DESIGN OF DNA-BINDING POLYAMIDES FOR REGULATION OF GENE EXPRESSION

Thesis by

John Wesley Trauger

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1999

(Submitted July 21, 1998)

To my Parents

Acknowledgements

I would like to thank my advisor, Professor Peter Dervan, for his support and guidance over the past five years. Peter's enthusiasm for science, his scientific vision and his committment to excellence have inspired me throughout my time at Caltech. I am particularly grateful to him for providing me with both the freedom to pursue my own ideas and for solid advice when I needed it, which allowed me to grow and accomplish more as a scientist than I could have imagined at the outset. I would also like to thank the members of my committee, Professors Doug Rees, Andy Myers, Bill Goddard and Jackie Barton for their valuable input at candidacy and in recent months.

To all of the members of the Dervan Group, past and present, go my thanks for providing a stimulating environment in which to work and learn. I would particularly like to thank Eldon Baird, with whom I enjoyed many stimulating scientific discussions, and who collaborated with me on much of the work presented in this thesis. It has been an honor to have worked with him. I would also like to thank my coworkers in 330 and 320 Church, Scott Priestley, Will Greenberg, Ramesh Baliga, David Liberles, David Herman, Adam Urbach, Paul Floreancig, Anna Mapp, Meridith Howard and Jason Belitsky, all of whom have been extraordinary colleagues and have enriched my experience at Caltech. I am especially grateful to Will Greenberg, Sue Swalley and David Herman for their friendship and support while I was writing this thesis.

To my mother and father, to whom this thesis is dedicated, go my deepest gratitude for their love and support from the very beginning.

Abstract

Polyamides containing *N*-methylimidazole and *N*-methylpyrrole amino acids are synthetic ligands that bind DNA by forming side-by-side complexes in the minor groove. Sequence-specificity depends on side-by-side aromatic amino acid pairings: imidazole (Im) opposite pyrrole (Py) targets a G•C base pair, Py/Im targets C•G and Py/Py targets T•A and A•T. This thesis describes work centered on the design of polyamide motifs that increase the binding site size limit, binding affinity and sequence repertoire of polyamides. In addition, polyamides were designed to bind within the regulatory regions of selected genes, which allowed investigations of polyamides as regulators of gene expression.

In Chapter Two, β-alanine was found to optimally link three-ring polyamide subunits in an extended conformation, allowing recognition of 9 and 13 base pair sequences. This result extended the binding site size limit of polyamides and identified a useful polyamide building block. Chapter Three describes the ability of C-terminal amino acids to modulate the DNA-binding affinity and specificity of polyamide dimers. Chapters Seven, Eight, Nine and Ten describe additional polyamide motifs for recognition of 9–16 base pair sequences.

Chapter Four describes eight-ring hairpin polyamides that bind respective six base pair target sequences with subnanomolar affinity and discriminate match from single-base pair mismatch sites. These results showed that polyamides can bind predetermined DNA sequences with affinity and specificity comparable to DNA-binding gene-regulatory proteins.

In Chapters Five and Six, studies of hairpin polyamides that incorporate β-alanine (β) residues to correctly position internal imidazoles showed that the choice of a Py/Py, Py/β or β/β pair depends on the sequence composition of the target site. These results expanded the sequence repertoire of polyamides and contributed to polyamide design guidelines that predict, for a given target sequence, when and where β-alanine residues are

needed.

Chapter Eleven describes polyamides that bind DNA sequences proximal to transcription factor binding sites within the promoter regions of the 5S rRNA gene, the HIV-1 genome and the HER2/neu gene. The results of biological assays using these polyamides demonstrated that polyamides are cell-permeable, can interfere with transcription factor-DNA interactions and can inhibit the transcription of specific genes and viral replication within living cells.

Table of Contents

page

List of Figures and Tables

CHAPTER ONE

page

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

CHAPTER SIX

CHAPTER SEVEN

CHAPTER EIGHT

CHAPTER NINE

CHAPTER TEN

CHAPTER ELEVEN

CHAPTER ONE

Introduction

The genetic blueprint of an individual organism is contained in its deoxyribonucleic acid (DNA). The information in DNA is organized into genes, each of which codes for a particular protein or ribonucleic acid (RNA). The proper development, growth and function of the cells within an organism depends on specific protein-DNA interactions that control gene expression. Synthetic molecules that bind to specific DNA sequences could be useful as pharmaceuticals that regulate gene expression. For such a therapeutic approach to be generally applicable, a class of synthetically accessible ligands that can recognize a wide range of predetermined DNA sequences and can permeate or be delivered into cells is needed.

DNA Structure. Double-helical DNA consists of two polydeoxyribonucleotide strands that pair in an antiparallel orientation through hydrogen bonds between heterocyclic bases (Fig. 1). There are four heterocyclic bases: adenine (A), guanine (G), cytosine (C) and thymine (T).[1] In duplex DNA, the two strands are joined by Watson-Crick pairing of guanine with cytosine (G•C base pair) and adenine with thymine (A•T base pair) (Fig. 2).[2] The helix is also stabilized by stacking interactions between the bases. The most common form of DNA in biological systems is B-DNA, which is characterized by a wide and shallow major groove and a narrow and deep minor groove.[1,2] Sequence-specific DNA-binding ligands generally recognize their target sites by a combination of specific hydrogen bonds with functional groups on the edges of the base pairs in the major or minor grooves, electrostatic interactions between positively-charged functional groups on the ligand and phosphate groups on the DNA backbone, and shape complementarity of the ligand with the binding site.

Protein-DNA Complexes. Within cells, gene transcription is regulated by sequence-specific DNA-binding proteins. Crystal structures of protein-DNA complexes reveal that nature chose a combinatorial approach for DNA recognition. There are a number of conserved motifs for DNA recognition by proteins, including the helix-turn-

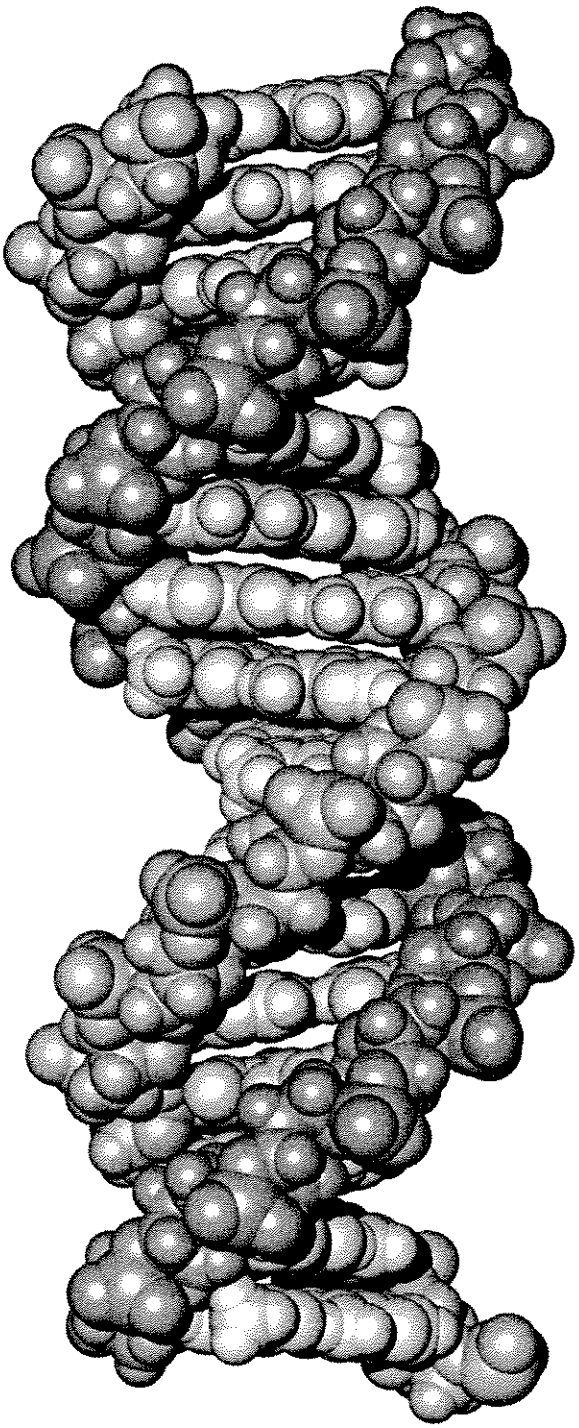


Figure 1. The structure of B-form double-helical DNA.

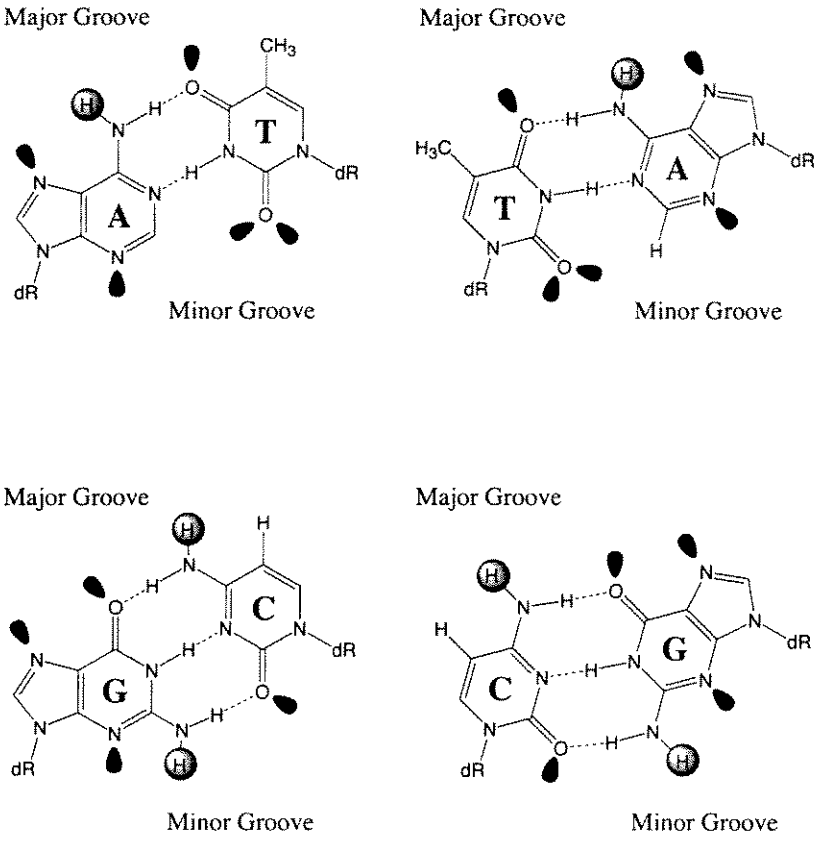


Figure 2. Structures of the four Watson-Crick base pairs. Hydrogen bond donors and acceptors are represented by gray circles and black lobes, respectively. dR denotes the sugar phosphate backbone. The groove location of functional groups is indicated.

helix, homeodomain, leucine zipper and zinc finger motifs (Fig. 3).[3] However, no single motif exists that generates an amino acid-base pair code for all DNA sequences.[4] Due to the complexity of protein-DNA interactions, the design of proteins for recognition of predetermined target sequences has been challenging. One encouraging success is the selection of zinc finger proteins having desired DNA-binding specificities using phage display methodology.[5-10] The ability of a selected zinc finger protein to inhibit gene transcription in transfected cells has been demonstrated.[10] Despite this advance in protein engineering, for therapeutic applications proteins have the drawbacks of being impermeable to cellular membranes and susceptible to degradation by proteases.

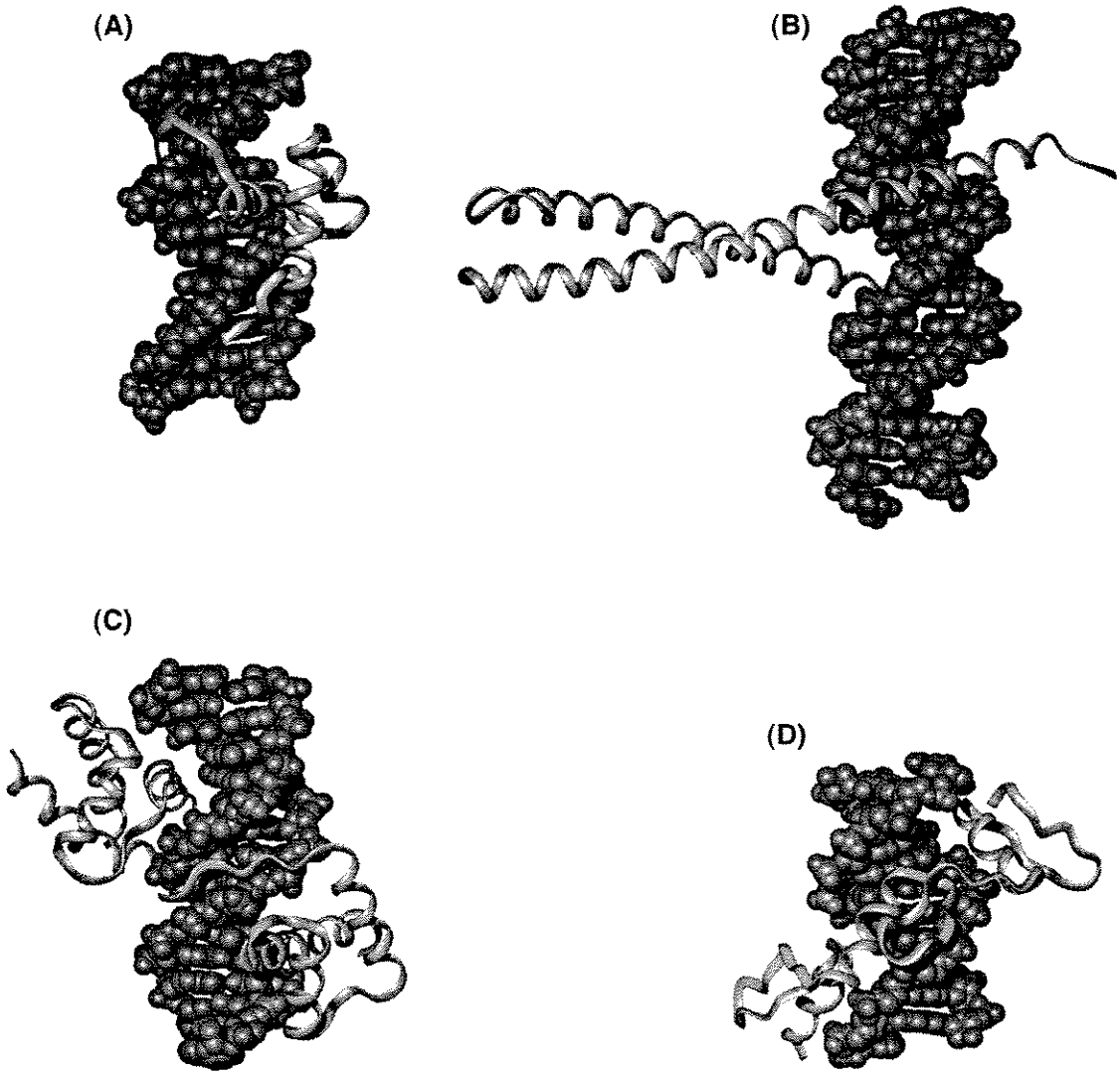
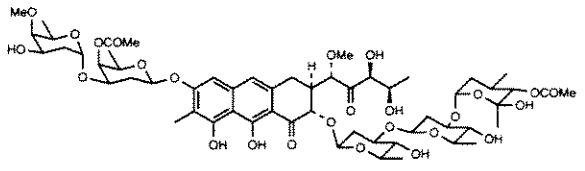
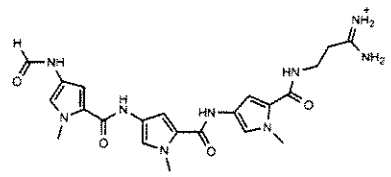


Figure 3. X-ray crystal structures of protein-DNA complexes. (A) Hin recombinase, helix-turn-helix motif;[68] (B) GCN4, leucine zipper motif;[69] (C) Oct-1 POU domain, homeodomain motif;[70] (D) Zif268, zinc finger motif.[71]

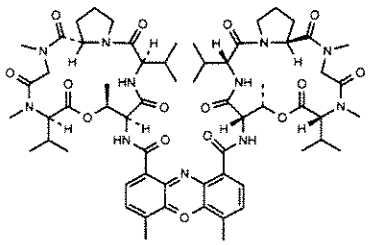
Oligonucleotide-Directed Triple Helix Formation. Moser and Dervan reported in 1987 that a 15 base pair homopyrimidine oligonucleotide could bind sequence-specifically in the major groove of double-helical DNA to a homopurine tract.[11] Specificity within the resulting "pyrimidine motif" triple helix derived from T•AT and C^+•GC base triplets wherein a pyrimidine base on the third strand forms hydrogen bonds with purine



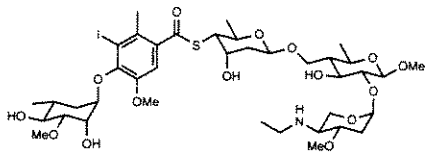
Chromomycin • 5'-GGCT-3'



Distamycin • 5'-(A,T)$_5$-3'



Actinomycin D • 5'-GC-3'



Calicheamicin oligosaccharide • 5'-TCCT-3'

Figure 4. Examples of DNA-binding natural products. An optimal recognition sequence is indicated for each compound.

bases of the Watson-Crick duplex.[12-15] A "purine motif" for triple helix formation was subsequently identified, in which a purine-rich oligonucleotide binds in the major groove through formation of G•GC, A•AT and T•AT triplets.[16,17] It has been demonstrated that triple helix-forming oligonucleotides can inhibit gene transcription *in vitro*.[18,19] The advantages of the triple helix approach are the ability to easily recognize DNA sequences 15–16 base pairs in size and very high specificity for match versus single-base pair mismatch sequences.[20,21] However, recognition is limited to purine tracts, and oligonucleotides are not readily cell-permeable and are susceptible to degradation by nucleases.

DNA-Binding Small Molecules. A structurally diverse array of naturally-occurring small molecules bind DNA.[22,23] Examples include chromomycin, actinomycin D, calicheamicin oligosaccharide and the polyamides netropsin and distamycin A (Fig. 4). These compounds generally recognize DNA either by binding in the minor groove and/or by intercalating between the base pairs. A limitation of most DNA-binding small molecules is their modest sequence specificity, relatively small binding site size and inability to target

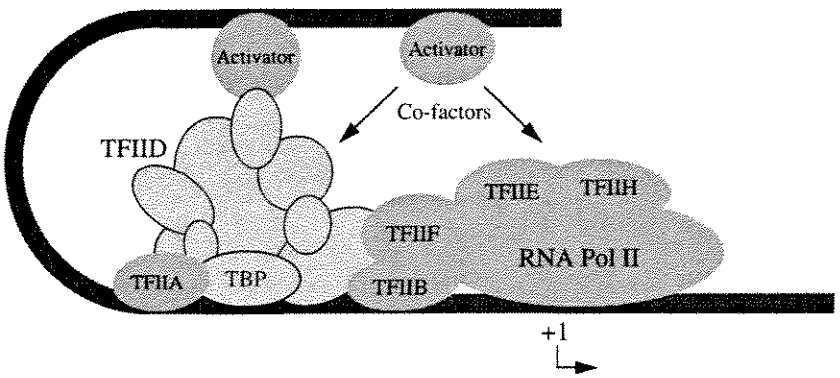


Figure 5. Schematic representation of an RNA polymerase II transcription preinitiation complex. The gray line and ovals represent DNA and proteins, respectively. The transcription start site (+1) is indicated.

predetermined DNA sequences. For therapeutic applications, however, small molecules are particularly attractive compared to proteins and oligonucleotides because they have the potential to be readily cell-permeable.

Regulation of Gene Expression by DNA-Binding Small Molecules. In biological systems, gene expression is regulated primarily at the level of transcription. The transcription of a particular gene is activated by formation of a multiprotein complex composed of "transcription factor" proteins that assembles on DNA, generally upstream of the site of transcription initiation.[24,25] This "transcription preinitiation complex" correctly positions an RNA polymerase to transcribe the gene (Fig. 5). Assembly of the preinitiation complex involves multiple sequence-specific DNA-binding transcription factors.

It has been demonstrated that some DNA-binding small molecules can permeate cells, disrupt transcription factor-DNA complexes and modulate gene expression.[26-33] For example, a carbohydrate-based ligand derived from the natural product calicheamicin has been shown to block the DNA-binding activity of the transcription factor NFAT and to inhibit NFAT-dependent transcription in cells.[33] However, oligosaccharides cannot currently recognize a broad range of DNA sequences. A broadly applicable approach to gene regulation by small molecules requires the development of a class of synthetically accessible molecules whose structure can be rationally tailored to provide recognition of a

broad range of predetermined DNA sequences.

DNA-Binding Assays. Several assays are used to examine the DNA-binding properties of small molecules. Equilibrium association constants of ligands for a particular DNA sequence are determined by quantitative DNase I footprinting experiments.[34-36] In this assay, the ligand is incubated with a radioactively end-labeled DNA fragment, the DNA partially digested with DNase I, and the resulting DNA fragments separated on a denaturing gel and visualized by autoradiography. Regions of cleavage protection reveal ligand binding sites. By quantitating the extent of cleavage protection as a function of ligand concentration, an equilibrium binding isotherm can be generated. The equilibrium association constant, K_a, is equal to the reciprocal of the ligand concentration at half-saturation (Fig. 6).

Since DNase I does not cleave all sequences equally, a different footprinting reagent is required to determine ligand binding sites to nucleotide resolution. The reagent methidiumpropyl-EDTA•Fe(II) (MPE•Fe(II)), which cleaves DNA with minimal sequence specificity in the presence of a reducing agent and O_2, is useful for this purpose.[37-39] The binding orientation of a ligand can be determined using a technique called affinity cleaving, in which the DNA-cleaving moeity EDTA•Fe(II) is covalently attached to the ligand (Fig. 7).[40-42] Both MPE•Fe(II) and affinity cleaving experiments are useful for determining the groove location of a ligand, since distinct cleavage patterns are observed for cleavage in the major versus the minor groove.

Distamycin and Netropsin. Distamycin and netropsin are naturally occurring oligopyrrolecarboxamide antibiotics that preferentially form 1:1 ligand:DNA complexes in the minor groove of DNA at A,T-rich sequences (Fig. 8).[37-39,43-48] Their structural simplicity relative to other DNA-binding natural products (see Fig. 4) makes these molecules attractive leads for the development of synthetic analogs having altered DNA-binding sequence-specificity. The compounds are crescent-shaped, providing for extensive

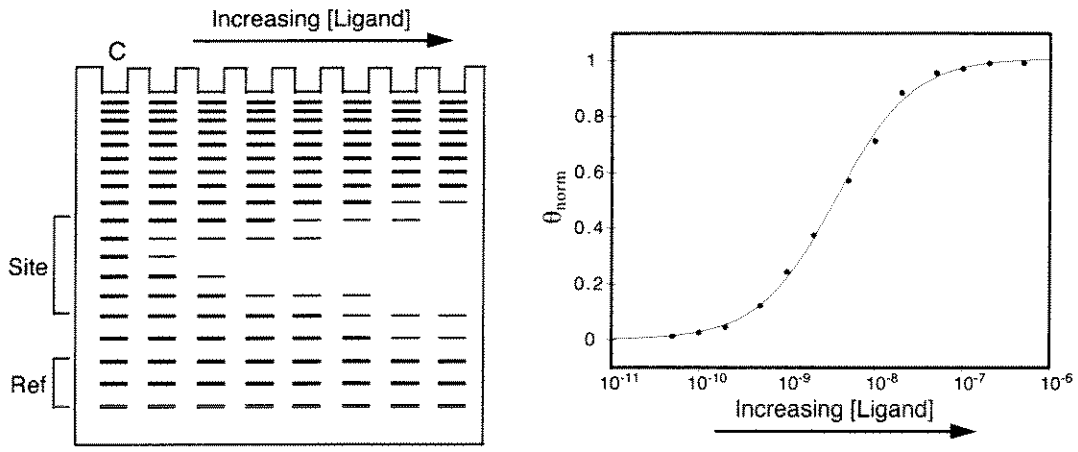


Figure 6. (Left) Graphical representation of a gel from a quantitative DNase I footprint titration experiment. (Right) Example of a binding isotherm derived from such a gel. Fractional occupancy, θ_{norm}, is plotted versus ligand concentration. The equilibrium association constant, K_a, is equal to the reciprocal of the ligand concentration at half occupancy ($\theta_{norm} = 0.5$).

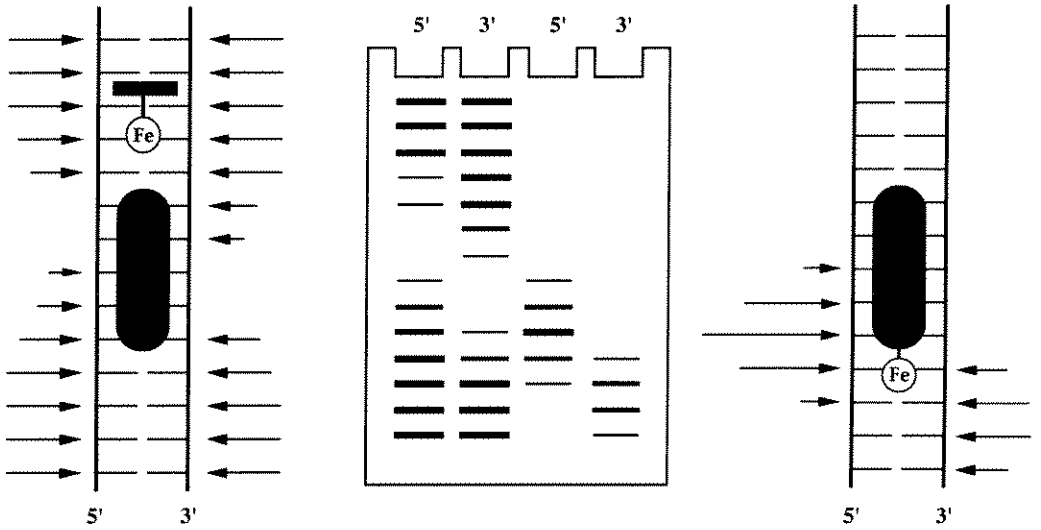


Figure 7. Schematic of (left) MPE•Fe(II) footprinting and (right) affinity cleaving techniques. (Center) Cleavage products obtained on a denaturing gel.

van der Waals contacts with the walls of the minor groove. Additional stabilization of polyamide:DNA complexes is provided by electrostatic interactions between positively-charged amidinium groups on the compounds and the negatively-charged DNA backbone, and by hydrogen bonding between the amide hydrogens on the polyamide with the N3 of

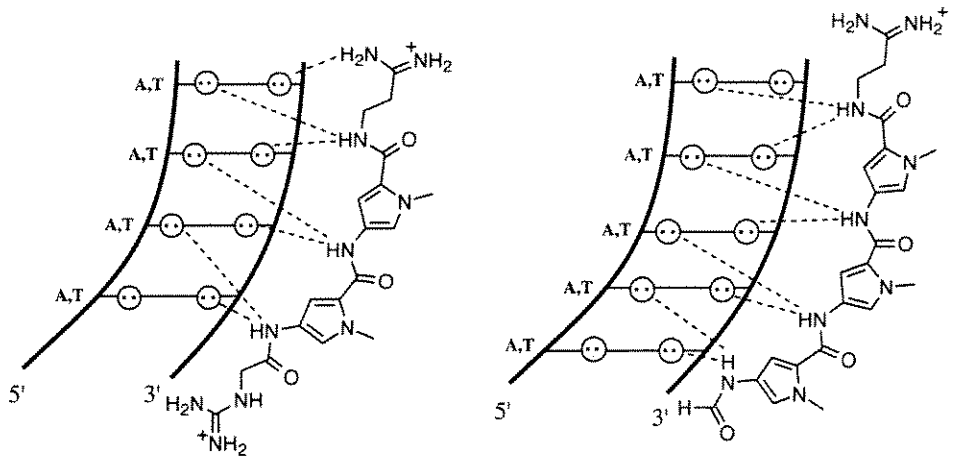


Figure 8. Binding models of the complexes formed between (left) netropsin and (right) distamycin with A,T-rich DNA sequences. Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dashed lines.

adenine and O2 of thymine. Specificity for A,T versus G,C base pairs results in part from a steric clash between the ligand and the exocyclic amino group of guanine. The guanine amino group also introduces a hydrogen bond mismatch.

Pyrrole-Imidazole Polyamides. Since the polyamides netropsin and distamycin are limited to recognition of A,T-containing sequences, an important goal was to design analogs that bind DNA sites containing G,C as well as A,T base pairs. G,C base pairs differ from A,T base pairs in the minor groove by the presence of the guanine exocyclic amino group, which protrudes from the floor of the groove and introduces a hydrogen bond donor (see Fig. 2). It was envisioned that analogs of distamycin containing a hydrogen bond acceptor could recognize a G,C base pair by formation of a hydrogen bond between the ligand and the guanine exocyclic amino group. In 1987, an analog of distamycin was prepared in which the N-terminal pyrrole was replaced with pyridine, (PyrPyPy-Dp, Pyr = pyridinecarboxamide, Py = *N*-methylpyrrolecarboxamide, and Dp = dimethylaminopropylamide) (Fig. 9).[49] While the molecule was expected to bind sequences having the form 5'-(A,T)G(A,T)$_3$-3', footprinting experiments revealed binding to 5'-(A,T)G(A,T)C(A,T)-3' sequences.[49] In addition, affinity cleavage experiments

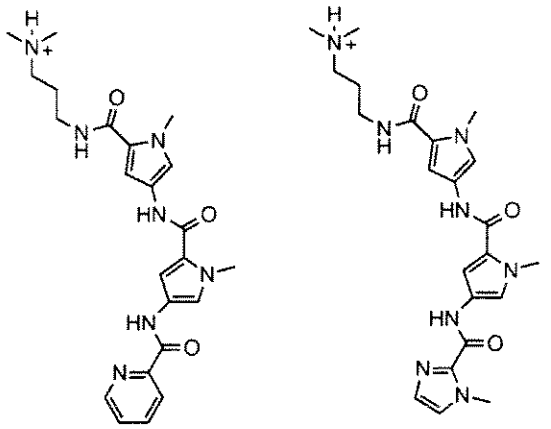


Figure 9. Structures of the polyamides (left) PyrPyPy-Dp and (right) ImPyPy-Dp.

indicated no orientation preference for the polyamide. The analogous *N*-methylimidazole containing compound ImPyPy-Dp (Im = *N*-methylimidazolecarboxamide) (Fig. 9) was found to bind 5'-(A,T)G(A,T)C(A,T)-3' sequences with higher sequence-specificity than PyrPyPy-Dp.[50] These results were inconsistent with the 1:1 polyamide:DNA model.

In 1989 the first 2:1 polyamide-DNA complex was characterized by NMR spectroscopy.[51] In this complex, two distamycin molecules stack side-by-side in an antiparallel orientation in the minor groove (Figs. 10 and 11). In light of this result, it was proposed, and subsequently verified by 2D NMR experiments, that the synthetic polyamides PyrPyPy-Dp and ImPyPy-Dp recognize 5'-(A,T)G(A,T)C(A,T)-3' sequences through formation of 2:1 complexes in the minor groove (Fig. 12).[52] The molecules bind in the minor groove in an antiparallel, side-by-side arrangement analogous to the 2:1 distamycin:DNA complex. Consistent with the original design, recognition of G,C base pairs occurs by formation of a hydrogen bond between a nitrogen atom on the polyamide pyridine or imidazole moeity with the guanine exocyclic amino group. No 1:1 complexes were detected in these experiments, indicating that the dimeric complex forms with high cooperativity.[53] The preference of the synthetic polyamides for 2:1 versus 1:1 complexes with DNA likely reflects the sequence-dependent width of the minor groove. For "mixed" DNA sequences containing both A,T and G,C base pairs, the minor groove width is about

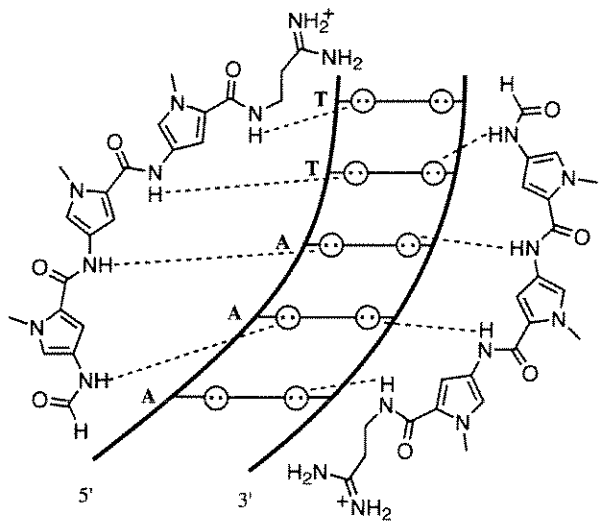


Figure 10. Model of the 2:1 complex of distamycin with 5'-AAATT-3'. Circles with dots represent the lone pairs on N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated with dashed lines.

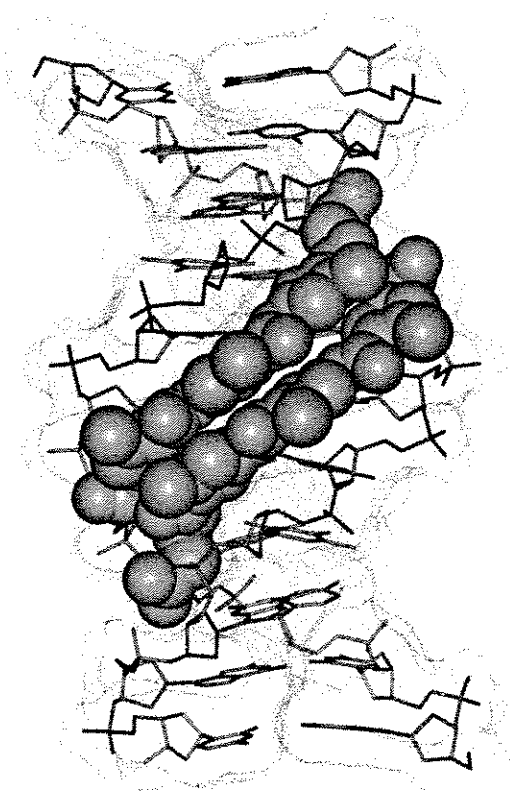


Figure 11. NMR structure of the $(distamycin)_2$•5'-AAATT-3' complex.

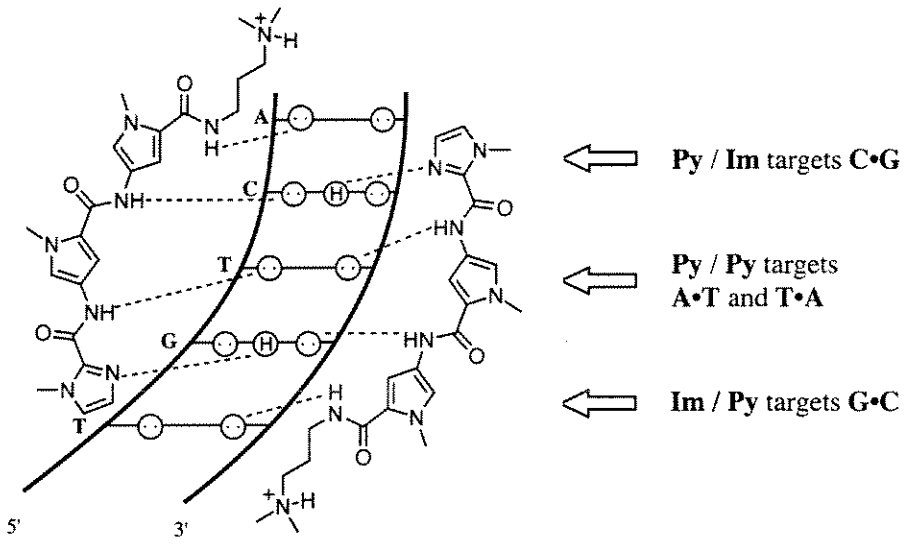


Figure 12. Model of the 2:1 complex of ImPyPy-Dp with 5'-TGTCA-3'. Circles with dots represent the lone pairs on N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated with dashed lines. Pairing rules for DNA recognition by polyamides are shown at right.

5–6 Å, while all-A,T sequences generally have a narrower groove width of about 3–4 Å.[54] Formation of 2:1 complexes with mixed DNA sequences is favored because this ligand arrangement maximizes van der Waals interactions with the walls of the minor groove.

In the (ImPyPy-Dp)$_2$•5'-(A,T)G(A,T)C(A,T)-3' complex, each polyamide makes specific contacts with one strand on the floor of the minor groove. The correspondence of each ring in the polyamide with a particular DNA base suggested general pairing rules for DNA recognition by polyamides (Fig. 12). A pairing of imidazole opposite pyrrole recognizes a G•C base pair, and pyrrole opposite imidazole recognizes C•G.[50,52,53] The pairing of pyrrole opposite pyrrole is partially degenerate and recognizes both A•T and T•A.[51,55] The generality of the 2:1 model was demonstrated by targeting other sequences of mixed A,T/G,C composition. ImPyPy-Dp and distamycin bind simultaneously to a 5'-(A,T)G(A,T)$_3$-3' site as an antiparallel heterodimer.[56,57] A PyImPy polyamide and distamycin bind 5'-(A,T)$_2$G(A,T)$_2$-3',[58,59] and ImPyImPy-Dp targets 5'-(A,T)GCGC(A,T)-3' (Figs. 13 and 14).[60,61]

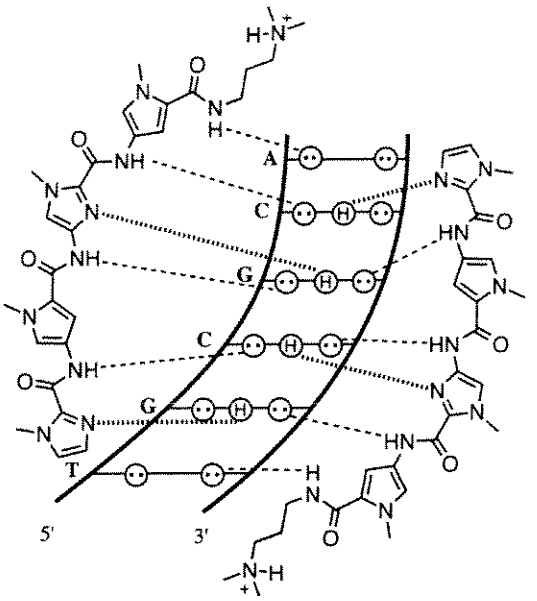


Figure 13. Model of the complex of ImPyImPy-Dp with 5'-TGCGCA-3'. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated with dashed lines.

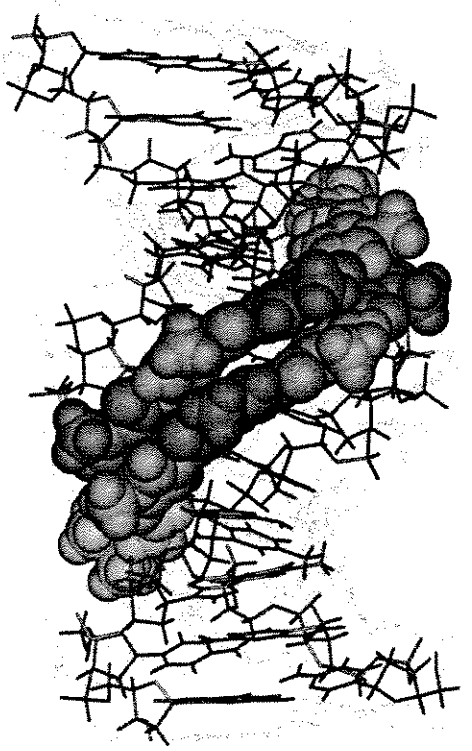


Figure 14. NMR structure of the $(ImPyImPy\text{-}Dp)_2$•5'-TGCGCA-3' complex.

Binding Site Size Limit. The binding site size limit of fully overlapped polyamide dimers was investigated using a series of compounds having the general sequence $Im(Py)_x$-Dp. Quantitative footprinting experiments revealed that, for polyamides longer than five rings (corresponding to a 7 bp binding site), the binding affinity ceases to increase with polyamide size, and the specificity decreases.[62] These results indicated that the *N*-methylimidazole and *N*-methylpyrrole residues fail to maintain an appropriate base-pair register across the entire length of the polyamide-DNA complex. With a view to use of polyamides for gene regulation, recognition of sequences longer than 5–7 base pairs may be necessary to provide optimal gene-specificity.

Hairpin Polyamides. In parallel with the development of the pairing rules described above have been efforts to increase the DNA-binding affinity and specificity of pyrrole-imidazole polyamides by linking polyamide subunits.[63-66] The binding affinities of unlinked polyamide-DNA complexes are relatively low. In addition, for heterodimeric complexes, one or both of the polyamides may independently bind other sequences with an affinity comparable to that of the heterodimeric complex. The second-generation linked polyamide ImPyPy-γ-PyPyPy-Dp (γ = γ-aminobutyric acid) was shown to specifically bind the sequence 5'-TGTTA-3' in a "hairpin" conformation with an equilibrium association constant, K_a, of 8 x 10^7 M^{-1}, an increase of about 300-fold over the unlinked polyamides (Fig. 15).[66,67] With regard to the use of polyamides as regulators of gene expression, it remained to develop polyamides that bind DNA with subnanomolar affinities comparable to natural DNA-binding transcription factors.

Description of this Work. The pairing rules for DNA recognition by pyrrole-imidazole polyamides function in the context of a polyamide motif that correctly positions the Py and Im residues. This thesis describes work centered on the design and characterization of new polyamide motifs that increase the binding site size limit, binding affinity and sequence repertoire of polyamides. In addition, polyamides were designed to bind within the regulatory regions of selected genes, which allowed initial investigations of

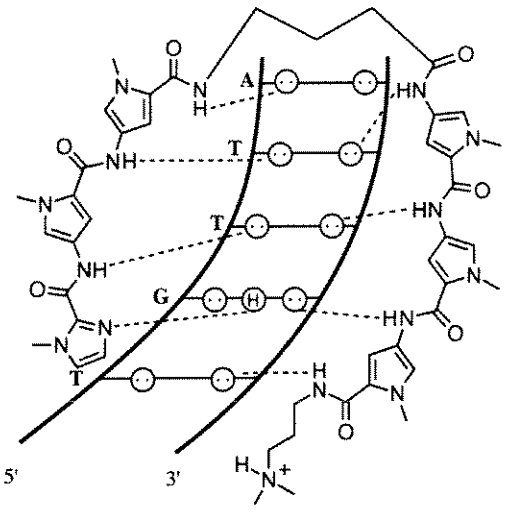


Figure 15. Model of the hairpin complex of ImPyPy-γ-PyPyPy-Dp with 5'-TGTTA-3'. Circles with dots represent the lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated with dashed lines.

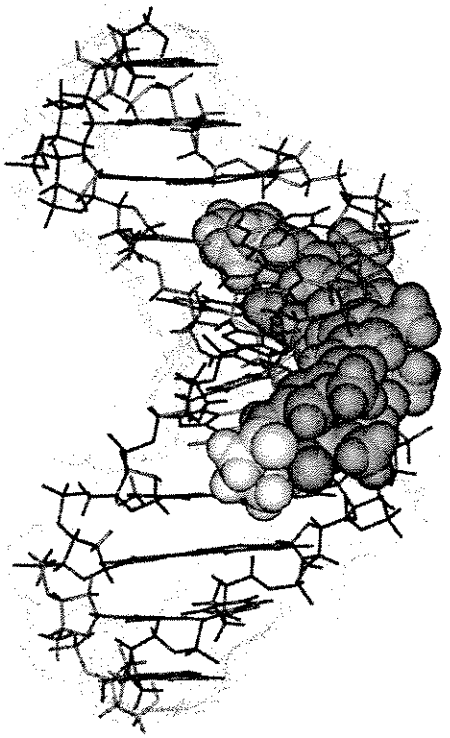


Figure 16. NMR structure of the ImPyPy-γ-PyPyPy-Dp•5'-TGTTA-3' complex.

polyamides as regulators of gene expression.

In Chapter Two, the ability of six-ring polyamides consisting of three-ring subunits linked by *N*-methylpyrrole, glycine, β-alanine or γ-aminobutyric acid to recognize 9 and 13 base pair target sequences was investigated. Results from quantitative DNase I footprint titration experiments demonstrated that β-alanine optimally links three-ring subunits in an extended conformation. This result extended the binding site size limit of polyamides and identified a useful polyamide building block. Chapter Three describes the ability of C-terminal amino acids to modulate the DNA-binding affinity and specificity of polyamide dimers.

Chapter Four describes the DNA-binding properties of two eight-ring hairpin polyamides, in which four-ring subunits are connected by γ-aminobutyric acid (4-γ-4 motif). Quantitative DNase I footprint titration experiments revealed that both polyamides bind respective 6 base pair target sequences with subnanomolar affinity, and discriminate match from single-base pair mismatch sites. These results showed that polyamides can bind predetermined DNA sequences with affinity and specificity comparable to natural DNA-binding gene-regulatory proteins.

Chapters Five and Six describe hairpin polyamides that incorporate β-alanine residues to correctly position internal imidazoles. In Chapter Five, the DNA-binding properties of eight-ring hairpin polyamides in which 2-β-2 subunits are connected by γ-aminobutyric acid are compared with those of analogous ten-ring hairpins. For two of the three 7 base pair target sequences tested, the β-alanine residues are essential for specific, high-affinity binding. In Chapter Six, ten hairpin polyamides that target four different 6–7 base pair sequences were synthesized to compare Py/β and β/β pairings. DNA-binding studies with these polyamides reveal that the choice of a Py/β or β/β pair depends on the sequence composition of the target site. These results contribute to polyamide design guidelines that predict, for a given target sequence, when and where β-alanine residues are needed.

Chapters Seven, Eight, Nine and Ten describe polyamide motifs for recognition of 9–16 base pair sequences. In Chapter Seven, an eight-ring polyamide is shown to recognize a 16 base pair target site through formation of a dimeric complex. Chapter Eight describes a hairpin polyamide having a C-terminal PyPyPy extension that binds to a 9 base pair target sequence with subnanomolar affinity. In Chapters Nine and Ten, hairpin polyamides having two- or three-ring extensions are shown to recognize 10–15 base pair sequences by formation of cooperative homodimeric complexes.

Chapter Eleven describes hairpin polyamides designed to bind DNA sequences proximal to transcription factor binding sites located within the promoter regions of the 5S rRNA gene, the HIV-1 genome and the HER2/neu gene. The results of transcription inhibition experiments and other biological assays carried out by our collaborators are highlighted. These experiments demonstrated that polyamides are cell-permeable, can interfere with transcription factor-DNA interactions, and can inhibit the transcription of specific genes and viral replication within living cells.

References

(1) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(2) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737-738.

(3) Branden, C.; Tooze, J. *Introduction to Protein Structure*, Garland Publishing, Inc.: New York and London, 1991.

(4) Pabo, C. O.; Sauer, R. T. *Ann. Rev. Biochem.* **1992**, *61*, 1053-95.

(5) Desjarlais, J. R.; Berg, J. M. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 2256-2260.

(6) Rebar, E. J.; Pabo, C. O. *Science* **1994**, *263*, 671-673.

(7) Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11163-11167.

(8) Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11168-11172.

(9) Jamieson, A. C.; Kim, S.-H.; Wells, J. A. *Biochem.* **1994**, *33*, 5689-5695.

(10) Choo, Y.; Sánchez-García, I.; Klug, A. *Nature* **1994**, *372*, 642-645.

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

(12) Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 637-640.

(13) Rajagopal, P.; Feigon, J. *Biochem.* **1989**, *28*, 7859-7870.

(14) Sklenár, V.; Feigon, J. *Nature* **1990**, *345*, 836-838.

(15) Radhakrishnan, I.; Gao, X.; de los Santos, C.; Live, D.; Patel, D. J. *Biochem.* **1991**, *30*, 9022-9030.

(16) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360-1363.

(17) Radhakrishnan, I.; de los Santos, C.; Patel, D. J. *J. Mol. Biol.* **1991**, *221*, 1403-1418.

(18) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. J. *Science* **1988**, *241*, 456-459.

(19) Maher, L. J.; Dervan, P. B.; Wold, B. *Biochem.* **1992**, *31*, 70-81.

(20) Best, G.C.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 1187–1193.

(21) Greenberg, W.A.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 5016–5022.

(22) Krugh, T. R. *Curr. Opin. Struct. Biol.* **1994**, *4*, 351-364.

(23) Kahne, K. *Chem. and Biol.* **1995**, *2*, 7-12.

(24) Roeder, R.G. *Trends Bioch.* **1996**, *21*, 327–335.

(25) Verrijzer, C.P.; Tjian, R. *Trends Bioch.* **1996**, *21*, 338–342.

(26) Welch, J.J.; Rauscher, F.J.; Beerman, T.A. *J. Biol. Chem.* **1994**, *49*, 31051-31058.

(27) Dorn, A.; Affolter, M.; Mueller, M.; Gehring, W.J.; Leupin, W. *EMBO J.* **1992**, *11*, 279-286.

(28) Chiang, S.Y.; Welch, J.; Rauscher, F.J.; Beerman, T.A. *Biochemistry* **1994**, *33*, 7033-7040.

(29) Bellorini, M.; Moncollin, V.; D'Incalci, M.; Mongelli, N.; Mantovani, R. *Nucleic Acids Res.* **1995**, *23*, 1657-1663.

(30) Blume, S.W.; Snyder, R.C.; Ray, R.; Thomas, S.; Koller, C.A.; Miller, D.M. *J. Clinical Investigation* **1991**, *88*, 1613-1621.

