Programmable Oligomers for DNA Recognition

Thesis by

Raymond Michael Doss

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2006

(Defended 23 September 2005)

For my family.

For BSE.

Acknowledgments

Over the course of five years you quickly learn that no one can go through graduate school on their own. For this very reason, there are many people in my life I wish to thank for all the help they have given me along the way.

First and foremost, I would like to thank my family. I don't think I have met two more selfless people in my life other than my mother and father, Pierrette and Emile. My parents have sacrificed so much just so they could give their children the opportunity to have a better life than they did. Since day one, they have taught me that I can do anything I want as long as I put my head to it and have been there every second to offer me any type of support I needed. For their constant love and support, I am forever indebted. I must also thank my brother Chris for putting up with my constant antics better than anyone I know. He has always been there when I needed him and know that he will continue to do so for a very long time.

I owe much of what I have achieved as a graduate student to my advisor Peter Dervan. Many people go for years in their career looking for direction and guidance before they find it. While my path in graduate school was not a direct one to the Dervan Group, Peter took me in when I thought I could never see the program through and has been an exemplary advisor every step of the way. In the day and age where many new advisors seem to be more concerned with engagements outside of their labs and students and forget about that piece of paper they signed committing to being a young student's

mentor, it is unbelievably refreshing to know that I can count on Peter for support now and in the future. For this, I say thank you.

To my committee of Brian Stoltz, John Bercaw and Doug Rees thank you for time and advice. It's an honor to learn from the best in science. I must especially thank Brian who helped me out quite a bit my first year. Being the only student in the Hsieh-Wilson lab for a month was rough and Brian would always invite me out with his group. He even went as far as taking me out with some friends to Lucky Baldwin's for my birthday in September 2000. Although it may have seemed like a small gesture, I truly appreciated your kindness – even if you said I looked like a slacker.

It would be hard to design a lab environment more conducive to excellence in science than the Dervan Lab at Caltech. From the very first day I stepped into lab, everyone from outgoing postdocs to first year students, were more than willing to help me get acclimated to the lab. I have benefited just as much from help with running gels to countless conversations about the most inane things at 3am over beers. I need to say thank you to Adam Urbach for teaching me many lessons in and out of lab, Bobby Arora for showing me how to be a good scientist while still having fun and Bogdan Olenyuk for advice on ANY scientific subject I chose. I'll never forget coffee with Shane Foister, where many of our ideas for research came about. In a place like Caltech where it's surprisingly difficult to get people to watch Monday Night Football, Shane was always up for that, late night poker and then hitting coffee early in the AM. To those who came here my year, Eric Fechter, Adam Kerstien and Michael Marques congratulations on making it through (even if it was a few months before me).

And the future of the Dervan Lab – of Dervana. It lies on the shoulders of the current students. I must point out Ryan Stafford, Justin Cohen, Jim Puckett, Carey Hsu, Britton Boras, Nick Nichols and Michelle Farkas in particular for they have been responsible for a decent amount of the late nights not spent in lab. You guys are the future of polyamides. May God help us. In all honesty, if I had to have a group students in my lab, I would surely grab all of you up. The only thing that needs to be worked on is making sure people survive their own birthdays. I won't name names, but huddling for warmth in restroom of an eating establishment in Old Town is just poor form.

Although things were very different back then, I will never forget the time I spent downstairs in 102 Church. There I made great friends with whom I have many memories over the past five years. To the original crew of Sarah Tully, Nelly Khidekel, Nathan LV, Cristal Gama, Katherine Poulin and Sherry Tsai – good luck in the future, although I know that you will not need it. I must say that if you ever want to have a interesting year in grad school with some of the most bizarre conversations ever, join a lab with six 22 year old women. Trust me, it will change you forever.

Out of all my friends I have made through Caltech, there are two which stand out from the rest. I have known Sarah Tully since the first few weeks of grad school. From the very beginning I knew that we would be friends for a long time to come. It's hard to find someone who can deal with my odd sense of humor and pension for anger, but Sarah has somehow done it. I only hope that our friendship will be as strong in the years to come. And then there is Michael Marques – the "Enigma," the Meaty Cheesy Boy, the Mercedes Boy, the "Saga Motor Inn Hater" – the list could go on. It feels very "middle

school" to say this, but his is certainly one of my best friends. For pulling me into the Dervan Group, working back-to-back-to-back 9AM-5AM days, taking the time out to teach me all the research techniques I would ever need and knowing how to have a good time and relax, I owe him endless thanks. If there is a friend that I know will always be there for me, it is Mike. Good luck on pursuing your second "Dr." degree in medical school. If you want, I'll call you "Dr, Dr." Riiiight....

I also have to thank Neil Garg, whom I've known for nine years now. From working late summer nights at NYU in the Walters lab to hanging out reminiscing at his wedding, it has been all good times. Neil has always been there for me during some very difficult times and I hope that I have reciprocated. Neil, you have an unhealthy love for all that is 80's, but I still love you.

There are countless people outside of Caltech that I could thank, but I'll stick to those who I've truly gotten close to. Benoit Ugeux, thanks for the good times out in Palm Springs and in NYC. I am glad you have found something that you truly enjoy and know that you will be greeted with much success. There are few people as compassionate, caring, and devoted to his friends and family as you are. You are truly a model friend. And one day we'll have a truly spectacular New Year's Eve. One day... And finally, to Alex Chen. Thank you for giving me a haven outside of Caltech. I've learned so much about countless things because of you, and I truly appreciate it. It doesn't matter if I haven't seen you for a day or for several months, I know that when we hang out it is always a good time. I truly cherish your friendship.

Abstract

As the amount of information about the genetic construct of the human body continues to grow, the ability to manipulate genes via the use of synthetic molecules becomes an increasingly attractive concept. Polyamides developed in the Dervan Lab are capable of doing just this by binding in the minor groove of DNA in a highly specific manner. Not only are polyamides able to specifically target sequences of DNA, but they are able to do so at affinities which make them competitive with endogenous transcriptional machinery.

The complex nature of the DNA minor groove structure, however, has forced the evolution of traditional imidazole, pyrrole and hydroxypyrrole polyamides into newly developed oligomers – compounds which have been shown to bind sequences of DNA that have been traditionally difficult to target. In going from polyamides to oligomers, these compounds have seen a variety of changes brought about by the search for ring systems capable of conveying improved binding properties. Several new recognition elements have been uncovered and characterized with respect to their DNA affinity and specificity. Experiments testing the capabilities of these oligomers have shown that such compounds demonstrate great potential for targeting many new, biologically relevant sequences of DNA thus showing promise as potential 2^{nd} generation therapeutics.

Table of Contents

List of Figures and Tables

Chapter 1 Page

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 7

Chapter 8

Towards DNA-Binding Programmable Oligomers

Introduction.

The recognition of DNA by polyamides is a young field which has been chiefly pioneered by Dervan Group here at Caltech. Many years of research have gone into developing a set of recognition elements that can be used to sequence specifically recognize DNA via the binding of ligands in the minor groove. Inspired by the natural products netropsin and distamycin which were capable of recognizing A-T tracts of DNA, our group developed a set of recognition elements initially based on the pyrrole carboxamides found in the original natural products. Several years later, a complete code matching a set of three rings systems, pyrrole (**Py**), imidazole (**Im**), and hydroxypyrrole (**Hp**), to DNA base pairs was developed. These "pairing rules" allowed for the facile design of polyamides that were able to bind their match sites with high levels of affinity and specificity.

At this point, the polyamide program had become sophisticated enough that research in several emerging fields was possible. Much effort was put into designing polyamide-based compounds capable of acting as transcriptional activators and repressors. In addition, the high levels of polyamide specificity allowed for their use as DNA-based molecular tools. As experiments became increasingly sophisticated, however, it became obvious that a single polyamide code based on three recognition elements would not be sufficient for the generation of compounds capable of binding all sequences of DNA. This is due, in large part, to the fact that the microstructure of DNA varied significantly depending on its sequence.

At this point, a new and exciting portion of the molecular recognition program in the Dervan Group emerged. What began as an endeavor to design new 5-membered ring

systems that possessed interesting and useful DNA recognition properties quickly moved to uncovering entirely new rings systems capable of conferring exciting levels of affinity and specificity. Our acceptance of the fact that all DNA is not alike, forced us to develop a new set of DNA-binding ligands which were no longer based on the original **Py**, **Im**, and **Hp** carboxamides. Over the past 2-3 years, efforts on designing these new "oligomers" have yielded compounds capable of binding to DNA sequences that were considered to be out of the targetable range of traditional polyamides.

My research, as summarized in this thesis, represents a portion of the work that was put into evolving our ligands from traditional polyamides to programmable oligomers. Such work represents only the first step in designing compounds capable of binding thousands of different DNA sequences mired in hundreds of different types of DNA microstructures. With the groundwork laid down, these oligomers will continue to evolve; and as they do so will be better able to target more and more biologically relevant DNA sequences.

Chapter 1

An Overview of DNA-Binding Polyamides

The text of this chapter was taken in part from a manuscript coauthored with Michael A. Marques and Professor Peter B. Dervan (Caltech)

(Dervan, P. B.; Doss, R. M. and Marques, M. A. "*Programmable DNA Binding Oligomers for Control of Transcription*" *Current Medicinal Chemistry: Anti-Cancer Agents* **2005**, 5(4), 373-387.)

Abstract.

Mapping and sequencing the genetic blueprint in human, mice, yeast and other model organisms has created challenges and opportunities for chemistry, biology and human medicine. An understanding of the function of each of the ~ 25,000 genes in humans, and the biological circuitry that controls these genes will be driven in part by new technologies from the world of chemistry. Many cellular events that lead to cancer and the progression of human disease represent aberrant gene expression. Small molecules that can be programmed to mimic transcription factors and bind a large repertoire of DNA sequences in the human genome would be useful tools in biology and potentially in human medicine. Polyamides are synthetic oligomers programmed to read the DNA double helix. They are cell permeable, bind chromatin and have been shown to downregulate endogenous genes in cell culture.

1.1 Introduction.

With the completion of the Human Genome Project, biology can be viewed as an informational science. The digital information in the genome (DNA) encodes the logic of life. These genes encode protein networks. Environmental cues from outside the genome (such as transcription factors) control these networks. A major goal for chemistry, biology and human medicine would be to ask the question whether human disease could be controlled by targeting gene expression, i.e., manipulate with small molecules the information and software programs encoded in nucleic acids that control protein networks. There are two general approaches; molecules that target mRNA (antisense, RNAi, PNA) or molecules that target the gene DNA, and inhibit mRNA synthesis by interfering with components of the RNA polymerase transcription machinery. For these approaches to be successful *in vivo*, the gene regulatory agent must be cell-permeable. In the case of DNA binding molecules, the regulatory molecule must traffic to the nucleus and bind its target sequence with high affinity and specificity in the context of cellular chromatin. DNA binding must interfere with transcription of the target gene by inhibition of key transcription factors in the promoter or alternatively a steric blockade in the coding region. Nucleic acid based approaches that target either DNA or RNA (antisense, triple helix forming oligonucleotides, ribozymes and siRNA) and engineered zinc fingers have the potential for sequence specificity and can effectively inhibit transcription or translation *in vitro*: however, oligonucleotides and proteins suffer from poor cell permeability and delivery strategies, such as viral vectors, must be used for effective therapeutic outcomes in animal studies.

Small molecule approaches for gene regulation could bypass the need for delivery strategies. A number of natural and synthetic DNA binding molecules have been explored for their ability to regulate gene expression *in vitro* and in cell culture. Our laboratory has explored the development of programmable oligomers for targeting double strand DNA with affinity and specificity comparable to transcription factors.[1, 2] These small molecules achieve affinities and specificities of DNA binding proteins, inhibit a broad range of transcription factors, are cell permeable, bind to chromatin and have been shown to downregulate endogenous gene expression in cell culture

Pairing Rules.

The original inspiration for programmable DNA binding oligomers is drawn from netropsin and distamycin A which are comprised of two and three aromatic N-methylpyrrole (**Py**) rings.[1] These crescent-shaped natural products bind A,T tracks with both 1:1 and 2:1 ligand:DNA stoichiometries.[2] Cofacial pairs of unsymmetrical heterocycles distinguish the edges of each Watson-Crick base pair in the minor groove.[1] DNA binding polyamides containing Py, Im and Hp amino acids form the basis of a modular code to control sequence specificity in a predictable way (**Figure 1.1**).[3] Pairs of pyrrole (**Py**), imidazole (**Im**), and hydroxypyrrole (**Hp**) rings distinguish the four Watson-Crick base pairs. Im/Py and Py/Im distinguishes G•C from C•G, Hp/Py distinguishes T•A from A•T. Each polyamide strand is usually oriented N→C with respect to the 5'→3' direction of the DNA helix.[4] The β-dimethylaminopropylamine (β-Dp) tail (from the method of synthesis) has a DNA sequence preference for A,T base pairs for steric reasons.[3]

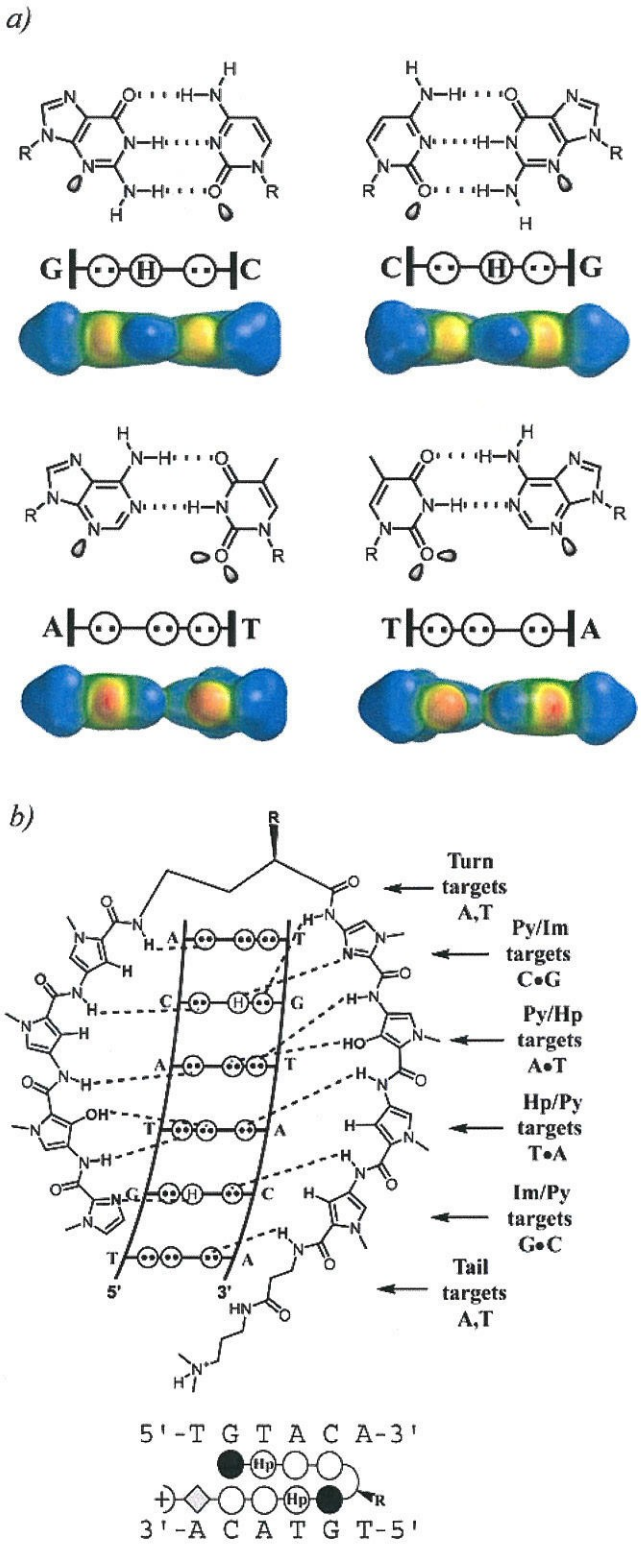


Figure 1.1 (a) Chemical structure of Watson-Crick base pairs. Circle with two dots indicates a hydrogen bond acceptor while a circle with an **H** indicates a hydrogen bond donor. Electronic surfaces of the base pair edges shown with red and blue indicating negative and positive electronic charges respectively. (b) Chemical structure of a polyamide with the putative hydrogen bonds to the DNA minor groove is shown. Pairing rules for the five-membered heterocycles shown to the right.

Affinity and Specificity.

Increasing polyamide affinity and specificity was accomplished by linking antiparallel polyamide dimers with a short alkyl chain γ to afford a single linear oligomer which folds into a hairpin structure in the minor groove of DNA.[5, 6] The turn unit enforces unambiguous ring pairing, eliminating slipped binding modes.[7, 8] Incorporation of a chiral amine at the α position of the turn unit increases affinity and disfavors against reverse binding.[9, 10] Acylation of the chiral amine is effective at further increasing sequence specificity and prevents hairpin oligomers from unfolding and binding the minor groove in a linear fashion.[11] The alphatic turn unit has a sequence preference for A,T base pairs due to an unfavorable steric clash with the exocyclic NH_2 of G,C base pairs.[11] Polyamide composition with respect to ring-number and type of heterocycles influences the affinity. Six-ring hairpin polyamides have equilibrium association constants on the order of Ka = 10^7-10^8 M^{-1} while for eight- and ten- ring hairpins, Ka = 10^9-10^{10} M^{-1}.[12] Despite the gain in specificity, polyamides with multiple **Hp/Py** pairs exhibit reduced affinity.[13] The reduction in affinity likely results from modest distortion of the DNA upon binding as well as differences in solvation in water vs. the minor groove of DNA.[14, 15]

Sequence dependent variations in the DNA microstructure play a role in the energetics of binding. A number of DNA sequences, such as purine tracts have emerged as "lower affinity" DNA targets. The relative rigidity of purine-purine steps, associated narrow minor groove, and negative propeller twist are thought to be governed by optimal base stacking interactions.[16] Variation in DNA groove width, curvature, bendability, hydration, or relative position of hydrogen bond donors/acceptors all influence the

DNA's ability to accommodate these shape selective ligands, and make it difficult to quantify incrementally the source of the affinity variations.[17, 18]

Sequence specific recognition of the DNA minor groove arises from the pairing of two different 5-membered heterocyclic amino acids and the interplay of a variety of direct and indirect recognition elements. The overall shape of the folded hairpin fits the shape (width, curvature, depth) of the DNA minor groove.[17, 18] The information face on the inside of the crescent shaped oligomer may be programmed by the incremental change of atoms on the corners of the ring pairs presented to the DNA minor groove floor. [17, 18] The corner of the ring pairs read in a digital way (not unlike Braille) each of the different edges (bumps and holes) of the four Watson-Crick base pairs.[19, 20] Stabilizing and destabilizing interactions with the distinct edges of the four Watson-Crick bases are modulated by shape complimentarity and specific hydrogen bonds (**Figure 1.2**). More specifically, the imidazole ring **Im**, which presents a lone pair of electrons via its N(3) to the DNA minor groove, can accept a hydrogen bond from the exocyclic amine of guanine.[19, 20] The 3-hydroxypyrrole ring in the **Hp/Py** pair projects an exocyclic OH group toward one side of the minor groove floor that is sterically accommodated in the asymmetric cleft of the T·A base pair, preferring to lie opposite T not A.[19, 20 21, 22] From x-ray structural analysis, it appears that **Hp** can hydrogen bond with the O(2) of thymine

Figure 1.2. Recognition of DNA by polyamides is also affected by a series of ligand-ligand or ligand-DNA interactions that take place away from the polyamide information face. Polyamide geometry with respect to overall curvature, ligand pre-organization, and Van der Walls contacts between the polyamide and the DNA minor groove walls are factors important in determining polyamide specificity.[23, 24]

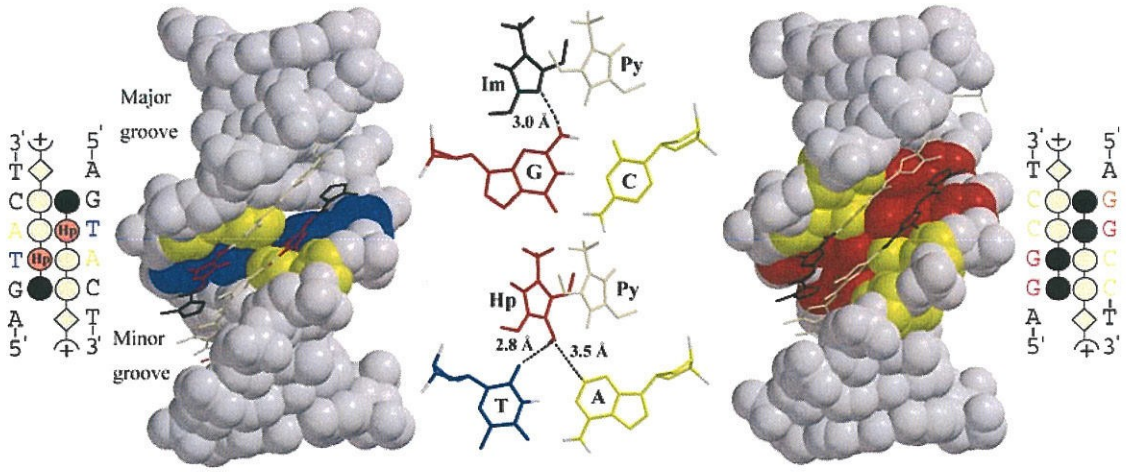


Figure 1.2 Crystal structures demonstrating the origin of G,C and T,A specificity. (a left) Ball and stick model of polyamide homodimer bound to 5'-GTAC-3' core sequence. (a center) Crystal structure of the Im-Hp-Py-Py-β-Dp homodimer bound in a 2:1 complex with the DNA. Adenine and thymine are colored yellow and blue respectively. Pyrrole (**Py**) and imidazole (**Im**) heterocycles are colored tan and black respectively. Hydroxypyrrole is colored red. (a right bottom) Structural basis for A-T vs. T-A discrimination by the hydroxypyrrole/pyrrole (**Hp/Py**) pair. (b right) Ball and stick model of polyamide homodimer bound to 5'-GGCC-3' core sequence. (b center) Crystal structure of the Im-Im-Py-Py-β-Dp homodimer bound in a 2:1 complex with DNA. Guanine and cytosine are colored red and yellow respectively. Pyrrole (Py) and imidazole (Im) heterocycles are colored tan and black respectively. (b left top) Structural basis for G-C vs. C-G discrimination by the imidazole/pyrrole (**Im/Py**) pair.

Binding Site Size.

Due to the large size of the human genome, short DNA sequences 6-8 bp in size would be expected to occur millions of times in gigabase size DNA. Thus, it is useful to maximize the size of the targeted binding site by creating oligomers that are capable of specifically recognizing long stretches of DNA. For example, an eight-ring hairpin binds six bp of DNA. It is intuitively obvious that longer oligomers should bind larger DNA sites. It has been shown that the polyamide ligand does not match the curvature of the DNA helix after five contiguous rings.[25] The polyamide is overcurved with respect to the DNA. One solution is to incorporate the flexible β-alanine linker as a discrete unit in the hairpin.[19, 20] The **β/Py** or **β/Im** pairings function similarly to the **Py/Py** and **Py/Im** pairings respectively, with the aliphatic C-H of the **β** residue targeting A, T, or C due to

steric occlusion of guanine's exocyclic amine. Unlike the rigid contiguous ring system, β-alanine provides more conformational freedom and allows relaxation of polyamide curvature (**Figure 1.3**). Subsequently, the polyamide is able to re-register itself and track the DNA helix.[26, 27] β-alanines dual functionality as an effective recognition element and as a flexible extension proves useful in the design of polyamides capable of binding extended 11 to 16 base pair sites.[28]

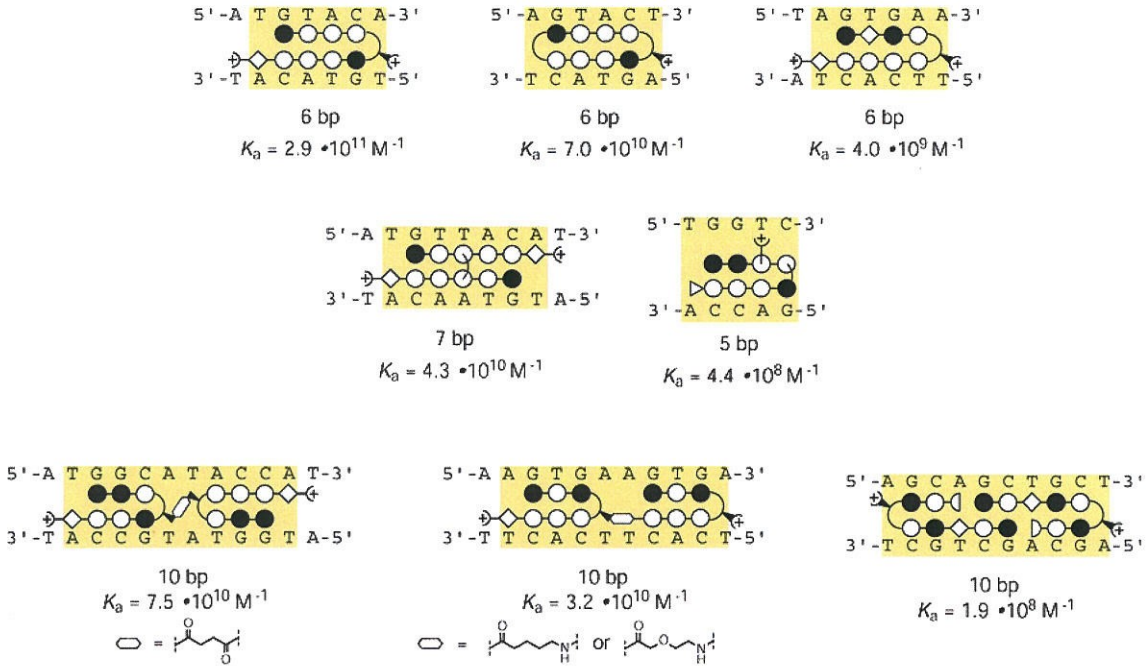


Figure 1.3 Selected polyamide binding motifs and their corresponding binding site size are shown. From left to right and top to bottom: hairpin, cycle, β-substituted, H-pin, U-Pin, turn-to-turn tandem, turn-to-tail tandem, and candy cane overlay.

The incorporation of β-alanine into hairpin polyamides allows for greater conformational freedom. Sequences such as 5'-GNG-3' that were originally targeted using rigid contiguous ring systems Im-**X**-Im (**X** = Im, Py) at low affinity (10^8 M^{-1}) can now be coded for at high affinity (10^{10} M^{-1}) by incorporation of a flexible **β** linker (Im-**β**-Im). The aliphatic linker relaxes the overall oligomer curvature and allows the

alternating imidazole rings to conform to the DNA minor groove microenvironment, maximizing hydrogen bonding with the exocyclic amine of guanine.[26, 29]

Polyamide Binding Motifs.

In addition to the hairpin motif, a series of new oligomer shapes were constructed. They include cycles, extended hairpins, H-pins, U-pins, and hairpin dimers **Figure 1.3**.[26, 29] Cycles are an attractive motif due to their symmetry and pre-organization but are difficult to synthesize. H-pins and U-pins are unique in their connectivity.[30, 31] Instead of linking polyamide strands at the turn position, H-pins and U-pins link the strands via an aliphatic bridge extending from the N-methyl substituent of paired rings. Unlike the γ turn in hairpins, the "bridging" linker is not sequence specific because it does not lie in the minor groove (**Figure 1.3**).

Transcription factors often bind as homo and heterodimers to contiguous DNA sites each 6 bp in size often in a highly cooperative manner. Inspired by biology, we could consider coupling hairpin modules for longer binding site size recognition.[32, 33] Hairpin dimers are linked turn-to-turn or turn-to-tail, occupying a larger 10 base pair binding site and bind DNA with high affinity ($Ka = 10^{10}$-10^{11} M^{-1}). The benefit of the dimer motifs is the targeting of larger binding sites in comparison to the standard hairpins. There are sequence restrictions. The alkyl linkers that join the two hairpins in the tandem formation specify for A,T base pairs for steric reasons. A concern regarding the dimer motifs is the overall oligomer size which results in poor cell uptake.

Alternative Heterocycles for DNA Minor Groove Recognition.

The pairing rules have proven useful for the recognition of hundreds of DNA sequences by designed polyamides. In addition, as this chemistry moves from cell culture to small animal models, one could imagine that a *library of heterocycles* for DNA recognition would be important regarding absorption, distribution, metabolism, toxicity and pharmacokinetics. A search was initiated for other novel heterocycle recognition elements that offer improved affinity/specificity and be chemically robust relative to the original Py analogs. Initial efforts were directed towards developing new pairs within the context of 5-membered heterocyclic amino acid pairs.[34, 35] This work is described in detail in future chapters.

N-Terminus Pairings.

The limited success at developing novel 5-membered heterocycle recognition elements at the internal position of hairpin polyamides prompted the search for new rings which could impart specificity at the N-terminus (cap) position. Earlier studies had shown that both **Im/Py** and **3-Pz/Py** pairs were capable of selecting for G > C at the cap position with good selectivity and affinity.[36] While **Im/Py** pairings show comparable specificity for G•C at both *terminal* and *internal* positions, and a 3-**Pz/Py** (pyrazole) shows near 100-fold specificity for G > C at the N-terminal position, N-terminal pairings capable of binding T•A with affinity and specificity comparable to those of the G•C specific residues had not been established.[37] Knowing that the specificity of cofacial aromatic amino acid pairings depend on their context (position) within a given hairpin

polyamide, a library of heterocyclic carboxylic acids was screened for favorable recognition properties.

Due to the absence of a second "groove-anchoring" carboxamide, rings at the cap position can adopt different conformations and are thought to be allowed to bind DNA in either of two rotamers. For example, a terminal **Hp** residue can in exists in two rotamer forms where one orients the hydroxyl group into the minor groove while a second orients the ring with the hydroxyl recognition element oriented away from the floor of the minor groove. This second orientation could be stabilized by intramolecular hydrogen bonding between the C3-OH and the carbonyl oxygen of the 2-carboxamide. For terminal 2-hydroxybenzamide residues, some measure of T•A selectivity can be recovered by creating steric bulk at the 6-position to force the hydroxyl recognition element into the groove.[37]

The fidelity of minor groove recognition by N-terminal **Im/Py** pairs suggests that the rotamer which projects N(3) in the groove is the preferred orientation in hairpin polyamides. This observation can be rationalized by a combination of both stabilizing and destabilizing forces which favors the rotamer with N(3) in the groove and N-methyl out. Rotation of a terminal **Im** residue in the opposite conformer, orienting N(3) away from the minor groove, would create unfavorable lone pair interactions with the proximal carboxamide oxygen, disrupt a favorable hydrogen bond with the exocyclic amine of G, and project an N-methyl group to the DNA floor which is sterically unfavorable.

Using the modest 3-fold specificity of the **Tn/Py** pair for T•A as a starting point, a library of thiophene rings derivatized at the 3-position was designed to impart a shape selective mode of recognition for thymine. From a series of thiophene caps including 3-

H, -CH_3, -NH_2, -NHAc, -OH, -OCH_3, -F and –Cl, it was found that N-terminal 3-methoxy (**Mt**) and 3-chloro (**Ct**) thiophene-2-carboxamide residues, when paired with **Py**, demonstrate selectivity for T•A versus A•T (**Figure 1.4**).[38] Three and four ring polyamides containing a variety of heterocycles at the N- and C-terminus have demonstrated good *in vitro* potency against Gram-positive bacteria.[39]

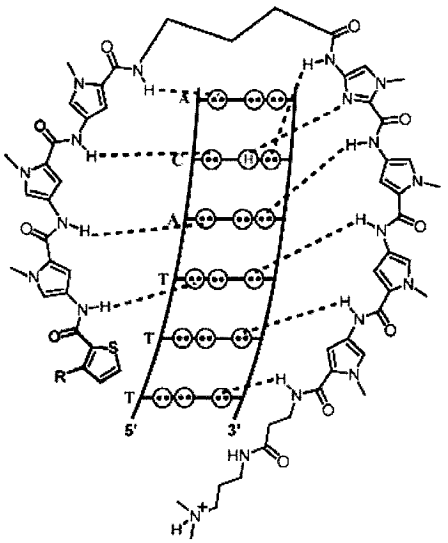


Figure 1.4 Proposed binding models for hairpin polyamides with 5'-T**X**TACA-3' site. A circle enclosing two dots represents lone pairs of N3 of purines and O2 of pyrimidines. A circle containing an H represents the exocyclic amine of guanine. Putative hydrogen bonds are indicated by dashed lines. N-terminal residue drawn in "sulfur down" syn conformation.

Fused Bicycles – Benzimidazole Analogues.

A movement away from the classic 5-membered heterocyclic carboxamides led to the incorporation of benzimidazole analogues as effective DNA recognition tools.[40, 41] The benzimidazole 6-5 bicyclic-ring structure, while having different curvature from the 5-membered heterocyclic carboxamides, presents an "inside edge" with a similar readout and shape to the DNA minor groove floor, effectively mimicking **Py**, **Im**, and **Hp** heterocycles (**Figure 1.5**). A series of experiments showed that the benzimidazole (**Bi**), imidazopyridine (**Ip**) and hydroxybenzimidazole (**Hz**) analogues, when placed into eight

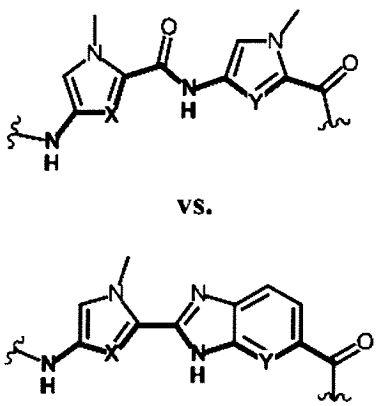


Figure 1.5 Structures of the fused 6-5 bicyclic benzimidazole building blocks in comparison with their respective five-membered ring systems. Hydrogen bonding surfaces presented to the DNA-minor-groove are bolded.

ring hairpins as **Bi/Py**, **Ip/Py** and **Hz/Py** pairs, are as effective at recognizing the DNA minor groove as their 5-membered counterparts (**Figure 1.6**).[42, 43]

Of particular interest is the **Hz/Py** pairing. Designed to be a hydroxypyrrole mimic, the **Hz/Py** pair places the same direct readout functionality to the floor of the DNA-minor groove as the **Hp/Py** pair. Like the **Hp/Py** pair, the **Hz/Py** pair is capable of effectively discriminating between A,T Watson-Crick base pairs such that **Hz/Py** codes for T·A and **Py/Hz** codes for A·T. The **Hz/Py** pair demonstrates an increase in binding affinity for its match sites and is

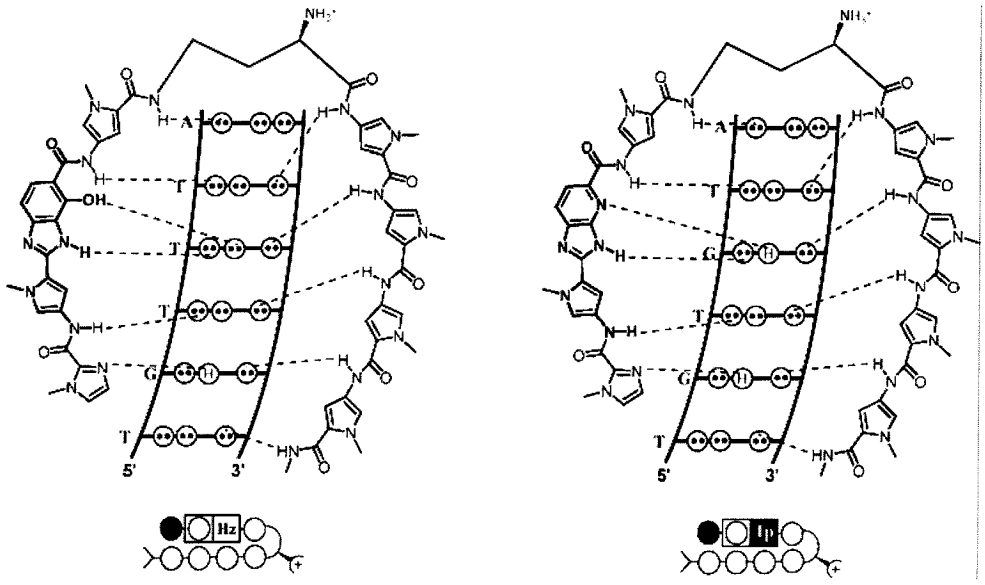


Figure 1.6 Proposed binding models for hairpin polyamides containing a hydoxybenzimidazole (**Hz**) ring system (Left) and a imidazopyridine (**Ip**) ring system (Right). A circle enclosing two dots represents lone pairs of N3 of purines and O2 of pyrimidines. A circle containing an H represents the exocyclic amine of guanine. Putative hydrogen bonds are indicated by dashed lines.

comparable in specificity to the **Hp/Py** pairing. The **Hz/Py** pair also discriminates more effectively against G,C base pairs than **Hp/Py**. While the "recognition edges" of **Hz** and **Hp** are the same, there are significant differences in overall ligand geometry and electronics. As is the case with all of the 6-5 ring systems **Bi**, **Ip** and **Hz**, a higher degree of rigidity and pre-organization of the fused hydroxybenzimidazole structure, coupled with a lower degree in curvature that may be more complimentary to the inherent curvature of the DNA helix, likely play roles in the increased affinity and specificity. Further, the benzimidazole moiety has a greater aromatic surface and hydrophobicity that may alter both the DNA-ligand van der Waals interactions, and the inter-strand π-stacking. Thus, by going from the 5-membered heterocyclic system to the fused 6-5 system, changes associated with the indirect readout of the DNA-minor groove may be responsible for the evident changes in recognition.

β-Linked Polyamides: A Special Case for DNA Recognition of $(GAA)_n$ Tracts.

Although the energetics and structure of the 2:1 complex has been explored extensively, there is less understood about 1:1 polyamide recognition beyond the initial studies on netropsin and distamycin.[44] Laemmli and co-workers reported that certain β-linked **Py/Im** polyamides bind GAGAA tracks in a 1:1 stoichiometry with a single orientation.[45] Previously, purine tracts have been difficult sequences to target using 2:1 binding hairpin polyamides. The 1:1 complex is important for expanding the sequence repertoire for DNA targeting, but the fact that β-linked **Py/Im** polyamides can bind *both* 1:1 and 2:1 in the minor groove raises important design issues for the field.[46] In an effort to further characterize the 1:1 mode of binding, specificity studies were conducted to

determine if β-linked polyamides in 1:1 stoichiometry complexes can discriminate any of the four Watson-Crick base pairs.[47] For 1:1 recognition we find that **Py** and **β** target A,T > G,C and **Im** targets G,C > A,T.

The structure of β-linked polyamides binding GAA purine tracts was studied by high resolution NMR (**Figure 1.7**).[48] The complex reveals B-form DNA with a narrow

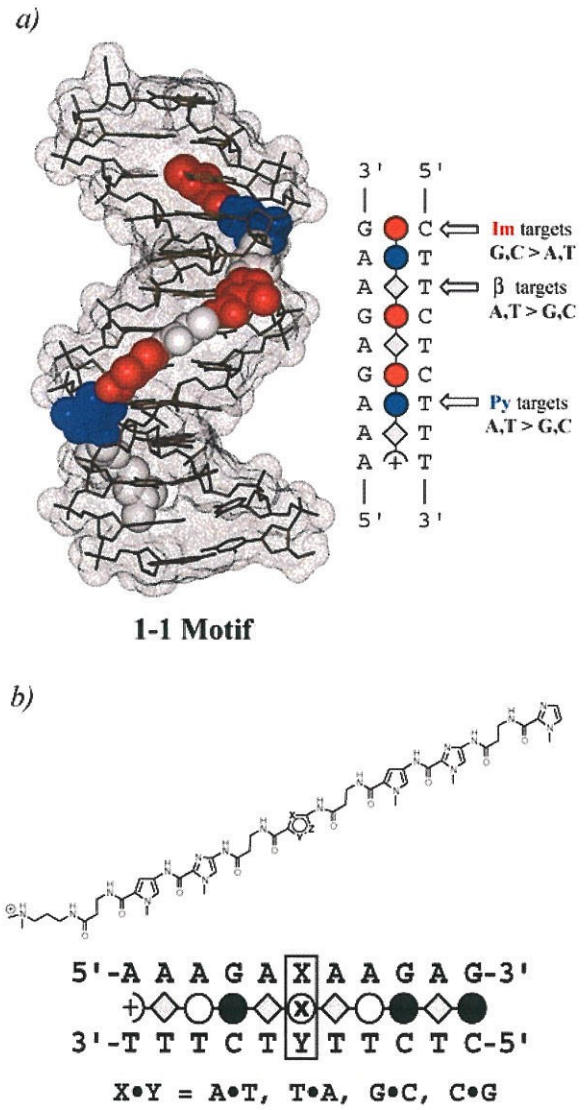


Figure 1.7 1:1 Polyamide:DNA motif. (a) (Left) High resolution 1:1 polyamide:DNA complex determined by NMR. (a) (Right) Ball and stick model of polyamide:DNA complex. Imidazole (**Im**), Pyrrole (**Py**), and β-alanine (**β**) are colored red, blue, and white respectively. (b) Chemical structure of the polyamide system used for examining the recognition profile of novel 5-membered heterocycles in a 1:1 polyamide:DNA complex.

minor groove and a large degree of negative propeller twist. Stabilization of the negative propeller twist by bifurcated hydrogen bonds donated from each polyamide NH group to proximal purine N(3) and pyrimidine O(2) atoms, in addition to the inherently rigid and narrow minor groove, is thought to be the reason polyamides bind 1:1 in polypurine sequences. Second, there is a G/C dependent orientation such that the polyamide is oriented N-C with respect to the 3'-5' direction of the guanine-containing strand. Finally, the ensemble reveals specific hydrogen bonds between Im-N(3) and G-NH2 that could only be made due to the flexibility imparted by the β residue.

Synthetic Methods.

The synthesis of polyamides is traditionally accomplished by the step-wise addition of Boc-protected amino acid monomers and dimers to either the Boc-β-Ala PAM resin or the Kaiser oxime resin (**Figure 1.8**).[49] The benzyl ester linkage, which binds the growing oligomer to the PAM resin, has been found to be stable to the established polyamide coupling and Boc-deprotection conditions. Monomers and dimers are coupled on to a growing polyamide chain by deprotection of the solid-phase bound amine with TFA, followed by addition of an activated ring with either DCC/HOBt or

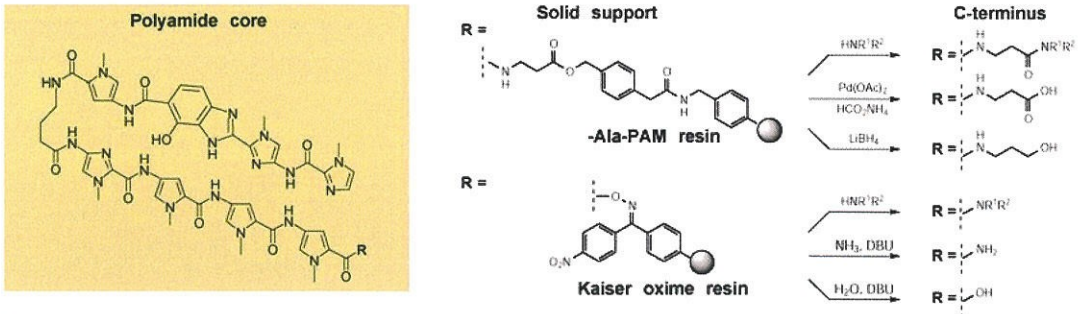


Figure 1.8 (Left) Oligomer synthesized using solid phase methodology. (Center) Resin systems used for polyamide synthesis. (Right) Cleavage conditions for modulating the chemical functionality at the C-terminus of a polyamide.

HBTU. Deprotection and addition steps are repeated in a step-wise fashion until the oligomer synthesis is complete, at which point the final product is cleaved off the resin by a nucleophile of choice.

Though use of the PAM resin allows for rapid preparation of a range of polyamides, it installs a T,A-selective β-alanine residue at the C-terminus, which places limits on the DNA sites that can be targeted. To address this problem, the Kaiser oxime resin was adapted to polyamide synthesis, allowing the preparation of polyamides with incrementally shorter C-termini.[50] Tails as shorts as methyl amide can be obtained. Molecules lacking the C-terminal β-alanine residue display the desired tolerance for G,C bases while maintaining high affinities. Moreover, removing the β residue has proven to be instrumental in nuclear uptake of polyamides in live cells. One caveat of the oxime resin is the significantly weaker aromatic oxime linker when compared to the beta-alanine benzyl ester used in PAM resin. High temperature monomer couplings needed to couple less reactive monomers – such as the less nucleophilic **Im** amine – and high concentrations of TFA for Boc deprotections tend to cleave the oligomer/resin bond, resulting in poor overall yields. In order to circumvent these shortcomings, lower concentrations of TFA are used for the Boc deprotection of solid-phase bound amines. Also, solution-prepared dimers and trimers are used to ensure that activated monomers and dimers are coupled only to reactive resin-bound amines.

Pessi and co-workers have recently used a sulfonamide-based safety-catch resin to prepare derivatives of hairpin polyamides. Unlike other resins that have been employed, activation of the safety-catch resin linker is necessary before nucleophilic cleavage of the compound can be accomplished. Pessi demonstrated that resin-bound polyamides were

readily cleaved with stoichiometric quantities of nucleophile providing a practical route to thioesters or polyamide-peptide conjugates.[51]

Although Boc-peptide chemistry is generally employed in solid phase polyamide synthesis, it should also be noted that that Fmoc chemistry has been effectively used with suitably protected monomers and Fmoc-β-Ala-Wang resin.[52]

Cell and Nuclear Uptake.

While DNA-binding polyamides have been shown to inhibit and influence a wide variety of protein–DNA interactions in solution, in order to see similar effects in cell culture, access to the nucleus is critical. Nuclear uptake of hairpin polyamides has proven to be dependent on cell type. In order to visualize the localization of polyamides in live cells, a series of fluorescently labeled polyamides have been prepared to analyze the intracellular distribution of these molecules in a panel of cell lines. It has been shown that polyamide–Bodipy conjugates stain the nuclei of T lymphocytes, a cell type which has shown robust responses to polyamides *in vivo*, but no other cell type tested.[53] In fact, polyamide–Bodipy conjugates most commonly produce a punctuate cytoplasmic staining pattern with no appreciable levels of nuclear staining. Other studies have shown that polyamide–fluorescein conjugates can uniformly exhibit favorable nuclear uptake properties in several human cancer cell lines.[54] Presence of a C-terminal β-alanine residue, a feature of polyamides synthesized on the PAM resin, seems to be a negative determinant for nuclear uptake. Polyamide–FITC conjugates, which only differ in whether or not they contain a β-alanine at the C-terminus, tend to exhibit different uptake properties with des-β compounds showing improved uptake properties.[54] It has also been

observed that the **Im** content and location within a polyamide can affect the level of nuclear localization. A clear trend, however, has yet to be determined. What has been concretely established is that manipulation of the linking residues, ring content and choice of dye, can generate compounds that display nuclear localization of polyamide-dye conjugates across a broad range of mammalian cells.[54]

Modulation of Gene Expression.

Diversity of cellular function within an organism is not dependent upon the genomic information contained within the cells but instead the choice of which genes, and at what frequency, they are transcribed. With the sequencing of the human genome complete, the stage is set to explore how manipulation of individual genes determines cell fate. In general, small molecule antigene therapeutics such as polyamides can be grouped into two different categories based on their function: transcriptional repression or transcriptional activation. Repression involves downregulation of a gene by inhibiting the assembly of the necessary transcriptional machinery on the DNA. Small molecules could inhibit repressor protein resulting in gene upregulation by derepression. Gene activation involves designing molecules with two separable modular domains. DNA binding oligomers with an activation domain (AD) recruit transcriptional machinery to the promoter of specific genes (**Figures 1.9 and 1.10**).

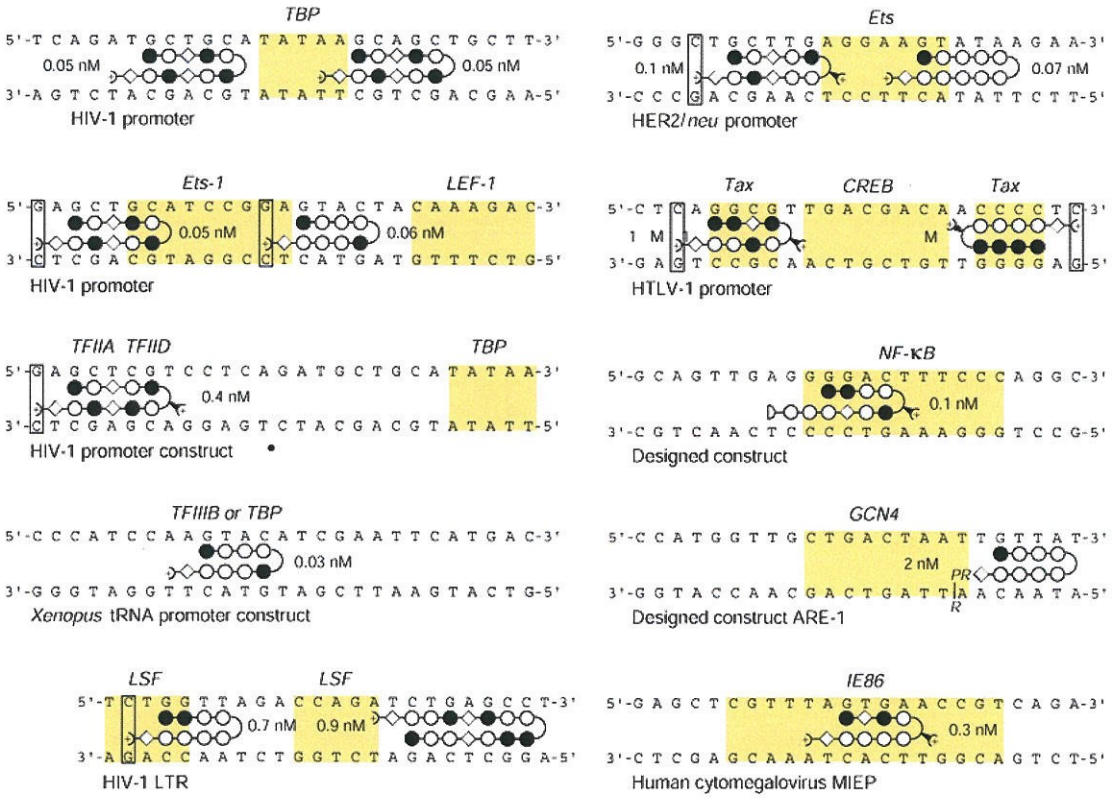


Figure 1.9 Binding sites for an array of important transcriptional regulatory proteins and the polyamides that modulate their binding are depicted.

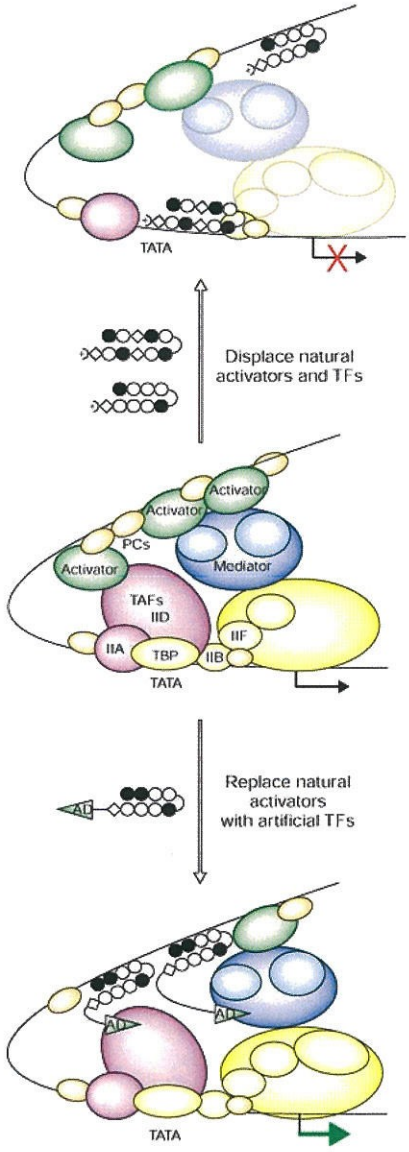


Figure 1.10 (Center) Assembly of transcription factor machinery prior to transcription is shown. (Top) Addition of polyamides that bind specific sequences in at the promoter site can disrupt the assembly of transcriptional machinery and block transcription. (Bottom) polyamides attached with small molecules that are specific for the recruitment of activator or mediator complexes can function as an artificial transcription factor (ATF) by promoting transcription.

Future Directions.

Inspired by the architecture of the natural products netropsin and distamycin A, a new class of programmable sequence-specific DNA-binding oligomers have been invented. Whether these synthetic molecules will allow external control of endogenous gene expression in living systems in a predictable manner, requires the successful integration of man-made chemistry with the complexity of living biological systems. Can we understand how these small molecules access DNA in the chromatin context and interact with the transcriptional machinery in the nucleus to reprogram gene expression. Currently, our research efforts are aimed in a context broader than medicinal chemistry. Perhaps our findings will contribute to the basic biology of gene regulation in eukaryotic cells, probe chromatin and promoter accessibility in the nucleus, and guide the way toward new thinking regarding the potential therapeutic value of small molecule regulation of gene expression.[79]

References

[1] Arcamone, F.; Nicolell.V; Penco, S.; Orezzi, P.; Pirelli, A., Structure and Synthesis of Distamycin A. *Nature*. **1964**, 203, (494), 1064-&.

[2] Pelton, J. G.; Wemmer, D. E., Structural Characterization of a 2:1 Distamycin A·D(CGCAAATTGGC) Complex by Two-Dimensional NMR. *Proceedings of the National Academy of Sciences of the United States of America*. **1989**, 86, (15), 5723-5727.

[3] Swalley, S. E.; Baird, E. E.; Dervan, P. B., Effects of γ-turn and β-tail amino acids on sequence-specific recognition of DNA by hairpin polyamides. *Journal of the American Chemical Society*. **1999**, 121, (6), 1113-1120.

[4] White, S.; Baird, E. E.; Dervan, P. B., Orientation preferences of pyrrole-imidazole polyamides in the minor groove of DNA. *Journal of the American Chemical Society*. **1997**, 119, (38), 8756-8765.

[5] Wade, W. S.; Mrksich, M.; Dervan, P. B., Design of peptides that bind in the minor groove of DNA at 5'-(A,T)G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif. *Journal of the American Chemical Society*. **1992**, 114, (23), 8783-8794.

[6] Mrksich, M.; Parks, M. E.; Dervan, P. B., Hairpin peptide motif. A new class of oligopeptides for sequence-specific recognition in the minor-groove of double-helical DNA. *Journal of the American Chemical Society*. **1994**, 116, (18), 7983-7988.

[7] Baliga, R.; Baird, E. E.; Herman, D. M.; Melander, C.; Dervan, P. B.; Crothers, D. M., Kinetic consequences of covalent linkage of DNA binding polyamides. *Biochemistry*. **2001**, 40, (1), 3-8.

[8] Pilch, D. S.; Poklar, N.; Baird, E. E.; Dervan, P. B.; Breslauer, K. J., The thermodynamics of polyamide-DNA recognition: Hairpin polyamide binding in the minor groove of duplex DNA. *Biochemistry*. **1999**, 38, (7), 2143-2151.

[9] Herman, D. M.; Baird, E. E.; Dervan, P. B., Stereochemical control of the DNA binding affinity, sequence specificity, and orientation preference of chiral hairpin polyamides in the minor groove. *Journal of the American Chemical Society*. **1998**, 120, (7), 1382-1391.

[10] Rucker, V. C.; Melander, C.; Dervan, P. B., Influence of β-alanine on hairpin polyamide orientation in the DNA minor groove. *Helvetica Chimica Acta*. **2003**, 86, (6), 1839-1851.

[11] Woods, C. R.; Ishii, T.; Wu, B.; Bair, K. W.; Boger, D. L., Hairpin versus extended DNA binding of a substituted β-alanine linked polyamide. *Journal of the American Chemical Society*. **2002**, 124, (10), 2148-2152.

[12] Turner, J. M.; Baird, E. E.; Dervan, P. B., Recognition of seven base pair sequences in the minor groove of DNA by ten-ring pyrrole-imidazole polyamide hairpins. *Journal of the American Chemical Society*. **1997**, 119, (33), 7636-7644.

[13] White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B., Affinity and specificity of multiple hydroxypyrrole/pyrrole ring pairings for coded recognition of DNA. *Journal of the American Chemical Society*. **1999**, 121, (1), 260-261.

[14] Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R., Complex of B-DNA with polyamides freezes DNA backbone flexibility. *Journal of the American Chemical Society*. **2001**, 123, (21), 5044-5049.

[15] Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C., Structural effects of DNA sequence on T·A recognition by hydroxypyrrole/pyrrole pairs in the minor groove. *Journal of Molecular Biology*. **2000**, 295, (3), 557-567.

[16] Hunter, C. A., Sequence-Dependent DNA-Structure - the Role of Base Stacking Interactions. *Journal of Molecular Biology*. **1993**, 230, (3), 1025-1054.

[17] Wemmer, D. E., Designed sequence-specific minor groove ligands. *Annual Review of Biophysics and Biomolecular Structure*. **2000**, 29, 439-461.

[18] Goodsell, D. S.; Kopka, M. L.; Cascio, D.; Dickerson, R. E., Crystal-Structure of Catggccatg and Its Implications for a- Tract Bending Models. *Proceedings of the National Academy of Sciences of the United States of America*. **1993**, 90, (7), 2930-2934.

[19] Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C., Structural basis for G·C recognition in the DNA minor groove. *Nature Structural Biology*. **1998**, 5, (2), 104-109.

[20] Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E., Design of a G·C-specific DNA minor groove-binding peptide. *Science*. **1994**, 266, (5185), 646-650.

[21] Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C., A structural basis for recognition of A·T and T·A base pairs in the minor groove of B-DNA. *Science*. **1998**, 282, (5386), 111-115.

[22] Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B., Sequence selectivity of 3-hydroxypyrrole/pyrrole ring pairings in the DNA minor groove. *Journal of the American Chemical Society*. **1999**, 121, (50), 11621-11629.

[23] Kool, E. T., Preorganization of DNA: Design principles for improving nucleic acid recognition by synthetic oligonucleotides. *Chemical Reviews*. **1997**, 97, (5), 1473-1487.

[24] Bremer, R. E.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B., Recognition of the DNA minor groove by pyrrole-imidazole polyamides: Comparison of desmethyl- and N-methylpyrrole. *Bioorganic & Medicinal Chemistry*. **2000**, 8, (8), 1947-1955.

[25] Kelly, J. J.; Baird, E. E.; Dervan, P. B., Binding site size limit of the 2:1 pyrrole-imidazole polyamide-DNA motif. *Proceedings of the National Academy of Sciences of the United States of America*. **1996**, 93, (14), 6981-6985.

[26] Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B., Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *Journal of the American Chemical Society*. **1998**, 120, (25), 6219-6226.

[27] Wang, C. C. C.; Ellervik, U.; Dervan, P. B., Expanding the recognition of the minor groove of DNA by incorporation of β-alanine in hairpin polyamides. *Bioorganic & Medicinal Chemistry*. **2001**, 9, (3), 653-657.

[28] Trauger, J. W.; Baird, E. E.; Dervan, P. B., Recognition of 16 base pairs in the minor groove of DNA by a pyrrole-imidazole polyamide dimer. *Journal of the American Chemical Society*. **1998**, 120, (14), 3534-3535.

[29] de Clairac, R. P. L.; Seel, C. J.; Geierstanger, B. H.; Mrksich, M.; Baird, E. E.; Dervan, P. B.; Wemmer, D. E., NMR characterization of the aliphatic β/β pairing for recognition of a A·T/T·A base pairs in the minor groove of DNA. *Journal of the American Chemical Society*. **1999**, 121, (13), 2956-2964.

[30] Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B., Cycle polyamide motif for recognition of the minor groove of DNA. *Journal of the American Chemical Society*. **1999**, 121, (6), 1121-1129.

[31] Melander, C.; Herman, D. M.; Dervan, P. B., Discrimination of A/T sequences in the minor groove of DNA within a cyclic polyamide motif. *Chemistry-a European Journal*. **2000**, 6, (24), 4487-4497.

[32] Weyermann, P.; Dervan, P. B., Recognition of ten base pairs of DNA by head-to-head hairpin dimers. *Journal of the American Chemical Society*. **2002**, 124, (24), 6872-6878.

[33] Kers, I.; Dervan, P. B., Search for the optimal linker in tandem hairpin polyamides. *Bioorganic & Medicinal Chemistry*. **2002**, 10, (10), 3339-3349.

[34] Nguyen, D. H.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B., Alternative heterocycles for DNA recognition: An N-methylpyrazole/N-methylpyrrole pair specifies for A·T/T·A base pairs. *Bioorganic & Medicinal Chemistry*. **2001**, 9, (1), 7-17.

[35] Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B., Toward an understanding of the chemical etiology for DNA minor-groove recognition by polyamides. *Helvetica Chimica Acta*. **2002**, 85, (12), 4485-4517.

[36] Zhan, Z. Y. J.; Dervan, P. B., Alternative heterocycles for DNA recognition: A 3-pyrazole/pyrrole pair specifies for G·C base pairs. *Bioorganic & Medicinal Chemistry*. **2000**, 8, (10), 2467-2474.

[37] Ellervik, U.; Wang, C. C. C.; Dervan, P. B., Hydroxybenzamide/pyrrole pair distinguishes T center dot A from A center dot T base pairs in the minor groove of DNA. *Journal of the American Chemical Society*. **2000**, 122, (39), 9354-9360.

[38] Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B., Shape selective recognition of T·A base pairs by hairpin polyamides containing N-terminal 3-methoxy (and 3-chloro) thiophene residues. *Bioorganic & Medicinal Chemistry*. **2003**, 11, (20), 4333-4340.

[39] Burli, R. W.; Jones, P.; McMinn, D.; Le, Q.; Duan, J. X.; Kaizerman, J. A.; Difuntorum, S.; Moser, H. E., DNA binding ligands targeting drug-resistant Gram-positive bacteria. Part 2: C-terminal benzimidazoles and derivatives. *Bioorganic & Medicinal Chemistry Letters*. **2004**, 14, (5), 1259-1263.

[40] Minehan, T. G.; Gottwald, K.; Dervan, P. B., Molecular recognition of DNA by Hoechst benzimidazoles: Exploring beyond the pyrrole-imidazole-hydroxypyrrole polyamide-pairing code. *Helvetica Chimica Acta*. **2000**, 83, (9), 2197-2213.

[41] Burli, R. W.; McMinn, D.; Kaizerman, J. A.; Hu, W. H.; Ge, Y. G.; Pack, Q.; Jiang, V.; Gross, M.; Garcia, M.; Tanaka, R.; Moser, H. E., DNA binding ligands targeting drug-resistant gram-positive bacteria. Part 1: Internal benzimidazole derivatives. *Bioorganic & Medicinal Chemistry Letters*. **2004**, 14, (5), 1253-1257.

[42] Briehn, C. A.; Weyermann, P.; Dervan, P. B., Alternative heterocycles for DNA recognition: The benzimidazole/imidazole pair. *Chemistry-a European Journal.* **2003**, 9, (9), 2110-2122.

[43] Renneberg, D.; Dervan, P. B., Imidazopyridine/pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove recognition. *Journal of the American Chemical Society.* **2003**, 125, (19), 5707-5716.

[44] Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E., The Molecular-Origin of DNA Drug Specificity in Netropsin and Distamycin. *Proceedings of the National Academy of Sciences of the United States of America.* **1985**, 82, (5), 1376-1380.

[45] Janssen, S.; Durussel, T.; Laemmli, U. K., Chromatin opening of DNA satellites by targeted sequence- specific drugs. *Molecular Cell.* **2000**, 6, (5), 999-1011.

[46] Dervan, P. B.; Urbach, A. R., *The Importance of β-Alanine for Recognition of the Minor Groove of DNA.* ed.; Verlag Helvetica Chimica Acta: Zurich, 2001; 'Vol.' p 327-339.

[47] Urbach, A. R.; Dervan, P. B., Toward rules for 1:1 polyamide:DNA recognition. *Proceedings of the National Academy of Sciences of the United States of America.* **2001**, 98, (8), 4343-4348.

[48] Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B., Structure of a β-alanine-linked polyamide bound to a full helical turn of purine tract DNA in the 1:1 motif. *Journal of Molecular Biology.* **2002**, 320, (1), 55-71.

[49] Baird, E. E.; Dervan, P. B., Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *Journal of the American Chemical Society.* **1996**, 118, (26), 6141-6146.

[50] Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B., Solid-phase synthesis of DNA binding polyamides on oxime resin. *Bioorganic & Medicinal Chemistry.* **2002**, 10, (8), 2767-2774.

[51] Fattori, D.; Kinzel, O.; Ingallinella, P.; Bianchi, E.; Pessi, A., A practical approach to the synthesis of hairpin polyamide-peptide conjugates through the use of a safety-catch linker. *Bioorganic & Medicinal Chemistry Letters.* **2002**, 12, (8), 1143-1147.

[52] Wurtz, N. R.; Turner, J. M.; Baird, E. E.; Dervan, P. B., Fmoc solid phase synthesis of polyamides containing pyrrole and imidazole amino acids. *Organic Letters.* **2001**, 3, (8), 1201-1203.

[53] Belitsky, J. M.; Leslie, S. J.; Arora, P. S.; Beerman, T. A.; Dervan, P. B., Cellular uptake of *N*-methylpyrrole/*N*-methylimidazole polyamide-dye conjugates. *Bioorganic & Medicinal Chemistry.* **2002**, 10, (10), 3313-3318.

[54] Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B., Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proceedings of the National Academy of Sciences of the United States of America.* **2003**, 100, (21), 12063-12068.

[55] Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B., Regulation of gene expression by small molecules. *Nature.* **1997**, 387, (6629), 202-205.

[56] Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B.; Gottesfeld, J. M., Importance of minor groove binding zinc fingers within the transcription factor IIIA-DNA complex. *Journal of Molecular Biology.* **1997**, 274, (4), 439-445.

[57] McBryant, S. J.; Baird, E. E.; Trauger, J. W.; Dervan, P. B.; Gottesfeld, J. M., Minor groove DNA-protein contacts upstream of a tRNA gene detected with a synthetic DNA binding ligand. *Journal of Molecular Biology*. **1999**, 286, (4), 973-981.

[58] Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B., Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proceedings of the National Academy of Sciences of the United States of America*. **1998**, 95, (22), 12890-12895.

[59] Ehley, J. A.; Melander, C.; Herman, D.; Baird, E. E.; Ferguson, H. A.; Goodrich, J. A.; Dervan, P. B.; Gottesfeld, J. M., Promoter scanning for transcription inhibition with DNA-binding polyamides. *Molecular and Cellular Biology*. **2002**, 22, (6), 1723-1733.

[60] Wurtz, N. R.; Dervan, P. B., Sequence specific alkylation of DNA by hairpin pyrrole- imidazole polyamide conjugates. *Chemistry & Biology*. **2000**, 7, (3), 153-161.

[61] Takahashi, R.; Bando, T.; Sugiyama, H., Specific alkylation of human telomere repeats by hairpin pyrrole-imidazole polyamide. *Bioorganic & Medicinal Chemistry*. **2003**, 11, (12), 2503-2509.

[62] Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Dervan, P. B.; Graves, B. J.; Gottesfeld, J. M., Inhibition of Ets-1 DNA binding and ternary complex formation between Ets-1, NF-κB, and DNA by a designed DNA-binding ligand. *Journal of Biological Chemistry*. **1999**, 274, (18), 12765-12773.

[63] Wurtz, N. R.; Pomerantz, J. L.; Baltimore, D.; Dervan, P. B., Inhibition of DNA binding by NF-κB with pyrrole-imidazole polyamides. *Biochemistry*. **2002**, 41, (24), 7604-7609.

[64] Chiang, S. Y.; Burli, R. W.; Benz, C. C.; Gawron, L.; Scott, G. K.; Dervan, P. B.; Beerman, T. A., Targeting the Ets binding site of the HER2/*neu* promoter with pyrrole-imidazole polyamides. *Journal of Biological Chemistry*. **2000**, 275, (32), 24246-24254.

[65] Lenzmeier, B. A.; Baird, E. E.; Dervan, P. B.; Nyborg, J. K., The Tax protein-DNA interaction is essential for HTLV-I transactivation in vitro. *Journal of Molecular Biology*. **1999**, 291, (4), 731-744.

[66] Bremer, R. E.; Baird, E. E.; Dervan, P. B., Inhibition of major-groove-binding proteins by pyrrole-imidazole polyamides with an Arg-Pro-Arg positive patch. *Chemistry & Biology*. **1998**, 5, (3), 119-133.

[67] Bremer, R. E.; Wurtz, N. R.; Szewczyk, J. W.; Dervan, P. B., Inhibition of major groove DNA binding bZIP proteins by positive patch polyamides. *Bioorganic & Medicinal Chemistry*. **2001**, 9, (8), 2093-2103.

[68] Fechter, E. J.; Dervan, P. B., Allosteric inhibition of protein-DNA complexes by polyamide-intercalator conjugates. *Journal of the American Chemical Society*. **2003**, 125, (28), 8476-8485.

[69] Nguyen-Hackley, D. H., Ramm E., Taylor C.M., Joung J. K., Dervan, P. B., and Pabo C. O., Allosteric Inhibition of Zinc-Finger Binding in the Major Groove of DNA by Minor-Groove Binding Ligands. *Biochemistry*. **2004**, ASAP.

[70] Olenyuk, B. Z.; Zhang, G. J.; Klco, J. M.; Nickols, N. G.; Kaelin, W. G.; Dervan, P. B., Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. *Proceedings of the National Academy of Sciences of the United States of America*. **2004**, 101, (48), 16768-16773.

[71] Dickinson, L. A.; Burnett, R.; Melander, C.; Edelson, B. S.; Arora, P. S.; Dervan, P. B.; Gottesfeld, J. M., Arresting cancer proliferation by small-molecule gene regulation. *Chemistry & Biology*. **2004**, 11, (11), 1583-1594.

