SEQUENCE-SPECIFIC RECOGNITION OF DNA BY PYRROLE-IMIDAZOLE HAIRPIN POLYAMIDES

Thesis by

Michelle E. Parks

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

California Institute of Technology

Pasadena, California

1996

(Submitted May 17, 1996)

To my family

Acknowledgements

Many people have contributed positively to my stay at Caltech, both intellectually and personally. First of all, I would like to thank my advisor Professor Peter B. Dervan for his incredible enthusiasm, support, and direction throughout my graduate career. My decision to pursue graduate studies at Caltech was largely a result of conversations with Peter, who visited Furman while I was still an undergraduate. Peter has provided a good balance of guidance and freedom, allowing the development of independent thinking. In addition, Peter genuinely cares about the development of his students, not only as scientists, but also as individuals. I also thank my undergraduate advisor Professor Moses Lee for providing a great introduction to research and a lasting friendship. In addition, I am grateful to my committee members, Professors Andy Myers, Bill Goddard, and Doug Rees, for helpful advice and support.

I would like to thank the Dervan group members, both past and present, for providing a stimulating scientific atmosphere in which to do research. Special thanks go to my labmates Milan Mrksich and George Best for showing me the ropes and helping me get started when I was a first-year student. Jim Kelly and Natalia Colocci have been very special to me both as colleagues and friends. Being able to work and chat with them on a daily basis made lab infinitely more interesting and enjoyable. I hope our friendship continues to grow as we begin our careers. Scientific discussions and friendships with Dave Liberles, Sue Swalley, Will Greenberg, Scott Priestly, and Bryan Takasaki were appreciated. John Trauger has provided tremendous support and friendship over the past year as well as welcome diversions to the beach. His optimism and constant encouragement have made even the most difficult tasks seem doable. I trust that our friendship will remain strong in spite of the distance between us. Thanks again to Sue, Dave, and John for proofreading this thesis and to Eldon Baird for technical assistance.

Finally, I thank Junhyeong Cho and Eldon Baird for their synthetic contributions to our collaborative work.

Other friendships outside of the Dervan group were also appreciated. In particular, Dom McGrath has been an integral part of my time in graduate school. His encouragement and confidence in me during my first couple of years at Caltech were invaluable. Thanks also to Dom for proofreading both my candidacy report and postdoctoral research proposal. Our friendship continues to be a source of strength for me. Kathy Pomykal has been a great running buddy and friend, in spite of my convincing her to run the Seattle marathon with me!! I have shared many fun times with her and her husband Andy including nights of Canasta and breakfasts after long runs. I will miss them when I go to Boston. Rob Li was also a great running partner and supporter when I was training for the LA marathon. I look forward to running Boston sometime with him! JoyAnne and Tyler Kelly were a welcome break from science. Babysitting Tyler and talking with JoyAnne helped me keep graduate school in perspective. Thanks to Marlys Hammond for her friendship and lunchtime discussions. I look forward to her arrival in Connecticut. Many fun times were shared with Tamara Hendrickson, and I appreciate her support, especially in our first couple of years at Caltech. I would also like to thank Mike Wyatt, my long distance supporter, for always being a great listener and friend. The Caltech Masters swim team provided a great place to socialize and exercise. My coach, Kenny Grace, never failed to put me in a good mood after a workout. Tom Dakan deserves a million thanks for accompanying me and Kathy on long runs and for planning my post-thesis party.

Finally, I would like to thank my parents–Mom, Dad, and Dick. Without their constant support I could never have made it through graduate school. They encouraged me to pursue studies at Caltech, even though they would have preferred a school closer to

home. Their unwavering confidence in me has meant more than they will ever know. Mom and Dick have answered more than their share of phone calls from a distressed daughter. Thanks again for always being there! My Dad's optimistic outlook and down-to-earth attitude has also served as an inspiration to me. I am lucky to have three parents that all love me and want the best in life for me. I would also like to thank all of my relatives back in South Carolina. Although they think I am crazy for moving to California, they have still encouraged and supported me throughout my time here. I look forward to spending more time with them soon.

Abstract

Polyamides composed of pyrrole (Py) and imidazole (Im) amino acids bind as antiparallel, side-by side dimers in the minor groove of DNA to sequences containing both (A,T) and (G,C) base pairs. Initial polyamides based on the 2:1 model have demonstrated new, predictable specificities but only modest affinities. Chapter 2 describes the thermodynamic characterization of a series of covalently linked polyamides, revealing an increase in affinity (>300-fold) and specificity of the hairpin compared to the unlinked dimers. Chapter 3 reports the characterization of the corresponding cyclic polyamide, which results in a 40-fold increase in affinity, although specificity decreases.

Chapter 4 explores the generality of the 2:1 polyamide:DNA model by systematically varying the imidazole content and position within a series of eight three-ring polyamides. Of ten homodimeric and heterodimeric combinations of polyamides characterized, four bind as expected, indicating that the 2:1 model, while not completely general, is a valuable predictive tool. This work adds three new sequences to the targetable repertoire for polyamides, 5'-WGGWW-3', 5'-WGGCW-3', and 5'-WCGCW-3'. Chapter 5 focuses on a series of linked hairpin analogs of the heterodimer AcImImPy-Dp/distamycin that recognize 5'-WGGWW-3' sequences. This study provides the first example of sequence-specific recognition of contiguous G•C base pairs using the side-by-side polyamide:DNA model. In addition, footprinting experiments reveal that linker position does not significantly affect the affinity or specificity of hairpin polyamides. Chapter 6 describes the optimization of polyamide N- and C-terminal groups. Thermodynamic characterizations of a series of ten polyamides show that an unacetylated N-terminus and a C-terminal β-alanine spacer are optimal for hairpin polyamides, providing important design guidelines.

Chapter 7 examines the simultaneous binding of an oligonucleotide in the major groove of DNA and a polyamide in the minor groove. Quantitative footprinting experiments indicate that the stability of a triple helix is not affected by simultaneous recognition in the minor groove, suggesting that oligonucleotide-polyamide conjugates could be designed for sequence-specific DNA recognition. Chapter 8 describes experiments designed to determine the DNA-binding orientation of a portion of a high mobility group protein. Although the synthetic peptides bind A,T-rich DNA, no specific cleavage was observed, precluding determination of binding orientation.

Table of Contents

List of Figures and Tables

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

CHAPTER SIX

CHAPTER SEVEN

CHAPTER EIGHT

APPENDIX ONE

CHAPTER ONE

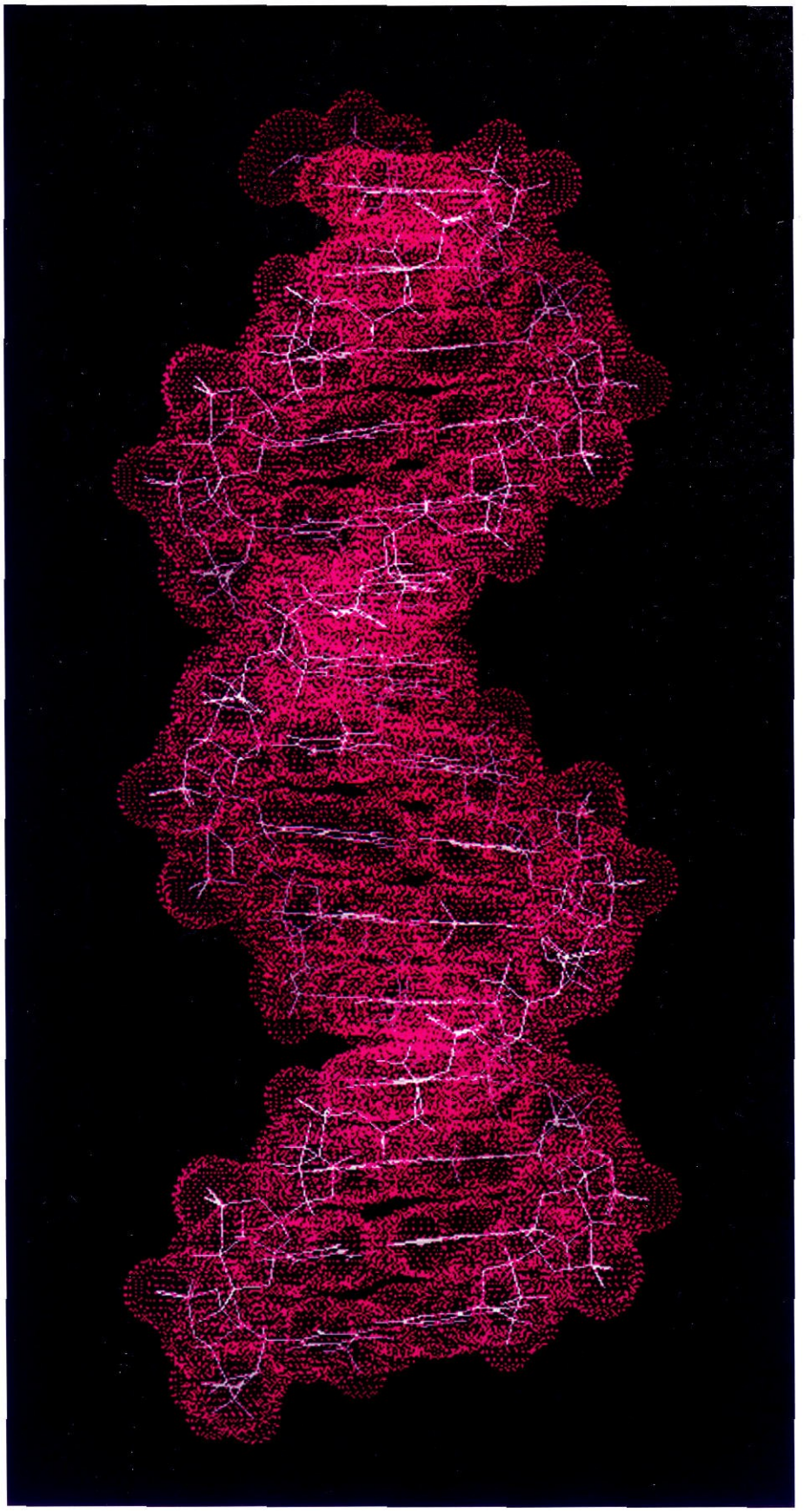
Introduction

Deoxyribonucleic acid plays a crucial role in many biological processes. The DNA sequence provides a unique blueprint for the physical and biochemical characteristics of each individual and mediates the transfer of genetic material to offspring. Protein-DNA interactions are critical for the regulation of DNA replication and transcription, which ultimately determine the phenotype of an organism. Exogenous molecules that could control the expression of certain genes would be useful tools in molecular biology and potentially in human medicine. Drug therapy at the gene level requires a clear understanding of the sequence-specific recognition of DNA. Through the design and synthesis of DNA-binding ligands, predictable recognition of single sites within megabase-size DNA may be realized, which is crucial for therapeutically targeting specific genes.

DNA Structure. Double-helical DNA arises from the antiparallel pairing of two polymeric strands of heterocyclic bases linked by a deoxyribose-phosphate backbone (Figure 1).[1-2] There are four heterocyclic bases in DNA: A, G, C, and T. Specific Watson-Crick hydrogen bonds are formed between guanine and cytosine (G•C base pair) and between adenine and thymine (A•T base pair) (Figure 2).[2] Vertical π-stacking interactions between bases also stabilize the helix. Double-helical B-form DNA contains two grooves, a wide, shallow major groove and a narrow, deep minor groove.[2] Each groove displays a distinct sequence-dependent pattern of hydrogen bond donors and acceptors (Figure 2). For example, in the minor groove an A•T or T•A base pair provides two hydrogen bond acceptors, adenine N3 and thymine O2, resulting in partial degeneracy in recognition. In addition to the analogous purine and pyrimidine hydrogen bond acceptors, a G•C base pair also contains a hydrogen bond donor, the guanine 2-exocyclic amino group. A C•G base pair is distinguished from a G•C base pair by the trajectory of the amino group. Many different types of ligands are known to interact with double-stranded DNA, including proteins, short oligonucleotides, and small molecules.

Protein-DNA Interactions. High resolution X-ray crystal and multidimensional NMR structures of protein-DNA complexes have revealed a diverse repertoire of structural

Figure 1. Structure of B-form double-helical DNA. The DNA is represented as a line model (white) with an overlaid van der Waals surface (magenta). The wide major groove and narrow minor grooves are easily visible.



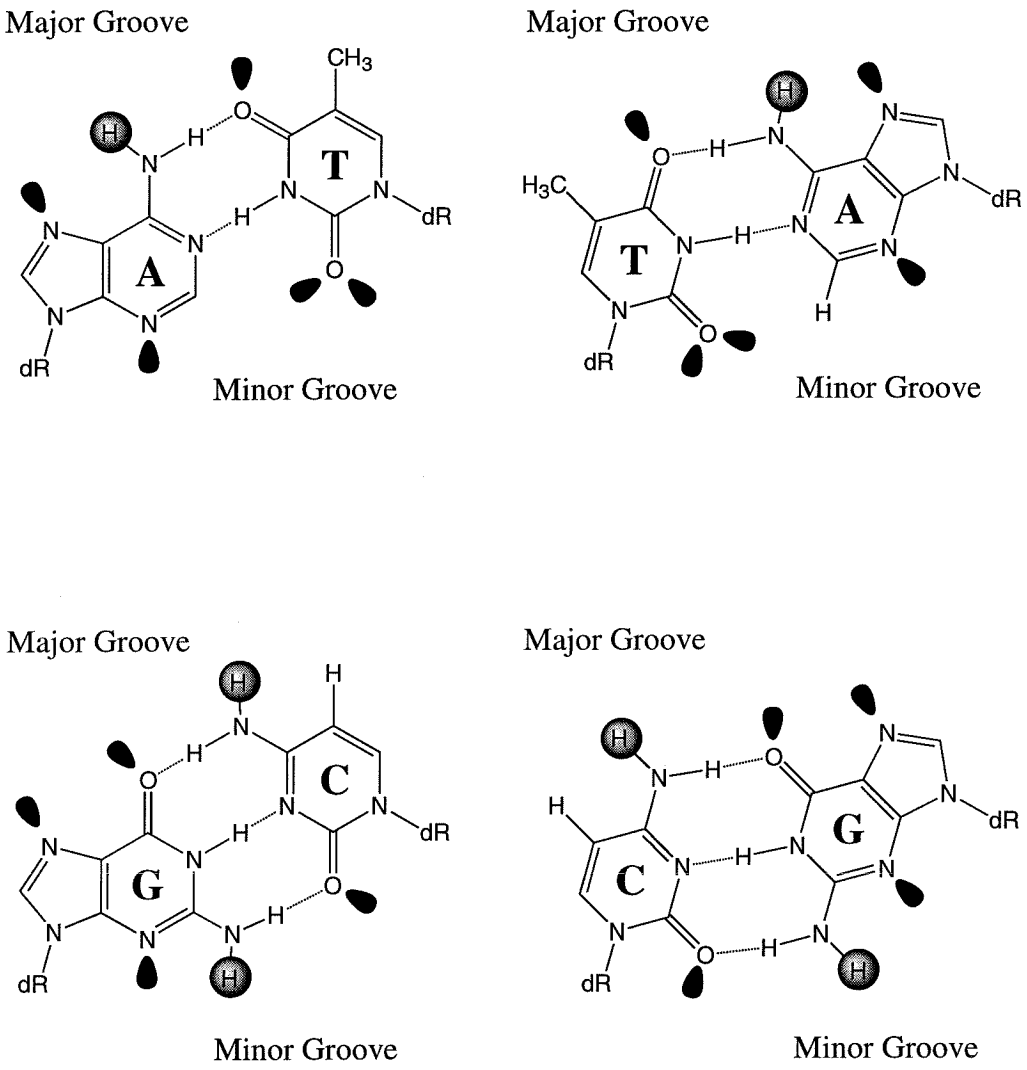


Figure 2. Structure of the four Watson-Crick base pairs. Hydrogen bond acceptors are shown as black lobes while hydrogen bond donors are indicated by gray circles. dR represents the sugar phosphate backbone. Groove location of functional groups is indicated.

motifs for DNA recognition, including the zinc finger, helix-turn-helix, homeodomain, and leucine zipper basic region motifs.[3] Protein-DNA interactions are based on specific hydrogen bonds formed between amino acid residues and base pairs as well as nonspecific protein contacts to the DNA sugar-phosphate backbone.[4] An example of a well-characterized motif is the zinc finger, which forms a compact globular structure consisting of a β-sheet and an α-helix stabilized by a central zinc ion, which is tetrahedrally coordinated to two cysteine residues in the β-sheet and two histidines in the α-helical region.[3a,5] Zinc fingers recognize a diverse set of DNA sequences by specific contacts with bases in the major groove. Through the use of selection techniques, a recognition code for selected zinc fingers has been elucidated, allowing specific recognition of predetermined DNA sites.[6] In contrast, for other binding motifs, no general codes have been deduced, making *de novo* prediction of protein specificity very difficult. Therefore, with the exception of the zinc finger motif, synthetic peptides are currently not very useful for predictable sequence-specific DNA recognition.

Oligonucleotide-Directed Triple Helix Formation. In 1957, Rich and coworkers observed a triple helical 2:1 complex of poly (rU) and poly (rA) in the presence of magnesium.[7] Since this discovery, other three-stranded complexes have been identified, including the 2:1 poly (rG)/poly (rC) triplex that forms at low pH.[8] Thirty years later Moser and Dervan demonstrated oligonucleotide-directed triple helix formation.[9] Using affinity cleavage techniques, Moser and Dervan showed that a pyrimidine-rich third strand binds parallel to a purine-rich DNA strand in what was designated the "pyrimidine motif" (Figure 3). Specificity derives from specific Hoogsteen hydrogen bonds between the pyrimidine third stand and the purine strand of the Watson-Crick duplex.[10] Protonated cytosine recognizes a G•C base pair, forming a C+GC triplet, while thymine recognizes an A•T base pair, forming a T•(AT) triplet (Figure 4). In another class of triple helical complexes, designated the "purine" motif, a purine-rich oligonucleotide binds in the major groove antiparallel to the purine strand of the Watson-Crick duplex.[11] Specificity in this

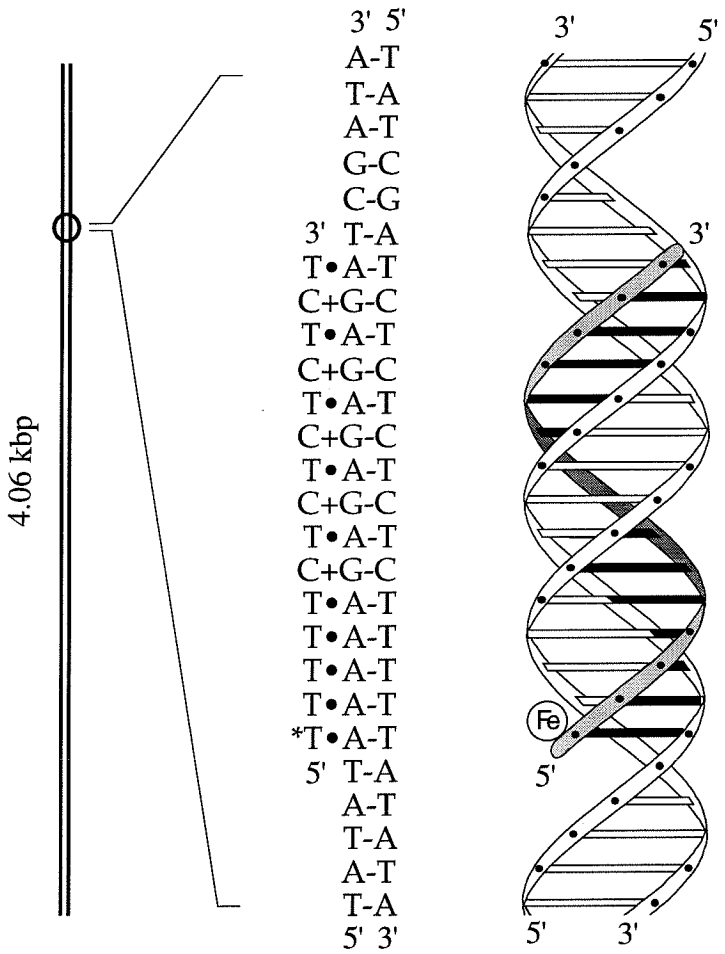
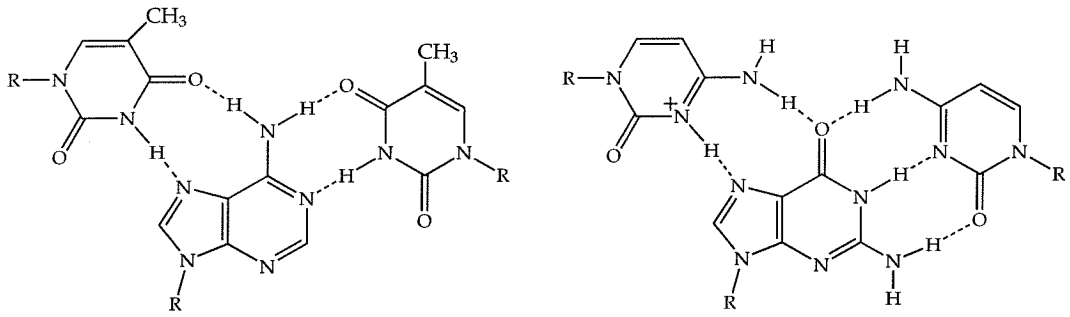


Figure 3. Ribbon model of oligonucleotide-directed triple helix formation on a 4 kbp restriction fragment. The 15 base third stand is indicated by shading and binds parallel to the purine-rich strand.

motif derives from guanine recognition of G•C base pairs (G•GC triplet) and adenine or thymine recognition of A•T base pairs (A•AT or T•AT triplets).

Oligonucleotide-directed triple helix formation has proven to be a powerful means of sequence-specific DNA recognition.[12] This approach toward DNA recognition has been used to mediate cleavage at a single site within human chromosomal DNA[13] as well as to block gene transcription *in vitro*.[14] In spite of the high stabilities and specificities of triple helical structures, currently only purine-rich sequences are targetable. Although a tremendous amount of effort has been dedicated to extending triple helical recognition to C•G and T•A base pairs through the design and synthesis of nonnatural bases,[15] to date no acceptable solution has been found. Therefore, the utility of oligonucleotide-directed triple helix formation is limited by its targetable sequence repertoire.



T•AT **C+GC**

Figure 4. Structures of T•AT and C+GC base triplets. The backbone (R) of the third strand is oriented parallel to the purine strand of the duplex.

Small Molecule Interactions with DNA. Small molecules (MW < 1500), including intercalators and naturally occurring polyamides, provide another means of DNA recognition.[16] Intercalators, such as actinomycin D and echinomycin, bind in the major groove by horizontal insertion between two base pairs, forming stabilizing π-stacking interactions.[17] Alternatively, the natural products netropsin and distamycin A, isolated from *Streptomyces netropsis* and *Streptomyces distallicus*, respectively, bind deep in the minor groove of A,T-rich DNA sequences (Figure 5).[18,19] Stabilization derives from the

formation of bifurcated hydrogen bonds between amide protons of the ligand and thymine O2 and adenine N3 atoms of DNA, electrostatic interactions between the positively charged amidinium or guanidinium end groups of the polyamide and the negatively charged DNA, and extensive van der Waals contacts formed between the ligand and the walls of the minor groove (Figure 6).[20,21]

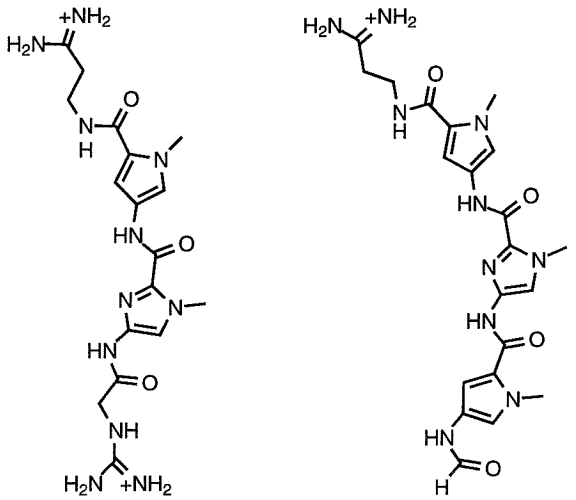


Figure 5. Structures of the natural products netropsin and distamycin.

Sequence-Specific DNA Binding Assays. MPE•Fe(II) footprinting and affinity cleavage are useful techniques for the analysis of sequence-specific DNA-binding molecules.[22] The reagent MPE•Fe(II) cleaves DNA nonspecifically generating a ladder with cleavage at each base pair.[23] Sequence-specific DNA binding by the ligand sterically protects the binding site from cleavage (Figure 7). Through analysis of 3'- and 5'-labeled strands of DNA, which generate asymmetric protection patterns, the binding site size, groove location, and qualitative binding affinity of a ligand can be determined. Affinity cleavage complements footprinting by confirming groove location and providing the additional information of ligand binding orientation within the groove. For affinity cleavage experiments, a ligand is derivatized on one end with a cleaving moiety such as EDTA•Fe(II) and allowed to equilibrate with DNA.[24] Addition of a reducing agent such as dithiothreitol (DTT) results in cleavage proximal to the EDTA•Fe(II) moiety presumably

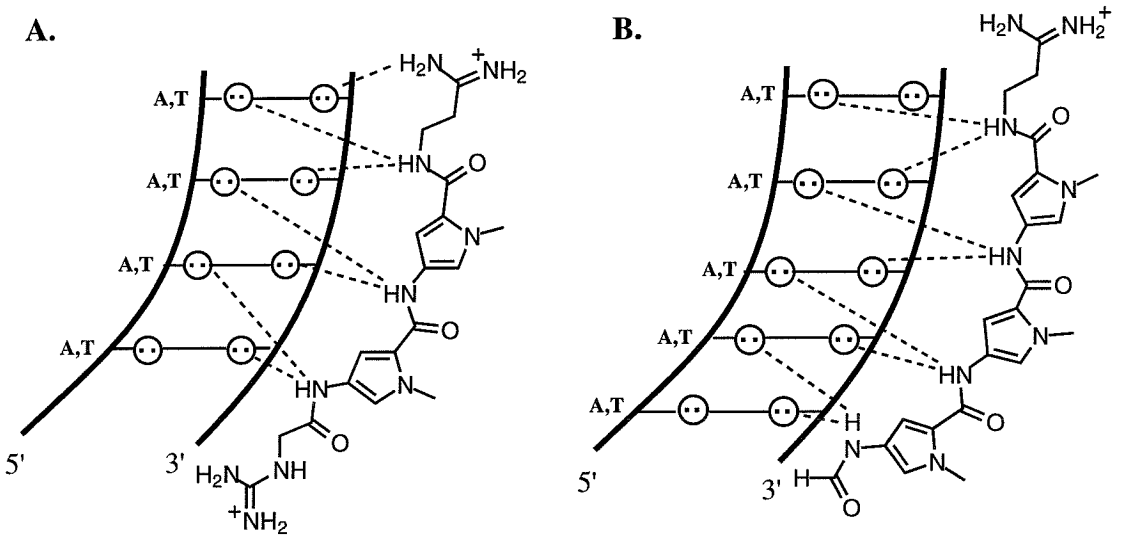


Figure 6. Binding models for the complex formed between (A) netropsin and (B) distamycin with A,T rich sequences. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

by hydroxyl radical formation followed by abstraction of sugar protons. Subsequent analysis of the asymmetric cleavage patterns confirms footprinting results and reveals the DNA-binding orientation of the ligand. For example, a ligand derivatized with EDTA•Fe(II) positioned in the minor groove produces asymmetric 3' shifted cleavage patterns, while an EDTA•Fe(II)-ligand in the major groove generates 5' shifted cleavage patterns (Figure 8).

Quantitative DNase I Footprint Titration Experiments. Quantitative DNase I footprint titration experiments are used to measure equilibrium association constants for ligand binding to DNA.[25] DNase I produces cleavage at base pairs not protected by bound ligand. Titrations with increasing concentrations of ligand provide a quantitatable increase in DNA protection (Figure 9). Quantitation of this protection provides a means of generating an equilibrium binding isotherm (Figure 10). From this isotherm, the equilibrium association constant can be determined.

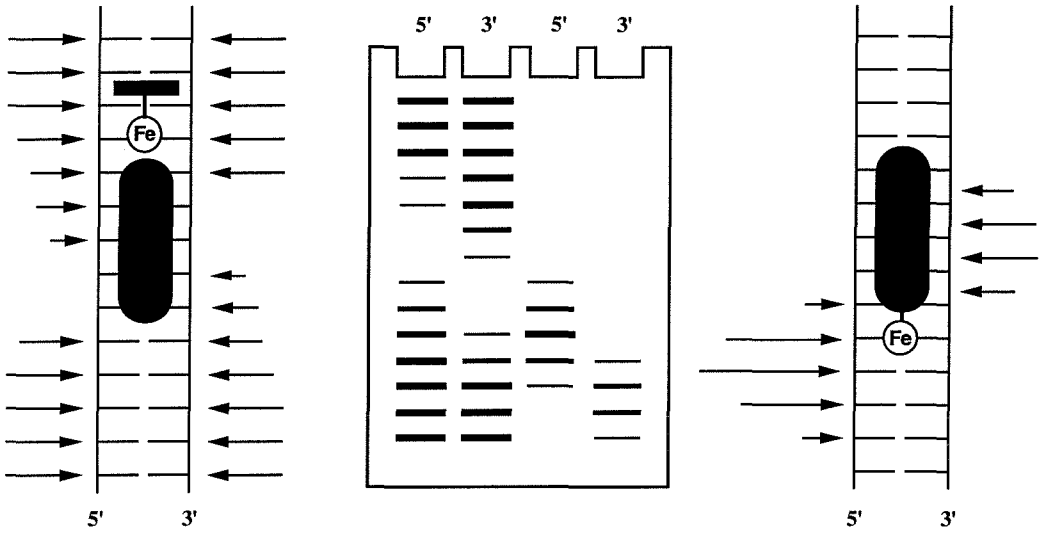


Figure 7. Schematic of (left) MPE•Fe(II) footprinting and (right) affinity cleaving techniques. (Center) Appearance of cleavage products on a typical denaturing gel.

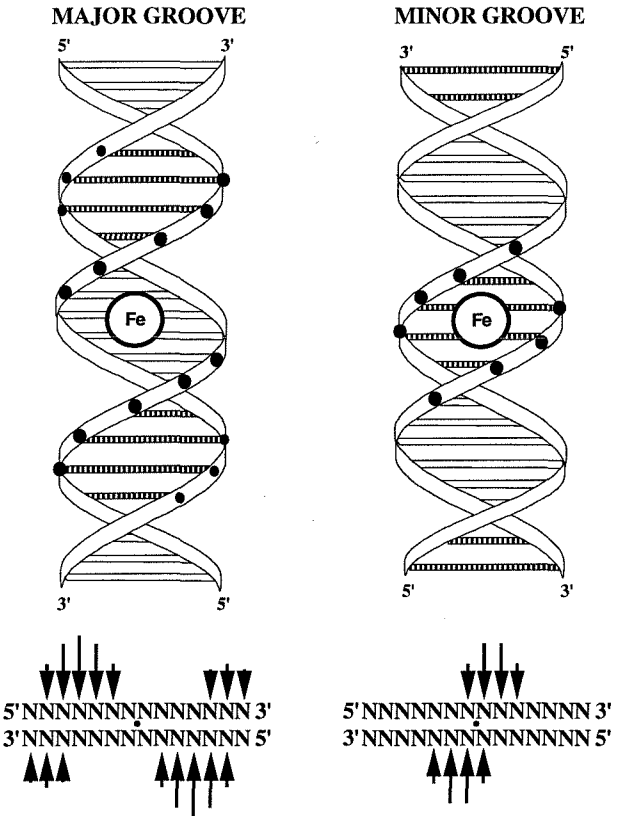


Figure 8. Ribbon models for the asymmetric DNA cleavage pattern generated by EDTA•Fe(II) positioned in the major and minor grooves of right-handed double helical DNA.

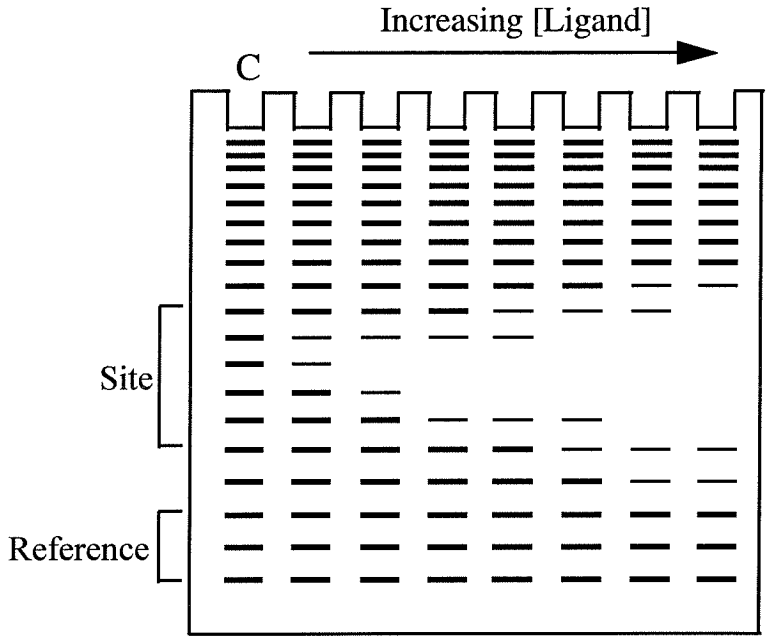


Figure 9. Graphical representation of a gel from a quantitative DNase I footprint titration experiment.

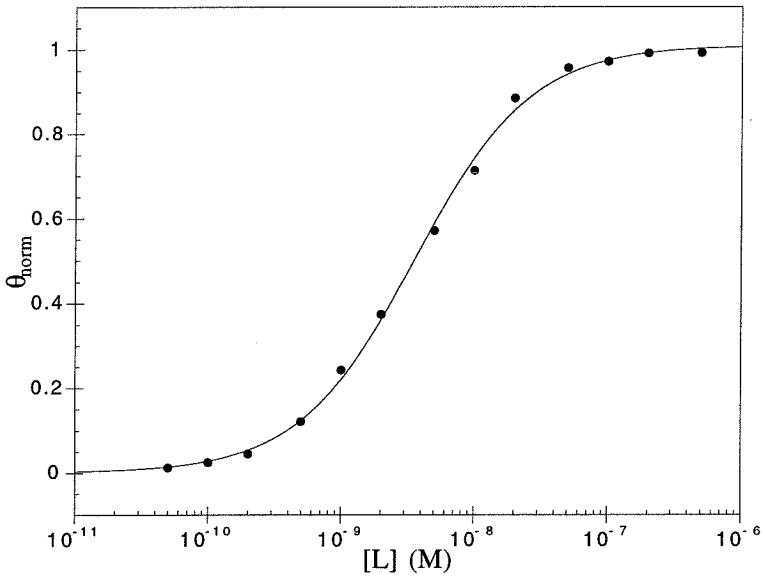


Figure 10. Typical binding isotherm generated from a quantitative DNase I footprint titration experiment. Fractional occupancy, θ_{norm}, and ligand concentration, [L], are plotted on the vertical and horizontal axes, respectively. Ligand concentration at half occupancy represents the equilibrium dissociation constant.

Pyrrole-Imidazole Polyamides Based on a 1:1 Polaymide:DNA Binding Model. Netropsin and distamycin bind to A,T-rich DNA primarily as 1:1 complexes as discussed earlier. Due to positive charges on both ends, netropsin binds solely as a 1:1 complex with DNA. X-ray crystallography[20] and NMR studies[21] of netropsin and distamycin show the crescent-shaped polyamides snugly seated deep in the minor groove. Based on these structures, new polyamides were designed to achieve recognition of sites containing a single G•C base pair.[26] The first successful sequence-specific recognition of G•C base pairs was achieved by the polyamide pyridine-2-carboxamide netropsin, PyrPyPy-Dp, (Pyr = pyridine, Py = pyrrole, Dp = dimethylaminopropylamide) (Figure 11).[27] Originally designed to recognize 5'-(A,T)G(A,T)$_3$-3' sequences in a single orientation, experiments demonstrated binding to a 5'-TGTCA-3' sequence with no orientation preference. In addition, this polyamide retained the parent specificity for A,T-rich DNA. An analogous polyamide, ImPyPy-Dp (Im = *N*-methylimidazole), was more specific for 5'-TGTCA-3' and did not retain substantial affinity for A,T-rich sequences (Figure 11).[28] The proposed 1:1 model for polyamide PyrPyPy-Dp binding to 5'-(A,T)G(A,T)C(A,T)-3' sites involved rotation of an amide bond to enable formation of a hydrogen bond between a carbonyl lone pair and the second guanine exocyclic amino group.[27]

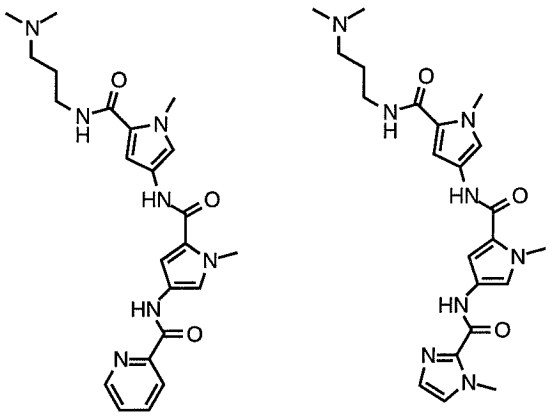


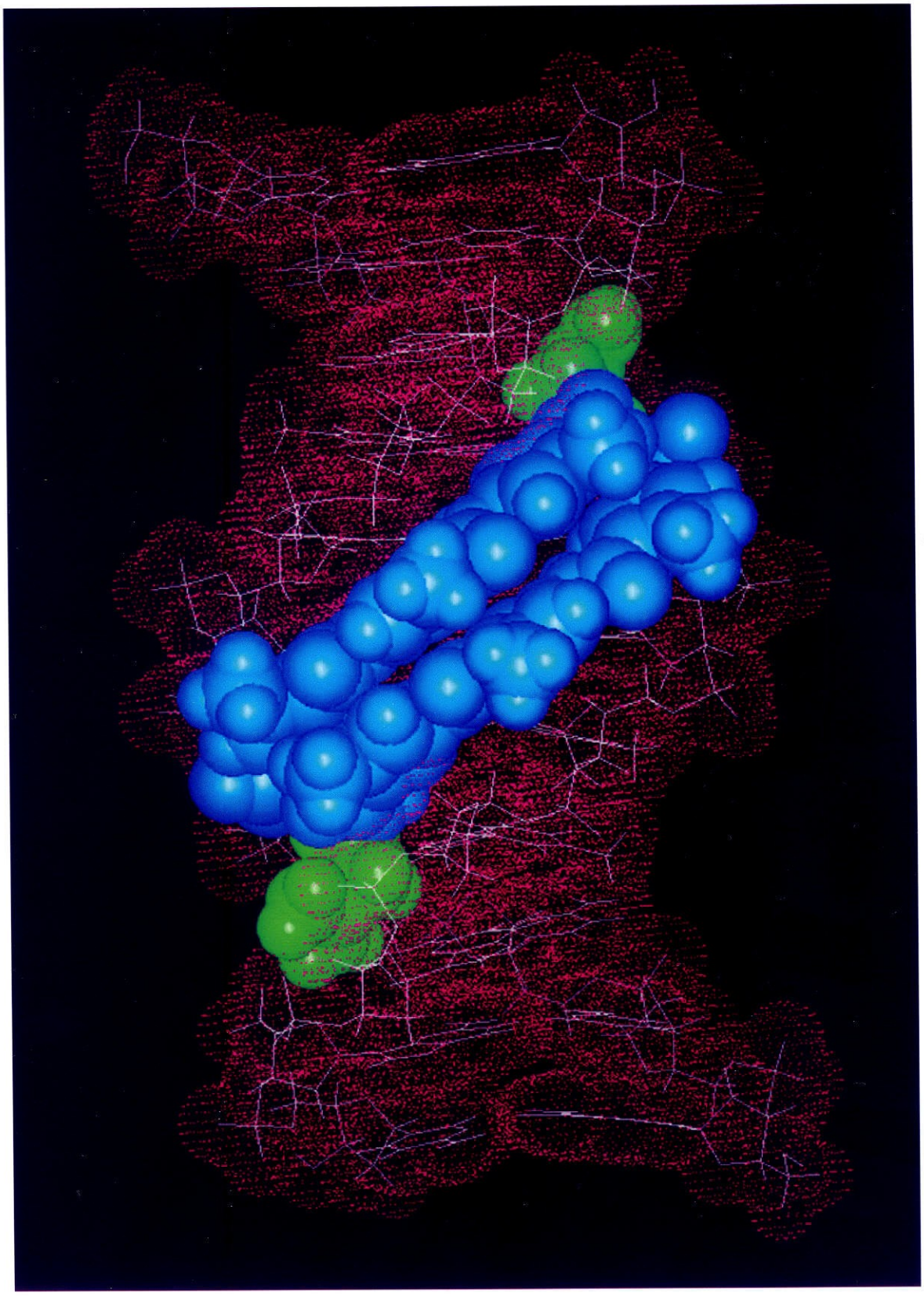
Figure 11. Structures of the synthetic polyamides PyrPyPy-Dp and ImPyPy-Dp.

2:1 Polyamide:DNA Binding Model. Soon after the 1:1 PyrPyPy-Dp model was published, Wemmer and coworkers demonstrated that distamycin forms an antiparallel, side-by-side 2:1 complex with a 5'-AAATT-3' site at millimolar concentrations (Figures 12 and 13).[29] Further studies showed that 2:1 complex formation occurs cooperatively for certain sequences.[30] In light of this new result, the data for PyrPyPy-Dp and ImPyPy-Dp were reexamined. 2:1 models for these polyamides explained the specificity for 5'-(A,T)G(A,T)C(A,T)-3' sequences and the lack of orientation preference (Figure 14).[27,28] Subsequent NMR studies of ImPyPy-Dp and PyrPyPy-Dp in collaboration with the Wemmer group confirmed that both polyamides form 2:1 complexes.[31] In addition, cooperative 2:1 binding was the only binding mode observed.

One explanation for a 2:1 binding preference is groove width. Mixed sequences, i.e. those containing A•T and G•C base pairs, typically have a width of 5-6 Å, while A,T-rich sequences are more narrow at 3-4 Å.[32] Therefore, for polyamides to make extensive van der Waals contacts with the wider minor groove of mixed sequences, binding in a 2:1 mode is optimal. In addition to the sequence-dependent minor groove width being a determinant of specificity, the 2:1 model allows specific contacts with *each strand* on the floor of the minor groove.[28-31,33-35] The revised "rules" dictate that the side-by-side combination of an imidazole ring on one ligand and a pyrrolecarboxamide on the second ligand recognizes G•C, while a pyrrolecarboxamide/imidazole pair targets a C•G base pair.[28,31,33-35] A pyrrolecarboxamide/pyrrolecarboxamide combination is partially degenerate and binds to either A•T or T•A base pairs.[28-31,33-35]

The first successful polyamide design based on the 2:1 motif involved the heterodimeric pairing of ImPyPy-Dp and distamycin to recognize a 5'-(A,T)G$(A,T)_3$-3' target sequence (Figure 15A).[33] MPE•Fe(II) footprinting determined sequence-specific binding to a 5'-TGTTA-3' site, and affinity cleavage confirmed the binding orientation of the 2:1 complex. The polyamide ImPyImPy-Dp was also designed based on the 2:1 motif to bind to a 5'-(A,T)GCGC(A,T)-3' sequence (Figure 15B).[35] Affinity cleavage,

Figure 12. Structure of the 2:1 complex formed between distamycin and a 5'-AAATT-3' site as determined by high resolution NMR studies. The DNA is represented as a line model (white) with an overlaid van der Waals surface (magenta). The *N*-methylpyrrolecarboxamides and the positively charged amidinium group are colored cyan and green, respectively.



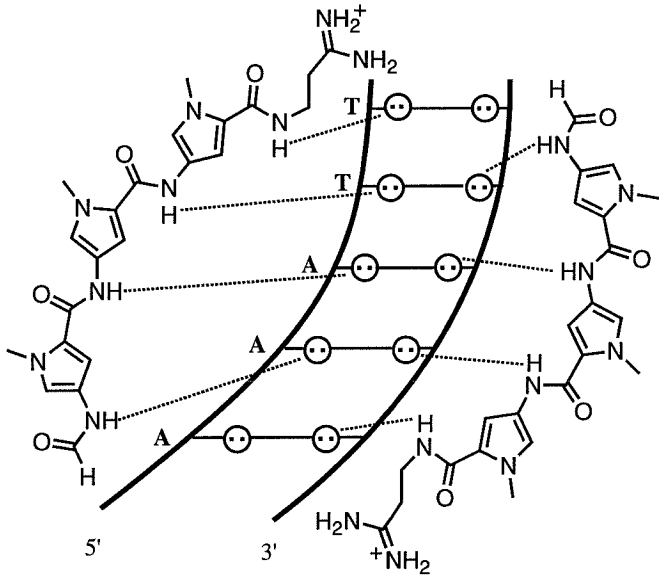


Figure 13. Binding model of the 2:1 complex formed between distamycin and a 5'-AAATT-3' sequence. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

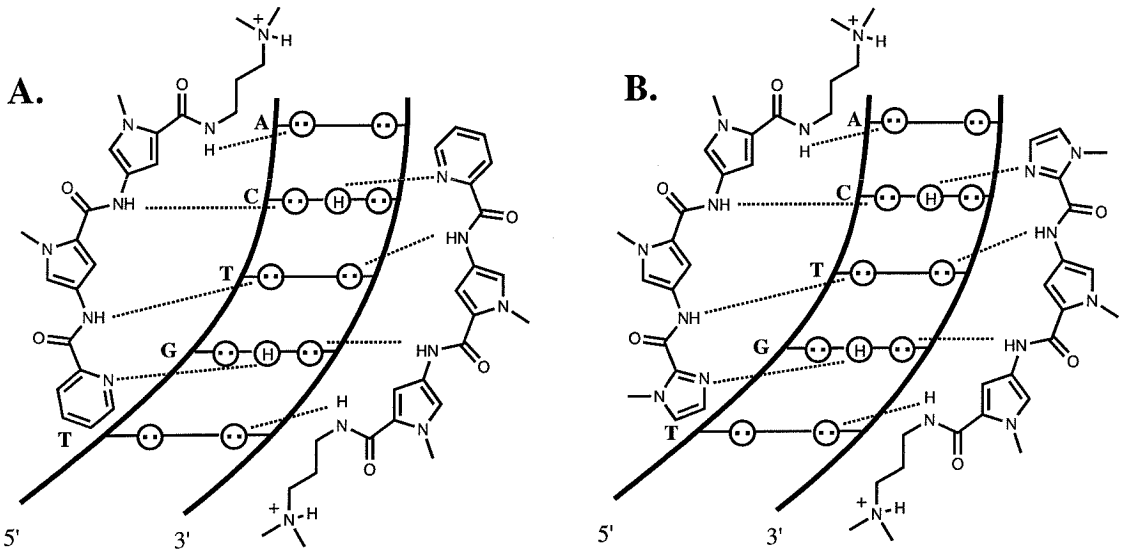


Figure 14. Binding model of the 2:1 complexes formed between either (A) PyrPyPy-Dp or (B) ImPyPy-Dp and a 5'-TGTCA-3' sequence. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

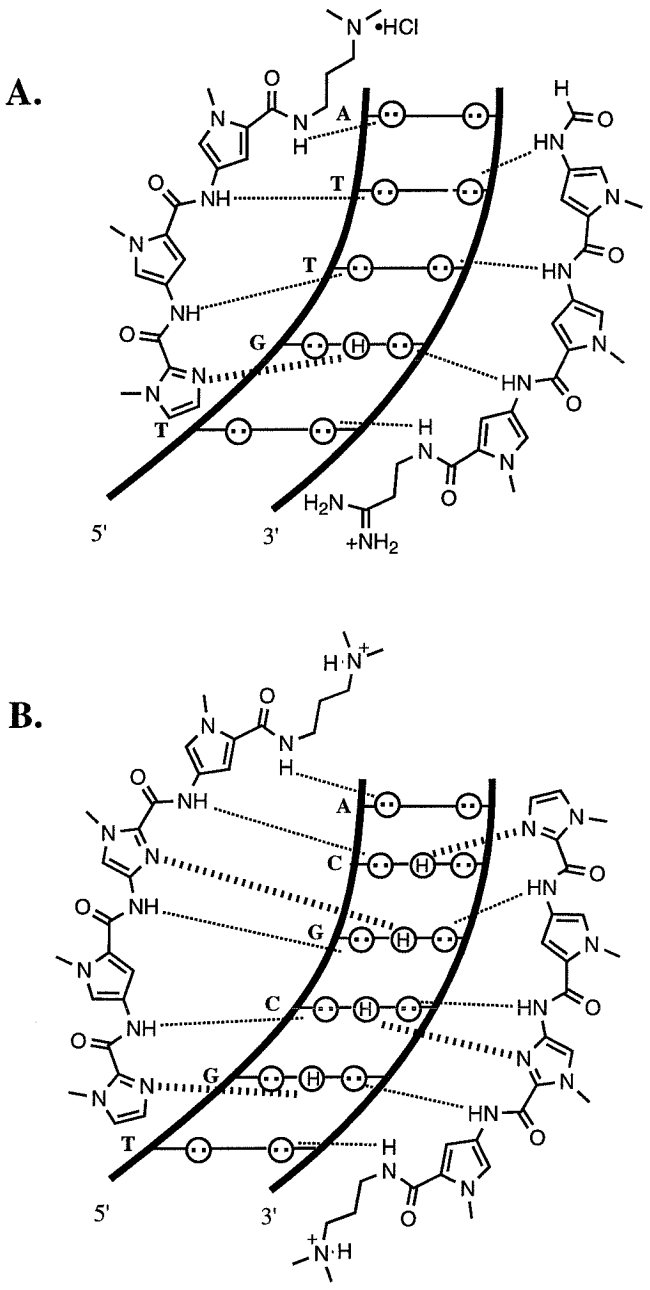


Figure 15. Binding model of the 2:1 complexes formed between (A) ImPyPy-Dp and distamycin with a 5'-TGTTA-3' sequence and (B) ImPyImPy-Dp with a 5'-TGCGCA-3' sequence. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds from ligand amides to pyrimidines are illustrated by dotted lines, while hydrogen bonds between an imidazole N3 and a guanine 2-amino group are illustrated by hashed lines.

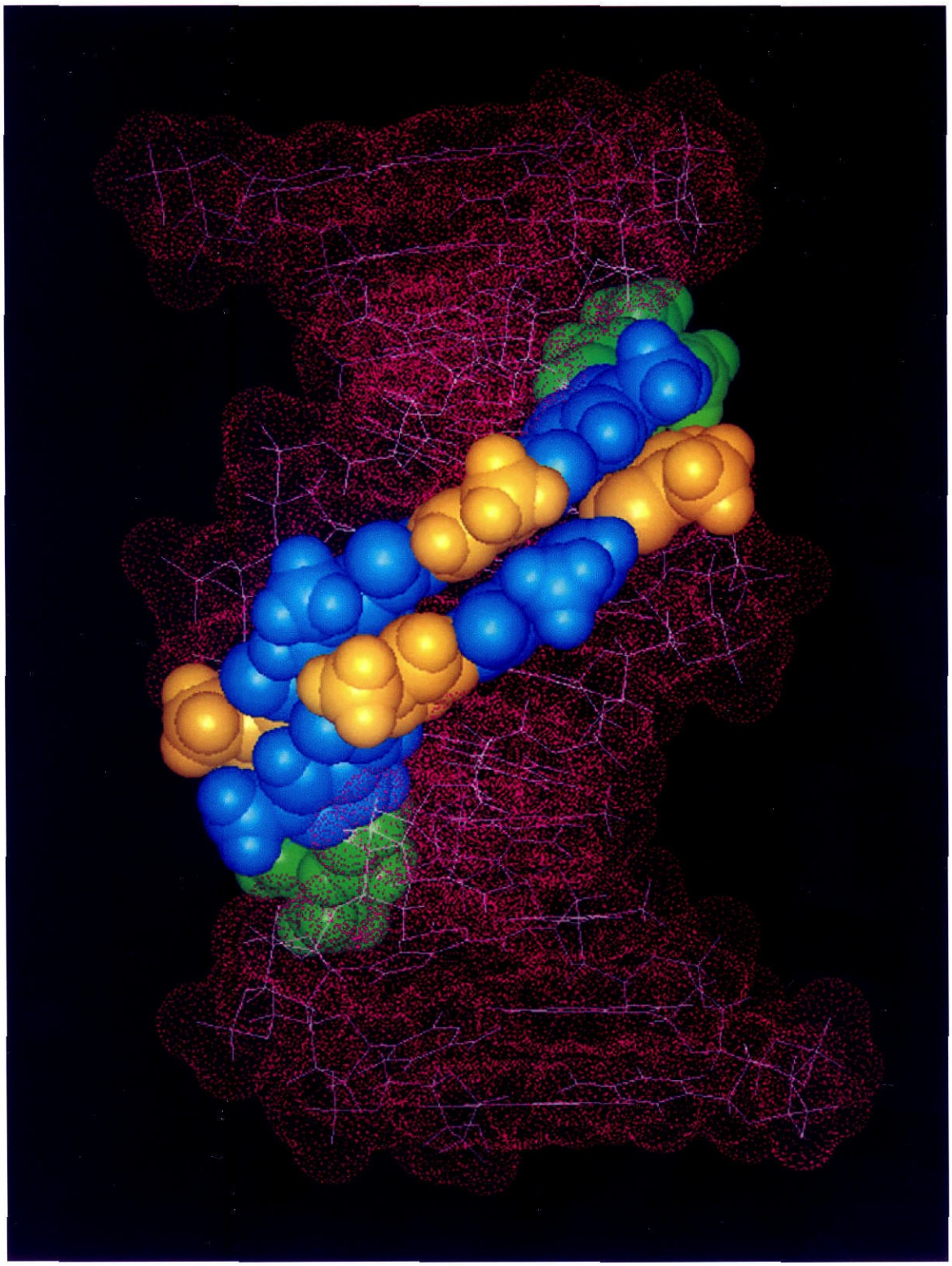
MPE•Fe(II) footprinting, and subsequent NMR studies showed 2:1 homodimeric binding to a 5'-TGCGCA-3' target sequence, a complete reversal of the A,T specificity of the natural products netropsin and distamycin (Figure 16).

