

# **Solid Phase Synthesis of DNA-Binding Small Molecules**

Thesis by

Eldon Eugene Baird

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1999

(Submitted August 25, 1998)

© 1999

Eldon Eugene Baird

All Rights Reserved

*To Sarah*

*and*

*to my family (past, present & future)*

## Acknowledgments

I would first like to thank my teachers (who have all become friends as well); Mr. Pik who first taught me to love chemistry, biology, and mushrooms; Dr. Tessi Kanavarioti and Prof. Claude F. Bernasconi who introduced me to chemical research; Prof. John D. Baldeschwieler who has introduced me to business, and Prof. Peter B. Dervan, who has provided friendship, teaching, freedom, and support that has made my time at Caltech gratifying and simply an amazing learning experience.

I thank the people who have supported me through the years at Caltech. First my relationship with Sarah White Baird, has shaped my life both in and out of the lab. Our trips to Santa Cruz and Monterey have been some of the best times during graduate school. I value very much the time we have shared together the last few years and the time that we will share in years to come. I thank Joe Hacia, who took the time to share some beers and teach me a lot of science; Matt Taylor, who was always excited by new ideas; Ramesh Baliga, who was always excited about science; Jim Turner, who was a great colleague and friend; Ryan Bremer, who always believed; and Jason Szewczyk, who has provided loyal support. I would like to thank Pete Beal, Scott Priestly, John Trauger, David Herman, Aileen Chang, and Nick Wurtz for making the third floor of Church more pleasant; and Milan Mrksich for providing a great inspiration to all of us on the minor groove project.

The last few years have provided an amazing opportunity to meet and work with an incredible group of scientists from a diverse group of fields. The results described here are the product of many collaborations, and I would like to thank some of the people who I have had the pleasure to work with, both at Caltech and elsewhere. It has been an especially great pleasure and honor to work with John Trauger, whose skill as a researcher and thoughtful approach to science were inspirational; and Joel Gottesfeld, whose excitement about polyamides, creative experimental designs, and numerous visits have contributed greatly to the excitement of the Caltech research projects over the past three years.



Footprinting experiments in the Dervan group, have been conducted in collaboration with Will Greenberg, David Herman, Jim Kelly, Michelle Parks, Sue Swalley, Jason Szewczyk, John Trauger, Jim Turner, and Sarah White.

Biology experiments have been conducted primarily in collaboration with Prof. Joel Gottesfeld and members of his research group. Joel has told many of his friends at Scripps about polyamides and experiments have been done with Prof. Don Mosier, Prof. Pete Ghazel, Prof. Peter Vogt and Masa Aoki. We have also collaborated with Prof. Mel Simon and Pam Eversol at Caltech and Prof. Julie Nyborg and Brian Lenzmier at Colorado State.

For structure studies we have had a great collaboration with Prof. Doug Rees and Clara Kielkopf at Caltech. Clara and Doug have been wonderful to work with, and have contributed greatly to the field's understanding of minor groove recognition. I have also had the pleasure of working with David Wemmer and his research group on continuing NMR collaborations.

For physical chemistry studies we have had a great collaboration with Dan Pilsch and Ken Breslauer at Rutgers University.

I also have to thank the support staff at Caltech, who have put up with large purchase orders, late night waxing of floors, large quantities of chemical and radioactive waste, big piles of trash, and enormous numbers of library and photocopy orders. Your patience and help through the years is appreciated. I would especially like to thank Tom Dunn, who has always provided thoughtful wisdom about life and electronics; Margot Hoyt, who keeps things organized and provides a pleasant ideology; and Gary Hathaway (and Dirk) who processed many hundreds of polyamide mass specs and provided great conversation to go with them.

I would like to thank the members of my committee, Prof. R. H Grubbs, Prof. J. D. Baldeschwieler, Prof. D. C. Rees, and Prof. P. B. Dervan for their time and helpful comments through the years.

Finally, I would like to thank my family for always supporting whatever I am doing, and who didn't complain even though there were some years when I didn't have much time to come home and visit.

## **Abstract**

Small molecules that bind to any predetermined DNA sequence in the human genome are potentially useful tools for molecular biology and human medicine. A twenty year research effort led by Dr. Peter B. Dervan at the California Institute of Technology has led to the development of “pairing rules” to control rationally the sequence-specificity of polyamides that bind in the DNA minor groove. During the course of my Ph.D. research, methodology was developed for the machine-assisted solid phase synthesis of these DNA-binding polyamides (Chapter 2). The large number of polyamides made available by the solid phase synthetic methodology has greatly accelerated the development of this class of molecules. Polyamides prepared by solid phase synthetic methodology have been used by a variety of collaborators to: extend the targetable binding-site size (Chapter 3), recognize predetermined DNA sequences with subnanomolar affinity (Chapter 6), regulate gene expression in human cells (Chapter 7), and recognize all four base pairs in the DNA minor groove (Chapter 8).

## Table of Contents

	page
Acknowledgments.....	iv
Abstract.....	vi
Table of Contents.....	vii
List of Figures and Tables.....	viii
<b>CHAPTER ONE:</b> Introduction.....	1
<b>CHAPTER TWO:</b> Solid Phase Synthesis of DNA-binding Polyamides Containing Aromatic Amino Acids. ....	14
<b>CHAPTER THREE:</b> Extending the Binding Site Size Limit of the 2:1 Pyrrole-Imidazole Polyamide-DNA Motif.....	41
<b>CHAPTER FOUR:</b> Optimization of the Hairpin Polyamide Design for Recognition of the Minor Groove of DNA .....	91
<b>CHAPTER FIVE:</b> Stereochemical Control of the DNA-Binding Affinity, Sequence-Specificity, and Orientation Preference of Hairpin Polyamides in the Minor Groove.....	145
<b>CHAPTER SIX:</b> Recognition of DNA by Designed Ligands at Subnanomolar Concentrations.....	187
<b>CHAPTER SEVEN:</b> Regulation of Gene Expression by Cell-Permeable DNA- Binding Small Molecules .....	249
<b>CHAPTER EIGHT:</b> Recognition of the Four Watson-Crick Base Pairs in the DNA Minor Groove by Synthetic Ligands.....	280
<b>CHAPTER NINE:</b> Sequence Specific Recognition of Double Helical DNA by Major-Minor Groove Binding Ligands .....	324