[72] Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Ghazal, P.; Dervan, P. B.; Gottesfeld, J. M., Anti-repression of RNA polymerase II transcription by pyrrole-imidazole polyamides. *Biochemistry*. **1999**, 38, (33), 10801-10807.

[73] Coull, J. J.; He, G. C.; Melander, C.; Rucker, V. C.; Dervan, P. B.; Margolis, D. M., Targeted derepression of the human immunodeficiency virus type 1 long terminal repeat by pyrrole-imidazole polyamides. *Journal of Virology*. **2002**, 76, (23), 12349-12354.

[74] Ansari, A. Z.; Mapp, A. K., Modular design of artificial transcription factors. *Current Opinion in Chemical Biology*. **2002**, 6, (6), 765-772.

[75] Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B., Activation of gene expression by small molecule transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*. **2000**, 97, (8), 3930-3935.

[76] Ansari, A. Z.; Mapp, A. K.; Nguyen, D. H.; Dervan, P. B.; Ptashne, M., Towards a minimal motif for artificial transcriptional activators. *Chemistry & Biology*. **2001**, 8, (6), 583-592.

[77] Arora, P. S.; Ansari, A. Z.; Best, T. P.; Ptashne, M.; Dervan, P. B., Design of artificial transcriptional activators with rigid poly-L-proline linkers. *Journal of the American Chemical Society*. **2002**, 124, (44), 13067-13071.

[78] McGinnis, W.; Krumlauf, R., Homeobox Genes and Axial Patterning. *Cell*. **1992**, 68, (2), 283-302.

[79] Arndt, H. D.; Hauschild, K. E.; Sullivan, D. P.; Lake, K.; Dervan, P. B.; Ansari, A. Z., Toward artificial developmental regulators. *Journal of the American Chemical Society*. **2003**, 125, (44), 13322-13323.

Chapter 2

Toward an Understanding of the Chemical Etiology for DNA Minor-Groove Recognition by Polyamides

The text of this chapter was taken in part from a manuscript coauthored with Michael A. Marques, Adam R. Urbach and Professor Peter B. Dervan (Caltech)

(Marques, M. A.; Urbach, A. R.; Doss, R. M. and Dervan, P. B. "*Toward A Chemical Etiology for DNA Minor Groove Recognition*" *Helvetica Chimica Acta* **2002,** 85, (12), 4485-4517.)

Abstract.

Crescent-shaped polyamides composed of aromatic amino acids, N-methylimidazole (**Im**), N-methylpyrrole (**Py**), and 3-hydroxypyrrole (**Hp**) bind in the minor groove of DNA as 2:1 and 1:1 ligand:DNA complexes. DNA sequence specificity can be attributed to shape-selective recognition and the unique corners or pairs of corners presented by each heterocycle(s) to the edges of the base pairs on the floor of the minor groove. Here we examine the relationship between heterocycle structure and DNA sequence specificity for a family of five-membered aromatic amino acids. Using quantitative DNase I footprinting, the recognition behavior of polyamides containing eight different aromatic amino acids: 1-methylpyrazole (**Pz**), 1H-pyrrole (**Nh**), 5-methylthiazole (**Nt**), 4-methylthiazole (**Th**), 4-methylthiophene (**Tn**), thiophene (**Tp**), 3-hydroxythiophene (**Ht**), and furan (**Fr**) were compared with the parent rings **Py**, **Im**, **Hp**, for their ability to discriminate between the four Watson-Crick base pairs in the DNA minor groove. From analysis of the data and molecular modeling, the geometry inherent to each heterocycle plays a significant role in the ability of polyamides to differentiate between DNA sequences. Binding appears sensitive to changes in curvature complementarity between the polyamide and DNA. The **Th/Py** pair affords a modest 3-fold discrimination of T•A vs. A•T and suggests that a sulfur atom in the thiophene ring prefers to lie opposite T not A.

2.1 Introduction.

Many diseases are related to aberrant gene expression and the ability to reprogram transcription in a cell by small molecules could be important in biology and human medicine. Minor groove binding polyamides which bind predetermined DNA sequences offer a chemical approach to artificial gene regulation. These molecules are based on analogues of the 1-methylpyrrolecarboxamide ring (**Py**) of the natural products netropsin and distamycin A which have been shown to bind in the minor groove of DNA in 1:1 and 2:1 ligand-DNA stoichiometries (**Figure 2.1**).[1-4] **Py** is specific for A•T and T•A base pairs due to steric exclusion of the guanine amino group (G-NH2).[1-3] Base pair specificity can be altered by changing the functional group(s) presented to the floor of the DNA minor groove. Stabilizing and destabilizing interactions with the different edges of

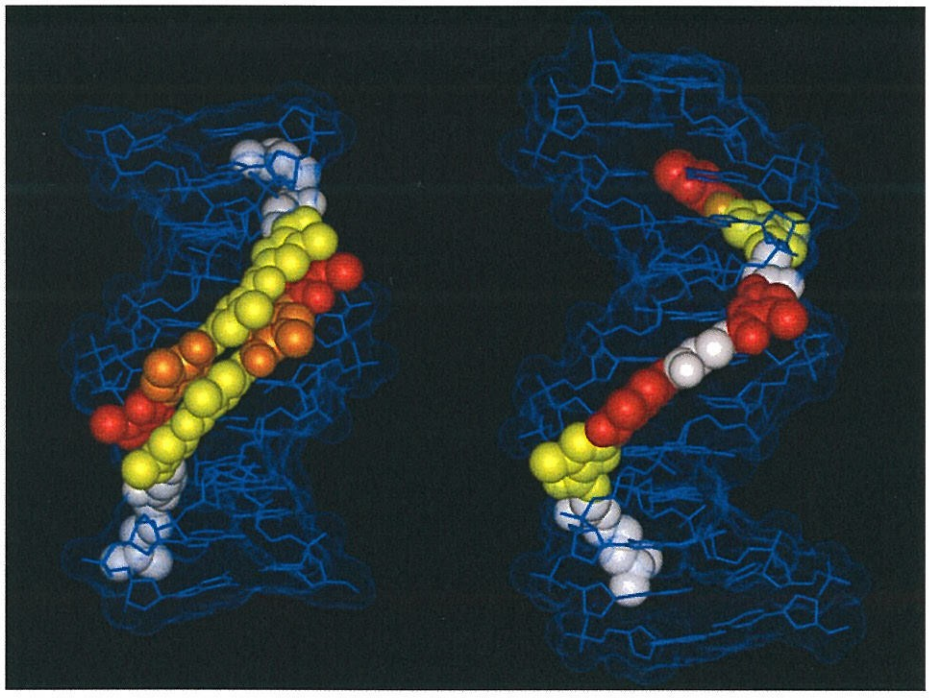


Figure 2.1 High-resolution structures of polyamides bound to DNA. (left) 2:1 motif determined by X-ray crystallography[9] (right) 1:1 motif determined by NMR.[8] DNA is shown as a stick model in blue. Polyamides are shown as space-filling models, with imidazole residues in red, hydroxypyrrole in orange, pyrrole in yellow, and aliphatic residues in white.

the four Watson-Crick base pairs are modulated by specific hydrogen bonds and, importantly, steric fit or shape complimentarity. For example, 1-methylimidazole (**Im**) presents the DNA with nitrogen and its lone pair sp^2 orbital which can accept a hydrogen bond from G-NH2.[1-6] Additionally, 3-hydroxypyrrole (**Hp**) presents a hydroxyl group that is sterically accommodated opposite T not A and, in addition, can donate hydrogen bonds to the O2 of thymine.[7-8] For discrimination of each of the Watson-Crick base pairs the 2:1 stoichiometry involving unsymmetrical antiparallel cofacial pairs appears to be the best solution such that **Im/Py** is specific for G•C, **Py/Im** for C•G, **Hp/Py** for T•A, and **Py/Hp** for A•T.

The pairing rules have proven useful for the recognition of hundreds of DNA sequences by polyamides. However, sequence-dependent DNA structural variation (such as minor groove width) makes binding affinity and specificity at many DNA sequences unpredictable, which leads us to continue our search for new aromatic amino acid residues of slightly different shape (curvature and twist) for minor groove recognition. Importantly, we find that the **Hp** residue can degrade over time in the presence of acid or free radicals and a robust replacement for the **Hp/Py** pair suitable for use in biological studies is desirable. Several five-membered heterocyclic residues other than **Py**, **Im**, and **Hp** have been investigated previously, including furan (**Fr**),[9] thiazole (**Nt**),[10,11] and pyrazole (**Pz**),[11,12] with no new specificity uncovered. This raises the issue whether there is something "special" about the N-methylpyrrole analogs **Py**, **Im** and **Hp** for minor groove recognition. We attempt here to broaden the repetoire of aromatic five membered heterocycles for DNA recognition by synthesizing and characterizing a family of five-membered heterocyclic carboxamides grouped by the type of functionality directed

toward the floor of the DNA minor groove (**Figure 2.2**). Analogs of **Py** would be pyrazole (**Pz**) and 1H-pyrrole (**Nh**) which project a hydrogen atom toward the floor of the minor groove. Analogs of **Im** would be 5-methylthiazole (**Nt**), and furan (**Fr**) which project lone pairs of electrons from nitrogen or oxygen. Analogs of **Hp** is 3-hydroxythiophene (**Ht**) which projects a hydroxyl group. 4-Methylthiazole (**Th**), 4-methylthiophene (**Tn**) and thiophene (**Tp**) project a large heteroatom sulfur to the floor of DNA and likely represent a new class for shape selective recognition. We chose to maintain a *five-membered heterocyclic framework* in our library to retain overall the crescent shape of the polyamide ligand and to observe the effects of small structural changes resulting from single atom substitution on DNA base pair specificity. One anticipates that substitution of atoms on the five-membered ring projecting away from the DNA minor groove (i.e. the non-reading frame) will have effects on bond lengths and

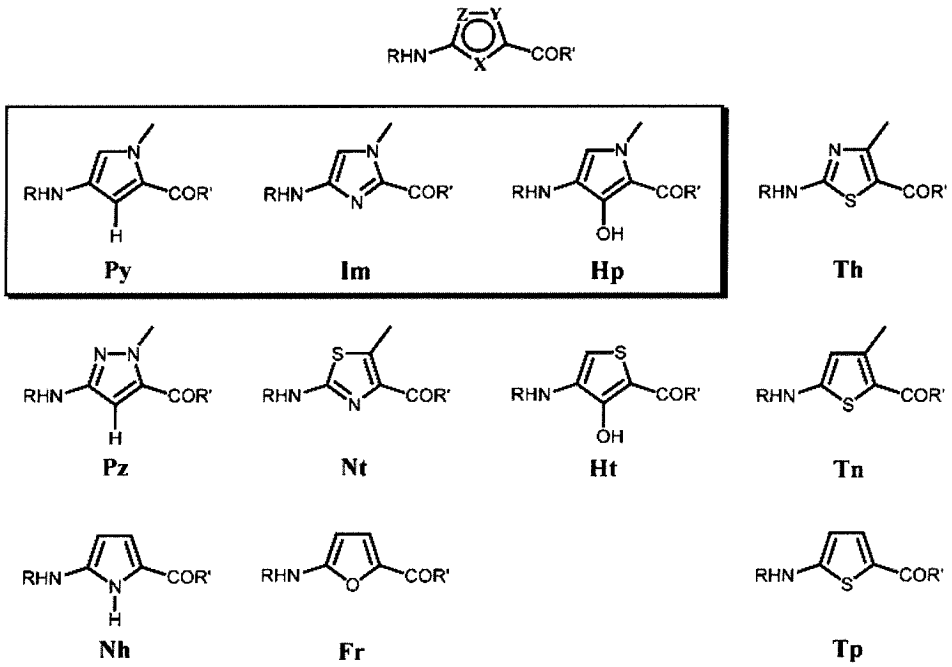


Figure 2.2 Family of heterocyclic amino acids studied here. Centered above is a schematic showing the five-membered heterocyclic framework with the variable positions labeled X, Y, and Z. The parent Im, Py, and Hp residues are boxed. All residues are shown with the functionality that faces the DNA minor groove pointed down (X).

bond angles of each heterocycle as well and allows us to ask how tolerant is DNA to subtle alterations in curvature and twist of the minor groove binding ligand.

The covalent head-to-tail linkage of polyamide subunits in a 2:1 complex results in a hairpin oligomer with increased DNA affinity and sequence specificity.[13,14] The hairpin motif avoids the ambiguity of slipping between the side-to-side stacked subunits,[15] and "locks" individual ring pairings in a predictable cofacial manner. In a formal sense, the hairpin oligomer provides a predictable foldamer for studying the DNA recognition characteristics of new ring pairings. We have reported previously the sequence specificities of **Py/Py**, **Hp/Py**, **Pz/Py**, and **Th/Py** pairings at a single position within the hairpin polyamide sequence context ImIm**X**Py-γ -Im**Py**PyPy-β-Dp (**X** = **Py**, **Hp**, **Pz**, and **Th**; γ = gamma amino butyric acid; β = beta alanine; Dp = dimethylaminopropylamide) opposite the four Watson-Crick base pairs within the sequence context 5'-ATGG**X**CA-3' (**X** = A, T, G, and C).[11,17] We found that **Pz/Py** can mimic **Py/Py** and, in a surprising result, **Th/Py** bound all four base pairs with low affinity and no specificity. Undaunted we broaden our data set here to include **Nh**, **Fr**, **Ht**, **Tp** and **Tn** residues each paired with **Py** in the same hairpin context for comparison (**Figure 2.3**). For the sake of completeness and to provide a comparative analysis of heterocycle behavior opposite the four Watson-Crick base pairs in *both stoichiometries* (**Figure 2.3**), we analyze polyamides in a 1:1 polyamide:DNA complex (1:1 motif) of type Im-β-ImPy-β-**X**-β-ImPy-β-Dp (**X** = **Py**, **Hp**, **Nh**, **Ht**, **Fr**, **Nt**, **Tn**, and **Th**) which bind DNA sequences 5'-AAAGA**X**AGAAG-3' (**X** = A, T, G, and C) in a single orientation.[16-19]

Quantitative DNase I footprinting was used to determine the equilibrium association constant for each complex. *Ab initio* computational modeling of the

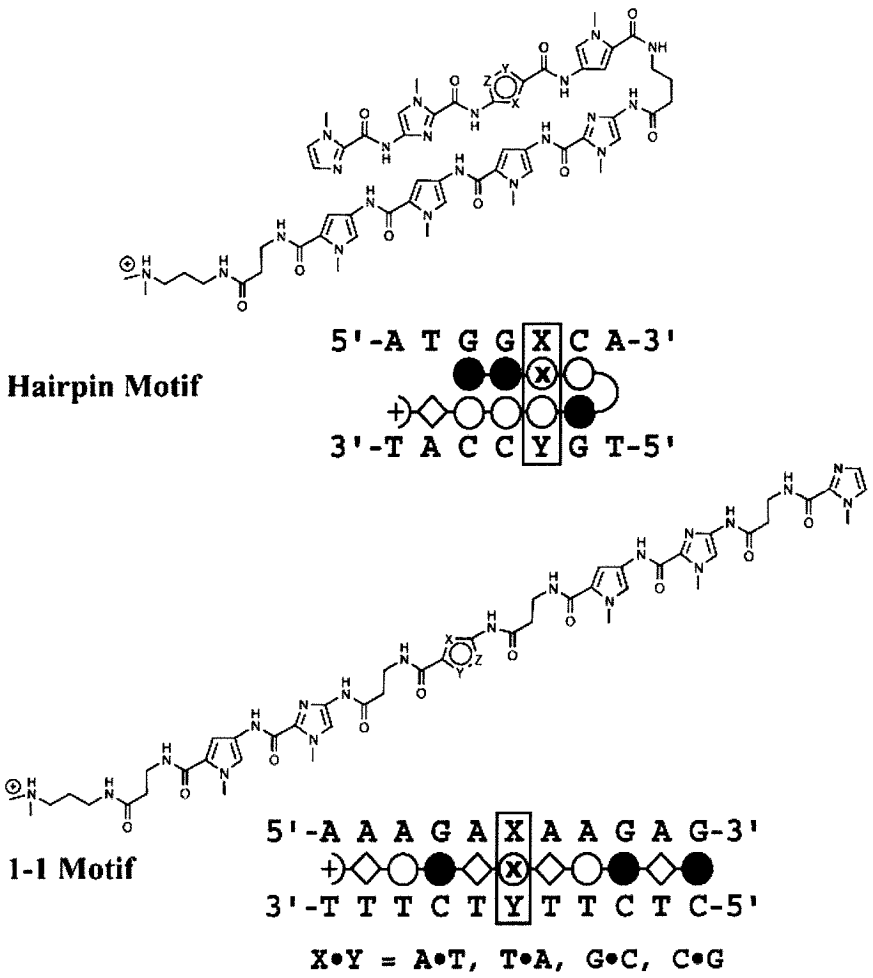


Figure 2.3 Schematic illustrating the examination of sequence selectivity against the four Watson-Crick base pairs within hairpin (top) and 1:1 (bottom) motifs. Each chemical structure has a variable residue containing X, Y, and Z-labeled positions, which are designated in Fig. 2. The dot models shown below each chemical structure illustrate the binding mode with the polyamide shown inside its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; the γ-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; and the novel heterocycles are indicated by a circle containing an X.

heterocyclic amino acids was implemented to derive their inherent geometric and electronic parameters to guide respectively interpretation of the experimental outcome. The combination of experiment and modeling provides insight to the origin of DNA sequence discrimination by polyamides.

2.2 Monomer, Dimer and Polyamide Synthesis.

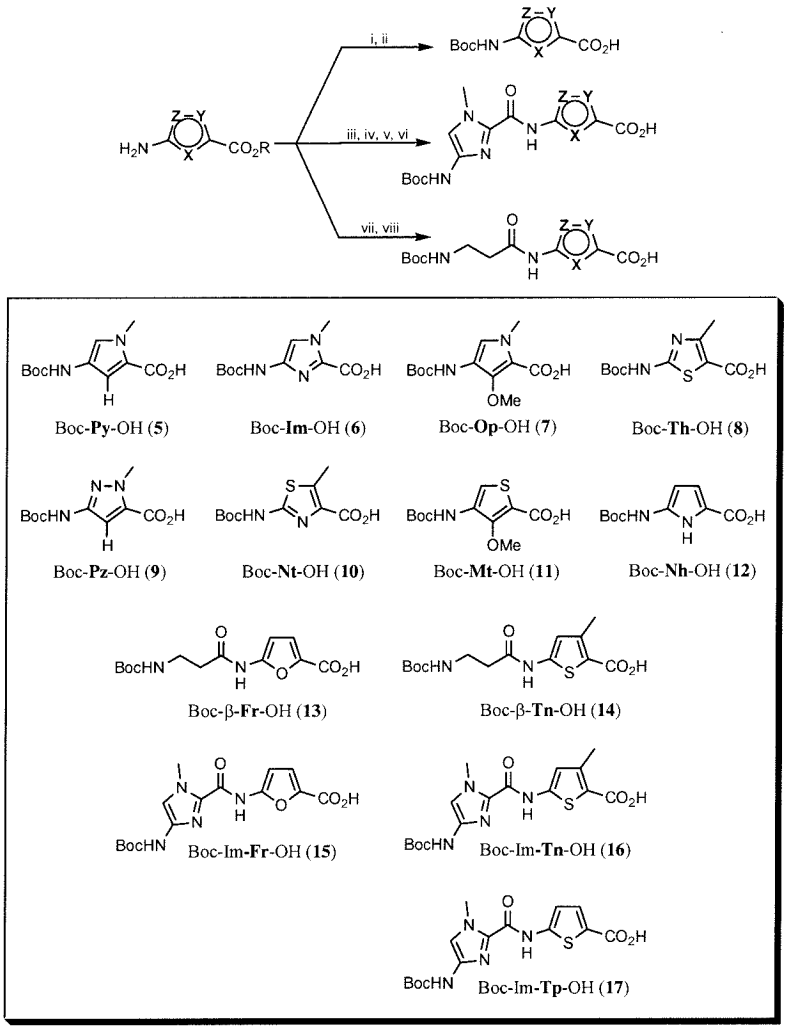


Figure 2.4 (Top) Formation of monomer and dimer Boc-protected amino acids. (Box) Monomeric and dimeric units: (i) (Py, X = C-H, Y = N-Me, Z = C-H, R = Me), (Im, X = N, Y = N-Me, Z, = C-H, R = Et), (Op, X = C-OMe, Y = N-Me, Z = C-H, R = Et), (Th, X = S, Y = C-Me, Z = N, R = Et), (Pz, X = C-H, Y = N-Me, Z = C-H, R = Et), (Nt, X = N, Y = C-Me, Z = S, R = H), and (Nh, X = N-H, Y = Z = C-H, R = Et), $(Boc)_2O$, DIEA, DMF, 60 °C, 12-18 h.; or (Hm, X = C-OMe, Y = S, Z = C-H, R = Me) $(Boc)_2O$, DIEA, TEA CH_2Cl_2, 60 °C, 12-18 h; (ii) (Py, X = C-H, Y = N-Me, Z = C-H, R = Me), (Im, X = N, Y = N-Me, Z, = C-H, R = Et), (Op, X = C-OMe, Y = N-Me, Z = C-H, R = Et), (Th, X = S, Y = C-Me, Z = N, R = OEt), (Pz, X = C-H, Y = N-Me, Z = C-H, R = Et), and (Nh, X = N-H, Y = Z = C-H, R = Et), 1N NaOH, MeOH, R. T., 3-4 h.; or (Nt, X = N, Y = C-Me, Z = S, R = H), 1N NaOH, MeOH, R. T., 1 h; or (Hm, X = C-OMe, Y = S, Z = C-H, R = Me), KOH, MeOH, 60 °C, 4-6 h; (iii) (Fu, X = O, Y = Z = C-H, R = Me), (Tn, X = S, Y = C-Me, Z = C-H, R = Me) and (Dt, X = S, Y = Z = C-H, R = Me), O_2N-Im-$COCCl_3$, EtOAc, DIEA, 35 °C, 10-12 h; (iv) (Fu, X = O, Y = Z = C-H, R = Me), (Tn, X = S, Y = C-Me, Z = C-H, R = Me) and (Dt, X = S, Y = Z = C-H, R = Me), H_2, Pd/C 10%, EtOAc, R. T., 1.5 h; (v) (Fu, X = O, Y = Z = C-H, R = Me), (Tn, X = S, Y = C-Me, Z = C-H, R = Me) and (Dt, X = S, Y = Z = C-H, R = Me), $(Boc)_2O$, DIEA, DMF, 60 °C, 12-18 h; (vi) (Fu, X = O, Y = Z = C-H, R = Me), 1N NaOH, MeOH, R. T., 3 h. (Tn, X = S, Y = C-Me, Z = C-H, R = Me) and (Dt, X = S, Y = Z = C-H, R = Me), 1N NaOH, MeOH, 60 °C, 4-6 h; (vii) (Fu, X = O, Y = Z = C-H, R = Me) and (Tn, X = S, Y = C-Me, Z = C-H, R = Me), (Boc-β-alanine)$_2$O, DMF, DIEA, 40 °C, 12-18 h; (viii) (Fu, X = O, Y = Z = C-H, R = Me), 1N NaOH, MeOH, R. T., 3 h. (Tn, X = S, Y = C-Me, Z = C-H, R = Me), 1N NaOH, MeOH, 60 °C, 4-6 h.

Polyamides were synthesized manually on solid support by the stepwise addition of monomeric and dimeric Boc-protected amino acids dimers, **5-17** (**Figure 2.4**). Boc-protected amino acid monomers and dimers for **Im**, **Py**, and **Tn-Im** were synthesized according to previously reported procedures. Syntheses of the **Tn-Py** dimer is shown in **Figure 2.5c**. Boc-protected monomeric amino acids **5-12** suitable for solid-phase synthesis were prepared in two steps from their NH_2-X-OR analogues (**Figure 2.4** and **Figures 2.5 a-f**). However, furan and thiophene amines were unreactive to coupling on solid support, and therefore dimers **13-17** were pre-formed in solution under strong acylation conditions prior to solid-phase coupling. Synthetic schemes for all Boc-protected amino acids used in this study follows (**Figures 2.5 a-f**).

Furan (Fr). The nitro-furan-ester (**18**) was prepared from methyl 2-furoate.[21] The amino ester of furan (**1**) was synthesized from **18** using H_2 (500 psi) and palladium on carbon (Pd/C) and isolated as the free base from ethyl acetate by precipitation from hexanes. The free base is a stable crystalline solid at room temperature. The nitro-imidazole-furan ester (NO_2-Im-Fr-OMe, **19**) was prepared by condensing the nitro-imidazole-trichloroketone (NO_2-Im-$COCCl_3$) with (**1**) in ethyl acetate at 35 °C. The dimer product (**19**) is rather insoluble in ethyl acetate and begins to precipitate upon formation. Reduction of **19** to the hydrochloride salt (HCl•H_2N-Im-Fr-OMe, **20**) was carried out using H_2 (500 psi) and Pd/C, followed by addition of 2M HCl in diethyl ether. **20** was Boc-protected using Boc-anhydride, DIEA and DMF at 60 °C for 12-18 h to give Boc-Im-Fr-OMe (**21**). Elevated temperatures and extended reaction times are necessary, presumably due to the poor nucleophilicity of the imidazole amine. Saponification of **21**

using 1N aqueous NaOH and methanol at room temperature provided the target dimer, Boc-Im-Fr-OH (**15**) (**Figure 2.5**).

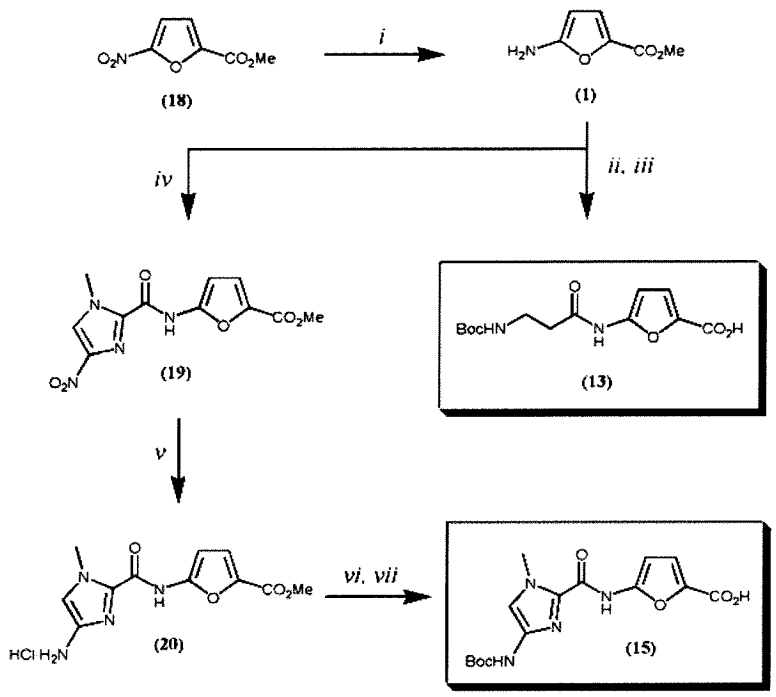


Figure 2.5 Synthesis of **Fr** containing dimers. (i) H_2 Pd/C, EtOAc, 500psi, R.T.; (ii) $(Boc\text{-}\beta\text{-}ala)_2O$, DMF, DIEA, DMAP, 40°C; (iii) 1N NaOH, MeOH, R.T.; (iv) NO_2-Im-$COCCl_3$, EtOAc, 35°C; (v) H_2 Pd/C, EtOAc, 500psi, R.T.; (vi) $(BOC)_2O$, DIEA, DMF, 60°C; (vii) 1N NaOH, MeOH, R.T.

Alternatively, the Boc-β-Fr-OMe dimer (**22**) can be synthesized from **1** by coupling to the symmetrical anhydride of Boc-beta-alanine in DMF, DIEA, and DMAP. The anhydride was pre-formed in minutes using DCC in methylene chloride at room temperature. The furan amino ester **1** was then added as a solution in DMF and DIEA, followed by the transacylation catalyst DMAP. The 5-amino group of furan is significantly unreactive,[22] and attempts to couple it on solid support using reagents such as DCC and HOBT, HBTU, HATU, PyBrOP, PyBOP, and TFFH, were unsuccessful. Formation of the Boc-β-Fr-OMe dimer (**22**) occurs slowly at elevated temperature. Heating the reaction above 40 °C did not affect the rate of reaction. Several different

solvent systems were tried, but DMF was found to be optimal. Saponification of **22** was accomplished using a mixture of MeOH and 1N NaOH at room temperature to provide the target Boc-β-Fr-OH (**13**), which is suitable for standard solid phase protocols (**Figure 2.5**).[20]

1-H Pyrrole (Nh). Pyrrole trichloromethyl ketone (**23**) was prepared by adding pyrrole to a mixture of trichloroacetyl chloride in diethyl ether at 0 °C, then warming to room temperature and stirring overnight, followed by precipitation from hexanes. Nitration of **23** was accomplished using acetic anhydride and nitric acid at -40 °C to provide the 5-nitro regioisomer (**24**) as the major product. Regiocontrol of the nitration appears to depend significantly on the reaction temperature. Warmer temperatures provide unfavorable mixtures of 4-nitro and 5-nitro regioisomers. Esterification of the trichloroketone (**24**) was accomplished using ethanolic sodium ethoxide at room temperature to provide the nitro-pyrrole-ester (NO_2-Nh-OEt, **25**) in good yield. Reduction of **25** using H_2 (500 psi) and Pd/C, followed by the addition of 2M hydrogen chloride in diethyl ether, provided the hydrochloride salt (HCl•H_2N-Nh-OEt, **2**), which was then Boc-protected using Boc-anhydride, DIEA and DMF to yield the Boc-pyrrole-ester (Boc-Nh-OEt, **26**). Saponification of **26** was accomplished using 1N NaOH and methanol at room temperature to provide the final monomer unit Boc-Nh-OH (**12**) (**Figure 2.6**).

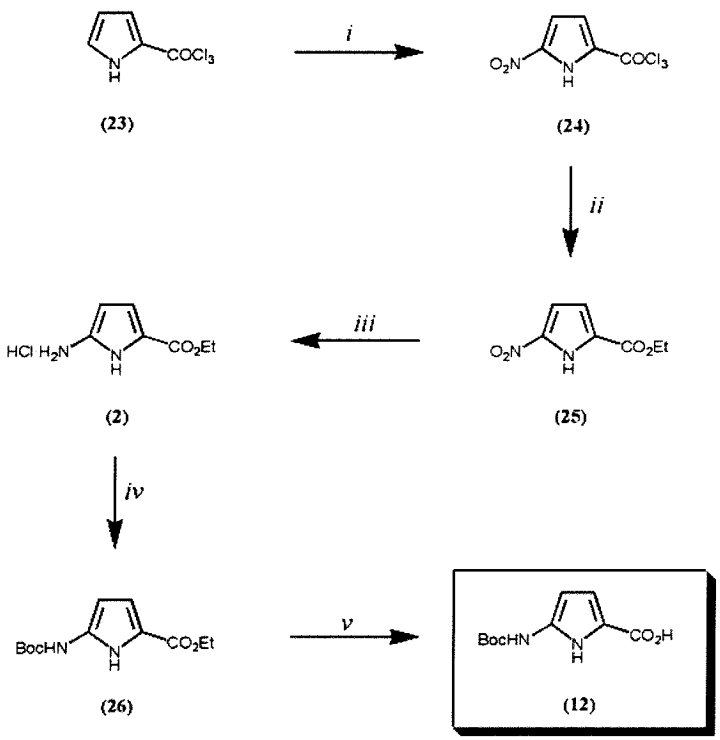


Figure 2.6 Synthesis of **Nh** monomer. (i) HNO_3, Ac_2O, -40°C; (ii) NaOEt, EtOH, Reflux; (iii) H_2 Pd/C, DMF, HCl, Et_2O, R.T.; (iv) $(BOC)_2O$, DIEA, DMAP, DMF, R.T.; (v) 1N NaOH, MeOH, R.T.

4-Methyl Thiophene (Tn). The acyclic precursor to the thiophene ring system was prepared by a Knovenagel reaction involving acetoacetate and cyanoacetic acid, providing **27** as a mixture of E and Z regioisomers in moderate yield after vacuum distillation. Treatment of **27** with sulfur flakes and diethylamine in ethanol yielded the cyclized aminothiophene, and addition of concentrated hydrochloric acid precipitated the hydrochloride salt (HCl•H_2N-Tn-OMe, **3**). Formation of the nitro-imidazole-thiophene ester (NO_2-Im-Tn-OMe, **28**), followed by reduction and Boc-protection to provide HCl•H_2N-Im-Tn-OMe (**29**) and Boc-Im-Tn-OMe (**30**), respectively, was accomplished using the procedures described above for the furan compounds **19**, **20**, and **21**. The Boc-β-Tn-OMe dimer (**31**) was prepared, as described above for **22**, by coupling **3** with the symmetric anhydride of Boc-beta-alanine. The thiophene amine displays low reactivity

comparable to furan. Elevated temperature was necessary to completely saponify methyl esters **30** and **31**. Saponification was carried out in a mixture of 1N NaOH and methanol at 60 °C for 4-6 hours to obtain the hetero dimers Boc-Im-Tn-OH (**16**) and Boc-β-Tn-OH (**14**) (**Figure 2.7**).

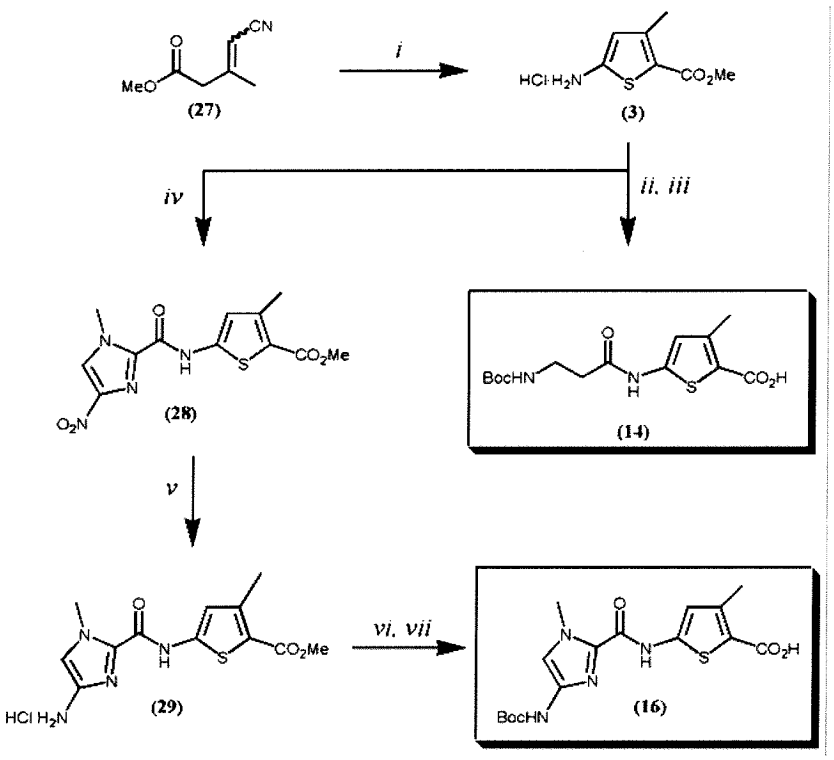


Figure 2.7 Synthesis of **Tn** containing dimers. (i) S, Diethylamine, EtOH, R.T.; (ii) $(Boc\text{-}\beta\text{-}ala)_2O$, DMF, DIEA, DMAP, 40°C; (iii) 1N NaOH, MeOH, 60°C; (iv) NO_2-Im-$COCCl_3$, EtOAc, 35°C; (v) H_2 Pd/C, EtOAc, 500psi, R.T.; (vi) $(BOC)_2O$, DIEA, DMF, 60°C; (vii) 1N NaOH, MeOH, 60°C.

Thiophene (Tp). Treatment of commercially available 5-nitrothiophene-2-carboxaldahyde in acetone with a mixture of sodium hypochlorite and sodium hydrogen phosphate in water, gave the nitro-acid of the des-methyl thiophene ring (NO_2-**Tp**-OH, **32**). Esterification of **32** by refluxing for 48 h in a mixture of sulfuric acid and methanol provided the nitro-ester (NO_2-Tp-OMe, **33**). Reduction of **33** using a mixture of tin(II) chloride dihydrate and hydrochloric acid in ethanol gave the hydrochloride salt

(HCl•H_2N-Tp-OMe, **34**). Formation of the dimer NO_2-Im-Tp-OMe (**35**), followed by reduction to HCl•H_2N-Im-Tp-OMe (**36**), Boc-protection to Boc-Im-Tp-OMe (**37**), and saponification to the target dimer Boc-Im-Tp-OH (**17**) proceeded as described above for furan compounds **19**, **20**, **21**, and **15** (**Figure 2.8**).

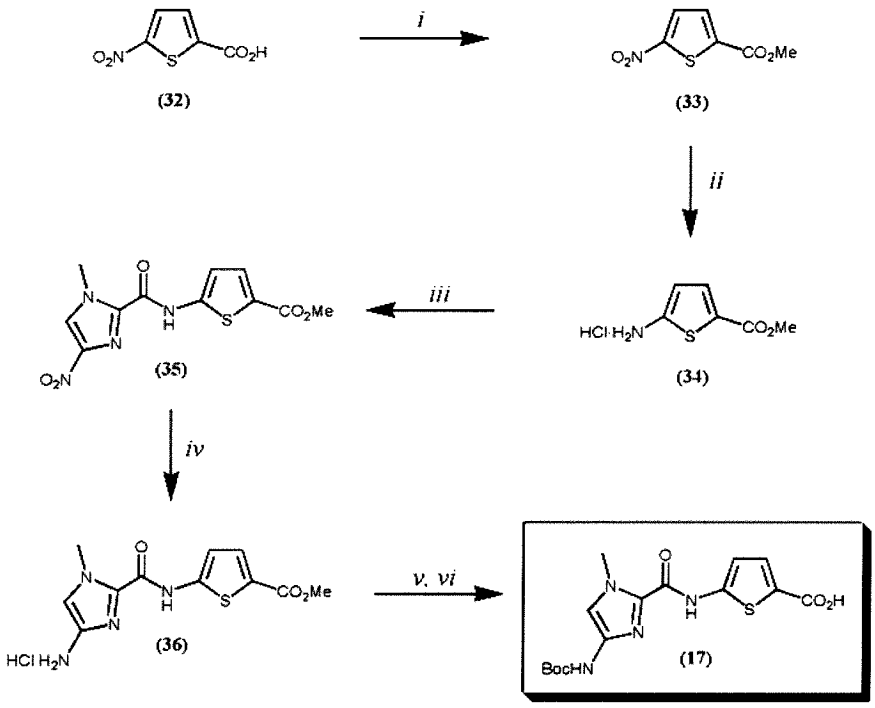


Figure 2.8 Synthesis of **Tp** containing dimers. (i) H_2SO_4, MeOH, Reflux; (ii) $SnCl_2.2H_2O$, 95% EtOH, 37% HCl, 35°C; (iii) NO_2-Im-$COCCl_3$, EtOAc, 35°C; (iv) H_2 Pd/C, EtOAc, 500psi, R.T.; (v) $(BOC)_2O$, DIEA, DMF, 60°C; (vi) 1N NaOH, MeOH, 60°C.

3-Methoxythiophene (Mt). The hydroxythiophene methyl ester (Ht-OMe, **38**) was synthesized by a cyclization reaction between methylthioglycolate and methyl-2-chloroacrylate in methanolic sodium methoxide.[23] **38** was nitrated using a mixture of concentrated sulfuric and nitric acid at -10 °C to give 4-nitro-3-hydroxythiophene (NO_2-Ht-OMe, **39**) as the major regioisomer after column chromatography. Treatment of **39** with diazomethane in diethyl ether afforded the methyl ether (NO_2-Mt-OMe, **40**) in near quantitative yield (**Ht** = hydroxythiophene, **Mt** = methoxythiophene). Reduction of **40**

using a mixture of tin(II) chloride dihydrate, hydrochloric acid and ethanol, gave the hydrochloride salt (HCl•H_2N-Mt-OMe, **4**). Boc-protection of (**4**) was accomplished by heating a mixture of Boc-anhydride, TEA and methylene chloride at 60 °C for 12 h to afford Boc-Mt-OMe (**41**). Saponification of **41** was achieved using methanolic potassium hydroxide and heating at 50 °C for 6 h to give the final monomer Boc-Mt-OH (**11**) (**Figure 2.9**).

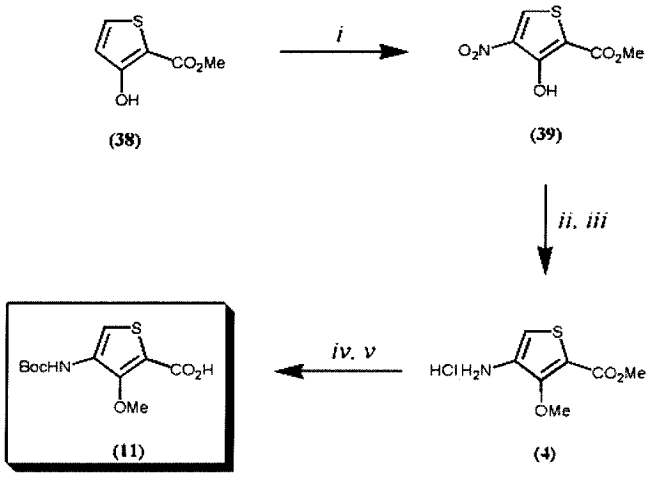


Figure 2.9 Synthesis of **Mt** monomer. (i) HNO_3, H_2SO_4, 0°C; (ii) CH_2N_2, R.T.; (iii) $SnCl_2$, HCl, EtOH, 40°C; (iv) $(BOC)_2O$, TEA, CH_2Cl_2, 60°C; (v) 1N NaOH, MeOH, 60°C.

5-Methylthiazole (Nt). The Boc-protected **5**-methylthiazole amino acid (Boc-Nt-OH, **10**) was synthesized on multi-gram scale by brominating 2-ketobutyric acid, followed by condensation with thiourea and Boc-protection of the amine. For best results, the bromine should be added dropwise over at least two hours, as the reaction is autocatalytic and highly exothermic. Also, the thiourea should be added in small portions with vigorous stirring. Boc-protection was accomplished by dissolving the crude material in DMF and DIEA, followed by the addition of Boc-anhydride and stirring at 60

°C for twelve hours. The material was then stirred in a solution of MeOH and 1N NaOH for ester saponification to provide the target Boc-Nt-OH (**10**) (**Figure 2.10**).

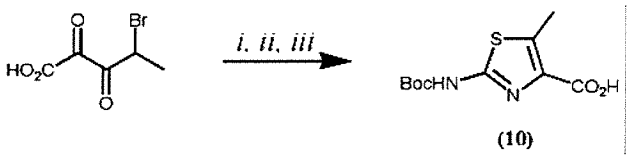


Figure 2.10 Synthesis of **Nt** monomer. (i) CSN_2H_4, Neat, R.T.; (ii) $(BOC)_2O$, DIEA, DMF, DMAP, 40°C; (iii) 1N NaOH, MeOH, 35°C.

Polyamides.

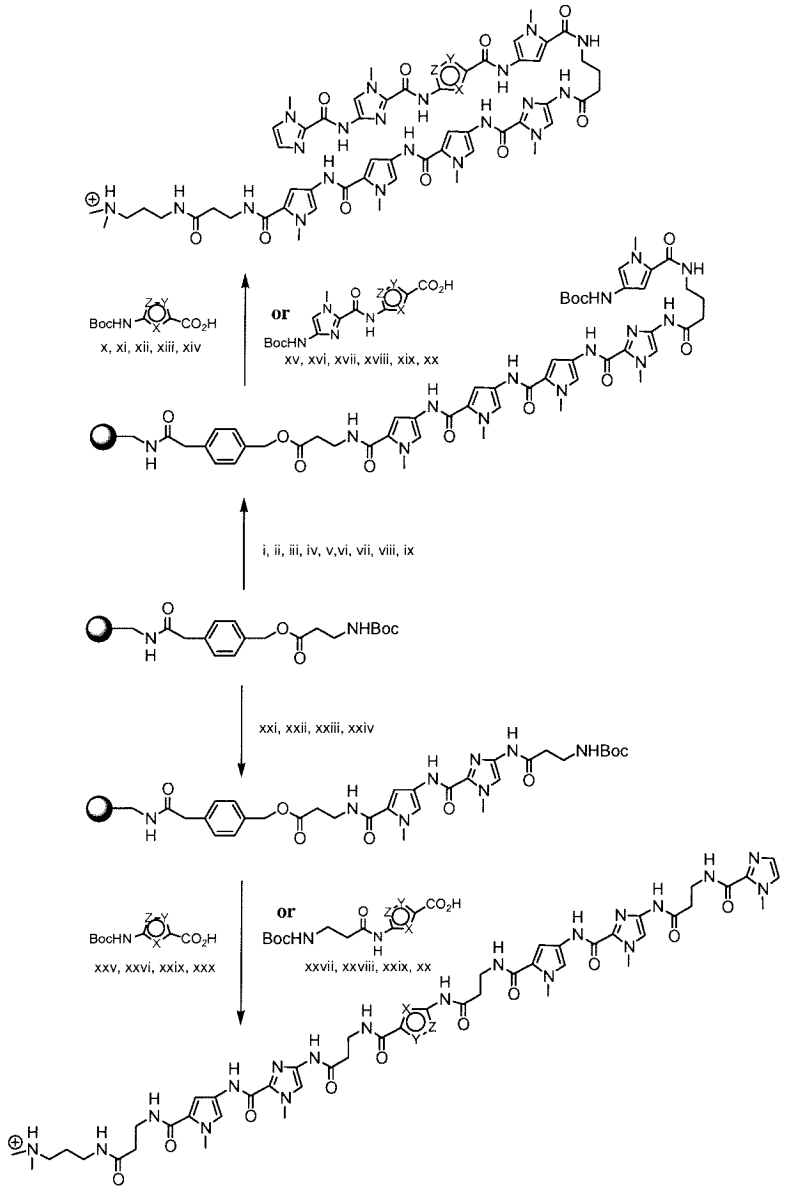


Figure 2.11 Solid phase synthetic scheme for Im-Im-**X**-Py-γ-Py-Py-Py-β-Dp (arrows up from center) and Im-β-ImPy-β-**X**-β-ImPy-β-Dp (arrows down from center) starting from commercially available Boc-β-Pam resin: (i) 80% TFA/DCM, 0.4M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) repeat steps (i) and (ii) x 2. (iv) 80% TFA/DCM, 0.4M PhSH; (v) Boc-Im-OH (HBTU, DIEA, DMF); (vi) 80% TFA/DCM, 0.4M PhSH; (vii) Boc-γ-OH (HBTU, DIEA, DMF); (viii) 80% TFA/DCM, 0.4M PhSH; (ix) Boc-Py-OBt, DIEA, DMF; (x) 80% TFA/DCM, 0.4M PhSH; (xi) Boc-**X**-OH (HBTU, DIEA, DMF); (xii) 80% TFA/DCM, 0.4M PhSH; (xiii) Boc-Im-OH (HBTU, DIEA, DMF); (xiv) Im-$COCCl_3$ (DIEA, DMF); (xv) 80% TFA/DCM, 0.4M PhSH; (xvi) Boc-**X**-OH (HBTU, DIEA, DMF); (xvii) 80% TFA/DCM, 0.4M PhSH; (xviii) Im-$COCCl_3$ (DIEA, DMF); (xix) Im-$COCCl_3$ (DIEA, DMF); (xx) (N,N-dimethylamino)propylamine, 85 °C; (xxi) 80% TFA/DCM, 0.4 M PhSH; (xxii) Boc-Py-OBt, DIEA, DMF; (xxiii) 80% TFA/DCM, 0.4 M PhSH; (xxiv) Boc-β-Im-OH [19] (HBTU, DIEA, DMF); (xxv) 80% TFA/DCM, 0.4 M PhSH; (xxvi) Boc-**X**-OH (HBTU, DIEA, DMF); (xxvii) 80% TFA/DCM, 0.4 M PhSH; (xxviii) Boc-**X**-OH (HBTU, DIEA, DMF); (xxix) elongation and termination according to standard procedures;[19] (xxx) cleave from resin using (N,N-dimethylamino)propylamine, 85 °C.

Hairpin and 1:1 motif polyamides were synthesized manually from Boc-β-PAM resin in a stepwise fashion using Boc-protected monomeric and dimeric amino acids (**Figure 2.11**) according to solid-phase protocols.[20] Polyamides containing 3-methoxythiophene (Mt) were deprotected by treatment with sodium thiophenoxide in DMF (100 °C, 2 h) to provide the **Ht** analogues after HPLC purification.

2.3 DNA Affinity and Sequence Specificity in the Hairpin Motif.

Quantitative DNase I footprinting titrations[24] were carried out for the following polyamides on the 278 bp PCR product of plasmid pDHN1[11]: Im-Im-**Im**-Py-γ-Im-**Py**-Py-Py-β-Dp (**Im/Py** pair, **42**), Im-Im-**Nh**-Py-γ-Im-**Py**-Py-Py-β-Dp (**Nh/Py** pair, **43**), Im-Im-**Tn**-Py-γ-Im-**Py**-Py-Py-β-Dp (**Th/Py** pair, **44**), Im-Im-**Tp**-Py-γ-Im-**Py**-Py-Py-β-Dp (**Tp/Py** pair, **45**), Im-Im-**Ht**-Py-γ-Im-**Py**-Py-Py-β-Dp (**Ht/Py** pair, **46**) and Im-Im-**Fr**-Py-γ-Im-**Py**-Py-Py-β-Dp (**Fr/Py** pair, **47**). The DNA sequence specificity of each polyamide at a single ring pairing position (bolded in the sequences listed above) was determined by varying a single DNA base pair within the parent sequence context, 5'-T G G **X** C A-3', to all four Watson-Crick base pairs (**X** = A, T, G, C) and comparing the relative affinities of the resulting complexes (**Figure 2.12**). The variable base pair position was installed opposite the novel heterocycle/pyrrole pairing in question, according to previously reported specificity studies on eight-ring hairpin polyamide.[11]

Equilibrium association constants (K_a) for eight-ring polyamides containing **Hp/Py**, **Py/Py**, **Pz/Py**, and **Th/Py**, pairings against the four DNA sites used in this study have been reported[11,17] and are included in **Table 1** for comparison with values presented

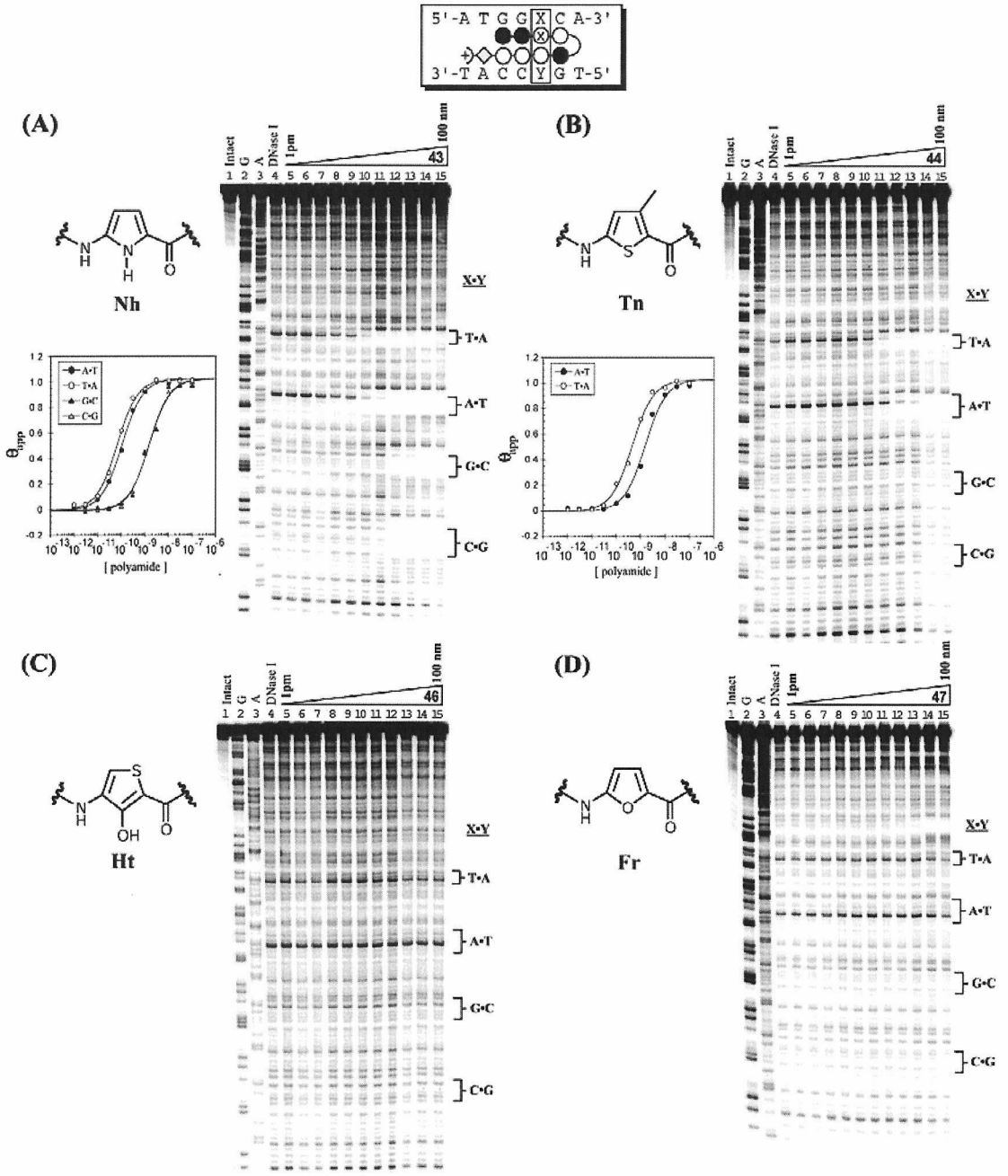


Figure 2.12 (A-D) Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **43**, **44**, **46** and **47**, respectively, on the 278 bp, 5'-end-labelled PCR product of plasmid DHN1: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (left, top) Chemical structure of the residue of interest; and for A and B (bottom left) Binding isotherms for the four designed sites. θ_{norm} values were obtained according to published methods.[23] A binding model for the hairpin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; the γ-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; and the novel heterocycles are indicated by a circle containing an X.

here. As expected, polyamide **42** (**Im/Py** pair) exhibited single site specificity (5'-TGG**X**CA-3', **X** = **G**) at modest affinity (K_a = 4.5 x 10^8 M^{-1}) with fourteen-fold preference over **X** = **C** and at least 400-fold preference over **X** = A, T. Polyamide **43** (**Nh/Py** pair) bound with high affinity to the **X** = A, T sites (K_a ~ 10^{10} M^{-1}) in preference to **X** = G, C by about ten-fold. Hairpin **44** (**Tn/Py** pair) bound with high affinity for X = T, A (K_a ~ 10^9 M^{-1}), a 3-fold preference for T•A > A•T, and 800-fold preference over the **X** = G, C sites. The thiophene analogue, **Tp**, lacking the methyl group was examined to probe possible effects on DNA binding caused by the 3-methyl group. The **Tp/Py** pair (**45**) was found to display similar recognition properties as **Tn/Py**. Remarkably, hairpins containing **Ht/Py** and **Fr/Py** pairs (**46** and **47**, respectively) demonstrated no binding to the designed sites at concentrations up to 1 μM.

Table 1. Hairpin Motif, K_a (M^{-1})[a,b]

Pair	A•T	T•A	G•C	C•G
Im/Py (42)[c]	≤ 10^6	≤ 10^6	4.5 (±0.7) x 10^8	3.2 (±0.5) x 10^7
Hp/Py	8.1 (±1.9) x 10^7	1.6 (±0.3) x 10^9	5.5 (±1.5) x 10^7	7.9 (±2.1) x 10^7
Py/Py	4.7 (±0.7) x 10^9	3.1 (±0.4) x 10^9	2.2 (±0.6) x 10^8	2.5 (±0.9) x 10^8
Nh/Py (43)	8.5 (±0.3) x 10^9	1.1 (±0.1) x 10^{10}	9.2 (±0.1) x 10^8	8.2 (±0.4) x 10^8
Pz/Py	1.0 (±0.5) x 10^9	2.0 (±0.3) x 10^9	≤ 2 x 10^7	≤ 2 x 10^7
Th/Py	≤ 2 x 10^7	≤ 2 x 10^7	≤ 2 x 10^7	≤ 2 x 10^7
Tn/Py (44)	8.0 (±0.4) x 10^8	2.7 (±0.2) x 10^9	≤ 10^6	≤ 10^6
Tp/Py (45)	3.8 (±0.5) x 10^8	1.0 (±0.3) x 10^9	≤ 10^6	≤ 10^6
Ht/Py (46)	≤ 10^6	≤ 10^6	≤ 10^6	≤ 10^6
Fr/Py (47)	≤ 10^6	≤ 10^6	≤ 10^6	≤ 10^6

[a] Values reported are the mean values from at least three DNase I footprint titration experiments, with the standard deviation given in parentheses. [b] Assays were performed at 22 °C in a buffer of 10 mM Tris•HCl, 10 mM KCl, 10 mM, $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0. [c] The number in parentheses indicates the compound containing the unique pairing.

2.4 DNA Affinity and Sequence Specificity in the 1:1 Motif.

Quantitative DNase I footprinting titrations were carried out for the following polyamides on the 298 bp PCR product of pAU8[18]: Im-β-Im-Py-β-**Py**-β-Im-Py-β-Dp (**8**), Im-β-Im-Py-β-**Nh**-β-Im-Py-β-Dp (**9**), Im-β-Im-Py-β-**Hp**-β-Im-Py-β-Dp (**10**), Im-β-Im-Py-β-**Ht**-β-Im-Py-β-Dp (**11**), Im-β-Im-Py-β-**Fr**-β-Im-Py-β-Dp (**12**), Im-β-Im-Py-β-**Nt**-β-Im-Py-β-Dp (**13**), Im-β-Im-Py-β-**Tn**-β-Im-Py-β-Dp (**14**), and Im-β-Im-Py-β-**Th**-β-Im-Py-β-Dp (**15**). The sequence specificity of each polyamide at a single carboxamide position (bolded in the sequences listed above) was determined by varying a single base pair within the parent DNA sequence context, 5'-AAAGA**X**AAGAG-3', to all four Watson-Crick base pairs (**X** = A, T, G, C) and comparing the relative affinities of the resulting complexes (**Figures 2.13** and **2.14**). The variable base pair position was installed opposite the novel heterocycle in question, according to previously described specificity studies on 1:1 polyamide:DNA complexes.[17]

Equilibrium association constants (K_a) for 1:1 polyamides containing **Im**, **Py**, and **Hp** residues tested against the four Watson-Crick base pairs have been reported. However, in that study only the **Im** specificity experiment was performed at the central residue, as with the new polyamides reported here. Therefore, new polyamides containing **Py** and **Hp** residues at the central position have been included in the study for a more controlled comparison. Polyamide **48** (**Py**) demonstrated very high affinity (K_a ~ 6 x 10^{10} M^{-1}) at the **X** = A, T sites (5'-AAAGA**X**AAGAG-3') with a 5- to 10-fold preference over **X** = G, C (**Table 2**). Polyamide **49** (**Hp**) bound with lower affinity (K_a ~ 3 x10^9 M^{-1}) but with similar specificity to **48**, preferring **X** = A, T > G, C by 5- to 10-fold. The **Nh**-containing polyamide (**50**) bound with very high affinity to the **X** = A, T

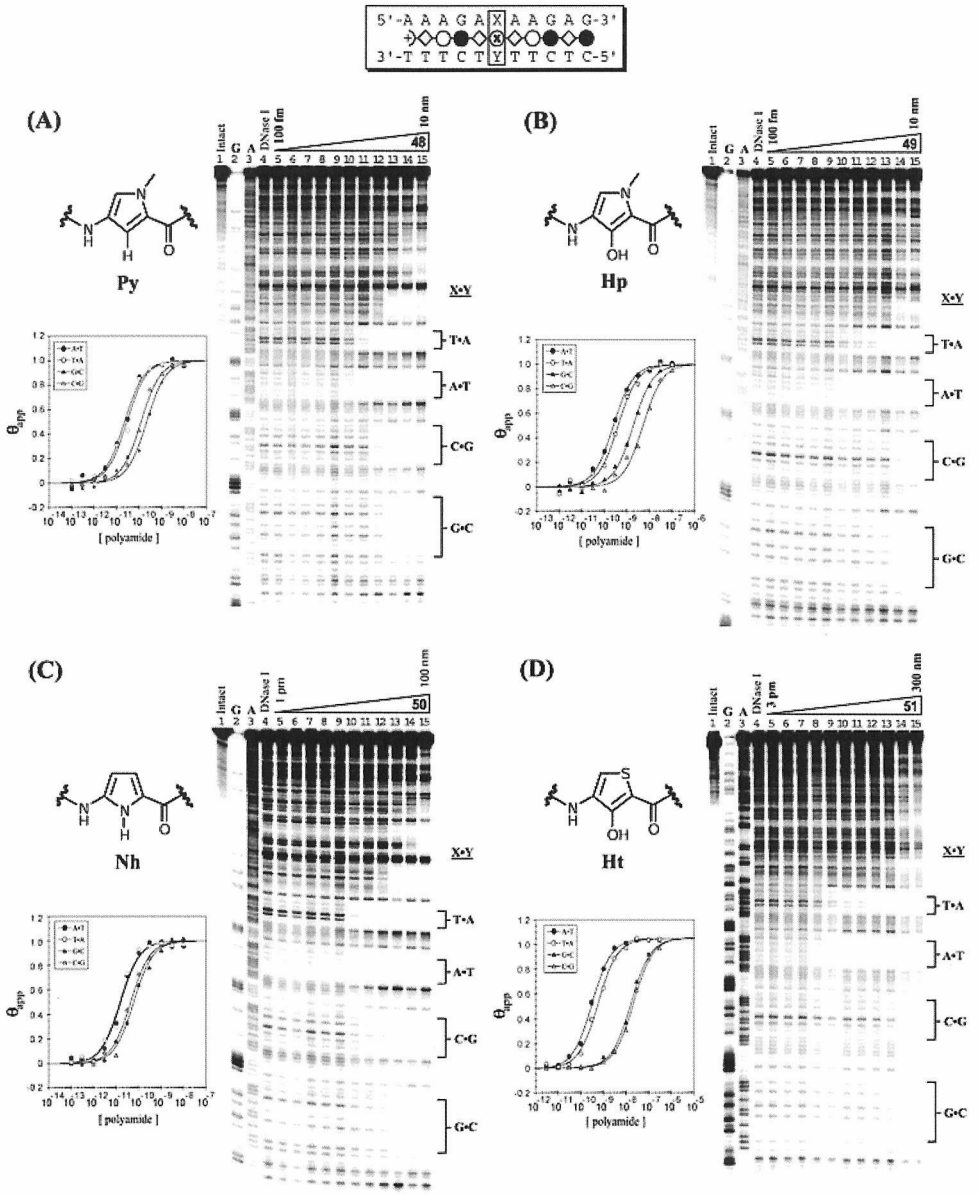


Figure 2.13 (A-D) Quantitative DNase I footprint titration experiments for polyamides **48-51**, respectively, on the 298 bp, 5'-end-labelled PCR product of plasmid pAU8: (A and B) lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM polyamide, respectively. (C) lane 1, intact DNA; lane 2, G reaction; lane 3 A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide respectively. (D) lane 1, intact DNA; lane 2, G reaction; lane 3 A reaction; lane 4, DNase I standard; lanes 5-15, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM, 300 nM polyamide respectively. Each footprinting gel is accompanied by the following: (left, top) chemical structure of the residue of interest; and (left bottom) binding isotherm for the four designed sites. θ_{norm} values were obtained according to published methods [23]. A binding model for the 1:1 motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; and the novel heterocycles are indicated by a circle containing an X.

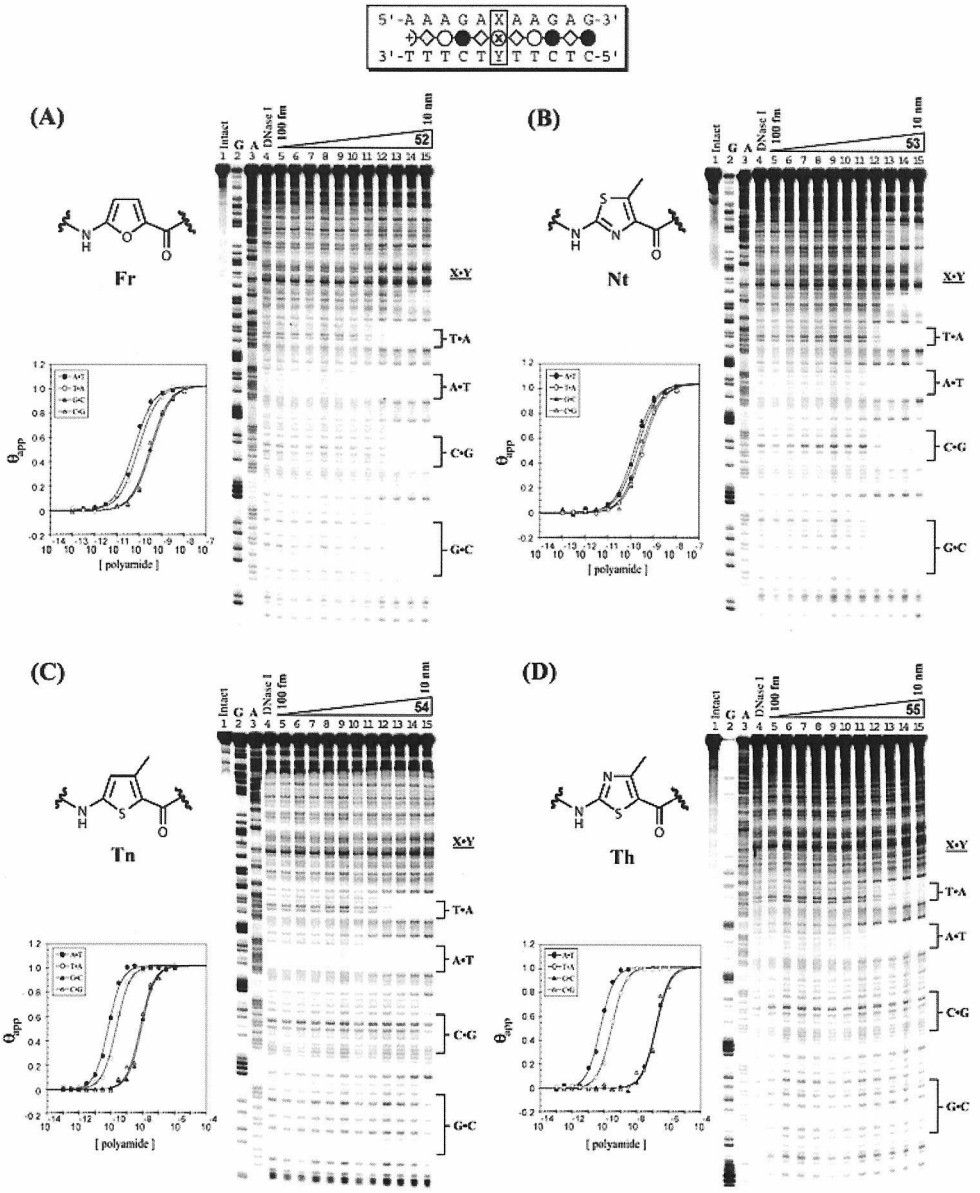


Figure 2.14 (A-D) Quantitative DNase I footprinting experiments for polyamides **52-55**, respectively, on the 298 bp, 5'-end-labelled PCR product of plasmid pAU8: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM respectively. Each footprinting gel is accompanied by the following: (left, top) Chemical structure of the residue of interest; and (left bottom) Binding isotherms for the four designed sites. Isotherms for C and D were generated from gels run out to a final concentration of 1 uM (not shown). θ_{norm} values were obtained according to published methods [23]. A binding model for the 1:1 motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; and the novel heterocycles are indicated by a circle containing an X.

sites (K_a = 7.5 x 10^{10} M^{-1}) but with a mere 3- to 5-fold selectivity over the high-affinity **X** = G, C sites. Polyamide **51** (**Ht**) bound with subnanomolar affinities to the **X** = A, T sites, similar to **49** but with ≥ 40-fold specificity for **X** = A, T > G, C. Polyamide **52** (**Fr**) showed high affinity for the **X** = A, T sites (K_a ~ 10^{10} M^{-1}) with a small 2- to 4-fold preference over **X** = **G**, **C**. The 5-methylthiazole-containing polyamide (**53**, **Nt**), which places the thiazole ring *nitrogen* into the floor of the minor groove, bound all four sites with similar high affinities (K_a ~ 5 x 10^9 M^{-1}). Thiophene-containing polyamide **54** (**Tn**) showed specificity for A, T versus G, C and a preference for a single A•T site. 4-Methylthiazole-containing polyamide (**55**, **Th**), which places the thiazole ring sulfur into the floor of the minor groove, bound with similar **X** = A, T affinity as **54** (**Tn**) but with > 400-fold preference over **X** = G, C. In all cases, binding isotherms fit well to an n = 1 Hill equation, supporting a 1:1 polyamide:DNA stoichiometry (**Figures 2.12, 2.13** and **2.14**).

