAATCCCT-3′ site and found to be $K_a \approx 3 \times 10^8\,M^{-1}$.[14] Internal $\beta$- and $\gamma$-amino acids are specific "guide residues" for binding of polyamides in extended and turn conformations, respectively.[7, 15] Supporting this, the deletion of a single methylene unit (MW = 14) from the linker region of ImImImPy-$\gamma$-PyPyPyPy-$\beta$-Dp to ImImImPy-$\beta$-PyPyPyPy-$\beta$-Dp changes the binding motif from 1:1 (hairpin–DNA) to 2:1 (dimer–DNA) and thereby enlarges the targeted binding site size from 6 to eleven base pairs.

The polyamide ImImImPy-$\beta$-PyPyPyPy-$\beta$-Dp (**3**) binds the sequence 5′-AGGGAGTCCCT-3′ with 20-fold lower affinity than it does 5′-AGGGAATCCCT-3′, demonstrating that a $\beta$–$\beta$ pairing has a preference for binding opposite an A·T or T·A base pair relative to a G·C or C·G base pair (Figure 9). The exocyclic amino group of guanine presents a steric bulk in the minor groove of B-form DNA. The specificity of a Py–Py pair for A,T base pairs relative to G,C base pairs probably arises from a steric clash between a pyrrole C3–H and the guanine exocyclic 2-amino group. An antiparallel-oriented $\beta$–$\beta$ pair may present an even more sterically demanding surface by directing four hydrogens toward the floor of the groove, as compared to the two hydrogens presented by a Py–Py pair. Although generality has yet to be examined, these results suggest that *$\beta$-alanine is a sequence-specific DNA binding element in the overlapped 2:1 motif, rather than a sequence neutral spacer residue.*

## Conclusions

Pyrrole–imidazole polyamides bind the minor groove of DNA as dimers with high affinity and specificity. Specific base pair recognition is governed by a simple set of ring–ring pairing rules. Here we expand these pairing rules by demonstrating that side-by-side $\beta$-alanine residues are a sequence-specific DNA recognition element with preference for A·T/T·A relative to a G·C/C·G base pair. Fully overlapped 4-$\beta$-4 dimers bind eleven base pair sequences at subnanomolar concentrations, expanding the DNA-binding site size limit for synthetic polyamides. Unlike the structural complexity of protein–DNA complexes,[16] simple aromatic and aliphatic amino acids (Im, Py, and $\beta$), rationally placed within a $(4\text{-}\beta\text{-}4)_2$·DNA motif, recognize predetermined DNA sequences. The specific recognition of eleven base pairs by synthetic ligands is another pivotal step in the development of biologically useful synthetic ligands for regulation of gene-expression in living cells.

## Experimental Section

**General:** NMR were recorded on a GE 300 instrument operating at 300 MHz ($^1$H). Spectra were recorded in [$D_6$]DMSO with chemical shifts reported in parts per million relative to residual [$D_5$]DMSO. UV spectra were measured on a Hewlett-Packard Model 8452 A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time-of-flight mass spectrometry was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed either on a HP 1090 M analytical HPLC or a Beckman Gold system using a Rainen C 18, Microsorb MV, 5 μm, 300 × 4.6 mm reversed-phase column in 0.1% (wt/v) TFA with acetonitrile as eluent and a flow rate of 1.0 mL min$^{-1}$, gradient elution 1.25% acetonitrile per min. Preparative HPLC was carried out on a Beckman instrument using a Waters DeltaPak 25 × 100 mm 100 μm $C_{18}$ column in 0.1% (wt/v) TFA, gradient elution 0.25% per min $CH_3CN$. Water was obtained from a Millipore Milli-Q water purification system. Reagent-grade chemicals were used unless otherwise stated.

