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

Thesis by

Sarah White

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1999

(Submitted August 24, 1998)

To Eldon

and

To my family

Acknowledgments

First and foremost, I would like to thank my advisor, Prof. Peter B. Dervan, for his endless enthusiasm, guidance, patience, and support during my graduate career. I feel extremely fortunate to have been able to study in his laboratory, where I have grown as a scientist and a person under his tutelage. I would also like to thank Prof. Brent L. Iverson, my undergraduate advisor, who introduced me to chemical research and his former advisor, Peter Dervan. Brent offered unwavering support and encouragement in my pursuit of a Ph.D. in chemistry, at a time when I needed it. Mr. Kurth, my high school chemistry teacher, also deserves special thanks for making chemistry fun and introducing me to the joys of teaching chemistry to others.

I thank all the people who have supported me through the years at Caltech, making the last five years incredibly fun, interesting, and unforgettable. First, I would like to thank my husband, Eldon Baird. I can barely see the computer screen through my tears as I type these words of thanks to the person who has shown me what true love and respect are. Our collaboration over the past few years has resulted in all the work presented herein, work that has special meaning to me because it came from our effort *together*. I look forward to a lifetime together with Eldon, starting a family and continuing our successful scientific collaboration. I thank all the group members who have taught me so much about life and science: Dr. Joe Hacia, who took the time and care to help me design and carry out my first project; Dr. Matt Taylor, with whom I share a love of music and Nancy Griffith and who gave me helpful advice on my first project; Dr. Scott Priestly, who taught me important concepts and techniques in footprinting; Dr. John Trauger, who also provided me with invaluable advice and tutelage on footprinting techniques, and who gave Dervan group members a level of perfection for which to strive; Jim Turner and Jason Szewczyk, whose friendship, help, and support for the hydroxypyrrole project made it possible (and fun ☺); Ryan Bremer (and Liz Stieg), who provided me with friendship, advice, and support, making

Caltech a very enjoyable experience; Aileen Chang, who gave us lots of surprise goodies over the years, making the 3rd floor all that much more fun. I would like to thank David Herman, Clay Wang, and Nick Wurtz also for making the Dervan Group fun and entertaining.

The hydroxypyrrole-imidazole-pyrrole polyamide-DNA structure was done in collaboration with Prof. Doug Rees and Clara Kielkopf. Clara and Doug have been extremely helpful and nice, providing the Dervan Group and the scientific community with important insight into how polyamides recognize the minor groove of DNA. Their hard work and incredible insight are much appreciated.

I also need to thank the special people I have met outside the chemistry department, who have made my tenure at Caltech that much more meaningful and fun. First, I want to thank Bill and Delores Bing, the directors of the concert band and chamber music. Their enthusiasm and love for music is contagious, and they have built a wonderful music program at Caltech. I also want to thank Monica Hubbard, director of the Women's Glee Club. She is an incredible person and musician who taught me a lot about singing, choral music, and important women in history. I want to thank all the people I have met while I was in those groups, especially Sarah Ngola, Gary Mines, Mike Werner, Sharon Laubach, Laura Wasylenki, and Julie Norris.

I would like to thank all the members of my thesis committee, Prof. R.H. Grubbs, Prof. J.D. Baldeschwieler, Prof. D.C. Rees, and Prof. P.B. Dervan for their time and guidance over the past year. I also want to thank the former members of my committee, Prof. A.G. Myers and Prof. H. Gray, who provided invaluable advice at my candidacy exam.

Finally, I would like to thank my family for all their love and support throughout the years. You have given everything I have ever needed and wanted, and words cannot express how grateful I am for that. I am very proud to add another Ph.D. to the family, following in your footsteps.

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.

Table of Contents

List of Figures and Tables

CHAPTER ONE page

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

Chapter 1

Introduction

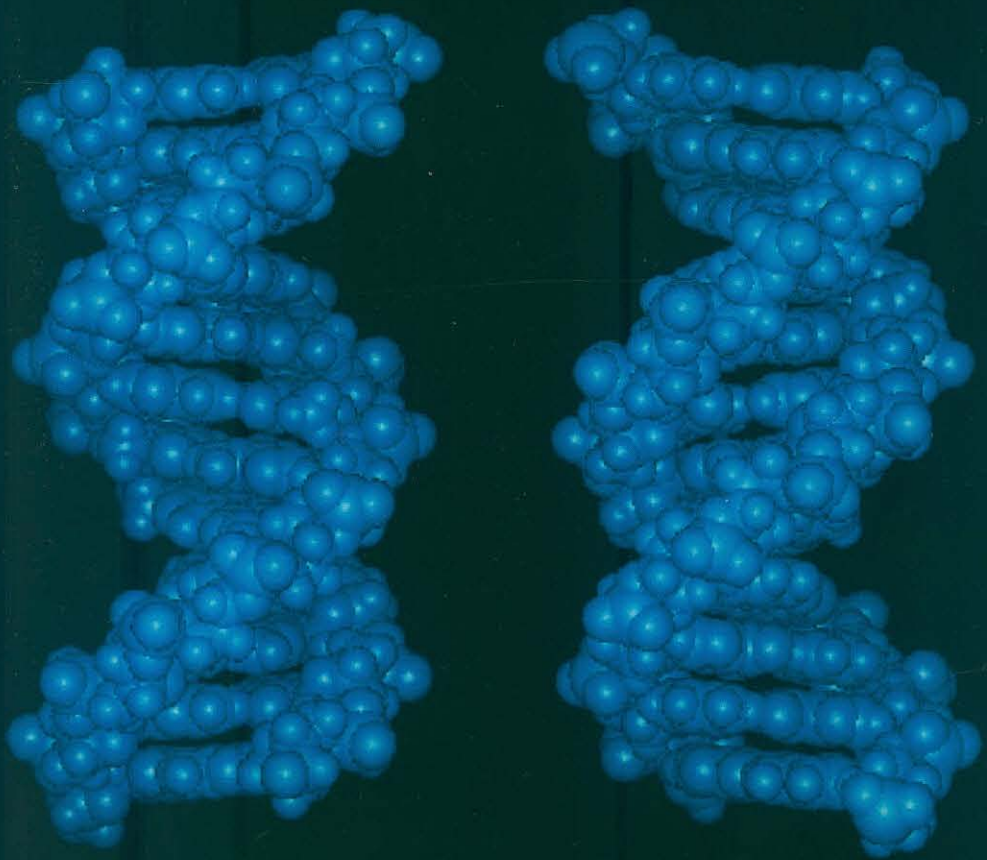
The design of synthetic ligands that read the information stored in the DNA (deoxyribonucleic acid) double helix has been a long-standing goal at the interface of chemistry and biology.[1] Double-helical DNA plays a central role in biological processes. In every human cell, the DNA polymer stores the hereditary information in the form of base pairs, encoding approximately 80,000 genes.[2] Specific non-covalent interactions of proteins with DNA determine cellular gene function and regulation.[3] Cell-permeable small molecules that target predetermined DNA sequences offer a potential approach for regulation of gene expression.[4]

Double-helical DNA. The genetic information is stored on two polydeoxyribonucleic acid strands which associate in an antiparallel orientation, forming a double-helical structure.[5] The DNA double helix consists of A,T and G,C base pairs held together via Watson-Crick hydrogen bonds.[6] The common B-form of DNA is characterized by a wide (12 Å) and shallow major groove and a narrow (4-6 Å) and deep minor groove (Figure 1.1).[7] In addition, sequence-dependent structural variations, conformational properties, and solvent and counterion organization can distinguish local DNA structures.[7]

Individual sequences may be distinguished by the pattern of hydrogen bond donors and acceptors displayed on the edges of the base pairs (Figure 1.2).[7] In the minor groove, the A,T base pair presents two symmetrically placed hydrogen bond acceptors, the purine N3 and the pyrimidine O2 atoms. The G,C base pair presents these two acceptors, but in addition presents a hydrogen bond donor, the exocyclic 2-amino group of guanine (Figure 1.3).[8]

Molecular recognition of DNA. The design of new sequence specific DNA-binding molecules is aided by principles learned from structural analyses of complexes of DNA with proteins and natural products.[1a,8] Sequence-specificity arises from the ability of these molecules to form specific hydrogen bonds or van der Waals contacts with functional groups in the grooves, by Coulombic attraction to the negatively charged phosphodiesters or to the electrostatic potential in the grooves, and/or through intercalation of aromatic functional groups between the DNA bases.[9]

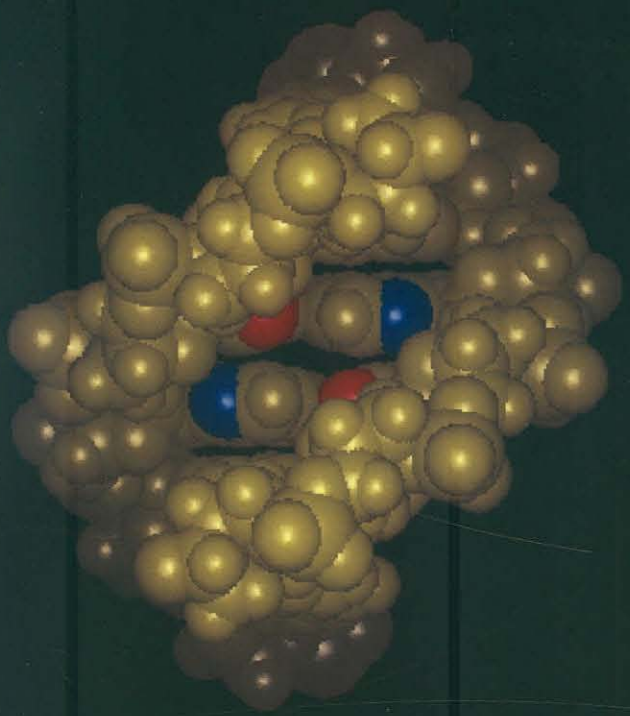
Figure 1.1 Molecular models of B-form double-helical DNA. On the left, the wide and shallow major groove is shown. On the right, the deep and narrow minor groove.



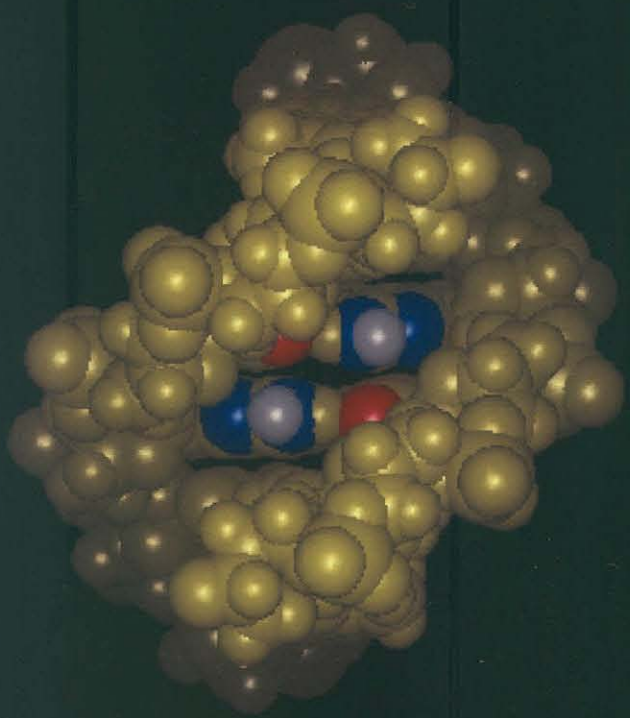
Major Groove

Minor Groove

Figure 1.2 Hydrogen bond partners on the floor of the minor groove. (top) The A,T base pair contains two hydrogen bond acceptors: the purine N3 (blue) and the pyrimidine O2 (red). (bottom) The G,C base pair contains these two acceptors, but in addition presents a hydrogen bond donor from the 2-amino group of guanine (white). The presence of the guanine amino group also introduces a "bump" on the floor of the minor groove.



T-A and A-T Base Pairs



C-G and G-C Base Pairs

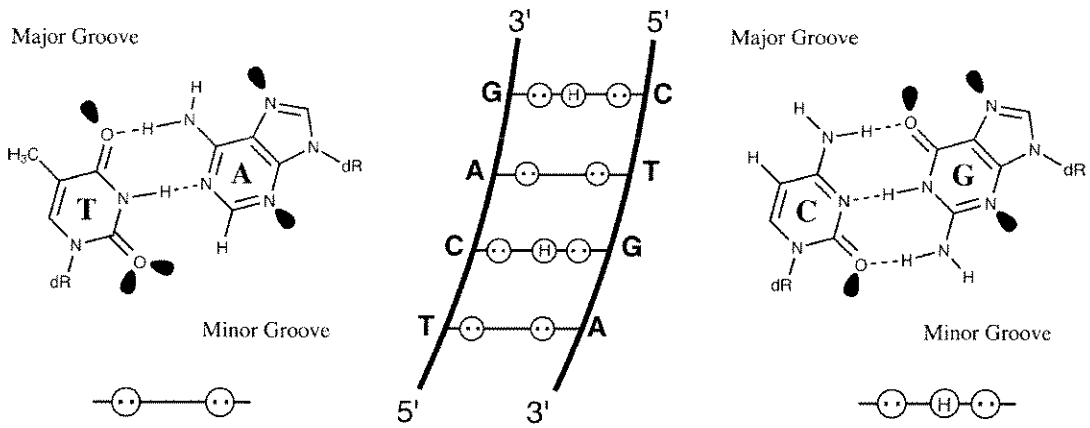
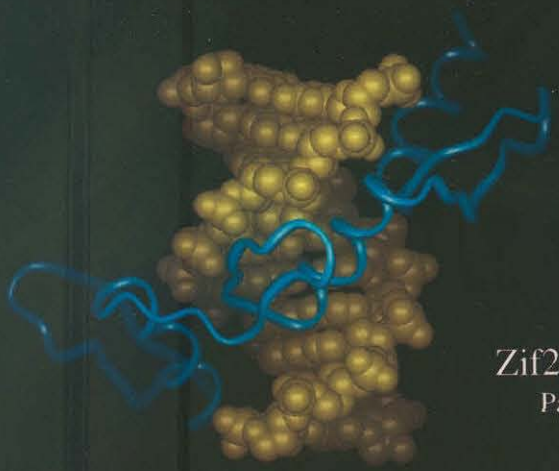


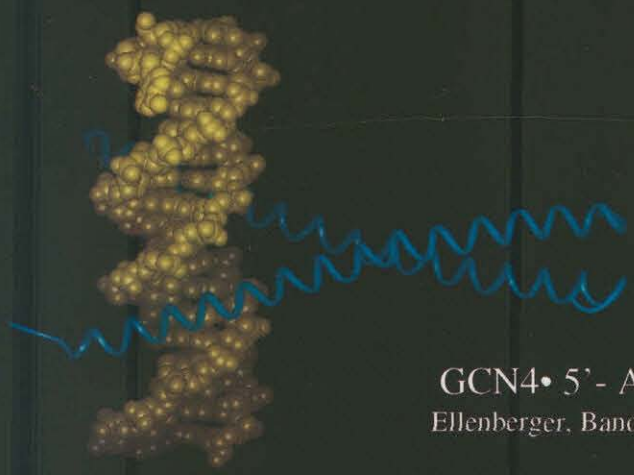
Figure 1.3 A schematic model for recognition of the minor groove, with hydrogen bond donors represented as (H) and hydrogen bond acceptors represented as two dots. This schematic underscores the potential coded reading of the DNA helix.

DNA binding proteins adopt several structural motifs for sequence-specific recognition which include the zinc finger[10], the leucine zipper[11], and the helix-turn-helix motifs (Figure 1.4).[12] Within these complexes, sequence-specificity is achieved through specific noncovalent interactions between the protein side chains and the nucleobases and phosphates of the DNA. However, these studies indicate that there is no "code" for protein-DNA interactions, that is no rigid rules determining which amino acids contact which base pairs.[13] Although a versatile recognition code is emerging from studies of certain zinc finger-DNA complexes[14], the de novo design of proteins for recognition of designated sequences proves to be difficult, due to the structural diversity available to a single protein, and limitations in predicting protein folding.

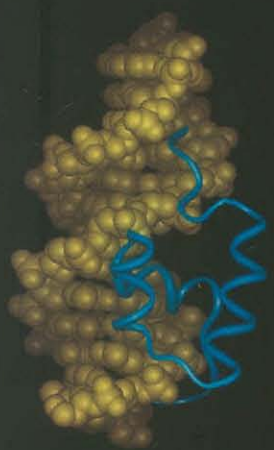
Figure 1.4 X-ray crystal structures of protein-DNA complexes. (top) Zinc finger Zif268 with 5'-GCGTGGGCG-3'.[10] (middle) Leucine zipper GCN4 with 5'-ATGACTCATCCA-3'.[11] (bottom) Helix-turn-helix Hin recombinase with 5'-TTTTGATAA-3'.[12]



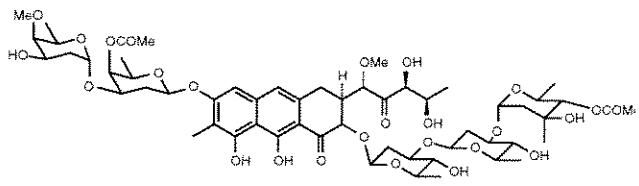
Zif268• 5'-GCGTGGGCG-3'
Pavletich & Pabo *Science* **1991**



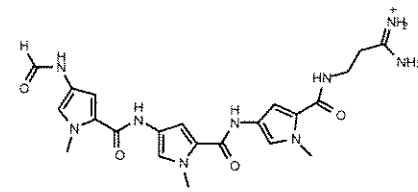
GCN4• 5'- ATGACTCATCCA -3'
Ellenberger, Bandl, Struhl & Harrison *Cell* **1992**



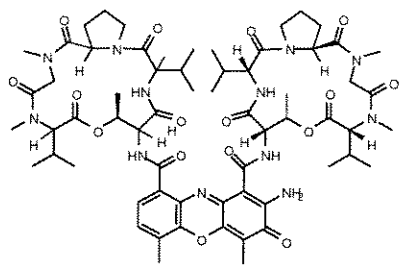
Hin • 5'-TTTTGATAA-3'
Feng, Johnson & Dickerson *Science* **1994**



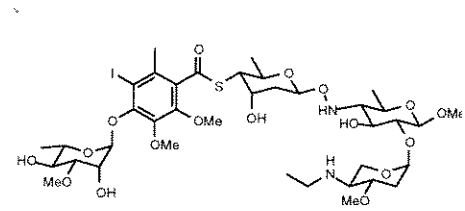
Chromomycin • 5'-GGCT-3'



Distamycin • 5'-(A,T)$_5$-3'



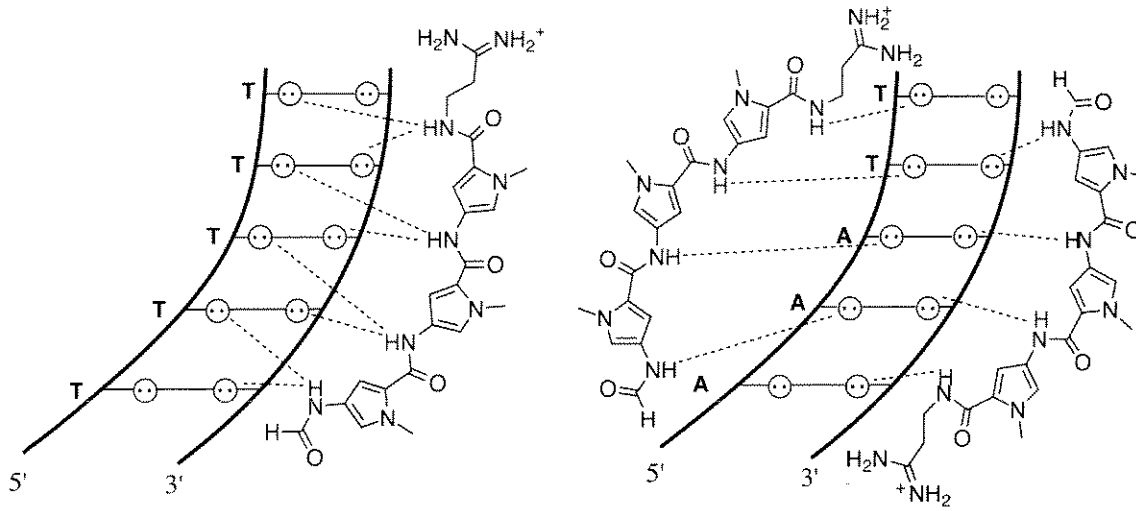
Actinomycin D • 5'-GC-3'



Calicheamicin oligosaccharide • 5'-TCCT-3'

Figure 1.5 Chemical structures of DNA-binding molecules from natural sources.

In addition to proteins, many small molecules have been isolated from natural sources which bind DNA with modest sequence-specificities. Examples of this structurally diverse class include chromomycin, actinomycin D, calicheamicin, and distamycin A (Figure 1.5).[15,16] These ligands generally bind in the minor groove and/or intercalate between the base pairs of DNA. And much like DNA-binding proteins, these molecules have no natural recognition code for readout of specific sequences of DNA. However, among these small molecules, distamycin is distinguished by its structural simplicity, having an oligopyrrolecarboxamide core structure with no chiral centers.[1a,17] Structural studies of distamycin-DNA complexes reveal modular recognition units, in which adjacent pyrrolecarboxamides make similar contacts with adjacent DNA base pairs (Figure 1.6).[16] The relative simplicity of distamycin, with respect both to its chemical structure and its complexes with DNA, makes this ligand an ideal system for guiding the design of polyamides having novel DNA-binding sequence specificity.[1b]



1:1 Distamycin•DNA Complex *2:1 Distamycin•DNA Complex*

Figure 1.6 A schematic representation of recognition of A,T rich sequences in the minor groove by 1:1 and 2:1 complexes of Distamycin.

Two distinct DNA binding modes exist for distamycin (Figure 1.6). In the first binding mode, a single molecule of distamycin binds in the middle of the minor groove of a 5 base pair A,T-rich sequence. The amide hydrogens of the *N*-methylpyrrolecarboxamides form bifurcated hydrogen bonds with the N3 of adenine and the O2 of thymine on the floor of the minor groove.[16] In the second binding mode, two distamycin ligands form an antiparallel side-by-side dimer in the minor groove of a 5 base pair A,T rich site.[18] In the 2:1 complex, each polyamide subunit forms hydrogen bonds to a unique DNA strand. In both binding motifs, the pyrrole rings fill the groove completely, forming extensive van der Waals contacts with the walls of the groove. The aromatic hydrogens of the *N*-methylpyrrole rings are set too deeply to allow room for the guanine 2-amino group, affording binding specificity for A,T sequences. The cationic tail(s) lies along the floor of the minor groove making favorable electrostatic interactions with the DNA.

DNA recognition by polyamides. In search of a general DNA recognition code, a twenty year research effort led by Prof. Peter B. Dervan of Caltech has resulted in the development of pairing rules to guide polyamide design for sequence-specific DNA recognition. A binary

code has been developed to correlate DNA sequence with side-by-side pairings between *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) carboxamides in the DNA minor groove (Figure 1.7).[19] A pairing of Im opposite Py (Im/Py) targets a G•C base pair, while Py/Im targets C•G.[19] A Py/Py pairing is degenerate, targeting both A•T and T•A base pairs.[1b,19,20] 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. Footprinting, NMR, and x-ray structure studies validate these pairing rules for DNA minor groove recognition.[21]

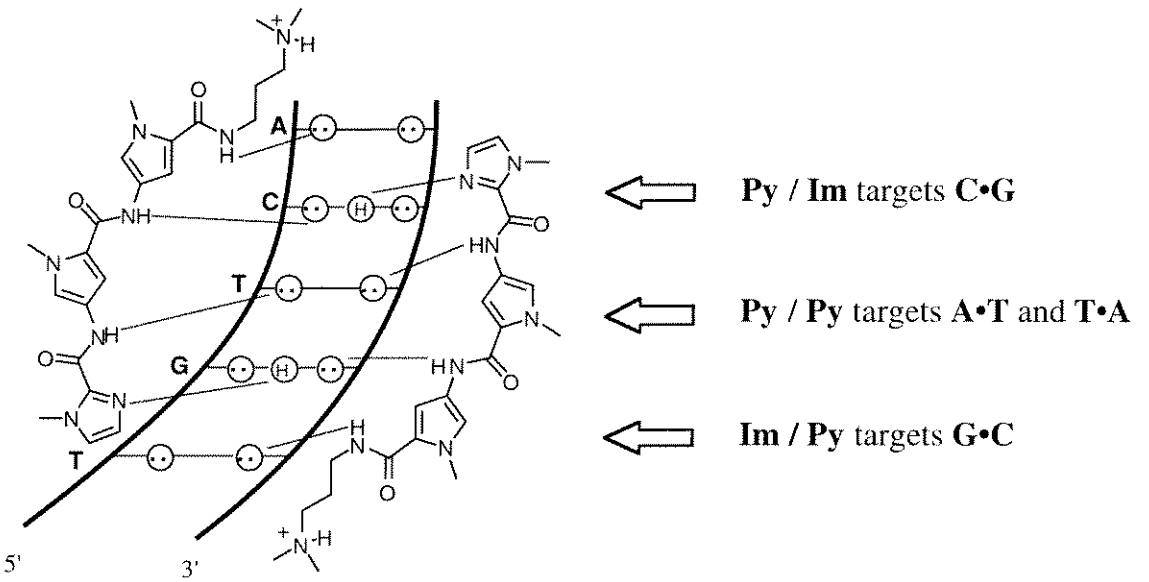
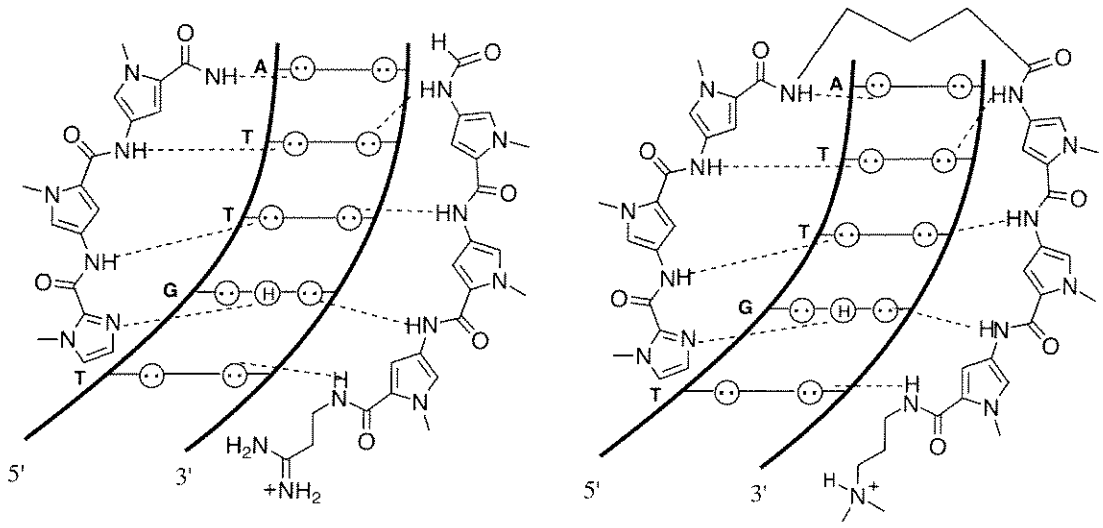


Figure 1.7 Pairing rules for polyamide recognition in the minor groove of DNA.

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). Linking the two polyamide subunits ImPyPy and PyPyPy to form ImPyPy-γ-PyPyPy increases the binding affinity 300-fold for the hairpin polyamide over the unlinked subunits.[24]

The increased affinity and specificity of hairpin polyamides allows for study of the energetics of aromatic ring pairings, polyamide orientation preferences, and chemical substitutions of the aromatic rings. The hairpin polyamide model is supported by footprinting, affinity cleaving, and NMR structure studies.[24,25] Increasing the hairpin polyamide subunit lengths to



ImPyPy-Dp / PyPyPy • TGTTA

$K_a \sim 2 \times 10^5\ M^{-1}$

ImPyPy-γ-PyPyPy-Dp • TGTTA

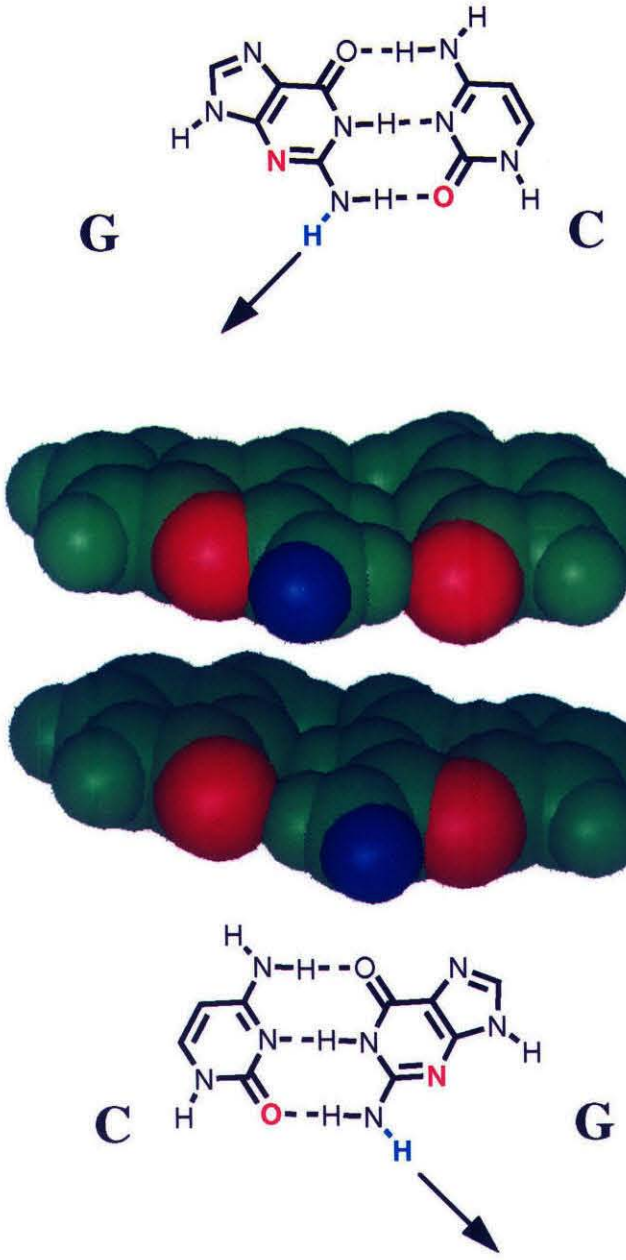
$K_a = 8 \times 10^7\ M^{-1}$

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).

four and five aromatic ring residues gives eight and ten ring hairpin polyamides with subnanomolar binding affinities, similar to naturally occurring DNA-binding proteins.[26]

Recognition of all four Watson-Crick base pairs in the minor groove. Because of the degeneracy of the hydrogen bond donors and acceptors displayed on the edges of the base pairs, the minor groove was believed to be a "veritable recognition desert," lacking sufficient information for a complete recognition code.[8,27] But, despite the central placement of the guanine exocyclic N2 amine group in the G,C minor groove (Figure 1.9), Py/Im and Im/Py pairings distinguish energetically between C•G and G•C.[19a,b;21c;26a;28]

Figure 1.9 Chemical structures and space-filling models of the G•C and C•G base pairs as viewed from the minor groove of DNA, generated using B-form DNA coordinates provided in InsightII. The O2 and N3 atoms are highlighted in red. The asymmetrically directed 2-amino proton is highlighted in blue.


G
C
C
G

The neighboring Py packs an Im to one side of the minor groove resulting in a precisely placed hydrogen bond between the Im N3 and the guanine N2 for specific recognition (Figure 1.7).[18d,24f] The remarkable sensitivity to single atomic replacement indicates that substitution at the 3 position of one Py within a Py/Py pair can complement 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 1.10).[1b,27] The existence of an asymmetrically placed cleft in the minor groove surface between the thymine O2 and the adenine C2-H suggests a possible shape-selective mechanism for T•A recognition.[29] 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. Footprinting studies reveal that replacement of a single hydrogen atom with a hydroxy group in a Hp/Py pairing regulates affinity and specificity of DNA-binding polyamides by an order of magnitude.[30] As shown by x-ray structure studies, selectivity arises from steric destabilization of polyamide binding via placement of Hp opposite A and stabilization by specific hydrogen bonds between Hp and T (Figures 1.11, 1.12).[31] By incorporation of this third 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 1.12).

Figure 1.10 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 shown in red.

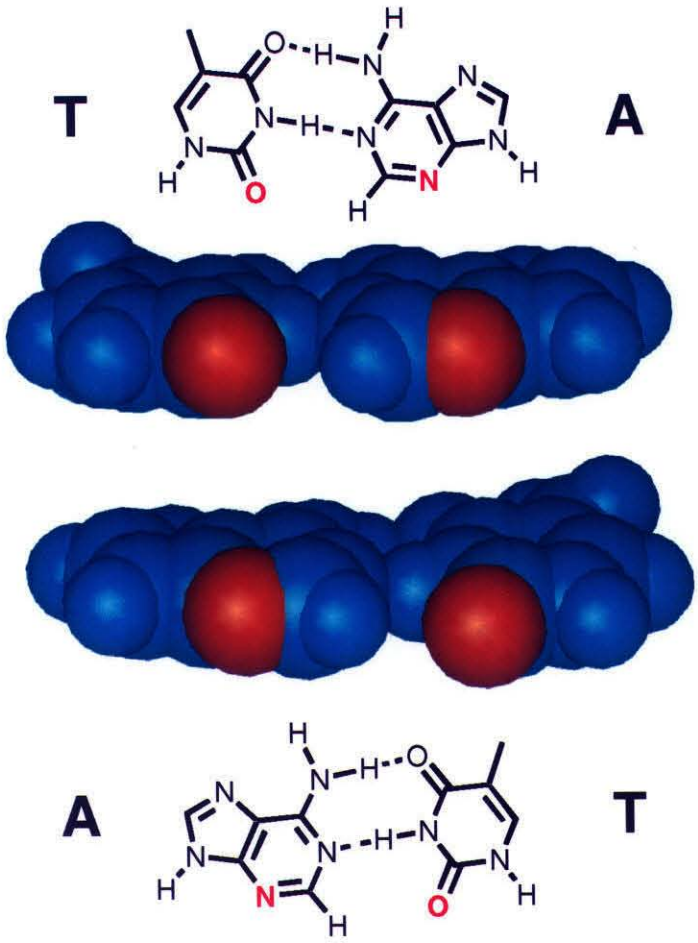
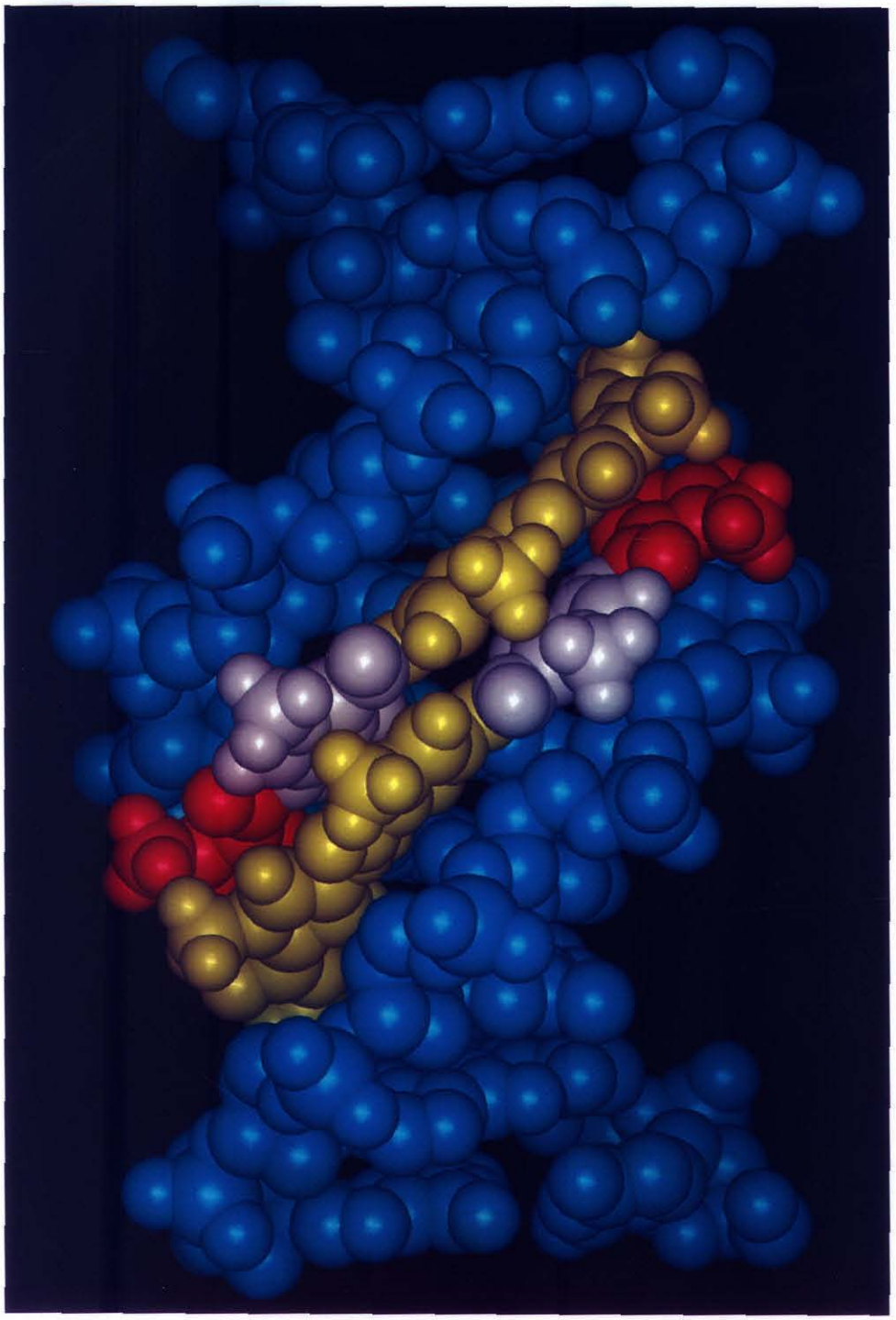

T
A
A
T

Figure 1.11 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.



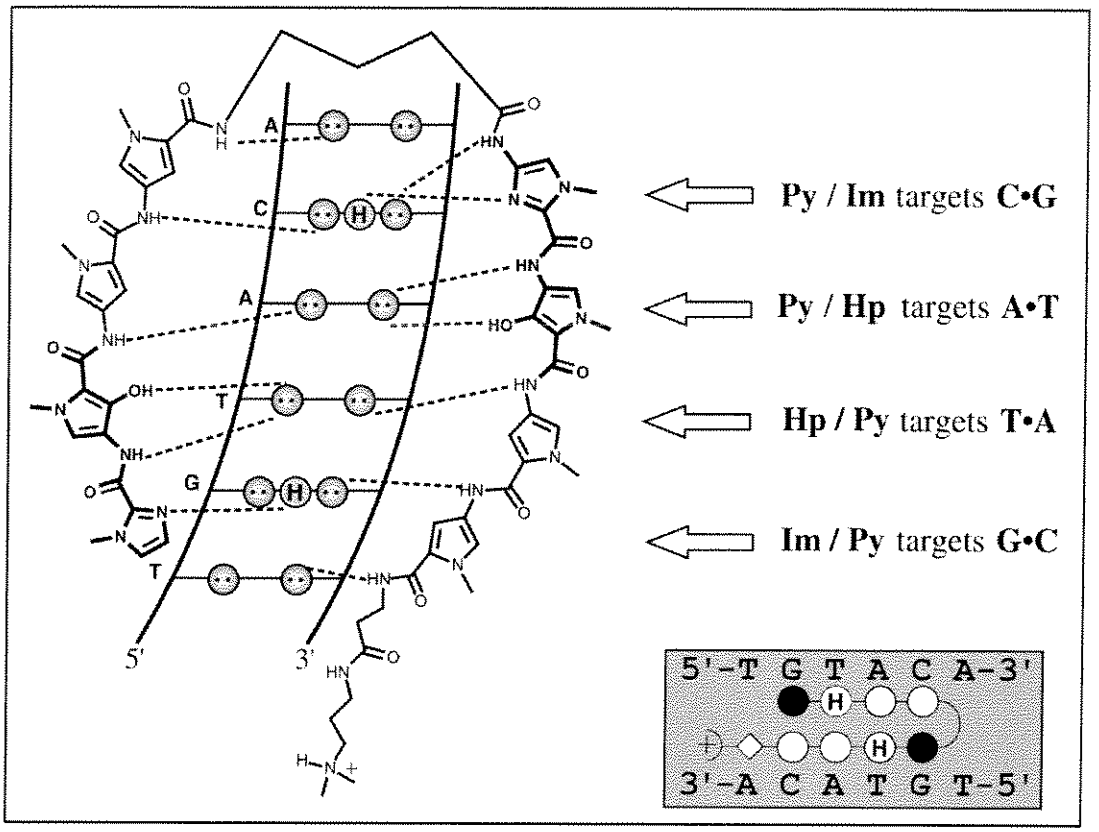


Figure 1.12 Recognition of all four Watson-Crick base pairs in the minor groove by hydroxypyrrole-imidazole-pyrrole polyamides. Circles with two dots represent hydrogen bond acceptors present on the floor of the minor groove. Circles with an H represent the 2-amino group of guanine, a hydrogen bond donor. Dotted lines represent putative hydrogen bonds. (inset) Ball and stick model of the hairpin polyamide-DNA complex. Black and white circles represent imidazole and pyrrole amino acids, respectively. Circles with an H represent hydroxypyrrole. Diamond represents β-alanine residue, while curved line represents the γ-aminobutyric acid. The positive charge represents the dimethylaminopropylamide tail group.