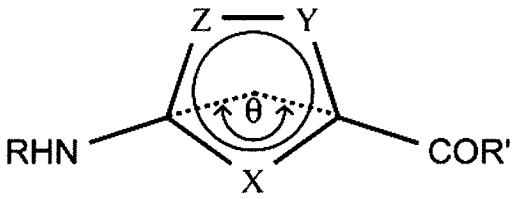
Table 2. 1:1 Motif, K_a (M^{-1})[a,b]

Ring	A•T	T•A	G•C	C•G
Im	2.5 (±0.2) x 10^{10}	1.1 (±0.1) x 10^{10}	2.6 (±0.4) x 10^{10}	1.3 (±0.3) x 10^{10}
Py (48)[c]	7.2 (±0.3) x 10^{10}	5.3 (±0.1) x 10^{10}	3.2 (±0.4) x 10^9	9.4 (±0.2) x 10^9
Hp (49)	3.9 (±0.1) x 10^9	2.5 (±0.3) x 10^9	5.3 (±0.5) x 10^8	1.9 (±0.5) x 10^8
Nh (50)	7.5 (±0.2) x 10^{10}	7.4 (±0.1) x 10^{10}	1.6 (±0.2) x 10^{10}	2.3 (±0.1) x 10^{10}
Ht (51)	2.8 (±0.5) x 10^9	1.6 (±0.6) x 10^9	3.8 (± 1.3) x 10^7	3.7 (±0.7) x 10^7
Fr (52)	2.2 (±0.5) x 10^{10}	1.0 (± 1.3) x 10^{10}	4.4 (±0.5) x 10^9	5.0 (±0.5) x 10^9
Nt (53)	5.4 (±0.9) x 10^9	2.9 (±0.6) x 10^9	8.0 (± 1.3) x 10^9	4.2 (±0.6) x 10^9
Tn (54)	3.0 (±0.2) x 10^{10}	5.7 (±0.4) x 10^9	8.1 (±0.4) x 10^7	8.3 (±0.2) x 10^7
Th (55)	1.5 (±0.2) x 10^{10}	3.0 (±0.7) x 10^9	6.5 (±0.5) x 10^6	7.4 (±0.5) x 10^6

[a] Values reported are the mean values from at least three DNase I footprint titration experiments, with the standard deviation given in parentheses. [b] Assays were performed at 22 °C in a buffer of 10 mM Tris•HCl, 10 mM KCl, 10 mM, $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0. [c] The number in parentheses indicates the compound containing the unique pairing.

2.5 Molecular Modeling Calculations.

Modeling calculations were preformed using *Spartan Essential* software package.[25] Each ring was first minimized using an AM1 model, followed by *Ab initio* calculations using the Hartree-Fock model and a 6-31G* polarization basis set. Each heterocycle exhibited a unique geometric and electronic profile (**Figure 2.15**). Bonding geometry for imidazole, pyrrole, and 3-hydroxypyrrole were in excellent agreement with coordinates derived from x-ray structures of polyamides containing these heterocycles.[7,9] The overall curvature of each monomer was calculated to be the sweep angle (θ) created by the theoretical intersection of the two ring-to-amide bonds in each ring. The structures were ranked by increasing θ as follows: **Fr > Nt > Ht > Nh > Im > Py > Hp > Tn > Pz > Tp > Th**. The ring atom in closest proximity to the floor of the DNA minor groove was



Ring	X	Y	Z	θ (*degrees*)	Charge on X (*e*)
Fr	O	C-H	C-H	*126*	*-0.31*
Nt	N	C-Me	S	*127*	*-0.60*
Ht	O-H	S	C-H	*133*	*+0.40*
Nh	N-H	C-H	C-H	*136*	*+0.34*
Im	N	N-Me	C-H	*137*	*-0.71*
Py	C-H	N-Me	C-H	*146*	*+0.21*
Hp	O-H	N-Me	C-H	*148*	*+0.50*
Tn	S	C-Me	C-H	*149*	*-0.21*
Pz	C-H	N-Me	N	*151*	*+0.23*
Dt	S	C-H	C-H	*152*	*-0.21*
Th	S	C-Me	N	*153*	*-0.25*

Figure 2.15 Geometric and electrostatic profiles for eleven heterocyclic amino acids, derived from *ab initio* molecular modeling calculations using *Spartan Essential* software.[24] (Top) Schematic illustrating the amide-ring-amide angle of curvature, θ. X, Y, and Z denote variable functionality at the different ring positions for each heterocycle. (Bottom) Table listing the functional groups at X, Y, and Z, along with the angle θ, and the electrostatic partial charge on X. For Ht, Nh, Py, Hp, and Pz the positive charge on X is listed for the H atom.

examined for partial charge. The structures were ranked by decreasing partial charge on this atom as follows: **Hp > Ht > Nh > Pz > Py > Tn = Tp > Th > Fr > Im**. Four-ring-subunits containing the sequence Im-Im-**X**-Py (**X** = **Py**, **Pz**, **Nh**, **Im**, **Fr**, **Hp**, **Ht**, **Th**, **Tn** and **Tp**) were constructed and subjected to AM1 and *ab initio* calculations as described above in order to examine overall subunit curvature and planarity (**Figures 2.16** and **2.17**).

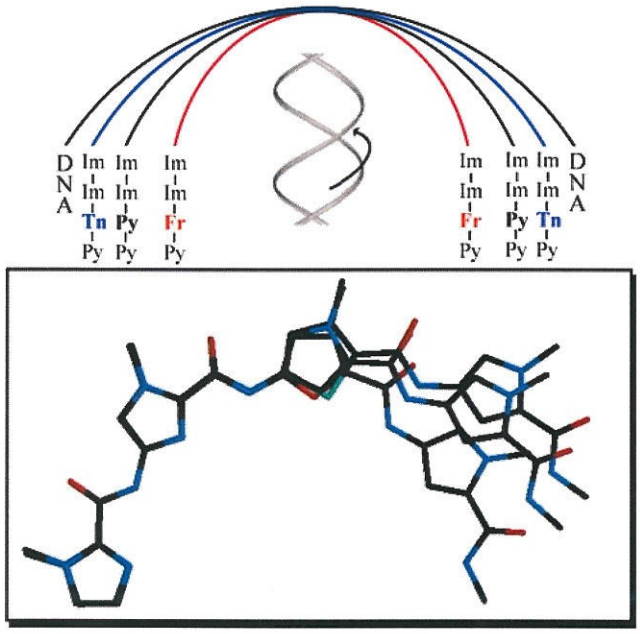


Figure 2.16 (Top) Schematic illustrating the curvatures of four-ring polyamide subunits containing Tn, Py and Fu heterocycles with respect to one another and the DNA helix. (Box) *Ab inito* models of polyamide subunits (Im-Im-**X**-Py, **X** = Tn, Py and Fu) superimposed to demonstrate the significant difference in curvature resulting from atomic substitution. Hydrogens are not shown.

2.6 Discussion.

Here we explore the effects of varying single atom positions in five-membered aromatic heterocycles on the ability of polyamides to discriminate the four Watson-Crick base pairs in the minor groove of DNA. In this experimental design the incremental scheme of DNA and polyamide sequence allows for the comparison of binding affinities

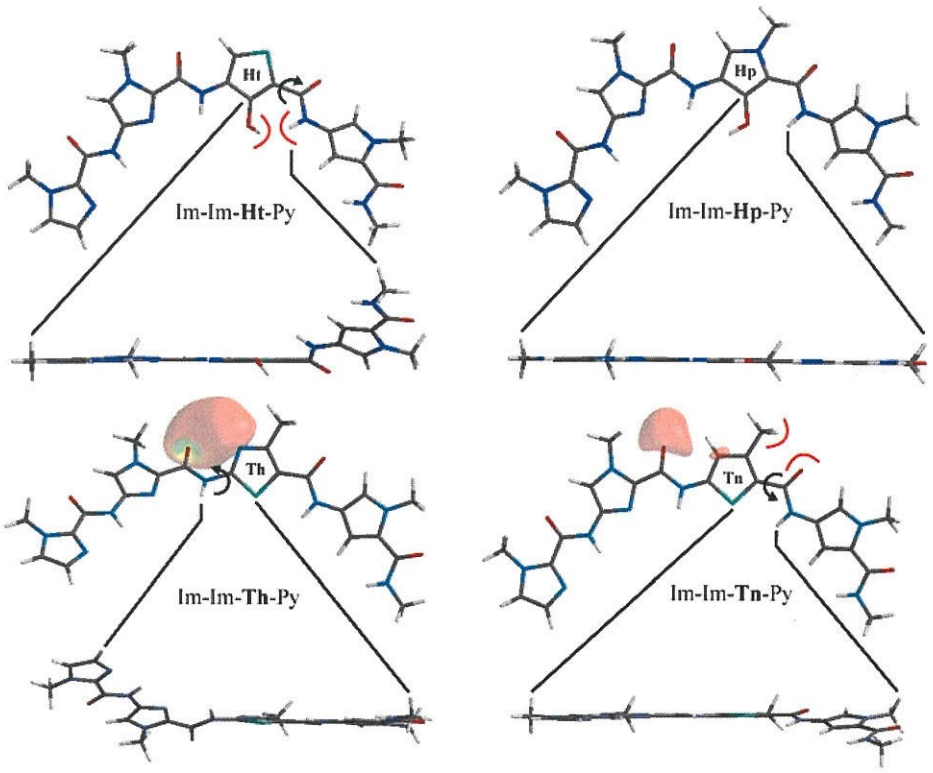


Figure 2.17 *Ab initio* models of four-ring polyamide subunits (Im-Im-**X**-Py, **X** = Ht, Hp, Th, and Tn): (Top Left) The Im-Im-**Ht**-Py subunit demonstrating a dihedral (shown as a curved arrow) created at the ring-carboxamide juncture due to destabilizing eclipsing interaction (shown as red arcs) between the C3-OH hydroxyl group and proximal carboxamide proton; (Top Right) The Im-Im-**Hp**-Py subunit demonstrating co-planarity of the contiguous ring system; (Bottom Left) The Im-Im-**Th**-Py subunit showing the negative isopotential surface (lone pair density) in red, and the dihedral (shown as a curved arrow) resulting from lone pair repulsion interaction between the thiazole nitrogen and the proximal carbonyl; and (Bottom Right) The Im-Im-**Tn**-Py subunit showing the negative isopotential surface (lone pair density) in red, and the small dihedral (shown as a curved arrow) resulting from a destabilizing eclipsing interaction (shown as red arcs) between the thiophene methyl group and the proximal carboxamide. Atomic substitution of N to C-H for **Th** to **Tn** removes lone pair repulsion interaction between the ring nitrogen and the proximal carbonyl.

for a {4} x {10} array of complexes containing unique combinations of {Watson-Crick base pair} x {five membered aromatic heterocycle} at a single position (**Tables 1 and 2**). This quantitative analysis combined with computational modeling of the different heterocycles has led to insight into the etiology of DNA sequence discrimination by polyamides.

The success of **Im/Py** and **Hp/Py** pairs at discriminating between the four Watson-Crick base pairs has been attributed to the shape of the functional groups directed

toward the floor of the minor groove. Therefore, the heterocycles discussed here will be divided into groups based on the types of "bumps and holes" presented to the complementary bumps and holes created by each Watson-Crick base pair on the floor of the DNA minor groove. For the sake of clarity and brevity, hairpin polyamides will be referred to in bold as their unique amino acid pair, *e.g.*, **Tn/Py** for polyamide ImIm**Tn**Py-γ-Im**Py**PyPy-β-Dp (**44**). 1:1 polyamides will be referred to in bold as their unique heterocycle, *e.g.*, **Tn** for polyamide Im-β-ImPy-β-**Tn**-β-ImPy-β-Dp (**54**). DNA sequences will be identified as the variable base position within each motif, *e.g.*, **X** = G in the hairpin motif for 5'-TGG**X**CA-3'.

The Pairing Rules. Discrimination of the four Watson-Crick base pairs using *unsymmetrical* cofacial pairs of aromatic amino acids has proven to be a key insight for minor groove DNA recognition. The four-ring contiguous subunits described here (e.g. ImPyPyPy) are π-conjugated, which limits their conformational flexibility. Consequently, small changes in individual ring curvature can cause greater effects on overall oligomer curvature.

Py/Py, **Pz/Py**, and **Nh/Py** present a hydrogen atom with positive potential to the minor groove floor. As reported previously, **Py/Py** exhibits preference for the **X** = A, T binding sites with > 10-fold selectivity for **X** = A, T > G, C. **Pz/Py** behave with similar affinity as **Py/Py**, but greater discrimination for A, T vs. G, C. **Nh/Py** is similar to **Py/Py** and binds all sites with higher affinity than **Py/Py** and **Pz/Py**. The A•T/T•A preference for the pairs **Py/Py**, **Pz/Py**, and **Nh/Py** originate from a steric interaction with G-NH2, yet there are subtle differences among these compounds which undoubtedly derive from

their slightly different shape. **Py** is known to be over-curved with respect to the DNA helix.[6,8,27-29] The calculations described here provide an amide-ring-amide intersection angle (θ in *Fig. 10*) of 146° and a partial positive charge on the C3-H atom of +0.21. By contrast, **Pz** is somewhat less curved than **Py** (θ = 151°) with similar charge, and **Nh** is considerably more curved (θ = 136°) with a greater charge of +0.34. The reduced curvature of **Pz/Py** should make it more complementary to the DNA curvature, and therefore the steric C4-H to G-NH2 clash would be exacerbated, resulting in lower affinity for the **X** = G, C sites. **Nh/Py** is more curved and more charged, thereby reducing negative steric effects and increasing binding affinity.

Im/Py and **Fr/Py** present a nitrogen or oxygen with sp^2 lone pair electrons directed toward the minor groove floor. Their DNA recognition behavior is strikingly different. **Im/Py** preferentially targets **X** = G. On the other hand, **Fr/Py** shows a complete loss of DNA-binding affinity. Based on the established principles for G•C recognition by an **Im/Py** pair, it was not unreasonable to expect that the **Fr/Py** pair could be a positive recognition element for G•C as well, with the **Fr** oxygen acting as a hydrogen bond acceptor to G-NH2. However, calculations for **Fr** reveal tight over-curvature, with the amide-ring-amide angle decreased by more than ten degrees with respect to imidazole and more than twenty degrees with respect to **Py**. This property causes a pronounced effect on the entire ImIm**Fr**Py subunit (**Figure 2.16**), such that complex formation is no longer energetically favorable. This view is further reinforced by results in the more flexible 1:1 motif, wherein the **Fr** containing polyamide binds all designed sites with high affinity.

Hp/Py and **Ht/Py** present a hydroxyl group (C3-OH) to the minor groove floor. **Hp/Py** displays a preference for **X** = T that breaks the A•T/T•A degeneracy of **Py/Py** pairs. The **Hp/Py** pair presents hydroxyl opposite T not A. Structural studies reveal the origin of this specificity to reside in the hydrogen bond formed between the 3C-OH of hydroxypyrrole and the O2 of thymine (T-O2), and to shape recognition of the asymmetric cleft in the T•A base pair.[8,29] However, a loss in affinity is typically observed for polyamides containing **Hp/Py** pairs compared to the **Py/Py** pair revealing that there is an energetic penalty for the gain in selectivity. This loss in affinity "trade-off" may be attributed to unfavorable steric interactions between the C3-OH and the minor groove floor or perhaps differential solvation between **Hp/Py** and **Py/Py** in water. The **Ht/Py** pair was designed with the hope to improve upon **Hp/Py** by increasing the ring curvature, thus reducing the steric interaction while maintaining the C3-OH to T-O2 hydrogen bond. Remarkably, DNase I footprinting for **Ht/Py** *reveals a complete loss in binding affinity*. As seen with **Fr/Py**, increased ligand curvature may be responsible for disrupting hairpin binding. *Ab initio* calculations on the ImIm**Ht**Py subunit reveal an unfavorable eclipsing interaction between the 3C-OH group and the proximal carboxamide proton. As shown in **Figure 2.17**, this steric clash may force rotation about the ring-amide bond, which would twist the subunit out of plane. Given the snug fit of stacked hairpin subunits within the DNA minor groove, a large distortion in ligand planarity may not be tolerated. **Ht** in the flexible 1:1 motif binds its target sites with high affinity, which further underscores the pronounced effects of ring geometry within the conformationally constrained hairpin motif.

Th/Py, **Tn/Py**, and **Tp/Py** pairs present a large sulfur atom with an sp^2 lone pair to the minor groove floor. The thiazole analog **Th/Py** has been shown previously to afford poor affinity overall and no discrimination of the Watson-Crick base pairs.[11] It was thought that the thiazole sulfur was too large to be easily accommodated within the closely packed cofacial rings in the hairpin•DNA complex. However, *ab initio* calculations on thiazole carboxamide reveal an unfavorable interaction between the lone pair of the thiazole N3 and the proximal carbonyl oxygen (**Figure 2.17**). Although this interaction exists for pyrazole N2, the effect on thiazole is much greater due to the large sulfur atom forcing the thiazole nitrogen into closer proximity with the carboxamide. Consequently, the polyamide subunit may twist out of plane about the amide-ring bond in order to alleviate electronic strain, as shown in **Figure 2.17**. As with **Ht/Py**, the diminished DNA binding affinity of **Th/Py** may be due to its non-planar conformation. These effects are not observed for **Th** in the 1:1 motif.

The negative electronic interaction and hence non-planar conformation should be alleviated if the nitrogen on the back corner of thiazole were replaced with C-H, as in the case of thiophene (**Tn** and **Tp**). Remarkably, **Tn/Py** binds to the **X** = A and T sites with high affinity, but with no observable binding to **X** = G, C. There is a modest 3-fold preference for T > A. Therefore, the sulfur is *not* too large to be accommodated within the tightly packed hairpin•DNA complex. It appears the large sulfur atom of thiophene prefers to sit opposite T not A for the **Th/Py** pair. This experimental finding is exactly *opposite* the result predicted by Lown and Dickerson from model building.[31] The reduced curvature of thiophene should exacerbate steric effects between the sulfur and the minor groove floor, sterically permissive for A, T recognition but resulting in a >

800-fold loss in binding affinity at the **X** = G, C sites. An *Ab inito* calculation of the ImIm**Tn**Py subunit reveals a likely steric interaction between the thiophene methyl group and the proximal carboxamide oxygen, which may force the subunit to twist slightly out of plane. Although structural data on polyamide•DNA complexes reveals a tolerance to small amide-ring dihedrals, a thiophene without the methyl substituent (**Tp**) and the **Tp/Py** pair was tested as a control. **Tp/Py** exhibited virtually identical DNA recognition behavior as **Tn/Py**.

In retrospect, in our search for a **Hp/Py** replacement, our group had great hopes for the **Ht/Py** to distinguish T•A from A•T. Yet this new **Ht** ring system was a complete disappointment. On the other hand, the earlier negative result with thiazole (**Th/Py**) pair suggested that sulfur on the corner of the ring was not a viable recognition element and that thiophene (**Tn/Py** or **Tp/Py** pairs) were not likely to lead to T•A vs. A•T discrimination. The fact that **Tn** or **Tp** prefers to bind opposite T not A is the most significant unanticipated lead to emerge from the study.

The 1:1 Motif. The 1:1 motif has emerged recently as a way to target certain purine-rich DNA sequences (e.g., GAGAA) with high affinity.[16,17] Although the cofacial pairing of rings in the hairpin motif offers a greater chance to differentiate between Watson-Crick base pairs, the single-subunit•DNA complexes of the 1:1 motif provide a relatively flexible system for the exploration of novel shape selective recognition elements. Due to the conformational freedom imparted by the beta alanine residues, changes in heterocycle geometry do not have such a pronounced impact on the rest of the molecule. Therefore, specificity may be more difficult to achieve in this motif. In fact, all 1:1 polyamides described here bind with high affinity to the **X** = A, T sites but with

varying degrees of **X** = A, T > G, C specificity. Structural studies reveal an important register of amide NH groups with the purine N3 and pyrmidine O2 groups on the floor of the DNA minor groove.[19] Given this alignment as a driving force for DNA recognition in the 1:1 motif, one may view the subtle differences in heterocycle curvature as merely placing the central ring atom (X in **Figure 2.15**) closer to or farther from the DNA. In this view, increasing the ring curvature decreases the polyamide•DNA intimacy, thereby diminishing DNA specificity. The results presented here fit well within this model.

Py and **Nh** present a hydrogen with a positive potential to the minor groove floor. Both compounds exhibit a modest 3- to 5-fold selectivity for **X** = A, T > G, C, but **Nh** binds with higher affinity to all sites. The selectivity is probably due to the unfavorable steric X-H to G-NH2 interaction (X = C3 for **Py** and N1 for **Nh**) postulated for netropsin and supported by recent NMR studies.[1,19] The higher affinity for **Nh** may be attributed to a combination of greater positive charge on N1-H and higher ring curvature, both of which should reduce specificity.

Im, **Fr,** and **Nt** present a small atom with an sp^2 lone pair directed toward the minor groove floor. **Im** has been reported previously,[16,17] binding all sites with high affinity and displaying virtually no discrimination between sites. **Fr** and **Nt** behave quite similarly. It is likely that the small atom (N for **Im** and **Nt** or O for **Fr**) presented to the DNA provides no steric clash with G-NH2, and therefore all sites are bound with similarly high affinity.

Hp and **Ht** present a hydroxyl group to the DNA minor groove. In a different polyamide context **Hp** discriminated between A•T and T•A base pairs in the 1:1 motif.[7,30] In this case **Hp** is flanked on both sides by β-alanine residues and specificity is lost.

This loss may be attributed to a larger degree of conformational freedom afforded to the **Hp** ring by the aliphatic linkers.[17] Nonetheless, both **Hp** and **Ht** exhibit significant **X** = A, T > G, C specificity, as expected from a negative 3C-OH to G-NH2 steric clash. **Ht** is slightly more specific, which may result from the non-planarity of this ligand as discussed above for **Ht/Py** in the hairpin motif.

Tn and **Th** present a sulfur atom with an sp^2 lone pair to the DNA minor groove. These compounds exhibit substantial **X** = A, T > G, C specificity ranging from ≥ 70 to ≥ 2300-fold. This remarkable selectivity may be attributed to the decreased curvature of thiazole and thiophene rings, which forces a more intimate interaction of the large sulfur atom and the minor groove floor. In the case of **X** = G, C, this interaction is very negative, resulting in a dramatic loss in binding affinity. **Th** is more specific than **Tn**, which is probably due to the curvature-induced non-planarity of ImIm**Tn**Py, as discussed above for the hairpin motif. (**Figure 2.15**).

2.7 Conclusions.

Our understanding of the origin of DNA sequence discrimination by polyamides has been improved by combining the tools of quantitative DNase I footprinting and computational molecular modeling to establish a correlation between polyamide structure and DNA sequence specificity (**Tables 3 and 4**). We believe that the footprinting results are best explained by differences in the overall heterocycle structure. Each heterocyclic amino acid has an inherently unique shape, which results in varying degrees of curvature complementarity between the polyamide and the DNA minor groove. Given that pyrrole is over-curved with respect to the DNA helix, reducing heterocycle curvature should

Table 3a. Pairing Specificity for Lone Pair to Floor

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

Table 3b. Pairing Specificity for C-H or N-H to Floor

Pair	A•T	T•A	G•C	C•G
Py/Py	+	+	-	-
Nh/Py	+	+	-	-
Pz/Py	+	+	-	-

Table 3c. Pairing Specificity for C-OH to Floor

Pair	A•T	T•A	G•C	C•G
Hp/Py	-	+	-	-
Ht/Py	-	-	-	-

Table 3d. Pairing Specificity for Sulfur to Floor

Pair	A•T	T•A	G•C	C•G
Th/Py	-	-	-	-
Tn/Py	+	+	-	-
Tp/Py	+	+	-	-

increase the polyamide-DNA fit. Consequently, the polyamide would have greater sensitivity to changes in DNA structure and therefore greater DNA sequence selectivity. On the other hand, increasing heterocycle curvature should decrease sensitivity to changes in DNA sequence. In addition, over-curvature can induce ligand non-planarity deriving from destabilizing eclipsing interactions. These results suggest that merely considering the functional group facing the minor groove floor is insufficient for an accurate prediction of DNA recognition behavior.

Curvature effects are amplified in contiguous ring polyamides, where continuous π-conjugation limits conformational flexibility. Furthermore, the packing of cofacial polyamide subunits in the minor groove, as with the hairpin motif, provides additional

Table 4. Specificity in the 1:1 Motif

Pair	A•T	T•A	G•C	C•G
Im	+	+	+	+
Py	+	+	-	-
Hp	+	+	-	-
Nh	+	+	+	+
Ht	+	+	-	-
Fr	+	+	+	+
Nt	+	+	+	+
Tn	+	+	-	-
Th	+	+	-	-

preorganization. The dense functional array offered to the DNA by ring pairs affords greater promise for sequence discrimination. By contrast, the 1:1 motif offers a high degree of flexibility but a less dense functional array and therefore a lower capacity for DNA sequence selectivity.

In retrospect, it is remarkable that a library of eight five-membered heterocycles (and hence, new heterocycle pairs) reveals very little new leads for sequence discrimination. This implies that the solution to DNA recognition by **Im/Py**, **Py/Im**, **Hp/Py** and **Py/Hp** could be a narrow structural window. Perhaps the most novel and useful lead is the possibility that **Tn/Py** pair will distinguish T•A from A•T. The next step is to test multiple **Tn/Py** pairs in different A, T rich contexts to validate whether this

is a potential breakthrough or not. This study is underway and results will be reported in due course.

2.8 Experimental

General. – N,N-dimethylformanide (DMF), N,N-diisopropylethylamine (DIEA), thiophenol (PhSH), N,N-diethylamine, N,N-dimethylaminopropylamine (Dp), Triethylamine (TEA), methyl 2-furoate, ketobutyric acid, methyl acetoacetate, cyanoacetic acid, trichloroacetyl chloride, pyrrole, sodium metal, methylthioglycolate, methyl-2-chloroacrylate, tin(II) chloride dihydrate, and thiourea were purchased from Aldrich. Boc-β-alanine-(4-carbonylaminomethyl)-benzyl-ester-copoly(styrene-divinylbenzene)resin (Boc-β-Pam-resin), dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBt), 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), N,N-dimethylaminopyridine (DMAP), and Boc-β-alanine were purchased from NOVA Biochem. Trifluoroacetic acid (TFA) was purchased from Halocarbon. All other solvents were reagent grade from EM. Oligonucleotide inserts were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology. Glycogen (20 mg/mL), dNTPs (PCR nucleotide mix), and all enzymes, unless otherwise stated, were purchased from Boehringer-Mannheim. pUC19 was purchased from New England Biolabs, and deoxyadenosine [γ-^{32}P]triphosphate was provided by ICN. Calf thymus DNA (sonicated, deproteinized) and DNaseI (7500 units/mL, FPLC pure) were from Amersham Pharmacia. AmpliTaq DNA polymerase was from Perkin-Elmer and used with the provided buffers. Tris.HCl, DTT, RNase-free water, and 0.5 M EDTA were from United States Biochemical. Calcium chloride, potassium chloride, and magnesium chloride were purchased from Fluka. Tris-borate-EDTA was from GIBCO and bromophenol blue was from Acros. All reagents were used without further purification.

NMR spectra were recorded on a Varian spectrometer at 300 MHz in DMSO-*d6* or $CDCl_3$ with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. High resolution FAB and EI mass spectra were recorded at the Mass Spectroscopy Laboratory at the University of California, Los Angeles. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) was conducted at the Protein and Peptide Microanalytical Facility at the California Institute of Technology.

Monomer and Dimer Synthesis. – *5-Amino-furan-2-carboxylic acid methyl ester* (NH_2-Fr-OMe, **1**). Methyl 5-nitro-2-furoate (**18**) was prepared by published methods[20] in 84% yield, TLC (5:2 hexanes/ethyl acetate) R_f 0.7; ^{1}H NMR (DMSO-d_6) δ 7.78 (d, J = 3.9 Hz, 1H), 7.59 (d, J = 3.9 Hz, 1H), 3.38 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 157.7, 152.8, 144.5, 120.4, 113.6, 53.5; EI-MS m/e 171.117 (M^+ calcd for 171.117 $C_6H_5NO_5$). A mixture of Methyl 5-nitro-2-furoate (**18**) (3 g, 17.5 mmol) and 10% palladium on carbon (0.3 g) in ethyl acetate (25 mL) was placed into a parr apparatus and hydrogenated at 500 psi and ambient temperature. The reaction was determined to be complete by TLC after 1.5 h. The mixture was filtered over a 1'' pad of celite to remove palladium on carbon. The filtrate was cooled to -20 °C and hexanes were added until a white precipitate was formed. The precipitate (2.1 g) was collected by vacuum filtration and washed with diethyl ether to give **1** in 85% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.25; IR (Thin Film) 3395, 3322, 1684, 1627, 1528, 1441, 1338, 1297, 1199, 1153 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 7.13 (d, J = 3.6 Hz, 1H), 6.60 (s, 2H), 5.09 (d, J = 3.6 Hz, 1H),

3.66, (s, 3H); ^{13}C NMR ($CDCl_3$) δ 123.8, 123.0, 121.9, 88.3, 86.3; EI-MS m/e 141.042 (M^+ calcd for 141.042 $C_6H_7NO_3$).

Methyl 5-[(1-methyl-4-nitroimidazole-2-yl)carbonylamino]furan-2-carboxylate (NO_2-Im-Fr-OMe, **19**). A mixture of **1** (1g, 7.08 mmol), NO_2-Im-$COCCl_3$ (2.31 g, 8.5 mmol), DIEA (1.1 g, 1.48 mL, 8.5 mmol) and ethyl acetate (14 mL) was stirred at 35 °C for 12 h. The reaction was cooled to ambient temperature and sufficient hexanes were added to completely precipitate a pale yellow solid. The precipitate was collected by vacuum filtration and washed with cold methanol and ether to give **19** (1.18 g) in 57% yield. TLC (5:2 hexanes/ ethyl acetate) R_f 0.20; IR (Thin Film) 3195, 3133, 1731, 1676, 1553, 1541, 1525, 1442, 1387, 1313, 1146 cm^{-1}; ^1H NMR (DMSO-d_6) δ 12.17 (br, 1H), 8.65 (s, 1H), 7.34 (d, J = 3.6 Hz, 1H), 6.52 (d, J = 3.6 Hz, 1H), 4.01 (s, 3H), 3.77 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 158.6, 155.8, 150.0, 145.0, 137.3, 137.0, 127.8, 121.4, 99.0, 52.3, 37.3; EI-MS m/e 294.060 (M^+ calcd for 294.060 $C_{11}H_{10}N_4O_6$).

Methyl 5-[(4-amino-1-methylimidazole-2-yl)carbonylamino]furan-2-carboxylate hydrochloride (HCl•H_2N-Im-Fr-OMe, **20**). A mixture of **19** (1 g, 3.4 mmol) and 10% palladium on carbon (0.2 g) in ethyl acetate (7 mL) was placed into a parr apparatus and hydrogenated at 500 psi and ambient temperature. The reaction was complete after 1.5 h. The mixture was filtered over a 1'' pad of celite to remove the palladium on carbon. The ethyl acetate was removed in vacuo and hydrogen chloride in diethylether (2M) was added to give the hydrochloride salt **20** (0.62 g) in 61% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.15 (amine) R_f 0.0 (salt); IR (Thin Film) 2985, 3008, 1708, 1689, 1537, 1318, 1193, 1142, 1119, 1018, 756, 668 cm^{-1}; ^1H NMR (DMSO-d_6) δ 11.73 (br, 1H), 7.54 (s, 1H), 7.33 (d, J = 3.6, 1H), 6.51 (s, J = 3.6 Hz, 1H), 3.96 (s, 3H), 3.76 (s, 3H); ^{13}C

NMR (DMSO-d_6) δ 157.9, 155.0, 149.5, 136.3, 134.8, 129.5, 120.9, 118.7, 97.9, 51.6, 35.8; EI-MS m/e 264.086 (M^+ calcd for 264.086 $C_{11}H_{12}N_4O_4$).

Methyl 5-({4-[(tert-butoxy)carbonylamino]-1-methylimidazole-2-yl}-carbonylamino)-furan-2-carboxylate (Boc-Im-Fr-OMe, **21**). A mixture of **20** (0.5 g, 1.6 mmol), Boc-anhydride (545 mg, 2.4mmol), and DIEA (258 mg, 348 µl, 2 mmol) in DMF (5 mL) was stirred at 60 °C for 18 h. The reaction mixture was added to ice water (0.5 L) and the precipitate extracted with ethyl acetate (50 mL). The organic layer was dried with sodium sulfate and removed in vacuo. The resulting residue was subjected to column chromatography (1:1 hexanes/ethyl acetate) to provide **21** (376 mg) as a white solid in 62% yield. TLC (1:1 hexanes/ethyl acetate) R_f 0.65; IR (Thin Film) 3243, 2978, 1722, 1577, 1533, 1436, 1368, 1311, 1163, 1138, 755 cm^{-1}; ^1H NMR (DMSO-d_6) δ 11.11 (s, 1H), 9.52 (s, 1H), 7.33 (d, J = 3.3 Hz, 1H), 6.47 (d, J = 3.3 Hz, 1H), 3.90 (s, 3H), 3.76 (s, 3H), 1.44 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 158.6, 156.1, 153.6, 150.4, 137.7, 136.8, 133.1, 121.7, 115.3, 97.8, 79.8, 52.3, 35.9, 28.9; EI-MS m/e 364.138 (M^+ calcd for 364.138 $C_{16}H_{20}N_4O_6$).

5-({4-[(tert-butoxy)carbonylamino]-1-methylimidazole-2-yl}carbonylamino)-furan-2-carboxylic acid (Boc-Im-Fr-OH, **15**). A mixture of **21** (0.3 g, 0.82 mmol), 1N NaOH (5 mL) and methanol (1 mL) was stirred at room temperature. The reaction was determined to be complete by TLC after 3 h. The methanol was removed in vacuo and the aqueous layer carefully adjusted to pH = 2 with 1N HCl. The milky white precipitate was extracted with ethyl acetate and the organics were dried over sodium sulfate. Filtration and evaporation of the organic layer gave **15** (270 mg) as a fine white powder in 94% yield. TLC (1:1 hexanes/ethyl acetate, 10% acetic acid) R_f 0.5; IR (Thin Film)

3231, 2917, 2856, 1688, 1542, 1311, 1259, 1163, 1119 cm^{-1}; ^1H NMR (DMSO-d_6) δ 10.94 (s, 1H), 9.52 (s, 1H), 7.30 (s, 1H), 7.22 (d, J = 3.6 Hz, 1H), 6.44 (d, J = 3.6 Hz, 1H), 3.91 (s, 3H), 1.44 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 160.1, 159.3, 149.9, 136.5, 135.5, 133.2, 123.6, 115.4, 101.4, 97.0, 82.3, 35.9, 28.9; EI-MS m/e 350.123 (M^+ calcd for 350.123 $C_{15}H_{18}N_4O_6$.

Methyl 5-{3-[(tert-butoxy)carbonylamino]propanoylamino}furan-2-carboxylate (Boc-β-Fr-OMe, **22**). A mixture of Boc-β-alanine (3.22 g, 17 mmol) and DCC (1.75 g, 8.5 mmol) in methylene chloride (25 mL) was stirred at ambient temperature for 30 min. To the above mixture was added **1** (0.6 g, 4.25 mmol) as a solution in DMF (5 mL) and DIEA (0.741 mL, 0.55 g, 4.25 mmol), followed by the addition of DMAP (0.155 g, 1.27 mmol). The reaction was heated to 40 °C and allowed to stir overnight. The reaction was filtered to remove the DCU and the filtrate poured into ice water (0.5 L) upon which time a crude white precipitate formed. The crude precipitate was extracted with ethyl acetate and subjected to column chromatography (5:2 hexanes/ethyl acetate) to give **22** (1.1 g) as a flaky white solid in 85% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.2; IR (Thin Film) 3372, 3234, 3036, 1957, 1728 cm^{-1}; ^1H NMR (DMSO-d_6) δ 11.58 (s, 2H), 7.29 (d, J = 3.8 Hz, 1H), 6.86 (s, 2H), 6.36 (d, J = 3.8 Hz, 1H), 3.76 (s, 3H), 3.32 (q, J = 6 Hz, 2H), 2.69 (t, J = 6 Hz, 3H), 1.35 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 167.9, 159.4, 155.8, 150.1, 137.0, 120.8, 95.2, 78.3, 51.6, 36.1, 35.8, 29.0; EI-MS m/e 312.132 (M^+ calcd for 312.132 $C_{14}H_{20}N_2O_6$).

5-{3-[(tert-butoxy)carbonylamino]propanoylamino}furan-2-carboxylic acid (Boc-β-Fr-OH, **13**). A mixture of **22** (1.1 g, 3.52 mmol), 1N NaOH (15 mL) and methanol (5 mL) was stirred at ambient temperature. The reaction was determined to be

complete by TLC after 4 h. The methanol was removed in vacuo and the aqueous layer carefully adjusted to pH = 2 with 1N HCl. The milky white precipitate was extracted with ethyl acetate and the organics were dried over sodium sulfate. Filtration and evaporation of the organic layer gave **13** (0.97 g) as an off white solid in 92% yield. TLC (5:2 hexanes/ethyl acetate, 10% acetic acid) R_f 0.6; IR (Thin Film) 3321, 3270, 3979, 1684, 1522 cm^{-1}; ^1H NMR (DMSO-d_6) δ 11.28 (s, 2H), 6.98 (d, J = 3.6 Hz, 1H), 6.82 (s, 2H), 6.23 (d, J = 3.6 Hz, 1H), 3.19 (q, J = 6 Hz, 2H), 2.45 (t, J = 6 Hz, 2H), 1.35 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 168.7, 159.7, 156.0, 150.6, 137.1, 120.8, 95.8, 78.3, 36.9, 36.6, 29.0; EI-MS m/e 298.117 (M^+ calcd for 298.116 $C_{13}H_{18}N_2O_6$).

2,2,2-Trichloro-1-pyrrol-2-ylethan-1-one (**23**). A solution of pyrrole (20.6 mL, 20 g, 298 mmol) and diethyl ether (86 mL) was added dropwise to trichloroacetyl chloride (71.9 mL, 117 g, 644 mmol) with stirring at 0 °C. The reaction was allowed to warm to room temperature and stirred overnight. The solvent was then removed in vacuo, and **23** (24 g) was recovered as a white solid in 38% yield upon precipitation from hexanes. TLC (5:2 hexanes/ethyl acetate) R_f 0.75; IR (Thin Film) 3322, 1656, 1388, 1136, 1035, 953, 842, 808, 754, 733, 688 cm^{-1}; ^1H NMR (DMSO-d_6) δ 12.4 (s, 1H), 7.32 (m, J = 2.1 Hz, 1H), 7.29 (m, J = 2.1 Hz, 1H), 6.34 (m, J = 2.1 Hz, 1H); ^{13}C NMR (DMSO-d_6) δ 172.5, 130.0, 122.3, 121.9, 112.0, 95.9; EI-MS m/e 210.936 (M^+ calcd for 210.936 $C_6H_4Cl_3NO$).

2,2,2-Trichloro-1-(5-nitropyrrol-2-yl)ethan-1-one (**24**). A solution of **23** (20 g, 95 mmol) and acetic anhydride (111 mL) was cooled to -40 °C and treated dropwise with 70% nitric acid (8.24 mL) over 2 h. After completion of addition, the reaction was warmed to room temperature over 2 h. The reaction was cooled back down to -40 °C and

sufficient ice water was added to precipitate **24** (16.5 g) as a white solid in 68% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.6; IR (Thin Film) 3316, 1676, 1551, 1518, 1405, 1379, 1317 cm^{-1}; ^1H NMR (DMSO-d_6) δ 13.62 (s, 1H), 8.33 (d, J = 3 Hz, 1H), 7.66 (d, J = 3 Hz, 1H); ^{13}C NMR (DMSO-d_6) δ 173.4, 137.6, 128.4, 122.0, 115.0, 94.4; EI-MS m/e 255.921 (M^+ calcd for 255.921 $C_6H_3Cl_3N_2O_3$).

Ethyl 5-nitropyrrole-2-carboxylate (NO_2-Nh-OEt, **25**). To a mixture of **24** (10 g, 39 mmol) in ethanol (35 mL) at room temperature was added sodium ethoxide (4 g, 59 mmol). The reaction was stirred for 2 h and quenched with sulfuric acid. The mixture was cooled to 0 °C and ice water was added (0.5 L) to precipitate **25** (7 g) as a tan solid in 97% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.5; IR (Thin Film) 3263, 3152, 2979, 1687, 1565, 1508, 1365, 1323, 1207, 1017, 752 cm^{-1}; ^1H NMR (DMSO-d_6) δ 8.04 (d, J = 1.5 Hz, 1H), 7.23 (d, J = 1.5 Hz, 1H), 4.29 (q, J = 7.2 Hz, 2H), 1.29 (t, J = 7.2 Hz, 3H); ^{13}C NMR (DMSO-d_6) δ 160.0, 137.2, 124.9, 123.6, 110.1, 61.4, 14.9; EI-MS m/e 184.048 (M^+ calcd for 184.048 $C_7H_8N_2O_4$).

Ethyl 5-aminopyrrole-2-carboxylate hydrochloride (HCl•H_2N-Nh-OEt, **2**). A mixture of **25** (3 g, 16 mmol) and 10% palladium on carbon (0.3 g) in ethyl acetate (25 mL) was placed into a parr apparatus and hydrogenated at 500 psi and ambient temperature for 1.5 h. The mixture was filtered over a 1'' pad of celite to remove palladium on carbon. The filtrate was cooled to 20 °C and HCl in diethyl ether was added. Upon addition, the hydrochloride salt precipitated out and was collected by vacuum filtration to give **2** (2.4 g) as an off white solid in 78% yield. TLC (1:1 hexanes/ethyl acetate) (amine) R_f 0.15, (hydrochloride) R_f 0.0; IR (Thin Film) 2914, 1694, 1495, 1429, 1376, 1345, 1284, 1224, 1106, 1020, 965 cm^{-1}; ^1H NMR (DMSO-d_6) δ

12.21 (s, 1H), 10.13 (s, 1H), 7.10 (d, J = 1.8 Hz, 1H), 6.74 (d, J = 1.8 Hz, 1H), 4.24 (q, J = 7.2 Hz, 2H), 1.26 (t, J = 7.2 Hz, 3H); ^{13}C NMR (DMSO-d_6) δ 159.6, 121.3, 117.7, 115.6, 109.2, 60.0, 14.3; EI-MS m/e 154.074 (M^+ calcd for 154.074 $C_7H_{10}N_2O_2$).

Ethyl 5-[(tert-butoxy)carbonylamino]pyrrole-2-carboxylate (Boc-Nh-OEt, **26**). A mixture of **2** (2 g, 11 mmole), Boc-anhydride (3.6 g, 16.5 mmol), and DIEA (2.1 mL, 1.56 g, 12.1 mmol) in DMF (15 mL) was stirred at 60 °C for 12 h. The mixture was then added to ice water (1 L) and extracted twice with ethyl acetate (150 mL). The organic layer was dried over sodium sulfate and evaporated to give a crude oil. Column chromatography of the oil (5:2 hexanes/ethyl acetate) afforded **26** (2 g) as a flaky white solid in 72% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.65; IR (Thin Film) 3296, 1683, 1570, 1384, 1315, 1264, 1249 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 11.48 (s, 1H), 9.06 (s, 1H), 6.93 (d, J = 1.8 Hz, 1H), 6.58 (d, J = 1.8 Hz, 1H), 4.21 (q, J = 7.2 Hz, 2H), 1.41 (s, 9H), 1.24 (t, J = 7.2 Hz, 3H); ^{13}C NMR (DMSO-d_6) δ 160.1, 152.5, 119.1, 112.4, 105.2, 78.3, 59.4, 28.2, 14.4; EI-MS m/e 254.127 (M^+ calcd for 254.127 $C_{12}H_{18}N_2O_4$).

5-[(tert-butoxy)carbonylamino]pyrrole-2-carboxylic acid (Boc-Nh-OH, **12**). A mixture of **26** (2 g, 7.9 mmol), 1N NaOH (15 mL) and methanol (5 mL) was stirred at ambient temperature. The reaction was determined to be complete by TLC after 3 h. The methanol was removed in vacuo and the aqueous layer carefully adjusted to pH = 2 with 1N HCl. The milky white precipitate was extracted with ethyl acetate and the organics were dried over sodium sulfate. Filtration and evaporation of the organic layer gave **12** (1.6 g) as an off white solid in 92% yield. TLC (5:2 hexanes/ethyl acetate, 10% acetic acid) R_f 0.5; IR (Thin Film) 3329, 3153, 2969, 1691, 1586, 1549, 1434, 1374, 1250, 1167, 1117, 1057, 961, 762 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 11.32 (s, 1H), 9.01 (s, 1H), 6.88

(s, 1H), 6.52 (s, 1H), 1.41 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 162.3, 153.3, 125.4, 120.7, 112.6, 106.0, 79.0, 28.9; EI-MS m/e 226.095 (M^+ calcd for 226.095 $C_{10}H_{14}N_2O_4$).

Methyl 4-cyano-3-methylbut-3-enoate (**27**). A mixture of acetoacetate (30 g, 258 mmol), cyanoacetic acid (24 g, 284 mmol), ammonium acetate (3.98 g, 51.6 mmol), acetic acid (6.65 mL, 6.98 g, 116 mmol) and benzene (75 mL) was stirred for 12 h at 145 °C in a round bottom equipped with a Dean Stark apparatus and condenser. The reaction was allowed to cool to ambient temperature, washed with brine (0.3 L), saturated sodium bicarbonate (0.3 L) and dried over magnesium sulfate. The reaction was filtered and solvent removed in vacuo. The crude product was distilled (60 °C, 0.1 mmHg) to give **27** (23 g) as a clear liquid and mixture of E and Z regioisomers in 65% yield. IR (Thin Film) 2957, 2221, 1741, 1437 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 5.69 (q, J = 0.6 Hz, 1H), 5.62 (q, J = 0.6 Hz, 1H), 3.61 (s, 3H), 3.60 (s, 3H), 3.42 (s, 2H), 3.35 (d, J = 1.2 Hz, 2H), 2.01 (d, J = 1.2 Hz, 3H), 1.93 (d, J = 1.2 z, 3H); ^{13}C NMR (DMSO-d_6) δ 170.1, 169.5, 158.4, 158.1, 117.4, 117.3, 99.6, 99.4, 52.8, 52.7, 42.8, 41.3, 23.6, 21.7; EI-MS m/e 139.063 (M^+ calcd for 139.063 $C_7H_9NO_2$).

Methyl 5-amino-3-methylthiophene-2-carboxylate (HCl•H_2N-Tn-OMe, **3**). Diethylamine (18.7 mL, 13.2 g, 181 mmol) was added dropwise to a mixture of **27** (23 g, 165 mmol) and sulfur flakes (5.28 g, 165 mmol), in ethanol (130 mL) and stirred at room temperature for 3 h. The reaction was concentrated to a minimal volume in vacuo and placed on an ice bath. Concentrated HCl was slowly added to the mixture to give a light orange solid. The precipitate was collected by vacuum filtration and washed repeatedly with diethyl ether to give **3** (19 g) in 68% yield. TLC (5:2 hexanes/ethyl acetate) (amine) R_f 0.55, (hydrochloride) R_f 0.0; IR (Thin Film) 3422, 3339, 3204, 2849, 1713, 1677,

1546, 1462, 1269, 1187, 1092 cm^{-1}; ^1H NMR (DMSO-d_6) δ 6.91 (s, 2H), 5.76 (s, 1H), 3.61 (s, 3H), 2.62 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 163.5, 146.7, 145.5, 114.9, 114.7, 52.0, 16.6; EI-MS m/e 171.035 (M^+ calcd for 171.035 $C_7H_9NO_2S$).

Methyl 3-methyl-5-[(1-methyl-4-nitroimidazole-2-yl)carbonylamino]thiophene-2-carboxylate (NO_2-Im-Tn-OMe, **28**). Compound **28** was synthesized from **3** (1g, 4.8 mmol) according to the procedure reported for **19**, providing **28** (0.87 g) as a yellow solid in 56% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.60; IR (Thin Film) 3125, 1649, 1543, 1506, 1382, 1312, 1267 cm^{-1}; ^1H NMR (DMSO-d_6) δ 12.43 (s, 1H), 8.66 (s, 1H), 6.95 (s, 1H), 4.03 (s, 3H), 3.79 (s, 3H), 2.41 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 163.5, 155.7, 145.1, 144.7, 143.6, 136.9, 128.0, 118.8, 117.2, 52.2, 37.5, 16.6; EI-MS m/e 324.053 (M^+ calcd for 324.053 $C_{12}H_{12}N_4O_5S$).

Methyl 5-[(4-amino-1-methylimidazol-2-yl)carbonylamino]-3-methylthiophene-2-carboxylate hydrochloride (HCl•H_2N-Im-Tn-OMe, **29**). Compound **29** was synthesized from **28** (0.5 g, 1.5 mmol) according to the procedure reported for **20**, providing **29** (321 mg) as a pale yellow solid in 63% yield. TLC (5:2 hexanes/ethyl acetate) (amine) R_f 0.25, (hydrochloride) R_f 0.0; IR (Thin Film) 3294, 1735, 1674, 1562, 1520, 1440, 1407, 1267, 1185, 1091 cm^{-1}; ^1H NMR (DMSO-d_6) δ 12.14 (s, 1H), 7.47 (s, 1H), 6.95 (s, 1H), 3.99 (s, 1H), 3.73 (s, 1H), 2.39 (s, 1H); ^{13}C NMR (DMSO-d_6) δ 177.9, 63.4, 157.6, 155.1, 144.1, 128.0, 118.6, 116.1, 52.2, 35.4, 16.4; EI-MS m/e 294.079 (M^+ calcd for 294.079 $C_{12}H_{14}N_4O_3S$).

Methyl 5-({4-[(tert-butoxy)carbonylamino]-1-methylimidazole-2-yl}-carbonylamino)-3-methylthiophene-2-carboxylate (Boc-Im-Tn-OMe, **30**). Compound **30** was synthesized from **29** (300 mg, 0.91 mmol) according to the procedure reported for

21, providing **30** (239 mg) as a pale yellow solid in 67% yield. TLC (1:1 hexanes/ethyl acetate) R_f 0.7; IR (Thin Film) 3424, 3219, 2995, 1750, 1704, 1677, 1571, 1251, 1141 cm^{-1}; ^1H NMR (DMSO-d_6) δ 11.78 (s, 1H), 9.33 (s, 1H), 7.32 (s, 1H), 6.90 (s, 1H), 3.91 (s, 1H), 3.73 (s, 1H), 2.39 (s, 1H), 1.43 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 162.8, 155.7, 152.8, 144.0, 143.4, 136.7, 132.4, 117.0, 115.6, 115.3, 79.4, 51.3, 35.2, 26.9, 15.9; EI-MS m/e 394.131 (M^+ calcd for 394.131 $C_{17}H_{22}N_4O_5S$).

5-({4-[(tert-butoxy)carbonylamino]-1-methylimidazole-2-yl}carbonylamino)-3-methylthiophene-2-carboxylic acid (Boc-Im-Tn-OH, **16**). A mixture of **30** (200 mg, 0.507 mmol), methanol (1 mL) and 1N NaOH (5 mL) was stirred at 60 °C for 6 h. The methanol was removed in vacuo and the aqueous layer carefully adjusted to pH = 2 with 1N HCl. The milky white precipitate was extracted with ethyl acetate and the organics were dried over sodium sulfate. Filtration and evaporation of the organic layer gave **16** (158 mg) as a pale tan solid in 82% yield. TLC (1:1 hexanes/ethyl acetate, 10% acetic acid) R_f 0.8; (Thin Film) 3400, 2976, 3231, 2961, 1722, 1678, 1589, 1253, 1179, 1091 cm^{-1}; ^1H NMR (DMSO-d_6) δ 11.68 (s, 1H), 9.35 (s, 1H), 7.32 (s, 1H), 6.87 (s 1H), 3.93 (s, 3H), 2.38 (s, 3H), 1.44 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 164.0, 158.3, 155.6, 142.8, 136.7, 117.3, 117.1, 115.2, 99.4, 81.4, 35.2, 28.1, 15.8; EI-MS m/e 380.115 (M^+ calcd for 380.115 $C_{16}H_{20}N_4O_5S$).

Methyl 5-{3-[(tert-butoxy)carbonylamino]propanoylamino}-3-methylthiophene-2-carboxylate (Boc-β-Tn-OMe, **31**). A mixture of Boc-β-alanine (1 g, 5.28 mmol) and DCC (545 mg, 2.64 mmol) in methylene chloride (10 mL) was stirred at ambient temperature for 30 min. The mixture was then filtered into a round bottom containing **3** (382 mg, 1.8 mmol), DIEA (322 μl, 239mg, 1.8 mmol), DMAP (100 mg, 0.8 mmol) and

DMF (8 mL). The mixture was then heated at 45 °C. Progress of the reaction was monitored by TLC with additional symmetrical anhydride (1.4 eq) added every 8 h, as needed, until completion. The mixture was then added to brine (0.2 L) and extracted twice with ethyl acetate (50 mL). The organic layer was then washed with saturated sodium bicarbonate (0.1 L), 10 mM HCl (0.1 L), and dried over sodium sulfate. The crude residue was subjected to column chromatography (5:2 hexanes/ethyl acetate) to give **31** (392 mg) as a flaky white powder in 62% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.32; IR (Thin Film) 3348, 3450, 2981, 1684, 1568, 1522, 1445, 1252 cm^{-1}; ^1H NMR ($CDCl_3$) δ 10.11 (s, 1H), 6.49 (s, 1H), 3.79 (s, 3H), 3.49 (q, J = 6 Hz, 2H), 2.64 (t, J = 6 Hz, 2H), 1.40 (s, 9H); ^{13}C NMR ($CDCl_3$) δ 168.7, 164.1, 157.0, 144.9, 143.4, 128.4, 117.3, 116.1, 80.4, 51.7, 37.0, 34.8, 16.4; EI-MS m/e 342.124 (M^+ calcd for 342.124 $C_{15}H_{22}N_2O_5S$).

5-{3-[(tert-butoxy)carbonylamino]propanoylamino}-3-methylthiophene-2-carboxylic acid (Boc-β-Tn-OH, **14**). A mixture of **31** (200 mg, 0.58 mmol), methanol (4 mL) and 1N NaOH (15 mL) was stirred at 60 °C for 6 h. The methanol was removed in vacuo and the aqueous layer carefully adjusted to pH = 2 with 1N HCl. The milky white precipitate was extracted with ethyl acetate and the organics were dried over sodium sulfate. Filtration and evaporation of the organic layer gave **14** (180 mg) as a light yellow solid in 94% yield. TLC (5:2 hexanes/ethyl acetate, 10% acetic acid) R_f 0.5; (Thin Film) 3255, 2976, 2954, 1674, 1569, 1522, 1445, 1253 cm^{-1}; ^1H NMR (DMSO-d_6) δ 12.35 (s, 1H), 11.42 (s, 1H), 6.89 (s, 1H), 6.48 (s 1H), 3.21 (q, J = 6 Hz, 2H), 2.50 (t, J = 6 Hz, 2H), 2.37 (s, 3H), 1.35 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 168.2, 164.1, 155.4, 143.3,

143.1, 116.8, 115.0, 77.7, 36.3, 35.6, 28.2, 15.7; EI-MS m/e 328.109 (M^+ calcd for 328.109 $C_{14}H_{20}N_2O_5S$).

5-nitrothiophene-2-carboxylic acid (NO_2-Tp-OH, **32**). A mixture of sodium hypochlorite (26.35 g, 291 mmol) and sodium hydrogen phosphate monohydrate (30.3 g, 219 mmol) in water (250 mL) was added dropwise to a solution of commercially available 5-nitrothiophene-2-carboxaldahyde (5g, 31.8 mmol) in acetone (0.6 L) at room temperature. Upon completion of addition, TLC showed total consumption of the starting aldahyde. The reaction was washed with hexanes (0.1 L) and acidified to pH = 2 with 1N HCl. The mixture was extracted three times with diethyl ether (0.1 L) and dried over sodium sulfate. Evacuation of the organic layer gave **32** (3.74 g) as a white solid in 68% yield. TLC (5:2 hexanes/ethyl acetate, 10% acetic acid) R_f 0.55; IR (Thin Film) 3118, 3109, 2876, 1688, 1680, 1512, 1350, 1336, 1274 cm^{-1}; ^1H NMR (DMSO-d_6) δ 8.12 (d, J = 4.2 Hz, 1H), 7.73 (d, J = 4.2 Hz, 1H); ^{13}C NMR (DMSO-d_6) δ 162.2, 154.5, 141.0, 132.6, 130.5; EI-MS m/e 172.978 (M^+ calcd for 172.978 $C_5H_3NO_4S$).

Methyl 5-nitrothiophene-2-carboxylate (NO_2-Tp-OMe, **33**). A mixture of **32** (3.5 g, 20.2 mmol), concentrated sulfuric acid (0.2 g, 110 µl, 2.0 mmol) and methanol (50 mL) was refluxed for 48 h. The methanol was removed in vacuo and the residue neutralized with 1N NaOH. The mixture was extracted twice with ethyl acetate (0.1 L) and the organics dried over sodium sulfate. Filtration and evaporation of the organic layer provided **33** (3.4 g) as a crystalline white solid in 91% yield. TLC (1:1 hexanes/ethyl acetate) R_f 0.8; IR (Thin Film) 3476, 3115, 1730, 1705, 1535, 1508, 1423, 1360, 1282, 1250, 1191, 997, 856, 748, 732 cm^{-1}; ^1H NMR (DMSO-*d*6) δ 8.21 (d, J = 3.9

Hz, 1H), 7.80 (d, J = 3.9 Hz, 1H), 3.87 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 161.2, 155.0, 138.4, 133.3, 130.4, 54.0; EI-MS m/e 186.994 (M^+ calcd for 186.994 $C_6H_5NO_4S$).

Methyl 5-aminothiophene-2-carboxylate hydrochloride (HCl•H_2N-Tp-OMe, **34**). Concentrated hydrochloric acid (5.8 mL) was added dropwise to a mixture of **33** (0.3 g, 1.6 mmol) and tin(II) chloride dihydrate (2.43 g, 12.8 mmol) in 95 % ethanol (5.8 mL) at room temperature. Sufficient cooling was necessary to keep the reaction temperature under 35 °C. The reaction was stirred at 35 °C for 2 h. The ethanol was removed in vacuo and the aqueous layer washed twice with hexanes (50 mL). The aqueous layer was neutralized with 1N NaOH to pH = 9, upon which time a milky white emulsion formed. The mixture was extracted several times with ethyl acetate (50 mL) and the organics were dried over sodium sulfate. The ethyl acetate was evaporated to give a thin yellow film. Addition of HCl in diethyl ether (2 M) gave the hydrochloride salt **34** (220 mg) as a white solid in 71% yield. TLC (1:1 hexanes/ethyl acetate) (amine) R_f 0.55, (hydrochloride) R_f 0.0; IR (Thin Film) 3219, 1731, 1706, 1471, 1272, 1088, 739 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 7.33 (d, J = 4.5 Hz, 1H), 6.70 (s, 2H), 5.87 (d, J = 4.5 Hz, 1H), 3.65 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 163.5, 162.8, 136.2, 105.1, 51.8; EI-MS m/e 157.020 (M^+ calcd for 157.020 $C_6H_7NO_2S$).

Methyl 5-[(1-methy-4-nitroimidazol-2-yl)-carbonylamino]-thiophene-2-carboxylate (NO_2-Im-Tp-OMe, **35**). Compound **35** was prepared from **34** (200 mg, 1.0 mmol) according to the procedure provided for **19**, providing **35** (164 mg) as a yellow solid in 51% yield. TLC (5:2 hexanes/ ethyl acetate) R_f 0.50; IR (Thin Film) 3133, 1698, 1672, 1560, 1543, 1521, 1455, 1379, 1313, 1265, 1098, 746.9 cm^{-1}; ^{1}H NMR (DMSO-*d6*) δ 12.55 (s, 1H), 8.68 (s, 1H), 7.64 (d, J = 3.9 Hz, 1H), 7.16 (d, J = 3.9 Hz, 1H), 4.04

(s, 3H), 3.77 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 163.1, 155.6, 145.8, 136.8, 132.6, 128.0, 123.5, 115.3, 52.6, 37.5; EI-MS m/e 310.037 (M^+ calcd for 310.037 $C_{11}H_{10}N_4O_5S$).

Methyl 5-[(4-amino-1-methylimidazol-2-yl)carbonylamino]thiophene-2-carboxylate hydrochloride (HCl•H_2N-Im-Tp-OMe, **36**). Compound **36** was synthesized from **35** (150 mg, 0.48 mmol) according to the procedure provided for **20**, providing **36** (107 mg) as an off white solid in 70% yield. TLC (ethyl acetate) (amine) R_f 0.45, (hydrochloride) R_f 0.0; IR (Thin Film) 3344, 3204, 2954, 1691, 1673, 1561, 1511, 1458, 1438, 1343, 1275, 1100 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 12.20 (s, 1H), 7.63 (d, J = 4.2 Hz, 1H), 7.40 (s, 1H), 7.11 (d, J = 4.2 Hz, 1H), 3.99 (s, 3H), 3.76 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 177.4, 162.3, 159.4, 155.0, 145.4, 131.8, 128.5, 113.8, 51.7, 35.6; EI-MS m/e 280.063 (M^+ calcd for 280.063 $C_{11}H_{12}N_4O_3S$).

Methyl 5-({4-[(tert-butoxy)carbonylamino]-1-methylimidazol-2-yl} carbonylamino) thiophene-2-carboxylate (Boc-Im-Tp-OMe, **37**). Compound **37** was prepared from **36** (100 mg, 0.35 mmol) according to the procedure provided for **21**, giving **37** (90 mg) as a white solid in 66% yield. TLC (1:1 hexanes/ethyl acetate) R_f 0.75; IR (Thin Film) 3282, 2964, 1707, 1692, 1673, 1573, 1550, 1368, 1341, 1273, 1159, 1096 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 11.91 (s, 1H), 9.32 (s, 3H), 7.62 (d, J = 4.2 Hz, 1H), 7.33 (s, 1H), 7.07 (d, J = 4.2 Hz, 1H), 3.94 (s, 3H), 3.76 (s, 3H), 1.44 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 162.3, 155.5, 152.8, 145.7, 136.7, 132.4, 131.8, 122.0, 115.4, 113.5, 51.7, 35.2, 28.1; EI-MS m/e 380.115 (M^+ calcd for 380.115 $C_{16}H_{20}N_4O_5S$).

5-({4-[(tert-butoxy)carbonylamino]-1-methylimidazol-2-yl}carbonylamino)-thiophene-2-carboxylic acid (Boc-Im-Tp-OH, **17**). Compound **17** was prepared from **37** (90 mg, 0.23 mmol) according to the procedure provided for **15**, giving **17** (80 mg) as a

white solid in 91% yield. TLC (5:2 hexanes/ethyl acetate, 10% acetic acid) R_f 0.6; IR (Thin Film) 3246, 2965, 1674, 1556, 1509, 1456, 1314, 1271, 1237, 1162, 1102, 1023, 750 cm^{-1}; ^1H NMR (DMSO-d_6) δ 11.79 (s, 1H), 9.31 (s, 1H), 7.52 (d, J = 3.9 Hz, 1H), 7.32 (s, 1H), 7.04 (d, J = 3.9 Hz, 1H), 3.93, (s, 3H), 1.43 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 163.4, 155.4, 152.8, 145.2, 136.7, 132.5, 131.3, 123.8, 115.2, 113.4, 79.1, 35.2, 28.1; EI-MS m/e 366.100 (M^+ calcd for 366.100 $C_{15}H_{18}N_4O_5S$).

Methyl 3-hydroxythiophene-2-carboxylate (Ht-OMe, **38**). To dry methanol (81 mL) under nitrogen was added sodium metal (3.68 g, 304 mmol). After H_2 evolution had stopped, the solution was cooled to 0 °C and methylthioglycolate (10 g, 179 mmol) was added dropwise. Methyl-2-chloroacrylate (10.88 g, 179 mmol) in MeOH (21 mL) was then added dropwise, resulting in the formation of a cloudy yellow precipitate. The solution was allowed to warm to room temperature and stirred for 2 hours, whereupon the precipitate turned dark brown. The solvent was removed in vacuo to give a dark yellow solid which was then acidified to pH = 2 with 4N HCl. The aqueous layer was extracted three times with methylene chloride (150 mL) and the resulting organic layer washed three times with water (150 mL) and dried over magnesium sulfate. Filtration and evaporation of the organic layer gave a dark oil that was subjected to column chromatography (20:1 hexanes/ethyl acetate) to give **38** (18.4 g) as a clear crystalline solid in 64.7% yield. TLC (20:1 hexanes/ethyl acetate) R_f 0.47; IR (Thin Film) 3334, 3112, 2955, 1716, 1664, 1552, 1444, 1415, 1350, 1296, 1208, 1104, 1032, 781 cm^{-1}; ^1H NMR ($CDCl_3$) δ 9.58 (s, 1H), 7.59 (d, J = 5.7 Hz, 1H), 6.75 (d, J = 4.8 Hz, 1H), 3.90 (s, 3H); ^{13}C NMR ($CDCl_3$) δ 131.7, 119.4, 52.2; EI-MS m/e 158.004 (M^+ calcd for 158.004 $C_6H_6O_3S$).

Methyl 3-hydroxy-4-nitrothiophene-2-carboxylate (NO_2-Ht-OMe, **39**). **38** (15 g, 98 mmol) was added to concentrated H_2SO_4 (48 mL) and stirred until homogeneous. The solution was then cooled to -10 – 0 °C and HNO_3 (4.3 mL) in H_2SO_4 (24 mL) was added dropwise with sufficient cooling to keep the temperature below 0° C. After the addition of HNO_3 was complete, the solution was allowed to stir at 0° C for 3 h. The resulting black solution was added to ice and extracted three times with methylene chloride (150 mL) and dried over magnesium sulfate. The solvent was removed in vacuo and the resulting residue chromatographed over silica (5:1 hexanes/ethyl acetate) to give **39** (8.2g) as a yellow solid in 37.6% yield. TLC (5:1 hexanes/ethyl acetate) R_f 0.12; IR (Thin Film) 3107, 1674, 1561, 1520, 1446, 1368, 1267, 1212, 1127, 974, 900, 842, 773 cm^{-1}; ^1H NMR ($CDCl_3$) δ 10.15 (s, 1H), 8.43 (s, 1H), 3.97 (s, 3H); ^{13}C NMR ($CDCl_3$) δ 164.6, 155.6, 132.6, 53.0; EI-MS m/e 202.989 (M^+ calcd for 202.989 $C_5H_5NO_5S$).

Methyl 3-methoxy-4-nitrothiophene-2-carboxylate (NO_2-Mt-OMe, **40**). A mixture of **39** (1.5g, 7.4mmol) and THF (29.5 mL) was cooled to 0 °C. Diazomethane (341 mg, 27 mL, 8.12 mmol) in diethyl ether was slowly added using a plastic funnel. After several seconds, nitrogen evolution ceased and the solution was allowed to warm to room temperature. A few drops of glacial acetic acid were added to ensure the complete consumption of diazomethane. The solvent was removed in vacuo to give **40** (1.49 g) as a yellow solid in 93% yield. TLC (1:1 hexanes/ethyl acetate) R_f 0.72; IR (Thin Film) 3115, 2960, 1725, 1555, 1506, 1453, 1435, 1387, 1353, 1281, 1199, 1110, 1059, 954, 772 cm^{-1}; ^1H NMR ($CDCl_3$) δ 8.38 (s, 1H), 4.10 (s, 1H), 3.94 (s, 3H); ^{13}C NMR ($CDCl_3$) δ 156.5, 131.3, 63.9, 53.0; EI-MS m/e 217.004 (M^+ calcd for 217.004 $C_7H_7NO_5S$).

Methyl 4-amino-3-methoxythiophene-2-carboxylate hydrochloride ($HCl \cdot H_2N$-Mt-OMe, **4**). A mixture of **4** (800 mg, 3.69 mmol) and tin(II) chloride dihydrate (13.3g, 58.9mmol) in 95% EtOH (29.5 mL) were stirred vigorously at room temperature. Concentrated HCl (29.5 mL) was added dropwise and the solution heated at 35 °C for 6 h. The solution was removed from heat and adjusted to pH = 9 with 4N NaOH (15mL). The resulting white emulsion was extracted three times with ethyl acetate (100mL) and dried over magnesium sulfate. The solvent was removed in vacuo and a small amount of fresh ethyl acetate added. 2M hydrogen chloride in diethyl ether was added to precipitate **4** as the crude hydrochloride salt. The salt was filtered and taken directly on to the next step.

Methyl 4-[(tert-butoxy)carbonylamino]-3-methoxythiophene-2-carboxylate (Boc-Mt-OMe, **41**). A mixture of **4** (1.0 g, 4.47 mmol), TEA (498 mg, 0.68 mL, 4.92 mmol) and Boc-anhydride (1.0 g, 4.92 mmol) in methylene chloride (9 mL) was stirred at 60 °C for 12 hours. The solution was washed three times with saturated ammonium chloride solution and dried over magnesium sulfate. The organics were filtered and removed in vacuo. The resulting red solid was subjected to chromatography (5:1 hexanes/ethyl acetate) providing **41** (528 mg) as a white solid in 41.3% yield. TLC (10:1 hexanes/ethyl acetate) R_f 0.24; IR (Thin Film) 3437, 3329, 2979, 1772, 1716, 1530, 1440, 1376, 1230, 1165, 1081, 1055, 993, 861, 778 cm^{-1}; ^1H NMR ($CDCl_3$) δ 7.58 (s, 1H), 6.82 (s, 1H), 4.06 (s, 3H), 3.85 (s, 3H), 1.52 (s, 9H); ^{13}C NMR ($CDCl_3$) δ 161.5, 152.6, 151.5, 130.2, 112.0, 111.2, 85.5, 81.3, 62.9, 52.3, 28.6, 28.0; EI-MS m/e 287.083 (M^+ calcd for 287.083 $C_{12}H_{17}NO_5S$).

4-[(tert-butoxy)carbonylamino]-3-methoxythiophene-2-carboxylic acid (Boc-Mt-OH , **11**). A mixture of **41** (250 mg, 0.87 mmol) and KOH (48.8 mg, 0.87 mmol) in dry methanol (1 mL), was heated to 50 °C for 6 hours. The solution was added to methylene chloride (5 mL) and water (5 mL). The aqueous layer was washed three times with methylene chloride (10 mL) and acidified to pH = 3 with 1N HCl (3.5 mL). The aqueous solution was then extracted three times with methylene chloride (10 mL) and dried over magnesium sulfate. Filtration and evaporation of the organic layer provided **11** (193 mg) as an off white solid in 81% yield. TLC (10:1 hexanes/ethyl acetate) R_f 0.24; IR (Thin Film) 3400, 1699, 1526, 1438, 1369, 1230, 1154, 1053 cm^{-1}; ^1H NMR ($CDCl_3$) δ 7.68 (s, 1H), 6.84 (s, 1H), 4.08 (s, 3H), 1.53 (s, 9H); ^{13}C NMR ($CDCl_3$) δ 165.8, 152.7, 130.4, 112.0, 111.2, 81.4, 63.3, 62.3, 28.7; EI-MS m/e 273.067 (M^+ calcd for 273.067 $C_{11}H_{15}NO_5S$).

