

Design of Ligands for Sequence-Specific Recognition of the Minor-Major Grooves of DNA

Thesis by
Jason W. Szewczyk

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

California Institute of Technology
Pasadena, California
1999

(Submitted November 25, 1998)

© 1999

Jason W. Szewczyk

All Rights Reserved

In memory of my Mom

and

To my family with love.

Acknowledgments

I would first like to thank my advisor Prof. Peter B. Dervan for his enthusiasm, vision, guidance, and support during my graduate career. I have been truly fortunate to have studied under his tutelage, which allowed me to grow and accomplish more than I would have thought possible at the outset.

I thank all the people who have supported me through the years at Caltech. First, I thank the group members who have taught me a great deal about life and science: Dr. Eldon Baird, who was a true friend; Dr. Sarah White who provided friendship, fun times, and support; Ryan Bremer (and L.S.), for quality midwest friendship and C.S. seminars; Jim Turner, who was both a friend and an incredible colleague; and Aileen S. Chang, who provided constant encouragement. I also thank Dave Herman, Clay Wang, and Nick Wurtz for enriching my experience at Caltech.

I have been fortunate to collaborate with a number of incredible scientists. Their hard work, insight, and dedication have inspired me and contributed greatly to the work described here. I would like to thank some of the people whom I have had the pleasure to work with at Caltech. The polyamide-oligonucleotides were developed in collaboration with Eldon Baird. The hydroxypyrrrole project was conducted in collaboration with a superstar team of scientists: Eldon Baird, Sarah White, and Jim Turner. For structure studies, Prof. Doug Rees and Clara Kielkopf have been great collaborators and contributed greatly to the field's understanding of minor

groove recognition. Lastly, the novel monomer program was in collaboration with Eldon Baird, Ryan Bremer, and Doan Nyguyen.

I also want to thank special people outside of Caltech for their encouragement and friendship, which made my time here much more meaningful. Thanks go out to: the boys from Chicago, E. O. and M. G., who are family; T. K., Dude, we have to get together! G. C., eight years and still going; Mr. and Mrs. C., for adopting me; and S. F., who taught me the *only* real skill I'll ever need.

I would like to thank all the members of my thesis committee: Prof. D. C. Rees, Prof. R. H. Grubbs, Prof. S. L. Mayo, and Prof. P. B. Dervan for their time and guidance. I also want to thank the former chairman of my committee, Prof. A. G. Myers, for his invaluable advice during my candidacy exam and in selecting a post-doctoral advisor.

Finally, I thank my family for all their love and support. I especially thank my brother, Greg, who always told it like it is; Nanni, who helped make me the man I am today, and my mom, whose unwavering encouragement, love and support carried me through my time at Caltech. Lastly, I thank the Lord, for without His love none of this would have been possible.

Abstract

The two most powerful chemical approaches to date for the sequence-specific recognition of double helical DNA are the pyrrole-imidazole polyamides in the minor groove and oligonucleotide-directed triple-helix formation in the major groove. Described here are the design and synthetic methods to combine the two models for DNA recognition in a single motif. A hairpin polyamide connected to a pyrimidine-oligonucleotide via a simple aliphatic linker was shown to simultaneously recognize the major and minor grooves of DNA at subnanomolar concentrations (Chapter 2). Development of versatile solid phase methodology expanded the protocols for the synthesis of pyrrole-imidazole polyamides to include polyamide-oligonucleotide conjugates. Utilizing the pyrrole-imidazole polyamide moiety as a sequence-specific dimerization domain afforded a class of cooperative polyamide-oligonucleotides which bound a 27 bp target with a 2.7 kcal mol⁻¹ increase in binding energy relative to the unlinked subunits (Chapter 3). Introduction of a new linker design created an extended polyamide-oligonucleotide motif to specifically target 31 contiguous base pairs of DNA at subnanomolar concentrations (Chapter 4). Polyamide-oligonucleotides combine the two binding models to create a paradigm for simultaneous recognition of the major-minor grooves of DNA with synthetic ligands.

For the pyrrole-imidazole polyamide model, sequence specificity depends on side-by-side aromatic amino acid pairings. An Im/Py pair

distinguishes G•C from C•G and both of these from A,T base pairs. A Py/Py pair specifies A,T from G,C but does not distinguish A•T from T•A. To break this degeneracy a new aromatic amino acid, 3-hydroxypyrrole (Hp), was synthesized to test for pairings which discriminate A•T from T•A. Replacement of a single hydrogen atom with a hydroxy group in a Hp/Py pairing regulates affinity and specificity by an order of magnitude. By incorporation of a third aromatic amino acid, hydroxypyrrole-imidazole-pyrrole polyamides form four ring pairings (Im/Py, Py/Im, Hp/Py, Py/Hp) which distinguish all four Watson-Crick base pairs in the minor groove of DNA (Chapter 5).

Table of Contents

	page
Acknowledgments.....	iv
Abstract.....	vi
Table of Contents	viii
List of Figures and Tables.....	ix
 CHAPTER ONE: Introduction	 1
 CHAPTER TWO: Sequence-Specific Recognition of DNA by a Major and Minor Groove Binding Ligand	 19
 CHAPTER THREE: Cooperative Triple-Helix Formation via a Sequence Specific Minor Groove Dimerization Domain.....	 40
 CHAPTER FOUR: Cooperative Binding of Extended Pyrrole- Imidazole Polyamide-Oligonucleotides to the Minor-Major Grooves of DNA.....	 62
 CHAPTER FIVE: Recognition of the Four Watson-Crick Base Pairs in the DNA Minor Groove by Synthetic-Ligands..	 94

List of Figures and Tables

CHAPTER ONE	page
Figure 1.1 Structure of double-helical B-form DNA.....	3
Figure 1.2 Schematic model for DNA recognition.....	4
Figure 1.3 X-ray crystal structures of protein-DNA complexes.....	5
Figure 1.4 Ribbon model of the pyrimidine triple helix	8
Figure 1.5 Pyrimidine motif T•AT and C+GC base triplets.....	9
Figure 1.6 Purine motif G•GC, A•AT, and T•AT base triplets	10
Figure 1.7 A schematic representation of the polyamide pairing rules.....	11
Figure 1.8 Recognition by unlinked and hairpin polyamides	12
Figure 1.9 Space filling model for Hp-Im-Py polyamide x-tal structure	14
 CHAPTER TWO	
Figure 2.1 Ribbon model of hairpin-polyamide-oligonucleotide	21
Figure 2.2 Structure of hairpin polyamide-oligonucleotide.....	22
Figure 2.3 Synthesis of 5'-N-alloc -thymidine phosphoramidite.....	23
Figure 2.4 Synthesis of hairpin-polyamide-oligonucleotide.....	25
Figure 2.5 DNase I footprinting gel.....	27
Table 2.1 Association constants for hairpin-polyamide-oligonucleotide.	28
 CHAPTER THREE	
Figure 3.1 Ribbon model of cooperative polyamide-oligonucleotide.....	42
Figure 3.2 Structure of cooperative polyamide-oligonucleotide.....	43
Figure 3.3 Solid phase synthesis of polyamide-oligonucleotide.....	46
Figure 3.4 DNase I footprinting gel.....	47
Figure 3.5 Structure of conjugates with a variable linker domain	50
Table 3.1 Cooperative-polyamide-oligonucleotide association constants	49
Table 3.2 Association constants for linker length dependence.....	50

CHAPTER FOUR

Figure 4.1	Two-dimensional models of specific DNA complexes	65
Figure 4.2	Ribbon model of extended polyamide-oligonucleotide.....	67
Figure 4.3	Structure of extended polyamide-oligonucleotide.....	70
Figure 4.4	Synthesis of extended-polyamide-oligonucleotide	72
Figure 4.5	DNase I footprinting gel.....	75
Figure 4.6	Binding isotherms for extended-polyamide-oligonucleotide....	77
Table 4.1	Association constants of extended polyamide-oligonucleotide	77

CHAPTER FIVE

Figure 5.1	Chemical structure and CPK models of A,T base pair.....	97
Figure 5.2	Eight-ring Hp-Im-Py polyamides with binding models.....	98
Figure 5.3	Schematic models and chemical structures of A,T base pair	104
Figure 5.4	Hydrogen bonding model for Hp-Im-Py polyamide.....	105
Figure 5.5	Models of the hairpin Hp-Im-Py polyamide complexes	107
Figure 5.6	Structures of eight-ring Hp-Im-Py polyamides	108
Figure 5.7	Synthesis of Hp monomer.....	109
Figure 5.8	Solid phase synthesis of Hp-Im-Py polyamides	111
Figure 5.9	MPE footprinting with eight-ring Hp-Im-Py hairpins.....	114
Figure 5.10	Affinity cleavage with eight-ring Hp-Im-Py hairpins.....	117
Figure 5.11	Chemical structure of ImHpPyPy and ImPyPyPy.....	124
Figure 5.12	Structure of ImHpPyPy bound to DNA double helix	127
Figure 5.13	Diagram of 2 H-bonds between T and Hp.....	129
Figure 5.14	T•A base pair is melted.	130
Figure 5.15	Interaction of Hp/Py pair with T•A base pair.....	131
Figure 5.16	All ligand-DNA hydrogen bonds.....	132
Table 5.1	Association constants for eight-ring Hp-Im-Py hairpins	99
Table 5.2	Pairing Code for Minor Groove Recognition.....	101

Table 5.3	Association constants for eight-ring Hp-Im-Py hairpins	119
Table 5.4	T•A/A•T specificity of Py and Hp Pairs	122
Table 5.5	Association constants for four-ring dimers GTAC v. GTAT	125
Table 5.6	Association constants for dimers GTAC v. GAAC v. GATC	126

Chapter 1

Introduction

Discussion of the Background. In every human cell the genetic information is stored on a string-like DNA polymer approximately 1 meter in length and contains 3×10^9 units of information in the form of base pairs, which encode 80,000 to 100,000 genes or sets of instructions.¹ The specific interaction of proteins, such as transcription factors, with DNA controls the regulation of genes and hence cellular processes.² The design of synthetic ligands that read the information stored in the DNA double helix has been a central goal at the interface of chemistry and biology.³ Synthetic molecules which target specific DNA sequences offer a potentially general approach for gene-specific regulation.⁴

The genetic information is stored on two polydeoxyribonucleic acid strands which associate in an antiparallel orientation to form a double helical structure.⁵ The DNA double helix consists of A,T and G,C base pairs held together by specific Watson-Crick hydrogen bonds resembling the rungs of a twisted ladder (Figure 1.1).⁶ The common B-form of DNA is characterized by a wide (12Å) and shallow major groove and a deep and narrow (4-6 Å) minor groove. Individual sequences may be distinguished by the pattern of hydrogen bond donors and acceptors displayed on the edges of the base pairs (Figure 1.2a).⁷ In the major groove the four Watson-Crick base pairs are distinguished by their individual pattern of hydrogen bond donors and acceptors (Figure 1.2b). In the minor groove, the A,T base pair presents two symmetrically placed hydrogen bond acceptors, the N3 of purines and the O2

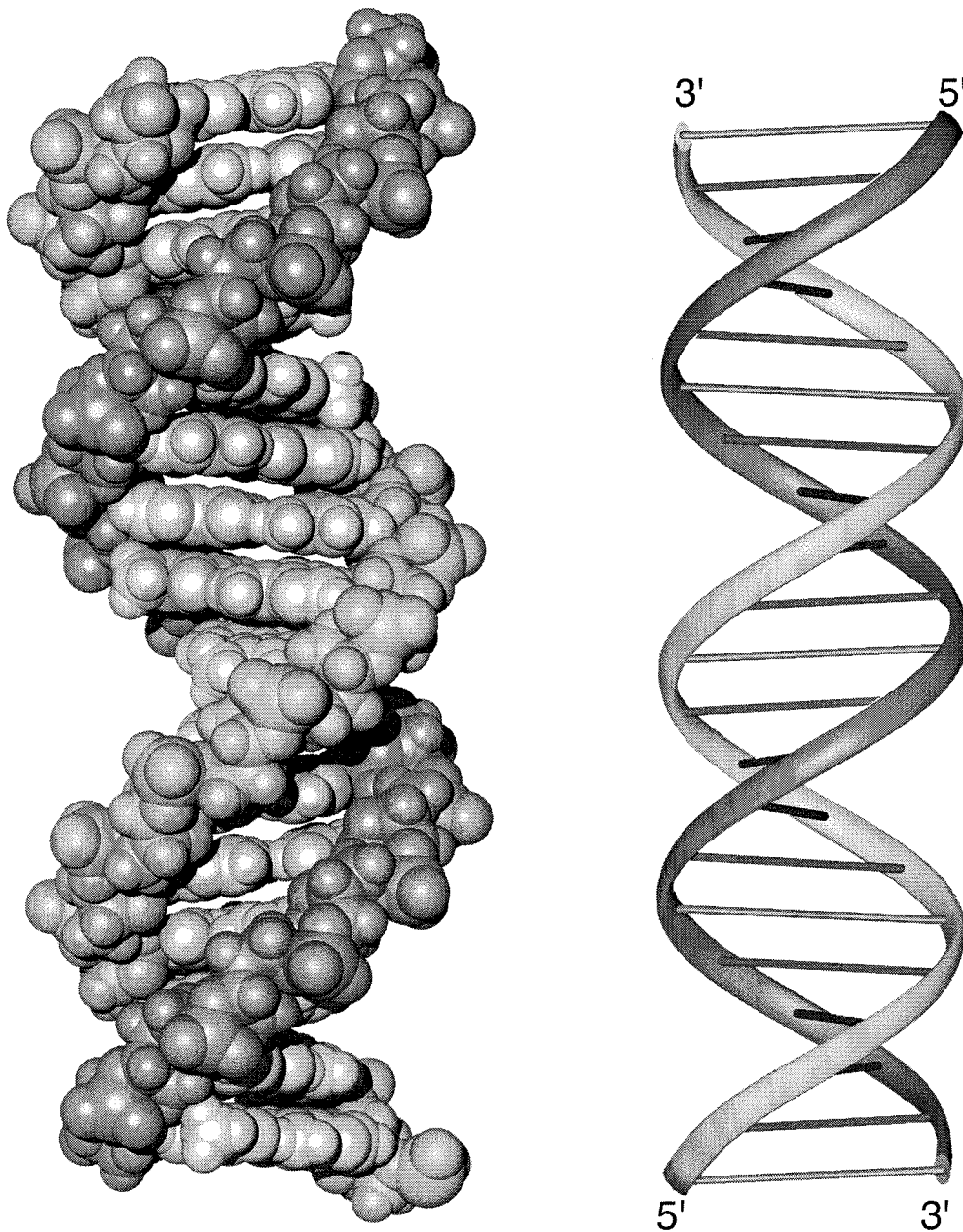


Figure 1.1 B-form double helical DNA. Antiparallel strands are indicated in dark and light gray. (left) space filling CPK model, (right) ribbon representation.

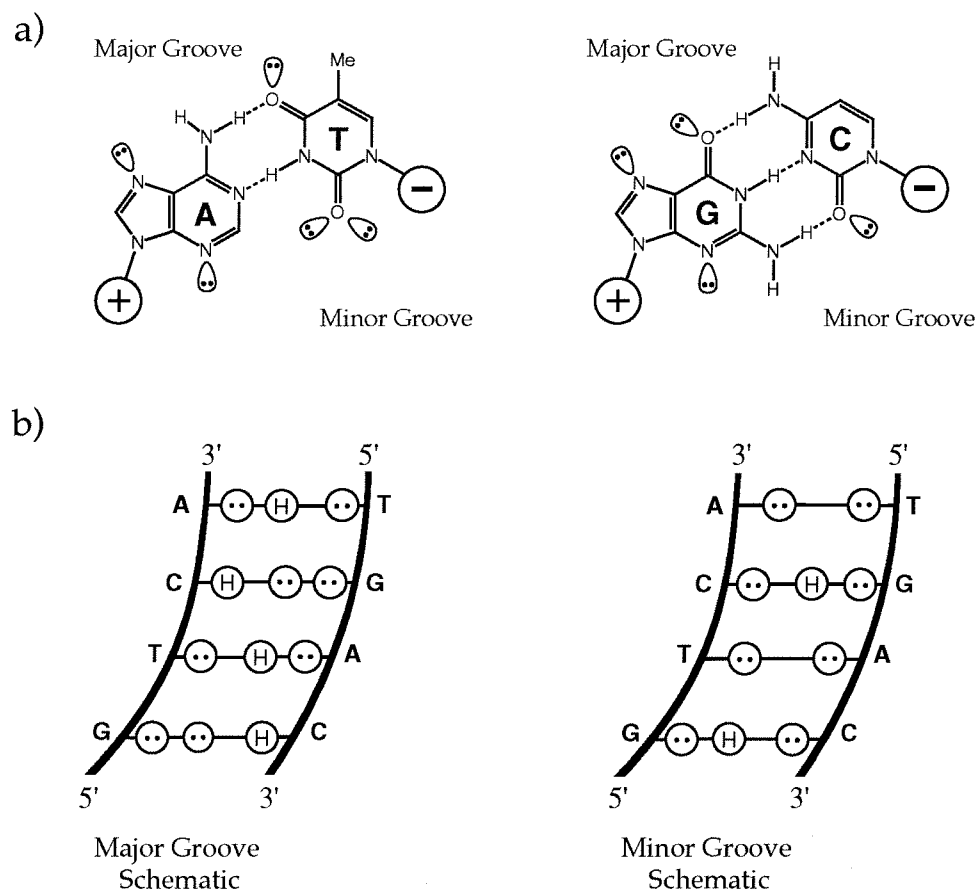


Figure 1.2 a) Chemical structure of the Watson-Crick base pairs. b) Schematic models for DNA recognition in the major and minor grooves. Circles containing an H represent hydrogen bond donors, while circles containing two dots represent hydrogen bond acceptors. The schematics underscore the potential coded reading of the DNA helix.

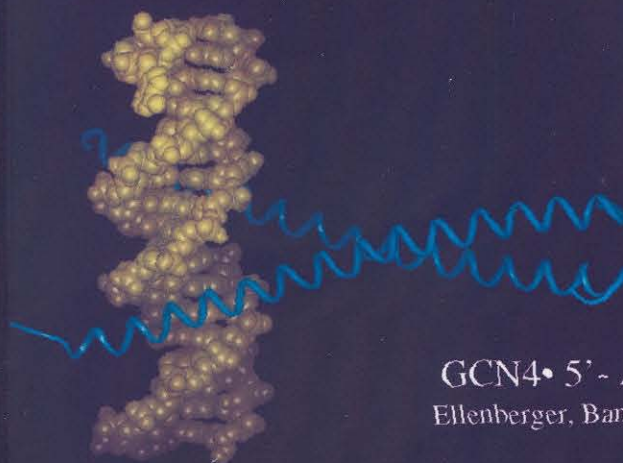
of pyrimidines. The G,C base pair presents these two acceptors, but in addition presents a hydrogen bond donor, the 2-amino group of guanine (Figure 1.2b).⁸

Molecular Recognition of DNA. Nature has developed a variety of protein motifs for the formation of specific DNA complexes (Figure 1.3).⁹ Sequence-specificity arises from the ability of these molecules to form specific

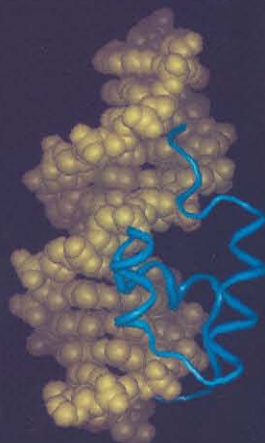
Figure 1.3 X-ray crystal structures of protein-DNA complexes. (top) Zinc finger Zif268 with 5'-GCGTGGGCG-3'. (middle) Leucine zipper GCN4 with 5'-ATGACTCATCCA-3'. (bottom) Helix-turn-helix Hin recombinase with 5'-TTTTGATAA-3'.



Zif268 • 5'-GCGTGGGCG-3'
Pavletich & Pabo *Science* 1991



GCN4 • 5'-ATGACTCATCCA-3'
Ellenberger, Bandl, Struhl & Harrison *Cell* 1992



Hin • 5'-TTTTGATAA-3'
Feng, Johnson & Dickerson *Science* 1994

interactions with the DNA helix: (i) hydrogen bonding, (ii) van der Waals contacts, (iii) Coulombic attraction to the negatively charged phosphodiester, and/or (iv) intercalation of the DNA bases. Crystal structures of DNA binding proteins reveals nature has evolved a specific protein solution for each sequence targeted, which indicates no simple 1:1 relationship exists between a single amino acid and a specific base pair.¹⁰ Although a versatile recognition code is emerging from studies of certain zinc-finger-DNA complexes,¹¹ the *de novo* design of proteins for recognition of designated sequences remains problematic due to limitations in predicting protein folding.

Chemical approaches to DNA recognition. In search of a general solution to DNA recognition, a twenty year research effort led by Professor Peter B. Dervan of Caltech has resulted in two distinct chemical methods which read a specific groove of the DNA helix. Synthetic oligonucleotides bind the major groove by formation of a local triple helix, while pyrrole-imidazole polyamides bind in the minor groove by formation side-by-side pairings of aromatic amino acids.

Oligonucleotide-directed triple-helix formation. In 1987, Moser and Dervan¹² reported that a 15 mer pyrimidine oligonucleotide specifically bound a 15 base target on a 4.0 kbp linearized plasmid. The synthetic oligonucleotide bound in the major groove oriented parallel to the purine rich strand of the duplex

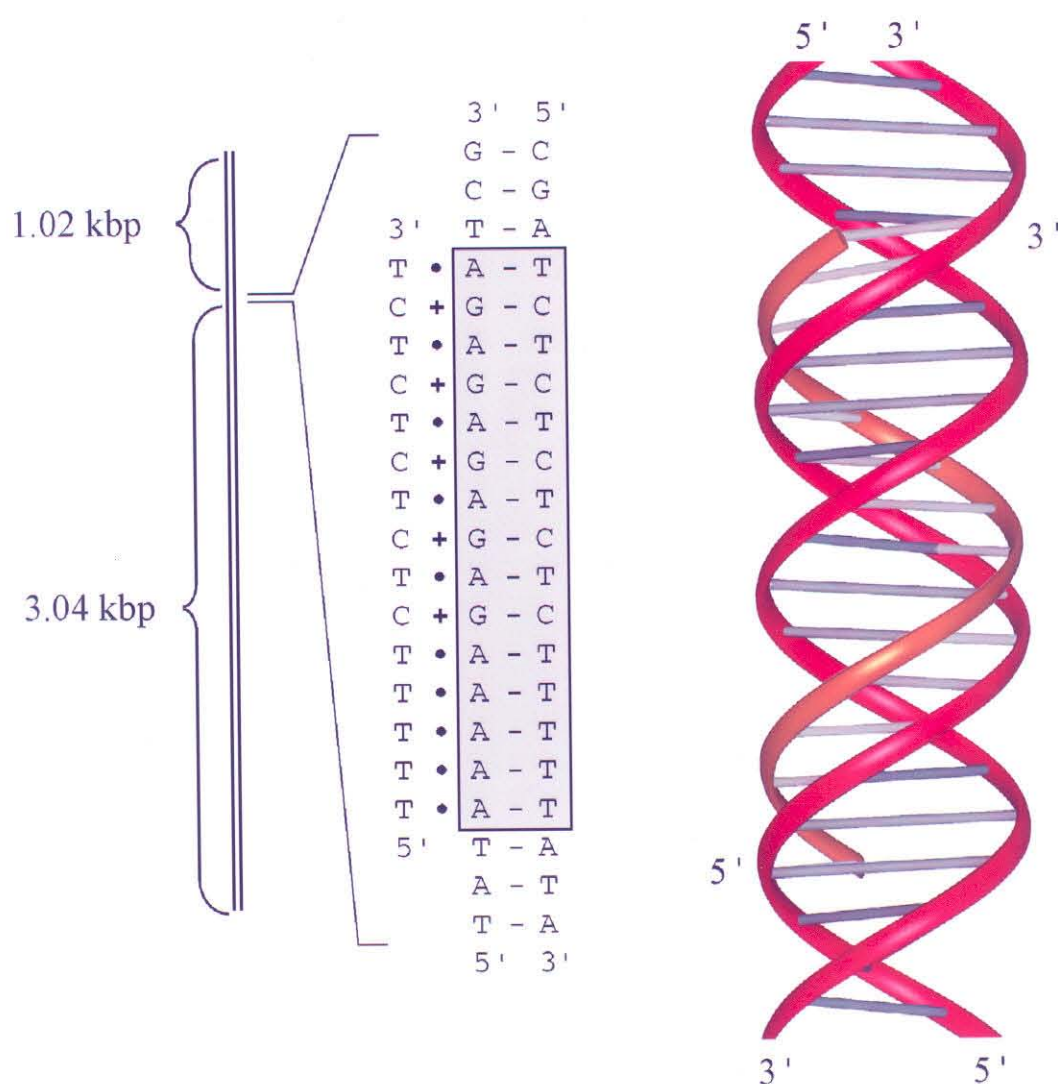


Figure 1.4 (Left) Schematic showing a 15-mer oligonucleotide binding a single site on a 4.0 kbp linearized plasmid.^{19a} The grey box surrounds the 15 base pairs specifically recognized. (Right) Ribbon model depicting a 15 base pyrimidine triple helix. Red and gold strands represent duplex DNA and the third strand, respectively.

placing each thymine residue opposite an A•T base pair and protonated cytosines opposite G•C base pairs (Figure 1.4). The specificity and stability of the complex was mediated by the formation of Hoogsteen hydrogen bonds between the third strand and the purine rich strand of the duplex. The hydrogen bonding interactions responsible for specific complex formation are summarized by two base triplet structures, T•AT, and C+GC (Figure 1.5).¹³

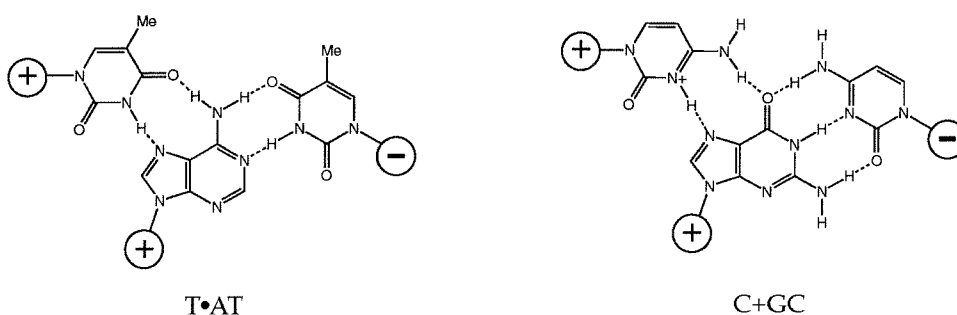


Figure 1.5 Pyrimidine motif T•AT and C+GC base triplets.

In 1991, a second triple helical motif was reported by Beal and Dervan,¹⁴ where a purine-rich oligonucleotide bound in the major groove of double helical DNA antiparallel to the Watson-Crick purine strand. Specific recognition is achieved by formation of reverse Hoogsteen hydrogen bonds between the third strand and the purine-rich strand of the duplex, depicted in three base triplet models, G•GC, A•AT, and T•AT (Figure 1.6). NMR structural studies have since confirmed the base triplet structures for both triplex motifs (Figures 1.5 and 1.6).¹⁵

The utility of oligonucleotide-directed triple-helix formation was demonstrated by cleavage of a single site in human chromosomal DNA,

inhibition of DNA binding proteins, and the regulation of gene expression.^{16,17} Triple helices are sensitive to oligonucleotide length, single base mismatch, pH, action concentration and valence, and temperature.¹⁸ Triple-helix formation targets a broad sequence repertoire with high affinity and specificity;¹⁹ however; the triple helix approach is limited to tracks of purines and suffers from poor cellular uptake.

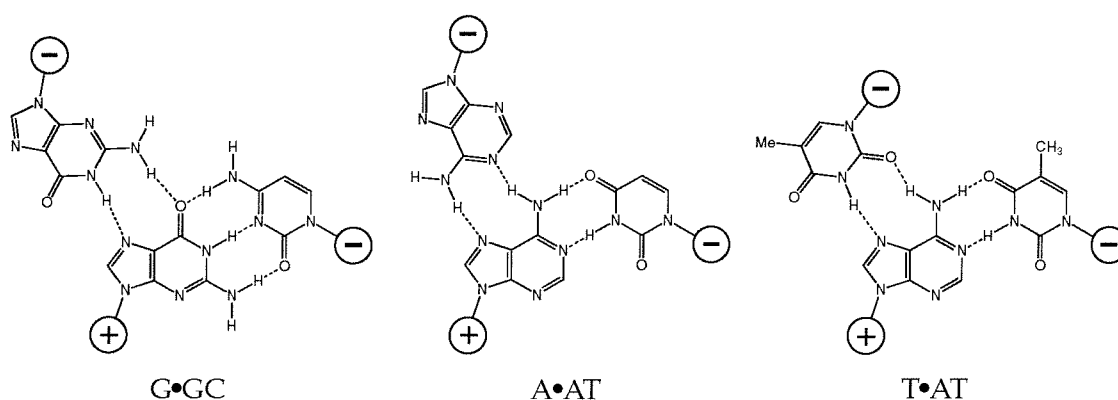


Figure 1.6 Purine motif G•GC, A•AT, and T•AT base triplets.

Pyrrole-imidazole polyamides. Paralleling the effort in the major groove, polyamides containing *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) amino acids provide a model for the design of ligands which target the minor groove. The DNA sequence specificity of these small molecules can be controlled by the linear sequence of pyrrole and imidazole amino acids.²⁰ A pairing of Im opposite Py targets a G•C base pair while a pairing of Py opposite Im targets a C•G base pair.¹⁴ A Py/Py combination is degenerate targeting both A•T and T•A base pairs (Figure 1.7).^{13,14} Specificity for G,C base pairs results from the formation of a putative hydrogen bond between the imidazole N3

and the exocyclic amine group of guanine. Validity of the pairing rules is supported by a variety of footprinting, NMR, and x-ray structure studies.²¹

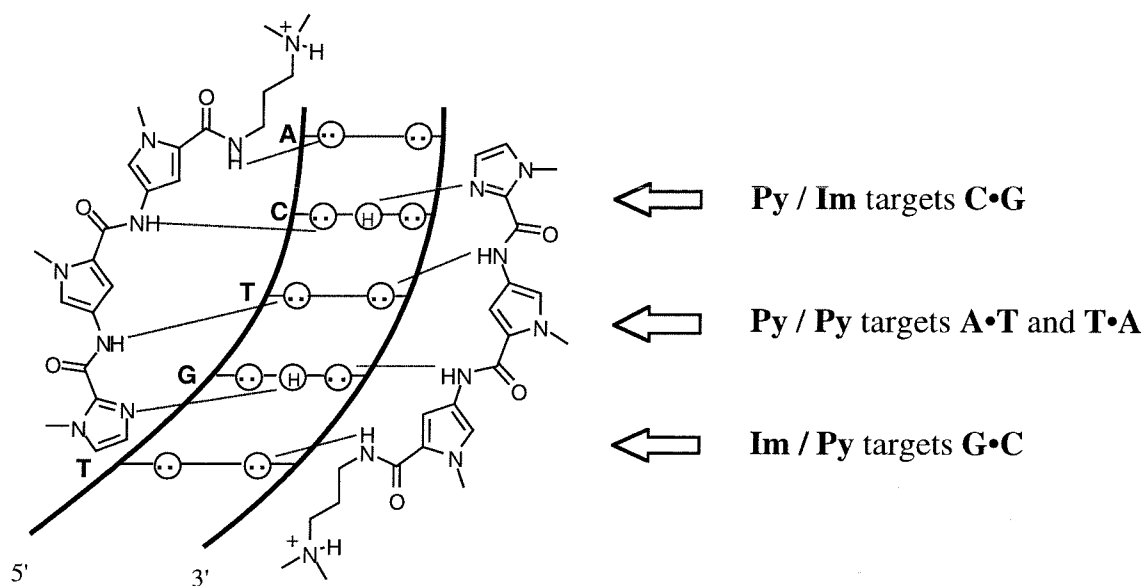


Figure 1.7 A schematic representation of the polyamide pairing rules.

In parallel with the elucidation of the scope and limitations of the pairing rules, efforts have been made to increase the DNA-binding affinity and specificity of pyrrole-imidazole polyamides by covalently linking polyamide subunits.^{22,23} A hairpin polyamide motif with γ -aminobutyric acid (γ) serving as a turn-specific internal-guide-residue provides a synthetically accessible method of linking polyamide subunits within the 2:1 motif (Figure 1.8). The head-to-tail linked polyamide ImPyPy- γ -PyPyPy-Dp was shown to specifically bind the designated target site 5'-TGTTA-3' with an equilibrium association constant of $K_a = 8 \times 10^7 \text{ M}^{-1}$, an increase of 300-fold relative to the

unlinked three-ring polyamide pair ImPyPy and PyPyPy.²⁴ The hairpin polyamide model is supported by footprinting, affinity cleaving, and NMR structure studies.^{18,25}

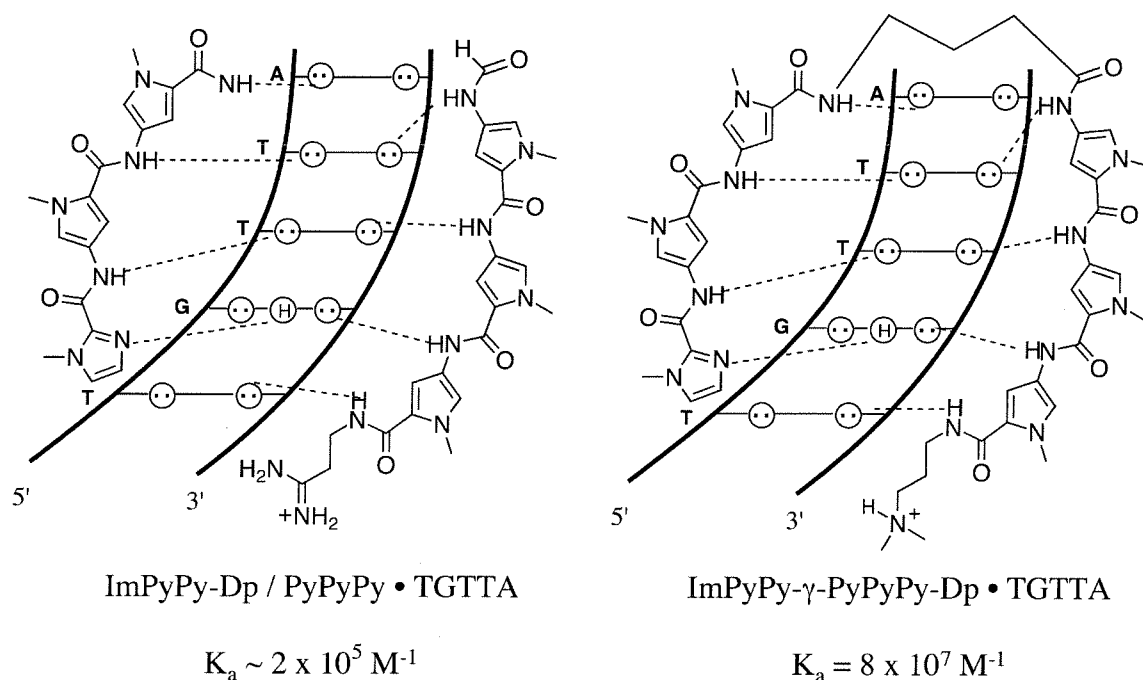


Figure 1.8 A schematic representation of recognition of a 5'-TGTTA-3' sequence by unlinked subunits (left) and γ -aminobutyric acid linked 'hairpin' subunits (right).

