VOLUME 121, NUMBER 6
FEBRUARY 17, 1999

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Effects of γ-Turn and β-Tail Amino Acids on Sequence-Specific Recognition of DNA by Hairpin Polyamides

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 August 27, 1998

Abstract: Three-ring polyamides containing pyrrole (Py) and imidazole (Im) amino acids covalently coupled by a turn-specific γ-aminobutyric acid linker (γ-turn) form six-ring hairpins that recognize predetermined 5-base pair (bp) sequences in the minor groove of DNA. To determine the sequence specificity of the γ-turn and C-terminal β-alanine (β-tail) amino acids, the DNA-binding properties of the hairpin polyamide ImImPy-γ-ImPyPy-β-Dp were analyzed by footprinting and affinity cleavage on DNA-restriction fragments containing the eight possible 5′-A*TGGCN*A-3′ and 5′-A*NGGCT*A-3′ sites (***N*** = A, T, G or C; 5-bp hairpin site is in italics). Quantitative footprint titrations demonstrate that both the γ-turn and β-tail amino acids have a >200−400-fold preference for A·T/T·A relative to G·C base pairs at these positions. Effects of the base pairs adjacent to the 5-bp hairpin-binding site were analyzed by footprinting experiments on a DNA-restriction fragment containing the eight possible 5′-A*TGGCT***N**-3′ and 5′-**N***TGGCT*A-3′ sites. Quantitative footprint titrations demonstrate that the turn and tail amino acids have reduced specificity (3−20-fold preference) for A·T/T·A relative to G·C base pairs at these positions. These results indicate that the turn and tail amino acids do not simply act as neutral linker residues but, in fact, are sequence-specific recognition elements with predictable DNA-binding specificity.

Introduction

Small molecules that target specific DNA sequences have the potential to control gene expression.[1] Polyamides containing the three aromatic amino acids 3-hydroxypyrrole (Hp), imidazole (Im) and pyrrole (Py) are synthetic ligands that have an affinity and specificity for DNA comparable to naturally occurring DNA-binding proteins.[2,3] DNA recognition depends on side-by-side amino acid pairings in the minor groove.[2−10]An anti-

(1) (a) Gottesfield, J. M.; Nealy, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202−205. (b) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12890.

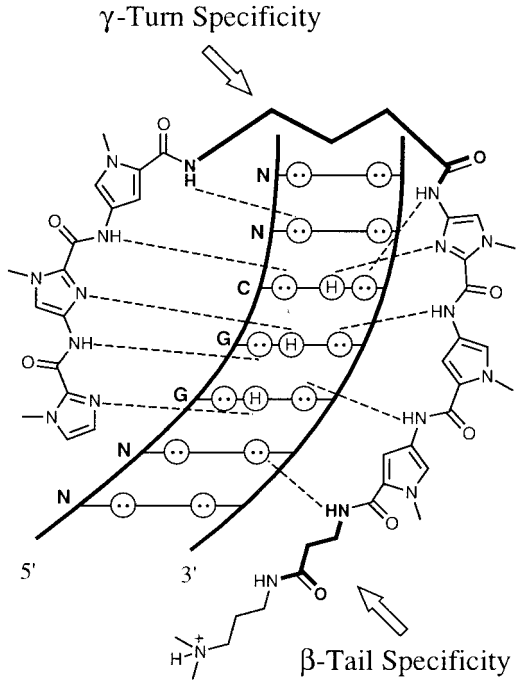
(2) (a) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559−561. (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953−6961. (c) Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 7636−7644. (d) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 1421−1423. (e) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219−6226.

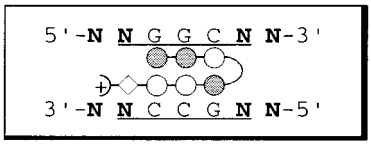
(3) White, S. E.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468−471.

(4) (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.; Geierstanger, 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) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572−2576. (e) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646−650. (f) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nature Struct. Biol.* **1998**, *5*, 104−109.

(5) (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.; Sundaralingham, M. *Nature Struct. Biol.* **1994**, *1*, 169−175. (d) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532−12537.

10.1021/ja9830905 CCC: $18.00
Published on Web 01/27/1999

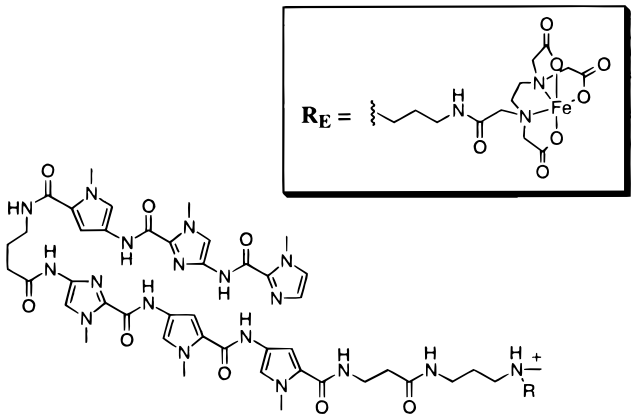




ImImPy-γ-ImPyPy-β-Dp•5′-N**N**GGCC**N**N-3′

Figure 1. Binding models for the complex formed between ImImPy-γ-ImPyPy-β-Dp (**1**) and 5′-**NN***GGCC***NN**-3′, where the tail or turn position is varied. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. N is represented as A or T but can be A, T, G, or C. Putative hydrogen bonds are illustrated by dotted lines. Ball and stick models are also shown. Shaded and nonshaded circles denote Im and Py, respectively. Nonshaded diamonds represent the β-residue.

parallel pairing of Im opposite Py (Im/Py pair) distinguishes G•C from C•G and both of these from A•T/T•A base pairs (bp).[4]



(1) R = CH_3, ImImPy-γ-ImPyPyy-β-Dp
(1-E•Fe(II)) R = R_E, ImImPy-γ-ImPyPy-β-Dp-EDTA•Fe(II)

Figure 2. Structures of the six-ring hairpin polyamides **1** and the corresponding Fe(II)•EDTA affinity cleaving derivative.