2-{(tert-butoxy)carbonylamino]-5-methyl-1,3-thiazole-4-carboxylic acid (Boc-Nt-OH, **10**). 2-ketobutyric acid (10 g, 98 mmol) was treated dropwise with bromine (8 mL, 25 g, 157 mmol) while stirring. Upon completion of addition, the reaction was allowed to stir until the red color of the bromine had dissipated. Thiourea (14.8 g, 196 mmol) was then added in portions and stirring was continued overnight. The mixture was acidified with conc. HCl and the precipitated hydrochloride was filtered and washed with cold ethanol. The crude solid was taken into DMF (50 mL), followed by the addition of DIEA (10 mL) and Boc anhydride (21.4 g, 98 mmol). The mixture was stirred at 60 °C for 12 h. The reaction was then diluted with ethyl acetate and washed three times with brine. The combined organics were dried over sodium sulfate and removed in vacuo to give a crude oil. The material was dissolved in methanol (0.1 L) and 1N NaOH (0.1 L), and

stirred at room temperature for 1 h. The methanol was then removed and the aqueous layer washed two times with diethyl ether (0.1 L). The aqueous phase was acidified to pH 2 with 1 N HCl and extracted three times with ethyl acetate (0.1 L). The combined organics were dried over sodium sulfate and concentrated in vacuo to afford **10** (11 g) as a white flaky solid in 44% yield. (5:2 hexanes/ethyl acetate, 10% acetic acid) R_f 0.4; (Thin Film) 3191, 2978, 1714, 1669, 1578, 1562, 1317, 1165 cm^{-1}; ^1H NMR (DMSO-d_6) δ 11.54 (s, 2H), 2.54 (s 3H), 1.44 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 164.1, 155.4, 137.1, 136.9, 81.9, 28.6, 12.94, 11.61; EI-MS m/e 258.067 (M^+ calcd for 258.067 $C_{10}H_{14}N_2O_4S$).

3. Hairpin Polyamide Synthesis. – Polyamides were synthesized from Boc-β-alanine-Pam resin (50 mg, 0.59 mmol/g) and purified by preparatory HPLC according to published manual solid phase protocols.[19]

Im-Im-Nh-Py-γ-Im-Py-Py-Py-β-Dp (**43**). (Boc-Nh-OH) (33 mg, 0.147 mmol) was incorporated by activation with HBTU (53 mg, 0.140 mmol), DIEA (50 μl) and DMF (300 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 1.5 h at room temperature. After Boc-deprotection, Boc-Im-OH (35 mg, 0.147 mmol) was activated using HBTU (53 mg, 0.140 mmol), DIEA (50 μl) and DMF (300 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-**Nh**-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 1.5 h at room temperature, and determined to be complete by analytical HPLC. After Boc-deprotection, the terminal imidazole residue was added using Im-$COCCl_3$. Im-$COCCl_3$ (67 mg, 0.295 mmol), DIEA (50 μl) and DMF

(600 μl) were added to the reaction vessel containing NH_2-Im-**Nh**-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 2 h at 37 °C, and determined complete by analytical HPLC. A sample of Im-Im-**Nh**-Py-γ-Im-Py-Py-Py-β-Pam resin (50 mg) was placed into a 20 mL scintillation vial, followed by Dp (1 mL). The mixture was allowed to stand for 2 h at 85 °C with occasional agitation. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was purified by reversed phase HPLC to provide Im-Im-**Nh**-Py-γ-Im-Py-Py-Py-β-Dp (**32**) (2 mg, 5.6% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1209.59 (M+H calcd for 1209.56 $C_{56}H_{69}N_{22}O_{10}$).

Im-Im-Im-Py-γ-Im-Py-Py-Py-β-Dp (**42**). Boc-Im-OH was incorporated according to previously described procedures.[19] The terminal imidazole residue was incorporated and the compound purified as described above for **43** to provide Im-Im-**Im**-Py-γ-Im-Py-Py-Py-β-Dp (**42**) (2.6 mg, 6.0% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1210.56 (M+H calcd for 1224.56 $C_{56}H_{70}N_{23}O_{10}$).

Im-Im-Tn-Py-γ-Im-Py-Py-Py-β-Dp (**44**). (Boc-Im-Tn-OH) (56 mg, 0.147 mmol) was incorporated by activation with HBTU (53 mg, 0.140 mmol), DIEA (50 μl) and DMF (300 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 24 h at 37 °C. After Boc-deprotection the terminal imidazole residue was incorporated as described for **43**. The compound was cleaved from resin and purified as described for **43** to provide Im-Im-**Tn**-Py-γ-Im-Py-Py-Py-β-Dp (**44**) (2.1 mg, 5.7% recovery) as a fine white powder under lyophilization of the

appropriate fractions. MALDI-TOF-MS (monoisotopic), 1240.53 (M+H calcd for 1240.53 $C_{57}H_{70}N_{21}O_{11}S$).

Im-Im-Tp-Py-γ-Im-Py-Py-Py-β-Dp (**45**). (Boc-Im-Tp-OH) was incorporated and Im-Im-**Tp**-Py-γ-Im-Py-Py-Py-β-Dp was synthesized, according to the procedure provided for **44** to give Im-Im-**Tp**-Py-γ-Im-Py-Py-Py-β-Dp (**45**) (1.8 mg, 4.9% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1226.53 (M+H calcd for 1226.52 $C_{57}H_{70}N_{21}O_{11}S$).

Im-Im-Ht-Py-γ-Im-Py-Py-Py-β-Dp (**46**). (Boc-Mt-OH) (42 mg, 0.147 mmol) was incorporated by activation with HBTU (53 mg, 0.140 mmol), DIEA (50 μl) and DMF (300 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 20 h at 37 °C. After Boc-deprotection, Boc-Im-OH (35 mg, 0.147 mmol) was activated using HBTU (53 mg, 0.140 mmol), DIEA (50 μl) and DMF (300 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-**Mt**-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 40 h at 37 °C, and determined to be complete by analytical HPLC. After Boc-deprotection the terminal imidazole residue was incorporated as described for **43**. The compound was cleaved from resin and purified as described for **43** to provide the methoxy protected Mt-containing polyamide Im-Im-**Mt**-Py-γ-Im-Py-Py-Py-β-Dp (2.0 mg, 5.4% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1256.54 (M+H calcd for 1256.53 $C_{57}H_{70}N_{21}O_{11}S$). The polyamide was then dissolved in DMF (200 μl) and added to a suspension of sodium hydride (40 mg, 60% oil dispersion) and

thiophenol in DMF (400 μl) that was pre-heated for 5 min at 100 °C. The mixture was heated for 2 h at 100 °C. The mixture was then cooled to 0 °C and 20% TFA (7.0 mL) was added. The aqueous layer was washed three times with diethyl ether (8 mL) and then diluted to a total volume of 9.5 mL using 0.1% TFA. The mixture was then purified by reverse-phase HPLC to give the deprotected **Ht**-containing polyamide Im-Im-**Ht**-Py-γ-Im-Py-Py-Py-β-Dp (**46**) (0.83 mg, 41% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1242.51 (M+H calcd for 1242.51 $C_{56}H_{68}N_{21}O_{11}S$).

Im-Im-Fr-Py-γ-Im-Py-Py-Py-β-Dp (**47**). (Boc-Im-Fr-OH) (51 mg, 0.147 mmol) was incorporated by activation with HBTU (53 mg, 0.140 mmol), DIEA (50 μl) and DMF (300 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 1.5 h at room temperature. After Boc-deprotection the terminal imidazole residue was incorporated as described for **43**. The compound was cleaved from resin and purified as described for **43** to provide Im-Im-**Fr**-Py-γ-Im-Py-Py-Py-β-Dp **47** (1.5 mg, 4.2% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1210.54 (M+H calcd for 1210.54 $C_{56}H_{68}N_{21}O_{11}$).

4. 1:1 Motif Polyamide Synthesis. – Polyamides were synthesized from Boc-β-alanine-Pam resin (50 mg, 0.59 mmol/g) and purified by preparatory HPLC according to published manual solid phase protocols.[19]

Im-β-Im-Py-β-Py-β-Im-Py-β-Dp (**48**). A sample of Im-β-Im-Py-β-**Py**-β-Im-Py-β-Pam resin (120 mg) was placed into a 20 mL scintillation vial, followed by Dp (2 mL). The mixture was allowed to stand for 2 h at 85 °C with occasional agitation. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was then purified by reversed phase HPLC to provide Im-β-Im-Py-β-**Py**-β-Im-Py-β-Dp (**48**) (12 mg, 15.3% recovery) was provided as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1107.70 (M+H calcd for 1107.53 $C_{50}H_{66}N_{20}O_{10}$).

Im-β-Im-Py-β-Hp-β-Im-Py-β-Dp (**49**). Polyamide **49** was synthesized, deprotected and purified according to the previously published protocol,[17] to provide Im-β-Im-Py-β-**Hp**-β-Im-Py-β-Dp (**49**) (5.6 mg 7.0% recovery) as a white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1124.20 (M+H calcd for 1124.19 $C_{50}H_{67}N_{20}O_{11}$).

Im-β-Im-Py-β-Nh-β-Im-Py-β-Dp (**50**). Boc-Nh-OH (271 mg, 1.2 mmol) was incorporated by activation with DCC (247 mg, 1.2 mmol) and HOBt (141 mg, 1.2 mmol) in DMF (2 mL). The mixture was shaken at 37 °C for 30 min and filtered into the reaction vessel containing NH_2-β-Im-Py-β-Pam resin. DIEA (400 μl) was added, and coupling was allowed to proceed for 1.5 h at ambient temperature. After Boc-deprotection, Boc-β-OH (227 mg, 1.2 mmol) was activated using HBTU (432 mg, 1.14 mmol), DIEA (400 μl) and DMF (2 mL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-**Nh**-β-Im-Py-β-Pam resin. Coupling was allowed to proceed for 2 h at 37 °C, and determined to be complete by analytical HPLC. The compound was cleaved from resin and purified as

described for **48**. Im-β-Im-Py-β-**Nh**-β-Im-Py-β-Dp (**50**) (9 mg, 11.4% recovery) was provided as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1107.70 (M+H calcd for 1107.53 $C_{49}H_{64}N_{20}O_{10}$).

Im-β-Im-Py-β-Ht-β-Im-Py-β-Dp (**51**). Polyamide **51** was synthesized using Boc-Mt-OH, deprotected and purified as described above for compound **46**, to provide Im-β-Im-Py-β-**Ht**-β-Im-Py-β-Dp (**51**) (1.1 mg 2.8% recovery) as a white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1126.43 (M+H calcd for 1126.47 $C_{49}H_{64}N_{19}O_{11}S$).

Im-β-Im-Py-β-Fr-β-Im-Py-β-Dp (**52**). Boc-β-Fr-OH (369 mg, 1.2 mmol) was incorporated by activation with DCC (247 mg, 1.2 mmol) and HOBt (141 mg, 1.2 mmol) in DMF (2 mL). The mixture was shaken at 37 °C for 30 min and filtered into the reaction vessel containing NH_2-β-Im-Py-β-Pam resin. DIEA (400 μl) was added, and coupling was allowed to proceed for 1.5 h at ambient temperature. The compound was cleaved from resin and purified as described for **48** to provide Im-β-Im-Py-β-**Fr**-β-Im-Py-β-Dp (**52**) (6 mg, 7.7% recovery) as a white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1094.50 (M+H calcd for 1094.60 $C_{49}H_{63}N_{19}O_{11}$).

Im-β-Im-Py-β-Nt-β-Im-Py-β-Dp (**53**). Boc-Nt-OH (309 mg, 1.2 mmol) was incorporated by activation with HBTU (432 mg, 1.14 mmol), DIEA (400 μl) and DMF (2 mL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-β-Im-Py-β-Pam resin. Coupling was allowed to proceed for 20 h at 37 °C. After Boc-deprotection, Boc-β-OH (227 mg, 1.2 mmol) was activated using HBTU (432 mg, 1.14 mmol), DIEA (400 μl) and DMF (2 mL). The

mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-**Nt**-β-Im-Py-β-Pam resin. Coupling was allowed to proceed for 48 h at 37 °C, and determined to be complete by analytical HPLC. The compound was cleaved from resin and purified as described for **48**. Im-β-Im-Py-β-**Nt**-β-Im-Py-β-Dp (**53**) (4.2 mg, 5.2% recovery) was provided as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1125.50 (M+H calcd for 1125.49 $C_{49}H_{64}N_{20}O_{10}S$).

Im-β-Im-Py-β-Tn-β-Im-Py-β-Dp (**54**). Boc-β-Tn-OH (393 mg, 1.2 mmol) was incorporated by activation with HBTU (432 mg, 1.14 mmol), DIEA (400 μl) and DMF (2 mL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-β-Im-Py-β-Pam resin. Coupling was allowed to proceed for 20 h at 37 °C. The compound was cleaved from resin and purified as described for **48**. Im-β-Im-Py-β-**Tn**-β-Im-Py-β-Dp (**54**) (5.8 mg, 7.2% recovery) was provided as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1124.50 (M+H calcd for 1124.49 $C_{50}H_{65}N_{19}O_{10}S$).

Im-β-Im-Py-β-Th-β-Im-Py-β-Dp (**55**). Compound **55** was prepared using the identical protocol provided for **53**. Im-β-Im-Py-β-**Th**-β-Im-Py-β-Dp (**55**) (6.0 mg, 7.5% recovery) was provided as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1125.60 (M+H calcd for 1125.49 $C_{49}H_{64}N_{20}O_{10}S$).

5. Footprinting Experiments. – Plasmids pDHN1 and pAU8 were constructed and 5'-radiolabeled as previously described.[11,18] DNase I footprint titrations were performed according to standard protocols.[23]

Acknowledgements

We are grateful for financial support from the National Institutes of Health and Caltech for a James Irvine Fellowship to R. M. D.

2.9 References

[1] Kopka, M. L., Yoo, C., Goodsell, D., Pjura, P., Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376.

[2] Lown, J. W., Krowicki, K., Bhat, U. G., Skorobogaty, A., Ward, B., Dabrroiak, J. C. *Biochemistry* **1986**, *25*, 7408.

[3] Pelton, J. G., Wemmer, D. E. *J. Am. Chem. Soc.* **1989**, *112*, 1393.

[4] Dervan, P. B. *Bioorg. & Med. Chem.* **2001**, *9*, 2215.

[5] Wade, W. S., Mrksich, M., Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783.

[6] Kielkopf, C. L., Baird, E. E., Dervan, P. B., Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104.

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

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

[9] Lee, M., Krowicki, K., Shea, R. G., Lown, J. W., Pon, R. T. *J. Mol. Recogn.* **1989**, *2*, 84.

[10] Sharma, S. K., Tandon, M., Lown, J. W. *J. Org. Chem.* **2000**, *65*, 1102.

[11] Nguyen, D. H., Szewczyk, J. W., Baird, E. E., Dervan, P. B. *Bioorg. & Med. Chem.* **2001**, *9*, 7.

[12] Zheng-Yun, J. Z., Dervan, P. B. *Bioorg. & Med. Chem.* **2000**, *8*, 2467.

[13] Mrksich, M., Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663.

[14] Trauger, J. W., Baird, E. E., Dervan, P. B. *Nature* **1996**, *382*, 559.

[15] Trauger, J. W., Baird, E. E., Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160.

[16] Janssen, S., Durussel, T., & Laemmli, U. K. *Mol. Cell.* **2000**, *6*, 999.

[17] Urbach, A. R., Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4343.

[18] Dervan, P. B., Urbach, A. R. in *Essays in Contemporary Chemistry*, From Molecular Structure Toward Biology, eds. Quinkert, G. & Kisakürek, M. V. (Verlag Helvetica Chimica Acta, Zurich) **2000**.

[19] Urbach, A. R., Love, J. J., Ross, S. A., Dervan, P. B. *J. Mol. Biol.*, **2002**, *320*, 55.

[20] Baird, E. E., Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.

[21] Dorsey, B. D., Hoffman, J. M. Jr., Joseph, S. A., McDaniel, S. L. *J. Heterocyclic Chem.* **1995**, *32*, 1283.

[22] Rao, K. E., Shea, R. G., Yadagiri, B., Lown, J. W. *Anti-Cancer Drug Des.* **1990**, *5*, 3.

[23] Huddleston, P. R. Barker, J. M. *Synth. Com.* **1979**, *9*, 731.

[24] Trauger, J. W., Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450.

[25] SPARTAN ESSENTIAL Copyright © 1991-2001 by Wavefunction Inc.

[26] Herman, D. M., Turner, J. M., Baird, E. E., Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1121.

[27] Kelly, J. J., Baird, E. E., Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6981.

[28] de Clairac, R. P. L., Seel, C. L., Geierstanger, B. H., Mrksich, M., Baird, E. E., Dervan, P. B., Wemmer, D. E. *J. Am. Chem. Soc.* **1999**, *121*, 2956.

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

[30] Urbach, A. R., Szewczyk, J. W., White, S., Turner, J. M., Baird, E. E., Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121,* 11621.
[31] Kopka, M. L., Goodsell, D. S., Han, G. W., Chiu, T. K., Lown, J. W., Dickerson, R. E. *Structure* **1997**, *5*, 1033.

Chapter 3

DNA Minor Groove Recognition by 3-Methylthiophene/Pyrrole Pair

The text of this chapter was taken in part from a manuscript coauthored with Michael A. Marques, Shane Foister and Professor Peter B. Dervan (Caltech)

(Doss, R. M.; Marques. M. A. and Dervan, P. B. "*DNA Minor Groove Recognition by 3-Methylthiophene/Pyrrole Pair" Chemistry & Biodiversity* **2004**, *1*, 886-899.)

Abstract.

Hairpin polyamides are synthetic oligomers which fold and bind to specific DNA sequences in a *programmable* manner. Internal side by side pairings of the aromatic amino acids N-methylpyrrole (**Py**), N-methylimidazole (**Im**) and N-methylhydroxypyrrole (**Hp**) confer the ability to distinguish between all four Watson-Crick base pairs in the minor groove of B-form DNA. In a broad search to expand the heterocycle repertoire we found that when 3-methylthiophene (**Tn**), which presents a sulfur atom to the minor groove, is paired with **Py**, it exhibits a modest 3-fold specificity for T•A > A•T presumably by shape selective recognition. In this study we explore the scope and limitations of this lead by incorporating multiple **Tn** residues within a single hairpin polyamide. It was found that hairpin polyamides containing more that one **Tn/Py** pair exhibit lowered affinities and specificities for their match sites. It appears that little deviation is permissible from the parent five-membered ring N-methylpyrrole-carboxamide scaffold for DNA recognition.

3.1 Introduction.

Polyamides, a class of crescent shaped oligomers inspired by the natural products netropsin and distamycin A, are able to bind a broad repertoire of DNA sequences with affinities similar to naturally occurring proteins.[1-5] In the first generation design, polyamide specificity can be attributed to the side-by-side pairings of N-methylpyrrole (**Py**) and N-methylimidazole (**Im**) aromatic rings in the minor groove of DNA where an **Im/Py** pair targets G•C and a **Py/Py** pair targets both A•T and T•A.[6, 7] With the addition of the N-methylhydroxypyrrole (**Hp**) ring it was shown that the **Hp/Py** pair distinguishes T•A from A•T.[6, 7] The **Hp** ring exhibits specificity for T through steric fit and specific hydrogen bonds. The bump presented by the exocyclic hydroxyl group of **Hp** docks comfortably in the asymmetric cleft opposite T in a T•A base pair rather than suffer a sterically unfavorable interaction opposite the larger purine ring of A. From X-ray crystal structure analysis, it appears that **Hp** forms two specific hydrogen bonds with the O2 carbonyl of T.[7]

Extensive studies were carried out on polyamides containing **Hp** in order to more fully explore how the incorporation of **Hp** affects ligand affinity and specificity.[8, 9] Towards this end, several eight ring hairpin polyamides were synthesized and their binding profiles assessed. According to the established pairing rules, a polyamide with the sequence Im-Im-**Hp**-Py-γ-Im-**Py**-Py-Py-β-Dp (**1**) (γ = gamma amino butyric acid; β = beta alanine; Dp = dimethylaminopropylamide) would be expected to target the six-base-pair site 5'-tGG**T**Ca-3' while a polyamide with the sequence Im-Im-**Py**-Py-γ-Im-**Hp**-Py-Py-β-Dp (**2**) should target 5'-tGG**A**Ca-3' (**Figure 3.1**). The parent compound, Im-Im-Py-Py-γ-Im-Py-**Py**-Py-β-Dp (**3**) should target both A•T and T•A (5'-tGG**W**Ca-3') with

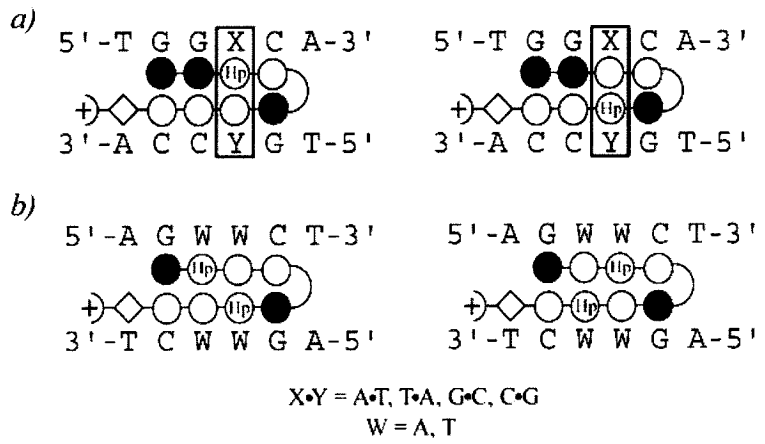


Figure 3.1 Proposed binding models for hairpin polyamides with 5'-T**X**TACA-3' site. A circle enclosing two dots represents lone pairs of N3 of purines and O2 of pyrimidines. A circle containing an H represents the exocyclic amine of guanine. Putative hydrogen bonds are indicated by dashed lines. (a) N-terminal residue drawn in "sulfur down" syn conformation. (b) N-terminal residue drawn in "sulfur up" anti conformation.

similar affinities. Hairpins **1-3** were tested within the sequence context 5'-tGG**X**Ca-3' (**X** = A,T,G,C) where all four Watson-Crick base pairs were varied under the third (in bold) polyamide residue. As expected **1** bound its match site 5'-tGG**T**Ca-3' with ~20 fold preference over its mismatch sequence 5'-tGG**A**Ca-3' while **2** bound its match site 5'-tGG**A**Ca-3' with ~11 fold preference over its mismatch sequence 5'-tGG**T**Ca-3' (**Table 1**). **3** bound both sites 5'-tGG**A**Ca-3' and 5'-tGG**T**Ca-3' with similar affinities.

While results indicated that one could distinguish a single T•A base pair within a six base-pair DNA site a crucial next step was to explore how the incorporation of multiple **Hp** rings would be tolerated within a single hairpin polyamide. To address this question, polyamides Im-**Hp**-Py-Py-γ-Im-**Hp**-Py-Py-β-Dp (**4**) and Im-Py-**Hp**-Py-γ-Im-Py-**Hp**-Py-β-Dp (**5**) were designed to target their respective binding sites 5'-aG**TA**Ct-3' and 5'-aG**AT**Ct-3' (**Figure 3.2**). We test whether all four ring pairings would code for a specific residue with each of the staggered **Hp** rings specifying for a T. As a control,

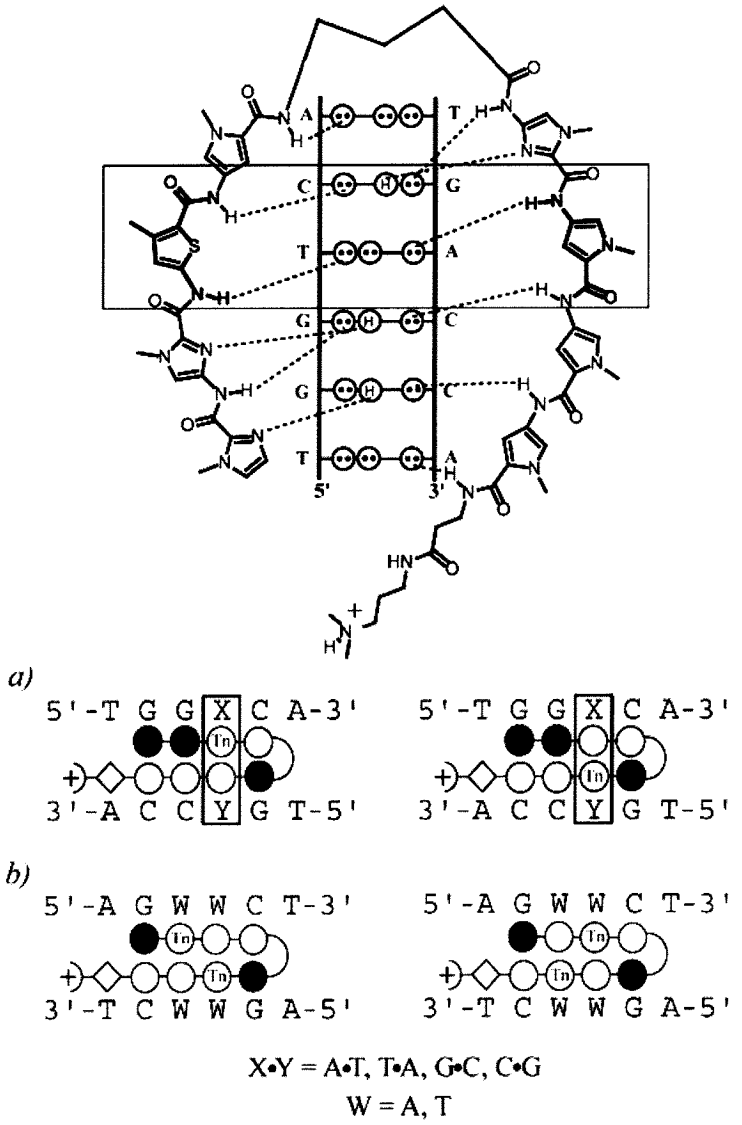


Figure 3.2 (top) pCW15 plasmid design. (bottom) Ball and stick model of hairpin polyamides varying the N-terminal residue. Shaded and non-shaded circles represent imidazole and pyrrole residues, respectively. A circle containing an S denotes an N-terminal thiophene (R = 1-8) residue.

polyamide Im-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**6**) was designed to bind both sequences with similar affinities. It was found that the specificities and affinities were significantly compromised by the incorporation of multiple **Hp** residues.[9] Polyamide **4** bound its match site with a 50-fold reduction in affinity when compared to the parent compound **6**. Polyamide **5** bound with a lower affinity of and preferred its double base-pair mismatch site over its designed match site.

While **Hp** was a breakthrough ring for completing the four base pair code, it was clear that use of **Hp** would be limited for some sequence contexts. In addition, it was observed that oligomers containing **Hp** slowly degraded in the presence of acids or free radicals. This prompts us to explore the properties of other 5-membered heterocyclic amino acids as potential recognition elements for minor groove DNA recognition.[10-14] Assuming that polyamide base pair specificity is derived, in part, from the functionality presented to the minor groove floor by heterocycle ring pairs, we sought to explore new heterocycles for selective recognition.[10-14]

We have previously reported the sequence specificities (or lack thereof) of several novel rings systems when paired with (**Py**) at a single position within the hairpin polyamide sequence context Im-Im-**X**-Py-γ-Im-**Py**-Py-Py-β-Dp (**X** = 1-methylpyrazole (**Pz**), 1H-pyrrole (**Nh**), 5-methylthiazole (**Nt**), 4-methylthiazole (**Th**), 4-methylthiophene (**Tn**), thiophene (**Tp**), 3-hydroxythiophene (**Ht**), and furan (**Fr**). After an exhaustive study, it was found that 3-methylthiophene (**Tn**), exhibited modest specificity (~3 fold) for a T•A base pair when paired against **Py** and was able to maintain a high binding affinity at its match site of K_a = 2.7 x 10^9 M^{-1} (**Figure 3.3**). **Tn** presents a large, polarizable sulfur atom to the minor groove and it is believed that its specificity for T is derived from the A•T base pair's ability to accommodate a large atom in the asymmetric cleft. The **Tn/Py** pairing was a potential step forward to replace **Hp** and we looked to explore the binding properties of hairpin polyamides containing more than one **Tn** residue.

While the selectivity of **Tn** for T•A > A•T was a modest, we were curious to see if there would be a multiplicity effect by targeting two T•A base pairs within a single hairpin binding site. Polyamides Im-**Tn**-Py-Py-γ-Im-**Tn**-Py-Py-β-Dp **(6)** and Im-Py-**Tn**-

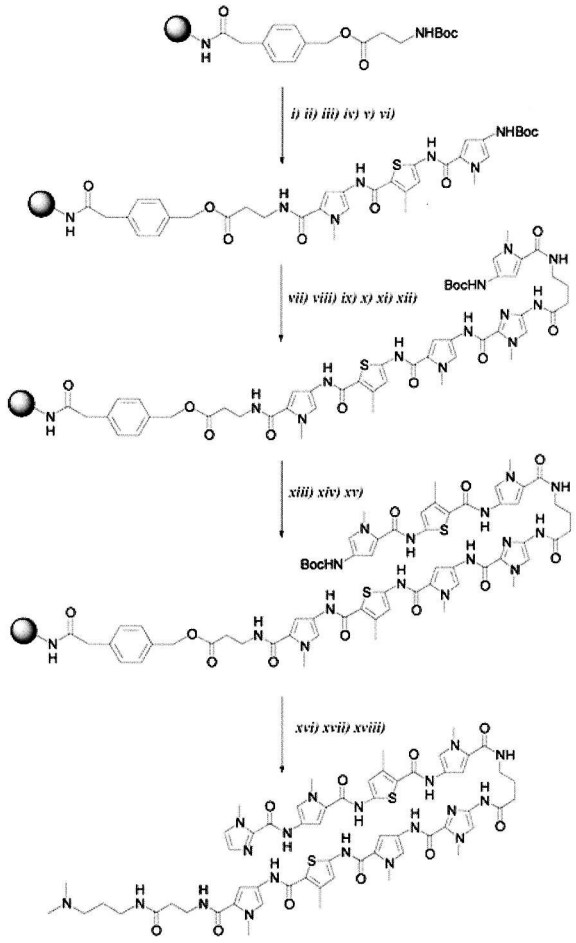


Figure 3.3 (a) Synthesis of *3*-[(tert-butoxy)carbonylamino]-2-thiophenecarboxylic acid (**11**). (i) Et_3N, Boc_2O, DMAP, acetone; (ii) 50% NaOH, MeOH. (b) Synthesis of 3-methoxy-2-thiophenecarboxylic acid (**13**). (iii) K_2CO_3, CH_3I, acetone, acetonitrile, Reflux; (iv) 50% NaOH, MeOH. (c) Synthesis of *3*-Fluoro-2-thiophenecarboxylic acid (**14**). (v) *n*BuLi (2.2 equ.), THF, -78 °C, 0.5 h.; (vi) $(PhSO_2)_2NF$, THF, -78 °C → RT.

Py-γ-Im-Py-**Tn**-Py-β-Dp **(7)** were synthesized to test whether the overall base pair specificity would benefit from the incorporation of two specific **Tn** rings (**Figure 3.4**).

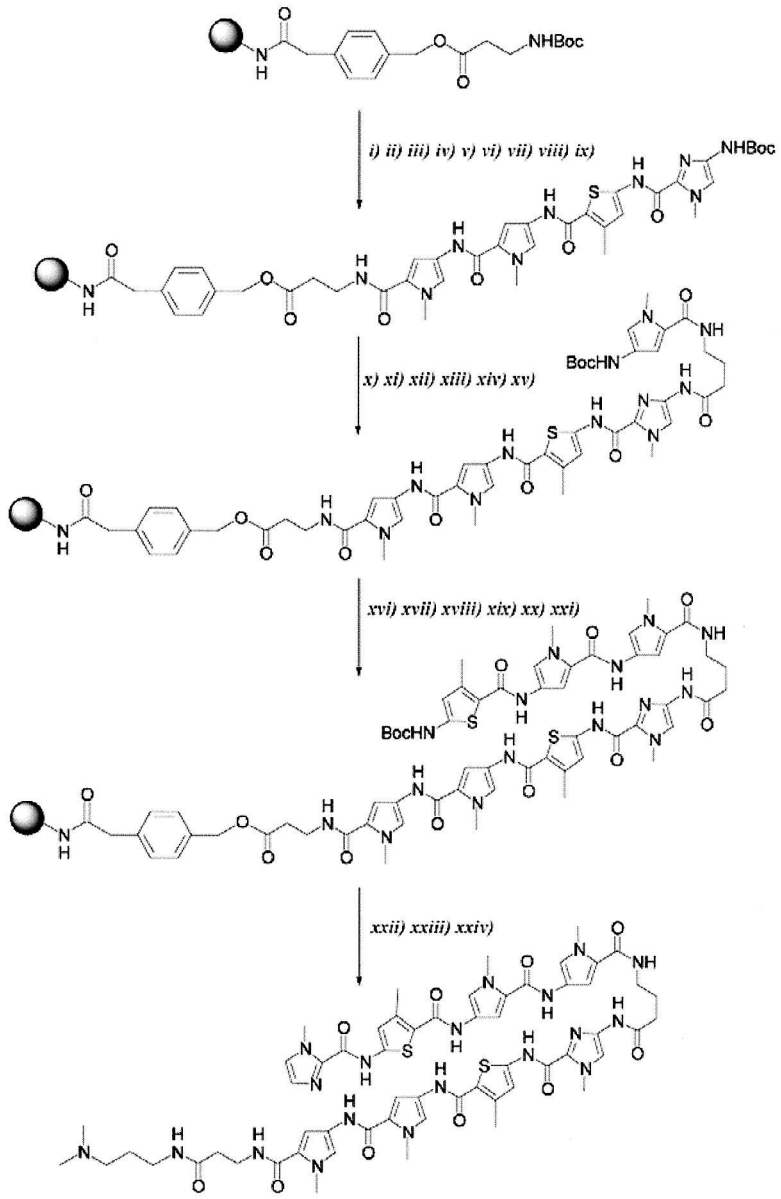


Figure 3.4 Solid phase synthetic scheme for Im-**Tn**-Py-Py-γ-Im-**Tn**-Py-Py-β-Dp starting from commercially available Boc-β-Pam resin: (i) 80% TFA/DCM, 0.4M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) Ac_2O, DIEA, DMF; (iv) 80% TFA/DCM, 0.4M PhSH; (v) Boc-Py-OBt, DIEA, DMF; (vi) Ac_2O, DIEA, DMF; (vii) 80% TFA/DCM, 0.4M PhSH; (viii) Boc-Im-Tn-OH, (HBTU, DIEA, DMF); (ix) Ac_2O, DIEA, DMF; (x) 80% TFA/DCM, 0.4M PhSH; (xi) Boc-γ-OH (HBTU, DIEA, DMF); (xii) Ac_2O, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4M PhSH; (xiv) Boc-Py-OBt, DIEA, DMF; (xv) Ac_2O, DIEA, DMF; (xvi) 80% TFA/DCM, 0.4M PhSH; (xvii) Boc-Py-OBt, DIEA, DMF; (xviii) Ac_2O, DIEA, DMF; (xix) 80% TFA/DCM, 0.4M PhSH; (xx) Boc-Tn-OH, (HBTU, DIEA, DMF); (xxi) Ac_2O, DIEA, DMF; (xxii) 80% TFA/DCM, 0.4M PhSH; (xxiii) Im-$COCCl_3$ (DIEA, DMF); (xxiv) cleave from resin using (N,N-dimethylamino)propylamine, 85 °C.

3.2 Monomer, Dimer and Polyamide Synthesis.

Hairpin polyamides were synthesized manually from Boc-β-PAM resin in a stepwise fashion using Boc-protected monomeric and dimeric amino acids according to solid-phase protocols (**Figures 4** and **5**). Boc-protected amino acid monomers and dimers for **Im**, **Py**, and **Tn-Im** were synthesized according to previously reported procedures.[8, 15] Synthesis of the **Tn-Py** amino acid dimer from the core amino ester (NH_2-Tn-OMe) is shown in **Figure 3.5**.

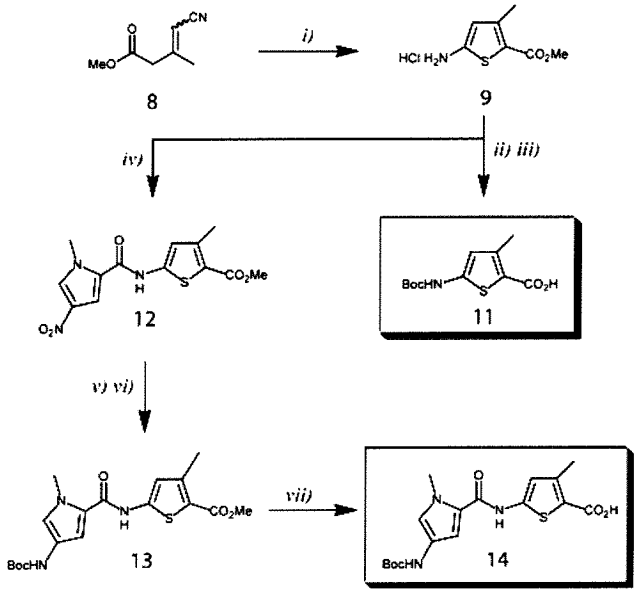


Figure 3.5 Synthesis of Boc-**Tn**-OH (**11**) and Boc-Py-**Tn**-OH (**14**). (i) S, Et_2NH, EtOH; (ii) $(Boc)_2O$, DMAP, DIEA, DMF; (iii) NaOH, MeOH; (iv) NO_2-Py-$COCCl_3$, DMAP, DIEA, EtOAc; (v) H_2 Pd/C, DIEA, DMF; (vi) $(Boc)_2O$; (vii) NaOH, MeOH.

Tn-Py Dimer (**Tn-Py**). The hydrochloride salt of **9** (HCl•H_2N-Tn-OMe) was formed directly via cyclization reaction between **8** and amorphous sulfur.[10] The amine of **9** was Boc-protected using t-butyldicarbonate and DMAP to provide the Boc-protected ester **10** (Boc-Tn-OMe). The use of heat and the transacylation catalyst was necessary for the reaction to occur due to the poor reactivity of the thiophene aryl amine. Saponification of **10** was accomplished by heating in an aqueous solution of sodium hydroxide to provide **11** (Boc-**Tn**-OH). Alternatively, **9** was condensed with NO_2-**Py**-$COCCl_3$ in the presence

of DMAP to provide the dimer **12** (NO_2-Py-**Tn**-OMe). The nitro group was reduced using a Parr apparatus (500 psi H_2) and Pd/C in a mixture of DMF and DIEA. Following reduction, t-butyl dicarbonate was added to the mixture to provide **13** (Boc-Py-**Tn**-OMe). **13** was saponified by heating in an aqueous solution of sodium hydroxide to provide **14** (Boc-Py-**Tn**-OH).

3.3 DNA Affinity and Sequence Specificity in the Hairpin Motif.

Quantitative DNase I footprinting titrations[16] were carried out for the following polyamides on the 285 bp PCR product of plasmids pDHN9 (polyamides **15** and **16**) and pDEH10 (polyamides **6** and **7**):[11] Im-Im-**Tn**-Py-γ-Im-**Py**-Py-Py-β-Dp (**15**), Im-Im-**Py**-Py-γ-Im-**Tn**-Py-Py-β-Dp (**16**), Im-**Tn**-**Py**-Py-γ-Im-**Tn**-**Py**-Py-β-Dp (**6**), Im-**Py**-**Tn**-Py-γ-Im-**Py**-**Tn**-Py-β-Dp (**7**) (ring pairings of interest in bold). The DNA sequence specificity of each polyamide was determined by varying the DNA base pairs within the sequence context, 5'-tGG**X**Ca-3' (**X** = A, T, G, and C) for compounds **15** and **16**, and 5'-aG**WW**Ct-3' (**W** = A and T) for compounds **6** and **7** and comparing the relative affinities of the resulting complexes. The Watson-Crick base-pairs were varied opposite the novel **Tn/Py** pairing in question, according to previously reported specificity studies on eight-ring hairpin polyamides.[10]

Hairpin **6** (**Tn/Py** pair) has been shown to bind with a high affinity for X = T, A ($K_a = 10^9\ M^{-1}$), a 3-fold preference for T•A > A•T, and an 800-fold preference over the **X** = G, C sites (**Figure 3.6**, **Table 1**). The hairpin control **7**, which places the **Tn** ring across the polyamide, bound it's match sequence with a reduced affinity ($K_a = 9.0 \times 10^8\ M^{-1}$) and a lowered specificity of ~2 fold for T•A > A•T. It was found that both Im-**Tn**-

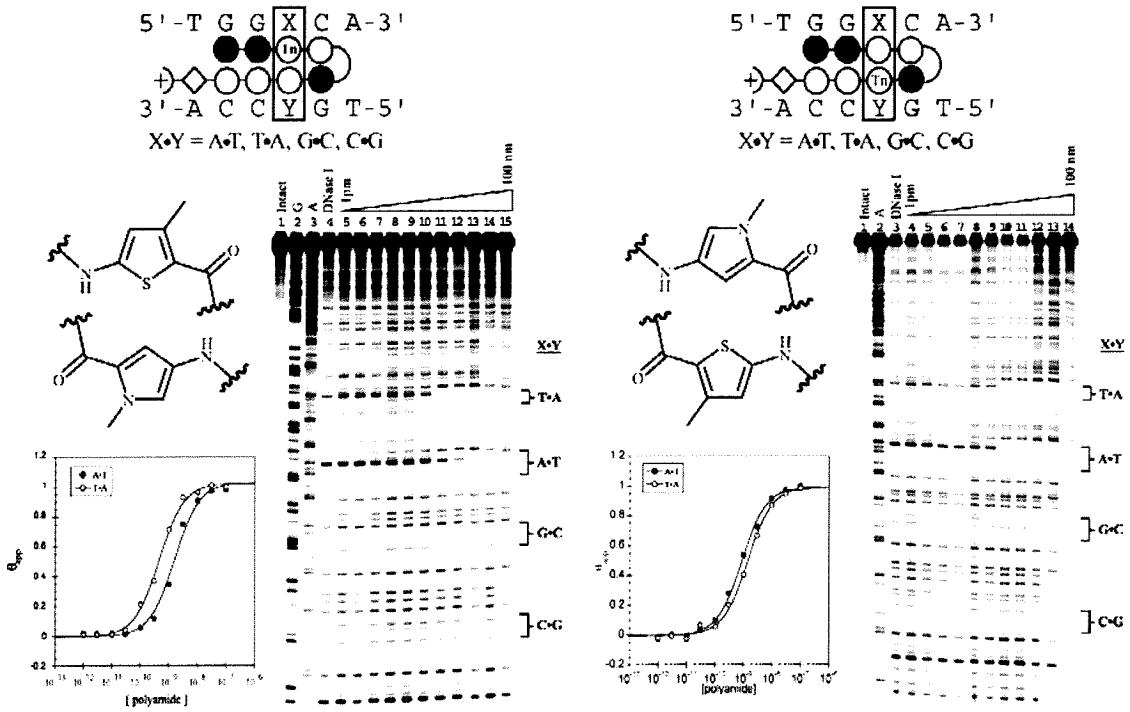


Figure 3.6 Quantitative DNase I footprint titration experiments for polyamides **15** and **16**, respectively, on the 298 bp, 5'-end-labelled PCR product of plasmid pDHN9: (A and B) lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (left, top) chemical structure of the residue of interest; and (left bottom) binding isotherm for the four designed sites. θ_{norm} values were obtained according to published methods. [16] A binding model for the hairpin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; beta alanine is shown as a diamond; and Tn is indicated by a circle containing a Tn.

Table 1. *Thiophene Hairpins:* K_a $[M^{-1}]^{a)b)}$

Polyamide	A•T	T•A	G•C	C•G
	$3.1(\pm 0.7)\times 10^{9}$	$4.7(\pm 0.4)\times 10^{9}$	$2.2(\pm 0.6)\times 10^{8}$	$2.5(\pm 0.9)\times 10^{8}$
	$8.0(\pm 0.4)\times 10^{8}$	$2.7(\pm 0.2)\times 10^{9}$	$\leq 10^{6}$	$\leq 10^{6}$
	$9.0(\pm 0.5)\times 10^{8}$	$5.4(\pm 0.6)\times 10^{8}$	$\leq 10^{6}$	$\leq 10^{6}$

[a)] Values reported are the mean values from at least three DNase-I-footprint titration experiments, with the standard deviation given in parentheses. [b)]Assays were performed at 22 °C in a buffer of 10 mM Tris HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0.

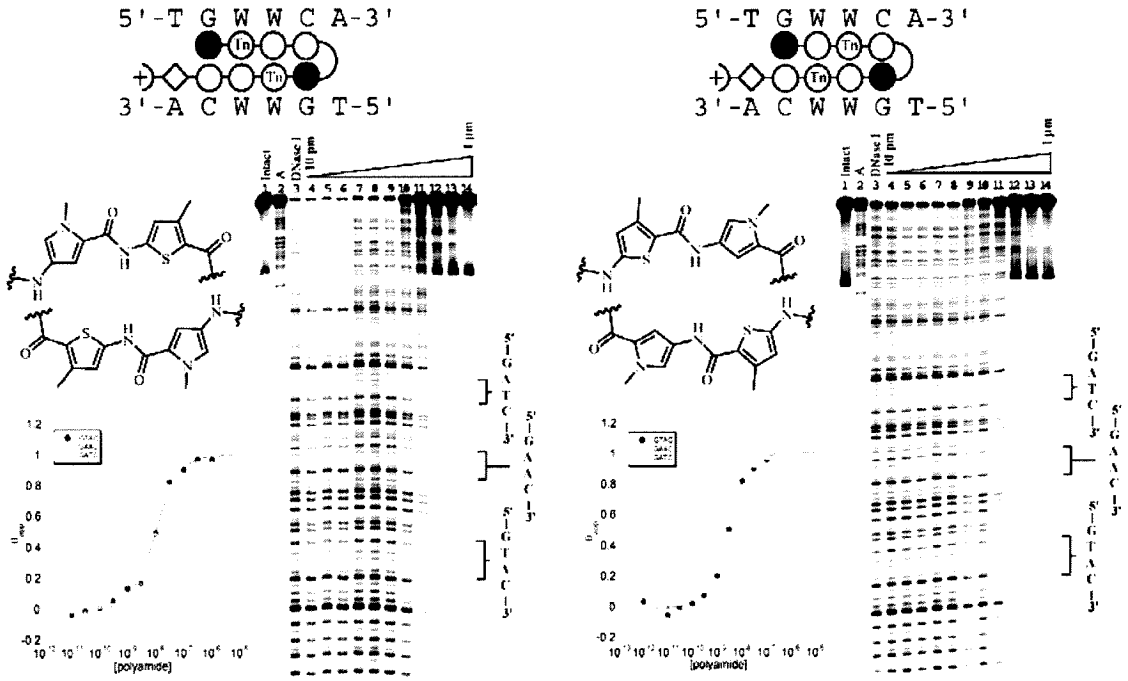


Figure 3.7 Quantitative DNase I footprint titration experiments for polyamides **6** and **7**, respectively, on the 298 bp, 5'-end-labelled PCR product of plasmid pDEH10: (A and B) lane 1, intact DNA; lane 2, G reaction; lane 3, DNase I standard; lanes 4-14, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (left, top) chemical structure of the residue of interest; and (left bottom) binding isotherm for the four designed sites. θ_{norm} values were obtained according to published methods. [16] A binding model for the hairpin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; beta alanine is shown as a diamond; and Tn is indicated by a circle containing a Tn.

Table 2. *Hairpins Containing Multiple Thiophene Rings:* K_a $[M^{-1}]^{a)b)}$

Polyamide	5'-aGTACt-3'	5'-aGAACt-3'	5'-aGATCt-3'
	$3.5(\pm 0.7)\times 10^{10}$	$4.7(\pm 0.7)\times 10^{9}$	$7.4(\pm 1.5)\times 10^{8}$
	$1.0(\pm 0.5)\times 10^{8}$	$1.0(\pm 0.3)\times 10^{8}$	$1.0(\pm 0.4)\times 10^{8}$
	$3.3(\pm 0.9)\times 10^{8}$	$4.7(\pm 0.6)\times 10^{8}$	$4.5(\pm 0.7)\times 10^{8}$

[a] Values reported are the mean values from at least three DNase-I-footprint titration experiments, with the standard deviation given in parentheses. [b] Assays were performed at 22 °C in a buffer of 10 mM Tris HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0.

Py-Py-γ-Im-**Tn**-**Py**-Py-β-Dp (**15**) and Im-**Py**-**Tn**-Py-γ-Im-**Py**-**Tn**-Py-β-Dp (**16**) bound their match sites with greatly reduced affinities (K_a = 1.0 x 10^8 M^{-1} and K_a = 4.5 x 10^8 M^{-1} respectively) and exhibited non-specific binding at concentrations above 10nM (**Figure 3.7**, **Table 2**).

3.4 Discussion.

The search for a ring pair system that can successfully discriminate between the T•A and A•T base pairs has garnered much attention. After several extensive studies, we found that our best lead within the 5-member ring heterocycle family for sequence specificity lay in the modest 3-fold preference of the 3-methylthiophene ring for T over A. While attempts to selectively target multiple T•A base pairs with **Hp** were unsuccessful, we hoped that the **Tn** ring system would not suffer from the same reductions in affinity and specificity. 3-Hydroxypyrrole uses an exocyclic hydroxyl group as a means of shape-selective discrimination and although the 3-OH group can be tolerated by the relatively flexible T•A base pair, its size is slightly larger than optimal and may contribute to the reductions in affinity through clashes with the floor of the minor groove. In addition to steric issues, **Hp** containing hairpins may suffer an energetic penalty which stems from the hydration of 3-OH group.[17] In binding the minor groove of DNA, the polyamide is sequestered from the aqueous solvent and the differential hydration of the bound and unbound hairpins may contribute to the lowered affinities. 3-Methylthiophene, however, presents an endocyclic sulfur atom to the minor groove and solvation issues could be different.

We first examined whether the **Tn** ring's specificity for T•A would be conserved if the recognition element was moved from the top strand of the hairpin to the lower strand. Im-Im-**Py**-Py-γ-Im-**Tn**-Py-Py-β-Dp (**16**) was found to bind its match site 5'-atGG**A**Ca-3' with a moderate affinity and specificity of ~2 fold over its mismatch site. The reduction in affinity and specificity compared to the parent compound was anticipated from earlier studies. In translocating the **Tn** ring to the bottom strand, the match site for the hairpin was changed from 5'-at*GGT*Ca-3' to 5'-at*GGA*Ca-3'. It has been shown that certain DNA sequences, such as 5'-*GGA*-3' have lower affinities for hairpin polyamides presumably due to altered B-form structure or lower intrinsic flexibility.

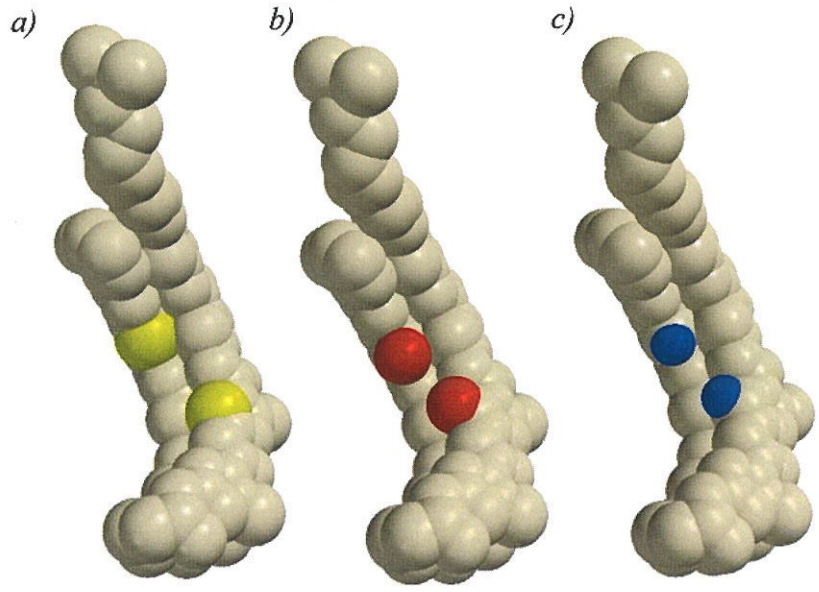


Fig. 3.8 *Ab initio* models illustrate the differences in steric crowding that occur at the polyamide surface which is presented to the minor groove. In each model the atom which is varied is highlighted in a different color. A). Hairpin containing two, staggered **Tn** rings (sulfur in yellow). B). Hairpin containing two, staggered **Hp** rings (hydroxyl in red). C). Hairpin containing two, staggered **Py** rings (hydrogen in blue).

Incorporation of two **Tn** rings in polyamides **6** and **7** led to a substantial reduction in affinity, an abolition of specificity at polyamide concentrations above 10nM. It may be that the minor groove is unable to accommodate two large sulfur atoms without disrupting the subtle steric interactions that confer the T selectivity of the **Tn** ring. *Ab initio* molecular modeling calculations using *Spartan Essential* software illustrate that there are significant steric differences between the **Py**, **Hp**, and **Tn** rings (**Figure 3.8**).[18] Binding affinities are consistent with the modeling trends as the **Tn** polyamide exhibits both the highest degree of steric crowding and the poorest binding affinity.

3.5 Conclusions.

Our search for novel recognition elements has again demonstrated that there is little room for deviation from the parent 5-membered ring N-methylpyrrole-carboxamide scaffold.[10] Although the **Tn/Py** pair can be used to selectively target a single T•A base pair, hairpin polyamides containing multiple **Tn/Py** pairs residues cannot be used to selectively target more than one T•A base pair. It should be noted that efforts to expand beyond N-methylpyrrole-carboxamide analogs to 6-5 fused bicycles (benzimidazole/hydroxybenzimidazole pairs) have shown promising levels of affinities and specificity for DNA.[19]

3.6 Experimental.

Methyl 5- [(tert-butoxy)carbonylamino]-3-methylthiophene-2-carboxylate (**Boc-Tn-OMe**, **10**). A mixture of **9** (0.5 g, 2.40 mmol), $(Boc)_2O$ (1.58 g, 7.22 mmol), DIEA (622 mg, 839 μL, 4.81 mmol), DMAP (58 mg, 0.48 mmol) and DMF (5 mL) was stirred at 50 °C for 12 h. The solvent was removed in vacuo and the resulting brown residue subject to column chromatography (5:2 Hex/EtOAc). Rotoevaporation of the appropriate fractions provided a pale yellow thin film, which when treated with hexanes and dried in vacuo gave **10** as an off white solid (346 mg, 53% Yield). TLC (5:2 Hex/EtOAc) R_f 0.6; ^{1}H NMR (DMSO-d_6) 10.89 (s, 1H), 6.39 (s, 1H), 3.70 (s, 3H), 2.35 (s, 3H), 1.46 (s, 9H); ^{13}C (DMSO-d_6) 163.4, 152.8, 146.7, 145.5, 114.9, 114.7, 81.6, 52.0, 28.7, 16.6; EI-MS m/e 271.088 (M^+ calcd for 271.088 $C_{12}H_{17}NO_4S$).

5- [(tert-butoxy)carbonylamino]-3-methylthiophene-2-carboxylate (**Boc-Tn-OH**, **11**). A mixture of **10** (300 mg, 1.1 mmol), MeOH (1 mL) and 1N NaOH (2 mL) was stirred at 60 °C for 4 h. The MeOH was removed in vacuo and the aqueous solution carefully adjusted to pH 2 with 1N HCl upon which time a milky suspension formed. The mixture was extracted with EtOAc (2 x 25 mL), dried over sodium sulfate. Removal of the organics in vacuo provided **11** as a tan powder (246 mg, 87% Yield). TLC (3:2 Hex/EtOAc, 10% AcOH) R_f 0.6; ^{1}H NMR (DMSO-d_6) 10.80 (s, 1H), 6.36 (s, 1H), 2.33 (s, 3H), 1.45 (s, 9H); ^{13}C (DMSO-d_6) 164.6, 152.7, 146.0, 144.6, 114.7, 81.4, 28.7, 16.5; EI-MS m/e 257.072 (M^+ calcd for 257.072 $C_{11}H_{15}NO_4S$).

Methyl 3-methyl-5-[(1-methyl-4-nitropyrrol-2-yl)carbonylamino]thiophene-2-carboxylate (**NO_2-Py-Tn-OMe**, **12**). A mixture of **9** (1 g, 4.8 mmol), NO_2-Py-$COCCl_3$ (1.96 g, 7.21 mmol), DIEA (652 mg, 880 μL, 5.05 mmol), and DMAP (60 mg, 0.48 mmol) was stirred in EtOAc (15 mL) at 40 °C overnight. The mixture was allowed to cool to room temperature and sufficient hexanes were added to completely precipitate a pale white solid. The precipitate was filtered, washed with cold EtOAc, and dried under vacuum to provide **12** (1.44 g, 93% Yield). TLC (5:2 Hex/EtOAc) R_f 0.55; ^{1}H NMR (DMSO-d_6) 11.73 (s, 1H), 8.27 (d, J = 1.8 Hz, 1H), 7.73 (d, J = 1.8 Hz, 1H), 3.95 (s, 3H), 3.73 (s, 3H), 2.42 (s, 3H); ^{13}C (DMSO-d_6) 163.5, 157.6, 144.8, 144.2, 134.6, 130.0, 125.0, 117.0, 116.5, 110.0, 52.1, 38.4, 16.3; EI-MS m/e 323.058 (M^+ calcd for 323.058 $C_{13}H_{13}N_3O_5S$).

Methyl 5-({4-[(tert-butoxy)carbonylamino]-1-methylpyrrol-2-yl}carbonylamino)-3-methylthiophene-2-carboxylate (**Boc-Py-Tn-OMe**, **13**). A mixture of **12** (500 mg, 1.54 mmol), DIEA (400 mg, 536 μL, 3.08 mmol), Pd/C (50 mg) and DMF (6 mL) was placed in a parr apparatus and hydrogenated (500 psi) for 1.5 h at ambient temperature. The mixture was removed from the parr apparatus and $(Boc)_2O$ (500 mg, 2.28 mmol) was added. The mixture was then stirred for 8 h at 50 °C. The solvent was removed in vacuo, followed by column chromatography of the brown residue (5:2 Hex/EtOAc) to provide **13** as a pale yellow film (205 mg, 34% Yield). TLC (5:2 Hex/EtOAc) R_f 0.37; ^{1}H NMR (DMSO-d_6) 11.36 (s, 1H), 9.20 (s, 1H), 7.05 (s, 1H), 7.00 (s, 1H), 6.70 (s, 1H), 3.81 (s, 3H), 3.71 (s, 3H), 2.40 (s, 3H), 1.43 (s, 9H); ^{13}C (DMSO-d_6) 162.9, 157.8, 152.6, 144.5,

143.9, 122.8, 120.6, 118.9, 115.3, 114.8, 105.0, 59.7, 31.0, 22.1, 20.7, 14.1; EI-MS m/e 393.136 (M^+ calcd for 393.136 $C_{18}H_{23}N_3O_5S$).

5-({4-[(tert-butoxy)carbonylamino]-1-methylpyrrol-2-yl}carbonylamino)-3-methylthiophene-2-carboxylic acid (**Boc-Py-Tn-OH, 14**). A mixture of **13** (200 mg, 0.51 mmol), MeOH (1 mL) and 1N NaOH (2 mL) was stirred at 60 °C for 4 h. The MeOH was removed in vacuo and the aqueous solution carefully adjusted to pH 2 with 1N HCl upon which time a milky white precipitate formed. The mixture was extracted with EtOAc (2 x 25 mL), dried over sodium sulfate. Removal of the organics in vacuo provided **14** as a tan solid (160 mg, 83% Yield). TLC (3:2 Hex/EtOAc, 10% AcOH) R_f 0.6; ^{1}H NMR (DMSO-d_6) 11.28 (s, 1H), 9.20 (s, 1H), 7.04 (s, 1H), 6.99 (s, 1H), 6.68 (s, 1H), 3.81 (s, 3H), 2.38 (s, 3H), 1.43 (s, 9H); ^{13}C (DMSO-d_6) 164.1, 157.7, 152.6, 143.9, 143.1, 122.8, 120.7, 118.8, 116.5, 115.4, 105.0, 78.4, 36.3, 28.2, 15.8; EI-MS m/e 379.120 (M^+ calcd for 379.120 $C_{17}H_{21}N_3O_5S$).

Hairpin Polyamide Synthesis. – Polyamides were synthesized from Boc-β-alanine-Pam resin (50 mg, 0.59 mmol/g) and purified by preparatory HPLC according to published manual solid phase protocols[10].

*Im-**Tn**-Py-Py-γ-Im-**Tn**-Py-Py-β-Dp*: (Boc-Im-**Tn**-OH) (34 mg, 89 μmol) was incorporated by activation with HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (300 μL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Py-Py-β-Pam resin.

Coupling was allowed to proceed for 24 h at 40 °C, followed by capping with acetic anhydride 20% in DMF. After Boc-deprotection, Boc-γ-OH (18 mg, 89 μmol) was activated using HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (300 μL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Im-Tn-Py-Py-β-Pam resin. Coupling was allowed to proceed for 2 h at 40 °C, followed by capping with acetic anhydride 20% in DMF. After Boc-deprotection, the next two Py residues were incorporated as previously described. [Ref] Boc-Tn-OH (23 mg, 89 μmol) was incorporated by activation with HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (300 μL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Py-Py-γ-Im-Tn-Py-Py-β-Pam resin. Coupling was allowed to proceed for 24 h at 40 °C followed by capping as described above. Boc-deprotection of the Boc-Tn-Py-Py-γ-Im-Tn-Py-Py-β-Pam resin was accomplished by shaking the resin in a 80% TFA in DCM mixture for 25 min at room temperature. The terminal **Im** residue was installed using Im-$COCCl_3$. Im-$COCCl_3$ (134 mg, 590 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (1 mL) were added to the H_2N-Tn-Py-Py-γ-Im-Tn-Py-Py-β-Pam resin and coupling was allowed to proceed for 48 h at 40 °C. The resin was then washed with DCM. Dp (1 mL) was added to the resin and the mixture was allowed to stand at 80 °C with occasional agitation for 2 h. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was purified by reversed phase HPLC to provide Im-**Tn**-Py-Py-γ-Im-**Tn**-Py-Py-β-Dp (**10**) (1.5 mg, 4.0% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1256.47 (M+H calcd for 1256.50 $C_{58}H_{70}N_{19}O_{10}S$).

*Im-Py-**Tn**-Py-γ-Im-Py-**Tn**-Py-β-Dp*: (Boc-Py-**Tn**-OH) (34 mg, 89 μmol) was incorporated by activation with HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (300 μL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Py-β-Pam resin. Coupling was allowed to proceed for 24 h at 40 °C, followed by capping with acetic anhydride 20% in DMF. After Boc-deprotection, Boc-γ-Im-OH (29 mg, 89 μmol) was activated using HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (300 μL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Im-Py-Tn-Py-β-Pam resin. Coupling was allowed to proceed for 4 h at room temperature, followed by capping. After Boc-deprotection, the Py residue was incorporated as previously described (Ref). The next Boc-Py-Tn-OH dimer was incorporated as described above. After Boc-deprotection, the final **Im** residue was added using Im-$COCCl_3$. Im-$COCCl_3$ (134 mg, 590 μmol), DIEA (23 mg, 31 μL, 177 μmol), and DMF (1 mL) were added to the H_2N-Py-Tn-Py-γ-Im-Py-Tn-Py-β-Pam resin and coupling was allowed to proceed for 2 h at 40 °C. The resin was then washed with DCM. Dp (1 mL) was added to the resin and the mixture was allowed to stand at 80 °C with occasional agitation for 2 h. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was purified by reversed phase HPLC to provide Im-Py-**Tn**-Py-γ-Im-Py-**Tn**-Py-β-Dp (**11**) (1.9 mg, 5.1% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1256.50 (M+H calcd for 1256.50 $C_{58}H_{70}N_{19}O_{10}S$).

3.7 References

[1] Arcamone F., N. V., Penco S., Orezzi P., Pirelli A., *Nature*. **1964**, 203, 1064.

[2] Kopka, M. L. Y., C. Goodsell, D. Pjura, P. Dickerson, R. E., *Proc. Natl. Acad. Sci. U.S.A.* **1985**, 82, 1376.

[3] Pelton, J. G., Wemmer, D.E., *Proc. Natl. Acad. Sci. U.S.A.* **1989**, 86, (15), 5723.

[4] Dervan, P. B., *Bioorg. & Med. Chem.* **2001**, 9, 2215.

[5] Edelson, B. S., Dervan, P. B., *Curr. Op. Struc. Bio.* **2003**, 13, 284.

[6] White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., *Nature*. **1998**, 391, 468.

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

[8] Urbach, A. R., Szewczyk J. W., White, S., Turner. J. M., Baird, E. E., and Dervan, P. B., *J. Am. Chem. Soc.* **1999**, 121, (50), 11621.

[9] Melander, C., Herman, D. M., Dervan, P. B., Discrimination of A/T sequences in the minor groove of DNA within a cyclic polyamide motif. *Chem. Eur. J.* **2000**, 6, (24), 4487-4497.

[10] Marques, M. A., Doss, R. M., Urbach, A. R., Dervan, P. B., Toward an understanding of the chemical etiology for DNA minor- groove recognition by polyamides. *Helvetica Chimica Acta*. **2002**, 85, (12), 4485-4517.

[11] Nguyen, D. H., Szewczyk, J. W., Baird, E. E., Dervan, P. B., *Bioorg. & Med. Chem.* **2001**, 9, 7.

[12] Foister, S. M., M. A.; Doss, R. M.; Dervan, P. B., *Bioorg. & Med. Chem.* 20, 4333.

[13] Ellervik, U. W., C. C.; Dervan, P. B., *J. Am. Chem. Soc.* **2000**, 122, 9354.

[14] Urbach, A. R., Dervan, P. B., *Proc. Natl. Acad. Sci. USA*. **2001**, 98, 4343.

[15] Baird, E. E., Dervan, P. B., *J. Am. Chem. Soc.* **1996**, 118, 6141.

[16] Trauger, J. W., Dervan, P. B., *Methods Enzymol.* **2001**, 340, 450.

[17] Wellenzohn, B., Loferer, M. J., Trieb, M., Rauch, C., Winger, R. H., Mayer, E., Liedl, K. R., Hydration of hydroxypyrrole influences binding of ImHpPyPy- beta-Dp polyamide to DNA. *J. Am. Chem. Soc.* **2003**, 125, (4), 1088.

[18] SPARTAN ESSENTIAL Copyright © 1991-2001 by Wavefunction Inc.

[19] Renneberg, D., Dervan, P., *J. Am. Chem. Soc.* **2003**, 125, 5707.

Chapter 4

Shape Selective Recognition of T·A Base Pairs by Hairpin Polyamides Containing N-Terminal 3-Methoxy (and 3-Chloro) Thiophene Residues

The text of this chapter was taken in part from a manuscript coauthored with Shane Foister, Michael A. Marques and Professor Peter B. Dervan (Caltech)

(Marques, M. A.; Urbach, A. R.; Doss, R. M. and Dervan, P. B. "*Shape Selective Recognition of T·A Base Pairs by Hairpin Polyamides Containing N-Terminal 3-Methoxy (and 3-Chloro) Thiophene Residues*" *Bioorg. Med. Chem.*, **2003**, *11*, 4333-4340.)

Abstract.

Hairpin polyamides selectively recognize predetermined DNA sequences with affinities comparable to naturally occurring proteins. Internal side-by-side pairs of unsymmetrical aromatic rings within the minor groove of DNA distinguish each of the four Watson-Crick base pairs. In contrast, N-terminal ring pairs exhibit less specificity, with the exception of Im/Py targeting G•C base pairs. In an effort to explore the sequence specificity of new ring pairs, a series of hairpin polyamides containing 3-substituted-thiophene-2-carboxamide residues at the N-terminus was synthesized. An N-terminal 3-methoxy (or 3-chloro) thiophene residue paired opposite Py displayed 6- (and 3-) fold selectivity for T•A relative to A•T base pair, while disfavoring G,C base pairs by > 200-fold. Our data suggests shape selective recognition with projection of the 3-thiophene substituent (methoxy or chloro) to the floor of the minor groove.

4.1 Introduction

Polyamides composed of N-methylpyrrole (Py), N-methylimidazole (Im), and N-methylhydroxypyrrole (Hp) amino acids are crescent-shaped ligands that bind sequence specifically in the minor groove of DNA and have the potential to modulate gene expression by chemical methods. The specificity of DNA recognition arises from interactions between the edges of the Watson-Crick base pairs and antiparallel aromatic amino acid ring pairs oriented N→C with respect to the 5'→3' direction of the DNA helix.[1-3] Covalent head-to-tail linkage of two polyamide strands by γ-aminobutyric acid constitutes the hairpin motif, in which opposing residues from each strand are locked into cofacial pairs.[4, 5] Im/Py distinguishes G•C from C•G and both of these from T•A / A•T base pairs while a Py/Py pair binds both T•A and A•T in preference to G•C / C•G. The exocyclic amino group of guanine imparts G•C specificity to Im/Py pairs through formation of a specific hydrogen bond with N3 of Im. Binding of Py/Py is disfavored at G,C base pairs by destabilizing steric interactions between the C3-H of Py and the guanine amino group.[6, 7] The replacement of C3-H of one Py with hydroxyl creates the Hp/Py pair which exploits the steric fit and hydrogen bond acceptor potential of thymine-O2 as well as the destabilizing steric interaction with the bulkier adenine ring to gain specificity for T•A.[8,9]

The above pairing rules have been used to design hundreds of synthetic ligands that bind predetermined DNA sequences. However, many sequences remain difficult to target, likely due to sequence dependent microstructure variations in minor groove width or curvature. Furthermore, the specificity of cofacial aromatic amino acid pairings depend on their context (position) within a given hairpin polyamide. For example, Im/Py

pairings show comparable specificity for G•C at both *terminal* and *internal* positions.[1] Conversely, Hp/Py pairings do not specify T•A at the N-terminus of hairpin polyamides.[10] The context dependence of Hp is presumably a result of the conformational freedom inherent to an N-terminal aromatic residue. The absence of a second "groove-anchoring" carboxamide allows terminal rings to bind DNA in either of two conformations. For a terminal Hp residue, a rotamer with the hydroxyl recognition element oriented away from the floor of the minor groove could be stabilized by intramolecular hydrogen bonding between the C3-OH and the carbonyl oxygen of the 2-carboxamide. For terminal 2-hydroxybenzamide residues, some measure of T•A selectivity was recovered by creating steric bulk at the 6-position to force the hydroxyl recognition element into the groove.[10] However, N-terminal pairings capable of binding T•A, with affinity and specificity comparable to those of Im/Py for G•C, remain to be devised.

