Allosteric Inhibition of Zinc Finger Proteins by DNA Binding Polyamides

Thesis by

Doan H. Nguyen

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2002

(Defended April 2, 2002)

Acknowledgements

I would first like to thank my advisor Professor Peter Dervan for providing a research environment where I had the freedom to pursue so many projects independently. I have learned so much about how to organize and execute an investigation to the very highest standards. These are invaluable lessons which I will undoubtedly always carry with me. Thanks also go to my other thesis committee members: Professors Harry Gray, Rich Roberts and Steve Mayo for all of their time and efforts on my behalf. I had the pleasure of being a teaching assistant for Professor Dennis Dougherty and would especially like to thank him for having so much confidence in me as to let me give many of his lectures. It was an invaluable experience that I truly enjoyed and I have learned so much about teaching from you.

I would like to thank all the members of the Dervan group, both past and present, for making work a more stimulating and entertaining place to be over the past few years. There are some former group members, and now good friends, to whom I must give special thanks; Anna Mapp for being such a great bay-mate and role model, Ryan Bremer who taught me a tremendous amount about critical scientific thinking, Clay Wang for teaching me to never waiver from the right way to footprint, and Ulf Ellervik for not only good scientific discussions but most importantly for great ones about food and travel. I want to thank Victor Rucker for our interesting discussions over the past couple of years, for making work a hysterical place and for teaching me about patience and change. I feel so thankful to my classmates Nick Wurtz and Jason Belitsky who were both such supportive resources when I first joined the Dervan group. Nick, I have enjoyed our collaborations and many discussions about science, careers and life in general. Jason, you

have been a truly amazing friend and colleague. I am indebted to you for your friendship, patience, insight and humor. I admire you as one of the best scientist I know and look forward to our continued friendship and to your every success.

Beyond the lab there are also many thanks to be given. First, to Dian Buchness who made my life so much more hassle free and was always there to listen. To the 1st year lunch crew; Soojin, Mike, Carlos, Liz, Isaac, Jeff, Niki, Catherine, Jason, and Shannon, I can't believe we made it this long despite so many people having to move to other places! I want to thank my good friend Niki Zacharias for always making sure we had a life beyond the lab and for all the great food over the years. Liz and Isaac Carrico, thank you both so much for your friendship, great discussions and all the fond memories of our trips together; we look forward to all the other trips with you in the future. I need to thank my longtime friends, Cara, Lisa, Barb and Melissa for always being so supportive even from such a long way away. My parents, I could never adequately thank them for having taught me the true value of education, loved me unconditionally and given me more than I ever need. I could not have gotten here without you both and am forever in your debt.

Finally and most importantly, I need to thank my husband Peter. You have been there as my strongest supporter and best friend for every step of this long road. You have always made me laugh and focus on the truly important things in life when times where tough and everything may have seemed bleak. I feel grateful everyday for your unconditional love and for always being able to come home to you. I am blessed to have you in my life and to share the future with you. I love you very much.....thank you.

Abstract

Small molecules that can bind selectively to any predetermined DNA sequence in the human genome could potentially be powerful tools for molecular biology and human medicine. Polyamides containing *N*-methylimidazole (Im) and *N*-methylpyrrole (Py) are small molecules that bind DNA according to a set of "pairing rules" with affinities and specificities similar to many naturally occurring DNA binding proteins. The study of DNA binding polyamides is further expanded by the development of new monomer pairings and new synthetic methods which allow access to polyamides with varying truncated tails. A new pairing of *N*-methylpyrazole with *N*-methylpyrrole increased specificity substantially without loss in affinity. This result indicates that other ring positions, besides the 3-position, can also greatly impact DNA recognition properties. Polyamides having truncated tails are shown to bind DNA with greater generality at the tail positions while maintaining high affinity, and may allow the targeting of a larger number of biologically relevant DNA sequences. Small molecules that bind DNA may offer a general approach to the chemical down- or up-regulation of gene expression by the inhibition or recruitment of transcription factors, respectively. Polyamide-peptide conjugates were synthesized and evaluated for their ability to activate transcription. A greater than 30-fold enhancement over basal levels was observed and activation could be correlated to DNA occupancy levels. Cys_2His_2 zinc finger proteins are the most common DNA binding motif in higher eukaryotes. We have elucidated an allosteric mechanism for the inhibition of zinc finger proteins, binding purely in the major groove, by Py/Im polyamides. The inhibition of this large class of proteins greatly enlarges the applicability of these minor groove ligands for gene regulation.

Table of Contents

List of Figures and Tables

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

To my husband, Peter

Chapter One

Introduction

Background

DNA is the universal storage material for genetic information in all living organisms. The flow of genetic information starts with DNA, which is transcribed into RNA, and the RNA is translated into proteins. DNA serves as the permanent storage material in the cell, while RNA is a transient copy of this genetic information and is synthesized as necessary. Proteins serve to perform specific functions such as signaling or maintenance in the cell, and like RNA, are often synthesized as needed and then degraded. It is remarkable, then, that all genetic information essential to sustaining living organisms is present in the form of DNA and that copies of this DNA are present in every living cell. The recent sequencing of the human genome has revealed that our genetic information is comprised of 3.2 gigabases of DNA organized into 30,000-40,000 protein-coding genes.[1,2]

Specific protein-DNA interactions are fundamental pieces to cell differentiation. For although almost all cells contain identical genetic information, the selective transcription of certain genes but not others allows a neural cell in the brain to function differently than an epithelial cell in the skin.[3] It is now understood that nature controls the expression of each of these genes by a remarkable three-dimensional switch, which is composed of over 50 proteins assembled on a few hundred specific base pairs of DNA, known as the promoter region, encoded upstream from the RNA polymerase start site and coding region. Transcription factors are DNA binding proteins that regulate transcription by recruiting the necessary transcriptional machinery. Typically, the promoter region of a gene contains the recognition sites of the required transcription factors. Upon binding, the transcription factor recruits the network of proteins necessary to initiate active

transcription. This multi-protein complex has a molecular weight of more than two million, and contains all the proteins required for activated transcription to occur.[4] The process of transcription elongation is just as complex, as more than 20 protein subunits that comprise RNA polymerase II must track along the DNA and synthesize the complementary RNA. Unlike in DNA synthesis, transcription elongation occurs in spurts and starts with elongation factors necessary to start the polymerase complex again once it has paused.[5,6]

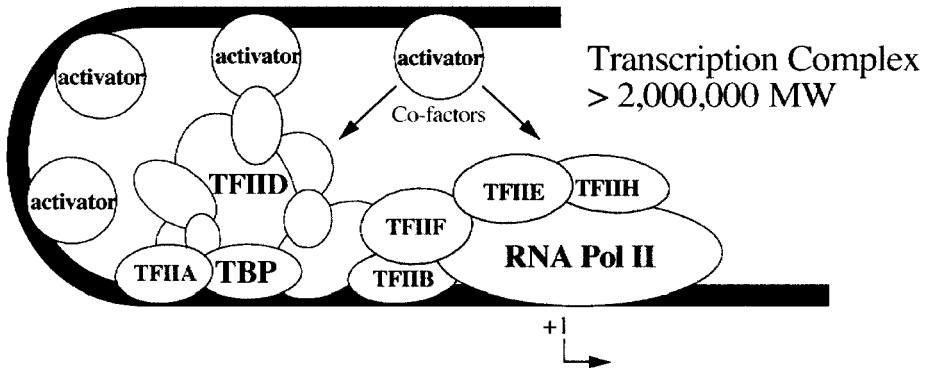


Figure 1.1. Model of protein regulation of gene transcription.

Since each cell contains a copy of the individual's DNA that is essentially identical to that of every other cell, yet each cell has a specialized function, it is required that certain genes be expressed while others remain silent.[3] Misregulation of gene expression resulting in aberrant biochemical behavior in cellular processes is responsible for numerous cancerous and viral disease states.[4] Therefore, artificial ligands that can sequence specifically bind DNA and restore the natural patterns of gene expression would potentially be valuable tools in human medicine.[7]

DNA Recognition. To develop general rules for sequence specific recognition by small molecule ligands, a careful examination of DNA structure is required. Double helical DNA consists of the two strands of deoxyribose-phosphate polymers that display four heterocycles bases, adenine (A), thymidine (T), guanine (G), and cytosine (C), bound in anti-parallel orientation. In duplex DNA, the two strands are held together via Watson-Crick hydrogen bonding of A, T and G,C base pairs.[8] Double helical DNA can be characterized as having three domains: the sugar/phosphate backbone, the major groove, and the minor groove. The common B-form of DNA is characterized by a wide (12 Å) and shallow major groove and a narrow (4-6 Å) and deep minor groove (Figure 1.2).[9] In addition, sequence-dependent structural variations, conformational properties, and solvent and counterion organization can distinguish local DNA structures.[9] Individual sequences may be distinguished by the pattern of hydrogen bond donors and acceptors, created by the base pairs.

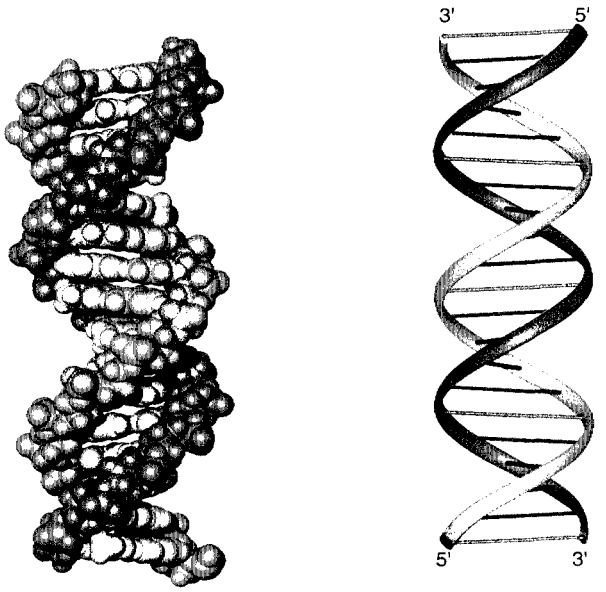


Figure 1.2. CPK model (left) and ribbon representation (right) of B-form duplex DNA. Separate strands are either darkly or lightly shaded.

The nature of the DNA helix dictates the characteristic structures of the major and minor grooves. The edges of the base pairs in the Watson-Crick hydrogen bond pairings result in the unique hydrogen bonding faces of the major and minor grooves. For a T•A base pair, in the major groove there is a pattern of hydrogen bond acceptor (lone pairs on the O4of thymine), and hydrogen bond donor (NH of exocyclic amine of adenine). The minor groove of a T•A base pair forms a pattern of two hydrogen bond acceptors (O2 of thymine and N3 of adenine). Likewise, for a C•G base pair, the major groove can be characterized as having a hydrogen bond donor (NH of exocyclic amine of cytosine) and a hydrogen bond acceptor (O6 of guanine). The minor groove of a C•G base pair has a hydrogen bond acceptor (O2 of cytosine), hydrogen bond donor (NH of exocyclic amine of guanine) and a hydrogen bond acceptor (N3 of guanine) (Figure 1.3). In the minor groove, it is the exocyclic amine of guanine that primarily distinguishes an A•T base pair from a G•C base pair.

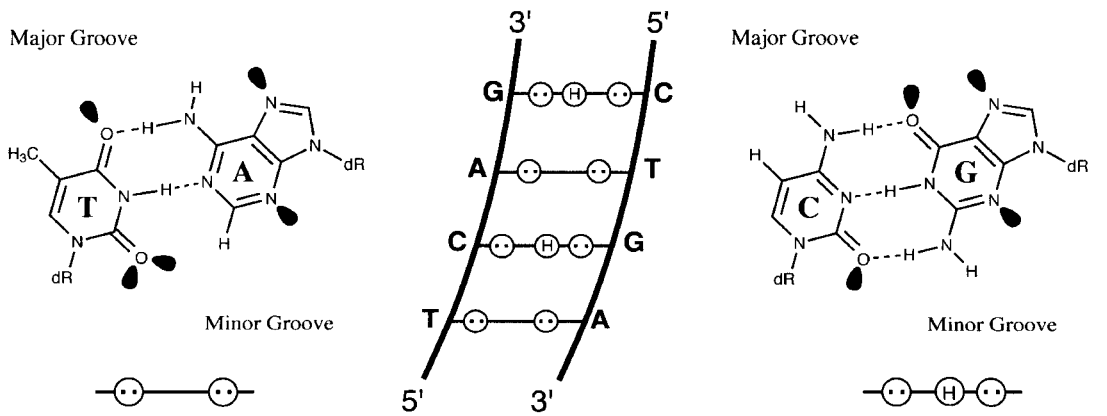


Figure 1.3. A schematic model for recognition of the minor groove, with hydrogen bond donors represented as (H) and hydrogen bond acceptors represented as two dots.

X-ray crystal and NMR structural analyses, of naturally occurring protein-DNA complexes and natural products, have provided insight into the design of novel DNA binding molecules.[10,11] Sequence-specificity arises from several methods of DNA interaction: specific hydrogen bonding or van der Waals contacts with functional groups in the grooves, Coulombic attraction to the negatively charged phosphodiester backbone or to the electrostatic potential in the grooves, and/or intercalation of aromatic functional groups between the DNA bases.[12] In cells, nature has devised a combinatorial approach for DNA recognition, utilizing a 20 amino acid code to form complex tertiary

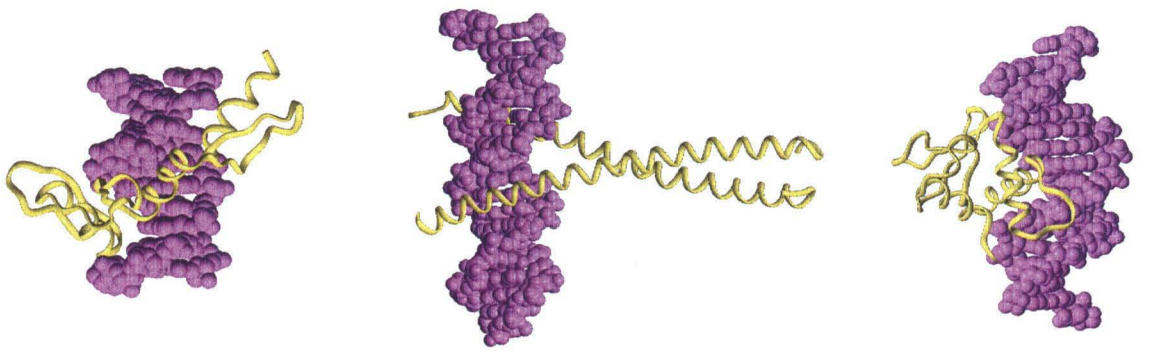
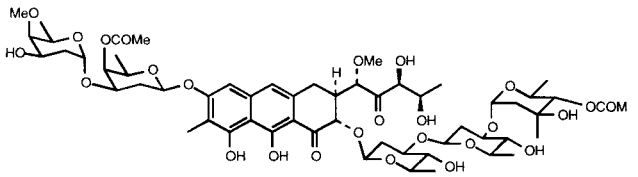


Figure 1.4. Representative X-ray crystal structures from a few families of DNA binding proteins. Proteins contact the DNA via interactions with the bases in the major groove and/or minor groove, as well as the sugar phosphate backbone.

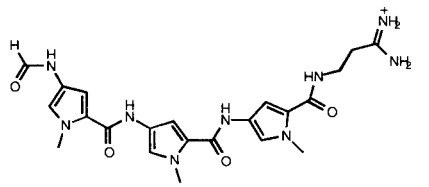
folded structures. DNA binding proteins adopt several structural motifs for sequence-specific recognition including the zinc finger,[13] the leucine zipper,[14] and the helix-turn-helix motifs[15] (Figure 1.4). Within these complexes, specificity for target sites is achieved through specific noncovalent interactions between the protein side chains and the nucleobases and phosphates of the DNA. However, no single motif exists that represents a general amino acid-base pair code for all DNA sequences.[16] Although

certain zinc finger-DNA complexes[17,18] provide a versatile recognition code, the *de novo* design of proteins for DNA recognition remains challenging due to protein structural diversity and limitations in predicting protein folding.

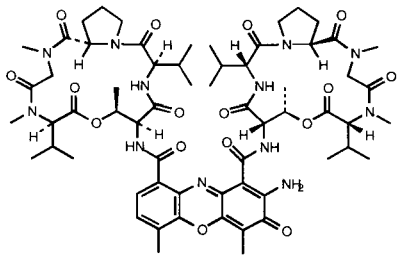
Small molecules isolated from natural sources are another class of DNA binding ligands that are structurally diverse, as evidenced by consideration of four representative molecules, chromomycin, distamycin, actinomycin D, and calicheamicin (Figure 1.5).[19-22]



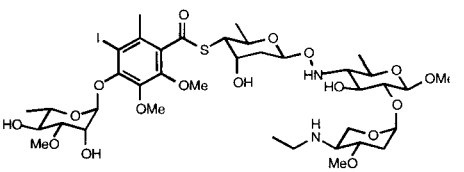
Chromomycin • 5'-GGCT-3'



Distamycin • 5'-$(A,T)_5$-3'



Actinomycin D • 5'-GC-3'



Calicheamicin oligosaccharide • 5'-TCCT-3'

Figure 1.5. The structures of four small molecules isolated from natural sources.

While these molecules can recognize short sequences of DNA through both intercalation between the bases and groove binding, there is no general recognition code for the readout of specific sequences of DNA. Analogs of calicheamicin have been shown to bind DNA and interfere with transcription factor function; however, oligosaccharides are not yet amenable to the recognition of a broad range of predetermined DNA sequences.[23,24] Oligodeoxynucleotides which recognize the major groove of double helical DNA via triple helix formation bind a broad sequence repertoire with high affinity

and specificity.[25,26] Although oligonucleotides and their analogs have been shown to interfere with gene expression, the triple helix approach is limited to purine tracks and suffers from poor cellular uptake.[27,28]

DNA Recognition by Polyamides. By examining the small molecule natural products that bind DNA, it is clear that distamycin is distinct from most DNA binding natural products by its simplicity. It consists of three *N*-methyl carboxamide units and contains no chiral centers. This crescent shaped molecule binds in the minor groove of DNA, preferentially at A-T rich regions. The crystal structure of distamycin bound to a DNA duplex showed that the carboxamides make specific hydrogen bonds to the bases in the minor groove.[29] The Dervan group has pursued distamycin as the starting point to formulate a general recognition code for small molecules which bind in the minor groove of DNA.[30]

An early lead in attempts to generate a general code for DNA recognition was the discovery of 2-imidazole netropsin. It had been suggested that the imidazole ring of this compound would make a specific hydrogen bond to a G•C base pair and that 2-imidazole netropsin would target the sequence 5'-GWWW-3'. Footprinting experiments showed that, unexpectedly, 2-imidazole netropsin instead bound the site 5'-WGWCW-3'. The two-fold symmetry of the binding site was puzzling since it suggested that there was not a single binding orientation for 2-imidazole netropsin. Earlier studies of distamycin/DNA complexes gave an interesting solution to this discrepancy. Distamycin was found to have two very distinct binding modes. In the first, a single molecule of distamycin binds in the center of the minor groove of A-T rich DNA. The amide hydrogens form bifurcated hydrogen bonds to the N3 of adenine and O2 of thymine on the floor of the minor groove. In the second binding mode, distamycin binds as an

antiparallel, side-by-side dimer in the minor groove. Instead of the bifurcated hydrogen bonds of the 1:1 complex, each strand of DNA forms hydrogen bonds with one molecule of distamycin. Using this 2:1 binding model for 2-imidazole netropsin explained why its preferred binding site is symmetrical (because it binds as a side-by-side antiparallel dimer) and does not match the predicted site.[31,32]

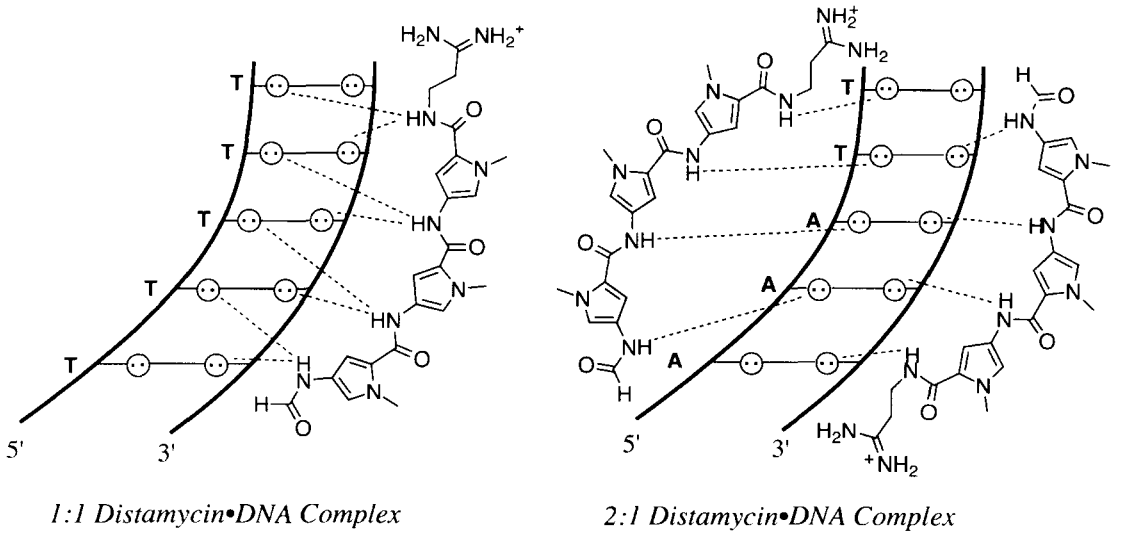
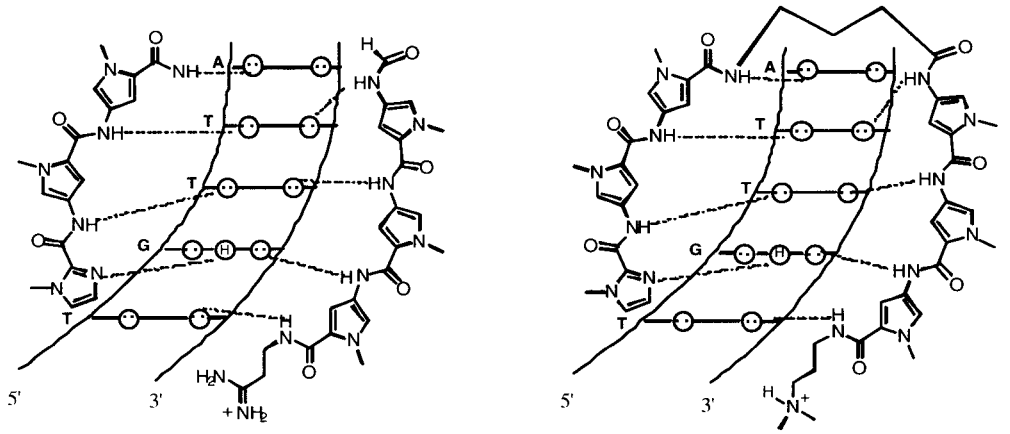


Figure 1.6. A schematic representation of recognition of A,T rich sequences in the minor groove by 1:1 and 2:1 complexes of Distamycin.

The affinity of the dimer 2-imidazole netropsin was still modest (micromolar binding constants) in comparison to DNA binding proteins (nanomolar or higher binding constants). To decrease the entropic factors influencing binding, two molecules of 2-imidazole netropsin were connected in a head to tail fashion with an alkyl linker, γ-amino-butyric acid (γ).[33] The resulting 'hairpin'-like structure yielded a compound with increased affinity for the binding site 5'-WGWCW-3' of ~100-fold. Connecting the two distamycin-like components also allowed for the design of heterodimeric polyamides to target asymmetric sequences. As an example, the compound ImPyPy-γ-PyPyPy was

found to target the sequence 5'-WGWWW-3'. The hairpin motif for this compound was found to increase the binding affinity by 400-fold (Figure 1.7).



ImPyPy-Dp / PyPyPy • TGTTA
$K_a \sim 2 \times 10^5\ M^{-1}$

ImPyPy-γ-PyPyPy-Dp • TGTTA
$K_a = 8 \times 10^7\ M^{-1}$

Figure 1.7. Example of the hairpin polyamide motif.

The results of 2-imidazole netropsin binding to the site 5'WGWCW-3' as an antiparallel dimer and the development of the hairpin motif allowed us to design a set of pairing rules for recognition by pyrrole (Py) and imidazole (Im) amino acids[34] (Figure 1.8). A Py/Im pair targets a C•G base pair while an Im/Py pair targets a G•C base pair.[35,36] The basis for discrimination of a G•C base pair is from the formation of a hydrogen bond between imidazole N3 and the exocyclic amine of guanine.[37,38] A Py/Py pair is partially degenerate and recognizes both A•T and T•A base pairs.[31,35,36,39,40] High resolution NMR and crystal structures have confirmed that specific hydrogen bonds are made from the carboxamide nitrogens to the base pairs and the imidazole nitrogen to guanine.

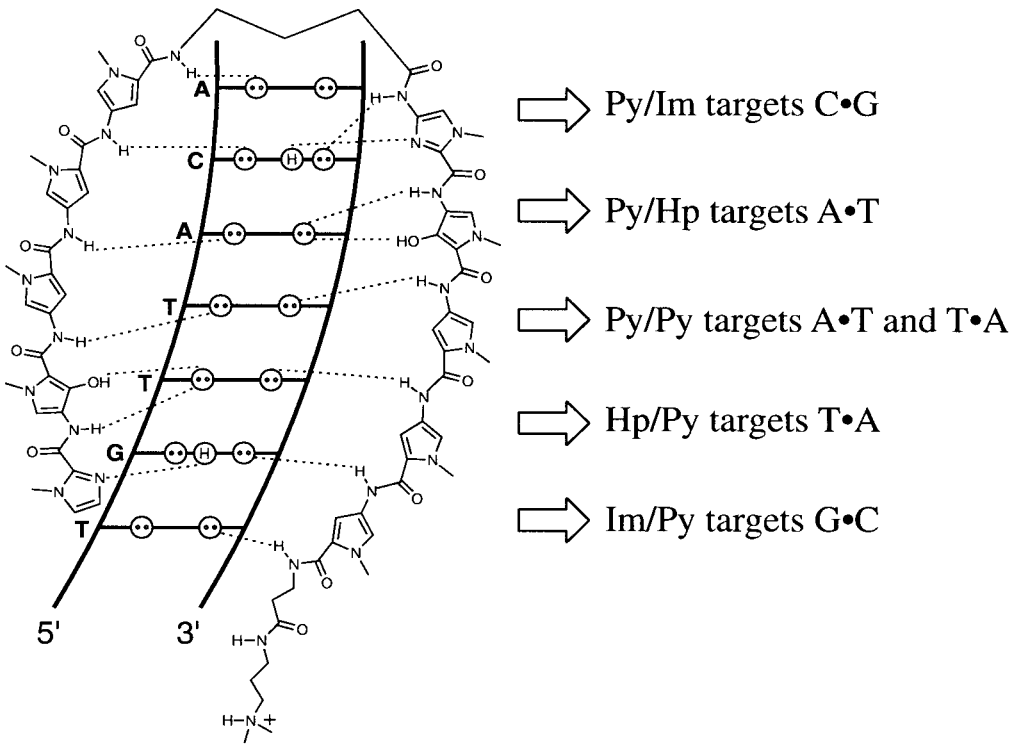


Figure 1.8. A schematic representation of the polyamide pairing rules.

To address the question of A•T/T•A degeneracy by a Py/Py pair, *N*-methyl-3-hydroxypyrrole (Hp) was synthesized and incorporated into a polyamide.[41,42] Hp places a hydroxy group in the asymmetric cleft between a thymine and adenine base pair. Thus, an Hp/Py pair recognizes the sequence T•A and a Py/Hp pair recognizes an A•T base pair (Figure 1.9). The binding of Hp has also been confirmed by a high resolution crystal structure, and shows that the hydroxy group of Hp binds in the asymmetric cleft as designed and also makes two specific hydrogen bonds to O2 of thymine.[43,44]

The development of a solid phase methodology to synthesize Py-Im polyamides has greatly facilitated the examination of other motifs and analogues of polyamides.[45] Another modification to the hairpin structures is the use of a chiral turn (diamino-butyric acid, DABA) which influences the binding orientation of the polyamide and can increase

affinity and specificity.[46] Increasing the hairpin polyamide subunit lengths to four and five aromatic ring residues gives eight- and ten-ring hairpin polyamides with subnanomolar binding affinities, similar to naturally occurring DNA-binding proteins.[47,48] Studies of polyamide site size limitations suggest that beyond five consecutive rings, the ligand curvature fails to match the pitch of the DNA helix, disrupting the hydrogen bonds and van der Waals interactions responsible for specific polyamide-DNA complex formation.[38] Addition of pairings with β-alanine (β) to form β/β, β/Py, and β/Im pairs has extended the targetable binding site size recognizable by the hairpin motif at high affinity.[49,50]

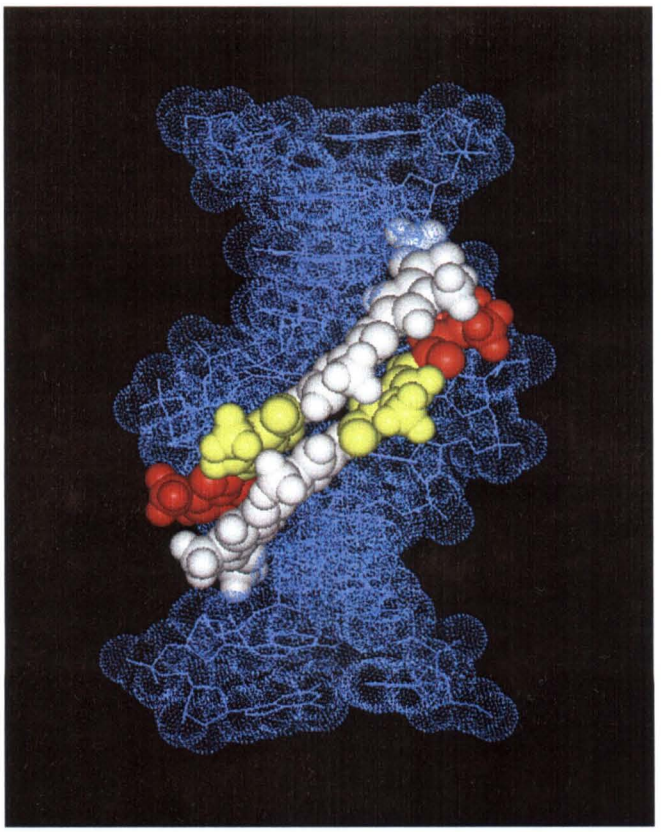


Figure 1.9. Space filling representation of the polyamide dimer ImHpPyPy-β-Dp bound in the minor groove of DNA. The figure was prepared using Insight II software and is derived from a high-resolution X-ray co-crystal structure of a polyamide dimer bound to DNA, which was obtained in collaboration with the Rees group at the California Institute of Technology.

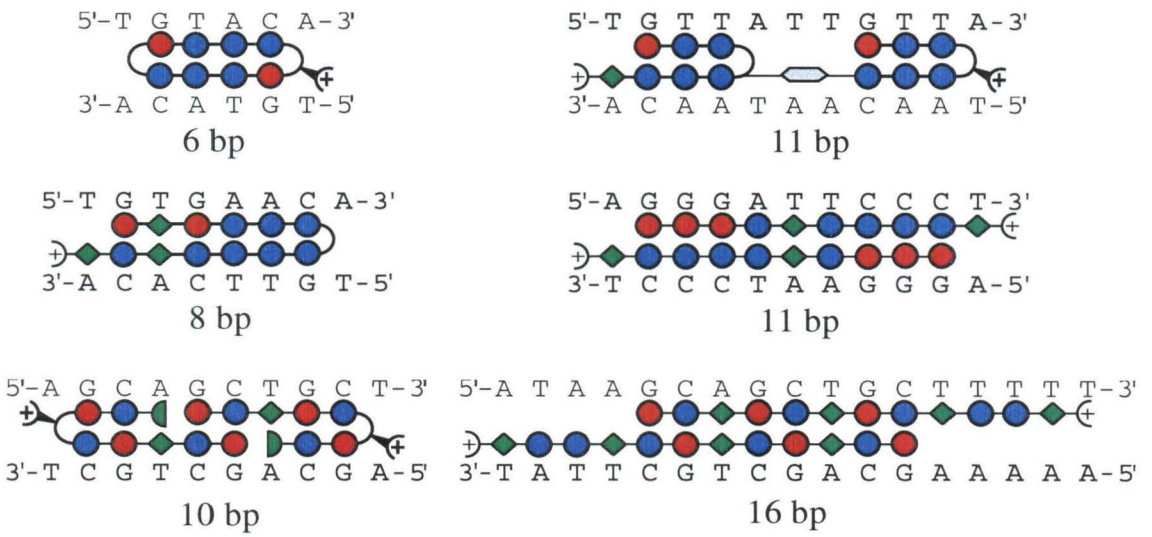


Figure 1.10. Representative motifs for polyamide:DNA recognition. Red and blue circles represent imidazole and pyrrole residues, respectively. β-Alanine and the γ-turn are depicted as green diamonds and curved lines, respectively. The plus sign represents the dimethylaminopropylamide tail.

A variety of motifs for sequence specific recognition using the polyamide scaffold have been successfully developed for target sites from 5-16 bp (Figure 1.10).[47,48,51-57] For longer sequences, the introduction of the flexible β-alanine unit allows the polyamide, that would otherwise be overwrought relative to the DNA helix, to track the minor groove. Cooperative 2:1 dimer motifs are particularly attractive for future applications due to the low molecular weight of the monomer units.[53] Binding of sites up to 16 bp, potentially sufficient for recognition of a unique sequence in the human genome, have been successfully targeted.[54] Closing the ends of a hairpin polyamide with a second γ affords the cycle motif which can increase polyamide affinity.[57,58] The incorporation of a stereospecific α-amino group into the γ turn also allows for functional moieties to be conjugated to the polyamide or the construction of a tandem motif by linking the C-terminus of another polyamide to this position.[46,56]

Gene Regulation by Polyamides. As an initial step toward asking whether cell-permeable small molecule transcription agonists might regulate gene expression in complex organism, regulation of transcription in specific genes by polyamides has been studied. The DNA-binding activity of the 5S RNA specific transcription factor TFIIIA, a nine-zinc-finger protein, was inhibited by an eight-ring hairpin polyamide that bound within the recognition site of the minor groove spanning fourth finger. Transcription of 5S RNA genes by RNA polymerase III was suppressed *in vitro* and in cultured *Xenopus* kidney cells.[7] Polyamides have also been shown to downregulate genes transcribed by RNA polymerase II. Two hairpin polyamides were targeted to DNA sequences adjacent to DNA sequences immediately adjacent to the binding site of transcription factors, Ets-1, LEF-1 and TBP in the HIV-1 enhancer/promoter (Figure 1.11). These synthetic ligands

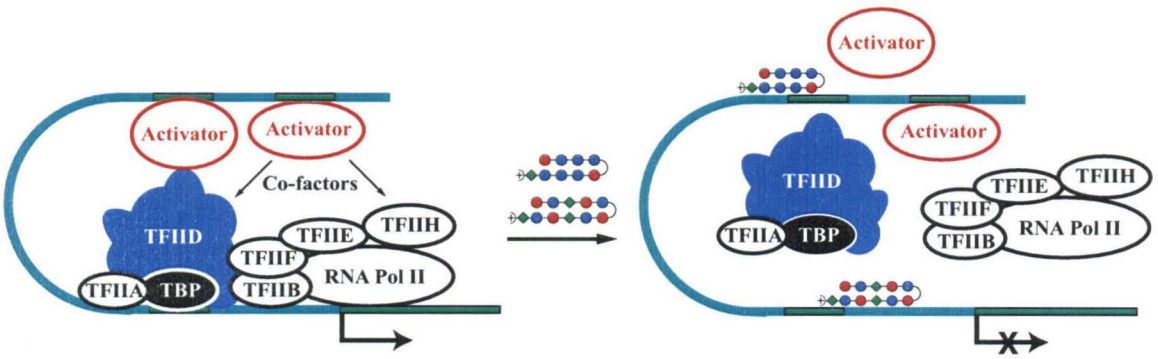


Figure 1.11. Inhibition of gene transcription by hairpin polyamides. A pair of hairpin polyamides targeted to the DNA sequences adjacent to the binding sites for LEF-1 and TBP inhibit assembly of the transcriptional machinery and transcription of the HIV-1 gene in human cell culture.

inhibit DNA-binding of each transcription factor and HIV-1 transcription in cell-free assays. Treatment of isolated human peripheral blood lymphocytes with the two hairpin polyamides demonstrated inhibition of viral replication.[49] In an effort to create an artificial activator, capable of upregulating gene expression, a hairpin polyamide was

tethered to a short peptide (20mer) activation domain (AH). This conjugate successfully upregulated transcription in a cell-free system.[59]

Scope of this Work

Polyamides are a class of small molecules capable of recognizing and binding sequence-specifically to DNA with high affinity. Their potential for application in molecular medicine and as a biological tool is tremendous. It is therefore imperative that investigations continue to probe the more subtle points of polyamide/DNA and polyamide/protein interactions to further increase the general applicability of this ligand system. This thesis describes work to further investigate how atomic changes to polyamides can impact their DNA recognition properties. Additionally, the ability of polyamides to impact protein/DNA interactions was also investigated.

Chapter Two describes the evaluation of two heterocyclic amino acids, thiazole and pyrazole, as new monomers within the polyamide framework. While the incorporation of a single thiazole monomer decreased binding affinity by several orders of magnitude, the pyrazole monomer conferred surprisingly high specificity for AT/TA base pairs over GC/CG pairs without loss in affinity. These results begin to highlight that positions on the aromatic ring, other than the corner 3-postion, may impact DNA recognition. Increasing the number of pairings that can be used in polyamides may impact the physical properties of polyamides such as cell permeability and metabolism which could greatly affect their application in biological systems. Chapter Three decribes several new synthetic methods to access polyamides with varying C-terminal tail motifs. Hydroxamic ester PAM resin, Kaiser oxime resin and Rink resin with Fmoc-

monomers were used to synthesize these new polyamides in high yield and purity. These truncated polyamides were found to bind DNA with comparable affinity to the standard β-Dp tail polyamides but were more general in their ability to target DNA as they did not require the two A/T base pairs at the C-terminus, normally necessary to accommodate the larger β-Dp tail. This increased generality coupled with a significant decrease in molecular weight should enhance polyamide applications in biological settings.

Chapter Four investigates in more detail the design of artificial transcriptional activators that use polyamides as the DNA recognition domain. Minimization of both the activation domain and the linker region between domains was achieved without significant diminution of the activation potential. The establishment of new DNase I footprinting protocols allowed a strong correlation to be made between promoter occupancy and in vitro levels of transcriptional activation. Chapter Five reports for the first time the allosteric inhibition of a purely major groove binding zinc finger protein by unmodified polyamides binding solely in the minor groove. A concern has often been expressed that while most DNA-binding proteins bind in the major groove of DNA, polyamides bind in the minor groove which may negatively impact their ability to inhibit transcription factor proteins. Clearly the results obtained from this study demonstrate that minor groove binding polyamides can inhibit proteins by a number of modes effectively.