A Py/Py pair binds both A•T and T•A in preference to G•C/C•G.[4,5] The discrimination of T•A from A•T using Hp/Py pairs completes the 4-bp code.[3,7] In contrast to dimers that can adopt a variety of "slipped" binding modes in the minor groove,[10] the "hairpin motif" locks the relative positions of the individual subunits.[9] The linker amino acid γ-aminobutyric acid (γ) connects polyamide subunits C → N, and these ligands bind to predetermined target sites with >100-fold enhanced affinity.[2,9] Addition of a C-terminal β-alanine (β) to hairpin polyamides increases both the affinity and specificity and facilitates solid-phase synthesis.[9b]

For a polyamide containing a number of consecutive ring pairings, *n*, the binding site size will be $n + 2$ base pairs.[4] Each side-by-side ring pairing recognizes a central base pair, and the outer base pairs are formally recognized by the terminal amides.[4] For example, according to the "pairing rules" the six-ring-hairpin polyamide ImImPy-γ-ImPyPy-β-Dp (Dp = dimethylaminopropylamine) will bind to a 5-bp 5′-**N**GGC**N**-3′ sequence (**N** = A, T G, C), where the GGC "core" is recognized by three corresponding Im/Py, Im/Py, and Py/Im ring pairings (Figure 1). From our prior observations, the γ-turn and β-tail amino acids appear to specify for A•T and T•A base pairs;[9] however, this has not yet been examined in detail, particularly with regard to a well-controlled study quantitating the flanking sequence effects of all 4 base pairs. Understanding the sequence specificity of the γ-turn and β-tail amino acids of hairpin polyamides for the base pairs surrounding the "core" sequence could potentially enhance the predictability of sequence targeting with regard to transcription inhibition experiments.

To investigate the sequence-specific effects of the γ-turn and β-tail amino acids, a hairpin polyamide, ImImPy-γ-ImPyPy-β-Dp (**1**) and the corresponding affinity cleaving analogue (**1-E**) were synthesized (Figures 1 and 2).[11] The single binding orientation expected for this six-ring hairpin polyamide at a 5-bp site (italics) allows the effects of the γ- and β-amino acids to be determined at the first position flanking the 3-bp core, 5′-A*TGGCN*A-3′ and 5′-A*NGGCT*A-3′ (***N*** = A, T, G, and C; see Figure 1). It was unknown whether base pairs immediately adjacent to the 5-bp recognition site would also be discriminated by the linker amino acids. Thus, for completeness, the binding of polyamide **1** was also examined on DNA-restriction fragments containing the eight possible sequences, 5′-A*TGGCT***N**-3′ and 5′-**N***TGGCT*A-3′. A hairpin polyamide ImImPy-$(R)^{H2N}\gamma$-ImPyPy-PrOH (**3**) (where $(R)^{H2N}\gamma$ = (*R*)-2,4-diaminobutyric

(6) (a) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474−4482. (b) Singh, S. B.; Ajay; Wemmer, D. E.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7673−7677. (c) White, S.; Baird, E. E.; Dervan, P. B. *Chem., Biol.* **1997**, *4*, 569−578.

(7) Kielkopf, C. L.; White, S. E.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111.

(8) (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900. (c) Chen, Y. H.; Lown, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 6995. (d) Greenberg, W. A., Baird, E. E., Dervan, P. B. *Chem. Eur. J.* **1998**, *4*, 796−805.

(9) (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983−7988. (b) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147−6152. (c) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6153−6159. (d) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198−8206. (e) Pilch, D. S.; Poklar, 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*, 8306−8311. (f) de Clairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909−7916. (g) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756−8765.

(10) (a) Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981−6985. (b) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160−6166. (c) Geierstanger, B. H.; Mrksich M.; Dervan, P. B.; Wemmer, D. E. *Nature Struct. Biol.* **1996**, *3*, 321−324. (d) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1997**, *3*, 1600−1607. (e) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534−3535.