(31) Ray, R.; Thomas, S.; Miller, D.M. *Am. J. Med. Sci.* **1990**, *299*, 203-208.

(32) Liu, C.; Smith, B.M.; Ajito, K.; Komatsu, H.; Gomez-Paloma, L.; Li, T.; Theodorakis, E.A.; Nicolaou, K.C.; Vogt, P.K. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 940-944.

(33) Ho, S.N.; Boyer, S.H.; Schreiber, S.L., Danishefsky, S.J.; Crabtree, G.R. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 9203-9207.

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

(35) Brenowitz, M.; Senear, D. R.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8462-8466.

(36) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochem.* **1986**, *25*, 7344-7354.

(37) Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 5470-5474.

(38) Van Dyke, M. W.; Dervan, P. B. *Cold Spring Harbor Symp. Quant. Biol.* **1982**, *47*, 347-353.

(39) Van Dyke, M. W.; Dervan, P. B. *Nucleic Acids Res.* **1983**, *11*, 5555-5567.

(40) Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861-6863.

(41) Schultz, P. G.; Dervan, P. B. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1133-1147.

(42) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *3*.

(43) For a review, see: Dervan, P. B. *Science* **1986**, *232*.

(44) Zimmer, C.; Wåhnert, U. *Prog. Biophys. & Molec. Biol.* **1986**, *47*, 31-112.

(45) Pelton, J. G.; Wemmer, D. E. *Biochem.* **1988**, *27*, 8088-8096.

(46) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376-1380.

(47) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*.

(48) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*.

(49) Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574-1575.

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

(51) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723-5727.

(52) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochem.* **1993**, *32*, 11385-11389.

(53) 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.

(54) Yoon, C.; Privé, G. G.; Goodsell, D. S.; Dickerson, R. E. *Proc. Natl. Acad. Sci.* **1988**, *85*.

(55) White, S.; Baird, E. E.; Dervan, P. B. *Biochem.* **1996**, *35*, 12532-12537.

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

(57) Geierstanger, B. H.; Jacobsen, J. P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochem.* **1994**, *33*, 3055-3062.

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

(59) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474-4482.

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

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

(62) Kelly, J.J.; Baird, E.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6981–6985.

(63) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892-9899.

(64) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900-9906.

(65) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663-3664.

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

(67) de Clairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909-7916.

(68) Marmorstein, R.; Carey, M.; Ptashne, M.; Harrison, S. C. *Nature* **1992**, *356*, 408-414.

(69) Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809-817.

(70) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223-1237.

(71) Kim, Y.; Gelger, J. H.; Hahn, S.; Sigler, P. B. *Nature* **1993**, *365*, 512-520.

CHAPTER TWO

Extension of Sequence-Specific Recognition in the Minor Groove of DNA by Pyrrole-Imidazole Polyamides to 9–13 Base Pairs

The text of this chapter is taken from a published paper that was coauthored with

Prof. Peter B. Dervan, Eldon E. Baird and Milan Mrksich

(Trauger, J.W. *et al. J. Am. Chem. Soc.* **1996**, *118*, 6160–6166)

Introduction

The recent discovery and development of 2:1 pyrrole-imidazole polyamide-DNA complexes provides a powerful model for the design of new molecules for sequence-specific recognition in the minor groove of DNA. The polyamide ImPyPy-Dp was shown to specifically bind the mixed A,T/G,C sequence 5'-(A,T)G(A,T)C(A,T)-3' as a side-by-side antiparallel dimer.[1] In this complex, each polyamide makes specific contacts with one strand on the floor of the minor groove such that the sequence-specificity depends on the sequence of side-by-side amino acid pairings. A side-by-side pairing of imidazole opposite pyrrole recognizes G•C base-pairs, while a side-by-side pairing of pyrrole opposite imidazole recognizes C•G base-pairs.[1] A side-by-side pyrrole-pyrrole pairing is partially degenerate and targets both A•T and T•A base pairs.[1,2] The generality of the 2:1 model has been demonstrated by targeting other sequences of mixed A,T/G,C composition.[3-5] ImPyPy-Dp and distamycin (PyPyPy) bind simultaneously to a 5'-(A,T)G(A,T)$_3$-3' site as an antiparallel heterodimer.[3] A PyImPy polyamide and distamycin bind 5'-(A,T)$_2$G(A,T)$_2$-3', and ImPyImPy-Dp targets 5'-(A,T)GCGC(A,T)-3'.[4,5]

The binding affinity and sequence-specificity of a non-covalent antiparallel homodimeric or heterodimeric polyamide-DNA complex can be increased by covalently linking the two polyamides.[6,7] The DNA-binding properties of the polyamides ImPyPy-G-PyPyPy-Dp, ImPyPy-β-PyPyPy-Dp, and ImPyPy-γ-PyPyPy-Dp, in which the terminal carboxyl group of ImPyPy and the terminal amine of PyPyPy-Dp are connected with glycine (G), β-alanine (β), and γ-aminobutyric acid (γ), respectively, were recently reported.[7,8] The γ-aminobutyric acid-linked polyamide bound the designated target site 5'-TGTTA-3' with high affinity and sequence-specificity, and exhibited a Langmuir binding isotherm in quantitative footprinting experiments consistent with formation of an intramolecular "hairpin" complex in which the polyamide folds back on itself.[7] Modeling suggested that the glycine- and β-alanine-linked polyamides could not favorably bind as "hairpins" in the minor groove of DNA. Moreover, these polyamides exhibited cooperative

binding isotherms in quantitative footprinting experiments, consistent with two polyamides binding in extended conformations as *intermolecular dimers*.[7,8] It appears that the glycine- and β-alanine-linked polyamides disfavor binding in the hairpin conformation and are restricted to binding in an extended conformation.[7,8] In a formal sense, there are multiple extended binding motifs (and hence multiple binding site sequences) for polyamides of sequence composition $Im(Py)_x$-Dp, as discussed below.

"Overlapped" and "Slipped" Binding Modes. We report here the DNA-binding affinities of four polyamides having the general sequence ImPyPy-X-PyPyPy-Dp, where X = Py, G, β, or γ, to the 9 bp site 5'-TGTTAAACA-3' and to the 13 bp sites 5'-AAAAAGACAAAAA-3' and 5'-ATATAGACATATA-3'. The polyamides having internal *N*-methylpyrrole, glycine and β-alanine residues were anticipated to bind the 9 and 13 bp sites in an extended conformation. It was not clear at the outset if the γ-aminobutyric acid-linked polyamide ImPyPy-γ-PyPyPy-Dp would bind in an extended or "hairpin" conformation to the targeted sites. For ImPyPy-X-PyPyPy-Dp polyamides binding in an extended conformation, the polyamide-DNA complexes expected to form at the 9 bp and 13 bp target sites represent two distinct binding modes, which we refer to as "overlapped" and "slipped," respectively. In the "overlapped" (9 bp) binding mode, two ImPyPy-X-PyPyPy-Dp polyamides bind directly opposite one another (Fig. 1A). The "slipped" (13 bp) binding mode integrates the 2:1 and 1:1 polyamide-DNA binding motifs at a single site. In this binding mode, the ImPyPy moieties of two ImPyPy-X-PyPyPy-Dp polyamides bind the central 5'-AGACA-3' sequence in a 2:1 manner as in the ImPyPy homodimer,[1] and the PyPyPy moeities of the polyamides bind the all-A,T flanking sequences as in the 1:1 complexes of distamycin (Fig. 1B). The structure of the complex formed by ImPyPy-G-PyPyPy-Dp with a 13 bp target site has been characterized by 2D NMR.[9]

In the 9 bp "overlapped" and 13 bp "slipped" binding sites described above, the G•C and C•G base pairs are separated by one and five A,T base pairs, respectively. While we have concentrated here on the these sites, we note that "partially slipped" sites of 10, 11

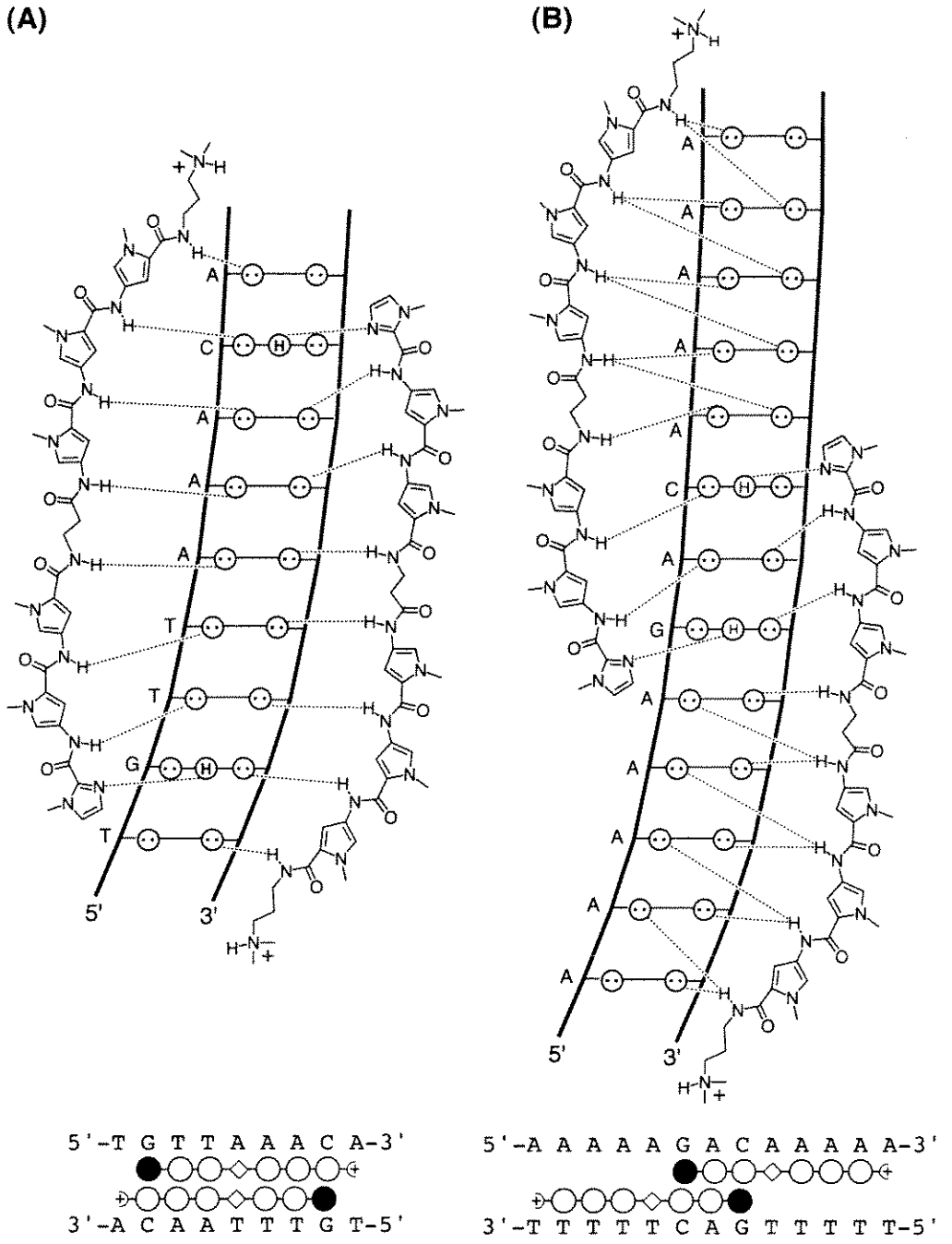


Figure 1. Models of the expected complexes of ImPyPy-β-PyPyPy-Dp (**3**) with the sites (A) 5'-TGTTAAACA-3' (9 bp, "overlapped") and (B) 5'-AAAAAGACAAAAA-3' (13 bp, "slipped"). (Top) Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dashed lines. (Bottom) The filled and open circles represent imidazole and pyrrole rings, respectively, and the diamonds represent β-alanine.

and 12 base pairs in which the G•C and C•G base pairs are separated by two, three, and four A,T base pairs, respectively, are also potential binding sites of the polyamides studied here.

Studies of the energetics of distamycin binding have shown that while the binding affinities are similar for complexation to poly[d(A-T)]•poly[d(A-T)] and poly[d(A)]•poly[d(T)], the origins of these binding affinities are different.[10] Binding to the alternating copolymer is enthalpy driven, while binding to the homopolymer is entropy driven.[10] However, not all 5 bp sites $(A,T)_5$ are bound with equal affinity. By quantitative footprinting experiments, the synthetic distamycin analog Ac-PyPyPy-Dp (Ac = acetyl) was shown to bind the sites 5'-AATAA-3' and 5'-TTAAT-3' with 2-fold and 14-fold lower affinity, respectively, relative to the site 5'-AAAAA-3'.[1c] On the basis of this result, we anticipated that the 13 bp site 5'-AAAAAGACAAAAA-3' may be bound with higher affinity than the 13 bp site 5'-ATATAGACATATA-3'.

Equilibrium association constants of polyamides **1–4** (Fig. 2) for the three targeted sites 5'-TGTTAAACA-3', 5'-AAAAAGACAAAAA-3' and 5'-ATATAGACATATA-3' were determined by quantitative DNase I footprint titration experiments.

Results

Synthesis of Polyamides. The polyamides ImPyPy-Py-PyPyPy-Dp (**1**), ImPyPy-G-PyPyPy-Dp (**2**), ImPyPy-β-PyPyPy-Dp (**3**) and ImPyPy-γ-PyPyPy-Dp (**4**) were prepared as described previously.[7,11]

Footprinting. Quantitative DNase I footprinting[12,13] was carried out with polyamides **1–4** on the 3'-^{32}P-labelled 281-bp pJT3 *Afl* II/*Fsp* I restriction fragment (Figs. 3 and 4) (10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, 22 °C). Of the four ImPyPy-X-PyPyPy-Dp polyamides, three (X = Py, G, β) bind to both the 9 bp "overlapped" site 5'-TGTTAAACA-3' and the 13 bp "slipped" site 5'-AAAAAGACAAAAA-3' with affinities greater than 5×10^7 M^{-1} (Table 1) and display

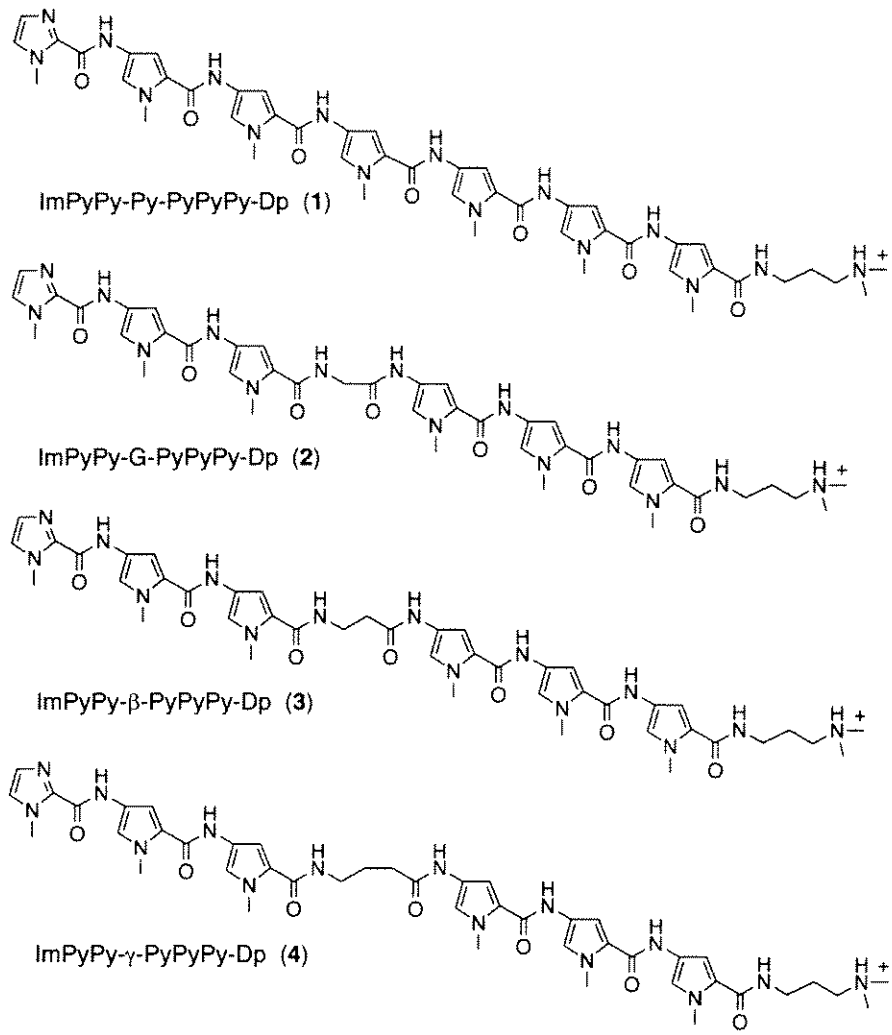
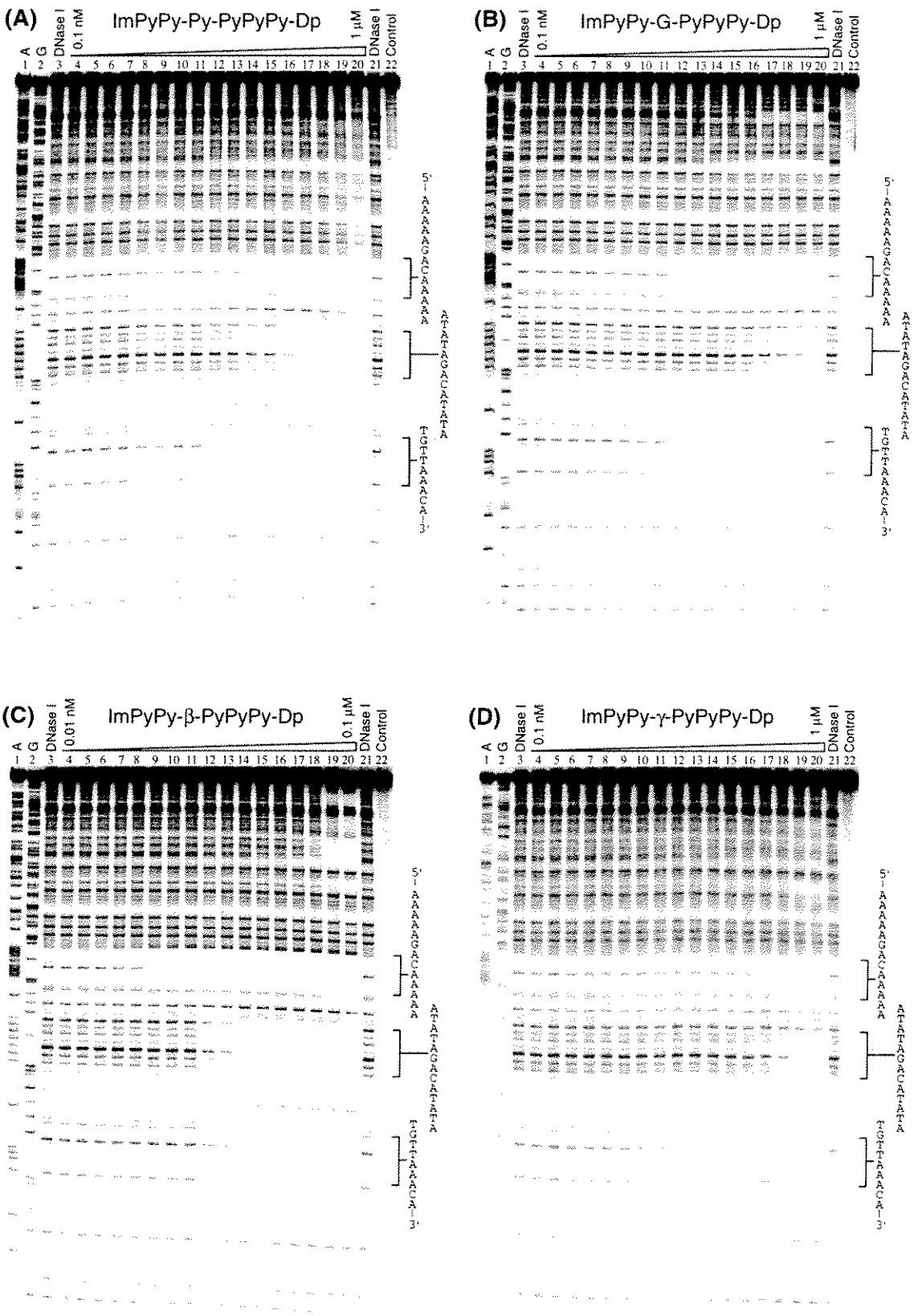


Figure 2. Structures of the polyamides **1–4**.

```
5'-GCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTG
3'-CGTTGACAACCCTTCCCGCTAGCCACGCCCGGAGAAGCGATAATGCGGTCGACCGCTTTCCCCCTACACGAC

CAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGC
GTTCCGCTAATTCAACCCATTGCGGTCCCAAAAGGGTCAGTGCTGCAACATTTTGCTGCCGGTCACTTAAGCTCG

TCGGTACCCGGGAACGTAGCGTACCGGTGCAAAAAGACAAAAAGGCTCGACGCCGCATATAGACATATAGGGCCG
AGCCATGGGCCCTTGCATCGCATGGCCACGTTTTTCTGTTTTTCCGAGCTGCGGCGTATATCTGTATATCCCGGC

TCGACGCTGTTAAACAGGCTCGACGCCAGCTCGTCCTAGCTAGCGTCGTAGCGTCTTAA-3'
AGCTGCGACAATTTGTCCGAGCTGCGGTCGAGCAGGATCGATCGCAGCATCGCAGAATT-5'
```

Figure 3. Sequence of the 281 bp pJT3 *Afl* II/*Fsp* I restriction fragment. The three binding sites that were analyzed in quantitative footprint titration experiments are indicated.

Figure 4. Storage phosphor autoradiogram of 8% denaturing polyacrylamide gels used to separate the fragments generated by DNase I digestion in quantitative footprint titration experiments: lanes 1-2, A and G sequencing lanes; lanes 3 and 21, DNase I digestion products obtained in the absence of polyamide; lanes 4-20, DNase I digestion products obtained in the presence of 0.1 nM (0.01 nM), 0.2 nM (0.02 nM), 0.5 nM (0.05 nM), 1 nM (0.1 nM), 1.5 nM (0.15 nM), 2.5 nM (0.25 nM), 4 nM (0.4 nM), 6.5 nM (0.65 nM), 10 nM (1 nM), 15 nM (1.5 nM), 25 nM (2.5 nM), 40 nM (4 nM), 65 nM (6.5 nM), 100 nM (10 nM), 200 nM (20 nM), 500 nM (50 nM), 1 μM (0.1 μM) (concentrations used for polyamide ImPyPy-β-PyPyPy-Dp only are in parentheses); lane 22, intact DNA. The targeted binding sites are indicated on the right side of the autoradiograms. All reactions contain 15 kcpm restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.


(A)
ImPyPy-Py-PyPyPy-Dp
DNase I
0.1 nM
1 μM
DNase I
Control
A
G
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
5'-AAAAAGACAAAAA
ATATAGACATATA
TGTTAAACA-3'
(B)
ImPyPy-G-PyPyPy-Dp
DNase I
0.1 nM
1 μM
DNase I
Control
A
G
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
5'-AAAAAGACAAAAA
ATATAGACATATA
TGTTAAACA-3'
(C)
ImPyPy-β-PyPyPy-Dp
DNase I
0.01 nM
0.1 μM
DNase I
Control
A
G
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
5'-AAAAAGACAAAAA
ATATAGACATATA
TGTTAAACA-3'
(D)
ImPyPy-γ-PyPyPy-Dp
DNase I
0.1 nM
1 μM
DNase I
Control
A
G
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
5'-AAAAAGACAAAAA
ATATAGACATATA
TGTTAAACA-3'

cooperative binding isotherms (eq. 2, n=2) at these sites (Fig. 5) *consistent with binding as intermolecular dimers*. The fact that the polyamides ImPyPy-G-PyPyPy-Dp and ImPyPy-β-PyPyPy-Dp bind in the 9 bp "overlapped" binding mode *indicates that the internal glycine and β-alanine amino acids are accommodated opposite a second ligand in a 2:1 polyamide-DNA complex.*

The polyamide ImPyPy-γ-PyPyPy-Dp binds the site 5'-TGTTAAACA-3' with an affinity of 1×10^8 M^{-1}, and also binds the site 5'-AAAAAGACAAAAA-3' with a lower affinity of 6×10^6 M^{-1}. This compound displays Langmuir binding isotherms (eq. 2, n=1) at these sites (Fig. 5), *consistent with binding as an intramolecular hairpin* to the 5 bp "matched" sites 5'-TGTTA-3' and 5'-AAACA-3' (Fig. 6) and to the 5 bp "single base pair mismatch" site 5'-AGACA-3'.[7] Significantly, it appears from these results that ImPyPy-γ-PyPyPy-Dp does not effectively link polyamide subunits in an extended conformation.

Binding Affinities of ImPyPy-X-PyPyPy-Dp, X = Py, G, β, γ. Comparison of the binding affinities of the four polyamides ImPyPy-X-PyPyPy-Dp, where X = Py, G, β, and γ, reveals that the internal amino acid X has a dramatic effect on complex stabilities (Table 1, Fig. 5). The formally *N*-methylpyrrole-linked polyamide ImPyPy-Py-PyPyPy-Dp binds 5'-TGTTAAACA-3' and 5'-AAAAAGACAAAAA-3' with affinities of 1×10^8 M^{-1} and 5×10^7 M^{-1}, respectively. The glycine-linked polyamide ImPyPy-G-PyPyPy-Dp binds both 5'-TGTTAAACA-3' and 5'-AAAAAGACAAAAA-3' with affinities similar (equal and 2-fold higher, respectively) to ImPyPy-Py-PyPyPy-Dp. In contrast, the β-alanine linked polyamide ImPyPy-β-PyPyPy-Dp binds 5'-TGTTAAACA-3' and 5'-AAAAAGACAAAAA-3' with affinities higher than ImPyPy-Py-PyPyPy-Dp by factors of approximately 8 ($K_a = 8 \times 10^8$ M^{-1}) and 85 ($K_a = 5 \times 10^9$ M^{-1}), respectively. Relative to ImPyPy-Py-PyPyPy-Dp, the hairpin-forming polyamide ImPyPy-γ-PyPyPy-Dp binds 5'-TGTTAAACA-3' (a matched hairpin binding site) and 5'-AAAAAGACAAAAA-3' (a mismatched hairpin binding site) with equal and 8-fold lower affinities, respectively.

Table 1. Equilibrium association constants (M^{-1}) for polyamides ImPyPy-X-PyPyPy-Dp, where X = Py, G, β, or γ.[a,b]

	Polyamide			
Binding Site	X = Py	X = G	X = β	X = γ
5'-TGTTAAACA-3'	9.7 (±2.3) x 10^7	1.4 (±0.1) x 10^8	7.8 (±0.6) x 10^8	1.4 (±0.3) x 10^8
5'-AAAAAGACAAAAA-3'	5.4 (±1.5) x 10^7	1.1 (±0.1) x 10^8	≥ 4.7 (±0.7) x 10^9	6.4 (±0.6) x 10^6
5'-ATATAGACATATA-3'	3.6 (±0.5) x 10^7	6.6 (±0.4) x 10^6	1.0 (±0.1) x 10^9	4.6 (±0.5) x 10^6

[a]The reported association constants are the mean values obtained from three DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses. [b]The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

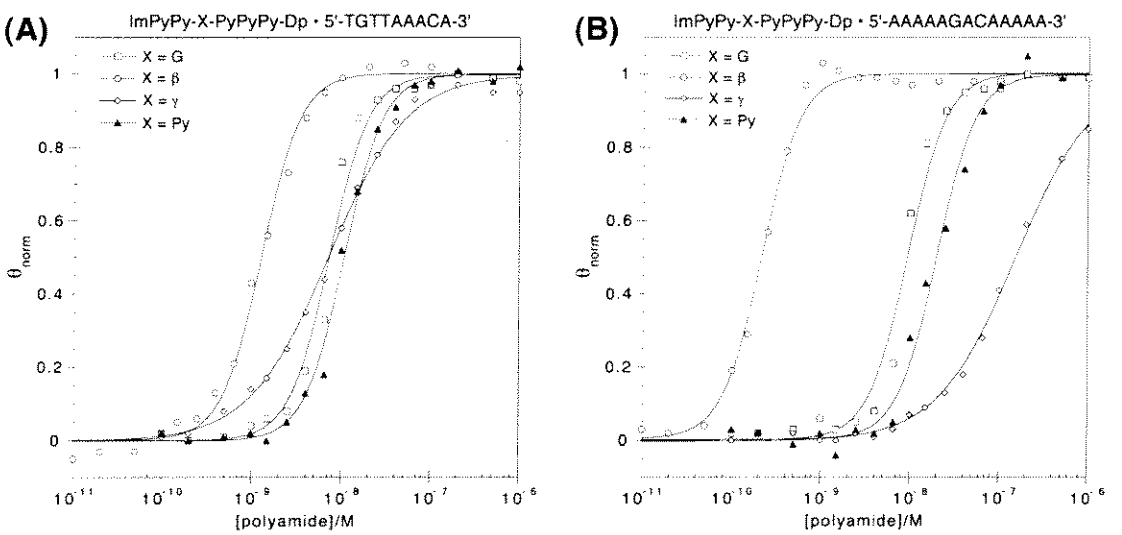


Figure 5. Data obtained from quantitative DNase I footprint titration experiments showing the effect of the internal amino acid X on binding of the polyamides ImPyPy-X-PyPyPy-Dp to (A) 5'-TGTTAAACA-3' and (B) 5'-AAAAAGACAAAAA-3', where X = G, β, γ, or Py. The (θ_{norm}, $[L]_{tot}$) data points were obtained as described in the Experimental Section. Each data point is the average value obtained from three quantitative footprint titration experiments.

Figure 6. Model of the complexes of ImPyPy-γ-PyPyPy-Dp with the site 5'-TGTTAAACA-3' (5 bp, "hairpin").

Specificity for 5'-AAAAAGACAAAAA-3' Versus 5'-ATATAGACATATA-3'. Comparison of the binding affinities of the four polyamides at the 13 bp sites 5'-AAAAAGACAAAAA-3' and 5'-ATATAGACATATA-3' indicates that the specificity between the sites depends on the internal amino acid (Table 1). ImPyPy-G-PyPyPy-Dp and ImPyPy-β-PyPyPy-Dp are approximately 20-fold and ≥5-fold specific, respectively, for 5'-AAAAAGACAAAAA-3' versus 5'-ATATAGACATATA-3'. The polyamides ImPyPy-Py-PyPyPy-Dp and ImPyPy-γ-PyPyPy-Dp bind 5'-AAAAAGACAAAAA-3' and 5'-ATATAGACATATA-3' with similar affinities.

Discussion

Implications for the Design of Minor-Groove Binding Molecules. The results presented here show that *β-alanine is an optimal linker for joining three-ring subunits in an extended conformation, providing a useful structural motif for the design of new polyamides targeted to sites longer than 7 base pairs.* Recently, it has been shown that as the length of a polyamide having the general sequence Im(Py)$_x$-Dp increases beyond 5 rings (corresponding to a 7 bp binding site), the binding affinity ceases to increase with increasing polyamide length, indicating that the *N*-methylimidazole and *N*-methylpyrrole residues fail to maintain an appropriate base pair register across the entire length of the polyamide-DNA complex.[11] Modeling suggests that this effect may be due largely to the greater curvature of polyamides compared to that of the minor groove floor. The higher binding affinities observed here for ImPyPy-β-PyPyPy-Dp relative to ImPyPy-Py-PyPyPy-Dp indicate that the flexible β-alanine linker relieves the register mismatch,

allowing both three-ring subunits to bind optimally. Notably, higher binding affinities are observed for ImPyPy-β-PyPyPy-Dp versus ImPyPy-Py-PyPyPy-Dp despite the higher conformational entropy and lower aromatic surface area of the β-alanine-linked polyamide. The observation here that β-alanine effectively links polyamide subunits within 2:1 polyamide-DNA complexes is consistent with the previously reported finding that β-alanine effectively links polyamide subunits within 1:1 polyamide-DNA complexes.[14]

From the standpoint of binding specificity, the observation here that a single compound can bind in multiple binding modes is problematic. The next generation of pyrrole-imidazole polyamides targeted to binding sites greater than 7 bp in length should incorporate constraints that specify a single binding mode.

The previously described γ-aminobutyric acid-based "hairpin" motif[7] complements the β-alanine-based "extended" motif described here. For ImPyPy-γ-PyPyPy-Dp, the binding isotherms and affinities observed here are consistent with the previous report that γ-aminobutyric acid effectively links polyamide subunits in a "hairpin" conformation[7] and indicate that γ-aminobutyric acid does not effectively link polyamide subunits in an extended conformation. Significantly, these observations indicate that 1) extended binding modes will not compromise the sequence-specificity of hairpin-forming, γ-aminobutyric acid-linked polyamides, and 2) β-alanine and γ-aminobutyric acid linkers could be used within a single polyamide with predictable results.

The results reported here expand the binding site size targetable with pyrrole-imidazole polyamides and provide structural motifs that will facilitate the design of new pyrrole-imidazole polyamides targeted to other sequences.

Experimental Section

Materials. *E. coli* XL-1 Blue competent cells were obtained from Stratagene. Restriction endonucleases were purchased from Boeringher-Mannheim or New England Biolabs. Plasmid isolation kits were obtained from Promega. Plasmid sequencing was

carried at the Sequence Analysis Facility at the California Institute of Technology. Sequenase (version 2.0) was obtained from United States Biochemical, and DNase I (FPLCpure) was obtained from Pharmacia. [α-^{32}P]-Thymidine-5'-triphosphate ($\geq$ 3000 C_i/mmol), [α-^{32}P]-deoxyadenosine-5'-triphosphate ($\geq$ 6000 C_i/mmol), and [γ-^{32}P]-adenosine-5'-triphosphate were purchased from Du Pont/NEN. Water was obtained from a Millipore Milli-Q water purification system.

Construction of Plasmid DNA. Plasmid pJT3 was prepared by standard methods. Briefly, plasmid pJT1 was prepared by hybridization of two complementary sets of synthetic oligonucleotides: 5'-CCGGGAACGTAGCGTACCGGTGCAAAAAGACA-AAAAGGCTCGA-3' and 5'-GGCGTCGAGCCTTTTTGTCTTTTTGCACCGGTAC-GCTACGTTC-3', and 5'-CGCCGCATATAGACATATAGGGCCCAGCTCGTCCT-AGCTAGCGTCGTAGCGTCTTAAG-3' and 5'-TCGACTTAAGACGCTACGACG-CTAGCTAGGACGAGCTGGGCCCTATATGTCTATATGC-3'. The resulting oligo-nucleotide duplexes were phosphorylated with ATP and T4 polynucleotide kinase and ligated to the large pUC19 *Ava I*/*Sal* I restriction fragment using T4 DNA ligase. *E. coli* XL-1 Blue competent cells were then transformed with the ligated plasmid, and plasmid DNA from ampicillin-resistant white colonies isolated using a Promega Maxi-Prep kit. Plasmid pJT3 was prepared in a similar manner, except that the following synthetic oligonucleotides were hybridized and cloned into the large *Apa I*/*Sal* I fragment of pJT1: 5'-GTCGACGCTGTTAAACAGGCTCGACGCCAGCTCGTCCTAGCTAGCGTCGTA-GCGTCTTAAGAG-3' and 5'-TCGACTCTTAAGACGCTACGACGCTAGCTAGG-ACGAGCTGGCGTCGAGCCTGTTTAACAGCGTCGACGGCC-3'. The presence of the desired insert was determined by restriction analysis and dideoxy sequencing. Plasmid DNA concentration was determined at 260 nm using the relation 1 OD unit = 50 μg/mL duplex DNA.

Preparation of ^{32}P-End-Labelled Restriction Fragments. Plasmid pJT3 was digested with *Afl* II, labeled at the 3'-end using Sequenase (version 2.0), [α-^{32}P]-

dTTP and [α-^{32}P]-dATP and digested with *Fsp* I. The 281 bp restriction fragment was isolated by nondenaturing gel electrophoresis and used in all quantitative footprinting experiments described here. Chemical sequencing reactions were performed as described.[15,16] Standard techniques were employed for DNA manipulations.[17]

Quantitative DNaseI Footprint Titration. All reactions were executed in a total volume of 40 μL. We note explicitly that no carrier DNA was used in these reactions. A polyamide stock solution or H_2O (for reference lanes) was added to an assay buffer containing radiolabeled restriction fragment (15,000 cpm), affording final solution conditions of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, and either (i) 0.1 nM–1 μM polyamide, for all polyamides except ImPyPy-β-PyPyPy-Dp, (ii) 0.01 nM–0.1 μM polyamide for ImPyPy-β-PyPyPy-Dp, and (iii) no polyamide (for reference lanes). The solutions were allowed to equilibrate for 5 hours at 22 °C. Footprinting reactions were initiated by the addition of 4 μL of a DNase I stock solution (at the appropriate concentration to give ~55% intact DNA) containing 1 mM dithiothreitol and allowed to proceed for 7 minutes at 22 °C. The reactions were stopped by the addition of 10 μL of a solution containing 1.25 M NaCl, 100 mM EDTA, and 0.2 mg/mL glycogen, and ethanol precipitated. The reactions were resuspended in 1X TBE/80% formamide loading buffer, denatured by heating at 85 °C for 10 minutes, and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% cross-link, 7 M urea) in 1X TBE at 2000 V. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics).

Data Analysis. Data from the footprint titration gels were obtained using a Molecular Dynamics 400S PhosphorImager followed by quantitation using ImageQuant software (Molecular Dynamics). Background-corrected volume integration of rectangles encompassing the footprint sites and a reference site at which DNase I reactivity was invariant across the titration generated values for the site intensities (I_{site}) and the reference

intensity (I_{ref}). The apparent fractional occupancy (θ_{app}) of the sites were calculated using the equation:

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

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

The ($[L]_{tot}$, θ_{app}) data points were fit to a general Hill equation (eq 2) by minimizing the difference between θ_{app} and θ_{fit}:

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

where $[L]_{tot}$ is the total polyamide concentration, K_a is the apparent first-order association constant, and θ_{min} and θ_{max} are the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. The data were fit using a nonlinear least-squares fitting procedure with K_a, θ_{max}, and θ_{min} as the adjustable parameters, and with either n=2 or n=1 depending on which value of n gave the better fit. 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 association constant, a parameter that represents the concentration of polyamide at which the binding site is half-saturated. The binding isotherms were normalized using the following equation:

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

Three sets of data were used in determining each association constant.

The method for determining association constants used here involves the assumption that $[L]_{tot} \approx [L]_{free}$, where $[L]_{free}$ is the concentration of polyamide free in solution (unbound). For very high association constants this assumption becomes invalid, resulting in underestimated association constants. In the experiments described here, the DNA concentration is estimated to be ~50 pM. As a consequence, measured association

constants of 1×10^9 M^{-1} and 5×10^9 M^{-1} underestimate the true association constants by factors of approximately ≤ 1.5 and 1.5–2, respectively.

References

(1) (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.; Geirstanger, 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.

(2) (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.

(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.

(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*, 3325–3332.

(6) (a) 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, (c) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 3663–3664. (d) Singh, M.P.; Plouvier, B.; Hill, G. C.; Gueck, J.; Pon, R.T.; Lown, J.W. *J. Am. Chem. Soc.* **1994**, *116*, 2006–2020.

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

(8) Mrksich, M. Ph.D. Thesis, California Institute of Technology, 1994.

(9) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E., *Nature Struct. Biol.* **1996**, *3*, 321–324.

(10) Breslauer, K.J.; Remeta, D.P.; Chou, W.-Y.; Ferrante, R.; Curry, J.; Zaunczkowski, D.; Snyder, J.G.; Marky, L.A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8922–8926, (b) Marky, L.A.; Breslauer, K.J. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4359–4363. (c) Marky, L.A.; Kupke, K.J. *Biochemistry* **1989**, *28*, 9982–9988.

(11) Kelly, J.J.; Baird, E.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6981–6985.

(12) (a) Galas, D.; Schmitz, A. *Nucleic Acids Res.* **1978**, *5*, 3157–3170. (b) Fox, K.R.; Waring, M.J. *Nucleic Acids Res.* **1984**, *12*, 9271–9285.

(13) (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.

(14) (a) Youngquist, R.S.; Dervan, P.B. *J. Am. Chem. Soc.* **1987**, *109*, 7564. (b) Griffin, J.H. Ph.D. Thesis, California Institute of Technology, 1990.

(15) Iverson, B.L.; Dervan, P.B. *Nucleic Acids Res.* **1987**, *15*, 7823–7830.

(16) 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 THREE

C-terminal Amino Acids Modulate the DNA-Binding Specificity of Polyamide Dimers: Optimization of "Slipped" Versus "Overlapped" Binding Modes

Introduction

Small molecules that specifically bind to any predetermined DNA sequence would be useful tools in molecular biology and potentially in human medicine. The three-ring polyamide ImPyPy-Dp was shown to specifically bind 5'-(A,T)G(A,T)C(A,T)-3' sequences as a side-by-side, antiparallel dimer.[1,2] In the $(ImPyPy\text{-}Dp)_2$•5'-TGACT-3' complex, each polyamide makes specific contacts with one strand on the floor of the minor groove such that the sequence-specificity depends on the sequence of side-by-side amino acid pairings.[1] A pairing of imidazole opposite pyrrole recognizes G•C base-pairs, while a pairing of pyrrole opposite imidazole recognizes C•G base-pairs.[1] A pyrrole/pyrrole pairing is partially degenerate and targets both A•T and T•A base pairs.[1,2] The generality of the 2:1 model has been demonstrated by targeting a wide variety of sequences.[3-5]

"Overlapped" and "Slipped" Binding Modes: We recently reported that three ImPyPy-X-PyPyPy-Dp polyamides, where X = G, β, or Py, can bind to the 9 bp site 5'-TGTTAAACA-3' and to the 13 bp site 5'-AAAAAGACAAAAA-3'.[6] The polyamides bind the 9 bp and 13 bp sites in two distinct binding motifs, which we refer to as "fully overlapped" and "slipped," respectively (Fig. 1). In the 9 bp "fully overlapped" binding motif, two ImPyPy-X-PyPyPy polyamides bind directly opposite one another. The 13 bp "slipped" binding mode integrates the 2:1 and 1:1 polyamide-DNA binding motifs at a single site. In this motif, the ImPyPy moieties of two ImPyPy-X-PyPyPy polyamides bind the central 5'-AGACA-3' sequence in a 2:1 manner as in the ImPyPy-Dp dimer, while the PyPyPy moieties bind the all-A,T flanking sequences as in 1:1 complexes of the natural product distamycin. Within the slipped motif, a preference for 5'-AAAAAGACAAAAA-3' over 5'-ATATAGACATATA-3' is observed.[6] The structure of the slipped complex formed by the glycine-linked polyamide ImPyPy-G-PyPyPy-Dp at a 13 bp binding site has been characterized by 2D NMR.[7]

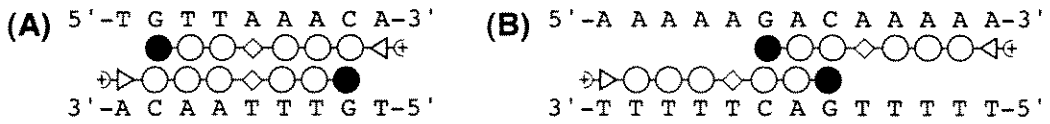


Figure 1. Models of the expected complexes of ImPyPy-β-PyPyPy-G-Dp (**3**) with the sites (A) 5'-TGTTAAACA-3' (9 bp, "overlapped") and (B) 5'-AAAAAGACAAAAA-3' (13 bp, "slipped"). The filled and open circles represent imidazole and pyrrole rings, respectively, diamonds represent β-alanine, and triangles represent glycine.

We report here the synthesis of six ImPyPy-X-PyPyPy-Y-Dp polyamides, where X = G, β, or Py, and Y = G or β (Fig. 2). Equilibrium association constants of these polyamides for the 9 bp site 5'-TGTTAAACA-3' and for the 13 bp sites 5'-AAAAAGACAAAAA-3' and 5'-ATATAGACATATA-3' located on a 281 bp restriction fragment were determined by quantitative DNase I footprint titration experiments. In addition, two EDTA•Fe(II) polyamides, ImPyPy-G-PyPyPy-G-Dp-EDTA•Fe(II) and ImPyPy-β-PyPyPy-G-Bp-EDTA•Fe(II), were synthesized and used in affinity cleavage reactions on both the 281 bp restriction fragment and its 2,799 bp parent plasmid.

