

Stereochemically Modified Polyamides for Recognition in the Minor Groove of DNA

Thesis by
David Herman

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2001

(Submitted February 22, 2001)

© 2001

David Herman

All Rights Reserved

To My Family

Acknowledgments

First and foremost I would like to thank my family for their support and understanding during all the trials and tribulations that we have survived. They have provided the solid foundation off which I have been able to strive for excellence no matter how seemingly unattainable. In particular, my deepest regards extend to my father, brother, and sister. My father, Dr. Arthur Lee Herman, has inspired me to no end with his exploration of and fascination for all the intricate facets of life stretching from science to art and back again. A shout goes out to my brother, Eric Herman, whose wisdom and spirit are as ageless as the wind. To my tenacious and courageous sister Leesl Herman, thanks for teaching me that there are no mountains too high to climb, no obstacles too big to overcome. To my best friend Jen, I am eternally grateful to the great goddess Fortuna for blessing me with someone whose stubbornness and impatience are second only to my own. Her smile and laughter are gems that bless our past, present, and future. I anxiously await the rest of our lives together.

I would also like to thank Professor Peter B. Dervan. His excitement and fervor for science has fueled me to some “bunt singles” and also some “home runs”. Thanks for giving a biologist a crack at physical organic chemistry. In addition my appreciation goes out to Professor Carl Parker, Professor Doug Rees, Professor Steve Mayo, and Professor Linda Hsieh-Wilson for residing on my thesis committee and providing me solid advice throughout the years here at CalTech.

My thanks goes out to the current residents of the Dervan labs: Adam Urbach for late night sessions and all the adventures both near and far; Victor “HNIC” Rucker for being himself; Clay Wang and Tammy for hosting fiestas throughout the years; Shane Foister, Jason Belitsky, Doan Nguyen, Nick Wurtz, and John Chevelliet for engaging discussions.

I also must give props to the gone and forgotten crew that has traveled on to graze greener pastures; Christian Melander for prolific collaborations that usually led to the Mexican Hat Dance; Eldon Baird for keeping me focused and teaching me the invaluable silent gaze for those “emergency situations”; Ramesh Baliga for introducing me to the volatile world of RecA and teaching me that fear and awe are a powerful combination; Will Greenberg for being a good friend and a great source information; John Trauger for showing me how it’s done right; David Liberles for teaching me new methods of pasta regeneration; and Ryan Bremer, Sarah White, Professor Paul Floreancig, Professor Anna Mapp, Dr. Aseem Ansari, and Sue Swalley for making the time spent at CalTech memorable. In closing I would also like to give thanks to Goleta the Goodland, Kutty, wasabi, D, Shiloh, Ash, Mickeys, Sammy David Jr., Takenya, chocolate milk, Jose, top-to-bottom over head offshore conditions, PIC lines, Splat, Ernie’s, Jon, VJ, whip cream, Brown Dave, Dre, Guinness, the Ath, hoops, Fishbone, 007, CC, Brown Sugar, Trejo, the Academy, vancomycin, Princess Bride, leftovers, P-32, Professional sports ortho, mom, magic, Abe, Dr. Freidman, sushi, -80 margaritas, Professor Campbell, and Ninja.

Abstract

The design of synthetic molecules that recognize specific sequences of DNA is an ongoing challenge in molecular medicine. Cell-permeable small molecules targeting predetermined DNA sequences offer a potential approach for offsetting the abnormal effects of misregulated gene-expression. Over the past twenty years, Professor Peter B. Dervan has developed a set of pairing rules for the rational design of minor groove binding polyamides containing pyrrole (Py), imidazole (Im), and hydroxypyrrole (Hp). Polyamides have illustrated the capability to permeate cells and inhibit transcription of specific genes *in vivo*. This provides impetus to identify structural elements that expand the repertoire of polyamide motifs with recognition properties comparable to naturally occurring DNA binding proteins. Through the introduction of chiral amino acids, we have developed chiral polyamides with stereochemically regulated binding characteristics. In addition, chiral substituents have facilitated the development of new polyamide motifs that broaden binding site sizes targetable by this class of ligands.

Table of Contents

	page
Acknowledgments	iv
Abstract	vi
Table of Contents	vii
List of Figures and Tables	viii
CHAPTER ONE: Introduction	1
CHAPTER TWO: Stereochemical Control of the DNA Binding Affinity, Sequence-Specificity, and Orientation-Preference of Chiral Hairpin Polyamides in the Minor Groove..	15
CHAPTER THREE: Tandem Hairpin Motif for Recognition in the Minor Groove of DNA by Pyrrole-Imidazole Polyamides.....	60
CHAPTER FOUR: Cycle-Polyamide Motif for Recognition of Minor Groove of DNA.....	86
CHAPTER FIVE: Discrimination of A/T Sequences in the Minor Groove of DNA within a Cyclic Polyamide Motif... ..	117
CHAPTER SIX: Hp/Py Hairpin Polyamides for the Discrimination of Core A•T Sequences in 5'-GWWC-3'.....	146
CHAPTER SEVEN: Targeting the HIV-1TATA Box with a 2-β-2 Cyclic Polyamide Motif... ..	171

List of Figures and Tables

CHAPTER ONE	page
Figure 1.1 Model of protein regulation of gene expression	2
Figure 1.2 Structure of double-helical B-form DNA	3
Figure 1.3 Model for DNA minor groove recognition	4
Figure 1.4 X-ray crystal structures of DNA-binding proteins	5
Figure 1.5 Structures of DNA-binding molecules from natural sources	5
Figure 1.6 Modes of Distamycin:DNA complexes	6
Figure 1.7 Schematic representation of the polyamide pairing rules.....	8
Figure 1.8 Structural comparison of hairpin and cycle motifs.....	9
Figure 1.9 Summary of polyamide chronology	11
CHAPTER TWO	
Figure 2.1 H-bonding model of chiral six-ring hairpin polyamides	19
Figure 2.2 CPK model of chiral hairpin polyamides	20
Figure 2.3 Structures of chiral hairpin polyamides	22
Figure 2.4 Synthesis of chiral hairpin polyamides.....	25
Figure 2.5 MPE footprinting of polyamides with match and mismatch sites.....	28
Figure 2.6 MPE data analysis.....	29
Figure 2.7 Affinity cleaving of polyamides with forward and reverse sites.....	32
Figure 2.8 Affinity cleavage data analysis.....	34
Figure 2.9 Affinity cleavage patterns for forward and reverse orientations	35
Figure 2.10 DNase I footprinting gels.....	37
Figure 2.11 Isotherms generated from DNase I footprinting gels	38
Figure 2.12 Model for chiral hairpin folding	43
Table 2.1 Equilibrium association constants for six-ring chiral hairpins	
	39

CHAPTER THREE

Figure 3.1	H-bonding model for tandem hairpin polyamide	63
Figure 3.2	Structures of tandem hairpin polyamides	64
Figure 3.3	Synthesis of tandem hairpin polyamides	67
Figure 3.4	Partial sequences of restriction fragments	68
Figure 3.5	Dnase I, MPE footprinting, and affinity cleavage experiments.....	69
Table 3.1	Equilibrium association constants for six ring tandem hairpins.....	71
Table 3.2	Specificities for six-ring tandem hairpins.....	73

CHAPTER FOUR

Figure 4.1	Binding models of unlinked, hairpin, and cyclic polyamides	90
Figure 4.2	Structures of unlinked, hairpin, and cyclic polyamides	91
Figure 4.3	Synthesis of Boc-Py-Pam-resin	92
Figure 4.4	Synthesis of hairpin and cyclic polyamides.....	94
Figure 4.5	MPE footprinting of polyamides with match and mismatch sites.....	97
Figure 4.6	MPE footprinting of polyamides with match and mismatch sites.....	98
Figure 4.7	Affinity cleaving data patterns for match and mismatch sites.....	99
Figure 4.8	DNase I footprinting experiments.....	101
Figure 4.9	Comparison of six-ring and eight-ring cycles	104
Table 4.1	Equilibrium association constants for three polyamide motifs.....	102

CHAPTER FIVE

Figure 5.1	H-bonding models of cyclic Hp polyamides	121
Figure 5.2	Structures of cyclic Hp polyamides	123
Figure 5.3	Synthesis of control hairpin Hp polyamides.....	124
Figure 5.4	Synthesis of cyclic Hp polyamides.....	127
Figure 5.5	Sequence and polyamide binding models on pDEH10	130
Figure 5.6	DNase I footprinting experiments.....	131

Figure 5.7	DNase I footprinting experiments.....	132
Table 5.1	Equilibrium association constants for cyclic Hp polyamides.....	133
Table 5.2	Specificities of cyclic Hp polyamides	134

CHAPTER SIX

Figure 6.1	Hp substitution sites in an eight-ring hairpin polyamide.....	150
Figure 6.2	Structures of hairpin Hp polyamides	151
Figure 6.3	Synthesis of hairpin Hp polyamides	153
Figure 6.4	DNase I footprinting experiments on control polyamide	155
Figure 6.5	DNase I footprinting experiments on a 5'- Hp polyamide.....	157
Figure 6.6	DNase I footprinting experiments on a 5'- internal Hp polyamide.....	158
Figure 6.7	DNase I footprinting experiments a double 5'- Hp polyamide	159
Figure 6.8	DNase I footprinting experiments 3'-shifted Hp polyamides	161
Table 6.1	Equilibrium association constants for control Hp polyamide.....	154
Table 6.2	Equilibrium association constants for 5'-shifted Hp polyamides	156
Table 6.3	Equilibrium association constants for 3'-shifted Hp polyamides	160
Table 6.4	Equilibrium association constants for contiguous Hp polyamides.....	162

CHAPTER SEVEN

Figure 7.1	2- β -2 hairpin and cycle binding models on the HIV-1 promoter	176
Figure 7.2	Structures of 2- β -2 cyclic polyamides	177
Figure 7.3	Synthesis of 2- β -2 cyclic polyamides.....	179
Figure 7.4	DNase I footprinting experiments on a singly charged 2- β -2 cycle.....	182
Figure 7.5	DNase I footprinting experiments on a doubly charged c	183
Figure 7.6	DNase I footprinting experiments on a control 2- β -2 cycle	184
Table 7.1	Equilibrium association constants for 2- β -2 cyclic polyamides.....	185

Chapter 1

Introduction

Background

The relentless drive to sequence the human genome has revealed that our genetic information, stored in helical DNA polymers, consists of 3.2 gigabase pairs that encode for 25,000-30,000 genes.¹ Specific classes of proteins such as activators interact with DNA to control the expression of genes (Figure 1.1).² Misregulation of gene expression resulting in aberrant biochemical behavior in cellular processes is responsible for numerous cancerous and viral disease states.³ Synthetic small cell permeable molecules that bind to predetermined DNA sequences and regulate gene expression⁴ would be important tools in molecular medicine for approaching such life-threatening diseases.

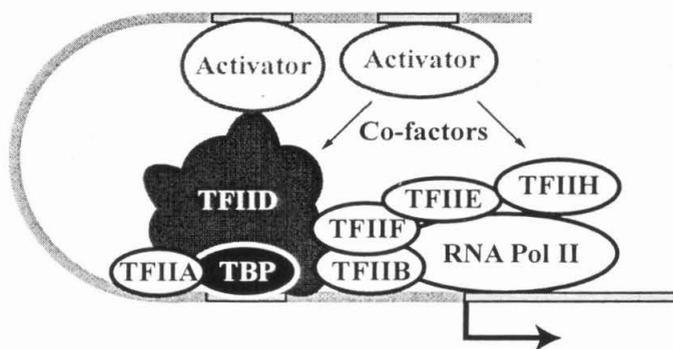


Figure 1.1. Model of protein regulated gene expression illustrating the myriad of protein complexes (> 2,000,000 MW) recruited by activators to initiate transcription.

Our genetic “blueprint” is stored on two antiparallel oligodeoxyribonucleic acid strands that form a double helical structure.⁵ The double helix consists of the four heterocycles bases adenine (A), thymidine (T), guanine (G), and cytosine (C). In duplex DNA, the two strands are held together via Watson-Crick hydrogen bonding of A,T and G,C base pairs.⁶ The common B-form of DNA is characterized by a wide (12 Å) and shallow major groove and a narrow (4-6 Å) and deep minor groove (Figure 1.2).⁷ In addition, sequence-dependent structural variations, conformational properties, and solvent and counterion organization can distinguish local DNA structures.⁷ Individual sequences may be distinguished by the pattern of hydrogen bond donors and acceptors

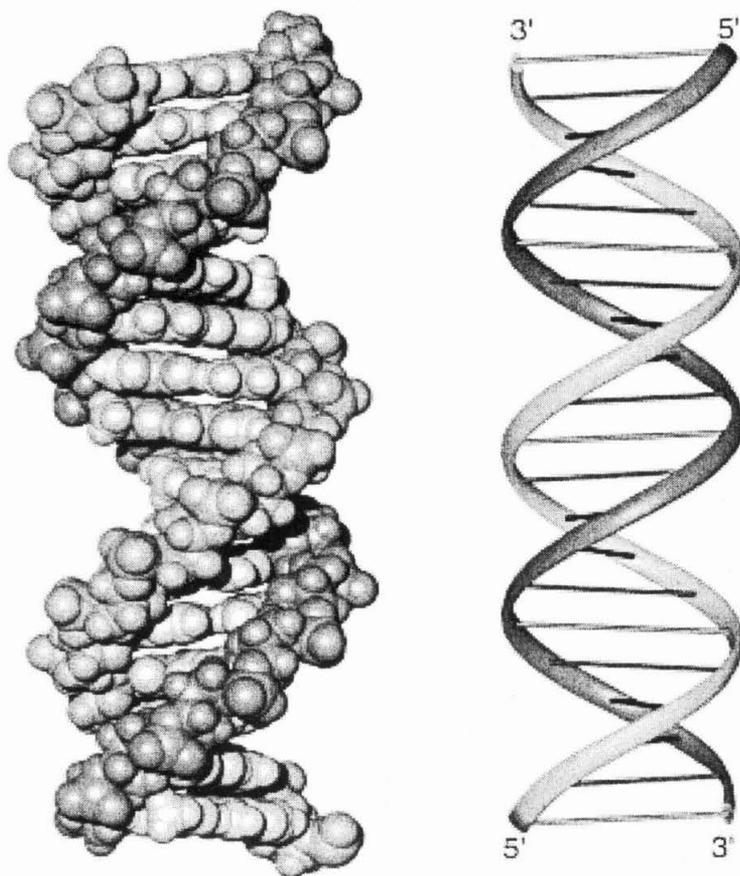


Figure 1.2. Structure of B-form DNA. (left) CPK model of double helix where separate strands are darkly and lightly shaded.

displayed on the edges of the base pairs (Figure 1.3).⁷ In the minor groove, the A,T base pair presents two symmetrically placed hydrogen bond acceptors, the purine N3 and the pyrimidine O2 atoms. The G,C base pair presents these two acceptors, but in addition presents a hydrogen bond donor, the exocyclic 2-amino group of guanine (Figure 1.3).⁸

DNA Recognition. X-ray crystal and NMR structural analyses of naturally occurring protein-DNA complexes and natural products has provided insight into the design of novel DNA binding molecules.^{8,9} Sequence-specificity arises from several methods of DNA interaction: specific hydrogen bonding or van der Waals contacts with functional

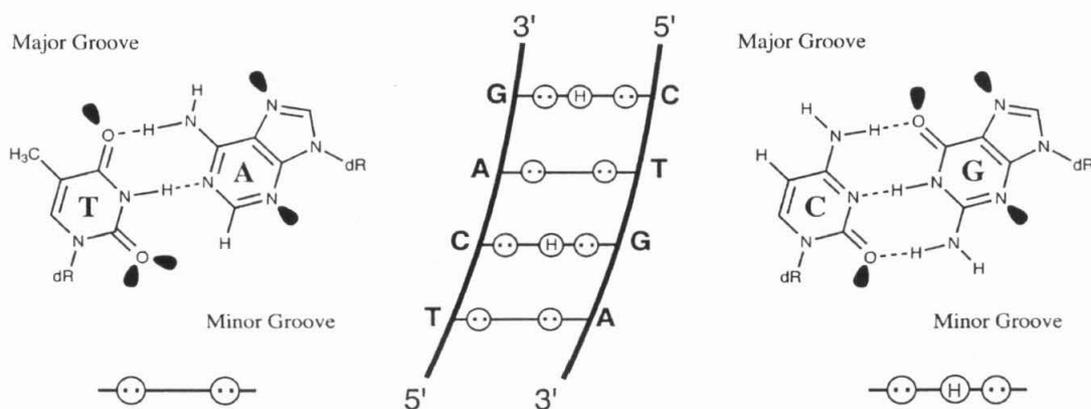


Figure 1.3 Schematic model for recognition of the minor groove, with hydrogen bond donors represented as (H) and hydrogen bond acceptors represented as two dots. This schematic underscores the potential coded reading of the DNA helix.

groups in the grooves, Coulombic attraction to the negatively charged phosphodiester backbone or to the electrostatic potential in the grooves, and/or intercalation of aromatic functional groups between the DNA bases.¹⁰ In cells, nature has devised a combinatorial approach for DNA recognition, utilizing a twenty amino acid code to form complex tertiary folded structures. DNA binding proteins adopt several structural motifs for sequence-specific recognition including the zinc finger,¹¹ the leucine zipper,¹² and the helix-turn-helix motifs (Figure 1.4).¹³ Within these complexes, specificity for target sites is achieved through specific noncovalent interactions between the protein side chains and the nucleobases and phosphates of the DNA. However, no single motif exists that represents a general amino acid-base pair code for all DNA sequences.¹⁴ Although certain zinc finger-DNA complexes¹⁵ are providing a versatile recognition code, the de novo design of proteins for DNA recognition proves to be challenging due to protein structural diversity and limitations in predicting protein folding.

DNA-Binding Small Molecules. In parallel with the evolution of proteins as recognition moieties, nature has provided an alternative class of structurally diverse small molecules

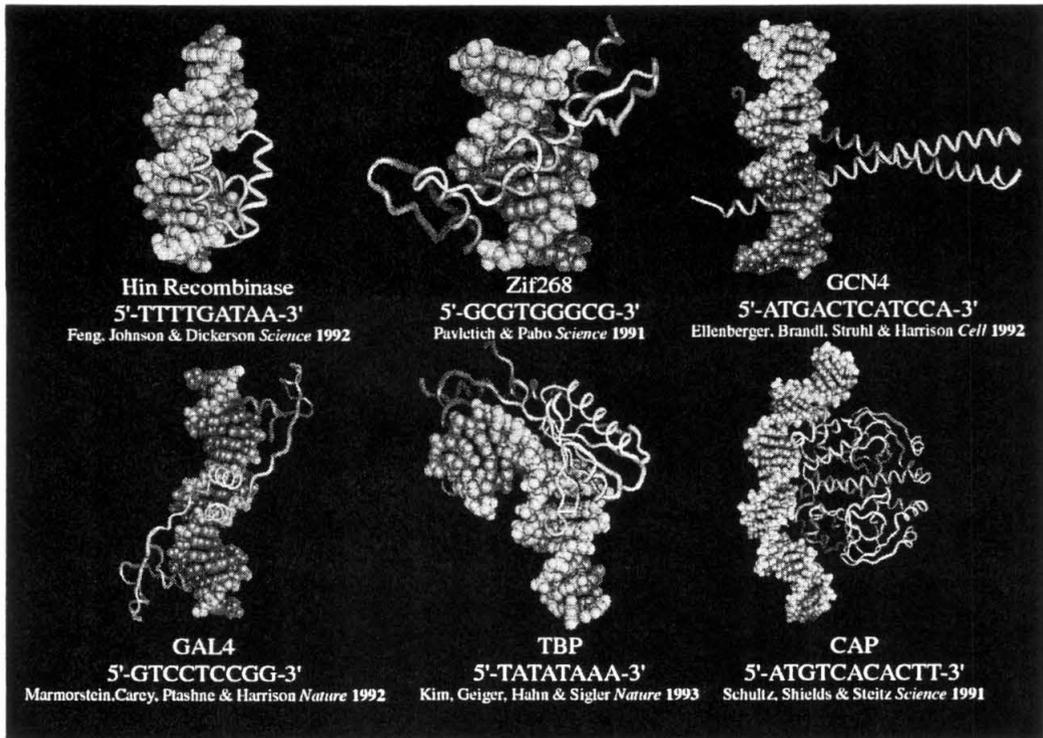


Figure 1.4. X-ray crystal structures of several DNA binding proteins illustrating the variety of motifs nature utilizes for DNA recognition.

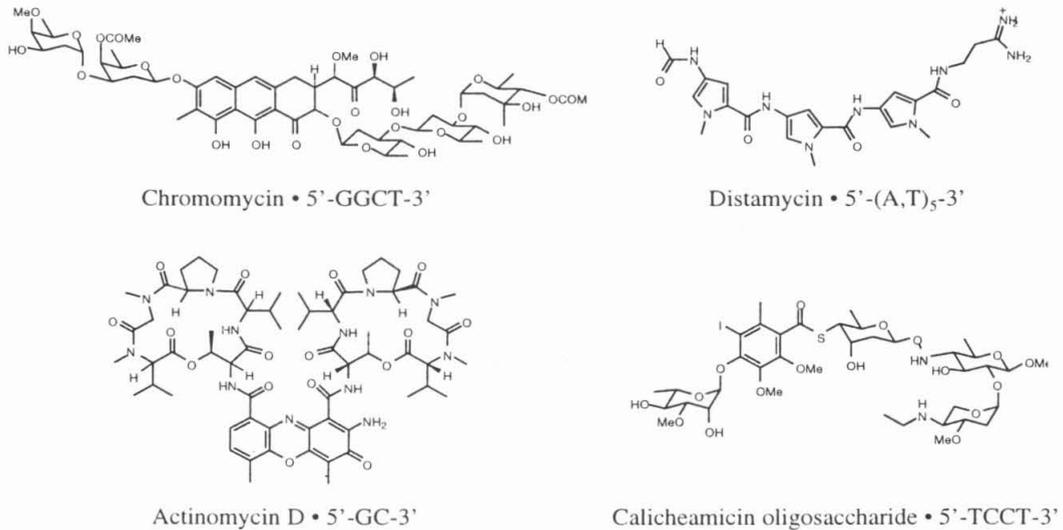


Figure 1.5. Structures of naturally occurring DNA binding small molecules.

that bind DNA. Examples include chromomycin, actinomycin D, calicheamicin, and distamycin A (Figure 1.5).^{16,17} The binding modes of these ligands ranges from minor groove recognition to base pair intercalation or both and display no general recognition code for specific DNA sequences. However, among these DNA binding molecules, distamycin is distinguished by its structural simplicity with an oligopyrrolicarboxamide core and no chiral centers.^{4,18} Structural studies of distamycin-DNA complexes reveal that distamycin is a crescent shaped ligand with modular recognition subunits, in which

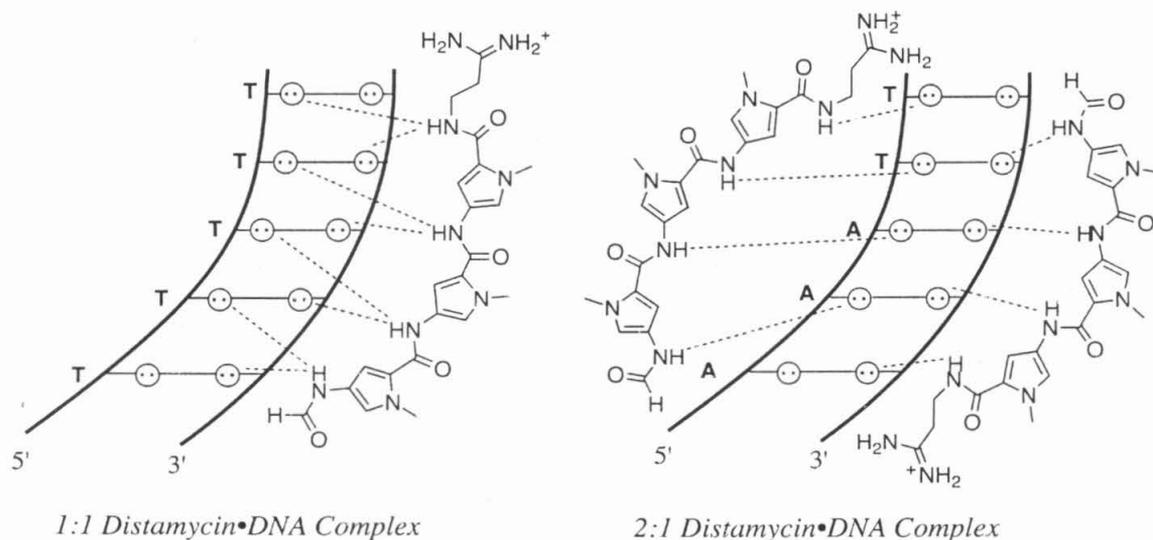


Figure 1.6. Schematic representation the two modes of distamycin:DNA complexes.

adjacent pyrrolicarboxamides make contacts with adjacent DNA base pairs (Figure 1.6).¹⁷ The structural simplicity of distamycin relative to other DNA binding molecules, with respect both to its chemical structure and its complexes with DNA, makes this ligand an attractive scaffold for the design of polyamides with novel DNA recognition properties.⁴

Distamycin forms two unique complexes with DNA (Figure 1.6). Initial observations revealed that a single molecule of distamycin forms a 1:1 ligand:DNA complex by binding in the middle of the minor groove of 5 base pair A,T-rich sequences.

The amide hydrogens of the *N*-methylpyrrolecarboxamides form bifurcated hydrogen bonds with the N3 of adenine and the O2 of thymine on the floor of the minor groove.¹⁷ At higher concentrations a 2:1 ligand:DNA complex is formed where two distamycin ligands form an antiparallel side-by-side dimer in the minor groove of A,T-rich sites.¹⁹ In the 2:1 complex, each ligand subunit selectively forms hydrogen bonds to an adjacent DNA strand. In both binding motifs, the pyrrole rings fill the groove completely, forming extensive van der Waals contacts with the walls of the groove. Specificity for A,T sequences arises from a steric clash between the deeply set *N*-methylpyrrole ring aromatic hydrogens and the exocyclic 2-amino group of guanine bases.

DNA Recognition by Polyamides. Over the past two decades, the research efforts led by Prof. Peter B. Dervan of Caltech have resulted in a set of pairing rules to guide polyamide design for sequence-specific recognition of DNA. A binary code has been developed to correlate DNA sequence with side-by-side pairings between pyrrole (Py), imidazole (Im), and 3-hydroxypyrrole (Hp) carboxamides in the DNA minor groove (Figure 1.7).²⁰ A pairing of Im opposite Py (Im/Py) targets a G•C base pair, while Py/Im targets C•G.²⁰ A Py/Py pairing is degenerate, targeting both A•T and T•A base pairs.²⁰ An Hp opposite a Py (Hp/Py) discriminates T•A from A•T, while Py/Hp targets A•T in preference to T•A and both of these from G•C and C•G. Footprinting, NMR, and X-ray structure studies validate these pairing rules for DNA minor groove recognition.²¹

Efforts have been made to increase the DNA-binding affinity and specificity of pyrrole-imidazole polyamides by covalently linking polyamide subunits.²² Polyamide dimers linked with a γ -aminobutyric acid (γ) turn residue provides a hairpin motif that binds target sites with two to three orders of magnitude enhanced affinity and has the important feature that ring pairings are unambiguously set in place as relative to homodimers which can afford slipped motifs.²³ Increasing the hairpin polyamide subunit

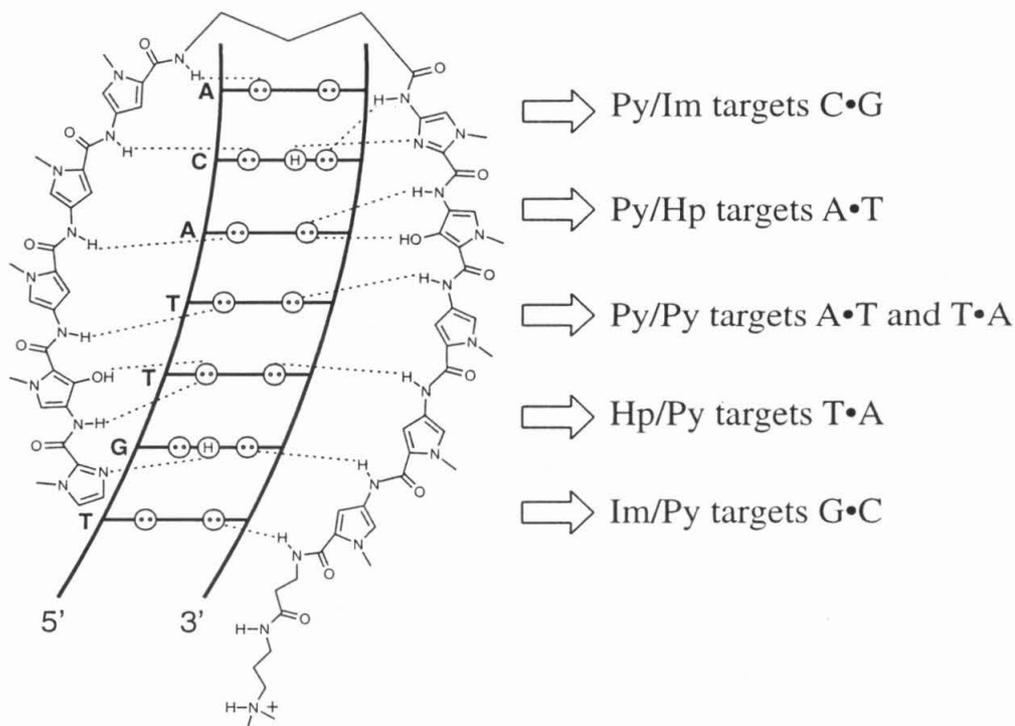


Figure 1.7. Schematic representation of polyamide pairing rules.

lengths to four and five aromatic ring residues gives eight- and ten-ring hairpin polyamides with subnanomolar binding affinities, similar to naturally occurring DNA-binding proteins.²⁴ Studies of polyamide site size limitations suggest that beyond five consecutive rings, the ligand curvature fails to match the pitch of the DNA helix, disrupting the hydrogen bonds and van der Waals interactions responsible for specific polyamide-DNA complex formation.²⁵ The recognition of seven base pairs by ten-ring hairpin polyamides containing five contiguous ring pairings represents the upper limit in binding site sizes targetable by the hairpin motif.²⁶ Addition of pairings with β -alanine (β) to form, β/β , β/Py , and β/Im pairs has extended the targetable binding site size recognizable by the hairpin motif.²⁷ The hairpin polyamide model is supported by footprinting, affinity cleaving, and NMR structure studies.^{22,28}

Covalently tethering the ends of a hairpin affords a cyclic polyamide motif that binds target sequences with higher affinities due to the limited number of conformers

available for the molecule.²⁹ The six-ring polyamide cyclo-(ImPyPy- γ -PyPyPy- γ) was shown to bind its target 5-bp sequence with a 40-fold increase in binding energetics relative to its hairpin analog (Figure 1.8). However, sequence specificity for a single base pair mismatch decreased by a factor of 15 for this cyclic polyamide.

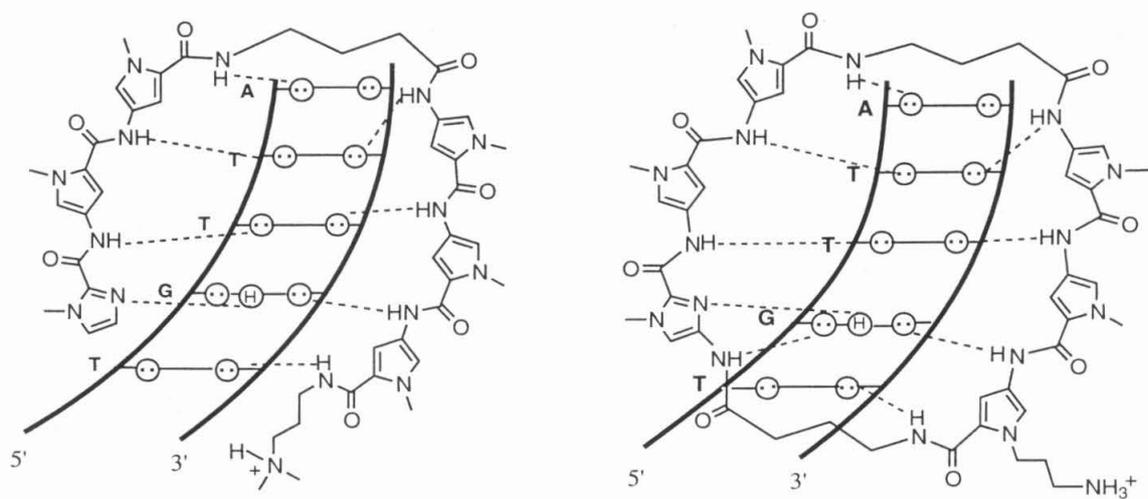


Figure 1.8. Hydrogen bonding model of hairpin-polyamide and cyclic-polyamide motifs.

Description of this Work. Polyamides as a class of cell permeable small molecules with potential for use in molecular medicine. The ongoing development of polyamide motifs with DNA binding properties comparable to naturally occurring proteins remains a high priority (Figure 1.9). This thesis describes work centered on the identification and use of chiral ligand-structure elements in the development of new polyamide motifs with improved DNA recognition properties. Polyamides were designed to inhibit regulatory proteins to expand the knowledge base of using this class of small molecules as regulators of gene expression in vivo.

Chapter Two describes the effects that replacement of the γ -turn with the chiral enantiomers of 2,4-diaminobutyric acid has on the DNA binding properties of a six ring hairpin polyamides. Quantitative footprint titration experiments demonstrated that we could stereochemically regulate polyamide binding through enantiomeric selection.

Consistent with modeling studies, substitution of (γ) with (*R*)-2,4-diaminobutyric acid was amenable to stable hairpin polyamide:DNA complexes and, furthermore, enhanced the binding properties of hairpins with respect to affinity and specificity. Chapter Three describes the use of chiral hairpins to facilitate the design of β -alanine or valeric acid linked tandem hairpin polyamide motif. Characterization of tandems illustrated that six ring hairpins linked with β -alanine off the chiral turn, bind 11 base pair sequences at picomolar concentrations.

Chapters Four, Five and Seven use the chiral turn to reinvestigate the cyclic polyamide motif. Chapter Four presents cycles a promising motif for DNA recognition as DNase I footprint titrations demonstrate that an eight ring cycle binds predetermined target sequence with higher affinities and specificities relative to its hairpin counterpart. Chapter Five investigates the eight ring hydroxypyrrole cycles that recognize all 4 base pairs. This is the first demonstration of the Hp monomer used in cycles and illustrates that cyclization is a viable method to offset compromised binding properties often seen due to the introduction of the Hp monomer. Chapter Seven introduces a β - β pair into the center of an eight-ring cycle to make a 2- β -2 cyclic polyamide. The design of this motif is the most recent addition to generations of polyamides targeting endogenous DNA sequences of regulatory proteins that bind the HIV-1 promoter.

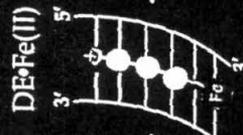
Chapter Six explores the diverse utility of the hydroxypyrrole monomer as a relatively unexplored recognition element. By rotating single and multiple Hp residues throughout the four central positions of an eight ring hairpin provides insight into the optimal placement, number, and orientation of Hp residues for polyamide design.



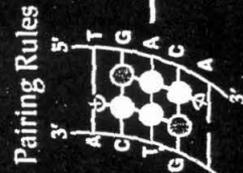
1978
scatchard plots
 $K_{d1} = 10^{11} M^{-1}$



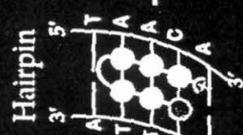
1982



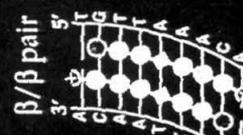
1982



1992
 $K_{d1} = 10^5 M^{-1}$



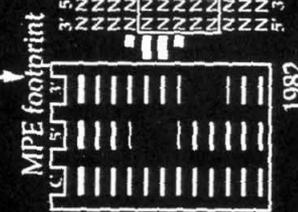
1993
 $K_{d1} = 10^7 M^{-1}$



1996
 $K_{d1} = 10^{10} M^{-1}$



1996
 $K_{d1} = 10^9 M^{-1}$

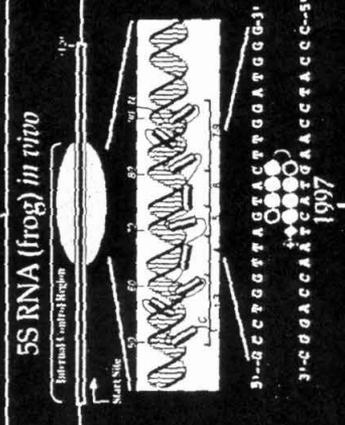


1982

- study weak bonds in water
- combine synthetic organic, physical organic, and biology

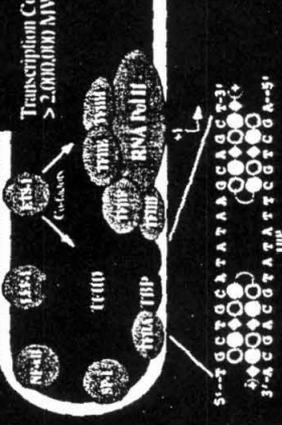
1975

Regulation of Gene Expression



1997

HIV-1 (human T-cell) *in vivo*

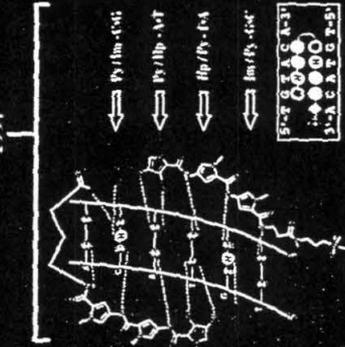


1998



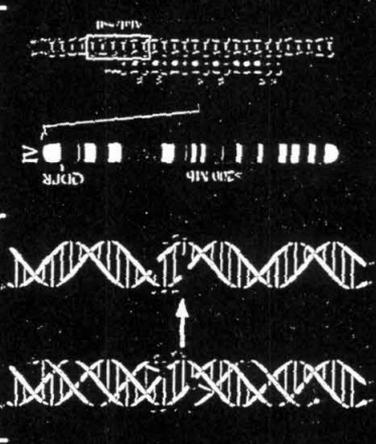
1992

X-ray structure



Hp/Py Pairing Code

1998



Cleavage in yeast

1992

References

1. (a) Ventner, C. et. al *Science* **2001**, 291, 1155 (b) International Human Genome Sequencing Consortium *Nature* **2001**, 409, 860.
2. Roeder, R. G. *TIBS* **1996**, 135, 309.
3. Tjian, R., *Sci. Am.* **1995**, 2, 54.
4. Gottesfeld, J.M.; Neely, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1997**, 387, 202.
5. Watson, J.D.; Crick, F.H.C. *Nature* **1953**, 171, 737.
6. Dickerson, R.E.; Drew, H.R.; Conner, B.N.; Wing, M.; Fratini, A.K.; Koopka, M.L. *Science* **1982**, 216, 475.
7. Saenger, W. In *Principles of Nucleic Acid Structure*; Springer-Verlag: New York.
8. Steitz, T.A. *Quart. Rev. Biophys.* **1990**, 23, 205.
9. Zimmer, C.; Wahnert, U. *Prog. Biophys. Mol. Biol.* **1986**, 47, 31-112.
10. (a) Pabo, C.O.; Sauer, R.T. *Annu. Rev. Biochem.* **1992**, 61, 1053. (b) Freemont, P.S.; Lane, A.N.; Sanderson, M.R. *Biochem. J.* **1991**, 278, 1. (c) Kim, Y.; Geiger, J.H.; Hahn, S.; Sigler, P.B. *Nature* **1993**, 365, 512. (d) Kim, J.L.; Burley, S.K. *Nature Struct. Biol.* **1994**, 1, 638. (e) Nielsen, P.E. *Bioconjugate Chem.* **1991**, 2, 1.
11. Pavletich, N.P.; Pabo, C.O. *Science* **1991**, 252, 809.
12. Ellenberger, T.E.; Brandl, C.J.; Struhl, K.; Harrison, S.C. *Cell* **1992**, 71, 1223.
13. Feng, J.A.; Johnson, R.C.; Dickerson, R.E. *Science* **1994**, 263, 348.
14. Kissinger, C.R.; Liu, B.; Martin-Blanco, E.; Kornberg, T.B.; Pabo, C.O. *Cell* **1990**, 63, 579.
15. Choo, Y.; Klug, A. *Curr. Opin. Struct. Biol.* **1997**, 7, 117.
16. (a) Gao, X.L.; Mirau, P.; Patel, D.J. *J. Mol. Biol.* **1992**, 223, 259. (b) Kamitori, S.; Takusagawa, F. *J. Mol. Biol.* **1992**, 225, 445. (c) Paloma, L.G.; Smith, J.A.; Chazin, W.J.; Nicolaou, K.C. *J. Am. Chem. Soc.* **1994**, 116, 3697.

17. Coll, M.; Frederick, C.A.; Wang, A.H.J.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8385.
18. (a) Zimmer, C. *Prog. Nucleic Acid Res. Mol. Biol.* **1975**, *15*, 285. (b) Baguley, B.C. *Molecular and Cellular Biochem.* **1982**, *43*, 167.
19. (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. *Nature Struct. Biol.* **1994**, *1*, 169. (d) Kielkopf, C.L.; Baird, E.E.; Dervan, P.B.; Rees, D.C. *Nature Struct. Biol.* **1998**, *5*, 104.
20. (a) Wade, W.S.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (c) 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. (d) Wade, W.S.; Mrksich, M.; Dervan, P.B. *Biochemistry* **1993**, *32*, 11385. (e) White, S.; Baird, E.E.; Dervan, P.B. *Biochemistry* **1996**, *35*, 12532.
21. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 9892. (b) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646. (c) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 3325.
22. (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, *J. Am. Chem. Soc.* **1996**, *118*, 6147. (c) M. E. Parks, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 6153. (d) J. W. Trauger, E. E. Baird, P. B. Dervan, *Chem. & Biol.* **1996**, *3*, 369. (e) S. E. Swalley, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 8198. (f) D. S. Pilch, N. A. Pokar, C. A. Gelfand, S. M. Law, K. J. Breslauer, E. E. Baird, P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306. (g) R. P. L. de Claire, B. H. Geierstanger M. Mrksich, P. B. Dervan, D. E. Wemmer, *J. Am. Chem. Soc.* **1997**, *119*, 7909. (h) S. White, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1997**, *119*, 8756. (i) S. White, E. E. Baird, P. B. Dervan, *Chem & Biol.* **1997**, *4*, 569.

23. J. W. Trauger, E. E. Baird, M. Mrksich, P. B. Dervan, *J. Am Chem. Soc.* **1996**, *118*, 6160-6166.
24. (a) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1996**, *382*, 559. (b) Turner, J.M.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, *119*, 7636.
25. C. L. Kielkopf, E. E. Baird, P. B. Dervan, D. C. Rees *Nature Struct. Biol.* **1998**, *5*, 104.
26. J. J. Kelly, E. E. Baird, P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981.
27. J. M. Turner, S. E. Swalley, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1998**, *120*, 6219.
28. De Claire, R.P.L.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1997**, *119*, 7909.
29. Cho, J.; Parks, M.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10389.

Chapter 2

**Stereochemical Control of the DNA Binding Affinity,
Sequence-Specificity, and Orientation-Preference of
Chiral Hairpin Polyamides in the Minor Groove**

Abstract: Three-ring polyamides containing pyrrole (Py) and imidazole (Im) amino acids covalently coupled by γ -aminobutyric acid (γ) form six-ring hairpins that recognize five base pair sequences in the minor groove of DNA. Selective chiral substitution of the " γ -turn" enhances the properties of polyamide-hairpins with regard to DNA affinity and sequence-specificity. Polyamides of core sequence composition ImPyPy- γ -PyPyPy- β which differ by selective stereochemical substitution of the prochiral α -position in the γ -turn were prepared. The DNA binding properties of two enantiomeric polyamides were analyzed by footprinting and affinity cleavage on a DNA fragment containing two match sites (5'-TGTTA-3' and 5'-ACATT-3') and one 5'-TGTC A-3' mismatch site. Quantitative footprint titrations demonstrate that replacement of γ -aminobutyric acid by (R)- γ -diaminobutyric acid enhances DNA binding affinity for the 5'-TGTTA-3' match site 13-fold ($K_a = 3.8 \times 10^9 M^{-1}$). The enhanced affinity is achieved without a compromise in sequence selectivity, which in fact increases and is found to be 100-fold higher relative to binding at a single base pair mismatch sequence, 5'-TGTC A-3'. An (S)- γ -diaminobutyric acid linked hairpin binds with 170-fold reduced affinity relative to the (R)-enantiomer, and only 5-fold sequence specificity versus a 5'-ACATT-3' reversed orientation site. These effects are modulated by acetylation of the chiral amine substituents. This study identifies structural elements which should facilitate the design of new hairpin-polyamides with improved DNA binding affinity, sequence-specificity, and orientational selectivity.

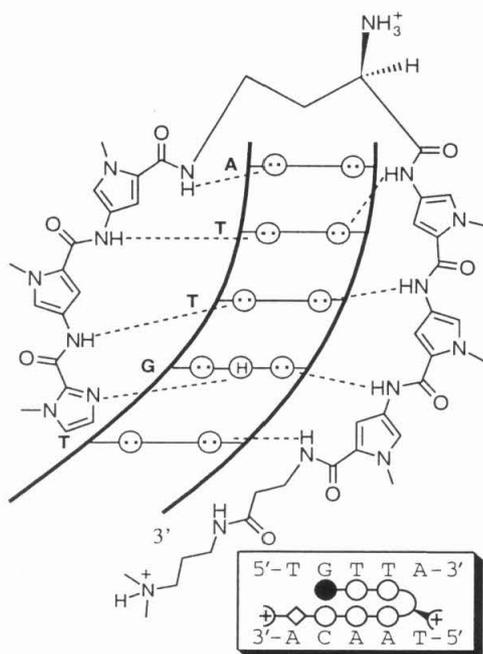
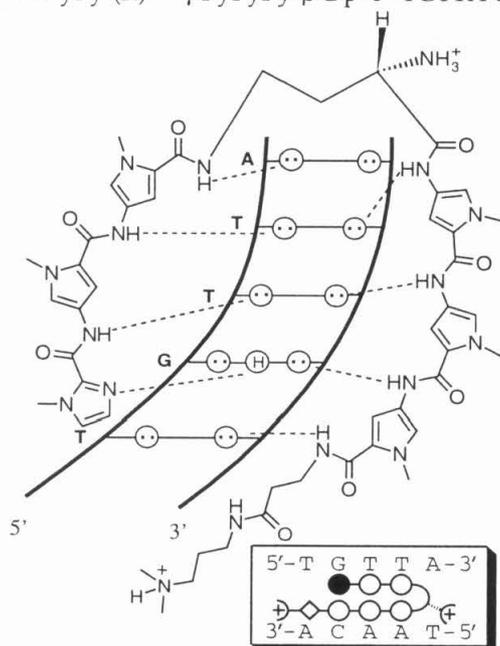
Publication: Herman, Baird & Dervan *J. Am. Chem. Soc.* **1998**, *118*, 6147-6152.

Introduction

Small molecules that target specific predetermined DNA sequences have the potential to control gene expression. 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 oriented N-C with respect to the 5'-3' direction of the DNA helix in the minor groove.² 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} An Im/Im pairing⁴ is disfavored,⁵ breaking a potential degeneracy for recognition.

In parallel with elucidation of the scope and limitations of the polyamide pairing rules, efforts have been made to prevent slipped binding motifs⁶ as well as increase DNA-binding affinity and sequence specificity by covalently linking polyamide subunits.⁶⁻⁹ A hairpin polyamide motif with γ -aminobutyric acid (γ) serving as a turn-specific internal-guide-residue provides a synthetically accessible method for C-N linkage of polyamide subunits (Figure 2.1).⁹ Head-to-tail linked polyamides bind specifically to designated target sites with 100-fold enhanced affinity relative to unlinked subunits.⁹ Eight-ring hairpin polyamides bearing a single positively charged tertiary amine group at the C-terminus have been shown to be cell-permeable and to inhibit the transcription of specific genes in cell culture.¹⁰ The relationship between biological regulation, and the placement, frequency, and nature of charged moieties within the hairpin structure has yet to be determined. This provided impetus to elucidate the effects on hairpin polyamide DNA-binding by selectively placed substituents within the γ -turn.

Figure 2.1 (Top) Hydrogen bonding model of the 1:1 polyamide:DNA complex formed between the hairpin polyamide ImPyPy-(R)^{H₂N}- γ -PyPyPy-b-Dp (**1-R**) with a 5'-TGTTA-3' site. 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. For schematic binding model, imidazole and pyrrole rings are represented as shaded and unshaded spheres respectively, and the β -alanine residue is represented as an unshaded diamond. (Bottom) Binding model for ImPyPy-(S)^{H₂N}- γ -PyPyPy- β -Dp (**1-S**) with a 5'-TGTTA-3' site.

ImPyPy-(R)^{H₂N}γ-PyPyPy-β-Dp•5'-TGTTA-3'ImPyPy-(S)^{H₂N}γ-PyPyPy-β-Dp•5'-TGTTA-3'

Analysis of the NMR structure⁹ⁱ of a hairpin polyamide of sequence composition ImPyPy- γ -PyPyPy complexed with a 5'-TGTTA-3' target site, suggested that substitutions at the γ -position of the γ -aminobutyric acid residue could be

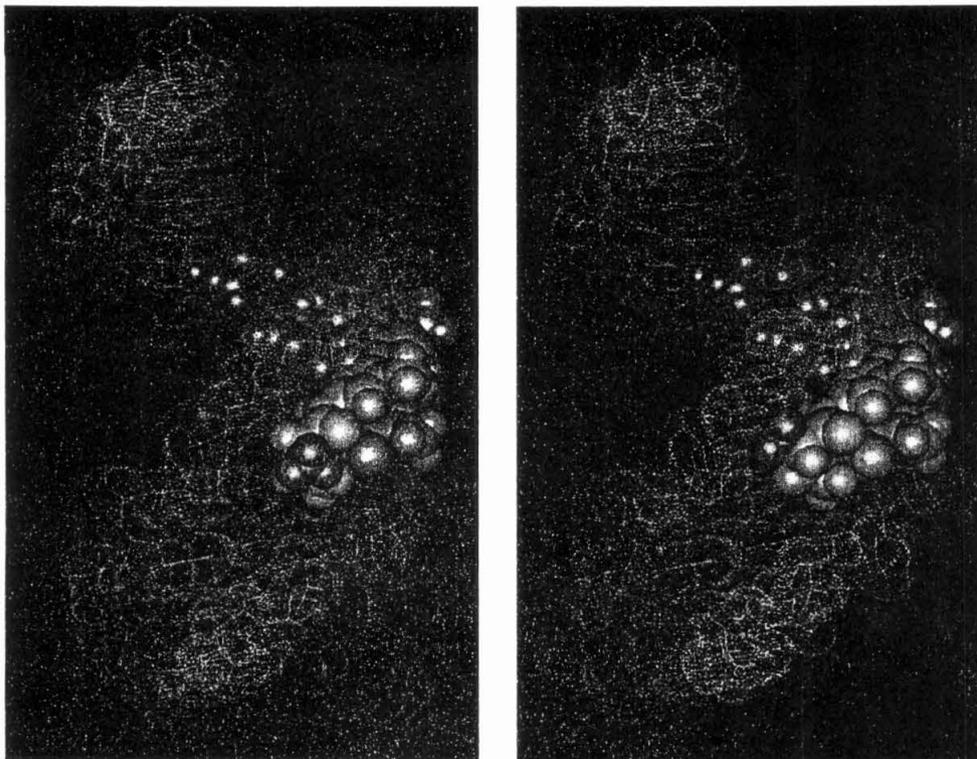


Figure 2.2 Computer generated models of (left) ImPyPy-(*R*)^{H₂N} γ -PyPyPy- β -Dp (gray) and (right) ImPyPy-(*S*)^{H₂N} γ -PyPyPy- β -Dp (gray) bound in the minor groove of double stranded DNA van der Waals surface (yellow). Respective -NH₃⁺ are shown in cyan. Models are derived from the NMR structure coordinates of ImPyPy- γ -PyPyPy -Dp•5'-TGTTA-3'⁹ⁱ using InsightII software.

accommodated within the hairpin-DNA complex. Modeling indicated that replacing the α -H of γ with an amino group that confers an *R*-configuration at the α -carbon could be accommodated within the floor and walls of the minor groove (Figures 2.1 and 2.2a). Formally this is accomplished by substitution of the γ -residue with (*R*)-2,4,-

diaminobutyric acid ($((R)^{H2N} \gamma)$). On the other hand, the (S) -2,4,-diaminobutyric acid ($((S)^{H2N} \gamma)$) linked hairpin is predicted to clash with the walls of the minor groove of the DNA helix (Figure 2.1 and 2.2b). We describe here the synthesis of a new class of chiral hairpin polyamides and their characterization with regard to DNA binding affinity and sequence specificity.

Substitution of the prochiral γ -turn with either enantiomer of 2,4-diaminobutyric acid provides the dicationic six-ring enantiomeric polyamides (+)-ImPyPy- $(R)^{H2N} \gamma$ -PyPyPy- β -Dp (1- R) and (-)-ImPyPy- $(S)^{H2N} \gamma$ -PyPyPy- β -Dp (1- S) which were synthesized by solid phase methods. As a control, the monocationic polyamide (+)-ImPyPy- $(R)^{H2N} \gamma$ -PyPyPy- β -EtOH (2- R) which lacks a charge at the C-terminus was prepared. In order to further study steric effects, the α -acetamido polyamides (+)-ImPyPy- $(R)^{Ac} \gamma$ -PyPyPy- β -Dp (3- R) and (-)-ImPyPy- $(S)^{Ac} \gamma$ -PyPyPy- β -Dp (3- S) were also studied (Figure 3).¹¹ The EDTA analogs ImPyPy- $(R)^{H2N} \gamma$ -PyPyPy- β -Dp-EDTA•Fe(II) (4- R •Fe(II)), ImPyPy- $(S)^{H2N} \gamma$ -PyPyPy- β -Dp-EDTA•Fe(II) (4- S •Fe(II)), ImPyPy- $(R)^{EDTA \bullet Fe(II)} \gamma$ -PyPyPy- β -Dp (5- R •Fe(II)), and ImPyPy- $(S)^{EDTA \bullet Fe(II)} \gamma$ -PyPyPy- β -Dp (5- S •Fe(II)) were constructed to confirm the binding orientation of the modified hairpins at each DNA binding site (Figure 3).