Covalently Linked Polyamides. Although new specificities have been achieved, the synthetic polyamides discussed above bind with significantly lower affinity than the natural products distamycin and netropsin. In addition, polyamides that form heterodimeric complexes at target sites are also able to form homodimeric or monomeric complexes at other sites.[33] To restrict binding motifs and increase DNA-binding affinity, polyamides were covalently linked through the N1 position of the central pyrrole units using simple polymethylene linkers.[36] The linker length was optimized using a series of linked PyrPyPy-Dp homodimers in which the number of methylenes was varied from 3 to 6. The linker having four methylenes (C4) was found to be the optimal linker for this system and was subsequently used to covalently link ImPyPy-Dp and AcPyPyPy-Dp, a distamycin analog (Ac = acetyl). As expected, the covalently linked heterodimer had higher affinity and specificity for a 5'-TGTTA-3' site than the unlinked polyamides. Although this linkage was functional, the synthesis was neither general nor efficient.

A second generation of covalently linked polyamides was designed and synthesized in which the N-terminus of one polyamide was linked to the C-terminus of a second polyamide.[37] The simple amino acids glycine, β-alanine, γ-aminobutyric acid, and 5-aminovaleric acid, which range from one to four methylene units, were chosen as linkers. Synthesis of these "head-to-tail" linked molecules was more efficient and general than molecules linked through the pyrrole N1 position. The optimal linker should allow the polyamides to fold back on each other, forming a "hairpin" polyamide. An analysis of the DNA-binding specificities and affinities of these polyamides is contained within this thesis.

Description of This Work. This thesis describes work centered on sequence-specific recognition of DNA by designed polyamides that bind in the minor groove. In addition, some work on simultaneous recognition of the major and minor grooves is

Figure 16. High resolution NMR structure of the (ImPyImPy-Dp)$_2$•5'-TGCGCA-3' complex. The DNA is represented as a line model (white) with an overlaid van der Waals surface (magenta). The imidazole rings, *N*-methylpyrrolecarboxamides, and the positively charged dimethylaminopropyl-amide groups are colored yellow, cyan, and green, respectively.



described. In Chapter 2, quantitative DNase I footprint titration experiments are employed to determine the apparent first-order association constants for the previously synthesized series of polyamides linked head to tail with the simple amino acids glycine, β-alanine, γ-aminobutyric acid, and 5-aminovaleric acid. Results demonstrate that γ is the optimal linker, providing a marked enhancement of affinity (~300-fold) and specificity for the heterodimeric binding site 5'-TGTTA-3' compared to the unlinked polyamides. The γ-aminobutyric acid linker has proven extremely useful in the subsequent design of hairpin polyamides targeted to other sequences. Chapter 3 describes further attempts to increase polyamide affinity and specificity through covalent linkage of the unlinked end of a hairpin with γ-aminobutyric acid to form a cyclic polyamide. Quantitative DNase I footprint titration experiments reveal a higher binding affinity for the cyclic polyamide; however, sequence specificity is decreased. This work suggests that further designs should stem from the hairpin instead of the cyclic polyamide.

Chapter 4 focuses on increasing the targetable sequence repertoire available to pyrrole-imidazole polyamides and testing the generality of the 2:1 polyamide:DNA model. A series of eight three-ring polyamides, in which the imidazole content and location are systematically varied, were synthesized. Ten different homo- and heterodimeric combinations of these polyamides were characterized by MPE•Fe(II) footprinting. Four of the combinations bind as predicted by the 2:1 model. The most problematic polyamide contains a C-terminal imidazole. This study demonstrates that while the 2:1 model is not completely general, it is an extremely valuable predictive tool. In addition, three new sequences were added to the targetable sequence repertoire for polyamides, 5'-WCGCW-3', 5'-WGGWW-3', and 5'-WGGCW-3'.

Chapter 5 describes further experiments with polyamides designed to recognize 5'-WGGWW-3' sequences. The hairpin analog of the heterodimer examined in Chapter 4 was synthesized to determine the thermodynamics of binding to a sequence containing contiguous G•C base pairs. In addition, the effect of the position of the γ-linker on affinity

and sequence specificity was investigated. Quantitative DNase I footprint titration experiments reveal that linker position does not significantly affect affinity or specificity for the fully matched target sequence. However, single base pair mismatch preferences for the polyamides were changed. Polyamides containing contiguous imidazole rings bind with affinities similar to the previously characterized hairpin polyamide, ImPyPy-γ-PyPyPy-Dp. This study demonstrates for the first time sequence-specific recognition of contiguous G•C base pairs based on the 2:1 polyamide:DNA model. This expansion of the targetable sequence repertoire will allow recognition of many new sequences.

Chapter 6 describes experiments designed to optimize the hairpin polyamide ImPyPy-γ-PyPyPy-Dp using recent solid phase synthesis methodology developed by Eldon Baird. The current solid phase methodology requires the use of an aliphatic amino acid spacer between the polyamide and the positively charged C-terminus. A series of polyamides were synthesized in which the N-terminal and C-terminal groups were varied to determine the optimal combination. The optimized polyamide ImPyPy-γ-PyPyPy-β-Dp contains an unacetylated N-terminus and a C-terminal β-alanine spacer. Remarkably, the addition of a β-alanine spacer increases both the equilibrium association constant and the sequence-specificity of this polyamide. This study provides important guidelines for selection of polyamide N- and C-terminal groups.

Chapter 7 explores the simultaneous binding of an oligonucleotide in the major groove of DNA and a polyamide in the minor groove. Quantitative DNase I footprint titration experiments reveal that the stability of the triple helix is not affected by simultaneous recognition of the DNA by either distamycin (1:1) or ImPyPy-Dp (2:1). This study suggests that a new class of hybrid molecules, polyamide-oligonucleotide conjugates, which interact with both the minor and major grooves of DNA, could be designed for sequence-specific recognition of DNA.

Chapter 8 describes experiments designed to determine the binding orientation of an eleven residue portion of the high mobility group I protein (HMG-I). Short peptide

segments of HMG-I that contain the cleaving moiety EDTA•Fe(II) at either the N- or C-terminus were synthesized. Affinity cleavage with the synthetic peptides produced only nonspecific cleavage instead of the desired sequence-specific cleavage. MPE•Fe(II) and DNase I footprinting experiments on the underivatized control peptide reveal binding to the predicted A,T rich sequences, but with very low affinity. Likely, affinity cleavage experiments were unsuccessful due to the low affinity and specificity of the derivatized HMG binding domain peptides.

References

1. Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.
2. Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag; New York, 1984.
3. (a) Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809. (b) Feng, J. A.; Johnson, R. C; Dickerson, R. E. *Science* **1994**, *263*, 348. (c) Li, T.; Stark, M. R.; Johnson, A. D.; Wolberger, C. *Science* **1995**, *270*, 262. (d) Wolberger, C. *Curr. Opin. Struct. Biol.* **1996**, *6*, 62. (e) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223.
4. (a) Steitz, T. A. *Q. Rev. Biophys.* **1990**, *23*, 205. (b) Pabo, C. O.; Sauer, R. T. *Ann. Rev. Biochem.* **1992**, *61*, 1053. (c) Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 804.
5. (a) Parraga, G.; Young, L.; Klevit, R. E. *Trends Biochem.* **1989**, *14*, 398. (b) Lee, M. S.; Gippert, G. P.; Soman, K. V.; Case, D. A.; Wright, P. E. *Science* **1989**, *245*, 635. (c) Klevit, R. E.; Herriott, J. R.; Horvath, S. J. *Proteins* **1990**, *7*, 215. (d) Omichinski, J. G.; Clore, G. M.; Appella, E.; Sakaguchi, K.; Gronenborn, A. M. *Biochemistry* **1990**, *29*, 9324.
6. (a) Choo, Y.; Klug, A.; *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11163. (b) Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11168. (c) Jamieson, A. C.; Kim, S.-H.; Wells, J. A. *Biochemistry* **1994**, *33*, 5689. (d) Rebar, E. J.; Pabo, C. O. *Science* **1994**, *263*, 671. (e) Choo, Y.; Sanchez-Garcia, I.; Klug, A. *Nature* **1994**, *372*, 642. (f) Desjarlais, J. R.; Berg, J. M. *Science* **1991**, *252*, 809.
7. Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023.
8. (a) Michelson, A. M.; Massoulie, J.; Guschlbauer, W. *Prog. Nucleic Acids Res. Mol. Biol.* **1967**, *6*,83. (b) Felsenfeld, G.; Miles, H. T. *Annu. Rev. Biochem.* **1967**, *36*, 407.

9. Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645.

10. (a) Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 637. (b) Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859. (c) de los Santos, C.; Rosen, M. Patel, D. J. *Biochemistry* **1989**, *28*, 7282. (d) Sklenar, V.; Feigon, J. *Nature* **1990**, *345*, 836. (e) Macaya, R. F.; Gilbert, D. E.; Malek, S.; Sinsheimer, J. S.; Feigon, J. *Science* **1991**, *254*, 270. (f) Radhakrishnan, I.; Gao, X.; de los Santos, C.; Live, D.; Patel, D. J. *Biochemistry* **1991**, *30*, 9022. (g) Macaya, R. F.; Schultze, P.; Feigon, J. *J. Am. Chem. Soc.* **1992**, *114*, 781. (h) Radhakrishnan, I.; Patel, D. J.; Gao, X. *Biochemistry* **1992**, *31*, 2514.

11. (a) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360. (b) Radhakrishnan, I.; de los Santos, C.; Patel, D. J. *J. Mol. Biol.* **1991**, *221*, 1403.

12. For a review, see Helene, C. *Angew. Chem. Intl. Ed. Engl.* **1993**, *32*, 666.

13. (a) Povsic, T. J.; Dervan, P. B.; *J. Am. Chem. Soc.* **1990**, *112*, 9428. (b) Strobel, S. A.; Doucette-Stamm, L. A.; Riba, L.; Housman, D. E.; Dervan, P. B. *Science* **1991**, *254*, 1639. (c) Povsic, T. J.; Strobel, S. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 5934.

14. Maher III, L. J.; Dervan, P. B.; Wold, B. *Biochemistry* **1992**, *31*, 70. (b) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456.

15. (a) Koh, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 1470. (b) Griffin, L. C.; Kiessling, L. L.; Beal, P. A.; Gillespie, P.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 7976. (c) Stilz, U.; Dervan, P. B. *Biochemistry* **1993**, *32*, 2177. (d) Koshlap, K. M.; Gillespie, P.; Dervan, P. B.; Feigon, J. *J. Am. Chem. Soc.* **1993**, *115*, 7908. (e) Radhakrishnan, I.; Patel, D. J.; Priestly, E. S.; Nash, H. M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11228. (f) Staubli, A. B.; Dervan, P. B. *Nucleic. Acids Res.* **1994**, *22*, 2637. (g) Hunziker, J.; Priestly, E. S.; Brunar, H.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 2661.

16. For reviews, see (a) Lown, J. W. *Chemtracts-Organic Chemistry* **1993**, *6*, 205. (b) Kahne, D. *Chem. & Biol.* **1995**, *2*, 7. (c) Lown, J. W. *Drug. Dev. Res.* **1995**, *34*, 145. (d) Geierstanger, B. H.; Wemmer, D. E. *Ann. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 463.

17. (a) Fox, K. R.; Marks, J. N.; Waterloh, K. *Nucl. Acids. Res.* **1991**, *19*, 6725. (b) Waring, M. *Ciba Foundation Symposia* **1991**, *158*, 128. (c) Waring, M. J. *Annu. Rev. Biochem.* **1981**, *30*, 159.

18. (a) Arcamone, F.; Bizioli, F.; Canevazzi, G.; Grein, A. German Pat. #1,027,667, 1958. (b) Finlay, A.; Hochstein, F.; Sobin, B.; Murphy, F. *J. Am. Chem. Soc.* **1951**, *73*, 342.

19. Hahn, F. E. in *Antibiotics III. Mechanisms of Action of Antimicrobial and Antitumor Agents*, Gottlieb, P. D.; Shaw, P. D.; Corcoran, J. W. Eds.; Springer: New York, 1975. (b) Zimmer, C. *Progress in Nucleic Acids and Molecular Biology* **1980**, *15*, 258. (c) Krey, A. *Prog. in Mol. Subcell. Biol.* **1980**, *7*, 43. (d) Zimmer, C.; Wähnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31.

20. (a) Berman, H. M.; Neidle, S.; Zimmer, C.; Thrum, H. *Biochim. Biophys. Acta* **1979**, *561*, 124. (b) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E.; *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376. (c) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553. (d) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385. (e) Coll, M.; Aymami, J.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Wang, H.-J. *Biochemistry* **1989**, *28*, 310. (f) Sriram, M.; van der Marel, G. A.; Roelen, H. L. P. F.; van Boom, J. H.; Wang, H.-J. *Biochemistry* **1992**, *21*, 11823. (g) Tabernero, L.; Verdauger, N.; Coll, M.; Fita, I.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Aymami, J. *Biochemistry* **1993**, *32*, 8403.

21. (a) Patel, D. J. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 6424. (b) Patel, D. J.; Shapiro, L. *Biochemie* **1985**, *67*, 887. (c) Sarma, M. H.; Gupta, G.; Sarma, R. H.

J. Biomol. Struct. Dyn. **1985**, *2*, 1085. (d) Patel, D. J.; Shapiro, L. *J. Biol. Chem.* **1986**, *261*, 1230. (e) Klevit, R. E.; Wemmer, D. E.; Reid, B. R. *Biochemistry* **1986**, *25*, 3296. (f) Pelton, J. G.; Wemmer, D. E. *Biochemistry* **1988**, *27*, 8088.

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

23. Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 5470. (b) Van Dyke, M. W.; Dervan, P. B. *Cold Spring Harbor Symposium on Quantitative Biology* **1982**, *47*, 347. (c) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373. (d) Harshman, K. D.; Dervan, P. B. *Nucl. Acids. Res.* **1985**, *13*, 4825.

24. (a) Sluka, J. P.; Horvath, S. J.; Bruist, M. F.; Simon, M. I.; Dervan, P. B. *Science* **1987**, *238*, 1129. (b) Mack, D. P.; Iverson, B. L.; Dervan, P. B. *J. Am. Chem. Soc.* **1988**, *110*, 7572. (c) Dreyer, G. B.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 968. (d) Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861. (e) Taylor, J. S.; Shultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457. (f) Schultz, P. G.; Dervan, P. B. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1133.

25. 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. USA* **1986**, *83*, 8462. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.

26. (a) Lown, J. W.; Krowicki, K.; Bhat, U. G.; Ward, B.; Dabrowiak, J. C. *Biochemistry* **1986**, *25*, 7408. (b) Kissinger, K.; Krowicki, K.; Dabrowiak, J. C.; Lown, J. W. *Biochemistry* **1987**, *26*, 5590. (c) Lee, M.; Chang, M. D.; Hartley, J. A.; Pon, R. T.; Krowicki, K.; Lown, J. W. *Biochemistry* **1988**, *27*, 445. (d) Rao, K. E.; Bathini, Y.; Lown, J. W. *J. Org. Chem.* **1990**, *55*, 728. (e) Plouvier, B.; Bailly, C.; Houssin, R.; Rao, K. E.; Lown, J. W.; Hemichar, J.-P.; Waring, M. J. *Nucleic Acids Res.* **1991**, *19*, 5821.

27. Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574,

28. (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.

29. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393.

30. Fagan, P.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 1080.

31. Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci., USA* **1992**, *89*, 7586.

32. Yoon, C.; Privé, G. G.; Goodsell, D. S.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 6332.

33. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (b) Geierstanger, B. H.; Jacobsen, J-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry*, **1994**, *33*, 3055.

34. Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.

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

36. (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.

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

CHAPTER TWO

Hairpin Polyamide Motif. A New Class of Oligopolyamides for Sequence-Specific Recognition in the Minor Groove of Double-Helical DNA

The text of this chapter is taken from a published paper that was coauthored with Prof. Peter B. Dervan and Milan Mrksich. Nomenclature has been updated to the current standard.

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

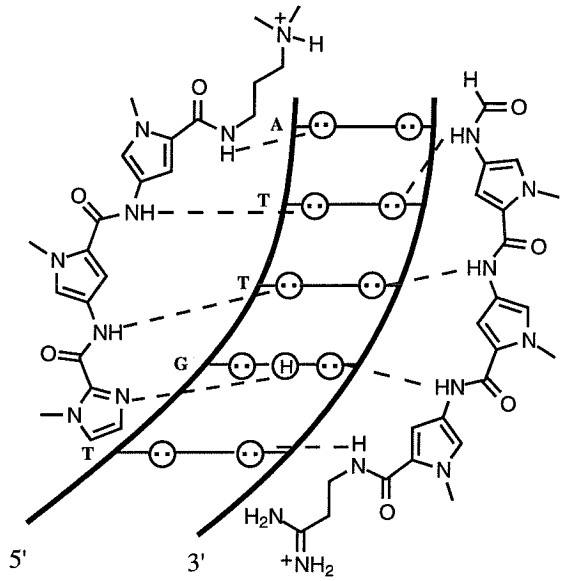
Introduction

Recent 2:1 polyamide-DNA complexes have created new models for the design of nonnatural ligands for sequence-specific recognition in the minor groove of DNA.[1-4] An imidazole-pyrrole-pyrrole tripolyamide, 1-methylimidazole-2-carboxamide-netropsin (ImPyPy-Dp), was shown to specifically bind the mixed sequence 5'-(A,T)G(A,T)C(A,T)-3' as a side-by-side antiparallel dimer.[2] In addition to the sequence-dependent minor groove width being a determinant of specificity, the 2:1 model allows specific contacts with *each strand* on the floor of the minor groove.[1-4] The side-by-side combination of one imidazole ring on one ligand and a pyrrolecarboxamide on the second ligand recognizes G•C, while a pyrrolecarboxamide/imidazole pair targets a C•G base pair.[2-4] A pyrrolecarboxamide/pyrrolecarboxamide combination is partially degenerate and binds to either A•T or T•A base pairs.[1-4]

For some DNA sequences, a single polyamide will suffice, such as the case of the crescent-shaped ImPyPy-Dp which binds 5'-TGACA-3' as a homodimer.[2] However, when two *different* polyamides bind a sequences as a *heterodimer*, there is the added complexity of *each* polyamide having affinity for other sequences as well.[3,4] For example, the two polyamides ImPyPy-Dp and distamycin (D) bind the sequence 5'-TGTTA-3' as a heterodimer and simultaneously bind with comparable affinity the sites 5'-$(A,T)_5$-3' (by D) and 5'-TGACA-3' (by ImPyPy-Dp) as well (Figure 1).[3] For those cases where *different* polyamides contact each strand of the minor groove, there is a need to design motifs that link the side-by-side polyamides to favor the heterodimeric binding site.

We recently reported that polyamides tethered through the central pyrrole rings with a butyl linker bind with increased affinity and specificity.[5,6] However, the synthetic methodology for preparation of this class of covalent polyamide dimers may not be easily extended to the synthesis of other polyamide analogs. We report here a second generation

polyamide motif that offers a more general synthetic methodology for the preparation of a wide variety of covalent polyamide dimers.



(ImPyPy-Dp / D) • TGTTA

Figure 1. Heterodimeric binding model for the complex formed between ImPyPy-Dp and distamycin (D) with a 5'-TGTTA-3' sequence.[3] Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

Hairpin Polyamide Motif. The design of the first-generation covalent polyamide dimers was based on the proximity of the two *N*-methyl groups of the central pyrroles in the 2:1 complex. Alternatively, the two polyamides could also be joined at their ends if a suitable "turn polyamide" could be found which fits in the minor groove of DNA and does not perturb the side-by-side polyamide-DNA complex geometry. Within the antiparallel ImPyPy-Dp/D • 5'-TGTTA-3' complex, the terminal carboxyl group of ImPyPy-Dp and the terminal amine of tris(*N*-methylpyrrolecarboxamide) (PyPyPy-Dp) are positioned for linking with amino acids which allow the linear polyamide "dimer" to *fold back* on itself (Figure 2). A series of polyamides were synthesized wherein the C-terminal carboxylic acid of ImPyPy-Dp and the N-terminal amine of PyPyPy-Dp were connected with simple

amino acids which differ incrementally in length by one methylene unit; glycine (G), β-alanine (β), γ-aminobutyric acid (γ), and 5-aminovaleric acid (Ava) (Figure 3). The binding affinities of the four hexapolyamides ImPyPy-G-PyPyPy-Dp, ImPyPy-β-PyPyPy-Dp, ImPyPy-γ-PyPyPy-Dp and ImPyPy-Ava-PyPyPy-Dp for three sites, 5'-TTTTT-3', 5'-TGTTA-3' and 5'-TGACA-3', on a 135 base pair DNA fragment were determined by quantitative DNase I footprint titration experiments.

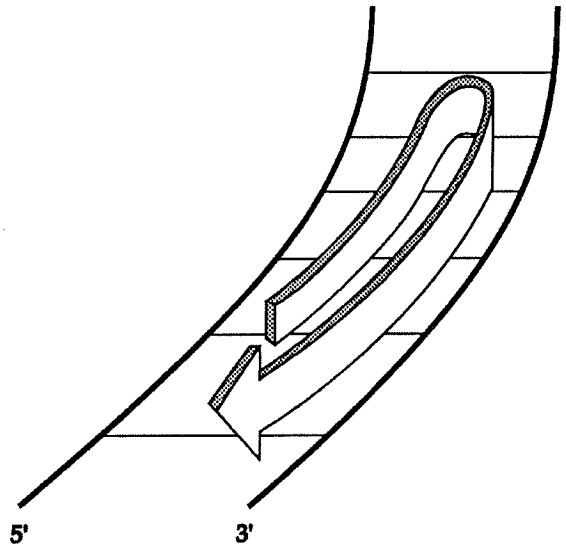
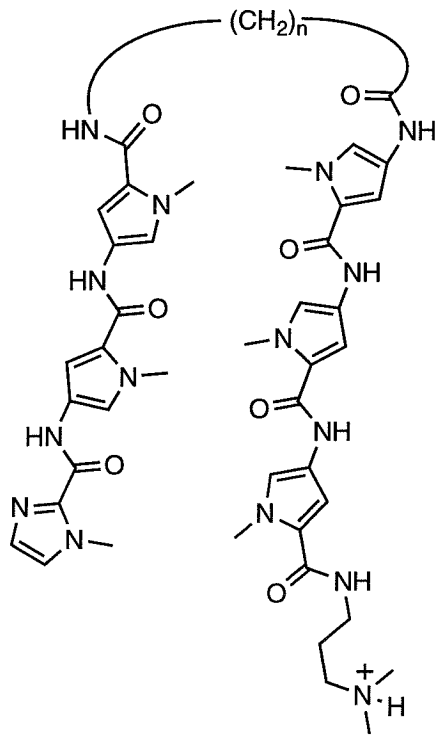


Figure 2. Model for binding of a hairpin polyamide-turn-polyamide in the minor groove of DNA.

Results

Synthesis of Hexapolyamides. The polyamides are synthesized in four steps from previously described intermediates (Figure 4). Reduction of the nitro-tris-(*N*-methylpyrrolecarboxamide)[7] (300 psi H_2, Pd/C) and coupling with the *N*-(*tert*-butoxycarbonyl) amino acid (DCC, HOBt) afforded polyamides **2a-d**. Deprotection of the amine (TFA, CH_2Cl_2) and coupling with *N*-methyl-4-(*N*-methyl-4-nitropyrrole-2-carboxamide)-pyrrole-2-carboxylic acid[8] (DCC, HOBt) yielded **4a-d** in 85% yield. Reduction of the nitropyrrole (300 psi H_2, Pd/C) and coupling to 1-methylimidazole-2-carboxylic acid (DCC, HOBt) provided the hexapolyamides ImPyPy-G-PyPyPy-Dp,



n = 1: ImPyPy-G-PyPyPy-Dp
n = 2: ImPyPy-β-PyPyPy-Dp
n = 3: ImPyPy-γ-PyPyPy-Dp
n = 4: ImPyPy-Ava-PyPyPy-Dp

Figure 3. Hexapolyamides ImPyPy-G-PyPyPy-Dp, ImPyPy-β-PyPyPy-Dp, ImPyPy-γ-PyPyPy-Dp and ImPyPy-Ava-PyPyPy-Dp wherein the terminal carboxylic acid of ImPyPy-Dp and the terminal amine of PyPyPy -Dp are connected with glycine, β-alanine, γ-aminobutyric acid and 5-aminovaleric acid, respectively.

ImPyPy-β-PyPyPy-Dp, ImPyPy-γ-PyPyPy-Dp and ImPyPy-Ava-PyPyPy-Dp. The solubility of the four polyamides has a strong even-odd dependence on the linker length. ImPyPy-G-PyPyPy-Dp and ImPyPy-γ-PyPyPy-Dp were freely soluble in aqueous solution. ImPyPy-β-PyPyPy-Dp and ImPyPy-Ava-PyPyPy-Dp were soluble only at concentrations less than 1.0 mM.

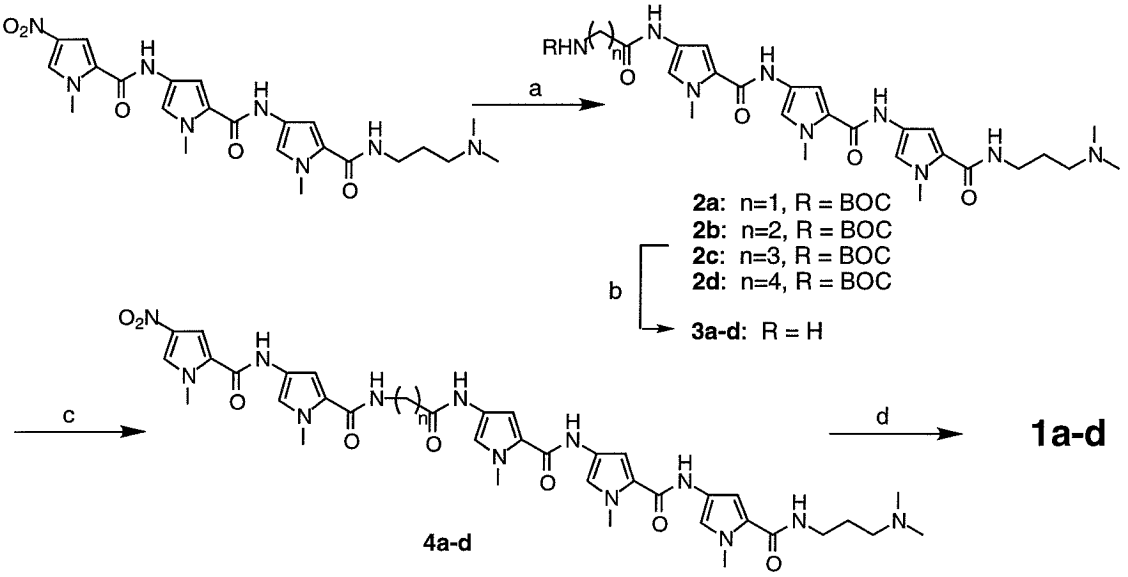


Figure 4. Synthetic scheme for ImPyPy-G-PyPyPy-Dp, ImPyPy-β-PyPyPy-Dp, ImPyPy-γ-PyPyPy-Dp and ImPyPy-Ava-PyPyPy-Dp. (a) (i) 300 psi H_2, 10% Pd/C; (ii) $BOCNH(CH_2)_nCO_2H$, DCC, HOBt; (b) TFA / CH_2Cl_2; (c) (i) 300 psi H_2, 10% Pd/C; (ii) *N*-methyl-4-(*N*- methyl-4-nitropyrrole-2-carboxamide)-pyrrole-2-carboxylic acid, DCC, HOBt; (d) (i) 300 psi H_2, 10% Pd/C; (ii) 1-methylimidazole-2-carboxylic acid, DCC, HOBt.

Footprinting. DNase I footprinting[9,10] on the 3'-^{32}P end-labeled 135 base pair *Eco*RI/*Bsr*BI restriction fragment from plasmid pMM5 (10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0 and 22 °C) reveal that two polyamides, ImPyPy-G-PyPyPy-Dp and ImPyPy-γ-PyPyPy-Dp, specifically bind the five base pair sites in the relative order 5'-TGTTA-3' > 5'-TGACA-3' > 5'-TTTTT-3' (Figures 5 and 6). Two polyamides, ImPyPy-β-PyPyPy-Dp and ImPyPy-Ava-PyPyPy-Dp, bound all sites with weak affinity. The apparent first-order binding affinities of the four polyamides for

the three sites were determined by quantitative DNase I footprint titration experiments (Table I). ImPyPy-γ-PyPyPy-Dp binds the 5'-TGTTA-3' site with the highest affinity, 7.6 x 10^7 M^{-1}. Moreover, this polyamide binds the single "mismatched" site 5'-TGACA-3' with a lower affinity of 3.2 x 10^6 M^{-1}, displaying the best specificity between these two sites. The θ_{app} points for ImPyPy-G-PyPyPy-Dp binding the 5'-TGTTA-3' and 5'-TGACA-3' sites are steeper than expected for a 1:1 complex of polyamide with DNA. The data points are adequately fit by a cooperative binding isotherm consistent with intermolecular dimeric binding.[2c] The data for ImPyPy-γ-PyPyPy-Dp binding these sites were best fit with a Langmuir binding isotherm consistent with a 1:1 complex of polyamide with DNA (Figure 7). For all polyamides binding the 5'-TTTTT-3' site, the binding affinities were less than 5 x 10^5 M^{-1} and are approximate since the quality of fits and the standard deviations for these data sets were poor.

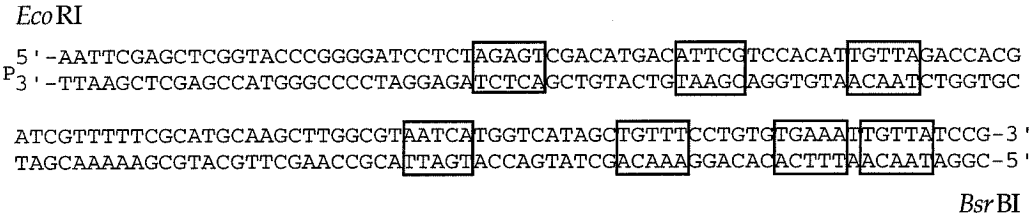
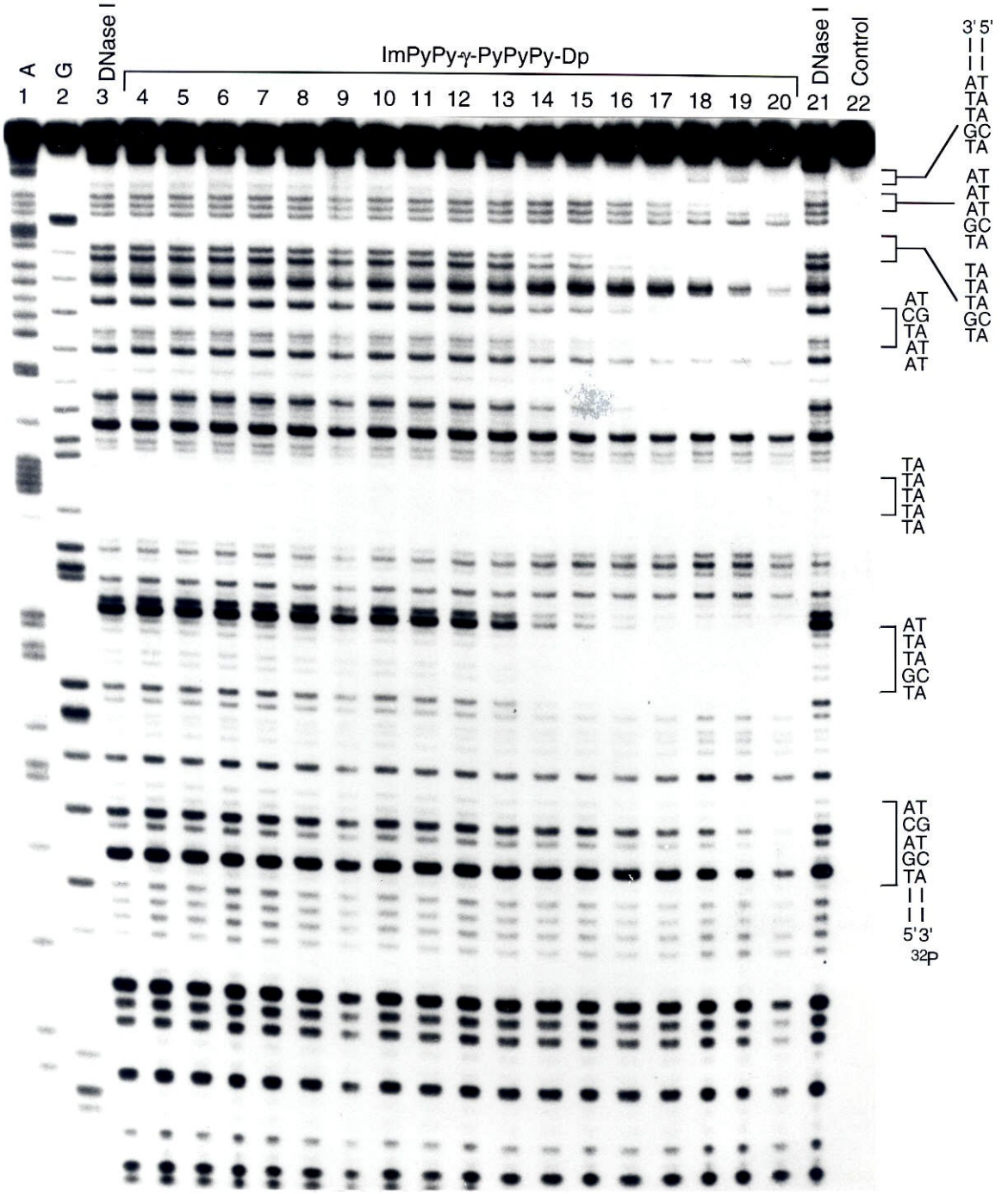


Figure 5. Sequence of the 135 base pair *Eco*RI/*Bsr*BI restriction fragment. Three five base pair sites, 5'-TTTTT-3', 5'-TGTTA-3', and 5'-TGACA-3', proximal to the ^{32}P label at the *Eco*RI site were analyzed by quantitative footprint titration analysis. Note that there are additional binding sites on the upper half of the autoradiogram (Figure 6) of sequence composition (A,T)G(A,T)$_3$ which are also occupied by ImPyPy-γ-PyPyPy-Dp.

Discussion

Binding Affinities. Among the four hexapolyamides, ImPyPy-γ-PyPyPy-Dp binds the 5'-TGTTA-3' site with the highest affinity (~10^8 M^{-1}) and sequence-specificity, which suggests that the hexapolyamide binds this site as an intramolecular hairpin structure (Figure 8). Importantly, the polyamide binds the mismatched 5'-TGACA-3' site with 24-

Figure 6. Quantitative DNase I footprint titration experiment with ImPyPy-γ-PyPyPy-Dp on the 3'-^{32}P-labeled 135 bp *Eco*RI/*Bsr*BI restriction fragment from plasmid pMM5: Lane 1, A reaction; lane 2, G reaction; lanes 3 and 21, DNase I standard; lanes 4-20, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM, 2 μM ImPyPy-γ-PyPyPy-Dp, respectively; lane 22, intact DNA. The 5'-TTTTT-3', 5'-TGTTA-3' and 5'-TGACA-3' binding sites are shown on the right side of the autoradiogram. Additional sites not analyzed are 5'-AATCA-3', 5'-TGTTT-3', 5'-TGAAA-3', and 5'-TGTTA-3'. All reactions contain 20 kcpm restriction fragment, 10 mM tris acetate , 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.


A
G
DNase I
ImPyPy-γ-PyPyPy-Dp
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
TA
GC
TA
AT
AT
AT
GC
TA
TA
TA
TA
GC
TA
AT
CG
TA
AT
AT
TA
TA
TA
TA
TA
AT
TA
TA
GC
TA
AT
CG
AT
GC
TA
5' 3'
32P

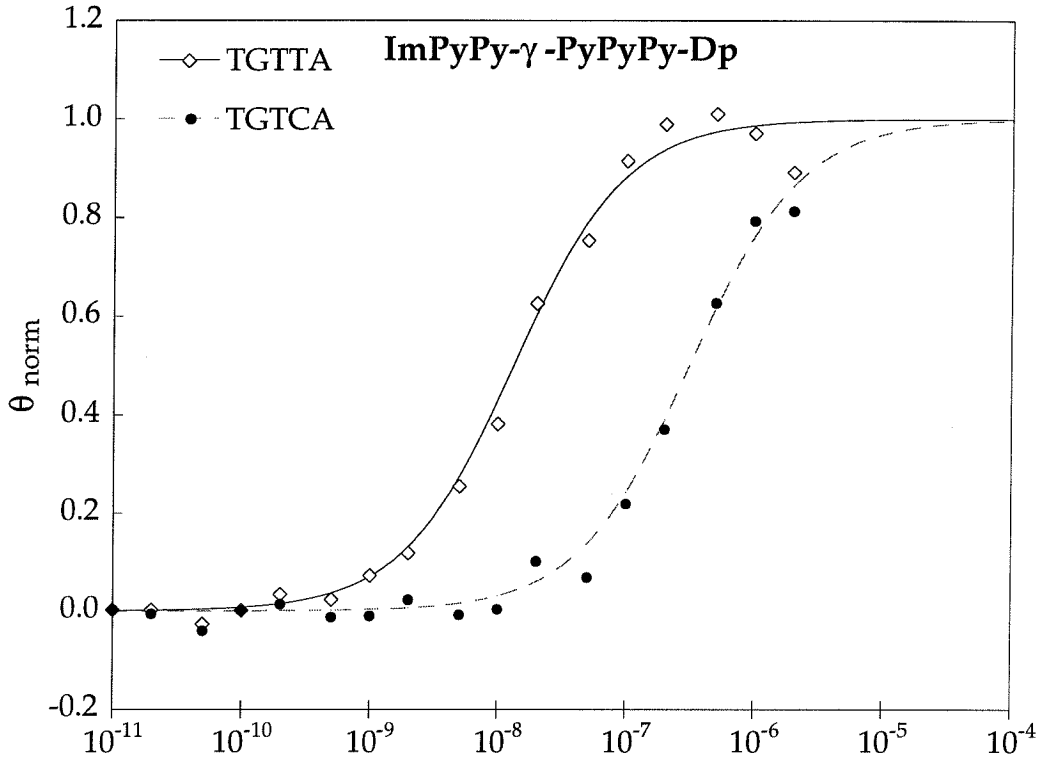


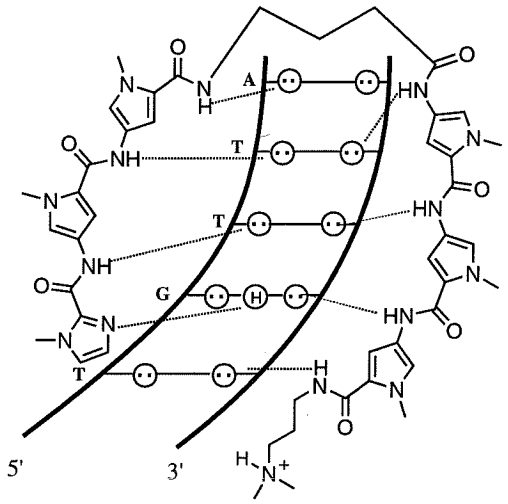
Figure 7. Data for the quantitative DNase I footprint titration experiments for ImPyPy-γ-PyPyPy-Dp in complex with the 5'-TGTTA-3' (match) and 5'-TGACA-3' (single mismatch) sites. The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography and processed as described in the Experimental section. The data points for the 5'-TGTTA-3' site are indicated by open diamonds (◊) and for the 5'-TGACA-3' site by filled circles (•). The solid and dashed curves are the best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm using eq. 2.

Table I. Apparent First-Order Binding Constants (M^{-1})[a, b]

	Binding Sites		
Polyamide	5'-TTTTT-3'	5'-TGTTA-3'	5'-TGACA-3'
ImPyPy-G-PyPyPy-Dp	$< 5 \times 10^5$ [c]	2.2×10^7 (0.2)	$\leq 6 \times 10^6$ [c]
ImPyPy-β-PyPyPy-Dp	$< 5 \times 10^5$ [c]	$\leq 2 \times 10^6$ [c]	$< 5 \times 10^5$ [c]
ImPyPy-γ-PyPyPy-Dp	$< 5 \times 10^5$ [c]	7.6×10^7 (0.8)	3.2×10^6 (0.7)
ImPyPy-Ava-PyPyPy-Dp	$< 5 \times 10^5$ [c]	$< 5 \times 10^5$ [c]	$< 5 \times 10^5$ [c]

[a]Values reported are the mean values measured from three footprint titration experiments, with the standard deviation for each data set indicated in parenthesis. [b]The assays were performed at 22 °C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [c]Approximate binding affinities are reported since the quality of fits and standard deviations for these data sets were poor.

fold lower affinity. Linking the polyamides through the terminal amine and carboxylic acid groups is an effective strategy for improving the sequence *specificities* of different antiparallel side-by-side polyamides for DNA sites. We estimate that the increase in affinity of these hairpin-linked polyamides over the uncoupled components is at least 2 orders of magnitude. This modest energetic gain is not unreasonable when one considers the decrease in the number of positive charges in the DNA-ligand complex in proceeding from the antiparallel dimer motif (two charges) to the hairpin motif (one charge) should be energetically unfavorable. In addition, restricting rotational freedom of the three methylene units in the γ-linker as a loop in the minor groove must have some entropic penalty.



ImPyPy-γ-PyPyPy-Dp • TGTTA

Figure 8. Model for the complex formed between the hexapolyamide ImPyPy-γ-PyPyPy-Dp with the 5'-TGTTA-3' site. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

A comparison of the binding affinites for the four polyamides for the 5'-TGTTA-3' site reveals that the linker length has a dramatic effect on overall complex stability. Modeling suggested that glycine and β-alanine linkers would not favorably accomodate binding of ImPyPy-G-PyPyPy-Dp and ImPyPy-β-PyPyPy-Dp in the minor groove of

DNA as hairpin polyamides. Consistent with the cooperative binding isotherms from the quantitative footprint titration experiments, ImPyPy-G-PyPyPy-Dp likely binds as an intermolecular dimer. Moreover, this polyamide displays a larger binding site size (data not shown) which may result from binding in the minor groove by extended conformations of two polyamides, creating a larger nonspecific steric blockade adjacent to the 5'-TGTTA-3' site to cleavage by DNase I. We defer further speculation about these complexes until structural data from NMR spectroscopy are available.

Implications for the Design of Minor-Groove Binding Molecules. We anticipate that the general and efficient synthetic methodology for preparation of γ-linked polyamides will allow the design of a new class of hairpin polyamides for specific recognition of many different sequences. This strategy involves the construction of a polyamide subunit for recognition in the minor groove of DNA of one strand of the binding site, the γ-linker for the turn, and a second polyamide subunit for recognition of the other strand. The next generation of hairpin polyamides should implement a linker which enforces a turn and disfavors binding of the polyamide in extended conformations which allow 1:1 or intermolecular 2:1 complexes by the tripolyamide subunits. Formally, this can be accomplished by either introducing rigid amino acid linkers or, alternatively, preparing cyclic polyamides incorporating two γ-linkers.

Experimental Section.

^{1}H NMR and ^{13}C NMR spectra were recorded on a General Electric-QE 300 NMR spectrometer in CD_3OD or DMSO-d_6, with chemical shifts reported in parts per million relative to residual CHD_2OD or DMSO-d_5, respectively. The ^{13}C resonances for pyrrole methyl groups are overlapped by resonances from solvent and in general are not listed. IR spectra were recorded on a Perkin-Elmer FTIR spectrometer. High-resolution mass spectra were recorded using fast atom bombardment (FABMS) techniques at the

Mass Spectrometry Laboratory at the University of California, Riverside. Reactions were executed under an inert argon atmosphere. Reagent grade chemicals were used as received unless otherwise noted. Tetrahydrofuran (THF) was distilled under nitrogen from sodium/benzophenone ketyl. Dichloromethane (CH_2Cl_2) and triethylamine were distilled under nitrogen from powdered calcium hydride. Dimethylformamide (DMF) was purchased as an anhydrous solvent from Aldrich. Flash chromatography was carried out using EM science Kieselgel 60 (230-400) mesh.[11] Thin-layer chromatography was performed on EM Reagents silica gel plates (0.5 mm thickness). All compounds were visualized with short-wave ultraviolet light.

***N*-(*tert*-butoxycarbonyl)-amino acid-tris(*N*-methylpyrrolecarboxamide)s 2a-d (Exemplified with 2a).** To a solution of *N*-(*tert*-butoxycarbonyl)glycine (780 mg, 4.46 mmol) in dimethylformamide (4.0 mL) was added a solution of dicyclohexylcarbodiimide (464 mg, 2.25 mmol) in methylene chloride (4.5 mL) and the solution was allowed to stir 30 min. Separately, a solution of tris(*N*-methylpyrrolecarboxamide) (300 mg, 0.60 mmol) and palladium on carbon (10%, 100 mg) in dimethylformamide (9.0 mL) was allowed to stir under a hydrogen atmosphere (300 psi) in a Parr bomb apparatus for 3 hr. The reaction mixture was filtered through celite to remove catalyst and the filtrate was added to the activated acid, followed by stirring for 2 hr. The mixture was filtered through celite, solvents were removed *in vacuo* and the residue was purified by flash column chromatography (0.5 % ammonium hydroxide in methanol) to afford **2a**.

2a: yield 86% (331 mg); ^{1}H NMR (CD_3OD) δ 7.18 (d, 1 H, J = 1.9 Hz), 7.16 (d, 1 H, J = 1.9 Hz), 7.14 (d, 1 H, J = 1.8 Hz), 6.91 (d, 1 H, J = 1.9 Hz), 6.82 (d, 1 H, J = 1.9 Hz), 6.78 (d, 1 H, J = 1.9 Hz), 3.95 (bs, 2 H), 3.89 (s, 3 H), 3.88 (s, 3 H), 3.86 (s, 3 H), 3.29 (m, 2 H), 2.37 (t, 2 H, J = 7.4 Hz), 2.23 (s, 6 H), 1.73 (m, 2 H), 1.43 (s, 9 H); ^{13}C NMR (CD_3OD) δ 169.5, 164.1, 161.2, 161.1, 158.4, 124.5, 123.2, 122.7, 120.7, 120.6, 120.4, 106.4, 106.1, 106.0, 80.7, 58.2, 44.9, 44.6, 38.5, 36.8, 28.7, 28.2; IR (thin film) 3300 (m), 2942 (w), 1643 (s), 1581 (m), 1538 (s), 1464 (m), 1435 (m), 1405 (m),

1258 (m), 1208 (w), 1164 (m), 1106 (w), 775 (w); FABMS *m/e* 626.3415 (M + H, 626.3432 calcd. for $C_{31}H_{46}N_9O_6$).

2b: yield 83% (320 mg); ^{1}H NMR (CD_3OD) δ 7.15 (d, 1 H, *J* = 1.8 Hz), 7.13 (d, 1 H, *J* = 1.8 Hz), 7.07 (d, 1 H, *J* = 1.8 Hz), 6.89 (d, 1 H, *J* = 1.7 Hz), 6.83 (d, 1 H, *J* = 1.7 Hz), 6.78 (d, 1 H, *J* = 1.9 Hz), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.82 (s, 3 H), 3.30 (m, 4 H), 2.49 (t, 2 H, *J* = 6.8 Hz), 2.41 (t, 2 H, *J* = 7.4 Hz), 2.27 (s, 6 H), 1.77 (m, 2 H), 1.42 (s, 9 H); ^{13}C NMR (CD_3OD) δ 170.9, 164.1, 161.3, 158.3, 124.6, 124.5, 123.3, 123.2, 120.8, 120.6, 120.4, 106.5, 106.0, 105.8, 80.1, 58.3, 45.4, 38.6, 38.0, 37.5, 36.8, 36.7, 28.7, 28.3; IR (thin film) 3293 (m), 2944 (w), 1637 (s), 1578 (m), 1542 (m), 1459 (s), 1406 (m), 1260 (m), 1208 (w), 1161 (m), 775 (w); FABMS *m/e* 640.3571 (M + H, 640.3573 calcd. for $C_{30}H_{44}N_9O_6$).

2c: yield 79% (70 mg); ^{1}H NMR (CD_3OD) δ 7.16 (d, 1 H, *J* = 1.9 Hz), 7.14 (d, 1 H, *J* = 1.9 Hz), 7.12 (d, 1 H, *J* = 1.6 Hz), 6.90 (d, 1 H, *J* = 1.7 Hz), 6.81 (d, 1 H, *J* = 1.6 Hz), 6.77 (d, 1 H, *J* = 1.7 Hz), 3.87 (s, 3 H), 3.86 (s, 3 H), 3.84 (s, 3 H), 3.30 (m, 2 H), 3.08 (t, 2 H, *J* = 6.9 Hz), 2.38 (t, 2 H, *J* = 7.8 Hz), 2.32 (t, 2 H, *J* = 7.5 Hz), 2.24 (s, 6 H), 1.77 (m, 4 H), 1.41 (s, 9 H); ^{13}C NMR (CD_3OD) δ 170.7, 162.2, 159.4, 156.5, 122.6, 122.5, 121.2, 118.8, 118.6, 118.4, 104.5, 104.0, 103.9, 78.0, 56.3, 43.4, 38.9, 34.8, 26.8, 26.3, 25.4; IR (thin film) 3301 (m), 2938 (w), 1638 (s), 1578 (m), 1534 (s), 1461 (s), 1436 (s), 1405 (m), 1257 (m), 1205 (w), 1164 (m); FABMS *m/e* 654.3728 (M + H, 654.3695 calcd. for $C_{32}H_{48}N_9O_6$).

2d: yield 84% (385 mg); ^{1}H NMR (CD_3OD) δ 7.17 (d, 1 H, *J* = 1.9 Hz), 7.16 (d, 1 H, *J* = 1.9 Hz), 7.13 (d, 1 H, *J* = 1.9 Hz), 6.91 (d, 1 H, *J* = 1.9 Hz), 6.82 (d, 1 H, *J* = 1.9 Hz), 6.78 (d, 1 H, *J* = 1.9 Hz), 3.89 (s, 3 H), 3.87 (s, 3 H), 3.85 (s, 3 H), 3.30 (m, 2 H), 3.05 (t, 2 H, *J* = 6.8 Hz), 2.41 (t, 2 H, *J* = 7.6 Hz), 2.32 (t, 2 H *J* = 7.4 Hz), 2.26 (s, 6 H), 1.76 (m, 2 H), 1.68 (m, 2 H), 1.51 (m, 2 H), 1.42 (s, 9 H); ^{13}C NMR (CD_3OD) δ 173.1, 164.1, 161.3, 158.5, 124.6, 124.5, 123.2, 120.8, 120.6, 120.4, 106.5, 106.0, 105.9, 79.8, 58.3, 45.4, 40.9, 38.6, 36.8, 30.5, 28.8, 28.3, 24.3; IR (thin film) 3298 (m), 2941

(w), 1642 (s), 1580 (m), 1536 (s), 1466 (m), 1435 (m), 1406 (m), 1258 (m), 1208 (w), 1168 (m), 776 (w); FABMS *m/e* 668.3884 (M + H, 668.3865 calcd. for $C_{33}H_{50}N_9O_6$).

Amino acid-tris(*N*-methylpyrrolecarboxamide)s 3a-d (Exemplified with 3b). To a solution of protected amine **2b** (335 mg, 0.535 mmol) in methylene chloride (13.0 mL) was added trifluoroacetic acid (3.0 mL) and the resulting solution was allowed to stir 20 min. Hexanes (200 mL) were added and the solution was allowed to stir 10 min during which an oil separated. The solvent was removed by decantation, the residue was dissolved in 6% NH_4OH/MeOH (100 mL) and solvent was removed *in vacuo*. Purification of the oil by flash column chromatography (8% ammonium hydroxide in methanol) afforded amine **3b**.

3a: yield 99% (256 mg); ^{1}H NMR (CD_3OD) δ 7.15 (d, 1 H, *J* = 1.8 Hz), 7.14 (d, 1 H, *J* = 1.8 Hz), 7.13 (d, 1 H, *J* = 1.9 Hz), 6.90 (d, 1 H, *J* = 1.8 Hz), 6.83 (d, 1 H, *J* = 1.8 Hz), 6.77 (d, 1 H, *J* = 1.8 Hz), 3.86 (s, 3 H), 3.85 (s, 3 H), 3.83 (s, 3 H), 3.37 (s, 2 H), 3.30 (m, 2 H), 2.41 (t, 2 H, *J* = 7.7 Hz), 2.27 (s, 6 H), 1.75 (m, 2 H); ^{13}C NMR (CD_3OD) δ 171.3, 163.3, 160.5, 160.4, 123.8, 122.4, 122.3, 120.0, 119.7, 119.6, 105.6, 105.2, 105.0, 57.4, 48.2, 44.5, 44.3, 37.7, 36.0, 27.3; IR (thin film) 3300 (s), 1638 (s), 1578 (s), 1546 (s), 1438 (m), 1406 (m), 1262 (w), 1208 (w), 1149 (w), 1108 (w); FABMS *m/e* 526.2912 (M + H, 526.2890 calcd. for $C_{25}H_{36}N_9O_4$).

3b: yield 91% (264 mg); ^{1}H NMR (CD_3OD) δ 7.16 (d, 1 H, *J* = 1.9 Hz), 7.15 (d, 1 H, *J* = 1.9 Hz), 7.14 (d, 1 H, *J* = 1.9 Hz), 6.91 (d, 1 H, *J* = 1.9 Hz), 6.82 (d, 1 H, *J* = 1.9 Hz), 6.77 (d, 1 H, *J* = 1.9 Hz), 3.88 (s, 3 H), 3.87 (s, 3 H), 3.85 (s, 3 H), 3.30 (m, 2 H), 2.95 (t, 2 H, *J* = 6.6 Hz), 2.48 (t, 2 H, *J* = 6.6 Hz), 2.38 (t, 2 H, *J* = 7.6 Hz), 2.25 (s, 6 H), 1.75 (m, 2 H); ^{13}C NMR (CD_3OD) δ 170.6, 163.3, 160.5, 160.4, 123.8, 123.7, 122.4, 122.3, 120.0, 119.7, 119.5, 105.6, 105.2, 105.0, 57.5, 47.3, 44.6, 38.6, 38.1, 37.8, 35.9, 27.4; IR (thin film) 3336 (s), 1638 (s), 1578 (s), 1560 (s), 1540 (s), 1466 (m), 1458 (m), 1406 (m), 1263 (w), 1209 (w), 1151 (w), 1108 (w); FABMS *m/e* 540.3047 (M + H, 540.3051 calcd. for $C_{26}H_{38}N_9O_4$).