References

1. (a) Zimmer, C.; Wahnert, U. *Prog. Biophys. Mol. Biol.* **1986**, *47*, 31-112. (b) Dervan, P.B. *Science* **1986**, *232*, 464. (c) Moser, H.E.; Dervan, P.B. *Science* **1987**, *238*, 645. (d) Thuong, N.T.; Helene, C. *Angew. Chem. Int. Edn. Engl.* **1993**, *32*, 666. (e) Nielsen, P.E. *Chem. Eur. J.* **1997**, *3*, 505.

2. Watson, J.D. *Gene* **1993**, *135*, 309.

3. Roeder, R.G. *TIBS* **1996**, *9*, 327.

4. Gottesfeld, J.M.; Neely, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1997**, *387*, 202.

5. Watson, J.D.; Crick, F.H.C. *Nature* **1953**, *171*, 737.

6. Dickerson, R.E.; Drew, H.R.; Conner, B.N.; Wing, M.; Fratini, A.K.; Koopka, M.L. *Science* **1982**, *216*, 475.

7. Saenger, W. In Principles of Nucleic Acid Structure; Springer-Verlag: New York, 1984.

8. Steitz, T.A. *Quart. Rev. Biophys.* **1990**, *23*, 205.

9. (a) Pabo, C.O.; Sauer, R.T. *Annu. Rev. Biochem.* **1992**, *61*, 1053. (b) Freemont, P.S.; Lane, A.N.; Sanderson, M.R. *Biochem. J.* **1991**, *278*, 1. (c) Kim, Y.; Geiger, J.H.; Hahn, S.; Sigler, P.B. *Nature* **1993**, *365*, 512. (d) Kim, J.L.; Burley, S.K. *Nature Struct. Biol.* **1994**, *1*, 638. (e) Nielsen, P.E. *Bioconjugate Chem.* **1991**, *2*, 1.

10. Pavletich, N.P.; Pabo, C.O. *Science* **1991**, *252*, 809.

11. Ellenberger, T.E.; Brandl, C.J.; Struhl, K.; Harrison, S.C. *Cell* **1992**, *71*, 1223.

12. Feng, J.A.; Johnson, R.C.; Dickerson, R.E. *Science* **1994**, *263*, 348.

13. Kissinger, C.R.; Liu, B.; Martin-Blanco, E.; Kornberg, T.B.; Pabo, C.O. *Cell* **1990**, *63*, 579.

14. Choo, Y.; Klug, A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 117.

15. (a) Gao, X.L.; Mirau, P.; Patel, D.J. *J. Mol. Biol.* **1992**, *223*, 259. (b) Kamitori, S.; Takusagawa, F. *J. Mol. Biol.* **1992**, *225*, 445. (c) Paloma, L.G.; Smith, J.A.; Chazin, W.J.; Nicolaou, K.C. *J. Am. Chem. Soc.* **1994**, *116*, 3697.

16. Coll, M.; Frederick, C.A.; Wang, A.H.J.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8385.

17. (a) Zimmer, C. *Prog. Nucleic Acid Res. Mol. Biol.* **1975**, *15*, 285. (b) Baguley, B.C. *Molecular and Cellular Biochem.* **1982**, *43*, 167.

18. (a) Pelton, J.G.; Wemmer, D.E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723. (b) Pelton, J.G.; Wemmer, D.E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S.T.; Sundaralingham, M. *Nature Struct. Biol.* **1994**, *1*, 169. (d) Kielkopf, C.L.; Baird, E.E.; Dervan, P.B.; Rees, D.C. *Nature Struct. Biol.* **1998**, *5*, 104.

19. (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.

20. White, S.; Baird, E.E.; Dervan, P.B. *Biochemistry* **1996**, *35*, 12532.

21. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 9892. (b) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646. (c) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 3325.

22. (a) 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.; Lang, 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.; Pokar, N.A.; Gelfand, C.A.; Law, S.M.; Breslauer, K.J.; Baird, E.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306. (h) White, S.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, *119*, 8756. (i) White, S.; Baird, E.E.; Dervan, P.B. *Chem. & Biol.* **1997**, *4*, 569.

25. De Claire, R.P.L.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1997**, *119*, 7909.

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

27. Seeman, N.C.; Rosenberg, J.M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804.

28. Swalley, S.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, *119*, 6953.

29. Wong, J.M.; Bateman, E. *Nucl. Acids Res.* **1994**, *22*, 1890.

30. White, S.; Szewczyk, J.W.; Turner, J.M.; Baird, E.E.; Dervan, P.B. *Nature* **1998**, *391*, 468.

31. Kielkopf, C.L.; White, S.; Szewczyk, J.W.; Turner, J.M.; Baird, E.E.; Dervan, P.B.; Rees, D.C. *manuscript in preparation.*

Chapter 2

Orientation Preferences of Pyrrole-Imidazole Polyamides in the Minor Groove of DNA

Abstract *In order to determine whether there is an orientation preference of pyrrole-imidazole (Py-Im) polyamide dimers with respect to the 5'-3' direction of the backbone in the DNA helix, equilibrium association constants (K_a) were determined for a series of six-ring hairpin polyamides which differ with respect to substitution at the N and C termini. Affinity cleaving experiments using hairpin polyamides of core sequence composition ImPyPy-γ-PyPyPy with an EDTA•Fe(II) moiety at the C-terminus reveal a single binding orientation at each formal match site, 5'-(A,T)G(A,T)$_3$-3' and 5'-(A,T)C(A,T)$_3$-3'. A positive charge at the C-terminus and no substitution at the N-terminus imidazole affords the maximum binding orientation preference, calculated from K_a(5'-TGTTA-3')/K_a(5'-TCTTA-3'), with the N-terminal end of each three-ring subunit located toward the 5' side of the target DNA strand. Removal of the positive charge, rearrangement of the positive charge to the N-terminus or substitution at the N-terminal imidazole decreases the orientation preference. These results suggest that second generation design principles superimposed on the simple pairing rules can further optimize the sequence-specificity of Py-Im polyamides for double helical DNA.*

Publication: White, Baird & Dervan *J. Am. Chem. Soc.* **1996**, *118*, 6147-6152.

Polyamides containing pyrrole (Py) and imidazole (Im) amino acids bind cooperatively as antiparallel dimers in the minor groove of the DNA helix.[1,2] Sequence-specificity depends on the side-by-side pairings of *N*-methylpyrrole and *N*-methylimidazole amino acids.[1] A pairing of Im opposite Py targets a G•C base-pair, while Py opposite Im targets a C•G base-pair.[1] A pyrrole/pyrrole combination is degenerate and targets both T•A and A•T base-pairs.[2] Py-Im polyamides have been shown to be cell permeable and to inhibit the transcription of genes in cell culture.[3] This provides impetus to develop second generation polyamide design rules that provide for enhanced sequence-specificity and perhaps optimal biological regulation.

Although the polyamides bind DNA antiparallel to each other, the "pairing rules" do not distinguish whether there should be any energetic preference for alignment of each polyamide (N-C) with respect to the backbone (5'-3') of the DNA double helix (Figure 2.1).

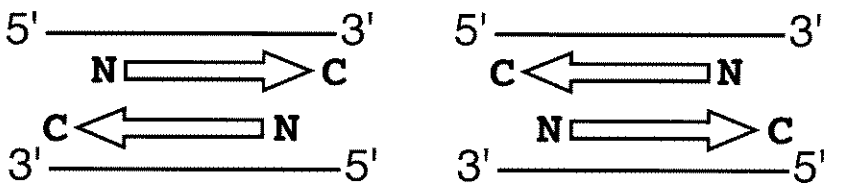


Figure 2.1 Antiparallel polyamide subunits are depicted as filled arrows. Arrowheads correspond to C-terminal end of the polyamide. (a) Polyamide binding with N-terminal end located towards 5'-side of the targeted DNA strand. (b) Binding with the C-terminal end of the polyamide located towards 5'-side of binding site.

In a formal sense the homodimer $(ImPyPy)_2$ could bind 5'-WGWCW-3' or 5'-WCWGW-3' and still not violate the binary code. Remarkably, even in the first report on the binding specificity of the three ring polyamide ImPyPy-Dp, there were qualitative data to suggest that there was indeed a binding preference 5'-WGWCW-3' > 5'-WCWGW-3'.[1a,4] This suggested that pyrrole-imidazole polyamide dimers align N-C with the 5'-3' direction of the DNA strand. This orientation preference superimposed on the pairing rules confers added

specificity by breaking a potential degeneracy for recognition. It would be useful to find out whether this preference is general and which aspects of the ligand design control the energetics of orientation preference. Therefore, we describe here a study to address the influence on orientation of (1) positive charge or lack of, (2) position of the positive charge at the N or C-terminus, and (3) substitution of the terminal imidazole.

Three-ring polyamide subunits covalently coupled by a γ-aminobutyric acid linker form 6-ring hairpin structures that bind to 5-bp target sequences with enhanced affinity and specificity relative to the unlinked polyamide pair.[5] In principle, a hairpin polyamide:DNA complex can form at two different DNA sequences depending on the N-C alignment of the polyamide with the walls of the minor groove of DNA (5'-3'). A six-ring hairpin polyamide of core sequence composition ImPyPy-γ-PyPyPy which places the N-terminus of each three-ring polyamide subunit at the 5'-side of each recognized DNA strand would bind 5'-TGTTA-3'. Placement of the polyamide N-terminus at the 3' side of each recognized strand would result in targeting of a 5'-TCTTA-3' sequence (Figure 2.2).

Four six-ring hairpin polyamides, ImPyPy-γ-PyPyPy-β-Dp **1**, ImPyPy-γ-PyPyPy-β-EtOH **2**, Ac-ImPyPy-γ-PyPyPy-β-Dp **3**, and Dp-ImPyPy-γ-PyPyPy-β-Me **4** were synthesized by solid phase methods (Figure 2.3).[6] The corresponding EDTA analogs ImPyPy-γ-PyPyPy-β-Dp-EDTA **1-E**, ImPyPy-γ-PyPyPy-β-C7-EDTA **2-E**, Ac-ImPyPy-γ-PyPyPy-β-Dp-EDTA **3-E**, and Dp-ImPyPy-γ-PyPyPy-β-C7-EDTA **4-E** were also constructed in order to confirm a single orientation of each hairpin:DNA complex. We report here the DNA-binding affinity, orientation, and sequence-selectivity of the four polyamides for the two match five base pair binding sites, 5'-TGTTA-3' and 5'-TCTTA-3'.

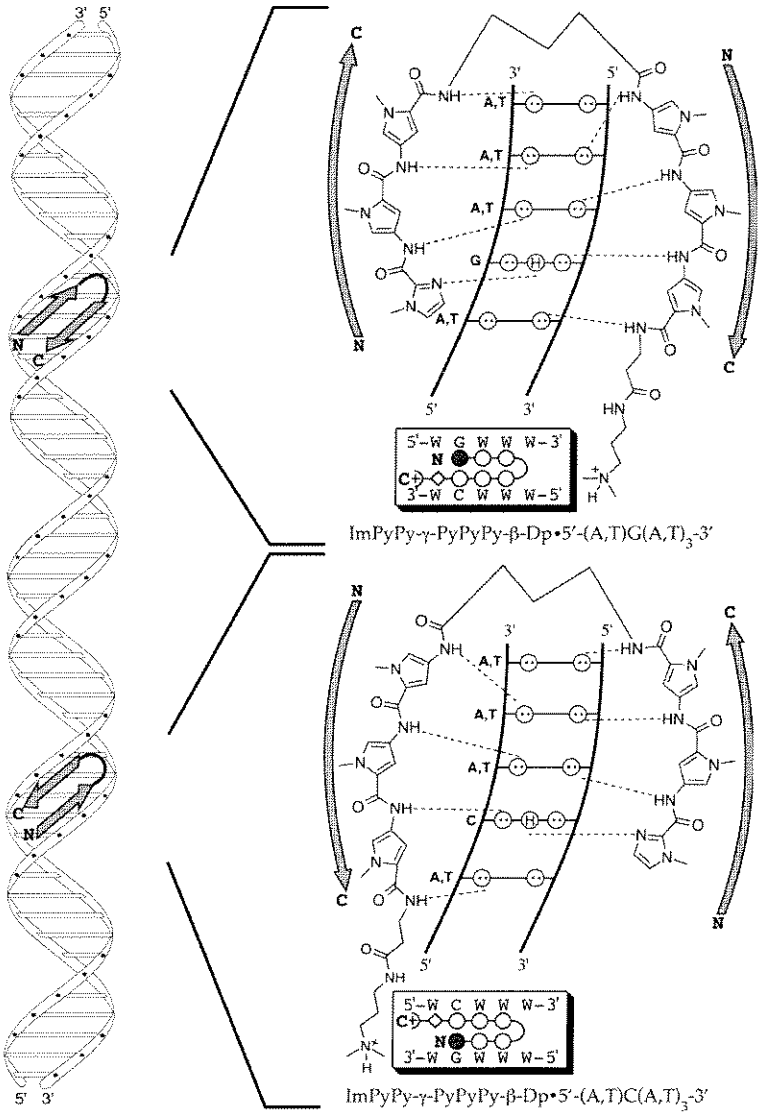


Figure 2.2 Binding models for ImPyPy-γ-PyPyPy-β-Dp in complex with 5'-WGWWW-3' (Top) and 5'-WCWWW-3' (Bottom) (W = A or T). Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. For schematic binding models, the imidazole and pyrrole rings are represented as shaded and unshaded spheres respectively, the curved line represents γ-aminobutyric acid, and the β-alanine residue is represented as an unshaded diamond. Shaded arrows represent the orientation of individual polyamide subunits. Arrowheads represent the polyamide C-terminus. (left) A ribbon model depicting the hairpin structure bound in the minor groove of the DNA helix with either the N or C-terminus located at the 5' side of the binding site.

Three separate techniques are used to characterize the DNA-binding properties of the polyamides: affinity cleaving[7] and MPE•Fe(II)[8] and DNase I[9] footprinting. Affinity cleavage studies determine the specific binding orientation and stoichiometry of each hairpin:DNA complex. Binding site size is accurately determined by MPE•Fe(II) footprinting, while quantitative DNase I footprint titration is more suitable for measurement of equilibrium association constants (K_a) for the polyamide binding to designated sequences.

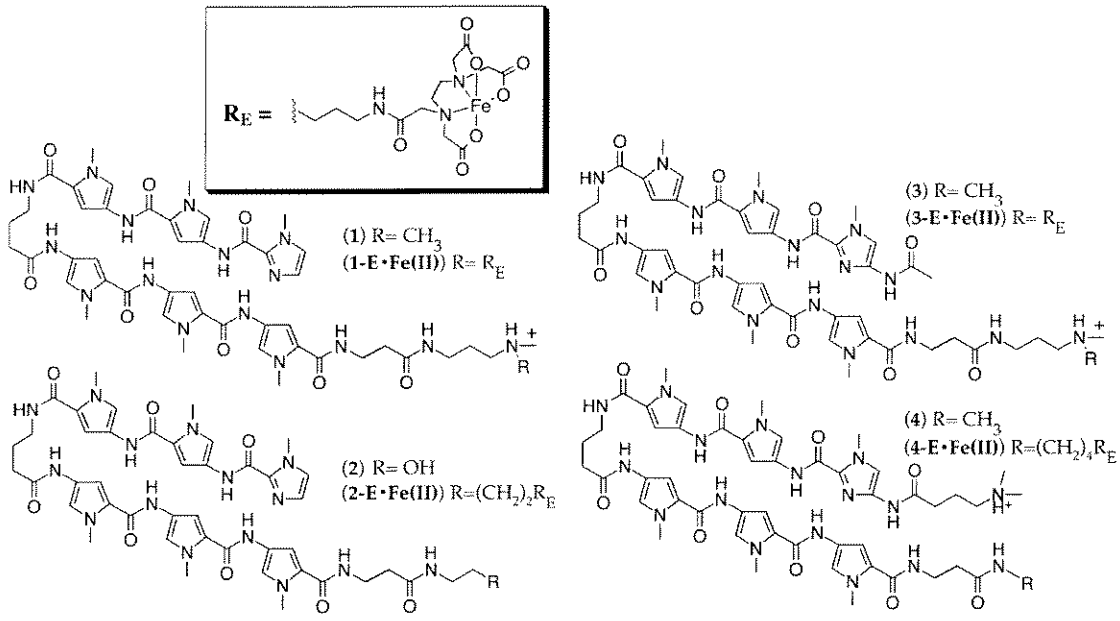


Figure 2.3 Structure of the hairpin polyamides **1-4** and the corresponding EDTA-modified derivatives; ImPyPy-γ-PyPyPy-β-Dp (**1**), ImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II) (**1-E**), ImPyPy-γ-PyPyPy-β-EtOH (**2**), ImPyPy-γ-PyPyPy-β-C7-EDTA (**3-E**), AcImPyPy-γ-PyPyPy-β-Dp (**3**), AcImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II) (**3-E**), Dp-ImPyPy-γ-PyPyPy-β-Me (**4**), Dp-ImPyPy-γ-PyPyPy-β-C7-EDTA (**4-E**).

Results

Synthesis. Polyamides were synthesized from Boc-β-alanine-Pam-Resin (0.2 mmol/gram substitution) using stepwise solid phase methods (Figure 2.4).[6] A C-terminal β-alanine residue facilitates solid phase synthesis and increases hairpin-polyamide DNA-binding affinity and sequence specificity.[5b] The synthesis of polyamides **1** and **3** has been described.[5b] A sample of ImPyPy-γ-PyPyPy-β-Pam-Resin was cleaved by a single-step aminolysis reaction with neat ethanolamine (55 °C, 16 h) to provide polyamide **2** after HPLC purification. Cleavage of Dp-ImPyPy-γ-PyPyPy-β-Resin with a satd. solution of methylamine in DMF (55 °C, 80 psi, 48 h) provided polyamide **4** after HPLC purification. For the synthesis of analogs modified with EDTA, a sample of resin was cleaved with 3,3'-diamino-*N*-methyldipropylamine or 1,7-diaminoheptane (55 °C, 18h) for polyamides **1-E** and **3-E** and polyamides **2-E** and **4-E** respectively. The amine modified polyamides were purified by reverse phase HPLC and then treated with an excess of the dianhydride of EDTA (DMSO/NMP, DIEA, 55 °C, 30 min) and the remaining anhydride hydrolyzed (0.1 M NaOH, 55 °C, 10 min). The EDTA modified polyamides **1-E, 2-E, 3-E** and **4-E** were then isolated by reverse phase HPLC. Polyamides were characterized by a combination of analytical HPLC, ^{1}H NMR spectroscopy, and MALDI-TOF mass spectroscopy.

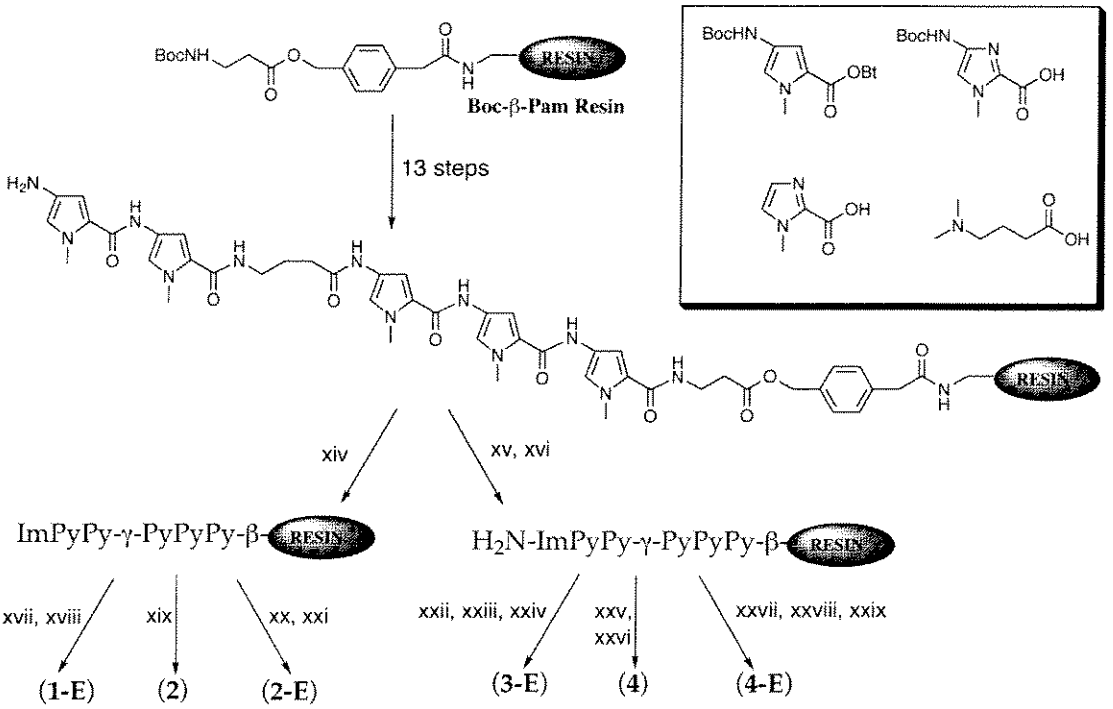


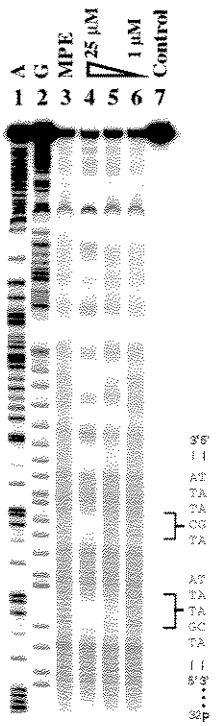
Figure 2.4 (Box) Pyrrole and imidazole monomers used for synthesis of the polyamides described here; Boc-Pyrrole-OBt ester **5**, Boc-Imidazole-acid **6** and imidazole-2-carboxylic acid **7**.[1a] Solid phase synthetic scheme for ImPyPy-γ-PyPyPy-β-Dp-EDTA (**1-E**), ImPyPy-γ-PyPyPy-β-EtOH (**2**), ImPyPy-γ-PyPyPy-C7-EDTA (**2-E**), AcImPyPy-γ-PyPyPy-β-Dp-EDTA (**3-E**), Dp-ImPyPy-γ-PyPyPy-β-Me (**4**), and Dp-ImPyPy-γ-PyPyPy-β-C7-EDTA (**4-E**). Synthesis is initiated from commercially available Boc-β-alanine-Pam-resin (0.2 mmol/gram): (i) 80% TFA/DCM, 0.4M PhSH; (ii) BocPy-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4M PhSH; (iv) BocPy-OBt, DIEA, DMF; (v) 80% TFA/DCM, 0.4M PhSH; (vi) BocPy-OBt, DIEA, DMF; (vii) 80% TFA/DCM, 0.4M PhSH; (viii) Boc-γ-aminobutyric acid (HBTU, DIEA), DMF; (ix) 80% TFA/DCM, 0.4M PhSH; (x) BocPy-OBt, DIEA, DMF; (xi) 80% TFA/DCM, 0.4M PhSH; (xii) BocPy-OBt, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4M PhSH; (xiv) imidazole-2-carboxylic acid (HBTU/DIEA); (xv) BocIm-OBt (DCC/HOBt), DIEA, DMF; (xvi) 80% TFA/DCM, 0.4M PhSH; (xvii) 3,3′-diamino-*N*-methyldipropylamine, 55 °C; (xviii) EDTA-dianhydride, DMSO/NMP, DIEA, 55 °C; 0.1M NaOH; (xix) dimethylaminopropylamine, 55 °C; (xx) 1,7-diaminoheptane, 55 °C; (xxi) EDTA-dianhydride, DMSO/NMP, DIEA, 55 °C; 0.1M NaOH (xxii) Acetic Anhydride, DMF, DIEA; (xxiii) 3,3′-diamino-*N*-methyldipropylamine, 55 °C; (xxiv) EDTA-dianhydride, DMSO/NMP, DIEA, 55 °C; 0.1M NaOH; (xxv) dimethylamino-γ (**8**), (HBTU/DIEA); (xxvi) $MeNH_2$ (satd.), DMF, 80 psi, 55 °C (xxvii) dimethylamino-γ (**8**), (HBTU/DIEA); (xxviii) 1,7-diaminoheptane, 55 °C; (xxix) EDTA-dianhydride, DMSO/NMP, DIEA, 55 °C; 0.1M NaOH.

Identification of Binding Sites by MPE•Fe(II) Footprinting. MPE•Fe(II) footprinting experiments were used to determine precise binding site location and size.[8] Footprinting on the 3' and 5'-^{32}P end labeled 266 base pair *Eco*RI/*Pvu*II restriction fragment from the plasmid pDEH2 (25 mM tris-acetate, 10 mM NaCl, 100 μM bp calf thymus DNA, 5 mM DTT, pH 7.0 and 22 °C) reveals that the polyamides are binding to the two match binding sites, 5'-TGTTA-3' and 5'-TCTTA-3' (Figure 2.5 and 2.6). Polyamides **1** and **2** at 5 μM concentration bind to the 5'-TGTTA-3' site and bind weakly to the 5'-TCTTA-3' site. Polyamides **3** and **4** at 10 μM concentration bind to both the 5'-TGTTA-3' site and the 5'-TCTTA-3' site.[10] The size of the footprint cleavage protection patterns for the polyamides are consistent with the expected binding location and 5-bp binding site size.

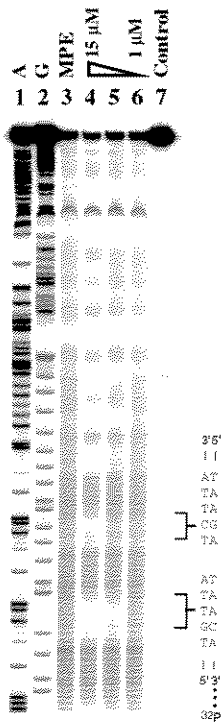
Identification of Binding Orientation and Stoichiometry by Affinity Cleaving. Affinity cleaving titration experiments[7] using hairpin polyamides modified with EDTA•Fe(II) at the C-terminus, were used to determine polyamide binding orientation and stoichiometry. Affinity cleaving experiments were performed on a 3' and 5'-^{32}P end labeled 266 base *Eco* RI/*Pvu* II restriction fragment from the plasmid pDEH2 (25 mM tris-acetate, 20 mM NaCl, 100 μM bp calf thymus DNA, pH 7, 22 °C, 10 mM DTT, 10 μM Fe(II)) (Figure 2.7 and 2.8). The observed cleavage patterns are in all cases 3'-shifted, consistent with minor groove occupancy. A single cleavage locus proximal to the 5'-side of both the 5'-TGTTA-3' and 5'-TCTTA-3' binding sites is consistent with a single orientation at each site (Figure 2.9). A number of additional strong cleavage sites present on the DNA restriction fragment which correspond to sequences of the form 5'-(A,T)G(A,T)$_3$-3' were not resolved.[2d]

Figure 2.5 MPE•Fe(II) footprinting experiments on the 3′-^{32}P-labeled 266-bp *Eco* RI/*Pvu* II restriction fragment from plasmid pDEH2. The 5′-TGTTA-3′, and 5′-TCTTA-3 sites are shown on the right side of the autoradiogram.[21] Lane 1, A reaction; lane 2, G reaction; lane 3 MPE•Fe(II) standard; lane 7, intact DNA; lanes 4-6: 25 μM, 5 μM, 1 μM ImPyPy-γ-PyPyPy-β-Dp (**1**), 15 μM, 5 μM, 1 μM ImPyPy-γ-PyPyPy-β-C2-OH (**2**), 25 μM, 10 μM, 1 μM AcImPyPy-γ-PyPyPy-β-Dp (**3**), or 30μM, 10 μM, 1 μM Dp-ImPyPy-γ-PyPyPy-β-Me (**4**). All lanes contain 15 kcpm 3′-radiolabeled DNA, 25 mM Tris-acetate buffer (pH 7.0), 10 mM NaCl, and 100 μM/base pair calf thymus DNA.

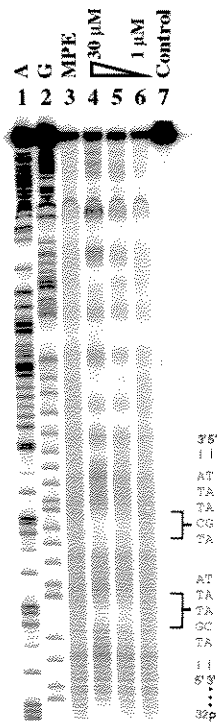
ImPyPy-γ-PyPyPy-β-Dp



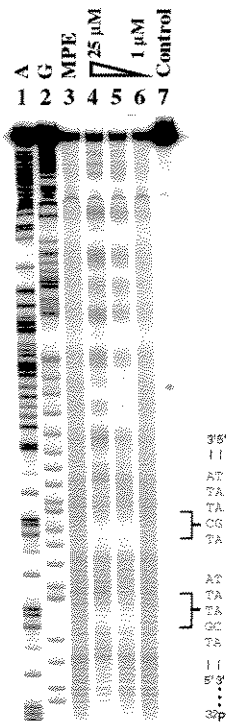
ImPyPy-γ-PyPyPy-β-EtOH



AcImPyPy-γ-PyPyPy-β-Dp



Dp-ImPyPy-γ-PyPyPy-β-Me



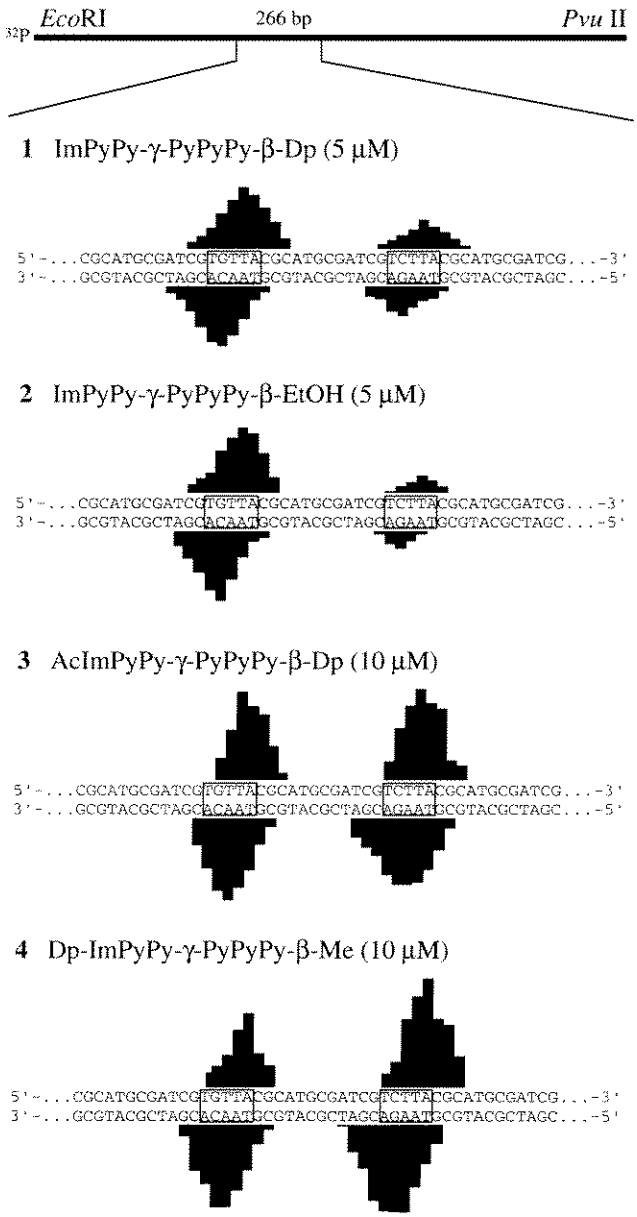
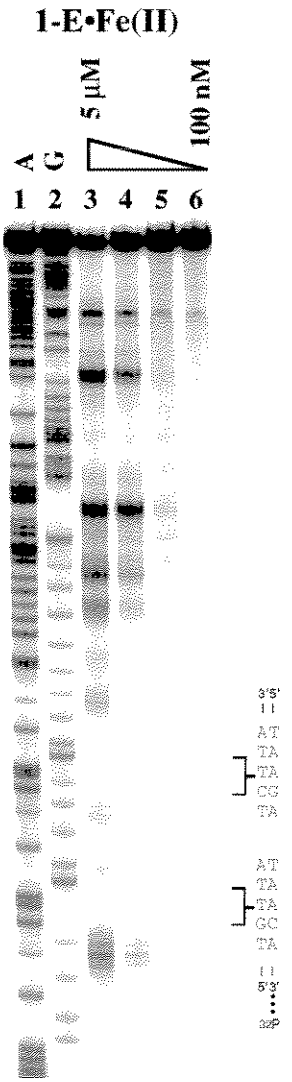
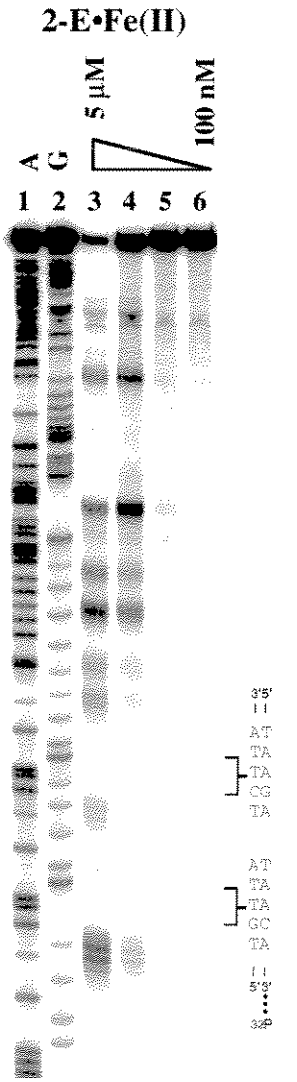


Figure 2.6 (Top) Illustration of the 266 bp restriction fragment with the position of the sequence indicated. (Bottom) MPE•Fe(II) protection patterns of ImPyPy-γ-PyPyPy-β-Dp (**1**) or ImPyPy-γ-PyPyPy-β-EtOH (**2**) for 5 μM concentration and AcImPyPy-γ-PyPyPy-β-Dp (**3**), or Dp-ImPyPy-γ-PyPyPy-β-Me (**4**) for 10 μM concentration. Bar heights are proportional to the relative protection from cleavage at each band.

Figure 2.7 Storage phosphor autoradiogram of 8% denaturing polyacrylamide gels used to separate the fragments[21] generated by affinity cleaving experiments performed with (a) ImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II) and (b) AcImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II): lanes 1 and 2, A and G sequencing lanes; lanes 3-6, digestion products obtained in the presence of **2**: 5 μM, 2.5 μM, 1 μM, and 100 nM polyamide. The targeted binding sites are indicated on the right side of the autoradiograms. All reactions contain 20 kcpm 3'-^{32}P restriction fragment, 25 mM tris-actetate, 20 mM NaCl, 100 μM bp calf thymus DNA, pH 7.


1-E•Fe(II)
A
G
5 μM
100 nM
1 2 3 4 5 6
3'5'
AT
TA
TA
CG
TA
AT
TA
TA
GC
TA
5'3'
32P


2-E•Fe(II)
A
G
5 μM
100 nM
1 2 3 4 5 6
3'5'
AT
TA
TA
CG
TA
AT
TA
TA
GC
TA
5'3'
32P

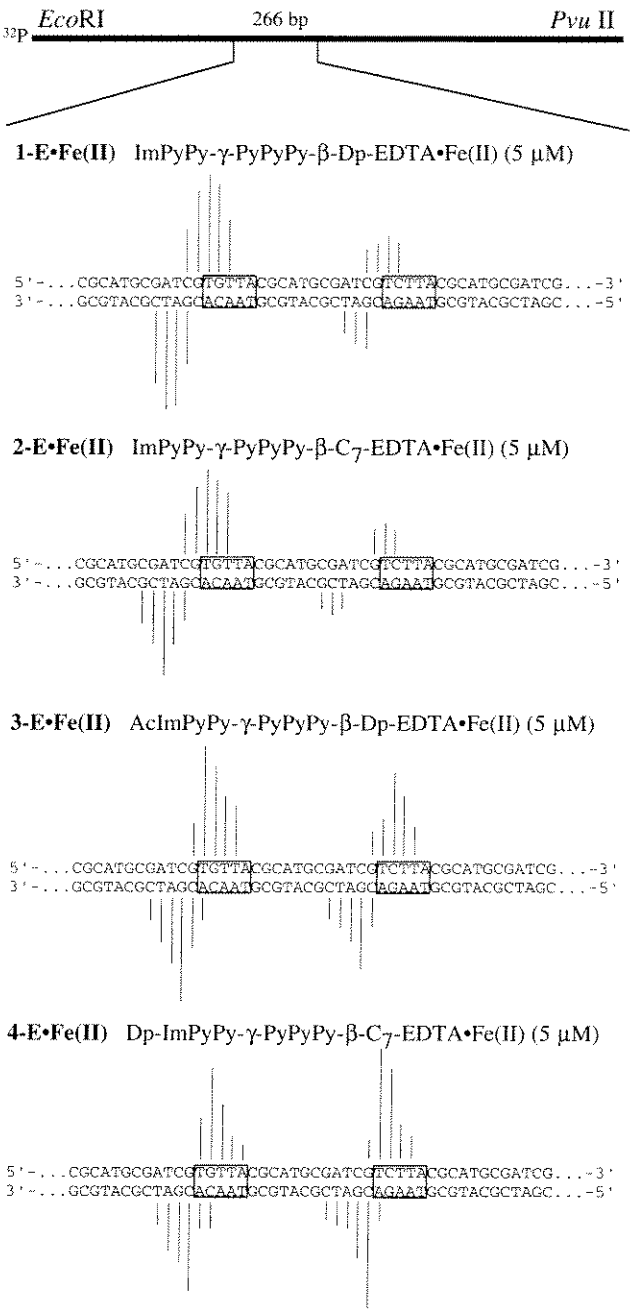


Figure 2.8 Results from affinity cleavage with ImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II) (**1-E**), ImPyPy-γ-PyPyPy-β-C7-EDTA (**2-E**) AcImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II) (**3-E**), and Dp-ImPyPy-γ-PyPyPy-β-C7-EDTA (**4-E**) at 5 μM concentration. (Top) Illustration of the 266 bp restriction fragment with the position of the sequence indicated. Only the two designated target sites are boxed. Arrow heights are proportional to the relative cleavage intensities at each base pair.

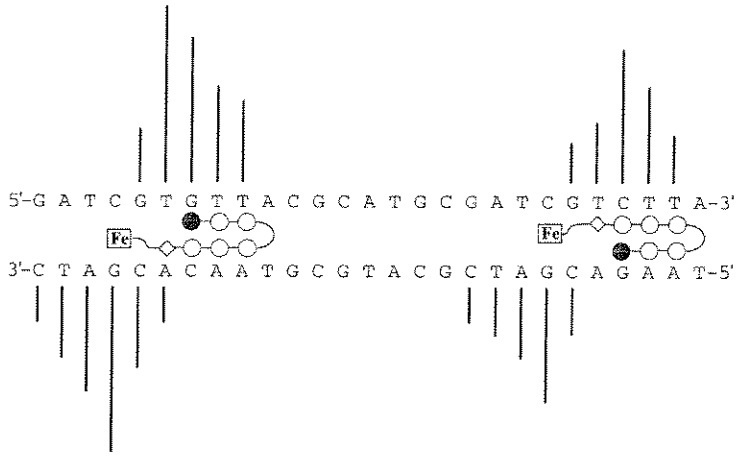


Figure 2.9 Affinity cleavage patterns and ball and stick models of the six-ring EDTA•Fe(II) analog **1-E•Fe(II)** bound to the match sites, 5'-TGTTA-3' and 5'-TCTTA-3'. Bar heights are proportional to the relative cleavage intensities at each base pair. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent the β-alanine residue. The boxed **Fe** denotes the EDTA•Fe(II) cleavage moiety. Cleavage patterns show a single binding orientation at each match site.