We report here the DNA-binding affinity, orientation, and sequence selectivity of five six-ring hairpin polyamides 1- R , 1- S , 2- R , 3- R , and 3- S for 5-bp 5'-TGTTA-3', 5'-ACATT-3' and 5'-TGTC A-3' sequences. Three separate techniques are used to characterize the DNA-binding properties of the designed polyamides: MPE•Fe(II) footprinting,¹² affinity cleaving,¹³ and DNase I footprinting.¹⁴ Information about binding site size is gained from MPE•Fe(II) footprinting, while affinity cleavage studies determine the binding orientation and stoichiometry of the 1:1 hairpin:DNA complex. Quantitative DNase I footprint titrations allow determination of equilibrium association constants (K_a) of the polyamides for respective match and mismatch binding sites.

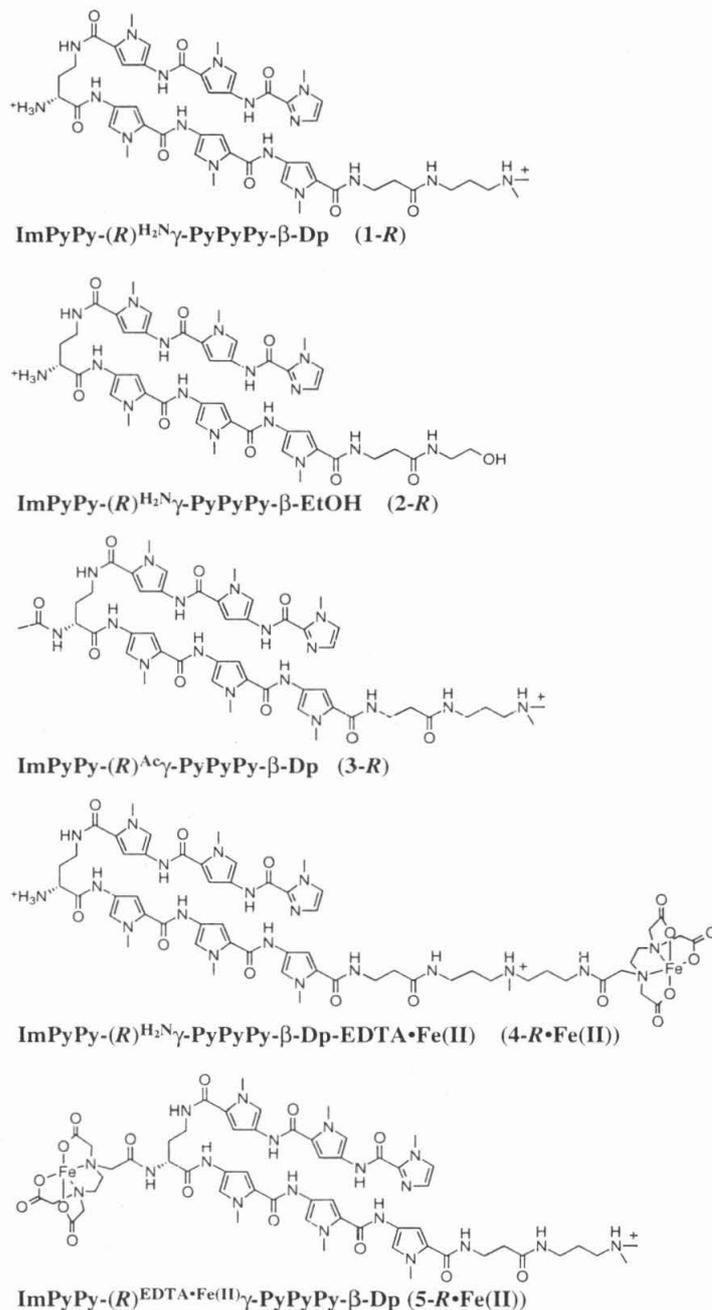
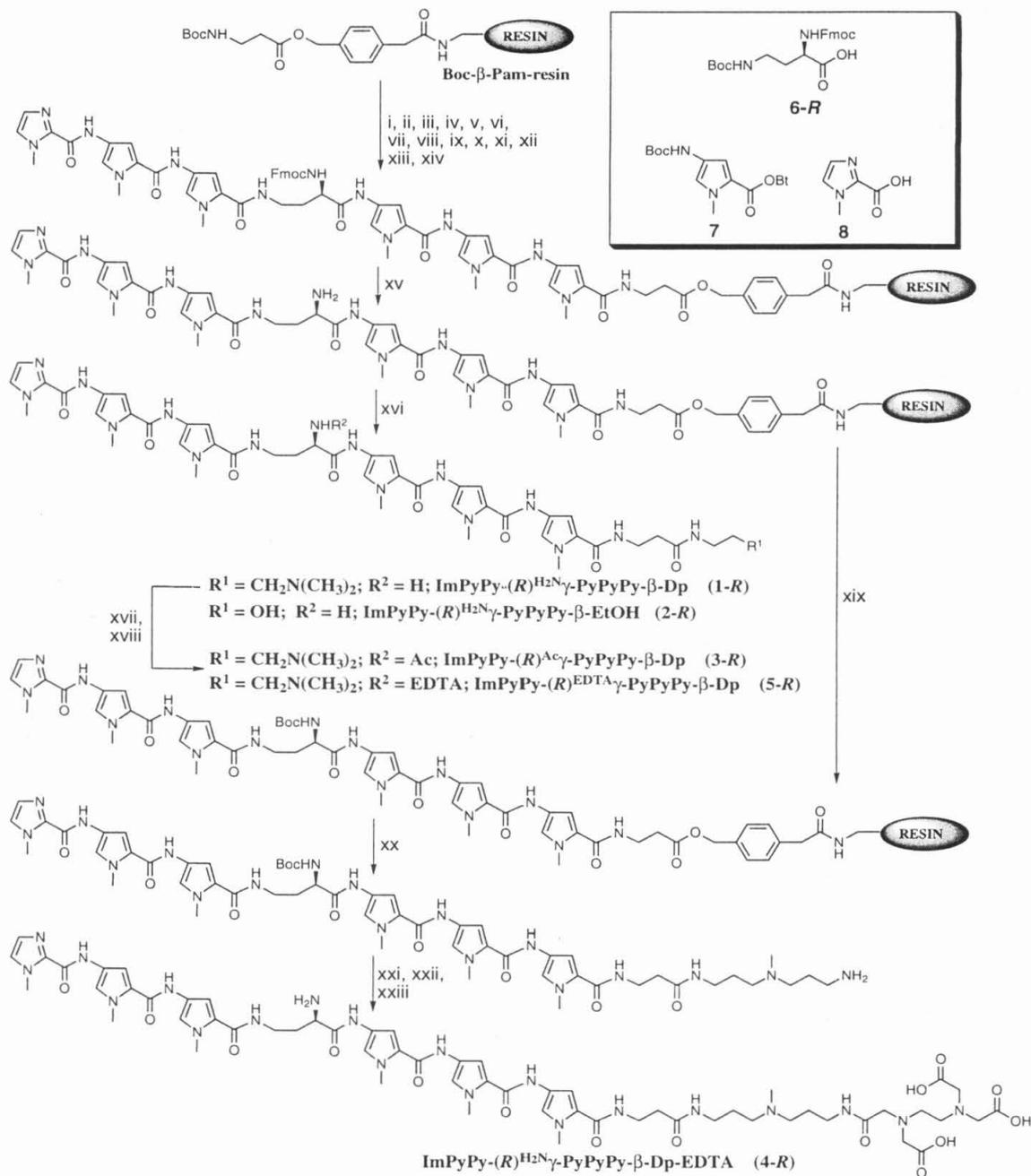


Figure 2.3 Structures of the six-ring hairpin polyamides ImPyPy-(R)^{H₂N}-γ-PyPyPy-β-Dp **1-R**, ImPyPy-(R)^{H₂N}-γ-PyPyPy-β-EtOH **2-R**, ImPyPy-(R)^{Ac}-γ-PyPyPy-β-Dp **3-R**, ImPyPy-(R)^{H₂N}-γ-PyPyPy-β-Dp-EDTA•Fe(II) **4-R•Fe(II)**, and ImPyPy-(R)^{EDTA•Fe(II)}-γ-PyPyPy-β-Dp **5-R•Fe(II)**. Structures of the corresponding (S)^{H₂N}-γ-linked hairpins are

Synthesis. Two polyamide-resins, ImPyPy-(R)^{Fmoc}- γ -PyPyPy- β -Pam-resin and ImPyPy-(S)^{Fmoc}- γ -PyPyPy- β -Pam-resin, were synthesized in 14 steps from Boc- β -alanine-Pam-resin (1 g resin, 0.2 mmol/g substitution) using previously described Boc-chemistry machine-assisted protocols (Figure 4).¹¹ (R)- and (S)-2,4-diaminobutyric acid residues were introduced as orthogonally protected *N*- α -Fmoc-*N*- γ -Boc derivatives (HBTU, DIEA). Fmoc protected polyamide resins ImPyPy-(R)^{Fmoc}- γ -PyPyPy- β -Pam-resin and ImPyPy-(S)^{Fmoc}- γ -PyPyPy- β -Pam-resin were treated with 1:4 DMF:Piperidine (22 °C, 30 min.) to provide ImPyPy-(R)^{H₂N}- γ -PyPyPy- β -Pam-resin and ImPyPy-(S)^{H₂N}- γ -PyPyPy- β -Pam-resin, respectively. A single-step aminolysis of the resin ester linkage was used to cleave the polyamide from the solid support. A sample of resin (240 mg) was treated with either dimethylaminopropylamine (55 °C, 18 h) to provide 1-R, 1-S, 3-R, and 3-S or ethanolamine (55 °C, 18 h) to provide 2-R. Resin cleavage products were purified by reverse phase HPLC to provide ImPyPy-(R)^{H₂N}- γ -PyPyPy- β -Dp (1-R), ImPyPy-(S)^{H₂N}- γ -PyPyPy- β -Dp (1-S), and ImPyPy-(R)^{H₂N}- γ -PyPyPy- β -EtOH (2-R). The stereochemical purity of 1-R was determined to be > 98% by Mosher amide analysis.¹⁵ 1-R,*R* and 1-R,*S* Mosher amides were prepared by reaction of 1-R with HOBt activated esters generated *in situ* from (R)- α -methoxy- γ -(trifluoromethyl)phenylacetic acid and (S)- α -methoxy- α -(trifluoromethyl)phenylacetic acid. For synthesis of analogs modified with EDTA at the carboxy-terminus, the amine-resin was treated with Boc-anhydride (DMF, DIEA, 55 °C, 30 min.) to provide ImPyPy-(R)^{Boc}- γ -PyPyPy- β -Pam-resin and ImPyPy-(S)^{Boc}- γ -PyPyPy- β -Pam-resin (Figure 2.4). A sample of Boc-resin was then cleaved with 3,3'-diamino-*N*-methylpropylamine (55 °C, 18 h) and purified by reversed phase HPLC to provide either ImPyPy-(R)^{Boc}- γ -PyPyPy- β -Dp-NH₂ (1-R-Boc-NH₂) or ImPyPy-(S)^{Boc}- γ -PyPyPy- β -Dp-NH₂ (1-S-Boc-NH₂) which afford free primary amine groups at the C-terminus suitable for post-synthetic modification. The polyamide-amines 1-R-Boc-NH₂ and 1-S-Boc-NH₂ were treated with an excess of EDTA-dianhydride (DMSO/NMP, DIEA, 55

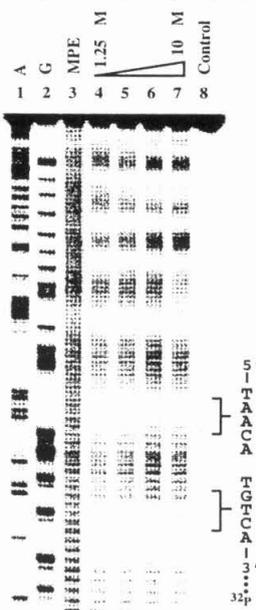
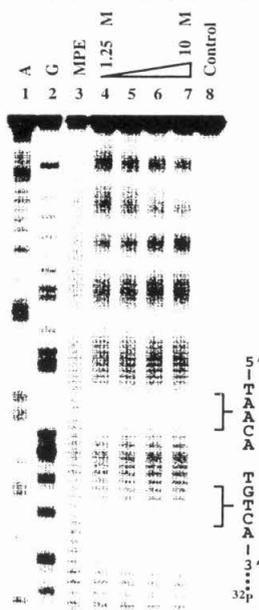
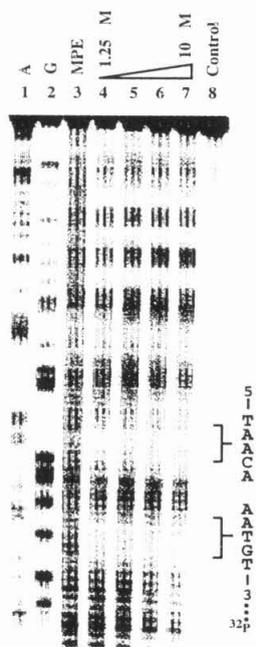
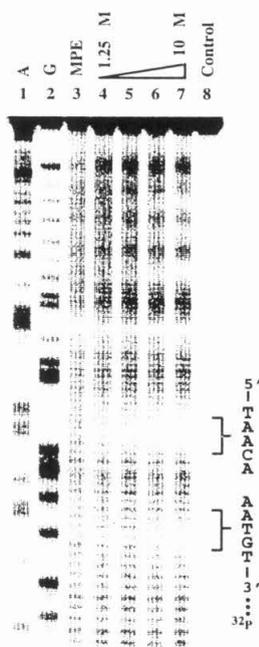
Figure 2.4 Solid phase synthetic scheme exemplified for ImPyPy-(*R*)^{H₂N}- γ -PyPyPy- β -Dp **1-R**, ImPyPy-(*R*)^{H₂N}- γ -PyPyPy- β -EtOH **2-R**, ImPyPy-(*R*)^{Ac}- γ -PyPyPy- β -Dp **3-R**, ImPyPy-(*R*)^{H₂N}- γ -PyPyPy- β -Dp-EDTA **4-R**, and ImPyPy-(*R*)^{EDTA}- γ -PyPyPy- β -Dp **5-R**: (i) 80 % TFA/DCM, 0.4 M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) 80 % TFA/DCM, 0.4 M PhSH; (iv) Boc-Py-OBt, DIEA, DMF; (v) 80 % TFA/DCM, 0.4 M PhSH; (vi) Boc-Py-OBt, DIEA, DMF; (vii) 80 % TFA/DCM, 0.4 M PhSH; (viii) Fmoc- α -Boc- γ -diaminobutyric acid (HBTU, DIEA); (ix) 80 % TFA/DCM, 0.4 M PhSH; (x) Boc-Py-OBt, DIEA, DMF; (xi) 80 % TFA/DCM, 0.4 M PhSH; (xii) Boc-Py-OBt, DIEA, DMF; (xiii) 80 % TFA/DCM, 0.4 M PhSH; (xiv) imidazole-2-carboxylic acid (HBTU/DIEA); (xv) 80 % Piperidine:DMF (25 °C, 30 min.) (xvi) *N,N*-dimethylaminopropylamine (55 °C, 18 h) for **1-R**; ethanolamine (55 °C, 18 h) for **2-R**; (xvii) Ac₂O (for **3-R**), EDTA-dianhydride (for **5-R**), DIEA, DMF (55 °C, 30 min.); (xviii) 0.1 NaOH (55 °C, 10 min.); (xix) Boc₂O, DIEA, DMF; (xx) 3,3'-diamino-*N*-methyldipropylamine (55 °C, 18 h); (xxi) EDTA-dianhydride, DMSO, NMP, DIEA (55 °C, 30 min.); (xxii) 0.1M NaOH, (55 °C, 10 min.) (xxiii) TFA; (Inset) Pyrrole, Imidazole, and diaminobutyric acid monomers for solid phase synthesis: (*R*)-Fmoc- α -Boc- γ -diaminobutyric acid **6-R**, Boc-Pyrrole-OBt ester¹¹ (Boc-Py-OBt) **7**, and Imidazole-2-Carboxylic acid^{2a} (Im-OH) **8**.



°C, 15 min.) and the remaining anhydride hydrolyzed (0.1 M NaOH, 55 °C, 10 min.). The Boc protected EDTA modified polyamides ImPyPy-(R)^{Boc}γ-PyPyPy-β-Dp-EDTA (4-R-Boc) and ImPyPy-(S)^{Boc}γ-PyPyPy-β-Dp-EDTA (4-S-Boc) were isolated by HPLC. Individual Boc-EDTA-polyamides were deprotected with neat TFA (22 °C, 1 h) to provide the respective C-terminal EDTA derivatives, ImPyPy-(R)^{H₂N}γ-PyPyPy-β-Dp-EDTA (4-R) and ImPyPy-(S)^{H₂N}γ-PyPyPy-β-Dp-EDTA (4-S). For the synthesis of acetamide-turn or EDTA-turn derivatives, a sample of the α-amino polyamide ImPyPy-(R)^{H₂N}γ-PyPyPy-β-Dp (1-R) or ImPyPy-(S)^{H₂N}γ-PyPyPy-β-Dp (1-S) was treated with an excess of either acetic anhydride or EDTA-dianhydride (DMSO/NMP, DIEA 55 °C, 30 min.) and the remaining anhydride hydrolyzed (0.1 M NaOH, 55 °C, 10 min.). The polyamides ImPyPy-(R)^{Ac}γ-PyPyPy-β-Dp (3-R), ImPyPy-(S)^{Ac}γ-PyPyPy-β-Dp (3-S), ImPyPy-(R)^{EDTA}γ-PyPyPy-β-Dp (5-R) and ImPyPy-(S)^{EDTA}γ-PyPyPy-β-Dp (5-S) were then isolated by reverse phase HPLC. The six-ring hairpin polyamides described here are soluble in aqueous solution at concentrations ≤ 10 mM.

Binding Site Size and Location by MPE•Fe(II) Footprinting. MPE•Fe(II) footprinting¹² on 3'- and 5'-³²P end-labeled 135 base pair restriction fragments reveals that the polyamides, each at 1 μM concentration, bind to the 5'-TGTTA-3' match site (25 mM Tris-acetate, 10 mM NaCl, 100 μM/base pair calf thymus DNA, pH 7.0 and 22 °C) (Figure 5 and 6). Compounds 1-R and 3-R, each at 1.25 μM, protect both the cognate 5'-TGTTA-3' site and the single base pair mismatch sequence 5'-TGTC A-3'. Remarkably, binding sequence preferences vary for the polyamides depending on the stereochemistry of the amine substituent. At 1.25 μM and 2.5 μM concentration respectively, polyamides 1-S and 3-S bind a 5'-ACATT-3' reverse orientation match site in addition to the target match site 5'-TGTTA-3'. The sizes of the asymmetrically 3'-shifted footprint cleavage protection patterns for the polyamides are consistent with 5 base pair binding sites.

Figure 2.5 MPE•Fe(II) footprinting experiments on a 3'-³²P-labeled 135 bp restriction fragment. 5'-TGTTA-3', 5'-TGTCA-3, and 5'-ACATT-3' sites are shown on the right side of the autoradiogram. Lane 1, A reaction; lane 2, G reaction; lane 3 MPE•Fe(II) standard; lane 8, intact DNA; lanes 4-7, 1.25 μ M, 2.5 μ M, 5 μ M, and 10 μ M polyamide. (a) ImPyPy-(R)^{H₂N} γ -PyPyPy- β -Dp **1-R**; (b) ImPyPy-(R)^{Ac} γ -PyPyPy- β -Dp **3-R**; (c) ImPyPy-(S)^{H₂N} γ -PyPyPy- β -Dp **3-S**; (d) ImPyPy-(S)^{Ac} γ -PyPyPy- β -Dp **3-S**. All lanes contain 15 kcpm 3'-radiolabeled DNA, 25 mM Tris-acetate buffer (pH 7.0), 10 mM NaCl, and 100 μ M/base pair calf thymus DNA.

ImPyPy-(R)^{H₂N}γ-PyPyPy-β-DpImPyPy-(R)^{Ac}γ-PyPyPy-β-DpImPyPy-(S)^{H₂N}γ-PyPyPy-β-DpImPyPy-(S)^{Ac}γ-PyPyPy-β-Dp

Binding Orientation by Affinity Cleaving. Affinity cleavage experiments¹³ using hairpin polyamides modified with EDTA•Fe(II) at either the C-terminus, or on the γ -turn, were used to determine polyamide binding orientation and stoichiometry. Affinity cleavage experiments were performed on the same 3'- and 5'-³²P end-labeled 135 base pair restriction fragment (25 mM Tris-acetate, 10 mM NaCl, 100 μ M/base pair calf thymus DNA, pH 7.0 and 22 °C). The observed cleavage patterns for ImPyPy-(R)^{H2N} γ -PyPyPy- β -Dp-EDTA•Fe(II) (4-R•Fe(II)), ImPyPy-(R)^{EDTA•Fe(II)} γ -PyPyPy- β -Dp (5-R•Fe(II)), ImPyPy-(S)^{H2N} γ -PyPyPy- β -Dp-EDTA•Fe(II) (4-S•Fe(II)), ImPyPy-(S)^{EDTA•Fe(II)} γ -PyPyPy- β -Dp (5-S•Fe(II)) (Figures 2.7, 2.8 and 2.9) are in all cases 3'-shifted, consistent with minor groove occupancy. In the presence of 3.3 μ M of 4-R•Fe(II) and 10 μ M 4-S•Fe(II) which have an EDTA•Fe(II) moiety at the C-terminus, a single cleavage locus proximal to the 5' side of the 5'-TGTTA-3' match sequence is revealed. In the presence of 3.3 μ M 5-R•Fe(II) and 10 μ M 5-S•Fe(II) which have an EDTA•Fe(II) moiety appended to the γ -turn, a single cleavage locus is revealed proximal to the 3' side of the 5'-TGTTA-3' match sequence. Cleavage loci are more concise for the g-turn EDTA•Fe(II) placement relative to carboxy terminal placement, consistent with the shorter tether. Cleavage loci are observed at both the 5' and 3' side of the 5'-TGTCA-3' single base pair mismatch site in the presence of 10 μ M of 4-R•Fe(II). The cleavage patterns observed at the 3' side of the site is approximately 3-fold more intense than cleavage at the 5'-side. For polyamide 4-S•Fe(II) at 10 μ M concentration, a single cleavage locus is revealed proximal to the 5' side of the 5'-ACATT-3' reverse orientation match site.

Figure 2.7 Affinity cleaving experiments on the 3'-³²P-labeled 135 bp restriction fragment. 5'-TGTTA-3', 5'-TGTC A-3, and 5'-ACATT-3' sites are shown on the right side of the autoradiograms as appropriate. Lane 1, A reaction; lane 2, G reaction; lane 8, intact DNA; lanes 3-7, 110 nM, 330 nM, 1.0 μM, and 3.3 μM, and 10 μM polyamide. (a) ImPyPy-(R)^{H2N}γ-PyPyPy-β-Dp-EDTA•Fe(II) (**4-R•Fe(II)**); (b) ImPyPy-(R)^{EDTA•Fe(II)}γ-PyPyPy-β-Dp (**5-R•Fe(II)**); (c) ImPyPy-(S)^{H2N}γ-PyPyPy-β-Dp-EDTA•Fe(II) (**4-S•Fe(II)**); (d) ImPyPy-(S)^{EDTAFe(II)}γ-PyPyPy-β-Dp (**5-S•Fe(II)**). 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.

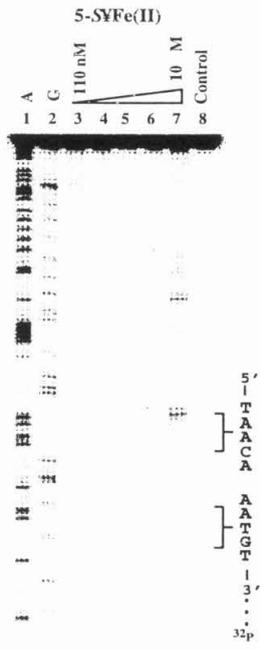
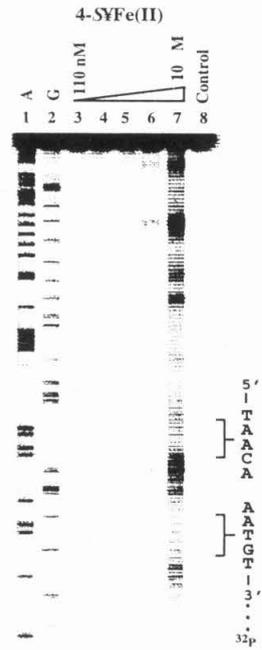
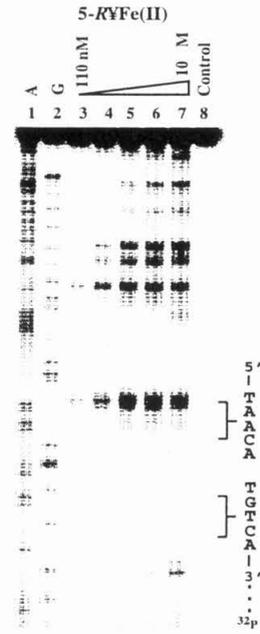
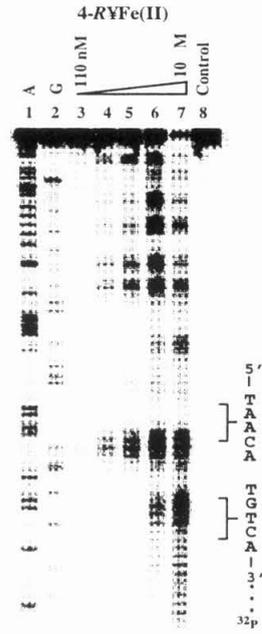
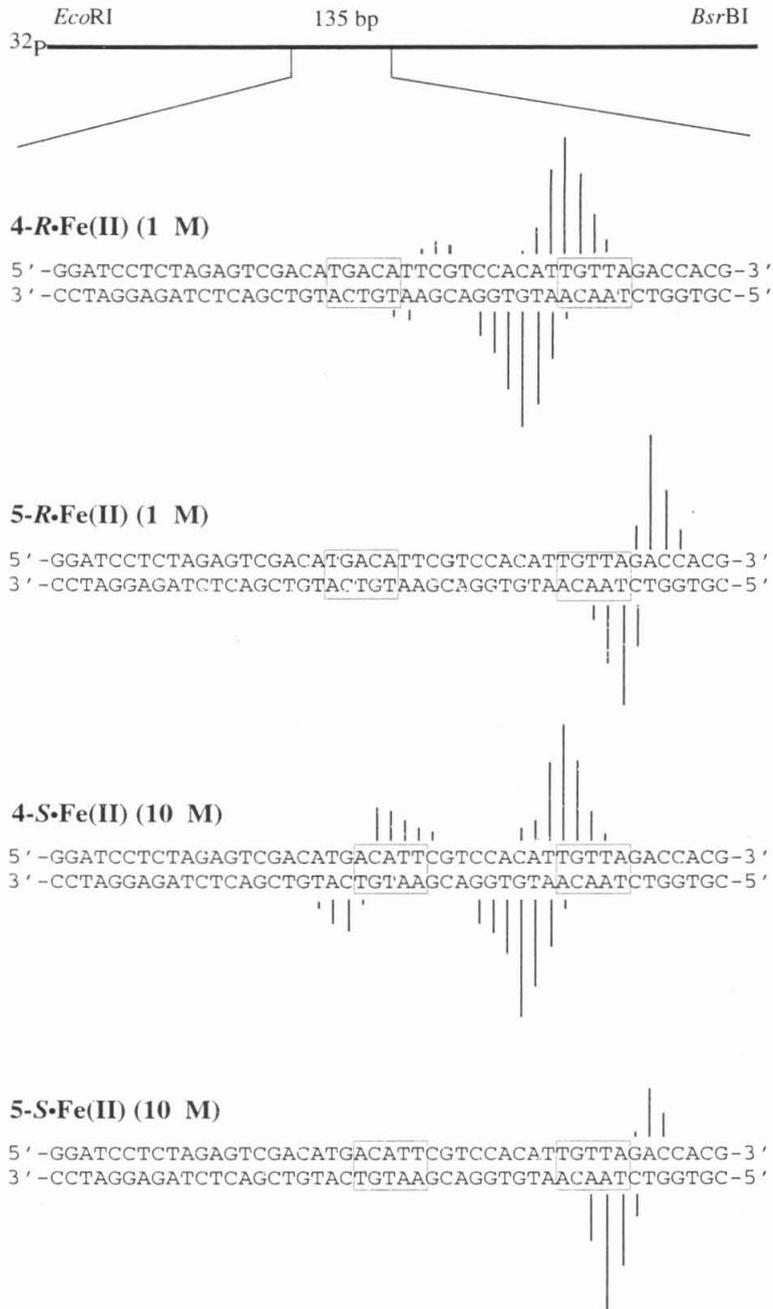
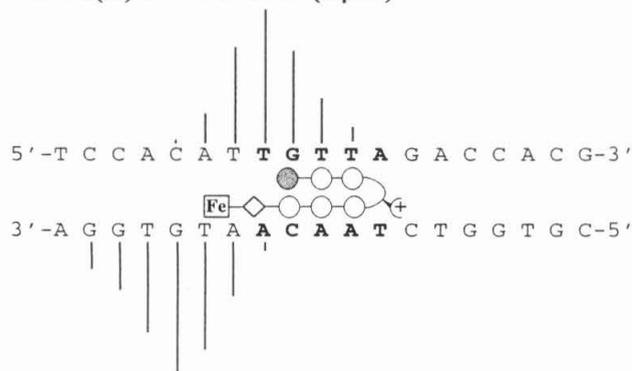


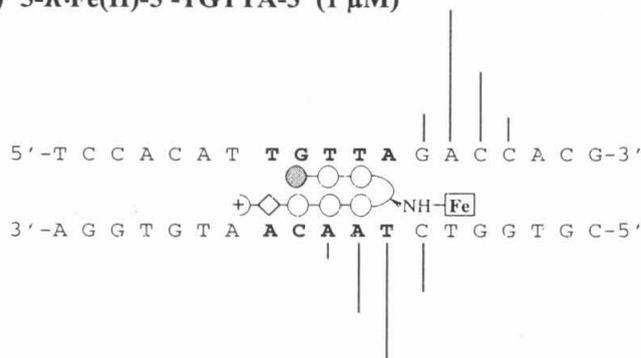
Figure 2.8. (Top) Illustration of the 135 bp restriction fragment with the position of the sequence indicated. Line heights are proportional to the relative cleavage at each band. Binding sites determined by MPE•Fe(II) footprinting and quantitated by DNase I footprint titrations are boxed. (Bottom) Affinity cleavage patterns for ImPyPy-(R)^{H₂N}γ-PyPyPy-β-Dp-EDTA•Fe(II) (**4-R•Fe(II)**) and ImPyPy-(R)^{EDTA•Fe(II)}γ-PyPyPy-β-Dp (**5-R•Fe(II)**) at 1 μM concentration; ImPyPy-(S)^{H₂N}γ-PyPyPy-β-Dp-EDTA•Fe(II) (**4-S•Fe(II)**) and ImPyPy-(S)^{EDTA•Fe(II)}γ-PyPyPy-β-Dp (**5-S•Fe(II)**) at 10 μM concentration.



a) 4-R-Fe(II)-5'-TGTTA-3' (1 μ M)



b) 5-R-Fe(II)-5'-TGTTA-3' (1 μ M)



c) 4-S-Fe(II)-5'-ACATT-3' (10 μ M)

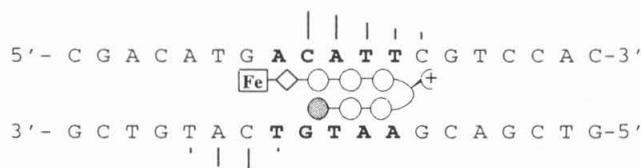


Figure 2.9 Affinity cleavage patterns and ball and stick models of the six-ring EDTA•Fe(II) analogs. Bar heights are proportional to the relative cleavage intensities at each base pair. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent β -alanine. The boxed Fe denotes the EDTA•Fe(II) cleavage moiety. (a) ImPyPy-(R)^{H₂N} γ -PyPyPy- β -Dp-EDTA•Fe(II) 4-R•Fe(II)•5'-TGTTA-3' at 1 μ M concentration; (b) 1 μ M ImPyPy-(R)^{EDTA•Fe(II)} γ -PyPyPy- β -Dp 5-R•Fe(II)•5'-TGTTA-3'; and (c) 10 μ M ImPyPy-(S)^{H₂N} γ -PyPyPy- β -Dp-EDTA•Fe(II) 4-S•Fe(II)•5'-ACATT-3'.

Energetics by Quantitative DNase I Footprinting. Quantitative DNase I footprint titrations¹⁴ (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂, pH 7.0 and 22 °C) were performed to determine the equilibrium association constant (K_a) of each six-ring hairpin polyamide for the three resolved sites (Figures 2.10 and 2.11). The 5'-TGTTA-3' site is bound by polyamides in the order: ImPyPy-(R)^{H2N}γ-PyPyPy-β-Dp (1-R) ($K_a = 3.8 \times 10^9 \text{ M}^{-1}$) \approx ImPyPy-(R)^{H2N}γ-PyPyPy-β-EtOH (2-R) ($K_a = 3.3 \times 10^9 \text{ M}^{-1}$) > ImPyPy-(R)^{Ac}γ-PyPyPy-β-Dp (3-R) ($K_a = 3.0 \times 10^8 \text{ M}^{-1}$) \approx ImPyPy-γ-PyPyPy-β-Dp ($K_a = 2.9 \times 10^8 \text{ M}^{-1}$) > ImPyPy-(S)^{H2N}γ-PyPyPy-β-Dp (1-S) ($K_a = 2.2 \times 10^7 \text{ M}^{-1}$) > ImPyPy-(S)^{Ac}γ-PyPyPy-β-Dp (3-S) ($K_a < 5.0 \times 10^6 \text{ M}^{-1}$). Equilibrium association constants for recognition of the 5'-TGACT-3' single base pair mismatch site are: ImPyPy-(R)^{H2N}γ-PyPyPy-β-Dp (1-R) ($K_a = 3.5 \times 10^7 \text{ M}^{-1}$) \approx ImPyPy-(R)^{H2N}γ-PyPyPy-β-EtOH (2-R) ($K_a = 3.1 \times 10^7 \text{ M}^{-1}$) > ImPyPy-(R)^{Ac}γ-PyPyPy-β-Dp (3-R) ($K_a < 5 \times 10^6 \text{ M}^{-1}$) \approx ImPyPy-γ-PyPyPy-β-Dp ($K_a = 4.8 \times 10^6 \text{ M}^{-1}$). The polyamides ImPyPy-(S)^{H2N}γ-PyPyPy-β-Dp (1-S) and ImPyPy-(S)^{Ac}γ-PyPyPy-β-Dp (3-S) recognize the 5'-ACATT-3' reverse orientation sequence with $K_a = 4.6 \times 10^6 \text{ M}^{-1}$ and $K_a < 5 \times 10^6 \text{ M}^{-1}$ respectively. It should be noted that a detailed comparison of the relative mismatch binding energetics cannot be made since the 5'-TGACA-3' and 5'-ACATT-3' binding sites overlap. The relative affinity of 5'-TGTTA-3' match site binding varies from 100-fold to 5-fold depending on the stereochemistry of the γ-turn substitutions (Table 2.1).

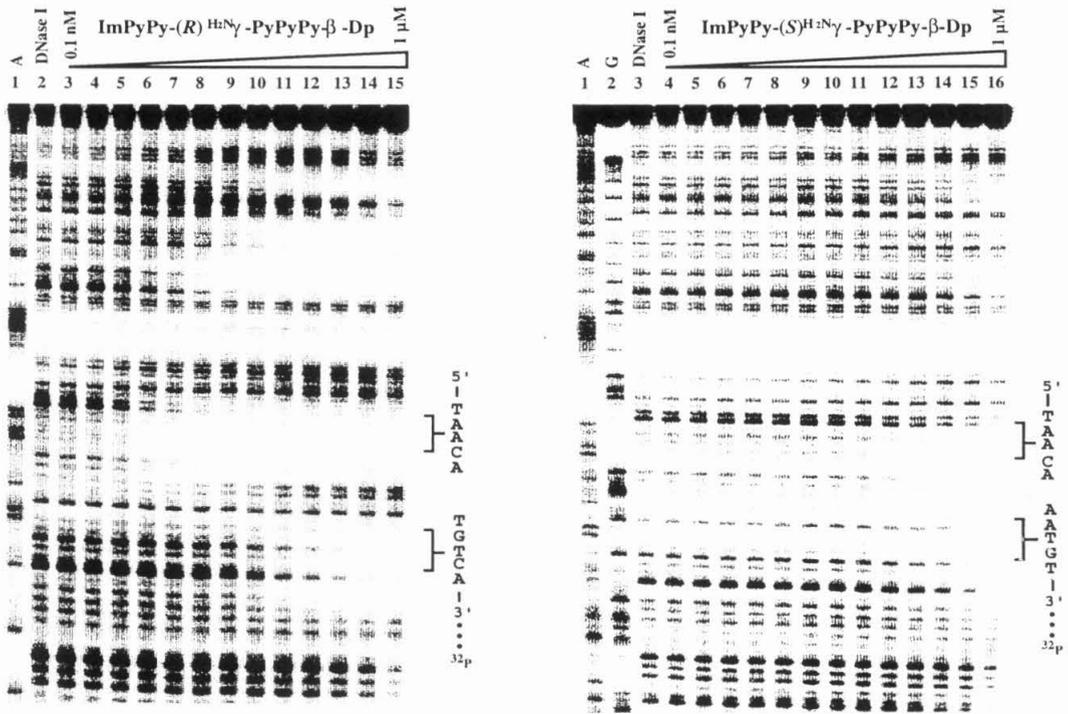


Figure 2.10 Quantitative DNase I footprint titration experiment with (left) ImPyPy-(R)^{H2N}γ-PyPyPy-β-Dp (1-R) and (right) ImPyPy-(S)^{H2N}γ-PyPyPy-β-Dp (1-S) on the 3'-end labeled 135-bp restriction fragment: lane 1, A reaction; lane 2, DNase I standard; lane 3-15, 0.1 nM, 0.2 nM, 0.5 nM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1.0 μM polyamide. The 5'-TGTTA-3' and 5'-TGACA-3' sites for 1-R and the 5'-TGTTA-3' and 5'-ACATT-3' sites for 1-S 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 MgCl₂ and 5 mM CaCl₂.

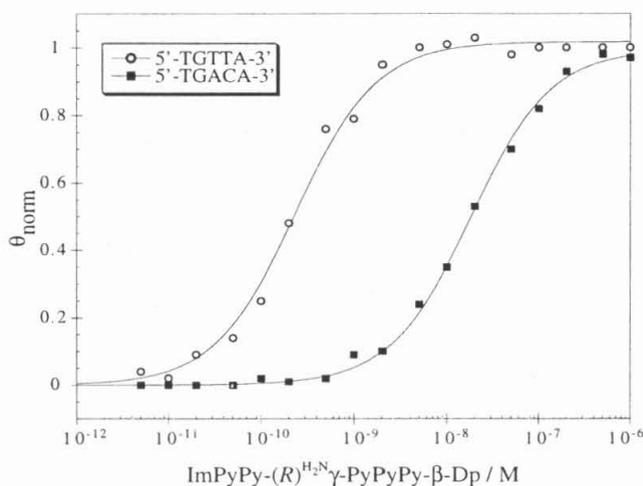


Figure 2.11 Data from quantitative DNase I footprint titration experiments for $\text{ImPyPy-(R)}^{\text{H}_2\text{N}}\gamma\text{-PyPyPy}\beta\text{-Dp}$ **1-R** binding to the 5'-TGTTA-3' and 5'-TGACA-3' sites. θ_{norm} points were obtained using storage phosphor autoradiography and processed as described in the experimental section. The data for the binding of **1-R** to the 5'-TGTTA-3' match site is indicated by open circles (○), and binding to the 5'-TGACA-3' mismatch site by closed squares (■). The solid curves are the best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm using eq. 2, $n=1$.

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

Polyamide	Match Site 5'-TGTTA-3'	Reverse Site 5'-ACATT-3'	Mismatch Site 5'-TGACA-3'	Specificity
	2.9×10^8	n/d	4.8×10^6	60
	3.8×10^9 (0.2)	n/d	3.5×10^7 (1.0)	100
	2.2×10^7 (0.8)	4.6×10^6 (2.0)	n/d	5
	3.3×10^9 (0.9)	n/d	3.1×10^7	100
	3.0×10^8 (1.3)	n/d	$< 5.0 \times 10^6$	<60
	$< 5.0 \times 10^6$	$< 5.0 \times 10^6$		n/d

^aThe reported equilibrium association constants are the mean values obtained from three DNase I footprint titration experiments. The standard deviation for each data set is indicated in parentheses. The assays were carried out at 22 °C, pH 7.0 in the presence of 10 mM Tris HCl, 10 mM •KCl, 10 mM MgCl₂, and 5 mM CaCl₂. ^bThe five base-pair binding sites are in capital letters. ^cSpecificity is calculated by $K_a(\text{match site})/K_a(\text{mismatch site})$. ^dMismatch site is 5'-ACATT-3' for ImPyPy-(S)^{H2N}γ-PyPyPy-β-Dp (1S) and ImPyPy-(S)^{Ac}γ-PyPyPy-β-Dp (3S) as determined by MPE•Fe(II) footprinting and affinity cleaving.

Discussion

Binding Site Size and Orientation. MPE•Fe(II) footprinting reveals that the polyamides bind with highest affinity to the 5'-TGTTA-3' match site, the 5'-TGACA-3' single base pair mismatch site for polyamides 1-R and 3-R, and the 5'-ACATT-3' reverse orientation match site for polyamides 1-S and 3-S (Figure 2.6). Affinity cleaving experiments using polyamides with EDTA•Fe(II) placed at either the carboxy terminus or the γ-turn confirm that polyamides derived from both (R) and (S)-2,4-diaminobutyric acid bind to the 5'-TGTTA-3' target site with a single orientation (Figure 2.10). The observation of a single cleavage locus is consistent only with an oriented 1:1

polyamide:DNA complex in the minor groove and rules out any dimeric overlapped or extended binding motifs. The hairpin binding model is further supported by the location of the cleavage locus at either the 5' or 3' side of the 5'-TGTTA-3' target site corresponding to EDTA•Fe(II) placement at the polyamide carboxy terminus or the γ -turn, respectively (Figure 2.10). Polyamide subunits linked by the (*R*)^{H2N} γ bind the symmetric single base pair mismatch sequence 5'-TGACA-3' in two distinct orientations. Polyamides linked with (*S*)^{H2N} γ bind to a 5'-ACATT-3' reverse orientation match sequence as revealed by a unique cleavage locus at the 5' side of the site.

Binding Affinity. All six polyamides bind to the 5'-TGTTA-3' target site with binding isotherms (eq. 2, $n = 1$) consistent with binding as an intramolecular hairpin (Figure 11). However, the relative match site binding affinity varies by nearly 1000-fold, depending on the stereochemistry of the γ -turn and the nature of the substituents. Among the six polyamides, ImPyPy-(*R*)^{H2N} γ -PyPyPy- β -Dp (1-*R*) binds to the targeted 5'-TGTTA-3' site with the highest affinity. ImPyPy-(*R*)^{H2N} γ -PyPyPy- β -Dp binds with an equilibrium association constant, ($K_a = 3 \times 10^9 \text{ M}^{-1}$)^{9b}, a factor of 10 greater than that of the parent polyamide, ImPyPy- γ -PyPyPy- β -Dp, ($K_a = 3 \times 10^8 \text{ M}^{-1}$). Replacement of the C-terminal dimethylaminopropylamide group of 1-*R* with an ethoxyamide group as in ImPyPy- γ -PyPyPy- β -EtOH (2-*R*) results in no decrease in binding affinity ($K_a = 3 \times 10^9 \text{ M}^{-1}$). Acetylation of the γ -turn amino group as in ImPyPy-(*R*)^{Ac} γ -PyPyPy- β -Dp (3-*R*) reduces binding affinity 10-fold ($K_a = 3 \times 10^8 \text{ M}^{-1}$) relative to 1-*R*.

The observation that polyamides which differ only by replacement of the dimethylaminopropylamide group 1-*R* with an ethoxyamide group 2-*R* bind with similar affinity indicates that interactions between the cationic dimethylaminopropyl tail group with anionic phosphate residues or the negative electrostatic potential in the floor of the minor groove¹⁶ may not contribute substantially to the energetics of hairpin-DNA binding. Furthermore, these results indicate that the observed binding enhancement of 1-*R*, in relation to ImPyPy- γ -PyPyPy- β -Dp, is not simply the difference between a

monocationic and dicationic ligand binding to the polycationic DNA helix.¹⁶ The modest increased binding affinity of polyamide 1-*R* may result from electrostatic interactions between the precisely placed amine group and the floor of the minor groove. Alternately, the increased affinity could indicate a reduction in the degrees of freedom accessible to the free hairpin in solution resulting from a steric effect, or an electrostatic interaction between the positively charged amine group and the negative potential of the α -carbonyl group.

Polyamides linked with $(S)^{H2N}\gamma$, ImPyPy- $(S)^{H2N}\gamma$ -PyPyPy- β -Dp (1-*S*) and ImPyPy- $(S)^{Ac}\gamma$ -PyPyPy- β -Dp (3-*S*) bind to the 5'-TGTTA-3' match site with 100-fold ($K_a = 2 \times 10^7 M^{-1}$) and 1000-fold ($K_a < 5 \times 10^6 M^{-1}$) reduced affinity relative to the $(R)^{H2N}\gamma$ linked polyamide 1-*R*. These results demonstrate that the DNA-binding affinity of chiral hairpin polyamides can be predictably regulated as a function of the stereochemistry of the turn residue.

Sequence-specificity. Polyamides with a variety of substitutions at the γ -turn bind preferentially to the 5'-TGTTA-3' match site, while overall specificity versus binding at reverse orientation and mismatch sites is modified. Replacing the α -proton in the γ residue of ImPyPy- γ -PyPyPy- β -Dp with an amino group that confers a chiral α -hydrogen (*R*) configuration provides the most specific polyamide ImPyPy- $(R)^{H2N}\gamma$ -PyPyPy- β -Dp (1-*R*). The ImPyPy- $(R)^{H2N}\gamma$ -PyPyPy- β -Dp•5'-TGTTA-3' complex forms with 100-fold preference relative to the ImPyPy- $(R)^{H2N}\gamma$ -PyPyPy- β -Dp•5'-TGTCA-3' mismatch complex. Substitution of the charged dimethylaminopropyl tail group with an ethoxyamide group as in (2-*R*) does not alter binding specificity. The modest increase in specificity against single base mismatch sequences for polyamides 1-*R* and 2-*R* (100-fold) relative to the parent unsubstituted hairpin polyamide (60-fold) implicates chiral hairpin polyamides as an optimized class of small molecules for recognition of the DNA minor groove.

Binding Orientation. In principle, a polyamide:DNA complex can form at two different DNA sequences depending on the alignment of the polyamide (N-C) with the walls of the minor groove (5'-3').^{2f} A six-ring hairpin polyamide of core sequence composition ImPyPy- γ -PyPyPy which places the N-terminus of each three-ring polyamide subunit at the 5'-side of individual recognized DNA strands would bind to 'forward match' 5'-WGWW-3' sequences (W = A or T). Placement of the polyamide N-terminus at the 3'-side of each recognized strand would result in targeting 'reverse match' 5'-WCWW-3' sequences. For hairpin polyamides there is an energetic preference for 'forward' alignment of each polyamide subunit (N-C) with respect to the backbone (5'-3') of the DNA double helix.^{2f}

In addition to decreasing the affinity for the 5'-TGTTA-3' match site, replacing the α -proton of γ -turn in ImPyPy- γ -PyPyPy- β -Dp with (*S*)^{H₂N} γ changes the mismatch sequence preference from the 5'-TGACA-3' site bound by the (*R*)^{H₂N} γ -linked polyamides to a 5'-ACATT-3' reverse match site. Binding to the reverse site may result from the presence of the steric bulk of the amino or acetamido groups in the floor of the minor groove preventing the deep polyamide binding required for specific DNA recognition. However, an analysis of hairpin folding requirements for 'forward' and 'reverse' binding reveals an additional model.

In principle, there exist two non-superimposable hairpin folds which are related by mirror plane symmetry (Figure 2.12). One hairpin fold is responsible for the preferred 5' to 3' N to C orientation, while the other fold corresponds to the 3' to 5' N to C reverse orientation binding. For an achiral hairpin polyamide in the absence of DNA, each non-superimposable fold should be energetically equivalent. However, an asymmetrically folded hairpin polyamide with a chiral substituent could potentially display differential energetics for oriented binding (Figure 2.12). In the forward folded hairpin (5' to 3' N to C), (*R*)^{H₂N} γ directs the amine functionality away from the DNA helix, while (*S*) enantiomer is predicted to direct the amine into the floor of the minor groove. For the

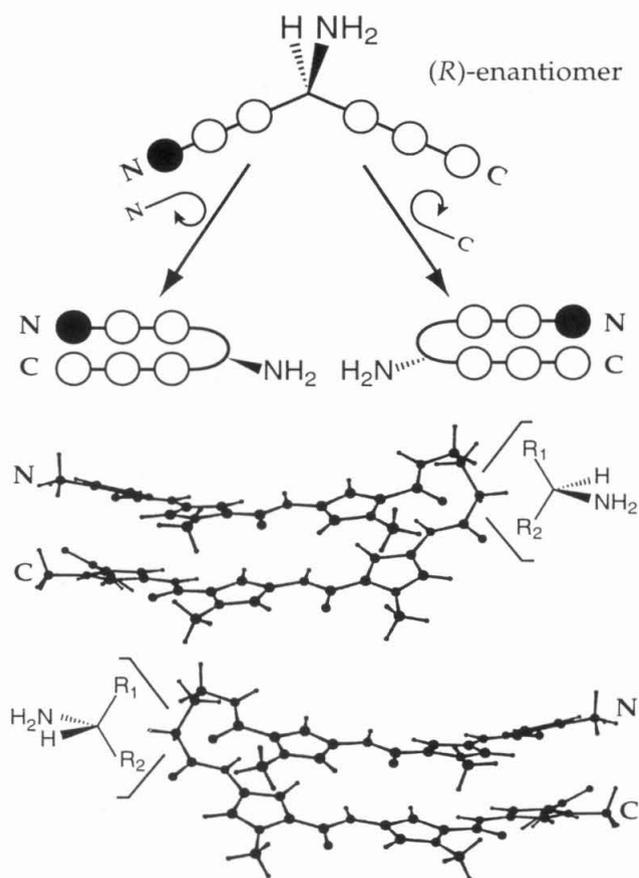


Figure 2.12 Model for chiral hairpin folding: filled and unfilled circles represent Py and Im residues respectively, α -amino, and α -H are highlighted and shown in the $(R)^{H_2N}$. Folding pathways leading to hairpin structures suitable for (left) polyamide (N-C) recognition of DNA 'forward orientation' (5'-3') and (right) polyamide (C-N) recognition of DNA 'reverse orientation' (5'-3'). The corresponding stereochemistry of the α -position of the γ -turn is highlighted for each fold. 'Forward' hairpin structure model for ImPyPy- γ -PyPyPy was generated from NMR structure coordinates⁹ⁱ using Chem3D software. 'Reverse' hairpin structure (bottom) was generated by performing a mirror transformation of the 'forward' hairpin.

'reverse' fold hairpin, (*S*)^{H₂N}γ directs the amine functionality away from the floor of the DNA helix, while the amine of the (*R*) enantiomer is predicted to clash with the floor of the helix. The modest enhanced specificity of chiral polyamides 1-*R* and 2-*R* relative to the unsubstituted parent hairpin may result from stabilization of the forward binding mode and/or destabilization of the reverse binding hairpin fold.

Implications for the Design of Minor Groove Binding Molecules. The results presented here reveal properties of chiral structure elements that will guide future polyamide design: (i) Amine substituents on the (*R*)^{H₂N}γ turn amino acid enhance DNA-binding affinity and specificity relative to the unsubstituted parent hairpin, providing for an optimized class of hairpin polyamides. (ii) Acetamido substituents at the (*R*)^{H₂N}γ turn do not compromise affinity or specificity relative to the parent hairpin, providing for a convenient synthetic attachment point at the 'capped' end of the molecule. (iii) (*S*)^{H₂N}γ linked hairpins bind with enhanced affinity to reverse orientation sights relative to the parent hairpin, while (*R*)^{H₂N}γ linked hairpin binds with enhanced specificity relative to the parent hairpin. These observations indicate that γ-turn substituents may regulate hairpin polyamide binding orientational preference. The results presented here set the stage for preparation of a variety of new chiral hairpin polyamide structures for specific recognition in the DNA minor groove.

Experimental Section

Dicyclohexylcarbodiimide (DCC), Hydroxybenzotriazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU) and 0.2 mmol/gram Boc-β-alanine-(4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-β-Pam-Resin) was purchased from Peptides International (0.2 mmol/gram) (*R*)-2-Fmoc-4-Boc-diaminobutyric acid, (*S*)-2-Fmoc-4-Boc-diaminobutyric acid, and (*R*)-2-amino-4-Boc-diaminobutyric acid were from Bachem.

N,N-diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF), *N*-methylpyrrolidone (NMP), DMSO/NMP, Acetic anhydride (Ac₂O), and 0.0002 M potassium cyanide/pyridine were purchased from Applied Biosystems. Dichloromethane (DCM) and triethylamine (TEA) were reagent grade from EM, thiophenol (PhSH), dimethylaminopropylamine (Dp), (*R*)- α -methoxy- α -(trifluoromethyl)phenylacetic acid ((*R*)MPTA) and (*S*)- α -methoxy- α -(trifluoromethyl)phenylacetic acid ((*S*)MPTA) were from Aldrich, trifluoroacetic acid (TFA) Biograde from Halocarbon, phenol from Fisher, and ninhydrin from Pierce. All reagents were used without further purification.

Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. A shaker for manual solid phase synthesis was obtained from St. John Associates, Inc. Screw-cap glass peptide synthesis reaction vessels (5 mL and 20 mL) with a #2 sintered glass frit were made as described by Kent.¹⁷ ¹H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Optical rotations were recorded on a JASCO Dip 1000 Digital Polarimeter. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on either a HP 1090M analytical HPLC or a Beckman Gold system using a RAINEN C₁₈, Microsorb MV, 5 μ m, 300 x 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, gradient elution 1.25% acetonitrile/min. Preparatory reverse phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100 μ m C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. 18M Ω water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μ m filtered.