**ImPyPyPy-$\beta$-PyPyPyPy-$\beta$-Dp (1)**: A sample of ImPyPyPy-$\beta$-PyPyPyPy-$\beta$-resin prepared by machine-assisted solid-phase synthesis (240 mg, 0.16 mmol g$^{-1}$[17]) was placed in a 20 mL glass scintillation vial, and treated with ((dimethylamino)propyl)amine (2 mL) at 55 °C for 18 h. Resin was removed by filtration, and the filtrate diluted to a total volume of 8 mL with 0.1% (wt/v) aqueous TFA. The resulting crude polyamide/amine solution was purified directly by reversed-phase HPLC to provide the trifluoroacetate salt of ImPyPyPy-$\beta$-PyPyPyPy-$\beta$-Dp (31 mg, 40% recovery) as a white powder. $^1$H NMR (300 MHz, [$D_6$]DMSO, 20 °C): $\delta$ = 10.49 (s, 1H; NH), 9.97 (s, 1H; NH), 9.95 (s, 1H; NH), 9.94 (s, 1H; NH), 9.93 (m, 2H; NH), 9.91 (s, 1H; NH), 9.4 (br s, 1H; $CF_3COOH$), 8.10 (m, 3H; NH), 7.38 (s, 1H; CH),

  0947-6539/97/0310-1605 $ 17.50+.50/0 

7.28 (d, $^2J$(H,H) = 1.6 Hz, 1H; CH), 7.22 (m, 3H; CH), 7.19 (m, 2H; CH), 7.16 (m, 2H; CH), 7.09 (m, 2H; CH), 7.04 (m, 2H; CH), 6.87 (d, $^2J$(H,H) = 1.6 Hz, 1H; CH), 6.86 (d, $^2J$(H,H) = 1.6 Hz, 1H; CH), 6.84 (d, $^2J$(H,H) = 1.5 Hz, 1H; CH), 3.97 (s, 3H; $NCH_3$), 3.82 (m, 15H; $NCH_3$), 3.80 (s, 3H; $NCH_3$), 3.78 (s, 3H; $NCH_3$), 3.4 (m, 6H; $CH_2$), 3.10 (q, $^4J$(H,H) = 5.4 Hz, 2H; $CH_2$), 2.98 (q, $^4J$(H,H) = 5.3 Hz, 2H; $CH_2$), 2.72 (d, $^2J$(H,H) = 4.7 Hz, 6H; $N(CH_3)_2$), 2.33 (t, $^3J$(H,H) = 7.0 Hz, 2H; $CH_2$), 1.71 (q, $^5J$(H,H) = 6.4 Hz, 2H; $CH_2$); UV/Vis ($H_2O$) $\lambda_{max}$ ($\varepsilon$) = 312 (66600, calculated from $\varepsilon$ = 8333 per ring[13d]), 244 nm; MALDI-TOF-MS [$M^-$ − H] 1208.2: calcd 1208.3.

**ImImPyPy-β-PyPyPyPy-β-Dp (2)**: A sample of ImImPyPy-β-PyPyPyPy-β-resin prepared by machine-assisted solid-phase synthesis (240 mg, 0.16 mmol g$^{-1}$[17]) was placed in a 20 mL glass scintillation vial, and treated with ((dimethylamino)propyl)amine (2 mL) at 55 °C for 18 h. Resin was removed by filtration, and the filtrate diluted to a total volume of 8 mL with 0.1% (wt/v) aqueous TFA. The resulting crude polyamide/amine solution was purified directly by reversed-phase HPLC to provide the trifluoroacetate salt of ImImPyPy-β-PyPyPyPy-β-Dp (31 mg, 40% recovery) as a white powder. $^1$H NMR (300 MHz, [D$_6$]DMSO, 20 °C): $\delta$ = 10.38 (s, 1H; NH), 9.95 (s, 1H; NH), 9.93 (s, 1H; NH), 9.91 (s, 1H; NH), 9.90 (m, 2H; NH), 9.76 (s, 1H; NH), 9.4 (br s, 1H; $CF_3COOH$), 8.09 (m, 3H; NH), 7.56 (s, 1H; CH), 7.46 (s, 1H; CH), 7.27 (d, $^2J$(H,H) = 1.8 Hz, 1H; CH), 7.21 (d, $^2J$(H,H) = 1.7 Hz, 1H; CH), 7.20 (d, $^2J$(H,H) = 1.9 Hz, 1H; CH), 7.19 (d, $^2J$(H,H) = 1.9 Hz, 1H; CH), 7.16 (d, $^2J$(H,H) = 1.9 Hz, 1H; CH), 7.15 (d, $^2J$(H,H) = 1.6 Hz, 1H; CH), 7.14 (d, $^2J$(H,H) = 1.9 Hz, 1H; CH), 7.12 (d, $^2J$(H,H) = 1.6 Hz, 1H; CH), 7.07 (s, 1H; CH), 7.05 (d, $^2J$(H,H) = 1.5 Hz, 1H; CH), 6.87 (d, $^2J$(H,H) = 1.9 Hz, 1H; CH), 6.86 (d, $^2J$(H,H) = 1.6 Hz, 1H; CH), 6.84 (d, $^2J$(H,H) = 1.6 Hz, 1H; CH), 3.99 (m, 6H; $NCH_3$), 3.82 (m, 12H; $NCH_3$), 3.80 (s, 3H; $NCH_3$), 3.78 (s, 3H; $NCH_3$), 3.4 (m, 6H; $CH_2$), 3.09 (q, $^4J$(H,H) = 5.6 Hz, 2H; $CH_2$), 2.97 (q, $^4J$(H,H) = 5.2 Hz, 2H; $CH_2$), 2.71 (d, $^2J$(H,H) = 4.2 Hz, 6H; $N(CH_3)_2$), 2.32 (t, $^3J$(H,H) = 5.1 Hz, 2H; $CH_2$), 1.71 (q, $^5J$(H,H) = 7.4 Hz, 2H; $CH_2$); UV/Vis ($H_2O$) $\lambda_{max}$ ($\varepsilon$) = 306 (66600, calculated from $\varepsilon$ = 8333 per ring[13d]), 243 nm; MALDI-TOF-MS [$M^+$ − H] 1209.1: calcd 1209.3