Analysis of Energetics by Quantitative DNase I Footprint Titrations. Quantitative DNase I footprint titration experiments[9] (10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5mM $CaCl_2$, pH 7.0, 22 °C) were performed to determine the equilibrium association constants K_a for recognition of the bound sites (Figure 2.10). The 5'-TGTTA-3' site is bound by the polyamides with decreasing affinity: ImPyPy-γ-PyPyPy-β-Dp **1** > ImPyPy-γ-PyPyPy-β-EtOH **2** > AcImPyPy-γ-PyPyPy-β-Dp **3** > Dp-ImPyPy-γ-PyPyPy-β-Me **4**. The 5'-TCTTA-3' site is bound with decreasing affinity: Dp-ImPyPy-γ-PyPyPy-β-Me **4** > AcImPyPy-γ-PyPyPy-β-Dp **3** > ImPyPy-γ-PyPyPy-β-EtOH **2** > ImPyPy-γ-PyPyPy-β-Dp

1. Remarkably, the ratio of association constants for each site varies from 16 to 1-fold between the four polyamides indicates a sensitivity to substitution at the N and C terminus (Table 2.1).

Table 2.1 Equilibrium Association Constants (M^{-1}) [a, b]

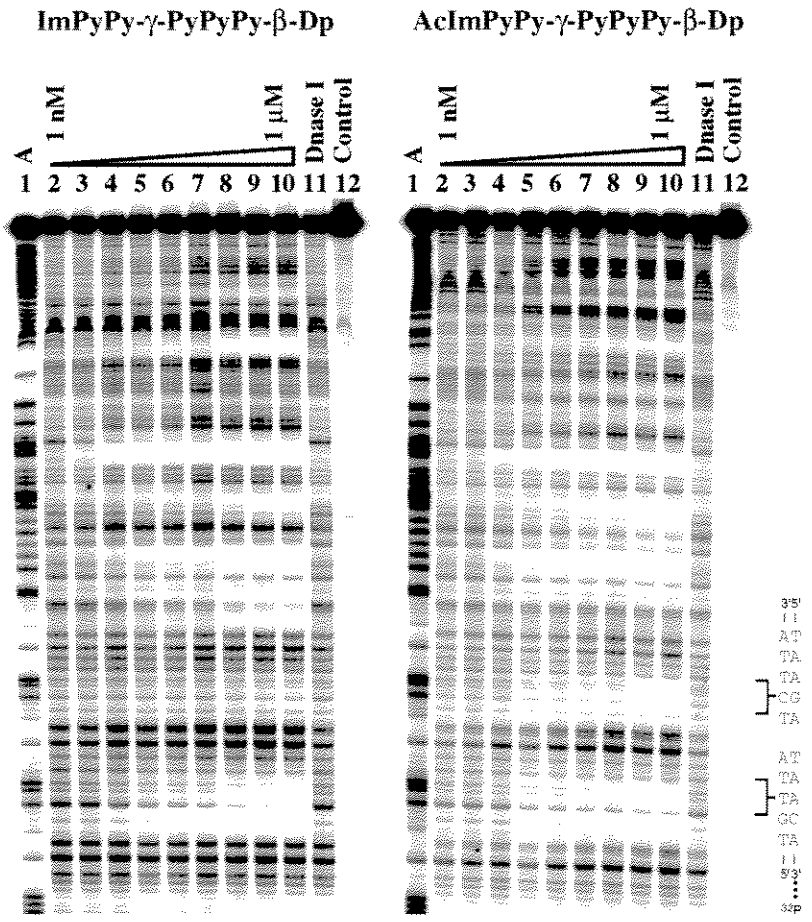
Polyamide (N→C)		5'-TGTTA -3'	5'-TCTTA-3'	Orientation[c]
ImPyPy-γ-PyPyPy-β-Dp	**1**	1.4×10^7 (0.4)	8.8×10^5 (1.0)	16
ImPyPy-γ-PyPyPy-β-EtOH	**2**	1.1×10^7 (0.2)	1.2×10^6 (1.0)	9
AcImPyPy-γ-PyPyPy-β-Dp	**3**	7.6×10^6 (1.1)	1.9×10^6 (1.2)	4
Dp-ImPyPy-γ-PyPyPy-β-Me	**4**	7.2×10^6 (1.4)	9.1×10^6 (0.1)	0.8

[a]Values reported are the mean values measured from at least three footprint titration experiments, with the standard deviation for each data set indicated in parentheses. [b]The assays were performed at 22 °C at pH 7.0 in the presence of 10 mM tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. Orientation preference calculated from K_a(5'-TGTTA-3')/K_a(5'-TCTTA-3')

Discussion.

Binding Site Size and Orientation. MPE•Fe(II) footprinting reveals that the four polyamides of core sequence composition ImPyPy-γ-PyPyPy bind with high affinity to both match sites 5'-TGTTA-3' and 5'-TCTTA-3'. Affinity cleavage experiments using polyamides with Fe(II)•EDTA at the carboxy terminus confirm that the four polyamides bind each discreet site with a single orientation (Figure 2.8 and 2.9). Asymmetric 3'-shifted cleavage patterns are consistent with the location of the 1:1 polyamide:DNA complex in the minor groove. The observation of a single cleavage locus is consistent only with an oriented 1:1 complex and rules out any 2:1 overlapped or extended binding motifs.[11] A 1:1 oriented but extended motif would require at least an eight base pair binding site, which is inconsistent with high-resolution MPE footprinting data on both target sites.

Figure 2.10 Quantitative DNase I footprint titration experiment with ImPyPy-γ-PyPyPy-β-Dp (**1**) and AcImPyPy-γ-PyPyPy-β-Dp on the 3'-^{32}P-labeled 266 bp *Eco* RI/ *Pvu* II restriction fragment from plasmid pDEH2. The binding sites 5'-TGTTA-3' and 5'-TCTTA-3' are shown on the right side of the storage phosphor autoradiogram.[21] All reactions contain 20 kcpm restriction fragment, 10 mM tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0. Lane 1, A reaction; lane 2, G reaction; lanes 3 and 21, DNase I standard; lanes 4-20 contain 0.1 nM, 10 nM, 65 nM, 100 nM, 150 nM, 250 nM, 500 nM, 750 nM, and 1 μM polyamide respectively.


ImPyPy-γ-PyPyPy-β-Dp
AcImPyPy-γ-PyPyPy-β-Dp
A
1 nM
1 μM
Dnase I
Control
1 2 3 4 5 6 7 8 9 10 11 12
3'
AT
TA
TA
CG
TA
AT
TA
TA
GC
TA
5'
32p

The hairpin structure is supported by direct NMR structure studies on a six-ring hairpin polyamide of sequence composition ImPyPy-γ-PyPyPy binding a core five base pair 5'-TGTTA-3' site.[5c]

A single cleavage locus is observed proximal to the 5'-side of both the 5'-TGTTA-3' and 5-TCTTA-3' binding sites, indicating that the carboxy terminus of the polyamide is located at the 5'-side of each site. Furthermore, the relative location of the observed cleavage maxima is unchanged for recognition of 5'-TGTTA-3' and 5'-TCTTA-3', indicating similar placement of the polyamide C-termini at both sites. *These results indicate that all four polyamides may adapt two unique binding orientations.* Each binding orientation may represent a unique and distinguishable hairpin fold (Figure 2.2).[12]

Binding Affinity. Four polyamides of core sequence composition ImPyPy-γ-PyPyPy-β, but varying at the N- and C- terminus bind both match sites 5'-TGTTA-3' and 5'-TCTTA-3' as a 1:1 complex (eq. 2, n = 1) consistent with the hairpin motif. However, the relative discrimination between the two sites varies by 16-fold. Among the four ligands, polyamide **1** binds the 5'-TGTTA-3' site with the highest affinity ($K_a = 1.4 \times 10^7$ M^{-1}) and orientation specificity (16-fold). Replacement of the charged dimethylaminopropylamide tail group with an uncharged ethoxyamide group as in polyamide **2** results in a negligible decrease in affinity ($K_a = 1.1 \times 10^7$ M^{-1}) but reduced specificity (9-fold). The decrease in orientational specificity indicates that the cationic tail group is necessary but not sufficient for optimal oriented polyamide binding.

In previous work we have observed that N-terminal acetylation reduces both the binding affinity and the sequence preference for target sites.[5b] The acetylated polyamide **3** binds the 5'-TGTTA-3' site with reduced affinity and orientational specificity compared to **1**. Replacement of the N-terminal acetyl group with a charged dimethylaminopropyl group results in a 4-fold increase in affinity at the 5'-TCTTA-3' site but a complete loss of

orientational specificity for polyamide **4**. These results indicate that the N or C terminus position of charged substituents plays an important role in orientation preference.

Subunit Dipoles. Dipole-dipole interactions could potentially account in part for the observed DNA-binding orientation preference (Figure 2.11). Calculated dipoles[13] for the

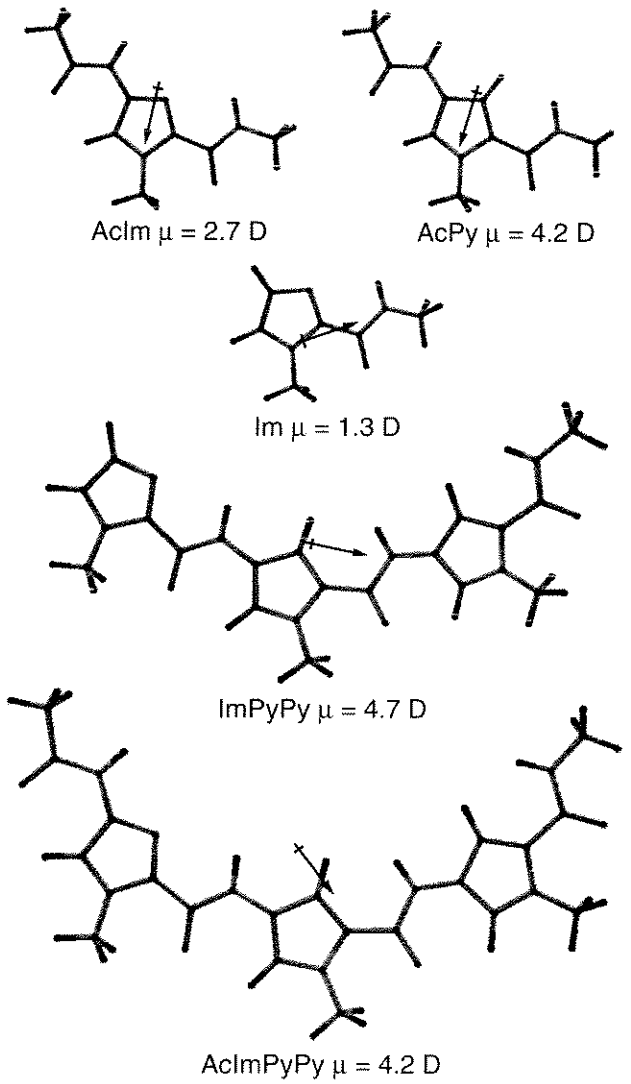


Figure 2.11 Dipoles calculated for individual polyamide monomers and subunits. All C-termini are the methylamide derivatives, all N-termini are either unmodified, Im and ImPyPy or acetamide derivatives, AcIm, AcPy, and AcImPyPy. Arrows depict direction of dipole moments.

individual monomer units AcIm, AcPy, and Im, as well as the three-ring subunits ImPyPy and AcImPyPy reveal that only monomers and subunits containing an unsubstituted Im residue

could have a dipole antiparallel to the walls of the helix. Subunit dipoles could stabilize side-by-side antiparallel placement of polyamide subunits in the minor groove. In addition, dipole interactions between the DNA backbone and the polyamide could lead to 5' to 3' aligned with N to C oriented binding *if there exists a net dipole on each antiparallel strand of the DNA double helix.*[14]

Induced Chirality. Pyrrole-imidazole polyamides are achiral molecules which form chiral complexes with DNA. In principle there exists two non-superimposable hairpin folds which are related by mirror plane symmetry (Figure 2.2). Fold 1 is responsible for the preferred 5' to 3' N to C orientation for pyrrole-imidazole polyamide DNA recognition. Fold 2 corresponds to the 3' to 5' N to C recognition observed here. In the absence of DNA each fold should be energetically equivalent. Upon binding to DNA, adjacent pyrrole and imidazole monomers twist to accommodate the right handed B-form DNA helix. Twisting of polyamide monomers results in an induced asymmetry upon binding. A hairpin polyamide-DNA complex composed of an asymmetrically folded polyamide may be expected to display differential energetics for oriented binding.

Implications for the Design of Minor Groove Binding Polyamides.

The results reported here establish that a hairpin polyamide binds with at least a 16-fold orientation preference with the N-termini of the polyamide subunits located at the 5'-end of the targeted DNA strand. Although the cationic dimethylaminopropyl end group is not required for oriented hairpin polyamide recognition in the minor groove of DNA, the N or C terminus position of the charge appears important. Pyrrole-imidazole polyamide DNA-binding orientation preference defines a new design feature which must be considered for application of the pairing rules for DNA targeting. For optimal orientation and, hence sequence-specificity, a positive charge at the C-terminus and no substitution of an N-terminal imidazole is preferred.

Experimental Section

Dicyclohexylcarbodiimide (DCC), Hydroxybenzotriazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU) and 0.2 mmol/gram Boc-β-alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-β-Pam-Resin) and Boc-γ-aminobutyric acid were purchased from Peptides International. *N,N*-diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF), *N*-methylpyrrolidone (NMP), DMSO/NMP, Acetic anhydride (Ac_2O), and 0.0002 M potassium cyanide/pyridine were purchased from Applied Biosystems. Dichloromethane (DCM) and triethylamine (TEA) was reagent grade from EM, thiophenol (PhSH) and 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. Disposable polypropylene filters were also used for washing resin for ninhydrin and picric acid tests, and for filtering pre-dissolved amino acids into reaction vessels. 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.[15] ^{1}H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz in 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 an HP 1090M analytical HPLC or a Beckman Gold system using a RAINEN C_{18}, Microsorb MV, 5μm, 300 x 4.6 mm reversed phase column in 0.1% (wt/v) TFA with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% acetonitrile/min. Preparatory reverse phase HPLC was performed on a Beckman

HPLC with a Waters DeltaPak 25 x 100 mm, 100 μm C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. 18MΩ water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μm filtered.

ImPyPy-γ-PyPyPy-β-EtOH (2) ImPyPy-γ-PyPyPy-β-Pam-Resin was synthesized by machine-assisted solid phase methods. A sample of resin (240 mg, 0.18 mmol/gram[16]) was placed in a glass scintillation vial and treated with neat ethanolamine (2 mL). The reaction mixture was placed in an oven and periodically agitated (55 °C, 24 h). Upon completion of polyamide cleavage, the reaction mixture was filtered to remove resin, 0.1 % (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC chromatography. ImPyPy-γ-PyPyPy-β-EtOH is recovered upon lyophyllization as a white powder (6.8 mg, 16 % recovery). UV λ_{max} 246, 306 (50,000); ^{1}H NMR (DMSO-d_6) δ 10.47 (s, 1 H), 9.90 (s, 1 H), 9.88 (s, 1 H), 9.87 (s, 1 H), 9.82 (s, 1 H), 8.03 (t, 1 H, J = 5.2 Hz), 7.95 (t, 1 H, J = 6.0 Hz), 7.85 (t, 1 H, J = 5.2 Hz), 7.38 (s, 1 H), 7.25 (d, 1 H, J = 1.4 Hz), 7.20 (d, 1 H, J = 1.3 Hz), 7.16 (d, 1 H, J = 1.4 Hz), 7.14 (m, 2 H), 7.11 (d, 1 H, J = 1.5 Hz), 7.05 (s, 1 H), 7.00 (d, 1 H, J = 1.3 Hz), 6.87 (d, 1 H, J = 1.4 Hz), 6.84 (d, 1 H, J = 1.4 Hz), 6.78 (d, 1 H, J = 1.4 Hz), 3.96 (s, 3 H), 3.81 (s, 3 H), 3.80 (s, 3 H), 3.79 (s, 3 H) 3.77 (s, 3 H), 3.76 (s, 3 H), 3.34 (m, 4 H), 3.17 (q, 2 H, J = 5.7 Hz), 3.05 (q, 2 H, J = 5.9 Hz), 2.27 (m, 4 H), 1.72 (quintet, 2 H, J = 6.9 Hz); MALDI-TOF-MS, 936.3 (937.0 calc. for M+H).

Dp-ImPyPy-γ-PyPyPy-β-Me (4) Dp-ImPyPy-γ-PyPyPy-β-Pam-Resin was synthesized by machine-assisted solid phase methods. A sample of resin (240 mg, 0.18 mmol/gram[16]) was placed in a glass scintillation vial and treated with a saturated solution of methylamine in DMF (20 ml). The reaction mixture was placed in sealed Parr apparatus and periodically agitated (55 °C, 24 h, 80 psi). Upon completion of cleavage, the reaction mixture was cooled to room temperature and filtered to remove resin. Excess DMF was removed *in vacuo*, 0.1% (wt/v) TFA added (6 mL), and the resulting solution purified by reversed phase HPLC chromatography. Upon lyophyllization Dp-ImPyPy-γ-PyPyPy-β-Me

is recovered as a white powder (20.6 mg, 47% recovery). UV λ_{max}, 248, 312 (50,000); ^{1}H NMR (DMSO-d_6) δ 10.34 (s, 1 H), 9.53 (s, 1 H) 9.90 (s, 3 H), 9.85 (s, 1 H), 9.5 (br s, 1 H), 8.03 (m, 2 H), 7.81 (q, 1 H, J = 4.2 Hz), 7.42 (s, 1 H), 7.24 (d, 1 H, J = 1.6 Hz), 7.22 (d, 1 H, J = 1.5 Hz), 7.17 (d, 1 H, J = 1.5 Hz), 7.17 (m, 2 H), 7.12 (d, 1 H, J = 1.5 Hz), 7.00 (d, 1 H, J = 1.6 Hz), 6.88 (d, 1 H, J = 1.6 Hz), 6.85 (d, 1 H, J = 1.6 Hz), 6.80 (d, 1 H, J = 1.5 Hz), 3.93 (s, 3 H), 3.83 (s, 3 H), 3.81 (m, 6 H), 3.77 (m, 6 H), 3.34 (q, 2 H, J = 6.1 Hz), 3.18 (q, 2 H, J = 5.4 Hz), 3.03 (q, 2 H, J = 5.2 Hz), 2.76 (d, 6 H, J = 4.5 Hz), 2.54 (d, 3 H, J = 4.3 Hz), 2.37 (t, 2 H, J = 6.3 Hz), 2.26 (m, 4 H), 1.87 (quintet, 2 H, J = 6.6 Hz), 1.76 (quintet, 2 H, J = 6.8 Hz), MALDI-TOF-MS, 1034.4 (1035.2 calc. for M+H).

ImPyPy-γ-PyPyPy-β-Dp-NH$_2$ (1-NH$_2$) A sample of ImPyPy-γ-PyPyPy-β-Pam-resin (350 mg, 0.18 mmol/gram[16]) was placed in a glass scintillation vial and treated with neat 3,3'-diamino-*N*-methyldipropylamine (2 ml). The reaction mixture was placed in an oven and periodically agitated (55 °C, 24 h). Resin was removed by filtration through a disposable propylene filter, and the resulting solution diluted with 0.1% (wt/v) TFA to a total volume of 8 mL, and purified directly by reversed phase HPLC to provide ImPyPy-γ-PyPyPy-β-Dp-NH$_2$ (28 mg, 40% recovery) as a white powder ^{1}H NMR (DMSO-d_6) δ 10.47 (s, 1 H), 9.91 (s, 1 H), 9.90 (s, 1 H), 9.88 (s, 1 H), 9.84 (s, 1 H), 9.2 (br s, 1 H), 8.0 (m, 3 H), 7.8 (br s, 3 H), 7.33 (s, 1 H), 7.26 (d, 1 H, J = 1.2 Hz), 7.20 (d, 1 H, J = 1.4 Hz), 7.14 (m, 4 H), 7.03 (m, 2 H), 6.88 (d, 1 H, J = 1.4 Hz), 6.85 (m, 2 H), 3.97 (s, 3 H), 3.82 (m, 9 H), 3.78 (m, 6 H), 3.33 (q, 2 H, J = 5.7 Hz), 3.2- 3.0 (m, 8 H), 2.81 (q, 2 H, J = 5.8 Hz), 2.71 (d, 3 H, J = 4.4 Hz), 2.39 (t, 2 H, J = 5.8 Hz), 2.22 (t, 2 H, J = 6.1 Hz), 1.85 (quintet, 2 H, J = 6.2 Hz), 1.78 (m, 4 H). MALDI-TOF-MS, 1022.1 (1021.2 calc. for M+H).

ImPyPy-γ-PyPyPy-β-C7-NH$_2$ (2-NH$_2$) A sample of ImPyPy-γ-PyPyPy-β-Pam-resin (350 mg, 0.18 mmol/gram[16]) was placed in a glass scintillation vial and treated with neat 1,7-diaminoheptane (2 ml). The reaction mixture was placed in an oven and

periodically agitated (55 °C, 24 h). Resin was removed by filtration through a disposable propylene filter, and the resulting solution dissolved with 0.1% (wt/v) TFA to a total volume of 8 mL, and purified directly by preparatory reversed phase HPLC to provide ImPyPy-γ-PyPyPy-β-Dp-NH_2 (28 mg, 39 % recovery) as a white powder. ^{1}H NMR (DMSO-d_6) δ 10.62 (s, 1 H), 9.95 (s, 1 H), 9.92 (s, 1 H), 9.91 (s, 1 H), 9.87 (s, 1 H), 8.09 (t, 1 H, J = 5.3 Hz), 8.05 (t, 1 H, J = 5.6 Hz), 7.90 (t, 1 H, J = 5.7 Hz), 7.6 (br s, 3 H), 7.46 (s, 1 H), 7.29 (d, 1 H, J = 1.7 Hz), 7.22 (d, 1 H, J = 1.5 Hz), 7.16 (m, 3 H), 7.13 (d, 1 H, J = 1.3 Hz), 7.02 (d, 1 H, J = 1.6 Hz), 6.88 (d, 1 H, J = 1.6 Hz), 6.86 (d, 1 H, J = 1.6 Hz), 6.82 (d, 1 H, J = 1.6 Hz), 3.98 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.34 (q, 2 H, J = 5.9 Hz), 3.18 (q, 2 H, J = 6.2 Hz), 3.03 (q, 2 H, J = 5.9 Hz), 2.73 (m, 2 H), 2.27 (m, 4 H), 1.77 (t, 2 H, J = 5.8 Hz), 1.46 (m, 2 H), 1.35 (m, 2 H), 1.23 (m, 6 H); MALDI-TOF-MS, 1006.6 (1006.2 calc. for M+H).

AcImPyPy-γ-PyPyPy-β-Dp-NH_2 (3-NH_2) AcImPyPy-γ-PyPyPy-β-resin was synthesized by machine assisted methods. A sample of resin (350 mg, 0.18 mmol/gram[17]) was placed in a 20 mL glass scintillation vial, and treated with 2 mL 3,3′-diamino-*N*-methyldipropylamine (55 °C, 18 h). Resin was removed by filtration through a propylene filter, and the resulting solution diluted with 0.1% (wt/v) TFA to a total volume of 8 mL, and purified directly by reversed phase HPLC to provide AcImPyPy-γ-PyPyPy-β-Dp-NH_2 (29 mg, 43% recovery) as a white powder. ^{1}H NMR (DMSO-d_6) δ 10.26 (s, 1 H), 10.17 (s, 1 H), 9.92 (m, 2 H), 9.90 (s, 1 H), 9.87 (s, 1 H), 9.5 (br s, 1 H), 8.12 (t, 2 H, J = 5.6 Hz), 8.09 (m, 2 H), 7.9 (br s , 3 H), 7.41 (s, 1 H), 7.25 (d, 1 H, J = 1.5 Hz), 7.21 (d, 1 H, J = 1.3 Hz), 7.16 (m, 3 H), 7.11 (d, 1 H, J = 1.6 Hz), 7.03 (d, 1 H, J = 1.3 Hz), 6.89 (d, 1 H, J = 1.6 Hz), 6.85 (m, 2 H), 3.93 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.37 (q, 2 H, J = 5.9 Hz), 3.2- 3.0 (m, 8 H), 2.84 (q, 2 H, J = 5.7 Hz), 2.71 (d, 3 H, J = 4.3 Hz), 2.33 (t, 2 H, J = 6.6 Hz), 2.26 (t, 2 H, J = 6.8 Hz), 2.00 (s, 3 H), 1.88 (quintet, 2 H, J = 6.8 Hz), 1.77 (m, 4 H). MALDI-TOF-MS, 1078.0 (1078.2 calc. for M+H).

Dp-ImPyPy-γ-PyPyPy-β-C7-NH_2 (4-NH_2) A sample of Dp-ImPyPy-γ-PyPyPy-β-resin (350 mg, 0.16 mmol/gram[17]) was placed in a 20 mL glass scintillation vial and treated with 2 mL 1,7-diaminoheptane (55 °C, 18 h). Resin was removed by filtration through a disposable propylene filter, and the resulting solution diluted with 0.1% (wt/v) TFA to a total volume of 8 mL, and purified directly by reversed phase HPLC to provide Dp-ImPyPy-γ-PyPyPy-β-C7-NH_2 (37 mg, 55% recovery) as a white powder. ^{1}H NMR (DMSO-d_6) δ 10.40 (s, 1 H), 9.98 (s, 1 H), 9.92 (d, 2 H), 9.91 (s, 1 H), 9.87 (s, 1 H), 9.6 (br s, 1 H), 8.09 (m, 1 H), 8.05 (t, 1 H, J = 5.1 Hz), 7.88 (t, 1 H, J = 5.4 Hz), 7.6 (br s , 3 H), 7.43 (s, 1 H), 7.25 (d, 1 H, J = 1.6 Hz), 7.21 (d, 1 H, J = 1.5 Hz), 7.17 (m, 3 H), 7.13 (d, 1 H, J= 1.6 Hz), 7.02 (d, 1 H, J = 1.6 Hz), 6.89 (d, 1 H, J = 1.7 Hz), 6.86 (d, 1 H, J = 1.6 Hz), 6.83 (d, 1 H, J = 1.6 Hz), 3.93 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.80 (s, 3 H), 3.78 (s, 3 H), 3.32 (q, 2 H, J = 5.7 Hz), 3.21 (q, 2 H, J = 6.2 Hz), 3.05 (m, 4 H), 2.76 (d, 6 H, J = 3.9 Hz), 2.72 (m, 2 H), 2.39 (t, 2 H, J = 6.1 Hz), 2.28 (m, 4 H), 1.91 (quintet, 2 H, J = 6.6 Hz), 1.77 (quintet, 2 H, J = 6.4 Hz), 1.47 (m, 2 H), 1.36 (m, 2 H), 1.24 (m, 6 H). MALDI-TOF-MS, 1133.8 (1134.3 calc. for M+H).

ImPyPy-γ-PyPyPy-β-Dp-EDTA (1-E) EDTA-dianhydride (50 mg) was dissolved in 1 mL DMSO/NMP solution and 1 mL DIEA by heating at 55 °C for 5 min. The dianhydride solution was added to ImPyPy-γ-PyPyPy-β-Dp-NH_2 (**1-NH_2**) (8.0 mg, 7 μmol) dissolved in 750 μL DMSO. The mixture was heated at 55 °C for 25 minutes, and treated with 3 mL 0.1M NaOH, and heated at 55 °C for 10 minutes. 0.1% aqueous TFA was added to adjust the total volume to 8 mL and the solution purified directly by reversed phase HPLC chromatography to provide **1-E** as a white powder. (4 mg, 40% recovery) ^{1}H NMR (DMSO-d_6) δ 10.45 (s, 1 H), 9.89 (s, 1 H), 9.88 (s, 1 H), 9.86 (s, 1 H), 9.82 (s, 1 H), 9.2 (br s, 1 H), 8.39 (t, 2 H, J = 6.0 Hz), 8.06 (m, 3 H), 7.37 (s, 1 H), 7.25 (d, 1 H, J = 1.5 Hz), 7.17 (d, 1 H, J = 1.3 Hz), 7.13 (m, 4 H), 7.02 (m, 2 H), 6.87 (d, 1 H, J = 1.2 Hz), 6.85 (m, 2 H), 3.96

(s, 3 H), 3.81 (s, 3 H), 3.80 (s, 3 H), 3.79 (s, 3H), 3.77 (s, 3 H), 3.76 (s, 3 H), 3.68 (m, 4 H), 3.33 (q, 2 H, J = 5.3 Hz), 3.3-3.0 (m, 16 H), 2.85 (m, 2 H), 2.69 (d, 3 H, J = 4.3 Hz), 2.33 (t, 2 H, J = 5.5 Hz), 2.24 (t, 2 H, J = 6.6 Hz), 1.73 (m, 6 H). MALDI-TOF-MS, 1295.4 (1295.3 calc. for M+H).

ImPyPy-γ-PyPyPy-β-C7-EDTA (2-E) Compound **2-E** was prepared from compound **2-NH$_2$** (8.0 mg, 7 μmol) as described for **1-E**. (2 mg, 20% recovery) ^{1}H NMR (DMSO-d_6) δ 10.51 (s, 1 H), 9.94 (s, 1 H), 9.92 (s, 1 H), 9.91 (s, 1 H), 9.86 (s, 1 H), 8.33 (t, 1 H, J = 5.5 Hz), 8.09 (t, 1 H, J = 5.8 Hz), 8.04 (t, 1 H, J = 5.0 Hz), 7.87 (t, 1 H, J = 5.2 Hz), 7.40 (s, 1 H), 7.27 (d, 1 H, J = 1.6 Hz), 7.21 (d, 1 H, J = 1.5 Hz), 7.17 (m, 3 H), 7.14 (d, 1 H, J = 1.7 Hz), 7.05 (s, 1 H), 7.02 (d, 1 H, J = 1.5 Hz), 6.89 (d, 1 H, J = 1.6 Hz), 6.86 (d, 1 H, J = 1.5 Hz), 6.82 (d, 1 H, J = 1.5 Hz), 3.97 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.67 (m, 4 H), 3.33 (q, 2 H, J = 5.6 Hz), 3.22 (m, 4 H), 3.10 (m, 4 H), 3.04 (m, 4 H), 2.92 (m, 4 H), 1.77 (t, 2 H, J = 6.4 Hz), 1.35 (m, 4 H), 1.22 (m, 6 H). MALDI-TOF-MS, 1280.4 (1280.3 calc. for M+H).

AcImPyPy-γ-PyPyPy-β-Dp-EDTA (3-E) Compound **3-E** was prepared from compound **3-NH$_2$** (8.0 mg, 7 μmol) as described for **1-E**. (3 mg, 30%). ^{1}H NMR (DMSO-d_6) δ 10.25 (s, 1 H), 10.00 (s, 1 H), 9.91 (m, 2 H), 9.90 (s, 1 H), 9.85 (s, 1 H), 8.42 (t, 1 H, J = 5.6 Hz), 8.08 (m , 3 H), 7.41 (s, 1 H), 7.26 (d, 1 H, J = 1.5 Hz), 7.20 (d, 1 H, J = 1.5 Hz), 7.16 (m, 3 H), 7.14 (d, 1 H, J = 1.5 Hz), 7.04 (d, 1 H, J = 1.2 Hz), 6.88 (d, 1 H, J = 1.6 Hz), 6.86 (m, 4 H), 3.93 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.71 (m, 4 H), 3.58 (q, 2 H, J = 5.2 Hz), 3.4- 3.0 (m, 18 H), 2.71 (d, 3 H, J = 4.3 Hz), 2.33 (t, 2 H, J = 5.4 Hz), 2.26 (t, 2 H, J = 5.7 Hz), 2.00 (s, 3 H), 1.74 (m, 6 H). MALDI-TOF-MS, 1352.8 (1352.4 calc. for M+H).

Dp-ImPyPy-γ-PyPyPy-β-C7-EDTA (4-E) Compound **4-E** was prepared from compound **4-NH$_2$** (8.0 mg, 7 μmol) as described for **1-E**. (4 mg, 40% recovery) ^{1}H NMR (DMSO-d_6) δ 10.36 (s, 1 H), 9.92 (s, 1 H), 9.88 (m, 2 H), 9.87 (s, 1 H), 9.82 (s, 1 H), 9.2 (br s, 1 H), 8.03 (m, 3 H), 7.98 (m, 1 H), 7.41 (s, 1 H), 7.22 (d, 1 H, J = 1.5 Hz), 7.20 (d, 1 H, J = 1.6 Hz), 7.14 (m, 4 H), 7.10 (d, 1 H, J = 1.6 Hz), 7.00 (d, 1 H, J = 1.5 Hz), 6.87 (d, 1 H, J = 1.3 Hz), 6.83 (d, 1 H, J = 1.4 Hz), 6.78 (d, 1 H, J = 1.6 Hz), 3.92 (s, 3 H), 3.82 (s, 3 H), 3.80 (s, 3 H), 3.79 (s, 3 H), 3.77 (s, 3 H), 3.76 (s, 3 H), 3.41 (m, 4 H), 3.3-3.0 (m, 16 H), 2.74 (d, 6 H, J = 4.3 Hz), 2.71 (m, 2 H), 2.31 (t, 2 H, J = 5.4 Hz), 2.22 (t, 2 H, J = 5.6 Hz), 1.77 (m, 2 H), 1.39 (m, 2 H), 1.28 (m, 4 H), 1.20 (m, 6 H). MALDI-TOF-MS, 1407.7 (1408.6 calc. for M+H).

DNA Reagents and Materials. Sonicated, deproteinized calf thymus DNA, from Pharmacia, was dissolved in filter sterilized water to a final concentration of 1 mM in base pairs and stored at 4 °C. Glycogen was purchased from Boehringer-Manheim as a 20 mg/mL aqueous solution. Nucleotide triphosphates were purchased from Pharmacia and used as supplied. Nucleoside triphosphates labeled with ^{32}P (≥ 3000 C_i /mmol) were obtained from Dupont-New England Nuclear. Cerenkov radioactivity was measured with a Beckman LS 2801 scintillation counter. All enzymes were purchased from Boehringer-Manheim and were used according to the supplier's recommended protocol in the activity buffer provided. Plasmid pUC19 was obtained from Worthington Biochemical. A solution of 0.5 M EDTA, pH 8.0 was purchased from Ultrapure. Phosphoramidites were from Glen Research. The pH of buffer solutions was recorded using a digital pH/millivolt meter (model no. 611, Orion Research) and a ROSS semimicro combination pH electrode. General manipulation of duplex DNA and oligonucleotides were performed according to established procedures.[17,18]

Construction of plasmid DNA. Oligodeoxynucleotides were synthesized by standard automated solid support chemistry using an Applied Biosystems model 380B DNA synthesizer and *O*-cyanoethyl-*N,N*-diisopropyl phosphoramidites. Plasmid pDEH2 was prepared by hybridization of a complementary set of synthetic oligonucleotides: 1) 5'-GATCCGCATGC GATCGTGTTACGCATGCGATCGTCTTACGCATGCGATCG-3'; 2) 5'-AGCT CGATCGCATGCGTAAGACGATCGCATGCGTAACACGATCGCATGCG-3'. The complementary oligonucleotides were annealed and then ligated to the large pUC19 *Bam* HI/ *Hind* III restriction fragment using T4 DNA ligase. The ligated plasmid was then used to transform Epicurean^{TM} Coli XL-1 Blue Supercompetent cells. Colonies were selected for α-complementation on 25 mL Luria-Bertani medium agar plates containing 50 mg/mL ampicillin and treated with XGAL and IPTG solutions. Large scale plasmid purification was performed using Qiagen purification kits. The presence of the desired insert was determined by dideoxy sequencing using a USB Sequenase v. 2.0 kit. Plasmid DNA concentration was determined at 260 nm using the relation 1 OD unit = 50 mg/mL duplex DNA.

Preparation of 3′- and 5′-End-Labeled Restriction Fragments. The plasmid pDEH2 was linearized with *Eco* RI and then treated with Sequenase v. 2.0, deoxyadenosine 5′-[α-^{32}P]triphosphate and thymidine 5′-[α-^{32}P]triphosphate. The 3'-end labeled fragment was then digested with *Pvu* II and loaded onto a 7% non-denaturing polyacrylamide gel. The desired 266 base pair band was visualized by autoradiography and isolated. For MPE and affinity cleaving reactions, pDEH2 also was 5'-^{32}P labeled. First, the plasmid pDEH2 was digested with *Eco* RI, then dephosphorylated with calf intestine alkaline phosphatase. The digested plasmid then was 5'-^{32}P labeled with T4 polynucleotide kinase and deoxyadenosine 5′-[γ-^{32}P]triphosphate, digested with *Pvu* II, and loaded onto a 7% non-denaturing polyacrylamide gel. The desired 266 base pair band was visualized by autoradiography and isolated. Chemical sequencing adenine-specific reactions were performed on all labeled fragments.[19]

MPE•Fe(II) Footprint titrations.[8] All reactions were performed in a total volume of 40 μL. A polyamide stock solution (ImPyPy-γ-PyPyPy-β-Dp, ImPyPy-γ-PyPyPy-β-EtOH, AcImPyPy-γ-PyPyPy-β-Dp, or Dp-ImPyPy-γ-PyPyPy-β-Me) or H_2O (for reference lanes) was added to an assay buffer containing either 3'-^{32}P labeled or 5'-^{32}P labeled restriction fragment (20,000 cpm), affording final solution conditions of 25 mM Tris-Acetate, 10 mM NaCl, 100 μM/bp calf thymus DNA, pH 7, and either (i) polyamide stock solution, or (ii) no polyamide (for reference lanes). Solutions were incubated at 22 °C for 24 hours. A fresh 50 μM MPE•Fe(II) solution was made from 100 μL of a 100 μM MPE solution and 100 μM ferrous ammonium sulfate ($Fe(NH_4)_2(SO_4)_2$•$6H_2O$) solution. Then, 4 μL of a 50 μM MPE•Fe(II) solution was added, and the solution was allowed to equilibrate for 10 minutes at 22 °C. Cleavage was initiated by the addition of 4 μL of a 50 mM dithiothreitol solution and allowed to proceed for 15 minutes at 22 °C. Reactions were stopped by ethanol precipitation, and resuspended in 1 × TBE/ 80% formamide loading buffer, denatured by heating at 85 °C for 15 minutes, and placed on ice. Reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% crosslinking, 7 M urea) in 1 × TBE at 2000 V for 2.5 hours. Gels were dried on a slab dryer and then exposed to a storage phophor screen at 22 °C. The data were analyzed by performing volume integrations of the cleavage bands and reference bands using the ImageQuant v. 3.3 software. Background-corrected volume integration of rectangles encompassing cleavage bands was normalized to a maximum value of 0.95.

Affinity Cleaving Titrations. All affinity cleavage reactions[7] were performed in a total volume of 40 μL. A polyamide stock solution (ImPyPy-γ-PyPyPy-β-Dp-EDTA, ImPyPy-γ-PyPyPy-β-C7-EDTA, AcImPyPy-γ-PyPyPy-β-Dp-EDTA, or Dp-ImPyPy-γ-PyPyPy-β-C7-EDTA) or H_2O (for reference lanes) was added to an assay buffer containing either 3'-^{32}P labeled or 5'-^{32}P labeled restriction fragment (20,000 cpm), affording final

solution conditions of 25 mM tris-acetate, 20 mM NaCl, 100 μM calf thymus DNA, pH 7, and either (i) 100 nM-5 μM polyamide or (ii) no polyamide (for reference lanes). Solutions were incubated at 22 °C for 24 hours. Then, 4 μL of a 100 μM ferrous ammonium sulfate solution was added, and the solution was allowed to equilibrate for 20 min at 22 °C. Affinity cleaving reactions were initiated by the addition of 4 μL of a 100 mM dithiothreitol solution, and reacted for 30 min at 22 °C. Reactions were stopped by the addition of 10 μL of a solution containing 1.5 M NaOAc (pH 5.5), 0.28 mg/mL glycogen, and 14 μM base pairs calf thymus DNA, and ethanol precipitated. Reactions were resuspended in 1 × TBE/ 80% formamide loading buffer, denatured by heating at 85 °C for 15 minutes, and placed on ice. Reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% crosslinking, 7 M urea) in 1 × TBE at 2000 V for 2.5 hours. Gels were dried on a slab dryer and then exposed to a storage phophor screen at 22 °C. The data were analyzed by performing volume integrations of the cleavage bands and reference bands using the ImageQuant v. 3.3 software. Background-corrected volume integration of rectangles encompassing cleavage bands was normalized to a maximum value of 3.95.

DNase I Footprinting. All reactions[9] were carried out in a volume of 40 μL. We note that no carrier DNA was used in these reactions. A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were: 10 mM tris•HCl buffer (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, and 20 kcpm 3′-radiolabeled DNA. The solutions were allowed to equilibrate for 24 hours at 22 °C. Cleavage was initiated by the addition of 4 μL of a DNase I stock solution (diluted with 1 mM DTT to give a stock concentration of 0.1 u/mL) and was allowed to proceed for 7 min at 22 °C. The reactions were stopped by the addition of 1.25 M sodium chloride solution containing 100 mM EDTA, 0.2 mg/mL glycogen, and 28 μM bp calf thymus DNA, and then ethanol precipitated. The cleavage products were resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer, denatured at 85 °C for 15 min, placed on ice, and immediately

loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V for 2.5 hour. The gels were dried under vacuum at 80 °C, then quantitated using storage phosphor technology.[20]

The data were analyzed by performing volume integrations of the 5′-TGTTA-3′ and 5′-TCTTA-3′ sites and a reference site. The apparent DNA target site saturation, θ_{app}, was calculated for each concentration of polyamide using the following equation:

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

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

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

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

data were used in determining each association constant. All lanes from each gel were used unless visual inspection revealed a data point to be obviously flawed relative to neighboring points. The data were normalized using the following equation:

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

Quantitation by Storage Phosphor Technology Autoradiography Photostimuable storage phosphor imaging plates (Kodak Storage Phosphor Screen SO230 obtained from Molecular Dynamics) were pressed flat against dried gel samples and exposed in the dark at 22 °C for 12-24 hours. A Molecular Dynamics 400S PhosphorImager was used to obtain all data from the storage screens.[20] The data were analyzed by performing volume integration of the target site and reference blocks using the ImageQuant v. 3.3 software.