3c: yield 91% (54 mg); ^{1}H NMR (CD_3OD) δ 7.17 (d, 1 H, J = 1.9 Hz), 7.16 (d, 1 H, J = 1.9 Hz), 7.13 (d, 1 H, J = 1.9 Hz), 6.91 (d, 1 H, J = 1.8 Hz), 6.82 (d, 1 H, J = 1.8 Hz), 6.77 (d, 1 H, J = 1.9 Hz), 3.88 (s, 3 H), 3.87 (s, 3 H), 3.85 (s, 3 H), 3.30 (m, 2 H), 2.67 (t, 2 H, J = 7.3 Hz), 2.41-2.32 (m, 4 H), 2.24 (s, 6 H), 1.20 (m, 4 H); ^{13}C NMR (CD_3OD) δ 170.5, 163.3, 160.4, 160.3, 123.8, 123.7, 121.1, 120.0, 119.7, 119.5, 105.7, 105.2, 105.0, 57.5, 44.6, 41.2, 37.8, 38.1, 36.0, 35.9, 33.8, 29.2, 27.4; IR (thin film) 3286 (s), 1637 (s), 1580 (s), 1560 (s), 1538 (s), 1466 (m), 1436 (s), 1406 (m), 1262 (w), 1209 (w), 1156 (w), 1108 (w); FABMS *m/e* 554.3203 (M + H, 554.3207 calcd. for $C_{27}H_{40}N_9O_4$).

3d: yield 96% (312 mg); ^{1}H NMR (CD_3OD) δ 7.17 (d, 1 H, J = 1.9 Hz), 7.16 (d, 1 H, J = 1.9 Hz), 7.14 (d, 1 H, J = 1.9 Hz), 6.92 (d, 1 H, J = 1.9 Hz), 6.83 (d, 1 H, J = 1.9 Hz), 6.78 (d, 1 H, J = 1.9 Hz), 3.87 (s, 3 H), 3.85 (s, 3 H), 3.83 (s, 3 H), 3.30 (m, 2 H), 3.86 (t, 2 H, J = 7.2 Hz), 2.37 (t, 2 H, J = 7.6 Hz), 2.32 (t, 2 H, J = 7.3 Hz), 2.23 (s, 6 H), 1.72 (m, 4 H), 1.53 (m, 2 H); ^{13}C NMR (CD_3OD) δ 172.9, 164.1, 161.3, 124.6, 124.5, 123.3, 120.8, 120.6, 120.4, 106.5, 106.0, 105.9, 58.3, 45.4, 41.7, 38.6, 36.8, 36.7, 32.2, 28.3, 24.4, 24.1; IR (thin film) 3287 (s), 2944 (w), 1644 (s), 1581 (s), 1556 (s), 1538 (s), 1467 (m), 1434 (s), 1406 (m), 1262 (m), 1209 (w), 1160 (w), 1107 (w); FABMS *m/e* 568.3360 (M + H, 568.3361 calcd. for $C_{28}H_{42}N_9O_4$).

Nitro-bis(*N*-methylpyrrolecarboxamide)amino acid-tris(*N*-methylpyrrole-carboxamide)s 4a-d (Exemplified with 4b). To a solution of dipyrrole carboxylic acid (303 mg, 1.04 mmol) and *N*-hydroxybenzotriazole hydrate (141 mg, 1.05 mmol) in dimethylformamide (2.1 mL) was added a solution of dicyclohexylcarbodiimide (217 mg, 1.05 mmol) in methylene chloride (2.1 mL) and the solution was allowed to stir 30 min. A solution of primary amine **3b** (242 mg, 0.449 mmol) in dimethylformamide (3.0 mL) was added and the mixture was allowed to stir for 8 hr. Methanol (1 mL) was added and the reaction was filtered through celite. The filtrate was concentrated *in vacuo* and purified by flash column chromatography (1% ammonium hydroxide in methanol) to afford **4b**.

4a: yield 86% (323 mg); ^{1}H NMR (DMSO-d_6) δ 10.29 (s, 1 H), 9.92 (bs, 2 H), 9.89 (bs, 1 H), 8.37 (t, 1 H, J = 5.8 Hz), 8.18 (d, 1 H, J = 1.6 Hz), 8.07 (t, 1 H, J = 6.1 Hz), 7.59 (d, 1 H, J = 1.9 Hz), 7.27 (d, 1 H, J = 1.8 Hz), 7.23 (d, 1 H, J = 1.8 Hz), 7.18 (d, 1 H, J = 1.9 Hz), 7.17 (d, 1 H, J = 1.9 Hz), 7.02 (d, 1 H, J = 1.8 Hz), 6.94 (d, 1 H, J = 1.9 Hz), 6.92 (d, 1 H, J = 1.9 Hz), 6.81 (d, 1 H, J = 1.8 Hz), 3.95 (s, 3 H), 3.90 (bs, 2 H), 3.83 (s, 6 H), 3.82 (s, 3 H), 3.78 (s, 3 H), 3.34 (m, 2 H), 2.22 (t, 2 H, J = 7.1 Hz), 2.12 (s, 6 H), 1.59 (m, 2 H); IR (KBr) 3310 (m), 2934 (w), 1638 (s), 1578 (s), 1534 (s), 1466 (m), 1438 (s), 1406 (m), 1311 (s), 1256 (m), 1208 (m), 1107 (m); FABMS *m/e* 800.3592 (M + H, 800.3578 calcd. for $C_{37}H_{46}N_{13}O_8$).

4b: yield 87% (319 mg); ^{1}H NMR (CD_3OD) δ 7.73 (d, 1 H, J = 1.8 Hz), 7.32 (d, 1 H, J = 1.9 Hz), 7.13 (d, 1 H, J = 1.8 Hz), 7.12 (d, 1 H, J = 1.9 Hz), 7.11 (d, 1 H, J = 1.9 Hz), 7.10 (d, 1 H, J = 1.9 Hz), 6.85 (d, 1 H, J = 1.9 Hz), 6.79 (d, 1 H, J = 1.9 Hz), 6.76 (d, 1 H, J = 1.9 Hz), 6.75 (d, 1 H, J = 1.9 Hz), 3.89 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 6 H), 3.79 (s, 3 H), 3.60 (bs, 2 H), 3.35 (m, 2 H), 2.58 (t, 2 H, J = 6.4 Hz), 2.35 (t, 2 H, J = 7.7 Hz), 2.22 (s, 6 H), 1.73 (m, 2 H); ^{13}C NMR (CD_3OD) δ 171.0, 164.1, 164.0, 161.3, 161.2, 159.5, 136.0, 128.6, 127.6, 124.6, 124.5, 123.2, 123.1, 122.8, 120.8, 120.7, 120.5, 120.4, 108.7, 106.4, 106.0, 105.9, 58.3, 45.4, 38.6, 38.0, 37.1, 37.0, 36.8, 36.7, 28.3; IR (neat) 3284 (m), 2942 (w), 1638 (s), 1578 (s), 1534 (s), 1466 (m), 1438 (s), 1402 (m), 1308 (s), 1259 (m), 1207 (m), 1109 (w); FABMS *m/e* 814.3749 (M + H, 814.3739 calcd. for $C_{38}H_{48}N_{13}O_8$).

4c: yield 81% (28 mg); ^{1}H NMR (CD_3OD) δ 7.78 (d, 1 H, J = 1.8 Hz), 7.35 (d, 1 H, J = 1.9 Hz), 7.15 (d, 1 H, J = 1.9 Hz), 7.14 (d, 1 H, J = 1.9 Hz), 7.13 (d, 1 H, J = 1.8 Hz), 7.06 (d, 1 H, J = 1.9 Hz), 6.86 (d, 1 H, J = 1.9 Hz), 6.79 (d, 2 H, J = 1.8 Hz), 6.76 (d, 1 H, J = 1.9 Hz), 3.94 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.83 (s, 3 H), 3.81 (s, 3 H), 3.35 (q, 2 H, J = 5.6 Hz), 3.30 (m, 2 H), 2.38 (t, 2 H, J = 7.6 Hz), 2.24 (s, 6 H), 1.94 (m, 2 H), 1.75 (m, 2 H); ^{13}C NMR (CD_3OD) δ 172.8, 164.2, 161.3, 161.2, 159.5, 136.1, 128.7, 127.7, 124.6, 124.5, 124.4, 123.2, 123.1, 122.8, 120.8, 120.6, 120.4, 108.7, 106.4,

106.0, 58.4, 45.4, 39.9, 38.6, 38.1, 36.8, 34.9, 28.3, 27.0; IR (neat) 3281 (m), 2931 (w), 1633 (s), 1578 (m), 1530 (s), 1463 (m), 1434 (s), 1400 (s), 1308 (s), 1255 (m), 1205 (m), 1105 (w); FABMS *m/e* 828.3905 (M + H, 828.3923 calcd. for $C_{39}H_{50}N_{13}O_8$).

4d: yield 87% (205 mg); ^{1}H NMR (CD_3OD) δ 7.76 (d, 1 H, *J* = 1.9 Hz), 7.35 (d, 1 H, *J* = 1.9 Hz), 7.14 (d, 1 H, *J* = 1.9 Hz), 7.13 (d, 2 H, *J* = 1.8 Hz), 7.09 (d, 1 H, *J* = 1.8 Hz), 6.87 (d, 1 H, *J* = 1.8 Hz), 6.81 (d, 1 H, *J* = 1.8 Hz), 6.79 (d, 1 H, *J* = 1.9 Hz), 6.75 (d, 1 H, *J* = 1.9 Hz), 3.92 (s, 3 H), 3.84 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.30 (m, 4 H), 2.34 (m, 4 H), 2.22 (s, 6 H), 1.72 (m, 4 H), 1.60 (m, 2 H); ^{13}C NMR (CD_3OD) δ 173.0, 164.1, 161.3, 159.5, 136.1, 128.7, 127.6, 124.8, 124.6, 124.5, 123.2, 122.8, 120.8, 120.6, 120.4, 108.7, 106.5, 106.0, 105.9, 105.8, 58.3, 45.4, 39.8, 38.6, 38.1, 36.8, 30.3, 28.3, 24.4; IR (neat) 3277 (m), 2942 (w), 1638 (s), 1578 (m), 1535 (s), 1460 (m), 1432 (s), 1401 (s), 1308 (s), 1258 (m), 1206 (m), 1109 (w); FABMS *m/e* 842.4062 (M + H, 842.4052 calcd. for $C_{40}H_{52}N_{13}O_8$).

ImPyPy- amino acid-PyPyPy-Dp 1a-d (Exemplified with 1d). To a solution of 1-methylimidazole-2-carboxylic acid (67 mg, 0.603 mmol) and *N*-hydroxybenzotriazole hydrate (82 mg, 0.60 mmol) in dimethylformamide (1.2 ml) was added a solution of 1,3-dicyclohexylcarbodiimide (124 mg, 0.60 mmol) in methylene chloride (1.2 ml) and the mixture was allowed to stir for 50 min. Separately, to a solution of **4d** (150 mg, 0.176 mmol) in dimethylformamide (4.0 ml) was added Pd/C catalyst (10%, 52 mg) and the mixture was hydrogenated in a Parr bomb apparatus (325 psi H_2) for 3 hr. The catalyst was removed by filtration through celite and the filtrate was immediately added to the activated acid and allowed to stir 2.5 hr. Methanol (1.0 ml) was added, the mixture was filtered through celite and the filtrate was concentrated *in vacuo*. Flash column chromatography of the residue (1% ammonium hydroxide in methanol) provided the hairpin polyamide ImPyPy-Ava-PyPyPy-Dp.

1a: yield 85% (81 mg); ^{1}H NMR (CD_3OD) δ 7.29 (d, 1 H, *J* = 1.9 Hz), 7.23 (d, 1 H, *J* = 1.1 Hz), 7.21 (d, 1 H, *J* = 1.9 Hz), 7.17 (d, 1 H, *J* = 1.9 Hz), 7.16 (d, 1 H, *J* =

1.9 Hz), 7.12 (d, 1 H, J = 1.9 Hz), 7.04 (d, 1 H, J = 1.1 Hz), 6.93 (d, 1 H, J = 1.9 Hz), 6.91 (d, 1 H, J = 1.9 Hz), 6.89 (d, 1 H, J = 1.9 Hz), 6.87 (d, 1 H, J = 1.9 Hz), 6.78 (d, 1 H, J = 1.9 Hz), 4.07 (bs, 2 H), 4.04 (s, 3 H), 3.91 (s, 3 H), 3.88 (s, 3 H), 3.86 (s, 6 H), 3.84 (s, 3 H), 3.32 (m, 2 H), 2.43 (t, 2 H, J = 7.6 Hz), 2.28 (s, 6 H), 1.77 (m, 2 H); IR (neat) 3334 (s), 2931 (w), 1636 (s), 1577 (m), 1559 (m), 1542 (m), 1466 (m), 1436 (m), 1405 (m), 1311 (s), 1260 (m), 1210 (m), 1026 (w); UV (H_20) l_{max} (e) 246 (36 600), 304 (42 500) nm; FABMS *m/e* 878.4174 (M + H, 878.4225 calcd. for $C_{42}H_{52}N_{15}O_7$).

1b: yield 78% (80 mg); ^{1}H NMR (DMSO-d_6) δ 7.38 (d, 1 H, J = 0.9 Hz), 7.27 (d, 1 H, J = 1.9 Hz), 7.23 (d, 1 H, J = 1.9 Hz), 7.19 (d, 1 H, J = 1.8 Hz), 7.18 (d, 1 H, J = 1.9 Hz), 7.17 (d, 1 H, J = 1.8 Hz), 7.14 (d, 1 H, J = 1.9 Hz), 7.03 (d, 1 H, J = 1.0 Hz), 7.02 (d, 1 H, J = 1.8 Hz), 6.87 (d, 1 H, J = 1.9 Hz), 6.84 (d, 1 H, J = 1.9 Hz), 6.81 (d, 1 H, J = 1.8 Hz), 3.98 (s, 3 H), 3.83 (bs, 9 H), 3.80 (s, 3 H), 3.78 (s, 3 H), 3.43 (m, 2 H), 3.17 (t, 2 H, J = 6.4 Hz), 2.52 (m, 2 H), 2.22 (t, 2 H, J = 7.0 Hz), 2.12 (s, 6 H), 1.59 (m, 2 H); IR (neat) 3338 (m), 2925 (w), 1638 (s), 1585 (s), 1534 (s), 1466 (m), 1438 (s), 1401 (m), 1260 (m), 1206 (m); UV (H_20) l_{max} (e) 244 (36 100), 306 (48 700) nm; FABMS *m/e* 892.4331 (M + H, 892.4329 calcd. for $C_{43}H_{54}N_{15}O_7$).

1c: yield 55% (13 mg); ^{1}H NMR (CD_3OD) δ 7.28 (d, 1 H, J = 1.9 Hz), 7.23 (d, 1 H, J = 1.1 Hz), 7.15 (d, 3 H, J = 1.8 Hz), 7.09 (d, 1 H, J = 1.9 Hz), 7.04 (d, 1 H, J = 1.1 Hz), 6.90 (d, 1 H, J = 1.9 Hz), 6.88 (d, 1 H, J = 1.9 Hz), 6.81 (d, 1 H, J = 1.9 Hz), 6.80 (d, 1 H, J = 1.9 Hz), 6.77 (d, 1 H, J = 1.9 Hz), 4.04 (s, 3 H), 3.90 (s, 3 H), 3.87 (s, 3 H), 3.86 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.38 (t, 2 H, J = 6.3 Hz), 3.30 (m, 2 H), 2.45-2.38 (m, 4 H), 2.28 (s, 6 H), 1.96 (m, 2 H), 1.77 (m, 2 H); IR (neat) 3384 (m), 2936 (w), 1637 (s), 1579 (m), 1542 (m), 1466 (s), 1406 (m), 1261 (m); UV (H_20) l_{max} (e) 244 (38 900), 308 (48 100) nm; FABMS *m/e* 906.4487 (M + H, 906.4500 calcd. for $C_{44}H_{56}N_{15}O_7$).

1d: yield 72% (116 mg); ^{1}H NMR (DMSO-d_6) δ 7.38 (d, 1 H, J = 0.6 Hz), 7.27 (d, 1 H, J = 1.8 Hz), 7.23 (d, 1 H, J = 1.9 Hz), 7.19 (bs, 3 H), 7.16 (d, 1 H, J = 1.9

Hz), 7.03 (d, 1 H, J = 0.8 Hz), 7.01 (d, 1 H, J = 1.8 Hz), 6.87 (d, 2 H, J = 1.7 Hz), 6.80 (d, 1 H, J = 1.7 Hz), 3.98 (s, 3 H), 3.83 (bs, 6 H), 3.81 (s, 3 H), 3.78 (bs, 6 H), 3.43 (m, 2 H), 3.17 (b, 2 H), 2.22 (m, 4 H), 2.12 (s, 6 H), 1.60-1.46 (m, 6 H); IR (neat) 3301 (m), 2935 (w), 1632 (s), 1581 (s), 1535 (s), 1506 (m), 1466 (m), 1432 (m), 1401 (m), 1260 (m), 1204 (m); UV (H_2O) l_{max} (e) 246 (38 700), 310 (52 900) nm; FABMS *m/e* 920.4644 (M + H, 920.4640 calcd. for $C_{45}H_{58}N_{15}O_7$).

Construction of Plasmid DNA. Using T4 DNA ligase, the plasmid pMM5 was constructed by ligation of an insert, 5'-TCGACATGACATTCGTCCACATTGTTAG-ACCACGATCGTTTTTCGCATG-3' and 5'-CGAAAAACGATCGTGGTCTAACAAT-GTGGACGAATGTCATG-3' into pUC19 previously cleaved with *Sal*I and *Sph*I. Ligation products were used to transform Epicurian^{TM} Coli XL 1 Bluecompetent 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. Plasmid DNA concentration was determined at 260 nm using the relation 1 OD unit = 50 mg/mL duplex DNA. The plasmid was digested with *Eco*RI, labeled at the 3'-end, and digested with *Bsr*BI. The 135-base pair restriction fragment was isolated by nondenaturing gel electrophoresis and used in all experiments described here. Chemical sequencing reactions were performed as described.[12,13] Standard techniques were employed for DNA manipulations.[14]

Quantitative DNase I Footprint Titration. All reactions were executed in a total volume of 40 μL with final concentrations of each species as indicated. The ligands, ranging from 2 μM to 10 pM, were added to solutions of radiolabeled restriction fragment (20,000 cpm), Tris•HCl (10 mM, pH 7.0), KCl (10 mM), $MgCl_2$ (10 mM) and $CaCl_2$ (5 mM) and incubated for 1 h at 22 °C. Footprinting reactions were initiated by the addition of 4 μL of a stock solution of DNase I (0.04 unit/mL) containing 1 mM dithiothreitol and allowed to proceed for 5 min at 22 °C. The reactions were stopped by addition of a 3 M

ammonium acetate solution containing 50 mM EDTA and ethanol precipitated. The reactions were resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer and electrophoresed on 8% polyacrylamide denaturing gels (5% crosslink, 7 M urea) at 2000 V for 1 h. The footprint titration gels were dried and quantitated using storage phosphor technology.

Apparent first-order association constants were determined as previously described.[2c,5a,10] The data were analyzed by performing volume integrations of the 5'-TGTTA-3' and 5'-TGACA-3' target sites and a 5'-AGAG-3' 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. At higher concentrations of polyamide (> 2 μM), the reference sites become partially protected, resulting in low θ_{app} values. For this reason, concentrations > 2 μM were not used. 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 apparent 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) running on a Macintosh IIfx computer 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 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:

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

The best fit isotherms for ImPyPy-G-PyPyPy-Dp binding the 5'-TGTTA-3' and 5'-TGACA-3' sites consistently give worse fits than those for ImPyPy-γ-PyPyPy-Dp binding these sites. Visual inspection of the binding curves for the former polyamide reveals that the increase in θ_{max} near half-saturation of the site is steeper than expected from the fitted curve. Fitting of these data to a general Hill model (eq 2) with K_a, θ_{max}, and θ_{min} and n as adjustable parameters afforded best fit values of n greater than 2.0 in all cases. Therefore, the reported binding affinities for ImPyPy-G-PyPyPy-Dp binding the 5'-TGTTA-3' and 5'-TGACA-3' sites were obtained by performing fits with a modified Hill eqation (eq 2, n=2). The ($[L]_{tot}$, θ_{app}) data points for ImPyPy-γ-PyPyPu-Dp binding these sites were adequately described by Langmuir binding isotherms consistent with a 1:1 polyamide-DNA complex (eq 2, n=1). We note explicitly that treatment of the data in this manner does not represent an attempt to model a binding mechanism. Rather, we have chosen to compare values of the apparent first order dissociation constant, because this paramter represents the concentration of polyamide at which the binding site is half-saturated.

References

1. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723-5727. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393-1399. (c) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Struct. Biol. Nat.* **1994**, *1*, 169-175.
2. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci., USA* **1992**, *89*, 7586-7590. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385-11389.
3. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572-2576. (b) Geierstanger, B. H.; Jacobsen, J-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry*, **1994**, *33*, 3055.
4. Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474-4482.
5. (a) For covalent homodimer polyamidesee: Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892-9899. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900-9906.
6. For covalent heterodimer polyamides see: Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663-3664..
7. Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457.
8. (a) Bialer, M.; Yagen, B.; Mechoulam, R. *Tetrahedron* **1978**, *34*, 2389-2391. (b) Lown, J. W.; Krowicki, K. J. *J. Org. Chem.* **1985**, *50*, 3774-3779.
9. (a) Galas, D.; Schmitz, A. *Nucl. Acids. Res.* **1978**, *5*, 3157-3170. (b) Fox, K. R.; Waring, M. J. *Nucl. Acids Res.* **1984**, *12*, 9271-9285.

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

11. Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, *40*, 2923-2925.

12. Iverson, B. L.; Dervan, P. B. *Nucl. Acids Res.* **1987**, *15*, 7823-7830.

13. Maxam, A. M.; Gilbert, W. S. *Methods in Enzymology* **1980**, *65*, 499-560.

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

CHAPTER THREE

Cyclic Polyamides for Recognition in the Minor Groove of DNA

The text of this chapter is taken from a published paper that was coauthored with Prof. Peter B. Dervan and Junhyeong Cho. Nomenclature has been updated to the current standard.

(Cho, J.; Parks, M. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10389-10392.)

Introduction

Footprinting, affinity cleavage, NMR and x-ray studies have established that 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 (1-16). The DNA sequence specificity of these small molecules can be controlled by the linear sequence of pyrrole and imidazole amino acids (1-13). An imidazole ring on one ligand complemented by a pyrrolecarboxamide ring on the neighboring ligand recognizes a G•C base pair, while a pyrrolecarboxamide/ imidazole combination targets a C•G base pair (2-4). A pyrrole/pyrrole pair is degenerate for A•T or T•A base pairs (1-4).

Despite this design breakthrough in molecular recognition of DNA, the binding affinities of this new class of polyamide dimers are modest (4). For example, a three ring polyamide dimer in complex with a five base pair DNA site, has a binding affinity typically in the range of $K = 2 \times 10^5\ M^{-1}$ (pH 7.0, 22 °C) (4). In an effort to improve the energetics, antiparallel dimers were connected by a central γ-aminobutyric acid (γ) residue to create a single molecule which could bind in the minor groove by folding to a hairpin shape (11). Second generation "hairpin" polyamides of sequence composition ImPyPy-γ-PyPyPy-Dp **1** were shown to have improved equilibrium association constants of $8 \times 10^7\ M^{-1}$ for designated five base pair target sites 5'-TGTTA-3', an increase of 400 over the unlinked dimers (Figure 1). In a formal sense, γ-aminobutyric acid linked pyrrole-imidazole polyamides could exist in at least two conformations, hairpin or extended, resulting in different DNA binding motifs, and hence possible different DNA specificities. Closing the ends of the hairpin to a circle would restrict conformational space for the DNA binding molecule and presumably further increase the overall energetics. We report here the synthesis of *cyclo*-(ImPyPy- γ-Py-Py-Py-γ) **2** and analysis of the binding affinity and

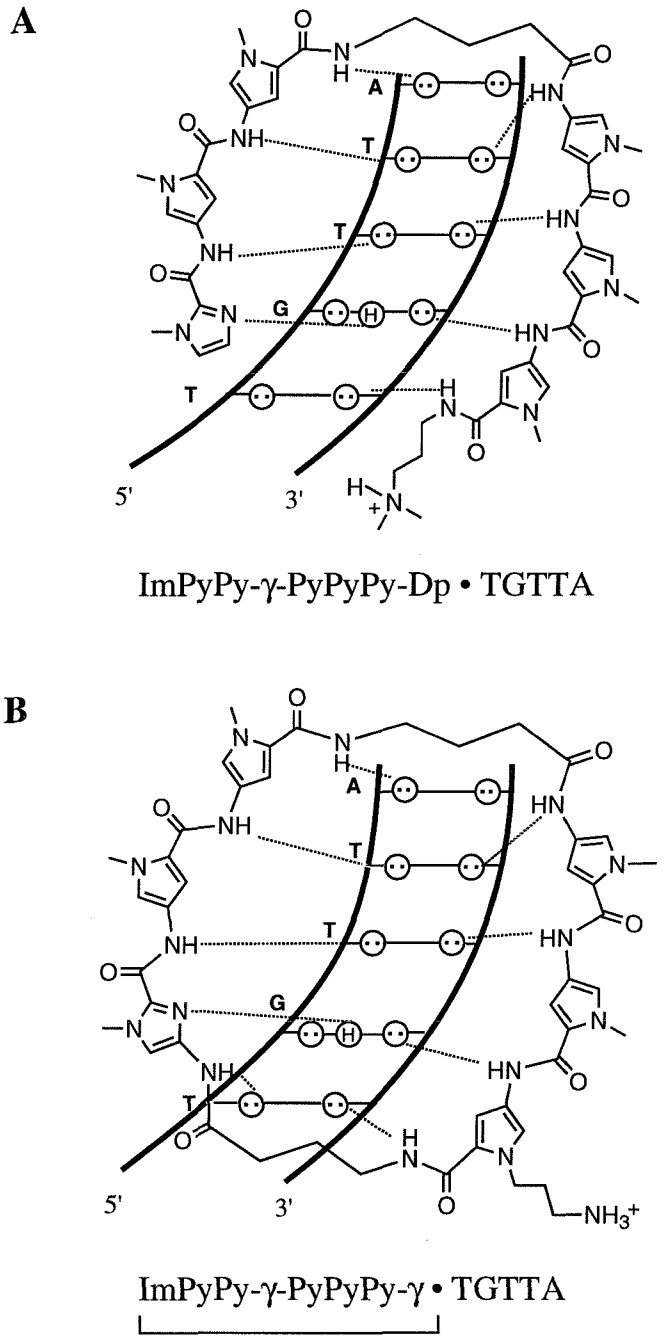


Figure 1. (A) ImPyPy-γ-PyPyPy-Dp **1** and (B) the cyclic polyamide *cyclo*-(ImPyPy-γ-PyPyPy-γ) **2** with a 5'-TGTTA-3' sequence. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines and circles containing an H represent the 2-amino group of guanine. Putative hydrogen bonds are illustrated by dashed lines.

specificity of this new class of cyclic DNA-binding small molecules (Figure 1).

Materials and Methods

General. ^{1}H NMR spectra were recorded at 300 MHz on a GE 300 NMR in $CDCl_3$, DMSO-d_6, or CD_3OD. Chemical shifts are reported in parts per million relative to tetramethylsilane or residual DMSO-d_5. IR spectra were recorded on a Perkin-Elmer FTIR spectrometer. High-resolution mass spectra (HRMS) were recorded using fast atom bombardment (FAB) techniques at the Mass Spectrometry Laboratory at the University of California, Riverside. Preparatory HPLC was carried out on a Beckman Instrument using a Waters DeltaPak 25 x 100 mm C_{18} column. Analytical HPLC was performed on a Hewlett Packard 1090 Series II analytical HPLC using a Vydac C_{18} reverse phase column (0.46 x 25 cm, 5 mm, HS silica). Flash column chromatography was carried out using silica gel 60 (230-400 mesh, Merck). Thin-layer chromatography (TLC) was performed on silica gel 60 F_{254} precoated plates (Merck). Chemicals for the syntheses were purchased from Aldrich unless otherwise specified. Dichloromethane and *N,N*-dimethylformamide (DMF) were purchased as anhydrous solvents from Aldrich. Boc-GABA and CbZ-GABA were obtained from Sigma. All enzymes were purchased from Boehringer Mannheim and used with the buffers supplied, with the exception of *Bsr*BI which was obtained from New England Biolabs. Deoxyadenosine 5'-[α-^{32}P] triphosphate and thymidine 5'-[α-^{32}P] triphosphate were obtained form Amersham. Storage phosphor technology autoradiography was performed using a Molecular Dynamics 400S Phosphorimager and ImageQuant software.

Boc-protected cyclic polyamide 4. To a solution of polyamide **3** (150 mg, 0.18 mmol) in DMF (70 mL) was added sequentially $NaHCO_3$ (78 mg, 0.93 mmol) and diphenylphosphoryl azide (DPPA, 154 mg, 0.56 mmol). The mixture was allowed to stir

at room temperature for 3 days and then filtered. The filtrate was concentrated in vacuo and the residue was purified by flash column chromatography (15% MeOH in CH_2Cl_2) to afford the cyclic polyamide **4** (90 mg, 58%) ^{1}H NMR [$CDCl_3$-CD_3OD (1:1)] δ 7.30 (s, 1H), 7.26 (s, 1H), 7.18 (s, 1H), 7.02 (s, 2H), 6.88 (s, 1H), 6.70 (s, 1H), 6.68 (s, 1H), 6.62 (s, 1H), 6.58 (s, 1H), 6.43 (s, 1H), 4.35 (m, 2H), 3.92 (s, 6H), 3.90 (s, 3H), 3.89 (s, 3H), 3.82 (s, 3H), 3.49 (m, 4H), 3.08 (m, 2H), 2.42 (m, 4H), 1.93-2.15 (m, 6H), 1.45 (s, 9H); IR (KBr) 3422 (s), 1654 (s), 1545 (s), 1438 (s), 1406 (m), 1260 (m), 1208 (m), 1105 (w), 919 (w), 777 (w) cm^{-1}; FABMS m/e 1047.4884 (M+H, 1047.4913 calcd. for $C_{50}H_{62}N_{16}O_{10}$).

Cyclic polyamide 2. To a mixture of trifluoroacetic acid (TFA) and thiophenol (4 mL, 0.5 M PhSH in TFA) was added the Boc-protected cyclic polyamide **4** (30 mg, 28.6 mmol) and the reaction mixture was allowed to stir at room temperature for 2 h. Excess TFA was removed in vacuo and the residue was purified by flash column chromatography (5% NH_4OH in MeOH) to afford cyclic polyamide **2** (23 mg, 87%). The product was purified by using reverse-phase HPLC on a preparatory Waters DeltaPak 25 x 100 mm C_{18} column with linear gradients of 60% acetonitrile plus 0.1% TFA versus 0.1% aqueous TFA. ^{1}H NMR (DMSO-d_6) δ 10.00 (s, 1H), 9.99 (s, 1H), 9.96 (s, 2H), 9.95 (s, 2H), 9.87 (s, 1H), 7.47 (s, 2H), 7.44 (s, 1H), 7.39 (s, 1H), 7.34 (s, 1H), 7.16 (s, 1H), 6.94 (s, 1H), 6.90 (s, 1H), 6.88 (s, 1H), 6.71 (s, 1H), 6.65 (s, 1H), 4.29 (m, 2H), 3.94 (s, 3H), 3.84 (s, 3H), 3.83 (s, 3H), 3.82 (s, 3H), 3.79 (s, 3H), 3.22 (brs, 4H), 2.52 (m, 2H), 2.39 (m, 2H), 2.30 (t, 2H, J=7.0Hz), 1.78 (m, 6H); IR 3383 (s), 1637 (s), 1560 (s), 1438 (s), 1407 (s), 1265 (m), 1207 (w), 1122 (w), 1062 (w), 777 (w) cm^{-1}; UV (H_2O) λmax (e) 240 (37,075 $cm^{-1}M^{-1}$), 310 (48,096 $cm^{-1}M^{-1}$) nm; FABMS m/e 947.4330 (M+H, 947.4389 calcd. for $C_{45}H_{54}N_{16}O_8$).

Preparation of Labeled DNA. The 135 base pair 3' end labeled *Eco*RI/*Bsr*BI restriction fragment from plasmid pMM5 was prepared and purified as follows. Plasmid DNA was linearized using *Eco*RI with simultaneous fill in by Klenow, deoxyadenosine 5'-[α-^{32}P] triphosphate, and thymidine 5'-[α-^{32}P] triphosphate. The linearized plasmid DNA was digested with *Bsr*BI and the 135 base pair *Eco*RI/*Bsr*BI restriction fragment was isolated by nondenaturing 5% polyacrylamide gel electrophoresis (PAGE). The gel bands were visualized by autoradiography, isolated, and filtered to remove the polyacrylamide. The resulting solution was further purified by phenol extraction followed by ethanol precipitation. The plasmid pJT3 *Afl*II/*Fsp*I 281 base pair fragment was prepared analogously. Chemical sequencing reactions were performed according to published protocols (17-18). Standard protocols were used for all DNA manipulations (19).

Quantitative DNase I Footprint Titration. All reactions were carried out in a total volume of 40 µL with final concentrations of each species as indicated. The polyamide ligands, ranging in concentration from 10 pM to 100 nM, were added to solutions of radiolabeled restriction fragment (10,000 cpm), tris•HCl (10 mM, pH 7.0), KCl (10 mM), $MgCl_2$ (10 mM) and $CaCl_2$ (5 mM) and incubated for 24 h at 22 °C. We explicitly note that no carrier calf thymus DNA is used in these reactions. Footprinting reactions were initiated by the addition of 4 µL of a stock solution of DNase I (0.03 units/mL) containing 1 mM dithiothreitol and allowed to proceed for 7 min at 22 °C. The reactions were stopped by addition of a 3 M sodium acetate solution containing 50 mM EDTA and ethanol precipitated. The reactions were resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer and electrophoresed on 8% polyacrylamide denaturing gels (5% crosslink, 7 M urea) at 2000 V for 1 h. The footprint titration gels were dried and quantitated using storage phosphor technology. Apparent first-order association constants were determined as previously described (11).

Results and Discussion

A cyclic polyamide **2** containing eight amino acids was synthesized in 12 steps from readily available starting materials. The key cyclization reaction was accomplished in good yield (58%) using diphenylphosphoryl azide in the presence of $NaHCO_3$ followed by the deprotection with trifluoroacetic acid to give the cyclic product *cyclo*-(ImPyPy-γ-PyPyPy-γ) **2** which was purified by reverse-phase HPLC (Figure 2). The observed molecular mass of the $C_{45}H_{54}N_{16}O_8$ polyamide is 947.4333 (FABMS) in good agreement with the calculated value [M + H, 947.0256 calcd.] (20).

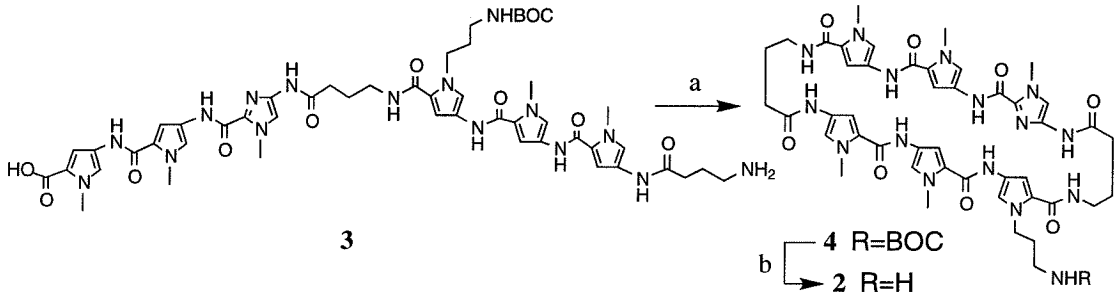


Figure 2. Synthetic scheme for cyclization to cyclic polyamide **2**. (a) DPPA, $NaHCO_3$, DMF; (b) TFA, thiophenol.

Quantitative DNase I footprint titration experiments (21-24) on a ^{32}P end-labeled 135-base pair restriction fragment were performed to obtain the binding affinity of the cyclic polyamide **2** for a designated match site, 5'-TGTTA-3' (Figure 3). The cyclic polyamide binds the 5'-TGTTA-3' site very tightly with an equilibrium association constant of $K \geq 2.9 \times 10^9\ M^{-1}$ (Figure 4). We attribute this increase of 40 in binding affinity over the corresponding hairpin polyamide **1** to a reduction in conformational entropy imparted via cyclization (Figure 5). The binding affinities to single base pair mismatch sites, 5'-ATTC**G**-3', 5'-AGA**G**T-3', and 5'-AGA**C**A-3' are lower, 7.4×10^8, 1.7×10^8, and $6.2 \times 10^8\ M^{-1}$, respectively (Table I). The cycle appears to bind the match over single base pair mismatch site by a factor of 4-17, although these values may be lower limits due to the uncertainty in the very high equilibrium association constant value for the cycle **2**:TGTTA

Figure 3. Sequence of the 135 base pair *Eco*RI/*Bsr*BI restriction fragment. Several five base pair sites were analyzed by quantitative footprint titration analysis. Note that there are additional binding sites on the upper half of the autoradiogram (Figure 4) of sequence composition (A,T)G(A,T)$_3$ which are also occupied by the cyclic polyamide.

complex. The hairpin binds more strongly to a match over single mismatch sites by a factor of twenty. The cyclic polyamide, though higher in affinity, is less than or only equal to the hairpin in specificity. It may be that there is an energetic price in specificity for closing the ends of the hairpin. Nevertheless, it is an encouraging step forward that wholly designed synthetic cyclic polyamides with a molecular weight of 950 bind designated five base pair sequences at subnanomolar concentrations. It remains a challenge to further optimize specificity.

Table I Apparent First-Order Association Constants (M^{-1})*,†

	Match Site	**Single Base Mismatch Sites**		
Polyamide	5'-TGTTA-3'	5'-ATTC**G**-3'	5'-AGA**G**T-3'	5'-AGA**C**A-3' §
Hairpin **1**	7.6×10^7 (0.8)	‡	7.8×10^5 (2.0)	2.6×10^6 (0.9)
Cyclic **2**	**2.9×10^9** (1.9)	7.4×10^8 (2.4)	1.7×10^8 (1.4)	6.2×10^8 (2.9)

*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.

†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 $CaCl_2$.

‡Not determined due to the fact that hairpin **1** binds neighboring 5'-TGACA-3' single mismatch site.

§Determined on different restriction fragment (pJT3 *Afl*III/*Fsp*I).

Figure 4. Quantitative DNase I footprint titration experiment with the cyclic polyamide **2** on the 3'-^{32}P-labeled 135 base pair *Eco*RI/*Bsr*BI restriction fragment from plasmid pMM5: lane 1, A reaction; lane 2, G reaction; lanes 3 and 17, DNase I standard; lanes 4-16, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM and 100 nM cyclic polyamide, respectively; lane 18, intact DNA. The 5'-TGTTA-3', 5'-ATTCG-3', and 5'-AGAGT-3' binding sites which were analyzed are shown on the right side of the autoradiogram. All reactions contain 10 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

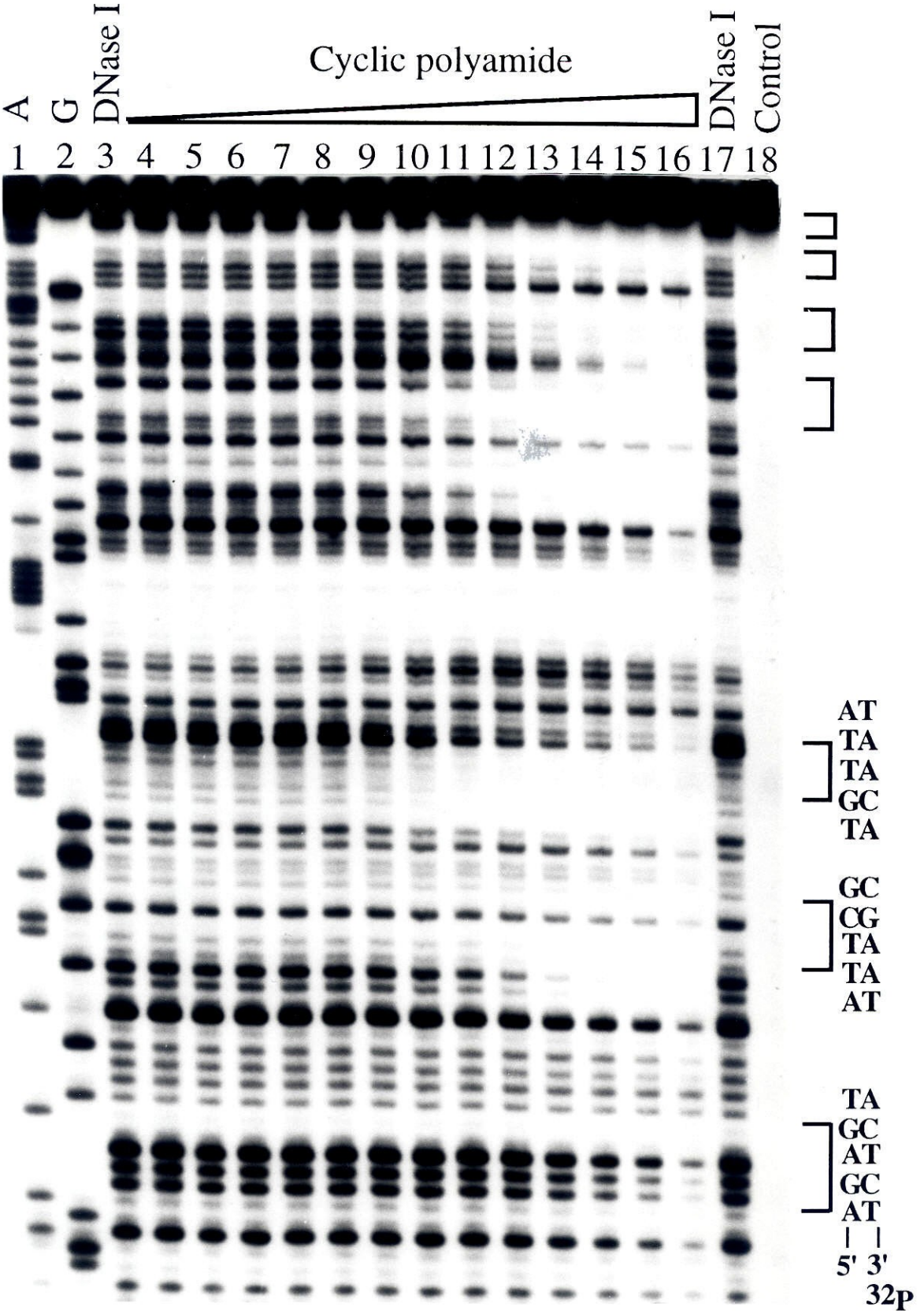
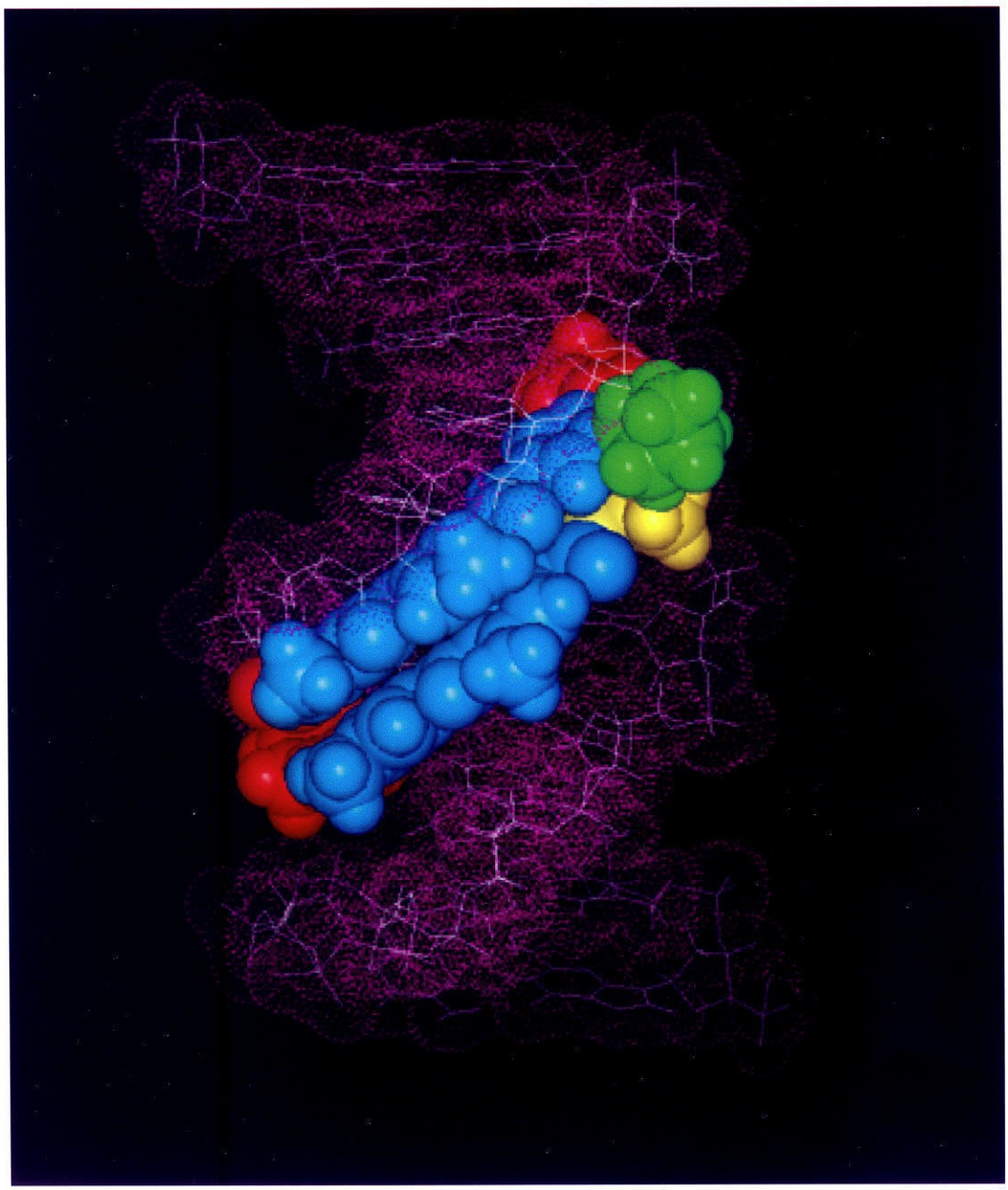

A
G
DNase I
Cyclic polyamide
DNase I
Control
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
AT
TA
TA
GC
TA
GC
CG
TA
TA
AT
TA
GC
AT
GC
AT
5'
3'
32P

Figure 5. Proposed model of the cyclic polyamide binding in the minor groove of DNA. Colors represent the following amino acids: cyan, Py; yellow, Im; magenta, γ; green, Dp side chains.



References

1. Pelton, J. G. & Wemmer, D. E. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5723-5727.
2. Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E. & Dervan, P. B. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7586-7590.
3. Wade, W. S., Mrksich, M. & Dervan, P. B. (1992) *J. Am. Chem. Soc.* **114**, 8783-8794.
4. Wade, W. S., Mrksich, M. & Dervan, P. B. (1993) *Biochemistry* **32**, 11385-11389.
5. Mrksich, M. & Dervan, P. B. (1993) *J. Am. Chem. Soc.* **115**, 2572-2576.
6. Geierstanger, B. H., Dwyer, T. J., Bathini, Y., Lown, J. W. & Wemmer, D. E. (1993) *J. Am. Chem. Soc.* **115**, 4474-4482.
7. Dwyer, T. J., Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1993) *J. Am. Chem. Soc.* **115**, 9900-9906.
8. Mrksich, M. & Dervan, P. B. (1993) *J. Am. Chem. Soc.* **115**, 9892-9899.
9. Mrksich, M. & Dervan, P. B. (1994) *J. Am. Chem. Soc.* **116**, 3663-3664.
10. Geierstanger, B. H., Jacobsen, J. P., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1994) *Biochemistry* **33**, 3055.
11. Mrksich, M., Parks, M. E. & Dervan, P. B. (1994) *J. Am. Chem.* Soc. **116**, 7983-7988.
12. Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1994) *Science* **266**, 646-650.
13. Mrksich, M. & Dervan, P. B. (1995) *J. Am. Chem. Soc.* **117**, 3325.
14. Pelton, J. G. & Wemmer, D. E. (1990) *J. Am. Chem. Soc.* **112**, 1393-1399.
15. Fagan, P. A. & Wemmer, D. E. (1992) *J. Am. Chem. Soc.* **114**, 1080-1081.

16. Chen, X., Ramakrishnan, B., Rao, S. T. & Sundaralingam, M. (1994) *Nature Struct. Biology* **1**, 169-175.

17. Iverson, B. L. & Dervan P.B. (1987) *Nucleic Acids Res.* **15**, 7823-7830.

18. Maxam, A. M. & Gilbert, W. S. (1980) *Methods in Enzymology* **65**, 499-560.

19. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning*, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY.

20. We thank UC Riverside Mass Spec Facility for the analysis.

21. Brenowitz, M., Senear, D. F., Shea, M. A. & Ackers, G. K. (1986) *Methods Enzymol.* **130**, 132-181.

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

23. Senear, D. F., Brenowitz, M., Shea, M. A. & Ackers, G. K. (1986) *Biochemistry* **25**, 7344-7354.

24. The quantitative footprint titration experiments and data analysis were performed as previously described (11), with unlabeled carrier absent from all reactions.

CHAPTER FOUR

Expansion of the Targetable Sequence Repertoire for Recognition of the Minor Groove of DNA by Pyrrole-Imidazole Polyamides

Introduction

Recent 2:1 polyamide-DNA complexes have created new models for the design of nonnatural ligands for sequence-specific recognition in the minor groove of DNA.[1-4] An imidazole-pyrrole-pyrrole polyamide, 1-methylimidazole-2-carboxamide-netropsin (ImPyPy-Dp), was shown to specifically bind the mixed sequence 5'-(A,T)G(A,T)C(A,T)-3' as a side-by-side antiparallel homodimer.[2] ImPyPy-Dp and distamycin were shown to bind to a 5'-TGTTA-3' sequence as an antiparallel heterodimer.[3] The 2:1 model allows each polyamide to form specific contacts with a single strand of DNA in the minor groove.[1-4] The side-by-side combination of an imidazole on one ligand and a pyrrolecarboxamide on the second ligand recognizes G•C, while a pyrrolecarboxamide/imidazole pair targets a C•G base pair.[2-4] A pyrrolecarboxamide /pyrrolecarboxamide combination is partially degenerate and binds to either A•T or T•A base pairs.[1-4]

Although specific recognition of the sequences 5'-(A,T)G(A,T)C(A,T)-3',[2] 5'-(A,T)G$(A,T)_3$-3',[3] 5'-$(A,T)_2$G$(A,T)_2$-3',[4] and 5'-(A,T)GCGC(A,T)-3'[5] has been achieved, the generality of the 2:1 model has not been explored systematically. Eight three-ring polyamides representing all possible combinations of imidazole-and pyrrole-carboxamides were synthesized or purchased (Figure 1). For ease of synthesis, we chose to study the combinations of polyamides as homo- or hetero-dimers rather than hairpin polyamides. For combinations of monomeric polyamides that result in new specificities, the corresponding hairpin could subsequently be synthesized.

In collaboration with Susanne Swalley, who synthesized AcPyImIm-Dp and AcImImIm-Dp, polyamides **1-8** were synthesized or obtained (**1** is the natural product distamycin A purchased from Sigma). Predicted specificities based on the 2:1 model are shown in Figure 2. In order to test these predictions, binding site specificities were determined by MPE•Fe(II) footprinting for all combinations of homo- and heterodimers

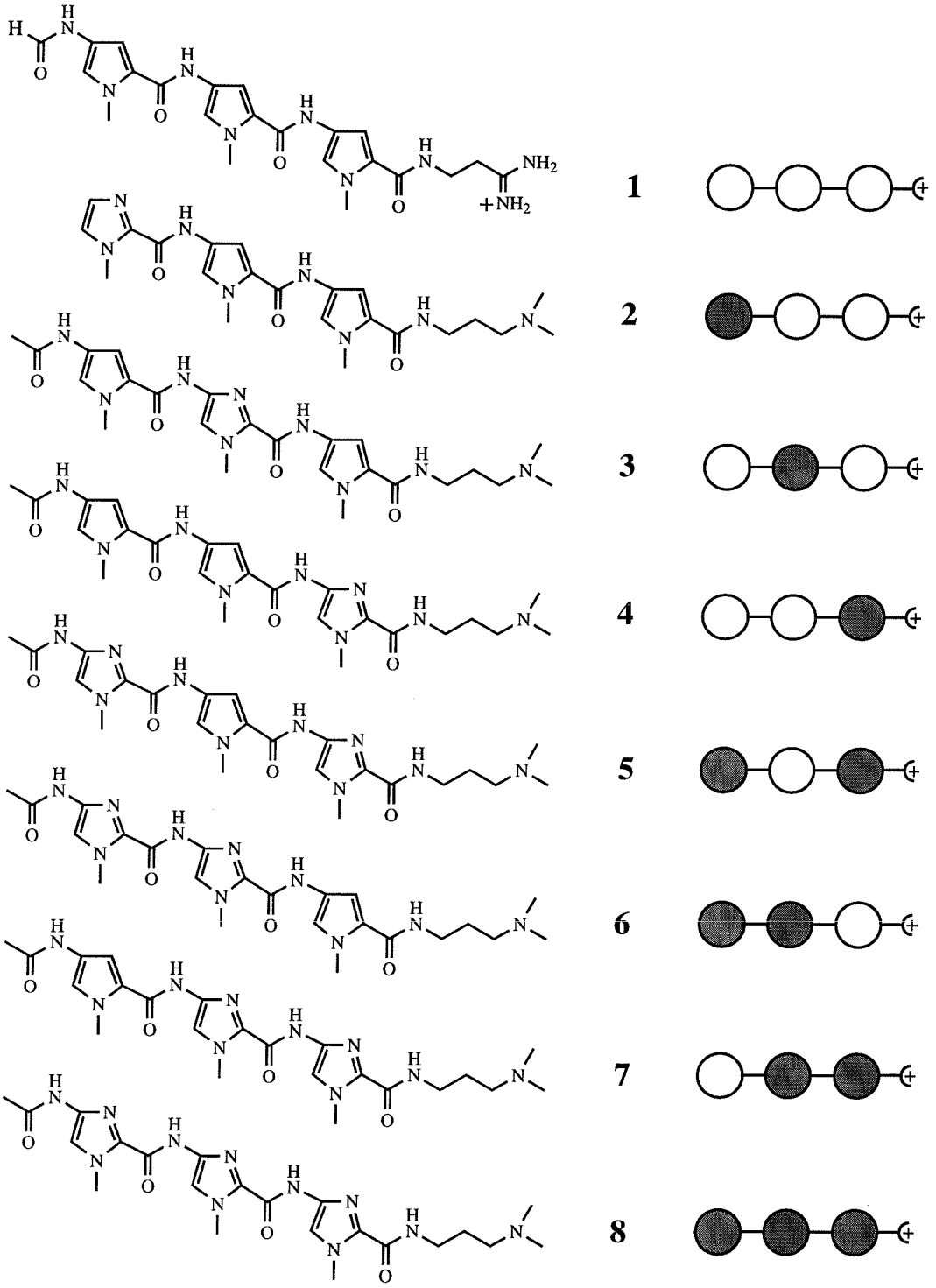


Figure 1. Chemical structures of polyamides **1-8** containing all combinations of 0-3 imidazole rings. Ball and stick models of the polyamides are shown to the right of each polyamide. Nonshaded and shaded circles represent pyrrole and imidazole carboxamides, respectively. Polyamides **7** and **8** were synthesized by Susanne Swalley.

