Inhibition of DNA Major Groove Binding Proteins by Hairpin Polyamides

Thesis by

Ryan Elwood Bremer

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2000

(Submitted April 17, 2000)

To Lis

and to my family

Acknowledgments

I have had the great fortune of having many wonderful teachers who deserve much of the credit for guiding me to this point. Mrs. Pettaway was the first to show me chemistry. Ron Henke and Kim Graves always kept me excited and I thank them for truly combining science and fun. Luther Erickson, Lee Sharpe, Leslie Lyons, Kevin Morris and Mary Mader were excellent teachers and mentors who always challenged me to do better. Most importantly, I thank Jim Swartz for teaching how to do successful research and showing me a balance of thoughts that I will always practice. Jon Boeckenstedt helped send me down this path with his few simple words of advice. Peter Dervan has been nothing less than an outstanding mentor. He has provided invaluable freedom, support and insight throughout my tenure at Caltech.

I would like to thank the various members of my committee; Prof. Doug Rees, Prof. Steve Mayo, Prof. David Tirrell and Prof. Barbara Imperialli for their time and advice during my graduate career. The staff at Caltech, Cora Carriedo, Steve Gould, Debby Miles, Margot Hoyt, Dian Buchness, Gary Hathaway and his staff, and Elly Noe all made it tremendously easier to get things done.

Many of the students and postdocs I have shared time and space with here have become great friends and colleagues. I thank Ramesh Baliga and Matt Taylor for not killing the stupid first-year; a trend that I hope Matt continues. I thank Jason Szewczyk for advice and training and many lunches and late night conversations, some more memorable than others. The quality of work John Trauger produced left us all in awe and striving to do better. Aileen Chang has become a good friend and I hope that we all might better follow her example. I can always count on David Herman for both a fun time and stimulating science. Clay Wang has been tireless and his commitment unwavering. Nick Wurtz and Doan Nguyen have become good friends and valued colleagues and I hope a short plane ride will not keep us apart. Clara Kielkopf is the best collaborator one could ask for and she always provides insightful new ideas. I thank Anna Mapp and Tom

Minehan for the countless times they have graciously tolerated my innane questions. I only hope I can remember what they taught me. Ulf Ellervik has been a great labmate and if his students share half his enthusiasm, he has a very exciting career ahead. Jason Belitsky always provides the terse dry comment that distills an issue to its humorous core. Most importantly, the friendship that Sarah White and Eldon Baird have provided has made my time at Caltech exceed my greatest expectations. I cannot thank Eldon enough for the hours he has spent with me over the years selflessly teaching me all matters of science; always looking out for my best interests. I will always marvel at his ability to find the silver lining in every cloud. I will probably never deserve the patience and guidance he provides me, but I hope that in some small part I can make it worthwhile in the years to come.

My parents have always been wonderful supporters of me. I thank them for finding a way to get it all done. I appreciate my family's understanding as my life takes me geographically away from them, but I hope that they know that I will always need them.

Even in the presence of such good fortunes as these, I have no doubt that I would not have completed graduate school intact if it were not for the love and unconditional support Lis has provided. She makes all of the other blessings of my life seem small. I only hope that in the years ahead I am able to be the person she deserves. In many ways, this thesis is largely hers. Thank you.

Abstract

Small molecules that bind to any predetermined DNA sequence in the human genome are potentially useful tools for molecular biology and human medicine. Polyamides containing N-methylimidazole (Im) and N-methylpyrrole (Py) are cell permeable small molecules that bind DNA according to a set of "pairing rules" with affinities and specificites similar to many naturally occurring DNA binding proteins. Im/Py polyamides offer a general approach to the chemical regulation of gene expression, provided inhibition of DNA binding for a variety of transcription factor families can be achieved. Polyamides bound in the minor groove have been shown to co-occupy the DNA helix with proteins in the major groove. We demonstrate here that polyamides containing a positively charged moiety directed to the DNA backbone can effectively inhibit DNA binding by an exclusively major groove protein, potentially by competing with the positively charged protein side chains for contacts to the negatively charged phosphate backbone. The requisite positive patch can be achieved with a naturally derived C-terminal Arg-Pro-Arg tripeptide (Chapter 2) or a simple synthetic diaminoalkyl chain delivered from the N-1 of a single pyrrole residue (Chapter 3). The functional repertoire of polyamides as synthetic ligands for the control of transcription factor binding has been expanded to include proteins which bind exclusively in the major groove of DNA. The broad targetable sequence repertoire of polyamides, coupled with the ubiquity of backbone contacts in protein recognition of DNA, make phosphate neutralization by a positive patch a promising approach for inhibition of major groove transcription factors. Other investigations into polyamide:DNA recognition have afforded N-aminoalkylpyrrole-containing polyamides that offer enhanced affinity without compromising specificity (Chapter 3), desmethylpyrrole-containing polyamides that increase water solubility while retaining subnanomolar affinity (Chapter 4), and structural insight into the lower affinities observed with Hp-containing polyamides (Chapter 5).

Table of Contents

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 1

Introduction

Background

Regulation of Gene Expression

In nearly every human cell, the genetic information is stored in three billion base pairs of DNA organized as 80,000 to 100,000 genes (*1*). The DNA sequence for each of these genes is the code for a corresponding protein that will be synthesized from the twenty naturally occurring amino acids and collectively is the blueprint by which a human being is constructed. The simple chain of only four bases, adenine, thymine, guanine, and cytosine, contains all of the information required to sustain a human being through an entire lifetime from infancy to maturity over a span of several decades and through countless challenges from the outside environment.

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 that requires it to express certain genes while leaving others silent (*2*). The regulated expression of specific genes is what allows, for example, a neuron and a red blood cell to start with the same DNA but to develop into very different products that cooperate with other cells to form tissue, organs and complex systems. At the molecular level, Nature accomplishes this complex task of gene regulation through a myriad of weak noncovalent interactions between DNA and proteins that are continuing to be understood. Typically, a gene is organized as an upstream region of DNA that contains short DNA sequences that serve as recognition sites for the proteins responsible for recruiting the catalytic machinery necessary to transcribe the adjacent coding region from DNA to RNA so that it can be translated to a protein chain at the ribosome (*3*). In order for a gene to be actively transcribed, a noncovalent network of more than fifty proteins with a combined molecular weight of greater than two million must assemble on the DNA (Figure 1.1) (*2*). These proteins consist of the general transcription factors that are common to many genes and responsible for many functions including releasing the double helix from chromatin and separating the DNA strands. RNA polymerase is the enzyme responsible for catalyzing the addition of each base to the

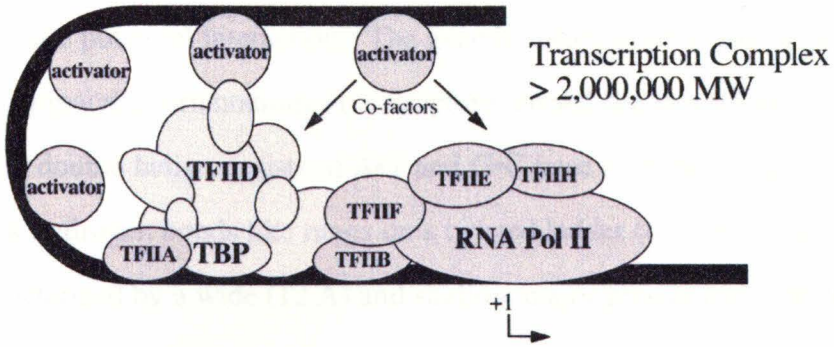


Figure 1.1. Model of protein regulation of gene transcription. A combinatorial assembly of transcription factors (> 2,000,000 MW) is required for initiation of gene-transcription. Inhibition of transcription factor binding by small molecule DNA-ligands provides a potential approach for the control of gene transcription in living cells.

growing RNA strand on the DNA template. Activator proteins that typically contain a domain responsible for DNA binding and a separate activation domain likely play a more specific role in recruiting the general machinery for transcription. Of the hundreds of known activators, a specific subset of several activators bind in a combinatorial fashion to the promoter region of a gene and recruit general transcription factors (*3*). Efficient transcription of a gene occurs only when the activators responsible for its expression assemble in concert. With many activators being under molecular control, either through their own expression or another signaling pathway, regulated expression of the desired genes is achieved.

Artificial Ligands for the Regulation of Gene Expression

Misregulation of gene expression can frequently result in a life-threatening disease (*3*). Altered activity of activator or repressor proteins either through their own altered expression, chromosomal translocations, or a flawed signal transduction pathway, may leave essential genes silent and/or result in undesired expression of genes with deleterious effects when turned on at this specific time in the given cell type. Artificial ligands that can restore the natural patterns of gene expression by sequence specifically binding DNA and interfering with DNA-binding proteins would be potentially valuable tools in human medicine (*4*).

As a target for noncovalent recognition, DNA is a rich source of repetitive and sequence specific potential interactions. The genetic information is stored on two strands of DNA (in antiparallel-orientation) in a structure termed the double helix (Figure 1.2) (*5*). The DNA double helix consists of A•T and G•C base pairs held together by specific Watson-Crick hydrogen bonds like rungs on a twisted ladder (*6*). The common B-form of DNA is characterized by a wide (12 Å) and shallow major groove and a deep and narrow (4-6 Å) minor groove. Individual sequences may be distinguished by the pattern of hydrogen bond donors and acceptors displayed on the edges of the base pairs (Figure 1.3) (*7*). In the minor groove, the A,T base pair presents two symmetrically placed hydrogen bond acceptors in the minor groove, the purine N3 and the pyrimidine O2 atoms. The G,C base pair presents these two acceptors, but in addition presents a hydrogen bond donor, the 2-amino group of guanine (*8*).

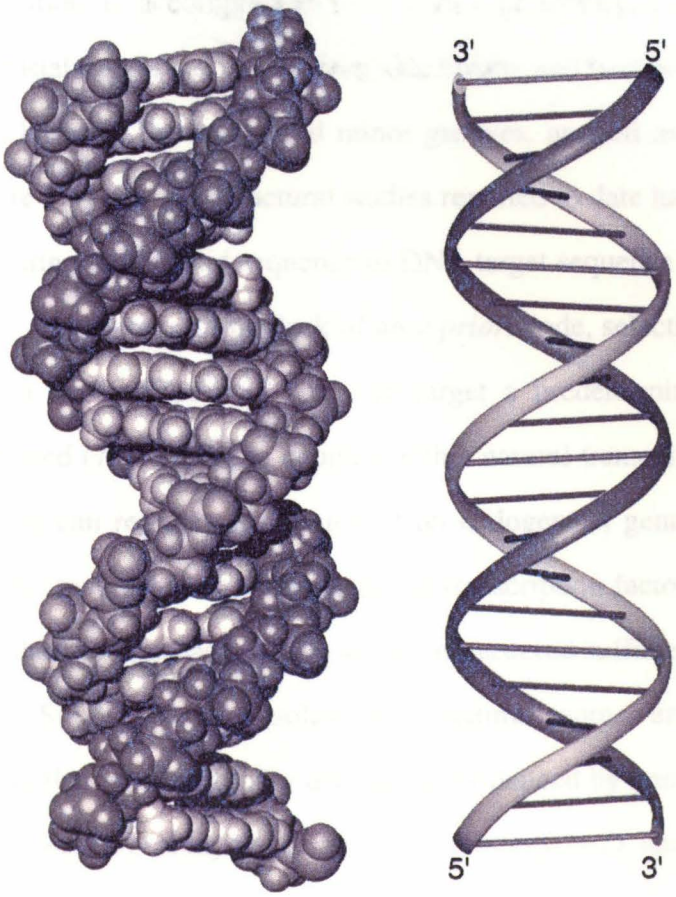


Figure 1.2. B-form double helical DNA. Antiparallel strands are indicated in dark and light gray. (left) space filling CPK model, (right) ribbon representation.

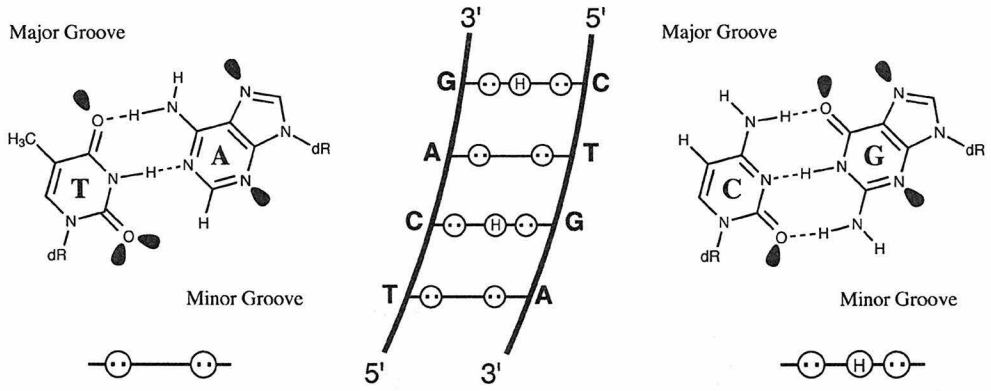


Figure 1.3. A schematic model for recognition of the minor groove, with hydrogen bond donors represented as (H) and hydrogen bond acceptors represented as two dots. This schematic underscores the potential coded reading of the DNA helix.

In pursuit of the goal to develop a general code for the recognition of predetermined DNA sequences with synthetic ligands, one can first look at the recognition of specific DNA sequences by the DNA-binding proteins themselves. Protein:DNA recognition is accomplished by a series of hydrogen bonds and van der Waals and electrostatic interactions between side chains and backbone atoms of the protein and the DNA bases in the major and minor grooves, as well as the sugar-phosphate backbone (Figure 1.4) (*9*). The structural studies reported to date have revealed that no simple code correlating amino acid sequence to DNA target sequence exists in Nature (*9*).

In response to the lack of an *a priori* code, selection methods to identify proteins from a randomized pool that can target a predetermined DNA sequence have been developed (*10-12*). When coupled with a natural transcriptional activating domain, these proteins can regulate expression of an endogenous gene in cell culture (*13*). However, efficient incorporation of the designed transcription factor and other obstacles inherent in gene therapy approaches have yet to be mastered sufficiently for therapeutic practice.

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 (Figure1. 5) (*14-17*). While these molecules can recognize short sequences of DNA through both

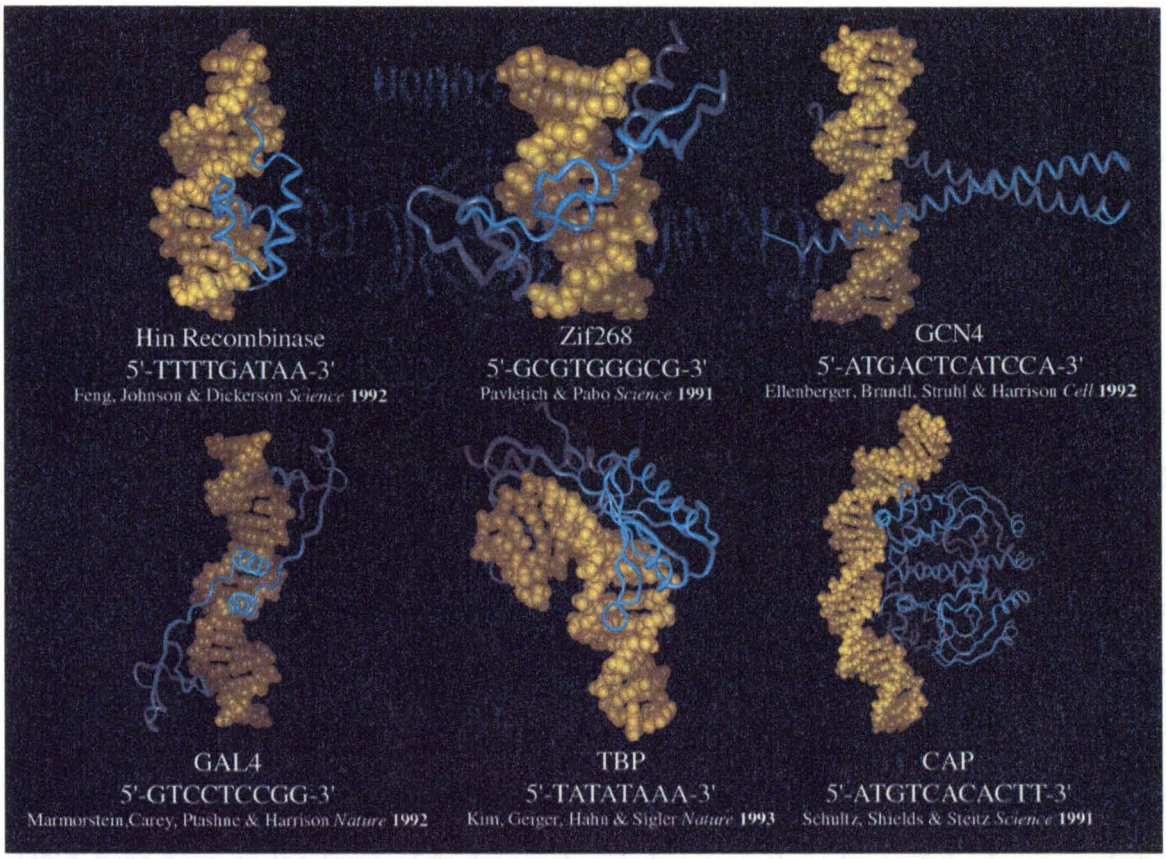
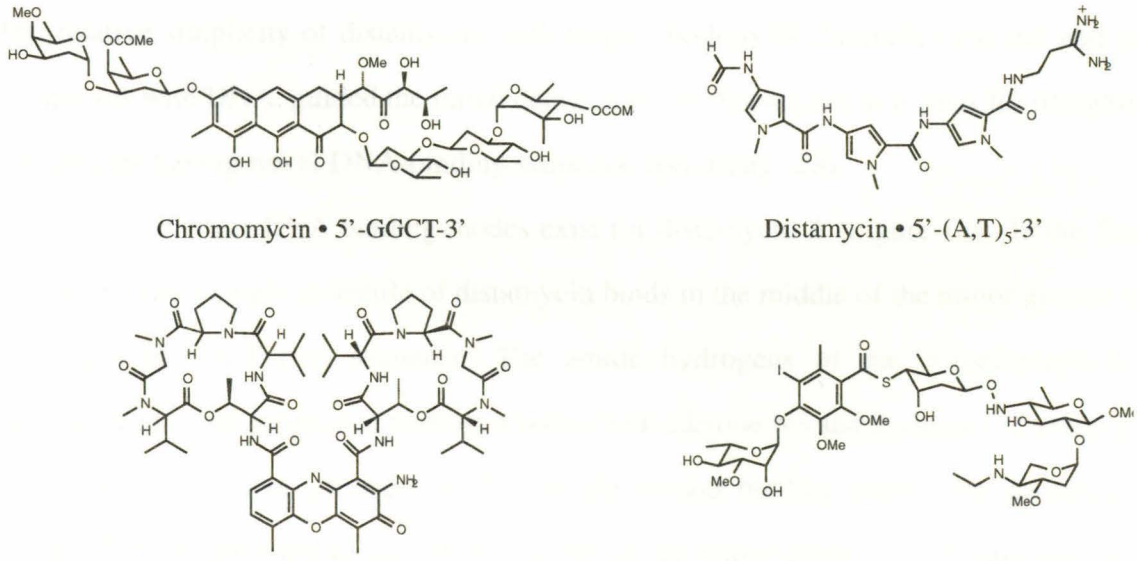


Figure 1.4. Representative X-ray crystal structures from several families of DNA binding proteins. Proteins can contact the DNA via interactions with the bases in the major groove and/or minor groove, as well as the sugar-phosphate backbone.



Chromomycin • 5'-GGCT-3'

Distamycin • 5'-$(A,T)_5$-3'

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

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 recognition of a broad range of predetermined DNA sequences (*18, 19*). Similarly, oligodeoxynucleotides which recognize the major groove of double helical DNA via triple helix formation bind a broad sequence repertoire with high affinity and specificity (*20, 21*). 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 (*22-24*).

Among these natural DNA-binding molecules, distamycin is distinguished by its structural simplicity, having no chiral centers and an oligopyrrolecarboxamide core structure (*25-27*). Structural studies of distamycin-DNA complexes reveal modular complexes in which adjacent pyrrole carboxamides make similar contacts with adjacent DNA base pairs in the form of hydrogen bonds between the edges of the bases and the amide bonds of distamycin, as well as van der Waals contacts with the walls of the minor groove, as the crescent shape of the ligand follows the curvature of the DNA helix (*14*). The relative simplicity of distamycin, with respect both to its chemical structure and its complexes with DNA, guided the initial decision to use distamycin as a basis for designed polyamides having novel DNA-binding sequence specificity (*28*).

Two distinct DNA binding modes exist for distamycin A (Figure 1.6). In the first binding mode, a single molecule of distamycin binds in the middle of the minor groove of a 5 base pair A,T rich sequence. The amide hydrogens of the *N*-methylpyrrole-carboxamides form bifurcated hydrogen bonds with adenine N3 and thymine O2 atoms on the floor of the minor groove (*14, 29*). In the second binding mode, two distamycin ligands form an antiparallel side-by-side dimer in the minor groove of a 5 base pair A,T rich site (*30, 31*). In the 2:1 model each polyamide subunit forms hydrogen bonds to a unique DNA strand in the minor groove.

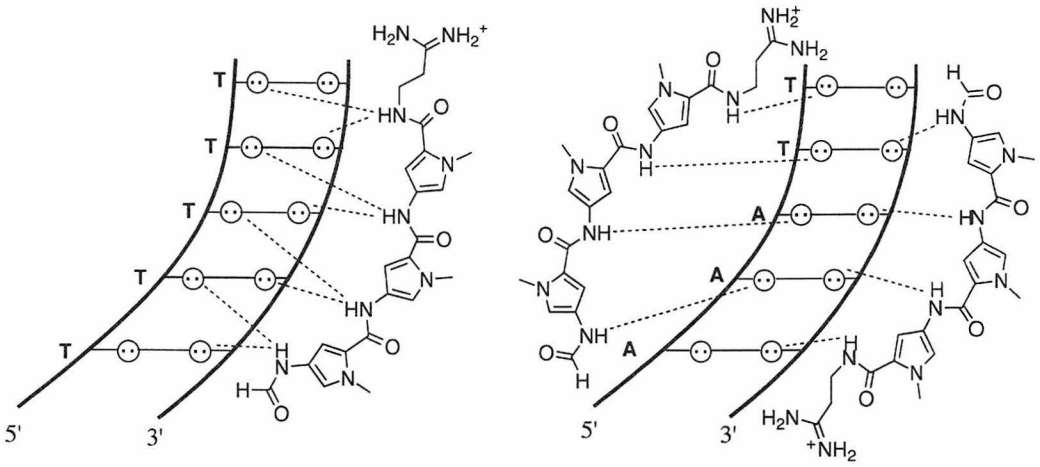


Figure 1.6. A schematic representation of recognition of A,T rich sequences in the minor groove by 1:1 and 2:1 complexes of Distamycin.

A twenty year search for a general DNA-recognition code based on minor groove binding analogs of distamycin, led by Prof. Peter B. Dervan of Caltech, has culminated in the development of pairing rules to guide polyamide design for recognition of double helical DNA (*32*). These rules correlate each Watson-Crick base pair with an anitparallel pairing of *N*-methylimidazole (Im), *N*-methylpyrrole (Py), and *N*-methyl-3-hydroxypyrrole (Hp), aromatic amino acids in side-by-side subunits for sequence specific recognition of the minor groove (Figure 1.7) (*33, 34*). A pairing of Im opposite Py targets a G•C base pair while a pairing of Py opposite Im targets a C•G base pair, both over A,T base pairs (*33, 35*). Specificity for G,C base pairs results from the formation of a hydrogen bond between the imidazole N3 and the exocyclic amine group of guanine (*36, 37*). A Py/Py combination is degenerate targeting both A•T and T•A base pairs, in preference to G,C base pairs (*30, 31, 33, 35, 38*). The use of Hp/Py and Py/Hp pairs for discrimination of T•A and A•T base pairs, respectively, completes the code (*34, 39*). Specificity for A,T base pairs is accomplished by a hydrogen bond between the 3-hydroxy of Hp and the second lone pair on the O2 of thymine, as well as a steric destabilization if Hp were placed over an adenine (*40, 41*). Validity of the pairing rules is supported by a variety of footprinting, NMR, and X-ray structure studies (*32, 33, 36, 37, 40-42*).

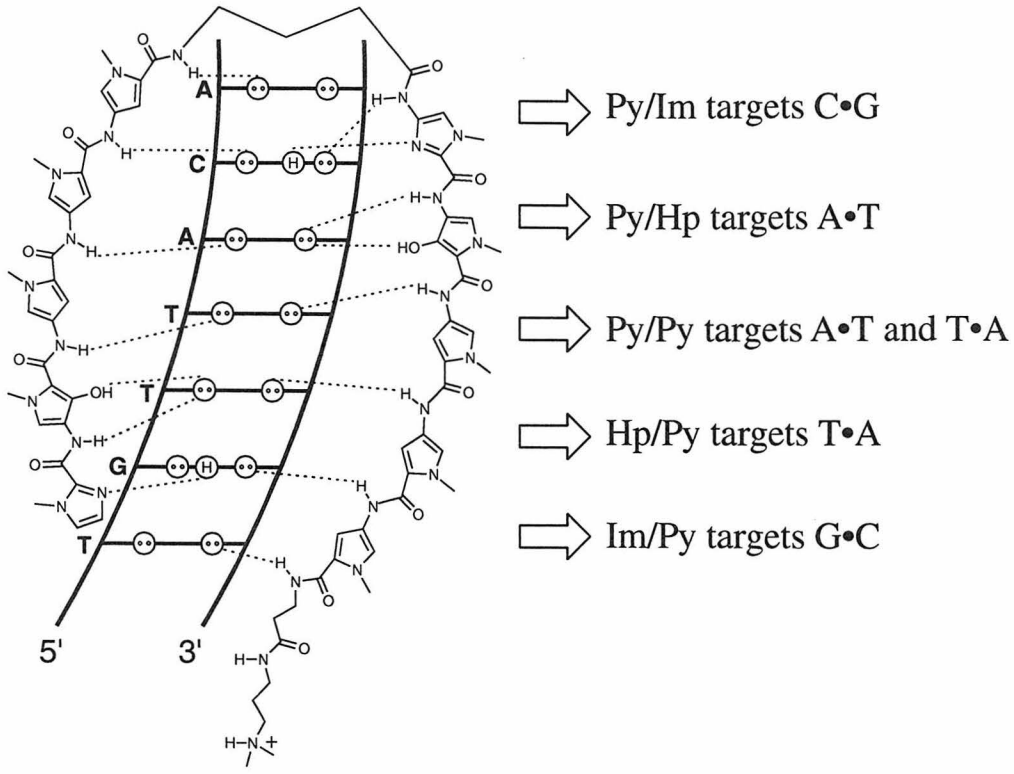


Figure 1.7. A schematic representation of the polyamide pairing rules.

Concurrently with explorations into the generality of the pairing rules, efforts have been underway in the Dervan group to increase polyamide affinity and specificity for DNA recognition and to expand their utility as synthetic regulators of gene expression. Linking the antiparallel polyamide subunits with an aliphatic γ-aminobutyric acid linker (γ) afford "hairpin" polyamides with enhanced affinity relative to the unlinked dimers (*43*). The development of solid phase synthesis protocols has dramatically reduced the time required for polyamide synthesis and made available the increased quantities necessary for cell culture studies, as well as allowed for the preparation of polyamide analogs and conjugates that expand the functional repertoire of this class of molecules (*44*). A C-terminal β-alanine (β) residue increases polyamide affinity and specificity and facilitates the solid phase synthesis (*44, 45*). The γ and β units are specific for A,T base pairs (*46*). The polyamide subunits recognize each DNA strand in a preferred orientation of N-C relative to the 5'-3' direction of the DNA strand (*47*).

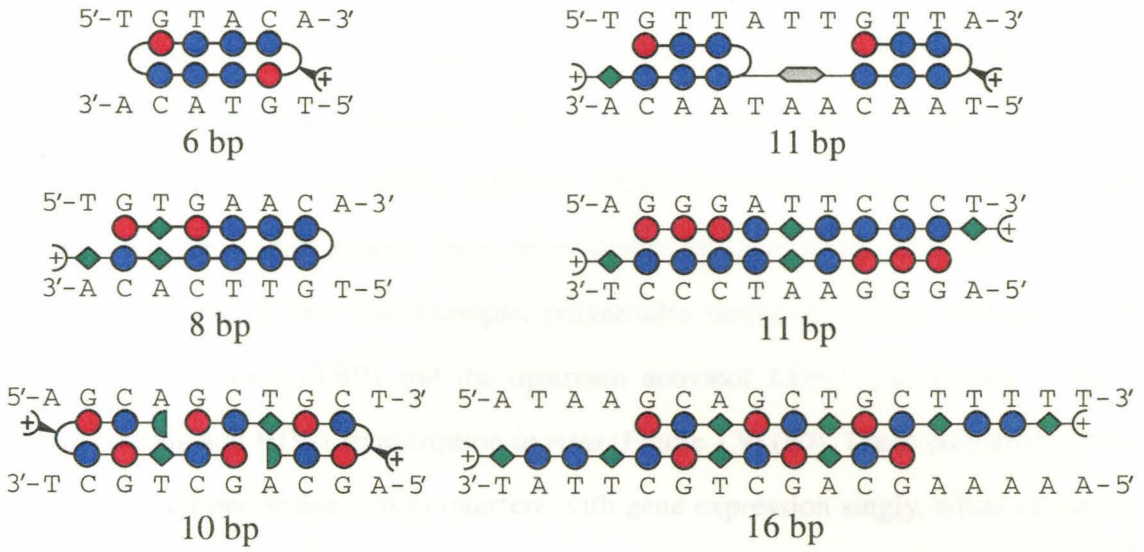


Figure 1.8. Representative motifs for polyamide:DNA recognition. Red and blue circles represent imidazole and pyrrole residues, respectively. β-Alanine and the γ-turn are depicted as green diamonds and curved lines, respectively. The plus sign represents the dimethylaminopropylamide tail or the α-amino-γ-turn.

A variety of motifs for sequence specific recognition using the polyamide scaffold have been successfully developed for target sites from 5-16 bp (Figure 1.8) (*48-57*). Increasing hairpin polyamide target size up to 7 bp can increase polyamide affinity (*49*). For longer sequences, the introduction of a flexible β-alanine allows the polyamide that would otherwise be overwrought relative to the DNA helix to track the minor groove (*37, 53, 54, 58, 59*). Selective incorporation of β-alanine can also enhance polyamide affinity for some target sequences (*58, 60*). Cooperative 2:1 dimer motifs are particularly attractive for future applications due to the low molecular weight of the monomer units (*51*). Target sites up to 16 bp, potentially sufficient for recognition of a unique sequence in the human genome, have also been achieved (*52*). Closing the ends of a hairpin polyamide with a second γ affords the cycle motif which can increase poylamide affinity (*56, 57*). The incorporation of a stereospecific α-amino group into the γ turn also increases polyamide affinity and specificity, while allowing for functional moieties to be linked or the construction of a tandem motif by linking the C-terminus of another polyamide to this position (*55, 61*).

Polyamide motifs that recognize their cognate DNA sites with subnanomolar affinities and specificities comparable to DNA binding proteins have allowed the potential of polyamides as artificial regulators of gene expression to be explored (*4, 48, 60*). Hairpin polyamides targeted to the DNA sequence where transcription factors make known contacts with the minor groove have been shown to specifically interfere with gene expression in cell culture. For example, polyamides designed to bind adjacent to the TATA-binding protein (TBP) and the upstream activator LEF-1 specifically inhibited protein binding and HIV-1 transcription *in vitro* (Figure 1.9) (*60*). These polyamides were found to be cell permeable and to interfere with gene expression singly, while in concert they lowered HIV-1 replication beyond detectable levels.

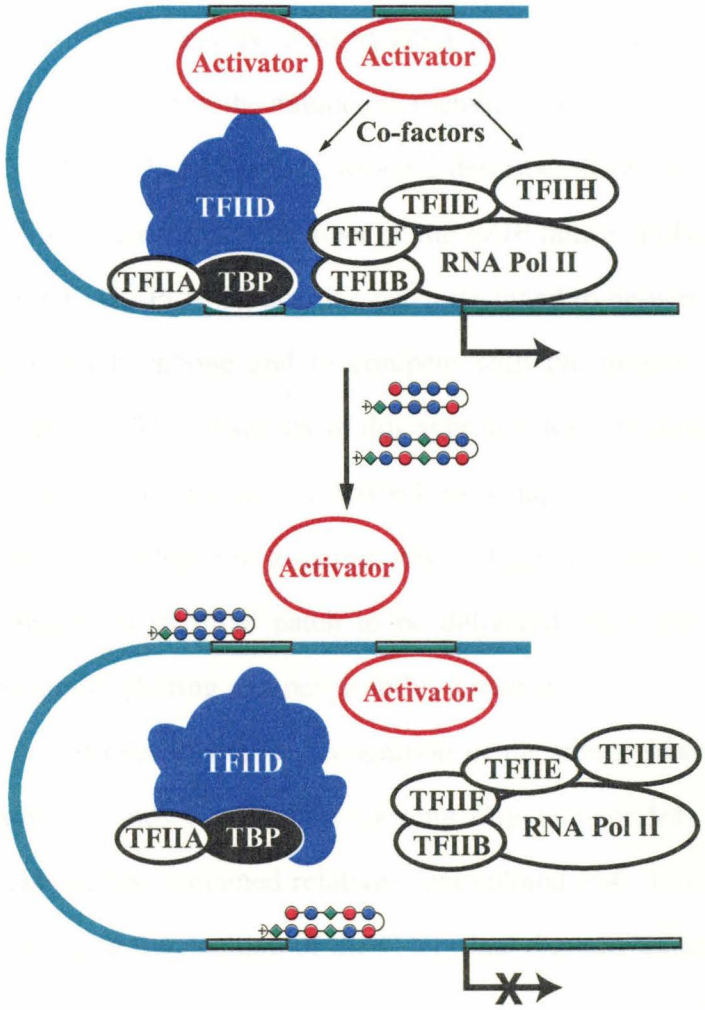


Figure 1.9. Inhibition of gene transcription by hairpin polyamides. A pair of hairpin polyamides targeted to the DNA sequences adjacent to the binding sites for LEF-1 and TBP inhibit assembly of the transcriptional machinery and transcription of the HIV-1 gene in human cell culture.

Scope of this Work

Inhibition of Protein Binding in the DNA Major Groove by Hairpin Polyamides

In contrast to the success of using polyamides to inhibit DNA binding by proteins that make contacts in the minor groove, some classes of transcription factors contact the DNA major groove exclusively (Figure 1.4) (*9*). X-Ray crystallography studies reveal that DNA bound by a 4-ring homodimeric polyamide is unaltered from its natural B-form structure, with all polyamide/DNA contacts confined to the minor groove (Figure 1.10) (*37, 40, 41*). Not surprisingly, polyamides have been found to bind simultaneously with ligands which exclusively occupy the major groove (*62-64*). The ubiquity of major groove contacts in protein-DNA recognition provides the impetus to develop approaches for inhibition of major groove proteins by polyamides which bind in the minor groove (Figure 1.4).

The negatively charged DNA phosphate backbone provides a target for ligands designed to disrupt the unique microenvironment which represents a protein's binding site on the DNA double helix. Chapter 2 describes efforts to inhibit the major groove protein, GCN4, a prototypical member of the bZIP family. Polyamides were modified with a C-terminal Arg-Pro-Arg positive patch designed to deliver a neutralizing positive charge to the DNA backbone and to compete with the protein side chains for contacts to the phosphates. The versatility of this approach was expanded with the N-aminoalkylpyrrole-containing polyamides described in Chapter 3. The α-amino acid tripeptide was essentially reduced to a diaminoalkyl chain delivered from the N-1 of apyrrole residue, allowing the positive patch to be delivered from any position of the polyamide and potentially altering cell permeability properties.

While polyamide recognition of the minor groove has been coupled with major groove ligands (i.e., triplex-forming oligonucleotides), interaction with the phosphate backbone has remained relatively unexplored (*64*). In Chapter 3, it is demonstrated that exchanging the position of the C-terminal dimethylaminopropyl tail and the N-methyl of a single pyrrole residue, to provide a structural isomer of a typical polyamide, affords a

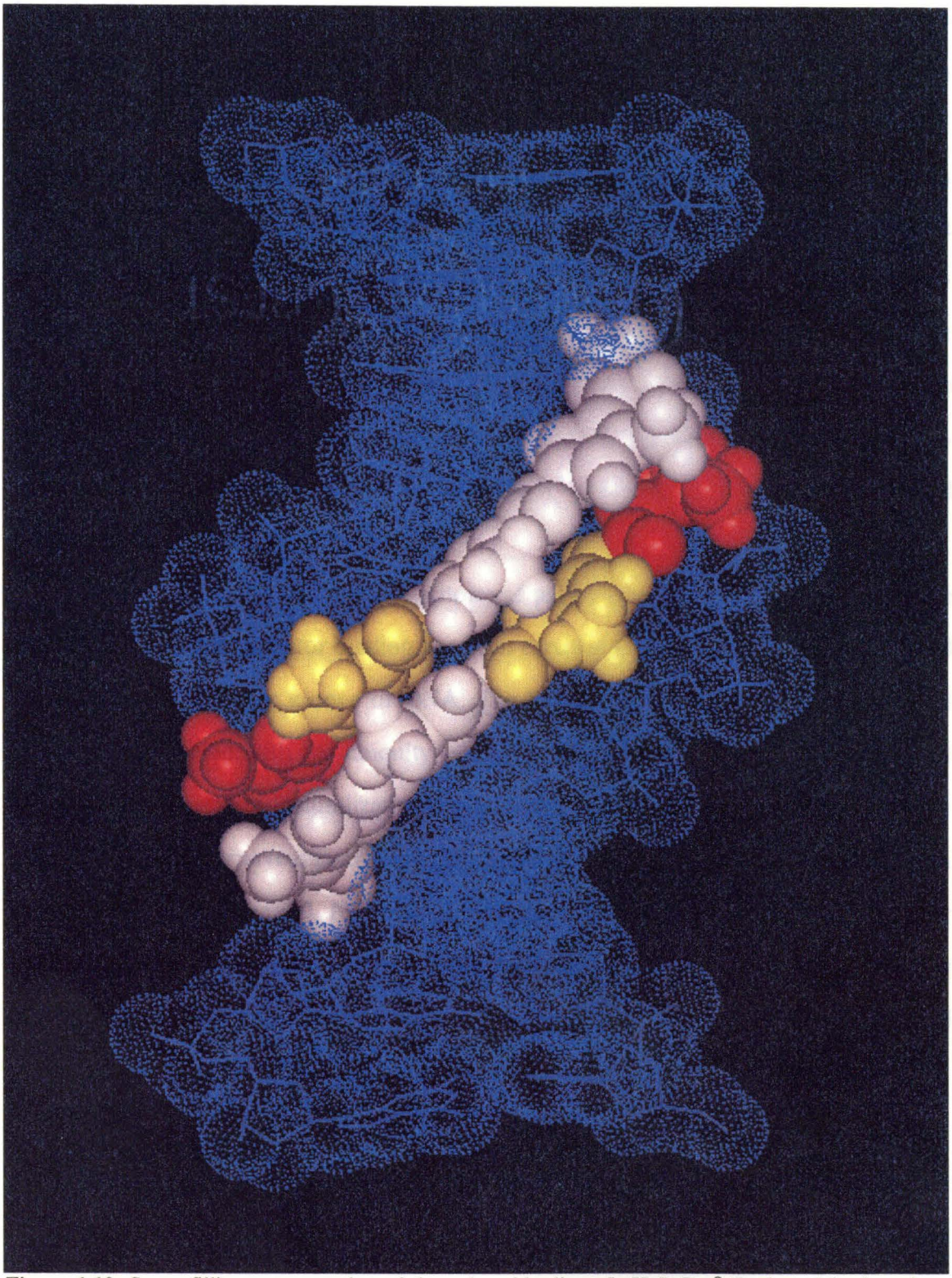


Figure 1.10. Space filling representation of the polyamide dimer ImHpPyPy-β-Dp bound in the minor groove of DNA. The figure was prepared using InsightII software and is derived from a high-resolution X-ray co-crystal structure of a polyamide dimer bound to DNA which was obtained in collaboration with the Rees group at the California Institute of Technology.

ligand with a 10-fold increase in affinity and uncompromised specificity. While a nonspecific electrostatic mechanism cannot be ruled out, the observed affinity enhancement may be the result of a specific interaction of the positively charged *N*, *N*-dimethylaminopropyl moiety with the phosphate backbone. The synthetic methodology developed for this class of molecules may prove useful in future applications. The investigation of the role of substituents at the N-1 of pyrrole in DNA recognition was continued in Chapter 4 with polyamides containing desmthylpyrrole residues. The absence of the N-methyl group resulted in no effect on affinity, but a 10-fold loss in specificity for this class of polyamides. Finally, the loss in polyamide affinity generally observed upon Hp introduction was investigated by comparing polyamide affinities and X-ray crystal structures in Chapter 5.

In this work, the functional repertoire of polyamides as synthetic ligands for the control of transcription factor binding has been expanded to include proteins which bind exclusively in the major groove of DNA. The broad targetable sequence repertoire of polyamides, coupled with the ubiquity of backbone contacts in protein recognition of DNA, make phosphate neutralization by a positive patch a promising approach for inhibition of major groove transcription factors. Other investigations into polyamide:DNA recognition have afforded N-aminoalkylpyrrole-containing polyamides that offer enhanced affinity without compromising specificity, desmethylpyrrole-containing polyamides that increase water solubility while retaining subnanomolar affinity, and structural insight into the lower affinities observed with Hp-containing polyamides.