**ImImImPy-β-PyPyPyPy-β-Dp (3)**: A sample of ImImImPy-β-PyPyPyPy-β-resin prepared by machine-assisted solid-phase synthesis (240 mg, 0.16 mmol g$^{-1}$[17]) was placed in a 20 mL glass scintillation vial, and treated with ((dimethylamino)propyl)amine (2 mL) at 55 °C for 18 h. Resin was removed by filtration, and the filtrate diluted to a total volume of 8 mL with 0.1% (wt/v) aqueous TFA. The resulting crude polyamide/amine solution was purified directly by reversed-phase HPLC to provide the trifluoroacetate salt of ImImImPy-β-PyPyPyPy-β-Dp (24 mg, 31% recovery) as a white powder. $^1$H NMR (300 MHz, [D$_6$]DMSO, 20 °C): $\delta$ = 10.37 (s, 1H; NH), 10.12 (s, 1H; NH), 9.95 (s, 1H; NH), 9.94 (s, 1H; NH), 9.93 (s, 1H; NH), 9.92 (s, 1H; NH), 9.59 (s, 1H; NH), 9.4 (br s, 1H; $CF_3COOH$), 8.09 (m, 3H; NH), 7.65 (s, 1H; CH), 7.56 (s, 1H; CH), 7.45 (s, 1H; CH), 7.27 (d, $^2J$(H,H) = 1.3 Hz, 1H; CH), 7.22 (m, 2H; CH), 7.18 (d, $^2J$(H,H) = 1.2 Hz, 1H; CH), 7.16 (d, $^2J$(H,H) = 1.0 Hz, 1H; CH), 7.07 (m, 2H; CH), 6.95 (d, $^2J$(H,H) = 1.1 Hz, 1H; CH), 6.88 (d, $^2J$(H,H) = 1.4 Hz, 1H; CH), 6.86 (d, $^2J$(H,H) = 1.3 Hz, 1H; CH), 4.01 (s, 3H; $NCH_3$), 3.98 (m, 2H; $NCH_3$), 3.83 (s, 3H; $NCH_3$), 3.82 (m, 6H; $NCH_3$), 3.80 (s, 3H; $NCH_3$), 3.78 (s, 3H; $NCH_3$), 3.4 (m, 6H; $CH_2$), 3.11 (q, $^4J$(H,H) = 5.2 Hz, 2H; $CH_2$), 2.94 (q, $^4J$(H,H) = 5.3 Hz, 2H; $CH_2$), 2.69 (d, $^2J$(H,H) = 4.4 Hz, 6H; $N(CH_3)_2$), 2.33 (t, $^3J$(H,H) = 5.4 Hz, 2H; $CH_2$), 1.75 (q, $^5J$(H,H) = 7.1 Hz, 2H; $CH_2$); UV/Vis ($H_2O$) $\lambda_{max}$ ($\varepsilon$) = 304 (66600, calculated from $\varepsilon$ = 8333 per ring[13d]), 241 nm; MALDI-TOF-MS [$M^+$ − H] 1210.4: calcd 1210.3.