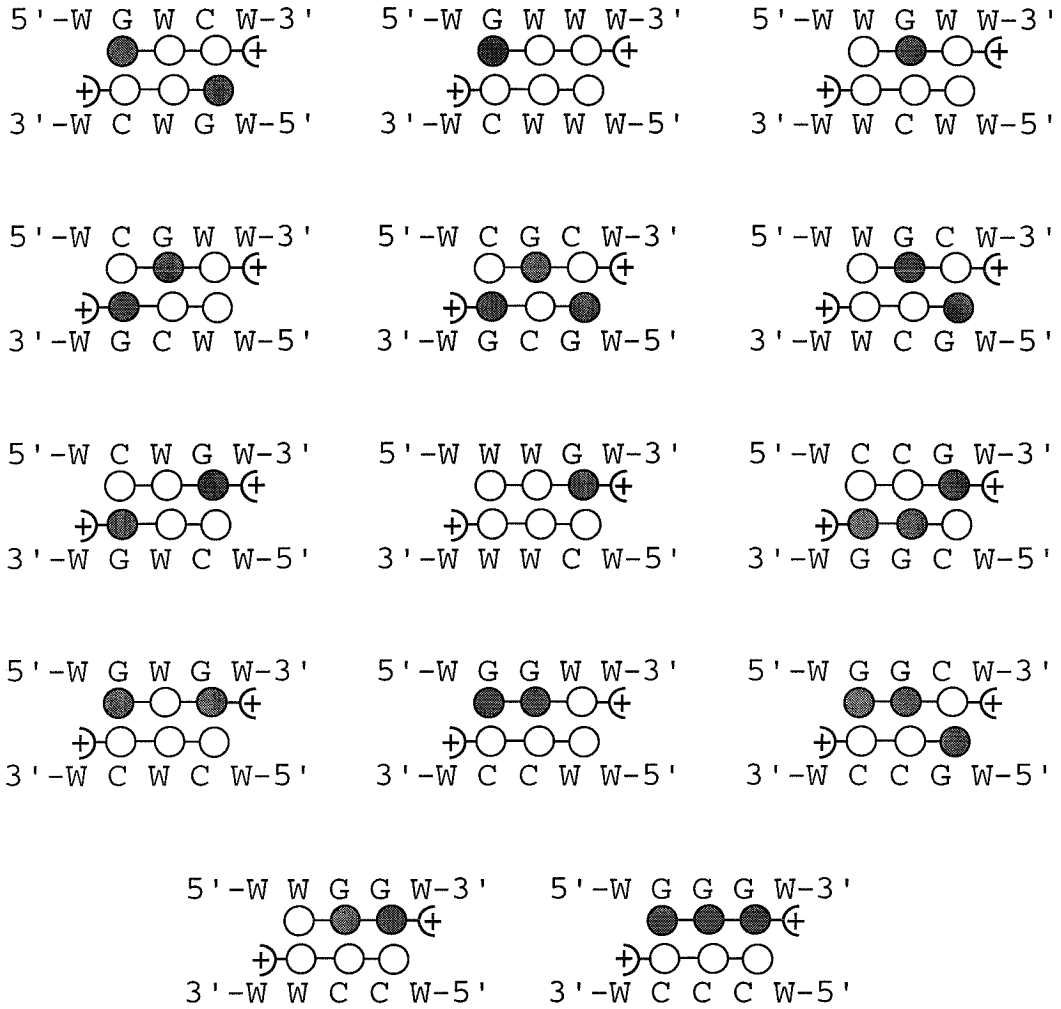


Figure 2. Ball and stick representations of all possible predicted binding combinations for polyamides **1-8** based on the current 2:1 model. The top row of sites represent sequences previously recognized. The bottom row of sites were studied by Susanne Swalley and will not be reported here. All other sites and combinations of polyamides represented here were examined by MPE•Fe(II) footprinting on either the 167- or 517-base pair *Eco*RI/*Rsa*I restriction fragment from pBR322. Nonshaded and shaded circles represent pyrrole and imidazole rings, respectively. W represents either A or T.

on either the 167 or 517 base pair *Eco*RI/*Rsa*I restriction fragment from pBR322. The experiments described here systematically examine the generality of the 2:1 model for DNA recognition by pyrrole-imidazole polyamides.

Results

Polyamide Synthesis. Polyamide ImPyPy-Dp **2** was synthesized as previously described.[2b] Polyamides **3-6** were synthesized in four steps from the previously described intermediates *N*-methyl-4-nitropyrrole-2-carboxylic acid **9**[6] and *N*-methyl-4-nitroimidazole-2-carboxylic acid **14**[7] (Figure 3). A representative synthesis begins with the reaction of **9** with thionyl chloride followed by coupling of dimethylaminopropylamine to form 3-[*N*-methyl-4-nitropyrrole-2-carboxamido]-dimethylpropylamine **10**. Subsequent reduction of **10** (300 psi H_2, Pd/C) to the amine derivative followed by coupling to the acid chloride of **14** yields nitro-PyImPy-Dp **12**. Reduction of **12** (350 psi H_2, Pd/C) and capping of the resulting amine with acetic anhydride in the presence of pyridine provides the desired polyamide AcPyImPy-Dp **3**. Polyamides **4-6** were prepared analogously.

Determination of Binding Sites by MPE•Fe(II) Footprinting. The binding site sizes and specificities of combinations of polyamides **1-8** were determined by MPE•Fe(II) footprinting on either the 167 or 517 base pair *Eco*RI/*Rsa*I restriction fragment derived from pBR322 (Boehringer Mannheim).[8] Each combination in Figure 2 was examined beginning with AcPyImPy-Dp **3**/distamycin **1**, which was previously studied by NMR, and ending with AcImImPy-Dp **6**/ImPyPy-Dp **2** moving in order from left to right.

The first heterodimer, AcPyImPy-Dp **3**/distamycin **1**, displayed the expected specificity, binding a 5'-TAGTT-3' sequence with the highest affinity of the sequences analyzed (Figure 4). However, the heterodimer pair, AcPyImPy-Dp **3**/AcPyPyIm-Dp **4** did not bind as predicted to a sequence of 5'-WCGWW-3' composition, instead giving a large footprint with no clearly defined site (Figure 5). An AcPyImPy-Dp **3**/AcImPyIm-Dp **5** combination binds as predicted to the site 5'-ACGCA-3' (Figure 6). A single base

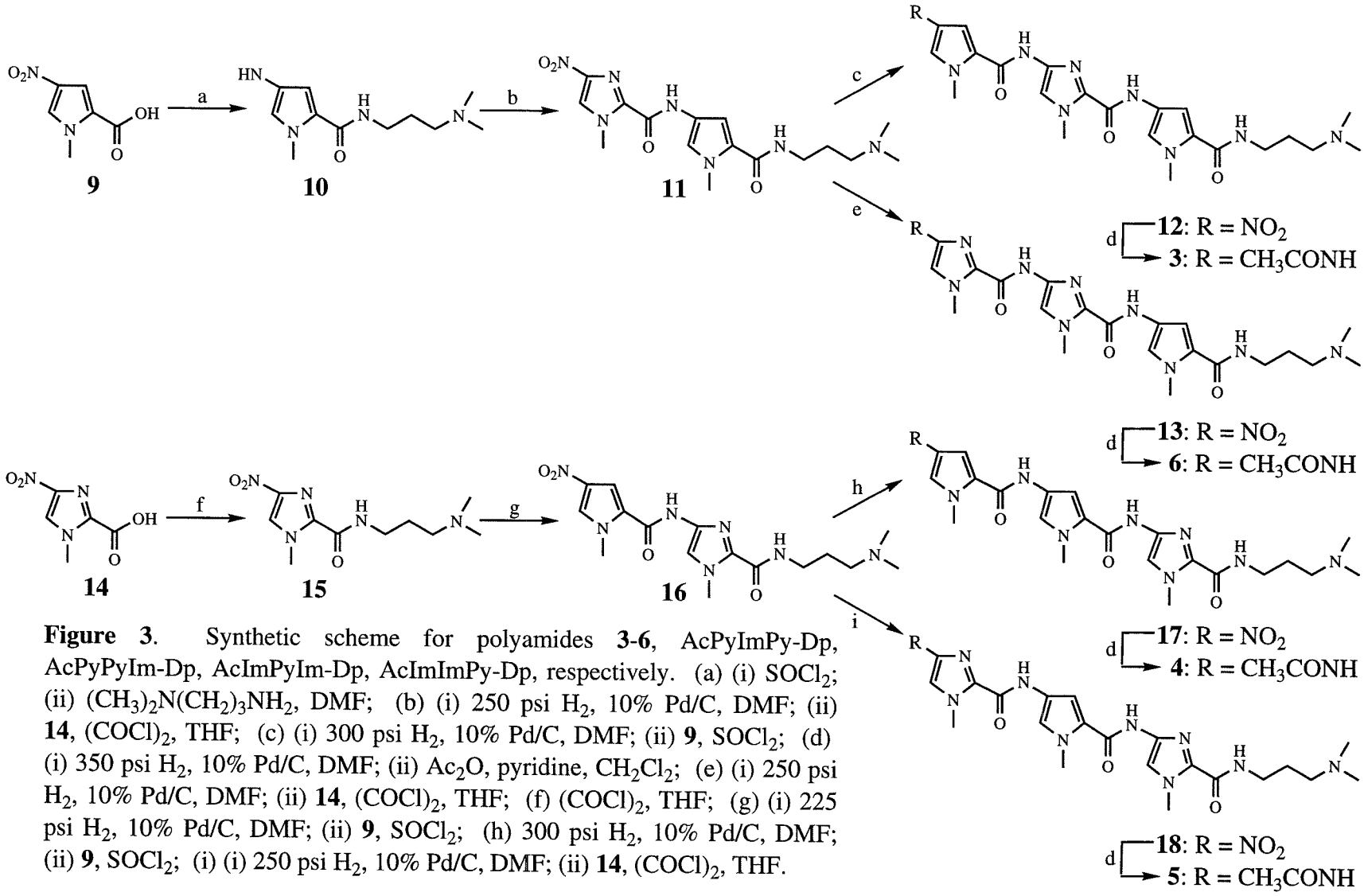
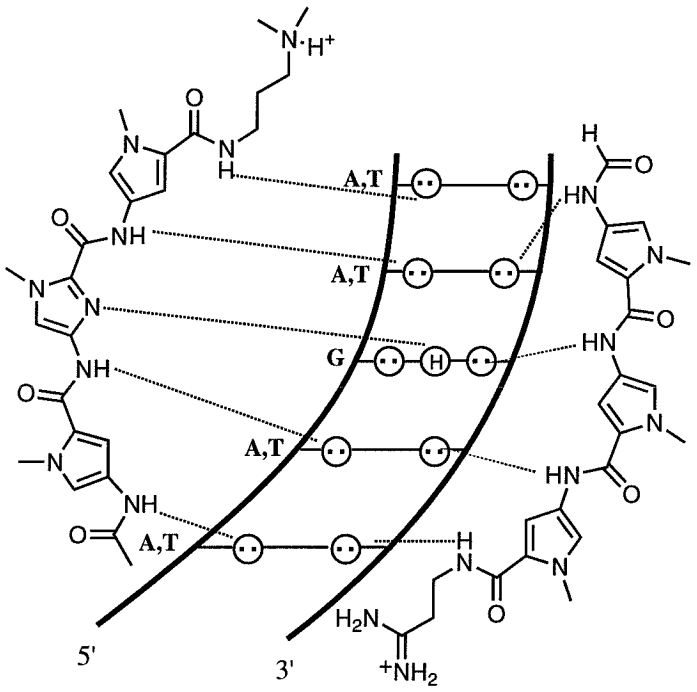
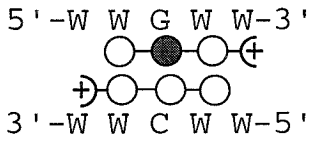


Figure 3. Synthetic scheme for polyamides **3-6**, AcPyImPy-Dp, AcPyPyIm-Dp, AcImPyIm-Dp, AcImImPy-Dp, respectively. (a) (i) $SOCl_2$; (ii) $(CH_3)_2N(CH_2)_3NH_2$, DMF; (b) (i) 250 psi H_2, 10% Pd/C, DMF; (ii) **14**, $(COCl)_2$, THF; (c) (i) 300 psi H_2, 10% Pd/C, DMF; (ii) **9**, $SOCl_2$; (d) (i) 350 psi H_2, 10% Pd/C, DMF; (ii) Ac_2O, pyridine, CH_2Cl_2; (e) (i) 250 psi H_2, 10% Pd/C, DMF; (ii) **14**, $(COCl)_2$, THF; (f) $(COCl)_2$, THF; (g) (i) 225 psi H_2, 10% Pd/C, DMF; (ii) **9**, $SOCl_2$; (h) 300 psi H_2, 10% Pd/C, DMF; (ii) **9**, $SOCl_2$; (i) (i) 250 psi H_2, 10% Pd/C, DMF; (ii) **14**, $(COCl)_2$, THF.

Figure 4. (Top) Binding model for the complex formed between polyamide AcPyImPy-Dp **3** and distamycin **1** with a 5'-WWGWW-3' sequence. 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. (Middle) Ball and stick model of predicted AcPyImPy-Dp/distamycin heterodimer. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. (Bottom) MPE•Fe(II) histograms of cleavage protection (footprinting) data for polyamide **3** (10 μM) in the presence and absence of distamycin **1** (2 μM) on the *Eco*RI/*Rsa*I 167 base pair restriction fragment from pBR322. Bar heights are proportional to the relative protection from cleavage at each band. Box represents expected binding site.

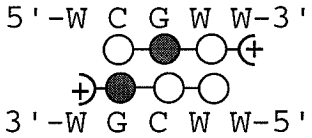




AcPyImPy-Dp **3** (10μM) / distamycin **1** (2μM)

AcPyImPy-Dp **3** (10μM)

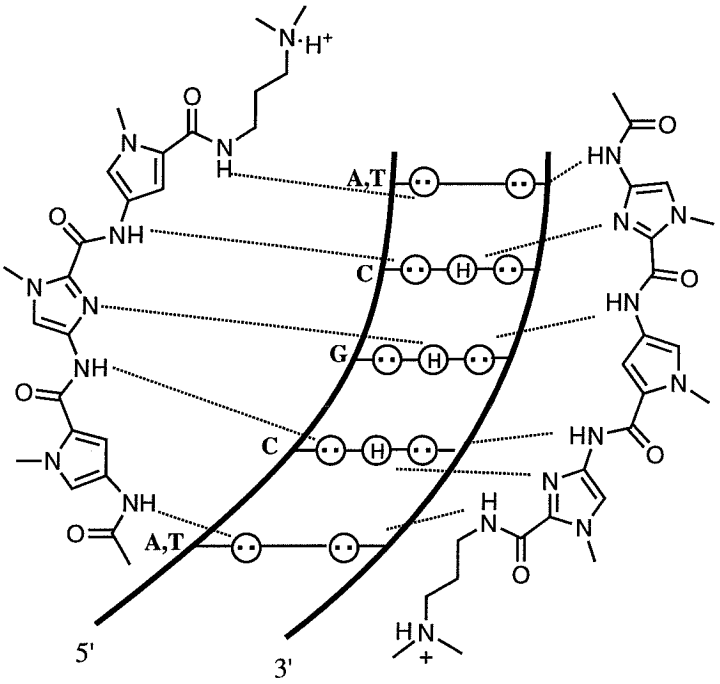
pair mismatch 5'-AGGCA-3' is also bound although with reduced affinity. The final possible pairing with polyamide **3** is to form a heterodimer with ImPyPy-Dp **2** which does not bind the predicted 5'-WWGCW-3' sequence (Figure 7). Instead we observe protection over a large area of the DNA, making assignment of binding sites ambiguous.

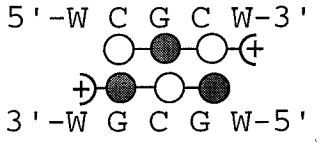


AcPyImPy-Dp **3** (10μM)/AcPyPyIm-Dp **4** (10μM)

Figure 5. Ball and stick model of predicted AcPyImPy-Dp/AcPyPyIm-Dp heterodimer. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. MPE•Fe(II) histograms of cleavage protection (footprinting) data for polyamides **3** and **4** at 10 μM concentration on the *Eco*RI/*Rsa*I 167 base pair restriction fragment from pBR322. Bar heights are proportional to the relative protection from cleavage at each band. Box represents expected heterodimer binding site.

Polyamide AcPyPyIm-Dp **4** does not recognize the predicted site 5'-WCWGW-3' as a homodimer, nor does the heterodimeric combination of AcPyPyIm-Dp **4**/distamycin **1** recognize its target sequence of 5'-WWWGW-3' composition (Figures 8 and 9). The sites recognized with highest affinity for the two systems are 5'-AGCTT-3' and 5'-AGTTT-3', respectively. In addition, the heterodimer pair, AcPyPyIm-Dp **4** and AcPyImIm-Dp **7**, binds most strongly to mismatched sequences and shows no detectable affinity for the match site (Figure 10). Polyamide AcImPyIm-Dp **5** binding as a heterodimer with distamycin **1** does not target the desired match site 5'-WGWGW-3', but instead recognizes a 5'-TGCGCT-3' site with the highest affinity (Figure 11).





AcPyImPy-Dp **3** (10μM)/AcImPyPy-Dp **5** (10μM)

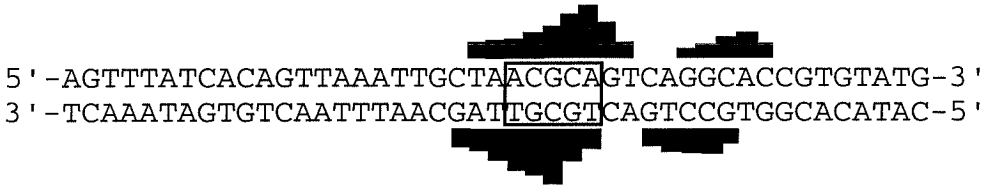
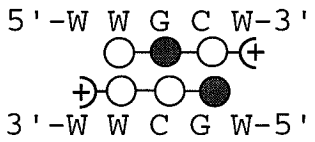


Figure 6. (Top) Binding model for the complex formed between polyamides AcPyImPy-Dp **3** and AcImPyImP-Dp **5** with a 5'-WCGCW-3' sequence. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. (Middle) Ball and stick model of predicted AcPyImPy-Dp/AcImPyImP-Dp heterodimer. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. (Bottom) MPE•Fe(II) histograms of cleavage protection (footprinting) data for polyamide **3** (10 μM) in the presence of **5** (10 μM) on the *Eco*RI/*Rsa*I 167 base pair restriction fragment from pBR322. Bar heights are proportional to the relative protection from cleavage at each band. Box represents expected binding site.



AcPyImPy-Dp **3** (20μM) / ImPyPy-Dp **2** (100μM)

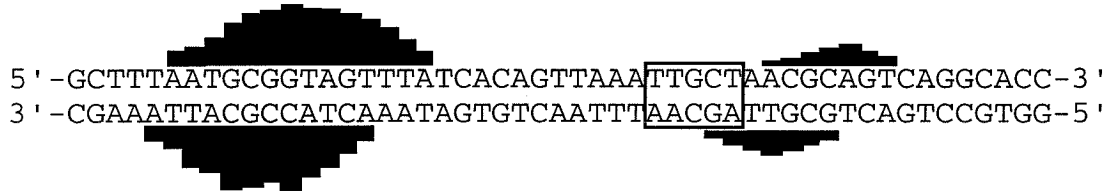
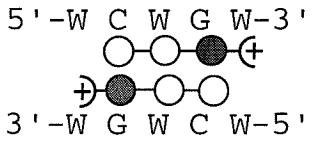


Figure 7. Ball and stick model of predicted AcPyImPy-Dp/ImPyPy-Dp heterodimer. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. MPE•Fe(II) histograms of cleavage protection (footprinting) data for polyamide **3** (20 μM) with **2** (100 μM) on the *Eco*RI/*Rsa*I 167 base pair restriction fragment from pBR322. Bar heights are proportional to the relative protection from cleavage at each band. Box represents expected binding site.



AcPyPyIm-Dp **4** (10μM)

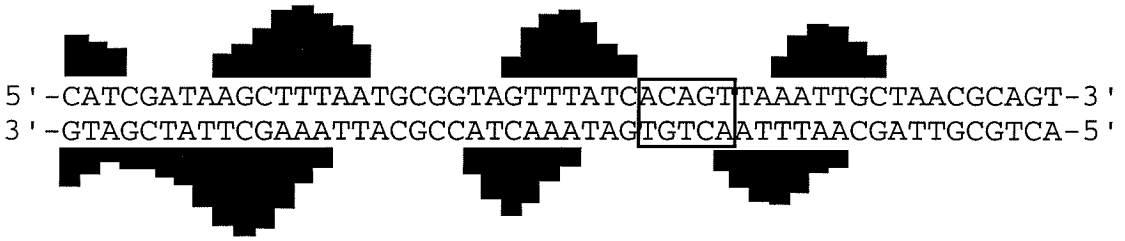
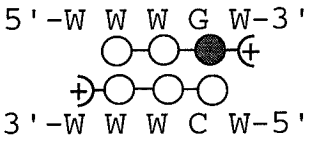


Figure 8. Ball and stick model of predicted AcPyPyIm-Dp homodimer. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. MPE•Fe(II) histograms of cleavage protection (footprinting) data for polyamide **4** (10 μM) on the *Eco*RI/*Rsa*I 167 base pair restriction fragment from pBR322. Bar heights are proportional to the relative protection from cleavage at each band. Box represents expected binding site.



AcPyPyIm-Dp **4** (10μM) / distamycin **1** (2μM)

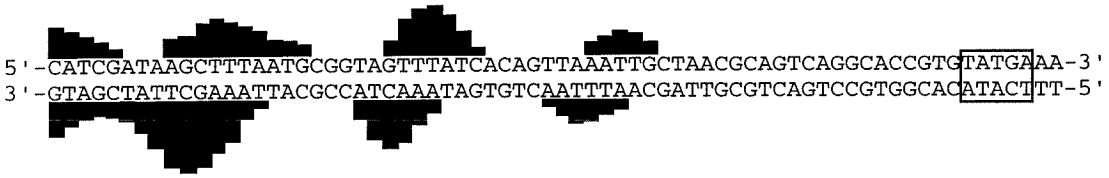
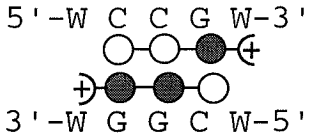


Figure 9. Ball and stick model of predicted AcPyPyIm-Dp/distamycin heterodimer. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. MPE•Fe(II) histograms of cleavage protection (footprinting) data for polyamide **4** (10 μM) with distamycin **1** (2 μM) on the *Eco*RI/*Rsa*I 167 base pair restriction fragment from pBR322. Bar heights are proportional to the relative protection from cleavage at each band. Box represents expected binding site.



AcPyPyIm-Dp **4** (10μM) / AcPyImIm-Dp **7** (10μM)

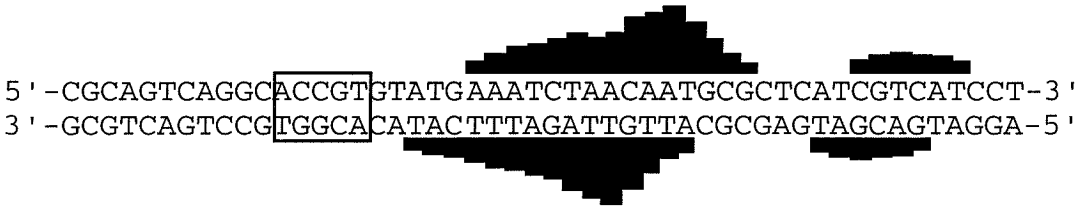


Figure 10. Ball and stick model of predicted AcPyPyIm-Dp/AcPyImIm-Dp heterodimer. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. MPE•Fe(II) histograms of cleavage protection (footprinting) data for polyamide **4** (10 μM) with **7** (10 μM) on the *Eco*RI/*Rsa*I 167 base pair restriction fragment from pBR322. Bar heights are proportional to the relative protection from cleavage at each band. Box represents expected binding site.

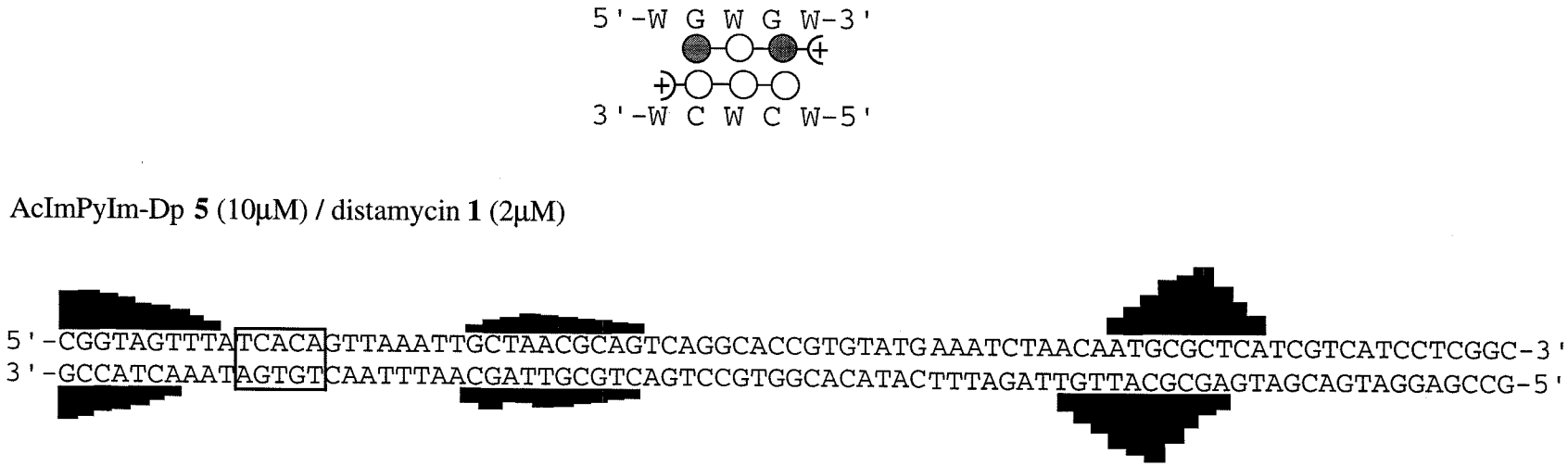


Figure 11. Ball and stick model of predicted AcImPyIm-Dp/distamycin heterodimer. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. MPE•Fe(II) histograms of cleavage protection (footprinting) data for polyamide **5** (10 μM) with distamycin **1** (2 μM) on the *Eco*RI/*Rsa*I 167 base pair restriction fragment from pBR322. Bar heights are proportional to the relative protection from cleavage at each band. Box represents expected binding site.

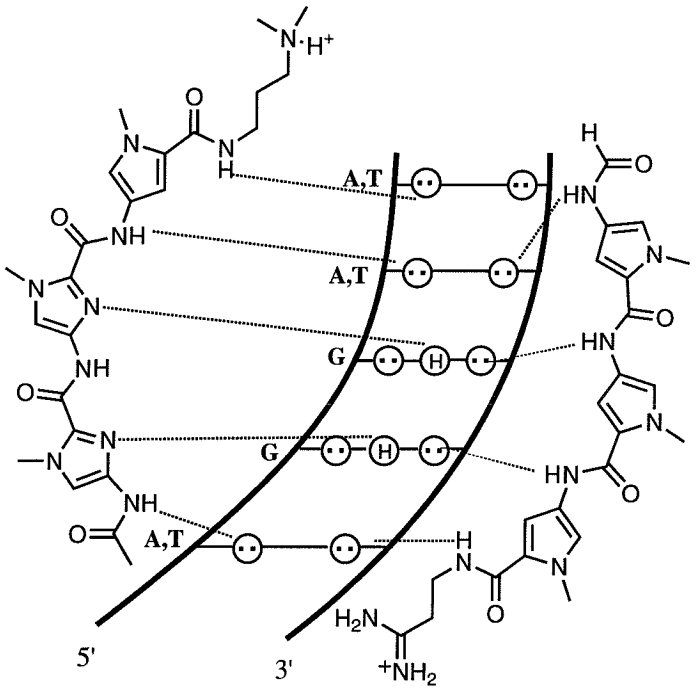
The final polyamide studied, AcImImPy-Dp **6**, binding as a heterodimer with distamycin **1** recognized both predicted match sites, 5'-AGGTT-3' and 5'-TGGTT-3', present on the 517 base pair restriction fragment (Figure 12). In addition, in the absence of distamycin **1**, the target sites were not bound indicating a dependence on distamycin. The binding orientation of this heterodimer was confirmed by affinity cleaving with EDTA-PyPyPy-Dp (Figure 13). Finally, the polyamide pair AcImImPy-Dp **6**/ ImPyPy-Dp **2** also recognizes the predicted heterodimer site 5'-AGGCA-3', although there is some overlap with another site (Figure 14).

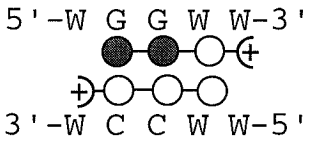
Discussion

Binding Site Specificities. The heterodimeric combination of polyamide AcPyImPy-Dp **3** with distamycin **1** binds to the expected site 5'-TAGTT-3' confirming previous NMR studies on the match sequence. Interestingly, polyamide **3** in the absence of distamycin **1** is still able to recognize the 5'-TAGTT-3' site, presumably as a mismatched homodimer. Our results are consistent with the mismatched homodimer being energetically less favorable than the fully matched complex as demonstrated by previous NMR studies.[9] Polyamide **3** is also able to form a heterodimeric complex with polyamide AcImPyIm-Dp **5** to recognize 5'-WCGCW-3' sequences. However, AcPyImPy-Dp **3** appears unable to recognize the predicted sequences, 5'-TCGAT-3' and 5'-TTGCT-3', which would result from heterodimers with AcPyPyIm-Dp **4** and ImPyPy-Dp **1**, respectively.

Surprisingly, polyamide AcPyPyIm-Dp **4** does not recognize any of the predicted match sequences either as a homodimer or heterodimer with distamycin **1**, AcPyImIm-Dp **7**, or AcPyImPy-Dp **3**. This deviation from the accepted 2:1 model may be due to the C-terminal position of the imidazole carboxamide. MPE•Fe(II) footprinting results for both AcPyImIm-Dp **7** and AcImImIm-Dp **8** in the presence of distamycin **1** support this

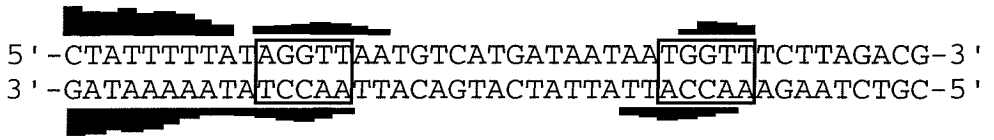
Figure 12. (Top) Binding model for the complex formed between polyamide AcImImPy-Dp **6** and distamycin **1** with a 5'-WGGWW-3' sequence. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. (Middle) Ball and stick model of predicted AcImImPy-Dp/distamycin heterodimer. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. (Bottom) MPE•Fe(II) histograms of cleavage protection (footprinting) data for polyamide **6** (10 μM) in the presence and absence of distamycin **1** (2 μM) on the *Eco*RI/*Rsa*I 517 base pair restriction fragment from pBR322. Bar heights are proportional to the relative protection from cleavage at each band. Boxes represent expected heterodimer binding sites.





AcImImPy-Dp **6** (10μM) / distamycin **1** (2μM)

AcImImPy-Dp **6** (10μM)



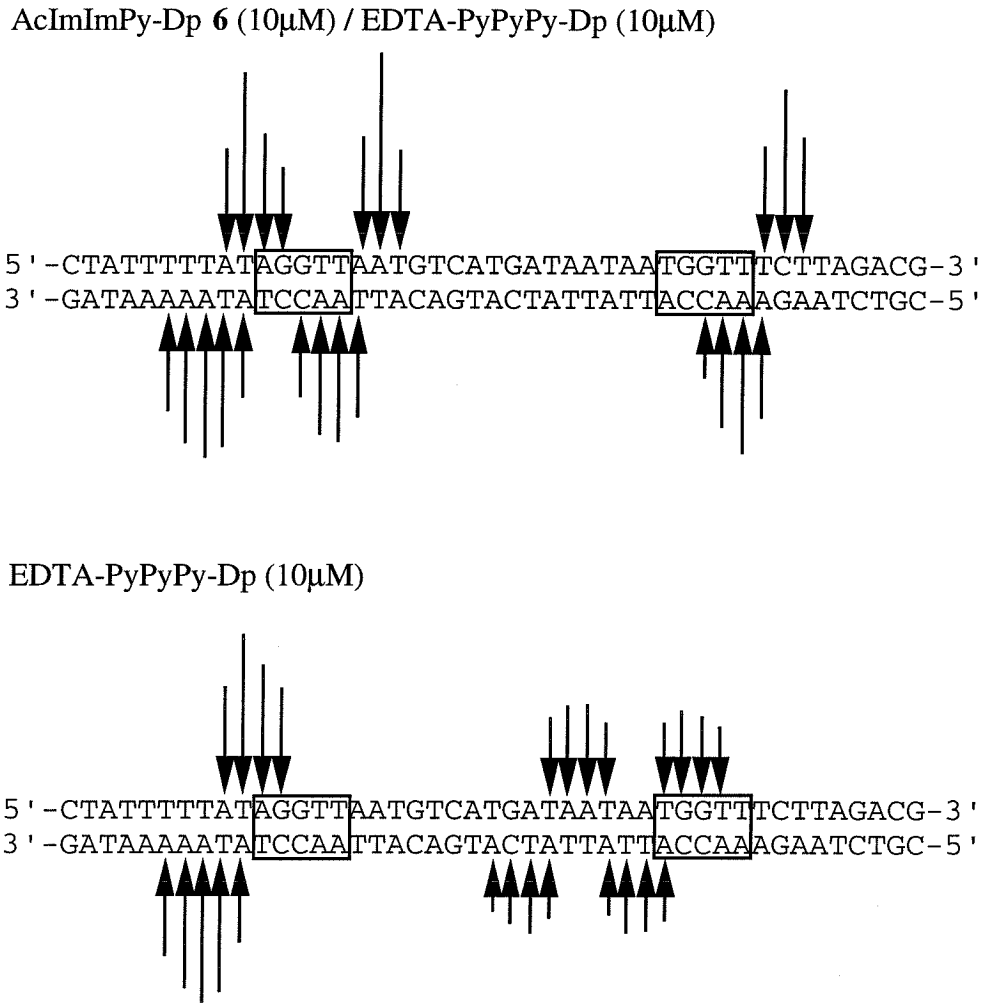
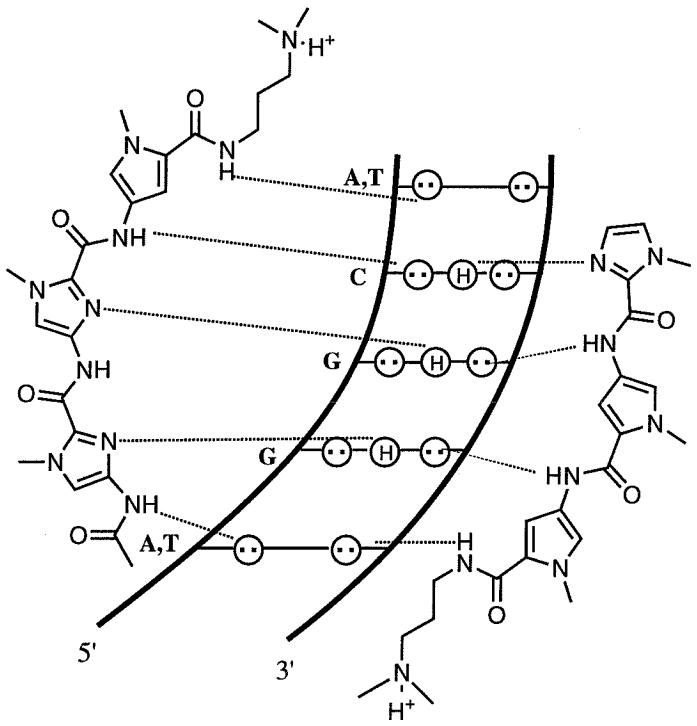
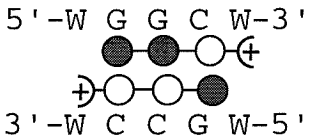


Figure 13. Affinity cleavage data for EDTA-PyPyPy-Dp (10μM) in the presence and absence of polyamide **6** (10 μM) on the *Eco*RI/*Rsa*I 517 base pair restriction fragment from pBR322. Arrow heights are proportional to the relative amount of cleavage at each band. Boxes represent expected heterodimer binding sites.





AcImImPy-Dp **6** (20μM) / ImPyPy-Dp **2** (100μM)

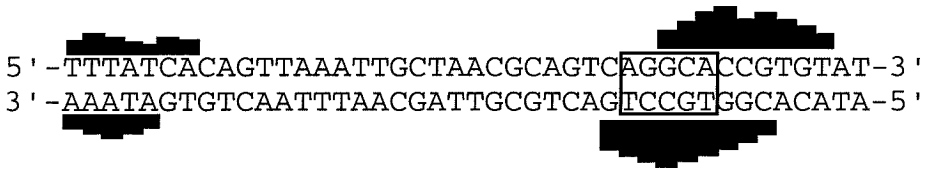


Figure 14. Ball and stick model of predicted AcImImPy-Dp/ImPyPy-Dp heterodimer. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. MPE•Fe(II) histograms of cleavage protection (footprinting) data for polyamide **6** (20 μM) with **2** (100 μM) on the *Eco*RI/*Rsa*I 167 base pair restriction fragment from pBR322. Bar heights are proportional to the relative protection from cleavage at each band. Box represents expected binding site.

conclusion.[10] In addition, the recently studied hairpin, ImImIm-γ-PyPyPy-β-Dp binds to the target site, 5'-TGGGT-3', with an apparent association constant of 8 x 10^6 M^{-1},[11] substantially lower than hairpins lacking a C-terminal imidazole,[12,13] indicating an energetic penalty for an imidazole in this position. Perhaps a C-terminal imidazole is not positioned properly to form an optimal hydrogen bond with the guanine exocyclic amine in the minor groove. Further speculation must await NMR studies being performed in collaboration with the Wemmer group at Berkeley.

Polyamide AcImPyIm-Dp **5**, in the presence of distamycin **1**, does not bind to the predicted heterodimer 5'-TGTGA-3' site; instead, the polyamide has a strong affinity for a 5'-TGCGCT-3' sequence. We propose that this site is bound by a "slipped" homodimer of AcImPyIm-Dp **5** (Figure 15). This slipped binding motif has precedence in both crystallography and studies with extended polyamides.[1c,14] While this binding specificity was not entirely unexpected, it was surprising that no binding was observed to the fully overlapped 5'-TGTGA-3' site. Polyamide AcImPyIm-Dp **5** is able to form a heterodimer with AcPyImPy-Dp **3**, however, indicating some dependence on the partner in the ability to form a heterodimer.

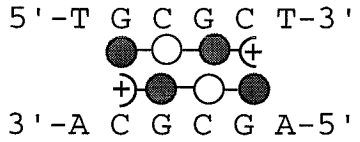


Figure 15. Ball and stick model of AcImPyIm-Dp **5** binding as a "slipped" homodimer. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively.

The final polyamide studied, AcImImPy-Dp **6**, binds heterodimerically as predicted with both distamycin **1** and ImPyPy-Dp **2**. Unlike the heterodimer pair AcPyImPy-Dp **3**/distamycin **1**, the target sites 5'-AGGTT-3' and 5'-TGGTT-3' are not bound in the absence of distamycin **1**. To confirm the binding orientation of the polyamides, affinity cleavage with EDTA-PyPyPy-Dp was performed. As expected, new

cleavage patterns consistent with heterodimeric binding in the predicted orientation appeared only in the presence of AcImImPy-Dp **6** (Figure 13). AcImImPy-Dp **6** and ImPyPy-Dp **2** also bind heterodimerically to a 5'-AGGCA-3' site as predicted.

Predictions based on the 2:1 polyamide:DNA model are realized for four of the ten combinations tested, adding 5'-WCGCW-3', 5'-WGGWW-3', and 5'-WGGCW-3' to the list of targetable DNA sequences. The fourth sequence 5'-WWGWW-3' was previously examined by Wemmer, Lown, and coworkers.[4] The AcPyImPy-Dp **3**/AcImPyIm-Dp **5** pair demonstrates recognition of 5'-ACGCA-3', while polyamide AcImImPy-Dp **6** binds heterodimerically with either distamycin **1** or ImPyPy-Dp **2** to recognize the sequences 5'-TGGTT-3' and 5'-AGGCA-3'. In contrast, polyamide AcPyPyIm-Dp **4** does not bind to any of the four predicted target sequences either as a homodimer or heterodimer. Additionally, AcImPyIm-Dp **5** appears to bind as a slipped homodimer rather than pair heterodimerically with distamycin **1** to bind a 5'-WGWGW-3' sequence. The heterodimer AcImImPy-Dp **6**/distamycin **1** is the most promising of the heterodimers studied here in that it demonstrates the expected specificity and provides the first example of recognition of contiguous G•C base pairs. Therefore, the hairpin analog of this heterodimer was synthesized and studied as reported in Chapter 5.

Conclusions. The 2:1 motif is clearly a valuable tool for predicting the sequences a given polyamide pair will bind to either homodimerically or heterodimerically. However, there are exceptions to the rules, indicating that the DNA sequence as well as individual polyamide composition must be taken into consideration. While the 2:1 motif is not yet completely generalizable, this study demonstrates the remarkable flexibility of polyamides for the recognition of many DNA sequences.

Experimental Section

^{1}H NMR spectra were recorded on a General Electric-QE 300 NMR spectrometer in $CDCl_3$ or DMSO-d_6, with chemical shifts reported in parts per million relative to

residual $CHCl_3$ or DMSO-d_5, respectively. High-resolution mass spectra were recorded using fast atom bombardment (FABMS) techniques at the Mass Spectrometry Laboratory at the University of California, Riverside. Reactions were executed under an inert argon atmosphere. Reagent grade chemicals were used as received unless otherwise noted. Tetrahydrofuran (THF) was distilled under nitrogen from sodium/benzophenone ketyl. Dichloromethane (CH_2Cl_2) and triethylamine were distilled under nitrogen from powdered calcium hydride. Dimethylformamide (DMF) was purchased as an anhydrous solvent from Aldrich. Flash chromatography was carried out using EM science Kieselgel 60 (230-400) mesh.[15] Preparative HPLC was carried out on a Beckman Instrument using a Waters DeltaPak 25 x 100 mm C_{18} column. Analytical HPLC was performed on a Hewlett Packard 1090 Series II analytical HPLC using a Vydac C_{18} reverse phase column (0.46 x 25 cm, 5 mm, HS silica). Thin-layer chromatography was performed on EM Reagents silica gel plates (0.5 mm thickness). All compounds were visualized with short-wave ultraviolet light.

3-[*N*-methyl-4-nitroimidazole-2-carboxamido]dimethylpropylamine (Im-Dp 15): To a solution of *N*-methyl-4-nitroimidazole-2-carboxylic acid **14** (1.02 g, 6.0 mmol) in anhydrous THF (8 mL) was slowly added oxalyl chloride (5 mL). This stirring solution was allowed to reflux (~75°C) for 45 min followed by lyophilization of the solvents. The resulting acid chloride was redissolved in THF (10 mL) and cooled to 0°C. Dimethylaminopropylamine (1.5 mL) was added slowly to the cooled solution. The solution was equilibrated to room temperature, allowed to stir for ~ 6 h, and then quenched with methanol (10 mL). Solvents were removed under reduced pressure, and the remaining solid was partitioned with 10% sodium bicarbonate (40 mL) and ethyl acetate (50 mL) followed by extraction with ethyl acetate (2 x 50 mL). The organic layers were then dried over magnesium sulfate and concentrated under reduced pressure to give **15** (1.29 g, 84%): ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.40 (s, 1 H), 7.79 (s, 1 H), 4.15 (s, 3 H),

3.50 (q, 2 H), 2.46 (t, 2 H), 2.30 (s, 6 H), 1.76, (quint, 2 H); FABMS *m/e* 256.1414 (M + H, 256.1410 calcd. for $C_{10}H_{18}N_5O_3$).

3-[*N*-methyl-4-nitropyrrole-2-carboxamido]dimethylpropylamine (Py-Dp 10): A solution of *N*-methyl-4-nitropyrrole-2-carboxylic acid **9** (4.04 g, 23.7 mmol) in thionyl chloride (10 mL) was allowed to reflux (~70°C) for 5 h followed by lyophilization of solvent. The resulting acid chloride was redissolved in DMF (5 mL) and cooled to 0°C. To the cooled acid chloride solution was added a solution of dimethylaminopropylamine (6 mL) in DMF (5 mL). The solution was then allowed to warm to room temperature, stirred for 6 h, and quenched with water (3 mL). After lyophilization of solvents, the residue was partitioned with ethyl acetate (150 mL), saturated sodium bicarbonate (100 mL) and water (50 mL), followed by extraction with ethyl acetate (1 x 100 mL). The ethyl acetate layer was dried over magnesium sulfate and concentrated under reduced pressure to give **10** (4.93 g, 82%): ^{1}H NMR (300 MHz, DMSO-d_6) δ 8.10 (d, 1 H), 7.40 (d, 1 H), 3.87 (s, 3 H), 3.20 (t, 2 H), 2.21 (t, 2 H), 2.12 (s, 6 H), 1.60 (quint, 2 H).

3-[4-(*N*-methyl-4-nitroimidazole-2-carboxamido)-*N*-methyl-4-nitropyrrole-2-carboxamido]dimethylpropylamine (ImPy-Dp 11): To a solution of *N*-methyl-4-nitroimidazole-2-carboxylic acid **14** (1.32 g, 7.71 mmol) in anhydrous THF (10 mL) was added oxalyl chloride (4 mL). The solution was heated to reflux for 1 h followed by lyophilization of solvents. The acid chloride residue was then redissolved in anhydrous DMF (2 mL). Separately, a solution of **10** (1.31 g, 5.14 mmol) and palladium on activated carbon (10%, 130 mg) in anhydrous DMF (15 mL) was hydrogenated in a Parr Bomb (250 psi) for 8 h. The amine mixture was filtered through celite and added to the acid chloride solution. After stirring at room temperature for 3 h anhydrous triethylamine (1 mL) was added, then allowed to stir for 5 additional hours, at which time water (3 mL) was added. Solvent was removed under reduced pressure. The product was purified by flash column chromatography (75% CH_3OH/CH_2Cl_2) to give ImPy-Dp **11** (0.260 g, 14%): ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.98 (s, 1 H), 8.03, t, 1 H), 7.82 (s, 1 H), 7.23 (d,

1 H), 6.42 (d, 1 H), 4.21 (s, 3 H), 3.98 (s, 3 H), 3.43 (q, 2 H), 2.44 (t, 2 H), 2.32 (s, 6 H), 1.73 (quint, 2 H).

3-[4-(*N*-methyl-4-nitropyrrole-2-carboxamido)-*N*-methyl-4-nitroimidazole-2-carboxamido]dimethylpropylamine (PyIm-Dp 16): To a solution of **9** (0.735 g, 4.0 mmol) was added thionyl chloride (5 mL) and allowed to reflux 5 h (~80°C), followed by lyophilization of the solvents. The resulting acid chloride was then redissolved in DMF (2 mL) and cooled to 0°C. Separately, a solution of **15** (0.510 g, 2.0 mmol) and palladium on activated carbon (10%, 135 mg) in anhydrous DMF (10 mL) was hydrogenated in a Parr Bomb (225 psi) for 3 h. The amine mixture was filtered through celite and added to the cooled acid chloride solution, followed by the addition of dry triethylamine (1 mL). The solution was allowed to stir for 16 h at which time methanol (5 mL) was added and solvents removed under reduced pressure. The product was purified by flash column chromatography (1% concentrated aqueous ammonia in methanol) to afford **16** (0.500 g, 66%): ^{1}H NMR (300 MHz, $CDCl_3$) δ 9.09 (s, 1 H), 7.60 (d, 1 H), 7.59 (t, 1 H), 7.40 (s, 1 H), 7.38 (d, 1 H), 4.01 (s, 6 H), 3.39 (q, 2 H), 2.38 (t, 2 H), 2.20, (s, 6 H), 1.72 (quint, 2 H); FABMS *m/e* 378.1903 (M + H, 378.1890 calcd. for $C_{16}H_{24}N_7O_4$).

Nitro-PyImPy-Dp (12): To **9** (0.135 g, 0.79 mmol) was added thionyl chloride (4 mL) and allowed to reflux for 5 h (~80°C), followed by lyophilization of the solvents. The resulting acid chloride was then redissolved in DMF (3 mL) and cooled to 0°C. Separately a solution of **11** (0.150 g, 0.40 mmol) and palladium on activated carbon (10%, 50 mg) in anhydrous DMF (8 mL) was hydrogenated in a Parr Bomb (300 psi) for 12 h. The amine mixture was filtered through celite and added to the cooled acid chloride solution, followed by the addition of dry triethylamine (1 mL). The solution was allowed to stir at room temperature for 16 h at which time solvents were removed under reduced pressure. The product was purified by flash column chromatography (1% concentrated aqueous ammonia in methanol) to afford **12** (0.150 g, 76%): ^{1}H NMR (300 MHz, $CDCl_3$) δ 9.51 (s, 1 H), 8.51 (s, 1H), 8.07 (t, 1 H), 7.77 (d, 1 H), 7.59 (d, 1H), 7.42 (s, 1H), 7.14

(d, 1 H), 6.54 (d. 1H), 4.03 (s, 3 H), 4.02 (s, 3 H), 3.82 (s, 3 H), 3.51 (q, 2 H), 2.46 (t, 2 H), 2.27 (s, 6 H), 1.74 (quint, 2 H); FABMS *m/e* 500.2372 (M + H, 500.2370 calcd. for $C_{22}H_{30}N_9O_5$).

Nitro-PyPyIm-Dp (17): To **9** (0.135 g, 0.79 mmol) was added thionyl chloride (4 mL) and allowed to reflux 5 h (~80°C), followed by lyophilization of the solvents. The resulting acid chloride was then redissolved in DMF (3 mL) and cooled to 0°C. Separately a solution of **16** (0.150 g, 0.40 mmol) and palladium on activated carbon (10%, 50 mg) in anhydrous DMF (8 mL) was hydrogenated in a Parr Bomb (300 psi) for 7 h. The amine mixture was filtered through celite and added to the cooled acid chloride solution, followed by the addition of dry triethylamine (1 mL). The solution was allowed to stir at room temperature for 16 h at which time solvents were removed under reduced pressure. The product was purified by flash column chromatography (0.5% concentrated aqueous ammonia in methanol) to afford **17** (0.145 g, 73%): ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.87 (s, 1 H), 8.55 (s, 1 H), 7.99 (t, 1 H), 7.65 (d, 1 H), 7.59 (d, 1 H), 7.40 (s, 1 H), 7.36 (d, 1 H), 6.45 (d, 1 H), 4.03 (s, 3 H), 3.97 (s, 3 H), 3.88 (s, 3 H), 3.47 (q, 2 H), 2.37 (t, 2 H), 2.20 (s, 6 H), 1.76 (quint, 2 H); FABMS *m/e* 500.2381 (M + H, 500.2370 calcd. for $C_{22}H_{30}N_9O_5$).

Nitro-ImPyIm-Dp (18): To a solution of **14** (0.136 g, 0.80 mmol) in anhydrous THF (4 mL) was added oxalyl chloride (3 mL) and allowed to reflux for 1 h followed by lyophilization of solvents. The resulting acid chloride was then redissolved in DMF (3 mL) and cooled to 0°C. Separately, a solution of **16** (0.150 g, 0.40 mmol) and palladium on activated carbon (10%, 50 mg) in anhydrous DMF (8 mL) was hydrogenated in a Parr Bomb (300 psi) for 5 h. The amine mixture was filtered through celite and added to the cooled acid chloride solution, followed by the addition of dry triethylamine (1 mL). The solution was allowed to stir at room temperature for 8 h at which time solvents were removed under reduced pressure. The product was purified by flash column chromatography (0.5% concentrated aqueous ammonia in methanol) to afford **18** (0.158 g,

79%): ^{1}H NMR (300 MHz, $CDCl_3$) δ 9.16 (s, 1 H), 8.28 (s, 1 H), 7.83 (s, 1 H), 7.66 (t, 1 H), 7.40 (s, 1 H), 7.27 (d, 1 H), 6.76 (d, 1 H), 4.18 (s, 3 H), 4.01 (s, 3 H), 3.94 (s, 3 H), 3.41 (q, 2 H), 2.36 (t, 2 H), 2.22 (s, 6 H), 1.74 (quint, 2 H); FABMS *m/e* 501.2298 (M + H, 501.2322 calcd. for $C_{21}H_{29}N_{10}O_5$).