Oligonucleotide-directed triple helix formation and the pyrrole-imidazole polyamides represent the two most powerful chemical approaches for the sequence-specific recognition of DNA. A key question remained if these two models could be combined in a single molecule without compromising the affinity or sequence-selectivity of either motif.

Described here are the design and synthetic methods to combine the two models for DNA recognition in a single motif. A hairpin polyamide connected to a pyrimidine-oligonucleotide via a simple aliphatic linker was shown to simultaneously recognize the major and minor grooves of DNA at subnanomolar concentrations. Development of versatile solid phase methodology expanded the protocols for the synthesis of pyrrole-imidazole polyamides to include polyamide-oligonucleotide conjugates. Utilizing the pyrrole-imidazole polyamide moiety as a sequence-specific dimerization domain afforded a class of cooperative polyamide-oligonucleotides which bound a 27 bp target with a $2.7 \text{ kcal mol}^{-1}$ increase in binding energy relative to the unlinked subunits. Introduction of a new linker design created an extended polyamide-oligonucleotide motif to specifically target 31 contiguous base pairs of DNA at subnanomolar concentrations. Polyamide-oligonucleotides combine the two binding models to create a paradigm for simultaneous recognition of the major-minor grooves of DNA with synthetic ligands.

For the pyrrole-imidazole polyamide model, sequence specificity depends on side-by-side aromatic amino acid pairings. An Im/Py pair distinguishes G•C from C•G and both of these from A,T base pairs. A Py/Py pair specifies A,T from G,C but does not distinguish A•T from T•A. To break this degeneracy a new aromatic amino acid, 3-hydroxypyrrole (Hp), was synthesized to test for pairings which discriminate A•T from T•A. Replacement of a single hydrogen atom with a hydroxy group in a Hp/Py

pairing regulates affinity and specificity by an order of magnitude. By incorporation of a third aromatic amino acid, hydroxypyrrole-imidazole-pyrrole polyamides form four ring pairings (Im/Py, Py/Im, Hp/Py, Py/Hp) which distinguish all four Watson-Crick base pairs in the minor groove of DNA (Figure 1.9).

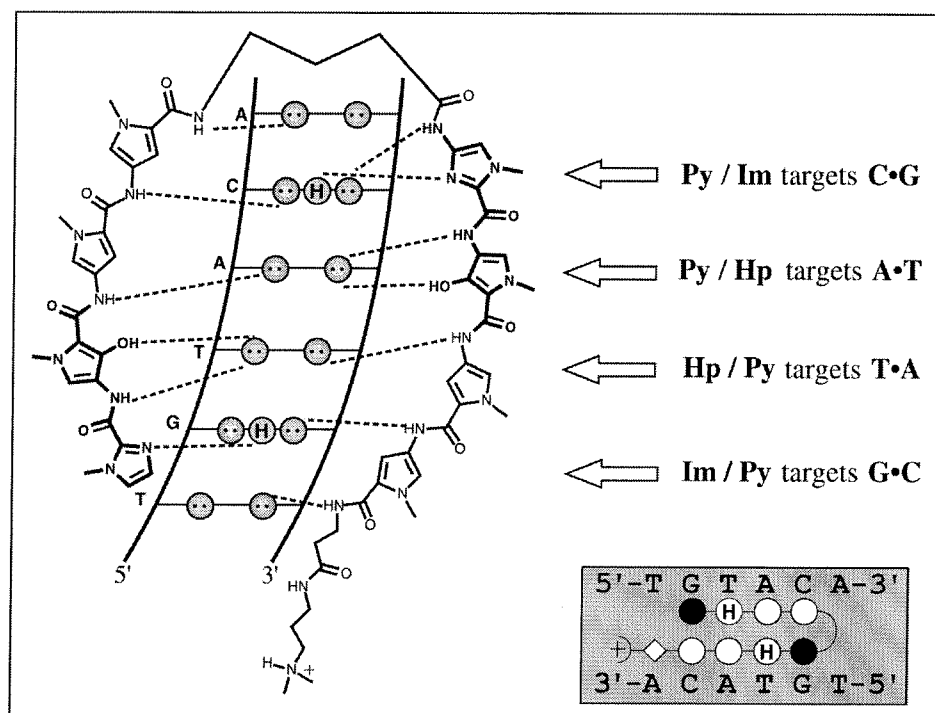
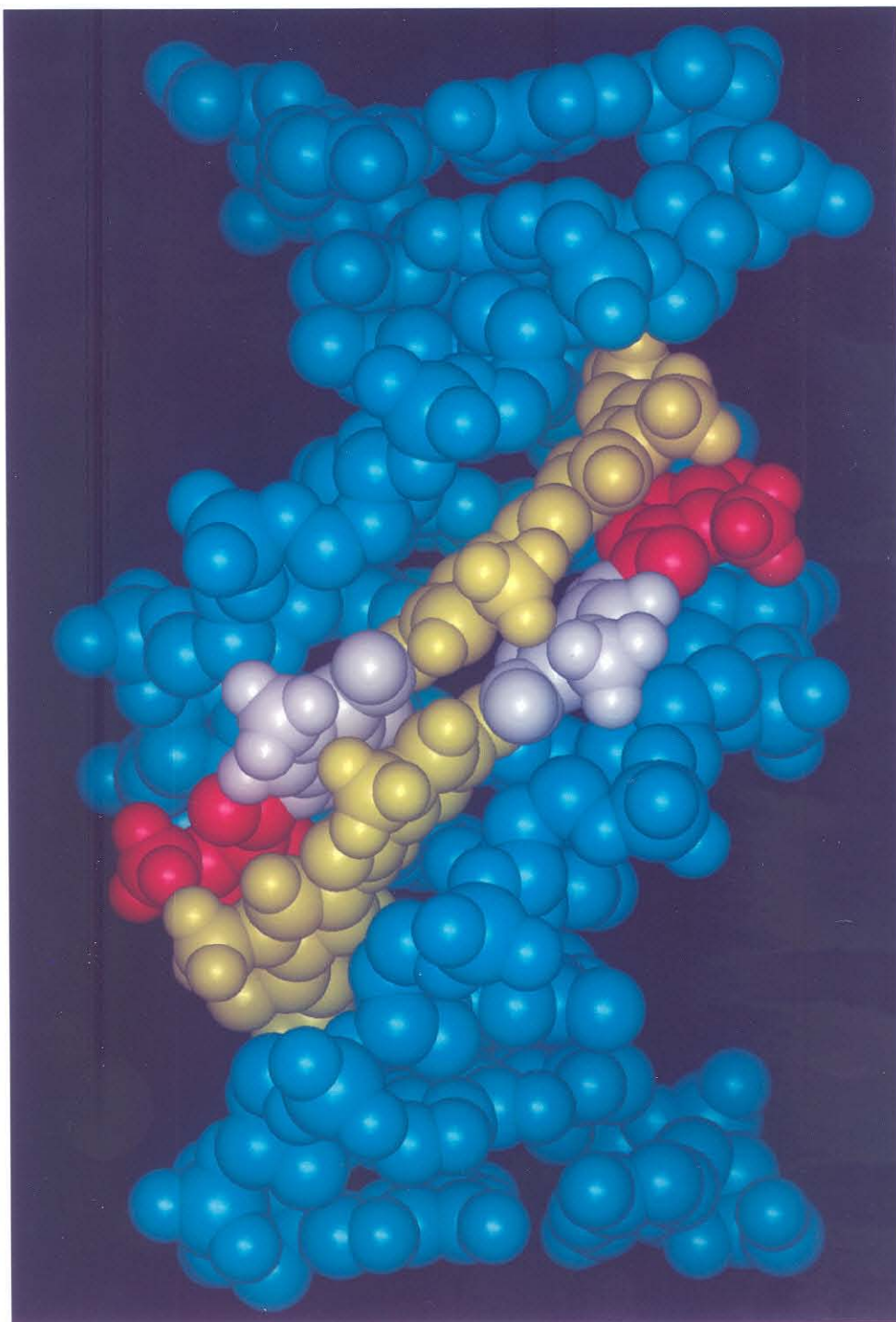


Figure 1.9. (Above) Small molecule polyamide, reads each of the four base-pairs in the DNA minor groove. (Next page) Space filling representation of the polyamide dimer ImHpPyPy- β -Dp bound in the minor groove of DNA. The figure was prepared using InsightII software and is derived from a high-resolution x-ray co-crystal structure of a polyamide dimer bound to DNA which was obtained in collaboration with the Rees group at the California Institute of Technology.



References

1. Watson, J.D. *Gene* **1993**, 135, 309.
2. Roeder, R.G. *TIBS* **1996**, 9, 327.
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.V.; Koopka, M.L. *Science* **1982**, 216, 475.
7. Principles of Nucliec Acid Stucture Saenger, W.; Springer-Verlag, New York, 1984.
8. Steitz, T.A. *Quart. Rev. Biophys.* **23**, 205.
9. (a) Pavletich, N.P.; Pabo, C.O. *Science* **1991**, 252, 809; (b) Ellenberger, T.E.; Brandl, C.J.; Struhl, K.; Harrison, S.C. *Cell* **1992**, 71, 1223; (c) Feng, J.A.; Johnson, R.C.; Dickerson, R.E. *Science* **1994**, 263, 348.
10. Kissinger, C.R.; Liu, B.; Martin-Blanco, E.; Kornberg, T.B.; Pabo, C.O. *Cell* **1990**, 63, 579.
11. Choo, Y.; Klug, A. *Curr. Opin. Struct. Biol.* **1997**, 7, 117.
12. Moser, H.E.; Dervan, P.B. *Science*, **1987**, 238, 645-650.
13. (a) Felsenfeld, G.; Davies, D.R.; Rich, A. *J. Am. Chem. Soc.* **1957**, 79, 2023-2024; (b) Lipsett, M.N. *J. Biol. Chem.* **1964**, 239, 1256-1260.
14. Beal, P.A.; Dervan, P.B. *Science*, **1991**, 251, 1360-1363.
15. (a) Sklenar, V.; Feigon, J. *Nature* **1990**, 345, 836-838; (b) Rajagapol, P.; Feigon, J. *Nature*, **1989**, 339, 637-640; (c) Koshlap, K.M.; Gillespie, P.; Dervan, P.B.; Feigon, J. *J. Am. Chem. Soc.* **1993**, 115, 7908-7909; (d) Radhakrishnan, I.; Patel, D.J. *Structure*, **1993**, 1, 135-152.
16. (a) Strobel, S.A.; Moser, H.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1998**, 110, 7927-7929; (b) Strobel, S.A., Dervan, P.B. *Science*, **1990**, 249, 73-75; (c) Strobel, S.A.; Doucette-Stamm, L.A.; Riba, L.; Housman, D.E.; Dervan, P.B. *Science*, **1991**, 254, 1639-2178.

17. (a) Maher, J.L.; Dervan, P.B.; Wold, B. *Biochemistry* **1992**, 31, 70; (b) Duvalvalentin, G.; Thuong, N.T.; Helene, C. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 504.
18. (a) Singleton, S.F.; Dervan, P.B.; *Biochemistry*, **1992**, 31, 10995-11003; (b) Singleton, S.F.; Dervan, P.B. *Biochemistry*, **1993**, 32, 13171-13179; (c) Singleton, S.F.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, 116, 10376-10382; (d) Best, G.C.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, 117, 1187-1193; (e) Greenberg, W.A.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, 117, 5016-5022.
19. (a) Moser, H.E.; Dervan, P.B. *Science* **1987**, 238, 645; (b) Thuong, N.T.; Helene, C. *Angew. Chem. Int. Ed. Engl.* **1993**, 32, 666.
20. (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.
21. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, 115, 2572; (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) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, 115, 9892; (b) Dwyer, T.J.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, 115, 9900; (c) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, 116, 3663; (d) Chen, Y.H.; Lown, J.W. *J. Am. Chem. Soc.* **1994**, 116, 6995; (e) Alsaid, N.H.; Lown, J.W. *Tett. Lett.* **1994**, 35, 7577; (f) Alsaid, N.H.; Lown, J.W. *Synth. Comm.* **1995**, 25, 1059; (g) Chen, Y.H.; Lown, J.W. *Biophys. J.* **1995**, 68, 2041; (h) Chen, Y.H.; Liu, J.X.; Lown, J.W. *Bioorg. Med. Chem. Lett.* **1995**, 5, 2223; (i) Chen, Y.H.; Yang, Y.W.; Lown, J.W. *J. Biomol. Struct. Dyn.* **1996**, 14, 341; (j) Singh, M.P.; Wylie, W.A.; Lown, J.W. *Mag. Res. Chem.* **1996**, 34, S55.
23. Cho, J.; Parks, M.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.*, **1995**, 92, 10389.
24. (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.; Poklar, N.A.; Gelfand, S.A.; Law, S.M.; Breslauer, K.J.; Baird, E.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.*

- 1996, 93, 8306; (g) White, S.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, 119, 8756; (h) White, S.; Baird, E.E.; Dervan, P.B. *Chem & Biol.* **1997**, 4, 569.
25. de Clairec, R.P.L.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1997**, 119, 7909.

Chapter 2

Sequence-Specific Recognition of DNA by a Major and Minor Groove Binding Ligand

***Abstract:** The two most powerful chemical approaches to date for the sequence-specific recognition of double helical DNA are the pyrrole-imidazole polyamides in the minor groove and oligonucleotide-directed triple-helix formation in the major groove. The design and synthesis of a hairpin pyrrole-imidazole polyamide-oligonucleotide, which specifically and simultaneously recognizes both the major and minor grooves of DNA at subnanomolar concentrations is reported. Binding affinity of the polyamide-oligonucleotide is enhanced at least 150-fold over either the individual oligonucleotide or polyamide alone.*

Publications: Szewczyk, Baird & Dervan *Angew. Chem. Intl. Ed.* **1996**, 35, 1487-1489.

Two nonbiological approaches for the sequence-specific recognition of double helical DNA that target a broad repertoire of sequences are oligonucleotide-directed triple helix formation¹ and pyrrole/imidazole polyamide-DNA complexes.² For recognition of the major groove of DNA, a pyrimidine oligonucleotide binds parallel to the purine Watson-Crick (W-C) strand.¹ Specificity is derived from thymine (T) recognition of adenine-thymine base pairs (T•AT base triplets) and protonated cytosine (C⁺) recognition of guanine-cytosine base pairs (C⁺•GC base triplets).¹ For recognition of the minor groove of DNA, polyamides containing *N*-methylimidazole (Im) and *N*-methylpyrrole (Py) amino acids are combined side-by-side and the DNA sequence specificity is controlled by the linear sequence of pyrrole and imidazole amino acids.^{2,3,4} Antiparallel pairing of imidazole (Im) opposite pyrrole (Py) targets a G•C base-pair, while pyrrole opposite imidazole targets a C•G base-pair.² A Py/Py pairing is partially degenerate and targets both T•A and A•T base-pairs.^{2,3} A hairpin polyamide motif with γ -aminobutyric acid (γ) serving as a turn-specific internal guide residue provides a versatile motif for high-affinity recognition of a wide variety of sequences in the minor groove of DNA.⁵ A key issue is to determine if the models for major and minor groove recognition can be combined within a single molecule without compromising the sequence-specificity of either recognition domain.

A hairpin polyamide linked to an 11-mer oligonucleotide specifically and simultaneously binds the major and minor grooves of DNA at

Figure 2.1 Left: A ribbon model depicting a hairpin pyrrole- imidazole polyamide-oligonucleotide simultaneously binding the major and minor grooves of double helical DNA. The shaded and unshaded circles represent *N*-methylimidazole and *N*-methylpyrrole rings, respectively, and the unshaded diamond represents β -alanine. Right: Two-dimensional models depicting the molecular interactions responsible for sequence specific recognition. (top) The T•AT and $^{\text{Me}}\text{C}^+\bullet\text{GC}$ base triplets formed by Hoogsteen hydrogen bonding of T to a Watson-Crick AT base pair and protonated $^{\text{Me}}\text{C}$ to a Watson-Crick GC base pair. The bases in the third strand are bold. (bottom) Expected complex of ImPyPy- γ -ImPyPy- β with the site 5'-TGACA-3'. Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines, and circles containing an H represent the N2 hydrogen on guanine. Putative hydrogen bonds are illustrated by dashed lines.

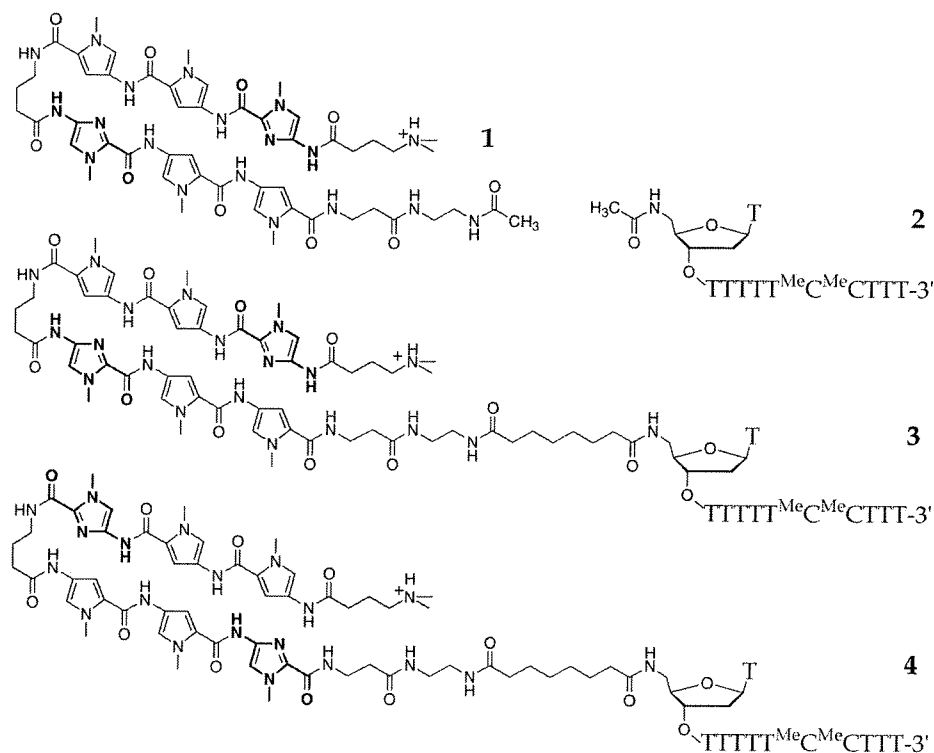


Figure 2.2 Structure of the hairpin polyamide, **1**; pyrimidine oligonucleotide, **2**; and the match and mismatch pyrrole-imidazole polyamide-oligonucleotides **3** and **4**, respectively.

subnanomolar concentration (Figures 2.1 and 2.2). A pyrimidine oligonucleotide, 5'-TTTTTT^{Me}C^{Me}CTTT-3', covalently linked by a 12 atom chain to the polyamide of sequence composition ImPyPy- γ -ImPyPy- β (β = β -alanine) recognizes 16 base pairs of double-helical DNA by formation of a specific hairpin pyrrole-imidazole polyamide-DNA complex in the minor groove and a specific triple helix in the major groove (Figure 2.1). According to the two binding models, the hybrid ligand-DNA complex affords ten discrete hydrogen bonds for recognition of 5 base pairs in the minor groove

(5'-TGACA-3') and 22 discrete hydrogen bonds for recognition of 11 base pairs in the major groove (5'-AAAAAAGGAAA-3') (Figure 2.1).

Results and Discussion.

Synthesis of extended polyamide-oligonucleotides. The alloc-protected 5'-amino-thymidine phosphoramidite **7** was synthesized in 3 steps from thymidine (Figure 2.3).^{6,7,8} First, the 5'-hydroxyl is converted to the iodide by treatment with methyl triphenoxyposphonium iodide and displaced with lithium azide. Catalytic reduction of the azido-nucleoside **5** afforded the 5'-amine, which was protected as the 5'-allyl carbamate **6** by reaction with allylchloroformate. Phosphitylation of the 3'-hydroxyl afforded the desired phosphoramidite **7** which was incorporated using standard β -cyanoethyl phosphoramidite chemistry to provide an 11-mer oligonucleotide, 5'-TTTTTT^{Me}C^{Me}CTTT-3', containing a 5'-alloc•protected amino group suitable for post-synthetic modification.

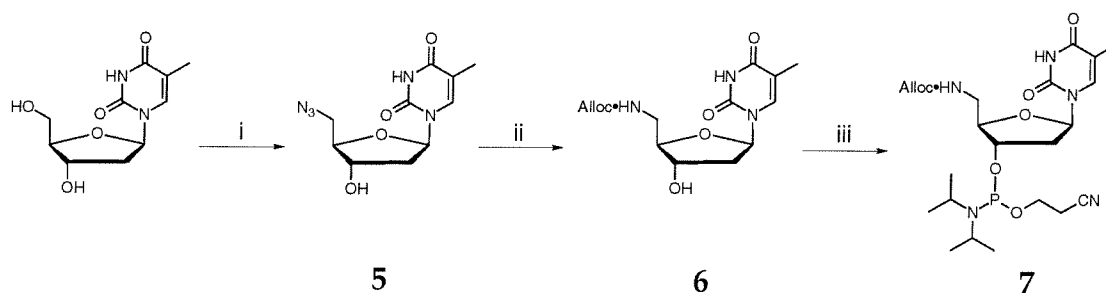


Figure 2.3 (i) a) methyl triphenoxyposphonium iodide, DMF; b) LiN_3 , 55 °C; (ii) a) 10% Pd/C, H_2 , MeOH; b) allylchloroformate, DIEA; (iii) DCM, TEA, 2-cyanoethyl *N,N*- diisopropyl- chlorophosphoramidite.

The alloc was removed on resin by treatment with a palladium catalyst⁹ followed by reaction with disuccinimidyl suberate to provide a support bound 5'-modified *N*-hydroxysuccinimide ester-oligonucleotide (Figure 2.4). The polyamide ImPyPy- γ -ImPyPy- β , modified with a free primary amine group at the carboxy terminus, was prepared by solid phase methods¹⁰ and allowed to react with either acetic anhydride to provide control hairpin **1** or the support bound *N*-hydroxysuccinimide ester-oligonucleotide to provide the conjugate, ImPyPy- γ -ImPyPy- β -(linker)-TTTTT^{Me}C^{Me}CTTT-3' **3**, upon deprotection (Figure 2.4).¹¹ Similarly, a control conjugate **4** was synthesized with a “reverse” polyamide sequence composition PyPyIm- γ -PyPyIm- β . Polyamide-oligonucleotides **3** and **4** were purified by reverse phase HPLC and the molecular composition was verified by MALDI-TOF mass spectroscopy, UV spectroscopy, and HPLC analysis of enzymatic phosphodiester hydrolysis products.

Energetics by quantitative DNase I footprinting. A DNA fragment was constructed which contains the 18 base pair site 5'-TGACATTAAAAAAGGAAA-3' composed of a 5 base pair polyamide binding site separated by 2 base pairs from an 11 base pair triple helix site. For controls, isolated 5 base pair polyamide and 11 base pair triple helix binding sites were also present on the fragment.

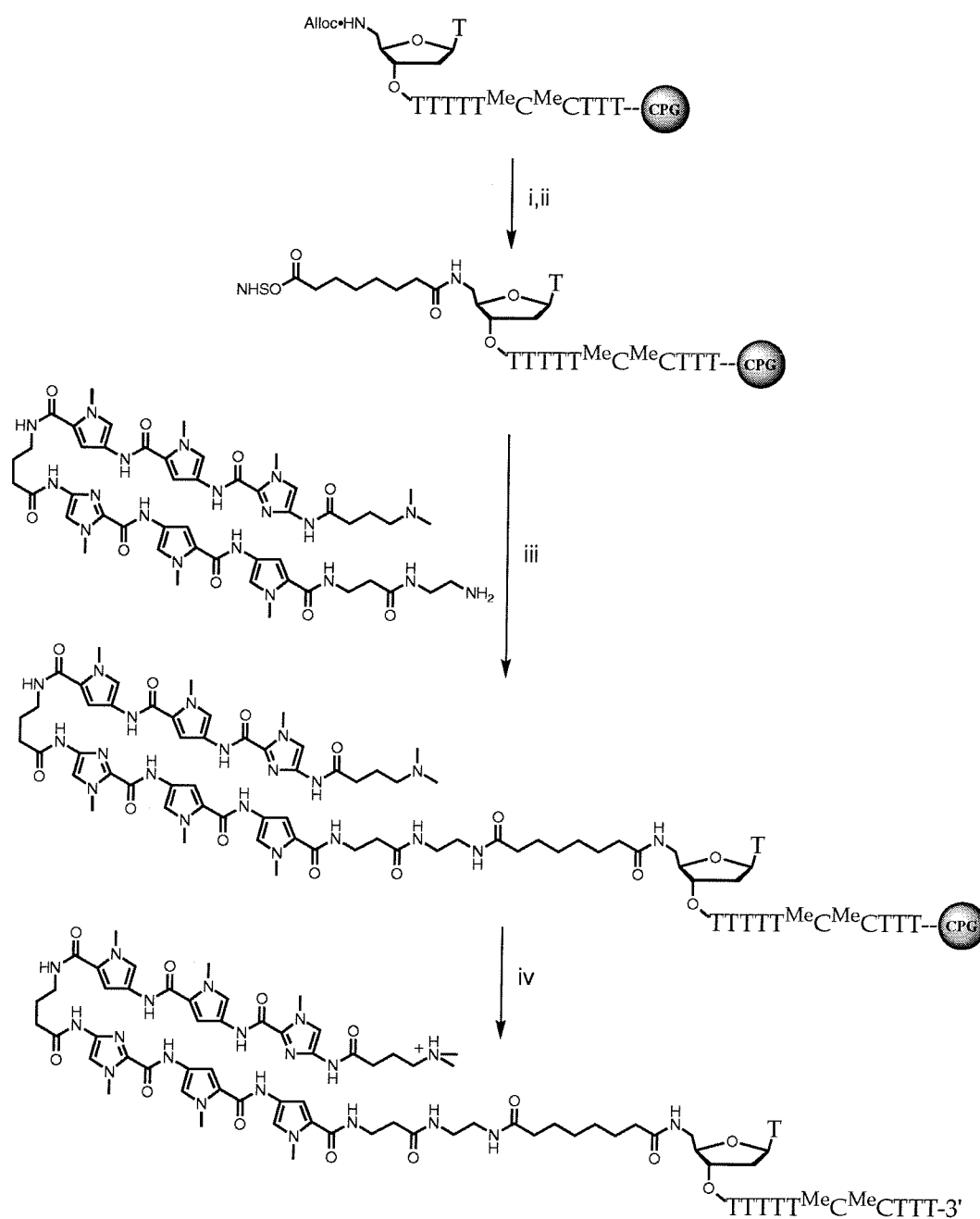


Figure 2.4 (i) $\text{Pd}_2(\text{dba})_2 \cdot \text{CHCl}_3$, PPh_3 ; (ii) disuccinimidyl suberate, DIEA, DMF; (iii) DIEA, DMF; (iv) 0.1 M NaOH, 55 °C.

Quantitative DNase I footprint titration experiments on a 302 base pair restriction fragment afford the binding affinity of the designed ligands (Table 2.1). Polyamide 1 binds to 5 base pairs of the target site 5'-TGACATTAAAAAAGGAAA-3' with an equilibrium association constant, $K_a = (1.3 \pm 0.3) \times 10^8 \text{ M}^{-1}$. Unmodified oligonucleotide 2 binds 11 base pairs 5'-TGACATTAAAAAAGGAAA-3' with a equilibrium association constant, $K_a = (2.5 \pm 0.8) \times 10^5 \text{ M}^{-1}$. The polyamide-oligonucleotide 3 binds the designated target site 5'-TGACATTAAAAAAGGAAA-3' with an equilibrium association constant, $K_a > 2 \times 10^{10} \text{ M}^{-1}$, representing at least a 150-fold increase in binding affinity relative to the unlinked subunits (Figure 2.5). In controls, the binding affinity of 3 at either the 5 base pair polyamide binding site, 5'-TGACATTAGCTCGTAATG-3', ($K_a < 1.0 \times 10^5 \text{ M}^{-1}$) or oligonucleotide 11 base pair site, 5'-ACCGGTTAAAAAAGGAAA-3', ($K_a = 1.9 \times 10^9 \text{ M}^{-1}$) is reduced by approximately 10-fold relative to binding at the designated match site. Mismatch polyamide-oligonucleotide 4 binds the 18 base pair target, 5'-TGACATTAAAAAAGGAAA-3, ($K_a = (6.1 \pm 0.2) \times 10^7 \text{ M}^{-1}$) with 300-fold reduced affinity relative to match conjugate 3.

Polyamide-oligonucleotide 3 binds with subnanomolar affinity only at the designated target site, indicating that the $3.0 \text{ kcal mol}^{-1}$ binding enhancement results from *simultaneous* sequence-specific recognition of the major and the minor grooves. By mimicking the function of certain sequence-specific DNA binding proteins,¹² hairpin polyamide-

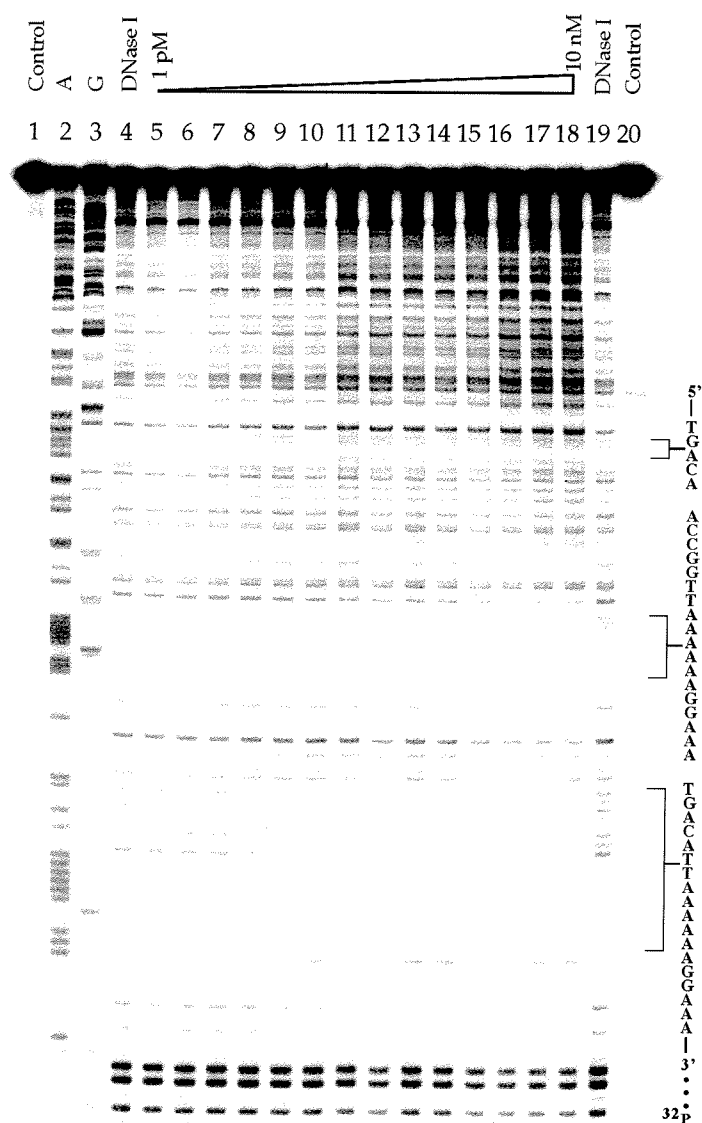


Figure 2.5 DNase I footprinting titration of pyrrole- imidazole polyamide-oligonucleotide **3** with a ^{32}P end-labeled 302 base pair restriction fragment. Gray scale representation of a storage phosphor autoradiogram of a 8% denaturing polyacrylamide gel representing a 10-fold change in the signal from the lowest (5 arbitrary units, white) to highest (50 arbitrary units, black) intensities. The binding site is shown to the right site of the autoradiogram. Lane 1 and 20, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4 and 19, DNase I standard; lanes 5-18 contain 1 pM, 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 150 pM, 250 pM, 400 pM, 650 pM, 1.5 nM, 2.5 nM, 6.5 nM, 10 nM polyamide-oligonucleotide **3**, respectively.

oligonucleotides expand the sequence composition repertoire targeted by artificial methods and create a new motif for the design of molecules for the sequence-specific recognition of both grooves of the DNA double helix.

Table 2.1 Equilibrium association constants (M^{-1})^{a,b}

ligand	K_a (M^{-1})
polyamide 1	$1.3 (\pm 0.3) \times 10^8$
oligonucleotide 2	$2.5 (\pm 0.8) \times 10^5$
polyamide-oligonucleotide 3	$> 2 \times 10^{10}$
(mismatch) polyamide-oligonucleotide 4	$6.6 (\pm 0.2) \times 10^7$

^aExperiments performed at 22 °C in the presence of 10 mM NaCl, 10 mM Bis Tris•HCl (pH 7.0), and 250 μ M spermine. ^b Values reported are the mean values obtained from three or more DNase I footprint titration experiments. The standard deviation for each data set is indicated in parentheses.

Experimental Section.

Polyamide synthesis and conjugation reactions for polyamide-oligonucleotides **3** and **4** were performed by Eldon Baird.

Materials. 0.45 M sublimed tetrazole in acetonitrile, THF/Lutidine/Ac₂O (8:1:1), 10% MeIm in THF, 0.1M I₂ in THF/Pyridine/H₂O, 3% TCA/DCM, 5-Methylcytidine CE (2-cyanoethyl) phosphoramidite, and bulk 500Å dT-Icaa-CPG were purchased from Glen Research. All 10 µmole synthesis columns were packed manually from bulk support. 1,2,3- Benzotriazol-1-yl-4-[(*tert*-Butyloxy)carbonyl]-amino]-1-methylpyrrole-2-carboxylate and 1,2,3- Benzotriazol-1-yl-4-[(*tert*-Butyloxy)carbonyl]-amino]-1-methylimidazole-2-carboxylate were prepared as previously described.^{16b,17} Trifluoroacetic acid (TFA) was purchased from Halocarbon. *N,N*-diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF), 0.0002M potassium cyanide/pyridine, and acetic anhydride (Ac₂O) were purchased from Applied Biosystems, and dicyclohexylcarbodiimide (DCC) and hydroxybenzotriazole (HOBt) were from Peptides International. Boc-GABA was from NOVABiochem, dichloromethane (DCM), reagent grade from EM. Thiophenol was purchased from Aldrich. A shaker for manual solid phase synthesis was obtained from Milligen. Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. and are used for filtration of CPG. Screw-cap glass peptide synthesis reaction vessels (5 ml) with a #2 scintered glass frit were made at the Caltech glass shop as described.¹⁰ NMR were recorded on a GE 300 instrument

operating at 300MHz (^1H). Chemical shifts are reported in ppm relative to the solvent residual signal; UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed either on a HP 1090 M analytical HPLC or a Beckman Gold system using a RAINEN C18, Microsorb MV, $5\mu\text{m}$, 300×4.6 mm reversed phase column in 100 mM Ammonium Acetate, pH 4.9 with acetonitrile as eluent and a flow rate of 1.0 ml/min, gradient elution 1.0% acetonitrile/min. Polyamide-oligonucleotides were purified by reverse phase HPLC using a linear gradient from 0 to 40% acetonitrile in 55 min., 100 mM triethylammonium acetate, pH 7.0.