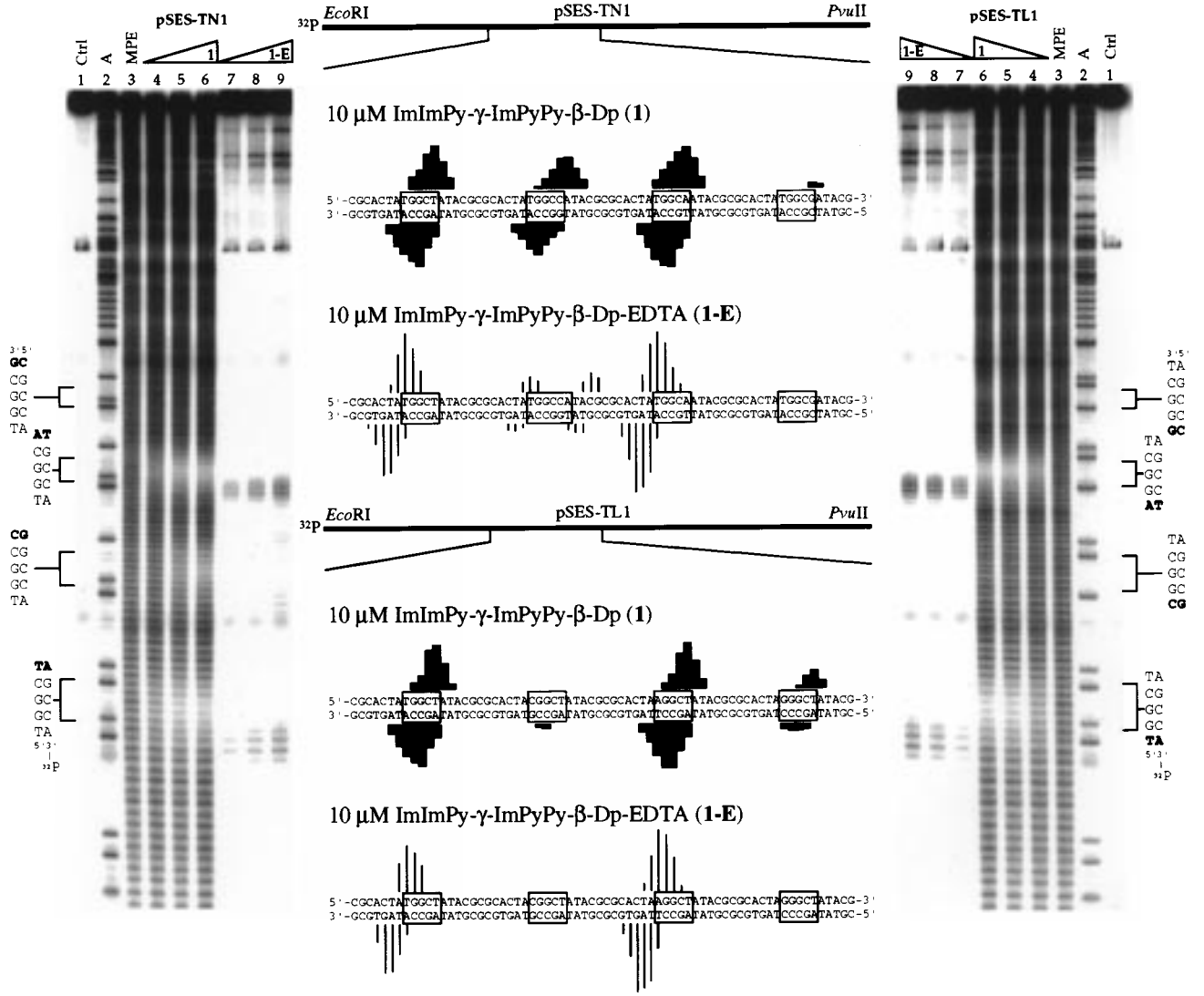


Figure 3. MPE·Fe(II) footprinting and affinity cleaving experiments on the 3′-^{32}P-labeled 286-bp *Eco*RI/*Pvu*II restriction fragments from the plasmids pSES-TN1 (left) and pSES-TL1 (right). For both gels: lane 1, intact DNA; lane 2, A reaction; lane 3, MPE·Fe(II) standard; lanes 4−6, 2, 5, and 10 μM **1**, respectively; lanes 7−9, 2, 5, and 10 μM **1-E**, respectively. The quantitated sites are shown on the side of the autoradiogram. All lanes contain 15 kcpm 3′-radiolabeled DNA. The control and MPE·Fe(II) lanes (1 and 3−6) contain 25 mM Tris-acetate buffer (pH 7.0), 10 mM NaCl, and 100 μM/base pair calf thymus DNA. The affinity cleaving lanes (7−9) contain 25 mM Tris-acetate buffer (pH 7.0), 10 mM NaCl, and 100 μM/base pair calf thymus DNA. (center) Results from MPE·Fe(II) footprinting of **1** and affinity cleavage with **1-E** on the plasmids pSES-TN1 (top) and pSES-TL1 (bottom). For both plasmids, an illustration of the 286-bp restriction fragment with the position of the sequence indicated is shown. 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 (footprinting) or the relative cleavage (affinity cleaving) at each band.

acid and PrOH = propanolamide)[2d,12] was analyzed as a "tailless" control to exclude the possibility of sequence-dependent DNA microstructure effects on the tail specificity and to determine the effect of the C-terminal amide bond.

We report here the affinities, binding locations, and relative selectivities of these polyamides as determined by three separate techniques: MPE·Fe(II) footprinting,[13] affinity cleaving,[14] and DNase I footprinting.[15] Information about binding site size and location is provided by MPE·Fe(II) footprinting, while binding orientations are demonstrated by affinity cleavage assays carried out with the Fe(II)·EDTA analogue (Figure 2). Quantitative DNase I footprint titrations allow the determination of equilibrium association constants (K_a) of the polyamide for its respective match sites.

Results and Discussion

MPE·Fe(II) Footprinting and Affinity Cleavage. MPE·Fe(II) footprinting (25 mM Tris-acetate, 10 mM NaCl, 100 μM/

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

(12) Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382−1391.

(13) (a) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373−2377. (b) Van Dyke, M. W.; Dervan, P. B. *Nucleic Acids Res.* **1983**, *11*, 5555−5567.

(14) (a) Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861−6863. (b) Schultz, P. G.; Dervan, P. B. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1133−1147. (c) Taylor, J. S.; Schultz, P. B.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457−465.

(15) (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132−181. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8462−8466. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344−7354.