References

1. Consortium, I. H. G. S. *Nature* **2001**, *409*, 860.
2. Venter, J. C. e. a. *Science* **2001**, *291*, 1304.
3. Roeder, R. G. *Trends in Biochemical Sciences* **1996**, *9*, 327-335.
4. Tjian, R. *Scientific American* **1995**, *2*, 54-61.
5. Cramer, P. e. a. *Science* **2000**, *288*, 640-649.
6. Conaway, J. W.; Conaway, R. C. *Science* **2000**, *288*, 632-633.
7. Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202-205.
8. Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, M.; Fratini, A. V.; Kopka, M. L. *Science* **1982**, *216*, 475.
9. Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
10. Zimmer, C.; Wahnaert, U. **1986**, *47*, 31-112.
11. Steitz, T. A. *Quarterly Review of Biophysics* **1990**, *23*, 205.
12. Pabo, C. O.; Sauer, R. T. *Annual Review of Biochemistry* **1992**, *61*, 1053-1095.
13. Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809.
14. Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223.
15. Feng, J. A.; Johnson, R. C.; Dickerson, R. E. *Science* **1994**, 348.
16. Kissinger, C. R.; Liu, B.; Martin-Blancco, E.; Kornberg, T. B.; Pabo, C. O. *Cell* **1990**, *63*, 579.
17. Greisman, H. A.; Pabo, C. O. *Science* **1997**, *275*, 657-661.
18. Segal, D. J.; Barbas, C. F. *Current Opinion in Chemical Biology* **2000**, *4*, 34-39.

19. Coll, M.; Frederick, C. A.; Wang, A. H. J.; Rich, A. *Proceedings of the National Academy of Sciences of the USA* **1987**, *84*, 8385-8389.

20. Gao, X. L.; Mirau, P.; Patel, D. J. *Journal of Molecular Biology* **1992**, *223*, 259-279.

21. Kamitori, S.; Takusagawa, F. *Journal of Molecular Biology* **1992**, *225*, 445-456.

22. Paloma, L. G.; Smith, J. A.; Chazin, W. J.; Nicolaou, K. C. *Journal of the American Chemical Society* **1994**, *116*, 3697-3708.

23. Ho, S. N.; Boyer, S. H.; Schreiber, S. L.; Danishefsky, S. J.; Crabtree, G. R. *Proceedings of the National Academy of Sciences of the USA* **1994**, *91*, 9203-9207.

24. Liu, C.; Smith, B. M.; Ajito, K.; Komatsu, H.; GomezPaloma, L.; Li, T. H.; Theodorakis, E. A.; Nicolaou, K. C.; Vogt, P. K. *Proceedings of the National Academy of Sciences of the USA* **1996**, *93*, 940-944.

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

26. Thuong, N. T.; Helene, C. *Angewandte Chemie-International Edition in English* **1993**, *32*, 666-690.

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

28. Nielsen, P. E. *Chemistry-a European Journal* **1997**, *3*, 505-508.

29. Coll, M.; Frederick, C. A.; Wang, A. H. J.; Rich, A. A. *Proceedings of the National Academy of Sciences of the USA* **1987**, *84*, 8385-8389.

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

31. Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Nature Structural Biology* **1994**, *1*, 169-175.

32. Pelton, J. G.; Wemmer, D. E. *Biochemistry* **1988**, *27*, 8088-8096.

33. Mrksich, M.; Dervan, P. B. *Journal of the American Chemical Society* **1994**, *116*, 3663-3664.

34. Dervan, P. B.; Bürli, R. W. *Current Opinion in Chemical Biology* **1999**, *3*, 688-693.

35. Wade, W. S.; Mrksich, M.; Dervan, P. B. *Journal of the American Chemical Society* **1992**, *114*, 8783-8794.

36. White, S.; Baird, E. E.; Dervan, P. B. *Chemistry & Biology* **1997**, *4*, 569-578.

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

38. Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C. *Nature Structural Biology* **1998**, *5*, 104-109.

39. Pelton, J. G.; Wemmer, D. E. *Proceedings of the National Academy of Sciences of the USA* **1989**, *86*, 5723-5727.

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

41. White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468-471.

42. White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1999**, *121*, 260-261.

43. Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Journal of Molecular Biology* **2000**, *295*, 557-567.

44. Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111-115.

45. Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6141-6146.

46. Herman, D. M.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1998**, *120*, 1382-1391.

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

48. Turner, J. M.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1997**, *119*, 7636-7644.

49. Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proceedings of the National Academy of Sciences of the USA* **1998**, *95*, 12890-12895.

50. Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1998**, *120*, 6219-6226.

51. Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6160-6166.

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

53. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Angewandte Chemie-International Edition* **1998**, *37*, 1421-1423.

54. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1998**, *120*, 3534-3535.

55. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chemistry-a European Journal* **1997**, *3*, 1600-1607.

56. Herman, D. M.; Baird, E. E.; Dervan, P. B. *Chemistry-a European Journal* **1999**, *5*, 975-983.

57. Cho, J.; Parks, M. E.; Dervan, P. B. *Proceedings of the National Academy of Sciences, USA* **1995**, *92*, 10389-10392.

58. Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1999**, *121*, 1121-1129.

59. Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B. *Proceedings of the National Academy of Sciences of the USA* **2000**, *97*, 3930.

Chapter Two

Alternative Heterocycles for DNA Recognition: An *N*-methylpyrazole/*N*-methylpyrrole Pair Specifies for A•T/T•A Base Pairs

The text of this chapter was taken in part from a publication co-authored with Jason W. Szewczyk, Eldon E. Baird and Professor Peter B. Dervan.

Publication: Nguyen, D.H.; Szewczyk, J.W.; Baird, E.E.; Dervan, P.B. *Bioorganic & Medicinal Chemistry*, **2001**, *9*, 7-17.

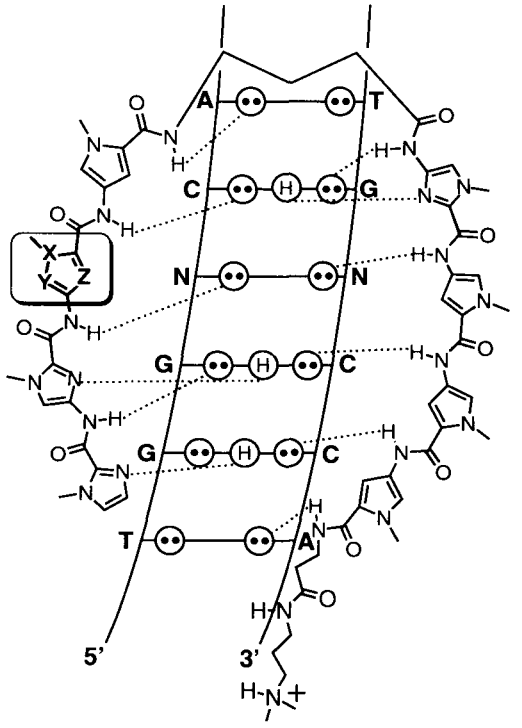
Abstract

Side-by-side pairs of three five membered rings, *N*-methylpyrrole (Py), *N*-methylimidazole (Im), and *N*-methylhydroxypyrrole (Hp) have been demonstrated to distinguish each of the four Watson-Crick base pairs in the minor groove of DNA. However, not all DNA sequences targeted by these pairing rules achieve affinities and specificities comparable to DNA binding proteins. We have initiated a search for new heterocycles which can expand the sequence repetoire currently available. Two heterocyclic aromatic amino acid, *N*-methylpyrazole (Pz) and 4-methylthiazole (Th), were incorporated into a single position of an eight ring polyamide of sequence ImIm**X**Py-γ-Im**Py**PyPy-β-Dp to examine the modulation of affinity and specificity for DNA binding by a Pz/Py pair and/or a Th/Py pair. The **X/Py** pairings Pz/Py and Th/Py were evaluated by quantitative DNase I footprint titrations on a DNA fragment with the four sites 5'-TGGNCA-3' (N=T, A, G, C). The Pz/Py pair binds T•A and A•T with similar affinity to a Py/Py pair but with improved specificity, disfavoring both G•C and C•G by about 100-fold. The Th/Py pair binds poorly to all four Watson Crick base pairs. These results demonstrate that in some instances new heterocyclic aromatic amino acid pairs can be incorporated into imidazole-pyrrole polyamides to mimic the DNA specificity of Py/Py pairs which may be relevant as biological criteria in animal studies become important.

Introduction

Small molecules that target specific predetermined DNA sequences have the potential to control gene expression.[1-3] Polyamides containing the three aromatic amino acids 3-hydroxypyrrole (Hp), imidazole (Im) and pyrrole (Py) are synthetic ligands that have an affinity and specificity for DNA comparable to naturally occurring DNA binding proteins.[4-6] Recently, eight-ring hairpin polyamides have been found to regulate transcription and permeate a variety of cell types in culture.[1,2] DNA recognition depends on the side-by-side amino acid pairings oriented N→C with respect to the 5'→3' direction of the DNA helix in the minor groove.[4-6] Antiparallel pairing of Im opposite Py (Im/Py pair) distinguishes G•C from C•G and both of these from A•T/T•A base pairs.[5,6] A Py/Py pair binds both A•T an T•A in preference to G•C/C•G.[5,6] The discrimination of T•A from A•T using Hp/Py pairs completes the four base pair (bp) code.[7-9]

Given the sequence dependent microstructure of the DNA helix,[10-16] it is significant that a simple recognition code can be developed.[17-20] In both published and unpublished work, over a hundred pyrrole-imidazole polyamides have been synthesized which recognize predetermined sequences. However, within this group, a number of DNA sequences have emerged as difficult targets for pyrrole-imidazole polyamides to bind with subnanomolar affinity. Sequence dependent DNA structural features such as intrinsic minor groove width, minor groove flexibility, and inherent DNA curvature may reduce polyamide binding at certain sites.[10-16] Since molecular composition could enhance affinity at difficult sequences, the discovery of new monomer subunits that provide polyamides with affinities and specificities comparable to naturally occurring DNA binding proteins remains a high priority.



	Z	Y	X
Py	C-H	C-H	N
Hp	C-OH	C-H	N
Im	N	C-H	N
Pz	C-H	N	N

Figure 2.1. Binding model for the 1:1 complexes formed between the DNA and ImIm**X**Py-γ-ImPyPyPy-β-Dp. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. N is represented as A or T but can be A, T, G, or C. Putative hydrogen bonds are illustrated by dotted lines. The atomic variations (at positions X, Y, and Z) between the pyrrole, hydroxypyrrole, imidazole, pyrazole and thiazole monomers are also shown.

The substitution of pyrrole C3-H with N or C3-OH to generate imidazole or hydroxypyrrole, respectively, clearly demonstrates the impact of atomic substitutions, in polyamide monomer units, on DNA recognition.[5-9] These substitutions also highlight the importance of the 3-position as a recognition element in the floor of the minor groove. The aromatic amino acid monomers of polyamides are also amenable to modification at the 1-and 5-postitions of the pyrrole/imidazole rings (Figure 2.1). Substitutions at the 1-position have been investigated,[5] but until now, substitutions at the pyrrole 5-position have not been evaluated. Replacement of the C5-H of *N*-methylpyrrole with N generates a new heterocyclic amino acid subunit, *N*-methylpyrazole (Pz). Pyrazole was thought to

be of interest in light of altered ring electronics, possible interactions between the newly introduced nitrogen and the polyamide backbone amide oxygen, differing cell permeation properties or degradation pathways in living animals. Additionally, a previously reported heterocyclic aromatic amino acid, 4-methylthiazole (Th) which substitutes an S for C3-H, N for C5-H and C for N1 relative to pyrrole, was added to this study for comparison (Figures 2.1 & 2.2).

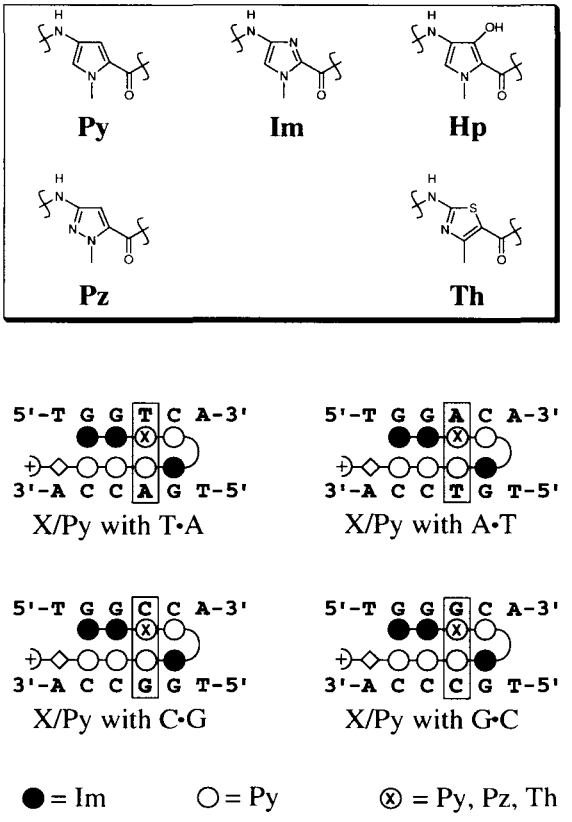


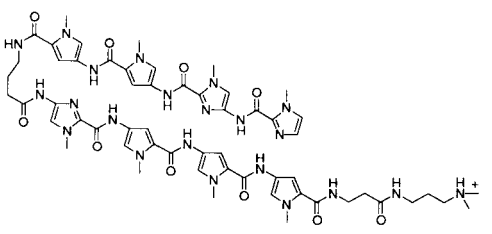
Figure 2.2. Monomer units incorporated into hairpin polyamides. Ball and stick polyamide: DNA binding models are also shown. Shaded circles denote Im while nonshaded circles represent Py. Nonshaded circles containing an X represent Py, Pz or Th monomer units. Nonshaded diamonds denote the β-residue. γ-aminobutyric acid (γ) and dimethylaminopropylamide (Dp) are depicted as a curved line and a plus sign, respectively.

When incorporated into three ring distamycin analogs, the 4-methylthiazole (Th) monomer,[21] which places a sulfur atom into the floor of the minor groove, conferred A/T

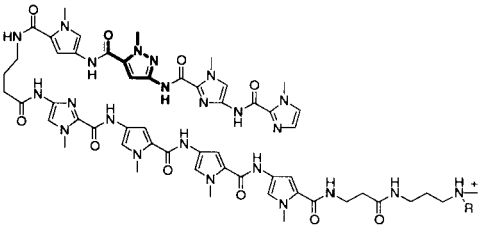
preference over G/C.[21,22] Thiazole (Th) has been proposed in the literature as an A-specific recognition element since the bulkiness of the sulfur atom was expected to sterically clash opposite the O2 of thymine and therefore be better accommodated opposite an adenine base.[23,24] However, it has been shown that the asymmetric cleft in the minor groove of A/T pairs preferentially places the bulky C3-OH group of Hp opposite T rather than A where there is a steric destabilization with the C2-H of adenine.[7,8] Since an additional hydrogen bond is also made between Hp and T, it remains quite interesting to investigate the discrimination of A•T/T•A base pairs by a purely steric means using the Th/Py pairing within the context of our hairpin motif.

To investigate these new pairings, three polyamides, ImIm**Py**Py-γ-Im**Py**PyPy-β-Dp (**1**), ImIm**Pz**Py-γ-Im**Py**PyPy-β-Dp (**2**), and ImIm**Th**Py-γ-Im**Py**PyPy-β-Dp (**3**), were synthesized by solid phase methods[25] to contain either a side-by-side pairing of two pyrrole rings (Py/Py), a pyrazole and pyrrole pairing (Pz/Py), or a thiazole and pyrrole pairing (Th/Py), respectively. The corresponding EDTA analogue ImImPzPy-γ-ImPyPyPy-β-Dp-EDTA (**2E**) was also constructed to confirm the orientation and stoichiometry of this hairpin at each binding site (Figure 2.3). We report here the DNA binding affinity and sequence selectivity of the previously reported[7,26] control polyamide ImIm**Py**Py-γ-Im**Py**PyPy-β-Dp (**1**) in relation to the new polyamides ImIm**Pz**Py-γ-Im**Py**PyPy-β-Dp (**2**), and ImIm**Th**Py-γ-Im**Py**PyPy-β-Dp (**3**) for the 6 bp sites, 5'-TGGTCA-3', 5'-TGGACA-3', 5'-TGGGCA-3', and 5'-TGGCCA-3'; thereby evaluating the three pairings Py/Py, Pz/Py and Th/Py, respectively, against all four Watson-Crick base pairs. Three separate techniques are used to characterize the DNA binding properties of the newly designed polyamides: DNase I footprinting, methidiumpropyl-

EDTA•Fe(II) (MPE•Fe(II)) footprinting, and affinity cleavage. Quantitative DNase I footprinting titrations allow determination of equilibrium association constants (K_a) of the polyamides for respective match and mismatch sites.[27-29] Information about precise binding site size is gained from MPE•Fe(II) footprinting,[30,31] while affinity cleavage studies confirm the binding orientation and stoichiometry of the hairpin:DNA complex.[17,32]

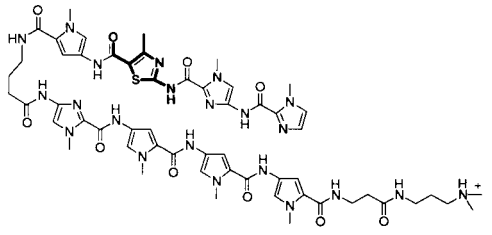


1 ImImPyPy-γ-ImPyPyPy-β-Dp



2 R=CH_3 ImImPzPy-γ-ImPyPyPy-β-Dp

2E R= R_E ImImPzPy-γ-ImPyPyPy-β-Dp-EDTA·Fe(II)



3 ImImThPy-γ-ImPyPyPy-β-Dp

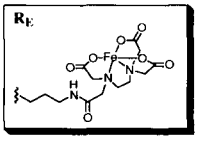


Figure 2.3. Structures of the eight-ring hairpin polyamides **1**, **2**, **3** and the corresponding Fe(II)•EDTA affinity cleavage derivative **2E** as synthesized by solid-phase methods.

Results

Monomer Synthesis. Ethyl-3-amino-1-methylpyrazole-5-carboxylate (**4**) and ethyl-2-amino-4-methylthiazole-5-carboxylate (**6**) were prepared according to published procedures[33-35] on multi-gram scale. These amino ethyl esters were then Boc-protected and the ethyl ester hydrolyzed to produce 3-[(*tert*-butoxycarbonyl) amino]-1-

methylpyrazole-5-carboxylic acid (**5**) (Boc-Pz-OH) and 2-[(*tert*-butoxycarbonyl) amino]-4-methylthiazole-5-carboxylic acid (**7**) (Boc-Th-OH) respectively (Figure 2.4).

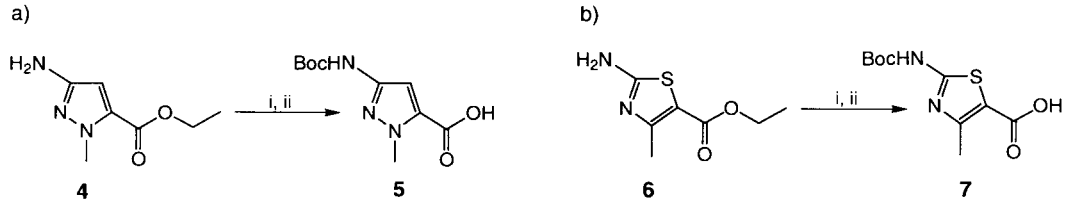


Figure 2.4. (a) Synthesis of Boc-Pz-OH (**5**). (i) Boc_2O, DIEA, DMF, (ii) 1 M NaOH; (b) Synthesis of Boc-Th-OH (**7**). (i) Boc_2O, DIEA, DMF, (ii) 1 M KOH.

Polyamide Synthesis. The polyamide resin ImImPyPy-γ-ImPyPyPy-β-Pam-resin was synthesized as previously described.[7,26] The polyamide resins ImImPzPy-γ-ImPyPyPy-β-Pam-resin and ImImThPy-γ-ImPyPyPy-β-Pam-resin were synthesized using previously described Boc-chemistry manual synthesis protocols.[25] Both Boc-Pz-OH and Boc-Th-OH monomers were activated with HBTU/DIEA and coupled for 20 h at 37 °C before termination with acetic anhydride. After Boc-deprotection of the resin-bound polyamide chain, the Boc-Im-OH was activated with HBTU/DIEA and coupled onto NH_2-PzPy-γ-ImPyPyPy-β-Pam-resin and NH_2-ThPy-γ-ImPyPyPy-β-Pam-resin at 37 °C for 24 and 40 h respectively. The terminal imidazole was incorporated as previously published.[25] A single-step aminolysis, using dimethylaminopropylamine (55 °C, 18 h), cleaved the resin ester linkage of all three polyamides from their solid support. The resin cleavage products were purified by reversed phase HPLC to give ImImPyPy-γ-ImPyPyPy-β-Dp (**1**), ImImPzPy-γ-ImPyPyPy-β-Dp (**2**), and ImImThPy-γ-ImPyPyPy-β-Dp (**3**). The polyamide ImImPzPy-γ-ImPyPyPy-β-Dp-EDTA (**2E**) was prepared by aminolysis of the

corresponding resin with 3,3'-diamino-*N*-methyldipropylamine (55 °C, 18 h) followed by reversed phase HPLC purification which afforded a free amine group at the C-terminus for postsynthetic modification. The polyamide-amine ImImPzPy-γ-ImPyPyPy-β-Dp-NH_2 (**2-NH_2**) was treated with an excess of EDTA-dianhydride (DMSO/NMP, DIEA, 55 °C, 15 min), the remaining anhydride hydrolyzed (0.1 N NaOH, 55 °C, 10 min) and the EDTA-modified polyamide purified by reversed phase HPLC.

Binding Energetics. Quantitative DNase I footprinting titration experiments[27-29] (10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, 22 °C) were performed on a 3'-^{32}P end labeled 278 bp *Eco*RI/*Pvu*II restriction fragment from the plasmid pDHN1 (Figure 2.5) to determine the equilibrium association constants for polyamides **1**, **2** and **3** at the four designed sites. On this restriction fragment, ImIm**Py**Py-γ-Im**Py**PyPy-β-Dp (**1**) bound the two match sites 5'-TGGTCA-3' and 5'-TGGACA-3'with equilibrium association constants of $K_a = 4.7\ (\pm 0.7) \times 10^9\ M^{-1}$ and $K_a = 3.1\ (\pm 0.4) \times 10^9\ M^{-1}$, respectively (Figures 2.6 & 2.7a). The mismatch sequences 5'-TGGGCA-3' and 5'-TGGCCA-3' are bound with at least 15-fold lower affinity ($K_a = 2.2\ (\pm 0.6) \times 10^8\ M^{-1}$ and $K_a = 2.5\ (\pm 0.9) \times 10^8\ M^{-1}$, respectively). ImIm**Pz**Py-γ-Im**Py**PyPy-β-Dp (**2**) bound the two sites 5'-TGGTCA-3'and 5'-TGGACA-3'with equilibrium association constants of $K_a = 2.0\ (\pm 0.5) \times 10^9\ M^{-1}$ and $K_a = 1.0\ (\pm 0.3) \times 10^9\ M^{-1}$, respectively (Figures 2.6 & 2.7b). No binding of polyamide **2** is observed within the concentration range of these DNase I footprinting titration experiments for the two sites 5'-TGGGCA-3' and 5'-TGGCCA-3' (K_a's $\leq 2.0 \times 10^7\ M^{-1}$ at both sites). Polyamide **3** containing a Th/Py pairing did not show any binding (K_a's $\leq 2.0 \times 10^7\ M^{-1}$ at all sites), at concentrations up to 50

nM, for any site (Figure 2.8). At concentrations greater than 50 nM, polyamide **3** exhibited non-specific binding for multiple DNA sites. While polyamide **2** bound its match sites with binding isotherms consistent with binding in a 1:1 polyamide-DNA complex (Figure 2.6), further characterization was necessary to assure similar behavior to pure imidazole-pyrrole polyamides.

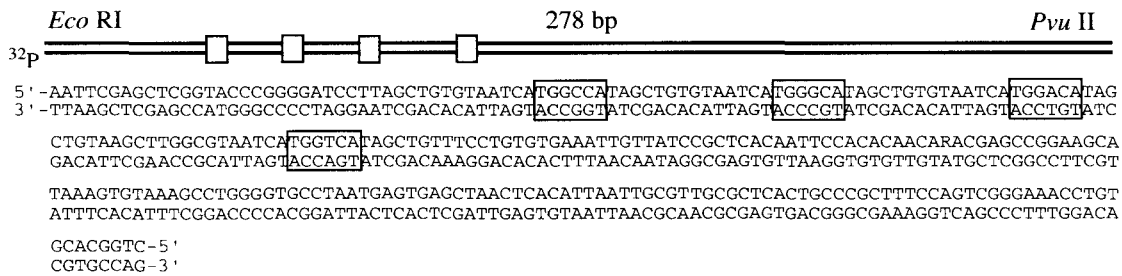


Figure 2.5. Illustration and complete sequence of the 278 bp pDHN1 *Eco*RI/*Pvu*II restriction fragment. The four designed binding sites that were analyzed in quantitative footprint titration experiments are indicated with a box surrounding each of the 6 bp sites.

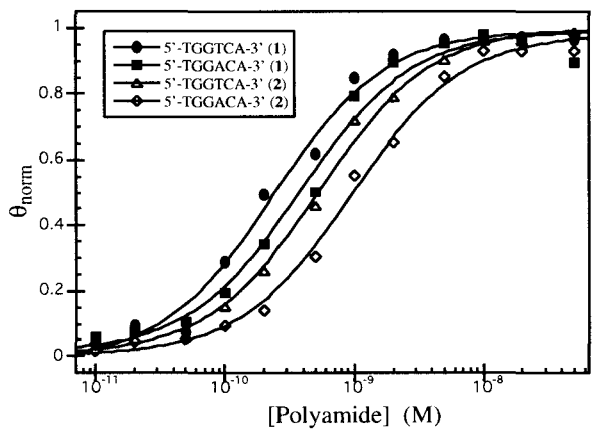


Figure 2.6. Data from quantitative DNase I footprint titration experiments for ImImPyPy-γ-ImPyPyPy-β-Dp (**1**) and ImImPzPy-γ-ImPyPyPy-β-Dp (**2**) binding to the two match sites 5'-TGGTCA-3' and 5'-TGGACA-3'. θ_{norm} points were obtained using storage phosphor autoradiography and processed by standard methods.[26, 36] The data for the binding of polyamide 1 to 5'-TGGTCA-3' is indicated by filled circles, of polyamide 1 to 5'-TGGACA-3' by filled squares, of polyamide 2 to 5'-TGGTCA-3' by open triangles, and of polyamide 2 to 5'-TGGACA-3' by open diamonds. Each data point shows the average value obtained from three footprinting experiments. The solid curves are best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm where n = 1 as previously described.[26, 36]

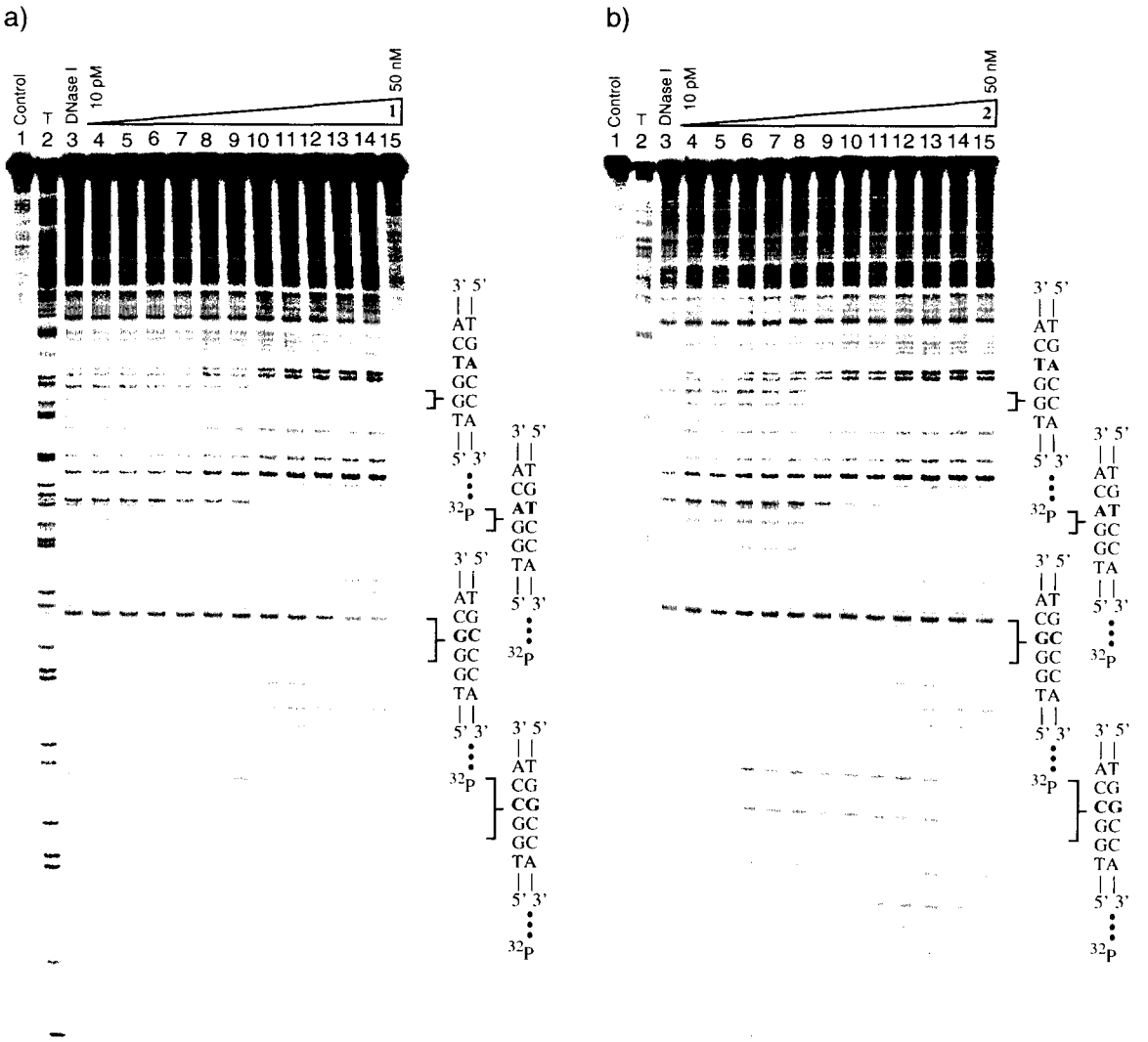


Figure 2.7. Quantitative DNase I footprint titration experiment with (a) ImImPyPy-γ-ImPyPyPy-β-Dp (**1**) and (b) ImImPzPy-γ-ImPyPyPy-β-Dp (**2**) on the 3'-end labeled 278 bp restriction fragment: lane 1, intact DNA; lane 2, T reaction; lane 3 DNase I standard; lanes 4-15 DNase I digestion products in the presence of 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM and 50 nM polyamide respectively. All four 6 bp binding sites 5'-TGGTCA-3', 5'-TGGACA-3', 5'-TGGGCA-3' and 5'-TGGCCA-3' were analyzed and are shown on the right side of the autoradiogram. All reactions contain 15 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$.

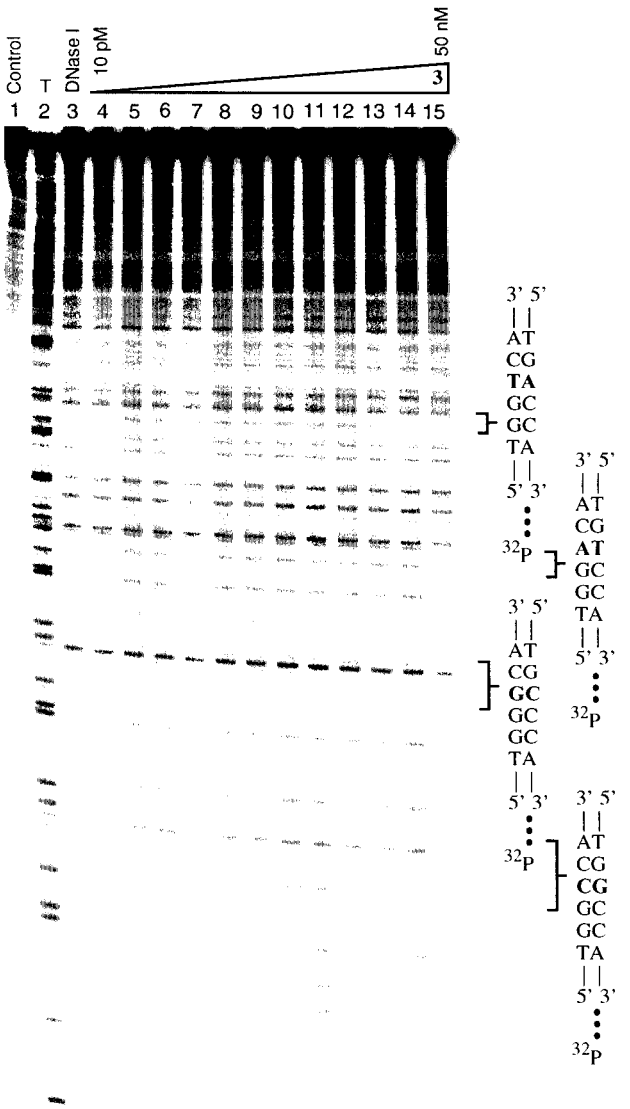


Figure 2.8. Quantitative DNase I footprint titration experiment with ImImThPy-γ-ImPyPyPy-β-Dp (**3**) on the 3'-end labeled 278 bp restriction fragment: lane 1, intact DNA; lane 2, T reaction; lane 3 DNase I standard; lanes 4-15 DNase I digestion products in the presence of 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM and 50 nM polyamide respectively. All four 6 bp binding sites 5'-TGGTCA-3', 5'-TGGACA-3', 5'-TGGGCA-3' and 5'-TGGCCA-3' were analyzed and are shown on the right side of the autoradiogram. All reactions contain 15 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$.

Binding Site Size. MPE•Fe(II) footprinting[30,31] (20 mM HEPES buffer, (pH 7.3), 200 mM NaCl, 50 μg/mL glycogen, 5 mM DTT, 0.5 μM MPE•Fe(II), and 22 °C) on 3' and 5'-^{32}P end labeled *Eco*RI/*Pvu*II 278 bp restriction fragments from the plasmid pDHN1 reveals that ImIm**Pz**Py-γ-Im**Py**PyPy-β-Dp (**2**), at 50 nM concentration, binds to and protects its designated six base pair match sites 5'-TGGTCA-3' and 5'-TGGACA-3'. Binding at the single base pair mismatch sites 5'-TGGGCA-3' and 5'-TGGCCA-3'could not be seen at the 50 nM concentration. The sizes of the asymmetrically 3'-shifted footprint cleavage patterns are consistent with a hairpin polyamide binding at six base pair site in the minor groove of DNA as observed for purely imidazole-pyrrole hairpin polyamides (Figures 2.9a and Figure 2.10).

Binding Orientation and Stoichiometry. Affinity cleavage experiments[17,32] were performed on 3' and 5'-^{32}P end labeled *Eco*RI/*Pvu*II 278 bp restriction fragments from the plasmid pDHN1 (20 mM HEPES buffer, pH 7.3, 200 mM NaCl, 50 μg/mL glycogen, 5 mM DTT, 1 μM Fe(II), and 22 °C). Affinity cleavage experiments were performed with a carboxy terminus EDTA•Fe(II) modified analog **2E** of polyamide **2**. The observed cleavage patterns for ImIm**Pz**Py-γ-Im**Py**PyPy-β-Dp-EDTA (**2E**) are 3'-shifted, consistent with minor groove occupancy (Figure 2.9b and Figure 2.10). A single cleavage locus proximal to the 5'-side of both the 5'-TGGTCA-3' and 5'-TGGACA-3' binding sites at 5 nM is consistent with a single orientation 1:1 polyamide:DNA complex.

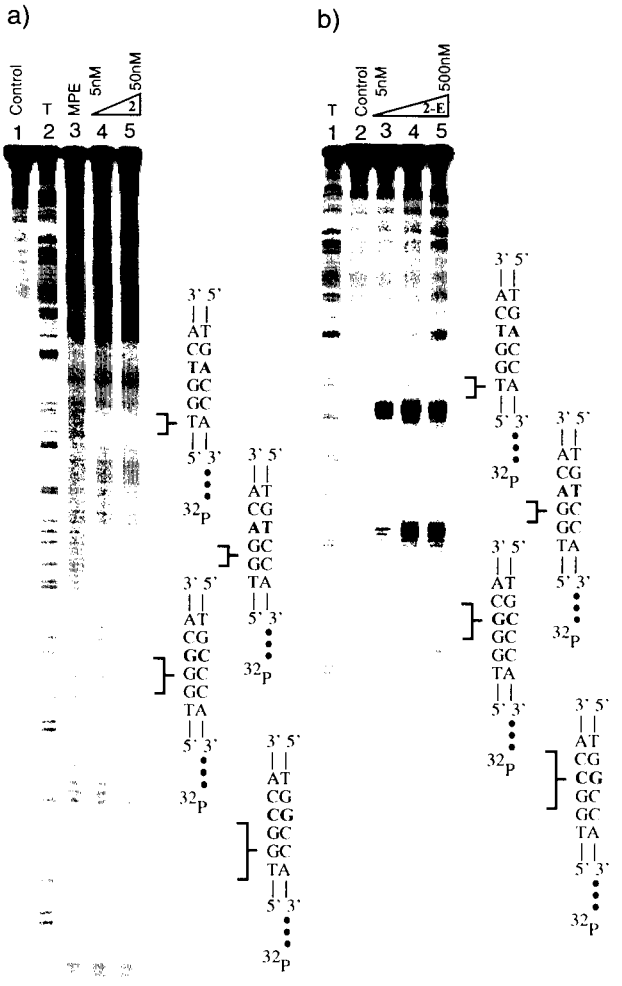


Figure 2.9. (a) MPE•Fe(II) footprinting experiment on the 3'-^{32}P end labeled 278 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pDHN1. The targeted 5'-TGGTCA-3', 5'-TGGACA-3', 5'-TGGGCA-3', and 5'-TGGCCA-3' sites are shown on the right side of the autoradiogram. Lane 1, intact DNA; lane 2, T reaction; lane 3 MPE•Fe(II) standard in the absence of any polyamide; lane 3 and 4, MPE•Fe(II) cleavage products in the presence of 5 nM and 50 nM ImImPzPy-γ-ImPyPyPy-β-Dp (**2**) respectively. All lanes contain 15 kcpm 3'-radiolabeled DNA, 20 mM HEPES buffer, (pH 7.3), 200 mM NaCl, 50 μg/mL glycogen, 5 mM DTT, and 0.5 μM MPE•Fe(II). (b) Affinity cleavage experiment on the 3'-^{32}P end labeled 278 bp *Eco*RI/*Pvu*II restriction fragment from plasmid pDHN1. The targeted 5'-TGGTCA-3', 5'-TGGACA-3', 5'-TGGGCA-3', and 5'-TGGCCA-3' sites are shown on the right side of the autoradiogram. Lane 1, T reaction; lane 2, intact DNA; lanes 3-5, affinity cleavage products in the presence of 5 nM, 50 nM, and 500 nM ImImPzPy-γ-ImPyPyPy-β-Dp-EDTA (**2E**) respectively. All lanes contain 15 kcpm 3'-radiolabeled DNA, 20 mM HEPES buffer, (pH 7.3), 200 mM NaCl, 50 μg/mL glycogen, 5 mM DTT, and 1.0 μM Fe(II).