## List of Figures and Tables

<b>CHAPTER ONE</b>	<b>page</b>
Figure 1.1 Model of protein regulation of gene transcription .....	2
Figure 1.2 Structure of double-helical B-form DNA .....	3
Figure 1.3 Schematic model for recognition of the DNA minor groove .....	4
Figure 1.4 Structures of small molecules isolated from natural sources .....	4
Figure 1.5 Recognition of A,T-rich sequences by complexes of Distamycin .....	5
Figure 1.6 The polyamide pairing rules .....	6
Figure 1.7 Recognition of 5'-TGTTA-3' by unlinked and hairpin polyamides .....	7
Figure 1.8 A comparison of cycle-polyamide and hairpin-polyamide motifs .....	8
Figure 1.9 Recognition of all four base pairs by hairpin-polyamide .....	10
Figure 1.10 Space filling model for polyamide x-tal structure .....	11
 <b>CHAPTER TWO</b>	
Figure 2.1 Structure of hairpin polyamides .....	16
Figure 2.2 Synthesis of Boc-Py-OBt .....	17
Figure 2.3 Synthesis of Boc-Im Acid .....	17
Figure 2.4 Synthesis of dimer building blocks .....	17
Figure 2.5 Protocol for solid phase synthesis .....	18
Figure 2.6 Solid phase synthesis scheme for hairpin polyamide .....	19
Figure 2.7 Coupling kinetics .....	21
Figure 2.8 HPLC and mass of hairpin polyamide .....	22
Figure 2.9 Linkers for synthesis of Py-resin .....	29
Figure 2.10 Synthesis of AcPyPyPy .....	32
Figure 2.11 Cleavage of resin ester linkages .....	32
Figure 2.12 Stepwise monitoring of synthesis with HPLC .....	37
Figure 2.13 Picric acid titration .....	38
Table 1.1 Protocol for manual solid phase synthesis .....	18

### CHAPTER THREE

Figure 3.1	Structures of length dependence polyamides.....	43
Figure 3.2	2:1 binding models.....	44
Figure 3.3	MPE footprinting.....	44
Figure 3.4	Boc-chemistry solution phase synthesis .....	45
Figure 3.5	Effects of polyamide length on DNA-binding affinity .....	47
Figure 3.6	Complexes of ImPyPy-X-PyPyPy-Dp.....	50
Figure 3.7	Structures of ImPyPy-X-PyPyPy-Dp.....	51
Figure 3.8	Complex of ImPyPy- $\beta$ -PyPyPy-Dp.....	52
Figure 3.9	Model for complex formed by ImPyPy- $\gamma$ -PyPyPy-Dp.....	54
Figure 3.10	Structures of 4- $\beta$ -4 polyamides .....	57
Figure 3.11	Binding model of 4- $\beta$ -4 polyamides.....	58
Figure 3.12	Solid phase synthesis of 4- $\beta$ -4 polyamides.....	59
Figure 3.13	MPE footprinting of 4- $\beta$ -4 polyamides.....	60
Figure 3.14	Binding affinity of 4- $\beta$ -4 polyamides .....	61
Figure 3.15	Model of 16-bp recognition .....	66
Figure 3.16	2- $\beta$ -2- $\beta$ -2- $\beta$ -2 binding model .....	67
Figure 3.17	Diagram of ImImPyPy .....	69
Figure 3.18	Structure of 2ImImPyPy:DNA complex .....	71
Figure 3.19	Extended polyamide helix.....	72
Figure 3.20	Assorted $\beta$ -linked polyamides.....	78
Table 3.1	Apparent first order binding affinities for Im(Py) <sub>n</sub> -Dp.....	46
Table 3.2	Apparent first order binding affinities for ImPyPy-X-PyPyPy-Dp.....	53

### CHAPTER FOUR

Figure 4.1	Hairpin-polyamide binding model.....	93
Figure 4.2	Hairpin polyamide structures .....	94
Figure 4.3	Modification of hairpin C-terminus.....	95

Figure 4.4	Structure of prior art hairpin.....	97
Figure 4.5	Model of ImIm hairpin-polyamide complexes.....	98
Figure 4.6	Structures of ImIm hairpins .....	99
Figure 4.7	Solid phase synthesis of ImIm hairpins.....	100
Figure 4.8	Binding models for ImImIm hairpins .....	103
Figure 4.9	Structures of ImImIm hairpins.....	104
Figure 4.10	Solid phase synthesis of ImImIm hairpins.....	106
Figure 4.11	MPE footprinting and affinity cleaving of ImImIm hairpins.....	107
Figure 4.12	Im/Im pairing models.....	110
Figure 4.13	Structures of Im/Im hairpins .....	111
Figure 4.14	Equilibrium association constants of Im/Im hairpins.....	112
Figure 4.15	Im/Im pairing rules .....	113
Figure 4.16	Space filling model of G,C base pair .....	115
Figure 4.17	$\gamma$ -turn and $\beta$ -tail binding model .....	117
Figure 4.18	$\gamma$ -turn and $\beta$ -tail hairpin structures .....	118
Figure 4.19	'tail-less' hairpin structures .....	123
Figure 4.20	hairpin complexes with DNA duplexes .....	127
Table 4.1	Equilibrium association constants for 1-Im polyamides.....	96
Table 4.2	Equilibrium association constants for 2-Im polyamides.....	101
Table 4.3	Equilibrium association constants for 3-Im polyamides.....	108
Table 4.4	Equilibrium association constants for $\gamma$ -turn and $\beta$ -tail polyamides.....	120
Table 4.5	More association constants for $\gamma$ -turn and $\beta$ -tail polyamides .....	121
Table 4.6	Even more association constants for $\gamma$ -turn and $\beta$ -tail polyamides .....	123
Table 4.7	Hairpin binding enthalpies .....	128
Table 4.8	$\Delta T_m$ derived hairpin binding affinities.....	130
Table 4.9	Thermodynamic parameters for hairpin binding.....	131
Table 4.10	Thermodynamic consequences for single base pair mismatches.....	133