Results and Discussion

Synthesis. Polyamides ImPyPy-G-PyPyPy-G-Dp (**1**), ImPyPy-G-PyPyPy-β-Dp (**2**), and ImPyPy-β-PyPyPy-G-Dp (**3**) were prepared by a manual solid phase synthetic methodology and a sample of polyamide cleaved from the resin with dimethylaminopropylamine.[8] ImPyPy-β-PyPyPy-β-Dp (**4**), ImPyPy-Py-PyPyPy-G-Dp (**5**) and ImPyPy-Py-PyPyPy-β-Dp (**6**) were prepared by a machine-assisted solid phase methodology and a sample of polyamide cleaved from the resin with dimethylaminopropylamine.[8] ImPyPy-G-PyPyPy-G-Dp-NH_2 (**1-NH_2**) and ImPyPy-β-PyPyPy-G-Dp-NH_2 (**3-NH_2**) were prepared from the same manual synthesis as **1** and **3**, respectively, by cleavage of a resin sample with 3,3′-diamino-*N*-methyldipropylamine, providing polyamides functionalized with a free primary amino group. Polyamides **1-E** and **3-E** were obtained by treatment of **1-NH_2** and **3-NH_2**, respectively, with EDTA

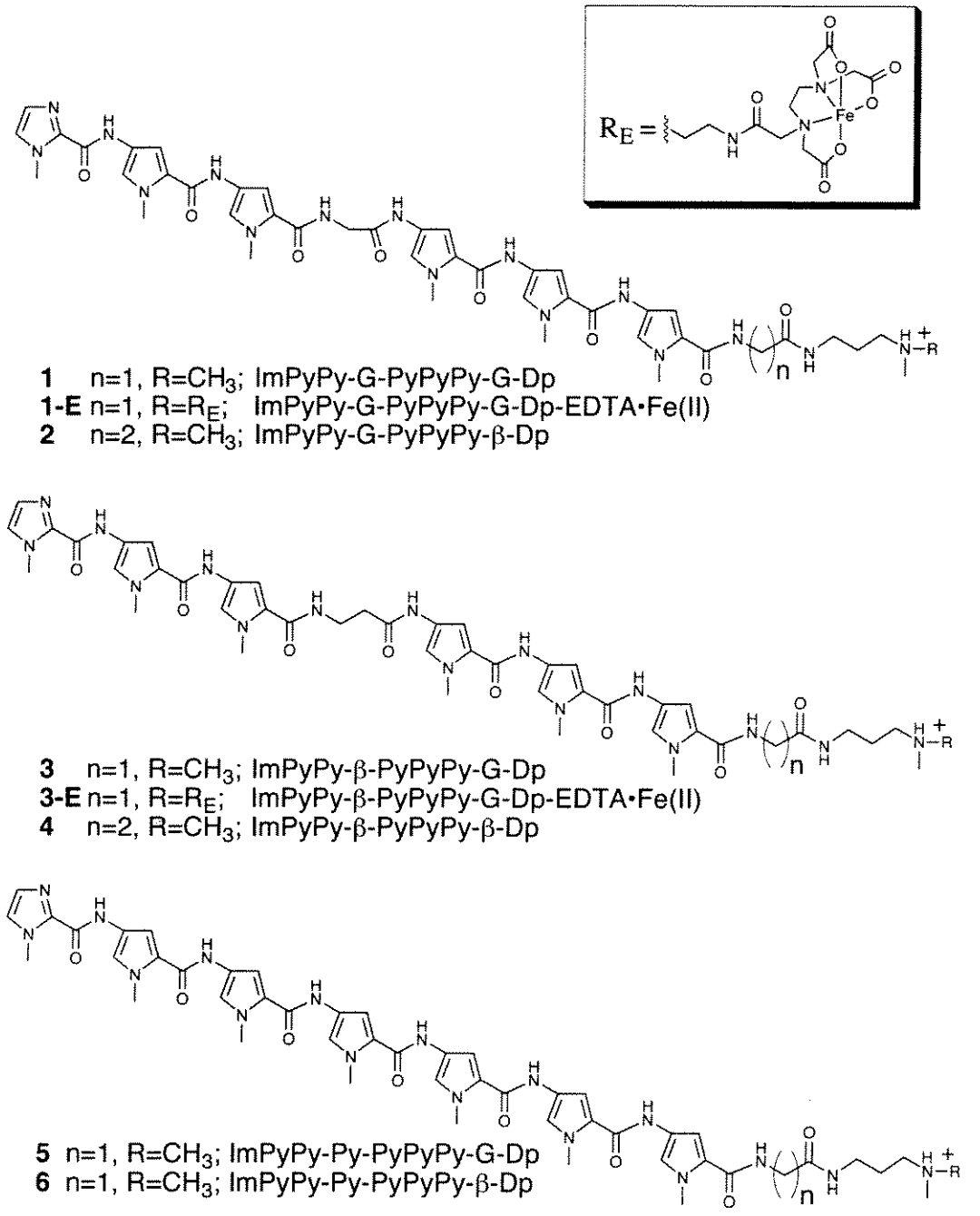


Figure 2. Structures of the polyamides used in this study.

dianhydride (DMSO/NMP, DIEA) followed by sodium hydroxide (Fig. 3). All compounds were purified by reverse phase HPLC as the trifluoroacetate salt and

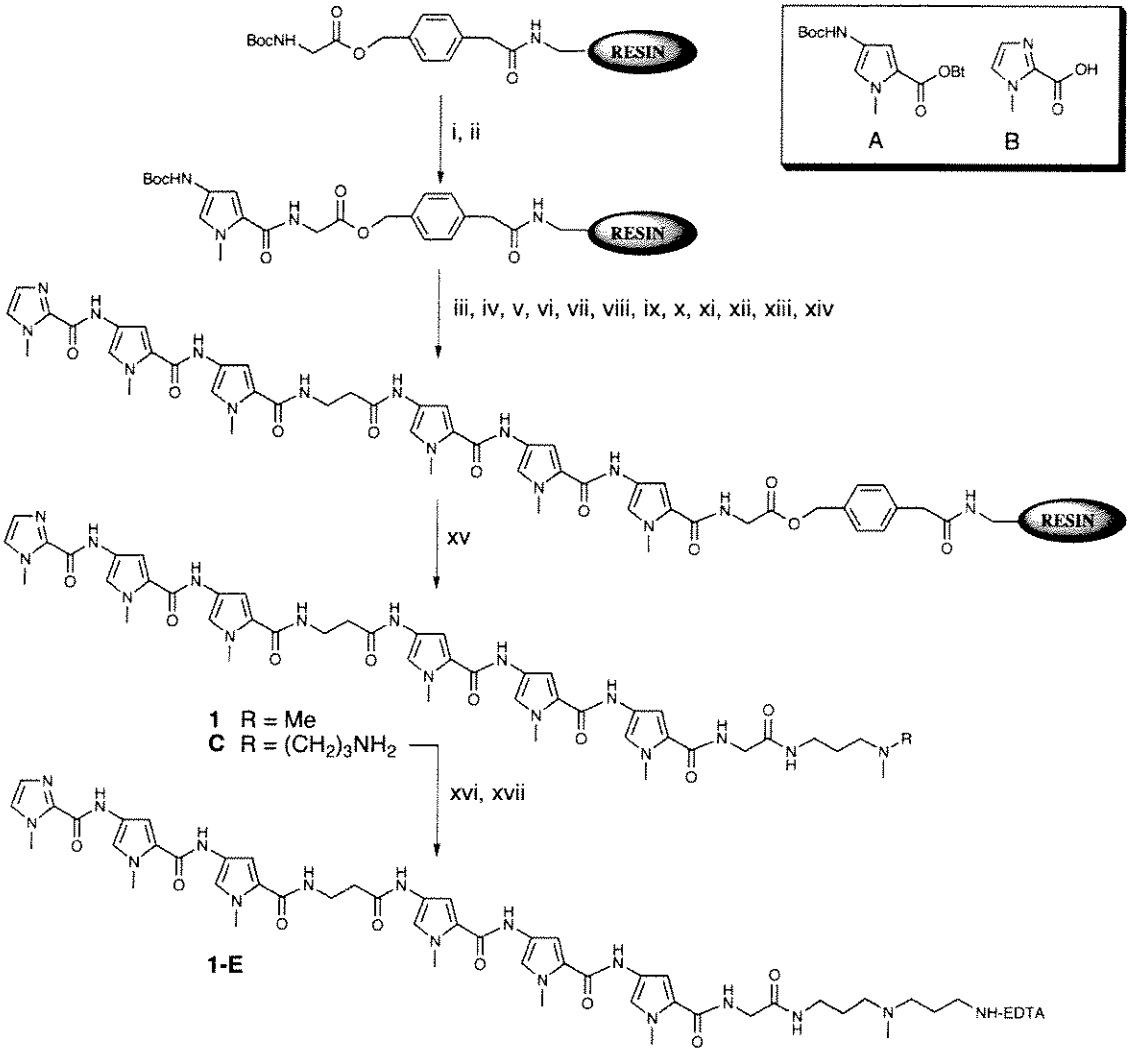


Figure 3. Solid phase polyamide synthesis: synthesis of **1** and **1-E**. (Box) Pyrrole and imidazole cap monomers; Boc-Pyrrole-OBt ester **A** and Imidazole-acid **B**. The polyamides were prepared from commercially available Boc-glycine-Pam-resin (0.2 mmol/gram): (i) 80% TFA/DCM, 0.4M PhSH; (ii) BocPy-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4M PhSH; (iv) BocPy-OBt, DIEA, DMF; (v) 80% TFA/DCM, 0.4M PhSH; (vi) BocPy-OBt, DIEA, DMF; (vii) 80% TFA/DCM, 0.4M PhSH; (viii) Boc-β-alanine (HBTU, DIEA), DMF; (ix) 80% TFA/DCM, 0.4M PhSH; (x) BocPy-OBt, DIEA, DMF; (xi) 80% TFA/DCM, 0.4M PhSH; (xii) BocPy-OBt, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4M PhSH; (xiv) imidazole-2-carboxylic acid (HBTU/DIEA); (xv) *N,N*-dimethylamino-propylamine or 3,3′-diamino-*N*-methyldipropylamine, 55 °C; (xvi) EDTA-dianhydride, DMSO/NMP, DIEA, 55 °C; (xvii) 0.1M NaOH.

characterized by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectroscopy, ^{1}H NMR and analytical HPLC.

DNA-Binding Specificity: Effect of C-Terminal Glycine and β-Alanine Residues. Quantitative DNase I footprinting experiments[9,10] on the 3'-^{32}P end-labeled 281 bp pJT3 *Afl* II/*Fsp* I restriction fragment (Fig. 4, Tables 1–3) reveal that all three ImPyPy-X-PyPyPy-β-Dp (X = G, β, and Py) polyamides having a C-terminal β-alanine residue have similar binding affinities (within a factor of 4) to the analogous parent compounds ImPyPy-X-PyPyPy-Dp at both the 9 bp site 5'-TGTTAAACA-3' and the 13 bp site 5'-AAAAAGACAAAAA-3'. For polyamides ImPyPy-G-PyPyPy-**G**-Dp and ImPyPy-β-PyPyPy-**G**-Dp having a C-terminal glycine residue, association constants for the 9 bp site 5'-TGTTAAACA-3' are decreased by about 80-fold and 10-fold, respectively, relative to the parent compounds, while association constants for the 13 bp site 5'-AAAAAGACAAAAA-3' remain similar (within a factor of 1.5) to the parent compounds. In these cases, *C-terminal glycine confers specificity for the 13 bp slipped complexes relative to the 9 bp fully overlapped complexes by selectively disrupting the fully overlapped complex*. For ImPyPy-Py-PyPyPy-**G**-Dp having a C-terminal glycine residue, association constants for both the 9 bp 5'-TGTTAAACA-3' and 13 bp 5'-AAAAAGACAAAAA-3' sites decrease, by factors of about 60 and 8, respectively.

Comparison of the binding affinities of polyamides ImPyPy-X-PyPyPy-**Y**-Dp with the parent compounds ImPyPy-X-PyPyPy-Dp indicates that addition of a C-terminal β-alanine residue has a relatively small effect on the stabilities of DNA-polyamide complexes, while addition of a C-terminal glycine residue can dramatically affect complex stabilities. Modeling indicates that ImPyPy-G-PyPyPy-Dp and ImPyPy-G-PyPyPy-β-Dp have similar DNA-binding surfaces at their C-terminal ends, in contrast to ImPyPy-G-PyPyPy-**G**-Dp (Fig. 5). The marked destabilization of some polyamide-DNA complexes by C-terminal glycine may result from (1) a steric interaction or hydrogen bond mismatch between the glycine carbonyl group and the floor of the minor groove and/or (2) from loss of a ligand-

Figure 4. Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the products of DNase I digestion in quantitative footprint titration experiments with polyamides: lanes 1–2, A and G sequencing lanes; lanes 3 and 21, DNase I digestion products obtained in the absence of polyamide; lanes 4–20, DNase I digestion products obtained in the presence of 0.1 nM (0.01 nM), 0.2 nM (0.02 nM), 0.5 nM (0.05 nM), 1 nM (0.1 nM), 1.5 nM (0.15 nM), 2.5 nM (0.25 nM), 4 nM (0.4 nM), 6.5 nM (0.65 nM), 10 nM (1 nM), 15 nM (1.5 nM), 25 nM (2.5 nM), 40 nM (4 nM), 65 nM (6.5 nM), 100 nM (10 nM), 200 nM (20 nM), 500 nM (50 nM), and 1 μM (100 nM) polyamide, respectively (concentrations used for ImPyPy-β-PyPyPy-G-Dp and ImPyPy-β-PyPyPy-β-Dp only are in parentheses); lane 22, intact DNA. The binding sites that were analyzed are indicated alongside the autoradiograms. All reactions contain 15 kcpm restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

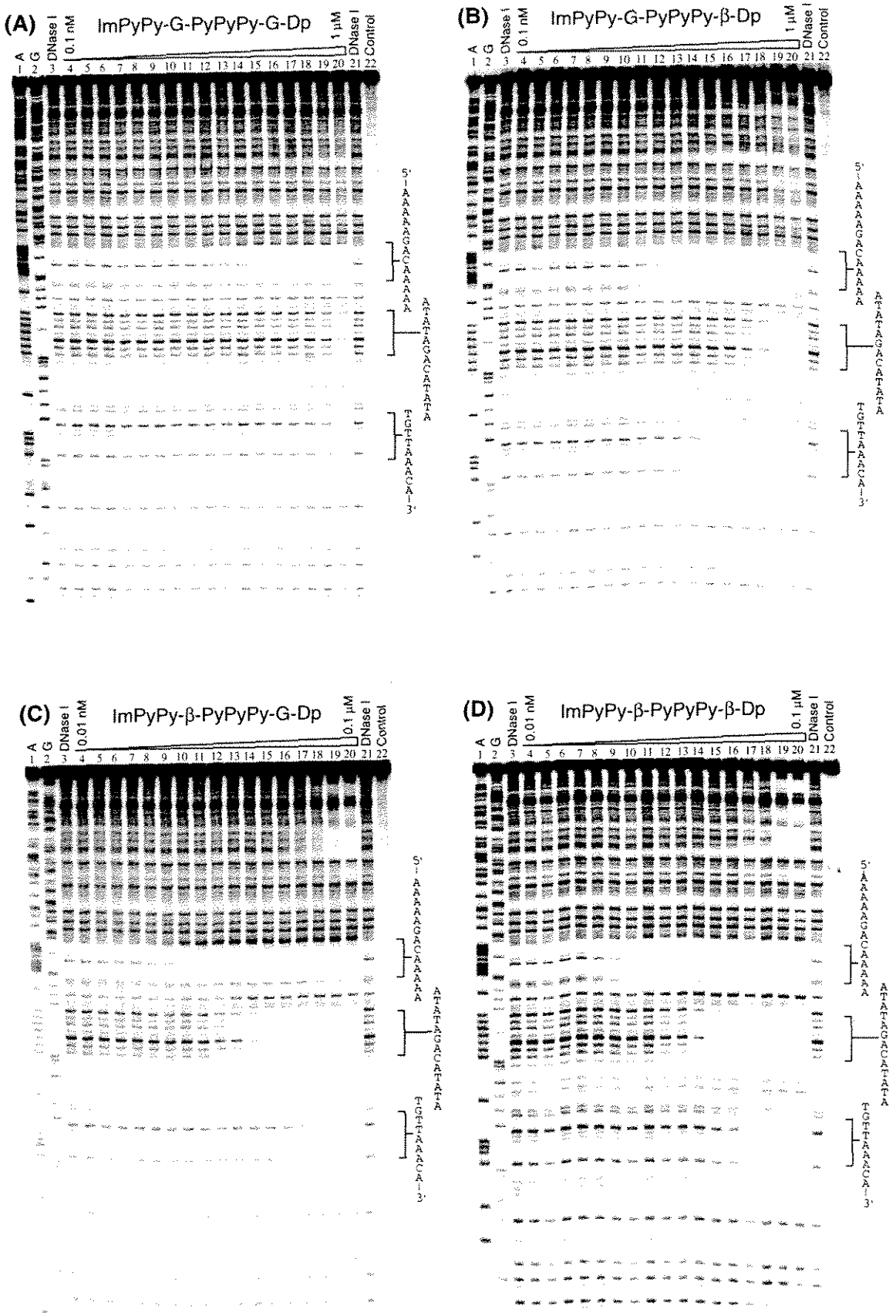

(A)
ImPyPy-G-PyPyPy-G-Dp
(B)
ImPyPy-G-PyPyPy-β-Dp
(C)
ImPyPy-β-PyPyPy-G-Dp
(D)
ImPyPy-β-PyPyPy-β-Dp
A
G
DNase I
0.1 nM
1 μM
0.01 nM
0.1 μM
Control
5'-AAAAAGACAAAAA
ATATAGACATATA
TGTTAAACA-3'

Table 1. Equilibrium association constants (M^{-1}) for polyamides having a central glycine (G) residue.[a,b,c,d]

Binding Site	ImPyPy-G-PyPyPy-Dp[c]	ImPyPy-G-PyPyPy-G-Dp	ImPyPy-G-PyPyPy-β-Dp
5'-AAAAAGACAAAAA-3'	**1.1 (±0.1) x 10^8**	**7.0 (±1.2) x 10^7**	**1.0 (±0.2) x 10^8**
5'-ATATAGACATATA-3'	6.6 (±0.4) x 10^6 [d]	≈ 3.5 (±1.4) x 10^6 [d]	1.0 (±0.2) x 10^7 [d]
5'-TGTTAAACA-3'	**1.4 (±0.1) x 10^8**	≈ 1.7 (±0.7) x 10^6 [d]	**3.4 (±0.5) x 10^7**

Table 2. Equilibrium association constants (M^{-1}) for polyamides having a central β-alanine (β) residue.[a,b,c,d]

Binding Site	ImPyPy-β-PyPyPy-Dp[c]	ImPyPy-β-PyPyPy-G-Dp	ImPyPy-β-PyPyPy-β-Dp
5'-AAAAAGACAAAAA-3'	**≥ 4.7 (±0.7) x 10^9**	**≥ 2.7 (±0.1) x 10^9**	**≥ 4.5 (±0.6) x 10^9**
5'-ATATAGACATATA-3'	**1.0 (±0.1) x 10^9**	6.2 (±0.5) x 10^8	**8.6 (±0.3) x 10^8**
5'-TGTTAAACA-3'	**7.8 (±0.6) x 10^8**	7.4 (±0.9) x 10^7	**2.6 (±0.3) x 10^8** [d]

Table 3. Equilibrium association constants (M^{-1}) for polyamides having a central N-methylpyrrole (Py) residue.[a,b,c,d]

Binding Site	ImPyPyPyPyPyPy-Dp[c]	ImPyPyPyPyPyPy-G-Dp	ImPyPyPyPyPyPy-β-Dp
5'-AAAAAGACAAAAA-3'	**5.4 (±1.5) x 10^7**	≈ 6.8 (±1.6) x 10^6 [d]	**4.9 (±0.2) x 10^7**
5'-ATATAGACATATA-3'	**3.6 (±0.5) x 10^7**	< 10^6	**3.5 (±0.3) x 10^7**
5'-TGTTAAACA-3'	**9.7 (±2.3) x 10^7**	≈ 1.6 (±1.2) x 10^6 [d]	**9.8 (±0.7) x 10^7**

[a]Values reported are the mean values obtained from three DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses. Assays were carried out at 22 °C and pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [b]Association constants >2 x 10^7 M^{-1} for glycine- and N-methylpyrrole-linked polyamides, and >2 x 10^8 M^{-1} for β-alanine-linked polyamides are in bold type. [c]Association constant from reference 6. [d]A binding isotherm shallower than expected for highly cooperative dimeric binding is observed for this complex.

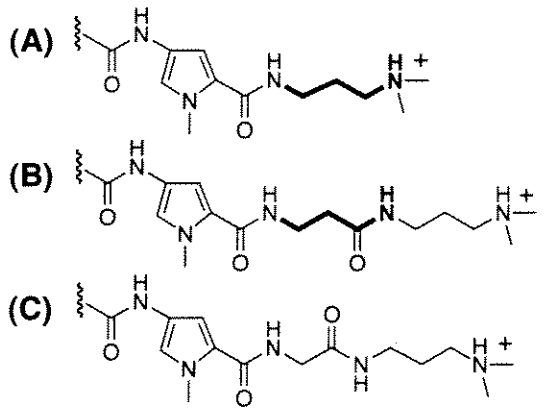


Figure 5. Structures of the C-termini of polyamides having the C-terminal sequence (A) Py-Dp, (B) Py-β-Dp and (C) Py-**G**-Dp.

DNA hydrogen bond. The slipped binding mode, in which the C-terminal part of the molecule is bound in a 1:1 manner (see Fig. 1A), appears to tolerate a C-terminal glycine in the cases of ImPyPy-G-PyPyPy-**G**-Dp and ImPyPy-β-PyPyPy-**G**-Dp, although the basis for this selectivity is unclear.

Sequence-Specific DNA Cleavage. The footprinting experiments described above show that, of the nine ImPyPy-X-PyPyPy-Y-Dp polyamides, ImPyPy-G-PyPyPy-G-Dp (**1**) and ImPyPy-β-PyPyPy-G-Dp (**3**) have the highest overall sequence-specificity, preferring 5'-AAAAAGACAAAAA-3' (slipped complex) over 5'-TGTTAACA-3' (overlapped complex) by ~40-fold and ≥35-fold, respectively. Of these, polyamide **3** has the highest binding affinity. Affinity cleavage[1a,11] experiments using EDTA•Fe(II)-polyamide **3-E** on the 5'- and 3'-^{32}P end-labeled 281 bp pJT3 *Afl* II/*Fsp* I restriction fragment show that this polyamide has cleavage specificity similar to the binding specificity of **3** (Fig. 6). At the lowest concentrations at which cleavage is observed, this compound is specific for 5'-AAAAAGACAAAAA-3' over 5'-ATATAGACATATA-3'.[6] The symmetric cleavage patterns observed for this compound are consistent with the expected side-by-side complexes.

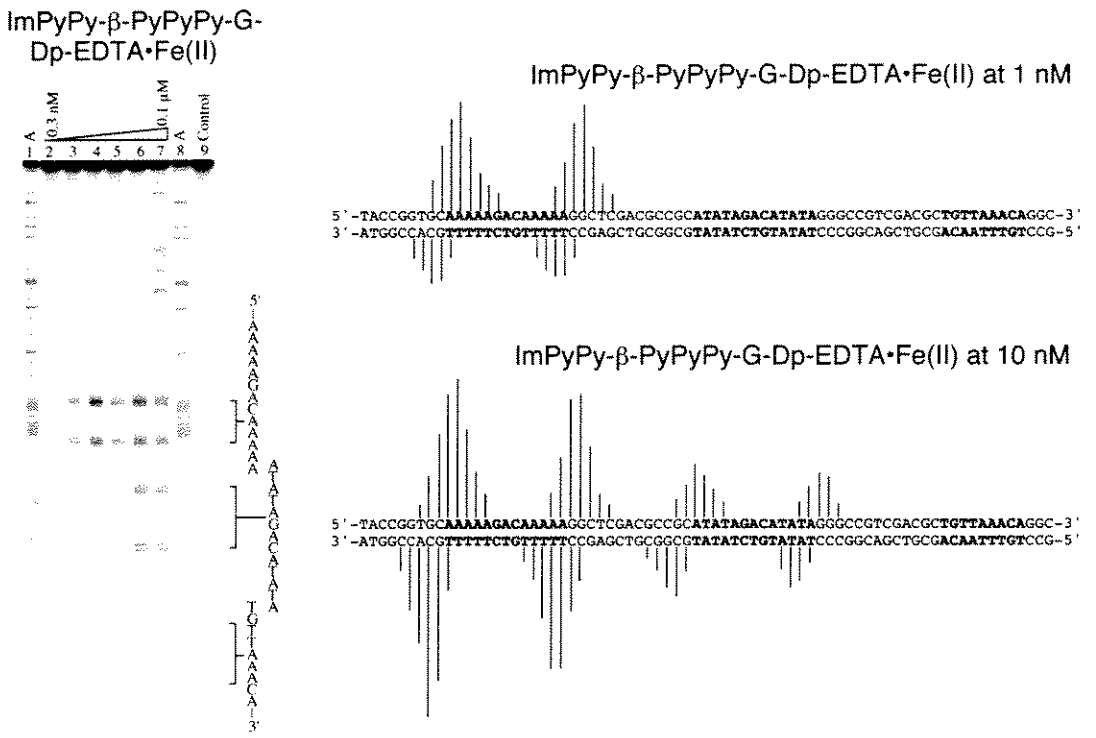


Figure 6. Affinity cleavage experiment using ImPyPy-β-PyPyPy-G-Dp-EDTA•Fe(II) on the 281 bp pJT3 *Afl* II/*Fsp* I restriction fragment. (Left) Storage phosphor autoradiogram of the 8% denaturing polyacrylamide gel used to separate cleavage products of the 3'-end-labeled restriction fragment. Lanes 1 and 8, A sequencing lanes; lanes 2–7, cleavage products obtained in the presence of 0.3 nM, 1 nM, 3 nM, 10 nM, 30 nM and 100 nM polyamide, respectively; lane 9, intact DNA. The designated binding sites are indicated on the right side of the autoradiogram. All reactions contain 15 kcpm restriction fragment, 20 mM HEPES, 200 mM NaCl, 50 μg/mL glycogen, 1 μM Fe(II) and 5 mM DTT. (Right) Line heigths are proportional to the amount of cleavage at the indicated base.

Double-strand cleavage experiments were carried out with **3-E** on the 2,799 bp plasmid pJT3 (Fig. 7). The linearized plasmid DNA was labeled selectively at one end (designated A) with [α-^{32}P]-dATP or at the other end (designated C) with [α-^{32}P]-dCTP using Sequenase. Polyamide **3-E** at 1–3 nM concentration effects double-strand cleavage of the linearized plasmid predominately at a single site, producing a 1.4 kb fragment consistent with cleavage at the 5'-AAAAAGACAAAAA-3' sequence. Cleavage sites were mapped to within ±40 bp by comparison of cleavage bands with molecular weight markers. Plasmid pJT3 is derived from pUC19 DNA by addition of a 100 bp insert containing the

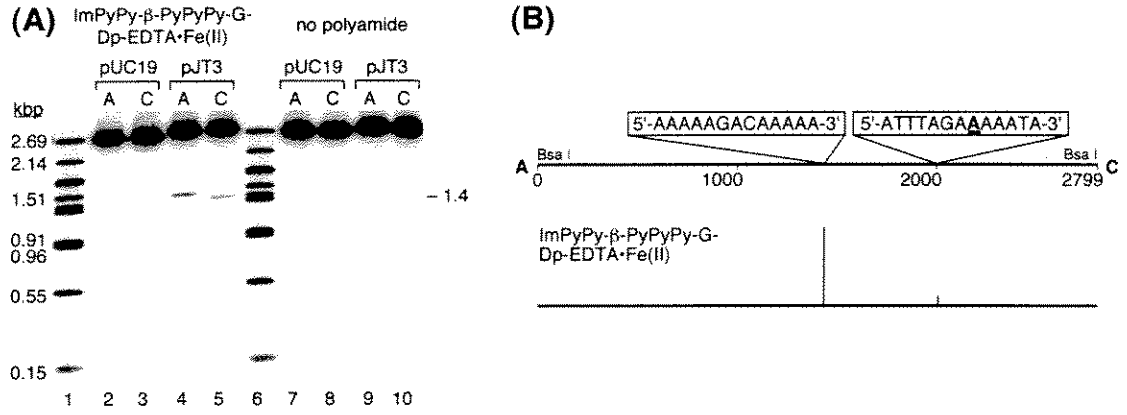


Figure 7. Double-strand cleavage of *Bsa* I-linearized plasmid pJT3 (2799 bp) by polyamide ImPyPy-β-PyPyPy-G-Dp-EDTA•Fe(II) (**3-E**). (A) Autoradiogram of a 1% agarose gel used to separate cleavage products. The plasmid was labeled at one end (designated A) by 3'-fill-in with [α-^{32}P]-dATP, and at the other end (designated C) by 3'-fill-in with [α-^{32}P]-dCTP. Lanes 1 and 6, molecular weight markers; lanes 2–5, 3 nM **3-E**; lanes 7–10, no polyamide. The plasmid (pUC19 or PJT3) and the labeled end (A or C) are indicated. All reactions contain 5 kcpm DNA, 20 mM HEPES, 200 mM NaCl, 50 μg/mL glycogen, 1 μM Fe(II) and 5 mM DTT. (B) Map of cleavage sites for **3-E** at 3 nM on the *Bsa* I-linearized plasmid pJT3. Line heigths are proportional to the amount of cleavage at the indicated position.

5'-AAAAAGACAAAAA-3' site. When pUC19 is substituted for pJT3 in the cleavage experiment, the 1.4 kb band is not observed, confirming that this fragment results from cleavage at 5'-AAAAAGACAAAAA-3' (significant cleavage at the adjacent 5'-ATATAGACATATA-3' and 5'-TGTTAAACA-3' sites does not occur at 1–3 nM concentrations, see Fig. 6). A second site is cleaved by **3-E** with ~20-fold lower efficiency than 5'-AAAAAGACAAAAA-3', producing a 2.0 kb fragment. Inspection of the plasmid sequence suggests that this fragment results from cleavage at the sequence 5'-ATTTAGAAAAATT-3', a single base pair mismatch site. The plasmid cleavage experiment was also carried out with polyamide **1-E**. Cleavage by **1-E** is observed at 100–300 nM concentrations predominately at the two sites described above, and with lower efficiency at several additional sites (data not shown).

Implications for the Design of Minor-Groove Binding Molecules. In the system described here, addition of a C-terminal glycine residue to polyamides ImPyPy-G-

PyPyPy-Dp and ImPyPy-β-PyPyPy-Dp markedly increases specificity for 13 bp "slipped" complexes versus a 9 bp "overlapped" complex by selectively disrupting the 9 bp complex. Addition of a C-terminal β-alanine residue has a smaller effect on DNA-binding affinity.

Solid-phase polyamide synthetic protocols have recently been reported that allow rapid polyamide preparation.[8] These protocols require the use of an aliphatic amino acid resin linkage. Cleavage of this resin linkage by aminolysis results in a polyamide having a C-terminal aliphatic amino acid. In this context, the finding here that a C-terminal β-alanine has only a small effect on DNA-binding affinities for a fully overlapped complex (in contrast to glycine) suggests that β-alanine is a suitable linker for the convenient solid-phase synthesis of polyamides designed to bind as fully overlapped dimers.

The findings reported here are consistent with another recent study in which the effects of added C-terminal glycine and β-alanine residues on the DNA-binding properties of the hairpin polyamide ImPyPy-γ-PyPyPy-Dp (γ = γ-aminobutyric acid) were studied.[12] In this system, addition of C-terminal β-alanine actually increased both DNA-binding affinity and sequence-specificity, while addition of C-terminal glycine reduced affinity and specificity. Both this study involving "hairpin" polyamide-DNA complexes and the study described here involving "extended dimer" polyamide-DNA complexes reveal that C-terminal aliphatic amino acids can be used to optimize the specificity of pyrrole-imidazole polyamides.

The observation here that β-alanine can in some cases decrease polyamide binding affinity has been clarified by a subsequent study. Using a six-ring hairpin polyamide and the binding sites TGGCTX (where X = either A, G, C or T), Swalley *et al.* showed that a C-terminal β-alanine residue is 20–30-fold selective at position "X" for A,T versus G,C base pairs.[16] The binding sites used in the experiments described in this chapter are flanked by G,C base pairs.

Experimental Procedure

General Methods. NMR were recorded on a GE 300 instrument operating at 300MHz (^{1}H) and 75 MHz (^{13}C). Spectra were recorded in DMSO-d_6 with chemical shifts reported in parts per million relative to residual DMSO-d_5. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. High resolution FAB mass spectra were recorded at the Mass Spectroscopy Laboratory at the University of California, Riverside. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed either on a HP 1090 M analytical HPLC or a Beckman Gold system using a Rainen C18, Microsorb MV, 5 μm, 300 × 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. Preparative HPLC was carried out on a Beckman instrument using a Waters DeltaPak 25 × 100 mm 100 μm C_{18} column in 0.1% (wt/v) TFA, gradient elution 0.25%/min. CH_3CN. Water was obtained from a Millipore Milli-Q water purification system. Reagent-grade chemicals were used unless otherwise stated. Restriction endonucleases were purchased from either New England Biolabs or Boeringher-Mannheim and used according to the manufacturer's protocol. *E. coli* XL-1 Blue competent cells were obtained from Stratagene. Sequenase (version 2.0) was obtained from United States Biochemicals, and DNase I was obtained from Pharmacia. [α-^{32}P]-Thymidine-5'-triphosphate (≥ 3000 C_i/mmol) and [α-^{32}P]-deoxyadenosine-5'-triphosphate (≥ 6000 C_i/mmol) were purchased from Du Pont NEN.

Synthesis. Polyamides were prepared using solid phase methods.[8]

Quantitative DNase I Footprint Titration. Quantitative DNase I footprint titration experiments were carried out as described previously.[6] Plasmid pJT3 and its ^{32}P end-labeled *Afl* II/*Fsp* I restriction fragment were prepared as described.[6] Sequencing reactions were performed as described.[13,14] Standard techniques were employed for DNA manipulations.[15] The method for determining association constants used here involves the

assumption that $[L]_{tot} \approx [L]_{free}$, where $[L]_{free}$ is the concentration of polyamide free in solution (unbound). For very high association constants this assumption becomes invalid, resulting in underestimated association constants. In the experiments described here, the DNA concentration is estimated to be ~50 pM. As a consequence, measured association constants of 1×10^9 M^{-1} and 5×10^9 M^{-1} underestimate the true association constants by factors of approximately ≤ 1.5 and 1.5-2, respectively. In cases where the binding affinity of a polyamide ImPyPy-X-PyPyPy-Y-Dp is decreased relative to the analogous polyamide ImPyPy-X-PyPyPy-Dp, a binding isotherm shallower than expected for a highly cooperative 2:1 complex is often observed (Fig. 5, Tables 1-3), presumably because the C-terminal substituent Y decreases the cooperativity of complexation.

Affinity Cleavage. All reactions were executed in a total volume of 40 μL. A stock solution of the EDTA-polyamide or H_2O (for control lanes) was added to a solution containing radiolabeled restriction fragment (15,000 cpm), affording final solution conditions of 20 mM HEPES, 200 mM NaCl, 50 μg/mL glycogen (Boeringer Mannheim), pH 7.3, and either 1 nM, 10 nM, and 100 nM ImPyPy-β-PyPyPy-G-Dp-EDTA or no polyamide (for control lanes). Subsequently, 2 μL of freshly prepared 20 μM $Fe(NH_4)_2(SO_4)_2$ was added and the solution allowed to equilibrate for 20 min. Cleavage reactions were initiated by the addition of 4 μL of 50 mM dithiothreitol, allowed to proceed for 12 min at 22 °C, then stopped by the addition of 130 μL of ethanol. Next, 10 μL of a solution containing calf thymus DNA (125 μM base-pair) (Pharmacia) and glycogen (0.6 mg/mL) was added, and the DNA precipitated. The reactions were resuspended in 1x TBE/80% formamide loading buffer, denatured by heating at 85 °C for 10 min, and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% cross-link, 7 M urea) in 1x TBE at 2000 V. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics). Relative cleavage intensities were determined by volume integration of individual cleavage bands using ImageQuant software (Molecular Dynamics).

Double-Strand Plasmid Cleavage. Plasmid pJT3 (a pUC19 derivative) was linearized with *Bsa* I and labeled selectively at one end (designated A) using [α-^{32}P] dATP and Sequenase (United States Biochemical) or at the other end (designated C) using [α-^{32}P] dCTP and Sequenase. The labeled fragment was purified by phenol/chloroform extraction and NICK column (Pharmacia), then ethanol precipitated. Molecular weight markers were prepared by digesting the labeled DNA with *Hin*dIII, *Alw*NI *Afl*III, *Eco*0109I, or *Fsp*I. Cleavage reactions were executed in a total volume of 400 μL. A stock solution of the EDTA-polyamide or H_2O (for control lanes) was added to a solution containing radiolabeled restriction fragment (15,000 cpm), affording final solution conditions of 20 mM HEPES, 200 mM NaCl, 50 μg/mL glycogen (Boeringer Mannheim), pH 7.3, and either 1–3 nM ImPyPy-β-PyPyPy-G-Dp-EDTA or no polyamide (for control lanes). 40 μL of freshly prepared 10 μM $Fe(NH_4)_2(SO_4)_2$ was added and the solution allowed to equilibrate for 1 h at 37 °C. Cleavage reactions were initiated by the addition of 40 μL of 50 mM dithiothreitol, allowed to proceed for 3 h at 37 °C, then stopped by the addition of 1 mL ethanol, and the DNA precipitated. The reactions were resuspended in 1.5X Ficoll/TE. The reaction products were separated by electrophoresis on a 1% agarose gel in 1X TAE. Gels were dried onto a nitrocellulose membrane and exposed to a storage phosphor screen. Relative cleavage intensities were determined by volume integration of individual cleavage bands using ImageQuant software.

References

(1) 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.

(2) 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.

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

(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*, 3325-3332; (c) Parks, M.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6153-6159; (d) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, in press.

(6) Trauger, J.W.; Baird, E.E.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6160-6166.

(7) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Nature Struct. Biol.* **1996**, *3*, 321-324.

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

(9) a) Galas, D.; Schmitz, A. *Nucleic Acids Res.* **1978**, *5*, 3157-3170; b) Fox, K.R.; Waring, M.J. *Nucleic 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) Schultz., P.G.; Taylor, J.S.; Dervan, P.B. *J. Am. Chem. Soc.* **1982**, *104*, 6861-6863.

(12) Parks, M.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.*, **1996**, *118*, 6147-6152.

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

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

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

(16) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1998**, submitted.

CHAPTER FOUR

Eight-Ring Hairpin Motif: Recognition of DNA by Synthetic Ligands at Subnanomolar Concentrations

The text of this chapter is taken in part from a published paper that was coauthored with

Prof. Peter B. Dervan and Eldon E. Baird

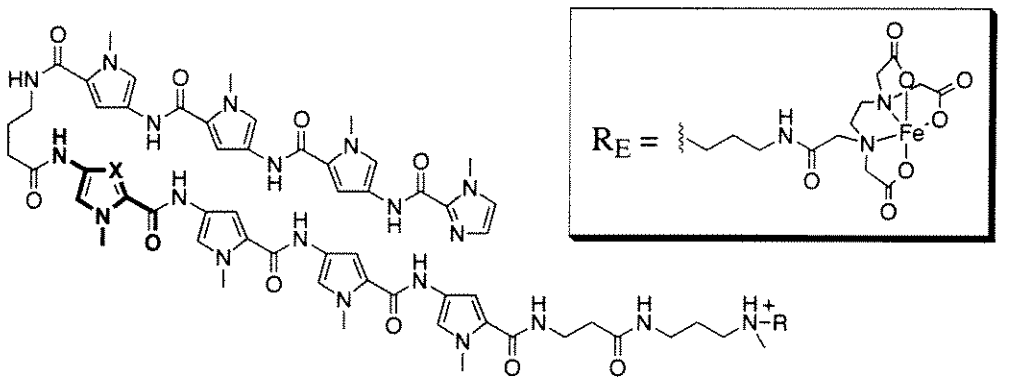
(Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1996**, *3*, 369–377)

Introduction

Small molecules that specifically bind with high affinity to any predetermined DNA sequence in the human genome would be useful tools in molecular biology and potentially in human medicine. Simple rules have been developed to rationally control the sequence-specificity of minor groove binding polyamides containing *N*-methylimidazole and *N*-methylpyrrole amino acids. We report here that two eight-ring pyrrole-imidazole polyamides differing in sequence by a single amino acid bind specifically to respective six base pair target sites which differ in sequence by a single base pair. Binding is observed at subnanomolar concentrations of ligand. The replacement of a single nitrogen atom with a C-H regulates affinity and specificity by two orders of magnitude. The broad range of sequences that can be specifically targeted with pyrrole-imidazole polyamides, coupled with an efficient solid-phase synthesis methodology, identify a powerful class of small molecules for sequence-specific recognition of double-helical DNA.

For side-by-side complexes of pyrrole-imidazole polyamides in the minor groove of DNA, the DNA-binding sequence-specificity depends on the sequence of side-by-side amino acid pairings.[1-3] A pairing of imidazole (Im) opposite pyrrole (Py) targets a G•C base-pair, while pyrrole opposite imidazole targets a C•G base-pair.[1-3] A pyrrole/pyrrole combination is degenerate and targets both T•A and A•T base-pairs.[1-5] Specificity for G,C base-pairs results from the formation of a hydrogen bond between the imidazole N3 and the exocyclic amino group of guanine.[1-3] The generality of these pairing rules has been demonstrated by targeting a wide variety of sequences and is supported directly by several NMR structure studies.[1-10]

In parallel with the elucidation of the scope and limitations of the pairing rules described above have been efforts to increase the DNA-binding affinity and specificity of pyrrole-imidazole polyamides by covalently linking polyamide subunits.[11-16] The polyamide ImPyPy-γ-PyPyPy-Dp containing a "turn" amino acid γ-aminobutyric acid (γ) was shown to specifically bind the designated target site 5'-TGTTA-3' in a "hairpin"



1 X=N, R=CH_3; ImPyPyPy-γ-ImPyPyPyPy-β-Dp
1-E X=N, R=R_E; ImPyPyPy-γ-ImPyPyPyPy-β-Dp-EDTA•Fe(II)
2 X=CH, R=CH_3; ImPyPyPy-γ-PyPyPyPy-β-Dp
2-E X=CH, R=R_E; ImPyPyPy-γ-ImPyPyPyPy-β-Dp-EDTA•Fe(II)

Figure 1. Structures of polyamides **1** and **2**, and their respective affinity cleavage analogs **1-E** and **2-E**. The identity and purity of polyamides **1** and **2** was verified by ^{1}H NMR, MALDI-TOF MS and analytical HPLC. MALDI-TOF MS: **1**, 1223.4 (1223.3 calculated for M+H); **2**, 1222.3 (1222.3 calculated for M+H).

conformation with an equilibrium association constant, K_a, of 8×10^7 M^{-1}, an increase of 300-fold relative to unlinked three-ring polyamide dimers. A key issue was to determine if low molecular weight (M.W. ≈ 1200) pyrrole-imidazole polyamides could be constructed which would bind DNA at subnanomolar concentration without compromising sequence selectivity.

Eight-Ring Hairpin Motif. We report the DNA-binding affinities of two eight-ring hairpin polyamides, ImPyPyPy-γ-ImPyPyPyPy-β-Dp (**1**) and ImPyPyPy-γ-PyPyPyPy-β-Dp (**2**), which differ by a single amino acid, for two 6 base pair (bp) target sites, 5'-AGTACT-3' and 5'-AGTATT-3', which differ by a single base pair. Based on the pairing rules for polyamide-DNA complexes, the sites 5'-AGTACA-3' and 5'-AGTATT-3' are for polyamide **1** "match" and "single base pair mismatch" sites, respectively, and for polyamide **2** "single base pair mismatch" and "match" sites,

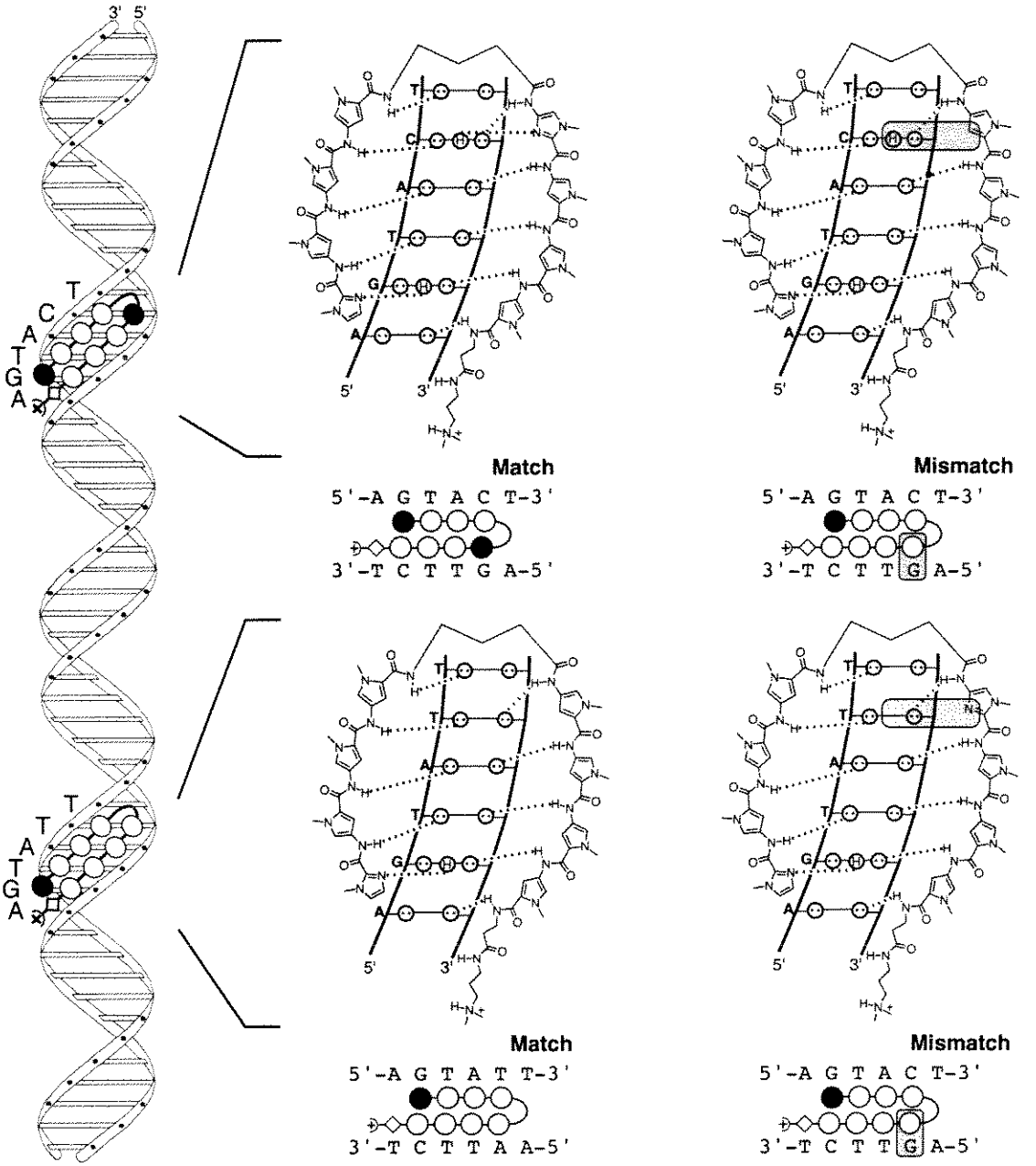


Figure 2. Binding models for (top) 5'-AGTACT-3' in complex with polyamides **1** (match) and **2** (mismatch), and (bottom) 5'-AGTATT-3' in complex with polyamides **2** (match) and **1** (mismatch). Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines, and circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dashed lines. The filled and unfilled circles represent imidazole and pyrrole rings, respectively, curved lines represent γ-aminobutyric acid, and diamonds represent β-alanine. Single hydrogen bond mismatches are highlighted with shading.

respectively (Figs. 1 and 2). Additional polyamide **1** and **2** binding sites present on the restriction fragment used in this study were also characterized.

Results and Discussion

Binding Sites and Orientations. Polyamides **1** and **2** were synthesized by solid phase methods and purified by reversed phase HPLC.[17] Binding sites for **1** and **2** on a 229 bp restriction fragment containing the 5'-AGTACT-3' and 5'-AGTATT-3' sites were determined to nucleotide resolution by MPE•Fe(II) footprinting experiments (Figs. 3 and 4). Polyamide **1**, at a concentration of 10 nM, binds its match site 5'-tAGTACTt-3', and the single-base pair mismatch sites 5'-tAGTA**T**Tt-3', 5'-tTGTA**A**Aa-3', and 5'-aTGT**G**CTg-3'. The "reverse orientation" mismatch site 5'-aT**T**AAGTt-3' is bound with lower apparent affinty (in the reverse orientation,[18] the polyamide N to C vector is aligned opposite the DNA 5' to 3' vector). Affinity cleavage with **1-E** at 10 nM reveals the expected, roughly symmetric cleavage patterns at all of the binding sites except 5'-aTGTGCTg-3', for which cleavage is observed only on one side of the binding site. The distinct orientation in this case suggests that the C-terminal β-alanine residue preferentially binds opposite T versus G at the base pair immediately flanking the six base pair binding site (the three other binding sites at which cleavage is observed are flanked on both sides by A,T base pairs).[19] Polyamide **2**, at a concentration of 10 nM, binds the match sites 5'-tAGTATTt-3', 5'-tTGTAAAa-3' and 5'-gTGAATTc-3', and also binds the reverse orientation match site[18] 5'-aTTAAGTt-3' and the single base pair mismatch site 5'-g**G**GTTTTc-3'. Affinity cleavage experiments reveal a single orientation at the match sites 5'-tAGTATTt-3' and 5'-tTGTAAAa-3' consistent with the expected hairpin complexes. A single orientation is also observed for the site 5'-aTTAAGTt-3', consistent with a reverse orientation hairpin complex.[18]

Equilibrium association constants for complexes of **1** and **2** with match and mismatch six base pair binding sites on the 229 bp restriction fragment were determined by

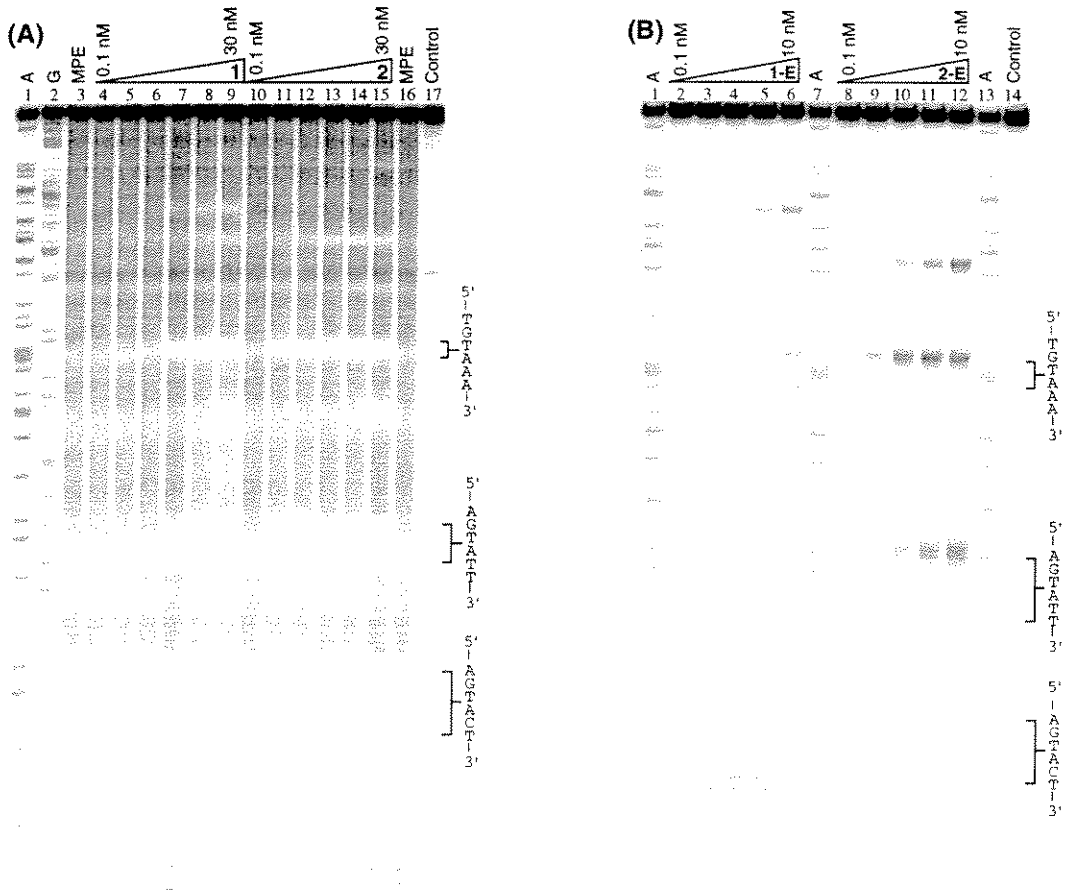


Figure 3. (A) MPE•Fe(II) footprinting experiment with polyamides **1** and **2**. Lanes 1 and 2, A and G sequencing lanes; lanes 3 and 16, cleavage products obtained in the absence of polyamides; lanes 4–9, cleavage products obtained in the presence 0.1 nM, 0.3 nM, 1 nM, 3 nM, 10 nM and 30 nM polyamide **1**, respectively; lanes 10–15, cleavage products obtained in the presence 0.1 nM, 0.3 nM, 1 nM, 3 nM, 10 nM and 30 nM polyamide **2**, respectively lane 17, intact DNA. The reactions contained 15 kcpm restriction fragment, 20 mM HEPES, 200 mM NaCl, 50 μg/mL glycogen, 0.5 μM MPE•Fe(II), and 5 mM DTT. (B) Affinity cleavage experiments with EDTA-polyamides **1-E** and **2-E**. Lanes 1, 7 and 13, A sequencing lanes; lanes 2–6, cleavage products obtained in the presence 0.1 nM, 0.3 nM, 1 nM, 3 nM and 10 nM **1-E**, respectively; lanes 8–12, cleavage products obtained in the presence 0.1 nM, 0.3 nM, 1 nM, 3 nM and 10 nM **2-E**; lane 14, intact DNA. The reactions contained 15 kcpm restriction fragment, 20 mM HEPES, 200 mM NaCl, 50 μg/mL glycogen, 1 μM Fe(II), and 5 mM DTT.