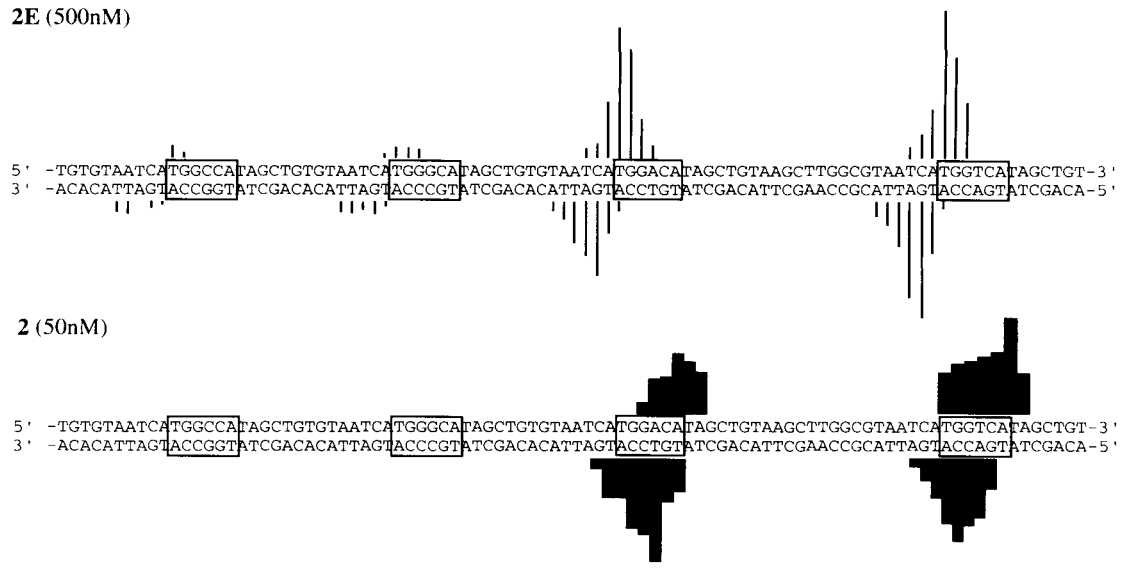


Figure 2.10. (top) Affinity cleavage patterns for ImImPzPy-γ-ImPyPyPy-β-Dp-EDTA (**2E**) at 500 nM. (bottom) MPE•Fe(II) protection patterns for ImImPzPy-γ-ImPyPyPy-β-Dp (**2**) at 50 nM.

Discussion

The Pz/Py Pair. Quantitative DNase I footprinting reveals that the incorporation of a pyrazole monomer into a hairpin polyamide results in a 100-fold increase in specificity for the match sites 5'-TGGTCA-3' and 5'-TGGACA-3' over the mismatch sites 5'-TGGGCA-3' and 5'-TGGCCA-3' due to significantly lower mismatch affinities (Table 2.1). The purely imidazole-pyrrole polyamide **1** only exhibits a 20-fold greater specificity for its match sites. Although a Pz/Py pairing exhibits slightly lower affinity for A•T/ T•A base pairs than a Py/Py pairing, polyamide **2** continues to bind to DNA with an affinity and specificity comparable to DNA binding proteins. The 2.3-fold decrease in affinity of the Pz/Py pairing relative to a Py/Py pairing may arise from some electrostatic repulsion between the N2 of pyrazole and the amide oxygen of the

polyamide backbone (Figure 2.11a). MPE•Fe(II) footprinting and affinity cleavage experiments demonstrate that ImImPzPy-γ-ImPyPyPy-β-Dp (**2**) binds its designated six base pair sites, in the minor groove of DNA, in a single orientation and as a 1:1 polyamide/DNA complex, consistent with the hairpin polyamide binding model.

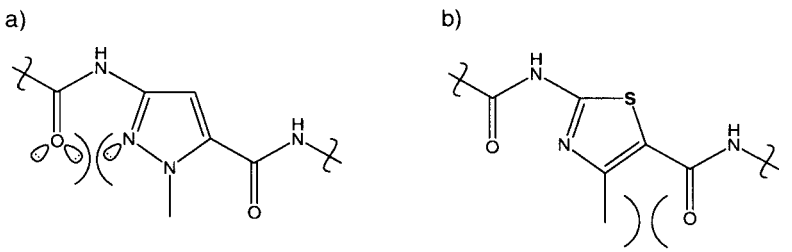


Figure 2.11. (a) Model for the electrostatic repulsion between the lone electron pairs on N2 of pyrazole and the lone electron pairs on the amide oxygen of the polyamide backbone. (b) Model for the possible steric clash between the 4-methyl of thiazole and the carbonyl of the proximal carboxamide.

Table 2.1 Equilibrium Association Constants* (M^{-1})

Polyamide	5′-caTGGTCAta-3′	5′-caTGGACAta-3′	5′-caTGGGCAta-3′	5′-caTGGCCAta-3′
Py/Py	5′-T G G T C A-3′ 3′-A C C A G T-5′	5′-T G G A C A-3′ 3′-A C C T G T-5′	5′-T G G G C A-3′ 3′-A C C C G T-5′	5′-T G G C C A-3′ 3′-A C C G G T-5′
	$K_a = 4.7 \times 10^9$	$K_a = 3.1 \times 10^9$	$K_a = 2.2 \times 10^8$	$K_a = 2.5 \times 10^8$
Pz/Py	5′-T G G T C A-3′ 3′-A C C A G T-5′	5′-T G G A C A-3′ 3′-A C C T G T-5′	5′-T G G G C A-3′ 3′-A C C C G T-5′	5′-T G G C C A-3′ 3′-A C C G G T-5′
	$K_a = 2.0 \times 10^9$	$K_a = 1.0 \times 10^9$	$K_a \leq 2 \times 10^7$	$K_a \leq 2 \times 10^7$
Th/Py	5′-T G G T C A-3′ 3′-A C C A G T-5′	5′-T G G A C A-3′ 3′-A C C T G T-5′	5′-T G G G C A-3′ 3′-A C C C G T-5′	5′-T G G C C A-3′ 3′-A C C G G T-5′
	$K_a \leq 2 \times 10^7$	$K_a \leq 2 \times 10^7$	$K_a \leq 2 \times 10^7$	$K_a \leq 2 \times 10^7$

*The reported equilibrium association constants are the mean values obtained from three DNase I footprint titration experiments. Assays were carried out in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0 and 22C. The new monomer pyrazole is indicated with an open cirlce containing a Z (Ⓩ) while thiazole is represented as an open circle with a T inside (Ⓣ).

The Th/Py Pair. It has been proposed in the literature that the degeneracy of A•T/ T•A recognition by polyamides could be overcome by using a thiazole (Th) monomer that would be sterically destabilized when placed opposite a thymine base.[23, 24] Thiazole was therefore proposed to be an *A-specific* recognition element.[23,24] Conversely, recent studies[7-9] on hydroxypyrrole (Hp) have shown that steric destabilization, arising from the bulky hydroxyl group, occurs only when Hp is located opposite adenine, resulting in the loss of DNA binding affinity. Since this steric destabilization is absent and an extra hydrogen bond is formed when hydroxylpyrrole is placed opposite thymine,[8] Hp is a *T-specific* recognition element. While the Hp structural data[8] reduces the plausibility of thiazole conferring A-specificity as proposed, it does allow for the possibility of thiazole being *T-specific* based solely on steric discrimination. Therefore, the singular steric discrimination by thiazole was interesting in light of the dual steric and hydrogen-bonding components of T/A differentiation by hydroxypyrrole (Hp). The quantitative DNase I footprinting titrations reveal that the binding affinity of an eight-ring hairpin polyamide is compromised by the incorporation of a single 4-methylthiazole monomer. It is puzzling that the 4-methylthiazole/*N*-methylpyrrole pair (Th/Py) has low affinity for both A•T and T•A. Perhaps different analogs of this ring system should be investigated further. For example, what is the role of the methyl substituent at the 4-position? With the change in both length and bond angle of the thiazole core,[21] one could imagine an energetically unfavorable steric clash between the 4-methyl and the carbonyl of the proximal carboxamide (Figure 2.11b).

Implications for the Design of Minor Groove Binding Polyamides. The results presented here reveal that the *N*-methylpyrazole ring might replace the *N*-methylpyrrole

moiety for minor groove DNA recognition. The ability to generate alternative pairing schemes for the minor groove with new monomers could greatly expand the targetable sequence repertoire of hairpin polyamides (Table 2.2). As issues of biodistribution in animal studies become important, an expanded set of pairing rules such as Pz/Pz or Im/Pz would be of interest to examine in this series for biological applications.

Table 2.2 Polyamide Ring Pairings

	T•A	A•T	G•C	C•G
Py/Py	+	+	-	-
Pz/Py	+	+	-	-
Th/Py	-	-	-	-

Experimental Section

Materials. Dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBt), 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), 0.75 mmol/g Boc-β-alanine- (4-carboxamidomethyl)-benzyl-ester-copoly (styrene-divinylbenzene) resin (Boc-β-Pam-Resin), and Boc-γ-aminobutyric acid were purchased from Peptides International. *N*, *N*-Diisopropylethylamine (DIEA), *N*, *N*-dimethylformamide (DMF), 1:1 dimethylsulfoxide: *N*-methylpyrrolidinone (DMSO/NMP) and acetic anhydride (Ac_2O) were purchased from Applied Biosystems. Dichloromethane (DCM) and triethylamine (TEA) were reagent grade from EM. Thiophenol (PhSH), dimethylaminopropylamine (Dp), and 3,3'diamino-N-

methyldipropylamine (Dp-NH_2) and EDTA-dianhydride were purchased from Aldrich. Trifluoroacetic acid (TFA) Biograde was from Halocarbon. All reagents were used without further purification.

^1H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz in DMSO-d_6 with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. High-resolution FAB mass spectra were recorded at the Mass Spectroscopy Laboratory at University of California, Riverside. DNA sequencing was performed at the Sequence/Structure Analysis Facility (SAF) at the California Institute of Technology. HPLC analysis was performed on a Beckman Gold System using a RAINEN C_{18}, Microsorb MV, 5 μm, 300 x 4.6 mm reversed phase column in 0.1% (wt/v) TFA with acetonitrile as eluent and a flow rate of 1.0 ml/min, gradient elution 1.25% acetonitrile/min. Preparatory reversed phase HPLC was performed on a Beckman HPLC with a Waters BondaPak 25 x 100 mm, 100 μm C_{18} column equipped with a guard, 0.1% (wt/v) TFA, 8.0 ml/min, 0.25% acetonitrile/min. 18MΩ water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μm filtered.

Monomer Synthesis. 3-[(*tert*-Butoxycarbonyl)amino]-1-methylpyrazole-5-carboxylic Acid (5). The amino ethyl ester **4** (6.0 g, 35.5 mmol) was dissolved in 100 ml of DMF and DIEA (12.3 ml, 70.6 mmol). To the stirred solution di-*tert*-butyl-dicarbonate (15.4 g, 70.6 mmol) was added and allowed to react overnight. The solution

was poured into water and extracted with diethyl ether. The organic layers were combined, dried with $MgSO_4$, and concentrated *in vacuo* to provide a white solid that was then dissolved in ethanol. 1 M NaOH was added and the solution heated at 45 °C for 1 h. The clear solution was diluted with water and the ethanol removed *in vacuo*. The solution was extracted with diethyl ether, the pH of the aqueous layer reduced to 3 with 10% (v/v) H_2SO_4 and the mixture extracted with ethyl acetate. The combined organic layers were dried with $MgSO_4$ and concentrated *in vacuo* to provide **5** as a white powder (3.24 g, 38% yield): ^{1}H NMR (DMSO-d_6) δ 13.32 (br s, 1H), 9.73 (s, 1H), 6.70 (s, 1H), 3.95 (s, 3H), 1.45 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 161.5, 153.7, 146.8, 133.8, 101.4, 80.2, 39.5, 29.0; FABMS m/e 264.0958 (M + NA 264.0960 calcd for $C_{10}H_{15}N_3O_4$ +Na).

2-[(*tert*-Butoxycarbonyl)amino]-4-methylthiazole-5-carboxylic Acid (7). The amino ethyl ester **6** (10 g, 53.5 mmol) and di-*tert*-butyl-dicarbonate (23.4 g, 107 mmol) were dissolved in 150 ml of DMF and DIEA (19 ml, 107 mmol). The stirred reaction was heated to 50 °C for 3 days. The reaction was allowed to cool to room temperature and diluted with 350 ml of water. The pH of the aqueous layer was reduced to 3 with 10% (v/v) H_2SO_4 and the mixture extracted with diethyl ether. The organic layers were combined, dried with $MgSO_4$, and concentrated *in vacuo* to provide a suspended solid. This solid was collected by vacuum filtration and washed with hexanes to provide a tan powder that was then dissolved in ethanol. 1 M KOH was added and the solution heated to 70 °C for 2 h. The mixture was poured into 250 ml of water and washed with diethyl ether, the pH of the aqueous layer reduced to 3 with 10% (v/v) H_2SO_4, and the milky white solution extracted with diethyl ether. The combined organic extracts were concentrated *in vacuo* to provide **7** as a white solid (7.1 g, 27.8 mmol, 52% yield): ^{1}H

NMR (DMSO-d_6) δ 12.5 (br s, 1H), 11.8 (br s, 1H), 2.4 (s, 3H), 1.4 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 164.6, 162.2, 156.8, 153.7, 115.9, 82.8, 28.8, 17.8; EIMS m/e 259.0751 (M + H 259.0752 calcd for $C_{10}H_{15}N_2O_4S$).

ImImPzPy-γ-ImPyPyPy-β-Dp (2). ImImPzPy-γ-ImPyPyPy-β-Pam resin was synthesized in a stepwise fashion by manual solid-phase protocols[25] from Boc-β-alanine-Pam resin (1.0 g, 0.75 mmol/g). Boc-Pz-OH (241 mg, 1 mmol) and HBTU (360 mg, 0.95 mmol) were combined in 2 mL of DMF. DIEA (1 ml) was then added and the reaction mixture allowed to stand for 5 min. The activated monomer was added to the reaction vessel, containing NH_2-Py-γ-ImPyPyPy-β-Pam resin, and the coupling was allowed to proceed for 20 h at 37 °C before termination with acetic anhydride. After Boc-deprotection of the resin-bound polyamide chain, the Boc-Im-OH was activated with HBTU/DIEA as previously described and coupled onto NH_2-PzPy-γ-ImPyPyPy-β-Pam-resin at 37 °C for 24 h. The terminal imidazole was incorporated as previously published. A sample of ImImPzPy-γ-ImPyPyPy-β-Pam resin (300 mg, 0.44 mmol/g) was placed in a 20 ml scintillation vial, 2 ml of dimethylaminopropylamine was added, and the mixture was allowed to stand at 55 °C for 18 h. Resin was removed through filtration through a disposable propylene filter, and the resulting solution diluted with water to a total volume of 8 ml and purified directly by reversed phase HPLC to provide ImImPzPy-γ-ImPyPyPy-β-Dp (**2**) (5.7 mg, 3.5% recovery) as a white powder upon lyophilization of the appropriate fractions. ^{1}H NMR (DMSO-d6) δ 10.47 (s, 1H), 10.29 (s, 1H), 10.14 (s, 1H), 10.04 (s, 1H), 10.02 (s, 1H), 9.97 (s, 1H), 9.92 (s, 1H), 9.25 (br s, 1H), 8.1 (m, 3H), 7.68 (s, 1H), 7.48 (s, 2H), 7.43 (s, 1H), 7.28 (s, 1H), 7.2 (m, 2H), 7.18 (s, 2H), 7.11 (s, 1H),

7.08 (s, 1H), 6.96 (s, 1H), 6.89 (s, 1H), 4.06 (s, 3H), 4.04 (s, 3H), 4.02 (s, 3H), 3.97 (s, 3H), 3.87 (s, 3H), 3.85 (s, 3H), 3.83 (s, 3H), 3.81 (s, 3H), 3.4 (m, 2H), 3.2(m, 2H), 3.1 (m, 2H), 3.0 (m, 2H), 2.7 (m, 6H), 2.4 (m, 4H), 1.8 (m, 4H); MALDI-TOF-MS (monoisotopic), 1224.63 (1224.57 calcd for M + H).

ImImPzPy-γ-ImPyPyPy-β-Dp-NH_2 (2-NH_2). A sample of ImImPzPy-γ-ImPyPyPy-β-Pam resin (250 mg, 0.44 mmol/g) was placed in a 20 ml scintillation vial, 2 ml of 3,3'-diamino-*N*-methyldipropylamine was added, and the mixture was allowed to stand at 55 °C for 18 h. Resin was removed through filtration through a disposable propylene filter, and the resulting solution diluted with water to a total volume of 8 ml and purified directly by reversed phase HPLC to provide ImImPzPy-γ-ImPyPyPy-β-Dp-NH_2 (**2-NH_2**) (8.2 mg, 5.9% recovery) as a white powder upon lyophilization of the appropriate fractions. ^{1}H NMR (DMSO-d6) δ 10.47 (s, 1H), 10.29 (s, 1H), 10.14 (s, 1H), 10.03 (s, 2H), 9.98 (s, 1H), 9.92 (s, 1H), 9.25 (br s, 1H), 8.1 (m, 3H), 7.82 (br s, 3H), 7.68 (s, 1H), 7.48 (s, 2H), 7.43 (s, 1H), 7.29 (s, 2H), 7.2 (m, 1H), 7.18 (s, 1H), 7.1 (m, 2H), 6.95 (s, 2H), 6.90 (s, 1H), 4.06 (s, 3H), 4.04 (s, 3H), 4.02 (s, 3H), 3.97 (s, 3H), 3.87 (s, 3H), 3.85 (s, 3H), 3.83 (s, 3H), 3.81 (s, 3H), 3.2-3.0 (m, 8H), 2.9 (m, 2H), 2.5 (m, 4H), 2.37 (br s, 4H), 2.0-1.7 (br m, 6H); MALDI-TOF-MS (monoisotopic), 1267.78 (1267.61 calcd for M + H).

ImImPzPy-γ-ImPyPyPy-β-Dp-EDTA (2E). EDTA-dianhydride (50 mg) was dissolved in 1 ml of DMSO/NMP solution and 1 ml of DIEA by heating at 55 °C for 5 min. The dianhydride solution was added to ImImPzPy-γ-ImPyPyPy-β-Dp-NH_2 (**2-NH_2**) (8.2 mg, 6.5 μmol) dissolved in 750 μl of DMSO. The mixture was heated to 55 °C for

25 min, treated with 3 ml of 0.1 N NaOH, and heated to 55 °C for 10 min. TFA (0.1%) was added to adjust the total volume to 8 ml, and the solution purified directly by preparatory reversed phase HPLC chromatography to provide **2E** as a white powder (3.6 mg, 36% recovery). MALDI-TOF-MS (monoisotopic), 1542.16 (1542.81 calcd for M + H).

ImImThPy-γ-ImPyPyPy-β-Dp (3). ImImPzPy-γ-ImPyPyPy-β-Pam resin was synthesized in a stepwise fashion by manual solid-phase protocols[25] from Boc-β-alanine-Pam resin (1.0 g, 0.75 mmol/g). Boc-Th-OH (258 mg, 1 mmol) and HBTU (360 mg, 0.95 mmol) were combined in 2 ml of DMF. DIEA (1 ml) was then added and the reaction mixture allowed to stand for 5 min. The activated monomer was added to the reaction vessel, containing NH_2-Py-γ-ImPyPyPy-β-Pam resin, and the coupling was allowed to proceed for 20 h at 37 °C before termination with acetic anhydride. After Boc-deprotection of the resin-bound polyamide chain, the Boc-Im-OH was activated with HBTU/DIEA as previously described and coupled onto NH_2-ThPy-γ-ImPyPyPy-β-Pam-resin at 37 °C for 40 h. The terminal imidazole was incorporated as previously published. A sample of resin (150 mg, 0.43 mmol/g) was placed in a 20 ml scintillation vial, 2 ml of dimethylaminopropylamine was added, and the mixture was allowed to stand at 55 °C for 18 h. Resin was removed by filtration through a disposable polypropylene filter, and the resulting solution diluted with water to a total volume of 8 ml and purified directly by reversed phase HPLC to provide ImImThPy-γ-ImPyPyPy-β-Dp (1.5 mg, 1.8% recovery) as a white powder upon lyophilization of the appropriate fractions. ^{1}H NMR (DMSO-d6) δ 10.29 (s, 1H), 10.05 (s, 1H), 10.02 (s, 1H), 9.98 (s, 3H), 9.92 (s, 1H), 9.2 (br s, 1H), 8.1

(m, 3H), 7.78 (s, 1H), 7.49 (s, 2H), 7.32 (s, 1H), 7.28 (s, 1H), 7.26 (s, 1H), 7.24 (s, 2H), 7.18 (s, 1H), 7.10 (s, 1H), 6.9 (m, 2H), 3.99 (s, 3H), 3.97 (s, 3H), 3.91 (s, 3H), 3.82 (s, 3H), 3.80 (s, 3H), 3.7 (m, 9H), 3.4 (m, 2H), 3.2(m, 2H), 3.1 (m, 2H), 3.0 (m, 2H), 2.7 (m, 6H), 2.4 (m, 4H), 1.8 (m, 4H); MALDI-TOF-MS (monoisotopic), 1241.67 (1241.53 calcd for M + H).

DNA Reagents and Materials. Enzymes were purchased from Boeringer-Mannheim and used with their supplied buffers. Deoxyadenosine and thymidine 5'-[α-32P] triphosphates were obtained from Dupont/NEN. Calf thymus DNA (sonicated, deproteinized), DNase I (7500 u/ml, FPLC pure), Tris•HCl, dithiothreitol (DTT), RNase-free water (used for all footprinting and affinity cleavage reactions), and 0.5 M EDTA were purchased from Amersham Pharmacia. XGal and IPTG were from ICN. Ampicillin trihydrate was acquired from Sigma. Absolute ethanol was purchased from Equistar. Calcium chloride, potassium chloride, and magnesium chloride were from Fluka. Formamide and pre-mixed tris-borate-EDTA (Gel-Mate) were from Gibco. All reagents were used without further purification. DNA manipulations were performed according to standard protocols.[36]

Construction of Plasmid DNA. The plasmid pDHN1 was prepared by hybridization of a complementary set of synthetic oligonucleotides: 5'-GATCCTTAGCT GTGTAATCATGGCCATAGCTGTGTAATCATGGGCATAGCTGTGTAATCATGGA CATAGCTGTA-3' and 5'-AGCTTACAGCTATGTCCATGATTACACAGCTATGCCCA TGATTACACAGCTATGGCCATGATTACACAGCTAAG-3'. The hybridized insert was ligated into *Bam*HI/*Hin*dIII linearized pUC19 using T4 DNA ligase. *Epicurean coli* XL-1 Blue supercompetent cells were then transformed with the ligated plasmid, and plasmid

DNA from ampicillin-resistant white colonies isolated using a Promega Maxi-Prep kit. The presence of the desired insert was determined by dideoxy sequencing.

Preparation of ^{32}P-End-Labeled Restriction Fragments. Plasmid pDHN1 was linearized with *Eco*RI and *Pvu*II restriction enzymes, then treated with Sequenase, deoxyadenosine 5'-[α-^{32}P] triphosphate, and thymidine 5'-[α-^{32}P] triphosphate for 3' labeling. The 5'-labeling of the desired restriction fragment was achieved through PCR amplification using the primers [γ-^{32}P]-DHN1-A (5'-[γ-^{32}P]-AATTCGAGCTCGGTACC CGGG-3') and DHN1-B (5'-CTGGCACGACAGGTTTCCCGA-3'). The labeled fragment (3' or 5') was loaded onto a 7% nondenaturing polyacrylamide gel, and the desired 278 bp band was visualized by autoradiography and isolated.

DNase I Footprinting. All reactions were carried out in a volume of 400 µl. We note explicitly that no carrier DNA was used in these reactions until after DNase I cleavage. A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were 10 mM Tris•HCl buffer (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, and 15 kcpm 3'-radiolabeled DNA. The solutions were equilibrated for a minimum of 12 h at 22 °C. Cleavage was initiated by the addition of 10 µl of a DNase I stock solution (diluted with 1 mM DTT to give a stock concentration of 0.75 u/ml) and was allowed to proceed for 7 min at 22 °C. The reactions were stopped by adding 50 µl of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/ml glycogen, and 30 µM base pair calf thymus DNA and then ethanol-precipitated. The cleavage products were resuspended in 100 mM Tris-borate-EDTA/80% formamide loading buffer, denatured at 85 °C for 10 min, and immediately loaded onto an 8%

denaturing polyacrylamide gel (5% cross-link, 7 M urea) at 2000 V for 1 h 45 min. The gels were dried under vacuum at 80 °C and then quantitated using storage phosphor technology. Equilibrium association constants were determined as previously described.[26-29] It should be noted that footprinting of polyamide **3** at concentrations higher than 50 nM resulted in non-specific binding to DNA and therefore an exact association constant could not be determined.

MPE•Fe(II) Footprinting. All reactions were carried out in a volume of 400 µl. A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were 20 mM HEPES buffer, (pH 7.3), 200 mM NaCl, 50 µg/ml glycogen, 5 mM DTT, and 0.5 µM MPE•Fe(II). Equilibration proceeded for 10 min after the addition of 40 µl of 5 µM MPE•Fe(II) (freshly made by combining equal volumes of 10 µM MPE and 10 µM $Fe(NH_4)_2(SO_4)_2$). The reactions were initiated with 10 µl of 200 mM DTT and cleavage allowed to proceed for 14 min. The reactions were stopped by ethanol precipitation, resuspended in 100 mM Tris-borate-EDTA/80% formamide loading buffer, denatured at 85 °C for 10 min, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% cross-link, 7 M urea) at 2000 V for 1 h 45 min.

Affinity Cleavage. All reactions were carried out in a volume of 400 µl. A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were 20 mM HEPES buffer, (pH 7.3), 200 mM NaCl, 50 µg/ml glycogen, 5 mM DTT, and 1.0 µM Fe(II). Equilibration proceeded for 20 min after the addition of 20 µl of 10 µM $Fe(NH_4)_2(SO_4)_2$. The reactions were initiated with 40 µl of 50 mM DTT and cleavage allowed to proceed for 30 min. The reactions were stopped

by ethanol precipitation, resuspended in 100 mM Tris-borate-EDTA/80% formamide loading buffer, denatured at 85 °C for 10 min, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% cross-link, 7 M urea) at 2000 V for 1 h 45 min.

Acknowledgements

We are grateful to the National Institutes of Health (GM 27681) for research support, National Institutes of Health for a research service award to J.W.S., The Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B. and the Natural Sciences and Engineering Research Council of Canada for a postgraduate scholarship to D.H.N.

References

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

2. Dickinson, L. A.; Guzilia, P.; Trauger, J. W.; Baird, E. E.; Mosier, D. M.; Gottesfeld, J. M.; Dervan, P. B. *Proceedings of the National Academy of Sciences USA* **1998**, *95*, 12890.

3. Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B. *Proceedings of the National Academy of Sciences USA* **2000**, *97*, 3930.

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

5. Dervan, P. B.; Burli, R. W. *Current Opinion in Chemical Biology* **1999**, *3*, 688.

6. Wemmer, D. E.; Dervan P. B. *Current Opinion in Structural Biology* **1997**, *7*, 355.

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

8. Kielkopf, C. L.; White, S. E.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111.

9. White, S. E.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1999**, *121*, 260.

10. Wu, H.; Crothers, D. M. *Nature* **1984**, *308*, 509.

11. Stietz, T. A. *Annual Reviews in Biophysics* **1990**, *23*, 205.

12. Goodsell, D. S.; Kopka, M. L.; Cascio, D.; Dickerson, R. E. *Proceedings of the National Academy of Sciences USA* **1993**, *90*, 2930.

13. Paolella, D. N.; Palmer, R.; Schepartz, A. *Science* **1994**, *264*, 1130.

14. Kahn, J. D.; Yun, E.; Crothers, D. M. *Nature* **1994**, *368*, 163.

15. Geierstanger, B. H.; Wemmer, D. E. *Annual Reviews in Biochemistry* **1995**, *24*, 463.

16. Hansen, M. R.; Hurley, L. H. *Accounts of Chemical Research* **1996**, *29*, 249.

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

18. Dervan, P. B. *In Nucleic Acids and Molecular Biology*; Springer-Verlag: Heidelberg, 1988; Vol. 2, 49-64.

19. Thuong, N. T.; Helene, C. *Angewendte Chemie* **1993**, *32*, 666.

20. Nielsen, P. E. *Chemistry-a European Journal* **1997**, *3*, 505.

21. Rao, K. E.; Shea, R. G.; Yadagiri, B.; Lown, J. W. *Anti-Cancer Drug and Design* **1990**, *5*, 3.

22. Sharma, S. K.; Tandon, M.; Lown, J. W. *Journal of Organic Chemistry* **2000**, *65*, 1102.

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

24. Walker, W. L.; Landaw, E. M.; Dickerson, R. E.; Goodsell, D. S. *Proceedings of the National Academy of Sciences USA* **1998**, *95*, 4315.

25. Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6141.

26. Bremer, R. E.; Baird, E. E.; Dervan, P. B. *Chemistry & Biology* **1998**, *5*, 119.

27. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods in Enzymology* **1986**, *130*, 132.

28. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proceedings of the National Academy of Sciences USA* **1986**, *83*, 8462.

29. Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.

30. Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proceedings of the National Academy of Sciences USA* **1982**, *79*, 5470.

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

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

33. Ding, L.; Grehn, L.; Ragnarsson, U. *Acta Chemica Scandia* **1990**, *44*, 75.

34. Chenard, B. L. *Journal of Organic Chemistry* **1984**, *49*, 1224.

35. Dodson, R. M.; King, L. C. *Journal of the American Chemical Society* **1945**, *67*, 2242.

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

Chapter Three

Influence of C-Terminal Tails on DNA Recognition by Pyrrole/Imidazole Polyamides

The text of this chapter was taken in part from a publication co-authored with Jason W. Belitsky, Nicholas R. Wurtz and Professor Peter B. Dervan.

Publication: Belitsky, J.W.; Nguyen, D.H.; Wurtz, N.R.; Dervan, P.B. *Bioorganic & Medicinal Chemistry*, in press.

Abstract

Pyrrole-imdazole polyamides are synthetic ligands that bind to predetermined sequences of DNA in the minor groove. Typically, solid phase methods using Boc monomers for synthesis have depended on Boc-β-Ala-PAM resin which affords a β-alanine-Dp tail at the C-terminus, after cleavage with *N*, *N*-dimethylaminopropylamine (Dp). New solid phase strategies, including Boc chemistry on Kaiser oxime resin, have permitted the synthesis of polyamides with truncated C-terminal tails in high yield. Thermodyanimc characterization of the DNA binding properties of polyamides having carboxylic acid, ethyl, methyl, and primary amide tails shows increased generality, in the tail region, without loss in affinity relative to the standard β-Dp tail obtained from Boc-β-Ala-PAM resin synthesis. The increased generality of the truncated tails expands the number of DNA sequences that can be targeted specifically by polyamides by removing the two A/T base pair requirement of the β-Dp tail. The high affinity, subnanomolar in the cases of the methyl and ethyl amide tails, of this series of polyamides, coupled with their lower molecular weights, may influence their utility in biological systems.

Introduction

Hairpin polyamides containing *N*-methylpyrrole (Py), *N*-methylimidazole (Im), and *N*-methyl-3-hydroxypyrrole (Hp) residues bind specific pre-determined sequences in the minor groove of DNA with affinities and specificities comparable to naturally occurring DNA-binding proteins.[1] DNA recognition depends on the side-by-side amino acid pairings in the minor groove that stack the aromatic rings against each other and the walls of the groove, allowing backbone amide hydrogens and the substituents at the 3-position of the Py, Im, and Hp residues to make specific contacts with the edges of the DNA base pairs. The antiparallel pairing of imidazole opposite pyrrole (Im/Py) selectively recognizes a G•C base pair, while a Py/Im pair recognizes a C•G base pair. A Py/Py pair binds with equal affinity to A•T and T•A and in preference to G•C and C•G. However, the unsymmetrical Hp/Py pair discriminates T•A over A•T base pairs. These heterocyclic amino acids make up what may be considered the "core" of the polyamide. An alkyl amino acid, either β-aminobutyric acid (GABA, γ), or the chiral, amine-functionalized derivative (*R*)-2,4,-diaminobutyric acid (DABA, $(R)^{H_2N}\gamma$), serves as the covalent linker region (referred to as the "turn") between the N-terminal and C-terminal strands. The "tail" or C-terminal portion of a typical hairpin polyamide is made up of β-alanine (β) and *N,N*-dimethylaminopropylamine (Dp).[1]

The β-Dp tail originated with the C→N synthesis of polyamides from Boc-β-alanine-PAM-resin.[2] An alkyl amino acid is necessary for efficient cleavage from the solid support with aminolysis by Dp, and β-alanine was found to be optimal for affinity and specificity.[3] Investigation of the β-Dp tail, GABA and DABA turns has shown that A•T or T•A are preferred for the base pairs immediately flanking the core of the

polyamide (tail-1 and turn-1 positions) as well as the second base pair from the core on the tail side of the polyamide (tail-2 position, Figure 3.1).[4] Swalley *et al.* also showed that the replacement of the β-Dp tail with a propanol amide, generated by the reduction of the β-alanine-PAM ester off the solid support,[5] removed the preference for an A•T or T•A base pair at the tail-2 position.[4]

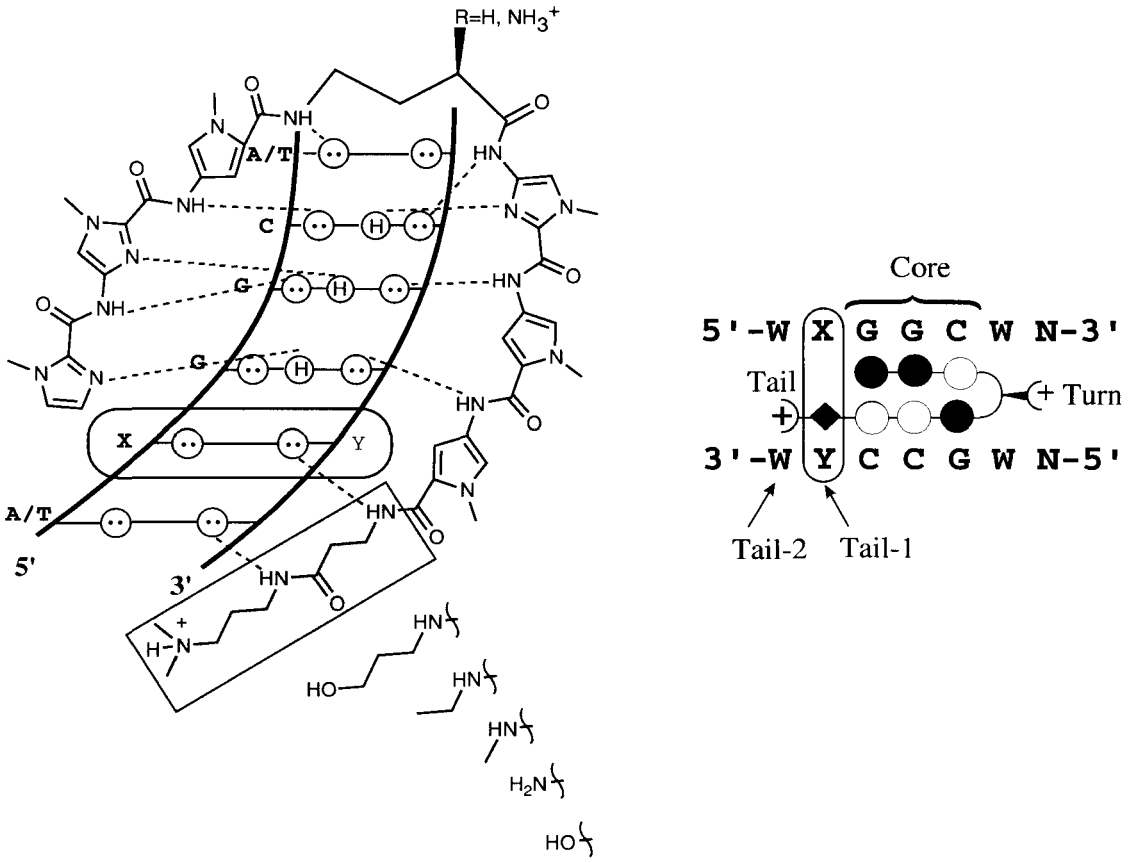
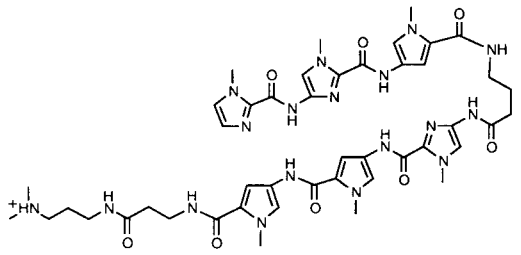
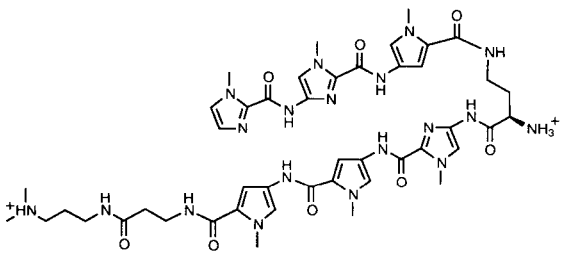


Figure 3.1. Binding model for polyamide **1** and **2**, with DNA. (Top) Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogens of guanine. Putative hydrogen bonds are illustrated by dotted lines. (Bottom) A ball and stick representation of polyamide **1** with DNA. Filled circles denote imidazole while open circles represent pyrrole. The diamond represents β-alanine. γ-aminobutyric acid (γ) and dimethylaminopropylamine (Dp) are depicted as a curved line and a plus sign, respectively. W signifies A or T; N represents any nucleotide. The Tail-2, Tail-1, Core, and Turn positions for the polyamide/DNA complex are defined as shown.

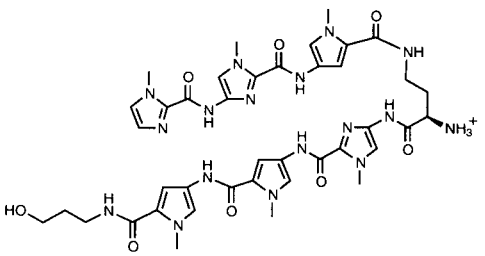
Since specificity for any combination of base pairs can, in principle, be encoded by the heterocyclic amino acids, it may not be necessary for the tail to have sequence specificity. In fact, while a large number of biologically relevant sequences have been targeted by the current generation of hairpin polyamides,[6] the necessity for A•T/T•A base pairs flanking the polyamide core is a limitation. Promoters are known to have high G,C content associated with CpG islands,[7] and many transcription factors bind G,C rich sequences.[8] In order to effectively compete with, and thus inhibit these transcription factors, polyamides having greater generality at the tail positions, while maintaining high affinity, are desirable.