Synthesis of Polyamides.

$^{\text{DM}}\gamma\text{-ImPyPy-}\gamma\text{-ImPyPy-}\beta\text{-ED-Ac}$ 1. $^{\text{DM}}\gamma\text{-ImPyPy-}\gamma\text{-ImPyPy-}\beta\text{-Pam-resin}$ was synthesized in stepwise fashion by machine-assisted solid phase methods.¹⁰ A sample of $^{\text{DM}}\gamma\text{-ImPyPy-}\gamma\text{-ImPyPy-}\beta\text{-Pam-resin}$ (240 mg, 0.2 mmol/g) was treated with neat ethylene diamine (2 ml) and heated to (55 °C) with periodic agitation for 18 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. $^{\text{DM}}\gamma\text{-ImPyPy-}\gamma\text{-ImPyPy-}\beta$ containing a free primary amine is recovered upon lyophilization of the appropriate fractions. A sample was

dissolved in DMF (0.2 M), treated with acetic anhydride (37 °C, 1 h), 0.1% (wt/v) TFA added (6 ml) and the resulting solution purified by reversed phase HPLC. $^{DM}\gamma$ -ImPyPy- γ -ImPyPy- β **1** is recovered upon lyophilization of the appropriate fractions as a white powder (17 mg, 36% recovery): UV (H₂O) λ_{max} 260, 310 (50 000) MALDI-TOF MS (monoisotopic) [M+H] 1065.0 (1065.1 calc for C₅₁H₆₇N₁₉O₁₀).

Synthesis of Polyamide Oligonucleotides.

$^{DM}\gamma$ -ImPyPy- γ -ImPyPy- β -(linker)-TTTTT^{Me}C^{Me}CTTT-3' **3**. N-alloc-5'-amino-thymine-TTTTT^{Me}C^{Me}CTTT-CPG was synthesized on an Applied Biosystems Model 380B DNA synthesizer using standard β -cyanoethyl phosphoramidite chemistry on 10 μ mol scale. The alloc protecting group was removed on solid support by combining the oligonucleotide on CPG with tris(dibenzylideneaceto)-di-palladium(0)•CHCl₃, (Pd₂(dba)₂•CHCl₃, 196 mg, 0.19 mmol), triphenyl phosphine (686 mg, 2.62 mmol), and a freshly made solution of 0.25 M n-butyl ammonium formate in THF (7 ml) in a reaction vessel with shaking for 6 h at 24 °C.⁹ The support was washed with sequentially with THF, acetone, 0.1 M sodium diethyldithiocarbamate (DTTC, aq, 2x), acetone, water, DTTC (2x), acetone, water, DCM, MeOH, DCM, DMF. A sample of support (2.8 μ mol) was treated with disuccinimidyl-suberate (65 mg, 250 μ l DIEA, 750 μ l DMF, 37 °C) to provide a support bound N-hydroxysuccinimidyl ester which was allowed to react with $^{DM}\gamma$ -ImPyPy- γ -

ImPyPy- β containing a free primary amine (17 mg, 0.3 ml DMF, 0.1 ml DIEA, 37 °C, 12 h). $^{DM}\gamma$ -ImPyPy- γ -ImPyPy- β -(linker)-TTTTTT^{Me}C^{Me}CTTT-CPG was isolated by filtration, washed sequentially with DMF, DCM, MeOH, diethyl ether and dried *in vacuo*. The support was treated with 0.1 M NaOH (2 ml) and heated (55 °C) for 12 h. The reaction mixture was then filtered to remove CPG, 0.1 M triethylammonium acetate (pH = 7.0) added (6 ml) and the resulting solution purified by reversed phase HPLC. $^{DM}\gamma$ -ImPyPy- γ -ImPyPy- β -(linker)-TTTTTT^{Me}C^{Me}CTTT-CPG is recovered upon lyophilization of the appropriate fractions (400 nmol, 14% recovery): UV (H₂O) λ_{max} 260 (138 000), 310 (54 000); MALDI-TOF MS [M-H] obsd. 4487 (calc 4483).

$^{DM}\gamma$ -PyPyIm- γ -PyPyIm- β -(linker)-TTTTTT^{Me}C^{Me}CTTT-3' 4. $^{DM}\gamma$ -PyPyIm- γ -PyPyIm- β -(linker)-TTTTTT^{Me}C^{Me}CTTT-CPG was synthesized by machine assisted solid phase methods as described for 3. A sample of $^{DM}\gamma$ -PyPyIm- γ -PyPyIm- β -(linker)-TTTTTT^{Me}C^{Me}CTTT-CPG was treated with 0.1 M NaOH (2 ml) and heated (55 °C) for 12 h. The reaction mixture was then filtered to remove CPG, 0.1 M triethylammonium acetate (pH = 7.0) added (6 ml) and the resulting solution purified by reversed phase HPLC. $^{DM}\gamma$ -PyPyIm- γ -PyPyIm- β -(linker)-TTTTTT^{Me}C^{Me}CTTT-3' is recovered upon lyophilization of the appropriate fractions (900 nmol, 32% recovery): UV (H₂O) λ_{max} 260 (138 000), 310 (54 000); MALDI-TOF MS [M-H] obsd. 4484 (calc 4483).

5'-azido-2',5'-dideoxythymidine 5. To a stirred solution of thymidine (25 g, 103 mmol) in DMF (250 ml) was added methyltriphenoxyphosphonium iodide (56.02 g, 123.6 mmol) and the mixture stirred for 30 minutes. Lithium azide (5 g, 310 mmol) was added and the resulting slurry stirred (55 °C, 2 h). Methanol (10 mL) was added and the solvent removed *in vacuo*. The resulting syrup was absorbed onto silica and chromatographed: SiO₂, gradient elution 50 to 100% ethyl acetate : hexanes, to afford **5** as a white powder (14 g, 52.4 mmol, 51%). ¹H NMR (DMSO-d₆); δ 11.33 (s, 1 H), 7.48 (s, 1 H), 6.19 (t, 1 H, *J* = 7.0 Hz), 5.42 (d, 1 H, *J* = 3.5 Hz), 4.18 (br d, 1 H, *J* = 2.4 Hz), 3.83 (m, 1 H), 3.54 (d, 2 H, *J* = 5.1 Hz), 2.22 (m, 1 H), 2.06 (m, 1 H), 1.77 (s, 3 H).

5'-N-[(allyloxy)carbonyl]-2',5'-dideoxythymidine 6. To a solution of **5** (10 g, 37.4 mmol) in MeOH (200 mL), 10% Pd/C (1 g) was added and the suspension stirred under hydrogen (1 atm) for two hours. Pd/C was removed by filtration through celite. Triethylamine (17 mL) and allylchloroformate (4.51 g, 37.4 mmol) were added and the solution stirred for 30 minutes. The mixture was absorbed onto silica and chromatographed: SiO₂, gradient elution 50 to 100% ethyl acetate : hexanes, to afford **6** as a white powder (11 g, 33.8 mmol, 90%). ¹H NMR (DMSO-d₆); δ 11.29 (s, 1 H), 7.48 (s, 1 H), 7.43 (t, 1 H, *J* = 6.5 Hz), 6.12 (t, 1 H, *J* = 7.0 Hz), 5.88 (m, 1 H), 5.29 (br s, 1 H), 5.25 (dd, 1 H, *J* = 1.5 Hz), 5.14 (dd, 1 H, *J* = 1.1 Hz), 4.46 (d, 2 H, *J* = 5.3 Hz), 4.14 (br s, 1 H), 3.74 (m, 1 H), 3.22 (m, 2 H), 2.05 (m, 2 H), 1.78 (s, 3 H).

5'-N-[(allyloxy)carbonyl]-2',5'-dideoxythymidine -3'-O-(2-cyanoethyl N,N-diisopropylphosphoramidite 7. To a stirred suspension of **6** (0.536 g, 1.65 mmol) in DCM (3 mL) and triethylamine (1 mL) was added 2-cyano-N,N-isopropylchlorophosphoramidite (0.467 g, 1.97 mmol) in DCM (1 mL), and allowed to react 30 minutes. The reaction was quenched by addition of MeOH (1 mL) and diluted with DCM (50 mL). The organic phase was dried over MgSO₄ and the solvent removed. The resulting oil was dissolved in a small amount of ethyl acetate and precipitated with cold hexanes and dried *in vacuo* to afford **7** (0.606 g, 1.15 mmol, 70%).

Preparation of ³²P-labeled DNA.

The plasmid pCON1 was prepared by hybridization of two complementary sets of synthetic oligonucleotides 5'-d(CCGGGAACGTAATGACATTAGCTCGTAATGTTATTAGCTCGCTAAG GTCATTACCGGTAAAAAAGGAAACGCTCGA)-3', 5'-d(GGCGTCGAGCGTTTCCTTTTTTAACCGGTAATGACCTTAGCGAGCTA ATAACATTACGAGCTAATGTCATTACGTTC)-3' and 5'-d(CGCCGCTAATGACATTAAAAAAGGAAACGCTCGAGCTCGTGGGCCC TAGCGTCGTAGCTAGCGTCGTCTTAAG)-3', 5'-(TCGACTTAAGACGACGCTAGCTACGACGCTAGGGCCCACGAGCTCGA GCGTTTCCTTTTTTAATGTCATTAGC)-3' which were subsequently ligated into pUC19 previously cleaved with *Ava* I and *Sal* I. The 3'-³²P end-labeled *Afl* II/*Fsp* I fragment was prepared by digesting the plasmid pCON1 with *Afl*

II I and *Fsp* I and simultaneously filling in using Sequenase, [α - 32 P]-deoxyadenosine-5'-triphosphate, and [α - 32 P]-thymidine-5'-triphosphate. The 302 bp fragment was isolated by nondenaturing gel electrophoresis. A and G sequencing were carried out as described.^{13,14} Standard methods were used for all DNA manipulations.¹⁵

Quantitative DNase I footprint titrations. The quantitative footprint titration experiments were executed in a total volume of 400 μ l with a final concentration of each species as indicated. The ligands were added to solutions of radiolabeled restriction fragment (20 000 cpm), NaCl (10 mM), Bis Tris•HCl (10 mM, pH 7.0), and spermine (250 μ M), incubated for 24 hours at 22 °C. Footprinting reactions were initiated by addition of 10 μ l stock solution of DNase I (2.4 units/ml) containing MgCl₂ (200 mM), CaCl₂ (200 mM), Bis Tris•HCl (10 mM), and glycerol (20%) and allowed to proceed for 6 minutes at 22 °C. The reactions were quenched by addition of 50 μ l of a solution made up of 107 μ l water, 788 μ l NaCl (4 M), 425 μ l EDTA (0.5 M, pH = 8.0), 40 μ l calf thymus DNA (1 mM bp), 40 μ l glycogen (20 mg / ml), and ethanol precipitated. The reactions resuspended in 80% formamide loading buffer, denatured with heat (85 °C, 10 min.) and electrophoresed on an 8% polyacrylamide denaturing gel at 2000 V for 1 hour.

Quantitation and data analysis. Data from the footprint titration gels were obtained using a Molecular Dynamics 400S PhosphorImager followed by quantitation using ImageQuant software (Molecular Dynamics). Background-corrected volume integration of rectangles encompassing the footprint sites and a reference site at which DNase I reactivity was invariant across the titration generated values for the site intensities (I_{site}) and the reference intensity (I_{ref}). The apparent fractional occupancy (θ_{app}) of the sites were calculated using the equation:

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

where I_{site}° and I_{ref}° are the site and reference intensities, respectively, from a control lane to which no polyamide was added. The $([L]_{\text{tot}}, \theta_{\text{app}})$ data points were fit to a general Hill equation (eq 2) by minimizing the difference between θ_{app} and θ_{fit} :

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

where $[L]_{\text{tot}}$ is the total ligand concentration, K_a is the apparent first-order association constant, and θ_{min} and θ_{max} are the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. The data were fit using a nonlinear least-squares fitting procedure with K_a , θ_{max} , and θ_{min} as the adjustable parameters. For polyamide ImPyPy- γ -ImPyPy- β 1

binding isotherms for the 5'-TGACA-3' target site were adequately fit by Langmuir isotherms (eq 2, $n=1$), consistent with formation of a 1:1 polyamide-DNA complexes. The mismatched polyamide- oligonucleotide **4**, only footprinted the 11 bp major groove binding site, 5'-AAAAAAGGAAA-3', and was fit by Langmuir isotherms (eq 2, $n=1$), consistent with formation of a local triple helix in the minor groove and no binding in the minor groove. For the matched conjugate **3**, binding isotherms for the 5'-TGACATTAAAAAAGGAAA-3' target site were adequately fit by Langmuir isotherms (eq 2, $n=1$), consistent with formation of a 1:1 polyamide-oligonucleotide:DNA complex. Three sets of data were used in determining each association constant. The method for determining association constants used here involves the assumption that $[L]_{\text{tot}} \approx [L]_{\text{free}}$, where $[L]_{\text{free}}$ is the concentration of polyamide- oligonucleotide free in solution (unbound). For very high association constants this assumption becomes invalid, resulting in underestimated association constants. In the experiments described here, the DNA concentration is estimated to be ~ 10 pM. As a consequence, a measured association constant of $1 \times 10^{10} \text{ M}^{-1}$ underestimates the true association constant by a factor of 4.

References

1. (a) Moser, H.E.; Dervan, P.B. *Science* **1987**, 238, 645; (b) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhouh, N.; Decout, J. L.; Thoung, N. T.; Lhomme, J.; Helene, C. *Nucleic Acids Res.* **1987**, 15, 7749.
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. *Biochem.* **1993**, 32, 11385.
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.
4. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, 115, 2572; (b) Geierstanger, B.H.; Dwyer, T.J.; Bathini, Y.; Lown, J.W.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, 115, 4474; (c) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, 115, 9892; (d) Dwyer, T.J.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, 115, 9900; (e) Geierstanger, B.H.; Jacobsen, J.P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E.; *Biochem.* **1994**, 33, 3055; (f) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, 116, 3663; (g) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, 266, 646; (h) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, 117, 3325.
5. (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, 6141.
6. (a) Tetzlaff, C.N.; Schwoppe, I.; Blecinski, C.F.; Steinberg, J.A.; Richert, C. *Tet. Let.* **1998**, 39, 4215; (b) James, K.D.; von Keidrowski, G.; Ellington, A.D. *Nucltsds. Nucltds.* **1997**, 16, 1821; (c) Bannwarth, W. *Helv. Chim. Acta.* **1998**, 71, 1517.
7. Bang-Hofman, N. *Acta. Chem. Scand.* **3**, 581-582.
8. Yamamoto, I.; Sekine, M.; Hata, T. *J. Chem. Soc. Perkin I.* **1978**, 306-310.
9. Hayakawa, Y.; Wakabayashi, S.; Kato, H.; Noyori, R. *J. Am. Chem. Soc.* **1990** 112, 1691.
10. Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996** 118, 6141.

11. (a) Haralambidis, J.; Duncan, L.; Angus, K.; Tregear, G.W. *Nuc. Acid. Res.* **1990**, *18*, 493-499; (b) Haralambidis, J. Duncan, L.; and Tregear, G.W. *Tet. Lett.* **1987**, *28*, 5199-5202; (c) Tong, G.; Lawlor, J.M.; Tregear, G.W.; Haralambidis, J. *J. Org. Chem.* **1993**, *58*, 2223-2231; (d) Tung, C.; Rudolph, J.; and Stein, S. *Bioconj. Chem.* **1991**, *2*, 464-465; (e) Bongratz, J.; Aubertin, A.; Milhaud, P.G.; and Lebleu, B. *Nuc. Acid. Res.* **1994**, *22*, 4681-4688; (f) Zhu, T.; Stein, S. *Bioconj. Chem.* **1994**, *5*, 312-315; (g) Sinyakov, A.N.; Lokhov, S.G.; Kutyavin, I.V.; Gamper, H.B.; Meyer, R.B. *J. Am. Chem. Soc.* **1995**, *117*, 4995.
12. (a) Sluka, J.P.; Horvath, S.J.; Bruist, M.F.; Simon, M.I.; Dervan, P.B. *Science* **1987**, *238*, 1129; (b) Sluka, J.P.; Horvath, S.J.; Glasgow, A.C.; Simon, M.I.; Dervan, P.B. *Biochem.* **1990** *29*, 6551; (c) Feng, J.; Johnson, R.C.; Dickerson, R.E. *Science* **1994**, *263*, 348.
13. Maxam, A.M.; Gilbert, W.S. *Methods Enzymol.* **1980**, *65*, 499-560.
14. Iverson, B.L.; Dervan, P.B. *Methods Enzymol.* **1987**, *15*, 7823-7830.
15. Sambrook, J., Fritsch, E.F.; Maniatis, T. (1989). *Molecular Cloning*. (2nd edn). Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

Chapter 3

Cooperative Triple-Helix Formation via a Sequence Specific Minor Groove Dimerization Domain

***Abstract.** Pyrrole- imidazole polyamide-oligonucleotides represent a new class of designed molecules for the sequence-specific recognition of the minor-major grooves of DNA. An 11 base pyrimidine oligonucleotide covalently tethered to a three ring pyrrole- imidazole polyamide is found to bind as a homodimer to a 31 base pair duplex-DNA target. Specific recognition is achieved by simultaneous formation of a 2:1 polyamide-DNA complex in the minor groove and oligonucleotide-directed triple-helix formation in the major groove. The cooperative polyamide-oligonucleotide binds with 100-fold increased affinity relative to unlinked-polyamide or oligonucleotide alone.*

Publications: Szewczyk, Baird & Dervan *J. Am. Chem. Soc.* **1996**, 118, 6778-6779.

Recently described 2:1 polyamide-DNA complexes¹⁻⁷ and oligonucleotide-directed triple-helix formation^{8,9} provide powerful *chemical* approaches for the sequence-specific recognition of DNA. In one triple helical model, a pyrimidine oligonucleotide binds specifically to a purine tract in the major groove of double helical DNA parallel to the purine Watson-Crick (W-C) strand.⁸ Specificity is derived from thymine (T) recognition of adenine-thymine base pairs (T•AT base triplets) and protonated cytosine (C⁺) recognition of guanine-cytosine base pairs (C⁺•GC base triplets).⁸ In the 2:1 polyamide-DNA model, polyamides containing *N*-methylimidazole (Im) and *N*-methylpyrrole (Py) amino acids are combined in antiparallel side-by-side dimeric complexes within the minor groove of DNA, and the DNA sequence specificity is controlled by the linear sequence of pyrrole and imidazole amino acids.¹ Pairing imidazole (Im) opposite pyrrole (Py) targets a G•C base-pair, while pyrrole opposite imidazole targets a C•G base-pair.¹ A Py/Py pairing is partially degenerate and targets both T•A and A•T base-pairs.¹⁻³ A key issue was to determine if a simple aliphatic linker could mediate cooperative binding without compromising the sequence specificity of either recognition element.

Cooperative ligand:DNA complexes. Cooperative interactions between DNA binding proteins are critical to the regulation of gene expression.¹⁰ When cooperative interactions occur between two oligonucleotides bound at neighboring sites on double helical DNA, the specific binding of each

Figure 3.1 Ribbon model depicting two pyrrole- imidazole polyamide-oligonucleotides binding as a dimer to double-helical DNA.

oligonucleotide is enhanced.¹¹ Pyrimidine oligodeoxyribonucleotides 11 bases in length bind cooperatively to abutting sites on double-helical DNA by triple-helix formation.^{11,12} Cooperative interactions up to 2.3 kcal mol⁻¹ were obtained by addition of dimerization domains to oligonucleotides such as those capable of forming a short Watson-Crick helix.¹³ The designed polyamide ImPyPy binds sequence specifically to the minor groove of DNA at 5'-(T,A)G(T,A)C(T,A)-3' sites as a cooperative side-by-side antiparallel homodimer.¹ We report here that a pyrrole- imidazole polyamide-oligonucleotide can bind cooperatively to a double stranded DNA template via a sequence-specific polyamide dimerization domain (Figure 3.1).

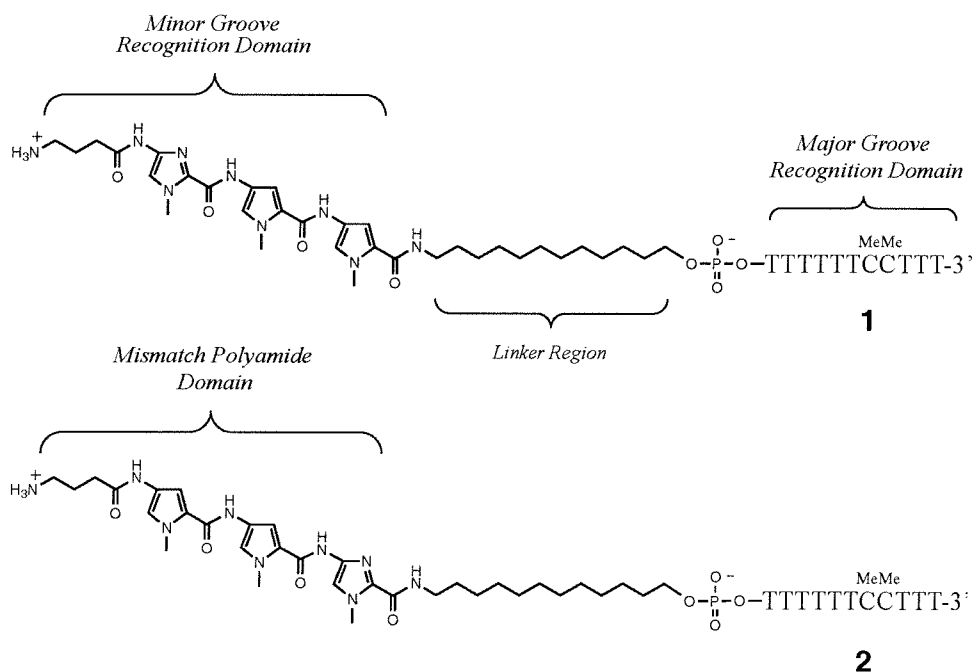


Figure 3.2 The cognate pyrrole- imidazole polyamide-oligonucleotide **1** and pyrrole- imidazole polyamide-oligonucleotide **2** containing a mismatch polyamide moiety.

A pyrimidine oligonucleotide 11 bases in length covalently tethered to the polyamide H₂N- γ -ImPyPy (γ = γ -aminobutyric acid), *cooperatively* binds as a homodimer to 27 non-contiguous base pairs of double helical DNA by simultaneous formation of a 2:1 polyamide-DNA complex in the minor groove and triple helix formation in the major groove (Figure 3.1). The putative recognition complex affords 44 discrete hydrogen bonds for recognition in the major groove and 8 discrete hydrogen bonds for recognition in the minor groove. A polyamide-oligonucleotide containing the mismatched polyamide H₂N- γ -PyPyIm shows no cooperative binding (Figure 3.2).

Results and Discussion

Synthesis of polyamide-oligonucleotides. Polyamide-oligonucleotides were prepared by automated synthesis of the amino-modified oligonucleotide, followed by manual stepwise Boc-chemistry solid phase synthesis of the polyamide.^{14,15} An oligonucleotide was assembled at 10 μ mol scale on an automated DNA synthesizer using standard DNA cycles, a 5'-MMT-C6-amino modifier was attached using an extended 10 min. synthesis cycle, and the synthesis removed from the synthesizer.

The MMT group was removed from the modified oligonucleotide by manual treatment with 3% TCA in dichloromethane until no yellow color was observed in the wash. The controlled pore glass support was then removed from the synthesis cartridge and transferred to a standard peptide

synthesis reaction vessel. The oligonucleotide was reacted with a 0.2 M solution of Boc-Py-OBt in DMF/0.4M DIEA for 45 min. (Figure 3.3). The reaction was determined complete by the quantitative ninhydrin test. A distinct blue color was observed for the oligo amine, consistent with a 0.05mmol/gram loading, and a lack of a blue color after reaction for 1 h. The Boc group was removed with 65% TFA/DCM/0.5M PhSH for 20 min., and a second pyrrole coupled to form the aromatic carboxamide. The Boc group is again removed with TFA, and Boc- γ -Im-OH incorporated with HBTU, DIEA. The polyamide-oligonucleotide is then simultaneously deprotected and cleaved from the resin by treatment with 0.1 M NaOH at 55 °C for 15 hours and purified by reverse phase HPLC (Figure 3.3). The molecular composition of purified polyamide-oligonucleotides were verified by MALDI-TOF mass spectroscopy, and HPLC analysis of enzymatic phosphodiester hydrolysis products.

Energetics by quantitative DNase I footprinting.¹⁶ A DNA fragment was constructed containing the symmetrical 31 base pair target site 5'-TTTCCTTTTAAATGTCATTAAAAAAGGAAA-3', composed of two 11 base pair triple helix sites separated by a 9 base pair region containing the 5 base pair polyamide binding site. Quantitative DNase I footprint titration experiments on the 268 base pair restriction fragment afford the binding affinities of the designed polyamide-oligonucleotides **1** and **2**.

The unmodified oligonucleotide, 5'-d(TTTTTT^{Me}C^{Me}CTTT)-3', binds each triple helix site with an equilibrium association constant of $K_a = 1.7 (\pm 0.1) \times 10^6 \text{ M}^{-1}$. The untethered polyamide ImPyPy was found to bind the 5'-TGTC A-3' target site with an equilibrium association constant, $K_a = 1.5 (\pm 0.3) \times 10^5 \text{ M}^{-1}$. Polyamide-oligonucleotide **1** binds the designated 31 base pair target site with an equilibrium association constant, $K_a = 1.7 \times 10^8 \text{ M}^{-1}$, representing a 100-fold increase in apparent affinity over the unmodified oligonucleotide (Figure 3.4 and Table 3.1). Polyamide-oligonucleotide **2**

Figure 3.4 Left: Storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel used to separate the products of DNase I digestion in quantitative footprint titration experiments with match pyrrole- imidazole polyamide-oligonucleotide **1**: Lane 1, A reaction; lane 2, G reaction; lane 3, intact DNA; lane 4 and 25, DNase I standard; lanes 5-24 contain 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, 1 μ M, 2 μ M, 5 μ M, 10 μ M, and 20 μ M polyamide-oligonucleotide **1**, respectively. All reactions contain 3'-³²P-end labeled *Eco* RI/*Pvu* II 268-bp restriction fragment from plasmid pJWS8 (15 kcpm) 10 mM Bis Tris•HCl (pH 7.0, 24 °C), 10 mM NaCl, and 250 μ M spermine. Right: The sequence of the restriction fragment in the region of the 31 bp target site and a model of the **1**₂•DNA complex are shown.

contains a mismatch in the polyamide moiety and is found to bind without positive cooperativity, demonstrating that the observed cooperative interaction is mediated by sequence specific recognition in the minor groove. The 2.7 kcal/mol binding enhancement reported here is achieved with an unconstrained 12 carbon linker domain for crossing from the major to the minor groove. The enhanced affinity is only consistent with formation of the designed homodimeric complex.

Table 3.1 Equilibrium association constants (M^{-1})^{a,b}

ligand	K_a (M^{-1})
1	$1.7 (\pm 0.9) \times 10^8$
2	$2.5 (\pm 0.3) \times 10^6$
5'-d(TTTTTT ^{Me} C ^{Me} CTTT)-3'	$1.7 (\pm 0.1) \times 10^6$

^aExperiments performed at 22 °C in the presence of 10 mM NaCl, 10 mM Bis Tris • HCl (pH 7.0), and 250 μ M spermine. ^b Values reported are the mean values obtained from three or more DNase I footprint titration experiments. The standard deviation for each data set is indicated in parentheses.

Linker distance dependence. Molecular modeling suggested that a 12 carbon linker is optimal to cross the phosphate backbone and connect the minor and major groove recognition domains. Two additional polyamide-oligonucleotides were synthesized, **3** and **4**, which contain a 2 or 3 atom change in linker length, respectively (Figure 3.5) An incremental increase or decrease in the linker domain results in a 10-fold reduction in affinity in both cases (Table 3.2). The dramatic loss in affinity produced by a small atomic

change demonstrates that the observed cooperativity is mediated by covalent linkage of the two recognition domains.

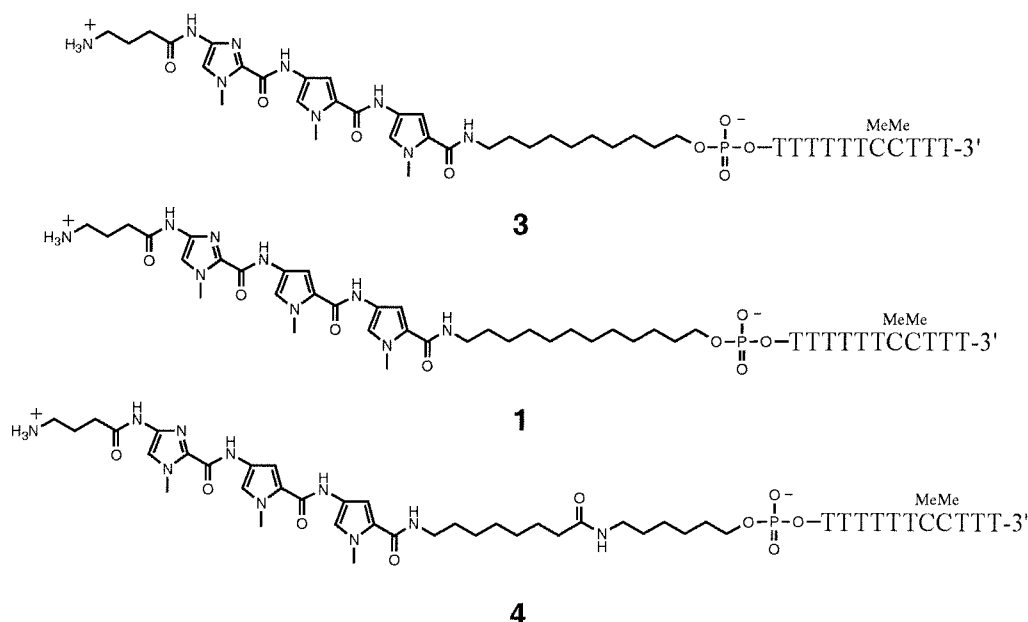


Figure 3.5 Pyrrole-imidazole polyamide-oligonucleotides containing a 10, 12, and 15 atom linker domain, respectively.

ligand	K_a (M^{-1})	atoms in linker
1	$1.7 (\pm 0.9) \times 10^8$	12
3	$1.8 (\pm 0.9) \times 10^7$	10
4	$1.6 (\pm 0.8) \times 10^7$	15

^aExperiments performed at 22 °C in the presence of 10 mM NaCl, 10 mM Bis Tris•HCl (pH 7.0), and 250 μM spermine. ^b Values reported are the mean values obtained from three or more DNase I footprint titration experiments. The standard deviation for each data set is indicated in parentheses.

It is interesting to compare the energetics of this artificial cooperative ligand-DNA complex with certain DNA-binding proteins which have discrete

dimerization domains.¹⁷ For example, the λ phage repressor can associate as a dimer at adjacent DNA binding sites with a ~ 2 kcal/mol cooperative interaction energy.^{10,17} By mimicking the complex behavior of such DNA-binding proteins, cooperative polyamide oligonucleotides provide a new model for the design of synthetic molecules for control of gene expression.¹¹

Experimental Section.

Polyamide-oligonucleotides **1** and **2** were synthesized by Eldon Baird as described.¹⁴

Materials. 6-(4-Monomethoxytritylamino)propyl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite, 5'-Amino-Modifier C6, 12-(4-Monomethoxytritylamino)propyl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite, 5'-Amino-Modifier C12, dT CE (2-cyanoethyl) phosphoramidite, 0.45M sublimed tetrazole in acetonitrile, THF/Lutidine/Ac₂O (8:1:1), 10% MeIm in THF, 0.1M I₂ in THF/Pyridine/H₂O, 3% TCA/DCM, 5-Methylcytidine CE (2-cyanoethyl) phosphoramidite, and bulk 500Å dT-Icaa-CPG were purchased from Glen Research. All 10 µmole synthesis columns were packed manually from bulk support. 1,2,3-Benzotriazol-1-yl-4-[(*tert*-Butyloxy)carbonyl]-amino]-1-methylpyrrole-2-carboxylate and 1,2,3-Benzotriazol-1-yl-4-[(*tert*-Butyloxy)carbonyl]-amino]-1-methylimidazole-2-carboxylate were prepared as previously described.¹⁴ Trifluoroacetic acid (TFA) was purchased from Halocarbon. N,N-diisopropylethylamine (DIEA), N,N-dimethylformamide (DMF), 0.0002M potassium cyanide/pyridine, and Acetic Anhydride (Ac₂O) were purchased from Applied Biosystems, and Dicyclohexylcarbodiimide (DCC) and Hydroxybenzotriazole (HOBt) were from Peptides International. Boc-GABA was from NOVABiochem, dichloromethane (DCM), reagent grade from EM. Thiophenol was purchased from Aldrich. A shaker for manual solid phase

synthesis was obtained from Milligen. Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. and are used for filtration of CPG. Screw-cap glass peptide synthesis reaction vessels (5 ml) with a #2 scintered glass frit were made at the Caltech glass shop as described by Kent.¹⁴ UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed either on a HP 1090 M analytical HPLC or a Beckman Gold system using a RAINEN C18, Microsorb MV, 5 μ m, 300 x 4.6 mm reversed phase column in 100 mM Ammonium Acetate, pH 4.9 with acetonitrile as eluent and a flow rate of 1.0 ml/min, gradient elution 1.0% acetonitrile/min. Polyamide-oligonucleotides were purified by reverse phase HPLC using a linear gradient from 0 to 40% acetonitrile in 55 min., 100 mM triethylammonium acetate, pH 7.0.