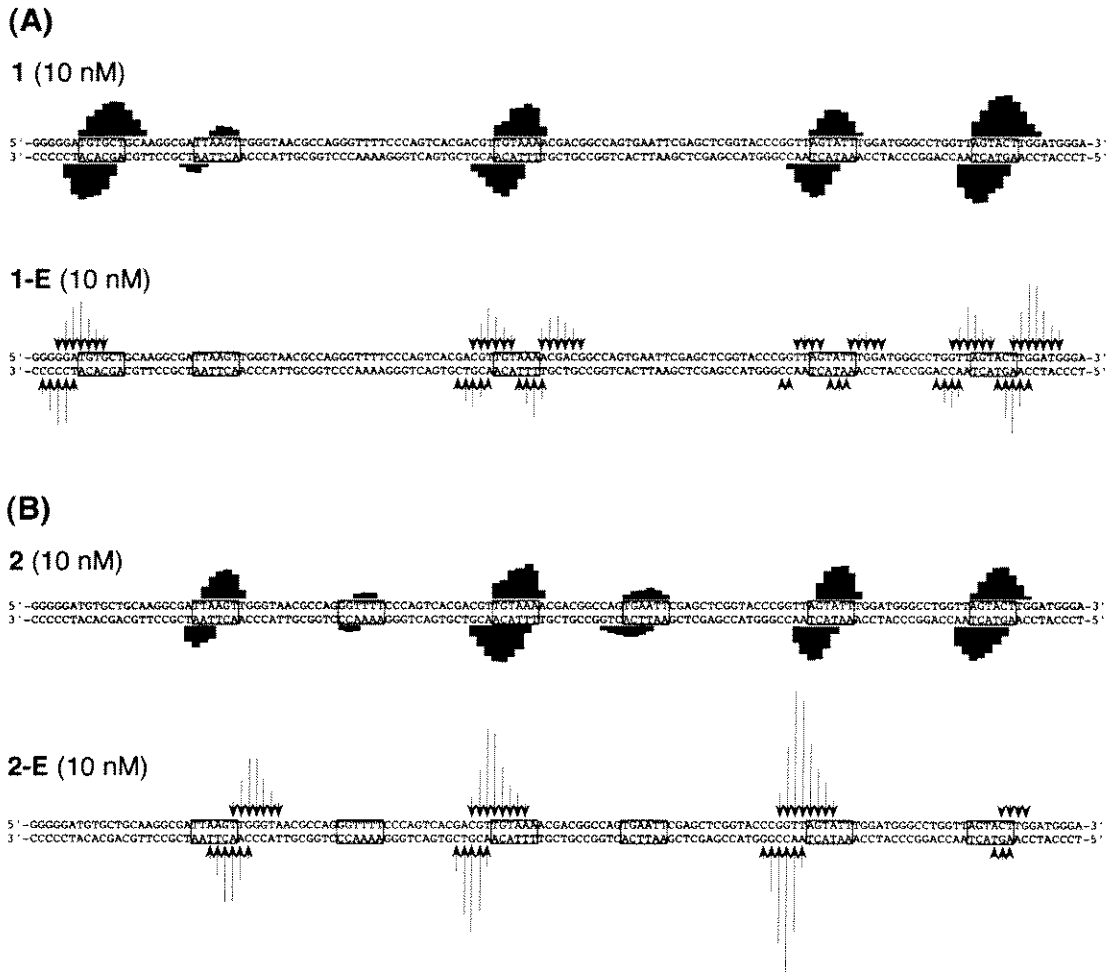


Figure 4. Results of MPE•Fe(II) footprinting and affinity cleavage experiments with (A) polyamides **1** and **1-E**, and (B) **2** and **2-E**. Bar and arrow heights are proportional to the amount of cleavage protection and cleavage, respectively, at the indicated base.

quantitative DNase I footprint titration experiments[20-22] (Fig. 5, Table 1). Polyamide **1** binds its match site 5'-AGTACT-3' at 0.03 nM concentration and its single base pair mismatch site 5'-AGTA<u>T</u>T-3' with nearly 100-fold lower affinity. Polyamide **2** binds its designated match site 5'-AGTATT-3' at 0.3 nM concentration and its single base pair mismatch site 5'-AGTA<u>C</u>T-3' with nearly 10-fold lower affinity. The specificity of **1** and **2** for their respective match sites results from very small structural changes (Fig. 1). Replacing a single nitrogen atom in **1** with C-H (as in **2**) reduces the affinity of the polyamide•5'-AGTACT-3' complex by ~75-fold representing a free energy difference of

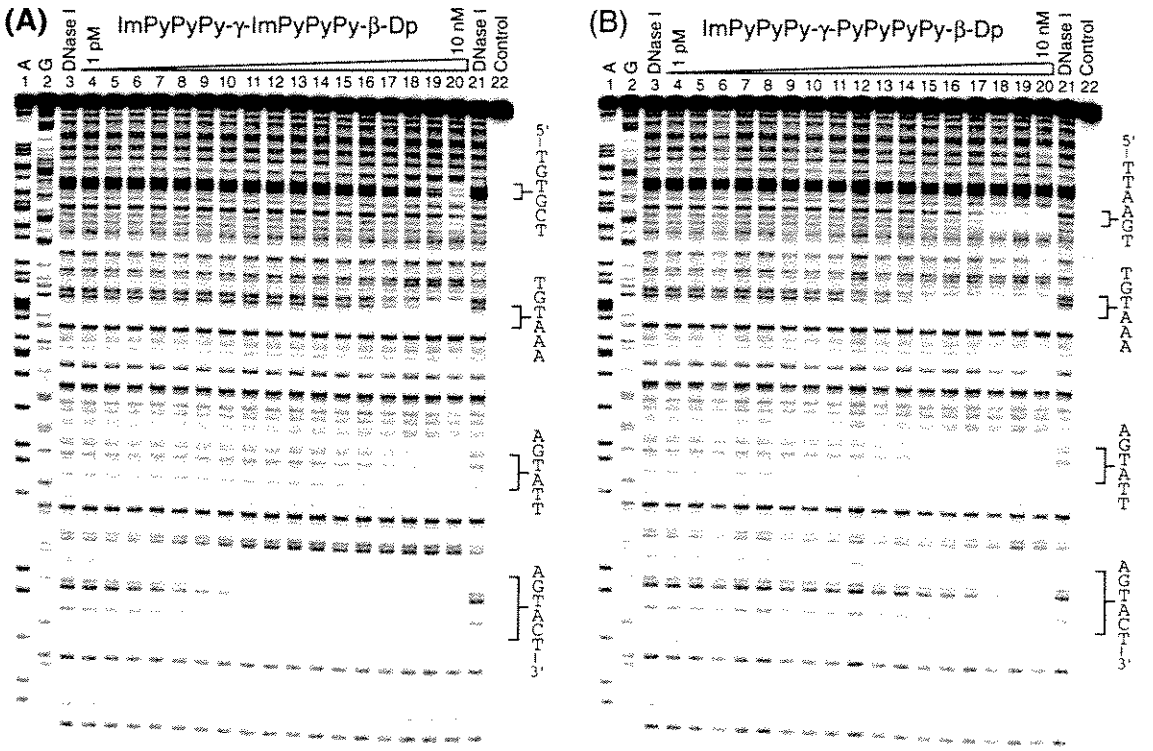


Figure 5. Quantitative DNase I footprint titration experiments with polyamides **1** and **2** on the 3'-^{32}P-labeled 229 bp pJT8 *Afl*II/*Fsp*I restriction fragment. Lanes 1-2: A and G sequencing lanes; lanes 3 and 21: DNase I digestion products obtained in the absence of polyamide; lanes 4-20: DNase I digestion products obtained in the presence of 1 pM, 2 pM, 5 pM, 10 pM, 15 pM, 25 pM, 40 pM, 65 pM, 0.1 nM, 0.15 nM, 0.25 nM, 0.4 nM, 0.65 nM, 1 nM, 2 nM, 5 nM, and 10 nM polyamide, respectively; lane 22: intact DNA. Polyamide binding sites for which association constants were determined are 5'-AGTACT-3' and 5'-AGTATT-3', 5'-TGTAAA-3', 5'-TGTGCT-3' and 5'-TTAAGT-3'. All reactions were executed in a total volume of 400 μL. A polyamide stock solution or H_2O was added to an assay buffer containing radiolabeled restriction fragment, affording final solution conditions of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, and pH 7.0. The solutions were allowed to equilibrate for 12-16 h at 22 °C prior to initiation of footprinting reactions (see Experimental Section).

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

	Polyamide	
Binding Site	**1**	**2**
5'-tAGTACTt-3'	3.7×10^{10} (0.8)	5.0×10^{8} (0.5)
5'-tAGTATTt-3'	4.1×10^{8} (0.5)	3.5×10^{9} (0.8)
5'-tTGTAAAa-3'	7.5×10^{8} (0.4)	6.5×10^{9} (1.8)
5'-aTGTGCTg-3'	6.2×10^{8} (0.9)	$< 10^{8}$
5'-gTGAATTc-3'	$< 10^{8}$	2.0×10^{8} (0.8)
5'-aTTAAGTt-3'	$< 10^{8}$	9.8×10^{8} (0.5)
5'-gGGTTTTc-3'	$< 10^{8}$	$< 10^{8}$

[a]The reported association constants are the average values obtained from three DNase I footprint titration experiments. The standard deviation for each data set is indicated in parentheses. The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [b]The 6 base pair binding sites are in capital letters, with flanking sequences in lower-case letters.

~2.5 kcal/mole. Similarly, replacing a C-H in **2** with N (as in **1**) reduces the affinity of the polyamide•5'-AGTATT-3' complex ~10-fold, a loss in binding energy of ~1.3 kcal/mol.

Association constants for additional binding sites present on the 229 base pair restriction fragment were determined. Polyamide **1** binds the single-base pair mismatch sites 5'-TGTA**A**A-3' and 5'-TGT**G**CT-3' with 50-fold and 60-fold lower affinity than the match site 5'-AGTACT-3'. Polyamide **2** binds the match site 5'-TGTAAA-3' with slightly higher affinity than the match site 5'-AGTATT-3'. The formally matched site 5'-gTGAATTc-3' site is bound by **2** with about 20-fold lower affinity than 5'-tAGTATTt-3'. The lower affinity in this case may be due to both the presence of a 5'-GA-3' step and a flanking G,C base pair opposite the C-terminal β-alanine residue. The reverse orientation site 5'-TTAAGT-3' is bound by **2** with 4–7-fold lower affinity than the match sites 5'-AGTATT-3' and 5'-TGTAAA-3'.

Crystal structures of protein-DNA complexes reveal that nature chose a combinatorial approach for specific DNA recognition. Although there are several highly

conserved structural modules that bind DNA, no single motif exists which generates an amino acid-base pair code for all DNA sequences.[23] One encouraging success is the development of zinc finger libraries with different DNA binding specificities based on a one finger-three nucleotide code.[24-27] The three-zinc finger DNA binding domain of the transcription factor Zif268 has a dissociation constant, K_d, of 0.5–3.0 nM for its 9 bp binding site and is 2–12 fold specific over single base pair mismatches.[26-29] Similarly, the three-zinc finger transcription factor Sp1 has a K_d = 0.5 nM for its 9 bp recognition sequence and is 3–30 fold specific over single base pair mismatches.[30] Using a simple molecular shape and a two letter aromatic amino acid code, pyrrole-imidazole polyamides achieve affinities and specificities comparable to these DNA-binding proteins and, in addition, have the potential to be general for any desired DNA sequence. This nonbiological approach to DNA recognition could provide an underpinning for the design of cell-permeable molecules for the control of gene expression *in vivo*.

Effect of Divalent Metal Cations. Mammalian cells have an intracellular Zn(II) concentration of ~1 μM. Polyamides contain imidazole and amide groups that could potentially coordinate Zn(II), and the formation of polyamide•Zn(II) complexes could potentially inhibit polyamide binding to DNA. To investigate this possibility, the effect of Zn(II) on the DNA-binding affinity of polyamide **1** and two other polyamides was investigated. The effects of Ni(II) and Cu(II) were also studied. Qualitative DNAse I footprint titration experiments were carried out with polyamide **1** on a restriction fragment (pXLO-wt *Eco*RI/*Hin*dIII, 160 base pairs) containing a 5'-AGTACT-3' binding site in the presence of either 20 μM $Zn(OAc)_2$, 20 μM $NiCl_2$, or 20 μM $CuSO_4$. These experiments revealed that none of the metal cations had an effect on polyamide **1** binding. Similar to **1**, binding of the six-ring hairpin polyamide ImImPy-γ-PyPyPy-β-Dp to several sites on the restriction fragment was unaffected by any of the metal ions. Binding of the hairpin dimer polyamide β-PyPyPy-γ-ImImPy-ββββ-PyPyPy-γ-ImImPy-β-Dp to two binding sites on the restriction fragment was unaffected by Zn(II) and Ni(II), but is strongly inhibited by 20

μM Cu(II). This result suggests that this polyamide selectively binds Cu(II) ions, possibly by formation of a coordination complex involving the β_4 linker.

Inhibition of 5S rRNA Gene Transcription. Polyamide **1** was originally designed to inhibit transcription of the 5S ribosomal RNA genes within *Xenopus* fibroblasts by blocking the DNA-binding activity of the 5S rRNA gene-specific transcription factor TFIIIA. The results of 5S rRNA gene transcription inhibition experiments with polyamide **1** are described in Chapter Eleven.

Experimental Section

Plasmid preparation. Plasmid pJT8 was prepared by hybridizing two 5'-phosphorylated complementary oligonucleotides, 5'-CCGGTTAGTATTTGGATG-GGCCTGGTTAGTACTTGGATGGGAGACCGCCTGGGAATACCAGGTGTCGTATC TTAAGAG-3' and 5'-TCGACTCTTAAGATACGACACCTGGTATTCCCAGGCGGTC-TCCCATCCAAGTACTAACCAGGCCCATCCAAATACTAA-3', and ligating the resulting duplex to the large pUC19 *AvaI*/*SalI* restriction fragment.

Preparation of ^{32}P-End-Labelled Restriction Fragments. Plasmid pJT8 was digested with *Afl*II, labeled at the 3'-end using Sequenase (version 2.0), and digested with *Fsp*I. The 229 bp restriction fragment was isolated by nondenaturing gel electrophoresis and used in all quantitative footprinting experiments described here. 5'-end labeling of the 229 bp restriction fragment was carried out using standard methods. Chemical sequencing reactions were performed as described.[31,32] Standard techniques were employed for DNA manipulations.

MPE•Fe(II) Footprinting and Affinity Cleavage. The protocol reported here avoids the use of carrier DNA and gives apparent affinities that are the same as those obtained in DNase I footprinting assays. The final solution conditions are 20 mM HEPES, 200 mM NaCl, 50 μg/mL glycogen, 5 mM DTT, pH 7.3, and either (i) 0.5 μM

MPE•Fe(II) for footprinting experiments, and (ii) 1 μM Fe(II) for affinity cleavage experiments.

1. Prepare a 1.43x DNA solution: To a 15 mL Falcon 2059 tube add 1,286 μL 5x "cleavage buffer" (100 mM HEPES, 300 mM NaCl, pH 7.3), 225 μL 4M NaCl, 3 mL water, and labeled DNA (250,000 cpm).

2. Prepare 10x polyamide (or EDTA-polyamide) stock solutions.

3. Set up equilibrations: To a 1.5 mL Eppendorf tube add: 280 μL 1.43x DNA solution, 40 μL 10x polyamide and 40 μL glycogen (0.5 mg/mL). Allow the solution to equilibrate for 1–18 hours.

4. Prepare a "precipitation buffer" by combining 35 μL glycogen (20 mg/mL), 35 μL calf thymus DNA (1 mM bp) and 180 μL water.

5. Prepare 5 μM MPE•Fe(II) by combining equal volumes of 10 μM MPE and freshly prepared 10 μM $Fe(NH_4)_2(SO_4)_2$.

6a. For MPE footprinting experiments: To the polyamide/DNA solution add 40 μL 5 μM MPE-Fe(II), allow to equilibrate for 10 minutes. Then add 10 μL 200 mM DTT (use a freshly thawed DTT aliquat). Allow cleavage to proceed for 30 minutes. Then add 1 mL ethanol, mix the tubes by inversion, and spin briefly.

6b. For affinity cleavage experiments: To the polyamide-EDTA/DNA solution add 20 μL freshly prepared $Fe(NH_4)_2(SO_4)_2$ and equilibrate for 10–30 minutes. Next add 40 μL DTT (use a freshly thawed aliquat) and allow cleavage to proceed for 30 minutes. Then add 1 mL ethanol, mix the tubes by inversion, and spin briefly.

7. Precipitate the DNA and run the gel: Add 10 μL "precipitation buffer" to each tube and mix by inversion. Spin the tubes at 14,000 rpm in a cold room for 30 minutes. Decant, wash with 75% ethanol (350 μL), and decant. Resuspend the DNA in 16 μL water, freeze and speed-vac dry. Resuspend in 7 μL 80% formamide/1x TBE loading buffer, heat denature (10 minutes at 85–90 °C, then place on ice) and load 5 μL per lane on a denaturing 8% gel and run the gel.

8. Transfer the gel to paper, cover with plastic wrap and dry on a gel dryer (45–60 minutes) at 80 °C. Expose the gel to a storage phosphor screen for 8–16 hours and image using the Phosphorimager. The extent of cleavage protection (for MPE•Fe(II) footprinting) and cleavage intensities (for affinity cleavage experiments) are determined by quantitation of the gel using ImageQuant software.

Quantitative DNase I Footprinting. Equilibrations are carried out in a total volume of 400 μL in the absence of carrier DNA. The final solution conditions for this protocol are 10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$ and pH 7.0.

1. Prepare a 1.14x DNA solution: To a 15 mL Falcon tube (Falcon 2059) add 2.06 mL 5x TKMC buffer (50 mM Tris-HCl, 50 mM KCl, 50 mM $MgCl_2$, and 25 mM $CaCl_2$, pH 7.0), 7.0 mL water, and ^{32}P-labeled DNA (~400,000 cpm). The amount of DNA used is calculated to give a final loading of ~15,500 cpm per lane.

2. Prepare polyamide stock solutions: Dissolve 25 nmol polyamide in 500 μL MilliQ water, and dilute 5-fold to give an approximately 10 μM solution. Determine the concentration of this solution by UV absorption, then dilute appropriately to give a final concentration of 1 μM. Prepare 10x polyamide stock solutions.

3. Prepare equilibrium binding mixtures: To a 1.5 mL Eppendorf tube (presiliconized) add 40 μL 10x polyamide solution (or water for control lanes) and 350 μL 1.14x DNA solution. Allow the mixtures to equilibrate at room temperature (24 °C) for 12–18 hours.

4. Prepare a DNase I "stop buffer:" Combine 40 μL glycogen (20 mg/mL), 40 μL 1 mM bp calf thymus DNA, 107 μL water, 788 μL 4M NaCl, and 425 μL 0.5 M EDTA (pH 8.0). The EDTA in this buffer quenches DNase I by chelating the essential Mg^{2+} and Ca^{2+} ions. The other ingredients provide for good EtOH precipitation and subsequent resuspension.

5. Prepare a DNase I stock solution: To a 1.5 mL Eppendorf tube add 975 μL water and 20 μL 50 mM DTT, and chill the solution on ice. Then add 5 μL DNase I

(Pharmacia FPLCPure, 7,500 units/mL) and mix the resulting 38 units/mL solution by inverting the tube several times. Prepare the final stock solution by adding 15–70 µL DNase I (38 units/mL) to a prechilled mixture of 20 µL 50 mM DTT and 930–965 µL water (the total final volume should be 1,000 µL). Keep the DNase I solution on ice throughout the experiment (this enzyme solution should be freshly prepared from the 7,500 unit/mL stock for each set of footprinting experiements). The exact amount of DNase I to use depends on the batch of DNase I and the restriction fragment being used. The goal is to achieve ~50% digestion of the restriction fragment. While the exact amount of DNase I can be determined by running a titration, it is often possible to estimate based on the length of the restriction fragment. For a 220 bp fragment, dilute ~50 µL of the 38 units/mL stock solution to 1 mL. Shorter restriction fragments require more DNase I; longer fragments less (e.g., for a 440 bp fragment, use ~25 µL of the 38 units/mL solution).

6. Digest the DNA: To each tube (except the intact DNA control) add 10 µL of the final DNase I solution and mix by vortexing. Allow the reaction to proceed for 7 minutes at room temperature. Then add 50 µL "stop buffer" and mix by vortexing. Then add 975 µL room temperature absolute ethanol and mix by inversion.

7. Precipitate the DNA: Spin the tubes in a microcentrifuge in a cold room at 14,000 rpm for 25–30 minutes. Carefully decant the supernatant. Next, add 300 µL 75% EtOH. Vortex the tube briefly to thoroughly wash out residual salt. Spin the tubes briefly (10–20 seconds) in a microcentrifuge. Carefully decant the supernatant.

8. Resuspend the DNA: To each tube add 15 µL water and vortex (5–10 seconds) to resuspend the DNA. Freeze the samples by placing them in crushed dry ice or in a -75 °C freezer (samples can be safely left in this state overnight). Dry in a speed-vac (to reduce the chances of hanging lanes, don't over-dry; it's best to take the samples out as soon as they are dry, usually 35–40 minutes). Next, add 7 µL 80% formamide/1x TBE loading buffer (loading buffer should be stored at 4 °C; loading buffer that smells of ammonia

should be discarded). Thoroughly vortex (15–20 seconds) each tube to resuspend the DNA.

9. Denature the DNA and run the gel: Denature the DNA by heating 10 minutes at 85–90 °C, then immediately place the samples on ice. Load 5 μL per lane on a pre-run 8% (19:1 acrylamide-bisacrylamide) denaturing polyacrylamide gel (pre-run the gel for 15–40 minutes). Run the gel at 2,000 V (running bromophenol blue, the faster dye, just to the bottom of the gel is good for resolving DNA sites 40–90 base pairs from the labeled end). Gel plates may be treated with SigmaCote occasionally.

10. Transfer the gel to paper, cover with plastic wrap, and dry on a gel dryer (45–60 minutes) at 80 °C. Expose the gel to a storage phosphor screen for 8–16 hours and image using the Phosphorimager. Analysis of data from quantitative footprinting experiments is carried out as described in Chapter 2.

References

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

(2) Mrksich, M.; Wade, W.S.; Dwyer, T.J.; Geirstanger, B.H.; Wemmer, D.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586–7590.

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

(4) Pelton, J.G.; Wemmer, D.E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723–5727.

(5) Pelton, J.G.; Wemmer, D.E. *J. Am. Chem. Soc.* **1990**, *112*, 1393–1399.

(6) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, ***115***, 2572–2576.

(7) Geierstanger, B.H.; Jacobsen, J.P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Biochemistry* **1994**, *33*, 3055.

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

(9) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646–650.

(10) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 3325–3332.

(11) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 9892–9899.

(12) Dwyer, T.J.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 9900–9906.

(13) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 3663–3664.

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

(15) Chen, Y.H.; Lown, J.W. *J. Am. Chem. Soc.* **1994**, *116*, 6995–7005.

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

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

(18) White, S.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, *119*, 8756–8765.

(19) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1998**, submitted.

(20) Galas, D.; Schmitz, A. *Nucleic Acids Res.* **1978**, *5*, 3157–3170.

(21) Fox, K.R.; Waring, M.J. *Nucleic Acids Res.* **1984**, *12*, 9271–9285.

(22) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods in Enzymology* **1986**, *130*, 132–181.

(23) Steitz, T.A. *Quart. Rev. Biophys.* **1990**, *23*, 205–280.

(24) Desjarlais, J.R.; Berg, J.M. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7345–7349.

(25) Desjarlais, J.R.; Berg, J.M. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 2256–2260.

(26) Jamieson, A.C.; Kim, S.-H.; Wells, J.A. *Biochemistry* **1994**, *33*, 5689–5695.

(27) Rebar, R.J.; Pabo, C.O. *Science* **1994**, *263*, 671–673.

(28) Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11163–11167.

(29) Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11168–11172.

(30) Letovsky, J.; Dynan, W.S. *Nucleic Acids Res.* **1989**, *17*, 2639–2653.

(31) Maxam, A.M.; Gilbert, W.S. *Methods in Enzymology* **1980**, *65*, 499–560.

(32) Iverson, B.L.; Dervan, P.B. *Nucleic Acids Res.* **1987**, *15*, 7823–7830.

CHAPTER FIVE

A Polyamide Motif for Recognition of 7 Base Pair DNA Sequences: Targeting the HIV-1 TATA Box

The text of this chapter is taken from a manuscript that was coauthored with

Prof. Peter B. Dervan and Eldon E. Baird

Introduction

Small molecules which bind specific DNA sequences and inhibit protein-DNA interactions have the potential to regulate gene expression *in vivo*.[1-5] Pyrrole-imidazole polyamides, which currently represent the only class of small molecule that can bind predetermined DNA sequences, are promising candidates for such applications.[5] Recent advances in polyamide design have led to synthetic ligands that bind target sites with specificities and affinities comparable to natural DNA-binding transcription factors.[6,7] In a 5S RNA gene model system, a polyamide was shown to be cell-permeable and to inhibit transcription of 5S RNA genes in Xenopus cells.[5] These results motivate further polyamide design efforts and further investigation of the potential of pyrrole-imidazole polyamides as regulators of gene expression.

The protein-encoding genes in mammalian cells are transcribed by RNA Polymerase II.[8] Furthermore, the promoters of most tissue-specific genes and viral protein-encoding genes contain binding sites (TATA boxes) for the TATA-binding protein (TBP).[8] Binding of the TBP subunit of TFIID to the TATA box is known to nucleate assembly of the RNA Polymerase II transcription machinery.[8] The X-ray crystal structure of TBP bound to DNA shows that TBP binds in the minor groove, inducing a severe bend and unwinding of the DNA.[9] This information indicates that minor groove-binding polyamides that block TBP/TFIID-DNA interactions in a gene-specific manner could provide a general approach to regulation of gene expression by small molecules. Support for this approach is provided by the observation that the minor groove-binding ligand distamycin can bind directly to TATA box sequences and inhibit transcription by interfering with TBP/TFIID binding.[4]

Targeting the HIV-1 TATA Box. With the ultimate goal of testing whether polyamides could specifically regulate RNA Polymerase II transcription by inhibiting TBP/TFIID-DNA interactions, we set out to design a polyamide to bind with subnanomolar affinity to a target sequence immediately adjacent to the HIV-1 TATA box. Transcription

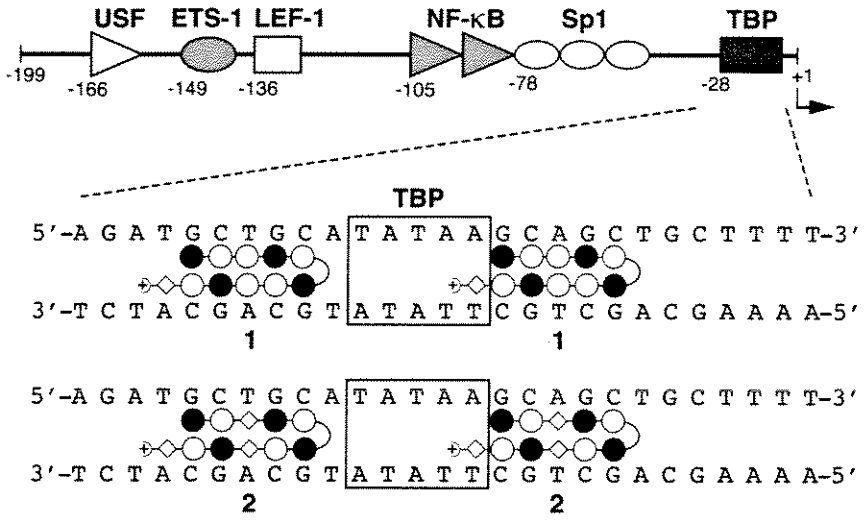


Figure. 1. (Top) Schematic representation of the HIV-1 promoter/enhancer. (Bottom) The nucleotide sequence of the TATA box region of a typical HIV-1 promoter is shown, along with binding models for polyamides **1** and **2**.

of HIV-1 by RNA Polymerase II is regulated through the viral promoter/enhancer, a 250 base pair DNA segment located within the left long terminal repeat of the HIV-1 genome (Fig. 1).[10] The promoter/enhancer contains a TATA box as well as binding sites for the cellular transcription factors USF, Ets-1, LEF-1, NF-κB, and Sp1.[10]

Inspection of the sequence of a typical HIV-1 promoter revealed 7 base pair 5'-(A,T)GC(A,T)GC(A,T)-3' sequences immediately flanking the TATA box on both sides (Fig. 1).[10] A search of GenBank revealed that this sequence is present on at least one side of the TATA box for all strains of HIV-1 in the data base. Notably, the 5'-AGCAGCTGCT-3' sequence immediately downstream of the TATA box is part of the conserved consensus sequence for all lentiviral TATA box sequences.[11] We note explicitly that, other factors aside, the high mutation rate of HIV-1 genomes may render sequence-specific DNA-binding molecules ineffective therapeutics for treatment of HIV-1 infection.[12] Nonetheless, the successful design of polyamides to target these sequences would allow determination of whether a polyamide binding adjacent to a TATA box can disrupt TBP-DNA complexes and transcription *in vitro* and viral replication in cell culture. We were

thus challenged to design a polyamide that specifically binds 5'-(A,T)GC(A,T)GC(A,T)-3' sequences with subnanomolar affinity.

Pairing Rules and Binding Motifs. The sequence specificity of pyrrole-imidazole polyamides depends on side-by-side amino acid pairings.[13-21] A pairing of imidazole (Im) opposite pyrrole (Py) targets a G•C base-pair, while Py/Im targets C•G.[13] A Py/Py combination is degenerate and targets both T•A and A•T base pairs.[17-19] The pairing of two β-alanines opposite one another is specific for A,T versus G,C base pairs.[22,23]

The pairing rules described above function within the context of a binding motif or "molecular template" that correctly positions imidazole and pyrrole rings. The first binding motif for DNA recognition by polyamides consisted of three-ring polyamides which bind 5 bp sites as antiparallel dimers.[13-15] Subsequent efforts led to the development of motifs which have increased the binding affinity and specificity, binding site size, and sequence repertoire of polyamides.[6,7,20,21]

Design of First-Generation Polyamide 1. Recently, it was shown that the ten-ring, "5-γ-5 motif" polyamide ImPyPyPyPy-γ-ImPyPyPyPy-β-Dp (**5**) (Im = *N*-methylimidazole, Py = *N*-methypyrrole, γ = γ-aminobutyric acid, β = β-alanine) specifically binds a 7 base pair 5'-(A,T)G(A,T)$_3$C(A,T)-3' target sequence with subnanomolar affinity in a "hairpin" conformation.[7] Applying the polyamide pairing rules suggested that the ten-ring hairpin polyamide ImPyPyImPy-γ-ImPyPyImPy-β-Dp (**1**) would bind the HIV-1 target sequence5'-(A,T)GC(A,T)GC(A,T)-3' (Figs. 1 and 2). As reported below, polyamide **1** specifically binds the target site 5'-TGCTGCA-3' with very low affinity, necessitating development of a second-generation polyamide.

Design of Second-Generation Polyamide 2. The amino acid β-alanine has been employed effectively as a single base pair-spanning linker in several different polyamide motifs.[20-22] In particular, it has been shown that replacing a pyrrole residue that precedes an imidazole residue with β-alanine can reset the imidazole binding register

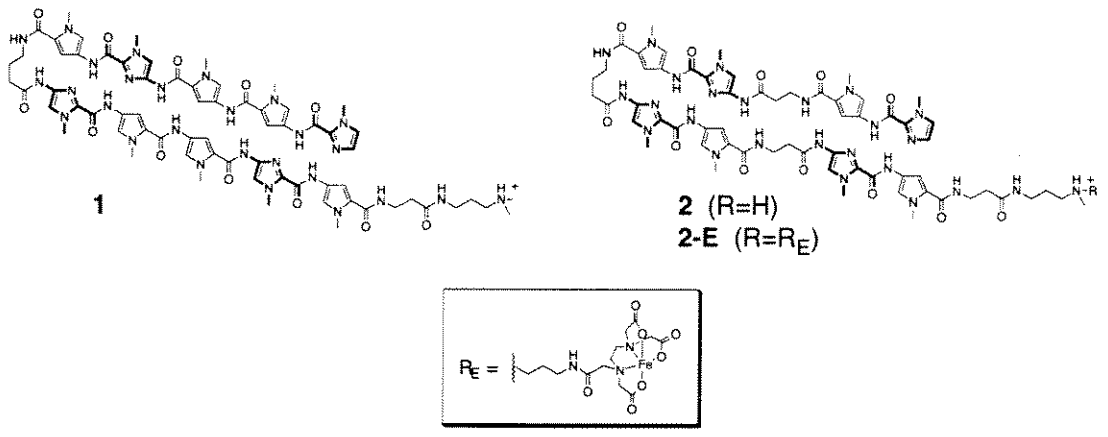


Figure. 2. Structures of 5-γ-5 polyamide **1** and 2-β-2-γ-2-β-2 polyamide **2**, and EDTA-polyamide **2-E**.

resulting in marked increases in binding affinity.[22] For example, the β-alanine-substituted hairpin polyamide Im-β-ImPy-γ-Im-β-ImPy-β-Dp binds 5'-TGCGCA-3' with 100-fold higher affinity than ImPyImPy-γ-ImPyImPy-β-Dp.[22] Additional results have demonstrated that β and γ preferentially bind in extended and hairpin conformations, respectively.[20,23,24] These results suggested that the polyamide ImPy-β-ImPy-γ-ImPy-β-ImPy-β-Dp (**2**), which differs from **1** by the replacement of two pyrrole residues with β-alanines, would specifically bind the 5'-(A,T)GC(A,T)GC(A,T)-3' HIV-1 target sequences with high affinity (Figs. 1 and 2).

We report here equilibrium association constants determined by quantitative DNase I footprint titration experiments of the formally matched (according to the polyamide pairing rules) polyamides **1** and **2** for the HIV-1 target sequence 5'-TGCTGCA-3' (see Fig. 1). To explore the generality of the 2-β-2-γ-2-β-2 motif, we investigated the DNA-binding properties of four additional polyamides to their respective formal match sequences 5'-TGGTGGA-3' and 5'-TGTTACA-3'. We also report the results of MPE•Fe(II) footprinting and affinity cleavage studies for a subset of these compounds. Finally, we compare the equilibrium association constants of polyamide **2** for the HIV-1 target

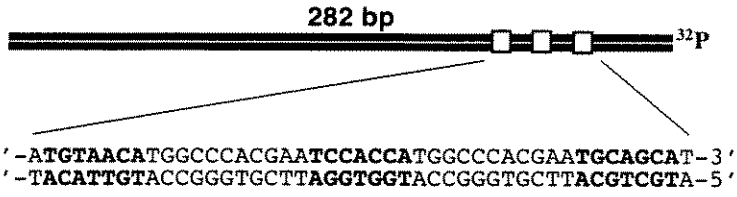


Figure 3. Partial sequence of the 282 bp restriction fragment used in quantitative DNase I footprinting experiments. Three 7 bp binding sites, the HIV-1 target site 5'-TGCTGCA-3' and the mismatch sites 5'-T**C**CA**C**CA-3' and 5'-TG**T**A**A**CA-3' are highlighted. Additional binding sites present on the restriction fragment are 5'-TCGTCGA-3' (reverse orientation binding site) and 5'-tAGCTG**T**Tt-3'.

sequence 5'-TGCTGCA-3' determined at either 24 °C or 37 °C, and using either standard polyamide assay solution conditions or model intracellular solution conditions.

Results and Discussion

Synthesis. Polyamides were synthesized by solid phase methods.[25]

Binding Affinities. Quantitative DNase I footprint titration experiments[26] (10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, 24 °C) were carried out on the 282 bp, 3'-^{32}P end-labeled pJT2B2 *Eco*RI/*Pvu*II restriction fragment which contains the three target sites 5'-TGCTGCA-3' (the HIV-1 target sequence), 5'-TGGTGGA-3', and 5'-TGTTACA-3' (Figs. 3, 4 and 5, Table 1). These experiments reveal that the 5-γ-5 motif polyamide **1** binds the HIV-1 target sequence 5'-TGCTGCA-3' with a relatively low equilibrium association constant, $K_a = 8 \times 10^7\ M^{-1}$. In contrast, polyamide **2** binds the HIV-1 target sequence 5'-TGCTGCA-3' with subnanomolar affinity, $K_a = 2 \times 10^{10}\ M^{-1}$, and is 10-fold specific for this site relative to the single-base pair mismatch site 5'-AGCTG**T**T-3' and the reverse orientation site 5'-TCGTCGA-3'. The two-base pair mismatch sites 5'-TG**G**TG**G**A-3' and 5'-TG**T**T**A**CA-3' are bound with >30-fold lower affinity compared to the match site. Thus, **2** represents a solution to the HIV-1 TATA box polyamide design problem.

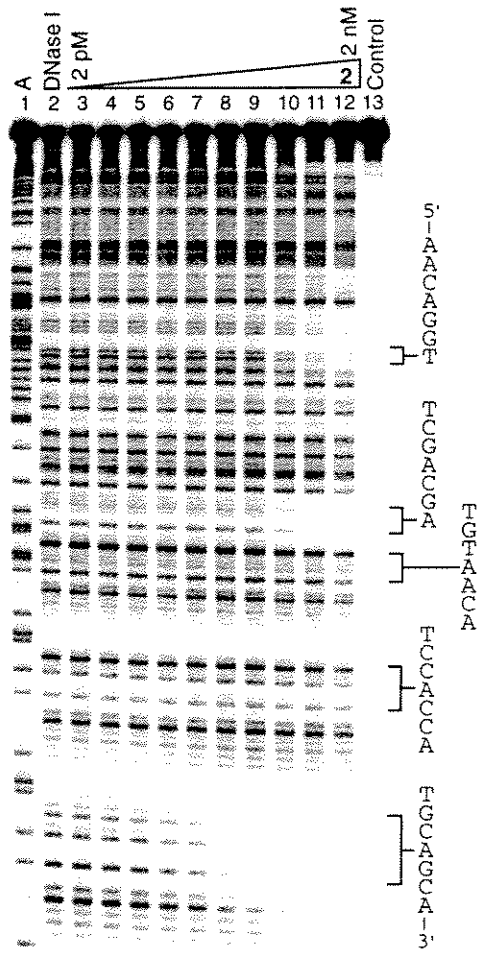


Figure 4. Quantitative DNase I footprinting experiment for polyamide **2**. Lane 1, A sequencing lane; lane 2, DNase I digestion products obtained in the absence of polyamide; lanes 3-12, DNase I digestion products obtained in the presence of 0.002 nM, 0.005 nM, 0.01 nM, 0.02 nM, 0.05 nM, 0.1 nM, 0.2 nM, 0.5 nM, 1 nM, and 2 nM, respectively; lane 13, intact DNA. Polyamide binding sites are indicated along the right side of the autoradiogram. All reactions contained 15 kcpm restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 5 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0 and 24 °C.

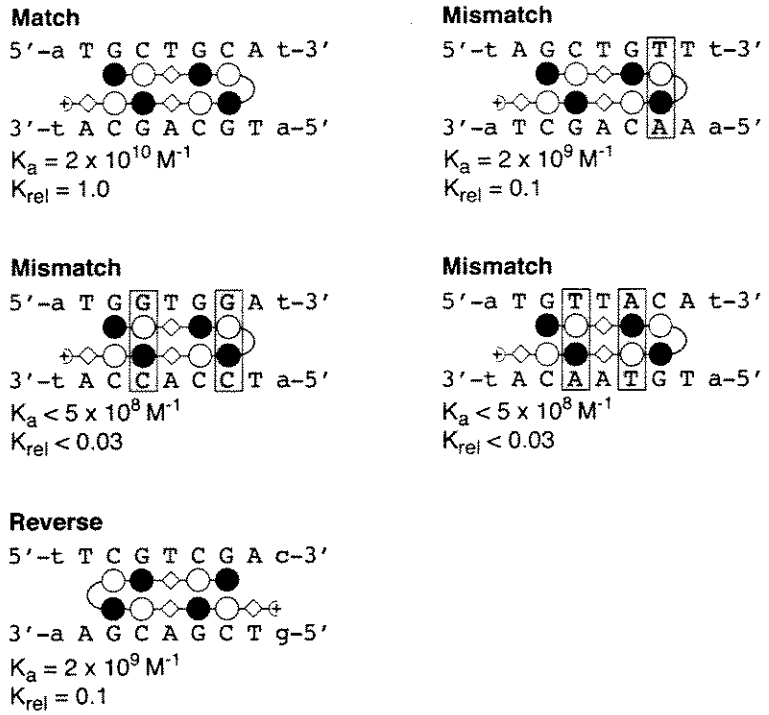


Figure 5. Association constants and binding models for complexes of polyamide **2**.

Generality of the 2-β-2-γ-2-β-2 Motif. To investigate the generality of the 2-β-2-γ-2-β-2 motif, polyamides designed to target the 5'-TGGTGGA-3' (**3** and **4**), and 5'-TGTTACA-3' (**5** and **6**) sequences present on the restriction fragment were synthesized (Fig. 7, Table 1). Polyamides **3** and **4** bind 5'-TGGTGGA-3' with equilibrium association constants of < 5 x 10^7 M-1 and 8 x 10^9 M-1, respectively, and polyamides **5** and **6** bind 5'-TGTTACA-3' with equilibrium association constants of 5 x 10^{10} M^{-1} and 3 x 10^9 M^{-1}, respectively.

The 5-γ-5 polyamides **1**, **3**, and **5** are analogs of **2**, **3**, and **4**, respectively, in which the β-alanine residues have been changed to pyrrole residues (Fig. 2). Polyamides **1** and **3**, which have imidazole residues beyond the second position of a polyamide subunit, have equilibrium association constants for their match sites more than 100-fold lower than their β-alanine-substituted analogs **2** and **4**, respectively (Table 1). Thus, *the 2-*

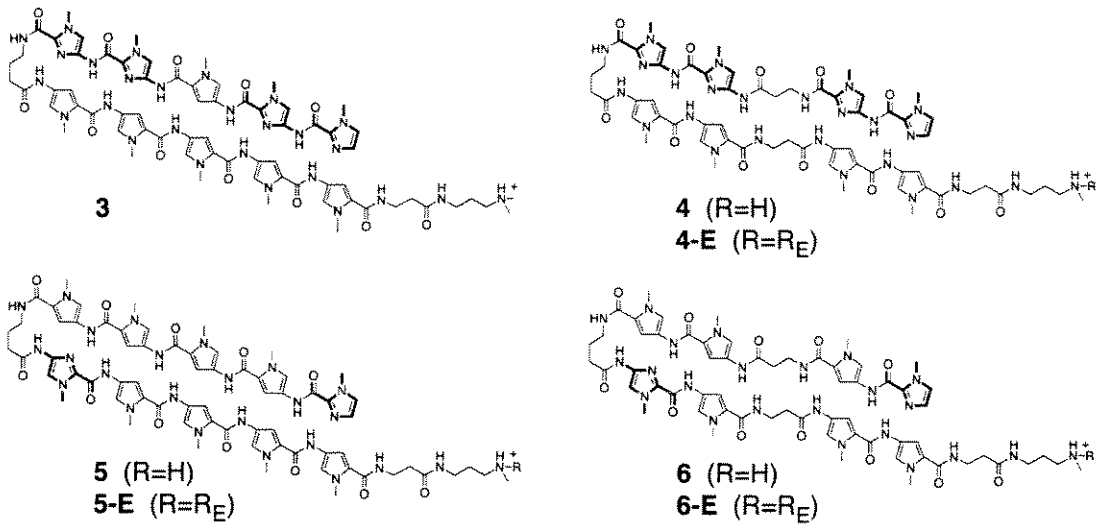


Figure 6. Structures of polyamides **3–6**, and EDTA-polyamides **4-E**, **5-E**, and **6-E**.

β-2-γ-2-β-2 motif is essential for high-affinity recognition in these cases. In contrast, polyamide **5** having no imidazoles beyond the second position of a polyamide subunit binds its match site with an equilibrium association constant ~20-fold higher than its β-alanine-substituted analog **6** (Table 1).

Role of the β-alanine Linkers in the 2-β-2-γ-2-β-2 Motif. The failure of polyamides **1** and **3** to bind their respective target sequence with high affinity is consistent with previous results. The 4-γ-4 motif polyamides ImPyPyPy-γ-ImPyPyPy-β-Dp, ImPyPyPy-γ-PyPyPyPy-β-Dp and ImImPyPy-γ-ImImPyPy-β-Dp bind respective 6 base pair target sequences with subnanomolar affinity ($K_a \geq 10^{10}$ M^{-1}),[6,27] while ImPyImPy-γ-ImPyImPy-β-Dp and ImImImIm-γ-PyPyPyPy-β-Dp bind respective target sequences with substantially lower affinity ($K_a < 10^8$ M^{-1}).[27] Based on these and additional results, the common feature of hairpin polyamides that bind with high affinity appears to be the placement of imidazole residues only at the first and second positions of polyamide subunits, counting from the N-terminal end, although in some cases substitution at the third position is tolerated.[6,21,27-29] Apparently imidazoles placed beyond the second position of a polyamide subunit are not positioned optimally for specific ligand-DNA contacts. The high

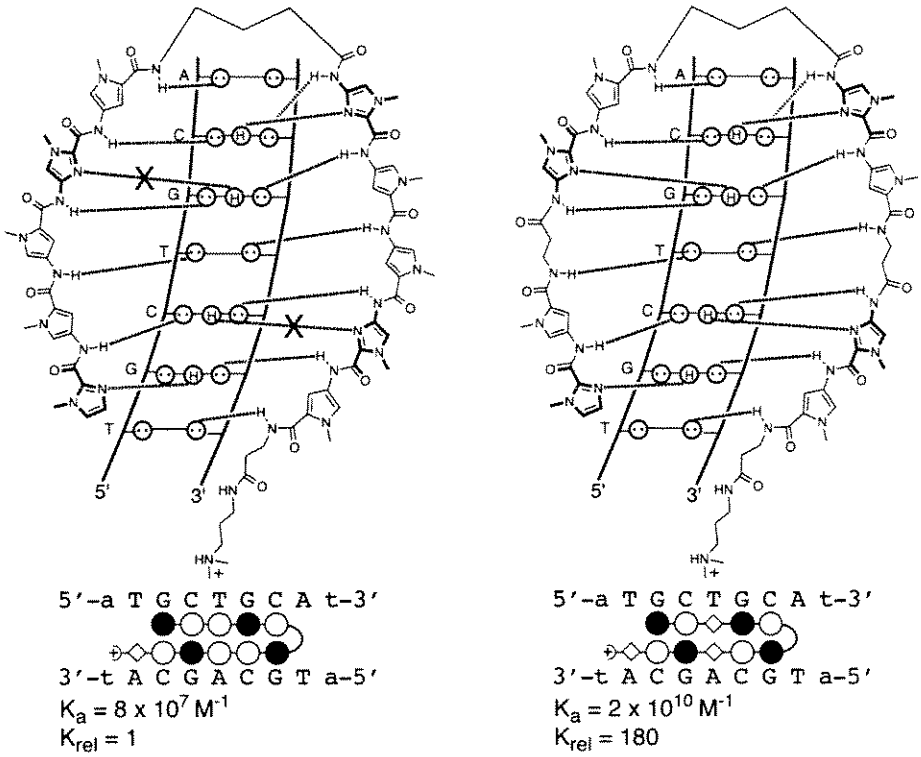


Figure 7. Models and experimentally determined equilibrium association constants for the complexes of polyamides **1** (left) and **2** (right) with the sequence 5'-TGCTGCA-3'.

affinity observed with the β-alanine-containing analogs **2** and **4** indicates that the flexible β-alanine residues reset the imidazole binding register (Fig. 7).

Relative Effects of Covalent Linkage with γ-Aminobutyric Acid. Equilibrium association constants for the "unlinked dimer motif" polyamides **7**, **8** and **9** were measured to allow comparison with their corresponding γ-aminobutyric acid-linked

Table 1. Equilibrium association constants $(M^{-1})^a$ polyamides **5**, **2**, and **6**, respectively. Polyamides **2** and **6**, which are composed of 2-β-2 subunits, bind with ~6000-fold and ~1000-fold higher affinity, respectively, relative to their respective unlinked analogs **8** and **9**. Polyamide **5**, which is composed five-ring subunits, binds with ~100-fold higher affinity than its unlinked analog **7**. The significantly larger increase in affinity upon linkage with γ-aminobutyric acid for **8** and **9** relative to **7** is consistent with the greater structural rigidity and greater preorganization of **7**.

Table 1. Equilibrium association constants (M^{-1}).