## CHAPTER FIVE

Figure 5.1	Polyamide subunit binding orientation.....	147
Figure 5.2	Binding models for both hairpin binding orientations.....	148
Figure 5.3	Structures of hairpin polyamides.....	150
Figure 5.4	Solid phase synthesis of hairpin polyamides.....	151
Figure 5.5	MPE•Fe(II) footprinting for forward and reverse sites .....	153
Figure 5.6	Affinity cleaving for forward and reverse sites.....	154
Figure 5.7	Hairpin binding models with affinity cleavage pattern .....	154
Figure 5.8	Polyamide-DNA dipole model.....	156
Figure 5.9	Hairpin folding model.....	157
Figure 5.10	Computer generated models of chiral hairpins.....	159
Figure 5.11	Hydrogen bond models for chiral hairpins.....	159
Figure 5.12	Structures of chiral hairpins .....	160
Figure 5.13	Solid phase synthesis of chiral hairpins.....	161
Figure 5.14	Mosher amide analysis of chiral hairpins.....	163
Figure 5.15	MPE footprinting of chiral hairpins.....	165
Figure 5.16	Affinity cleaving by chiral hairpins.....	166
Figure 5.17	Affinity cleaving models for chiral hairpins.....	166
Figure 5.18	Folding model for chiral hairpins.....	170
Table 5.1	Equilibrium association constants for forward and reverse sites .....	155
Table 5.2	Equilibrium association constants for chiral hairpins.....	167

## CHAPTER SIX

Figure 6.1	Structures of eight-ring hairpin polyamides.....	191
Figure 6.2	Hydrogen bond models for eight-ring hairpin polyamides.....	192
Figure 6.3	Binding models for all G,C hairpin complexes.....	195
Figure 6.4	Structures of 4-Im, 4-Py hairpin polyamides .....	196
Figure 6.5	Solid phase synthesis of 4-Im, 4-Py hairpin polyamides.....	198
Figure 6.6	MPE footprinting of 4-Im, 4-Py hairpin polyamides.....	199

Figure 6.7	Affinity Cleaving by 4-Im, 4-Py hairpin polyamides.....	200
Figure 6.8	Structures of ten-ring hairpin polyamides.....	204
Figure 6.9	Binding models for ten-ring hairpin polyamides.....	205
Figure 6.10	Solid phase synthesis of ten-ring hairpin polyamides .....	207
Figure 6.11	MPE footprinting of ten-ring hairpin polyamides.....	208
Figure 6.12	Affinity cleaving of ten-ring hairpin polyamides.....	209
Figure 6.13	Affinity cleaving binding models of ten-ring hairpin polyamides.....	210
Figure 6.14	Hairpin ribbon model with $\beta$ -springs.....	213
Figure 6.15	Binding model for twelve-ring hairpins with $\beta$ -springs .....	214
Figure 6.16	Structures of twelve-ring hairpins with $\beta$ -springs .....	215
Figure 6.17	MPE footprinting of twelve-ring hairpins with $\beta$ -springs.....	216
Figure 6.18	Structures of twelve-ring hairpins with Im- $\beta$ -Im.....	218
Figure 6.19	MPE footprinting of twelve-ring hairpins with Im- $\beta$ -Im .....	219
Figure 6.20	Space filling model of Im- $\beta$ -Im.....	221
Figure 6.21	Structures of eight-ring hairpins with Im- $\beta$ -Im.....	222
Figure 6.22	Schematic binding model for extended hairpin .....	225
Figure 6.23	Structure of extended hairpin .....	227
Figure 6.24	Solid phase synthesis of extended hairpin .....	228
Figure 6.25	Affinity cleavage by extended hairpin.....	229
Figure 6.26	Structures and binding models for cooperative hairpins .....	232
Figure 6.27	Equilibrium association constants for series of cooperative hairpins .....	234
Table 6.1	Equilibrium association constants for eight-ring hairpins .....	193
Table 6.2	Equilibrium association constants for 4-Im, 4-Py eight-ring hairpins.....	210
Table 6.3	Equilibrium association constants for ten-ring hairpins.....	212
Table 6.4	Equilibrium association constants for twelve-ring hairpins .....	217
Table 6.5	Association constants for twelve-ring hairpins with Im- $\beta$ -Im.....	220
Table 6.6	Association constants for eight-ring hairpins with Im- $\beta$ -Im.....	223
Table 6.7	Aliphatic/Aromatic pairing code .....	224
Table 6.8	Equilibrium association constants for extended hairpins .....	230



## CHAPTER SEVEN

Figure 7.1	Model of TFIIIA•DNA complex with eight-ring hairpin target.....	251
Figure 7.2	Structures of match and mismatch polyamides.....	252
Figure 7.3	Inhibition of 5S RNA gene transcription <i>in vitro</i> .....	253
Figure 7.4	Inhibition of 5S RNA gene transcription <i>in vivo</i> .....	254
Figure 7.5	HIV-1 TATA sequence.....	257
Figure 7.6	Structure of hairpin polyamides for HIV-1 targeting.....	258
Figure 7.7	Binding models for hairpin polyamides .....	260
Figure 7.8	MPE footprinting of HIV-1 hairpins.....	261
Figure 7.9	Affinity cleavage of HIV-1 hairpins.....	262
Figure 7.10	Binding model derived from affinity cleavage of HIV-1 hairpins.....	263
Figure 7.11	Association constants for HIV-1 hairpins .....	264
Figure 7.12	Targeting the HIV-1 promoter .....	267
Figure 7.13	Inhibition of HIV-1 transcription <i>in vitro</i> .....	270
Figure 7.14	Kinetics of inhibition of HIV-1 replication <i>in vivo</i> .....	271
Figure 7.15	Inhibition of HIV-1 replication <i>in vivo</i> .....	272
Table 7.1	Equilibrium association constants for HIV-1 polyamides .....	259
Table 7.2	Salt and temperature dependence.....	265

## CHAPTER EIGHT

Figure 8.1	Model for hairpin-DNA complex .....	282
Figure 8.2	Structures of six-ring hairpins .....	283
Figure 8.3	Partial sequence of restriction fragment .....	284
Figure 8.4	Chemical structure of A,T base pair .....	286
Figure 8.5	Eight-ring Hp-Im-Py polyamides .....	287
Figure 8.6	Eight-ring affinity cleaving analogs.....	288
Figure 8.7	Six-ring Hp-Im-Py polyamides.....	289