Acknowledgments.

We are grateful to the National Institutes of Health (GM-2768) for research support, the National Science Foundation for a predoctoral fellowship to S.W. and The Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B. We thank G. M. Hathaway for MALDI-TOF mass spectrometry.

References

1. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci.U.S.A.* **1992**, *89*, 7586. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385. (d) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 2572 (e) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559.

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

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

4. See Table I, reference 1a.

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*, 6147. (c) de Claire, R. P. L.; Geierstanger B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc. submitted.*

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

7. (a) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457. (b) Dervan, P. B. *Science* **1986**, *232*, 464.

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

9. (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.

10. We note that no conclusion regarding polyamide binding affinities can be drawn from these experiments due to specific and non-specific interactions of the polyamides with the calf thymus carrier DNA.

11. (a) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160. (b) Geierstanger, B. H.; Mrksich M.; Dervan, P. B.; Wemmer, D. E. *Nature Struct. Biol.* **1996**, *3*, 321.

12. Mismatched pairing of Im/Py opposite C•G cannot be ruled out for recognition of 5'-TCTTA-3' sites; however, mismatches of this class are energetically unfavorable and have been found to reduce DNA binding by at least 3 kcal/mol.

13. Dipoles were calculated using MacSpartan v 1.0. See: Mecozzi, S., West, A. P., Dougherty, D. A. *J. Am. Chem. Soc.* **1996**, *93*, 10566-10571.

14. B. Norden, personal communication

15. Kent, S.B.H. *Ann. Rev. Biochem.* **1988**, *57*, 957.

16. Resin substitution has been corrected for the weight of the polyamide chain. The change in resin substitution can be calculated as $L_{new}(mmol/g) = L_{old}/(1 + L_{old}(W_{new} - W_{old}) \times 10^{-3})$, where L is the loading (mmol of amine per gram of resin), and W is the weight ($gmol^{-1}$) of the growing polyamide attached to the resin. See: Barlos, K.; Chatzi, O.; Gatos, D.; Stravropoulos, G. *Int. J. Peptide Protein Res.* **1991**, *37*, 513.

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

18. Gait, M.J. (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, Oxford.

19. (a) Iverson, B. L.; Dervan, P. B. *Nucl. Acids Res.* **1987**, *15*, 7823-7830. (b) Maxam, A. M.; Gilbert, W. S. *Methods in Enzymology* **1980**, *65*, 499-560.

20. Johnston, R.F.; Picket, S.C.; Barker, D.L. (1990) *Electrophoresis 11*, 355.

21. We note that residual salt is most likely responsible for the discontinuity present at the top of gels and does not interfere with analysis of the target sites.

Chapter 3

On the Pairing Rules for Recognition in the Minor Groove of DNA by Pyrrole-Imidazole Polyamides

Abstract *Synthetic polyamides containing aromatic amino acids bind to predetermined sequences of double helical DNA. Sequence-specificity depends on side-by-side pairings of N-methylpyrrole and N-methylimidazole carboxamides in the DNA minor groove. We set out to determine the relative energetics of pairings of Im/Py, Py/Im, Im/Im, and Py/Py for targeting G•C and A•T base pairs. A key specificity issue which has not been previously addressed is whether an Im/Im pair is energetically equivalent to an Im/Py pair for targeting G•C base pairs. Equilibrium association constants (K_a) were determined at two five base pair sites for a series of four six-ring hairpin polyamides in order to test the relative energetics of the four aromatic amino acid pairings opposite G•C and A•T base pairs in the central position. We observed that a G•C base pair is effectively targeted with Im/Py but not Py/Im, Py/Py, or Im/Im. The A•T base pair is effectively targeted with Py/Py but not Im/Py, Py/Im, or Im/Im. An Im/Im pairing is energetically disfavored for recognition of both A•T and G•C. This specificity will create limitations on slipped motifs available for unlinked dimers in the minor groove resulting in enhanced predictability of the current pairing rules for specific molecular recognition of double helical DNA.*

Publication: White, Baird & Dervan *Chemistry & Biology* **1997**, *4*, 569-578.

Crescent shaped polyamides containing pyrrole (Py) and imidazole (Im) amino acids bind cooperatively as antiparallel dimers in the minor groove of the DNA helix.[1] Sequence-specificity depends on the side-by-side pairings of *N*-methylpyrrole and *N*-methylimidazole amino acids. A pairing of Im opposite Py targets a G•C base-pair, while Py/Im targets C•G.[1-5] A Py/Py pairing is degenerate and targets both T•A and A•T base-pairs.[1-9] Pyrrole-imidazole polyamides have been shown to be cell permeable and to inhibit the transcription of genes in cell culture.[10] This provides impetus to explore the scope and limitations of this approach for DNA recognition, particularly the energetics and structural details of this remarkably simple binary code.

Recognition of G•C by the Im/Py pairing requires precise positioning for the key hydrogen bond between the Im N3 and the exocyclic amine of guanine.[11,12] Given the central location of the guanine exocyclic amine group in the DNA minor groove[13-15], the question arises whether an Im/Im pairing might also be expected to target G•C.[16] Remarkably, even in the first report on the binding specificity of the three ring polyamide homodimer (ImPyPy-Dp)$_2$, there was qualitative data to suggest that there was indeed a binding preference for placement of the Im/Py pair opposite G•C (Figure 3.1).[1] It would be useful to determine the generality of the aromatic amino acid pairing preferences and to compare the relative energetics of the four possible pairings of Im and Py for recognition of G•C and T•A base pairs. Therefore, we describe here an experimental design utilizing a single oriented six-ring hairpin polyamide which allows the relative energetic preferences of four different binary combinations, Im/Py, Py/Im, Im/Im, and Py/Py, to be tested opposite a G•C or an A•T base pair (Figure 3.2).

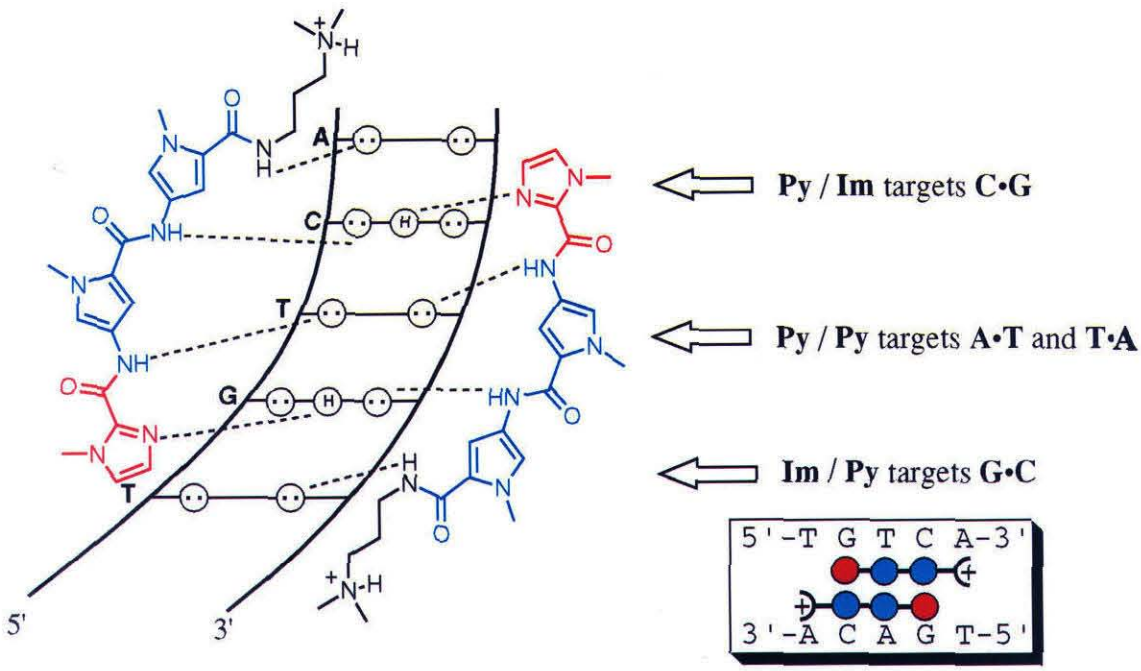


Figure 3.1 Binding models for antiparallel three-ring polyamide subunits, ImPyPy-Dp, in complex with 5'-TGTCA-3'. (left) Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an *H* represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. For schematic binding model (right), the Im and Py rings are represented as red and blue spheres respectively.

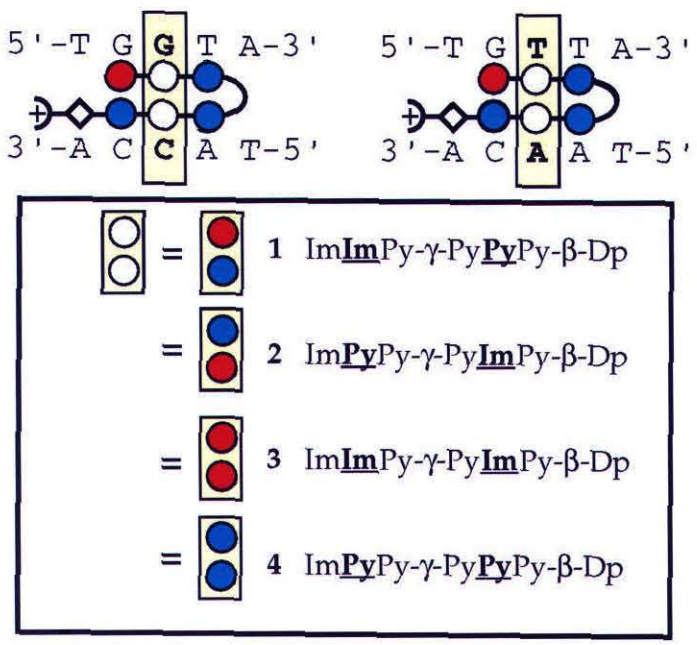


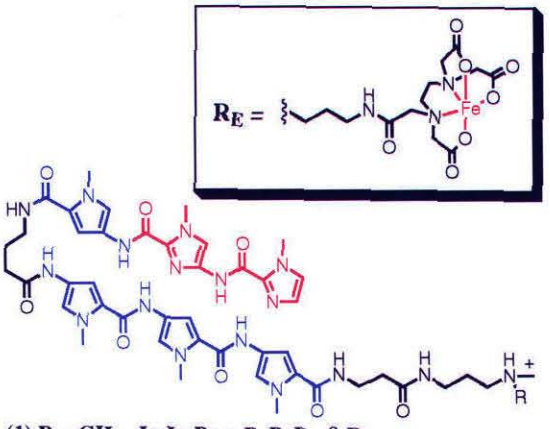
Figure 3.2 Models of the expected "hairpin" complex of ImImPy-γ-PyPyPy-β-Dp **1**, ImPyPy-γ-PyImPy-β-Dp **2**, ImImPy-γ-PyImPy-β-Dp **3**, and ImPyPy-γ-PyPyPy-β-Dp **4** with 5'-TGGTA-3' and 5'-TGTTA-3'. Unfilled white circles may be either Im or Py residues which are represented by red and blue circles respectively. Diamonds represent β-alanine, and γ-aminobutyric acid is represented as a curved line. The central pairing of Im/Py, Py/Im, Im/Im, and Py/Py with G•C and T•A is highlighted with a yellow box.

Three-ring polyamide subunits covalently coupled C-N by a central γ-aminobutyric acid linker form hairpin structures at 5-bp target sequences.[17-20] For example, according to the pairing rules, polyamides of sequence composition ImImPy-γ-PyPyPy-β and ImPyPy-γ-PyPyPy-β (central amino acid pairing underlined) would be expected to bind to 5'-TGGTA-3' and 5'-TGTTA-3' sequences respectively. Selective substitution of the central amino acid of each three ring polyamide subunit allows for four ring pairings at a unique central location within the hairpin structure placed opposite a G•C or T•A base pair (Figure 3.2).

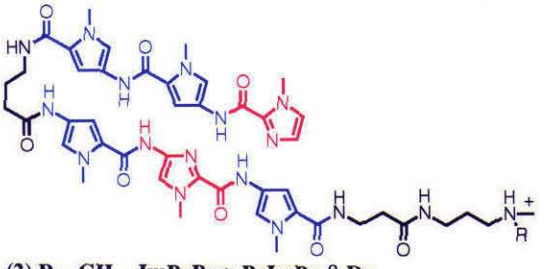
Four six-ring polyamides, ImImPy-γ-PyPyPy-β-Dp **1**, ImPyPy-γ-PyImPy-β-Dp **2**, ImImPy-γ-PyImPy-β-Dp **3**, and ImPyPy-γ-PyPyPy-β-Dp **4**, containing central amino pairings of Im/Py, Py/Im, Im/Im, and Py/Py in a hairpin structure were synthesized by solid phase methods (Figure 3.3).[21] The corresponding EDTA analogs ImImPy-γ-PyPyPy-β-Dp-EDTA **1-E**, ImPyPy-γ-PyImPy-β-Dp-EDTA **2-E**, ImImPy-γ-PyImPy-β-Dp-EDTA **3-E**, and ImPyPy-γ-PyPyPy-β-Dp-EDTA **4-E** were also constructed in order to confirm the single binding orientation of each hairpin:DNA complex.

We report here the DNA-binding affinities, orientations, and sequence-selectivity of the four polyamides for two five base pair binding sites, 5'-TGTTA-3' and 5'-TGGTA-3', which vary at one unique third position. Three separate techniques are used to characterize the DNA-binding properties of the polyamides: affinity cleaving[22,23] and MPE•Fe(II)[24,25] and DNase I footprinting.[26-28] Affinity cleavage studies determine the specific binding orientation and stoichiometry of each hairpin:DNA complex. Binding location and site size is accurately determined by MPE•Fe(II) footprinting, while quantitative DNase I footprint titration is more suitable for measurement of equilibrium association constants (K_a) for the polyamide binding to designated sequences.

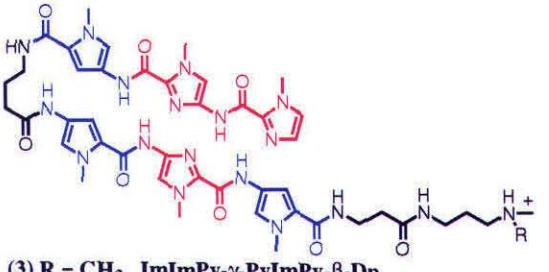
Figure 3.3 Structures of polyamides ImImPy-γ-PyPyPy-β-Dp **1**, ImPyPy-γ-PyImPy-β-Dp **2**, ImImPy-γ-PyImPy-β-Dp **3**, ImPyPy-γ-PyPyPy-β-Dp **4**, ImImPy-γ-PyPyPy-β-Dp-EDTA• Fe(II) **1-E•Fe(II)**, ImPyPy-γ-PyImPy-β-Dp-EDTA• Fe(II) **2-E•Fe(II)**, ImImPy-γ-PyImPy-β-Dp-EDTA• Fe(II) **3-E•Fe(II)**, ImPyPy-γ-PyPyPy-β-Dp-EDTA• Fe(II) **4-E•Fe(II)**.



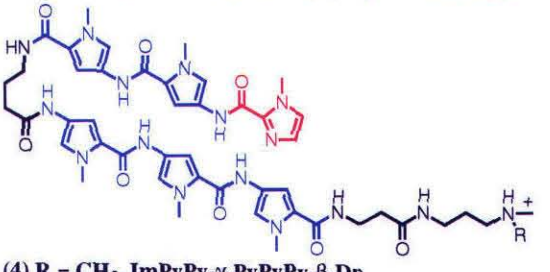
(1) R = CH_3, ImImPy-γ-PyPyPy-β-Dp
(1-E) R = R_E, ImImPy-γ-PyPyPy-β-Dp-EDTA•Fe(II)



(2) R = CH_3, ImPyPy-γ-PyImPy-β-Dp
(2-E) R = R_E, ImPyPy-γ-PyImPy-β-Dp-EDTA•Fe(II)



(3) R = CH_3, ImImPy-γ-PyImPy-β-Dp
(3-E) R = R_E, ImImPy-γ-PyImPy-β-Dp-EDTA•Fe(II)



(4) R = CH_3, ImPyPy-γ-PyPyPy-β-Dp
(4-E) R = R_E, ImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II)

Results and discussion

Binding Site Size by MPE•Fe(II) Footprinting. MPE•Fe(II) footprinting on 3' and 5'-^{32}P end labeled 302 base pair restriction fragments (25 mM tris-acetate, 10 mM NaCl, 100 μM bp calf thymus DNA, 5 mM DTT, pH 7.0 and 22 °C) reveals that the polyamides bind and discriminate the two five base pair sites, 5'-TGTTA-3' and 5'-TGGTA-3' (Figure 3.4). Polyamides **1** and **3** at μM concentrations bind in decreasing affinity to 5'-TGGTA-3' > 5'-TGTTA-3'. Polyamides **2** and **4** at μM concentrations bind in reverse order with decreasing affinity 5'-TGTTA-3' > 5'-TGGTA-3'.

Binding Orientation by Affinity Cleaving. Affinity cleaving experiments using hairpin polyamides modified with EDTA•Fe(II) at the C-terminus, were used to determine polyamide binding orientation and stoichiometry. Experiments were performed on the same 3' and 5'-^{32}P end labeled 302 base restriction fragments (25 mM tris-acetate, 20 mM NaCl, 100 μM bp calf thymus DNA, pH 7, 22 °C, 10 mM DTT, 10 μM Fe(II)) (Figure 3.5). The observed cleavage patterns are in all cases 3'-shifted, consistent with minor groove occupancy.[22] A single cleavage locus proximal to the 5'-side of both the 5'-TGTTA-3' and 5'-TGGTA-3' binding sites confirm that the four polyamides bind each discreet site with a single 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.[29] A 1:1 oriented but extended motif would require at least an eight base pair binding site, which is inconsistent with the high-resolution MPE footprinting data on both target sites. The hairpin complex ImPyPy-γ-PyPyPy-Dp•5'-TGTTA-3' has recently been characterized by direct NMR methods.[30]

Figure 3.4 (left) MPE•Fe(II) footprinting experiments on the 3′-^{32}P-labeled 302-bp *Eco* RI/*Pvu* II restriction fragment derived from the plasmid pDEH4. The 5′-TGGTA-3′, and 5′-TGTTA-3 sites are shown on the right side of each autoradiogram. Lane 1, A reaction; lane 2, G reaction; lane 3 MPE•Fe(II) standard; lane 7, intact DNA; lanes 4-6: 20 μM, 10 μM, 1 μM polyamide. All lanes contain 30 kcpm 3′-radiolabeled DNA, 25 mM Tris-acetate buffer (pH 7.0), 10 mM NaCl, and 100 μM/base pair calf thymus DNA. (Top) Illustration of the 302 bp restriction fragment with the position of the sequence indicated. (right) MPE•Fe(II) protection patterns of 10 μM ImImPy-γ-PyPyPy-β-Dp (**1**), 20 μM ImPyPy-γ-PyImPy-β-Dp (**2**), 20 μM ImImPy-γ-PyImPy-β-Dp (**3**), 10 μM ImImPy-γ-PyImPy-β-Dp (**4**). Bar heights are proportional to the relative protection from cleavage at each band. The binding sites 5'-TGGTA-3' and 5'-TGTTA-3' are highlighted in blue.

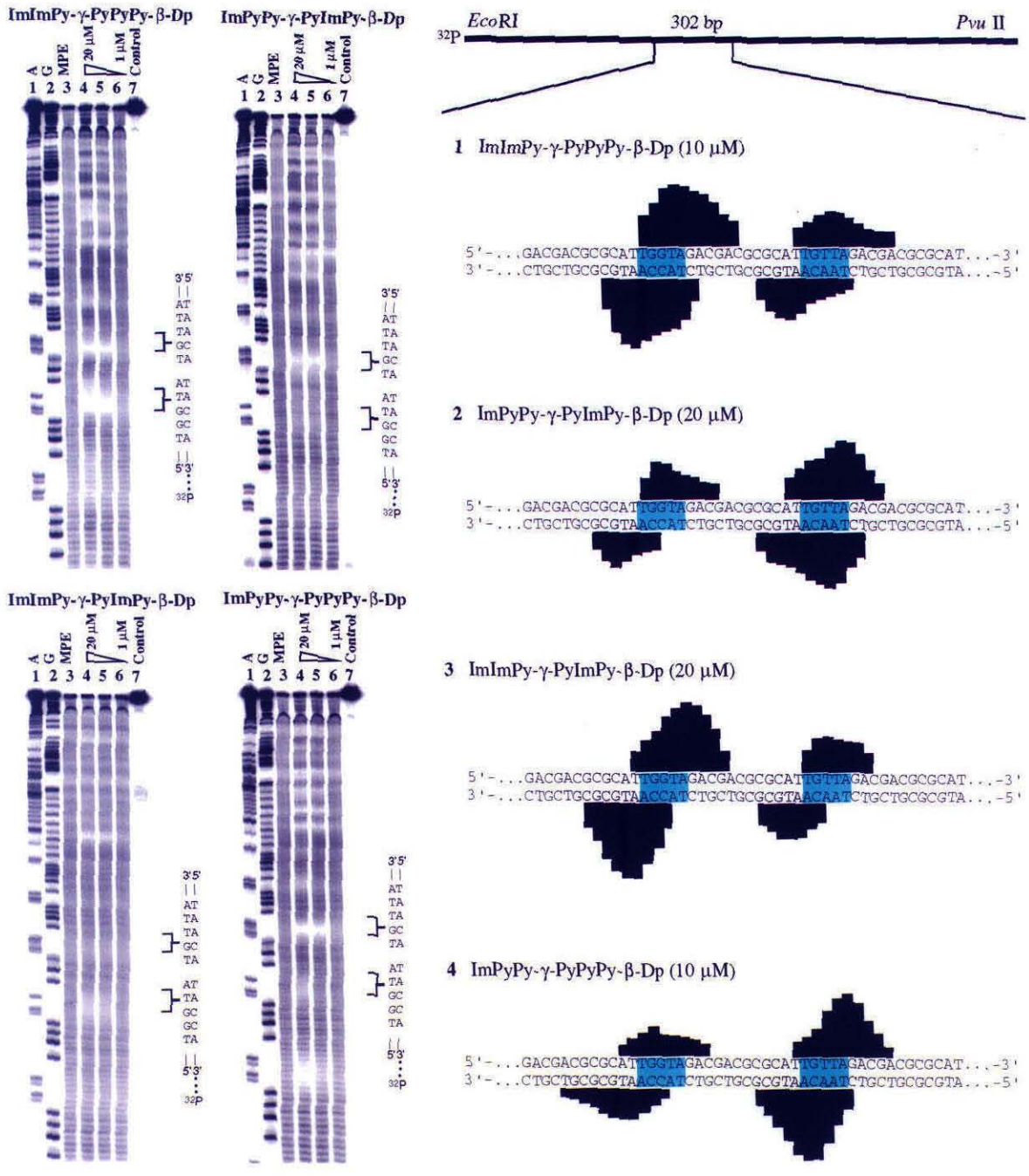
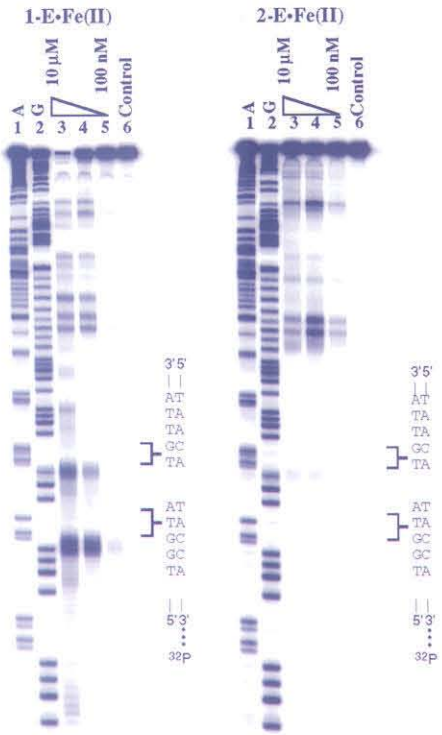
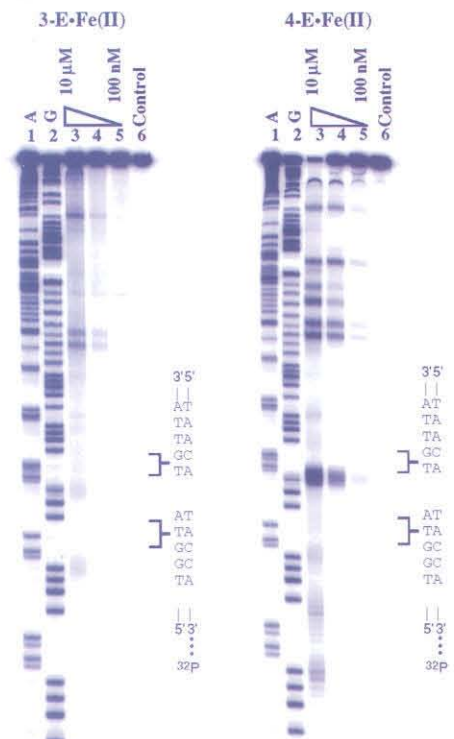
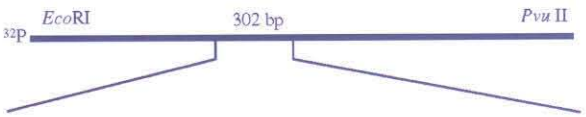

ImImPy-γ-PyPyPyPy-β-Dp
ImPyPy-γ-PyImPy-β-Dp
ImImPy-γ-PyImPy-β-Dp
ImPyPy-γ-PyPyPy-β-Dp
A G MPE 20 μM 1 μM Control
1 2 3 4 5 6 7
3'5'
5'3'
32P
EcoRI
302 bp
Pvu II
1 ImImPy-γ-PyPyPy-β-Dp (10 μM)
2 ImPyPy-γ-PyImPy-β-Dp (20 μM)
3 ImImPy-γ-PyImPy-β-Dp (20 μM)
4 ImPyPy-γ-PyPyPy-β-Dp (10 μM)
5'-...GACGACGCGCATTGGTAGACGACGCGCATTGTTAGACGACGCGCAT...-3'
3'-...CTGCTGCGCGTAACCATCTGCTGCGCGTAACAATCTGCTGCGCGTA...-5'

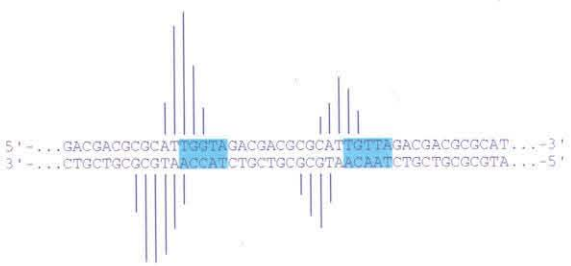
Figure 3.5 (left) Affinity cleavage experiments on the 3′-^{32}P-labeled 302-bp *Eco* RI/*Pvu* II restriction fragment derived from the plasmid pDEH4. The targeted binding sites are indicated on the right side of the storage phosphor autoradiograms obtained from the 8% denaturing polyacrylamide gels used to separate the fragments generated by affinity cleaving experiments: lanes 1 and 2, A and G sequencing lanes; lane 7 intact DNA, lanes 3-6, digestion products obtained in the presence of **1-E, 2-E, 3-E,** or **4-E**: 10 μM, 1 μM, and 100 nM polyamide. All reactions contain 20 kcpm 3'-^{32}P restriction fragment, 25 mM Tris-acetate, 20 mM NaCl, 100 μM bp calf thymus DNA, pH 7. (Top) Illustration of the 302 bp restriction fragment with the position of the sequence indicated. (right) Results from affinity cleavage with ImImPy-γ-PyPyPy-β-Dp-EDTA•Fe(II) **1-E•Fe(II)**, ImPyPy-γ-PyImPy-β-Dp-EDTA•Fe(II) **2-E•Fe(II)**, ImImPy-γ-PyImPy-β-Dp-EDTA•Fe(II) **3-E•Fe(II)**, and ImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II) **4-E•Fe(II)** at 10 μM concentration. Line heights are proportional to the relative cleavage intensities at each base pair. The binding sites 5'-TGGTA-3' and 5'-TGTTA-3' are highlighted in blue.







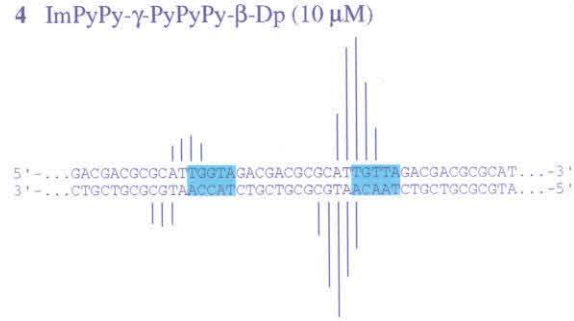
1 ImImPy-γ-PyPyPy-β-Dp (10 µM)



2 ImPyPy-γ-PyImPy-β-Dp (10 µM)

3 ImImPy-γ-PyImPy-β-Dp (10 µM)

4 ImPyPy-γ-PyPyPy-β-Dp (10 µM)



Energetics by Quantitative DNase I Footprint Titrations. MPE•Fe(II) footprinting combined with affinity cleaving experiments indicate that each polyamide binds the designated five base pair target site as a 1:1 hairpin complex in the minor groove. This single and consistent mode of binding allows valid thermodynamic comparison for the central ring amino acid pairing of each polyamide recognizing the central base pair of each designated target site. Quantitative DNase I footprint titration experiments[24-26] (10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5mM $CaCl_2$, pH 7.0, 22 °C) were performed to determine the equilibrium association constants K_a for recognition of the bound sites (Figure 3.6). The 5'-TGGTA-3' site is bound by the polyamides with decreasing affinity: ImImPy-γ-PyPyPy-β-Dp (**1**) >> ImPyPy-γ-PyImPy-β-Dp (**2**) ≈ ImImPy-γ-PyImPy-β-Dp (**3**) ≈ ImPyPy-γ-PyPyPy-β-Dp (**4**). The 5'-TGTTA-3' site is bound with decreasing affinity ImPyPy-γ-PyPyPy-β-Dp (**4**) > ImPyPy-γ-PyImPy-β-Dp (**2**) ≈ ImImPy-γ-PyPyPy-β-Dp (**1**) > ImImPy-γ-PyImPy-β-Dp (**3**). Remarkably, the association constant for recognition of each site varies 100-fold between the four polyamides, indicating a remarkable sensitivity to a single atomic substitution within the central ring amino acids (Figure 3.7).

The Im/Py Pair. Among the four ligands, ImImPy-γ-PyPyPy-β-Dp **1** (central Im/Py pairing) binds to the 5'-TGGTA-3' site which contains a central G•C base with the highest affinity ($K_a = 9.0 \times 10^7\ M^{-1}$). This selectivity indicates that Im/Py is the optimal ring pairing for recognition of G•C. The sequence-specificity of the Im/Py pairing is underscored by the 50-fold reduced affinity ($K_a = 1.7 \times 10^6\ M^{-1}$) for placement of the Im/Py pair opposite a T•A base pair at 5'-TGTTA-3'.

Figure 3.6 Quantitative DNase I footprint titration experiment with ImImPy-γ-PyPyPy-β-Dp **1**, ImPyPy-γ-PyImPy-β-Dp **2**, ImImPy-γ-PyImPy-β-Dp **3**, ImPyPy-γ-PyPyPy-β-Dp **4**, on the 3' ^{32}P end-labeled 302-base pair *Eco* RI/ *Pvu* II restriction fragment from plasmid pDEH4. All reactions contained 20 kcpm restriction fragment, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$. Lane 1, intact DNA; lane 2 DNase I standard; lanes 3-11 contain 1 μM, 100 nM, 65 nM, 10 nM, 6.5 nM, 4 nM, 2.5 nM, 500 pM, and 100 pM polyamide, respectively; lane 12, G reaction. The portion of the restriction fragment containing the targeted binding sites is indicated on the left-hand side of the storage phosphor autoradiograms that were obtained from the 8% denaturing polyacrylamide gels used to separate the fragments generated by DNase I footprinting experiments. Hairpin polyamide binding models are indicated at each binding site; blue and red circles represent Py and Im rings, respectively. Diamonds represent β-alanine and the curved line between the two polyamide subunits represents the γ-aminobutyric acid linker. The central pairings of Im/Py, Py/Im, Im/Im, and Py/Py with G•C and T•A are highlighted with yellow boxes.

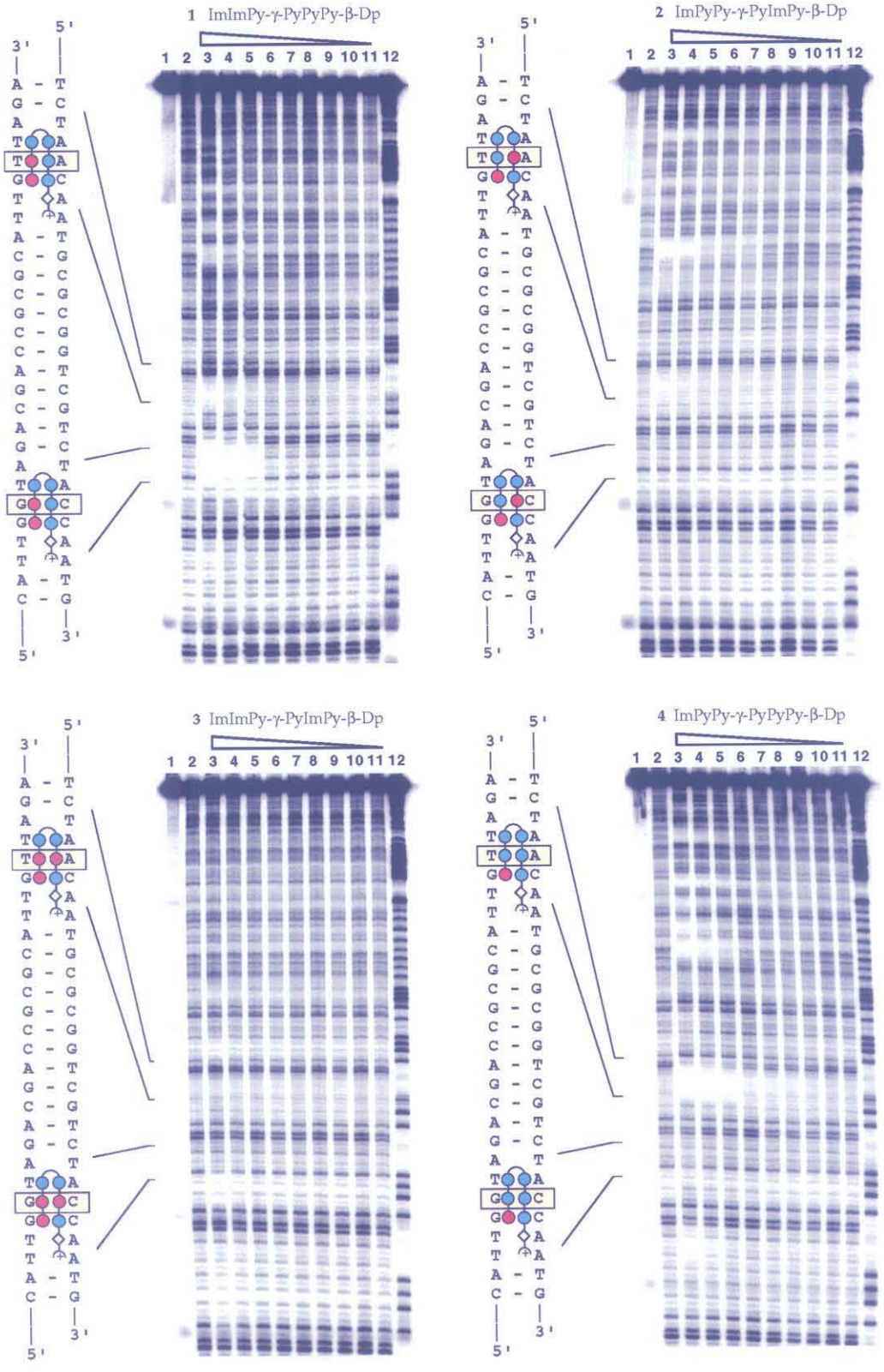
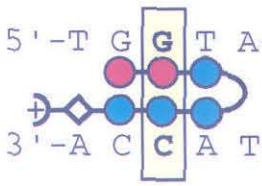
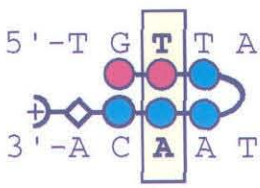

1 ImImPy-γ-PyPyPy-β-Dp
2 ImPyPy-γ-PyImPy-β-Dp
3 ImImPy-γ-PyImPy-β-Dp
4 ImPyPy-γ-PyPyPy-β-Dp
1 2 3 4 5 6 7 8 9 10 11 12
5'
3'

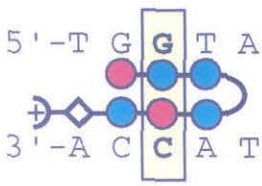
Figure 3.7 Models of the expected hairpin complex of ImImPy-γ-PyPyPy-β-Dp **1**, ImPyPy-γ-PyImPy-β-Dp **2**, ImImPy-γ-PyImPy-β-Dp **3**, and ImPyPy-γ-PyPyPy-β-Dp **4** complexed with 5'-TGGTA-3' and 5'-TGTTA-3'. Blue and red circles represent Py and Im rings, respectively. Diamonds represent β-alanine and the curved line between the two polyamide subunits represents the γ-aminobutyric acid linker. The central pairings of Im/Py, Py/Im, Im/Im, and Py/Py with G•C and T•A are highlighted with yellow boxes. The binding affinity for each complex as determined by quantitative DNase I footprinting is shown to the right of each complex. (bottom) A summary of the binding affinities: +, favorable; -, no binding.



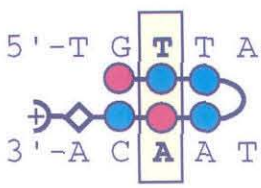
Im/Py vs. G•C
$K_a = 9.0 \times 10^7\ M^{-1}$



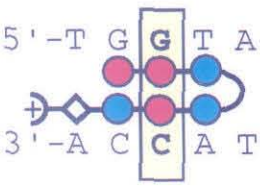
Im/Py vs. T•A
$K_a = 1.7 \times 10^6\ M^{-1}$



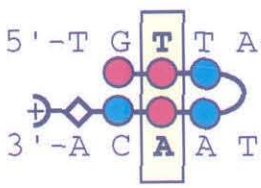
Py/Im vs. G•C
$K_a < 1.0 \times 10^6\ M^{-1}$



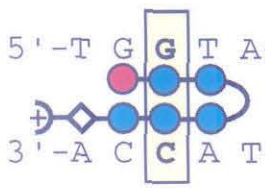
Py/Im vs. T•A
$K_a = 1.8 \times 10^6\ M^{-1}$



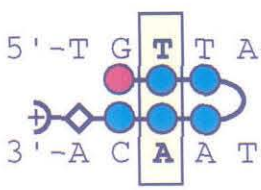
Im/Im vs. G•C
$K_a < 1.0 \times 10^6\ M^{-1}$



Im/Im vs. T•A
$K_a < 1.0 \times 10^6\ M^{-1}$



Py/Py vs. G•C
$K_a < 1.0 \times 10^6\ M^{-1}$



Py/Py vs. T•A
$K_a = 9.0 \times 10^7\ M^{-1}$

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

The Py/Im Pair. The polyamide ImPyPy-γ-PyImPy-β-Dp **2** (central Py/Im pairing) binds to the 5'-TGGTA-3' site (central G•C base pair) with 100-fold reduced affinity relative to polyamide **1** (central Im/Py pairing). Remarkably, given the central location of the exocyclic 2-amino group of guanine in the minor groove, the Py/Im pairing is disfavored relative to the Im/Py pair for recognition of G•C. Although the nitrogen of the 2-amino group is displayed in a similar location for both G•C and C•G, the proton available for hydrogen bond recognition in the minor groove has a *strand specific directionality* (Figure 3.8). The directionality requirements for effective hydrogen bond formation allow discrimination of G•C by the Im/Py and Py/Im pairs. High resolution x-ray structure studies validate this model for G•C recognition by an Im/Py pair.[31] Placement of the Py/Im pair of **2** opposite T•A at the 5'-TGTTA-3' target site results in similar affinity ($K_a = 1.7 \times 10^6$ M^{-1}) to placement of the Im/Py pair **1** opposite T•A. It should be noted that the "Py/Im" hairpin polyamide **2** recognizes a 5'-AGCTT-3' match site present on the restriction fragment (central C•G pair) with high affinity.