Polyamide	5'-aTGCTGCAt-3'	5'-aTGGTGGAt-3'	5'-aTGTTACAt-3'
5-γ-5motif:			
1 ImPyPyImPy-γ-ImPyPyImPy-β-Dp	**8.3 x 10^7 (1.5)**	< 1 x 10^7	< 1 x 10^7
3 ImImPyImIm-γ-PyPyPyPyPy-β-Dp	< 5 x 10^7	**< 5 x 10^7**	< 5 x 10^7
5 ImPyPyPyPy-γ-ImPyPyPyPy-β-Dp	1.2 x 10^9 (0.4)	< 5 x 10^8	**5.1 x 10^{10} (1.1)**
2-β-2-γ-2-β-2 motif:			
2 ImPy-β-ImPy-γ-ImPy-β-ImPy-β-Dp[b]	**1.5 x 10^{10} (0.3)**	< 5 x 10^8	< 5 x 10^8
4 ImIm-β-ImIm-γ-PyPy-β-PyPy-β-Dp	1.7 x 10^8 (0.3)	**7.6 x 10^9 (1.0)**	< 1 x 10^8
6 ImPy-β-PyPy-γ-ImPy-β-PyPy-β-Dp	< 1 x 10^8	< 1 x 10^8	**2.6 x 10^9 (1.0)**
Unlinked dimer motif:			
7 ImPyPyPyPy-β-Dp	1.1 x 10^7 (0.2)	< 1 x 10^7	**3.9 x 10^8 (0.6)**
8 ImPy-β-ImPy-β-Dp	**3.0 x 10^6 (0.2)**	< 2 x 10^5	1.5 x 10^6 (0.3)
9 ImPy-β-PyPy-β-Dp	< 5 x 10^5	< 5 x 10^5	**2.8 x 10^6 (0.1)**

[a]Values reported are the mean values from at least three DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses. Assays were carried out at 24 °C, pH 7.0, in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. Nucleotides flanking polyamide binding sites are in lowercase type. Association constants corresponding to formally matched complexes are in bold type. [b]This polyamide binds the sites 5'-TCGTCGA-3' and 5'-tAGCTGTTt-3' with equilibrium association constants of 2.1 x 10^9 M^{-1} (0.1) and 1.7 x 10^9 M^{-1} (0.1), respectively.

Exact Binding Sites and Binding Orientations. MPE•Fe(II) footprinting experiments[30] carried out with polyamides **2**, **4**, **5** and **6**, which have subnanomolar binding affinities, confirm that these compounds specifically bind their respective 7 base pair match sites (Figs. 8A and 9A). Affinity cleavage experiments[31] were performed with the EDTA-polyamides **2-E**, **4-E**, **5-E** and **6-E** to identify the location of the C-termini of these polyamides when bound to their target sites (Figs. 8B and 9B). The symmetric polyamides **2-E**, **5-E**, and **6-E** produce cleavage patterns at both sides of their binding sites as expected, consistent with two distinct binding orientations. It appears that, in general, side-by-side polyamide-DNA complexes preferentially bind in the 5' to 3' (N- to C-terminus) orientation.[13] Recent results with hairpin polyamides are consistent with

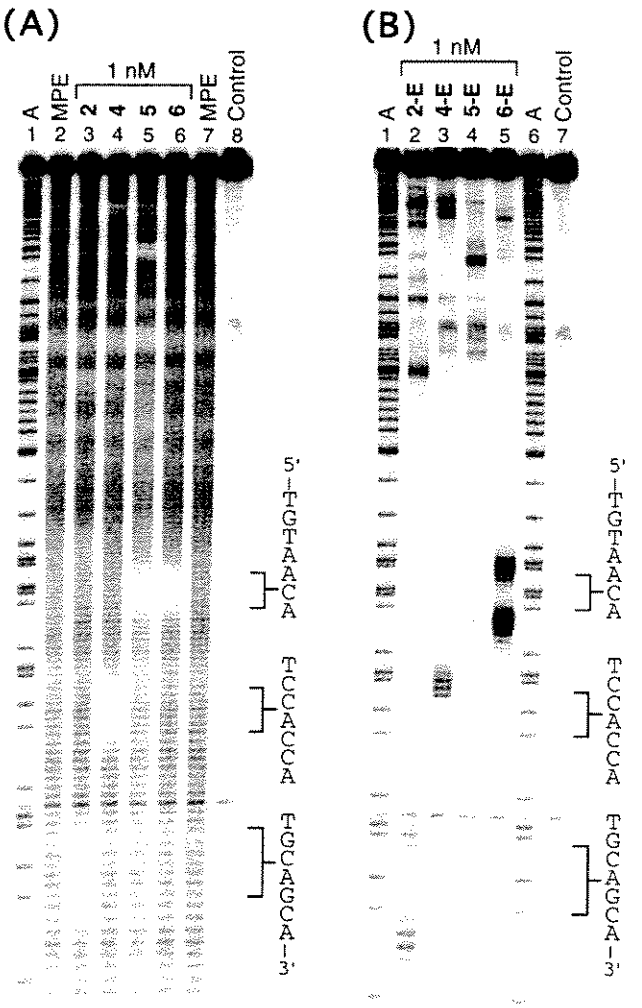


Figure 8. (A) MPE•Fe(II) footprinting of polyamides **2**, **4**, **5** and **6** at 1 nM. The reactions contained 15 kcpm restriction fragment, 20 mM HEPES, 200 mM NaCl, 50 μg/mL glycogen, 0.5 μM MPE•Fe(II) and 5 mM DTT. (B) Affinity cleavage experiments with polyamides **2-E**, **4-E**, **5-E**, and **6-E** at 1 nM. The reactions contained 15 kcpm restriction fragment, 20 mM HEPES, 300 mM NaCl, 50 μg/mL glycogen, 1 μM Fe(II) and 5 mM DTT at pH 7.0 and 24 °C.

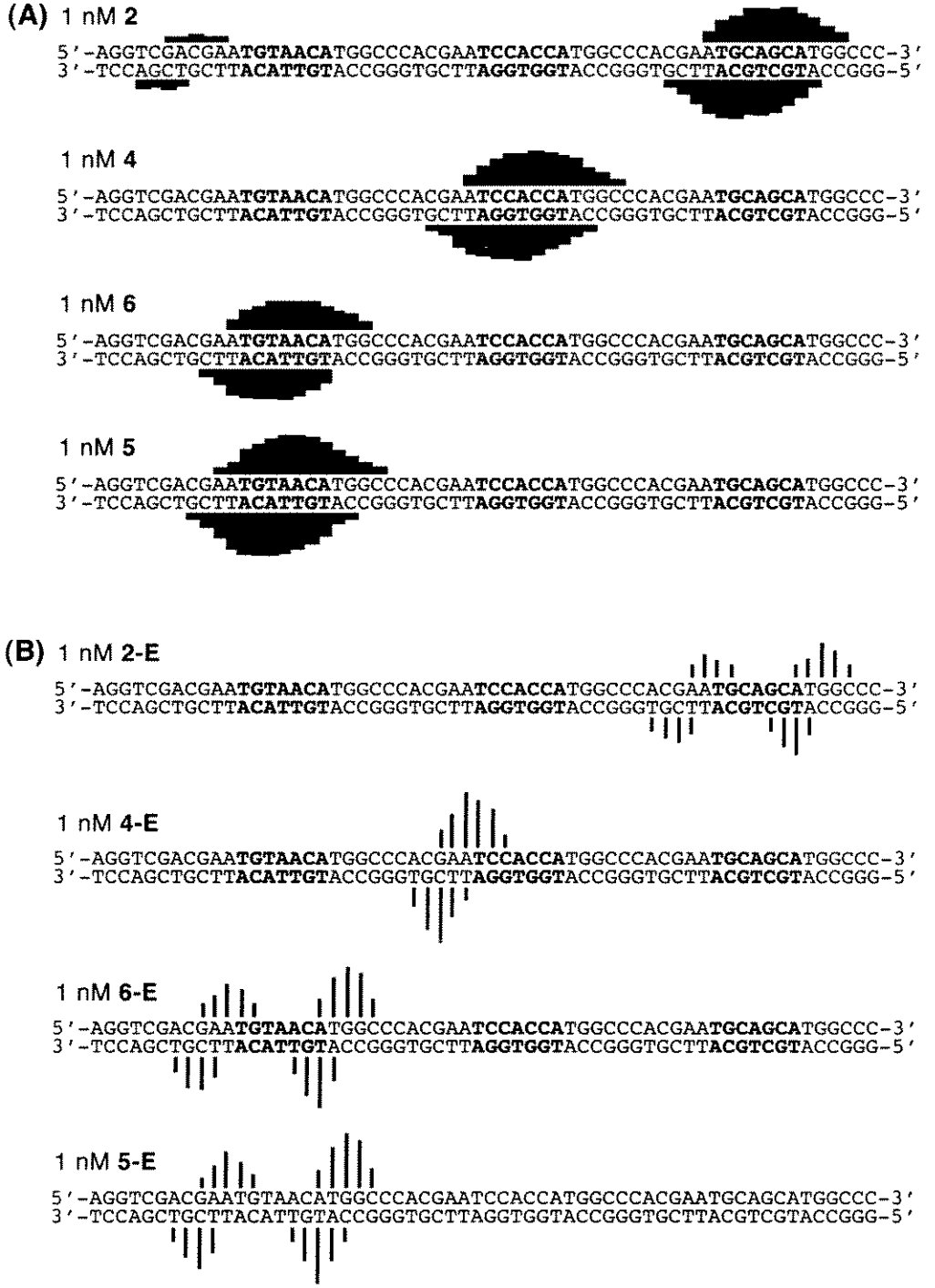


Figure 9. (A) Results of MPE•Fe(II) footprinting experiments: binding of respective 7 bp target sequences by polyamides **2**, **4**, **5**, and **6**. Bar heights are proportional to the extent of cleavage protection at the indicated base. (B) Results of affinity cleavage experiments with polyamides **2-E**, **4-E**, **5-E**, and **6-E**. Line heights are proportional to the extent of cleavage at the indicated base.

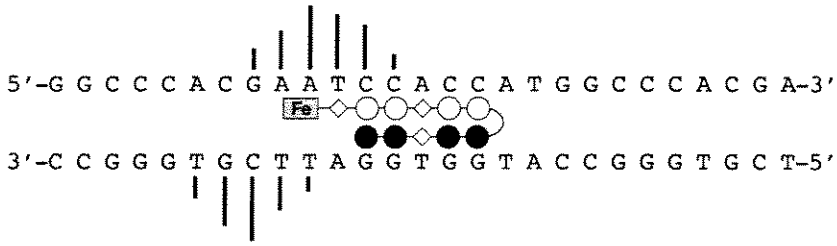


Figure 10. Polyamide **4-E** prefers the "reverse" binding orientaion. The affinity cleavage results for 1 nM **4-E** at the site 5'-caTGGTGGAtt-3' are shown together with a model of polyamide-DNA complex. Line heights are proportional to the extent of cleavage at the indicated base.

formation of 3' to 5'-oriented polyamide-DNA complexes, and suggest that such "reversed orientation" complexes typically have affinities ~10-fold lower than analogous 5' to 3'-oriented complexes.[32] Consistent with this trend, we find here that polyamide **2** binds the "reversed orientation" match site 5'-**TCGTCG**A-3' with an affinity about 10-fold lower than it binds the site 5'-**TGCTGC**A-3' (see footnote to Table 1). Inconsistent with this trend, however, the asymmetric polyamide **4-E** produces a cleavage pattern consistent with the polyamide binding preferentially in the reversed orientation (3' to 5', N- to C-terminus) (Fig. 10).

Effect of Model Intracellular Solution Conditions. Quantitative DNase I footprinting experiments indicate that increasing the equilibration temperature from 24 °C to 37 °C, and changing the solution conditions from standard polyamide assay conditions (10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0 at 24 °C) to conditions modeling those within a typical mammalian cell[33] (140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 1 mM spermine, pH 7.2) has very little effect on polyamide binding affinity (Table 2).

Implications for Polyamide Design. Cell-permeable, sequence-specific DNA-binding ligands have the potential to regulate gene expression *in vivo*. For this approach to be generally useful, a class of ligands is needed which can target *predetermined*

Table 2. Equilibrium association constants (M^{-1}) of polyamide **2** for its match site 5'-TGCTGCA-3'.[a]

Buffer[b]	Temp.	5'-aTGCTGCAt-3'
A	24 °C	1.5×10^{10} (0.3)
	37 °C	8.4×10^{9} (2.3)
B	24 °C	1.9×10^{10} (0.6)
	37 °C	1.1×10^{10} (0.2)

[a]Values reported are the mean values from at least three DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses. [b]Buffer A: 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0 at 24 °C; Buffer B: 10 mM HEPES•HCl, 140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 1 mM spermine, pH 7.2.

DNA sequences with high affinity and specificity. We report here the development of a small molecule that binds with subnanomolar affinity to a conserved 7 base pair sequence within the HIV-1 promoter using a novel polyamide motif, and demonstrate the unique ability of this motif to target certain G,C-rich sequences including the HIV-1 target sequence. These results demonstrate the ability of polyamides to bind predetermined DNA sequences with high affinity and specificity. Finally, these results provide a basis for investigation of the ability of polyamides targeted adjacent to a TATA box to inhibit RNA Polymerase II-mediated transcription and to block viral replication. The results of transcription inhibition and viral inhibition experiments with polyamide **2** are highlighted in Chapter Eleven.

Experimental Section

Materials. *E. coli* XL-1 Blue competent cells were obtained from Stratagene. Restriction endonucleases were purchased from Boeringher-Mannheim or New England Biolabs. Plasmid isolation kits were obtained from Promega. Plasmid sequencing was carried at the Sequence Analysis Facility at the California Institute of Technology.

Sequenase (version 2.0) was obtained from United States Biochemical, and DNase I (FPLCpure) was obtained from Pharmacia. [α-^{32}P]-Thymidine-5'-triphosphate (≥ 3000 C_i/mmol), [α-^{32}P]-deoxyadenosine-5'-triphosphate (≥ 6000 C_i/mmol), and [γ-^{32}P]-adenosine-5'-triphosphate were purchased from Du Pont/NEN. Water was obtained from a Millipore Milli-Q water purification system.

Synthesis. Polyamides were prepared by stepwise solid-phase synthesis using BOC-protected monomers as described.[25]

Preparation of Plasmids and ^{32}P-labeled DNA. Plasmid pJT2B2 was prepared by hybridizing the complementary oligonucleotides, 5'-CCGGCTTAAGTTCGTGGGCCATGCTGCATTCGTGGGCCATGGTGGATTCGTGG GCCATGTTACATTCG-3' and 5'-TCGACGAATGTAACATGGCCCACGAATCCAC-CATGGCCCACGAATGCAGCATGGCCCACGAACTTAAG-3', and ligating the resulting duplexes to the large pUC19 *Ava* I/*Sal* I restriction fragment. The plasmid was transformed into *E.coli.* and plasmid DNA isolated using standard methods, and the sequence of the insert confirmed by sequencing. The 3'-^{32}P end-labeled *Eco* RI/*Pvu* II restriction fragment was prepared by simultaneous digestion with *Eco* RI and *Pvu* II and 3'-fill-in using Sequenase, [α-^{32}P]-deoxyadenosine-5'-triphosphate, and [α-^{32}P]-thymidine-5'-triphosphate. The 282 bp fragment was isolated by nondenaturing gel electrophoresis. The 5'-^{32}P-end-labeled *EcoRI*/*Pvu*II fragment was prepared using standard methods. A-specific chemical sequencing was carried out as described. Standard methods were used for all DNA manipulations.

Quantitative DNase I Footprinting. Equilibrium association constants for polyamide-DNA complexes were determined by quantitative DNase I footprint titration experiments.[26] Reactions were carried out in a total volume of 400 μL in the absence of carrier DNA using exactly the procedure described in Chapter Four. For hairpin polyamides **1**-**6**, binding isotherms were adequately fit by Langmuir isotherms (eq. 2 in Chapter Four, n=1), consistent with formation of a 1:1 polyamide-DNA complexes. For

unlinked polyamides **7-9** binding to their match sequences, binding isotherms were adequately fit by a cooperative isotherm (eq. 2 in Chapter Four, n=2) consistent with cooperative dimeric binding. Reported association constants are the average value obtained from at least three independent footprinting experiments.

MPE•Fe(II) Footprinting and Affinity Cleavage. Exact binding sites were determined by MPE•Fe(II) footprinting.[30] Binding orientation was probed using affinity cleavage experiments.[31] Footprinting and affinity cleavage experiments were carried out using the procedures described in Chapter Four. A stock solution of polyamide (for MPE•Fe(II) footprinting), polyamide-EDTA (for affinity cleavage), or H_2O (for reference lanes) was added to a solution containing 3'- or 5'-end-labeled restriction fragment (15,000 cpm) affording final solution conditions of 20 mM HEPES, 200 mM NaCl, 50 μg/mL glycogen, and pH 7.3, and the solution allowed to equilibrate for 12 hr at 24 °C. Next, for MPE•Fe(II) footprinting reactions, MPE•Fe(II) was added to a final concentration of 0.5 μM and the solution allowed to equilibrate for 10 min. (a 5 μM MPE•Fe(II) stock solution was prepared by mixing equal volumes of 10 μM MPE and freshly prepared 10 μM $Fe(NH_4)_2(SO_4)_2$), and for affinity cleavage reactions, freshly prepared $Fe(NH_4)_2(SO_4)_2$ was added to a final concentration of 1 μM and the solution allowed to equilibrate for 30 min. For both MPE•Fe(II) footprinting and affinity cleaving, cleavage was initiated by the addition of dithiothreitol to a final concentration of 5 mM and allowed to proceed for 30 min at 24 °C, then stopped by adding 1 mL ethanol. Next, 10 μL of a solution containing calf thymus DNA (140 μM base-pair) (Pharmacia) and glycogen (Boehringer-Mannheim) (2.8 mg/mL) was added, and the DNA precipitated. The reactions were resuspended in 1X TBE/80% formamide, heated at 85 °C for 10 min, placed on ice, and the reaction products separated by electrophoresis on an 8% polyacrylamide gel (5% cross-link, 7 M urea) in 1X TBE at 2000 V. Gels were dried and imaged as described above. Relative cleavage intensities were determined by volume integration of individual cleavage bands using ImageQuant software (Molecular Dynamics).

References

(1) Ho, S.N.; Boyer, S.H.; Schreiber, S.L.; Danishefsky, S.J.; Crabtree, G.R. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 9203–9207.

(2) Liu, C.; Smith, B.M.; Ajito, K.; Komatsu, H.; Gomez-Paloma, L.; Li, T.H.; Theodorakis, E.A.; Nicolaou, K.C.; Vogt, P.K. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 940–944.

(3) Bianchi, N.; Passadore, M.; Rutigliano, C.; Feriotto, G.; Mischiati, C.; Gambari, R. *Biochem. Pharm.* **1996**, *52*, 1489–1498.

(4) Bellorini, M.; Moncollin, V.; D'Incalci, M.D.; Mongelli, N.; Mantovani, R. *Nucleic Acids Res.* **23**, 1657–1663.

(5) Gottesfeld, J.M.; Neely, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1997**, *387*, 202–205.

(6) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1996**, *382*, 559–561.

(7) Turner, J.M.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, *119*, 7636–7644.

(8) Goodrich, L.A.; Tjian, R. *Curr. Op. Cell Biol.* **1994**, *6*, 403–409.

(9) Kim, J.L.; Nikolev, D.B.; Burley, S.K. *Nature* **1993**, *365*, 520–527.

(10) Jones, K.A.; Peterlin, B.M. *Ann. Rev. Biochem.* **1994**, *63*, 717–743.

(11) Frech, K.; Brack-Werner, R.; Werner, T. *Virology* **1996**, *224*, 256–267.

(12) Coffin, J.M. *Science* **1995**, *267*, 483–489.

(13) Wade, W.S.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1992**, *114*, 8783–8794.

(14) Mrksich, M.; Wade, W.S.; Dwyer, T.J.; Geirstanger, B.H.; Wemmer, D.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586–7590.

(15) Wade, W.S.; Mrksich, M.; Dervan, P.B. *Biochemistry* **1993**, *32*, 11385–11389.

(16) White, S.; Baird, E.E.; Dervan, P.B. *Chem. & Biol.* **1997**, *4*, 569–578.

(17) Pelton, J.G.; Wemmer, D.E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723–5727.

(18) Pelton, J.G.; Wemmer, D.E. *J. Am. Chem. Soc.* **1990**, *112*, 1393–1399.

(19) White, S.; Baird, E.E.; Dervan, P.B. *Biochemistry* **1996**, *35*, 12532–12537.

(20) Trauger, J.W.; Baird, E.E.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6160–6166.

(21) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *Chem. Eur. J.* **1997**, *3*, 1600–1607.

(22) Turner, J.M.; Swalley, S.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1998**, in press.

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

(24) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Chem. Biol.* **1996**, *3*, 369–377.

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

(26) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods in Enzymology* **1986**, *130*, 132–181.

(27) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, *119*, 6953–6961.

(28) Parks, M.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6153–6159.

(29) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *118*, 8198–8206.

(30) Van Dyke, M.W.; Dervan, P.B. *Biochemistry* **1983**, *22*, 2373–2377.

(31) Taylor, J.S.; Schultz, P.G.; Dervan, P.B. *Tetrahedron* **1984**, *40*, 457–465.

(32) White, S.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 8756–8765.

(33) Jones, R.J.; Lin, K.Y.; Milligan, J.F.; Wadwani, S.; Matteucci, M.D. *J. Org. Chem.* **1993**, *58*, 2983–2991.

CHAPTER SIX

Comparison of Aliphatic/Aromatic and Aliphatic/Aliphatic Pairings in Hairpin Polyamides that Recognize 6–7 Base Pair DNA Sequences

The text of this chapter is taken from a manuscript that was coauthored with

Prof. Peter B. Dervan and Eldon E. Baird

Introduction

Pyrrole-imidazole polyamides are synthetic ligands that bind DNA by forming side-by-side complexes in the minor groove. Sequence specificity depends on side-by-side amino acid pairings.[1-9] A pairing of *N*-methylimidazole (Im) opposite *N*-methylpyrrole (Py) recognizes a G•C base pair and Py/Im recognizes C•G,[1-5] A Py/Py combination recognizes T•A and A•T.[6-8] Recently, it has been found that the pairing of *N*-methyl-3-hydroxypyrrole opposite Py is specific T•A base pairs.[9]

The pairing rules described above function in the context of a polyamide motif that correctly positions the Im and Py residues. The first motif for DNA recognition by polyamides consisted of three-ring polyamides that bind 5 base pair sites as antiparallel dimers.[1-3] Subsequently, the polyamide ImPyPy-γ-PyPyPy-Dp (Dp = dimethylaminopropylamide) containing a "turn" amino acid γ-aminobutyric acid (γ) was shown to bind the target site 5'-TGTTA-3' in a "hairpin" conformation with an equilibrium association constant, K_a, of 8×10^7 M^{-1}, an increase of 300-fold relative to unlinked three-ring polyamide dimers.[10] Eight-ring and ten-ring hairpin polyamides can bind 6–7 base pair target sequences at subnanomolar concentrations and discriminate match from single-base pair mismatch sites.[11,12]

To be generally useful as regulators of gene expression, polyamides must be capable of recognizing a broad range of DNA sequences with high specificity and subnanomolar affinity.[11,13] However, hairpin polyamides do not bind all sequences with equal affinity and specificity. Models of previously characterized hairpin polyamide-DNA complexes, together with their experimentally determined association constants, are illustrated in Figure 1.[14-20] The boxed complexes bind with optimal affinity (>10^7 M^{-1} for six-ring hairpins, and >10^9 M^{-1} for eight- and ten-ring hairpins), while unboxed complexes have lower affinity. For six-, eight- and ten-ring hairpins, a common feature of those having suboptimal DNA-binding affinity is the presence of one or more imidazole residues

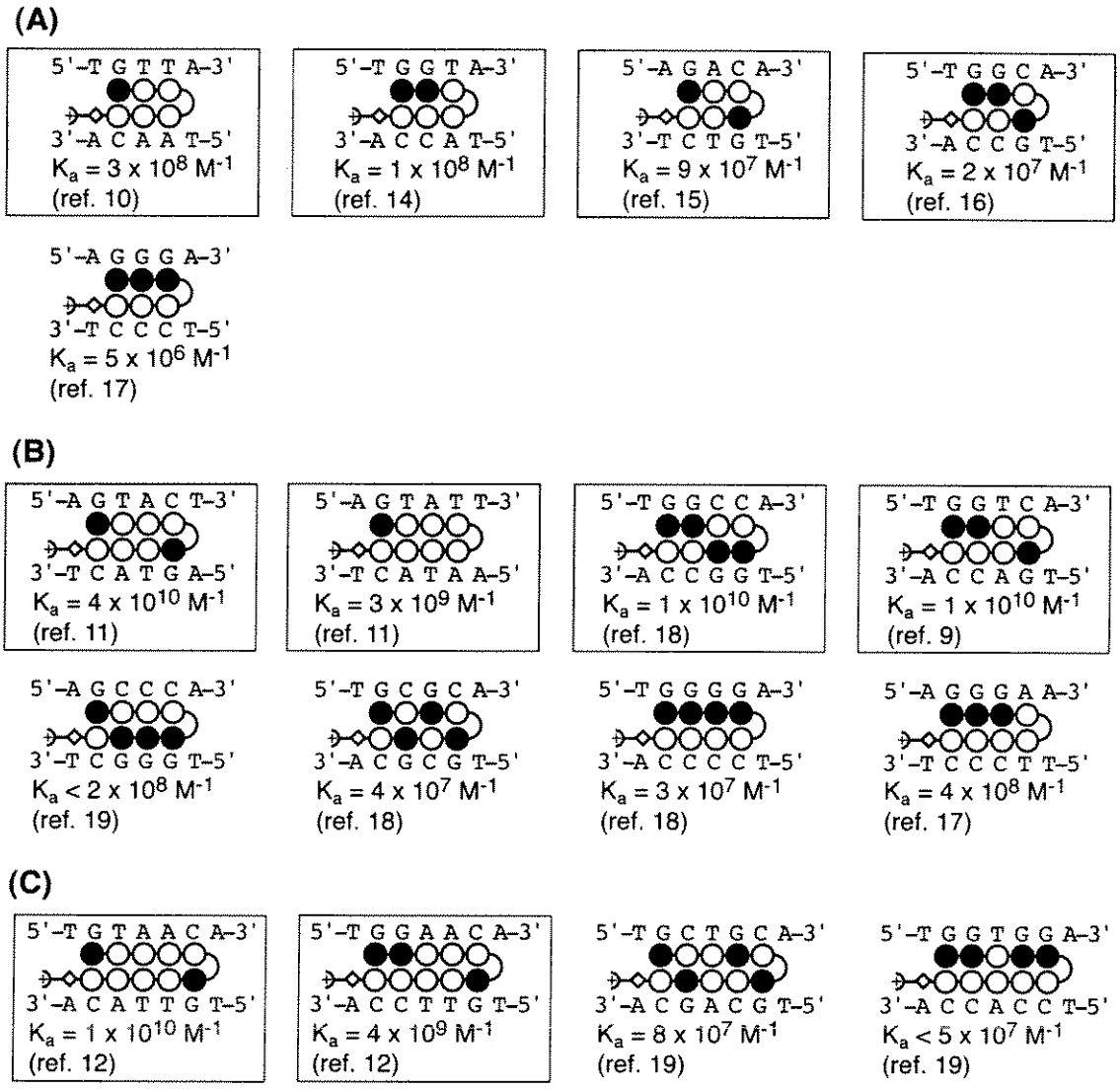


Figure 1. Models of previously characterized polyamide-DNA complexes. (A) six-ring, (B) eight-ring, and (C) ten-ring hairpin polyamide-DNA complexes. The experimentally determined equilibrium association constants are indicated for each complex.

beyond the second position of a polyamide subunit, counting from the N-terminal end (a "polyamide subunit" is defined as one side of a hairpin polyamide; a hairpin thus consists of two subunits connected by γ-aminobutyric acid). These results suggest that "internal imidazoles," i.e., imidazole residues beyond the second position of a polyamide subunit, are not correctly positioned to hydrogen bond with their target guanine bases.[18]

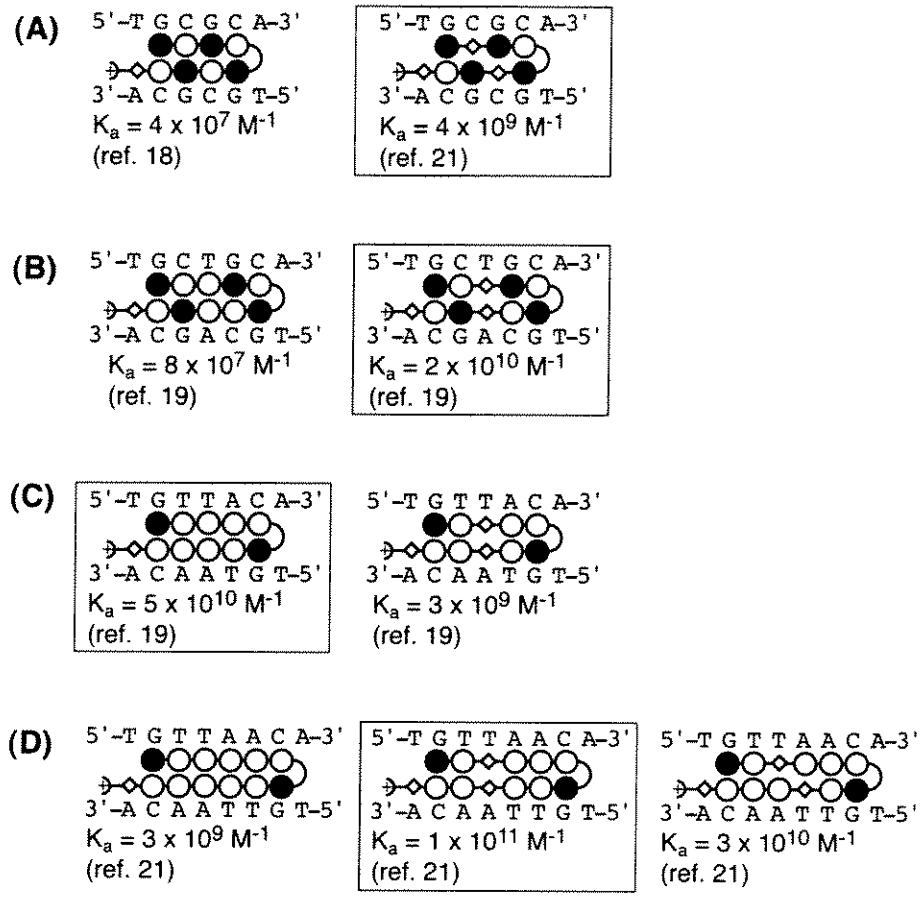


Figure 2. Replacement of pyrrole residues with β-alanine can improve the DNA-binding affinity of polyamides.

It has been found that β-alanine residues[20] can be used to reposition internal imidazoles. For example, the polyamide Im-β-ImPy-γ-Im-β-ImPy-β-Dp binds its match site 5'-TGCGCA-3' with K_a = 4 x 10^9 M^{-1}, an increase of 100-fold compared to ImPyImPy-γ-ImPyImPy-β-Dp (Fig. 2A).[21] Similarly, the polyamide ImPy-β-ImPy-γ-ImPy-β-ImPy-β-Dp binds its match site 5'-TGCTGCA-3' with K_a = 2 x 10^{10} M^{-1}, an increase of more than 100-fold compared to ImPyPyImPy-γ-ImPyPyImPy-β-Dp (Fig. 2B).[19] The polyamide ImPy-β-PyPy-γ-ImPy-β-PyPy-β-Dp, which does not have any internal imidazole residues, binds its match site 5'-TGTTACA-3' with 20-fold lower affinity than ImPyPyPyPy-γ-ImPyPyPyPy-β-Dp (Fig. 2C).[19]

Results obtained to date suggest that polyamides in which β-alanine residues are paired generally have higher DNA-binding specificity than analogous polyamides in which the β-alanines are unpaired. For example, in the series of "ten-ring" polyamides shown in Fig. 2D, the compound having the paired β-alanines (center) is 55-fold specific for a single-base pair mismatch site, while the compound with two β/Py pairs (right) is only 5-fold specific versus the mismatch site.[21] The β/β pairing is specific for A,T versus G,C base pairs.[20]

The results described above for eight- and ten-ring hairpins indicate that, for polyamide subunits that contain one or more internal imidazole residues, the imidazole will be correctly positioned only if preceded by β-alanine (see Figs. 2A and 2B). However, for polyamide subunits that do not have any internal imidazole residues, β-alanine substitutions can decrease binding affinity (see Fig. 2C). The question arises for unsymmetrical hairpins in which one subunit has an internal imidazole and requires β-alanine, while the other subunit does not, whether a β/Py or a β/β pairing is optimal. To address this question, the present study investigates hairpin polyamides **1–10**, which target four different 6–7 base pair DNA sequences, to compare Py/β and β/β pairings (Fig. 3). We note that these polyamides were synthesized to bind proximal to transcription factor binding sites within the CMV major-immediate early promoter (**1–3**), the Vκ gene promoter (**4–6**), the HER2/neu gene promoter (**7** and **8**) and the ribonucleotide reductase gene promoter (**9** and **10**). A preliminary investigation of the ability of polyamide **7** to inhibit transcription of the HER2/neu gene is described in Chapter 11.

Results and Discussion

DNA-binding Affinities. Polyamides **1–10** were synthesized by solid phase methods.[22] Polyamides **1–3** target the sequence 5'-AGTGAA-3', **4–6** target 5'-TGCAAA-3', **7** and **8** target 5'-AGAATGA-3', and **9** and **10** target 5'-AGAGCCA-3'. The DNA-binding affinities of the polyamides were determined by quantitative DNase I

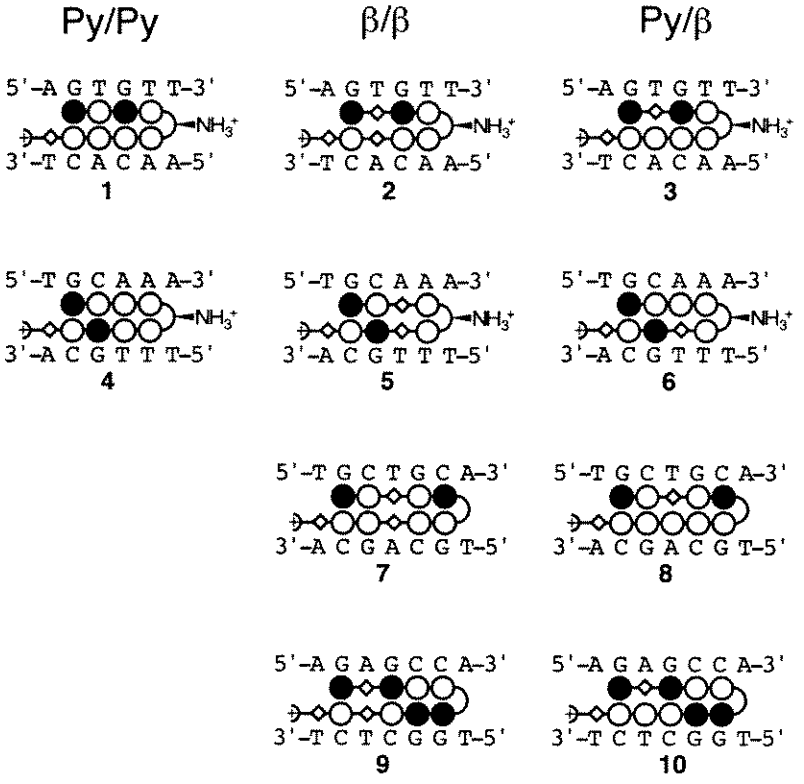


Figure 3. Models of complexes of polyamides **1–10** with their target sequences 5'-AGTGAA-3' (**1–3**), 5'-TGCAAA-3' (**4–6**), 5'-TGCTGCA-3' (**7**,**8**) and 5'-AGAGCCA-3' (**9**, **10**). Curved lines represent γ-aminobutyric acid, and curved lines with an amino group represent (*R*)-2,4-diaminobutyric acid.

footprinting experiments carried out on ^{32}P-end-labeled restriction fragments or PCR products (Figs. 4 and 5).[23]

For polyamides **1–3**, compound **3**, which has a β/Py pair, has optimal affinity (Fig. 5). Similarly, for polyamides **4–6**, compound **4** having a β/Py pair has optimal affinity (Fig. 5). The results obtained with "eight-ring" polyamides **1–6** indicate a relatively modest energetic penalty (2–10-fold reduction in affinity) for a Py/Py versus a β/Py pair, and a larger penalty (>50-fold reduction in affinity) for a β/β versus a β/Py pair. The similar DNA-binding properties of polyamides **1** and **3** indicate that, for the case of a hairpin polyamide containing a single imidazole at the third position of one subunit, a β-alanine residue is not essential in all cases. For "ten-ring" polyamides **9** and **10**, an energetic penalty (10-fold reduction in affinity) is observed for a β/β versus a β/Py pairing

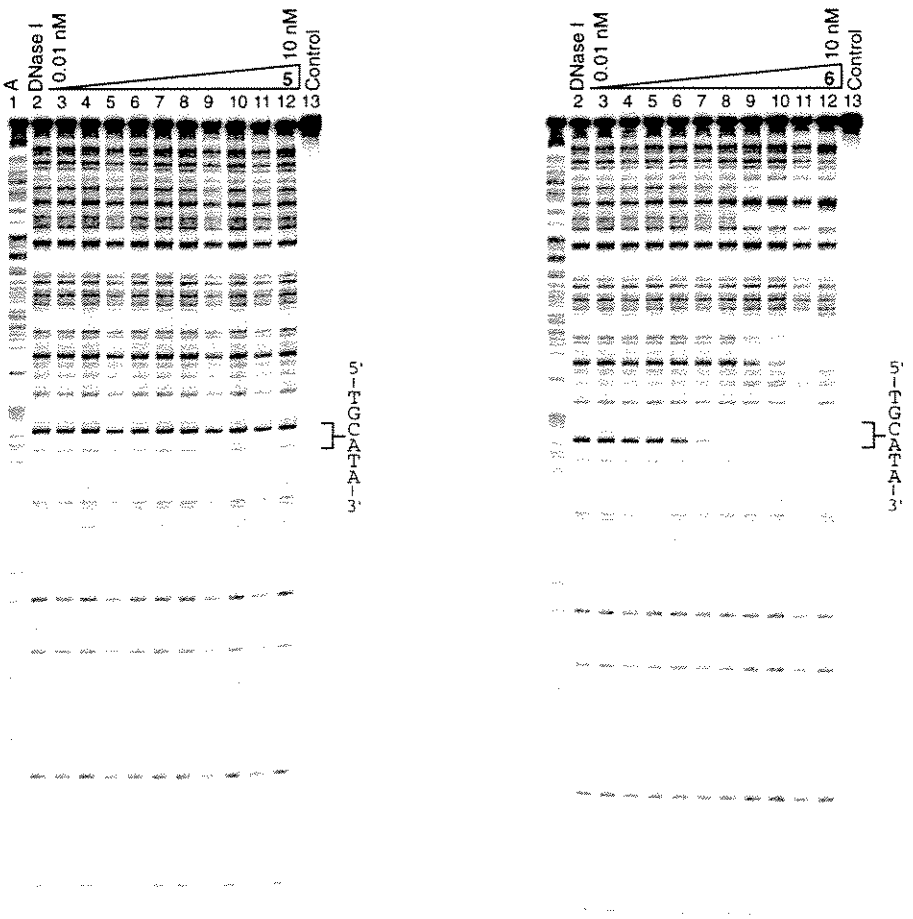


Figure 4. Repesentative quantitative DNase I footprint titration experiments. Comaparison of the binding of polyamides **5** and **6** to the Vκ gene promoter. The target sequence 5'-TGCATA-3' is indicated. All reactions contained 15 kcpm labeled restriction fragment, 10 mM Tris•HCl, 10 mM $MgCl_2$, 10 mM KCl and 5 mM $CaCl_2$.

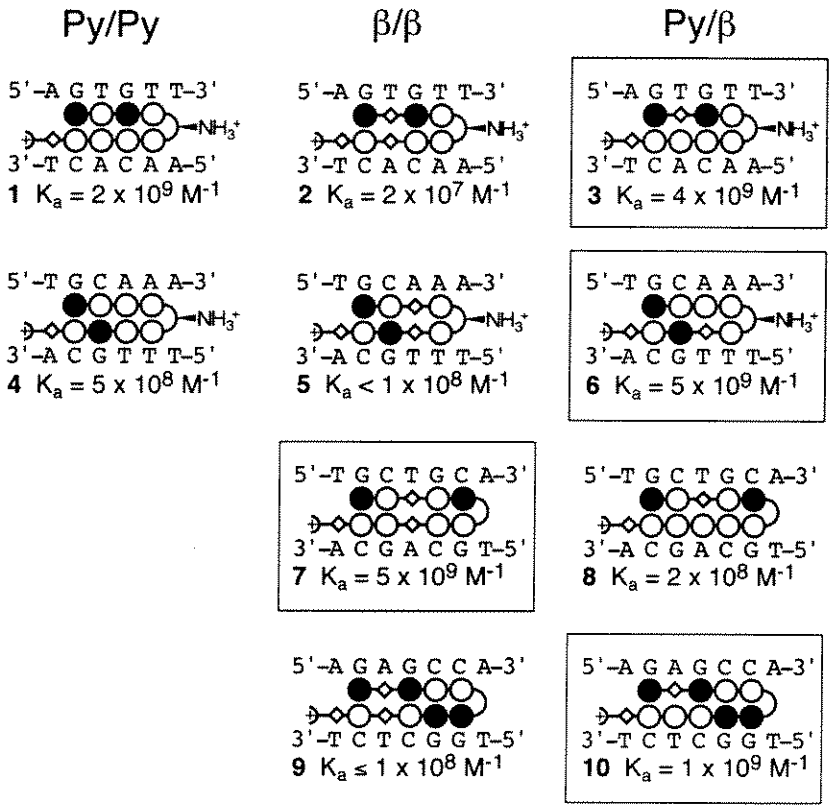


Figure 5. Equilibrium association constants (K_a) for complexes of polyamides **1**–**10** with their target sequences. For each target sequence, the polyamide-DNA complex having the highest affinity is boxed.

(Fig. 5). Although not tested, it is likely that an analog of **10** having a Py/Py pair would have lower affinity than **10**, by analogy to the results obtained with **1**–**6**. In contrast to these results, for "ten-ring" polyamides **7** and **8**, compound **7** having a β/β pair has 20-fold higher affinity than its analog **8** having a β/Py pair. Again, while not tested, it is expected that the analog of **7** having a Py/Py pair would have lower affinity than **7** (see Fig. 1C).

The results obtained with polyamides **1**–**10** can be systematized as follows: For a hairpin polyamide targeted to a 6 or 7 base pair DNA sequence in which one subunit has an internal imidazole and requires β-alanine, while the other subunit does not, either a β/Py or a β/β pairing is optimal, depending on the structure of the subunit having the internal

STEP 1: initial design using ring pairings

The N-terminal position of the hairpin must be Im.
- Im/Py targets G•C, Py/Im targets C•G, and Py/Py targets A,T base pairs.

STEP 2: classify hairpin subunits

a. For each hairpin subunit:
- if there is no Im beyond position 2, the subunit is class **A**, otherwise it is class **B**.

b. For class **B** subunits, the internal Im should be preceded if possible by β (by replacing a Py) using:
- 1-β-2, 2-β-1, 1-β-3 (class **B1**),
- or 2-β-2 (class **B2**) subunits.
- If there is no Py to replace with β, the design is not optimal and may have low DNA-binding affinity and/or specificity.

STEP 3: subunit pairing rules

For an **A/A** pairing, no β is needed.
For a **B/B** pairing, the βs should be paired if possible.
For an **A/B1** pairing, a β/Py pair should be used.
For an **A/B2** pairing, a β/β pair should be used.

Example:

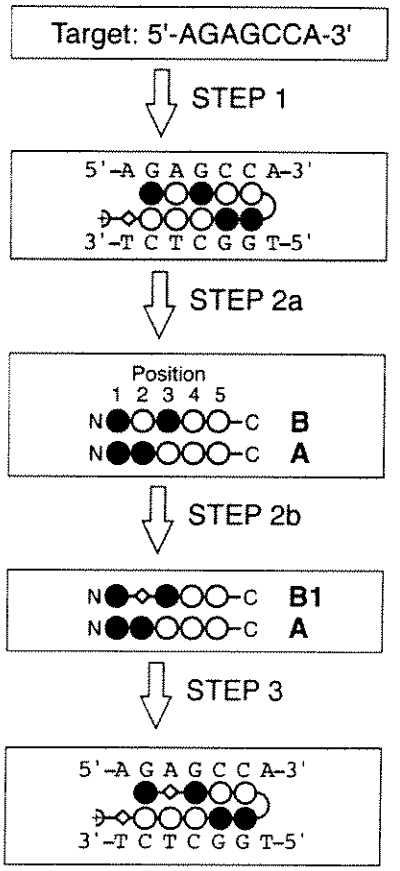


Figure 6. Design guidelines for hairpin polyamides that recognize 6–7 base pair DNA sequences.

imidazole(s). If the subunit having the internal imidazole(s) has a 1-β-x (i.e., one ring, then β-alanine, then x rings) or x-β-1 structure, a β/Py pairing is optimal; otherwise, for a 2-β-2 structure, a β/β pairing is optimal. Thus, the choice of a Py/β or β/β pair depends on the sequence composition of the target site in a predictable manner.

Analysis of the results of the present study, together with the database of previously characterized polyamide-DNA complexes (see Figs. 1 and 2), suggest guidelines that predict, for a given target sequence, when and where β-alanine residues are needed. These guidelines are presented in a systematic form in Figure 6. The key features of the

guidelines are: 1) the N-terminal position of the hairpin must be Im,[14,19] 2) imidazole residues beyond the second position of a hairpin subunit should be preceded by β-alanine (by replacing a Py),[14-20] 3) in cases where only one of the two subunits has an internal imidazole, the choice a Py/β or β/β pair depends on the sequence composition of the target site in a predictable manner, and 4) in cases where both subunits have internal imidazoles, the β-alanine residues should be paired if possible.[20] It must be emphasized that, as the database of characterized β-alanine-containing hairpins upon which the guidelines are based is relatively small, further studies are necessary to either support their generality or to identify exceptions.

Applications. The design guidelines in Figure 7 were applied to target a polyamide to the sequence 5'-AGCAGA-3' (Fig. 8), which is immediately downstream of the TATA box within the CMV major immediate-early promoter. The designed polyamide, ImPy-β-Im-Daba-Py-β-ImPy-β-Dp (Daba = (*R*)-2,4-diaminobutyric acid), was synthesized and shown by quantitative DNase I footprinting to bind the target sequence with $K_a = 3 \times 10^8\ M^{-1}$ (Fig. 9). The relatively low affinity of this polyamide likely results from the high conformational entropy due to the presence of two β-alanines within the "eight-ring" hairpin polyamide framework.

Polyamides **1–3** were originally designed to bind a 5'-AGTGTT-3' sequence within the CMV major-immediate early promoter (MIEP). Subsequently, we needed a polyamide to target a 5'-AGTGTT-3' sequence within the promoter region of the CMV polymerase gene. Based on the results for **1–3** binding to the CMV MIEP, compound **2** was selected and shown to bind the CMV polymerase gene target site with high affinity ($K_a = 2 \times 10^{10}\ M^{-1}$) and specificity (Fig. 9).

Conclusion

The present study demonstrates that, for polyamides that recognize 6–7 base pair sequences, the choice of an aliphatic/aliphatic versus an aliphatic/aromatic pairing depends

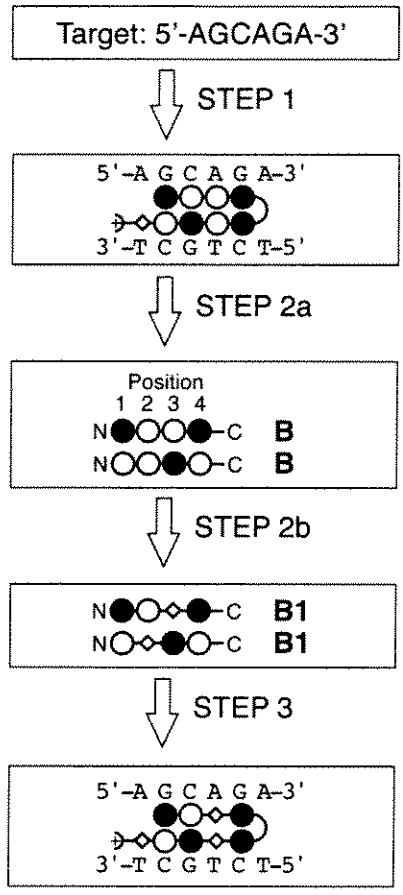


Figure 7. Application of the design rules to select a polyamide to bind the sequence 5'-AGCAGA-3', which is immediately downstream from the TATA box present within the CMV major-immediate early promoter (MIEP).

on the sequence composition of the target site. The results reported here, together with previously published results, suggest systematic guidelines for the design of new hairpins that target predetermined DNA sequences. Additional studies are needed to investigate the generality of these guidelines.