The fidelity of minor groove recognition by N-terminal Im/Py pairings in hairpin polyamides can be rationalized by a combination of both stabilizing and destabilizing forces which favors the rotamer with N3 in the groove and N-methyl out. Rotation of a terminal Im residue in the opposite conformer, orienting N3 away from the minor groove, would create unfavorable lone pair interactions with the proximal carboxamide oxygen, disrupt a favorable hydrogen bond with the exocyclic amine of G, and project an N-methyl group to the DNA floor which is presumably sterically unfavorable. We address in this paper whether T recognition element could be designed using the asymmetric cleft of a T•A base pair as the basis for *shape selective discrimination*. Recent work from our

group has indicated that the polarizable sulfur atom of thiophene heterocycles might serve this purpose.[11]

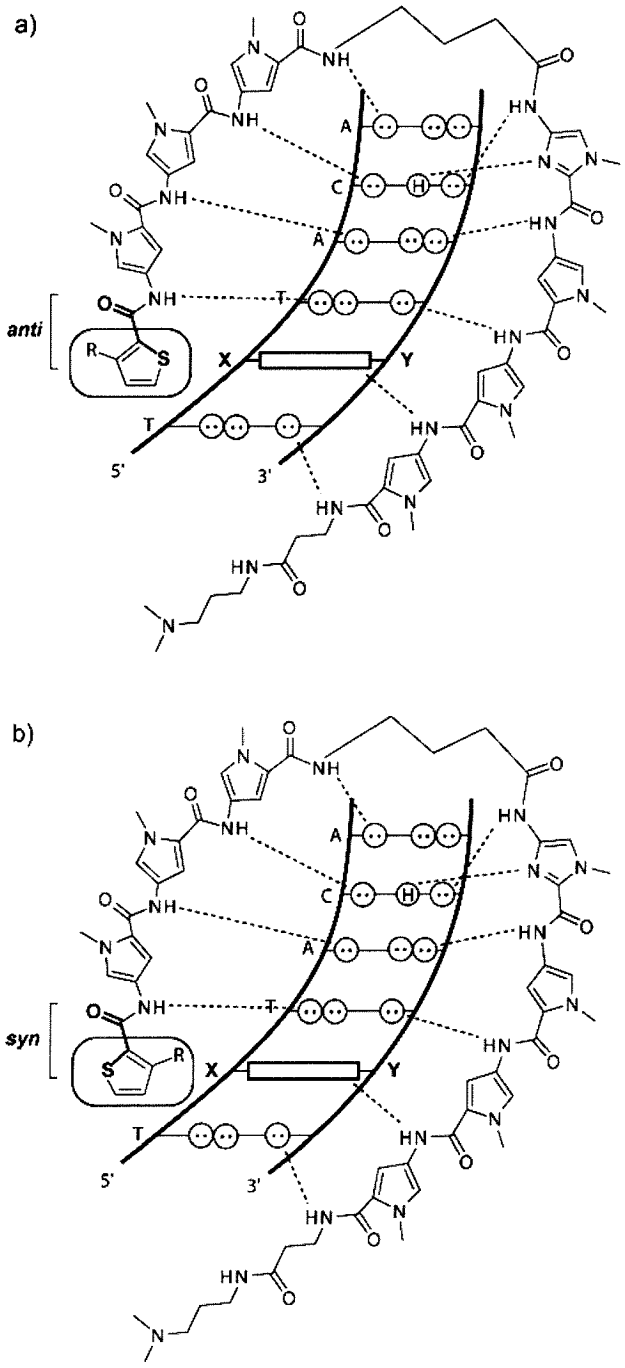


Figure 4.1 Proposed binding models for hairpin polyamides with 5'-TXTACA-3' site. A circle enclosing two dots represents lone pairs of N3 of purines and O2 of pyrimidines. A circle containing an H represents the exocyclic amine of guanine. Putative hydrogen bonds are indicated by dashed lines. (a) N-terminal residue drawn in "sulfur down" syn conformation. (b) N-terminal residue drawn in "sulfur up" anti conformation.

Our experimental design anticipated that substitution of the 3-position of a thiophene-2-carboxamide scaffold could be used to favor an anti ("sulfur down") conformation at the N-terminus by disfavoring contact of the 3-substituent with the floor of the minor groove (**Figure 4.1**). It was envisioned that the electronic properties of the 3-substituent might be used to tune the polarization of the sulfur atom, allowing a more complementary fit with thymine in the minor groove. We attempt here to expand the repertoire of DNA sequences that can be targeted using hairpin polyamides by investigating the DNA recognition properties of a series of N-terminal residues consisting of 3-substituted-thiophene-2-carboxamide heterocycles. Quantitative DNAse I footprinting was used to determine the affinity of eight novel N-terminal 3-substituted thiophene rings residues, paired opposite Py, for each of the four Watson-Crick base pairs (**Figure 4.2**). *Ab initio* computational modeling was used to guide interpretation of the experimental results.

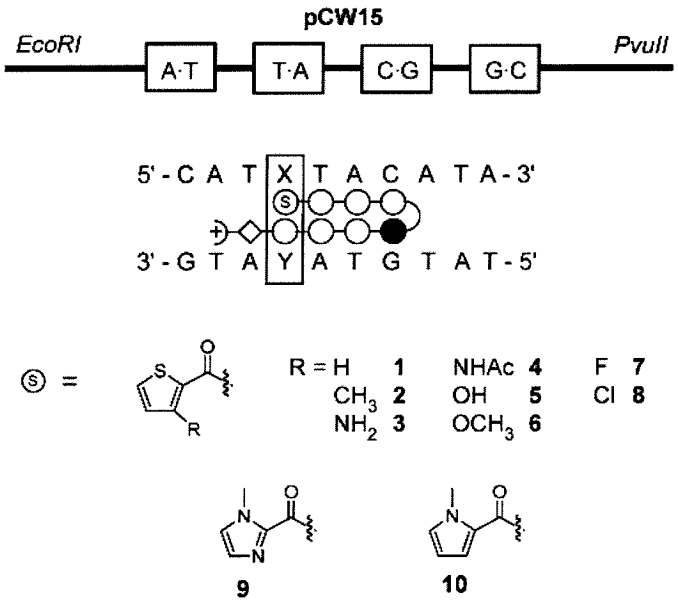


Figure 4.2 (top) pCW15 plasmid design. (bottom) Ball and stick model of hairpin polyamides varying the N-terminal residue. Shaded and non-shaded circles represent imidazole and pyrrole residues, respectively. A circle containing an S denotes an N-terminal thiophene (R = 1-8) residue.

4.2 Results

Monomer synthesis (Figure 4.3)

Methyl 3-aminothiophene-2-carboxylate was Boc-protected and the resulting ester was saponified to yield 3-[(*tert*-butoxy)carbonylamino]-2-thiophenecarboxylic acid (**11**). Methyl 3-hydroxythiophene-2-carboxylate (**12**) was prepared by cyclization of methylthioglycolate and methyl-2-chloroacrylate in methanolic sodium methoxide.[12] Alkylation of (**12**) with iodomethane and subsequent hydrolysis of the methyl ester gave 3-methoxy-2-thiophenecarboxylic acid (**13**). 3-Fluorothiophene-2-carboxylic acid (**14**) was synthesized as described previously.[13] The remaining 3-substituted-thiophene-2-carboxylic acids were obtained from commercial sources.

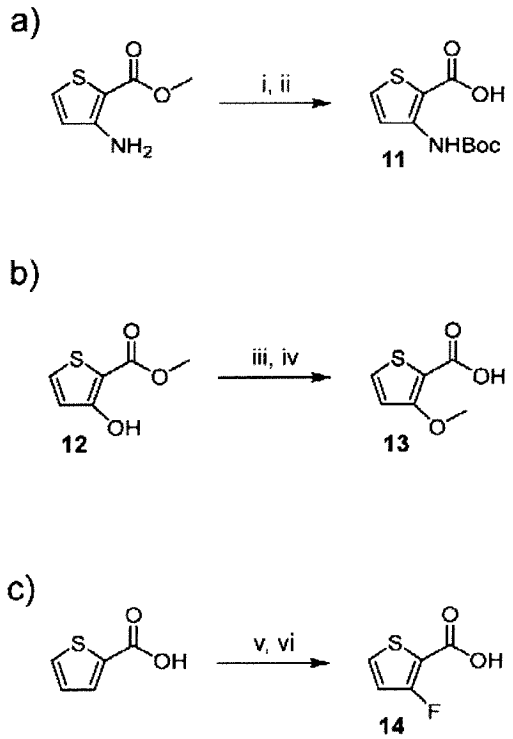


Figure 4.3 (**a**) Synthesis of *3*-[(tert-butoxy)carbonylamino]-2-thiophenecarboxylic acid (**11**). (i) Et_3N, Boc_2O, DMAP, acetone; (ii) 50% NaOH, MeOH. (**b**) Synthesis of 3-methoxy-2-thiophenecarboxylic acid (**13**). (iii) K_2CO_3, CH_3I, acetone, acetonitrile, Reflux; (iv) 50% NaOH, MeOH. (**c**) Synthesis of *3*-Fluoro-2-thiophenecarboxylic acid (**14**). (v) *n*BuLi (2.2 equ.), THF, -78 °C, 0.5 h.; (vi) $(PhSO_2)_2NF$, THF, -78 °C → RT.

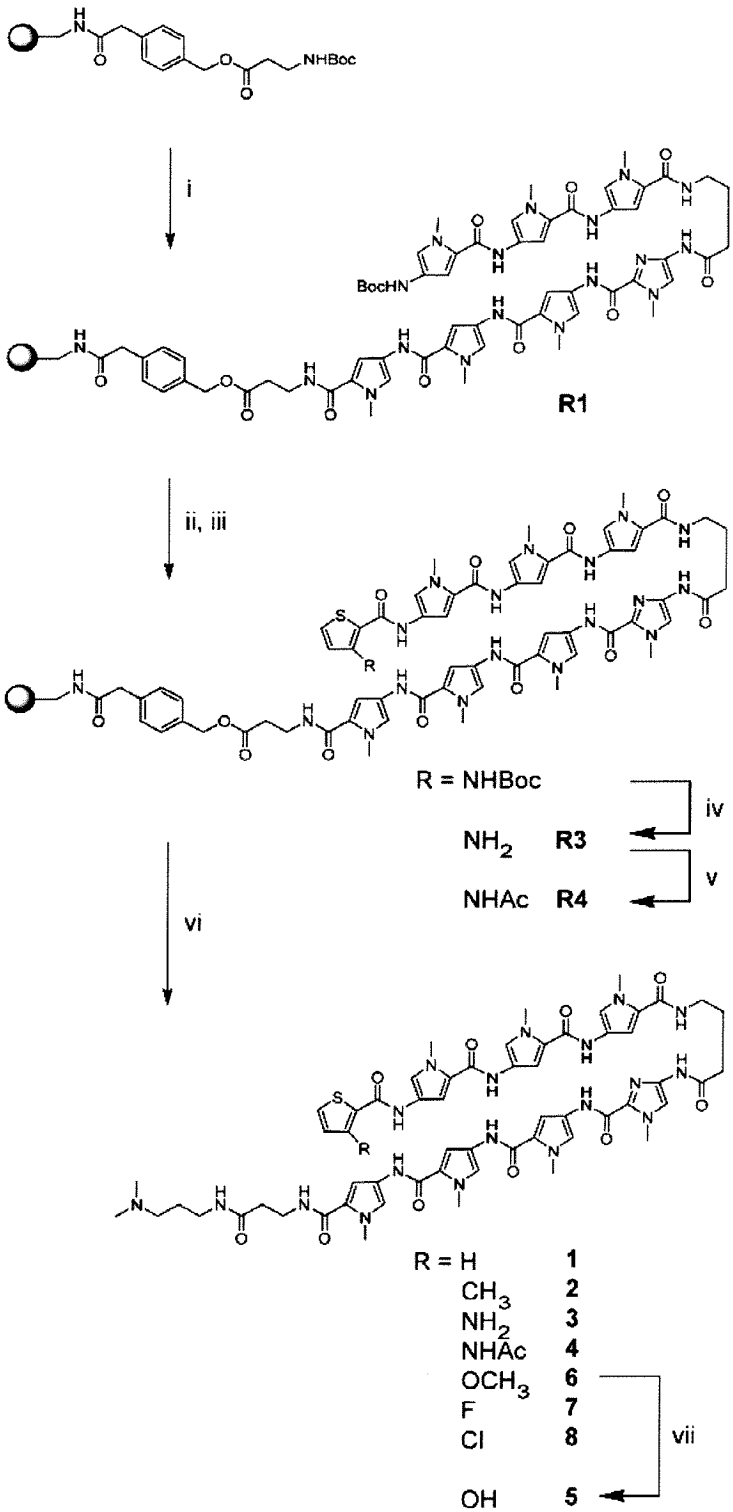


Figure 4.4. Synthesis of hairpin polyamides. (i) Synthesis of polyamide resin by standard solid phase techniques;[14] (ii) TFA, CH_2Cl_2; (iii) 3-R-thiophene-2-CO_2H, HBTU, DMF, DIEA; (iv) TFA, CH_2Cl_2; (v) Ac_2O, DMF, DIEA; (vi) Dp, 40 °C; (vii) PhSH, NaH, DMF, 100 °C.

Polyamide synthesis (Figure 4.4)

Polyamide resin (**R1**) was prepared using manual solid phase synthetic techniques described previously.[14] Treatment of this resin with trifluoroacetic acid (80% TFA in CH_2Cl_2) yielded a support-bound amine that was subsequently acylated with the appropriate, HBTU-activated, thiophene-2-carboxylic acids. Acylation of (**R1**) by (**11**) and removal of the Boc protecting group with TFA yielded resin (**R3**) which was cleaved with dimethylaminopropylamine (Dp) to give polyamide (**3**). Treatment of (**R3**) with acetic anhydride prior to cleavage with Dp gave polyamide (**4**). The remaining polyamides (**1, 2, 6-8**) were cleaved from resin with Dp immediately following acylation of the carboxylic acid. Treatment of (**6**) with sodium thiophenoxide in DMF gave (**5**). Crude products were purified by reversed-phase HPLC and characterized by MALDI-TOF mass spectrometry.

DNA binding energetics

Quantitative DNase I footprinting titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, 22 °C) were performed on 5'-^{32}P end-labeled, 285 bp PCR product from plasmid pCW15.[10] This plasmid contains four binding sites that vary at a single N-terminal position, 5'-A T **N** T A C A-3', where N = T, A, G, C. The DNA sequence specificity of novel thiophene-2-carboxamides was evaluated by comparing their affinities for each Watson-Crick base pair to those of N-methylimidazole (Im) and N-methylpyrrole (Py) (**Figure 4.5** and **Table 1**). The divergent behavior of control polyamides **9** and **10** illustrate the need for development of new N-terminal residues. A terminal Im/Py pairing **9** binds it match sequence, 5'-A T **G** T A C A-3',

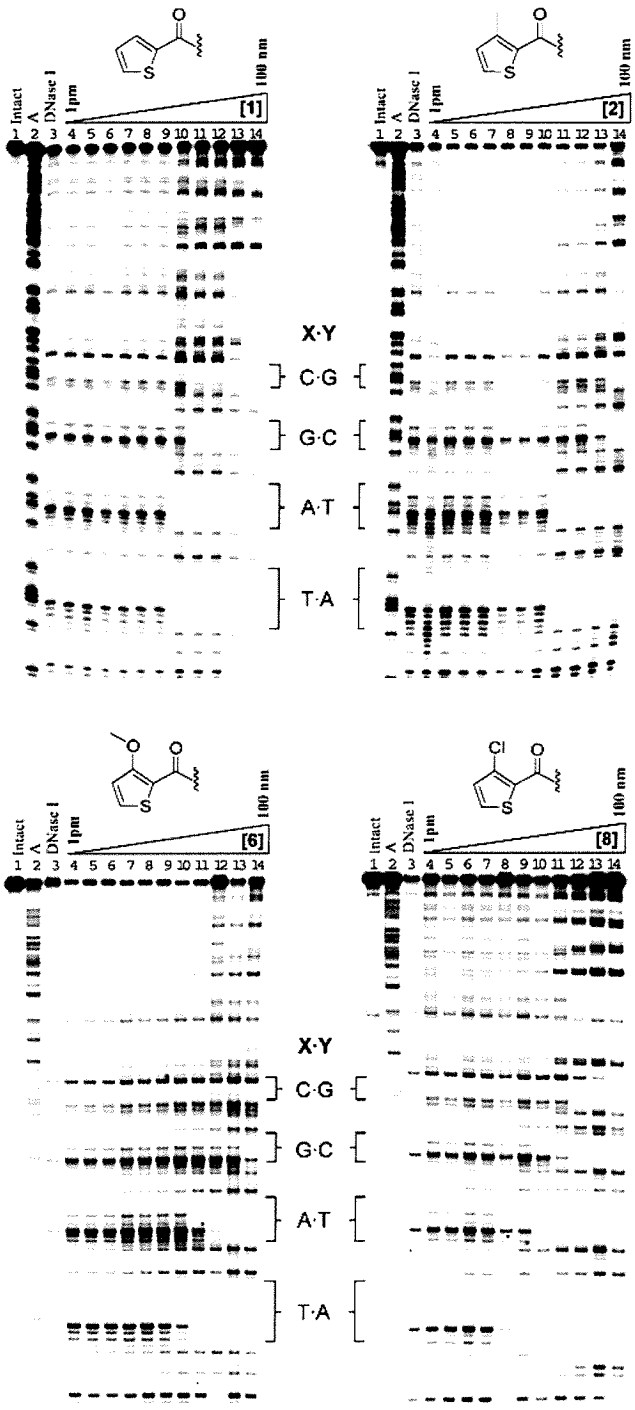


Figure 4.5 Quantitative DNase I footprint titration experiments for polyamides **1**, **2**, **6**, and **8** on pCW15 PCR product. Lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. The chemical structure of each N-terminal residue is included at the top of the gel and the four binding sites are labeled.

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

Polyamide	Ring Pairing	R	T·A	A·T	G·C	C·G
1	**Tp(1) / Py**	H	6.0 (0.7) x 10^9	4.7 (0.7) x 10^9	4.3 (0.4) x 10^8	2.2 (0.3) x 10^9
2	**Tp(2) / Py**	CH_3	2.3 (0.4) x 10^9	1.4 (0.2) x 10^9	1.0 (0.4) x 10^7	1.0 (0.3) x 10^7
3	**Tp(3) / Py**	NH_2	6.3 (1.0) x 10^9	4.6 (0.6) x 10^9	7.8 (0.9) x 10^8	2.2 (0.3) x 10^8
4	**Tp(4) / Py**	NHAc	5.9 (0.3) x 10^9	2.9 (0.1) x 10^9	6.6 (0.4) x 10^8	6.0 (0.2) x 10^8
5	**Tp(5) / Py**	OH	6.2 (0.6) x 10^9	4.5 (0.6) x 10^9	2.1 (0.3) x 10^8	8.4 (0.1) x 10^7
6	**Tp(6) / Py**	OCH_3	2.0 (0.4) x 10^9	3.2 (0.6) x 10^8	$\leq$ 1.0 x 10^7	$\leq$ 1.0 x 10^7
7	**Tp(7) / Py**	F	1.2 (0.2) x 10^{10}	3.9 (0.3) x 10^9	3.7 (0.4) x 10^8	2.9 (0.3) x 10^8
8	**Tp(8) / Py**	Cl	1.3 (0.2) x 10^{10}	3.7 (0.2) x 10^9	3.1 (0.6) x 10^8	2.1 (1.1) x 10^8
9	**Im / Py**		3.8 (0.3) x 10^9	2.8 (0.2) x 10^9	7.0 (0.9) x 10^{10}	3.2 (0.4) x 10^9
10	**Py / Py**		5.1 (0.6) x 10^9	3.1 (0.3) x 10^9	1.1 (0.1) x 10^9	2.6 (0.3) x 10^8

[a] Values reported are mean results determined by at least three DNase I footprint titrations, with standard deviation given in parentheses. Assays were performed at 22 °C in a buffer containing 10 mM Tris·HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0.

with high affinity (K_a = 7 x 10^{10} M^{-1}) while showing > 15-fold preference for G•C relative to T•A, A•T, and C•G base pairs. Terminal Py/Py pairings **10**, on the other hand, are characterized by little sequence specificity, binding T•A, A•T, and G•C with comparable affinity.

Within the thiophene-2-carboxamide series, *an unsubstituted thiophene ring* Tp **1** paired with Py shows little sequence specificity. Addition of a methyl group at the 3-position exerts a dramatic effect on sequence specificity; A,T favored over G,C. Polyamide **2** binds both T•A and A•T with a 140-fold preference for T,A relative to G,C. Amino **3**, acetamido **4**, or hydroxyl **5** substituents at the 3-position of thiophene all distinguish T,A from G,C but again do not distinguish T•A from A•T. Remarkably, a 3-methoxythiophene **6** paired with Py shows good affinity for T•A (K_a = 2 x 10^9 M^{-1}) with 6-fold selectivity for T•A relative to A•T and > 200-fold specificity relative to G,C. Fluoro **7** and chloro **8** substituted thiophene paired with Py afford higher binding affinities for T•A but a lower selectivity (3-fold) for T•A over A•T.

Molecular modeling (Table 2)

Molecular modeling was performed using the *Spartan Essential* software package.[15] N-terminal residues were first minimized as methyl-2-carboxamides using an AM1 model. The resulting geometry was then subjected to *ab initio* calculation using the Hartree-Fock model with a 6-31G* polarization basis set. The partial electrostatic charge of the sulfur atom, δ_S, and the partial charge of the peripheral atom of the 3-substituent, δ_R, were examined for each novel thiophene residue. The electronic influences of 3-substituents on the polarization of the sulfur atom follow expected trends, with partial electronic charge, δ_S, decreasing as follows: **4 > 7 > 1 > 6 > 5 > 8 > 2 > 3**. The electronic surfaces presented by the 3-substituents, δ_R, were also calculated and found to decrease as follows: **5 > 3 > 4 > 1 > 6 > 2 > 7 > 8**.

Table 2. Physical Properties Determined by Molecular Modeling[a]

Polyamide	R	δ_S[b]	δ_R[b], (R)[c]	$E_{syn} - E_{anti}$[d]	A_R / A_H[e]
1	H	-0.065	0.124, (**CH**)	0.262	1.00
2	CH_3	-0.093	0.036, ($\mathbf{CH_3}$)	1.739	1.11
3	NH_2	-0.117	0.426, ($\mathbf{NH_2}$)	10.289	1.07
4	NHAc	-0.057	0.320, (**NH**)	10.308	1.31
5	OH	-0.069	0.512, (**OH**)	7.142	1.04
6	OCH_3	-0.068	0.063, ($\mathbf{OCH_3}$)	- 7.298	1.17
7	F	-0.061	-0.227, (**F**)	- 5.043	1.03
8	Cl	-0.076	-0.106, (**Cl**)	-13.293	1.09

[a] *Ab initio* calculations were performed with Spartan Essential software package using Hartree-Fock model with 6-31G* polarization basis set. [b] Partial electrostatic charges are given in abrbitrary units. [c] Partial charges given for atoms in bold. [d] Energy differences are reported in kcal/mol. [e] Ratio of surface area, A, of 3-substituent to hydrogen.

The relative energy differences between minimized syn and anti conformations was also examined for each new thiophene ring. Hairpins **1-5** show a preference for the anti, or "sulfur down," conformation which may be attributed to lone pair repulsions

between the sulfur atom and the carbonyl oxygen of the 2-carboxamide moiety. This bias can be reinforced by favorable hydrogen bonding interactions between 3-substituents and the carboxamide as in polyamides **3-5**. By contrast, polyamides **6-8** display a bias for the syn, or "sulfur up," conformation, possibly owing to more severe electronic clashes between the electron rich 3-substituents and the carboxamide relative to those of the sulfur atom. Finally, the solvent exposed surface area of each 3-substituted thiophene was compared to the unsubstituted thiophene ring to assess the steric contribution of the 3-substituent and surface area was found to increase in the following order: **1 < 7 < 5 < 3 < 8 < 2 < 6 < 4**.

4.3 Discussion

The observed equilibrium association constants for polyamides **1-5** support an anti conformation for the N-terminal thiophene residue. The binding preference of these compounds for T•A / A•T relative to G•C / C•G might be a result of unfavorable steric clashes between the sulfur atom and the exocyclic amino group of guanine. The binding properties of N-terminal, 3-methylthiophene-2-carboxamide residues also correlate well with values derived from internal contexts, where the sulfur down conformation is stringently enforced.[11]

It was envisioned that polyamides **6-8** would assume an anti conformation by sterically disfavoring contact between the bulky 3-substituents and the floor of the minor groove. Quantitative DNase I footprinting revealed modest selectivity for T•A relative to A•T and excellent specificity for both of these over G•C / C•G. However, binding properties of 3-methoxy- and 3-chlorothiophene residues determined at the N-terminus

do not correlate with those derived from internal positions. Furthermore, molecular modeling indicated that unfavorable lone pair interactions favor the sulfur up conformation. Taken together, these results could suggest that the *electron rich methoxy and halogen groups are projected toward the minor groove* (**Figure 4.6**). The greater size of the methoxy group relative to the halogens might account for the lower affinity of polyamide **6** relative to **7** and **8**. The greater T•A selectivity of **6** could stem from the more complementary positive electronic surface presented to the thymine carbonyl by the methoxy protons relative to the negatively polarized halogen atoms.

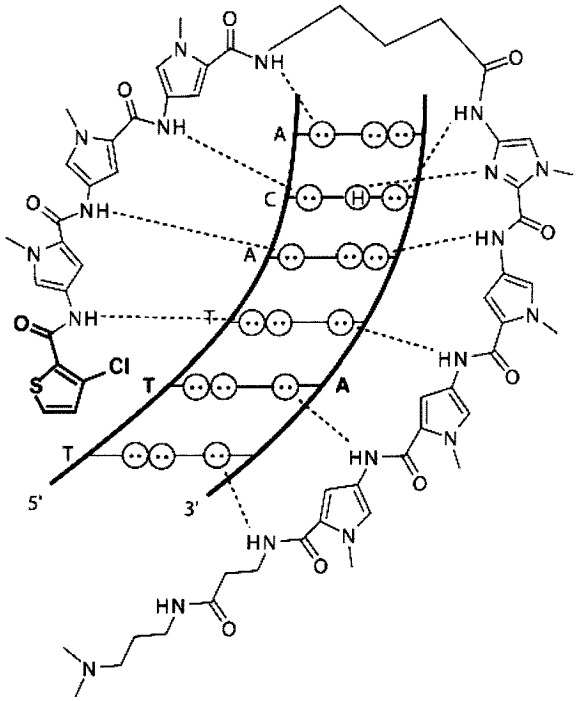


Figure 4.6 Hypothetical binding model to explain selectivity for T•A over A•T. The thiophene is in the *anti* conformation (sulfur away from the groove) and the 3-chloro substituent points to the minor groove floor.

N-terminal 3-methoxy (or 3-chloro) thiophene-2-carboxamide residues when paired with Py demonstrate selectivity for T•A versus A•T. This represents an important step toward expanding the array of DNA sequences that can be targeted by minor groove-binding polyamides. Confirmation of the novel binding model with chloro (or methoxy)

responsible for minor groove shape selective recognition, not sulfur, must await structure studies in solution by NMR.

4.4 Experimental

Materials

Methyl-2-chloroacrylate was obtained from Acros. Benzenethiol, di-*tert*-butyl dicarbonate (Boc_2O), fluorotrichloromethane (NMR grade), iodomethane, methyl thioglycolate, 3-methyl-2-thiophenecarboxylic acid, N,N-diisopropylethylamine (DIEA), N,N-dimethylaminopropylamine (Dp), N,N-dimethylaminopyridine (DMAP), N,N-dimethylformamide (DMF), N-fluorobenzenesulfonimide, potassium carbonate, sodium metal, tetrahydrofuran (THF), 2-thiophenecarboxylic acid, and triethylamine (TEA) were purchased from Aldrich. Methyl 3-amino-2-thiophenecarboxylate and 3-chloro-2-thiophenecarboxylic acid were obtained from Alfa Aesar. Boc-β-alanine-(4-carbonylaminomethyl)-benzyl-ester-copoly(styrene-divinylbenzene)resin (Boc-β-Pam-resin) and 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were purchased from NOVA Biochem. Trifluoroacetic acid (TFA) was purchased from Halocarbon. *N*-Butyllithium was obtained as a solution in hexanes from Strem. All other solvents were reagent grade from EM.

NMR spectra were recorded on a Varian spectrometer at 300 MHz in DMSO-d_6 or $CDCl_3$ with chemical shifts reported in parts per million relative to residual solvent. Fluorotrichloromethane was used as an internal standard for ^{19}F NMR. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. High-resolution EI mass spectra were recorded at the Mass Spectrometry Laboratory at the California Institute of Technology. Matrix-assisted, laser desorption/ionization time of

flight mass spectrometry (MALDI-TOF-MS) was conducted at the Protein and Peptide Micronanalytical Facility at the California Institute of Technology.

Monomer synthesis

3-[(tert-butoxy)carbonylamino]-2-thiophenecarboxylic acid (**11**). A mixture of methyl 3-amino-2-thiophenecarboxylate (2.53 g, 15.9 mmol), Boc_2O (7.64 g, 35 mmol), and DMAP (2.04 g, 16.7 mmol) was dissolved in acetone (15 mL) and TEA (5 mL). The reaction mixture was stirred vigorously for 4 hours and diluted to a volume of 75 mL with dichloromethane. The resulting solution was washed with cold 1N HCl (3 x 50 mL), 1N NaOH (3 x 50 mL), and brine (50 mL). The dichloromethane solution was then dried over $MgSO_4$, filtered, and concentrated *in vacuo* to yield a yellow oil. The crude product was loaded onto a short plug of silica and eluted with 9:1 hexanes/ethyl acetate to yield a pale yellow solid (1.2 g) that was used without further purification. The solid was dissolved in methanol (76 mL) and 50% NaOH (4 mL) and the mixture was stirred for 4 hours. The reaction was diluted to a volume of 160 mL with water and concentrated briefly *in vacuo*. The remaining aqueous solution was washed with diethyl ether (2 x 80 mL), cooled in an ice bath, and cautiously acidified to pH 2 with sulfuric acid. The suspension was washed with ethyl acetate (3 x 50 mL) and the combined organic washes were dried over $MgSO_4$, filtered, and concentrated *in vacuo* to yield (**11**) as a white solid (0.79 g) in 69% yield over two steps. ^{1}H NMR (d_6-DMSO) δ 9.43 (s, 1H), 7.80 (d, J = 5.4 Hz, 1H), 7.72 (d, J = 5.4 Hz, 1H), 1.46 (s, 9H) ; ^{13}C NMR (75 MHz, d_6-DMSO) δ 165.8, 151.8, 144.9, 133.1, 121.2, 109.9, 81.5, 28.6 ; EI-MS m/e 243.0563 (M^+ calculated 243.0565 for $C_{10}H_{13}NO_4S$).

Methyl 3-hydroxy-2-thiophenecarboxylate (**12**). To dry methanol (81 mL), under nitrogen, was added sodium metal (3.68 g, 304 mmol). After H_2 evolution has ceased, the solution was cooled to 0 °C and methyl thioglycolate (10 g, 179 mmol) was added dropwise. A solution of methyl-2-chloroacrylate (10.88 g, 179 mmol) in methanol (21 mL) was then added slowly, resulting in the formation of yellow precipitate. The solution was allowed to warm to ambient temperature and stirred for 2 hours. The solvent was removed in vacuo to give a dark yellow solid that was acidified to pH 2 with 4N HCl. The resulting aqueous solution was extracted with dichloromethane (3 x 150 mL) and the combined organic solutions were washed with water (3 x 150 mL), dried over $MgSO_4$, filtered, and concentrated to give a dark oil. The oil was subjected to column chromatography on silica gel (20:1 hexanes/ethyl acetate) to give (**12**) (18.4 g) as a crystalline solid in 64% yield. TLC (20:1 hexanes/ethyl acetate) R_f 0.47; ^{1}H NMR ($CDCl_3$) δ 9.58 (s, 1H), 7.59 (d, J = 5.7 Hz, 1H), 6.75 (d, J = 4.8 Hz, 1H), 3.90 (s, 3H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 166.8, 164.7, 131.7, 119.4, 52.2; EI-MS m/e 158.0039 (M+ calculated 158.0038 for $C_6H_6O_3S$).

3-Methoxy-2-thiophenecarboxylic acid (**13**). A mixture of (**12**) (2.3 g, 14.5 mmol), K_2CO_3 (5.02 g, 36.3 mmol), and iodomethane (10.4 g, 73 mmol) was suspended in acetone (25 mL) and acetonitrile (5 mL). The resulting mixture was stirred vigorously at reflux for 3 hours. The reaction was filtered and the resulting solid was washed with acetone and dichloromethane. The reaction and washes were combined and concentrated *in vacuo* to yield a yellow solid (1.9 g) that was used without further purification. The yellow solid was dissolved in methanol (17 mL) and 50% NaOH (3 mL) and was stirred for 3 hours. The reaction was diluted to 40 mL with water and concentrated briefly *in*

vacuo to yield a suspension. The aqueous suspension was washed with diethyl ether (2 x 25 mL), cooled to 0 °C, and acidified to pH 2 with 10% sulfuric acid. The aqueous mixture was then washed with dichloromethane (3x 50 mL) and the combined organic washes were dried over sodium sulfate, filtered, and concentrated *in vacuo* to give a yellow oil. The oil was suspended in 3:1 petroleum ether/dichloromethane at –20 °C overnight. Filtration gave (**13**) as a finely divided white solid (0.736 g) in 33% yield over two steps. TLC (4:1 ethyl acetate/hexanes) R_f 0.5; ^{1}H NMR (DMSO-d_6) δ 12.4 (s, 1H), 7.74 (d, J = 5.7 Hz, 1H), 7.06 (d, J = 5.4 Hz, 1H), 3.85 (s, 3H); ^{13}C NMR (75 MHz, DMSO-d_6) δ 163.0, 161.9, 131.9, 118.0, 109.9, 59.4; EI-MS m/e 158.0034 (M+ calculated 158.0038 for $C_6H_6O_3S$).

3-Fluoro-2-thiophenecarboxylic acid (**14**). 2-Thiophenecarboxylic acid (1.7 g, 13.3 mmol) was dissolved in anhydrous THF (30 mL) and the solution was cooled to –78 °C under Ar, with stirring. *n*-Butyllithium (18.3 mL, 29.3 mmol) in hexanes was added to the above solution and the mixture was stirred for 30 minutes. A solution of N-fluorobenzenesulfonimide (5 g, 15.9 mmol) in THF (30 mL) was then added and the resulting solution was stirred at –78 °C for 4 hours and allowed to warm to ambient temperature over a period of 6 hours. The reaction was diluted with diethyl ether (100 mL), cooled to 0 °C, and 1N HCl (15 mL) was added to give a biphasic mixture. The aqueous layer was isolated and washed with diethyl ether (3 x 50 mL). The combined ethereal layers were dried over $MgSO_4$, filtered, and concentrated *in vacuo* to yield an orange oil. The oil was subjected to column chromatography on silica gel using 1:1 hexanes/ethyl acetate as the eluent. (**14**) was obtained as a slightly brown solid (0.777 g) in 40% yield. TLC (1:1 ethyl acetate/hexanes) R_f 0.17; ^{1}H NMR ($CDCl_3$) δ 10.7 (s, 1H),

7.53 (dd, J = 5.4, 3.6 Hz, 1H), 6.89 (d, J = 5.4 Hz, 1H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 166.2 (d, J = 3.5 Hz), 161.5 (d, J = 278 Hz), 132.0 (d, J = 10 Hz), 118.9 (d, J = 24.7 Hz), 113.6; ^{19}F NMR (282 MHz, $CDCl_3$, $CFCl_3$) δ - 65.2 (d, J = 6 Hz) ; EI-MS m/e 145.9838 (M+ calculated 145.9838 for $C_5H_3FO_2S$).

Polyamide synthesis

Hairpin polyamides were synthesized from intermediate resin (**R1**) that was prepared according to published protocols using Boc-β-alanine-Pam resin (50 mg, 0.59 mmol/g).[14] Products were purified by reversed-phase HPLC and characterized by MALDI-TOF mass spectrometry.

(**1**). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of 2-thiophenecarboxylic acid (19 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 minutes and poured onto the deprotected resin. The resin slurry was shaken for 4 hours at room temperature and filtered. After washing with DMF, the resin was cleaved with Dp (1 mL) at 40 °C for 4 hours. The crude product was purified by reversed phase HPLC to afford (**1**) as a white solid upon lyophilization (3.3 mg, 9% recovery). MALDI-TOF-MS m/z 1224.23 (1224.53 calcd for M + H).

(**2**). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of 3-methyl-2-thiophenecarboxylic acid (21 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 minutes and poured onto the deprotected resin. The resin slurry was shaken for 4 hours at room temperature and filtered. After washing with DMF, the resin was cleaved

with Dp (1 mL) at 40 °C for 4 hours. The crude product was purified by reversed phase HPLC to afford (**1**) as a white solid upon lyophilization (3.0 mg, 8.2% recovery). MALDI-TOF-MS m/z 1238.35 (1238.54 calcd for M + H).

(**2**). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of 3-methyl-2-thiophenecarboxylic acid (21 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 minutes and poured onto the deprotected resin. The resin slurry was shaken for 4 hours at room temperature and filtered. After washing with DMF, the resin was cleaved with Dp (1 mL) at 40 °C for 4 hours. The crude product was purified by reversed phase HPLC to afford (**2**) as a white solid upon lyophilization (3.0 mg, 8.2% recovery). MALDI-TOF-MS m/z 1238.35 (1238.54 calcd for M + H).

(**3**). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of (**11**) (36 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 minutes and poured onto the deprotected resin. The resin slurry was shaken for 4 hours at room temperature and filtered. After washing with DMF and dichloromethane, the resin was treated with 80% TFA in dichloromethane. The resin was filtered and washed before cleavage with Dp (1 mL) at 40 °C for 4 hours. The crude product was purified by reversed phase HPLC to afford (**3**) as a slightly yellow solid upon lyophilization (3.4 mg, 9.4% recovery). MALDI-TOF-MS m/z 1239.46 (1239.54 calcd for M + H).

(**4**). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of (**11**) (36 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 minutes and poured onto

the deprotected resin. The resin slurry was shaken for 4 hours at room temperature and filtered. After washing with DMF and dichloromethane, the resin was treated with 80% TFA in dichloromethane. The resin was filtered, neutralized and shaken in a solution of acetic anhydride (0.2 mL), DIEA (0.2 mL) and DMF (1.6 mL) for 30 minutes. The resin was then filtered and washed with DMF before cleavage with Dp (1 mL) at 40 °C for 4 hours. The crude product was purified by reversed phase HPLC to afford (**4**) as a pale yellow solid upon lyophilization (4.2 mg, 11.2% recovery). MALDI-TOF-MS m/z 1281.62 (1281.55 calcd for M + H).

(**5**). A solution of sodium hydride (40 mg, 60% oil dispersion) and thiophenol (0.1 mL) in DMF (0.15 mL) was heated to 100 °C and a solution of (**6**) (1.3 mg, 1 μmol) in DMF (0.25 mL) was added. After 2 hours, the reaction mixture was cooled to 0 °C and 20% TFA in water (7 mL) was added. The aqueous solution was washed three times with diethyl ether (8 mL) and was subjected to preparative, reversed-phase HPLC to afford (**5**) as a white solid upon lyophilization (0.6 mg, 50% recovery). MALDI-TOF-MS m/z 1241.09 (1240.52 calcd for M + H).

(**6**). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of (**13**) (23 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 minutes and poured onto the deprotected resin. The resin slurry was shaken for 4 hours at room temperature and filtered. After washing with DMF and dichloromethane, the resin was cleaved with Dp (1 mL) at 40 °C for 4 hours. The crude product was purified by reversed phase HPLC to afford (**6**) as a white solid upon lyophilization (3.3 mg, 8.9% recovery). MALDI-TOF-MS m/z 1255.96 (1255.39 calcd for M + H).

(**7**). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of (**14**) (22 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 minutes and poured onto the deprotected resin. The resin slurry was shaken for 4 hours at room temperature and filtered. After washing with DMF and dichloromethane, the resin was cleaved with Dp (1 mL) at 40 °C for 4 hours. The crude product was purified by reversed phase HPLC to afford (**7**) as a white solid upon lyophilization (2.6 mg, 7.0% recovery). MALDI-TOF-MS m/z 1242.20 (1242.52 calcd for M + H).

(**8**). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of 3-chloro-2-thiophenecarboxylic acid (24 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 minutes and poured onto the deprotected resin. The resin slurry was shaken for 4 hours at room temperature and filtered. After washing with DMF and dichloromethane, the resin was cleaved with Dp (1 mL) at 40 °C for 4 hours. The crude product was purified by reversed phase HPLC to afford (**8**) as a white solid upon lyophilization (3.8 mg, 10.1% recovery). MALDI-TOF-MS m/z 1258.86 (1258.49 calcd for M + H).

(**9**). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of 2-trichloroacetyl-1-methylimidazole (34 mg, 0.148 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was poured onto the deprotected resin. The resin slurry was shaken for 4 hours at 40 °C and filtered. After washing with DMF and dichloromethane, the resin was cleaved with Dp (1 mL) at 40 °C for 4 hours. The crude product was purified by reversed phase HPLC to afford (**9**) as a yellow solid upon

lyophilization (2.5 mg, 6.9% recovery). MALDI-TOF-MS m/z 1222.03 (1222.58 calcd for M + H).

(**10**). Resin (**R1**) was treated with 80% TFA in dichloromethane and washed thoroughly. A solution of N-methylpyrrole-2-carboxylic acid (19 mg, 0.148 mmol) and HBTU (28 mg, 0.079 mmol) in DMF (0.45 mL) and DIEA (0.5 mL) was mixed at 40 °C for 25 minutes and poured onto the deprotected resin. The resin slurry was shaken for 4 hours at room temperature and filtered. After washing with DMF and dichloromethane, the resin was cleaved with Dp (1 mL) at 40 °C for 4 hours. The crude product was purified by reversed phase HPLC to afford (**10**) as a white solid upon lyophilization (2.7 mg, 7.5% recovery). MALDI-TOF-MS m/z 1222.12 (1221.58 calcd for M + H).

DNA reagents and materials

Oligonucleotide primers SF1 (5'-AATTCGAGCTCGGTACCGGGG-3') and SF2 (5'-CTGGCACGACAGGTTTCCCGA-3') were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology. Products from PCR amplification of the pCW15 using 5'-[γ-^{32}P]-labeled SF1 and SF2 were purified on a 7% non-denaturing polyacrylamide gel. Glycogen (20 mg/mL), dNTPs (PCR nucleotide mix), and all enzymes, unless otherwise stated, were purchased from Boehringer-Mannheim. Deoxyadenosine [γ-^{32}P]triphosphate was obtained from ICN. Calf thymus DNA (sonicated, deproteinized) and DNase I (7500 units/mL, FPLC pure) were from Amersham Pharmacia. AmpliTaq DNA polymerase was obtained from Perkin-Elmer and was used with provided buffers. Tris•HCl, DTT, RNase-free water, and 0.5 M EDTA were from United States Biochemical. Calcium chloride, potassium chloride, and

magnesium chloride were purchased from Fluka. Tris-borate-EDTA was from GIBCO and bromophenol blue was from Acros. All reagents were used without further purification.

DNase I footprinting experiments were performed according to standard protocols.[16]

4.5 References

[1] Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783.
[2] Dervan, P. B.; Burli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688.
[3] Dervan, P. B. *Bioorg. & Med. Chem.* **2001**, *9*, 2215.
[4] Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663.
[5] Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559.
[6] White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569.
[7] Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104.
[8] White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; *Nature* **1998**, *391*, 468.
[9] Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111.
[10] Ellervik, U.; Wang, C. C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 9354.
[11] Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B. *Helv. Chim. Acta.* **2002**, *85*, 4485.
[12] Huddleston, P. R.; Barker, J. M. *Synth. Com.* **1979**, *9*, 731.
[13] Taylor, E. C.; Ping, Z. *Org. Prep. Proc. Int.* **1997**, *29*, 221.
[14] Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.
[15] SPARTAN ESSENTIAL Copyright © 1991-2001 by Wavefunction Inc.
[16] Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450.

Chapter 5

Expanding the Repertoire of Heterocycle Ring Pairs for Programmable Minor Groove DNA Recognition

The text of this chapter was taken in part from a manuscript coauthored with Michael A. Marques, Shane Foister and Professor Peter B. Dervan (Caltech).

(Marques, M. A.; Doss, R. M.; Foister, S. F.; and Dervan, P. B. "*Expanding the Repertoire of Heterocycle Ring Pairs for Programmable Minor Groove DNA Recognition*" *JACS*, **2004**, *126*, 10339-10349.)

Abstract.

The discrimination of the four Watson-Crick base pairs by minor groove DNA binding polyamides have been attributed to the specificity of three five-membered aromatic amino acid subunits, 1-methyl-1*H*-imidazole (**Im**), 1-methyl-1*H*-pyrrole (**Py**), and 3-hydroxy-1*H*-pyrrole (**Hp**) paired four different ways. The search for additional ring pairs that demonstrate DNA sequence specificity has led us to a new class of 6-5 fused bicycle rings as minor groove recognition elements. The affinities and specificities of the hydroxybenzimidazole/pyrrole (**Hz/Py**) and hydroxybenzimidazole/benzimidazole (**Hz/Bi**) pairs for each of the respective Watson-Crick base pairs within the sequence context 5'-TGG**X**CA-3' (**X** = A, T, G, C) were measured by quantitative DNase-I footprinting titrations. The **Hz/Py** and **Hz/Bi** distinguish T·A from A·T. Hairpin polyamides containing multiple **Hz/Py** pairs were examined and shown to mimic the **Hp/Py** pair with regard to affinity and specificity. Therefore the **Hz/Py** pair may be considered a second generation replacement for the **Hp/Py** pair.

5.1 Introduction.

Aberrant gene expression is the cause of many diseases and the ability to reprogram transcriptional pathways using cell permeable small molecules may one day have an impact on human medicine.[1] DNA-binding polyamides, which are based on the architecture of the natural products netropsin and distamycin A,[2,3] are capable of distinguishing all four Watson-Crick base pairs in the DNA minor groove.[4,5] Sequence specific recognition of the minor groove of DNA arises from the pairing of two different antiparallel 5-membered heterocyclic amino acids and the interplay of a variety of direct and indirect recognition elements.[4] The direct read out, or information face, on the inside of the crescent shaped polyamide may be programmed by the incremental change of atoms on the corners of the ring pairs presented to the DNA minor groove floor. Stabilizing and destabilizing interactions with the different edges of the four Watson-Crick bases are modulated by shape complimentarity and specific hydrogen bonds.[6,7] For example, the imidazole ring **Im**, which presents a lone pair of electrons to the DNA minor groove, can accept a hydrogen bond from the exocyclic amine of guanine.[6] Additionally, the 3-hydroxypyrrole ring **Hp** projects an exocyclic OH group toward the minor groove floor that is sterically accommodated in the cleft of the T·A base pair, preferring to lie opposite T not A.[6c] From x-ray structural analysis, it appears that **Hp** can form two hydrogen bonds with the O(2) of thymine. [6c] Molecular recognition of DNA by polyamides is also affected by a series of critical ligand-DNA interactions that take place away from the polyamide recognition face. Polyamide geometry with respect to overall curvature, rise per residue and contacts between the polyamide and the walls of the minor groove, as well as sequence dependent energetic penalties which arise from distorting DNA from its low energy unbound conformation, are examples of how

sequence dependant microstructure and flexibility of DNA may influence indirectly the affinity and specificity of polyamides.[7]

Previously, pairings using **Im**, **Py**, and **Hp** have been used to discriminate the Watson-Crick base pairs such that: **Im/Py** is specific for G·C and **Hp/Py** for T·A.[4] These pairing rules have proven useful for the recognition of hundreds of DNA sequences by these programmable oligomers. However, sequence dependant changes in the microenvironment of the DNA makes the targeting of certain sequences difficult, leading us to explore whether other novel heterocyclic recognition elements could be discovered for use in DNA groove recognition within the pairing paradigm.[8] Furthermore, replacement of the **Hp** ring system, which was a benchmark for the field with regard to its ability to distinguish T·A from A·T, was a priority due to the subsequent observation that hairpin oligomers containing **Hp** degrade over time in the presence of acid or trace free radicals.[9] We recently reported that the benzimidazole architecture can be an effective platform for the development of novel modular recognition elements for the minor groove of DNA.[10] The benzimidazole 6-5 bicyclic-ring structure, while having different curvature from the classic 5-membered heterocyclic carboxamides, presents an "inside edge" with a similar readout shape to the DNA minor groove floor, effectively mimicking **Py**, **Im**, and **Hp** heterocycle-carboxamides (**Figure 5.1**).

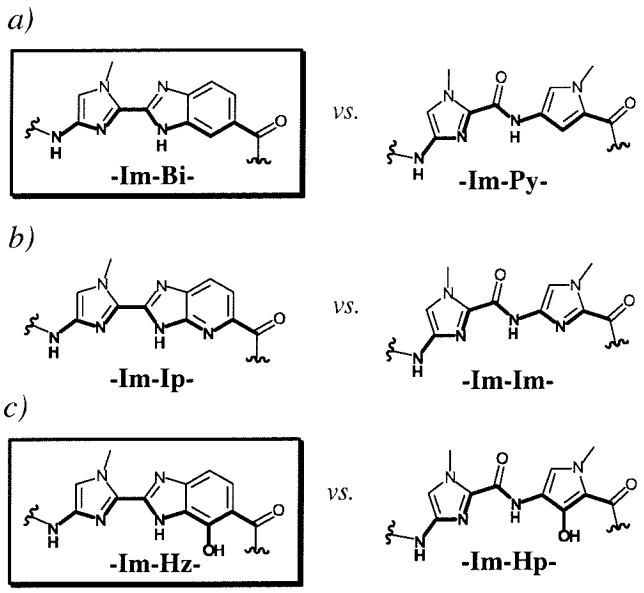


Figure 5.1 Structures of the (a) benzimidazole (**Bi**), (b) imidazolpyridine (**Ip**), and (c) hydroxybenzimidazole (**Hz**) building blocks in comparison with their respective five-membered ring dimers. Hydrogen bonding surfaces presented to the DNA-minor-groove floor are bolded.

Here we compare the hydroxybenzimidazole bicycle **Hz** with respect to the **Hp** amino acid at discriminating between the Watson-Crick bases within several different sequence contexts (**Figure 5.2 a,b**). Experiments were designed to elucidate the DNA recognition properties of the **Hz/Py** pair, and the ability of multiple **Hz/Py** pairs to distinguish multiple A·T sequences. We also report that the 6-5/6-5 bicyclic pair **Hz/Bi** is effective for minor groove recognition (**Figure 5.2c**). For the series of experiments, this required the synthesis of two new fused ring dimers **Bi-Im** and **Hz-Im** which would mimic the **Py-Im** and **Hp-Im** dimers. These were incorporated into the synthesis of hairpin oligomers **1** – **6** (**Figure 5.3**). DNaseI footprinting titrations were used to determine DNA binding affinities of hairpins **1** – **6** which will reveal the energetic preferences of the **Hz/Py** and **Hz/Bi** pairs for the four Watson-Crick bases, as well as determine the fidelity of hairpin polyamides containing multiple **Hz/Py** pairs. Molecular modeling was used to further analyze the binding data.

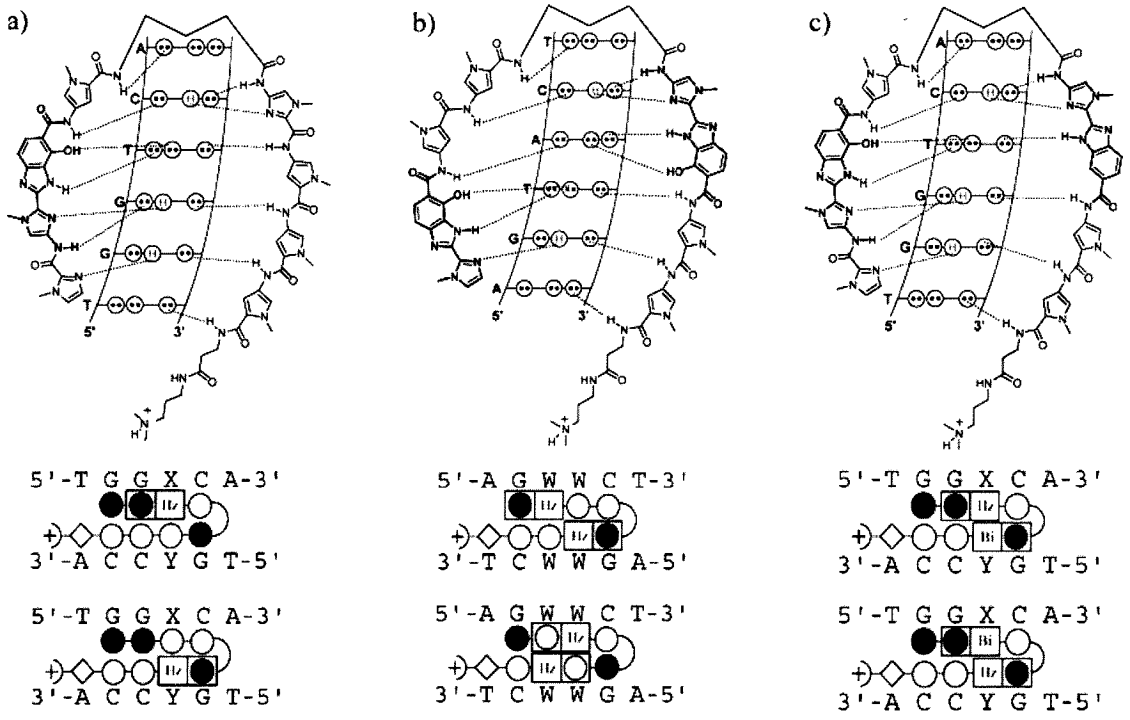


Figure 5.2 Hairpin polyamides and the minor groove contacts for their match sequences along with experimental schemes showing the DNA sequences that each polyamide will be tested against. (a) Hz/Py and Py/Hz pairs tested against the four Watson-Crick base pairs. (b) multiple Hz/Py pairs tested against different core A,T sequences. (c) Hz/Bi and Bi/Hz pairs tested against the four Watson-Crick base pairs. Dark circle = imidazole; Light circle = pyrrole; Half circle = γ; Diamond = β-Alanine; Half circle with a plus = dimethylaminopropylamide tail.

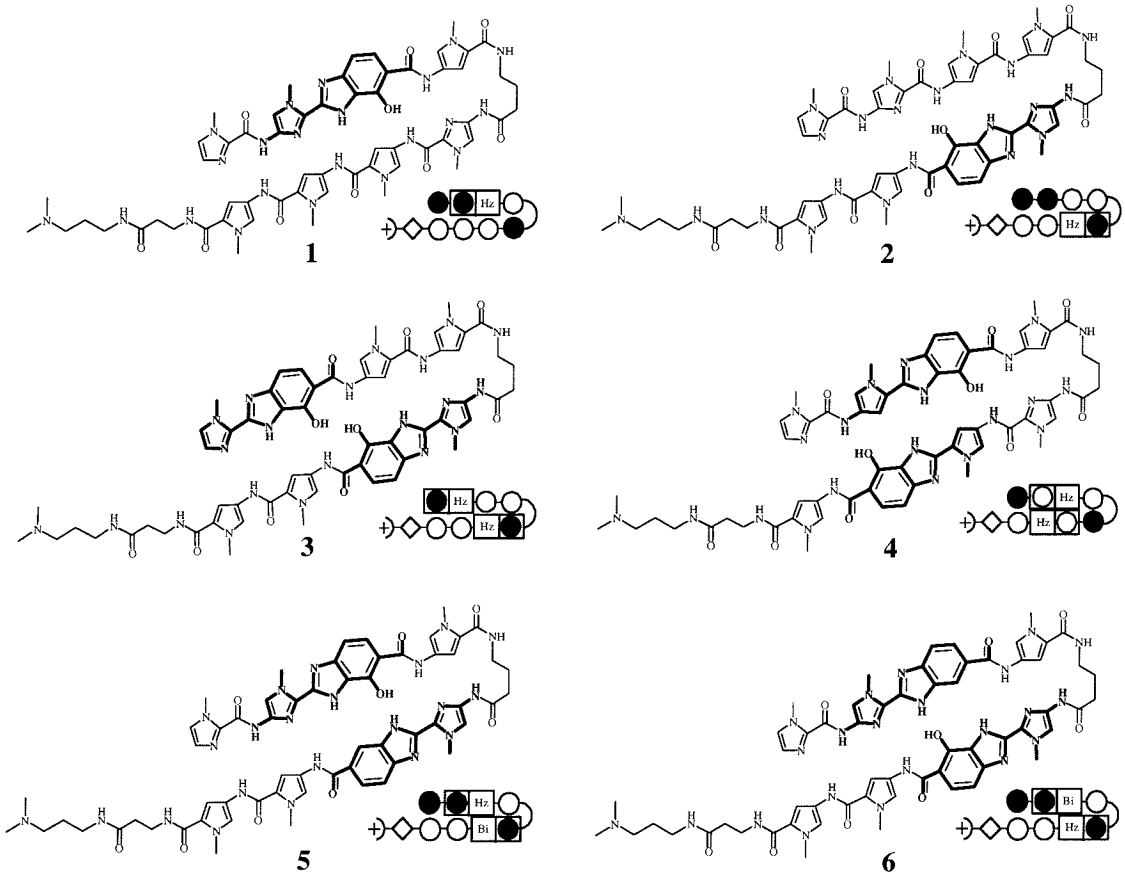


Figure 5.3 Structures of the polyamides containing Hz/Py, Hz/Bi, and multiple Hz/Py pairs shown along with their ball and stick representation. Shaded and non-shaded circles indicate imidazole and pyrrole respectively, whereas hydroxybenzimidazole and benzimidazole are indicated as Hz and Bi. The half circle represents γ-aminobutyric acid linker, while the diamond indicates β-Alanine. The half circle with a plus indicates the dimethylaminopropylamide tail.

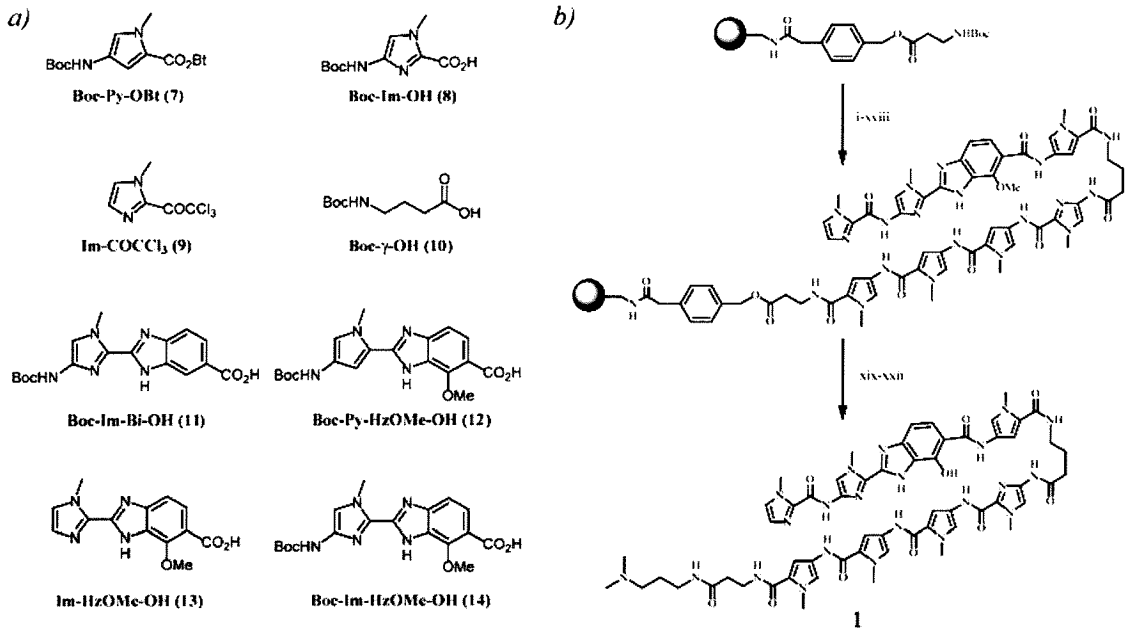


Figure 5.4 (a) Amino Acid Building Blocks for Polyamide Synthesis. (b) Representative solid phase synthesis of polyamide 1. Reaction conditions: (i) 80% TFA/DCM; (ii) Boc-Py-OBt, DIEA, DMF; (iii) Ac_2O, DIEA, DMF; (iv) Repeat i-iii x2; (v) 80% TFA/DCM; (vi) Boc-Im-OH, HBTU, DIEA, DMF; (vii) Ac_2O, DIEA, DMF; (viii) 80% TFA/DCM; (ix) Boc-γ-OH, HBTU, DIEA, DMF; (x) Ac_2O, DIEA, DMF; (xi) 80% TFA/DCM; (xii) Boc-Py-OBt, DIEA, DMF; (xiii) Ac_2O, DIEA, DMF; (xiv) 80% TFA/DCM; (xv) Boc-Im-HzOMe-OH, HBTU, DIEA, DMF; (xvi) Ac_2O, DIEA, DMF; (xvii) 80% TFA/DCM; (xviii) Im-$COCCl_3$, DIEA, DMF; (xix) dimethylaminopropylamine (Dp), 80 °C 2 h.; (xx) prep. HPLC; (xxi) thiophenol, NaH, DMF, 80 °C 2 h.; (xxii) prep. HPLC.

5.2 Results.

Polyamide Synthesis. Polyamides **1-6** were synthesized in stepwise fashion on β-Pam resin following manual solid phase methods.[11] The Boc-protected amino acids utilized for polyamide synthesis were Boc-**Py**-OBt (**7**), Boc-**Im**-OH (**8**), **Im**-$COCCl_3$ (**9**), Boc-γ-OH (**10**), Boc-Im-**Bi**-OH (**11**), Boc-Py-**Hz**OMe-OH (**12**), Im-**Hz**OMe-OH (**13**) and Boc-Im-**Hz**OMe-OH (**14**) (**Figure 5.4**). The synthetic route for the monomer units **13** and **14** are shown in **Figure 5.5**. Couplings were realized using pre-activated monomers (**7**) or HBTU activation in a DIEA and DMF mixture. Coupling times ran from 3-24 h at 25-40 °C. Deprotection of the polyamide was accomplished using 80% TFA/DCM. Polyamides were cleaved from the resin by treatment with dimethylaminopropylamine

(Dp) neat at 80 °C for 2 h, and purified by preparatory reverse phase HPLC. The methoxy substituted polyamides were deprotected with thiophenoxide in DMF at 80 °C, to provide the free hydroxy derivatives after a second HPLC purification: Im-Im-**Hz**-Py-γ-Im-**Py**-Py-Py-β-Dp (**1**), Im-Im-**Py**-Py-γ-Im-**Hz**-Py-Py-β-Dp (**2**), Im-**Hz**-**Py**-Py-γ-Im-**Hz**-**Py**-Py-β-Dp (**3**), Im-**Py**-**Hz**-Py-γ-Im-**Py**-**Hz**-Py-β-Dp (**4**), Im-Im-**Hz**-Py-γ-Im-**Bi**-Py-Py-β-Dp (**5**) and Im-Im-**Bi**-Py-γ-Im-**Hz**-Py-Py-β-Dp (**6**).

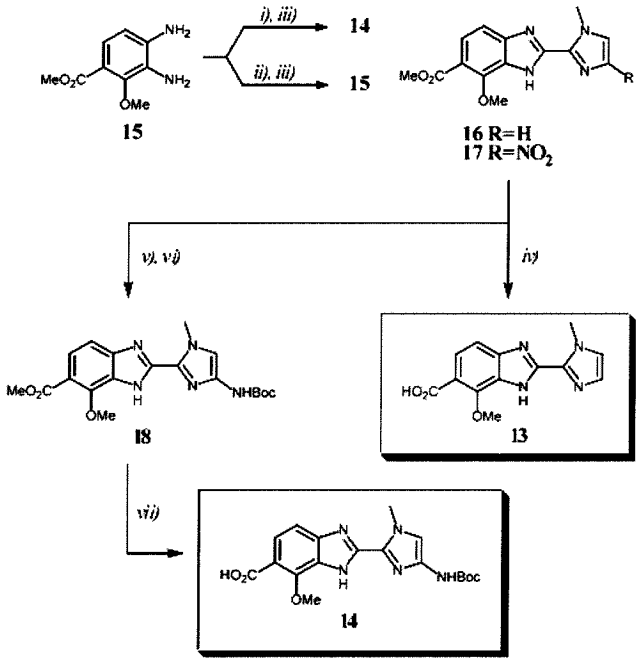


Figure 5.5 Synthesis of Im-HzOMe-OH (16) and Boc-Im-HzOMe-OH (18); (i) Im-$COCCl_3$(9)[12,13], HBTU, DIEA, DMF; (ii) NO_2-Im-OH (19), DIEA, EtOAc; (iii) AcOH; (iv) NaOH, MeOH; (v) H_2 Pd/C, DIEA, DMF; (vi) $(Boc)_2O$; (vii) NaOH, MeOH.

DNA Affinity and Sequence Specificity. Quantitative DNase-I-footprinting titrations were carried out for hairpin polyamides **1-6.** Polyamides **1**, **2**, **5** and **6** were footprinted on the 278-base-pair PCR product of plasmid pDHN1 (**Figure 5.6a**). Polyamides **3** and **4** were footprinted on the 285-base-pair PCR product of plasmid pDEH10 (**Figure 5.6b**).

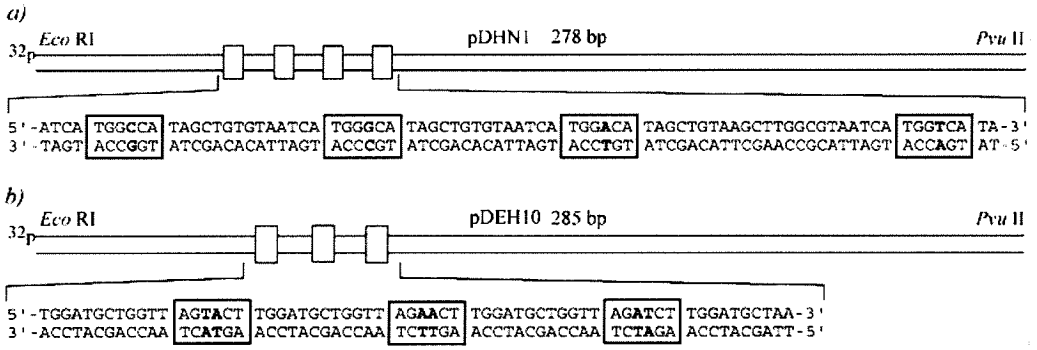


Figure 5.6 Illustration and complete sequence of the *Eco*RI/*Pvu*II restriction fragment derived from plasmids (a) pDHN1 and (b) pDEH10. For pDHN1, the four designed 6-base-pair binding sites that were analyzed in quantitative DNaseI-footprinting titrations are shown with the variable Watson-Crick base pairs bolded and the binding site boxed. For pDEH10, the three designed 6-base-pair binding sites that were analyzed in quantitative DNaseI-footprinting titrations are shown with the variable Watson-Crick base pairs bolded and the binding site boxed.

For polyamides Im-Im-**Hz**-Py-γ-Im-**Py**-Py-Py-β-Dp (**1**) and Im-Im-**Py**-Py-γ-Im-**Hz**-Py-Py-β-Dp (**2**), the DNA-sequence specificity at a single ring-pairing position (bolded in the sequences listed above) was determined by varying a single DNA base pair within the parent-sequence context, 5'-TGG**X**CA-3', to all four Watson-Crick base pairs (**X** = A, T, G, C) and comparing the relative affinities of the resulting complexes (**Figure 5.7**). The variable base-pair position was installed opposite the novel **Hz/Py** and **Py/Hz** pairs, designed to target T·A and A·T respectively. Equilibrium association constants (K_a) for eight-ring polyamides **19-21** containing **Py/Py**, **Hp/Py** and **Py/Hp** pairs have been reported and are included for comparison with values presented here (**Table 1**).[9b] Polyamide **1** (**Hz/Py** pair) bound with high affinity and demonstrated sequence specificity, preferring T·A over A·T by 10-fold, and A,T over G,C by more than 50-fold. In comparison to the **Hp/Py** pair, the **Hz/Py** exhibited a higher affinity, similar T vs. A specificity, and much greater A,T over G,C specificity. Polyamide **2** (**Py/Hz** pair) bound with high affinity and demonstrated modest specificity, preferring A·T over T·A by more than 4-fold, and A,T over G,C by more than 30-fold. In comparison to the **Py/Hp** pair,

the **Py/Hz** pair within this sequence context shows slightly lower T vs. A specificity, but improved A,T over G,C specificity. Both polyamides **1** and **2**, containing the **Hz/Py** and **Py/Hz** pairs respectively, bound with comparable affinity and single site specificity in comparison to the **Py/Py** pair.