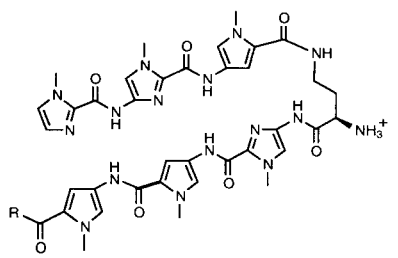
1 ImImPy-γ-ImPyPy-β-Dp



2 ImImPy-(*R*)^{H_2N}γ-ImPyPy-β-Dp



3 ImImPy-(*R*)^{H_2N}γ-ImPyPy-PrOH



4 ImImPy-(*R*)^{H_2N}γ-ImPyPy-CO- R=NHEt
5 ImImPy-(*R*)^{H_2N}γ-ImPyPy-CO- R=NHMe
6 ImImPy-(*R*)^{H_2N}γ-ImPyPy-CO- R=NH_2
7 ImImPy-(*R*)^{H_2N}γ-ImPyPy-CO- R=OH

Figure 3.2. Structures of polyamides ImImPy-γ-ImPyPy-β-Dp (**1**), ImImPy-(*R*) ^{H_2N}γ-ImPyPy-β-Dp (**2**), ImImPy-(*R*)^{H_2N}γ-ImPyPy-PrOH (**3**), ImImPy-(*R*)^{H_2N}γ-ImPyPy-CONHEt (**4**), ImImPy-(*R*)^{H_2N}γ-ImPyPy-CONHMe (**5**), ImImPy-(*R*)^{H_2N}ImPyPy-$CONH_2$ (**6**), and ImImPy-(*R*)^{H_2N}γ-ImPyPy-CO_2H (**7**) as synthesized by solid phase methods.

In an attempt to extend this generality to the tail-1 position, hairpin polyamides with C-termini further truncated to ethyl amide, methyl amide, carboxylic acid, and primary amide tails were examined. Since Boc-β-Ala-PAM resin does not allow these tails to be accessed readily, new synthetic strategies were also investigated. We describe here the synthesis of a series of six-ring polyamide hairpins with the general sequence ImImPy-$(R)^{H_2N}$γ-ImPyPy-**X** where X represents the various truncated C-terminal tails. The DNA-binding properties of these new compounds were also investigated; in particular, their specificity at the tail-1 position was compared to the specificity of the same polyamide core having either a β-Dp or propanol amide tail.

Results

Polyamide Synthesis. The synthesis of both polyamides ImImPy-γ-ImPyPy-β-Dp (**1**), ImImPy-$(R)^{H_2N}$γ-ImPyPy-β-Dp (**2**) and ImImPy-$(R)^{H_2N}$γ-ImPyPy-PrOH (**3**) has previously been described.[4] The core of the entire series of polyamides with novel truncated tails **4-7** (Figure 3.2) was designed to be the same as that of compounds **1-3** in order to facilitate direct comparison with the data found in the tail study by Swalley *et al.*[4] There is no straightforward way to synthesize compounds **4-7** from the standard Boc-β-Ala-Pam resin. While the adoption of other resins to polyamide synthesis may be the most significant aspect of this study, initially polyamide **6** was synthesized using a derivative of Boc-β-Ala-Pam resin. During an investigation of the DNA binding properties of an aminoxy acid analog[9-13] of β-alanine,[14] it was noted that the chemistry of the N-O bond could be exploited to produce the desired C-terminal pyrrole primary amide. Aminoxy acids are hydroxamic ester analogs of β-amino acids where the β-carbon is replaced by oxygen. The well known reduction of hydroxamic esters to amides[15,16] was envisioned as a method to cleave the novel polyamide **6** from β-Ala-Pam

resin. The successful implementation of this strategy is shown below; however, it may suffer from limited generality. Upon further investigation, the straightforward synthesis of polyamide **6** was also readily accomplished by two different solid phase routes: utilizing recently developed Fmoc-protected monomers[17] with Rink resin, or using standard Boc-protected monomers with Kaiser oxime resin. The latter is particularly significant because the oxime resin proved a versatile support for the synthesis of polyamides with a variety of different C-termini (compounds **4-7**).

Boc/Hydroxamic Ester Route. In order to cleave a hydroxamic ester unit to the desired primary amide, it first had to be incorporated in a manner consistent with the standard solid phase conditions.[2] Difficulties with primary aminoxy groups on solid phase have been previously noted.[18] After considerable experimentation, this problem was circumvented by the one step, solution phase synthesis of the Boc-pyrrole-hydroxamic ester (**8**) derived from the standard HOBt activated pyrrole monomer[2] (**9**) and commercially available carboxymethoxylamine hemihydrochloride (Figure 3.3). The synthesis of **6** (Figure 3.4) began with the deprotection of commercially available Boc-β-Ala-PAM resin, followed by coupling of *N*-hydroxysuccinimide (NHS) activated **8**, in DMF for six hours at room temperature with DIEA as a base. Removal of the Boc group of dimer **8** was accomplished as previously reported and five standard couplings[2,19] followed to yield ImImPy-(*R*)^{FmocHN}γ-ImPyPyCONHOCH$_2$CO-β-Ala-PAM resin. Some degradation was observed in a coupling which requires heating at 37 °C, suggesting that the hydroxamic ester unit is best handled at room temperature, which may limit the generality of this method. Attempts to remove the Fmoc group with piperidine led to unexpected cleavage from the solid support.[20] Therefore, the Fmoc group was not removed prior to cleavage from resin. After surveying various reducing agents including zinc dust[21] and triphenylphosphine, best results for the reduction of the hydroxamic ester in the full length polyamide resin were obtained with palladium acetate and ammonium

formate.[22] The Fmoc group was also removed under these reductive conditions[23] to yield the novel truncated tail polyamide, ImImPy-$(R)^{H_2N}\gamma$-ImPyPyCONH$_2$ (**6**).

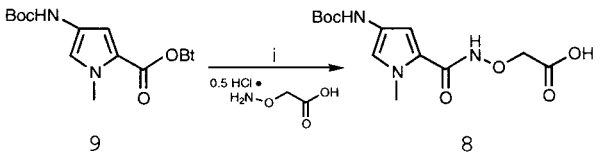


Figure 3.3. Synthesis of Boc-Pyrrole-Hydroxamic Ester **8**: i) DIEA, DMF, room temperature, 10 hours.

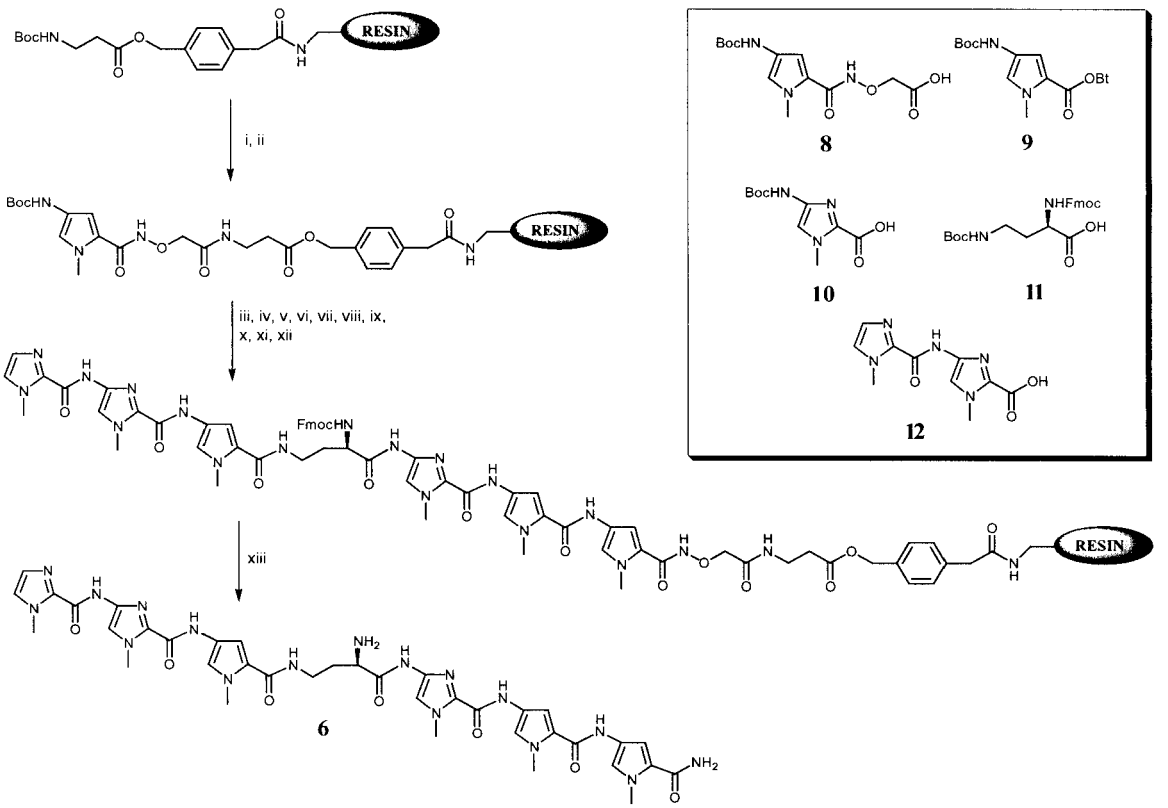


Figure 3.4. Boc/Hydroxamic Ester route to polyamide **6** starting from Boc-β-Ala-PAM resin: (i) 80% TFA/DCM, 0.4 M PhSH; (ii) **8**, DCC, NHS, DIEA, DMF; (iii) 80% TFA/DCM, 0.4 M PhSH; (iv) **9**, DIEA, DMF; (v) 80% TFA/DCM, 0.4 M PhSH; (vi) **10**, HBTU, DIEA, DMF, 37 °C; (vii) 80% TFA/DCM, 0.4 M PhSH; (viii) **11**, HBTU, DIEA, DMF; (ix) 80% TFA/DCM, 0.4 M PhSH; (x) **9**, DIEA, DMF; (xi) 80% TFA/DCM, 0.4 M PhSH; (xii) **12**, HBTU, DIEA, DMF (xiii) $Pd(OAc)_2$, HCO_2NH_4, H_2O, DMF, 37 °C.

Fmoc/Rink Resin Route. The synthesis of Fmoc-protected Py (**13**) and Im (**14**) monomers and the production of β-Dp tail polyamides from Fmoc-β-Ala-Wang resin has recently been described.[17] The Fmoc-protected monomers can also be used for the straightforward synthesis of polyamides with C-terminal primary amides. Commercially available Fmoc-Rink resin was deprotected with piperidine, and reacted with Fmoc-Py-OBt (**13**). Six standard cycles of deprotection and coupling[17] yielded ImImPy-(*R*)^{BocHN}γ-ImPyPyCONH-Rink resin which was cleaved with TFA to yield polyamide **6** (Figure 3.5). Similar cleavage yields and purities were obtained compared to synthesis of the analogous polyamide with standard Boc protocols on Boc-β-Ala-PAM resin followed by aminolysis with Dp.[2]

Boc/Oxime Resin Route. The oxime resin proved a versatile support for the synthesis of polyamides with a variety of different C-termini (compounds **4-7**) while still permitting the use of Boc-monomers. Developed by Kaiser and DeGrado,[24,25] the oxime resin is a versatile polystyrene solid support that is amenable to Boc chemistry and allows for the synthesis of a variety of carboxylic acid derivatives by direct nucleolytic cleavage of the peptide from resin. Peptides can be cleaved from oxime resin to yield primary amide,[25] alkyl amide,[26] and carboxylic acid,[27,28] C-termini. The oxime linker is reported to be somewhat acid labile, and solutions of less than 25% TFA are recommended for the deprotection of Boc groups. Consequently, a 20% TFA/CH_2Cl_2 solution was used, for 30 minutes, to remove the Boc groups from pyrrole and aliphatic amines. However, it was found that Boc-imidazole residues could not be fully deprotected under these conditions. Longer deprotection times were attempted but resulted in increasing degradation of the resin bound polyamide. After screening various conditions, it was found that Boc groups could be cleanly and efficiently removed from imidazoles in 30 minutes with a 50% TFA/CH_2Cl_2 solution. Surprisingly, no premature cleavage of the oxime linker was observed at this higher concentration of acid. Apparently, attachment of the first pyrrole

unit makes a bulky aromatic oxime that stabilizes the linker relative to the aliphatic oxime present when conventional amino acids are used.

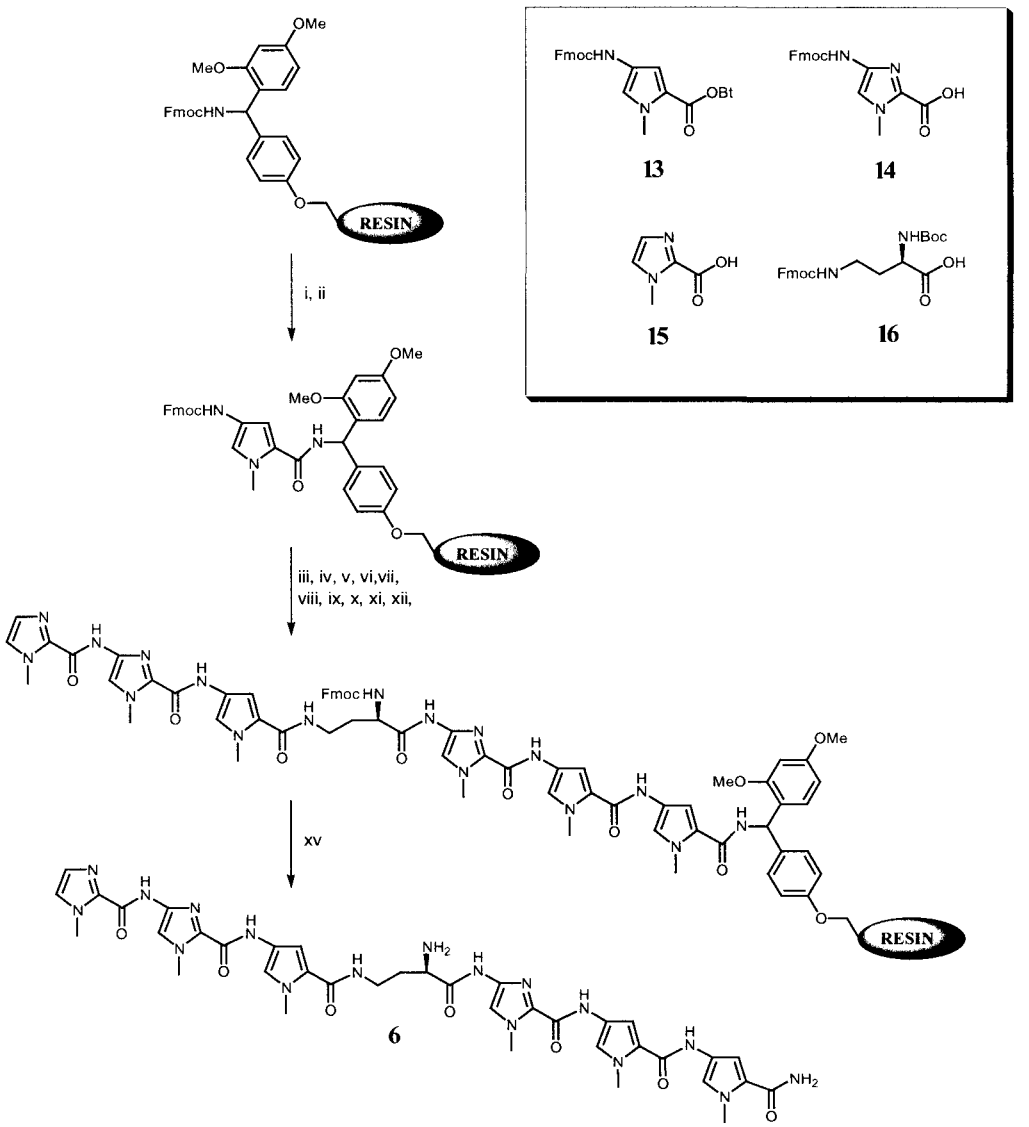


Figure 3.5. Synthesis of polyamide **6** from Fmoc-Rink resin using Fmoc-protected monomers. (i) 20% piperidine/NMP; (ii) **13**, HBTU, DIEA; (iii) 20% piperidine/NMP; (iv) **13**, HBTU, DIEA; (v) 20% piperidine/NMP; (vi) **14**, HBTU, DIEA; (vii) 20% piperidine/NMP; (viii)**16** acid, HBTU, DIEA; (ix) 20% piperidine/NMP; (x) **13**, HBTU, DIEA; (xi) 20% piperidine/NMP; (xii) **14**, HBTU, DIEA; (xiii) 20% piperidine/NMP; (xiv) **15**, HBTU, DIEA; (xv) 20% TFA/DCM.

Polyamides **4-7** were generated from the common intermediate, ImImPy-$(R)^{FmocHN}\gamma$-ImPyPy-Oxime resin **17** (Figure 3.6). Commercially available oxime resin was reacted directly with Boc-Py-OBt (**9**) in *N*-methylpyrrolidone (NMP) and diisopropylethylamine (DIEA) overnight at room temperature. The subsequent deprotections and couplings were carried out in a stepwise manner using the previously outlined deprotection conditions. All other monomers used, namely **10**, **11** and **12**, were activated with HBTU and coupled in NMP/DIEA for 1.5-8 h at either room temperature or 37 °C. Unlike the hydroxamic ester route, no degradation of the oxime resin was observed upon heating. All cleavages of resin **17** required a co-solvent (CH_2Cl_2 or DMF) to swell the resin and/or solubilize the polyamide. Based on literature methods,[25] cleavage of resin **17** with a saturated solution of ammonia in THF/CH_2Cl_2 yielded polyamide **6**. However, even after more than 60 h at 37 °C, only 60% cleavage was observed. Addition of a large excess of DBU, which had been used as a transacylation catalyst to generate peptide acids and esters from oxime resin previously,[27] resulted in quantitative cleavage of the resin to generate polyamide **6**. The C-terminal carboxylic acid polyamide **7** was generated in a similar manner with 1:1 H_2O:DMF, 30 eq. DBU for 60 h at 37 °C. This method was superior to acidic, basic, or reductive cleavage conditions. It is interesting to note that the classic transacylation catalyst, DMAP, was totally ineffective as a replacement for DBU in these cleavages. Both polyamides **4** and **5** were quantitatively cleaved from resin **17** with a 1.0 M solution in THF/CH_2Cl_2 of ethylamine or methylamine, respectively overnight at 37 °C. The oxime resin is clearly a highly versatile way to generate modified polyamides cleanly and in good yield while remaining amenable to Boc chemistry.

Quantitative DNase I Footprinting. It has been shown previously that replacement of the β-Dp tail in polyamide **1** with propanol (**3**) diminishes specificity at the tail-2 position (Figure 3.7). Furthermore, thermodynamic characterization of doubly

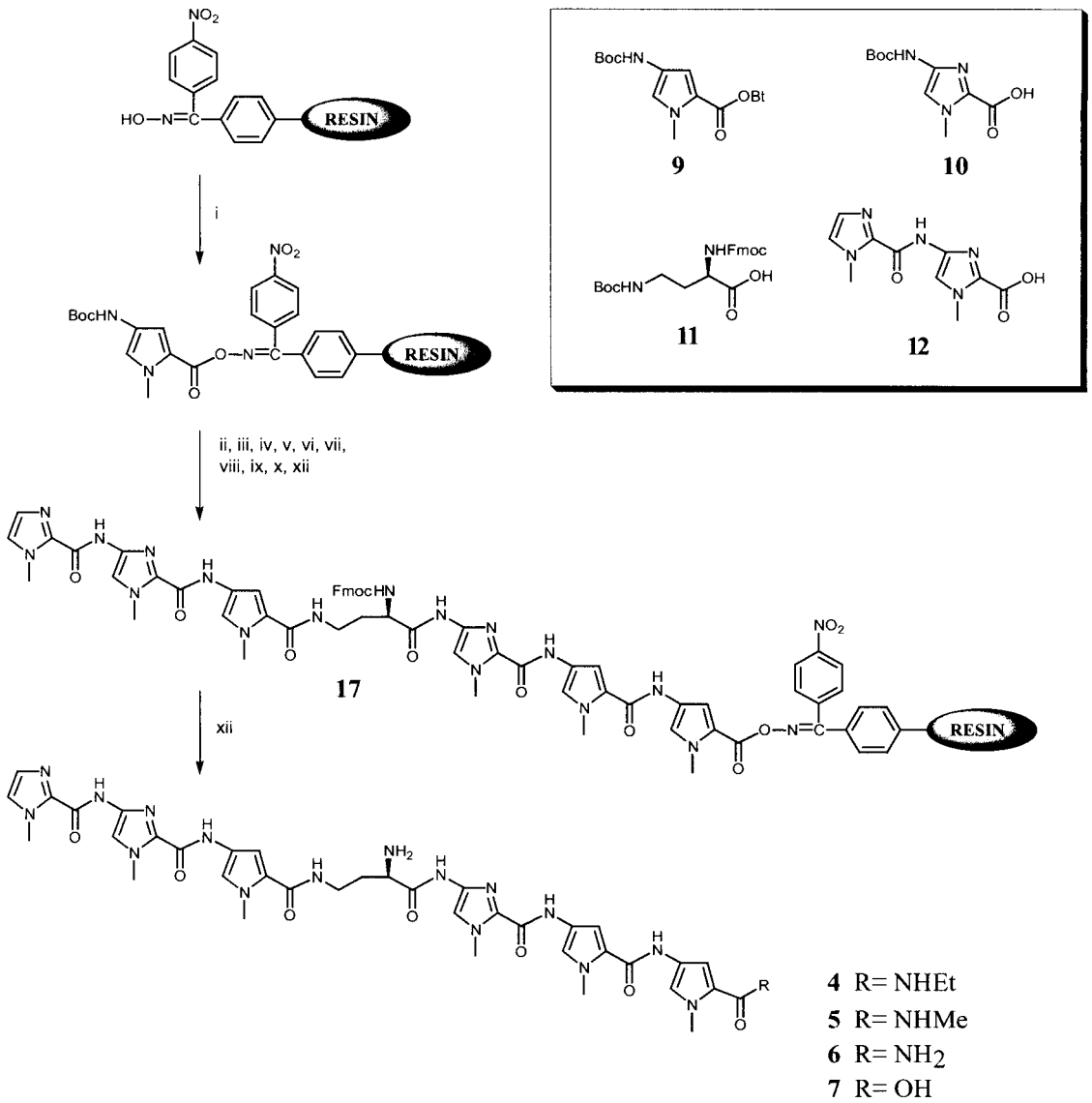
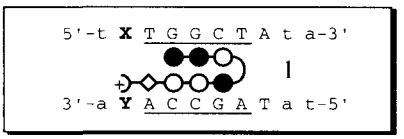


Figure 3.6. Synthesis of polyamide **4-7** from oxime resin using Boc-protected monomers: (i) **9**, DIEA, NMP; (ii) 20% TFA/DCM; (iii) **9**, DIEA, NMP; (v) 20% TFA/DCM; (v) **10**, HBTU, DIEA, NMP; (vi) 50% TFA/DCM; (vii) **11**, HBTU, DIEA, NMP, 37 °C; (viii) 20% TFA/DCM; (ix) **9**, DIEA, NMP; (x) 20% TFA/DCM; (xi) **12**, HBTU, DIEA,NMP; (xii) a) R = NHEt, 2.0 M $CH_3CH_2NH_2$/THF, DCM, 37 °C; or b) R = NHMe, 2.0 M CH_3NH_2/THF, DCM, 37 °C; or c) R = NH_2, NH_3/THF, DCM, DBU, 37 °C; or d) R = OH, H_2O, DMF, DBU, 37 °C.

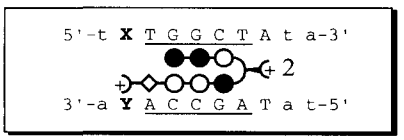
charged β-Dp tail polyamide **2** has shown a two-fold decrease in specificity from **1** at the tail-2 position. While empirical observations[29] have suggested that the propanol amide tail, the β-Dp tail with the chiral turn, is A•T/T•A specific at the tail-1 position, these had

not been characterized in detail. The DNA binding affinity and specificity of the four new C-terminal polyamides at the tail-1 position against all four Watson Crick base pairings were also evaluated.

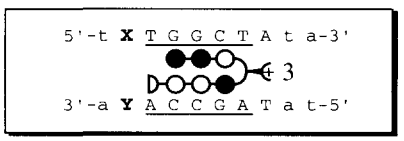
Tail 2 Position



X·Y	Ka's	Specificity
A·T	6.2 X 10^7(±0.2)	44
T·A	2.8 X 10^7(±0.4)	20
G·C	1.4 X 10^6(±1.4)	1
C·G	2.0 X 10^6(±0.8)	1

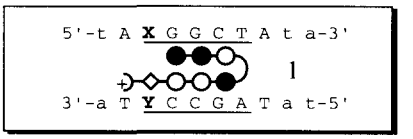


X·Y	Ka's	Specificity
A·T	2.3 X 10^9(±0.8)	23
T·A	1.1 X 10^9(±0.4)	11
G·C	1.0 X 10^8(±0.3)	1
C·G	1.0 X 10^8(±0.8)	1

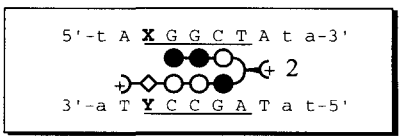


X·Y	Ka's	Specificity
A·T	1.1 X 10^9(±0.3)	4
T·A	3.4 X 10^8(±0.8)	1
G·C	7.0 X 10^8(±2.1)	2
C·G	2.7 X 10^8(±0.4)	1

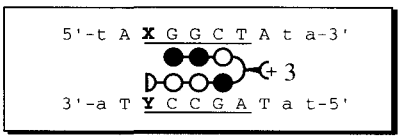
Tail 1 Position



X·Y	Ka's	Specificity
A·T	3.1 X 10^7(±1.1)	> 310
T·A	2.1 X 10^7(±0.8)	> 210
G·C	< 10^5	1
C·G	< 10^5	1



X·Y	Ka's	Specificity
A·T	1.7 X 10^9(±0.5)	53
T·A	1.6 X 10^9(±0.2)	50
G·C	5.8 X 10^7(±1.4)	2
C·G	3.2 X 10^7(±0.3)	1



X·Y	Ka's	Specificity
A·T	5.9 X 10^8(±0.4)	140
T·A	4.7 X 10^8(±0.4)	112
G·C	2.2 X 10^7(±0.3)	5
C·G	4.2 X 10^6(±0.7)	1

Figure 3.7. Specificity of polyamides **1**, **2** and **3**. a) Equilibrium association constants at the tail-2 position with restriction fragment pSES-TL2 as determined by Swalley *et al.*[4] b) Equilibrium association constants at the tail-1 position with restriction fragment pSES-TL1 as previously determined for polyamide **1** and as determined in this present study for polyamide **2** and **3**. Values reported are the mean values from at least three DNase I footprinting titration experiments with the standard deviation for each data set in parentheses. Assays were performed at 22 °C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. Specificity is calculated as $K_a(X•Y)/K_a(C•G)$.

β-Dp Tail with Chiral Turn. Quantitative DNase I footprinting titrations[30-32] were performed to evaluate polyamide **2** at the tail-1 position (data not shown). Equilibrium association constants (K_a) of polyamide **2** were determined for four sites on the restriction fragment pSES-TL1,[4] corresponding to the sequences 5'-ANGGCTA-3', where N represents each DNA base (Figure 3.7). As expected, doubly charged polyamide 2 continues to show specificity for A•T (K_a = 1.7 $\times 10^9$ M^{-1}) and T•A (K_a = 1.6 $\times 10^9$ M^{-1}), with 52- and 50-fold higher affinity for those sites than for C•G (K_a = 3.2 $\times 10^2$ M^{-1}), at that same position. The additional positive charge results in a substantial gain in affinity (about a factor of 50) over the singly charged β-Dp tail parent polyamide **1**. These changes in affinity and specificity for a doubly charged polyamide have been observed frequently.

Propanol Amide Tail. Quantitative DNase I footprinting titrations[30-32] were performed to evaluate polyamide **3** at the tail-1 position (Figure 3.8A). As expected, the propanol amide tail confers specificity for A•T (K_a = 5.9 $\times 10^8$ M^{-1}) and T•A (K_a = 4.7 $\times 10^8$ M^{-1}), with 140- and 112-fold higher affinity for those sites than for C•G (K_a = 4.2 $\times 10^6$ M^{-1}), and 28- and 22-fold higher affinity than G•C (K_a = 2.2 $\times$ 10^7 M^{-1}) at that position. Polyamide **3** showed a small 5-fold preference for G•C over C•G, consistent with a greater steric clash with the exocyclic amine of guanine when guanine is located on the bottom strand of DNA. At this tail-1 position, the parent β-Dp polyamide **1** displays 2-fold greater selectivity for A•T /T•A versus C•G base pairs relative to polyamide **3** (Figure 3.7) while **3** displays 3-fold greater selectivity over the doubly charged polyamide **2**. However, the magnitudes of the equilibrium association constants for **3** are approximately 20-fold higher than those observed for **1**. This increase in affinity is consistent with the displacement of the charge from the C-terminus to either the turn or the *N*-methyl position on a ring[19,33] and was also observed on the restriction fragment pSES-TL2.[4]

Methyl and Ethyl Amide Tails. From the footprinting titrations of polyamides with either a ethyl (**4**) or methyl (**5**) amide tail (Figure 3.8B) several features are notable. Both polyamides **4** and **5** exhibit much higher affinity for A•T ($K_a = 1.1 \times 10^9$ M^{-1} and $K_a = 1.1 \times 10^9$ M^{-1}, respectively) and T•A ($K_a = 9.2 \times 10^8$ M^{-1} and $K_a = 1.4 \times 10^9$ M^{-1}, respectively) base pairs compared to G•C ($K_a = 8.2 \times 10^7$ M^{-1} and $K_a = 2.4 \times 10^8$ M^{-1}, respectively) and C•G ($K_a = 3.0 \times 10^7$ M^{-1} and $K_a = 1.1 \times 10^8$ M^{-1}, respectively). Both of these alkyl amide tails have specificity ranges intermediate between those of the primary amide/acid tails and that of the propanol amide tail (Figure 3.10). The methyl amide tail has a 13-fold range between its highest and lowest affinity sites (T•A and C•G, respectively) while the ethyl amide tail shows a 37-fold range between the A•T and C•G sites (highest and lowest affinity, respectively). However, the surprising result was that both of these alkyl amide tails generated singly charged, six-ring hairpin polyamides that bind to their match sites at subnanomolar concentrations. This is about 35-fold higher affinity than the parent β-Dp polyamide (**1**) and comparable to the doubly charged polyamide (**2**).

Primary Amide Tail. Quantitative DNase I footprinting titrations of polyamide **6** (Figure 3.9A) revealed equilibrium association constants ranging from 2.1×10^8 M^{-1} for T•A to 3.5×10^7 M^{-1} for C•G (Figure 3.10). Within this 6-fold range, A•T ($K_a = 1.9 \times 10^8$ M^{-1}) and T•A remain the preferred sites. While a 6-fold range is not normally considered degenerate, it is clear that the primary amide tail is more permissive for C•G and G•C ($K_a = 6.8 \times 10^7$ M^{-1}) base pairs at the tail-1 position than either the β-Dp or propanol amide tails. This 6-fold decrease in affinity of primary amide tail (**6**) for the C•G site from its most preferred site is quite small compared to the ≥310-fold, 53-fold

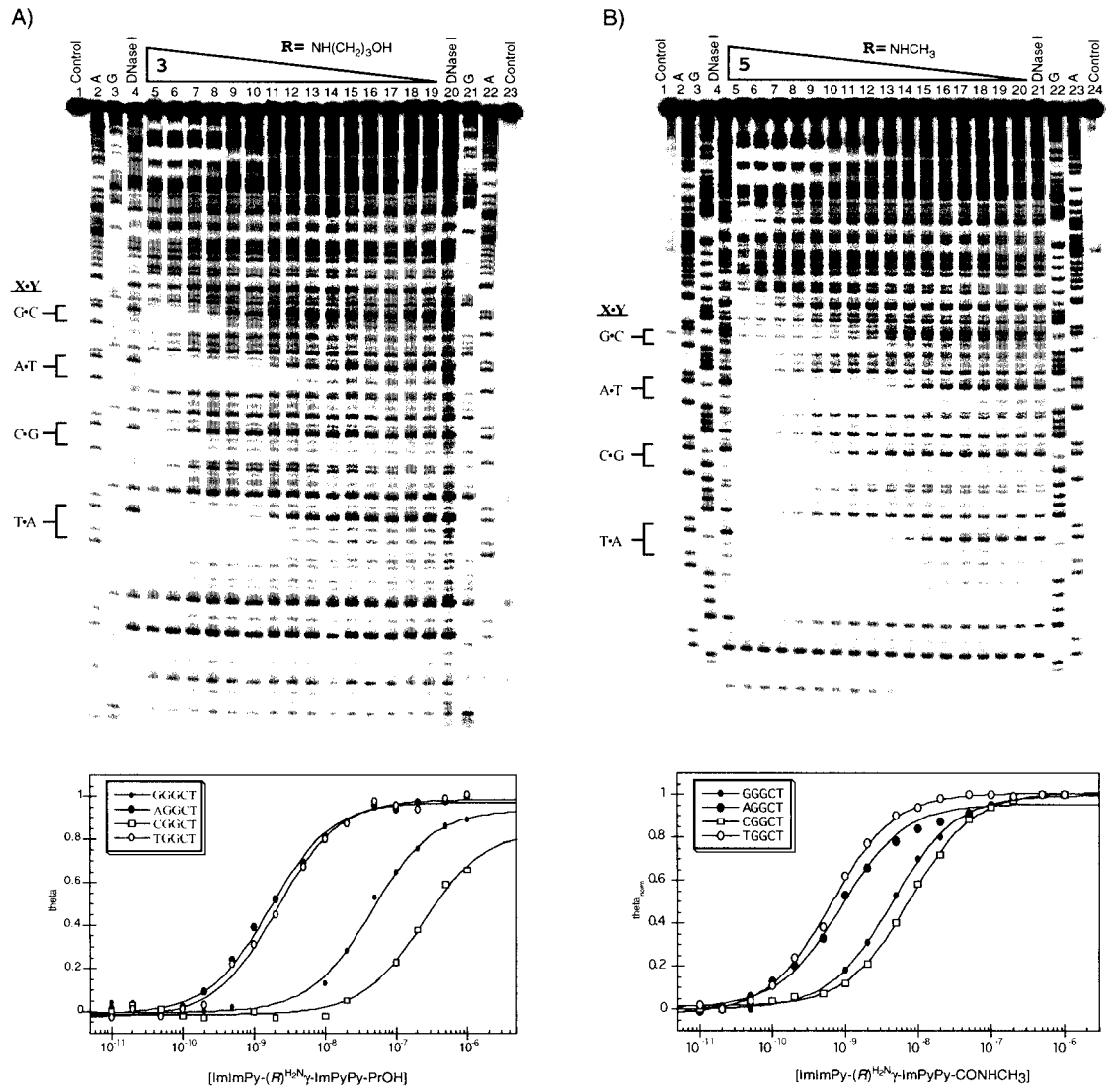


Figure 3.8. A) Quantitative DNase I footprint titration experiment with ImImPy-$(R)^{H_2N}\gamma$-ImPyPyPrOH (**3**) on the 3'-^{32}P-labeled 286-bp restriction fragment pSES-TL1[4]: lane 1 & 23, intact DNA; lane 2 & 22, A reaction; lane 3 & 21, G reaction; lane 4 & 20 DNase I standard; lanes 5-19 DNase I digestion products in the presence of 1 μM, 500 nM, 200 nM, 100 nM, 20 nM, 10 nM, 5 nM, 2 nM, 1 nM, 500 pM, 200 pM, 100 pM, 50 pM, 20 pM, 10 pM polyamide respectively. B) Quantitative DNase I footprint titration experiment with ImImPy-$(R)^{H_2N}\gamma$-ImPyPyCONHCH$_3$ (**5**) on the 3'-^{32}P-labeled 286-bp restriction fragment pSES-TL1[4]: lane 1 & 24, intact DNA; lane 2 & 23, A reaction; lane 3 & 22, G reaction; lane 4 & 21 DNase I standard; lanes 5-20 DNase I digestion products in the presence of 1 μM, 500 nM, 200 nM, 100 nM, 50 nM, 20 nM, 10 nM, 5 nM, 2 nM, 1 nM, 500 pM, 200 pM, 100 pM, 50 pM, 20 pM, 10 pM polyamide respectively. All reactions contain 15 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. Data was obtained for the four binding sites indicated at the left of the gel, 5'-GGGCT-3', 5'-AGGCT-3', 5'-CGGCT-3', and 5'-TGGCT-3', and is shown in the isotherm plot below. θ_{norm} points were obtained using storage phosphor autoradiography and processed by standard methods. Each data point shows the average value obtained from three footprinting experiments. The solid curves are best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm where n = 1 as previously described.[3]

and 140-fold disfavoring of C•G base pairs by polyamide **1**, **2** and **3**, respectively. Furthermore, while the values of the equilibrium association constants of polyamide **6** at the A•T and T•A sites are mildly reduced (3- and 2-fold respectively) compared to those observed for **3**, the K_a's for G•C and C•G are in fact 3- and 8-fold higher for polyamide **6** than for **3**. It is also worth noting that the magnitude of the equilibrium association constants for polyamide **6** with its least preferred site (C•G, $K_a = 3.5 \times 10^7\ M^{-1}$) is quite similar to that of β-Dp polyamide **1** for its most preferred site (A•T, $K_a = 3.1 \times 10^7\ M^{-1}$).

Carboxylic Acid Tail. Polyamides with C-terminal carboxylic acid tails have previously been synthesized as intermediates to cyclic polyamides; however, their DNA binding properties have never been investigated.[22,34] DNase I footprinting of polyamide **7** (Figure 3.9B) revealed that the carboxylic acid tail was also degenerate in its affinity for all four Watson Crick base pairings. The range between its most preferred site, A•T ($K_a = 2.1 \times 10^7\ M^{-1}$), and its least preferred site, C•G ($K_a = 4.9 \times 10^6\ M^{-1}$), was only 4-fold (Table 2). While the affinity of polyamide **7** for G•C ($K_a = 1.1 \times 10^7\ M^{-1}$) and C•G base pairs is very similar to polyamides **3** and **6**, its affinity for A•T ($K_a = 2.1 \times 10^7\ M^{-1}$) and T•A ($K_a = 1.2 \times 10^7\ M^{-1}$) were significantly lower by at least 10-fold relative to either the propanol (**3**) or primary amide tail (**6**) at those sites. However, the carboxylic acid tail allows binding at all sites with affinities comparable to the β-Dp tail (**1**) for its match A,T sites (Figure 3.10).

Discussion

Structural Basis of Recognition. In sequences with an A•T or T•A to the C-terminal side of the C-terminal pyrrole ring, it has long been argued that recognition

originates from the C-terminal pyrrole secondary amide making a hydrogen bond with the N3 of adenine and/or the O2 of thymine.[35] NMR and crystal structures have provided evidence for this supposition.[36-40] The similar affinities of polyamides **6** for all four base pairs suggest that the C-terminal pyrrole primary amide can also make hydrogen bonds to the N3 of guanine and/or the O2 of cytosine. Additionally, increased affinity for G•C and C•G sites was observed for the primary amide tail (**6**) relative to the propanol amide tail (**3**). Thus, the methylene groups of the propanol amide and β-Dp tails likely account for their A,T specificity, presumably via a steric clash with the exocyclic amine of guanine. The influence of the exocyclic amine on specificity can be inferred from the observation that the C•G site was the lowest affinity site for all of the polyamides examined in this study. This site places a guanine base on the bottom strand of the DNA where the steric clash between the tail and the exocyclic amine of guanine should be strongest.