References

1. Watson, J. D. (1993) *Gene 135*, 309.

2. Roeder, R. G. (1996) *Trends Biocehm. Sci. 9*, 327.

3. Tjian, R. (1995) *Sci. Am. 2*, 54.

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

5. Watson, J. D., and Crick, F. H. C. (1953) *Nature 171*, 737.

6. Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, M., Fratini, A. V., and Kopka, M. L. (1982) *Science 216*, 475.

7. Saenger, W. (1984) *Princliples of Nucleic Acid Structure*, Springer-Verlag, New York.

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

9. Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053-1095.

10. Greisman, H. A., and Pabo, C. O. (1997) *Science 275*, 657-661.

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

12. Segal, D. J., and Barbas, C. F. (2000) *Curr. Opin. Chem. Biol. 4*, 34-39.

13. Beerli, R. R., Dreier, B., and Barbas, C. F. (2000) *Proc. Natl. Acad. Sci. USA 97*, 1495-1500.

14. Coll, M., Frederick, C. A., Wang, A. H. J., and Rich, A. (1987) *Proc. Natl. Acad. Sci. USA 84*, 8385-8389.

15. Gao, X. L., Mirau, P., and Patel, D. J. (1992) *J. Mol. Biol. 223*, 259-279.

16. Kamitori, S., and Takusagawa, F. (1992) *J. Mol. Biol. 225*, 445-456.

17. Paloma, L. G., Smith, J. A., Chazin, W. J., and Nicolaou, K. C. (1994) *J. Am. Chem. Soc. 116*, 3697-3708.

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

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

940-944.

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

21. Thuong, N. T., and Helene, C. (1993) *Angew. Chem.-Int. Edit. Engl. 32*, 666-690.

22. Duvalvalentin, G., Thuong, N. T., and Helene, C. (1992) *Proc. Natl. Acad. Sci. USA 89*, 504-508.

23. Maher, L. J., Dervan, P. B., and Wold, B. (1992) *Biochemistry 31*, 70-81.

24. Nielsen, P. E. (1997) *Chem.-Eur. J. 3*, 505-508.

25. Arcamone, F., Penco, S., Prezzi, P. G., Nicolella, V., and Pirelli, A. (1964) *Nature 203*, 1064.

26. Baguley, B. C. (1982) *Mol. Cell Biochem. 43*, 167.

27. Zimmer, C. (1975) *Prog. Nucleic Acids Res. Mol. Biol. 15*, 285.

28. Dervan, P. B. (1986) *Science 232*, 464-471.

29. Pelton, J. G., and Wemmer, D. E. (1988) *Biochemistry 27*, 8088-8096.

30. Chen, X., Ramakrishnan, B., Rao, S. T., and Sundaralingam, M. (1994) *Nat. Struct. Biol. 1*, 169-175.

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

32. Dervan, P. B., and Bürli, R. W. (1999) *Curr. Opin. Chem. Biol. 3*, 688-693.

33. White, S., Baird, E. E., and Dervan, P. B. (1997) *Chem. Biol. 4*, 569-578.

34. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. (1998) *Nature 391*, 468-471.

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

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

37. Kielkopf, C. L., Baird, E. E., Dervan, P. D., and Rees, D. C. (1998) *Nat. Struct. Biol. 5*, 104-109.

38. White, S., Baird, E. E., and Dervan, P. B. (1996) *Biochemistry 35*, 12532-12537.

39. White, S., Turner, J. M., Szewczyk, J. W., Baird, E. E., and Dervan, P. B. (1999) *J. Am. Chem. Soc. 121*, 260-261.

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

41. Kielkopf, C. L., Bremer, R. E., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. (2000) *J. Mol. Biol. 295*, 557-567.

42. deClairac, R. P. L., Geierstanger, B. H., Mrksich, M., Dervan, P. B., and Wemmer, D. E. (1997) *J. Am. Chem. Soc. 119*, 7909-7916.

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

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

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

46. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1999) *J. Am. Chem. Soc. 121*, 1113-1120.

47. White, S., Baird, E. E., and Dervan, P. B. (1997) *J. Am. Chem. Soc. 119*, 8756-8765.

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

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

50. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1996) *Chem. Biol. 3*, 369-377.

51. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1998) *Angew. Chem.-Int. Edit. Engl. 37*, 1421-1423.

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

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

54. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) *Chem.-Eur. J. 3*, 1600-1607.

55. Herman, D. M., Baird, E. E., and Dervan, P. B. (1999) *Chem.-Eur. J. 5*, 975-983.

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

57. Herman, D. M., Turner, J. M., Baird, E. E., and Dervan, P. B. (1999) *J. Am. Chem. Soc. 121*, 1121-1129.

58. Turner, J. M., Swalley, S. E., Baird, E. E., and Dervan, P. B. (1998) *J. Am. Chem. Soc. 120*, 6219-6226.

59. deClairac, R. P. L., Seel, C. J., Geierstanger, B. H., Mrksich, M., Baird, E. E., Dervan, P. B., and Wemmer, D. E. (1999) *J. Am. Chem. Soc. 121*, 2956-2964.

60. Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan, P. B. (1998) *Proc. Natl. Acad. Sci. USA 95*, 12890-12895.

61. Herman, D. M., Baird, E. E., and Dervan, P. B. (1998) *J. Am. Chem. Soc. 120*, 1382-1391.

62. Oakley, M. G., Mrksich, M., and Dervan, P. B. (1992) *Biochemistry 31*, 10969-10975.

63. Park, Y. W., and Breslauier, K. J. (1992) *Proc. Natl. Acad. Sci. USA 89*, 6653-6657.

64. Szewczyk, J. W., Baird, E. E., and Dervan, P. B. (1996) *Angew. Chem.-Int. Edit. Engl. 35*, 1487-1489.

Chapter 2

Inhibition of Major Groove Binding Proteins by Pyrrole-Imidazole Polyamides with an Arg-Pro-Arg Positive Patch

The text of this chapter was taken in part from a publication coauthored with Eldon E. Baird & Prof. Peter B. Dervan.

(R. E. Bremer, E. E. Baird & P. B. Dervan *Chem. & Biol.,* **1998**, *5*, 119-133.)

Abstract: Gene-specific targeting of any protein-DNA complex by small molecules is a challenging goal at the interface of chemistry and biology. Polyamides containing *N*-methylimidazole and *N*-methylpyrrole amino acids are synthetic ligands that have an affinity and specificity for DNA comparable to many naturally occurring DNA binding proteins. It has been demonstrated that an 8-ring hairpin polyamide targeted to a specific minor groove contact within a transcription factor binding site can inhibit protein-DNA binding and gene transcription. However, polyamides and certain major groove binding proteins have been found to co-occupy the DNA helix. In order to expand the number of genes targetable by pyrrole/imidazole polyamides, we set out to develop a class of polyamides which can selectively inhibit major groove binding proteins. An 8-ring hairpin polyamide conjugated to a C-terminal Arg-Pro-Arg tripeptide was designed to deliver a positive residue to the DNA backbone and interfere with protein-phosphate contacts. Gel mobility shift analysis demonstrated that a polyamide hairpin-Arg-Pro-Arg binding in the minor groove selectively inhibits binding of the bZIP protein, GCN4 (222-281), in the adjacent major groove. Substitution within the Arg-Pro-Arg revealed that each residue was required for optimal GCN4 inhibition. A pyrrole-imidazole polyamide which binds a predetermined site in the DNA minor groove and delivers a positive patch to the DNA backbone can selectively inhibit a DNA-binding protein which recognizes the adjacent major groove. A subtle alteration of the DNA microenvironment targeted to a precise location within a specific DNA sequence could achieve both gene- and protein-specific targeting.

Introduction

Small molecules specifically targeted to any predetermined DNA sequence would be useful tools in molecular biology and potentially human medicine. Polyamides containing *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) amino acids bind to predetermined sequences in the minor groove of DNA with affinities and specificities comparable to naturally occurring DNA binding proteins *(1-3)*. Sequence specificity is determined by a code of oriented side-by-side pairings of the Py and Im amino acids *(4-9)*. An Im/Py pairing targets a G•C base pair, while Py/Im pair recognizes C•G *(4-6)*. The Py/Py pair is degenerate and targets both A•T and T•A base pairs *(4-12)*. The validity of the pairing rules for ligand design is supported by a variety of polyamide structural motifs which have been characterized by footprinting, affinity cleaving, 2-D NMR, and X-ray methods *(1-14)*. Polyamides have been found to be cell permeable and to inhibit transcription factor binding and expression of a designated gene *(15-16)*. Py/Im polyamides offer a potentially general approach for gene regulation, provided that efficient inhibition of DNA-binding can be achieved for a variety of transcription factors.

Protein inhibition by synthetic ligands

Several approaches for the development of synthetic ligands which interfere with protein-DNA recognition have been reported. Oligodeoxyribonucleotides which recognize the major groove of double-helical DNA via triple-helix formation bind to a broad range of sequences with high affinity and specificity *(17-18)*. Although oligonucleotides and their analogs have been shown to disrupt protein-DNA binding *(19-21)*, the triple-helix approach is limited to purine tracts and suffers from poor cellular uptake. There are a few examples of carbohydrate-based ligands which interfere with protein-DNA recognition, but oligosaccharides cannot currently recognize a broad range of DNA sequences *(22-23)*. Perhaps most relevant to this work, analogs of distamycin (PyPyPy) appended with multiple cationic substituents have been found to inhibit protein

binding *(24-26)*. Based on these encouraging results, we wished to identify similar charged residues which could be appended to a Py/Im polyamide via linear solid phase synthesis and would not compromise polyamide binding specificity.

Pyrrole-imidazole polyamides inhibit proteins which recognize the minor groove

Proteins use a diverse structural library to recognize their target sequences *(27)*. Proteins such as TBP bind exclusively in the minor groove *(28)*, others, such as GCN4 *(29-31)*, bind exclusively in the major groove, and certain proteins such as Hin recombinase recognize both grooves *(32-33)*. Polyamides have been found to interfere with protein-DNA recognition in cases where contacts in the minor groove are important for protein-DNA binding affinity. For example, within the nine zinc-finger protein TFIIIA, fingers 4 and 6 bind in or across the minor groove and are required for high affinity binding ($K_a = 5 \times 10^9$ M^{-1}). An 8-ring hairpin polyamide ($K_a = 3 \times 10^{10}$ M^{-1}) targeted to the minor groove contact region of finger 4 has been recently found to efficiently inhibit protein binding *(15-16)*.

Pyrrole-imidazole polyamides bind simultaneously in the minor groove with proteins which recognize the major groove

X-Ray crystallography studies reveal that DNA bound by a 4-ring homodimeric polyamide is unaltered from its natural B-form structure, with all polyamide/DNA contacts confined to the minor groove *(34)*. Not surprisingly, polyamides have been found to bind simultaneously with ligands which exclusively occupy the major groove *(16, 35-36)*. For example, an 8-ring hairpin polyamide and a recombinant protein containing only the three amino-terminal zinc fingers of TFIIIA which are in the major groove were found to co-occupy the TFIIIA binding site *(16)*. Similarly, the 3-ring homodimer ImPyPy bound simultaneously with the bZIP protein GCN4 (226-281) *(35)*. We subsequently found that a variety of polyamide motifs co-occupy the DNA helix at sites both overlapping and

adjacent to GCN4 (Figure 2.1). The ubiquity of major groove contacts in protein-DNA recognition provides the impetus to develop approaches for inhibition of major groove proteins by polyamides which bind in the minor groove.

Design of the positive patch

The negatively charged DNA phosphate backbone provides a target for ligands designed to disrupt the unique microenvironment which represents a protein's binding site on the DNA double helix. Polyamides which deliver a positive patch to the DNA

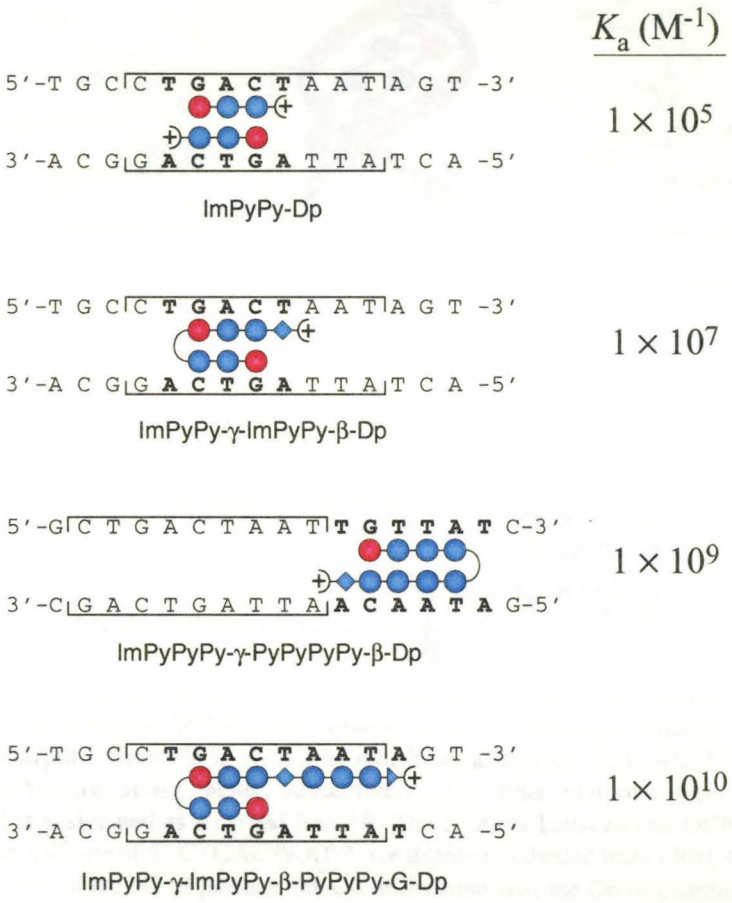


Figure 2.1. Schematic models of polyamides targeted to the binding site of the bZIP transcriptional activator, GCN4. (a) ImPyPy-Dp *(4, 35)*, (b) ImPyPy-γ-ImPyPy-β-Dp *(56)*, (c) ImPyPyPy-γ-PyPyPyPy-β-Dp *(1)*, (d) ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp *(56)*. The nine-base pair (5'-CTGACTAAT-3') GCN4 binding site is indicated by brackets above and below the base pairs. Red and blue circles represent imidazole (Im) and pyrrole (Py) polyamide rings, respectively. Blue diamonds and triangles represent β-alanine (β) and glycine (G), respectively. γ-Aminobutyric acid (γ) and dimethylaminopropylamide (Dp) are depicted as a curved line and a plus sign, respectively. Polyamide binding sites are shown in bold. Equilibrium association constants (K_a) for each polyamide binding to the indicated match site are shown at the right. Association constants were determined by DNase I footprinting; simultaneous binding was determined by gel mobility shift assay.

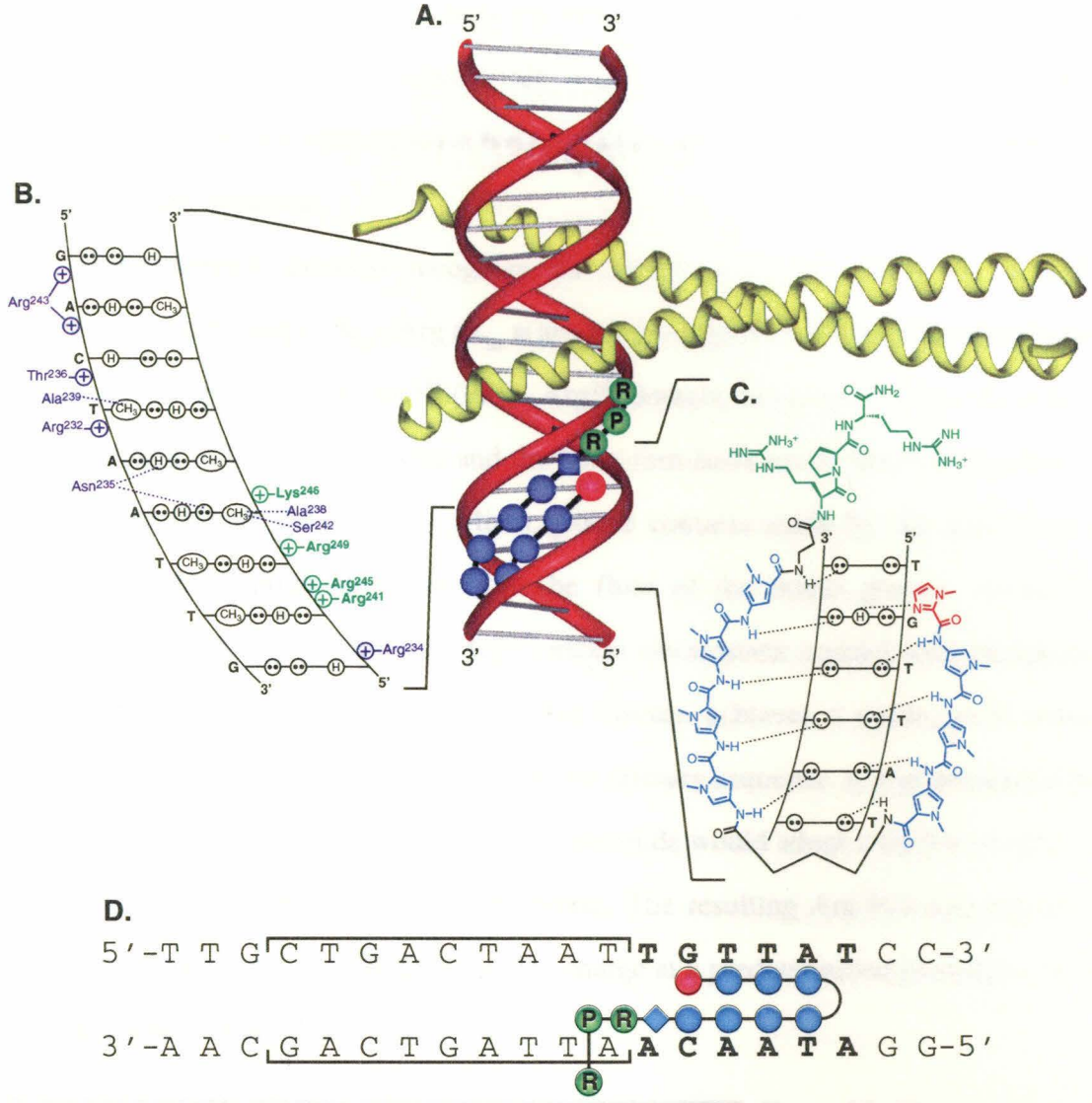


Figure 2.2. A schematic model of Arg-Pro-Arg-polyamides targeted to the major groove transcription factor, GCN4. (A) The α-helical GCN4 dimer (yellow) is shown binding to adjacent major grooves *(30)*. The Arg-Pro-Arg-hairpin polyamide is shown as red, blue, and green balls which represent imidazole, pyrrole, and Arg-Pro-Arg amino acids, respectively. The blue diamond represents β-alanine. γ-Aminobutyric acid is designated as a curved line. (B) The contacts between one GCN4 monomer and the major groove of one half-site of 5'-CTGACTAAT-3' are depicted (adapted from *(30)*). Circles with two dots represent the lone pairs of the N7 of purines, the O4 of thymine, and the O6 of guanine. Circles containing an H represent the N6 and N4 hydrogens of the exocyclic amines of adenine and cytosine, respectively. The C5 methyl group of thymine is depicted as a circle with CH_3 inside. Protein side chains which make hydrogen bonds or van der Waals contacts to the bases are shown in purple and connected to the DNA via a dotted line. Green and purple plus signs represent protein residues which electrostatically contact the phosphate backbone. The residues which are predicted to be disrupted by an Arg-Pro-Arg-polyamide are shown in green. (C) The hydrogen bonding model of the eight-ring hairpin polyamide ImPyPyPy-γ-PyPyPyPy-β-RPR bound to the minor groove of 5'-TGTTAT-3'. Circles with two dots represent the lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogens of guanines. Putative hydrogen bonds are illustrated by dotted lines. Py and Im rings are represented as blue and red rings, respectively. The Arg-Pro-Arg moiety is green. (D) The model of the polyamide binding its target site (bold) adjacent to the GCN4 binding site (brackets). Polyamide residues are as in A.

backbone may destabilize contacts between protein side chains and phosphate residues and consequently inhibit protein binding *(24-26)* (Figure 2.2). We looked to nature for an α-amino acid model of a minor groove bound domain which delivers a cationic side chain to the phosphate backbone.

Homeodomain proteins recognize the minor groove of DNA via a highly conserved Arginine (Arg) - X_{aa} - Arg (X_{aa} = any amino acid) *(37-38)*. In Hin recombinase, the corresponding Arg^{140} - $Proline^{141}$ (Pro) - Arg^{142} domain serves as a bridge between the N-terminal arm in the minor groove and the helix-turn-helix motif which recognizes the major groove (Figure 2.3) *(32-33)*. Minor groove contacts made by the side chain of Arg^{140} direct the peptide chain up from the floor of the minor groove, toward the backbone, where the guanidinium of Arg^{142} makes electrostatic contact with phosphates. Upon interaction with DNA, the Arg-Pro-Arg domain achieves a stable, local tertiary structure which is potentially based solely on the primary sequence. It was postulated that Arg-Pro-Arg attached at the C-terminus of a polyamide would adopt a similar structure to that of Arg^{140}-Pro^{141}-Arg^{142} in Hin recombinase. The resulting Arg-Pro-Arg-polyamide could be used to place a neutralizing positive charge at a predetermined phosphate on the DNA backbone (Figure 2.2).

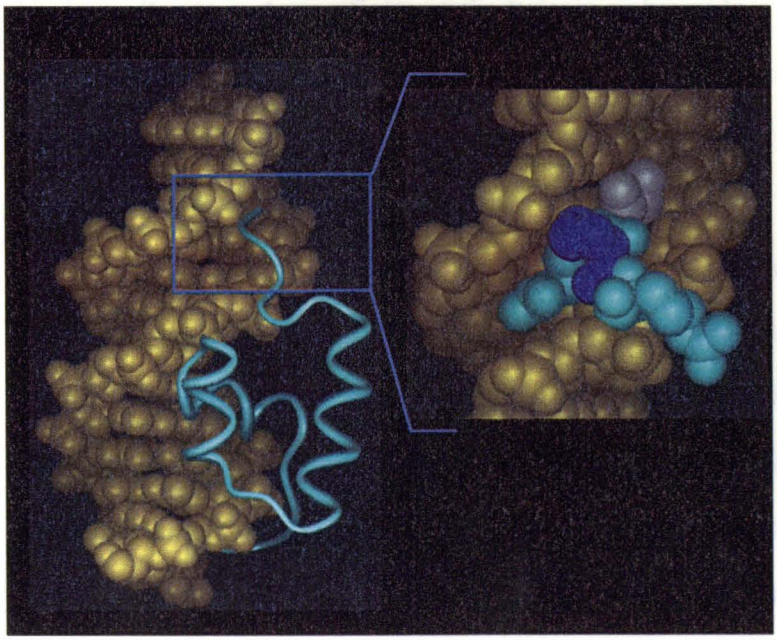


Figure 2.3. The structure of the DNA binding domain of Hin recombinase bound to DNA. (Left) Ribbon model of residues 139-190 of Hin recombinase (cyan) bound to the 5'-GTTTTTGATAAGA-3' DNA duplex (yellow) *(33)*. (Right) Enlarged view of the N-terminal Gly^{139}-Arg^{140}-Pro^{141}-Arg^{142} domain reaching from the floor of the minor groove to the phosphate backbone. Glycine, arginine, and proline are white, cyan and purple, respectively.

Arg-Pro-Arg-polyamides

A series of polyamides with Arg-Pro-Arg tripeptides at the C-terminus have been synthesized by solid phase methods. The polyamides were evaluated as inhibitors of the major groove transcription factor GCN4, the prototypical member of the basic region-leucine zipper (bZIP) family of transcriptional regulators *(39-41)*. The C-terminal sixty amino acids (222-281) of GCN4 contain the "leucine zipper" dimerization domain and the "basic region" which is responsible for DNA binding. GCN4 (222-281) has been shown to be sufficient for sequence specific binding *(29, 42-43)*. The basic region of each α-helical monomer makes specific hydrogen bonds, van der Waals contacts, and phosphate interactions with one half-site of the nine base pair pseudosymmetrical GCN4 binding site (Figure 2.2 A & B) *(29-31)*. The protein-DNA electrostatic interactions which are targeted for disruption by the Arg-Pro-Arg-polyamides are highlighted in Figure 2B.

We report here the ability of Arg-Pro-Arg-polyamides to selectively inhibit DNA binding by the major groove transcription factor, GCN4, as measured by gel mobility shift analysis. Substitutions in the Arg-Pro-Arg domain revealed the relative importance of each residue for protein inhibition. Quantitative DNase I footprint titration experiments were performed in order to measure the effect of net ligand charge on both the DNA binding affinity and specificity of the hairpin polyamide.

Results

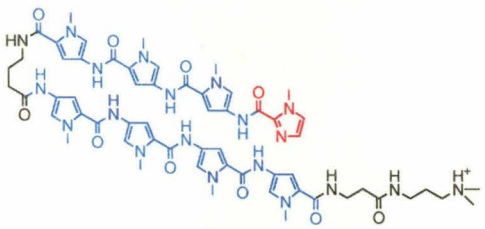
Synthesis of Arg-Pro-Arg-polyamides

The polyamides ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**) and ImImPyPy-γ-ImPyPyPy-β-Dp (**2**) were synthesized in a stepwise manner from Boc-β-alanine-Pam resin using Boc-chemistry machine-assisted protocols as previously described *(44)*. Polyamides with C-terminal aliphatic amino acids were synthesized on MBHA resin from Im and Py monomer units and commercially available aliphatic amino acids in 26 steps (Figures 2.4 and 2.5). Treatment with HF:*p*-cresol (9:1) followed by precipitation with ethyl ether and

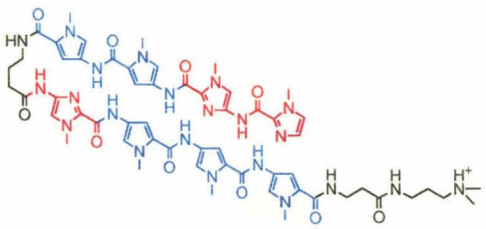
extraction with 0.1% TFA:CH_3CN (50:50) afforded the deprotected polyamide which was purified by reverse phase HPLC.

Arg-Pro-Arg polyamides selectively inhibit GCN4 (222-281) binding

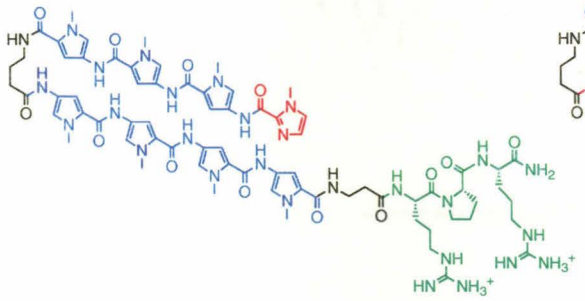
Synthetic radiolabeled DNA duplexes, ARE-1 and ARE-2, containing a GCN4 binding site (5'-CTGACTAAT-3') *(29, 35)*, were bound near saturation at 200 nM GCN4 (222-281) as revealed by gel mobility shift analysis (10 mM bisTris pH 7.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 50 µg/mL poly(dI-dC)•poly(dI-dC), 22 °C). ImPyPyPy-γ-PyPyPyPy-β-polyamides (**1** and **3**) target the six base pair 5'-TGTTAT-3' site adjacent to the GCN4 binding site of ARE-1 (Figures 2.2, 2.4 & 2.6A & C). ImImPyPy-γ-ImPyPyPy-β-polyamides (**2** and **4**) were designed to bind 5'-TGGTCT-3' adjacent to the GCN4 site in ARE-2 (Figures 2.4 & 2.6B & F).



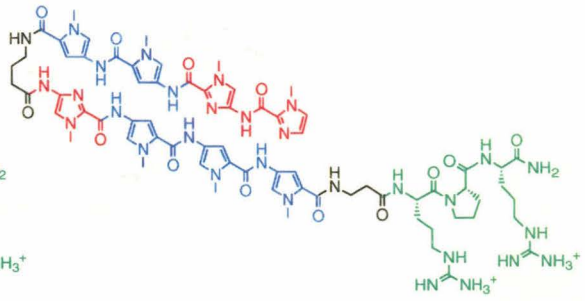
(**1**) ImPyPyPy-γ-PyPyPyPy-β-Dp



(**2**) ImImPyPy-γ-ImPyPyPy-β-Dp



(**3**) ImPyPyPy-γ-PyPyPyPy-β-RPR



(**4**) ImImPyPy-γ-ImPyPyPy-β-RPR

Figure 2.4. Structures of eight-ring hairpin polyamides, ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**) and ImImPyPy-γ-ImPyPyPy-β-Dp (**2**), and their Arg-Pro-Arg analogues, ImPyPyPy-γ-PyPyPyPy-β-RPR (**3**) and ImImPyPy-γ-ImPyPyPy-β-RPR (**4**). Imidazole and pyrrole rings and the Arg-Pro-Arg domain are shown in red, blue, and green, respectively. MALDI-TOF MS: **3**, 1546.3 (1545.7 calc); **4**, 1547.8 (1547.7 calc).

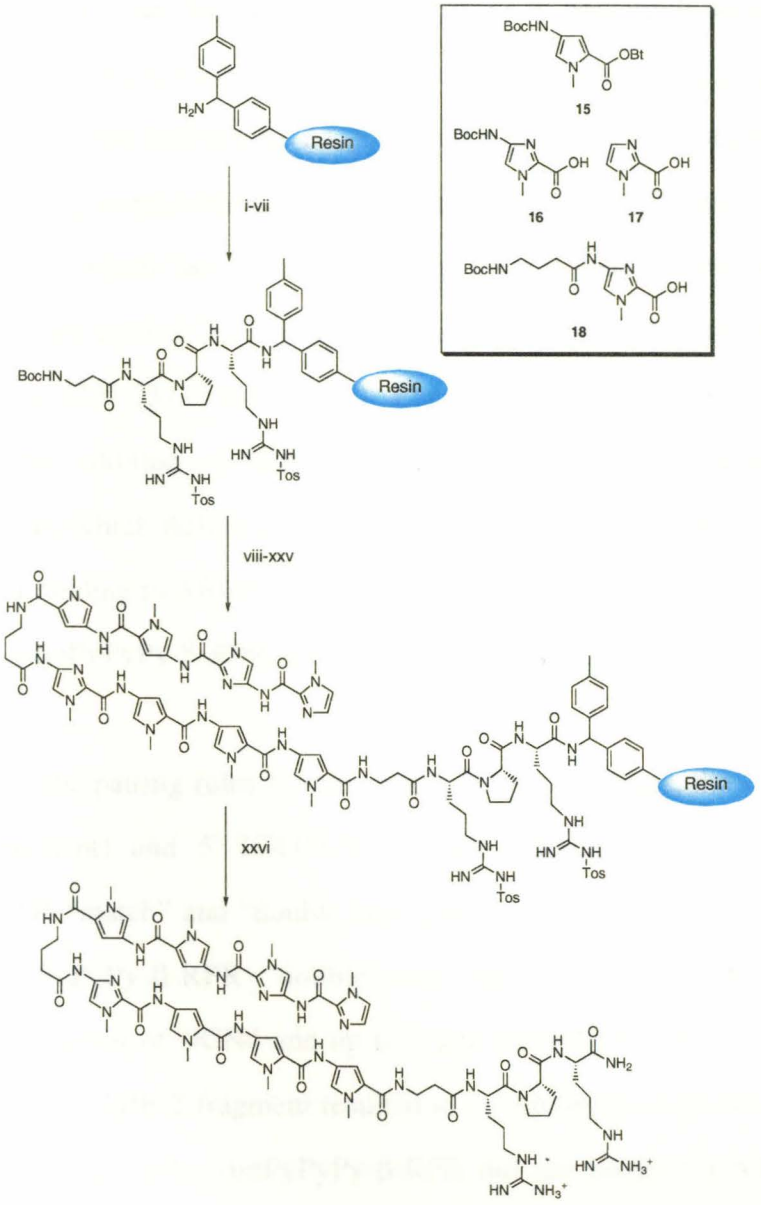


Figure 2.5. Synthetic scheme for solid phase preparation of Arg-Pro-Arg-polyamides. Cycling protocols consist of trifluoroacetic acid (TFA) deprotection followed by coupling with HOBt-activated Boc-pyrrole (**15**), Boc-imidazole (**16-18**), or aliphatic amino acid ester (steps i-xxv). The MBHA resin is then cleaved by treatment with HF:cresol (9:1) and purified by reverse phase HPLC (step xxvi). The synthesis of ImImPyPy-γ-ImPyPyPy-β-RPR (**4**) is shown as a representative example.

The ability of polyamides to inhibit GCN4 (222-281) binding was evaluated using the gel mobility shift assay. Increasing concentrations of polyamide were incubated with the desired radiolabeled synthetic DNA duplex followed by the addition of a constant concentration of 200 nM GCN4 (222-281). DNA fragments bound and unbound by GCN4 were separated using nondenaturing polyacrylamide gel electrophoresis.

Polyamides which lack the Arg-Pro-Arg moiety were unable to inhibit GCN4 binding (**1**-**2**, Figure 2.6A & B). The upper band in Figure 2.6A is the ARE-1 fragment bound by GCN4 (222-281) (lanes 2-12). Lanes 3-11 show that GCN4 binding was unaffected by the addition of ImPyPyPy-γ-PyPyPyPy-β-Dp. However, ImPyPyPy-γ-PyPyPyPy-β-RPR, which differs from **1** by the addition of the C-terminal Arg-Pro-Arg, inhibited GCN4 binding to ARE-1 (Figure 2.6C). When bound to its match site on ARE-2, ImImPyPy-γ-ImPyPyPy-β-RPR also successfully inhibited GCN4 binding (Figure 2.6F).

Based on the pairing rules for polyamide-DNA complexes, the sites 5'-TGTTAT-3' (ARE-1 fragment) and 5'-TGGTCT-3' (ARE-2 fragment) are for ImPyPyPy-γ-PyPyPyPy-β-RPR "match" and "double base pair mismatch" sites, respectively, and for ImImPyPy-γ-ImPyPyPy-β-RPR "double base pair mismatch" and "match" sites, respectively. Incubation of GCN4 and up to 2 μM ImPyPyPy-γ-PyPyPyPy-β-RPR with the double mismatch ARE-2 fragment resulted in no inhibition of GCN4 binding (Figure 2.6D). Likewise, ImImPyPy-γ-ImPyPyPy-β-RPR did not inhibit GCN4 binding to the mismatched ARE-1 fragment (Figure 2.6E).

When bound to their respective match sites, ImPyPyPy-γ-PyPyPyPy-β-RPRRR and ImImPyPy-γ-ImPyPyPy-β-RPRRR, which contain an additional three C-terminal arginine residues relative to **3** and **4**, were found to fully inhibit GCN4 (222-281) binding (Figure 2.7 and 2.8). The gel mobility shift experiments depicted in Figure 2.8 demonstrate that **5** and **6** selectively provided complete inhibition of GCN4 binding with no apparent loss in specificity for double base pair mismatches.

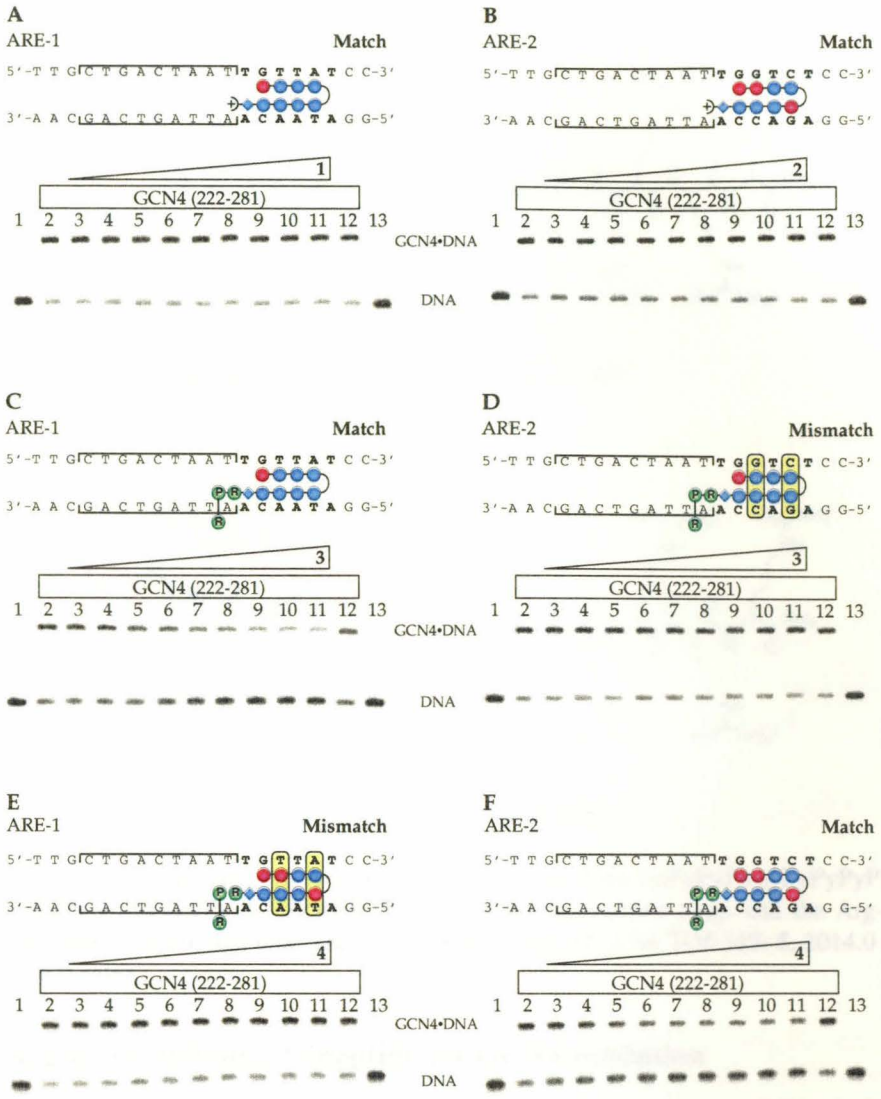
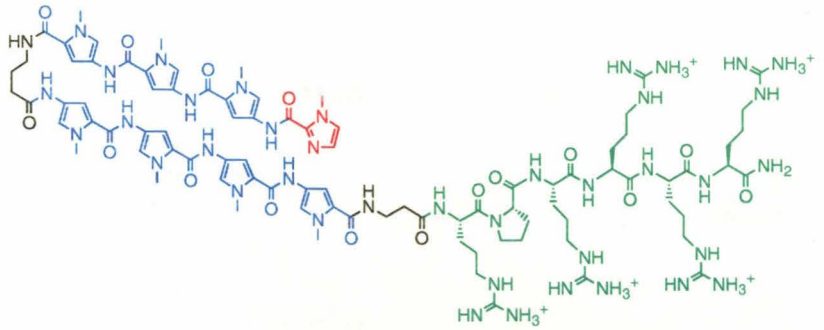
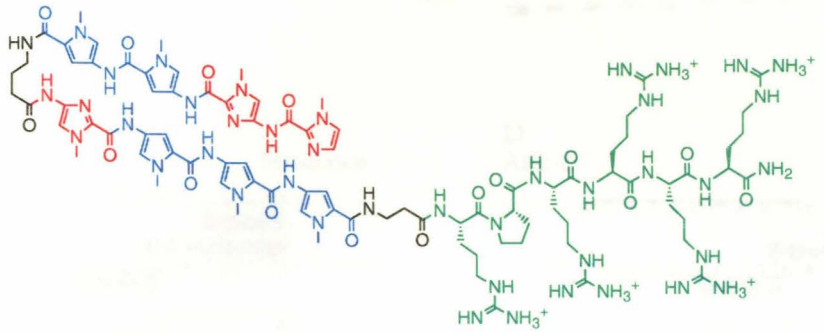


Figure 2.6. GCN4 (222-281) gel mobility shift experiments in the presence of Dp and Arg-Pro-Arg-polyamides. (A-F) Top, model of polyamide binding the match or mismatch site of DNA fragment ARE-1 or ARE-2 adjacent to the GCN4 binding site. The polyamide is colored as in Figure 1A. Mismatches are indicated by a yellow box. Bottom, storage phosphor autoradiogram of non-denaturing polyacrylamide gel showing GCN4 (222-281) binding to the radiolabeled ARE-1 fragment in the presence of increasing concentrations of polyamide (10 mM bisTris pH 7.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 50 μg/mL poly(dI-dC)•poly(dI-dC), 22 °C). The upper band is the GCN4 (222-281)-DNA complex and the lower band is free DNA. Lanes 1 and 13 are DNA only. Lanes 2 and 12 contain DNA incubated with 200 nM GCN4 (222-281). (A, C-D) Lanes 3-11 are 200 nM GCN4 (222-281) and 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM, and 2 μM **1** or **3**. (B, E-F) Lanes 3-11 are 200 nM GCN4 (222-281) and 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, and 500 nM **2** or **4**. (A) ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**) binding the match site 5'-TGTTAT-3' of ARE-1. (B) ImImPyPy-γ-ImPyPyPy-β-Dp (**2**) binding the match site 5'-TGGTCT-3' of ARE-2. (C) ImPyPyPy-γ-PyPyPyPy-β-RPR (**3**) binding the match site 5'-TGTTAT-3' of ARE-1. (D) ImPyPyPy-γ-PyPyPyPy-β-RPR (**3**) incubated with the double-mismatch site 5'-TGGTCT-3' of ARE-2 (mismatch sites underlined). (E) ImImPyPy-γ-ImPyPyPy-β-RPR (**4**) incubated with the double-mismatch site 5'-TGTTAT-3' of ARE-1. (F) ImImPyPy-γ-ImPyPyPy-β-RPR (**4**) binding the match site 5'-TGGTCT-3' of ARE-2. Lower K_a values ($K_a \approx 1 \times 10^7$ M^{-1}) are observed for polyamides under the gel shift conditions due to the carrier DNA which artificially depresses polyamide binding constants. The polyamide concentrations required for GCN4 inhibition are within the expected range based on the K_a under gel shift conditions.