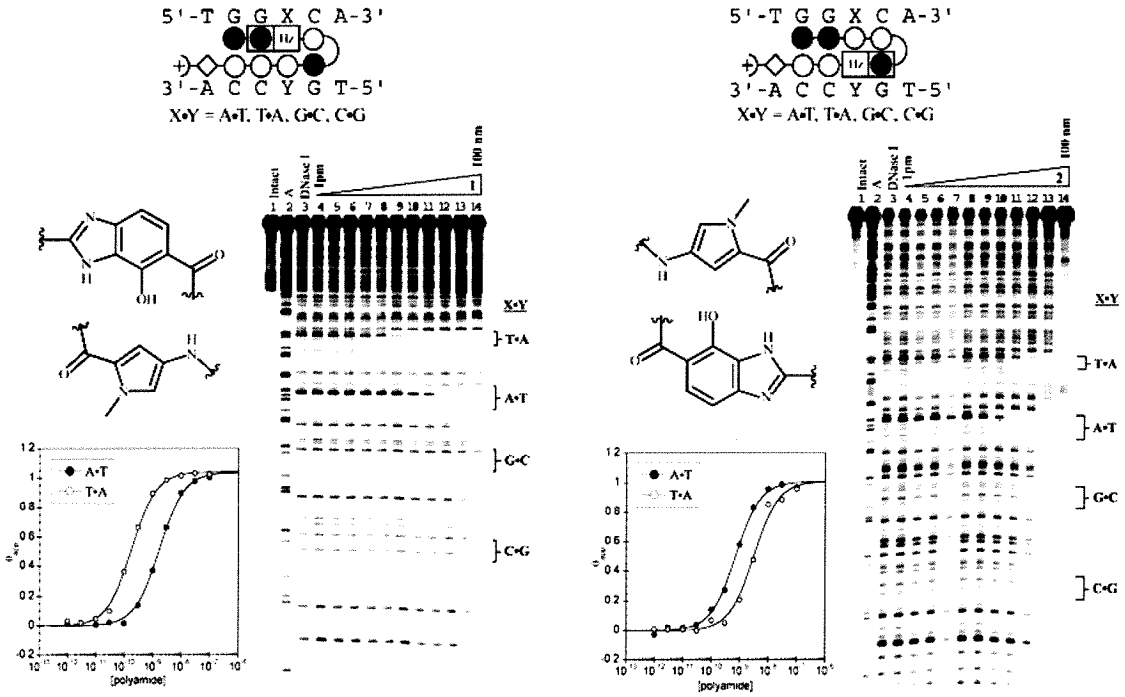


Figure 5.7 Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **1** and **2**, respectively, on the 278 bp, 5'-end-labelled PCR product of plasmid DHN1: lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (left, top) Chemical structure of the pairing of interest; and (bottom left) Binding isotherms for the four designed sites. θ_{norm} values were obtained according to published methods.[16] A binding model for the hairpin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; the hydroxybenzimidazole residue is indicated by a square containing Hz.

Table 1. *Hydroxypyrrole & Hydroxybenzimidazole Hairpins:* K_a $[M^{-1}]^{a,b}$

Polyamide	A•T	T•A	G•C	C•G
 19	**3.1 (± 0.7)x 10^9**	**4.7 (± 0.4)x 10^9**	2.2 (± 0.6)x 10^8	2.5 (± 0.9)x 10^8
 20	8.1 (± 1.9)x 10^7	**1.6 (± 0.3)x 10^9**	5.5 (± 1.5)x 10^7	7.9 (± 2.1)x 10^7
 1	5.7 (± 0.4)x 10^8	**5.5 (± 0.2)x 10^9**	≤ 1.0 x 10^7	≤ 1.0 x 10^7
 21	**1.1 (± 0.2)x 10^9**	9.8 (± 0.9)x 10^7	2.5 (± 0.3)x 10^7	3.3 (± 1.0)x 10^7
 2	**1.4 (± 0.3)x 10^9**	3.2 (± 0.6)x 10^8	≤ 1.0 x 10^7	≤ 1.0 x 10^7

a) Values reported are the mean values from at least three DNase I footprinting titration experiments, with the standard deviation given in parentheses. b) Assays were performed at 22 °C in a buffer of 10 mM Tris.HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0.

For polyamides Im-**Hz**-**Py**-Py-γ-Im-**Hz**-**Py**-Py-β-Dp (**3**) and Im-**Py**-**Hz**-Py-γ-Im-**Py**-**Hz**-Py-β-Dp (**4**) the ability of multiple **Hz/Py** pairs to distinguish multiple A,T sequences was tested by varying two DNA base pairs across from the **Hz/Py** and **Py/Hz** pairs (bolded in the sequences listed above) within the parent-sequence context, 5'-AG**WW**CT-3', (**W** = A, T) and comparing the relative affinities of the resulting complexes (**Figure 5.8**). Equilibrium association constants (K_a) for eight-ring polyamides **22-24** containing multiple **Py/Py** and **Hp/Py** pairs have been reported and are included for comparison with values presented here (**Table 2**).[14] Polyamide **3** bound at moderate affinity and showed good specificity (greater than 14-fold) for its match sequence 5'-AG**T**A**C**T-3' over 5'-AG**AA**CT-3' and 5'-AG**AT**CT-3'. Polyamide **4** bound all three sites at moderate affinity, demonstrating poor site selectivity. The recognition profiles of polyamides containing multiple **Hz/Py** pairs (**3** and **4**) are similar to those reported for multiple **Hp/Py** pairs.

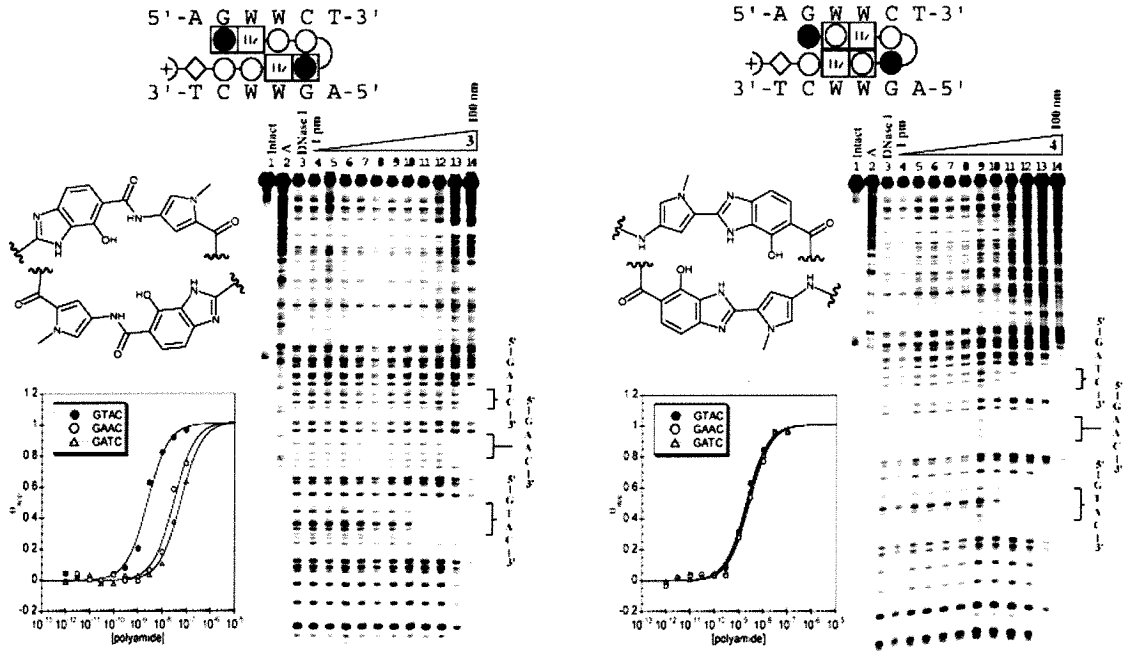


Figure 5.8 Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **3** and **4**, respectively, on the 285 bp, 5'-end-labelled PCR product of plasmid DEH10: lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (left, top) Chemical structure of the pairing of interest; and (bottom left) Binding isotherms for the four designed sites. θ_{norm} values were obtained according to published methods.[16] A binding model for the hairpin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; the hydroxybenzimidazole residue is indicated by a square containing Hz.

Table 2. *Multiple Hyroxypyrrole & Hydroxybenzimidazole Ring Pairings:* K_a $[M^{-1}]^{a,b}$

Polyamide	5'-aGTACt-3'	5'-aGAACt-3'	5'-aGATCt-3'
22	3.5 (± 0.7) x 10^{10}	4.7 (± 0.7) x 10^{9}	7.4 (± 1.5) x 10^{8}
23	**7.0 (± 1.8) x 10^{8}**	≤ 1.0 x 10^{7}	≤ 1.0 x 10^{7}
3	**4.6 (± 0.8) x 10^{8}**	3.2 (± 0.4) x 10^{7}	1.7 (± 0.5) x 10^{7}
24	1.0 (± 0.2) x 10^{8}	2.6 (± 0.6) x 10^{7}	3.3 (± 0.7) x 10^{7}
4	4.5 (± 0.7) x 10^{8}	3.3 (± 0.7) x 10^{8}	4.4 (± 0.9) x 10^{8}

a) Values reported are the mean values from at least three DNase I footprinting titration experiments, with the standard deviation given in parentheses. b) Assays were performed at 22 °C in a buffer of 10 mM Tris.HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0.

For polyamides Im-Im-**Hz**-Py-γ-Im-**Bi**-Py-Py-β-Dp (**5**) and Im-Im-**Bi**-Py-γ-Im-**Hz**-Py-Py-β-Dp (**6**) the DNA-sequence specificity at a single ring-pairing position (bolded in the sequences listed above) was determined by varying a single DNA base pair within the parent-sequence context, 5'-TGG**X**CA-3', to all four Watson-Crick base pairs (**X** = A, T, G, C) and comparing the relative affinities of the resulting complexes (**Figure 5.9 and Table 3**). The variable base-pair position was installed opposite the novel **Hz/Bi** and **Bi/Hz** pairs, designed to target T·A and A·T respectively. Polyamide **5** (**Hz/Bi** pair) bound with a markedly high affinity, demonstrating a 13-fold selectivity for A,T over G,C and 2.5-fold preference for T·A over A·T. Polyamide **6** (**Bi/Hz** pair) also bound with high affinity, lowered A,T over G,C selectivity (4.5-fold), and a 2.5-fold preference for A·T over T·A. Polyamides containing **Hz/Bi** pairs bind with significantly higher affinity than those containing the **Hz/Py** pairings, but with a loss in A,T selectivity.

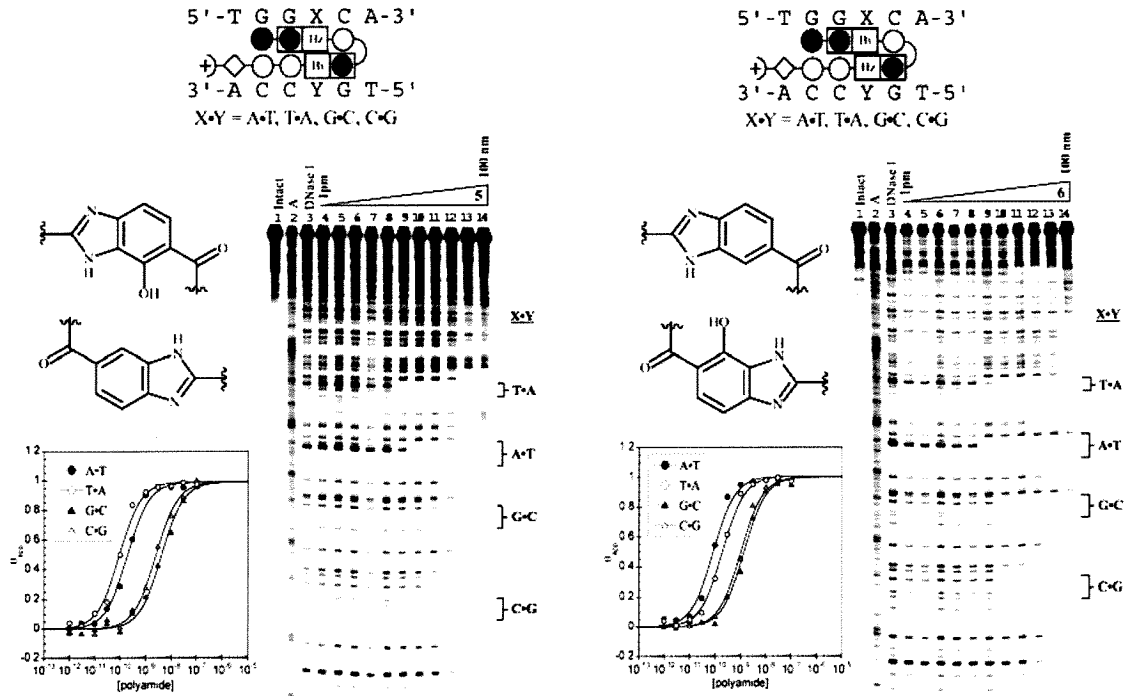


Figure 5.9 Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **5** and **6**, respectively, on the 278 bp, 5'-end-labelled PCR product of plasmid DHN1: lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (left, top) Chemical structure of the pairing of interest; and (bottom left) Binding isotherms for the four designed sites. θ_{norm} values were obtained according to published methods.[16] A binding model for the hairpin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; the hydroxybenzimidazole residue is indicated by a square containing Hz; the benzimidazole residue is indicated by a square containing Bi.

Table 3. *Hydroxybenzimidazole/Benzimidazole Pairings:* K_a $[M^{-1}]^{a,b}$

Polyamide	A•T	T•A	G•C	C•G
19	$3.1 (\pm 0.7) \times 10^{9}$	$4.7 (\pm 0.4) \times 10^{9}$	$2.2 (\pm 0.6) \times 10^{8}$	$2.5 (\pm 0.9) \times 10^{8}$
5	$4.1 (\pm 0.4) \times 10^{9}$	$\mathbf{1.0 (\pm 0.3) \times 10^{10}}$	$2.4 (\pm 0.7) \times 10^{8}$	$3.2 (\pm 0.5) \times 10^{8}$
6	$\mathbf{1.1 (\pm 0.4) \times 10^{10}}$	$4.5 (\pm 0.4) \times 10^{9}$	$8.1 (\pm 0.8) \times 10^{8}$	$1.0 (\pm 0.7) \times 10^{9}$

a) Values reported are the mean values from at least three DNase I footprinting titration experiments, with the standard deviation given in parentheses. b) Assays were performed at 22 °C in a buffer of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0.

Molecular Modeling. Modeling calculations were performed with the 'Spartan Essential' software package.[15] *Ab initio* calculations were done using a *Hartree-Fock* model and a 6-31G* polarization basis set. Four-ring subunits containing the sequence Im-Im-**X**-Py (**X = Hp, Hz**) were constructed to examine respective overall ligand geometry and curvature (**Figure 5.10**). Furthermore, dimeric subunits containing the sequence Im-**X** (**X = Hp, Hz, Py**, and **Bi**) were constructed and evaluated using electron density, and isopotential plots (**Figures 5.11 and 5.12**).

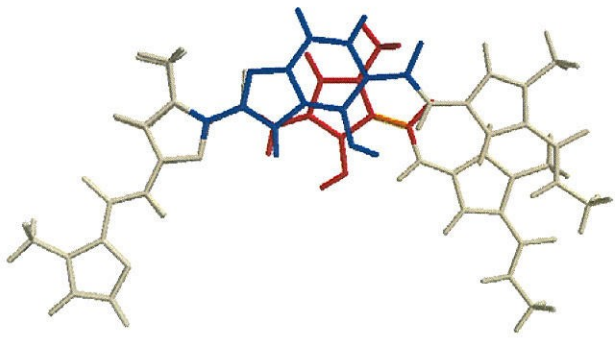


Figure 5.10 Schematic illustrating the respective curvatures of four-ring polyamide subunits Im-Im-Hp-Py and Im-Im-Hz-Py. Overall polyamide curvature effects how well the ligand can track the DNA minor groove. Hypercurvature negatively effects polyamide binding while curvature more complimentary to that of the DNA helix may allow for recognition of longer DNA sequences.

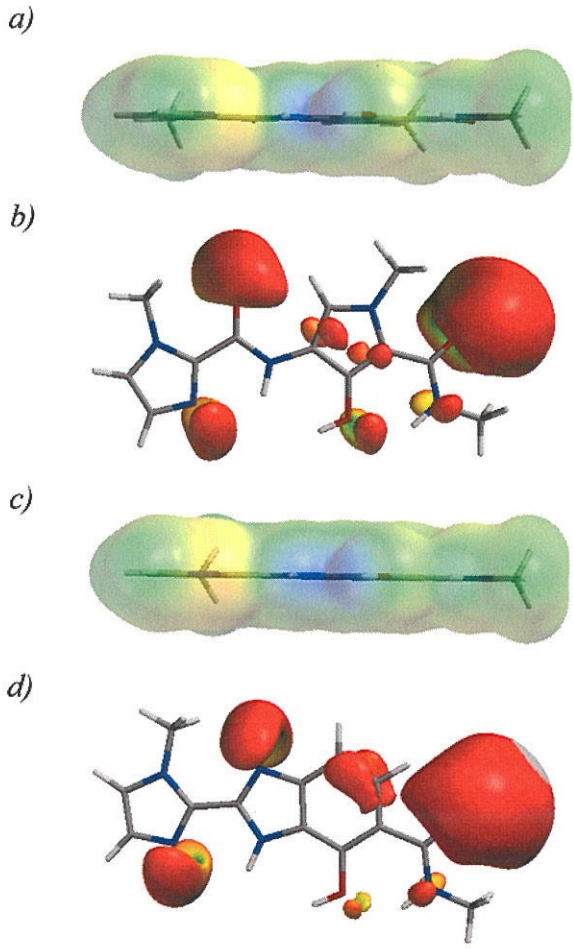


Figure 5.11 Calculated electronic surfaces of Im-Hp and Im-Hz dimeric units. *Ab initio* calculations done using 6-31G* basis set. (a) Electronic surface presented to the floor of the DNA-minor groove by Im-Hp dimer. (b) Isopotential surface of the Im-Hp dimer. (c) Electronic surface presented to the floor of the DNA-minor groove by Im-Hz dimer. (c) Isopotential surface of the Im-Hz dimer. Positive potentials are indicated by blue while negative potentials are indicated by red.

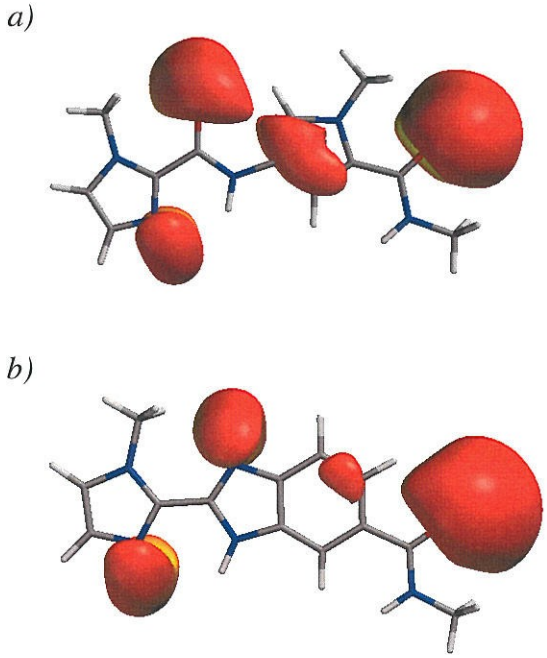


Figure 5.12 Calculated isopotential surfaces of Im-Py and Im-Bi dimeric units. *Ab initio* calculations done using 6-31G* basis set. (a) Isopotential surface of the Im-Py dimer. (b) Isopotential surface of the Im-Bi dimer. Positive potentials are indicated by blue while negative potentials are indicated by red.

5.3 Discussion.

Previously, **Hp/Py** pairs have been the state of the art for distinguishing between T·A and A·T Watson-Crick pairs, with **Hp/Py** coding for T·A and **Py/Hp** coding for A·T.[9] The specificity imparted by **Hp** is attributed to the interaction of the exocyclic hydroxyl group with the asymmetric A,T base-pair cleft. More specifically, **Hp** functions through two modes: first, the hydroxyl group provides a steric bump that is better accommodated across from the thymine base, and second, the hydroxyl proton may make a hydrogen bond with the O(2) of thymine. While **Hp** allows for single base specificity, utilization of the hydroxypyrrole ring system results in polyamides with lower affinities, only moderate success at discriminating between multiple A,T sequences, and chemical instability over extended periods of time in the presence of acid. This motivated us to

explore the molecular recognition capability of polyamides containing novel **Hz/Py** and **Hz/Bi** heterocyclic pairs.

Hydroxybenzimidazole/Pyrrole Pair (Hz/Py). The **Hz/Py** pair places the same direct readout functionality to the floor of the DNA-minor groove as the **Hp/Py** pair. Like the **Hp/Py** pair, the **Hz/Py** pair is capable of discriminating between A,T Watson-Crick base pairs such that **Hz/Py** codes for T·A and **Py/Hz** codes for A·T (**Table 4**). The **Hz/Py** pair demonstrates an increase in binding affinity for its match sites and is comparable in specificity to the **Hp/Py** pairing. The **Hz/Py** pair also discriminates more effectively against G,C base pairs than **Hp/Py**. While the substituents presented to the floor of the DNA minor groove appear identical between **Hz** and **Hp**, there are significant differences in ligand geometry (**Figure 5.10**) and electronics (**Figure 5.11**). A higher degree of rigidity and pre-organization of the fused hydroxybenzimidazole structure, coupled with a lower degree in curvature that may be more complimentary to the curvature of the DNA helix, likely play roles in the increased affinity and specificity.[8c,10b] Furthermore, examination of the isopotential surfaces of **Hz** and **Hp** indicates that side by side aromatic stacking interactions may be altered between the two different heterocyclic systems. Thus, by going from the 5-membered heterocyclic system of **Hp** to the fused 6-5 system of hydroxybenzimidazole **Hz**, changes associated with the indirect readout of the DNA-minor groove may be responsible for changes in affinity and specificity.

Table 4. *Specificity of the **Hz/Py** and **Hz/Bi** Pairings*[a]

Pair	A•T	T•A	G•C	C•G
Hz/Py	-	++	-	-
Py/Hz	++	-	-	-
Hz/Bi	-	+	-	-
Bi/Hz	+	-	-	-

a) Highly Favored (++), Favored (+), Disfavored (-)

Polyamides incorporated with multiple **Hz/Py** pairs are only modestly effective at differentiating between DNA sequences with multiple A,T sites, the specificity appearing to have a high degree of DNA sequence dependence. For example, polyamide **3** binds its match site 5'-AG**TA**CT-3' with high fidelity. However, polyamide **4**, which is designed to target the site 5'-AG**AT**CT-3' does so with no specificity. Similarly, only modest specificity for the 5'-AG**AT**CT-3' sequence is noted with multiple **Hp/Py** pairs – an effect that may be attributed to the different DNA microstructure of the 5'-GAT-3'step in comparison to the 5'-GTA-3' step. This inherent sequence bias is evident upon examination of the binding affinities of polyamides containing multiple symmetrical **Py/Py** pairs. The symmetrical **Py/Py** pair binds the 5'-AG**TA**CT-3' sequence with a substantially higher affinity than the 5'-AG**AT**CT-3' sequence, demonstrating the significant effect of DNA sequence on polyamide binding affinity.

Hydroxybenzimidazole/Benzimidazole Pair (Hz/Bi). The **Hz/Bi** pair presents the same functionality to the DNA-minor groove as the **Hz/Py** pair. While the functionality directed at the minor groove is the same, the **Hz/Bi** pair is the first example of a 6-5/6-5 bicyclic ring-ring pair. Accordingly, the overall geometry and electronics of the polyamide containing the **Hz/Bi** pair are substantially different (**Figure 5.12**). Like the **Hz/Py** pair, the **Hz/Bi** pair is capable of discriminating between A,T Watson-Crick base pairs such that **Hz/Bi** codes for T·A and **Bi/Hz** codes for A·T. The affinity of the **Hz/Bi** pair is significantly higher than the **Hz/Py** pair, albeit demonstrates lower selectivity. Polyamides containing two fused benzimidazole analogues have substantially lower conformational freedom and a higher degree of pre-organization, favorably orienting the polyamides in the proper orientation to bind the DNA. Further, the

benzimidiazole moiety has a greater aromatic surface and hydrophobicity that may alter both the DNA-ligand van der Waals interactions, and the intra-polyamide π-stacking. The lower degree of curvature of the **Hz/Bi** pair may also allow for a more intimate polyamide/DNA interaction, contributing to the noted increase in binding affinity. These changes in structure, while providing a marked increase in affinity, do come at the cost of lowered specificity. The lowered selectivity of the **Hz/Bi** pairing may be attributed to the greater rigidity of the system. More specifically, the lack of conformational freedom of the fused heterocyclic pairs may lower the sensitivity of the reading face of the polyamide, making it difficult to adjust to the small structural changes in the minor groove microenvironment that are the source of specificity.

5.4 Conclusion.

Previously, we have relied on hydroxypyrrole **Hp** as a T specific recognition element, differentiating between T and A when paired with pyrrole **Py,** such that **Hp/Py** codes for T·A and **Py/Hp** codes for A·T. In an effort to find a replacement for **Hp**, we have examined the DNA recognition properties of polyamides containing **Hz/Py** and **Hz/Bi** pairs. It was determined that the **Hz/Py** pair is an effective and stable replacement for the previously reported **Hp/Py** pair, successfully differentiating A,T base pairs at high affinity and specificity. It was also determined that multiple **Hz/Py** pairs are moderately effective at discriminating some multiple A,T sequences. Importantly, we report the use of the **Hz/Bi** pair, the first example of a 6-membered/6-membered ring-ring pair that is also capable of discriminating T·A from A·T base pairs. *What is remarkable is that we are now on a pathway to constructing oligomers for minor groove recognition that are*

not based on N-methylpyrrole-carboxamides. We visualize a time when sequence specific programmable oligomers may no longer be "polyamides" but are still based on the concept of pairing rules for DNA recognition. Future work will focus on the cellular uptake properties of these oligomers and their use in biological systems.

5.5 Experimental

General. N,N-dimethylformanide (DMF), N,N-diisopropylethylamine (DIEA), thiophenol (PhSH), N,N-dimethylaminopropylamine (Dp), Triethylamine (TEA), and thiourea were purchased from Aldrich. Boc-β-alanine-(4-carbonylaminomethyl)-benzyl-ester-copoly(styrene-divinylbenzene)resin (Boc-β-Pam-resin), dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBt), 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), N,N-dimethylaminopyridine (DMAP), and Boc-β-alanine were purchased from NOVA Biochem. Trifluoroacetic acid (TFA) was purchased from Halocarbon. All other solvents were reagent grade from EM. Oligonucleotide inserts were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology. Precoated silica gel plates $60F_{254}$ for TLC and silica gel 60 (40 μm) for flash chromatography were from Merck. Glycogen (20 mg/mL), dNTPs (PCR nucleotide mix), and all enzymes, unless otherwise stated, were purchased from Boehringer-Mannheim. pUC19 was purchased from New England Biolabs, and deoxyadenosine [γ-^{32}P]triphosphate was provided by ICN. Calf thymus DNA (sonicated, deproteinized) and DNaseI (7500 units/mL, FPLC pure) were from Amersham Pharmacia. AmpliTaq DNA polymerase was from Perkin-Elmer and used with the provided buffers. Tris.HCl, DTT, RNase-free water, and 0.5 M EDTA were from United States Biochemical. Calcium chloride, potassium chloride, and magnesium chloride were purchased from Fluka. Tris-borate-EDTA was from GIBCO and bromophenol blue was from Acros. All reagents were used without further purification.

NMR spectra were recorded on a Varian spectrometer at 300 MHz in DMSO-*d6* or $CDCl_3$ with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. High resolution FAB and EI mass spectra were recorded at the Mass Spectroscopy Laboratory at the University of California, Los Angeles. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) was conducted at the Protein and Peptide Microanalytical Facility at the California Institute of Technology.

Monomer Synthesis. Compounds **7-12, 15** and **19** were synthesized according to previously published protocols.[10b,10c,11]

Methy 7-methoxy-2-(1-methylimidazol-2-yl)benzimidazole-6-carboxylate **(Im-HzOMe-OMe 16).** A mixture of **15** (0.2 g, 1.02 mMol), Im-$COCCl_3$, (345 mg, 1.53 mMol), DIEA (132 mg, 178 μL, 1.02 mMol), and DMAP (25 mg, 204 μMol) in EtOAc (5 mL) was stirred for 12 h at 60 °C, over which time a precipitate formed. The reaction was cooled to room temperature, filtered, and washed with cold Et_2O. The off white solid was collected and dissolved in AcOH (5 mL). The mixture was then heated to 90 °C for 6 h. The solvent was then removed by rotoevaporation and the remaining white solid dried under high vacuum to provide **16** (210 mg, 72% Yield). TLC (4:1 EtOAc/Hex) R_f 0.4; ^{1}H NMR (DMSO-d_6) 7.55 (s, 1H), 7.53 (s, 1H) 7.43 (s, 1H), 7.13 (s, 1H), 4.36 (s, 3H), 4.16 (s, 3H), 3.78 (s, 3H); ^{13}C (DMSO-d_6) 166.4, 151.2, 143.8, 138.9, 136.9, 134.6,

128.3, 125.4, 125.3, 113.9, 105.1, 60.9, 51.7, 35.1; EI-MS m/e 286.107 (M^+ calcd for 286.107 $C_{14}H_{14}N_4O_3$).

7-methoxy-2-(1-methylimidazol-2-yl)benzimidazole-6-carboxylic acid **(Im-HzOMe-OH 13).** A mixture of **16** (200 mg, 699 μMol), MeOH (3 mL), and 1N NaOH (4 mL) was stirred at 35 °C for 3 h. The methanol was removed in vacuo and the mixture was taken to pH 2 using 1N HCl, upon which time a white precipitate formed. The mixture was poured into a 50 mL Falcon tube and spun down in a centrifuge (10 min x 14,000 rpm). The tube was decanted, leaving a white solid that was dried under high vacuum to provide **16** (165 mg, 87% Yield). TLC (3:2 EtOAc/Hex, 10% AcOH) R_f 0.4; ^{1}H NMR (DMSO-d_6) 7.58 (s, 1H), 7.55 (s, 1H) 7.46 (s, 1H), 7.16 (s, 1H), 4.33 (s, 3H), 4.16 (s, 3H); ^{13}C (DMSO-d_6) 167.4, 151.0, 143.6, 138.6, 136.8, 135.2, 127.8, 125.7, 125.3, 115.1, 105.2, 61.0, 35.2; EI-MS m/e 272.260 (M^+ calcd for 272.260 $C_{13}H_{12}N_4O_3$).

Methyl 7-methoxy-2-(1-methyl-4-nitroimidazol-2-yl)benzimidazole-6-carboxylate **(NO_2-Im-HzOMe-OMe 17).** Diamine **15** (0.5 g, 2.54 mmol), NO_2-Im-OH **19** (480 mg, 2.80 mmol), HBTU (1 g, 2.66 mmol), DIEA (362 mg, 488 μL, 2.80 mmol), and DMF (7 mL) were stirred for 2 days at room temperature. The mixture was then added to a 50 mL Falcon tube containing water (20 mL), resulting in a precipitate. The Falcon tube was centrifuged (10 min x 14,000 rpm) and the mother liquor decanted, leaving a tan solid that was dried under high vac. The solid was then dissolved in AcOH (8 mL) and heated to 90 °C with stirring. It is noteworthy that the solid was not completely soluble in AcOH. The reaction was stirred for 6 hours and the precipitate that was present was filtered over a fine fritted funnel. The solid was washed with Et_2O and dried under high vacuum to provide **17** (481 mg, 57% Yield) as a powdery yellow solid. TLC (3:2

EtOAc/Hex) R_f 0.5; ^{1}H NMR (DMSO-d_6) 8.68 (s, 1H), 7.60 (d, J = 8.4 Hz, 1H) 7.18 (d, J = 8.4 Hz, 1H), 4.38 (s, 3H), 4.24 (s, 3H), 3.78 (s, 3H); ^{13}C (DMSO-d_6) 166.2, 147.0, 142.1, 138.2, 137.7, 133.4, 132.7, 128.1, 117.5, 109.5, 107.3, 61.8, 52.1, 36.5; EI-MS m/e 331.092 (M^+ calcd for 331.092 $C_{14}H_{13}N_5O_5$).

Methyl 2-{4-[(tert-butoxy)carbonylamino]-1-methylimidazol-2-yl}-7-methoxybenzimidazole-6-carboxylate **(Boc-Im-HzOMe-OMe 18)** A mixture of **17** (400 mg, 1.21 mmol), DIEA (400 mg, 536 μL, 3.08 mmol), Pd/C (50 mg) and DMF (5 mL) was placed in a parr apparatus and hydrogenated (500 psi) for 1.5 h at ambient temperature. The mixture was removed from the parr apparatus and $(Boc)_2O$ (396 mg, 1.82 mmol) was added. The mixture was then stirred for 8 h at 50 °C. The solvent was removed in vacuo, followed by column chromatography of the brown residue (3:2 Hex/EtOAc) to provide **18** as a thin film. The thin film was treated with hexanes and then the solvent was removed by rotoevaporation, followed by drying under high vacuum to provide **18** as a white solid (228 mg, 47% Yield). TLC (3:2 EtOAc/Hex) R_f 0.6; ^{1}H NMR (DMSO-d_6) 9.55 (s, 1H), 7.75 (d, J = 8.7 Hz, 1H) 7.60 (d, J = 8.7 Hz, 1H), 7.21 (s, 1H) 4.27 (s, 3H), 3.81 (s, 3H), 3.70 (s, 3H), 1.44 (s, 9H); ^{13}C (DMSO-d_6) 165.8, 151.0, 147.0, 142.1, 137.9, 137.0, 133.3, 132.4, 127.8, 117.3, 109.4, 107.2, 85.9, 61.7, 52.0, 33.5, 28.2; EI-MS m/e 401.170 (M^+ calcd for 401.170 $C_{14}H_{13}N_5O_5$).

2-{4-[(tert-butoxy)carbonylamino]-1-methylimidazol-2-yl}-7-methoxybenzimidazole-6-carboxylic acid **(Boc-Im-HzOMe-OH 14).** A mixture of **18** (200 mg, 498 μMol), 1N NaOH (2 mL) and MeOH (2 mL) was stirred at 30 °C for 4 h. The MeOH was removed by rotoevaporation and the pH carefully adjusted to pH = 2 with 1N HCl. The precipitate was extracted with EtOAc (3 x 10 mL), the organics dried over sodium sulfate and

removed by rotoevaporation to provide **14** (166 mg, 86% Yield) as a fine white solid. TLC (3:2 EtOAc/Hex, 10% AcOH) R_f 0.65; ^{1}H NMR (DMSO-d_6) 9.55 (s, 1H), 7.77 (d, J = 8.7 Hz, 1H) 7.63 (d, J = 8.7 Hz, 1H), 7.23 (s, 1H) 4.24 (s, 3H), 3.77 (s, 3H), 1.44 (s, 9H); ^{13}C (DMSO-d_6) 166.7, 150.8, 146.7, 142.0, 135.9, 137.0, 133.3, 132.4, 127.4, 117.3, 109.4, 107.3, 86.0, 61.8, 33.6, 28.2; EI-MS m/e 387.154 (M^+ calcd for 387.154 $C_{18}H_{21}N_5O_5$).

Polyamide Synthesis: Polyamides were synthesized from Boc-β-alanine-Pam resin (50 mg, 0.59 mmol/g) and purified by preparatory HPLC according to published manual solid phase protocols.[10]

Im-Im-**Hz**-Py-γ-Im-Py-Py-Py-β-Dp (**1**): (Boc-Im-**HzOMe**-OH) (34 mg, 88.5 μmol) was incorporated by activation with HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μl, 177 μmol) and DMF (250 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 12 h at room temperature. After Boc-deprotection, the terminal imidazole residue was incorporated using Im-$COCCl_3$. Im-$COCCl_3$ (67 mg, 295 μmol), DIEA (23 mg, 31 μl, 177 μmol) and DMF (400 μl) were added to the reaction vessel containing H_2N-Im-**HzOMe**-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 2 h at 37 °C, and determined complete by analytical HPLC. The resin-bound polyamide was then washed with DCM and subjected to the cleavage, O-methyl deprotection and purification protocol described below to provide Im-Im-**Hz**-Py-γ-Im-Py-Py-Py-β-Dp (**1**) (1.1 mg, 3.1% recovery) as a fine white powder

under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1233.56 (M+H calcd for 1233.56 $C_{58}H_{69}N_{22}O_{10}$).

Im-Im-Py-Py-γ-Im-**Hz**-Py-Py-β-Dp (**2**): (Boc-Im-**HzOMe**-OH) was incorporated as described for **1** . The polyamide was cleaved from resin and treated as described in the deprotection protocol below to provide (**2**) (0.9 mg, 2.5 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1233.55 (M+H calcd for 1233.56 $C_{58}H_{69}N_{22}O_{10}$).

Im-**Hz**-Py-Py-γ-Im-**Hz**-Py-Py-β-Dp (**3**): (Im-**HzOMe**-OH) (25 mg, 88.5 μmol) was incorporated by activation with HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μl, 177 μmol) and DMF (250 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Py-Py-γ-Im-**Hz**-Py-Py-β-Pam resin. Coupling was allowed to proceed for 12 h at room temperature. The resin-bound polyamide was then washed with DCM and treated as described in the deprotection protocol below to provide Im-**Hz**-Py-Py-γ-Im-**Hz**-Py-Py-β-Dp (**3**) (0.7 mg, 1.9 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1242.56 (M+H calcd for 1242.55 $C_{60}H_{68}N_{21}O_{10}$).

Im-Py-**Hz**-Py-γ-Im-Py-**Hz**-Py-β-Dp (**4**): (Boc-Py-**HzOMe**-OH) (34 mg, 88.5 μmol) was incorporated by activation with HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μl, 177 μmol) and DMF (250 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Py-β-Pam resin. Coupling was allowed to proceed for 12 h at room temperature. After Boc-deprotection, the additional **Im**, **γ**, and **Py** units were incorporated as previously described. [Ref] The

second Boc-Py-**HzOMe**-OH unit was activated as described above and added to the reaction vessel containing H_2N-Py-γ-Im-Py-**Hz**-Py-β-Pam resin. Coupling was allowed to proceed for 12 h at room temperature. After Boc-deprotection, the terminal imidazole residue was added as described for **1**. The resin-bound polyamide was then washed with DCM and treated as described in the deprotection protocol below to provide Im-Py-**Hz**-Py-γ-Im-Py-**Hz**-Py-β-Dp (**4**) (1.1 mg, 3.0 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1242.55 (M+H calcd for 1242.55 $C_{60}H_{68}N_{21}O_{10}$).

Im-Im-**Hz**-Py-γ-Im-**Bi**-Py-Py-β-Dp (**5**): (Boc-Im-**Bi**-OH) (32 mg, 88.5 μmol) was incorporated by activation with HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μl, 177 μmol) and DMF (250 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing H_2N-Py-Py-β-Pam resin. Coupling was allowed to proceed for 12 h at room temperature. After Boc-deprotection, the additional **Im**, **γ**, and **Py** units were incorporated as previously described. [Ref] The Boc-Im-**HzOMe**-OH residue and the terminal imidazole residue were incorporated as described for **1**. The resin-bound polyamide was then washed with DCM and treated as described in the deprotection protocol below to provide Im-Im-**Hz**-Py-γ-Im-**Bi**-Py-Py-β-Dp (**5**) (1.1 mg, 3.0 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1226.55 (M+H calcd for 1226.54 $C_{59}H_{66}N_{22}O_9$).

Im-Im-**Bi**-Py-γ-Im-**Hz**-Py-Py-β-Dp (**6**): (Boc-Im-**HzOMe**-OH) and (Boc-Im-**Bi**-OH) were incorporated as described in **1** and **5**. The terminal imidzole residue was

incorporated as described in **1**. Upon completion of the synthesis, the resin-bound polyamide was then washed with DCM and treated as described in the deprotection protocol below to provide Im-Im-**Bi**-Py-γ-Im-**Hz**-Py-Py-β-Dp (**6**) (1.3 mg, 3.5 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1226.54 (M+H calcd for 1226.54 $C_{59}H_{66}N_{22}O_9$).

Deprotection of the *O*-Methyl-Protected Polyamides. All of the above polyamides were cleaved from resin, purified, deprotected and subject to further purification using the following general procedure. Upon completion of solid phase synthesis, Dp (500 μL) was added to the synthesis vessel containing the resin (50 mg). The mixture was allowed to stand for 2 h at 85 °C with occasional agitation. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was purified by reversed phase HPLC and lyophilized to provide polyamides containing the *O*-methyl protected hydroxybenzimidazole unit (-**HzOMe-**) as a dry solid. The polyamides were then dissolved in DMF (200 μl) and added to a suspension of sodium hydride (40 mg, 60% oil dispersion) and thiophenol (200 μl) in DMF (400 μl) that was pre-heated for 5 min at 85 °C. The mixture was heated for 2 h at 85 °C. The mixture was then cooled to 0 °C and 20% TFA (7.0 mL) was added. The aqueous layer was washed three times with diethyl ether (8 mL) and then diluted to a total volume of 9.5 mL using 0.1% TFA. The mixture was then purified by reverse-phase HPLC to give the deprotected **Hz**-containing polyamides.

Footprinting Experiments. – Plasmids pDHN1 and pDEH10 were constructed and 5'-radiolabeled as previously described.[8b,9d] DNase I footprint titrations were performed according to standard protocols.[16]

5.6 References

[1] (a) Pandolfi, P. P. *Oncogene* 2001, *20*, 3116-3127. (b) Darnell, J. E. *Nature Rev. Cancer* **2002**, *2*, 740-748.

[2] (a) Finlay, A. C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F.X. *J. Am. Chem. Soc.* **1951**, *73*, 341. (b) Arcamone F., N. V.; Penco S., Orezzi P.; Nicolella V.; Pirelli A., *Nature* **1964**, 203, 1064- 1065

[3] (a) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376-1380. (b) Pelton, J. G.; Wemmer, D.E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723-5727.

[4] Dervan, P. B., *Bioorg. & Med. Chem.* **2001**, *9*, 2215-2235. (b) Dervan, P. B.; Edelson, B. S. *Curr. Op. Struc. Bio.* **2003**, *13*, 284-299.

[5] (a) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202-205. (b) Dickenson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, 95, 12890-12895. (c) Janssen, S.; Durussel, T.; Laemmli, U. K. *Mol. Cell.* **2000**, *6*, 999-1011. (d) Ansari, A. Z.; Mapp, A. K.; Nguyen, D. H.; Dervan, P. B.; Ptashne, M. *Chem. Biol.* **2001**, *8*, 583-592. (e) Coull, J. J.; He, G.; Melander, C.; Rucker, V. C.; Dervan, P. B.; Margolis, D. M. *J. Virology* **2002**, *76*, 12349-12354. (f) Oyoshi, T.; Kawakami, W.; Narita, A. Bando, T.; Sugiyama, H. *J. Am. Chem. Soc.* **2003**, *125*, 4752-4754. (g) Dudouet, B; Burnett, R.; Dickinson, L. A.; Wood, M. R.; Melander, C.; Belitsky, J. M.; Edelson, B.; Wurtz, N; Briehn, C.; Dervan, P. B.; Gottesfeld, J. M. *Chem. Biol.* **2003**, *10*, 859-867.

[6] (a) Geierstanger, B. H.; Mrksich M.; Dervan, P. B.; Wemmer D. E. *Science* **1994**, *226,* 646-650. (b) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104-109. (c) Kielkopf, C. L.; White, S; Szewcyzk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111-115. (d) Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *J. Mol. Biol.* **2000**, *295*, 557-567.

[7] (a) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. *J. Am. Chem. Soc.* **2001** *123*, 5044-5049. (c) Wellenzohn, B.; Loferer, M. J.; Trieb, M.; Rauch, C.; Winger, R. H.; Mayer, E.; Liedl, K. R. *J. Am. Chem. Soc.* **2003**, *125*, 1088-1095.

[8] (a) Ellervik, U.; Wang, C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 9354-9360. (b) Nguyen, D. H.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 7-17. (c) Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B. *Helv. Chim. Acta* **2002**, *85*, 4485-4517. (d) Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B. *Bioorg. Med. Chem.* **2003**. *11*, 4333-4340

[9] (a) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P.B. *Nature* **1998**, *391*, 468-471. (b) Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J.

M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 11621-11629. (c) White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 260-261. (d) Melander, C.; Herman, D. M.; Dervan, P. B. *Chem. Eur. J.* **2000**, *24*, 4487-4497.

[10] (a) Minehan, T. G.; Gottwald, K.; Dervan, P. B. *Helv. Chim. Acta* **2000**, *83*, 2197-2213. (b) Briehn, C. A.; Weyermann, P.; Dervan, P. B. *Chem. Eur. J.* **2003**, *9*, 2110-2122. (c) Renneberg, D.; Dervan, P. B. *J. Am. Chem. Soc.* **2003**, *125*, 5707-5716. (d) Reddy, P. M.; Jindra, P. T.; Satz, A. L.; Bruice, T. C. *J. Am. Chem. Soc.* **2003**, *125*, 7843-7848.

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

[12] Nishiwaki, E; Tanaka, S.; Lee, H.; Sibuya, M. *Heterocycles* **1988**, *27*, 1945-1952.

[13] (a) Harbuck, J. W.; Rapoport, J. *J. Org. Chem.* **1972**, *37*, 3618-3622. (b) Bailey, D. M.; Johnson, R. E. *J. Med. Chem.* **1973**, *16*, 1300-1302.

[14] Herman, D. E. *Thesis* 2001, California Institute of Technology, Pasadena, CA 91125.

[15] 'Spartan Essential', Wavefunction Inc., 1991-2001.

[16] Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450-466.

Chapter 6

Programmable Oligomers for DNA Recognition

The text of this chapter was taken in part from a manuscript coauthored with Michael A. Marques, Shane Foister and Professor Peter B. Dervan (Caltech).

(Doss, R. M.; Marques, M. A.; Foister, S. F.; and Dervan, P. B. "*Programmable Oligomers for DNA Recognition*" *JACS, in preparation.*)

Abstract.

Hairpin polyamides have been shown to selectively recognize predetermined DNA sequences with high affinities and sequence specificities. Polyamides are able to discriminate between each of the four Watson-Crick Base pairs through the side-by-side pairing of aromatic rings in the minor groove of DNA. While the established polyamide/DNA pairing rules have been effective at targeting hundreds of sequences *in vivo* and *in vitro*, we herein present a step towards designing polyamide-based oligomers which contain an emerging set of recognition elements capable of targeting a new selection of DNA sequences. These elements include the N-terminal **No-Hz** and **Ct-Hz** dimer caps which are able to target G-T and T-T sequences, respectively, with good affinities and specificities. The combination of these new N-terminal dimers with previously described 6-5 fused systems has yielded oligomers which are able to bind DNA with comparable specificities and improved affinities.

6.1 Introduction.

Hairpin polyamides represent a growing class of sequence-specific, DNA binding molecules.[1-3] First generation polyamides comprised of the heterocyclic amino acid subunits 1-methyl-1*H*-imidazole (**Im**), 1-methyl-1*H*-pyrrole (**Py**) and 3-hydroxy-1*H*-pyrrole (**Hp**) are able to bind in the minor groove of DNA with sub-nanomolar affinities and significant sequence selectivity. While co-facial pairings of **Im**, **Py**, and **Hp** where Im/Py is specific for G•C, Hp/Py is specific for T•A and Py/Py is specific for A•T and T•A, are able to code for the four Watson-Crick base pairs,[4-6] the high level of structural variation between DNA sequences makes it especially difficult for such a small repertoire of recognition elements to effectively target a diverse range of sequences. Even with substantial, sequence-specific differences in the DNA microstructure, the "pairing rules" have proven effective in recognizing hundreds of predetermined and biological DNA targets.[7, 8]

Much inspiration for the structural design of standard **Im**, **Py**, and **Hp** polyamides was drawn from the natural products netropsin and dystamycin A.[9] Comprised of amide linked pyrrole rings, these small molecules are capable of selectively binding A,T tracks of DNA in the minor groove. Subsequent, deliberate atomic changes to the edges of the antiparallel aromatic amino acid ring pairs which interact with the target DNA led to the advent of the G specific **Im** residue and the T specific **Hp** residue.[10, 11] Placing an aromatic nitrogen in the minor groove allowed **Im** to accept a hydrogen bond from the exocyclic amine of guanine, thus imparting excellent G > C specificity. Incorporating an exocyclic hydroxyl group in **Hp** allows the ring to specify for T > A; the hydroxyl groups

prefer to lie across the less bulky thymine base in an A•T base pair while also forming hydrogen bonds with the O2 carbonyl of thymine.

Our initial efforts to further expand the polyamide alphabet utilized the same approach of single atom substitutions on the existing five-membered heterocycle scaffold.[12] We believed that evaluating many heterocycles, each with relatively subtle structural differences, would lead us to a ring pairing that could serve as a novel mode of DNA recognition. To our surprise, an extensive study of many new ring systems led only to a single result wherein we found that a thiophene ring (**Tn**) was able to present a sulfur atom to the minor groove thus imparting a modest selectivity for T > A in a T•A base pair.

It was not until we explored the properties of the thiophene rings at the polyamide N-terminal (cap) position that the full significance of the **Tn** result became apparent: internal pairings and cap pairings were not equivalent.[13] The notion that ring pairings were context dependent led to a veritable paradigm shift in our approach to polyamide design (**Figure 6.1**). Ring pairings that exhibited poor binding properties and little or no specificity internally, such as the (chlorothiophene) **Ct/Py** pair, were much more successful at inferring an appreciable level of specificity at the cap position.

We herein report a new set of heterocyclic dimers which represent a step away from single base-pair recognition and towards a more global approach to molecular recognition. The new recognition elements were designed by combining the T-specific hydroxybenzimidazole 6-5 fused ring system (**Hz**) at the internal position with the newly

developed thiophene ring caps (**Figure 6.2**).[14, 15] When used at the cap position of

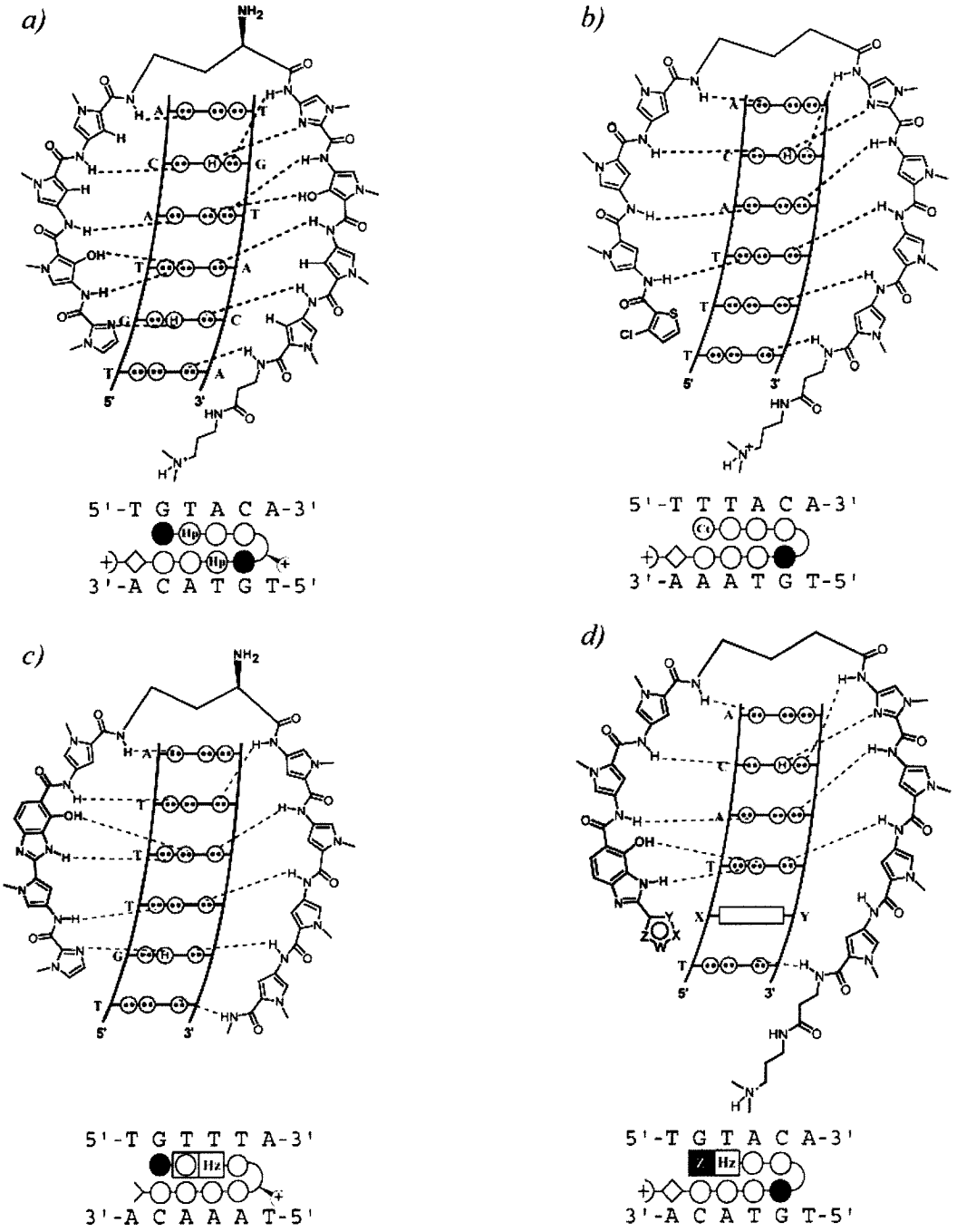


Figure 6.1 The success of new ring systems depends strongly upon their positional context in a polyamide or oligomer. Examples of novel rings systems and position where they have been proven to be effective are shown above.

hairpin polyamides, the **Ct-Hz** and **No-Hz** dimers are able to target T-T and G-T sequences respectively with high affinity and appreciable specificity. Evolution of the **No-Hz** dimer came about due to the inability of **Im-Hz** to specify for its designed G-T site. DNase I footprinting titrations were used to determine DNA binding affinities of the **Ct-Hz** and **No-Hz** dimers for the four Watson-Crick bases. Molecular modeling was used to further compliment the thermodynamic results.

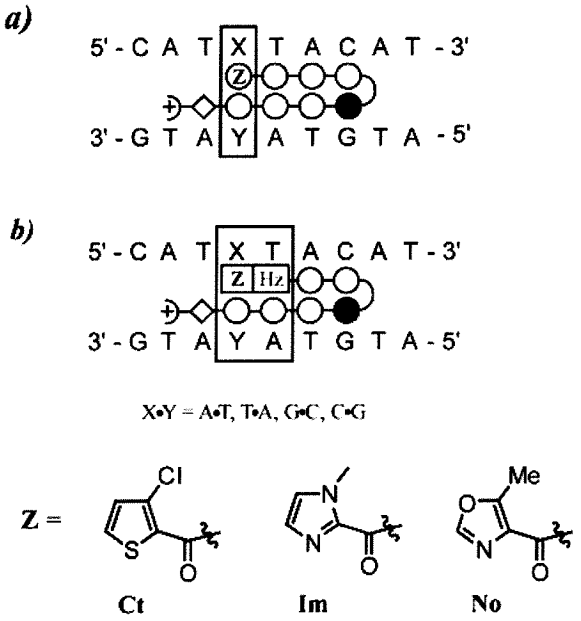


Figure 2. Experimental scheme showing the DNA sequences that each polyamide will be tested against. (a) Hairpin polyamides containing the rings (**Z** = **Ct**, **Im**, **No**) proximal to a pyrrole (**Py**) residue. (b) Hairpin polyamides containing the rings (**Z**) proximal to a hydroxybenzimidazole residue (**Hz**). Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits.

The development of a dimer system capable of recognizing a G-T sequence of DNA provided the opportunity to create our first oligomeric system designed to bind the site 5`-GTAC-3` (**Figure 6.3**). Such an oligomer represents our first compound which demonstrates excellent DNA binding properties without containing a single pyrrole or imidazole carboxamide.

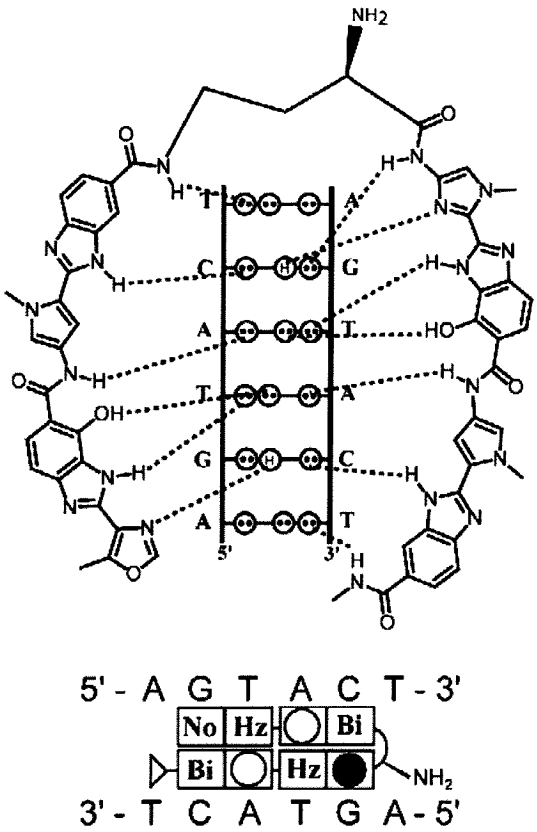


Figure 6.3 Chemical structure and cartoon representation of the first, complete oligomer generated with the new recognition elements. Oligomers such as this one contain no pyrrole, imidazole, or hydroxypyrrole carboxamides.

6.2 Monomer, Dimer and Polyamide Synthesis (Figure 6.4)

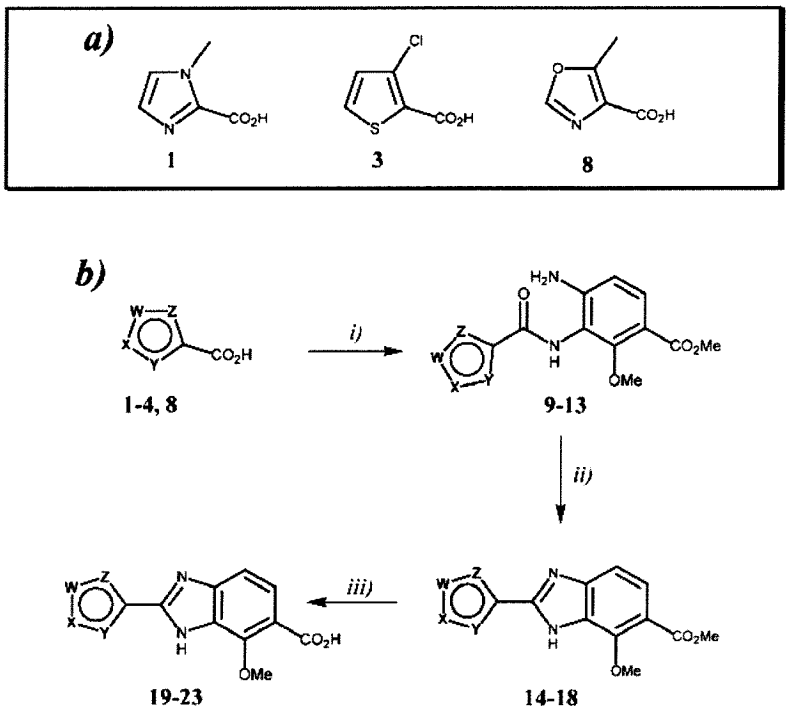


Figure 4. Synthesis of heterocyclic amino acid monomers and dimers. (a) Commercially available heterocycles **1-4** and synthesized oxazole **8**. General procedure for the synthesis of fused heterocyclic dimers: (i) Heterocyclic Acids, HBTU, DIEA, DMF, 35 ^{o}C, 24 h; (ii) AcOH, 90 ^{o}C, 12 h; (iii) NaOH, MeOH, 60 ^{o}C, 4-6 h. (**1**, **9**, **14**, **19**) W = C-H, X = C-H, Y = N, Z = N-Me; (**2**, **10**, **15**, **20**) W = C-H, X = C-H, Y = S, Z = C-Me; (**3**, **11**, **16**, **21**) W = C-H, X = C-H, Y = S, Z = C-H; (**4**, **12**, **17**, **22**) W = C-Me, X = O, Y = N, Z = C-H; (**8**, **13**, **18**, **23**) W = O, X = C-H, Y = N, Z = C-Me.

Is-OH (4). A solution of 2,4-dioxopentanoic acid methyl ester (**5**), hydroxylamine•HCl, $MgSO_4$ and H_2SO_4 in methanol was stirred at R.T. for 8 hours in order to afford the isoxazole ring (**Is-OMe** (**6**)) cap precursor. The methyl ester was subsequently exposed to 1N NaOH in methanol at 40°C for 4 hours at which the saponification of the ester was confirmed by TLC. After rotoevaporation of the methanol solvent, the pH of the aqueous solution was adjusted to pH = 2 with 1N HCl. The precipitate was extracted with EtOAc, dried over sodium sulfate and removed by rotoevaporation to provide **4** as a fibrous white solid in 88% yield.

No-OH (8). To a mixture of ethyl-alpha-isocyanoacetate and DBU in dry THF at 10°C was added a solution of acetic anhydride in dry THF dropwise. After warming to R.T the reaction was allowed to stir for 10 hours at which point the solvent was removed by rotoevaporation, water was added and the product extracted with EtOAc (2 x 100 mL). The organic layer was dried over sodium sulfate and removed by rotoevaporation to yield **No-OMe** (**7**) as a crude amber oil which after column chromatography (3:2 Hex/EtOAc) provided the ester as a crystalline white solid in 64% yield. **7** was then exposed to 1N NaOH in methanol at 40°C for 4 hours. After rotoevaporation of the methanol solvent, the pH of the aqueous solution was adjusted to pH = 2 with 1N HCl. The precipitate was extracted with EtOAc, dried over sodium sulfate and removed by rotoevaporation to provide **No-OH (8)** as a fibrous white solid in 82% yield.

General Synthetic Procedure for dimeric caps **19-23**. To a solution of heterocyclic carboxylic acids **1-4**, and **8** in DMF was added DIEA and HBTU. The solution was allowed to stir at R.T. for 1 hour allowing for conversion of the carboxylic acid to the activated, aryl HBTU-ester. The aryl diamine (**37**) was then added and the reaction heated to 35°C for 24 hours. The reaction was allowed to cool to room temperature and poured into a separatory funnel containing water. The water was then extracted with EtOAc (2 x 100 mL), the organic layer dried over sodium sulfate and removed by rotoevaporation to provide the amides as crude solids. The amides were then dissolved in acetic acid and heated to 90°C for 12 hours at which point the AcOH was removed by rotoevaporation. The resultant residues were subjected to column chromatography (EtOAc/Hex) to provide the dimeric methyl ester caps as a thin films. Addition of hexanes, followed by rotoevaporation and drying under high vacuum provided the dimers

as white solids. Subjection of the dimeric esters to 1N NaOH and MeOH at 40 °C for 4 h and workup afforded the dimer cap acids as white solids in good yields.

Polyamide Synthesis. Hairpin polyamides were synthesized manually from Boc-β-PAM resin in a stepwise fashion using Boc-protected monomeric and dimeric amino acids according to established solid-phase protocols. Base Resin 1 (**BR1**) (H_2N-Py-Py-γ-Im-Py-Py-Py-β-Pam) and Base Resin 2 (**BR2**) (H_2N-Py-Py-Py-γ-Im-Py-Py-Py-β-Pam) were synthesized in gram quantities using the following amino acid building blocks: Boc-**Py**-OBt, Boc-**Im**-OH and Boc-γ-OH (**Figure 6.5**). The base resins were then split into smaller batches for coupling to the final monomeric and dimeric caps. Boc-protected amino acid monomers and dimers for **Im**, **Py**, and **Ct** were synthesized according to previously reported procedures.[13, 16, 17] Couplings were realized using pre-activated monomers (Boc-**Py**-OBt) or HBTU activation in a DIEA and DMF mixture. Coupling times ran from 3-24 h at 25-40 °C. Deprotection of the growing polyamide was accomplished using 80% TFA/DCM. Polyamides were cleaved from the resin by treatment with dimethylaminopropylamine (Dp) neat at 80 °C for 2 h, and purified by preparatory reverse phase HPLC. Deprotection of the methoxy-protected polyamides was done using a mix of thiophenoxide in DMF at 80 °C, to provide the free hydroxy derivatives after a second HPLC purification: **Im**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**24**), **Tn**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**25**), **Ct**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**26**), **Is**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**27**), **No**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**28**), **Im**-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**29**), **Tn**-**Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**30**), **Ct**-**Hz**-Py-Py-γ-

Im-Py-Py-Py-β-Dp (**31**), **Is-Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**32**), **No-Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**33**).

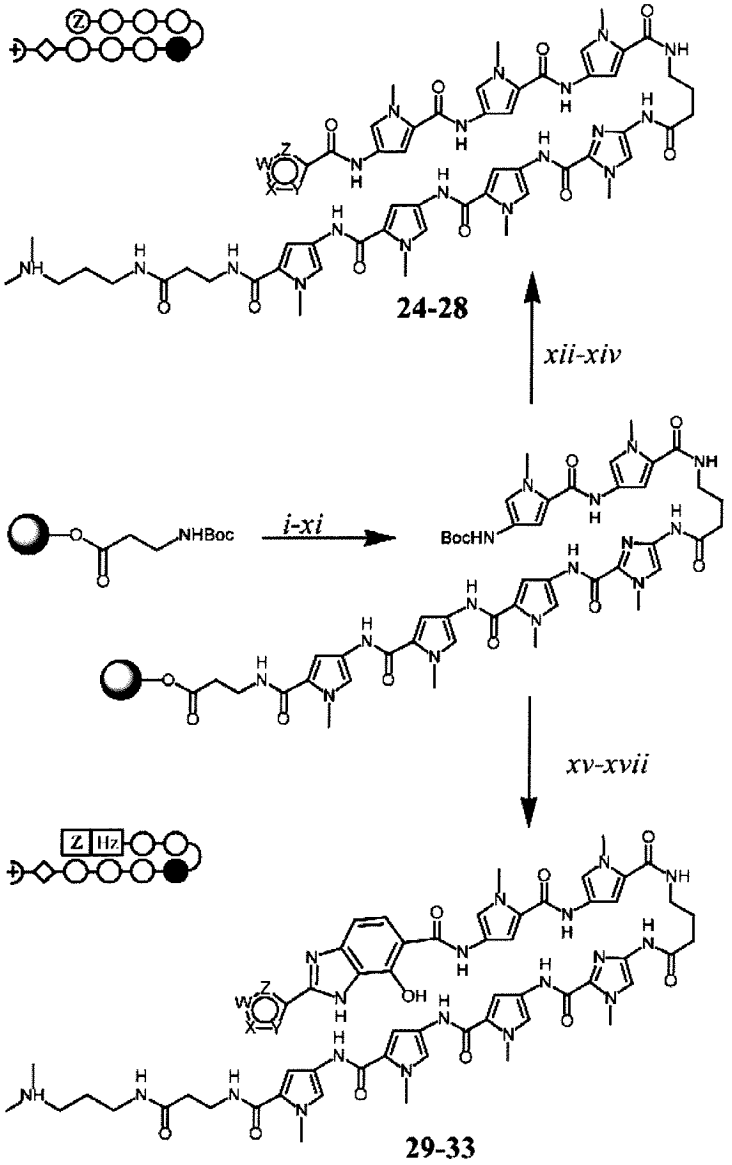


Figure 5. Solid Phase Synthesis of Hairpin Polyamides. i) TFA/DCM; ii) Boc-Py-OBt, DIEA, DMF; iii) Ac_2O, DIEA, DMF; iv) Steps i-iii (x2); v) TFA/DCM; vi) Boc-Im-OH, HBTU, DIEA, DMF; vii) Ac_2O, DIEA, DMF; viii) TFA/DCM; ix) Boc-γ-OH, HBTU, DIEA, DMF; x) Ac_2O, DIEA, DMF; xi) Steps i-iii (x2); xii) Steps i-iii (x1); xiii) Terminal Heterocyclic Acid, HBTU, DIEA, DMF; xiv) Dp, 80 ^{o}C, 2h; xv) TFA/DCM; xvi) Terminal Heterocyclic Dimer, HBTU, DIEA, DMF; xvii) Dp, 80 ^{o}C, 2h. Polyamides: (**24, 29**) W = C-H, X = C-H, Y = N, Z = N-Me; (**25, 30**) W = C-H, X = C-H, Y = S, Z = C-Me; (**26, 31**) W = C-H, X = C H, Y = S, Z = C-H; (**27, 32**) W = C-Me, X = O, Y = N, Z = C-H; (**28, 33**) W = O, X = C-H, Y = N, Z = C-Me.

6.3 Results

DNA Affinity and Sequence Specificity. Quantitative DNase-I footprinting titrations were carried out for polyamides **24-33**. All polyamides were footprinted on the 285-base-pair PCR product of plasmid pCW15. In all cases, the DNA-sequence specificity at cap position (in bold) was determined by varying a single DNA base pair within the sequence, 5'-T**X**TACA-3', to all four Watson-Crick base pairs (**X** = A, T, G, C) and comparing the relative affinities of the resulting complexes. The variable base-pair position was designed to be adjacent to the **Hz** ring, which has been shown to specify for T when paired across from **Py**, so as to be able to determine the binding properties of each compound to the following two base-pair *sequences*: **AT**, **TT**, **GT** and **CT**.

Hairpins **Im**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**24**) and **Ct**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**26**) serve as the parent compounds from which the dimer-cap containing polyamides are derived. As shown by their binding profiles (**Figure 6.6**), both systems are well characterized with the **Im** cap hairpin exhibiting specificity for G at the cap position and the **Ct** cap hairpin exhibiting specificity for T (**Table 1**).