Figure 8.8	Synthesis of Hp monomer .....	290
Figure 8.9	Solid phase synthesis of Hp-Im-Py polyamides.....	291
Figure 8.10	Synthesis of affinity cleaving hairpin.....	292
Figure 8.11	Affinity cleavage with eight-ring Hp-Im-Py hairpin.....	293
Figure 8.12	Structures of ten-ring Hp-Im-Py hairpins.....	301
Figure 8.13	Structure of ImHpPyPy and ImPyPyPy.....	305
Figure 8.14	Structure of ImHpPyPy bound to DNA double helix. ....	308
Figure 8.15	Diagram of 2 H-bonds between T and Hp.....	309
Figure 8.16	T•A base pair is melted .....	310
Figure 8.17	Interaction of Hp/Py pair with T•A base pair .....	311
Figure 8.18	All ligand-DNA hydrogen bonds .....	312
Table 8.1	Equilibrium association constants for ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp.....	285
Table 8.2	Association constants for eight-ring Hp-Im-Py hairpins.....	294
Table 8.3	Association constants for six-ring Hp-Im-Py hairpins for TGWTT.....	296
Table 8.4	Association constants for six-ring Hp-Im-Py hairpins for TGWTA .....	296
Table 8.5	Association constants for six-ring Hp-Im-Py hairpins for TGWAT .....	297
Table 8.6	Association constants for six-ring Hp-Im-Py hairpins for TGWAA.....	297
Table 8.7	Association constants for six-ring Hp-Im-Py hairpins for TGTWA .....	298
Table 8.8	Association constants for six-ring Hp-Im-Py hairpins for TGTWT.....	298
Table 8.9	Association constants for six-ring Hp-Im-Py hairpins for TGAWA.....	299
Table 8.10	Association constants for six-ring Hp-Im-Py hairpins for TGAWT .....	299
Table 8.11	Pairing Code for Minor Groove Recognition .....	300
Table 8.12	Association constants for ten-ring Hp-Im-Py hairpins TTA v. TAT.....	302
Table 8.13	Association constants for ten-ring Hp-Im-Py hairpins TWA v. TWT...	303
Table 8.14	Association constants for ten-ring Hp-Im-Py hairpins TWA v. GWT...	303
Table 8.15	Association constants for four-ring dimers GTAC v. GTAT .....	306
Table 8.16	Association constants for four-ring dimers GTAC v. GAAC v. GATC.	307

## CHAPTER NINE

Figure 9.1	Recognition of DNA by polyamide-oligonucleotide.....	326
Figure 9.2	Structures of polyamide-oligonucleotide.....	327
Figure 9.3	Synthesis of hairpin-polyamide-oligonucleotide.....	327
Figure 9.4	Mass analysis of hairpin-polyamide-oligonucleotide .....	328
Figure 9.5	Ribbon model of hairpin-polyamide-oligonucleotide.....	329
Figure 9.6	Solid phase synthesis of polyamide-oligonucleotide .....	331
Figure 9.7	Mass analysis of polyamide-oligonucleotide.....	332
Figure 9.8	HPLC analysis of polyamide-oligonucleotide.....	333
Figure 9.9	UV-vis analysis of polyamide-oligonucleotide .....	334
Figure 9.10	NMR analysis of polyamide-oligonucleotide .....	335
Figure 9.11	Ribbon model of cooperative polyamide-oligonucleotide.....	336
Table 9.1	Association constants for hairpin-polyamide-oligonucleotide.....	330
Table 9.2	Association constants for cooperative-polyamide-oligonucleotide .....	337