Nitro-ImImPy-Dp (13): To a solution of **14** (0.136 g, 0.80 mmol) in anhydrous THF (4 mL) was added oxalyl chloride (3 mL) and allowed to reflux for 1 h followed by lyophilization of solvents. The resulting acid chloride was then redissolved in DMF (3 mL) and cooled to 0°C. Separately, a solution of **11** (0.106 g, 0.28 mmol) and palladium on activated carbon (10%, 50 mg) in anhydrous DMF (8 mL) was hydrogenated in a Parr Bomb (250 psi) for 7 h. The amine mixture was filtered through celite and added to the cooled acid chloride solution, followed by the addition of dry triethylamine (1 mL). The solution was allowed to stir at room temperature for 12 h at which time solvents were removed under reduced pressure. The product was purified by flash column chromatography (column 1: 6:3.5:0.5 $CHCl_3$:CH_3OH:NH_4OH, column 2: 1% concentrated aqueous ammonia in methanol) to afford **13** (0.86 g, 61%): ^{1}H NMR (300 MHz, $CDCl_3$/CD_3OD) δ 7.88 (s, 1 H), 7.30 (s, 1 H), 7.09 (d, 1 H), 6.55 (d, 1 H), 4.03 (s, 3 H), 3.92 (s, 3 H), 3.74 (s, 3 H), 3.20 (t, 2 H), 2.32 (t, 2 H), 2.17 (s, 6 H), 1.63 (quint, 2 H); FABMS *m/e* 501.2335 (M + H, 501.2322 calcd. for $C_{21}H_{29}N_{10}O_5$).

AcPyImPy-Dp (3): A solution of **12** (0.028 g, 0.056 mmol) and palladium on activated carbon (10%, 30 mg) in anhydrous DMF (4 mL) was hydrogenated in a Parr Bomb (350 psi) for 3 h. The amine mixture was filtered through celite and concentrated under reduced pressure. The resulting amine was redissolved in anhydrous CH_2Cl_2 (4 mL). To this solution was added acetic anhydride (100 μL) and dry pyridine (100 μL) and allowed to stir at room temperature for 3 h at which time solvents were removed under reduced pressure. The product was purified by flash column chromatography (1% concentrated aqueous ammonia in methanol) to afford **3** (0.020 g, 70%). Further purification was performed using reverse phase HPLC on a preparatory Waters DeltaPak

25 x 100 mm C_{18} column with linear gradients of 60% acetonitrile plus 0.1% TFA versus 0.1% aqueous TFA. ^{1}H NMR (300 MHz, $CDCl_3$) δ 9.19 (s, 1 H), 8.26 (s, 1 H), 8.22 (s, 1 H), 7.89 (t, 1 H), 7.39 (s, 1 H), 7.22 (d, 1 H), 7.20 (d, 1 H), 6.64 (d, 1 H), 6.4 (d, 1 H), 4.00 (s, 3 H), 3.99 (d, 6 H), 3.46 (q, 2 H), 2.40 (t, 2 H), 2.22 (s, 6 H), 2.14, (s, 3 H), 1.70 (quint, 2 H); FABMS *m/e* 512.2734 (M + H, 512.2734 calcd. for $C_{24}H_{34}N_9O_4$).

AcPyPyIm-Dp (4): A solution of **17** (0.068 g, 0.14 mmol) and palladium on activated carbon (10%, 40 mg) in anhydrous DMF (6 mL) was hydrogenated in a Parr Bomb (350 psi) for 3 h. The amine mixture was filtered through celite and concentrated under reduced pressure. The resulting amine was redissolved in anhydrous CH_2Cl_2 (5 mL). To this solution was added acetic anhydride (100 μL) and dry pyridine (100 μL) and allowed to stir at room temperature for 3 h at which time solvents were removed under reduced pressure. The product was purified by flash column chromatography (1% concentrated aqueous ammonia in methanol) to afford **4** (0.096 g, 83%). Further purification was performed using reverse phase HPLC on a preparatory Waters DeltaPak 25 x 100 mm C_{18} column with linear gradients of 60% acetonitrile plus 0.1% TFA versus 0.1% aqueous TFA. ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.68 (s, 1 H), 8.49 (s, 1 H), 8.39 (s, 1 H), 7.80 (t, 1 H), 7.40 (s, 1 H), 7.30 (d, 1 H), 7.07 (d, 1 H), 6.79 (d, 1 H), 6.61 (d, 1 H), 3.95 (s, 3 H), 3.82 (s, 3 H), 3.80 (s, 3 H), 3.40 q, 2 H), 2.38 (t, 2 H), 2.20 (s, 6 H), 2.06 (s, 3 H), 1.72 (quint, 2 H); FABMS *m/e* 512.2737 (M + H, 512.2734 calcd. for $C_{24}H_{34}N_9O_4$).

AcImPyIm-Dp (5): A solution of **18** (0.075 g, 0.15 mmol) and palladium on activated carbon (10%, 45 mg) in anhydrous DMF (4 mL) was hydrogenated in a Parr Bomb (350 psi) for 5 h. The amine mixture was filtered through celite and concentrated under reduced pressure. The resulting amine was redissolved in anhydrous CH_2Cl_2 (5 mL). To this solution was added acetic anhydride (150 μL) and dry pyridine (100 μL) and allowed to stir at room temperature for 3 h at which time solvents were removed under reduced pressure. The product was purified by flash column chromatography (1%

concentrated aqueous ammonia in methanol) to afford **5** (0.061 g, 79%). Further purification was performed using reverse phase HPLC on a preparatory Waters DeltaPak 25 x 100 mm C_{18} column with linear gradients of 60% acetonitrile plus 0.1% TFA versus 0.1% aqueous TFA. ^{1}H NMR (300 MHz, $CDCl_3$) δ 9.04 (s, 1 H), 8.86 (s, 1 H), 8.80 (s, 1 H), 7.81 (t, 1 H), 7.46 (s, 1 H), 7.39 (s, 1 H), 7.29 (d, 1 H), 6.70 (d, 1 H), 3.98 (s, 3 H), 3.94 (s, 3 H), 3.86, (s, 3 H), 3.36 (q, 2 H), 2.29 (t, 2 H), 2.14 (d, 9 H), 1.66 (quint, 2 H); FABMS *m/e* 513.2670 (M + H, 513.2686 calcd. for $C_{23}H_{33}N_{10}O_4$).

AcImImPy-Dp (6): A solution of **13** (0.041 g, 0.08 mmol) and palladium on activated carbon (10%, 38 mg) in anhydrous DMF (4 mL) was hydrogenated in a Parr Bomb (350 psi) for 3.5 h. The amine mixture was filtered through celite and concentrated under reduced pressure. The resulting amine was redissolved in anhydrous CH_2Cl_2 (5 mL). To this solution was added acetic anhydride (100 μL) and dry pyridine (300 μL) and allowed to stir at room temperature for 3 h at which time solvents were removed under reduced pressure. The product was purified by flash column chromatography (1% concentrated aqueous ammonia in methanol) to afford **6** (0.029 g, 71%). Further purification was performed using reverse phase HPLC on a preparatory Waters DeltaPak 25 x 100 mm C_{18} column with linear gradients of 60% acetonitrile plus 0.1% TFA versus 0.1% aqueous TFA. ^{1}H NMR (300 MHz, $CDCl_3$) δ 9.34 (s, 1 H), 9.30 (s, 1 H), 8.77 (s, 1 H), 7.72 (t, 1 H), 7.43 (s, 1 H), 7.39 (s, 1 H), 7.20 (d, 1 H), 6.56 (d, 1 H), 4.02 (s, 3 H), 3.99 (s, 3 H), 3.89 (s, 3 H), 3.43 (q, 2 H), 2.41 (t, 2 H), 2.23 (s, 6 H), 2.18 (s, 3 H), 1.71 (quint, 2 H); FABMS *m/e* 513.2678 (M + H, 513.2686 calcd. for $C_{23}H_{33}N_{10}O_4$).

DNA Reagents and Materials. Doubly distilled water was further purified through the Milli Q filtration system from Millipore. Sonicated, deproteinized calf thymus DNA was obtained from Pharmacia. Plasmid pBR322 and enzymes were purchased from Boehringer-Mannheim and used with the buffers supplied. Deoxyadenosine 5'-[α-^{32}P] triphosphate, thymidine 5'-[α-^{32}P] triphosphate, and adenosine 5'-[γ-^{32}P] triphoshphate were obtained from Amersham. Storage phosphor technology autoradiography was

performed using a Molecular Dynamics 400S Phosphorimager and ImageQuant software. The 167 and 517 base pair 3' and 5' end labeled *Eco*RI/*Rsa*I restriction fragments from plasmid pBR322 were prepared and purified as follows. Plasmid DNA was linearized using *Eco*RI, followed by treatment with either Klenow, deoxyadenosine 5'-[α-^{32}P] triphosphate, and thymidine 5'-[α-^{32}P] triphosphate for 3' labeling or calf alkaline phosphatase and subsequent 5' end labeling with T4 polynucleotide kinase and γ-^{32}P dATP. The linearized plasmid DNA was digested with *Rsa*I and the 167 and 517 base pair *Eco*RI/*Rsa*I restriction fragments were isolated by nondenaturing 5% polyacrylamide gel electrophoresis (PAGE). The gel bands were visualized by autoradiography, isolated, and filtered to remove the polyacrylamide. The resulting solution was further purified by phenol extraction followed by ethanol precipitation. Chemical sequencing reactions were performed as described.[16] Standard techniques were employed for DNA manipulations.[17]

Determination of Binding Sites By MPE•Fe(II) Footprinting. All reactions were carried out in a total volume of 40 μL with final concentrations of species as indicated in parentheses. The ligands were added to solutions of radiolabeled restriction fragment (10,000 cpm), calf thymus DNA (100 μM bp), tris-acetate (25 mM, pH 7.0), and NaCl (10 mM) and incubated for 1 h at 22°C. A 50 μM MPE•Fe(II) solution was prepared by mixing 100 μL of a 100 μM MPE solution with a freshly prepared 100 μM ferrous ammonium sulfate solution. Footprinting reactions were initiated by the addition of MPE•Fe(II) (5μM), followed 5 min later by the addition of dithiothreitol (5 mM), and allowed to proceed for 15 min at 22°C. Reactions were stopped by ethanol precipitation, resuspended in 100 mM Tris-borate-EDTA/80% formamide loading buffer, and electrophoresed on 8% polyacrylamide denaturing gels (5% crosslink, 7 M urea) at 2000 V for 1 h. The gels were analyzed using storage phosphor technology.

Affinity Cleavage with EDTA-PyPyPy-Dp. All reactions were carried out in a total volume of 40 μL with final concentrations of species as indicated in parentheses. The ligands were added to solutions of radiolabeled restriction fragment (10,000 cpm), calf

thymus DNA (100 μM bp), tris-acetate (25 mM, pH 7.0), and NaCl (10 mM) and incubated for 1 h at 22°C. A 100 μM EDTA•Fe(II)-PyPyPy-Dp solution was prepared by mixing 50 μL of a 200 μM EDTA-PyPyPy-Dp solution with 50 μL of a 200 μM freshly prepared ferrous ammonium sulfate solution. Footprinting reactions were initiated by the addition of 4 μL of dithiothreitol (5 mM), and allowed to proceed for 30 min at 22°C. Reactions were stopped by ethanol precipitation, resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer, and electrophoresed on 8% polyacrylamide denaturing gels (5% crosslink, 7 M urea) at 2000 V for 1 h. The gels were analyzed using storage phosphor technology.

References

1. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **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.
2. (a) Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574-1575. (b) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (c) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci., USA* **1992**, *89*, 7586. (d) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385.
3. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (b) Geierstanger, B. H.; Jacobsen, J-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry*, **1994**, *33*, 3055.
4. Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.
5. (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646-650. (b) Mrksich, M.; Dervan, P.B.; *J. Am. Chem. Soc.* 1995, 117, 3325.
6. Bialer, M.; Yagen, B.; Mechoulam, R. *Tetrahedron* **1978**, *34*, 2389.
7. Krowicki, K.; Lown, J. W. *J. Org. Chem.* **1987**, *52*, 3493-3501.
8. (a) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373. (b) Van Dyke, M. W.; Dervan, P. B. *Nucl. Acids Res.* **1983**, *11*, 5555.
9. Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 5911-5919.
10. Susanne Swalley, *unpublished results.*
11. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc., submitted.*

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

13. Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc., in press.*

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

15. Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, *43*, 2923-2925.

16. (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.

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

CHAPTER FIVE

Recognition of 5'-(A,T)GG(A,T)$_2$-3' Sequences in the Minor Groove of DNA by Hairpin Polyamides

The text of this chapter is partially taken from a submitted paper that was coauthored with Prof. Peter B. Dervan and Eldon E. Baird.

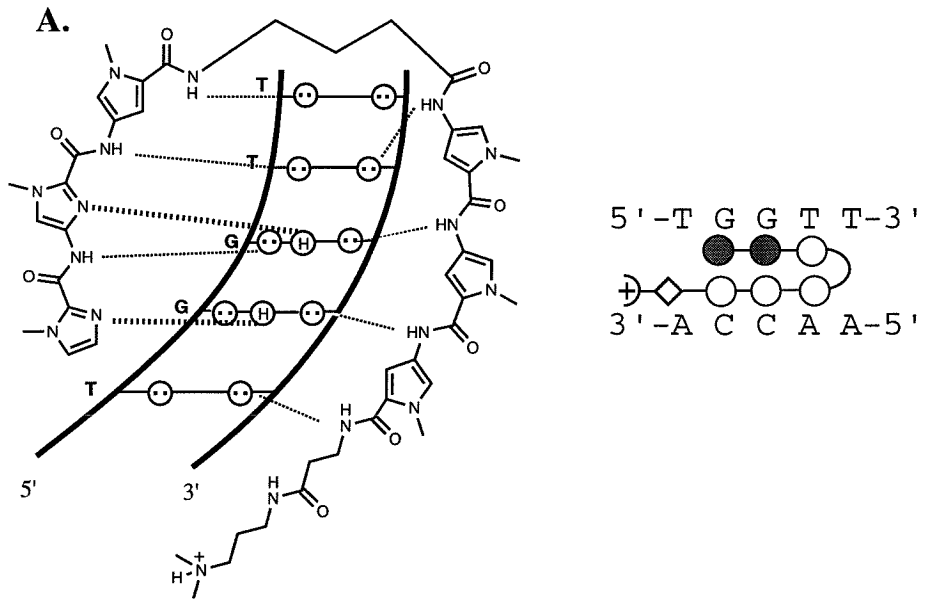
(Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *in press.*)

Introduction

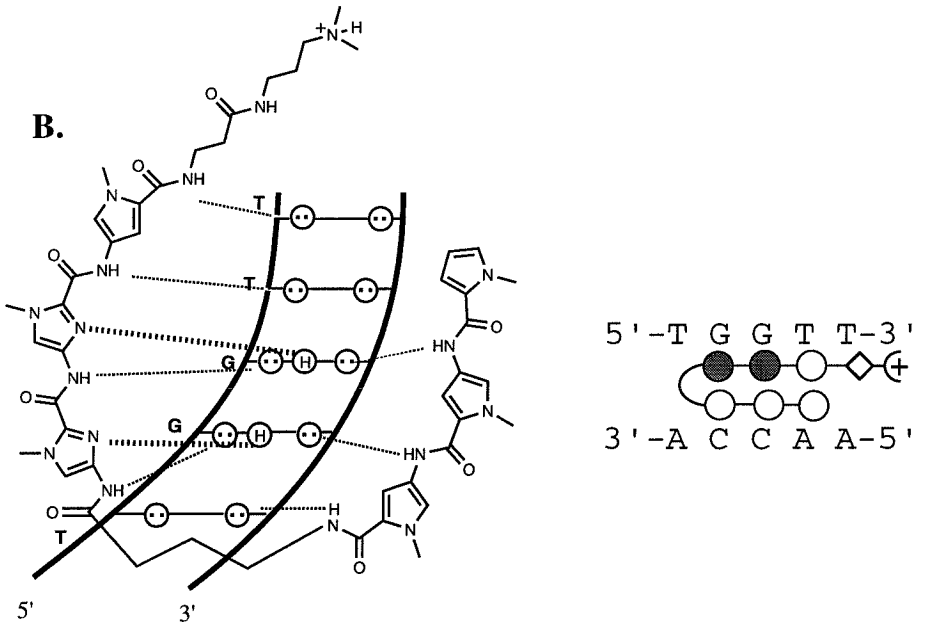
Pyrrole-imidazole polyamides offer a general method for the design of non-natural molecules for sequence-specific recognition in the minor groove of DNA.[1-3] Within the 2:1 polyamide:DNA model, an imidazole (Im) on one ligand opposite a pyrrole carboxamide (Py) on the second ligand recognizes a G•C base pair, while a pyrrolecarboxamide/imidazole combination targets a C•G base pair.[1,3] A pyrrole carboxamide/pyrrole carboxamide pair is partially degenerate for A•T or T•A base pairs.[1-3] Based on this model, the recognition of the sequences 5'-(A,T)G(A,T)C(A,T)-3',[1] 5'-(A,T)G(A,T)$_3$-3',[3] (A,T)$_2$G(A,T)$_2$-3',[4] and 5'-(A,T)GCGC(A,T)-3'[5] has been achieved. However, sequences containing *contiguous* G•C base pairs are notably absent from this list.

Formation of a hairpin polyamide by covalently linking a polyamide heterodimer with a γ-aminobutyric acid (γ) residue provides an approximate 300-fold enhancement in affinity over the unlinked polyamides, ImPyPy-Dp and AcPyPyPy-Dp.[6] Moreover, the specificity of the hairpin is greatly improved. The initial placement of the γ-amino acid turn was chosen for synthetic ease and was not varied. With the development of solid phase methodology for polyamide synthesis, we can now rapidly assess the effect of varying the position of the γ-turn monomer.[7]

In order to explore the recognition of 5'-(A,T)GG(A,T)(A,T)-3' sequences, a series of eight head-to-tail linked hairpin polyamides containing *neighboring* imidazole rings, ImImPy-γ-PyPyPy-β-Dp **1**, PyPyPy-γ-ImImPy-β-Dp **2**, AcImImPy-γ-PyPyPy-β-Dp **3**, AcPyPyPy-γ-ImImPy-β-Dp **4**, ImImPy-γ-PyPyPy-G-Dp **5**, PyPyPy-γ-ImImPy-G-Dp **6**, AcImImPy-γ-PyPyPy-G-Dp **7**, and AcPyPyPy-γ-ImImPy-G-Dp **8**, were prepared using solid phase methods (Figures 1 and 2).[7] The polyamides were all synthesized with either Boc-β-alanine-Pam or Boc-glycine-Pam resin. Each imidazole is expected to form a specific hydrogen bond with a guanine amino group allowing the recognition of contiguous

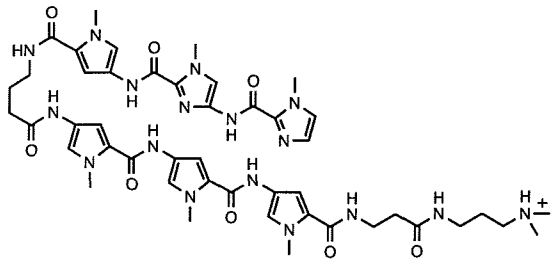


ImImPy-γ-PyPyPy-β-Dp • TGGTT

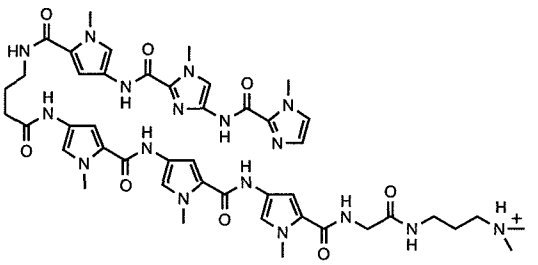


PyPyPy-γ-ImImPy-β-Dp • TGGTT

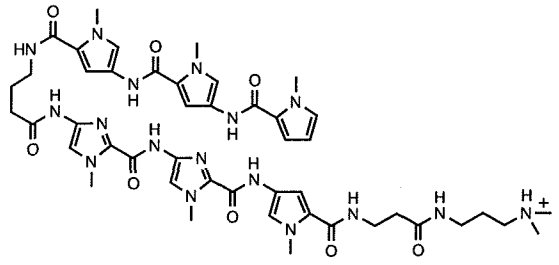
Figure 1. Binding model for the complexes formed between polyamides ImImPy-γ-PyPyPy-β-Dp **1** (A) and PyPyPy-γ-ImImPy-β-Dp **2** (B) and a 5'-TGGTT-3' sequence. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. Ball and stick models are also shown. Shaded and unshaded circles denote imidazole and pyrrole carboxamides, respectively, the β-alanine residue is represented as an unshaded diamond.



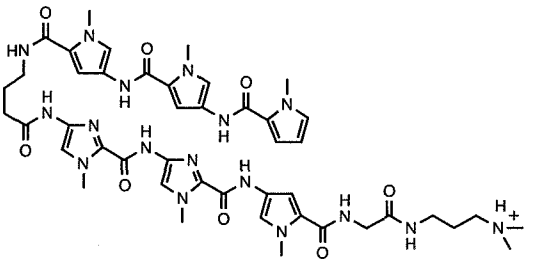
ImImPy-γ-PyPyPy-β-Dp (1)



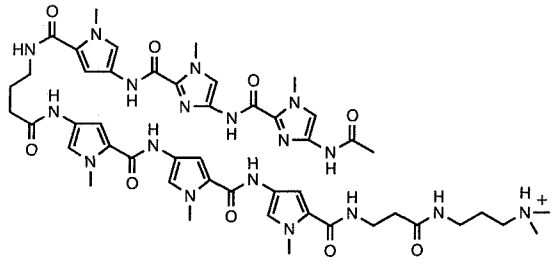
ImImPy-γ-PyPyPy-G-Dp (5)



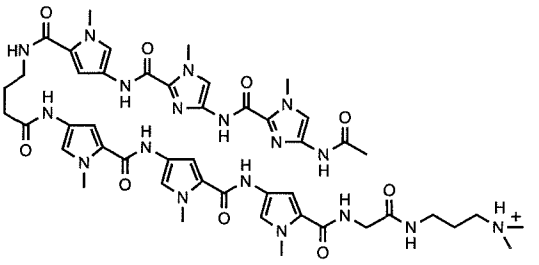
PyPyPy-γ-ImImPy-β-Dp (2)



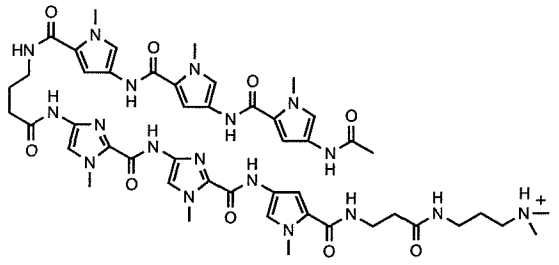
PyPyPy-γ-ImImPy-G-Dp (6)



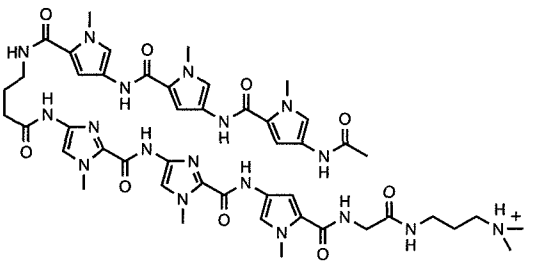
AcImImPy-γ-PyPyPy-β-Dp (3)



AcImImPy-γ-PyPyPy-G-Dp (7)



AcPyPyPy-γ-ImImPy-β-Dp (4)



AcPyPyPy-γ-ImImPy-G-Dp (8)

Figure 2. Series of polyamides synthesized using solid phase methodology.[7]

G•C base pairs (Figure 1). In addition, the linker turn position is varied within the nonacetylated and acetylated pairs of polyamides to determine the effect on the sequence specificity and binding affinity, similar to work with the ImPyPy-γ-PyPyPy-Dp series.[8] We report here the binding specificity and affinity of the polyamides as determined by the complementary techniques, MPE•Fe(II) footprinting[9] and quantitative DNase I footprinting.[10] MPE•Fe(II) footprinting verifies that sequence-specific recognition of the expected 5'-TGGTT-3' target site has been achieved. In addition, quantitative DNase I footprint titration experiments reveal that the position of the γ-linker does not dramatically affect either affinity or specificity of polyamides, especially the pair containing acetylated N-termini.

Results

Synthesis of polyamides. The polyamides ImImPy-γ-PyPyPy-β-Dp **1**, PyPyPy-γ-ImImPy-β-Dp **2**, AcImImPy-γ-PyPyPy-β-Dp **3**, AcPyPyPy-γ-ImImPy-β-Dp **4**, ImImPy-γ-PyPyPy-G-Dp **5**, PyPyPy-γ-ImImPy-G-Dp **6**, AcImImPy-γ-PyPyPy-G-Dp **7**, and AcPyPyPy-γ-ImImPy-G-Dp **8** were prepared by solid phase methodolgy (Figure 2). Four unique pyrrole and imidazole building blocks were combined in a stepwise manner on a solid support using Boc- chemistry protocols (Figure 3). For example, polyamide **2,** PyPyPy-γ-ImImPy-β-Dp, was prepared in 16 steps on the resin, and then cleaved with a single step aminolysis reaction (Figure 3). All polyamides were found to be soluble to at least 1 mM concentration in aqueous solution.

Footprinting. MPE•Fe(II) footprinting on a 3'- or 5'-^{32}P end-labeled 266 base pair *Eco*RI/*Pvu*II restriction fragment from plasmid pMEPGG (25 mM tris-acetate, 100 μM bp calf thymus DNA, 10 mM NaCl) reveals that the synthetic polyamides **1-8**, at 10 μM concentration, bind the designated target site 5'-TGGTT-3' (Figures 4,5, and 6). In addition, several single base pair mismatch sites are bound with lower affinity.

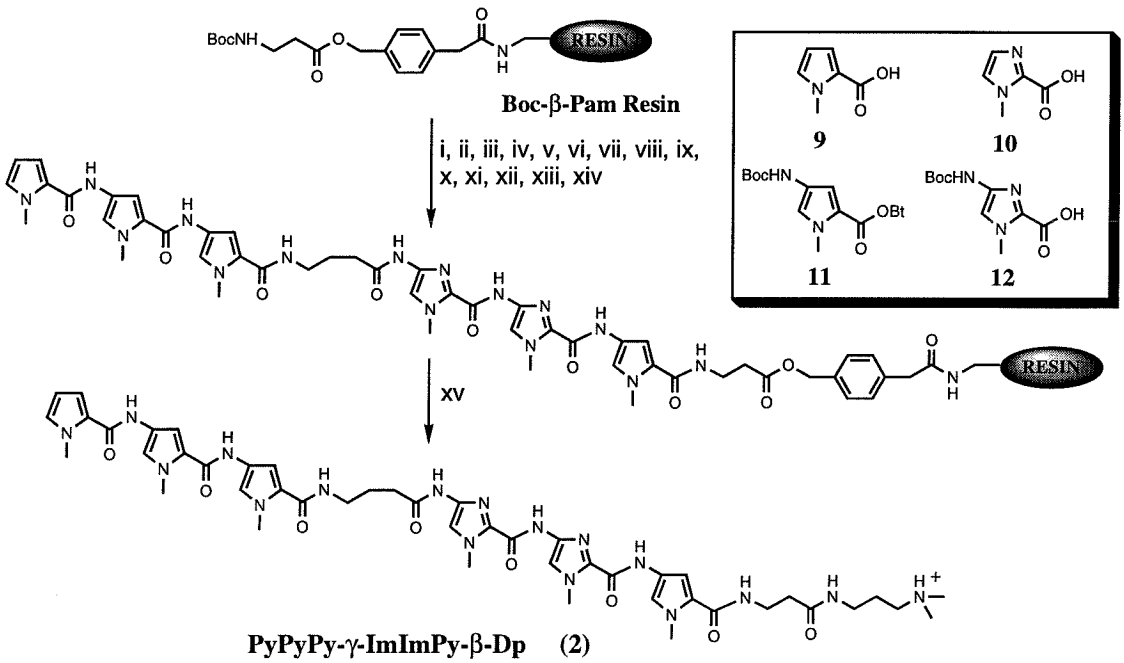
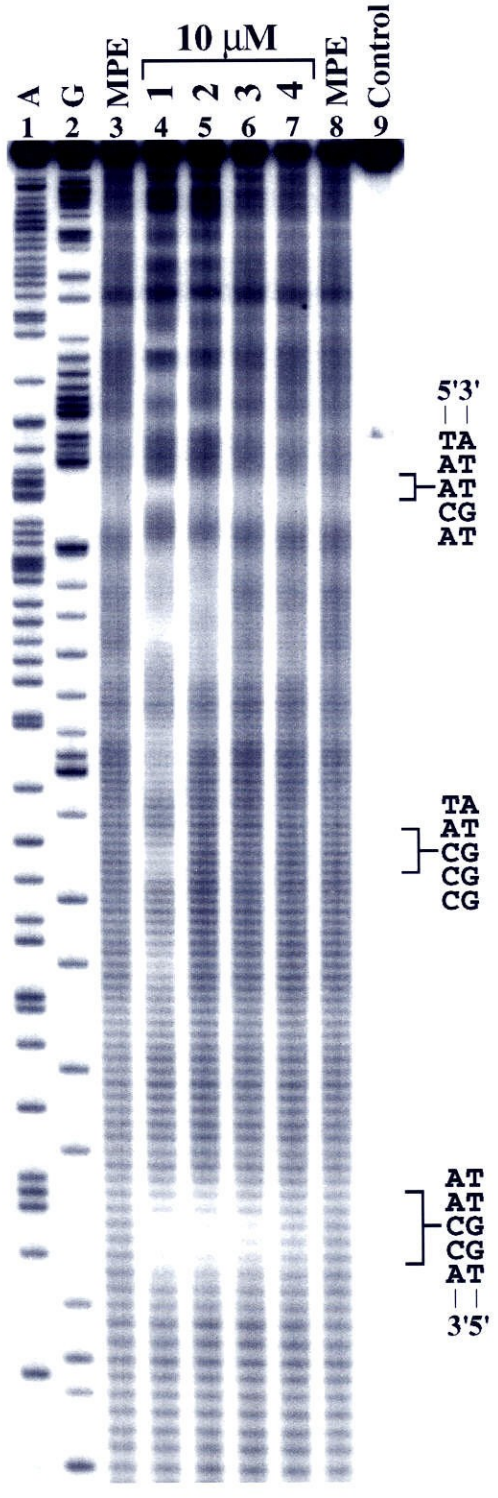


Figure 3. (Box) Pyrrole and imidazole monomers for synthesis of all compounds described here; pyrrole-2-carboxylic acid **9**, imidazole-2-carboxylic acid **10**,[1b] Boc-Pyrrole-OBt ester **11**,[7] and Boc-Imidazole-acid **12**.[7] Solid phase synthetic scheme for PyPyPy-γ-ImImPy-β-Dp starting from commercially available Boc-β-Pam-resin: (i) 80% TFA/DCM, 0.4M PhSH; (ii) BocPy-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4M PhSH; (iv) BocIm-OBt (DCC/HOBt), DIEA, DMF; (v) 80% TFA/DCM, 0.4M PhSH; (vi) BocIm-OBt (DCC/HOBt), DIEA, DMF; (vii) 80% TFA/DCM, 0.4M PhSH; (viii) Boc-g-aminobutyric acid (HBTU, DIEA); (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); pyrrole-2-carboxylic acid (HBTU/DIEA); (xv) *N,N*- dimethylaminopropylamine, 55 °C.

Figure 4. MPE•Fe(II) footprinting experiment on a 3'-^{32}P-labeled 266 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pMEPGG. The 5'-TGGTT-3' , 5'-GGGTA-3', and 5'-TGTTA-3' sites are shown on the right side of the autoradiogram. All reactions contain 10 kcpm restriction fragment, 25 mM tris-acetate, 10 mM NaCl, 100 μM calf thymus DNA (bp), and 5 mM DTT. Lane 1, A reaction; lane 2, G reaction; lanes 3 and 8, MPE•Fe(II) standard; lane 4, 10 μM ImImPy-γ-PyPyPy-β-Dp **1**; lane 5, 10 μM PyPyPy-γ-ImImPy-β-Dp **2**; lane 6, 10 μM AcImImPy-γ-PyPyPy-β-Dp **3**; lane 7, 10 μM AcPyPyPy-γ-ImImPy-β-Dp **4**; lane 9, intact DNA.


A
G
MPE
10 μM
1
2
3
4
MPE
Control
1
2
3
4
5
6
7
8
9
5'3'
TA
AT
AT
CG
AT
TA
AT
CG
CG
CG
AT
AT
CG
CG
AT
3'5'

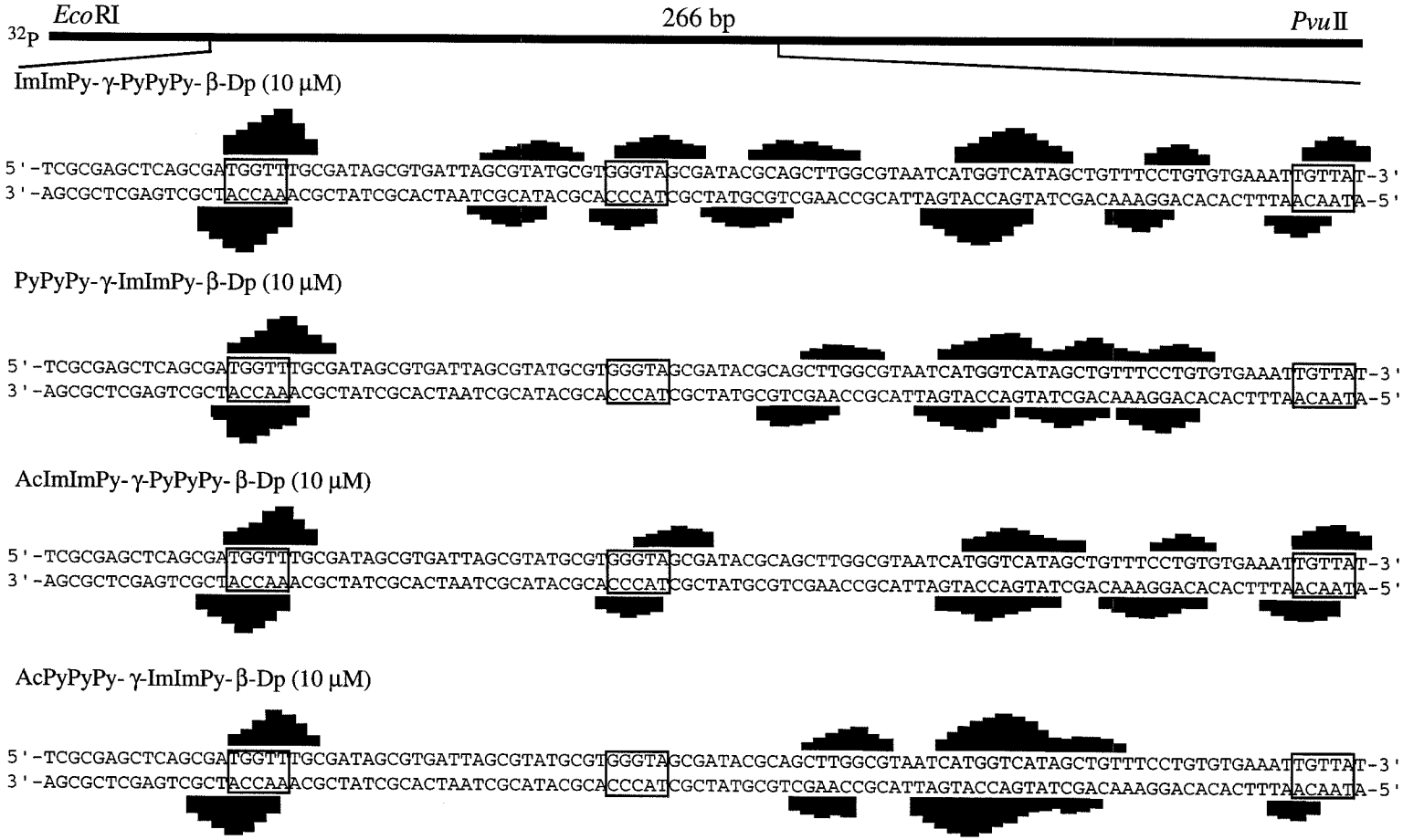


Figure 5. Histograms of cleavage protection (footprinting) data. (Top) Illustration of the 266 bp restriction fragment with the position of the sequence indicated. MPE•Fe(II) protection patterns for polyamides at 10 μM concentration. Bar heights are proportional to the relative protection from cleavage at each band. Boxes represent equilibrium binding sites determined by the published model.[9] Only sites that were quantitated by DNase I footprint titrations are boxed.

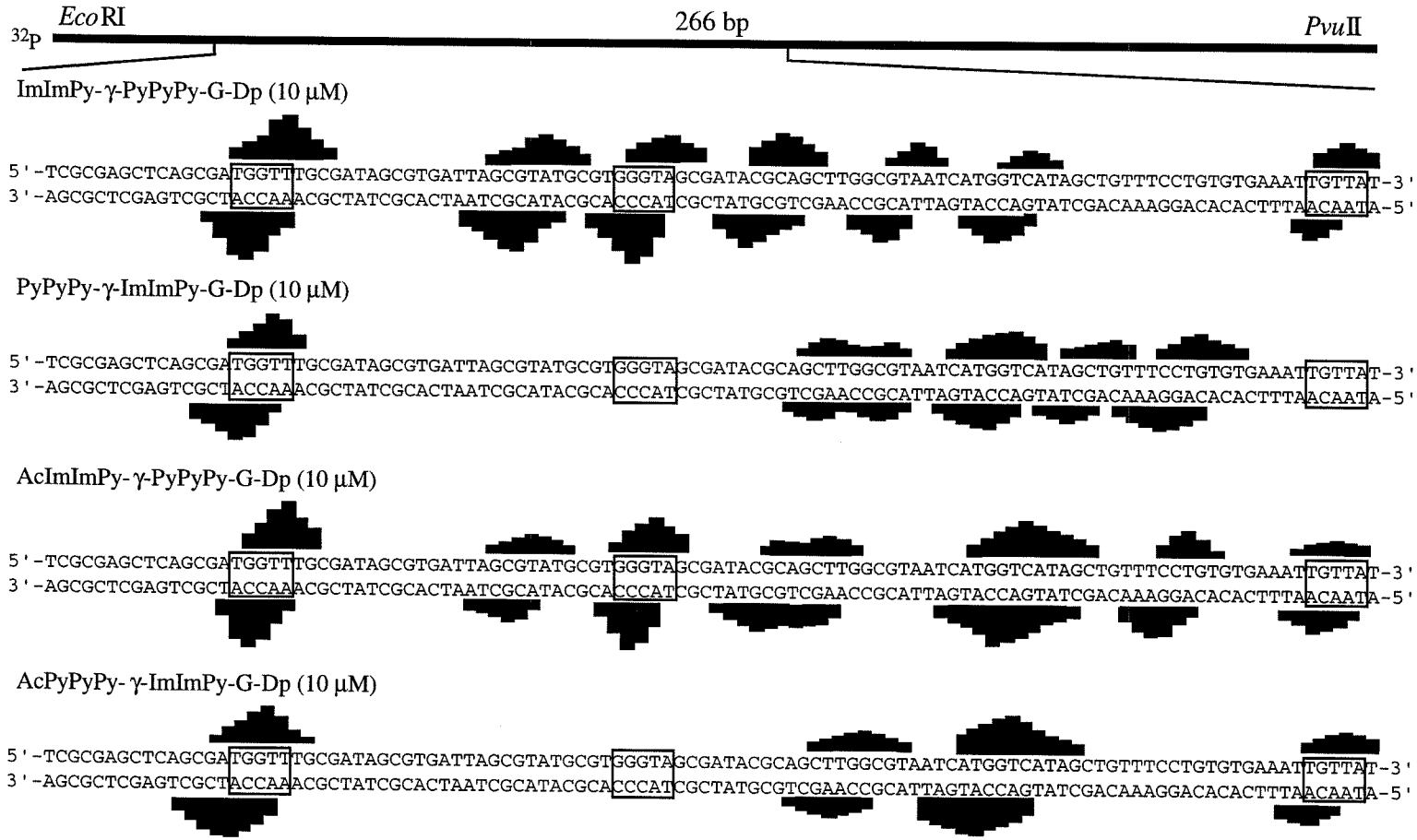


Figure 6. Histograms of cleavage protection (footprinting) data. (Top) Illustration of the 266 bp restriction fragment with the position of the sequence indicated. MPE•Fe(II) protection patterns for polyamides at 10 μM concentration. Bar heights are proportional to the relative protection from cleavage at each band. Boxes represent equilibrium binding sites determined by the published model.[9] Only sites that were quantitated by DNase I footprint titrations are boxed.

Quantitative DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0 and 22°C) were performed to determine the apparent first-order binding affinities of the eight polyamides **1-8** for a designated match site, 5'-TGGTT-3', as well as for two single base pair mismatch sites, 5'-TGTTA-3' and 5'-GGGTA-3' (Tables 1 and 2).[10] The polyamide ImImPy-γ-PyPyPy-β-Dp binds the target site 5'-TGGTT-3' with the highest affinity (first order association constant K_a = 1.0 x 10^8 M^{-1}) (Figures 7 and 8). Polyamides **2-4**, **7**, and **8** have lower but approximately equal first-order association constants of K_a = ~2 x 10^7 M^{-1} for the target site. The nonacetylated polyamides **5** and **6** containing G-Dp have substantially reduced affinities of ~2 x 10^6 M^1. The non-acetylated polyamides in the β-Dp series are > 50-fold specific for the 5'-TGGTT-3' match site over either of the single base pair mismatch sites analyzed. The acetylated pair of β-Dp polyamides exhibit lower sequence specificity for the analyzed sites. The non-acetylated G-Dp polyamides show both low affinity and specificity, while the acetylated G-Dp polyamides have reasonable affinity, but lower specificity than the optimal polyamide ImImPy-γ-PyPyPy-β-Dp.

Discussion

Each polyamide within this series specifically binds the five base pair designated target sequence 5'-TGGTT-3', as shown by MPE•Fe(II) footprinting experiments, providing the first example of contiguous G•C recognition in the polyamide-DNA motif. Interestingly, the polyamides prefer different mismatch sequences, indicating that the position of the turn alters sequence selectivity, although only for the mismatches.

Quantitative DNase I footprint titration experiments reveal that ImImPy-γ-PyPyPy-β-Dp **1** is optimal within this series of eight polyamides. This hairpin binds a 5'-TGGTT-3' match site with an apparent first-order association constant of K_a = 1 x 10^8 M^{-1}, while the corresponding hairpin PyPyPy-γ-ImImPy-β-Dp **2**, which differs only in the position of

Table I. Apparent First-Order Association Constants (M^{-1})[a, b]

	Match Site	Single Mismatch Sites	
Polyamide	5'-aTGGTTt -3'	5'-tTGTTAt-3'	5'-tGGGTAg-3'
ImImPy-γ-PyPyPy-β-Dp	1.0×10^8 (0.1)	1.7×10^6 (0.6)	$\leq 1 \times 10^6$
PyPyPy-γ-ImImPy-β-Dp	1.6×10^7 (0.2)	$< 1 \times 10^5$	$< 1 \times 10^5$
AcImImPy-γ-PyPyPy-β-Dp	1.3×10^7 (0.7)	1.6×10^6 (1.1)	1.3×10^6 (0.8)
AcPyPyPy-γ-ImImPy-β-Dp	2.0×10^7 (0.3)	1.3×10^6 (0.5)	$\leq 1 \times 10^6$

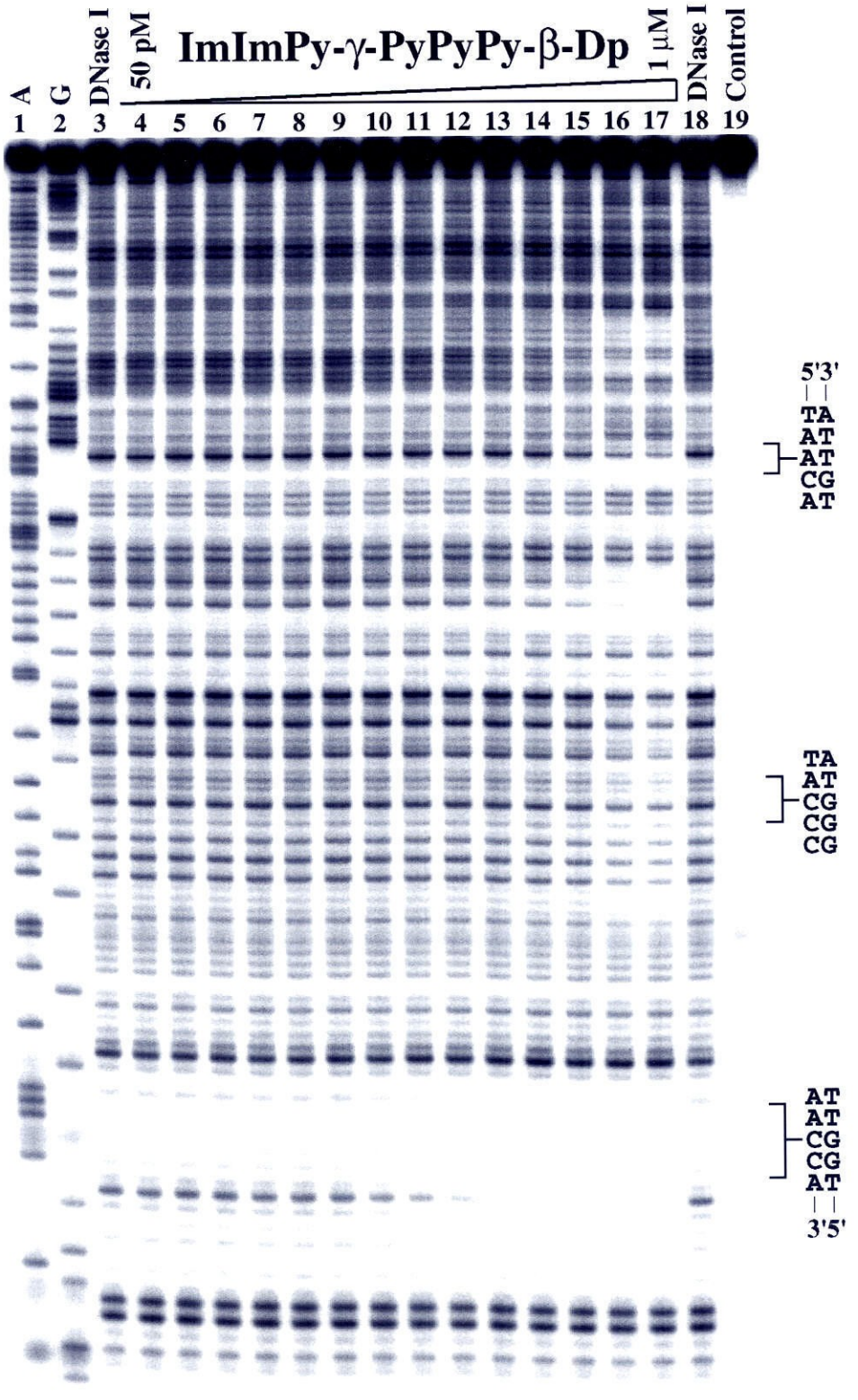
[a]Values reported are the mean values measured from at least three footprint titration experiments, with the standard deviation for each data set indicated in parentheses. [b]The assays were performed at 22 °C at pH 7.0 in the presence of 10 mM tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

Table II. Apparent First-Order Association Constants (M^{-1})[a, b]

	Match Site	Single Mismatch Sites	
Polyamide	5'-aTGGTTt -3'	5'-tTGTTAt-3'	5'-tGGGTAg-3'
ImImPy-γ-PyPyPy-G-Dp	3.6×10^6 (0.8)	2.5×10^6 (0.2)	2.9×10^6 (0.5)
PyPyPy-γ-ImImPy-G-Dp	2.0×10^6 (0.6)	$< 1 \times 10^5$	$< 1 \times 10^5$
AcImImPy-γ-PyPyPy-G-Dp	1.5×10^7 (0.3)	$\leq 1 \times 10^5$	1.5×10^6 (0.2)
AcPyPyPy-γ-ImImPy-G-Dp	3.6×10^7 (1.1)	5.0×10^6 (1.7)	5.9×10^6 (1.7)

[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 $CaCl_2$.

Figure 7. Quantitative DNase I footprint titration experiment with ImImPy-γ-PyPyPy-β-Dp **1** on the 3'-^{32}P-labeled 266 base pair *Eco*RI/*Pvu*II restriction fragment from plasmid pMEPGG: lane 1, A reaction; lane 2, G reaction; lanes 3 and 18, DNase I standard; lanes 4-17, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM ImImPy-γ-PyPyPy-β-Dp **1**, respectively; lane 19, intact DNA. The 5'-TGGTT-3', 5'-GGGTA-3', and 5'-TGTTA-3' sites which were analyzed are shown on the right side of the autoradiogram. All reactions contain 10 kpm restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.


A
G
DNase I
50 pM
ImImPy-γ-PyPyPy-β-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
5'3'
TA
AT
AT
CG
AT
TA
AT
CG
CG
CG
AT
AT
CG
CG
AT
3'5'

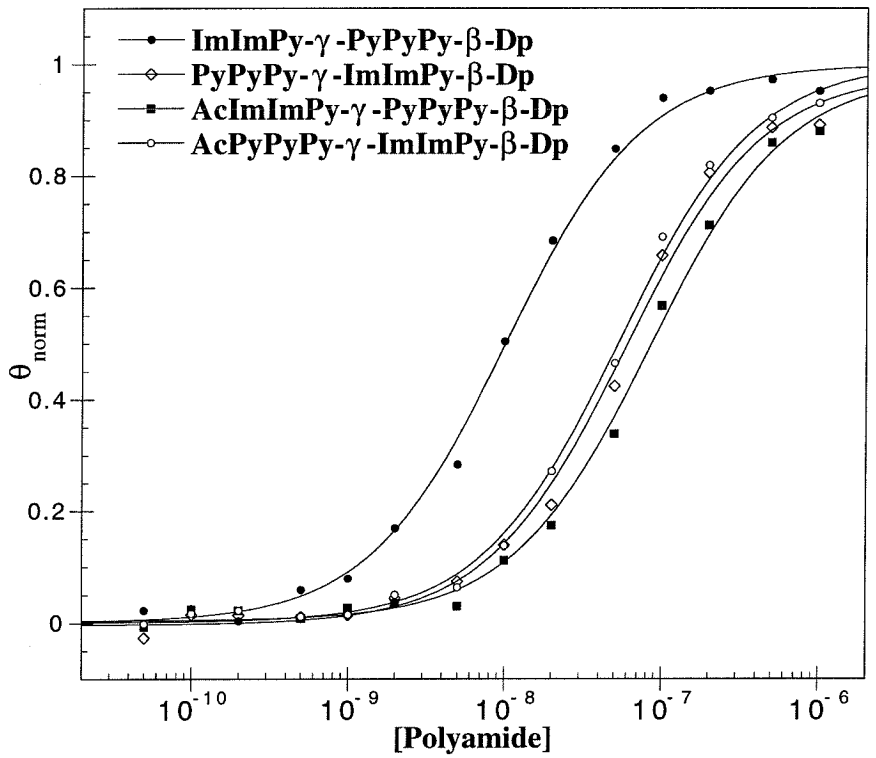


Figure 8. Data for the quantitative DNase I footprint titration experiments for the four polyamides **1-4** in complex with the designated 5'-TGGTT-3' site. The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography and processed as described in the experimental section. The data points for ImImPy-γ-PyPyPy-β-Dp **1**, PyPyPy-γ-ImImPy-β-Dp **2**, AcImImPy-γ-PyPyPy-β-Dp **3**, and AcPyPyPy-γ-ImImPy-β-Dp **4** are indicated by filled circles (●), open diamonds (◊), filled squares (■), and open circles (○), respectively. The solid curves are the best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm using eq. 2.

the γ turn, shows lower affinity (K_a = ~2 x 10^7 M^{-1}) for the 5'-TGGTT-3' site. Both unacetylated polyamides demonstrate good specificity (> 10-fold) for the target match site over the single base pair mismatch sites. The acetylated polyamides are similar in affinity to PyPyPy-γ-ImImPy-β-Dp **2**, but exhibit lower specificity. AcImImPy-γ-PyPyPy-β-Dp **3** and AcPyPyPy-γ-ImImPy-β-Dp **4** are virtually indistinguishable from each other based on affinity and specificity for the analyzed target sequences, indicating little preference for turn position. The acetyl group may sterically prevent the polyamide from sitting deep in the minor groove, thus lowering the penalty for a single base pair mismatch and resulting in lower specificity. However, affinity is maintained, perhaps by the addition of a hydrogen bond between the acetyl amide proton and a flanking base pair. The polyamides containing glycine-Dp also show lower affinity and no preference for turn position. Similar to the acetyl group, glycine may have negative interactions with the floor of the minor groove preventing the formation of optimal van der Waals contacts, resulting in lower affinity and specificity. We defer further speculation until NMR structural studies are complete.

The optimal contiguous imidazole-containing polyamide is remarkably similar in affinity and specificity to the single imidazole-containing hairpin polyamide, ImPyPy-γ-PyPyPy-β-Dp, indicating little or no energetic penalty in this system for *adjacent* imidazoles.[8] Importantly, the position of the hairpin turn does not significantly affect the recognition of the target 5'-TGGTT-3' match site, although single base pair mismatch relative affinities are altered.

Implications for the Design of Minor Groove Binding Molecules. The 2:1 motif has previously been used to specifically target several sequences: 5'-TGTCA-3',[1] 5'-TGTTA-3',[3] 5'-AAGTT-3',[4] and 5'-TGCGCA-3'.[5] The results reported herein add sequences containing two contiguous G•C base pairs to the list, expanding the sequence repertoire for DNA recognition by polyamides. Furthermore, turn position showed minimal effects on the specificity and affinity of the polyamides, indicating a new degree of

flexibility within the 2:1 motif. The expansion of the polyamide sequence repertoire through contiguous G•C recognition coupled with solid phase synthetic advances allowing the rapid assembly and characterization of polyamides brings the goal of sequence-specific recognition of any DNA sequence by designed molecules closer to fruition.

Experimental Section.

Materials. 0.2 mmol/gram Boc-Glycine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-G-Pam-Resin), 0.2 mmol/gram Boc-β-alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-β-Pam-Resin), Dicyclohexylcarbodiimide (DCC), Hydroxybenzo-triazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU), Boc-Glycine and Boc-β-alanine were purchased from Peptides International. *N,N*-diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF), *N*-methylpyrrolidone (NMP), DMSO/NMP, Acetic anhydride (Ac_2O), and 0.0002M potassium cyanide/pyridine were purchased from Applied Biosystems. Boc-γ-aminobutyric acid was from NOVA Biochem, dichloromethane (DCM) and triethylamine (TEA) was reagent grade from EM, thiophenol (PhSH), dimethylaminopropylamine from Aldrich, trifluoroacetic acid (TFA) from Halocarbon, phenol from Fisher, and ninhydrin from Pierce. All reagents were used without further purification.

Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. and are used for filtration of DCU. Disposable polypropylene filters are also used for washing resin for ninhydrin and picric acid tests, and for filtering pre-dissolved amino acids into synthesis cartridges. A shaker for manual solid phase synthesis was obtained from Milligen. A rotary evaporator was also modified for use as a shaker. Screw- cap glass peptide synthesis reaction vessels (5 ml and 20 ml) with a #2 sintered glass frit were made

at the Caltech glass shop as described by Kent.[11] ^{1}H NMR were recorded in d_6-DMSO on a GE 300 instrument operating at 300 MHz. Chemical shifts are reported in ppm relative to the solvent residual signal. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix- assisted, laser desorption/ionization time of flight mass spectrometry was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed either on a HP 1090M analytical HPLC or a Beckman Gold system using a Rainin 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 HPLC was carried out on a Beckman HPLC using a Waters DeltaPak 25 x 100 mm , 100μm C_{18} 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 0.2μm filtered. Reagent-grade chemicals were used unless otherwise stated.

Activation of Boc-γ-aminobutyric, imidazole-2-carboxylic acid and pyrrole-2-carboxylic acid. The appropriate amino acid or acid (2 mmol) was dissolved in 2 ml DMF. HBTU (720 mg, 1.9 mmol) was added followed by DIEA (1 ml) and the solution lightly shaken for at least 5 min.

Activation of Boc-Imidazole acid. Boc imidazole acid (257 mg, 1 mmol) and HOBt (135 mg, 1 mmol) were dissolved in 2 ml DMF, DCC (202 mg, 1 mmol) is then added and the solution allowed to stand for at least 5 minutes.

Typical Manual Synthesis Protocol: **AcPyPyPy-γ-ImImPy-β-Dp**. Boc-β-Pam-resin (1.25 g, 0.25 mmol amine) was shaken in DMF for 30 min and drained. The N-Boc group removed by washing with DCM for 2 x 30 s, followed by a 1 min shake in 80% TFA/DCM/0.5M PhSH, draining the reaction vessel and a brief 80% TFA/DCM/ 0.5 M PhSH wash, and 20 min shaking in 80% TFA/DCM/0.5M PhSH solution. The

resin was washed 1 min with DCM and 30 s with DMF. A resin sample (8-10 mg) was taken for analysis. The resin was drained completely and Boc-pyrrole-OBt monomer (357 mg, 1 mmol) dissolved in 2 ml DMF added followed by DIEA (1 ml) and the resin shaken vigorously to make a slurry. The coupling was allowed to proceed for 45 min. A resin sample (8-10 mg) was taken after 40 min to check reaction progress. The reaction vessel was washed with DMF for 30 s and dichloromethane for 1 min to complete a single reaction cycle. Six additional cycles were performed adding, BocIm-OH (DCC/HOBt), BocIm-OH (DCC/HOBt), Boc-γ-aminobutyric acid (HBTU/DIEA) and allowed to couple for 2 hours, BocPy-OBt, BocPy-OBt, and pyrrole-2-carboxylic acid (HBTU/DIEA). The resin was washed with DMF, DCM, MeOH, and ethyl ether and then dried *in vacuo*. PyPyPy-γ-ImImPy-β-Pam-Resin (180 mg, 29 μmol)[12] was weighed into a glass scintillation vial, 1.5 ml of *N,N*-dimethylaminopropylamine added, and the mixture heated at 55 °C for 18 hours. The resin was removed by filtration through a disposable polypropylene filter and washed with 5 ml of water, the amine solution and the water washes combined, and the solution loaded on a C_{18} preparatory HPLC column, the column allowed to wash for 4 min in 0.1% TFA at 8 ml/min, the polyamide was then eluted in 100 min. as a well defined peak with a gradient of 0.25% acetonitrile per min. The polyamide was collected in four separate 8 ml fractions, the purity of the individual fractions verified by HPLC and ^{1}H NMR, to provide purified PyPyPy-γ-ImImPy-β-Dp **2**. (11.2 mg, 39% recovery), UV λ_{max}, 246 (31,100), 312 (51,200) HPLC, r.t. 23.6, ^{1}H NMR (DMSO-d_6) δ 10.30 (s, 1 H), 10.26 (s, 1 H), 9.88 (s, 1 H), 9.80 (s, 1 H), 9.30 (s, 1 H), 9.2 (br s, 1 H), 8.01 (m, 3 H), 7.82 (br s 1 H), 7.54 (s, 1 H), 7.52 (s, 1 H), 7.20 (d, 1 H, *J* = 1.3 Hz), 7.18 (d, 1 H, *J* = 1.2 Hz), 7.15 (d, 1 H, *J* = 1.3 Hz), 7.01 (d, 1 H, *J* = 1.4 Hz), 6.96 (d, 1 H, *J* = 1.4 Hz), 6.92 (d, 1 H, *J* = 1.8 Hz), 6.89 (m, 2 H), 6.03 (t, 1 H, *J* = 2.4 Hz), 3.97 (s, 3 H), 3.96 (s, 3 H), 3.85 (s, 3 H), 3.82 (s, 3 H), 3.78 (m, 6 H), 3.37 (m, 2 H), 3.20 (q, 2 H, *J* =

5.7 Hz), 3.08 (q, 2 H J = 6.6 Hz), 2.94 (q, 2 H J = 5.3 Hz), 2.71 (d, 6 H J = 5.8 Hz), 2.32 (m, 4 H), 1.83 (m, 4 H); MALDI-TOF-MS, 978.7 (979.1 calc. for M+H).

ImImPy-γ-PyPyPy-β-Dp 1. Polyamide was prepared by machine assisted solid phase synthesis protocols[7] and 900 mg resin cleaved and purifed to provide **1** as a white powder. (69 mg, 48% recovery), UV λ_{max}, 246 (43,300), 308 (54,200) HPLC, r.t. 23.9, ^{1}H NMR (DMSO-d_6) δ 10.31 (s, 1 H), 9.91 (s, 1 H), 9.90 (s, 1 H), 9.85 (s, 1 H), 9.75 (s, 1 H), 9.34 (br s, 1 H), 8.03 (m, 3 H), 7.56 (s, 1 H), 7.46 (s, 1 H), 7.21 (m, 2 H), 7.15 (m, 2 H), 7.07 (d, 1 H J = 1.2 Hz), 7.03 (d, 1 H, J = 1.3 Hz), 6.98 (d, 1 H, J = 1.2 Hz), 6.87 (m, 2 H), 4.02 (m, 6 H), 3.96 (m, 6 H), 3.87 (m, 6 H), 3.75 (q, 2 H, J = 4.9 Hz), 3.36 (q, 2 H, J = 4.0 Hz), 3.20 (q, 2 H, J = 4.7 Hz), 3.01 (q, 2 H J = 5.1 Hz), 2.71 (d, 6H, J = 4.8 Hz), 2.42 (m, 4 H), 1.80 (m, 4 H) MALDI-TOF-MS 978.8, (979.1 calc. for M + H).

AcImImPy-γ-PyPyPy-β-Dp 3. Polyamide was prepared by manual solid phase protocols and isolated as a white powder. (8 mg, 28% recovery), UV λ_{max}, 246 (43,400), 312 (50,200) HPLC, r.t. 24.8, ^{1}H NMR (DMSO-d_6) δ 10.35 (s, 1 H), 10.30 (s, 1 H), 9.97 (s, 1 H), 9.90 (s, 1 H), 9.82 (s, 1 H), 9.30 (s, 1 H), 9.2 (br s , 1H), 8.02 (m, 3 H), 7.52 (s, 1 H). 7.48 (s, 1 H), 7.21 (m, 2H), 7.16 (d, 1 H, J = 1.1 Hz), 7.11 (d, 1 H, J = 1.2 Hz), 7.04 (d, 1 H, J = 1.1 Hz), 6.97 (d, 1 H, J = 1.3 Hz), 6.92 (d, 1 H, J = 1.4 Hz), 6.87 (d, 1 H, J = 1.2 Hz), 3.99 (s, 3 H), 3.97 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.80 (s, 3 H), 3.79 (s, 3 H), 3.47 (q, 2 H, J = 4.7 Hz), 3.30 (q, 2 H, J = 4.6 Hz), 3.20 (q, 2 H, J = 5.0 Hz), 3.05 (q, 2 H, J = 5.1 Hz), 2.75 (d, 6 H, J = 4.1 Hz), 2.27 (m, 4 H), 2.03 (s, 3 H), 1.74 (m, 4 H) MALDI-TOF-MS, 1036.4 (1036.1 calc. for M+H).

AcPyPyPy-γ-ImImPy-β-Dp 4. Polyamide was prepared by machine assisted solid phase methods protocols[7] as a white powder. (14 mg, 48% recovery), UV λ_{max}, 246 (44,400), 312 (52,300) HPLC, r.t. 23.8, ^{1}H NMR (DMSO-d_6) δ 10.32 (s, 1 H), 10.28 (s, 1 H), 9.89 (m, 2 H), 9.82 (s, 1 H), 9.18 (s, 1 H), 9.10 (br s, 1 H), 8.03 (m, 3 H), 7.55 (s, 1

H), 7.52 (s, 1 H), 7.21 (d, 1 H, J = 1.1 Hz), 7.18 (d, 1 H, J = 7.16), 7.15 (d, 1 H, J = 1.0 Hz), 7.12 (d, 1H, J = 1.0 Hz), 7.02 (d, 1 H, J = 1.0 Hz), 6.92 (d, 1 H, J = 1.1 Hz), 6.87 (d, 1H, J = 1.1 Hz), 6.84 (d, 1H, J = 1.0 Hz), 3.97 (s, 3 H), 3.93 (s, 3 H), 3.87 (s, 3 H), 3.80 (s, 3 H), 3.78 (m, 6 H), 3.35 (q, 2 H, J = 5.6 Hz), 3.19 (q, 2 H, J = 5.3 Hz), 3.08 (q, 2 H, J = 5.7 Hz), 2.87 (q, 2 H, J = 5.8 Hz), 2.71 (d, 6 H, J = 4.0 Hz), 2.33 (m, 4 H), 1.99 (s, 3 H), 1.74 (m, 4 H). MALDI-TOF-MS, 1036.2 (1036.1 calc for M+H).

Characterization of polyamides **5-8** will be reported in Eldon Baird's thesis.

Construction of plasmid DNA. Using T4 DNA ligase, the plasmid pMEPGG was constructed by ligation of an insert, 5'-GATCGCGAGCTCAGCGATGGTTTGC-GATAGCGTGATTAGCGTATGCGTGGGTAGCGATACGC-3' and 5'-GCGTATCG-CTACCCACGCATACGCTAATCACGCTATCGCAAACCATCGCTGAGCTCGCGA TC-3', into pUC 19 previously cleaved with *Bam*HI and *Hind*III. Ligation products were used to transform Epicurian Coli XL 1 Blucompetent cells (Stratagene). Colonies were selected for α-complementation on 25 mL Luria-Bertani medium agar plates containing 50 μg/mL ampicillin and treated with XGAL and IPTG solutions. Large-scale plasmid purification was performed with Qiagen purification kits. Plasmid DNA concentration was determined at 260 nm using the relation 1 OD unit = 50 μg/mL duplex DNA. The plasmid was linearized with *Eco*RI, followed by treatment with either Klenow, deoxyadenosine 5'-[α-^{32}P]-triphosphate (Amersham), and thymidine 5'-[α-^{32}P]-triphosphate for 3' labeling or calf alkaline phosphatase and subsequent 5' end labeling with T4 polynucleotide kinase and γ-^{32}P-dATP. The 3' or 5' end-labeled fragment was then digested with *Pvu*II and isolated by nondenaturing gel electrophoresis. The 3'- or 5' ^{32}P-end-labeled 266 base pair *Eco*RI/*Pvu*II restriction fragment was used in all experiments described here. Chemical sequencing reactions were performed according to published protocols.[13] Standard protocols were used for all DNA manipulations.[14]

Identification of Binding Sites by MPE•Fe(II) Footprinting. All reactions were carried out in a total volume of 40 μL with final concentrations of species as indicated in parentheses. The ligands were added to solutions of radiolabeled restriction fragment (10 000 cpm), calf thymus DNA (100 μM bp), tris-acetate (25 mM, pH 7.0), and NaCl (10 mM) and incubated for 1 h at 22°C. A 50 μM MPE•Fe(II) solution was prepared by mixing 100 μL of a 100 μM MPE solution with a freshly prepared 100 μM ferrous ammonium sulfate solution. Footprinting reactions were initiated by the addition of MPE•Fe(II) (5μM), followed 5 min later by the addition of dithiothreitol (5 mM), and allowed to proceed for 15 min at 22°C. Reactions were stopped by ethanol precipitation, resuspended in 100 mM Tris-borate-EDTA/80% formamide loading buffer, and electrophoresed on 8% polyacrylamide denaturing gels (5% crosslink, 7 M urea) at 2000 V for 1 h. The gels were analyzed using storage phosphor technology.

Analysis of Energetics by Quantitative DNase I Footprint Titration. All reactions were executed in a total volume of 40 μL with final concentrations of each species as indicated. The ligands, ranging from 50 pM to 1 μM, were added to solutions of radiolabeled restriction fragment (10,000 cpm), tris-HCl (10 mM, pH 7.0), KCl (10 mM), $MgCl_2$ (10 mM) and $CaCl_2$ (5 mM) and incubated for 4 h at 22 °C. Footprinting reactions were initiated by the addition of 4 μL of a stock solution of DNase I (0.025 units/mL) containing 1 mM dithiothreitol and allowed to proceed for 6 min at 22 °C. The reactions were stopped by addition of a 3 M sodium acetate solution containing 50 mM EDTA and ethanol precipitated. The reactions were resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer and electrophoresed on 8% polyacrylamide denaturing gels (5% crosslink, 7 M urea) at 2000 V for 1 hr. The footprint titration gels were dried and quantitated using storage phosphor technology.

Apparent first-order association constants were determined as previously described.[6,10] The data were analyzed by performing volume integrations of the 5'-TGGTT-3', 5'-TGTTA-3', and 5'-GGGTA-3' sites and a reference site. Binding sites are assumed to be independent and noninteracting as they are separated by at least one full turn of the double helix. 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 apparent monomeric association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. The concentration of DNA used for quantitative footprint titrations is $\leq$ 50 pM, which justifies the assumption that free ligand concentration is approximately equal to total ligand concentration.[10] Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software (version 2.1, Abelbeck software) running on a Power Macintosh 6100/60AV computer 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. 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:

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

Quantitation by Storage Phosphor Technology Autoradiography. Photostimulable storage phosphorimaging plates (Kodak Storage Phosphor Screen S0230 obtained from Molecular Dynamics) were pressed flat against gel samples and exposed in the dark at 22°C for 12-16 h. A Molecular Dynamics 400S PhosphorImager was used to obtain all data from the storage screens. The data were analyzed by performing volume integrations of all bands using the ImageQuant v. 3.2 software running on an AST Premium 386/33 computer.

References

1. (a) Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574-1575. (b) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (c) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci., USA* **1992**, *89*, 7586. (d) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385.
2. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **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. *Struct. Biol. Nature* **1994**, *1*, 169.
3. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (b) Geierstanger, B. H.; Jacobsen, J-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry*, **1994**, *33*, 3055.
4. Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.
5. (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646-650. (b) Mrksich, M.; Dervan, P.B.; *J. Am. Chem. Soc.* **1995**, *117*, 3325.
6. Mrksich, M.; Parks, M. E.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 7983.
7. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc., in press.*
8. Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc., in press.*
9. (a) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373. (b) Van Dyke, M. W.; Dervan, P. B. *Nucl. Acids Res.* **1983**, *11*, 5555.
10. (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. USA **1986**, *83*, 8462. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.

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

12. Resin substitution has been corrected for the weight of the polyamide chain. The change in the substitution level of the resin as the molecular weight of the polyamide increases is small for low substitution resin such as the 0.2 mmol/gram resin used here. The change in substitution during a specific coupling or for the entire synthesis 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.

13. (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.

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

CHAPTER SIX

Optimization of the Hairpin Polyamide Design for Recognition of the Minor Groove of DNA

The text of this chapter is partially taken from a submitted paper that was coauthored with Prof. Peter B. Dervan and Eldon E. Baird.

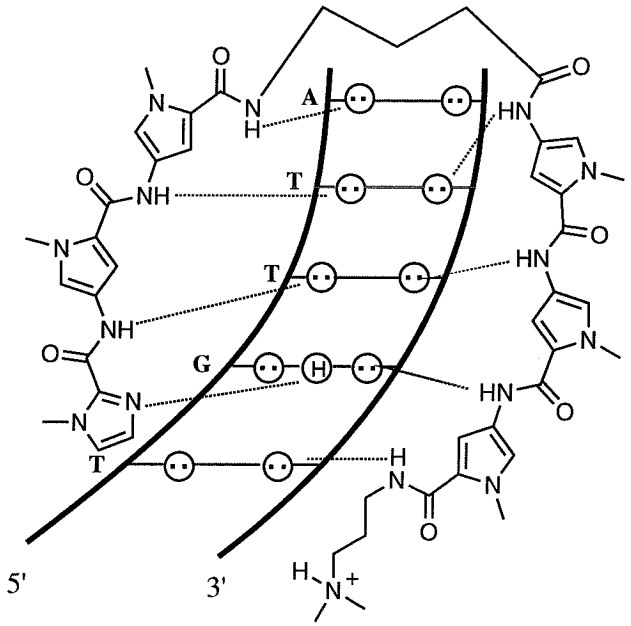
(Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *in press.*)

Introduction

Recently described 2:1 pyrrole-imidazole polyamide-DNA complexes combined with a convenient solid phase synthesis provide a powerful new paradigm for the design of artificial molecules for the sequence-specific recognition of DNA.[1,2] 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.[1,3-5] The DNA sequence-specificity of these small molecules can be controlled by the linear sequence of pyrrole and imidazole amino acids. An imidazole ring on one ligand complemented by a pyrrolecarboxamide ring on the second ligand recognizes a G•C base pair, while a pyrrolecarboxamide/ imidazole combination targets a C•G base pair.[1,4] A pyrrolecarboxamide/pyrrolecarboxamide pair is degenerate for A•T or T•A base pairs.[1,3-4]

Hairpin Polyamide Motif. A simple polyamide hairpin motif with γ-aminobutyric acid (γ) serving as a turn monomer provides a synthetically accessible method of covalently linking polyamide units within the 2:1 motif (Figure 1).[6] Covalently linked polyamide heterodimers and homodimers have both increased affinities and sequence specificity.[6,7] The polyamide ImPyPy-γ-PyPyPy-Dp **3** was found to bind a designated 5'-TGTTA-3' target site with high specificity and an approximate 300-fold binding enhancement over the individual unlinked polyamide pair; ImPyPy-Dp and AcPyPyPy-Dp.[6]

Solid Phase Polyamide Synthesis. During the process of developing solid phase methodology for the synthesis of the pyrrole-imidazole polyamides, it was necessary to identify a stable resin linkage agent that can be cleaved in high yield upon completion of the synthesis. Because the reactivity of an imidazole aromatic ester, a pyrrole aromatic ester, and the ester of an aliphatic amino acid are different, separate novel synthetic linkage agents would have to be developed to prepare polyamides with C-terminal imidazole and C-terminal pyrrole residues. Furthermore, negatively charged or uncharged pyrrole-



ImPyPy-γ-PyPyPy-Dp • TGTTA

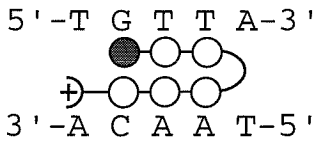
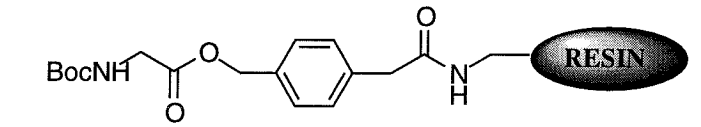


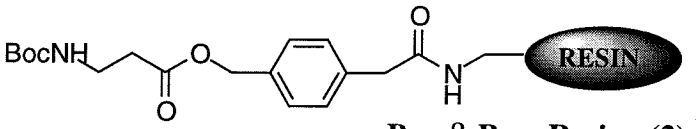
Figure 1. Hairpin-polyamide binding model for the complex formed between ImPyPy-γ-PyPyPy-Dp and a 5'-TGTTA-3' sequence (Top). Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. (Bottom) Ball and stick binding model, the imidazole and pyrrole rings are represented as shaded and unshaded circles respectively.

imidazole polyamides tend to be highly insoluble in both organic and aqueous solvents. Therefore, any resin cleavage strategy has to be a single step process which introduces a positive charge into the polyamide.

The addition of an aliphatic amino acid at the C-terminus of the pyrrole-imidazole polyamides provides a convenient alternative to developing new synthetic chemistry for linking aromatic esters to a solid support. *Tert*-butyloxycarbonylaminoacyl-4-(oxymethyl)phenylacetamidomethyl-resin (PAM resin) has been reported to be cleaved in high yield by aminolysis.[8] Preloaded Boc-β-alanine-Pam-Resin and Boc-glycine-Pam-Resin are commercially available in a variety of substitution levels (Figure 2).[9] A series of eight pyrrole-imidazole polyamides containing either a C-terminal glycine or β-alanine residue were prepared (Figure 3). We find that *N*-terminal acetylation can alter the relative affinities of single base pair mismatches. We report here that the polyamide synthesized from Boc-β-alanine-Pam-Resin, ImPyPy-γ-PyPyPy-β-Dp **7** binds with both enhanced affinity and specificity relative to the parent compound, ImPyPy-γ-PyPyPy-Dp **3**,[6] which lacks the C-terminal β-alanine residue.



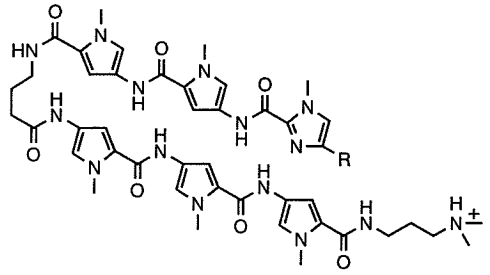
Boc-G-Pam Resin (1)



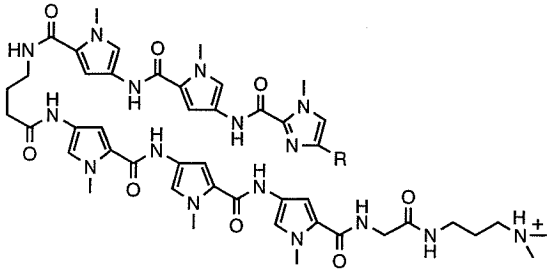
Boc-β-Pam Resin (2)

Figure 2. Commercially available preloaded amino acid resins (**1**) Boc-glycine-Pam-resin and (**2**) Boc-β-alanine-Pam-Resin.

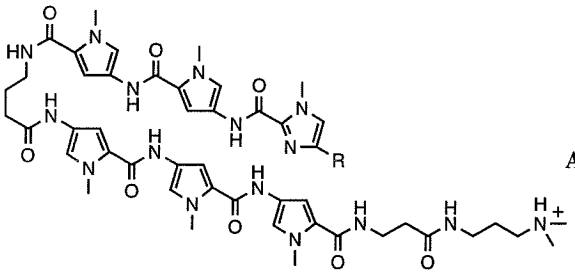
Synthesis of Polyamides. Eight polyamides containing C-terminal amino acids were prepared by solid phase methodology.[2] Three pyrrole and imidazole building blocks were combined in a stepwise manner on a solid support (Figure 4). Synthesis is



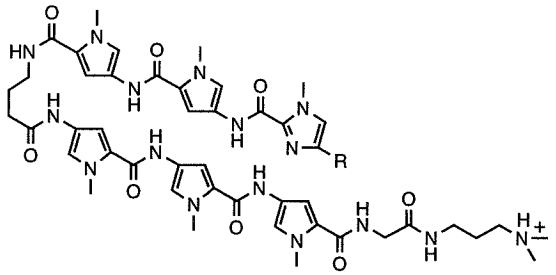
ImPyPy-γ-PyPyPy-Dp; R = H (3)
AcImPyPy-γ-PyPyPy-Dp; R = $NHCOCH_3$ (4)



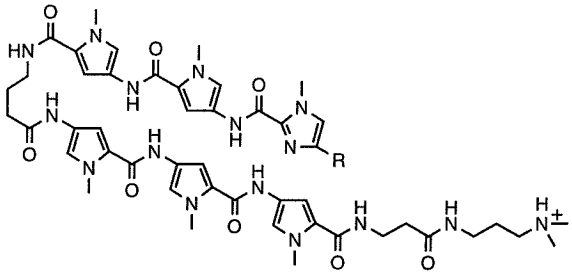
ImPyPy-γ-PyPyPy-G-Dp; R = H (5)
AcImPyPy-γ-PyPyPy-G-Dp; R = $NHCOCH_3$ (6)



ImPyPy-γ-PyPyPy-β-Dp; R = H (7)
AcImPyPy-γ-PyPyPy-β-Dp; R = $NHCOCH_3$ (8)



PyPyPy-γ-ImPyPy-G-Dp; R = H (9)
AcPyPyPy-γ-ImPyPy-G-Dp; R = $NHCOCH_3$ (10)



PyPyPy-γ-ImPyPy-β-Dp; R = H (11)
AcPyPyPy-γ-ImPyPy-β-Dp; R = $NHCOCH_3$ (12)

Figure 3. Hairpin pyrrole-imidazole polyamides, with C-terminal -Dp **3** and **4**, G-Dp **5**, **6**, **9** and **10**, and β-Dp, **7**, **8**, **11**, and **12** end groups.

exemplified for polyamide **8,** AcImPyPy-γ-PyPyPy-β-Dp, which was prepared in 17 steps on the resin, and then cleaved with a single step aminolysis reaction (Figure 4). All polyamides were found to be soluble up to at least 1 mM concentration in aqueous solution.

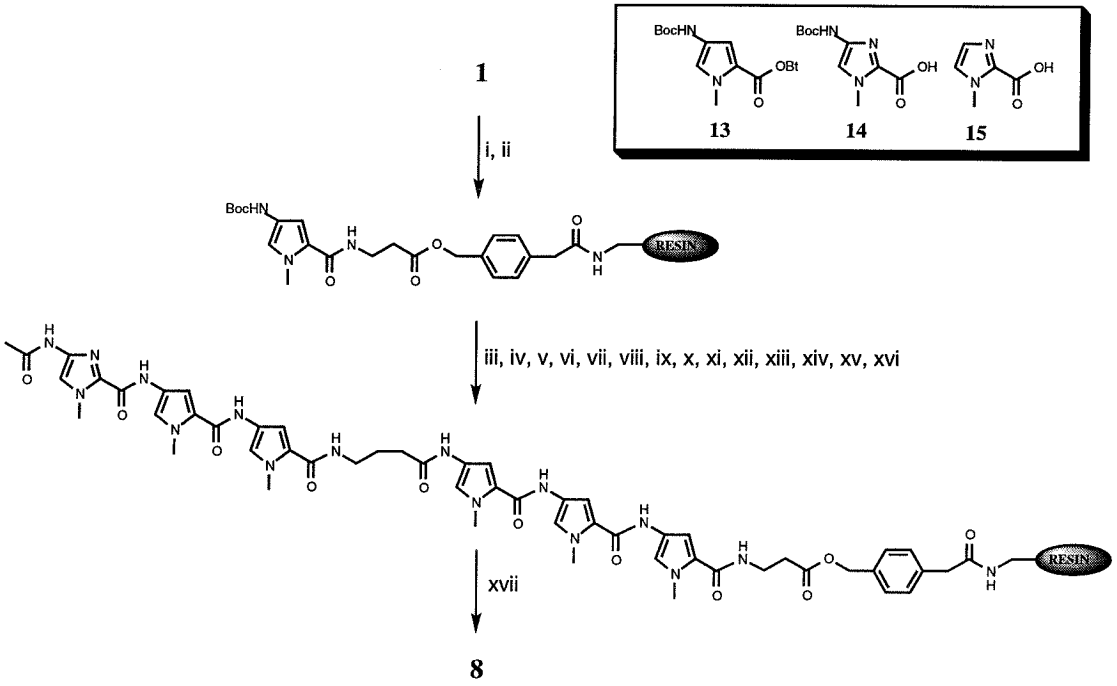
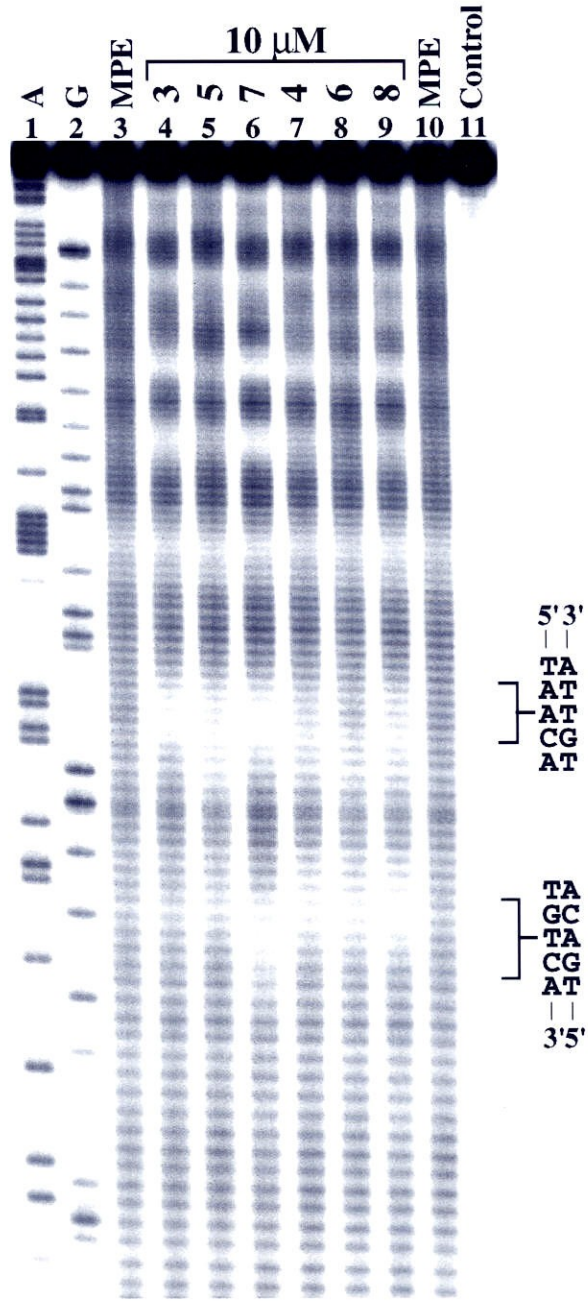


Figure 4. (Box) Pyrrole-imidazole monomers for synthesis of all compounds described here; Boc-Pyrrole-OBt ester **13**,[2] Boc-Imidazole-acid **14**,[2] and Imidazole-2-carboxylic acid **15**.[1a] Solid phase synthetic scheme for AcImPyPy-γ-PyPyPy-β-Dp starting from commercially available Boc-β-Pam-resin: (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); (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); BocPy-OBt, DIEA, DMF; (xv) 80% TFA/DCM, 0.4M PhSH; (xvi) Acetic anhydride, DMF, DIEA; (xvii) *N,N*- dimethylaminopropylamine, 55 °C.

Identification of Binding Sites by MPE•Fe (II) Footprinting. MPE•Fe (II) footprinting[10] on the 3' and 5'-^{32}P-end-labeled 135 base pair *Eco*RI/*Bsr*BI restriction fragments (25 mM Tris-acetate, 10 mM NaCl, 100 μM/ base pair calf thymus DNA at pH 7.0 and 22°C) reveal that the polyamides all bind to the 5'-TGTTA-3' match site (Figures 5,

Figure 5. MPE•Fe(II) footprinting of ImPyPy-γ-PyPyPy-Dp **3**, ImPyPy-γ-PyPyPy-G-Dp **5**, ImPyPy-γ-PyPyPy-β-Dp **7**, AcImPyPy-γ-PyPyPy-Dp **4**, AcImPyPy-γ-PyPyPy-G-Dp **6**, and AcImPyPy-γ-PyPyPy-β-Dp **8** on a 135 base pair *Eco*RI/*Bsr*BI restriction fragment. All reactions contain 10 kcpm restriction fragment, 25 mM tris-acetate, 10 mM NaCl, 100 μM calf thymus DNA (bp), and 5 mM DTT. Lane 1, A reaction; lane 2, G reaction; lanes 3 and 10, MPE•Fe(II) standard; lane 4, 10 μM **3**; lane 5, 10 μM **5**; lane 6, 10 μM **7**; lane 7, 10 μM **4**; lane 8, 10 μM **6**; lane 9, 10 μM **8**; lane 11, intact DNA.


A
G
MPE
10 μM
3
5
7
4
6
8
MPE
Control
1
2
3
4
5
6
7
8
9
10
11
5'3'
TA
AT
AT
CG
AT
TA
GC
TA
CG
AT
3'5'

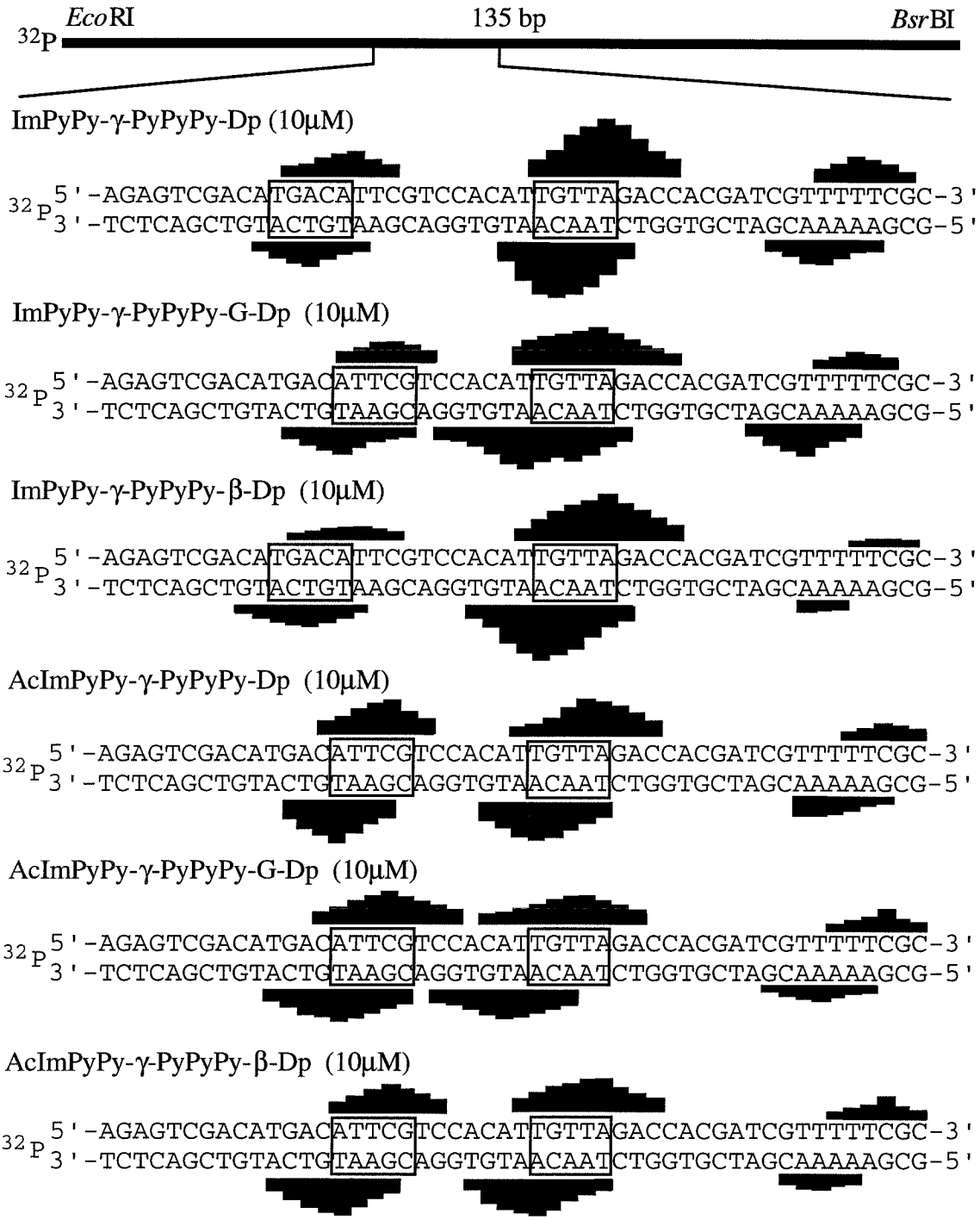


Figure 6. MPE•Fe(II) protection patterns for ImPyPy-γ-PyPyPy-Dp **3**, ImPyPy-γ-PyPyPy-G-Dp **5**, ImPyPy-γ-PyPyPy-β-Dp **7**, AcImPyPy-γ-PyPyPy-Dp **4**, AcImPyPy-γ-PyPyPy-G-Dp **6**, and AcImPyPy-γ-PyPyPy-β-Dp **8** on a 135 base pair *Eco*RI/*Bsr*BI restriction fragment. Bar heights are proportional to the protection from cleavage at each band. Boxes represent binding sites determined by the published model.[10]

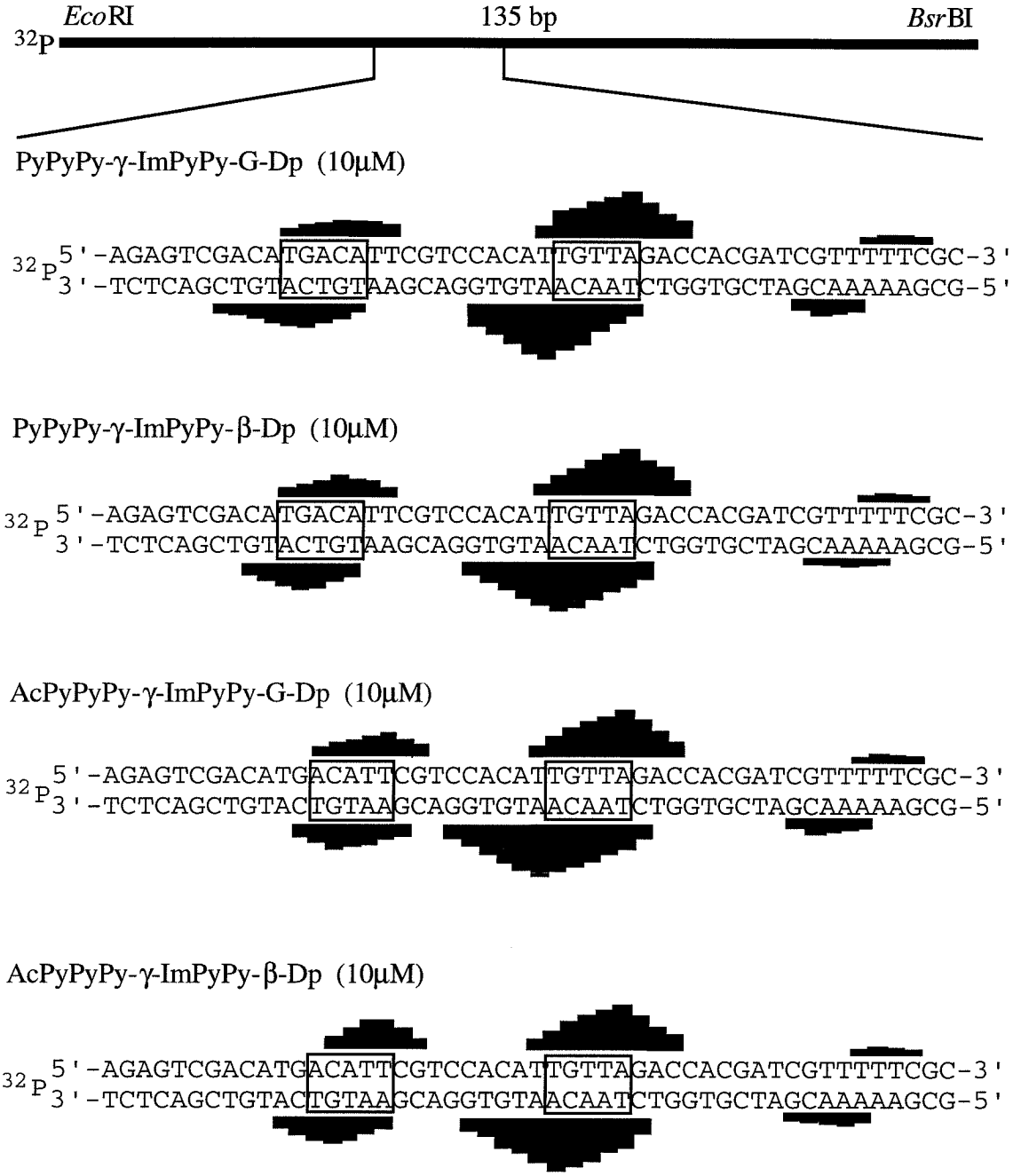


Figure 7. MPE•Fe(II) protection patterns for PyPyPy-γ-ImPyPyPy-G-Dp **9**, PyPyPy-γ-ImPyPy-β-Dp **11**, AcPyPyPy-γ-ImPyPy-G-Dp **10**, and AcPyPyPy-γ-ImPyPy-β-Dp **12** on a 135 base pair *Eco*RI/*Bsr*BI restriction fragment. Bar heights are proportional to the protection from cleavage at each band. Boxes represent binding sites determined by the published model.[10]

6, and 7). Single base pair mismatch site preferences varied for the polyamides as follows: ImPyPy-γ-PyPyPy-Dp **3**, ImPyPy-γ-PyPyPy-β-Dp **7**, PyPyPy-γ-ImPyPy-G-Dp **9**, and PyPyPy-γ-ImPyPy-β-Dp **11** bind a 5'-TGACA-3' site; AcImPyPy-γ-PyPyPy-Dp **4**, ImPyPy-γ-PyPyPy-G-Dp **5**, AcImPyPy-γ-PyPyPy-G-Dp **6**, and AcImPyPy-γ-PyPyPy-β-Dp **8** bind to a 5'-ATTCG-3' site; and AcPyPyPy-γ-ImPyPy-G-Dp **10** and AcPyPyPy-γ-ImPyPy-β-Dp **12** bind to a 5'-ACATT-3' site.

Analysis of Energetics by Quantitative DNase I Footprint Titrations. Quantitative DNase I footprinting[11] on the 3'-^{32}P-end-labeled 135 base pair *Eco*RI/*Bsr*BI restriction fragment (10 mM Tris•HCl, 10 mM KCl. 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0 and 22°C) reveals the following equilibrium association constants for the polyamides studied: ImPyPy-γ-PyPyPy-β-Dp **7** (2.9 x 10^8 M^{-1}) > ImPyPy-γ-PyPyPy-Dp **3** (7.6 x 10^7 M^{-1}) > AcPyPyPy-γ-ImPyPy-β-Dp **12** (6.8 x 10^7 M^{-1}) > AcImPyPy-γ-PyPyPy-Dp **4** (6.4 x 10^7 M^{-1}) > AcPyPyPy-γ-ImPyPy-G-Dp **10** (5.1 x 10^7 M^{-1}) > AcImPyPy-γ-PyPyPy-G-Dp **6** (4.0 x 10^7 M^{-1}) > AcImPyPy-γ-PyPyPy-β-Dp **8** (2.9 x 10^7 M^{-1}) > PyPyPy-γ-ImPyPy-β-Dp **11** (2.0 x 10^7 M^{-1}) > PyPyPy-γ-ImPyPy-G-Dp **9** (5.1 x 10^6 M^{-1}) > ImPyPy-γ-PyPyPy-G-Dp **5** (3.3 x 10^6 M^{-1}) (Tables I and II). Equilibrium association constants for the single base pair mismatch site 5'-TGACA-3' for ImPyPy-γ-PyPyPy-Dp **3**, ImPyPy-γ-PyPyPy-β-Dp **7**, PyPyPy-γ-ImPyPy-G-Dp **9**, and PyPyPy-γ-ImPyPy-β-Dp **11** are 3.2 x 10^6 M^{-1}, 4.8 x 10^6 M^{-1}, $\leq$ 1 x 10^5 M^{-1}, and 1.2 x 10^7 M^{-1}, respectively (Figures 8, 9, and 10). AcImPyPy-γ-PyPyPy-Dp **4**, ImPyPy-γ-PyPyPy-G-Dp **5**, AcImPyPy-γ-PyPyPy-G-Dp **6**, and AcImPyPy-γ-PyPyPy-β-Dp **8** bind to 5'-ATTCG-3' with affinities of 2.8 x 10^7 M^{-1}, 1.0 x 10^6 M^{-1}, 1.2 x 10^7 M^{-1}, and 1.8 x 10^7 M^{-1}, respectively. Finally, AcPyPyPy-γ-ImPyPy-G-Dp **10** and AcPyPyPy-γ-ImPyPy-β-Dp **12** bind to a 5'-ACATT-3' site with affinities of 1.6 x 10^7 M^{-1} and 8.7 x 10^6 M^{-1}, respectively.

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

	Match Site	**Single Mismatch Sites**	
Polyamide	5'-tTGTTAg-3'	5'-aTGACAt-3'	5'-cATTCGt-3'
ImPyPy-γ-PyPyPy-Dp	7.6×10^7 (0.8)	3.2×10^6 (0.7)	*c*
ImPyPy-γ-PyPyPy-G-Dp	3.3×10^6 (0.6)	*c*	1.0×10^6 (0.2)
ImPyPy-γ-PyPyPy-β-Dp	2.9×10^8 (0.5)	4.8×10^6 (1.1)	*c*
AcImPyPy-γ-PyPyPy-Dp	6.4×10^7 (3.0)	*c*	2.8×10^7 (0.5)
AcImPyPy-γ-PyPyPy-G-Dp	2.4×10^7 (2.2)	*c*	1.5×10^7 (1.2)
AcImPyPy-γ-PyPyPy-β-Dp	2.9×10^7 (0.5)	*c*	1.8×10^7 (0.2)

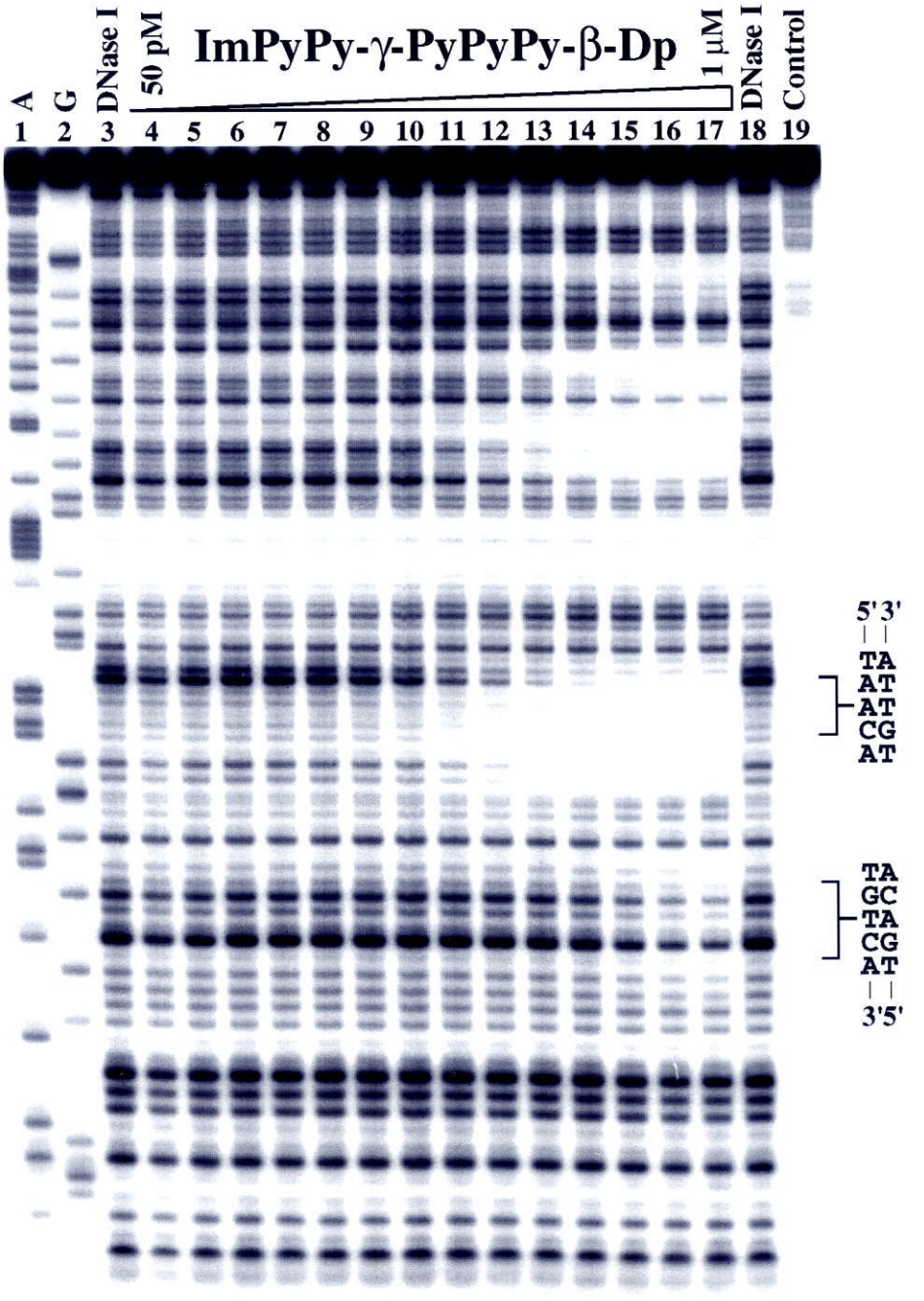
[a]Values reported are the mean values measured from at least three footprint titration experiments, with the standard deviation for each data set indicated in parentheses. [b]The assays were performed at 22 °C at pH 7.0 in the presence of 10 mM tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [c]The affinity for this site was not determined.

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

	Match Site	**Single Mismatch Sites**	
Polyamide	5'-tTGTTAg -3'	5'-aTGACAt-3'	5'-gACATTc-3'
PyPyPy-γ-ImPyPy-G-Dp	5.1×10^6 (1.2)	$\leq 1 \times 10^5$	*c*
PyPyPy-γ-ImPyPy-β-Dp	2.0×10^7 (0.5)	1.2×10^7	*c*
AcPyPyPy-γ-ImPyPy-G-Dp	5.1×10^7 (1.8)	*c*	1.6×10^7 (0.1)
AcPyPyPy-γ-ImPyPy-β-Dp	6.8×10^7 (2.1)	*c*	8.7×10^6 (2.9)

[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 $CaCl_2$. [c]The affinity for this site was not determined.

Figure 8. Quantitative DNase I footprint titration experiment with ImPyPy-γ-PyPyPy-β-Dp **7** on the 3'-^{32}P-labeled 135 base pair *Eco*RI/*BsrBI* restriction fragment from plasmid pMM5: lane 1, A reaction; lane 2, G reaction; lanes 3 and 18, DNase I standard; lanes 4-17, 50 pM, 100 pM, 200 pM, 500 pM, 1nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM ImPyPy-γ-PyPyPy-β-Dp, respectively; lane 19, intact DNA. The 5'-TGTTA-3' and 5'-TGACA-3' binding sites which were analyzed are shown on the right. All reaction contain 10 kcpm restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.


A
G
DNase I
50 pM
ImPyPy-γ-PyPyPy-β-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
5'3'
TA
AT
AT
CG
AT
TA
GC
TA
CG
AT
3'5'

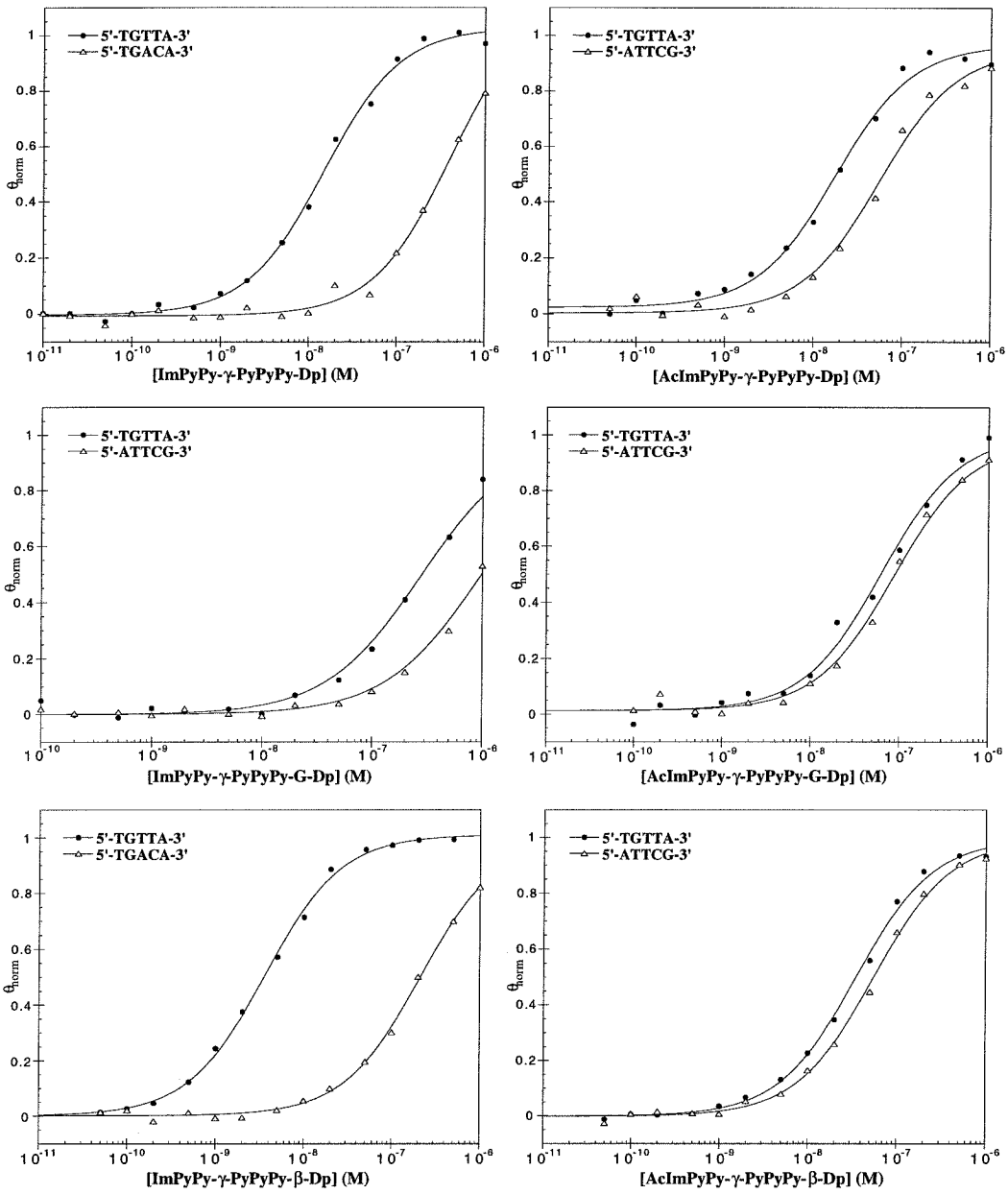


Figure 9. Data for the quantitative DNase I footprint titration experiments for the six polyamides in complex with the designated 5'-TGTTA-3' target site. The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography and processed as described in the experimental section. The data points for ImPyPy-γ-PyPyPy-Dp **3**, ImPyPy-γ-PyPyPy-G-Dp **5**, ImPyPy-γ-PyPyPy-β-Dp **7**, AcImPyPy-γ-PyPyPy-Dp **4**, AcImPyPy-γ-PyPyPy-G-Dp **6**, and AcImPyPy-γ-PyPyPy-b-Dp **8** are indicated by filled circles for the 5-TGTTA-3' match site, and open triangles for the 5-TGACA-3' or 5'-ATTCG-3' single base pair mismatch sites. The solid curves are the best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm using eq. 2.