The Py/Py pair. The polyamide ImPyPy-γ-PyPyPy-β-Dp **4** (central Py/Py pairing) binds to the 5'-TGTTA-3' site (central T•A base pair) with 100-fold enhanced affinity relative to the 5'-TGGTA-3' site (central G•C base pair). A Py/Py pairing discriminates A•T/T•A base pairs from G•C/C•G, likely due to the exocyclic amine groups of guanine which present a steric hindrance to deep polyamide binding in the minor groove. However, Im rich hairpin polyamides recognize G,C sequences with affinities and specificities similar to Py rich polyamides that recognize A•T rich sequences[32] indicating, as will become evident below, that additional energetic parameters may be important for high affinity recognition.

Figure 3.8 Space-filling models of the G•C and C•G base pairs as viewed from the minor groove of DNA, generated using B-form DNA coordinates provided in InsightII. The O2 and N3 atoms are highlighted in red. The asymmetrically directed amino proton is highlighted in blue.

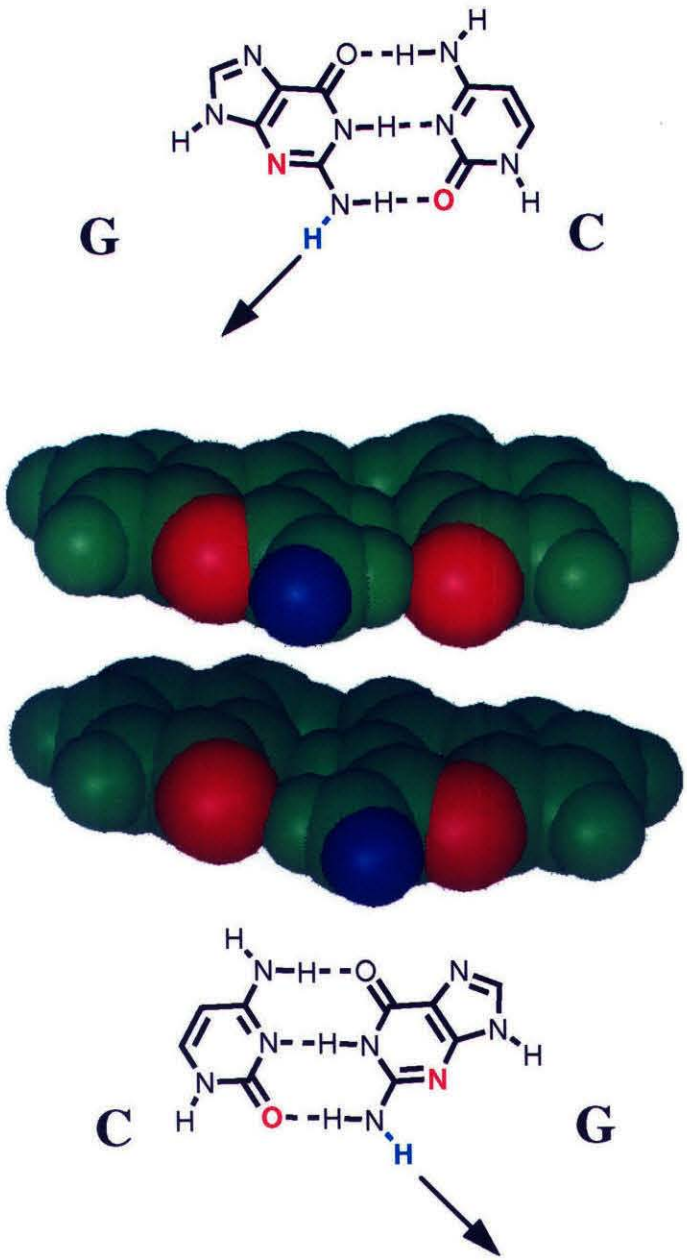
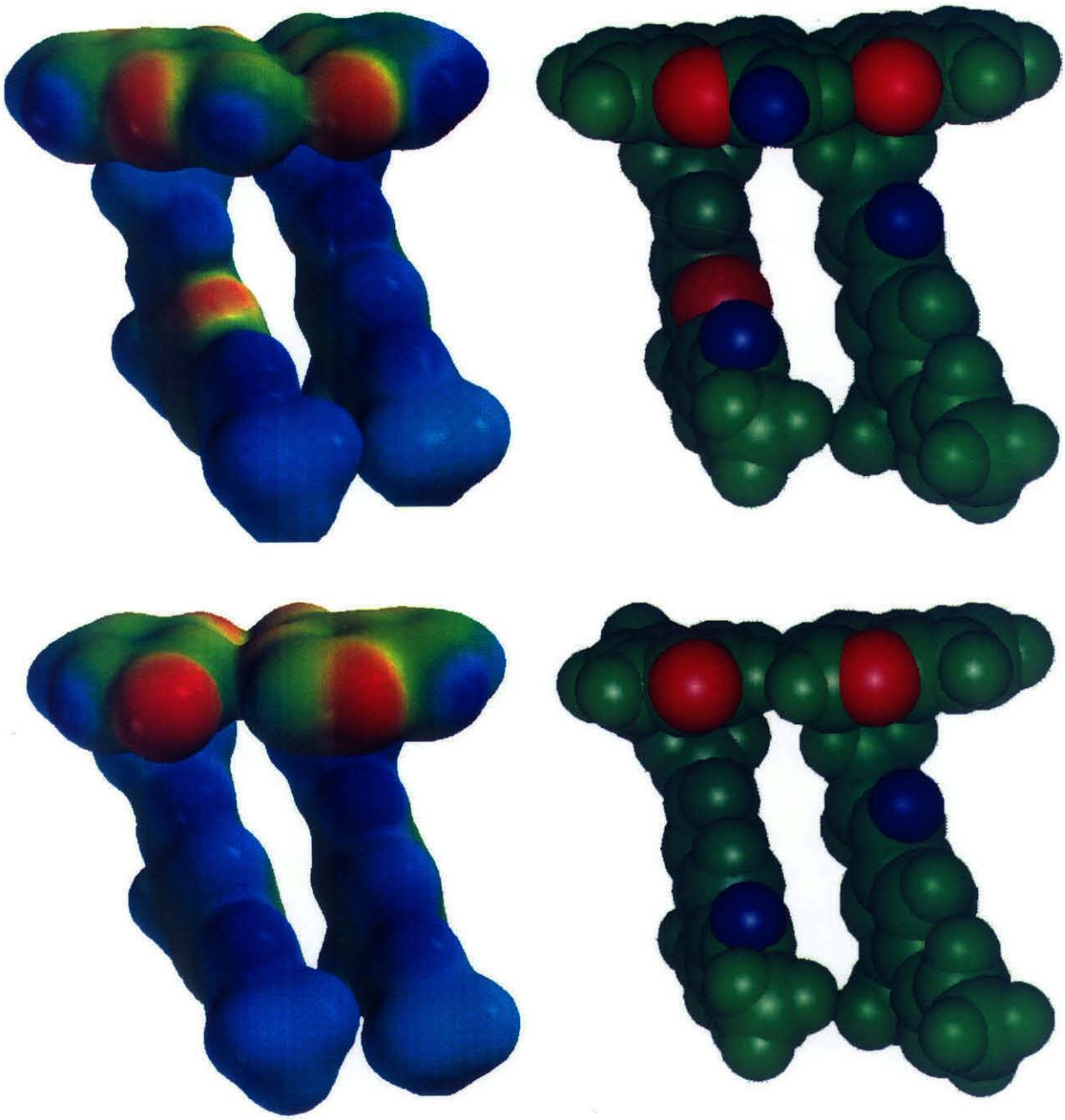

G
C
C
G

Figure 3.9 Electrostatic potential maps (left) and space-filling models (right) for Im/Py with G•C (top) and Py/Py with T•A (bottom). Electrostatic potentials were calculated using MacSpartan version 1.0 as described in ref. 40. Areas of positive potential are represented as blue surface and areas of negative potential as red surface; neutral regions appear green. Space-filling models were generated using InsightII software. Atoms involved in hydrogen bond formations are highlighted, with donors and acceptors represented as blue and red, respectively.



The Im/Im pair. The hairpin ImImPy-γ-PyImPy-β-Dp (central Im/Im pairing) binds to both designated sites, 5'-TGTTA-3' and 5'-TGGTA-3' with > 100-fold reduced affinity relative to polyamides **1** and **4** (central Py/Py and Im/Py pairs, respectively). The reduced binding energetics of the Im/Im pair may result from desolvation of the N-3 of Im upon binding to DNA.[33] Such an unfavorable desolvation could be compensated when a hydrogen bond is formed between Im and the exocyclic 2-amino group of guanine, but not compensated upon placement of Im opposite A, T, or C bases.

Alternatively, specificity may result from electrostatic interactions between the aromatic ring carboxamides and the floor of the DNA minor groove.[34] The DNA minor groove displays a negative electrostatic potential at A•T and T•A base pairs.[35] The only positive potential located on the minor groove floor is at the exocyclic 2-amino group of guanine.[34] The Py ring displays a positive potential across the interface with the floor of the minor groove, providing complementary electrostatic interactions at A, T, and C bases (Figure 3.9). The Im ring displays a negative potential which can interact favorably with the 2-amino group of G, but may interact unfavorably with the other three bases.

Significance

Cell-permeable small molecules that target predetermined DNA sequences with high affinity and specificity have the potential to control gene expression. A binary code has been developed to correlate DNA sequence with polyamide sequence composition. The results described here demonstrate the sequence-specificity of the four individual ring amino acid pairings: (1) the Im/Py pairing recognizes G•C, but is disfavored when placed opposite T•A, (2) the Py/Im pairing targets C•G, but is disfavored for placement opposite G•C and T•A base pairs, (3) the Py/Py pairing recognizes T•A/A•T base pairs but is disfavored opposite

G•C/C•G, (4) an Im/Im pairing is disfavored at all four base pairs breaking a potential degeneracy for recognition by preventing unlinked polyamide dimers from binding in certain slipped motifs.[29] These results create baseline energetic parameters which will guide further second-generation polyamide design for DNA recognition.

Experimental Section

The solid phase synthesis of six-ring hairpin polyamides has been described.[9,17] *E. coli* XL-1 Blue supercompetent cells were obtained from Stratagene. All enzymes were purchased from Boeringher-Mannheim and were used according to the supplier's recommended protocol in the activity buffer provided. Sequenase (version 2.0) was obtained from United States Biochemical. [α-^{32}P]-Thymidine-5'-triphosphate (≥ 3000 C_i/mmol), [α-^{32}P]-deoxyadenosine-5'-triphosphate (≥ 6000 C_i/mmol), and [γ-^{32}P]-adenosine-5'-triphosphate were purchased from Du Pont/NEN.

Preparation of ^{32}P-labeled DNA. Plasmid pDEH4 was prepared by hybridizing a complementary set of synthetic oligonucleotides, 5'-CTAGACGACGCGCATTATTAGACGACGCGCATTGGTAGACGACGCGCATTGTTA GACGACGCGCATTGTCAGACGACGCGCATTGCA-3' and 5'-ATGCGCGTCGTCTG ACAATGCGCGTCGTCTAACAATGCGCGTCGTCTACCAATGCGCGTCGTCTAATAA TGCGCGTCGT-3' and ligating the resulting duplex to the large pUC19 *Xba* I/*Pst* I restriction fragment. The 3'-^{32}P end-labeled *Eco* RI/*Pvu* II fragment was prepared by digesting the plasmid with *Eco* RI and simultaneously filling in using Sequenase, [α-^{32}P]-deoxyadenosine-5'-triphosphate, and [α-^{32}P]-thymidine-5'-triphosphate, digesting with *Pvu* II, and isolating the 302 bp fragment by nondenaturing gel electrophoresis. The 5'-^{32}P-end-

labeled *Eco* RI /*Pvu* II fragment was prepared using standard methods. A and G sequencing were carried out as described.[36,37] Standard methods were used for all DNA manipulations.[38]

MPE•Fe(II) Footprint Titrations. All reactions were performed in a total volume of 40 μL. A polyamide stock solution (**1**,**2**,**3**, or **4**) or H_2O (for reference lanes) was added to an assay buffer containing labeled restriction fragment (30,000 cpm), affording final solution conditions of 25 mM Tris-Acetate, 10 mM NaCl, 100 μM/bp calf thymus DNA, pH 7. Solutions were incubated at 22 °C for 24 hours. A fresh 50 μM MPE•Fe(II) solution was made from 100 μL of a 100 μM MPE solution and 100 μM ferrous ammonium sulfate ($Fe(NH_4)_2(SO_4)_2$•$6H_2O$) solution. Then, 4 μL of the 50 μM MPE•Fe(II) solution was added, and the solution was allowed to equilibrate for 10 minutes at 22 °C. Cleavage was initiated by the addition of 4 μL of a 50 mM dithiothreitol solution and allowed to proceed for 15 minutes 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.[39] Relative cleavage intensities were determined by volume integration of individual cleavage bands using ImageQuant software.

Affinity cleavage reactions. All reactions were executed in a total volume of 40 μL. A stock solution of polyamide (1-E,2-E,3-E, or 4-E) or H_2O 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 μM/bp calf thymus DNA, and pH 7.0. Solutions were incubated for a minimum of 4 hours at 22 °C. Subsequently, 4 μL of freshly prepared 100 μM

$Fe(NH_4)_2(SO_4)_2$ was added and the solution allowed to equilibrate for 20 min. Cleavage reactions were initiated by the addition of 4 μL of 100 mM dithiothreitol, allowed to proceed for 30 min at 22 °C, then stopped by the addition of 10 μL of a solution containing 1.5 M NaOAc (pH 5.5), 0.28 mg/mL glycogen, and 14 μM 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 titration experiments. All reactions were executed in a total volume of 40 μL. A polyamide stock solution (**1**, **2**, **3**, or **4**) or H_2O (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 $MgCl_2$, 5 mM $CaCl_2$, 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 24h. 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 1.5h. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics).

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:

$$\theta app = 1 - \frac{Isite/Iref}{Isite^{\circ}/Iref^{\circ}} \quad (1)$$

where I°_{site} and I°_{ref} are the site and reference intensities, respectively, from a control lane to which no polyamide was added. The ($[L]_{tot}$, 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:

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

where $[L]_{tot}$ 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. The binding isotherms were normalized using the following equation:

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

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

References

1. Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794.

2. Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586-7590.

3. Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385-11389.

4. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572-2576.

5. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559-561.

6. Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723-5727.

7. Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393-1399.

8. Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingham, M. *Nature Struct. Biol.* **1994**, *1*, 169.

9. White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532-12537.

10. Gottesfeld, J. M.; Nealy, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202-205.

11. Pilch, D. S.; Poklar, N. A.; Gelfand, C.A.; Law, S. M.; Breslauer, K J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306.

12. Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646-650.

13. Wing, R.; Drew, T.; Takano, C.; Broka, S.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Nature* **1980**, *287*, 755-758.

14. Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804-808.

15. Steitz, T. A. *Quart. Rev. Biophys.* **1990**, *23*, 205-280.

16. Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 5911-5919.

17. Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988.

18. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Chemistry & Biology* **1996**, *3*, 369-377.

19. Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147.

20. Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6153.

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

22. Taylor, J.S.; Schultz, P.G.; Dervan, P.B. *Tetrahedron* **1984**, *40*, 457.

23. Dervan, P. B. *Science* **1986**, *232*, 464.

24. Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.*, **1982**, *79*, 5470.

25. Van Dyke, M. W.; Dervan, P. B. *Science* **1984**, *225*, 1122.

26. Fox, K. R.; Waring, M. J. *Nucleic Acids Res.* **1984**, *12*, 9271-9285.

27. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132-181.

28. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8462-8466.

29. Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160.

30. de Clairac, R.P.L.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1997**, *119*, 7906.

31. Kielkopf, C.L.; Baird, E.E.; Dervan, P.B.; Rees, D.C. *Nature Struct. Biol.* **1998**, *5*, 104.

32. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953-6961.

33. Singh, S. B.; Wemmer, D. E.; Kollman, P.A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7673-7677.

34. Zimmer, C.; Wahnert, U. *Prog. Biophys. molec. Biol.* **1986**, *47*, 31-112.

35. Pullman, B. *Advances in Drug Research* **1990**, *18*, 1-113.

36. Maxam, A.M.; Gilbert, W. S. *Methods Enzymol.* **1980**, *65*, 499-560.

37. Iverson, B. L.; Dervan, P. B. *Methods Enzymol.* **1987**, *15*, 7823-7830.

38. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning*. (2nd edn). Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1989.

39. Johnston, R. F.; Picket, S. C.; Barker, D. L. *Electrophoresis* **1990**, *11*, 355.

40. Mecozzi, S.; West, A. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10566-10571.

Chapter 4

Effects of the A•T/ T•A Degeneracy of Pyrrole-Imidazole Polyamide Recognition in the Minor Groove of DNA

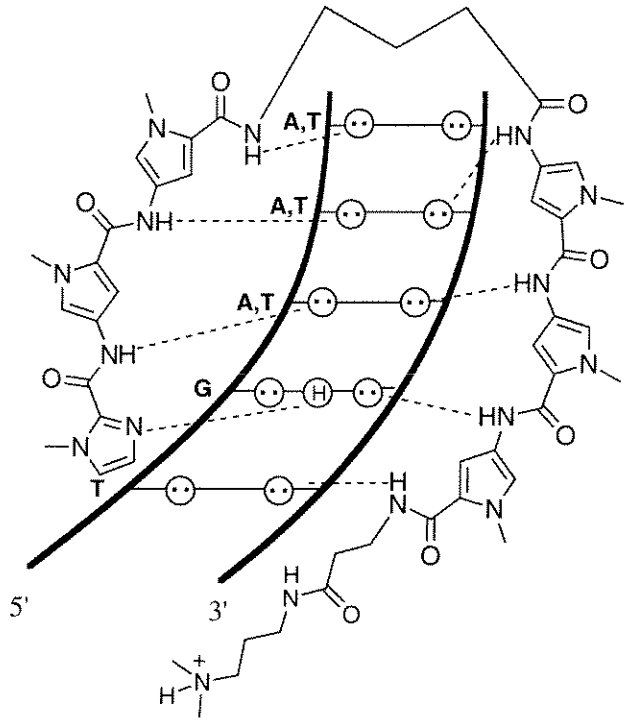
Abstract *Pairing rules have been developed to predict the sequence specificity of minor groove binding polyamides containing pyrrole (Py) and imidazole (Im) amino acids. An Im/Py pair distinguishes G•C from C•G and both of these from A•T/ T•A base pairs. A Py/Py pair appears not to distinguish A•T from T•A base pairs. To test the extent of this degeneracy, the affinity and binding orientation of the hairpin polyamide ImPyPy-γ-PyPyPy-β-Dp was measured for eight possible five base pair 5'-TG(A,T)$_3$-3' match sites. Affinity cleavage experiments using a polyamide with an EDTA•Fe(II) moiety at the carboxy terminus, ImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II), are consistent with formation of an oriented 1:1 hairpin polyamide complex at all eight 5'-TG(A,T)$_3$-3' binding sites. Quantitative DNase I footprint titration experiments reveal that ImPyPy-γ-PyPyPy-β-Dp binds all eight 5'-TG(A,T)$_3$-3' target sites with only a 12-fold difference in the equilibrium association constants between the strongest site, 5'-TGTTT-3' ($K_a = 2.1 \times 10^8$ M^{-1}), and the weakest site, 5'-TGAAT-3' ($K_a = 1.8 \times 10^7$ M^{-1}). This relatively small range indicates that the Py/Py pair is approximately degenerate for recognition of A,T base pairs, affording generality with regard to targeting sequences of mixed A•T/T•A composition.*

Publication: White, Baird, & Dervan *Biochemistry* **1996**, *35*, 12532-12537.

Pyrrole-imidazole polyamide-DNA complexes provide a paradigm for the design of artificial molecules for recognition of double helical DNA.[1] Polyamides containing *N*-methylimidazole (Im) and *N*-methylpyrrole (Py) amino acids can be combined in antiparallel side-by-side dimeric complexes with the minor groove of DNA. The DNA-binding sequence specificity of these small molecules depends on the sequence of side-by-side amino acid pairings.[2] A pairing of imidazole opposite pyrrole recognizes a G•C base pair, while a Py/Im combination targets a C•G base pair.[1] A Py/Py pair has apparent degeneracy for A•T/ T•A base pairs.[1a,3]

Covalently linking polyamide subunits has led to designed ligands with increased affinity and specificity.[4] The polyamide ImPyPy-γ-PyPyPy-Dp containing γ-aminobutyric acid (γ) as an internal guide residue was found to specifically bind as a "hairpin" to a designated 5'-TGTTA-3' target site with 300-fold enhancement relative to the binding affinities of the individual unlinked polyamide pair, ImPyPy and PyPyPy (Figure 4.1 and Figure 4.2).

The discrimination of G•C from C•G base pairs and both of these from A•T/ T•A base pairs by pyrrole-imidazole polyamides has been demonstrated.[1a] A key issue is to determine if a polyamide of core sequence composition ImPyPy-γ-PyPyPy would bind all possible 5'-(A,T)G(A,T)$_3$-3' target sequences. Alternatively, given the sequence dependent variation in DNA groove width[5], perhaps only a smaller subset of 5'-(A,T)G(A,T)$_3$-3' sequences would be structurally compatible with polyamide-DNA complex formation. To address this question, a plasmid was designed containing eight binding sites of the form 5'-TG(A,T)$_3$-3'; 5'-TGTTT-3', 5'-TGTTA-3', 5'-TGTAA-3', 5'-TGTAT-3', 5'-TGAAA-3', 5'-TGATT-3', 5'-TGAAT-3', and 5'-TGATA-3', with each site flanked by the same 12 base-pair sequence (Figure 4.3). Quantitative DNase I footprint titration experiments afford a



ImPyPy-γ-PyPyPy-β-Dp•5'-TG(A,T)$_3$-3'

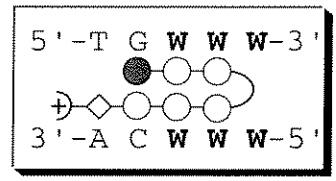
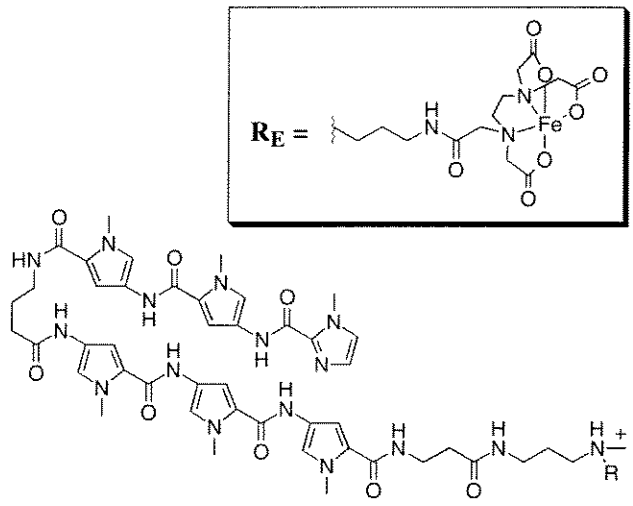


Figure 4.1 (Top) Model for the complex formed between the hairpin polyamide ImPyPy-γ-PyPyPy-β-Dp with a 5'-TG(A,T)-3' site. Circles with dots represent lone pairs of N3 purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. (bottom) Schematic binding model. The imidazole and pyrrole rings are represented as shaded and unshaded circles, respectively; the β-alanine residue is represented as an unshaded diamond. W is either A•T or T•A.



R = CH_3 ImPyPy-γ-PyPyPy-β-Dp

R = R_E ImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II)

Figure 4.2 Structure of the hairpin polyamides ImPyPy-γ-PyPyPy-β-Dp and ImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II).

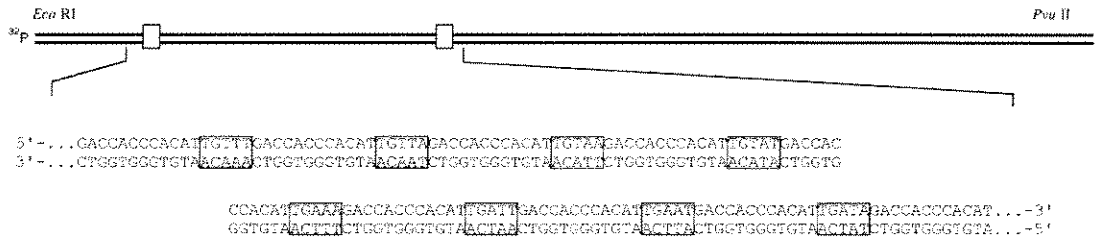


Figure 4.3 Partial sequence of the 370 base pair *Eco* RI/ *Pvu* II restriction fragment. Eight five base pair binding sites having the sequence 5'-TG(A,T)$_3$-3' proximal to the ^{32}P label at the *Eco* RI site were analyzed by quantitative footprint titration analysis.

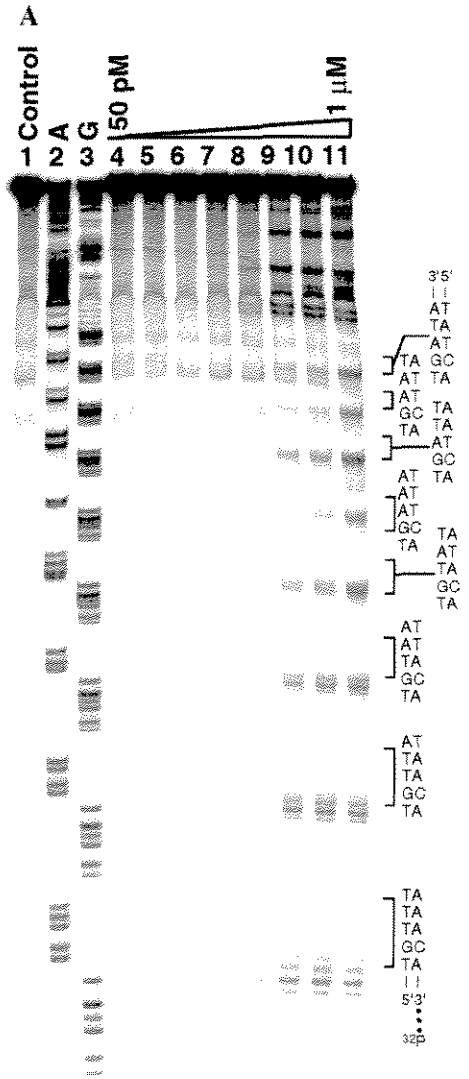
comparison of the equilibrium association constant for binding of each site by the polyamide ImPyPy-γ-PyPyPy-β-Dp, previously optimized for a hairpin motif.[4e,f] For controls, affinity cleavage experiments with ImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II) confirm that the polyamide binds each five base pair site in a single orientation, supporting the hairpin model. We report here that ImPyPy-γ-PyPyPy-β-Dp binds *all sites* of the form 5'-TG(A,T)$_3$-3' with a 12-fold range between the highest and lowest observed affinities.

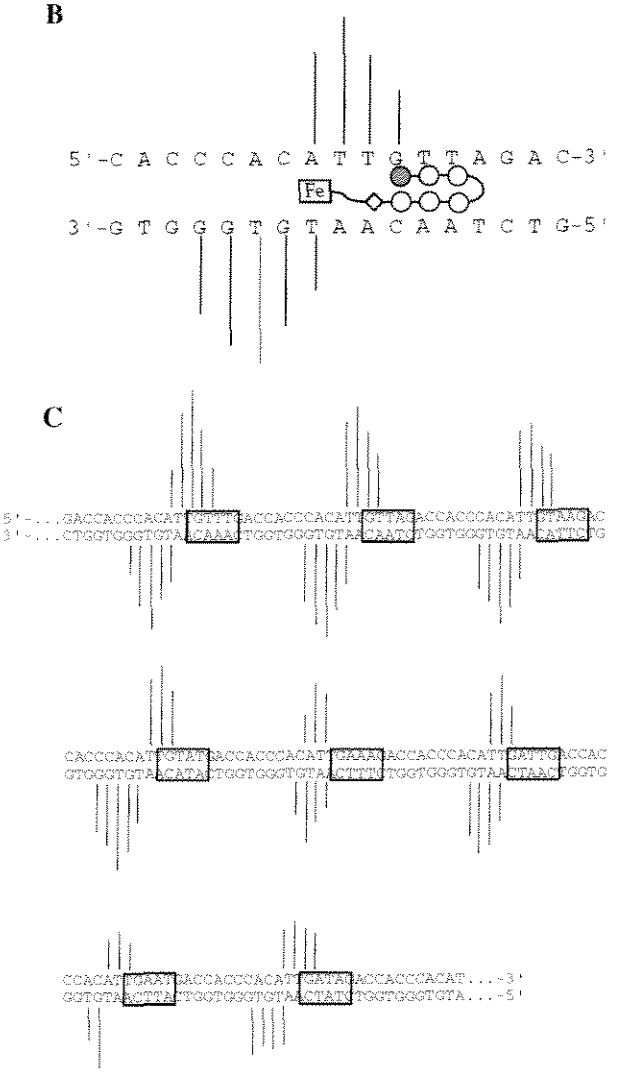
Results and Discussion

Hairpin Motif. Affinity Cleavage experiments using a polyamide with Fe(II)•EDTA at the carboxy terminus confirm that ImPyPy-γ-PyPyPy-β-Dp binds each discrete site with a *single orientation*. Cleavage experiments with ImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II) were performed on the 370 base pair restriction fragment radiolabeled at either the 5' or 3' end (20 mM HEPES, 200 mM NaCl, 50 mg/ml glycogen, pH 7.0, 22 °C, 5 mM DTT, 1 mM Fe(II)). A single cleavage locus is observed proximal to the 5'-side of each of the eight 5'-TG(A,T)$_3$-3' binding sites indicating that the carboxy terminus of the polyamide is located at the 5'-side of each binding site. A 3'-shifted asymmetric cleavage pattern is consistent with location of the 1:1 polyamide complex in the minor groove (Figure 4.4).

The observation of a single cleavage locus is consistent only with an oriented 1:1 complex and rules out any 2:1 overlapped or extended binding motifs.[6] A 1:1 oriented but extended motif would require at least an 8 base pair binding site which is inconsistent with high resolution MPE footprinting data.[4e] The hairpin structure is supported by direct NMR structure studies on a six ring hairpin polyamide of sequence composition ImPyPy-γ-PyPyPy binding to a core 5 base pair 5'-TGTTA-3' site.[7]

Figure 4.4 (A) Storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel used to separate the fragments generated by affinity cleaving experiments performed with ImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II): lanes 1 digestion products obtained in the absence of polyamide, lanes 2 and 3, A and G sequencing lanes; lanes 4-11, digestion products obtained in the presence of **2**, 50 pM, 100 pM, 1 nM, 5 nM, 10 nM, 50 nM, 100nM, and 1 μM. The targeted binding sites are indicated on the right side of the autoradiograms. All reactions contain 20 kcpm 3'-^{32}P restriction fragment, 20 mM HEPES, 200 mM NaCl, 50 mg/mL glycogen, pH 7. (B) Schematic binding model of affinity cleaving at the 5'-TGTTA-3' binding site. The imidazole and pyrrole rings are represented as shaded and unshaded spheres respectively, the β-alanine residue is represented as an unshaded diamond, the position of the iron is represented as a rectangle. (C) Cleavage of the 370 bp restriction fragment from pDEH1 at 100 nM concentration of polyamide. Lines are proportional to the integrated densities of the cleavage bands.


A
Control
A
G
50 pM
1 µM
1 2 3 4 5 6 7 8 9 10 11
3'5'
5'3'
32p


B
5'-C A C C C A C A T T G T T A G A C-3'
Fe
3'-G T G G G T G T A A C A A T C T G-5'
C

Relative Energetics. DNase I footprinting on the 3'-^{32}P end-labeled 370 base pair *EcoR* I/ *Pvu* II restriction fragment from the plasmid pDEH1 (10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, 22°C) reveals the equilibrium association constants for ImPyPy-γ-PyPyPy-β-Dp binding each of the five base pair sites, 5'-TG(A,T)$_3$-3', ranges from $K_a = 2.1 \times 10^8$ M^{-1} to $K_a = 1.8 \times 10^7$ M^{-1} in decreasing order: 5'-TGTTT-3' > 5'-TGTTA-3' > 5'-TGTAA-3' > 5'-TGTAT-3' > 5'-TGATT-3' > 5'-TGATA-3' > 5'-TGAAA-3' > 5'-TGAAT-3' (Table 4.1, Figure 4.5). The polyamide displays a binding isotherm (eq. 2, n = 1) consistent with binding as an intramolecular hairpin at all sites (Figure 4.6).

The affinities of ImPyPy-γ-PyPyPy-β-Dp for binding sites of the type 5'-TG(A,T)$_3$-3' may be grouped into two sets according to sequence composition: 5'-TG**T**(A,T)$_2$-3' and 5'-TG**A**(A,T)$_2$-3'. ImPyPy-γ-PyPyPy-β-Dp binds 5'-TG**T**(A,T)$_2$-3' sites with between 2-fold and 12-fold higher affinity than 5'-TG**A**(A,T)$_2$-3' sites. Overall, sequence composition results in a difference of up to 1.5 kcal/mole for recognition of 5'-TG(A,T)$_3$-3' sites. X-ray diffraction data suggest that a G-A step acts to narrow the minor groove of B-form DNA.[5a,5b,8] A decrease in minor groove width and flexibility would act to disfavor binding of a hairpin polyamide, which prefers a wide, flexible minor groove for favorable binding.[1c]

Implications for the Design of Minor Groove Binding Polyamides.

The results reported here indicate that A•T and T•A base pairs are degenerate in the hairpin polyamide-DNA motif with a 12-fold difference in binding affinities. These results indicate that at least a 10-fold *range* of binding affinities and sequence specificities will be observed for a polyamide binding to a designated set of match sites containing A•T base pairs. The similarity of the polyamide binding affinities for the eight 5'-TG(A,T)$_3$-3' match sites reflects a limit to the specificity of the hairpin polyamide binding motif. Because G•C is

Figure 4.5 Quantitative DNase I footprint titration experiment with ImPyPy-γ-PyPyPy-β-Dp on the 3'-^{32}P-labeled 370 bp *EcoR* I/ *Pvu* II restriction fragment from plasmid pDEH1. The eight binding sites 5'-TG(A,T)$_3$-3' are shown on the right side of the storage phosphor autoradiogram. All reactions contain 20 kcpm restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0. Lane 1, A reaction; lane 2, G reaction; lanes 3 and 21, DNase I standard; lanes 4-20 contain 0.1 nM, 0.2 nM, 0.5 nM, 1 nM, 1.5 nM, 2.5 nM, 4.0 nM, 6.5 nM, 10 nM, 15 nM, 25 nM, 40 nM, 65 nM, 100 nM, 200 nM, 500 nM, 1 μM ImPyPy-γ-PyPyPy-β-Dp, respectively.

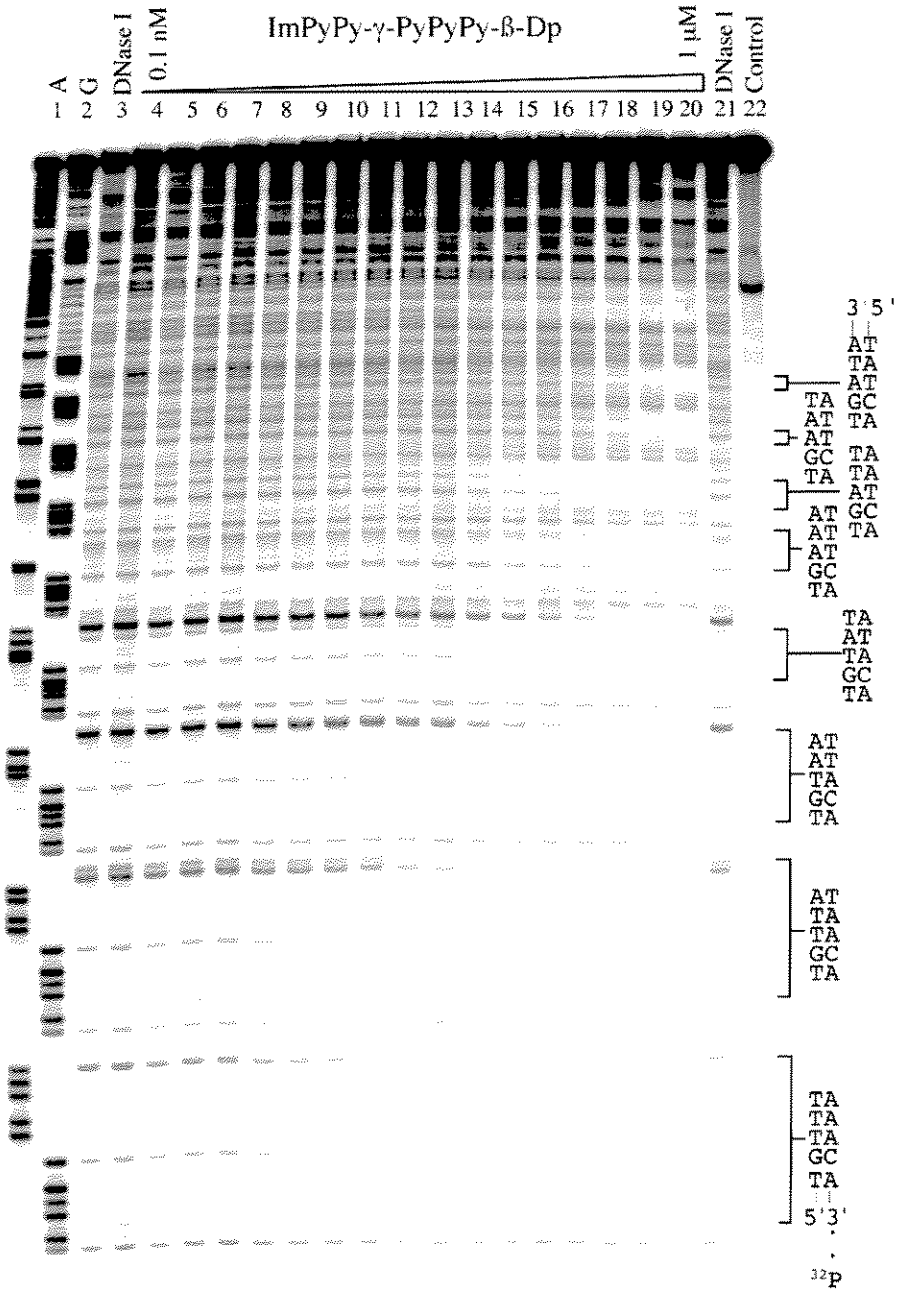

A
G
DNase I
0.1 nM
ImPyPy-γ-PyPyPy-ß-Dp
1 μM
DNase I
Control
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
3'
5'
AT
TA
AT GC
TA
TA AT
AT GC
TA
AT AT
AT GC
TA
TA AT
TA GC
TA
AT AT
TA GC
TA
AT TA
TA GC
TA
TA TA
TA GC
TA
5'
3'
32P

Table 4.1 Equilibrium Association Constants and Binding Energies for ImPyPy-γ-PyPyPy-β-Dp [a,b]

Binding Site	K_a (M^{-1})	K_a (rel)[c]	ΔG(kcal/mol)
5'-TGTTT-3'	$2.1 (\pm 0.7) \times 10^8$	12	-11.2
5'-TGTTA-3'	$1.5 (\pm 0.4) \times 10^8$	8.4	-11.0
5'-TGTAA-3'	$7.3 (\pm 1.0) \times 10^7$	4.1	-10.6
5'-TGTAT-3'	$4.7 (\pm 0.8) \times 10^7$	2.6	-10.4
5'-TGATT-3'	$3.9 (\pm 1.0) \times 10^7$	2.2	-10.3
5'-TGATA-3'	$2.5 (\pm 0.9) \times 10^7$	1.4	-10.0
5'-TGAAA-3'	$2.2 (\pm 0.9) \times 10^7$	1.2	-9.9
5'-TGAAT-3'	$1.8 (\pm 0.8) \times 10^7$	1	-9.8

[a]Values reported are the mean values measured from four footprint titration experiments, with the standard deviation for each data set indicated in parentheses. [b]The assays were performed at 22 °C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM $MgCl_2$, 10 mM KCl, and 5 mM $CaCl_2$. [c]K_a (rel) = K_a(5'-TGWWW-3')/K_a(5'-TGAAT-3'), W is either A or T.