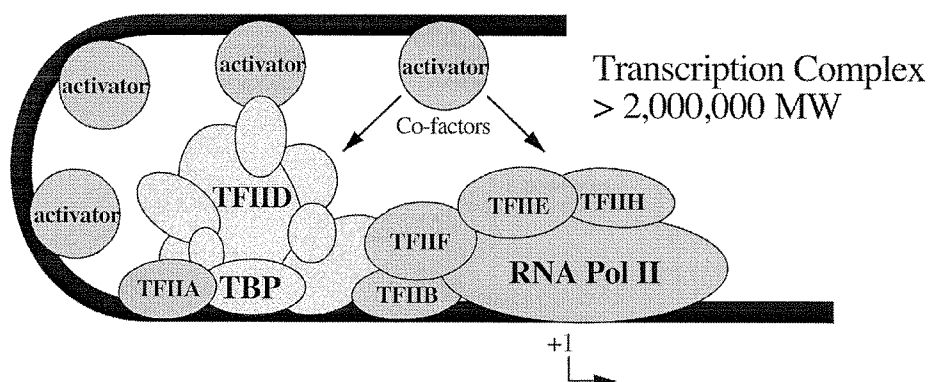
# **CHAPTER 1**

## **Introduction**

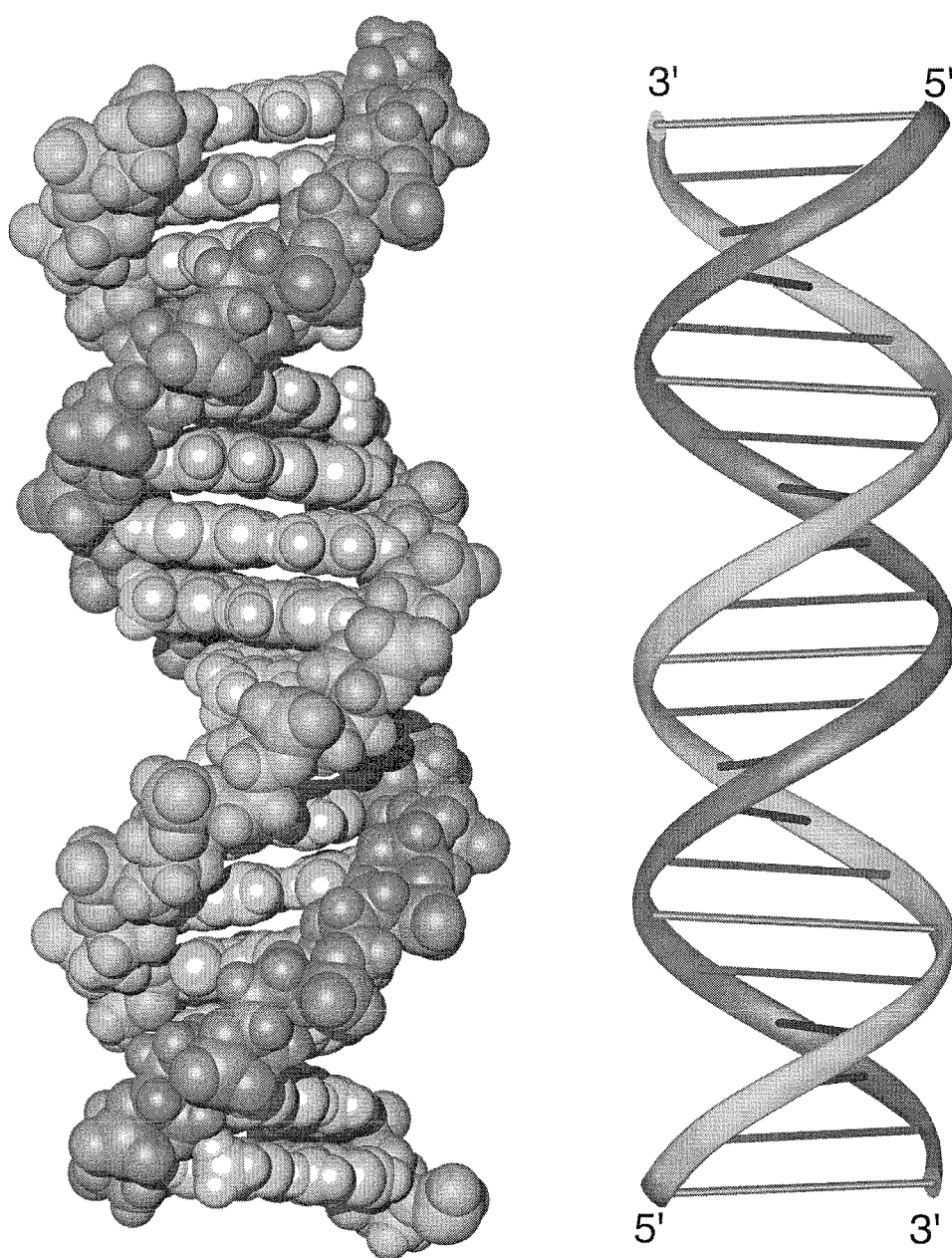
## Discussion of the Background.

In every human cell, the genetic information is stored on a string-like DNA polymer which is approximately 1 meter in length and contains  $3 \times 10^9$  units of information in the form of base pairs, within which is encoded approximately 80,000 to 100,000 genes or sets of instructions.<sup>1</sup> The specific interaction of proteins such as transcription factors with DNA controls the regulation of genes and hence cellular processes (Figure 1.1).<sup>2</sup> A wide variety of human conditions ranging from cancer to viral infection arise from malfunctions in the biochemical machinery that regulates gene-expression.<sup>3</sup> Designed small molecules which target specific DNA sequences offer a potentially general approach for gene-specific regulation.<sup>4</sup> Such molecules could be powerful therapeutics for combating life-threatening diseases which result from misregulation in transcription.

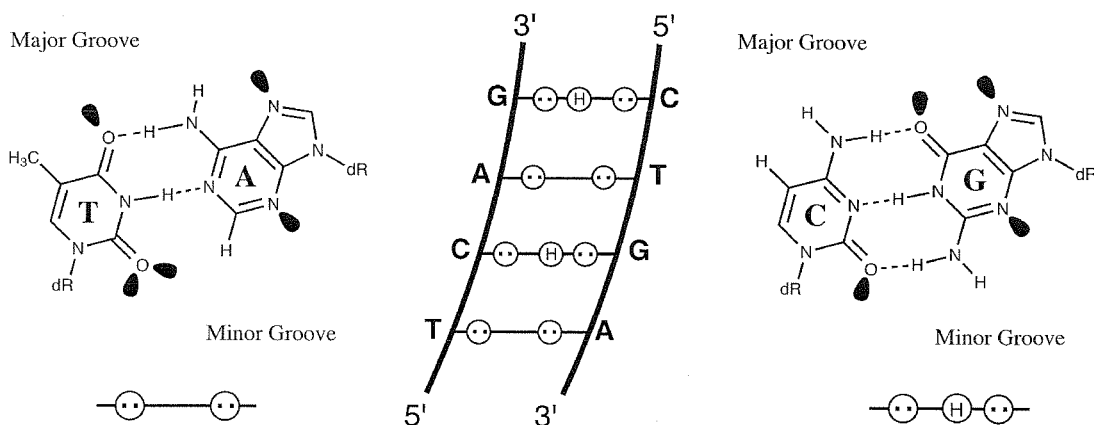
The genetic information is, in fact, stored on two strands of DNA (in antiparallel-orientation) in a structure termed the double helix.<sup>5</sup> The DNA double helix consists of A,T and G,C base pairs held together by specific Watson-Crick hydrogen bonds like rungs on a twisted ladder (Figure 1.2).<sup>6</sup> The common B-form of DNA is characterized by a wide (12Å) and shallow major groove and a deep and narrow (4-6 Å) minor groove.