Enzymes were purchased from Boehringer-Mannheim and used with their

supplied buffers. Deoxyadenosine and thymidine 5'- α - 32 P] triphosphates were obtained from Amersham, deoxyadenosine 5'-[α - 32 P]triphosphate was purchased from I.C.N. Sonicated, and 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.¹⁸

ImPyPy-(R)^{H2N} γ -PyPyPy- β -Dp (1-R). ImPyPy-(R)^{Fmoc} γ -PyPyPy- β -Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods.¹¹ (R)-2-Fmoc-4-Boc-diaminobutyric acid (0.7 mmol) was incorporated as previously described for Boc- α -aminobutyric acid. ImPyPy-(R)^{Fmoc} γ -PyPyPy- β -Pam-Resin was placed in a glass 20 mL peptide synthesis vessel and treated with DMF (2 mL), followed by piperidine (8 mL) and agitated (22 °C, 30 min.). ImPyPy-(R)^{H2N} γ -PyPyPy- β -Pam-resin was isolated by filtration, and washed sequentially with an excess of DMF, DCM, MeOH, and ethyl ether and the amine-resin dried *in vacuo*. A sample of ImPyPy-(R)^{H2N} γ -PyPyPy- β -Pam-resin (240 mg, 0.18 mmol/gram¹⁹) was treated with neat dimethylaminopropylamine (2 mL) and heated (55 °C) with periodic agitation for 16 h. The reaction mixture was placed in an oven and periodically agitated (55 °C, 16 h). The reaction mixture was then filtered to remove resin, 0.1% (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC. ImPyPy-(R)^{H2N} γ -PyPyPy- β -Dp is recovered upon lyophilization of the appropriate as a white powder (32 mg, 66% recovery). $[\alpha]_D^{20} +14.6$ (c 0.05, H₂O); UV (H₂O) λ_{max} 246, 310 (50,000); ¹H NMR (DMSO-*d*₆) δ 10.56 (s, 1 H), 10.47 (s, 1 H), 9.97 (s, 1 H), 9.94 (s, 1 H), 9.88 (s, 1 H), 9.4 (br s, 1 H), 8.28 (s, 3 H), 8.22 (m, 1 H), 8.03 (m, 2 H), 7.38 (s, 1 H), 7.25 (d, 1 H, *J* = 1.6 Hz), 7.22 (d, 1 H, *J* = 1.5 Hz), 7.19 (d, 1 H, *J* = 1.5 Hz), 7.16 (d, 1 H, *J* = 1.6 Hz), 7.14 (d, 1 H, *J* = 1.8 Hz), 7.12 (d, 1 H, *J* = 1.7 Hz), 7.03 (m, 2 H), 6.95 (d, 1 H, *J* = 1.6 Hz), 6.91 (d, 1 H, *J* = 1.6 Hz), 6.85 (d, 1 H, *J* = 1.6 Hz), 3.96 (s, 3 H), 3.83 (s, 3 H), 3.81 (m,

6 H), 3.79 (s, 3 H), 3.76 (s, 3 H), 3.33 (q, 2 H, $J = 6.3$ Hz), 3.25 (q, 2 H, $J = 5.7$ Hz), 3.05 (q, 2 H, $J = 5.9$ Hz), 2.96 (q, 2 H, $J = 5.3$ Hz), 2.71 (d, 6 H, $J = 4.9$ Hz), 2.32 (t, 2 H, $J = 7.1$ Hz), 1.95 (q, 2 H, $J = 5.9$ Hz), 1.70 (quintet, 2 H, $J = 7.3$ Hz); MALDI-TOF-MS (monoisotopic), 992.5 (992.5 calc. for $C_{47}H_{62}N_{17}O_8$).

ImPyPy-(R)^{(R)MTPA}- γ -PyPyPy- β -Dp (1-*R,R*) (R)- α -methoxy- α -(trifluoromethyl)phenylacetic acid (117 mg, 0.5 mmol) and HOBt (70 mg, 0.5 mmol) were dissolved in DMF (1 mL), DCC (100 mg, 0.5 mmol) added and the solution agitated for 30 min. at 22 °C. A sample of the activated ester solution (100 μ L, 0.05 mmol) was added to ImPyPy-(R)^{H₂N}- γ -PyPyPy- β -Dp 1-*R* (10 mg, 0.01 mmol), DIEA (50 μ L) added, and the solution agitated for 3 h (22 °C). DMF (1 mL) followed by 0.1% (wt/v) TFA (6 mL) was then added to the reaction mixture and the resulting solution purified by reversed phase HPLC (1% acetonitrile/min.) under conditions which were determined to the diastereomers. ImPyPy-(R)^{(R)MTPA}- γ -PyPyPy- β -Dp is recovered as a white powder upon lyophilization of the appropriate fractions (6 mg, 53% recovery). ¹H NMR (DMSO-*d*₆) δ 10.50 (s, 1 H), 10.14 (s, 1 H), 9.92 (s, 2 H), 9.88 (s, 1 H), 9.2 (br s, 1 H), 8.43 (d, 1 H, $J = 7.0$ Hz), 8.02 (m, 3 H), 7.92 (m, 1 H), 7.47 (m, 2 H), 7.41 (m, 2 H), 7.36 (s, 1 H), 7.24 (m, 1 H), 7.19 (m, 1 H), 7.15 (m, 1 H), 7.12 (m, 3 H), 7.01 (m, 2 H), 6.90 (m, 3 H), 6.83 (m, 1 H), 4.46 (q, 1 H, $J = 5.5$ Hz), 3.94 (s, 3 H), 3.79 (m, 9 H), 3.75 (m, 6 H), 3.32 (m, 4 H), 3.05 (m, 2 H), 2.94 (m, 2 H), 2.68 (d, 6H, $J = 4.0$ Hz), 2.28 (t, 2 H, $J = 6.3$ Hz), 1.93 (q, 2 H, $J = 6.1$ Hz), 1.66 (quintet, 2 H, $J = 6.0$ Hz), 1.18 (s, 3 H); MALDI-TOF-MS (monoisotopic), 1208.5 (1208.5 calc. for $C_{57}H_{68}F_3N_{17}O_{10}$).

ImPyPy-(R)^{(S)MTPA}- γ -PyPyPy- β -Dp 1-*R,S* ImPyPy-(R)^{(S)MTPA}- γ -PyPyPy- β -Dp was prepared from (S)- α -methoxy- α -(trifluoromethyl)phenylacetic acid as described for 1-*R,R* (5 mg, 45% recovery). ¹H NMR (DMSO-*d*₆) δ 10.47 (s, 1 H), 10.08 (s, 1 H), 9.92 (s, 2 H), 9.88 (s, 1 H), 9.2 (br s, 1 H), 8.43 (d, 1 H, $J = 6.9$ Hz), 8.02 (m, 3 H), 7.46 (m, 2 H), 7.40 (m,

2 H), 7.36 (s, 1 H), 7.23 (m, 1 H), 7.19 (m, 1 H), 7.14 (m, 1 H), 7.12 (m, 3 H), 7.01 (m, 2 H), 6.87 (m, 3 H), 6.83 (m, 1 H), 4.44 (q, 1 H, $J = 6.5$ Hz), 3.94 (s, 3 H), 3.79 (m, 9 H), 3.75 (m, 6 H), 3.28 (m, 4 H), 3.06 (m, 4 H), 2.94 (m, 2 H), 2.69 (d, 6H, $J = 4.5$ Hz), 2.28 (t, 2 H, $J = 6.5$ Hz), 1.93 (q, 2 H, $J = 6.1$ Hz), 1.66 (quintet, 2 H, $J = 6.0$ Hz), 1.18 (s, 3 H); MALDI-TOF-MS (monoisotopic), 1209.0 (1208.5 calc. for $C_{57}H_{68}F_3N_{17}O_{10}$).

ImPyPy-(S)^{H2N}γ-PyPyPy-γ-Dp (1-S). ImPyPy-(S)^{H2N}γ-PyPyPy-β-Dp was prepared as described for 1-R (23 mg, 49% recovery). $[\alpha]_D^{20}$ -14.2 (c 0.04, H₂O); ¹H NMR (DMSO-*d*₆) identical to 1-R; MALDI-TOF-MS (monoisotopic), 992.5 (992.5 calc. for $C_{47}H_{62}N_{17}O_8$).

ImPyPy-(R)^{H2N}γ-PyPyPy-β-EtOH (2-R). A sample of ImPyPy-(R)^{H2N}γ-PyPyPy-β-Pam-resin (240 mg, 0.18 mmol/gram¹⁹) was treated with neat ethanolamine (2 mL) and heated (55 °C) with periodic agitation for 16 h. The reaction mixture was then filtered to remove resin, 0.1% (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC to provide ImPyPy-(R)^{H2N}γ-PyPyPy-β-EtOH as a white powder upon lyophilization of the appropriate fractions (21 mg, 46% recovery). $[\alpha]_D^{20}$ +18.6 (c 0.04, H₂O); UV (H₂O) λ_{max} 246, 310 (50,000); ¹H NMR (DMSO-*d*₆) δ 10.55 (s, 1 H), 10.48 (s, 1 H), 9.97 (s, 1 H), 9.94 (s, 1 H), 9.89 (s, 1 H), 8.24 (m, 4 H), 8.00 (t, 1 H, $J = 4.1$ Hz), 7.89 (t, 1 H, $J = 5.8$ Hz), 7.38 (s, 1 H), 7.25 (d, 1 H, $J = 1.6$ Hz), 7.22 (d, 1 H, $J = 1.6$ Hz), 7.21 (d, 1 H, $J = 1.5$ Hz), 7.16 (m, 2 H), 7.14 (d, 1 H, $J = 1.6$ Hz), 7.03 (d, 1 H, $J = 1.7$ Hz), 6.99 (d, 1 H, $J = 1.4$ Hz), 6.95 (d, 1 H, $J = 1.6$ Hz), 6.91 (d, 1 H, $J = 1.5$ Hz), 6.78 (d, 1 H, $J = 1.5$ Hz), 5.33 (m, 1 H), 3.95 (s, 3 H), 3.83 (s, 3 H), 3.81 (m, 6 H), 3.79 (s, 3 H), 3.76 (s, 3 H), 3.37 (q, 2 H, $J = 6.2$ Hz), 3.07 (q, 2 H, $J = 5.9$ Hz), 2.29 (t, 2 H, $J = 7.1$ Hz), 1.93 (q, 2 H, $J = 5.8$ Hz), 1.20 (m, 4 H); MALDI-TOF-MS (monoisotopic), 951.4 (951.4 calc. for $C_{44}H_{55}N_{16}O_9$).

ImPyPy-(R)^{Ac}γ-PyPyPy-β-Dp (3-R). A sample of ImPyPy-(R)^{H₂N}γ-PyPyPy-β-Dp (4 mg) in DMSO (1 mL) was treated with a solution of acetic anhydride (1 mL) and DIEA (1 mL) in DMF (1 mL) and heated (55 °C) with periodic agitation for 30 min. Residual acetic anhydride was hydrolyzed (0.1 M NaOH, 1 mL, 55 °C, 10 min.), 0.1% (wt/v) TFA was added (6 mL) and the resulting solution purified by reversed phase HPLC to provide ImPyPy-(R)^{H₂N}γ-PyPyPy-β-Dp is recovered as a white powder upon lyophilization of the appropriate fractions (2 mg, 50% recovery). $[\alpha]_D^{20} +20.5$ (c 0.06, H₂O); UV (H₂O) λ_{\max} 242, 304 (50,000); ¹H NMR (DMSO-*d*₆) δ 10.49 (s, 1 H), 10.06 (s, 1 H), 9.94 (m, 2 H), 9.00 (s, 1 H), 9.4 (br s, 1 H), 8.21 (d, 1 H, *J* = 7.8 Hz), 8.06 (m, 2 H), 8.00 (t, 1 H, *J* = 6.2 Hz), 7.39 (s, 1 H), 7.27 (d, 1 H, *J* = 1.7 Hz), 7.21 (d, 1 H, *J* = 1.6 Hz), 7.18 (m, 2 H), 7.14 (m, 2 H), 7.03 (m, 2 H), 6.90 (d, 1 H, *J* = 1.6 Hz), 6.86 (m, 2 H), 4.43 (q, 1 H, *J* = 7.5 Hz), 3.96 (s, 3 H), 3.82 (m, 9 H), 3.73 (m, 6 H), 3.37 (q, 2 H, *J* = 5.8 Hz), 3.11 (q, 2 H, *J* = 6.9 Hz), 2.98 (q, 2 H, *J* = 5.4 Hz), 2.79 (q, 2 H, *J* = 5.3 Hz), 2.71 (d, 6 H, *J* = 4.7 Hz), 2.33 (t, 2 H, *J* = 6.2 Hz), 1.97 (s, 3 H), 1.70 (quintet, 2 H, *J* = 6.0 Hz) MALDI-TOF-MS (average), 1035.1 (1035.2 calc. for M+H).

ImPyPy-(S)^{Ac}γ-PyPyPy-β-Dp (3-S). ImPyPy-(S)^{Ac}γ-PyPyPy-β-Dp was prepared as described for 3-R. (2 mg, 50% recovery). $[\alpha]_D^{20} -16.4$ (c 0.07, H₂O); ¹H NMR (DMSO-*d*₆) is identical to 3-R; MALDI-TOF-MS (monoisotopic), 1034.6 (1034.5 calc. for C₄₉H₆₄N₁₇O₉).

ImPyPy-(R)^{Boc}γ-PyPyPy-β-Dp-NH₂ (4-R-Boc-NH₂). A sample of ImPyPy-(R)^{H₂N}γ-PyPyPy-β-Pam-resin (300 mg, 0.18 mmol/gram¹⁹) was treated with a solution of Boc-anhydride (500 mg) and DIEA (1 mL) in DMF (4 mL) and heated (55 °C) with periodic agitation for 30 min. ImPyPy-(R)^{Boc}γ-PyPyPy-β-Pam-resin was isolated by filtration, and washed sequentially with an excess of DMF, DCM, MeOH, and ethyl ether and the dried *in vacuo*. A sample of ImPyPy-(R)^{Boc}γ-PyPyPy-β-Pam-resin (240 mg, 0.18 mmol/gram¹⁹)

was treated with neat 3,3'-diamino-*N*-methylpropylamine (2 mL) and heated (55 °C) with periodic agitation for 16 h. The reaction mixture was then filtered to remove resin, 0.1% (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC to provide ImPyPy-(*R*)^{Boc}- γ -PyPyPy- β -Dp-NH₂ as a white powder upon lyophilization of the appropriate fractions (18 mg, 36% recovery); $[\alpha]_D^{20}$ -30 (c 0.05, H₂O); UV (H₂O) λ_{\max} 240, 306 (50,000); ¹H NMR (DMSO-*d*₆) δ 10.59 (s, 1 H), 10.16 (s, 1 H), 10.04 (m, 2 H), 10.00 (s, 1 H), 9.4 (br s, 1 H), 8.31 (d, 1 H, *J* = 7.8 Hz), 8.16 (m, 2 H), 8.10 (t, 1 H, *J* = 6.2 Hz), 7.89 (t, 1 H, *J* = 5.8 Hz), 7.49 (s, 1 H), 7.37 (d, 1 H, *J* = 1.7 Hz), 7.22 (d, 1 H, *J* = 1.6 Hz), 7.21 (d, 1 H, *J* = 1.5 Hz), 7.16 (m, 2 H), 7.14 (d, 1 H, *J* = 1.6 Hz), 7.03 (d, 1 H, *J* = 1.7 Hz), 6.99 (d, 1 H, *J* = 1.4 Hz), 6.95 (d, 1 H, *J* = 1.6 Hz), 6.91 (d, 1 H, *J* = 1.5 Hz), 6.78 (d, 1 H, *J* = 1.5 Hz), 5.33 (m, 1 H), 3.95 (s, 3 H), 3.83 (s, 3 H), 3.81 (m, 6 H), 3.79 (s, 3 H), 3.76 (s, 3 H), 3.37 (q, 2 H, *J* = 6.2 Hz), 3.07 (q, 2 H, *J* = 5.9 Hz), 2.29 (t, 2 H, *J* = 7.1 Hz), 1.93 (q, 2 H, *J* = 5.8 Hz), 1.20 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1135.3 (1135.6 calc. for C₅₄H₇₅N₁₈O₁₀).

ImPyPy-(*S*)^{Boc}- γ -PyPyPy- β -Dp-NH₂ (4-*S*-Boc-NH₂). ImPyPy-(*S*)^{Boc}- γ -PyPyPy- β -Dp-NH₂ was prepared as described for 4-*R*. (16 mg, 32% recovery). $[\alpha]_D^{20}$ -30 (c 0.05, H₂O); ¹H NMR (DMSO-*d*₆) is identical to 4-*R*-Boc-NH₂; MALDI-TOF-MS (monoisotopic), 1135.4 (1135.6 calc. for C₅₄H₇₅N₁₈O₁₀).

ImPyPy-(*R*)^{Boc}- γ -PyPyPy- β -Dp-EDTA (4-*R*-Boc). Excess EDTA-dianhydride (50 mg) was dissolved in DMSO/NMP (1 mL) and DIEA (1 mL) by heating at 55 °C for 5 min. The dianhydride solution was added to ImPyPy-(*R*)^{Boc}- γ -PyPyPy- β -Dp-NH₂ (10.4 mg, 10 μ mol) dissolved in DMSO (750 μ L). The mixture was heated (55 °C, 25 min.) and the remaining EDTA-anhydride hydrolyzed (0.1M NaOH, 3 mL, 55 °C, 10 min.). Aqueous TFA (0.1% wt/v) was added to adjust the total volume to 8 mL and the solution purified directly by reversed phase HPLC to provide ImPyPy-(*R*)^{Boc}- γ -PyPyPy- β -Dp-

EDTA (4-*R*-Boc) as a white powder upon lyophilization of the appropriate fractions (4 mg, 40% recovery). MALD-TOF-MS (monoisotopic), 1409.6 (1409.7 calc. for $C_{64}H_{89}N_{20}O_{17}$).

ImPyPy-(S)^{Boc}- γ -PyPyPy- β -Dp-EDTA (4-*S*-Boc). ImPyPy-(S)^{Boc}- γ -PyPyPy- β -Dp-NH₂ (12.0 mg, 12 μ mol) was converted to 4-*S*-Boc as described for 4-*R*-Boc (4 mg, 33% recovery). MALDI-TOF-MS (monoisotopic), 1409.7 (1409.7 calc. for $C_{64}H_{89}N_{20}O_{17}$).

ImPyPy-(R)^{H₂N}- γ -PyPyPy- β -Dp-EDTA (4-*R*). A sample of ImPyPy-(R)^{Boc}- γ -PyPyPy- β -Dp-EDTA (2.1 mg) in DMSO (750 μ L) was placed in a 50 mL flask and treated with TFA (15 mL, 22 °C, 2 h). Excess TFA was removed *in vacuo*, water added (6 mL) and the resulting solution purified by reversed phase HPLC to provide ImPyPy-(R)^{H₂N}- γ -PyPyPy- β -Dp-EDTA as a white powder upon lyophilization of the appropriate fractions (1.3 mg, 50% recovery). MALDI-TOF-MS (monoisotopic), 1309.5 (1309.6 calc. for $C_{59}H_{81}N_{20}O_{15}$).

ImPyPy-(S)^{H₂N}- γ -PyPyPy- β -Dp-EDTA (4-*S*). ImPyPy-(S)^{Boc}- γ -PyPyPy- β -Dp-EDTA (3.0 mg) was converted to 4-*S* as described for 4-*R* (1 mg, 33% recovery). MALDI-TOF-MS (monoisotopic), 1309.5 (1309.6 calc. for $C_{59}H_{81}N_{20}O_{15}$).

ImPyPy-(R)^{EDTA}- γ -PyPyPy- β -Dp (5-*R*). Excess EDTA-dianhydride (50 mg) was dissolved in DMSO/NMP (1 mL) and DIEA (1 mL) by heating at 55 °C for 5 min. The dianhydride solution was added to ImPyPy-(R)^{H₂N}- γ -PyPyPy- β -Dp (1.0 mg, 1 μ mol) dissolved in DMSO (750 μ L). The mixture was heated (55 °C, 25 min.) and remaining EDTA-anhydride was hydrolyzed (0.1M NaOH, 3mL, 55 °C, 10 min.). Aqueous TFA (0.1% wt/v) was added to adjust the total volume to 8 mL and the solution purified directly by reversed phase HPLC to provide 5-*R* as a white powder upon lyophilization

of the appropriate fractions (0.6 mg, 60% recovery). MALDI-TOF-MS (monoisotopic), 1266.4 (1266.6 calc. for C₅₇H₇₆N₁₉O₁₅).

ImPyPy-(S)^{EDTA}γ-PyPyPy-β-Dp (5-S). ImPyPy-(S)^{EDTA}γ-PyPyPy-β-Dp was prepared from 1-S as described for 5-R (6.8 mg, 16% recovery). MALDI-TOF-MS (monoisotopic), 1266.5 (1266.6 calc. for C₅₇H₇₆N₁₉O₁₅).

Preparation of 3'- and 5'-End-Labeled Restriction Fragments. The plasmid pMM5 was linearized with *Eco*RI and *Bsr*BI, then treated with the Sequenase enzyme, deoxyadenosine 5'-α³²P]triphosphate and thymidine 5'-α³²P]triphosphate for 3' labeling. Alternatively, pMM5 was linearized with *Eco*RI, treated with calf alkaline phosphatase, and then 5' labeled with T4 polynucleotide kinase and deoxyadenosine 5'-[α-³²P]triphosphate. The 5' labeled fragment was then digested with *Bsr*BI. The labeled fragment (3' or 5') was loaded onto a 6% non-denaturing polyacrylamide gel, and the desired 135 base pair band was visualized by autoradiography and isolated. Chemical sequencing reactions were performed according to published methods.²⁰

MPE•Fe(II) Footprinting. 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 hours. A fresh 50 μM MPE•Fe(II) solution was prepared from 100 μL of a 100 μM MPE solution and 100 μL of a 100 μM ferrous ammonium sulfate (Fe(NH₄)₂(SO₄)₂•6H₂O) solution. MPE•Fe(II) solution (5 μM) was added to the equilibrated DNA, and the reactions were allowed to equilibrate 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 6 min., and a 5 µL sample (~ 15 kcpm) was immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V.

Affinity Cleaving. 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), 20 mM NaCl, 100 µM/base pair calf thymus DNA, and 20 kcpm 3'- or 5'-radiolabeled DNA. The solutions were allowed to equilibrate for 8 hours. A fresh solution of ferrous ammonium sulfate ($\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$) (10 µM) was added to the equilibrated DNA, and the reactions were allowed to equilibrate for 15 min. Cleavage was initiated by the addition of dithiothreitol (10 mM) and allowed to proceed for 30 min. Reactions were stopped by ethanol precipitation, resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer, denatured at 85 °C for 6 min., and the entire sample was immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V.

DNase I Footprinting. All reactions were carried out in a volume of 400 µL. We note explicitly that no carrier DNA was used in these reactions until after DNase I cleavage. 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 MgCl₂, 5 mM CaCl₂, and 30 kcpm 3'-radiolabeled DNA. The solutions were allowed to equilibrate for a minimum of 12 hours 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 1.875 u/mL) and was allowed to proceed for 7 min. at 22 °C. The reactions were stopped by adding 50 µL of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/mL 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 6 min, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V for 1 hour. The gels were dried under vacuum at 80 °C, then quantitated using storage phosphor technology.

Equilibrium association constants were determined as previously described.^{9a} The data were analyzed by performing volume integrations of the 5'-TGTTA-3' and 5'-TGACA-3 sites and a reference site. The apparent DNA target site saturation, θ_{app} , was calculated for each concentration of polyamide using the following equation:

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

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{ref}° correspond to those values for a DNase I control lane to which no polyamide has been added. The ($[L]_{tot}$, θ_{app}) data points were fit to a Langmuir binding isotherm (eq 2, $n=1$ for polyamides 1-3, $n=2$ for polyamides 4 and 5) by minimizing the difference between θ_{app} and θ_{fit} , using the modified Hill equation:

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

where $[L]_{tot}$ corresponds to the total polyamide concentration, K_a corresponds to the equilibrium association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software

(version 2.1, Abelbeck software) with K_a , θ_{\max} , and θ_{\min} as the adjustable parameters. All acceptable fits had a correlation coefficient of $R > 0.97$. At least three 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. The data were normalized using the following equation:

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

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 12-20 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.

References

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.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646. (f) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756.
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) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532. (d) Chen, X.; Ramakrishnan, Sundaralingham, M. *J. Mol. Biol.* **1997**, *267*, 1157.
4. (a) Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 5911; (b) Al-Said, N.; Lown, J. W. *Tet. Lett.* **1994**, *35*, 7577. (c) Walker, W. L.; Landlaw, E. M., Dickerson, R. E.; Goodsell, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 5634. (d) Kopka, M. L.; Goodsell, D. S.; Han, G. W.; Chiu, T. K.; Lown, J. W.; Dickerson, R. E. *Structure* **1997**, *5*, 1033.
5. (a) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474. (b) Singh, S. B.; Ajay; Wemmer, D. E.; Kollman,

- P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7673; (c) White, S.; Baird, E. E.; Dervan, P. B. *Chem. & Biol.* **1997**, *4*, 569.
6. (a) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160. (b) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Nature Struct. Biol.* **1996**, *3*, 321. (c) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1997**, *3*, 1608.
7. For a review, see: Wemmer, D. E. and Dervan, P. B. *Curr. Opin. Struct. Biol.* **1997**, *7*, 355.
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.
9. (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. (i) de Claire, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909.

10. Gottesfeld, J. M.; Nealy, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202.
11. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.
12. (a) Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5470. (b) Van Dyke, M. W.; Dervan, P. B. *Science* **1984**, *225*, 1122.
13. (a) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457. (b) Dervan, P. B. *Science* **1986**, *232*, 464
14. (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.
15. (a) Dale, J. A.; Mosher, H. S. *J. Am. Chem. Soc.* **1973**, *95*, 512. (b) Yamaguchi, S. *Asymmetric Synthesis (Vol. 1), Analytical Methods* p.125-152, J. D. Morrison (ed.) Academic Press (1983).
16. (a) Zimmer, C.; Wahnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31. (b) Pullman, B. *Adv. Drug. Res.* **1990**, *18*, 1. (c) Breslauer, K. J.; Ferrante, R.; Marky, L. A.; Dervan, P. B.; Youngquist, R. S. *Structure and Expression (Vol. 2), DNA and Its Drug Complexes* p. 273-289, R. H. Sarma and M. H. Sarma (eds.) Academic Press (1988).

17. Kent, S.B.H. *Annu. Rev. Biochem.* **1988**, *57*, 957.
18. Sambrook, J.; Fritsch, E.F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.
19. Resin substitution can be calculated as $L_{\text{new}}(\text{mmol/g}) = L_{\text{old}} / (1 + L_{\text{old}}(W_{\text{new}} - W_{\text{old}}) \times 10^{-3})$, where L is the loading (mmol of amine per gram of resin), and W is the weight (gmol^{-1}) of the growing polyamide attached to the resin. see: Barlos, K.; Chatzi, O.; Gatos, D.; Stravropoulos, G. *Int. J. Peptide Protein Res.* **1991**, *37*, 513.
20. (a) Iverson, B.L.; Dervan, P.B. *Nucl. Acids Res.* **1987**, *15*, 7823. (b) Maxam, A.M.; Gilbert, W.S. *Methods Enzymol.* **1980**, *65*, 499.

Chapter 3

Tandem Hairpin Motif for Recognition in the Minor Groove of DNA by Pyrrole-Imidazole Polyamides

Abstract: Two six-ring hairpin polyamides linked “tail-to-turn” by a five-carbon tether recognize a predetermined 11 base pair (bp) site in the minor groove of DNA. Polyamide subunits containing three pyrrole (Py) or imidazole (Im) aromatic amino acids covalently linked by a turn-specific γ -aminobutyric acid (γ -turn) residue, form six-ring hairpin structures that recognize designated five base pair sequences. Replacement of the γ -turn residue with (R)-2,4-diaminobutyric acid ((R)^{H2N} γ) provides for enhanced hairpin-DNA-binding affinity and sequence specificity. In order to extend the targetable binding site size of the hairpin motif, two tandem hairpin polyamides, ImPyPy-(R)[ImPyPy-(R)^{H2N} γ -PyPyPy- β -J^{H2N} γ -PyPyPy- β -Dp (1) and ImPyPy-(R)[ImPyPy-(R)^{H2N} γ -PyPyPy- δ -J^{H2N} γ -PyPyPy- β -Dp (2), were designed such that the carboxy tail of one six-ring hairpin is covalently tethered to the (R)^{H2N} γ -turn of the second using β -alanine (β) or 5-aminovaleric acid (δ) respectively. Fmoc-protection of the turn amino group makes the tandem-hairpin polyamides accessible with existing Boc-chemistry machine-assisted solid phase synthesis protocols. The DNA-binding affinity of each polyamide was characterized by quantitative footprint titration experiments on DNA fragments containing 10-, 11-, or 12-bp match and mismatch sequences. The parent six-ring hairpin ImPyPy-(R)^{H2N} γ -PyPyPy- β -Dp binds to a 5-bp 5'-TGTTA-3' half site with an equilibrium association constant (K_a) = $5 \times 10^9 M^{-1}$ and 100-fold specificity versus a 5'-TGTCA-3' mismatch site. Valeric acid linked tandem-hairpin polyamide 2 binds the 11 bp site 5'-TGTTATTGTTA-3' (individual 6-ring hairpin target sites underlined) with $K_a \geq 1 \times 10^{12} M^{-1}$ and > 4500-fold specificity versus the double mismatch sequence 5'-TGTCATTGTCA-3'. The 10-bp and 12-bp sites 5'-TGTTATGTTA-3' and 5'-TGTTATTTGTTA-3' are bound with at least 70 and 1000-fold reduced affinity, respectively. β -linked polyamide 1 binds to both the 10- and 11-bp sites with $K_a = 2 \times 10^{10} M^{-1}$ and to the 12-bp site with $K_a = 9 \times 10^8 M^{-1}$. The results presented here identify structure elements which expand polyamide binding site size by linking previously described hairpin recognition units. Remarkably, a simple aliphatic 5-carbon tether is sufficient to provide increased binding affinity without compromising hairpin sequence-selectivity. **Publication:** Herman, Baird & Dervan *Chem. Eur. J.* **1999**, *5*, 975-983.

Introduction

Small synthetic ligands that target predetermined 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 small molecules that have an affinity and specificity for DNA comparable to naturally occurring DNA binding proteins.^{[2],[3]} DNA recognition depends on a code of side-by-side aromatic amino acid pairings that are oriented N-C with respect to the 5'-3' direction of the DNA helix in the minor groove.^{[2]-[11]} An antiparallel 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.^[4] 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 four base pair code.^[3] Eight-ring polyamides have been shown to be cell permeable and to inhibit transcription of a specific gene in cell culture.^[1] This provides impetus to develop an ensemble of motifs which recognize a broad binding site size repertoire.^{[2]-[11]} It is particularly important to identify ligand-structure elements which amplify existing recognition motifs for binding to DNA sequences 10-16 bp in size.

Hairpin Polyamide: A hairpin polyamide motif with γ -aminobutyric acid (γ) serving as a turn-specific internal-guide-residue provides specific binding to designated target sites with > 100-fold enhanced affinity relative to the unlinked subunits.^[9] Studies of polyamide site size limitations suggest that beyond five consecutive rings, the ligand curvature fails to match the pitch of the DNA helix, disrupting the hydrogen bonds and van der Waals interactions responsible for specific polyamide-DNA complex formation.^{[6],[7a]} The recognition of seven base pairs by ten-ring hairpin polyamides containing five contiguous ring pairings represents the upper limit in binding site sizes targetable by the hairpin motif.^[2c] Addition of pairings with β -alanine (β) to form β/β , β /Py, and β /Im pairs has allowed extension of the hairpin motif to 8-bp recognition.^[2e]

Cooperative-binding extended-hairpins provide one motif for expanding the hairpin recognition for targeting 10-bp and 12-bp sites.^[2d] An alternate approach to increase the targetable binding site size of hairpins would be to identify a strategy for covalently linking existing hairpin motifs without compromising DNA-binding and sequence-specificity.

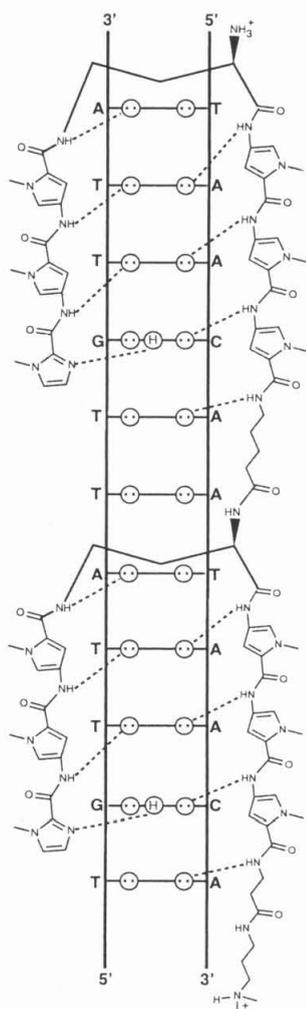
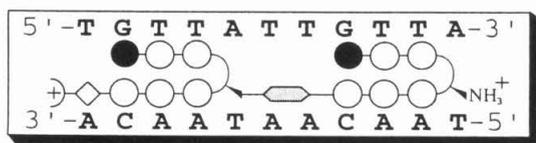


Figure 3.1 (Top) Hydrogen bonding model of the 1:1 polyamide:DNA complex formed between the tandem hairpin ImPyPy-(R)[ImPyPy-(R)^{H^{2N}}- γ -PyPyPy- δ -]^{H^N}- γ -PyPyPy- β -Dp (**2**) and the 11-bp 5'-TGTTATTGTTA-3' site. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. (Bottom) For schematic binding model, Im and Py rings are represented as shaded and unshaded spheres respectively. The b-residue and valeric acid linker are represented as an unshaded diamond and an unshaded hexagon, respectively.



Within the hairpin structure, replacement of the γ -turn residue with (*R*)-2,4-diaminobutyric acid ((*R*)^{H₂N} γ) has recently been shown to enhance hairpin DNA-binding affinity and sequence specificity.^[10] The primary turn-amino group provides a potential site for covalently connecting two hairpins. In one potential linkage arrangement, the C-terminus of the first hairpin is coupled to the α -amino group of the γ -turn of the second by an amino acid linker (Figure 3.1). To determine preferred binding site size and linker length effects for tandem hairpins, two 12-ring polyamides, ImPyPy-(*R*)[ImPyPy-(*R*)^{H₂N} γ -PyPyPy- β -]^{H_N} γ -PyPyPy- β -Dp (**1**) and ImPyPy-(*R*)[ImPyPy-(*R*)^{H₂N} γ -PyPyPy- δ -]^{H_N} γ -PyPyPy- β -Dp (**2**) (Figure 3.2), were synthesized and their DNA binding properties

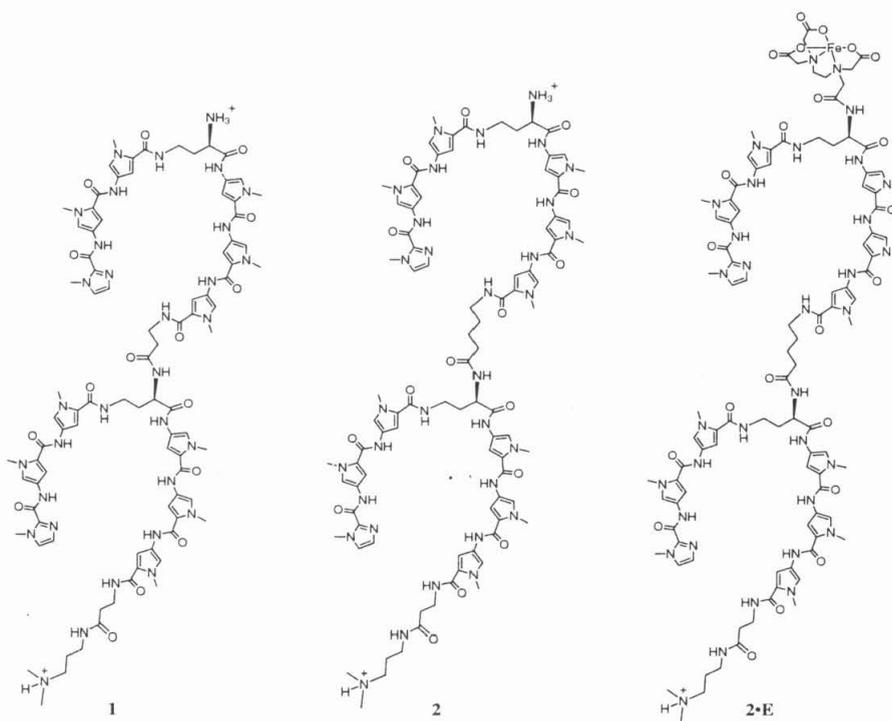


Figure 3.2 Structures of the 12-ring hairpin polyamides ImPyPy-(*R*)[ImPyPy-(*R*)^{H₂N} γ -PyPyPy- β -]^{H_N} γ -PyPyPy- β -Dp (**1**), ImPyPy-(*R*)[ImPyPy-(*R*)^{H₂N} γ -PyPyPy- δ -]^{H_N} γ -PyPyPy- β -Dp (**2**), and ImPyPy-(*R*)[ImPyPy-(*R*)^{EDTA} γ -PyPyPy- δ -]^{H_N} γ -PyPyPy- β -Dp (**2-E**) synthesized by solid phase methods.

determined on a series of DNA fragments containing 10-, 11-, and 12-bp target sites.

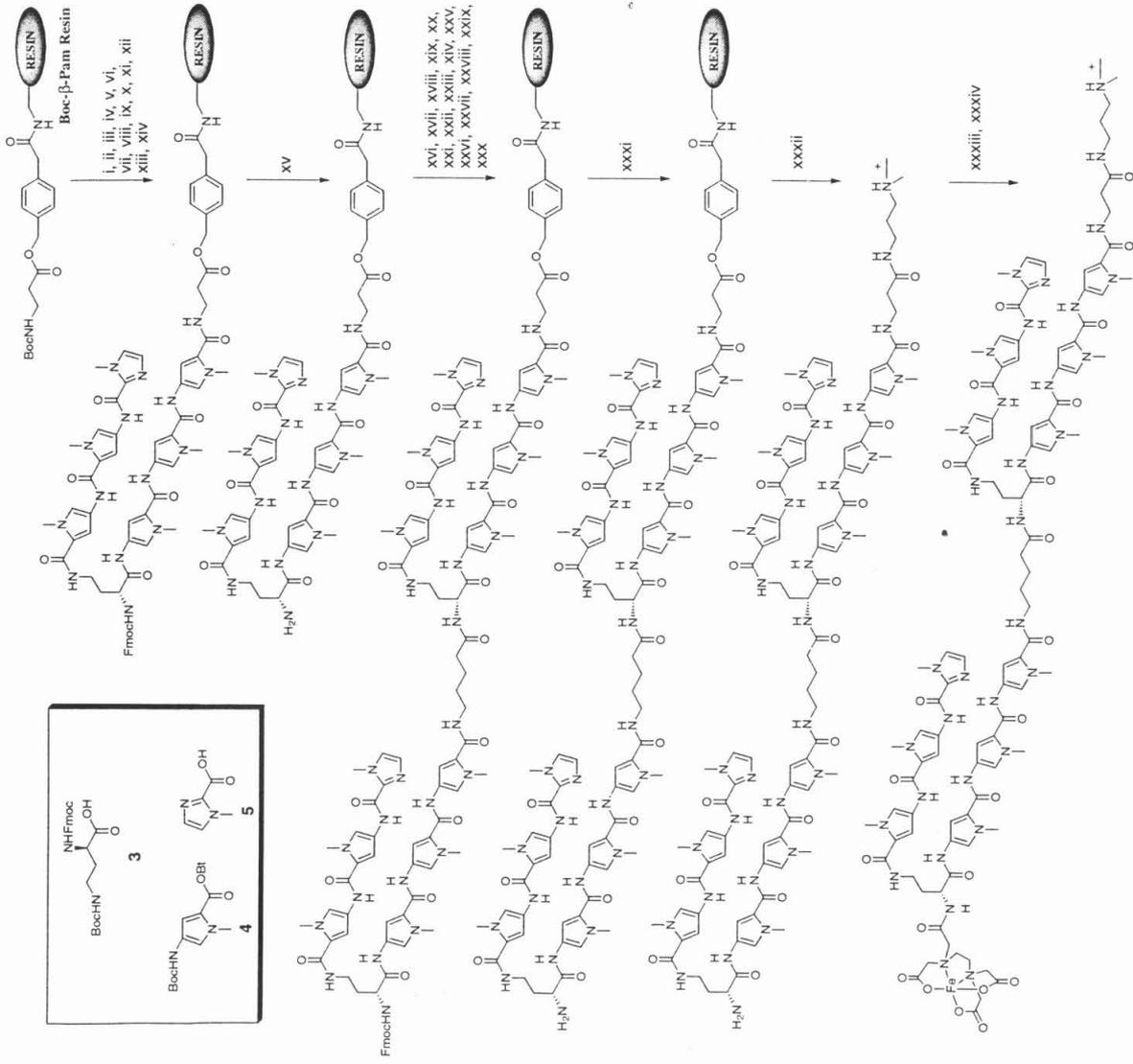
Polyamides were synthesized by solid phase methods,^[12] and their purity and identity confirmed by ¹H NMR and MALDI-TOF MS, and analytical HPLC. An affinity cleaving

derivative $\text{ImPyPy-(R)[ImPyPy-(R)}^{\text{EDTA}}\text{-}\gamma\text{-PyPyPy-}\delta\text{-]}^{\text{HN}}\text{-}\gamma\text{-PyPyPy-}\beta\text{-Dp}$ (**2-E**) was synthesized in order to confirm a single predicted binding orientation for the tandem hairpin polyamide. We report here the DNA-binding affinity and sequence selectivity of **1** and **2** for the 10-, 11-, and 12- bp match sites 5'-TGTTATGTTA-3', 5'-TGTTATTGTTA-3', and 5'-TGTTATATGTTA-3' (5-bp hairpin target sites are underlined) and double mismatch sites 5'-TGTCATGTCA-3', 5'-TGTCATTGTCA-3', and 5'-TGTCATATGTCA-3' (mismatched base pairs are bold) respectively. Precise binding site sizes were determined by MPE•Fe(II) footprinting,^[13] and binding orientation and stoichiometry confirmed by affinity cleaving experiments.^[14] Equilibrium association constants (K_a) of the polyamides for respective match and mismatch binding sites were determined by quantitative DNase I footprint titration.^[15]

Results and Discussion

Synthesis: $\text{ImPyPy-(R)[ImPyPy-(R)}^{\text{H}_2\text{N}}\text{-}\gamma\text{-PyPyPy-}\beta\text{-]}^{\text{HN}}\text{-}\gamma\text{-PyPyPy-}\beta\text{-Dp}$ (**1**) and $\text{ImPyPy-(R)[ImPyPy-(R)}^{\text{H}_2\text{N}}\text{-}\gamma\text{-PyPyPy-}\delta\text{-]}^{\text{HN}}\text{-}\gamma\text{-PyPyPy-}\beta\text{-Dp}$ (**2**) were synthesized from Boc- β -alanine-Pam resin (0.6 g resin, 0.6 mmol/g substitution) using Boc-chemistry machine-assisted protocols in 31 steps (Figure 3.3).^[12] $\text{ImPyPy-(R)}^{\text{FmocHN}}\text{-}\gamma\text{-PyPyPy-}\beta\text{-Pam-Pam}$ resin was prepared as described.^[10] The Fmoc protecting group was then removed by treatment with (4:1) piperidine/DMF. The remaining amino acid sequence was then synthesized in a stepwise manner using Boc-chemistry machine assisted protocols to provide $\text{ImPyPy-(R)[ImPyPy-(R)}^{\text{FmocHN}}\text{-}\gamma\text{-PyPyPy-}\beta\text{-]}^{\text{HN}}\text{-}\gamma\text{-PyPyPy-}\beta\text{-Pam-Resin}$ and $\text{ImPyPy-(R)[ImPyPy-(R)}^{\text{FmocHN}}\text{-}\gamma\text{-PyPyPy-}\delta\text{-]}^{\text{HN}}\text{-}\gamma\text{-PyPyPy-}\beta\text{-Pam-Resin}$. The Fmoc group was removed with (4:1) piperidine/DMF. A sample of resin was then cleaved by a single-step aminolysis reaction with ((dimethylamino)propylamine (55 °C, 18 h) and the reaction mixture subsequently purified by reversed phase HPLC to provide $\text{ImPyPy-(R)[ImPyPy-(R)}^{\text{H}_2\text{N}}\text{-}\gamma\text{-PyPyPy-}\beta\text{-]}^{\text{HN}}\text{-}\gamma\text{-PyPyPy-}\beta\text{-Dp}$ (**1**) and $\text{ImPyPy-(R)[ImPyPy-(R)}^{\text{H}_2\text{N}}\text{-}\gamma\text{-PyPyPy-}\delta\text{-]}^{\text{HN}}\text{-}\gamma\text{-PyPyPy-}\beta\text{-Dp}$ (**2**). For the synthesis of the EDTA-turn

Figure 3.3 Solid phase synthetic scheme exemplified for ImPyPy-(*R*)[ImPyPy-(*R*)^{H²N}- γ -PyPyPy- β -]^{H^N}- γ -PyPyPy- β -Dp (**1**), ImPyPy-(*R*)[ImPyPy-(*R*)^{H²N}- γ -PyPyPy- δ -]^{H^N}- γ -PyPyPy- β -Dp (**2**): (i) 80 % TFA/DCM, 0.4 M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) 80 % TFA/DCM, 0.4 M PhSH; (iv) Boc-Py-OBt, DIEA, DMF; (v) 80 % TFA/DCM, 0.4 M PhSH; (vi) Boc-Py-OBt, DIEA, DMF; (vii) 80 % TFA/DCM, 0.4 M PhSH; (viii) Fmoc- α -Boc- γ -diaminobutyric acid (HBTU, DIEA); (ix) 80 % TFA/DCM, 0.4 M PhSH; (x) Boc-Py-OBt, DIEA, DMF; (xi) 80 % TFA/DCM, 0.4 M PhSH; (xii) Boc-Py-OBt, DIEA, DMF; (xiii) 80 % TFA/DCM, 0.4 M PhSH; (xiv) imidazole-2-carboxylic acid (HBTU/DIEA); (xv) 80 % Piperidine:DMF (25 °C, 30 min) (xvi) Boc- β -alanine (HOBT/DIEA) for **1**; Boc-valeric acid (HOBT/DIEA) for **2** (xvii) 80 % TFA/DCM, 0.4 M PhSH; (xviii) Boc-Py-OBt, DIEA, DMF; (xix) 80 % TFA/DCM, 0.4 M PhSH; (xx) Boc-Py-OBt, DIEA, DMF; (xxi) 80 % TFA/DCM, 0.4 M PhSH; (xxii) Boc-Py-OBt, DIEA, DMF; (xxiii) 80 % TFA/DCM, 0.4 M PhSH; (xxiv) Fmoc- α -Boc- γ -diaminobutyric acid (HBTU, DIEA); (xxv) 80 % TFA/DCM, 0.4 M PhSH; (xxvi) Boc-Py-OBt, DIEA, DMF; (xxvii) 80 % TFA/DCM, 0.4 M PhSH; (xxviii) Boc-Py-OBt, DIEA, DMF; (xxix) 80 % TFA/DCM, 0.4 M PhSH; (xxx) imidazole-2-carboxylic acid (HBTU/DIEA); (xxxi) 80 % Piperidine:DMF (25 °C, 30 min.); (xxxii) N-N-((dimethylamio)propyl)amine, 55 °C; (Inset) Py, Im, and diaminobutyric acid monomers for solid phase synthesis: (*R*)-Fmoc- α -Boc- γ -diaminobutyric acid **3-R**, Boc-Pyrrole-OBt ester^[12] (Boc-Py-OBt) **4**, and Imidazole-2-Carboxylic acid^[14a] (Im-OH) **5**.



derivative **2-E**, a sample ImPyPy-(R)[ImPyPy-(R)^{H₂N}γ-PyPyPy-δ-]^{HN}γ-PyPyPy-β-Dp (**2**) was treated with an excess of EDTA-dianhydride (DMSO/NMP, DIEA 55 °C, 30 min.) and the remaining anhydride hydrolyzed (0.1 M NaOH, 55 °C, 10 min.). The polyamide ImPyPy-(R)[ImPyPy-(R)^{EDTA}γ-PyPyPy-δ-]^{HN}γ-PyPyPy-β-Dp (**2-E**) was then isolated by reverse phase HPLC. The dicationic twelve-ring tandem hairpins are soluble in aqueous solution at concentrations up to 1 mM. The solubility of the tandem hairpins is 10- to 100-fold greater than that found for extended or hairpin twelve-ring polyamides.

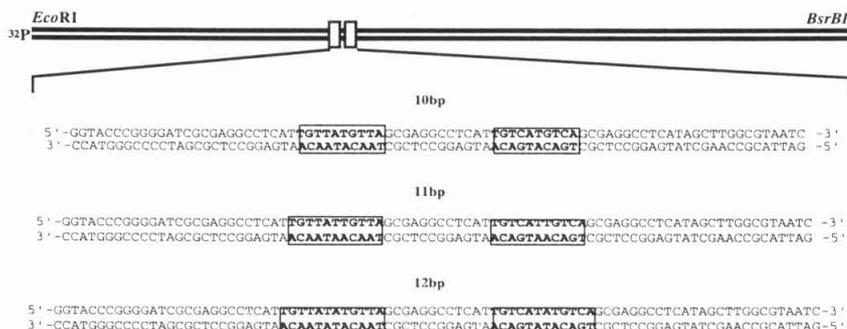


Figure 3.4 Sequence of the synthesized inserts from the pDH10, pDH11, and pDH12 plasmids containing 10-bp, 11-bp, and 12-bp match and mismatch target sites. Top: illustration of the *Eco*RI/*Bsr*BI restriction fragments containing the *Bam*HI and *Hind*III inserts as indicated below. Only the boxed sites were analyzed by quantitative DNase I footprint titrations.

Binding Site Size: MPE•Fe(II) footprinting^[13] on 3'- or 5'-³²P end-labeled 135 base pair *Eco*RI/*Bsr*BI restriction fragments from the plasmid pDH11 reveals that polyamide **2**, at 100 pM concentration, binds to the designated 11-bp match site 5'-TGTTATTGTTA-3' (25 mM Tris-acetate, 10 mM NaCl, pH 7.0 and 22 °C) (Figures 4 and 5b). Binding of the mismatch site 5'-TGTCATTGTCA-3' is only observed at much higher polyamide concentrations. The size of the asymmetrically 3'-shifted cleavage protection pattern for polyamide **2** at the designated match site 5'-TGTTATTGTTA-3' is consistent with formation of the predicted hairpin-δ-hairpin•DNA complex.

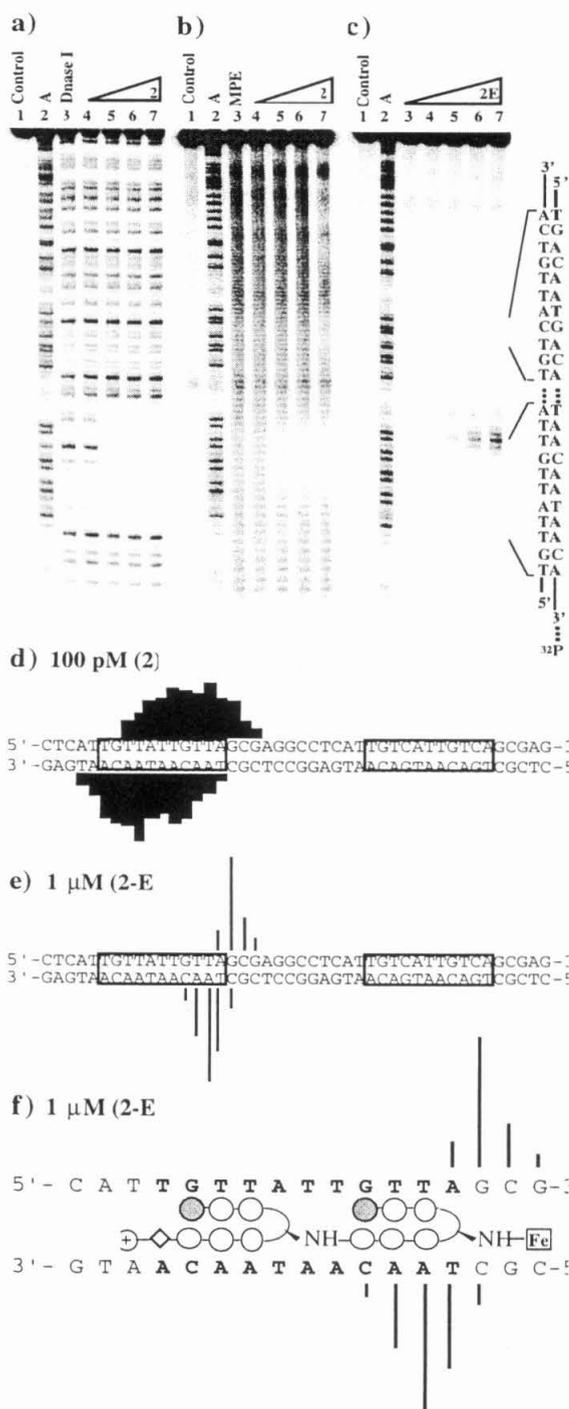
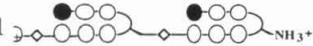
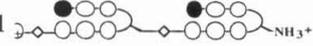
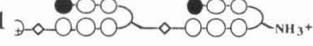


Figure 3.5 Footprinting experiments on the on a 3'-³²P-labeled 147 bp DNA restriction fragment derived from the plasmid pDH11. (a) Quantitative DNase I footprint titration experiment with ImPyPy-(R)[ImPyPy-(R)^{H2N}γ-PyPyPy-δ-]^{HN}γ-PyPyPy-β-Dp (2): lane 1, intact; lane 2, A reaction; lane 3, DNase I standard, lanes 4-7, 1 pM, 3 pM, 10 pM, and 30 pM. All reactions contain 10 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂. (b) MPE•Fe(II) footprinting of tandem-hairpin 2: Lane 1, intact; lane 2, A reaction; lane 3, MPE•Fe(II) standard; lanes 4-7; 10 pM, 100 pM, 1 nM, and 10 nM polyamide. (c) affinity cleaving titration experiment using tandem-hairpin 2-E: Lane 1, control reaction; lane 2, A reaction; lanes 3-7, 10 nM, 30 nM, 100 nM, 300 nM, and 1 μM polyamide. 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. 5'-TGTTATTGTTA-3' and 5'-TGTCATTGTCA-3' sites are shown on the right side of the autoradiograms. (d) MPE•Fe(II) protection patterns for 2 at 100 pM concentration. Bar heights are proportional to the relative protection from cleavage at each band. (e) Affinity cleaving pattern for 2-E at 1 μM concentration. Bar heights are proportional to the relative cleavage intensities at each base pair. (f) Ball and stick schematic of compound 2-E•11-bp 5'-TGTTATTGTTA-3' complex showing the affinity cleaving data on the right side of the autoradiogram.

Binding Orientation: Affinity cleavage experiments^[14] using **2-E** which has an EDTA•Fe(II) moiety appended to the γ -turn, were used to confirm polyamide binding orientation and stoichiometry. Affinity cleavage experiments were performed on the same 3'- or 5'-³²P end-labeled 135 base pair DNA restriction fragment from the plasmid pDH11 (25 mM Tris-acetate, 10 mM NaCl, 100 μ M/base pair calf thymus DNA, pH 7.0 and 22 °C). The observed cleavage pattern for **2-E** (Figures 3.5b and 3.5d) are 3'-shifted, consistent with minor groove occupancy. In the presence of 1 μ M **2-E**, a single cleavage locus proximal to the 3' side of the 5'-TGTTATTGTTA-3' match sequence is revealed, consistent with formation of an oriented 1:1 hairpin- δ -hairpin•DNA complex.