H₂N- γ -ImPyPy-C₁₀-TTTTTT^{Me}C^{Me}CTTT-3' 3. The oligonucleotide DMT-TTTTTT^{Me}C^{Me}CTTT-CPG was prepared on an Applied Biosystems Model 394B DNA synthesizer using a manually prepared 10 μ mole synthesis column and a standard 10 μ mole synthesis cycle. C₁₀-AminoModifier-MMT (100 μ mole) was dissolved in 1,100 μ l of anhydrous acetonitrile, vortexed vigorously, and placed on the synthesizer. The amino modifier was added by machine synthesis using a modified 10 μ mole synthesis cycle with an extended 10 min. coupling time and the MMT group left on. The column is manually washed

with 50 ml of 3% TCA/dichloromethane for a period of 12 min., until yellow color is no longer observed in the wash. The column is washed with 15 ml dichloromethane and dried *in vacuo*. The CPG is transferred to a 5 ml glass manual peptide reaction vessel and washed with DMF (30 sec.). A sample is taken for ninhydrin test and an absorbance consistent with 50 μ mole/ gram substitution found. Boc-Pyrrole-OBt ester (70 mg, 200 μ mole) is dissolved in DMF (600 μ l) and DIEA (68 μ l) added. The coupling mixture is added to the reaction vessel and allowed to shake for 60 min. A sample is taken for ninhydrin test and no blue color observed, indicating at least 90% reaction. The resin is washed with DMF (30 sec.) followed by dichloromethane (30 sec.) and 65% TFA/DCM/0.5M PhSH (30 sec.). The resin is shaken in 65% TFA/DCM/0.5M PhSH for 20 min., drained, washed with DCM (30 sec.) followed by DMF 30 sec. A second equivalent of Boc-Pyrrole-OBt is added under identical conditions to the first, and the reaction shaken for 1 hour, washed DMF (30 sec.), DCM (30 sec.) and treated with 65% TFA/DCM/0.5M PhSH as described for the first deprotection. The resin was treated with a solution of Boc- γ -Im-OBt (80 mg) DIEA (68 μ l) and DMF (600 μ l), and shaken for 1 hour. The CPG is washed, DMF (30 sec.), DCM (30 sec.), and 65% TFA/DCM/0.5M PhSH. The CPG is washed with DCM (30 sec.), DMF (30 sec.), DCM (30 sec.), and dried *in vacuo*. The support was treated with 0.1 M NaOH (2 ml) and heated (55 °C) for 12 h. The reaction mixture was then filtered to remove CPG, 0.1 M triethylammonium acetate (pH = 7.0) added (6 ml) and the resulting solution purified by reversed phase HPLC. H₂N- γ -ImPyPy-C₁₀-

TTTTTT^{Me}C^{Me}CTTT-3' is recovered upon lyophilization of the appropriate fractions: UV 260 (118,000), 304 (33,000) MALDI-TOF MS, calc. M-1 3969, found 3969.

H₂N-γ-ImPyPy-C₇-(CONH)-C₆-TTTTTT^{Me}C^{Me}CTTT-3' 4. Polyamide-oligonucleotide **4** was prepared as described for **3**. The support was treated with 0.1 M NaOH (2 ml) and heated (55 °C) for 12 h. The reaction mixture was then filtered to remove CPG, 0.1 M triethylammonium acetate (pH = 7.0) added (6 ml) and the resulting solution purified by reversed phase HPLC. H₂N-γ-ImPyPy-C₇-(CONH)-C₆-TTTTTT^{Me}C^{Me}CTTT-3' is recovered upon lyophilization of the appropriate fractions: UV 260 (118,000), 304 (33,000) MALDI-TOF MS, calc. M-1 4054, found 4055.

Preparation of ³²P-labeled DNA. The plasmid pJWS8 was constructed by ligation of an insert, 5'-d(GATCCTTTCCTTTTTTAATGACATTAAAAAAGGAAATTA)-3' and 5'-d(AGCTTAATTTTCCTTTTTTAATGTCATTAAAAAAGGAAAG)-3', into pUC19 previously cleaved with *Bam* HI and *Hind* III. The 3'-³²P end-labeled *Eco* RI/ *Pvu* II fragment was prepared by digesting the plasmid pJWS8 with *Eco* RI and *Pvu* II and simultaneously filling in using Sequenase, [α-³²P]-deoxyadenosine-5'-triphosphate, and [α-³²P]-thymidine-5'-triphosphate. The 268 bp fragment was isolated by nondenaturing gel electrophoresis. A and G

sequencing were carried out as described.^{18,19} Standard methods were used for all DNA manipulations.²⁰

Quantitative DNase I footprint titrations.¹⁶ The quantitative footprint titration experiments were executed in a total volume of 40 μ L with a final concentration of each species as indicated. The ligands were added to solutions of radiolabeled restriction fragment (15 000 cpm), NaCl (10 mM), Bis Tris•HCl (10 mM, pH 7.0), and spermine (250 μ M), incubated for 24 hours at 22 °C. Footprinting reactions were initiated by addition of 4 μ L stock solution of DNase I (0.8 units/ml) containing $MgCl_2$ (50 mM), $CaCl_2$ (50 mM), Bis Tris•HCl (10 mM), and glycerol (5%) and allowed to proceed for 6 minutes at 22 °C. The reactions were quenched by addition of 12.8 μ L of a solution made up of 205 μ L NaOAc (3M), 160 μ L glycogen, and 160 μ L NaEDTA (50 mM) and ethanol precipitated. The reactions were resuspended in 80% formamide loading buffer, denatured (85 °C, 10 min.), and electrophoresed on an 8% polyacrylamide denaturing gel at 2000 V for 1 hour.

Quantitation and data analysis. Data from the footprint titration gels were obtained using a Molecular Dynamics 400S PhosphorImager followed by quantitation using ImageQuant software (Molecular Dynamics). Background-corrected volume integration of rectangles encompassing the footprint sites and a reference site at which DNase I reactivity was invariant across the

titration generated values for the site intensities (I_{site}) and the reference intensity (I_{ref}). The apparent fractional occupancy (θ_{app}) of the sites were calculated using the equation:

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

where I_{site}° and I_{ref}° are the site and reference intensities, respectively, from a control lane to which no polyamide was added. The ($[L]_{\text{tot}}$, θ_{app}) data points were fit to a general Hill equation (eq 2) by minimizing the difference between θ_{app} and θ_{fit} :

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

where $[L]_{\text{tot}}$ is the total ligand concentration, K_a is the apparent first-order association constant, and θ_{min} and θ_{max} are the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. The data were fit using a nonlinear least-squares fitting procedure with K_a , θ_{max} , and θ_{min} as the adjustable parameters. For polyamide ImPyPy binding isotherms for the 5'-TGTCA-3' target site were adequately fit by Langmuir isotherms (eq 2, $n=2$), consistent with formation of a 2:1 polyamide-DNA complexes. The mismatched polyamide- oligonucleotide **2** only footprinted the two 11 bp major groove binding sites, 5'-AAAAAAGGAAA-3', and was fit by Langmuir isotherms (eq 2, $n=1$), consistent with formation of a local triple helix in the major groove and no binding in the minor groove. For the matched polyamide oligonucleotide **1**, and control conjugates **3** and **4**,

shallower binding isotherms ($n=1$) were observed at the 31 base pair target site. We believe that the shape of these isotherms is consistent with binding being comprised of a series of binding events, and emphasize that treatment of the data in this manner does not represent an attempt to model a binding mechanism. Rather, we have chosen to compare values of the apparent first-order association constant, a value that represents the concentration of ligand at which a site is half-saturated. Three sets of data were used in determining each association constant. The method for determining association constants used here involves the assumption that $[L]_{\text{tot}} \approx [L]_{\text{free}}$, where $[L]_{\text{free}}$ is the concentration of polyamide- oligonucleotide free in solution (unbound).

References

1. (a) Wade, W.S.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1992**, *114*, 8783; (b) Wade, W.S.; Mrksich, M.; Dervan, P.B. *Biochemistry* **1993**, *32*, 11385; (c) Mrksich, M.; Wade, W.S.; Dwyer, T.J.; Geierstanger, B.H.; Wemmer, D.H.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586.
2. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 2572; (b) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 3325; (c) Geierstanger, B.H.; Dwyer, T.J.; Bathini, Y.; Lown, J.W.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 4474; (d) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646; (e) Geierstanger, B.H.; Jacobsen, J.P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Biochemistry* **1994**, *33*, 3055.
3. (a) Pelton, J.G.; Wemmer, D.E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723; (b) Pelton, J.G.; Wemmer, D.E. *J. Am. Chem. Soc.* **1990**, *112*, 1393; (c) Chen, X.; Ramakrishnan, B.; Rao, S.T.; Sundaralingham, M. *Nature Struct. Biol.* **1994**, *1*, 169.
4. (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) de Claire, R. P. L.; Geierstanger B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909; (g) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756; (h) White, S.; Baird, E. E.; Dervan, P. B. *Chem & Biol.* **1997**, *4*, 569.
5. (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; (d) Turner, J. M. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219.
6. (a) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan. P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160; (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1997**, *3*, 1600; (c) Trauger, J.W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534.
7. De Clairac, R.P.L.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1997**, *119*, 7906.

8. (a) Moser, H.E.; Dervan, P.B. *Science* **1987**, 238, 645–650; (b) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.L.; Thoung, N.T.; Lhomme, J.; Helene, C. *Nucleic Acids Res.* **1987**, 15, 7749; (c) Strobel, S.A.; Doucetestamm, L.A.; Riba, L.; Housman, D.E.; Dervan, P.B. *Science* **1991**, 254, 1639–1642; (d) Thuong, N.T.; Helene, C. *Angew. Chem. Int. Ed. Engl.* **1993**, 32, 666–690.
9. (a) Maher, J. L.; Dervan, P. B.; Wold, B. *Biochemistry* **1992**, 31, 71; (b) Duvalentin, G.; Thuong, N.T.; Helene, C.; *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 504.
10. (a) Hogan, M.; Dattagupta, N.; Crothers, D.M. *Nature* **1979**, 278, 521; (b) Hill, T.L. *Cooperativity Theory in Biochemistry: Steady State and Equilibrium Systems*; Springer- Verlag: New York, N.Y. 1985; (c) Ptashne, M. *Nature* **1988**, 355, 683; (d) Ptashne, M. *A Gentic Switch*; Blackwell Scientific Publications and Cell Press: Cambridge, MA, **1992**.
11. (a) Strobel, S.A.; Dervan, P.B. *J. Am. Chem. Soc.* **1989**, 111, 7286; (b) Colocci, N.; Distefano, M.D.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, 115, 4468; (c) Colocci, N.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, 116, 785; (d) *J. Am. Chem. Soc.* **1995**, 117, 4781-4787.
12. (a) Moser, H.E.; Dervan, P.B. *Science* **1987**, 238, 645; (b) For a review, see: Thuong, N.T.; Helene, C. *Angew. Chem., Int. Ed. Engl.* **1993**, 32, 666 and references cited therein.
13. (a) Distefano, M.D., Shin, J.A., Dervan, P.B. *J. Am. Chem. Soc.* **1991**, 113, 5901; (b) Distefano, M.D.; Dervan, P.B. *J. Am. Chem. Soc.* **1992**, 114, 11006; (c) Distefano, M.D.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, 90, 1179.
14. (a) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, 118, 6141; (b) Szewczyk, Baird & Dervan *J. Am. Chem. Soc.* **1996**, 118, 6778-6779.
15. For other synthetic methodology see (a) Sinyakov, A.N.; Lokhov, S.G.; Kutuyavin, I.V.; Gamper, H.B.; Meyer, R.B. *J. Am. Chem. Soc.* **1995**, 117, 4995; (b) Szewczyk, J.W.; Baird, E.E.; Dervan P.B. *Angew. Chemie. Int. Ed.* **1996**, 35, 1487.
16. (a) Breslauer, K.J.; Remeta, D.P.; Chou, W.-Y.; Ferrante, R.; Curry, J.; Zaunczkowski, D.; Snyder, J.G.; Marky, L.A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, 84, 8922; (b) Marky, L.A.; Breslauer, K.J.; *Proc. Natl. Acad. Sci. U.S.A.* **1987**, 84, 4359; (c) Marky, L.A.; Kupke, K.J. *Biochemistry* **1989**, 28, 9982.

17. (a) Johnson, A.D.; Meyer, B.J.; Ptashne, M. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, 76, 5061; (b) Ackers, G.K.; Johnson, A.D.; Shea, M.A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, 79, 1129.
18. Maxam, A.M.; Gilbert, W.S. *Methods Enzymol.* **1980**, 65, 499-560.
19. Iverson, B.L.; Dervan, P.B. *Methods Enzymol.* **1987**, 15, 7823-7830.
20. Sambrook, J.; Fritsch, E.F.; Maniatis, T. (1989). *Molecular Cloning*. (2nd edn). Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

Chapter 4

Cooperative Binding of Extended Pyrrole-Imidazole Polyamide-Oligonucleotides to the Minor-Major Grooves of DNA

***Abstract:** Covalent attachment of pyrrole-imidazole polyamides to oligonucleotides forms a class of cooperative ligands which simultaneously recognize the minor and major grooves of DNA. Two extended pyrrole-imidazole polyamide-oligonucleotides specifically recognize 31 contiguous base pairs of double helical DNA with subnanomolar affinity. Quantitative footprint titrations demonstrate that the match polyamide-oligonucleotide binds the target site with over 3.2 kcal mol⁻¹ ($K_a \geq 1.0 \times 10^{10} \text{ M}^{-1}$) in cooperative enhancement. The unlinked polyamide and oligonucleotide bind with > 300 or > 5000-fold lower affinity, respectively. Introduction of a new linker design has facilitated the development of cooperative polyamide-oligonucleotides which specifically recognize all bases within the match site.*

Pyrrole-imidazole polyamide-oligonucleotides are synthetic ligands which combine the two most powerful chemical approaches for DNA recognition to simultaneously recognize both the minor and major grooves.^{1,2} Specific recognition occurs by formation of a side-by-side pyrrole-imidazole polyamide-DNA complex in the minor groove³⁻⁵ and oligonucleotide-directed triple-helix formation in the major groove.^{6,7}

Polyamides containing the three aromatic amino acids 3-hydroxypyrrole (Hp), pyrrole (Py) and imidazole (Im) are cell permeable small molecules⁸ that bind predetermined DNA sequences in the minor groove with affinities and specificities comparable to those of naturally occurring DNA-binding proteins.⁹⁻¹¹ 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.^{3-5,12-16} 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.^{3-5,13} A Py/Py pair binds both A•T and T•A in preference to G•C/C•G.¹²⁻¹⁹ The discrimination of T•A from A•T using Hp/Py pairs completes the four base pair code.^{20,21} The generality of this approach has been demonstrated by targeting a wide variety of DNA sequences²²⁻²⁷ and is directly supported by several crystallographic^{21,28} and NMR structure studies.^{14,29-31}

In one triple helical motif, a pyrimidine oligonucleotide binds in the major groove parallel to the purine Watson-Crick strand.^{6,7,32,33} Recognition occurs by formation of specific hydrogen bonds between the bases in the third strand and a purine rich strand of the duplex. Specificity is derived from

thymine (T) recognition of adenine-thymine base pairs (T•AT base triplets) and protonated methyl cytosine (^{Me}C⁺) recognition of guanine-cytosine base pairs (^{Me}C⁺+G•C).^{6,7,32,33} However, triplex forming oligonucleotides suffer from poor cellular uptake and possess a targetable sequence repertoire restricted to tracts of purines, limiting the utility of this strategy.

Cellular transporter peptides have been shown to efficiently promote the cellular uptake of both oligonucleotides³⁴ and peptide nucleic acids (PNAs).³⁵ Unfortunately, the consequences of the addition of a transporter peptide on DNA-binding properties of the ligand have not been reported. A method combining the functions of enhanced permeation and recognition in a single domain would be an attractive addition to current approaches. Recently, an eight-ring hairpin pyrrole-imidazole polyamide binding to 6 base pairs of DNA was shown to inhibit transcription of a specific gene in cell culture, demonstrating that polyamides have the ability to transverse the cellular membrane.^{8,36} An alternative strategy to promote the cellular uptake of oligonucleotides may be the covalent attachment of cell-permeable pyrrole-imidazole polyamides to triplex-forming oligonucleotides.

Cooperative polyamide-oligonucleotide:DNA complexes.

Previously, two three-ring pyrrole-imidazole polyamide-oligonucleotides were shown to cooperatively bind to 27 noncontiguous base pairs of DNA by formation of a 2:1 polyamide-DNA complex in the minor groove and two local triple helices in the major groove (Figure 4.1a).¹ According to the two

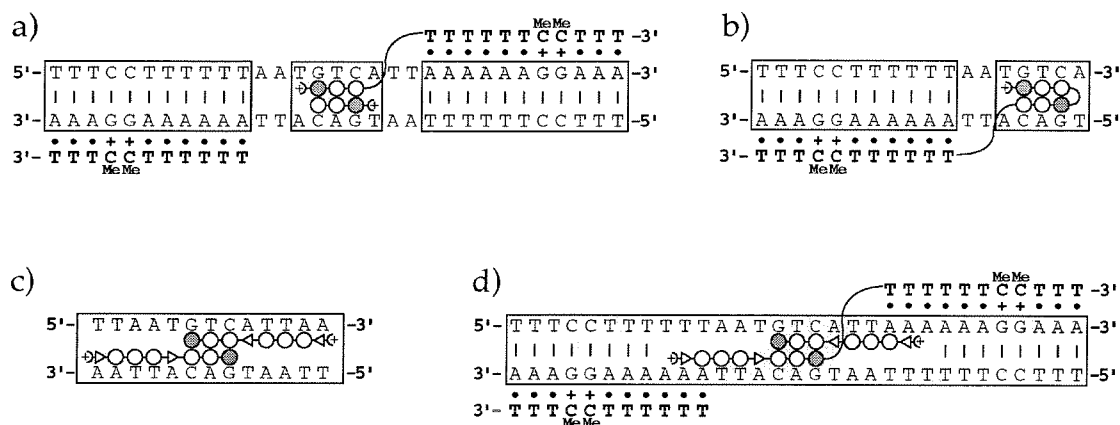


Figure 4.1 Models of specific DNA complexes: (a) Two polyamide-oligonucleotides cooperatively binding 27 noncontiguous base pairs.¹ (b) Recognition of 16 noncontiguous bp by a hairpin pyrrole-imidazole polyamide-oligonucleotide simultaneously binding the minor and major grooves.² (c) A six-ring polyamide forms a discrete 2:1 extended complex in the DNA minor groove binding a 13 base target.³⁷ (d) An extended pyrrole-imidazole polyamide oligonucleotide complex binding to the major and minor grooves specifically recognizes 31 contiguous base pairs. Shaded and nonshaded circles represent imidazole and pyrrole rings, respectively. Nonshaded triangles represent glycine residues. Linkers are represented as curved lines. Shaded boxes surround base pairs specifically recognized.

binding models, the homodimeric complex afforded 10 discrete hydrogen bonds for recognition of 5 base pairs in the minor groove and 44 discrete hydrogen bonds for recognition of 22 bases in the major groove. Alternatively, a hairpin polyamide-oligonucleotide binds an 18 bp target with subnanomolar affinity ($K_a > 2 \times 10^{10} \text{ M}^{-1}$) by simultaneous recognition of the major and minor grooves (Figure 4.1b).² Although the putative cooperative

recognition complex afforded twice as many hydrogen bonds for major groove recognition, a 100-fold decrease in affinity was observed, $K_a = 2 \times 10^8$, potentially due to addition of a second linker domain. It remained unclear if cooperative polyamide-oligonucleotides could be designed with the desired affinity due to increased rotational entropy introduced by a second linker.

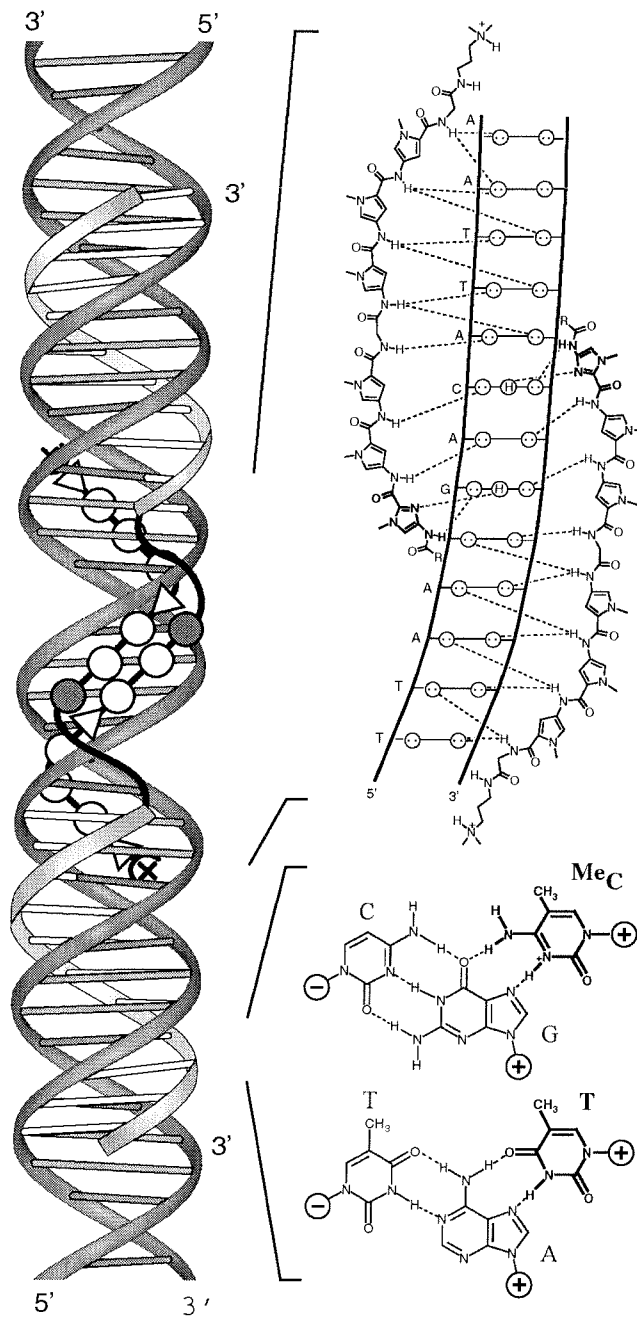
Extended polyamide binding motifs.

The six-ring polyamide AcImPyPy-G-PyPyPy **1** cooperatively binds the 13 bp site, 5'-TTAATGTCATTAA-3' as a 2:1 antiparallel extended complex (Figure 4.1c).³⁷ The ImPyPy polyamide subunits bind the central 5'-TGTC-3' sequence as a 2:1 polyamide:DNA complex, while the PyPyPy moiety binds the A,T flanking sequences as in the 1:1 complexes of distamycin (Figure 4.2).^{17,27} The ability of extended polyamides to target the AT/TA sequences composing the triplex suggest the two motifs be combined to allow overlap of the minor and major groove recognition domains.

Extended pyrrole-imidazole polyamide-oligonucleotides.

Design of a new linker topology allows incorporation of the extended polyamide motif into polyamide-oligonucleotides to specifically recognize all bases within the target site (Figure 4.1d). Polyamides recognize DNA with a 5' to 3', N to C binding orientation,^{3-5,13,15,16,21,23,28} while pyrimidine oligodeoxyribonucleotides recognize the major groove of homopurine tracts parallel to the purine strand binding 3' to 5' (Figure 4.2).^{3,4,33} These rules

Figure 4.2 Right: Two-dimensional models depicting the molecular interactions responsible for sequence specific recognition. Upper right: (a) Binding model for the complex formed between the DNA and the extended polyamide AcImPyPy-G-PyPyPy-G-Dp. Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen on guanine. Putative hydrogen bonds are illustrated by dashed lines. Lower right: (b) The T•AT and ^{Me}C+GC base triplets formed by Hoogsteen hydrogen bonding of T to a Watson-Crick AT base pair and protonated ^{Me}C to a Watson-Crick GC base pair. The bases in the third strand are bold. Left: A ribbon model depicting two extended pyrrole- imidazole polyamide-oligonucleotides cooperatively binding the major and minor grooves of double helical DNA. The shaded and nonshaded circles represent *N*-methylimidazole and *N*-methylpyrrole rings, respectively. The nonshaded triangles represent glycine. Linkers are represented by curved black lines. The third strand is shaded gray with white bars depicting specific Hoogsteen hydrogen bonds to the purine Watson-Crick bases.



require the polyamide moiety be directed to a defined strand when linked to the 5'-terminus of an oligonucleotide. Modeling indicated that linkage 5' to N constrains the rotational freedom of the aliphatic chain to a small set of defined geometries to potentially increase the binding energy afforded by simultaneous recognition. The new linker topology connected 5' to N for the “extended” polyamide-oligonucleotides (Figure 4.1d) requires the linker to cross over the minor groove directing the polyamide moiety to bind the strand opposite the one recognized by the oligonucleotide (Figure 4.2).

We report here the synthesis of an extended, six ring polyamide, ImPyPy-G-PyPyPy, covalently attached to the 5'-amino modified oligonucleotide, 5'(NH)-d(TTTTTT^{Me}C^{Me}CTTT)-3', by an aliphatic amino acid linker. The extended polyamide-oligonucleotide cooperatively binds 31 contiguous bases by formation of an extended polyamide dimer in the minor groove which overlaps two local triple helices formed in the major groove. As controls unlinked polyamide **1**, AcImPyPy-G-PyPyPy-G-Dp, and the oligonucleotide **2**, 5'-d(TTTTTT^{Me}C^{Me}CTTT)-3' were also synthesized. A convergent solid phase approach provided both the match **3** and mismatch polyamide-oligonucleotide **4**, which contains a mismatched oligonucleotide, 5'-d(TTTTTTT^{Me}CT^{Me}CT)-3' (Figure 4.3). To characterize the cooperative system, equilibrium association constants were determined by quantitative DNase I footprint titration experiments³⁸⁻⁴⁰ carried out on a 3'-³²P end labeled 268 bp restriction fragment containing the 31 base pair target site (5'-TTTCCTTTTTTAATGTCATTAAAAAAGGAAA-3') composed of a 13 bp

polyamide binding site (5'-TTAATGTCATTAA-3') overlapping by two 11 bp oligonucleotide binding sites (Figure 4.1d).

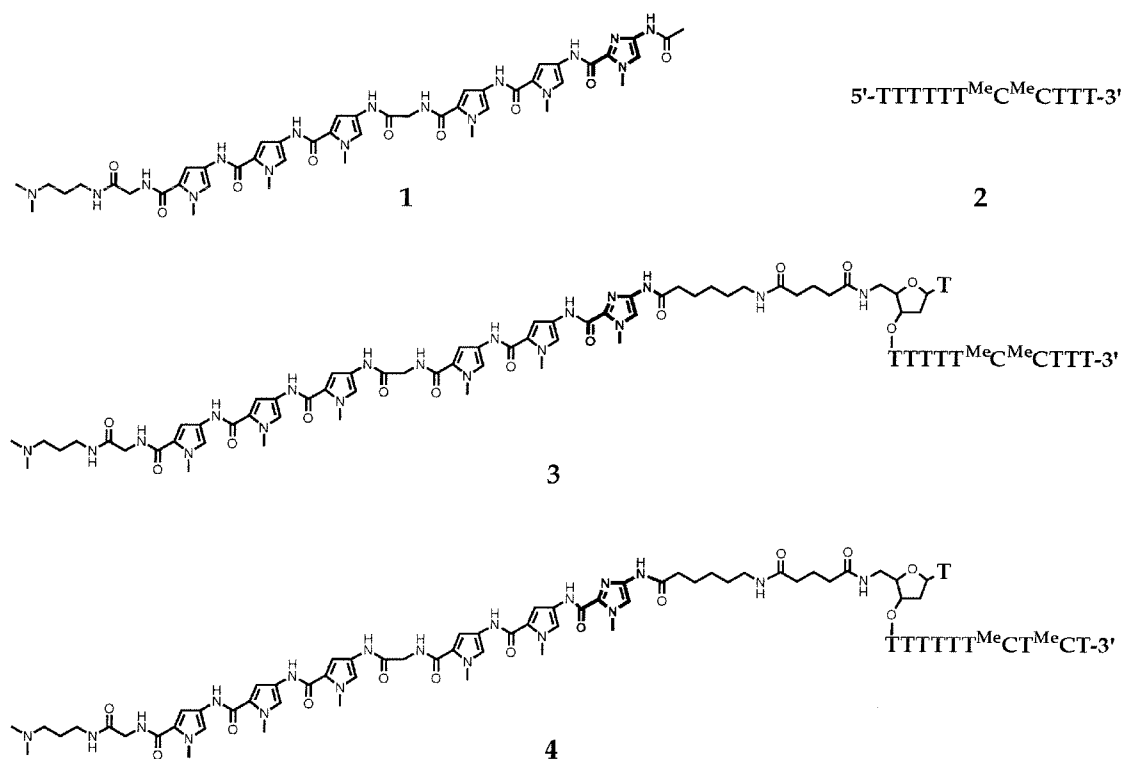


Figure 4.3 Structure of the unlinked pyrrole- imidazole polyamide **1**, the unmodified oligonucleotide **2**, the match pyrrole-imidazole polyamide-oligonucleotide **3**, and pyrrole-imidazole polyamide-oligonucleotide **4** containing a mismatch oligonucleotide.

Results and Discussion

Solid phase synthesis of pyrrole-imidazole polyamides.

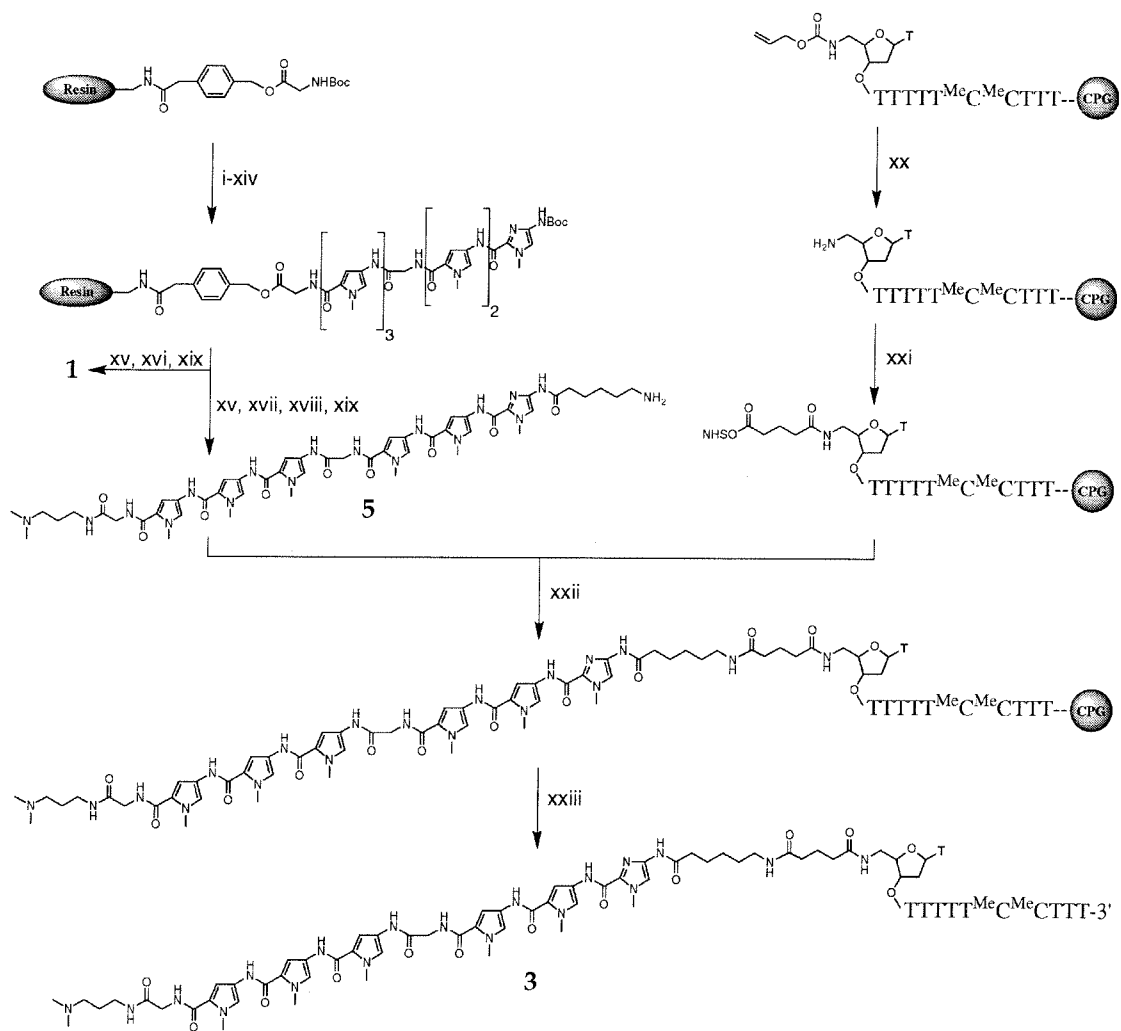
Two polyamide resins R-ImPyPy-G-PyPyPy-G-Pam-resin (R = COCH₃, **1**; or CO(CH₂)₅NH₂, **5**) were synthesized in 16 and 17 steps, respectively, from Boc-Gly-Pam-resin (1 g of resin 0.2 mmol/g of substitution) using previously

described Boc-chemistry machine assisted protocols (Figure 4.4).⁴¹ A single-step aminolysis of the resin ester linkage was used to cleave the polyamide from the solid support to provide polyamides **1** and **5**. Resin cleavage products were purified by reversed phase HPLC to provide the desired polyamides R-ImPyPy-G-PyPyPy-G-Dp **1** and **5** (R = COCH₃ or CO(CH₂)₅NH₂, respectively).

Synthesis of extended polyamide-oligonucleotides. All polyamides-oligonucleotides were prepared in high purity using a convergent solid phase synthetic methodology (Figure 4.4).^{2,41-45} An 11-mer oligonucleotide containing a 5'-alloc-protected amino group suitable for post-synthetic modification was prepared using standard β -cyanoethyl phosphoramidite chemistry.^{42,43,46} The alloc group was removed on support by reaction with a palladium catalyst⁴⁷ followed by reaction with disuccinimidyl glutarate to provide a support bound NHS ester.²

Reaction of the support with polyamide **5** afforded CPG-TTT^{Me}C^{Me}CTTTTTT-5'-DSG-(CH₂)₅-CONHImPyPy-G-PyPyPy-G-Dp. A one step cleavage and deprotection was achieved by treatment with 0.1 N NaOH (55 °C, 16 h). A single reverse phase HPLC purification (TEAA, pH 7.0) provided polyamide-oligonucleotide **3** in >98% purity as determined by a combination of MALDI-TOF mass spectrometry, reverse-phase analytical HPLC, and HPLC analysis of enzymatic hydrolysis reactions. Similarly, the control conjugate **4**,

Figure 4.4 Solid phase synthetic scheme for polyamide- oligonucleotide **3** starting from commercially available Boc-Gly-Pam-resin: (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; (vii) 80% TFA/DCM, 0.4 M PhSH; (viii) Boc-Gly-OH, 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) BocImOH, HBTU, DIEA; (xv) 80% TFA/DCM, 0.4 M PhSH; (xvi) acetic anhydride, DIEA, DMF; (xvii) Boc-6-aminocaproic acid, HBTU, DIEA; (xviii) 80% TFA/DCM, 0.4 M PhSH; (xix) *N,N*-dimethylaminopropylamine, 55 °C; (xx) $\text{Pd}_2(\text{dba})_2 \cdot \text{CHCl}_3$, PPh_3 ; (xxi) disuccinimidyl glutarate, DIEA, DMF; (xxii) **5**, DIEA, DMF; (xxiii) 0.1 M NaOH, 55 °C.



containing an mismatched oligonucleotide, 5'-TTTTTTT^{Me}CT^{Me}CT-3', was synthesized using the methodology described for 3.