The sequence specificity of the **Ct-Hz** and **Im-Hz** dimers for 5'-T**XT**ACA-3' were evaluated in polyamides **31** and **29** respectively. As expected polyamide **31** bound its designed match site 5'-T**TT**ACA-3' (K_a = 2.4 x 10^9) (**Figure 6.6**) with both the **Ct** and the **Hz** halves of the dimer preferring to rest over the less bulky T in a T•A base pair. To our delight, placing the **Ct** ring adjacent to the **Hz** seemingly increased the specificity of the thiophene ring for T > A from ~4-fold in a **Ct**-Py system to 10-fold on the **Ct-Hz** system. Polyamide **29**, which contains the **Im-Hz** dimer, did not bind its designed match

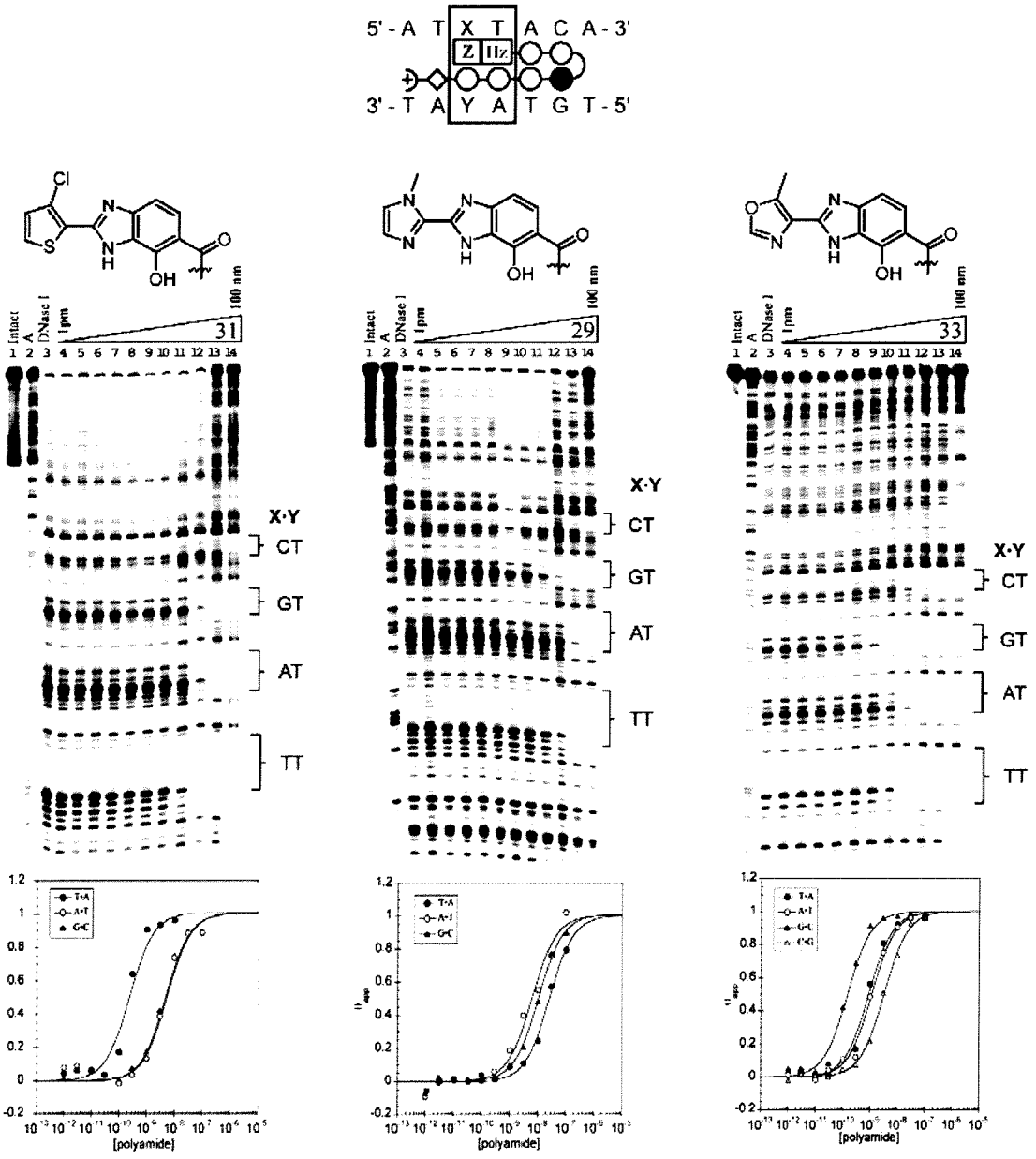


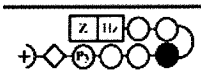
Figure 6. Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **31**, **29**, and **33** respectively, on the 278 bp, 5'-end-labelled PCR product of plasmid CW15: lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (top) Chemical structure of the pairing of interest; and (bottom) Binding isotherms for the four designed sites. θ norm values were obtained according to published methods.[X] A binding model for the hairpin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits.

site 5'-T**GT**ACA-3' with any appreciable level of specificity exhibiting affinities of (K_a = 1.6 x 10^8) and (K_a = 4.0 x 10^8) for the **GT** and **AT** sites respectively.

The binding properties of **Is** and **No** were accessed in the very well characterized 8-ring hairpin system as a means of screening them for any specificity for G•C over C•G for eventual use as **Im** replacements in the dimer caps. Polyamide **27**, which contains the isoxazole end-cap showed a complete lack of specificity for any of the Watson-Crick base pairs. Polyamide **28** which contains the oxazole end-cap, however, displayed a 5.5-fold preference for G > C.

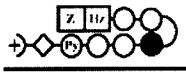
Both the isoxazole (polyamide **32**) and oxazole caps (polyamide **33**) were incorporated into the dimer cap system and their affinities for their designed match site, 5'-T**GT**ACA-3', accessed. Polyamide **32** was unable to discriminate between the **AT**, **TT** and **GT** sites with affinities of 1.1, 1.0 and 2.0 x 10^9 M^{-1} respectively. Polyamide **33** successfully targeted its designed match site with an appreciable level of specificity (25-fold) and a match site affinity of K_a = 6.8 x 10^9 M^{-1} (**Table 1**). Thanks to the development of the chlorothiophene and oxazole dimeric caps, the range of targetable sequences by polyamides has been expanded (**Table 2**).

Table 1. Affinities of Z/Py Ring Pairs Proximal to a Hydroxybenzimidazole Bicycle K_a $[M^{-1}]^{a,b}$

	5'-**A**TAC-3'	5'-**T**TAC-3'	5'-**G**TAC-3'	5'-**C**TAC-3'
ImHz/Py	9.7(±0.7)x 10^7	4.5(±0.6)x 10^8	1.7(±0.4)x 10^8	≤ 1.0 x 10^7
CtHz/Py	2.1(±0.3)x 10^8	**2.4(±0.2)x 10^9**	2.6(±0.4)x 10^8	≤ 1.0 x 10^7
NoHz/Py	8.6(±0.3)x 10^8	9.5(±0.3)x 10^8	**6.8(±0.4)x 10^9**	2.7(±0.5)x 10^8

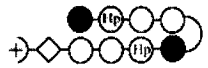
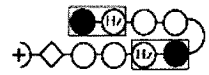
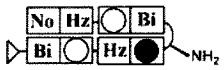
a) Values reported are the mean values from at least three DNase I footprinting titration experiments, with the standard deviation given in parentheses. b) Assays were performed at 22 °C in a buffer of 10 mM Tris.HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0.

Table 2. Specificities of Z/Py Ring Pairs Proximal to a Hydroxybenzimidazole Bicycle K_a $[M^{-1}]^{a,b}$

	5'-ATAC-3'	5'-TTAC-3'	5'-GTAC-3'	5'-CTAC-3'
ImHz/Py	-	-	-	-
CtHz/Py	-	+	-	-
NoHz/Py	-	-	+	-

The synthesis of oligomer **36** was smoothly achieved via the stepwise addition of Boc-amino acid dimers in the same manner as previously described polyamide syntheses. The oligomer's binding properties were assessed in the same context as the compound it was derived from, polyamide **34**. Footprinting of the oligomer on the previously characterized plasmid DEH10 showed a binding affinity of K_a = 2.3 x 10^{10} M^{-1} for the match site 5`-GTAC-3` and affinities of 3.5 x 10^9 M^{-1} and 9.8 x 10^8 M^{-1} for the mismatch sites 5`-GAAC-3` and 5`-GATC-3` (**Table 3**). Such a result demonstrates that a compound consisting exclusively of 6-5 fused ring systems is able to maintain good levels of specificity and an excellent binding affinity.

Table 3. Affinities of Oligomer and Parent Polyamides K_a $[M^{-1}]^{a,b}$

	Polyamide	5'-aGTACt-3'	5'-aGAACt-3'	5'-aGATCt-3'
(34)		7.0 (± 1.8) x 10^8	≤ 1.0 x 10^7	≤ 1.0 x 10^7
(35)		4.6 (± 0.8) x 10^8	3.2 (± 0.4) x 10^7	1.7 (± 0.5) x 10^7
(36)		2.3 (± 0.2) x 10^{10}	3.5 (± 0.6) x 10^9	9.8 (± 0.7) x 10^8

a) Values reported are the mean values from at least three DNase I footprinting titration experiments, with the standard deviation given in parentheses. b) Assays were performed at 22 °C in a buffer of 10 mM Tris.HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0.

Molecular Modeling. Modeling calculations were preformed using the 'Spartan Essential' software package. All *Ab initio* calculations were conducted using the *Hartree-Fock* model and a 6-31G* polarization basis set. Two dimers, **Im-Hz** and **Im-Py**, were modeled to examine the differences in polyamide curvature and proximity to the DNA bases and to establish a rational for the decreased ability of Im to bind G > C when adjacent to the 6-5 fused hydroxybenzimidazole ring (**Figure 6.7**).

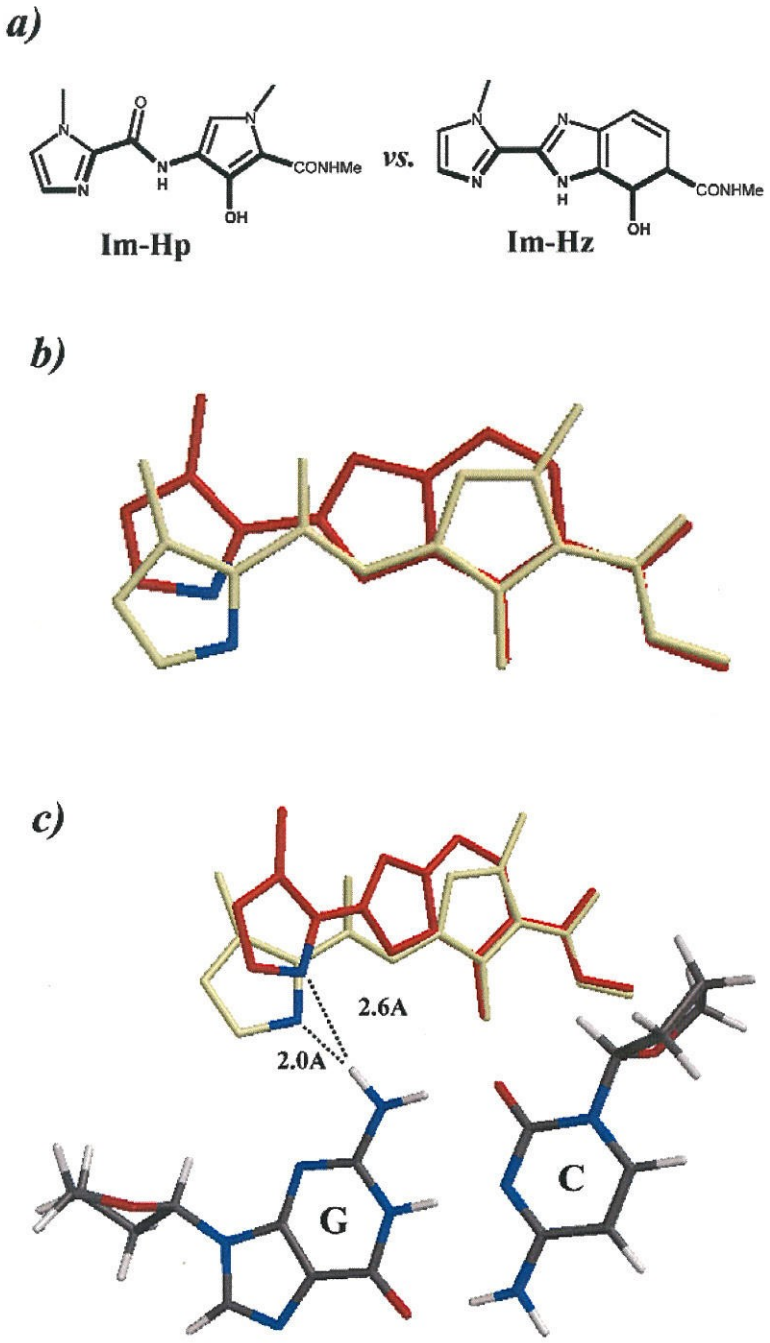


Figure 7. Comparative Geometry of Imidazole-Hydroxypyrrole (Im-Hp) and Imidazole-Hydroxybenzimidazole (Im-Hz) dimeric units. (a) Structure of Im-Hp and Im-Hz with DNA grove contacts shown in bold. (b) Overlay shows Im-Hp in tan, Im-Hz in red and the imidazole nitrogen in blue. Hydrogen not shown. (c) Overlay of Im-Hp:DNA and Im-Hz:DNA complexes showing the difference in distance of the imidazole nitrogen to the exocylic amine of guanine. Coordinates for the Im-Hp:DNA complex taken from preveiously reported crystal structure data.[X]

6.4 Discussion.

Recent advances in hairpin polyamide designs have traditionally focused on developing new modes of single base pair recognition by heterocyclic ring pairings. Our studies, however, have highlighted the fact that the microstructure of DNA highly depends on the sequence in question.[18] Thus we have taken a more global view of molecular recognition where our efforts have shifted from designing molecules that target DNA base-pairs to designing those that target DNA sequences.

The **Ct-Hz** dimer cap represents our first effort to target a short sequence of DNA using sequence-inspired recognition elements. Previous efforts to target two consecutive thymine bases using adjacent **Hp** rings yielded compounds with poor affinities and no specificity. Studies have also shown that **Hz** exhibits specificity for T > A at the N-1 position – relative to the polyamide N-terminus – and that **Ct** polyamides exhibited specificity for T > A at the cap position with excellent polyamide affinities. We hoped that a hybrid dimer would impart excellent specificity for the TT sequence while maintaining a biologically relevant affinity – something that compounds containing multiple **Hp** were incapable of doing. Polyamide **31** bound its designed match site 5'-T**TT**ACA-3' with a sub-nanomolar affinity of K_a = 2.4 x 10^9 M^{-1} with an increased specificity of the chlorothiophene ring for T•A over A•T from ~4-fold in a **Ct**-Py system to 10-fold on the **Ct-Hz** system. This seemingly synergistic effect is attributed to the fact that both the exocyclic chlorine and hydroxyl groups prefer to lie across the less bulky thymine base in an A•T base pair and that the –OH of the **Hz** ring is able to form an energetically favorable hydrogen bond with the O2 carbonyl of thymine.[11] Combined,

these attributes makes this dimer the preferred system for targeting consecutive thymine residues.

We next looked to evaluate how the G specific **Im** ring would perform once incorporated in the **Im-Hz** dimer cap (polyamide **29**). We had hoped that the hairpin would prefer it's match sequence of 5'-T**GT**ACA-3' and were surprised when it failed to demonstrate any preference for its designed site in addition to displaying a significantly decreased average affinity of ~10^8 M^{-1}. Since the specificity of the **Im** ring arises from its ability to accept a hydrogen bond from the exocyclic amine of guanine, we hypothesize that the less curved **Hz** ring could be responsible for the apparent loss of **Im** specificity. To test our hypothesis we took to modeling both the Im-Hp dimer and the **Im-Hz** dimer cap and compared the resulting differences in the Im position (**Figure 6.7**). As expected, linking the Im residue to the **Hz** 6-5 fused ring system via a linear carbon-carbon bond results in a dimer which effectively pulls the terminal Im residue away from the minor groove floor and, in turn, away from the exocyclic amine of guanine (**Figure 6.7b**, red structure). **Figure 6.7c** illustrates how this structural change could affect the cap's ability to hydrogen bond and thus specifically bind across from guanine.

The shortcomings of the **Im-Hz** dimer prompted a search for a ring system that was capable of specifying for G > C within the **Xx-Hz** context. The isoxazole (**Is**) and oxazole (**No**) caps (**Figure 6.1**) were considered because of their structural resemblance to **Im** -- all three rings present a nitrogen atom capable of hydrogen bonding to the minor groove. The binding profiles of both rings were accessed within the standard 8-ring hairpin design and it was found that **No** exhibited a 5.5-fold preference for G > C, while **Is** showed no appreciable specificity for any of the four base pairs. After evaluating the

electron densities on the atom presented to the minor groove of several end-caps it was found that of the nitrogen presenting rings, **No** presented the most electron rich nitrogen to the minor groove and would be the most likely candidate to successfully hydrogen bond. Interestingly, **Is**, which was found to be degenerate at the cap position, was calculated to have even less electron density that **Im** thus backing the thermodynamics results. Further substantiating our claims was the inability of the **No-Hz** dimer to selectively target 5'-T**GT**ACA-3'. To our delight, when the **No-Hz** dimer was incorporated into polyamide **33**, it was found to be specific for its designed sequence of 5'-T**GT**ACA-3' with a 25-fold preference for G > C and an affinity of $K_a = 6.8 \times 10^9 M^{-1}$ at its match site (**Figure 6.6**). The **No-Hz** dimer presents the same functionality to the minor groove as the **Im-Hz** dimer, but with an increased negative charge on the guanine selective, endocyclic nitrogen. This increased charge may enhance it's ability to hydrogen bond to the exocyclic amine of guanine thus explaining it's ability to target a G-T site.

As a first step in the design of polyamide-based oligomers, we incorporated the newly synthesized **No-Hz** dimer into a compound completely consisting of 6-5 fused ring systems. The oligomer, which was designed to target the site 5`-GTAC-3`, is a direct evolution from the parent polyamide **34**. In order to evaluate the impact of removing four amide bonds and moving to a system consisting of only 6-5 fused recognition elements, its binding properties were evaluated and compared to polyamide **34**. Oligomer **36** was found to bind its match site with an impressive affinity of $2.3 \times 10^{10} M^{-1}$ while discriminating against its mismatch sites of 5`-GAAC-3` and 5`-GATC-3` with specificities of ~7 and ~ 23 fold respectively (**Figure 6.8**). The oligomer's affinity for its

match site represents a 33 fold increase in binding affinity when compared to the parent polyamide **34** and a 50 fold increase compared to the second generation polyamide **35** (**Table 3**). A stepwise evolution of polyamide **34** to polyamide **35** and finally to oligomer **36** shows a removal of 2 amide bonds per molecule. Removing amide bonds has previously been hypothesized to be detrimental to polyamide binding affinities since it represents the abolition of energetically favorable hydrogen bonds formed with the base pairs in the minor groove. While this may be the case, we believe that oligomer **36** is able to overcome the loss of four hydrogen bonds due to the fact that the 6-5 fused rings represent a system that is pre-organized to bind in the minor groove. Since the recognition elements in the oligomer are linked by carbon-carbon bonds, there are less degrees of freedom and the compound may incur a reduced entropic penalty for arranging itself to bind DNA. In addition, the oligomer's large hydrophobic face may prefer the environment of the minor groove which, when compared to the surrounding aqueous solution, is relatively water-free.

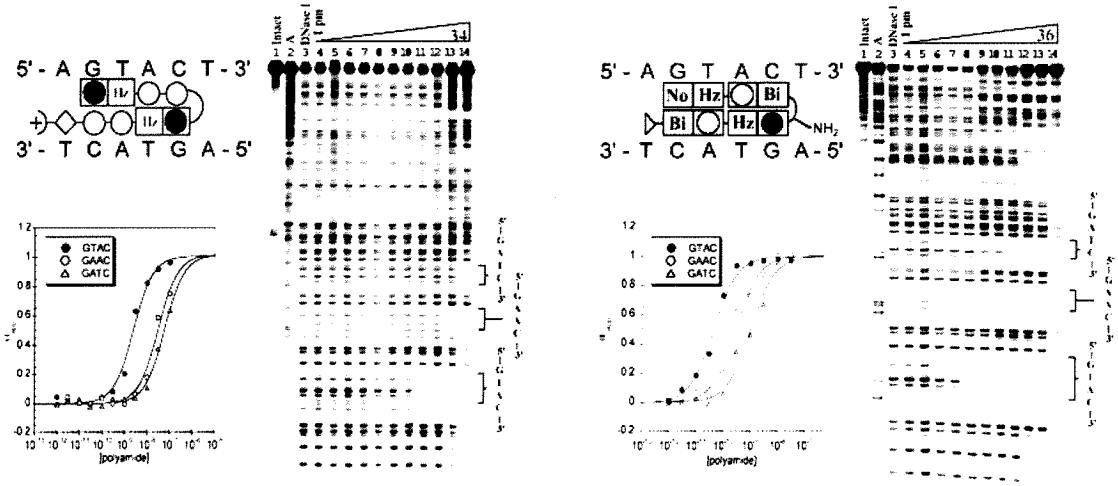


Figure 8. Quantitative DNase I footprinting experiments in the hairpin motif for polyamide **34** and oligomer **36** respectively, on the 5'-end-labelled PCR product of plasmid DEH10: lane 1, intact DNA; lane 2, A reaction; lane 3, DNase I standard; lanes 4-14, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (top) Chemical structure of the pairing of interest; and (bottom) Binding isotherms for the four designed sites. θ norm values were obtained according to published methods.[X] A binding model for the hairpin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; the methyl amide tail is shown as a triangle; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits.

6.5 Conclusion.

Eight-ring hairpin polyamides containing the **No-Hz** and **Ct-Hz** dimer caps are able to target G-T and T-T sequences at the polyamide N-terminus with good specificity and affinity and represent one of our most effective recognition systems. The **No-Hz** and **Ct-Hz** dimer caps denote our first successful attempts to shift the focus of our recognition element designs away from single base pair interactions and towards polyamide/DNA sequence recognition. In addition, the development of the **No-Hz** cap has allowed for the design, synthesis and characterization of our first DNA binding oligomer. We are encouraged by the fact that this oligomer demonstrates good levels of specificity while exhibiting an excellent affinity for DNA. We hope to apply the development of these new technologies to the targeting of sequences that have proven to be problematic in the past.

6.6 Experimental

Heterocycle Synthesis. The synthesis of compounds **1** (**Im**-OH), **2** (**Tn**-OH), **3** (**Ct**-OH), **10**, **15** (**Im-Hz**(OMe)-OMe), **20** (**Im-Hz**(OMe)-OH) and **35** (aryl diamine) have previously been reported (**Figure 6.3**).[13, 15]

Methyl 5-methylisoxazole-3-carboxylate (Is-OMe) 5: Compound **5** was synthesized as previously described in the literature from 2,4-dioxopentanoic acid methyl ester **4**.

5-methylisoxazole-3-carboxylic acid (Is-OH) 6: A mixture of **5** (1 g, 7.08 mMol), 1N NaOH (10 mL) and MeOH (5 mL) was stirred at 40 °C for 4 h. The MeOH was removed by rotoevaporation and the pH adjusted to pH = 2 with 1N HCl. The precipitate was extracted with EtOAc (3 x 10 mL), the organics dried over sodium sulfate and removed by rotoevaporation to provide **6** (792 mg, 88% Yield) as a fibrous white solid. TLC (3:2 EtOAc/Hex +10% AcOH) R_f 0.4; ^{1}H NMR (DMSO-d_6) 6.54 (s, 1H), 2.45 (s, 3H); ^{13}C (DMSO-d_6) 171.74, 161.21, 157.17, 102.47, 11.88; EI-MS m/e 127.027 (M^+ calcd for 127.027 $C_5H_5NO_3$).

Methyl 5-methyl-1,3-oxazole-4-carboxylate (No-OMe) 8: To a mixture of methyl-alpha-isocyanoacetate (3 g, 30.2 mmol) and DBU (4.5 g, 30.2 mmol) in dry THF (40 mL) at 10 °C, was added acetic anhydride (3.06 g, 3.0 mmol) in dry THF (10 mL) dropwise. The reaction was allowed to warm to room temperature and stirred for 10 h. The solvent was removed by rotoevaporation and water (100 mL) was added. The mixture was extracted with EtOAc (2 x 100 mL). The organic layer was collected and dried over sodium sulfate. Removal of organics by rotoevaporation provided **8** as a crude amber oil. The oil was subjected to column chromatography using (3:2 Hex/EtOAc) to provide **8**

(2.73 g, 64% Yield) as a crystalline white solid. TLC (3:2 Hex/EtOAc) R_f 0.5; ^{1}H NMR (DMSO-d_6) 8.32 (s, 1H), 3.77 (s, 3H), 2.55 (s, 3H); ^{13}C (DMSO-d_6) 161.9, 156.1, 150.3, 126.2, 51.5, 11.5; EI-MS m/e 141.043 (M^+ calcd for 141.043 $C_6H_7NO_3$).

5-Methyl-1,3-oxazole-4-carboxylic acid (No-OH) 9: A mixture of **8** (1 g, 7.08 mMol), 1N NaOH (10 mL) and MeOH (5 mL) was stirred at 40 °C for 4 h. The MeOH was removed by rotoevaporation and the pH adjusted to pH = 2 with 1N HCl. The precipitate was extracted with EtOAc (3 x 10 mL), the organics dried over sodium sulfate and removed by rotoevaporation to provide **9** (738 mg, 82% Yield) as a fibrous white solid. TLC (3:2 EtOAc/Hex +10% AcOH) R_f 0.4; ^{1}H NMR (DMSO-d_6) 8.27 (s, 1H), 2.45 (s, 3H); ^{13}C (DMSO-d_6) 163.0, 155.6, 150.1, 145.6, 11.6; EI-MS m/e 127.027 (M^+ calcd for 127.027 $C_5H_5NO_3$).

Tn-Hz(OMe)OMe 16: To a solution of **2** (500 mg, 3.52 mmol) in DMF (6 mL) was added DIEA (.676 mL, 3.87 mmol) and HBTU (1.26 g, 3.34 mmol). The mixture was stirred at room temperature for 1 h, followed by the addition of the aryl diamine **35** (690 mg, 3.52 mmol). The reaction was then heated to 35 °C and stirred for an additional 24 h. The reaction was allowed to cool to room temperature and then poured into a Falcon tube containing cold water (40 mL) upon which time a cloudy precipitate formed. The Falcon tube was centrifuged at 14000 rpm for 10 min, the mother liquor decanted, and the precipitated solid dried under high vacuum. After drying, the crude solid (**11**) was dissolved in acetic acid (5 mL) and heated to 90 °C. The reaction was stirred for 12 h, followed by removal of the solvent by rotoevaporation. The resultant residue was subjected to column chromatography (3:2 EtOAc/Hex) to provide **16** as an off-white solid (447 mg, 42% Yield).

Tn-Hz(OMe)OMe 17: To a solution of **3** (294 mg, 1.80 mmol) in DMF (5 mL) was added DIEA (.348 mL, 1.98 mmol) and HBTU (650 mg, 1.72 mmol). The mixture was stirred at room temperature for 1 h, followed by the addition of the aryl diamine **35** (300 mg, 1.80 mmol). The reaction was then heated to 35 °C and stirred for an additional 24 h. The reaction was allowed to cool to room temperature and then poured into a Falcon tube containing cold water (40 mL) upon which time a cloudy precipitate formed. The Falcon tube was centrifuged at 14000 rpm for 10 min, the mother liquor decanted, and the precipitated solid dried under high vacuum. After drying, the crude solid (**12**) was dissolved in acetic acid (5 mL) and heated to 90 °C. The reaction was stirred for 12 h, followed by removal of the solvent by rotoevaporation. The resultant residue was subjected to column chromatography (3:2 EtOAc/Hex) to provide **17** as an off-white solid (647 mg, 57% Yield).

Methyl 7-methoxy-2-(5-methylisoxazole-3-yl)benzimidazole-6-carboxylate (Is-Hz(OMe)OMe) 18: To a solution of **6** (0.3 g, 2.36 mmol) in DMF (4 mL) was added DIEA (915 mg, 1.23 mL, 7.08 mmol) and HBTU (895 mg, 2.36 mmol). The mixture was stirred at room temperature for 1 h, followed by the addition of the aryl diamine **35** (463 mg, 2.36 mmol). The reaction was then heated to 35 °C and stirred for an additional 24 h. The reaction was allowed to cool to room temperature and then poured into a Falcon tube containing cold water (40 mL) upon which time a cloudy precipitate formed. The Falcon tube was centrifuged at 14000 rpm for 10 min, the mother liquor decanted, and the precipitated solid dried under high vacuum. After drying, the crude solid (**13**) was dissolved in acetic acid (5 mL) and heated to 90 °C. The reaction was stirred for 12 h, followed by removal of the solvent by rotoevaporation. The resultant residue was

subjected to column chromatography (3:2 EtOAc/Hex) to provide **18** as a thin film. Addition of hexanes to the film, followed by rotoevaporation and drying under high vacuum provided **18** as a white solid (305 mg, 45% Yield). TLC (3:2 EtOAc/Hex) R_f 0.5; ^{1}H NMR (DMSO-d_6) 7.60 (d, 1H, J = 8.4 Hz), 7.21 (d, 1H, J = 8.4 Hz), 6.86 (s, 1H), 4.31 (s, 3H), 3.79 (s, 3H) 2.51 (s, 3H); ^{13}C (DMSO-d_6) 171.3, 168.0, 154.9, 151.7, 142.2, 141.0, 135.2, 127.1, 115.8, 105.8, 101.4, 61.2, 52.1, 11.8; EI-MS m/e 287.091 (M^+ calcd for 287.090 $C_{14}H_{13}N_3O_4$).

Methyl 7-methoxy-2-(5-methyl(1,3-oxazole-4-yl))benzimidazole-6-carboxylate (NoHz(OMe)OMe) 19: To a solution of **9** (0.3 g, 2.36 mmol) in DMF (4 mL) was added DIEA (915 mg, 1.23 mL, 7.08 mmol) and HBTU (895 mg, 2.36 mmol). The mixture was stirred at room temperature for 1 h, followed by the addition of the aryl diamine **35** (463 mg, 2.36 mmol). The reaction was then heated to 35 °C and stirred for an additional 24 h. The reaction was allowed to cool to room temperature and then poured into a separatory funnel containing water (200 mL). The water was then extracted with EtOAc (2 x 100 mL). The organic layer was dried over sodium sulfate and removed by rotoevaporation to provide **14** as a crude residue. **14** was then dissolved in acetic acid (5 mL) and heated to 90 °C. The reaction was stirred for 12 h, followed by removal of the solvent by rotoevaporation. The resultant residue was subjected to column chromatography (4:1 EtOAc/Hex) to provide **19** as a thin film. Addition of hexanes to the film, followed by rotoevaporation and drying under high vacuum provided **19** as a white solid (379 mg, 56% Yield). TLC (4:1 EtOAc/Hex) R_f 0.75; ^{1}H NMR (DMSO-d_6) 8.49 (s, 1H), 7.51 (d, 1H, J = 8.7 Hz), 7.16 (d, 1H, J = 8.7 Hz), 4.34 (s, 3H), 3.77 (s, 3H) 2.78 (s, 3H); ^{13}C

(DMSO-d_6) 166.7, 151.0, 149.3, 145.5, 139.0, 135.3, 125.6, 124.8, 114.2, 105.3, 60.8, 51.6, 11.3; EI-MS m/e 287.091 (M^+ calcd for 287.090 $C_{14}H_{13}N_3O_4$).

(Tn-Hz(OMe)OH) 21: A mixture of **16** (500 mg, 0.97 mMol), 1N NaOH (8 mL) and MeOH (8 mL) was stirred at 40 °C for 6 h. The MeOH was removed by rotoevaporation and the aqueous layer washed with EtOAc (3 x 10 mL). The precipitate was extracted with EtOAc (3 x 10 mL), the organics dried over sodium sulfate and removed by rotoevaporation to provide **21** (338 mg, 71% Yield) as an off-white solid.

(Ct-Hz(OMe)OH) 22: A mixture of **17** (500 mg, 1.55 mMol), 1N NaOH (8 mL) and MeOH (8 mL) was stirred at 40 °C for 5 h. The MeOH was removed by rotoevaporation and the aqueous layer washed with EtOAc (3 x 10 mL). The precipitate was extracted with EtOAc (3 x 10 mL), the organics dried over sodium sulfate and removed by rotoevaporation to provide **22** (330 mg, 69% Yield) as a light yellow solid.

(Is-Hz(OMe)OH) 23: A mixture of **18** (280 mg, 0.97 mMol), 1N NaOH (2 mL) and MeOH (1 mL) was stirred at 40 °C for 4 h. The MeOH was removed by rotoevaporation and the pH adjusted to pH = 2 with 1N HCl. The precipitate was extracted with EtOAc (3 x 10 mL), the organics dried over sodium sulfate and removed by rotoevaporation to provide **23** (226 mg, 85% Yield) as a white solid. TLC (3:2 EtOAc/Hex +10% AcOH) R_f 0.4; ^{1}H NMR (DMSO-d_6) 7.62 (d, 1H, J = 8.7 Hz), 7.21 (d, 1H, J = 8.7 Hz), 6.87 (s, 1H), 4.30 (s, 3H), 2.51 (s, 3H); ^{13}C (DMSO-d_6) 171.2, 167.5, 155.3, 151.7, 142.0, 139.3, 135.3, 126.5, 115.7, 105.7, 101.0, 61.2, 11.8; EI-MS m/e 273.075 (M^+ calcd for 273.075 $C_{13}H_{11}N_3O_4$).

7-methoxy-2-(5-methyl(1,3-oxazole-4-yl))benzimidazole-6-carboxylic acid (NoHz(OMe)OH) 24: A mixture of **19** (200 mg, 0.69 mMol), 1N NaOH (2 mL) and MeOH (1 mL) was stirred at 40 °C for 4 h. The MeOH was removed by rotoevaporation and the pH adjusted to pH = 2 with 1N HCl. The precipitate was extracted with EtOAc (3 x 10 mL), the organics dried over sodium sulfate and removed by rotoevaporation to provide **24** (156 mg, 82% Yield) as a white solid. TLC (3:2 EtOAc/Hex +10% AcOH) R_f 0.35; ^{1}H NMR (DMSO-d_6) 8.50 (s, 1H), 7.54 (d, 1H, J = 8.4 Hz), 7.18 (d, 1H, J = 8.4 Hz), 4.29 (s, 3H), 2.78 (s, 3H); ^{13}C (DMSO-d_6) 168.1, 151.5, 149.8, 138.8, 136.1, 126.2, 125.6, 61.5, 11.8; EI-MS m/e 273.075 (M^+ calcd for 273.075 $C_{13}H_{11}N_3O_4$).

Polyamide Synthesis: Polyamides were synthesized from Boc-β-alanine-Pam resin (50 mg, 0.59 mmol/g) and purified by preparatory HPLC according to published manual solid phase protocols.[17] The synthesis of batch resin **BR1** (H_2N-Py-Py-γ-Im-Py-Py-Py-β-Pam), **BR2** (H_2N-Py-Py-Py-γ-Im-Py-Py-Py-β-Pam) and polyamides **25-27** has previously been reported.[13] (**Figure 6.4**).

Is-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**28**): (**Is**-OH) (11.2 mg, 88.5 μmol) was incorporated by activation with HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μl, 177 μmol) and DMF (250 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing base resin **BR2** H_2N-Py-Py-Py-γ-Im-Py-Py-Py-β-Pam. Coupling was allowed to proceed for 12 h at 37 °C. The resin was then washed with DCM followed by the addition of Dp (500 μL). The mixture was allowed to stand for 2 h at 85 °C with occasional agitation. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was purified by reversed phase HPLC to

provide **Is**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**28**) (2.8 mg, 7.7 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1223.56 (M+H calcd for 1223.56 $C_{58}H_{71}N_{20}O_{11}$).

No-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**29**): (**No**-OH) was incorporated as described above for **Is**-OH (polyamide **28**) to provide **No**-Py-Py-Py-γ-Im-Py-Py-Py-β-Dp (**29**) (3.1 mg, 8.5 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1223.58 (M+H calcd for 1223.56 $C_{58}H_{71}N_{20}O_{11}$).

Im-Hz-Py-Py-γ-Im-Py-Py-Py-β-Dp (**30**): (**Im-HzOMe**-OH) (25 mg, 88.5 μmol) was incorporated by activation with HBTU (32 mg, 84 μmol), DIEA (23 mg, 31 μl, 177 μmol) and DMF (250 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing base resin **BR1** H_2N-Py-Py-γ-Im-Py-Py-Py-β-Pam. Coupling was allowed to proceed for 12 h at room temperature. The resin-bound polyamide was then washed with DCM and treated as described in the deprotection protocol below to provide **Im-Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**30**) (0.9 mg, 2.4 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1232.55 (M+H calcd for 1232.56 $C_{59}H_{70}N_{21}O_{10}$).

Tn-Hz-Py-Py-γ-Im-Py-Py-Py-β-Dp (**31**): (**Tn-HzOMe**-OH) was incorporated as described above for **Im-Hz**-OH (polyamide **30**) to provide **Tn-Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**31**) (1.1 mg, 3.0 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1248.55 (M+H calcd for 1248.53 $C_{60}H_{70}N_{19}O_{10}S$).

Ct-Hz-Py-Py-γ-Im-Py-Py-Py-β-Dp (**32**): (**Ct-HzOMe**-OH) was incorporated as described above for **Im-Hz**-OH (polyamide **30**) to provide **Ct-Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**32**) (1.1 mg, 2.9 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1269.47 (M+H calcd for 1269.47 $C_{59}H_{67}ClN_{19}O_{10}S$).

Is-Hz-Py-Py-γ-Im-Py-Py-Py-β-Dp (**33**): (**Is-HzOMe**-OH) was incorporated as described above for **Im-Hz**-OH (polyamide **30**) to provide **Is-Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**33**) (0.8 mg, 2.2 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1233.55 (M+H calcd for 1233.55 $C_{59}H_{69}N_{20}O_{11}$).

No-Hz-Py-Py-γ-Im-Py-Py-Py-β-Dp (**34**): (**No-HzOMe**-OH) was incorporated as described above for **Im-Hz**-OH (polyamide **30**) to provide **No-Hz**-Py-Py-γ-Im-Py-Py-Py-β-Dp (**34**) (1.5 mg, 4.1 % recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS 1233.54 (M+H calcd for 1233.55 $C_{59}H_{69}N_{20}O_{11}$).

Deprotection of the *O*-Methyl-Protected Polyamides. O-Methyl protected polyamides were cleaved from resin, purified, deprotected and subject to further purification using the following general procedure. Upon completion of solid phase synthesis, Dp (500 μL) was added to the synthesis vessel containing the resin (50 mg). The mixture was allowed to stand for 2 h at 85 °C with occasional agitation. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was purified by reversed phase HPLC and lyophilized to provide polyamides containing the *O*-methyl protected hydroxybenzimidazole unit (-**HzOMe-**) as a dry solid. The

polyamides were then dissolved in DMF (200 μl) and added to a suspension of sodium hydride (40 mg, 60% oil dispersion) and thiophenol (200 μl) in DMF (400 μl) that was pre-heated for 5 min at 85 °C. The mixture was heated for 2 h at 85 °C. The mixture was then cooled to 0 °C and 20% TFA (7.0 mL) was added. The aqueous layer was washed three times with diethyl ether (8 mL) and then diluted to a total volume of 9.5 mL using 0.1% TFA. The mixture was then purified by reverse-phase HPLC to give the deprotected **Hz**-containing polyamides **30**-**34**.

6.7 References

[1] Thrum, H., *Naturwissenschaften* **1959,** 46, (2), 87-87.

[2] Finlay, A. C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F. X., *Journal of the American Chemical Society* **1951,** 73, (1), 341-343.

[3] Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E., *Proceedings of the National Academy of Sciences of the United States of America* **1985,** 82, (5), 1376-1380.

[4] Pelton, J. G.; Wemmer, D. E., *Journal of the American Chemical Society* **1990,** 112, (4), 1393-1399.

[5] Lown, J. W.; Krowicki, K.; Bhat, U. G.; Skorobogaty, A.; Ward, B.; Dabrowiak, J. C., *Biochemistry* **1986,** 25, (23), 7408-7416.

[6] Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2001,** 9, (9), 2215-2235.

[7] Helene, C., *Nature* **1998,** 391, (6666), 436-438.

[8] White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B., *Nature* **1998,** 391, (6666), 468-471.

[9] Arcamone, F.; Nicolell.V; Penco, S.; Orezzi, P.; Pirelli, A., *Nature* **1964,** 203, (494), 1064-&.

[10] Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America* **1992,** 89, (16), 7586-7590.

[11] Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1999,** 121, (50), 11621-11629.

[12] Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B., *Helvetica Chimica Acta* **2002,** 85, (12), 4485-4517.

[13] Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2003,** 11, (20), 4333-4340.

[14] Briehn, C. A.; Weyermann, P.; Dervan, P. B., *Chemistry-a European Journal* **2003,** 9, (9), 2110-2122.

[15] Renneberg, D.; Dervan, P. B., *Journal of the American Chemical Society* **2003,** 125, (19), 5707-5716.

[16] Urbach, A. R.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America* **2001,** 98, (8), 4343-4348.

[17] Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1996,** 118, (26), 6141-6146.

[18] Urbach, A. R. L., J. J.; Ross, S. A.; Dervan, P. B., *Journal of Molecular Biology* **2002,** 320, (1), 55-71.

Chapter 7

Targeting Poly G Tracts of DNA

The text of this chapter was taken in part from a manuscript in preparation with David Chenoweth, Michael A. Marques and Professor Peter B. Dervan (Caltech).

(Chenoweth, D.; Marques, M. A.; Doss, R. M.; and Dervan, P. B.)

Abstract.

The evolution of pyrrole (**Py**), hydroxypyrrole (**Hp**) and imidazole (**Im**) based polyamides into DNA-binding oligomers has greatly expanded the targetable range of DNA sequences. Incorporation of the hydroxybenzimidazole (**Hz**), benzimidazole (**Bi**), and imidazopyridine (**Ip**) 6-5 fused rings into these minor groove binding compounds has resulted in systems with improved binding affinities, levels of specificity and stability. In addition to displaying improved molecular recognition properties, these new oligomers posses a unique level of curvature which may be useful for targeting DNA sequences which continue to be problematic. Because the microstructure of DNA has been shown to vary markedly depending on sequence, we are interested in seeing if **Ip** containing oligomers will be able to match the wide, shallow and undercurved minor groove of DNA tracts containing several consecutive guanines. While previously synthesized polyamides containing four consecutive imidazoles designed to target 5`-GGGG-3` were unsuccessful are targeting their match sites with useful affinities, we believe that compounds containing the 6-5 fused recognition elements **Ip** and **Bi** may have improved binding properties.

7.1 Introduction.

Current DNA-binding oligomers have been rationally evolved from the natural products netropsin and distamycin. These small, minor groove biding molecules are capable of targeting A,T tracts of DNA and discriminating against guanine thanks to a clash between the protruding C-H of the pyrrole ring and the exocyclic amine of guanine. The ability to target guanine was achieved by replacing the steric bulk of the pyrrole C-H with a lone pair donated from the N3 of imidazole. This substitution not only relieved the steric clash, but also allowed for the formation of the energetically favorable hydrogen bond between **Im** and the N-H of guanine.

Previous work has shown that recognition elements with heteroatoms presented in the minor groove tend to be overcurved with respect to DNA. This phenomenon can be explained by the fact that the small heteroatom on the DNA recognition side of the ligand "pulls" the polyamide crescent into a tighter semicircle.[2] A single atomic change can have marked effects on the overall shape of the ligand due to the rigidity of the conjugated system – a characteristic which is exacerbated in polyamides containing multiple **Im** residues. Such overcurvature manifests itself in a rapid drop in binding affinity which, in turn, limits the usefulness of the compounds. When targeting a sequence such as 5`-GGGG-3` – one which has been shown to be undercurved – it becomes clear that **Im** containing polyamides are not suitable ligands.

With this in mind, compounds containing a varying amount of undercurved 6-5 fused ring systems were designed and their ability to target G-rich sequences of DNA in the minor groove tested by DNAse I footprinting analysis. In designing compounds to target 5`-GGGG-3`, several 6-5 system containing oligomers were designed to improve

upon the original, unsuccessful four **Im** polyamide. These new oligomers were designed to replace all four "bottom strand" pyrroles with two **Bi-Py** dimers in order to decrease the overall curvature of the compounds and match the unique microstructure of the poly-G tract (**Figure 7.1**). In addition to the introduction of the **Bi-Py** dimers, **Ip-Im** dimers

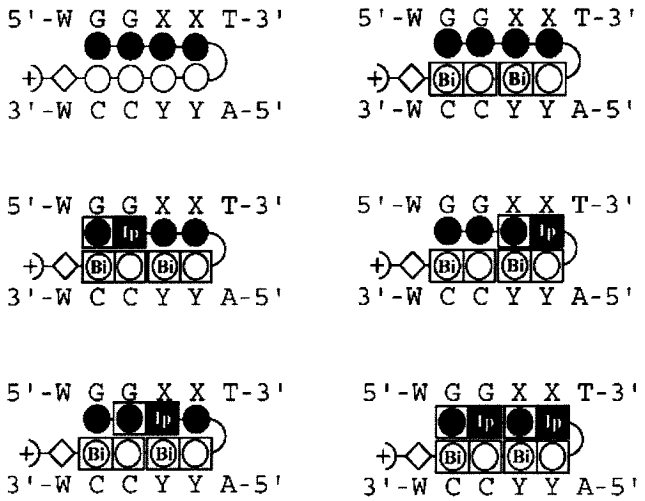


Figure 7.1 Ball and stick models of oligomers in the context of a variable DNA sequence. Examination of flanking sequence effects on oligomer binding are tested by varying W to both A and T Watson-Crick bases. Examination of mismatch specificity is tested by varying X•Y in the core binding sequence to both G•C and A•T respectively. Imidazole is indicated by a dark circle. Pyrrole is indicated by a white circle. Imidazopyridine-imidazole (**Ip-Im**) and benzimidazole-pyrrole (**Bi-Py**) dimers are indicated by dark and light rectangles, respectively. β-alanine is depicted as a diamond and γ-aminobutyric acid is depicted as a half circle.

were introduced into the top strands in various positions in order to probe for the best combination of guanine-specific recognition elements. Recently, Renneberg *et al.* reported the utility of imidazopyridine/pyrrole pair (**Ip/Py**) as a functional imidazole/pyrrole (**Im/Py**) pair mimic.[3] A fused 6-5 bicycle, **Ip** is able to present the same recognition face to the DNA minor groove, coding for the exocyclic amine of guanine via hydrogen bond formation, while offering a significantly different geometry. While we believe that the relaxed shape of the 6-5 systems will be the most important

characteristic of the newly developed ligands, it remains unclear as to whether the best approach would be to replace all eight recognition elements or to employ a combination of traditional and new ring systems (**Figure 7.2**).

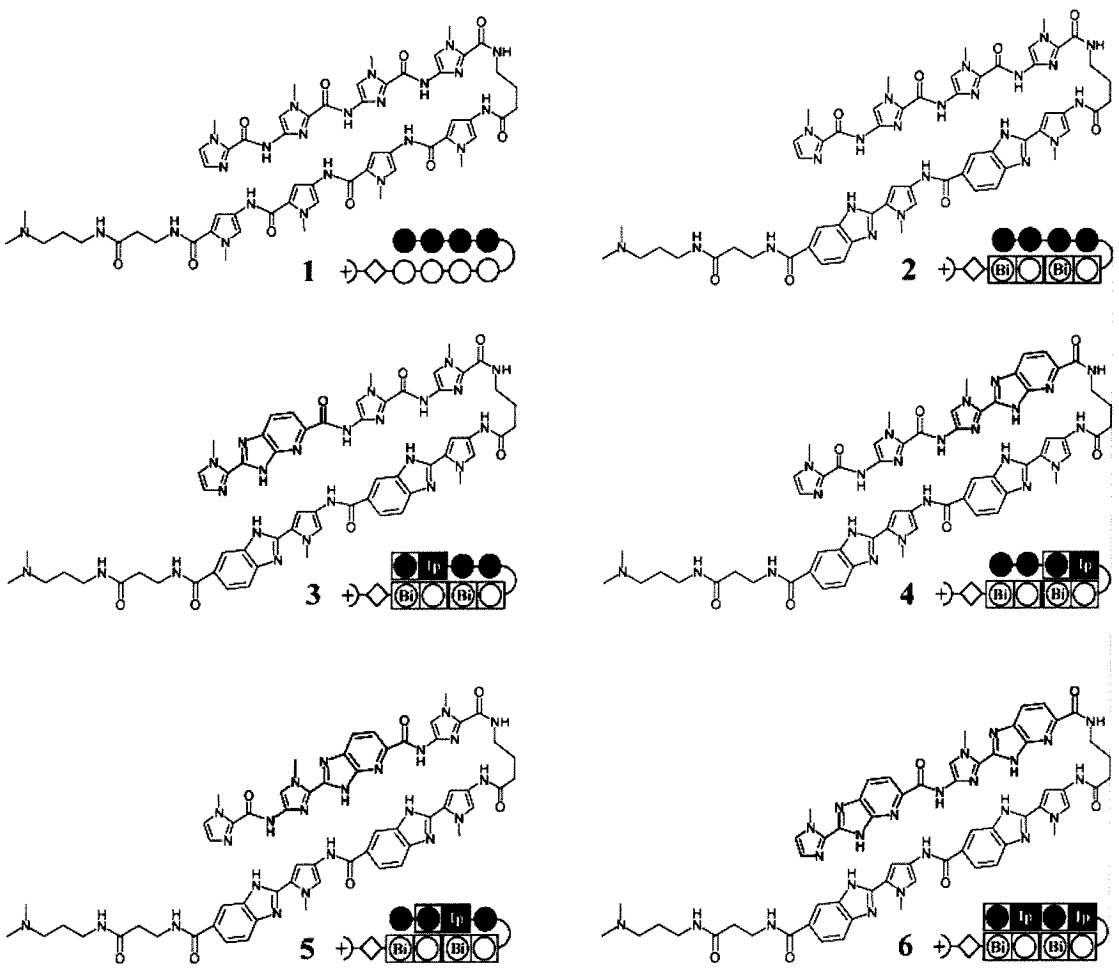


Figure 7.2 Chemical structure of oligomers **1-6** shown with imidazopyridine-imidazole (**Ip-Im**) dimers in bold. With each chemical structure is the accompanying ball and stick model for the oligomer. Imidazopyridine-imidazole (**Ip-Im**) and benzimidazole-pyrrole (**Bi-Py**) dimers are indicated by dark ation of 5'-GGGG-3', 5'-alanine is depicted as a diamond and γ-aminobutyric acid is depicted as a half circle.

7.2 Results

It should be noted that a major portion of the following synthetic efforts were made possible by synthetic contributions from graduate student David Chenoweth while the procedural write-up was compiled by graduate student Michael Marques.

Heterocycle Synthesis (Figure 7.3)

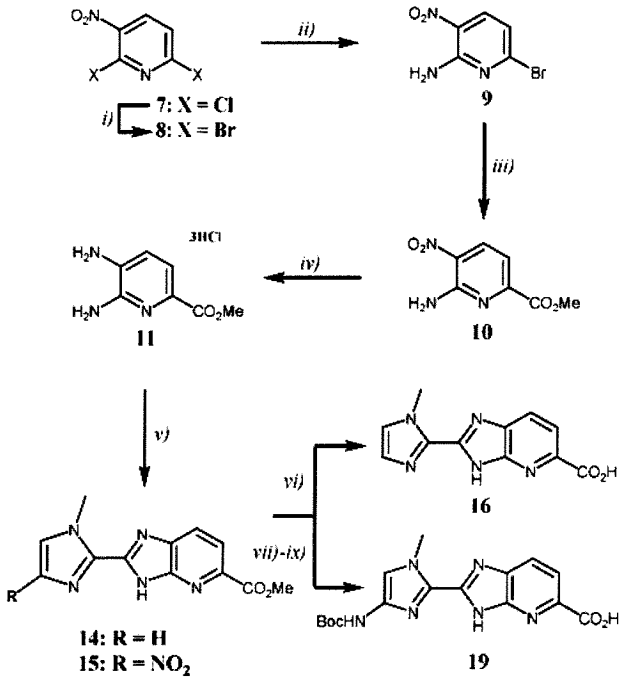


Figure 7.3 Synthesis of imidazopyridine-imidazole dimers Im-Ip-OH (**16**) and Boc-Im-Ip-OH (**19**). i) 30% HBr in AcOH, NaBr, H_2SO_4; ii) NH_3 in MeOH; iii) $Pd(OAc)_2$, MeOH, CO; iv) H_2, Pd/C, DMF; v) Im-CHO (**12**) or NO_2-Im-CHO (**13**), NO_2Ph; vi) 4N KOH, MeOH; vii) H_2, Pd/C, DMF; viii) $(Boc)_2O$, DIEA, DMAP, DMF; ix) 4N KOH, MeOH.

Dimeric units Im-Ip-OH (**16**) and Boc-Im-Ip-OH (**19**) were synthesized starting from commercially available 2,6-dichloro-3-nitropyridine (**7**). Nitropyridine (**7**) was converted to its dibromo derivative (**8**) by halogen exchange at both *ortho*-positions using a combination of HBr in acetic acid and anhydrous HBr.[4] Regioselective amination of (**8**) using a saturated solution of NH_3 in MeOH provided (**9**).[3] Continuous bubbling of anhydrous NH_3 into the reaction vessel provides improved yields. Palladium-catalyzed carbonylation of (**9**) provided (**10**) in moderate yield as previously described.[5, 6]

Reduction of **10** using Pd/C in the presence of hydrogen provided the corresponding diamine (**11**) cleanly and in near quantitative yields. It is noteworthy that increased hydrogen pressure (~500 psi) is not necessary and complete conversion of starting material can be accomplished by equipping the reaction flask with a hydrogen balloon.

From **11**, a variety of conditions can be employed to effect the transformation to fused bicyclic derivatives such as **14** and **15**. First, heterocyclic carboxylic acids can be activated and mixed with their corresponding diamine coupling partners to form the initial amide bond, followed by cyclodehydration to provide the fused benzimidazole system.[7, 8] system.[7, 8] This initial approach did not allow access to the desired imidazopyridine intermediate **14**. A second approach, mixing diaminopyridine intermediate (**11**) and commercially available imidazole aldehyde (**12**) in the presence of an oxidant such as $FeCl_3$ or benzoquinone was only moderately successful, providing **14** in low yield following chromatography.[9] Ultimately, mixing **12** and **11** in refluxing nitrobenzene overnight provided **14** in excellent yield without the need for chromatography.[10-12] **14** was then saponified using a mixture of KOH (4M) in MeOH under heating to provide the final Im-Ip-OH dimer (**16**).

Nitroimidazole (**13**) was synthesized from **12** using a mixture of oleum and neat red fuming nitric acid.[13] All attempts at nitration using a variety of other conditions, such as 70% nitric acid or 90% nitric acid with sulfuric acid in acetic anhydride over a broad range of temperatures (-30-60 ^{o}C) were unsuccessful. Nitronium reagents were also unsuccessful. **13** was then added to a mixture of **11** in nitrobenzene and refluxed at 140 ^{o}C open to the atmosphere overnight to provide NO_2-Im-Ip-OMe (**15**) cleanly. **15** was reduced using Pd/C in the presence of hydrogen to provide **17.** The amine (**17**) was

then Boc-protected using a mixture of Boc anhydride and DMAP in DMF to give **18**. It is noteworthy that the amine **17** is sufficiently unreactive and must be heated with excess $(Boc)_2O$ in the presence of DMAP for the reaction to procede. Following a single chromatography, saponification of **18** provided the final dimer Boc-Im-Ip-OH (**19**).

7.3 Oligomer Synthesis (Figure 7.4)

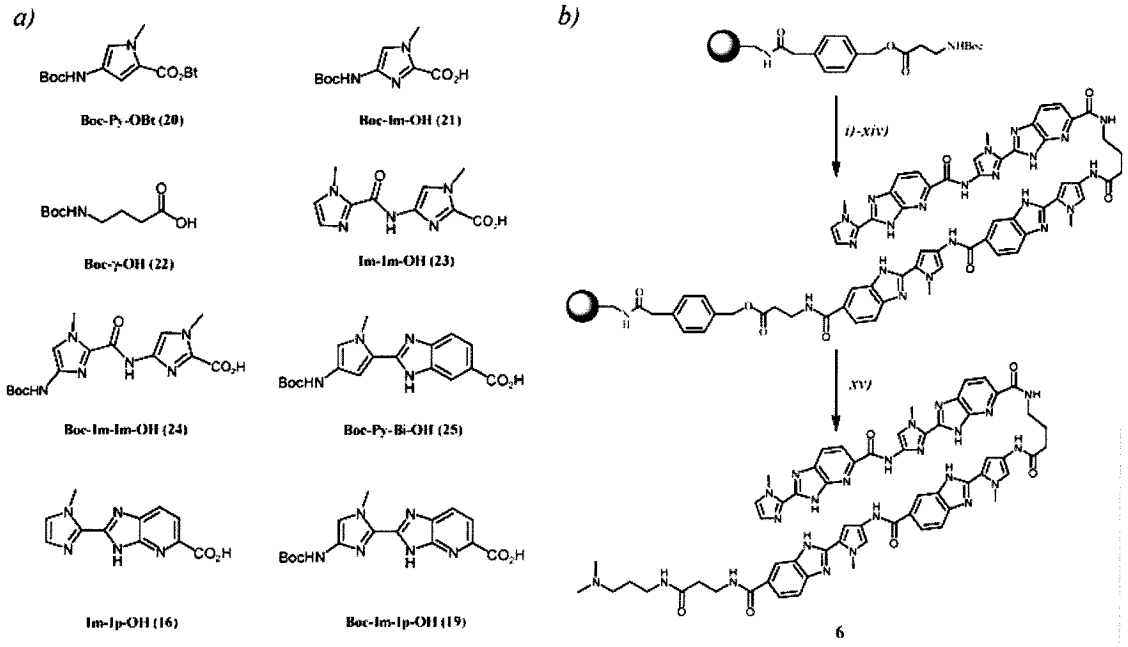


Figure 7.4 Solid phase synthesis of oligomers. a) heterocyclic and aliphatic monomers and dimers used to synthesize oligomers **1-6**. b) solid phase synthesis of oligomer **6**: i) 80% TFA in DCM; ii) Boc-Py-Bi-OH, HBTU, DIEA, DMF; iii) Ac_2O, DIEA, DMF; iv) 80% TFA in DCM; v) Boc-Py-Bi-OH, HBTU, DIEA, DMF; vi) Ac_2O, DIEA, DMF; vii) 80% TFA in DCM; viii) Boc-γ-OH, HBTU, DIEA, DMF; ix) Ac_2O, DIEA, DMF; x) 80% TFA in DCM; xi) Boc-Im-Ip-OH, HBTU, DIEA, DMF; xii) Ac_2O, DIEA, DMF; xiii) 80% TFA in DCM; xiv) Im-Ip-OH, HBTU, DIEA, DMF; xv) dimethylaminopropylamine (**Dp**) 2h 80 °C.

Oligomers **1-6** were synthesized using manual solid phase synthesis methodology on commercially available β-Ala-PAM resin as previously described.[14] Starting from base resin, monomeric and dimeric heterocyclic units were appended onto the resin in

stepwise fashion using HBTU activation. Couplings were allowed to proceed for several hours between 25-40 °C. Unreacted amines were acylated between coupling rounds using acetic anhydride. Deprotection of the Boc-protected amines was accomplished using 80% TFA in DCM. After completion of solid-phase synthesis, the resin was treated with dimethylaminopropylamine (Dp) and the oligomers were purified by reverse phase preparatory HPLC: Im-Im-Im-Im-γ-Py-Py-Py-Py-β-Dp (**1**), Im-Im-Im-Im-γ-PyBi-PyBi-β-Dp (**2**), ImIp-Im-Im-γ-PyBi-Py-Bi-β-Dp (**3**), Im-Im-ImIp-γ-PyBi-PyBi-β-Dp (**4**), Im-ImIp-Im-γ-PyBi-PyBi-β-Dp (**5**), ImIp-ImIp-γ-PyBi-PyBi-β-Dp (**6**). Oligomers were characterized using MALDI-TOF spectrometry.

7.4 DNA Affinity and Sequence Specificity. Quantitative DNase-I-footprinting titrations were carried out for oligomers **1-6** on the PCR product of plasmid pEF16 (**Figure 7.6**). Plasmid pEF16 was constructed containing two designed match sites (5'-**X**GGGGT-3' **X** = A, T) and two mismatch sites (5'-A**GGGA**T-3' and 5'-A**GAGG**T-3') (**Figure 7.5**). Careful attention was devoted to designing a plasmid which would probe

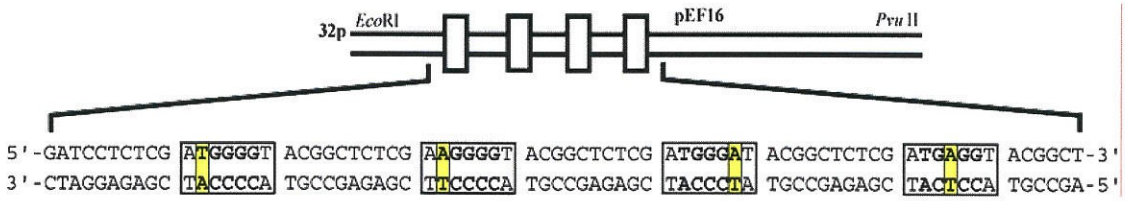


Figure 7.5 Sequence of plasmid pEF16 used for DNaseI footprinting titrations. The four designed binding sites are shown boxed with the variable base pair depicted in yellow.

not only the match and mismatch sequences of the designed oligomers, but would also look at the effect of flanking sequences. In particular, we were curious to see if an "AG" tract of DNA would present itself as a more problematic sequence to target when

compared to a "TA" tract. It has been shown in previous studies that DNA sequences containing 5'-AG-3' tracts are more difficult to target at high affinities.[15-17] The remaining two binding sites, 5'-A**GGGA**T-3' and 5'-A**GAGG**T-3' were designed to elucidate the specificity of the oligomers for single base-pair mismatch sites. In this case, mismatch evaluation involves determining the energetic penalty for the loss of a favorable hydrogen bond between the exocyclic amine of guanine and the lone pair nitrogen on the oligomer in question.

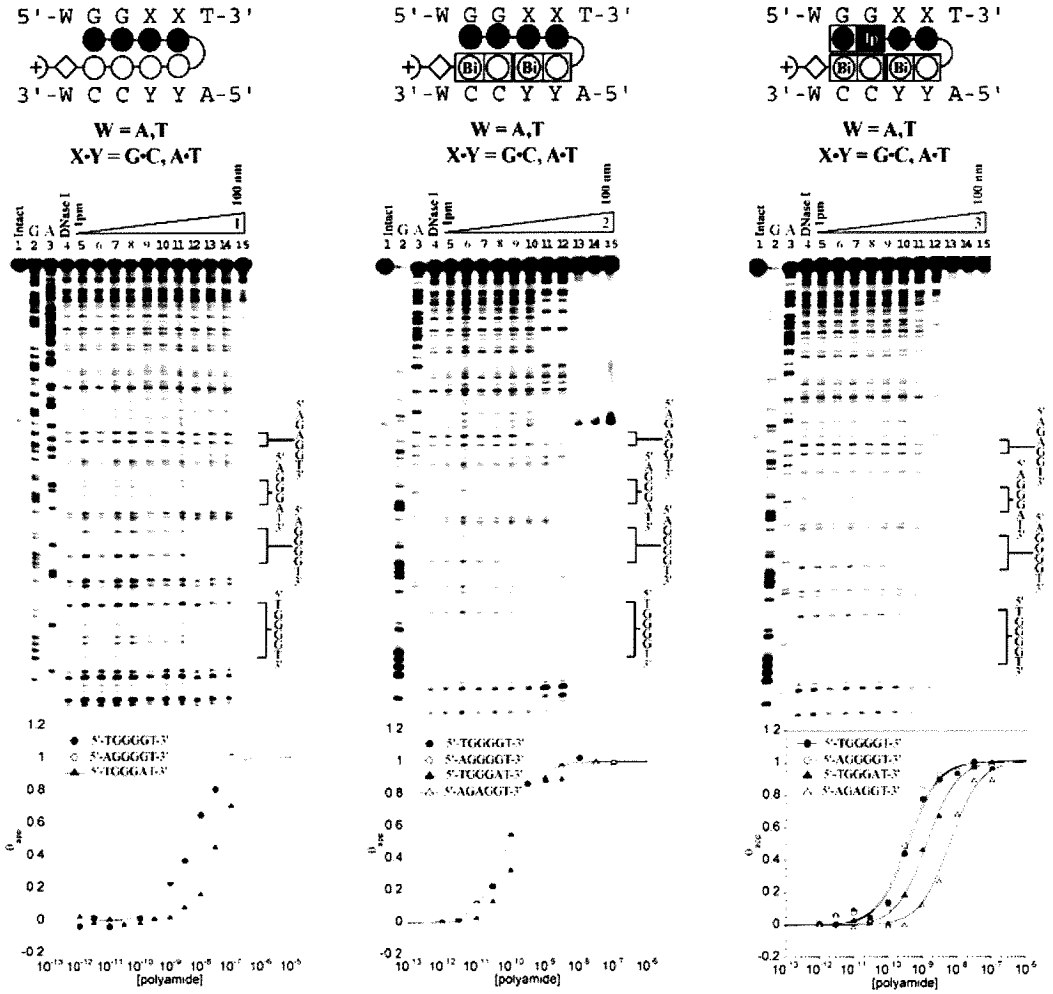


Figure 7.6 DNase I footprinting titrations for control compound **1** and oligomers **2** and **3**. (Top) Ball and stick structure of each oligomer. Imidazopyridine-imidazole (**Ip-Im**) and benzimidazole-pyrrole (**Bi-Py**) dimers are indicated by dark and light rectangles, respectively. β-alanine is depicted as a diamond and γ-aminobutyric acid is depicted as a half circle. (Middle) DNase I footprinting titration shown with lanes from left to right: Intact, G and A sequencing lanes, DNase I control, lanes 5-15 oligomer concentrations 1pM to 100 nM. (Bottom) Quantitative isotherms depicting the affinity and specificity of each oligomer. Isotherms were generated using previously published methods.[1]

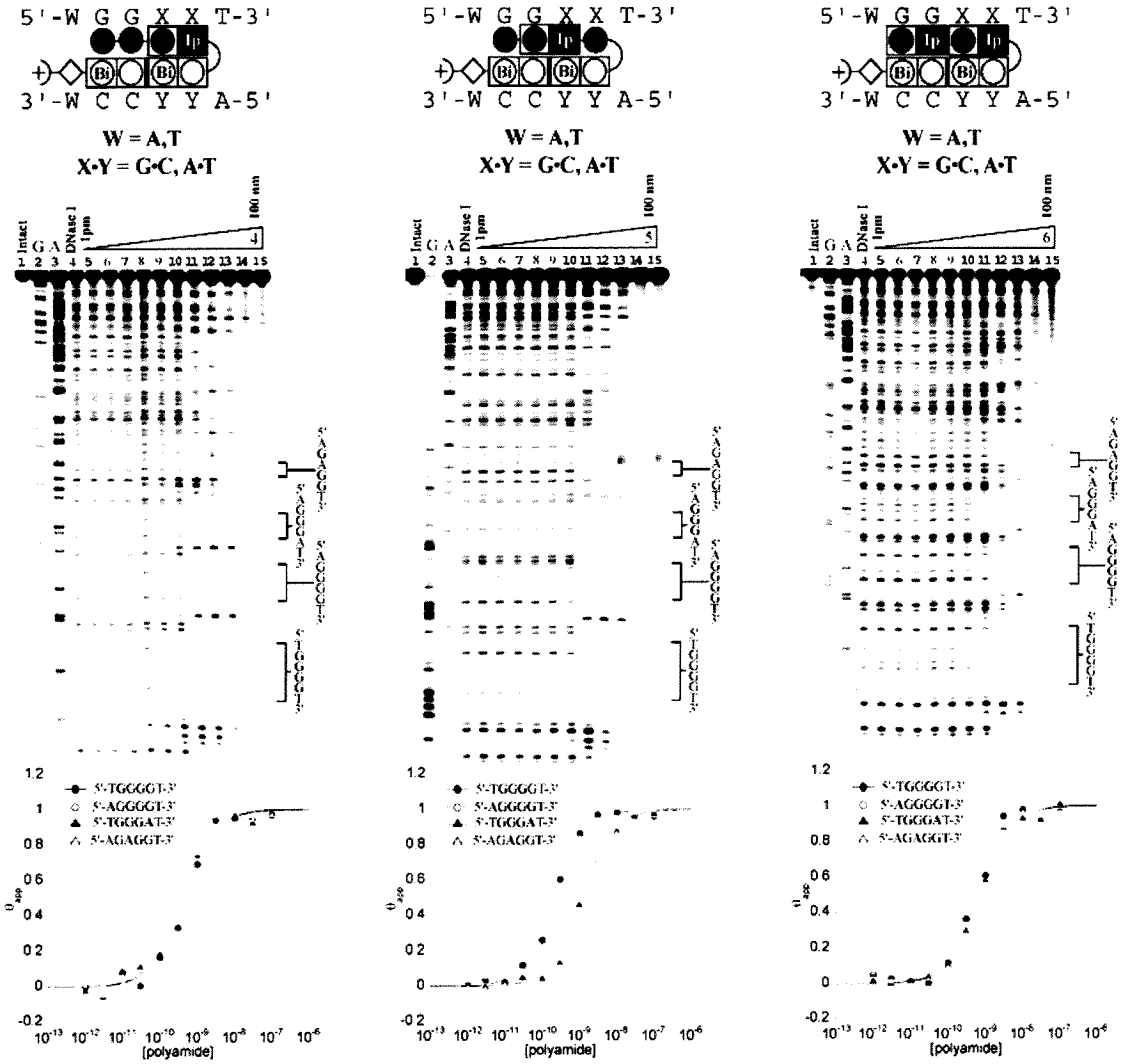
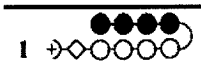


Figure 7.7 DNase I footprinting titrations for oligomers **4-6**. (Top) Ball and stick structure of each oligomer. Imidazopyridine-imidazole (**Ip-Im**) and benzimidazole-pyrrole (**Bi-Py**) dimers are indicated by dark and light rectangles, respectively. β-alanine is depicted as a diamond and γ-aminobutyric acid is depicted as a half circle. (Middle) DNase I footprinting titration shown with lanes from left to right: Intact, G and A sequencing lanes, DNaseI control, lanes 5-15 oligomer concentrations 1pM to 100 nM. (Bottom) Quantitative isotherms depicting the affinity and specificity of each oligomer. Isotherms were generated using previously published methods.[1]

Control compound **1** bound both 4-G match sequences (5'-**X**GGGGT-3', **X** = A,T) with comparably low affinity ($K_a \sim 10^8\ M^{-1}$), showing no bias for either site. **1** distinguished against mismatch sequences (5'-AGGGAT-3' and 5'-AGAGGT-3') with roughly 10-fold specificity (**Figure 7.6**). Oligomer **2** demonstrated a large increase in affinity ($K_a \sim 10^{10}\ M^{-1}$) as compared to **1** for both match sites but showed lower

specificity (4-fold) over the designed mismatch sites. Oligomers **3** and **4** showed a moderate increase in affinity ($K_a \sim 10^9\ M^{-1}$) but demonstrated only minor selectivity over mismatch sites. Oligomer **5** demonstrated high affinity for the 4-G match sequences ($K_a \sim$ mid $10^9\ M^{-1}$) with a 5-fold selectivity over mismatch sequences (5'-AGGGAT-3' and 5'-AGGAGT-3') (**Figure 7.7**). **6** bound all designed sequences with similar affinity. Thermodynamic data for oligomers **1-6** is summarized in **Table 7.1**.