(**5**) ImPyPyPy-γ-PyPyPyPy-β-RPRRRR



(**6**) ImImPyPy-γ-ImPyPyPy-β-RPRRRR

Figure 2.7. Structures of the Arg-Pro-Arg-Arg-Arg-Arg-polyamides ImPyPyPy-γ-PyPyPyPy-β-RPRRRR (**5**) and ImImPyPy-γ-ImPyPyPy-β-RPRRRR (**6**). Imidazole and pyrrole rings and the Arg-Pro-Arg-Arg-Arg-Arg domain are shown in red, blue, and green, respectively. MALDI-TOF MS: **5**, 2014.0 (2014.1 calc); **6**, 2017.1 (2016.0 calc).

Arg-Pro-Arg is the optimum tripeptide for GCN4 inhibition

Polyamides with deletions and/or substitutions in the Arg-Pro-Arg domain were prepared in order to determine the elements which were essential for GCN4 inhibition (Figure 2.9). Each of these polyamides was based on the ImPyPyPy-γ-PyPyPyPy-β polyamide targeted to 5'-TGTTAT-3' of ARE-1. The ability of polyamides **7**-**14** to bind their DNA target sites and inhibit GCN4 binding to ARE-1 was evaluated using DNase I footprinting and gel mobility shift analysis.

DNase I footprinting of polyamides **1**-**14** was performed on restriction fragments containing the appropriate ARE-1 or ARE-2 sequences under conditions identical to those used for the gel mobility shift experiments. In every case, (except **11**, see below) the polyamide was found to specifically bind the target site with $K_a \approx 1 \times 10^7$ M^{-1}. Lower K_a

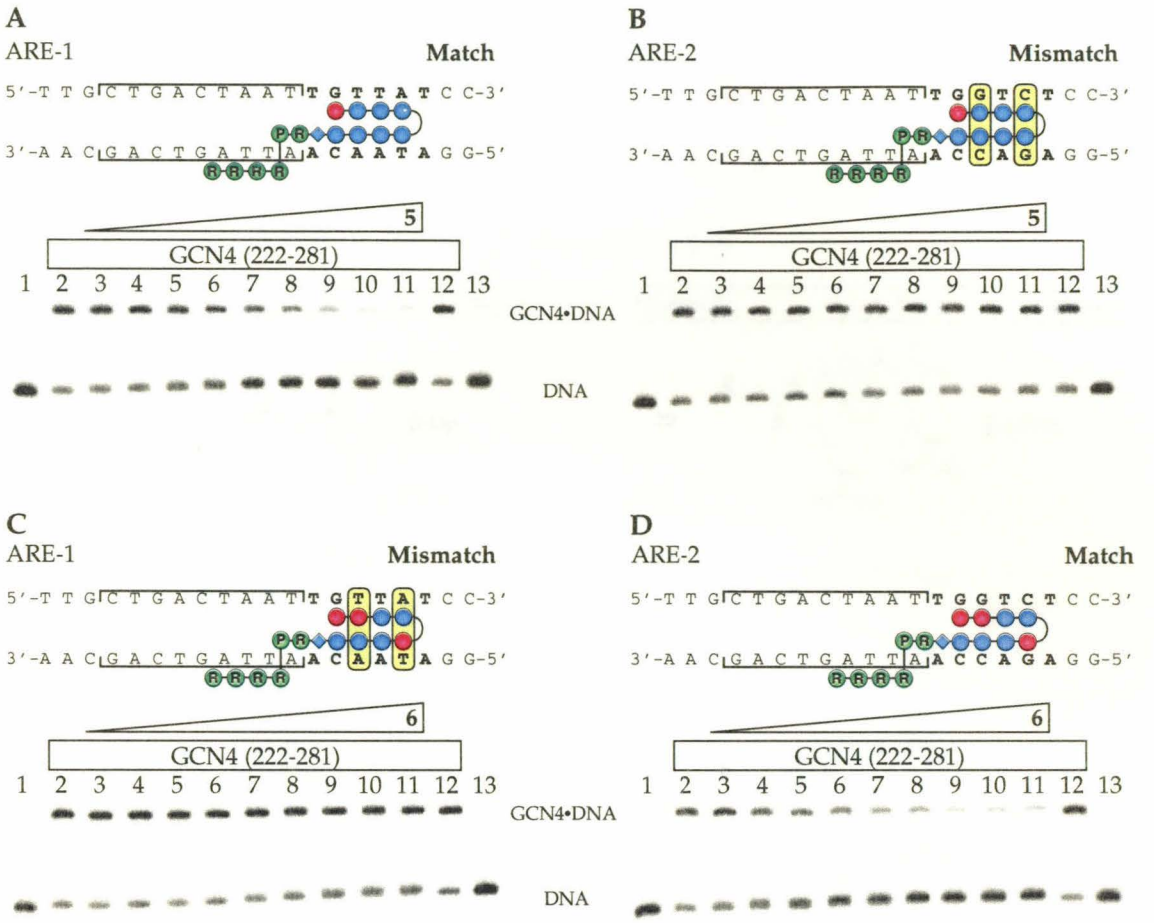
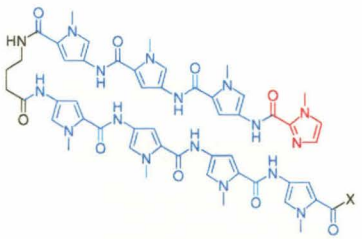
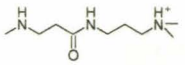
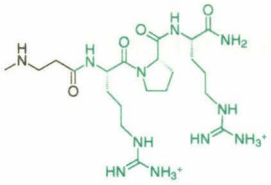
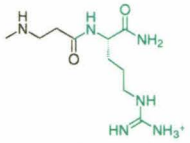
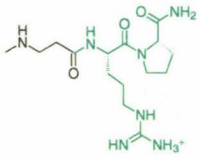
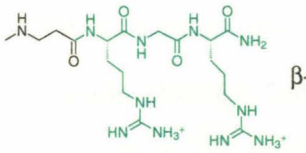


Figure 2.8. GCN4 (222-281) gel mobility shift experiments in the presence of Arg-Pro-Arg-Arg-Arg-Arg-polyamides. (A-D) Top, model of polyamide binding the match or mismatch site of DNA fragment ARE-1 or ARE-2 adjacent to the GCN4 binding site. The polyamide is colored as in Figure 1A. Mismatches are indicated by a yellow box. Bottom, storage phosphor autoradiogram of non-denaturing polyacrylamide gel showing GCN4 (222-281) binding to the DNA fragment in the presence of increasing concentrations of polyamide (10 mM bisTris pH 7.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 50 μg/mL poly(dI-dC)•poly(dI-dC), 22 °C). The upper band is the GCN4 (222-281)-DNA complex and the lower band is free DNA. Lanes 1 and 13 are DNA only. Lanes 2 and 12 contain DNA incubated with 200 nM GCN4 (222-281). (A-B) Lanes 3-11 are 200 nM GCN4 (222-281) and 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM, and 2 μM **5**. (C-D) Lanes 3-11 are 200 nM GCN4 (222-281) and 0.5 nM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, and 200 nM **6**. (A) ImPyPyPy-γ-PyPyPyPy-β-RPRRRR (**5**) binding the match site 5'-TGTTAT-3' of ARE-1. (B) ImPyPyPy-γ-PyPyPyPy-β-RPRRRR (**5**) incubated with the double-mismatch site 5'-TGGTCT-3' of ARE-2 (mismatch sites underlined). (C) ImImPyPy-γ-ImPyPyPy-β-RPRRRR (**6**) incubated with the double-mismatch site 5'-TGTTAT-3' of ARE-1. (D) ImImPyPy-γ-ImPyPyPy-β-RPRRRR (**6**) binding the match site 5'-TGGTCT-3' of ARE-2. Lower K_a values ($K_a \approx 1 \times 10^7$ M^{-1}) are observed for polyamides under the gel shift conditions due to the carrier DNA which artificially depresses polyamide binding constants. The polyamide concentrations required for GCN4 inhibition are within the expected range based on the K_a under gel shift conditions.



ImPyPyPy-γ-PyPyPyPy-**X**

Polyamide	X =	Inhibition
1		β-Dp -
3		β-RPR ++
7		β-R -
8		β-RP -
9		β-RGR +

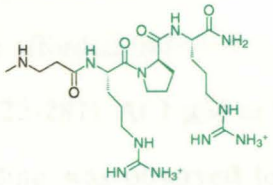
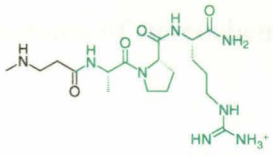
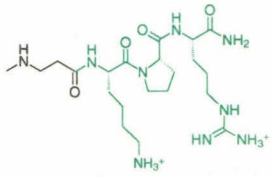
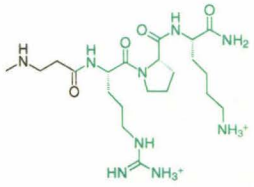
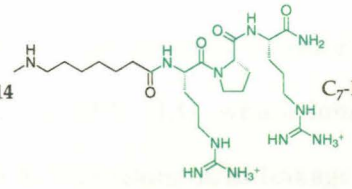
Polyamide	X =	Inhibition
10		β-R^{D}PR -
11		β-APR -
12		β-KPR +
13		β-RPK ++
14		C_7-RPR -

Figure 2.9. Aliphatic amino acid substitutions in the Arg-Pro-Arg domain. The structure of the ImPyPyPy-γ-PyPyPyPy-X polyamide is shown where X= β-Dp (β-Dp, **1**), β-Arg-Pro-Arg (β-RPR, **3**), β-Arg (β-R, **7**), β-Arg-Pro (β-RP, **8**), β-Arg-Gly-Arg (β-RGR, **9**), β-Arg-D-Pro-Arg (β-R^{D}PR, **10**), β-Ala-Pro-Arg (β-APR, **11**), β-Lys-Pro-Arg (β-KPR, **12**), β-Arg-Pro-Lys (β-RPK, **13**) or C_7-Arg-Pro-Arg (C_7-RPR, **14**). The ability of each polyamide to inhibit GCN4 (222-281) binding in a gel mobility shift assay performed identical to Figure 6A is reported. A lack of protein inhibition is indicated by a minus sign, reduced protein inhibition relative to **3** is shown by a plus sign, and two plus signs indicate protein inhibition identical to **3**. MALDI-TOF MS: **7**, 1292.6 (1292.6 calc); **8**, 1389.7 (1389.6 calc); **9**, 1505.6 (1505.7 calc); **10**, 1546.9 (1545.7 calc); **11**, 1461.2 (1460.7 calc); **12**, 1517.8 (1517.7 calc); **13**, 1517.6 (1517.7 calc); **14**, 1601.8 (1601.8 calc).

values are observed for polyamides under the gel shift conditions due to the carrier DNA which artificially depresses polyamide binding constants. The polyamide concentrations required for GCN4 inhibition are within the expected range based on the K_a under gel shift conditions.

Deletion of the terminal Pro-Arg or Arg, as in ImPyPyPy-γ-PyPyPyPy-β-R, (**7**) and ImPyPyPy-γ-PyPyPyPy-β-RP (**8**), results in polyamides which are unable to inhibit GCN4 binding. Substituting the proline with glycine afforded ImPyPyPy-γ-PyPyPyPy-β-RGR (**9**), which did not effectively inhibit GCN4 (222-281). At 1 μM of **9**, < 50% of the GCN4 was inhibited. No inhibition of GCN4 binding was observed for ImPyPyPy-γ-PyPyPyPy-β-R^{D}PR (**10**) which contained a single inversion of stereochemistry relative to **3**.

The internal arginine was replaced with an alanine residue to provide ImPyPyPy-γ-PyPyPyPy-β-APR (**11**). **11** was unable to inhibit GCN4 binding under these conditions. This Arg to Ala substitution was the only alteration which was found to affect polyamide binding affinity. By DNase I footprinting, **11** binds the 5'-TGTTAT-3' target site with 10-fold lower affinity than **3** under conditions identical to those used for gel shift analysis. The conservative substitution of lysine for arginine in ImPyPyPy-γ-PyPyPyPy-β-KPR (**12**) also compromised the polyamide's ability to inhibit GCN4. At 1 μM **12**, < 50% of the bound GCN4 was inhibited, similar to **9**. However, the identical substitution in the C-terminal position afforded ImPyPyPy-γ-PyPyPyPy-β-RPK (**13**), which inhibited GCN4 binding identically to the Arg-Pro-Arg polyamide **3**. The amino acid linkage between the final Py amino acid and the initial arginine was also crucial for GCN4 inhibition. A polyamide in which the β-alanine linker was replaced with a 7-aminoheptanoic acid linker, ImPyPyPy-γ-PyPyPyPy-C7-RPR (**14**), was unable to inhibit GCN4 binding. Protein inhibition did not require prebinding of polyamide. Preincubation of ARE-1 with GCN4 followed by addition of **5** afforded inhibition identical to that of prebound polyamide.

Protein inhibition was confirmed by affinity cleavage

The ability of the Arg-Pro-Arg polyamides to inhibit DNA binding by GCN4 was further investigated using affinity cleavage. Fe(II)•EDTA tethered to the N-terminus of the GCN4 peptide has been previously shown to result in a tripartite cleavage pattern upon activation with DTT *(29, 35)*. Polyamides that bind simultaneously with the GCN4 protect a region of the DNA from cleavage, which appears as a footprint on the cleavage pattern relative to that in the absence of polyamide. Figure 2.10 shows that ImImPyPy-γ-ImPyPyPy-β-Dp (**2**) has little effect upon protein binding when the polyamide is targeted

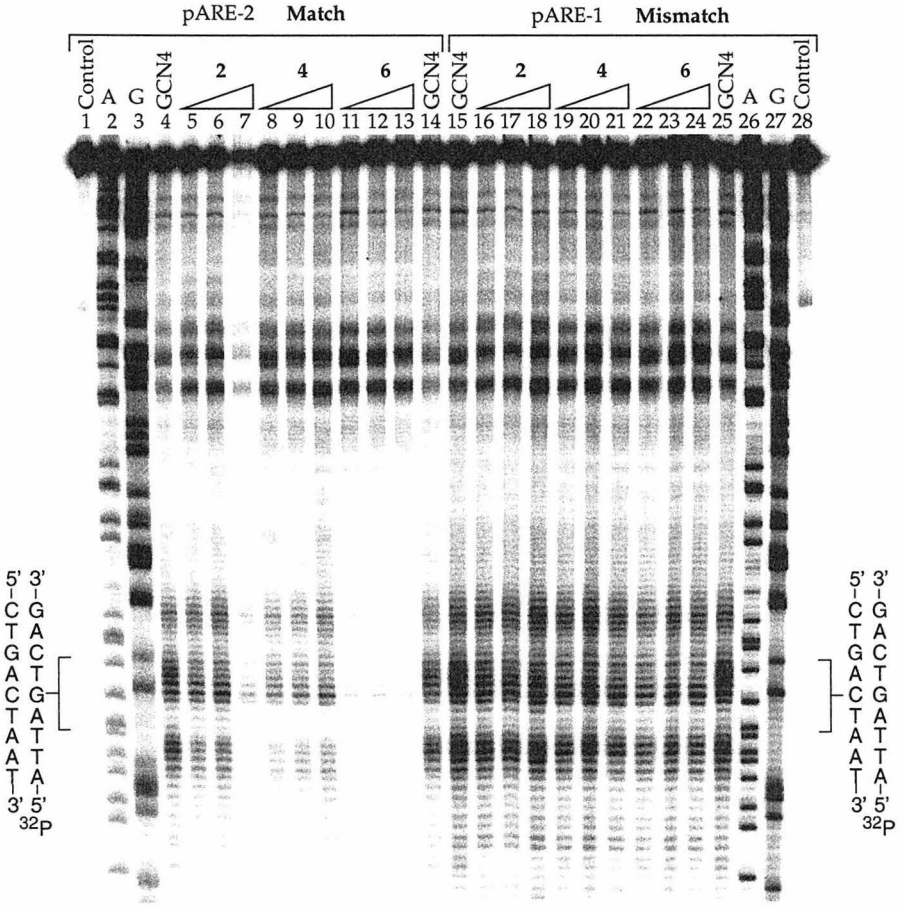


Figure 2.10. Fe(II)•EDTA-GCN4 (222-281) affinity cleavage experiments in the absence and presence of polyamides ImImPyPy-γ-ImPyPyPy-β-Dp (**2**), ImImPyPy-γ-ImPyPyPy-β-RPR (**4**), and ImImPyPy-γ-ImPyPyPy-β-RPRRRR (**6**). Lanes 1-14 and 15-28 are the 5'-end-labeled *Eco*RI/*Pvu*II restriction fragments of pARE-2 and pARE-1, repectively. Lanes 1 and 28, Intact DNA; lanes 2 and 26, A-specific reaction; lanes 3 and 27, G-specific reaction; lanes, 4, 14, 15, and 25, 200 nM Fe(II)•EDTA-GCN4 (222-281) only; lanes 4-13 and 16-24 contain 200 nM Fe(II)•EDTA-GCN4 (222-281) and 500 nM, 1 μM or 2 μM polyamide in each set of three lanes. Lanes 5-7 and 16-18, **2**; lanes 8-10 and 19-21, **4**; lanes 11-13 and 22-24, **6**.

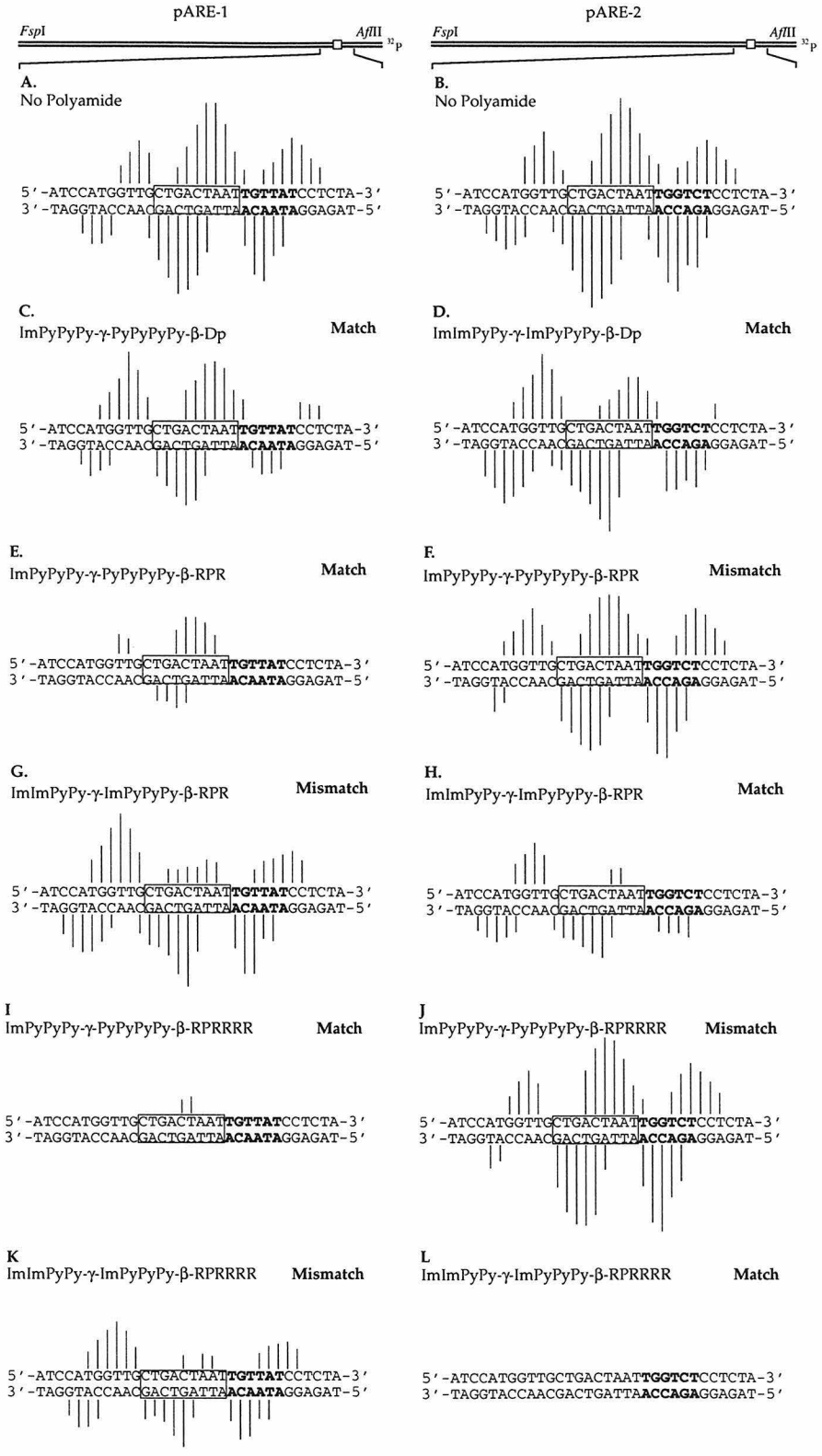


Figure 2.11. Histograms of Fe(II)•EDTA-GCN4 (222-281) affinity cleavage patterns in the absence and presence of ArgProArg polyamides. The *Eco*RI/*Pvu*II restriction fragments of pARE-1 and pARE-2 are shown on the left and right, respectively. Each panel is labeled with the polyamide employed at a concentration of 1 μM. Line heights are proportional to cleavage intensity.

adjacent to the GCN4 binding site on pARE-2 (lane 4 vs. 6). ImImPyPy-γ-ImPyPyPy-β-RPR (**4**) provides some inhibition of GCN4 binding (lanes 4 vs. 10), while ImImPyPy-γ-ImPyPyPy-β-RPRRRR (**6**) completely inhibits GCN4 binding to the target site (lanes 4 vs. 13). Polyamides **4** and **6** retain their specificity over the double base pair mismatch in pARE-1 as evidenced by a lack of significant inhibition in lanes 21 and 24.

The affinity cleavage patterns of Fe(II)•EDTA-GCN4 on pARE-1 and pARE-2 in the absence of polyamide were consistent with those previously observed (Figure 11 A & B) *(29, 35)*. The addition of a polyamide that binds the adjacent target site, but does not prevent GCN4 binding, resulted in a protection of cleavage at the polyamide binding site (Figure 2.11 C & D) *(35)*. The 5' and 3' shifts of the GCN4 and polyamide were consistent with the major and minor groove occupancy of each ligand, respectively. The presence of the Arg-Pro-Arg tail afforded GCN4 inhibition with little effect on protein binding to the polyamide mismatch target (Figure 2.11 E-H). Similarly, the presence of the additional terminal arginine residues resulted in complete inhibition of protein binding near the polyamide match site and slight inhibition on the mismatch target. The inhibition of GCN4 binding by Arg-Pro-Arg polyamides as measured by Fe(II)•EDTA-GCN4 affinity cleavage was consistent with the results obtained by gel mobility shift assays.

The DNA binding of the Arg-Pro-Arg polyamide was also probed using an analog containing a C-terminal lysine with a ε-Fe(II)•EDTA moiety, ImPyPyPy-γ-PyPyPyPy-β-RPRK(EDTA). This polyamide bound the target site on pARE-1 and afforded a 3'-shifted cleavage pattern consistent with minor groove binding (Figure 2.12). The locus of the pattern was indicative of a single ligand bound in a 5'-3'/N-C orientation.

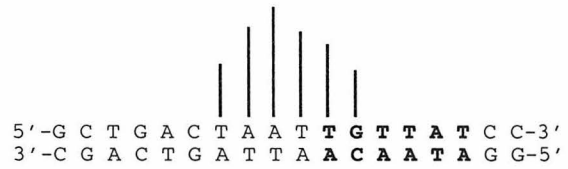


Figure 2.12. Histogram of affinity cleavage by ImPyPyPy-γ-PyPyPyPy-β-RPRK(EDTA•Fe(II)) at 1 μM.

Salt dependence

In order to evaluate the sensitivity of positive patch mediated major groove protein inhibition to the nature of the compensating electrolyte, as well as the overall ionic strength, gel mobility shift analysis was performed using a buffer which models the environment of the cellular nucleus (20 mM MOPS, pH 7.2, 140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 1 mM spermine) *(45)*. Arg-Pro-Arg-polyamide **3** was found to inhibit GCN4 (222-281) binding under the *in vivo* ionic conditions which feature KCl as the primary compensating electrolyte (Figure 2.13) and the conditions optimized for protein binding which feature NaCl as the predominant compensating electrolyte (Figure 2.6C). Further biophysical characterization of major groove protein inhibition by positive patch polyamides will be reported in due course.

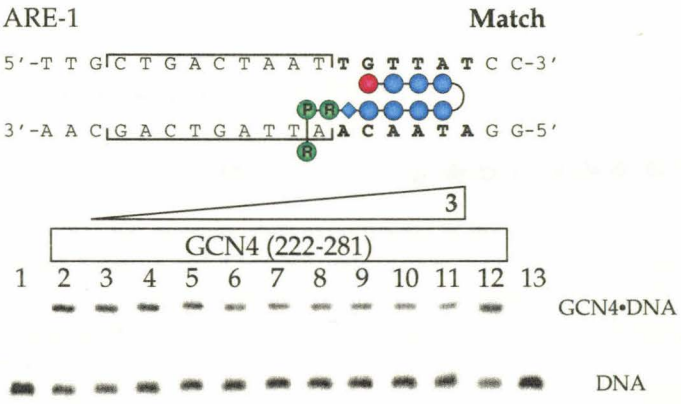


Figure 2.13. GCN4 (222-281) gel mobility shift experiments in the presence of ImPyPyPy-γ-PyPyPyPy-β-RPR under purported physiological ionic conditions. Top, model of polyamide binding the match site of DNA fragment ARE-1 adjacent to the GCN4 binding site. Bottom, storage phosphor autoradiogram of non-denaturing polyacrylamide gel showing GCN4 (222-281) binding to the radiolabeled ARE-1 fragment in the presence of increasing concentrations of **3** under ionic conditions modeling those found *in vivo* (20 mM MOPS pH 7.2, 140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 1 mM spermine, 50 μg/mL poly(dI-dC)•poly(dI-dC), 22 °C). The upper band is the GCN4 (222-281)-DNA complex and the lower band is free DNA. Lanes 1 and 13 are DNA only. Lanes 2 and 12 contain DNA incubated with 200 nM GCN4 (222-281). Lanes 3-11 are 200 nM GCN4 (222-281) and 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM, and 2 μM **3**. Lower K_a values ($K_a \approx 1 \times 10^7$ M^{-1}) are observed for polyamides under the gel shift conditions due to the carrier DNA which artificially depresses polyamide binding constants. The polyamide concentrations required for GCN4 inhibition are within the expected range based on the K_a under gel shift conditions.

Arg-Pro-Arg polyamides lower GCN4 binding affinity

The dissociation constant of GCN4 for the ARE-2 probe was measured by gel mobility in the absence and presence of saturating amounts of polyamide (Figure 2.14). Under the conditions employed here with poly(dI-dC)•poly(dI-dC) required as carrier to minimize non-specific binding by the GCN4 peptide, the K_d of GCN4 (222-281) for ARE-2 was found to be 181 nM. Titration of **6** had revealed that maximal inhibition of GCN4 (222-281) binding was observed by 1 μM **6** (Figure 2.6). The K_d of GCN4 (222-281) in the presence of 1 μM **6** was found to be 2.6 μM, approximately a 14-fold lower affinity than in the absence of **6**.

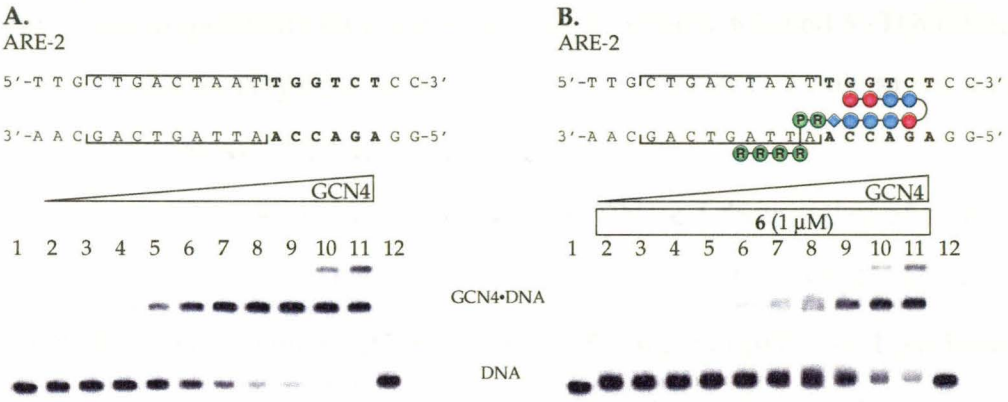


Figure 2.14. GCN4 (222-281) gel mobility shift experiments in the absence and presence of ImImPyPy-γ-ImPyPyPy-β-RPRRRR (**6**). Lanes 1 and 12: DNA only, Lanes 2-11: 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM, 2 μM, 5 μM and 10 μM GCN4 (222-281), respectively. (A) GCN4 (222-281) titration with ARE-2 in the absence of **6**. (B) GCN4 (222-281) titration with ARE-2 in the presence of a constant concentration of 1 μM **6**.

Polyamide binding affinity and specificity are unaltered by Arg-Pro-Arg

In order to evaluate the effects of the Arg-Pro-Arg moiety on the DNA binding properties of the polyamides, quantitative DNase I footprint titration experiments were performed to determine the equilibrium association constants of polyamides **1-6** for their respective six base pair match and single base pair mismatch sites (10 mM Tris pH 7.0, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 22 °C) *(46)*. DNase I footprinting of ImImPyPy-γ-ImPyPyPy-β-Dp (**2**), ImImPyPy-γ-ImPyPyPy-β-RPR (**4**), and ImImPyPy-γ-ImPyPyPy-β-RPRRRR (**6**) was performed on the 250 bp *Eco*RI/*Pvu*II restriction fragment

of pJK6 (Figure 2.15) *(47)*. **2** bound the match sites, 5'-TGGTCA-3' and 5'-TGGACA-3', with identical affinities within experimental error (K_a = 1.3 (±0.1) × 10^{10} M^{-1} and 6.4 (±1.2) × 10^9 M^{-1}, respectively). **2** also demonstrated greater than 100-fold specificity for a single base pair mismatch site 5'-TG<u>T</u>ACA-3' ($K_a \leq 5 \times 10^7$ M^{-1}, mismatched base pair underlined). Similar affinity and a slight increase in specificity were observed for Arg-Pro-Arg polyamide **4**. 5'-TGGTCA-3' and 5'-TGGACA-3' were bound by **4** (K_a = 4.6 (±0.2) × 10^9 M^{-1} and 6.6 (±1.0) × 10^9 M^{-1}, respectively) with greater than 450-fold specificity versus the mismatch site (K_a = ≤1 × 10^7 M^{-1}). The additional three terminal arginines of **6** generated a 10-fold increase in affinity relative to **2** coupled with a significant loss in specificity for a single base pair mismatch. **6** bound 5'-TGGTCA-3', 5'-TGGACA-3', and 5'-TG<u>T</u>ACA-3' with affinities of 2.6 (±0.4) × 10^{10} M^{-1}, 2.8 (±0.5) × 10^{10} M^{-1} and 1.9 (±0.8) × 10^{10} M^{-1}, respectively.

Corresponding results were observed for DNase I footprinting of ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**), ImPyPyPy-γ-PyPyPyPy-β-RPR (**3**) and ImPyPyPy-γ-PyPyPyPy-β-RPRRRR (**5**) on the 229 bp *Afl*II/*Fsp*I restriction fragment of pJT8 *(1)*. **1** has been shown to bind its six base pair match site, 5'-AGTTAT-3', with an affinity of 3.5 (±0.8) × 10^9 M^{-1} and 7-fold specificity versus a single base pair mismatch site 5'-AGTA<u>C</u>T-3' (K_a = 5.0 (±0.5) × 10^8 M^{-1}, Table 2.1) *(1)*. The Arg-Pro-Arg polyamide **3** demonstrated only a slight loss in affinity and a similar specificity (K_a = 5.5 (±1.5) × 10^8 M^{-1} for 5'-AGTATT-3' and 9.2 (±0.4) × 10^7 M^{-1} for 5'-AGTA<u>C</u>T-3'). Analogous to **6**, the additional terminal arginines of **5** provided a 10-fold increase in affinity for the match site (K_a = 1.0 (±0.2) × 10^{10} M^{-1}) and a severe loss in specificity for a single base pair mismatch (K_a = 3.4 (±0.5) × 10^9 M^{-1}).

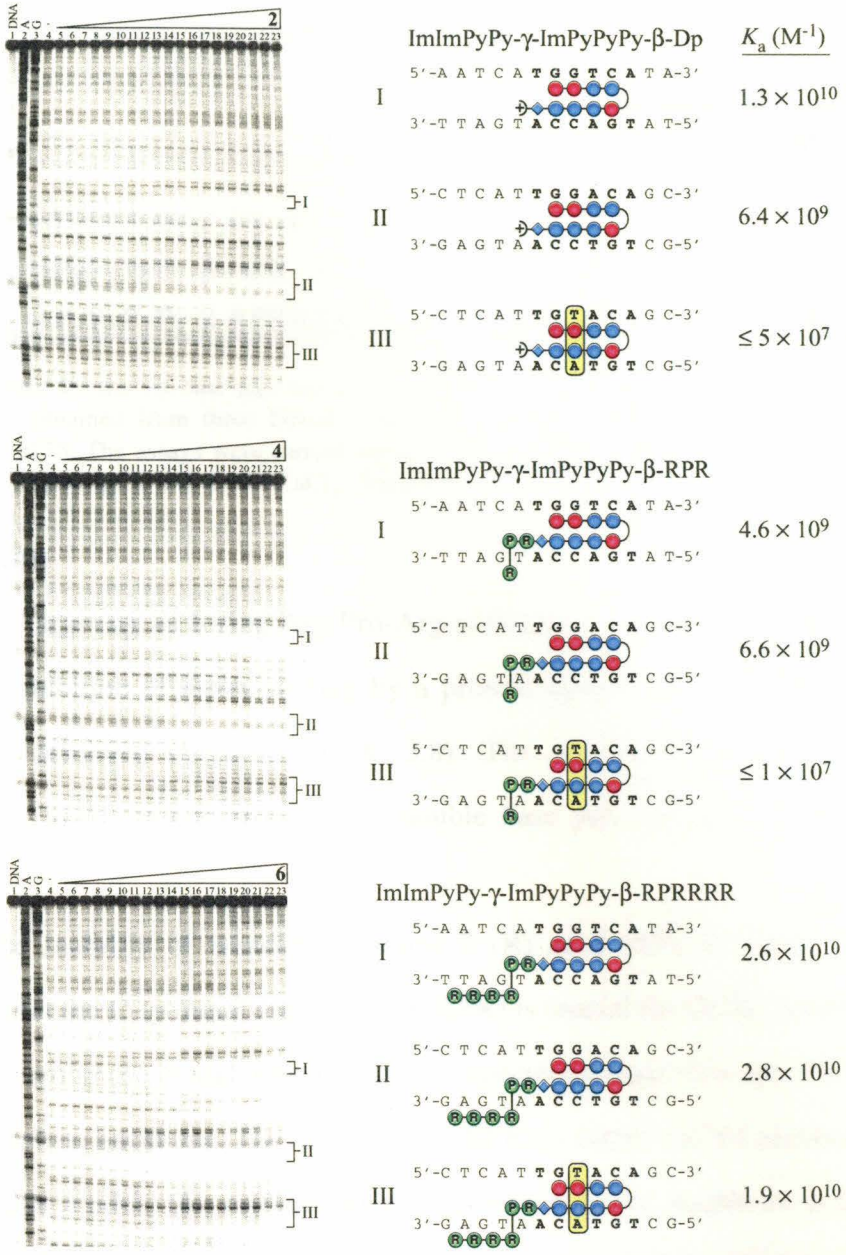


Figure 2.15. Quantitative DNase I footprint titration experiments with ImImPyPy-γ-ImPyPyPy-β-Dp (**2**, top), ImImPyPy-γ-ImPyPyPy-β-RPR (**4**, middle), and ImImPyPy-γ-ImPyPyPy-β-RPRRRR (**6**, bottom), on the 3′-^{32}P-end-labeled 250 base pair *Eco* RI/*Pvu* II restriction fragment from plasmid pJK6 *(47)*. Left, storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated from DNase I digestion. All reactions contained 20 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$ and were performed at 22 °C. Lane 1, intact DNA; lane 2, A-specific reaction; lane 3, G-specific reaction; lane 4, DNase I standard; lanes 5-23, 1 pM, 2 pM, 5 pM, 10 pM, 15 pM, 25 pM, 40 pM, 65 pM, 100 pM, 150 pM, 250 pM, 400 pM, 650 pM, 1 nM, 1.5 nM, 2.5 nM, 4 nM, 6.5 nM, 10 nM, respectively. Right, models of polyamide bound to the match or mismatch sites on the DNA fragment. I (5′-TGGTCT-3′) and II (5′-TGGACT-3′) are both match sites and III (5′-TGTACA-3′) is a single base pair mismatch site (mismatch underlined and indicated by a yellow box). The polyamide is colored as in Figure 1A. The restriction fragment was 3′-^{32}P-end labeled on the bottom strand. Equilibrium association constants of each polyamide binding to the match or mismatch site is reported on the right. Each value is the average of at least three experiments.

Table 2.1. Equilibrium Assocation Constants (M^{-1})

	Binding Site	
Polyamide	5'-AGTATT-3'	5'-AGTACT-3'
ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**)	$3.5 \times 10^{9*}$	$5.0 \times 10^{8*}$
ImPyPyPy-γ-PyPyPyPy-β-RPR (**3**)	5.5×10^{8}	9.2×10^{7}
ImPyPyPy-γ-PyPyPyPy-β-RPRRRR (**5**)	1.0×10^{10}	3.4×10^{9}

Values reported for the six-base-pair match (5'-AGTATT-3') and mismatch (5'-AGTACT-3') sites are the mean values obtained from three DNase I footprint titration experiments on the *Afl*III/*Fsp*I restriction fragment of pJT8. The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$. *From *(1)*.

Discussion

By targeting an 8-ring Arg-Pro-Arg-polyamide adjacent to a GCN4 binding site, selective inhibition of DNA binding by a protein which exclusively contacts the major groove is achieved (Figures 2.6 & 2.8). The polyamide domain binds sequence specifically in the minor groove with double base pair mismatches preventing GCN4 inhibition.

The inability of truncated analogs **7** (R) or **8** (RP) to inhibit GCN4 binding indicates that the C-terminal arginine in **3** (RPR) is crucial for GCN4 inhibition. Based on the Hin recombinase model, this arginine is expected to make non-specific contacts to the DNA phosphate backbone. The ability of **13** (RPK) to inhibit GCN4 identically to **3** (RPR) supports this model. The neutralization of a portion of the backbone is the most likely mechanism by which Arg-Pro-Arg polyamides achieve GCN4 inhibition. Other models, such as steric blockage of the major groove or DNA distortion, cannot be ruled out *(48-49)*. Modeling suggests that Arg-Pro-Arg is insufficient to cross the DNA backbone and block the major groove. Determination of the exact mechanism of inhibition awaits high-resolution structure studies which are in progress. Targeting protein:DNA contacts at other phosphates for inhibition may enhance the ability of ArgProArg polyamides to inhibit GCN4 binding (see Chapter 3).

Arg-Pro directs the terminal Arg to the phosphate backbone

The results of GCN4 inhibition experiments with polyamides **7**-**14** suggest that the Arg-Pro-Arg of **3** and **4** adopts a stable and well-defined structure similar to Arg[140]-Pro[141]-Arg[142] of Hin recombinase (Figure 2.16). The internal Arg-Pro of **3** and **4** (**RP**R) is required for GCN4 inhibition. Polyamide **8** does not inhibit GCN4, suggesting that these two residues play a structural role in the placement of the terminal arginine near the phosphate backbone. Replacing the rigid proline of **3** (RPR) with a flexible glycine (**9**) (RGR) allows significant amounts of GCN4 to remain bound in the presence of saturating concentrations of **9**. The glycine in **9** may permit the C-terminal arginine to shift to a position which permits simultaneous binding with GCN4.

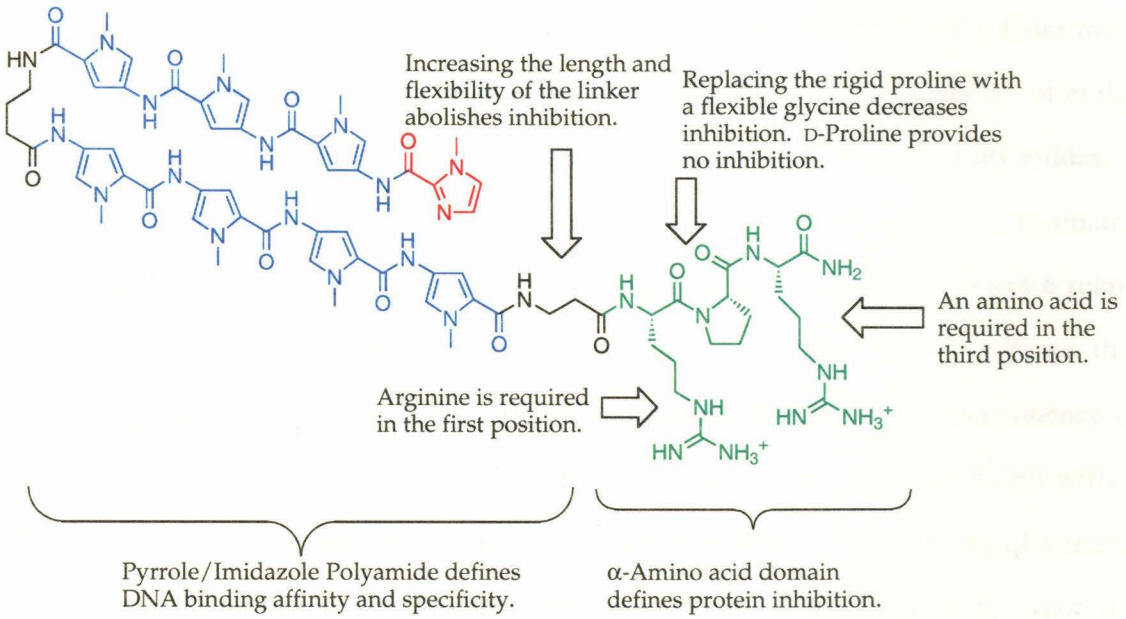


Figure 2.16. Summary of the structural requirements for optimal GCN4 inhibition by Arg-Pro-Arg polyamides.