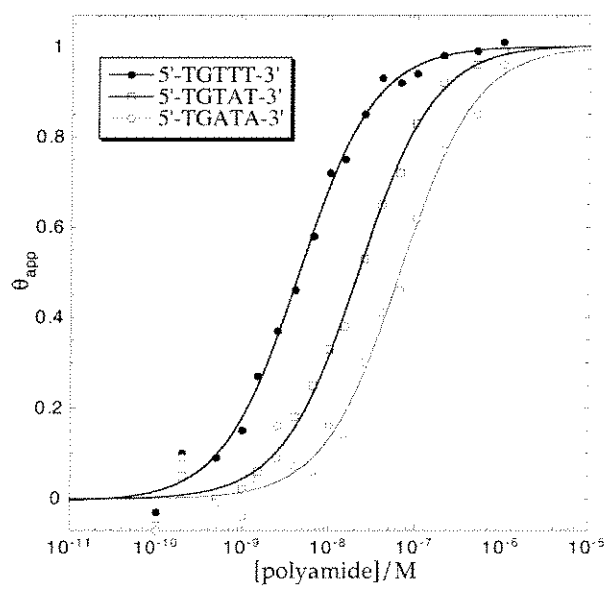


Figure 4.6 Data for the quantitative DNase I footprint titration experiments for ImPyPy-γ-PyPyPy-β-Dp with 5'-TGTTT-3', 5'-TGTAT-3', and 5'-TGATA-3'. The θapp points were obtained using phosphostimulable storage phosphor autoradiography and processed as described in Experimental Section. = 5'-TGTTT-3', = 5'-TGTAT-3', = 5'-TGATA-3'. The solid lines are the best fit Langmuir binding titration isotherms obtained from a nonlinear least-squares algorithm using eq. 2.

distinct from C•G, the most specific recognition will be G,C rich sequences. In order to increase the specificity of pyrrole-imidazole polyamides for sequences of rich A,T composition, methods of recognition to discriminate A•T and T•A base pairs are needed.

Experimental Section

Materials Sonicated, deproteinized calf thymus DNA, from Pharmacia, was dissolved in filter sterilized water to a final concentration of 800 μM in base pairs and stored at 4 °C. Glycogen was purchased from Boehringer-Manheim as a 20 mg/mL aqueous solution. Nucleotide triphosphates were purchased from Pharmacia and used as supplied. Nucleoside triphosphates labeled with ^{32}P ($\geq$ 3000 C_i /mmol) were obtained from Dupont-New England Nuclear. Cerenkov radioactivity was measured with a Beckman LS 2801 scintillation counter. All enzymes were purchased from Boehringer-Manheim and were used according to the supplier's recommended protocol in the activity buffer provided. Plasmid pUC19 was obtained from Worthington Biochemical. A solution of 0.5 M EDTA, pH 8.0 was purchased from Ultrapure. Phosphoramidites were from Glen Research. The pH of buffer solutions was recorded using a digital pH/millivolt meter (model no. 611, Orion Research) and a ROSS semimicro combination pH electrode. General manipulation of duplex DNA[9] and oligonucleotides[10] were performed according to established procedures.

Synthesis of ImPyPy-γ-PyPyPy-β-Dp and ImPyPy-γ-PyPyPy-β-Dp-EDTA. Polyamides optimized for hairpin formation[4e,f] were synthesized from β-alanine-PAM resin using solid phase methods[11], and characterized by a combination of analytical HPLC, ^1H NMR spectroscopy, and MALDI-TOF mass spectroscopy. Mass observed for ImPyPy-γ-PyPyPy-β-

Dp, 978.0; 978.1 calculated; 1295.4 observed for ImPyPy-γ-PyPyPy-β-Dp-EDTA; 1295.3 calculated.

Construction of Plasmid DNA. Oligodeoxynucleotides were synthesized by standard automated solid support chemistry using an Applied Biosystems model 380B DNA synthesizer and *O*-cyanoethyl-*N,N*-diisopropyl phosphoramidites. Plasmid pDEH1 was prepared by hybridization of two complementary sets of synthetic oligonucleotides: 1) 5'-CTA-GACCACCATTGTTTGACCACCCACATTGTTAGACCACCCACATTGTAAGAC CACCCACATTGTATGACCACC-3'; 2) 5'-GTCATACAATGTGGGTGGTCTTACAAT GTGGGTGGTCTAACAATGTGGGTGGTCAAACAATGTGGGTGGT-3'; 3) 5'-CACAT TGAAAGACCACCCACATTGATTGACCACCCACATTGAATGACCACCCACATTGAT AGACCACCCACATTGCA-3'; 4) 5'-ATGTGGGTGGTCTATCAATGTGGGTGGTCATT CAATGTGGGTGGTCAATCAATGTGGGTGGTCTTTCAATGTGGGTG-3'. Oligonucleotides 2) and 3) were phosphorylated with dATP and T4 polynucleotide kinase, and then annealed to their respective complementary strands 1) and 4). The two sets of duplexes were then ligated to the large pUC19 *Xba* I/ *Pst* I restriction fragment using T4 DNA ligase. The ligated plasmid was then used to transform Epicurean^{TM} Coli XL-1 Blue Supercompetant cells. Colonies were selected for α-complementation on 25 mL Luria-Bertani medium agar plates containing 50 mg/mL ampicillin and treated with XGAL and IPTG solutions. Large scale plasmid purification was performed using Qiagen purification kits. The presence of the desired insert was determined by dideoxy sequencing using a USB Sequenase v. 2.0 kit. Plasmid DNA concentration was determined at 260 nm using the relation 1 OD unit = 50 mg/mL duplex DNA. The plasmid pDEH1 was digested with *Eco* RI, labeled at the 3' end using Sequenase v. 2.0 and digested with *Pvu* II. The 370 base pair

fragment was isolated by nondenaturing gel electrophoresis and used in all experiments described here. For affinity cleaving reactions, pDEH1 also was 5'-^{32}P labeled. First, the plasmid pDEH1 was digested with *Eco* RI, then dephosphorylated with calf intestine alkaline phosphatase. The digested plasmid then was 5'-^{32}P labeled with γ-^{32}P dATP using T4 polynucleotide kinase and digested with *Pvu* II. The 5' labeled fragment was isolated by nondenaturing gel electrophoresis and used in affinity cleaving experiments. Chemical sequencing adenine-specific reactions were carried out as previously described.[12]

Quantitative DNase I Footprint Titrations. All reactions were executed in a total volume of 400 μL.[13] A polyamide stock solution or H_2O (for reference lanes) was added to an assay buffer containing 3'-^{32}P radiolabeled restriction fragment (20,000 cpm), affording final solution conditions of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, and either (i) 0.1 nM- 1 μM polyamide ImPyPy-γ-PyPyPy-β-Dp, or (ii) no polyamide (for reference lanes). The solutions were allowed to equilibrate for 24 hours at 22 °C. Footprinting reactions were initiated by the addition of 10 μL of a stock solution of DNase I (at the appropriate concentration to give ~55% intact DNA) containing 1 mM dithiothreitol and allowed to proceed for 7 minutes at 22 °C. The reactions were stopped by the addition of 50 μL of a solution containing 2.25 M NaCl, 150 mM EDTA, 23 μM base pair calf thymus DNA, and 0.6 mg/ml glycogen, and ethanol precipitated. The reactions were resuspended in 1 × TBE/ 80% formamide loading buffer, denatured by heating at 85 °C for 15 minutes, and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% crosslinking, 7 M urea) in 1 × TBE at 2000 V for 1.5 h. Gels were dried on a slab dryer and then exposed to a storage phosphor screen at 22 °C.

Quantitation by Storage Phosphor Technology Autoradiography. Photostimuable storage phosphor imaging plates (Kodak Storage Phosphor Screen SO230 obtained from Molecular Dynamics) were pressed flat against dried gel samples and exposed in the dark at 22 °C for 12-24 hours. A Molecular Dynamics 400S PhosphorImager was used to obtain all data from the storage screens.[14] The data were analyzed by performing volume integration of the target site and reference blocks using the ImageQuant v. 3.3 software running on a Compaq Pentium 80.

Quantitative DNase I Footprint Titration Data Analysis. 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:

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

where I_{site}° and I_{ref}° are the site and reference intensities, respectively, from a DNase I control lane to which no polyamide was added.

The ($[L]_{tot}$, θ_{app}) data 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:

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

where $[L_{tot}]$ 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 of KaleidaGraph software (v. 3.0.1, Abelbeck Software) with K_a, θ_{max}, and θ_{min} as the adjustable parameters. The goodness of fit of the binding curve to the data points is evaluated by the correlation coefficient, with $R > 0.97$ as the criterion for an acceptable fit. Four sets of acceptable data were used in determining each association constant. All lanes from a gel were used unless a visual inspection revealed a data point to be obviously flawed relative to neighboring points. The data were normalized using the following equation:

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

At higher concentrations of polyamide, (> 1 μM for ImPyPy-γ-PyPyPy-β-Dp), the reference site becomes partially protected due to nonspecific DNA binding, resulting in low θ_{app} values. For this reason, higher concentrations were not used.

Affinity Cleaving Titrations. All affinity cleavage reactions[15] were performed in a total volume of 400 μL. A polyamide stock solution (ImPyPy-γ-PyPyPy-β-Dp-EDTA) or H_2O (for reference lanes) was added to an assay buffer containing either 3'-^{32}P labeled or 5'-^{32}P labeled restriction fragment (20,000 cpm), affording final solution conditions of 20 mM HEPES, 200 mM NaCl, 50 mg/mL glycogen, pH 7, and either (i) 0.1 nM-1 μM polyamide ImPyPy-γ-PyPyPy-β-Dp-EDTA, or (ii) no polyamide (for reference lanes). Solutions were incubated at 22 °C for 24 hours. Then, 20 μL of a 20 mM ferrous ammonium sulfate solution was added, and the solution was allowed to equilibrate for 20 minutes at 22 °C. Affinity cleaving reactions were initiated by the addition of 40 μL of a 50 mM dithiothreitol solution, and reacted for 11 minutes at 22 °C. Reactions were stopped by the addition of 10 μL of a solution containing 2.8 mg/mL glycogen and 112 μM base pairs calf thymus DNA, and ethanol precipitated. Reactions were resuspended in 1 × TBE/ 80% formamide loading buffer, denatured by heating at 85 °C for 15 minutes, and placed on ice. Reaction products

buffer, denatured by heating at 85 °C for 15 minutes, and placed on ice. Reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% crosslinking, 7 M urea) in 1 × TBE at 2000 V for either 1.5 hours (for visualization of first four binding sites) or 3 hours (for visualization of last four binding sites). Gels were dried on a slab dryer and then exposed to a storage phophor screen at 22 °C.

Data Analysis of Affinity Cleavage Titrations. The data were analyzed by performing volume integrations of the cleavage bands and reference bands using the ImageQuant v. 3.3 software running on a Compaq Pentium 80. Background-corrected volume integration of rectangles encompassing cleavage bands was normalized to a maximum value of 3.95.

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.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 9892. (b) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *115*, 3663. (c) Dwyer, T.J.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 9900. (d) Mrksich, M.; Parks, M.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 7983. (e) Parks, M.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6147. (f) Parks, M.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6153. (g) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Chem. Biol.* **1996**, *3*, 369. (h) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature 1996*, **382**, 559.

5. (a) Dickerson, R.E.; Goodsell, D.S.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3579. (b) Yoon, C.; Prive, G.G.; Goodsell, D.S.; Dickerson, R.E. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6332.

6. Trauger, J.W.; Baird, E.E.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6160.

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) Larsen, T.A.; Kopka, M.L.; Dickerson, R.E. *Biochemistry* **1991**, *30*, 4443. (b) Narayana, N.; Ginell, S.L.; Russu, I.M.; Berman, H.M. *Biochemistry* **1991**, *30*, 4449. (c) Yanagi, K.; Prive, G.G.; Dickerson, R.E. *J. Mol. Biol.* **1991**, *217*, 201.

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

10. Gait, M.J. *Oligonucleotide Synthesis: A Practical Approach*; IRL Press: Oxford, 1984.

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

12. Iverson, B.L.; Dervan, P.B. *Nucleic Acids Res. 1987, 15, 7823.*

13. Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods Enzymol. 1986, 130, 132.*

14. Johnston, R.F.; Picket, S.C.; Barker, D.L. *Electrophoresis 1990, 11, 355.*

15. Schultz, P.G.; Taylor, J.S.; Dervan, P.B. *J. Am. Chem. Soc. 1982, 104, 6861.*

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.*[1-5] *Cell-permeable small molecules which target predetermined DNA sequences offer a potential approach for the regulation of gene-expression.*[6] *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.*[3-4] *Although oligonucleotides and their analogs have been shown to interfere with gene expression*[7-8]*, 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.*[9-11] *An Im/Py pair distinguishes G•C from C•G and both of these from A•T/T•A base pairs.*[9-11] *A Py/Py pair specifies A,T from G,C but does not distinguish A•T from T•A.*[9-14] *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.
White, Turner, Szewczyk, Baird, & Dervan manuscript in preparation.
Kielkopf, White, Szewczyk, Turner, Baird, Dervan, & Rees submitted.

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.[15] However, despite the central placement of the guanine exocyclic N2 amine group in the G,C minor groove[15-16], Py/Im and Im/Py pairings distinguish energetically G•C and C•G.[9-11,17-18] 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.[19-20] 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).[15] 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.[21] 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 concentrations.[11,18] 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 5.2b). 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 experiments[22] on a 3' ^{32}P-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 5.9, Table 5.1). The binding orientations and stoichiometries of the

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 red.

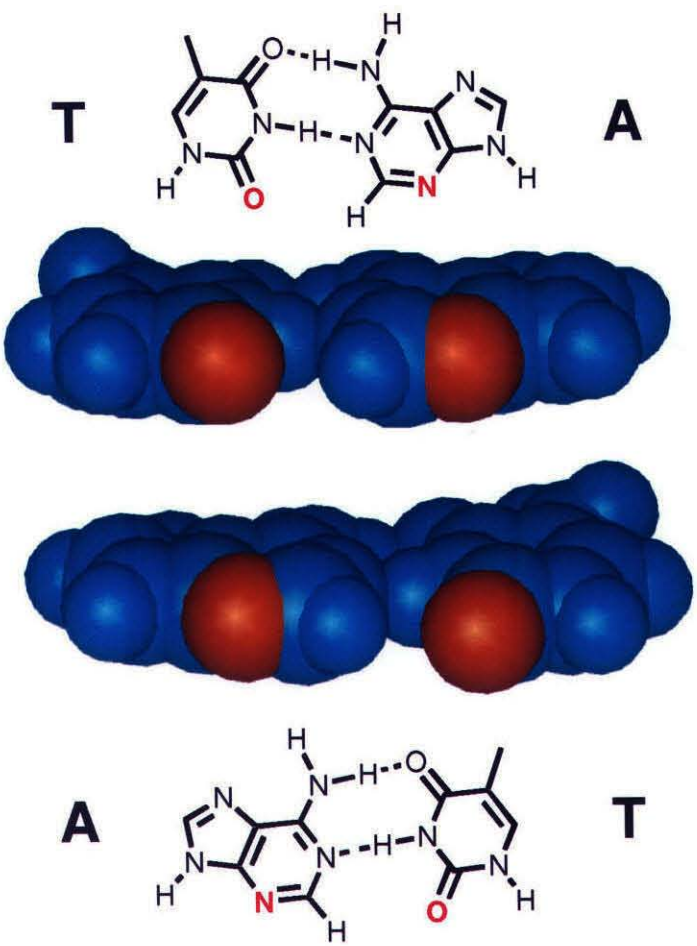
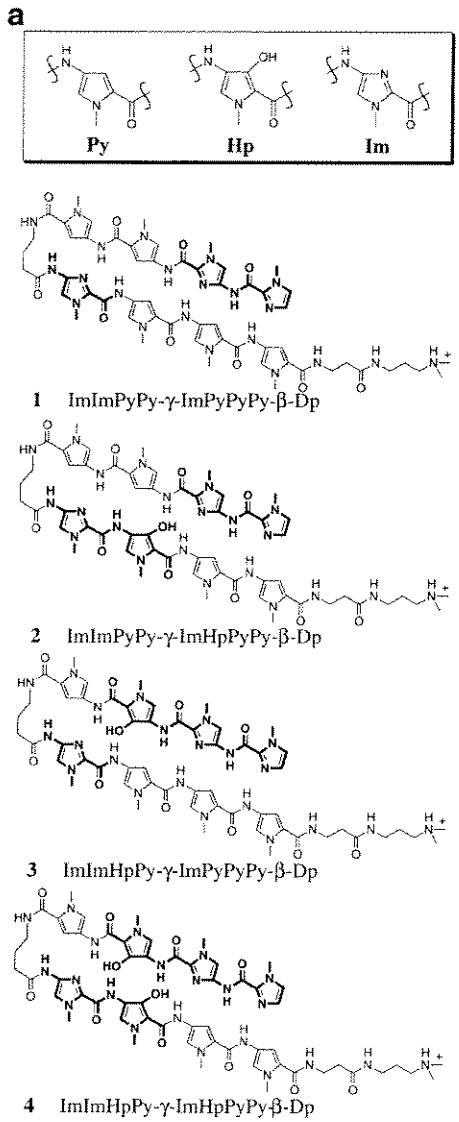
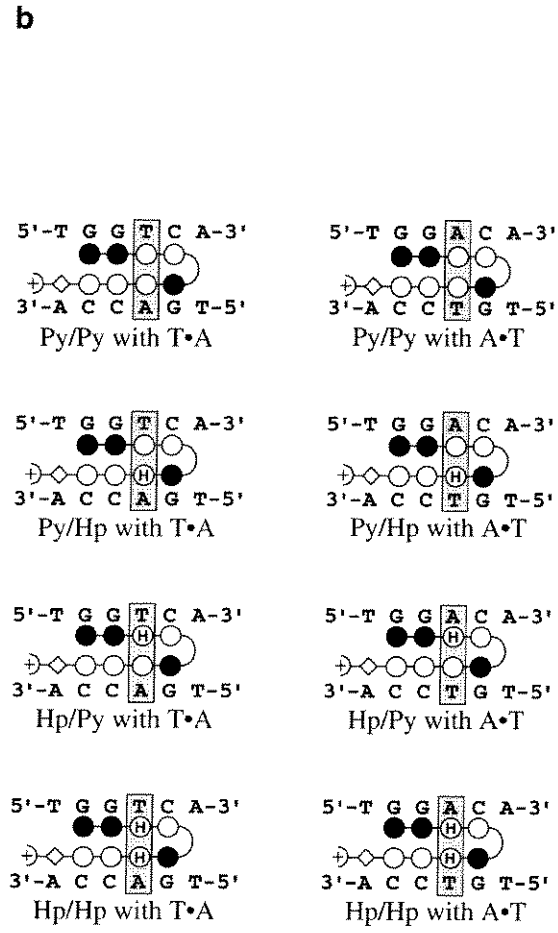
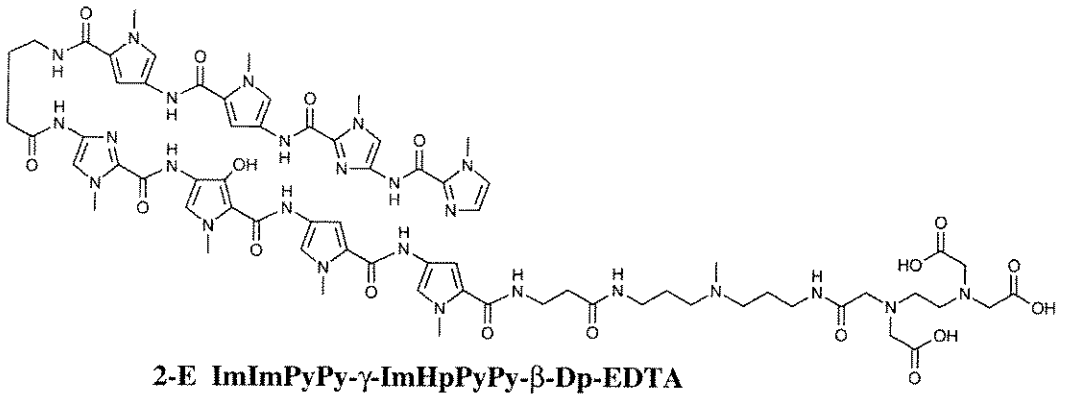

T
A
A
T

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**), ImImHpPy-γ-ImPyPyPy-β-Dp (**3**), and ImImHpPy-γ-ImHpPyPy-β-Dp (**4**). (Hp = 3-hydroxypyrrole, Im = imidazole, Py = pyrrole, β = β-alanine, γ = γ-aminobutyric acid, Dp = dimethylaminopropylamide). **b**, Binding models for polyamides **1-4** 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 and 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.

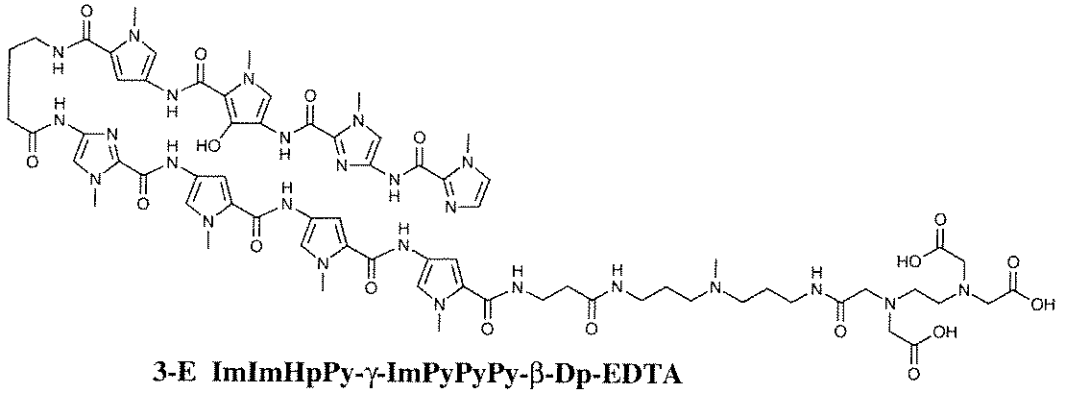

a
Py
Hp
Im
1 ImImPyPy-γ-ImPyPyPy-β-Dp
2 ImImPyPy-γ-ImHpPyPy-β-Dp
3 ImImHpPy-γ-ImPyPyPy-β-Dp
4 ImImHpPy-γ-ImHpPyPy-β-Dp


b
5'-T G G T C A-3'
3'-A C C A G T-5'
Py/Py with T•A
5'-T G G A C A-3'
3'-A C C T G T-5'
Py/Py with A•T
Py/Hp with T•A
Py/Hp with A•T
Hp/Py with T•A
Hp/Py with A•T
Hp/Hp with T•A
Hp/Hp with A•T

corresponding EDTA analogs ImImPyPy-γ-ImHpPyPy-β-Dp-EDTA•Fe(II) **2-E•Fe(II)** and ImImHpPy-γ-ImPyPyPy-β-Dp-EDTA•Fe(II) **3-E•Fe(II)** were determined by affinity cleaving footprint titration experiments (Figures 5.3, 5.8).[2,23] To test whether a Hp/Hp pair



2-E ImImPyPy-γ-ImHpPyPy-β-Dp-EDTA



3-E ImImHpPy-γ-ImPyPyPy-β-Dp-EDTA

Figure 5.3 Chemical structures of affinity cleaving polyamides ImImPyPy-γ-ImHpPyPy-β-Dp-EDTA **2-E** and ImImHpPy-γ-ImPyPyPy-β-Dp-EDTA **3-E**.

is degenerate for recognition of A•T and T•A base pairs, the polyamide ImImHpPy-γ-ImHpPyPy-β-Dp **4** (Figure 5.2) was studied for binding versus the sites 5'-TGGACA-3' and 5'-TGGTCA-3' (Table 5.1). In addition, a series of six-ring hairpins containing a single Hp/Py ring pairing was studied versus T•A and A•T base pairs in a variety of sequence contexts to test the generality of the 3 amino acid pairing rules. Equilibrium association constants were determined for ImHpPy-γ-PyPyPy-β-Dp **5**, ImPyPy-γ-PyHpPy-β-Dp **6**, ImPyHp-γ-PyPyPy-β-Dp **7**, and ImPyPy-γ-HpPyPy-β-Dp **8** on a 3'-^{32}P end-labeled DNA fragment containing eight possible five base pair 5'-TG(A,T)$_3$-3' match sites (Figure 5.4).[13]

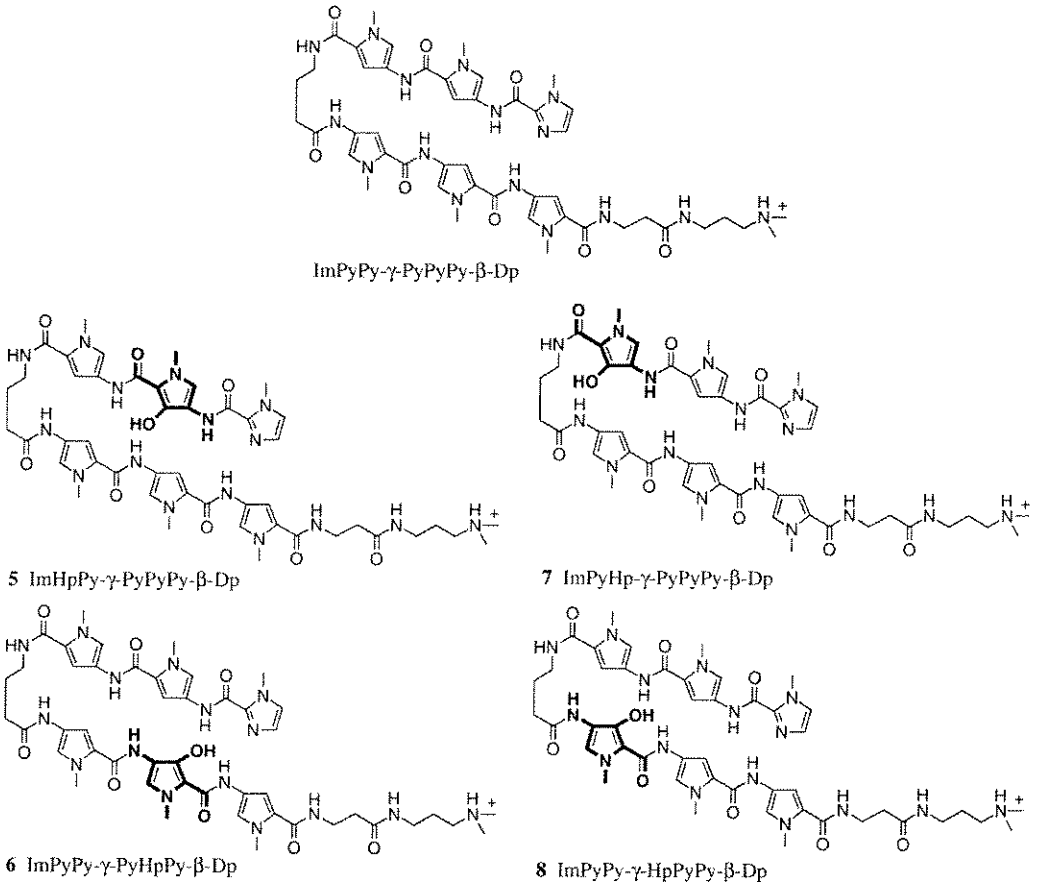


Figure 5.4 Chemical structures of 6-ring hairpin hydroxypyrrole-imidazole-pyrrole polyamides. ImPyPy-γ-PyPyPy-β-Dp has been previously reported.[13]

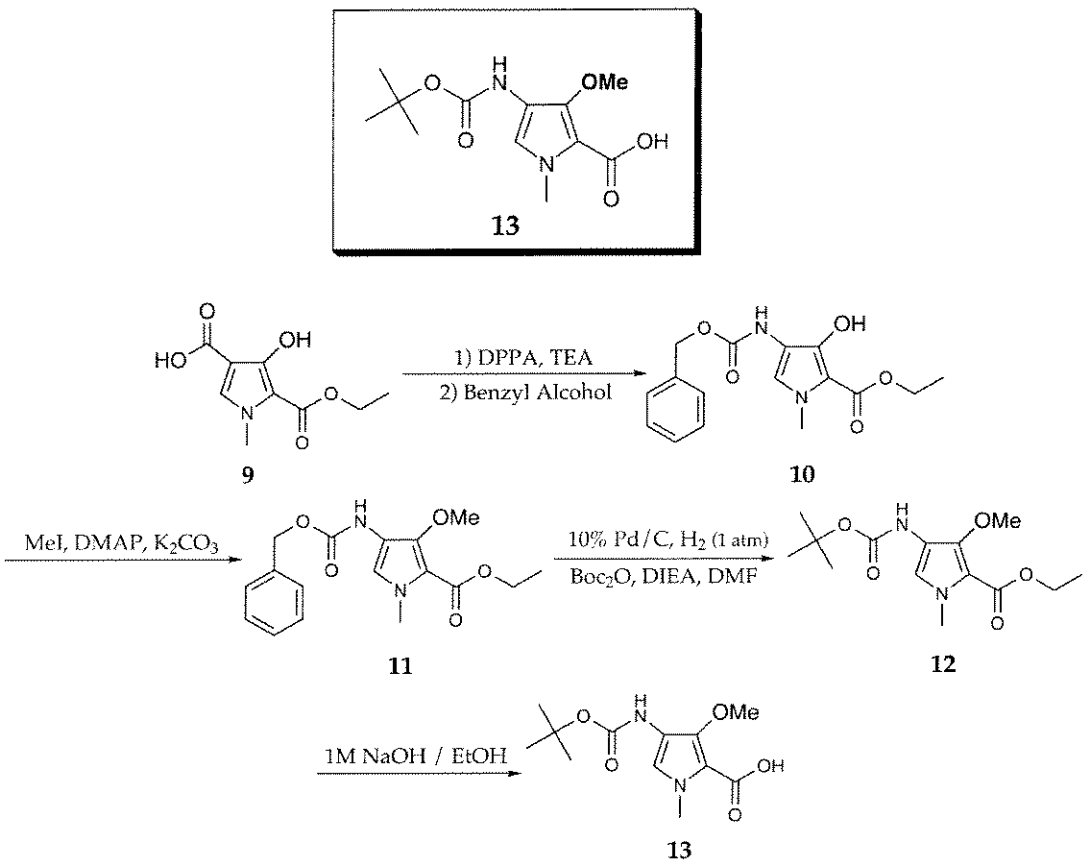


Figure 5.5 The synthetic scheme for 3-O-methyl-N-Boc protected pyrrole-2-carboxylate **13**. The hydroxypyrrole monoester **9** can be prepared in 0.5 kg quantity using published procedures on enlarged scale.[24]

Results and Discussion

Synthesis of 3-hydroxypyrrole amino acid The 3-O-methyl-N-Boc protected pyrrole-2-carboxylate was synthesized on multigram scale starting from the previously described hydroxypyrrole monoester **9** (Figure 5.5).[24] First, the 4-carboxy-3-hydroxypyrrole monoester was converted to 4-benzyl carbamate 3-hydroxypyrrole **10** using DPPA (TEA, CH_3CN, reflux, 5 h), followed by the addition of benzyl alcohol (reflux, 12 h). Then, the 3-hydroxy group of the pyrrole was protected as the methyl ether, giving **11** via alkylation of **10** using methyliodide (anhydrous K_2CO_3, DMAP, 22 °C, 12 h). The 4-benzyl carbamate 3-methoxypyrrole **11** was converted to the 4-Boc protected 3-methoxypyrrole **12** by removal of the benzyl group via reduction, then in situ reaction of the deprotected 4-amine group with Boc anhydride (H_2, 1 atm, Pd/C, DIEA, 2.1 h). Finally, the 2-ethyl ester of Boc-protected 3-methoxypyrrole **12** was hydrolysed (1M NaOH, 22 °C, 4 days) to give the 3-O-methyl-N-Boc protected pyrrole 2-carboxylate **13** for use in solid phase protocols.

Solid phase synthesis of hydroxypyrrole-imidazole-pyrrole polyamides Hydroxypyrrole-imidazole-pyrrole polyamides were then synthesized using previously described solid phase protocols starting from the commercially available Boc-β-Pam-Resin (Figure 5.6).[25] In machine synthesis protocols, 3-hydroxypyrrole-Boc-amino 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. The methyl-protected polyamides **1** through **8** were synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc-β-Pam-Resin, cleaved from the resin by aminolysis with dimethylaminopropylamine, and then purified by reversed phase HPLC (Figure 5.6). The 3-methoxypyrrole amino acids were deprotected by treating the polyamide with sodium thiophenoxide in DMF at 100 °C for 2 h. For affinity cleaving analogs of polyamides **2** and **3** (Figure 5.7), however, the polyamides were cleaved from the resin using neat 3,3'-diamino-*N*-methyldipropylamine (55 °C, 16 h). The methyl-protected amine modified polyamides **2-NH$_2$** and **3-NH$_2$**, after purification by reversed phase HPLC, were

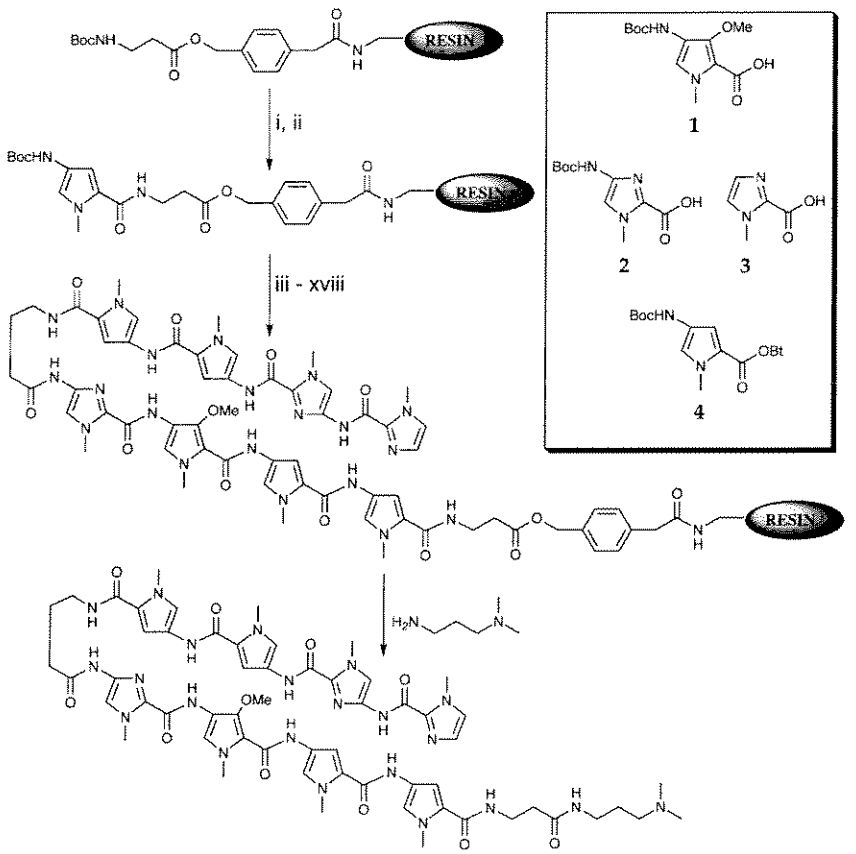


Figure 5.6 Solid phase synthetic scheme for ImImPyPy-γ-ImHpPyPy-β-Dp starting from Boc-β-Pam-Resin: (i) 80% TFA/DCM, 0.4 M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4 M PhSH; (iv) Boc-Py-OBt, DIEA, DMF; (v) 80% TFA/DCM, 0.4 M PhSH; (vi) Boc-3-OMe-Py-OH, HBTU, DMF, DIEA; (vii) 80% TFA/DCM, 0.4 M PhSH; (viii) Boc-Im-OH, DCC, HOBt; (ix) 80% TFA/DCM, 0.4 M PhSH; (x) Boc-γ-aminobutyric acid, 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) Boc-Py-OBt, DIEA, DMF; (xv) 80% TFA/DCM, 0.4 M PhSH; (xvi) Boc-Im-OH, DCC, HOBt; (xvii) 80% TFA/DCM, 0.4 M PhSH; (xviii) imidazole-2-carboxylic acid, HBTU, DIEA; (xix) dimethylaminopropylamine, 55 °C, 18 h. Purification by reversed phase HPLC provides ImImPyPy-γ-ImOpPyPy-β-Dp. (Op = 3-methoxypyrrole). Treatment of the 3-methoxypyrrole polyamide with thiophenol, NaH, DMF at 100 °C for 120 min provides polyamide **2** after purification.

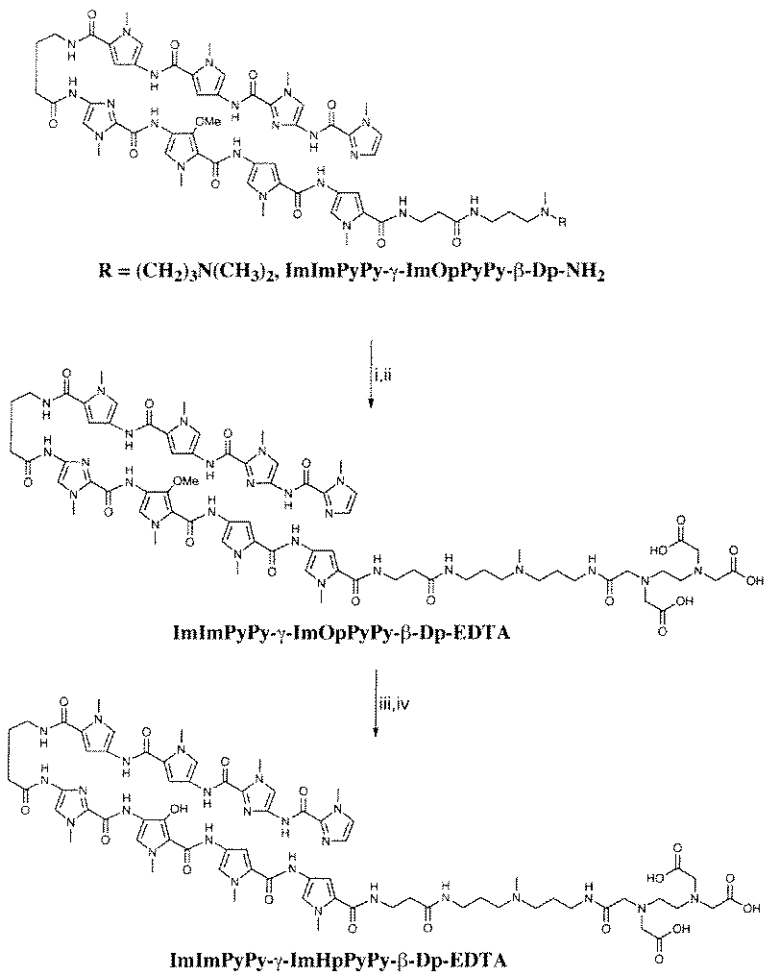


Figure 5.7 The synthesis of the bifunctional polyamide which incorporates the Hp/Py pair. Treatment of a sample of ImImPyPy-γ-ImOpPyPy-β-Pam-Resin with 3,3'-diamino-*N*-methyldipropylamine, 55 °C, 18 h followed by reversed phase HPLC purification provides the Op polyamide with a free primary amine group which can be coupled to an activated carboxylic acid derivative. Treatment with (I) EDTA-dianhydride, DMSO/NMP, DIEA, 55 °C; (ii) 0.1 M NaOH, followed by reversed phase HPLC purification provides the Op-Py-Im-polyamide-EDTA conjugate. Treatment of the 3-methoxypyrrole polyamide with thiophenol, NaH, DMF, at 100 °C for 2 h provides polyamide **2-E** after reversed phase HPLC purification.

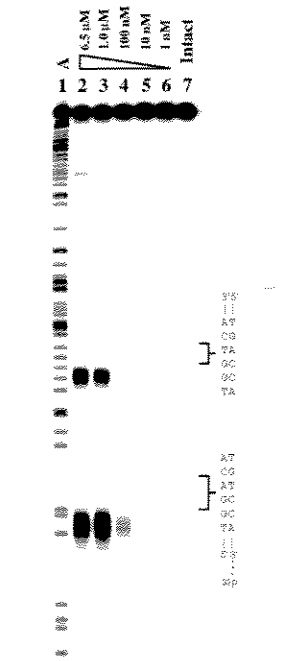
reacted with excess EDTA-dianhydride (DMSO/NMP, DIEA, 55 °C, 25 min.) and the remaining anhydride was hydrolyzed (0.1 M NaOH, 55 °C, 10 min.). After HPLC purification, the EDTA modified polyamides were deprotected with sodium thiophenoxide to give the affinity cleaving analogs **2-E** and **3-E**. Polyamide identity and purity were confirmed by ^{1}H NMR, analytical HPLC, and matrix-assisted laser-desorption ionization time-of-flight mass spectrometry.

Identification of binding orientation and stoichiometry by affinity cleaving Affinity cleaving titration experiments using hairpin polyamides modified with EDTA•Fe(II) at the C-terminus were used to determine polyamide binding orientation and stoichiometry (Figure 5.8). Affinity cleaving experiments were performed on a 3'-^{32}P end-labeled 250-bp pJK6 *Eco*RI/ *Pvu*II restriction fragment (25 mM Tris-acetate, 20 mM NaCl, 100 μM/bp calf thymus DNA, pH 7.0). The observed cleavage patterns are in both cases 3'-shifted, indicating minor groove occupancy.[23] A single cleavage locus proximal to the 5'-side of the binding site is consistent with binding of one hairpin polyamide in a single orientation at each site.