Table 7.1. *Oligomers Targeting Guanine Rich Sequences* $K_a\ [M^{-1}]^{a,b}$

Oligomer	5'-ATGGGGT-3'	5'-AAGGGGT-3'	5'-ATGGGAT-3'	5'-ATGAGGT-3'
1 	$1.4(\pm 1.0)\times 10^8$	$2.6(\pm 1.1)\times 10^8$	$2.3(\pm 0.8)\times 10^7$	$< 1.0 \times 10^7$
2 	$1.9(\pm 1.4)\times 10^{10}$	$2.0(\pm 1.1)\times 10^{10}$	$4.8(\pm 1.1)\times 10^9$	$3.6(\pm 0.9)\times 10^9$
3 	$2.6(\pm 0.5)\times 10^9$	$2.9(\pm 0.2)\times 10^9$	$8.6(\pm 2.1)\times 10^8$	$2.5(\pm 0.8)\times 10^8$
4 	$2.6(\pm 0.6)\times 10^9$	$2.7(\pm 0.4)\times 10^9$	$2.6(\pm 0.2)\times 10^9$	$2.9(\pm 0.1)\times 10^9$
5 	$4.4(\pm 0.9)\times 10^9$	$4.1(\pm 1.2)\times 10^9$	$8.1(\pm 2.3)\times 10^8$	$8.2(\pm 1.3)\times 10^8$
6 	$1.6(\pm 0.2)\times 10^9$	$1.1(\pm 0.4)\times 10^9$	$1.3(\pm 0.1)\times 10^9$	$1.1(\pm 0.1)\times 10^9$

a) Values reported are the mean values from at least three DNase I footprinting titration experiments, with the standard deviation given in parentheses.
b) Assays were performed at 22 °C in a buffer of 10 mM Tris.HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0.

7.5 Molecular Modeling.

Modeling calculations were preformed with the *Spartan Essential* software package. *Ab initio* calculations were done using a Hartree-Fock model and a 6-31G* polarization basis set. Four-ring subunits containing the sequences Im-Im-Im-Im, **ImIp**-Im-Im, Im-Im-**ImIp**, Im-**ImIp**-Im and **ImIp-ImIp** were constructed to examine respective overall ligand geometry and relative location of the nitrogen lone pair presented to the DNA minor groove (**Figure 7.8**). Because of the importance of the hydrogen bonding interactions between the exocyclic amine of guanine and the G-

recognizing ring systems, we were very interested to see how adding 6-5 fused ring systems in various positions of the oligomer would affect the positioning of the hydrogen bond accepting nitrogen lone pair. To get a sense of the lone pair placement, the nitrogens on the oligomer were exaggerated as large spheres and overlaid to show a comparative oligomer analysis (**Figure 7.8b**).

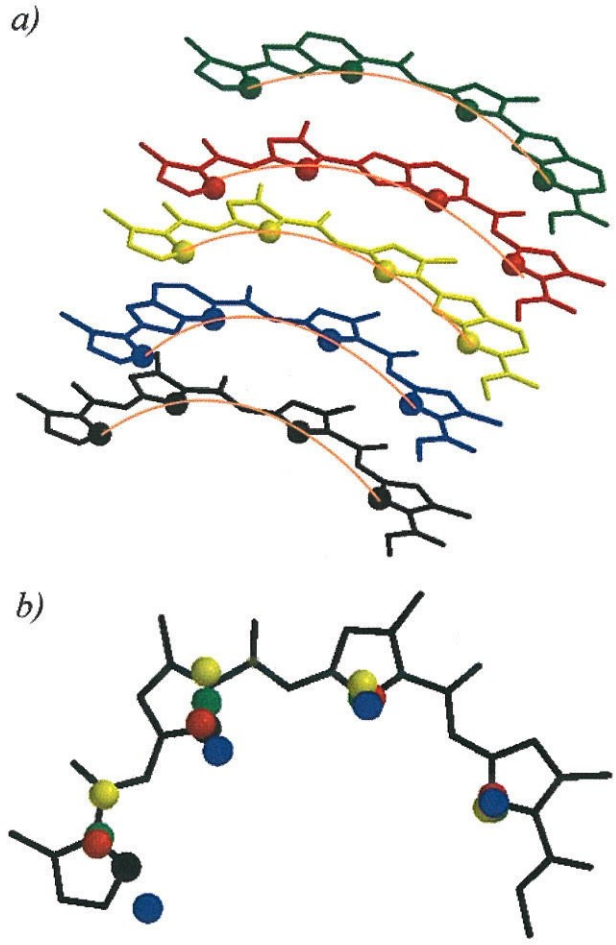


Figure 7.8 *Ab initio* modeling of oligomeric top strands. a) differential curvature of: Im-Im-Im-Im (black), Im-Im-**IpIm** (blue), **IpIm**-Im-Im (yellow), Im-**IpIm**-Im (red), and **IpIm-IpIm** (green). b) Overlay of oligomeric top strands with position of hydrogen bond accepting nitrogen shown. Color scheme is the same as described above.

7.6 Discussion

The weak binding affinity of the four imidazole compound **1**, prompted the search for oligomers that could target four consecutive guanines in the minor groove. By designing compounds with better shape complementarity with the uniquely shaped binding site, we have seen that it is possible to improve the average affinity of the ligands. Unfortunately what is gained by a boost in affinity is mitigated by the fact that the first round of compounds are not capable of targeting their match sites with appreciable levels of specificity.

Of the five oligomers we examined, oligomer **5** was the most successful at discriminating against its mismatch sites with a 5-fold level of specificity. This system places an **Ip-Im** recognition system in the middle of the top strand. It may be possible that the specificity of the compound stems from the fact that the placement of the 6-5 element generates two crescents which are closely related with regards to curvature. One can imagine that a single eight ring oligomer, comprised of two strands sitting adjacent to each other in the minor groove, would be better able to situate itself in its binding site if its components were similar in curvature. Binding data for oligomer **6**, a compound comprised completely of 6-5 ring systems and would have very similar curvature for both of its strands, seems to counter this hypothesis and it shows no preference for its match site whatsoever. With only five characterized oligomers, however, it is difficult to draw conclusions from the limited data set.

One issue encountered by oligomers consisting of less curved recognition elements is that the decrease in curvature is met with a less intimate level of contact

between the ligand and the minor groove floor. The level of contact becomes especially important when dealing with systems designed to target poly guanine stretches of DNA. Here, binding site recognition is based completely upon the ability of the compound to hydrogen bond with the exocyclic amine of guanine. Ring systems designed to recognize guanine exhibit a higher binding affinity for their match sites because of their ability to form an energetically favorable hydrogen bond. That is, systems recognizing guanine exhibit a boost in affinity when binding their match sites as opposed to systems like **Py** which exhibit a decrease in binding affinity thanks to steric clashes with the atoms in the minor groove. Knowing this, one can imagine that ligands demonstrating a decreased level of curvature may have a diminished capacity to hydrogen bond. This phenomenon may explain why the oligomers in this study demonstrated a mild, at best, preference for their match sites.

One striking characteristic of oligomers **2-6**, however, is their marked increase in binding affinity. This phenomenon may be attributed to the fact that the 6-5 fused benzimidazole analogues have a larger hydrophobic surface. Such a surface could explain their preference for binding in the minor groove where they are removed from the surrounding aqueous solution.[18] In addition, the benzimidazole derivatives are a more rigid structure with a lower degree of rotational freedom. Such pre-organization may decreases the entropic cost of DNA complexation. Finally, the overall geometry of oligomers incorporated with benzimidazole derivatives is less curved, possibly allowing a more complementary fit with the DNA helix.

7.7 Conclusion

The current study has laid the groundwork for future investigations. The initial round of oligomers has demonstrated that it is possible to design compounds that bind traditionally difficult sequences to target such as 5`-GGGG-3`. This new development can be attributed to a better ligand/DNA fit with regards to shape, an increase in the degree of ligand pre-organization and the large hydrophobic surface associated with the benzimidazole-based recognition elements.

Improving the specificity of oligomers designed to target guanine rich sequences is currently the biggest challenge that faces these types of ligands. Second generation compounds may look to varying the positioning of the benzimidazole rings in the bottom strand of the oligomers. Replacing all four pyrroles on that strand may not be the most effective methodology and care should be given to finding the appropriate balance between traditional 5-membered ring systems and the new 6-5 fused recognition elements.

7.8 Experimental

General. N,N-dimethylformamide (DMF), N,N-diisopropylethylamine (DIEA), thiophenol (PhSH), N,N-dimethylaminopropylamine (Dp), triethylamine (TEA), nitrobenzene (NO_2Ph), 2-formyl-N-methylimidazole, red fuming nitric acid, 1,3-dichlo-4-nitroropyridine, 30% bromine in acetic acid, palladium acetate ($Pd(OAc)_2$), and 10% palladium on carbon, were purchased from Aldrich. Boc-β-alanine-(4-carbonylaminomethyl)-benzyl-ester-copoly(styrene-divinylbenzene)resin (Boc-β-Pam-

resin), dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBt), 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), N,N-dimethylaminopyridine (DMAP), and Boc-β-alanine were purchased from NOVA Biochem. Trifluoroacetic acid (TFA) was purchased from Halocarbon. All other solvents were reagent-grade from EM. Oligonucleotide inserts were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology. Precoated silica gel plates $60F_{254}$ for TLC and silica gel 60 (40 μm) for flash chromatography were from Merck. Glycogen (20 mg/mL), dNTPs (PCR nucleotide mix), and all enzymes, unless otherwise stated, were purchased from Boehringer-Mannheim. pUC19 was purchased from New England Biolabs, and deoxyadenosine [γ-^{32}P]triphosphate was provided by ICN. Calf thymus DNA (sonicated, deproteinized) was from Amersham Pharmacia. DNaseI (7500 units/mL, FPLC pure) was purchased from Roche. AmpliTaq DNA polymerase was from Perkin-Elmer and used with the provided buffers. Tris.HCl, DTT, RNase-free water, and 0.5 M EDTA were from United States Biochemical. Calcium chloride, potassium chloride, and magnesium chloride were purchased from Fluka. Tris-borate-EDTA was from GIBCO and bromophenol blue was from Acros. All reagents were used without further purification.

NMR spectra were recorded on a Varian spectrometer at 300 MHz in DMSO-*d6* or $CDCl_3$ with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. High resolution FAB and EI mass spectra were recorded at the Mass Spectroscopy Laboratory at the California Institute of Technology. Matrix-assisted, laser

desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) was conducted at the Mass Spectroscopy Laboratory at the California Institute of Technology.

Heterocycle Synthesis. Heterocyclic monomers Boc-Py-OBt (**20**), Boc-Im-OH (**21**) and dimer Boc-Py-Bi-OH (**25**) were synthesized as reported.[14] Im-Im-OH (**21**) and Boc-Im-Im-OH (**24**), dimers were prepared using solution-phase dimerization methodology previously described.[2]

1-methyl-4-nitro-1H-imidazole-2-carbaldehyde (NO_2-Im-CHO) (**13**). A cooled flask (0°C) of 1-methyl-2-imidazole-carboxaldehyde (**12**) (8g, 72.6 mmol, Aldrich) was treated dropwise with a precooled (0 °C) solution of red fuming nitric acid (75 ml) in conc. $H_2SO_4 \bullet SO_3$ (30%) (75 ml). The mixture was warmed to room temperature and stirred for 12 hours open to the atmosphere. Next, the mixture was poured over ice, neutralized with solid sodium carbonate, extracted four times with dichloromethane, dried over anhydrous sodium sulfate, and concentrated in vacuo to give a brownish-yellow oil. The oil was recrystallized from iPrOH/Et_2O or EtOH/Et_2O to give 1-methyl-4-nitro-1H-imidazole-2-carbaldehyde **(13)** as a tan crystalline solid (4.5 g, 40% Yield). TLC (1:1 EtOAc/Hex) R_f 0.4; ^1H NMR (300 MHz, DMSO-d_6) δ 9.74 (s, 1H), 8.71 (s, 1H), 3.99 (s, 3H); ^{13}C (75 MHz, DMSO-d_6) δ 182.31, 146.14, 140.56, 127.33, 35.70; HR-MS (EI+): calcd. for $C_5H_5O_3N_3$: 155.033; found: 155.035.

2-(1-Methyl-4-nitroimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester (NO_2-Im-Ip-OMe) **(15).** 1-Methyl-4-nitro-1H-imidazole-2-carbaldehyde (2.01 g, 13.0 mmol) **(13)** and methyl 5,6-diaminopyridine-2-carboxylate **(11)** (2.17 g, 13.0 mmol) suspended in 120 ml of nitrobenzene was heated to 140 °C for 48 hours open

to the atmosphere. The reaction mixture was cooled to 23 °C and the precipitate collected by vacuum filtration. The solid was washed with diethyl ether and dried under high vacuum to provide 2-(1-methyl-4-nitroimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester **(15)** (3.6 g, 92% Yield) as a powdery tan solid. ^{1}H NMR (300 MHz, DMSO-d_6) δ 14.02 (broad s, 1H), 8.03-8.06 (m, 2H), 4.27 (s, 3H), 3.91 (s, 3H); ^{13}C (75 MHz, DMSO-d_6) δ 166.01, 146.44, 142.95, 136.59, 135.80, 130.36, 127.35, 123.81, 121.17, 121.04, 52.85, 37.43; HR-MS (EI+): calcd. for $C_{12}H_{10}N_6O_4$: 302.076; found: 302.076.

2-{4-amino-1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester (H_2N-Im-Ip-OMe) **(17).** 2-(1-methyl-4-nitroimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester **(15)** (3 g, 9.9 mmol) was dissolved in anhydrous DMF (150 ml) and the solution was degassed with Ar. After the addition of Pd/C (10 wt. %, 600 mg) the reaction mixture was purged three times with hydrogen and then left to stir at 23°C for 9 hours under a hydrogen balloon atmosphere. After filtering through a pad of Celite and washing with copious amounts EtOAc the filtrate was concentrated in vacuo to give 2-{4-amino-1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester **(17)** without further purification (2.7 g, 100% Yield). ^{1}H NMR (300 MHz, DMSO-d_6) δ ; ^{13}C (75 MHz, DMSO-d_6) δ 166.17, 149.03, 148.21, 141.40, 131.50, 120.12, 107.77; HR-MS (EI+): calcd. for $C_{12}H_{12}N_6O_2$: 272.102; found: 272.103.

2-{4-[(tert-Butoxy)carbonylamino]-1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester (Boc-Im-Ip-OMe) **(18).** 2-{4-amino-1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester (**17**) (2.0

g, 7.3 mmol) dissolved in DMF (25 ml) was treated with Boc2O (5.3 g, 24.3 mmol), DIEA (5.2 ml), and DMAP (95 mg, 0.73 mmol). The reaction mixture was then heated to 80 °C for 72 hours, cooled to 23 °C, and flashed through a plug of silica gel eluting with EtOAc to give a mixture of mono- and di-bocked (2-{4-[(tert-Butoxy)carbonylamino]-1-methylimidazol-2-yl)-3-[(tert-Butoxy)carbonylamino]-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester) products which were carried on for saponification.

2-{4-[(tert-Butoxy)carbonylamino]-1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic (Boc-Im-Ip-OH) **(19).** 2-{4-[(tert-Butoxy)carbonylamino]-1-methylimidazol-2-yl)-3-[(tert-Butoxy)carbonylamino]-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester dissolved in MeOH (10 ml) and NaOH (1 N, 25 ml) was heated to 50 °C for 4 hours. The reaction mixture was cooled to 0 °C and the pH adjusted slowly to pH = 4 with 1 N HCl. The reaction mixture was then extracted with ethyl acetate (four times), dried over anhydrous sodium sulfate, concentrated in vacuo, and dried under high vacuum to give 2-{4-[(tert-Butoxy)carbonylamino]-1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic **(19)** (258 mg, 60% Yield) as a light yellow solid. ^{1}H NMR (300 MHz, DMSO-d_6) δ 13.19 (broad s, 2H), 9.53 (s, 1H), 8.00 (m, 2H), 7.36 (s, 1H), 4.15 (s, 3H), 1.47 (s, 9H); ^{13}C (75 MHz, DMSO-d_6) δ 166.61, 152.92, 147.81, 142.48, 138.38, 132.01, 128.91, 119.74, 113.50, 79.07, 35.35, 28.12; HR-MS (EI+): calcd. for $C_{16}H_{18}N_6O_4$: 358.139; found: 358.137.

2-(1-Methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester (Im-Ip-OMe **(14).** 1-Methylimidazole-2-carbaldehyde (**12**) (214 mg, 1.85 mmol) and methyl 5,6-diaminopyridine-2-carboxylate **(11)** (310 mg, 1.85 mmol) suspended in

17 ml of nitrobenzene was heated to 140 °C for 24 hours open to the atmosphere. The reaction mixture was cooled to room temperature. To the reaction was added diethyl ether (1 mL), upon which a tan precipitate formed. The precipitate was collected by filtration and washed with cold diethyl ether. The material was dissolved in hot iso propanol, cooled to room temperature and re-precipitated with diethyl ether. The solid was washed with diethyl ether and dried under high vacuum to provide 2-(1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester **(14)** (250 mg, 52% Yield) as s tan solid. ^{1}H NMR (300 MHz, DMSO-d_6) δ 7.99 (s, 2 H), 7.51 (d, 1H, J = 0.9 Hz), 7.20 (s, 1H, J = 0.9 Hz), 4.18 (s, 3H), 3.88 (s, 3H); ^{13}C (75 MHz, DMSO-d_6) δ 147.7, 142.0, 137.1, 131.8, 129.6, 126.9, 119.9, 109.3, 52.8, 35.8; HR-MS (EI+): calcd. for $C_{12}H_{11}N_5O_2$: 257.091; found: 257.092.

2-(1-Methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid (Im-Ip-OH). **(16).** 2-(1-methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid methyl ester **(14)** (250 mg, 0.97 mmol) dissolved in MeOH (2 ml) and KOH (4 N, 3 ml) was heated to 50 °C for 4 hours. TLC showed baseline product formation with some residual non-polar impurities. The methanol was removed in vacuo and the aqueous layer washed with EtOAc (2 x 10 mL) to remove any starting material and trace impurities. The pH of the aqueous layer was then adjusted slowly to pH = 4 with 1 N HCl upon which time a cloudy beige precipitate formed. The mixture was placed in a falcon tube and the precipitate concentrated by centrifugation. The supernatant was decanted and the solid dried under high vacuum to give 2-(1-Methylimidazol-2-yl)-3H-imidazo[4,5-b]pyridine-5-carboxylic acid **(16)** (154 mg, 65% Yield) as a brown solid. ^{1}H

NMR (300 MHz, DMSO-d_6) δ 7.99 (s, 2H), 7.51 (d, 1H, J = 0.9 Hz), 7.20 (d, 1H, J = 0.9 Hz), 4.18 (s, 3H); HR-MS (EI+): calcd. for $C_{11}H_9N_5O_2$: 243.075; found: 243.074.

Oligomer Synthesis. Oligomers were synthesized on solid support using Boc-β-PAM resin (0.59 meq/g). Stepwise elongation of the oligomers was done according to previously published protocols.[14] The synthesis of compound **1** has been previously reported.[19]

Preparation of Base Resin **R**-β-BiPy-BiPy-γ-NHBoc (**BR1**): To a manual solid phase synthesis vessel was added Boc-β-PAM resin (0.3 g). The resin was washed with DMF (15 mL) and allowed to swell for 15 minutes while shaking at room temperature. The resin was then washed with DCM (~30 mL), followed by 80% TFA in DCM (~30 mL) to deblock the Boc-group. The resin was then agitated at room temperature in 80% TFA/DCM for another 25 minutes to provide the deblocked resin bound amine (**R**–β-NH_2). Following Boc-deprotection, the resin was washed with DCM and 10% DIEA in DMF to neutralize and prepare for coupling. Simultaneously, in a separate reaction vessel, Boc-PyBi-OH (**23**) (189 mg, 531 μM), HBTU (191 mg, 504 μM), DIEA (137 mg, 185 μL, 1.06 mM) and DMF (1.2 mL) was mixed and allowed to activate at room temperature for 25 minutes. This mixture was then added to the solid-phase synthesis vessel containing **R**–β-NH_2. Coupling was allowed to proceed at room temperature with agitation for 3-6 hours. Initial loading of the resin requires elongated coupling times. Following coupling, the resin was acylated by the addition of acetic anhydride to the mixture and shaking for 15 minutes. The addition of the next Boc-PyBi-OH (**25**) dimer was incorporated and deprotected as described above to provide the resin bound fragment

(**R**-β-BiPy-Bi-Py-NH_2). To this fragment was added a preactivated mixture of Boc-γ-OH (180 mg, 885 μM), HBTU (319 mg, 841 μM), DIEA (229 mg, 308 μL, 1.77 mM). Coupling was allowed to proceed for 3 h at room temperature with agitation. The resin was then caped with acetic anhydride as described above to provide the base resin **R**-β-BiPy-Bi-Py-γ-NHBoc (**BR1**). **BR1** was then washed with DCM followed by MeOH and Et_2O. The resin was then dried under high vacuum and stored for subsequent use.

Im-Im-Im-Im-γ-PyBi-Py-Bi-β-Dp (**2**): **BR1** (50 mg) was added to a manual solid-phase synthesis vessel. The resin was washed with DCM (~15 mL), followed by deprotection with 80% TFA in DCM. The resin was shaken at room temperature in the 80% TFA solution for 25 minutes. The resin was then drained, washed with DCM, and neutralized with 10% DIEA in DMF. A pre-activated mixture of Boc-Im-Im-OH (**24**) (54 mg, 148 μM), HBTU (53 mg, 140 μM), DIEA (38 mg, 52 μL, 295 μM) and DMF (400 μL) was then added to the reaction vessel and coupling was allowed to proceed for three hours at room temperature, followed by capping with acetic anhydride as described for **BR1** to give **R**-β-BiPy-BiPy-γ-Im-Im-NHBoc. Following resin deprotection as described above, Im-Im-OH (**23**) was activated as described for Boc-Im-Im-OH (**24**). Coupling of **23** to the resin was allowed to proceed overnight at room temperature to provide **R**-β-BiPy-BiPy-γ-Im-Im-Im-Im. The resin was treated with the cleavage protocol outlined below to provide *Im-Im-Im-Im-γ-PyBi-Py-Bi-β-Dp* **2** in 5% Yield: [MALDI-TOF] $C_{58}H_{66}N_{23}O_8$ 1212.55 calcd. 1212.5 found. $[M+H]^+$

ImIp-Im-Im-γ-PyBi-Py-Bi-β-Dp (**3**): **BR1** (50 mg) was added to a manual solid phase synthesis vessel and **R**-β-BiPy-BiPy-γ-Im-Im-NHBoc was prepared as described

above for **2**. Following deprotection, washing and neutralization as described above, a pre activated mixture of ImIp-OH (**14**) (21.5 mg, 88.5 μM), HBTU (32 mg, 84 μM), DIEA (23 mg, 31μL, 177 μM), DMF (400 μL) was added to the vessel containing **R**-β-BiPy-BiPy-γ-Im-Im-NH_2. Coupling was allowed to proceed overnight at room temperature to provide **R**-β-BiPy-BiPy-γ-Im-Im-IpIm. The resin was treated with the cleavage protocol outlined below to provide *ImIp-Im-Im-γ-PyBi-Py-Bi-β-Dp* **3** in 2.2% Yield: [MALDI-TOF] $C_{59}H_{64}N_{23}O_7$ 1206.54 calcd. 1206.60 found. $[M+H]^+$

Im-Im-ImIp-γ-PyBi-PyBi-β-Dp (**4**): **BR1** (50 mg, 0.81 meq/g) was added to a manual solid phase synthesis vessel. The resin was treated with the cleavage protocol outlined below to provide *Im-Im-ImIp-γ-PyBi-Py-Bi-β-Dp* **4** in 3% Yield: MALDI-TOF-MS $C_{59}H_{64}N_{23}O_7$ 1206.5 calcd. 1206.5 found. $[M+H]^+$

Im-ImIp-Im-γ-PyBi-PyBi-β-Dp (**5**): **BR1** (70 mg, 0.59 meq/g) was added to a manual solid phase synthesis vessel. The resin was washed with DCM (~15 mL), followed by deprotection with 80% TFA in DCM. The resin was shaken at room temperature in the 80% TFA solution for 25 minutes. The resin was then drained, washed with DCM, neutralized with 50% DIEA in DCM, and washed with DMF. A pre-activated mixture of Boc-Im-OH (**21**) (50 mg, 207 μmol), HBTU (79 mg, 208 μmol), DIEA (53 mg, 72 μL, 413 μmol) and DMF (900 μL) was then added to the reaction vessel and coupling was allowed to proceed for 12 hours at room temperature, followed by capping with acetic anhydride as described for **BR1** to give **R**-β-BiPy-BiPy-γ-Im-NHBoc. Following resin deprotection as described above, Boc-ImIp-OH (**19**) was activated as described for Boc-Im-OH (**21**). Coupling of **19** to the resin was allowed to

proceed overnight at room temperature to provide **R**-β-BiPy-BiPy-γ-Im-IpIm-NHBoc. Following resin deprotection as described above, Im-CCl_3 (*2-Trichloroacetyl-1-methylpyrrole*) (47 mg, 207 μmol) and DIEA (53 mg, 72 μL, 413 μmol) were dissolved in NMP (900 μL) and added to the reaction vessel. Coupling of Im-CCl_3 to the resin was allowed to proceed overnight at 32 °C to provide **R**-β-BiPy-BiPy-γ-Im-IpIm-Im. The resin was treated with the cleavage protocol outlined below to provide *Im-ImIp-Im-γ-PyBi-Py-Bi-β-Dp* **5** in 3% Yield: MALDI-TOF-MS $C_{59}H_{64}N_{23}O_7$ 1206.54 calcd. 1206.5 found. $[M+H]^+$

ImIp-ImIp-γ-PyBi-PyBi-β-Dp (**6**): **BR1** (50 mg) was added to a manual solid phase synthesis vessel. The resin was treated with the cleavage protocol outlined below to provide *ImIp-ImIp-γ-PyBi-Py-Bi-β-Dp* **6** in 2% Yield: MALDI-TOF-MS $C_{60}H_{62}N_{23}O_6$ 1200.52 calcd. 1200.5 found. $[M+H]^+$

Resin Cleavage Procedure: A sample of resin (20-100 mg) was washed with DCM followed by the addition of dimethylaminopropylamine (Dp) (1 mL). The mixture was heated to 80 °C for 2 h with occasional agitation. The resin was then filtered and washed with 0.1% TFA in water (7 mL). The combined filtrate was collected and subjected to purification by reverse phase preparatory HPLC using a Waters C18 column and 0.1% TFA/ACN solvent system. Appropriate fractions from the HPLC purification were checked for purity by analytical HPLC and characterized by MALDI-TOF spectroscopy. Pure fractions were then pooled, flash frozen using liquid nitrogen and lyophilized to a dry solid for later use.

Footprinting Experiments. Plasmids pEF16 was constructed using standard methods.[19] DNase I footprint titrations were performed according to standard protocols.[1]

7.9 References

[1] Trauger, J. W.; Dervan, P. B., Footprinting methods for analysis of pyrrole-imidazole polyamide/DNA complexes. In *Drug-Nucleic Acid Interactions*, ed.; 'Ed.'^'Eds.' 2001; 'Vol.' 340, p^pp 450-466.

[2] Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B., Toward an understanding of the chemical etiology for DNA minor-groove recognition by polyamides. *Helvetica Chimica Acta.* **2002**, 85, (12), 4485-4517.

[3] Renneberg, D.; Dervan, P. B., Imidazopyridine/pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove recognition. *Journal of the American Chemical Society.* **2003**, 125, (19), 5707-5716.

[4] Mutterer, F.; Weis, C. D., Halogenated Pyridines .5. Fluorinated and Brominated Pyridine Compounds. *Helvetica Chimica Acta.* **1976**, 59, (1), 229-235.

[5] Magerlein, W.; Indolese, A. F.; Beller, M., A more efficient catalyst for the carbonylation of chloroarenes. *Angewandte Chemie-International Edition.* **2001**, 40, (15), 2856-2859.

[6] Schoenbe.A; Bartolet.I; Heck, R. F., Palladium-Catalyzed Carboalkoxylation of Aryl, Benzyl, and Vinylic Halides. *Journal of Organic Chemistry.* **1974**, 39, (23), 3318-3326.

[7] Briehn, C. A.; Weyermann, P.; Dervan, P. B., Alternative heterocycles for DNA recognition: The benzimidazole/imidazole pair. *Chemistry-a European Journal.* **2003**, 9, (9), 2110-2122.

[8] Marques, M. A.; Doss, R. M.; Foister, S.; Dervan, P. B., Expanding the repertoire of heterocycle ring pairs for programmable minor groove DNA recognition. *Journal of the American Chemical Society.* **2004**, 126, (33), 10339-10349.

[9] Lombardy, R. L.; Tanious, F. A.; Ramachandran, K.; Tidwell, R. R.; Wilson, W. D., Synthesis and DNA interactions of benzimidazole dications which have activity against opportunistic infections. *Journal of Medicinal Chemistry.* **1996**, 39, (7), 1452-1462.

[10] Ji, Y. H.; Bur, D.; Hasler, W.; Schmitt, V. R.; Dorn, A.; Bailly, C.; Waring, M. J.; Hochstrasser, R.; Leupin, W., Tris-benzimidazole derivatives: Design, synthesis and DNA sequence recognition. *Bioorganic & Medicinal Chemistry.* **2001**, 9, (11), 2905-2919.

[11] Behrens, C.; Harrit, N.; Nielsen, P. E., Synthesis of a hoechst 32258 analogue amino acid building block for direct incorporation of a fluorescent, high-affinity DNA binding motif into peptides. *Bioconjugate Chemistry.* **2001**, 12, (6), 1021-1027.

[12] Yadagiri, B.; Lown, J. W., Convenient Routes to Substituted Benzimidazoles and Imidazolo 4,5-B Pyridines Using Nitrobenzene as Oxidant. *Synthetic Communications.* **1990**, 20, (7), 955-963.

[13] Austin, M. W.; Blackbor.Jr; Ridd, J. H.; Smith, B. V., Kinetics and Mechanism of Heteroaromatic Nitration .2. Pyrazole and Imidazole. *Journal of the Chemical Society.* **1965**, (FEB), 1051-&.

[14] Baird, E. E.; Dervan, P. B., Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *Journal of the American Chemical Society.* **1996**, 118, (26), 6141-6146.

[15] Qu, X. G.; Ren, J. S.; Riccelli, P. V.; Benight, A. S.; Chaires, J. B., Enthalpy/entropy compensation: Influence of DNA flanking sequence on the binding of 7-amino actinomycin D to its primary binding site in short DNA duplexes. *Biochemistry*. **2003**, 42, (41), 11960-11967.

[16] Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B., Structure of a β-alanine-linked polyamide bound to a full helical turn of purine tract DNA in the 1:1 motif. *Journal of Molecular Biology*. **2002**, 320, (1), 55-71.

[17] Melander, C.; Herman, D. M.; Dervan, P. B., Discrimination of A/T sequences in the minor groove of DNA within a cyclic polyamide motif. *Chemistry-a European Journal*. **2000**, 6, (24), 4487-4497.

[18] Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chaires, J. B., Specific binding of Hoechst 33258 to the d(CGCAAATTTGCG)(2) duplex: Calorimetric and spectroscopic studies. *Journal of Molecular Biology*. **1997**, 271, (2), 244-257.

[19] Swalley, S. E.; Baird, E. E.; Dervan, P. B., Discrimination of 5'-GGGG-3', 5'-GCGC-3', and 5'-GGCC-3' sequences in the minor groove of DNA by eight-ring hairpin polyamides. *Journal of the American Chemical Society*. **1997**, 119, (30), 6953-6961.

Chapter 8

Enforcing the Hairpin Motif in Polyamide-DNA Recognition

The text of this chapter was taken in part from a manuscript coauthored with Michael A. Marques, Adam R. Urbach and Professor Peter B. Dervan (Caltech)

(Urbach, A. R.; Marques, M. A. and Dervan, P. B. "*Enforcing the Hairpin Motif in Polyamide-DNA Recognition,*" *in preparation.*)

Abstract.

Polyamides containing imidazole (Im) and pyrrole (Py) amino acids linked by beta-alanine (β) and gamma-aminobutyric acid (γ) target predetermined DNA sequences in hairpin or extended 1:1 conformations depending on the rules for recognition inherent to each binding mode. In order to eliminate this ambiguity of sequence targeting based on conformation, we present design principles for predicting the preferred binding mode based on the sequence composition of the polyamide. To this end, a series of aliphatic linkages were employed to confer a preference for hairpin or extended conformation and the relative preference for each binding mode determined. DNase I and MPE footprinting experiments were performed for polyamides Im-β-ImPy-**β**-Im-β-ImPy-β-Dp (**1**), Im-β-ImPy-**γ**-Im-β-ImPy-β-Dp (**2**), Im-β-ImPy-^{H_2N}**γ**-Im-β-ImPy-β-Dp (**3**), and Im-β-ImPy-^{Ac}**γ**-Im-β-ImPy-β-Dp (**4**) (Dp = dimethylaminopropylamide) in complex with their target hairpin and extended binding sites, 5'-TAGCGCT-3' and 5'-AAAGAGAAGAG-3', respectively. In addition, the preferred binding modes of the eight-ring polyamides ImPyPyPy-**β**-ImPyPyPy-β-Dp (**5**) and ImPyPyPy-**γ**-ImPyPyPy-β-Dp (**6**) in complex with their target hairpin and 1:1 binding sites, 5'-TAGTACT-3' and 5'-AAAAAGAAAAG-3', respectively were assessed. We find for the β-rich series, **1** – **4**, that the γ-linked polyamide is least specific. However, by functionalizing the γ residue at the prochiral α-(R)-proton with amino or acetamide moiety, the hairpin conformation is favored over 1:1 by 150- or 560-fold, respectively. Alternatively, a β linker can be employed to favor the 1:1 mode by >150-fold. Therefore we can controllably reverse the DNA sequence specificity from extended to hairpin conformation by 82,000-fold or ~7 kcal/mol. The γ-linked eight-ring hairpin polyamide exhibits no 1:1 binding and, in fact, prefers binding

as a hairpin to a single base-pair mismatch site rather than opening up into extended conformation. These results demonstrate that the hairpin and extended modes of polyamide-DNA binding, which are dependent on ligand conformation, can be effectively controlled by incorporating the appropriate linkage between polyamide subunits. These design principles should greatly improve the overall sequence fidelity of the next generation of polyamides for DNA recognition in larger, genomic contexts.

8.1 Introduction.

Polyamides composed of pyrrole (Py), imidazole (Im), hydroxypyrrole (Hp), and beta alanine (β) amino acids target predetermined DNA sequences in the minor groove with affinities and specificities that rival native transcription factors, offering a potentially powerful weapon against human disease.[1] There exist 1:1 and 2:1 ligand-DNA stoichiometries with quite different rules for recognition.[1-4] The 2:1 complex, in which a pair of polyamide subunits is arranged in antiparallel fashion, allows for discrimination between the four Watson-Crick base pairs.[5,6] The covalent head-to-tail linkage of polyamide subunits in a 2:1 complex results in a hairpin structure with substantially enhanced DNA affinity and sequence specificity over the analogous, unlinked polyamide subunits.[7,8] To date, hairpin polyamides are the most sequence-selective class of minor groove-binding small molecules, due in part to the high density array of functional groups presented by the closely packed pairs of polyamide subunits.[2]

Polyamides composed of multiple contiguous heterocyclic residues are overcurved with respect to the DNA helix.[9] However, the flexible β-alanine residue (β) can be incorporated to relax ligand curvature and, in turn, restore ligand-DNA alignment register and binding affinity.[10,11] The β residue has proven useful in the extended 1:1 motif for recognizing purine-rich DNA sequences using polyamides of type -Im-β-ImPy-β-.[12] Analysis of this motif revealed a different set of rules for DNA recognition.[13] In addition, it was found that a polyamide may target different DNA sequences depending on its mode of binding, i.e., 1:1 or 2:1.[14] In an effort to maximize sequence fidelity in larger, genomic contexts, it will be necessary to efficiently control the ligand's choice of binding modes (**Figure 1**).

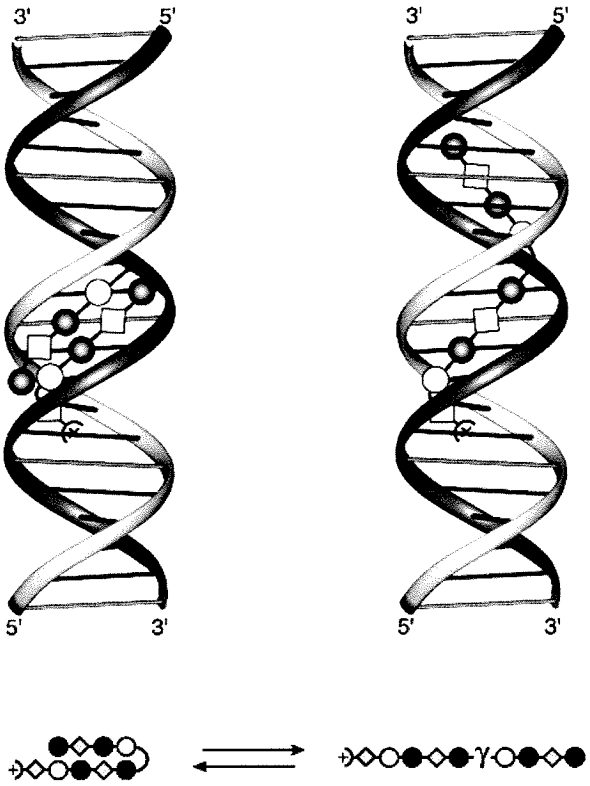


Figure 8.1 Illustration of the equilibrium between hairpin (left) and extended (right) conformational binding modes. The cartoons at top represent DNA with polyamide bound in the respective conformation. DNA is shown as a ladder. Polyamides are shown as dot models, with shaded and non-shaded circles representing imidazole and pyrrole, respectively, and gray diamonds indicating beta-alanine. The gamma turn residue is shown both as a semicircle connecting the two subunits and as the symbol, γ.

Control over polyamide binding modes, in order to favor the preferred hairpin motif, has been demonstrated by increasing the length of the central residue from three carbons (β) to four (γ).[15] Solution-phase structural studies showed that β-linked polyamides can form hairpins, although they are more conformationally constrained than those linked by γ.[16] Boger and coworkers reported that the turn conformation of a β linker can be reinforced by substituting the prochiral α-(R)-proton of β with $-OCH_3$.[17] Prior to this work, Herman et al. demonstrated the use of α-(R)-amino-substituted γ ($^{H_2N}\gamma$) to increase the binding affinity of hairpin polyamides.[18]

Until now, it was believed that polyamides containing γ residues should have a strong preference for binding in hairpin over extended conformation. We find, however, that highly β-substituted polyamides can bind with high affinity as a hairpin and in an extended to 1:1 conformation. For example, the polyamide of sequence Im-β-ImPy-**β**-Im-β-ImPy-β-Dp (**1**) (Dp = dimethylaminopropylamide) targets 5'-AAAGAGAAGAG-3' as a 1:1 complex and 5'-TAGCGCAGCGCTA-3' as a 2:1 complex.[14] Prior work has shown that the γ-linked analogue, Im-β-ImPy-**γ**-Im-β-ImPy-β-Dp (**2**), which differs from **1** by a single methylene unit in the central residue, binds with high affinity as a hairpin to its target sequence, 5'-TAGCGCT-3'.[10] Given the structural similarity between **1** and **2**, it was a simple prediction that **2** could also bind in a 1:1 mode. We find that, indeed, **2** binds to 5'-TAGCGCA-3' in a folded, hairpin conformation and to 5'-AAGAGAAGAG-3' in an extended, 1:1 conformation, with surprisingly similar affinities.

This finding prompted us to develop design principles for polyamides that would favor the hairpin binding mode. We hypothesized that the $^{H_2N}\gamma$ turn could be used to disfavor the extended 1:1 binding mode, while favoring the hairpin mode. Accordingly, the polyamide Im-β-ImPy-$^{\mathbf{H_2N}}$**γ**-Im-β-ImPy-β-Dp (**3**) and its acetamide analogue, Im-β-ImPy-$^{\mathbf{Ac}}$**γ**-Im-β-ImPy-β-Dp (**4**), were prepared (**Figure 2**). In addition to β-rich polyamides, we were interested in studying conformational preferences in eight-ring polyamides, such as ImPyPyPy-**β**-ImPyPyPy-β-Dp (**5**) and ImPyPyPy-**γ**-ImPyPyPy-β-Dp (**6**). Based on established recognition rules for the 1:1 and 2:1 motifs,[1,8,10,13] we constructed a plasmid containing the sequences 5'-TAGCGCT-3' and 5'-AAAGAGAAGAG-3', for compounds **1** – **4,** as well as 5'-TAGTACT-3' and 5'-AAAAAGAAAAG-3', for compounds **5** and **6** (**Figure 3**). Equilibrium association

constants and binding site sizes for compounds **1** – **6** at their target sites were determined in order to derive the relative preference of each compound for hairpin versus 1:1 binding modes.

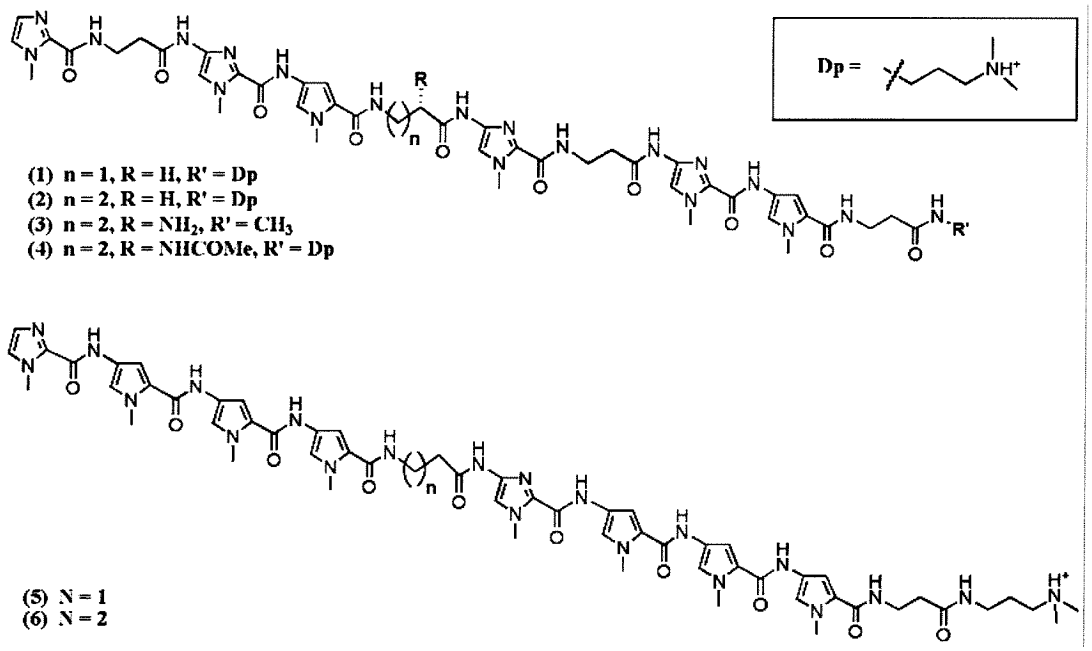


Figure 8.2 Chemical structures of polyamides **1** – **6**.

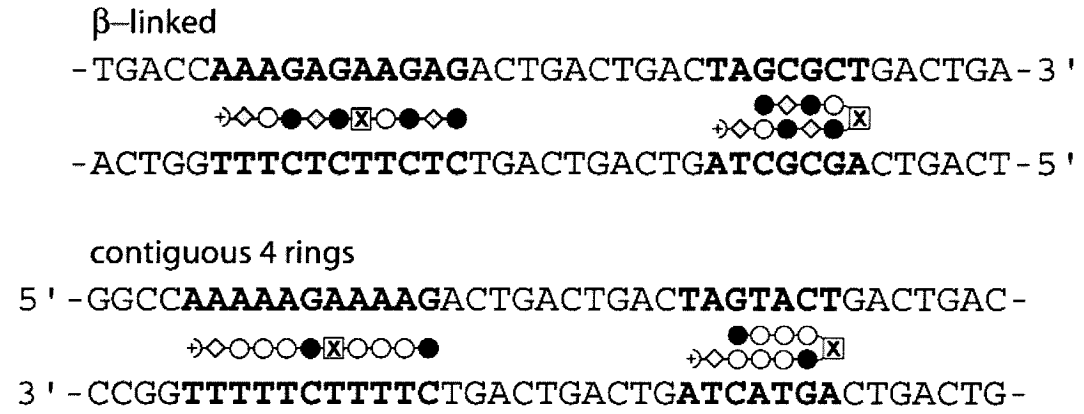


Figure 8.3 The designed insert cloned into plasmid pAU27. The targeted recognition sites are shown in bold type. Two sets of polyamides are shown as dot models. The two at left represent compounds **5** and **6** bound in the predicted extended and hairpin conformations. The two at right represent compounds **1** – **4** bound in the predicted extended and hairpin conformations. Polyamides are shown as dot models, with shaded and non-shaded circles representing imidazole and pyrrole, respectively, and gray diamonds indicating beta-alanine. The variable linker position is shown as a square containing the letter X.

8.2 Results.

DNA Binding Affinity and Sequence Specificity. Quantitative DNase I footprint titrations[19] were performed for polyamides **1** – **6** on the 288 bp PCR product of pAU27, in order to compare the equilibrium association constants for the resulting complexes. Polyamides **1** – **4** were designed to bind at DNA sites 5'-TAGCGCT-3' (hairpin) and 5'-AAAGAGAAGAG-3' (extended 1:1) (**Figure 4**). The β-linked compound, **1**, bound to the extended site with very high affinity (K_a = 1.5 x 10^{10} M^{-1}) and more than 150-fold preference over the hairpin site (K_a = 9.7 x 10^7 M^{-1}) (Table 1). The γ-linked compound, **2**, bound with similar affinity to the hairpin (K_a = 7.6 x 10^8 M^{-1}) and extended (K_a = 1.3 x 10^8 M^{-1}) sites. The ^{H_2N}γ-linked polyamide **3** bound to the hairpin site with an almost exact reversal of specificity in comparison to compound **1**, very high affinity (K_a = 1.2 x 10^{10} M^{-1}) and more than 150-fold specificity versus the extended site (K_a = 7.6 x 10^7 M^{-1}). Polyamide **4** (^{Ac}γ-linker) also bound with very high affinity to the hairpin site (K_a = 8.5 x 10^9 M^{-1}), but with greater than 500-fold specificity versus the extended site (K_a = 1.6 x 10^7 M^{-1}).

Eight-ring polyamides **5** and **6** were designed to bind DNA sites 5'-TAGTACT-3' (hairpin) and 5'-AAAAAGAAAAG-3' (extended 1:1) (**Figure 5**). The β-linked compound, **5**, bound with similar affinity ($K_a \approx 3 \times 10^9$ M^{-1}) to both designed sites (Table 1). The γ-linked polyamide, **6**, bound the hairpin site with very high affinity (K_a = 1.0 x 10^{10} M^{-1}) but only 5-fold selectivity over its single base-pair mismatch hairpin site, 5'-AAGAA**A**A-3' (K_a = 1.8 x 10^9 M^{-1}) (mismatch base bolded), which is located within the extended site (see MPE results below). All binding isotherms fit well to an n = 1 Hill equation, supporting 1:1 polyamide:DNA stoichiometry (**Figures 4 and 5**).

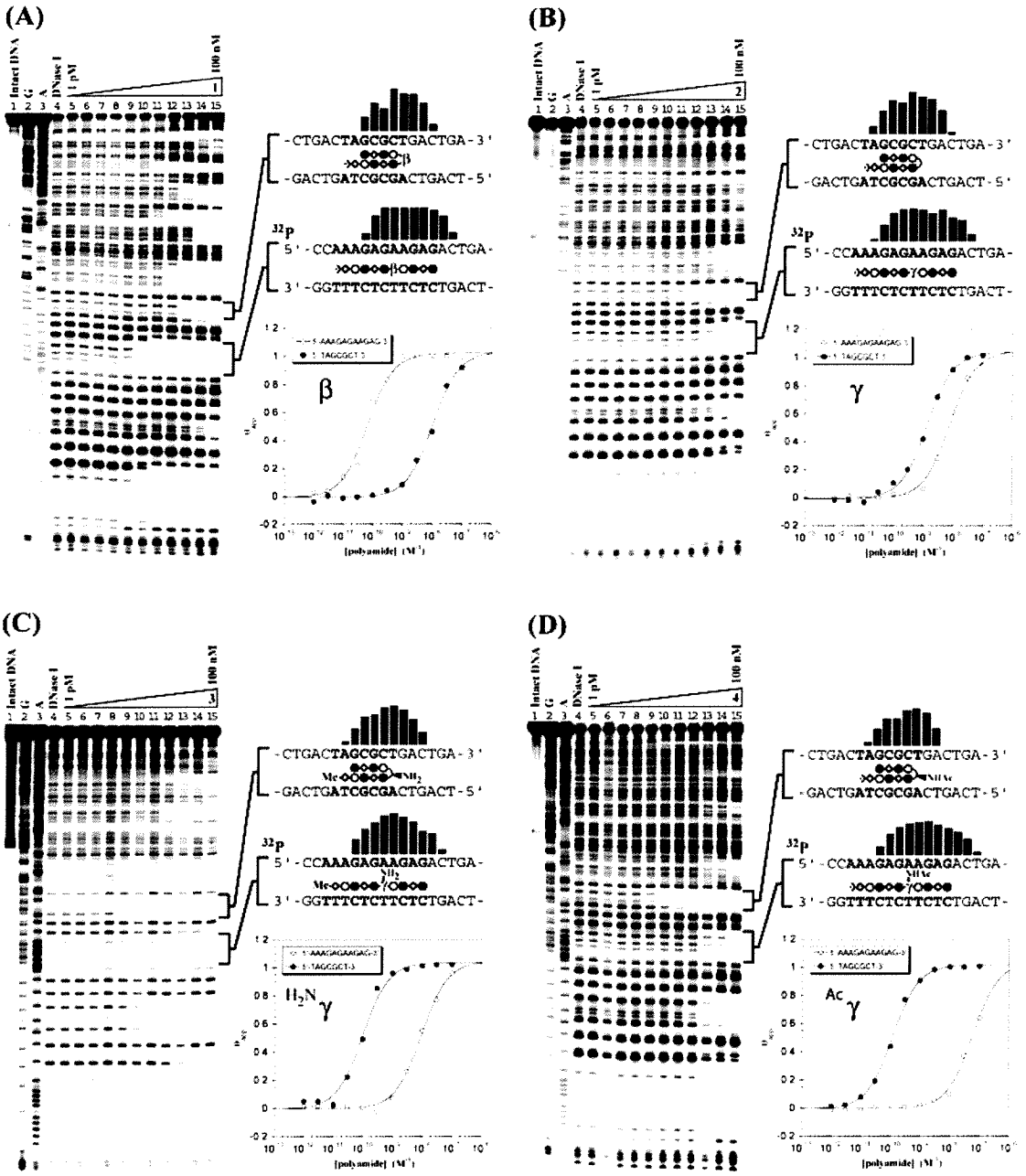


Figure 8.4 (A – D) Quantitative DNase I footprinting[21] for polyamides **1** – **4**, respectively, on the 288 bp, 5'-end-labeled PCR product of plasmid pAU27: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (right top) Schematic illustrating the observed protection pattern derived from the MPE footprinting experiment, with the polyamides shown in the observed conformation; (right bottom) binding isotherms derived from the DNase I footprinting experiment for the two designed sites, 5'-AAAGAGAAGAG-3' and 5'-TAGCGCT-3', as determined from a non-linear least squares fit.

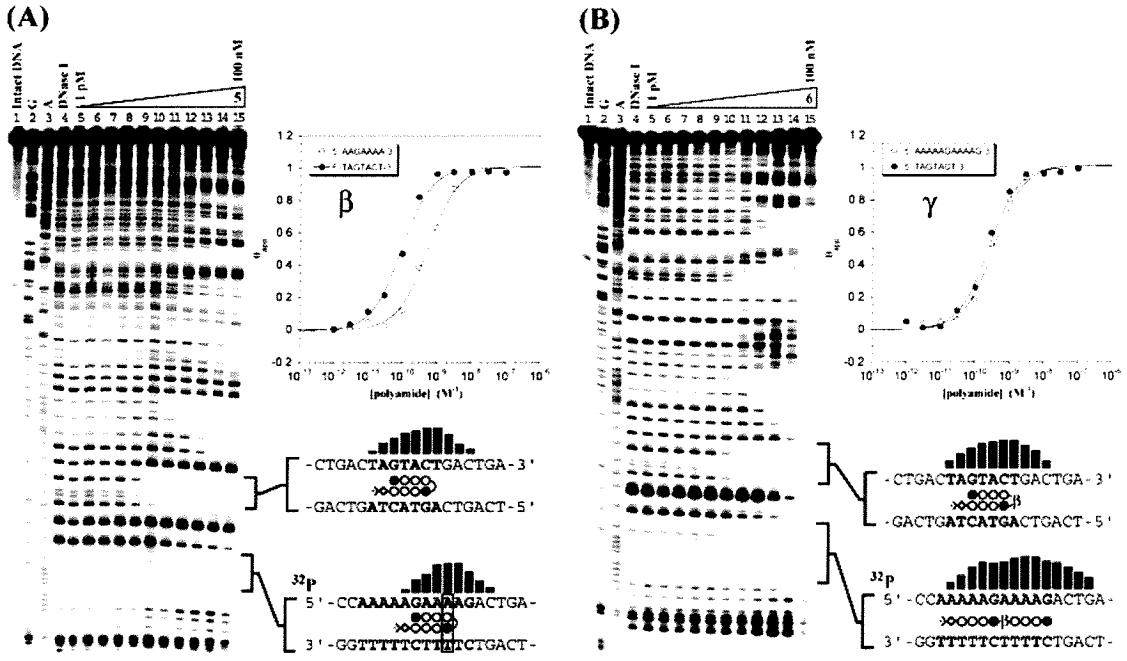


Figure 8.5 (A and B) Quantitative DNase I footprinting[21] for polyamides **5** and **6**, respectively, on the 288 bp, 5'-end-labeled PCR product of plasmid pAU27: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (right top) binding isotherms derived from the DNase I footprinting experiment for the two designed sites, 5'-AAAAAGAAAAG-3' and 5'-TAGTACT-3', as determined from a non-linear least squares fit; (right bottom) schematic illustrating the observed protection pattern derived from the MPE footprinting experiment, with the polyamides shown in the observed conformation.

Binding Site Size Determination. Binding modes were deduced from methidium propyl-EDTA (MPE) footprinting[19] on the basis of previously characterized hairpin and 1:1 complexes for polyamides **1**, **2**, and **6** at identical sites.[8,10,13] For example, the eight-ring hairpin, **6**, occupies a seven base-pair binding site,[3] whereas the prototypical extended binder, **1**, occupies an eleven base-pair binding site.[13] **Figures 4 and 5** display occupation histograms, which are derived from the MPE gels (not shown), depicted at the right of the respective DNase gel. In all cases, polyamides occupied their target hairpin sites as hairpins. Polyamides **1** – **5** bound to their target 1:1 sites in an extended 1:1 binding mode. However, polyamide **6** occupied a smaller site within its designed 1:1 site. Closer inspection of the occupied site suggests that **6** binds as a hairpin to a single

base-pair mismatch site (**Figure 5 and Table 1**). Inspection of the relative footprint sizes in the DNase gels further supports the MPE results.

Table 1. Equilibrium Association Constants, K_a (M^{-1})*

Polyamide	Linker	Hairpin	1:1	Specificity (Hairpin/1-1)
Im-β-ImPy-		5'-TAGCGCT-3'	5'-AAAGAGAAGAG-3'	
(1)	β	$9.7(\pm 0.8) \times 10^{7}$	$1.5(\pm 0.5) \times 10^{10}$	0.0065
(2)	γ	$7.6(\pm 0.6) \times 10^{8}$	$1.3(\pm 0.5) \times 10^{8}$	5.8
(3)	H_2Nγ	$1.2(\pm 0.2) \times 10^{10}$	$7.6(\pm 0.2) \times 10^{7}$	160
(4)	^{Ac}γ	$8.5(\pm 0.5) \times 10^{9}$	$1.6(\pm 0.4) \times 10^{7}$	530
ImPyPyPy-		5'-TAGTACT-3'	5'-AAAAAGAAAAG-3'**	
(5)	β	$3.9(\pm 0.7) \times 10^{9}$	$2.7(\pm 0.7) \times 10^{9}$ *a*	1.4
(6)	γ	$1.0(\pm 0.2) \times 10^{10}$	$1.8(\pm 0.2) \times 10^{9}$ *b*	5.6

* Values reported are the mean values from at least three DNase I footprint titration experiments, with the standard deviation given in parentheses. Assays were performed at 22 °C in a buffer of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0
** Observed binding sites are indicated as boxes *a* and *b*.

8.3 Discussion.

Previous studies of polyamide **1** demonstrated its high affinity for the extended 1:1 binding mode.[13,14] The >150-fold specificity of **1** for 1:1 versus hairpin binding sites, presented here, may be attributed to steric destabilization incurred when the β residue adopts a hairpin conformation.[16] Structural studies of 1:1 and 2:1 complexes support the preference of β to extend in order to allow optimal alignment between amino acids residues and DNA base pairs.[11,16,20] Inspection of the DNase gel reveals that **1** bound with very high affinity to the 5'-AAAAAGAAAAG-3' site, in accordance with the rules for recognition in the 1:1 motif.[13] The γ-linked polyamide, **2**, discriminates the least between the hairpin and 1:1 sites. This is problematic because hairpin polyamides, which are the most sequence-specific class of DNA binding polyamides, most commonly contain γ as the turn residue. Therefore, a linker that favors hairpin formation over alternative binding modes would be of great utility.

Based on molecular modeling using available solution structures,[16,20] we postulated that incorporation of $^{H_2N}\gamma$ at the turn position should promote hairpin binding while disfavoring extended binding due to unfavorable steric interactions of the α-(R)-amino group with the wall of the minor groove. Indeed, this result was observed for the $^{H_2N}\gamma$-linked polyamide **3**, which shows >150-fold preference for hairpin versus extended binding—a complete reversal of preference from compound **1**. It is worth noting that **3** contains a truncated tail in order to maintain a single positive charge. We have also tested the doubly charged Dp-tail analogue of **3** and find that its recognition properties are virtually identical (data not shown). Compounds **1** – **3** are relatively non-specific, binding with high affinity to many other sites on the plasmid. Compound **4**, which has an acetylated Herman turn and a Dp tail, binds with high affinity as a hairpin but with markedly reduced nonspecific binding. Moreover, **4** exhibits >500-fold preference for hairpin versus extended binding modes, which is an effective 82,000-fold or ~7 kcal/mol reversal of specificity in comparison to **1**. The exceptional specificity of **4** may be attributed to the limited mobility of the acetamide group in the hairpin conformation, which should disfavor alternative binding modes.

Polyamides **5** and **6** contain contiguous four-ring subunits, which are inherently limited in conformational flexibility. It is expected that these compounds should display different recognition properties than polyamides **1** – **4**. The β-linked polyamide (**5**) binds in an extended mode to the sequence, 5'-AAAAAGAAAAG-3', and as a hairpin to the designed site, 5'-TAGTACT-3'. However, unlike the β-linked polyamide **1**, described above, **5** shows no preference for hairpin or 1:1 modes, binding both sites with similar affinities. On the other hand, the γ-linked compound (**6**) *binds as a hairpin to both sites*,

tolerating a single base-pair mismatch within the 1:1 binding site, 5'-AAGAAAA-3', rather than adopting an extended conformation. This result provides further evidence that contiguous ring polyamides linked by γ are not inclined to adopt an extended conformation upon binding.[15]

Implications for the Design of Minor Groove Binding Polyamides. The results presented here indicate that the hairpin mode of binding can be preferred for both flexible (β-rich) and rigid (contiguous ring) polyamides. For the β-rich series, the hairpin mode can be favored by incorporating an amino or acetamide substituent at the α-(R) position of γ. The $^{Ac}\gamma$ residue substantially favors hairpin binding, while maintaining high affinity. Alternatively, the extended 1:1 binding mode can be favored by replacing γ with a β residue. For the conformationally rigid, γ-linked polyamide **6**, we do not observe binding in an extended 1:1 mode. We are currently most interested in favoring hairpin binding because of its higher information density and therefore higher capacity for programmable DNA sequence selection.[2] The design principles elucidated herein should greatly improve the fidelity of sequence recognition for hairpin polyamides in larger, genomic contexts.

8.4 Materials and Methods.

Materials. Methylamine, piperidine, and dimethylaminopropylamine were purchased from Aldrich. Dimethylformamide (DMF) and diisopropylethylamine (DIEA) were purchased from Applied Biosystems. Acetic anhydride and acetonitrile were from EM. Trifluoroacetic acid (TFA) was from Halocarbon. (R)-2-Fmoc-4-Boc-diaminobutyric acid was purchased from Bachem. Boc-β-Pam resin was purchased from

Peptides International. HPLC analysis was performed on a Beckman Gold system using a Rainin C_{18}, Microsorb MV, 5 μm, 300 x 4.6 mm reversed phase column in 0.1% (wt/v) TFA with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% acetonitrile/min. Preparatory reversed phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100 μm C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. Oligonucleotide inserts were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology and used without further purification. Plasmids were sequenced by Davis Sequencing (Davis, CA). Glycogen (20 mg/mL), dNTP's (PCR nucleotide mix), and all enzymes (unless otherwise stated) were purchased from Boehringer-Mannheim and used with their supplied buffers. pUC19 was from New England Biolabs. Deoxyadenosine [γ-^{32}P] triphosphate was from ICN. Calf thymus DNA (sonicated, deproteinized) and DNase I (7500 u/mL, FPLC pure) were from Amersham-Pharmacia. AmpliTaq DNA polymerase was from Perkin Elmer and used with the supplied buffers. HEPES was from Sigma. Tris-HCl, dithiothreitol (DTT), RNase-free water (used for all DNA manipulations), and 0.5 M EDTA were from US Biochemicals. Ethanol (200 proof) was from Equistar. Calcium chloride, potassium chloride, and magnesium chloride were from Fluka. Formamide and pre-mixed tris-borate-EDTA (Gel-Mate, used for gel running buffer) were from Gibco. Bromophenol blue was from Acros. All reagents were used without further purification.

Polyamide Synthesis. Polyamides **1**-**6** (**Figure 2**) were prepared according to solid-phase protocols.[23] The synthesis and characterization of polyamides **1**, **2**, and **6** have been reported previously.[8,10,13]

Im-β-ImPy-(R)^{H2N}γ-Im-β-ImPy-β-Me (**3**). Im-β-ImPy-(R)^{H2N}γ-Im-β-ImPy-β-Pam resin was synthesized in a stepwise fashion from Boc-β-Pam resin (0.59 mmol/g) (Peptides International) using manual solid-phase protocols.[18,23] The chiral diaminobutyric acid "turn" residue was incorporated by coupling (R)-2-Fmoc-4-Boc-diaminobutyric acid (10 equivalents) to 300 mg Boc-Im-β-Im-Py-β-Pam resin in 2 mL DMF with 1.1 equivalents of DIEA at 37 °C for 2 h, followed by an acetylation wash.[23] Subsequent coupling steps used 1.1 equivalents of DIEA and 45 minute coupling times at room temperature to minimize Fmoc deprotection. Im-β-ImPy-(R)^{Fmoc}γ-Im-β-ImPy-β-Pam resin was treated with piperidine for 20 minutes at room temperature to remove the Fmoc group. 100 mg (38 μmol) of vacuum-dried Im-β-ImPy-(R)^{H2N}γ-Im-β-ImPy-β-Pam resin was cleaved in 30 mL condensed methylamine in a Parr bomb apparatus at 50 °C for 2 h, then overnight at room temperature. The methylamine was allowed to evaporate at ambient pressure and temperature, and the resin was suspended in 2 mL acetonitrile, followed by 7 mL 0.1% (wt/v) $TFA_{(aq)}$. The suspension was filtered, and the filtrate was purified by reversed phase preparatory HPLC to afford **3** as a white powder (7.1 μmol, 19% recovery) upon lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1066.49 (1066.48 calcd. for $C_{46}H_{60}N_{21}O_{10}^{+}$).

Im-β-ImPy-(R)^{Ac}γ-Im-β-ImPy-β-Dp (**4**). Im-β-ImPy-(R)^{H2N}γ-Im-β-ImPy-β-Pam resin was synthesized as described above for **3**. The resin was washed with acetic anhydride in DMF and DIEA for ten minutes at room temperature and then dried *in vacuo*. 100 mg (37 μmol) Im-β-ImPy-(R)^{Ac}γ-Im-β-ImPy-β-Pam resin was treated with dimethylaminopropylamine at 100 C for 2 h. The resin was removed by filtration, and the filtrate was diluted to 10 mL with 0.1 % (wt/v) $TFA_{(aq)}$ and purified by reversed phase

HPLC. **4** was obtained as a white powder (4.0 μmol, 11% recovery) upon lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1179.57 (1179.57 calcd. for $C_{52}H_{71}N_{22}O_{11}^{+}$).

ImPyPyPy-β-ImPyPyPy-β-Dp (**5**). **5** was synthesized and purified according to standard methods.[23] MALDI-TOF-MS (monoisotopic), 1208.6 (calcd. 1208.6 for $C_{57}H_{70}N_{21}O_{10}^{+}$).

Construction of Plasmid DNA. Plasmid pAU27 (**Figure 3**) was constructed by inserting the hybridized oligonucleotide, 5'-G A T C C G G G G C C A A A A A G A A A A G A C T G A C T G A C T A G T A C T G A C G A C T G A C C A A A G A G A A G A G A C T G A C T G A C T A G C G C T G A C T G A-3' • 5'-A G C T T C A G T C A G C G C T A G T C A G T C A G T C T C T T C T C T T T G G T C A G T C G T C A G T A C T A G T C A G T C A G T C T T T T C T T T T T G G C C C C G-3', into the *Bam*HI/*Hin*DIII polycloning site in pUC19, with subsequent transformation in to JM109 subcompetent cells (Promega), according to standard protocols.[24] Plasmid DNA was isolated using WizardPlus Midi Preps from Promega.

DNA Radiolabeling and Footprinting Experiments. The 5' end-labeling of plasmid pAU27 as well as DNase I and MPE footprinting experiments were performed exactly in accordance with published protocols.[21] The PCR method was chosen for 5' end-labeling, employing two primer oligonucleotides, **^{32}P**-5'- A A T T C G A G C T C G G T A C C C G G -3' (forward) and 5'- C T G G C A C G A C A G G T T T C C C G -3' (reverse) to complement the pUC19 *Eco*RI and *Pvu*II sites, respectively, such that amplification by PCR generates the 288-bp, 3'-filled E*co*RI/*Pvu*II restriction fragment.

Acknowledgements. We wish to thank Dr. Eldon Baird for synthesizing compound **5**.

8.5 References.

[1] Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215-2235.

[2] Marques, M. A.; Doss, R. A.; Urbach, A. R.; Dervan, P. B. *Helv. Chim. Acta* **2002**, *85*, 4485-4517.

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

[4] Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1989**, *112*, 1393-1399.

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

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

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

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

[9] Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6981-6985.

[10] Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.

[11] de Clairac, R. P. L.; Seel, C. L.; Geierstanger, B. H.; Mrksich, M.; Baird, E. E.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1999**, *121*, 2956-2964.

[12] Janssen, S.; Durussel, T.; Laemmli, U. K. *Mol. Cell.* **2000**, *6*, 999-1011.

[13] Urbach, A. R.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4343-4348.

[14] Dervan, P. B.; Urbach, A. R. in *Essays in Contemporary Chemistry*, From Molecular Structure Toward Biology, eds. Quinkert, G. & Kisakürek, M. V. (Verlag Helvetica Chimica Acta, Zurich) **2000**.

[15] Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160-6166.

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

[17] Woods, C. R.; Ishii, T.; Wu, B.; Bair, K. W.; Boger, D. L. *J. Am. Chem. Soc.* **2002**, *124*, 2148-2152.

[18] Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391.

[19] Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450-466.

[20] Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B. *J. Mol. Biol.* **2002**, *320*, 55-71.

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

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