ImPyPyPy-γ-PyPyPyPy-β-R^{D}PR (**10**) is a diastereomer of ImPyPyPy-γ-PyPyPyPy-β-RPR (**3**), but is unable to inhibit GCN4. Modeling indicates that substitution of D-proline for L-proline may in fact direct the neutralizing terminal arginine to the backbone of the opposite DNA strand. However, confirmation of this prediction awaits studies with other protein systems.

Replacement of the internal arginine with alanine, as in **11**, reduces binding affinity by a factor of ten and prevents GCN4 inhibition. Furthermore, the Lys-Pro-Arg polyamide, **12**, exhibits a binding affinity comparable to **3**, yet it is a less effective inhibitor of GCN4. Together, these results suggest that the guanidinium of the internal arginine makes specific contacts with the DNA which are required for the proper positioning of the remaining residues. Replacement of the β-alanine linker (**3**) with 7-aminoheptanoic acid (**14**) eliminates inhibition, further implicating the placement of the Arg-Pro-Arg moiety as a requirement for effective inhibition.

Arg-Pro-Arg-polyamides retain high DNA binding affinity and specificity

Quantitative DNase I footprinting demonstrates that the addition of a C-terminal Arg-Pro-Arg tripeptide as in **3** and **4** does not alter the DNA binding properties of eight-ring hairpin polyamides (**1** and **2**). However, Arg-Pro-Arg-Arg-Arg-Arg-polyamides, **5** and **6**, have increased binding affinity but no specificity for a single base pair mismatch site. DNase I footprinting and gel mobility shift analysis demonstrate that **5** and **6** retain their specificity versus double base pair mismatch sites. These results indicate that synthetic ligands may balance the benefits of additional charge with the consequence of lowered sequence specificity *(50)*. For example, a distamycin analog modified with a decaaza decabutylamine moiety on a pyrrole nitrogen interferes with binding of a major groove transcription factor *(26)*. Unfortunately, the sequence specificity of this molecule which contains potentially eleven positive charges has not been reported. The results described here suggest that such a molecule may bind DNA with reduced sequence specificity.

Conclusions

The functional repertoire of polyamides as synthetic ligands for the control of transcription factor binding has been expanded to include proteins which bind exclusively in the major groove of DNA. The addition of an Arg-Pro-Arg tripeptide results in no significant alteration of polyamide-DNA binding affinity or specificity. By targeting sequences outside the protein binding site, selective inhibition of GCN4 at two sites was demonstrated using Arg-Pro-Arg-polyamides. The Arg-Pro-Arg domain appears to adopt a stable and defined structure which delivers a neutralizing positive charge to the DNA backbone where it competes with GCN4 for contact to the phosphates. The broad targetable sequence repertoire of polyamides, coupled with the ubiquity of backbone contacts in protein recognition of DNA, make phosphate neutralization by a positive patch a promising approach for inhibition of major groove transcription factors.

Experimental

All buffers for gel mobility shift and footprinting experiments were prepared from J.T. Baker reagents and 0.2 μM filtered. EDTA and DTT were obtained from Gibco BRL. Poly(dI-dC)•poly(dI-dC) was from Pharmacia Biotech. Ficoll (MW 400,000) was purchased from Sigma. T4 polynucleotide kinase, *Eco*RI, *Pvu*II, and DNase I were from Boehringer Mannheim. *Afl*II and *Fsp*I were purchased from New England Biolabs. Sequenase (version 2.0) was obtained from United States Biochemical. [α-^{32}P]-Thymidine-5'-triphosphate (≥ 3000 Ci/mmol), [α-^{32}P]-deoxyadenosine-5'-triphosphate (≥ 6000 Ci/mmol), and were purchased from Du Pont/NEN. [γ-^{32}P]-Adenosine-5'-triphosphate (≥ 7000 Ci/mmol) was obtained from ICN. Standard methods were used for all DNA manipulations *(51)*. GCN4 (222-281) prepared by solid phase synthesis was generously provided by Martha G. Oakley *(29, 35)*. MBHA resin (0.57 mmol/g) was from Calbiochem. Boc-β-Ala, Boc-γ-aminobutyric acid, Boc-(Tos)Arg, Boc-Ala, Boc(CBz)Lys, Boc-Gly, Boc-Pro and Boc-D-Pro were from Peptides International. *p*-Cresol was purchased from Aldrich. All other chemicals, as well as the purification and characterization of polyamides were as previously described *(44)*. Arg-Pro-Arg-polyamide synthesis is exemplified here for polyamide **3**.

ImPyPyPy-γ-PyPyPyPy-β-RPR (compound 3)

ImPyPyPy-γ-PyPyPyPy-β-RPR-MBHA-resin was synthesized in a stepwise fashion by machine-assisted solid phase methods *(44)* from MBHA resin (600 mg, 0.57 mmol/g, calculated as $L_{new}(mmol/g) = L_{old}/(1 + L_{old}(W_{new} - W_{old}) \times 10^{-3})$, where L is the loading (mmol of amine per gram of resin) and W is the weight (g mol^{-1}) of the growing peptide attached to the resin *(52)*). A sample of polyamide resin (300 mg, 0.30 mmol/g) was placed in a Kel-F reaction vessel, *p*-cresol (1 g) added, and the vessel cooled to -60 °C. HF was then condensed into the vessel. The solution was stirred vigorously for one hour

(0 °C) and the excess HF was removed *in vacuo*. The reaction mixture was then treated with cold ethyl ether (50 mL) and the resulting resin/polyamide coprecipitate was collected by vacuum filtration. The polyamide was then extracted with CH_3CN:H_2O:TFA (10:9:1), 0.1% (w/v) TFA added (6 mL) and the resulting solution purified by reverse phase HPLC using a Waters DeltaPak 25 × 100 mm 100 μm C_{18} column in 0.1% (w/v) TFA, gradient elution 0.25%/min. CH_3CN. ImPyPyPy-γ-PyPyPyPy-β-RPR-NH_2 was recovered upon lyophilization of the appropriate fractions as a white powder (84 mg, 60% recovery); UV (H_2O) λ_{max} 244, 312 (66,000); ^{1}H NMR (DMSO-d_6): δ 10.49 (s, 1H), 9.96 (s, 1H), 9.93 (s, 1H), 9.90 (s, 3H), 9.84 (s, 1H), 8.19 (d, 1H, J=7.6 Hz), 8.07 (m, 1H), 8.02 (m, 1H), 7.95 (m, 1H), 7.91 (m, 1H), 7.53 (m, 3H), 7.38 (s, 1H), 7.26 (d, 1H, J=1.6 Hz), 7.25 (s, 1H), 7.21 (d, 1H, J=1.6 Hz), 7.20 (s, 1H), 7.19 (s, 1H), 7.14 (m, 3H), 7.03 (m, 4H), 6.87 (d, 1H, J=1.6 Hz), 6.85 (m, 2H), 4.48 (t, 1H, J=4.8 Hz), 4.27 (q, 1H, J=4.4 Hz), 4.77 (m, 1H), 3.96 (s, 3H), 3.81 (m, 12H), 3.80 (s, 3H), 3.77 (s, 3H), 3.76 (s, 3H), 3.60 (m, 2H), 3.32 (q, 2H, J=4.8 Hz), 3.17 (q, 2H, J=6.1 Hz), 3.0 (m, 4H), 2.36 (t, 2H, J=6.9 Hz), 2.25 (t, 2H, J=6.9 Hz), 2.0 (m, 2H), 1.77 (m, 4H), 1.64 (m, 2H), 1.46 (m, 6H).

Preparation of ^{32}P-labeled synthetic duplex DNA for gel mobility shift assay

Synthetic DNA fragments were prepared on an ABI 380B Automated DNA Synthesizer and purified by preparative denaturing polyacrylamide gel electrophoresis. ARE-1 and ARE-2, 5'-CCGGATCCATGGTTGCTGACTAATTGTTATCCTCTAGAGTCGACC-3' and 5'- CCGGATCCATGGTTGCTGACTAATTGGTCTCCTCTAGAGTCGACC -3', respectively, were radiolabeled at the 5'-terminus with γ-^{32}P-ATP and T4 polynucleotide kinase, annealed to an equimolar amount of the unlabeled complement, and purified by nondenaturing polyacrylamide gel electrophoresis *(51)*.

Gel mobility shift assay

For gel mobility shifts, polyamide was incubated with radiolabeled synthetic DNA duplex (10 kcpm) in 40 mL reaction volumes of bisTris (10 mM, pH 7.0), NaCl (100 mM), DTT (1 mM), EDTA (1 mM), and poly(dI-dC)•poly(dI-dC) (50 μg/mL) for 16 hours at 22 °C (20 mM MOPS, pH 7.0, 140 mM KCl, 10 mM NaCl, 1mM $MgCl_2$, 1 mM spermine was used to model ionic conditions *in vivo*). GCN4 (222-281) was added and equilibrated for 30 minutes. Loading buffer (15% Ficoll, 0.025% bromophenol blue) (10 mL) was added and 10 mL was immediately loaded onto a running 8% (29:1, acrylamide:bis-acrylamide) polyacrylamide gel (0.5 X TBE, 280 V, 0.8 mm, 13 cm). Sufficient separation of the free DNA and the DNA•GCN4 (222-281) complexes was achieved within 45 minutes. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics) *(53)*.

Preparation of pARE-1 and pARE-2 for affinity cleavage studies

The plasmids pARE-1 and pARE-2 were constructed by ligating the following oligonucleotides into the *Ava*I/*Sal*I fragment of pUC19: 5'-CCGGATCCATGGTTGCTGACTAATTG**N_1**T**N_2**TCCTCTAGAGTCGACCCTGGGA ATACCAGGTGTCGTATCTTAAGAG-3' and 5'-TCGACTCTTAAGATACGACA CCTGGTATTCCCAGGGTCGACTCTAGAGGA**N_3**A**N_4**CAATTAGTCA GCAACCATGGAT-3', where N_1, N_2, N_3 and N_4 are T, A, T and A, respectively for pARE-1 and G, C, G, and C, respectively for pARE-2. The *Eco*RI/*Pvu*II restriction fragments of pARE-1 and pARE-2 were 3'-^{32}P-end-labeled as described below. 3'-^{32}P-End-labeling was accomplished by sequential treatment with *Eco*RI, snake venom phosphatase, polynucleotide kinase with [γ-^{32}P]-adenosine-5'-triphosphate, and *Pvu*II followed by isolation with nondenaturing gel electrophoresis.

Affinity cleavage experiments

Affinity cleavage with Fe(II)•EDTA-GCN4 (222-281) was performed in a final volume of 40 mL with final conditions of 10 mM bisTris, pH 7.0, 100 mM NaCl, 50 μg/mL poly(dI-dC)•poly(dI-dC), 200 nM Fe(II)•EDTA-GCN4 (222-281), 1 mM sodium ascorbate, 0.06% hydrogen peroxide and 20 kcpm end-labeled restriction fragment. Polyamide was allowed to equilibrate with the DNA for 12 hours. After 30 min. of incubation of Fe(II)•EDTA-GCN4 (222-281), sodium ascorbate and hydrogen peroxide were added and 45 min. later the cleavage was stopped by ethanol precipitation. The samples were prepared and resolved by denaturing polyacrylamide gel electrophoresis as described below. Polyamide affinity cleavage was performed by incubating ImPyPyPy-γ-PyPyPyPy-β-RPRK(EDTA) with radiolabeled fragment at the initial solution conditions described above for 24 hours. Ferrous ammonium sulfate and DTT were added to a final concentrations of 10 mM and 10 mM, respectively, and cleavage was allowed to proceed for 30 min. Samples were ethanol precipitated then prepared and resolved as described below.

Preparation of ^{32}P-labeled restriction fragments for DNaseI footprinting

The *Afl* II/*Fsp* I restriction fragment of pJT8 *(1)* was 3'-^{32}P-end-labeled by digesting the plasmid with *Afl*II and *Fsp*I and simultaneously filling in using Sequenase, [α-^{32}P]-deoxyadenosine-5'-triphosphate, and [α-^{32}P]-thymidine-5'-triphosphate, and isolating the 229 bp fragment by nondenaturing gel electrophoresis. The 250 bp *Eco*RI/*Pvu*II restriction fragment of pJK6 *(40)* was prepared in an analogous manner. A and G sequencing were carried out as described *(54-55)*.

Quantitative DNase I footprint titration experiments

All reactions were executed in a total volume of 400 μL. A polyamide stock solution or H_2O (for reference lanes) was added to an assay buffer containing radiolabeled restriction fragment (20 kcpm), affording final solution conditions of 10 mM Tris•HCl (pH 7.0), 10

mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, and either (i) 0.001 nM-100 nM polyamide or (ii) no polyamide (for reference lanes). The solutions were allowed to equilibrate at 22 °C for 18h. Footprinting reactions were initiated by the addition of 4 μL of a DNase I stock solution (at the appropriate concentration to give ~ 55% intact DNA) containing 1 mM dithiothreitol and allowed to proceed for seven min. at 22 °C. The reactions were stopped by the addition of 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. Reactions were resuspended in 1X TBE/80% formamide loading buffer, denatured by heating at 85 °C for 10 min., and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% cross-link, 7 M urea) in 1X TBE at 2000 V for 1.5 h. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics) *(53)*.

Quantitation and data analysis

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

$$\theta_{app} = 1 - \frac{Isite/Iref}{Isite°/Iref°} \qquad (1)$$

where I^o_{site} and I^o_{ref} are the site and reference intensities, respectively, from a control lane to which no polyamide was added. The ($[L]_{tot}$, θ_{app}) data points were fit to a Langmuir binding isotherm (eq. 2, n=1) by minimizing the difference between θ_{app} and θ_{fit}, using the modified Hill equation:

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

where $[L]_{tot}$ is the total polyamide concentration, K_a is the equilibrium association constant, and θ_{min} and θ_{max} are the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. The data were fit using a nonlinear least-squares fitting procedure with K_a, θ_{max}, and θ_{min} as the adjustable parameters. All acceptable fits had a correlation coefficient of $R > 0.97$. At least three sets of data were used in determining each association constant. All lanes from each gel were used unless visual inspection revealed a data point to be obviously flawed relative to neighboring points.

Acknowledgments

We are grateful to the National Institutes of Health General Medical Sciences for research support, to the National Science Foundation for a predoctoral fellowship to R.E.B., and to the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B.

References

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

2. Swalley, S. E., Baird, E. E. and Dervan, P. B. (1997) *J. Am. Chem. Soc. 119*, 6953-6961.

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

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

5. Mrksich, M., Wade, W. S., Dwyer, T. J., Geirstanger, B. H., Wemmer, D. E. and Dervan, P. B (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7586-7590.

6. Wade, W. S., Mrksich, M. and Dervan, P. B. (1993) *Biochemistry 32*, 11385-11389.

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

8. White, S., Baird, E. E. and Dervan, P.B. (1997) *Chem. Biol. 4*, 569-578.

9. White, S., Baird, E. E. and Dervan, P.B. (1997) *J. Am. Chem. Soc. 119*, 8756-8765.

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

11. Chen, X., Ramakrishnan, B., Rao, S. T. and Sundaralingham, M. (1994) *Nature Struct. Biol. 1*, 169-175.

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

13. Pilch, D. S., *et al.*, and Dervan, P. B. (1996) *Proc. Natl. Acad. Sci. USA 93*, 8306-8311.

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

15. Gottesfeld, J. M., Nealy, L., Trauger, J. W., Baird, E. E. and Dervan, P. B. (1997) *Nature 387*, 202-205.

16. Neely, L., Trauger, J. W., Baird, E. E., Dervan, P. B. and Gottesfeld, J. M. (1997) *J. Mol. Biol. 274*, 439-445.

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

18. Thuong, N. T. and Helene, C. (1993) *Angew. Chem. Int. Ed. Engl. 32*, 666-690.

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

20. Duval-Valentin, G., Thuong, N.T., and Helene, C. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 504-508.

21. Nielsen, P. E. (1997) *Chem. Eur. J. 3*, 505-508.

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

23. Liu, C. *et al.*, and Vogt, P. K. (1996) *Proc. Natl. Acad. Sci. USA 93*, 940-944.

24. Bruice, T. C., Mei. H.-Y., He, G.-X. and Lopez, V. (1992) *Proc. Natl. Acad. Sci. USA 89*, 1700-1704.

25. Chiang, S.-Y., Bruice, T. C., Azizkhan, J.C., Gawron, L. and Beerman, T. A. (1997) *Proc. Natl. Acad. Sci. USA 94,* 2811-2816.

26. Bruice, T. C., Sengupta, D., Blasko, A., Chiang, S.-Y. and Beerman, T. A. (1997) *Bioorg. Med. Chem. 5*, 685-692.

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

28. Kim, Y., Geiger, J. H., Hahn, S. and Sigler, P. B. (1993) *Nature 365*, 512-520.

29. Oakley, M. G. and Dervan, P. B. (1990) *Science 248*, 847-850.

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

31. König, P. and Richmond, T. (1993) *J. Mol. Biol. 233*, 139-154.

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

33. Feng, J-A., Johnson, R. C. and Dickerson, R. E. (1994) *Science 263*, 348-355.

34. Kielkopf, C. L., Baird, E. E., Dervan, P.B. and Rees, D.C. (1998) *Nature Struct. Biol.*, *5*, 104-109.

35. Oakley, M. G., Mrksich, M. and Dervan, P. B. (1992) *Biochemistry 31*, 10969-10975.

36. Park, Y. W. and Breslauer, K. J. (1992) *Proc. Natl. Acad. Sci. USA 89*, 6653-6657.

37. Gehring, W. J., *et al.*, and Wüthrich, K. (1994) *Cell 78*, 211-223.

38. Gehring, W. J., Affolter, M. and Burglin, T. (1994) *Annu. Rev. Biochem. 63*, 487-526.

39. Hurst, H. C. (1995) *Protein Profile 2*, 105-168.

40. Struhl, K. (1992) Yeast GCN4 transcriptional activator protein. In *Transcriptional Regulation*. (McKnight, S. L. and Yamamoto, K. R., eds), pp. 833-859, Cold Spring Harbor Laboratory Press, New York.

41. Curran, T. and Vogt, P. (1992) Dangerous liasions: fos and jun, oncogenic transcription factors. In *Transcriptional Regulation*. (McKnight, S. L. and Yamamoto, K. R., eds), pp. 797-832, Cold Spring Harbor Laboratory Press, New York.

42. Hope, I. A. and Struhl, K. (1986) *Cell 46*, 885-894.

43. Paolella D. N., Palmer, C. R. and Schepartz, A. (1994) *Science 264*, 1130-1133.

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

45. Jones, R. J., Lin, K.-Y., Milligan, J.F., Wadawani, S. and Matteucci, M.D. (1993) *J. Org. Chem. 58*, 2983-2991.

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

47. Kelly, J. J., Baird, E. E. and Dervan, P.B. (1993) *Proc. Natl. Acad. Sci., USA 93*, 6981-6985.

48. Strauss, J. K. and Maher, L. J. (1994) *Science 266*, 1829-1834.

49. Welch, J. J., Rauscher III, F. J., and Beerman, T. A. (1994) *J. Biol. Chem. 269*, 31051-31058.

50. Breslauer, K. J., Ferrante, R., Marky, L.A., Dervan, P. B. and Youngquist, R. S. (1988) The origins of the DNA binding affinity and specificity of minor groove directed ligands: correlations of thermodynamic and structural data. In*Structure and Expression (Vol. 2), DNA and Its Drug Complexes* (Sarma, R. H. and Sarma, M. H. eds), pp. 273-289, Academic Press.

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

52. Barlos, K., Chatzi, O., Gatos, D. and Stravropoulos, G. (1991) *Int. J. Peptide Protein Res. 37*, 513-520.

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

54. Maxam, A.M. and Gilbert, W. S. (1980) *Methods Enzymol. 65*, 499-560.

55. Iverson, B. L. and Dervan, P. B. (1987) *Methods Enzymol. 15*, 7823-7830.

56. Trauger, J. W., Baird, E. E. and Dervan, P. B. (1996) *Chem. Biol. 3*, 369-377.

Chapter 3

Hairpin Polyamides Containing *N*-Aminoalkylpyrrole Substitutions: Recognition of the DNA Minor Groove and Inhibition of a Major Groove Binding Protein

The work described in this chapter was a collaboration with Nicholas R. Wurtz, Jason W. Szewczyk & Prof. Peter B. Dervan.

Abstract: Pyrrole/imidazole polyamides are cell permeable synthetic ligands that bind predetermined DNAsequences. Polyamides offer a potentially general chemical approach to gene regulation, provided a diverse set of DNA sequences can be targeted with subnanomolar affinity and inhibition of DNA binding for a variety of transcription factors can be achieved. We report here a new class of hairpin polyamides containing a dimethylaminopropyl (Dp) moiety at the N-1 of a single pyrrole residue that targets DNA with increased affinity and uncompromised specificity relative to isomers containing the Dp at the polyamide C-terminus. The affinity enhancement may result from a contact of the positively charged 1-(dimethylaminopropyl)pyrrole with the phosphate backbone or a nonspecific electrostatic mechanism. A general modification that increases DNA binding affinity without compromising specificity may prove useful for several applications, including enhancement of affinities for weak binding polyamide:DNA complexes and the development of polyamides with new functions. Polyamides displaying an Arg-Pro-Arg tripeptide derived from a natural protein:DNA complex have been shown previously to inhibit proteins bound exclusively in the major groove, potentially by directing the positive charges to the DNA backbone and competing with the protein side chains for contacts to the phosphates. Similar protein inhibition has been accomplished here with a positively charged diaminoalkyl chain at the N-1 of a single pyrrole residue. The polyamide modification necessary for major groove protein inhibition has been reduced from a natural α-amino acid tripeptide to a simple synthetic diamine. By altering the presentation of the aminoalkyl moiety, a new class of polyamides with enhanced affinity have been developed that may expand the functional utility of polyamides as regulators of gene expression.

Introduction

Hairpin polyamides containing pyrrole (Py) and imidazole (Im) amino acids are cell permeable synthetic ligands that bind predetermined DNA sequences in the minor groove with affinities and specificities comparable to many DNA binding proteins (*1*). A set of rules have been developed that allow for sequence specific recognition of the DNA minor groove by relating each Watson Crick base pair with a particular pairing of the Py and Im rings (*2, 3*). The crescent shaped Py/Im polyamides bind in the minor groove of DNA with pairs of aromatic rings stacked against each other and the walls of the groove, allowing the backbone amide hydrogens and the substituents at the 3-position of the Py and Im residues to make specific contacts with the edges of the base pairs (*4-6*). A G•C base pair is recognized by a side-by-side pairing of Im with Py (Im/Py), while a C•G base pair is targeted by a Py/Im pairing (*2, 4, 7*). The Py/Py pair is degenerate and binds both A•T and T•Abase pairs in preference to G•C and C•G (*2, 4, 8*). The discrimination of T•A and A•T base pairs using 3-hydroxypyrrole (Hp)/Py and Py/Hp pairs, respectively, completes the four base pair code (*3, 5*). In practice, an γ-aminobutyric acid residue (γ) connects the polyamide subunits in a "hairpin" motif and enhances affinity >100-fold relative to the unlinked dimers (*6, 9*). A C-terminal β-alanine (β) increases both affinity and specificity and facilitates solid-phase synthesis (*10-12*).

Recognition of the DNA Minor Groove by N-Aminoalkylpyrrole Polyamides

Hairpin polyamides targeted to the DNA sequences adjacent to transcriptional activators have been shown to be cell permeable and to specifically inhibit gene transcription in cell culture (*13, 14*). Polyamides offer a potentially general approach to gene regulation, provided that polyamide motifs can be developed to target a variety of DNA sequences with the subnanomolar affinity necessary to compete with DNA binding proteins. Toward this end, polyamides that target longer sequences and the covalent linking of polyamide subunits in the hairpin conformation have led to second generation

designs with both enhanced DNA binding affinity and specificity (*9, 15-17*). The development of an α-amino substitution in the γ-turn also increased polyamide affinity and allowed the utility of "cycle" polyamides to be reinvestigated (*18-20*). Previously, the cycle polyamide had been designed as a 6-ring hairpin polyamide with an additional γ-turn linking the free N-terminus and C-terminus (*19*). The propylamino charge, placed at the C-terminus of the hairpin polyamide in analogy to distamycin, was attached at the N-1 of a pyrrole residue in the cycle polyamide. Interestingly, this six-ring cycle exhibited ≈10-fold enhanced affinity but lowered specificity relative to the hairpin analog. Possible explanations of these results are that the cyclization of the polyamide provides an entropic advantage toward DNA binding or the new N-aminopropyl moiety makes an electrostatic interaction with the polyanionic DNA phosphate backbone or a combination of both, coupled with a loss in the ability to discriminate against mismatch sites. Subsequent investigations have revealed that 8-ring cycle polyamides containing the α-amino-γ-turn also exhibit ≈10-fold higher affinity for their match site and *enhanced* specificity relative to the hairpin analog, demonstrating a significant advantage for polyamide cyclization (*20*). The effects of the N-aminopropylpyrrole on the DNAbinding affinity and specificity of a hairpin polyamide had yet to be determined.

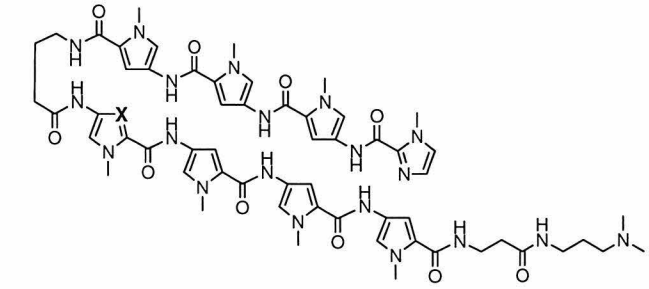
The addition of a positively charged alkyl amine moiety designed to interact with the phosphate backbone to other synthetic DNA binding ligands including a triple helix forming oligonucleotide and the minor groove binder PyPyPy resulted in an increase in DNA binding affinity for these ligands, as well as an induced bend in the DNA helix, in the latter case (*21-23*). Similarly, the placement of an aminoalkyl substituent at the N-1 of a pyrrole residue in a hairpin polyamide will result in a positively charged protonated amine that may increase the DNAbinding affinity of the polyamide through a contact with the phosphate backbone or a nonspecific electrostatic mechanism. Such an effect may be generally applicable to many polyamides, independent of sequence.

Inhibition of a Major Groove Protein

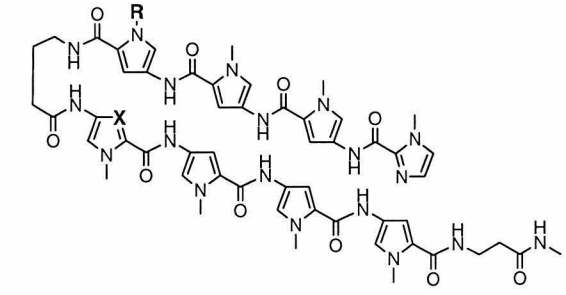
A further requirement for the general application of polyamides as regulators gene transcription is polyamide motifs that can effectively inhibit DNA binding for a variety of classes of transcription factors. Polyamides have been found to interfere with protein-DNA recognition in cases where contacts in the *minor* groove are important for protein-DNAbinding affinity (*13, 14, 24*). In contrast, Py/Im polyamides have been shown to bind simultaneously with a variety of proteins that exclusively occupy the DNA *major* groove (*24-26*). The ubiquity of major groove contacts in protein:DNA recognition provides the impetus to develop approaches for the inhibition of major groove proteins by polyamides that bind in the minor groove. Hairpin polyamides with a C-terminal Arg-Pro-Arg positive patch have been shown to inhibit a protein bound exclusively in the major groove, potentially by competing with the protein side chains for electrostatic contacts to the DNA phosphate backbone (*27*). The general utility of Arg-Pro-Arg polyamides is limited by both the requirement that the positive patch be placed on the C-terminus and the use of α-amino acids which may be a disadvantage in terms of cell permeability and biostability. The N-aminoalkylpyrrole residue may serve as a simple synthetic substitute for the Arg-Pro-Arg positive patch that can be placed at any position in the polyamide, potentially expanding the transcription factors targetable for inhibition and providing a class of molecules with the necessary cell permeability properties. Other attempts to inhibit major groove binding transcription factors with minor groove binding ligands have required the addition of a multiply positively charged deccaaza dacabutylamine substituent to PyPyPy, greatly adding to the size and charge of the molecule, potentially resulting in a ligand with significantly reduced sequence specificity (*28*).

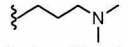
We report here the synthesis and DNA binding properties of a series of hairpin polyamides containing N-aminoalkyl substitutions at a single pyrrole residue and their ability to inhibit DNAbinding by an exclusively major groove binding protein. Two 8-ring hairpins were prepared with a *N*-(*N'*, *N'*-dimethylaminopropyl)pyrrole residue and an N-

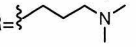
methyl amide at the C-terminus to provide compounds that are structural isomers of those reported previously, differing only in the placement of the substituents (Figure 3.1) (*29*). The synthetic methodology was extended to afford a series of polyamides containing multiple amines and selected ether substitutions in the alkyl moiety of the N-1 position of pyrrole. The DNA binding affinity and specificity of each of these compounds was evaluated using quantitative DNase I footprinting and compared to analogs containing the traditional N-methyl substituents. Gel mobility shift assays were employed to investigate the ability of these polyamides to inhibit DNA binding by a major groove binding protein, GCN4 (222-281), a member of the bZIP family of transcriptional activators.

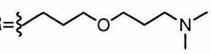


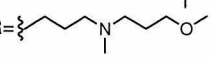
1) X=N, ImPyPyPy-γ-ImPyPyPyPy-β-Dp
2) X=CH, ImPyPyPy-γ-PyPyPyPyPy-β-Dp

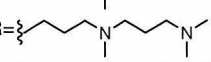


3) X=N R=  ImPyPyPy(Dp)-γ-ImPyPyPy-β-Me

4) X=CH, R=  ImPyPyPy(Dp)-γ-PyPyPyPy-β-Me

5) X=CH, R=  ImPyPyPy(C_3ODp)-γ-PyPyPyPy-β-Me

6) X=CH, R=  ImPyPyPy(DpC_3OMe)-γ-PyPyPyPy-β-Me

7) X=CH, R=  ImPyPyPy(DpDp)-γ-PyPyPyPy-β-Me

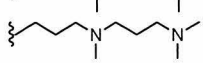
8) X=N, R=  ImPyPyPy(DpDp)-γ-ImPyPyPy-β-Me

Figure 3.1. Structures of 8-ring hairpin polyamides containing N-methylpyrrole residues and a C-terminal Dp tail (top) and polyamides with N-aminoalkylpyrrole residues and a C-terminal N-methylamide. ImPyPyPy-γ-ImPyPyPyPy-β-Dp (**1**), and ImPyPyPy-γ-PyPyPyPyPy-β-Dp (**2**) (top). ImPyPyPy(Dp)-γ-ImPyPyPy-β-Me (**3**), ImPyPyPy(Dp)-γ-PyPyPyPy-β-Me (**4**), ImPyPyPy(C_3ODp)-γ-PyPyPyPy-β-Me (**5**), ImPyPyPy(DpC_3OMe)-γ-PyPyPyPy-β-Me (**6**), ImPyPyPy(DpDp)-γ-PyPyPyPy-β-Me (**7**) and ImPyPyPy(DpDp)-γ-ImPyPyPy-β-Me (**8**).

Results

Polyamide Synthesis

The synthesis of N-aminoalkylpyrrole-containing polyamides required the preparation of a new pyrrole monomer suitable for solid phase synthesis and subsequent modification at the N-1 position. (3-Hydroxypropyl)-4-[(*tert*-butoxycarbonyl)amino]-pyrrole-2-carboxylic acid, (**11**), introduces a 3-hydroxypropyl moiety at the N-1 of pyrrole which can be modified by sulfonylation followed by nucleophilic displacement to afford a variety of derivatives from a single polyamide synthesis. By cleaving the polyamide from resin with methylamine, polyamides (**3** and **4**) can be obtained that are structural isomers of those previously reported (**1** and **2**) and limit the overall charge of the N-diaminoalkylpyrrole-containing polyamides (**7** and **8**) (*29*). Ethyl 4-nitropyrrole-2-carboxylate (**9**) is available in 500 g quantities as described (Figure 3.2) (*26*). Alkylation

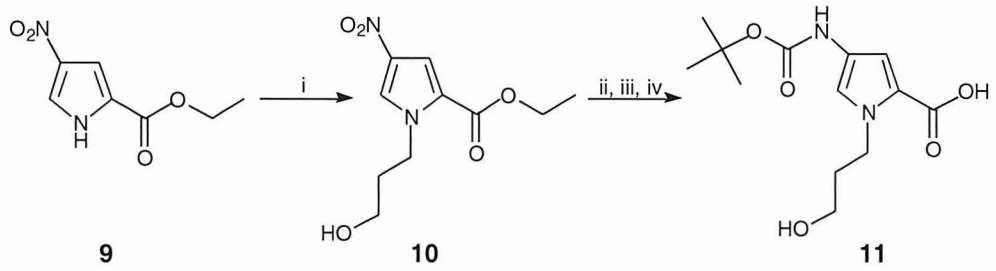


Figure 3.2. Synthesis of the Boc-*N*-(3-hydroxypropyl)pyrrole-acid monomer (**11**) for solid phase synthesis. (i) Iodopropanol, acetone, K_2CO_3. (ii) H_2, 10% Pd/C, DMF. (iii) Boc-anhyride, DIEA. (iv) 1 M KOH, EtOH, 70 °C.

of the pyrrole N-1 with 3-iodoproponol provided **10**. Reduction and Boc protection of the amine followed by hydrolysis of the ethyl ester afforded **11** in gram quantities in 53% yield. The *N*-(3-hydroxypropyl)pyrrole monomer **11** was used without further protection in the Boc-chemistry machine assisted solid phase synthesis of polyamides **12** and **13** from Boc-β-Ala-PAM resin (Figure 3.3) (*12*). In order to provide polyamides that differ from those previously reported only by the placement of the positively charged N,N-dimethylaminopropyl moiety and the N-methyl group, it was necessary to cleave the polyamide from resin using methylamine. This was accomplished by condensing

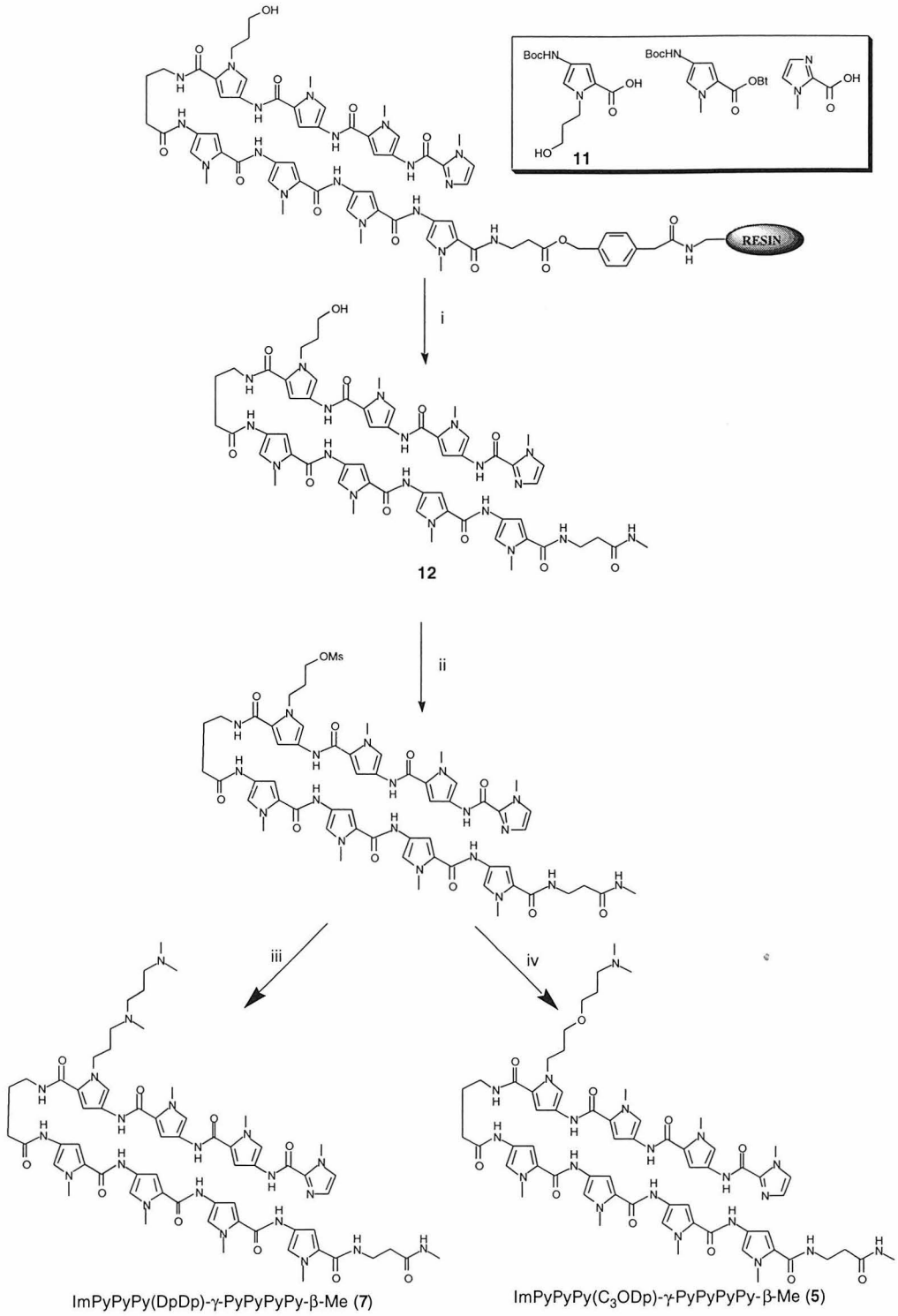


Figure 3.3. Synthetic scheme for solid phase preparation of *N*-aminoalkylpyrrole containing polyamides *(12)*. Cycling protocols consisted of trifluoroacetic acid (TFA) deprotection followed by coupling with HOBt-activated Boc-pyrrole, Boc-imidazole, Boc-*N*-(3-hydroxypropyl)pyrrole (**11**) or aliphatic amino acid ester. (i) The PAM resin is then cleaved by treatment with methylamine. (ii) Activation with methanesulfonyl chloride or toluenesulfonyl chloride followed by reaction with the appropriate amine (iii) or alkoxide (iv) and reverse phase HPLC provided the desired products. The syntheses of ImPyPyPy(DpDp)-γ-PyPyPyPy-β-Me (**7**) ImPyPyPy(C_3ODp)-γ-PyPyPyPy-β-Me (**5**) are shown as representative examples.

methylamine in a Parr apparatus with resin-bound polyamide, followed by heating overnight (55 °C) to provide crude polyamides ImPyPyPy(C_3OH)-γ-PyPyPyPy-β-Me, (**12**), and ImPyPyPy(C_3OH)-γ-ImPyPyPy-β-Me, (**13**), containing the N-(3-hydroxypropyl)pyrrole residue suitable for further modification. The free hydroxyl of the crude polyamide was activated with *p*-toluenesulfonyl chloride or methanesulfonyl chloride in pyridine. Nucleophilic displacement of the tosylate or mesylate with the appropriate amine (**1**-**4**, **6**-**8**) or sodium salt (**5**) followed by preparatory reverse phase HPLC provided the desired N-aminoalkyl substituted polyamides in recoveries and purities similar to those observed for N-methyl derivatives cleaved with dimethylaminopropylamine.

N-Aminoalkylpyrrole Substitutions Increase Polyamide Binding Affinity

The DNA binding affinity and specificity for the six base pair target sites were measured for the N-aminoalkylpyrrole-containing polyamides by quantitative DNase I footprinting on the 229-bp *Afl*III/*Fsp*I restriction fragment of pJT8 (*29*). Polyamides of the basic composition ImPyPyPy-γ-ImPyPyPy (**1**, **3** and **8**) were designed to bind the site 5'-AGTACT-3'according to the pairing rules. Conversely, the single atomic substitution that converts an imidazole to a pyrrole residue provides polyamides based on ImPyPyPy-γ-PyPyPyPy (**2**, **4** and **5**-**7**), targeted to 5'-AGTATT-3'. The sites 5'-AGTACT-3' and 5'-AGTATT-3' are for ImPyPyPy-γ-ImPyPyPy polyamides "match" and "single base pair mismatch" sites, respectively, and for ImPyPyPy-γ-PyPyPyPy polyamides "single base pair mismatch" and "match" sites, respectively.