Equilibrium Association Constants: Quantitative DNase I footprint titrations (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂, pH 7.0 and 22 °C) were performed to determine the equilibrium association constants of **1** and **2** for the 10-, 11- and 12-bp match and mismatch sites (Table 3.1, Figure 3.4). Polyamide **2** preferentially binds the 11-bp 5'-TGTTATTGTTA-3' target sequence with, $K_a > 1 \times 10^{11} \text{ M}^{-1}$. The corresponding 11-bp mismatch 5'-TGTCATTGTCA-3' site is bound by **2** with > 4500-fold lower affinity ($K_a = 2.2 \times 10^8 \text{ M}^{-1}$). Polyamide **2** binds the 10-bp site 5'-TGTTATGTTA-3' ($K_a = 1.5 \times 10^{10} \text{ M}^{-1}$) and the 12-bp site 5'-TGTTATATGTTA-3' ($K_a = 1.0 \times 10^9 \text{ M}^{-1}$) with 70- and 1000-fold lower affinity, respectively. Polyamide **1** binds the 10-bp 5'-TGTTATGTTA-3' site and 11-bp 5'-TGTTATTGTTA-3' site with $K_a = 2 \times 10^{10} \text{ M}^{-1}$, and also binds the 12-bp 5'-TGTTATATGTTA-3' site with 16-fold lower affinity ($K_a = 9.0 \times 10^9 \text{ M}^{-1}$). The parent hairpin ImPyPy-(R)^{H₂N} γ -PyPyPy- β -Dp was found to bind to the 5'-TGTTA-3' match site with $K_a = 5 \times 10^9 \text{ M}^{-1}$.

Table 1. Equilibrium Association Constants [M^{-1}]^[a-c]

Polyamide	5'- <u>TGTTAT</u> GT <u>TA</u> -3'	5'-TGTC <u>ATGT</u> <u>CA</u> -3'	Specificity ^[d]
1 	2.0×10^{10}	1.5×10^8	133
2 	1.5×10^{10}	1.9×10^8	80
Polyamide	5'- <u>TGTTAT</u> <u>TGTTA</u> -3'	5'-TGTC <u>ATTGT</u> <u>CA</u> -3'	Specificity
1 	1.5×10^{10}	2.7×10^8	55
2 	$\geq 1 \times 10^{12}$	2.2×10^8	≥ 4500
Polyamide	5'- <u>TGTTA</u> <u>TATGTTA</u> -3'	5'-TGTC <u>ATATGT</u> <u>CA</u> -3'	Specificity
1 	9.4×10^9	3.1×10^7	30
2 	1.3×10^9	2.5×10^7	52

[a] The reported association constants are the average 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 MgCl₂, and 5 mM CaCl₂. The ten, eleven, and twelve base-pair sites are in capital letters. Match site association constants are shown in boldtype. [c] Half sites for tandem hairpin polyamide binding are underlined and mismatch bases are underlined in boldtype for all target sequences. [d] Specificity is calculated as $K_a(\text{match})/K_a(\text{mismatch})$.

Linker Dependence: Site size preferences of polyamides **1** and **2** are modulated by the length of the turn-to-tail linker. Modeling indicated that β and δ linkers would provide sufficient length for recognition of either 10- or 11- base pairs, but would be too short to span the 12-bp binding site. Polyamide **2** displays optimal recognition of the 11-bp binding site, 5'-TGTTATTGTTA-3', $K_a \geq 1 \times 10^{12} M^{-1}$. Replacing the δ linker in **2** with the two-carbon shorter β - residue in **1** results in a reduction of affinity at the 11-bp site by > 6-fold ($K_a = 1.5 \times 10^{10} M^{-1}$). The unfavorable binding affinities of **1** and **2** at the 12-bp site indicates that the covalent constraint of the linker subunit prevents alignment of hairpin subunits at this longer recognition sequence.

Conclusions

It might have been expected that tandem-hairpins would bind by a mechanism with one hairpin binding its 5-bp target site and the second hairpin providing non-specific binding

enhancement from van der Waals and electrostatic interactions. Surprisingly, a large affinity increase is observed only at the 11-bp target site, while affinity at isolated 5-bp hairpin sites does not increase substantially and in some cases decreases (Table 3.2). These results indicate that a simple aliphatic 5-carbon linker is sufficient to provide for synergistic tandem-hairpin binding affinity and specificity. Although it is remarkable that simple aliphatic linkers provide six orders of magnitude enhancement in DNA-binding affinity, this still remains substantially lower than the nine order of magnitude enhancement predicted for a 'perfect linker'. Therefore, although the structural elements reported here for tail-to-turn coupling of hairpin polyamides expand the binding site size targetable by the motif, the generality of the approach as well as second generation 'rigid linkers' will need to be explored and reported in due course

Experimental Procedure

General: Dicyclohexylcarbodiimide (DCC), Hydroxybenzotriazole (HOBT), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU) and 0.2 mmol/gram Boc- β -alanine-(4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc- β -Pam-Resin) was purchased from Peptides International (0.2 mmol/gram) (*R*)-2-Fmoc-4-Boc-diaminobutyric acid, (*S*)-2-Fmoc-4-Boc-diaminobutyric acid, and (*R*)-2-amino-4-Boc-diaminobutyric acid were from Bachem. *N,N*-diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF), *N*-methylpyrrolidone (NMP), DMSO/NMP, Acetic anhydride (Ac₂O), and 0.0002 M potassium cyanide/pyridine were purchased from Applied Biosystems. Dichloromethane (DCM) and triethylamine (TEA) were reagent grade from EM, thiophenol (PhSH), dimethylaminopropylamine (Dp), (*R*)- α -methoxy- α -(trifluoromethyl)phenylacetic acid ((*R*)MPTA) and (*S*)- α -methoxy- α -(trifluoromethyl)phenylacetic acid ((*S*)MPTA) were from Aldrich, trifluoroacetic acid

Table 2. Equilibrium Association Constants [M^{-1}] for binding of the parent hairpin to 5-bp half sites and tandem hairpin at 11-bp mismatch Sites.^[a-c]

Site	Hairpin Motif (H)	Tandem Hairpin Motif (T)	$K_a(T)/K_a(H)$	Specificity ^[d]
I	 $K_a = 3.8 \times 10^9 M^{-1}$	 $K_a \geq 1.0 \times 10^{12} M^{-1}$	≥ 260	$\geq 20,000$
II	 $K_a = 3.8 \times 10^9 M^{-1}$	 $K_a = 1.5 \times 10^{10} M^{-1}$	4	300
III	 $K_a = 3.8 \times 10^9 M^{-1}$	 $K_a = 1.0 \times 10^9 M^{-1}$.26	20
IV	 $K_a = 3.5 \times 10^7 M^{-1}$	 $K_a = 2.2 \times 10^8 M^{-1}$	6	4.4
V	 $K_a = 3 \times 10^8 M^{-1}$	 $K_a = 1 \times 10^8 M^{-1}$.33	2
VI	 $K_a = 1 \times 10^8 M^{-1}$	 $K_a = 5 \times 10^7 M^{-1}$.5	1

[a] The reported association constants are the average 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 MgCl₂ and 5 mM CaCl₂. Mismatch base pairings are in shaded regions and brackets enclose the binding site and half binding sites for the parent and tandem hairpins respectively. [c] Sites V - VI were less accurately measured because they were located in the compressed region of the sequencing gel. [d] Specificity is calculated as $K_a(\text{sites I - VI})/K_a(\text{site VI})$.

(TFA) Biograde from Halocarbon, phenol from Fisher, and ninhydrin from Pierce. All reagents were used without further purification.

Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. A shaker for manual solid phase synthesis was obtained from St. John Associates, Inc. Screw-cap glass peptide synthesis reaction vessels (5 mL and 20 mL) with a #2 sintered glass frit were made as described by Kent.^[16] ¹H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Optical rotations were recorded on a JASCO Dip 1000 Digital Polarimeter. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on either a HP 1090M analytical HPLC or a Beckman Gold system using a RAINEN C₁₈, Microsorb MV, 5µm, 300 x 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, gradient elution 1.25% acetonitrile/min. Preparatory reverse phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100 µm C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. 18MΩ water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 µm filtered.

ImPyPy-(R)[ImPyPy-(R)^{H₂N}γ-PyPyPy-β-]^{HN}γ-PyPyPy-β-Dp (1): ImPyPy-(R)^{FmocHN}γ-PyPyPy-β-Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc-β-Pam-Resin (0.6 mmol/g).^[12] (R)-2-Fmoc-4-Boc-diaminobutyric acid (0.7 mmol) was incorporated as previously described for Boc-γ-aminobutyric acid. ImPyPy-(R)^{FmocHN}γ-PyPyPy-β-Pam-Resin was placed in a glass 20 mL peptide synthesis vessel and treated with DMF (2 mL), followed by piperidine (8 mL)

and agitated (22 °C, 30 min.). ImPyPy-(R)^{H₂N}γ-PyPyPy-β-Pam-resin was isolated by filtration, and washed sequentially with an excess of DMF, DCM, MeOH, and ethyl ether and the amine-resin dried *in vacuo*. ImPyPy-(R)[ImPyPy-(R)^{FMocHN}γ-PyPyPy-β-]^{HN}γ-PyPyPy-β-Pam-Resin was then synthesized in a stepwise fashion by machine-assisted solid phase methods from ImPyPy-(R)^{H₂N}γ-PyPyPy-β-Pam-resin (0.38 mmol/g^[18]). ImPyPy-(R)[ImPyPy-(R)^{FMocHN}γ-PyPyPy-β-]^{HN}γ-PyPyPy-β-Pam-Resin was placed in a glass 20 mL peptide synthesis vessel and treated with DMF (2 mL), followed by piperidine (8 mL) and agitated (22 °C, 30 min.). ImPyPy-(R)[ImPyPy-(R)^{H₂N}γ-PyPyPy-β-]^{HN}γ-PyPyPy-β-Pam-Resin was isolated by filtration, and washed sequentially with an excess of DMF, DCM, MeOH, and ethyl ether and the amine-resin dried *in vacuo*. A sample of ImPyPy-(R)[ImPyPy-(R)^{H₂N}γ-PyPyPy-β-]^{HN}γ-PyPyPy-β-Pam-Resin (240 mg, 0.29 mmol/gram¹⁹) was treated with neat dimethylaminopropylamine (2 mL) and heated (55 °C) with periodic agitation for 16 h. The reaction mixture was then filtered to remove resin, 0.1% (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC. ImPyPy-(R)[ImPyPy-(R)^{H₂N}γ-PyPyPy-β-]^{HN}γ-PyPyPy-β-Dp is recovered upon lyophilization of the appropriate fractions as a white powder (28 mg, 22% recovery). $[\alpha]_D^{20} +14.6$ (c 0.05, H₂O); UV (H₂O) λ_{max} 246, 306 (100,000); ¹H NMR (300 MHz, [D₆]DMSO, 20 °C): δ = 10.54 (s, 1 H; aromatic NH); 10.45 (s, 1 H; aromatic NH); 10.44 (s, 1 H; aromatic NH); 10.02 (s, 1 H; aromatic NH); 9.95 (s, 1 H; aromatic NH); 9.92 (s, 1 H; aromatic NH); 9.90 (d, 2 H; aromatic NH); 9.86 (d, 2 H; aromatic NH); 9.2 (br s, 1 H; CF₃COOH); 8.25 (m, 4 H; aliphatic NH, NH₃); 8.11 (d, 1 H; J = 8.5 Hz, aliphatic NH); 8.04 (m, 4H, aliphatic NH), 7.37 (s, 2 H; CH); 7.25 (m, 2 H; CH); 7.22 (d, 1 H; CH); 7.18 (m, 2 H; CH); 7.16 (m, 3 H; CH); 7.12 (m, 4 H; CH); 7.02 (m, 4 H; CH); 6.95 (d, 1 H; J = 1.6 Hz; CH); 6.91 (d, 1 H; J = 1.5 Hz; CH); 6.88 (d, 1 H, J = 1.3 Hz; CH); 6.85 (m, 3 H; CH); 5.32 (t, 1 H; aliphatic CH), 4.45 (m, 1 H, aliphatic CH), 3.96 (s, 6 H; NCH₃); 3.83 (s, 3 H; NCH₃); 3.80 (s, 18 H; NCH₃); 3.79 (s, 3 H; NCH₃); 3.76 (s, 3 H; NCH₃); 3.39 (m, 4 H; CH₂); 3.28 (m, 2 H; CH₂); 3.15 (m, 4 H; CH₂); 3.07

(m, 2 H; CH₂); 2.97 (m, 2 H; CH₂); 2.70 (d, 6 H; N(CH₃)₂); 2.32 (m, 2 H; CH₂); 1.93 (m, 2 H; CH₂); 1.71 (m, 2 H; CH₂); 1.47 (m, 2 H; CH₂); 1.20 (m, 4 H; CH₂); MALDI-TOF-MS [M⁺-H] (monoisotopic), 1881.9: 1881.9 calc. for C₈₉H₁₀₉N₃₂O₁₆

ImPyPy-(R)[ImPyPy-(R)^{H₂N}γ-PyPyPy-δ-]^{H^N}γ-PyPyPy-β-Dp (2): ImPyPy-(R)[ImPyPy-(R)^{H₂N}γ-PyPyPy-δ-]^{H^N}γ-PyPyPy-β-Pam-Resin was prepared as described for ImPyPy-(R)[ImPyPy-(R)^{H₂N}γ-PyPyPy-β-]^{H^N}γ-PyPyPy-β-Pam-Resin. A sample of ImPyPy-(R)[ImPyPy-(R)^{H₂N}γ-PyPyPy-δ-]^{H^N}γ-PyPyPy-β-Pam-Resin (240 mg, 0.29 mmol/gram^[18]) was treated with neat dimethylaminopropylamine (2 mL) and heated (55 °C) with periodic agitation for 16 h. The reaction mixture was then filtered to remove resin, 0.1% (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC. ImPyPy-(R)[ImPyPy-(R)^{H₂N}γ-PyPyPy-δ-]^{H^N}γ-PyPyPy-β-Dp is recovered upon lyophilization of the appropriate fractions as a white powder (32 mg, 25% recovery). [α]_D²⁰ +14.6 (c 0.05, H₂O); UV (H₂O) λ_{max} 246, 306 (100,000); ¹H NMR (300 MHz, [D₆]DMSO, 20 °C): δ = 10.54 (s, 1 H; aromatic NH); 10.45 (s, 1 H; aromatic NH); 10.44 (s, 1 H; aromatic NH); 10.02 (s, 1 H; aromatic NH); 9.95 (s, 1 H; aromatic NH); 9.92 (s, 1 H; aromatic NH); 9.90 (d, 2 H; aromatic NH); 9.86 (d, 2 H; aromatic NH); 9.2 (br s, 1 H; CF₃COOH); 8.25 (m, 4 H; aliphatic NH, NH₃); 8.11 (d, 1 H; J = 8.5 Hz, aliphatic NH); 8.04 (m, 4H, aliphatic NH), 7.37 (s, 2 H; CH); 7.25 (m, 2 H; CH); 7.22 (d, 1 H; CH); 7.18 (m, 2 H; CH); 7.16 (m, 3 H; CH); 7.12 (m, 4 H; CH); 7.02 (m, 4 H; CH); 6.95 (d, 1 H; J = 1.6 Hz; CH); 6.91 (d, 1 H; J = 1.5 Hz; CH); 6.88 (d, 1 H, J = 1.3 Hz; CH); 6.85 (m, 3 H; CH); 5.32 (t, 1 H; aliphatic CH), 4.45 (m, 1 H, aliphatic CH), 3.96 (s, 6 H; NCH₃); 3.83 (s, 3 H; NCH₃); 3.80 (s, 18 H; NCH₃); 3.79 (s, 3 H; NCH₃); 3.76 (s, 3 H; NCH₃); 3.39 (m, 4 H; CH₂); 3.28 (m, 2 H; CH₂); 3.15 (m, 4 H; CH₂); 3.07 (m, 2 H; CH₂); 2.97 (m, 2 H; CH₂); 2.70 (d, 6 H; N(CH₃)₂); 2.32 (m, 2 H; CH₂); 1.93 (m, 2 H; CH₂); 1.71 (m, 2 H; CH₂); 1.47 (m, 2 H; CH₂); 1.20 (m, 4 H; CH₂); MALDI-TOF-MS [M⁺-H] (monoisotopic), 1910.2: 1909.9 calc. for C₉₁H₁₁₃N₃₂O₁₆

ImPyPy-(R)[ImPyPy-(R)^{EDTA} γ -PyPyPy- δ -]^{HN} γ -PyPyPy- β -Dp (2-E): Excess EDTA-dianhydride (50 mg) was dissolved in DMSO/NMP (1 mL) and DIEA (1 mL) by heating at 55 °C for 5 min. The dianhydride solution was added to ImPyPy-(R)[ImPyPy-(R)^{H₂N} γ -PyPyPy- δ -]^{HN} γ -PyPyPy- β -Dp (10 mg, 5 μ mol) dissolved in DMSO (750 μ L). The mixture was heated (55 °C, 25 min.) and the remaining EDTA-anhydride hydrolyzed (0.1M NaOH, 3 mL, 55 °C, 10 min.). Aqueous TFA (0.1% wt/v) was added to adjust the total volume to 8 mL and the solution purified directly by reversed phase HPLC to provide ImPyPy-(R)[ImPyPy-(R)^{EDTA} γ -PyPyPy- δ -]^{HN} γ -PyPyPy- β -Dp (2-E) as a white powder upon lyophilization of the appropriate fractions (2 mg, 20% recovery). MALDI-TOF-MS [M⁺-H] (monoisotopic), 2184.3: 2184.0 calc. for C₁₀₁H₁₂₇N₃₄O₂₃

DNA Reagents and Materials. Enzymes were purchased from Boehringer-Mannheim and used with their supplied buffers. Deoxyadenosine and thymidine 5'-[α -³²P] triphosphates were obtained from Amersham, deoxyadenosine 5'-[γ -³²P]triphosphate was purchased from I.C.N. Sonicated, and 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.^[17]

Construction of plasmid DNA. The plasmids pDH10, pDH11, and pDH12 were constructed by hybridization of the inserts listed in Figure 4. Each hybridized insert was ligated individually into linearized pUC19 *Bam*HI/*Hind*III plasmid 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 μ g/mL 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 duplex DNA.

Preparation of 3'- and 5'-End-Labeled Restriction Fragments. The plasmids pDH(11-12) were linearized with *EcoRI* and *BsrBI*, then treated with the Sequenase enzyme, deoxyadenosine 5'-[α -³²P]triphosphate and thymidine 5'-[α -³²P]triphosphate for 3' labeling. Alternatively, these plasmids were linearized with *EcoRI*, treated with calf alkaline phosphatase, and then 5' labeled with T4 polynucleotide kinase and deoxyadenosine 5'-[γ -³²P]triphosphate. The 5' labeled fragment was then digested with *BsrBI*. The labeled fragment (3' or 5') was loaded onto a 6% non-denaturing polyacrylamide gel, and the desired 147 base pair band was visualized by autoradiography and isolated. Chemical sequencing reactions were performed according to published methods.^[19]

MPE•Fe(II) Footprinting.^[13] 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 hours. A fresh 50 µM MPE•Fe(II) solution was prepared from 100 µL of a 100 µM MPE solution and 100 µL of a 100 µM ferrous ammonium sulfate ($\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$) solution. MPE•Fe(II) solution (5 µM) was added to the equilibrated DNA, and the reactions were allowed to equilibrate 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 6 min., and a

5 μL sample (~ 15 kcpm) was immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V.

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), 20 mM NaCl, 100 μM /base pair calf thymus DNA, and 20 kcpm 3'- or 5'-radiolabeled DNA. The solutions were allowed to equilibrate for 8 hours. A fresh solution of ferrous ammonium sulfate ($\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$) (10 μM) was added to the equilibrated DNA, and the reactions were allowed to equilibrate for 15 min. Cleavage was initiated by the addition of dithiothreitol (10 mM) and allowed to proceed for 30 min. Reactions were stopped by ethanol precipitation, resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer, denatured at 85 $^\circ\text{C}$ for 6 min., and the entire sample was immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V.

DNase I Footprinting.^[15] All reactions were carried out in a volume of 400 μL . We note explicitly that no carrier DNA was used in these reactions until after DNase I cleavage. 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 30 kcpm 3'-radiolabeled DNA. The solutions were allowed to equilibrate for a minimum of 12 hours at 22 $^\circ\text{C}$. 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 1.875 u/mL) and was allowed to proceed for 7 min at 22 $^\circ\text{C}$. The reactions were stopped by adding 50 μL of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/mL 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 $^\circ\text{C}$ for 6 min, and immediately

loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V for 1 hour. The gels were dried under vacuum at 80 °C, then quantitated using storage phosphor technology.

Equilibrium association constants were determined as previously described.^{11a} The data were analyzed by performing volume integrations of the 5'-TGTTAxTGTTA-3' and 5'-TGACAxTGACA-3 sites and a reference site. The apparent DNA target site saturation, θ_{app} , was calculated for each concentration of polyamide using the following equation:

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

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{ref}° correspond to those values for a DNase I control lane to which no polyamide has been added. The ($[L]_{tot}$, θ_{app}) data points were fit to a Langmuir binding isotherm (eq 2, $n=1$ for polyamides 1 and 2, by minimizing the difference between θ_{app} and θ_{fit} , using the modified Hill equation:

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

where $[L]_{tot}$ corresponds to the total polyamide concentration, K_a corresponds to the equilibrium association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software

(version 2.1, Abelbeck software) with K_a , θ_{\max} , and θ_{\min} as the adjustable parameters. All acceptable fits had a correlation coefficient of $R > 0.97$. At least three 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. The data were normalized using the following equation:

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

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 12-20 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.

References

1. (a) J. M. Gottesfeld, L. Nealy, J. W. Trauger, E. E. Baird, P. B. Dervan, *Nature* **1997**, *387*, 202. (b) L. A. Dickinson, R. J. Gulizia, J. W. Trauger, E. E. Baird, D. E. Mosier, J. M. Gottesfeld, P. B. Dervan *Proc. Natl. Acad. Sci. USA* *in press*.
2. (a) J. W. Trauger, E. E. Baird, P. B. Dervan, *Nature* **1996**, *382*, 559. (b) S. E. Swalley, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1997**, *119*, 6953. (c) J. M. Turner, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1997**, *119*, 7636. (d) J. W. Trauger, E. E. Baird, P. B. Dervan, *Angew. Chem.* **1998**, *37*, 1421. (e) J. M. Turner, S. E. Swalley, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1998**, *120*, 6219.
3. (a) S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, *Nature* **1998**, *391*, 468. (b) C. L. Kielkopf, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, D. C. Rees *Science* *submitted*.
4. (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. (d) M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc.* **1993**, *115*, 2572. (e) B. H. Geierstanger, M. Mrksich P. B. Dervan, D. E. Wemmer, *Science* **1994**, *266*, 646. (f) S. White, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1997**, *119*, 8756.
5. (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) S. White, E. E.

Baird, P. B. Dervan, *Biochemistry* **1996**, *35*, 12532. (d) X. Chen, Sundaralingham Ramakrishnan, *M. J. Mol. Biol.* **1997**, *267*, 1157.

6. C. L. Kielkopf, E. E. Baird, P. B. Dervan, D. C. Rees *Nature Struct. Biol.* **1998**, *5*, 104.
7. (a) J. J. Kelly, E. E. Baird, P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981. (b) J. W. Trauger, E. E. Baird, M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 6160. (c) B. H. Geierstanger, M. Mrksich, P. B. Dervan, D. E. Wemmer, *Nature Struct. Biol.* **1996**, *3*, 321. (d) S. E. Swalley, E. E. Baird, P. B. Dervan, *Chem. Eur. J.* **1997**, *3*, 1608. (e) J. W. Trauger, E. E. Baird, P. B. Dervan, *J. Am. Chem.* **1998**, *120*, 3534.
8. M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc.* **1993**, *115*, 9892. (b) T.J. Dwyer, B.H. Geierstanger, M. Mrksich, P. B. Dervan, D. E. Wemmer, *J. Am. Chem. Soc.* **1993**, *115*, 9900. (c) M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc.* **1994**, *116*, 3663. (d) Y.H. Chen, J.W. Lown, *J. Am. Chem. Soc.* **1994**, *116*, 6995. (e) N.H. Alsaïd, J.W. Lown, *Tett. Lett.* **1994**, *35*, 7577. (f) N. H. Alsaïd, J.W. Lown, *Synth. Comm.* **1995**, *25*, 1059. (g) Y.H. Chen, J.W. Lown, *Biophys. J.* **1995**, *68*, 2041. (h) Y.H. Chen, J.X. Liu, J.W. Lown, *Bioorg. Med. Chem. Lett.* **1995**, *5*, 2223. (i) Y.H. Chen, Y.W. Yang, J.W. Lown, *J. Biomol. Struct. Dyn.* **1996**, *14*, 341. (j) M.P. Singh, W.A. Wylie, J.W. Lown, *Mag. Res. Chem.* **1996**, *34*, S55. (k) W. A. Greenberg, E. E. Baird, P. B. Dervan, *Chem. Eur. J.* **1998**, *4*, 796.
9. (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, *J. Am. Chem. Soc.* **1996**, *118*, 6147. (c) M. E. Parks, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 6153. (d) J. W.

- Trauger, E. E. Baird, P. B. Dervan, *Chem. & Biol.* **1996**, *3*, 369. (e) S. E. Swalley, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 8198. (f) D. S. Pilch, N. A. Pokar, C. A. Gelfand, S. M. Law, K. J. Breslauer, E. E. Baird, P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306. (g) R. P. L. de Claire, B. H. Geierstanger M. Mrksich, P. B. Dervan, D. E. Wemmer, *J. Am. Chem. Soc.* **1997**, *119*, 7909. (h) S. White, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1997**, *119*, 8756. (i) S. White, E. E. Baird, P. B. Dervan, *Chem & Biol.* **1997**, *4*, 569.
10. D. H. Herman, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1998**, *120*, 1382.
11. For a review, see: D. E. Wemmer and P. B. Dervan, *Curr. Opin. Struct. Biol.* **1997**, *7*, 355.
12. E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 6141.
13. (a) M. W. Van Dyke, R. P. Hertzberg, P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5470. (b) M. W. Van Dyke, P. B. Dervan, *Science* **1984**, *225*, 1122.
14. (a) J. S. Taylor, P. G. Schultz, P. B. Dervan, *Tetrahedron* **1984**, *40*, 457. (b) P. B. Dervan, *Science* **1986**, *232*, 464
15. (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. U.S.A.* **1986**, *83*, 8462. (c) D. F. Senear, M. Brenowitz, M. A. Shea, G. K. Ackers, *Biochemistry* **1986**, *25*, 7344.
16. S.B.H. Kent, *Annu. Rev. Biochem.* **1988**, *57*, 957.

17. J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.
18. Resin substitution can be calculated as $L_{\text{new}}(\text{mmol/g}) = L_{\text{old}} / (1 + L_{\text{old}}(W_{\text{new}} - W_{\text{old}}) \times 10^{-3})$, where L is the loading (mmol of amine per gram of resin), and W is the weight (gmol^{-1}) of the growing polyamide attached to the resin. see: Barlos, K.; Chatzi, O.; Gatos, D.; Stravropoulos, G. *Int. J. Peptide Protein Res.* **1991**, 37, 513.
19. (a) B.L. Iverson, P. B. Dervan, *Nucl. Acids Res.* **1987**, 15, 7823. (b) A.M. Maxam, W.S. Gilbert, *Methods Enzymol.* **1980**, 65, 499.

Chapter 4

Cycle-Polyamide Motif for Recognition of the Minor Groove of DNA

Abstract: Motifs for covalent linkage of side-by-side complexes of pyrrole-imidazole (Py-Im) polyamides in the DNA minor groove provide for small molecules that specifically recognize predetermined sequences with subnanomolar affinity. Polyamide subunits linked by a turn-specific γ -aminobutyric acid (γ) residue form hairpin-polyamide structures. Selective amino-substitution of the prochiral α -position of the γ -turn residue relocates the cationic charge from the hairpin C-terminus. Here we report the synthesis of pyrrole-resin as well as a solid phase strategy for the preparation of cycle-polyamides. The DNA binding properties of two 8-ring cycle-polyamides were analyzed on a DNA restriction fragment containing 6-bp match and mismatch binding sites. Quantitative footprint titrations demonstrate that a cycle-polyamide of sequence composition $\text{cyclo}-(\gamma\text{-ImPyPyPy}-(\text{R})^{\text{H}2\text{N}}\gamma\text{-ImPyPyPy}-)$ binds a 5'-AGTACT-3' site with an equilibrium association constant $K_a = 7.6 \times 10^{10} \text{ M}^{-1}$, a 3600-fold enhancement relative to the unlinked homodimer $(\text{ImPyPyPy}-\beta\text{-Dp})_2 \bullet 5'\text{-AGTACT-3}'$, and an 8-fold enhancement relative to hairpin analog $\text{ImPyPyPy}-(\text{R})^{\text{H}2\text{N}}\gamma\text{-ImPyPyPy-C3-OH} \bullet 5'\text{-AGTACT-3}'$. Replacement of a single nitrogen atom with a C-H ($\text{Im} \rightarrow \text{Py}$) regulates affinity and specificity of the cycle-polyamide by two orders of magnitude. The results presented here suggest that addition of a chiral γ -turn combined with placement of a second γ -turn within the hairpin structure provides a cycle-polyamide motif with favorable DNA-binding properties.

Publication: Herman, Turner, Baird, Dervan J. Am. Chem. Soc. **1999**, 121, 1121-1129.

Introduction

Small molecules that target predetermined DNA sequences have the potential to control gene expression. Polyamides containing the three aromatic amino acids 3-hydroxyproline (Hp), imidazole (Im) and pyrrole (Py) are synthetic ligands that bind to predetermined DNA sequences with subnanomolar affinity.^{2,3} DNA recognition depends on a code of side-by-side amino acid pairings oriented N-C with respect to the 5'-3' direction of the DNA helix in the minor groove.^{2,9} An antiparallel 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.⁴ 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 four base pair code.^{3,7} The linker amino acid γ -aminobutyric acid (γ) connects polyamide subunits C \rightarrow N in a "hairpin motif" and these ligands bind to predetermined target sites with > 100-fold enhanced affinity relative to dimers.^{2,9} In both published and unpublished work, eight-ring hairpin-polyamides have been found to regulate transcription and permeate a variety of cell-types in culture.¹ *Because topology could potentially regulate cell-permeation properties*, discovery of new motifs for covalent linkage that provide polyamides with affinities and specificities comparable to naturally occurring DNA-binding proteins remains a high priority.

Design of Cycle-Polyamides. In a formal sense, addition of a second γ -turn at the C and N-termini of a hairpin-polyamide allows covalent closure to form a cycle.¹⁰ An initial report described a 6-ring (- γ -3- γ -3-) cycle-polyamide which bound to a 5-bp DNA sequence with higher affinity than a corresponding hairpin-polyamide;¹⁰ however, sequence-specificity versus mismatch DNA-sequences was extremely poor (\approx 3-fold), compared to 40-fold observed for the hairpin.¹¹ It was initially thought that the cycle restricted polyamide flexibility, limiting the available conformers to prevent formation of a specific recognition complex. It remained to be determined if γ -turn cycle-polyamides

could be designed that have comparable DNA binding properties to polyamide-hairpins.

Because of the labor intensive solution phase cycle-polyamide synthesis and the initial discouraging thermodynamics with regard to sequence-specificity, the cycle-polyamides have not been investigated further until this report. We describe here a pyrrole-resin which enables cycle-polyamide linear precursors to be synthesized by solid phase methods, reducing the synthetic effort from weeks to days. Two 8-ring (- γ -4- γ -4-) cycle-polyamides have been prepared for the studies described here. Two key design elements from the original 6-ring cycle have been altered. First, the number of ring pairings has been increased from 3 to 4. Polyamide-DNA binding affinity is predicted to increase as the number of consecutive ring pairings increases from 3 to 4.^{9a} In addition, the charge has been moved from a Py-N-methyl group to a γ -turn. Although the detailed effects of the placement of charge remain to be determined, substitution of the prochiral α -position of the γ -turn residue to provide (*R*)-2,4,-diaminobutyric acid has been previously found to yield chiral linked hairpins with enhanced DNA-binding sequence specificity and orientation preference.¹² Cycle-polyamides must be capped at the N-terminus in order to complete the cycle; however, N-terminal, acetylation has been found to reduce hairpin-binding specificity and orientation preference.^{4g,8b} Analogous to hairpin-polyamides, cycle-polyamides are capable of forming two mirror image folded structures, one of which is responsible for 5'-3', N-C match-DNA binding^{4g} (Figure 4.1). It is likely that the chiral amine group could offset the 'acetylation effect' by controlling cycle-polyamide binding orientation preference and hence binding-specificity.¹²

Here we report the DNA-binding affinities and sequence specificities of the eight-ring cycle- polyamides, cyclo-(γ -ImPyPyPy-(*R*)^{H₂N} γ -ImPyPyPy-) (3) and cyclo-(γ -ImPyPyPy-(*R*)^{H₂N} γ -PyPyPyPy-) (4) that differ by a single amino acid substitution (underlined), for their respective 6-bp match sites, 5'-AGTACT-3' and 5'-AGTAIT-3,

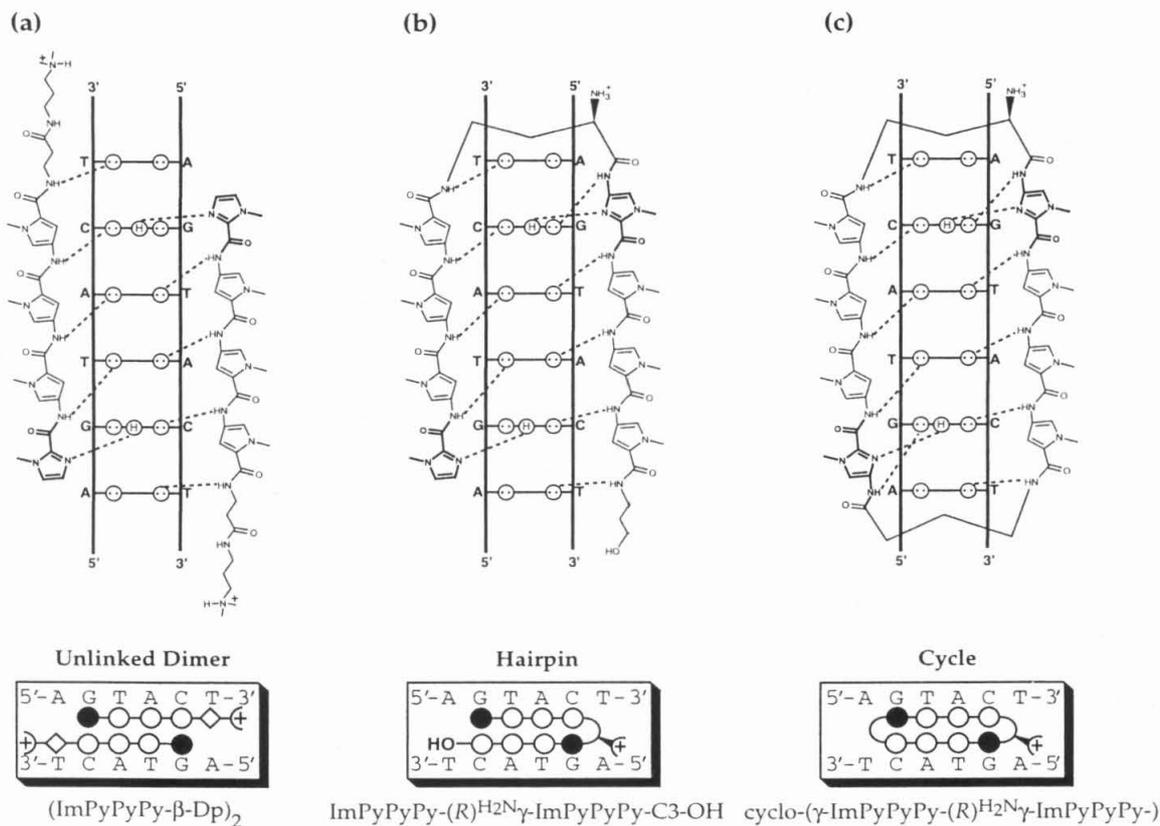


Figure 4.1 (Top a-c) Hydrogen bond models of the polyamide:DNA complexes formed between the 2:1 dimer ImPyPyPy- β -Dp (1), the 1:1 hairpin ImPyPyPy-(R)^{H₂N} γ -ImPyPyPy-OH (2), and the 1:1 cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -ImPyPyPy) (3) with the 6-bp 5'-AGTACT-3' match site. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of G. Putative hydrogen bonds are illustrated by dotted lines. (Bottom a-c) Schematic binding models for (1), (2) and (3). Im and Py rings are represented as shaded and unshaded spheres respectively.

which differ by a single base pair (underlined). In control experiments, the binding affinity and sequence-specificity of the ImPyPyPy- β -Dp (1) dimer and a hairpin analog ImPyPyPy-(R)^{H₂N}- γ -ImPyPyPy-C3-OH (2) were also studied (Figure 4.2). An EDTA analogue cyclo-(γ -ImPyPyPy-(R)^{EDTA•Fe(II)}- γ -PyPyPyPy-) (4-E•Fe(II)) was constructed to confirm the binding orientation of cycle-polyamide 4 at its 5'-AGTAIT-3' match and 5'-AGTACT-3' mismatch site. All polyamides were synthesized by solid phase methods,¹³ and their purity and identity confirmed by ¹H NMR, MALDI-TOF MS, and analytical HPLC. Precise binding site sizes were determined by MPE•Fe(II) footprinting,¹⁴ and binding orientation and stoichiometry confirmed by affinity cleaving experiments.¹⁵ Equilibrium association constants (K_a) of the polyamides for respective match and mismatch binding sites were determined by quantitative DNase I footprint titration.¹⁶

Results and Discussion

Resin Synthesis. The Py-Pam ester (5) was prepared according to the published procedures of Merrifield,¹⁷ with Boc-Py acid substituted for the standard Boc protected α -amino acid (Figure 4.3). The phenacyl ester (6) was selectively cleaved (Zn, AcOH) and the resultant acid (6) activated (DCC, HOBT) followed by reaction of the activated

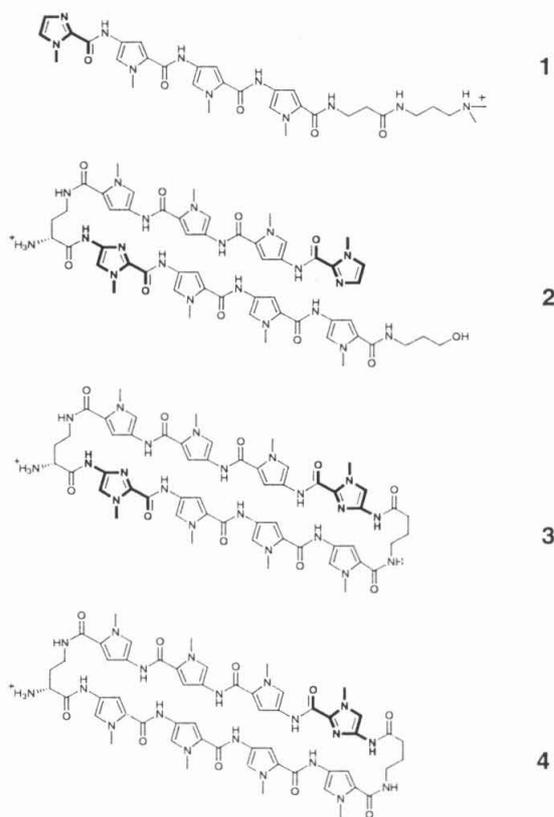


Figure 4.2 Structures of polyamides ImPyPyPy- β -Dp (1), ImPyPyPy-(R)^{H₂N}- γ -ImPyPyPy-C3-OH (2), and cyclo-(γ -ImPyPyPy-(R)^{H₂N}- γ -ImPyPyPy) (3), and cyclo-(γ -ImPyPyPy-(R)^{H₂N}- γ -PyPyPyPy) (4) as synthesized by solid phase methods.

ester with an excess of 0.7 mmol/g aminomethylated polystyrene for 24 hours (DIEA, DMF) to give Boc-Py-Pam-resin (7). Reactions were stopped at 0.1 mmol/g substitution as determined by quantitative ninhydrin analysis of free amine groups.¹⁸ Unreacted amine groups were capped by acetylation (Ac₂O, DIEA, DMF). Picric acid titration¹⁹ of Py-amino groups was used to verify resin loading of 0.1 mmol/g.

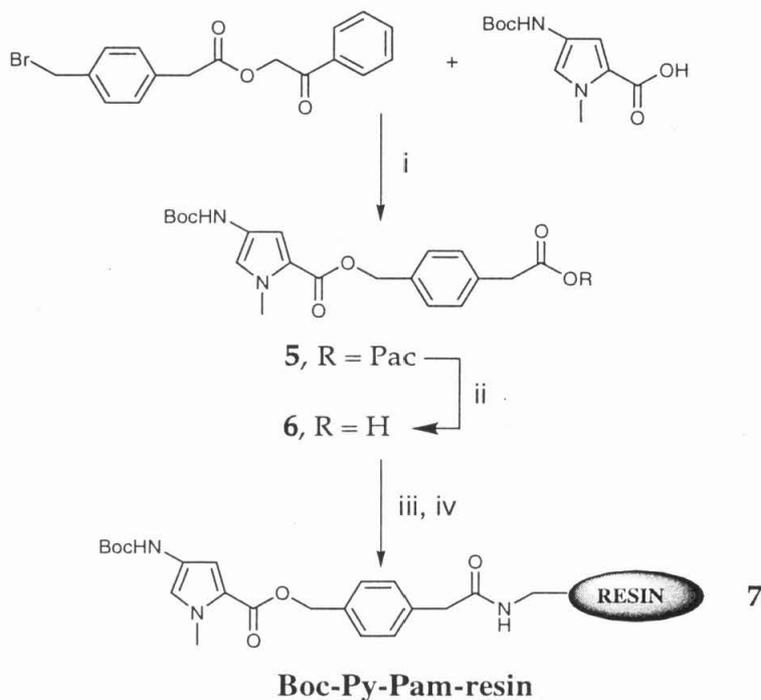


Figure 4.3 Synthesis of Boc-Py-PAM-resin (7). (i) K₂CO₃, DMF, (ii) Zn/AcOOH (iii) DCC/HOBT, DMF; (iv) aminomethylated-polystyrene); DIEA.

Synthesis of Hairpin Control 2. ImPyPyPy-(R)^{H₂N}-γ-ImPyPyPy-β-Pam-resin was synthesized by machine assisted protocols in 18 steps from commercially available Boc-β-Ala-Pam resin (Figure 4.4a).¹³ The polyamide was cleaved from the resin by a single step reduction with lithium borohydride (EtOH, 60 °C), followed by reversed phase HPLC purification to yield the hairpin-polyamide ImPyPyPy-(R)^{H₂N}-γ-ImPyPyPy-OH (2).

Cycle-Polyamide Synthesis. Two polyamide resins, Cbz γ -ImPyPyPy-(R)^{Fmoc} γ -PyPyPyPy-Pam-resin, and Cbz γ -ImPyPyPy-(R)^{Fmoc} γ -ImPyPyPy-Pam-resin were synthesized in 18 steps from Boc-Py-Pam-resin (600 mg of resin, 0.1 mmol/g of substitution) using Boc-chemistry machine-assisted protocols (Figure 4.4b).¹³ The (R)-2,4-diaminobutyric acid residue was introduced as an orthogonally protected *N*- α -Fmoc-*N*- γ -Boc derivative **10** (HBTU, DIEA). The final step was introduction of Cbz γ -Im acid **11** as a dimer block (HBTU, DIEA).¹³ Fmoc protected polyamide-resins Cbz γ -ImPyPyPy-(R)^{Fmoc} γ -PyPyPyPy-Pam-resin, and Cbz γ -ImPyPyPy-(R)^{Fmoc} γ -ImPyPyPy-Pam-resin were treated with 1:4 DMF:Piperidine (22 °C, 30 min.) to provide Cbz γ -ImPyPyPy-(R)^{H₂N} γ -PyPyPyPy-Pam-resin and Cbz γ -ImPyPyPy-(R)^{H₂N} γ -ImPyPyPy-Pam-resin, respectively. The amine-resins were then treated with Boc-anhydride (DIEA, DMF, 55 °C, 30 min.) providing Cbz γ -ImPyPyPy-(R)^{Boc} γ -PyPyPyPy-Pam-resin and Cbz γ -ImPyPyPy-(R)^{Boc} γ -ImPyPyPy-Pam-resin. A single step catalytic transfer hydrogenolysis was used to cleave the polyamide from the solid support and remove the Cbz protecting group from the N-terminal γ residue. A sample of the resin (240 mg) was treated with palladium acetate (2 ml DMF, 240 mg Pd(OAc)₂, 37 °C, 10 min.). Ammonium formate was added (500 mg, 8 hr) and the reaction mixture purified by reversed phase HPLC to provide H₂N- γ -ImPyPyPy-(R)^{Boc} γ -PyPyPyPy-COOH (**13**) and H₂N- γ -ImPyPyPy-(R)^{Boc} γ -ImPyPyPy-COOH (**12**). Cyclization of H₂N- γ -ImPyPyPy-(R)^{Boc} γ -PyPyPyPy-COOH (**13**) and H₂N- γ -ImPyPyPy-(R)^{Boc} γ -ImPyPyPy-COOH (**12**) was achieved with DPPA and potassium carbonate, as described previously.¹⁰ The Boc-protecting group was then removed *in situ* by treatment with neat TFA to yield the cyclic compounds cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -ImPyPyPy-) (**3**) and cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -PyPyPyPy-) (**4**) after subsequent purification by reversed phase HPLC. The cycle-polyamides are obtained with similar yield and purity and have similar solubility as their hairpin counterparts.

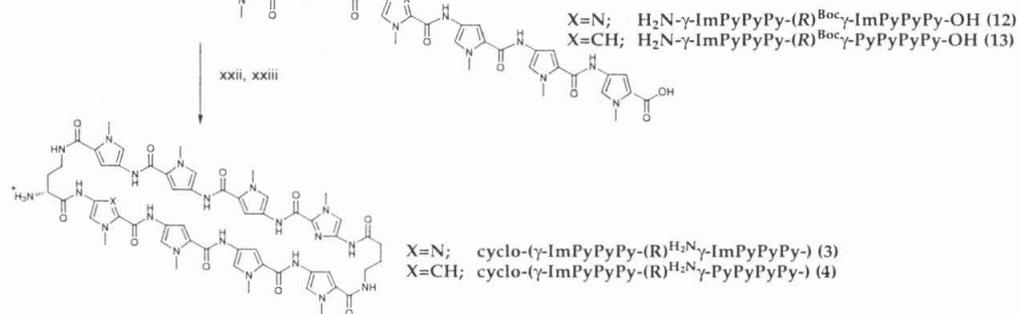
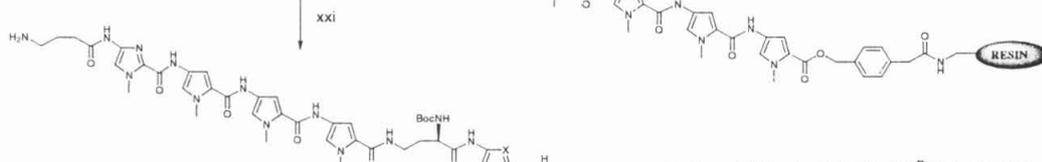
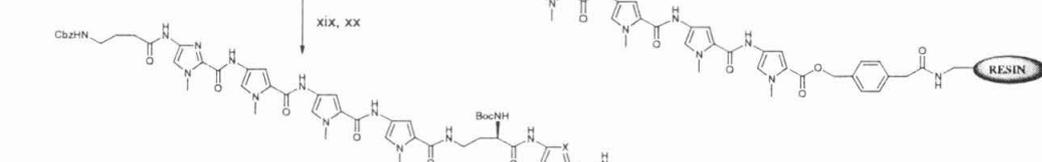
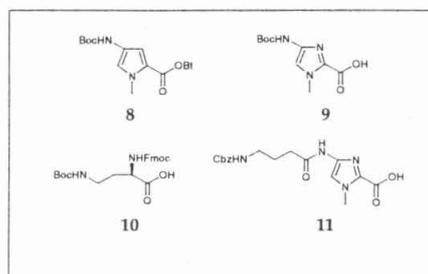
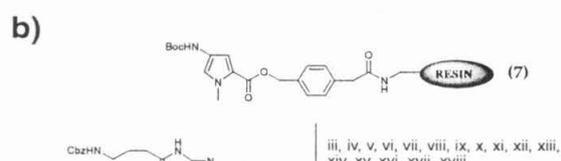
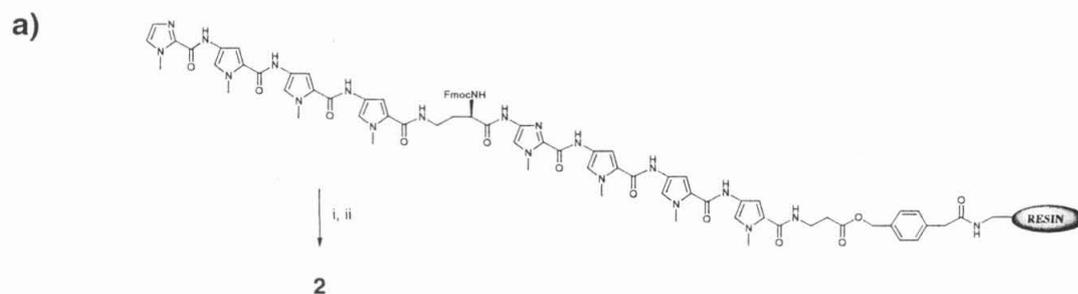


Figure 4.4 (Box) Monomers for synthesis of compounds described here; Boc-Py-OBt ester **8**, Boc-Im acid **9**, (*R*)-Fmoc- α -Boc- γ -diaminobutyric acid **10**, and Cbz- γ -Im acid **11**. (i) 1:4 DMF:Piperidine (22 °C, 30 min); (ii) LiBH₄, EtOH, reflux 16 hr. (iii) 80 % TFA/DCM, 0.4 M PhSH; (iv) Boc-Py-OBt, DIEA, DMF; (v) 80 % TFA/DCM, 0.4 M PhSH; (vi) Boc-Py-OBt, DIEA, DMF; (vii) 80 % TFA/DCM, 0.4 M PhSH; (viii) Boc-Im acid (DCC, HOBT) for **3**, Boc-Py-OBt, DIEA, DMF for **4**; (ix) 80 % TFA/DCM, 0.4 M PhSH; (x) (*R*)-Fmoc- α -Boc- γ -diaminobutyric acid (HBTU, DIEA); (xi) 80 % TFA/DCM, 0.4 M PhSH; (xii) Boc-Py-OBt, DIEA, DMF; (xiii) 80 % TFA/DCM, 0.4 M PhSH; (xiv) Boc-Py-OBt, DIEA, DMF; (xv) 80 % TFA/DCM, 0.4 M PhSH; (xvi) Boc-Py-OBt, DIEA, DMF; (xvii) 80 % TFA/DCM, 0.4 M PhSH; (xviii) Cbz- γ -Im acid (HBTU, DIEA); (xix) 80 % Piperidine:DMF (25 °C, 30 min); (xx) Boc anhydride, DIEA, DMF; (xxi) Pd(OAc)₂, HCO₂NH₄, DMF (37 °C, 8 hr); (xxii) DPPA, K₂CO₃ (xxiii) TFA (1 hr). (Inset) Pyrrole, Imidazole, and diaminobutyric acid monomers for solid phase synthesis: Boc-Pyrrole-OBt ester¹³ (Boc-Py-OBt) **8**, imidazole-2-Carboxylic acid^{2a} (Im-OH) **9**, (*R*)-Fmoc- α -Boc- γ -diaminobutyric acid **10**, and CBZ- γ -aminobutyric acid-imidazole dimer **11**.

Binding Site Size. MPE•Fe(II) footprinting¹⁴ on 3'- or 5'-³²P end-labeled 229 base pair restriction fragments reveals that each cycle-polyamide, at 10 nM concentration, binds to its designated 6-bp match sites (25 mM HEPES buffer (pH 7.3), 200 mM NaCl, 50 µg/ml glycogen, 5 mM DTT, 0.5 µM MPE•Fe(II), and 22 °C) (Figure 4.5 and 4.6). The polyamide cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -ImPyPyPy-) (3) which contains an Im/Py and a Py/Im pair protects the cognate 5'-AGTACT-3' match site. Binding of the single base pair mismatch site 5'-AGTATT-3' is only seen at much higher polyamide concentrations. The polyamide cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -PyPyPyPy-) (4) which contains a single Im/Py pair protects the designated match site 5'-AGTATT-3' and the single base pair mismatch site 5'-AGTACT-3'. The sizes of the asymmetrically 3'-shifted footprint cleavage patterns are consistent with 1:1 cycle-polyamide:DNA complex formation at 6-bp binding sites.

Binding Orientation. Affinity cleavage experiments¹⁵ using Cyclo-(γ -ImPyPyPy-(R)^{EDTA•Fe(II)} γ -PyPyPyPy-) (4-E•Fe(II)), which has an EDTA•Fe(II) moiety appended to the γ -turn, were used to confirm polyamide binding orientation and stoichiometry. For synthesis of the EDTA analogue, cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -PyPyPyPy-) (4) was treated with an excess of EDTA-dianhydride (DMSO/NMP, DIEA, 55 °C, 15 min) and the remaining anhydride was hydrolyzed (0.1 M NaOH, 55 °C, 10 min). Cyclo-(γ -ImPyPyPy-(R)^{EDTA} γ -PyPyPyPy-) (4-E•Fe(II)) was then isolated by reversed phase HPLC (Figure 6a). Affinity cleavage experiments were performed on the same 3'- or 5'-³²P end-labeled 229 base pair DNA restriction fragment from the plasmid pJT8 (20 mM HEPES buffer (pH 7.3), 200 mM NaCl, 50 µg/ml glycogen, 5 mM DTT, 1 µM Fe(II), pH 7.0 and 22 °C).