Energetics by quantitative DNase I footprinting.

Quantitative DNase I footprint titration experiments³⁸⁻⁴⁰ (10 mM Bis Tris•HCl, 10 mM NaCl, 250 μ M spermine, pH 7.0, 22 °C) were performed to determine the apparent equilibrium association constant (K_a) for each of the ligands **1-4** (Figures 4.5 and 4.6). A DNA fragment was constructed containing the 31 bp pair target site (5'-TTTCCTTTTTTAATGTCATTAAAAAAGGAAA-3') composed of a 13 bp polyamide binding site overlapping two 11 bp triple helix sites. Polyamide **1** specifically binds the 13 base target 5'-TTTCCTTTTTTTAATGTCATTAAAAAAGGAAA-3' with an equilibrium association constant, $K_a = 3.0 (\pm 0.8) \times 10^7 \text{ M}^{-1}$. Oligonucleotide **2** binds each 11 bp triple-helix site with an equilibrium association constant of $K_a = 1.7 (\pm 0.1) \times 10^6 \text{ M}^{-1}$. The match polyamide-oligonucleotide **3** targets the 31 bp site 5'-TTTCCTTTTTTTAATGTCATTAAAAAAGGAAA-3' with an apparent first-order equilibrium association constant, $K_a \geq 1 (\pm 0.2) \times 10^{10} \text{ M}^{-1}$, representing at least a 300-fold increase in binding affinity relative to the unlinked subunits. The mismatched polyamide-oligonucleotide **4** binds the target site ($K_a = 6.6 (\pm 0.5) \times 10^6 \text{ M}^{-1}$) with 1500-fold reduced affinity relative to match conjugate **3** (Table 4.1).

Figure 4.5 Left: Storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel used to separate the products of DNase I digestion in quantitative footprint titration experiments with polyamide-oligonucleotide **3**: Lane 1: intact DNA, lanes 2-3: A and G sequencing lanes; lanes 4 and 22: DNase I digestion products obtained in the absence of ligand; lanes 5-21: DNase I digestion products obtained in the presence of 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 1.5 nM, 2.5 nM, 4.0 nM, 6.5 nM, 10 nM, 15 nM, 25 nM, 40 nM, 65 nM, and 100 nM polyamide- oligonucleotide **3**, respectively. All reactions contain 3'-³²P-end labeled *Eco* RI/*Pvu* II 268-bp restriction fragment from plasmide pJWS8 (15 kcpm) 10 mM Bis Tris•HCl (pH 7.0, 24 °C), 10 mM NaCl, and 250 μM spermine. Right: The sequence of the restriction fragment in the region of the 31 bp target site and a model of the **3**₂•DNA complex are shown.

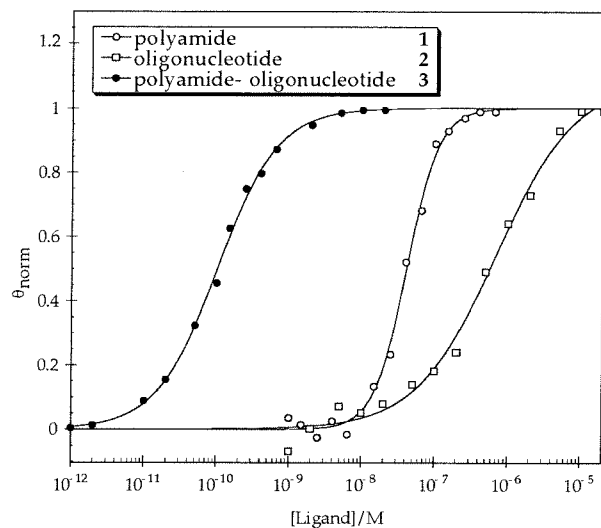


Figure 4.6 Data from quantitative DNase I footprinting experiments with unlinked polyamide **1** (white circles), unmodified oligonucleotide **2** (white squares), and the polyamide- oligonucleotide **3** (black circles) at their match sites. θ_{norm} points were obtained using storage phosphor autoradiography and processed as described in the Experimental Section. The solid curves are the best fit Langmuir binding titration isotherms obtained from a nonlinear least squares algorithm.

Table 4.1 Equilibrium association constants (M^{-1})^{a,b}

ligand	K_a (M^{-1})
1	$3.0 (\pm 0.8) \times 10^7$
2	$1.7 (\pm 0.1) \times 10^6$
3	$\geq 1 (\pm 0.2) \times 10^{10}$
4	$6.6 (\pm 0.5) \times 10^6$

^aExperiments performed at 22 °C in the presence of 10 mM NaCl, 10 mM Bis Tris•HCl (pH 7.0), and 250 μM spermine. ^b Values reported are the mean values obtained from three or more DNase I footprint titration experiments. The standard deviation for each data set is indicated in parentheses.

Extended polyamide-oligonucleotides bind at subnanomolar concentrations.

Two six-ring pyrrole-imidazole polyamide-oligonucleotides specifically recognize 31 contiguous bases of double-helical DNA by simultaneous formation of a 2:1 extended polyamide-DNA complex in the minor groove and triple-helix formation in the major groove. According to the two binding models the complex affords 26 discrete hydrogen bonds for recognition of 13 base pairs in the minor groove and 44 discrete hydrogen bonds for recognition of 22 bases in the major groove. For the previous cooperative polyamide-oligonucleotide linked 5' to C (Figure 4.1a), the polyamide and oligonucleotide domains recognize the same strand leaving 4 bases unrecognized. The extended polyamide-oligonucleotide linked 5' to N directs the simple aliphatic linker over the adjacent polyamide and the phosphodiester backbone to deliver the oligonucleotide to the alternate strand in the major groove (Figure 4.1d). Overlap of the two recognition domains allows the two PyPyPy polyamide-subunits to bind in the minor groove of the two local triple helices specifically recognizing all bases within the target site (Figure 4.2).

Polyamide-oligonucleotide **3** binds with subnanomolar affinity only at the 31 base target. Polyamide-oligonucleotide **4** contains a mismatched oligonucleotide and binds without positive cooperativity, demonstrating that the observed cooperative interaction is mediated by sequence-specific recognition of both the major and minor grooves. Local triple helix formation at a binding site flanking a 2:1 polyamide:DNA site was reported to

be energetically neutral⁴⁸ indicating that the enhanced affinity is only consistent with formation of the dimeric complex.

The 3.2 kcal mol⁻¹ binding enhancement reported here is achieved with an optimized linker design to incorporate the extended polyamide motif. Requiring the linker to cross both the phosphate backbone and the neighboring polyamide restricts the degrees of rotational freedom when bound potentially affording the increased affinity. Comparing the two designed cooperative systems (Figure 4.1c and 4.1d), the 0.5 kcal mol⁻¹ increase in cooperative interaction energy indicates that overlap of the major/minor groove recognition domains by 4 bases is an allowed binding arrangement.

Incorporation of the extended polyamide motif for recognition of the minor groove creates a new class of pyrrole-imidazole polyamide-oligonucleotides. Design of a new linker topology allows overlap of the minor and major groove recognition domains specifically targeting all bases within the binding site. The affinity increase (3.2 kcal mol⁻¹) provides the first example for the compatibility of polyamide binding when overlapping a local triple helix. The affinity and sequence selectivity of cooperative polyamide-oligonucleotide:DNA complexes suggests exploring the potential of this class of ligands to enhance the cell permeability of triplex-forming oligonucleotides. Extended pyrrole-imidazole polyamide-oligonucleotides combine a versatile solid phase synthetic methodology with a broad targetable sequence repertoire to provide a paradigm for the design of artificial

molecules which sequence-specifically recognize both grooves of duplex DNA.

Experimental Section.

Polyamide synthesis and conjugation reactions for polyamide-oligonucleotides **3** and **4** were performed by Eldon Baird.

Materials. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed either on a HP 1090 M analytical HPLC or a Beckman Gold system using a Rainen C18, Microsorb MV, 5 μ m, 300 x 4.6 mm reversed phase column in 0.1% (wt/v) TFA (for polyamides) or 100 mM triethylammonium acetate (pH = 7.0) (for polyamide-oligonucleotides) with acetonitrile as eluent and a flow rate of 1.0 ml/min, gradient elution 1.25% acetonitrile/min. Preparatory HPLC was carried out on a Beckman instrument using a Waters DeltaPak 25 x 100 mm 100 μ m C₁₈ column in 0.1% (wt/v) TFA (for polyamides) or 100 mM triethylammonium acetate (pH = 7.0) (for polyamide-oligonucleotides), gradient elution 0.25%/min. CH₃CN. Water was obtained from a Millipore Milli-Q water purification system. Reagent-grade chemicals were used unless otherwise stated. Di-N-hydroxy succinimidyl glutarate (DSG) was obtained from Pierce. *E. coli* XL-1 Blue competent cells were obtained from Stratagene. Restriction endonucleases were purchased from either New England Biolabs or Boehringer-Mannheim and used according to the manufacturer's protocol. Sequenase (version 2.0)

was obtained from United States Biochemical, and DNase I (FPLCpure) was obtained from Pharmacia. $[\alpha\text{-}^{32}\text{P}]$ -Thymidine-5'-triphosphate (≥ 3000 C_i/mmol), $[\alpha\text{-}^{32}\text{P}]$ -deoxyadenosine-5'-triphosphate (≥ 6000 C_i/mmol), and $[\gamma\text{-}^{32}\text{P}]$ -adenosine-5'-triphosphate were purchased from Du Pont/NEN.

Synthesis of Polyamides.

AcImPyPy-G-PyPyPy-G-Dp 1. AcImPyPy-G-PyPyPy-G-Pam-resin was synthesized in stepwise fashion by machine-assisted solid phase methods.⁴¹ A sample of AcImPyPy-G-PyPyPy-G-Pam-resin (240 mg, 0.2 mmol/g) was treated with neat (dimethylamino)-propylamine (2 ml) and heated to (55 °C) with periodic agitation for 18 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. AcImPyPy-G-PyPyPy-G-Dp **1** is recovered upon lyophilization of the appropriate fractions as a white powder (17 mg, 36% recovery): UV (H₂O) λ_{max} 260, 310 (50 000) MALDI-TOF MS (monoisotopic) [M+H] 993.02 (992.06 calc for C₄₆H₅₇N₁₇O₉).

H₂N-ε-(CH₂)₅ImPyPy-G-PyPyPy-G-Dp 5. H₂N-ε-(CH₂)₅-ImPyPy-G-PyPyPy-G-Pam-resin was synthesized in stepwise fashion by machine-assisted solid phase methods.⁴¹ A sample of H₂N-ε-(CH₂)₅-ImPyPy-G-PyPyPy-G-Pam-resin (1.5 g, 0.2 mmol/g) was treated with neat (dimethylamino)-propylamine (4 ml) and heated to (55 °C) with periodic agitation for 18 h. The reaction

mixture was then filtered to remove resin, 0.1% (wt/v) TFA added (12 ml) and the resulting solution purified by reversed phase HPLC. $\text{H}_2\text{N}-\epsilon-(\text{CH}_2)_5\text{-ImPyPy-G-PyPyPy-G-Dp 5}$ is recovered upon lyophilization of the appropriate fractions as a white powder (102 mg, 32% recovery): UV (H_2O) λ_{max} 260, 310 (50 000) MALDI-TOF MS (monoisotopic) $[\text{M}+\text{H}]$ 1063.82 (1063.19 calc for $\text{C}_{50}\text{H}_{55}\text{N}_{18}\text{O}_9$).

Synthesis of Polyamide Oligonucleotides.

3'-TTT^{Me}C^{Me}CTTTTTT-5'-DSG-(CH_2)₅-CONHImPyPy-G-PyPyPy-G-Dp 3.

N-alloc-5'-amino-thymidine-TTTTT^{Me}C^{Me}CTTT-CPG was synthesized on an Applied Biosystems Model 380B DNA synthesizer using standard β -cyanoethyl phosphoramidite chemistry on 10 μmol scale.^{42,43} The alloc protecting group was removed on solid support by combining the CPG bound oligonucleotide with tris(dibenzylideneaceto)-di-palladium(0)• CHCl_3 , ($\text{Pd}_2(\text{dba})_2 \cdot \text{CHCl}_3$, 196 mg, 0.19 mmol), triphenyl phosphine (686 mg, 2.62 mmol), and a freshly made solution of 0.25 M *n*-butyl ammonium formate in THF (7 ml) in a reaction vessel with shaking for 6 h at 24 °C.⁴⁷ The support was washed with sequentially with THF, acetone, 0.1 M sodium diethyldithiocarbamate (DTTC, aq, 2x), acetone, water, DTTC (2x), acetone, water, DCM, MeOH, DCM, DMF. A sample of support (2.8 μmol) was treated with disuccinimidyl-glutarate (65 mg, 250 μl DIEA, 750 μl DMF, 37 °C) to provide a support bound *N*-hydroxysuccinimidyl ester which was allowed to

react with a polyamide containing a free primary amine **5** (17 mg, 0.3 ml DMF, 0.1 ml DIEA, 37 °C, 12 h). CPG-TTT^{Me}C^{Me}CTTTTTT-5'-DSG-(CH₂)₅-CONHImPyPy-G-PyPyPy-G-Dp was isolated by filtration, washed sequentially with DMF, DCM, MeOH, diethyl ether and dried *in vacuo*. The support was treated with 0.1 M NaOH (2 ml) and heated (55 °C) for 12 h. The reaction mixture was then filtered to remove CPG, 0.1 M triethylammonium acetate (pH = 7.0) added (6 ml) and the resulting solution purified by reversed phase HPLC. 3'-TTT^{Me}C^{Me}CTTTTTT-5'-DSG-(CH₂)₅-CONHImPyPy-G-PyPyPy-G-Dp is recovered upon lyophilization of the appropriate fractions (400 NOME, 14% recovery): UV (H₂O) λ_{\max} 260 (138 000), 310 (50 000); MALDI-TOF MS [M-H] obsd. 4511 (calc 4516).

3'-T^{Me}CT^{Me}CTTTTTTT-5'-DSG-(CH₂)₅-CONHImPyPy-G-PyPyPy-G-Dp **4.**

CPG-T^{Me}CT^{Me}CTTTTTTT-5'-DSG-(CH₂)₅-CONHImPyPy-G-PyPyPy-G-Dp was synthesized by machine assisted solid phase methods as described for **3**. A sample of CPG-T^{Me}CT^{Me}CTTTTTTT-5'-DSG-(CH₂)₅-CONHImPyPy-G-PyPyPy-G-Dp was treated with 0.1 M NaOH (2 ml) and heated (55 °C) for 12 h. The reaction mixture was then filtered to remove CPG, 0.1 M triethylammonium acetate (pH = 7.0) added (6 ml) and the resulting solution purified by reversed phase HPLC. 3'-T^{Me}CT^{Me}CTTTTTTT-5'-DSG-(CH₂)₅-CONHImPyPy-G-PyPyPy-G-Dp is recovered upon lyophilization of the appropriate fractions (900 nmol, 32% recovery): UV (H₂O) λ_{\max} 260 (138 000), 310 (50 000); MALDI-TOF MS [M-H] obsd. 4514 (calc 4516).

Preparation of ^{32}P -labeled DNA. The 3'- ^{32}P end-labeled *Eco* RI/ *Pvu* II fragment was prepared by digesting the plasmid pJWS8² with *Eco* RI and *Pvu* II and simultaneously filling in using Sequenase, [α - ^{32}P]-deoxyadenosine-5'-triphosphate, and [α - ^{32}P]-thymidine-5'-triphosphate. The 268 bp fragment was isolated by nondenaturing gel electrophoresis. A and G sequencing were carried out as described.^{49,50} Standard methods were used for all DNA manipulations.⁵¹

Quantitative DNase I footprint titration experiments. All reactions were executed in a total volume of (i) 40 μl or (ii) 400 μl .³²⁻³⁴ A ligand stock solution or H_2O (for reference lanes) was added to an assay buffer containing radiolabeled restriction fragment (15,000 cpm), affording final solution conditions of 10 mM Bis Tris•HCl (pH 7.0), 10 mM NaCl, 250 μM spermine, and either (a) serially diluted ligands **1-4** or (b) no ligand (for reference lanes). The solutions were allowed to equilibrate at 22 °C for 12 h for polyamide **1** or 24 h for compounds **2-4**. Footprinting reactions were initiated by the addition of (i) 4 or (ii) 10 μl of a DNase I stock solution (i) 4 or (ii) 10 μl ; at the appropriate concentration to give ~ 55% intact DNA) containing MgCl_2 (i) 50 or (ii) 200 mM), CaCl_2 (i) 50 or (ii) 200 mM), glycerol (i) 5 or (ii) 20%), and Bis Tris•HCl (10 mM) and allowed to proceed for six minutes at 22 °C. The reactions were stopped by the addition of a solution (i) 10 or (ii) 50 μl),

containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/mL glycogen, and 30 μ M base-pair calf thymus DNA, and ethanol precipitated. Reactions were resuspended in 1X TBE/80% formamide loading buffer, denatured by heating at 85 °C for 10 min, and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% cross-link, 7 M urea) in 1X TBE at 2000 V. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics).⁵²

Quantitation and data analysis. Data from the footprint titration gels were obtained using a Molecular Dynamics 400S PhosphorImager followed by quantitation using ImageQuant software (Molecular Dynamics). Background-corrected volume integration of rectangles encompassing the footprint sites and a reference site at which DNase I reactivity was invariant across the titration generated values for the site intensities (I_{site}) and the reference intensity (I_{ref}). The apparent fractional occupancy (θ_{app}) of the sites were calculated using the equation:

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

where I_{site}° and I_{ref}° are the site and reference intensities, respectively, from a control lane to which no polyamide was added. The ($[L]_{\text{tot}}$, θ_{app}) data points were fit to a general Hill equation (eq 2) by minimizing the difference between θ_{app} and θ_{fit} :

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

where $[L]_{\text{tot}}$ is the total ligand concentration, K_a is the apparent first-order association constant, and θ_{min} and θ_{max} are the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. The data were fit using a nonlinear least-squares fitting procedure with K_a , θ_{max} , and θ_{min} as the adjustable parameters. For polyamide ImPyPy-G-PyPyPy-G-Dp **1** binding isotherms for the 5'-TTAATGTCATTAA-3' target site were adequately fit by Langmuir isotherms (eq 2, $n=2$), consistent with formation of a 2:1 polyamide-DNA complexes. The mismatched polyamide-oligonucleotide **4**, only footprinted the 13 bp minor groove binding site, 5'-TTAATGTCATTAA-3', and was fit by Langmuir isotherms (eq 2, $n=2$), consistent with formation of a 2:1 polyamide complex in the minor groove and no binding in the major groove. For the matched conjugate **3**, shallower binding isotherms ($n=1$) were observed at the 31 base pair target site. We believe that the shape of these isotherms is consistent with binding being comprised of a series of binding events, and emphasize that treatment of the data in this manner does not represent an attempt to model a binding mechanism. Rather, we have chosen to compare values of the apparent first-order association constant, a value that represents the concentration of ligand at which a site is half-saturated. The binding isotherms were normalized using the following equation:

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

Three sets of data were used in determining each association constant. The method for determining association constants used here involves the assumption that $[L]_{\text{tot}} \approx [L]_{\text{free}}$, where $[L]_{\text{free}}$ is the concentration of polyamide-oligonucleotide free in solution (unbound). For very high association constants this assumption becomes invalid, resulting in underestimated association constants. In the experiments described here, the DNA concentration is estimated to be ~ 10 pM. As a consequence, a measured association constant of $1 \times 10^{10} \text{ M}^{-1}$ underestimates the true association constant by a factor of 4.

References

1. Szewczyk, J.W.; Baird, E.E.; Dervan, P.B. (1996) Cooperative Triple-Helix Formation via a Minor Groove Dimerization Domain *J. Am. Chem. Soc.* **117**, 7863-7864.
2. Szewczyk, J. W.; Baird, E.E.; Dervan, P.B. (1996) Sequence Specific Recognition of DNA by a Major and Minor Groove Binding Ligand. *Angew. Chem. Int. Ed.* **35**, 1487-1489.
3. Wade, W.S.; Mrksich, M.; Dervan, P.B. (1992) Design of peptides that bind in the minor groove of DNA at 5'-(A,T)₅G(A,T)C(A,T) sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114**, 8783-8794.
4. Wade, W.S.; Mrksich, M.; Dervan, P.B. (1993) Binding affinities of synthetic peptides, pyridine-2-carboxamidenetropsin and 1-methylimidazole-2-carboxamidenetropsin, that form 2/1 complexes in the minor groove of double-helical DNA. *Biochemistry* **32**, 11385-11389.
5. Mrksich, M.; Wade, W.S.; Dwyer, T.J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. (1992) Antiparallel Side-by-Side Dimeric Motif for Sequence-Specific Recognition in the Minor Groove of DNA by the Designed Peptide 1-Methylimidazole-2-carboxamidenetropsin. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 7586-7590.
6. Moser, H.E.; Dervan, P.B. (1987) Sequence-Specific Cleavage of Double Helical DNA by Triple Helix Formation *Science*, **238**, 645-650.
7. For a review, see: Thuong, N.T.; Helene, C. Sequence-Specific Recognition and Modification of Double-Helical DNA by Oligonucleotides. (1993) *Angew. Chem., Int. Ed. Engl.* **32**, 666-690, and references cited therein.
8. Gottesfeld, J.M.; Neely, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B. (1997) Regulation of gene-expression by small molecules. *Nature* **387**, 202-205
9. Trauger, J. W.; Baird, E.E.; Dervan, P.B. (1996) Recognition of DNA by Designed Ligands at Subnanomolar Concentrations. *Nature*, **382**, 559-561.
10. Turner, J. M.; Baird, E. E.; Dervan, P. B. (1997) Recognition of 7 base-pair sequences in the minor groove of DNA by 10-ring pyrrole-imidazole polyamides. *J. Am. Chem. Soc.* **119**, 7636-7644.

11. Herman, D.H.; Baird, E.E.; Dervan, P.B. (1998) Tandem hairpin motif: for recognition of the minor groove of DNA by pyrrole imidazole polyamides. *Chem. Eur. J.* *in press*.
12. Pelton, J.G.; Wemmer, D.E. (1989) Structural characterization of a 2-1 distamycin A•d(CGCAAATTTGGC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. USA* **86**, 5723-5727.
13. Mrksich, M.; Dervan, P.B. (1993) Antiparallel side-by-side heterodimer for sequence-specific recognition in the minor groove of DNA by a distamycin 1-methylimidazole-2-carboxamide pair. *J. Am. Chem. Soc.* **115**, 2572-2576.
14. Geierstanger, B.H.; Dwyer, T.J.; Bathini, Y.; Lown, J.W.; Wemmer, D.E. (1993) NMR characterization of a heterocomplex formed by distamycin and its analog 2-ImD with d(CGCAAGTTGGC)-d(GCCAACTTGCG) - preference for the 1/1/1 2-ImD-Dst-DNA complex over the 2/1 2-ImD-DNA and the 2/1 Dst-DNA complexes. *J. Am. Chem. Soc.* **115**, 4474-4482.
15. White, S.; Baird, E. E.; Dervan, P.B. (1997) On the pairing rules for recognition in the minor-groove of DNA by pyrrole-imidazole polyamides. *Chem. Biol.* **4**, 569-578.
16. White, S.; Baird, E.E.; Dervan, P.B. (1997) Orientation preferences of pyrrole-imidazole polyamides in the minor-groove of DNA. *J. Am. Chem. Soc.* **119**, 8756-8765.
17. Pelton, J.G.; Wemmer, D.E. (1990) Binding modes of distamycin-A with d(CGCAAATTTGCG)₂ determined by Two-dimensional NMR *J. Am. Chem. Soc.* **112**, 1393-1399.
18. Chen, X.; Ramakrishnan, B.; Rao, S.T.; Sundaralingam, M. N. (1994) Binding of 2 distamycin-A molecules in the minor groove of an alternating B-DNA duplex. *Nature Struct. Biol.* **1**, 169-175.
19. White, S.; Baird, E.E.; Dervan, P.B. (1996) Sequence Composition Effects on a Hairpin Polyamide for Recognition of the Minor Groove of DNA *Biochem.* **35**, 12532-12537.
20. White, S. E.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. (1998) Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature*, **391**, 468-471.

21. Kielkopf, C.L.; White, S.W.; Szewczyk, J.W.; Turner, J.M.; Baird, E.E.; Dervan, P.B.; Rees, D.C. (1998) A structural basis for recognition of A•T and T•A base pairs in the minor groove of B-DNA. *Science* **282**, 111-115.
22. Pilch, D. S., *et al.*; Dervan, P. B. (1996) Binding of a hairpin polyamide in the minor groove of DNA: sequence-specific enthalpic discrimination. *Proc. Natl. Acad. Sci. USA* **93**, 8306-8311.
23. Mrksich, M.; Dervan, P.B. (1995) Recognition in the minor-groove of DNA at 5'-(A,T)GCGC(A,T)-3' by a 4-ring tripeptide dimer - reversal of the specificity of the natural product distamycin. *J. Am. Chem. Soc.* **117**, 3325-3332.
24. Trauger, J. W.; Baird, E.E.; Dervan, P.B. (1996) Extended Hairpin Polyamide Motif for Sequence Specific Recognition in the Minor Groove. *Chem. and Bio.* **3**, 369-377.
25. Parks, M. E.; Baird, E. E.; Dervan, P. B. (1996) Optimization of the Hairpin Polyamide Design for Recognition of the Minor Groove of DNA. *J. Am. Chem. Soc.* **118**, 6147-6152.
26. Parks, M. E.; Baird, E. E.; Dervan, P. B. (1996) Recognition of 5'-(A,T)GG(A,T)₂-3' Sequences in the Minor Groove of DNA by Hairpin Polyamides. *J. Am. Chem. Soc.* **118**, 6153-6159.
27. Kelly, J. J.; Baird, E. E.; Dervan, P.B. (1995) Binding Site Size Limit of the 2:1 Pyrrole- Imidazole Polyamide-DNA Motif. *Proc. Natl. Acad. Sci. USA* **93**, 6981-6985.
28. Kielkopf, C.L.; Baird, E.E.; Dervan, P.B.; Rees, D.C. (1998) Structural basis for G•C recognition in the DNA minor groove. *Nat. Struc. Biol.* **5**, 104-109.
29. de Clairac, R.P.L.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. (1997) NMR characterization of hairpin polyamide complexes with the minor groove of DNA. *J. Am. Chem. Soc.* **119**, 7909-7916.
30. Geierstanger, B.H.; Jacobsen, J.P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. (1994) Structural and dynamic characterization of the heterodimeric and homodimeric complexes of distamycin and 1-methylimidazole-2-carboxamide-netropsin bound to the minor-groove of DNA. *Biochemistry* **33**, 3055-3062.

31. Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. (1994) Design of a G•C-specific minor groove-binding peptide. *Science* **266**, 646-650.
32. Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Pecout, J.L.; Thoung, N.T.; Lhomme, J.; Helene, C. (1987) Sequence-specific recognition, photocrosslinking and cleavage of the DNA double helix by an oligo-alpha-thymidylate covalently linked to an azidoproflavine derivative. *Nucleic Acids Res.* **15**, 7749-7760.
33. Strobel, S.A.; Doucettstamm, L.A.; Riba, L.; Housmann, D.E.; Dervan, P.B. (1991) Site-specific cleavage of human chromosome-4 mediated by triple-helix formation. *Science* **254**, 1639-1642.
34. Allinquant, B.; Hantraye, P.; Mailleux, P.; Moya, K.; Bouillot, C.; Prochiantz, A. (1994) Down-regulation of amyloid precursor protein inhibits neurite outgrowth *in-vitro*. *J. Cell. Biol.* **128**, 919-927.
35. Pooga, M.; Soomets, U.; Hallbrink, M.; Valkna, A.; Saar, K.; Rezaei, K.; Kahl, U.; Hao, J.X.; Xu, X.J.; et al. (1998) Penetrating PNA constructs regulate galanin receptor levels and modify pain transmission *in-vivo*. *Nat. Biotech.* **16**, 857-861.
36. Nealy, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B.; Gottesfeld, J. M. (1997) Importance of minor groove binding zinc fingers within the transcription factor IIIA•DNA complex. *J. Mol. Biol.* **274**, 439-445.
37. Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. (1996) Extension of Sequence-Specific Recognition in the Minor Groove of DNA by Pyrrole-Imidazole Polyamides to 9-13 Base Pairs. *J. Am. Chem. Soc.* **118**, 6147-6152.
38. Fox, K.R.; Waring, M.J. (1984) DNA structural variations produced by actinomycin and distamycin as revealed by DNase-I footprinting. *Nucleic Acids Res.* **12**, 9271-9285.
39. Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. (1986) Quantitative DNase footprint titration - a method for studying protein-DNA interactions. *Methods Enzymol.* **130**, 132-181.
40. Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. (1986) Footprint titrations yield valid thermodynamic isotherms. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8462-8466.

41. Baird, E. E.; Dervan, P. B. (1996) Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141-6146.
42. Yamamoto, I.; Sekine, M.; Hata, T. (1978) One-step synthesis of 5'-azido-nucleosides. *J. Chem. Soc. Perkin I.* 306-310.
43. Brannwarth, W. (1988) Solid-phase synthesis of oligodeoxynucleotides containing phosphoramidate internucleotide linkages and their specific chemical cleavage. *Helv. Chim. Acta* **71**, 1517-1527.
44. Haralambidis, J.; Duncan, L.; Angus, K.; Tregear, G.W. (1990) The synthesis of polyamide oligonucleotide conjugate molecules. *Nucleic Acids Res.* **18**, 493-499.
45. Sinyakov, A.N.; Lokhov, S.G.; Kutuyavin, I.V.; Gamper, H.B.; Meyer, R.B. (1995) Exceptional and selective stabilization of A-T rich DNA •DNA duplexes by *N*-methylpyrrole carboxamide peptides conjugated to oligodeoxynucleotides *J. Am. Chem. Soc.* **117**, 4995-4996.
46. Tetzlaff, C.N.; Schwöpe, I.; Bleczyński, C.F.; Steinberg, J.A.; Richert, C. (1998) A convenient synthesis of 5'-amino-5'-deoxythymidine and preparation of peptide-DNA hybrids. *Tet. Let.* **39**, 4215-4218.
47. Hayakawa, Y.; Wakabayashi, S.; Kato, H.; Noyori, R. (1990) The allylic protection method in solid-phase oligonucleotide synthesis- an efficient preparation of solid-anchored DNA oligomers. *J. Am. Chem. Soc.* **112**, 1691-1696.
48. Parks, M.E.; Dervan, P.B. (1996) Simultaneous Binding of a polyamide dimer and an oligonucleotide in the minor and major grooves of DNA. *Bioorg. Med. Chem.* **4**, 1045-1050.
49. Maxam, A.M.; Gilbert, W.S. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 499-560.
50. Iverson, B.L.; Dervan, P.B. (1987) Adenine-specific DNA chemical sequencing reaction. *Methods Enzymol.* **15**, 7823-7830.
51. Sambrook, J.; Fritsch, E.F.; Maniatis, T. (1989) *Molecular Cloning*. (2nd edn). Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
52. Johnston, R. F.; Picket, S. C.; Barker, D. L. (1990) Autoradiography using storage phosphor technology. *Electrophoresis* **11**, 355.

Chapter 5

Recognition of the Four Watson-Crick Base Pairs in the DNA Minor Groove by Synthetic-Ligands

***Abstract** The design of synthetic ligands that read the information stored in the DNA double helix has been a long standing goal at the interface of chemistry and biology. Cell-permeable small molecules which target predetermined DNA sequences offer a potential approach for the regulation of gene-expression. Oligodeoxynucleotides that recognize the major groove of double-helical DNA via triple-helix formation bind to a broad range of sequences with high affinity and specificity. Although oligonucleotides and their analogs have been shown to interfere with gene expression, the triple helix approach is limited to purine tracks and suffers from poor cellular uptake. The subsequent development of pairing rules for minor groove binding polyamides containing pyrrole (Py) and imidazole (Im) amino acids offers a second code to control sequence specificity. An Im/Py pair distinguishes G•C from C•G and both of these from A•T/T•A base pairs. A Py/Py pair specifies A,T from G,C but does not distinguish A•T from T•A. In order to break this degeneracy, a new aromatic amino acid, 3-hydroxypyrrole (Hp), has been added to the repertoire to test for pairings which discriminate*

A•T from T•A. We find that replacement of a single hydrogen atom with a hydroxy group in a Hp/Py pairing regulates affinity and specificity by an order of magnitude. By incorporation of a third amino acid, hydroxypyrrole-imidazole-pyrrole polyamides form four ring-pairings (Im/Py, Py/Im, Hp/Py, and Py/Hp) which distinguish all four Watson-Crick base pairs in the minor groove of DNA.

Publications: White, Szewczyk, Turner, Baird, & Dervan; *Nature* **1998**, 391, 468.

Kielkopf, White, Szewczyk, Turner, Baird, Dervan, & Rees; *Science* **1998**, 282, 111.

Due to degeneracy of the hydrogen bond donors and acceptors displayed on the edges of the base pairs, the minor groove was thought to lack sufficient information for a complete recognition code.¹ However, despite the central placement of the guanine exocyclic N2 amine group in the G,C minor groove,^{2,3} Py/Im and Im/Py pairings distinguish energetically G•C and C•G.^{1,4-9} The neighboring Py packs an Im to one side of the minor groove resulting in a precisely placed hydrogen bond between Im N3 and guanine N2 for specific recognition.¹⁰⁻¹¹ This remarkable sensitivity to single atomic replacement indicates that substitution at the 3 position of one Py within a Py/Py pair can compliment small structural differences at the edges of the base pairs in the center of the minor groove.

For A,T base pairs, the hydrogen bond acceptors at N3 of adenine and O2 of thymine are almost identically placed in the minor groove, making hydrogen bond discrimination a challenge (Figure 5.1).² The existence of an asymmetrically placed cleft on the minor groove surface between the thymine O2 and the adenine 2H suggests a possible shape-selective mechanism for A•T recognition.¹² We reasoned that substitution of C3-H by C3-OH within a Py/Py pair would create 3-hydroxypyrrole (Hp)/Py pairings to discriminate T•A from A•T (Figure 5.2). Selectivity could potentially arise from steric destabilization of polyamide binding via placement of Hp opposite A or stabilization by a specific hydrogen bond between Hp and T.