DNase I footprinting revealed that ImPyPyPy(Dp)-γ-ImPyPyPy-β-Me (**3**) bound the 5'-AGTACT-3' match site with an affinity of 1.9 (±0.4) × 10^{11} M^{-1}, and a 56-fold preference over the single base pair mismatch site (K_a=3.3 (±0.4) × 10^9 M^{-1}) (Figure 3.4A, Table 3.1). This represents ≈10-fold increase in affinity relative to the parent polyamide differing in the placement of the dimethylaminopropyl and N-methyl moieties,

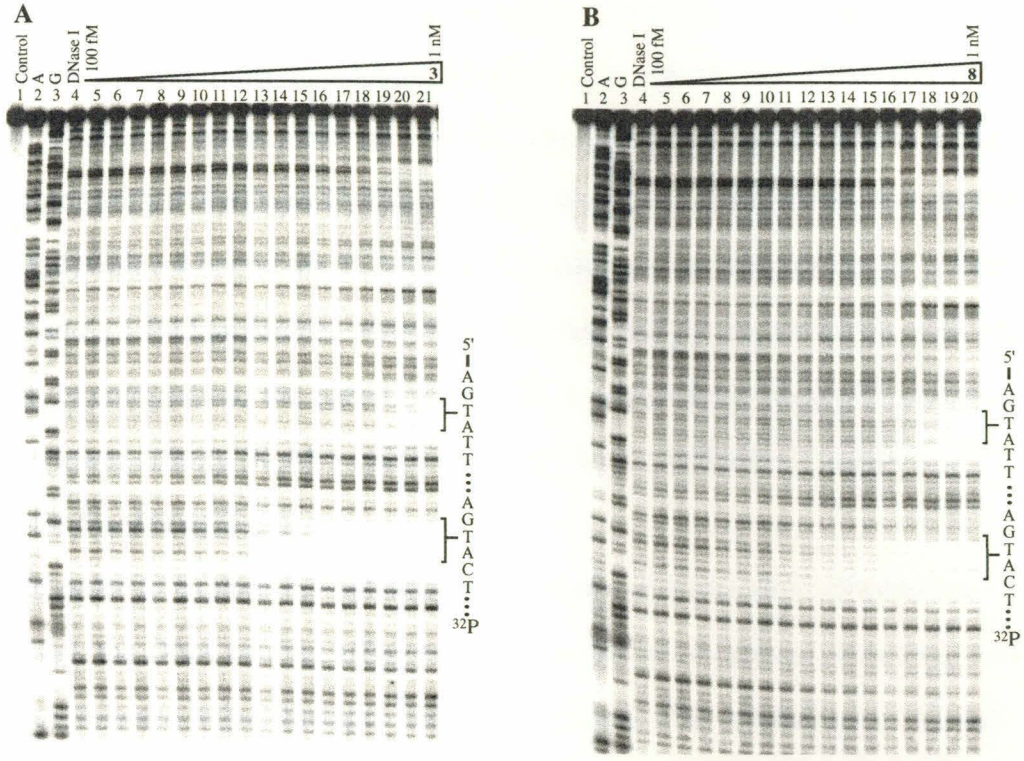


Figure 3.4. Storage phosphor autoradiograms of quantitative DNase I footprint titration experiments with (A) ImPyPyPy(Dp)-γ-ImPyPyPy-β-Dp (**3**) and (B) ImPyPyPy(DpDp)-γ-ImPyPyPy-β-Dp (**8**) on the 3'-^{32}P-end-labeled *Afl*III/*Fsp*I restriction fragment from pJT8 *(29)*. All reactions contained 8 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$ and were performed at 22 °C. Lane 1, intact DNA; lane 2, A-specific reaction; lane 3, G-specific reaction; lane 4, DNase I standard; lanes 5-21, 100 fM, 200 fM, 500 fM, 1 pM, 1.5 pM, 2.5 pM, 4.0 pM, 6.5 pM, 10 pM, 15 pM, 25 pM, 40 pM, 65 pM, 100 pM, 200 pM, 500 pM, 1 nM, respectively of **3** or **8**. The 65 pM lane has been omitted for **8**. The positions of the match (5'-AGTACT-3') and single base pair mismatch (5'-AGTA<u>T</u>T-3') sites are indicated.

Table 3.1. Equilibrium Association Constants (M^{-1})[a]

Polyamide	Match	Mismatch
ImPyPyPy-γ-ImPyPyPy-β-Dp (**1**)	1.4×10^{10}	2.3×10^{8}
ImPyPyPy(Dp)-γ-ImPyPyPy-β-Me (**3**)	1.9×10^{11}	3.3×10^{9}
ImPyPyPy-γ-PyPyPyPy-β-Dp (**2**)	1.5×10^{9}	2.4×10^{8}
ImPyPyPy(Dp)-γ-PyPyPyPy-β-Dp (**4**)	1.3×10^{10}	2.9×10^{9}

[a]Values reported are the mean values obtained from at least three DNase I footprint titration experiments. The assays were carried out at 22 ¡C, 10 mM Tris¥HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. The match and mismatch sites for **1** and **3** are 5'-ttAGTACTtg-3' and 5'-ttAGTA<u>T</u>Ttg-3', respectively. The match and mismatch sites for **2** and **4** are 5'-ttAGTATTtg-3' and 5'-ttAGTA<u>C</u>Ttg-3', respectively.

ImPyPyPy-γ-ImPyPyPy-β-Dp (**1**) (K_a=1.4 (±0.5) × 10^{10} M^{-1} for 5'-AGTACT-3'), and uncompromised specificity (61-fold, K_a=2.3 (±0.5) × 10^8 M^{-1} for 5'-AGTATT-3'). The increase in affinity upon placement of the alkylamine on the pyrrole ring was also observed for ImPyPyPy(Dp)-γ-PyPyPyPy-β-Me (**4**) relative to ImPyPyPy-γ-PyPyPyPy-β-Dp (**2**). Polyamide **4** bound the 5'-AGTATT-3' match site with an affinity of 1.3 (±0.6) × 10^{10} M^{-1} and a 5-fold preference over the single base pair mismatch site (K_a=2.9 (±0.6) × 10^9 M^{-1}). In comparison, the parent polyamide, **2**, bound both the match and mismatch sites with lower affinity (K_a=1.5 (±0.3) × 10^9 M^{-1} for 5'-AGTATT-3' and K_a=2.4 (±1.2) × 10^8 M^{-1} for 5'-AGTACT-3').

In order to explore the potential applications of aminoalkyl substitutions on the pyrrole ring, the DNA binding affinities and specificities of analogs containing multiple amines or selected ether linkages were determined. The presence of an additional aminopropyl moiety to provide a polyamide with potentially two positive charges as in ImPyPyPy(DpDp)-γ-ImPyPyPy-β-Me (**8**), resulted in an additional slight increase in affinity for the match site relative to **3** (K_a=3.0 (±0.4) × 10^{11} M^{-1}), and a 36-fold preference over the mismatch site (K_a=8.2 (±3.1) × 10^9 M^{-1}) (Figure 3.4B, Table 3.2). This represents only a modest decrease in specificity for **8** relative to **3** (36-fold and 56-fold, respectively). A similar decrease in specificity was observed for ImPyPyPy(DpDp)-γ-PyPyPyPy-β-Me

Table 3.2. Equilibrium Association Constants (M^{-1})[a]

Polyamide	Match	Mismatch
ImPyPyPy(C3ODp)-γ-PyPyPyPy-β-Me (**5**)	1.2 × 10^{10}	1.8 × 10^9
ImPyPyPy(DpC3OMe)-γ-PyPyPyPy-β-Me (**6**)	1.5 × 10^{10}	2.2 × 10^9
ImPyPyPy(DpDp)-γ-PyPyPyPy-β-Me (**7**)	6.9 × 10^9	3.1 × 10^9
ImPyPyPy(DpDp)-γ-ImPyPyPy-β-Me (**8**)	3.0 × 10^{11}	8.2 × 10^9

[a]Values reported are the mean values obtained from three DNase I footprint titration experiments. The assays were carried out at 22 °C, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. The match and mismatch sites for **5-7** are 5'-ttAGTATTtg-3' and 5'-ttAGTACTtg-3', respectively. The match and mismatch sites for **8** are 5'-ttAGTACTtg-3' and 5'-ttAGTATTtg-3', respectively.

(**7**) relative to polyamide **4** (K_a=6.9 (±1.2) × 10^9 M^{-1} for 5'-AGTATT-3'and K_a=3.1 (±1.0) × 10^9 M^{-1} for 5'-AGTACT-3'). Substitution of either tertiary amine in **7** to provide the singly charged polyamides ImPyPyPy(C_3ODp)-γ-PyPyPyPy-β-Me (**5**) and ImPyPyPy(DpC_3OMe)-γ-PyPyPyPy-β-Me (**6**) restored the optimal affinity and specificity observed with **4**, while retaining the overall size and length of **7**. Polyamide **5**, containing an internal ether linkage and a terminal dimethylamine, bound the 5'-AGTATT-3' match site with an affinity of 1.2 (±0.1) × 10^{10} M^{-1} with a 7-fold preference over the 5'-AGTACT-3' mismatch (1.8 (±0.9) × 10^9 M^{-1}). Similarly, the introduction of an internal amine and a terminal methyl ether in **6** resulted in affinities of 1.5 (±0.3) × 10^{10} M^{-1} and 2.2 (±0.7) × 10^9 M^{-1} (7-fold specificity) for the match and mismatch sites, respectively.

N-aminoalkylpyrrole Polyamides Inhibit a Major Groove Binding Protein

The ability of N-aminoalkylpyrrole-containing polyamides to inhibit DNAbinding by GCN4 (222-281) was investigated using gel mobility shift assays. The carboxy-terminal 60 amino acids of GCN4 (222-281) contain the "leucine zipper" dimerization domain and the basic region DNA binding domain and have been shown to be sufficient for sequence specific DNA binding (*30, 31*). GCN4 (222-281) binds its target site as a dimer of α-helicies that make specific hydrogen bonds, van der Waal contacts and phosphate interactions in the DNA major groove (Figure 3.5A,B) (*32, 33*).

Analysis of an X-ray crystal structure of a polyamide bound to its cognate DNA site reveals that the N-1 substituent of each pyrrole ring makes its closest phosphate contact two phosphates to the 3' side of the base the ring pair recognizes (*4*). Specifically, in Figure 3.5D, the N-1 of the Py residue immediately to the N-terminal side of the γ-turn would be closest in space to the phosphate of the 3'-TpT-5'step. This 3'-shift is consistent with the minor groove binding nature of the polyamide. Polyamides displaying an N-aminoalkyl positive patch from this Py ring could thus interfere with the GCN4 contact to this phosphate and consequently inhibit protein binding.

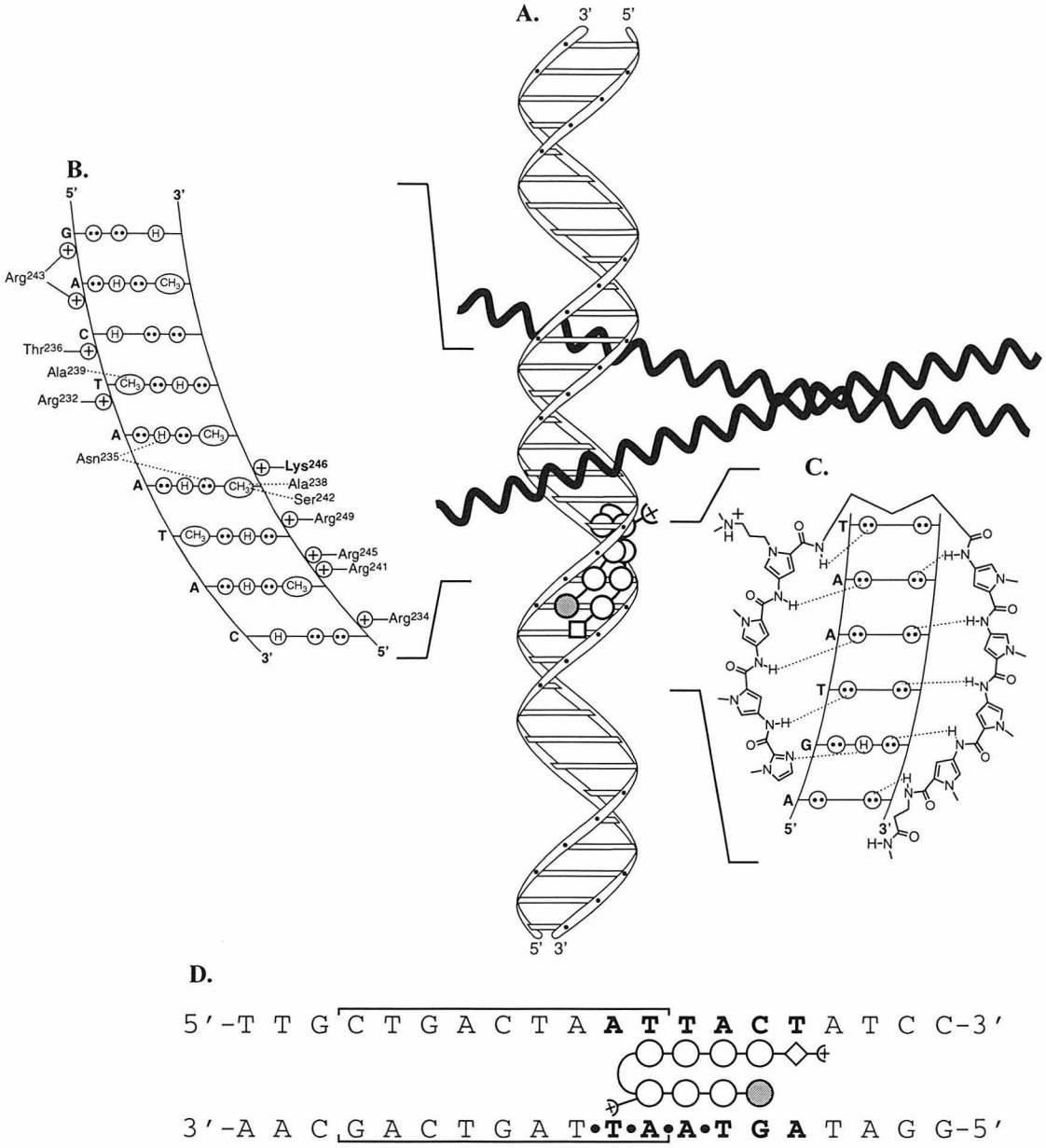


Figure 3.5. A schematic model of N-(aminoalkyl)pyrrole polyamides targeted to the major groove transcription factor, GCN4. (A) The α-helical GCN4 dimer is shown binding to adjacent major grooves *(32)*. The N-(aminoalkyl)pyrrole polyamide is shown as black and white balls which represent imidazole and pyrrole amino acids, respectively. The diamond represents β-alanine; γ-aminobutyric acid is designated as a curved line and the aminoalkyl chain is represented by a plus sign in a circle. (B) The contacts between one GCN4 monomer and the major groove of one half-site of 5'-CTGACTAAT-3' are depicted (adapted from Ref. #32). Circles with two dots represent the lone pairs of the N7 of purines, the O4 of thymine, and the O6 of guanine. Circles containing an H represent the N6 and N4 hydrogens of the exocyclic amines of adenine and cytosine, respectively. The C5 methyl group of thymine is depicted as a circle with CH_3 inside. Plus signs represent protein residues which electrostatically contact the phosphate backbone. (C) The hydrogen bonding model of the eight-ring hairpin polyamide ImPyPyPy(Dp)-γ-PyPyPyPy-β-Me bound to the minor groove of 5'-AGTAAT-3'. Circles with two dots represent the lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogens of guanines. Putative hydrogen bonds are illustrated by dotted lines. (D) The model of the polyamide binding its target site (bold) adjacent to the GCN4 binding site (brackets). Polyamide residues are as in A.

The ImPyPyPy-γ-PyPyPyPy polyamides were designed to bind the 5'-AGTAAT-3' site adjacent to the GCN4 binding site (5'-CTGACTAAT-3') on the synthetic DNA duplex ARE-4. Gel mobility shift assays were performed as previously described (*25, 27, 31*). Remarkably, incubation of the radiolabeled DNA fragment with the N-methyl containing polyamide ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**) resulted in a slight inhibition of protein binding, providing the first evidence that standard Py/Im polyamides can affect the DNA binding of an exclusively major groove binding protein (Figure 3.6A). The placement of *N'*, *N'*-dimethylaminopropyl at the N-1 position of the specified pyrrole in **4** did not offer a significant improvement in GCN4 (222-281) inhibition over **2** (Figure 3.6B). Incorporation of a second positive charge and a longer tether, as in **7**, provided efficient inhibition of GCN4 binding, ≈50% of that originally bound (Figures 3.6C, 3.7). The specificity of GCN4 inhibition was also tested using ImPyPyPy(DpDp)-γ-ImPyPyPy-β-Me (**8**), which is a single base pair mismatch for the polyamide site on ARE-4 and therefore should bind the DNA with significantly reduced affinity relative to **7**. Indeed, polyamide **8** did not inhibit GCN4 binding, demonstrating that the mechanism of GCN4 inhibition requires a polyamide specifically bound in the DNAminor groove (Figure 3.6D).

Optimal GCN4 Inhibition Requires N-diaminoalkylpyrrole

It was important to determine if the basis for GCN4 inhibition by **7** was dependent upon the presence of both potentially charged amines or if perhaps the added length of the second aminopropyl was sufficient for inhibition. Polyamide **5** contains an ether linkage in place of the internal amine in **7**, resulting in a single charge placed distal to the polyamide. This change resulted in a lack of GCN4 inhibition (Figure 3.6E). Furthermore, **6**, containing a single charge internal and a terminal methyl ether, exhibited similar behavior with only slight inhibition on GCN4 binding (Figure 3.6F). Together, these results suggest that both of the positive charges present in **7** are necessary for efficient inhibition of GCN4.

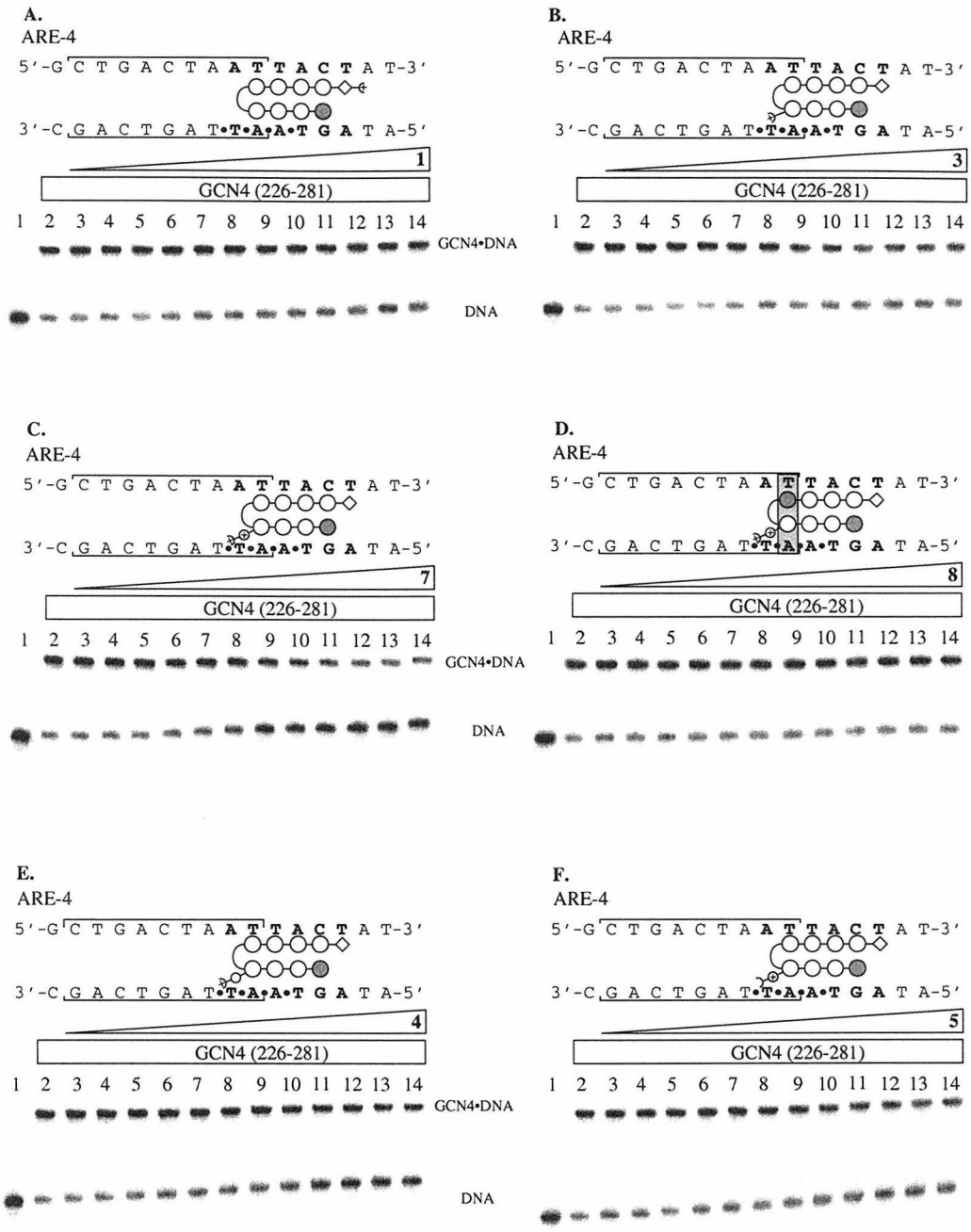


Figure 3.6. GCN4 (222-281) gel mobility shift experiments in the presence of N-methylpyrrole and N-aminoalkylpyrrole polyamides. (A-F) Top, model of polyamide binding the DNA fragment ARE-4. Mismatches are indicated by a grey box. Bottom, storage phosphor autoradiogram of non-denaturing polyacrylamide gel showing GCN4 (222-281) binding to the radiolabeled ARE-4 fragment in the presence of increasing concentrations of polyamide (10 mM bisTris pH 7.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 50 μg/mL poly(dI-dC)•poly(dI-dC), 22 °C). The upper band is the GCN4 (222-281)-DNA complex and the lower band is free DNA. Lane 1 is DNA only. Lane 2 contains DNA incubated with 200 nM GCN4 (222-281). Lanes 3-14 are 200 nM GCN4 (222-281) and 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, and 500 nM polyamide: (A) ImPyPyPy-γ-PyPyPyPy-β-Dp (**2**), (B) ImPyPyPy(Dp)-γ-PyPyPyPy-β-Me (**4**), (C) ImPyPyPy(DpDp)-γ-PyPyPyPy-β-Me (**7**), (D) ImPyPyPy(DpDp)-γ-ImPyPyPy-β-Me (**8**), (E) ImPyPyPy(C_3ODp)-γ-ImPyPyPy-β-Me (**5**), (F) ImPyPyPy(DpC_3OMe)-γ-ImPyPyPy-β-Me (**6**).

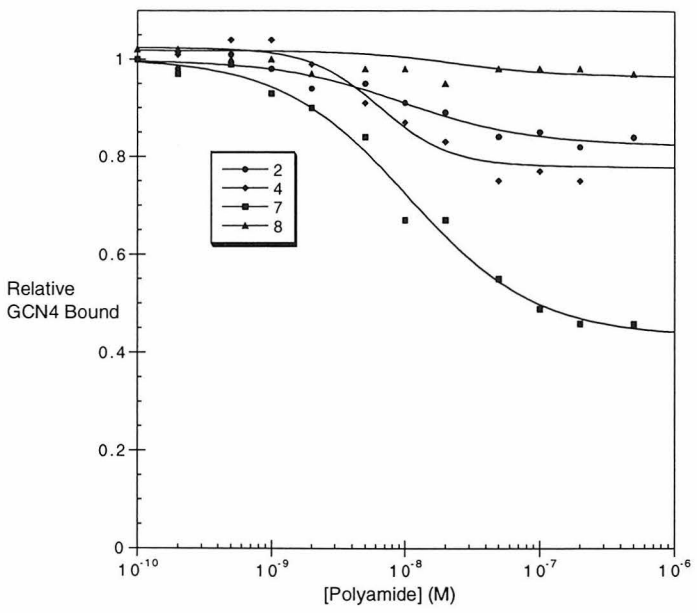


Figure 3.7. The level of GCN4 inhibition achieved by selected polyamides is compared. For each concentration of polyamide, the fraction of bound probe was determined by phosphorimager analysis followed by dividing the signal of the bound band by the sum of the bound and free bands. This value was normalized across experiments by dividing the value for each concentration by that for a control lane on the gel performed only with GCN4 (222-281). Typically, the fraction of bound probe in the absence of polyamide was ≈0.70. The optimal GCN4 inhibition observed for **7** is depicted with squares. The slight inhibition provided by **2** and **4** is shown as circles and diamonds, respectively. The absence of inhibition with **8** is depicted with triangles.

Directing the Positive Patch to Adjacent Phosphates does not Inhibit GCN4

To further probe GCN4 binding inhibition by polyamide **7**, DNA probes were designed to shift the polyamide binding site by one or two base pairs relative to that in ARE-4. It was necessary to not alter the DNA sequence of the GCN4 binding site in this new design. To accomplish this, the A/T-tract at the 3' side of the GCN4 binding site and the degeneracy of Py/Py pairs for A•T and T•A base pairs was exploited. By simply shifting the G•C bp bound by the polyamide to the 3' or 5' side, probes were obtained that would deliver the polyamide positive patch to a different phosphate on the DNA backbone, allowing other protein-DNA contacts to be examined. ARE-3 and ARE-6 shift the polyamide to one or two base pairs deeper into the GCN4 binding site, respectively, while ARE-5 moves the polyamide one bp distal to the GCN4 cognate sequence, relative to ARE-4. Significant inhibition of GCN4 binding to any of these probes was not obtained at concentrations of **7** up to 20 times higher than that required for the maximum inhibition of GCN4 on ARE-4 (Figure 3.8). On ARE-3, **7** showed a slight amount of GCN4 inhibition (Figure 3.8C), while an apparent enhancement in GCN4 binding was observed on ARE-5 with as little as 100 nM **7** (Figure 3.8A).

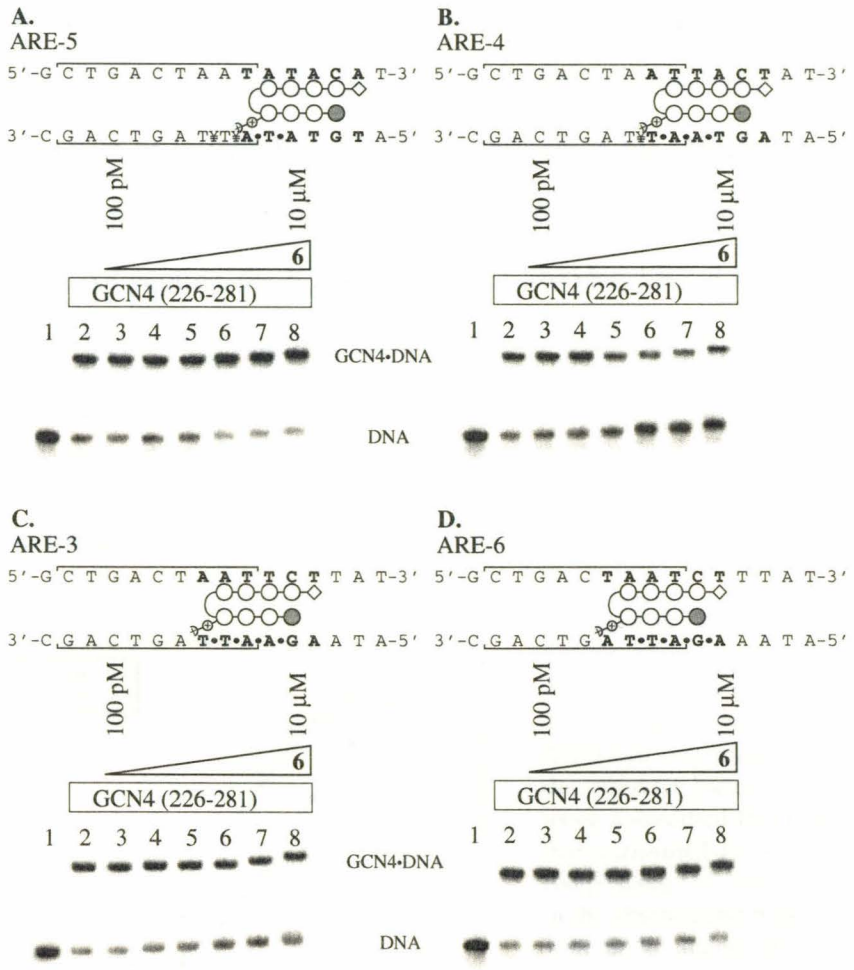


Figure 3.8. The effect of moving the polyamide binding site on GCN4 inhibition is investigated. Results are presented as in Figure 6. All panels include Lane 1: DNA only, Lane 2: 200 nM GCN4(222-281), Lane 3-8: 100 pM, 1 nM, 10 nM, 100 nM, 1 μM, and 10 μM **7**, respectively. (A) ARE-5 probe, (B) ARE-4 probe, (C) ARE-3 probe, (D) ARE-6 probe.

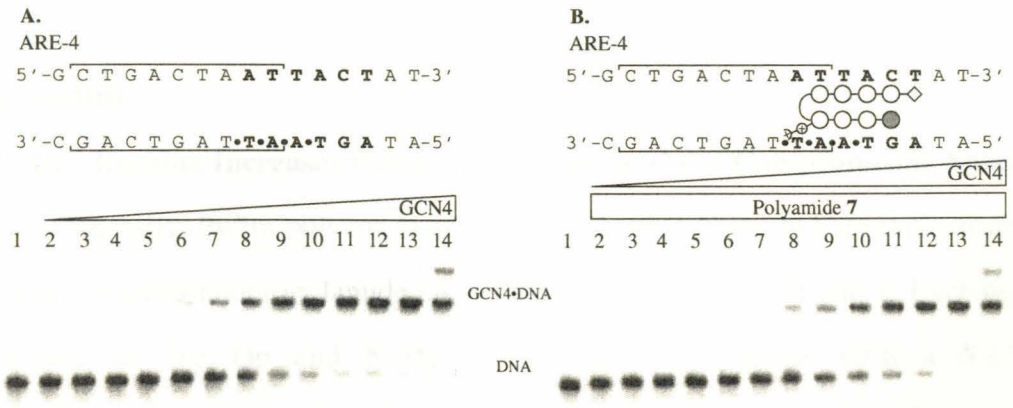


Figure 3.9. GCN4 (222-281) gel mobility shift experiments in the absence and presence of ImPyPyPy(DpDp)-γ-PyPyPyPy-β-Me (**7**). Lane 1: DNA only, Lanes 2-14: 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM, 2 μM, 5 μM and 10 μM GCN4 (222-281), respectively. (A) GCN4 (222-281) titration with ARE-4 in the absence of **7**. (B) GCN4 (222-281) titration with ARE-4 in the presence of a constant concentration of 200 nM **7**.

Effect of N-diaminoalkylpyrrole Polyamides on GCN4 affinity

The affinity of GCN4 (222-281) for the ARE-4 probe was measured in the presence and absence of ImPyPyPy(DpDp)-γ-PyPyPyPy-β-Me (**7**). Under the conditions employed here with poly(dI-dC)•poly(dI-dC) required as carrier to minimize non-specific binding by the GCN4 peptide, the K_d of GCN4 (222-281) for ARE-4 was measured as 110 nM (Figures 3.9, 3.10). Maximal inhibition of GCN4 (222-281) binding to ARE-4 was observed at a concentration of 200 nM **7** (Figures 3.6D, 3.7). The K_d of GCN4 (222-281) in the presence of 200 nM **7** was found to be 392 nM, approximately a 4-fold lower affinity than in the absence of **7** (Figures 3.9, 3.10).

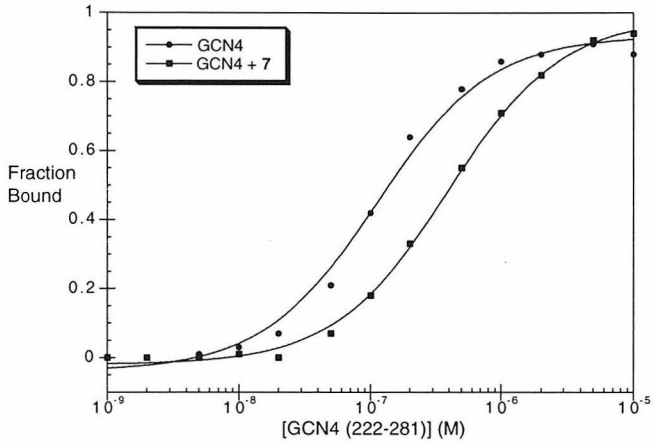


Figure 3.10. The K_d of GCN4 (222-281) for the ARE-4 probe under the gel shift conditions in the absence and presence of **7** was determined by calculating a fractional occupancy of the probe as described for Figure 3.7. The data were fit using a modified Hill equation, similar to that performed for quantitative DNase I footprinting *(15, 16)*. The GCN4 isotherm in the absence and presence of **7** is shown with circles and squares, respectively.

Discussion

A Py(Dp) Residue Increases Polyamide Affinity without Compromising Specificity

Generally, Py/Im polyamides are prepared using N-methyl aromatic amino acids and a dimethyalaminopropylamide tail on the C-terminus, as in **1** and **2**. Exchanging the placement of the Dp and N-Me to provide a polyamide with a *N*-(*N'*, *N'*-dimethylaminopropyl)pyrrole residue and a C-terminal N-methyl amide (**3** and **4**) affords an isomer of identical molecular weight and composition; however, affinity for the DNA cognate site is increased 10-fold without compromising specificity. The cell permeability

properties of the N-aminoalkylpyrrole polyamides will be interesting considering their structural similarity to those known to be cell permeable (*13, 14*).

As the generality of the pairing rules are explored for the sequence specific recognition of DNA, the incorporation of N-aminoalkylpyrrole residues to increase affinity may be useful for DNA sequences that otherwise cannot be targeted with subnanomolar affinity (*34*). The investigation of the effects on DNA binding affinity, specificity and kinetics upon incorporation of multiple *N*-(*N'*, *N'*-dimethylaminopropyl)pyrrole rings may also prove interesting. Furthermore, the benefits of N-aminoalkylpyrrole incorporation could potentially be combined with other polyamide motifs known to enhance recognition, such as the α-amino-γ-turn and cycle polyamides, and may prove to be more universally applicable (*18, 20*). N-aminoalkylpyrrole units may also be useful in polyamides containing Hp residues, which generally demonstrate a loss in affinity relative to the Py analogs (*3, 35*). A general feature that increases affinity, such as the N-aminoalkylpyrrole, may allow for expansion of the utility of low molecular polyamide motifs such as unlinked dimers and cooperative hairpin dimers (*17, 36*). Current efforts to design polyamides conjugated to moieties that endow additional function may be aided by the affinity enhancement provided by N-aminoalkylpyrrole-containing polyamides (*37*). Finally, synthetic ligands that can sequence specifically bend DNAmay be useful tools in probing protein:DNAinteractions. One or more polyamides designed to deliver an asymetrical set of positive patches to the DNA backbone may be able to bend the DNA helix by phosphate neutralization (*38*).

The model of a *N*-(*N'*, *N'*-dimethylaminopropyl)pyrrole containing polyamide in Figure 3.11 suggests that the affinity enhancement observed may result from a specific electrostatic interaction between the dimethylamino group and the phosphate oxygen on the minor groove side, two phosphates to the 3'side of the base recognized by the pyrrole. The minimized structure affords a distance between the terminal amine and phosphate oxygen of 3.6 Å. The polyamide:DNA model also suggests that alkyl linkers shorter than

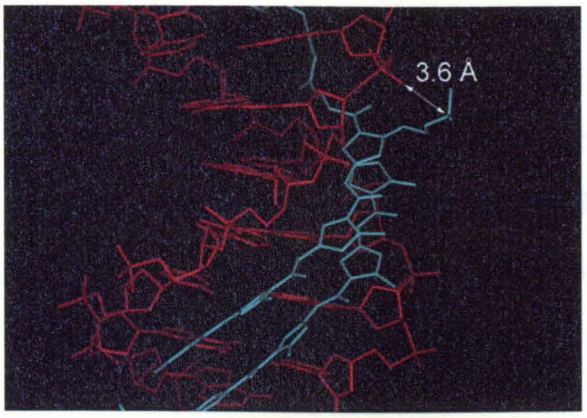


Figure 3.11. A model of a *N*-(*N'*, *N'*-dimethylaminopropyl)pyrrole residue is shown. The dimethylaminopropyl chain was built into the C-terminal pyrrole residue in one monomer of the structure obtained by X-ray crystallography of ImPyPyPy-β-Dp bound to its DNA target *(4)*. The structure was minimized by V. Rucker using Cerius software.

three methylenes may be sufficient to allow for interaction between the amine and the phosphate backbone. Upon addition of a second positive charge (**7** and **8**), the change in polyamide affinity was less than that observed by moving the Dp moiety to a pyrrole residue in **3** and **4**. It is interesting to note that simply altering the placement of the Dp chain from **1** and **2** to provide **3** and **4** had a greater net impact on DNA binding than the addition of a second positive charge as in **7** and **8**. Modeling suggests that the distal Dp moiety may wrap around the phosphate to make an electrostatic contact with the major groove oxygen, or extend down the DNA helix to an adjacent phosphate. The lower affinity may be accounted for by the energetic penalty for polyamide binding to a DNA target that has been bent or distorted by phosphate neutralization (*38*). In future studies, in will be interesting to evaluate the effects of linker length between each of the positive charges and the pyrrole ring.

It has been shown previously that the Dp tail can be removed with no effect on polyamide affinity (*39*). The *N*-(*N'*, *N'*-dimethylaminopropyl)pyrrole is clearly interacting with the DNA in a way different than that of the Dp at the C-terminus. The increase in affinity observed for **5** and **6** similar to **4** might argue in favor of a non-specific electrostatic mechanism for affinity enhancement. While this may be the case for **5** and **6**, it cannot be assumed that **3** and **4** would be expected to follow a mechanism identical to **5** and **6**. Finally, perhaps the strongest evidence for a specific phosphate contact comes from the protein inhibition data. It is not clear how the GCN4 inhibition by N-

aminoalkylpyrrole polyamides could be explained in the absence of a specific polyamide:phosphate contact that partially neutralizes the negative charge of the phosphate. Modeling suggests that the DpDp tail does not possess the necessary length to significantly block the DNA major groove. If the protein were inhibited simply by steric bulk extending into the major groove, one would have expected **5** and/or **6** to effectively inhibit the protein.

N-diaminoalkylpyrrole Polyamides Effectively Inhibit DNA Binding by GCN4

The incorporation of a doubly positively charged alkyl chain on the N-1 of a pyrrole residue has afforded a new class of hairpin polyamides that can effectively inhibit DNA binding by an exclusively major groove binding protein. The presence of the positive patch is designed to compete with protein side chains that make electrostatic contacts to the phosphate backbone, potentially by partially neutralizing the negative charge of the phosphate. The maximal GCN4 inhibition achieved with **7** is equivalent to that of the Arg-Pro-Arg analog; however, it is observed at 10-fold lower polyamide concentration for **7** than for the Arg-Pro-Arg polyamide (*27*). Furthermore, the N-dimainoalkyl moiety could be delivered from any Py residue in the polyamide, providing a versatility in design that allows for a polyamide optimized for recognition to be combined with the positive patch, offering a greater number of targetable major groove proteins.

The optimal location of the positive patch for GCN4 inhibition is consistent with X-ray crystal structure data and phosphate ethylation interference experiments (*32, 40*). As described above, the positive patch of the polyamide on ARE-4 is likely directed to the phosphate of the 3'-TpT-5' step. X-ray crystallography of GCN4 bound to a different DNA sequence reveals the residue Lys246 makes an electrostatic contact to the corresponding phosphate. Ethylation at the analogous phosphate on another GCN4 binding site resulted in a severe loss in affinity of GCN4 for the DNA target (*40*). While

structural studies show other electrostatic contacts between protein side chains and the phosphates on either side of the 3'-TpT-5' step, ethylation of these phosphates has less of an effect on GCN4 binding affinity (*32, 40*). This latter observation is consistent with a lack of significant inhibition for **7** on ARE-3 and ARE-5. The slight inhibition observed for ARE-3 may be due to the flexible linker of **7** interacting with the 3'-TpT-5' step shown to inhibit GCN4 binding. Ethylation of the phosphate corresponding to the 3'-GpT-5' (ARE-6) reduced GCN4 affinity; however, no protein contacts to this phosphate have been identified by crystallography (*32, 40*). Polyamide **7** did not inhibit GCN4 on ARE-6. Finally, modeling suggests that the terminal arginine of the reported Arg-Pro-Arg compounds may be directed not to the 3'-TpT-5' step, but rather to the adjacent 3'-TpA-5', which was shown to be less effective for GCN4 inhibition with **7**. Thus, the Arg-Pro-Arg polyamides may prove more potent if the polyamide were directed to a different phosphate. Taken together, these results suggest that the inhibition of a given protein:phosphate contact may not be universally equivalent in its effect on protein binding. Each contact plays a unique role in DNA recognition and inhibition of protein binding by positive patch polyamides will need to be optimized for each application.

It is interesting to note that the level of GCN4 inhibition achieved with **7** does not increase beyond a polyamide concentration of 200 nM. The flexibility of the linker may allow for an alternative complex with both polyamide and GCN4 bound to the DNA simultaneously. A similar effect was observed when conformational flexibility was introduced into the Arg-Pro-Arg system with an Arg-Gly-Arg substitution (*27*). It has been demonstrated that the right half-site of the GCN4 binding site, which was targeted in the probes employed here, consists of weaker protein:DNA contacts and is less important for DNA binding (*41*). Therefore, it may be beneficial to target the left half-site alone or in combination with a polyamide targeted to the right side such as **7**. Furthermore, the observation that increasing polyamide appears to enhance GCN4 binding to ARE-3 also provides insight into the level of GCN4 inhibition. The solution conditions used here

include poly(dI-dC)•poly(dI-dC) as a competitor to prevent non-specific GCN4 binding to the radiolabeled probe. Significant amounts of protein and polyamide are thus bound to the unlabeled competitor; resulting in increased concentrations of GCN4 and polyamide required for binding to the probe relative to their affinites under optimized conditions with minimal competitor. As a result, **7** likely binds the poly(dI-dC)•poly(dI-dC) and releases some of the bound GCN4 into solution, effectively increasing the concentration of free GCN4 and pushing the equilibrium toward bound GCN4. This effect might be observed as a static amount of GCN4 bound probe in the presence of higher concentrations of polyamide.