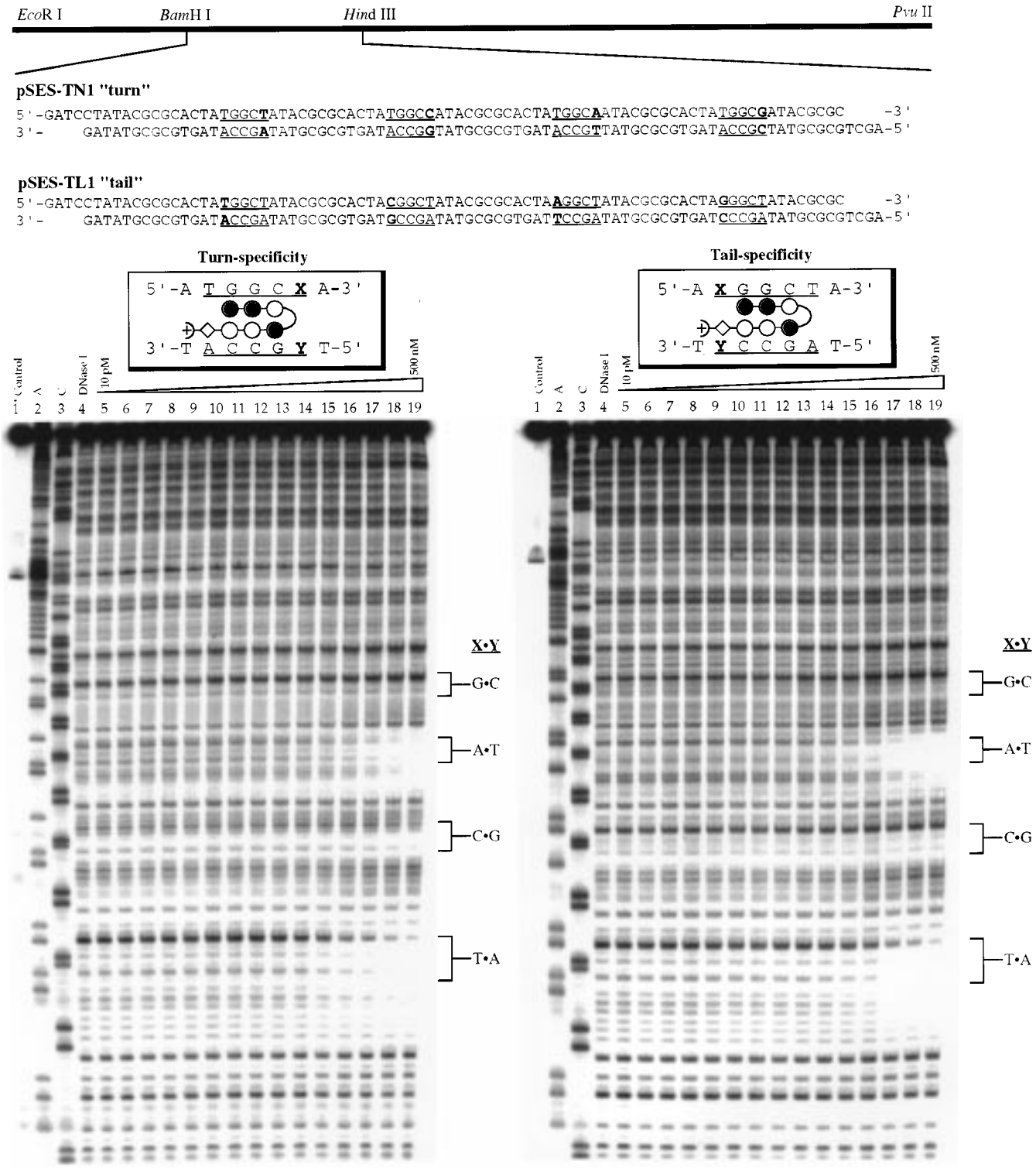


Figure 4. Quantitative DNase I footprint titration experiment with ImImPy-γ-ImPyPy-β-Dp on the *Eco*RI/*Pvu*II restriction fragments derived from plasmids pSES-TN1 (left) and pSES-TL1. (top) Illustration of the *Eco*RI/*Pvu*II restriction fragment with the *Bam*HI and *Hin*dIII insertion sites indicated. Sequences of the synthesized inserts from the plasmids pSES-TN1 "turn" and pSES-TL1 "tail" are shown with the polyamide **1** binding sites underlined and the tail and turn base pairs indicated in bold type. (right) Lane 1, intact DNA; lane 2, A reaction; lane 3, C reaction; lane 4, DNase I standard; lanes 5−19, 10, 20, 50, 100, 200, and 500 pM and 1, 2, 5, 10, 20, 50, 100, 200, and 500 nM **1**, respectively. The 5′-A*TGGCN*A-3′ and 5′-A*NGGCT*A-3′ sites that were analyzed are shown on the right side of the autoradiograms. Ball and stick models are also shown. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent the β-alanine residue. W represents either an A or T base. The 5-bp hairpin binding sites are underlined. All reactions contain 20 kcpm restriction fragment, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

base pair calf thymus DNA, pH 7.0, and 22 °C)[13] was performed on the 3′- and 5′-^{32}P end-labeled 286-bp restriction fragments from the cloned plasmids pSES-TN1 and pSES-TL1 (Figures 3 and 4). pSES-TN1 contains the four possible 5′-A*TGGCN*A-3′ (turn specificity) sites while pSES-TL1 contains the four possible 5′-A*NGGCT*A-3′ (tail specificity) sites, with identical sequences between sites. MPE·Fe(II) footprinting reveals that polyamide **1**, at 10 μM concentration, binds to three of the four

Table 1. Equilibrium Association Constants (M^{-1})[a,b]

binding site	ImImPy-γ-ImPyPy-β-Dp	specificity[c]
5′-A*TGGCA*A-3′	2.0×10^7 (0.3)	>400
5′-A*TGGCT*A-3′	2.1×10^7 (0.1)	>420
5′-A*TGGCC*A-3′	7.4×10^5 (1.4)[d]	>15
5′-A*TGGCG*A-3′	$<5 \times 10^4$[d]	1
5′-A*AGGCT*A-3′	3.1×10^7 (1.1)	>310
5′-A*TGGCT*A-3′	2.1×10^7 (0.8)	>210
5′-A*CGGCT*A-3′	$<10^5$[d]	1
5′-A*GGGCT*A-3′	$<10^5$[d]	1

[a] Values reported are the mean values measured from at least three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. [b] 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 $CaCl_2$. [c] Specificity is calculated as K_a(5′-A*TGGCN*A-3′)/K_a(5′-A*TGGCG*A-3′) for the data from pSES-TN1 and as K_a(5′-A*NGGCT*A-3′)/K_a(5′-A*GGGCT*A-3′) for the data from pSES-TL1. [d] These equilibrium association constants were determined in a 40-μL reaction volume.

binding sites: 5′-A*TGGCT*A-3′, 5′-A*TGGCA*A-3′, and 5′-A*TGGCC*A-3′ (A·T, T·A, and C·G) base pairs at the turn position but not 5′-A*TGGCG*A-3′, which has a G·C base pair at the turn position. For the tail position, MPE·Fe(II) footprinting reveals that polyamide **1**, at 10 μM concentration, binds to only two of the four binding sites on the restriction fragment, 5′-A*TGGCT*A-3′ and 5′-A*AGGCT*A-3′ (A·T and T·A at the tail position), but not 5′-A*GGGCT*A-3′ and 5′-A*CGGCT*A-3′ (G·C and C·G at the turn position). The size and 3′-shift of the protection patterns is consistent with the polyamide binding as a hairpin in the DNA minor groove.

Affinity cleavage experiments were performed (25 mM Tris-acetate, 200 mM NaCl, 50 μg/mL glycogen, pH 7.0, and 22 °C)[15] on the "turn" restriction fragment pSES-TN1 which contains the four possible 5′-A*TGGCN*A-3′ sites (Figure 3). The affinity cleavage assays reveal cleavage patterns that are 3′-shifted and appear on only the 5′-side of the 5′-A*TGGCT*A-3′ and 5′-A*TGGCA*A-3′ sites, as is expected for hairpin formation. Strong affinity cleavage is seen at only the 5′-side of the two recognized sites, consistent with formation of an oriented hairpin polyamide complex in the minor groove. Equal amounts of cleavage are seen on both sides of the 5′-ATGGC**C**A-3′ site, which can be explained by the fact the site contains the palindromic sequence 5′-TGGC**C**A-3′, where two identical sites are present, 5′-AT*GGCCA*-3′ and 5′-A*TGGCC*A-3′.