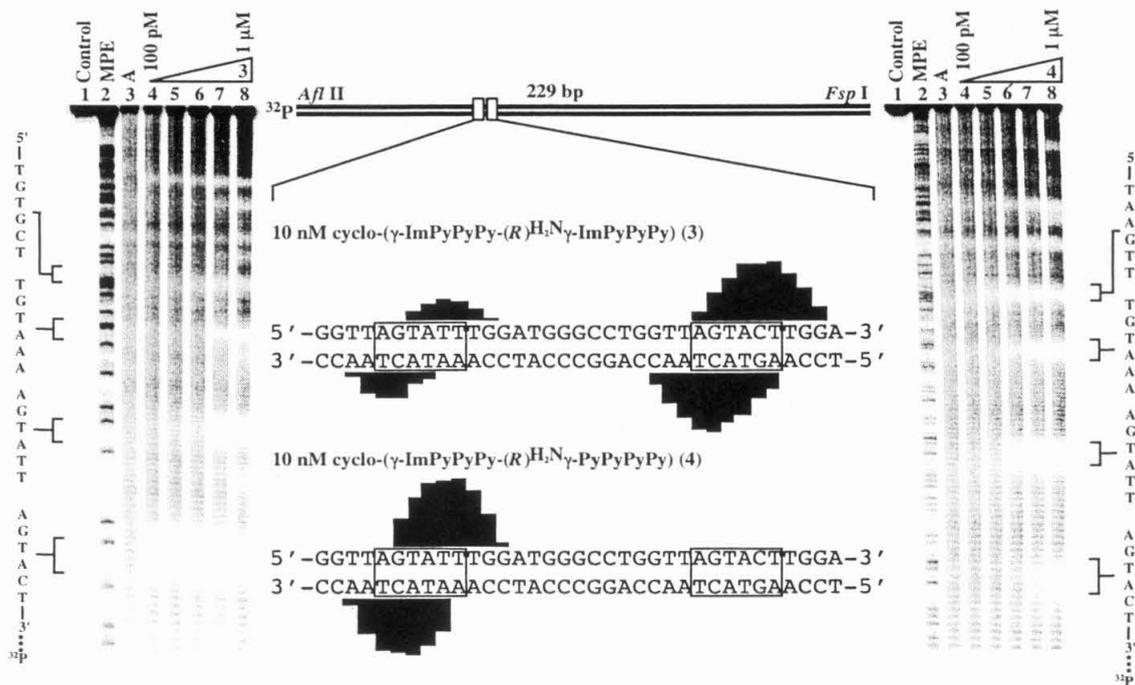


Figure 4.5 (Middle) Illustration of the 229-bp restriction fragment with the position of the sequence indicated. MPE•Fe(II) protection patterns of 10 nM cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -ImPyPyPy) (3) and 10 nM cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -PyPyPyPy) (4). Bar heights are proportional to the relative protection from cleavage at each band. Binding sites determined by MPE•Fe(II) footprinting are boxed. MPE•Fe(II) footprinting experiments of cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -ImPyPyPy) (3) (left) and cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -PyPyPyPy) (4) (right) on the 3' ³²P-labeled 229 bp restriction fragment of pJT8. 5'-AGTATT-3' and 5'-AGTACT-3' sites are shown adjacent to the autoradiograms. Additional 5'-TGTA AAA-3', 5'-TGTGCT-3', and 5'-TTAAGT-3' mismatch sites are not analyzed. Lane 1, intact; lane 2, A reaction; lane 3, MPE•Fe(II) standard; lanes 4-8, 100 pM, 1 nM, 10 nM, 100 nM and 1 μ M polyamide. All lanes contain 15 kcpm of either 3' or 5'-radiolabeled DNA, 25 mM HEPES buffer (pH 7.3), 200 mM NaCl, 50 mg/ml glycogen, 5 mM DTT, 0.5 mM MPE•Fe(II), and 22 °C.

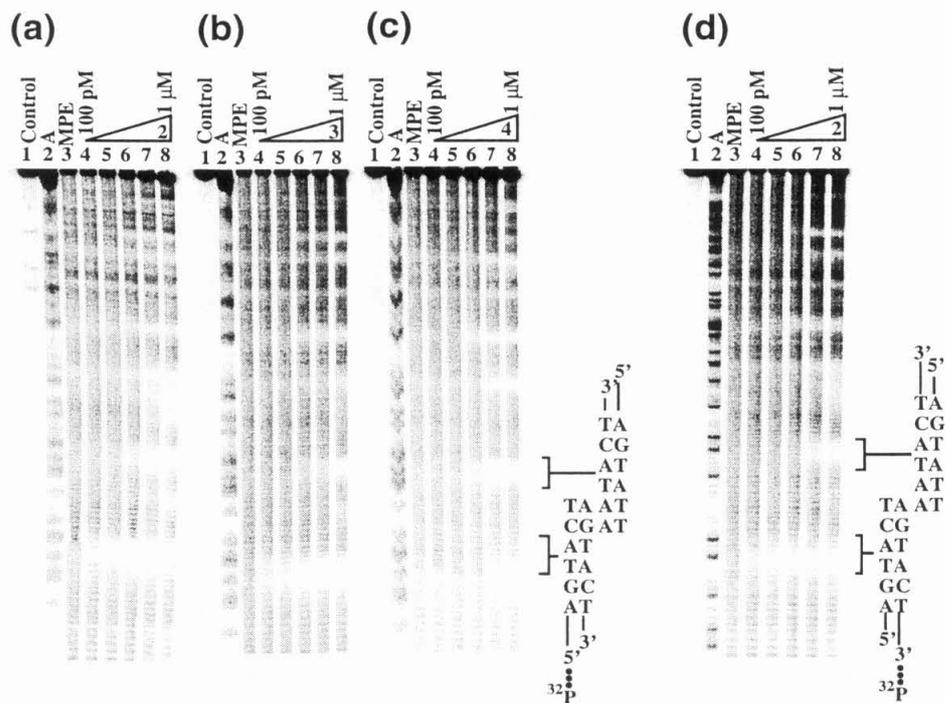


Figure 4.6 (a-c) MPE•Fe(II) footprinting experiments of ImPyPyPy-(R)^{H₂N}γ-ImPyPyPy-C3-OH (2), cyclo-(γ-ImPyPyPy-(R)^{H₂N}γ-ImPyPyPy) (3) and cyclo-(γ-ImPyPyPy-(R)^{H₂N}γ-PyPyPyPy) (4) on the 3'-³²P-labeled 229 bp restriction fragment of pJT8. 5'-AGTATT-3' and 5'-AGTACT-3' sites are shown adjacent to the autoradiograms. Additional 5'-TGATAA-3', 5'-TGTGCT-3', and 5'-TTAAGT-3' mismatch sites are not analyzed. Lane 1, intact; lane 2, A reaction; lane 3, MPE•Fe(II) standard; lanes 4-8, 100 pM, 1 nM, 10 nM, 100 nM and 1 μM polyamide. All lanes contain 15 kcpm of either 3' or 5'-radiolabeled DNA, 25 mM HEPES buffer (pH 7.3), 200 mM NaCl, 50 mg/ml glycogen, 5 mM DTT, 0.5 mM MPE•Fe(II), and 22 °C. (d) MPE•Fe(II) footprinting experiments of ImPyPyPy-(R)^{H₂N}γ-ImPyPyPy-C3-OH (2) on the 5'-³²P-labeled 229 bp restriction fragment of pJT8. Lane 1, intact; lane 2, A reaction; lane 3, MPE•Fe(II) standard; lanes 4-8, 100 pM, 1 nM, 10 nM, 100 nM and 1 μM polyamide.

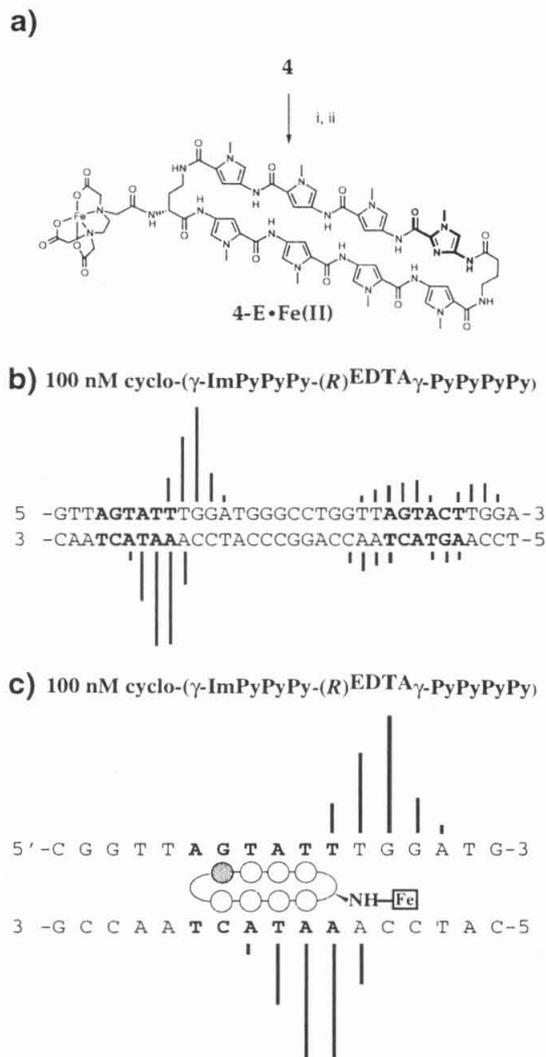


Figure 4.7 (a) Synthesis of **4-E•Fe(II)**: (i) EDTA dianhydride (DMSO/NMP, DIEA, 55 °C, 15 min); (ii) 0.1 M NaOH (55 °C, 10 min). (b) Affinity cleavage pattern for cyclo-(γ -ImPyPyPyPy-(*R*)^{EDTA•Fe(II)} γ -PyPyPyPyPy) (**4-E•Fe(II)**) at 100 nM concentration depicting a single binding orientation at the 5'-AGTATT-3' match site and no orientational preference at the 5'-AGTACT-3' mismatch site. (c) Ball-and-stick model of **4-E•Fe(II)**•5'AGTATT-3' complex. Bar heights are proportional to the relative cleavage intensities at each base pair. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. The boxed Fe denotes the EDTA•Fe(II) cleavage moiety. See supplemental material for autoradiograms.

The observed cleavage pattern for cyclo-(γ -ImPyPyPy-(R)^{EDTA•Fe(II)} γ -PyPyPyPy) (**4-E•Fe(II)**) (Figures 4.7b and 4.7c, supplemental material) are 3'-shifted, consistent with minor groove occupancy. In the presence of 100 nM **4-E•Fe(II)**, a major cleavage locus proximal to the 3' side of the 5'-AGTATT-3' match sequence is revealed, consistent with formation of an oriented 1:1 cycle-polyamide•DNA complex. At the same ligand concentration, minor cleavage loci located 3' and 5' adjacent to the single base pair mismatch 5'-AGTACT-3' site appear, consistent with dual binding orientations at this symmetrical binding site. The cyclo-polyamide binding model is further supported by the location of cleavage loci at the 5'-side of the 5'-AGTATT-3' match site (Figure 4.7c), and at the 5'- and 3'- sides of the 5'-AGTACT-3' mismatch site corresponding to the EDTA•Fe(II) moiety placement off the ((R)^{H₂N} γ)-turn residue.

Binding Energetics. Quantitative DNase I footprint titrations¹⁶ (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂, pH 7.0 and 22 °C) were performed to determine the equilibrium association constants (K_a) of ImPyPyPy- β -Dp (**1**), ImPyPyPy-(R)^{H₂N} γ -ImPyPyPy-OH (**2**), cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -ImPyPyPy) (**3**), and cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -PyPyPyPy) (**4**) for the 6-bp match and mismatch sites (Figure 4.8 and Table 4.1). Polyamide **1** binds the respective match and mismatch sites with apparent first order association constants (eq 2, $n = 2$) consistent with 2:1 dimer formation.²³ Hairpin-polyamide (**2**) and cycle-polyamides (**3**) and (**4**) bind their respective match and mismatch sequences with binding isotherms (eq 2, $n = 1$) consistent with binding in a 1:1 polyamide•DNA complex. Polyamides bind the 5'-AGTACT-3' site with decreasing affinity; match cycle (**3**) > match hairpin (**2**) > mismatch cycle (**4**) > match dimer (**1**). Polyamides bind the 5'-AGTATT-3' sequence with decreasing affinity; match cycle (**4**) > mismatch cycle (**3**) > mismatch hairpin (**2**) > mismatch dimer (**1**).

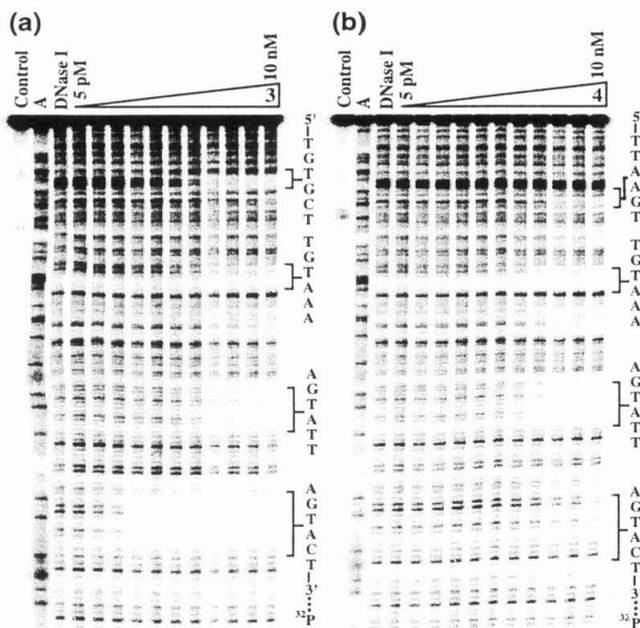


Figure 4.8 Quantitative DNase I footprint titration experiments with (a) cyclo-(γ -ImPyPyPy-(R)^{H₂N}- γ -ImPyPyPy) (3) and (b) cyclo-(γ -ImPyPyPy-(R)^{H₂N}- γ -PyPyPyPy) (4) on the 3'-end labeled 229-bp restriction fragment: lane 1, intact; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, and 10 nM. The 5'-AGTACT-3' and 5'-AGTATT-3' sites were analyzed and are shown on the right side of the autoradiogram. Additional sites not analyzed are 5'-TGTA AAA-3', 5'-TGTGCT-3', and 5'-TTAAGT-3'. All reactions contain 20 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂.

Table 1. Association Constants (M^{-1}) for polyamides **1-4**. ^{a-d}

Polyamide	Motif	5'-AGTACT-3'	5'-AGTATT-3'	Specificity	
	1	dimer	$K_a = 2.1 \times 10^7$	$K_a = 1.4 \times 10^6$	15
	2	hairpin	$K_a = 9.0 \times 10^9$	$K_a = 5.0 \times 10^8$	18
	3	cycle	$K_a = 7.6 \times 10^{10}$	$K_a = 1.3 \times 10^9$	55
	4	cycle	$K_a = 4.2 \times 10^8$	$K_a = 3.1 \times 10^9$	0.14

^aThe reported equilibrium association constants are the mean values obtained from three DNase I footprint titration experiments. ^bThe assays were carried out at 22 °C, pH 7.0 in the presence of 10 mM Tris·HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂. ^cApparent monomeric association constants were determined for polyamide homodimers.²³ ^dspecificity calculated as $K_a(5'-AGTACT-3')/K_a(5'-AGTATT-3')$.

Covalent coupling of dimer (**1**) to form hairpin-polyamide (**2**) results in a 428-fold increase in the DNA-binding affinity and comparable DNA-binding sequence-specificity. It is interesting to compare hairpin-polyamide (**2**), ImPyPyPy-(R)^{H₂N}γ-ImPyPyPy-OH, to the previously reported hairpin ImPyPyPy-γ-ImPyPyPy-β-Dp. Each hairpin contains eight aromatic rings and a single C-terminal charge located either on the γ-turn or a C-terminal β-Dp group. Although hairpin-polyamide (**2**) binds to DNA with affinity and specificity comparable to DNA-binding proteins, it does binds with 4-fold lower affinity and 5-lower sequence-specificity than the previously described hairpin ImPyPyPy-γ-ImPyPyPy-β-Dp. This probably results from loss of favorable interactions between the β-Dp group and A,T rich flanking sequences. Since cycle-polyamide have no C-terminal β-Dp group, hairpin-polyamide (**2**) is a more applicable control for the study described here.

On the basis of the pairing rules for polyamide•DNA complexes, the 5'-AGTACT-3' and 5'-AGTATT-3' sites represent 'match' and 'single base pair mismatch' sites for cycle-3, respectively, and 'single base pair mismatch' and 'match' sites for cycle-4, respectively. Cycle-polyamide (3), cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -PyPyPyPy), binds the 6-bp 5'-AGTACT-3' target sequence with an equilibrium association constant, $K_a = 7.6 \times 10^{10} \text{ M}^{-1}$, and 55-fold specificity over the single base pair mismatch 5'-AGTATT-3' site ($K_a = 1.3 \times 10^9 \text{ M}^{-1}$). These affinities represent a 3,600-fold increase relative to dimer (1) and an 8-fold enhancement relative to hairpin-polyamide (2). Furthermore, the affinity and specificity of (3) are comparable to the previously described hairpin ImPyPyPy- γ -ImPyPyPy- β -Dp. The cycle-polyamide (2), cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -PyPyPyPy), which contains a single Im/Py pair, preferentially binds the 5'-AGTATT-3' match site ($K_a = 3.1 \times 10^9 \text{ M}^{-1}$) versus the single base pair mismatch 5'-AGTACT-3' ($K_a = 4.2 \times 10^8 \text{ M}^{-1}$) with a 7-fold preference. Therefore, replacing a single pyrrole amino acid in cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -PyPyPyPy) (4) with an imidazole residue in cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -ImPyPyPy) (3) regulates cycle-polyamide specificity and affinity by 2 orders of magnitude.

Conclusions. A second generation cycle-polyamide motif has been designed and characterized. Two 8-ring (- γ -4- γ -4-) cycle-polyamides were found here to bind to DNA with affinity and specificity comparable to naturally occurring DNA-binding proteins. The polyamide-DNA binding affinity increases as expected as the number of consecutive ring pairings increases from 3 to 4.^{9a} Important key design factors likely contributed to the improved specificity of the eight-ring cycles compared with the original design (Figure 4.9).¹⁰ Moving the charge to the γ -turn may enhance the cycle-polyamide binding orientation preference and hence binding-specificity. Determination of the exact molecular basis for the optimization of the cycle-polyamides awaits further footprinting efforts as well as high resolution structure studies. However, it is clear from

comparison of first and second generation cycles that polyamide design for DNA recognition can be continually optimized using affinity and *specificity* as two key criteria in parallel with continued investment in synthetic methodology.

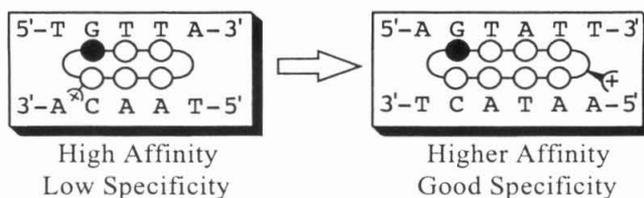


Figure 4.9 Comparison of six-ring and eight-ring cycles.

Experimental Procedure. Dicyclohexylcarbodiimide (DCC), Hydroxybenzotriazole (HOBT), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU); aminomethylated polystyrene, 4-(bromomethyl)phenylacetic Acid Phenacyl Ester, and 0.6 mmol/gram Boc- β -alanine-(4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc- β -Pam-Resin) was purchased from Peptides International (0.2 mmol/gram) (R)-2-Fmoc-4-Boc-diaminobutyric acid, and (R)-2-amino-4-Boc-diaminobutyric acid were from Bachem. N,N-diisopropylethylamine (DIEA), N,N-dimethylformamide (DMF), N-methylpyrrolidone (NMP), DMSO/NMP, Acetic anhydride (Ac₂O), and 0.0002 M potassium cyanide/pyridine were purchased from Applied Biosystems. Dichloromethane (DCM) and triethylamine (TEA) were reagent grade from EM; thiophenol (PhSH), and dimethylaminopropylamine (Dp) were from Aldrich; trifluoroacetic acid (TFA) Biograde from Halocarbon; phenol from Fisher; and ninhydrin from Pierce. All reagents were used without further purification. Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. A shaker for manual solid phase synthesis was obtained from St. John Associates, Inc. Screw-cap glass peptide synthesis reaction vessels (5 mL and 20 mL) with a #2 sintered glass frit were made as described by Kent.²⁰ ¹H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz with chemical shifts reported in parts per million

relative to residual solvent. UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Optical rotations were recorded on a JASCO Dip 1000 Digital Polarimeter. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on either a HP 1090M analytical HPLC or a Beckman Gold system using a RAINEN C18, Microsorb MV, 5 μ m, 300 x 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, gradient elution 1.25% acetonitrile/min. Preparatory reverse phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100 μ m C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. 18 Ω water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μ m filtered.

Resin Substitution. Resin substitution can be calculated as $L_{\text{new}}(\text{mmol/g}) = L_{\text{old}} / (1 + L_{\text{old}}(W_{\text{new}} - W_{\text{old}}) \times 10^{-3})$, where L is the loading (mmol of amine per gram of resin), and W is the weight (g mol⁻¹) of the growing polyamide attached to the resin.²¹

Control Hairpin ImPyPyPy-(R)^{H₂N} γ -ImPyPyPy-C3-OH (2). ImPyPyPy-(R)^{Fmoc} γ -ImPyPyPy- β -Pam-resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc- β -alanine-Pam-resin (0.6 mmol/g).¹³ (R)-2-Fmoc-4-Boc-diaminobutyric acid (0.7 mmol) was incorporated as previously described for Boc- γ -aminobutyric acid. ImPyPyPy-(R)^{Fmoc} γ -ImPyPyPy- β -Pam-resin was placed in a glass 20 mL peptide synthesis vessel and treated with DMF (2 mL), followed by piperidine (8 mL) and agitated (22 °C, 30 min). ImPyPyPy-(R)^{H₂N} γ -ImPyPyPy- β -Pam-resin was isolated by filtration, and washed sequentially with an excess of DMF, DCM, MeOH, and ethyl ether, and the amine-resin dried *in vacuo*. A sample of resin (200 mg 0.40 mmol/gram¹⁷) was suspended in absolute ethanol (25 ml). LiBH₄ (200 mg) was added,

and the mixture refluxed for 16 hr. The reaction mixture was then filtered to remove resin, neat TFA added (6 ml), and the resulting solution concentrated *in vacuo*, resuspended in 0.1% (wt/v) TFA (8 ml), and purified twice by reversed phase HPLC to provide the trifluoroacetate salt of ImPyPyPy-(R)^{H₂N}γ-ImPyPyPy-C3-OH (**2**) as a white powder upon lyophilization of the appropriate fractions. (2.5 mg, 3% recovery). ¹H NMR (DMSO-d₆) δ 11.03 (s, 1 H); 10.47 (s, 1 H); 10.13 (s, 1 H); 9.97 (s, 2 H); 9.94 (s, 1 H); 9.90 (s, 1 H); 8.35 (br s, 3 H); 8.29 (m, 1 H); 8.01 (m, 1 H); 7.39 (s, 2 H), 7.32 (s, 1 H), 7.26 (m, 2 H), 7.22 (s, 1 H), 7.17 (m, 2 H), 7.08 (s, 2 H), 7.05 (s, 1 H), 7.02 (s, 2 H); 6.93 (s, 1 H), 6.84 (s, 1 H); 5.32 (m, 1 H); 3.98 (s, 6 H); 3.85 (s, 3 H); 3.83 (m, 9 H); 3.80 (s, 3 H); 3.78 (s, 3 H); 2.73 (m, 2 H), 2.35 (m, 2 H); 1.95 (m, 4 H); 1.73 (m, 2 H); MALDI-TOF-MS (monoisotopic) [M+H] 1139.5 (1139.5 calc. for C₅₃H₆₃N₂₀O₁₀).

Boc-Pyrrolyl-4-(oxymethyl)phenylacetic Acid Phenacyl Ester (5). A solution of Boc-Py acid (6.9 g, 29 mmol), 4-(bromomethyl)phenylacetic Acid Phenacyl Ester (10 g, 29 mmol), and DIEA (7.2 ml, 41 mmol) in 60 ml DMF were stirred at 50 °C for 6 hr. The solution was cooled and partitioned between 400 ml water and 400 ml ethyl ether. The ether layer was washed sequentially (2 × 200 ml each) with 10% citric acid, brine, satd. NaHCO₃, and brine. The organic phase was dried (sodium sulfate) and concentrated *in vacuo*. The crude product was recrystallized from 3:1 ethyl acetate:hexanes to yield **5** as a fluffy white foam (6.1 g, 42% yield). TLC (2:3 hexanes/ethyl acetate v/v) R_f 0.6 ¹H NMR (DMSO-d₆) δ 9.07 (s, 1 H), 7.90 (d, 2 H, *J* = 7.6 Hz), 7.61 (t, 1 H, *J* = 7.2 Hz), 7.50 (t, 2 H, *J* = 7.5 Hz), 7.31 (m, 4 H), 7.06 (s, 1 H), 6.62 (s, 1 H), 5.47 (s, 2 H), 5.16 (s, 2 H), 3.80 (s, 2 H), 3.76 (s, 3 H), 1.42 (s, 9H) ¹³C NMR (DMSO-d₆) δ 193.1, 171.2, 160.6, 153.2, 135.7, 134.4, 130.1, 129.4, 128.5, 128.3, 123.7, 120.0, 119.0, 108.0, 79.0, 67.4, 65.1, 36.7, 28.6; FABMS *m/e* 506.205 (506.205 calc. for C₂₈H₃₀N₂O₇).

Boc-Pyrrolyl-4-(oxymethyl)phenylacetic Acid (6). Zinc dust was activated with 1M HCl (aq.) as described.¹⁷ Boc-Pyrrolyl-4-(oxymethyl)phenylacetic Acid Phenacyl Ester (3 g, 5.9 mmol) was dissolved in 90 ml 4:1 Acetic Acid:water (v/v). Zinc dust (9.6 g, 147 mmol) was added, and the reaction stirred for 18 hours at room temperature. The zinc was removed by filtration, and the reaction mixture partitioned between 200 ml ethyl ether and 200 ml water. The layers were separated, the aqueous layer extracted (ethyl ether, 1 x 200 ml), and the combined ether layers were washed (water, 5 x 100 ml), dried (sodium sulfate), concentrated *in vacuo*, and azeotroped (benzene, 6 x 100 ml). The crude acid product was purified by flash chromatography (2:1 hexanes :ethyl acetate) to yield a yellow oil (2.0 g, 86 % yield). TLC (ethyl acetate) R_f 0.7; ¹H NMR (DMSO-d₆) δ 9.06 (s, 1 H); 7.29 (d, 2 H, J = 7.8 Hz); 7.21 (d, 2 H, J = 7.8 Hz); 7.06 (s, 1 H); 6.6 (s, 1 H); 5.14 (s, 2 H); 3.74 (s, 3 H); 3.52 (s, 2 H); 1.38 (s, 9 H).

Boc-aminoacyl-Pyrrolyl-4-(oxymethyl)-Pam-resin (7). Boc-Pyrrolyl-4-(oxymethyl)phenylacetic Acid (6) (1 g, 2.57 mmol) was dissolved in 6.5 ml DMF. HOBt (382 mg, 2.8 mmol) followed by DCC (735 mg, 3.2 mmol) was added and the reaction mixture shaken for 4 hr at room temperature. The precipitated DCU was filtered and the reaction mixture added to 10 grams aminomethyl-polystyrene-resin (0.7 mmol/g substitution) previously swollen for 30 min in DMF. DIEA (1 ml) was added and the reaction shaken until the resin was determined by the ninhydrin test and picric acid titration to be approximately 0.1 mmol/g substituted. The resin was washed with DMF, and the remaining amine groups capped by acetylation (2x) with excess acetic anhydride capping solution (2:2:1 DMF:Ac₂O:DIEA). The resin was washed with DMF (1 x 20 ml), DCM (1 x 20 ml), and MeOH (1 x 20 ml) and dried *in vacuo*.

H₂N-γ-ImPyPyPy-(R)^{Boc}-γ-PyPyPyPy-COOH (13). Cbz-γ-ImPyPyPy-(R)^{Fmoc}-γ-PyPyPyPy-Pam-resin was synthesized in a stepwise fashion by machine-assisted solid phase

methods from Boc-Py-Pam-resin (0.1 mmol/g). (*R*)-2-Fmoc-4-Boc-diaminobutyric acid (0.7 mmol) was incorporated as previously described for Boc- γ -aminobutyric acid. Cbz γ -ImPyPyPy-(*R*)^{Fmoc} γ -PyPyPyPy-Pam-resin was placed in a glass 20 mL peptide synthesis vessel and treated with DMF (2 mL), followed by piperidine (8 mL) and agitated (22 °C, 30 min). Cbz γ -ImPyPyPy-(*R*)^{H₂N} γ -PyPyPyPy-Pam-resin was isolated by filtration, and washed sequentially with an excess of DMF, DCM, MeOH, and ethyl ether and the amine-resin dried *in vacuo*. Cbz γ -ImPyPyPy-(*R*)^{H₂N} γ -PyPyPyPy-Pam-resin was then treated with 300 mg Boc-anhydride (2 ml DMF, 1 ml DIEA, 55 °C, 30 min) to provide Cbz γ -ImPyPyPy-(*R*)^{Boc} γ -PyPyPyPy-Pam-resin. Palladium acetate (200 mg, 37 °C, 10 min) in 2 mL DMF was added to a sample of the Boc-resin (200 mg, 0.09 mmol/gram) ammonium formate (500 mg) was then added, and the reaction was allowed to shake for 8 hr at 37 °C. The reaction mixture was then filtered to remove resin and precipitated palladium metal, then diluted with 0.1% (wt/v) TFA (8 ml), and the resulting solution purified by reversed phase HPLC to yield the trifluoroacetate salt of H₂N- γ -ImPyPyPy-(*R*)^{Boc} γ -PyPyPyPy-COOH (17 mg, 66% recovery). ¹H NMR (DMSO-*d*₆) δ 10.35 (s, 1 H); 9.98 (s, 2 H); 9.96 (s, 3 H); 9.92 (s, 2 H); 7.98 (m, 1 H); 7.75 (br s 3 H); 7.44 (m, 1 H); 7.42 (s, 1 H); 7.30 (s, 1 H); 7.26 (s, 1 H); 7.25 (m, 2 H); 7.18 (m, 2 H); 7.15 (d, 1 H, *J* = 1.3 Hz); 7.13 (s, 1 H); 7.06 (m, 3 H); 6.96 (s, 1 H); 6.93 (s, 1 H); 6.90 (s, 1 H); 6.84 (d, 1 H, *J* = 1.8 Hz); 4.07 (q, 1 H, *J* = 6.0 Hz); 3.95 (s, 3 H); 3.84 (m, 15 H); 3.82 (s, 3 H); 3.79 (s, 3 H); 3.22 (m, 2 H); 2.81 (m, 2 H), 2.39 (t, 2 H, *J* = 6.9 Hz); 1.83 (m, 4 H); 1.39 (s, 9 H); MALDI-TOF-MS (monoisotopic) [M + H] 1281.5 (1281.6 calc. for C₆₀H₇₃N₂₀O₁₃).

H₂N- γ -ImPyPyPy-(*R*)^{Boc} γ -ImPyPyPy-COOH (12). Cbz- γ -ImPyPyPy-(*R*)^{Boc} γ -ImPyPyPy-Pam-resin was prepared as described for (13). A sample of the Boc-resin (330 mg, 0.09 mmol/gram) was treated with palladium acetate (330 mg, 37 °C, 10 min) followed by the addition of ammonium formate (500 mg, 37 °C, 8 hr). The reaction mixture was then

filtered to remove resin, diluted with 0.1% (wt/v) TFA (8 ml), and the resulting solution purified by reversed phase HPLC to yield $\text{H}_2\text{N-}\gamma\text{-ImPyPyPy-(R)}^{\text{Boc}}\text{-}\gamma\text{-ImPyPyPy-COOH}$ as the trifluoroacetate salt (4.2 mg, 10% recovery). $^1\text{H NMR}$ (DMSO-d_6) δ 10.36 (s, 1 H); 10.22 (s, 1 H); 10.14 (s, 1 H); 9.99 (s, 1 H); 9.97 (s, 2 H); 9.92 (s, 2 H); 7.97 (m, 1 H); 7.74 (br s 3 H); 7.45 (m, 1 H); 7.43 (s, 1 H); 7.42 (s, 1 H); 7.29 (s, 1 H); 7.26 (d, 1 H, $J = 1.7$ Hz); 7.23 (m, 2 H); 7.17 (s, 1 H); 7.14 (s, 1 H); 7.12 (s, 1 H); 7.06 (m, 2 H); 6.95 (s, 1 H); 6.90 (s, 1 H); 6.83 (d, 1 H, $J = 1.3$ Hz); 4.20 (m, 1 H); 3.94 (s, 6 H); 3.84 (m, 12 H); 3.81 (s, 3 H); 3.78 (s, 3 H); 3.22 (m, 2 H); 2.83 (quintet, 2 H, $J = 6.2$ Hz), 2.41 (t, 2 H, $J = 6.9$ Hz); 1.83 (m, 4 H); 1.38 (s, 9 H); MALDI-TOF-MS (monoisotopic) $[\text{M} + \text{H}]$ 1282.6 (1282.6 calc. for $\text{C}_{59}\text{H}_{72}\text{N}_{21}\text{O}_{13}$).

Cyclo-($\gamma\text{-ImPyPyPy-(R)}^{\text{H}_2\text{N}}\text{-}\gamma\text{-ImPyPyPy-}$) (3). The amine-polyamide **12** (2.8 mg, 2.0 μmol) was dissolved in DMF (7 ml), and treated with DPPA (12.5 μl) and K_2CO_3 (100 mg) for 3 hours. The reaction mixture was concentrated *in vacuo*, treated with TFA (3 ml, 1 hr), and purified by reversed phase HPLC to provide the trifluoroacetate salt of cyclo-($\gamma\text{-ImPyPyPy-(R)}^{\text{H}_2\text{N}}\text{-}\gamma\text{-ImPyPyPy-}$) (**3**). Cyclo-($\gamma\text{-ImPyPyPy-(R)}^{\text{H}_2\text{N}}\text{-}\gamma\text{-ImPyPyPy-}$) was recovered as a white powder upon lyophilization of the appropriate fractions (1.0 mg, 38% recovery). MALDI-TOF-MS (monoisotopic) $[\text{M} + \text{H}]$ 1164.5 (1164.5 calc. for $\text{C}_{54}\text{H}_{62}\text{N}_{21}\text{O}_{10}$).

Cyclo-($\gamma\text{-ImPyPyPy-(R)}^{\text{H}_2\text{N}}\text{-}\gamma\text{-PyPyPyPy-}$) (4). The amine-polyamide **13** (7 mg, 5.0 μmol) was dissolved in DMF (7 ml), and treated with DPPA (12.5 μl) and K_2CO_3 (100 mg) for 3 hours. The reaction mixture was concentrated *in vacuo*, treated with TFA (3 ml, 1 hr), diluted to 8 ml with 0.1% (wt/v) TFA, and purified by reversed phase HPLC to provide cyclo-($\gamma\text{-ImPyPyPy-(R)}^{\text{H}_2\text{N}}\text{-}\gamma\text{-PyPyPyPy-}$) (**4**). Cyclo-($\gamma\text{-ImPyPyPy-(R)}^{\text{H}_2\text{N}}\text{-}\gamma\text{-PyPyPyPy-}$) was recovered as a white powder upon lyophilization of the appropriate fractions as the trifluoroacetate salt (2.6 mg, 41% recovery). $^1\text{H NMR}$ (DMSO-d_6) δ 10.56 (s, 1 H); 10.26

(s, 1 H); 9.96 (s, 2 H); 9.95 (s, 3 H); 9.92 (s, 1 H); 8.27 (m, 4 H); 9.00 (m, 1 H); 7.45 (s, 1 H); 7.39 (s, 1 H); 7.38 (s, 1 H); 7.34 (s, 1 H); 7.29 (s, 1 H); 7.26 (s, 1 H); 7.25 (s, 1 H); 7.18 (s, 1 H); 7.08 (s, 1 H); 6.93 (m, 1 H); 6.91 (s, 1 H); 6.89 (s, 1 H); 6.85 (s, 2 H); 6.82 (s, 1 H); 5.31 (m, 1 H); 3.93 (s, 3 H); 3.85 (s, 3 H), 3.83 (m, 9 H); 3.80 (s, 3 H); 3.78 (s, 3 H); 3.76 (s, 3 H); 3.23 (m, 2 H); 2.72 (m, 2 H); 2.36 (m, 2 H); 1.97 (m, 4 H); 1.38 (s, 9 H); MALDI-TOF-MS (monoisotopic) [M + H] 1163.4 (1163.5 calc. for C₅₅H₆₃N₂₀O₁₀).

Cyclo-(γ -ImPyPyPy-(R)^{EDTA•Fe(II)} γ -PyPyPyPy-) (4-E•Fe(II)). Excess EDTA-dianhydride (180 mg) was dissolved in 1:1 DMSO/NMP (1 mL) and DIEA (1 mL) by heating at 60 °C for 10 min. The dianhydride solution was added to cyclo(γ -ImPyPyPy-(R)^{H₂N} γ -PyPyPyPy-) (**4**) (1.3 mg, 1.0 μ mol) dissolved in DMSO (1 ml). The mixture was heated (60 °C, 2 hr) and the remaining EDTA-anhydride hydrolyzed with 0.1M NaOH (2 mL, 60 °C, 15 min). Aqueous TFA (0.1% wt/v) was added to adjust the total volume to 8 mL and the solution purified directly by reversed phase HPLC to provide to cyclo-(γ -ImPyPyPy-(R)^{EDTA•Fe(II)} γ -PyPyPyPy-) (**4-E•Fe(II)**) as a white powder upon lyophilization of the appropriate fractions (0.25 mg, 18% recovery). MALDI-TOF-MS (monoisotopic) [M+H] 1438.1 (1437.6 calc. for C₆₅H₇₇N₂₂O₁₇).

DNA Reagents and Materials. Enzymes were purchased from Boehringer-Mannheim and used with their supplied buffers. Deoxyadenosine and thymidine 5'-[α -³²P] triphosphates were obtained from Amersham, deoxyadenosine 5'-[γ -³²P]triphosphate was purchased from I.C.N. Sonicated, and 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.²²

Preparation of 3'- and 5'-End-Labeled Restriction Fragments. The plasmid pJT8 was constructed as previously reported. pJT8 was linearized with *AflIII* and *FspI* restriction enzymes, then treated with the Sequenase enzyme, deoxyadenosine 5'- α - ^{32}P]triphosphate and thymidine 5'-[α - ^{32}P]triphosphate for 3' labeling. Alternatively, these plasmids were linearized with *AflIII*, treated with calf alkaline phosphatase, and then 5' labeled with T4 polynucleotide kinase and deoxyadenosine 5'-[γ - ^{32}P]triphosphate. The 5' labeled fragment was then digested with *FspI*. The labeled fragment (3' or 5') was loaded onto a 6% non-denaturing polyacrylamide gel, and the desired 229 base pair band was visualized by autoradiography and isolated.

MPE•Fe(II) Footprinting.¹⁴ All reactions were carried out in a volume of 400 μL . A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were: 20 mM HEPES 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 hr. A fresh 50 μM MPE•Fe(II) solution was prepared from 100 μL of a 100 μM MPE solution and 100 μL of a 100 μM ferrous ammonium sulfate ($\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$) solution. MPE•Fe(II) solution (5 μM) was added to the equilibrated DNA, and the reactions were allowed to equilibrate for 5 minutes. 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 $^\circ\text{C}$ for 6 min, and a 5 μL sample (~ 15 kcpm) was immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V.

Affinity Cleaving.¹⁵ All reactions were carried out in a volume of 400 μL . A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were: 20 mM HEPES buffer (pH 7.0), 20 mM NaCl, 100 μM /base

pair calf thymus DNA, and 20 kcpm 3'- or 5'-radiolabeled DNA. The solutions were allowed to equilibrate for 8 hours. A fresh solution of ferrous ammonium sulfate ($\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$) (10 μM) was added to the equilibrated DNA, and the reactions were allowed to equilibrate for 15 minutes. Cleavage was initiated by the addition of dithiothreitol (10 mM) and allowed to proceed for 30 min. Reactions were stopped by ethanol precipitation, resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer, denatured at 85 °C for 6 min, and the entire sample was immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V.

DNase I Footprinting.¹⁶ All reactions were carried out in a volume of 400 μL . We note explicitly that no carrier DNA was used in these reactions until after DNase I cleavage. 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 30 kcpm 3'-radiolabeled DNA. The solutions were allowed to equilibrate for a minimum of 12 hours 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 1.875 u/mL) and was allowed to proceed for 7 min at 22 °C. The reactions were stopped by adding 50 μL of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/mL 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 6 min, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V for 1 hour. The gels were dried under vacuum at 80 °C, then quantitated using storage phosphor technology. Equilibrium association constants were determined as previously described.¹²

References:

1. (a) Gottesfield, J. M.; Nealy, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1997**, *387*, 202-205. (b) Dickenson; L. A., Guzilia; P. Trauger, J. W.; Baird, E. E.; Mosier, D. M.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA*. **1998**, *95*, 12890.
2. For subnanomolar binding see: (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. Chemie. Int. Ed. Eng.* **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. For Hp/Py pairings see: (a) White, S. E.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468-471. (b) Kielkopf, C. L.; White, S. E.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *391*, 468-471.
4. For specificity of Im/Py pairings see: (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., USA* **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. (g) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756-8765.

5. For Py/Py pairing see: (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **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.
6. For Im/Im pairing see: (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. For H-pin motif see: (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.
8. For hairpin motif see: (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. USA* **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) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.*, submitted.

9. For polyamide dimers see: (a) Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **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.
10. For cycle-polyamides see: Cho, J. Y.; Parks, M. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10389.
11. For a review, see: Wemmer, D. E. and Dervan, P. B. *Curr. Opin. Struct. Biol.* **1997**, *7*, 355.
12. Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382.
13. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.
14. (a) Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5470. (b) Van Dyke, M. W.; Dervan, P. B. *Science* **1984**, *225*, 1122.
15. (a) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457. (b) Dervan, P. B. *Science* **1986**, *232*, 464
16. (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.
17. Mitchell, A. R.; Kent, S. B. H.; Engelhard, M.; Merrifield, R. B. *J. Org. Chem.* **1978**, *43*, 2845.
18. Sarin, V. K.; Kent, S. B. H.; Tam, J. P.; Merrifield, R. B. *Anal. Biochem.* **1981**, *117*, 147.
19. Gisin, B. F. *Anal. Chim. Acta.* **1972**, *58*, 248.
20. Kent, S. B. H.; *Annu. Rev. Biochem.* **1988**, *57*, 957.
21. Barlos, K.; Chatzi, O.; Gatos, D.; Stravropoulos, G. *Int. J. Peptide Protein Res.* **1991**, *37*, 513.
22. Sambrook, J.; Fritsch, E.F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.
23. 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 and Company, New York, NY, **1980**. p. 863.

Chapter 5

**Discrimination of A/T Sequences in the Minor Groove
of DNA within a Cyclic Polyamide Motif**

Abstract: Eight-ring cyclic polyamides containing pyrrole (Py), imidazole (Im), and hydroxypyrrole (Hp) aromatic amino acids recognize predetermined 6 base pair sites in the minor groove of DNA. Two four-ring polyamide subunits linked by a turn-specific (R)-2,4-diaminobutyric acid ((R)^{H₂N}γ) residue form hairpin polyamide structures with enhanced DNA binding properties. In hairpin polyamides, substitution of Hp/Py for Py/Py pairings enhances selectivity but compromises binding affinity for specific sequences. In an effort to enhance the binding properties of polyamides containing Hp/Py pairings, four eight-ring cyclic polyamides were synthesized and analyzed on a DNA restriction fragment containing three 6-bp sites 5'-tAGNNCTt-3', where N = AA, TA, or AT. Quantitative footprint titration experiments demonstrate that contiguous placement of Hp/Py pairs in cyclo-(γImPyPyPy-(R)^{H₂N}γImHpHpPy-) (4) provides a 20-fold increase in affinity for the 5'-tAGAACTt-3' site ($K_a = 7.5 \times 10^7 \text{ M}^{-1}$) relative to ImPyPyPy-(R)^{H₂N}γImHpHpPy-C3-OH (3). A cycle polyamide of sequence composition cyclo-(γImHpPyPy-(R)^{H₂N}γImHpPyPy-) (6) binds a 5'-tAGTACTt-3' site with an equilibrium association constant $K_a = 3.2 \times 10^9 \text{ M}^{-1}$, representing a 5-fold increase relative to the hairpin analog ImHpPyPy-(R)^{H₂N}γImHpPyPy-C3-OH (5). Arrangement of Hp/Py pairs in a 3'-stagger regulates specificity of cyclo-(γImPyHpPy-(R)^{H₂N}γImPyHpPy-) (8) for the 5'-tAGATCTt-3' site ($K_a = 7.5 \times 10^7 \text{ M}^{-1}$), representing a 3-fold increase in affinity relative to the hairpin analog ImPyHpPy-(R)^{H₂N}γImPyHpPy-C3-OH (7), respectively. This study identifies cyclic polyamides as a viable strategy for restoring recognition properties of polyamides containing Hp/Py pairs.

Publication: Melander, Herman & Dervan *Chem. Eur. j.* **2000**, *6*, 4487-4497.

Introduction

Polyamides containing the three aromatic amino acids pyrrole (Py), imidazole (Im), and 3-hydroxypyrrole (Hp) are cell permeable synthetic ligands^[1] which recognize predetermined sequences of DNA at subnanomolar concentrations and may be useful for gene regulation studies.^[2] DNA recognition depends on a code of side-by-side aromatic amino acid pairings that are oriented N-C with respect to the 5'-3' direction of the DNA helix in the minor groove. An antiparallel pairing of Im opposite Py (Im/Py pair) distinguishes G•C from C•G, pairing of Py opposite Im (Py/Im) distinguishes C•G from G•C and both of these from A•T/T•A base pairs.^[2] A Py/Py pair binds both A•T and T•A in preference to G•C/C•G.^[2] A Hp/Py pair specifies T•A from A•T, while Py/Hp targets A•T in preference to T•A and both of these from G•C/C•G, completing recognition of the four Watson-Crick base pairs.^[2]

In parallel with elucidation of the scope and limitations of the polyamide pairing rules, efforts have been made to increase DNA binding affinity and sequence specificity by covalently linking polyamide subunits. A hairpin polyamide motif with γ -aminobutyric acid (γ) serving as a turn-specific internal-guide-residue provides specific binding to designated target sites with > 100-fold enhanced affinity relative to the unlinked subunits.^[2] Hairpins have the important feature that ring pairings are set in place unambiguously as compared to homodimers which can afford "slipped motifs".^[4] Replacement of the γ -turn residue with the chiral subunit (R)-2,4-diaminobutyric acid ($((R)^{H_2N}\gamma)$) enhances hairpin DNA-binding affinity, sequence specificity and orientational preference.^[2] Further modification of hairpin polyamides by covalently tethering the N- and C- termini with a second γ -turn provides a cyclic polyamide motif that recognizes target sequences with increased affinity and specificity.^[2]

Hp/Py Hairpin Polyamides: High resolution X-ray diffraction data reveals that the T•A selectivity of the Hp/Py pair arises from a combination of (i) shape selection of an asymmetric cleft of the floor of the minor groove by the O2 of thymine and C2-H of adenine and (ii) specific hydrogen bonds between the 3-hydroxy and 4-carboximido groups of Hp with the O2 of T (Figure 5.1).^[3b] The gain in specificity, however, is accompanied by an energetic penalty. Replacement of a single Py/Py pair with a Hp/Py pair results in a 5-fold destabilization of an eight-ring hairpin for an identical match site.^[3a-c] Surprisingly, addition of multiple Hp/Py pairs for Py/Py pairs in a 10-ring hairpin polyamide results in a modest 10-fold reduction in affinity for the same site, illustrating that loss in binding affinity is not always additive as the number of Hp/Py substitutions is increased.^[3d] Crystallographic data reveals that Hp/Py polyamides bind undistorted B-form DNA, however, a localized 0.5 Å melting of the T•A Watson-Crick base pair is observed when the ImHpPyPy-Dp dimer is bound to 5'-AGTACT-3' and is potentially responsible for the energetic destabilization of the Hp/Py pair relative to the Py/Py pair.^[3b] A subsequent crystallographic study of an (ImPyHpPy-Dp)₂•5'-AGATCT-3' complex did not reveal such a melting and it is uncertain at this time the extent base pair melting contributes to binding destabilization.^[3e] In either case, reduction of affinity is observed and thus provides impetus to elucidate structural elements that restore polyamide binding properties to those comparable to naturally occurring DNA binding proteins without loss of specificity.

Cycle Motif: As a test case we consider discrimination of the three six base pair sequence 5'-WGNNCW-3', where NN = AA, TA and AT, which are bound by parent compounds ImPyPyPy-(R)^{H₂N}γ-ImPyPyPy-C3-OH (1) and cyclo-(γ-ImPyPyPy-(R)^{H₂N}γ-ImPyPyPy-) (2) containing central Py/Py pairs with *high affinity but modest discrimination* of the core AA, TA, and AT sequences as expected by Py/Py degenerate recognition of A•T/T•A base pairs. Replacement of the two central Py/Py pairs with

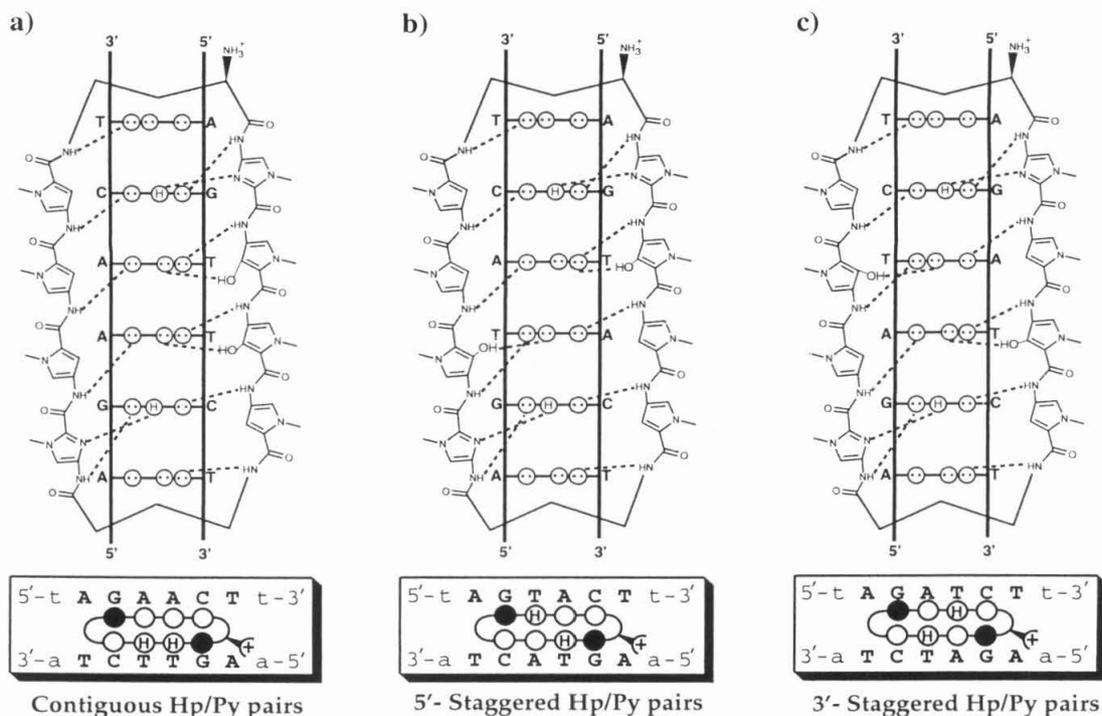


Figure 5.1 (Top a-c) Hydrogen bond models of the 1:1 polyamide-DNA complexes formed between cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -ImHpHpPy-) (4), cyclo-(γ -ImHpPyPy-(R)^{H₂N} γ -ImHpPyPy-) (6), and cyclo-(γ -ImPyHpPy-(R)^{H₂N} γ -ImPyHpPy-) (8) and their respective six base pair match sites 5'-tAGAACTt-3', 5'-tAGTACTt-3', and 5'-tAGTACTt-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 a-c) For schematic binding model, Im and Py rings are represented as shaded and unshaded spheres respectively, while Hp rings are annotated with an "H" in the center of an unshaded sphere.

Hp/Py pairs in varying spatial arrangements within the hairpin and cycle motifs allows us to compare the magnitude of the energetic penalty versus the gain or loss in sequence specificity for each motif. To determine if cyclization is a viable strategy for restoring the DNA binding properties of multiple Hp/Py pairings, three eight ring hydroxypyrrrole hairpins and their cyclic analogs were synthesized by solid phase methods,^[5] ImPyPyPy-(R)^{H₂N} γ -ImHpHpPy-C3-OH (3), cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -ImHpHpPy-) (4), ImHpPyPy-(R)^{H₂N} γ -ImHpPyPy-C3-OH (5), cyclo-(γ -ImHpPyPy-(R)^{H₂N} γ -ImHpPyPy-) (6),

ImPyHpPy-(R)^{H₂N}γ-ImPyHpPy-C3-OH (7), and cyclo-(γ-ImPyHpPy-(R)^{H₂N}γ-ImPyHpPy) (8) (Figure 5.2). Equilibrium association constants (K_a) for the eight polyamides were determined by quantitative DNase I footprint titration^[6] experiments on a DNA fragment containing three six base pair binding sites 5'-AGAACT-3', 5'-AGTACT-3, and 5'-AGATCT-3'. Based on the pairing rules, polyamides 3 and 4 with contiguous Hp residues bind 5'-AGAACT-3' as a match site and 5'-AGATCT-3' and 5'-AGTACT-3' as single base pair mismatch sites. Polyamides 5 and 6 which provide Hp residues staggered towards the 5'-direction relative to the DNA helix, target 5'-AGTACT-3' as a match, 5'-AGAACT-3' as a single base pair mismatch, and 5'-AGATCT-3' as a double base pair mismatch. Compounds 7 and 8, containing a 3'-stagger of Hp residues, preferentially target 5'-AGATCT-3', relative to the single and double base pair mismatch sites, 5'-AGAACT-3' and 5'-AGTACT-3', respectively. It might have been anticipated that the flexible hairpin can slide for optimal ligand:minor groove contacts, and that the constraints of preorganized cycles would not be beneficial for the Hp/Py pair:T•A recognition. Remarkably our findings are just the opposite and form the basis for our report here.