Experimental Section

Restriction Fragments. The following ^{32}P-end-labeled DNA fragments were used in this study: the pCMV-Δ517 *Hin*dIII/*Eco*RI restriction fragment for polyamides **1**–**3**, the PCR product amplified from p58-3-G using the primers VK-A (5'-ACTCTCAATC-

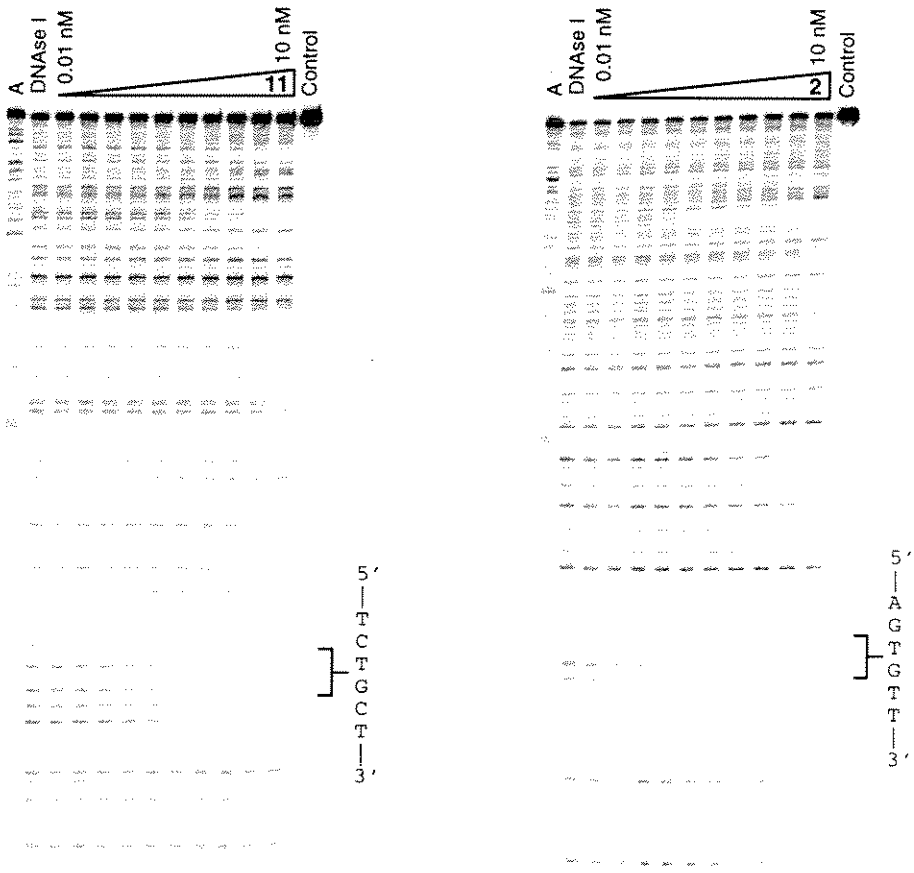


Figure 8. Quantitative DNase I footprint titration experiments for (left) the designed polyamide ImPy-β-Im-Daba-Py-β-ImPy-β-Dp (**11**) binding to the CMV MIEP and (right) polyamide Im-β-ImPy-Daba-PyPyPyPy-β-Dp (**2**) binding to the CMV polymerase gene and (Daba = (*R*)-2,4-diaminobutyric acid). All reactions contained 15 kcpm labeled restriction fragment, 10 mM Tris•HCl, 10 mM $MgCl_2$, 10 mM KCl and 5 mM $CaCl_2$.

GTCTGGGTGAA-3') and VK-B (5'-TGAGGAGACTGATCAGTCACT-3') for polyamides **4–6**, the pGEM-HNP *Hin*dIII/*Eco*RI restriction fragment for polyamides **7** and **8**, and the pSBR2a *Apa*LI/*Eco*RI restriction fragment for polyamides **9** and **10**. Restriction fragments were labeled by 3'-fill-in with [α-^{32}P]-dATP, [α-^{32}P]-dTTP, dGTP and dCTP using Sequenase. The footprinting experiments in Figure 9 were carried out using the following DNA fragments: for **2**, the pPol-CAT *Hind*III/*Sal*I fragment, and for **11**, the pCMV-Δ517 *Hin*dIII/*Eco*RI fragment.

PCR labeling protocol. The desired fragment of p58-3-G was amplified using the primers VK-A and VK-B as described below.

1. To 60 pmol primer VK-B (the primer to be labeled) add: 80 µL water, 10 µL 10x kinase buffer, 4 µL [γ-^{32}P]-ATP (~250 µC$_i$, ICN end-labeling grade) and 6 µL polynucleotide kinase. The total volume should be 100 µL. Incubate at 37 °C for 30 minutes. Then add 5 µL 0.5 M EDTA and phenol/$CHCl_3$ extract (4 x 100 µL). Purify the DNA using a NICK column and EtOH precipitate (to the 400 µL from the NICK column add 24 µL 4M NaCl, 1 µL glycogen (20 mg/mL), mix, and spin for 30 minutes in a cold room. Decant, wash with 100 µL 75% EtOH, decant and dry in a speed-vac.

2. Dissolve 60 pmol primer VK-A in 60 µL water. To the pellet from step 3, add 50 µL of the primer VK-A solution, 33 µL water, 10 µL PCR buffer (B.M.), 3.7 µL plasmid DNA (0.003 µg/mL), 2 µL dNTP mix (each at 10 mM), and 1 µL 100x BSA (New England Biolabs). Transfer the mixture to a 0.7 mL tube.

3. Heat the tube at 70 °C in a thermocycler for 5 minutes ("hot-start PCR"). Carefully add 4 units (0.8 µL) Taq DNA Polymerase (B.M.) and mix by tapping the tube. Then add oil (enough to at least cover the solution).

4. Thermocycle: 30 cycles of 1) 94 °C, 1 minute, 2) 54 °C, 1 minute, 3) 72 °C, 1.5 minutes. Then heat at 72 °C for 10 minutes to insure that all ends are filled in completely.

5. Carefully transfer the aqueous phase to a new tube (wipe the pipet tip with a KimWipe to get the oil off). For each 100 µL of solution, add 30 µL 5x Ficoll loading buffer. Load onto a 7% non-denaturing gel.

6. Run the gel at 180–250 V for 1–2 hours. Excise the desired band using a razor blade and transfer the gel slice to a 1.5 mL tube. Crush the gel slice with a yellow pipet tip and add 700 µL elution buffer (25 mM Tris•HCl, 250 mM NaCl, 1 mM EDTA, pH 8.0). Place the tube in a shielded container and soak overnight (8–12 hours) in a 37 °C shaker. Then filter the DNA using a plastic centrifugal filter into a Falcon 2059 tube, and transfer the eluate to a 1.5 mL tube. Add 1.5 volumes isopropanol, mix by inversion, and precipitate by spinning 30 minutes (14,000 rpm) in a cold room. Decant, then add 100 µL 75% EtOH, spin briefly (30–90 seconds), decant, and dry the DNA pellet in a speed vac (dry to 700 mtorr, about 2 minutes).

7. Dissolve the DNA in 100 µL RNase-free water (U.S.B.) and put it through a NICK column (use RNase-free water to equilibrate the NICK column; the label may degrade more rapidly if stored in Milli-Q water). Divide the label into 4–6 aliquats, count and store at –80 °C until use.

Quantitative DNase I footprinting. Footprinting was carried out as described in Chapter Four.

References

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

(2) Mrksich, M.; Wade, W.S.; Dwyer, T.J.; Geirstanger, B.H.; Wemmer, D.E.; Dervan, P.B. (1992). *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586–7590.

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

(4) (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*, 3325–3332.

(5) White, S.; Baird, E.E.; Dervan, P.B. *Chem. Biol.* **1997**, *4*, 569–578.

(6) Pelton, J.G.; Wemmer, D.E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723–5727.

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

(8) White, S.; Baird, E.E.; Dervan, P.B. (1996). *Biochemistry* **1996**, *35*, 12532–12537.

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

(10) (a) Mrksich, M.; Parks, M.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 7983–7988. (b) Parks, M.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6147.

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

(12) Turner, J.M.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, *119*, 7636–7644.

(13) Gottesfeld, J.M.; Neely, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1997**, *387*, 202–205.

(14) Parks, M.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6153–6159.

(15) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Chem. & Biol.* **1996**, *3*, 369–377.

(16) Swalley, S.E.; Baird, E.E.; Dervan, P.B. Unpublished results.

(17) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 8198–8206.

(18) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1998**, *119*, 6953–6961.

(19) Trauger, J.W.; Baird, E.E.; Dervan, P.B. Unpublished results.

(20) Turner, J.M.; Swalley, S.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1998**, in press.

(21) (a) Trauger, J.W.; Baird, E.E.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6160–6166. (b) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *Chem. Eur. J.* **1997**, *3*, 1600–1607.

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

(23) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods in Enzymology* **1986**, *130*, 132–181.

CHAPTER SEVEN

Recognition of 16 Base Pairs in the Minor Groove of DNA by a Polyamide Dimer

The text of this chapter is taken from a published paper that was coauthored with

Prof. Peter B. Dervan and Eldon E. Baird

(Trauger, J.W.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1998**, *120*, 3534–3535)

Cell-permeable small molecules that bind predetermined DNA sequences with affinity and specificity comparable to natural DNA-binding proteins have the potential to regulate the expression of specific genes. Recently, an 8-ring hairpin Py-Im polyamide that binds 6 base pairs of DNA was shown to inhibit transcription of a specific gene in cell culture.[1] Polyamides recognizing longer DNA sequences should provide more specific biological activity. To specify a single site within the 3 billion base pair human genome, ligands that specifically recognize 15–16 base pairs are necessary.[2] For this reason, recognition of 16 base pairs represents a milestone in the development of chemical approaches to DNA recognition.[2,3] We examine here the affinity and specificity of a Py-Im polyamide dimer that targets 16 contiguous base pairs in the minor groove of DNA.[4]

As the length of a polyamide dimer having the general sequence $ImPy_{2\text{-}6}$ increases beyond 5 rings (corresponding to a 7 base pair binding site), the DNA-binding affinity ceases to increase with polyamide length.[5] A structural basis for this observation is provided by a recently determined X-ray crystal structure of a polyamide-DNA complex, which reveals a perfect match of polyamide rise-per-residue with the pitch of the DNA duplex, but overwound ligand curvature.[6] The curvature mismatch explains the observation that flexible β-alanine residues reset an optimum fit of polyamide dimers with the DNA helix at long binding sites.[7]

We chose as our binding site the 16 base pair sequence 5'-ATAAGCAGCTGCTTTT-3' present in the regulatory region of the HIV-1 genome.[8] Consideration of the previously published polyamide ring pairing rules,[9,10] the A,T specificity of β/β pairs,[7] and the "slipped" dimer motif[7a,11] suggested that the 8-ring polyamide ImPy-β-ImPy-β-ImPy-β-PyPy-β-Dp (**1**) would specifically bind the target sequence as a cooperative antiparallel dimer (Fig. 1). The polyamide was synthesized using solid-phase methods,[12] purified by HPLC, and its identity and purity confirmed by ^{1}H NMR, analytical HPLC and MALDI-TOF MS. A quantitative DNase I footprinting experiment[13] carried out on a 245 base pair 3'-^{32}P-end-labeled restriction fragment revealed

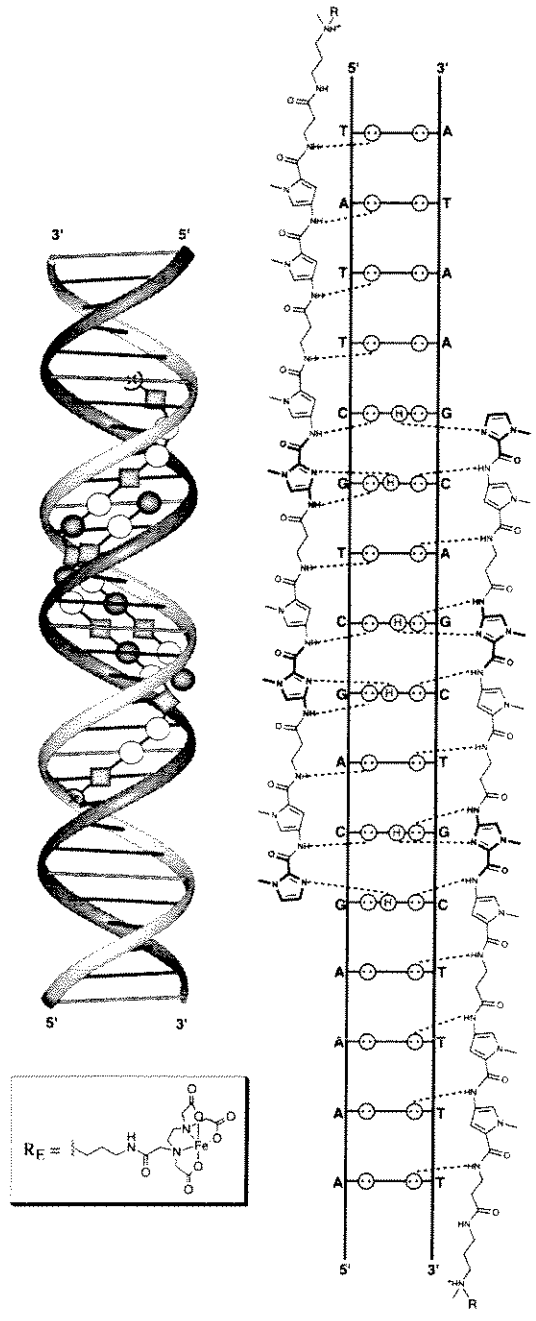


Figure 1. Model of the complex of ImPy-β-ImPy-β-ImPy-β-PyPy-β-Dp (**1**, R = Me) or ImPy-β-ImPy-β-ImPy-β-PyPy-β-Dp-EDTA•Fe(II) (**1-E**, R = R_E) (Im = *N*-methylimidazole, Py = *N*-methylpyrrole, β = β-alanine, Dp = dimethylaminopropylamide) with 5'-ATAAGCAGCTGCTTTT-3'. The shaded and open circles represent imidazole and pyrrole rings, respectively, and the diamonds represents β-alanine. Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines, and circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dashed lines.

that the polyamide specifically binds it target site at subnanomolar concentrations (apparent monomeric association constant, $K_a \geq 3.5 \times 10^{10}\ M^{-1}$) (Fig. 2).[14-15] The binding data were well-fit by a cooperative binding isotherm, consistent with formation of a cooperative 2:1 complex.[14] To provide further evidence that **1** binds as an extended dimer, an affinity cleavage experiment[2a] was carried out with the polyamide-EDTA•Fe(II) conjugate **1-E** (Fig. 2). Cleavage was observed at each end of the match sequence, consistent with a dimeric, antiparallel binding mode. With regard to sequence specificity, there is a proximal two-base pair mismatch site, 5'-cAGATGCTGCATATa-3', to the 5' side of the ^{32}P-labeled strand that is bound with at least 35-fold lower affinity than the match site. However, other mismatch sites on the restriction fragment are bound with 10–20-fold lower affinity, revealing limitations of this first effort at 16 base pair recognition. Undoubtedly there is ample room for further optimization of sequence specificity.

The results reported here have two implications. First, the high binding affinity and the affinity cleavage pattern observed for the 16 base pair polyamide•DNA complex suggests that 8 pairs of amide residues form a fully overlapped core that properly positions the 6 Im/Py pairs for recognition of 6 G,C base pairs and 2 β/β pairs for recognition of 2 A,T base pairs. Polyamides composed of two-ring subunits connected by β-alanine appear to be isohelical with B-DNA, and allow placement of imidazole residues at any ring position, thus providing a generalizable motif for recognition of predetermined DNA sequences. Importantly, these results demonstrate that polyamides of similar size to those shown to permeate cells (i.e., MW ~ 1,200) can now bind 16 base pairs of DNA at subnanomolar concentrations, paving the way for investigation of the optimal polyamide binding site size required for specific biological activity. However, the specificity is sub-optimal and this is likely a minimum step forward. This data provides a baseline for comparison with other cooperative binding motifs which combine minimum polyamide size (for cell-permeation) with maximum DNA sequence size (for biological specificity) that will be reported in due course.[16]

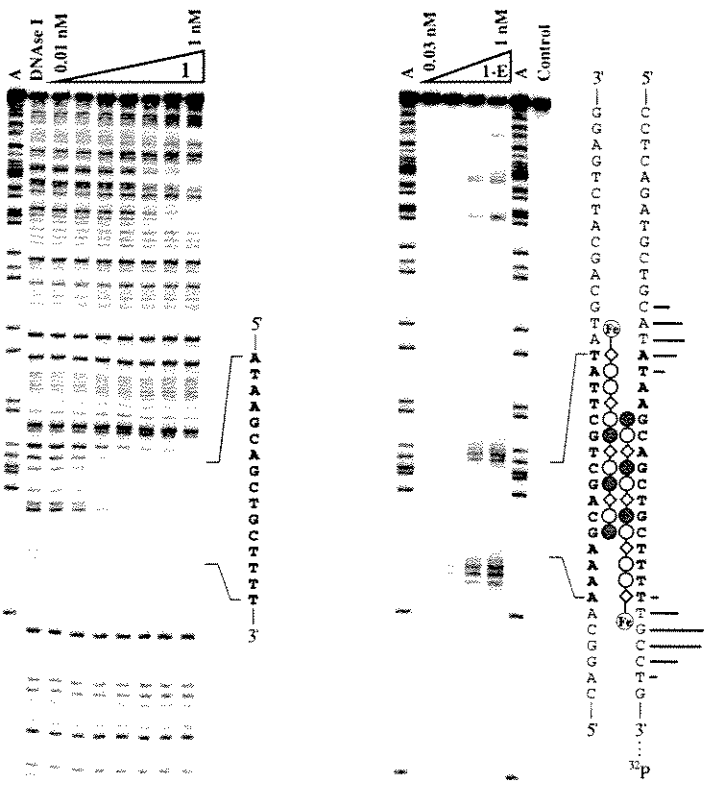


Figure 2. (Left) Storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel used to separate the fragments generated by DNase I digestion in a quantitative footprint titration experiment with polyamide **1**: lane 1, A lane; lane 2, DNase I digestion products obtained in the absence of polyamide; lanes 3–10, DNase I digestion products obtained in the presence of 0.01, 0.02, 0.05, 0.1, 0.2, 0.5 and 1 nM polyamide **1**, respectively; lane 11, intact DNA. All reactions contain 3'-^{32}P-end-labeled *Eco*RI/*Hin*dIII restriction fragment from plasmid pJT-LTR (15 kcpm), 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$ (pH 7.0, 24 °C). (Right) Autoradiogram of a gel used to separate the fragments generated by an affinity cleavage reaction using polyamide **1-E**. Lanes 1 and 5: A sequencing lanes; lanes 2–4: cleavage products obtained in the presence of 0.03, 0.1, 0.3 and 1 nM **1-E**, respectively; lane 6: intact DNA. All reactions contain labeled restriction fragment (7 kcpm), 20 mM HEPES, 300 mM NaCl, 50 μg/mL glycogen, 1 μM Fe(II) and 5 mM DTT (pH 7.3, 24 °C). The sequence of the restriction fragment in the region of the 16 base pair target site and a model of the $(\mathbf{1\text{-}E})_2$•DNA complex are shown along the right side of the autoradiogram. Line heights are proportional to the observed cleavage intensity at the indicated base.

References and Notes

(1) Gottesfeld, J.M.; Neely, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1997**, *387*, 202–205.

(2) (a) Dervan, P.B. *Science* **1986**, *232*, 464. (b) Dervan, P.B. In *The Robert A. Welch Foundation Conference on Chemical Research XXXI. Design of Enzymes and Enzyme Models*; Houston, Texas, November 2–4, 1987; pp 93–109. (c) Dervan, P.B. In *Nucleic Acids and Molecular Biology*, Vol. 2; Springer-Verlag: Heidelberg, 1988; pp 49–64.

(3) (a) Moser, H.E.; Dervan, P.B. *Science* **1987**, 238, 645–650. (b) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.L.; Thoung, N.T.; Lhomme, J.; Helene, C. *Nucleic Acids Res.* **1987**, *15*, 7749. (c) Strobel, S.A.; Doucettestamm, L.A.; Riba, L.; Housman, D.E.; Dervan, P.B. *Science* **1991**, *254*, 1639–1642. (d) Thuong, N.T.; Helene, C. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 666–690.

(4) For early examples of minor groove recognition of extended A,T-tracts based on the 1:1 distamycin-DNA model, see: (a) Youngquist, R.S.; Dervan, P.B. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 2565–2569. (b) Youngquist, R.S.; Dervan, P.B. *J. Am. Chem. Soc.* **1987**, *109*, 7564.

(5) Kelly, J.J.; Baird, E.E.; Dervan, P.B., *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6981–6985.

(6) Keilkopf, C.L.; Baird, E.E.; Dervan, P.B.; Rees, D.C. *Nature Struct. Biol.* **1998**, *5*, 104-109.

(7) β/β pairs specify for A,T base pairs: (a) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6160–6166. (b) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *Chem. Eur. J.* **1997**, *3*, 1600–1607.

(8) (a) Jones, K.A.; Peterlin, B.M. *Ann. Rev. Biochem.* **1994**, *63*, 717–743. (b) Frech, K.; Brack-Werner, R.; Werner, T. *Virology* **1996**, *224*, 256–267.

(9) Im/Py and Py/Im pairs specify for G•C and C•G, respectively: (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.; Geirstanger, B.H.; Wemmer, D.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586–7590. (c) Wade, W.S.; Mrksich, M.; Dervan, P.B. *Biochemistry* **1993**, *32*, 11385–11389. (d) Geierstanger, B.H.; Dwyer, T.J.; Bathini, Y.; Lown, J.W.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 4474–4482. (e) White, S.; Baird, E.E.; Dervan, P.B. *Chem. & Biol.* **1997**, *4*, 569–578.

(10) Py/Py pairs specify for A,T base pairs: (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. *Nature Struct. Biol.* **1994**, *1*, 169–175. (d) White, S.; Baird, E.E.; Dervan, P.B. *Biochemistry* **1996**, *35*, 12532–12537.

(11) (a) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Nature Struct. Biol.* **1996**, *3*, 321–324. (b) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Chem. & Biol.* **1996**, *3*, 369–377.

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

(13) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods Enzymol.* **1986**, *130*, 132–181.

(14) For the treatment of data on cooperative association of ligands, see: Cantor, C.R.; Schimmel, P.R., *Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules*; W.H. Freeman, New York, NY, 1980, p 863.

(15) The method used here for determining association constants involves the assumption that $[L]_{tot} \approx [L]_{free}$, where $[L]_{free}$ is the concentration of polyamide free in solution (unbound). For very high association constants this assumption becomes invalid, resulting in underestimated association constants. In the experiments described here, the DNA concentration is estimated to be ~ 5 pM. As a consequence, apparent

association constants greater than 1–2 × 10^{10} M^{-1} represent a lower limit of the true association constant.

(16) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Angew. Chem. Int. Ed. Eng.* **1998**, *37*, 1421–1423.

CHAPTER EIGHT

Extended Hairpin Polyamide Motif

The text of this chapter is taken from a published paper that was coauthored with

Prof. Peter B. Dervan and Eldon E. Baird

(Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Chem. Biol.* **1996**, *3*, 369–377)

Introduction

The three-ring polyamide ImPyPy-Dp (Im = *N*-methylimidazole, Py = *N*-methylpyrrole, Dp = dimethylaminopropylamide) was found by footprinting and affinity cleavage studies to specifically bind to the sequence 5'-TGTCA-3'.[1] In the $(ImPyPy\text{-}Dp)_2$•5'-(A,T)G(A,T)C(A,T)-3' complex, each polyamide makes specific contacts with one strand on the floor of the minor groove such that the sequence-specificity depends on the sequence of side-by-side amino acid pairings.[1-3] A pairing of imidazole opposite pyrrole targets a G•C base-pair, and a pairing of pyrrole opposite imidazole targets a C•G base-pair.[1-3] A pyrrole/pyrrole combination is partially degenerate and targets both T•A and A•T base-pairs.[1-6] Specificity for G,C base pairs results from the formation of a hydrogen bond between the imidazole N3 and the exocyclic amino group of guanine.[1-3] The generality of these pairing rules has been demonstrated by targeting a wide variety of sequences.[7-11]

Hairpin and Extended Binding Motifs. In addition to elucidating the scope and limitations of the sequence-specificity rules, we have made efforts to increase the DNA-binding affinity of polyamides by covalently linking polyamide subunits.[12-15] Some of the polyamide molecules that have been studied and their interactions with DNA are depicted schematically in Figure 1. The polyamide ImPyPy-γ-PyPyPy-Dp, containing a "turn" amino acid, γ-aminobutyric acid (γ), specifically binds the sequence 5'-TGTTA-3' in a "hairpin" conformation (Fig. 1C) with an equilibrium association constant, K_a, of 8×10^7 M^{-1}. This is an increase of about 300-fold relative to unlinked three-ring polyamides.[15] Addition of a C-terminal β-alanine residue enhances both the binding affinity and sequence specificity of hairpin polyamides.[16]

The six-ring polyamide ImPyPy-β-PyPyPy-Dp containing an internal β-alanine (β) residue specifically binds as a dimer in an extended conformation to two designated target sites, 5'-TGTTAAACA-3' (9 bp) and 5'-AAAAAGACAAAAA-3' (13 bp), with equilibrium association constants of $K_a = 8 \times 10^8$ M^{-1} and $K_a = 5 \times 10^9$ M^{-1}, respectively.

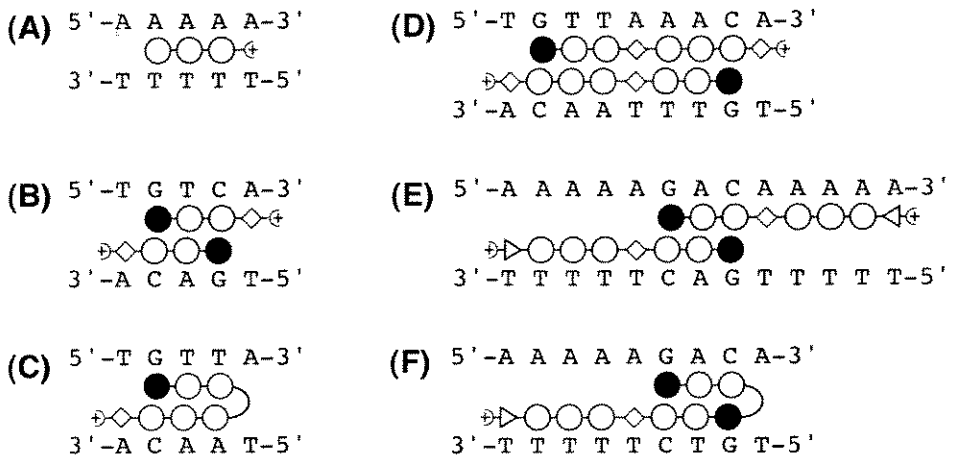


Figure 1. Schematic models of polyamide-DNA complexes: (A) 1:1 complex of distamycin, (B) 2:1 complex of ImPyPy-Dp, (C) "hairpin" complex of ImPyPy-γ-PyPyPy-β-Dp, (D) 9 bp and (E) 13 bp "extended" complexes formed by ImPyPy-β-PyPyPy-β-Dp and ImPyPy-β-PyPyPy-G-Dp, respectively, and (F) 9bp "extended hairpin" complex of ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp. Filled and unfilled circles represent imidazole and pyrrole rings, respectively, triangles and diamonds represent glycine and β-alanine, respectively, and curved lines represent γ-aminobutyric acid.

This represents increases of 10-fold for the 9 bp sequence and 100-fold for the 13 bp sequence over the formally *N*-methylpyrrole-linked polyamide ImPyPy-Py-PyPyPy-Dp.[17] Addition of a C-terminal amino acid regulates the specificity between the 13 bp and 9 bp binding modes. Two ImPyPy-β-PyPyPy-β-Dp polyamides, each containing a C-terminal β-alanine residue, can bind directly opposite one another (Fig. 1D), whereas two ImPyPy-β-PyPyPy-G-Dp polyamides, each containing a C-terminal glycine residue strongly favor binding in the 13 bp binding mode (Fig 1E). The 13 bp binding mode integrates the 2:1 and 1:1 polyamide-DNA binding motifs at a single site.[18] The ImPyPy moieties of two ImPyPy-β-PyPyPy-G-Dp polyamides bind the central 5'-AGACA-3' sequence in a 2:1 manner as in the ImPyPy homodimer, and the PyPyPy moieties of the polyamides bind the all-A,T flanking sequences as in the 1:1 complexes of distamycin (Fig 1A).

It has been demonstrated that γ-aminobutyric does not optimally link polyamide subunits in extended conformations, and β-alanine does not optimally link polyamide

subunits in hairpin conformations. The polyamide ImPyPy-β-PyPyPy-Dp binds the hairpin site 5'-TGTTAgacc-3' with an association constant of $K_a \leq 2 \times 10^6$ M^{-1}, a decrease of ≥ 40-fold relative to ImPyPy-γ-PyPyPy-Dp, and displays a cooperative binding isotherm (eq 2, n=2) in quantitative footprinting experiments at this site, consistent with mismatched binding as an intermolecular dimer.[15] The polyamide ImPyPy-γ-PyPyPy-Dp binds the 13 bp site 5'-AAAAAGACAAAAA-3' with an association constant of $K_a = 6 \times 10^6$ M^{-1}, a decrease of 700-fold relative to ImPyPy-β-PyPyPy-Dp, and displays a Langmuir binding isotherm (eq 2, n=1) at this site consistent with mismatched binding as a hairpin. The site 5'-TGTTAAACA-3' is a match site for both hairpin (5'-TGTTA-3' and 5'-AAACA-3') and extended binding by ImPyPy-X-PyPyPy-Dp polyamides. ImPyPy-γ-PyPyPy-Dp and ImPyPy-β-PyPyPy-Dp both bind this sequence with high affinity ($K_a = 1 \times 10^8$ M^{-1} and $K_a = 8 \times 10^8$ M^{-1}, respectively), but display, respectively, a Langmuir isotherm consistent with hairpin binding and a cooperative isotherm consistent with extended binding. Remarkably, given their simplicity and similarity, γ-aminobutyric acid and β-alanine selectively link polyamide subunits in hairpin and extended conformations, respectively.

Extended Hairpin Motif. The results described above suggest that γ-aminobutyric acid and β-alanine could be combined within a single polyamide with predictable results. We report here the synthesis of the nine-ring polyamide ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp and its association constant for the designated 9 bp target site 5'-AAAAAGACA-3' (Fig. 1F). The expected ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp•5'-AAAAAGACA-3' "extended hairpin" complex integrates the 1:1 and 2:1 polyamide-DNA motifs at a single site. Each of the three linkers within the polyamide is expected to fulfill a specific function: γ links subunits in a hairpin conformation, β links subunits in an extended conformation, and G confers specificity for the 1:1 binding mode at the C-terminal end of the polyamide.

In principle, the "extended hairpin" motif could provide a general motif for targeting

hairpin binding sites having an $(A,T)_4$ flanking sequence. Since six-ring hairpin polyamides already have high DNA-binding affinity ($K_a \approx 10^8$ M^{-1}), adding an optimally linked three-ring subunit should provide a polyamide with very high binding affinity. To determine the increase in binding affinity provided by the C-terminal PyPyPy subunit, the six-ring hairpin polyamide ImPyPy-γ-ImPyPy-β-Dp was synthesized and its association constant for 5'-AGACA-3' determined.

Binding to the 9 bp target site 5'-ATATAGACA-3' was also investigated. Based on previous results with the distamycin analog Ac-PyPyPy-Dp (Ac = acetyl), a preference of ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp for 5'-AAAAAGACA-3' over 5'-ATATAGACA-3' is predicted.

Equilibrium association constants for the polyamides ImPyPy-γ-ImPyPy-β-Dp (**1**) and ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp (**2**) for two 9 bp target sites 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' located on a 247 bp restriction fragment, and to additional binding sites identified on the fragment, were determined by quantitative DNase I footprint titration experiments. We also carried out affinity cleavage experiments using the EDTA•Fe(II)-polyamide ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp-EDTA•Fe(II) (**2-E**) to determine the DNA-binding orientation and qualitative DNA-binding affinity and specificity of this polyamide (Fig. 2).

Results and Discussion

Synthesis. All polyamides were prepared in high purity using a solid phase synthetic methodology.[19] Polyamides **1** and **2** were assembled in a stepwise manner on Boc-β-alanine-Pam resin and Boc-glycine-Pam-resin, respectively. Polyamides **1**, **2** and **2-NH$_2$** were cleaved from the support with an appropriate primary amine and purified by reversed-phase HPLC to provide 10–30 mg of polyamide. Polyamide **2-NH$_2$** was treated with an excess of the dianhydride of EDTA, unreacted anhydride was hydrolyzed, and polyamide **2-E** isolated by reversed-phase HPLC.

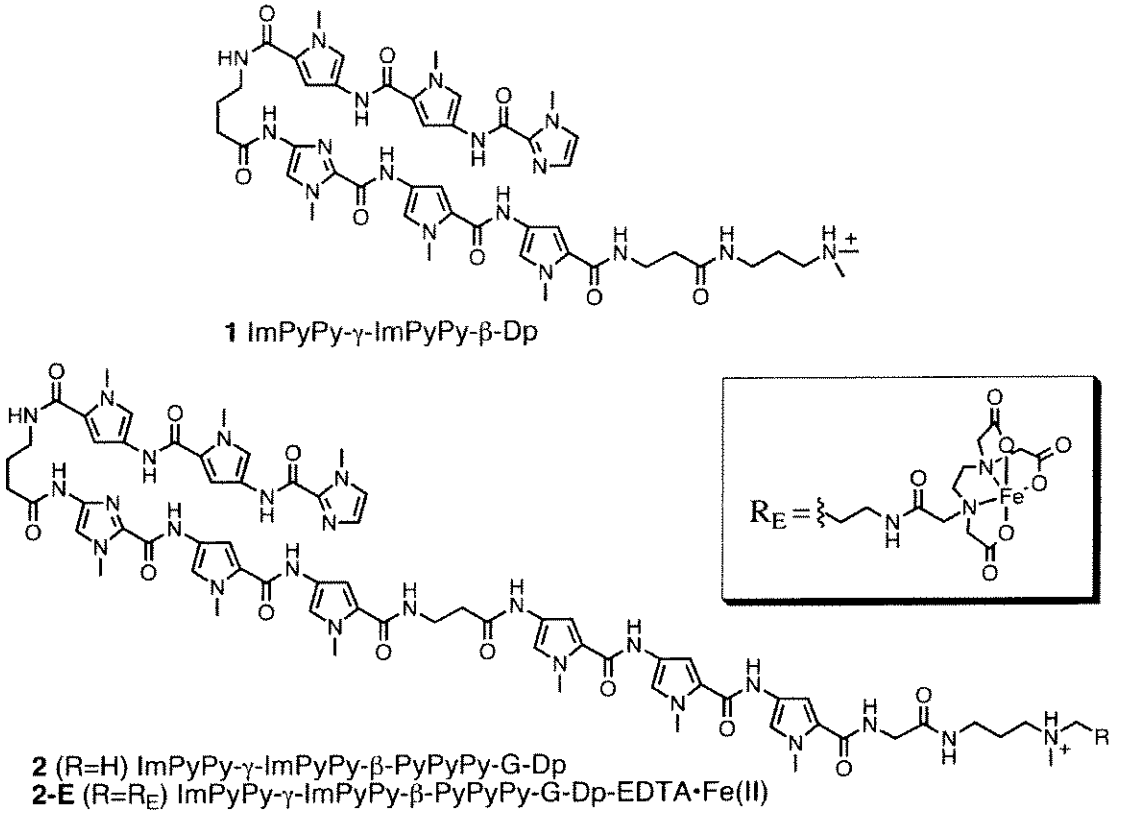


Figure 2. Structures of polyamides **1**, **2** and **2-E**.

```
5'-GCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGCTGC
3'-CGTTGACAACCCTTCCCGCTAGCCACGCCCGGAGAAGCGATAATGCGGTCGACCGCTTTCCCCCTACGACG

AAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGC
TTCCGCTAATTCAACCCATTGCGGTCCCAAAAGGGTCAGTGCTGCAACATTTTGCTGCCGGTCACTTAAGCTCG

TCGGTACCCGGGAACGTAGCGTACCGGTCGCAAAAAGACAGGCTCGACGCCGCATATAGACAGGCCCAGCTGCG
AGCCATGGGCCCTTGCATCGCATGGCCAGCGTTTTTCTGTCCGAGCTGCGGCGTATATCTGTCCGGGTCGACGC

TCCTAGCTAGCGTCGTAGCGTCTTAA-3'
AGGATCGATCGCAGCATCGCAGAATT-5'
```

Figure 3. Sequence of the 247 bp restriction fragment used for footprinting and affinity cleaving experiments. The two match sites are highlighted in bold type.

DNA-Binding Orientation. We carried out affinity cleavage experiments[1,20] with ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp-EDTA•Fe(II) (**2-E**) on the 5'- or 3'-^{32}P end-labeled 247 bp pJT4 *Afl* II/*Fsp* I restriction fragment (Fig. 3). This polyamide selectively binds to the 5'-AAAAAGACA-3' and 5'-ATATAGACATATA-3' target sequences at subnanomolar concentrations. A single 3'-shifted cleavage pattern is observed at each 9 bp binding site, indicating that the polyamide is bound in one orientation with the C-terminus at the 5' end of the 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' sequences (Fig. 4).

DNA-Binding Affinity and Specificity. The exact locations and sizes of all binding sites were determined first by preliminary footprinting experiments using the cleavage reagent methidiumpropyl EDTA•Fe(II)[21] (data not shown). Quantitative DNase I footprint titration experiments[22-24] on the 3'-^{32}P-labeled 247 bp restriction fragment (10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, 22 °C) revealed that ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp specifically binds the 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' target sequences with equilibrium association constants of $K_a \sim 2 \times 10^{10}$ M^{-1} and $K_a = 8 \times 10^9$ M^{-1}, respectively (Table 1). The polyamide also binds to additional sites on the restriction fragment with lower affinity. The six-ring hairpin polyamide ImPyPy-γ-ImPyPy-β-Dp binds 5'-aaaaAGACA-3' and 5'-atatAGACA-3' with association constants of $K_a = 5 \times 10^7$ M^{-1} and $K_a = 9 \times 10^7$ M^{-1}, respectively.

Relative to the six-ring polyamide ImPyPy-γ-ImPyPy-β-Dp, the nine-ring polyamide ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp binds 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' with ~400-fold and ~100-fold higher affinity, respectively. Similar binding enhancements have recently been reported in a separate system.[25] Addition of a C-terminal PyPyPy subunit using a β-alanine linker is thus an effective strategy for increasing the DNA-binding affinity of hairpin polyamides that bind adjacent to an $(A,T)_4$ sequence.

Polyamide ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp binds to several mismatch sites

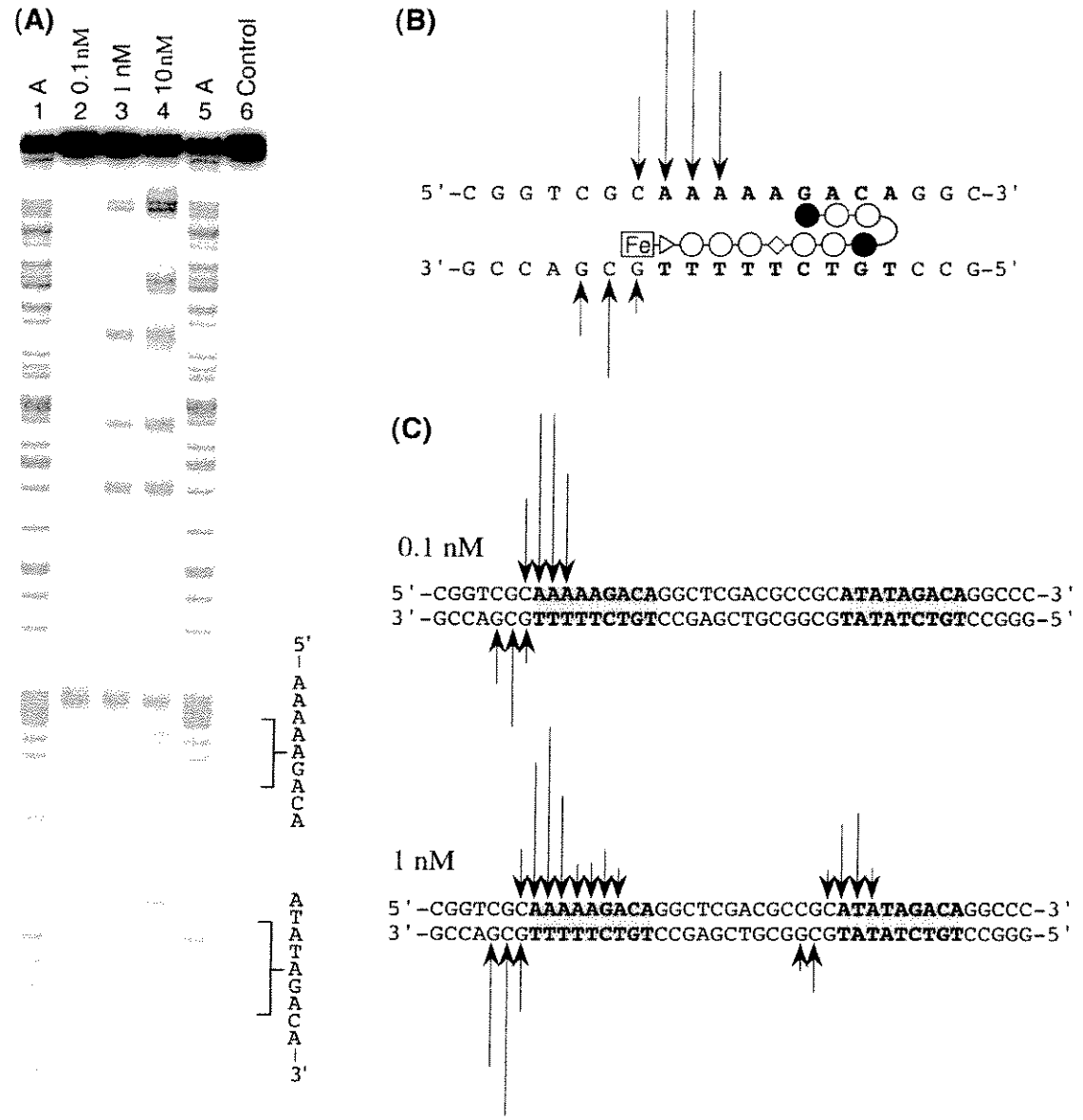


Figure 4. Compound **2-E** cleaves the 247 bp pJT4 *Afl* II/*Fsp* I restriction fragment at specific sites. (A) Storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel used to separate the fragments generated by cleavage reactions using polyamide **2**•Fe(II) on the 3'-^{32}P-end-labeled restriction fragment. Lanes 1 and 5: A sequencing lanes; lanes 2-4: cleavage products obtained in the presence of 0.1 nM, 1 nM and 10 nM **2-E**, respectively; lane 6: intact DNA. (B) Schematic representation of affinity cleaving experiment. Arrow lengths are proportional to the amount of cleavage at the indicated base. (C) Observed cleavage intensities at the designated match sites at 0.1 nM and 1 nM concentrations.

Table 1. Apparent first-order association constants (M^{-1})[a,b,c,d]

	Polyamide	
Binding Site	ImPyPy-γ-ImPyPy-β-Dp	ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp
Match sites:		
5'-AAAAA**GACA**-3'	5.2×10^7 (0.5)	2.0×10^{10} (0.4)
5'-ATATA**GACA**-3'	9.1×10^7 (1.1)	8.1×10^9 (0.1)
Mismatch sites:		
5'-G**AATTCA**CT-3'	$\leq 8 \times 10^6$	4.5×10^9 (1.0)
5'-C**GTTTTACA**-3'	9.2×10^6 (1.6)	1.6×10^9 (0.3)
5'-GTTTTCCCA-3'	$< 10^6$	2.5×10^9 (0.7)
5'-GGCGATTAAGTTG-3'	$< 10^6$	8.9×10^8 (0.6)
5'-TCGCTATTACGCCA-3'	$< 10^6$	1.5×10^9 (0.2)

[a]Values reported are the mean values obtained from four DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses. [b]The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [c]The 5 base-pair ImPyPy-γ-ImPyPy-β-Dp binding sites are in bold type. [d]The portion of each site occupied by the C-terminal PyPyPy subunit of ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp, as evidenced by affinity cleavage experiments with ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp-EDTA•Fe(II), is underlined.

present on the 247 bp restriction fragment with high affinity (Table 1). The two highest affinity mismatch sites, 5'-**G**AATT**C**ACT-3' (4.5×10^9 M^{-1}) and 5'-**G**TTTT**CC**CA-3' (2.5×10^9 M^{-1}), are bound with at least 5-fold reduced affinity relative to the optimal match site 5'-AAAAAGACA-3' (formally mismatched base-pairs are highlighted). This value may, however, be a lower limit due to the uncertainty in the very high association constant for the optimal match site. In contrast, the six-ring polyamide ImPyPy-γ-ImPyPy-β-Dp binds more strongly to the match site 5'-AGACA-3' than the single base-pair mismatch sites 5'-A**T**TCA-3' and 5'-T**T**ACA-3' by a factor of 10. Although it remains a challenge to optimize the specificity of the extended hairpin polyamide motif, it is significant that a wholly designed, synthetic polyamide specifically binds a designated nine base-pair site at subnanomolar concentrations.

Experimental Section

Synthesis. Polyamides were prepared using solid phase methods.[19]

Preparation of ^{32}P-Labeled DNA. Plasmid pJT4 was prepared by hybridizing two sets of 5'-phosphorylated complementary oligonucleotides, 5'-CCGGGAACGTAGC-GTACCGGTCGCAAAAAGACAGGCTCGA-3' and 5'-GGCGTCGAGCCTGTCTTTT-TGCGACCGGTACGCTACGTTC-3', and 5'-CGCCGCATATAGACAGGCCCAGCT-GCGTCCTAGCTAGCGTCGTAGCGTCTTAAGAG-3' and 5'-TCGACTCTTAAGACG-CTACGACGCTAGCTAGGACGCAGCTGGGCCTGTCTATATGC-3', and ligating the resulting duplexes to the large pUC19 *AvaI*/*SalI* restriction fragment. The 3'-^{32}P end-labeled *AflII*/ *Fsp*I fragment was prepared by digesting the plasmid with *AflII* and simultaneously filling in using Sequenase, [α-^{32}P]-deoxyadenosine-5'-triphosphate, and [α-^{32}P]-thymidine-5'-triphosphate, digesting with *Fsp*I, and isolating the 247 bp fragment by nondenaturing gel electrophoresis. The 5'-^{32}P-end-labeled *AflII*/*Fsp*I fragment was prepared using standard methods. A and G sequencing were carried out as described.[26,27] Standard methods were used for all DNA manipulations.[28]

Affinity Cleavage Reactions. All reactions were executed in a total volume of 400 μL. A stock solution of **2-E** or H_2O was added to a solution containing labeled restriction fragment (15,000 cpm), affording final solution conditions of 20 mM HEPES, 200 mM NaCl, 50 μg/mL glycogen, and pH 7.3. Subsequently, 20 μL of freshly prepared 20 μM $Fe(NH_4)_2(SO_4)_2$ was added and the solution allowed to equilibrate for 20 min. Cleavage reactions were initiated by the addition of 40 μL of 50 mM dithiothreitol, allowed to proceed for 12 min at 22 °C, then stopped by the addition of 1 mL of ethanol. Reactions were precipitated and the cleavage products separated using standard methods.[25] Next, 10 μL of a solution containing calf thymus DNA (140 μM base-pair) (Pharmacia) and glycogen (2.8 mg/mL) was added, and the DNA precipitated. The reactions were resuspended in 1X TBE/80% formamide loading buffer, denatured by heating at 85 °C for 10 minutes, and placed on ice. The reaction products were separated by electrophoresis on

an 8% polyacrylamide gel (5% cross-link, 7 M urea) in 1X TBE at 2,000 V. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics). Relative cleavage intensities were determined by volume integration of individual cleavage bands using ImageQuant software.

Quantitative DNase I Footprint Titration Experiments. All reactions were executed in a total volume of 400 μL. A polyamide stock solution or H_2O (for reference lanes) was added to an assay buffer containing radiolabeled restriction fragment (15,000 cpm), affording final solution conditions of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, and either (i) 1 pM–10 nM polyamide or (ii) no polyamide (for reference lanes). The solutions were allowed to equilibrate at 22 ℃ for (i) 12 hours for polyamide **1** or (ii) 36 hours for polyamide **2**. Footprinting reactions were initiated by the addition of 10 μL of a DNase I stock solution (at the appropriate concentration to give ~ 55% intact DNA) containing 1 mM dithiothreitol and allowed to proceed for 7 minutes at 22 °C. The reactions were stopped by the addition of 50 μL of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/mL glycogen, and 30 μM base-pair calf thymus DNA, and ethanol precipitated. Reactions were resuspended in 1X TBE/80% formamide loading buffer, denatured by heating at 85 °C for 10 min, and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% cross-link, 7 M urea) in 1X TBE at 2000 V. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics). Data analysis was carried out as described in Chapter Two.

References

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

(2) Mrksich, M.; Wade, W.S.; Dwyer, T.J.; Geirstanger, B.H.; Wemmer, D.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586–7590.

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

(4) Pelton, J.G.; Wemmer, D.E. *Proc. Natl. Acad. Sci.* **1989**, *86*, 5723–5727.

(5) Pelton, J.G.; Wemmer, D.E. *J. Am. Chem. Soc.* **1990**, *112*, 1393–1399.

(6) Chen, X.; Ramakrishnan, B.; Rao, S.T.; Sundaralingam, M.N. *Nature Struct. Biol.* **1994**, *1*, 169-175.

(7) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 2572–2576.

(8) Geierstanger, B.H.; Jacobsen, J-P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Biochemistry* **1994**, *33*, 3055–3062.

(9) Geierstanger, B.H.; Dwyer, T.J.; Bathini, Y.; Lown, J.W.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 4474–4482.

(10) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646–650.

(11) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 3325–3332.

(12) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 9892–9899.

(13) Dwyer, T.J.; Geirstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 9900–9906.

(14) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 3663–3664.

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

(16) Parks, M.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6147–6152.

(17) Trauger, J.W.; Baird, E.E.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**,

118, 6160–6166.

(18) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Nature Struct. Biol.* **3**, 321–324.

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

(20) Schultz, P.G.; Taylor, J.S.; Dervan, P.B. *J. Am. Chem. Soc.* **1982**, *104*, 6861–6863.