Analysis of energetics by quantitative DNase I footprint titrations DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0, 22 °C) were performed to determine the equilibrium association constants K_a for recognition of bound sites (Figure 5.9, Table 5.1).[22] 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 (±0.9) × 10^{10} M^{-1} or 6.8 (±0.4) × 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 (±1.0) × 10^7 M^{-1} and 1.2 (±0.9) × 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 (±0.7) × 10^9 M^{-1} and

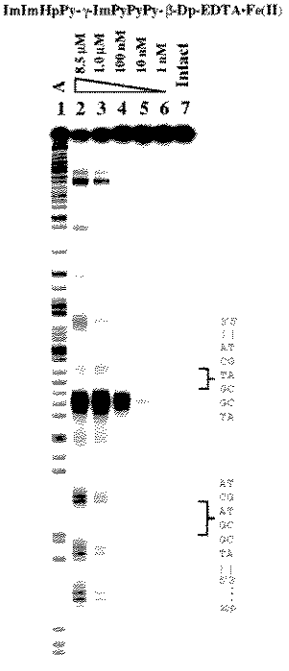
Figure 5.8 Affinity cleaving footprint titrations of hairpin polyamides ImImPyPy-γ-ImHpPyPy-β-Dp-EDTA•Fe(II) **2-E•Fe(II)** and ImImHpPy-γ-ImPyPyPy-β-Dp-EDTA•Fe(II) **3-E•Fe(II)**. Top and bottom left: Affinity cleavage experiments on a 3' ^{32}P labeled 250-bp pJK6 *EcoRI/ Pvu* II restriction fragment. The 5'-TGGACA-3' and 5'-TGGTCA-3' sites are shown on the right side of the autoradiograms. Top left: lane 1, adenine-specific chemical sequencing reaction; lanes 2-6, 6.5 μM, 1.0 μM, 100 nM, 10 nM, 1 nM polyamide **2-E•Fe(II)**; lane 7, intact DNA, no polyamide added. Bottom left: lane 1, A reaction; lanes 2-6, 8.5 μM, 1.0 μM, 100 nM, 10 nM, 1 nM polyamide **3-E•Fe(II)**; lane 7, intact DNA. Top and bottom right: Affinity cleavage patterns of **2-E•Fe(II)** and **3-E•Fe(II)** at 100 nM bound to 5'-TGGACA-3' and 5'-TGGTCA-3'. Bar heights are proportional to the relative cleavage intensities at each base pair. Shaded and unshaded circles denote imidazole and pyrrole carboxamides, respectively. Circles with an H represent the 3-hydroxypyrrole. Nonshaded diamonds represent the β-alanine moiety. A curved line represents the γ-aminobutryic acid, and the + represents the positively charged dimethylaminopropylamide tail group. The boxed **Fe** denotes the EDTA•Fe(II) moiety.

ImImPyPy-γ-ImHpPyPy-β-Dp-EDTA•Fe(II)



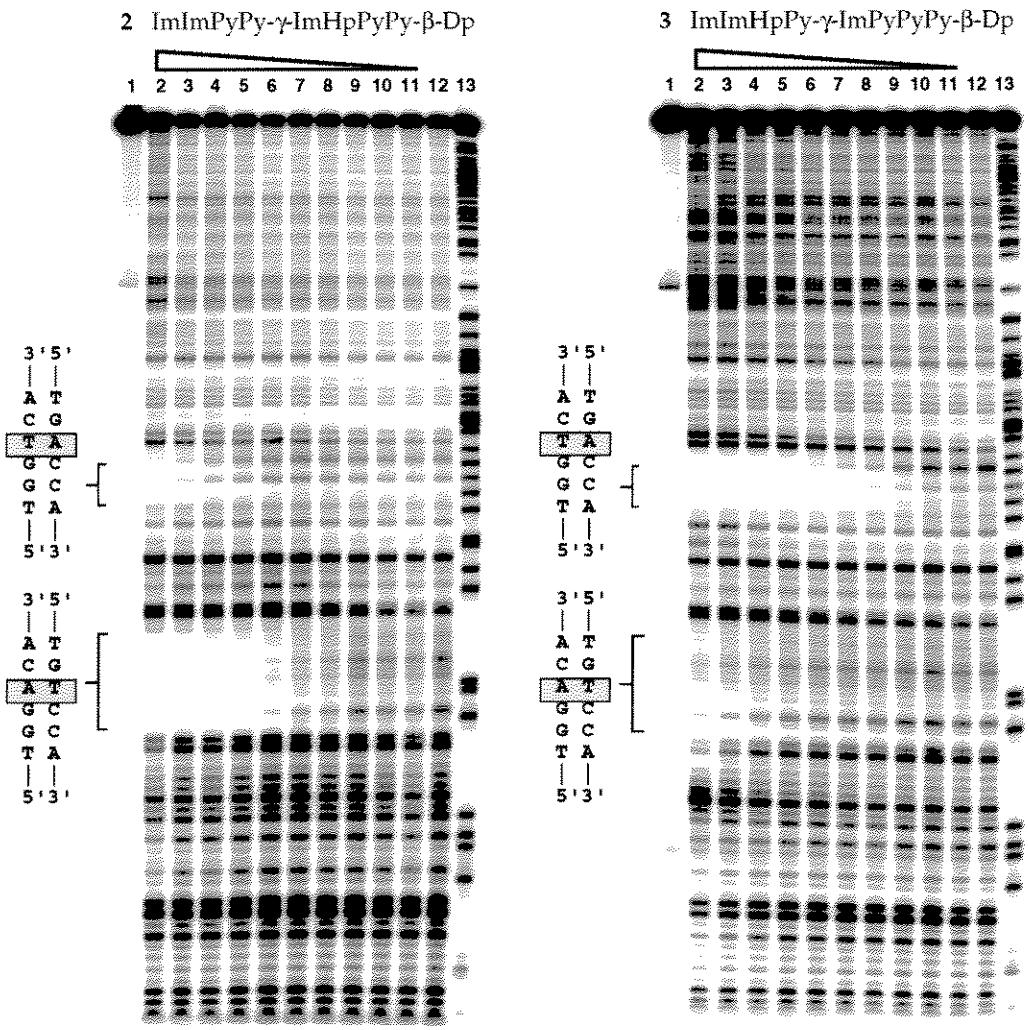
ImImPyPy-γ-ImHpPyPy-β-Dp-EDTA•Fe(II) 100 nM

ImImHpPy-γ-ImPyPyPy-β-Dp-EDTA•Fe(II)



ImImHpPy-γ-ImPyPyPy-β-Dp-EDTA•Fe(II) 100 nM

Figure 5.9 Quantitative DNase I footprint titration experiments. These experiments were performed with polyamides **2** and **3** on the 3' ^{32}P labeled 250-bp pJK6 *Eco* RI/ *Pvu* II restriction fragment. Lane 1, intact DNA; lanes 2-11, DNase I digestion products in the presence of 100, 50, 20, 10, 5, 2, 1, 0.5, 0.2, 0.1 nM polyamide, respectively; lane 12, DNase I digestion products in the absence of polyamide; lane 13, adenine-specific chemical sequencing. Footprinting reactions, separation of cleavage products, and data analysis were carried out as previously described.[13]


2 ImImPyPy-γ-ImHpPyPy-β-Dp
3 ImImHpPy-γ-ImPyPyPy-β-Dp
1 2 3 4 5 6 7 8 9 10 11 12 13
3'5'
ACTG
TAGC
GCCA
GT
5'3'
ACAT
GTGC
CA

2.7 (±0.9) × 10^7 M^{-1} respectively). Control experiments performed on separate DNA fragments[18,26] reveal that neither a 5'-TGGGCA-3' nor a 5'-TGGCCA-3' site is bound by polyamide **2** or **3** with a $K_a \leq 1.0 \times 10^7$ M^{-1}, indicating that the Hp/Py and Py/Hp ring pairings do not bind opposite G•C or C•G. Additional experiments with polyamide ImImHpPy-γ-ImHpPyPy-β-Dp **4**, which contains a central Hp/Hp pair, show that the polyamide binds with a $K_a \leq 1.0 \times 10^7$ M^{-1} to either 5'-TGGTCA-3' or 5'-TGGACA-3', indicating that a Hp/Hp pair is not degenerate for A•T/T•A base pairs and is in fact disfavored (Table 5.1).

Table 5.1 Equilibrium Association Constants (M^{-1})[a]

	Polyamide[b]	K_a 5'-TGGTCA-3'	K_a 5'-TGGACA-3'	K_{rel}[c]
1	Py/Py	1.3 (±0.9) × 10^{10}	6.8 (±0.4) × 10^9	2
2	Py/Hp	6.8 (±1.0) × 10^7	1.2 (±0.9) × 10^9	0.06
3	Hp/Py	2.1 (±0.7) × 10^9	2.7 (±0.9) × 10^7	77
4	Hp/Hp	≤ 1 × 10^7	≤ 1 × 10^7	--

[a]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 $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0 and 22 °C. [b]Ring pairing opposite T•A and A•T in the fourth position of six base pair binding site. [c]K_{rel} is calculated as K_a (TGGACA)/ K_a (TGGTCA).

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 **5**, **6**, **7**, and **8** confirm A•T/T•A discrimination at sequence

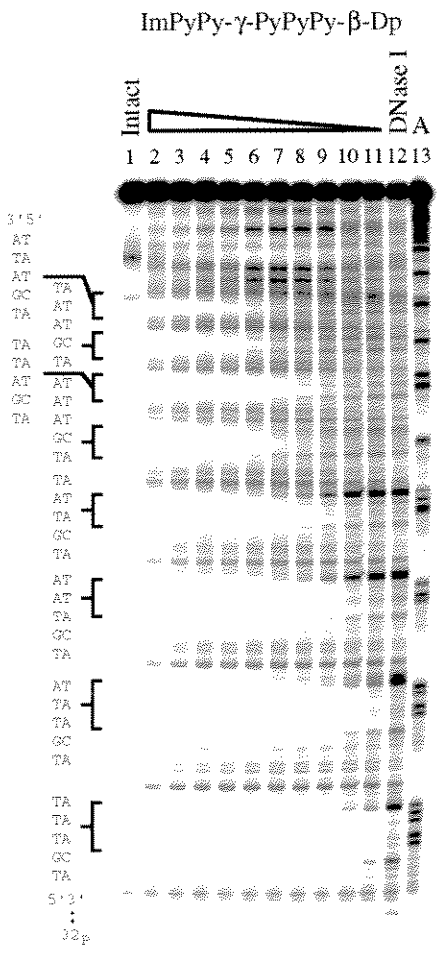
contexts which include TTT, TAA, TAT, TTA, ATT, GTT, GAT, and GTA (Figure 5.10, Tables 5.2-5.9).

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/Hp pairing as in **2**, a single substitution of C3-OH for C3-H, destabilizes interaction with 5'-TGGTCA-3' by 191-fold, a free energy difference of 3.1 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 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 kcal mol^{-1}. Interaction of **3** with 5'-TGGTCA-3' is destabilized only 6-fold relative to **1**, a free energy difference of 1.0 kcal mol^{-1}.

Implications for the design of minor groove binding polyamides

Molecular recognition in nature occurs via an ensemble of stabilizing and destabilizing forces.[16] 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.[27] For DNA-bending proteins such as TBP, practical sequence-specificity can be enhanced via the sequence-dependent energetics of DNA distortion.[28] Homeodomain proteins recognize target sequences via a combination of specific interactions with both the major and minor grooves.[29] 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.[30] 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.[32] 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.[11] Hydroxypyrrole-imidazole-pyrrole polyamides complete the minor groove recognition code using three aromatic amino acids which

Figure 5.10 DNase I footprint titrations for 6-ring hairpin polyamides ImPyPy-γ-PyPyPy-β-Dp and ImHpPy-γ-PyPyPy-β-Dp **5** on the 3' ^{32}P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II restriction fragment. Intact lane, labeled restriction fragment no polyamide or DNase I added; lanes 1-10, DNase I digestion products in the presence of 10 μM, 5 μM, 2 μM, 1 μM, 500 nM, 200 nM, 100 nM, 50 nM, 20 nM, 10 nM ImPyPy-γ-PyPyPy-β-Dp, respectively, or 1 μM, 500 nM, 200 nM, 100 nM, 50 nM, 20 nM, 10 nM, 5 nM, 2 nM, 1 nM ImHpPy-γ-PyPyPy-β-Dp **5,** respectively; DNase I lane, DNase I digestion products in the absence of polyamide; A lane, adenine-specific chemical sequencing. Footprinting reactions, separation of cleavage products, and data analysis were carried out as previously described.[14]


ImPyPy-γ-PyPyPy-β-Dp
Intact
DNase I
A
1 2 3 4 5 6 7 8 9 10 11 12 13
3'5'
5'3'
32p

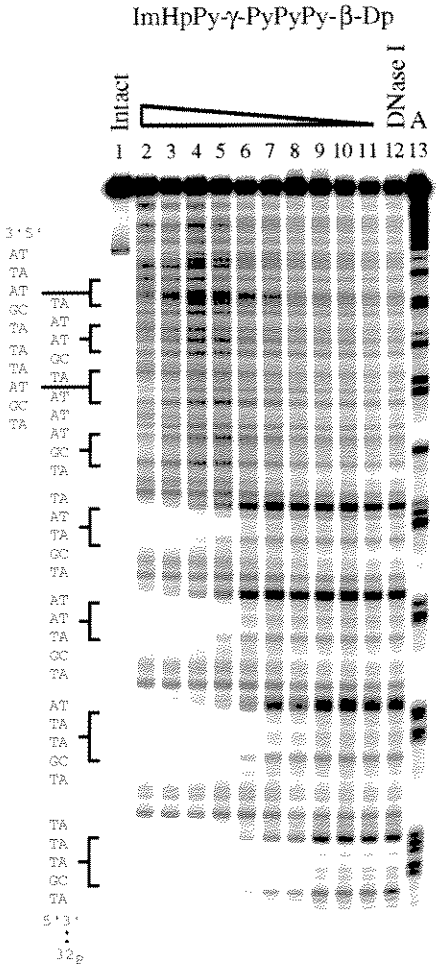
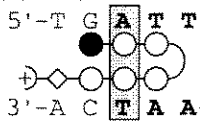
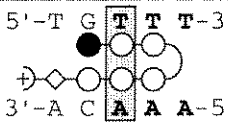
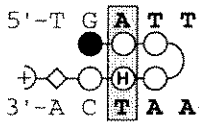
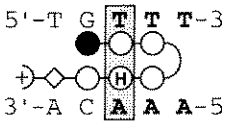
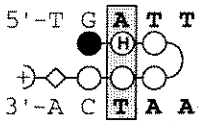
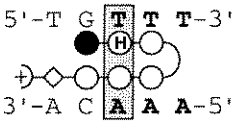
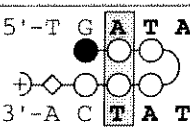
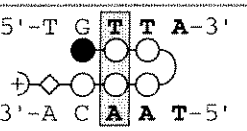
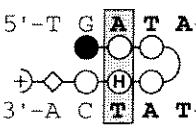
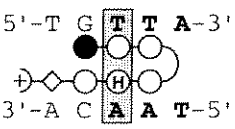
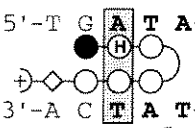
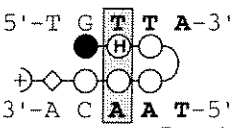

ImHpPy-γ-PyPyPy-β-Dp
Intact
DNase I
A
1 2 3 4 5 6 7 8 9 10 11 12 13
3'5'
5'3'
32p

Table 5.2 Discrimination of 5′-TGATT-3′ and 5′-TGTTT-3′*

Pair†	5′-TGATT-3′	5′-TGTTT-3′	K_{rel}‡
Py/Py	 $K_a = 3.8 \times 10^7\ M^{-1}$	 $K_a = 2.0 \times 10^8\ M^{-1}$	5.2
Py/Hp	 $K_a = 1.5 \times 10^6\ M^{-1}$	 $K_a = 8.8 \times 10^5\ M^{-1}$	0.58
Hp/Py	 $K_a = 1.3 \times 10^6\ M^{-1}$	 $K_a = 1.8 \times 10^8\ M^{-1}$	133

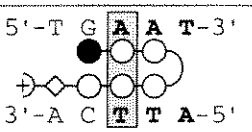
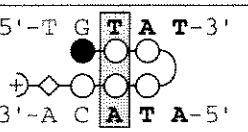
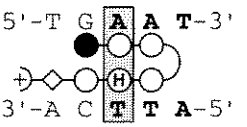
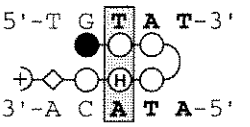
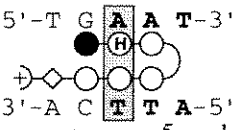
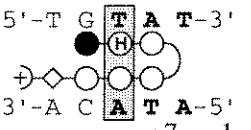
*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3′ ^{32}P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment[13]. The assays were carried out at 22 °C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.
†Ring pairing opposite T•A and A•T in the second position.
‡Calculated as K_a(5′-TGTTT-3′)/K_a(5′-TGATT-3′).

Table 5.3 Discrimination of 5′-TGATA-3′ and 5′-TGTTA-3′*

Pair†	5′-TGATA-3′	5′-TGTTA-3′	K_{rel}‡
Py/Py	 $K_a = 2.5 \times 10^7\ M^{-1}$	 $K_a = 1.4 \times 10^8\ M^{-1}$	5.7
Py/Hp	 $K_a = 5.6 \times 10^5\ M^{-1}$	 $K_a = 5.6 \times 10^5\ M^{-1}$	1.02
Hp/Py	 $K_a = 7.0 \times 10^5\ M^{-1}$	 $K_a = 5.9 \times 10^7\ M^{-1}$	84

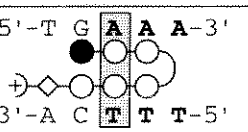
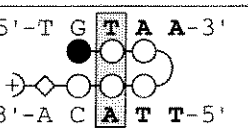
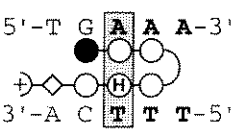
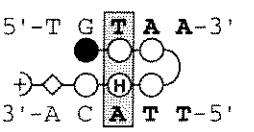
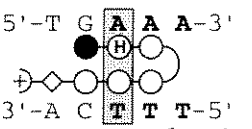
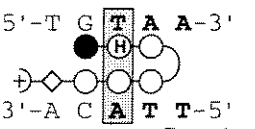
*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3′ ^{32}P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment[13]. The assays were carried out at 22 °C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.
†Ring pairing opposite T•A and A•T in the second position.
‡Calculated as K_a(5′-TGTTA-3′)/K_a(5′-TGATA-3′).

Table 5.4 Discrimination of 5′-TGAAT-3′ and 5′-TGTAT-3′*

Pair†	5′-TGAAT-3′	5′-TGTAT-3′	K_{rel}‡
Py/Py	 $K_a = 1.8 \times 10^7\ M^{-1}$	 $K_a = 4.8 \times 10^7\ M^{-1}$	2.7
Py/Hp	 $K_a = 3.8 \times 10^5\ M^{-1}$	 $K_a = 2.8 \times 10^5\ M^{-1}$	0.73
Hp/Py	 $K_a = 6.9 \times 10^5\ M^{-1}$	 $K_a = 1.7 \times 10^7\ M^{-1}$	25

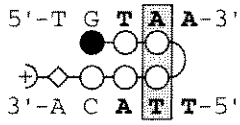
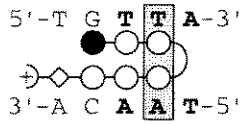
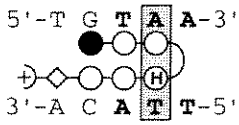
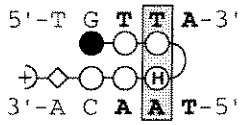
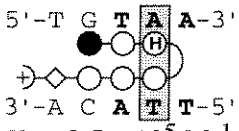
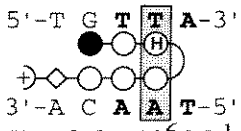
*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3′ ^{32}P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment[13]. The assays were carried out at 22 °C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.
†Ring pairing opposite T•A and A•T in the second position.
‡Calculated as K_a(5′-TGTAT-3′)/K_a(5′-TGAAT-3′).

Table 5.5 Discrimination of 5′-TGAAA-3′ and 5′-TGTAA-3′*

Pair†	5′-TGAAA-3′	5′-TGTAA-3′	K_{rel}‡
Py/Py	 $K_a = 2.2 \times 10^7\ M^{-1}$	 $K_a = 7.1 \times 10^7\ M^{-1}$	3.3
Py/Hp	 $K_a = 6.5 \times 10^5\ M^{-1}$	 $K_a = 3.7 \times 10^5\ M^{-1}$	0.57
Hp/Py	 $K_a = 1.0 \times 10^6\ M^{-1}$	 $K_a = 1.9 \times 10^7\ M^{-1}$	18

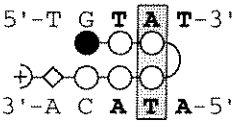
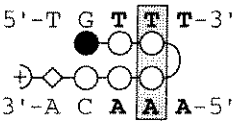
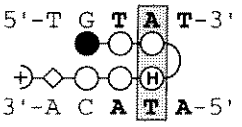
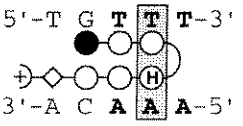
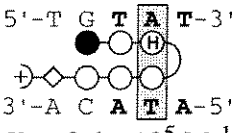
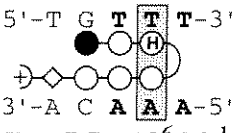
*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3′ ^{32}P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment[13]. The assays were carried out at 22 °C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.
†Ring pairing opposite T•A and A•T in the second position.
‡Calculated as K_a(5′-TGTAA-3′)/K_a(5′-TGAAA-3′).

Table 5.6 Discrimination of 5′-TGTAA-3′ and 5′-TGTTA-3′*

Pair†	5′-TGTAA-3′	5′-TGTTA-3′	K_{rel}‡
Py/Py	 $K_a = 7.1 \times 10^7\ M^{-1}$	 $K_a = 1.4 \times 10^8\ M^{-1}$	2.0
Py/Hp	 $K_a = 5.0 \times 10^6\ M^{-1}$	 $K_a = 1.8 \times 10^6\ M^{-1}$	0.36
Hp/Py	 $K_a = 2.5 \times 10^5\ M^{-1}$	 $K_a = 3.6 \times 10^6\ M^{-1}$	14

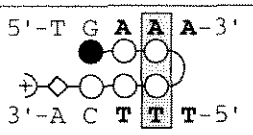
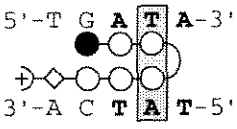
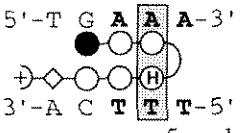
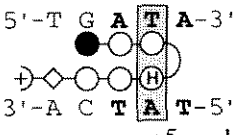
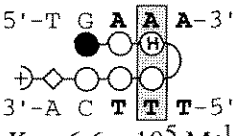
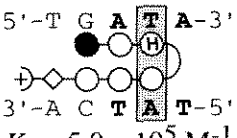
*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3′ ^{32}P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment[13]. The assays were carried out at 22 °C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.
†Ring pairing opposite T•A and A•T in the third position.
‡Calculated as K_a(5′-TGTTA-3′)/K_a(5′-TGTAA-3′).

Table 5.7 Discrimination of 5′-TGTAT-3′ and 5′-TGTTT-3′*

Pair†	5′-TGTAT-3′	5′-TGTTT-3′	K_{rel}‡
Py/Py	 $K_a = 4.8 \times 10^7\ M^{-1}$	 $K_a = 2.0 \times 10^8\ M^{-1}$	4.2
Py/Hp	 $K_a = 1.2 \times 10^6\ M^{-1}$	 $K_a = 2.4 \times 10^6\ M^{-1}$	2.0
Hp/Py	 $K_a = 3.6 \times 10^5\ M^{-1}$	 $K_a = 7.7 \times 10^6\ M^{-1}$	21

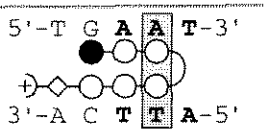
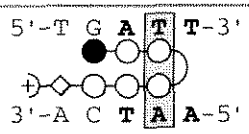
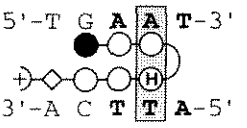
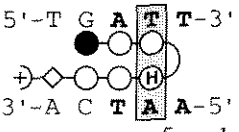
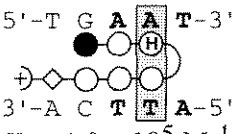
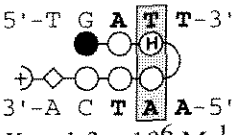
*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3′ ^{32}P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment[13]. The assays were carried out at 22 °C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.
†Ring pairing opposite T•A and A•T in the third position.
‡Calculated as K_a(5′-TGTTT-3′)/K_a(5′-TGTAT-3′).

Table 5.8 Discrimination of 5′-TGAAA-3′ and 5′-TGATA-3′*

Pair†	5′-TGAAA-3′	5′-TGATA-3′	K_{rel}‡
Py/Py	 $K_a = 2.2 \times 10^7\ M^{-1}$	 $K_a = 2.5 \times 10^7\ M^{-1}$	1.1
Py/Hp	 $K_a = 3.7 \times 10^5\ M^{-1}$	 $K_a = 1.9 \times 10^5\ M^{-1}$	0.53
Hp/Py	 $K_a = 6.6 \times 10^5\ M^{-1}$	 $K_a = 5.0 \times 10^5\ M^{-1}$	0.76

*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3′ ^{32}P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment[13]. The assays were carried out at 22 °C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.
†Ring pairing opposite T•A and A•T in the third position.
‡Calculated as K_a(5′-TGATA-3′)/K_a(5′-TGAAA-3′).

Table 5.9 Discrimination of 5′-TGAAT-3′ and 5′-TGATT-3′*

Pair†	5′-TGAAT-3′	5′-TGATT-3′	K_{rel}‡
Py/Py	 $K_a = 1.8 \times 10^7\ M^{-1}$	 $K_a = 3.8 \times 10^7\ M^{-1}$	2.2
Py/Hp	 $K_a = 2.2 \times 10^5\ M^{-1}$	 $K_a = 1.9 \times 10^5\ M^{-1}$	0.88
Hp/Py	 $K_a = 1.9 \times 10^5\ M^{-1}$	 $K_a = 1.3 \times 10^6\ M^{-1}$	6.7

*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3′ ^{32}P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment[13]. The assays were carried out at 22 °C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.
†Ring pairing opposite T•A and A•T in the third position.
‡Calculated as K_a(5′-TGATT-3′)/K_a(5′-TGAAT-3′).

combine to form four ring pairings (Im/Py, Py/Im, Hp/Py, and Py/Hp) which complement the four Watson-Crick base pairs (Table 5.10).

Table 5.10 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 (--).

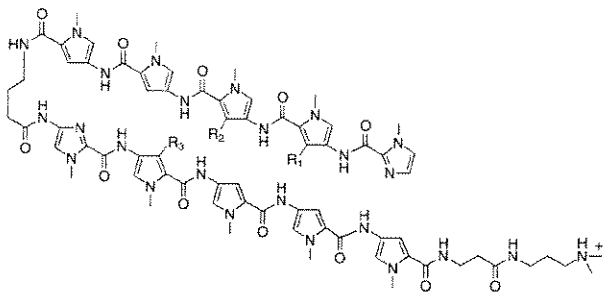
HYDROXYPYRROLE-IMIDAZOLE-PYRROLE HAIRPIN POLYAMIDES CODED FOR RECOGNITION OF A 5'-TGTTACA-3' SEQUENCE IN THE MINOR GROOVE OF DNA

Pairing rules have been developed to guide the design of synthetic polyamides for recognition of predetermined sequences in the minor groove of DNA.[9,10,14,19] We recently reported that eight ring hairpin-polyamides containing 3-hydroxypyrrole (Hp), imidazole (Im), and pyrrole (Py) amino acids form four ring pairings (Im/Py, Py/Im, Hp/Py and Py/Hp) which distinguish the four Watson-Crick base pairs in the minor groove of DNA.[33] An Im/Py pair distinguishes G•C from C•G and both of these from A•T/T•A base pairs.[9,10,14,19] A Hp/Py specifies T•A from A•T, and both of these from G•C/C•G.[33] The T•A selectivity of the Hp/Py pair likely arises from a combination of differential destabilization of polyamide binding via placement of Hp/Py opposite A•T or T•A, and specific hydrogen bonds between the 3-hydroxy and 4-carboxamido groups of Hp with the O2 of thymine (Figure 5.11).[34] A general pairing rule would require the same discrimination to be observed for the recognition of *multiple* T•A base pairs within other sequence contexts, including A•T rich sequences. It remains to be determined whether consecutive Hp/Py pairings could target binding sites varying in their A•T base pair sequence composition without compromising polyamide affinity and sequence specificity. We report here that two and three Hp/Py pairs can be combined within a hairpin template to distinguish core sequence 5'-TTA-3' from 5'-TAT-3' in the DNA minor groove.

Results and Discussion

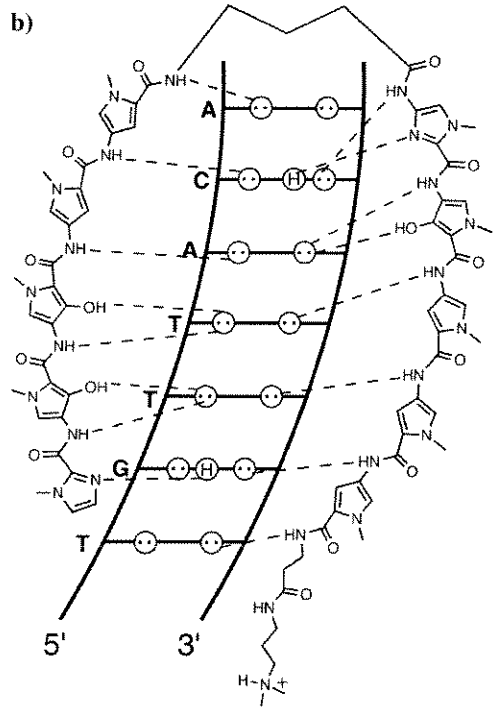
Synthesis of hydroxypyrrole-imidazole-pyrrole polyamides Polyamides ImHpHpPyPy-γ-ImHpPyPyPy-β-Dp (**14**) and ImHpPyPyPy-γ-ImHpPyPyPy-β-Dp (**15**) were synthesized by solid-phase methods using Boc-protected 3-methoxypyrrole, imidazole, and pyrrole aromatic amino acids.[25] Polyamide identity and purity were verified by ^{1}H NMR, analytical HPLC,

a)



14 $R_{1\text{-}3}$= OH, ImHpHpPyPy-γ-ImHpPyPyPy-β-Dp
15 $R_{1,3}$= OH, R_2= H ImHpPyPyPy-γ-ImHpPyPyPy-β-Dp
16 $R_{1\text{-}3}$= H, ImPyPyPyPy-γ-ImPyPyPyPy-β-Dp

b)



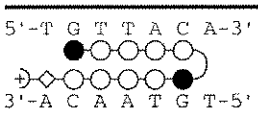
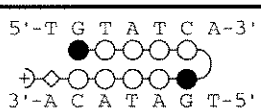
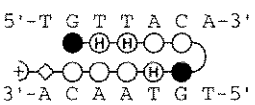
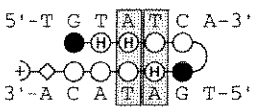
14•5'-TGTTACA-3'

Figure 5.11 **a)** Structures of the 10-ring hairpin polyamides used for this study (Im = imidazole, Py = pyrrole, Hp = 3-hydroxypyrrole, γ = γ-aminobutryic acid, β = β-alanine, Dp = dimethylaminopropylamide). b) Proposed binding model for the complex formed between the DNA and ImHpHpPyPy-γ-ImHpPyPyPy-β-Dp **14**. Circles with dots represent the lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

and matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS-monoisotopic): **14** 1514.7 (1514.7 calculated), **15** 1498.0 (1498.7 calculated).

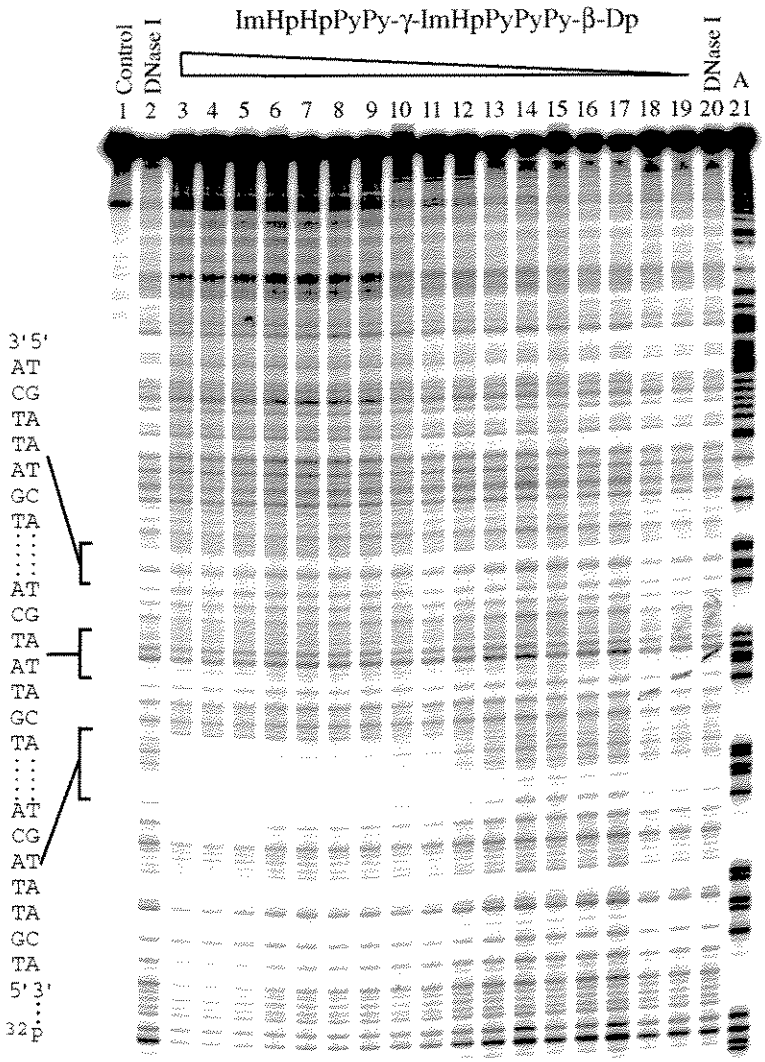
DNase I footprint titration experiments Application of the pairing rules predicts that the ten-ring hairpin polyamides ImHpHpPyPy-γ-ImHpPyPyPy-β-Dp (**14**) and ImHpPyPyPy-γ-ImHpPyPyPy-β-Dp (**15**) will bind to the 7-base pair site 5'-TGTTACA-3' but will be mismatched with the site 5'-TGTATCA-3' (Figure 5.11).[11,20,35] Quantitative DNase I footprint titration experiments[22] performed on a 361-base pair DNA fragment revealed that polyamide **14** prefers the match site 5'-TGTTACA-3' (K_a = 9.4 (±1.0) × 10^8 M^{-1}) over 5'=TGTATCA-3' ($K_a \leq 1 \times 10^7$ M^{-1}) by a factor of at least 94-fold (Figure 5.12, Table 5.11).

Table 5.11 Equilibrium Association Constants (K_a)[a,b,c]

5'-TTA-3'[d]	5'-TAT-3'	Specificity[e]
		4-fold
1.0 (±0.4) × 10^{10} M^{-1}	2.5 (±0.9) × 10^9 M^{-1}	
		≥ 94-fold
9.4 (±1.0) × 10^8 M^{-1}	≤ 1 × 10^7 M^{-1}	

[a]Reported mean values were obtained from three DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses. [b]The assays were carried out at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [c]Binding models for polyamides **1** and **2** with 5'-TGTTACA-3' and 5'-TGTATCA-3'. Filled and unfilled circles represent Im and Py rings respectively; circles containing an H represent Hp, the curved line connecting the polyamide subunits represents γ, the diamond represents β-alanine, and the + represents the positively charged dimethylaminopropylamide tail group. [d]Central three base pairs of the 7-base pair binding sites. [e]Specificity is calculated as K_a (5'-TGTTACA-3') / K_a (5'-TGTATCA-3').

Figure 5.12 Storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel used to separate the fragments generated by DNase I digestion in a quantitative footprint titration experiment with polyamide ImHpHpPyPy-γ-ImHpPyPyPy-β-Dp **14**; lane 1, intact DNA; lanes 2 and 20, DNase I digestion products in the absence of polyamide; lanes 3-19, DNase I digestion products in the presence of 100, 65, 40, 25, 15, 10, 6.5, 4.0, 2.5, 1.5, 1.0, 0.5, 0.2, 0.1, 0.05, 0.02, 0.01 nM polyamide **14**; lane 21, adenine-specific chemical sequencing reaction. Footprinting reactions, separation of cleavage products, and data analysis were carried out as previously described.[13]


ImHpHpPyPy-γ-ImHpPyPyPy-β-Dp
Control
DNase I
DNase I
A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
3'5'
AT
CG
TA
TA
AT
GC
TA
AT
CG
TA
AT
TA
GC
TA
AT
CG
AT
TA
TA
GC
TA
5'3'
32p

Control experiments on a separate fragment reveal that polyamide **14** binds to a single base mismatch site 5'-TGTTCCA-3' with a $K_a \leq 1 \times 10^7$ M^{-1}, indicating that the Hp/Py pairing is disfavored for placement opposite a G•C base pair. In comparison, polyamide **15** prefers the match site 5'-TGTTACA-3' ($K_a = 4.8 \times 10^9$ M^{-1}) over single base pair mismatch site 5'-TGTATCA-3' ($K_a = 1.5 \times 10^8$ M^{-1}) by a factor of 32-fold (Table 5.12). Remarkably, polyamide **15** exhibits much increased specificity for the match site 5'-TGTAACA-3' versus a single G•C base pair mismatch site 5'-TGGAACA-3' ($K_a \leq 1 \times 10^6$ M^{-1}), a preference of *over 2900-fold* (Table 5.13).

A symmetric pairing of Py/Py is degenerate for placement opposite A•T and T•A base pairs in the minor groove of DNA.[13,36] We find that control polyamide ImPyPyPyPy-γ-ImPyPyPyPy-β-Dp **16**, which contains Py/Py pairs rather that Hp/Py pairings, binds within a factor of four to the sites 5'-TGTTACA-3' ($K_a = 1.0$ (±0.4) $\times 10^{10}$ M^{-1}) and 5'-TGTATCA-3' ($K_a = 2.5$ (±0.9) $\times 10^9$ M^{-1}), consistent with the assignment of Py/Py as degenerate for A•T and T•A (Table 5.11). In addition, only 18-fold specificity is observed when polyamide **16** is placed opposite the site 5'-TGGAACA-3' versus 5'-TGTAACA-3' (Table 5.13).[37]

In the original report, we observed that a single Hp/Py pair replacing a Py/Py pair destablizes an eight-ring hairpin polyamide by 5-fold for an identical match site.[33] High resolution x-ray structure analysis reveals that Hp/Py polyamides bind undistorted B-form DNA; however, a localized 0.5 Å melting of the T•A Watson-Crick base pair is observed that is potentially responsible for the energetic destabilization of Hp/Py relative to the Py/Py pair.[34] Remarkably, three Hp/Py pairs destabilize polyamide **14** for 5'-TGTTACA-3' by only 11-fold relative to polyamide **16**, while two Hp/Py pairs destabilize polyamide **15** for 5'-TGTTACA-3' by 3-fold. Therefore, multiple Hp/Py pair substitutions within this ten-ring polyamide do not appear to have an additive effect on binding affinity.

Table 5.12 Equilibrium Association Constants[a,b]

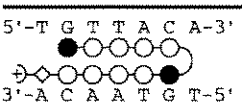
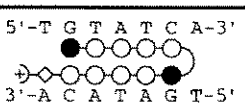
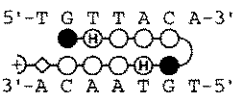
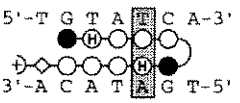
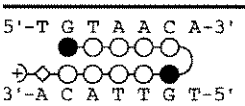
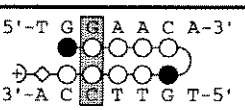
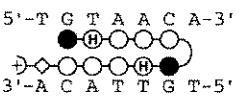
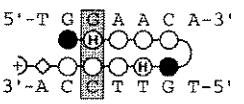
5'-TTA-3'[c]	5'-TAT-3'	Specificity[d]
		4-fold
1.2×10^{10} M^{-1}	3.0×10^{9} M^{-1}	
		32-fold
4.8×10^{9} M^{-1}	1.5×10^{8} M^{-1}	

Table 5.13 Equilibrium Association Constants[a,b]

5'-TAA-3'[c]	5'-GAA-3'	Specificity[d]
		18-fold
1.2×10^{10} M^{-1}	6.8×10^{8} M^{-1}	
		≥2900-fold
2.9×10^{9} M^{-1}	$\leq 1 \times 10^{6}$ M^{-1}	

[a]The assays were carried out at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [b]Binding models for polyamides **1** and **2** with 5'-TGTTACA-3' and 5'-TGTATCA-3'. Filled and unfilled circles represent Im and Py rings respectively; circles containing an H represent Hp, the curved line connecting the polyamide subunits represents γ, the diamond represents β-alanine, and the + represents the positively charged dimethylaminopropylamide tail group. [c]Central three base pairs of the 7-base pair binding sites. [d]Specificity is calculated as K_a (5'-TGTTACA-3') / K_a (5'-TGTATCA-3').