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

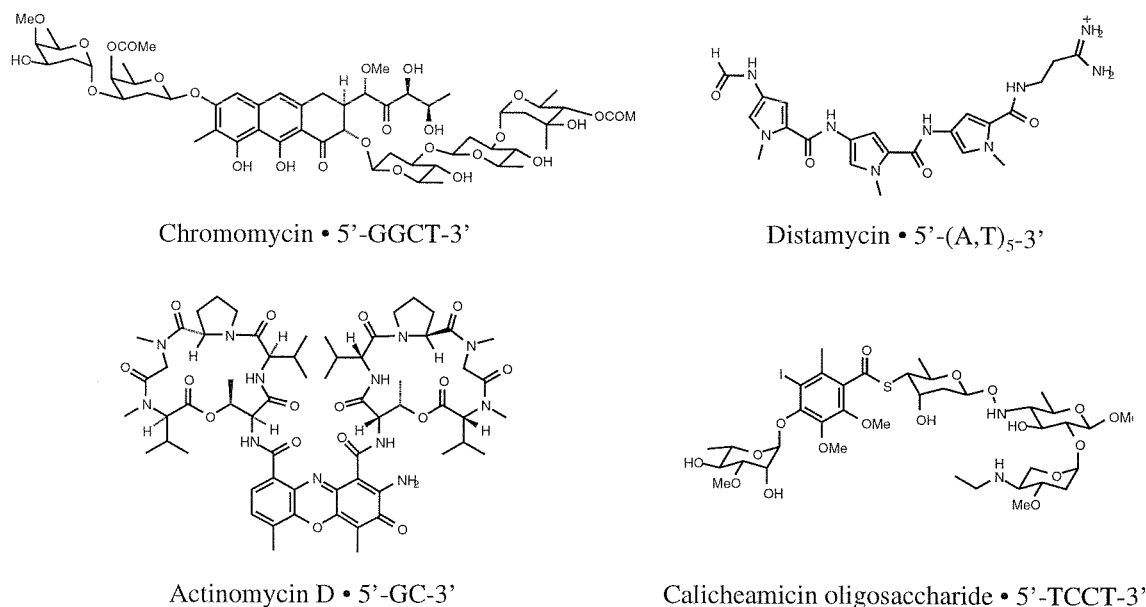


**Figure 1.2.** B-form double helical DNA. Antiparallel strands are indicated in dark and light gray. (left) space filling CPK model, (right) ribbon representation.



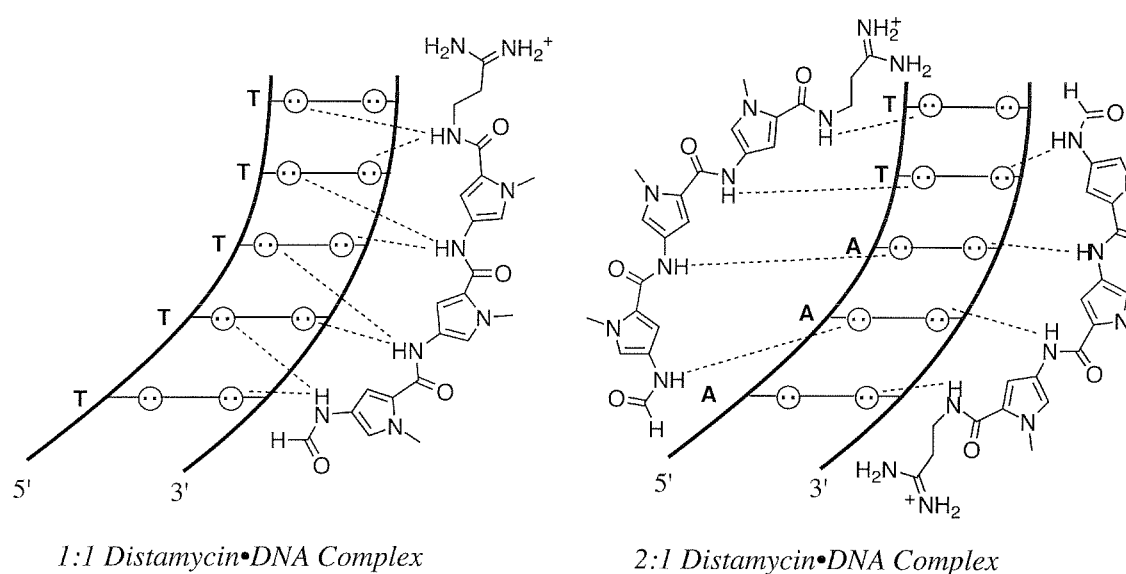
**Figure 1.3.** A schematic model for recognition of the minor groove, with hydrogen bond donors represented as (H) and hydrogen bond acceptors represented as two dots. This schematic underscores the potential coded reading of the DNA helix.

Individual sequences may be distinguished by the pattern of hydrogen bond donors and acceptors displayed on the edges of the base pairs (Figure 1.3).<sup>7</sup> In the minor groove, the A,T base pair presents two symmetrically placed hydrogen bond acceptors in the minor groove, the purine N3 and the pyrimidine O2 atoms. The G,C base pair presents these two acceptors, but in addition presents a hydrogen bond donor, the 2-amino group of guanine.<sup>8</sup>



**Figure 1.4** The structures of four small molecules isolated from natural sources.

Small molecules isolated from natural sources which bind DNA are found to be a structurally diverse class, as evidenced by consideration of four representative molecules, chromomycin, distamycin, actinomycin D, and calicheamicin.<sup>9,10</sup> There is no natural recognition code for the readout of specific sequences of DNA (Figure 1.4). Among these DNA-binding molecules, distamycin is distinguished by its structural simplicity, having no chiral centers and an oligopyrrolecarboxamide core structure.<sup>11</sup> Structural studies of distamycin-DNA complexes reveal modular complexes in which adjacent pyrrolecarboxamides makes similar contacts with adjacent DNA base pairs (Figure 1.5).<sup>10</sup> The relative simplicity of distamycin, with respect both



**Figure 1.5.** A schematic representation of recognition of A,T rich sequences in the minor groove by 1:1 and 2:1 complexes of Distamycin.

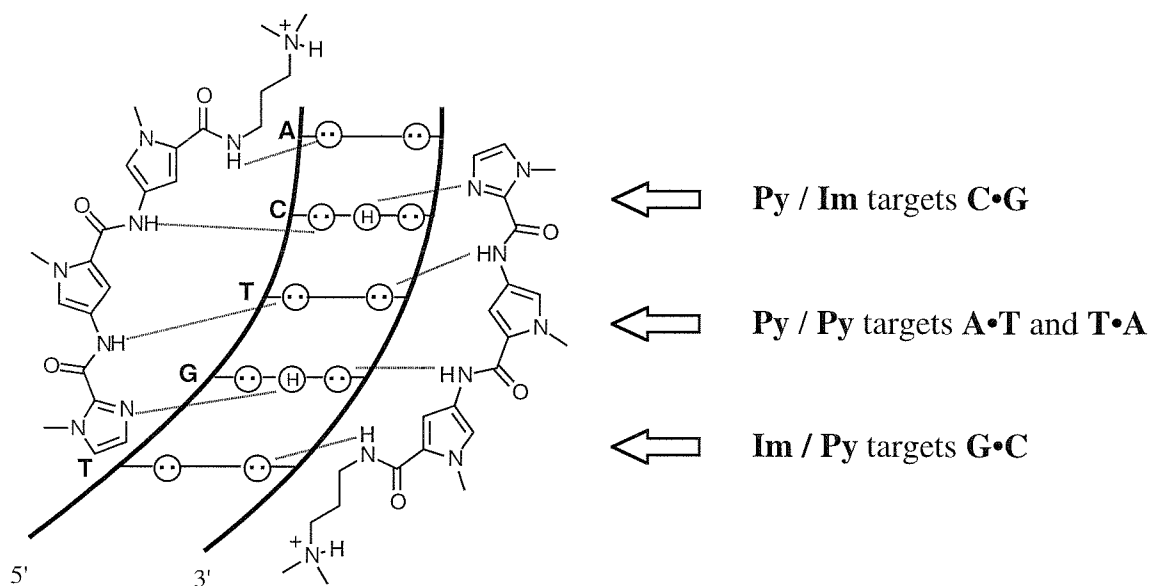
to its chemical structure and its complexes with DNA, guided the initial decision to use distamycin as a basis for designed polyamides having novel DNA-binding sequence specificity.<sup>12</sup>