Four-ring polyamide subunits, covalently coupled to form eight-ring hairpin structures, bind specifically to 6-bp target sequences at subnanomolar

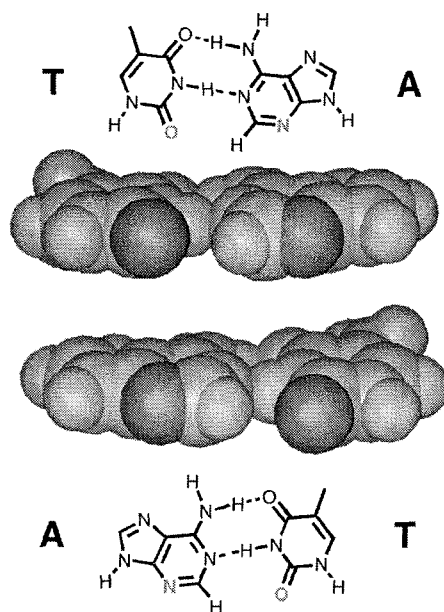


Figure 5.1 Chemical structures and space-filling models of the T•A and A•T base pairs as viewed from the minor groove of DNA. Models generated using B-form DNA coordinates provided in InsightII. The hydrogen-bond acceptors (N3 of adenine and O2 of thymine) are in dark gray.

concentrations.⁷ We report here the DNA-binding affinities of three eight-ring hairpin polyamides containing pairings of Im/Py, Py/Im opposite G•C, C•G and either Py/Py, Hp/Py or Py/Hp at a common single point opposite T•A and A•T (Figure 5.2a). Equilibrium association constants (K_a) for ImImPyPy- γ -ImPyPyPy- β -Dp **1**, ImImPyPy- γ -ImHpPyPy- β -Dp **2**, and ImImHpPy- γ -ImPyPyPy- β -Dp **3** were determined by quantitative DNase I footprint titration experiments¹³ on a 3' ³²P-labeled 250-bp DNA fragment containing the target sites, 5'-TGGACA-3' and 5'-TGGTCA-3', which differ by a single A,T base pair in the fourth position (Figure 5.2b).

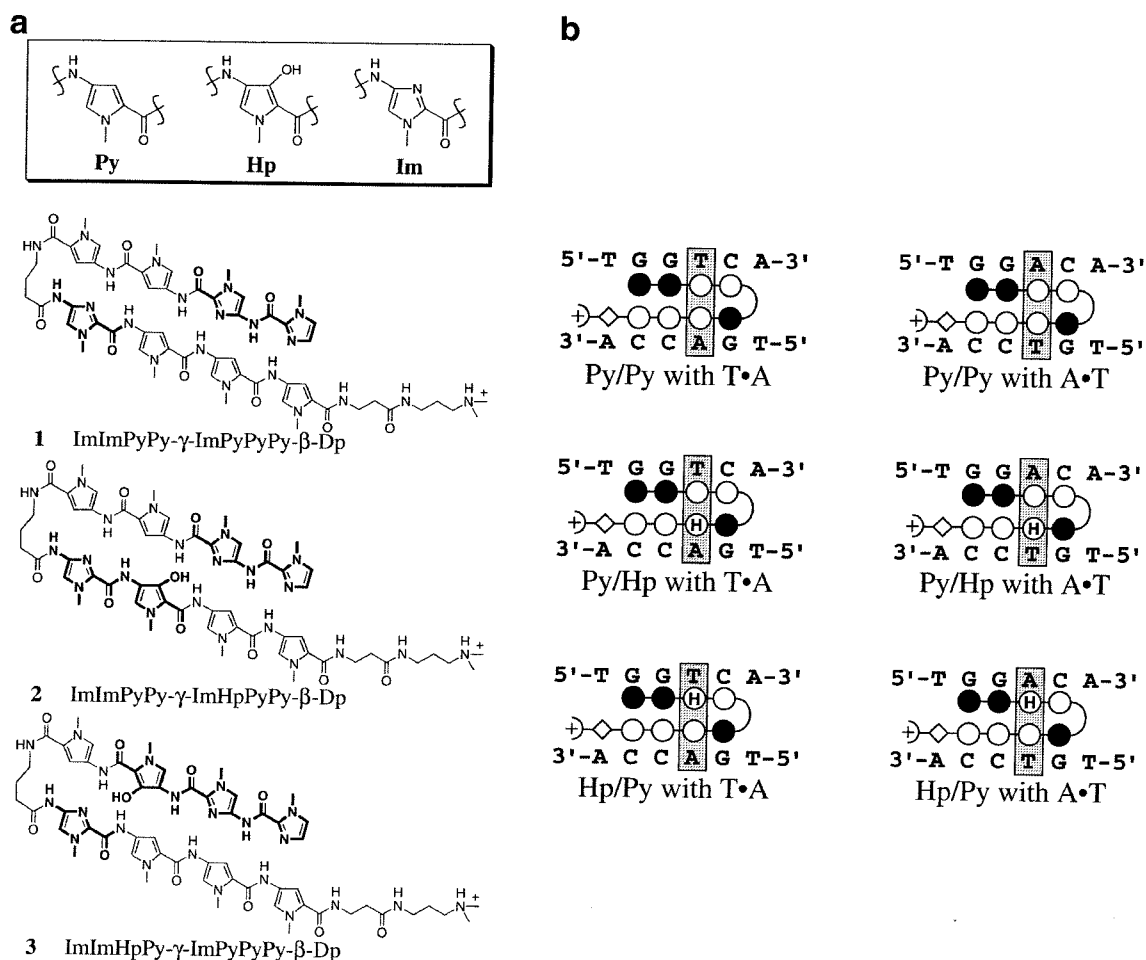


Figure 5.2 Eight-ring hairpin polyamides containing three aromatic amino acids (Py, Hp, Im). **a**, Structures of polyamides ImImPyPy-γ-ImPyPyPy-β-Dp (1), ImImPyPy-γ-ImHpPyPy-β-Dp (2), and ImImHpPy-γ-ImPyPyPy-β-Dp (3). (Hp = 3-hydroxypyrrole, Im = imidazole, Py = pyrrole, β= b-alanine, γ= γ-aminobutyric acid, Dp = dimethylaminopropylamide). **b**, Binding models for polyamides 1-3 in complex with 5'-TGGTCA-3' and 5'-TGGACA-3' (A•T and T•A in the fourth position highlighted). Filled and unfilled circles represent imidazole and pyrrole rings respectively; circles containing an H represent 3-hydroxypyrrole, the curved line connecting the polyamide subunits represents γ-aminobutyric acid, the diamond represents β-alanine, and the + represents the positively charged dimethylaminopropylamide tail group.

Table 5.1		Equilibrium Association Constants (M^{-1}) ^a		
Polyamide ^b		K_a		K_{rel} ^c
		5'-TGG <u>T</u> CA-3'	5'-TGG <u>A</u> CA-3'	
1	Py/Py	$1.3 (\pm 0.9) \times 10^{10}$	$6.8 (\pm 0.4) \times 10^9$	2
2	Py/Hp	$6.8 (\pm 1.0) \times 10^7$	$1.2 (\pm 0.9) \times 10^9$	0.06
3	Hp/Py	$2.1 (\pm 0.7) \times 10^9$	$2.7 (\pm 0.9) \times 10^7$	77

^aThe reported association constants are the average values obtained from three DNase I footprint titration experiments. The standard deviation for each data set is less than 15% of the reported number. Assays were carried out in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂ at pH 7.0 and 22 °C. ^bRing pairing opposite T•A and A•T in the fourth position of 6 base pair binding site. ^c K_{rel} is calculated as K_a (TGGACA)/ K_a (TGGTCA).

DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂, pH 7.0, 22 °C) were performed to determine the equilibrium association constants K_a for recognition of bound sites (Table 5.1). Based on the pairing rules for polyamide-DNA complexes, the sequences 5'-TGGACA-3' and 5'-TGGTCA-3' are match sites for control polyamide 1 which places a Py/Py pairing opposite A•T and T•A at both sites. We find that polyamide 1 (Py/Py) binds to 5'-TGGTCA-3' and 5'-TGGACA-3' within a factor of 2 ($K_a = 1.3 (\pm 0.9) \times 10^{10} M^{-1}$ or $6.8 (\pm 0.4) \times 10^9 M^{-1}$, respectively). In contrast, polyamide 2 (Py/Hp) binds to 5'-TGGTCA-3' and 5'-TGGACA-3' with dissociation constants which differ by a factor of 18 ($K_a = 6.8 (\pm 1.0) \times 10^7 M^{-1}$ and $1.2 (\pm 0.9) \times 10^9 M^{-1}$, respectively). By reversing the pairing in polyamide 3 (Hp/Py), the association constants differ again in the opposite direction by a factor of 77 ($K_a = 2.1 (\pm 0.7) \times 10^9 M^{-1}$ and $2.7 (\pm 0.9) \times 10^7 M^{-1}$ respectively).

Control experiments performed on separate DNA fragments reveal that neither a 5'-TGGGCA-3' nor a 5'-TGGCCA-3' site is bound by polyamide **2** or **3** with a $K_a \leq 1.0 \times 10^7 \text{ M}^{-1}$, indicating that the Hp/Py and Py/Py ring pairings do not bind opposite G•C or C•G.

The discrimination between A•T vs. T•A. is achieved when the two neighboring base pairs are G•C and C•G (GTC vs. GAC). A general rule would require that the same discrimination be observed when adjacent A•T/T•A base pairs are present in the target sequence. Further sequence composition studies of six-ring hydroxypyrrole-imidazole-pyrrole hairpin polyamides confirm A•T/T•A discrimination at sequence contexts which include TTT, TAA, TAT, TTA, ATT, GTT, GAT, and GTA.

The specificity of **2** and **3** for sites which differ by a single A•T/T•A base pair results from small chemical changes. Replacing the Py/Py pair in **1** with a Py/Py pairing as in **2**, a single substitution of C3-OH for C3-H, destabilizes interaction with 5'-TGGICA-3' by 191-fold, a free energy difference of $3.1 \text{ kcal mol}^{-1}$ (13.0 kJ mol^{-1}). Interaction of **2** with 5'-TGGACA-3' is destabilized only 6-fold relative to **1**, a free energy difference of $1.1 \text{ kcal mol}^{-1}$ (4.6 kJ mol^{-1}). Similarly, replacing the Py/Py pair in **1** with Hp/Py as in **3** destabilizes interaction with 5'-TGGACA-3' by 252-fold, a free energy difference of $3.2 \text{ kcal mol}^{-1}$. Interaction of **3** with 5'-TGGICA-3' is destabilized only 6-fold relative to **1**, a free energy difference of $1.0 \text{ kcal mol}^{-1}$.

Molecular recognition in nature occurs via an ensemble of stabilizing and destabilizing forces.³ For example, the DNA-binding transcription factor,

TBP, recognizes 5'-TATA-3' sequences in the DNA minor groove via specific contacts and a large DNA-bend.¹⁴ For DNA-bending proteins such as TBP, practical sequence-specificity can be enhanced via the sequence-dependent energetics of DNA distortion.¹⁵ Homeodomain proteins recognize target sequences via a combination of specific interactions with both the major and minor grooves.¹⁶ An N-terminal arm recognizes 5'-TAAT-3 sequences in the minor groove such that a single substitution of T•A for A•T reduces binding at 5'-TTAT-3' by a factor of 7.¹⁷ However, no single protein structure motif has been identified which provides a general amino acid-base pair code for the minor groove, although there has been remarkable progress with zinc fingers in the major groove.¹⁸

Table 5.2 Pairing code for minor groove recognition

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

Favored (+), disfavored (--).

Polyamides use a single molecular shape to provide for coded targeting of predetermined DNA sequences with affinity and specificity comparable to sequence-specific DNA binding proteins.⁷ Hydroxypyrrole-imidazole-pyrrole

polyamides complete the minor groove recognition code using three aromatic amino acids, which combine to form four ring pairings (Im/Py, Py/Im, Hp/Py, and Py/Hp) that complement the four Watson-Crick base pairs (Table 5.2).

ENERGETICS OF 3-HYDROXYPYRROLE RING PAIRINGS FOR PLACEMENT OPPOSITE T•A, A•T, G•C, AND C•G BASE PAIRS IN THE DNA MINOR GROOVE.

Polyamides containing the three aromatic amino acids 3-hydroxypyrrole (Hp), imidazole (Im) and pyrrole (Py) are synthetic ligands that bind to predetermined DNA sequences with subnanomolar affinity.^{7,19-20} 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.^{1,4-9} 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.^{4,6} A Py/Py pair binds both A•T and T•A in preference to G•C/C•G.⁵ A Hp/Py specifies T•A from A•T, and both of these from G•C/C•G.¹⁹⁻²⁰ The T•A selectivity of the Hp/Py pair arises from two discrete mechanisms for recognition: (i) shape selection of an asymmetric cleft on the floor of the minor groove formed by the O2 of thymine and 2H of adenine, and (ii) formation of two specific hydrogen bonds between the 3-hydroxyl and 4-carboxamido groups of Hp with the two lone pairs on the O2 of thymine (Figure 5.3 and 5.4).²⁰ A potential degeneracy which has not been previously addressed is the energetics of a Hp/Hp pairing opposite the four Watson-Crick base pairs.

Four eight-ring hairpin polyamides^{6a-c,11} containing a single binary combination of Py and Hp form all four possible ring pairings (Py/Py, Py/Hp, Hp/Py, or Hp/Hp). Relative energetic preferences of each ring pairing opposite each of the four Watson-Crick base pairs are reported here (Figure

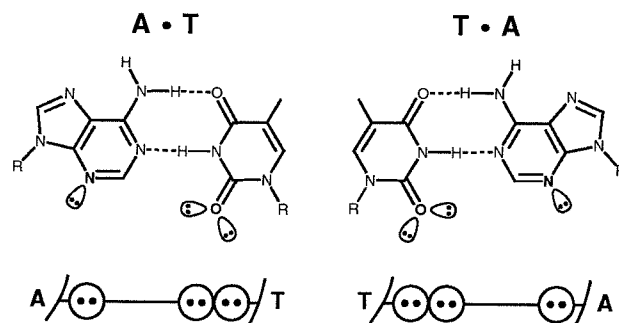
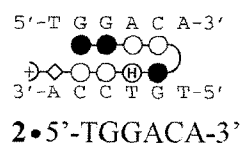
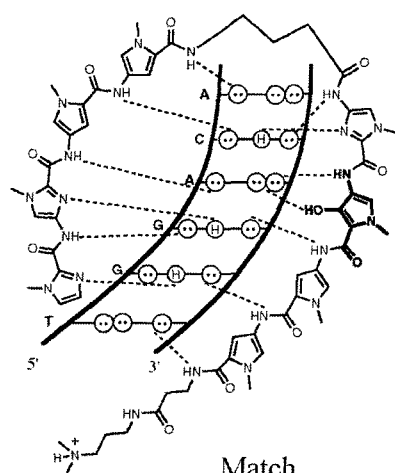


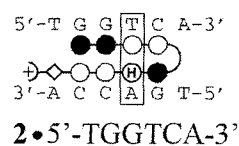
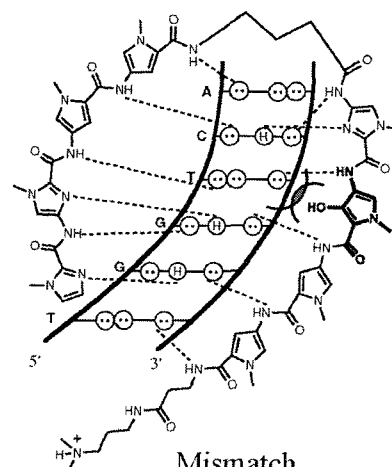
Figure 5.3 (Top) Chemical structures for the A•T and T•A base pairs. Lone pair electrons in the minor groove are shown as ovals. (Bottom) Schematic representations of the A•T and T•A bases in the minor groove. A circle with two dots represents the lone pair of the N3 of adenine. Two touching circles with dots represent the two lone pairs of the O2 of thymine.

5.5). According to the pairing rules, polyamides with sequences ImImPyPy- γ -ImPyPyPy- β (β : β -alanine) are expected to target the 6 base pair sites 5'-TGGTCA-3' and 5'-TGGACA-3' with comparable affinity. Selective substitution into each four-ring subunit modifies a single pairing within the hairpin structure placing each of the four combinations (Py/Py, Py/Hp, Hp/Py, or Hp/Hp) opposite each of the four base pairs (T•A, A•T, G•C, C•G). The four eight-ring polyamides, ImImPyPy- γ -ImPyPyPy- β **1**, ImImPyPy- γ -ImHpPyPy- β **2**, ImImHpPy- γ -ImPyPyPy- β **3**, and ImImHpPy- γ -ImHpPyPy- β **4**, were synthesized by solid phase methods²¹ containing each of the ring pairings (Py/Py, Py/Hp, Hp/Py, and Hp/Hp, respectively) in a hairpin structure (Figure 5.6). We describe the synthesis of the appropriately protected

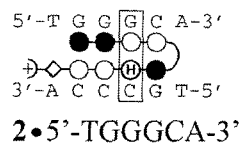
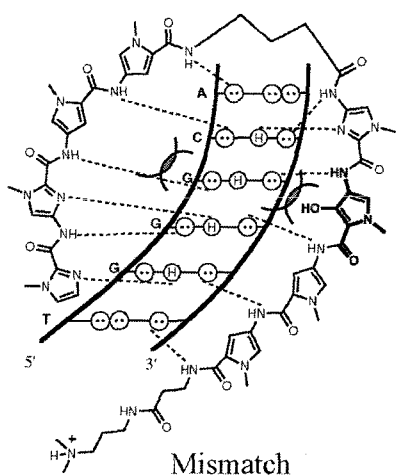
Figure 5.4 Hydrogen bonding model of the 1:1 polyamide:DNA complex formed between the eight-ring hairpin polyamide ImImPyPy- γ -ImHpPyPy- β -Dp 2 with the four sites: 5'-TGGACA-3', 5'-TGGTCA-3', 5'-TGGGCA-3', 5'-TGGCCA-3'. A circle with two dots represents the lone pair of N3 of purines and the O2 of cytosine. Two touching circles with dots represent the two lone pairs of the O2 of thymine. 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 a circle containing an H represents 3-hydroxypyrrole. The β -alanine residue is represented as an unshaded diamond. A steric clash between the DNA and polyamide is indicated by overlapped half circles.



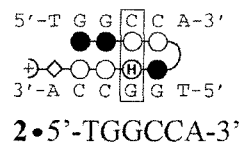
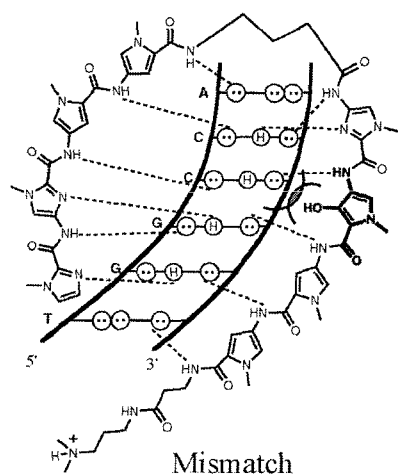
←
H-bond
Match



←
Steric
Clash



←
Steric
Clash



←
Steric
Clash

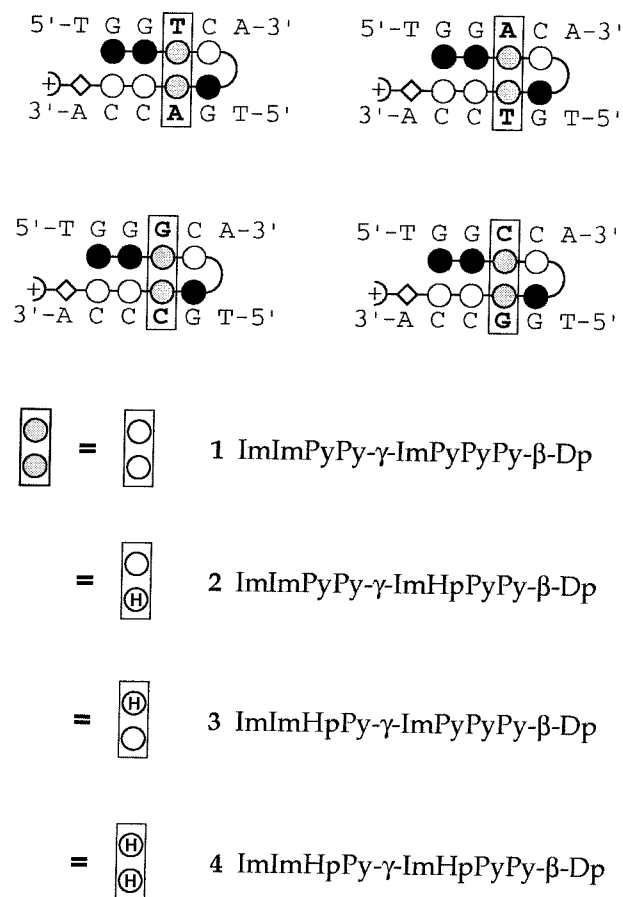


Figure 5.5 Models of the expected hairpin complexes of ImImPyPy-γ-ImPyPyPy-β-Dp **1**, ImImPyPy-γ-ImHpPyPy-β-Dp **2**, ImImHpPy-γ-ImPyPyPy-β-Dp **3**, and ImImHpPy-γ-ImHpPyPy-β-Dp **4** complexed with 5'-TGGACA-3', 5'-TGGTCA-3', 5'-TGGGCA-3', 5'-TGGCCA-3'. The four pairings of Py/Py, Py/Hp, Hp/Py, and Hp/Hp with the 4 Watson-Crick base pairs are highlighted by boxes. Gray circles may be either Py or Hp; white and black circles are pyrrole and imidazole, respectively; and circles containing an H represent 3-hydroxypyrrole. Diamonds represent β-alanine, the curved line connecting the polyamide subunits represents γ-aminobutyric acid, and the + represents the positively charged dimethylaminopropylamide tail.

Hp monomer and Boc-chemistry machine-assisted protocols for synthesis of hydroxypyrrole-imidazole-pyrrole polyamides and EDTA derivatives.

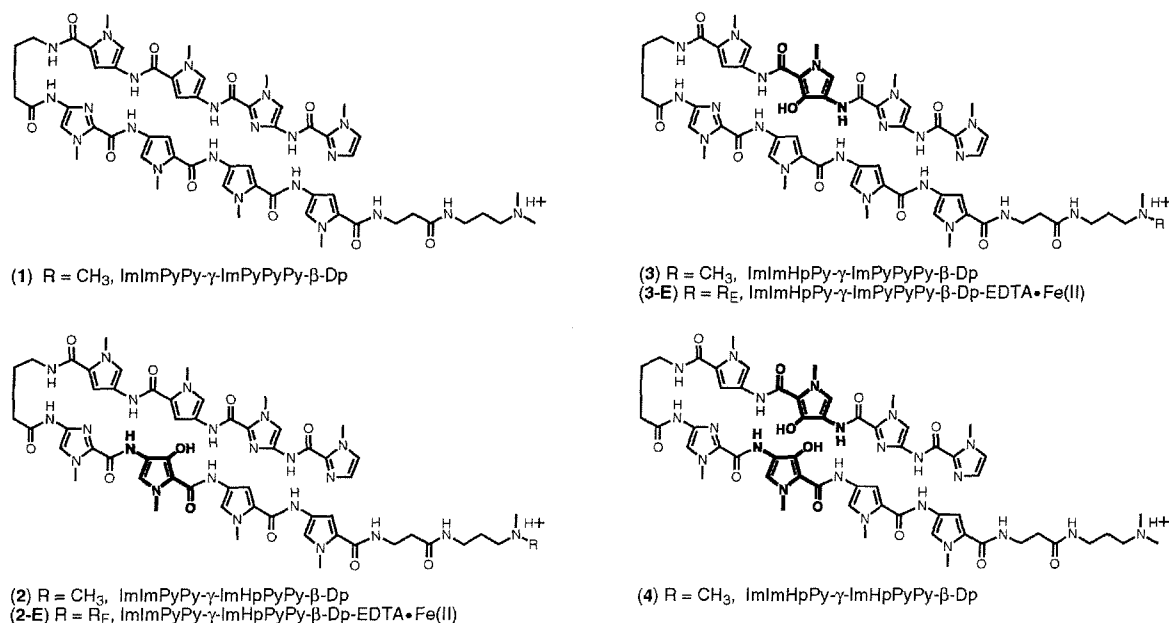


Figure 5.6 Structures of the eight-ring hairpin hydroxypyrrole-imidazole-pyrrole polyamides **1-4**.

Results

Monomer synthesis. The hydroxypyrrole mono-acid **5** may be prepared in 0.5 kg quantity using published procedures on enlarged scale.²² The appropriately protected monomer, 3-methoxypyrrole (Op) **8**, is prepared from **5** in 4 steps on 50 g scale (Figure 5.7). The 4-carboxy-3-hydroxypyrrole monoester **5** was converted to the 4-benzyl carbamate 3-hydroxypyrrole **6** by treatment with DPPA followed by reaction of the resulting isocyanate with benzyl alcohol. The 3-hydroxy group was protected as the methyl ether using methyl iodide,

and the 4-benzylcarbamate (Cbz) converted to the 4-*t*-butylcarbamate (Boc) **7** by reduction of the benzyl group with *in situ* reaction of the resulting amine with Boc anhydride. Lastly, the 4-Boc protected 3-methoxypyrrole 2-ethyl ester **7** is hydrolyzed to give the 3-O-methyl-*N*-Boc protected pyrrole 2-carboxylate **8** for use in solid phase protocols.

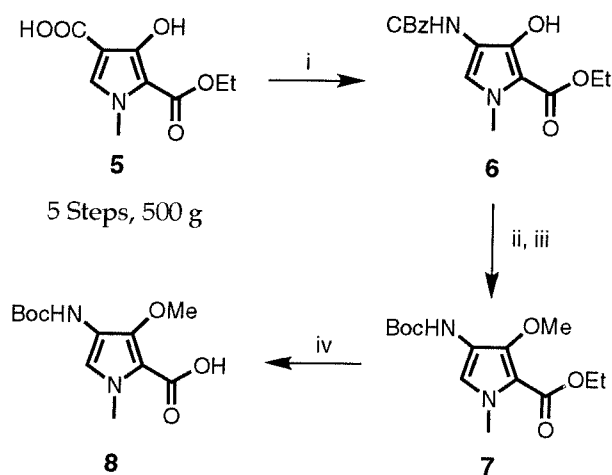


Figure 5.7 (i) a) DPPA, TEA, CH₃CN. b) benzyl alcohol. (ii) MeI, acetone, DMAP, K₂CO₃. (iii) 10% Pd/C, H₂ (1 atm), Boc₂O, DIEA, DMF. (iv) 1 M NaOH (aq), EtOH.
(CBz = benzyloxycarbonyl)

Polyamide Synthesis. Four polyamide resins, ImImPyPy- γ -ImPyPyPy- β -Pam-resin, ImImPyPy- γ -ImHpPyPy- β -Pam-resin, ImImHpPy- γ -ImPyPyPy- β -Pam-resin, and ImImHpPy- γ -ImHpPyPy- β -Pam-resin were synthesized in 18 steps from Boc- β -alanine-Pam-resin (1 g of resin, 0.2 mmol/g of substitution) using

previously described Boc-chemistry machine-assisted protocols (Figure 5.8).²¹ Hydroxypyrrole amino acid residues were introduced as orthogonal protected 3-methoxypyrrole derivatives (HBTU/DIEA). In machine synthesis protocols, 4-*N*-Boc-3-methoxypyrrole acid **8** (Boc-Op acid) was incorporated by placing the amino acid and an equivalent of HBTU in a machine synthesis cartridge. Upon automated delivery of DMF (2 ml) and DIEA (1 ml) activation occurs. 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 (dimethylamino)propylamine (55 °C, 18 h) to provide **1** and the methoxy-protected polyamides **2-Me**, **3-Me**, **4-Me**. Resin cleavage products were purified by reverse phase HPLC to provide ImImPyPy- γ -ImPyPyPy- β -Dp **1**, and the methoxy-protected polyamides ImImPyPy- γ -ImOpPyPy- β -Dp **2-Me**, ImImOpPy- γ -ImPyPyPy- β -Dp **3-Me**, and ImImOpPy- γ -ImOpPyPy- β -Dp **4-Me**. Polyamides containing 3-methoxypyrrole were subsequently deprotected by treatment with sodium thiophenoxide in DMF (100 °C, 2 h) and purified by reverse phase HPLC to provide ImImPyPy- γ -ImHpPyPy- β -Dp **2**, ImImHpPy- γ -ImPyPyPy- β -Dp **3**, and ImImHpPy- γ -ImHpPyPy- β -Dp **4**.

For synthesis of analogues modified with EDTA at the carboxy terminus, the resin was cleaved with 3,3'-diamino-*N*-methyldipropylamine (55 °C, 18 h) and purified by reversed phase HPLC to provide **2-Me-E** and **3-Me-E**, which affords a free primary amine group at the C-terminus suitable for postsynthetic modification. The polyamide amines were treated with an

Figure 5.8 Solid phase synthetic scheme for ImImPyPy- γ -ImHpPyPy- β -Dp starting from commercially available Boc- β -Pam-resin: (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-Op acid, HBTU, DIEA; (vii) 80% TFA/DCM, 0.4 M PhSH; (viii) Boc-Im-OH, HBTU, DIEA; (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-Im-OH, HBTU, DIEA; (xvii) 80% TFA/DCM, 0.4 M PhSH; (xviii) imidazole-2-carboxylic acid, HBTU, DIEA; (xix) *N,N*-dimethylaminopropylamine or 3,3'-diamino-*N*-methyldipropylamine, 55 °C; (xx) EDTA-dianhydride, DMSO/NMP, DIEA, 55 °C; (xxi) 0.1 M NaOH; (xxii) NaSPh, DMF, 100 °C.

excess of EDTA-dianhydride (DMSO/NMP, DIEA, 55 °C, 15 min.) and isolated by HPLC to afford the EDTA-modified polyamides ImImPyPy- γ -ImOpPyPy- β -Dp-EDTA and ImImOpPy- γ -ImPyPyPy- β -Dp-EDTA. Treatment with sodium thiophenoxide in DMF (100 °C, 2 h) and HPLC purification provides the respective C-terminal EDTA derivatives: ImImPyPy- γ -ImHpPyPy- β -Dp-EDTA **2-E** and ImImHpPy- γ -ImPyPyPy- β -Dp-EDTA **3-E**. The 8-ring hairpin polyamides here are soluble in aqueous solution at concentration of ≤ 10 mM.

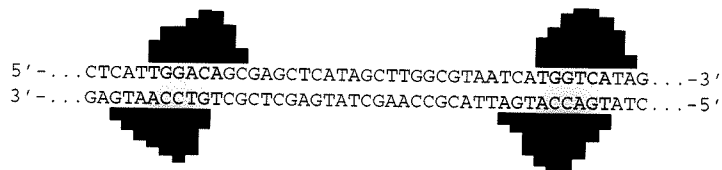
Binding site size and location by MPE•Fe(II) footprinting. MPE•Fe(II) footprinting²³ on 3' and 5'-³²P end labeled 250-bp restriction fragments (25 mM tris-acetate, 10 mM NaCl, 100 μ M bp calf thymus DNA, 5 mM DTT, pH 7.0 and 22 °C) demonstrate that the polyamides bind and discriminate the two 6 base pair sites 5'-TGGACA-3' and 5'-TGGTCA-3' (Figure 5.9). At 1 μ M concentration polyamide **1** bound both sites, 5'-TGGACA-3' \approx 5'-TGGTCA-3'; whereas polyamides **2** and **3** selectively targeted one of the 6 base pair sites. At 1 μ M concentration polyamide **2** bound 5'-TGGACA-3' $>$ 5'-TGGTCA-3', while compound **3** bound 5'-TGGACA-3' $<$ 5'-TGGTCA-3' as expected. No binding was observed for polyamide **4**, containing a Hp/Hp pair. The sizes of the asymmetrically 3'-shifted footprint cleavage protection patterns for **1-3** are consistent with the polyamides binding 6 base pair sites in the minor groove.

Figure 5.9 (top): Results from MPE•Fe(II) footprinting of ImPyPyPy-β-Dp **1**, ImImPyPy-γ-ImHpPyPy-β-Dp **2**, ImImHpPy-γ-ImPyPyPy-β-Dp **3**. Boxes represent equilibrium binding sites determined by the published model. Only sites that were quantitated by DNase I footprint titrations are boxed. Bar heights are proportional to the relative protection from cleavage at each band.

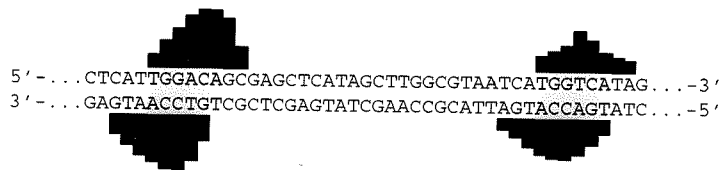
(bottom) MPE•Fe(II) footprinting on a 3'-³²P-labeled 250-bp *Eco* RI/*Pvu* II restriction fragment from plasmid pJK6. The 5'-TGGTCA-3' and 5'-TGGACA-3' sites are shown on the right side of the autoradiograms: lane 1, A reaction; lane 2, MPE•Fe(II) standard; lane 3, 100 nM **1**, **2**, or **3**; lane 4, 1 μM **1**, **2**, or **3**; lane 5, 10 μM **1**, **2**, or **3**; lane 6, intact DNA. All lanes contain 15 kcpm 3'-radiolabeled DNA and 25 mM Tris-acetate buffer (pH 7.0), 10 mM NaCl, and 100 μM/base pair calf thymus DNA.

32P *Eco*RI 302 bp *Pvu* II

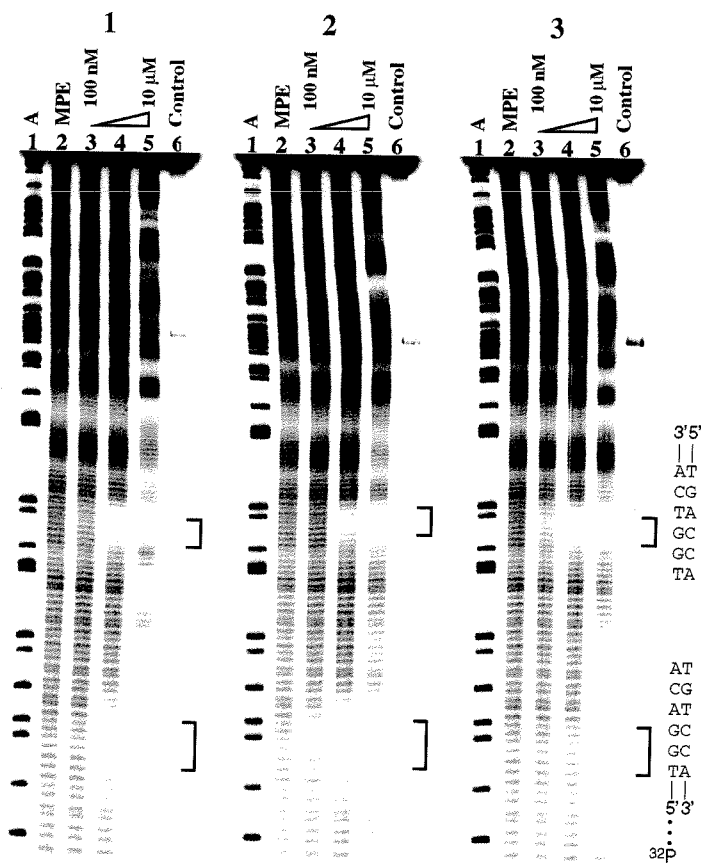
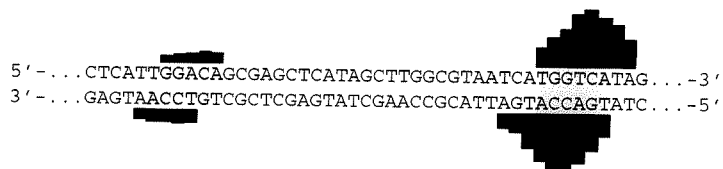
1 ImImPyPy- γ -ImPyPyPy- β -Dp (1 μ M)



2 ImImPyPy- γ -ImHpPyPy- β -Dp (1 μ M)



3 ImImHpPy- γ -ImPyPyPy- β -Dp (1 μ M)



Binding orientation by affinity cleavage. Affinity cleavage experiments²⁴ using hairpin polyamides modified with EDTA•Fe(II) at the C-terminus were used to determine polyamide binding orientation and stoichiometry. Affinity cleaving experiments were performed on 3' and 5'-³²P end-labeled 250-base restriction fragments (25 mM tris-acetate, 20 mM NaCl, 100 μ M bp calf thymus DNA, pH 7.0, 22 °C) (Figure 5.10). The cleavage patterns for ImImPyPy- γ -ImHpPyPy- β -Dp-EDTA•Fe(II) **2-E** and ImImHpPy- γ -ImPyPyPy- β -Dp-EDTA•Fe(II) **3-E** are 3'-shifted, consistent with minor groove occupancy. A single cleavage locus proximal to the 5'-side of both the 5'-TGGTCA-3' and 5'-TGGACA-3' binding sites confirmed both polyamides bound each site with a single discrete orientation. The observation of a single cleavage locus is consistent only with an oriented 1:1 complex in the minor groove of DNA and rules out dimeric overlapped or extended binding motifs. A 1:1 oriented but extended motif would require at least an 8 base pair binding site which is inconsistent with the high resolution MPE footprinting data on both sites.