Lacking methylene groups, the carboxylic acid tail polyamide (**7**) recognizes all four base pairs degenerately (within 4-fold of the C•G site). To act as a hydrogen bond donor, in analogy to the amides, the carboxylic acid would have to be protonated in the minor groove. Thus, the 10-fold lower affinity than the primary amide tail may be due to loss of this hydrogen bond. However, the protonation state of the acid in the minor groove is unknown, and several factors, including a larger desolvation penalty, could account for the reduced affinity. In water **7** is expected to be a zwitterion with a net charge of zero, which one would expect to have a lower affinity for polyanionic DNA than the positively charged polyamides **1**-**6**. Given these factors, **7** binds with higher

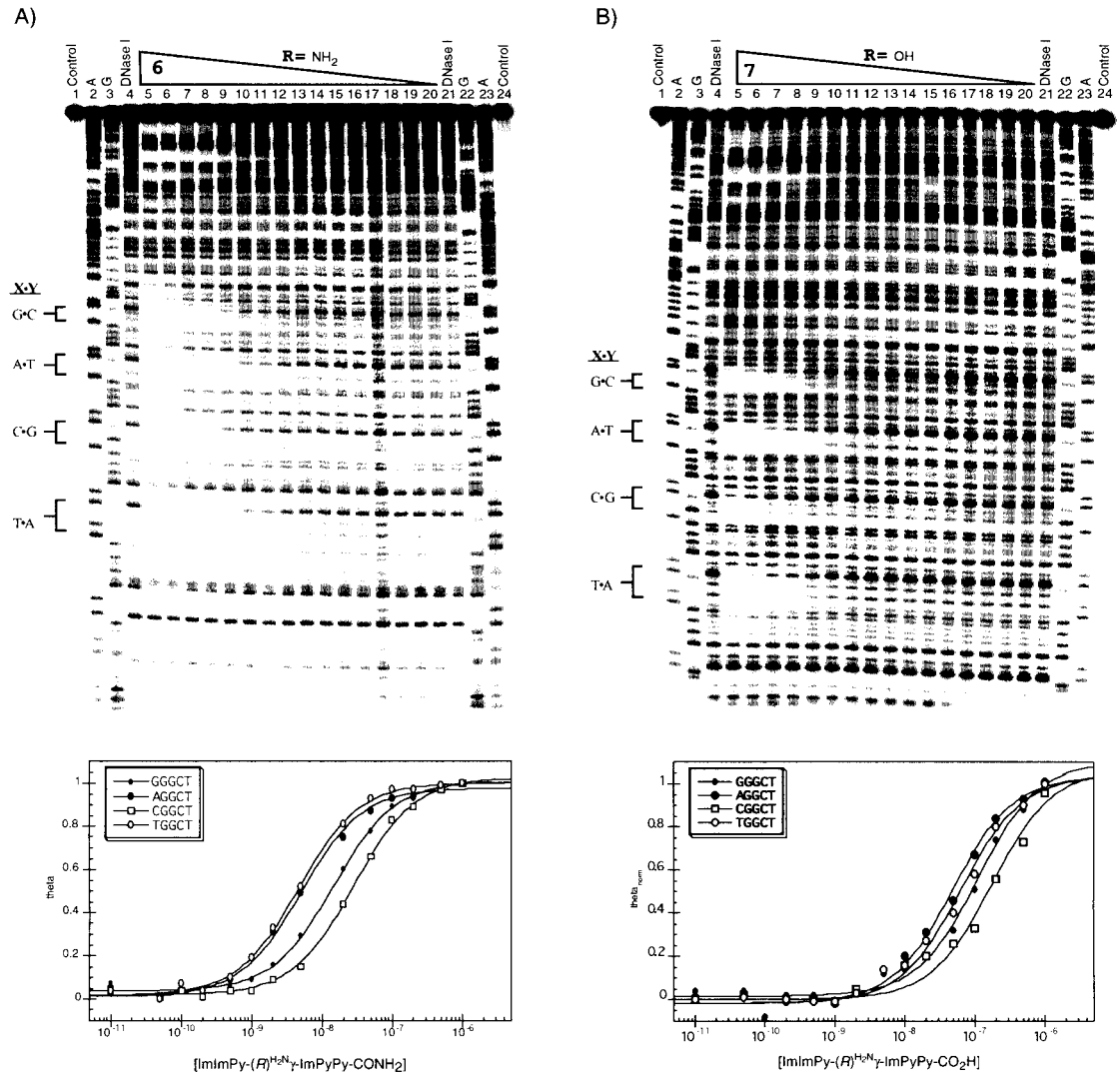
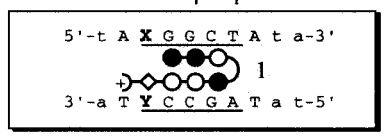


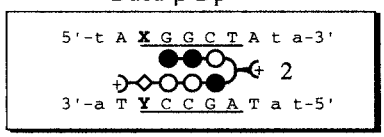
Figure 3.9. A) Quantitative DNase I footprint titration experiment with ImImPy-$(R)^{H_2N}$γ-ImPyPyCONH$_2$ (**6**) on the 3'-^{32}P-labeled 286-bp restriction fragment pSES-TL1[4]: lane 1 & 24, intact DNA; lane 2 & 23, A reaction; lane 3 & 22, G reaction; lane 4 & 21 DNase I standard; lanes 5-20 DNase I digestion products in the presence of 1 μM, 500 nM, 200 nM, 100 nM, 50 nM, 20 nM, 10 nM, 5 nM, 2 nM, 1 nM, 500 pM, 200 pM, 100 pM, 50 pM, 20 pM, 10 pM polyamide respectively. Quantitative DNase I footprint titration experiment with ImImPy-$(R)^{H_2N}$γ-ImPyPyCO$_2$H (**7**) on the 3'-^{32}P-labeled 286-bp restriction fragment pSES-TL1[4]: lane 1 & 24, intact DNA; lane 2 & 23, A reaction; lane 3 & 22, G reaction; lane 4 & 21 DNase I standard; lanes 5-20 DNase I digestion products in the presence of 1 μM, 500 nM, 200 nM, 100 nM, 50 nM, 20 nM, 10 nM, 5 nM, 2 nM, 1 nM, 500 pM, 200 pM, 100 pM, 50 pM, 20 pM, 10 pM polyamide respectively. Conditions and data analysis as outlined in Figure 3.8.

Gaba-β-Dp



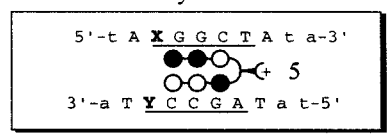
X·Y	Ka's	Specificity
A·T	3.1 X 10^7(±1.1)	> 310
T·A	2.1 X 10^7(±0.8)	> 210
G·C	< 10^5	1
C·G	< 10^5	1

Daba-β-Dp



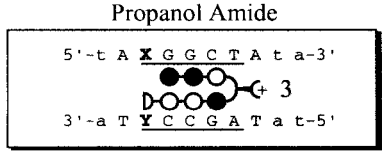
X·Y	Ka's	Specificity
A·T	1.7 X 10^9(±0.5)	53
T·A	1.6 X 10^9(±0.2)	50
G·C	5.8 X 10^7(±1.4)	2
C·G	3.2 X 10^7(±0.3)	1

Methyl Amide



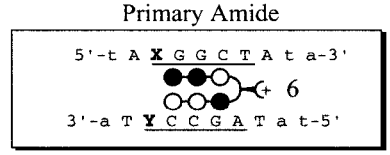
X·Y	Ka's	Specificity
A·T	1.1 X 10^9(±0.1)	10
T·A	1.4 X 10^9(±0.3)	12
G·C	2.4 X 10^8(±0.1)	2
C·G	1.1 X 10^8(±0.2)	1

Propanol Amide



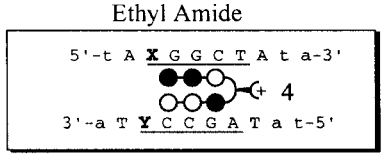
X·Y	Ka's	Specificity
A·T	5.9 X 10^8(±0.4)	140
T·A	4.7 X 10^8(±0.4)	112
G·C	2.2 X 10^7(±0.3)	5
C·G	4.2 X 10^6(±0.7)	1

Primary Amide



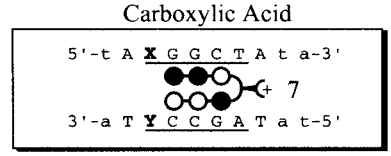
X·Y	Ka's	Specificity
A·T	1.9 X 10^8(±0.2)	5
T·A	2.1 X 10^8(±0.2)	6
G·C	6.8 X 10^7(±0.5)	2
C·G	3.5 X 10^7(±0.2)	1

Ethyl Amide



X·Y	Ka's	Specificity
A·T	1.1 X 10^9(±0.1)	35
T·A	9.2 X 10^8(±0.3)	30
G·C	8.2 X 10^7(±1.3)	2
C·G	3.0 X 10^7(±0.5)	1

Carboxylic Acid



X·Y	Ka's	Specificity
A·T	2.1 X 10^7(±0.1)	4
T·A	1.2 X 10^7(±0.2)	2
G·C	1.1 X 10^7(±0.2)	2
C·G	4.9 X 10^6(±0.2)	1

Figure 3.10. Polyamide **1-7** at the tail-1 position. Equilibrium association constants (M^{-1}) for polyamide **1-7** with restriction fragment pSES-TL1.[4] Values reported are the mean values from at least three DNase I footprinting titration experiments with the standard deviation for each data set in parentheses. Assays were performed at 22 °C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. Specificity is calculated as K_a(X•Y)/K_a(C•G).

affinity than expected, an affinity that is in fact comparable to the parent β-Dp polyamide (**1**).

Polyamides **4** and **5**, with the ethyl and methyl amide tails, provided a way to investigate the molecular recognition of DNA by the tail portion of polyamides more closely. At the outset of this study, it was possible that specificity could be assigned to a particular methylene group of the β-Dp and propanol amide tails. The first methylene group should be bulky enough to disfavor binding to G,C sites. Alternatively, the specific hydrogen bond made by the amide of the C-terminal pyrrole (Figure 3.1) would point the first methylene group away from the floor of the minor groove, such that specificity might be attributable to the second methylene group. In actuality, while polyamides **4** and **5** both show A,T specificity, compared to **6** and **7**, they are less specific than **1**, **2** and **3**. While both **4** and **5** bind with similar affinity for the A,T sites, the ethyl amide tail disfavors both G•C and C•G sites by about 5-fold more than the methyl amide tail. Both alkyl groups appear to contribute to A,T specificity, and even greater steric bulk beyond the second carbon is likely necessary for the specificity observed for the β-Dp and propanol amide tails. However, given the greater affinity of the shorter alkyl amides relative to **1** and **2**, it is likely that the increased steric bulk, while leading to greater discrimination against G,C sites, would also lead to lower affinity for the match sites. The unexpectedly high affinity of polyamides **4** and **5** for their match sites would suggest an energetic balance between increased Van der Waals contacts relative to the primary amide/acid tails and decreased steric clashing relative to the β-Dp tail.

Implications for polyamide design. A series of polyamides (**4-7**) with novel truncated C-terminal tails have been synthesized using oxime resin and Boc chemistry.

This solid phase methodology is an extremely versatile and efficient route to polyamides of various structures. Polyamide **6** was synthesized by three different routes, employing newly developed Fmoc monomers and a novel hydroxamic ester linker, in addition to the versatile oxime route. The ethyl (**4**) and methyl (**5**) amide tails are both A,T specific but exhibit higher affinity than either the propanol amide or β-Dp tails. Their subnanomolar affinity, usually seen in eight-ring rather than six-ring hairpin polyamides, will enable them to compete with proteins for DNA binding. The low molecular weights (below 900) of the methyl/ethyl amide tail polyamides coupled with their high affinity may influence their utility in biological systems. The primary amide and acid tail polyamides are 85% the molecular weight of the corresponding β-Dp polyamide, and both exhibit generality at the tail-1 position. While G,C sites are strongly disfavored by β-Dp and propanol amide tail polyamides (by >310- and 140-fold, respectively), the acid and primary amide tails have only a 4- and 6-fold range, respectively, between all four base pairings. Increased generality is attributed to removal of methylene groups which likely clash with the exocyclic amine of guanine. While the acid tail is slightly more general, the C-terminal pyrrole primary amide confers 10-fold higher affinity allowing the targeting of all four base pairs with similar affinity to the propanol tail for A,T sites, and provides increased affinity as well as generality relative to the β-Dp tail. The absence of a tertiary amine, coupled with the low molecular weights of these new truncated tail polyamides, may influence their cellular uptake and localization. Increased generality at the tail position, without loss of affinity, will enable targeting of new biologically relevant sequences by hairpin polyamides.

Experimental

Materials. Polyamides **1**, **2** and **3**, and the restriction fragment pSES-TL1 have been previously described.[4] DNase I footprinting titrations on the ^{32}P-labeled restriction fragment pSES-TL1 were performed as previously described.[4] Carboxymethoxylamine

hemihydrochloride, 2.0 M methylamine in THF, 2.0 M ethylamine in THF, DBU and anhydrous ammonia gas was purchased from Aldrich. Fmoc-Rink Amide resin and Oxime resin were purchased from Nova Biochem. All other synthetic reagents were as previously described.[2,17,19] ^{1}H NMR were recorded on a Varian Mercury 300 instrument. UV spectra were measured on a Beckman Coulter DU 7400 diode array spectrophotometer. Autoradiography was performed with a Molecular Dynamics Tyhoon Phosphorimager. Matrix-assisted, laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry was carried out at the Peptide and Protein Microanalytical Facility at the California Institute of Technology. High resolution electrospray ionization mass spectrometry was carried out at the University of California, Los Angeles Mass Spectrometry Facility. HPLC analysis was performed on a Beckman Gold system using a RAINEN C_{18}, Microsorb MV, 5 μm, 300 x 4.6 mm reversed-phase column in 0.1% (w/v) TFA with CH_3CN as eluent and a flow rate of 1.0 ml/min, gradient elution 1.25% CH_3CN/min. Preparatory HPLC was carried out on a Beckman HPLC using a Waters DeltaPak 25 x 100 mm, 100 μm C_{18} column, 0.1% (w/v) TFA, 0.56% CH_3CN/min. 18MΩ water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μm filtered. Reagent-grade chemicals were used unless otherwise stated.

Boc-Pyrrole-Hydroxamic Ester (8). 1,2,3-Benzotriazol-1-yl 4-[(*tert*-Butoxycarbonyl)amino]-1-methylpyrrole-2-carboxylate[2] (3.56 g, 10 mmol) and carboxymethoxylamine hemihydrochloride (1.09 g, 10 mmol) were dissolved in DMF (50 ml). DIEA (7 ml, 40 mmol) was added and the mixture was stirred at room temperature for 10 h. The solution was concentrated *in vacuo* and the resulting oil was partitioned between 250 ml saturated $NaHCO_3$ and 250 mL ether. The aqueous layer was acidified to pH 3 with concentrated HCl, and extracted with EtOAc (2 x 600 ml). The combined ethyl acetate extracts were concentrated *in vacuo* to 200 ml and washed with 0.1 N HCl (2 x 400 ml). The organic layer was dried over Na_2SO_4, and concentrated *in*

vacuo to yield **8** as a yellow oil (2.64 g, 8.43 mmol, 84% yield): ^{1}H NMR (DMSO-d_6) δ 11.40 (s, 1H), 9.06 (s, 1H), 6.93 (s, 1H), 6.57 (s, 1H), 4.40 (s, 2H), 3.71 (s, 3H), 1.41 (s, 9H); EIMS m/e 313.1278 (313.1274 calc. for $C_{13}H_{19}N_3O_6$).

ImImPy-(*R*)^{H_2N}γ-ImPyPyCONH$_2$ (6) (Boc/Hydroxamic Ester Route). Boc-β-Ala-PAM resin (1 g, 0.25 mmol/g) was deprotected with 80% TFA/DCM and 0.4 M PhSH. Boc-pyrrole-hydroxamic ester **8** (313 mg, 1 mmol) was activated by treatment with DCC (206 mg, 1 mmol) and *N*-hydroxysuccinimide (116 mg, 1 mmol) in 2 ml DMF for 1 h. The solution was filtered and added to the resin, 1 ml of DIEA was added and the mixture was shaken at room temperature for 6 h to yield Boc-Py-CONHOCH$_2$CO-β-Ala-PAM resin, which was elaborated to ImImPy-(*R*)^{FmocHN}γ-ImPyPyCONHOCH$_2$CO-β-Ala-PAM resin in a stepwise fashion by manual solid phase synthesis.[2,19] A sample of the polyamide resin (155 mg) was treated with 160 mg Pd(OAc)$_2$ in 2 ml DMF. The mixture was shaken at 37 °C for 2 h, 600 mg of ammonium formate in 900 μl H_2O was added, and shaking at 37 °C continued for 7 h. The reaction mixture was filtered and diluted to 8 ml with 0.1% (w/v) TFA, and the crude polyamide was purified by reverse-phase HPLC as previously described.[2] ImImPy-(*R*)^{H_2N}γ-ImPyPyCONH$_2$ (302 μg, 360 nmol, 1.2% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder; UV (H_2O) λ_{max} 310 (52140); ^{1}H NMR (DMSO-d_6) δ 10.65 (s, 1H), 10.32 (s, 1H), 10.20 (s, 1H), 9.90 (s, 1H), 9.75 (s, 1H), 8.09 (bs, 1H), 7.55 (s, 3H), 7.47 (s, 1H), 7.44 (s, 1H), 7.25 (d, 1H, J = 2.1 Hz), 7.21 (d, 1H, J = 2.1 Hz), 7.19 (bs 2H), 7.11 (d, 1H, J = 1.8 Hz), 7.05 (d, 1H, J = 1.2 Hz), 6.98 (d, 1H, J = 1.2 Hz), 6.83 (d, 1H, J = 1.5 Hz), 6.65 (d, 1H, J = 1.2 Hz), 6.53 (d, 1H, J = 2.4 Hz), 3.99 (s, 3H), 3.98 (s, 3H), 3.96 (s, 3H), 3.82 (s, 3H), 3.79 (s, 1H), 3.77 (s, 3H), 2.97 (m, 1H), 1.96 (m, 2H), 1.43 (m, 2H), MALDI-TOF-MS 838.4 (838.35 calc for M+H).

ImImPy-(*R*)^{H_2N}γ-ImPyPyCONH$_2$ (6) (Fmoc/Rink Resin Route). ImImPy-(*R*)^{BocHN}γ-ImPyPyCONH-Rink resin was generated by manual solid phase synthesis from Fmoc-Rink resin (0.6 mmol/g) using Fmoc-protected monomers as previously

described.[17] A sample of the polyamide resin (30 mg) was cleaved by treatment of 20% TFA in DCM at room temperature for 2 h. After filtration the cleavage solution was concentrated *in vacuo* and purified by reverse phase HPLC as previously described.[2] ImImPy-$(R)^{H_2N}\gamma$-ImPyPyCONH$_2$ (630 μg, 750 nmol, 4.2% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder. Spectroscopic analysis confirmed this product was identical to **6** produced by the Boc/Hydroxamic Ester route.

ImImPy-$(R)^{H_2N}\gamma$-ImPyPyCONH$_2$ (6) (Boc/Oxime Resin Route). ImImPy-$(R)^{FmocHN}\gamma$-ImPyPyCO-Oxime resin (**17**) was generated by manual solid phase synthesis from oxime resin (1 g, 0.48 mmol/g) using previously described Boc-protected monomers.[2] Boc-Py-OBt (**9**) (358 mg, 1 mmol) was dissolved in 2 ml of NMP and added to 1 g of oxime resin followed by 1 ml of DIEA. The coupling was allowed to proceed overnight at room temperature. The resin was then acetylated with 3 ml of acetic anhydride (Ac_2O), 4 ml of NMP and 1 ml of DIEA for 30 minutes at room temperature. The Boc group was removed upon treatment with 20% TFA/DCM for 30 minutes. The second pyrrole residue was coupled in the same fashion as the first but complete coupling could be achieved in 2 h at room temperature followed by the acetylation and deprotection steps outlined above. Boc-Im-OH (**10**) (241 mg, 1 mmol) was dissolved in 2 ml of NMP to which 360 mg (1 mmol) of HBTU and 1 ml of DIEA was added for activation of this monomer. Coupling was allowed to proceed for 1.5 h at room temperature followed by acetylation. The Boc-Im residue was deprotected using a 50% TFA/DCM solution for 30 minutes at room temperature. The next residue, α-Fmoc-γ-Boc-(*R*)-diaminobutyric acid (DABA) (**11**) (660 mg, 1.5 mmol) was activated with HBTU (540 mg, 1.5 mmol) in 2 ml of NMP and 1 ml of DIEA. Coupling of this residue onto the resin took 2 h at 37 °C. After acetylation and treatment with 20% TFA/DCM for 30 minutes, the next Boc-Py-OBt was attached in exactly the same manner as the second residue. For ease of synthesis, the terminal two imidazoles were added as an ImIm-OH dimer.[2] 249 mg (1 mmol) of the ImIm-OH dimer (**12**) was activated with HBTU (360

mg, 1 mmol) in 2 ml of NMP and 1 ml of DIEA and allowed to couple overnight at room temperature. It should be noted that the progress of the stepwise couplings were all monitored by analytical HPLC. The resin was washed thoroughly with DMF, DCM, MeOH and Et_2O then dried *in vacuo*. A small sample of the dried resin (≈ 20 mg) was placed into a pressure tolerant screw cap test tube, suspended in 4 ml of dry THF and cooled for 20 minutes in a -10 °C ice/brine bath. Anhydrous NH_3 was bubble into the suspension at a steady rate for 30 minutes to generate a saturated ammonia solution. 1.5 ml of a 10%v/v solution of DBU in CH_2Cl_2, cooled to -20 °C was quickly added to the cleavage mixture and the tube immediately sealed. The cleavage was allowed to proceed in a shaker at 37 °C for 72 h. The mixture is filtered and concentrated *in vacuo*, then purified by reverse phase HPLC. The recovery upon lyophilization of the appropriate fractions was comparable to the Fmoc/Rink resin route. Spectroscopic analysis confirmed this product was identical to polyamide **3** produced in both of the previously outlined routes.

ImImPy-(*R*)^{H_2N}γ-ImPyPyCONHEt (4). A 75 mg sample of dried resin **17** was suspended in 2 ml of CH_2Cl_2 to which was added 2 ml of 2.0 M ethylamine in THF. This cleavage mixture was placed in a 37 °C oven and allowed to stand overnight in a sealed scintillation vial. The resin was filtered, the eluant concentrated *in vacuo*, then purified by reverse phase HPLC. ImImPy-(*R*)^{H_2N}γ-ImPyPyCONHCH$_3$ (1.9 mg, 2.2 μmol, 10.6% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder; UV (H_2O) λ_{max} 310 (52140); ^{1}H NMR (DMSO-d_6) δ 11.02 (s, 1H), 10.36 (s, 1H), 10.10 (s, 1H), 9.89 (s, 1H), 9.71 (s, 1H), 8.20 (m, 1H), 7.98 (m, 1H), 7.56 (s, 1H), 7.52 (s, 1H), 7.45 (d, 1H, J = 0.9 Hz), 7.25 (d, 1H, J = 1.5 Hz), 7.23 (d, 1H, J = 1.5 Hz), 7.15 (d, 1H, J = 1.8 Hz), 7.11 (d, 1H, J = 1.8 Hz), 7.06 (d, 1H, J = 0.9 Hz), 7.03 (d, 1H, J = 1.8 Hz), 6.83 (d, 1H, J = 1.8 Hz), 3.99 (s, 3H), 3.98 (s, 3H), 3.96 (s, 3H), 3.83 (s, 3H), 3.80 (s, 3H), 3.78 (s, 3H), 3.17 (dd, 2H, J = 6.6 Hz), 2.42 (m, 2H), 2.26 (m, 1H), 1.99 (m, 2H),

1.62 (m, 1H), 1.05 (t, 3H J = 6.6 Hz), MALDI-TOF-MS 866.5 (866.39 calc for [M+H] $C_{39}H_{48}N_{17}O_7^+$).

ImImPy-(R)^{H_2N}γ-ImPyPyCONHMe (5). A 35 mg sample of dried resin **17** was suspended in 2 ml of CH_2Cl_2 to which was added 2 ml of 2.0 M methylamine in THF. This cleavage mixture was placed in a 37 °C oven and allowed to stand overnight in a sealed scintillation vial. The resin was filtered, the eluant concentrated *in vacuo*, then purified by reverse phase HPLC. ImImPy-(R)^{H_2N}γ-ImPyPyCONHCH$_3$ (850 μg, 996 nmol, 8.9% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder; UV (H_2O) λ_{max} 310 (52140); ^{1}H NMR (DMSO-d_6) δ 11.01 (s, 1H), 10.35 (s, 1H), 10.09 (s, 1H), 9.90 (s, 1H), 9.70 (s, 1H), 8.19 (m, 1H), 7.92(m, 1H), 7.56 (s, 1H), 7.51 (s, 1H), 7.45 (d, 1H, J = 0.9 Hz), 7.25 (d, 1H, J = 1.8 Hz), 7.23 (d, 1H, J = 1.8 Hz), 7.16 (d, 1H, J = 1.5 Hz), 7.10 (d, 1H, J = 1.5 Hz), 7.05 (d, 1H, J = 0.9 Hz), 7.03 (d, 1H, J = 1.8 Hz), 6.79 (d, 1H, J = 1.8 Hz), 3.99 (s, 3H), 3.98 (s, 3H), 3.96 (s, 3H), 3.82 (s, 3H), 3.79 (s, 3H), 3.77 (s, 3H), 2.66 (m, 1H), 2.41 (m, 2H), 2.25 (m, 1H), 1.98 (m, 2H), 1.62 (m, 2H), MALDI-TOF-MS 852.5 (852.38calc for [M+H] $C_{38}H_{46}N_{17}O_7^+$).

ImImPy-(R)^{H_2N}γ-ImPyPyCO$_2$H (7). A 35 mg sample of dried resin **17** was cleaved by treatment with 1 ml of H2O, 1 ml DMF, and 75 μl DBU (0.48 mmol) at 37 °C for 60 h. After filtration the cleavage solution was concentrated in vacuo and purified by reverse phase HPLC. ImImPy-(R)^{H_2N}γ-ImPyPyCO$_2$H (294 μg, 350 nmol, 3.1% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder; UV (H_2O) δ_{max} 310 (52140); ^{1}H NMR (DMSO-d_6) δ 12.67 (bs, 1H), 11.53 (s, 1H), 10.88 (s, 1H), 10.62 (s, 1H), 10.43 (s, 1H), 10.22 (s, 1H), 8.83 (m, 1H), 8.07 (s, 1H), 8.03 (s, 1H), 7.96 (d, 1H, J = 0.9 Hz), 7.93 (d, 1H, J = 1.8 Hz), 7.76 (d, 1H, J = 0.9 Hz), 7.75 (d, 1H, J = 1.8 Hz), 7.64 (d, 1H, J = 1.2 Hz), 7.57 (d, 1H, J = 1.2 Hz), 7.54 (d, 1H, J = 1.8 Hz), 7.34 (d, 1H, J = 1.8 Hz), 4.50 (s, 3H), 4.50 (s, 3H), 4.47 (s, 3H), 4.34 (s, 3H), 4.31 (s, 3H), 4.31 (s, 3H), 2.50 (m, 2H), 2.11 (m, 2H), 2.06 (m, 3H), MALDI-TOF-MS 839.4 (839.34calc for M+H).

Acknowledgements

We thank Dr. Tom Minehan, Dr. Anna Mapp, and Dr. Paul Foreancig for helpful advice and discussions concerning the hydroxamic ester portion of this work. We are grateful to the National Institutes of Health GM27681 for research support, the Ralph M. Parsons Foundation for a predoctoral fellowship to J.M.B., the Natural Sciences and Engineering Research Council of Canada for a postgraduate scholarship to D.H.N., and Bristol-Myers Squibb for a predoctoral fellowship to N.R.W.

References

1. Dervan, P. B.; Bürli, R. W. *Current Opinion in Chemical Biology* **1999**, *3*, 688-693.
2. Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6141-6146.
3. Parks, M. E.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6147-6152.
4. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1999**, *121*, 1113-1120.
5. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Angewendte Chemie-International Edition English* **1998**, *37*, 1421-1423.
6. Gottesfeld, J. M.; Turner, J. M.; Dervan, P. B. *Gene Expression* **2000**, *9*, 77-91.
7. Ioshikhes, I. P.; Zhang, M. Q. *Nature Genetics* **2000**, *26*, 61-63.
8. Pabo, C. O.; Sauer, R. T. *Annual Reviews in Biochemistry* **1992**, *61*, 1053-1095.
9. Peter, C.; Daura, X.; van Gunsteren, W. F. *Journal of the American Chemical Society* **2000**, *122*, 7461-7466.
10. Shin, I.; Lee, M.; Lee, J.; Jung, M.; Lee, W.; Yoon, J. *Journal of Organic Chemistry* **2000**, *65*, 7667-7675.
11. Wu, Y.-D.; Wang, D.-P.; Chan, K. W. K.; Yang, D. *Journal of the American Chemical Society* **1999**, *121*, 11189-11196.
12. Yang, D.; Ng, F.-F.; Li, Z.-J.; Wu, Y. D.; Chan, K. W. K.; Wang, D.-P. *Journal of the American Chemical Society* **1996**, *118*, 9794-9795.

13. Yang, D.; Qu, J.; Li, B.; Ng, F.-F.; Wang, X.-C.; Cheung, K.-K.; Wang, D.-P.; Wu, Y.-D. *Journal of the American Chemical Society* **1999**, *121*, 589-590.

14. Belitsky, J. M.; Floreancig, P. F.; Dervan, P. B. Unpublished Results.

15. Meloni, M.; Taddei, M. *Organic Letters* **2001**, *3*, 337-340.

16. Keck, G. E.; McHardy, S. F.; Wager, T. T. *Tetrahedron Letters* **1995**, *36*, 7419-7422.

17. Wurtz, N. R.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Organic Letters* **2001**, *3*, 1201-1203.

18. Bark, S. J.; Schmid, S.; Hahn, K. M. *Journal of the American Chemical Society* **2000**, *122*, 3567-3573.

19. Herman, D. M.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1998**, *120*, 1382-1391.

20. This unexpected reactivity is under investigation.

21. Canne, L. E.; Bark, S. J.; Kent, S. B. H. *Journal of the American Chemical Society* **1996**, *118*, 5891-5896.

22. Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1999**, *121*, 1121-1129.

23. Carpino, L. A. *Accounts of Chemical Research* **1987**, *20*, 401-407.

24. DeGrado, W. F., Kaiser, E.T. *Journal of Organic Chemistry* **1980**, *45*, 1295-1300.

25. DeGrado, W. F., Kaiser, E.T. *Journal of Organic Chemistry* **1982**, *47*, 3258-3261.

26. Voyer, N., Lavoie, A., Pinette, M., Bernier, J. *Tetrahedron Letters* **1994**, *35*, 355-358.

27. Pichette, A., Voyer, N., Larouche, R., Meillon, J-C. *Tetrahedron Letters* **1997**, *38*, 1279-1282.

28. Nakagawa, S. H., Kaiser, E.T. *Journal of Organic Chemistry* **1983**, *48*, 678-685.

29. Swalley, S.E., Dervan, P. B. Unpublished Results.

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

31. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proceedings of the National Academy of Sciences USA* **1986**, *83*, 8462-8466.

32. Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344-7354.

33. Bremer, R. E., Wurtz, N.R., Szewczyk, J. W., Dervan, P.B. *Bioorganic & Medicinal Chemistry* **2001**, in press.

34. Cho, J.; Parks, M. E.; Dervan, P. B. *Proceedings of the National Academy of Sciences USA* **1995**, *92*, 10389-10392.

35. Bailly, C.; Chaires, J. B. *Bioconjugate Chemistry* **1998**, *9*, 513-538.

36. deClairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Journal of the American Chemical Society* **1997**, *119*, 7909-7916.

37. deClairac, R. P. L.; Seel, C. J.; Geierstanger, B. H.; Mrksich, M.; Baird, E. E.; Dervan, P. B.; Wemmer, D. E. *Journal of the American Chemical Society* **1999**, *121*, 2956-2964.

38. Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C. *Nature Structural Bioogy.* **1998**, *5*, 104-109.

39. Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Journal of Molecular Biology* **2000**, *295*, 557-567.

40. Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P.B.; Rees, D. C. *Science* **1998**, *282*, 111-115.

Chapter Four

Towards a Minimal Motif for Artificial Transcriptional Activators

The text of this chapter was taken in part from a publication co-authored with Aseem Z. Ansari, Anna K. Mapp, Professor Peter B. Dervan and Professor Mark Ptashne.

Publication: Ansari, A.Z.; Mapp, A.K.; Nguyen, D.H.; Dervan, P.B.; Ptashne, M. *Cheimstry & Biology*, **2001**, *8*, 583-592.

Abstract

Most transcriptional activators minimally comprise two functional modules, one for DNA binding and the other for activation. In a previous study we generated an artificial transcriptional activator, 4.2 kDa in size, comprised of a DNA binding hairpin polyamide tethered to a 20 residue activating peptide (AH) capable of stimulating promoter specific transcription.[1] The question arises as to the general nature and the versatility of this minimal activator motif and whether smaller ligands can be designed which maintain potent activation function. Here we have first replaced the 20 amino acid AH peptide with 8 or 16 residues derived from the activation domain of the potent viral activator VP16 to give comparable or better levels of activation. Altering the site-of-attachment of the activation module on the polyamide allowed reduction of the intervening linker from 36 atoms to 8 without significant diminution of the activation potential. We also exchanged the polyamide to target a different DNA sequence without compromising the activation function further demonstrating the generality of this design. The polyamide-activator conjugates described here represent a class of DNA binding ligands which are tethered to a second functional moiety; an activation domain which recruits elements of the endogenous transcriptional machinery. Our results define the minimal structural elements required to construct artificial, small molecule activators. If such activators are cell-permeable and can be targeted to designated sites in the genome, they may then serve as a tool to study mechanistic aspects of transcriptional regulation and eventually to modulate gene expression relevant to human diseases.

Introduction

Positive regulation (activation) of gene expression requires factors called transcriptional activators. An economical 'recruitment' model posits that activator proteins bind to DNA and recruit the transcriptional machinery to a proximal promoter, thereby stimulating gene expression.[2] These steps define the initial regulatory decisions in a transcriptional circuit, and misregulation at any stage can result in a variety of human diseases. Given the critical role of activators in transcriptional regulation, the development of artificial counterparts that could be used to rectify errors in gene expression has been a long-standing goal at the interface of chemistry and biology. Activators achieve specificity in targeting genes by a DNA recognition module which binds to cognate DNA sequences near a promoter and in most cases binding specificity is further enhanced by dimerization.[1,2] The key functional module, the activating region, is thought to bind several components of the transcriptional machinery (2, 3 and references therein). Many of these components of the machinery exist in large multi-subunit complexes which associate with the RNA polymerase II, and are known as the RNA polymerase II holoenzyme.[4,5] The holoenzyme along with a few additional factors, like TBP, constitutes the transcriptional machinery that is recruited by activators to most promoters in vivo.[6] It is believed that weak interactions between an activating region and several components of the holoenzyme result in high avidity 'multi-dentate' binding by the activator to the holoenzyme. In addition, acidic activating regions (e.g., those used here) are believed to contact and recruit nucleosome modifying activities to promoters.[7,8]

Artificial transcriptional activators. Attempts to generate artificial activators have relied on this principle of recruitment. In one example, a dimer of the natural

product FK506 was used to couple a chimeric DNA binding protein to an activating region, thus up-regulating a gene bearing upstream cognate binding sites.[9-11] More recently attachment of an activating region to a triplex-forming oligonucleotide or designed Zn(II) finger motifs were shown to up-regulate gene expression.[12,13] Our approach towards creating a general synthetic motif for artificial activators that are capable of targeting a wide variety of DNA sequences relies upon a versatile non-protein DNA binding motif, namely the polyamides which are composed of heterocyclic residues.[14] Obeying a set of pairing rules, the polyamides bind to pre-determined DNA sequences in the minor groove with affinities and specificities comparable to typical DNA binding proteins.[14-16] Finally, polyamides are cell-permeable and can mediate their regulatory effects on gene expression in primary cells and even drosophila larvae.[17-19]

In our previous studies, we substituted key functional domains of naturally occurring activator proteins with synthetic counterparts to create an entirely synthetic transcriptional activator.[1] One focus of that study was to test the importance of dimerization of activators in their ability to activate transcription. In that study a flexible polyether linker (the linker domain) replaced the typical protein dimerization domain and it served to tether the designed activating region AH to a hairpin polyamide that binds the cognate sequence 5'-TGTTAT-3' (Figure 4.1). This created our smallest functional activator, with a molecular weight of 4.2 kDa.[1] Given that polyamides composed of pyrrole, imidazole, and hydroxypyrrole hetrocycles can be designed to target a wide variety of DNA sequences, this motif can potentially be used to generate compounds that regulate expression of designated genes.[14]

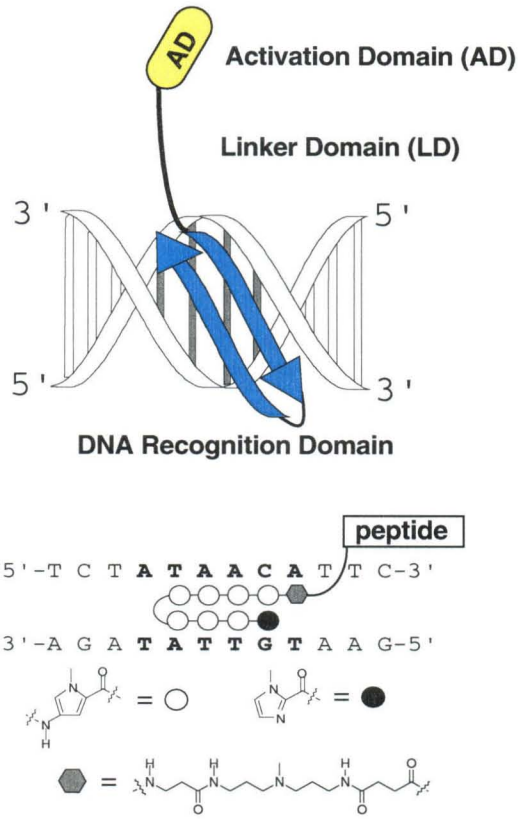


Figure 4.1. Design of a synthetic activator. The synthetic activator is composed of three functional domains. The activation domain (AD), a peptide, is tethered to the DNA binding region by a linker domain (LD). The DNA recognition domain is a hairpin polyamide composed of imidazole and pyrrole amino acids that selectively binds the sequence 5'-TGTTAT-3'.

A smaller, more potent artificial activator. Our original conjugate (PA-1L-AH, **2**) functioned demonstrably (15-20 fold over polyamide levels) in a cell-free system[1] and the next long-term goal was to use this motif for upregulation in cell culture (Figure 4.2). Some immediate concerns remained to be addressed prior to undertaking these experiments, however. Although our activator is much smaller than a typical protein activator such as Gal4 (~101 kDa versus 4.2 kDa), it was desirable to further decrease the molecular weight of the motif in order to increase the probability of membrane permeability.