**DNA Reagents and Materials:** Enzymes were purchased from Boehringer-Mannheim or New England Biolabs and were used with their supplied buffers. Deoxyadenosine and thymidine 5′-[α-$^{32}$P]triphosphates were obtained from Amersham. RNase-free water was obtained from USB. Sonicated, deproteinized calf thymus DNA was acquired from Pharmacia. All other reagents and materials were used as received. All DNA manipulations were performed according to standard protocols.[18]

**Construction of Plasmid DNA:** The plasmids pSES4β4 and pSES123 were constructed by hybridization of the inserts listed in Figures 4a and 4b. Each hybridized insert was ligated individually into linearized pUC19 *Bam*HI/*Hind*III plasmid by using T4 DNA ligase. The resultant constructs were used to transform Top10F′ OneShot competent cells from Invitrogen. Ampicillin-resistant white colonies were selected from 25 mL Luria-Bertani medium agar plates containing 50 μg mL$^{-1}$ ampicillin and treated with XGAL and IPTG solutions. Large-scale plasmid purification was performed with Qiagen Maxi purification kits. Dideoxy sequencing was used to verify the presence of the desired insert. Concentration of the prepared plasmid was determined at 260 nm using the relationship of 1 OD unit = 50 μg mL$^{-1}$ duplex DNA.

**Preparation of 3′- and 5′-End-Labeled Restriction Fragments:** Either the plasmid pSES123 or pSES4β4 was linearized with *Eco*RI and then treated with either Klenow fragment, deoxyadenosine 5′-[α-$^{32}$P]triphosphate and thymidine 5′-[α-$^{32}$P]triphosphate for 3′ labeling, or with calf alkaline phosphatase and then 5′ labeled with T4 polynucleotide kinase and deoxyadenosine 5′-[γ-$^{32}$P]triphosphate (for pSES123 only). The labeled fragment (3′ or 5′) was then digested with *Pvu*II and loaded onto a 5% non-denaturing polyacrylamide gel. The desired band was visualized by autoradiography and isolated. Chemical sequencing reactions were performed according to published methods.[19]

**MPE·Fe$^{II}$ Footprinting**:[9] All reactions were carried out in a volume of 40 μL. A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were: 25 mM Tris-acetate buffer (pH 7.0), 10 mM NaCl, 100 μM/base pair calf thymus DNA, and 30 kcpm 3′- or 5′-radiolabeled DNA. The solutions were allowed to equilibrate for 4 h. A fresh 50 μM MPE·Fe$^{II}$ solution was made from 100 μL of a 100 μM MPE solution and 100 μL of a 100 μM ferrous ammonium sulfate ($Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$) solution. After the 24 h equilibration, MPE·Fe$^{II}$ solution (5 μM) was added, and the reactions were equilibrated for 5 min. Cleavage was initiated by the addition of dithiothreitol (5 mM) and allowed to proceed for 14 min. Reactions were stopped by ethanol precipitation, resuspended in 100 mM Tris-borate-EDTA/80% formamide loading buffer, denatured at 85 °C for 5 min, placed on ice, and half of each reaction (≈15 kcpm) immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V.

**DNase I Footprinting**:[10] All reactions were carried out in a volume of 400 μL. We note explicitly that no carrier DNA was used in these reactions prior to the DNase I cleavage step. A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were: 10 mM Tris·HCl buffer (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, and 20 kcpm 3′-radiolabeled DNA. The solutions were allowed to equilibrate (36 h for studies on pSES4β4, and a minimum of 48 h for all studies on pSES123) at 22 °C. Cleavage was initiated by the addition of 10 μL of a DNase I stock solution (diluted with 1 mM DTT to give a stock concentration of 0.28 u mL$^{-1}$) and was allowed to proceed for 5 min at 22 °C. The reactions were stopped by the addition of 50 μL of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg mL$^{-1}$ glycogen, and 30 μM base-pair calf thymus DNA, and then ethanol precipitated. The cleavage products were resuspended in 100 mM Tris-borate-EDTA/80% formamide loading buffer, denatured at 85 °C for 5 min, placed on ice, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V. The gels were dried under vacuum at 80 °C, then quantitated using storage phosphor technology.