Results and Discussion

Hydroxypyrrole Hairpin Synthesis: Polyamides 1 and 2 were prepared as previously described.^[7a] Three polyamide resins, ImPyPyPy-(R)^{Fmoc}γ-ImOpOpPy-β-Pam resin, ImOpPyPy-(R)^{Fmoc}γ-ImOpPy-β-Pam resin, and ImPyOpPy-(R)^{Fmoc}γ-ImPyOpPy-β-Pam resin were synthesized from commercially available Boc-β-alanine-Pam resin (0.5 g resin, 0.25 mmol/g subst.) using manual, solid phase protocols in 18 steps (Figure 5.3).^[5] The Fmoc protecting group was then removed by treatment with (4:1) piperidine/DMF (22 °C, 30 min) to provide ImPyPyPy-(R)^{H₂N}γ-ImOpOpPy-β-Pam, ImOpPyPy-(R)^{H₂N}γ-ImOpPyPy-β-Pam, and ImPyOpPy-(R)^{H₂N}γ-ImPyOpPy-β-Pam resins. The polyamide

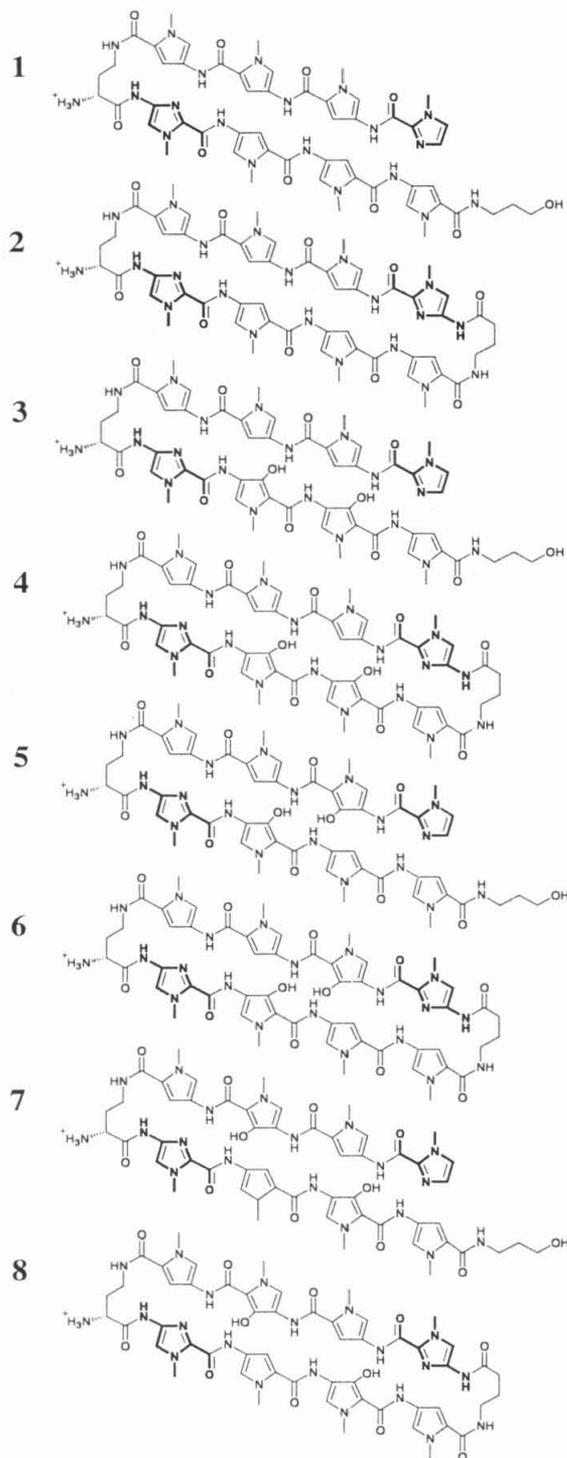


Figure 5.2 Structures of the eight-ring polyamides ImPyPyPy-(R)^{H₂N}γImPyPyPy-C3-OH (1), cyclo-(γ-ImPyPyPy-(R)^{H₂N}γ-ImPyPyPy-) (2), ImPyPyPy-(R)^{H₂N}γ-ImHpHpPy-C3-OH (3), cyclo-(γ-ImPyPyPy-(R)^{H₂N}γ-ImHpHpPy-) (4), ImHpPyPy-(R)^{H₂N}γ-ImHpPyPy-C3-OH (5), cyclo-(γ-ImHpPyPy-(R)^{H₂N}γ-ImHpPyPy-) (6), ImPyHpPy-(R)^{H₂N}γ-ImPyHpPy-C3-OH (7), and cyclo-(γ-ImPyHpPy-(R)^{H₂N}γ-ImPyHpPy-) (8) as synthesized by solid phase methods.

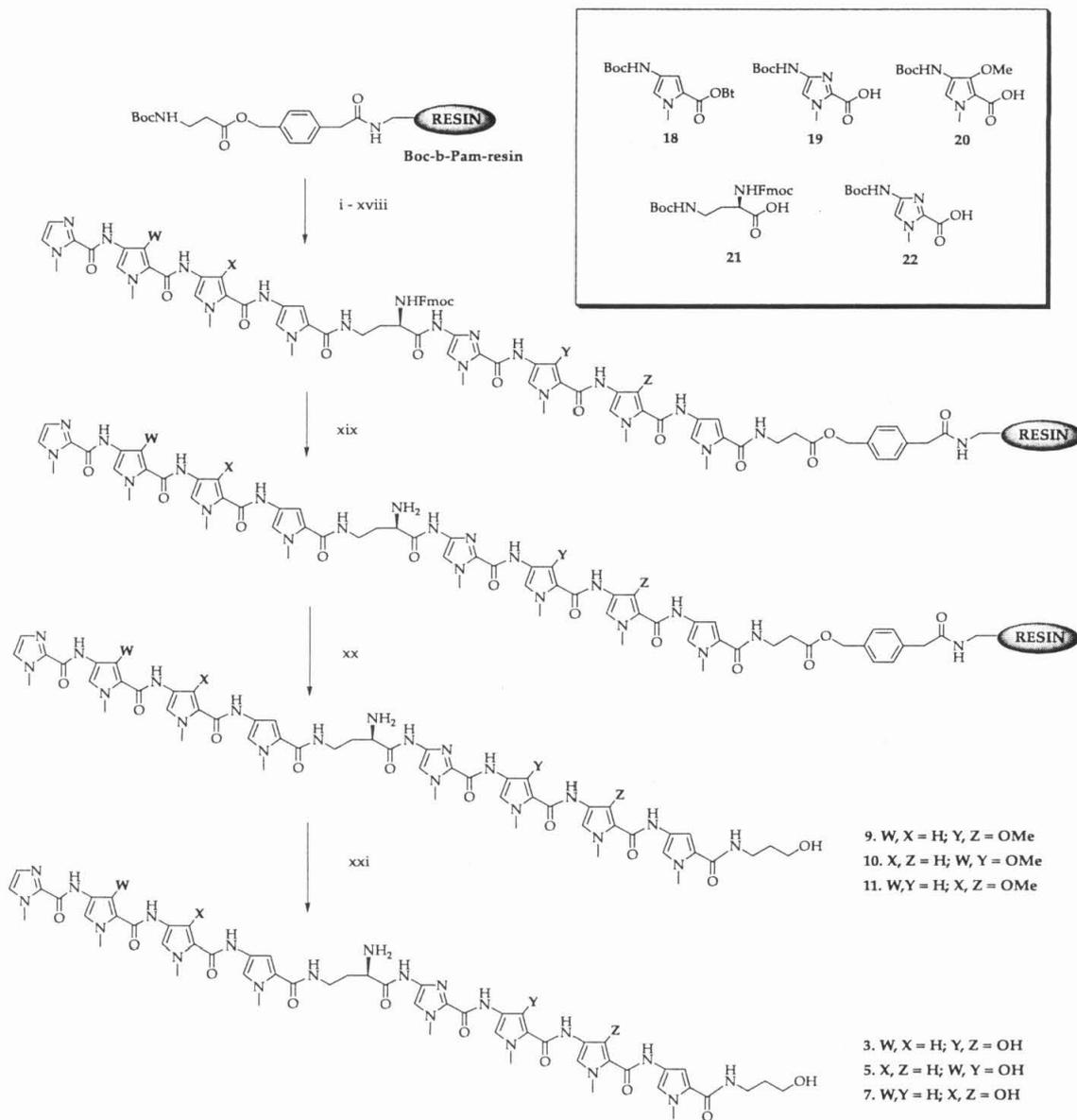


Figure 5.3 Hairpin synthetic scheme. ImPyPyPy-(*R*)^{H₂N}γImHpHpPy-C3-OH (**3**): (i) 80 % TFA/DCM, 0.4 M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) 80 % TFA/DCM, 0.4 M PhSH; (iv) Boc-Op-OH, HBTU, DIEA, DMF; (v) 80 % TFA/DCM, 0.4 M PhSH; (vi) Boc-Op-OH, HBTU, DIEA, DMF; (vii) 80 % TFA/DCM, 0.4 M PhSH; (viii) Boc-Im-OH, DIEA, HBTU, DMF; (ix) 80 % TFA/DCM, 0.4 M PhSH; (x) Fmoc-α-Boc-γ-diaminobutyric acid (HBTU, DIEA); (xi) 80 % TFA/DCM, 0.4 M PhSH; (xii) Boc-Py-OBt, DIEA, DMF; (xiii) 80 % TFA/DCM, 0.4 M PhSH; (xiv) Boc-Py-OBt, DIEA, DMF; (xv) 80 % TFA/DCM, 0.4 M PhSH; (xvi) Boc-Py-OBt, DIEA, DMF; (xvii) 80 % TFA/DCM, 0.4 M PhSH; (xviii) imidazole-2-carboxylic acid (HBTU/DIEA); (xix) 80 % Piperidine:DMF (25 °C, 30 min) (xx) NaBH₄/THF; (xxi) Sodium thiophenoxide/DMF. ImHpPyPy-(*R*)^{H₂N}γImHpPyPy-C3-OH (**5**): (i) - (iii), (v) - (xiii), (xv) - (xxi) same as above; (iv) Boc-Py-OH, HBTU, DIEA, DMF; (xiv) Boc-Op-OH, HBTU, DIEA, DMF. ImPyHpPy-(*R*)^{H₂N}γImPyHpPy-C3-OH (**7**): (i) - (v), (vii) - (xiii), (xiv) - (xxi) same as above; (vi) Boc-Py-OBt, DIEA, DMF; (xiv) Boc-Op-OH, HBTU, DIEA, DMF. (Inset) Py, Im, Op, and diaminobutyric acid monomers for solid phase synthesis: Boc-Pyrrole-OBt ester^[12] (Boc-Py-OBt) **18**, Boc-Imidazole-OH **19**, Boc-Op-OH **20**, (*R*)-Fmoc-α-Boc-γ-diaminobutyric acid **21**, Imidazole-2-Carboxylic acid^[4a] (Im-OH) **22**.

was then cleaved from resin by a single-step reductive cleavage reaction with NaBH_4/THF ($60\text{ }^\circ\text{C}$, 6 hrs.).^[8] The reaction mixture was subsequently purified by reversed phase HPLC to provide the methoxy-protected analogs $\text{ImPyPyPy-(R)}^{\text{H}_2\text{N}\gamma}$ - ImOpOpPy-C3-OH (9), $\text{ImOpPyPy-(R)}^{\text{H}_2\text{N}\gamma}$ - ImOpPyPy-C3-OH (10), and $\text{ImPyOpPy-(R)}^{\text{H}_2\text{N}\gamma}$ - ImPyOpPy-C3-OH (11). Deprotection with sodium thiophenoxide/DMF ($100\text{ }^\circ\text{C}$, 2 hrs)^[3e] and purification by reversed phase HPLC yields polyamides $\text{ImPyPyPy-(R)}^{\text{H}_2\text{N}\gamma}$ - ImHpHpPy-C3-OH (3), $\text{ImHpPyPy-(R)}^{\text{H}_2\text{N}\gamma}$ - ImHpPyPy-C3-OH (5), and $\text{ImPyOpPy-(R)}^{\text{H}_2\text{N}\gamma}$ - ImPyOpPy-C3-OH (7).

Cyclo Hydroxypyrrole Polyamide Synthesis: Polyamide 2 was synthesized as previously described.^[7a] Three polyamide resins, $\text{Boc}\gamma\text{-ImPyPyPy-(R)}^{\text{Fmoc}\gamma}$ - ImOpOpPy-Pam resin, $\text{Boc}\gamma\text{-ImOpPyPy-(R)}^{\text{Fmoc}\gamma}$ - ImOpPyPy-Pam resin, and $\text{Boc}\gamma\text{-ImPyOpPy-(R)}^{\text{Fmoc}\gamma}$ - ImPyOpPy-Pam resin, were synthesized in 18 steps from Boc-Py-Pam resin (1.0 g resin, 0.5 mmol/g substitution) using Boc-chemistry and manual solid phase synthesis protocols (Figure 5.4).^[5,7a] The (R) -2,4-diaminobutyric acid residue was introduced as an orthogonally protected N - α -Fmoc- N - γ -Boc derivative (HBTU, DIEA). Boc protected resins $\text{Boc}\gamma\text{-ImPyPyPy-(R)}^{\text{Fmoc}\gamma}$ - ImOpOpPy-Pam resin, $\text{Boc}\gamma\text{-ImOpPyPy-(R)}^{\text{Fmoc}\gamma}$ - ImOpPyPy-Pam resin, and $\text{Boc}\gamma\text{-ImPyOpPy-(R)}^{\text{Fmoc}\gamma}$ - ImPyOpPy-Pam resin, were treated with $\text{TFA}/\text{CH}_2\text{Cl}_2/\text{thiophenol}$ ($22\text{ }^\circ\text{C}$, 30 min.) to provide $\text{H}_2\text{N}\gamma\text{-ImPyPyPy-(R)}^{\text{Fmoc}\gamma}$ - ImOpOpPy-Pam resin, $\text{H}_2\text{N}\gamma\text{-ImOpPyPy-(R)}^{\text{Fmoc}\gamma}$ - ImOpPyPy-Pam resin, and $\text{H}_2\text{N}\gamma\text{-ImPyOpPy-(R)}^{\text{Fmoc}\gamma}$ - ImPyOpPy-Pam resin. Following protection as the benzyl carbamate using CBZ-OSuc (DMF, $22\text{ }^\circ\text{C}$, 90 min.), $\text{Cbz}\gamma\text{-ImPyPyPy-(R)}^{\text{Fmoc}\gamma}$ - ImOpOpPy-Pam resin, $\text{Cbz}\gamma\text{-ImOpPyPy-(R)}^{\text{Fmoc}\gamma}$ - ImOpPyPy-Pam resin, and $\text{Cbz}\gamma\text{-ImPyOpPy-(R)}^{\text{Fmoc}\gamma}$ - ImPyOpPy-Pam resin, were treated with 4:1 piperidine:DMF ($22\text{ }^\circ\text{C}$, 30 min.) to provide $\text{Cbz}\gamma\text{-ImPyPyPy-(R)}^{\text{H}_2\text{N}\gamma}$ - ImOpOpPy-Pam , $\text{Cbz}\gamma\text{-ImOpPyPy-(R)}^{\text{H}_2\text{N}\gamma}$ - ImOpPyPy-Pam , and $\text{Cbz}\gamma\text{-ImPyOpPy-(R)}^{\text{H}_2\text{N}\gamma}$ - ImPyOpPy-Pam resins, respectively. The resulting amine resins were then treated with Boc-anhydride (DMF, DIEA, $22\text{ }^\circ\text{C}$, 50 min.) producing

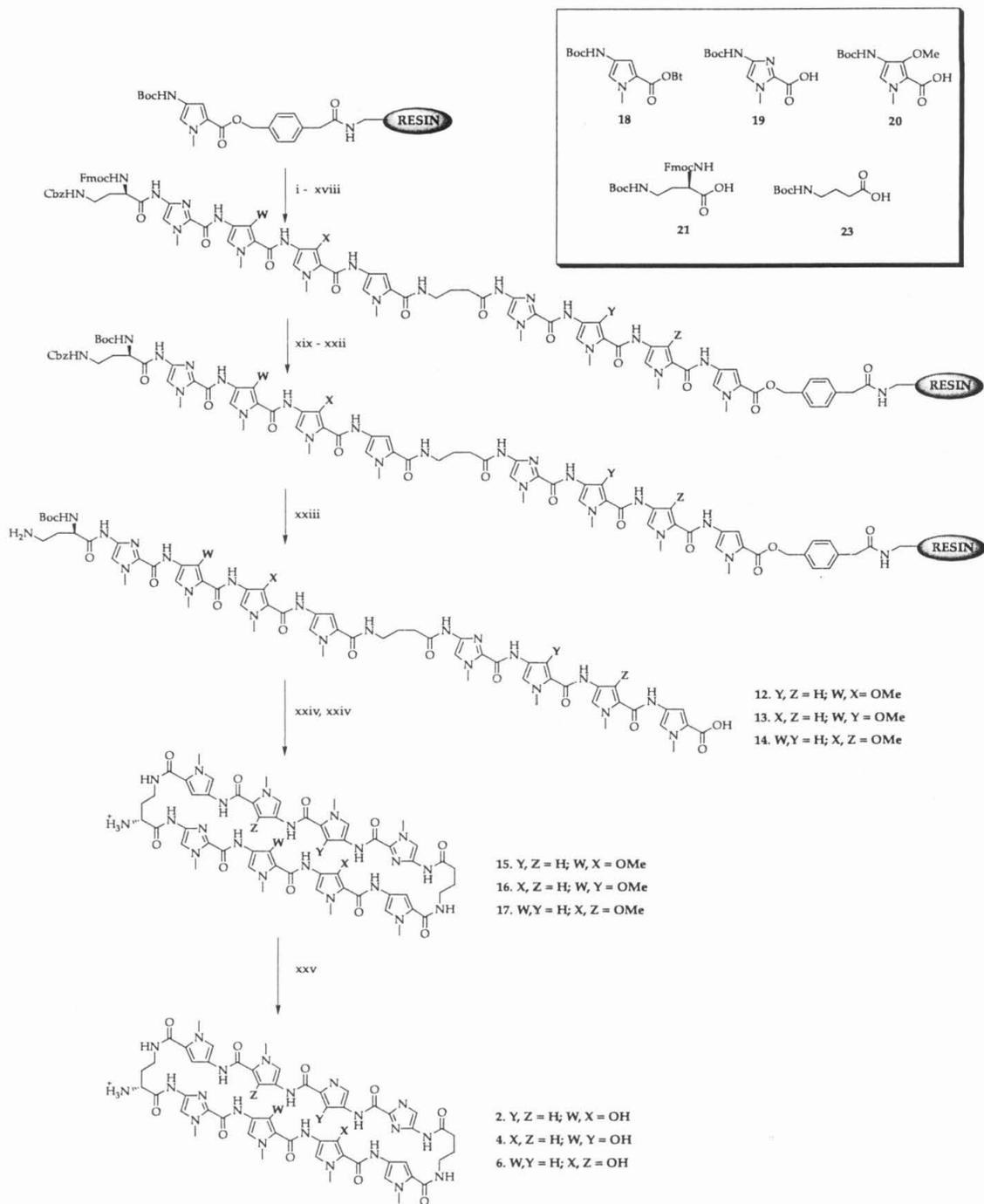


Figure 5.4 Cycle synthetic scheme. Cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -ImHpHpPy-) (**4**): (i) 80 % TFA/DCM, 0.4 M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) 80 % TFA/DCM, 0.4 M PhSH; (iv) Boc-Py-OBt, HBTU, DIEA, DMF; (v) 80 % TFA/DCM, 0.4 M PhSH; (vi) Boc-Im-OH, DIEA, HBTU, DMF; (vii) 80 % TFA/DCM, 0.4 M PhSH; (viii) Boc- γ -diaminobutyric acid (HBTU, DIEA); (ix) 80 % TFA/DCM, 0.4 M PhSH; (x) Boc-Py-OBt, DIEA, DMF; (xi) 80 % TFA/DCM, 0.4 M PhSH; (xii) Boc-Op-OH, HBTU, DIEA, DMF; (xiii) 80 % TFA/DCM, 0.4 M PhSH; (xiv) Boc-Op-OH, HBTU, DIEA, DMF; (xv) 80 % TFA/DCM, 0.4 M PhSH; (xvi) Boc-Im-OH, DIEA, HBTU, DMF; (xvii) 80 % TFA/DCM, 0.4 M PhSH; (xviii) Fmoc- α -Boc- γ -diaminobutyric acid (HBTU, DIEA) (xix) 80 % TFA/DCM, 0.4 M PhSH; (xx) CBZ-OSuc, DMF, (xxi) 80 % Piperidine:DMF (25 °C, 30 min) (xxii) Boc anhydride, DMF, DIEA; (xxiii) Pd(OAc)₂/DMF/ammonium fromate/H₂O; (xxiv) DPPA/K₂CO₃/DIEA; (xxv) TFA; (xxvi) sodium thiophenoxide/DMF. Cyclo-(γ -ImHpPyPy-(R)^{H₂N} γ -ImHpPyPy) (**6**): (i) - (iii), (v) - (xi), (xiii) - (xxvi) same as above; (iv) Boc-Op-OH, HBTU, DIEA, DMF; (xii) Boc-Py-OBt, DIEA, DMF. Cyclo-(γ -ImPyHpPy-(R)^{H₂N} γ -ImPyHpPy-) (**8**): (i), (iii) - (xiii), (xv) - (xxvi) same as above; (vi) Boc-Op-OH, HBTU, DIEA, DMF; (xiv) Boc-Py-OBt, DIEA, DMF. (Inset) Py, Im, Op, and diaminobutyric acid monomers for solid phase synthesis: Boc-Pyrrole-OBt ester^[12] (Boc-Py-OBt) **18**, Boc-Imidazole-OH **19**, Boc-Op-OH **20**, (R)-Fmoc- α -Boc- γ -diaminobutyric acid **21**, Boc-g-aminobutyric acid **23**.

Cbz γ -ImPyPyPy-(R)^{Boc} γ -ImOpOpPy-Pam, Cbz γ -ImOpPyPy-(R)^{Boc} γ -ImOpPyPy-Pam, and Cbz γ -ImPyOpPy-(R)^{Boc} γ -ImPyOpPy-Pam resins. A sample of each resin was then taken and the respective peptides were liberated from the resin with concomitant removal of the CBZ protecting group by reductive cleavage employing Pd(OAc)₂/DMF/water/ammonium formate (37 °C, 14 hrs.).^[7a] Following removal of the resin by filtration, the crude reaction mixtures were purified by reversed phase HPLC to yield polyamides H₂N- γ -ImOpOpPy-(R)^{Boc} γ -ImPyPyPy-OH (**12**), H₂N- γ -ImOpPyPy-(R)^{Boc} γ -ImOpPyPy-OH (**13**), and H₂N- γ -ImPyOpPy-(R)^{Boc} γ -ImPyOpPy-OH (**14**). Each polyamide was then cyclized by treatment with DPPA/K₂CO₃/DMF (22 °C, 5 hrs.),^[7b] all volatiles subsequently removed, the resulting product deprotected in neat TFA, and the crude product purified by reversed phase HPLC to yield Op polyamides cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -ImOpOpPy-) (**15**) cyclo-(γ -ImOpPyPy-(R)^{H₂N} γ -ImOpPyPy-) (**16**) and cyclo-(γ -ImPyOpPy-(R)^{H₂N} γ -ImPyOpPy-) (**17**). Finally, polyamides **15**, **16**, and **17** were O-demethylated with sodium thiophenoxide/DMF (100 °C, 2 hrs.)^[3e] and purified by reversed phase HPLC to yield cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -ImHpHpPy-) (**4**), cyclo-(γ -ImHpPyPy-(R)^{H₂N} γ -ImHpPyPy-) (**6**), and cyclo-(γ -ImPyHpPy-(R)^{H₂N} γ -ImPyHpPy-) (**8**).

Equilibrium Association Constants: Quantitative DNase I footprint titrations (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂, pH 7.0 and 22 °C)^[6] were performed on a ³²P end-labeled 285-bp DNA restriction fragment of pDEH10 to determine equilibrium association constants (K_a) for polyamides **1-8** on three 6-bp binding sites (Figures 5.5, 5.6 and 5.7). The match sites for the parent hairpin **1** and cyclo **2** are bound at subnanomolar concentrations in the same order: 5'-tAGTACTt-3' > 5'-tAGAACTt-3' > 5'-tAGATCTt-3' (Table 1). Hairpin **3** and cyclo **4**, containing contiguous Hp/Py pairs, bind the 6-bp sites (Table 1) with two to three orders of magnitude lower affinity and in the order: 5'-tAGAACTt-3' (match) > 5'-tAGTACTt-3' (mismatch) > 5'-tAGATTCTt-3' (mismatch), where mismatch base pairs are underlined.

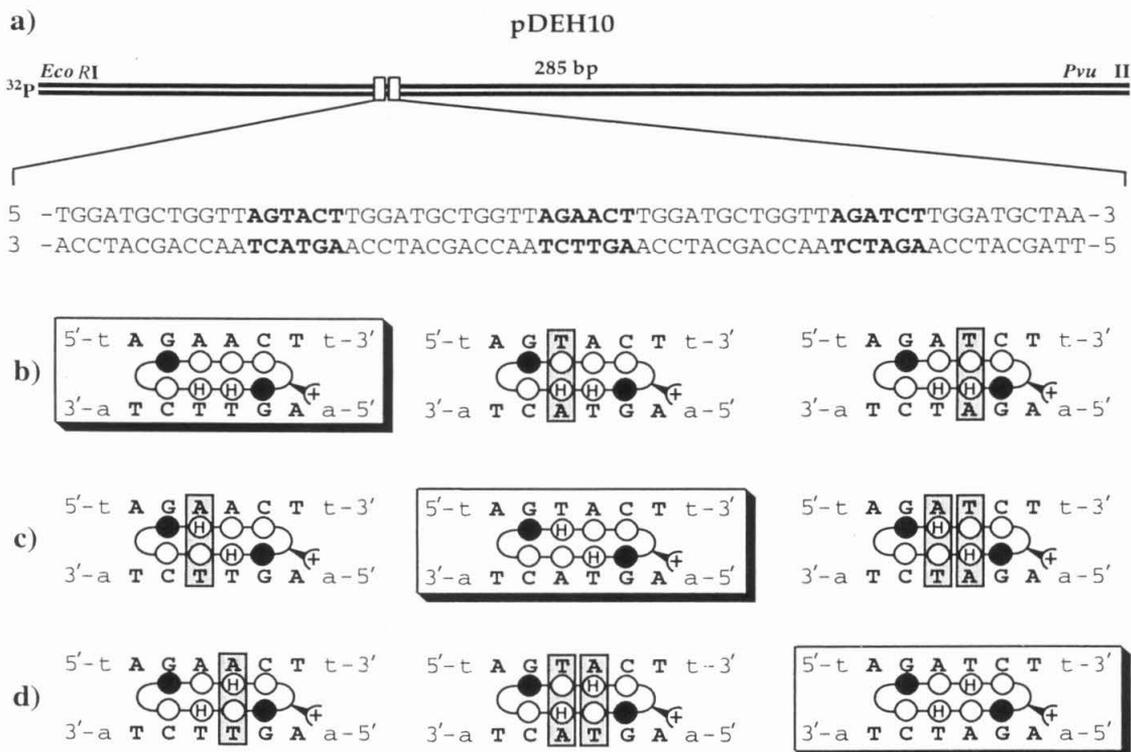


Figure 5.5 Partial sequence of the synthesized insert from the pDEH10 match and mismatch target sites. (a) Illustration of the *Eco* RI/*Pvu* II restriction fragment containing the *Bam* HI and *Hind* III insert. The sequences in bold were the only sites analyzed by quantitative DNase I footprint titrations. (b-d) Schematic binding models of cyclo-(γ -ImPyPyPy-(R)^{H2N} γ -ImHpHpPy-) (4), cyclo-(γ -ImPyPyPy-(R)^{H2N} γ -ImHpHpPy-) (6), and cyclo-(γ -ImPyPyPy-(R)^{H2N} γ -ImHpHpPy-) (8) with their putative match and mismatch sites. Im and Py rings are represented as shaded and unshaded spheres respectively, while Hp rings are annotated with an "H" in the center of an unshaded sphere. Match sites are boxed, and mismatch base pairs are marked with a shaded rectangle.

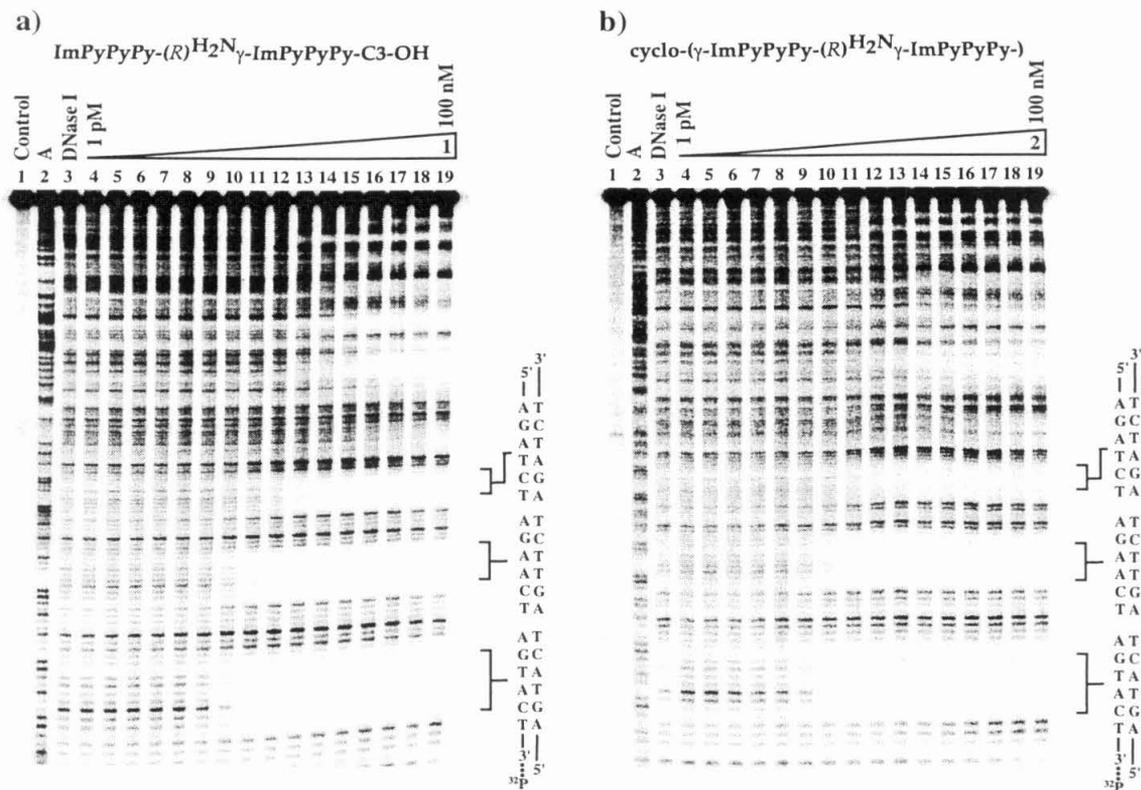


Figure 5.6 Footprinting experiments on the 3'-³²P-labeled 285-bp DNA restriction fragment derived from the plasmid pDEH10. Quantitative DNase I footprint titration experiment with (a) $\text{ImPyPyPy-(R)}^{\text{H}_2\text{N}}\gamma\text{ImHpHpPy-C3-OH}$ (3): lane 1, intact; lane 2, A reaction; lane 3, DNase I standard, lanes 4-19, 1 pM, 2 pM, 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, and 100 nM; (b) $\text{cyclo-(}\gamma\text{-ImPyPyPy-(R)}^{\text{H}_2\text{N}}\gamma\text{ImHpHpPy-)}$ (4): lane 1, intact; lane 2, A reaction; lane 3, DNase I standard, lanes 4-19, 1 pM, 2 pM, 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, and 100 nM; (c) $\text{ImHpPyPy-(R)}^{\text{H}_2\text{N}}\gamma\text{ImHpPyPy-C3-OH}$ (5): lane 2, A reaction; lane 3, DNase I standard, lanes 4-19, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, and 1 μM ; (d) $\text{cyclo-(}\gamma\text{-ImPyHpPy-(R)}^{\text{H}_2\text{N}}\gamma\text{-ImPyHpPy-)}$ (4): lane 1, intact; lane 2, A reaction; lane 3, DNase I standard, lanes 4-19, 1 pM, 2 pM, 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, and 100 nM. All reactions contain 30 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂. 5'-tAGAACTt-3', 5'-tAGTACTt-3', and 5'-tAGTACTt-3' binding sites are shown on the right side of the autoradiograms.

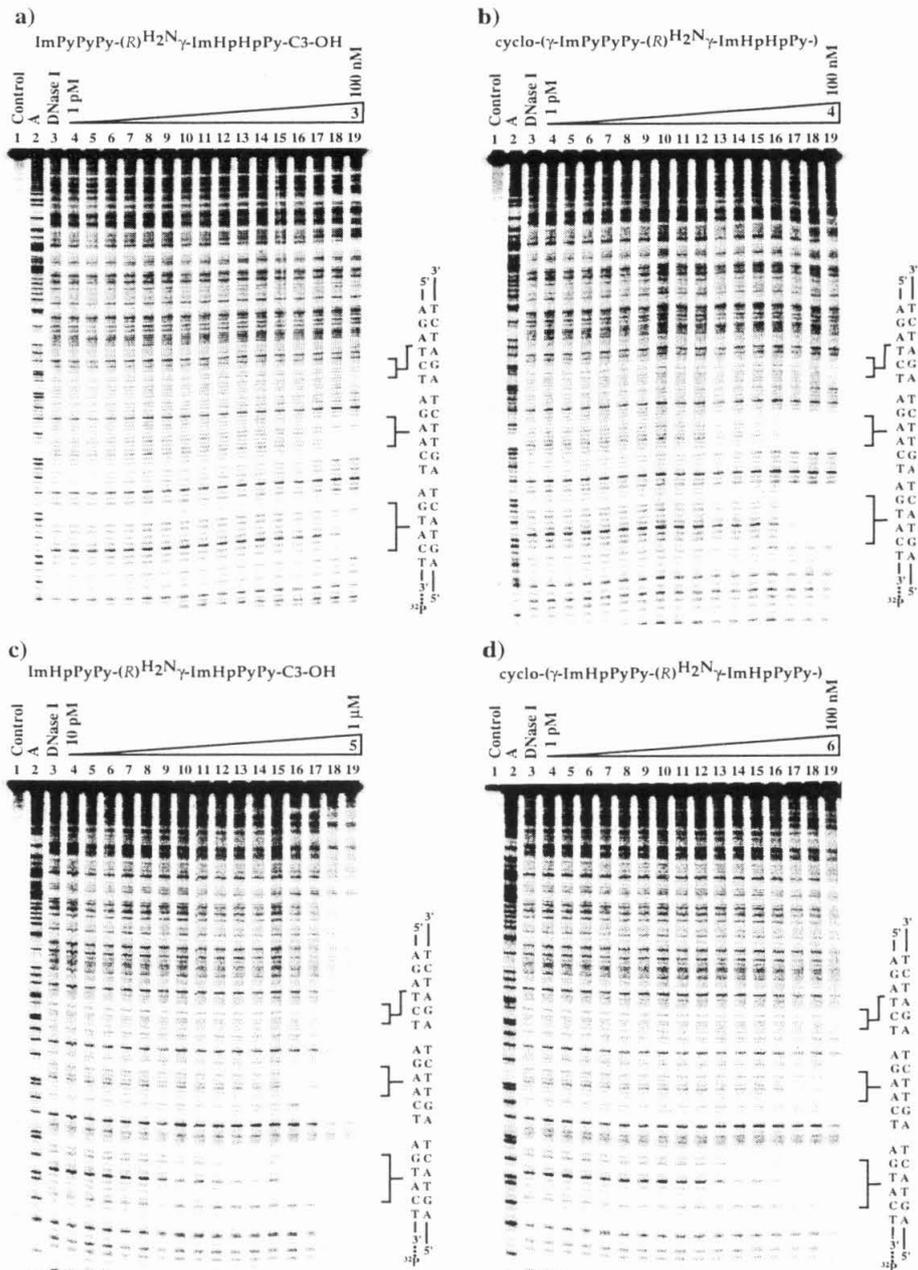
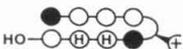
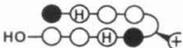
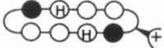
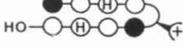
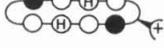


Figure 5.7 Footprinting experiments on the 3'-³²P-labeled 285-bp DNA restriction fragment derived from the plasmid pDEH10. Quantitative DNase I footprint titration experiment with (a) ImPyPyPy-(R)^{H2N}-γ-ImPyPyPy-C3-OH (3): lane 1, intact; lane 2, A reaction; lane 3, DNase I standard, lanes 4-19, 1 pM, 2 pM, 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, and 100 nM; (b) cyclo-(γ-ImPyPyPy-(R)^{H2N}-γ-ImPyPyPy-) (4): lane 1, intact; lane 2, A reaction; lane 3, DNase I standard, lanes 4-19, 1 pM, 2 pM, 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, and 100 nM; All reactions contain 30 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂. 5'-tAGAACTt-3', 5'-tAGTACTt-3', and 5'-tAGTACTt-3' binding sites are shown on the right side of the autoradiograms.

Table 1. Equilibration Association Constants (M^{-1}) for Polyamides^{a-c}

Polyamide	5'-tAGAACTt-3'	5'-tAGTACTt-3'	5'-tAGATCTt-3'
1 	1.0×10^{10} (0.8)	2.1×10^{10} (0.3)	4.0×10^9 (0.3)
2 	4.0×10^{10} (0.6)	7.0×10^{10} (0.4)	5.0×10^9 (0.3)
3 	4.4×10^6 (0.3)	1.4×10^6 (0.5)	$\leq 1 \times 10^6$
4 	7.5×10^7 (1.2)	1.4×10^7 (0.3)	$\leq 1 \times 10^6$
5 	2.6×10^7 (0.5)	7.0×10^8 (1.2)	$\leq 1 \times 10^7$
6 	3.5×10^8 (0.6)	3.2×10^9 (1.1)	$\leq 1 \times 10^7$
7 	2.0×10^7 (0.5)	5.4×10^7 (1.1)	2.7×10^7 (0.9)
8 	3.5×10^7 (1.1)	1.4×10^7 (0.5)	6.2×10^7 (1.7)

^aValues reported are the mean values obtained from three DNase I titration experiments. ^bThe assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂. ^cMatch site association constants are shown in bold type.

Hairpin 5 and cyclo 6, with a 5'-stagger of Hp rings, preferentially bind their match sites with higher affinities relative to polyamides 3 and 4, and all sites the following order: 5'-tAGTACTt-3' (match) > 5'-tAGAACTt-3' (mismatch) > 5'-tAGATCTt-3' (double mismatch) (Table 1). Hairpin 7 and cyclo 8, with a 3'-stagger of Hp rings, binds 5'-tAGATCTt-3' (match), 5'-tAGAACTt-3' (mismatch), and 5'-AGTACTt-3' (double mismatch) sites (Table 1) with lower overall affinity than the 5'-stagger, but cyclo 8 (not hairpin 7) establishes preference for the cognate match site 5'-tAGATCTt-3' by a factor of 2.

Specificity of Hp/Py Cycles: Comparison of the binding affinities reveals the effects introduction of multiple Hp/Py pairs has on sequence specificity (Table 2). On the

contiguous Hp/Py pairings of cycle 4 affords a 16-fold enhanced binding to the 5'-tAGAACTt-3' match site ($K_a = 7.5 \times 10^7 \text{ M}^{-1}$), relative to hairpin 3, and 5- to 75-fold specificity versus the other two single base pair mismatch sites. Cycle 6 with 5'-staggered Hp/Py pairs restores subnanomolar binding to the cognate 5'-tAGTACTt-3' match site ($K_a = 3.2 \times 10^9 \text{ M}^{-1}$) with 10- and 320-fold specificity toward the single base pair mismatch site 5'-tAGAACTt-3' and double base pair mismatch 5'-tAGATCCTt-3', respectively. It is surprising to note that a 3'-stagger of Hp/Py pairs in hairpin 7 selectively binds the double base pair mismatch 5'-tAGTACT-3' site with a 2-fold preference over the designed 5'-tAGATCTt-3' match site. Perhaps preferential positioning of Hp/Py pairs within hairpin polyamides or DNA microstructure prevents favorable hairpin•DNA interactions at the match site. However, the cyclic analog 8 reverses this mismatch preference, and effectively discriminates 5'-tAGATCT-3' from 5'-tAGTACT-3' by enhancing affinity to the match sequence 2-fold, and providing 4-fold specificity versus the double base pair mismatch site. These results illustrate that substitution of Py/Py with Hp/Py pairs in a cyclic polyamide motif may be an effective strategy for discriminating multiple A•T/T•A base pairs with enhanced affinities but that 5'- versus 3'- stagger of Hp rings are quite different energetically.

It might have been expected that increasing the demand for correct ligand/DNA interaction by introduction of multiple Hp pairs would have been detrimental to the binding affinity and specificity properties of the rigid cyclic polyamide motif in comparison to their respective hairpin counterparts. In contrast, we have demonstrated that introduction of multiple Hp residues according to the "pairing rules" in a cyclic polyamide motif enhances DNA binding properties of 8-ring polyamides by increasing overall affinity while augmenting specificity towards 5'-tTGAACTt-3' and 5'-tTGATCTt-3' sites. Although these results indicate that employing a cyclic polyamide motif containing Hp/Py pairs to discriminate A/T sequences through minor groove

recognition is viable, future studies will be required to test the general applicability of this approach and will be reported in due course.

Experimental Section

General: Dicyclohexylcarbodiimide (DCC), Hydroxybenzotriazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU), Boc anhydride (Boc₂O) and 0.25 mmol/gram Boc-β – alanine - (-4-carboxamidomethyl) - benzyl - ester-copoly(styrene-divinylbenzene) resin (Boc-β-Pam-Resin) was purchased from Peptides International. (*R*)-2-Fmoc-4-Boc-diaminobutyric acid and γ-aminobutyric acid (γ) were purchased from Bachem. *N,N*-diisopropylethylamine (DIEA) and *N,N*-dimethylformamide (DMF) were purchased from Applied Biosystems. Dichloromethane (DCM) was reagent grade from EM; thiophenol (PhSH), dimethylaminopropylamine (Dp), diphenylphosphoryl azide (DPPA), piperidine, palladium acetate, potassium carbonate, and ammonium formate were from Aldrich. Trifluoroacetic acid (TFA) Biograde was from Halocarbon and *N*-(benzyloxycarbonyloxy) succinimide (CBZ-OSuc) was from Fluka. All reagents were used without further purification.

Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. A shaker for manual solid phase synthesis was obtained from St. John Associates, Inc. Screw-cap glass peptide synthesis reaction vessels (5 mL and 20 mL) with a #2 sintered glass frit were made as described by Kent.^[10] UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on a Beckman Gold system using a RAINEN C₁₈, Microsorb MV, 5μm, 300 x 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, gradient elution 1.25% acetonitrile/min. Preparatory reversed phase HPLC was performed on a Beckman

HPLC with a Waters DeltaPak 25 x 100 mm, 100 μ m C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. Milli-Q water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μ m filtered.

Polyamide Synthesis: Reagents and protocols for polyamide synthesis were as previously described.^{[3c],[5],[7a]} Polyamides were purified by reversed phase HPLC on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100 μ m C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. Extinction coefficients were calculated based on $\epsilon = 8333/\text{ring}$ at 304 nm.^[11]

General procedure for NaBH₄ cleavage:^[8] 100 mg of the appropriate resin was placed in a sealable container and swollen in 1 mL dry THF. 15 mg of NaBH₄ was subsequently added, the vessel then sealed and the reaction heated at 60 °C for 5 hours. After cooling to room temperature, the reaction was quenched by addition of 4 mL 20% TFA/80% H₂O; 4 mL CH₃CN was then added and the supernatant was collected by filtration. The resulting solution was frozen in liquid nitrogen, lyophilized, resuspended in .1% TFA/H₂O and purified by reversed phase HPLC to yield the appropriate polyamide.

ImPyPyPy-(R)^{H₂N} γ -ImOpOpPy-C3-OH (9): Recovered upon lyophilization of the appropriate fractions as a white powder (1 mg, 3% yield). MALDI-TOF-MS [M⁺-H] (monoisotopic), 1199.5: 1199.5 calc. for C₅₅H₆₇N₂₀O₁₂.

ImOpPyPy-(R)^{H₂N} γ -ImOpPyPy-C3-OH (10): Recovered upon lyophilization of the appropriate fractions as a white powder (.9 mg, 3% yield). MALDI-TOF-MS [M⁺-H] (monoisotopic), 1199.5: 1199.5 calc. for C₅₅H₆₇N₂₀O₁₂.

ImPyOpPy-(R)^{H₂N} γ -ImPyOpPy-C3-OH (11): Recovered upon lyophilization of the

appropriate fractions as a white powder (.5 mg, 2% yield). MALDI-TOF-MS [M^+ -H] (monoisotopic), 1199.7: 1199.5 calc. for $C_{55}H_{67}N_{20}O_{12}$.

General procedure for $Pd(OAc)_2/NH_4CO_2H$ cleavage.^[7a] 300 mg of the appropriate resin (.5 mmol/g as synthesized in reference 7a) and 300 mg of $Pd(OAc)_2$ were placed in a sealable container, suspended in 1 mL DMF, and shaken at 37 °C for 3 hours. 700 mg NH_4CO_2H , dissolved in 1 mL H_2O , was then added slowly over 3 minutes, the vessel sealed, and the resulting solution shaken at 37 °C for 12 hours. The supernatant was collected by filtration and subsequently purified by reversed phase HPLC to yield the appropriate polyamide.

$H_2N-\gamma-ImOpOpPy-(R)^{Boc}-\gamma-ImPyPyPy-OH$ (**12**). Recovered upon lyophilization of the appropriate fractions as a white powder (7 mg, 4% yield). MALDI-TOF-MS [M^+ -H] (monoisotopic), 1342.7: 1342.6 calc. for $C_{61}H_{76}N_{21}O_{15}$.

$H_2N-\gamma-ImOpPyPy-(R)^{Boc}-\gamma-ImOpPyPy-OH$ (**13**). Recovered upon lyophilization of the appropriate fractions as a white powder (13 mg, 7% yield). MALDI-TOF-MS [M^+ -H] (monoisotopic), 1342.7: 1342.6 calc. for $C_{61}H_{76}N_{21}O_{15}$.

$H_2N-\gamma-ImPyOpPy-(R)^{Boc}-\gamma-ImPyOpPy-OH$ (**14**). Recovered upon lyophilization of the appropriate fractions as a white powder (5 mg, 3% yield). MALDI-TOF-MS [M^+ -H] (monoisotopic), 1342.7: 1342.6 calc. for $C_{61}H_{76}N_{21}O_{15}$.

General cyclization procedure.^[7b] The appropriate precycle was dissolved in DMF (1 μ mole/5 mL), K_2CO_3 was added (10 mg/ μ mole of polyamide) and the resulting solution stirred at RT for 30 minutes. DPPA (30 μ L/ μ mole of polyamide) was then added and the resulting solution stirred at room temperature for 5 hours. The solution was filtered

to remove excess K_2CO_3 , and all volatiles in the filtrate were removed in vacuo. The resulting white powder was dissolved in 2 mL TFA and allowed to stand at RT for 30 minutes. The volume was subsequently adjusted to 10 mL with .1% TFA/ H_2O and purification by reversed phase HPLC yielded the appropriate cycle.

cyclo-(γ -ImPyPyPy-(R) H_2N γ -ImOpOpPy-) (15). Recovered upon lyophilization of the appropriate fractions as a white powder (1.2 mg, 20% yield). MALDI-TOF-MS [M^+ -H] (monoisotopic), 1224.6: 1224.5 calc. for $C_{56}H_{66}N_{21}O_{12}$.

cyclo-(γ -ImOpPyPy-(R) H_2N γ -ImOpPyPy-) (16). Recovered upon lyophilization of the appropriate fractions as a white powder (1.5 mg, 25% yield). MALDI-TOF-MS [M^+ -H] (monoisotopic), 1224.5: 1224.5 calc. for $C_{56}H_{66}N_{21}O_{12}$.

cyclo-(γ -ImPyOpPy-(R) H_2N γ -ImPyOpPy-) (17). Recovered upon lyophilization of the appropriate fractions as a white powder (1.4 mg, 20% yield). MALDI-TOF-MS [M^+ -H] (monoisotopic), 1224.8: 1224.5 calc. for $C_{56}H_{66}N_{21}O_{12}$.

General O-demethylation procedure.^[3e] 100 mg NaH, 1 mL DMF, .5 mL thiophenol were heated at 100 °C for 10 minutes. The polyamide, dissolved in 1 mL DMF was subsequently added and the resulting mixture heated at 100 °C for 2 hours. After cooling to 22 °C, the solution was diluted to 10 mL with 20% TFA/80% H_2O , extracted x3 with EtOAc and x2 with Et_2O , and the aqueous phase purified by reversed phase HPLC to yield the appropriate polyamide.

ImPyPyPy-(R) H_2N γ -ImHpHpPy-C3-OH (3). Recovered upon lyophilization of the appropriate fractions as a white powder (.2 mg, 30% yield). MALDI-TOF-MS [M^+ -H] (monoisotopic), 1171.4: 1171.5 calc. for $C_{53}H_{63}N_{20}O_{12}$.

cyclo-(γ -ImPyPyPy-(R)^{H₂N} γ -ImHpHpPy-) (4). Recovered upon lyophilization of the appropriate fractions as a white powder (.25 mg, 25% yield). MALDI-TOF-MS [M⁺-H] (monoisotopic), 1196.5: 1196.5 calc. for C₅₄H₆₂N₂₁O₁₂.

ImHpPyPy-(R)^{H₂N} γ -ImHpPyPy-C3-OH (5). Recovered upon lyophilization of the appropriate fractions as a white powder (.18 mg, 30% yield). MALDI-TOF-MS [M⁺-H] (monoisotopic), 1171.5: 1171.6 calc. for C₅₃H₆₃N₂₀O₁₂.

cyclo-(γ -ImHpPyPy-(R)^{H₂N} γ -ImHpPyPy-) (6). Recovered upon lyophilization of the appropriate fractions as a white powder (.2 mg, 35% yield). MALDI-TOF-MS [M⁺-H] (monoisotopic), 1196.5: 1196.5 calc. for C₅₄H₆₂N₂₁O₁₂.

ImPyOpPy-(R)^{H₂N} γ -ImPyOpPy-C3-OH (7). Recovered upon lyophilization of the appropriate fractions as a white powder (.2 mg, 30% yield). MALDI-TOF-MS [M⁺-H] (monoisotopic), 1171.4: 1171.5 calc. for C₅₃H₆₃N₂₀O₁₂.

cyclo-(γ -ImPyHpPy-(R)^{H₂N} γ -ImPyHpPy-) (8). Recovered upon lyophilization of the appropriate fractions as a white powder (.12 mg, 20% yield). MALDI-TOF-MS [M⁺-H] (monoisotopic), 1196.4: 1196.5 calc. for C₅₄H₆₂N₂₁O₁₂.

DNA Reagents and Materials. Enzymes were purchased from Boehringer-Mannheim and used with their supplied buffers. Deoxyadenosine and thymidine 5'-[α -³²P] triphosphates were obtained from Amersham, 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.^[12]

Construction of plasmid DNA. The plasmid pDEH10 was constructed by hybridization of oligos 5'-CTAGTGGATGCTGGTTAGTACTTGGATGCTGGTAGA ACTTGGATGCTGGTTAGATCTTGGATGCTGGTTGCA-3' and 5'-ACCAGCATCCA AGATCTAACCAGCATCCAAGTTCTAACCAGCATCCAAGTACTAACCAGCATC CA-3' followed by insertion into *Xba* I/*Pst* I linearized pUC19 plasmid using T4 DNA ligase. The resultant construct was 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 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 duplex DNA.

Preparation of 3'-End-Labeled Restriction Fragments. The plasmid pDEH10 was linearized with *Eco* RI and *Pvu* II, then treated with the Sequenase enzyme, deoxyadenosine 5'-[α -³²P]triphosphate and thymidine 5'-[α -³²P]triphosphate for 3' labeling. The labeled 3' fragment was loaded onto a 6% non-denaturing polyacrylamide gel, and the desired 285 base pair band was visualized by autoradiography and isolated. Chemical sequencing reactions were performed according to published methods.^[13]

DNase I Footprinting.^[15] All reactions were carried out in a volume of 400 µL. We note explicitly that no carrier DNA was used in these reactions until after DNase I cleavage. 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 MgCl₂, 5 mM CaCl₂, and 30 kcpm 3'-radiolabeled DNA. The solutions were allowed to equilibrate for 4 hours 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 1.875 u/mL) and was allowed to proceed for 7 min at 22 °C. The reactions were stopped by adding 50 μL of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/mL 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 6 min, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V for 1 hour. The gels were dried under vacuum at 80 °C, then quantitated using storage phosphor technology.

Equilibrium association constants were determined as previously described.^{11a} The data were analyzed by performing volume integrations of the 5'-AGAACT-3', 5'-AGTACT-3', and 5'-AGATCT-5' sites and a reference site. The apparent DNA target site saturation, θ_{app} , was calculated for each concentration of polyamide using the following equation:

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

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{ref}° correspond to those values for a DNase I control lane to which no polyamide has been added. The ($[L]_{tot}$, θ_{app}) data points were fit to a Langmuir binding isotherm (eq 2, $n=1$ for polyamides **1** and **2**, by minimizing the difference between θ_{app} and θ_{fit} , using the modified Hill equation:

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

where $[L]_{\text{tot}}$ corresponds to the total polyamide concentration, K_a corresponds to the equilibrium association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software (version 2.1, Abelbeck software) with K_a , θ_{max} , and θ_{min} as the adjustable parameters. All acceptable fits had a correlation coefficient of $R > 0.97$. At least three 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. The data were normalized using the following equation:

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

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 12-20 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 Treadway Foundation (Bio.47123-1-ENDOW.471230) for a

research award to D.M.H., and the National Institutes of Health (GM19789-02) for a postdoctoral fellowship to C.M. We thank G.M. Hathaway for MALDI-TOF mass spectrometry and S. White for construction of pDEH10.

References

1. a) J. M. Gottesfeld, L. Nealy, J. W. Trauger, E. E. Baird, P. B. Dervan, *Nature* **1997**, 387, 202-205. b) L. A. Dickinson, R. J. Gulizia, J. W. Trauger, E. E. Baird, D. E. Mosier, J. M. Gottesfeld, P. B. Dervan *Proc. Natl. Acad. Sci. USA* **1998**, 95, 12890-12895.
2. P. B. Dervan, R. W. Burli, *Curr. Opin. Chem. Biol* **1999**, 3, 688-693.
3. a) S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, *Nature* **1998**, 391, 468-471. b) C. L. Kielkopf, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, D. C. Rees *Science* **1998**, 282, 111-115. c) A. R. Urbach, J. W. Szewczyk, S. White, J. M. Turner, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1999**, 121, 11621-11629. d) S. White, J. M. Turner, J. W. Szewczyk, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1999**, 121, 260-261. e) C. L. Kielkopf, R. E., Bremer, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, D. C. Rees *J. Mol. Biol.* **2000**, 295, 557-567.
4. J. W. Trauger, E. E. Baird, M. Mrksich, P. B. Dervan, *J. Am Chem. Soc.* **1996**, 118, 6160-6166.
5. E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, 118, 6141-6146.
6. a) M. Brenowitz, D. F. Senear, M. A. Shea, G. K. Ackers, *Methods Enzymol.* **1986**, 130, 132-181. b) M. Brenowitz, D. F. Senear, M. A. Shea, G. K. Ackers, *Proc. Natl. Acad. Sci. U.S.A.* **1986**, 83, 8462-8466. c) D. F. Senear, M. Brenowitz, M. A. Shea, G. K. Ackers, *Biochemistry* **1986**, 25, 7344-7354.
7. a) D. M. Herman, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1999**, 121 1121-1129 b) J. Cho, M. E. Parks and P. B. Dervan *Proc. Natl. Acad. Sci. USA* **1995**, 92, 10389-10392.
8. a) A. R. Mitchell, S. B. Kent, M. Engelhard, R. B. J. Merrifield, *J. Org. Chem.* **1978**, 43, 2845-2852. b) J. M. Stewart, J. D. Young, *Solid Phase Peptide Synthesis*. Pierce Chemical Company, rockfor IL. **1984**.
9. S. White, E. E. Baird, P. B. Dervan, *Biochemistry* **1996**, 35, 12532-12537.
10. S.B.H. Kent, *Annu. Rev. Biochem.* **1988**, 57, 957-989.
11. D. S. Pilch, N. A. Pokar, C. A. Gelfand, S. M. Law, K. J. Breslauer, E. E. Baird, P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, 93, 8306-8311.
12. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.
13. a) B. L. Iverson, P. B. Dervan, *Nucl. Acids Res.* **1987**, 15, 7823-7830. b) A. M. Maxam, W.S. Gilbert, *Methods Enzymol.* **1980**, 65, 499-560.
14. W. S. Wade, M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc.* **1992**, 114, 8783-8794.