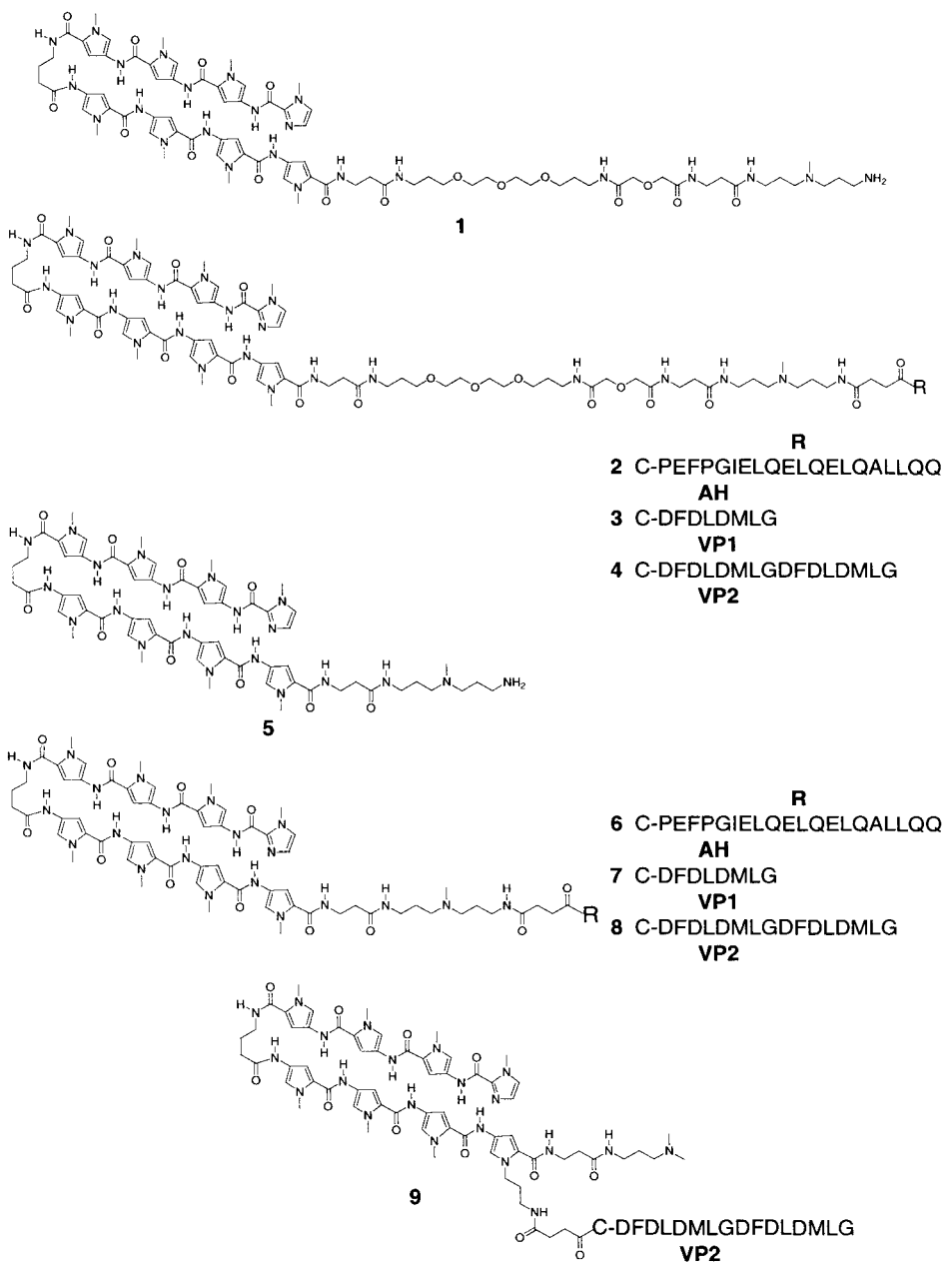


Figure 4.2. Structures of hairpin polyamides **1** and **5** and conjugates **2-4** and **6-9** prepared to explore the effect of activating region identity, linker domain motif, and site-of-attachment, on activation potential. All are designed to bind to the DNA sequence 5'-TGTTAT-3' and differ in the identity of the activating region and the linking domain. In conjugates **2-4**, the activating peptide is covalently linked to the hairpin polyamide by a polyether tether 36 atoms in length. Conjugates **6-8** have a shorter, 12 atom tether. The three different activating regions are AH,[29] a designed peptide used in our earlier study, and one (VP1) or two (VP2) tandem repeats of an eight amino acid sequence derived from the viral activator protein VP16.[23,24] For conjugate **9**, the activator peptide VP2 was attached to an internal pyrrole residue. The C at the junction between the peptide and the tether is cysteine, a key residue for the ligation of peptides with polyamide substrates.[34]

Ideally, this would be accomplished with no loss in function. This could be carried out by varying the identity of the activating region or by shortening the linker domain. In addition, we wished to characterize the binding of the conjugates to the cognate sites within the promoter region and correlate this with function. Finally it remained to be demonstrated that the identity of the DNA recognition module, the polyamide, could be changed without compromising function.

Here we describe artificial activators with tunable potency effected by the size and identity of the activating region as well as the site of attachment to the polyamide. Most notably, we demonstrate a potency of 50% to 150% greater than PA-1L-AH (**2**) with concomitant decreases in size of 21% and 12%, respectively (Table 4.1). We also report that all polyamide-peptide conjugates of this minimal motif bind to their cognate sites upstream of the AdML promoter. Consistent with our predictions, we find that the concentration of conjugate required for full promoter occupancy corresponds to the concentration required to elicit maximal activation *in vitro*. Finally, we show that changing the DNA recognition domain can be accomplished while retaining function.

Conjugate	Molecular Weight
PA-1L-AH (**2**)	4164
PA-1L-VP1 (**3**)	2764
PA-1L-VP2 (**4**)	3670
PA-VP1 (**7**)	2375
PA-VP2 (**8**)	3282
PA-(py)-VP2 (**9**)	3282

Table 4.1. Summary of conjugate size containing identical eight-ring polyamide but different activation domains and tether lengths.

Results and Discussion

Activation modules and their site-of-attachment. The activation domain (residues 411-490) of VP16, a viral protein, fused to a DNA binding protein yields a chimeric activator with the potency comparable to strong natural activators.[20] Dissection of this and other activation domains has identified minimal units that activate transcription; these modules are often surreptitiously reiterated in natural activating regions.[2,21,22] In VP16 one such minimal module consists of eight amino acids.[23,24] When reiterated the activation potential of the consequent peptide increased in a synergistic rather than an additive manner.[23,25] This was also reported for other activating modules.[26-28] We adopted this property of activating regions in order to design artificial activators of varying strengths. It appeared likely that decreasing the activating region size to a minimal eight amino acid module (VP1) would provide a small activator that would retain some level of function. In order to probe these questions, we designed a series of polyamide-peptide conjugates, each of which consists of a hairpin polyamide designed to target the cognate sequence 5'-TGTTAT-3', a flexible tether of varying length (12 or 36 atoms), and one of three different activating regions (Figure 4.2). The three activating regions are AH,[29] a designed peptide used in our earlier studies, and one (VP1) or two (VP2) tandem repeats of an eight amino acid sequence derived from VP16.[24,25]

Next, from our earlier studies we determined that a critical role of the linking or dimerization domain is to facilitate projection of the activating module fully away from the DNA for productive interaction with the transcriptional machinery.[1] To date this had been accommodated by conjugation through a polyether linker at the C-terminus of the hairpin polyamides, but conjugation via linkage at an internal pyrrole residue appeared an

attractive alternative. Solution studies as well as X-ray crystallographic data have demonstrated that the *N*-methyl group of the pyrrole residues is directed outward from the minor groove.[30-32] We hypothesized that extension to an 8 atom linker at this position as in conjugate **9** (Figure 4.2) should provide effective projection of any peptide conjugated at that position, perhaps eliminating the need for a longer tethering linker.

Polyamides **1** and **5** were prepared by solid-phase synthesis and subsequently elaborated to provide conjugates **2-4** and **6-9** in accordance with established procedures.[1,34] All polyamides and conjugates thus prepared were purified by reversed-phase HPLC and their identities were confirmed by matrix-assisted laser-desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. Conjugates **2**, **3**, and **4** incorporate a 36 atom polyether tether as the linker domain. Conjugates **6**, **7**, and **8** utilize a shorter 12 atom linker.

Promoter occupancy under transcriptional conditions. The dissociation constant (K_D) for conjugate **2** binding to its cognate site (5'-TGTTAT-3') has been measured as 32 nM using quantitative DNase I footprinting titrations.[1] However, almost 10-fold higher concentrations of this conjugate were required to achieve maximal transcriptional activation *in vitro* (Figure 4.3). The conditions for these experiments are substantially different than those of the quantitative footprinting assays. Perhaps the most pertinent difference is the concentration of DNA. In quantitative DNase I footprinting experiments, a greater than 10-fold excess of polyamide relative to DNA is necessary for accurate determination of K_D for the ligand - DNA complex.[35]

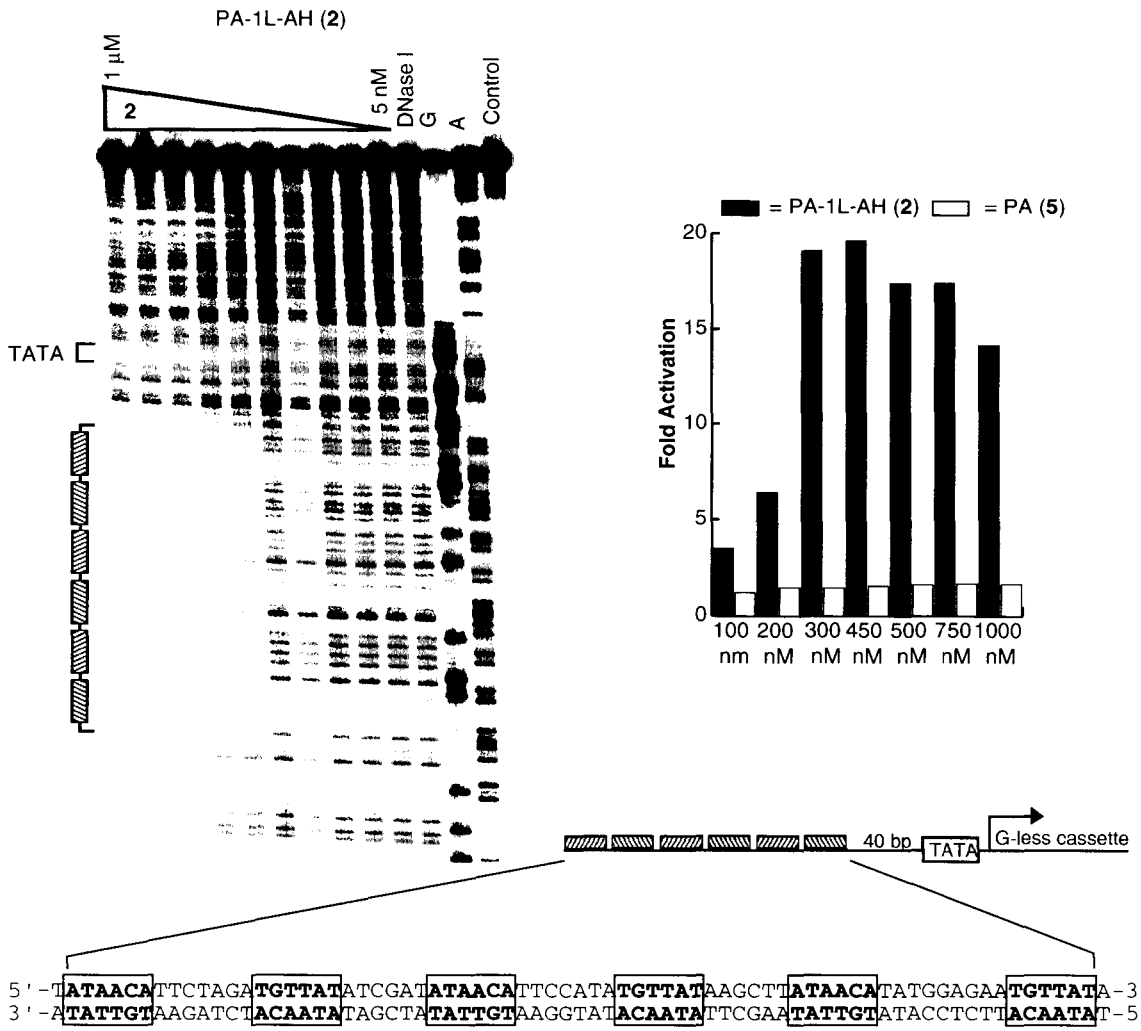


Figure 4.3. Promoter occupancy corresponds to activator function. (Upper left) Storage phosphor autoradiogram of a DNase I footprinting titration of **2** (PA-1L-AH) carried out under transcription conditions on a 5'-^{32}P-labeled 363 bp DNA fragment containing both the promoter region and 140 bp of the G-less cassette reporter. To each footprinting reaction, unlabeled plasmid DNA was added to bring the total concentration of polyamide binding sites to a level equivalent to those utilized in transcription assays. From left to right the lanes are DNase I digestion products in the presence of **2** at concentrations of 1 M, 750 nM, 500 nM, 300 nM, 200 nM, 100 nM, 50 nM, 20 nM, 10 nM, and 5 nM respectively; DNase I digestion products with no **2** present; the G sequencing lane; the A sequencing lane; undigested DNA. (Upper right) Titration of conjugate **2** in transcription assays shows that maximal activation corresponds to maximal occupancy of the binding sites. Relative fold activation was determined by comparing the levels of transcript produced by increasing concentrations of **2** or **5** with that obtained from transcription reactions performed in the absence of any compound (basal levels). Increasing concentrations of **5** did not have any effect on transcription whereas conjugate **2** shows maximal activation at 300-450 nM. (Lower) Design of the plasmid DNA template used for *in vitro* transcription experiments and the DNase I footprinting titration experiments. The promoter region contains six cognate binding sites for the hairpin polyamide upstream of a G-less cassette reporter.

In vitro transcription assays, however, employ a higher concentration of plasmid DNA (0.8 ng/μL), and this has been demonstrated in other cases to decrease the occupancy of a binding site by 10- to 100-fold.[36] To investigate the binding behavior of **2** under such conditions, a 5'-^{32}P-labeled 363 bp DNA fragment containing both the promoter region and 140 bp of the G-less cassette reporter was generated. To each footprinting reaction, unlabeled plasmid DNA was added to bring the total concentration of polyamide binding sites to a level equivalent to those utilized in transcription assays. As anticipated, DNase I footprinting titrations performed under the conditions of the *in vitro* transcription assays revealed that 50% occupancy of the three dimeric binding sites occurs at a concentration of 215 nM, approximately a 7-fold increase over the measured K_D (Figure 4.3).

Promoter occupancy and the level of activation. An *in vitro* transcription titration experiment with conjugate **2** demonstrates that full occupancy of the promoter is necessary for maximal activator function (Figure 4.3). At a concentration of 100 nM, detectable levels (>4-fold) of transcriptional activation are observed, and as the concentration increases to provide full saturation of the dimeric sites, activation levels climb to almost 20-fold over basal transcription. Transcription levels were not influenced by a similar increase in the concentrations of **5** which does not bear an activation module. Taken together, the data suggest that optimal conjugate concentration for *in vitro* transcription assays can be predicted using data from DNase I footprinting titrations under such conditions. Therefore, additional DNase I footprinting titration experiments were carried out, revealing that the polyamide-peptide conjugates require concentrations of 75 to 215 nM to attain 50% occupancy (Table 4.2). Consistent with previous studies we

observe that tethering a peptide to the C-terminus of the hairpin polyamide leads to a 5- to 14-fold decrease in overall binding affinity. We had anticipated that attachment of a peptide through an internal pyrrole residue, as in conjugate **9**, would have a lesser impact upon DNA binding. However, the concentration of conjugate **9** required for 50% occupancy, 175 nM, is similar to that of its C-terminus-linked counterparts. Based on these data, all subsequent *in vitro* transcription experiments were carried out at conjugate concentrations of 300-400 nM.

Conjugate	50% Occupancy (nM)
PA-1L-AH (**2**)	215 nM
PA-1L-VP1 (**3**)	75 nM
PA-1L-VP2 (**4**)	210 nM
PA(**5**)	15 nM
PA-VP1 (**7**)	110 nM
PA-VP2 (**8**)	75 nM
PA-(py)-VP2 (**9**)	175 nM

Table 4.2. Summary of data from DNase I footprinting titration experiments for conjugates **2-4** and **7-9** and polyamide **5**. The reported 50% occupancies are the average values obtained from a minimum of two DNase I footprint titration experiments. 50% occupancies were determined by best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm (n=1) as previously described.[35] All experiments were carried out under transcription conditions as described in the methods section and depicted in Figure 4.3.

Activating potential corresponds to composition and the site-of-attachment of the activation module. The *in vitro* transcription experiments of Figures 4.4 and 4.5 reveal that the strength of the activator-polyamide conjugates is proportional to the size and composition of the activating region, and that projection of the activating region away from the DNA enhances its functional potency. Results from five independent sets of reactions summarized in Figure 4.4 show that when compared with conjugate **2** (PA-1L-AH), we find that conjugate **7** (PA-VP1) is a weaker activator and that substituting VP1

with VP2 (conjugate **8**) increases the activation strength of the resulting conjugate. While the composition of the activating region plays a role in determining transcriptional potency, it was surprising that reiteration of the VP1 module in this context only improved transcriptional potency by two-fold. The absence of appreciable synergy in transcriptional activation *in vitro* suggests that the synergy observed when a similar motif was reiterated in vivo may arise from the relative accessibility of the two VP1 modules when presented in the context of a chimeric fusion protein[23] or that there is an activation potency threshold in vivo below which activators do not function efficiently. Another possible explanation is the differences in the secondary structure of the 8 versus the 16 residue peptide. However, based on studies demonstrating resilience of the activating regions to structural perturbations,[10,37] it is more likely that difference in activation potential of these peptides stems from the fact that there is simply insufficient surface area available for high-avidity interactions with the targets in the transcriptional machinery.

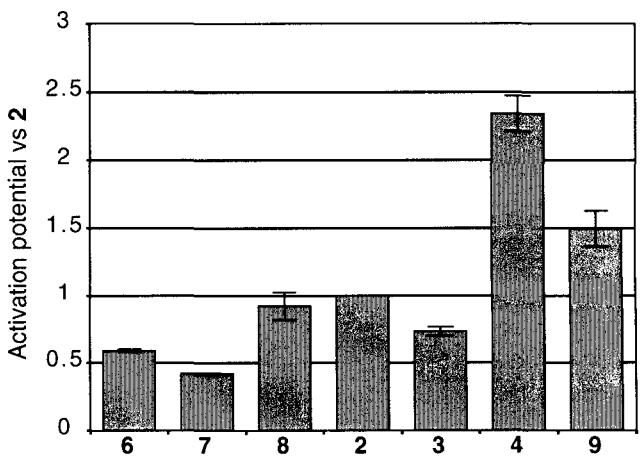


Figure 4.4. Summary of five independent *in vitro* transcription reactions showing the relative potency of each compound in comparison to PA-1L-AH. The absolute fold-activation mediated by all conjugates varies slightly between different experiments. To effectively compare the activation potential of the various conjugates, we normalized data from five separate experiments for each conjugate to that of **2**.

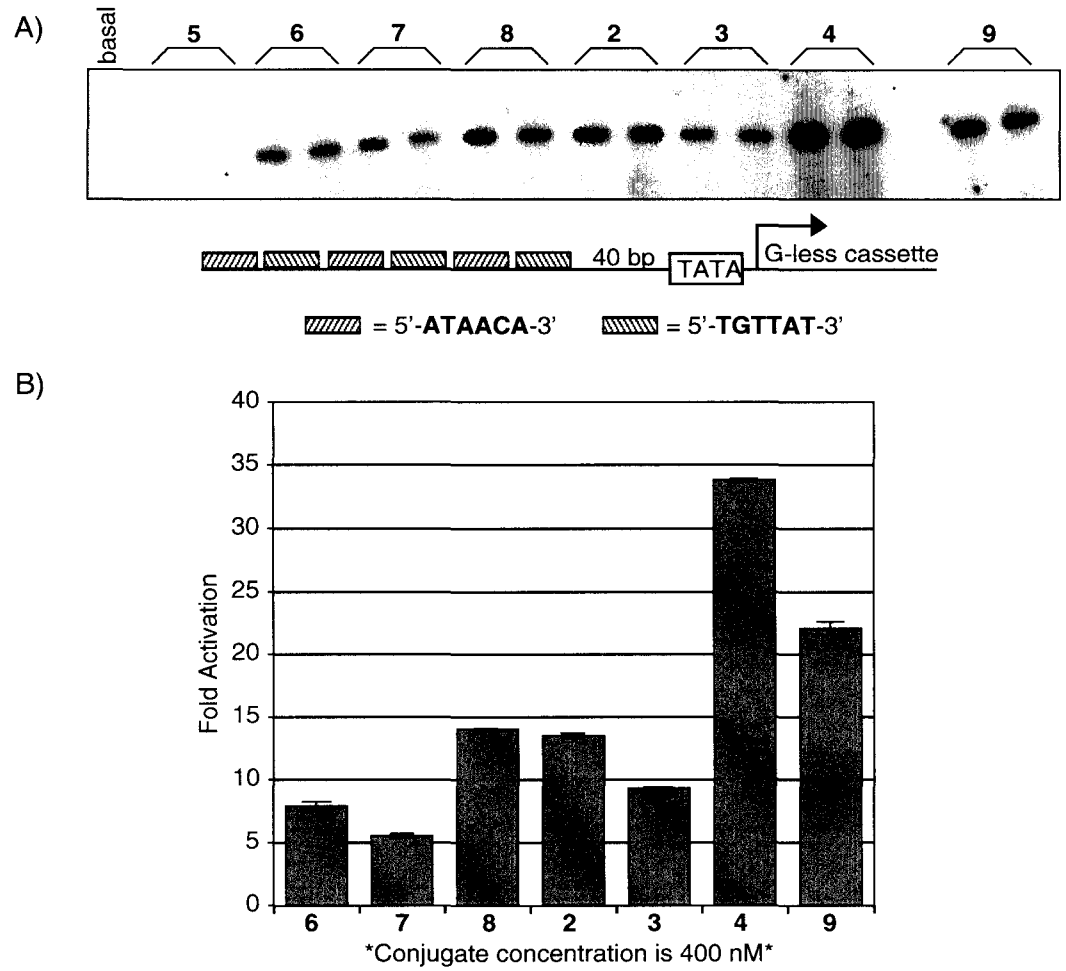


Figure 4.5. *In vitro* transcription reactions with the compounds listed in Figure 4.2. Transcription reaction conditions are described in the methods section. (A) Storage phosphor autoradiogram of the reactions which were performed in duplicate with 400 nM concentrations of each compound including **9**, the 'internal pyrrole' conjugated polyamide. The template configuration showing the sequence and number of the polyamide binding sites is depicted below the gel. (B) Fold activation was determined by comparing the amount of transcription elicited by each compound with that of the parent polyamide, **5**. The standard deviations were determined from the measurement of the duplicates of each compound.

Consistent with the requirement for an activating module to access targets in the transcriptional machinery, projecting the VP1 or VP2 module via a longer linker further improves their activation potential (Figure 4.5B). The normalized data from five independent experiments summarized in Figure 4.4 demonstrates that all of the activator

peptides benefit from the inclusion of a flexible hydrophilic linker. In this configuration, VP2 is now more than 3-fold more potent than VP1. It is possible that the additive increase in transcriptional potency, upon reiteration of VP1 module, in the absence of a linker, occurs because less-than-optimal access of the activation modules to their targets in the transcriptional machinery. Simply improving the reach of the activator modules now leads to a more than additive (~3-fold) increase in activation potency as the VP1 module is reiterated. Moreover, from Figure 4.4 it is apparent that VP2 peptide attached to the C-terminus of the polyamide via a 36 atom linker (PA-1L-VP2, **4**) is the most potent of the polyamide-activating region conjugates synthesized.

Figures 4.4 and 4.5 also show that conjugate **9**, where VP2 is attached to an internal pyrrole residue via an 8 atom linker, activates transcription robustly despite the absence of a long linker. Structural studies of polyamide-DNA complexes predict that projection of the activating region from this internal position should be particularly effective. The ability of this conjugate (**9**) to activate ~60% better than PA-VP2 (**8**), which does not bear a linker, and ~50% less well than one that has a long polyether linker (**4**) further supports the suggestion that projection of the activating module a certain distance away from the DNA plays an important role in determining the efficacy of activation.

Replacing the DNA binding module alters specificity. The final requirement for generality of the polyamide-peptide motif of artificial activators is the ability to alter the identity of the DNA recognition domain without compromising function. To test this we designed conjugate **11** to target the sequence 5'-AGGTCA-3', incorporated the polyether linker as the linking domain, and chose VP2 as the activating region since this

motif had proven to be the most active of all tested (Figure 4.6). Control polyamide **10** and conjugate **11** were prepared by established protocols.[33] As shown in Figure 4.7 a template bearing five cognate binding sites 40bp upstream of the AdML-G less cassette reporter was constructed. As in earlier studies, conjugate **11** activated transcription 15- to 20-fold on Template 2 bearing the cognate binding sites whereas polyamide **10**, which lacks an activation module, had little impact on basal transcription. Furthermore, transcription was dependent upon the presence of cognate DNA binding sites as **11** failed to activate transcription on Template 1 at 400 nM (Figure 4.7). Conversely, conjugate **4** activated transcription from Template 1 but not Template 2. In additional experiments, it was found that when the concentration of conjugate **4** or **11** was increased to 600 nM or greater, low levels (2- to 3-fold) of transcription stimulation were observed on templates lacking the cognate polyamide binding sites (data not shown). Taken together, these data suggest that it is possible to change the identity of the polyamide DNA binding domain without functional penalty.

Significance

Here we describe a series of artificial regulators which were generated by tethering one or two tandem repeats of the minimal activating module of the viral activator VP16. The activating region is either conjugated directly to the polyamide via a short linker or has a longer polyethylene glycol linker between the two modules. These features generate a family of regulators with tunable activation potentials. We also show that tethering the VP2 activating module to an internal pyrrole projects the peptide directly out of the minor groove of the DNA thereby increasing its activation potential

despite the absence of an extended linker. While further increasing the number of reiterated minimal VP modules would increase the activation potency of these conjugates, we sought to maintain the compactness of activator and in this respect conjugate **9** represents a perfect compromise between strength and size. The results described here thus provide evidence that artificial activators of differing strengths can be generated by varying the activation module and its accessibility and that these activators can be targeted to different sites by altering the composition of the polyamides. Taken together our results define the basic design principles for general construction of a minimal artificial regulatory factor. Whether these minimal design elements will be sufficient to allow specific gene activation in cell culture remains to be determined and will be reported in due course.

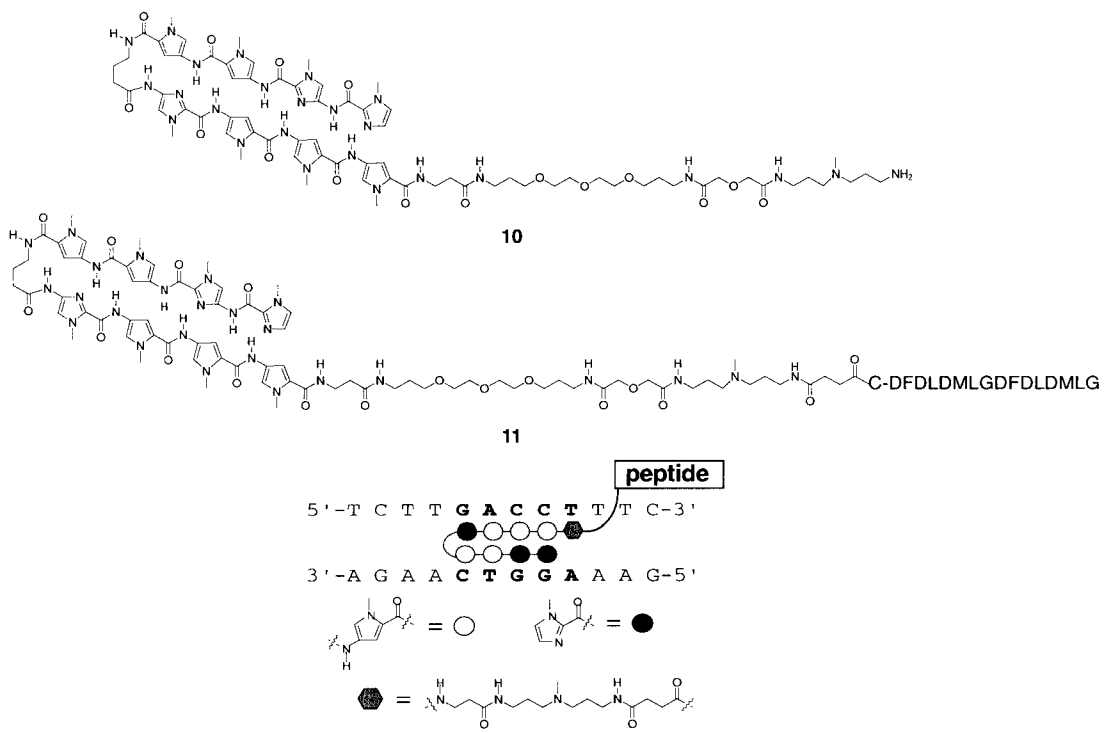


Figure 4.6. Structures of hairpin polyamide **10** and conjugate **11**. Both polyamide **10** and conjugate **11** were designed to target the DNA sequence 5'-TGACCT-3'.

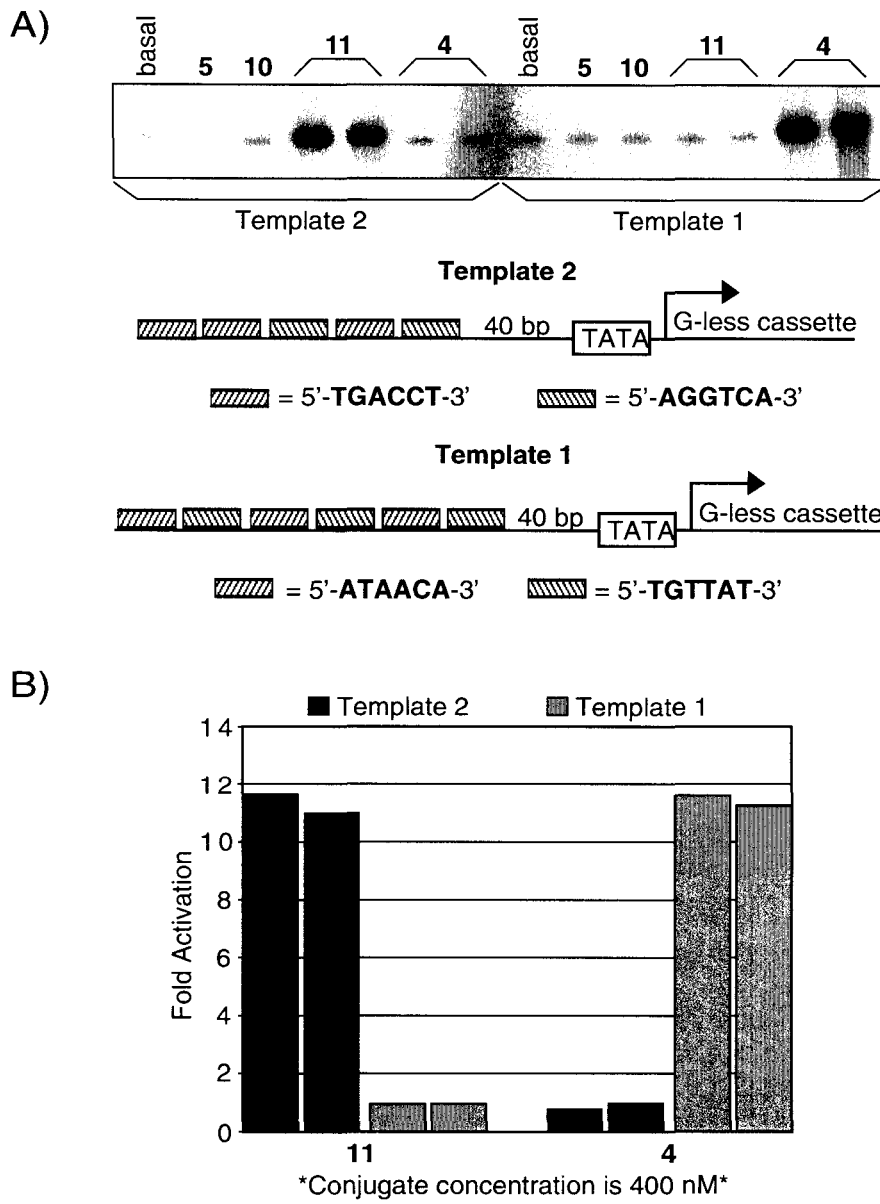


Figure 4.7. Each conjugate activates transcription from a template bearing its own cognate DNA binding site. Conjugates bearing identical linker and activator modules but differing hairpin polyamides target and stimulate transcription from their respective templates. Both templates bear the AdML TATA box and identical transcriptional units. The sequences and the orientation of each polyamide binding site on either template is depicted below the gel. Compounds **10** and **11** are designed to target Template 2 whereas **3** and **5** bind Template 1. Transcription reactions were performed as described in the methods section using 400 nM concentrations of the respective compounds.

Materials and methods

Synthesis of polyamide-peptide conjugates. Polyamides **1**, **5**, and **10** were prepared according to established solid phase synthesis protocols[33] and subsequently transformed into conjugates **2-4**, **6-8**, and **11** by previously reported methods.[1,34] Conjugate **9** was prepared in an analogous fashion. The identity of all conjugates was verified by MALDI-TOF mass spectrometry. Characterization: **2** (PA-1L-AH): MALDI-TOF [M+H] (average mass) calcd 4164.7, obsd 4164.6; **3** (PA-1L-VP1) MALDI-TOF [M+H] (monoisotopic mass) calcd 2763.25, obsd 2763.53; **4** (PA-1L-VP2): MALDI-TOF: [M-H] (average mass) calcd 3670.1, obsd 3670.1; **6** (PA-AH): MALDI-TOF: [M+H] (average mass) calcd 3774.2, obsd 3774.5; **7** (PA-VP1): MALDI-TOF [M+H] (monoisotopic mass) calcd 2374.0, obsd 2374.0; **8** (PA-VP2): MALDI-TOF [M+H] (monoisotopic mass) calcd 3280.4, obsd 3280.5; **9** (PA-(py)-VP2): MALDI-TOF [M+H] (monoisotpic mass) calcd 3280.4, obsd 3280.7; **11**: MALDI-TOF [M+H] (monoisotopic mass) calcd 3670.6, obsd 3670.7.

DNase I Footprinting Titration Experiments. A 363-bp 5'-^{32}P-labeled PCR fragment was generated from template plasmid pAZA812 in accordance with standard protocols and isolated by nondenaturing gel electrophoresis. All DNase I footprinting reactions were carried out in a volume of 40 μl. We note explicitly that 0.8 ng/μl of plasmid pPT7 was used in these reactions as unlabeled carrier DNA. A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were 50 mM HEPES, 100 mM KOAc, 15 mM $Mg(OAc)_2$, 5 mM $CaCl_2$, 6.5% glycerol, 1 mM DTT, pH 7.0 and 15 kcpm 5'-radiolabeled DNA. The solutions were equilibrated for 75 min at 22 °C. Cleavage was initiated by the addition of 4 μl of a

DNase I stock solution and was allowed to proceed for 7 min at 22 °C. The reactions were stopped by adding 10 µl of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/mL glycogen, and 30 µM base pair calf thymus DNA and then ethanol-precipitated. The cleavage products were resuspended in 100 mM Tris-borate-EDTA/80% formamide loading buffer, denatured at 85 °C for 10 min, and immediately loaded onto an 8% denaturing polyacrylamide gel (5% cross-link, 7 M urea) at 2000 V for 2 h 15 min. The gels were dried under vacuum at 80 °C and quantitated using storage phosphor technology.

***In vitro* transcription assays.** Template plasmid pAZA812 was constructed by cloning a 78 bp oligomer bearing three cognate palindromic sequences for conjugates 2-4 and 6-9 into a Bgl2 site 30 bp upstream of the TATA box of pMLΔ53. This plasmid has the AdML TATA box 30 bp upstream of a 277 bp G-less cassette. Template pAZA813 was constructed by cloning a 78 bp oligomer containing five cognate sites for conjugate 11 into a Bgl2 site 30 bp upstream of the TATA box of pMLΔ53.1 For each reaction, 20 ng of plasmid (30 femtomoles of palindromic sites) was pre-incubated with 400 nM concentrations of the compound for 75 minutes prior to the addition of 90 ng of yeast nuclear extract in a 25 µl reaction volume. The reactions were performed as previously described and resolved on 8% 30:1 polyacrylamide gels containing 8 M urea. Gels were dried and exposed to photostimulatable phosphorimaging plates (Fuji Photo Film Co.). Data were visualized using a Fuji phosphorimager followed by quantitation using MacBAS software (Fuji Photo Film Co.).

Acknowledgements

We are grateful to the National Institutes of Health for research support.

References

1. Mapp AK, Ansari AZ, Ptashne M, Dervan PB: Activation of gene expression by small molecule transcription factors. *Proceedings of the National Academy of Sciences USA* 2000; 97:3930-3935.

2. Ptashne M, Gann A: Transcriptional activation by recruitment. *Nature* 1997; 386:569 -577.

3. Koh SS, Ansari AZ, Ptashne M, Young RA: An activator target in the RNA polymerase II holoenzyme. *Mol Cell* 1998; 1:895-904.

4. Kim YJ, Bjorklund S, Li Y, Sayre MH, Kornberg RD: A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 1994; 77:599-608.

5. Koleske AJ, Young RA: An RNA polymerase II holoenzyme responsive to activators. *Nature* 1994; 368:466-469.

6. Lee TI, Young RA: Transcription of eukaryotic protein-coding genes. *Annual Reviews in Genetics* 2000; 34:77-137.

7. Utley RT, Ikeda K, Grant PA, Cote J, Steger DJ, Eberharter A, John S, Workman JL: Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature* 1998; 394:498-502.

8. Peterson CL, Workman JL: Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Current Opinion in Genetic Development 2000 Apr;10(2):187-92* 2000; 10:187-192.

9. Ho SN, Biggar SR, Spencer DM, Schreiber SL, Crabtree GR: Dimeric Ligands Define a Role for Transcriptional Activation Domains in Reinitiation. *Nature* 1996; 382:822-826.

10. Nyanguile O, Uesugi M, Austin DJ, Verdine GL: A Nonnatural Transcriptional Coactivator. *Proceedings of the National Academy of Sciences USA* 1997; 94:13402-13406.

11. Natesan S, Molinari E, Rivera VM, Rickles RJ, Gilman M: A general strategy to enhance the potency of chimeric transcriptional activators. *Proceedings of the National Academy of Sciences USA* 1999; 96:13898-13903.

12. Kuznetsova S, Ait-Si-Ali S, Nagibneva I, Troalan F, Le Villain J-P, Harel-Bellan A, Svinarchuk F: Gene activation by triplex-forming oligonucleotide coupled to the activating domain of protein VP16. *Nucleic Acids Research* 1999; 27:3995-4000.