Apparent equilibrium association constants were determined as previously described.[12b] The data were analyzed by performing volume integrations of the 5′-atAGTAATTTACTgc-3′, 5′-atAGGTATTACCTgc-3′, and 5′-atAGGGATTCCCTgc-3 sites (for pSES123) or the 5′-aaAGGGAATCCCTta-3′, 5′-aaAGGGAGTCCCTta-3′, and 5′-aaAGGTATTACCTta-3′ sites (for pSES4β4) and a reference site. The apparent DNA target site saturation, $\theta_{app}$, was calculated for each concentration of polyamide by means of Equation (1), where $I_{tot}$ and $I_{ref}$ are the integrated volumes of the target and

$$\theta_{app} = 1 - \frac{I_{tot}/I_{ref}}{I^{\circ}_{tot}/I^{\circ}_{ref}} \quad (1)$$

reference sites, respectively, and $I^{\circ}_{tot}$ and $I^{\circ}_{ref}$ correspond to those values for a DNase I control lane to which no polyamide has been added. The ($[L]_{tot}$, $\theta_{app}$) data points were fitted to a Langmuir binding isotherm by minimizing the difference between $\theta_{app}$ and $\theta_{fit}$, by using the modified Hill Equation (2),

$$\theta_{fit} = \theta_{min} + (\theta_{max} - \theta_{min})\frac{K_a^n[L]_{tot}^n}{1 + K_a^n[L]_{tot}^n} \quad (2)$$



0947-6539/97/0310-1606 $ 17.50+.50/0 *Chem. Eur. J.* **1997**, *3*, No. 10

where $[L]_{tot}$ corresponds to the total polyamide concentration, $K_a$ corresponds to the apparent monomeric association constant,[20] and $\theta_{min}$ and $\theta_{max}$ represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. For pSES4$\beta$4, Equation (2) was solved with $n = 2$, while for pSES123, $n$ was treated as another variable.

Data were fitted by using a nonlinear least-squares fitting procedure of KaleidaGraph software (version 2.1, Abelbeck software) with $K_a$, $\theta_{max}$, and $\theta_{min}$ as the adjustable parameters (for pSES123 data, $n$ was also an adjustable parameter). All acceptable fits had a correlation coefficient of $R > 0.97$. For pSES4$\beta$4, each target site was adequately fitted by a fully cooperative Langmuir isotherm [Eq. (2), $n = 2$], consistent with the hypothesis that ImIm-ImPy-$\beta$-PyPyPyPy-$\beta$-Dp (**3**) binds in a 2:1 manner, though we emphasize that this analysis does not attempt to model an actual binding mechanism. For data obtained at 48 h equilibration on pSES123, some sites were fitted with lower values of $n$ (though $n$ never fell below 1.2). The use of the modified Hill equation [Eq. (2)] makes the assumption that $[L]_{tot} \approx [L]_{free}$, where $[L]_{free}$ represents the amount of unbound polyamide in solution. For extremely high-affinity ligands, this assumption no longer holds and the association constants are underestimated. Here, the DNA concentration is estimated at approximately 2 pM, and therefore equilibrium association constants of $> 3 \times 10^{10}\,M^{-1}$ should be considered lower limits. The data were normalized by means of Equation (3). Four sets of acceptable data were used in determining each association constant. All lanes from each gel were used unless visual inspection revealed a data point to be obviously flawed relative to neighboring points.

$$\theta_{norm} = \frac{\theta_{app} - \theta_{min}}{\theta_{max} - \theta_{min}} \qquad (3)$$

**Quantitation by Storage Phosphor Technology Autoradiography:** 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 16–20 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 the ImageQuant v. 3.2.

**Acknowledgments:** We are grateful to the National Institutes of Health (GM-27681) for research support, the National Institutes of Health for a research traineeship award (GM-08501) to S. E. S., and the Howard Hughes Medical Institute for a predoctoral fellowship to E. E. B.