The inability of polyamides containing ether linkages in place of the amines in **7** (**5** and **6**) to effectively inhibit GCN4 binding demonstrates at a minimum that an alkyl chain of this length is insufficient to obtain protein inhibition and that the amine nitrogens play a specific role in inhibiting GCN4 binding. The amines may be partially neutralizing one or more phosphates to interfere with a protein:DNA contact or potentially distort the DNA by phosphate neutraliztion (*38*). Furthermore, the lack of inhibition provided by the mismatched polyamide **8** demonstrates that the mechanism of inhibition requires a polyamide specifically bound in the minor groove.

Previous experiments have demonstrated that Arg-Pro-Arg polyamides are not permeable to some cell types in culture (unpublished results, D. Mosier, J. Gottesfeld, P. Dervan, E. Baird, R. Bremer). The lack of permeability appears to result from the proximity or presentation of the two positive charges of the residues rather than the simple presence of two charges, considering that a doubly charged polyamide (Dp and α-amino-γ-turn) was cell permeable and specifically interfered with gene transcription (unpublished results, D. Mosier, J. Gottesfeld, P. Dervan, E. Baird, D. Herman). N-Diaminoalkylpyrrole containing polyamides may provide a new class of cell permeable positive patch polyamides that can inhibit major groove binding proteins and regulate gene expression.

The synthetic methodology developed here may also prove useful in future applications. Preparation of the N-(3-hydroxypropyl)pyrrole monomer is similar to that of the Py and Im monomers and provides material suitable for machine assisted solid phase synthesis without further protection. The use of a different monomer unit, such as the *o*-nitrophenylsulfonamide protected analog, would be a more direct synthetic approach when incorporation of N-(3-aminopropyl)pyrrole residues to enhance affinity is desired. Sequence specific DNA phosphate alkylation may also be achievable using the mesylate or tosylate derivatives for DNA binding. The availability of a readily functionalizable activated sulfonate intermediate may allow for moieties that offer additional functions to be tethered to a polyamide, as well as for the preparation of libraries of polyamides with a variety of N-1 derivatives to optimize cell permeability, bioavailability, stability or other features.

Conclusions

A new class of hairpin polyamides containing a positively charged alkyl amine at the N-1 of a pyrrole residue have been shown to sequence specifically recognize DNA with enhanced affinity and uncompromised specificity relative to analogs with the amine at the C-terminus. These polyamides are obtained by exchanging the position of the N-methyl of a pyrrole ring and the C-terminal dimethylaaminopropylamide to provide an isomer of identical molecular weight containing a terminal N-methyl amide and a *N*-(*N'*, *N'*-dimethylaminopropyl)pyrrole residue with potentially interesting cell permeability properties. The positive patch provided by a N-diaminoalkylpyrrole residue proved effective in the inhibition of DNA binding by the exclusively major groove binding protein GCN4 (222-281), potentially by competing with the protein side chains for electrostatic contacts to the DNA phosphate backbone. The naturally occurring Arg-Pro-Arg motif that recognizes the major groove and electrostatically contacts the phosphate backbone has been functionally reduced to a simple synthetic alkyl chain. N-aminoalkylpyrrole-containing polyamides may expand both the DNA sequences targetable with subnanomolar affinity and the general utility of certain polyamide motifs, as well as augment the functional repertoire of polyamides as synthetic ligands for the regulation of gene expression.

Experimental

All synthetic reagents were as previously described or obtained from Aldrich (*12*). Analytical HPLC was performed on a Beckman *Gold Nouveau* system with a model 126 pump and model 168 diode array detector. A Rainen C_{18}, Microsorb MV, 5 μm, 300 × 4.6 mm reverse phase column was employed with 0.1% (w/v) TFA:H_2O and 1.5% acetonitrile/min. Preparatory HPLC was performed on a Waters DeltaPak 25 × 100 mm 100 μm C_{18} column in 0.1% (w/v) TFA, gradient elution 0.25%/min. CH_3CN. Resin substitution of synthesized polyamides was calculated as $L_{new}(mmol/g) = L_{old}/(1 + L_{old}(W_{new} - W_{old}) \times 10^{-3})$ where L is the loading (mmol of amine per gram of resin) and W is the weight (g mol^{-1}) of the growing peptide attached to the resin (*12*). DNA restriction fragment labeling, DNase I footprinting, and determination of equilibrium association constants were accomplished using previously described protocols (*15, 16*). Chemical sequencing reactions were performed according to published methods (*42, 43*).

Monomer Synthesis

Ethyl (3-hydroxypropyl)-4-nitropyrrole-2-carboxylate (10). Potassium carbonate (22.8 g, 163 mmol) was added to ethyl 4-nitropyrrole-2-carboxylate (**9**) (10 g, 54.3 mmol) in acetone (740 mL) and the slurry stirred for 2 hours, followed by the addition of 3-iodopropanol (23.5 g, 109 mmol) and refluxing for 2 hours. The reaction was cooled to room temperature, filtered and concentrated *in vacuo*. Water (200 mL) was added to the resulting oil, the pH of the aqueous layer was reduced to 3 with 10% H_2SO_4, and the mixture extracted with ethyl acetate. The combined ethyl acetate extracts were dried ($MgSO_4$) and concentrated *in vacuo*. The solid was chromatographed with silica gel (ethyl acetate:hexanes) to provide a clear oil (12.8 g, 52.6 mmol, 99% yield): ^1H NMR (DMSO-d_6): δ 8.21 (s, 1H), 7.32 (s, 1H), 4.62 (t, 1H, J=4.9 Hz), 4.47 (t, 2H, J=6.8 Hz), 4.21 (q, 2H, J=7.0 Hz), 3.31 (q, 2H, J=6.1 Hz), 1.80 (m, 2H), 1.30 (t, 3H, J=7.0 Hz).

(3-Hydroxypropyl)-4-[(*tert*-butoxycarbonyl)amino]-pyrrole-2-carboxylic acid (11). Ethyl (3-hydroxypropyl)-4-nitropyrrole-2-carboxylate (6.08 g, 25.1 mmol) was dissolved in DMF (20 mL). 10% Pd/C (1.0 g) was added and the mixture was stirred under hydrogen (300 psi) for 3 hours. Pd/C was removed by filtering through Celite. Di-*tert*-butyl dicarbonate (5.41 g, 24.8 mmol) and DIEA (8.6 mL, 50 mmol) was added and stirred for 2 hours. The reaction was cooled to room temperature, filtered, washed with methanol and concentrated *in vacuo*. Chromatography with silica gel (ethyl acetate:hexanes) provided a clear oil (5.10 g, 16.3 mmol, 65% yield): ^{1}H NMR (DMSO-d_6): δ 9.08 (s, 1H), 7.06 (s, 1H), 6.60 (s, 1H), 4.49 (t, 1H, J=5.1 Hz), 4.20 (t, 2H, J=6.8 Hz), 4.12 (q, 2H, J=7.5 Hz), 3.30 (q, 2H, J=5.5 Hz), 1.72 (m, 2H), 1.45 (s, 9H), 1.23 (t, 3H, J=7.5 Hz). Ethyl (3-hydroxypropyl)-4-[(*tert*-butoycarbonyl)amino]-pyrrole-2-carboxylate (2.0 g, 6.4 mmol) was dissolved in ethanol (10 mL), followed by the addition of 1M KOH (40 mL) and heating (70 °C) for 2 hours. The reaction was cooled to room temperature and washed with ethyl ether. The pH of the aqueous layer was reduced to 3 with 10% H_2SO_4, and the mixture extracted with diethyl ether. The combined ethyl acetate extracts were dried ($MgSO_4$) and concentrated *in vacuo* to provide a white, crystalline solid (1.5 g, 5.3 mmol, 83% yield): ^{1}H NMR (DMSO-d_6): δ 12.04 (s, 1H), 9.05 (s, 1H), 7.04 (s, 1H), 6.54 (s, 1H), 4.20 (t, 2H, J=6.8 Hz), 3.27 (q, 2H, J=6.1 Hz), 1.73 (m, 2H), 1.40 (s, 9H).

Polyamide Synthesis

ImPyPyPy(C_3OH)-γ-PyPyPyPy-β-Me (12). ImPyPyPy(C_3OH)-γ-PyPyPyPy-β-PAM-resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc-β-PAM-resin (600 mg, 0.75 mmol/g). (3-hydroxylpropyl)-4-[(*tert*-butoxycarbonyl)amino]-pyrrole-2-carboxylic acid (212 mg, 0.75 mmol), HBTU (284 mg, 0.75 mmol) and HOBt (304 mg, 2.25 mmol) were added dry to a synthesis cartridge. Upon delivery of DMF and DIEA activation occurs as described. A sample of polyamide resin (380 mg, 0.43 mmol/g) was cleaved with methylamine (55 °C, 28 hr.) in a Parr apparatus.

The methylamine was allowed to evaporate and the crude polyamide was redissolved in 1:1 1M NH_4OH:CH_3CN (20 mL), filtered to remove the resin and lyophilized dry (146 μmol).

ImPyPyPy(C_3OH)-γ-ImPyPyPy-β-Me (13). ImPyPyPy(C_3OH)-γ-ImPyPyPy-β-PAM-resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc-β-PAM-resin (450mg, 0.75 mmol/g) as described above. A sample of polyamide resin (178 mg, 0.43 mmol/g) was cleaved with methylamine (55 °C, 18 hr.) in a Parr apparatus. The methylamine was allowed to evaporate and the crude polyamide was redissolved in 1:1 1M NH_4OH:CH_3CN (20 mL), filtered to remove the resin and lyophilized dry (56 μmol).

ImPyPyPy(Dp)-γ-ImPyPyPy-β-Me (3). Crude ImPyPyPy(C_3OH)-γ-ImPyPyPy-β-Me (25 μmol) was dissolved in pyridine (600 μL) followed by the addition of methanesulfonyl chloride (100 μL). After 30 min., the reaction was precipitated with ethyl ether (1 mL), decanted, and redissolved in DMF (1 mL). 1:1 Dimethylamine (2.0M in THF):DMF (3 mL) was added and the reaction allowed to proceed for 45 min. at 90 °C. The reaction was diluted to 8 mL with 0.1% (w/v) TFA and purified by preparatory reverse phase HPLC to afford ImPyPyPy(Dp)-γ-ImPyPyPy-β-Me (**3**) upon lyophilization of the appropriate fractions (1.3 mg, 1.0 μmol, 4.0 % recovery); UV (H_2O) λ_{max} 312 (66,000); ^{1}H NMR (DMSO-d_6): δ 10.45 (s, 1H), 10.27 (s, 1H), 9.95 (s, 3H), 9.93 (s, 1H), 9.88 (s, 1H), 9.27 (m, 1H), 8.19 (m, 1H), 8.00 (m, 1H), 7.79 (m, 1H), 7.42 (s, 1H), 7.36 (s, 1H), 7.30 (s, 1H,), 7.25 (s, 1H), 7.22 (s, 1H), 7.19 (s, 1H), 7.15 (s, 1H), 7.13 (s, 1H), 7.10 (s, 1H), 7.05 (s, 1H), 7.00 (s, 2H), 6.87 (s, 1H), 6.78 (s, 1H), 4.26 (m, 2H), 3.94 (s, 3H), 3.91 (s, 3H), 3.80 (s, 12H), 3.75 (s, 3H), 3.35 (m, 2H), 3.20 (m, 2H), 2.97 (m, 2H), 2.74 (d, 6H, J=4.8 Hz), 2.52 (d, 3H, J=4.5 Hz), 2.29 (m, 4H), 2.02 (m, 2H), 1.75 (m, 2H); MS (MALDI-TOF): $[M+H]^+$ 1222.6, $[M+Na]^+$ 1244.5, $[M+K]^+$ 1260.5. ($C_{58}H_{72}N_{21}O_{10}^+$, $[M+H]^+$, calc. 1222.6; $C_{58}H_{71}N_{21}NaO_{10}^+$, $[M+Na]^+$, calc. 1244.6; $C_{58}H_{71}KN_{21}O_{10}^+$, $[M+K]^+$, calc. 1260.5).

ImPyPyPy(Dp)-γ-PyPyPyPy-β-Me (4). Crude ImPyPyPy(C_3OH)-γ-PyPyPyPy-β-Me (49 μmol) was dissolved in pyridine followed by the addition of toluenesulfonyl chloride (2.2 M in pyridine, 350 μL) and kept at 0 °C for 10 hours. After the addition of dimethylamine (2.0 M in THF, 2.0 mL), the reaction was heated at increasing temperatures until it was determined complete by analytical HPLC after 5 hours at 80 °C. the reaction mixture was diluted to 8 mL with 0.1% (w/v) TFA and purified by preparatory reverse phase HPLC to provide ImPyPyPy(Dp)-γ-PyPyPyPy-β-Me (**4**) upon lyophilization of the appropriate fractions (5.8 mg, 4.8 μmol, 9.8% recovery); UV (H_2O) λ_{max} 312 (66,000); ^{1}H NMR (DMSO-d_6): δ 10.46 (s, 1H), 9.96 (s, 1H), 9.95 (s, 1H), 9.92 (s, 1H), 9.89 (s, 1H), 9.88 (s, 1H), 9.83 (s, 1H), 9.2 (br s, 1H), 8.18 (m, 1H), 7.99 (m, 1H), 7.80 (m, 1H), 7.38 (s, 1H), 7.31 (d, 1H, J=1.8 Hz), 7.26 (d, 1H, J=1.8 Hz), 7.20 (d, 1H, J=1.5 Hz), 7.16 (m, 3H), 7.06 (d, 1H, J=1.8 Hz), 7.03 (s, 1H), 7.02 (s, 1H), 7.01 (s, 1H), 6.89 (d, 1H, J=1.5 Hz), 6.84 (d, 1H, J=1.5 Hz), 6.79 (d, 1H, J=1.5 Hz), 4.24 (m, 2H), 3.96 (s, 3H), 3.82 (s, 9H), 3.81 (s, 3H), 3.80 (s, 3H), 3.76 (s, 3H), 3.35 (m, 2H), 3.20 (m, 2H), 3.00 (m, 2H), 2.74 (d, 6H, J=5.1 Hz), 2.53 (d, 3H, J=4.5 Hz), 2.27 (m, 2H), 2.03 (m, 2H), 1.97 (m, 2H), 1.78 (m, 2H); MS (MALDI-TOF): $[M+H]^+$ 1221.6. ($C_{59}H_{73}N_{20}O_{10}^+$, $[M+H]^+$, calc. 1221.6).

ImPyPyPy(C_3ODp)-γ-PyPyPyPy-β-Me (5). Crude ImPyPyPy(C_3OH)-γ-PyPyPyPy-β-Me (24 μmol) in pyridine (0.4 mL) at 0 °C was activated with MsCl (75 μL) at 0° C for three hours, followed by precipitation with ether (1 mL). The supernatant was redissolved in DMF (0.5 mL) and cooled to 0 °C. *N*,*N*-dimethyl-3-amino-propanol (14 μL, 120 μmol) was added to NaH (60% dispersion in mineral oil, 4.7 mg, 114 μmol) in DMF (150 μL) and added to the reaction. Repeating the addition of NaH and *N*,*N*-dimethyl-3-amino-propanol and warming to room temperature did not provide sufficient reaction after 27 hours as determined by analytical HPLC. More NaH (60% dispersion in mineral oil, 60 mg, 1.5 mmol) and *N*,*N*-dimethyl-3-amino-propanol (180 μL, 1.5 mmol) was added, followed by reaction at room temperature for 22 hours. The reaction mixture was diluted

to 8 mL with 0.1% (w/v) TFA and purified by preparatory reverse phase HPLC to afford ImPyPyPy(C_3ODp)-γ-PyPyPyPy-β-Me (**5**) upon lyophilization of the appropriate fractions (1.4 mg, 1.1 μmol, 4.6% recovery); UV (H_2O) λ_{max} 312 (66,000); ^{1}H NMR (DMSO-d_6): δ 10.43 (s, 1H), 9.94 (s, 1H), 9.93 (s, 1H), 9.90 (s, 1H), 9.87 (s, 2H), 9.81 (s, 1H), 9.25 (m, 1H), 8.16 (m, 1H), 7.97 (m, 1H), 7.77 (m, 1H), 7.35 (s, 1H), 7.32 (s, 1H), 7.30 (s, 1H,), 7.24 (s, 1H), 7.18 (s, 3H), 7.14 (s, 1H), 7.13 (s, 2H), 7.05 (s, 1H), 7.00 (s, 2H), 6.88 (s, 1H), 6.82 (s, 1H), 6.77 (s, 1H), 4.25 (m, 2H), 3.94 (s, 3H), 3.80 (s, 15H), 3.75 (s, 3H), 3.34 (m, 4H), 3.20 (m, 2H), 3.00 (m, 2H), 2.72 (d, 6H, J=4.8 Hz), 2.52 (d, 3H, J=4.2 Hz), 2.24 (m, 6H), 2.02 (m, 2H), 1.76 (m, 4H); MS (MALDI-TOF): $[M+H]^+$ 1279.7. ($C_{62}H_{79}N_{20}O_{11}^+$, $[M+H]^+$, calc. 1279.6).

ImPyPyPy(DpC_3OMe)-γ-PyPyPyPy-β-Me (6). MsCl activation of crude ImPyPyPy(C_3OH)-γ-PyPyPyPy-β-Me (24 μmol) as for **4** followed by redissolving of the precipitate in DMF (3 mL), reaction with 3-methoxypropylamine (1 mL, 90 °C, 30 min.) and preparatory reverse phase HPLC afforded ImPyPyPy(DpC_3OMe)-γ-PyPyPyPy-β-Me (**6**) upon lyophilization of the appropriate fractions (1.5 mg, 1.2 μmol, 5.0% recovery); UV (H_2O) λ_{max} 312 (66,000); ^{1}H NMR (DMSO-d_6): δ 10.44 (s, 1H), 9.96 (s, 1H), 9.95 (s, 1H), 9.91 (s, 1H), 9.88 (s, 1H), 9.88 (s, 1H), 9.83 (s, 1H), 8.24 (m, 2H), 8.20 (m, 1H), 7.99 (m, 1H), 7.80 (m, 1H), 7.37 (d, 1H, J=0.6 Hz), 7.32 (d, 1H, J=1.8 Hz), 7.26 (d, 1H, J=1.8 Hz), 7.20 (m, 3H), 7.16 (m, 3H), 7.06 (d, 1H, J=1.8 Hz), 7.01 (m, 3H), 6.89 (d, 1H, J=1.5 Hz), 6.84 (d, 1H, J=1.8 Hz), 6.79 (d, 1H, J=2.1 Hz), 4.30 (m, 2H), 3.96 (s, 3H), 3.82 (s, 9H), 3.81 (s, 3H), 3.81 (s, 3H), 3.76 (s, 3H), 3.19 (s, 3H), 2.90 (m, 4H), 2.52 (m, 5H), 2.42 (m, 2H), 2.27 (m, 6H), 2.00 (m, 2H), 1.78 (m, 4H); MS (MALDI-TOF): $[M+H]^+$ 1265.8, $[M+Na]^+$ 1287.8, $[M+K]^+$ 1303.8. ($C_{61}H_{76}N_{20}O_{11}^+$, $[M+H]^+$, calc. 1265.6; $C_{61}H_{76}N_{20}NaO_{11}^+$, $[M+Na]^+$, calc. 1287.6; $C_{61}H_{76}KN_{20}O_{11}^+$, $[M+K]^+$, calc. 1303.6).

ImPyPyPy(DpDp)-γ-PyPyPyPy-β-Me (7). MsCl activation of crude ImPyPyPy(C_3OH)-γ-PyPyPyPy-β-Me (23 μmol) as for **4** followed by redissolving of the precipitate in pyridine (0.75 mL), reaction with *N,N,N'*-trimethyl-1,3-propanediamine (0.75 mL, 90 °C,

40 min.) and preparatory reverse phase HPLC afforded ImPyPyPy(DpDp)-γ-PyPyPyPy-β-Me (**7**) upon lyophilization of the appropriate fractions (6.7 mg, 5.2 μmol, 23% recovery); UV (H_2O) λ_{max} 312 (66,000); ^{1}H NMR (DMSO-d_6): δ 10.45 (s, 1H), 9.97 (s, 1H), 9.95 (s, 1H), 9.92 (s, 1H), 9.89 (s, 1H), 9.88 (s, 1H), 9.84 (s, 1H), 9.55 (m, 1H), 8.44 (m, 1H), 8.19 (m, 1H), 7.98 (m, 1H), 7.80 (m, 1H), 7.37 (m, 1H), 7.34 (d, 1H, J=1.8 Hz), 7.26 (d, 1H, J=1.8 Hz), 7.20 (s, 3H), 7.15 (m, 2H), 7.07 (d, 1H, J=1.5 Hz), 7.01 (m, 3H), 6.88 (d, 1H, J=1.8 Hz), 6.84 (d, 1H, J=1.8 Hz), 6.79 (d, 1H, J=1.8 Hz), 4.30 (m, 2H), 3.96 (s, 3H), 3.82 (s, 9H), 3.81 (s, 6H), 3.80 (s, 3H), 3.35 (m, 4H), 3.20 (m, 2H), 3.00 (m, 2H), 2.75 (m, 9H), 2.54 (d, 3H, J=4.5 Hz), 2.26 (m, 6H), 2.04 (m, 2H), 1.96 (m, 2H), 1.77 (m, 2H); MS (MALDI-TOF): $[M+H]^+$ 1292.7 ($C_{63}H_{82}N_{21}O_{10}^+$, $[M+H]^+$, calc. 1292.7).

ImPyPyPy(DpDp)-γ-ImPyPyPyPy-β-Me (8). Crude ImPyPyPy(C_3OH)-γ-ImPyPyPyPy-β-Me (23 μmol) was dissolved in pyridine (150 μL) and activated with TsCl (4.8M in pyridine, 200 μL) for 1 hour at 0 °C. The addition of *N,N,N'*-trimethyl-1,3-propanediamine (0.30 mL, 55 °C, 30 min.) followed by preparatory reverse phase HPLC afforded ImPyPyPy(DpDp)-γ-ImPyPyPyPy-β-Me (**8**) upon lyophilization of the appropriate fractions (0.4 mg, 0.3 μmol, 1.3% recovery); UV (H_2O) λ_{max} 312 (66,000); ^{1}H NMR (DMSO-d_6): δ 10.43 (s, 1H), 10.26 (s, 1H), 9.96 (s, 1H), 9.93 (s, 1H), 9.91 (s, 2H), 9.86 (s, 1H), 9.45 (m, 1H), 8.17 (m, 1H), 7.97 (m, 1H), 7.78 (m, 1H), 7.41 (s, 1H), 7.36 (s, 2H), 7.33 (s, 1H,), 7.25 (s, 1H), 7.22 (s, 1H), 7.18 (s, 2H), 7.14 (s, 1H), 7.14 (s, 1H), 7.11 (s, 1H), 7.00 (s, 2H), 6.86 (s, 1H), 6.77 (s, 1H), 4.24 (m, 2H), 3.96 (s, 3H), 3.91 (s, 3H), 3.80 (s, 12H), 3.75 (s, 3H), 3.35 (m, 4H), 3.20 (m, 2H), 3.00 (m, 2H), 2.74 (d, 9H, J=3.6 Hz), 2.52 (d, 3H, J=4.8 Hz), 2.27 (m, 6H), 2.05 (m, 2H), 1.94 (m, 2H), 1.78 (m, 2H); MS (MALDI-TOF): $[M+H]^+$ 1293.9 ($C_{62}H_{81}N_{22}O_{10}^+$, $[M+H]^+$, calc. 1293.7).

Acknowledgments

We are grateful to the National Institutes of Health for research support, the National Science Foundation for a predoctoral fellowship to R.E.B., Bristol-Myers Squibb and the Ralph M. Parsons Foundation for predoctoral fellowships to R.E.B. and N.R.W., and the National Institutes of Health for a research service award to J.W.S. We thank G.M. Hathaway and the Caltech Protein/Peptide Microanalytical Laboratory for MALDI-TOF mass spectrometry.

References

1. Dervan, P. B., and Bürli, R. W. (1999) *Curr. Opin. Chem. Biol. 3*, 688-693.

2. White, S., Baird, E. E., and Dervan, P. B. (1997) *Chem. Biol. 4*, 569-578.

3. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. (1998) *Nature 391*, 468-471.

4. Kielkopf, C. L., Baird, E. E., Dervan, P. D., and Rees, D. C. (1998) *Nat. Struct. Biol. 5*, 104-109.

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

6. deClairac, R. P. L., Geierstanger, B. H., Mrksich, M., Dervan, P. B., and Wemmer, D. E. (1997) *J. Am. Chem. Soc. 119*, 7909-7916.

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

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

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

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

11. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1999) *J. Am. Chem. Soc. 121*, 1113-1120.

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

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

14. Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan, P. B. (1998) *Proc. Natl. Acad. Sci. USA 95*, 12890-12895.

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

16. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) *Chem.-Eur. J. 3*, 1600-1607.

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

18. Herman, D. M., Baird, E. E., and Dervan, P. B. (1998) *J. Am. Chem. Soc. 120*, 1382-1391.

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

20. Herman, D. M., Turner, J. M., Baird, E. E., and Dervan, P. B. (1999) *J. Am. Chem. Soc. 121*, 1121-1129.

21. Cuenoud, B., Casset, F., Hüsken, D., Natt, F., Wolf, R. M., Altman, K.-H., Martin, P., and Moser, H. E. (1998) *Angew. Chem.-Int. Ed. 37*, 1288-1291.

22. Bruice, T. C., Mei, H. Y., He, G. X., and Lopez, V. (1992) *Proc. Natl. Acad. Sci. USA 89*, 1700-1704.

23. Blasko, A., Browne, K. A., and Bruice, T. C. (1994) *J. Am. Chem. Soc. 116*, 3726-3737.

24. Neely, L., Trauger, J. W., Baird, E. E., Dervan, P. B., and Gottesfeld, J. M. (1997) *J. Mol. Biol. 274*, 439-445.

25. Oakley, M. G., Mrksich, M., and Dervan, P. B. (1992) *Biochemistry 31*, 10969-10975.

26. Bremer, R. E., Szewczyk, J. W., Baird, E. E., and Dervan, P. B. (2000) *submitted.*

27. Bremer, R. E., Baird, E. E., and Dervan, P. B. (1998) *Chem. Biol. 5*, 119-133.

28. Bruice, T. C., Sengupta, D., Blasko, A., Chiang, S. Y., and Beerman, T. A. (1997) *Bioorg. & Med. Chem. 5*, 685-692.

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

30. Hope, I. A., and Struhl, K. (1986) *Cell 46*, 885-894.

31. Oakley, M. G., and Dervan, P. B. (1990) *Science 248*, 847-850.

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

33. Konig, P., and Richmond, T. J. (1993) *J. Mol. Biol. 233*, 139-154.

34. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) *J. Am. Chem. Soc. 119*, 6953-6961.

35. White, S., Turner, J. M., Szewczyk, J. W., Baird, E. E., and Dervan, P. B. (1999) *J. Am. Chem. Soc. 121*, 260-261.

36. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1998) *Angew. Chem.-Int. Edit. Engl. 37*, 1421-1423.

37. Mapp, A. K., Ansari, A., Ptashne, M., and Dervan, P. B. (2000) *Proc. Natl. Acad. Sci. USA in press.*

38. Maher III, L. J. (1998) *Curr. Opin. Chem. Biol. 2*, 688-694.

39. White, S., Baird, E. E., and Dervan, P. B. (1997) *J. Am. Chem. Soc. 119*, 8756-8765.

40. Gartenberg, M. R., Ampe, C., Steitz, T. A., and Crothers, D. M. (1990) *Proc. Natl. Acad. Sci. USA 87*, 6034-6038.

41. Struhl, K. (1992) in *Transcriptional Regulation* (McKnight, S. L., and Yamamoto, K. R., Eds.) pp 833-860, Cold Spring Harbor Laboratory Press, Plainview, NY.

42. Maxam, A. M., and Gilbert, W. S. (1980) *Methods Enzymol. 65*, 499-560.

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

Chapter 4

Recognition of the DNA Minor Groove by Pyrrole-Imidazole Polyamides: Comparison of Desmethyl- and *N*-Methylpyrrole

The text of this chapter was taken in part from a submitted manuscript coauthored with Jason W. Szewczyk, Eldon E. Baird and Prof. Peter B. Dervan.

Abstract: Polyamides consisting of N-methylpyrrole (Py), N-methylimidazole (Im), and N-methyl-3-hydroxypyrrole (Hp) are synthetic ligands that recognize predetermined DNA sequences with affinities and specificities comparable to many DNA-binding proteins. As derivatives of the natural products distamycin and netropsin, Py/Im/Hp polyamides have retained the N-methyl substituent, although structural studies of polyamide:DNA complexes have not revealed an obvious function for the N-methyl. In order to assess the role of the N-methyl moiety in polyamide:DNA recognition, a new monomer, desmethylpyrrole (Ds), where the N-methyl moiety has been replaced with hydrogen, was incorporated into an eight-ring hairpin polyamide by solid phase synthesis. MPE footprinting, affinity cleavage, and quantitative DNase I footprinting revelaved that replacement of each Py residue with Ds resulted in identical binding site size and orientation and similar binding affinity for the six base pair target DNA sequence. Remarkably, the Ds-containing polyamide exhibited an 8-fold loss in specificity for the match site versus a mismatched DNA site, relative to the all-Py parent. Polyamides with Ds exhibit increased water solubility, which may alter the cell membrane permeability properties of the polyamide. The addition of Ds to the repertoire of available monomers may prove useful as polyamides are applied to gene regulation *in vivo*. However, the benefits of Ds incorporation must be balanced with a potential loss in specificity.

Introduction

Cell permeable small molecules with the ability to target predetermined DNA sequences and interfere with gene expression would be valuable tools in molecular biology and potentially in human medicine. Recent efforts in our laboratories have been inspired by the natural products netropsin and distamycin, crescent-shaped di- and tripeptides composed of N-methylpyrrole (Py) carboxamides that recognize four or five successive A•T base pairs (*1-8*). Distamycin and netropsin normally form 1:1 complexes in the minor groove of DNA (*9-12*). However, structural studies have revealed that at high concentrations distamycin forms an antiparallel side-by-side dimer tightly packed into the minor groove, with each monomer making hydrogen bonds from the backbone amides to the bases on one DNA strand (*13, 14*). Atomic substitutions to the N-methylpyrrole scaffold of distamycin have provided new aromatic amino acids, N-methylimidazole (Im) and N-methyl-3-hydroxypyrrole (Hp), that allow for the recognition of predetermined DNA sequences (*15-17*). A set of pairing rules have been developed that correlate each Watson-Crick base pair with a side-by-side pairing of the Py, Hp, and Im aromatic amino acids (Figure 4.1) (*16-18*). An antiparallel pairing of Im opposite Py (Im/Py pair) distinguishes G•C from C•G and both of these from A•T and T•A base pairs (*16, 18-21*). A Py/Py pair binds both A•T and T•A in preference to G•C and C•G (*13, 14, 16, 18-20, 22, 23*). The discrimination of T•A from A•T using Hp/Py and Py/Hp pairs, respectively, completes the four base pair code (*17, 24, 25*). The linker amino acid γ-aminobutyric acid (γ) connects the polyamide subunits in a "hairpin motif" and enhances affinity >100-fold relative to the unlinked dimers (*26-29*). A C-terminal β-alanine (β) increases both affinity and specificity and facilitates solid-phase synthesis (*27, 30*). The γ and β residues both favor recognition of A•T and T•A base pairs (*31*). Each polyamide subunit recognizes the minor groove with a N-C orientation relative to the 5'-3' direction of the DNA helix (*21, 32*). Eight-ring hairpin polyamides with affinities and specificities comparable to those of DNA-binding proteins have been shown to permeate a variety of cell types in culture and specifically interfere with gene expression (*33-35*).

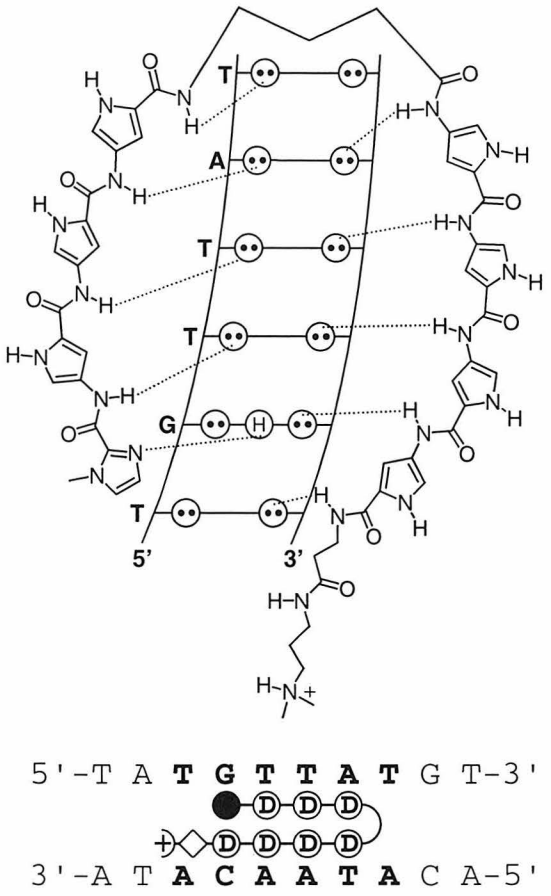


Figure 4.1. (Top) Hydrogen bonding model of ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**) bound in the minor groove to the 5'-TGTTAT-3' match site. Circles with two dots represent the lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogens of guanines. Putative hydrogen bonds are illustrated by dotted lines. (Bottom) Model of **2** bound to the 5'-TGTTAT-3' match site (bold). N-Methylimidazole and desmethylpyrrole are represented by black circles and white circles with a "D" inside, respectively. The γ-turn is shown as a curved line. β-Alanine and Dp are depicted as a diamond and a plus sign, respectively.

Extensive biochemical studies have revealed the incremental contributions to DNA affinity and specificity of many features of polyamides, most importantly the effects of substitutions at the 3-position of the aromatic rings (*16-18, 22*). However, in every case, the N-methyl substituent present in the distamycin and netropsin has been retained. Structural studies do not reveal an obvious role for the N-methyls in polyamide:DNA recognition, as they make no DNA contacts and point away from the minor groove, toward the solvent (*20, 24, 29, 36*). As natural products, distamycin and netropsin may have faced selective pressures based on many factors, including biosynthesis, solubility, permeability, and DNA affinity and specificity. It remained to be determined if the N-methyl is a substituent that can be discarded, or if it plays an important role in polyamide:DNA recognition, in particular in the side-by-side pairings. This question provided the impetus to examine a desmethylpyrrole ring (Ds), that replaces the N-methyl

group of the pyrrole with hydrogen. Desmethylpyrrole should follow the polyamide pairing rules similarly to pyrrole. The sequence specific elements of Py/Im/Hp polyamides are the two atoms at the C-3 position of each ring pair, whilst the N-1 position projects out of the minor groove and does not make sequence specific DNA contacts.

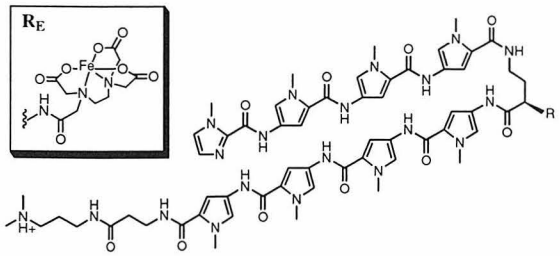
Distamycin analogs with single desmethylpyrrole substitutions exhibited enhanced antiviral and cytotoxic properties relative to distamycin (*37, 38*). It was postulated that the enhanced activity derived from increased cell permeability for the desmethylpyrrole analogs (*37*). The potential benefits of desmethylpyrrole substitutions on biological activity provided further motivation to examine the DNA binding energetics of Ds-containing hairpin polyamides.

Here we report the DNA binding properties of two eight-ring hairpin polyamides: a control polyamide containing all Im/Py and Py/Py pairs, ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**), and a polyamide completely substituted with Ds, ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**) (Figure 4.2). Binding site size and location were determined by MPE footprinting, while affinity cleavage revealed binding site orientation and stoichiometry (*4, 8, 39*). Quantitative DNase I footprint titrations were employed to measure the equilibrium association constants (K_a) of each polyamide at its match and mismatch sites (*40-42*). Finally, not only will the Ds monomer reveal the role of the N-methyl substituent in polyamide:DNA recognition, but each incorporation will decrease the polyamide molecular weight and add a hydrogen bond donor that can interact with the solvent, potentially increasing polyamide solubility in water and altering cell permeability, as well.

Results

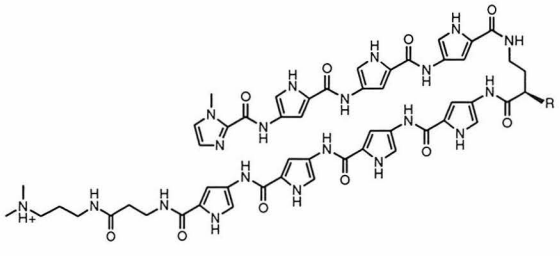
Desmethylpyrrole Monomer and Polyamide Synthesis

Analogs of distamycin have been prepared previously by solution phase methods with a single desmethylpyrrole substitution using the Boc-Ds-acid (**5**) (*37*). It remained to be determined if a hairpin polyamide fully substituted with desmethylpyrrole in place of



(**1**) R=H, ImPyPyPy-γ-PyPyPyPy-β-Dp

(**1-E**) R=R_E, ImPyPyPy-(*R*)$^{EDTA•Fe(II)}$γ-PyPyPyPy-β-Dp



(**2**) R=H, ImDsDsDs-γ-DsDsDsDs-β-Dp

(**2-E**) R=R_E, ImDsDsDs-(*R*)$^{EDTA•Fe(II)}$γ-DsDsDsDs-β-Dp

Figure 4.2. Structures of eight-ring hairpin polyamides ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**), and ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**), and the corresponding EDTA-modified derivatives, **1-E**, and **2-E**. Schematics of each polyamide bound to the 5'-TGTTAT-3' match site are depicted as in Figure 1, with an open circle representing N-methylpyrrole.

each pyrrole residue could be synthesized by solid phase protocols (*30*). Towards this end, we report here a convenient synthesis of **5,** based on the preparation of the N-methylpyrrole derivative, that provides **5** in 25 g scale without column chromatography (Figure 4.3) (*30*). Polyamide **1** has been previously reported (*33*). The solid-phase synthesis of desmethylpyrrole containing polyamides, **2**, was performed on β-Ala-Pam resin using DCC/HOBt activated **5**, Boc-Py-OBt, and Boc-Im-acid, according to previously reported procedures (*30*). Cleavage of the resin bound polyamide by aminolysis with dimethylaminopropylamine (Dp) afforded polyamide **2**, following reversed phase HPLC purification. Although the recovery of polyamide **2** was lower than that generally observed for Py-containing polyamides, the solid phase protocols did

provide Ds-containing polyamides of equal purity to the Py-containing polyamides. For the synthesis of affinity cleavage analogs with EDTA, the desired polyamides were synthesized using (*R*)-2-Fmoc-4-Boc-diaminobutyric acid in place of Boc-γ-aminobutyric acid (*32*). After aminolysis and reversed phase HPLC purification, the deprotected 2-amino group was treated with an excess of EDTA-dianhydride, and the remaining anhydride hydrolyzed to provide **1-E** and **2-E**, following another reversed phase HPLC purification (*32*). Polyamides were characterized by analytical reversed phase HPLC, ^{1}H-NMR, and MALDI-TOF mass spectroscopy. Polyamide **2** showed a significantly reduced retention time in reversed phase HPLC (21.9 min), relative to the all-Py analog, **1** (28.8 min).

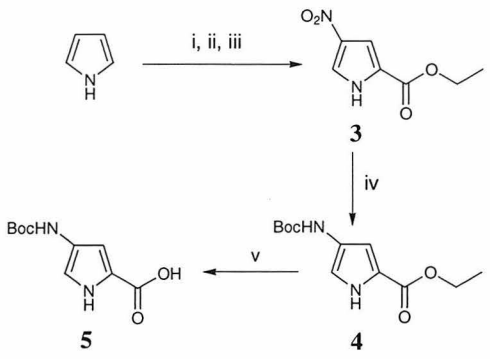


Figure 4.3. Synthesis of the Boc-Ds-acid monomer (**5**) for solid phase synthesis. (i) Trichloroacetyl chloride, ethyl ether. (ii) HNO_3 (fuming), H_2SO_4, acetic anhyride. (iii) NaH, ethanol. (iv) a) H_2 (1 atm), 10% Pd/C, DMF. b) Boc-anhyride, DIEA. (v) 1 M KOH, MeOH, water, 60 °C.

Identification of Binding Sites by MPE•Fe(II) Footprinting

MPE•Fe(II) footprinting was performed on the 3'- and 5'-^{32}P end-labeled 279 bp *Eco*RI/*Pvu*II restriction fragment of pSES4 (Figures 4.4, 4.5) (*4*). On the basis of the polyamide pairing rules, the sites 5'-TGTTAT-3' and 5'-TGGTAT-3' are for **1** and **2** "match" and "single base pair mismatch" sites, respectively (mismatch underlined). The footprint cleavage protection patterns of **1** and **2** revealed the polyamides bound the expected match site, 5'-TGTTAT-3' (Figure 4.6A-B). The location and size of the 3'-shifted footprints is consistent with the eight-ring polyamides binding as hairpins in the minor groove to the expected 6 bp target. A MPE cleavage protection pattern was also

observed at the mismatch site, with what appeared to be a potentially elongated footprint to the 3'-side of the bottom strand. Subsequent affinity cleavage experiments revealed the significance of this observation (*vide infra*). A number of additional footprints on the restriction fragment corresponding to match sequences of the form 5'-(A,T)G(A,T)$_4$-3' and a single base pair mismatch have been identified but were not quantitated.