13. Beerli RR, Dreier B, Barbas CF: Positive and negative regulation of endogenous genes by designed transcription factors. *Proceedings of the National Academy of Sciences USA* 2000; 97:1495-1500.

14. Dervan PB, Burli RW: Sequence-specific DNA recognition by polyamides. *Current Opinion in Chemical Biology* 1999; 3:688-693.

15. White S, Baird EE, Dervan PB: On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chemistry & Biology* 1997; 4:569-578.

16. White S, Szewczyk JW, Turner JM, Baird EE, Dervan PB: Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* 1998; 391:468-471.

17. Gottesfeld JM, Neely L, Trauger JW, Baird EE, Dervan PB: Regulation of gene expression by small molecules. *Nature* 1997; 387:202-205.

18. Dickinson LA, Gulizia RJ, Trauger JW, Baird EE, Mosier DE, Gottesfeld JM, Dervan PB: Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proceedings of the National Academy of Sciences of the USA* 1998; 95:12890-12895.

19. Janssen S, Cuvier O, Muller M, Laemmli UK: Specific gain- and loss-of-function phenotypes induced by satellite-specific DNA-binding drugs fed to drosophila melanogaster. *Mol Cell* 2000; 6:1013-1024.

20. Sadowski I, Ma J, Triezenberg S, Ptashne M: GAL4-VP16 is an unusually potent transcriptional activator. *Nature* 1988; 335:563-564.

21. Wu Y, Reece RJ, Ptashne M: Quantitation of putative activator-target affinities predicts transcriptional activating potentials. *EMBO J.* 1996; 15:3951-3963.

22. Drysdale CM, Duenas E, Jackson BM, Reusser U, Braus GH, Hinnebusch AG: The transcriptional activator GCN4 contains multiple activation domains that are critically dependent on hydrophobic amino acids. *Molecular Cellular Biology* 1995; 15:1220-1233.

23. Seipel K, Georgiev O, Schaffner W: Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. *EMBO J.* 1992; 11:4961-4968.

24. Seipel K, Georgiev O, Schaffner W: A minimal transcription activation domain consisting of a specific array of aspartic acid and leucine residues. *Biol Chem Hoppe Seyler* 1994; 375:463-470.

25. Tanaka M: Modulation of promoter occupancy by cooperative DNA/binding and activation-domain function is a major determinant of transcriptional regulation by activators *in vivo*. *Proceedings of the National Academy of Sciences USA* 1996; 93:4311-4315.

26. Ohashi Y, Brickman JM, Furman E, Middleton B, Carey M: Modulating the potency of an activator in a yeast *in vitro* transcription system. *Molecular Cellular Biology* 1994; 13:2731-2739.

27. Gerber HP, Seipel K, Georgiev O, Hofferer M, Hug M, Russconi S, Schaffner W: Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* 1994; 263:808-811.

28. Blair WS, Bogerd HP, Madore SJ, Cullen BR: Mutational analysis of the transcriptional activation domain of RelA: identification of a highly synergistic minimal acidic activation domain. *Molecular Cellular Biology* 1994; 14:7226-7234.

29. Giniger E, Ptashne M: Transcription in yeast activated by a putative amphipathic α helix linked to a DNA binding unit. *Nature* 1987; 330:670-672.

30. Wemmer DE, Dervan PB: *Current Opinion in Structural Biology* 1997; 7:355.

31. Kielkopf CL, White S, Szewczyk JW, Turner JM, Baird EE, Dervan PB, Rees DC: A structural basis for recognition of A center dot T and T center dot A base pairs in the minor groove of B-DNA. *Science* 1998; 282:111-115.

32. Kielkopf CL, Baird EE, Dervan PD, Rees DC: Structural basis for G center dot C recognition in the DNA minor groove. *Nature Structural Biology* 1998; 5:104-109.

33. Baird EE, Dervan PB: Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *Journal of the American Chemical Society* 1996; 118:6141-6146.

34. Mapp AK, Dervan PB: Preparation of thioesters for the ligation of peptides with non-native substrates. *Tetrahedron Lett.* 2000; 41:9451-9454.

35. Brenowitz M, Senear DF, Shea MA, Ackers GK: Quantitative Dnase Footprint Titration - a Method For Studying Protein-Dna Interactions. *Methods in Enzymology* 1986; 130:132-181.

36. Bremer RE, Baird EE, Dervan PB: Inhibition of major-groove-binding proteins by pyrrole- imidazole polyamides with an Arg-Pro-Arg positive patch. *Chemistry & Biology* 1998; 5:119-133.

37. Ansari AZ, Reece RJ, Ptashne M: A transcriptional activating region with two contrasting modes of protein interaction. *Proceedings of the National Academy of Sciences USA* 1998; 95:13543-13548.

Chapter Five

Allosteric Inhibition of Zinc Fingers Binding in the Major Groove of DNA by Minor Groove Binding Ligands

The text of this chapter was taken in part from a publication co-authored with Elizabeth Ramm, Christina M. Collins, J. Keith Joung, and Professors Peter B. Dervan and Carl O. Pabo.

Publication: Nguyen, D.H.; Ramm, E.; Collins, C.M.; Joung, J.K.; Dervan, P.B.; Pabo, C.O. *Biochemistry*, submitted 2002.

Abstract

In recent years, two methods have been developed that may eventually allow the targeted regulation of any desired gene. The protein strategy involves selecting Cys_2His_2 zinc finger proteins that will recognize specific sites in the major groove of DNA. The small molecule approach utilizes pairing rules for pyrrole-imidazole polyamides that target specific sites in the minor groove. To understand how these two methods might complement each other, we have begun exploring how polyamides and zinc fingers interact when they bind the same site on opposite grooves of DNA. Although structural comparisons show no obvious source of van der Waals collisions, we have found a significant "negative cooperativity" when the two classes of compounds are directed to the overlapping sites. Examining available crystal structures suggests that this may reflect differences in the precise DNA conformation, especially with regard to width and depth of the grooves, which is preferred for binding. These results may give new insights into the structural requirements for zinc finger and polyamide binding and may eventually lead to the development of more powerful and flexible schemes for regulating gene expression.

Introduction

Recent developments in the design and selection of Cys_2His_2 zinc fingers,[1,2] and in the design of novel polyamides,[3,4] suggest that each method can be used to target specific sites in the genome. Cys_2His_2 zinc finger proteins, in which each domain contains 28-30 amino acid residues, are the most common DNA-binding motif in higher eukaryotes. Each zinc finger has a conserved $\beta\beta\alpha$ motif, and amino acids near the N-terminus of the α-helix contact bases in the major groove of B-DNA.[2,5,6,7] The X-ray crystal structure of the three-finger Zif268 protein (Figure 5.1A, references 6,7) illustrates the basic features of recognition and has provided the basis for most of the subsequent work in this field. The Cys_2His_2 zinc finger framework appears to be very adaptable, and specific variants have been selected that bind to many different desired target sites in duplex DNA.[8-11]

Another strategy for recognition and regulation has been developed that uses polyamides consisting of *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) rings which can be linked together to recognize a predetermined DNA sequence.[3,4] Sequence-specific polyamide-DNA recognition depends on binding in the minor groove with side-by-side amino acid pairings. Simple rules were developed for designing polyamides that can target desired DNA sequences.[4] These rules have been carefully validated through characterization of synthesized polyamides via DNase I footprinting, affinity cleavage, two-dimensional nuclear magnetic resonance[12] (NMR), and X-ray crystallographic methods.[13-15] To illustrate the docking arrangement, Figure 5.1B shows the X-ray crystal structure of a polyamide[15] bound in the minor groove of B-DNA.

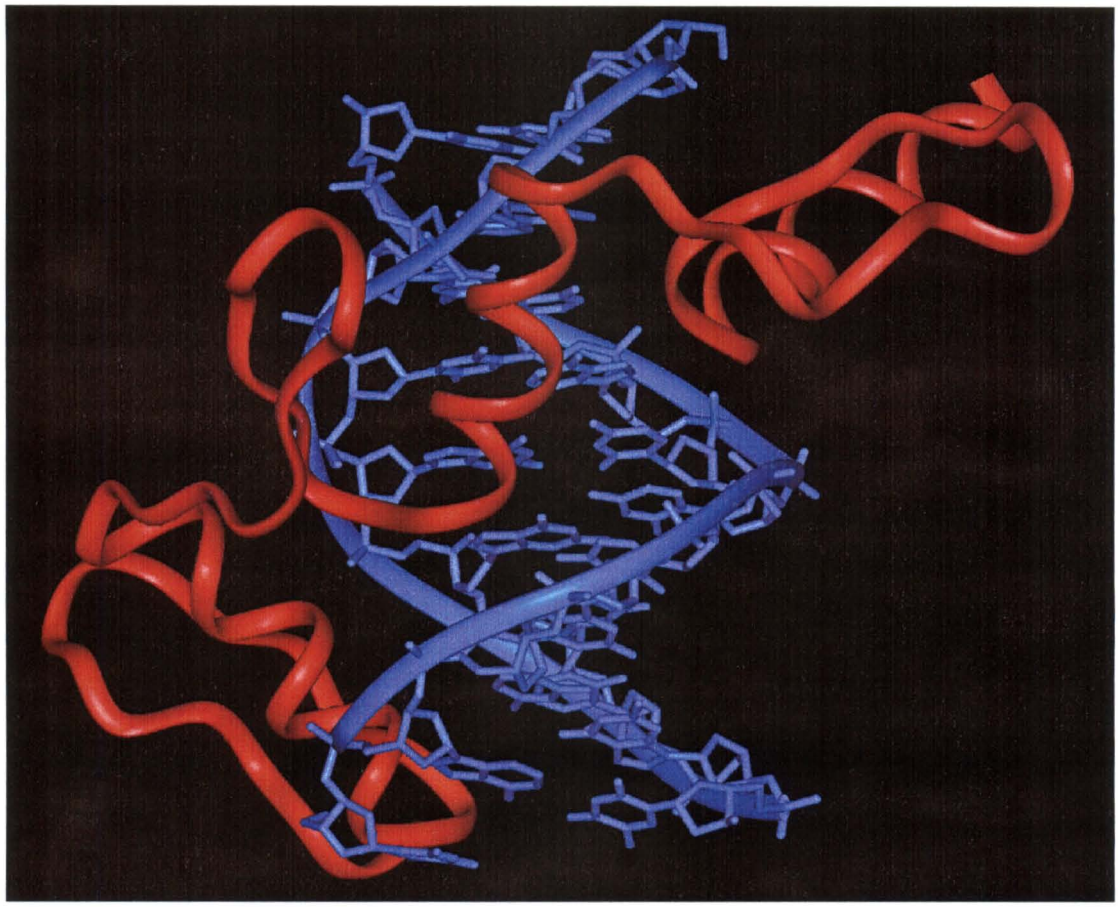


Figure 5.1A. (A) Crystal structure of the Zif268 zinc finger protein (red, with ribbon representing protein backbone) bound in the major groove of B-DNA (blue, with ribbon connecting phosphates). [Coordinates are from the reference 6.]

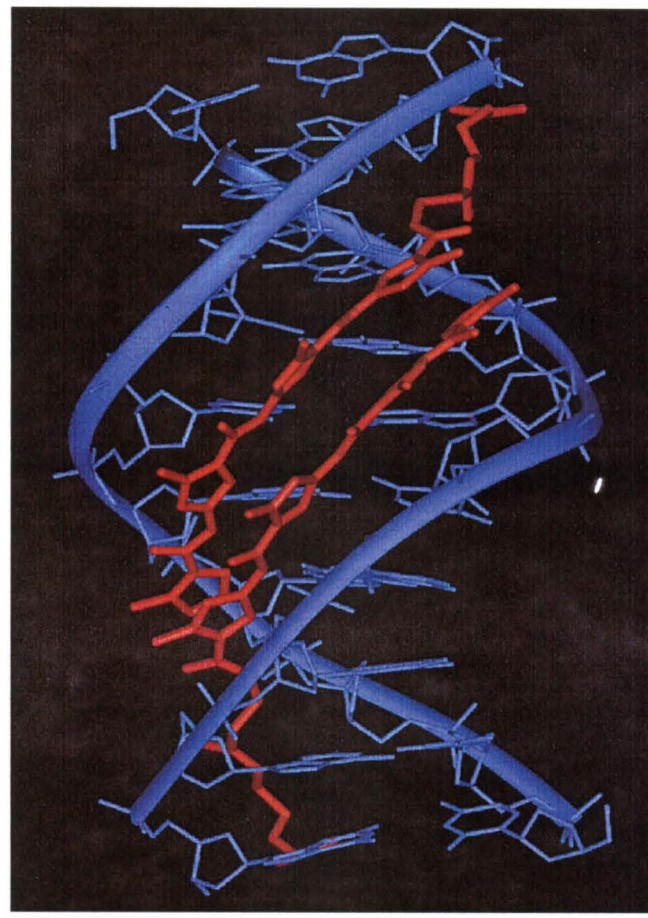


Figure 5.1B. (B) Crystal structure of a polyamide (red) bound in minor groove of B-DNA (blue). [Coordinates are from reference 15 – 1CVY.]

In this work, we have begun studies to investigate how polyamides and zinc fingers may interact as they bind to overlapping sites on double-stranded DNA. Some earlier studies have explored how polyamides interfere with the binding of other transcription factors.[16-23] Polyamides have inhibited minor groove binding proteins such as TATA-binding protein[17] (TBP) as well as minor groove contacting proteins like Hin Recombinase.[18] In the case of the purely major groove binding protein GCN4, polyamides were clearly shown to co-occupy in the minor groove of the GCN4 binding site.[21] An extended Arg-Pro-Arg tripeptide attached to a polyamide did inhibit GCN4 binding by neutralizing a phosphate contact made by the protein with the DNA backbone.[22] Polyamides have been shown to inhibit the binding of a zinc finger protein, TFIIIA.[16,20] However, inhibition of this zinc finger protein was the result of the polyamide being targeted to and displacing the minor groove-spanning fourth finger of the nine finger protein. Therefore, studies to date on polyamide/protein interactions would suggest that an unmodified polyamide should co-occupy the minor groove face of a purely major groove zinc finger protein binding site.

In this study, the interaction of polyamides and zinc fingers that bind to opposing grooves of the same DNA target site was explored. The zinc finger proteins chosen for this study included Zif268 and a set of Zif268 variants that had been selected to recognize rather different DNA sequences.[9] These three other variants had been selected to recognize the TATA box, the nuclear receptor element, and the p53 binding sites, and they are referred to (respectively) as $TATA_{ZF}$, NRE_{ZF}, and $p53_{ZF}$. All these proteins bind their sites with nanomolar dissociation constants, recognizing sites in the DNA major groove and discriminating effectively against nonspecific DNA.[9,24] For this project, polyamides **1-4** were designed to specifically target the very same TATA, NRE, p53, and

Zif268 binding sites. These hairpin polyamides bind their sites with at least nanomolar dissociation constants, as determined by quantitative DNase I footprinting experiments.[25-28] Biochemical studies, using gel mobility shift experiments and DNase I analysis, give information about how these compounds interact as they bind. Computer modeling based on previous structural studies of zinc finger-DNA and polyamide-DNA complexes help us interpret these results.

Results

Gel mobility shift experiments and DNase I footprinting experiments were used to analyze binding of the zinc finger proteins, binding of polyamides, and interactions between the zinc fingers and the match polyamides that recognized overlapping sites. In competition studies, control experiments also were performed with mismatch polyamides to help ensure the specificity of the observed effects. To facilitate comparison of different studies, all experiments were carried out with identical buffer conditions, temperature, equilibration times of the binding reactions, and the same order of the addition of the components into the reaction mixtures.

Polyamide equilibrium dissociation constants. A set of pyrrole-imidazole polyamides was designed, synthesized, and purified for our competitive binding studies. Structures of these polyamides and their expected modes of binding with the DNA minor groove, as predicted on the basis of "pairing rules,[4]" are shown in Figures 5.2 and 5.3. Polyamides **1-4** are referred to as PA_{p53}, PA_{TATA}, PA_{NRE}, and PA_{Zif268} (where our nomenclature indicates which binding site each polyamide has been designed to recognize).

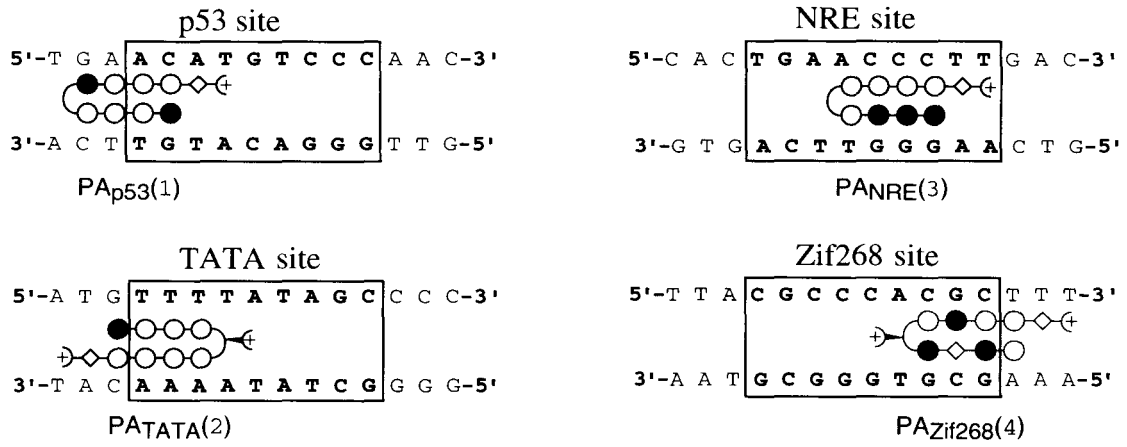


Figure 5.2. Minor-groove binding models expected for hairpin complexes of ImPyPyPy-γ-ImPyPyPy-β-Dp (**1**) (PA_{p53}), ImPyPyPy-(*R*)$^{H}2^{N}$γ-PyPyPyPy-β-Dp (**2**) (PA_{TATA}), ImImImPy-γ-PyPyPyPy-β-Dp (**3**) (PA_{NRE}), and ImPyPyPy-(*R*)$^{H}2^{N}$γ-PyPyPyPy-β-Dp (**4**) (PA_{Zif268}) targeted to their match sites in the minor groove opposite the $p53_{ZF}$, $TATA_{ZF}$, NRE_{ZF}, and $Zif268_{ZF}$ major groove binding sites. Shaded and unshaded circles represent imidazole and pyrrole carboxamides, respectively, and the β-alanine residue is represented by an unshaded diamond. Boxes enclose the 9 bp sites that are recognized by the corresponding zinc finger proteins.

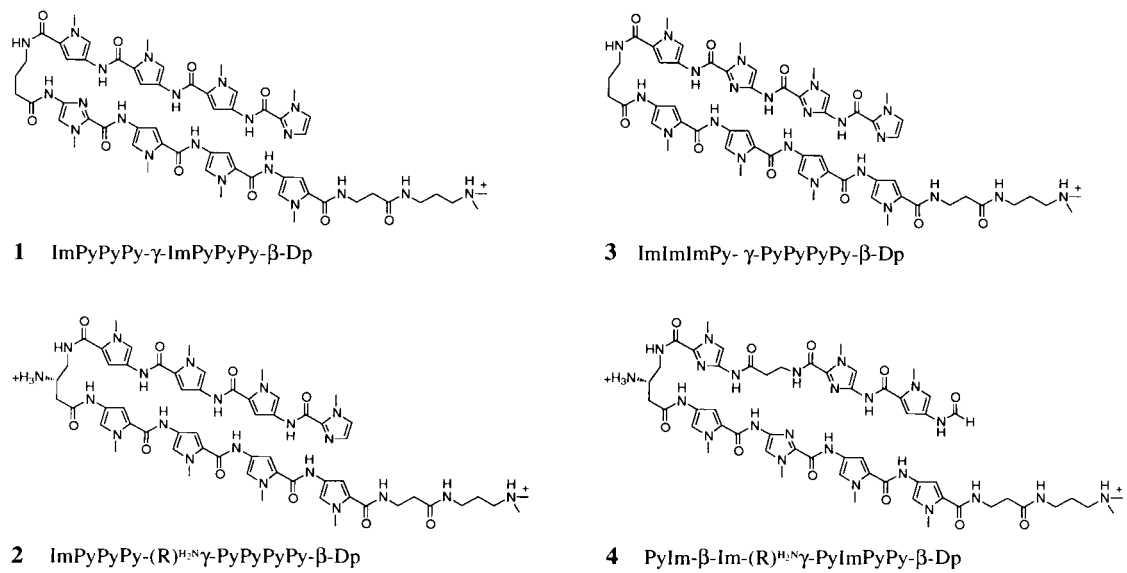


Figure 5.3. Full chemical structures of the hairpin polyamides: PA_{p53} ImPyPyPy-γ-ImPyPyPy-β-Dp (**1**), PA_{TATA} ImPyPyPy-(*R*)$^{H}2^{N}$γ-PyPyPyPy-β-Dp (**2**), PA_{NRE} ImImImPy-γ-PyPyPyPy-β-Dp (3), and PA_{Zif268} ImPyPyPy-(*R*)$^{H}2^{N}$γ-PyPyPyPy-β-Dp (**4**).

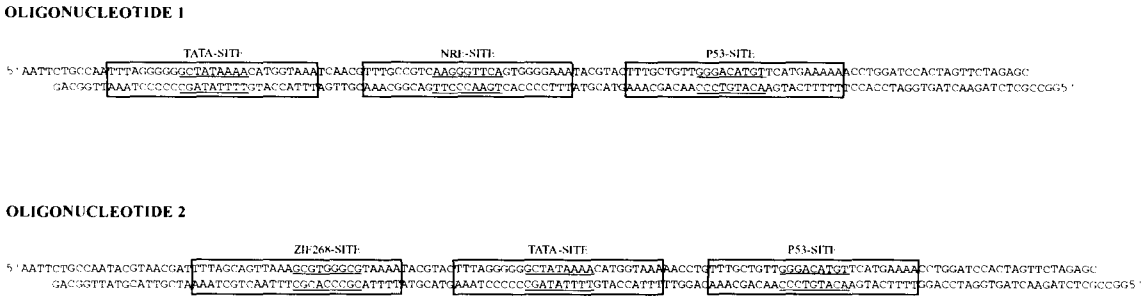
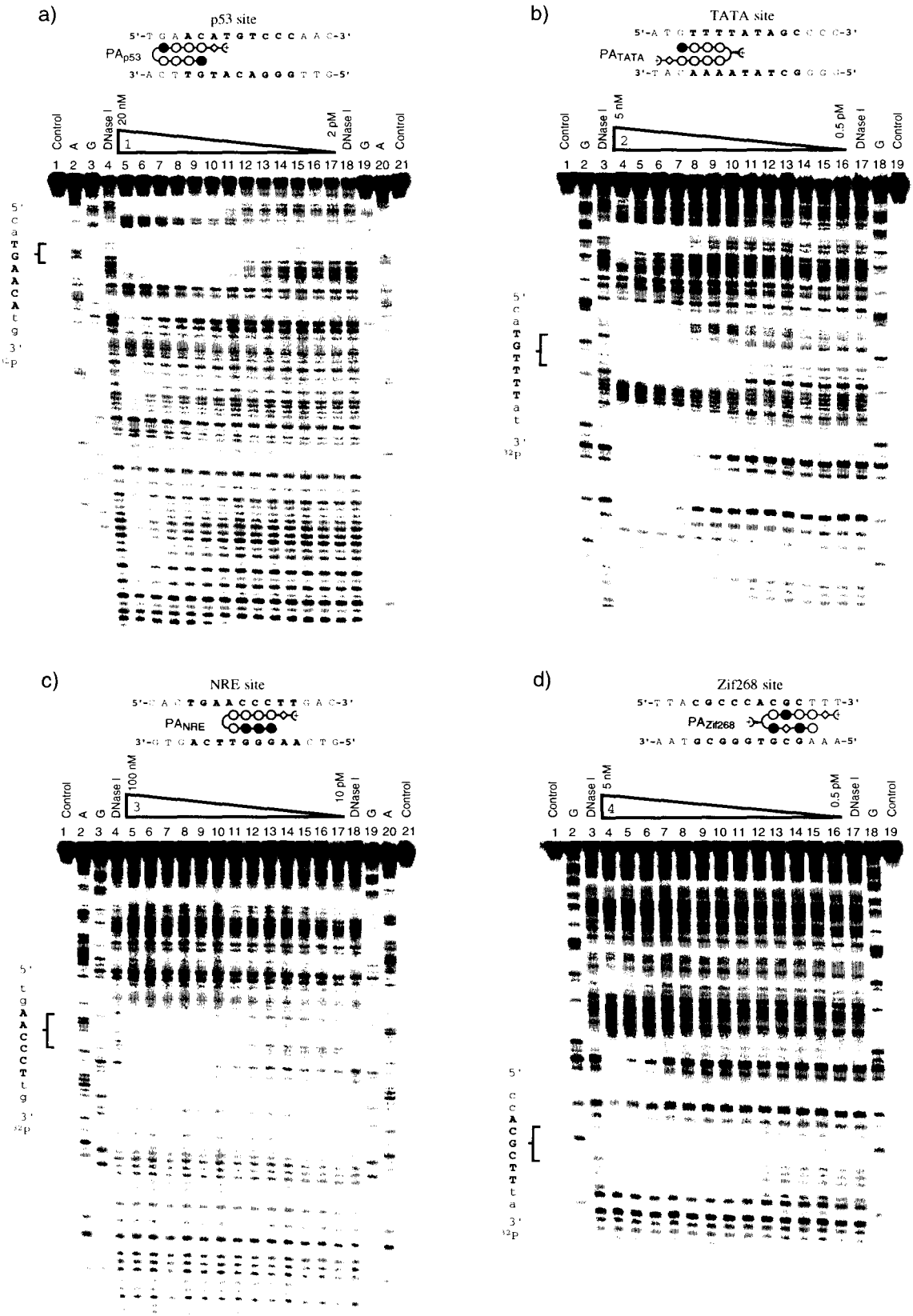


Figure 5.4. Oligonucleotide sequences used for DNase I footprinting experiments, with each of the nine base pair target sites for the zinc finger proteins underlined. Oligonucleotide 1 contains the $TATA_{ZF}$, NRE_{ZF}, and $p53_{ZF}$ binding sites; oligonucleotide 2 contains the Zif268, $TATA_{ZF}$, and $p53_{ZF}$ binding sites. Gel mobility assays used shorter oligonucleotides (27 bp segments each containing a single binding site), and the corresponding segments used in these experiments are shown as boxed regions surrounding each of the 9 bp binding sites. (In this orientation the site that has been cleaved by *Eco*RI is at the left end of the oligonucleotide; the site cleaved by *Not*I is on the right.)

We used quantitative DNase I footprint titration analysis to determine the apparent dissociation constants for the polyamides. Oligonucleotide 1 was used for PA_{p53} and PA_{NRE}; oligonucleotide 2 was used for PA_{TATA} and PA_{Zif268} (Figure 5.4). The oligonucleotides were labeled from the 3'-end. DNase I footprinting experiments (Figure 5.5), done at a series of different polyamide concentrations, reveal the location of the binding site for each polyamide and allow determination of the binding constants. These footprinting experiments (Figure 5.5) show PA_{TATA} and PA_{Zif268} polyamides also bind specifically to other match and single base pair mismatch sites (as defined by the pairing rules), found on the restriction fragment, at the same or higher concentration, respectively. Quantitative analysis of the degree of protection observed at each polyamide concentration allowed us to present the data of the DNase I footprinting experiments as binding isotherms (Figure 5.6). The apparent K_D, equal to $1/K_A$, was determined from the binding isotherms by fitting the data points with a modified Hill equation, as previously described.[28] Theses binding isotherms and apparent dissociation

Figure 5.5. Quantitative DNase I footprint titration experiments used to monitor binding of polyamides to the oligonucleotides shown in Figure 5.4. Oligonucleotides were obtained as *Eco*RI/*Not*I restriction fragments from the pBluescript II SK+ plasmid and were 3'-^{32}P-labelled. a) PA_{p53} (**1**) on oligonucleotide 1: lanes 1 & 21, intact DNA; lanes 2 & 20, A-specific chemical sequencing reaction; lanes 3 & 19, G-specific chemical sequencing reaction; lanes 4 & 18, DNase I digestion products in the absence of polyamide; lanes 5-17 DNase I digestion products in the presence of 20, 10, 5, 2, 1, 0.5, 0.2, 0.1 nM, 50, 20, 10, 5, 2 pM polyamide, respectively. b) PA_{TATA} (**2**) on oligonucleotide 2: lanes 1 & 19, intact DNA; lanes 2 & 18, G-specific chemical sequencing reaction; lanes 3 & 17, DNase I digestion products in the absence of polyamide; lanes 4-16 DNase I digestion products in the presence of 5, 2, 1, 0.5, 0.2, 0.1 nM, 50, 20, 10, 5, 2, 1, 0.5 pM polyamide, respectively. c) PA_{NRE} (**3**) on oligonucleotide 1: lanes 1 & 21, intact DNA; lanes 2 & 20, A-specific chemical sequencing reaction; lanes 3 & 19, G-specific chemical sequencing reaction; lanes 4 & 18, DNase I digestion products in the absence of polyamide; lanes 5-17 DNase I digestion products in the presence of 100, 50, 20, 10, 5, 2, 1, 0.5, 0.2, 0.1 nM, 50, 20, 10 pM polyamide, respectively. d) PA_{Zif268} (**4**) on oligonucleotide 2: lanes 1 & 19, intact DNA; lanes 2 & 18, G-specific chemical sequencing reaction; lanes 3 & 17, DNase I digestion products in the absence of polyamide; lanes 4-16 DNase I digestion products in the presence of 5, 2, 1, 0.5, 0.2, 0.1 nM, 50, 20, 10, 5, 2, 1, 0.5 pM polyamide, respectively.


a)
p53 site
PA$_{p53}$
Control
A
G
DNase I
20 nM
2 pM
DNase I
G
A
Control
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
b)
TATA site
PA$_{TATA}$
Control
G
DNase I
5 nM
0.5 pM
DNase I
G
Control
2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
c)
NRE site
PA$_{NRE}$
Control
A
G
DNase I
100 nM
10 pM
DNase I
G
A
Control
3
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
d)
Zif268 site
PA$_{Zif268}$
Control
G
DNase I
5 nM
0.5 pM
DNase I
G
Control
4
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

constants for each polyamide represent the average of three independent experiments, and all these polyamides have at least nanomolar affinity for their expected target sites. DNase I footprinting of polyamides 1-4 was also carried out in triplicate under standard TKMC conditions. No difference was seen in the binding affinity of the polyamides as a result of the change from TKMC buffering conditions to those used in the gel mobility shift assays.

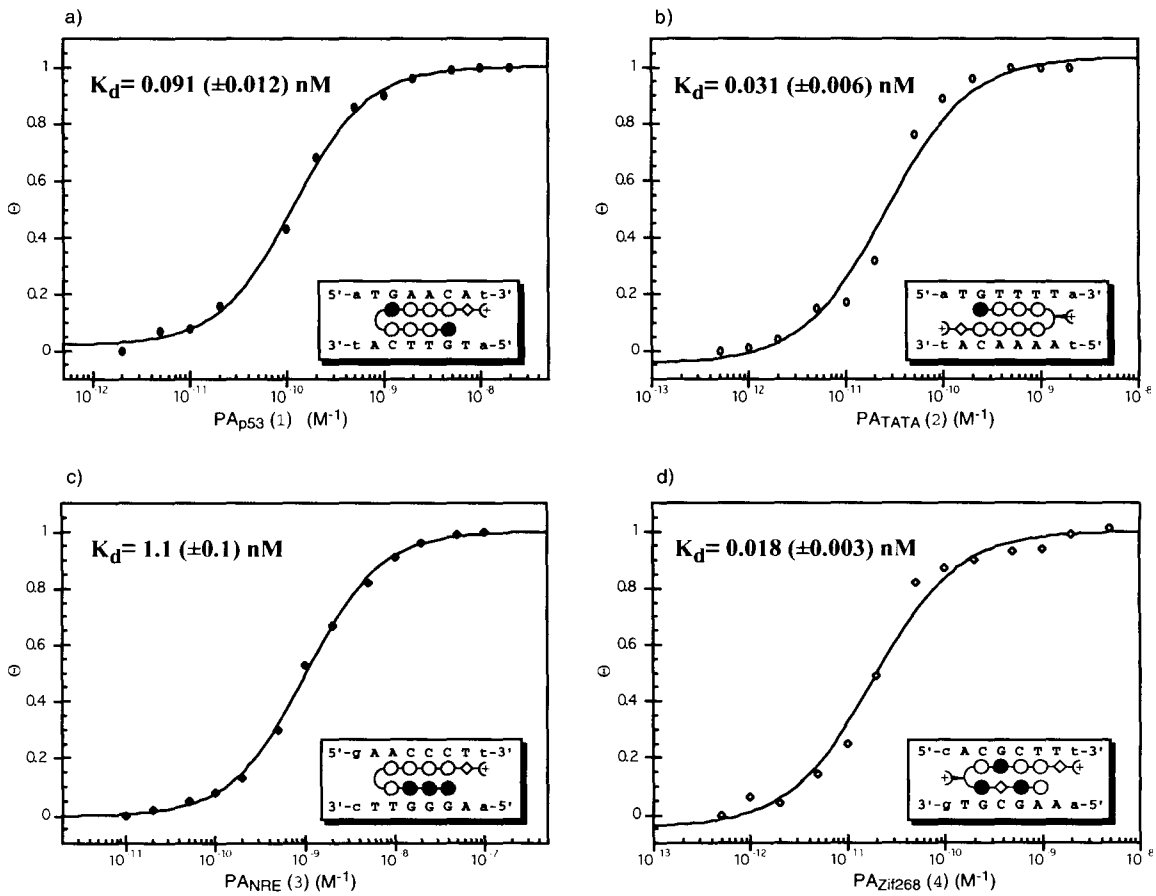


Figure 5.6. Binding isotherms derived from the DNase I quantitative footprint titration experiments (Figure 5.5) as shown for the PA_{p53} (a), PA_{TATA} (b), PA_{NRE} (c), and PA_{Zif268} (d) polyamides. In each case, isotherms represent binding at the match target site in the minor groove, which has the sequence shown in the lower right panel of the figure, and which overlaps with the 9 bp site in the major groove recognized by the corresponding zinc finger protein. Θ_{norm} points were obtained using storage phosphor autoradiography and processed as previously described.[28] Each data point represents the average of three independent quantitative footprint titration experiments and the solid curves are the best-fit Langmuir binding titration isotherms obtained from a nonlinear least-squares algorithm where n=1.

Determination of zinc finger dissociation constants. This study used Zif268 and three variants, specific for the p53, NRE, and TATA binding sites, that had been selected via phage display. (Note that the sequential selection protocol[9] used to obtain these variants allowed extensive changes in the recognition site, since six amino acids in each finger had been randomized.) Quantitative gel shift analysis was used to determine the fraction of the DNA fragment bound at a series of protein concentrations (Figure 5.7), and K_D values were calculated from the slope of Scatchard plots. Corrections were made for the active concentration of each protein, which was estimated using stoichiometric competition experiments. The sequences of the 27 bp oligonucleotides used in gel mobility shift experiments are boxed in Figure 5.4. (Note: In this figure, the oligonucleotides are shown embedded within larger sites that were later used for DNase I footprinting). The dissociation constants for the zinc fingers (Table 5.1)

Table 5.1. Equilibrium dissociation constants of DNA-zinc finger protein binding in the absence and in the presence of polyamides.

Polyamide/ Protein	**K_D (nM)[a]**
$p53_{ZF}$	0.240 ± 0.018
$PA_{p53}/p53_{ZF}$	3.200 ± 0.283
$TATA_{ZF}$	0.204 ± 0.030
$PA_{TATA}/TATA_{ZF}$	4.320 ± 0.582
NRE_{ZF}	0.101 ± 0.009
PA_{NRE}/NRE_{ZF}	3.520 ± 0.248
Zif268	0.010 ± 0.003
PA_{Zif268}/Zif268	0.182 ± 0.009

[a] The reported equilibrium dissociation constants are apparent for the proteins in the presence of polyamides. All constants are the mean values obtained from three or more gel mobility shift experiments.

were determined from the average of three independent experiments, and the values are comparable to those previously reported,[9] even though slightly different buffers were used.

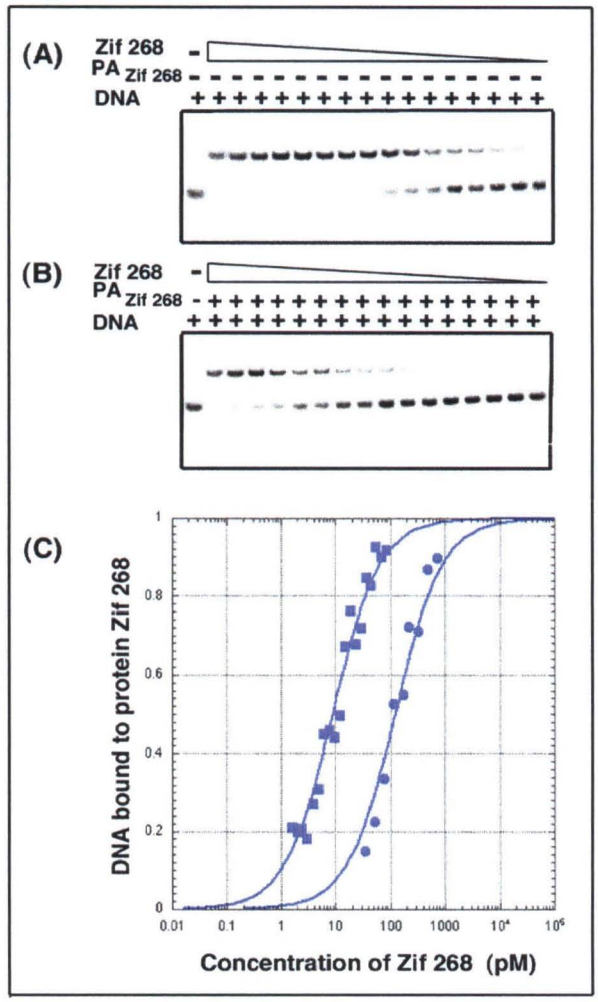


Figure 5.7. Gel mobility shift experiment showing that the PA_{Zif268} polyamide interferes with binding of Zif268 protein. (A) Binding of Zif268 protein in the absence of polyamide, where upper band on the gel represents the protein-DNA complex. Protein was diluted 1.5 X between lanes, and the 27 bp. DNA probe had a concentration of 0.5 pM. (B) Similar binding experiment, but this was performed in the presence of a constant amount of polyamide. (The concentration was chosen as $10 \times K_D^{PA}$ and thus is expected to give about 90% saturation of the site in the absence of protein.) (C) Zif268 binding isotherms in the absence (squares) and in the presence (circles) of PA_{Zif268}. Each data point represents the average of the three independent gel mobility shift experiments. The binding curves show that a protein concentration of 10 pM gives 50% occupancy when no polyamide is present; a protein concentration of 182 pM is required to give 50% occupancy when the polyamide is present.