Quantitative DNase I Footprinting. Quantitative DNase I footprint titrations (10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0, and 22 °C) were performed to determine the equilibrium association constants (K_a) of polyamide **1** for the eight sites on the two restriction fragments (Figure 4 and Table 1).[15−17] For the turn position, the 5′-A*TGGCA*A-3′ and 5′-A*TGGCT*A-3′ sites are bound with equivalent equilibrium association constants of $K_a = 2.0 \times 10^7$ and 2.1×10^7 M^{-1}, respectively. The 5′-A*TGGCC*A-3′ site is bound with 28-fold lower affinity ($K_a = 7.4 \times 10^5$ M^{-1}), while no binding is observable at the 5′-A*TGGCG*A-3′ site ($K_a < 5 \times 10^4$ M^{-1}). The β-tail residue has comparable sequence preference with the 5′-A*AGGCT*A-3′ and 5′-A*TGGCT*A-3′ sites being bound with similar affinities, $K_a = 3.1 \times 10^7$ M^{-1} and $K_a = 2.1 \times 10^7$ M^{-1}, respectively. No binding is seen at either the 5′-A*CGGCT*A-3′ and 5′-A*GGGCT*A-3′ sites ($K_a < 10^5$ M^{-1}), such that turn sites containing A·T and T·A base pairs are bound with >200-fold higher affinity than those with G·C and C·G base pairs.

Footprinting experiments were also performed on the 286-bp restriction fragments (pSES-TN2 and pSES-TL2) which contain the eight possible 5′-**N***TGGCT*A-3′ and 5′-A*TGGCT***N**-3′ sites (Figure 5 and Table 2). **1** again shows preference for A·T and T·A at the turn position adjacent to the 5-bp binding site. However, the magnitude of the discrimination at this position is greatly reduced with equilibrium association constants at the four sites of 5′-A*TGGCT***G**-3′ ($K_a = 1.2 \times 10^7$ M^{-1}) < 5′-A*TGGCT***C**-3′ ($K_a = 2.6 \times 10^7$ M^{-1}) < 5′-A*TGGCT***T**-3′ ($K_a = 3.6 \times 10^7$ M^{-1}) < 5′-A*TGGCTA*-3′ ($K_a = 4.1 \times 10^7$ M^{-1}). Similar effects are seen for polyamide **1** binding at the tail sites with a preference for A·T and T·A over G·C and C·G base pairs. The 5′-**T***TGGCT*A-3′ ($K_a = 2.8 \times 10^7$ M^{-1}) and 5′-**A***TGGCT*A-3′ ($K_a = 6.2 \times 10^7$ M^{-1}) sites are bound with the highest affinity. The and 5′-**C***TGGCT*A-3′ ($K_a = 2.0 \times 10^6$ M^{-1}) and 5′-**G***TGGCT*A-3′ ($K_a = 1.4 \times 10^6$ M^{-1}) sites are bound with 20-fold and 44-fold reduced affinity, respectively (relative to the 5′-**G***TGGCT*A-3′ site).

Sequence-Specificity of the γ-Turn. The energetics of **1** binding against the eight possible sequences which vary at the 3′-end of the site, 5′-A*TGGCN*A-3′ and 5′-A*TGGCT***N**-3′, was examined by quantitative DNase I footprinting. At the turn position 5′-A*TGGCN*A-3′, the hairpin polyamide demonstrates a large preference for A·T base pairs, with both 5′-A*TGGCT*A-3′ and 5′-A*TGGCA*A-3′ bound with an equilibrium association constant of $K_a = 2 \times 10^7$ M^{-1}. A sequence with a C·G base pair at the turn, 5′-A*TGGCC*A-3′, is recognized with 28-fold reduced affinity. A sequence with a G·C base pair in this position, 5′-A*TGGCG*A-3′ is bound with >400-fold reduced affinity. The preference of the γ-turn for A·T and T·A base pairs is undoubtedly due to the steric factors. Presumably, the C-terminal aromatic amide at the turn makes a hydrogen bond to the N3 of A or O2 of T, as was observed for the C-terminal aromatic amide in the crystal structure of the dimeric complex of ImImPyPy (Figure 1).[4f] The preference for C·G over G·C base pairs has not been observed before. This effect is consistent with the preference of paired Py amide residues for placement opposite T, A, or C relative to G. Placement of the turn amide opposite G would likely result in unfavorable steric interaction or an unfavorable electrostatic effect.[6c]

At the turn sequence adjacent to the 5-bp binding site, 5′-A*TGGCT***N**-3′, the hairpin polyamide demonstrates a much smaller preference for A·T base pairs than at the sequences, 5′-A*TGGCN*A-3′. Both 5′-A*TGGCT***A**-3′ and 5′-A*TGGCT***A**-3′ are bound with 2-fold higher affinity than 5′-A*TGGCT***C**-3′ and 3-fold higher affinity than 5′-A*TGGCT***G**-3′. It is interesting to note that the relative preference (A ≈ T > C > G) is observed for polyamide **1** for binding to both 5′-A*TGGCT***N**-3′ and 5′-A*TGGCN*A-3′ sequences. The reduced magnitude of the discrimination is consistent with the compact γ-turn observed in an NMR structure of ImPyPy-γ-PyPyPy-β-Dp bound to a 5′-C*TGTTA*G-3′ site.[9f]

Sequence Specificity of the β-Tail. The energetics of **1** was also examined against the eight possible sequences which vary at the 5′-end of the site, 5′-A***N***GGCT**A-3′ and 5′-**N***TGGCT*A-3′. As seen for the γ-turn, the β-tail amino acid is specific for A·T and T·A base pairs. The 5′-A***A***GGCT**A-3′ and 5′-A***T***G-GCT**A-3′ sequences are bound with $K_a = 3 \times 10^7$ and 2×10^7 M^{-1}, respectively, while the 5′-A***G***GGCT**A-3′ and 5′-A***C***G-GCT**A-3′ sites are not bound appreciably (>300-fold reduced affinity). The preference of C over G, observed for the turn, is

(16) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.
(17) (a) Maxam, A. M.; Gilbert, W. S. *Methods Enzymol.* **1980**, *65*, 499−560. (b) Iverson, B. L.; Dervan, P. B. *Nucleic Acids Res.* **1987**, *15*, 7823−7830.