Energetics by quantitative DNase I footprinting. MPE•Fe(II) footprinting combined with affinity cleavage experiments indicated that polyamides **1-3** bound the designated 6 base pair target site as a 1:1 hairpin complex in the minor groove. Quantitative DNase I footprint titrations¹³ (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂, 5 mM CaCl₂, pH 7.0, and 22 °C) were performed to determine the equilibrium association constants (K_a) of each eight-ring

Figure 5.10 (a) Affinity cleavage experiments on the 3'-³²P-labeled pJK6-bp *Eco*RI/*Pvu* II restriction fragment derived from the plasmid pJK6. The 5'-TGGTCA-3' and 5'-TGGACA-3' are shown on the right of each storage phosphor autoradiogram obtained from the 8% denaturing polyacrylamide gels used to separate the fragments generated by affinity cleaving experiments: Lane 1, A reaction; lanes 2-6, digestion products obtained in the presence of **2-E** and **3-E**: 6.5 μ M, 1.0 μ M, 100 nM, 10 nM, and 1.0 nM polyamide, respectively, lane 7, intact DNA. All reactions contain 20 kcpm 3'-³²P restriction fragment, 25 mM tris-acetate (pH 7.0), 20 mM NaCl, 100 μ M bp calf thymus DNA. (b) Affinity cleavage patterns with ball and stick models of the eight-ring EDTA•Fe(II) analogues and an illustration of the 250 base pair restriction fragment with the position of the sequence indicated. Line heights are proportional to the relative cleavage intensities at each base pair. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Circles containing an H represent hydroxypyrrole. Nonshaded diamonds represent β -alanine. The boxed Fe denotes the EDTA•Fe(II) cleavage moiety.

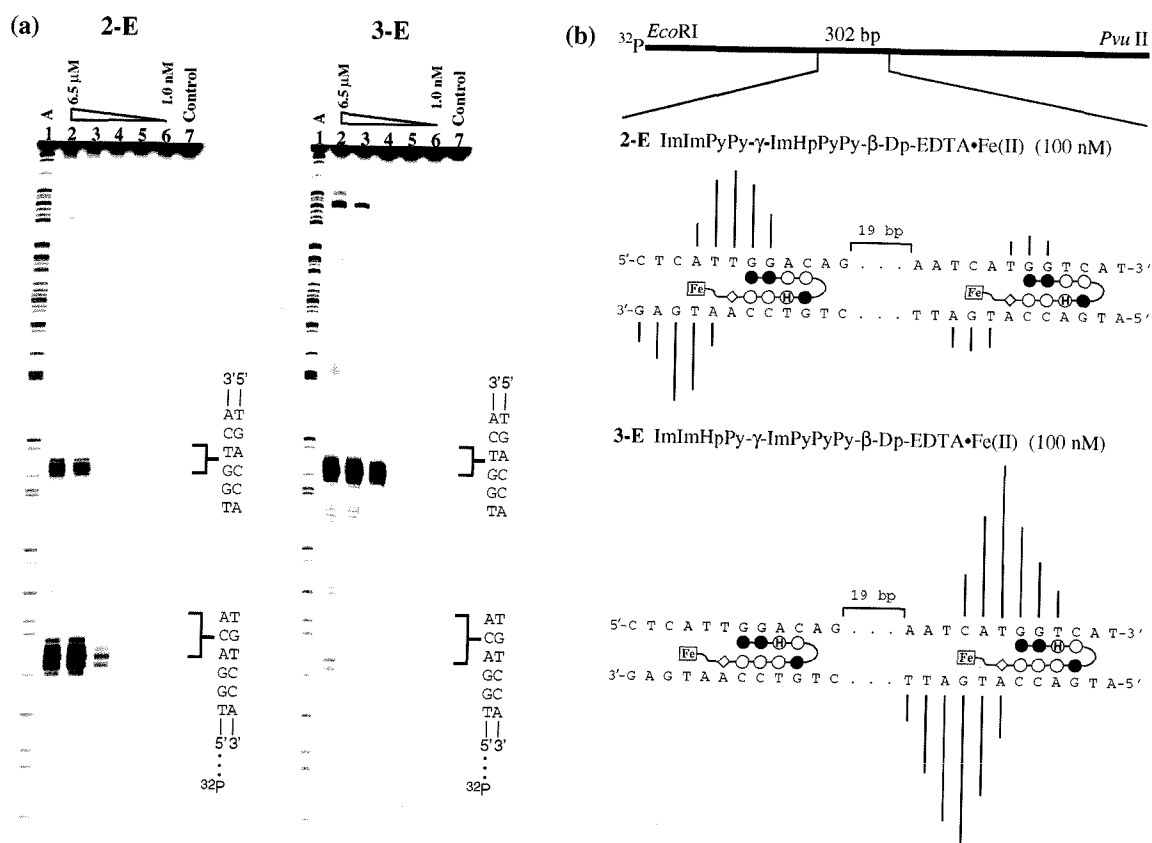


Table 5.3 Equilibrium association constants*

Polyamide†	5'-TGGTCA-3'	5'-TGGACA-3'	5'-TGGGCA-3'	5'-TGGCCA-3'
1 Py/Py	 $K_a = 1.3 \times 10^{10} \text{ M}^{-1}$	 $K_a = 6.7 \times 10^9 \text{ M}^{-1}$	 $K_a \leq 1 \times 10^7 \text{ M}^{-1}$	 $K_a \leq 1 \times 10^7 \text{ M}^{-1}$
2 Py/Hp	 $K_a = 6.7 \times 10^7 \text{ M}^{-1}$	 $K_a = 1.2 \times 10^9 \text{ M}^{-1}$	 $K_a \leq 1 \times 10^7 \text{ M}^{-1}$	 $K_a \leq 1 \times 10^7 \text{ M}^{-1}$
3 Hp/Py	 $K_a = 2.1 \times 10^9 \text{ M}^{-1}$	 $K_a = 2.7 \times 10^7 \text{ M}^{-1}$	 $K_a \leq 1 \times 10^7 \text{ M}^{-1}$	 $K_a \leq 1 \times 10^7 \text{ M}^{-1}$
4 Hp/Hp	 $K_a \leq 1 \times 10^7 \text{ M}^{-1}$	 $K_a \leq 1 \times 10^7 \text{ M}^{-1}$	 $K_a \leq 2 \times 10^7 \text{ M}^{-1}$	 $K_a \leq 2 \times 10^7 \text{ M}^{-1}$

*The reported association constants are the average values obtained from three DNase I footprint titration experiments. The standard deviation for each data set is less than 15% of the reported number. Assays were carried out in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂ at pH 7.0 and 22 °C.

†Ring pairing opposite T•A and A•T in the fourth position.

hairpin polyamide for the four sites 5'-TGGTCA-3', 5'-TGGACA-3', 5'-TGGGCA-3', and 5'-TGGCCA-3' (Table 5.3). The 5'-TGGTCA-3' site was bound by the polyamides with decreasing affinity, ImImPyPy-γ-ImPyPyPy-β-Dp 1 ($K_a = 1.3 \times 10^{10}$) \approx ImImHpPy-γ-ImPyPyPy-β-Dp 3 ($K_a = 2.1 \times 10^9$) \gg ImImPyPy-γ-ImHpPyPy-β-Dp 2 ($K_a = 6.7 \times 10^7$) \approx ImImHpPy-γ-ImHpPyPy-β-Dp 4 ($K_a \leq 1 \times 10^7$). The 5'-TGGACA-3' site was bound with decreasing affinity ImImPyPy-γ-ImPyPyPy-β-Dp 1 ($K_a = 6.7 \times 10^9$) \approx ImImPyPy-γ-ImHpPyPy-β-Dp 2 ($K_a = 1.2 \times 10^9$) \gg ImImHpPy-γ-ImPyPyPy-β-Dp 3 ($K_a = 2.7 \times 10^7$) $>$ ImImHpPy-γ-ImHpPyPy-β-Dp 4 ($K_a \leq 1 \times 10^7$). Equilibrium association constants for each of the four polyamides placing a ring pairing opposite G•C or C•G (5'-TGGGCA-

3', and 5'-TGGCCA-3', respectively) decreased affinity over 200-fold ($K_a \leq 1 \times 10^7$ for polyamides **1-3** and $K_a \leq 2 \times 10^7$ for **4**). Remarkably, association constants for polyamides **2** and **3** varied at each of the four sites by over an order of magnitude indicating a sensitivity to a single atomic substitution within each ring pairing (Table 5.3).

Discussion

The Py/Py pair. The polyamide ImImPyPy- γ -ImPyPyPy- β -Dp **1** (containing a Py/Py pair) is degenerate binding opposite both T•A (5'-TGGTCA-3') and A•T (5'-TGGACA-3') with only a 2-fold selectivity; however, placement opposite a G•C (5'-TGGGCA-3') or C•G (5'-TGGCCA-3') base pair decreases affinity by > 600-fold. The discrimination of A•T/T•A base pairs from G•C/G•C base pairs by Py/Py is likely to be due to the exocyclic amine groups of guanine which present a steric hindrance to deep polyamide binding in the minor groove. The Py/Py pairing provides a baseline for which to evaluate hydroxypyrrole ring pairings.

The Hp/Py pair. Among the four ligands ImImHpPy- γ -ImPyPyPy- β -Dp **3** (Hp/Py pair italicized) bound the 5'-TGGTCA-3' site with the highest specificity. The 5'-TGGACA-3' mismatch site is bound with 77-fold lower affinity indicating that Hp/Py is the optimal ring pairing for recognition of a T•A base pair. The sequence specificity of the Hp/Py pairing was underscored

by the > 200-fold reduced affinity when the Hp/Py pair was placed opposite G•C/C•G base pairs demonstrating a 1:1 correspondence between a Hp/Py pairing and T•A.

The Py/Hp pair. The polyamide ImImPyPy- γ -ImHpPyPy- β -Dp **2** bound with 18-fold enhanced affinity opposite an A•T base pair (5'-TGGACA-3'), relative to a T•A base pair (5'-TGGTCA-3'). Placement of a Py/Hp pair opposite G•C (5'-TGGGCA-3') or C•G (5'-TGGCCA-3') disfavors binding by over 120-fold. The selectivity observed demonstrates that a Py/Hp pairing efficiently discriminates A•T from the other three Watson-Crick base pairs forming the second hydroxypyrrole ring pairing breaking polyamide A,T degeneracy in the minor groove (Table 5.3).

The Hp/Hp pair. The hairpin ImImHpPy- γ -ImHpPyPy- β -Dp **4** (Hp/Hp pairing italicized) bound with > 500-fold reduced affinity opposite all four Watson-Crick base pairs relative to polyamide 1. The expected reduction in binding energy by a potential steric clash of the 3-hydroxy group and the floor of the minor groove provides further evidence for shape selective recognition of an asymmetric cleft in the minor groove.²⁰ The dramatic reduction in affinity by a single Hp/Hp pair demonstrates that Hp cannot be accommodated opposite A, G, or C (Table 5.4).

A ternary code has been developed to correlate DNA sequence with polyamide sequence composition. The results described here demonstrate the

sequence specificity of the four binary combinations of pyrrole and hydroxypyrrole; the Hp/Py pairing recognizes T•A, but is disfavored when placed opposite G•C, C•G, and A•T; the Py/Hp pairing targets A•T, but is disfavored opposite G•C, C•G, and T•A; Py/Py pairing is degenerate recognizing T•A/A•T base pairs but is disfavored opposite G•C/C•G, and a Hp/Hp pairing is disfavored with all four base pairs, breaking any potential degeneracy for recognition by preventing unlinked polyamide dimers from binding in certain slipped motifs (Table 5.4). These results will guide further second generation design of hydroxypyrrole-imidazole-pyrrole polyamides.

Table 5.4 T•A/A•T specificity of Py and Hp pairs*

Pair	T•A	A•T	G•C	C•G
Py/Py	+	+	-	-
Py/Hp	-	+	-	-
Hp/Py	+	-	-	-
Hp/Hp	-	-	-	-

* favored (+), disfavored (-)

A STRUCTURAL BASIS FOR RECOGNITION OF A•T AND T•A BASE PAIRS IN THE MINOR GROOVE OF B-DNA

Before the first structure of a molecule bound to DNA had been determined, specific recognition of double helical B-form DNA was predicted to occur primarily in the major, rather than the minor groove.² This proposal was based upon the observation that for A,T base pairs, the hydrogen bond acceptors at N3 of adenine and O2 of thymine are almost symmetrically placed in the minor groove (Figure 5.1).² Subsequent structures of DNA-binding domains co-crystallized with DNA supported this idea, since most of the specific contacts were made with the major groove.³ The principle that the major groove is a better candidate for sequence-specific recognition than the minor groove²⁵ continues to provide the basis for strategies to decipher rules for protein-DNA recognition. Although there has been remarkable progress in the design of zinc fingers to recognize the major groove,¹⁸ no protein structure motif has been identified which provides an α -amino acid-base pair code for the minor groove.

The discovery of pairing rules for dimers of minor groove binding polyamides containing three aromatic amino acids, pyrrole (Py), imidazole (Im), and 3-hydroxypyrrole (Hp), affords a recognition code for the discrimination of all four Watson-Crick base pairs in the minor groove of DNA.¹⁹ The side-by-side pairing of the aromatic residues in the polyamidedimer determines the DNA sequence recognized. An Im/Py pair

distinguishes G•C from C•G and both of these from A•T/T•A,¹ and the structural basis of this discrimination has been determined through x-ray structure studies.¹⁰ However, a structural understanding of how an Hp/Py pair distinguishes T•A from A•T, and both from G•C/C•G, has yet to be established. To address this question, the co-crystal structure of a polyamide of sequence ImHpPyPy-β-Dp **9** (Figure 5.11) bound as a dimer to a self-complementary ten base pair oligonucleotide containing all four Watson-Crick base pairs, 5'-CCAGTACTGG-3' (binding site in bold), has been determined.

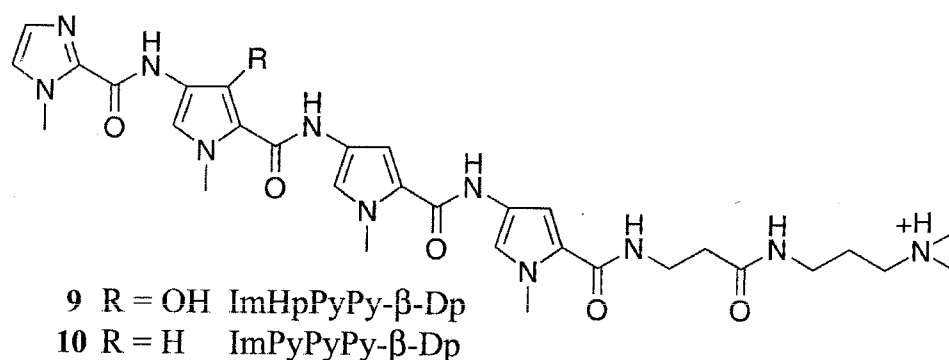
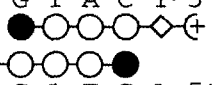
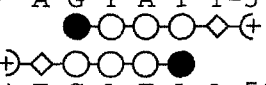
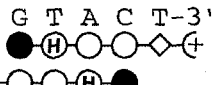
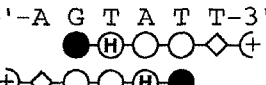


Figure 5.11 Chemical structures of the polyamides ImHpPyPy-β-Dp **9** and ImPyPyPy-β-Dp **10**.

The structure of the polyamide ImPyPyPy-β-Dp **10**, containing Py/Py pairs which do not distinguish T•A from A•T,²⁶ bound to the same duplex was solved for comparison. In addition, the equilibrium association constants (K_a) for polyamides ImHpPyPy-β-Dp **9** and ImPyPyPy-β-Dp **10**

(Figure 5.11) were determined using DNase I quantitative footprint titration experiments in order to correlate energetic analysis of minor groove recognition with three-dimensional structural information (Tables 5.5, 5.6).

Table 5.5 Equilibrium Association Constants (K_a)^{a,b}

Match	Mismatch	Specificity ^c
5'-A G T A C T-3'  3'-T C A T G A-5' $2.1 (\pm 0.1) \times 10^7 \text{ M}^{-1}$	5'-A G T A T T-3'  3'-T C A T A A-5' $1.4 (\pm 0.3) \times 10^6 \text{ M}^{-1}$	15-fold
5'-A G T A C T-3'  3'-T C A T G A-5' $2.9 (\pm 0.7) \times 10^6 \text{ M}^{-1}$	5'-A G T A T T-3'  3'-T C A T A A-5' $7.6 (\pm 0.6) \times 10^5 \text{ M}^{-1}$	4-fold

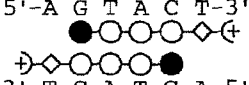
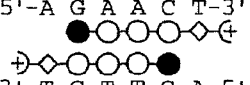
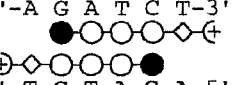
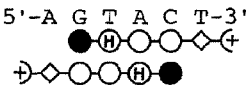
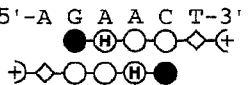
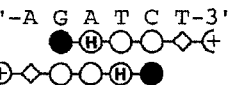
^aValues reported are the mean values obtained from three DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses.

^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.

^cSpecificity is calculated as K_a (5'-AGTACT-3') / K_a (5'-AGTATT-3').

The polyamides ImHpPyPy-β-Dp **9** and ImPyPyPy-β-Dp **10** were synthesized by solid-phase methods using Boc-protected 3-methoxypyrrole, imidazole, and pyrrole amino acids. The identity and purity of the polyamides were confirmed by ¹H NMR and matrix-assisted laser-desorption ionization time-of-flight mass spectrometry. Synthetic deoxyoligonucleotides were synthesized with the 5'-trityl on, and purified with two rounds of reverse phase on a C8 column.

Table 5.6 Equilibrium Association Constants (K_a)^{a,b}

GTAC	GAAC	GATC	Specificity
5'-A G T A C T-3'  $2.5 \times 10^7 \text{ M}^{-1}$	5'-A G A A C T-3'  $4.0 \times 10^6 \text{ M}^{-1}$	5'-A G A T C T-3'  $4.0 \times 10^6 \text{ M}^{-1}$	6-fold
5'-A G T A C T-3'  $4.0 \times 10^6 \text{ M}^{-1}$	5'-A G A A C T-3'  $3.0 \times 10^5 \text{ M}^{-1}$	5'-A G A T C T-3'  $1.0 \times 10^5 \text{ M}^{-1}$	13-40-fold

^aThe 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.

^bSpecificity is calculated as K_a (5'-AGTACT-3') / K_a (5'-AGAACT-3') and K_a (5'-AGTACT-3') / K_a (5'-AGATCT-3').

DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂, pH 7.0, 22 °C) were performed to determine the equilibrium association constants K_a for recognition of bound sites. The hydroxypyrrole-containing polyamide, ImHpPyPy-β-Dp **9**, prefers the match site 5'-AGTACT-3' ($K_a = 2.9 (\pm 0.7) \times 10^6 \text{ M}^{-1}$) over the single base pair mismatch site 5'-AGTATT-3' ($K_a = 7.6 (\pm 0.6) \times 10^5 \text{ M}^{-1}$) by a factor of 4 (Table 5.5). Polyamide **9** has increased A•T/T•A specificity, preferring the match site 5'-AGTACT-3' ($K_a = 2.9 (\pm 0.7) \times 10^6 \text{ M}^{-1}$) over the mismatch sites 5'-AGAACT-3' ($K_a = 3.0 \times 10^5 \text{ M}^{-1}$) and 5'-AGATCT-3' ($K_a = 1.0 \times 10^5 \text{ M}^{-1}$) by a factor of 10- and 29-fold respectively (Table 5.6). The control polyamide ImPyPyPy-β-Dp **10**, containing only Py/Py pairs which are approximately degenerate for A•T and T•A base pairs, binds to all three sites 5'-AGTACT-3', 5'-AGAACT-3', and 5'-

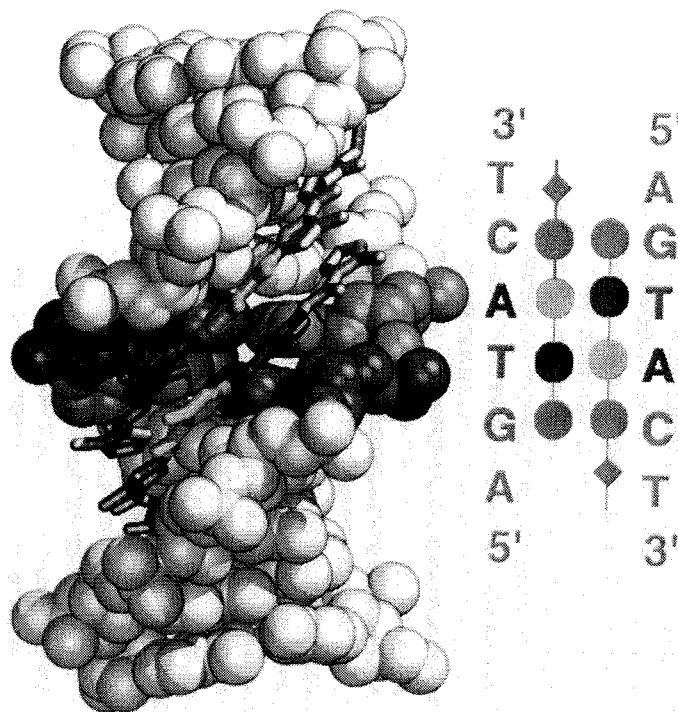


Figure 5.12. Structure of $(\text{ImHpPyPy-}\beta\text{-Dp})_2 \cdot 5'\text{-CCAGTACTGG-}3'$. Adenosine is very dark gray and thymidine is medium gray. A schematic DNA-binding model is shown to the side, with the residues of the polyamide indicated by filled circles. Im/Py pairs = medium gray, Hp = black, Py = light gray. The overall structure of $(\text{ImPyPyPy-}\beta\text{-Dp})_2 \cdot 5'\text{-CCAGTACTGG-}3'$ is similar.

AGATCT-3' within a factor of six (Table 5.6). In addition, ImPyPyPy- β -Dp 10 prefers the site 5'-AGTACT-3' ($K_a = 2.1 (\pm 0.1) \times 10^7 \text{ M}^{-1}$) over the site 5'-AGTATT-3' ($K_a = 1.4 (\pm 0.3) \times 10^6 \text{ M}^{-1}$) by at least 15-fold (Table 5.5). Within these polyamides, the replacement of a Py/Py pair by a Hp/Py pair decreases the G•C specificity of an adjacent Im/Py pair. Also, there is a loss in binding affinity of about 7-fold with the substitution of two Py/Py pairs in polyamide

10 by two Hp/Py pairs to give polyamide **9**. The structural studies described herein provide insight into how hydroxypyrrole-imidazole-pyrrole polyamides discriminate T•A from A•T base pairs via hydrogen bond formation and differential destabilization.

In both the ImHpPyPy and ImPyPyPy structures, the polyamides bind as antiparallel dimers centered over the target AGTACT sequence in the minor groove of a B-form DNA duplex (Figure 5.12). The N to C-terminal orientation of each fully overlapped polyamide is parallel to the adjacent 5' to 3' strand of DNA, consistent with previous chemical¹ and structural studies of polyamide dimers.¹⁰

Although the functional groups of adenine and thymine are very similar in the minor groove, the number of lone pairs on the hydrogen bond acceptors is different: a thymine-O2 has two free lone pairs, whereas an adenine-N3 has only one (Figure 5.1). The amide nitrogens of the ligand form hydrogen bonds with the purine-N3 (A or G) or pyrimidine-O2 (T or C). Therefore, the hydrogen bond potential of adenine-N3 is filled when an imidazole-pyrrole polyamide is bound, but the thymine-O2 has the capacity to accept an additional hydrogen bond. We find that both the hydroxyl group of the Hp, and the amide-NH of the preceding residue, form hydrogen bonds with the target thymine-O2 of the adjacent DNA strand (Figure 5.13). A similar interaction between the Hp and the adenine-N3 would be impossible without the loss of the hydrogen bond from the preceding amide-NH.

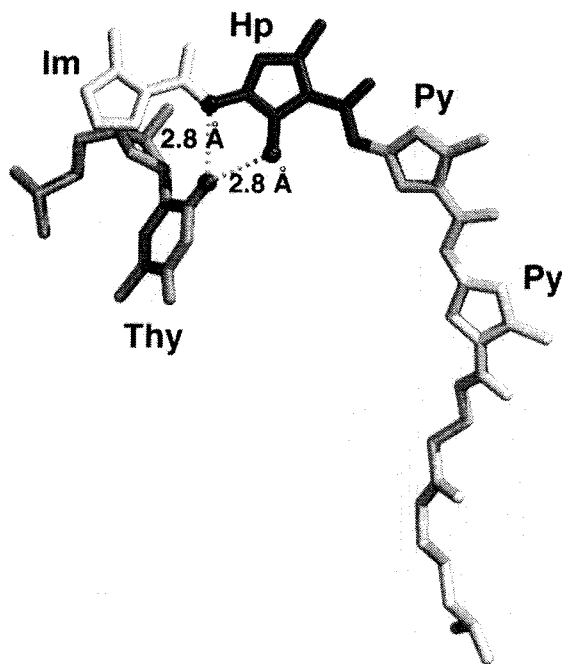


Figure 5.13 Diagram of the hydrogen bonds between the hydroxyl and the preceding amide and the two lone pairs of thymine-O2.

In addition to the difference in number of lone pairs of the adenine-N3 versus thymine-O2, adenine is also distinguished from thymine by a bulkier aromatic ring. Although the adenine-C2-H does not protrude into the minor groove like the guanine exocyclic amine, the adenine-C2-H does not protrude into the minor groove like the guanine exocyclic amine, the additional carbon results in an asymmetric cleft in the minor groove of a T•A base pair (Figure 5.1).¹² The adenine-C2 of the ImHpPyPy structure contacts the Hp hydroxyl. Modeling the target thymine as an adenine reveals that the C2 carbon of a mismatch 'adenine' opposite an Hp residue would sterically overlap the hydroxyl by 1-2 Å (depending on the hydrogen positions). Shape

selective recognition of the asymmetric cleft is the second feature that allows the Hp/Py pair to discriminate T•A from A•T.

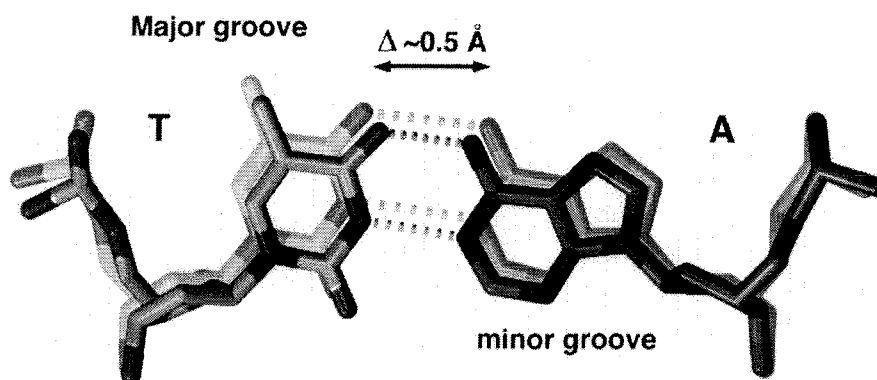


Figure 5.14 T•A base pair of the ImPyPyPy structure superimposed on the corresponding base pair from the ImHpPyPy structure, showing the slight elongation of the hydrogen bonds of the target T•A base pair.

In both structures, the oligonucleotides have standard B-DNA features, 35° twist, 3.4 Å rise per residue, and C2'-endo sugar pucker. However, the oligonucleotide structure deviates from ideal B-form by having a strong propeller twist and opening of the target T•A base pairs. The Hp/Py pairs induce a change in the T•A base pairs from no shear (displacement between the bases in the base pair, perpendicular to the helix axis) to a large positive shear. The movement of the bases past one another may result from the Hp-O6 contact with the adenine-C2 pressing the adenine of the target base pair back into the major groove. The increased displacement between the bases stretches the Watson-Crick hydrogen bonds between them by 0.5 Å, on

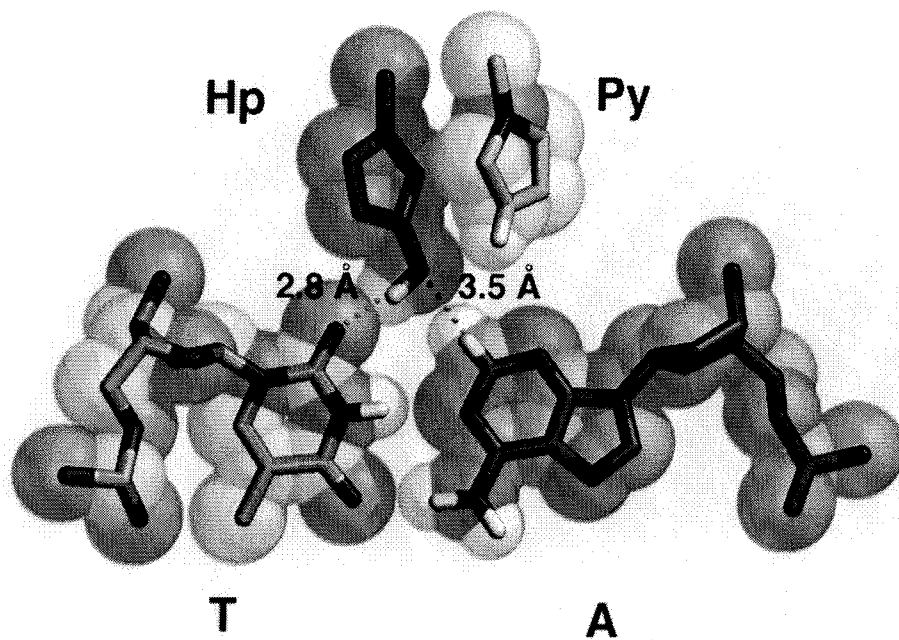


Figure 5.15 Interaction of Hp/Py pair with T•A base pair.

average (Figure 5.14 and 5.15). Although the specificity of hydroxypyrrole-containing polyamides is greatly increased for T•A over A•T, the affinities are slightly reduced relative to the pyrrole counterparts (Table 5.5). For example, ImHpPyPy- β -Dp and ImPyPyPy- β -Dp bind a 5'-AGTACT-3' site with equilibrium association constants of $2.9 (\pm 0.7) \times 10^6 \text{ M}^{-1}$ and $2.1 (\pm 0.1) \times 10^7 \text{ M}^{-1}$ respectively, a 7-fold difference. The energetic penalty due to the partial 'melting' of the target T•A base pairs could account for the 1.2 kcal/mol reduction in binding affinity.²⁷

The hydrogen bonds between the amides of each ImPyPyPy polyamide and the purine-N3 or pyrimidine-O2 of the adjacent DNA strand are

maintained for the ImHpPyPy polyamide (Figure 5.16). However, the hydrogen bonds between the DNA and the ImHpPyPy amides are longer for the residues that follow the Hp than those observed for the ImPyPyPy complex. The hydroxyl forms an intramolecular hydrogen bond with the following amide, causing the hydrogen bond of that amide with the adenine-N3 to become bifurcated and therefore weaker. This may be an additional source of the slightly decreased affinity of the Hp-containing polyamides relative to the pyrrole counterparts.

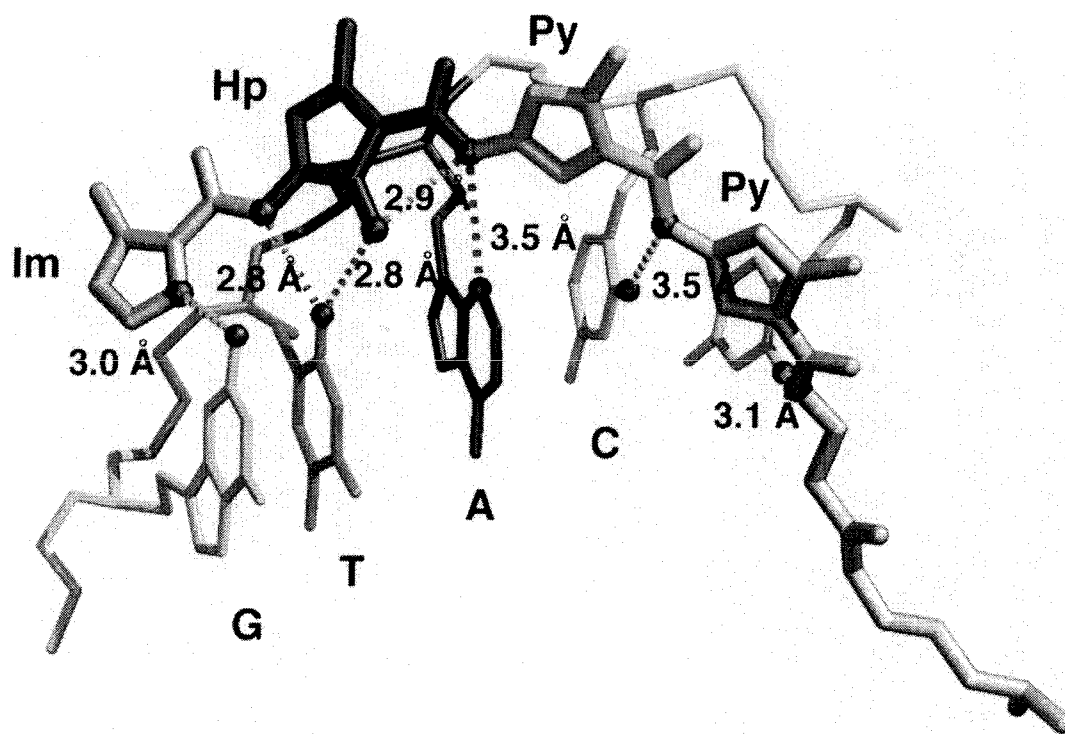


Figure 5.16 Crystal structure reveals that all potential aromatic amino acid – base pair H-bonds are formed.