Hydroxypyrrole-imidazole-pyrrole polyamides complete the minor groove recognition code using the four ring pairings (Im/Py, Py/Im, Hp/Py, and Py/Hp) to complement the four Watson-Crick base pairs. The observation of substantially increased A•T/T•A specificity without severe energetic penalty by incorporation of multiple Hp/Py pairings is a minimal first step toward full integration of multiple Hp/Py ring pairings for Py/Py pairs. Because the sequence dependent microstructure of DNA is still not well understood, the scope and limitations of the 3-ring pairing rules with regards to energetics and specificity remains to be elucidated and will be reported in due course.

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.[15] 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).[15] 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.[16] The principle that 'the major groove is a better candidate for sequence-specific recognition than the minor groove'[39] 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[32], 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.[9-11,14,33] The side-by-side pairing of the aromatic residues in the polyamide dimer determines the DNA sequence recognized. An Im/Py pair distinguishes G•C from C•G and both of these from A•T/T•A[9-11,14], and the structural basis of this discrimination has been determined through x-ray structure studies.[19,40] 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.[33] To address this question, the co-crystal structure of a polyamide of sequence ImHpPyPy-β-Dp **17** (Figure 5.13) bound as a dimer to a self-complementary ten base pair oligonucleotide containing all four Watson-Crick base pairs, 5'-CC**AGTACT**GG-3' (binding site in bold) has been determined (Figure 5.14). The structure

of the polyamide ImPyPyPy-β-Dp **18**, containing Py/Py pairs which do not distinguish T•A from A•T[13,36], bound to the same duplex was solved for comparison. In addition, the equilibrium association constants (K_a) for polyamides ImHpPyPy-β-Dp **17** and ImPyPyPy-β-Dp **18** (Figure 5.13) were determined using DNase I quantitative footprint titration experiments in order to correlate energetic analysis of minor groove recognition with three-dimensional structural information (Figure 5.15, Tables 5.14, 5.15).

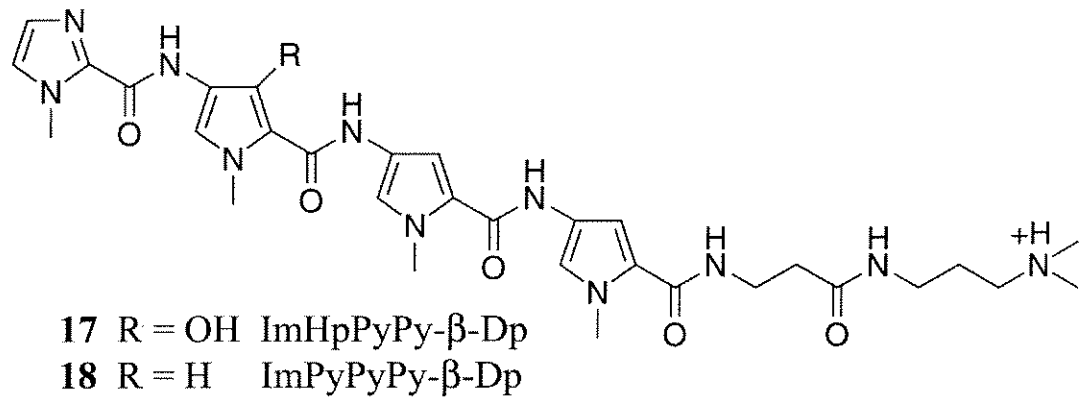


Figure 5.13 Chemical structures of the polyamides ImHpPyPy-β-Dp 17 and ImPyPyPy-β-Dp 18.

Results and Discussion

Synthesis of hydroxypyrrole-imidazole-pyrrole polyamides The polyamides ImHpPyPy-β-Dp **17** and ImPyPyPy-β-Dp **18** were synthesized by solid-phase methods using Boc-protected 3-methoxypyrrole, imidazole, and pyrrole amino acids.[25] The identity and purity of the polyamides were confirmed by ^{1}H 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.

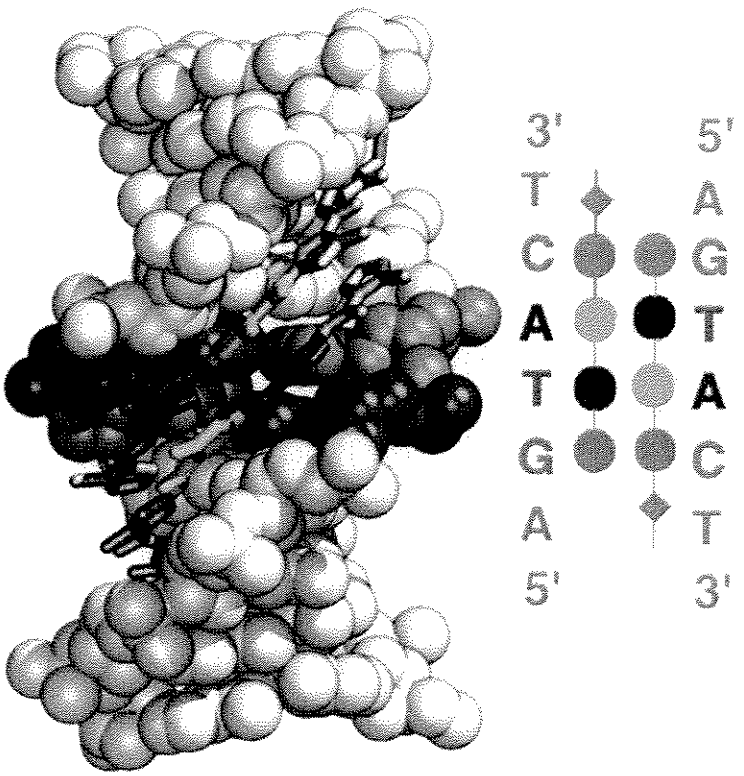


Figure 5.14 Structure of (ImHpPyPy-β-Dp)$_2$•5'-CC**AGTACT**GG-3'. Adenosine is very dark gray and thymidine is medium gray. A schematic DNA-binding model is shown to the side, with the residues of the polyamide indicated by filled circles. Im/Py pairs = medium gray, Hp = black, Py = light gray. The overall structure of (ImPyPyPy-β-Dp)$_2$•5'-CC**AGTACT**GG-3' is similar.

Analysis of energetics by quantitative DNase I footprint titrations DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0, 22 °C) were performed to determine the equilibrium association constants K_a for recognition of bound sites (Figure 5.15).[22] The hydroxypyrrole-containing polyamide, ImHpPyPy-β-Dp **17**, prefers the match site 5'-AGTACT-3' (K_a = 2.9 (±0.7) × 10^6 M^{-1}) over the single base pair mismatch site 5'-AGTATT-3' (K_a = 7.6 (±0.6) × 10^5 M^{-1}) by a factor of 4 (Table 5.14). Polyamide **17** has increased A•T/T•A specificity, preferring the match site 5'-AGTACT-3' (K_a = 2.9 (±0.7) × 10^6 M^{-1}) over the mismatch sites 5'-AGAACT-3' (K_a = 3.0 × 10^5 M^{-1}) and 5'-AGATCT-3' (K_a = 1.0 × 10^5 M^{-1}) by a factor of 10- and 29-fold respectively (Table 5.15). The control polyamide ImPyPyPy-β-Dp **18**, 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'-AGATCT-3' within a factor of six (Table 5.15). In addition, ImPyPyPy-β-Dp **18** prefers the site 5'-AGTACT-3' (K_a = 2.1 (±0.1) × 10^7 M^{-1}) over the site 5'-AGTATT-3' (K_a = 1.4 (±0.3) × 10^6 M^{-1}) by at least 15-fold (Table 5.14). 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 **18** by two Hp/Py pairs to give polyamide **17**. 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.

Figure 5.15 DNase I footprint titration experiment for polyamide ImHpPyPy-β-Dp 17 on the 3' ^{32}P end-labeled *Fsp*I/*Afl*II restriction fragment of pJT8.[11] Lane 1, adenine-specific chemical sequencing reaction; lane 2, DNase I digestion products in the absence of polyamide; lanes 3-19, DNase I digestion products in the presence of 100 μM, 50 μM, 20 μM, 10 μM, 6.5 μM, 4.0 μM, 2.5 μM, 1.5 μM, 1.0 μM, 650 nM, 400 nM, 250 nM, 150 nM, 100 nM, 50 nM, 20 nM, 10 nM polyamide **17**; lane 20, DNase I digestion in absence of polyamide; lane 21, intact DNA. Footprinting reactions, separation of cleavage products, and data analysis were carried out as described.[14]

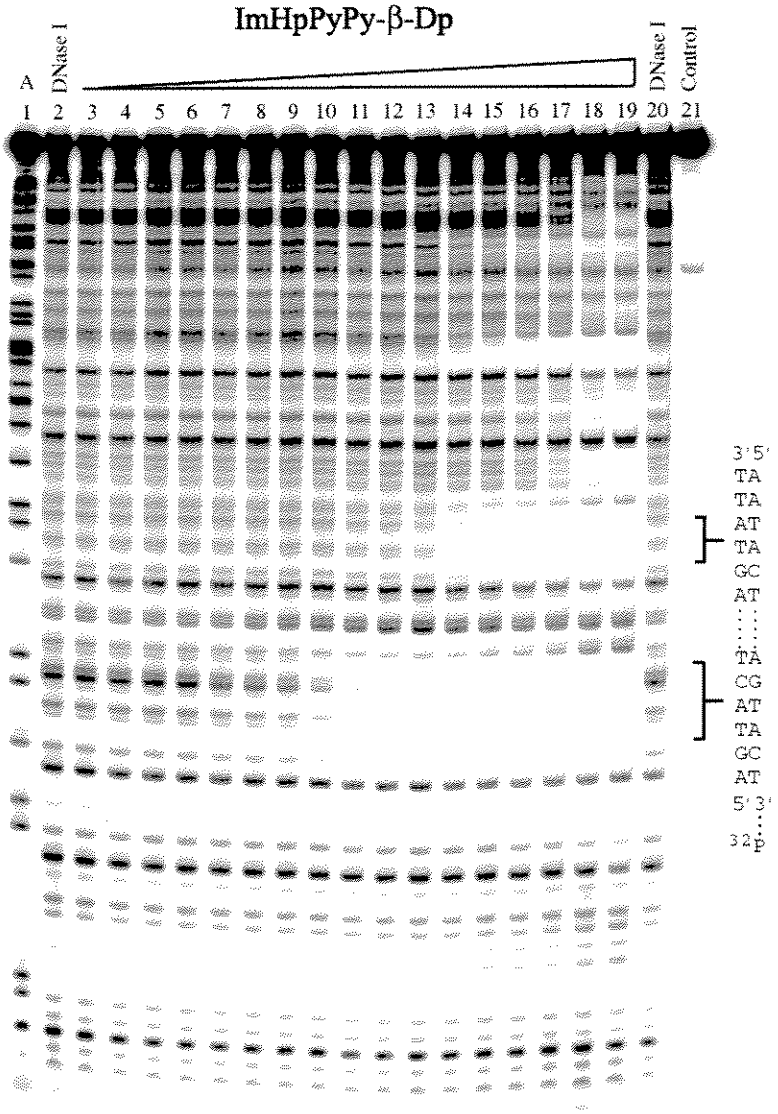
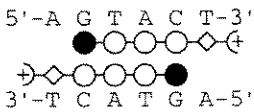
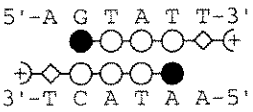
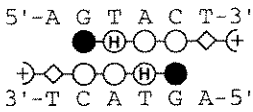
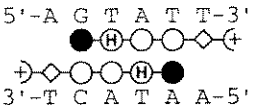
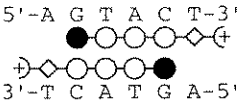
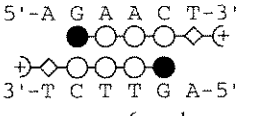
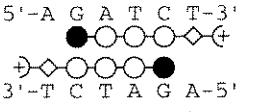
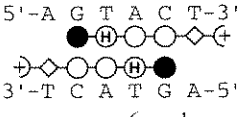
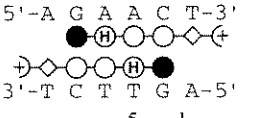
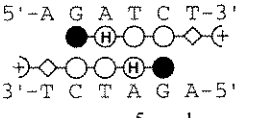

ImHpPyPy-β-Dp
A
DNase I
DNase I
Control
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
3'5'
TA
TA
AT
TA
GC
AT
TA
CG
AT
TA
GC
AT
5'3'
32P

Table 5.14 Equilibrium Association Constants (K_a)[a,b]

Match	Mismatch	Specificity[c]
 $2.1\ (\pm 0.1) \times 10^7\ M^{-1}$	 $1.4\ (\pm 0.3) \times 10^6\ M^{-1}$	15-fold
 $2.9\ (\pm 0.7) \times 10^6\ M^{-1}$	 $7.6\ (\pm 0.6) \times 10^5\ M^{-1}$	4-fold

[a]Values reported are the mean values obtained from three DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses.
[b]The assays were carried out at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.
[c]Specificity is calculated as K_a (5'-AGTACT-3') / K_a (5'-AGTATT-3').

Table 5.15 Equilibrium Association Constants (K_a)[a,b]

GTAC	GAAC	GATC	Specificity
 $2.5 \times 10^7\ M^{-1}$	 $4.0 \times 10^6\ M^{-1}$	 $4.0 \times 10^6\ M^{-1}$	6-fold
 $4.0 \times 10^6\ M^{-1}$	 $3.0 \times 10^5\ M^{-1}$	 $1.0 \times 10^5\ M^{-1}$	13-40-fold

[a]The assays were carried out at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.
[b]Specificity is calculated as K_a (5'-AGTACT-3') / K_a (5'-AGAACT-3') and K_a (5'-AGTACT-3') / K_a (5'-AGATCT-3').

Three-dimensional structural analysis of polyamide-DNA recognition 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 5.14). 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 chemical[41] and structural studies of polyamide dimers.[34,36c,40,42]

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 5.16). 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.

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 additional carbon results in an asymmetric cleft in the minor groove of a T•A base pair (Figure 5.1).[21,33] 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.

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 average (Figure 5.17). 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 5.14). For example, ImHpPyPy-β-Dp and ImPyPyPy-β-Dp bind a 5'-AGTACT-3' site with equilibrium association constants of 2.9 (±0.7) × 10^6 M^{-1} and 2.1 (±0.1) × 10^7 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.[43]

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. 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

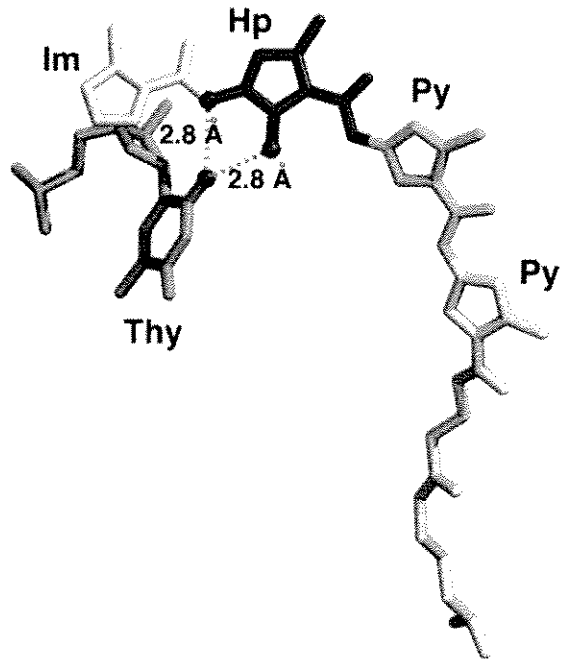


Figure 5.16 Diagram of the hydrogen bonds between the hydroxyl and the preceding amide and the two lone pairs of thymine-O2.

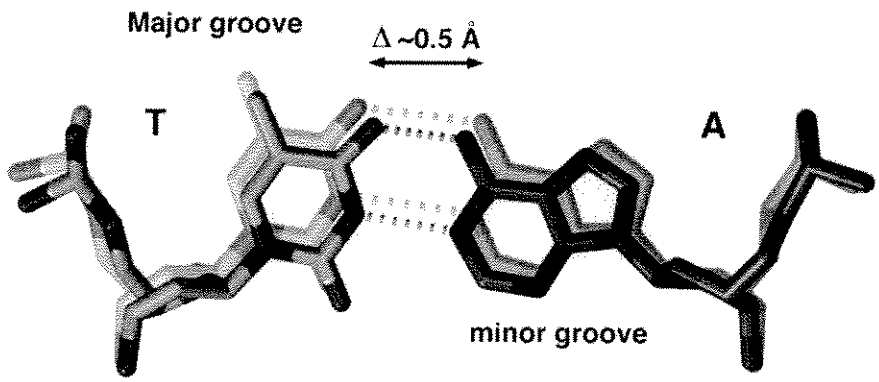


Figure 5.17 T•A base pair of the ImPyPyPy structure superimposed on the corresponding base pair from the ImHpPyPy structure, showing the slight elongation of the hydrogen bonds of the target T•A base pair.

weaker. This may be an additional source of the slightly decreased affinity of the Hp-containing polyamides relative to the pyrrole counterparts.

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.[21,33] The structure eliminates the possibilities that a bulky substitution at the pyrrole 3-position might 1) sterically clash with the thymine-O2,[42] or 2) cause a gross distortion of the DNA duplex.[44] Whereas a predictable relationship between amino acids of DNA binding proteins and nucleotide sequence targeted has proven elusive, simple structural principles have allowed polyamides to provide the first amino acid-base pair code for recognition of all four base pairs in the minor groove.

Experimental Section

Materials Dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBt), 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophophate (HBTU) and Boc-β-alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-β-Pam-Resin) was purchased from Peptides International (0.2 mmol/gram), NovaBiochem (0.6 mmol/gram), or Peninsula (0.6 mmol/gram). N,N-diisopropylethylamine (DIEA), N,N-dimethylformamide (DMF), N-methylpyrrolidone (NMP), DMSO/NMP, acetic anhydride (Ac2O), and 0.0002 M potassium cyanide/pyridine were purchased from Applied Biosystems. Dichloromethane (DCM) and triethylamine (TEA) were reagent grade from EM, thiophenol (PhSH), dimethylaminopropylamine (Dp), and sodium hydride were from Aldrich, trifluroacetic acid (TFA) Biograde from Halocarbon, phenol from Fisher, and ninhydrin from Pierce. All reagents were used without further purification.

Sonicated, deproteinized calf thymus DNA, from Pharmacia, was dissolved in filter sterilized water to a final concentration of 1 mM in base pairs and stored at 4 °C. Glycogen was purchased from Boehringer-Manheim as a 20 mg/mL aqueous solution. Nucleotide triphosphates were purchased from Pharmacia and used as supplied. Nucleoside triphosphates labeled with ^{32}P ($\geq$ 3000 C_i/mmol) were obtained from Dupont-New England Nuclear. All enzymes were purchased from Boehringer-Manheim and were used according to the supplier's recommended protocol in the activity buffer provided. Plasmid pUC19 was obtained from New England Biolabs. A solution of 0.5 M EDTA, pH 8.0 was purchased from Ultrapure. Phosphoramidites were from Glen Research. General manipulation of duplex DNA and oligonucleotides were performed according to established procedures.

Ethyl 4-[benzyloxycarbonyl)amino]-3-hydroxy-1-methylpyrrole-2-carboxylate 10 Ethyl-4-carboxy-3-hydroxy-1-methylpyrrole-2-carboxylate **9** (60 g, 281.7 mmol) was dissolved in

282 mL acetonitrile. TEA (28.53 g, 282 mmol) was added, followed by diphenylphosphorylazide (77.61 g, 282 mmol). The mixture was refluxed for 5 hours, followed by addition of benzyl alcohol (270 mL) and reflux continued overnight. The solution was cooled and volatiles removed *in vacuo*. The residue was absorbed onto silica and chromatographed, 4:1 hexanes:ethyl acetate, to give a white solid (21.58 g, 24 %). ^{1}H NMR (DMSO-d_6) 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 $C_{16}H_{18}N_2O_5$).

Ethyl 4-[(*tert*-butoxycarbonyl)amino]-3-methoxy-1-methylpyrrole-2-carboxylate 12

Ethyl 4-[(benzyloxycarbonyl)amino]-3-hydroxy-1-methylpyrrole-2-carboxylate **10** (13.4 g, 42.3 mmol) was dissolved in 110 mL acetone. Anhydrous K_2CO_3 (11.67 g, 84.5 mmol) was added, followed by methyliodide (5.96 g, 42.3 mmol) and dimethylaminopyridine (0.5 g, 4.23 mmol) and the mixture stirred overnight. The solid K_2CO_3 was removed by filtration and 200 mL water added. Volatiles were removed *in vacuo* and the solution made acidic with addition of 1*N* H_2SO_4. The aqueous layer was extracted with diethyl ether. Organic layers were combined, washed with 10% H_2SO_4, dried over $MgSO_4$, and dried to give a white solid **11**. The solid was used without further purification and dissolved in 38 mL DMF. DIEA (11 mL), Boc anhydride (9.23 g, 42.3 mmol), and 10% Pd/C (500 mg) were added and the solution stirred under hydrogen (1 atm) for 2.1 h. The slurry was filtered through celite which was washed with methanol. Water (250 mL) was added and volatiles removed *in vacuo*. The aqueous layer was extracted with ether. Organic layers were combined, washed with water and brine, and dried over $MgSO_4$. Solvent was removed *in vacuo* to give a white solid **12**(8.94 g, 71%). ^{1}H 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 Hz); MS *m/e* 299.161 (M+H 299.153 calcd. for $C_{14}H_{22}N_2O_5$).

Ethyl 4-[(*tert*-butoxycarbonyl)amino]-3-hydroxy-1-methylpyrrole-2-carboxylate 13 Ethyl 4-[(*tert*-butoxycarbonyl)amino]-3-methoxy-1-methylpyrrole-2-carboxylate **12** (9.0 g, 30.2 mmol) was dissolved in 30 mL ethanol. NaOH (30 mL, 1 M, aq.) was added and the solution stirred for 4 days. Water (200 mL) was added and ethanol removed *in vacuo*. The solution was extracted with diethyl ether, aqueous layer acidified to pH = 2-3, and extracted again with diethyl ether. Organic layers were dried over $MgSO_4$, and solvent *removed in vacuo* to give a white solid **13** (6.0 g, 20.5 mmol, 87% based on recovered SM). ^{1}H 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 $C_{12}H_{18}N_2O_5$).

ImImOpPy-γ-ImPyPyPy-β-Dp 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).[25] 3-hydroxypyrrole-Boc amino 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 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 appropiate fractions as a white powder (97 mg, 49% recovery). UV (H_2O) λmax 246, 316 (66,000); ^{1}H 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 In order to remove the methoxy protecting group, a sample of ImImOpPy-γ-ImPyPyPy-β-Dp (5 mg, 3.9 μmol) was treated with sodium thiophenoxide at 100 °C for 2 h. DMF (1000 μL) and thiophenol (500 μL) 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, ImImOpPy-γ-ImPyPyPy-β-Dp (5 mg) dissolved in DMF (500 mL) 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. 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) λmax 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 $C_{57}H_{71}N_{22}O_{11}$).

ImImPyPy-γ-ImOpPyPy-β-Dp 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 for ImImOpPy-γ-ImPyPyPy-β-Dp.[25] 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).

ImImPyPy-γ-ImHpPyPy-β-Dp 2 A sample of ImImPyPy-γ-ImOpPyPy-β-Dp (5 mg, 3.9 μmol) was treated with sodium thiophenoxide and purified by reversed phase HPLC as described for ImImHpPy-γ-ImPyPyPy-β-Dp. ImImPyPy-β-ImHpPyPy-β-Dp is recovered upon lyophilization of the appropriate fractions as a white powder (3.2 mg, 66% recovery). UV (H_2O) λmax 246, 312 (66,000); MALDI-TOF-MS (monoisotopic), 1239.6 (1239.6 calcd. for $C_{57}H_{71}N_{22}O_{11}$).

ImPyPy-γ-OpPyPy-β-Dp ImPyPy-γ-OpPyPy-β-Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc-β-Pam-Resin (0.66 mmol/g).[25] 3-hydroxypyrrole-Boc amino 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 ImPyPy-γ-OpPyPy-β-Pam-Resin (400 mg, 0.45 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. ImPyPy-γ-OpPyPy-β-Dp is

recovered upon lyophilization of the appropiate fractions as a white powder (45 mg, 25% recovery). UV (H_2O) λmax 246, 310 (50,000); ^{1}H NMR (DMSO-d_6) δ 10.45 (s, 1 H), 9.90 (s, 1 H), 9.82 (s, 1 H), 9.5 (br s, 1 H), 9.38 (s, 1 H), 9.04 (s, 1 H), 8.02 (m, 3 H), 7.37 (s, 1 H), 7.25 (m, 2 H), 7.15 (d, 1 H, J = 1.6 Hz), 7.11 (m, 2 H), 7.09 (d, 1 H), 7.03 (d, 1 H), 6.99 (d, 1 H), 6.87 (d, 1 H), 6.84 (d, 1 H), 3.96 (s, 3 H), 3.81 (s, 6 H), 3.77 (s, 6 H), 3.76 (s, 3 H), 3.74 (s, 1 H), 3.34 (q, 2 H, J = 5.6 Hz), 3.20 (q, 2 H, J = 5.8 Hz), 3.09 (q, 2 H, J = 6.1 Hz), 2.97 (quintet, 2 H, J = 5.3 Hz), 2.70 (d, 6 H, J = 3.9 Hz), 2.34 (m, 4 H), 1.73 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1007.6 (1007.5 calcd. for $C_{48}H_{63}N_{16}O_9$).

ImPyPy-γ-HpPyPy-β-Dp 8 In order to remove the methoxy protecting group, a sample of ImPyPy-γ-OpPyPy-β-Dp (5 mg, 4.8 mmol) was treated with sodium thiophenoxide 100°C for 2 h and reverse phase HPLC purified using the same protocol as for ImImHpPy-γ-ImPyPyPy-β-Dp. ImPyPy-γ-HpPyPy-β-Dp is recovered upon lyophilization of the appropriate fractions as a white powder (2.5 mg, 52% recovery). UV (H_2O) λmax 246, 310 (50,000); ^{1}H NMR (DMSO-d_6) δ 10.44 (s, 1 H), 10.16 (s, 1 H), 9.90 (s, 1 H), 9.77 (s, 1 H), 9.5 (br s, 1 H), 9.00 (s, 1 H), 8.03 (m, 3 H), 7.37 (s, 1 H), 7.26 (m, 2 H), 7.14 (d, 1 H, J = 1.7 Hz), 7.12 (m, 2 H), 7.02 (d, 1 H), 6.93 (d, 1 H), 6.88 (d, 1 H), 6.82 (d, 1 H), 6.72 (d, 1 H), 3.96 (s, 3 H), 3.81 (s, 6 H), 3.77 (s, 3 H), 3.76 (s, 3 H), 3.74 (s, 1 H), 3.36 (q, 2 H, J = Hz), 3.22 (q, 2 H, J = 5.9 Hz), 3.09 (q, 2 H, J = 5.5 Hz), 2.98 (quintet, 2 H, J = 5.3 Hz), 2.70 (d, 6 H, J = 4.3 Hz), 2.34 (m, 4 H), 1.78 (m, 4 H); MALDI-TOF-MS (monoisotopic), 994.2 (993.5 calcd. for $C_{47}H_{61}N_{16}O_9$).

ImImOpPy-γ-ImPyPyPy-β-Dp-NH_2 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 previously for hydroxypyrrole-imidazole-pyrrole polyamides.[25] A sample of ImImOpPy-γ-ImPyPyPy-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. ImImOpPy-γ-ImPyPyPy-β-Dp-NH_2 is recovered upon lyophilization of the appropriate fractions as a white powder (93 mg, 46% recovery). UV (H_2O) λmax 246, 316 (66,000); ^{1}H NMR (DMSO-d_6) 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 $C_{60}H_{78}N_{23}O_{11}$).

ImImOpPy-γ-ImPyPyPy-β-Dp-EDTA 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-NH_2 (13 mg, 10 μmol) dissolved in DMSO (750 μL). 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 $C_{70}H_{92}N_{25}O_{18}$).

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 μmol) 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 (H_2O) λmax 246, 312 (66,000); MALDI-TOF-MS (monoisotopic), 1556.6 (1556.7 calcd. for $C_{69}H_{90}N_{25}O_{18}$).

ImImPyPy-γ-ImOpPyPy-β-Dp-NH$_2$ 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)[25], 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, ImImPyPy-γ-ImOpPyPy-β-Dp-NH_2 is recovered upon lyophilization of the appropriate fractions as a white powder (104 mg, 54 % recovery). UV (H_2O) λmax 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 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-NH_2 (13 mg, 10 μmol) dissolved in DMSO (750 μL). 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 $C_{70}H_{92}N_{25}O_{18}$).

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 μmol) 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 (H_2O) λmax 246, 312 (66,000); MALDI-TOF-MS (monoisotopic), 1555.9 (1556.7 calcd. for $C_{69}H_{90}N_{25}O_{18}$).

Table 5.16 Mass spectral characterization of Op and Hp polyamides, synthesized and purified as described for ImImOpPy-γ-ImPyPyPy-β-Dp and ImImHpPy-γ-ImPyPyPy-β-Dp.

POLYAMIDE	FORMULA	(M+H)CALCD	FOUND
ImOpPy-γ-PyPyPy-β-Dp	$C_{48}H_{63}N_{16}O_9$	1007.5	1007.5
ImHpPy-γ-PyPyPy-β-Dp	$C_{47}H_{61}N_{16}O_9$	993.5	993.2
ImPyOp-γ-PyPyPy-β-Dp	$C_{48}H_{63}N_{16}O_9$	1007.5	1007.5
ImPyHp-γ-PyPyPy-β-Dp	$C_{47}H_{61}N_{16}O_9$	993.5	993.4
ImPyPy-γ-OpPyPy-β-Dp	$C_{48}H_{63}N_{16}O_9$	1007.5	1007.6
ImPyPy-γ-HpPyPy-β-Dp	$C_{47}H_{61}N_{16}O_9$	993.5	993.2
ImPyPy-γ-PyOpPy-β-Dp	$C_{48}H_{63}N_{16}O_9$	1007.5	1007.5
ImPyPy-γ-PyHpPy-β-Dp	$C_{47}H_{61}N_{16}O_9$	993.5	993.4
ImImOpPy-γ-ImPyPyPy-β-Dp	$C_{58}H_{72}N_{22}O_{11}$	1253.6	1253.5
ImImHpPy-γ-ImPyPyPy-β-Dp	$C_{57}H_{71}N_{22}O_{11}$	1239.6	1239.6
ImImPyPy-γ-ImOpPyPy-β-Dp	$C_{58}H_{72}N_{22}O_{11}$	1253.6	1253.6
ImImPyPy-γ-ImHpPyPy-β-Dp	$C_{57}H_{71}N_{22}O_{11}$	1239.6	1239.6
ImImOpPy-γ-ImOpPyPy-β-Dp	$C_{59}H_{75}N_{22}O_{12}$	1283.6	1283.6
ImImHpPy-γ-ImHpPyPy-β-Dp	$C_{57}H_{71}N_{22}O_{12}$	1255.6	1255.5
ImOpPyPyPy-γ-ImOpPyPyPy-β-Dp	$C_{72}H_{88}N_{25}O_{14}$	1526.7	1526.6
ImHpPyPyPy-γ-ImHpPyPyPy-β-Dp	$C_{70}H_{84}N_{25}O_{14}$	1498.7	1498.0
ImHpHpPyPy-γ-ImHpPyPyPy-β-Dp	$C_{71}H_{87}N_{26}O_{14}$	1514.7	1514.7

Preparation of ^{32}P-labeled DNA Plasmids were prepared by hydridizing a complementary set of synthetic oligonucleotides and ligating the resulting duplex to the large pUC19 *Xba* I/ *Pst* I restriction fragment. The oligonucleotides for plasmid pDEH11 5'-d(CTAGGCGAGCGCTCAT**TGTTACA**GCGAGCGCTCAT**TGTATCA**GCGAGCG CTCATTGATTCAGCGAGCGCTCATGCA)-3' and 5'-d(TGAGCGCTCGCTGAATC AATGAGCGCTCGCTGATACAATGAGCGCTCGCTGTAACAATGAGCGCTCGC)-3'. The oligonucleotides for pDEH10 are 5'-d(CTAGTGGATGCTGGTT**AGTACT**TGGA

TGCTGGTT**AGAACT**TGGATGCTGGTT**AGATCT**TGGATGCTGGTTGCA)-3' and 5'-d(ACCAGCATCCAAGATCTAACCAGCATCCAAGTTCTAACCAGCATCCAAGTACT AACCAGCATCCA)-3'. The designed binding sites are shown in bold in the above DNA oligonucleotides. The 3'-^{32}P end-labeled *Eco* RI/ *Pvu* II fragments were prepared by digesting the plasmid with *Eco* RI and simultaneously filling in using Sequenase (v. 2.0, USB), [α-^{32}P]-deoxyadenosine-5'-triphosphate, and [α-^{32}P]-thymidine-5'-triphosphate, digesting with *Pvu* II, and isolating the fragments containing the binding sites by nondenaturing gel electrophoresis. A and G sequencing were carried out as described.[38,45] Standard methods were used for all DNA manipulations.[46]

Affinity cleavage footprint titrations All reactions were executed in a total volume of 40 µL. A stock solution of polyamide or H_2O 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 µM/bp calf thymus DNA, and pH 7.0. Solutions were incubated for a minimum of 4 hours at 22 °C. Subsequently, 4 µL of freshly prepared 100 µM $Fe(NH_4)_2(SO_4)_2$ was added and the solution allowed to equilibrate for 20 min. Cleavage reactions were initiated by the addition of 4 µL of 100 mM dithiothreitol, allowed to proceed for 30 min at 22 °C, then stopped by the addition of 10 µL of a solution containing 1.5 M NaOAc (pH 5.5), 0.28 mg/mL glycogen, and 14 µM 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.

Quantative DNase I footprint titrations All reactions were executed in a total volume of 400 μL. A polyamide stock solution or H_2O (for reference lanes) was added to an assay buffer containing 3'-^{32}P radiolabeled restriction fragment (20,000 cpm), affording final solution conditions of 10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, and either (i) a suitable concentration range of polyamide or (ii) no polyamide (for reference lanes). The solutions were allowed to equilibrate for 4 hours at 22°C. Footprinting reactions were initiated by the addition of 10 μL of a stock solution of DNase I (at the appropriate concentration to give ~55% intact DNA) containing 1 mM dithiothreitol and allowed to proceed for 7 minutes at 22 °C. The reactions were stopped by the addition of 50 μL of a solution containing 2.25 M NaCl, 150 mM EDTA, 23 μM base pair calf thymus DNA, and 0.6 mg/mL glycogen, and ethanol precipitated. The reactions were resuspended in 1 x TBE/ 80% formamide loading buffer, denatured by heating at 85 °C for 15 minutes, and placed on ice. The reactions products were separated by electrophoresis on an 8% polyacrylamide gel (5% crosslinking, 7 M urea) in 1 x TBE at 2000 V for 1.5 h. Gels were dried on a slab dryer and then exposed to a storage phosphor screen at 22 °C.

Phosphostimuable storage phosphor imaging plates (Kodak Storage Phosphor Screen SO320 obtained from Molecular Dynamics) were pressed flat against dried gel samples and exposed in the dark at 22 °C for 12-24 hours. A Molecular Dynamics 400S PhosphorImager was used to obtain all data from the storage screens. The data were analyzed by performing volume integration of the target site and reference blocks using ImageQuant v. 1.0 for Macintosh.

Quantitative DNase I footprint titration data analysis was performed by taking a 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:

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

where I_{site}° and I_{ref}° are the site and reference intensities, respectively, from a DNase I control lane to which no polyamide was added.

The ($[L]_{tot}$, θ_{app}) data 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:

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

where $[L]_{tot}$ 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 of KaleidaGraph software (v. 3.0.1, Abelbeck Software) with K_a, θ_{max}, and θ_{min} as the adjustable parameters. The goodness of fit of the binding curve to the data points is evaluated by the correlation coefficient, with $R > 0.97$ as the criterion for an acceptable fit. Three sets of acceptable data were used in determining each association constant. All lanes from a gel were used unless a visual inspection revealed a data point to be obviously flawed relative to neighboring points. The data were normalized using the following equation:

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

At higher concentrations of polyamide, the reference site becomes partially protected due to nonspecific DNA binding, resulting in low θ_{app} values. For this reason, higher concentrations were not used.

References

1. Zimmer, C.; Wähnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31-112.

2. Dervan, P. B. *Science* **1986**, *232*, 464-471.

3. Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645-650.

4. Thuong, N. T.; Helene, C. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 666-690.

5. Nielsen, P. E. *Chem. Eur. J.* **1997**, *3*, 505-508.

6. Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202-205.

7. Maher, J. L.; Dervan, P. B.; Wold, B. *Biochemistry* **1992**, *31*, 70-81.

8. Duvalentin, G.; Thuong, N.T.; Helene, C. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 504-508.

9. Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794.

10. Mrksich, M. *et al. Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586-7590.

11. Trauger, J.W.; Baird, E. E.; Dervan, P.B. *Nature* **1996**, *382*, 559-561.

12. Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723-5727.

13. White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532.

14. White, S.; Baird, E.E.; Dervan, P. B. *Chem. & Biol.* **1997**, *4*, 569-578.

15. Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804-808.

16. Steitz, T. A. *Quart. Rev. Biophys.* **1990**, *23*, 203-280.

17. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 3325-3332.

18. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953-6961.

19. Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nature Struct. Biol.* **1998**, *5*, 104.

20. Pilch, D. S. *et al. Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306-8311.

21. Wong, J. M.; Bateman, E. *Nucl. Acids. Res.* **1994**, *22*, 1890-1896.

22. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132-181.

23. Taylor, J.S.; Schultz, P.B.; Dervan, P.B. *Tetrahedron* **1984**, *40*, 457.

24. Momose, T.; Tamaka, T.; Yokota, T.; Nagamoto, N.; Yamada, K. *Chem. Pharm. Bull.* **1978**, *26*, 2224.

25. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.

26. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198-8206.

27. Kim, Y.; Geiger, J. H.; Hahn, S.; Sigler, P. B. *Nature* **1993**, *365*, 512-520.

28. Gartenberg, M. R.; Crothers, D. M. *Nature* **1988**, *333*, 824-829.

29. Sluka, J. P.; Horvath, S. J.; Glasgow, A. C.; Simon, M. I.; Dervan, P. B. *Biochemistry* **1990**, *29*, 6551-6561.

30. Ades, S. E.; Sauer, R. T. *Biochemistry* **1995**, *34*, 14601-14608.

31. Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981-6985.

32. Choo, Y.; Klug, A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 117-125.

33. White, S.; Szewczyk, J.W.; Turner, J.M.; Baird, E.E.; Dervan, P.B. *Nature* **1998**, *391*, 468.

34. Kielkopf, C.L.; White, S.; Szewczyk, J.W.; Turner, J.M.; Baird, E.E.; Dervan, P.B.; Rees, D.C. *manuscript submitted.*

35. (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) de Claire, R.P.L.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1997**, *119*, 7906.

36. (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.

37. Turner, J.M.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, *119*, 7636.

38. Iverson, B.L.; Dervan, P.B. *Nucleic Acids Res.* **1987**, *15*, 7823.

39. Branden, C.; Tooze, J. *Introduction to Protein Structure*; Garland: New York, 1991.

40. Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646.

41. White, S.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, *119*, 6953.

42. Kopka, M.L.; Goodsell, D.S.; Han, G.W.; Chiu, T.K.; Lown, J.W.; Dickerson, R. *Structure* **1997**, *5*, 1033.

43. Brameld, K.; Dasgupta, S.; Goddard, W.A. *J. Amer. Phys. Chem. B.* **1997**, *101*, 4851.

44. (a) Kim, Y.; Geiger, J.H.; Hahn, S.; Sigler, P.B. *Nature* **1993**, *365*, 512. (b) Kim, J.L.; Nikolov, D.B.; Burley, S.K. *Nature* **1993**, *365*, 520.

45. Maxam, A.M.; Gilbert, W.S. *Methods Enzymol.* **1980**, *65*, 499.

46. Sambrook, J.; Fritsch, E.F.; Maniatis, T. *Molecular Cloning* (2^{nd} edn). Cold Spring Harbor Laboratory Press: New York, 1989.