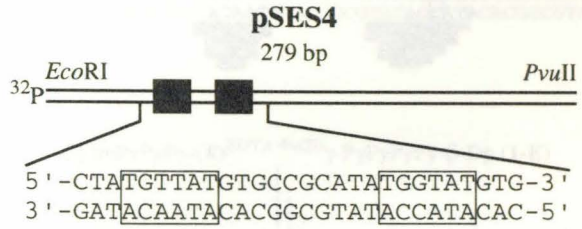


Figure 4.4. The 279 bp *Eco*RI/*Pvu*II restriction fragment of pSES4. The match and single base pair mismatch sequences, 5'-TGTTAT-3' and 5'-TGGTAT-3' (mismatch underlined), respectively, are boxed.

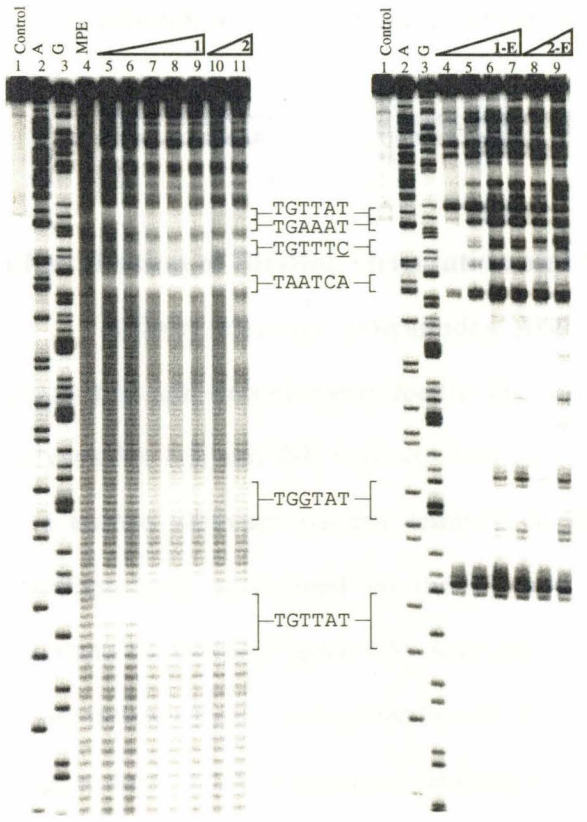


Figure 4.5. Storage phosphor autoradiograms of MPE•Fe(II) footprinting (left) and affinity cleavage experiments (right) on the 5'-^{32}P end-labeled 279 bp *Eco*RI/*Pvu*II restriction fragment of pSES4. Match and single base pair mismatch sites (mismatch underlined) are indicated in the middle. MPE•Fe(II) footprinting (left): Lane 1, Intact DNA; Lane 2, A reaction; Lane 3 G reaction; Lane 4, MPE•Fe(II) standard; Lane 5-9, 10 nM, 100 nM, 200 nM, 500 nM, 1 μM, **1**; Lane 10-11, 10 nM, 100 nM **2**. All reactions were performed in a final volume of 400 μL containing 20 kcpm 5'-radiolabeled DNA, 20 mM HEPES (pH 7.3), 200 mM NaCl, 50 μg/mL glycogen, 0.5 μM MPE•Fe(II), and 5 mM DTT. Linear storage phosphor range 160-1600. Affinity cleavage (right): Lane 1, Intact DNA; Lane 2, A reaction; Lane 3 G reaction; Lane 4-7, 1 nM, 10 nM, 100 nM, 1 μM **1-E**; Lane 8-9, 1 nM, 10 nM **2-E**. All reactions were performed in a final volume of 400 μL containing 20 kcpm 5'-radiolabeled DNA, 20 mM HEPES (pH 7.3), 200 mM NaCl, 50 μg/mL glycogen, 0.5 μM Fe(II)$(NH_4)_2SO_4$, and 5 mM DTT. Linear storage phosphor range 200-2000.

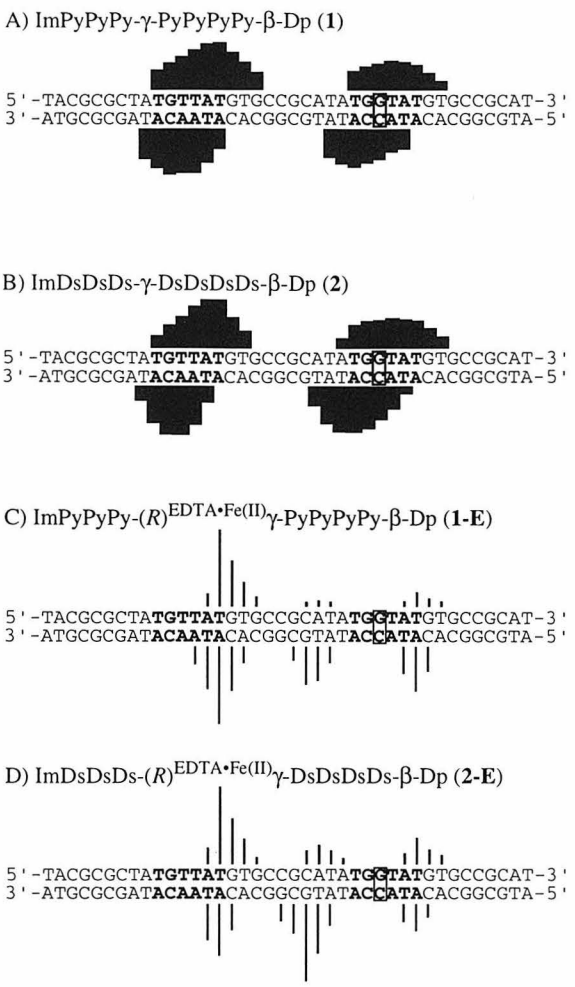


Figure 4.6. (A-B) MPE•Fe(II) protection patterns with (A) ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**) (500 nM), and (B) ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**) (100 nM). Bar heights are proportional to the relative protection from cleavage for each band. (C-D) Affinity cleavage patterns of (C) ImPyPyPy-γ-PyPyPyPy-β-Dp-EDTA•Fe (**1-E**) 1 μM, and (D) ImDsDsDs-γ-DsDsDsDs-β-Dp-EDTA•Fe (**2**) (10 nM). Line heights are proportional to the cleavage intensity of each band. The 6 bp polyamide binding site 5'-TGTTAT-3' and the designed single base pair mismatch 5'-TGGTAT-3'are shown in bold.

Identification of Binding Orientation and Stoichiometry by Affinity Cleavage

Affinity cleavage polyamides, **1-E** and **2-E**, were used to determine binding orientation and stoichiometry for the Ds polyamides (*8, 39*). The EDTA•Fe(II) moiety of polyamides **1-E** and **2-E** was attached via the γ-turn. The lack of a flexible linker results in a narrow cleavage pattern centered on the position of the γ-turn. Affinity cleavage titrations were performed on the 3'- and 5'-^{32}P end-labeled *Eco*RI/*Pvu*II restriction fragment of pSES4 (Figure 4.5). A single cleavage locus at the 3'-side of the 5'-TGTTAT-3' site indicated each polyamide bound the match site as a monomer in a single orientation (Figure 4.6 C-D). The expected cleavage pattern at the 3'-side of the 5'-TGGTAT-3' single base pair mismatch site was also observed for **1-E** and **2-E**. An unexpected cleavage locus between the designed match and mismatch sites was believed to result from the polyamide

bound in a reverse orientation (3'-5', N-C) with the γ-turn centered on the bold **T** in 5'-ATATGG-3'. This binding mode is also a single base pair mismatch according to the pairing rules, because it places the β-alanine residue over a C•G base pair. Binding to the designed mismatch and the reverse oritentation mismatch was observed for both the Py- and Ds-containing polyamides.

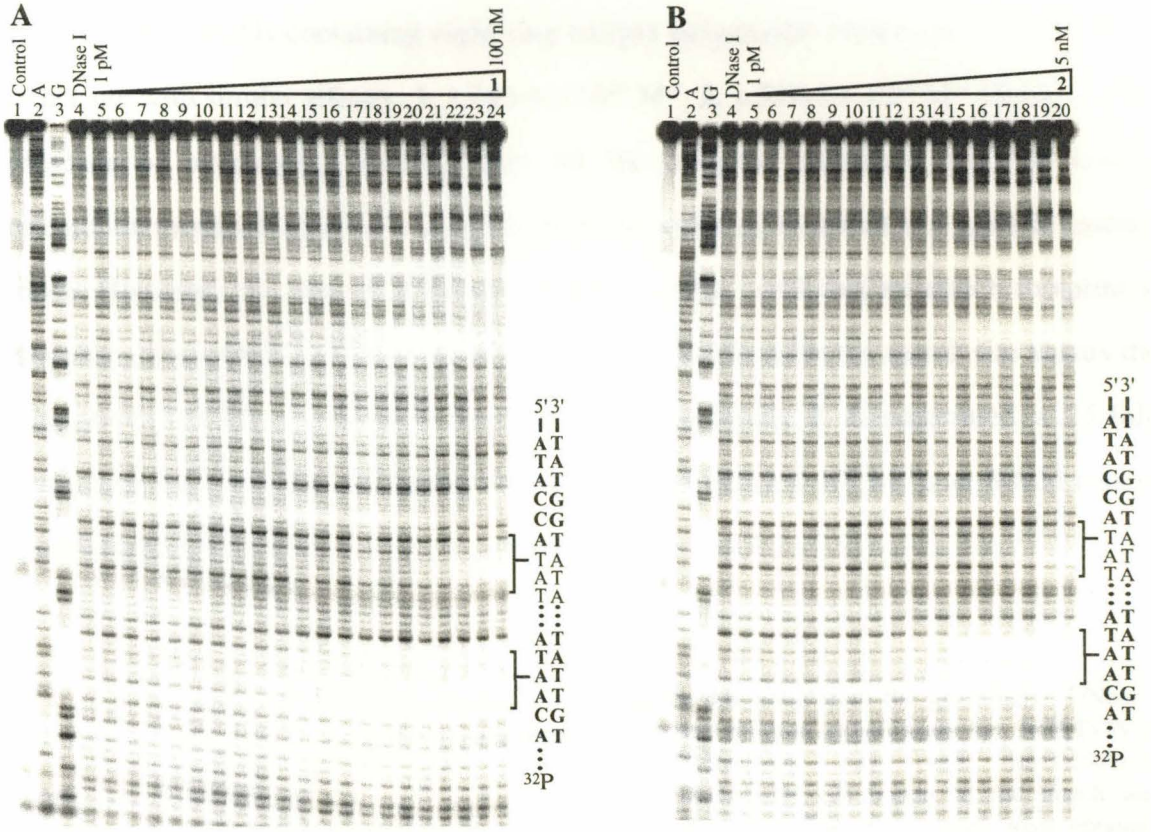


Figure 4.7. Storage phosphor autoradiograms of quantitative DNase I footprint titration experiments with (A) ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**) and (B) ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**), on the 3'-^{32}P-end-labeled 279 bp *Eco*RI/*Pvu*II restriction fragment from pSES4. All reactions contained 20 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$ and were performed at 22 °C. Lane 1, intact DNA; lane 2, A-specific reaction; lane 3, G-specific reaction; lane 4, DNase I standard; lanes 5-20, 1 pM, 2 pM, 5 pM, 10 pM, 15 pM, 25 pM, 40 pM, 65 pM, 100 pM, 150 pM, 250 pM, 400 pM, 650 pM, 1 nM, 2 nM, 5 nM, respectively of **1** or **2**. Footrpinting gels of **1** contained the additional lanes 21-24 of 10 nM, 20 nM, 50 nM, 100 nM, respectively. The positions of the match and mismatch sequences are indicated. The designed match (5'-TGTTAT) and single base pair mismatch (5'-TGGTAT-3') mismatch sites are in bold. The overlapping reverse orientation mismatch site observed by affinity cleavage (5'ATATGG-3') is also labeled. Linear storage phosphor ranges for A and B were 30-500, 80-1000, respectively.

Quantitative DNase I Footprint Titrations

Quantitative DNase I footprint titrations were performed on the 3'-^{32}P end-labeled *Eco*RI/*Pvu*II restriction fragment of pSES4 (Figure 4.7) (*40-42*). Equilibrium association constants (K_a) for each polyamide bound to the match and mismatch sites were determined by calculating a fractional saturation value at both sites for each polyamide concentration and fitting the data to a modified form of the Hill equation (Figure 4.8). Both the Py- and Ds-containing eight-ring hairpin polyamides bound the 5'-TGTTAT-3' target site with similar affinity: **1**, $1.2\pm0.4 \times 10^{10}$ M^{-1}; **2**, $1.5\pm0.2 \times 10^{10}$ M^{-1} (Table 1). The presence of overlapping binding sites for the designed mismatch and the reverse orientation mismatch precludes assignment of the resulting footprint to a specific binding mode. However, the relative affinity of **1** and **2** at this combined mismatch footprint is revealing. Polyamide **1** demonstrated >200-fold preference for the match site versus the mismatch ($5.9\pm0.9 \times 10^7$ M^{-1}). Ds-containing polyamide **2** demonstrated a 25-fold preference for its match site versus the mismatch ($5.9\pm2.3 \times 10^8$ M^{-1}), an 8-fold loss in specificity relative to **1**.

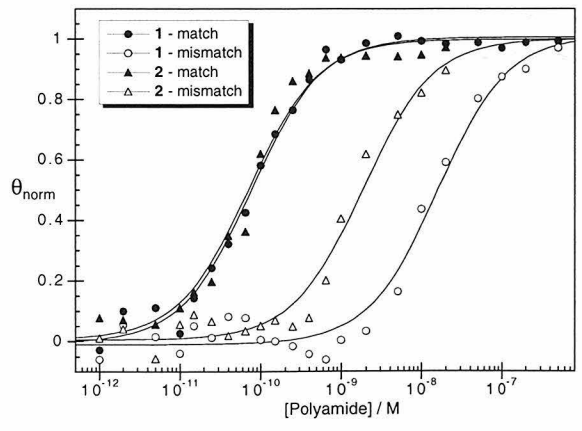


Figure 4.8. Data from the quantitative DNase I footprint titration experiments for ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**), and ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**) bound to the match and mismatch sites. The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography and processed as described in the experimental section. The data for the binding of **1** to the match and mismatch is indicated by filled and unfilled circles, respectively. The data for the binding of **2** to the match and mismatch is indicated by filled and unfilled triangles, respectively. The solid curves are the best-fit binding titration isotherms obtained from nonlinear least squares algorithm using eq. 2, n=1.

Table 4.1. Equilibrium Association Constants (M^{-1})[a]

Polyamide	Match[b]	Mismatch[c]	Specificity[d]
ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**)	1.2×10^{10}	5.9×10^{7}	203
ImDsDsDs-γ-DsDsDsDs-β-Dp (**2**)	1.5×10^{10}	5.9×10^{8}	25

[a]Values reported are the mean values obtained from three DNase I footprint titration experiments. The assays were carried out at 22 °C, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [b]The "match" affinity is for the polyamide binding the 5'-TGTTAT-3' target site. [c]The "mismatch" affinity is derived from the footprint arising from the binding of the overlapping reverse orientation and single base pair mismatch sites (see text). [d]Specificity is defined as K_a(match)/K_a(mismatch).

Discussion

In the reported syntheses of desmethylpyrrole analogs of distamycin, the N-1 position proved reactive under the coupling conditions, resulting in side reactions and lowering the yield of the desired product (*37*). The solid phase synthesis of Ds-containing polyamides also suffered from lower recoveries than the Py-containing analogs, suggesting that future application of Ds may require protection of the N-1 position.

MPE•Fe(II) footprinting and affinity cleavage studies revealed that replacement of the N-methyl of the pyrrole carboxamide with hydrogen had little effect on polyamide binding site size, orientation or stoichiometry at the 6-bp target site. The footprint protection patterns and cleavage loci at the match site were consistent with 1:1 polyamide:DNA complexes bound as hairpins in the minor groove. The expected affinity cleavage pattern was also observed at the 3'-side of the designed single base pair mismatch site. Affinity cleavage revealed the polyamide exhibited an additional unexpected binding mode assigned as reverse orientation (3'-5', N-C) binding to 5'-ATATGG-3'. The MPE footprints of **1** and **2** at the mismatch site are consistent with polyamide binding primarily at the designed mismatch site, 5'-TGGTAT-3', with only a slightly expanded footprint on the bottom strand perhaps resulting from polyamide binding to the reverse orientation site. Polyamides have demonstrated >10-fold preference for binding with the polyamide oriented N-C along the 5'-3' direction of the targeted DNA

strand (*21*). Furthermore, this orientation preference would be expected to be enhanced by the *R*-enantiomer of the affinity cleavage analogs (**1-E** and **2-E**), which would direct the bulky EDTA moiety into the floor of the minor groove if the polyamide bound in the reverse orientation (*32*). Thus, it was surprising that **1-E** and **2-E** bound both the designed mismatch and the reverse orientation mismatch with similar relative affinities, as revealed by affinity cleavage. It is important to note that the incorporation of Ds resulted in enhanced binding relative to **1** for both of these mismatch modes, as evidenced by cleavage at *both* mismatch sites at lower concentration for **2** (10 nM) than **1** (1 μM) (Figure 4.5). In other words, the removal of the N-methyl substituent does not result in a loss of specficity versus one mismatch binding mode, but rather versus both mismatches observed here.

As expected, polyamide **1** bound the 5'-TGTTAT-3' site with subnanomolar affinity and excellent specificity. Remarkably, replacing all of the N-methylpyrroles with desmethylpyrrole in **2** did not have a significant effect on the polyamide's DNA binding affinity to the target site. Polyamide **2** bound the target site with similar affinity to **1**, revealing the N-methyl group is not necessary for subnanomolar binding by hairpin polyamides. However, removal of the N-methyl substituents resulted in an 8-fold loss in specificity for **2**, relative to **1**. The overlapping binding modes at the mismatch site precludes the development of a model based on the structure of the polyamide:DNA complex to account for the higher affinity at the mismatch site exhibited by Ds-containing polyamides. However, in the DNase I footprint titrations, the mismatch footprint clearly appears at approximately 10-fold lower polyamide concentrations for **2** than **1**. The affinity cleavage results strongly suggest that this DNase I footprint arises from a combination of both mismatch binding modes.

It was unexpected that incorporation of Ds would result in differential polyamide-DNA interactions at the match and mismatch sites. Further studies, including structural studies of Ds-containing polyamides bound to match and mismatch sites, will be

necessary to elucidate the basis for the decrease in specificity upon Ds-introduction. The N-methyl moiety in the ring-ring pairs may be necessary to keep the pyrrole rings set deeply in the minor groove. In the absence of the N-methyl, the energetic penalty may be decreased for the exocyclic amine of guanine pushing the ring away from the DNA, allowing for an increased tolerance of G•C base pairs over Ds/Ds pairs. Such an arrangement would align the positive electrostatic potential of the N-H on Ds opposite the negative potential in the center of the Py ring (Figure 4.9). The N-methyl may also play a role in aligning the side-by-side register of the ring pairings. Furthermore, it may be possible for some or all of the Ds carboxamides to rotate about the amide backbone and place the N-1 toward the floor of the minor groove. The implications on specificity of such a structural alteration at the polyamide:DNA interface may be significant.

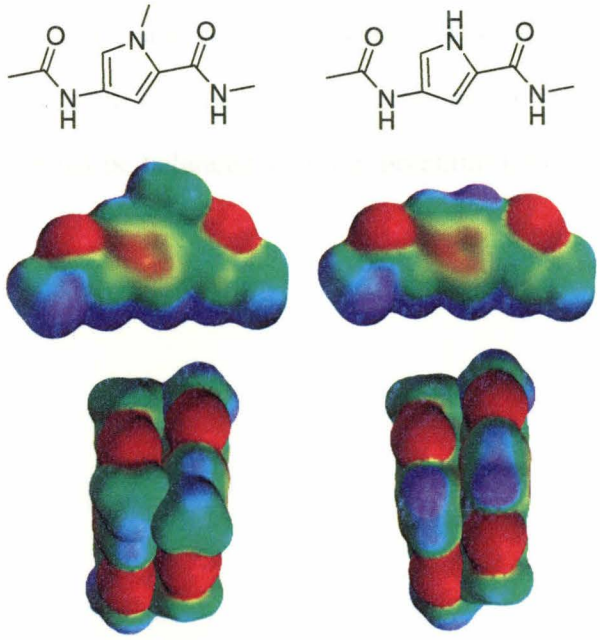


Figure 4.9. Electrostatic potential maps calculated for Ac-Py-Me and Ac-Ds-Me. (top) Chemical structures used for calculations. (middle) Side view of electrostatic potentials mapped onto electron density. Electrostatic potentials were calculated using Spartan SGI 5.0 as described *(18)*. Areas of positive potential are represented as blue surface, areas of negative potential are depicted as red surface. Neutral regions appear green. (bottom) Electrostatic potential maps superimposed on a Py/Py pair from ImPyPyPy X-ray crystal structure viewed from above *(20)*.

Conclusions

The introduction of desmethylpyrrole into the hairpin polyamide motif has provided insight into the role of the N-methyl substituent in polyamide:DNA recognition. While the N-methyl is not necessary for subnanomolar binding by hairpin polyamides, it is clearly important for attaining optimal target site specificity. It is now becoming clear that a complete understanding of polyamide:DNA recognition will require not only structural studies of a variety of polyamides bound to their DNA targets, but of equal importance will be an analysis of the structural basis of specificity achieved by analyzing structures of polyamides bound to mismatch DNA sites. Current efforts to develop larger polyamides that bind longer sequences may face difficulty with issues of solubility and cell permeability. Desmethylpyrrole containing polyamides exhibit higher solubility than their N-methylpyrrole analogs and may potentially alter polyamide cell permeability, making them useful alternatives in the design of second-generation polyamides to recognize larger binding sites for genomic applications. In the future, the benefits of each incorporation of Ds must be balanced with the potential loss in sequence specificity.

Experimental

All synthetic reagents were as previously described (*30*). Analytical HPLC was performed on a Beckman *Gold Nouveau* system with a model 126 pump and model 168 diode array detector. A Rainen C_{18}, Microsorb MV, 5 μm, 300 × 4.6 mm reverse phase column was employed with 0.1% (w/v) TFA:H_2O and 1.5% acetonitrile/min. Preparatory HPLC was performed on a Hamilton PRP-1 reversed phase C_{18}, 250 mm × 21.5 mm, 12-20 μm column with 0.5% (w/v) acetic acid and 0.56%/min acetonitrile unless otherwise noted. Resin substitution of synthesized polyamides was calculated as $L_{new}(mmol/g) = L_{old}/(1 + L_{old}(W_{new} - W_{old}) \times 10^{-3})$ where L is the loading (mmol of amine per gram of resin) and W is the weight (g mol^{-1}) of the growing peptide attached to the resin (*30*). DNA restriction fragment labeling, DNase I footrpinting, and determination of equilibrium association constants were accomplished using previously described protocols (*41, 42*). Chemical sequncing reactions were performed according to published methods (*43, 44*). Plasmid pSES4 was previously prepared by S. Swalley using reported procedures with the following inserts: 5'-GATCGGCTACGCGCTATGTTATGTGCCGCATATGGTAT GTGCCGCATATGGGATGTGCGCCATATGGGGTGTGCCG-3' and 5'-AGCTCG GCACACCCCATATGGCGCACATCCCATATGCGGCACATACCATATGCGGC ACATAACATAGCGCGTAGCC (*31, 42*).

Monomer Synthesis

Ethyl 4-nitropyrrole-2-carboxylate (**3**). 2-(trichloroacetyl)pyrrole was available in kilogram quantities by scaling the literature procedure (*45*). Nitration of 2-(trichloroacetyl)pyrrole was accomplished using the previously reported procedure for 1-methyl-2-(trichloroacetyl)pyrrole with minor alterations (*30*). The addition of nitric acid was performed at –20 °C and the reaction was quenched with water. Esterification of the 2-(trichloroacetyl)-4-nitropyrrole was accomplshied using the procedure reported for the N-methyl analog without reflux (*30*). The solid isolated after precipitaiton with water was

recrystalized from ethyl acetate:hexanes:methanol to provide a light tan powder (551 g, 3.0 mol, 22% yield from pyrrole): ^{1}H NMR (DMSO-*d6*) δ 13.13 (s, 1H), 8.01 (d, 1H, J=1.5 Hz), 7.20 (d, 1H, J=1.9 Hz), 4.25 (q, 2H, J=7.2 Hz), 1.27 (t, 3H, J=7.2 Hz); EIMS m/e 184.0482 (184.0484 calcd for $C_7H_8N_2O_4$).

Ethyl 4-[(*tert*-butoxycarbonyl)amino]pyrrole-2carboxylate (4). Ethyl 4-nitropyrrole-2-carboxylate (128 g, 0.75 mol) was dissolved in DMF (640 mL) and 23 g of 10% Pd/C added. The mixture was stirred under H_2 (50 psi) for 2 h. The Pd/C was removed by filtration through Celite, followed by washing with DMF (100 mL). Di-*tert*-butyl dicarbonate (164 g, 0.75 mol) and triethylamine (315 mL, 2.26 mol) was added and stirred for 1.25 h. The mixture was poured into 1 L of water and extracted with diethyl ether. The ether was passed over solid sodium bicarbonate, decanted, dried ($MgSO_4$) and concentrated to *in vacuo* to provide a white solid (125 g, 0.49 mol, 65% yield): ^{1}H NMR (DMSO-*d6*) δ 11.50 (s, 1H), 9.08 (s, 1H), 6.96 (s, 1H), 6.59 (s, 1H), 4.18 (q, 2H, J=7.2 Hz), 1.42 (s, 9H), 1.24 (t, 3H, J=7.1 Hz); FABMS m/e (254.1266 calcd for $C_{12}H_{18}N_2O_4$).

4-[(*tert*-Butoxycarbonyl)amino]pyrrole-2 carboxylic acid (5). Ethyl 4-[(*tert*-butoxycarbonyl)amino]pyrrole-2carboxylate (31.0 g, 122 mmol) was dissolved in ethanol (124 mL); 1 M KOH (620 mL) was added and the solution heated at 65 °C for 4 h. The solution was cooled to room temperature and the ethanol removed *in vacuo*. The solution was diluted with water (800 mL) and extracted with diethyl ether. The pH of the aqueous layer was reduced to ca. 3 with 10% (v/v) H_2SO_4 and the mixture extracted with diethyl ether. The combined ether extracts were dried ($MgSO_4$) and concentrated *in vacuo* to provide a white powder (25.6 g, 113 mmol, 93% yield): ^{1}H NMR (DMSO-*d6*) δ 12.17 (s, 1H), 11.34 (s, 1H), 9.03 (s, 1H), 6.90 (s, 1H), 6.53 (s, 1H), 1.42 (s, 9H); FABMS m/e 249.0841 (M+Na 249.0851 calcd for $C_{10}H_{14}N_2NaO_4$).

Polyamide Synthesis

ImDsDsDs-γ-DsDsDsDs-β-Dp (2). ImDsDsDs-γ-DsDsDsDs-β-PAM-resin was synthesized in a stepwise fashion by manual solid phase methods from Boc-β-PAM-resin (600 mg, 0.75 mmol/g) (*30*). A sample of polyamide resin (217 mg, 0.46 mmol/g) was cleaved with dimethylaminopropylamine (2 mL, 55 °C, 18 hr.). The crude polyamide was diluted to 8 mL with 0.1% (w/v) TFA then purified by reverse phase HPLC using a Waters DeltaPak 25 × 100 mm 100 μm C_{18} column in 0.1% (w/v) TFA, gradient elution 0.25%/min. CH_3CN. ImDsDsDs-γ-DsDsDsDs-β-Dp was recovered upon lyophilization of the appropriate fractions as a white powder (1.7 mg, 1.5 μmol, 1.5% recovery); UV (H_20) λ_{max} 312 (66,000); analytical reverse phase HPLC rt=21.9 min; ^{1}H NMR (DMSO-d_6): δ 11.28 (s, 1H), 11.20 (s, 4H), 11.11 (s, 2H), 10.42 (s, 1H), 9.93 (s, 2H), 9.88 (s, 3H), 9.82 (s, 1H), 9.2 (br s, 1H), 8.04 (m, 3H), 7.35 (s, 1H), 7.23 (s, 1H), 7.17 (s, 3H), 7.14 (s, 1H), 7.13 (s, 1H), 7.06 (s, 1H), 7.04 (s, 1H), 7.02 (s, 2H), 7.01 (s, 2H), 6.90 (s, 1H), 6.83 (s, 1H), 6.81 (s, 1H), 3.95 (s, 3H), 3.61 (m, 2H), 3.19 (m, 2H), 3.06 (quartet, 2H, J=5.8 Hz), 2.93 (quintet, 2H, J=4.6 Hz), 2.68 (d, 6H, J=4.6 Hz), 2.2 (br m, 4H), 1.92 (m, 2H), 1.7 (br m, 2H); MALDI-TOF-MS (monoisotopic) [M+H] 1123.6, 1123.5 calc for $C_{52}H_{59}N_{20}O_{10}^+$.

ImDsDsDs-(*R*)^{EDTA}γ-DsDsDsDs-β-Dp (2-E). ImDsDsDs-(*R*)^{NH2}γ-DsDsDsDs-β-PAM-resin was synthesized in a stepwise fashion by manual solid phase methods from Boc-β-PAM-resin (600 mg, 0.75 mmol/g) (*30*). A sample of polyamide resin (215 mg, 0.46 mmol/g) was cleaved with dimethylaminopropylamine (2 mL, 55 °C, 18 hr.). The crude polyamide was diluted to 8 mL with 0.1% (w/v) TFA then purified by reverse phase HPLC using a Waters DeltaPak 25 × 100 mm 100 μm C_{18} column in 0.1% (w/v) TFA, gradient elution 0.25%/min. CH_3CN. ImDsDsDs-(*R*)^{NH2}γ-DsDsDsDs-β-Dp was recovered upon lyophilization of the appropriate fractions as a white powder (2.1 mg, 1.9 μmol, 1.9% recovery); UV (H_20) λ_{max} 312 (66,000); ^{1}H NMR (DMSO-d_6): δ 11.42 (s, 1H), 11.20 (m, 5H), 11.11 (s, 1H), 10.60 (s, 1H), 10.41 (s, 1H), 9.95 (s, 1H), 9.92 (s, 3H), 9.87 (s, 1H), 9.2 (br s, 1H), 8.27 (m, 2H), 8.04 (m, 1H), 7.35 (s, 1H), 7.23 (s, 1H), 7.17 (s, 4H), 7.12

(s, 1H), 7.02 (m, 6H), 6.95 (s, 1H), 6.90 (s, 1H), 6.82 (s, 1H), 3.95 (s, 3H), 3.31 (m, 2H), 3.20 (s, 2H), 3.06 (quartet, 2H, J=5.6 Hz), 2.93 (quintet, 2H, J=4.9 Hz), 2.69 (d, 6H, J=4.5 Hz), 2.2 (br m, 4H), 1.95 (m, 2H), 1.64 (m, 2H); MALDI-TOF-MS (monoisotopic) [M+H] 1138.5, 1138.5 calc for $C_{52}H_{60}N_{21}O_{10}^{+}$. ImDsDsDs-(*R*)^{NH2}γ-DsDsDsDs-β-Dp (700 nmol) was functionalized with EDTA as described (*32*) (68 nmol, 9.7% recovery) MALDI-TOF-MS (monoisotopic) [M+H] 1412.7, 1412.6 calc for $C_{62}H_{74}N_{23}O_{17}^{+}$.

ImPyPyPy-(*R*)^{EDTA}γ-PyPyPyPy-β-Dp (1-E). ImPyPyPy-(*R*)^{NH2}γ-PyPyPyPy-β-PAM-resin was synthesized in a stepwise fashion by manual solid phase methods from Boc-β-PAM-resin (600 mg, 0.75 mmol/g) (*30*). A sample of polyamide resin (350 mg, 0.44 mmol/g) was cleaved with dimethylaminopropylamine (2 mL, 55 °C, 18 hr.). The crude polyamide was diluted to 8 mL with 0.1% (w/v) TFA then purified by reverse phase HPLC. ImPyPyPy-(*R*)^{NH2}γ-PyPyPyPy-β-Dp was recovered upon lyophilization of the appropriate fractions as a white powder (25.3 mg, 20.5 μmol, 13.3% recovery); UV (H_20) λ_{max} 312 (66,000); ^{1}H NMR (DMSO-d_6): δ 10.40 (s, 1H), 9.93 (s, 1H), 9.92 (s, 1H), 9.89 (s, 2H), 9.87 (s, 2H), 9.75 (br s, 1H), 8.05 (m, 1H), 7.97 (m, 1H), 7.84 (t, 1H, J=5.4 Hz), 7.35 (s, 1H), 7.24 (d, 1H, J=0.9 Hz), 7.20 (s, 2H), 7.16 (s, 1H), 7.14 (s, 2H), 7.12 (m, 2H), 7.01 (s, 2H), 6.99 (s, 2H), 6.92 (s, 1H), 6.83 (s, 1H), 6.78 (s, 1H), 3.94 (s, 3H), 3.80 (s, 12H), 3.75 (m, 9H), 3.23 (m, 5H), 3.00 (m, 2H), 2.23 (m, 2H), 2.16 (m, 2H), 2.04 (m, 2H), 2.03 (s, 6H), 1.42 (m, 2H); MALDI-TOF-MS (monoisotopic) [M+H] 1236.7, 1236.6 calc for $C_{59}H_{74}N_{21}O_{10}^{+}$. ImPyPyPy-(*R*)^{NH2}γ-PyPyPyPy-β-Dp (6.4 mg, 5.2 μmol) was functionalized with EDTA as described (*32*) (3.5 mg, 2.3 μmol, 43% recovery) MALDI-TOF-MS (monoisotopic) [M+H] 1510.9, 1510.7 calc for $C_{69}H_{88}N_{23}O_{17}^{+}$.

MPE Footprinting and Affinity Cleavage

All reactions were carried out in a final volume of 400 μL. A polyamide stock solution (or water for reference lanes) was added to an assay buffer where the final solution conditions were as follows: 20 mM HEPES (pH 7.3), 200 mM NaCl, 50 μg/mL

glycogen and 20 kcpm 3'- or 5'-end-labeled *Eco*RI/*Pvu*II fragement of pSES4. The reactions were equilibrated for 16 hours at 22 °C. For MPE footprinting, MPE•Fe(II) was added to a final concentration of 0.5 μM and equilibrated for 5 minutes (*4*). For affinity cleavage experiments, $Fe(NH_4)_2(SO_4)_2$ was added to a final concentration of 0.5 μM and equilibrated for 20 minutes (*8, 39*). Cleavage was initiated by the addition of DTT to a final concentration of 5 mM and allowed to proceed for 30 minutes. Reactions were terminated with ethanol (1 mL) and 10 μL of a precipitation buffer (2.8 mg/mL glycogen, 140 μM bp calf thymus DNA). The labeled DNA was precipitated and analyzed on an 8% denaturing polyacrylamide gel as describded for DNase I footprinting (*41, 42*).

Acknowledgments

We are grateful to the National Institutes of Health for research support, the National Science Foundation, Bristol-Myers Squibb and the Ralph M. Parsons Foundation for predoctoral fellowships to R.E.B., the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B., and the National Institutes of Health for a research service award to J.W.S.

References

1. Arcamone, F., Penco, S., Prezzi, P. G., Nicolella, V., and Pirelli, A. (1964) *Nature 203*, 1064.

2. Zasedatelev, A. S., Gursky, G. V., Zimmer, C., and Thrum, H. (1974) *Mol. Biol. Rep. 1*, 337-342.

3. Krylov, A. S., Grokhovsky, S. L., Zasedatelev, A. S., Zhuze, A. L., Gursky, G. V., and Gottikh, B. P. (1979) *Nucleic Acids Res. 6*, 289-304.

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

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

6. Fox, K. R., and Waring, M. J. (1984) *Nucleic Acids Res. 12*, 9271-9285.

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

8. Dervan, P. B. (1986) *Science 232*, 464-471.

9. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., and Dickerson, R. E. (1985) *J. Mol. Biol. 183*, 553-563.

10. Coll, M., Frederick, C. A., Wang, A. H. J., and Rich, A. (1987) *Proc. Natl. Acad. Sci. USA 84*, 8385-8389.

11. Klevit, R. E., Wemmer, D. E., and Reid, B. R. (1986) *Biochemistry 25*, 3296-3303.

12. Pelton, J. G., and Wemmer, D. E. (1988) *Biochemistry 27*, 8088-8096.

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

14. Chen, X., Ramakrishnan, B., Rao, S. T., and Sundaralingam, M. (1994) *Nat. Struct. Biol. 1*, 169-175.

15. Dervan, P. B., and Bürli, R. W. (1999) *Curr. Opin. Chem. Biol. 3*, 688-693.

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

17. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. (1998)

Nature 391, 468-471.

18. White, S., Baird, E. E., and Dervan, P. B. (1997) *Chem. Biol. 4*, 569-578.

19. Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E., and Dervan, P. B. (1992) *Proc. Natl. Acad. Sci. USA 89*, 7586-7590.

20. Kielkopf, C. L., Baird, E. E., Dervan, P. D., and Rees, D. C. (1998) *Nat. Struct. Biol. 5*, 104-109.

21. White, S., Baird, E. E., and Dervan, P. B. (1997) *J. Am. Chem. Soc. 119*, 8756-8765.

22. Pilch, D. S., Poklar, N., Baird, E. E., Dervan, P. B., and Breslauer, K. J. (1999) *Biochemistry 38*, 2143-2151.

23. Pelton, J. G., and Wemmer, D. E. (1990) *J. Am. Chem. Soc. 112*, 1393-1399.

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

25. Urbach, A. R., Szewczyk, J. W., White, S., Turner, J. M., Baird, E. E., and Dervan, P. B. (1999) *J. Am. Chem. Soc. 121*, 11621-11629.

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

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

28. Pilch, D. S., Poklar, N., Gelfand, C. A., Law, S. M., Breslauer, K. J., Baird, E. E., and Dervan, P. B. (1996) *Proc. Natl. Acad. Sci. USA 93*, 8306-8311.

29. deClairac, R. P. L., Geierstanger, B. H., Mrksich, M., Dervan, P. B., and Wemmer, D. E. (1997) *J. Am. Chem. Soc. 119*, 7909-7916.

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

31. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1999) *J. Am. Chem. Soc. 121*, 1113-1120.

32. Herman, D. M., Baird, E. E., and Dervan, P. B. (1998) *J. Am. Chem. Soc. 120*, 1382-1391.

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

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

35. Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan, P. B. (1998) *Proc. Natl. Acad. Sci. USA 95*, 12890-12895.

36. deClairac, R. P. L., Seel, C. J., Geierstanger, B. H., Mrksich, M., Baird, E. E., Dervan, P. B., and Wemmer, D. E. (1999) *J. Am. Chem. Soc. 121*, 2956-2964.

37. Grehn, L., Ragnarsson, U., Eriksson, B., and Oberg, B. (1983) *J. Med. Chem. 26*, 1042-1049.

38. Grehn, L., Ragnarsson, U., and Datema, R. (1986) *Acta Chem. Scand. B 40*, 145-151.

39. Taylor, J. S., Schultz, P. G., and Dervan, P. B. (1984) *Tetrahedron 40*, 457-465.

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

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

42. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) *Chem.-Eur. J. 3*, 1600-1607.

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

44. Maxam, A. M., and Gilbert, W. S. (1980) *Methods Enzymol. 65*, 499-560.

45. Bailey, D. M., Johnson, R. E., and Albertson, N. F. (1988) *Org. Synth. 50*, 618-619.

Chapter 5

Structural Effects of DNA Sequence on T•A Recognition by Hydroxypyrrole/Pyrrole Pairs in the Minor Groove

The text of this chapter was taken in part from a publication coauthored with Clara L. Kielkopf, Sarah White, Jason W. Szewczyk, James M. Turner, Eldon E. Baird, Prof. Peter B. Dervan & Prof. Douglas C. Rees. X-ray crystallography and structure analysis were performed by C.L.K. in the Rees laboratory at Caltech. Polyamide synthesis and footprinting were performed in the Dervan laboratory.

(C. L. Kielkopf, R. E. Bremer, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan & D. C. Rees *J. Mol. Biol.*, **2000**, *295*, 557-567.)

Abstract: Synthetic polyamides composed of three types of aromatic amino acids, N-methylimidazole (Im), N-methylpyrrole (Py) and N-methyl-3-hydroxypyrrole (Hp) bind specific DNA sequences as antiparallel dimers in the minor groove. The side-by-side pairings of aromatic rings in the dimer afford a general recognition code that allows all four base pairs to be distinguished. To examine the structural consequences of changing the DNA sequence context on T•A recognition by Hp/Py pairs in the minor groove, crystal structures of polyamide dimers $(ImPyHpPy)_2$ and the pyrrole counterpart $(ImPyPyPy)_2$ bound to the six base pair target site 5'-AGATCT-3' in a ten base pair oligonucleotide have been determined to a resolution of 2.27 and 2.15 Å, respectively. The structures demonstrate that the principles of Hp/Py recognition of T•A are consistent between different sequence contexts. However, a general structural explanation for the non-additive reduction in binding affinity due to introduction of the hydroxyl group is less clear. Comparison with other polyamide-DNA cocrystal structures reveals structural themes and differences that may relate to sequence preference.