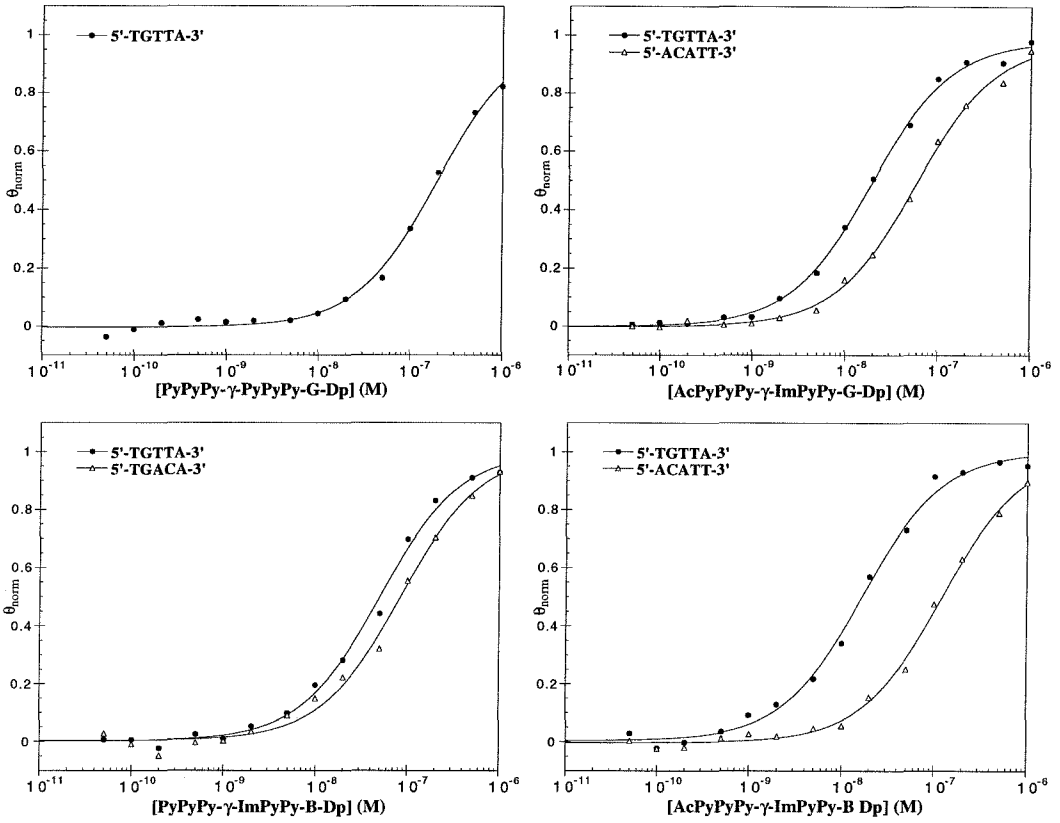


Figure 10. Data for the quantitative DNase I footprint titration experiments for the four polyamides in complex with the designated 5'-TGTTA-3' target site. The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography and processed as described in the experimental section. The data points for PyPyPy-γ-ImPyPy-G-Dp **9**, PyPyPy-γ-ImPyPy-β-Dp **11**, AcPyPyPy-γ-ImPyPy-G-Dp **10**, and AcPyPyPy-γ-ImPyPy-β-Dp **12** are indicated by filled circles for the 5-TGTTA-3' match site, and open triangles for the 5-TGACA-3' or 5'-ACATT-3' single base pair mismatch sites. The solid curves are the best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm using eq. 2.

Discussion

Binding Affinities. Among the ten polyamides, ImPyPy-γ-PyPyPy-β-Dp binds the targeted 5'-TGTTA-3' site with the highest affinity. This suggests that addition of a C-terminal linker *β-alanine* residue to facilitate solid phase polyamide synthesis is not merely an acceptable strategy, but serendipitously, *designates an optimized hairpin polyamide.* ImPyPy-γ-PyPyPy-β-Dp binds with an apparent equilibrium association constant, K_a = 3 x 10^8 M^{-1}, a factor of four greater than the parent polyamide, ImPyPy-γ-PyPyPy-Dp, K_a = 8 x 10^7 M^{-1}. A C-terminal glycine residue *reduces* binding affinity at the 5'-TGTTA-3' match site by a factor of 88 for ImPyPy-γ-PyPyPy-G-Dp binding relative to ImPyPy-γ-PyPyPy-β-Dp. The glycine residue may create a steric clash placing the glycine carbonyl in the floor of the minor groove (Figure 11). A high resolution NMR study of the 2:1 polyamide:DNA complex for ImPyPy-G-PyPyPy-Dp indicates considerable distortion of the polyamides to avoid placing the glycine carbonyl in the floor of the minor groove.[12] In contrast to glycine, a C- terminal β–alanine residue presents a steric surface which resembles that of the C-terminus of the original hairpin polyamide (Figure 11). The modest increased binding affinity of the C-terminal β-alanine polyamide, may result from an additional hydrogen bond between the β-alanine carboxamide and a 'sixth' base pair of the binding site. Evidence for such an interaction must await further structural studies.

Specificity. All ten polyamides containing a common sequence component, either ImPyPy-γ-PyPyPy or the reverse PyPyPy-γ-ImPyPy, but having different substitutions at the N and C -termini were found to be highly specific for the designated 5'-TGTTA-3' target site, indicating the pyrrole-imidazole polyamides can tolerate a variety of substitutions. The optimal was ImPyPy-γ-PyPyPy-β-Dp which binds the target 5'-TGTTA-3' match site with 60-fold specificity relative to a single base pair 5'-TGACA-3' mismatch site. This can be compared with the parent polyamide ImPyPy-γ-PyPyPy-Dp which has a 24-fold specific binding relative to the same sites.[6] An N-terminal acetyl

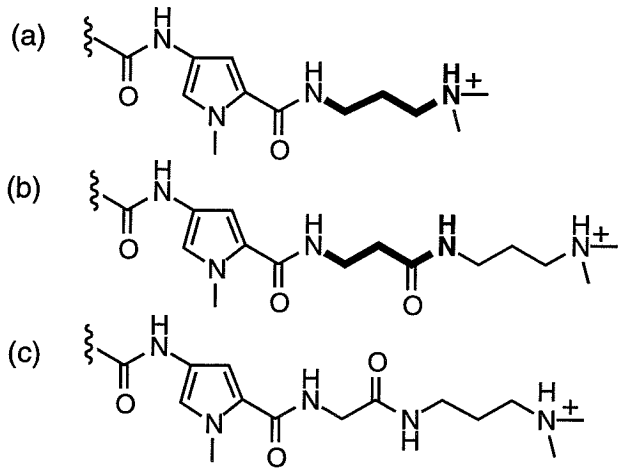


Figure 11. Modification of the C-terminus of pyrrole-imidazole polyamides: a) -Dp, b) G-Dp, c) β-Dp.

group or a C-terminal glycine group reduces the observed relative sequence specificity to between 1 and 3 fold. Furthermore, MPE•Fe(II) footprinting reveals the sequence of the mismatch binding site changes from 5'-TGACA-3', 5'-ATTCG-3', and 5'-ACATT-3' (Figures 6 and 7). For example, the acetylated polyamide, Ac-ImPyPy-γ-PyPyPy-β-Dp, binds the 5'-TGTTA-3' match site with 10-fold reduced affinity compared to ImPyPy-γ-PyPyPy-β-Dp, while binding a 5'-ATTCG-3' mismatch site with a 1.6-fold *increase* in affinity relative to the original 5'-TGACA-3' mismatch site (Figure 12). Binding at the 5'-TGACA-3' single base pair mismatch is greatly reduced by introducing the acetyl group, or C-terminal glycine group in all cases. The steric bulk of the acetyl group or glycine residue may simply be preventing the polyamide from sitting deeply in the minor groove, reducing the affinity for the match, while increasing the tolerance at the mismatch. Alternatively, there may be a favorable interaction between the acetyl group or glycine and the terminal G•C base pair of the 5'-ATTCG-3' mismatch site. Again, structural studies will be necessary to describe this interaction.

Implications for the Design of Minor-Groove Binding Molecules. Ten polyamides, with a variety of substitutions at the N and C -termini are found specific for a designated 5'-TGTTA-3' target site, indicating pyrrole-imidazole polyamides are tolerant to

a variety of chemical substitutions. Remarkably, the addition of a C-terminal β-alanine residue is found not only to facilitate synthesis, but also to enhance both the affinity and sequence-specificity of the pyrrole-imidazole polyamides. Coupling the hairpin polyamide model to rapid solid phase synthetic methodology will allow the design of pyrrole-imidazole polyamides for sequence-specific recognition of many different DNA sequences at nanomolar concentrations.

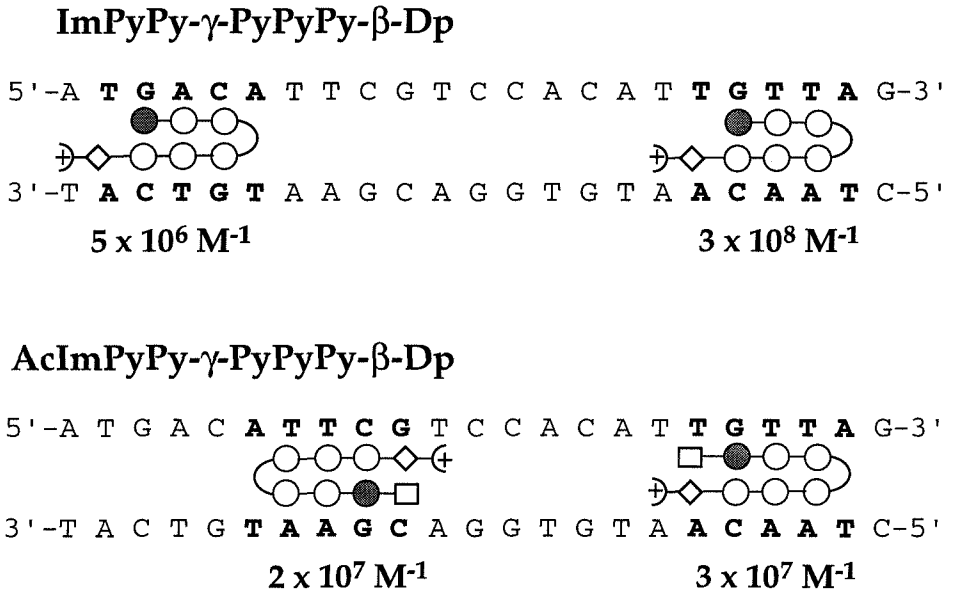


Figure 12. Schematic binding model for ImPyPy-γ-PyPyPy-β-Dp **7** and AcImPyPy-γ-PyPyPy-β-Dp **8** binding to the 5'-TGTTA-3' match site and either a 5'-TGTCA-3' or 5'-ATTCG-3' mismatch site, the imidazole and pyrrole rings are represented as shaded and unshaded circles, respectively; the β-alanine residue is represented as an unshaded diamond, and the acetyl group as an unshaded square.

Experimental Section

Materials. 0.2 mmol/gram Boc-Glycine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-G-Pam-Resin), 0.2 mmol/gram Boc-β-alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-β-Pam-Resin), Dicyclohexylcarbodiimide (DCC), Hydroxybenzo-triazole (HOBt), 2-(1H-

Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU), Boc-Glycine, and Boc-β-alanine were purchased from Peptides International. *N,N*-diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF), *N*-methylpyrrolidone (NMP), DMSO/NMP, Acetic anhydride (Ac_2O), and 0.0002M potassium cyanide/pyridine were purchased from Applied Biosystems. Boc-γ-aminobutyric acid was from NOVA Biochem, dichloromethane (DCM) and triethylamine (TEA) was reagent grade from EM, thiophenol (PhSH), dimethylaminopropylamine from Aldrich, trifluoroacetic acid (TFA) from Halocarbon, phenol from Fisher, and ninhydrin from Pierce. All reagents were used without further purification.

Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. and are used for filtration of DCU. Disposable polypropylene filters are also used for washing resin for ninhydrin and picric acid tests, and for filtering pre-dissolved amino acids into synthesis cartridges. A shaker for manual solid phase synthesis was obtained from Milligen. A rotary evaporator was also modified for use as a shaker. Screw- cap glass peptide synthesis reaction vessels (5 ml and 20 ml) with a #2 sintered glass frit were made at the Caltech glass shop as described by Kent.[12c] ^1H NMR were recorded in d_6-DMSO on a GE 300 instrument operating at 300 MHz. Chemical shifts are reported in ppm relative to the solvent residual signal. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix- assisted, laser desorption/ionization time of flight mass spectrometry was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed either on a HP 1090M analytical HPLC or a Beckman Gold system using a Rainin 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 HPLC was carried out on a Beckman HPLC using a Waters DeltaPak 25 x 100 mm, 100 μm C_{18} 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. Reagent-grade chemicals were used unless otherwise stated.

Activation of Boc-γ-aminobutyric acid. The appropriate amino acid or acid (2 mmol) was dissolved in 2 ml DMF. HBTU (720 mg, 1.9 mmol) was added followed by DIEA (1 ml) and the solution lightly shaken for at least 5 min.

Activation of Boc-Imidazole acid. Boc imidazole acid (257 mg, 1 mmol) and HOBt (135 mg, 1 mmol) were dissolved in 2 ml DMF, DCC (202 mg, 1 mmol) is then added and the solution allowed to stand for at least 5 minutes.

Typical Manual Synthesis Protocol: AcImPyPy-γ-PyPyPy-β-Dp. Boc-β-Pam-resin (1.25 g, 0.25 mmol amine) was shaken in DMF for 30 min and drained. The N-Boc group removed by washing with DCM for 2 x 30 s, followed by a 1 min shake in 80% TFA/DCM/0.5M PhSH, draining the reaction vessel and a brief 80% TFA/DCM/ 0.5M PhSH wash, and 20 min shaking in 80% TFA/DCM/0.5M PhSH solution. The Resin was washed 1 min with DCM and 30 s with DMF. A resin sample (8-10 mg) was taken for analysis. The resin was drained completely and Boc-pyrrole-OBt monomer (357 mg, 1 mmol) dissolved in 2 ml DMF added followed by DIEA (1 ml.) and the resin shaken vigorously to make a slurry. The coupling was allowed to proceed for 45 min. A resin sample (8-10 mg) was taken after 40 min to check reaction progress. The reaction vessel was washed with DMF for 30 s and dichloromethane for 1 min to complete a single reaction cycle. Six additional cycles were performed adding, BocPy-OBt, BocPy-OBt, Boc-γ-aminobutyric acid-OBt (activated *in situ* with HBTU), BocPy-OBt, BocPy-OBt, and BocIm-OBt. The resin was washed with DMF, DCM, MeOH, and ethyl ether and then dried *in vacuo*. Approximately 300 mg of the BocImPyPy-γ-PyPyPy-β-Pam resin was removed from the reaction vessel, the remaining resin was stored at -20 °C for future use. The N-terminal Boc group was removed as described above, the resin washed with DCM for 30 sec., DMF, 1 min, a resin sample (8-10 mg) taken for analysis, and the resin treated with 2 ml DMF, 500 μl acetic anhydride, and 1M DIEA for 1 hour, and a resin

sample (8-10 mg) taken for analysis. The reaction vessel was washed with DMF (1 min), DCM (1 min), MeOH (1 min), ethyl ether (1 min.) and dried *in vacuo*. AcImPyPy-γ-PyPyPy-β-Pam-Resin (180 mg, 29 μmol)[13] was weighed into a glass scintillation vial, 1.5 ml of *N,N*-dimethylaminopropylamine added, and the mixture heated at 55 °C for 18 hours. The resin was removed by filtration through a disposable polypropylene filter and washed with 5 ml of water, the amine solution and the water washes combined, and the solution loaded on a C_{18} preparatory HPLC column, the column allowed to wash for 4 min in 0.1% TFA at 8 ml/min, the polyamide was then eluted in 100 min. as a well defined peak with a gradient of 0.25% acetonitrile per min. The polyamide was collected in two separate 8 ml fractions, the purity of the individual fractions verified by HPLC and ^{1}H NMR, to provide purified AcImPyPy-γ-PyPyPy-β-Dp (**8**). (9.2 mg 31% yield), UV λ_{max}, 246 (42,800), 312 (50,400) HPLC, r.t. 24.9, ^{1}H NMR (DMSO-d_6) δ 10.25 (s, 1 H), 10.01 (s, 1 H), 9.92 (m, 3 H), 9.86 (s, 1 H), 9.3 (br s, 1 H), 8.10 (m, 3 H), 7.42 (s, 1 H), 7.25 (d, 1 H, J = 1.5 Hz), 7.20 (d, 1 H, J = 1.6 Hz), 7.16 (m, 3 H), 7.12 (d, 1 H, J = 1.4 Hz), 7.03 (d, 1 H J = 1.7), 6.89 (d, 1 H, J = 1.6 Hz), 6.86 (m, 2 H), 3.92 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.80 (s, 6H), 3.78 (s, 3 H), 3.35 (q, 2 H, J = 5.5 Hz), 3.20 (q, 2 H, J = 3.8 Hz), 3.08 (q, 2 H, J = 3.3 Hz), 2.97 (q, 2 H, J = 3.8 Hz), 2.75 (d, 6 H J = 4.8 Hz), 2.34 (t, 2 H, J = 5.0 Hz), 2.24 (t, 2 H, J = 4.4 Hz), 2.00 (s, 3 H), 1.71 (m, 4 H); MALDI-TOF-MS, 1035.4 (1035.1 calc. for M+H).

ImPyPy-γ-PyPyPy-G-Dp (4) (12 mg, 40% recovery). HPLC, r.t. 26.9, UV λ_{max}, 246 (41,100), 306 (51,300) ^{1}H NMR (DMSO-d_6) δ 10.50 (s, 1 H), 9.95 (s, 1 H), 9.93 (s, 1 H), 9.92 (s, 1 H), 9.86 (s, 1 H), 9.2 (br s, 1H), 8.29 (t, 1 H, J = 4.4 Hz), 8.07 (t, 1 H, J = 5.2 Hz), 8.03 (t, 1 H, J = 5.4 Hz), 7.39 (s, 1 H), 7.27, (d, 1 H, J = 1.6 Hz), 7.22 (m, 2 H), 7.16 (m, 2 H), 7.04 (m, 2 H), 6.92 (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), 3.97 (s, 3 H), 3.82 (m, 6 H), 3.81 (s, 3 H), 3.78 (m, 6 H), 3.70 (d, 2 H, J = 5.7 Hz), 3.20 (q, 2 H, J = 5.7), 3.11 (q, 2 H, J = 4.2 Hz), 3.00 (q, 2 H, J = 4.4

Hz), 2.76 (d, 6 H, J = 4.7 Hz), 2.24 (t, 2 H, J = 4.8 Hz), 1.77 (m, 4 H); MALDI-TOF-MS, 964.3 (964.1 calc. for M+H).

AcImPyPy-γ-PyPyPy-G-Dp (5), (13.1 mg, 30% yield) HPLC, r.t. 24.0, UV λ_{max}, 246 (35,900), 312 (48,800) ^{1}H NMR (DMSO-d_6) δ 10.23 (s, 1 H), 9.98 (s, 1 H), 9.32 (s, 1 H), 9.90 (m, 2 H), 9.84 (s, 1 H), 9.2 (br s, 1 H), 8.27 (t, 1 H, J = 5.0), 8.05 (m, 2 H), 7.41 (s, 1 H), 7.25 (d, 1 H, J = 1.4 Hz), 7.22 (m, 2 H), 7.16 (m, 2 H), 7.12 (d, 1 H, J = 1.7 Hz), 7.05 (d, 1 H, J = 1.5 Hz), 6.94 (d, 1 H, J = 1.6 Hz), 6.89 (d, 1 H, J = 1.7 Hz) 6.87 (d, 1 H, J = 1.6 Hz), 3.93 (s, 3 H), 3.83 (s, 3 H), 3.82 (m, 6 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.71 (d, 2 H, J = 5.1 Hz), 3.19 (q, 2 H, J = 5.8 Hz), 3.12 (q, 2 H, J = 5.0 Hz), 3.01 (q, 2 H, J = 4.2 Hz), 2.74 (d, 6 H, J = 4.6 Hz), 2.26 (t 2 H, J = 4.6 Hz), 2.00 (s, 3 H), 1.75 (m, 4 H); MALDI-TOF-MS, 1021.6 (1021.1 calc. for M+H).

Characterization of polyamides not listed here can be found in Eldon Baird's thesis.

DNA Reagents and Materials. Sonicated, deproteinized calf thymus DNA was obtained from Pharmacia. Enzymes were purchased from Boehringer-Mannheim and used with the buffers supplied. Deoxyadenosine 5'-[α-^{32}P]-triphosphate, thymidine 5'-[α-^{32}P]-triphosphate, and adenosine 5'-[γ-^{32}P]-triphosphate were obtained from Amersham. Storage phosphor autoradiography was performed using a Molecular Dynamics 400S Phosphorimager and ImageQuant software. The 135 base pair 3' and 5' end labeled *Eco*RI/*Bsr*BI restriction fragments from plasmid pMM5 were prepared and purified as follows. Plasmid DNA was linearized using *Eco*RI, followed by treatment with either Klenow, deoxyadenosine 5'-[α-^{32}P]-triphosphate, and thymidine 5'-[α-^{32}P]-triphosphate for 3' labeling or calf alkaline phosphatase and subsequent 5' end labeling with T4 polynucleotide kinase and γ-^{32}P-dATP. The linearized plasmid DNA was digested with *Bsr*BI and the 135 base pair *Eco*RI/*Bsr*BI restriction fragment was isolated by nondenaturing 5% polyacrylamide gel electrophoresis (PAGE). The gel bands were visualized by autoradiography, isolated, and filtered to remove the polyacrylamide. The resulting solution was further purified by phenol extraction followed by ethanol

precipitation. Chemical sequencing reactions were performed according to published protocols.[14] Standard protocols were used for all DNA manipulations.[15]

MPE•Fe(II) Footprinting. All reactions were carried out in a total volume of 40 μL with final concentrations of species as indicated in parentheses. The ligands were added to solutions of radiolabeled restriction fragment (10 000 cpm), calf thymus DNA (100 μM bp), tris-acetate (25 mM, pH 7.0), and NaCl (10 mM) and incubated for 1 h at 22°C. A 50 μM MPE•Fe(II) solution was prepared by mixing 100 μL of a 100 μM MPE solution with a freshly prepared 100 μM ferrous ammonium sulfate solution. Footprinting reactions were initiated by the addition of MPE•Fe(II) (5μM), followed 5 min later by the addition of dithiothreitol (5 mM), and allowed to proceed for 15 min at 22°C. Reactions were stopped by ethanol precipitation, resuspended in 100 mM Tris-borate-EDTA/80% formamide loading buffer, and electrophoresed on 8% polyacrylamide denaturing gels (5% crosslink, 7 M urea) at 2000 V for 1 h. The gels were analyzed using storage phosphor technology.

Analysis of Energetics by Quantitative DNase I Footprint Titration. All reactions were executed in a total volume of 40 μL with final concentrations of each species as indicated. The ligands, ranging from 50 pM to 1 μM, were added to solutions of radiolabeled restriction fragment (10,000 cpm), tris-HCl (10 mM, pH 7.0), KCl (10 mM), $MgCl_2$ (10 mM) and $CaCl_2$ (5 mM) and incubated for 4 h at 22 °C. Footprinting reactions were initiated by the addition of 4 μL of a stock solution of DNase I (0.025 units/mL) containing 1 mM dithiothreitol and allowed to proceed for 6 min at 22 °C. The reactions were stopped by addition of a 3 M sodium acetate solution containing 50 mM EDTA and ethanol precipitated. The reactions were resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer and electrophoresed on 8% polyacrylamide denaturing gels (5% crosslink, 7 M urea) at 2000 V for 1 hr. The footprint titration gels were dried and quantitated using storage phosphor technology.

Apparent equilibrium association constants were determined as previously described.[6,11] The data were analyzed by performing volume integrations of the 5'-

TGTTA-3', 5'-TGACA-3', 5'-ATTCG-3', and 5'-ACATT-3' sites and a reference site. The apparent DNA target site saturation, θ_{app}, was calculated for each concentration of polyamide using the following equation:

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

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{ref}° correspond to those values for a DNase I control lane to which no polyamide has been added. The ($[L]_{tot}$, θ_{app}) data points were fit to a Langmuir binding isotherm (eq 2, n=1) 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 $[L]_{tot}$ corresponds to the total polyamide concentration, K_a corresponds to the apparent 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) running on a Power Macintosh 6100/60AV computer 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. 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:

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

Quantitation by Storage Phosphor Technology Autoradiography. Photostimulable storage phosphorimaging plates (Kodak Storage Phosphor Screen S0230 obtained from Molecular Dynamics) were pressed flat against gel samples and exposed in the dark at 22°C for 12-16 h. A Molecular Dynamics 400S PhosphorImager was used to obtain all data from the storage screens. The data were analyzed by performing volume integrations of all bands using the ImageQuant v. 3.2 software running on an AST Premium 386/33 computer.

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., USA* **1992**, *89*, 7586. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385.
2. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc. in press.*
3. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **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. *Struct. Biol. Nature* **1994**, *1*, 169.
4. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (b) Geierstanger, B. H.; Jacobsen, J. P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E.; *Biochemistry* **1994**, *33*, 3055. (c) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.
5. (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646. (b) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 3325.
6. Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983.
7. (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. (b) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663.
8. Mitchell, A. R.; Kent, S. B. H.; Engelhard, M.; Merrifield, R. B. *J. Org. Chem.* **1978**, *43*, 2845.
9. Available in bulk from: Peptides International, Louisville, Kentucky.

10. (a) Van Dyke, M. W.; Hertzberg, R. P.; and Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 5470. (b) Van Dyke, M. W.; Dervan, P. B. *Science* **1984**, *225*, 1122.

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

12. Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Nature Struct. Biol.* **1996**, *3*, 321.

13. Resin substitution has been corrected for the weight of the polyamide chain. The change in the substitution level of the resin as the molecular weight of the polyamide increases is small for low substitution resin such as the 0.2 mmol/gram resin used here. The change in substitution during a specific coupling or for the entire synthesis 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.

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

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

CHAPTER SEVEN

Simultaneous Binding of a Polyamide Dimer and an Oligonucleotide in the Minor and Major Grooves of DNA

The text of this chapter is partially taken from a paper that was coauthored with Prof. Peter B. Dervan.

(Parks, M. E.; Dervan, P. B. *Bioorg. & Med. Chem.* **1996**, *in press*.)

Introduction

High resolution X-ray crystal structures of protein-DNA complexes have revealed a diverse repertoire of structures for DNA recognition.[1-3] Proteins such as BZIP bind sequence-specifically in the major groove,[1] while others, such as TBP, bind exclusively in the minor groove.[2] Still others, such as Hin, make specific contacts *simultaneously* in the major and minor grooves of double helical DNA.[3] In the area of non-natural ligand design for DNA recognition, it has been demonstrated that oligonucleotides bind sequence specifically in the major groove of DNA by triple helix formation.[4] Antiparallel polyamide dimers bind sequence-specifically in the minor groove.[5-7] The compatibility of oligonucleotide-directed triple helix formation with simultaneous binding of a polyamide dimer in the minor groove has not yet been explored.

Pyrimidine Motif Triple Helix Formation. Oligonucleotide-directed triple helix formation is a valuable technique for sequence-specific recognition of double helical DNA.[4] One recognition motif involves a short pyrimidine oligonucleotide binding parallel to a purine-rich strand in the major groove of Watson-Crick duplex DNA through the formation of a local triple helix.[4a,8,9] Sequence specificity is derived from Hoogsteen base pairing between the pyrimidine bases of the oligonucleotide and the purine strand of duplex DNA to form T•(AT) and C^+(GC) triplets.[8,9]

Minor Groove Binding Polyamides. Recently, polyamides containing *N*-methylpyrrole and *N*-methylimidazole amino acids have been shown to bind as antiparallel, side-by-side dimers in the minor groove of DNA.[5-7] Within the 2:1 polyamide:DNA model, an imidazole on one ligand opposite a pyrrole carboxamide on the second ligand recognizes a G•C base pair, while a pyrrole carboxamide/imidazole combination targets a C•G base pair.[5,7] A pyrrole carboxamide/pyrrole carboxamide pair is partially degenerate

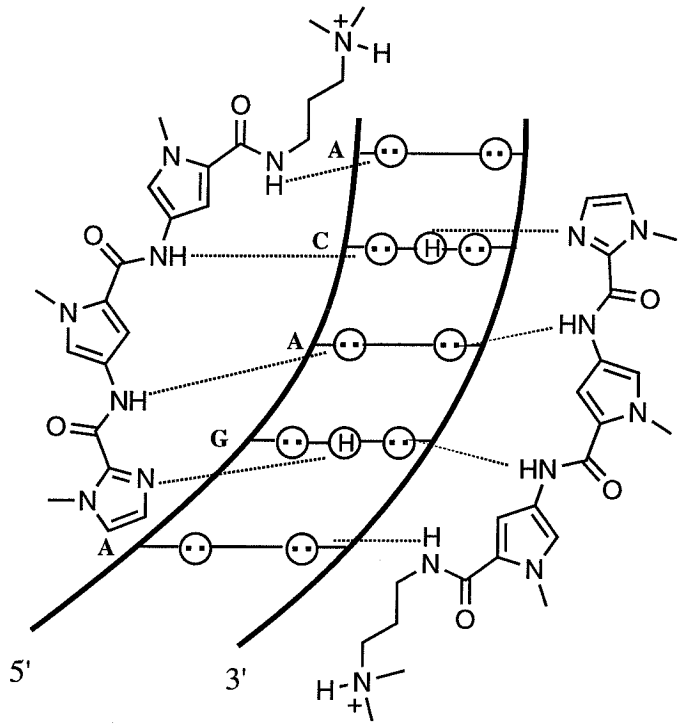
for A•T and T•A base pairs.[5-7] The polyamide ImPyPy-Dp has been shown to bind to the mixed sequence 5'-(A,T)G(A,T)C(A,T)-3' with a first-order association constant of ~2 x 10^5 M^{-1} (Figure 1).[5]

Simultaneous Binding of Ligands to DNA. Several studies have reported that binding of the natural product distamycin or netropsin in the minor groove of DNA is compatible with triple helix formation, although the small molecule thermally destabilizes the triple helix.[10] However, these studies were generally performed with short A,T rich duplexes. The minor groove width of DNA is approximately 3.4 Å for netropsin binding as a 1:1 complex to A,T rich sequences[11] compared with 6.8 Å for models of 2:1 polyamide-DNA complexes.[5,6] Therefore, 2:1 binding polyamides may have a very different effect on triple helix stability than the 1:1 binding natural products.

As a minimum first step toward the creation of non-natural ligands which bind specifically and simultaneously in the major and minor grooves of DNA, we have carried out a study aimed at determining whether major groove binding oligonucleotides and minor groove binding polyamides are structurally compatible in overlapping sequence space on opposite sides of a double helix. We report here the energetics of triple helix formation by a pyrimidine oligonucleotide bound in the major groove in the absence and presence of a polyamide dimer $(ImPyPy\text{-}Dp)_2$ or a distamycin monomer bound in the minor groove of an overlapping binding site.

Results and Discussion

Two plasmids (pMEP-2-ImN and pMEP-dist) have been constructed which contain two identical oligonucleotide binding sites, one overlapping a minor groove binding site (proximal **B** or **E**) and a control site thirteen or eleven base pairs away, respectively



$(ImPyPy\text{-}Dp)_2 \bullet AGACA$

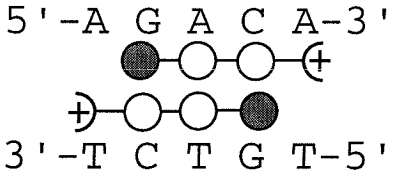


Figure 1. Binding model for the 2:1 complex of polyamide $(ImPyPy\text{-}Dp)_2$ with a 5'-AGACA-3' sequence. Circles with dots represent lone pairs of N3 purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dashed lines. Ball and stick models are also shown. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively.

(distal **C** or **F**) (Figures 2 and 3). The target binding site for ImPyPy-Dp **A** ovelaps site **B** by one base pair, while the distamycin target site **D** overlaps site **E** by two base pairs. Quantitative DNase I footprint titrations[12] were performed on the 255 base pair *Eco*RI/*Pvu*II and the 800 base pair *Eco*RI/*Xmn*I base pair restriction fragment from pMEP-2-ImN and pMEP-dist, respectively. The results of a typical experiment with oligonucleotide **1** in the absence of ImPyPy-Dp are shown in Figure 4A. The DNase I cleavage patterns are the same for both triple helix binding sites. In addition, the apparent first-order association constants of oligonucleotide **1** for both site **B** and site **C** in the absence of ImPyPy-Dp are K_a = ~2.7 x 10^7 M^{-1} (Table 1). In the presence of 50 μM ImPyPy-Dp, the binding constant for oligonucleotide **1** remains the same within experimental error for both sites **B** and **C** (Table 1). Results of footprinting oligonucleotide **1** in the presence of 100 μM ImPyPy-Dp are shown in Figure 4B. ImPyPy-Dp protects site **A** from cleavage, indicating that the dimeric polyamide complex forms in the presence of **1**. Also, the equilibrium association constant of oligonucleotide **1** for both **B** and **C** sites is unchanged in the presence of ImPyPy-Dp, consistent with simultaneous binding of the triple helix and ImPyPy-Dp (Figure 4). The overlaid binding isotherms for oligonucleotide **1** for sites **B** and **C** in the absence and presence of ImPyPy-Dp are shown in Figure 5. Similarly, the binding constants for oligonucleotide **2** in the absence and presence of distamycin do not change significantly (Table 2).

The results described above show that binding of ImPyPy-Dp or distamycin A in the minor groove is compatible with triple helix formation in the major groove at an overlapping site, although no cooperativity was observed in this system. These findings are consistent with previous results, which showed that ImPyPy-Dp is compatible with the major groove binding protein GCN4, but no cooperative interactions are observed.[13]

Figure 2. Ribbon model of fragment showing binding sites for ImPyPy-Dp **A**, oligonucleotide **1** (proximal **B**), and oligonucleotide **1** (distal **C**). ImPyPy-Dp is represented in the ribbon model by antiparallel black arrows.

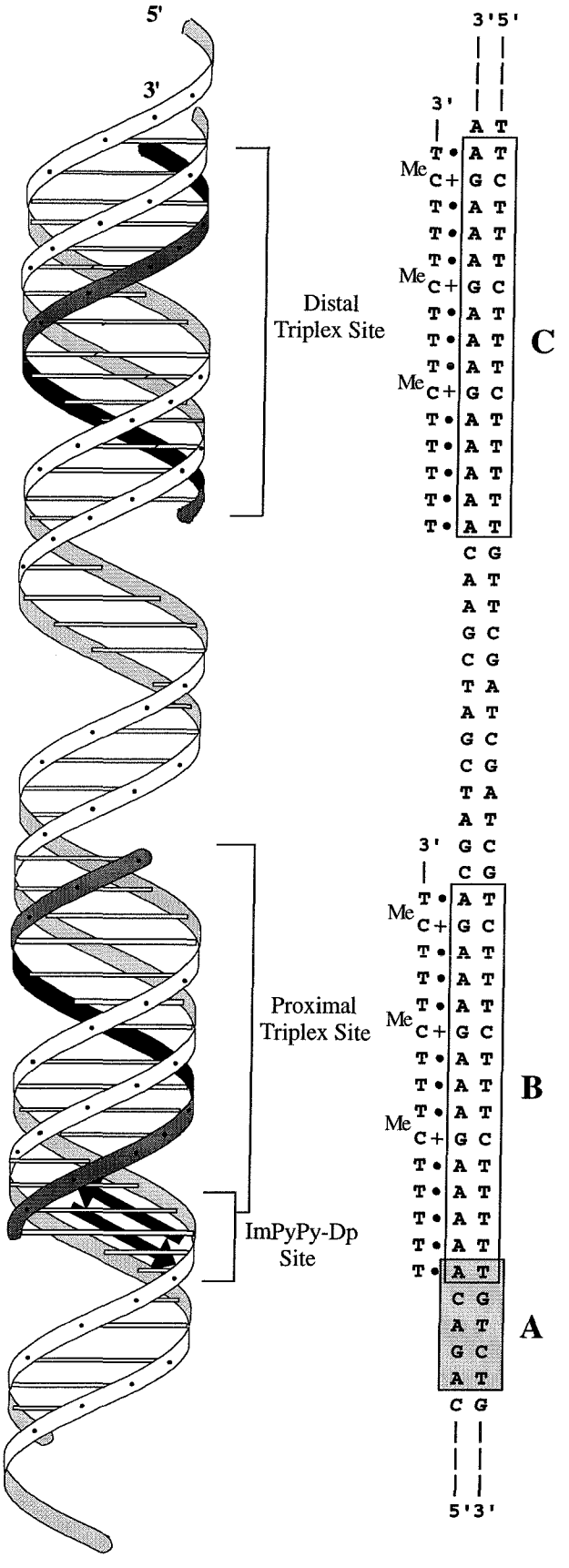

5'
3'
Distal
Triplex Site
Proximal
Triplex Site
ImPyPy-Dp
Site
3'5'
3'
Me
C
B
A
5'3'

Figure 3. Ribbon model of fragment showing binding sites for distamycin A **D**, oligonucleotide **2** (proximal **E**), and oligonucleotide **2** (distal **F**). Distamycin A is represented in the ribbon model by a black arrow.

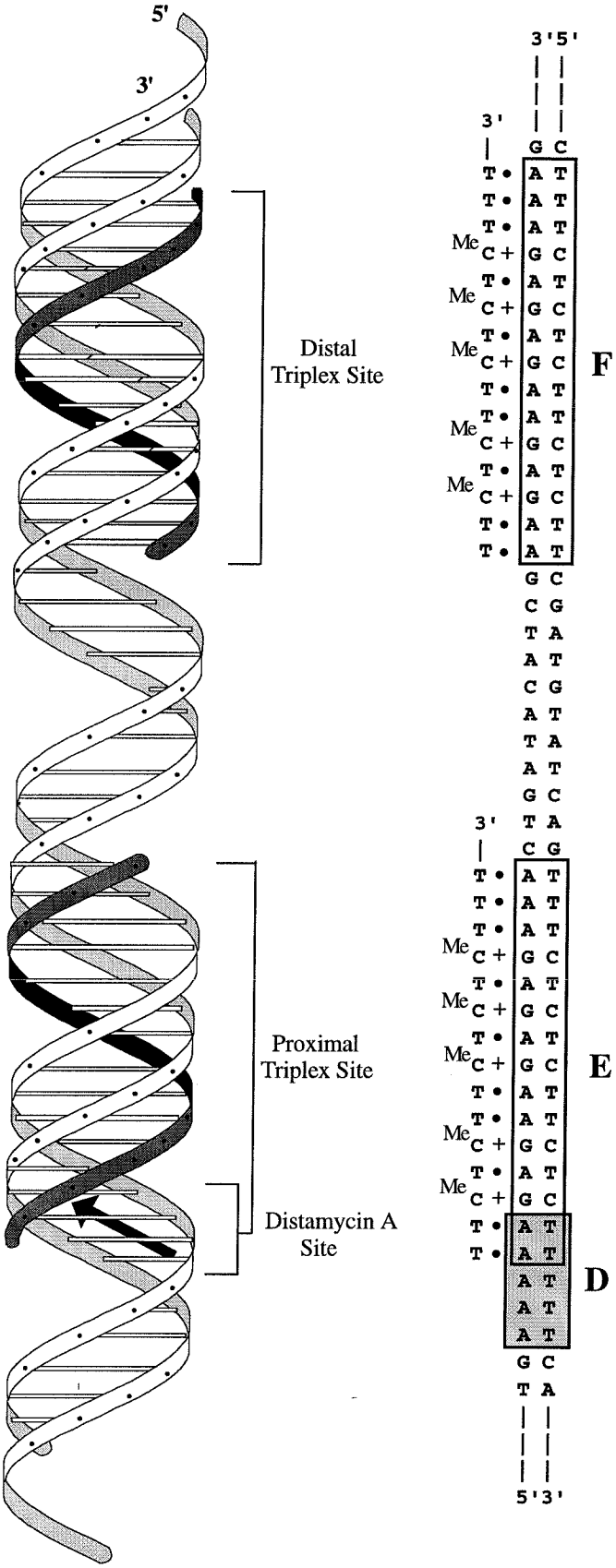
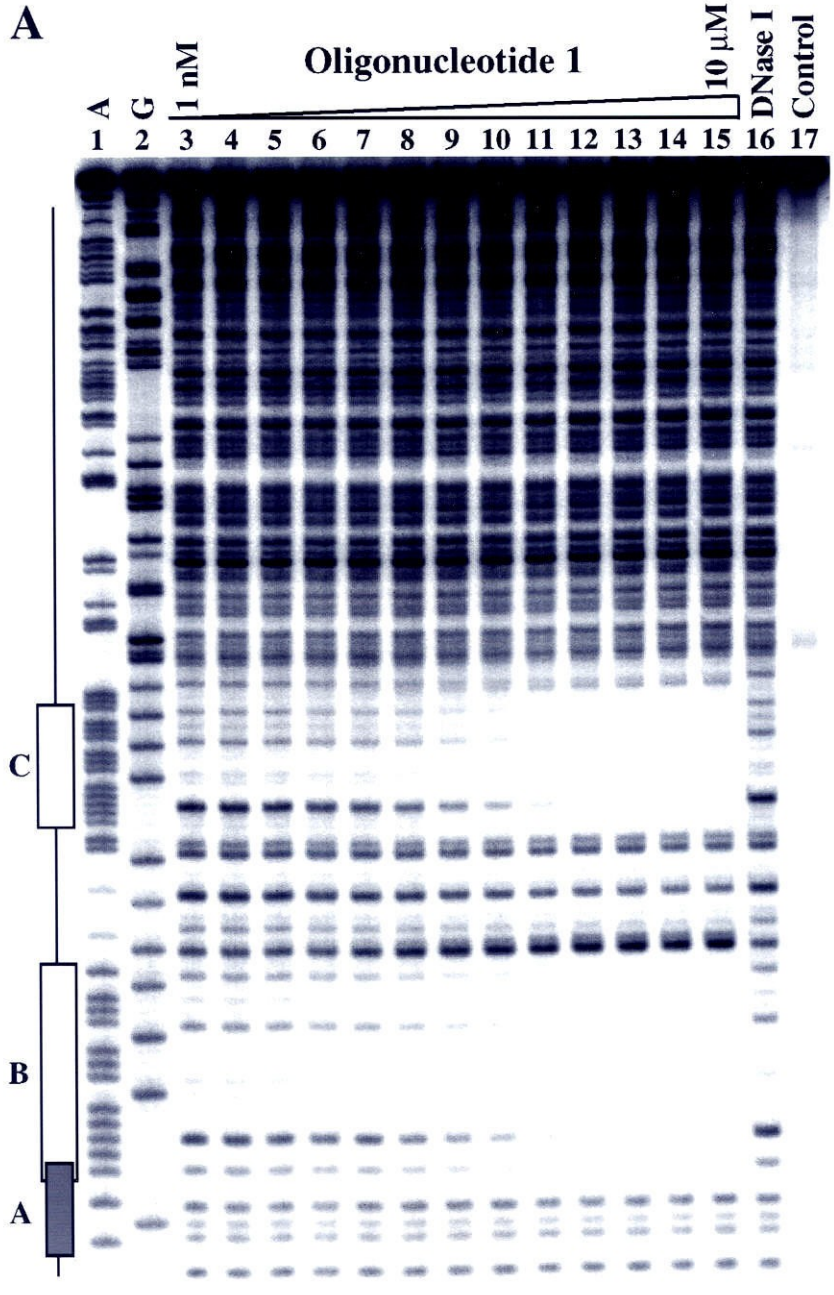
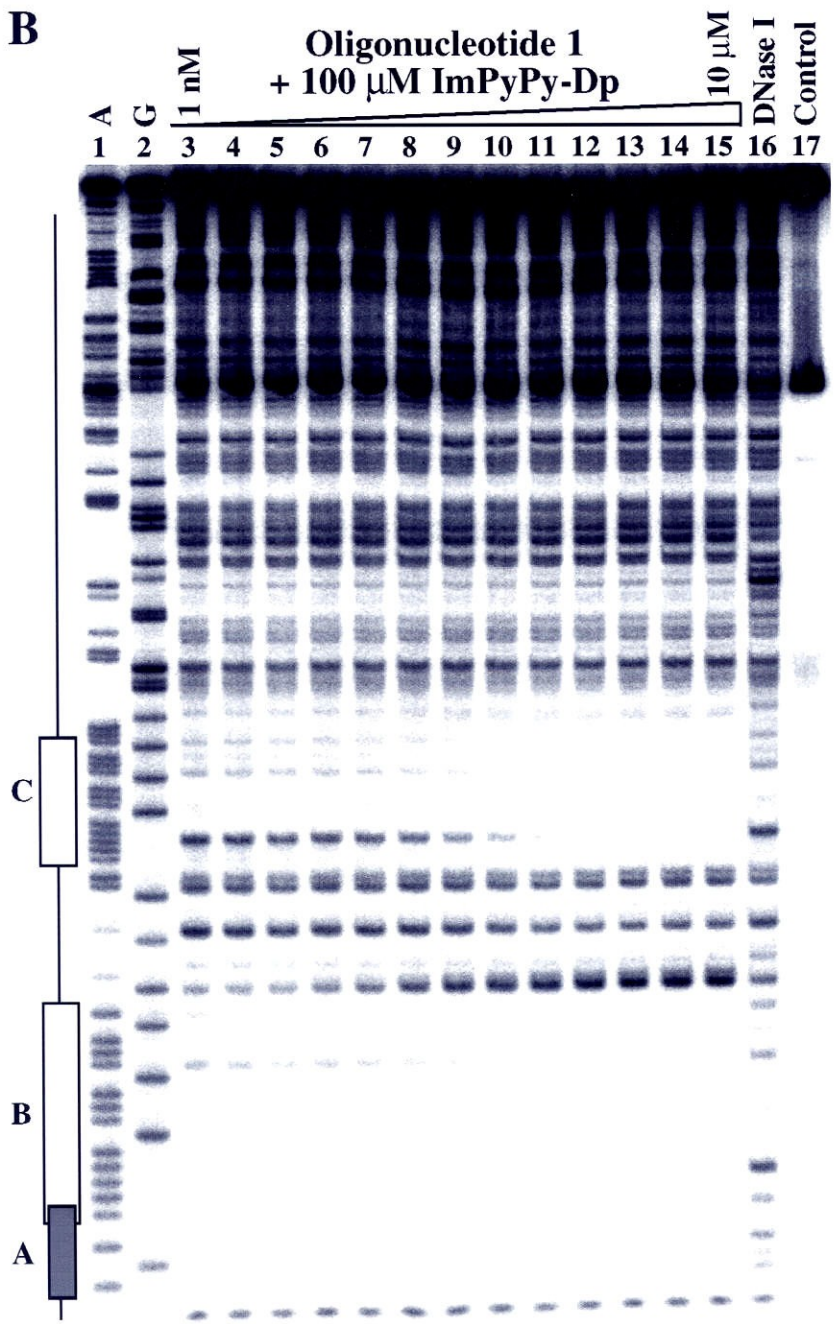

5'
3'
Distal
Triplex Site
Proximal
Triplex Site
Distamycin A
Site
3'5'
3'
F
E
D
5'3'

Figure 4. Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate products of DNase I digestion in quantitative footprint titration experiments with oligonucleotide **1** in the absence (**A**) and presence of 100 μM ImPyPy-Dp (**B**) on the 5'-^{32}P-labeled 255 base pair *Eco*RI/*Pvu*II restriction fragment from plasmid pMEP-2-ImN (25 mM Tris-acetate, pH 7.0, 10 mM NaCl, 250 μM spermine, 100 μM calf thymus DNA in 40 μL total volume). Lane 1, A sequencing reaction; lane 2, G sequencing reaction; lanes 3-14, DNase I digestion products produced in the presence of increasing concentrations of oligonucleotide **1** alone (**A**) and in the presence of 100 μM ImPyPy-Dp (**B**): lane 3, 1 nM; lane 4, 2 nM; lane 5, 5 nM; lane 6, 10 nM; lane 7, 20 nM; lane 8, 50 nM; lane 9, 100 nM; lane 10, 200 nM; lane 11, 500 nM; lane 12, 1 μM; lane 13, 2 μM; lane 14, 5 μM; lane 15, 10 μM,lane 16, DNase I standard, lane 17, intact DNA. Location of oligonuceotide **1** binding sites **B** and **C** are indicated by a nonshaded rectangle, while polyamide ImPyPy-Dp binding site **A** is indicated by a shaded rectangle. Note that sites **A** and **B** are protected from DNase I cleavage in (**B**) indicating simultaneous binding.


A
A
G
1 nM
Oligonucleotide 1
10 μM
DNase I
Control
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
C
B
A


B
A
G
1 nM
Oligonucleotide 1
+ 100 μM ImPyPy-Dp
10 μM
DNase I
Control
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
C
B
A

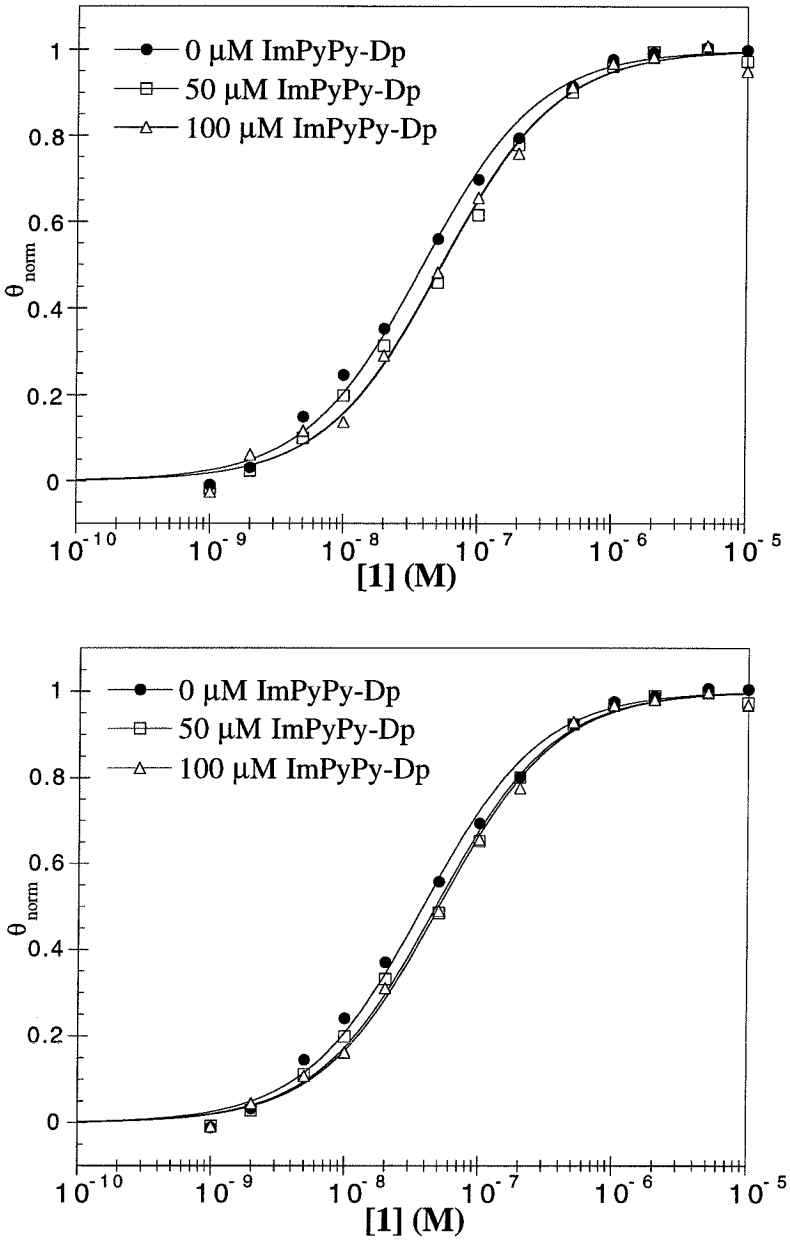


Figure 5. Data for the quantitative DNase I footprint titration experiments for oligonucleotide **1** in complex with the designated proximal **B** and distal **C** triplex sites (top and bottom, respectively). The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography and processed as described in the experimental section. The data points for oligonucleotide **1** in the presence of no ImPyPy-Dp, 50 μM ImPyPy-Dp and 100 μM ImPyPy-Dp are indicated by filled circles (●), open squares (□), and open triangles (Δ), respectively. The solid curves are the best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm using eq. 2.

Table I. Apparent First-Order Association Constants for Oligonucleotide **1** in the Absence and Presence of ImPyPy-Dp (M^{-1})[a, b]

ImPyPy-Dp (μM)	Proximal Site **B**	Distal Site **C**
0	2.7×10^7 (0.5)	2.7×10^7 (0.3)
50	1.9×10^7 (0.2)	2.1×10^7 (0.1)
100	1.9×10^7 (0.2)	2.0×10^7 (0.2)

[a]Values reported are the mean values of 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 25 mM Tris acetate, 10 mM NaCl, 250 μM spermine and 100 μM bp calf thymus DNA.

Table II. Apparent First-Order Association Constants for Oligonucleotide **2** in the Absence and Presence of distamycin A (M^{-1})[a, b]

Distamycin (μM)	Proximal Site **E**	Distal Site **F**
0	4.8×10^6 (0.8)	1.1×10^7 (1.0)
1	3.6×10^6 (1.2)	6.7×10^6 (1.6)
10	7.4×10^6 (1.0)	1.1×10^7 (1.9)

[a]Values reported are the mean values of 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 25 mM Tris acetate, 10 mM NaCl, 250 μM spermine and 100 μM bp calf thymus DNA.