Two distinct DNA binding modes exist for Distamycin A. In the first binding mode, a single molecule of Distamycin binds in the middle of the minor groove of a 5 base pair A,T rich sequence (Figure 1.5). The amide hydrogens of the *N*-methylpyrrole-carboxamides form bifurcated hydrogen bonds with Adenine N3 and thymine O2 atoms on the floor of the minor



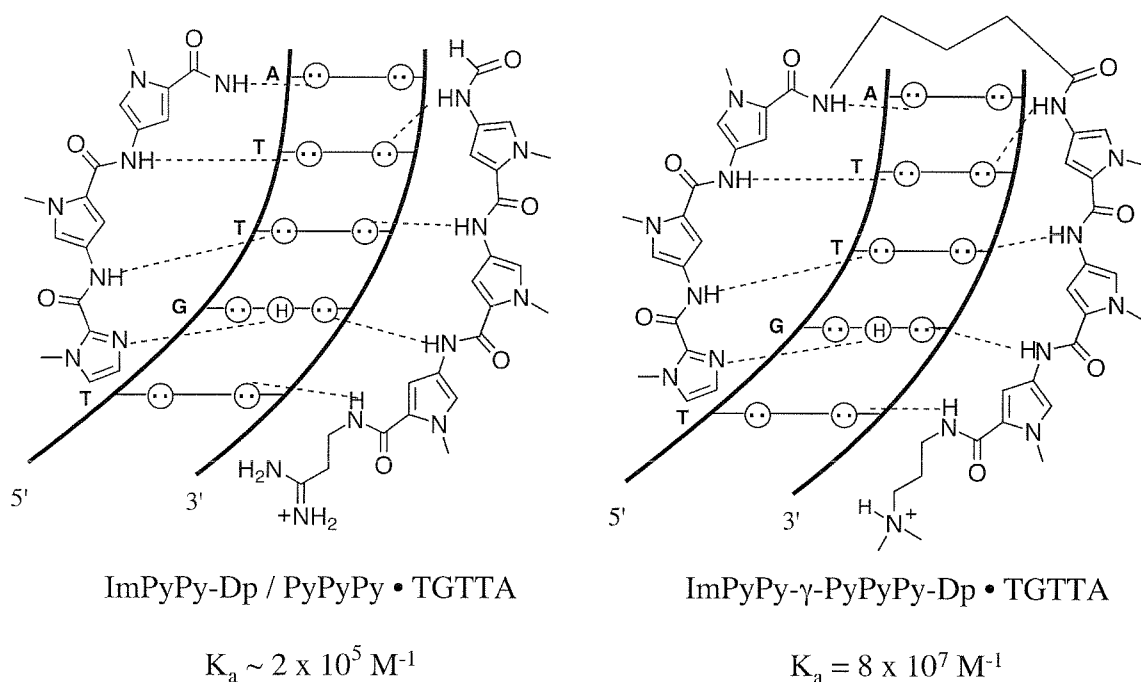
groove.<sup>10</sup> In the second binding mode, two distamycin ligands form an antiparallel side-by-side dimer in the minor groove of a 5 base pair A,T rich site.<sup>13</sup> In the 2:1 model each polyamide subunit forms hydrogen bonds to a unique DNA strand in the minor groove (Figure 1.5).

A twenty year search for a general DNA-recognition code, led by Prof. Peter B. Dervan of Caltech, led to the development of pairing rules to guide polyamide design for recognition of double helical DNA. Polyamides containing *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) amino acids provide a model for the design of artificial molecules for recognition of double helical DNA. For side-by-side complexes of Py/Im-polyamides in the minor groove of DNA, the DNA binding sequence specificity depends on the sequence of side-by-side amino acid pairings.<sup>14</sup> A pairing of Im opposite Py targets a G•C base pair while a pairing of Py opposite Im targets a C•G base pair.<sup>14</sup> A Py/Py combination is degenerate targeting both A•T and T•A base pairs (Figure 1.6).<sup>13,14</sup> Specificity for G,C base pairs results from the formation of a putative hydrogen bond between the imidazole N3 and the exocyclic amine group of guanine. Validity of the pairing rules is supported by a variety of footprinting, NMR, and x-ray structure studies.<sup>15</sup>