Chapter 6

**Hp/Py Hairpin Polyamides for the Discrimination of
Core A•T Sequences in 5'-GWWC-3'**

Abstract: *In hairpin polyamides, substitution of Hp/Py for Py/Py pairings enhances selectivity but compromises binding affinity for specific sequences. The differential destabilization of polyamide•DNA complexes upon addition of Hp/Py partially depends on the position of the Hp residue within the linear polyamide sequence. In an effort to elucidate how placement of Hp/Py pairs in specific positions of hairpin polyamides affects DNA binding properties with respect to affinity and sequence specificity, eight hairpin polyamides containing single or double Hp residues were synthesized. Quantitative footprint titration experiments on a DNA restriction fragment containing the 6-base pair sites 5'-tAGTACTt-3', 5'-tAGATCTt-3', and 5'-tAGAACTt-3', demonstrate that a single Hp substitution as in ImHpPyPy- γ -ImPyPyPy- β -Dp (2) and ImPyPyPy- γ -ImHpPyPy- β -Dp (3), compromises affinity for the match site 5'-tAGTACTt-3' by a factor of 5 relative to the unsubstituted control hairpin ImPyPyPy- γ -ImPyPyPy- β -Dp (1), while enhancing specificity versus the single base pair mismatch site 5'-tAGATCTt-3' to ≥ 720 -fold and ≥ 590 -fold, respectively. The di-Hp-substituted polyamide ImHpPyPy- γ -ImHpPyPy- β -Dp (4) recognizes the 5'-tAGTACTt-3' match site with 50-fold reduced affinity relative to (1), but maintains a ≥ 70 -fold preference versus the single (5'-tAGAACTt-3') and double (5'-tAGATCTt-3') base pair mismatch sites. Polyamides ImPyHpPy- γ -ImPyPyPy- β -Dp (5), ImPyPyPy- γ -ImPyHpPy- β -Dp (6), and ImPyHpPy- γ -ImPyHpPy- β -Dp (7) reduce specificity for the mismatch site 5'-tAGTACTt-3' over the match sites site 5'-tAGAACTt-3 and site 5'-tAGATCTt-3' by a factor of 11. The results presented herein help to identify guidelines for the utilization of Hp residues in the design of eight ring hairpin polyamides.*

Introduction

Cell permeable synthetic ligands that regulate gene expression⁽²⁾ are implicitly potent therapeutics for medicinal purposes. Polyamides containing the three aromatic amino acids pyrrole (Py), imidazole (Im), and 3-hydroxypyrrole (Hp), are a class of small molecules that recognize predetermined sequences of DNA with subnanomolar affinities.⁽¹⁾ DNA recognition depends on a code of side-by-side aromatic amino acid pairings that are oriented N-C with respect to the 5'-3' direction of the DNA helix in the minor groove. An Im/Py pair distinguishes G•C from C•G, a Py/Im distinguishes C•G from G•C and both of these from A•T/T•A base pairs.^[2] A Py/Py pair binds both A•T and T•A in preference to G•C/C•G.^[2] The recent introduction of the Hp monomer into Hp/Py pairs discriminates T•A from A•T, while Py/Hp targets A•T in preference to T•A and both of these from G•C/C•G completing recognition of the four Watson-Crick base pairs^[3] and expanding the repertoire of DNA sequences efficiently discriminated by polyamides. Polyamide dimers linked with a γ -aminobutyric acid (γ) or a γ -diaminobutyric acid ($((R)^{H_2N} \gamma)$ turn residue provides a hairpin motif that binds target sites with two to three orders of magnitude enhanced affinity and has the important feature that ring pairings are set in place unambiguously as compared to homodimers which can afford "slipped motifs".^(2,4)

Hp/Py Hairpin Polyamides: High resolution X-ray diffraction data reveals that the T•A selectivity of the Hp/Py pair arises from a combination of shape selection of an asymmetric cleft of the floor of the minor groove formed by the O2 of thymine and C2-H of adenine and specific hydrogen bonds between the 3-hydroxy and 4-carboximido groups of Hp with the O2 of T.^(3b) Crystallographic data also illustrates that Hp/Py polyamides bind undistorted B-form DNA; however, a localized 0.5 Å melting of the T•A Watson-Crick base pair is observed when the ImHpPyPy-Dp dimer is bound to 5'-AGTACT-3' and is potentially responsible for the energetic destabilization of the

Hp/Py pair relative to the Py/Py pair.^(3b) A subsequent crystallographic study of an (ImPyHpPy-Dp)₂•5'-AGATCT-3' complex did not reveal such a melting, and it is uncertain at this time the extent base pair melting contributes to binding destabilization.^(3e) While incorporation of Hp/Py pairs into polyamides affords a substantial enhancement of sequence specificity, there is an energetic penalty associated with the number and location of Hp/Py substitutions. Replacement of a single Py/Py pair with an Hp/Py pair in ImImHpPy- γ -ImPyPyPy- β -Dp results in a 5-fold decrease in affinity for the 5'-tGGTCTt-3' site.^(3a-c) Placement of two Hp/Py pairs in a "5'-staggered" or "3'-staggered" orientation (with respect to the DNA helix) as in ImHpPyPy-(R)^{H₂N} γ -ImHpPyPy-C3-OH and ImPyHpPy-(R)^{H₂N} γ -ImPyHpPy-C3-OH affords one to two orders of magnitude reduction in binding affinity for respective 5'-tGTACTt-3' and 5'-tGATCTt-3' target sites.^(3f) Contiguous arrangement of two Hp/Py pairs in ImHpHpPy-(R)^{H₂N} γ -ImPyPyPy-C3-OH reduces recognition of 5'-tGAACTt-3' by three orders of magnitude.⁽⁵⁾ Surprisingly, addition of three Hp/Py pairs in ImHpHpPyPy- γ -ImHpPyPyPy- β -Dp reduces affinity for the 5'-tAGTTACTt-3' target site by a modest factor of 10 suggesting that all Hp substitutions do not equally contribute to complex destabilization.^(3d) In light of these observations, it remains to be determined if we can develop guidelines for the effective placement of Hp/Py pairs in polyamides which preserve DNA recognition properties with respect to affinity and sequence selectivity.

As a first step towards understanding how the implementation of Hp residues in hairpin polyamides discriminates three 6-base pair sequences 5'-WGNNCW-3', where NN = TA, AT, or AA, we consider the parent hairpin polyamide ImPyPyPy- γ -ImPyPyPy- β -Dp (1) which binds the aforementioned sites with high affinities and modest sequence selectivities as anticipated from Py/Py degenerate recognition of A•T/T•A base pairs. Incremental substitution of the central Py rings in positions 1 – 4 (Figure 5.1) with Hp residues varying in number and orientation within an eight ring

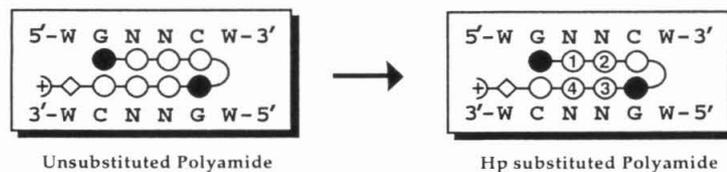


Figure 6.1. Schematic binding model eight ring hairpin bound to the respective six base pair match sites 5'-WGNNCW-3' site where NN = TA, AA, or AT. Im and Py rings are represented as shaded and unshaded spheres respectively. The numbering scheme of 1 – 4 represents core polyamide positions for Py to Hp substitutions.

hairpin allows us to compare the relative contributions that each Hp substitution makes to both the discrimination of core TA, AT, and AA sequences and to the energetic destabilization of hairpin•DNA complexes.

To evaluate the extent to which each Hp substitution affects the stability of polyamide•DNA complexes, eight ring hydroxypyrrole hairpins were synthesized by solid phase methods;^[5] ImHpPyPy- γ -ImPyPyPy- β -Dp (2), ImPyPyPy- γ -ImHpPyPy- β -Dp (3), ImHpPyPy- γ -ImHpPyPy- β -Dp (4), ImPyHpPy- γ -ImPyPyPy- β -Dp (5), ImPyPyPy- γ -ImPyHpPy- β -Dp (6), ImPyHpPy- γ -ImPyHpPy- β -Dp (7), ImHpHpPy- γ -ImPyPyPy- β -Dp (8), and ImPyPyPy- γ -ImHpHpPy- β -Dp (9) (Figure 6.2). Equilibrium association constants (K_a) for the eight polyamides were determined by quantitative DNase I footprint titration^[6] experiments on a DNA fragment containing three 6-base pair binding sites 5'-tAGTACTt-3', 5'-tAGATCTt-3, and 5'-tAGAACTt-3'. Based on the pairing rules, polyamides 2 and 3 with Hp residues in positions 1 or 4, bind 5'-tAGTACTt-3' and 5'-tGAACTt-3' as match sites and 5'-tAGATCTt-3' as a single base pair mismatch site (where mismatched base pairs are underlined). Polyamide 4 with a 5'-stagger of Hp residues in positions 1 and 4 binds 5'-tAGTACTt-3' as a match site, 5'-tAGAACTt-3' as a single base pair mismatch site, and 5'-tAGATCTt-3' as a double base pair mismatch site. Polyamides 5 and 6, with Hp residues in positions 2 or 4, bind 5'-tAGATCTt-3' and 5'-tAGAACTt-3' as match sites and 5'-tAGTACTt-3' as a single base

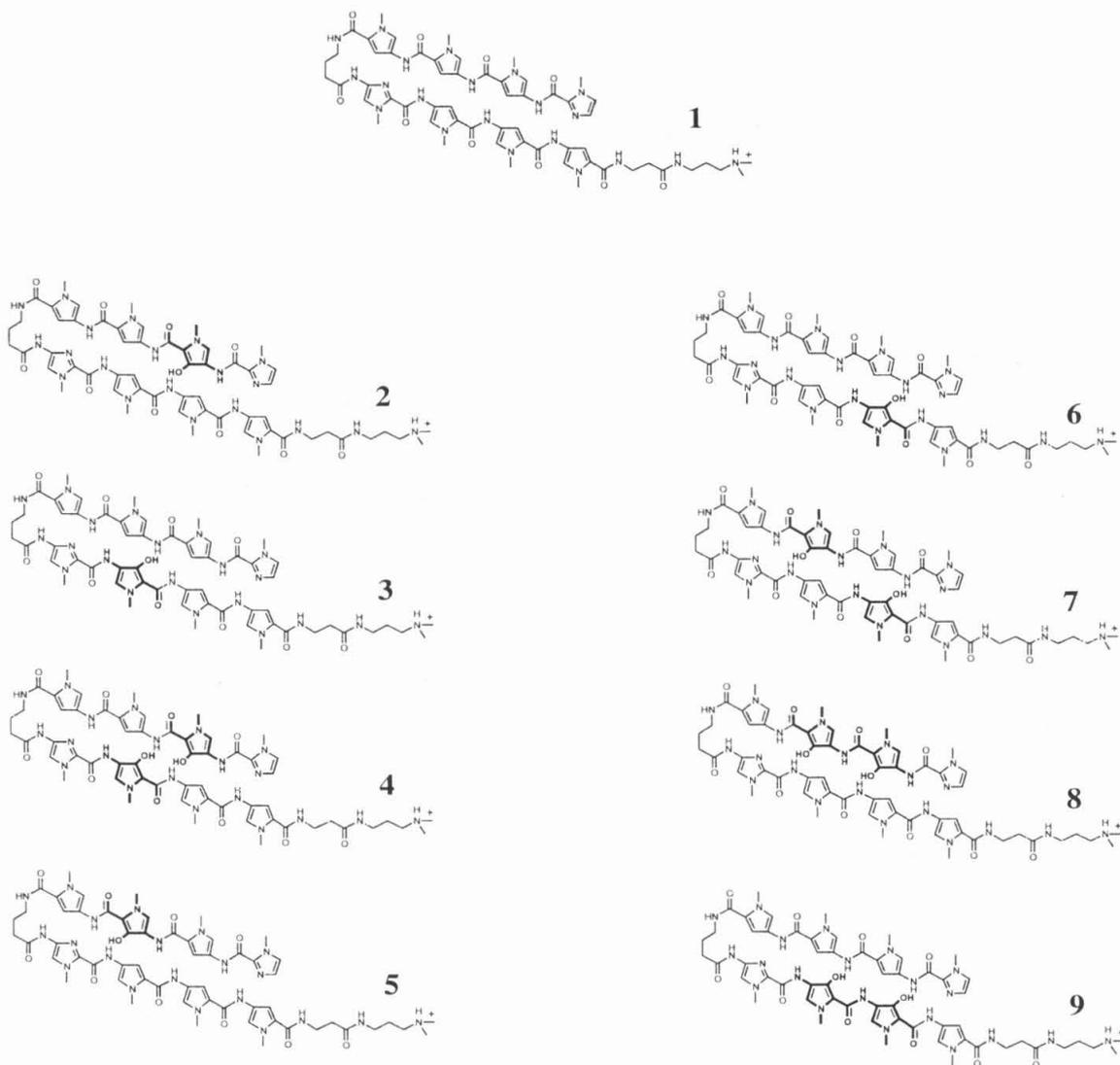


Figure 6.2. Structures of the eight ring polyamides ImPyPyPy- γ -ImPyPyPy- β -Dp (1), ImHpPyPy- γ -ImPyPyPy- β -Dp (2), ImPyPyPy- γ -ImHpPyPy- β -Dp (3), ImHpPyPy- γ -ImHpPyPy- β -Dp (4), ImPyHpPy- γ -ImPyPyPy- β -Dp (5), ImPyPyPy- γ -ImPyHpPy- β -Dp (6), ImPyHpPy- γ -ImPyHpPy- β -Dp (7), ImHpHpPy- γ -ImPyPyPy- β -Dp (8), and ImPyPyPy- γ -ImPyPyPy- β -Dp (9).

pair mismatch site. Polyamide 7 with a 3'-stagger of Hp residues in positions 2 and 4 binds 5'-tAGATCTt-3' as a match site, 5'-tAGAACTt-3' as a single base pair mismatch site, and 5'-tAGTACTt-3' as a double base pair mismatch site. Polyamides 8 and 9 with contiguous Hp rings in positions 1-2 and 3-4 respectively, bind 5'-tAGAACTt-3' as a match, and 5'-tAGTACTt-3' and 5'-tAGATCTt-3' as single base pair mismatches. Our findings presented here help define the scope and limitations of the Hp amino acid in polyamide design for discriminating core A•T sequences in the minor groove of DNA.

Results and Discussion

Hydroxypyrrole Hairpin Synthesis: Polyamide 1 was prepared as previously described.⁽⁷⁾ Eight polyamide resins, ImOpPyPy- γ -ImPyPyPy- β -Pam, ImPyPyPy- γ -ImOpPyPy- β -Pam, ImOpPyPy- γ -ImOpPyPy- β -Pam, ImPyOpPy- γ -ImPyPyPy- β -Pam, ImPyPyPy- γ -ImPyOpPy- β -Pam, ImPyOpPy- γ -ImPyOpPy- β -Pam, ImOpOpPy- γ -ImPyPyPy- β -Pam, and ImPyPyPy- γ -ImOpOpPy- β -Pam, were synthesized from commercially available Boc- β -alanine-Pam resin (0.5 g resin, 0.25 mmol/g substitution) using manual, solid phase protocols in 18 steps (for a representative example, polyamide 2, see Figure 6.3).⁽⁵⁾ A sample of resin was then cleaved by a single step aminolysis reaction with ((dimethylamino)propylamine (55 °C, 18 hr.) and the reaction mixture subsequently purified by reversed phase HPLC to provide ImOpPyPy- γ -ImPyPyPy- β -Dp, ImPyPyPy- γ -ImOpPyPy- β -Dp, ImOpPyPy- γ -ImOpPyPy- β -Dp, ImPyOpPy- γ -ImPyPyPy- β -Dp, ImPyPyPy- γ -ImPyOpPy- β -Dp, ImPyOpPy- γ -ImPyOpPy- β -Dp, ImOpOpPy- γ -ImPyPyPy- β -Dp, and ImPyPyPy- γ -ImOpOpPy- β -Dp. Deprotection of the Op monomer with sodium thiophenoxide/DMF (100 °C, 2 hrs.)^[3e] and subsequent purification by reversed phase HPLC yields polyamides: ImHpPyPy- γ -ImPyPyPy- β -Dp (2), ImPyPyPy- γ -ImHpPyPy- β -Dp (3), ImHpPyPy- γ -ImHpPyPy- β -Dp (4), ImPyHpPy- γ -ImPyPyPy- β -Dp (5), ImPyPyPy- γ -ImPyHpPy- β -Dp (6), ImPyHpPy- γ -ImPyHpPy- β -Dp (7), ImHpHpPy- γ -ImPyPyPy- β -Dp (8), and ImPyPyPy- γ -ImHpHpPy-

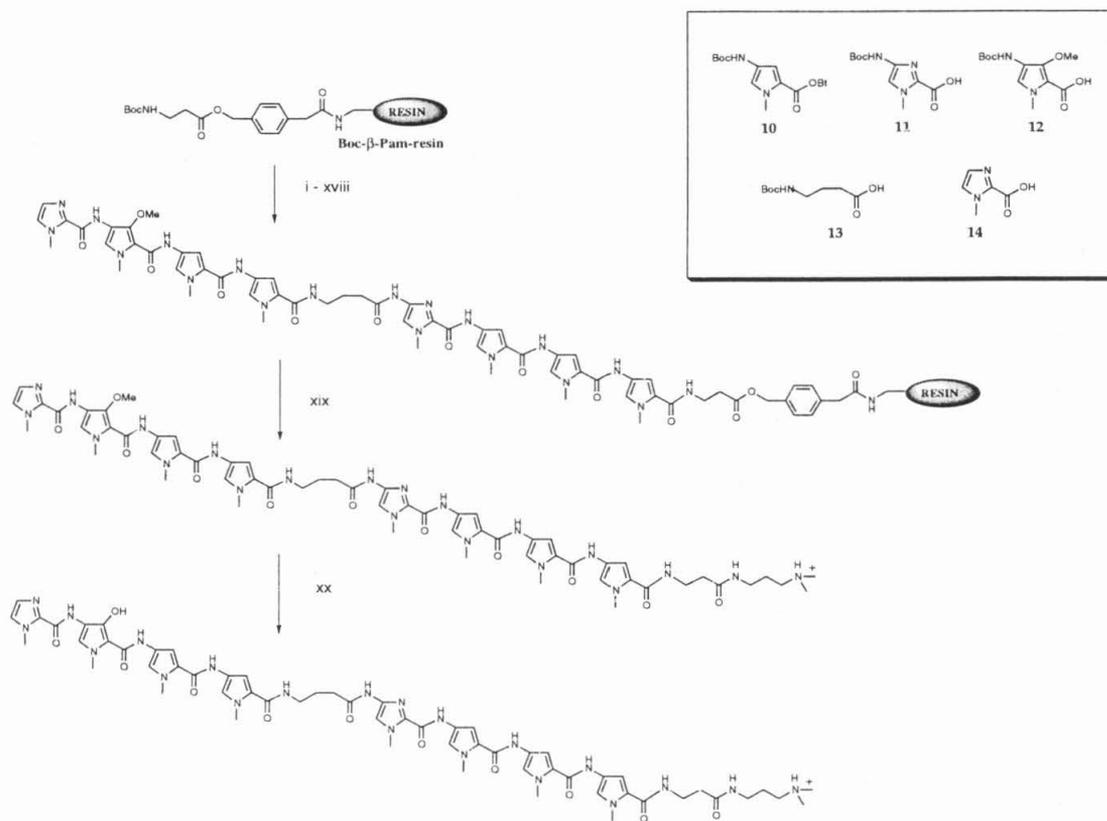


Figure 6.3. Hairpin synthetic scheme for ImHpPyPy- γ -ImPyPyPy- β -Dp (2): (i) 80 % TFA/DCM, 0.4 M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) 80 % TFA/DCM, 0.4 M PhSH; (iv) Boc-Py-OBt, DIEA, DMF; (v) 80 % TFA/DCM, 0.4 M PhSH; (vi) Boc-Py-OBt, DIEA, DMF; (vii) 80 % TFA/DCM, 0.4 M PhSH; (viii) Boc-Im-OH, DIEA, HBTU, DMF; (ix) 80 % TFA/DCM, 0.4 M PhSH; (x) Boc- γ -aminobutyric acid (HBTU, DIEA); (xi) 80 % TFA/DCM, 0.4 M PhSH; (xii) Boc-Py-OBt, DIEA, DMF; (xiii) 80 % TFA/DCM, 0.4 M PhSH; (xiv) Boc-Py-OBt, DIEA, DMF; (xv) 80 % TFA/DCM, 0.4 M PhSH; (xvi) Boc-Op-OH, DIEA, DMF; (xvii) 80 % TFA/DCM, 0.4 M PhSH; (xviii) Im-OH (HBTU/DIEA); (xix) N-N-((dimethylamio)propyl)amine, 55 °C; (xxi) Sodium thiophenoxide/DMF. (Inset) Pyrrole, Imidazole, Hydroxypyrrole, and diaminobutyric acid monomers for solid phase synthesis: Boc- γ -diaminobutyric acid 3-R, Boc-Pyrrole-OBt ester¹¹ (Boc-Py-OBt) 4, Boc-Imidazole acid (Boc-Im-OH), Imidazole-2-Carboxylic acid^{2a} (Im-OH) 5 and Boc-Methoxypyrrole acid (Boc-Op-OH).

β -Dp (9). Eight ring hairpin polyamides containing single or multiple Hp residues are soluble in aqueous solution at concentrations up to 1 mM.

Table 6.1. Equilibration Association Constants (M^{-1})^{a-c}

Polyamide	5'-tAGTACTt-3'	5'-tAGAACTt-3'	5'-tAGATCTt-3'
1 ImPyPyPy- γ -ImPyPyPy- β -Dp	<i>3.5×10^{10}</i> (0.7)	<i>4.7×10^9</i> (0.7)	<i>7.4×10^8</i> (1.5)

^aValues reported are the mean values obtained from three DNase I titration experiments. ^bThe assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂. ^cMatch site association constants are shown in bold type and italicized.

Equilibrium Association Constants: Quantitative DNase I footprint titrations (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂, pH 7.0 and 22 °C)⁽⁶⁾ were performed on ³²P end-labeled 285-bp DNA restriction fragment of pDEH10 to determine equilibrium association constants (K_a) for the polyamides 1 - 9 on three 6-bp binding sites. It might have been expected that the compound 1 would not discriminate 5'-NN-3' sequences based on the Py/Py degenerate recognition of A•T base pairs, but we find that the three 6-bp sites are bound at nanomolar concentrations with a 50-fold preference in the following order: 5'-tAGTACTt-3' ($K_a = 3.5 \times 10^{10} M^{-1}$) > 5'-tAGAACTt-3' ($K_a = 4.7 \times 10^9 M^{-1}$) > 5'-tAGATCTt-3' ($K_a = 7.4 \times 10^8 M^{-1}$). The inherent discrimination of binding sites is most likely due to DNA microstructure. Previous reports have shown that sequences containing 5'-GA-3' steps have narrower minor groove widths than those with 5'-GT-3' steps,⁽⁸⁾ and this narrow groove width may prevent optimal H-bonding between polyamide amino acids and the edges of base pairs in the minor groove. This provides a probable explanation for the observed binding preference for the three match sites which include a GT step > GA step (single) > GA step (double).⁽⁹⁾

Compounds 2 and 3, ImHpPyPy- γ -ImPyPyPy- β -Dp and ImPyPyPy- γ -ImHpPyPy- β -Dp, which contain single Hp substitutions in positions 1 or 3, respectively, bind the 6-bp sites with lower affinity and enhanced specificity, relative to polyamide 1

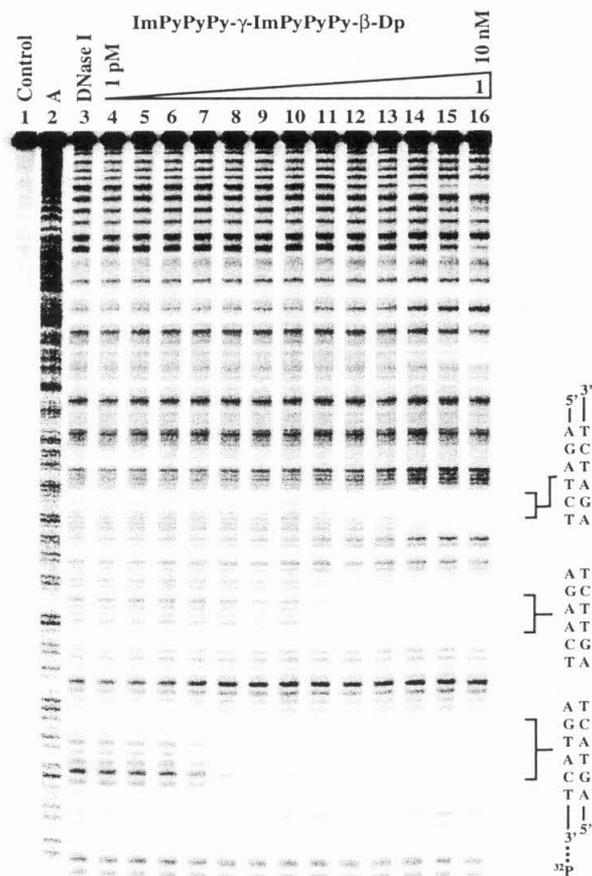


Figure 6.4. Footprinting experiment on the 3'- 32 P-labeled 285-bp DNA restriction fragment derived from the plasmid pDEH10. Quantitative DNase I footprint titration experiment with (a) ImPyPyPy- γ -ImPyPyPy- β -Dp (**1**): lane 1, intact; lane 2, A reaction; lane 3, DNase I standard, lanes 4-16, 1 pM, 2 pM, 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, and 10 nM. All reactions contain 30 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂. 5'-tAGTACTt-3', 5'-tAGAACTt-3', and 5'-tAGATCTt-3' binding sites are shown on the right side of the autoradiograms.

(Figures 6.5 and 6.6), and in the following order: 5'-tAGTACTt-3' (match) > 5'-tAGAACTt-3' (match) > 5'-tAGATCTt-3' (mismatch), where mismatched base pairs are underlined (Table 6.2). Discrimination of the 5'-tAGTACTt-3' match site using Hp residues in positions 1 or 3 results in a modest 5- to 6-fold decrease in affinity ($K_a = 7.9 \times 10^9 \text{ M}^{-1}$ and $K_a = 5.9 \times 10^9 \text{ M}^{-1}$ respectively). However, this modest compromise in

Table 6.2. Equilibration Association Constants (M^{-1})^{a-c}

Polyamide	5'-tAGTACTt-3'	5'-tAGAACTt-3'	5'-tAGATCTt-3'
2 ImHpPyPy- γ -ImPyPyPy- β -Dp	<i>7.2×10^9 (1.2)</i>	<i>3.6×10^8 (0.7)</i>	$\leq 1.0 \times 10^7$
3 ImPyPyPy- γ -ImHpPyPy- β -Dp	<i>5.9×10^9 (0.8)</i>	<i>3.9×10^8 (0.8)</i>	$\leq 1.0 \times 10^7$
4 ImHpPyPy- γ -ImHpPyPy- β -Dp	<i>7.0×10^8 (1.8)</i>	$\leq 1.0 \times 10^7$	$\leq 1.0 \times 10^7$

^aValues reported are the mean values obtained from three DNase I titration experiments. ^bThe assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂. ^cMatch site association constants are shown in bold type and italicized.

affinity is accompanied by a ≥ 720 - to ≥ 590 -fold specificity versus the single base pair mismatch 5'-tAGATCTt-3' site. Discrimination of the 5'-tAGAACTt-3' match site results in a 13-fold reduction in binding affinity for polyamides 2 and 3 in comparison to polyamide 1; however, this energetic penalty is accompanied by a ≥ 36 - to ≥ 39 -fold binding specificity for the single base pair mismatch 5'-tAGATCTt-3'. Compound 4, ImHpPyPy- γ -ImHpPyPy- β -Dp, with a 5'-stagger of Hp rings in positions 1 and 3, binds its match site with enhanced specificity relative to the parent polyamide 1, and bound all sites in the following order (Figure 6.7, Table 6.2): 5'-tAGTACTt-3' (match) > 5'-tAGAACTt-3' (mismatch) = 5'-tAGATCTt-3' (double mismatch). Positioning multiple Hp rings in this manner lowers affinity for the 5'-tAGTACTt-3' match site by a factor of 50 ($K_a = 7.0 \times 10^8 \text{ M}^{-1}$), but establishes a ≥ 70 -fold preference over both the single and double base pair mismatch sites.

ImPyHpPy- γ -ImPyPyPy- β -Dp (5) and ImPyPyPy- γ -ImPyHpPy- β -Dp (6), with Hp residues in positions 2 or 4, bind the 6-bp sites with overall compromised DNA

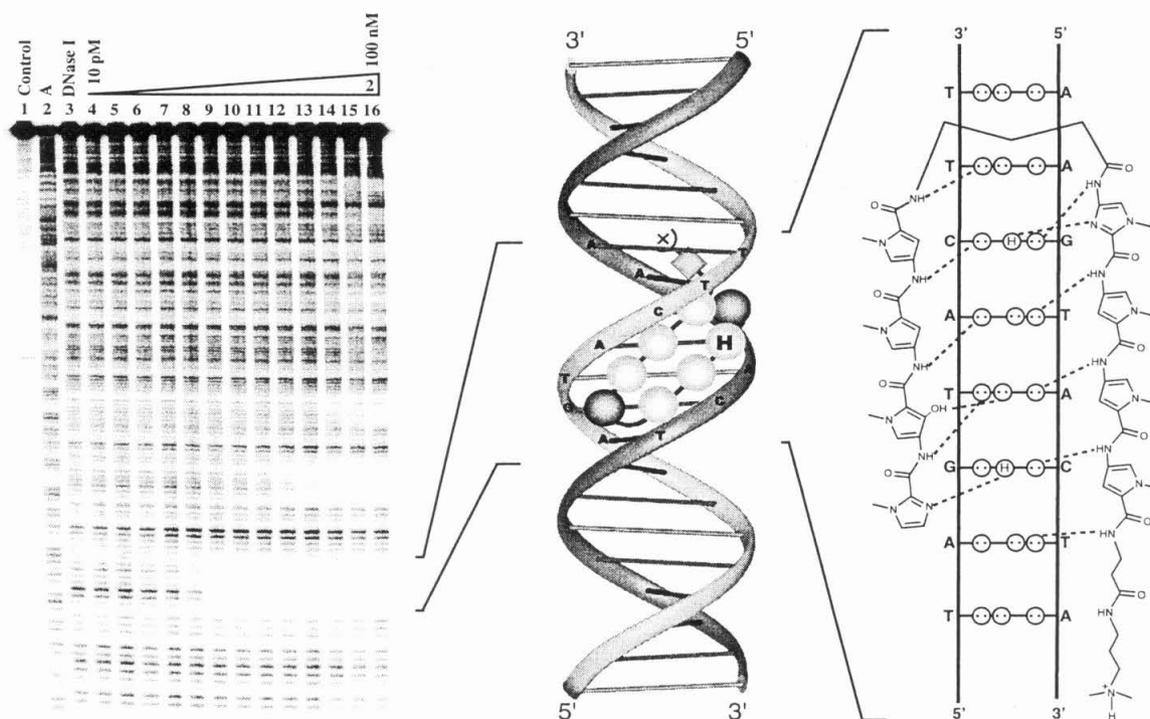


Figure 6.5. (Left) Footprinting experiment on the $3'$ - 32 P-labeled 285-bp DNA restriction fragment derived from the plasmid pDEH10. Quantitative DNase I footprint titration experiment with (a) ImHpPyPy- γ -ImPyPyPy- β -Dp (**2**): lane 1, intact; lane 2, A reaction; lane 3, DNase I standard, lanes 4-16, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, and 100 nM. All reactions contain 30 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂. Brackets on the right side of the auto radiograms outline 5'-tAGTACTt-3' match site. (Middle) Ribbon diagram model of hairpin•DNA complex. For schematic binding model, Im and Py rings are represented as shaded and unshaded spheres respectively, while Hp rings are annotated with an "H" in the center of an unshaded sphere. The β -alanine amino acid is represented by a partially shaded diamond. Binding site sequence is notated by bold capital letters along the helix. (Right) Hydrogen bonding models of the 1:1 polyamide-DNA complexes formed between ImHpPyPy- γ -ImPyPyPy- β -Dp (**2**) and 5'-tAGTACTt-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.

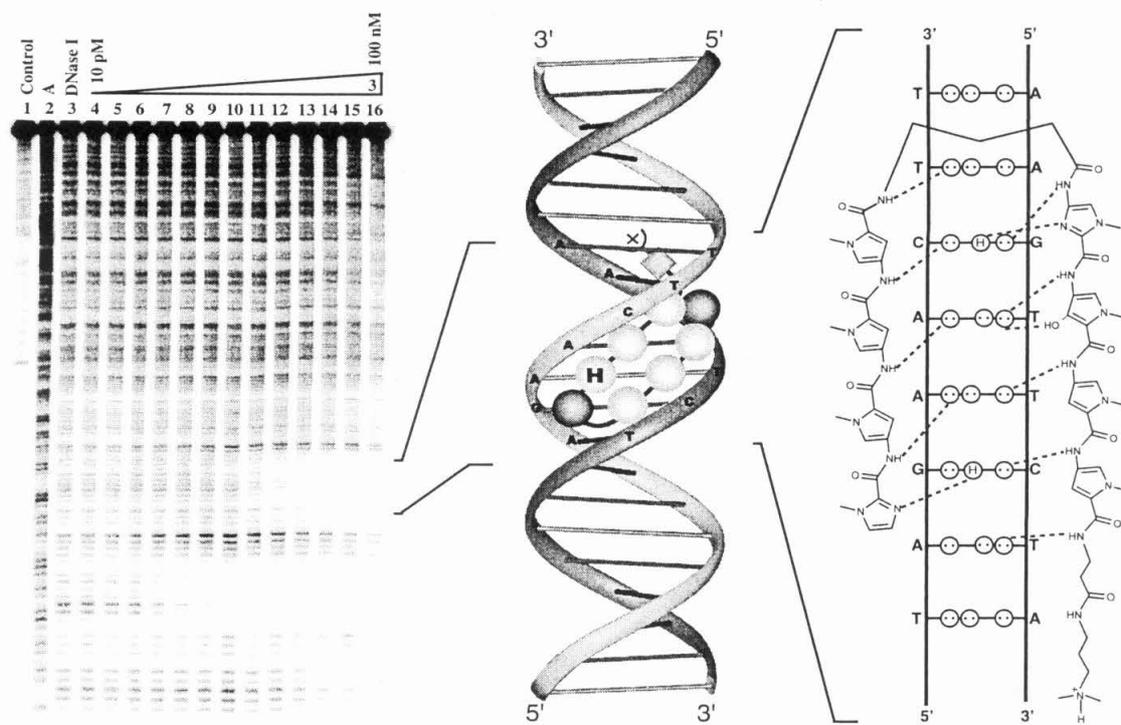


Figure 6.6. (Left) Footprinting experiment on the 3'-³²P-labeled 285-bp DNA restriction fragment derived from the plasmid pDEH10. Quantitative DNase I footprint titration experiment with (a) ImPyPyPy-γ-ImHpPyPy-β-Dp (3): lane 1, intact; lane 2, A reaction; lane 3, DNase I standard, lanes 4-16, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, and 100 nM. All reactions contain 30 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂. Brackets on the right side of the autoradiograms outline 5'-tAGTACTt-3' match site. (Middle) Ribbon diagram model of hairpin•DNA complex. For schematic binding model, Im and Py rings are represented as shaded and unshaded spheres respectively, while Hp rings are annotated with an "H" in the center of an unshaded sphere. The β-alanine amino acid is represented by a partially shaded diamond. Binding site sequence is annotated by bold capital letters along the helix. (Right) Hydrogen bonding models of the 1:1 polyamide-DNA complexes formed between ImHpPyPy-γ-ImPyPyPy-β-Dp (2) and 5'-tAGTACTt-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.

Table 6.3. Equilibration Association Constants (M^{-1})^{a-c}

	Polyamide	5'-tAGTACTt-3'	5'-tAGAACTt-3'	5'-tAGATCTt-3'
5	ImPyHpPy- γ -ImPyPyPy- β -Dp	2.7×10^8 (0.5)	<i>1.1×10^8 (0.3)</i>	<i>4.0×10^7 (0.4)</i>
6	ImPyPyPy- γ -ImPyHpPy- β -Dp	7.3×10^7 (1.4)	<i>7.4×10^7 (1.5)</i>	<i>2.5×10^7 (0.6)</i>
7	ImPyHpPy- γ -ImPyHpPy- β -Dp	1.0×10^8 (0.2)	2.6×10^7 (0.6)	<i>3.1×10^7 (0.7)</i>

^aValues reported are the mean values obtained from three DNase I titration experiments. ^bThe assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂. ^cMatch site association constants are shown in bold type and italicized.

binding properties in comparison to polyamides 2 and 3, and in the order (Figure 6.8, Table 6.3): 5'-tAGTACTt-3' (mismatch) > 5'-tAGAACTt-3' (match) > 5'-tAGATCTt-3' (match). Although compounds 5 and 6 do not exhibit complete specificity for the designed match sites, addition of Hp residues in position 2 or 4 reduces the inherent 50-fold preference of polyamide 1 for 5'-tAGTACTt-3' versus 5'-tAGATCTt-3' to a modest 3- to 5- fold preference by lowering affinity for 5'-tAGTACTt-3' site by a factor of 220 and the 5'-tAGATCTt-3' site only by a factor of 20. This “compression” of affinities represents a 10- to 12- fold enhancement in specificity for the 5'-tAGATCTt-3' match site for compounds 5 and 6, an increase analogous to that observed for compounds 1 and 3. Placement of the Hp amino acid in positions 2 or 4 also “compresses” the inherent binding preference seen in parent polyamide 1 by reducing the binding of the 5'-tAGTACTt-3' site versus the 5'-tAGAACTt-3' site from a factor of 7.4 to 2.5 (polyamide 5) or 1.0 (polyamide 6). This discrimination is accompanied by a larger energetic penalty than observed for polyamides 2 or 3, notably a 43- and 64-fold reduction in binding affinity for polyamides 5 and 6 respectively in comparison to the parent polyamide 1. Polyamide 7, ImPyHpPy- γ -ImPyHpPy- β -Dp, with a 3'- stagger of multiple Hp rings in positions 2 and 4, recognizes the binding sites with similar DNA binding properties to compounds 5 and 6 (Table 6.3): 5'-tAGTACTt-3' (double mismatch) > 5'-tAGATCTt-3' (match) > 5'-tAGAACTt-3' (mismatch). The double mismatch 5'-tAGTACTt-3' site is bound with only 3- fold higher affinity, ($K_a = 1.0 \times 10^8$

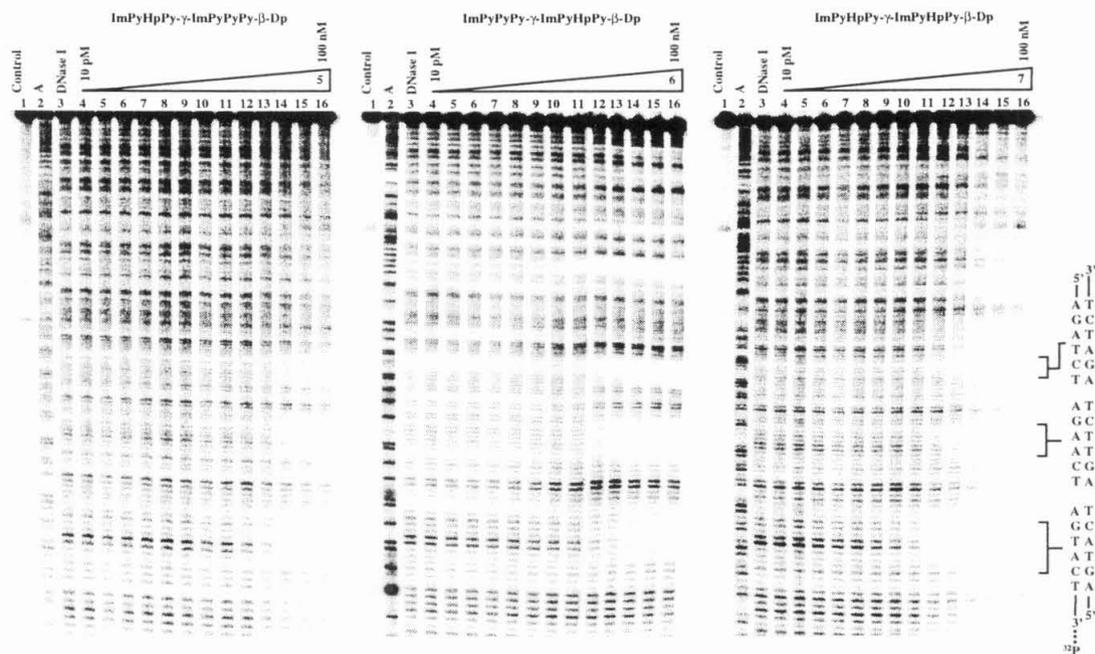


Figure 6.8. Footprinting experiments on the 3'-³²P-labeled 285-bp DNA restriction fragment derived from the plasmid pDEH10. Quantitative DNase I footprint titration experiments with (a) ImPyHpPy- γ -ImPyPyPy- β -Dp (5), (b) ImPyPyPy- γ -ImPyHpPy- β -Dp (5), and (c) ImPyHpPy- γ -ImPyHpPy- β -Dp (7): lane 1, intact; lane 2, A reaction; lane 3, DNase I standard, lanes 4-16, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, and 100 nM. All reactions contain 30 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂. The 5'-tAGAACTt-3', 5'-tAGTACTt-3', and 5'-tAGTACTt-3' binding sites are shown on the right side of the autoradiograms.

M^{-1}), relative to the match site 5'-tAGATCT-3', ($K_a = 3.1 \times 10^7 M^{-1}$), illustrating the compression of affinities displayed by **5** and **6**. The similar binding profiles of polyamides **5** and **6**, containing Hp residues in positions 2 or 4, and polyamide **7**, containing Hp/Py for the designed sites, suggests that the energetic destabilization of these polyamide•DNA complexes is caused by the initial placement of a single Hp ring in positions 2 or 4, and that addition of a second Hp residue in either position occurs with minimal energetic cost.

Table 6.4. Equilibration Association Constants (M^{-1})^{a-c}

Polyamide	5'-tAGTACTt-3'	5'-tAGAACTt-3'	5'-tAGATCTt-3'
8 ImHpHpPy-γ-ImPyPyPy-β-Dp	5.0×10^7 (2.4)	$\leq 1.0 \times 10^7$	$\leq 1.0 \times 10^7$
9 ImPyPyPy-γ-ImHpHpPy-β-Dp	$\leq 1.0 \times 10^7$	$\leq 1.0 \times 10^7$	$\leq 1.0 \times 10^7$

^aValues reported are the mean values obtained from three DNase I titration experiments. ^bThe assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂. ^cMatch site association constants are shown in bold type and italicized.

Contiguous arrangement of multiple Hp residues in positions 1-2 for compound **8**, affords binding to the three designated sites with two orders of magnitude less affinity than the parent hairpin **1**, and in the following order: (Table 6.4): 5'-tAGTACTt-3' (mismatch) > 5'-tAGAACTt-3' (match) = 5'-tAGATCTt-3' (mismatch). Compound **8** displays a 5-fold preference for the single base pair mismatch 5'-tAGTACTt-3' site as opposed to compound **9**, with Hp rings in positions 3 and 4, which establishes no preference for any binding site and binds in the following order (Table 6.4): 5'-tAGAACTt-3' (match) = 5'-tAGTACTt-3' (mismatch) = 5'-tAGATCTt-3' (mismatch). The slight preference of compound **8** is most likely due to flexibility of the N-terminus versus the C-terminus of the hairpin allowing for favorable polyamide•DNA contacts.

Specificity Enhancement of Py to Hp Substitutions: We observe a variety of effects on hairpin specificity depending on the number and position of Hp substitutions. Compounds **2**, **3**, **5**, and **6**, with single Hp substitutions, display a 4- to ≥ 15 -fold *gain* in

specificity for designated match sites over mismatch sites relative to polyamide 1. Substituting a Py ring for a Hp ring at positions 1 (polyamide 2) or 3 (polyamide 3), provides a ≥ 15 - fold and ≥ 13 - fold gain in specificity for the 5'-tGTACTt-3' match site versus the 5'-tGATCTt-3' single base pair mismatch site. These substitutions also provide a ≥ 6 - fold enhancement for the 5'-tGAACTt-3' match site versus the identical mismatch sequence. Replacement of Py with Hp at positions 2 (polyamide 5) or 4 (polyamide 6), yields a 7- and 15- fold specificity enhancement for the 5'-tGATCt-3' match site, and a 4- and 8-fold gain for the 5'-tGAACTt-3' match site versus the 5'-tGTACTt-3' mismatch site. For multiple Hp substitutions in positions 1 and 3 (polyamide 4), we observe a ≥ 9 -fold and ≥ 2 -fold gain in specificity for the 5'-tGTACTt-3' versus the 5'-tGAACTt-3' single base pair mismatch and the 5'-tGATCTt-3' double base pair mismatch sites, respectively. Multiple substitutions at 2 and 4 (polyamide 7) provides for 6- and 15- fold specificity gain for 5'-tGTACTt-3' versus the 5'-tGAACTt-3' single and 5'-tGTACTt-3' double base pair mismatch sites. Contiguous substitution of Hp pairs in positions 1-2 (polyamide 8) or 3-4 (polyamide 9) yields no increase in specificity for any of the target sequences.

Implications for Polyamide Design: We have presented a first step towards understanding Hp/Py recognition properties within DNA sequence context. By substituting core Py amino acids with Hp residues, we observe that a Py to Hp modification at positions 1 or 3 target the T•A base pair of the 5'-GT-3' step in 5'-tGTACTt-3' with a small energetic penalty relative to a Py ring in these positions. Single Py to Hp substitution at positions 2 or 4 that target A•T pairs of the 5'-GA-3' step in 5'-tGATCTt-3', results in a slightly larger (3- to 5-fold) energetic penalty. Introduction of a second Hp ring in a 5'- stagger arrangement, synergistically destabilizes the hairpin•DNA complex at the 5'-tGTACTt-3' match site. In contrast to the 5'- stagger, the subsequent addition of another Hp ring in the 3'- stagger does not further

destabilize polyamide binding beyond the initial introduction of an Hp ring at either of these positions. Contiguous arrangement of multiple Hp/Py pairs in an eight ring hairpin is an ineffectual method for recognizing the 5'-tAGAACTt-3' site in an eight ring hairpin. However, use of a cyclic polyamide motif containing contiguous Hp/Py pairs efficiently discriminates the 5'-tAGAACTt-3' binding site.^(3f)

Overall, we observe comparable specificity enhancements for polyamides containing single Py to Hp substitutions when discriminating 5'-WGNNCW-3' where NN = AT or TA. However, it has been demonstrated that binding site context and flanking sequences affect minor groove width⁸ and potentially play a significant role in the formation of 1:1 hairpin:DNA complexes. Therefore, a more comprehensive analysis of how sequence context effects Hp/Py discrimination of A•T base pairs awaits further research.

Experimental Section

General: Dicyclohexylcarbodiimide (DCC), Hydroxybenzotriazole (HOBT), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU), Boc anhydride (Boc2O) and 0.25 mmol/gram Boc- β -alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc- β -Pam-Resin) was purchased from Peptides International. γ -Aminobutyric acid (γ) were purchased from Bachem. *N,N*-diisopropylethylamine (DIEA) and *N,N*-dimethylformamide (DMF) were purchased from Applied Biosystems. Dichloromethane (DCM) was reagent grade from EM; thiophenol (PhSH), dimethylaminopropylamine (Dp), were from Aldrich. Trifluoroacetic acid (TFA) Biograde was from Halocarbon. All reagents were used without further purification.

Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. A shaker for manual solid phase synthesis was obtained from St. John Associates, Inc. Screw-cap glass peptide synthesis reaction vessels (5 mL and 20 mL) with a #2 sintered

glass frit were made as described by Kent.^[10] UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on a Beckman Gold system using a RAINEN C₁₈, Microsorb MV, 5 μ m, 300 x 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, gradient elution 1.25% acetonitrile/min. Preparatory reversed phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100 μ m C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. Milli-Q water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μ m filtered.

Polyamide Synthesis: Reagents and protocols for polyamide synthesis were as previously described.^{(3c),(5),(7)} Polyamides were purified by reversed phase preparatory HPLC on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100 μ m C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. Extinction coefficients were calculated based on $\epsilon = 8333/\text{ring}$ at 304 nm.⁽¹¹⁾

ImHpPyPy- γ -ImPyPyPy- β -Dp (2): Recovered upon lyophilization of the appropriate fractions as a white powder (1 mg, 10% yield, 2 steps). MALDI-TOF-MS [M+H]⁺ (monoisotopic), 1238.6: 1238.6 calc. for C₅₈H₇₂N₂₁O₁₁⁺.

ImPyPyPy- γ -ImHpPyPy- β -Dp (3): Recovered upon lyophilization of the appropriate fractions as a white powder (1 mg, 10% yield, 2 steps). MALDI-TOF-MS [M+H]⁺ (monoisotopic), 1238.9: 1238.6 calc. for C₅₈H₇₂N₂₁O₁₁⁺.

ImHpPyPy- γ -ImHpPyPy- β -Dp (4): Recovered upon lyophilization of the appropriate

fractions as a white powder (5 mg, 20% yield, 2 steps). MALDI-TOF-MS $[M+H]^+$ (monoisotopic), 1254.5: 1254.6 calc. for $C_{58}H_{72}N_{21}O_{12}^+$.

ImPyHpPy- γ -ImPyPyPy- β -Dp (5): Recovered upon lyophilization of the appropriate fractions as a white powder (.5 mg, 4% yield, 2 steps). MALDI-TOF-MS $[M+H]^+$ (monoisotopic), 1238.8: 1238.6 calc. for $C_{58}H_{72}N_{21}O_{11}^+$.

ImPyPyPy- γ -ImPyHpPy- β -Dp (6): Recovered upon lyophilization of the appropriate fractions as a white powder (.2 mg, 5% yield, 2 steps). MALDI-TOF-MS $[M+H]^+$ (monoisotopic), 1238.6: 1238.6 calc. for $C_{58}H_{72}N_{21}O_{11}^+$.

ImPyHpPy- γ -ImPyHpPy-- β -Dp (7): Recovered upon lyophilization of the appropriate fractions as a white powder (.5 mg, 7% yield, 2 steps). MALDI-TOF-MS $[M+H]^+$ (monoisotopic), 1254.5: 1254.6 calc. for $C_{58}H_{72}N_{21}O_{12}^+$.

ImHpHpPy- γ -ImPyPyPy- β -Dp (8): Recovered upon lyophilization of the appropriate fractions as a white powder (6 mg, 15% yield, 2 steps). MALDI-TOF-MS $[M+H]^+$ (monoisotopic), 1254.4: 1254.6 calc. for $C_{58}H_{72}N_{21}O_{12}^+$.

ImPyPyPy- γ -ImHpHpPy- β -Dp (9): Recovered upon lyophilization of the appropriate fractions as a white powder (1 mg, 9% yield, 2 steps). MALDI-TOF-MS $[M+H]^+$ (monoisotopic), 1254.5: 1254.6 calc. for $C_{58}H_{72}N_{21}O_{12}^+$.

DNA Reagents and Materials: Enzymes were purchased from Boehringer-Mannheim and used with their supplied buffers. Deoxyadenosine and thymidine 5'- $[\alpha\text{-}^{32}\text{P}]$ triphosphates were obtained from Amersham, 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.⁽¹²⁾

Construction of Plasmid DNA: The plasmid pDEH10 was constructed as previously described.

Preparation of 3'-End-Labeled Restriction Fragments. The plasmid pDEH10 was linearized with *Eco* RI and *Pvu* II, then treated with the Sequenase enzyme, deoxyadenosine 5'-[α -³²P]triphosphate and thymidine 5'-[α -³²P]triphosphate for 3' labeling. The labeled 3' fragment was loaded onto a 6% non-denaturing polyacrylamide gel, and the desired 285 base pair band was visualized by autoradiography and isolated. Chemical sequencing reactions were performed according to published methods.⁽¹³⁾

DNase I Footprinting:⁽⁶⁾ All reactions were carried out in a volume of 400 μ L. We note explicitly that no carrier DNA was used in these reactions until after DNase I cleavage. 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 MgCl₂, 5 mM CaCl₂, and 30 kcpm 3'-radiolabeled DNA. The solutions were allowed to equilibrate for 4 hours at 22 °C. Cleavage was initiated by the addition of 10 μ L of a DNase I stock solution (diluted with 1 mM DTT to give a stock concentration of 1.875 u/mL) and was allowed to proceed for 7 min at 22 °C. The reactions were stopped by adding 50 μ L of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/mL glycogen, and 30 μ M 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 6 min, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V for

1 hour. The gels were dried under vacuum at 80°C, then quantitated using storage phosphor technology.

Equilibrium association constants were determined as previously described.^[6a] The data were analyzed by performing volume integration of the 5'-AGAACT-3', 5'-AGTACT-3', and 5'-AGATCT-5' sites and a reference site. The apparent DNA target site saturation, θ_{app} , was calculated for each concentration of polyamide using the following equation:

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

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{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 fit to a Langmuir binding isotherm (eq 2, $n=1$ for polyamides **1 - 9**, by minimizing the difference between θ_{app} and θ_{fit} , using the modified Hill equation:

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

where $[L]_{tot}$ corresponds to the total polyamide concentration, K_a corresponds to the equilibrium association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software (version 2.1, Abelbeck software) with K_a , θ_{max} , and θ_{min} as the adjustable parameters.