Dissociation constant of zinc finger protein in the presence of specific polyamides. Binding of the polyamides could not be directly observed in our gel shift experiments, but competition studies clearly revealed that the polyamides affected formation and/or stability of the zinc finger-DNA complexes. Two types of competition experiments were done to analyze the effects of polyamides on zinc finger binding. The first set of experiments used a fixed polyamide concentration that would have given ~ 90% occupancy of the free DNA site, and gel shift experiments were used to determine the apparent binding constant of the zinc finger proteins under these conditions. These gel shift experiments were essentially identical to those done with the protein alone. The long equilibration time, done for all gel shift experiments, ensured that the apparent binding did not depend on the order of addition of the components. Figure 5.7 shows one set of gel shift experiments, where the apparent binding constant of Zif268 is determined in the presence and absence of the polyamide that recognizes an overlapping site. Binding isotherms for this reaction are shown in Figure 5.7C, and similar results were obtained for the other polyamide/zinc finger competitive binding titrations. In every case, the presence of the corresponding polyamide significantly decreases (by 13 to 35-fold) the apparent affinity of the zinc finger protein for the binding site (Table 5.1).

Interactions between the polyamides and zinc fingers were also studied in experiments that used fixed zinc finger protein concentrations with variable amounts of polyamide. In each of these experiments, the zinc finger proteins were present at active concentrations equal to 10× their respective dissociation constants, and thus about 90% of the DNA was initially shifted by the proteins. Similar binding reactions with increasing amounts of polyamide were conducted, and after allowing for full equilibration, we monitored the effect of the polyamides on formation of the zinc finger-DNA complexes.

Gel shift results obtained with the PA_{Zif268} polyamide and the Zif268 protein (Figure 5.8A) show that the polyamide interferes with the formation or stability of the zinc finger-DNA complex. Similar results were obtained using other zinc finger

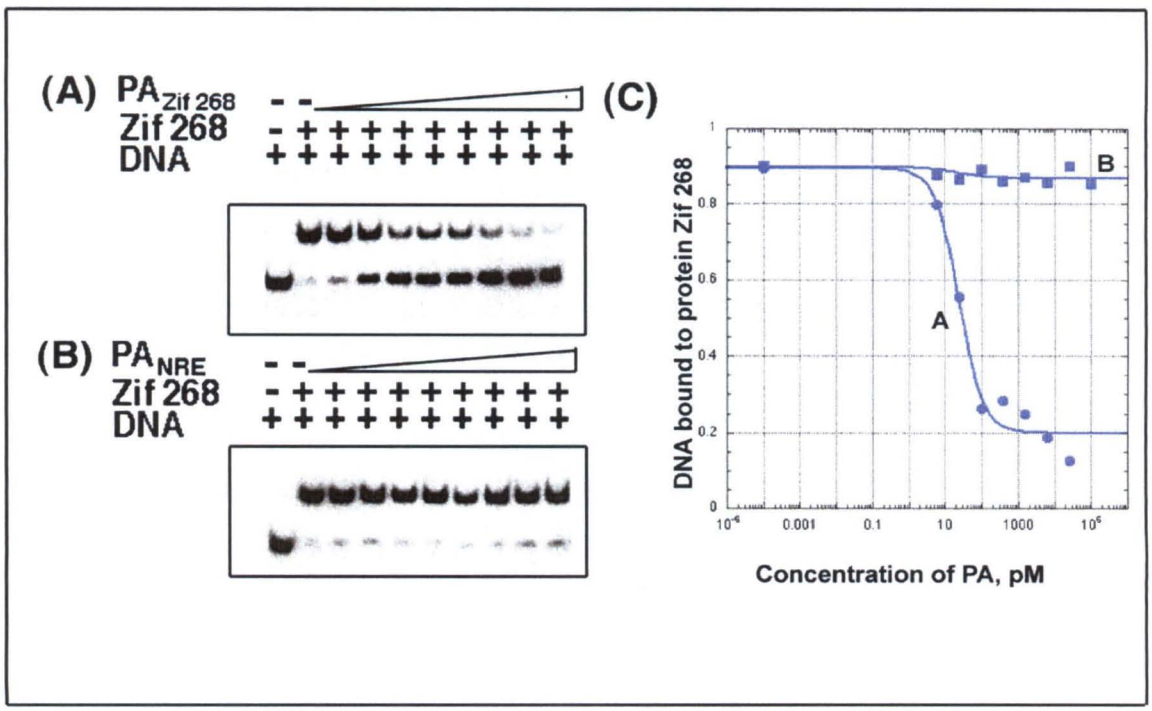


Figure 5.8. Specificity of polyamide interference effects. In these experiments the active concentration of protein was constant and equal to $10 \times K_D^{prot.}$ (giving ~90% of saturation of DNA binding sites when no polyamide was present). The DNA probe was present at 0.5 pM, and the polyamide concentration varies 4-fold between adjacent lanes. (A) Representative gel mobility shift results for Zif268 protein binding to DNA probe in the presence of increasing concentrations of PA_{Zif268} polyamide. (B) Corresponding experiments using the Zif268 protein, the Zif268 binding site, and the polyamide PA_{NRE}. [Note: This polyamide will not bind to the site and thus represents a "mismatch" control for the purposes of this experiment.] (C) Effects of polyamides PA_{Zif268} (circles) and PA_{NRE} (squares) on Zif268 protein binding. Each data point represents the average of three independent experiments, and difference between these curves emphasizes the specificity of polyamide binding and interference.

proteins and polyamides which compete for the same binding sites. Protein bound was calculated as a function of polyamide concentration, with each data point representing the average of three independent experiments. Data was fit with the modified Hill equation,[28] enabling us to estimate "inhibition constants." These represent the polyamide concentrations that give 50% inhibition of formation of the respective protein-DNA

complex, and we find that these "inhibition constants" are in the sub-nanomolar range (Table 5.2).

Table 5.2. Equilibrium Inhibition Constants.

Polyamide/Protein	app**Ki** (nM)[a]
$PA_{p53}/p53_{ZF}$	0.20 ± 0.07
$PA_{TATA}/TATA_{ZF}$	0.04 ± 0.02
PA_{NRE}/NRE_{ZF}	16.40 ± 3.90
PA_{Zif268}/Zif268	0.03 ± 0.02

[a] The reported equilibrium inhibition constants are the mean values obtained from three gel mobility shift experiments.

Testing for specificity of zinc finger/polyamide interference effects. Several types of experiments were done to test the specificity of the observed interference effects between polyamides and zinc fingers. For example, Figure 5.8 shows gel shift data obtained with the Zif268 protein-DNA complex. This figure compares the interference effects observed with PA_{Zif268} (the match polyamide with specific binding for this site) and interference effects observed with PA_{NRE} (which is specific for the NRE site and a mismatch polyamide with no binding site on the 27 bp Zif268 oligonucleotide). There is marked interference by PA_{Zif268}, as indicated by the gel shift results in Figure 5.8A and the sigmodal plot shown in Figure 5.8C. However, similar concentrations of PA_{NRE} have no measurable effect at this site, as there is no reduction in Zif268 binding even at high polyamide concentrations (Figures 5.8B and 5.8C). Similar data were obtained using other zinc finger-polyamide combinations, and these results show that interference is dependent on having a polyamide that can compete for the same binding site.

Competitive DNase I footprinting experiments. Since binding of polyamides could not be directly observed in any of our gel shift experiments, we used DNase I footprinting to further explore the mechanism of polyamide/zinc finger interference effects. In principle, these experiments should allow us to directly monitor polyamide-DNA interactions, zinc finger-DNA interactions, and possible formation of the ternary complex. Figure 5.9 shows the results of DNase I footprinting experiments that explored interference effects between the Zif268 protein and the PA_{Zif268} polyamide. These experiments used oligonucleotide 2 (Figure 5.4), labeled from the 5' end, to provide a Zif268 protein and PA_{ZF} polyamide binding site. The PA_{NRE} polyamide, which does not have a binding site on this oligonucleotide, was used as a "mismatch" control. Six sets of experiments were performed, using the PA_{Zif268} polyamide at concentrations of 1, 5, and 10 nM (panel A) and then using the PA_{NRE} polyamide at corresponding concentrations (panel B). When the Zif268 protein was present, it was used at a concentration that would have been expected to give approximately 90% occupancy of the free DNA site. As indicated at the top of the respective lanes (Figure 5.9), four conditions were tested within each set of reactions: 1) the first lane within each set has no protein and no polyamide, 2) the second lane has Zif268 protein but no polyamide, 3) the third lane has Zif268 protein and one of the polyamides, and 4) the fourth lane has the same polyamide, but contains no protein.

We used Gilbert-Maxam G+A sequencing reactions to determine the precise base positions where cleavage occurred in these footprinting experiments, and to locate the binding sites for the Zif268 protein and for the PA_{Zif268} polyamide, which are marked on the left side of the gels. (Note that experiments in panel B used the PA_{NRE} polyamide, which does not have a binding site on this oligonucleotide.) As expected, there is

significant overlap between the region where Zif268 footprints and the region where PA_{Zif268} footprints: the PA_{Zif268} binding site includes bases 32-39, while the Zif268 protein binding site includes bases 35-43 (when counting from the 5' end). However, it appears that band 43, at the far end of the Zif268 binding site, can only be effectively protected by the protein.

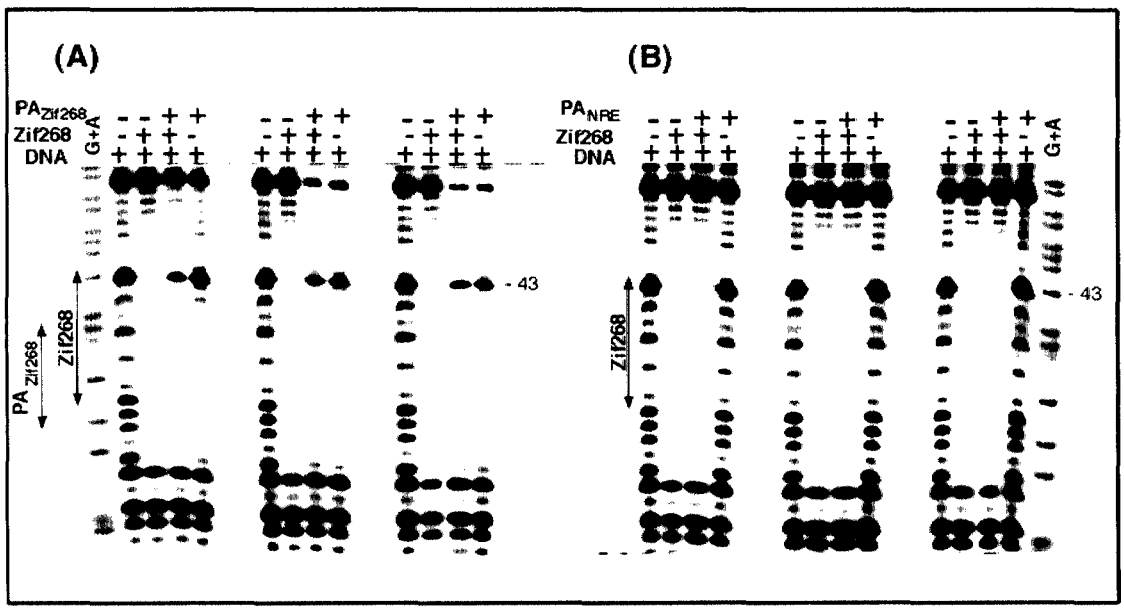


Figure 5.9. DNase I footprinting analysis of polyamide/zinc finger protein interference effects. (A) Reactions using Zif268 protein, Zif268 DNA site, and PA_{Zif268} polyamide. Left lane within each group of four has no protein and no polyamide; second lane has protein but no polyamide; third lane has protein and polyamide; forth lane has polyamide alone. When present, concentration of Zif268 protein was equal to 10 × K_D; concentrations of polyamides were 1 nM in first set of reactions (i.e., the first four lanes); 5 nM in second set of reactions; and 10 nM in third set of reactions. Lanes labeled G+A show markers prepared with the Gilbert-Maxam G+A reaction protocol, and binding sites for the Zif268 protein and PA_{Zif268} polyamide are marked on the left side of the gel. (B) Control experiment (with a similar arrangement of lanes and choice of concentration) using the Zif268 protein, Zif268 DNA and PA_{NRE} polyamide. [Note that the PA_{NRE} polyamide does not have a binding site on this oligonucleotide and thus serves as a "mismatch" control for these experiments.]

Results obtained with the mismatch polyamide (panel B) are entirely straightforward: Zif268 occupies its binding site in lanes 2 and 3 of each experiment, and the PA_{NRE} polyamide (which is present in lanes 3 and 4 but has no binding site in this oligonucleotide) has no effect on the footprinting patterns. The results with the PA_{Zif268}

polyamide (panel A) are somewhat more complicated, but they are the real crux of the experiment. In this panel, lane 2 of each set (protein alone) shows the expected footprint for the Zif268 protein. Lane 4 of each set (polyamide alone) shows the footprints obtained when the PA_{Zif268} polyamide is present at concentrations of 1nM, 5 nM and 10 nM. Effects are localized to the expected binding site at low concentration (1 nM), but additional protection due to specific binding of a single base pair mismatch site, present in the restriction fragment (near the top of the gel), is seen at higher concentrations.

Obviously, the most interesting results in each set of reactions from panel A involve lane 3, in which both the Zif268 protein and the PA_{Zif268} polyamide were present. At the higher polyamide concentrations (5nM and 10nM), the footprinting patterns obtained with the polyamide/zinc finger combinations (lane 3) appear surprisingly similar to those obtained with the polyamide alone (lane 4). The simplest interpretation of these patterns is that binding of the polyamide displaces the zinc fingers, making the patterns look very similar in lanes 3 and 4. It is somewhat more problematic to explain the pattern obtained in lane 3 of the first set (at a 1 nM concentration of PA_{Zif268}), but this may indicate a mixture of species, with the Zif268 protein bound to some fraction of the DNA sites and the PA_{Zif268} polyamide bound to most of the remaining DNA sites.

Computer modeling of polyamide and zinc finger structures. In a further attempt to understand the mechanisms of interference, we examined the structure of various zinc finger-DNA and polyamide-DNA complexes. X-ray crystal structures are now available for the $TATA_{ZF}$,[34] $p53_{ZF}$,[35] and Zif268[6] zinc finger-DNA complexes. X-ray crystal structures also are available for several polyamide-DNA complexes,[12,13,15] and we thought that these structures would be useful for initial modeling, even though the exact sequences of the polyamides and of the binding sites are somewhat different than

those used in our current experiments. Our first major conclusion is that we do not see any obvious basis for a steric collision between the zinc fingers and the polyamides. All the polyamide-DNA complexes indicate that these compounds bind exclusively in the minor groove, while the zinc finger-DNA complexes show that these proteins bind in the major groove. [The only exception here involves one lysine (Lys 189) in the TATA structure with a high temperature factor that might reach into the minor groove, but the polyamide binds at the opposite end of this site, far from Lys 189].

Various zinc finger-DNA and polyamide-DNA complexes were superimposed, and these were compared in an attempt to understand the basis for the "negative cooperativity" that we had observed in biochemical studies. Alignments done with the program PROTEUS,[36] typically used 3-4 sets of phosphates or 3-4 sets of C1' atoms to superimpose the complexes. The rms deviations in these alignments varied from about 0.6Å to about 2Å, depending on 1) which structures were aligned, 2) which atoms were used for alignment, and 3) how many base pairs were superimposed. However, there were striking and consistent differences in the groove width, and these were large enough that they could readily explain how the polyamides interfere with zinc finger binding. In nearly every alignment we carried out, the polyamide-DNA complex had a wider and deeper minor groove than observed in the zinc finger-DNA complexes. There were also a number of cases where superimposing the complexes indicated shifts in the precise arrangement of the base pairs, and these could affect the position of hydrogen bond donors and acceptors that are critical for recognition.

These initial qualitative observations of the structures were confirmed by using the Curves program[38] to calculate dimensions of the major and minor grooves. We find that the major grooves in zinc finger-DNA complexes are consistently wider and deeper

than the major grooves in the polyamide-DNA complexes. In this set of structures, the Zif268 complex has the deepest major groove (6 to 8Å), while the $TATA_{ZF}$ and $p53_{ZF}$ complexes have the widest major grooves (13 to 14.5Å). The major groove width of the polyamide-DNA complexes ranged from 9-10.5Å, while the major groove depth ranged from 3 to 5.5Å. Differences in minor groove width are not quite as consistent when comparing the two classes of complexes, but there is a general tendency for the polyamide/DNA complexes to have wider minor grooves than the zinc finger-DNA complexes.

The initial superimpositions (Figure 5.10) had indicated that the orientations of the bases were slightly different in the two types of complexes, and these differences were confirmed and analyzed using the program Curves. In general, we find that 1) The base pairs have negative opening angles in the polyamide-DNA complexes, but positive opening angles in the zinc finger-DNA complexes. 2) The zinc finger complexes have a larger negative X-displacement of the base pairs than do the polyamide complexes. 3) Parameters for the polyamide-DNA complexes are generally closer to those of B-DNA than are parameters for the zinc finger-DNA complexes. No significant differences in shear or buckle were noted when comparing these two classes of complexes.

Discussion

Compounds that can target specific sites on double-stranded DNA may provide powerful new tools for the regulation of gene expression. Excellent progress has been reported in 1) the design and selection of Cys_2His_2 zinc fingers for these purposes[1,2] and in 2) the design of pyrrole-imidazole polyamides that will bind according to well-defined "recognition rules.[4]" These two broad classes of compounds may provide powerful new

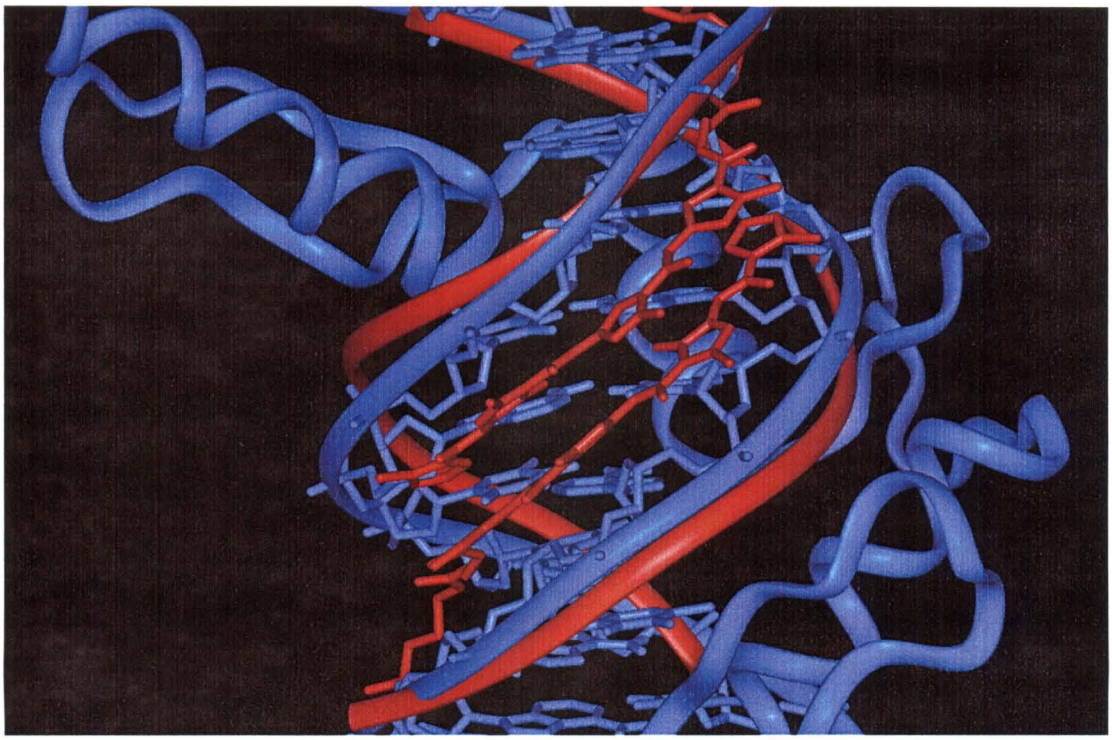


Figure 5.10. Superposition of zinc finger – DNA complex [Coordinates are from the reference 34] (blue) and polyamide – DNA complex [Coordinates are from the reference 13] (red), oriented to highlight differences in minor groove width. Such "allosteric" changes in the groove dimensions may explain the "negative cooperativity" observed in our binding studies with polyamides and zinc fingers.

reagents for molecular medicine and gene therapy, and interest in this type of strategy has been further heightened by recent publication of a draft sequence of the human genome. There is a real prospect that fundamental principles of molecular recognition, including a careful analysis of the relevant affinity and specificity of these compounds, can prepare the way for bold new medical applications.

Our central goal in this paper was to explore how zinc fingers and polyamides might interact when targeted to overlapping recognition sites. We thought this would be interesting from a structural and physical/chemical perspective, and we thought this work also might provide a basis for the eventual development of new regulatory schemes (perhaps where polyamides were used to help regulate the binding of zinc finger proteins or visa versa). Our strategy in these studies was to 1) pick a set of zinc finger proteins that had been carefully characterized, 2) design polyamides that would target sequences which overlap the binding sites of these proteins, and 3) carefully characterize the DNA-binding affinity for each set of compounds under identical buffer conditions. We then used a set of biochemical studies (involving gel mobility shifts and DNase I footprinting experiments) to see what happened when zinc fingers and polyamides, that recognized the same DNA site, were mixed together.

Our central conclusion from these biochemical studies is that polyamides interfere with the binding of zinc finger proteins when the two compounds recognize overlapping, or partially overlapping, binding sites on the minor and major groove sides of the DNA, respectively. This is clearly demonstrated by interference experiments that use the gel mobility shift assay to monitor zinc finger binding (Figures 5.7 and 5.8), and this result holds for every combination of zinc fingers and polyamides that we have tested, when they recognize overlapping or partially overlapping binding sites. It also is clear that this

"negative cooperativity" requires direct interactions of the polyamide with the zinc finger binding site: polyamides directed to other DNA sites show no interference with the binding of a given zinc finger protein (as indicated, for example, in the data of Figure 5.8B).

In an attempt to understand the structural and energetic basis for this "negative cooperativity," we have examined and compared crystal structures that are available for a set of zinc finger-DNA and polyamide-DNA complexes. In every case, structures of the relevant zinc finger-DNA complexes show that the proteins bind in the major groove, while structures of polyamide-DNA complexes show that these compounds bind in the minor groove. There does not appear to be any basis for a van der Waals collision (or any other direct contact) between the protein and the polyamide when they bind to overlapping sites. However, structural comparisons do show striking differences between the DNA conformations in the polyamide-DNA complexes and those in the zinc finger-DNA complexes. Differences in the groove dimensions are quite clear: 1) The minor groove tends to be wider in the polyamide-DNA complexes than in the zinc finger-DNA complexes (Figure 5.10); while 2) The major groove tends to be wider (and often deeper) in the zinc finger-DNA complexes than in the polyamide-DNA complexes. Quantitative comparison of the DNA structures (using output from the program Curves[38]) reveals many other differences between zinc finger-DNA and polyamide-DNA complexes, and many of these differences will affect the precise position and orientation of key hydrogen bond donors and acceptors within the DNA site. At this stage, it seems very plausible that differences in the preferred DNA conformations could explain the "negative cooperativity" that we have observed when zinc fingers and polyamides are targeted to the same site. We assume that 1) the free DNA is somewhat "plastic" and that each

compound can induce an appropriate conformation when it binds alone, but that 2) simultaneous binding is difficult because the two types of compounds prefer somewhat different DNA conformations. (Note: If the conformational preferences of a given DNA sequence were more rigid, it might be hard to design polyamides for one class of sites and it might be hard to design zinc fingers for another class of sites. This does not seem to be the primary problem.) In short, our modeling suggests that negative cooperativity may involve "allosteric" effects of changes in the DNA structure, and considering differences in the relative dimensions of the major and minor groove (Figure 5.10) makes it easy to picture why these changes may be so important.

We undertook this study to investigate how polyamides and zinc fingers interact when they bind to overlapping sites on double-stranded DNA and show here the first example of allosteric inhibition of a major groove binding protein by unmodified polyamides. One could envision that designed zinc finger proteins could be displaced by small molecule polyamides, thereby providing both an on and off switch for gene regulation. Furthermore, phage selection technology may allow for the generation of artificial zinc finger proteins that only bind their target DNA sites with high affinity in the presence of a bound polyamide in the minor groove of the site. Given the therapeutic potential of sequence-specific DNA targeting, the results of this study lend interesting insight to how a zinc finger/polyamide system could be designed to regulate gene expression.

Materials and Methods

Protein Production and Purification. The Zif268 zinc finger region (residues 349-421) was subcloned into a pET-21d expression vector (Novagen) and was

transformed into the *Escherichia coli* strain BL21(DE3) containing the pLysE plasmid. Cultures were grown and induced as described (Novagen). After the cells were harvested, they were lysed and sonicated as recommended (Novagen). The peptides were denatured and reduced in 6.4M guanidine·HCl, 150mM DTT, and 50mM Tris-HCl, pH 8.0 at 75 °C for 30 minutes and acidified to pH ~ 2.0 by addition of 10% trifluoroacetic acid (TFA). The peptides were purified by reverse-phase batch extraction on Sep-pack C-18 cartridges (Waters) as described,[29] followed by purification on a C4 reversed-phase (Vydac) high performance liquid chromatography column using a gradient of 22 to 35% acetonitrile (ACN) containing 0.1% TFA. The purified peptide fractions were then refolded anaerobically in the binding buffer that we used for gel shifts, after supplementing it with a 0.5 molar excess of $ZnCl_2$. Refolded peptides were stored at -80 °C in 10 µl aliquots; each aliquot was used once for binding studies and then discarded. The active concentrations of peptides were determined in stoichiometric competition experiments.

$TATA_{ZF}$, NRE_{ZF}, and $p53_{ZF}$ zinc finger peptides, selected from residues 333-414 or Zif268[24] used in these studies, were purified by Scot A. Wolfe, Robert Grant, and Sandra Fay-Richard, respectively. Since they were preparing these samples for crystallographic studies, they used several additional purification steps, essentially as described.[30]

Hairpin Polyamide Syntheses and Characterization. The pyrrole-imidazole hairpin polyamides **1-4** were prepared by manual and machine-assisted solid-phase methods[31] (Figures 5.2 and 5.3). Purity and identity of each compound were verified by a combination of analytical HPLC, ^1H NMR, and matrix-assisted laser desorption ionization time-of-flight mass-spectrometry (MALDI-TOF). Polyamides **1-3** have been

described previously[32.33] and polyamide **4**, PyIm-β-Im-$(R)^{H}{}_2{}^{N}$γ-PyImPyPy-β-Dp, is described here; UV (H_2O) λ_{max}304 (60800); MALDI-TOF-MS 1230.65 (1230.59 calc. for [M+H] $C_{55}H_{72}N_{23}O_{11}^{+}$). Lyophilized samples of polyamides were stored in at -80 °C. Polyamide concentrations were determined spectrophotometrically with extinction coefficients estimated based on the number of aromatic rings using the relation 8690 M^{-1} cm^{-1}/aromatic ring for the absorption maximum at 290-315 nM.[28]

Gel Mobility Shift Assay. Double-stranded oligonucleotides used for gel mobility shift assays were essentially identical to DNA sites used[9] when selecting the $TATA_{ZF}$, NRE_{ZF}, and $p53_{ZF}$ zinc finger proteins. A few nucleotides in the sequence flanking the zinc finger binding sites were also changed so that the zinc fingers and polyamides would have overlapping binding sites. The 27 bp DNA duplexes that were used for gel mobility shift assays are boxed in Figure 5.4, and the nine base pair zinc finger target sites near the center of each duplex is underlined. (Note: In the form shown in Figure 5.4, 27 bp duplexes are embedded within larger DNA segments that were later used for footprinting studies.) The oligonucleotide binding site for wt Zif268 was slightly different from that used by Greisman et al.[9] containing GCGGGGGCG rather than GCGTGGGCG. However, this construct was advantageous because it readily provided an overlapping polyamide binding site and should have no other effect. (Zif268 binds extremely well to either site and there is no direct contact with this base in the crystal structure[6].)

Each 27 bp strand of these binding sites was synthesized using standard phosphoramidite chemistry on an Applied Biosystems Model 392 DNA Synthesizer. Following deprotection, oligonucleotides were purified by denaturing polyacrylamide gel

electrophoresis. Duplexes were annealed and quantified, and then 0.3 pmol of each DNA fragment were end-labeled with [α-^{32}P] ATP (New England Biolabs), using Klenow exo$^-$ to fill-in the overhangs. Unincorporated nucleotides were removed by use of Sephadex G-25 quick spin columns (Boehringer Mannheim), and the DNA was resuspended in 1 ml of the buffer used for all gel shift experiments, which contained: 5 mM HEPES, pH 7.8; 50 mM KCl, 50 mM KGlu, 50 uM KoAc, 5 mM $MgCl_2$, 5% Glycerol, 0.1% NP-40, 20 uM $ZnSO_4$, 100 ug/ml BSA, 1 mM $CaCl_2$, and 1 mM DTT. The nucleotides were stored at –20 °C.

Using the buffer above, binding reactions and equilibrium binding studies for zinc finger proteins were performed in both the presence and absence of polyamides. Zinc finger dissociation constants were determined as previously described,[9] except that 0.5 pM of labeled duplex DNA was used and equilibration was done at room temperature for 16-18 hours for all gel shift experiments. The reaction mixtures were then subjected to 12% native polyacrylamide gel electrophoresis for about 2 hours with 0.5×TBE running buffer. Radioactive signals were quantitated by PhosphorImage analysis (Molecular Dynamics).

Equilibrium dissociation constants (K_Ds) were determined by linear regression using the Scatchard equation:

$$\Theta/[P] = 1 / K_D - \Theta/ K_D \qquad (1)$$

in which Θ equals the fraction of DNA bound ([PD]/([PD]+[D])), and [P] equals the free protein concentration (after applying corrections to account for the percent of the protein that was active for DNA binding).

DNase I Footprinting. Oligonucleotides used for footprinting experiments (Figure 5.4) were designed so that they included the sequences of the 27 bp duplexes that had been used for gel mobility shift assays, and additional 6 to 7 bp "spacers" were added so the protein-binding sites would be slightly further apart. Oligonucleotide 1 contains binding sites for $TATA_{ZF}$, NRE_{ZF}, and $p53_{ZF}$; oligonucleotide 2 contains binding sites for Zif268, $TATA_{ZF}$, and $p53_{ZF}$. The synthetic duplexes were designed such that the ends were ready for cloning into a pBluescript II SK(+) plasmid (Stratagene) that had been cut with *Eco*RI and *Bam*HI. After growth in *E. coli*, DNA probes for footprinting were prepared by digestion of the appropriate plasmids with *Eco*RI and *Not*I restriction enzymes, giving a 134 bp fragment for oligonucleotide 1 and a 143 bp fragment for oligonucleotide 2. After digestion, these DNA fragments were radioactively labeled using Klenow enzyme (New England Biolabs) and α-^{32}P labeled nucleotides to fill in the overhanging ends. The labeled oligonucleotides were purified using 5% native PAGE, were precipitated with EtOH, and were counted for specific activity.

Footprinting experiments were done essentially as previously described.[25-28] Quantitative footprinting for polyamides **1-4** were performed using the appropriate restriction fragment (oligonucleotide 1 or 2), in triplicate, under both conventional TKMC buffer: 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$ as well as the same binding buffer used in the gel shift experiments. Polyamide/DNA solutions were allowed to equilibrate at 22 °C for 16-18 hours. Footprinting reactions were initiated by the addition of the appropriate amount of DNase I to give ~50% intact DNA and allowed to proceed for 7 minutes at 22 °C. The reactions was stopped by addition of 50 µl of a solution containing 1.25 M NaCl, 100 mM EDTA, 0.2 mg/ml glycogen, and 28 µM base-pair calf thymus DNA and ethanol precipitated. The reactions

were resuspended in 1 × TBE/80% formamide loading buffer, denatured by heating at 80 °C for 10 minutes, and placed on ice. The reaction products were separated by electrophoresis. Radioactive signals were visualized with a Molecular Dynamics Typhoon Phosphorimager followed by quantitation using ImageQuant software (Molecular Dynamics). Equilibrium association constants for the polyamides were calculated as previously described[29] and found to be identical in both TKMC buffer and the gel shift buffering conditions.

Computer modeling experiments. Structures of polyamide-DNA complexes were taken from the protein databank[13-15] and were compared with known zinc finger-DNA complexes for $TATA_{ZF}$,[34] $p53_{ZF}$,[35] and Zif268.[6] The structures were aligned (using the PROTEUS Program[36]) via phosphorous atoms, C1' atoms, or common atoms in the set of superimposed base pairs. Several different alignment strategies were tested, including schemes in which: 1) the entire zinc finger DNA binding site was aligned with the polyamide DNA; 2) the polyamide binding sites in the known polyamide-DNA structures were aligned with the expected polyamide binding sites on the zinc finger DNA; 3) the GC or AT base pair in the polyamide/DNA structure was aligned with the corresponding base pair in the DNA/zinc finger structure; or 4) all the common atoms in a base pair were aligned. The RMS deviation was noted for each of the alignments, and a visual inspection of the alignment was made using Insight.[37]

As another strategy for comparison of the DNA structures, Curves[38] was run on all the polyamide-DNA complexes and all the zinc finger-DNA complexes. In addition, Curves also was run on B-DNA.[39] Comparisons of the output from Curves focused on the width and depth of the major and minor grooves and on base pair parameters such as buckle, shear, propeller twist, opening, and X and Y displacement. For major and minor

groove comparisons, the groove dimensions were taken from 1) the region of the polyamide-DNA complex closest to the polyamide binding site and 2) from the region on the zinc finger-DNA complex that was closest to the expected polyamide binding site present in our experiments. The average of the buckle, shear, propeller twist, and opening parameters were taken from the global base-base parameter output of Curves. Parameters for individual base pair were also compared separately.

Acknowledgements

$TATA_{ZF}$, NRE_{ZF} and $p53_{ZF}$ zinc finger peptides were overexpressed, purified, and submitted from Scot A. Wolfe, Robert Grant, and Sandra Fay-Richard, respectively. The pET-21d expression vector transformed in BL21(DE3) was submitted by Bryan S. Wang. We would like to thank Ezra Peisach for a critical reading of this manuscript, and we would like to thank members of the Pabo lab for their support and useful discussions throughout the experimental process and writing of this manuscript. We are grateful to the National Institutes of Health (GM 51747) for research support to P.B.D. and the Natural Sciences and Engineering Research Council of Canada for a postgraduate scholarship to D.H.N.

References

1. Rebar, E.J., Greisman, H.A., Pabo, C.O. *Methods in Enzymology* **1996**, *267,* 129-149.
2. Wolfe, S.A., Nekludova, L., Pabo, C.O. (2000) *Annual Reviews in Biophysical and Biomolecular Structure* **2000**, *29,* 183-212.
3. Dervan, P.B. *Bioorganic & Medicinal Chemistry* **2001**, 9, 2215-2235.
4. Dervan, P. B., Burli, R. W., *Current Opinion in Chemical Biology* **1999**, *3,* 688-693.
5. Pabo, C.O., Peisach, E., Grant, R., *Annual Reviews in Biochem*istry, in press.
6. Elrod-Erickson, M., Rauld, M.A., Nekludova, L., Pabo, C.O., *Structure* **1996**, *4,* 1171.
7. Pavletich, N. P., and Pabo, C. O., *Science* **1991**, *252,* 809-817.
8. Rebar, E. J. and Pabo, C. O. *Science* **1994**, *263,* 671-673.
9. Greisman, H. A. and Pabo, C. O. *Science* **1997**, *275,* 657-661.
10. Choo, Y., and Klug, A. *Proceedings of the National Academy of Sciences USA* 1994, *91,* 11168-11172.
11. Segal, D.J., Dreier, B., Beerli, R.R., and Barbas, C.F. *Proceedings of the National Academy of Sciences USA* **1999**, *96,* 2758-2763.
12. DeClairac, R. P. L., Geierstanger, B. N., Mrksich, M. Dervan, P. B., Wemmer, D. E., *Journal of the American Chemical Society* **1997**, *119,* 7909-7916.
13. Keilkopf, C. L., Baird, E. E., Dervan, P. B., Rees, D. C., *Nature Structural Biology* **1998**, *5,* 104-109.
14. Keilkopf, C. L., White, S., Szewczyk, Y. W., Turner, J. M., Baird, E. E., Dervan, P. B., Rees, D. C., *Science* **1998**, *282,* 111-115.

15. Kielkopf, C.L., Bermer, R.E., White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E., Dervan, P.B., Rees, D.C., *Journal of Molecular Biology* **2000**, *295,* 557-567.

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

17. Dickinson, L. A., Gulizia, R. J., Traunger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., Dervan, P. B., *Proceedings of the National Academy of Sciences USA* **1998**, *95,*12890-12895.

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

19. Dickinson, L. A., Trauger, J. W., Baird, E. E., Dervan, P. B., Graves, B. J., Gottesfeld, J. M., *Journal of Biological Chemistry* **1999**, *274,*12765-12773.

20. Nealy, L. Trauger, J. W., Baird, E. E., Dervan, P. B., Gottesfeld, J. M., *Journal of Molecular Biology* **1997**, *274,*439-445.

21. Oakley, M. G., Mrksich, M., Dervan, P. B., *Biochemistry* **1992**, *31,* 10969-10975.

22. Bremer R. E., Baird, E. E., Dervan, P. B., *Chemistry and Biology* **1998**, *5,* 119-133.

23. Winston, R.L., Ehley, J.A., Baird, E.E., Dervan, P.B., and Gottesfeld, J.M. *Biochemistry* **2000**, *39,* 9092-9098.

24. Wolfe, S. A., Greisman, H. A., Ramm, E. I., Pabo, C. O., *Journal of Molecular Biology* **1999**, *285,* 1917-1934.

25. Brenowitz, M., Senear, D.F., Shea, M.A., Ackers, G.K., *Methods in Enzymology* **1986**, *130*, 132-181.

26. Brenowitz, M., Senear, D.F., Shea, M.A., Ackers, G.K. *Proceedings of the National Academy of Sciences USA* **1986**, *83*, 8462-8466.

27. Senear, D.F., Brenowitz, M., Shea, M.A., Ackers, G.K., *Biochemistry* **1986**, *25*, 7344-7354.

28. Trauger, J.W., Dervan, P.B., *Methods in Enzymology* **2001**, *340*, 450-466.

29. Elrod-Erickson, M., and Pabo, C.O., *Journal of Biological Chemistry* **1999**, *274 (27),* 19281-19285.

30. Elrod-Erickson, M., Benson, T.E., Pabo, C.O., *Structure* **1998**, *6(4)*, 451-464.

31. Baird, E. E., Dervan, P. B., *Journal of the American Chemical Society* **1996**, *118,* 6141-6146.

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

33. Swalley, S.E., Baird, E.E., Dervan, P.B., (1996) *Journal of the American Chemical Society* **1996**, *118*, 8198-8206.

34. Wolfe, S.A., Grant, R., Pabo, C.O, *Structure*, in press.

35. Grant, Robert, Richards, S.F., Pabo, C.O., personal communication.

36. Pabo, C.O., and Nekludova, L., *Journal of Molecular Biology* **2000**, *301,* 597-624.

37. MSI Insight II Program

38. Stofer, E., and Lavery, R., *Biopolymers* **1994**, *34,* 337-346.

39. Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K., Dickerson, R.E., *Proceedings of the National Academy of Sciences USA 78,* 2179-2183.