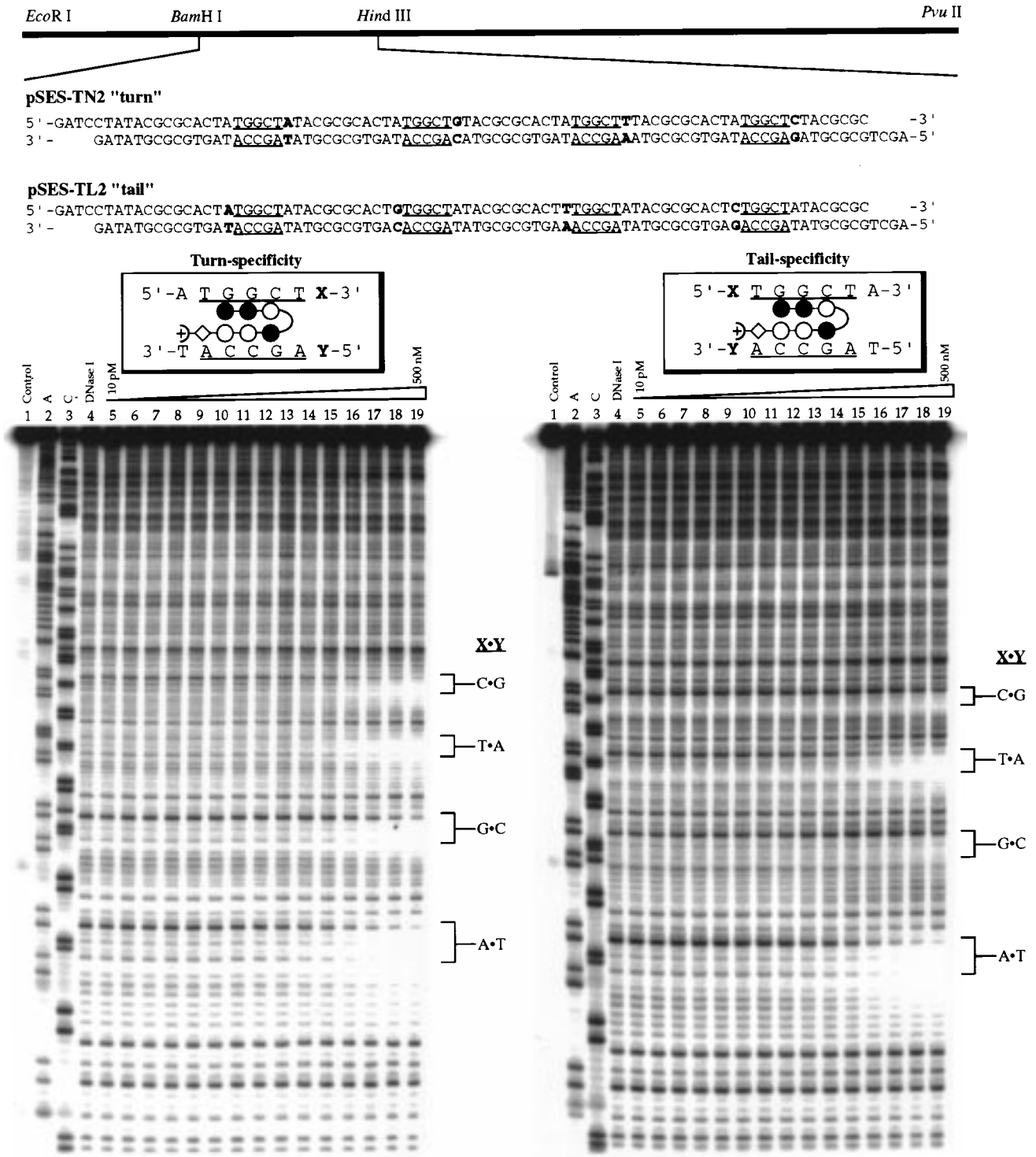


Figure 5. Quantitative DNase I footprint titration experiment with ImImPy-γ-ImPyPy-β-Dp on the *Eco*RI/*Pvu*II restriction fragments from plasmids pSES-TN2 (left) and pSES-TL2 (right). (top) Illustration of the *Eco*RI/*Pvu*II restriction fragment with the sequences of the *Bam*HI and *Hin*dIII inserts from the plasmids pSES-TN2 "turn" and pSES-TL2 "tail". Polyamide **1** binding sites are underlined and the tail and turn base pairs adjacent to the 5-bp binding site are in bold type. Lane 1, intact DNA; lane 2, A reaction; lane 3, C reaction; lane 4, DNase I standard; lanes 5−19, 10, 20, 50, 100, 200, and 500 pM and 1, 2, 5, 10, 20, 50, 100, 200, and 500 nM **1**, respectively. The 5′-A*TGGCT***N**-3′ and 5′-**N***TGGCT*A-3′ sites that were analyzed are shown on the right side of the autoradiograms. All reactions contain 20 kcpm restriction fragment, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. Ball and stick models are also shown as described in Figure 4.

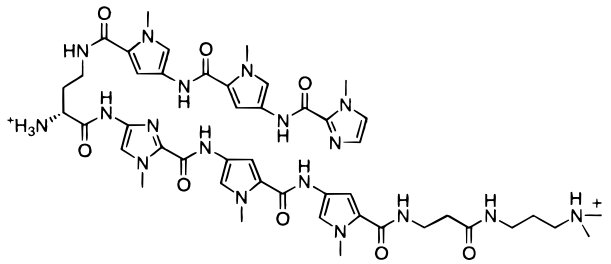
not observed for the tail. This increased specificity likely results from the greater steric bulk of the flexible β-alanine residue relative to the compact folded γ-turn residue.[4f,9f]

Specificity is also observed at the tail position adjacent to the 5-bp binding site, 5′-**N***TGGCT*A-3′. Sequences with A·T and T·A base pairs at this position, 5′-**A***TGGCT*A-3′ and 5′-**T***TGGCT*A-3′, are bound with the highest affinity. The sequences containing G·C and C·G base pairs 5′-**G***TGGCT*A-3′ and 5′-**C***TGGCT*A-3′ are bound with 20-fold lower affinity compared to 5′-**T***TGGCT*A-3′. In the first study of a hairpin polyamide with a C-terminal β-Dp group, it was observed that the hairpin polyamide ImPyPy-γ-PyPyPy-β-Dp bound a 5′-**T***TGTTA*G-3′ sequence with 3-fold higher affinity than the corresponding hairpin ImPyPy-γ-PyPyPy-Dp, which has a

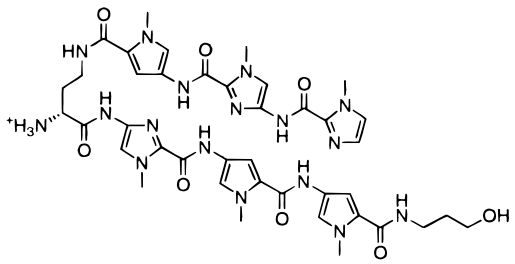
Table 2. Equilibrium Association Constants (M^{-1})[a,b]

binding site	ImImPy-γ-ImPyPy-β-Dp	specificity[c]
5′-A*TGGCT***A**-3′	4.1×10^7 (0.9)	3
5′-A*TGGCT***T**-3′	3.6×10^7 (1.3)	3
5′-A*TGGCT***C**-3′	2.6×10^7 (0.6)	2
5′-A*TGGCT***G**-3′	1.2×10^7 (0.3)	1
5′-**A***TGGCT*A-3′	6.2×10^7 (0.2)	44
5′-**T***TGGCT*A-3′	2.8×10^7 (0.4)	20
5′-**C***TGGCT*A-3′	2.0×10^6 (0.8)	1
5′-**G***TGGCT*A-3′	1.4×10^6 (1.4)	1

[a] Values reported are the mean values measured from at least three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. [b] 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 $CaCl_2$. [c] Specificity is calculated as K_a(5′-ATGGCT**N**-3′)/K_a(5′-ATGGCT**G**-3′) for the data from pSES-TN2 and as K_a(5′-**N**TGGCTA-3′)/K_a(5′-**G**TGGCTA-3′) for the data from pSES-TL2.