Simultaneous binding of ImPyPy-Dp or distamycin A with a triple helix is also consistent with current literature studies demonstrating the compatibility of the minor groove binders, distamycin A and netropsin, with triple helix formation.[10]

ImPyPy-Dp differs from the natural products, distamycin and netropsin, in that it binds in the minor groove as a dimer, even at low concentrations.[5] In addition, the minor groove width for 1:1 distamycin or netropsin binding to A,T rich sites is approximately 3.4 Å[11] compared with 6.8 Å for models of the 2:1 complexes.[5,6] Since the oligonucleotide **1** site **B** overlaps the ImPyPy-Dp binding site **A**, positive or negative cooperativity between **1** and ImPyPy-Dp might be expected. Based on the first order equilibrium association constants for oligonucleotide **1** in the absence and presence of ImPyPy-Dp, no cooperativity of binding is observed (Table 1). In a similar experiment, we also found the binding of distamycin A to be compatible with triple helix formation, but not cooperative (Table 2). The fact that distamycin A and ImPyPy-Dp give similar results is surprising since the minor groove widths for these two ligand binding sites are expected to be different. This suggests that the triple helix can accommodate changes in the width of the minor groove of overlapping duplex DNA.

Early fiber diffraction studies on triple helical poly(dT)•poly(dA)•poly(dT) complexes indicated that all three strands in this complex are in the A DNA conformation and therefore have C3'-endo sugar puckers instead of C2'-exo puckers seen in B DNA.[14] However, NMR data indicate that not all of the sugars in the three strands are in the C3'-endo configuration.[8,15] In addition, recent fiber type X-ray diffraction patterns of triple helical DNA[16] as well as infrared spectroscopy studies[10c] indicate that triple helical DNA has B type helices rather than A type.

Our data supports a B-type conformation because the minor groove binding ligands ImPyPy-Dp and distamycin bind to the duplex DNA target simultaneously with a third

strand oligonucleotide binding. It is known that netropsin and distamycin do not bind appreciably to A-form DNA.[17] Furthermore, ImPyPy-Dp binds only to B-DNA as shown by NMR studies.[5] In our systems both ImPyPy-Dp and distamycin bind simultaneously with the triple helix strongly suggesting that at least the duplex, if not all three strands, exist in the B-form conformation. Our data also suggest that the triple helix is compatible with slightly different minor groove widths since the ImPyPy-Dp 2:1 binding site is wider than the distamycin 1:1 binding site.[5,6,11]

Conclusions

In summary, binding of an oligonucleotide to double helical DNA by triple helix formation in the major groove is compatible with either the polyamide ImPyPy-Dp binding as an antiparallel dimer or the natural product distamycin A binding as a monomer in the minor groove of *an overlapping sequence*. Remarkably, no binding cooperativity between the two partners is observed under the experimental conditions described, implying that triple helix formation is unaffected by different minor groove widths. This data suggests that a new class of hybrid molecules, polyamide-oligonucleotide conjugates, which interact with both the minor and major grooves of DNA, could be designed for sequence-specific recognition of DNA. This should further expand the sequence repertoire available for targeting DNA by artificial methods.

Experimental Section

Materials. Doubly distilled water was further purified using the Milli Q filtration system from Millipore. Spermine-tetrachloride was purchased from Sigma and dissolved in water to a concentration of 2.5 mM. Sonicated, deproteinized calf thymus DNA was

obtained from Pharmacia and dissolved in water to a concentration of 2 mM in base pairs. Enzymes were purchased from Boehringer-Mannheim or New England Biolabs and used with the buffers supplied. DNase I was obtained from Pharmacia. Deoxyadenosine-5'-[α-^{32}P]-triphosphate was obtained from Amersham. Phosphoramidites were supplied by Cruachem. pUC 19 plasmid was purchased from Boehringer -Mannheim Biochemical. Epicurian XL-1 Blue Supercompetent cells were obtained from Stratagene. UV-vis spectra were recorded on a Hewlett-Packard Diode Array spectrophotometer. Cerenkov radioactivity was measured with a Beckman LS 3801 Scintillation Counter. Gels were dried using a Bio-Rad Model 483 Slab Dryer. Storage phosphor autoradiography was performed using a Molecular Dynamics 400S Phosphorimager and ImageQuant software. Kodak S0230 storage phosphor screens were purchased from Molecular Dynamics.

Methods

Preparation of Plasmid DNA. The plasmid pMEP-2-ImN was prepared by ligation of the duplex formed between 5'-GATCCAGACAAAAAGAAAGAAAGA-CGATCGATCGAAGCAAAAAGAAAGAAAGAA-3' and 5'-AGCTTTCTTTCTTTC-TTTTTGCTTCGATCGATCGTCTTTCTTTCTTTTTGTCTG-3' with T4 DNA ligase into pUC 19, previously digested with *Bam*HI and *Hind*III. The plasmid pMEP-dist was similarly prepared by ligation of the duplex formed between 5'-AGCTTGAAAAAGA-GAAGAGAGAAACTGATACATCGAAGAGAAGAGAGAAAG-3' and 5'-GATCC-TTTCTCTCTTCTCTTCGATGTATCAGTTTCTCTCTTCTCTTTTTCA-3' with T4 DNA ligase into pUC 18, which was previously digested with *Bam*HI and *Hind*III. Resulting ligation products were used to transform competent cells. Colonies were selected for α-complementation on 25 mL Luria-Bertani agar plates containing 50 μg/mL ampicillin and treated with IPTG and XGAL solutions. Large scale plasmid purifications were carried out using Qiagen purification kits according to the provided protocol. The

resulting DNA was sequenced using a Sequenase Version 2.0 sequencing kit (United States Biolabs) according to the manufacturer's protocol. Concentration of plasmid DNA was determined at an absorbance of 260 nm using the relation 1 OD unit = 50 μg/mL duplex DNA.

Preparation of Labeled Restriction Fragment. Labeled restriction fragment from pMEP-2-ImN or pMEP-dist was generated as follows. Plasmid DNA was linearized using *Eco*RI, followed by treatment with either calf alkaline phosphatase and subsequent 5' end labeling with T4 polynucleotide kinase and γ-^{32}P-dATP for pMEP-2-ImN or 3' end labeling for pMEP-dist with Sequenase and α-^{32}P-dATP and α-^{32}P-TTP. The linearized pMEP-2-ImN plasmid DNA was digested with *Pvu*II and the 255 base pair *Eco*RI/*Pvu*II restriction fragment was isolated by nondenaturing 5% polyacrylamide gel electrophoresis (PAGE). Linearized pMEP-dist was digested with *Xmn*I and subsequently treated identically to pMEP-2-ImN. The gel bands were visualized by autoradiography, isolated, and eluted with 10 mM Tris•HCl and 1 mM EDTA (TE). Polyacrylamide was removed from the resulting mixture by filtration through a 0.45 micron Centrex filter. The DNA was then isopropanol precipitated, resuspended in TE and phenol extracted several times. The DNA was further purified by ethanol precipitation. Chemical sequencing adenine- and guanine-specific reactions were carried out as previously described.[18] All DNA manipulations were performed according to standard protocols.[19]

Quantitative DNase I Footprint Titrations. Footprint titration reaction conditions were 25 mM Tris-acetate, pH 7.0, 10 mM NaCl, 250 μM spermine, 100 μM calf thymus DNA, ~10,000 cpm of 5' labeled DNA in 40 μL total volume. Oligonucleotide **1** or **2** was allowed to equilibrate with the labeled restriction fragment in the above buffer conditions for 24 h at 24°C. For titrations in the presence of ImPyPy-Dp or distamycin, the polyamide was added after the 24 hour equilibration and allowed to

equilibrate for an additional hour. (Order of addition of oligonucleotide **1** and ImPyPy-Dp was reversed as a control. Results were identical for both orders of addition). After equilibration, DNase I was added to give a final concentration of 38 μunits/μL in 10 mM $MgCl_2$, 10mM $CaCl_2$, and 1 μM nonspecific oligonucleotide to maintain uniform DNase I reactivity. DNase I digestion was quenched after 8 minutes by the addition of EDTA, glycogen and NaOAc (pH 5.2) to give final concentrations of 5 mM, 60 μg/mL, and 300 mM, respectively. The reactions were then precipitated with 2.5 volumes of ethanol. The resulting pellets were washed with 70% ethanol, redissolved in 35 μL of water and lyophilized to dryness. The residue was then resuspended in 5 μL of denaturing 80% formamide loading buffer. The DNA was denatured at 90°C for 10 min, then placed on ice until loading. Reaction products were analyzed using 8% denaturing polyacrylamide gel (5% crosslink, 7M urea) electrophoresis. Gels were then dried and autoradiographed and/or exposed to phosphor screens for quantitative analysis.

Footprint Titration Fitting Procedure. The footprint titration gels were quantitated using storage phosphor technology (Molecular Dynamics), and data were analyzed by performing volume integrations of the target and reference sites using the ImageQuant v 3.1 software running on an AST Premium 386/33 computer. DNase I footprint titration experiments were analyzed according to the previously described protocol.[12,20] Briefly, θ_{app} was determined for each site 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} correspond to the integrated volumes of the target and reference sites, respectively. I_{tot}° and I_{ref}° represent the amount of DNase I digestion at the target site and the reference site, respectively, in a DNase I control lane to which no oligonucleotide or ImPyPy-Dp has been added. The ($[O]_{tot}$, θ_{app}) data points were fit by minimizing the

difference between θ_{app} and θ_{fit}, using the following equation, where θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively:

$$\theta_{app} = \theta_{min} + (\theta_{max} - \theta_{min}) \bullet \frac{K_T[O]_{tot}}{1 + K_T[O]_{tot}} \qquad (2)$$

Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software (version 3.0.1, Abelbeck software) running on a Macintosh IIfx computer with K_T, θ_{max}, and θ_{min} as the adjustable parameters and without weighting the individual data points. All lanes from each gel were used unless visual inspection of the computer image from a storage phosphor screen revealed a flaw at the target or reference site or unless the θ_{app} value was greater than two standard errors away from the initial θ_{fit}. Data for experiments in which fewer than 80% of the lanes were usable were discarded. Data were normalized using the following equation:

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

At least three sets of acceptable data were used in determining each association constant.

References

1. (a) Keller, W.; Konig, P.; Richmond, T.J. *J. Mol. Biol.* **1995**, *254*, 657. (b) Glover, J.N.M., Harrison, S.C. *Nature* **1995**, *373*, 257. (c) Konig, P.; Richmond, T.J. *J. Mol. Biol.* **1993**, *233*, 139. (d) Ellenberger, T.E.; Brandl, C.J.; Struhl, K.; Harrison, S.C. *Cell* **1992**, *71*, 1223.
2. (a) Kim, J.L.; Burley, S.K. *Nature Struct. Biol.* **1994**, *1*, 638. (b) Kim, J.L.; Nikolov, D.B.; Burley, S.K. *Nature* **1993**, *365*, 520. (c) Nikolov, D.B.; Hu, S.H.; Lin, J.; Gasch, A.; Hoffman, A.; Horikoshi, M.; Chua, N.H.; Roeder, R.G.; Burley, S.K. *Nature* **1992**, *360*, 40.
3. (a) Xu, W.G.; Rould, M.A.; Jun, S.; Desplau, C.; Pabo, C.O. *Cell* **1995**, *80*, 639. (b) Feng, J.A.; Johnson, R.C.; Dickerson, R.E. *Science* **1994**, *263*, 348.
4. (a) Moser, H.E.; Dervan, P.B. *Science* **1987**, *238*, 645. (b) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.; Thuong, N.T.; Lhomme, J.; Helene, C. *Nucleic Acids Res.* **1987**, *15*, 7749. (c) Cooney, M.; Czernuszewicz, G.; Postel, E.H.; Flint, S.J.; Hogan, M. E. *Science* **1988**, *241*, 456.
5. (a) Wade, W.S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574-1575. (b) Wade, W.S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (c) Mrksich, M.; Wade, W.S.; Dwyer, T.J.; Geierstanger, B.H.; Wemmer, D.E.; Dervan, P.B. *Proc. Natl. Acad. Sci., USA* **1992**, *89*, 7586. (d) Wade, W.S.; Mrksich, M.; Dervan, P.B. *Biochemistry* **1993**, *32*, 11385.
6. (a) Pelton, J.G.; Wemmer, D.E. *Proc. Natl. Acad. Sci. USA* **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. *Struct. Biol. Nature* **1994**, *1*, 169.

7. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (b) Geierstanger, B.H.; Jacobsen, J-P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Biochemistry*, **1994**, *33*, 3055.

8. (a) Praseuth, D.; Perrouault, L.; Le Doan, T.; Chassignol, M.; Thuong, N.; Helene, C. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1349. (b) de los Santos, C.; Rosen, M.; Patel, D. *Biochemistry* **1989**, *28*, 7282. (c) Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859. (d) Rajagopal, P., & Feigon, J. *Nature* **1989**, *339*, 637.

9. (a) Felsenfeld, G.; Davies, D.R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023. (b) Howard, F.B.; Frazier, J.; Lipsett, M.N.; Miles, H.T. *Biochem. Biophys. Res. Commun.* **1964**, *23*, 861. (c) Live, D.H.; Radhakrishnan, I.; Misra, V.; Patel, D.J. *J. Am. Chem. Soc.* **1991**, *113*, 4687.

10. (a) Umemoto, K.; Sarma, M.H.; Gupta, G.; Luo, J.; Sarma, R.H. *J. Am. Chem. Soc.* **1990**, *112*, 4539. (b) Park, Y.; Breslauer, K.J. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6653. (c) Howard, F.B.; Miles, H.T.; Liu, K.; Frazier, J.; Raghunathan, G.; Sasisekharan, V. *Biochemistry* **1992**, *31*, 10671. (d) Durand, M.; Thuong, N.T.; Maurizot, J. C. *J. Biol. Chem.* **1992**, *267*, 24394.

11. (a) Kopka, M.L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R.E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376. (b) Kopka, M.L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R.E. *J. Mol. Biol.* **1985**, *183*, 553.

12. (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. USA* **1986**, *83*, 8462. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.

13. Oakley, M.G.; Mrksich, M.; Dervan, P.B. *Biochemistry* **1992**, *31*, 10969.

14. Arnott, S.; Selsing, E. *J. Mol. Biol.* **1974**, *88*, 509.

15. (a) Macaya, R.F.; Schultze, P.; Feigon, J. *J. Am. Chem. Soc.* **1992**, *114*, 781. (b) Macaya, R.; Wang, E.; Schultze, P.; Sklenar, V.; Feigon, J. *J. Mol. Biol.* **1992**, *225*, 755.

16. Liu, K.; Miles, H.T.; Parris, K.D.; Sasisekharan, V. *Struct. Biol.* **1994**, *1*, 11.

17. (a) Zimmer, C. *Progr. Nucleic Acid Res. Mol. Biol.* **1975** *15*, 285. (b) Wartrell, R.M.; Larson, J.E.; Wells, R. D. *J. Biol. Chem.* **1974**, *249*, 6719. (c) Zimmer, C.; Reinert, K., E.; Luck, G., Wahnert, U.; Lober, G.; Thrum, H. *J. Mol. Biol.* **1971**, *72*, 329.

18. (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.

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

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

CHAPTER EIGHT

Affinity Cleaving and Footprinting Studies on the High Mobility Group I Binding Domain Peptide

Introduction

Protein-DNA Interactions. Proteins bind to DNA through a number of well-characterized structural motifs.[1] One important motif, the helix-turn-helix (HTH) is present in many prokaryotic transcriptional regulatory proteins and eukaryotic homeodomains.[2] The prokaryotic proteins usually bind as dimers, whereas the eukaryotic homeodomain proteins bind as monomers. Sequence-specific DNA recognition is mediated by an α-helix which binds in the major groove. Some of the HTH proteins also contact the minor groove of DNA with N-terminal extensions. Another common motif is the leucine zipper-basic region (bZIP) motif, in which proteins bind as dimers with the basic regions forming two alpha helices which traverse the major groove.[3] These helices continue in a straight line until their carboxy termini meet, forming a coiled-coil dimerization domain. Other structural motifs for protein-DNA recognition are discussed in the reviews of Pabo and Harrison.[1]

Protein recognition in the minor groove of DNA also occurs, but is less extensively studied. An example of minor groove binding proteins is a subsection of the high mobility group (HMG) proteins, which are a large class of mammalian non-histone proteins thought to be important structural components in the conformation and function of chromatin.[4] This subgroup of proteins, collectively called the HMG-I family, binds preferentially to A•T rich DNA *in vitro*.[5]

Previous Work in the Dervan Group on Small Peptides. Early attempts to recognize DNA with short peptides containing naturally occurring amino acids met with little success. Initial work by J. Sluka involved the ten amino-terminal residues of Hin recombinase.[6] The ten residue peptide derivatized with EDTA was prepared, but produced no sequence-specific cleavage. In addition, no footprints were observed with these peptides. In addition, G. Best prepared EDTA•Fe(II)-containing peptides having repeating SPKK units based on the report of Suzuki and coworkers, which proposed that such

repeats could sequence-specifically recognize DNA by binding in the minor groove.[7,8] Again, no sequence-specific cleavage or footprints were observed. Consistent with this result, recent NMR studies by Geierstanger et al. found that peptides containing repeating SPRK units did not bind specifically to a 5'-CGCAAATTTGCG-3' DNA sequence.[9]

High Mobility Group Proteins. HMG-I proteins are of considerable biological interest because they are expressed at elevated levels in actively proliferating cells.[10] High HMG-I levels have also been found consistently in rat and mouse malignant tumors[11] and are suggested to be a protein marker for neoplastic transformation and metastatic potential.[12] The specificity of HMG-I for A•T base pairs has led to other postulated functions, such as nucleosome phasing[5d] and the 3' end processing of mRNA transcripts.[13] In addition, HMG-I has been demonstrated to alter the conformation and stability of A,T-rich regions of DNA upon binding *in vitro*,[14] a property often associated with DNA-binding gene regulatory proteins.

Based on primary sequence information, the HMG-I proteins possess both a DNA binding domain and a highly acidic carboxy terminus,15 the typical structure of Ptashne-type transcriptional activator proteins.[5b,15,16] Reeves and coworkers propose that an 11-amino acid DNA-binding domain also contains a new structural motif, the "A•T hook," presumably used by HMG-I proteins to specifically bind to the narrow minor groove of A,T-rich DNA.[15b] The hypothetical "A•T hook" consists of a sharp turn at the amino end of the conserved binding domain. Positively charged arginine and lysine side chains contribute to binding by electrostatic interactions. Further stabilization occurs through the predicted crescent shape, similar to netropsin and distamycin, which places the amide hydrogens of the polypeptide backbone in positions compatible with hydrogen bonding with the minor groove face of the base pairs (Figure 1).[17]

Experimental Design. Previous studies have shown that HMG-I proteins bind to A,T-rich sequences in the minor groove.[5a-c,15b] To investigate the binding orientation and groove location of the HMG-I protein, an 11-residue peptide **T-P-K-R-P-R-G-R-P-K-K**

(**1**), proposed by Reeves and coworkers to be the HMG binding domain (HMG-BD),[15b] was synthesized (Figure 2). In addition, peptides of identical sequence containing an

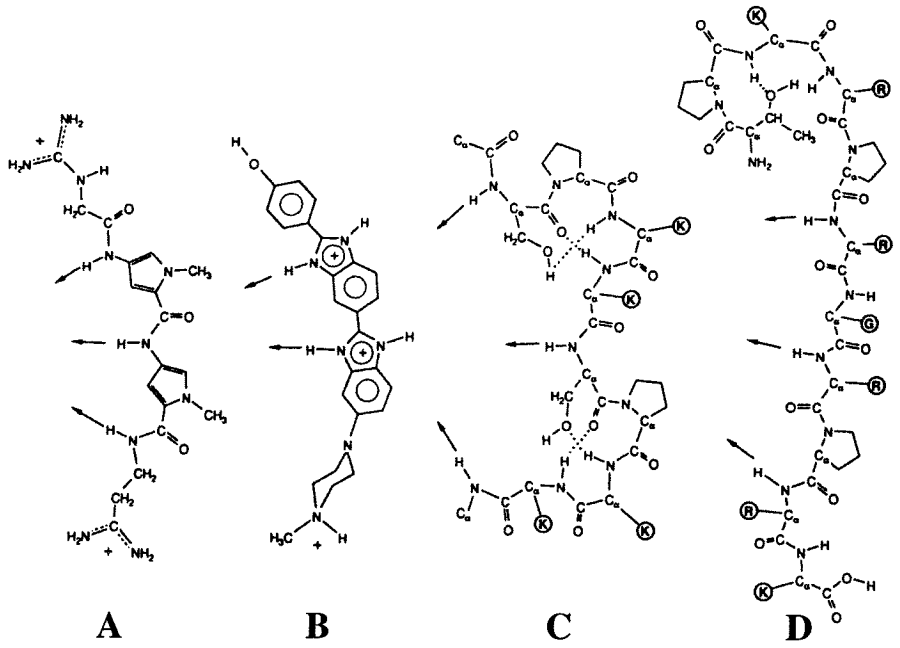


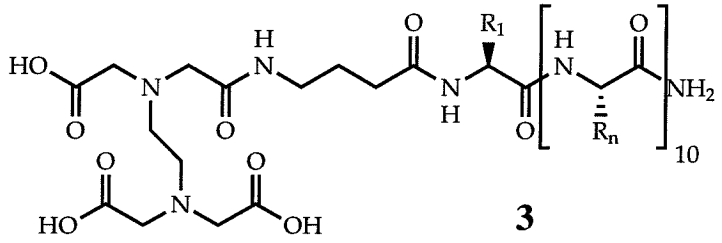
Figure 1. Comparison of the structures of the minor groove binding ligands (A) netropsin and (B) Hoechst 33258 with the proposed structures of the (C) SPKKSPKK and (D) HMG-BD peptides. Arrows indicate hydrogen bond donors. Lysine, arginine, and glycine side chains are represented by circles containing K, R, or G, respectively. Reproduced from reference 17.

EDTA•Fe(II) moiety at either the N-terminus (**3**) or the C-terminus (**4**) were synthesized (Figure 2). The C-termini of **1**-**3** are primary amides; the N-termini of **1** and **4** are acetylated. Peptide **5** is a second generation attempt to achieve sequence-specific cleavage using the more specific cleaving moiety Ni(II)•Gly-Gly-His.[18] Affinity cleavage experiments with **3** and **4** were performed to determine the orientation and binding sites of the peptide on double-stranded DNA.[19] MPE•Fe(II) footprinting was used to confirm that peptide **1** binds to the DNA target sequences chosen.[20]

A. TPKRPRGRPKK

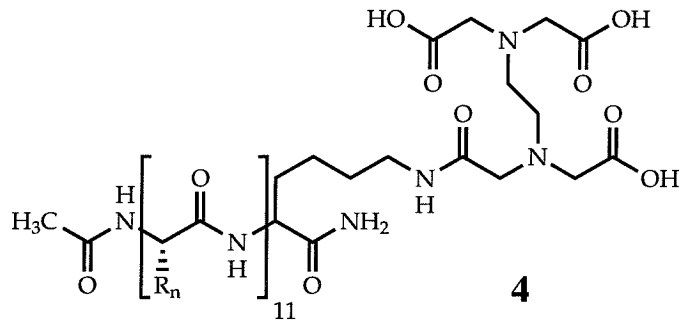
1: acetylated N-terminus
2: free N-terminus

B.



3

C.



4

D. GGH-TPKRPRGRPKK **5**

Figure 2. (A) An eleven residue HMG BD peptide with either acetylated (**1**) or free (**2**) N-terminus and C-terminal primary amide. (B) HMG BD peptide **3** with EDTA linked to the N-terminus through γ-aminobutyric acid. (C) HMG BD peptide **4** with EDTA linked to the C-terminus through the side chain of an added lysine residue. (D) Gly-Gly-His-HMG BD peptide **5** with free N-terminus and C-terminal amide.

Results

MPE•Fe(II) Footprinting of the HMG Binding Domain Peptides and GGH(HMG) Derivatives 1, 2 and 5.[20] Footprinting assays were performed on the 517 base pair *Eco*RI/*Rsa*I restriction fragment generated from pBR322, which contains several A,T-rich DNA regions. MPE•Fe(II) footprinting of peptides **1**, **2**, and **5** reveals that all three peptides bind the A,T-rich sites, 5'-TTAAT-3' and 5'-ATAATAAT-3' as expected (Figure 3). GGH peptide **5** generated similar footprints in the absence or presence of Ni(II).

Quantitative DNase I Footprinting. DNase I footprinting was performed to determine equilkibrium association constants for HMG BD peptide **1** (Table I).[21] Peptide **1** binds sites 5'-TTAAT-3' and 5'-ATAATAAT-3' with apparent equilibrium association constants of ~8 x 10^4 M^{-1}. Affinity for a 5'-TTTTT-3' site was too weak for quantitation (< 10^4 M^{-1}).

Affinity Cleaving Experiments.[19] Affinity cleaving experiments with Fe(II)•EDTA-HMG BD **3** and HMG BD-Fe(II)•EDTA **4** showed no sequence-specific cleavage at any concentration. Cleavage bands occurred at every base pair. No specific protection from this cleavage was observed at the target sites.

Ni(II)•GGH(HMG) Affinity Cleavage.[18] Affinity cleaving experiments with Ni(II)•GGH(HMG) **5** showed cleavage at all G residues, probably due to the fact that G is the most easily oxidized base. No sequence-specific cleavage around HMG binding sites was observed.

Discussion

Binding Affinity of HMG Peptides. The HMG binding domain peptides **1** and **2** bind as expected to A,T-rich DNA with equilibrium association constants of ~8x10^4 M^{-1}. The two strongest sites for peptides **1** and **2** are 5'-TTAAT-3' and 5'-ATAATAAT-3'.

HMG peptide **1** (acetylated N-terminus) (10 μM):

HMG peptide **2** (free N-terminus) (10 μM):

GGH-HMG peptide **5** (free N-terminus) (10 μM):

GGH-HMG peptide **5** (with Ni(II)) (10 μM):

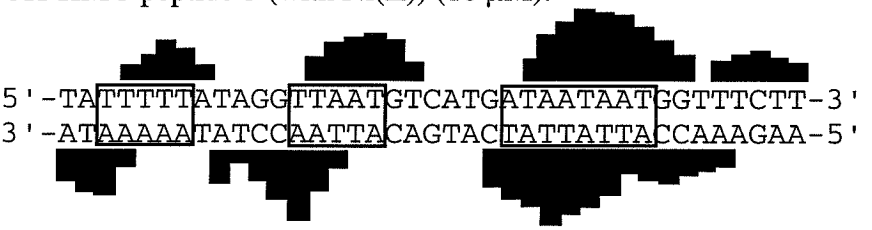


Figure 3. Histograms of cleavage protection (footprinting) data. MPE•Fe(II) protection patterns for HMG peptides. Bar heights are proportional to the relative protection from cleavage at each band. Boxes represent equilibrium binding sites determined by the published model.

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

Binding Site	K_a (M^{-1})
5'-TTTTT-3'	$< 10^4$
5'-TTAAT-3'	8.2×10^4 (1.5)
5'-ATAATAAT-3'	8.9×10^4 (3.5)

[a]Reaction conditions were 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, 1 mM $CaCl_2$, 50 mM NaCl, 100 μM bp calf thymus DNA, 4°C, and 20,000 cpm radiolabeled restriction fragment in a volume of 40 μL. [b]Three sets of acceptable data were used in determining each association constant.

The second site may be two overlapping five base pair sites. Unfortunately, EDTA•Fe(II) derivatives **3** and **4** did not produce sequence-specific cleavage. This may be due to either low affinity or lack of specificity for the target sequences. Nonspecific interactions of the six positively charged lysine and arginine residues with the DNA likely decrease the sequence specificity. In spite of the discouraging affinity cleaving results, we were able to detect and characterize binding sites by footprinting and confirm the preference for A,T-rich DNA.

Ni(II)•GGH(HMG) Peptide. GGH has previously been shown to produce a very specific cleavage pattern of one or two bands.[18] By using GGH as the cleaving moiety, we hoped to see less diffuse cleavage patterns, allowing the determination of peptide binding orientation. Unfortunately, no specific cleavage patterns around the known binding sites of HMG were observed. Instead, cleavage at all guanine bases was observed. This was not surprising as Burrows and coworkers have observed similar behavior with square planar Ni complexes.[22]

Conclusions. Although the eleven residue peptide proposed by Reeves and coworkers[15b] binds to A,T-rich DNA, the affinity and specificity are too low to allow the binding orientation to be determined. However, recent NMR studies by Geierstanger et al.

demonstrate specific complex formation between the eleven residue peptide and the central region of a 5'-CGCAAATTTGCG-3' sequence.[9] No evidence of the A,T-hook motif is seen. Instead, the PRGRP core anchors the peptide specifically into the minor groove by contacts from the RGR segment as predicted by Lund et al.[10a] and shown by 2D NOE experiments.

Experimental

Manual peptide synthesis was carried out in 20 mL reaction vessels fitted with coarse glass frits as described by Kent.[23] The synthetic protocols used were developed at the California Institute of Technology.[24] Protected amino acid derivatives were purchased from Peninsula Laboratories. *t*Boc-L-His (DNP protected) was purchased from Fluka. N^{α}-*t*-butoxycarbonyl(Boc)-N^{ε}-Fmoc-L-Lys was purchased from Chemical Dynamics Corp. Methyl-benzhydrylamine resin (MBHA) was purchased from U.S. Biochemical Corp. Dimethylformamide (DMF), diisopropylethylamine (DIEA), and trifluoroacetic acid (TFA) were obtained from Applied Biosystems. Dicyclohexylcarbodiimide (DCC) was purchased from Peptides International. Hydroxybenzotriazole (HOBt) was obtained from Aldrich. Dichloromethane (DCM) (HPLC grade) was purchased from Mallinkrodt.

N^{α}-*t*Boc-L-amino acids were used with the following side chain protecting groups: Arg(Tos), His(DNP), Lys(Cl-Z) and Thr(Bzl). Manual assembly of the protected peptide on the solid support was carried out at 25°C via a three-step reaction cycle. First the Boc protecting group was removed from the α-amino group of the resin-bound amino acid by using TFA (65% TFA in DCM for 1 and 15 min). The deprotected peptide resin was then neutralized with 10% DIEA in DCM (2 X 1 min). All amino acids were coupled to the free α-amino group as the HOBt esters. Coupling yields were determined by quantitative ninhydrin monitoring with acceptable values being ≥99.5%.[25] Average coupling yields were 99.8%. If a second coupling was necessary, the resin was neutralized with 10%

DIEA in DMF. The HOBt ester was formed outside the reaction vessel in an absolute minimum of DCM/DMF. The solution was filtered into the vessel to remove the dicyclohexylurea, topped off with DMF, and allowed to react for 1-2 hours. If the yield was not acceptable at this point, a third coupling was performed by the same protocol as the second coupling.

Tricyclohexyl-EDTA (TCE) from EDTA-Dianhydride. A modified synthesis of TCE was used as follows.[19b] Ethylenediaminetetraacetic dianhydride (10.0 g, 39.0 mmol), cyclohexanol (39.1 g, 390 mmol), and 4-dimethylaminopyridine (0.476 g, 3.9 mmol) were suspended in dimethylformamide (DMF) (100 mL) and stirred at 45° C for 11 h. At this time 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (7.85 g, 40.9 mmol) was dissolved/suspended in DMF (50 mL) and added to the stirring solution, then allowed to stir at room temperature for 6 h. The reaction mixture was then added to diethyl ether (200 mL) and washed with water (4x200 mL). The ether layer was then dried over magnesium sulfate and concentrated by rotary evaporation. Remaining DMF and cyclohexanol were removed by vacuum distillation. The remaining yellow oil was purified by flash column chromatography (3% methanol/dichloromethane-10% methanol/dichloromethane gradient) to yield 2.59 g (12%) as a thick gum. Characterization correlated with previously published data.

TCE Coupling to the NH_2-terminus of Resin Bound Peptides. A typical procedure for the coupling of TCE to the NH_2-terminus of resin bound peptide is as follows. A 300 mg sample of peptide-resin (0.292 mmol/g, total peptide 0.073 mmol) was placed in a 12 x 80 mm reactor, and the resin was swollen in DCM for 15 minutes. The resin was washed with DCM (5X), the terminal Boc protecting group was removed, and the resin neutralized using standard procedures. TCE (125 mg, 0.19 mmol) and HOBt (50 mg, 0.33 mmol) were dissolved in DMF (2 mL) and DCC (42 mg, 0.20 mmol) was added. The solution was stirred 30 min at room temperature and added to the peptide resin

along with sufficient DMF to fill the reactor two-thirds full. Ninhydrin analysis indicated ≥99.9% reaction yield after 2 h. The resin was washed with DMF (5X) and DCM (5X).

TCE Coupling to the NH_2 Side Chain of a Lysine Residue at the Carboxy Terminus of Resin Bound Peptides. N^{α}-tBoc-N^{ε}-Fmoc-L-lysine was coupled to the growing peptide as an HOBt ester using the standard procedure. The resin was washed with DCM (5X) followed by DMF (5X). The N^{ε}-Fmoc group was selectively removed using 20% piperidine in DMF (a 3 min reaction step followed by a 20 min step) and the resin was washed with DMF (10X). TCE (1.61 g, 3 mmol, 6 equiv. based on a 0.5 mmol synthesis) and HOBt (0.405 g, 3.0 mmol, 6.0 equiv.) were dissolved in DMF (2 mL) and DCC (0.619 g, 3.0 mmol, 6 equiv.) was added. The solution was stirred 30 min at room temperature and added to the peptide resin along with sufficient DMF to fill the reactor two-thirds full. Ninhydrin analysis indicated 99.9% reaction after 2 h. The peptide resin was washed with DMF (5X) and DCM (5X). The N^{α}-Boc was removed as usual with TFA and the synthesis continued with the standard protocols.

Deprotection and Purification. After removal of the N^{α}-*t*Boc group with TFA and drying of the resin, all other side-chain protecting groups as well as the peptide-resin bond were cleaved with anhydrous hydrofluoric acid (HF), in the presence of *p*-cresol and *p*-thiocresol as radical scavengers, for 60 min at 0°C. The HF was removed under vacuum. The crude protein was precipitated with diethyl ether, collected on a fritted funnel, dissolved in 5% acetic acid/water, and washed through, leaving the resin on the frit. A small sample was then removed, filtered, and subjected to analytical HPLC (Vydac 25 cm X 4.6 mm C_{18} column, 0-60% acetonitrile/0.1% TFA over 60 min). The remaining solution was frozen and lyophilized. When the histidine protecting group, dinitrophenyl (DNP), was present it was removed by treatment in 4 M guanidine hydrochloride, 50 mM tris, pH 8.5, and 20% 2-mercaptoethanol for 1 hour at 50 °C. This solution was injected directly onto a semipreparative C_{18} HPLC column 25 cm X 1 cm) and run in H_2O/0.1% TFA until the guanidine and 2-mercaptoethanol had eluted. A gradient of 0-30%

acetonitrile/0.1% TFA was run over 120 min, and fractions were collected. Fractions were checked by analytical HPLC, and those containing the desired peak were pooled and lyophilized. Protein concentration was determined using the bicinchoninic acid (BCA) assay.

Characterization of Peptides. HMG BD control peptides **1** and **2** were characterized by mass spectrometry performed by Jane Sanders at Caltech. HMG BD•EDTA **4** was characterized by electrospray mass spectrometry performed by New York University. EDTA•HMG BD **3** was not amenable to characterization by mass spectrometry. GGH(HMG) **5** was sequenced by the Caltech peptide facility.

Determination of Binding Sites by MPE•Fe(II) Footprinting[20] MPE•Fe(II) footprinting conditions were 10 mM Tris•HCl pH 7.5, 100 mM NaCl, 100 μM calf thymus DNA, labeled DNA (20,000 cpm), and 5 μM MPE•Fe(II) in a volume of 40 μL. The peptides were equilibrated with either the 3' or 5'-labeled 517 base pair *Eco*RI/*Rsa*I restriction fragment for 30 minutes, followed by addition of MPE•Fe(II) and further equilibration for 15 min. Cleavage was initiated with 4 mM DTT, allowed to proceed for 10 min, then terminated by ethanol precipitation. One microliter of glycogen was added to each tube to aid in precipitation and resuspension. The residue after precipitation was dried and resuspended in 80% formamide loading buffer. Reaction products were analyzed using 8% denaturing polyacrylamide gel (5% crosslink, 7M urea) electrophoresis. Gels were dried using a Bio-Rad Model 483 Slab Dryer followed by exposure to phosphor screens for quantitative analysis. Analysis was performed by drawing a rectangle around each resolvable band and integrating the volume. The bands were then compared to an MPE•Fe(II) control lane with no ligand to determine if footprints were present.

Determination of Energetics by Quantitative DNase I Footprint Titration. Apparent equilibrium association constants were determined by quantitative DNase I footprint titration experiments as previously described.[21] Reaction conditions were 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 1 mM $CaCl_2$, 50 mM NaCl, 100 μM bp calf

thymus DNA and 20,000 cpm radiolabeled restriction fragment in a volume of 40 μL. The peptides were added to the DNA and allowed to equilibrate for 30 min at 4°C. Reactions were initiated by the addition of 3 μL of a stock solution of DNase I (7 units/mL) containing 1 mM DTT and allowed to proceed for 10 minutes at 4°C. The reactions were stopped by addition of a 3 M ammonium acetate solution containing 250 mM EDTA and ethanol precipitated. The reactions were resuspended and analyzed as previously discussed. The data was analyzed by performing volume integrations of the target sites shown in Table 1 and a 5'-TAGGCG-3' reference site. The apparent DNA target site saturation, θ_{app}, was calculated for each concentration of peptide using the following equation:

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

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 which contains no peptide. At higher concentrations of peptide (>20 μM), the reference sites become partially protected, resulting in low θ_{app} values. For these data points, the reference value was calculated from the amount of radioactivity loaded in each lane, using the mean value for all data points from lane with < 20 μM peptide. The ($[L]_{tot}$, θ_{app}) data points were fit by minimizing the difference between θ_{app} and θ_{fit} using the modified Hill equation:

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

where $[L]_{tot}$ corresponds to the total peptide concentration, K_a corresponds to the apparent monomeric association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. A monomeric association was assumed (n =1).

Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software (version 2.1, Abelbeck software) running on a Macintosh IIfx computer with K_a, θ_{max}, and θ_{min} as the adjustable parameters. All lanes from each gel were used unless visual inspection revealed a flaw. The data were normalized using the following equation:

$$\theta_{norm} = \frac{\theta_{app} - \theta_{min}}{\theta_{max} - \theta_{min}}$$

The goodness of fit of the binding curve to the data points was 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.

Affinity Cleaving Reactions.[19] Reaction conditions were 10 mM Tris•HCl pH 7.5, 100 mM NaCl, 100 μM calf thymus DNA, labeled DNA (20,000 cpm), and various concentrations of peptide (1 pM-100 μM). The peptides were equilibrated with the DNA for 30 min at which time 4 mM DTT was added to initiate cleavage. Cleavage reactions were allowed to proceed for 30 min, then terminated by ethanol precipitation. The residue was resuspended and analyzed by gel electrophoresis as previously reported.

Ni(II)•GGH(HMG BD) Reactions.[18] Cleavage reactions were performed in a volume of 40 μL. Final concentrations were 20 mM phosphate, pH 7.5, 20 mM NaCl, 100 μM (in base pairs) calf thymus DNA, ~20,000 cpm ^{32}P-radiolabeled restriction fragment, 5 or 50 μM monoperoxypthalic acid, and various concentrations of Ni(II)•GGH-(HMG BD). The peptide was equilibrated with the DNA for 30 min prior to the addition of oxidant. Cleavage reactions were initiated by the addition of magnesium monoperoxypthalate and allowed to proceed for 15 min at 24°C. The reactions were terminated by ethanol precipitation, dried, resuspended in 50 μL of 0.1 N *n*-butylamine, heated to 90°C for 30 min, frozen, and lyophilized as described previously.[18] Reaction products were resuspended and analyzed by gel electrophoresis as previously discussed.

References

1. (a) Pabo, C. O.; Sauer, R. T. *Annu Rev. Biochem.* **1992**, *61*, 1053. (b) Harrison, S. C. *Nature* **1991**, *353*, 715.
2. Feng, J. A.; Johnson, R. C.; Dickerson, R. E. *Science* **1994**, *263*, 348.
3. Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223.
4. Goodwin, G.; Bustin, M. in *Architecture of Eukarytoic Genes*, Kahl, G. Ed.; VCH, Weinheim, F. R. G., 1988.
5. (a) Elton, T. S.; Nissen, M. S.; Reeves, R. *Biochem. Biophys. Res. Chem.* **1987**, *143*, 260. (b) Reeves, R.; Elton, T. S.; Nissen, M. S.; Lehn, D.; Johnson, K. R. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6531. (c) Solomon, M.; Strauss, F.; Varshavsky, A. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 1276. (d) Strauss, F.; Varshacsky, A. *Cell* **1984**, *37*, 889.
6. Sluka, J. P. Ph.D. Thesis, California Institute of Technology, 1988.
7. Best, G. C.; Dervan, P. B. Unpublished observations.
8. (a) Suzuki, M. *J. Mol. Biol.* **1989**, *207*, 61. (b) Churchill, M. E. A.; Suzuki, M. *EMBO J.* **1989**, *8*, 4189.
9. Geierstanger, B. H.; Volkman, B. F.; Kremer, W.; Wemmer, D. E. *Biochemistry* **1994**, *33*, 5347.
10. (a) Lund, T.; Holtlund, J.; Fredriksen, M.; Laland, S. G. *FEBS Lett.* **1983**, *152*, 163. (b) Elton, T. S.; Reeves, R. *Anal. Biochem.* **1986**, *157*, 53.
11. (a) Vartiainen, E.; Palvimo, J.; Mahonen, A,; Linnala-Kankkunen, A.; Maenpaa, P. H. *FEBS Lett.* **1988**, *228*, 45. (b) Giancotti, V.; Pani, B.; D'Andrea, P.; Berlingieri, M. T.; Di Fiore, P. et al. *EMBO J.* **1987**, *6*, 1981. (c) Giancotti, V.; Buratti, E.; Perissin, L.; Zorvet, S.; Balmain, A. et al. *Exp. Cell Res.* **1989**, *184*, 538.

12. Bussemakers, M. J. G.; van de Ven, W. J. M.; Debruyne, F. M. J.; Schalken, J. A. *Cancer Res.* **1991**, *51*, 606.

13. Russnak, R. H.; Candido, E. P. M.; Astell, C. R. *J. Biol. Chem.* **1988**, *263*, 6392.

14. Lehn, D. A.; Elton, T. S.; Johnson, K. R.; Reeves, R. *Biochem. Int.* **1988**, *16*, 963.

15. (a) Johnson, K. R.; Lehn, D. A.; Reeves, R. *Mol. Cell. Biol.* **1989**, *9*, 2114. (b) Reeves, R.; Nissen, M. S. *J. Biol. Chem.* **1990**, *265*, 8573.

16. Ptashne, M. *Nature* **1988**, *335*, 683.

17. Churchill, M. E. A.; Travers, A. A. *Trends. Biochem. Sci.* **1991**, *16*, 92.

18. (a) Mack, D. P.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**. *112*, 4604. (b) Mack, D. P.; Dervan, P. B. *Biochemistry* **1992**, *31*, 9399.

19. (a) Sluka, J. P.; Horvath, S. J.; Bruist, M. F.; Simon, M. I.; Dervan, P. B. *Science* **1987**, *238*, 1129. (b) Sluka, J. P.; Griffin, J. H.; Mack, D. P.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 6369. (c) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457. (d) Dervan, P. B. *Science* **1986**, *232*, 464.

20. (a) Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 5470. (b) Van Dyke, M. W.; Dervan, P. B. *Cold Spring Harbor Symposium on Quantitative Biology* **1982**, *47*, 347. (c) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373.

21. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132. (b) Brenowitz, M.; Senear, M. A.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8462. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.

22. (a) Chen, X.; Rokita, S. E.; Burrows, C. J. *J. Am. Chem. Soc.* **1991**, *113*, 5884. (b) Chen, X.; Burrows, C. J.; Rokita, S. E. *J. Am. Chem. Soc.* **1992**, *114*, 322.

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

24. (a) Kent, S. B. H.; Clark-Lewis, I. in *Synthetic Peptides in Biology and Medicine*, Alitalo, K.; Pertanen, P.; Vaheri, A. Eds.; Elsevier: Amsterdam, 1985. (b) Clark-

Lewis, I.; Aebersold, R.; Ziltener, H.; Schrader, J.; Hood, L. E., Kent, S. B. H. *Science* **1986**, *231*, 134.

25. Sarin, V. K.; Kent, S. B. H.; Tam, J. P.; Merrifield, R. B. *Anal. Biochem.* **1981**, *117*, 147.

APPENDIX ONE

Synthesis of ^{15}N-Labeled ImPyPy-Dp for NMR Studies

Introduction

Recent 2:1 polyamide-DNA complexes have created new models for the design of nonnatural ligands for sequence-specific recognition in the minor groove of DNA.[1-4] An imidazole-pyrrole-pyrrole tripolyamide, 1-methylimidazole-2-carboxamide-netropsin (ImPyPy-Dp), was shown to specifically bind the mixed sequence 5'-(A,T)G(A,T)C(A,T)-3' as a side-by-side antiparallel dimer.[2] In addition to the sequence-dependent minor groove width being a determinant of specificity, the 2:1 model allows specific contacts with *each strand* on the floor of the minor groove.[1-4] The side-by-side combination of one imidazole ring on one ligand and a pyrrolecarboxamide on the second ligan reacognizes G•C, while a pyrrolecarboxamide/imidazole pair targets a C•G base pair.[2-4] A pyrrolecarboxamide/pyrrolecarboxamide combination is partially degenerate and binds to either A•T or T•A base pairs.[1-4]

Experimental Design. In an effort to obtain direct physical evidence of the hydrogen bond formed between the imidazole N3 and a guanine exocyclic amine, ImPyPy-Dp containing ^{15}N at the 1 and 3 positions was synthesized. The goal was to study the polyamide:DNA complex by NMR in collaboration with the Wemmer group and try to observe NOEs between the nitrogen and the amino group. In addition, the chemical shift of the ^{15}N-labeled N3 position should change upon hydrogen bonding with the DNA.

Results

Synthesis. The ^{15}N-labeled compound **4** was synthesized as shown in Figure 1. [1,3-^{15}N] imidazole was methylated using trimethylsulfonium hydroxide,[5] then converted to the carboxylic acid **2** by deprotonation with *n*-butyl lithium followed by quenching with carbon dioxide. Carboxylic acid **2** was activated (DCC and HOBt) and coupled with the dipyrrole unit **3**, synthesized as previously reported,[2a] to afford [1,3-^{15}N]-ImPyPy-Dp **4**.

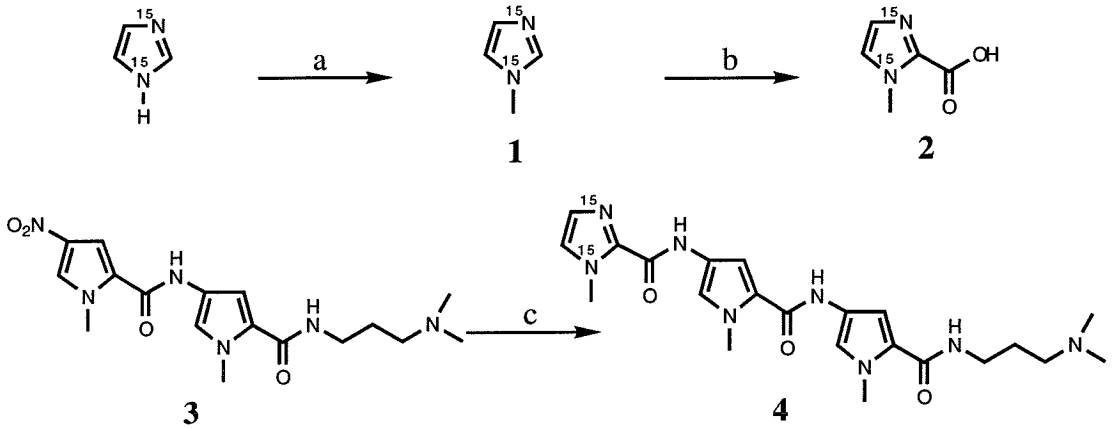


Figure 1. Synthetic scheme for [1,3-^{15}N]- ImPyPy-Dp. (a) $(CH_3)_3SOH$; (b) (i) *n*-butyllithium; (ii) CO_2; (iii) HCl; (c) (i) 50 psi H_2, 10% Pd/C; (ii) DCC, HOBt, **2**.

NMR Characterization. NMR characterization performed by the Wemmer group demonstrates a chemical shift change (Δ=87.5 ppm) of the polyamide N3 imidazole position upon addition of the DNA, consistent with formation of a hydrogen bond. However, they were unable to observe direct NOEs between the ^{15}N3 and the guanine exocyclic amine protons.

Experimental Section

^{1}H NMR spectra were recorded at 300 MHz on a General Electric-QE 300 NMR spectrometer in DMSO-d_6 or $CDCl_3$. High-resolution mass spectra (HRMS) were recorded using electron ionization (EI) or fast atom bombardment (FAB) techniques at the Mass Spectrometry Laboratory at the University of California, Riverside. Reagent grade chemicals were used unless otherwise noted. Tetrahydrofuran (THF) was distilled under nitrogen from sodium/benzophenone ketyl. Dimethylformamide (DMF) was purchased as an anhydrous solvent from Aldrich. [1, 3-^{15}N]-imidazole was purchased from Cambridge Isotopes and used as received. Trimethylsulfonium hydroxide was prepared as previously reported.[5] Flash chromatography was carried out using EM Science Kieselgel

60 (230-400) mesh.[6] All the pyrrole derivatives were visualized with short-wave ultraviolet light. *N*-Methylimidazole was visualized by I_2 staining. The nitropyrrole **3** was synthesized as previously reported.[1]

***N*-Methyl-[1,3-^{15}N]-imidazole 1.** A solution of trimethylsulfonium hydroxide (0.95N, 2.2 ml, 2.06 mmol) and [1,3-^{15}N]-imidazole (0.100 g, 1.43 mmol) was concentrated under reduced pressure, then heated to reflux (75°C) for 4h. The reaction mixture was then mixed with water (5 mL), saturated with sodium chloride, and extracted with dichloromethane (5x10 mL). The dichloromethane layer was dried over magnesium sulfate and concentrated under reduced pressure. The product was purified by flash column chromatography (10% methanol in dichloromethane) to give an oil (0.033 g, 30%): ^{1}H NMR ($CDCl_3$) δ 7.3 (s, 1H), 6.9 (s, 1H) 6.7 (s, 1H), 3.5 (s, 3H). Purified material cleanly co-spotted by TLC (1% methanol in chloroform) with authentic unlabeled material. All material was carried through to the labeled acid **2**.

***N*-Methyl-[1,3-^{15}N]-imidazole-2-carboxylic Acid 2.** *n*-BuLi (0.16 ml, 0.41 mmol in hexanes) was added to a cooled solution (-70°C) of *N*-Methyl-[1,3-^{15}N]-imidazole (0.033 g, 0.39 mmol) in anhydrous THF (15 mL) and allowed to stir for 30 min. Dry carbon dioxide was bubbled through the solution for 1 h at -70°C, then the solution was allowed to warm to room temperature. The reaction was quenched with HCl (1N, 1 mL), at which time a yellow oil separated. Lyophilization of this oil from water, followed by lyophilization from HCl (1N) afforded the [1, 3-^{15}N]-imidazole carboxylic acid (0.048 mg, 95%) as a white solid: m/z for $C_5H_7O_2{}^{15}N_2$, calcd 129.045, obsd 129.045.[7]

1-Methyl-[1,3-^{15}N]-imidazole-pyrrole-pyrrole-dimethylaminopropyl ([1,3-^{15}N]-ImPyPy-Dp) 4. To a solution of *N*-methyl-[1,3-^{15}N]-imidazole-2-carboxylic acid **2** (0.046 g, 0.36 mmol) and *N*-hydroxybenzotriazole (0.074 g, 0.55 mmol) in DMF (25 mL) was added a solution of dicyclohexylcarbodiimide (0.128 g, 0.62 mmol) in CH_2Cl_2 (2 mL). The mixture was allowed to stir 10 h. Separately, a solution of nitropyrrole **3** (0.203

g, 0.59 mmol) and palladium on activated carbon (10%, 0.075 g) was allowed to stir under a hydrogen atmosphere (50 psi) for 23 h to form the amine. The mixture was filtered through celite and added to the activated acid. The solution was allowed to stir for 2.5 h at which time methanol (3 mL) was added; the solvent was then removed under reduced pressure. The crude residue was dissolved in CH_2Cl_2 (35 mL) and washed once with 10% $NaHCO_3$. The organic layer was dried over $MgSO_4$ and concentrated under reduced pressure. The product was purified by flash column chromatography (gradient 0-1% ammonium hydroxide in methanol) to afford [1,3-^{15}N]-ImPyPy-Dp (0.053 g, 30%). The molecule was further purified using reverse phase FPLC (Pro/RPC 10/10 (C2-C8) column (Pharmacia) and a 0-40% acetonitrile gradient in 100 mM triethylammonium acetate, pH 7): ^{1}H NMR ($CDCl_3$) δ 9.15 (s, 1H), 7.95 (s, 1H) 7.60 (t, 1H), 7.20 (d, 1H), 7.15 (d, 1H), 7.00 (d, 1H), 6.95 (d, 1H), 6.75 (d, 1H), 6.50 (d, 1H), 4.10 (s, 3H), 3.92 (s, 3H), 3.90 (s, 3H), 3.45 (q, 2H), 2.50 (t, 2H), 2.30 (s, 6H), 1.75 (quint, 2H); FABMS m/z for $C_{22}H_{31}O_3N_6{}^{15}N_2$ calcd 457.2459 obsd 457.2456.

References

1. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723-5727. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393-1399. (c) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Struct. Biol. Nat.* **1994**, *1*, 169-175.
2. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci., USA* **1992**, *89*, 7586-7590. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385-11389.
3. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572-2576. (b) Geierstanger, B. H.; Jacobsen, J-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry*, **1994**, *33*, 3055.
4. Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474-4482.
5. Yamauchi, K.; Tanabe, T.; Kinoshita, M. *J. Org. Chem.* **1979**, *44*, 638-639.
6. Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, *43*, 2923-2925.
7. ^{1}H NMR on the *N*-methyl-[1,3-^{15}N]-imidazole-2-carboxylic acid **2** was not performed due to relatively rapid decarboxylation of the compound in solution.