Introduction

Polyamides are synthetic ligands composed of N-methylimidazole (Im), N-methylpyrrole (Py) and N-methyl-3-hydroxypyrrole (Hp) amino acids, that target specific DNA sequences with affinities and specificities that can rival those of DNA binding proteins (*1-6*). These molecules bind as antiparallel dimers in the minor groove, and use orbital asymmetry to distinguish each of the four base pairs in the minor groove (*7-13*). An Im/Py pair discriminates a G•C from C•G, and both from A•T/T•A base pairs by forming a directional hydrogen bond from the Im-N3 to the free guanine-(*9, 13-17*). Although a Py/Py pair is partially degenerate for T•A and A•T base pairs, an Hp/Py pair distinguishes T•A from A•T and G•C/C•G base pairs using a combination of specific hydrogen bonds between the hydroxyl with the thymine-O2, and shape selective recognition of an asymmetric cleft between the thymine-O2 and adenine-C2 (*12, 18-22*). Additionally, a C-terminal β-alanine residue (β) preceding a dimethylpropylamine (Dp) enhances both DNA-binding affinity and specificity and displays a strong preference for flanking A•T/T•A base pairs (*23, 24*).

The generality of these pairing rules for minor groove recognition is somewhat remarkable given the sequence-dependent microstructure of the minor groove of B-DNA (*25-27*). In fact, we now know that certain classes of base pair combinations are more difficult to distinguish, including 5'-G*A*-3' steps and 5'-GN*G*-3' sequences (problematic nucleotide italicized) (*3, 20*). Both Py/Py and Hp/Py pairs display higher affinity for 5'-G*T*-3' than 5'-G*A*-3' sequences (*20-22*). Although recognition of 5'-GN*G*-3' is also difficult, it was found empirically that a flexible β-alanine residue preceding the Im residue restores specificity and high affinity for these sequences (*28*). Although the preference for 5'-G*T*-3' over 5'-G*A*-3' steps has not yet been addressed, higher affinity recognition of 5'-GN*G*-3' sequences shows that even in difficult cases polyamide motifs can be fine-tuned to provide minor structural adjustments that restore predictable specificity.

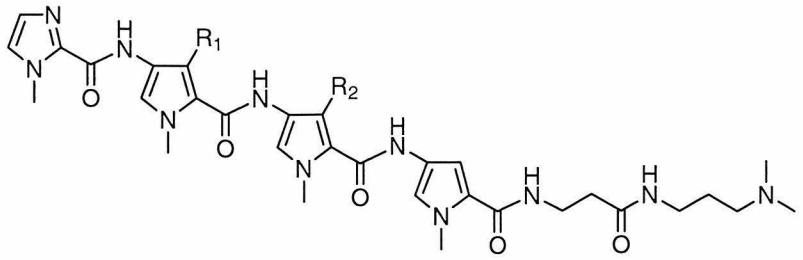
The X-ray structural analysis of the polyamide ImHpPyPyβDp bound to 5'-CC**AGTACT**GG-3' (binding site in bold) suggested that elongation of the Watson-Crick hydrogen bonds of the target T•A base pairs or bifurcation of the amide-DNA hydrogen bond following the Hp residue by the hydroxyl group may contribute to a slight reduction in binding affinity for Hp- containing polyamide to the Py- counterpart (*12*). The generality of these observations for different TA sequences and the structural basis of the preference for 5'-G*T*-3' over 5'-G*A*-3' steps was unknown. To investigate this, DNA binding affinities and the crystal structures of ImPyHpPyβDp•5'-CC**AG*A*TCT**GG-3' and its pyrrole-counterpart ImPyPyPyβDp•5'-CC**AG*A*TCT**GG-3' are presented here. A detailed structural understanding of polyamide-DNA interactions will contribute to development of these molecules as reagents for biological research, for which the ability to target specific genes is essential.

Results

Polyamide Binding Affinities

In order to attempt to correlate relative binding affinities of Py/Py and Hp/Py containing polyamides with the observed differences in the cocrystal structures, apparent equilibrium association constants were determined by quantitative DNase I footprinting for ImPyPyPy-β-Dp (**1**), ImHpPyPy-β-Dp (**2**), and ImPyHpPy-β-Dp (**3**) bound to the sites 5'-AGTACT-3', 5'-AGAACT-3', and 5'-AGATCT-3' (Figures 5.1 and 5.2). The cocrystal structures of ImPyPyPy-β-Dp and ImHpPyPy-β-Dp bound to 5'-AGTACT-3' have been described, while the structures ImPyPyPy-β-Dp•5'-AGATCT-3' and ImHpPyPy-β-Dp•5'-AGATCT-3' are reported here (*12*). ImPyPyPy-β-Dp bound with highest affinity to the 5'-AGTACT-3' site with an apparent equilibrium association constant of 2.2×10^7 M^{-1} (Table 5.1). While the 5'-AGAACT-3' and 5'-AGATCT-3' sites are also match sites for **1** according to the pairing rules, lower affinity was observed for **1** binding to the these sites ($K_{a, app}$=2.0×10^6 M^{-1}, and 2.6×10^6 M^{-1}, respectively). This observation is consistent with

the range of affinities observed for Py/Py recognition of A•T and T•A base pairs in different sequence contexts (*20*). Lower relative binding affinity is frequently observed when a 5'-GA-3' step is present in the DNA target site (*20*).



1) R_1=H, R_2=H, ImPyPyPy-β-Dp

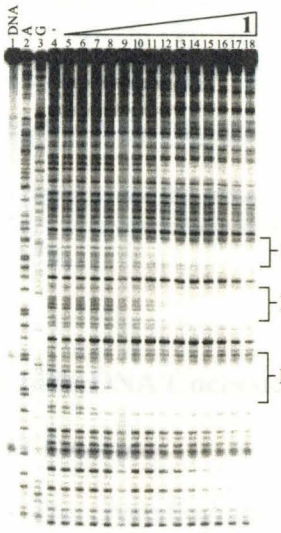
2) R_1=OH, R_2=H, ImHpPyPy-β-Dp

3) R_1=H, R_2=OH, ImPyHpPy-β-Dp

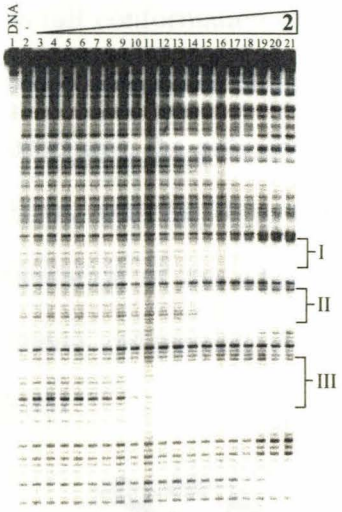
Figure 5.1. Chemical structures of the polyamides ImPyPyPy-β-Dp (**1**), ImHpPyPy-β-Dp (**2**), and ImPyHpPy-β-Dp (**3**).

The introduction of an Hp residue affords ImHpPyPy-β-Dp, which was designed to bind 5'-AGTACT-3' with a preference over 5'-AGAACT-3' and 5'-AGATCT-3'. Polyamide **2** bound these sites with affinities of 1.8×10^6 M^{-1}, 1.8×10^5 M^{-1}, and 1.9×10^5 M^{-1}, respectively (Table 5.1). The addition of the Hp residue resulted in a 12-fold loss in affinity for the match site relative to **1**. ImPyHpPy-β-Dp bound its designed match site, 5'-AGATCT-3', with an apparent equilibrium association constant 9.3×10^4 M^{-1}, a 27-fold loss in affinity relative to **1** bound to this site. While ImPyHpPy-β-Dp bound the formal mismatch sites 5'-AGTACT-3' ($K_{a,\,app}$=1.9×10^5 M^{-1}) and 5'-AGAACT-3' ($K_{a,\,app}$=7.7×10^4 M^{-1}) with similar affinities, these represent 115-fold and 25-fold losses in affinity relative to **1** bound to these sites, respectively. The Hp residue appears to discriminate against the Hp/Py pair mismatched with an A•T base pair; however, the Hp destabilization over an adenine did not overcome the inherent polyamide preference for the 5'-AGTACT-3' site. Structural features of polyamide:DNA recognition (5'-GT-3' preference over 5'-GA-3') can rival the formal pairing rules.

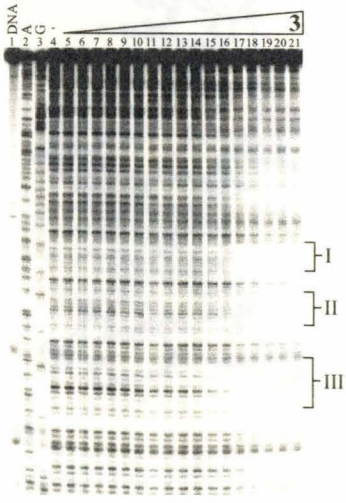
Figure 5.2. Quantitative DNase I footprint titration experiments with ImPyPyPy-β-Dp (**1**, top), ImHpPyPy-β-Dp (**2**, middle), and ImPyHpPy-β-Dp (**3**, bottom), on the 3'-^{32}P-end-labeled *Eco* RI/*Pvu* II restriction fragment from plasmid pDEH10. Left, storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated from DNase I digestion. All reactions contained 20 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$ and were performed in a volume of 40 μL at 22 °C. Top: Lane 1, intact DNA; lane 2, A-specific reaction; lane 3, G-specific reaction; lane 4, DNase I standard; lanes 5-18, 10 nM, 20 nM, 50 nM, 100 nM, 150 nM, 250 nM, 400 nM, 650 nM, 1 μM, 1.5 μM, 2.5 μM, 4 μM, 6.5 μM, 10 μM, respectively. Middle: Lane 1, intact DNA; lane 2, DNase I standard; lanes 3-21, 10 nM, 20 nM, 50 nM, 100 nM, 150 nM, 250 nM, 400 nM, 650 nM, 1 μM, 1.5 μM, 2.5 μM, 4 μM, 6.5 μM, 10 μM, 15 μM, 25 μM, 40 μM, 65 μM, 100 μM, respectively. Bottom: Lane 1, intact DNA; lane 2, A-specific reaction; lane 3, G-specific reaction; lane 4, DNase I standard; lanes 5-21, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM, 1.5 μM, 2.5 μM, 4 μM, 6.5 μM, 10 μM, 15 μM, 25 μM, 40 μM, 65 μM, 100 μM, respectively. Right, models of polyamide bound to the match or mismatch sites on the DNA fragment. Red, blue and green circles represent imidazole, pyrrole, and 3-hydroxypyrrole rings, respectively. The restriction fragment was 3'-^{32}P-end labeled on the bottom strand. Apparent equilibrium association constants of each polyamide binding to the match or mismatch site is reported on the right. Each value is the average of at least three experiments.



	ImPyPyPy-β-Dp	$K_{a,\,app}$ (M^{-1})
I	5′-T T **A G A T C T** T G-3′ 3′-A A **T C T A G A** A C-5′	2.6×10^6
II	5′-T T **A G A A C T** T G-3′ 3′-A A **T C T T G A** A C-5′	2.0×10^6
III	5′-T T **A G T A C T** T G-3′ 3′-A A **T C A T G A** A C-5′	2.2×10^7



	ImHpPyPy-β-Dp	$K_{a,\,app}$ (M^{-1})
I	5′-T T **A G** A T **C T** T G-3′ 3′-A A **T C** T A **G A** A C-5′	1.9×10^5
II	5′-T T **A G** A A **C T** T G-3′ 3′-A A **T C** T T **G A** A C-5′	1.8×10^5
III	5′-T T **A G T A C T** T G-3′ 3′-A A **T C A T G A** A C-5′	1.8×10^6



	ImPyHpPy-β-Dp	$K_{a,\,app}$ (M^{-1})
I	5′-T T **A G A T C T** T G-3′ 3′-A A **T C T A G A** A C-5′	9.3×10^4
II	5′-T T **A G A** A **C T** T G-3′ 3′-A A **T C T** T **G A** A C-5′	7.7×10^4
III	5′-T T **A G** T A **C T** T G-3′ 3′-A A **T C** A T **G A** A C-5′	1.9×10^5

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

Polyamide	5'-AGTACT-3'	5'-AGAACT-3'	5'-AGATCT-3'
ImPyPyPy-β-Dp (**1**)	2.2 (±0.4) × 10^7	2.0 (±0.1) × 10^6	2.6 (±0.2) × 10^6
ImHpPyPy-β-Dp (**2**)	1.8 (±0.3) × 10^6	1.8 (±0.1) × 10^5	1.9 (±0.6) × 10^5
ImPyHpPy-β-Dp (**3**)	1.9 (±0.2) × 10^5	7.7 (±0.8) × 10^4	9.3 (±2.0) × 10^4

[a]Values reported are the mean values obtained from at least three DNase I footprint titration experiments. The assays were carried out in 40 μL at 22 °C, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

Polyamide:DNA Cocrystal Strucutre

Overall Structure

The crystallographic details of the structure are described in the publication (*29*). The overall conformation of ImPyHpPyβDp and ImPyPyPyβDp•5'-CC**AGATCT**GG-3' are similar to those of other four-ring polyamide dimers bound to DNA (Figure 5.3). In both structures, the polyamides bind as an antiparallel dimer in the minor groove of B-form DNA, with the N- to C-terminal orientation of the polyamide parallel to the 5'- to 3'- orientation of the adjacent DNA strand. This is consistent with affinity cleaving experiments and previous structures of polyamide dimer-DNA complexes (*7-9, 11-13, 18, 19, 30-36*).

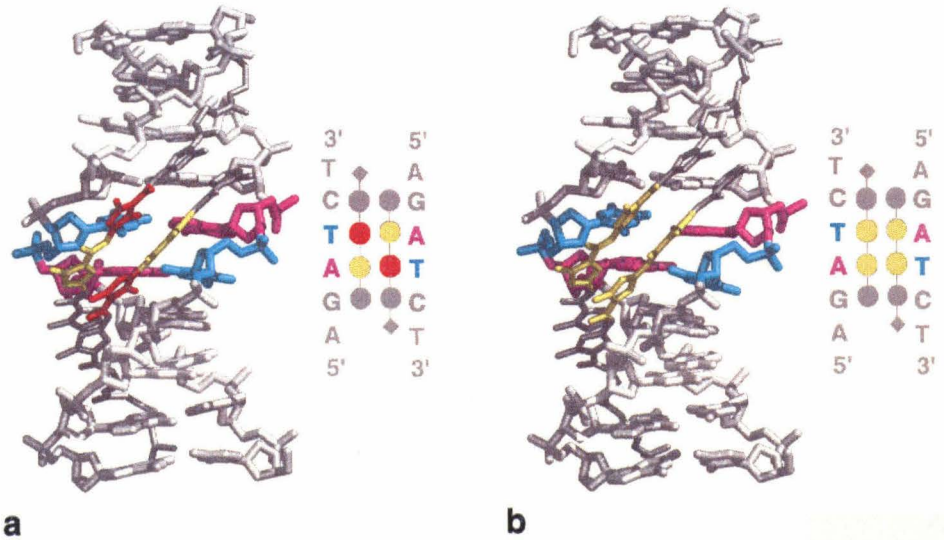


Figure 5.3. (a) Structure of $(ImPyHpPy)_2$•5'-CC**AGATCT**GG-3'. Adenosine is purple and thymidine cyan; Hp is red, Py residues paired with Hp or Py are yellow, and the residues in Im/Py pairs are grey. (b) Overall structure of $(ImPyPyPy)_2$•5'-CC**AGATCT**GG-3'. Schematic models of the bound sites are shown to one side.

DNA Helical Parameters

For the ImPyHpPyβDp structure and its pyrrole counterpart, the twist, rise per residue and sugar pucker of the bound DNA remains B-form. The ImPyPyPyβDp and ImImPyPyβDp polyamides reverse the buckle of the unliganded DNA, whereas by comparison the Hp of ImPyHpPyβDp and ImHpPyPyβDp simply flattens the base pairs.

Table 5.2. Minor Groove Width[1,3] (Å)

Base Pair:	A•T	G•C	T•A	A•T	C•G	T•A
ImPyPyPy•GTAC	7.2	7.9	7.7	7.8	8.4	7.9
ImHpPyPy•GTAC	7.7	7.7	7.4	7.4	8.7	8.5
Base Pair:	A•T	G•C	A•T	T•A	C•G	T•A
ImPyPyPy•GATC	8.2	7.6	6.8	7.2	8.0	8.2
ImPyHpPy•GATC	7.5	7.2	6.7	7.2	7.9	8.3
Average Difference, polyamide•GTAC-polyamide•GATC	-0.4	0.4	0.7	0.4	0.6	-0.1
Base Pair:	A•T	G•C	T•A	A•T	C•G	T•A
GTAC, DNA[4]	10.5	7.9	5.1	3.9	6.3	8.5
Base Pair:	A•T	G•C	A•T	T•A	C•G	T•A
GATC, DNA[2]	8.5	7.5	4.8	3.2	5.2	8.3
Average Difference, GTAC DNA-GATC DNA	2.0	0.4	0.3	0.7	1.1	0.2

[1]Calculated using the program Curves4 (Lavery & Sklenar).
[2]Leonard & Hunter, 1993.
[3]Average of minor groove width for indicated base step.
[4]Ref. 37. Average of two alternate conformations.

In addition, the Hp residue consistently induces a positive displacement between the target TA bases rather than the small negative shear observed both in the presence and absence of other polyamide residues. All of the polyamide dimers consistently increase the opening of the target base pairs toward the major groove, and induce a large negative propeller twist, regardless of the sequence recognized. The presence of the polyamide dimer widens the minor groove to ≥7.0 Å from the ~4 Å minor groove width of AT rich sequences. This difference is much more significant than the difference when comparison is made with the inherently wider minor groove of GC rich sequences (~3.0 Å versus 1.0 Å differences). The high propeller twisting of the base pairs in light of the widened minor groove is surprising, since propeller twist often accompanies a narrow minor groove in B-DNA structures (*26, 27*). This discrepancy supports the idea that many factors interact to determine DNA conformation, so that strict relationships between DNA helical parameters and sequence are difficult to define (*25-27*).

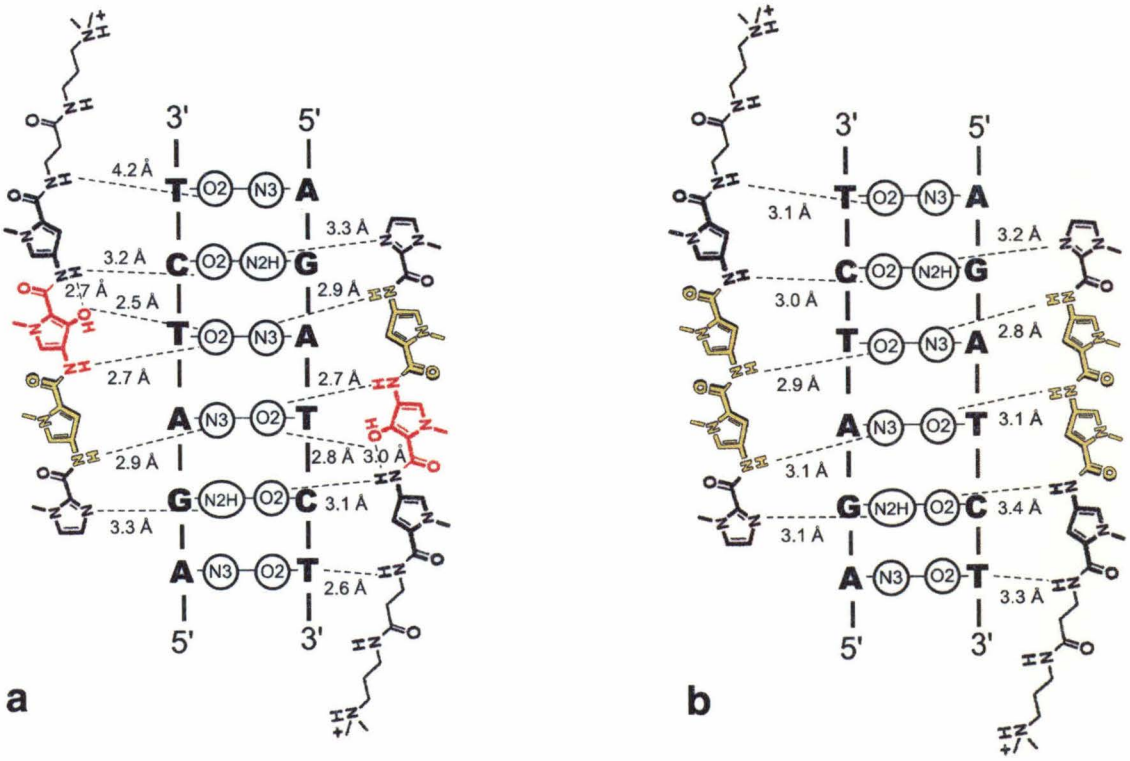


Figure 5.4. Chemical structure of the polyamides and schematic diagram of hydrogen bonding interactions with the 5'-AGATCT-3' binding site, for (a) ImHpPyPy and (b) ImPyPyPy. Hydrogen bonds are shown by dashed lines, and numbers indicate distances between non-hydrogen atoms.

Likewise, from analysis of several B-DNA structures, it was concluded that the minor groove width of G-A steps tends to be narrower than that of G-T steps (*25*). Since this sequence-dependent variation in groove width is a possible explanation for the 5'-G*T*-3' preference by polyamides (for example, 10-fold preference for 5'-G*T*AC-3' over 5'-G*A*TC-3' by ImPyPyPyPyβDp), the minor groove widths of the polyamide binding sites were compared (Table 5.2). Indeed, there appears to be a difference between the groove widths of the GATC and the GTAC sequences in the presence of both the Hp-containing polyamides and the Py-parent compounds, with the bound GTAC site ~0.5 Å wider on average. A difference is also observed between the structures of the unliganded DNA sites (*27, 37*). Since this effect is observed between all of the polyamide complexes, it is unlikely to result from variations in crystal packing. However, given the interaction between the O4' of the deoxyribose and the polyamide aromatic rings, it is uncertain whether the small differences contribute unfavorably to the overall binding affinity for GATC versus GTAC sequences.

Polyamide-DNA hydrogen bonds

The N3 and O2 atoms of the bases form hydrogen bonds with the amide nitrogens of the ImPyHpPyβDp and ImPyPyPyβDp polyamides (Figure 5.4). Reversing the TA sequence of the binding site does not affect these hydrogen bonds, nor the additional specific hydrogen bond between the hydroxyl and the thymine. The ImPyHpPyβDp polyamide recognizes the thymine-O2 in an identical manner to ImHpPyPyβDp. Between the amide nitrogen and the hydroxyl group of the Hp, both the lone pairs of the thymine-O2 accept hydrogen bonds from the polyamide, whereas the single lone pair of adenine-N3 can accept only one. Secondly, the hydroxyl oxygen contacts the C2 carbon of the bulky adenine ring, and would overlap the carbon by ~2 Å if the target thymine is modeled as a mismatch adenine. Therefore, the 20-fold difference in DNA binding affinity between the ImPyHpPyβDp and ImHpPyPyβDp polyamides cannot be due to direct perturbation of the interactions between the Hp residue and target T•A pairs.

As was observed in the ImHpPyPyβDp structure, the hydroxyl is positioned to form an intramolecular hydrogen bond with the amide of the following residue (2.9 Å distance). This forces the amide to donate two hydrogen bonds, one with the DNA and the other to the hydroxyl oxygen, thereby weakening the interaction with the DNA. Accordingly, in both Hp-containing structures, this amide-DNA hydrogen bond is slightly elongated (3.2-3.8 Å), and could contribute to the destabilization of binding affinity upon introduction of the hydroxyl group (12-fold and 28-fold for ImHpPyPyβDp and ImPyHpPyβDp, respectively).

Table 5.3. Watson-Crick Hydrogen Bond Lengths for Central A,T Base Pairs (Å)

Base Pair		ImPyHpPy	ImPyPyPy	Difference
Ade5-N1	Thy16-N3	3.0	3.2	-0.2
Ade5-N6	Thy16-O4	2.9	2.9	0.0
Thy6-N3	Ade15-N1	3.1	3.3	-0.2
Thy6-O4	Ade15-N6	2.9	3.1	-0.2

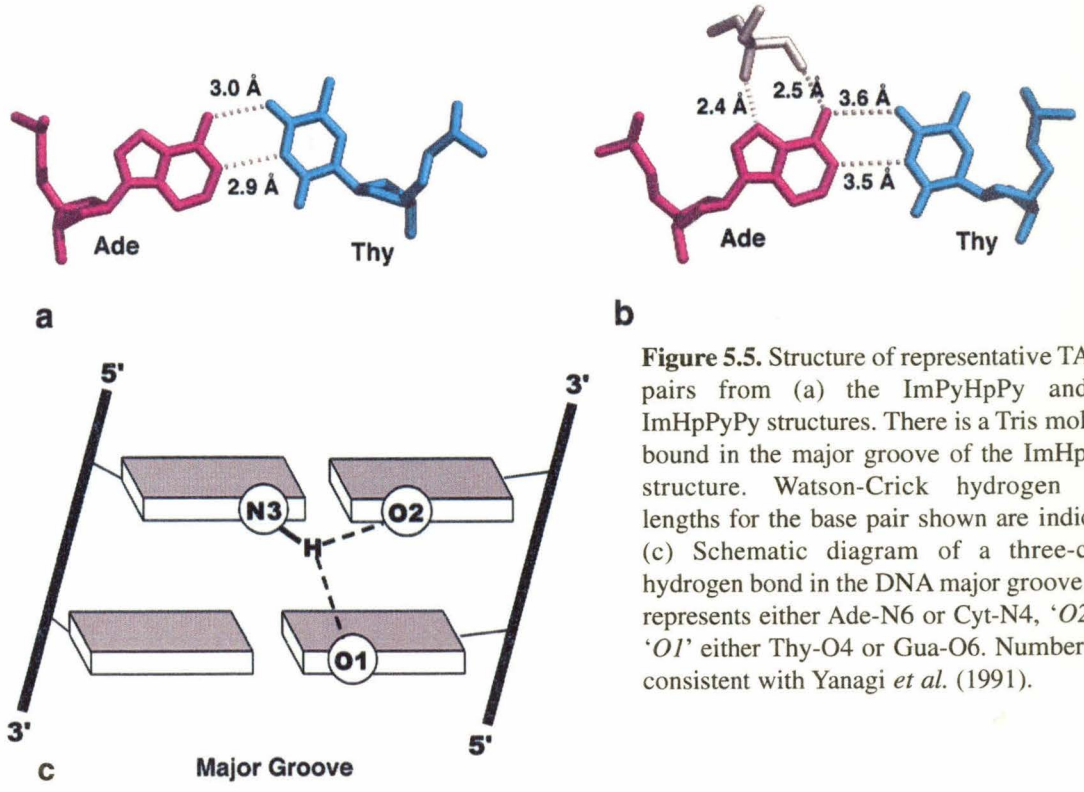


Figure 5.5. Structure of representative TA base pairs from (a) the ImPyHpPy and (b) ImHpPyPy structures. There is a Tris molecule bound in the major groove of the ImHpPyPy structure. Watson-Crick hydrogen bond lengths for the base pair shown are indicated. (c) Schematic diagram of a three-center hydrogen bond in the DNA major groove. '*N3*' represents either Ade-N6 or Cyt-N4, '*O2*' and '*O1*' either Thy-O4 or Gua-O6. Numbering is consistent with Yanagi *et al.* (1991).

Intra-duplex hydrogen bonds

In the structure of the Hp-containing polyamide ImHpPyPyβDp bound to 5'-AG*T*ACT-3', this slight decrease in binding affinity upon introduction of the hydroxyl group was attributed to lengthening of the Watson Crick hydrogen bonds of the target TA base pairs in the presence of the Hp/Py pairs. However, ImPyHpPyβDp recognizes 5'-AG*A*TCT-3' without perturbing these base pair hydrogen bonds relative to the pyrrole-counterpart (Table 5.3). Instead of resulting directly from the presence of the Hp residue,

Table 5.4. Comparison of Three-Center Interactions in the Major Groove

Distances R_{xy} are the separation between the atoms x and y, as defined in Figure 5.5, where atoms 1 and 2 are either Thy-O4 or Gua-O6, and atom 3 is either Ade-N6 or Cyt-N4. The angle A_{xy} is measured between x, hydrogen H, and y. Hydrogens were assigned using X-PLOR v3.857 (Brünger *et al.*, 1998). For structures containing more than one DNA strand in the crystallographic asymmetric unit, values are listed for both strands of the palindromic DNA site, with the second italicized and indicated by a prime.

		Distance (Å)		Angle (°)			
Structure	Base Step	R_{13}	R_{23}	A_{12}	A_{13}	A_{23}	Sum
ImPyPyPy•GATC	Gua4-Ade5	2.9	2.9	-	-	-	-
ImPyPyPy•GATC'	*Gua14-Ade15'*	2.7	2.7	-	-	-	-
ImPyHpPy•GATC	Gua4-Ade5	2.6	3.0	-	-	-	-
ImPyHpPy•GATC'	*Gua14-Ade15'*	2.9	2.9	-	-	-	-
GATC	Gua4-Ade5	3.4	2.6	-	-	-	-
GATC'	*Gua14-Ade15'*	3.1	2.7	-	-	-	-
ImPyPyPy•GTAC	Gua4-Thy5	3.1	2.6	128	91	157	376
ImPyPyPy•GTAC'	*Gua14-Thy15'*	3.3	3.0	114	95	167	376
ImHpPyPy•GTAC	Gua4-Thy5	2.8	3.6	143	87	150	370
ImHpPyPy•GTAC'	*Gua14-Thy15'*	3.2	3.2	112	80	166	358
GTAC	Gua4-Thy5	3.1	2.9	117	84	158	359
ImImPyPy•GGCC	Gua4-Gua5	2.7	2.8	133	85	157	375
ImImPyPy•GGCC'	*Gua14-Gua15'*	2.8	2.9	126	91	165	382
GGCC	Gua4-Gua5	3.1	2.9	105	87	168	360
GCGC	Gua4-Cyt5	3.1	2.9	-	-	-	-

the lengthened hydrogen bonds may have been a secondary effect of a buffer molecule, tris-(hydroxymethyl)-aminomethane (Tris), from the crystallization conditions bound in the major groove of the ImHpPyPyβDp structure (Figure 5.5(b)). Although such a buffer molecule has not been observed in any of the other polyamide-DNA complex structures, the structure of the oligonucleotide 5'-CCAGTACTGG-3', determined in the absence of polyamide, shows a hydrated calcium molecule interacting via bound water molecules to the same site (*37*). Therefore, the functional groups displayed on the edges of the bases may serve to bind small molecules in the major groove of the 5'-GT-3' step.

Cross-strand hydrogen bonds between a member of one base pair and another in the following step are conceptually possible whenever a donor amine is opposed diagonally with an acceptor oxygen across the DNA groove (Figure 5.5(c)) (*38, 39*). In the major groove, these intrastrand hydrogen bonds can occur for 5'-C-C-3', 5'-C-A-3', 5'-A-A-3' and 5'-A-C-3' sequences (and complementary strands 5'-G-G-3', 5'-G-T-3', 5'-T-T-3' and 5'-T-G-3'). Although not required for cross-strand hydrogen bond formation, base pairs with a large propeller twist, such as observed in the polyamide-DNA structures, are usually engaged in these three-center hydrogen bonds (*25, 38-40*).

The binding sites of the polyamide dimer-DNA structures were analyzed for these three-center hydrogen bonds using the criteria of Heinemann and Alings; the three angles about the hydrogen position should be ~90° or greater, and the sum of the angles ~360° to indicate coplanarity (*40*). The ImImPyPyβDp•GGCC complex, at the 5'-G-G-3' and 5'-C-C-3' steps, and ImHpPyPyβDp or ImPyPyPyβDp•GTAC at 5'-G-T-3' and 5'-A-C-3', are theoretically capable of forming cross-strand hydrogen bonds. Indeed, these anticipated interactions are observed, although borderline in some cases where the N-H---O angle is 80-87° (Table 5.4). Interestingly, a similar close approach between equivalent atoms is observed even in the absence of an appropriate hydrogen bond donor, possibly induced by the propeller twisting of the base pairs in the presence of the polyamides. Thus, for GATC sequences, the Gua-O6 and Thy-O4 are forced into a close contact that would

presumably be uncomfortable due to the electronegative quality of both atoms.

Accordingly, it is these sequences that are unable to form the intraduplex hydrogen bonds, 5'-GCG-3' and 5'-GAT-3', that are more problematic to target with high affinity. For example, the unlinked compound ImPyHpPyβDp displays 19-fold lower affinity for its match site 5'-AG*A*TCT-3' than does ImHpPyPyβDp for 5'-AG*T*ACT-3' (this study), and ImPyImPyβDp has two orders of magnitude lower affinity for 5'-TGC*G*CA-3' than does the analogous polyamide ImImPyPyβDp for its respective match site 5'-TGG*C*CA-3' (*3*). These trends in sequence preferences are shared by the hairpin counterparts that are used in practice (*3, 20, 21*). Therefore, a correlation exists between the ability of some sequence steps to form three-center hydrogen bonds in the presence of highly propeller twisted base pairs, and the DNA binding affinity of the corresponding polyamide.

Conclusions

The structures of ImPyHpPyβDp and ImPyPyPyβDp bound to 5'-CC**AG*AT*CT**GG-3' were undertaken to achieve a better understanding of TA recognition and the detailed affinities of polyamides for different TA sequences. From these structures, T•A recognition by Hp/Py follows two principles that are general in different sequence contexts; formation of specific hydrogen bonds with the thymine-O2 and shape selective recognition of the adenine-C2. The structural basis of the non-additive, slight reduction in binding affinity upon addition of the hydroxyl group is more difficult to address. Originally, it was calculated that the partial melting of the target TA hydrogen bonds in the structure of ImHpPyPyPyβDp bound to DNA would produce sufficient energetic penalty to account for the difference in affinity. However, the elongation of the TA hydrogen bonds is not observed in the ImPyHpPyβDp•5'-AGATCT-3' structure, so this cannot be a general mechanism.

Although perhaps inadequate to account for the energetic difference between Hp-polyamides and their Py- counterparts, lengthening of the amide-DNA hydrogen bond following the hydroxyl group is observed in both Hp-containing structures. One attractive feature of this as an explanation for the reduction in affinity is that it is not necessarily additive. Since optimization of the hydrogen bonds between polyamides composed of aromatic amino acids and DNA becomes a compromise as the number of aromatic residues is increased beyond five, interfering with the amide-DNA hydrogen bond following the Hp residue may have less of a consequence for longer polyamides (*11, 41*). This fits the observation that addition of multiple Hp residues in a ten-ring hairpin polyamide does not additively reduce the binding constant (*22*). Nevertheless, the structural basis of such a small energetic difference is difficult to unambiguously pinpoint without higher resolution structures coupled with further thermodynamic studies.

The basis of the higher affinity and specificity of Hp-Im-Py polyamides for 5'-G*T*-

3' versus 5'-G*A*-3' and 5'-G*GC*-3' over 5'-G*CG*-3' sequences is also difficult to ascribe to a single type of interaction. Comparison of the different polyamide dimer-DNA structures shows that favorable cross-strand hydrogen bonds form in the major groove of certain sequences in the presence of the propeller-twisted base pairs induced by the polyamide, whereas for others, an unfavorable electrostatic interaction results instead. This may contribute to the difficulty of targeting 5'-G*A*-3' and 5'-G*CG*-3' sequences, although a trend for the 5'-GATC-3' sequences to display a narrower minor groove than 5'-GTAC-3' may also relate to differences in affinity.

Problematic 5'-G*CG*-3' sequences can be efficiently targeted by introducing a flexible β-alanine residue prior to the specific imidazole (*28*). Internal β-alanines allow long sequences to be targeted with high affinity, by a mechanism that most probably involves resetting the register of the polyamide to that of the DNA (*2, 11, 13*). By analogy, disparity between the helical parameters of the polyamide and the DNA may contribute to difficulty targeting 5'-G*CG*-3'. Additionally, the β-alanine may relieve the requirement for propeller twisted base pairs, so that unfavorable interactions are not incurred for these sequences. Further assessment of this speculation awaits a high resolution structure of a polyamide-DNA complex containing β-alanine residues.

Although many potentially important interactions may be observed structurally, it is difficult to deduce which provide important contributions to the overall binding affinity (*42*). This is clearly demonstrated by the absence of partial melting of the AT base pairs in the second structure of an Hp-containing polyamide, a feature that was previously believed to be the basis of the slightly reduced affinity for these compounds relative to the Py- counterparts (*12*). Therefore, the coupling of multiple, high-resolution structures with thermodynamic analysis may provide a more reliable picture of the underlying interactions than would a single structure alone. The increasing efficiency of DNA crystallography makes the routine determination of multiple structures for this purpose an increasingly realistic goal (*43*).

Experimental

The polyamides ImPyHpPyβDp and ImPyHpPyβDp were synthesized by solid phase methods using Boc-protected 3-methoxypyrrole, imidazole and pyrrole amino acids and purified by reverse phase chromatography (*44, 45*). Identity and purity of the polyamides was confirmed by ^{1}H nuclear magnetic resonance and matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Apparent equilibrium association constants were determined by quantitative DNase I footprint titration experiments on the *Eco*RI/*Pvu*II restriction fragment of pDEH10 as described (*3*). Crystallization and X-ray crystallography were performed as described (*29*).

Protein Data Bank accession number

The coordinates have been deposited in the Nucleic Acid Database with accession numbers DD0020 (ImPyHpPy) and DD0021 (ImPyPyPy).

Acknowledgments

We are grateful to the NIH for research support, to the NSF and NIH for predoctoral fellowships to C.L.K. and R. E. B., to J. Edward Richter for an undergraduate fellowship to J.M.T., and to the HHMI for a predoctoral fellowship to E.E.B. We thank A. Chirino and M. Williamson for assistance with computational and X-ray equipment respectively, and S. Horvath for oligonucleotide synthesis.

References

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

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

3. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) *J. Am. Chem. Soc. 119*, 6953-6961.

4. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) *Chem.-Eur. J. 3*, 1600-1607.

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

6. Herman, D. M., Baird, E. E., and Dervan, P. B. (1999) *Chem.-Eur. J. 5*, 975-983.

7. Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E., and Dervan, P. B. (1992) *Proc. Natl. Acad. Sci. USA 89*, 7586-7590.

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

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

10. deClairac, R. P. L., Geierstanger, B. H., Mrksich, M., Dervan, P. B., and Wemmer, D. E. (1997) *J. Am. Chem. Soc. 119*, 7909-7916.

11. deClairac, R. P. L., Seel, C. J., Geierstanger, B. H., Mrksich, M., Baird, E. E., Dervan, P. B., and Wemmer, D. E. (1999) *J. Am. Chem. Soc. 121*, 2956-2964.

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

13. Kielkopf, C. L., Baird, E. E., Dervan, P. D., and Rees, D. C. (1998) *Nat. Struct. Biol. 5*, 104-109.

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

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

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

17. White, S., Baird, E. E., and Dervan, P. B. (1997) *Chem. Biol. 4*, 569-578.

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

19. Pelton, J. G., and Wemmer, D. E. (1990) *J. Am. Chem. Soc. 112*, 1393-1399.

20. White, S., Baird, E. E., and Dervan, P. B. (1996) *Biochemistry 35*, 12532-12537.

21. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. (1998) *Nature 391*, 468-471.

22. White, S., Turner, J. M., Szewczyk, J. W., Baird, E. E., and Dervan, P. B. (1999) *J. Am. Chem. Soc. 121*, 260-261.

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

24. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1999) *J. Am. Chem. Soc. 121*, 1113-1120.

25. Yanagi, K., Prive, G. G., and Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 201-214.

26. Yuan, H., Quintana, J., and Dickerson, R. E. (1992) *Biochemistry 31*, 8009-8021.

27. Leonard, G. A., and Hunter, W. N. (1993) *J. Mol. Biol. 234*, 198-208.

28. Turner, J. M., Swalley, S. E., Baird, E. E., and Dervan, P. B. (1998) *J. Am. Chem. Soc. 120*, 6219-6226.

29. Kielkopf, C. L., Bremer, R. E., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. (2000) *J. Mol. Biol. 295*, 557-567.

30. White, S., Baird, E. E., and Dervan, P. B. (1997) *J. Am. Chem. Soc. 119*, 8756-8765.

31. Chen, X., Ramakrishnan, B., Rao, S. T., and Sundaralingam, M. (1994) *Nat. Struct. Biol. 1*, 169-175.

32. Chen, X., Ramakrishnan, B., Rao, S. T., and Sundaralingam, M. (1995) *Nat. Struct. Biol. 2*, 733-735.

33. Chen, X., Ramakrishnan, B., and Sundaralingam, M. (1997) *J. Mol. Biol. 267*, 1157-1170.

34. Kopka, M. L., Goodsell, D. S., Han, G. W., Chiu, T. K., Lown, J. W., and Dickerson, R. E. (1997) *Structure 5*, 1033-1046.

35. Mitra, S. N., Wahl, M. C., and Sundaralingam, M. (1999) *Acta Crystallog. sect. D 55*, 602-609.

36. Yang, X. L., Kaenzig, C., Lee, M., and Wang, A. H. J. (1999) *Eur. J. Biochemistry 263*, 646-655.

37. Kielkopf, C. L., Ding, S., Kuhn, P., and Rees, D. C. (2000) *J. Mol. Biol. 296*, 787-801.

38. Calladine, C. R. (1982) *J. Mol. Biol. 161*, 343-352.

39. Dickerson, R. E. (1983) *J. Mol. Biol. 166*, 419-441.

40. Heinemann, U., and Alings, C. (1989) *J. Mol. Biol. 210*, 369-381.

41. Kelly, J. J., Baird, E. E., and Dervan, P. B. (1996) *Proc. Natl. Acad. Sci. USA 93*, 6981-6985.

42. Clackson, T., and Wells, J. A. (1995) *Science 267*, 383-386.

43. Joshua-Tor, L., and Sussman, J. L. (1993) *Curr. Opin. Struct. Biol. 3*, 323-335.

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

45. Urbach, A. R., Szewczyk, J. W., White, S., Turner, J. M., Baird, E. E., and Dervan, P. B. (1999) *J. Am. Chem. Soc. 121*, 11621-11629.