ImImPy-(*R*)^{H_2N}γ-ImPyPy-β-Dp (2)



ImImPy-(*R*)^{H_2N}γ-ImPyPy-PrOH (3)

Figure 6. Structures of ImImPy-(*R*)$^{H2N}\gamma$-ImPyPy-β-Dp (**2**) and ImImPy-(*R*)$^{H2N}\gamma$-ImPyPy-PrOH (**3**).

C-terminal Dp group. It was proposed based on this observed binding enhancement that the C-terminal tail could make a favorable interaction with the "sixth" base pair of the binding site.[9b] Subsequently, the hairpin ImPyPy-γ-PyPyPy-β-Dp was found to bind a 5′-**G***TGTTA*C-3′ sequence with 10-fold reduced affinity relative to binding at 5′-**T***TGTTA*G-3′.[9g] Crystal structure analysis of the ImImPyPy-β-Dp dimer bound to the 6-bp sequence 5′-G*TGGCCA*C-3′ revealed that the β-Dp tails (placed opposite the G·C and C·G adjacent to the 6-bp binding site) were disordered and did not sit deeply within the minor groove.[4f] No corresponding high-resolution structure of a polyamide with an A·T or T·A base pair at this position is available; however, the study reported here indicates that there are favorable interactions (perhaps an extra hydrogen bond) between the β-tail and A·T flanking sequences.

Effects of Removing the β-Dp Tail. To confirm that favorable interactions exist between the β-Dp tail and flanking A·T or T·A base pairs, the polyamide ImImPy-(*R*)$^{H2N}\gamma$-ImPyPy-PrOH (**3**) was prepared such that the β-dimethylaminopropylamide tail is replaced by a smaller propanolamide moiety (Figure 6). To maintain a positive charge on the resulting polyamide, (*R*)-2,4-diaminobutyric acid was substituted for the γ-turn.[12] A control polyamide that contains only the turn substitution,

Table 3. Equilibrium Association Constants (M^{-1})[a,b]

binding site	ImImPy-(*R*)$^{H2N}\gamma$-ImPyPy-β-Dp	specificity[c]
5′-**A***TGGCT*A-3′	2.3×10^9 (0.8)	23
5′-**T***TGGCT*A-3′	1.1×10^9 (0.4)	11
5′-**C***TGGCT*A-3′	1.0×10^8 (0.8)	1
5′-**G***TGGCT*A-3′	1.0×10^8 (0.3)	1
binding site	**ImImPy-(*R*)$^{H2N}\gamma$-ImPyPy-PrOH**	**specificity[c]**
5′-**A***TGGCT*A-3′	1.1×10^9 (0.3)	2
5′-**T***TGGCT*A-3′	3.4×10^8 (0.8)	0.5
5′-**C***TGGCT*A-3′	2.7×10^8 (0.4)	0.4
5′-**G***TGGCT*A-3′	7.0×10^8 (2.1)	1

[a] Values reported are the mean values measured from at least three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. [b] 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 $CaCl_2$. [c] Specificity is calculated as K_a(5′-**N***TGGCT*A-3′)/K_a(5′-**N***TGGCT*A-3′) for both compounds.

ImImPy-(*R*)$^{H2N}\gamma$-ImPyPy-β-Dp-Dp (**2**), shows similar tail specificity to polyamide **1** with a 20-fold preference for A·T and T·A base pairs observed at the 5′-**N***TGGCT*A-3′ sites. The binding affinity of polyamide **2** is increased by 50-fold relative to polyamide **1** at each of the 5′-**N***TGGCT*A-3′ sites, consistent with the substitution of (*R*)$^{H2N}\gamma$ for γ.[12] The presence of the chiral α-amino group on the γ-turn did not alter the sequence specificity of the turn position.

In contrast to polyamides **1** and **2**, the polyamide **3**, tolerates all 4 base pairs at the 5′-end of the binding site, 5′-**N***TGGCT*A-3′ (Table 3, Figure 7). Furthermore, elimination of the β-Dp moiety reduces the binding affinity of polyamide **3** by 3-fold relative to polyamide **2** binding at the 5′-**A***TGGCT*A-3′ and 5′-**T***TGGCT*A-3′ sites. This effect is similar in magnitude to the reported difference for ImPyPy-γ-PyPyPy-β-Dp and ImPyPy-γ-PyPyPy-Dp binding to the sequence 5′-**T***TGTTA*G-3′[9b] and is consistent with the β-Dp tail of polyamides **1** and **2** making favorable contact with A·T-rich flanking sequences.

Implications for the Design of Minor Groove-Binding Molecules. Hp-Im-Py polyamides provide a chemical method for targeting predetermined DNA sequences. The recent demonstration that hairpin polyamides are cell permeable[1] and inhibit the transcription of specific genes increases the importance of understanding the sequence specificity of the hairpin structure in detail. Both the γ-turn and β-tail amino acids show strong preferences for A·T and T·A base pairs at the turn and tail positions, respectively. Remarkably, the β-Dp tail results in additional specificity, i.e., not 1 but 2 base pairs flank the core. This binding effect can be regulated by elimination of the C-terminal amide bond. The recent demonstration that certain ring pairings of 3-hydoxypyrrole with pyrrole can discriminate A·T from T·A,[3,7] combined with the observation that polyamide tails, like the Py/Py pair,[5] are degenerate for A·T, T·A base pairs, provides an interesting future challenge to design additional A·T specificity into these linker amino acids. The polyamide end effects observed here now explain the previously reported 10-fold difference in affinity for the hairpin polyamide ImPyPy-γ-PyPyPy-β-Dp binding to 5′-**T***TGTTA*G-3′ and 5′-