These structural studies have established how a designed ligand can predictably discriminate A•T from T•A in the minor groove, utilizing the double hydrogen bond acceptor potential of thymine-O2 and the asymmetry of the adenine-C2 cleft. The structure eliminates the possibilities that a bulky substitution at the pyrrole 3-position might 1) sterically clash with the thymine-O2,²⁸ or 2) cause a gross distortion of the DNA duplex.²⁹

Experimental Section.

Polyamide synthesis performed by Eldon Baird. Polyamide deprotection reactions were performed by Sarah White.¹⁹⁻²⁰ Monomer synthesis was by Jason Szewczyk and Jim Turner.¹⁹⁻²⁰ Gel electrophoresis and footprinting assays were performed by Sarah White¹⁹⁻²⁰ and Jason Szewczyk. All crystal growing and structure analysis was performed by Clara Kielkopf.²⁶

Materials. DNA restriction fragment labeling protocols, MPE•Fe(II) footprinting, affinity cleaving, DNase I footprinting, determination of equilibrium association constants, and quantitation by storage phosphor autoradiography were as previously described. Dicyclohexylcarbodiimide (DCC), Hydroxybenzo-triazole (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) were purchased from Peptides International. *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. Boc- γ -aminobutyric acid was from NOVA Biochem, dichloromethane (DCM) and triethylamine (TEA) was reagent grade from EM, thiophenol (PhSH), dimethylaminopropylamine from Aldrich, trifluoroacetic acid (TFA) 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. and were used for filtration of DCU. 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.³⁰ ^1H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz in $\text{DMSO-}d_6$, 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. 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 C_{18} , Microsorb MV, 5 μm , 300 \times 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 \times 100 mm, 100 μm C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. 18MW water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μm filtered.

Monomer Synthesis. Ethyl 4-[(benzyloxycarbonyl)amino]-3-hydroxy-1-methylpyrrole-2-carboxylate (6). Pyrrole acid 5 (125 g, 586.85 mmol) was

dissolved in 587 mL acetonitrile under argon. Triethylamine (59.40 g, 587 mmol, 81.87 ml) was added, followed by diphenylphosphorylazide (161.54 g, 587 mmol, 126.2 ml). The mixture was refluxed for 4.75 hours, followed by addition of benzyl alcohol (562 ml) and reflux continued for 17 hours. The solution was allowed to cool and the volatiles removed *in vacuo*. The black tar was partitioned between 1:1 water:ethyl acetate and the layers separated. The aqueous phase was acidified to pH = 3.0 and extracted with ethyl acetate. Extractions were combined, washed with 0.1 N H₂SO₄, dried with MgSO₄, and solvent removed *in vacuo* to yield a dark red tar. The residue was then triturated with diethyl ether (10 times), decanted, dried over MgSO₄, and the solvent removed *in vacuo* to give a crude solid. The residue was absorbed onto silica and chromatographed with 4:1 hexanes : ethyl acetate to give a white solid. The solid was redissolved in acetone, filtered, and dried *in vacuo* to provide a white solid **6** (98.7 g, 311 mmol, 53%). ¹H NMR (DMSO-d₆) δ 8.73 (s, 1H), 8.31 (s, 1H), 7.31 (m, 5H), 6.96 (s, 1H), 5.08 (s, 2H), 4.21 (q, 2H, J = 7.1 Hz), 3.66 (s, 3H), 1.25 (t, 3H, J = 7.1 Hz); MS *m/e* 319.163 (M+H 319.122 calcd for C₁₆H₁₈N₂O₅).

Ethyl 4-[(*tert*-butoxycarbonyl)amino]-3-methoxy-1-methylpyrrole-2-carboxylate (7). Benzylhydroxypyrrole **6** (98.7 g, 311 mmol) was dissolved in 810 mL acetone. Anhydrous K₂CO₃ (86.0 g, 622 mmol) was added and the solution allowed to stir for ~ 15 minutes. Methyl iodide (44.14g, 311 mmol, 19.36 ml) was then added followed by dimethylaminopyridine (3.50 g, 31.1 mmol). Progress of the reaction was monitored by tlc with additional methyl iodide

(0.25 eq) added every two hours, as needed, until 100% conversion. The solid K_2CO_3 was removed by filtration and 1 L of water was added. Volatiles were removed *in vacuo* and the solution made acidic with addition of 1N H_2SO_4 to pH = 3.0. The aqueous layer was extracted with diethyl ether. The ether extractions were combined, washed with 10% H_2SO_4 , dried over MgSO_4 , and volatiles removed to give a white solid sufficiently pure to be used without further purification. The solid was redissolved in DMF (280 ml), DIEA (60g, 464 mmol, 81 ml). Boc anhydride (67.88g, 311 mmol) and 10% Pd/C (3.5 g) were added and the solution stirred under hydrogen (1 atm) for 2.1 h. The slurry was filtered through celite which was rinsed with methanol. Water (1 L) was added to the solution and volatiles removed *in vacuo*. The aqueous layer was extracted with diethyl ether, organic layers combined, washed with water, brine, and dried over MgSO_4 . Solvents removed *in vacuo* to provide a white solid **7** (65.8 g, 221 mmol, 71%) ^1H NMR (DMSO-d_6) δ 8.43 (s, 1H), 7.03 (s, 1H), 4.19 (q, 2H, J = 7.1 Hz), 3.70 (s, 3H), 3.67 (s, 3H), 1.42 (s, 9H), 1.26 (t, 3H, J = 7.1); MS *m/e* 299.161 (M+H 299.153 calcd for $\text{C}_{14}\text{H}_{22}\text{N}_2\text{O}_5$).

4-[(*tert*-butoxycarbonyl)amino]-3-methoxy-1-methylpyrrole-2-carboxylic Acid (8). The ester **7** (65.8 g, 221 mmol) was dissolved in 221 mL ethanol. NaOH (aq, 1 M, 221 ml) was added and the solution stirred for 4 days. Additional water (1 L) was added and the ethanol removed *in vacuo*. The aqueous solution was extracted with diethyl ether, acidified to pH = 2-3, and reextracted with diethyl ether. The extractions were combined, dried over MgSO_4 , and the solvent removed *in vacuo* to yield a fine white powder **8** (54.3 g, 201

mmol, 91% based on recovered SM) ^1H NMR (DMSO- d_6) δ 12.14 (s, 1H), 8.37 (s, 1H), 6.98 (s, 1H), 3.69 (s, 3H), 3.66 (s, 3H), 1.42 (s, 9H); MS m/e 293.112 (M+H) 293.104 calcd for $\text{C}_{12}\text{H}_{18}\text{N}_2\text{O}_5$).

Solid Phase Synthesis. 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 (gmol^{-1}) of the growing polyamide attached to the resin.³¹

ImImPyPy- γ -ImOpPyPy- β -Dp (2-Me) ImImPyPy- γ -ImOpPyPy- β -Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc- β -Pam-Resin (0.66 mmol/g).²¹ 3-methoxypyrrole-Boc amino acid (Boc-Op acid) was incorporated by placing the amino acid (0.5 mmol) and HBTU (0.5 mmol) in a machine synthesis cartridge. Upon automated delivery of DMF (2 mL) and DIEA (1 mL), activation occurs. A sample of ImImPyPy- γ -ImOpPyPy- β -Pam-Resin (400 mg, 0.40 mmol/gram) was placed in a glass 20 mL peptide synthesis vessel and 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. ImImPyPy- γ -ImOpPyPy- β -Dp is recovered upon lyophilization of the appropriate fractions as a white powder (101 mg, 50% recovery). UV (H_2O) λ_{max} 246, 316 (66,000); ^1H NMR (DMSO- d_6) δ 10.47 (s, 1 H), 10.36 (s, 1 H), 9.94 (s, 1 H), 9.87 (s, 1 H), 9.78 (s, 1 H), 9.23 (s, 1 H), 8.99 (s, 1 H), 8.07 (m, 3 H),

7.57 (s, 1 H), 7.51 (s, 1 H), 7.46 (s, 1 H), 7.28 (d, 1 H), 7.27 (d, 1 H), 7.22 (d, 1 H), 7.17 (d, 1 H), 7.15 (m, 2 H), 7.08 (s, 1 H), 7.03 (d, 1 H), 6.88 (m, 2 H), 4.00 (s, 6 H), 3.96 (s, 3 H), 3.84 (s, 6 H), 3.82 (m, 3 H), 3.80 (s, 9 H), 3.39 (q, 2 H, $J = 6.0$ Hz), 3.17 (q, 2 H, $J = 5.6$ Hz), 3.14 (q, 2 H, $J = 6.2$ Hz), 3.03 (quintet, 2 H, $J = 5.5$ Hz), 2.74 (s, 3 H), 2.72 (s, 3 H), 2.33 (m, 4 H), 1.75 (m, 4H); MALDI-TOF-MS (monoisotopic), 1253.5 (1253.6 calcd. for $C_{58}H_{72}N_{22}O_{11}$).

ImImPyPy- γ -ImHpPyPy- β -Dp (2) In order to remove the methoxy protecting group, a sample of ImImPyPy- γ -ImOpPyPy- β -Dp (5 mg, 3.9 mmol) was treated with sodium thiophenoxide at 100 °C for 2 h. DMF (1000 mL) and thiophenol (500 mL) were placed in a (13 x 100 mm) disposable Pyrex screw cap culture tube. A 60% dispersion of sodium hydride in mineral oil (100 mg) was slowly added. Upon completion of the addition of the sodium hydride, ImImPyPy- γ -ImOpPyPy- β -Dp (5 mg) dissolved in DMF (500 μ L) was added. The solution was agitated, and placed in a 100 °C heat block, and deprotected for 2 h. Upon completion of the reaction, the culture tube was cooled to 0 °C, and 7 mL of a 20% (wt/v) solution of trifluoroacetic acid added. The aqueous layer is separated from the resulting biphasic solution and purified by reversed phase HPLC. ImImPyPy- γ -ImHpPyPy- β -Dp is recovered upon lyophilization of the appropriate fractions as a white powder (3.2 mg, 66% recovery). UV (H_2O) λ_{max} 246, 312 (66,000); 1H NMR ($DMSO-d_6$) δ 10.37 (s, 1 H), 10.34 (s, 1 H), 10.27 (s, 1 H), 9.95 (s, 1 H), 9.81 (s, 1 H), 9.77 (s, 1 H), 9.70 (s, 1 H), 9.12 (s, 1 H), 8.05 (br s, 3 H), 7.57 (s, 1 H), 7.51 (s, 1 H), 7.47 (s, 1 H), 7.30 (d, 1 H), 7.28 (s, 1 H), 7.24 (s, 1

H), 7.17 (d, 1 H), 7.15 (d, 2 H), 7.09 (s, 1 H), 7.00 (s, 1 H), 6.90 (s, 1 H), 6.86 (s, 1 H), 4.00 (s, 6 H), 3.95 (s, 3 H), 3.84 (s, 6 H), 3.79 (s, 3 H), 3.68 (s, 6 H), 3.40 (q, 2 H, $J = 5.8$ Hz), 3.20 (q, 2 H, $J = 5.6$ Hz), 3.13 (q, 2 H, $J = 5.9$ Hz), 3.00 (quintet, 2 H, $J = 5.4$ Hz), 2.74 (s, 3 H), 2.72 (s, 3 H), 2.34 (m, 4 H), 1.76 (m, 4H); MALDI-TOF-MS (monoisotopic), 1239.6 (1239.6 calcd. for $C_{57}H_{71}N_{22}O_{11}$).

ImImOpPy- γ -ImPyPyPy- β -Dp (3-Me) ImImOpPy- γ -ImPyPyPy- β -Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc- β -Pam-Resin (0.66 mmol/g) as described for ImImPyPy- γ -ImOpPyPy- β -Dp.²¹ A sample of ImImOpPy- γ -ImPyPyPy- β -Pam-Resin (400 mg, 0.40 mmol/gram) was placed in a glass 20 mL peptide synthesis vessel and 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. ImImOpPy- γ -ImPyPyPy- β -Dp is recovered upon lyophilization of the appropriate fractions as a white powder (97 mg, 49% recovery). UV (H_2O) λ_{max} 246, 316 (66,000); 1H NMR ($DMSO-d_6$) δ 10.24 (s, 1 H), 10.14 (s, 1 H), 9.99 (s, 1 H), 9.94 (s, 1 H), 9.88 (s, 1 H), 9.4 (br s, 1 H), 9.25 (s, 1 H), 9.11 (s, 1 H), 8.05 (m, 3 H), 7.60 (s, 1 H), 7.46 (s, 1 H), 7.41 (s, 1 H), 7.23 (d, 1 H), 7.21 (d, 1 H), 7.19 (d, 1 H), 7.13 (m, 2 H), 7.11 (m, 2 H), 7.02 (d, 1 H), 6.83 (m, 2 H), 3.96 (s, 6 H), 3.90 (s, 3 H), 3.81 (m, 6 H), 3.79 (s, 3 H), 3.75 (d, 9 H), 3.33 (q, 2 H, $J = 5.4$ Hz), 3.15 (q, 2 H, $J = 5.5$ Hz), 3.08 (q, 2 H, $J = 6.0$ Hz), 2.96 (quintet, 2 H, $J =$

5.6 Hz), 2.70 (d, 6 H, $J = 4.5$ Hz), 2.32 (m, 4 H), 1.71 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1253.5 (1253.6 calcd. for $C_{58}H_{72}N_{22}O_{11}$).

ImImHpPy- γ -ImPyPyPy- β -Dp (3) A sample of ImImOpPy- γ -ImPyPyPy- β -Dp (5 mg, 3.9 mmol) was treated with sodium thiophenoxide and purified by reversed phase HPLC as described for ImImPyPy- γ -ImHpPyPy- β -Dp. ImImHpPy- γ -ImPyPyPy- β -Dp is recovered as a white powder upon lyophilization of the appropriate fractions (3.8 mg, 77% recovery). UV (H_2O) λ_{max} 246, 312 (66,000); 1H NMR ($DMSO-d_6$) δ 10.34 (s, 1 H), 10.24 (s, 1 H), 10.00 (s, 2 H), 9.93 (s, 1 H), 9.87 (s, 1 H), 9.83 (s, 1 H), 9.4 (br s, 1 H), 9.04 (s, 1 H), 8.03 (m, 3 H), 7.58 (s, 1 H), 7.44 (s, 1 H), 7.42 (s, 1 H), 7.23 (s, 1 H), 7.20 (m, 3 H), 7.12 (m, 2 H), 7.05 (d, 1 H), 7.02 (d, 1 H), 6.83 (s, 1 H), 6.79 (s, 1 H), 3.96 (s, 6 H), 3.90 (s, 3 H), 3.81 (s, 6 H), 3.79 (s, 3 H), 3.75 (d, 6 H), 3.33 (q, 2 H, $J = 5.4$ Hz), 3.14 (q, 2 H, $J = 5.4$ Hz), 3.08 (q, 2 H, $J = 6.1$ Hz), 2.99 (quintet, 2 H, $J = 5.4$ Hz), 2.69 (d, 6 H, $J = 4.2$ Hz), 2.31 (m, 4 H), 1.72 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1239.6 (1239.6 calcd. for $C_{57}H_{71}N_{22}O_{11}$).

ImImOpPy- γ -ImOpPyPy- β -Dp (4-Me) ImImOpPy- γ -ImOpPyPy- β -Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc- β -Pam-Resin (0.66 mmol/g) as described for ImImPyPy- γ -ImOpPyPy- β -Dp.²¹ A sample of ImImOpPy- γ -ImOpPyPy- β -Pam-Resin (400 mg, 0.40 mmol/gram) was placed in a glass 20 mL peptide synthesis vessel and 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. ImImOpPy- γ -ImOpPyPy- β -Dp is recovered upon lyophilization of the appropriate fractions as a white powder (97 mg, 49% recovery). UV (H₂O) λ_{max} 246, 316 (66,000); ¹H NMR (DMSO-d₆) δ 10.46 (s, 1 H), 10.00 (s, 1 H), 9.88 (s, 1 H), 9.30 (s, 1 H), 9.23 (s, 1 H), 9.13 (s, 1 H), 8.99 (s, 1 H), 8.06 (m, 3 H), 7.62 (s, 1 H), 7.50 (s, 1 H), 7.46 (s, 1 H), 7.28 (s, 1 H), 7.24 (d, 1 H), 7.22 (s, 1 H), 7.16 (s, 1 H), 7.15 (d, 2 H), 7.08 (d, 1 H), 7.03 (d, 1 H), 6.92 (s, 1 H), 6.86 (s, 1 H), 3.99 (s, 9 H), 3.96 (s, 3 H), 3.85 (s, 6 H), 3.82 (s, 3 H), 3.79 (s, 9 H), 3.15 (q, 2 H, J = 7.5 Hz), 3.05 (q, 2 H, J = 6.2 Hz), 2.98 (q, 2 H, J = 4.2 Hz), 2.84 (quintet, 2 H, J = 4.2 Hz), 2.74 (s, 3 H), 2.73 (s, 3 H), 2.34 (m, 4 H), 1.75 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1282.61 (1282.59 calcd. for C₅₉H₇₄N₂₂O₁₂).

ImImHpPy- γ -ImHpPyPy- β -Dp (4) A sample of ImImOpPy- γ -ImOpPyPy- β -Dp (5 mg, 3.9 mmol) was treated with sodium thiophenoxide and purified by reversed phase HPLC as described for ImImPyPy- γ -ImHpPyPy- β -Dp. ImImHpPy- γ -ImHpPyPy- β -Dp is recovered as a white powder upon lyophilization of the appropriate fractions (3.8 mg, 77% recovery). UV (H₂O) λ_{max} 246, 312 (66,000); ¹H NMR (DMSO-d₆) δ 10.34 (s, 1 H), 10.24 (s, 1 H), 10.00 (s, 2 H), 9.93 (s, 1 H), 9.87 (s, 1 H), 9.83 (s, 1 H), 9.4 (br s, 1 H), 9.04 (s, 1 H), 8.03 (m, 3 H), 7.58 (s, 1 H), 7.44 (s, 1 H), 7.42 (s, 1 H), 7.23 (s, 1 H), 7.20 (m, 3 H), 7.12 (m, 2 H), 7.05 (d, 1 H), 7.02 (d, 1 H), 6.83 (s, 1 H), 6.79 (s, 1 H), 3.96 (s, 6 H), 3.90 (s, 3 H), 3.81 (s, 6 H), 3.79 (s, 3 H), 3.75 (d, 6 H), 3.33 (q, 2 H, J = 5.4 Hz), 3.14 (q, 2 H, J = 5.4 Hz), 3.08 (q, 2 H, J = 6.1 Hz), 2.99 (quintet, 2 H, J = 5.4 Hz), 2.69 (d, 6 H, J =

4.2 Hz), 2.31 (m, 4 H), 1.72 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1254.57 (1254.55 calcd. for $C_{57}H_{70}N_{22}O_{12}$).

ImImPyPy- γ -ImOpPyPy- β -Dp-NH₂ (2-Me-NH₂) ImImPyPy- γ -ImOpPyPy- β -Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc- β -Pam-Resin (0.66 mmol/g) as described previously for hydroxypyrrole-imidazole-pyrrole polyamides.²¹ A sample of ImImPyPy- γ -ImOpPyPy- β -Pam-Resin (400 mg, 0.40 mmol/g) was placed in a glass 20 mL peptide synthesis vessel and treated with neat 3,3'-diamino-*N*-methyldipropylamine (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. ImImPyPy- γ -ImOpPyPy- β -Dp-NH₂ is recovered upon lyophilization of the appropriate fractions as a white powder (104 mg, 54% recovery). UV (H₂O) λ_{max} 246, 316 (66,000); MALDI-TOF-MS (monoisotopic), 1296.6 (1296.6 calcd. for $C_{60}H_{78}N_{23}O_{11}$).

ImImPyPy- γ -ImOpPyPy- β -Dp-EDTA (2-Me-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 ImImPyPy- γ -ImOpPyPy- β -Dp-NH₂ (13 mg, 10 mmol) dissolved in DMSO (750 mL). The mixture was heated (55 °C, 25 min.) and the remaining EDTA-dianhydride hydrolyzed (0.1 M 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 ImImPyPy- γ -ImOpPyPy- β -Dp-EDTA as a white powder upon lyophilization of the appropriate fractions (5.9 mg, 42% recovery). MALDI-TOF-MS (monoisotopic), 1570.8 (1570.7 calcd. for C₇₀H₉₂N₂₅O₁₈).

ImImPyPy- γ -ImHpPyPy- β -Dp-EDTA (2-E) In order to remove the methoxy protecting group, a sample of ImImPyPy- γ -ImOpPyPy- β -Dp-EDTA (5 mg, 3.1 mmol) was treated with sodium thiophenoxide at 100 °C for 2 h and purified as before to give as a white powder ImImPyPy- γ -ImHpPyPy- β -Dp-EDTA (3.2 mg, 72% recovery). UV (H₂O) λ_{max} 246, 312 (66,000); ¹H NMR (DMSO-d₆) δ 10.40 (s, 1 H), 10.26 (s, 1 H), 10.07 (s, 1 H), 10.02 (s, 1 H), 9.95 (s, 1 H), 9.90 (s, 1 H), 9.86 (s, 1 H), 9.07 (s, 1 H), 8.43 (s, 1 H), 8.08 (m, 3 H), 7.61 (s, 1 H), 7.46 (s, 1 H), 7.45 (s, 1 H), 7.26 (s, 1 H), 7.24 (d, 1 H), 7.21 (d, 1 H), 7.15 (s, 1 H), 7.10 (s, 1 H), 7.07 (s, 1 H), 6.93 (s, 1 H), 6.87 (d, 1 H), 6.82 (d, 1 H), 6.64 (br s, 1 H), 4.00 (s, 6 H), 3.94 (s, 3 H), 3.84 (s, 3 H), 3.83 (s, 3 H), 3.81 (s, 3 H), 3.80 (s, 3 H), 3.79 (s, 3 H), 3.67 (s, 3 H), 3.5-3.1 (m, 14 H), 3.04 (m, 2 H), 2.71 (s, 3 H), 2.35 (t, 2 H, J = 5.4 Hz), 1.99 (m, 4 H), 1.78 (m, 2 H), 1.44 (m, 2 H), 1.23 (s, 6 H) 0.84 (m, 2 H); MALDI-TOF-MS (monoisotopic), 1555.9 (1556.7 calcd. for C₆₉H₉₀N₂₅O₁₈).

ImImOpPy- γ -ImPyPyPy- β -Dp-NH₂ (3-Me-NH₂) ImImOpPy- γ -ImPyPyPy- β -Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc-b-Pam-Resin (0.66 mmol/g),²¹ and the polyamide was cleaved from the resin using neat 3,3'-diamino-*N*-methyldipropylamine (2 mL) as before. After purification by reversed phase HPLC, ImImOpPy- γ -

ImPyPyPy- β -Dp-NH₂ is recovered upon lyophilization of the appropriate fractions as a white powder (93 mg, 46% recovery). UV (H₂O) λ_{max} 246, 316 (66,000); ¹H NMR (DMSO-d₆) δ 10.34 (s, 1 H), 10.30 (br s, 1 H), 10.25 (s, 1 H), 9.96 (s, 1 H), 9.95 (s, 1 H), 9.89 (s, 1 H), 9.24 (s, 1 H), 9.11 (s, 1 H), 8.08 (t, 1 H, *J* = 5.6 Hz), 8.0 (m, 5 H), 7.62 (s, 1 H), 7.53 (s, 1 H), 7.42 (s, 1 H), 7.23 (d, 1 H, *J* = 1.2 Hz), 7.21 (m, 2 H), 7.15 (m, 2 H), 7.13 (d, 1 H), 7.11 (m, 2 H), 7.04 (d, 1 H), 6.84 (m, 3 H), 3.98 (s, 3 H), 3.97 (s, 3 H), 3.92 (s, 3 H), 3.82 (m, 6 H), 3.80 (s, 3 H), 3.77 (d, 6 H), 3.35 (q, 2 H, *J* = 5.8 Hz), 3.0-3.3 (m, 8 H), 2.86 (q, 2 H, *J* = 5.4 Hz), 2.66 (d, 3 H, *J* = 4.5 Hz), 2.31 (m, 4 H), 1.94 (quintet, 2 H, *J* = 6.2 Hz), 1.74 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1296.0 (1296.6 calcd. for C₆₀H₇₈N₂₃O₁₁).

ImImOpPy- γ -ImPyPyPy- β -Dp-EDTA (3-Me-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 ImImOpPy- γ -ImPyPyPy- β -Dp-NH₂ (13 mg, 10 mmol) dissolved in DMSO (750 mL). The mixture was heated (55 °C, 25 min.) and the remaining EDTA-dianhydride hydrolyzed (0.1 M 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 ImImOpPy- γ -ImPyPyPy- β -Dp-EDTA as a white powder upon lyophilization of the appropriate fractions (5.5 mg, 40% recovery). MALDI-TOF-MS (monoisotopic), 1570.9 (1570.7 calcd. for C₇₀H₉₂N₂₅O₁₈).

ImImHpPy- γ -ImPyPyPy- β -Dp-EDTA (3-E) In order to remove the methoxy protecting group, a sample of ImImOpPy- γ -ImPyPyPy- β -Dp-EDTA (5 mg, 3.1

mmol) was treated with sodium thiophenoxide at 100 °C for 2 h and purified as before to give as a white powder ImImHpPy- γ -ImPyPyPy- β -Dp-EDTA (3.2 mg, 72% recovery). UV (H₂O) λ_{max} 246, 312 (66,000); MALDI-TOF-MS (monoisotopic), 1556.6 (1556.7 calcd. for C₆₉H₉₀N₂₅O₁₈).

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.³²

Preparation of 3'- and 5'-End-Labeled Restriction Fragments. The plasmids pJT6 and pDEH9 were constructed as previously reported. pJT6 and pDEH9 were linearized with *Eco* RI and *Pvu* II restriction enzymes, 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 *Eco* RI, 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 *Pvu* II. The

labeled fragment (3' or 5') was loaded onto a 6% non-denaturing polyacrylamide gel, and the desired 250 base pair band was visualized by autoradiography and isolated.

MPE•Fe(II) Footprinting.²³ All reactions were performed in a total volume of 40 μ L. A polyamide stock solution (1,2,3, or 4) or H₂O (for reference lanes) was added to an assay buffer containing labeled restriction fragment (20,000 cpm), affording final solution conditions of 25 mM Tris-Acetate, 10 mM NaCl, 100 μ M/bp calf thymus DNA, pH 7. Solutions were incubated at 22 °C for 24 hours. A fresh 50 μ M MPE•Fe(II) solution was made from 100 μ L of a 100 μ M MPE solution and 100 μ M ferrous ammonium sulfate ($\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$) solution. Then, 4 μ L of the 50 μ M MPE•Fe(II) solution was added, and the solution was allowed to equilibrate for 10 min. at 22 °C. Cleavage was initiated by the addition of 4 μ L of a 50 mM dithiothreitol solution and allowed to proceed for 15 min. at 22 °C. Reactions were stopped by ethanol precipitation, and were resuspended in 1x TBE/80% formamide loading buffer, denatured by heating at 85 °C for 15 min., and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% cross-link, 7 M urea) in 1x TBE at 2000 V for 1.5 hours. Gels were dried and exposed to a storage phosphor screen. Relative cleavage intensities were determined by volume integration of individual cleavage bands using ImageQuant software.

Affinity Cleaving.²⁴ All reactions were executed in a total volume of 40 μ L. A stock solution of polyamide (**1-E,2-E,3-E, or 4-E**) or H₂O was added to a solution containing labeled restriction fragment (20,000 cpm), affording final solution conditions of 25 mM Tris-Acetate, 20 mM NaCl, 100 μ M/bp calf thymus DNA, and pH 7.0. Solutions were incubated for a minimum of 4 hours at 22 °C. Subsequently, 4 μ L of freshly prepared 100 μ M Fe(NH₄)₂(SO₄)₂ was added and the solution allowed to equilibrate for 20 min. Cleavage reactions were initiated by the addition of 4 μ L of 100 mM dithiothreitol, allowed to proceed for 30 min. at 22 °C, then stopped by the addition of 10 μ L of a solution containing 1.5 M NaOAc (pH 5.5), 0.28 mg/mL glycogen, and 14 μ M base pairs calf thymus DNA, and ethanol precipitated. The reactions were resuspended in 1x TBE/80% formamide loading buffer, denatured by heating at 85 °C for 15 min., and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% cross-link, 7 M urea) in 1x TBE at 2000 V for 1.5 hours. Gels were dried and exposed to a storage phosphor screen. Relative cleavage intensities were determined by volume integration of individual cleavage bands using ImageQuant software.

Quantitative DNase I footprint titrations.¹³ All reactions were executed in a total volume of 40 μ L. A polyamide stock solution (**1, 2, 3, or 4**) or H₂O (for reference lanes) was added to an assay buffer containing radiolabeled

restriction fragment (20,000 cpm), affording final solution conditions of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂, 5 mM CaCl₂, pH 7.0, and either (i) 0.1 nM-1 μM polyamide or (ii) no polyamide (for reference lanes). The solutions were allowed to equilibrate at 22 °C for 24 h. Footprinting reactions were initiated by the addition of 4 μL of a DNase I stock solution (at the appropriate concentration to give ~ 55% intact DNA) containing 1 mM dithiothreitol and allowed to proceed for seven min. at 22 °C. The reactions were stopped by the addition of 10 μL of a solution containing 1.25 M NaCl, 100 mM EDTA, 0.2 mg/mL glycogen, and 28 μM base-pair calf thymus DNA, and ethanol precipitated. Reactions were resuspended in 1x TBE/80% formamide loading buffer, denatured by heating at 85 °C for 15 min., and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% cross-link, 7 M urea) in 1x TBE at 2000 V for 1.5 h. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics). Equilibrium association constants were determined as previously described.^{6h}

Quantitation and data analysis. Data from the footprint titration gels were obtained using a Molecular Dynamics 400S PhosphorImager followed by quantitation using ImageQuant software (Molecular Dynamics). Background-corrected volume integration of rectangles encompassing the footprint sites and a reference site at which DNase I reactivity was invariant across the

titration generated values for the site intensities (I_{site}) and the reference intensity (I_{ref}). The apparent fractional occupancy (θ_{app}) of the sites were calculated using the equation:

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

where I_{site}° and I_{ref}° are the site and reference intensities, respectively, from a control lane to which no polyamide was added. The ($[L]_{\text{tot}}$, θ_{app}) data points were fit to a Langmuir binding isotherm (eq 2, $n=1$) 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 $[L]_{\text{tot}}$ is the total polyamide concentration, K_a is the equilibrium association constant, and θ_{min} and θ_{max} are the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. The data were fit using a nonlinear least-squares fitting procedure 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 data were used in determining each association constant. All lanes from each gel were used unless visual inspection revealed a data point to obviously flawed relative to neighboring points.

References

1. (a) Wade, W.S.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1992**, 114, 8783. (b) Wade, W.S.; Mrksich, M.; Dervan, P.B. *Biochemistry* **1993**, 32, 11385. (c) Mrksich, M.; Wade, W.S.; Dwyer, T.J.; Geierstanger, B.H.; Wemmer, D.H.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 7586.
2. Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, 73, 804-808.
3. Steitz, T. A. *Quart. Rev. Biophys.* **1990**, 23, 203-280.
4. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, 115, 2572. (b) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, 117, 3325. (c) Geierstanger, B.H.; Dwyer, T.J.; Bathini, Y.; Lown, J.W.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, 115, 4474. (d) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, 266, 646. (e) Geierstanger, B.H.; Jacobsen, J.P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Biochemistry* **1994**, 33, 3055.
5. (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.
6. (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) de Claire, R. P. L.; Geierstanger B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, 119, 7909. (g) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, 119, 8756. (h) White, S.; Baird, E. E.; Dervan, P. B. *Chem & Biol.* **1997**, 4, 569.
7. (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. (d) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, 120, 6219.
8. (a) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, 118, 6160. (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem.*

- Eur. J.* **1997**, 3, 1600. (c) Trauger, J.W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, 120, 3534.
9. De Clairac, R.P.L.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1997**, 119, 7906.
 10. Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nature Struct. Biol.* **1998**, 5, 104.
 11. Pilch, D. S. *et al.* *Proc. Natl. Acad. Sci. U.S.A.* **1996**, 93, 8306.
 12. Wong, J. M.; Bateman, E. *Nucl. Acids. Res.* **1994**, 22, 1890.
 13. (a) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods Enzymol.* **1986**, 130, 132. (b) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, 83, 8462. (c) Senear, D.F.; Brenowitz, M.; Shea, M.A.; Ackers, G.K. *Biochemistry* **1986**, 25, 7344.
 14. Kim, Y.; Geiger, J. H.; Hahn, S.; Sigler, P. B. *Nature* **1993**, 365, 512.
 15. Gartenberg, M. R.; Crothers, D. M. *Nature* **1988**, 333, 824.
 16. Sluka, J. P.; Horvath, S. J.; Glasgow, A. C.; Simon, M. I.; Dervan, P. B. *Biochemistry* **1990**, 29, 6551.
 17. Ades, S. E.; Sauer, R. T. *Biochemistry* **1995**, 34, 14601-14608.
 18. Choo, Y.; Klug, A. *Curr. Opin. Struct. Biol.* **1997**, 7, 117-125.
 19. White, S.; Szewczyk, J.W.; Turner, J.M.; Baird, E.E.; Dervan, P.B. *Nature* **1998**, 391, 468.
 20. Kielkopf, C.L.; White, S.; Szewczyk, J.W.; Turner, J.M.; Baird, E.E.; Dervan, P.B.; Rees, D.C. **1998**, 282, 111.
 21. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, 118, 6141.
 22. Momose, T.; Tamaka, T.; Yokota, T.; Nagamoto, N.; Yamada, K. *Chem. Pharm. Bull.* **1978**, 26, 2224.
 23. (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.

24. (a)Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457. (b) Dervan, P. B. *Science* **1986**, *232*, 464.
25. Branden, C.; Tooze, J. *Introduction to Protein Structure*; Garland: New York, 1991.
26. White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532.
27. Brameld, K.; Dasgupta, S.; Goddard, W.A. *J. Amer. Phys. Chem. B.* **1997**, *101*, 4851.
28. Kopka, M.L.; Goodsell, D.S.; Han, G.W.; Chiu, T.K.; Lown, J.W.; Dickerson, R. *Structure* **1997**, *5*, 1033.
29. Kim, J.L.; Nikolov, D.B.; Burley, S.K. *Nature* **1993**, *365*, 520.
30. Kent, S. B. H.; *Annu. Rev. Biochem.* **1988**, *57*, 957.
31. Barlos, K.; Chatzi, O.; Gatos, D.; Stravropoulos, G. *Int. J. Peptide Protein Res.* **1991**, *37*, 513.
32. Sambrook, J.; Fritsch, E.F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.