At least three 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. The data were normalized using the following equation:

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

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 12-20 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 Treadway Foundation (Bio.47123-1-ENDOW.471230) for a research award to D.M.H., and the National Institutes of Health (GM19789-02) for a postdoctoral fellowship to C.M. We thank G.M. Hathaway for MALDI-TOF mass spectrometry and S. White for construction of pDEH10.

References

1. a) J. M. Gottesfeld, L. Nealy, J. W. Trauger, E. E. Baird, P. B. Dervan, *Nature* **1997**, 387, 202. b) L. A. Dickinson, R. J. Gulizia, J. W. Trauger, E. E. Baird, D. E. Mosier, J. M. Gottesfeld, P. B. Dervan *Proc. Natl. Acad. Sci. USA* **1998**, 95, 12890.
2. P. B. Dervan, R. W. Burli, *Curr. Opin. Chem. Biol* **1999**, 3, 688.
3. a) S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, *Nature* **1998**, 391, 468. b) C. L. Kielkopf, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, D. C. Rees *Science* **1998**, 282, 111. c) A. R. Urbach, J. W. Szewczyk, S. White, J. M. Turner, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1999**, 121, 11621. d) S. White, J. M. Turner, J. W. Szewczyk, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1999**, 121, 260. e) C. L. Kielkopf, R. E., Bremer, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, D. C. Rees *J. Mol. Biol.* **1999**, 295, 557. f) C. Melander, D. M. Herman, P. B. Dervan, *Chem. Eur. J* **2000**, in press.
4. J. W. Trauger, E. E. Baird, M. Mrksich, P. B. Dervan, *J. Am Chem. Soc.* **1996**, 118, 6160.
5. E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, 118, 6141.
6. 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. U.S.A.* **1986**, 83, 8462. c) D. F. Senear, M. Brenowitz, M. A. Shea, G. K. Ackers, *Biochemistry* **1986**, 25, 7344.
7. J. W. Trauger, E. E. Baird, P. B. Dervan, *Nature* **1996**, 382, 559.
8. a) G. G. Prive, K. Yanagi, R. E. Dickerson, *J. Mol. Biol.* **1991**, 217, 177. b) K. Grzeskowiak, K. Yanagi, G. G. Prive, R. E. Dickerson, *J. Biol. Chem.* **1991**, 217, 8861.
9. S. White, E. E. Baird, P. B. Dervan, *Biochemistry* **1996**, 35, 12532.
10. S.B.H. Kent, *Annu. Rev. Biochem.* **1988**, 57, 957.
11. D. S. Pilch, N. A. Pokar, C. A. Gelfand, S. M. Law, K. J. Breslauer, E. E. Baird, P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, 93, 8306.
12. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.
13. a) B. L. Iverson, P. B. Dervan, *Nucl. Acids Res.* **1987**, 15, 7823. b) A. M. Maxam, W.S. Gilbert, *Methods Enzymol.* **1980**, 65, 499.
14. W. S. Wade, M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc.* **1992**, 114, 8783.

Chapter 7

Targeting the HIV-1 TATA Box with a 2- β -2 Cyclic Polyamide Motif

Abstract: *Hairpin polyamides containing pyrrole (Py) and imidazole (Im) amino acids have been designed to regulate gene expression of HIV-1 promoter by inhibiting TATA-binding protein (TBP) recognition of the TATA-box element. First generation polyamide targeting the TATA-box flanking sequences 5'-WGCWGCW-3' consisted of an unsubstituted ten-ring hairpin that recognized the aforementioned sites with low affinity. Introduction of β -alanine residues into the second generation polyamide yields a 2- β -2 hairpin motif capable of binding target sequences with high affinity and specificity and inhibiting TBP binding in vitro and in vivo. Here we introduce a cyclic 2- β -2 motif capable of recognizing the TATA-box flanking 5'-WGCWGCW-3 sites with enhanced DNA binding properties. Quantitative footprint titrations demonstrate that cyclization of 2- β -2 hairpin motif modestly increases affinity for the target site. The enhanced affinity is achieved with a compromise in sequence selectivity, which is 2-fold relative to the hairpin analog for binding the single base pair mismatch sequence 5'-AGCTGTT-3'. The design and characterization of the 2- β -2 cycle serves as first step in the development of new cyclic polyamide motifs for recognizing endogenous promoters for in vitro and in vivo studies.*

Introduction

Cell permeable small molecules that recognize specific sequences of DNA are potentially useful therapeutic agents in molecular medicine.¹ Among these molecules, antisense oligonucleotides,² peptide nucleic acids (PNA),³ calicheamiycin oligosacharrides,⁴ metallointercalators,⁵ and minor groove binding compounds chromomycin,⁶ doxotubicin,⁷ and polyamides¹ have all shown promise for their ability to control biological processes at the genetic level. Of these DNA binding ligands, polyamides represent the only class of small molecules that are rationally designed to target predetermined sequences of DNA with affinities and specificities comparable to naturally occurring DNA binding proteins.¹ The simple rules that control the recognition properties of the minor groove binding polyamides are based on a code of side-by-side amino acid pairings of pyrrole (Py), imidazole (Im), and 3-hydroxypyrrole (Hp).¹ A pyrrole opposite an imidazole (Py/Im) targets a C•G bp, and an imidazole opposite a pyrrole (Im/Py) targets a G•C bp. A Py/Py pair degenerately recognizes A•T and T•A base pairs. An Hp/Py pair breaks this degeneracy by recognizing a T•A bp, while a Py/Hp pair specifies an A•T bp. The b/b pair is degenerate for A•T/T•A recognition, but perhaps more importantly acts as a “molecular spring” in allowing proper register of polyamide subunits with their respective base pairs.¹ These pairing rules are supported by NMR structural studies⁸ as well as X-ray crystallography.^{9,10} The linear combination of these amino acids utilizing a solid phase synthesis¹¹ strategy form oligomers that bind as side-by-side dimers in the minor groove of DNA. Covalently tethering unlinked dimers with aminobutyric acid (γ) turn residue forms hairpin structures that bind target sequences with enhanced binding properties.¹ Replacement of γ with diaminobutyric acid ($((R)^{H_2N})\gamma$) relocates the charge to the turn and frees the C-terminus of the polyamide for modification.¹ Addition of a second γ -turn to link hairpin termini forms a cyclic polyamide structure which further enhances the affinity and specificity for designed sequences relative to their hairpin counterparts.¹² The implementation of structural

elements in the ongoing effort to develop novel polyamide motifs is critical to evaluating the biological relevance of polyamides as potential therapeutics in both in vitro and in vivo systems.

It has been shown that polyamides effectively block eukaryotic cellular and viral transcription factors from binding their target sites and inhibit transcription, both in vitro and in cell culture experiments. In the model *Xenopus* 5S RNA system, an eight ring hairpin polyamide designed to bind the 5s RNA promoter at subnanomolar concentrations inhibits binding of the zinc finger transcription factor TFIIIA and downregulates transcription by RNA polymerase III. Moreover this polyamide disrupts transcription complexes on the chromosomal 5S RNA genes in vivo with *Xenopus* fibroblasts in culture.¹³ The 2- β -2 hairpin polyamide motif was constructed to target the adjacent sequences of the HIV-1 TATA-box element (Figure 7.1). Binding of the TBP subunit of TFIID in the minor groove nucleates assembly of the RNA polymerase II machinery for genes containing the TATA-box element. The 2- β -2 hairpin polyamide motif effectively blocks TBP/TFIID binding and inhibits basal transcription by RNA pol II.¹³ On this same HIV-1 promoter, an eight ring hairpin polyamide was designed to target the transactivator lymphoid enhancer factor LEF-1, and was shown to occlude LEF-1 DNA binding activity. Most importantly, when the 2- β -2 and the eight ring hairpin were used in combination, they synergistically inhibited HIV-1 replication in isolated human lymphoid cells.¹⁴ This provides impetus to explore the utility of several polyamide motifs for gene regulation studies.

Polyamide Design for Targeting HIV-1 TATA-Box: Inspection of the HIV-1 promoter, revealed two 7-bp sequences immediately flanking the TATA-box, 5'-TGCTGCA-3' and 5'-AGCAGCT-3', that are present on at least one side of the TATA-box for all strains of HIV-1 and therefore are candidates for targeting with polyamides. First generation

hairpin polyamide targeting these sites was a ten-ring unsubstituted 5-g-5 motif polyamide ImPyPyImPy- γ -ImPyPyImPy- β -Dp (γ -aminobutyric acid, β -beta alanine, and Dp-dimethylaminopropylamine). Unfortunately, the sites were bound with very low affinity, requiring the development of a second-generation motif. It has been shown that substituting the pyrrole residue that precedes the imidazole residue with β -alanine resets the imidazole binding register and increases affinity.¹ Replacing two pyrrole residues with the single base pair spanning β -alanine amino acid generated the 2- β -2 hairpin motif polyamide ImPy- β -ImPy- γ -ImPy- β -ImPy- β -Dp, that binds both the cognate match HIV-1 flanking sequences with good affinity and specificity.¹⁴ The addition of a second γ -turn residue tethering the termini of hairpin yields a cyclic polyamide motif with enhanced DNA binding properties. This addition represents the third generation polyamide to target the HIV-1 TATA-box.

Here we present a new 2- β -2 cyclic polyamide motif for targeting the flanking sequences of the HIV promoter TATA-box element. Four polyamides: a control reduced end hairpin ImPy- β -ImPy- γ -ImPy- β -ImPy-C3OH (**1**), a singly charged cyclo-(γ -ImPy- β -ImPy-(R)^{H2N}- γ -ImPy- β -ImPy-) (**2**), a doubly charged cyclo-((R)^{H2N}- γ -ImPy- β -ImPy-(R)^{H2N}- γ -ImPy- β -ImPy-) (**3**), and a control mismatched cyclo-(γ -PyPy- β -PyPy-(R)^{H2N}- γ -ImIm- β -ImIm-) (**4**) were synthesized by solid phase protocols¹¹ and their DNA binding properties were characterized by quantitative DNase I footprint¹⁵ titration experiments on a restriction fragment containing the 7-bp 5'-TGCTGCA-3', 5'-AGCTGTT-3', and 5'-TGGTGGA-3' binding sites (Figures 7.1 and 7.2). Based on the polyamide pairing rules, polyamide **1**, cyclo-**2**, and cyclo-**3** are designed to recognize 5'-TGCTGCA-3' as a match, 5'-AGCTGTT-3' as a single base pair mismatch, and 5'-TGGTGGA-3' as a double base pair mismatch. Cyclo-**4** targets 5'-TGGTGGA-3' as its cognate match site, and 5'-TGCTGCA-3' and 5'-AGCTGTT-3' as double base pair mismatches.

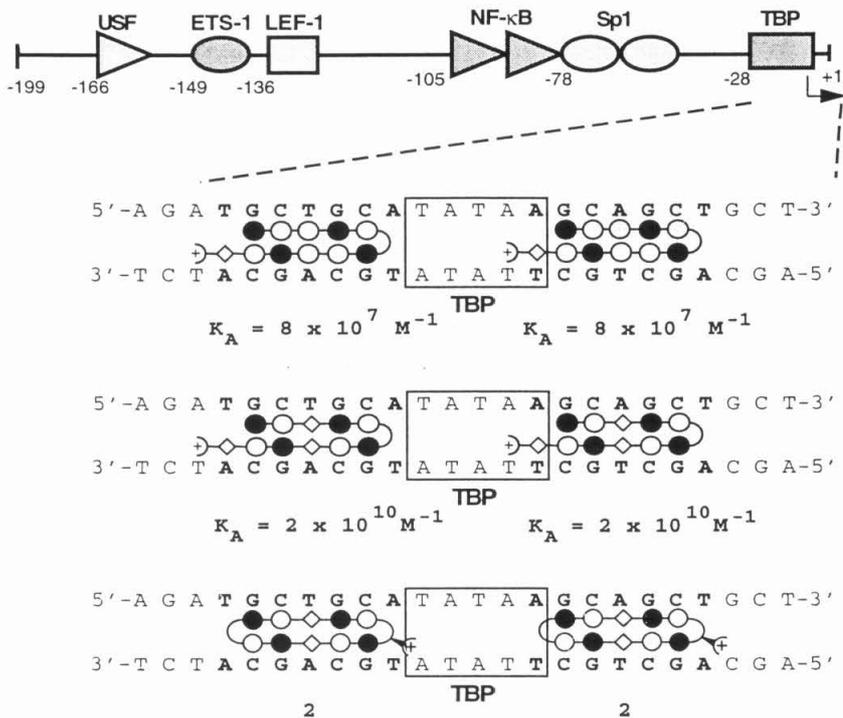
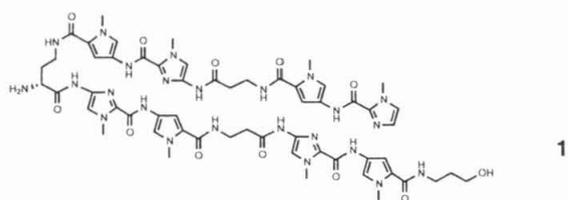
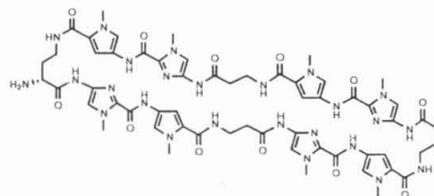


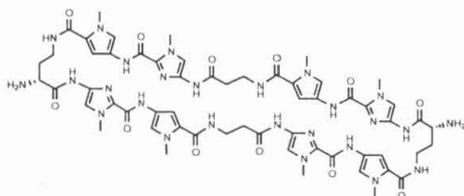
Figure 7.1 (Top) Schematic illustration of the HIV-1 promoter/enhancer. (Bottom) The DNA sequence of the TATA-box element of the HIV-1 promoter is shown along with the binding models for ImPyPyImPy- γ -ImPyPyImPy- β -Dp, ImPy- β -ImPy- γ -ImPy- β -ImPy- β -Dp, and cyclo-(γ -ImPy- β -ImPy-(R)^{H2N}- γ -ImPy- β -ImPy-) (2).



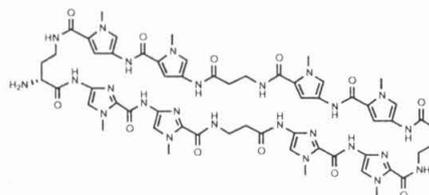
ImPy-β-ImPy-(R)^{H₂N}γ-ImPy-β-ImPy-C3-OH



cyclo-(γ-ImPy-β-ImPy-(R)^{H₂N}γ-ImPy-β-ImPy)



cyclo-((R)^{H₂N}γ-ImPy-β-ImPy-(R)^{H₂N}γ-ImPy-β-ImPy)



cyclo-(γ-PyPy-β-PyPy-(R)^{H₂N}γ-ImIm-β-ImIm)

Figure 7.2 Structure of the ten-ring hairpin control ImPyPyImPy-γ-ImPyPyImPy-β-Dp (**1**), and three 2-β-2 cyclic polyamide analogs cyclo-(γ-ImPy-β-ImPy-(R)^{H₂N}γ-ImPy-β-ImPy-) (**2**), cyclo-((R)^{H₂N}γ-ImPy-β-ImPy-(R)^{H₂N}γ-ImPy-β-ImPy-) (**3**), and cyclo-(γ-PyPy-β-PyPy-(R)^{H₂N}γ-ImIm-β-ImIm-) (**4**) as synthesized by solid phase protocols.

Results and Discussion

Synthesis of 2-β-2 Reduced-end Hairpin. ImPy-β-ImPy-(R)^{H₂N}γ-ImPy-β-ImPy-β-Pam-resin was synthesized by machine assisted protocols¹¹ in 22 steps from commercially available Boc-β-Ala-Pam resin.^{12a} The polyamide was cleaved from the resin by a single step reduction with lithium borohydride (EtOH, 60 °C), followed by reversed phase HPLC purification to yield the hairpin-polyamide ImPy-β-ImPy-(R)^{H₂N}γ-ImPy-β-ImPy-OH (1).

2-β-2 Cycle-Polyamide Synthesis. Three polyamide resins, Cbzγ-ImPy-β-ImPy-(R)^{Fmoc}γ-ImPy-β-ImPy-Pam-resin, Cbz-(R)^{Fmoc}γ-ImPy-β-ImPy-(R)^{Fmoc}γ-ImPy-β-ImPy-Pam-resin, and Cbzγ-PyPy-β-PyPy-(R)^{Fmoc}γ-ImIm-β-ImIm-Pam-resin, were synthesized in 20 steps from Boc-Py-Pam-resin (600 mg of resin, 0.1 mmol/g of substitution) using Boc-chemistry and manual solid phase synthesis protocols (Figure 7.3).^{11,12a} The (R)-2,4-diaminobutyric acid residue was introduced as an orthogonally protected *N*-α-Fmoc-*N*-γ-Boc derivative **10** (HBTU, DIEA). The final step was introduction of Cbzγ-Im acid **11** as a dimer block (HBTU, DIEA).¹¹ Fmoc protected polyamide-resins Cbzγ-ImPy-β-ImPy-(R)^{Fmoc}γ-ImPy-β-ImPy-Pam-resin, Cbz-(R)^{Fmoc}γ-ImPy-β-ImPy-(R)^{Fmoc}γ-ImPy-β-ImPy-Pam-resin, and Cbzγ-PyPy-β-PyPy-(R)^{Fmoc}γ-ImIm-β-ImIm-Pam-resin were treated with 1:4 DMF:Piperidine (22 °C, 30 min) to provide Cbzγ-ImPy-β-ImPy-(R)^{H₂N}γ-ImPy-β-ImPy-Pam-resin, Cbz(R)^{H₂N}γ-ImPy-β-ImPy-(R)^{H₂N}γ-ImPy-β-ImPy-Pam-resin, and Cbzγ-PyPy-β-PyPy-(R)^{H₂N}γ-ImIm-β-ImIm-Pam-resin, respectively. The amine-resins were then treated with Boc-anhydride (DIEA, DMF, 55 °C, 30 min) providing Cbzγ-ImPy-β-ImPy-(R)^{Boc}γ-ImPy-β-ImPy-Pam-resin, Cbz(R)^{Boc}γ-ImPy-β-ImPy-(R)^{Boc}γ-ImPy-β-ImPy-Pam-resin, and Cbzγ-PyPy-β-PyPy-(R)^{Boc}γ-ImIm-β-ImIm-Pam-resin. A sample of each resin was then taken and the respective peptides were liberated from the resin with concomitant removal of the CBZ protecting group by reductive cleavage employing Pd(OAc)₂/DMF/water/ammonium formate (37 °C, 14 hrs.).^{12a} Following removal of the

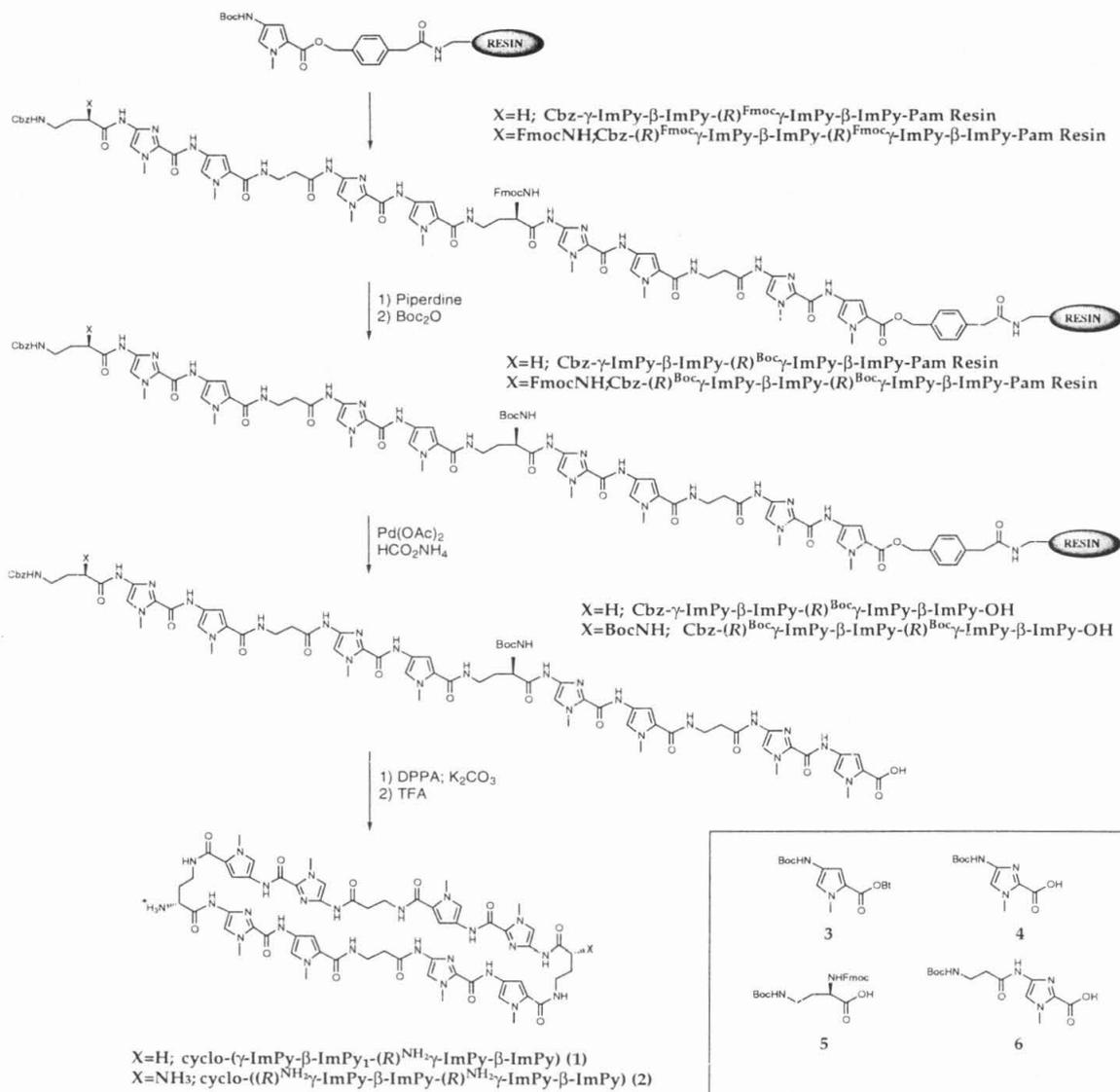


Figure 7.3. 2- β -2 Cycle synthetic scheme of cyclo-(γ -ImPy- β -ImPy-(R)^{H₂N} γ -ImPy- β -ImPy-) (2) and cyclo-((R)^{H₂N} γ -ImPy- β -ImPy-(R)^{H₂N} γ -ImPy- β -ImPy-) (3): (i) 80 % TFA/DCM, 0.4 M PhSH; (ii) Boc-Im-OH, DIEA, HBTU, DMF; (iii) 80 % TFA/DCM, 0.4 M PhSH; (iv) Boc- β -alanine-OH, HBTU, DIEA, DMF; (v) 80 % TFA/DCM, 0.4 M PhSH; (vi) Boc-Py-OBt, DIEA, DMF; (vii) 80 % TFA/DCM, 0.4 M PhSH; (viii) Boc-Im-OH, DIEA, HBTU, DMF; (ix) 80 % TFA/DCM, 0.4 M PhSH; (x) Fmoc- α -Boc- γ -diaminobutyric acid (HBTU, DIEA); (xi) 80 % TFA/DCM, 0.4 M PhSH; (xii) Boc-Py-OBt, DIEA, DMF; (xiii) 80 % TFA/DCM, 0.4 M PhSH; (xiv) Boc-Im-OH, HBTU, DIEA, DMF; (xv) 80 % TFA/DCM, 0.4 M PhSH; (xvi) Boc-Py-OBt, DIEA, DMF; (xvii) 80 % TFA/DCM, 0.4 M PhSH; (xviii) Boc- β -alanine-OH, HBTU, DIEA, DMF (xix) Boc-Py-OBt, DIEA, DMF; (xx) 80 % TFA/DCM, 0.4 M PhSH; (xxi) Boc- γ -Im-OH (2) or Boc- γ -(R)^{Fmoc} γ -Im-OH (3) HBTU, DIEA, DMF; (xxii) 80 % TFA/DCM, 0.4 M PhSH; (xxiii) CBZ-OSuc, DMF, (xxiiii) 80 % Piperidine:DMF (25 °C, 30 min) (xxiv) Boc anhydride, DMF, DIEA; (xxv) Pd(OAc)₂/DMF/ammonium fromate/H₂O; (xxvi) DPPA/K₂CO₃/DIEA; (xxviii) TFA. (Inset) Py, Im, and diaminobutyric acid monomers for solid phase synthesis: Boc-Pyrrole-OBt ester^[12] (Boc-Py-OBt) 11, Boc-Imidazole-OH (Boc-Im-OH) 12, (R)-Fmoc- α -Boc- γ -diaminobutyric acid 13, and Boc-aminobutyric acid-imidazole-2-carboxylic acid^[4a] (Bic- γ -Im-OH) 14.

resin by filtration, the crude reaction mixtures were purified by reverse phase HPLC to yield polyamides $\text{H}_2\text{N}-\gamma\text{-ImPy}-\beta\text{-ImPy}-(\text{R})^{\text{Boc}}\gamma\text{-ImPy}-\beta\text{-ImPy}-\text{OH}$ (**5**), $\text{H}_2\text{N}-(\text{R})^{\text{Boc}}\gamma\text{-ImPy}-\beta\text{-ImPy}-(\text{R})^{\text{Boc}}\gamma\text{-ImPy}-\beta\text{-ImPy}-\text{OH}$ (**6**), and $\text{H}_2\text{N}-\gamma\text{-PyPy}-\beta\text{-PyPy}-(\text{R})^{\text{Boc}}\gamma\text{-ImIm}-\beta\text{-ImIm}-\text{OH}$ (**7**). Each polyamide was then cyclized by treatment with DPPA/ K_2CO_3 /DMF (22 °C, 5 hours), all volatiles subsequently removed, the resulting product deprotected in neat TFA, and the crude product purified by reverse phase HPLC to yield cyclo-($\gamma\text{-ImPy}-\beta\text{-ImPy}-(\text{R})^{\text{H}_2\text{N}}\gamma\text{-ImPy}-\beta\text{-ImPy}$ -) (**2**), cyclo-($(\text{R})^{\text{H}_2\text{N}}\gamma\text{-ImPy}-\beta\text{-ImPy}-(\text{R})^{\text{H}_2\text{N}}\gamma\text{-ImPy}-\beta\text{-ImPy}$ -) (**3**), and cyclo-($\gamma\text{-PyPy}-\beta\text{-PyPy}-(\text{R})^{\text{H}_2\text{N}}\gamma\text{-ImIm}-\beta\text{-ImIm}$ -) (**4**).

Binding Energetics. Quantitative DNase I footprint titration¹⁵ experiments (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl_2 and 5 mM CaCl_2 , pH 7.0 and 22 °C) were performed on the 282-bp, 3'-32P end-labeled pJT2B2 Eco RI/Pvu II restriction fragment to determine the equilibrium association constants (K_a) of compound **1**, cyclo-**2**, cyclo-**3**, and cyclo-**4** for the 7-bp 5'-TGCTGCA-3', 5'-AGCTGTT-3', and 5'-TGGTGGA-3' match and mismatch sites (Figures 7.4, 7.5, 7.6, and Table 7.1). All polyamides bind their respective match and mismatch sequences with binding isotherms (eq 2, $n = 1$) consistent with binding in a 1:1 polyamide•DNA complex. The results reveal that hairpin polyamide **1** binds the sites in the following decreasing order: 5'-TGCTGCA-3' ($K_a = 6.7 \times 10^{10} \text{ M}^{-1}$) > 5'-AGCTGTT-3' ($K_a \sim 6.0 \times 10^9 \text{ M}^{-1}$) > 5'-TGGTGGA-3' ($K_a < 1.0 \times 10^9 \text{ M}^{-1}$). Cyclization generates cyclo-**2**, which binds target sequences with slightly higher affinities than compound **1** and in the following order (Figure 7.4): 5'-TGCTGCA-3' ($K_a = 7.5 \times 10^{10} \text{ M}^{-1}$) > 5'-AGCTGTT-3' ($K_a \sim 1.5 \times 10^{10} \text{ M}^{-1}$) > 5'-TGGTGGA-3' ($K_a < 1.0 \times 10^9 \text{ M}^{-1}$). Cyclo-**3** recognizes the designed sequences with similar affinities to cyclo-**2** (Figure 7.5): 5'-TGCTGCA-3' ($K_a = 4.1 \times 10^{10} \text{ M}^{-1}$) > 5'-AGCTGTT-3' ($K_a \sim 1.5 \times 10^{10} \text{ M}^{-1}$) > 5'-TGGTGGA-3' ($K_a < 1.0 \times 10^9 \text{ M}^{-1}$). The control polyamide cyclo-**4** also binds its cognate match with high affinities and in the following order (Figure 7.6): 5'-TGGTGGA-3' ($K_a =$

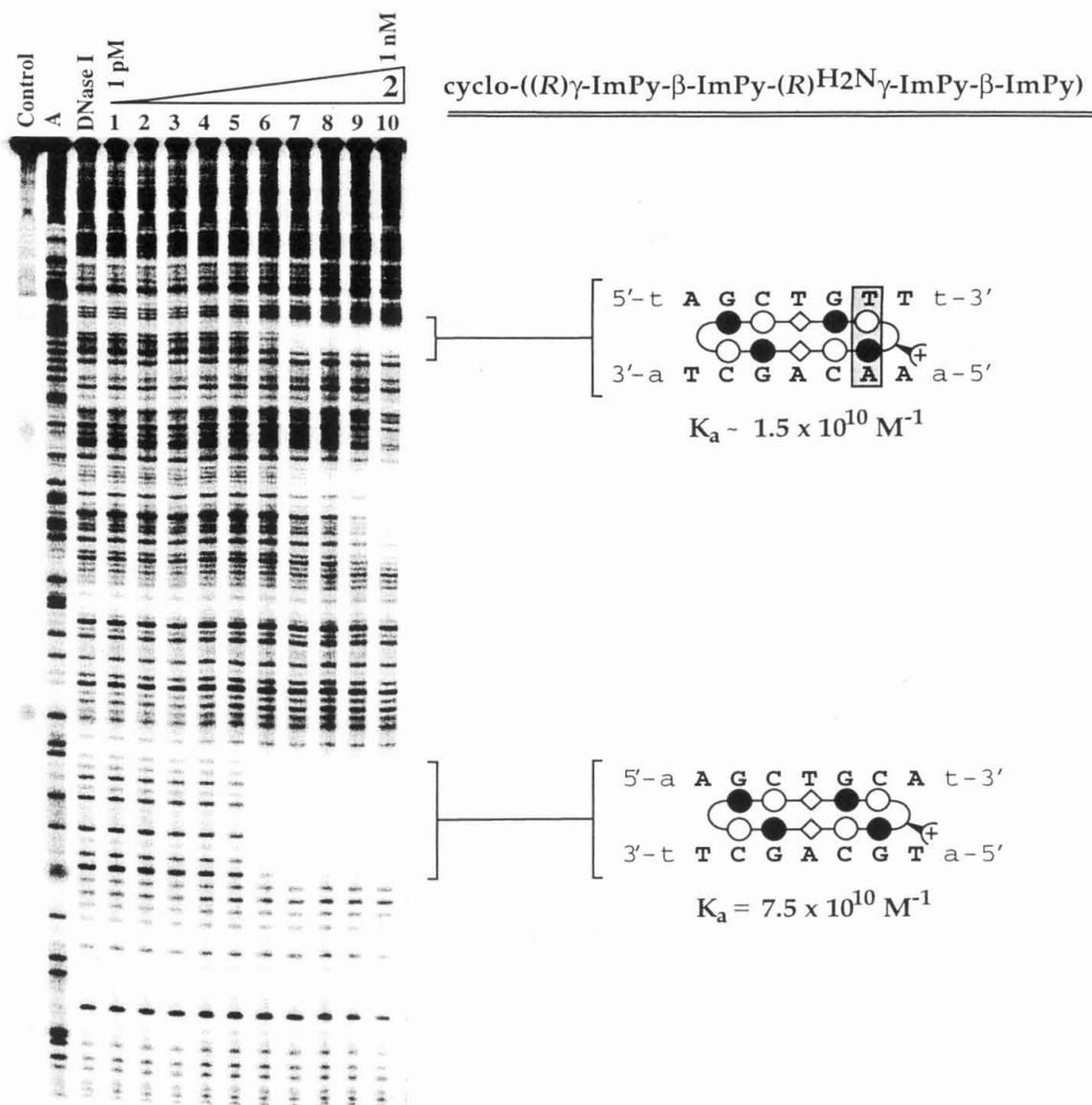


Figure 7.4 Quantitative DNase I footprinting titration experiments on polyamide 2. Lanes left to right; first lane, undigested control DNA; second lane, A sequencing lane; third lane, DNase I digested control lane; lanes 1 - 10, DNase I digestion products obtained in the presence of 1 pM, 2 pM, 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, and 1 nM, respectively. (Right) Ball and stick representation of 2 bound to the match 5'-AGCTGCA-3' and single base pair mismatch 5'-AGTGTTC-3'.

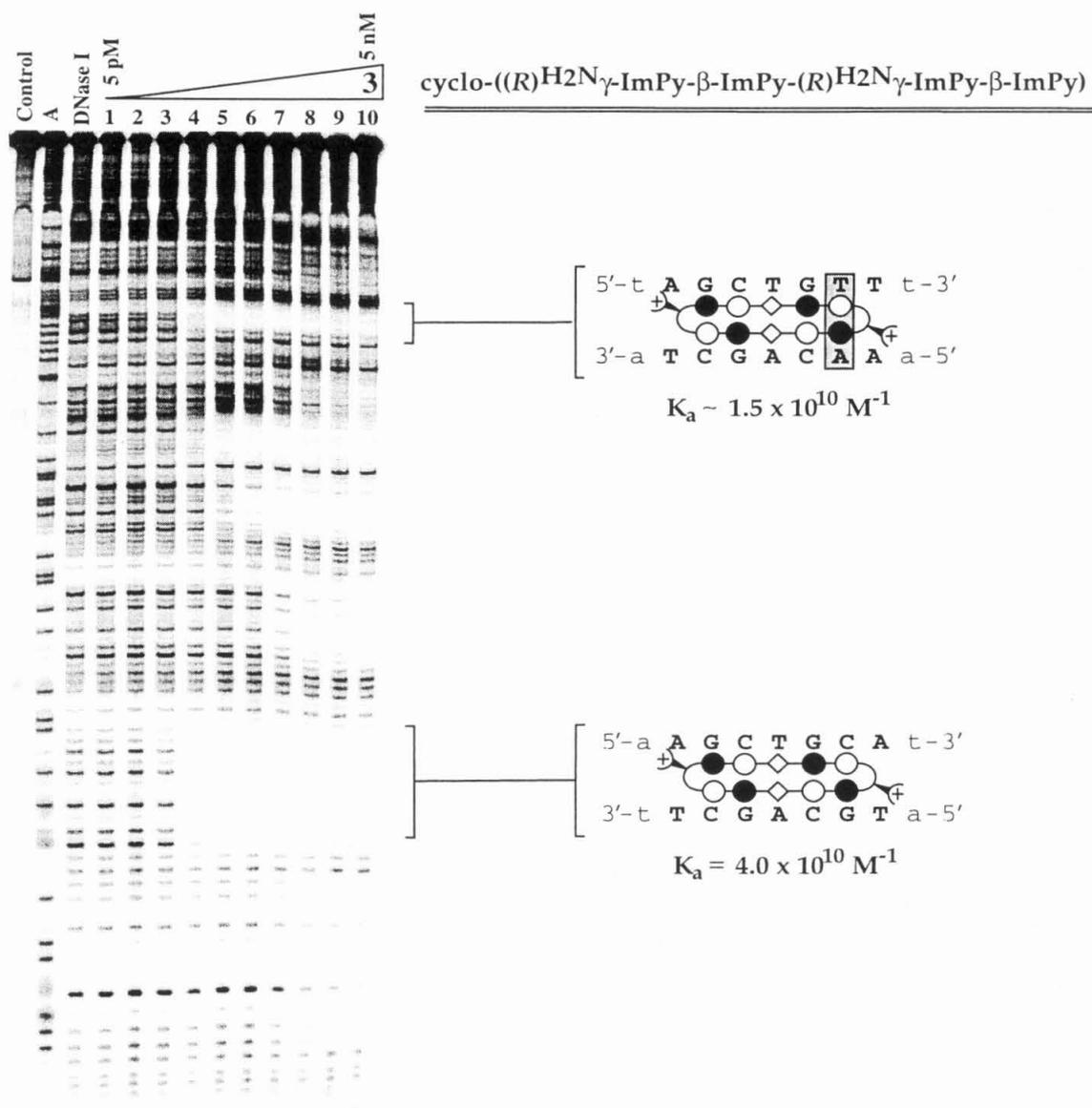


Figure 7.5 Quantitative DNase I footprinting titration experiments on polyamide 3. Lanes left to right; first lane, undigested control DNA; second lane, A sequencing lane; third lane, DNase I digested control lane; lanes 1 - 10, DNase I digestion products obtained in the presence of 1 pM, 2 pM, 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, and 1 nM, respectively. (Right) Ball and stick representation of 3 bound to the match 5'-AGCTGCA-3' and single base pair mismatch 5'-AGTGTTC-3'.

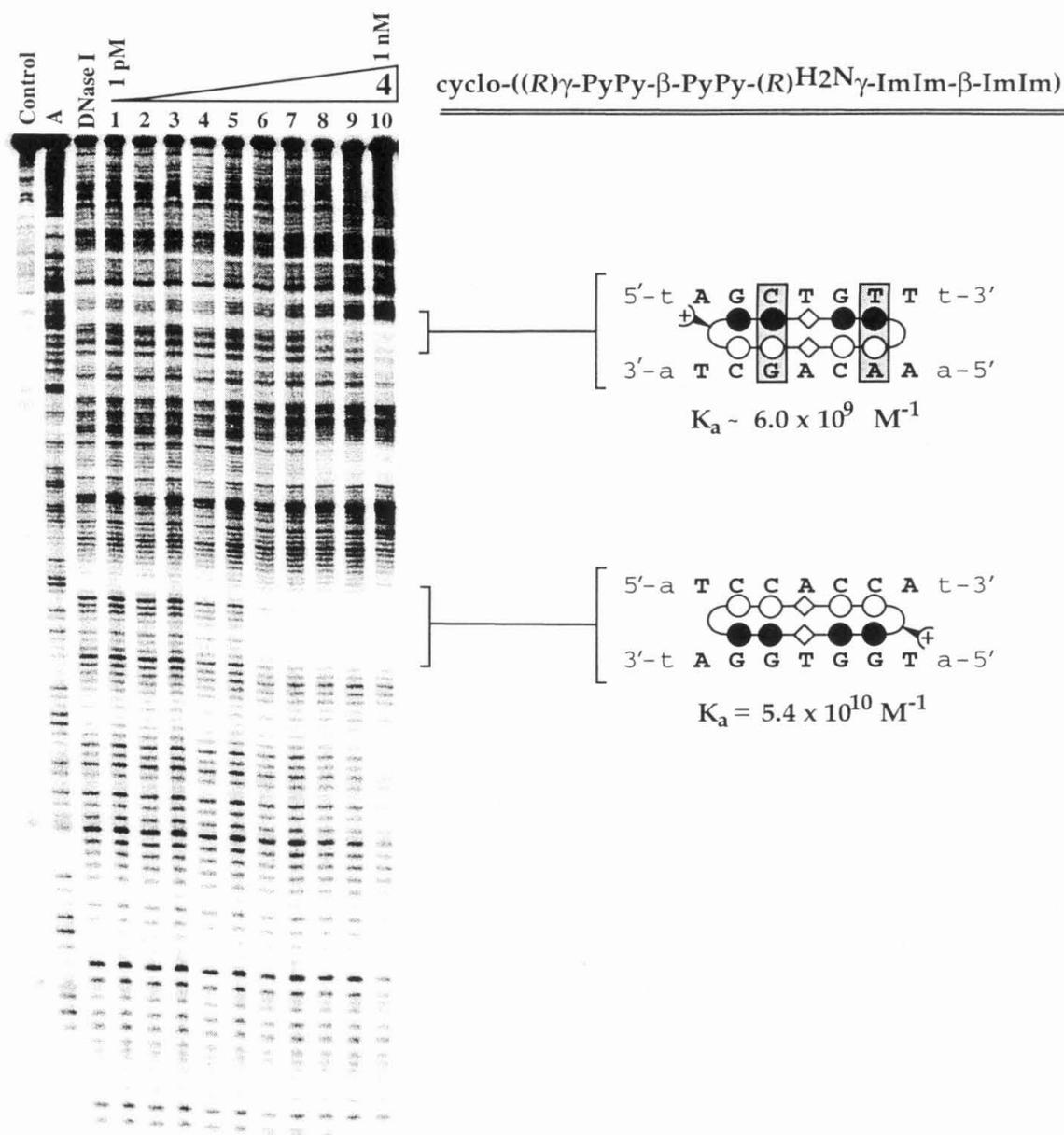


Figure 7.6 Quantitative DNase I footprinting titration experiments on polyamide 2. Lanes left to right; first lane, undigested control DNA; second lane, A sequencing lane; third lane, DNase I digested control lane; lanes 1 - 10, DNase I digestion products obtained in the presence of 1 pM, 2 pM, 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, and 1 nM, respectively. (Right) Ball and stick representation of **2** bound to the match 5'-TCCACCA-3' and double base pair mismatch 5'-AGCTGTT-3'.

$5.4 \times 10^{10} \text{ M}^{-1}) > 5\text{'-AGCTGTT-3'}$ ($K_a \sim 6.0 \times 10^9 \text{ M}^{-1}$) $> 5\text{'-TGCTGCA-3'}$ ($K_a < 1.0 \times 10^9 \text{ M}^{-1}$).

Equilibrium Association Constants (M^{-1})

Polyamide	5'-aTGCTGCA-3'	5'-tAGCTGTT-3'	5'-aTGGTGG-3'	Specificity
	6.7×10^{10} (0.3)	$\sim 6.0 \times 10^9$	-	11
	7.5×10^{10} (0.6)	$\sim 1.5 \times 10^{10}$	-	5
	4.1×10^{10} (0.8)	$\sim 1.5 \times 10^{10}$	-	2
	-	$\sim 6.0 \times 10^9$	5.4×10^{10} (0.6)	9

In this study we have developed a new viable binding motif for targeting the HIV-TATA box. As a control to evaluate effects that cyclization has upon the DNA binding properties of 2- β -2 polyamides, compound **1** was constructed with a "reduced-end" C-terminus which lacks the charged Dp moiety of conventional hairpins. Polyamide **1** binds its cognate match 5'-TGCTGCA-3' site with 10-fold and ≥ 67 -fold specificity versus the single and double base pair mismatch sites, respectively. Cyclo-**2** illustrates a modest increase in affinity for both the match site 5'-TGCTGCA-3' ($K_a = 7.5 \times 10^{10} \text{ M}^{-1}$) and the single base pair mismatch 5'-AGCTGTT-3' ($K_a \sim 1.5 \times 10^{10} \text{ M}^{-1}$), but this enhancement is accompanied by a 2-fold loss in specificity versus hairpin **1**. Cyclo-**3** contains two charges via incorporation of a second (R)^{H2N} γ -turn residue, binds the aforementioned sites with lower affinity and specificity. It is possible that multiple charges contribute to lower specificity by making non-specific contacts with the phosphate backbone of the DNA helix. However, this does not explain the decrease in affinity for the target match sequence. The control polyamide cyclo-**4** binds the match site 5'-TGGTGG-3' ($K_a = 5.4 \times 10^{10} \text{ M}^{-1}$), with 9- and 54-fold specificity versus the

double base pair mismatch sites 5'-AGCTGTT-3' ($K_a \sim 6.0 \times 10^9 \text{ M}^{-1}$) and 5'-TGCTGCA-3' ($K_a \leq 1.0 \times 10^9 \text{ M}^{-1}$).

Conclusion: The 2- β -2 motif is the first example of cyclic polyamides designed to target endogenous promoter sequences of DNA and, furthermore, represent promising candidates for addressing cell permeability of this class of small molecules. In light of the success in the 2- β -2 motif, it remains a challenge to design cycles that target longer sequences with high affinities and specificities. The in vitro and in vivo applications of these molecules will be reported in due course.

Experimental Procedure

General: Dicyclohexylcarbodiimide (DCC), Hydroxybenzotriazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU), Boc anhydride (Boc_2O) and 0.2 mmol/gram Boc- β -alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc- β -Pam-Resin) was purchased from Peptides International. (*R*)-2-Fmoc-4-Boc-diaminobutyric acid and γ -aminobutyric acid (γ) were from Bachem. *N,N*-diisopropylethylamine (DIEA) and *N,N*-dimethylformamide (DMF) were purchased from Applied Biosystems. Dichloromethane (DCM) was reagent grade from EM; thiophenol (PhSH), dimethylaminopropylamine (Dp), diphenylphosphoryl azide (DPPA), piperidine, palladium acetate, potassium carbonate, and ammonium formate were from Aldrich. Trifluoroacetic acid (TFA) Biograde from Halocarbon and *N*-(benzyloxycarbonyloxy) succinimide (CBZ-OSuc) from Fluka. All reagents were used without further purification.

Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. A shaker for manual solid phase synthesis was obtained from St. John Associates, Inc. Screw-cap glass peptide synthesis reaction vessels (5 mL and 20 mL) with a #2 sintered glass frit were made as described by Kent.¹⁶ UV spectra were measured in water on a

Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on a Beckman Gold system using a RAINEN C₁₈, Microsorb MV, 5 μ m, 300 x 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, gradient elution 1.25% acetonitrile/min. Preparatory reverse phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100 μ m C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. 18M Ω water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μ m filtered.

Polyamide Synthesis: Reagents and protocols for polyamide synthesis were as previously described.^{[3c],[5],[7a]} Polyamides were purified by reversed phase HPLC on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100 μ m C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. Extinction coefficients were calculated based on $\epsilon = 8333/\text{ring}$ at 304 nm.^[11]

General procedure for NaBH₄ cleavage:^[8] 100 mg of the appropriate resin was placed in a sealable container and swollen in 1 mL dry THF. 15 mg of NaBH₄ was subsequently added, the vessel then sealed and the reaction heated at 60 °C for 5 hours. After cooling to room temperature, the reaction was quenched by addition of 4 mL 20% TFA/80% H₂O; 4 mL CH₃CN was then added and the supernatant was collected by filtration. The resulting solution was frozen in liquid nitrogen, lyophilized, resuspended in .1% TFA/H₂O and purified by reversed phase HPLC to yield the appropriate polyamide.

ImPy- β -ImPy-(R)^{H₂N} γ -ImPy- β -ImPy-C3-OH (1): Recovered upon lyophilization of the appropriate fractions as a white powder (2 mg, 6% yield). MALDI-TOF-MS [M⁺-H]

(monoisotopic), 1283.3: 1283.3 calc. for $C_{57}H_{70}N_{24}O_{12}$.

General procedure for $Pd(OAc)_2/NH_4CO_2H$ cleavage:^[7a] 300 mg of the appropriate resin (.5 mmol/g as synthesized in reference 7a) and 300 mg of $Pd(OAc)_2$ were placed in a sealable container, suspended in 1 mL DMF, and shaken at 37 °C for 3 hours. 700 mg NH_4CO_2H , dissolved in 1 mL H_2O , was then added slowly over 3 minutes, the vessel sealed, and the resulting solution shaken at 37 °C for 12 hours. The supernatant was collected by filtration and subsequently purified by reversed phase HPLC to yield the appropriate polyamide.

$H_2N-\gamma-ImPy-\beta-ImPy-(R)^{Boc}-\gamma-ImPy-\beta-ImPy-OH$ (5): Recovered upon lyophilization of the appropriate fractions as a white powder (7 mg, 4% yield). MALDI-TOF-MS [M^+-H] (monoisotopic), 1426.4: 1426.4 calc. for $C_{63}H_{79}N_{21}O_{15}$.

$H_2N-(R)^{Boc}-\gamma-ImPy-\beta-ImPy-(R)^{Boc}-\gamma-ImPy-\beta-ImPy-OH$ (6): Recovered upon lyophilization of the appropriate fractions as a white powder (13 mg, 7% yield). MALDI-TOF-MS [M^+-H] (monoisotopic), 1541.5: 1541.5 calc. for $C_{68}H_{88}N_{26}O_{17}$.

$H_2N-\gamma-PyPy-\beta-PyPy-(R)^{Boc}-\gamma-ImIm-\beta-ImIm-OH$ (7): Recovered upon lyophilization of the appropriate fractions as a white powder (5 mg, 3% yield). MALDI-TOF-MS [M^+-H] (monoisotopic), 1426.4: 1426.4 calc. for $C_{63}H_{79}N_{25}O_{15}$.

General cyclization procedure:^[7b] The appropriate precycle was dissolved in DMF (1 μ mole/5 mL), K_2CO_3 was added (10 mg/ μ mole of polyamide) and the resulting solution stirred at RT for 30 minutes. DPPA (30 μ L/ μ mole of polyamide) was then added and the resulting solution stirred at room temperature for 5 hours. The solution was filtered to remove excess K_2CO_3 , and all volatiles in the filtrate were removed in vacuo. The resulting white powder was dissolved in 2 mL TFA and allowed to stand at RT for 30

minutes. The volume was subsequently adjusted to 10 mL with .1% TFA/H₂O and purification by reversed phase HPLC yielded the appropriate cycle.

cyclo-(γ -ImPy- β -ImPy-(R)^{H²N} γ -ImPy- β -ImPy-) (2): Recovered upon lyophilization of the appropriate fractions as a white powder (1.2 mg, 20% yield). MALDI-TOF-MS [M⁺-H] (monoisotopic), 1308.3: 1308.3 calc. for C₅₈H₆₉N₂₅O₁₂.

cyclo-((R)^{H²N} γ -ImPy- β -ImPy-(R)^{H²N} γ -ImPy- β -ImPy-) (3): Recovered upon lyophilization of the appropriate fractions as a white powder (1.5 mg, 25% yield). MALDI-TOF-MS [M⁺-H] (monoisotopic), 1323.3: 1323.3 calc. for C₅₈H₇₀N₂₆O₁₂.

cyclo-(γ -PyPy- β -PyPy-(R)^{H²N} γ -ImIm- β -ImIm-) (4): Recovered upon lyophilization of the appropriate fractions as a white powder (1.4 mg, 20% yield). MALDI-TOF-MS [M⁺-H] (monoisotopic), 1308.3: 1308.3 calc. for C₅₈H₆₉N₂₅O₁₂.

DNA Reagents and Materials: Enzymes were purchased from Boehringer-Mannheim and used with their supplied buffers. Deoxyadenosine and thymidine 5'-[α -³²P] triphosphates were obtained from Amersham, and deoxyadenosine 5'-[γ -³²P]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.^[17]

Preparation of 3'-End-Labeled Restriction Fragments. The plasmid pJT2B2 was linearized with *Eco* RI and *Pvu* II, then treated with the Sequenase enzyme, deoxyadenosine 5'-[α -³²P]triphosphate and thymidine 5'-[α -³²P]triphosphate for 3' labeling. The labeled 3' fragment was loaded onto a 6% non-denaturing polyacrylamide

gel, and the desired 282 base pair band was visualized by autoradiography and isolated.

DNase I Footprinting:¹⁵ All reactions were carried out in a volume of 400 μL . We note explicitly that no carrier DNA was used in these reactions until after DNase I cleavage. 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 30 kcpm 3'-radiolabeled DNA. The solutions were allowed to equilibrate for 4 hours 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 1.875 $\mu\text{g}/\text{mL}$) and was allowed to proceed for 7 min at 22 °C. The reactions were stopped by adding 50 μL of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/mL 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 6 min, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V for 1 hour. The gels were dried under vacuum at 80 °C, then quantitated using storage phosphor technology.

Equilibrium association constants were determined as previously described.^{11a} The data were analyzed by performing volume integrations of the 5'-TGCTGCA-3', 5'-AGCTGTT-3', and 5'-TGGTGGGA-5' sites and a reference site. The apparent DNA target site saturation, θ_{app} , was calculated for each concentration of polyamide using the following equation:

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

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{ref}° correspond to those values for a DNase I control lane to which no polyamide has been added. The ($[L]_{\text{tot}}$, θ_{app}) data points were fit to a Langmuir binding isotherm (eq 2, $n=1$ for polyamides 1 and 2, by minimizing the difference between θ_{app} and θ_{fit} using the modified Hill equation:

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

where $[L]_{\text{tot}}$ corresponds to the total polyamide concentration, K_a corresponds to the equilibrium association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software (version 2.1, Abelbeck software) with K_a , θ_{max} , and θ_{min} as the adjustable parameters. All acceptable fits had a correlation coefficient of $R > 0.97$. At least three 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. The data were normalized using the following equation:

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

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 12-20 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.

References

1. Dervan, P. B., and Burli, R. W. (1999) *Curr. Opin. Chem. Biol.* 3, 688-693.
2. Galderisi, U.; Cascino, A.; Giordani, A., *J. Cell. Physiol.* **1999**, 181, 251-257.
3. Good, L. M., Nielson, P. E.; *Nat. Biotechnol.* **1998**, 16, 355-358.
4. Smith, A. L.; Nicolauo, K. C., *The Eneidyne Antibiotics* **1996**, 39, 2103-2117
5. Odom, D. T.; Parker, C. S.; Barton, J. K., *Biochemistry* **1999**, 38, 5155-5163
6. Chiang, S. Y.; Welch, J. J.; Rauscher, F. J., 3rd; Beerman, T. A. *J. Biol. Chem.* **1996**, 271, 23999-234004
7. Welch, J. J.; Rauscher, F. J., 3rd; Beerman, T. A. *J. Biol. Chem.* **1994**, 269, 31051-31058
8. Geierstanger, B. H., Mrksich, M., Dervan, P. B., and Wemmer, D. E. (1994) *Science* 266, 646-650.
9. Kielkopf, C. L., Baird, E. E., Dervan, P. B., and Rees, D. C. (1998) *Nature Struct. Biol.* 5, 104-109.
10. Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. (1998) *Science (Washington, D.C.)* 282, 111-115.
11. Baird, E. E., and Dervan, P. B. (1996) *J. Am. Chem. Soc.* 118, 6141-6146.
12. a) D. M. Herman, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1999**, 121 1121-1129 b) J. Cho, M. E. Parks and P. B. Dervan *Proc. Natl. Acad. Sci. USA* **1995**, 92, 10389-10392.
13. Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E., and Dervan, P. B. (1997) *Nature (London)* 387, 202-205.
14. Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan, P. B. (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95, 12890-12895.

15. (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. U.S.A.* **1986**, 83, 8462. (c) D. F. Senear, M. Brenowitz, M. A. Shea, G. K. Ackers, *Biochemistry* **1986**, 25, 7344.
16. S.B.H. Kent, *Annu. Rev. Biochem.* **1988**, 57, 957-989.