(18) (a) Wu, H.; Crothers, D. M. *Nature* **1984**, *308*, 509. (b) Koopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553. (c) Coll, M.; Fredrick, C. A.; Wang, A. H.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8385. (d) Steitz, T. A. *Q. Rev. Biophys.* **1990**, *23*, 205. (e) Goodsell, D. S.; Kopka, M. L.; Cascio, D.; Dickerson, R. E. *Proc. Natl. Sci. U.S.A.* **1993**, *90*, 2930. (f) Paolella, D. N.; Palmer, R.; Schepartz, A. *Science* **1994**, *264*, 1130. (g) Kahn, J. D.; Yun, E.; Crothers, D. M. *Nature* **1994**, *368* 163. (h) Geierstanger, B. H.; Wemmer, D. E. *Annu. Rev. Biochem.* **1995**, *24*, 463. (i) Hansen, M. R.; Hurley, L. H. *Acc. Chem. Res.* **1996**, *29*, 249.

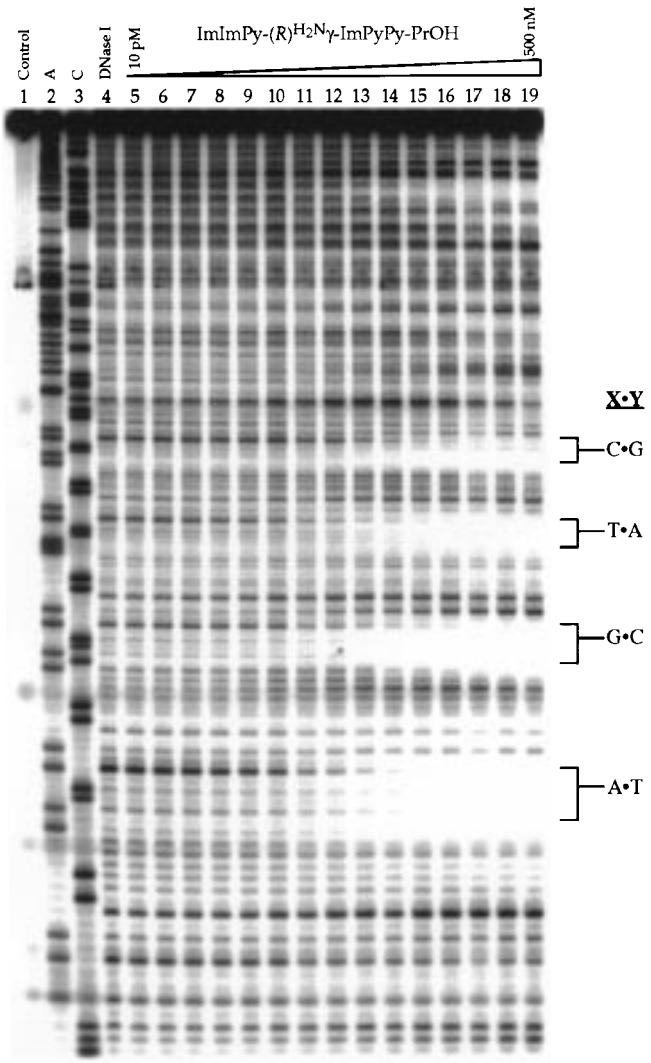


Figure 7. Quantitative DNase I footprint titration experiment with ImImPy-(*R*)$^{H2N}\gamma$-ImPyPy-PrOH (**3**) on the *Eco*RI/*Pvu*II restriction fragment from plasmid pSES-TN2: lane 1, intact DNA; lane 2, A reaction; lane 3, C reaction; lane 4, DNase I standard; lanes 5−19, 10, 20, 50, 100, 200, and 500 pM 1, 2, 5, 10, 20, 50, 100, 200, and 500 nM **3**, respectively. The 5′-**N***TGGCTA*-3′ sites that were analyzed are shown on the right side of the autoradiogram. All reactions contain 20 kcpm restriction fragment, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

G*TGTTAC*-3′ sequences,[9b,g] suggesting that polyamide end effects must be considered for successful use of the pairing rules in conjunction with the less well understood sequence-dependent microstructure of DNA.[2b,18]

Experimental Section

Materials. Enzymes were purchased from Boehringer-Mannheim and used with their supplied buffers. Deoxyadenosine and thymidine 5′-[α-^{32}P]triphosphates were obtained from Amersham, and deoxyadenosine 5′-[γ-^{32}P]triphosphate was purchased from I.C.N. Sonicated, deproteinized calf thymus DNA was acquired from Pharmacia. RNase-free water was obtained from USB and used for all footprinting reactions. All other reagents and materials were used as received. All DNA manipulations were performed according to standard protocols.[16]

Generation of Restriction Fragments. The plasmids pSES-TN1, pSES-TL1, pSES-TN2, and pSES-TL2 were constructed using previously described methods.[2b] Fluorescent sequencing was performed at the DNA Sequencing Facility at the California Institute of Technology and was used to verify the presence of the desired insert. Plasmid concentration was determined at 260 nm using the relationship of 1 OD unit = 50 μg/mL duplex DNA. 3′- and 5′-end-labeled restriction fragments were prepared using published procedures,[2b] as were chemical sequencing reactions.[17]

MPE·Fe(II) Footprinting[13] and Affinity Cleaving.[14] 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 15 kcpm 3′- or 5′-radiolabeled DNA. The solutions were allowed to equilibrate for 4 h and then processed as previously described.[2b]

DNase I Footprinting.[15] All reactions were carried out in a volume of 400 μL, except when noted otherwise. No carrier DNA was used in the binding reaction. 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 for a minimum of 12 h at 22 °C and then processed as previously described. 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 unit/mL) and was allowed to proceed for 5 min at 22 °C. The reactions were stopped by adding 50 mL of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/mL glycogen, and 30 mM bp calf thymus DNA, and then ethanol precipitated. Cleavage products were processed as described previously.[2b] For reactions done in a 40-μL volume, cleavage was initiated by addition of 4 μL of DNase I stock solution (diluted with 1 mM DTT to 0.225 unit/mL) and halted by addition of 10 μL of stop solution.[9d] Equilibrium association constants were determined as previously described.[2b,15]

Acknowledgment. We are grateful to the National Institutes of Health (Grant GM-27681) for research support and National Research Service Awards to S.E.S. and the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B.

JA9830905