(21) Hertzberg, R.P.; Dervan, P.B. *J. Am. Chem. Soc.* **1982**, *104*, 313–315.

(22) Fox, K.R.; Waring, M.J. *Nucleic Acids Res.* **1984**, *12*, 9271–9285.

(23) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods Enzymology* **1986**, *130*, 132–181.

(24) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8462–8466.

(25) Nicolaou, K.C.; Smith, B.M.; Ajito, K.; Komatsu, H.; Gomez-Paloma, L.; Tor, Y. *J. Am. Chem. Soc.* **1996**, *118*, 2303–2304.

(26) Maxam, A.M.; Gilbert, W.S *Methods Enzymology* **1980**, *65*, 499–560.

(27) Iverson, B.L.; Dervan, P.B. *Nucleic Acids Res.* **1987**, *15*, 7823–7830.

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

CHAPTER NINE

Cooperative Hairpin Dimers for DNA Recognition by Pyrrole-Imidazole Polyamides

The text of this chapter is taken in part from a published paper that was coauthored with

Prof. Peter B. Dervan and Eldon E. Baird

(Trauger, J.W. *et al. Angewandte Chemie Int. Ed. Eng.* **1998**, *37*, 1421–1423)

Small molecules that permeate cells and bind predetermined DNA sequences have the potential to control the expression of specific genes.[1,2] Recently, an eight-ring polyamide that binds to a 6 base pair target site was shown to inhibit gene transcription in cell culture.[2] Polyamides recognizing longer DNA sequences should provide more specific biological activity,[3] which could be achieved by synthesizing larger hairpins.[4] However, the upper limit of polyamide size with regard to efficient cell permeation is not known.

Alternatively, a more biomimetic approach is to bind larger DNA sequences while maintaining the size of the polyamide. Nature's transcription factors often bind large DNA sequences by formation of cooperative protein dimers at adjacent half-sites.[5] For cooperatively binding extended Py-Im polyamide dimers, the two ligands can slip sideways with respect to one another, allowing recognition of other sequences.[6] Hairpin polyamides utilizing the turn-specific γ-aminobutyric acid linker[7] are constrained to be fully overlapped and preclude the "slipped motif" option. Here we report a cooperative six-ring extended hairpin polyamide which dimerizes to specifically bind a predetermined 10 base pair sequence.

For our target site, we chose a sequence contained in the regulatory region of the HIV-1 genome.[8] To design the ligand we considered the polyamide ring pairing rules,[9-13] the need for β-alanine (β) to relax ligand curvature,[6,13] and the preference of γ-aminobutyric acid (γ) for a "hairpin turn" conformation within polyamide-DNA complexes.[6,7a,e] This analysis suggested that the six-ring polyamide having the core sequence ImPy-β-ImPy-γ-ImPy might bind the target sequence 5'-AGCAGCTGCT-3' through formation of a cooperative hairpin dimer (Figs. 1 and 2). To avoid a collision between the N-terminal end of one ligand and the C-terminal end of the second within the complex, the positively-charged β-alanine-dimethylaminoproplyamide C-terminus used in standard polyamides has been replaced with the shorter, uncharged $(CH_2)_2OH$ group (C_2-OH). The cationic "turn"

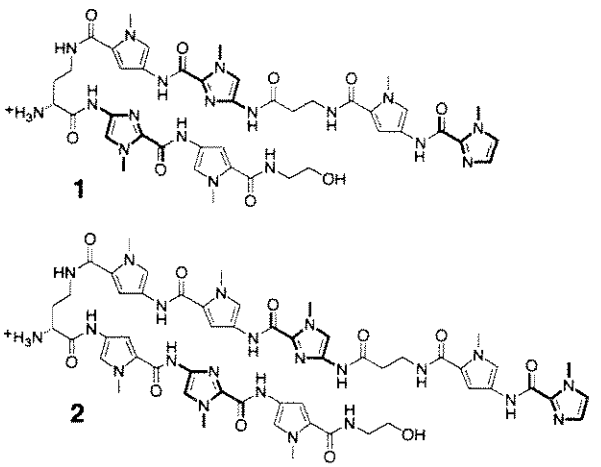


Figure 1. Structures of polyamides **1** and **2**.

residue (*R*)-2,4-diaminobutyric acid ($(R)^{H_2N}\gamma$)[14] maintains the overall +1 charge needed for optimal solubility in water.

Polyamide ImPy-β-ImPy-$(R)^{H_2N}\gamma$-ImPy-C_2-OH (**1**) was synthesized using solid-phase methods[15] on glycine-PAM resin,[16] reductively cleaved from the solid support using $LiBH_4$,[17] and purified by HPLC (reverse-phase).[18] Quantitative DNase I footprinting[19] on a 245 base pair 3'-^{32}P-end-labeled restriction fragment showed that **1** binds its match site 5'-AGCAGCTGCT-3' at nanomolar concentrations (apparent monomeric association constant, K_a = 1.9 (±0.3) x 10^8 M^{-1}), and also binds a single-base pair mismatch site 5'-AG<u>A</u>TGCTGCA-3' with 9-fold lower affinity, K_a = 2.2 (±0.5) x 10^7 M^{-1} (Fig. 1).[20] The binding data for match and single-base pair mismatch sites were well-fit by cooperative binding isotherms, consistent with formation of cooperative 2:1 polyamide-DNA complexes.[5] A double-base pair mismatch site, 5'-AGCTGCA<u>T</u>C<u>C</u>-3', is also bound with 65-fold lower affinity. The fact that this mismatch site, which contains the "half-site" 5'-AGCTGCA-3', is not effectively bound indicates that recognition of the match site occurs *through cooperative dimerization, and not due to formation of 1:1 hairpin* complexes.

Further study of the generality and sequence specificity of this motif is in progress and will be reported in due course. For example, we found that the eight-ring polyamide

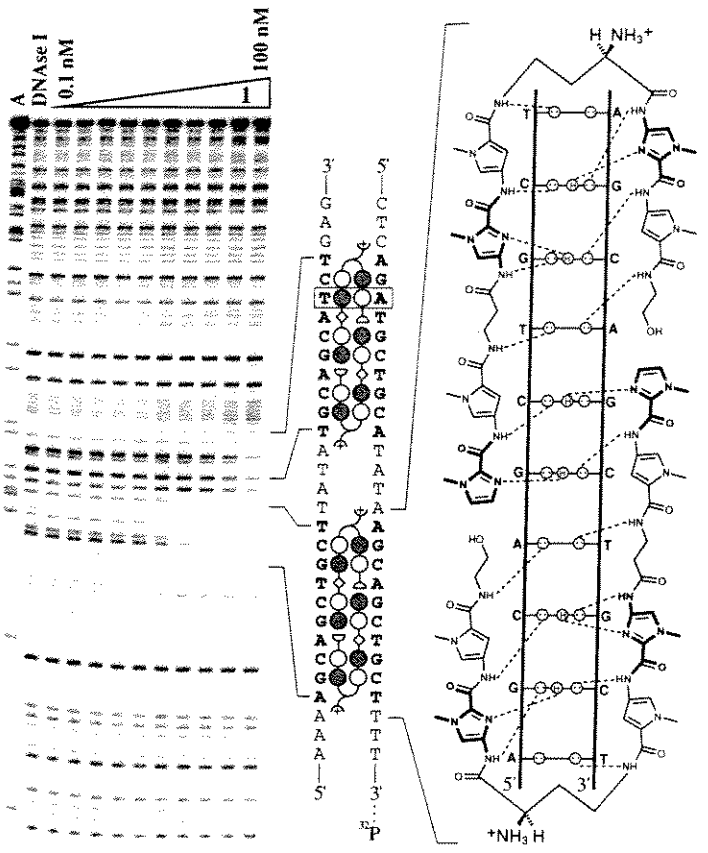


Figure 2. (Right) Model of the $(\mathbf{1})_2$ • 5'-AGCAGCTGCT-3' complex. Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines, and circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dashed lines. (Middle) Binding models for complexes of **1** at 10 base pair match and single-base pair mismatch sites (the mismatched base pair is highlighted by shading). The filled and open circles represent imidazole and pyrrole rings, respectively, diamonds represent β-alanine, half-circles represent $(CH_2)_2OH$ groups, and curved lines represent (*R*)-2,4-diaminobutyric acid. (Left) Storage phosphor autoradiogram of the 8% denaturing polyacrylamide gel used to separate the fragments generated by DNase I digestion in a quantitative footprint titration experiment with polyamide **1**: lane 1, A lane; lane 2, DNase I digestion products obtained in the absence of polyamide; lanes 3-12, DNase I digestion products obtained in the presence of 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, and 100 nM polyamide **1**, respectively. All reactions contain 3'-^{32}P-end-labeled pJT-LTR *Eco*RI/*Hin*dIII restriction fragment (15 kcpm), 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$ (pH 7.0, 24 °C). Plasmid pJT-LTR was prepared by ligating an insert having the sequence 5'-CCGGTAACCAGAGAGACCCAGTACAGGCAA-AAAGCAGCTGCTTATATGCAGCATCTGAGGGACGCCACTCCCCAGTCCCGCCC AGGCCACGCCTCCCTGGAAAGTCCCCAGCGGAAAGTCCCTTGTAGAAAGCTCG ATGTCAGCAGTCTTTGTAGTACTCCGGATGCAGCTCTCGGGCCACGTGATGAA ATGCTAGGCGGCTGTCAATCGA-3' to the large *AvaI*/*Sal*I fragment of pUC19.

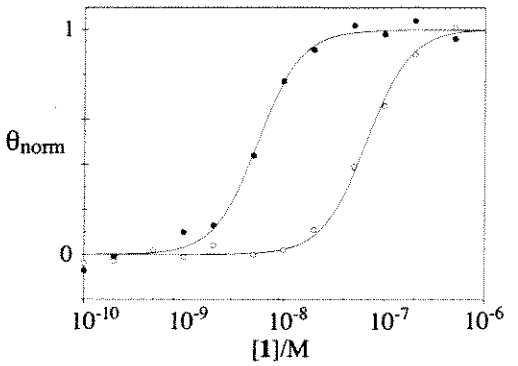


Figure 3. Binding isotherms obtained from a single gel for binding of **1** to the match site 5'-AGCAGCTGCT-3' (filled circles) and single-base pair mismatch site 5'-AGATGCTGCA-3' (open circles). The reported association constants are the average values obtained from four independent footprint titration experiments.

Figure 4. Models of (A) the 10 base pair $(\mathbf{1})_2$•5'-AGCAGCTGCT-3' complex and (B) the 12 base pair $(\mathbf{2})_2$•5'-AAGCAGCTGCTT-3' complex.

ImPy-β-ImPyPy-$(R)^{H_2N}$γ-PyImPy-C_2-OH (**2**) binds the twelve base pair match site 5'-AAGCAGCTGCTT-3' with 10-fold higher affinity than **1**, and is ~100-fold specific for this site versus the double-base pair mismatch site 5'-CAGATGCTGCAT-3'.

It is interesting to note that the DNA-binding affinity and specificity of the six-ring polyamide **1** for its ten base pair binding site are typical of standard six-ring hairpins that recognize five base pairs.[7] Thus, use of a the cooperative hairpin dimer motif doubles the binding site size relative to the standard hairpin motif without sacrificing affinity or specificity, and without increasing the molecular weight of the ligand. The results reported here show that by using a novel cooperative hairpin dimer motif, relatively low molecular weight pyrrole-imidazole polyamides (MW ~ 950-1,200) can specifically recognize 10-12 base pairs of DNA. Such polyamides will be useful in determining the optimal ligand size and optimal binding site size required for specific biological activity.

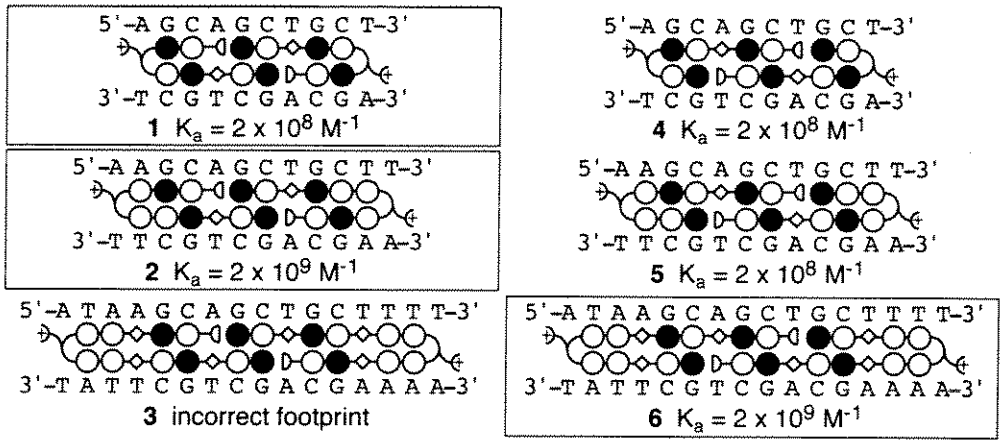


Figure 5. Models of complexes of polyamides **1**–**6** with the site 5'-ATAAGCAGCTGCTTTT-3'. Polyamides that are at least 5-fold specific for this sequence versus any other sites present on the pJT-LTR *Eco*RI/*Hin*dIII restriction fragment are boxed.

Additional Polyamides. In addition to polyamides **1** and **2** described above, four additional cooperative hairpin polyamides **3**–**6** were synthesized and their equilibrium association constants for the site 5'-ATAAGCAGCTGCTTTT-3' determined by quantitative DNase I footprinting experiments. Out of the series of six polyamides, three are at least 5-fold specific for the match site versus all other sites on the restriction fragment used (Fig. 5).

References

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

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

(3) Dervan, P.B. *Science* **1986**, *232*, 464.

(4) Turner, J.M.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, *119*, 7636-7644.

(5) a) Ptashne, M. *A Genetic Switch*, Blackwell Scientific Publications and Cell Press: Palo Alto, CA, **1986**; b) Pabo, C.O.; Sauer, R.T. *Ann. Rev. Biochem.* **1992**, *61*, 1053-1095; c) Marmorstein, R.; Carey, M.; Ptashne, M.; Harrison, S.C. *Nature* **1992**, *356*, 408-414; d) Klemm, J.D.; Rould, M.A.; Aurora, R.; Herr, W.; Pabo, C.O. *Cell* **1994**, *77*, 21-32; e) Bellon, S.F.; Rodgers, K.K.; Schatz, D.G.; Coleman, J.E.; Steitz, T.A. *Nature Struct. Biol.* **1997**, *4*, 586-591.

(6) a) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6160-6166; b) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *Chem. Eur. J.* **1997**, *3*, 1600-1607.

(7) a) Mrksich, M.; Parks, M.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988; b) Parks, M.E.; Baird, E.E.; Dervan, P.B. *ibid.* **1996**, *118*, 6153-6159; c) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *ibid.* **1996**, *118*, 8198-8206; d) Swalley, S. E.; Baird, E.E.; Dervan, P.B. *ibid.* **1997**, *119*, 6953-6961; e) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Chem. Biol.* **1996**, *3*, 369-377; f) Declairac, R.P.L.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1997**, *119*, 7909-7916.

(8) (a) Jones, K.A.; Peterlin, B.M. *Ann. Rev. Biochem.* **1994**, *63*, 717–743; (b) Frech, K.; Brack-Werner; R.; Werner, T. *Virology* **1996**, *224*, 256–267.

(9) 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.; Geirstanger, 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.

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

(11) a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 2572-2576; b) Geierstanger, B.H.; Dwyer, T.J.; Bathini, Y.; Lown, J.W.; Wemmer, D.E. *ibid.* **1993**, *115*, 4474; c) Geierstanger, B.H.; Jacobsen, J.P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Biochemistry* **1994**, *33*, 3055; d) White, S.; Baird, E.E. *Chem. Biol.* **1997**, *4*, 569-578.

(12) 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*, 3325-3332.

(13) Keilkopf, C.L.; Baird, E.E.; Dervan, P.B.; Rees, D.C. *Nature Struct. Biol.* **1998**, *5*, 104-109.

(14) Herman, D.M.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1998**, in press.

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

(16) Available in 0.3 mmol/g substitution from Peptides International, Louisville, KY.

(17) (a) Mitchell, A.R.; Kent, S.B.; Engelhard, M.; Merrifield, R.B.J. *J. Org. Chem.* **1978**, *43*, 2845; (b) Stewart, J.M.; Young, J.D. *Solid Phase Peptide Synthesis*, Pierce Chemical Company, Rockford, IL, **1984**.

(18) The identity and purity of polyamides **1** and **2** was confirmed by ^{1}H NMR, analytical HPLC, and MALDI-TOF MS. MALDI-TOF MS (monoisotopic) (M+H): **1**, obsd 953.3, calcd ($C_{42}H_{53}N_{18}O_9$) 953.4; **2**, obsd 1197.5, calcd ($C_{54}H_{65}N_{22}O_{11}$) 1197.5.

(19) a) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods in Enzymology* **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.

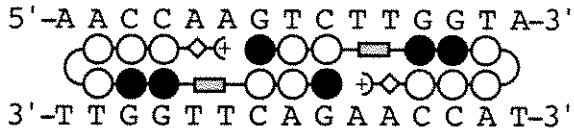
(20) For the treatment of data on cooperative association of ligands, see: Cantor, C.R.; Schimmel, P.R. *Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules*, W.H. Freeman, New York, NY, **1980**, 863.

CHAPTER TEN

Recognition of 15 Base Pairs by Cooperative Hairpin Polyamides

Cell-permeable small molecules that bind predetermined DNA sequences have the potential to control the transcription of specific genes.[1,2] Recently, an 8-ring polyamide that recognizes a 6 base pair target site was shown to inhibit gene transcription in cell culture.[2] Polyamides that recognize longer DNA sequences should provide more specific biological activity,[3] which could be achieved by synthesizing larger hairpins.[4] Cooperative polyamide dimers represent another approach to the recognition of larger DNA sequences. This approach has the potential advantage of keeping the size of the polyamide similar to that of those known to permeate cells.[2] Transcription factors often recognize large DNA sequences by formation of cooperative dimers.[5] For cooperatively binding extended Py-Im polyamide dimers, the two ligands can slip sideways with respect to one another, allowing recognition of other sequences.[6] Hairpin polyamides utilizing the linker γ-aminobutyric acid[7] are constrained to be fully overlapped and preclude the "slipped motif" option. These results suggested that hairpins linked to a cooperative polyamide dimerization subunit would recognize large DNA sequences with very high sequence-specificity. We recently reported that four- and six-ring hairpin polyamides linked to a two-ring dimerization domain cooperatively dimerize to recognize predetermined 10- and 12-base pair sequences, respectively.[8] We report here the design of cooperative hairpins that recognize the 15 base pair target sequence 5'-AACCAAGTCTTGGTA-3'.

To design the ligands we considered the polyamide ring pairing rules[9-13] and the need for an appropriate linker between the hairpin and the dimerization domain. This analysis suggested that the 9-ring polyamide having the sequence ImPyPy-X-ImImPy-γ-PyPyPy-β-Dp, where X is an appropriate linker amino acid, could bind the target sequence through formation of a cooperative hairpin dimer (Fig. 1). Seven polyamides (**1**–**7**, Fig. 2) with linkers that range in length from 5 to 10 atoms were synthesized using solid phase



Center mismatch:
5'-A A C C A A **C** T **G** T T G G T A-3'

Edge mismatch:
5'-A A C C A A G T C T T G **C** **G** A-3'

Figure 1. (Top) Model of the complex of an ImPyPy-X-ImImPy-γ-PyPyPy-β-Dp polyamide (X denotes a variable linker amino acid) with the 15 base pair sequence 5'-AACCAAGTCTTGGTA-3'. (Bottom) Sequences of the "center" and "edge" two-base pair mismatch sites. Mismatched base pairs are in bold type.

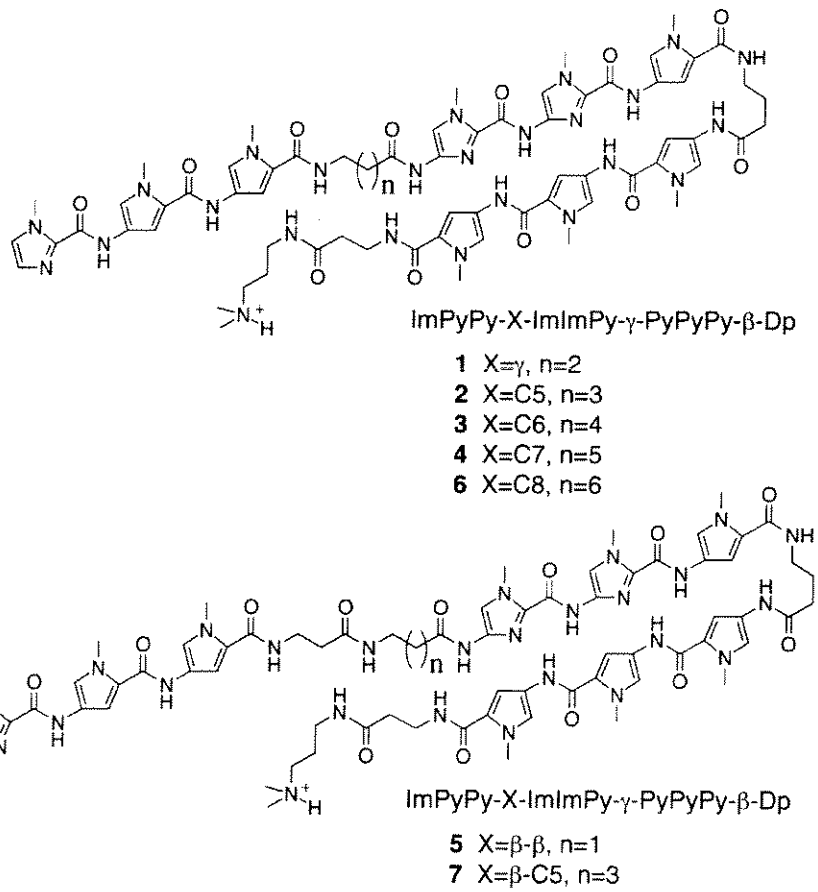


Figure 2. Structures of ImPyPy-X-ImImPy-γ-PyPyPy-β-Dp polyamides **1–7**.

methods.[14] The identity and purity of the polyamides were verified by analytical HPLC, ^{1}H NMR and MALDI-TOF-MS.

The DNA-binding affinity and specificity of ImPyPy-X-ImImPy-γ-PyPyPy-β-Dp polyamides **1**–**7** were studied by quantitative DNase I footprinting[15] on a 3'-^{32}P-end-labeled restriction fragment containing the match site AACCAAGTCTTGGTA-3' and the two-base pair mismatch sequences 5'-AACCAA**C**T**G**TTGGTA-3' (center mismatch) and 5'-AACCAAGTCTTG**CG**A-3' (edge mismatch) (Fig. 3). Polyamides **2**–**7** bind the match site with K_a = 1–4 x 10^8 M^{-1}, while **1** (X = γ) binds with lower affinity (K_a = 1 x 10^7 M^{-1}) (Table 1). The parent hairpin Ac-ImImPy-γ-PyPyPy-β-Dp binds the 5'-AACCA-3' and 5'-TGGTA-3' subsites with K_a = 3 x 10^7 M^{-1}. The low affinity of **1**, which is less than that of the parent hairpin, supports that γ-aminobutyric acid is a turn-specific linker. Despite the similar binding affinities of **2**–**7** for the match site, the specificity of these ligands versus the "center" and "edge" mismatch sites varies significantly: **2**, **3** and **6** (X = C5, C6 and C8) are 10–19-fold specific, while **4**, **5** and **7** (X = C7, β-β and β-C5) are only 2-fold specific. In sum, of seven ligands synthesized, three recognize the 15 base pair match sequence with ≥10-fold specificity over the designated mismatch sites.

The fact that the optimal polyamides **2**, **3** and **6** bind the edge mismatch sequence with lower affinity supports that the increased affinity of this ligand for the match site is the result of *cooperative* complex formation. In addition, the data obtained from quantitative footprinting experiments for binding of **2**, **3** and **6** are well-fit by cooperative binding isotherms.[6,16] Comparison of the equilibrium association constants of **2**, **3** and **6** for the match site with that of the parent hairpin Ac-ImImPy-γ-PyPyPy-β-Dp reveals relatively modest cooperative enhancements in binding affinity of 7–13-fold. However, the cooperative enhancement is entirely sequence-specific, resulting in good overall specificity.

The modest cooperative enhancement observed with polyamides **2**, **3** and **6** suggests that the geometry of the dimeric complexes is imperfect. This could be due to incorrect positioning of the dimerization domains within the complexes, or to distortion of the ligands

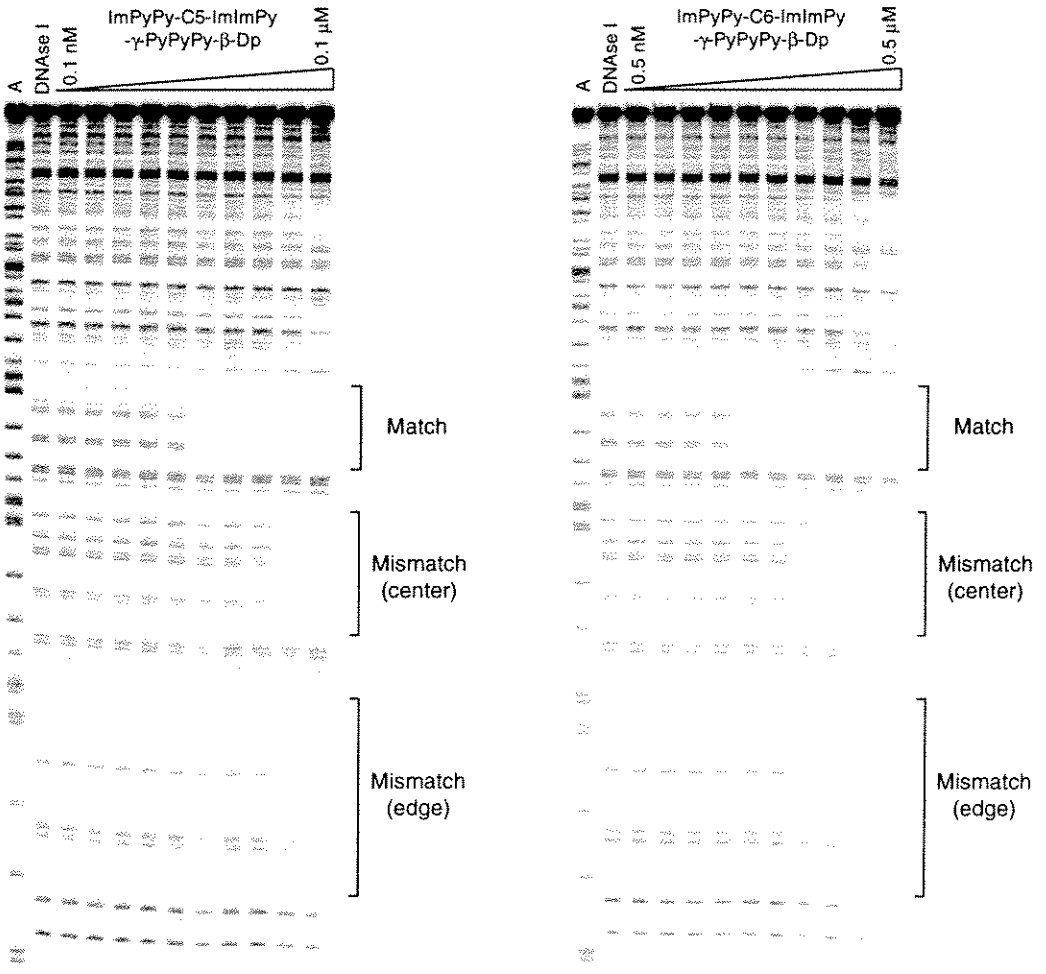


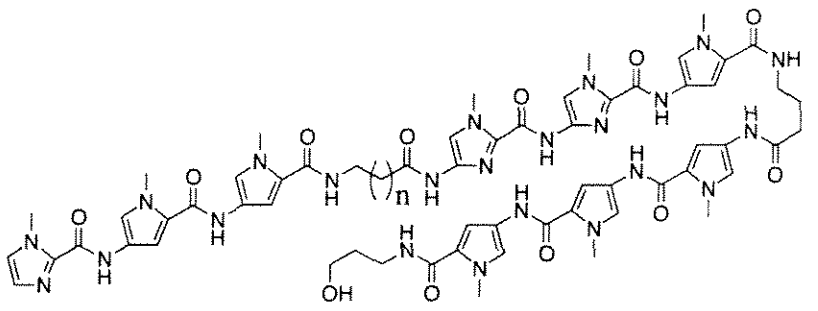
Figure 3. Storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel used to separate the fragments generated by DNase I digestion in a quantitative footprint titration experiment with polyamides **2** and **3**: lane 1, A lane; lane 2, DNase I digestion products obtained in the absence of polyamide; lanes 3–10, DNase I digestion products obtained in the presence of 0.01, 0.02, 0.05, 0.1, 0.2, 0.5 and 1 nM polyamide, respectively; lane 11, intact DNA. All reactions contain 3'-^{32}P-end-labeled *Afl*II/*Fsp*I restriction fragment from plasmid pJTCC15 (15 kcpm), 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$ (pH 7.0, 24 °C).

Table 1. Recognition of 15 base pairs by ImPyPy-X-ImImPy-γ-PyPyPy-β-Dp polyamides. Association constants (K_a) for the match site 5'-AACCAAGTCTTGGTA-3' and specificities for the match site versus center (5'-AACCAA**C**T**G**TTGGTA-3') and edge (5'-AACCAAGTCTTG**CG**A-3') mismatch sites are shown.

		Length of X	K_a (M^{-1})	Specificity	
	X =	(# atoms)	Match	Center mismatch	Edge mismatch
1	γ	5	1×10^7	1	1
2	C5	6	**4×10^8**	**11**	**19**
3	C6	7	**2×10^8**	**12**	**15**
4	C7	8	2×10^8	2	2
5	β-β	8	3×10^8	2	2
6	C8	9	**2×10^8**	**10**	**≥ 10**
7	β-C5	10	1×10^8	2	2

Solution conditions: 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at 24 °C and pH 7.0. The parent hairpin Ac-ImImPy-γ-PyPyPy-β-Dp binds both 5 base pair binding sites within the 15 base pair target site 5'-AACCAAGTCTTGGTA-3' with $K_a = 3 \times 10^7$ M^{-1}.

and/or the DNA within the complex. Alternatively, a steric clash between the C-terminus of one ligand and the N-terminus of the second ligand within the complex could reduce the cooperative affinity enhancement. Modeling suggests that this may potentially be a problem for the shorter linkers in **2** (X = C5) and **3** (X = C6), but is unlikely to be a problem for **6** (X = C8). To control for this possibility the analogs **2-OH** and **3-OH**, which have shorter C-terminal groups, were prepared (Fig. 4). The preparation of these analogs was accomplished by synthesis of the polyamides on β-alanine-PAM resin followed by reductive cleavage from the solid support using $LiBH_4$[17] and HPLC purification (Fig. 4). The corresponding analog of the parent hairpin, ImImPy-γ-PyPyPy-C3-OH, was also prepared. Quantitative footprinting of **2-OH**, **3-OH** and the parent hairpin analog revealed association complexes for the match site of 3×10^7 M^{-1}, 3×10^7 M^{-1} and 7×10^6 M^{-1}, repectively. The cooperative binding enhancements for **2-OH** and **3-OH** are 4-fold, which is less than the 7–13-fold enhancement observed with **2** and **3**. This



ImPyPy-X-ImImPy-γ-PyPyPy-C3-OH

2-OH X=C5, n=3
3-OH X=C6, n=4

Figure 4. Structures of ImPyPy-X-ImImPy-γ-PyPyPy-C3-OH polyamides **2-OH** and **3-OH**.

result suggests that the modest cooperative enhancements observed with **2** and **3** result not from a steric clash between ligands, but from incorrect positioning of the dimerization domains or conformational strain within the complexes.

The fact that polyamides **4**, **5** and **7** have similar affinities for the match site to **2**, **3** and **6**, but lower specificity inidicates that the optimal linkers not only position the dimerization domain to bind cooperatively, but also prevent the ligands from accessing mismatch binding modes.

To test whether attachment of the dimerization domain to the N-terminus of the hairpin (as in **1**–**7**) or to the C-terminus (Fig. 5) is preferable, polyamides **8** and **9** were synthesized (Fig. 6). The results of quantitative footprinting experiments with these ligands are shown in Table 2. Both polyamides have affinities for the 15 base pair match site similar to their oppositely-oriented analogs **2** and **3**, but very poor specificity versus the center and edge mismatches (≤2-fold). The low specificity of **8** and **9** compared to **2** and **3**, respectively, may result from lower specificity of the polyamide fragment PyPyPy-X-ImPyPy compared to the more imidazole-rich fragment ImPyPy-X-ImImPy, or from the effect of acylation of the dimerization domain Im residue instead of the hairpin imidazole.[18]

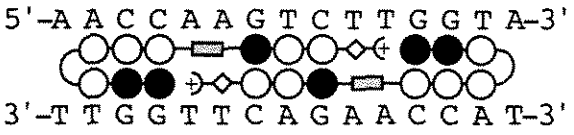


Figure 5. Model of the complex of an ImImPy-γ-PyPyPy-X-ImPyPy-β-Dp polyamide (X denotes a variable linker amino acid) with the 15 base pair sequence 5'-AACCAAGTCTTGGTA-3'.

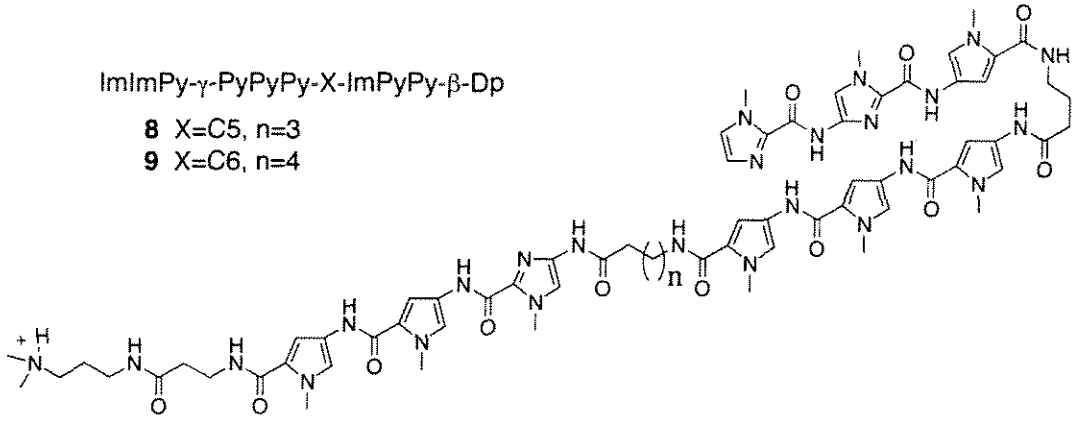


Figure 6. Structures of ImImPy-γ-PyPyPy-X-ImPyPy-β-Dp polyamides **8** and **9**.

Table 2. Recognition of 15 base pairs by ImImPy-γ-PyPyPy-X-ImPyPy-β-Dp polyamides. Association constants (K_a) for the match site 5'-AACCAAGTCTTGGTA-3' and specificities for the match site versus center (5'-AACCAA<u>C</u>T<u>G</u>TTGGTA-3') and edge (5'-AACCAAGTCTTG<u>CG</u>A-3') mismatch sites are shown.

		Length of X	K_a (M^{-1})	Specificity	
	X =	(# atoms)	Match	Center mismatch	Edge mismatch
8	C5	6	1×10^9	2	2
9	C6	7	2×10^8	1	2

Solution conditions: 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at 24 °C and pH 7.0. The parent hairpin ImImPy-γ-PyPyPy-β-Dp binds both 5 base pair binding sites within the 15 base pair target site 5'-<u>AACCA</u>AGTCT<u>TGGTA</u>-3' with $K_a = 1 \times 10^8\ M^{-1}$.

We attempted to recognize the 16 base pair sequence 5'-AACCAAGTACTTGGTA-3' using two ImPyPyPy-X-ImImPy-γ-PyPyPy-β-Dp polyamides, where X = C5 or C6.

Quantitative footprinting were carried out on a labeled restriction fragment containing the match site and two base pair mismatch sites 5'-AACCAA**C**TA**G**TTGGTA-3' (center) and 5'-AACCAAGTACTTG**CG**A-3' (edge). The C5- and C6-linked ligands bind with K_a = 7 x 10^8 M^{-1} and K_a = 1 x 10^8 M^{-1}, respectively. These binding affinities are similar to those of the analogous ligands **2** and **3**. Moreover, the overall DNA-binding specificity of these compounds is poor: a number of sites present on the restriction fragment are bound with affinity comparable to the match site (data not shown). Similar results were obtained with the polyamide ImPyPyPy-β-β-ImImPy-γ-PyPyPy-β-Dp.

The results reported here extend the maximum binding site size of the cooperative hairpin polyamide motif to 15 base pairs. While the specificity of the optimal ligands is quite good (≥10-fold versus two base pair mismatch sites), their cooperative affinity enhancement, and consequently their binding affinity, is modest. The low binding affinity likely represents imperfect complex geometry and/or conformational strain within the dimeric complex. An attempt to adapt the motif to recognize a 16 base pair sequence was unsuccessful. The challenge for the future is to create cooperative polyamide motifs that combine high match site binding affinity, sufficiently constrained structures that preclude mismatch binding modes and provide high sequence-specificity, and sequence generality.

References

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

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

(3) Dervan, P.B. *Science* **1986**, *232*, 464.

(4) Turner, J.M.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, *119*, 7636-7644.

(5) a) Ptashne, M. *A Genetic Switch*, Blackwell Scientific Publications and Cell Press: Palo Alto, CA, **1986**; b) Pabo, C.O.; Sauer, R.T. *Ann. Rev. Biochem.* **1992**, *61*, 1053-1095; c) Marmorstein, R.; Carey, M.; Ptashne, M.; Harrison, S.C. *Nature* **1992**, *356*, 408-414; d) Klemm, J.D.; Rould, M.A.; Aurora, R.; Herr, W.; Pabo, C.O. *Cell* **1994**, *77*, 21-32; e) Bellon, S.F.; Rodgers, K.K.; Schatz, D.G.; Coleman, J.E.; Steitz, T.A. *Nature Struct. Biol.* **1997**, *4*, 586-591.

(6) a) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6160-6166; b) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *Chem. Eur. J.* **1997**, *3*, 1600-1607.

(7) a) Mrksich, M.; Parks, M.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988; b) Parks, M.E.; Baird, E.E.; Dervan, P.B. *ibid.* **1996**, *118*, 6153-6159; c) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *ibid.* **1996**, *118*, 8198-8206; d) Swalley, S. E.; Baird, E.E.; Dervan, P.B. *ibid.* **1997**, *119*, 6953-6961; e) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Chem. Biol.* **1996**, *3*, 369-377; f) Declairac, R.P.L.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1997**, *119*, 7909-7916.

(8) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Angewandte Chemie Int. Ed. Eng.* **1998**, *37*, 1421–1423.

(9) 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.; Geirstanger, 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.

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

(11) a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 2572-2576; b) Geierstanger, B.H.; Dwyer, T.J.; Bathini, Y.; Lown, J.W.; Wemmer, D.E. *ibid.* **1993**, *115*, 4474; c) Geierstanger, B.H.; Jacobsen, J.P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Biochemistry* **1994**, *33*, 3055; d) White, S.; Baird, E.E. *Chem. Biol.* **1997**, *4*, 569-578.

(12) 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*, 3325-3332.

(13) Keilkopf, C.L.; Baird, E.E.; Dervan, P.B.; Rees, D.C. *Nature Struct. Biol.* **1998**, *5*, 104-109.

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

(15) a) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods in Enzymology* **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.

(16) For the treatment of data on cooperative association of ligands see: Cantor, C.R.; Schimmel, P.R. *Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules*, W.H. Freeman, New York, NY, **1980**, 863.

(17) (a) Mitchell, A.R.; Kent, S.B.; Engelhard, M.; Merrifield, R.B.J. *J. Org. Chem.* **1978**, *43*, 2845; (b) Stewart, J.M.; Young, J.D. *Solid Phase Peptide Synthesis*, Pierce Chemical Company, Rockford, IL, **1984**.

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

CHAPTER ELEVEN

Regulation of Gene Expression by Polyamides

The protein-DNA complex inhibition, transcription inhibition and viral replication inhibition experiments described in this chapter were carried out by our collaborators at The Scripps Research Institute.

Part I of this chapter is taken in part from a published paper that was coauthored with Eldon E. Baird, Prof. Peter B. Dervan, Prof. Joel M. Gottesfeld and Laura Neely

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

Part II of this chapter is taken from a manuscript that was coauthored with Eldon E. Baird, Prof. Joel M. Gottesfeld, Liliane A. Dickinson, Richard J. Gulizia, Prof. Barbara J. Graves and Prof. Donald E. Mosier

(Dickinson, L.A.; Gulizia, R.J.; Trauger, J.W.; Baird, E.E.; Graves, B.J.; Mosier, D.E.; Gottesfeld, J.M.; Dervan, P.B., *in preparation*).

Part III of this chapter describes work carried out in collaboration with Eldon E. Baird, Prof. Peter B. Dervan, John Long and Prof. Joel M. Gottesfeld

Introduction

Small molecules that target specific DNA sequences offer a potentially general approach to the control of gene expression. Ligands designed for therapeutic applications must bind a predetermined DNA sequence with high affinity and permeate living cells.[5] Synthetic polyamides containing *N*-methylimidazole and *N*-methylpyrrole amino acids have an affinity and specificity for DNA comparable to naturally occurring DNA-binding proteins.[1,2] This chapter describes projects that have investigated the ability of polyamides designed to bind proximal to transcription factor binding sites within the promoter region of a target gene to permeate cells and regulate gene expression. In Part I of this chapter, an eight-ring hairpin polyamide targeted to a specific region of the transcription factor TFIIIA binding site is shown to inhibit 5S RNA gene transcription by RNA polymerase III in *Xenopus* kidney cells. This study provided the first demonstration that pyrolle-imidazole polyamides are cell-permeable and can regulate gene expression in cells. The question arose whether polyamides can permeate human cells and specifically regulate genes transcribed by RNA polymerase II. Part II of this chapter addresses this question by investigating the ability of hairpin polyamides to inhibit HIV-1 transcription in cell-free assays and viral replication in human lymphocytes. Part III of this chapter describes initial studies directed toward using polyamides to inhibit transcription of the human breast cancer oncogene HER2/neu.

Two approaches to the development of synthetic transcriptional antagonists have been reported.[3-8] Oligodeoxyribonucleotides that recognize the major groove of DNA through triple helix formation bind to a broad range of sequences with high affinity and specificity.[3,4] Although oligonucleotides and their analogs interfere with gene expression,[5,6] the triple helix approach is limited to purine tracts and suffers from poor cellular uptake. There are a few examples of cell-permeable carbohydrate-based ligands that interfere with transcription factor function,[6,7] but oligosaccharides cannot currently recognize a broad range of DNA sequences.

Pyrrole-imidazole polyamides represent the only class of small molecules so far that can bind a wide range of predetermined DNA sequences. DNA recognition depends on side-by-side amino acid pairings in the minor groove.[9,10] A pairing of imidazole (Im) opposite pyrrole (Py) targets a G•C base pair, and Py/Im targets C•G.[9,10] A Py/Py combination is degenerate and targets both T•A and A•T base pairs.[9-12] The generality of the pairing rules has been demonstrated by targeting a variety of 5–16 base pair[9-17] sequences and is supported by NMR and X-ray structural data.[18,19]

Part I: Inhibition of 5S RNA Gene Transcription

The relatively small number of genes transcribed by RNA polymerase III provides a simple system in which to explore the selectivity and efficiency of polyamides as regulators of gene expression. Well established methods exist for assessing in live cells the status of RNA polymerase III transcription complexes on the genes encoding the small 5S ribosomal RNA.[20] We report here that an eight-ring hairpin polyamide targeted to a specific region of the transcription factor TFIIIA binding site interferes with 5S RNA gene expression in *Xenopus* kidney cells. These results suggest that pyrrole-imidazole polyamides are cell-permeable and can inhibit the transcription of specific genes.

The initial step in activation of 5S RNA gene transcription is binding of the 5S RNA gene-specific transcription factor TFIIIA to a 50 base pair "internal control region" (ICR) within the coding sequence of the gene.[21,22] Subsequently, TFIIIB, TFIIIC and finally RNA polymerase III are recruited. Given this mechanism, inhibition of the DNA-binding activity of TFIIIA will block 5S RNA gene expression.

Biochemical studies of the nine-zinc finger DNA-binding domain of TFIIIA have led to a model of the TFIIIA-DNA complex. Zinc fingers 1–3, 5, and 7–9 bind the internal control region (ICR) of the gene through base-specific interactions in the major groove.[23-27] Fingers 4 and 6 are essential for high affinity DNA binding[23] and bind in or across the minor groove[25,26] (Fig. 1). Since polyamides bind in the minor groove and have

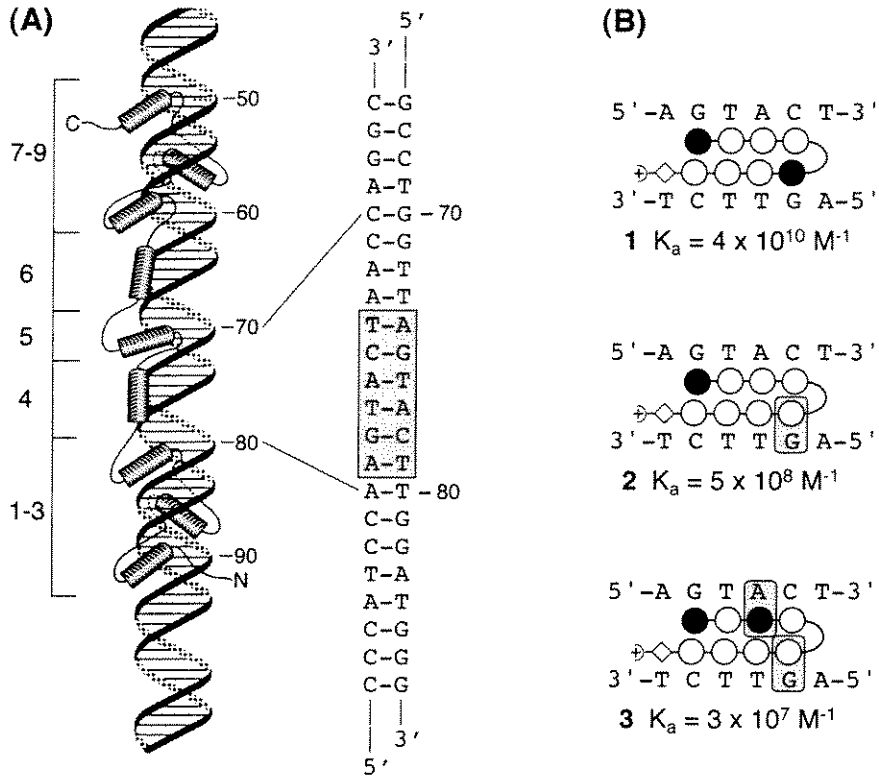


Figure 1. (A) Model of the nine-zinc finger protein TFIIIA in complex with the 5S rRNA gene internal control region. The DNA sequence recognized by zinc finger 4 is shown, with the six base pair polyamide **1** binding site 5'-AGTACT-3' highlighted. (B) Binding models and experimentally determined association constants for polyamides **1**, **2** and **3** in complex with 5'-AGTACT-3'. Hydrogen bond mismatches are indicated with shading.

been shown not to affect binding of a major groove-binding transcription factor,[28] we chose to target the binding site of finger 4. Polyamide ImPyPyPy-γ-ImPyPyPy-β-Dp (**1**) (β = β-alanine, γ = γ-aminobutyric acid) was designed to bind the 6 base pair DNA sequence 5'-AGTACT-3' within the finger 4 binding site (Fig 1A). As described in Chapter Four, polyamide **1** selectively binds this sequence with an association constant, K_a, of 4 x 10^{10} M^{-1}, an affinity higher than that of TFIIIA for its 50 base pair site (K_a ~ 1 x 10^9 M^{-1}).[25] Mismatch polyamides **2** and **3** have 100-fold (K_a = 5.0 x 10^8 M^{-1}) and 1,000-fold (K_a = 3.0 x 10^7 M^{-1}) lower affinities, respectively, for the 5'-AGTACT-3' site (Fig. 1B). Polyamide **1** does not bind any sites within the coding region of the 5S RNA gene

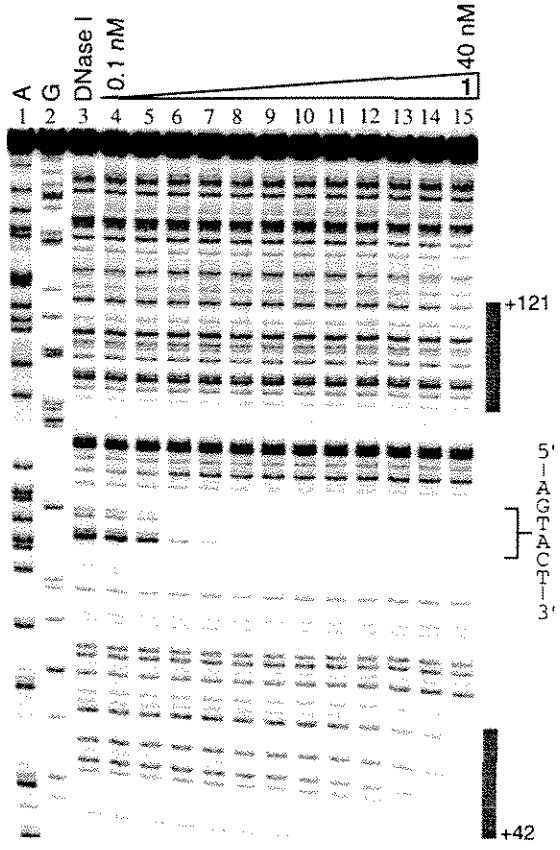


Figure 2. DNase I footprinting analysis of the binding of polyamide **1** to the 5S RNA gene. A polyamide stock solution or H_2O was added to an assay buffer containing radiolabeled restriction fragment, affording final solution conditions of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$ and pH 7.0. This experiment was carried out using a total solution volume of only 40 μL and underestimates the actual association constant by about 6-fold.