Received: April 28, 1997 [F681]

[1] J. W. Trauger, E. E. Baird, P. B. Dervan, *Nature* **1996**, *382*, 559.
[2] a) W. S. Wade, M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc.* **1992**, *114*, 8783; b) M. Mrksich, W. S. Wade, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. B. Dervan, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586; c) W. S. Wade, M. Mrksich, P. B. Dervan, *Biochemistry* **1993**, *32*, 11385.
[3] a) J. G. Pelton, D. E. Wemmer, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723; b) J. G. Pelton, D. E. Wemmer, *J. Am. Chem. Soc.* **1990**, *112*, 1393; c) X. Chen, B. Ramakrishnan, S. T. Rao, M. Sundaralingham, *Nature Struct. Biol.* **1994**, *1*, 169; d) S. White, E. E. Baird, P. B. Dervan, *Biochemistry* **1996**, *35*, 12532.
[4] J. M. Gottesfield, L. Nealy, J. W. Trauger, E. E. Baird, P. B. Dervan, *Nature* **1997**, *387*, 202.
[5] P. B. Dervan, *Science* **1986**, *232*, 464.
[6] J. J. Kelly, E. E. Baird, P. B. Dervan, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6981.
[7] a) J. W. Trauger, E. E. Baird, M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 6160.
[8] E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 6141.
[9] a) M. W. Van Dyke, P. B. Dervan, *Biochemistry* **1983**, *22*, 2373; b) M. W. Van Dyke, P. B. Dervan, *Nucleic Acids Res.* **1983**, *11*, 5555.
[10] a) M. Brenowitz, D. F. Senear, M. A. Shea, G. K. Ackers, *Methods Enzymol.* **1986**, *130*, 132; b) M. Brenowitz, D. F. Senear, M. A. Shea, G. K. Ackers, *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8462; c) D. F. Senear, M. Brenowitz, M. A. Shea, G. K. Ackers, *Biochemistry* **1986**, *25*, 7344.
[11] S. E. Swalley, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 8198.
[12] a) B. H Geierstanger, M. Mrksich, P. B. Dervan, D. E. Wemmer, *Science*, **1994**, *266*, 646; b) M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc.* **1995**, *117*, 3325.
[13] a) M. Mrksich, M. E. Parks, P. B. Dervan, *J. Am. Chem. Soc.* **1994**, *116*, 7983; b) M. E. Parks, E. E. Baird, P. B. Dervan, *ibid.* **1996**, *118*, 6147; c) *ibid.* **1996**, *118*, 6153; d) D. S. Pilch, N. A. Poklar, C. A. Gelfaned, S. M. Law, K. J. Breslauer, E. E. Baird, P. B. Dervan, *Proc. Natl. Acad. Sci. USA*, **1996**, *93*, 8306. e) R. P. L. de Claire, B. H. Geierstanger, M. Mrksich, P. B. Dervan, D. E. Wemmer, *J. Am. Chem. Soc.* **1997**, *34*, 7909..
[14] The eleven base pair site contains two overlapping six base pair 5′-AGGGA(A,T)-3′ sites. The affinity of the 4-$\gamma$-4 hairpin was measured for the entire eleven base pairs, since the two individual six base pair sites cannot be considered separately. However, the determined value of $K_a \approx 3 \times 10^8\,M^{-1}$ is consistent with the previously published finding of $K_a = 4 \times 10^8\,M^{-1}$ for the single site 5′-AGGGAA-3′ (see ref. [11]).
[15] J. W. Trauger, E. E. Baird, P. B. Dervan, *Chem. Biol.* **1996**, *3*, 369.
[16] T.A Steitz. *Quart. Rev. Biophys* **1990**, *23*, 205.
[17] Resin substitution has been corrected acording to $L_{new}$ (mmol g$^{-1}$) = $L_{old}$/$(1 + L_{old}(W_{new} - W_{old}) \times 10^{-3})$, where $L$ is the loading (mmol of amine per gram of resin), and $W$ is the weight (g mol$^{-1}$) of the polyamide. See: K. Barlos, O. Chatzi, D. Gatos, G. Stravropoulos, *Int. J. Peptide Protein Res.* **1991**, *37*, 513.
[18] J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning*, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, **1989**.
[19] a) A. M. Maxam, W. S. Gilbert, *Methods Enzymol.* **1980**, *65*, 499; b) B. L. Iverson, P . B. Dervan, *Nucleic Acids Res.* **1987**, *15*, 7823.
[20] For the treatment of data on cooperative association of ligands see: C. R. Cantor, P. R. Schimmel, *Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules*, W. H. Freeman, New York, NY, **1980**, p. 863.

  0947-6539/97/0310-1607 $ 17.50+.50/0 