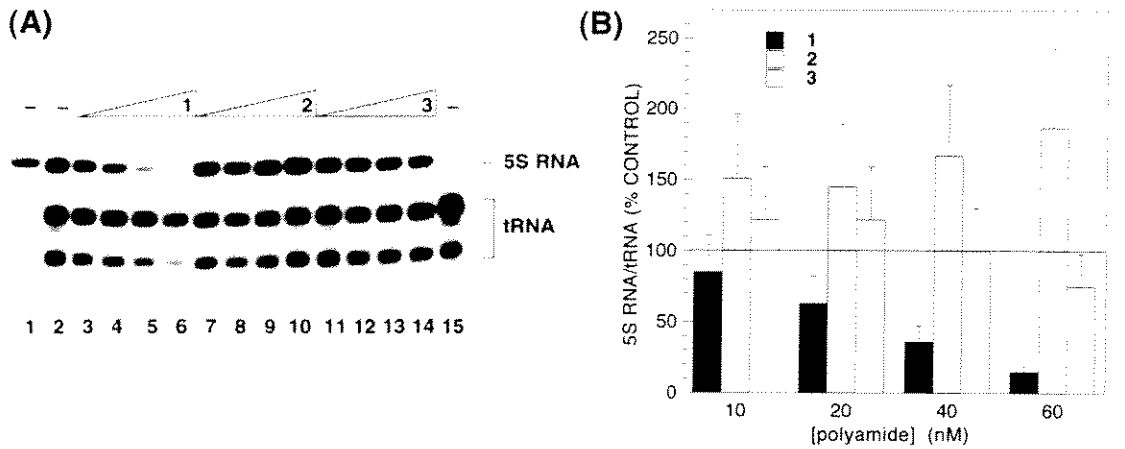


Figure 3. Inhibition of 5S RNA gene transcription *in vitro*. (A) DNA templates were incubated with polyamide for 30 minutes prior to the addition of TFIIIA, a cytoplasmic extract prepared from unfertilized *Xenopus* eggs, and labeled and unlabeled nucleoside triphosphates. Reactions contained either the somatic-type 5S RNA gene (lane 1), a $tRNA^{tyrD}$ gene (lane 15), or a mixture of both genes (lanes 2-14). Reactions contained the following final concentrations of the indicated polyamide: 0 nM, lanes 1, 2, 15; 10 nM, lanes 3, 7, 11; 20 nM, lanes 4, 8, 12; 40 nM, lanes 5, 9, 13; 60 nM, lanes 6, 10, 14. Transcription reactions were for 1 hour at ambient temperature and RNA products were analyzed on a denaturing polyacrylamide gel. The positions of 5S RNA and tRNA primary and processed transcripts are indicated at the right of the autoradiogram. (B) Graphic representation of inhibition results expressed as the ratio of 5S RNA to tRNA transcription relative to the ratio obtained in the absence of polyamide. Error bars represent the estimated standard deviation.

besides the target site at concentrations up to 10 nM (Fig. 2). Polyamides were synthesized by solid phase methods as described.[29]

Inhibition of TFIIIA-DNA Interactions. We examined using a DNase I footprinting assay the effect of polyamide **1** on binding of TFIIIA to a restriction fragment isolated from a 5S RNA gene-containing plasmid. When 5 nM polyamide **1** was pre-incubated with the DNA target, binding of TFIIIA was inhibited by >90%.

Inhibition of Transcription. Transcription of the 5S RNA gene in an *in vitro* system was monitored in the presence of 10–60 nM concentrations of polyamide **1** (Fig. 3). In these experiments, polyamide **1** was added to a 5S RNA gene-containing plasmid prior to the addition of exogenous TFIIIA (12 nM) and a crude extract derived from

unfertilized *Xenopus* eggs. As a control, a tyrosine tRNA gene was included on a separate plasmid in these reactions. The tRNA gene is transcribed by RNA polymerase III, but its activation is not controlled by TFIIIA. The tRNA gene has an upstream binding site for **1**, but lacks a predicted protein-polyamide interaction. Both genes are actively transcribed in this system, either individually (lanes 1 and 15) or in mixed template reactions (lane 2). Addition of 60 nM polyamide **1** inhibits 5S rRNA gene transcription by >80% (lane 6). The targeted 5S RNA gene is inhibited approximately 10-fold more effectively than the control tRNA gene.

Mismatch polyamides **2** and **3** do not inhibit 5S RNA transcription at concentrations up to 60 nM. If the TFIIIA-DNA complex is first allowed to form, 30 nM polyamide **1** added, and the mixture incubated for 90 minutes prior to adding egg extract, efficient inhibition (80%) of 5S RNA transcription is also observed. Shorter incubation times result in less inhibition. The required incubation time of 90 minutes is similar to the measured half-life of the TFIIIA-DNA complex[27] and supports that **1** forms a more stable complex with DNA than does TFIIIA.

The effect of the polyamides on 5S gene transcription *in vivo* was monitored (Fig. 4). *Xenopus* kidney-derived fibroblasts were grown with polyamide **1** in the culture medium at concentrations of 0.1–1 μM for various times. We found that **1** at concentrations of up to 1 μM were not toxic, as measured by cell density, if growth was limited to less than 72 hours. Nuclei were prepared from cells by hypotonic lysis and equivalent amounts of the isolated nuclei from control and treated cells were used as templates for transcription with exogenous RNA polymerase III and labeled and unlabeled nucleoside triphosphates. This experiment monitors the occupancy of class III genes with active transcription complexes.[20] 5S RNA transcription can easily be assessed since the repetitive 5S genes give rise to a prominent band on a denaturing polyacrylamide gel. Concentrations of **1** as low as 100 nM have a pronounced and selective effect on 5S transcription. At higher polyamide concentration, a general decrease in the transcriptional

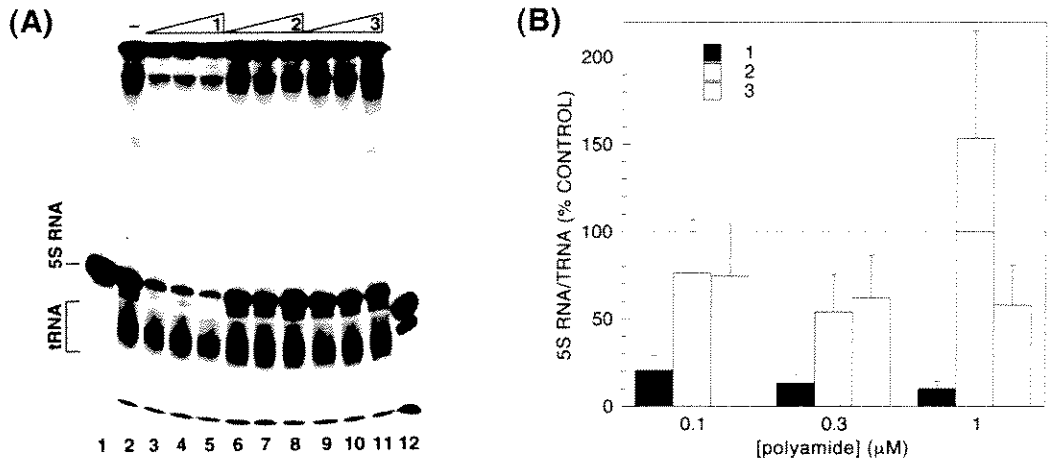


Figure 4. Inhibition of 5S RNA gene transcription complex formation *in vivo*. (A) Nuclei were prepared from *Xenopus* kidney-derived fibroblasts grown in culture in the presence or absence of polyamides. Polyamides were included in culture medium (in 2.5 mL of media per 25 cm^2 flask) for 24 h prior to harvesting cells and isolation of nuclei. Equal amounts of nuclei (containing 5 μg of DNA) were incubated for 2 h in 20 μL reactions containing *Xenopus* RNA polymerase III and labeled and unlabeled nucleoside triphosphates. RNA was isolated from these reactions and analyzed on a denaturing polyacrylamide gel. Lane 1, 5S RNA marker, lane 12, $tRNA^{tyrD}$ gene marker, lane 2, cells not exposed to polyamide, lanes 3-11, cells exposed to the indicated polyamides at the following concentrations: 100 nM, lanes 3, 6, 9; 300 nM, lanes 4, 7, 10; 1 μM, lanes 5, 8, 11. (B) Graphic representation of inhibition results expressed as the ratio of 5S RNA to tRNA transcription relative to the ratio obtained in the absence of polyamide. Error bars represent the estimated standard deviation.

activity of the nuclei is observed; however, at each concentration tested, the effects of the polyamide are much greater on 5S RNA transcription than on tRNA transcription. Having established that nearly maximal inhibition of 5S RNA transcription is achieved with 1 μM polyamide **1**, we monitored nuclear transcription after various times of cell growth in the presence of the polyamide. No inhibition is observed for zero time incubation with **1** at 1 μM concentration, indicating that disruption of transcription complexes does not occur during or after the isolation or work-up of cell nuclei. Statistically equivalent levels of 5S transcription inhibition were observed when the cells were exposed to **1** for 24, 48, or 72 hours.

These nuclear transcription experiments suggest that polyamide **1** is able to enter cells, transit to the nucleus and disrupt transcription complexes on the chromosomal 5S RNA genes. To rule out the possibility that the observed inhibition is due to some nonspecific toxicity of the polyamide rather than to direct binding to the 5S RNA genes, the effects of mismatch polyamides **2** and **3** in the nuclear transcription assay were monitored. Only a small effect on 5S RNA synthesis relative to tRNA synthesis is observed with 1 μM of the mismatch polyamides **2** or **3** in the culture medium for 24 hours. This result suggests that the general inhibition of transcription observed with high concentrations of polyamide **1** may be a secondary effect of the inhibition of 5S RNA synthesis *in vivo* rather than the result of non-specific polyamide interactions. Polyamide **2** effects a small enhancement of 5S RNA transcription *in vitro* and *in vivo*, indicating that polyamides may be able to upregulate transcription in certain cases.

Selective inhibition of RNA polymerase III transcription of a 5S RNA gene relative to a tRNA gene is a first step towards exploring the potential of polyamides as artificial regulators of gene expression. It remained to be determined if a broad panel of genes can be selectively targeted within a variety of cell types.

Part II: Inhibition of HIV-1 Transcription and Replication

At least six cellular DNA-binding transcription factors are required by human immunodeficiency virus type 1 (HIV-1) for efficient RNA synthesis. Two pyrrole-imidazole polyamides were designed to bind DNA sequences proximal to binding sites for the cellular transcription factors Ets-1, LEF-1 and TBP. We report here that both synthetic ligands were found to specifically inhibit protein binding as well as HIV-1 transcription in cell-free assays. When used in combination, the synthetic ligands effectively inhibit virus replication in isolated human peripheral blood lymphocytes. The ability of small molecules to target predetermined DNA sequences located within RNA polymerase II promoters

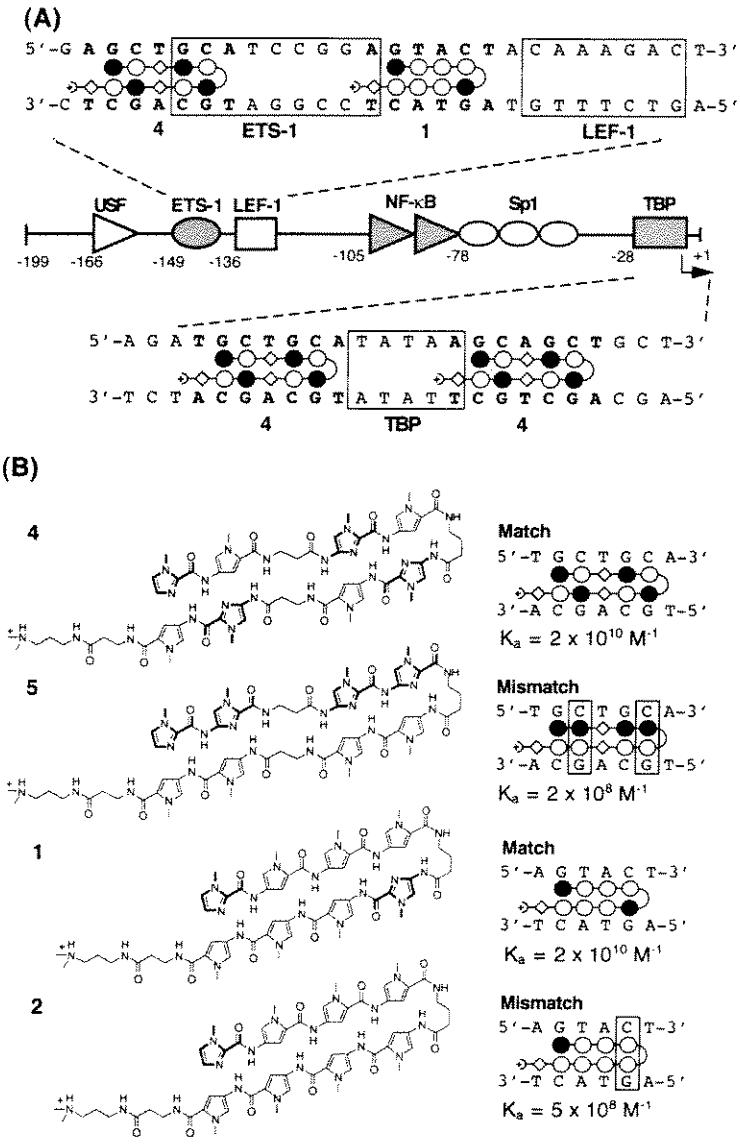


Figure 5. (A) DNA binding sites for hairpin polyamides designed to target the HIV-1 enhancer and promoter. Shaded and unshaded circles represent imidazole (Im) and pyrrole (Py) rings, respectively, curved lines represent γ-aminobutyric acid (γ) and the diamonds represent β-alanine (β). Schematic of the HIV-1 enhancer and promoter (nucleotide positions -170 to the transcription start site at +1). The binding sites for the transcription factors USF, Ets-1, LEF-1, NF-kB, SP1 and TFIID (TBP) are indicated. The binding sites for polyamides **1** and **4** are indicated. (B) Structures of polyamides ImPy-β-ImPy-γ-ImPy-β-ImPy-β-Dp (**4**), ImIm-β-ImIm-γ-PyPy-β-PyPy-β-Dp (**5**), ImPyPyPy-γ-ImPyPyPy-β-Dp (**1**) and ImPyPyPy-γ-PyPyPyPy-β-Dp (**2**), where Dp = dimethyaminopropylamide. Binding models and measured association constants (K_a) are shown. Mismatched base pairs are highlighted. Synthesis of the polyamides was by solid phase methods as described.[29]

suggests a general approach for regulation of gene expression, as well as a mechanism for the inhibition of viral replication.

The HIV-1 enhancer/promoter element contains binding sites for the cellular transcription factors USF, Ets-1, LEF-1, NF-kB, Sp1 and TBP (Fig. 5).[30-34] These transcription factor binding sites are not optimal polyamide target sequences because they are found in the promoters of many cellular protein-coding genes. However, the sequences immediately flanking these transcription factor binding sites are often conserved for a particular gene providing an address for gene-specific targeting. For example, the 5'-AGTACT-3' sequence and three 5'-(A,T)GC(A,T)GC(A,T)-3' sequences immediately adjacent to the transcription factors Ets-1, LEF-1 and TBP are conserved for most reported strains of HIV-1.[35,36] Although the propensity for mutation at these sites is unknown, allowed sequence changes in the promoter could be targeted with new polyamides designed by the pairing rules. In order to regulate HIV-1 transcription specifically, polyamides were designed to bind immediately adjacent to rather than directly to the binding sites for three transcription factors with known minor groove contacts, LEF-1, Ets-1 and TBP.

The design of the hairpin polyamide ImPy-β-ImPy-γ-ImPy-β-ImPy-β-Dp (**4**) was described in Chapter Five. Quantitative footprint titration experiments reveal that **4** binds to both 5'-(A,T)GC(A,T)GC(A,T)-3' sites adjacent to the TBP binding site as well as the 5'-AGCTGCA-3' match site adjacent to the binding site for Ets-1 with an equilibrium association constant, K_a, of 2 x 10^{10} M^{-1} (Figs. 5 and 6). A mismatch control polyamide ImIm-β-ImIm-γ-PyPy-β-PyPy-β-Dp (**5**) that differs only in the placement of the imidazole and pyrrole amino acids binds the 5'-(A,T)GC(A,T)GC(A,T)-3' sites with 100-fold reduced affinity relative to polyamide **4**. The LEF-1 and Ets-1 binding sites are immediately flanked by the sequence 5'-AGTACT-3'. According to the pairing rules, this sequence will be bound by polyamide **1** of sequence composition ImPyPyPy-γ-ImPyPyPy-β-Dp (Fig. 5).[7] Quantitative footprint titration experiments reveal that polyamide **1** binds the 5'-AGTACT-3' site between the Ets-1 and LEF-1 binding sites with

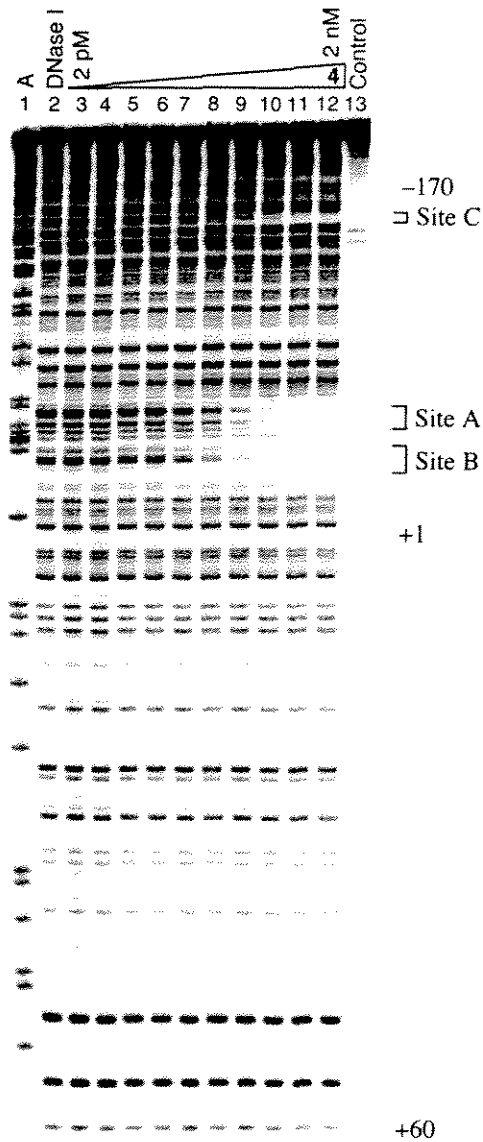


Figure 6. DNase I footprinting analysis of the binding of polyamide **4** to the HIV-1 promoter. Sites A and B are the TATA box-proximal sequences 5'-TGCTGCA-3' and 5'-AGCAGCTGCT-3', respectively, and site C is the Ets-1 binding site-proximal sequence 5'-AGCTGCA-3'. A polyamide stock solution or H_2O was added to an assay buffer containing radiolabeled restriction fragment, affording final solution conditions of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$ and pH 7.0.

$K_a = 2 \times 10^{10}\ M^{-1}$. Mismatch polyamide **2** binds this sequence with >100-fold reduced affinity.

Inhibition of Protein-DNA Interactions. Binding of the TBP subunit of TFIID in the minor groove nucleates assembly of the pol II transcription machinery for TATA-containing genes.[37-41] TBP binds the HIV-1 TATA element with a K_d of ~1-3 nM.[39] DNase I footprinting assays reveal that polyamide **4** inhibits TBP binding at a concentration of 2.5 nM, while mismatch polyamide **5** fails to inhibit TBP binding. A gel mobility shift assay reveals that polyamide **4** inhibits TBP binding to a double-stranded oligonucleotide corresponding to the HIV-1 TATA box region, while no inhibition is observed for control polyamide **5** (Fig. 7A). Additionally, polyamide **4** does not inhibit TBP binding to the TATA box region of the adenovirus major late (AdML) promoter (5'-GGGGGCTATAAAAGGGGGT-3') which contains mismatch flanking sequences. The half-life of the polyamide **4**-DNA complex was determined by competition experiments to be in excess of 2.5 hours.

The Ets-1 recognition site in the HIV-1 enhancer is flanked by binding sites for polyamides **1** and **4** (Fig. 5). The DNA-binding domain of Ets-1 belongs to the class of 'winged' helix-turn-helix proteins (wHTH), where the β-sheet wings extend into the minor groove while helices H2 and H3 contact the GGA core consensus in the major groove.[42,43] Polyamide **4** is predicted to bind immediately upstream and polyamide **1** is predicted to bind immediately downstream of the Ets-1 binding site. The Ets-1 DNA binding domain (331-416) binds to the HIV-1 enhancer at nanomolar concentration.[44] Gel mobility shift experiments revealed that when polyamides were preincubated with the labeled double-stranded HIV-1 oligonucleotide, polyamide **1** had no effect on Ets-1 DNA-binding, even at a concentration as high as 400 nM (Fig. 7B, lanes 3-5). Polyamide **4** present at > 5nM concentration prevented formation of the Ets-1/DNA complex (Fig. 7B, lanes 9-11). When both polyamides were combined, the degree of inhibition was very similar to polyamide **4** alone. The two mismatch polyamides **2** and **5** did not prevent complex formation at up to

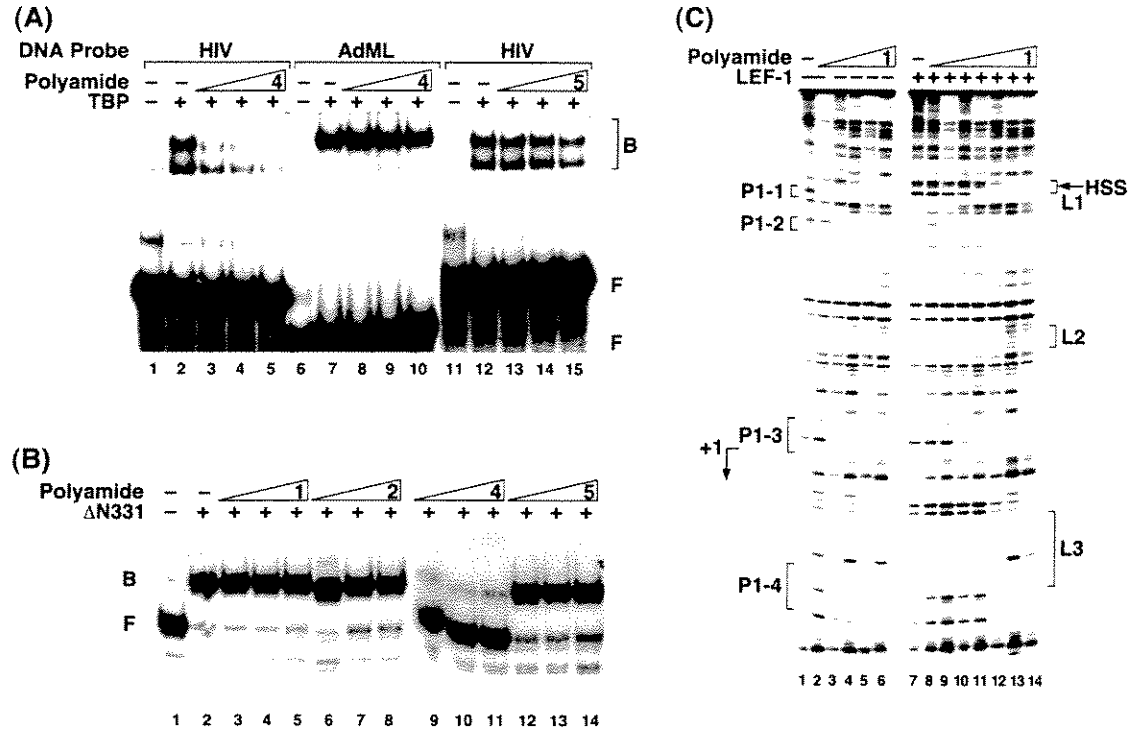


Figure 7. Inhibition of transcription factor binding to the HIV-1 promoter. (A) Gel mobility shift assay for inhibition of TBP-DNA binding with polyamides **4** and **5**. Radiolabeled HIV-1 TATA box (lanes 1-5 and 11-15) or adenovirus major late promoter TATA box duplex oligonucleotides (lanes 6-10) were incubated with the following concentrations of polyamide prior to addition of rhTBP, where indicated (+): no polyamide, lanes 1, 2, 6, 7, 11, 12; 50 nM polyamide, lanes 3, 8, 13; 100 nM, lanes 4, 9, 14; 200 nM, lanes 5, 10, 15. After an additional 30 minutes incubation, samples were subjected to electrophoresis. The positions of free DNA probes (F) and the monomer and dimer TBP-DNA complexes (B) are indicated alongside the figure. (B) Differential effect of polyamides **1** and **4** on Ets-1 DNA-binding. An oligonucleotide corresponding to the Ets-1 binding site with flanking regions within the HIV-1 enhancer/promoter (position -160 to -117) was used as a probe in the gel mobility shift assays. Polyamides were preincubated with the DNA before addition of ΔN331, where indicated. The concentrations of polyamides were 100 nM (lanes 3, 6, 9, 12), 200 nM (lanes 4, 7, 10, 13) and 400 nM (lanes 5, 8, 11, 14). ΔN331 protein was added at a concentration of 12 nM. Positions of the free probe (F) and bound probe (B) are indicated. (C) Inhibition of LEF-1 binding to the HIV-1 enhancer by polyamide **1**. The 3'-^{32}P end-labeled HIV-1 LTR restriction fragment was incubated with 10 pM–10 nM polyamide **1** for 45 minutes prior to the addition of LEF-1 DBD to a final concentration of 8 nM (lanes 7 to 14). After an additional 45 minutes, the samples were subjected to DNase I digestion and analyzed by gel electrophoresis. The polyamide concentrations were: no polyamide (lanes 1 and 7), 10 pM (lanes 2 and 8), 30 pM (lane 9), 0.1 nM (lanes 3 and 10), 0.3 nM (lane 11), 1 nM (lanes 4 and 12), 3 nM (lanes 5 and 13) and 10 nM (lanes 6 and 14).

400 nM concentrations (Fig. 7B, lanes 6-8, 12-14). DNase I footprinting reveals that polyamide **4** prevents Ets-1 DNA-binding while polyamide **1** co-occupies the DNA with Ets-1.

LEF-1 is a member of the HMG family of minor-groove binding proteins.[45-47] In addition to acting as an architectural transcription factor, LEF-1 possesses a strong trans-activation domain and has been shown to be essential for viral transcription and replication in lymphoid cells.[46,47] A DNase I footprinting assay reveals that LEF-1 binding is inhibited by 50% at a polyamide **1** concentration of approximately 60 pM (Fig. 7C). Inhibition was observed either by adding the polyamide to the DNA before LEF-1 or after preincubation of the DNA with LEF-1, consistent with the relative association constants for the two binding reactions. Mismatch polyamide **2** fails to inhibit LEF-1 binding.

Inhibition of Transcription. The effects of polyamides **4** and **5** on HIV-1 transcription were tested in an *in vitro* transcription assay with a HeLa cell nuclear extract. Polyamide **4**, which targets the HIV-1 TATA box and the Ets-1 site, inhibits basal transcription from the HIV-1 promoter mediated by the general transcription factors Sp1 and pol II.[30] HIV-1 transcription was inhibited 50% in the presence of between 50 and 100 nM polyamide **4** in several independent experiments. Polyamide **4** does not inhibit transcription from the CMV major intermediate early promoter (MIEP), which contains a mismatched TATA-flanking sequence (5'-GAGGTCTATATAAGCAGA-3'). The mismatch polyamide **5** does not inhibit transcription from either promoter. We performed titrations of polyamide **4** over a wide range of concentrations (1 to 200 nM) with both the HIV-1 and CMV templates in the same reaction. Under these conditions, we observed 50% inhibition of HIV-1 transcription at 30 nM polyamide **4** which corresponds to a 3–5-fold excess of polyamide over binding sites. No inhibition of CMV transcription was observed even at the highest concentrations of polyamide **4**.

We next tested the effects of polyamides **1** and **2** on HIV-1 transcription in an *in vitro* system consisting of a cell-free extract prepared from cultured human lymphoid H9

cells supplemented with HeLa cell extract. The H9 extract contains high levels of LEF-1 protein but was found to support only low levels of transcription, suggesting a limited amount of other transcription components in this extract. We supplemented the H9 cell extract with small amounts of a HeLa cell-derived nuclear extract in order to obtain high levels of transcription. The mixture of H9 and HeLa cell extracts stimulates HIV-1 transcription 2.5–3-fold over the level of transcription observed with the HeLa extract alone. Immunodepletion of LEF-1 protein from the H9 extract abolishes this activated transcription. Polyamide **1** inhibits HIV-1 transcription in this system with a 50% reduction of transcription observed at 10-30 nM polyamide. Polyamide **2** fails to inhibit HIV-1 transcription in the LEF-1-depleted extract. We also monitored the activity of the CMV MIEP in both the mock-depleted and LEF-1 depleted H9 cell extract, with the result that LEF-1 depletion had no effect on CMV transcription. As additional controls, we tested the effect of the mismatch polyamide **2** on HIV-1 transcription and polyamides **1** and **2** on CMV transcription. No potential binding sites for either polyamide **1** or **2** are present in the CMV MIEP sequence. As expected, polyamide **1** fails to inhibit CMV transcription and polyamide **2** fails to inhibit either CMV or HIV-1 transcription.

Notably, polyamide **1** does not inhibit basal transcription with the HeLa nuclear extract although binding sites for this polyamide are present at the start-site for transcription and within the HIV-1 mRNA coding sequence present in plasmid pHIV-CAT. These observations suggest that RNA polymerase II can transcribe DNA with a polyamide bound in the minor groove and that polyamides are only inhibitory to transcription when these compounds interfere with the DNA binding activity of a required transcription factor.

Inhibition of Viral Replication. We set out to determine whether the polyamides could inhibit HIV-1 transcription in virus-infected human cells. Since there are multiple spliced and unspliced species of HIV-1 RNA with different turn-over kinetics,[48] we chose not to monitor viral transcription by measuring levels of RNA. Instead we examine the effects of the polyamides on the levels and kinetics of HIV-1 replication in

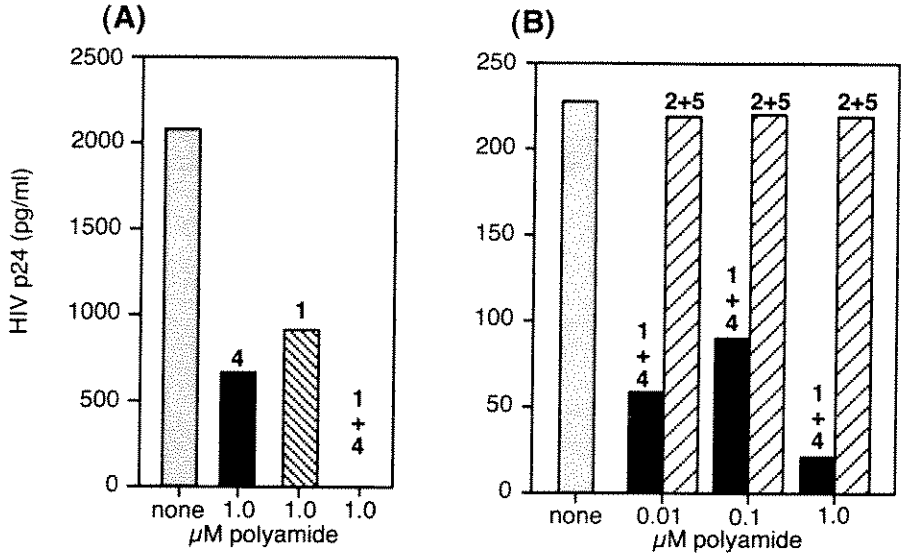


Figure 8. Inhibition of HIV-1 replication in peripheral blood mononuclear cells by polyamides. Panels depict experiments in which polyamides alone or in combination were added to cultures of human peripheral blood mononuclear cells (PBMC) stimulated 3 days earlier with phytohemagglutinin and interleukin-2. PBMC cultures were infected with the primary HIV-1 isolate WEAU 1.6 (kindly provided by G. Shaw), and virus replication measured by release of p24 capsid antigen into the medium that was detected by antibody-capture ELISA (Dupont). Each experiment involved a separate PBMC donor. Values shown are for day 6 or day 8 after virus infection. When two polyamides were combined, the concentration shown is for each component of the mixture. Assays of p24 concentration were performed in duplicate and showed less than 5% variation from the mean value reported.

isolated human peripheral blood mononuclear cells (PBMC) in culture. PBMC were infected with the T cell-tropic HIV-1 strain WEAU1.6,[49,50] or the macrophage-tropic strain SF162.[51,52] Polyamides were added to the culture medium and the levels of HIV-1 p24 viral capsid protein released into the culture media (primarily as virions) was determined on subsequent days using a standard ELISA method after 4-10 days of culture. Assays of HIV-1 replication representative of five replicate experiments with five human PBL donors are shown in Figure 8. In control PBMC cultures with no added polyamide, viral replication resulted in increasing p24 levels between day 4 and day 10 of culture. Polyamide **4** at 1 μM concentration caused an 80% reduction in virus, while polyamide **1** at 1 μM concentration caused a 60% reduction. The combination of polyamides **1** and **4** at 1

μM each can act in synergy to reduce viral p24 levels to below the threshold of detection (<10 pg/ml; greater than 99.9% inhibition of viral replication), and was as effective as 1 μM azidothymidine (AZT) in blocking HIV-1 replication. The macrophage-tropic SF162 isolate, which replicates in both macrophages and CD4+ T lymphocytes, was more difficult to inhibit, but the combination of 1 μM polyamides **4** and **1** was able to reduce and eventually block its replication (data not shown). These results demonstrate that cell-permeable synthetic DNA ligands can effectively inhibit HIV-1 replication in isolated human lymphocytes in culture and suggest that inhibition of virus replication is due to inhibition of transcription factor-DNA interactions and gene transcription of viral and possibly cellular genes by RNA polymerase II. The observed polyamide inhibition of virus replication is likely due to interference with the DNA-binding activities of TBP and Ets-1 by polyamide **4** and the binding activity of LEF-1 by polyamide **1**. The inhibitory effects of polyamides singly or in combination was not due to obvious toxicity. No significant decrease in cell viability was apparent in PBMC cultures treated with polyamides **4** and **1** for 10 days, in contrast to 90% mortality observed for PBMC cells treated with 1 μM AZT for the same period. Cell recovery was not impacted by polyamide treatment, but was reduced by AZT treatment.

The combination of polyamides **1** + **4** inhibited HIV-1 replication at 10 nM to 1 μM concentration, but the closely related polyamides **2** + **5** did not. The simplest explanation for this finding is that polyamides **1** + **4** inhibited HIV-1 RNA transcription in cells as well as in the *in vitro* assays, but it is possible that inhibition of cellular genes involved in T cell activation could have an indirect effect on HIV-1 replication. To assess this possibility, we performed an RNase protection assay for transcripts of a number of cytokine genes, including IL-2, IL-5 and IL-13 which differ in the target sequences flanking the TATA

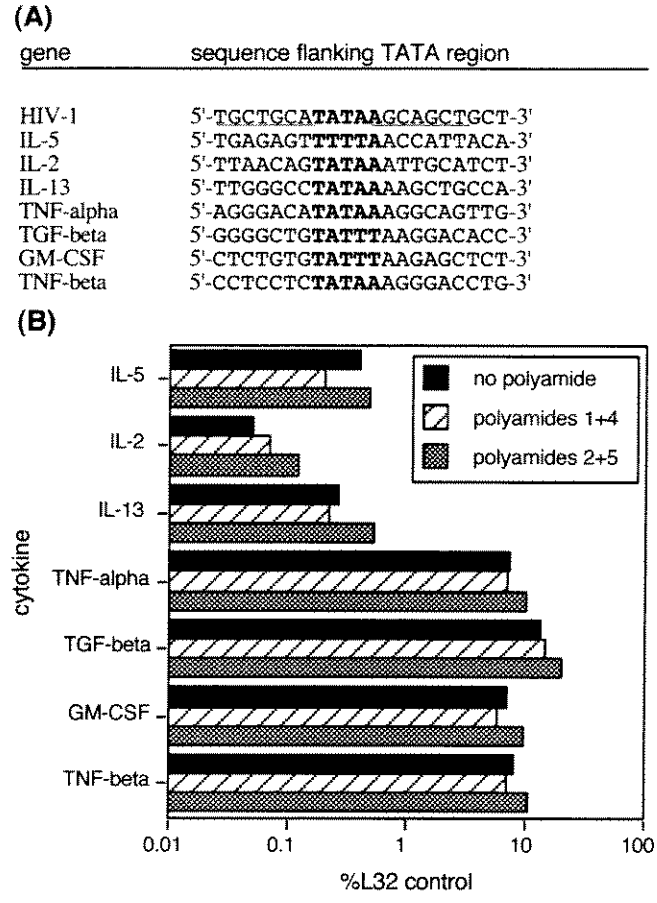


Figure 9. Lack of inhibition of cytokine gene expression with polyamides targeted to the HIV-1 promoter and enhancer sequence. (A) The sequences of the TATA-box region (taken from GenBank listings) of each of the cytokine/growth factor genes examined are shown with the TATA box in bold and the binding sites for polyamide **4** underlined. Single base mismatches are indicated in lower case. (B) Human peripheral blood mononuclear cells were cultured under the same conditions used for HIV-1 infection. Cultures were either left untreated, or 10 μM of polyamides **1** + **4** or polyamides **2** + **5** added. After six days, cells were harvested, total RNA extracted, and a ribonuclease protection assay performed with riboprobes specific for the indicated cytokines as well as for CD4, CD8 and ribosomal protein L32. After digestion with RNase T1, protected fragments were separated by polyacrylamide gel electrophoresis, and the amount of labeled probe quantitated by PhosphorImager analysis. Data are expressed as the intensity of each cytokine RNA relative to the intensity of the ribosomal L32 RNA band to standardize for RNA loading. There was no difference in the intensity of CD4 and CD8 RNA bands between groups, indicating equivalent recovery of CD4 and CD8 T cells in all cultures. Similar results were seen when the cultured cells were analyzed after 10 days, although the levels of RNA for most cytokines had declined in both polyamide treated and untreated cells.

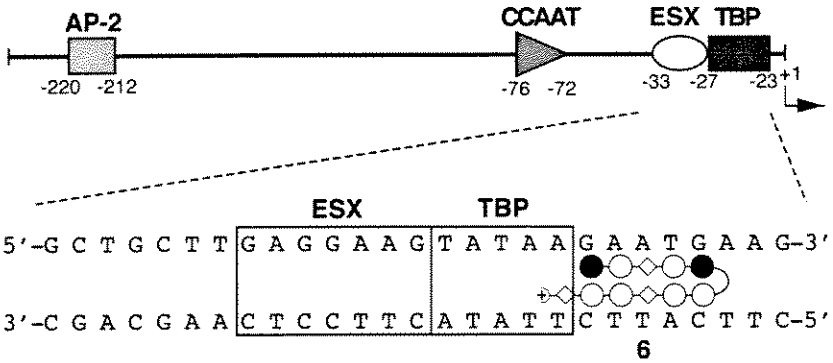


Figure 10. Binding site of polyamide **6** on the HER2/neu promoter. The polyamide recognizes a sequence proximal to the TBP binding site.

box. Four other cytokine genes that lack binding sites for either polyamide **1** or **4** in their promoters were also examined (Fig. 9). The results show that exposure of activated human PBMC to a combination of either polyamides **1** + **4** or **2** + **5** (1 μM each) for 6 days failed to inhibit cytokine RNA expression. This lack of inhibition of cytokine gene transcription suggests that the polyamides reduce virus replication in cells by a direct effect on HIV-1 transcription.

Part III: Transcription Inhibition of Human Breast Cancer Gene HER2/neu

The growth factor receptor HER2/neu is overexpressed in about 25% of human breast cancers.[53] HER2/neu overexpression is associated with tumor metastasis and a poor prognosis for the patient. Sequence-specific DNA-binding small molecules that are cell-permeable could potentially inhibit this overexpression, providing a new treatment for breast cancer. We report here initial results that demonstrate the ability of a hairpin polyamide designed to bind a DNA sequence proximal to the TBP binding site within the HER2/neu promoter to downregulate transcription of the HER2/neu gene.

The transcription factors TBP, ESX and AP-2 are known to bind within the promoter region of the HER2/neu gene and to activate gene expression.[54,55] The results reported in Part II of this chapter demonstrate that polyamides can inhibit TBP-DNA

complexes and TBP-dependent transcription *in vitro*. The hairpin polyamide **6** was synthesized to bind immediately downstream of the TBP binding site (Fig. 10). Quantitative DNase I footprinting experiments on a ^{32}P-end-labeled restriction fragment isolated from a HER2/neu gene-containing plasmid revealed that **6** binds its target sequence with an equilibrium association constant of 5 x 10^9 M^{-1}. The footprinting experiment indicated that this polyamide also binds the 5'-AGGAAGT-3' single-base pair mismatch sequence proximal to the ESX binding site with comparable affinity.

Preliminary experiments carried out by John Long and Joel Gottesfeld at the Scripps Research Institute show that polyamide **6** inhibits HER2/neu gene transcription by 50% at 100 nM concentration in HER2/neu-overexpressing human breast cancer cells.

Future studies will investigate the effects of additional polyamides that target the TBP and ESX binding sites, as well as polyamides that target other regulatory elements within the HER2/neu promoter such as the AP-2 binding site and the CCAAT box. The polyamides that provide optimal inhibition could then be tested in a mouse model of human breast cancer. The long-term view is that these studies will investigate whether polyamides could provide a new approach to the treatment of human breast cancer.

Future Plans. Our present studies have utilized polyamides designed to target DNA sequences 6–7 bp in length and have shown that these compounds are effective inhibitors of gene transcription in cell-free systems and viral replication in human cells. Because sequences of these lengths would be highly redundant in the human genome, it had seemed likely that these ligands would have deleterious effects on cell metabolism due to interference with the activity of cellular genes. However, the results described here indicate that a battery of polyamides which recognize 6–7 bp sequences will be sufficient for gene-specific regulation *in vivo*. It is interesting to compare these small molecule transcription factors to eukaryotic transcriptional regulatory proteins that also recognize multiple short sequences in tandem in order to increase functional specificity.[39-41] The

observations that polyamides do not interfere with pol II elongation, and that polyamides can bind simultaneously with certain major groove proteins should further enhance gene-specificity.[28,53] In addition, polyamides are not limited to 6-7 base-pair recognition. For example, polyamides of similar size to those described here have been shown to bind as cooperative dimers to sites 10-16 base pairs in length.[15-17,57] The polyamide binding site size required to elicit optimal biological function will be reported in due course.

The specific inhibition of genes transcribed by pol II represents an important first step toward asking whether cell-permeable small molecule transcription antagonists might regulate gene expression in complex organisms.[10,39,40] We have chosen TBP and sequences adjacent to the TATA element for inhibition of basal transcription by RNA polymerase II. Since most tissue-specific cellular genes and viral genes contain TATA elements, this approach may be generally applicable for the inhibition of most target genes.

References

(1) Trauger, J.W., Baird, E. E.; Dervan, P.B. *Nature* **1996**, *382*, 559–561.

(2) Turner, J. M., Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 7636–7644.

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

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

(5) Maher, J. L.; Dervan, P. B.; Wold, B. *Biochemistry* **1992**, *31*, 70–81.

(6) Duvalvalentin, G.; Thuong, N.T.; Helene, C. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 504–508.

(7) Ho, S.N.; Boyer, S. H.; Schreiber, S.L.; Danishefsky, S. J.; Crabtree, G. R. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 9203–9207.

(8) Liu, C. *et al. Proc. Natl. Acad. Sci. USA* **1996**, *93*, 940–944.

(9) Wade, W.S., Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1992**, *114*, 8784–8794.

(10) Mrksich, M. *et al. Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586–7590.

(11) Pelton, J.G.; Wemmer, D.E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723–5727.

(12) White, S.; Baird, E. E.; Dervan, P.B. *Biochemistry* **1996**, *35*, 12532–12537.

(13) Parks, M.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6147–6152.

(14) Kelly, J.J.; Baird, E. E.; Dervan, P.B. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6981–6985.

(15) Trauger, J.W.; Baird, E. E.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6160–6166.

(16) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1998**, *120*, 3534–3535.

(17) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Eur. J. Chem.* **1997**, *3*, 1600–1607.

(18) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646–650.

(19) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nature Struct. Biol.* **1998**, *5*, 104–109.

(20) Schlissel, M.S.; Brown, D.D. *Cell* **1984**, *37*, 903–913.

(21) Engelke, D.R.; Ng, S.-Y.; Shastry, B.S.; Roeder, R.G. *Cell* **1980**, *19*, 717–728.

(22) Wolffe, A.P.; Brown, D. D. *Science* **1988**, *241*, 1626–1632.

(23) Fairall, L.; Rhodes, D.; Klug, A. *J. Mol. Biol.* **1986**, *192*, 577–591.

(24) Sakonju, S.; Brown, D. D. *Cell* **1982**, *31*, 395–405.

(25) Clemens, K.R.; Liao, X.B.; Wolf, V.; Wright, P.E.; Gottesfeld, J.M. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 10822–10826.

(26) Hayes, J.J.; Tullius, T.D. *J. Mol. Biol.* **1992**, *227*, 407–417.

(27) Clemens, K.R. *et al. J. Mol. Biol.* **1994**, *244*, 23–35.

(28) Oakley, M. G.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10969–10975.

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

(30) Jones, K. A.; Peterlin, B. M. *Annu. Rev. Biochem.* **1994**, *63*, 717.

(31) Berkhout, B.; Jeang, K.-T. *J. Virol.* **1992**, *66*, 139–149.

(32) Kamine, J.; Subramanian, T.; Chinnadurai, G. *J. Virol.* **1993**, *67*, 6828–6834.

(33) Olsen, H. S.; Rosen, C. A. *J. Virol.* **1992**, *66*, 5594–5597.

(34) Sadie, M.R.; Benter, T.; Eong-Staal, F. *Science* **1988**, *239*, 910–913.

(35) Frech, K.; Brack-Werner, R.; Werner, T. *Virology* **1996**, *224*, 256–267.

(36) Estable, M.C. *et al. J. Virol.* **1986**, *70*, 4053–4062.

(37) Kim, Y.; Geiger, J. H.; Hahn, S.; Sigler, P. B. *Nature* **1993**, *365*, 512–520.

(38) Kim, J.L.; Nikolov, D. B.; Burley, S. K. *Nature* **1993**, *365*, 520–527.

(39) Burley, S. K.; Roeder, R. G. *Ann. Rev. Biochem.* **1996**, *65*, 769–799.

(40) Verrijzer, C. P.; Tjian, R. *Trends Bioc.* **1996**, *21*, 338–342.

(41) Roeder, R.G. *Trends Bioc.* **1996**, *21*, 327–335.

(42) Donaldson, L.W. *et al. EMBO J.* **1996**, *15*, 125–134.

(43) Kodandapani, R. *et al. Nature* **1996**, *380*, 456–460.

(44) Petersen, J.M. *et al. Science* **1995**, *269*, 1866–1869.

(45) Love, J. J. *et al. Nature* **376**, 791-795 (1995).

(46) Kim, J.; Gonzales-Scarano, F.; Zeichner, S.; Alwine, J. C. *J. Virol* **1993**, *67*, 1658–1662.

(47) Giese, K.; Kingsley, C.; Kirschner, J.; Grosschedl, R. *Genes Dev.* **1995**, *9*, 995–1008.

(48) Klotman, M. E.; Kim, S.; Buchbinder, A.; DeRossi, A.; Baltimore, D.; Wong-Staal, F. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5011–5015.

(49) Clark, S. J. *et al. N. Engl. J. Med.* **1991**, *324*, 954–960.

(50) Borrow, P. *et al. Nature Med.* **1997**, *3*, 205–211.

(51) Cheng-Mayer, C.; Seto, D.; Tateno, M.; Levy, J. A. *Science* **1988**, *240*, 80–82.

(52) Cheng-Mayer, C.; Weiss, C.; Seto, D.; Levy, J. A. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 8575–8579.

(53) Slamon, D.J.; Clark, G.M.; Wong, S.G.; Levin, W.J.; Ullrich, A.; McGuire, W.L. *Science* **1987**, *235*, 177.

(54) Chang, C.-H. *et al.*; Benz, C.C. *Oncogene* **1997**, *14*, 1617.

(55) Baert, J.-L.; Monte, D.; Musgrove, E.A.; Albagli, O.; Sutherland, R.L.; de Launoit, Y. *Int. J. Cancer* **1997**, *70*, 590.

(56) Bremer, R.E.; Baird, E.E.; Dervan, P.B. *Chem. Biol.* **1998**, *5*, 119–133.

(57) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Angewandte Chemie Int. Ed. Eng.* **1998**, *37*, 1421–1423.