**Figure 1.6** A schematic representation of the polyamide pairing rules.

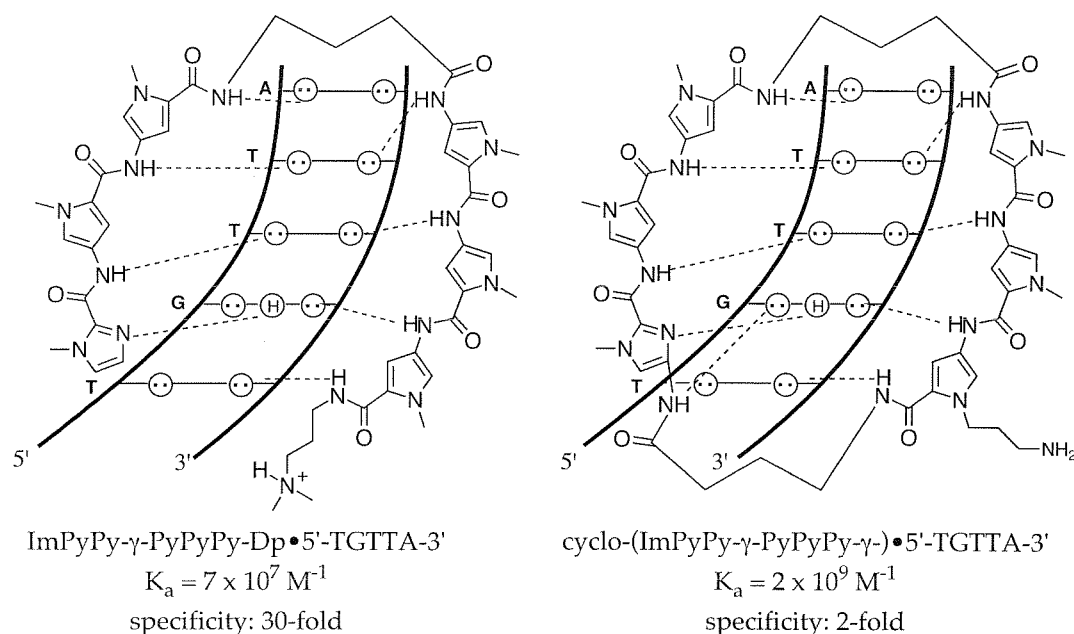
In parallel with the elucidation of the scope and limitations of the pairing rules, efforts have been made to increase the DNA-binding affinity and specificity of pyrrole-imidazole polyamides by covalently linking polyamide subunits.<sup>16,17</sup> A hairpin polyamide motif with  $\gamma$ -aminobutyric acid ( $\gamma$ ) serving as a turn-specific internal-guide-residue provides a synthetically accessible method of linking polyamide subunits within the 2:1 motif (Figure 1.7). The head-to-tail linked polyamide ImPyPy- $\gamma$ -PyPyPy-Dp was shown to specifically bind the designated target site 5'-TGTTA-3' with an equilibrium association constant of  $K_a = 8 \times 10^7 \text{ M}^{-1}$ , an increase of 300-fold relative to the unlinked three-ring polyamide pair ImPyPy and PyPyPy.<sup>18</sup> The hairpin polyamide model is supported by footprinting, affinity cleaving and NMR structure studies.<sup>18,19</sup>



**Figure 1.7** A schematic representation of recognition of a 5'-TGTTA-3' sequence by unlinked subunits (left) and  $\gamma$ -aminobutyric acid linked 'hairpin' subunits (right).

Closing the ends of the hairpin to form a cyclic polyamide increases the overall energetics for DNA-binding presumably by restricting conformational space for the molecule.<sup>17</sup> A cyclic polyamide cyclo-(ImPyPy- $\gamma$ -PyPyPy- $\gamma$ -) was shown to specifically bind the designated

target site 5'-TGTTA-3' with an equilibrium association constant of  $K_a = 2.9 \times 10^9 \text{ M}^{-1}$ , an increase of 40-fold relative to the corresponding hairpin polyamide of sequence composition ImPyPy- $\gamma$ -PyPyPy. The sequence-specificity versus single base pair mismatch sites drops from 30-fold for the hairpin polyamide to 2-fold for the cyclic polyamide.



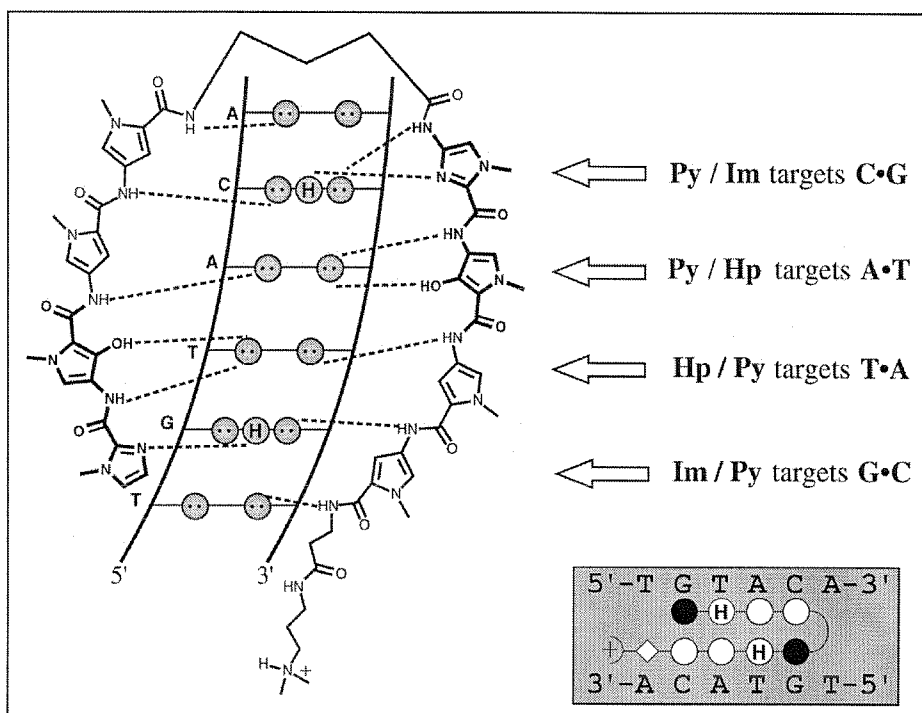
**Figure 1.8** A comparison of cyclic-polyamide and hairpin-polyamide motifs for recognition of the minor groove of DNA.

Despite the design breakthrough in molecular recognition of DNA, the binding affinities of linked and unlinked polyamide dimers of the prior art are modest when compared to those found with natural DNA binding proteins.<sup>20</sup> For example DNA-binding transcription factors recognize their cognate sites at subnanomolar concentrations.<sup>21</sup> 6-ring hairpin polyamides require concentrations greater than 10 nM to occupy their target sites. The only class of polyamides described in the prior art with affinities similar to DNA-binding proteins are the 6-ring cyclic polyamides; however, this class of molecules lacks the sequence-specificity of proteins (i.e., an energetic penalty for binding a single base pair mismatch site) and therefore currently has no potential for therapeutic applications.

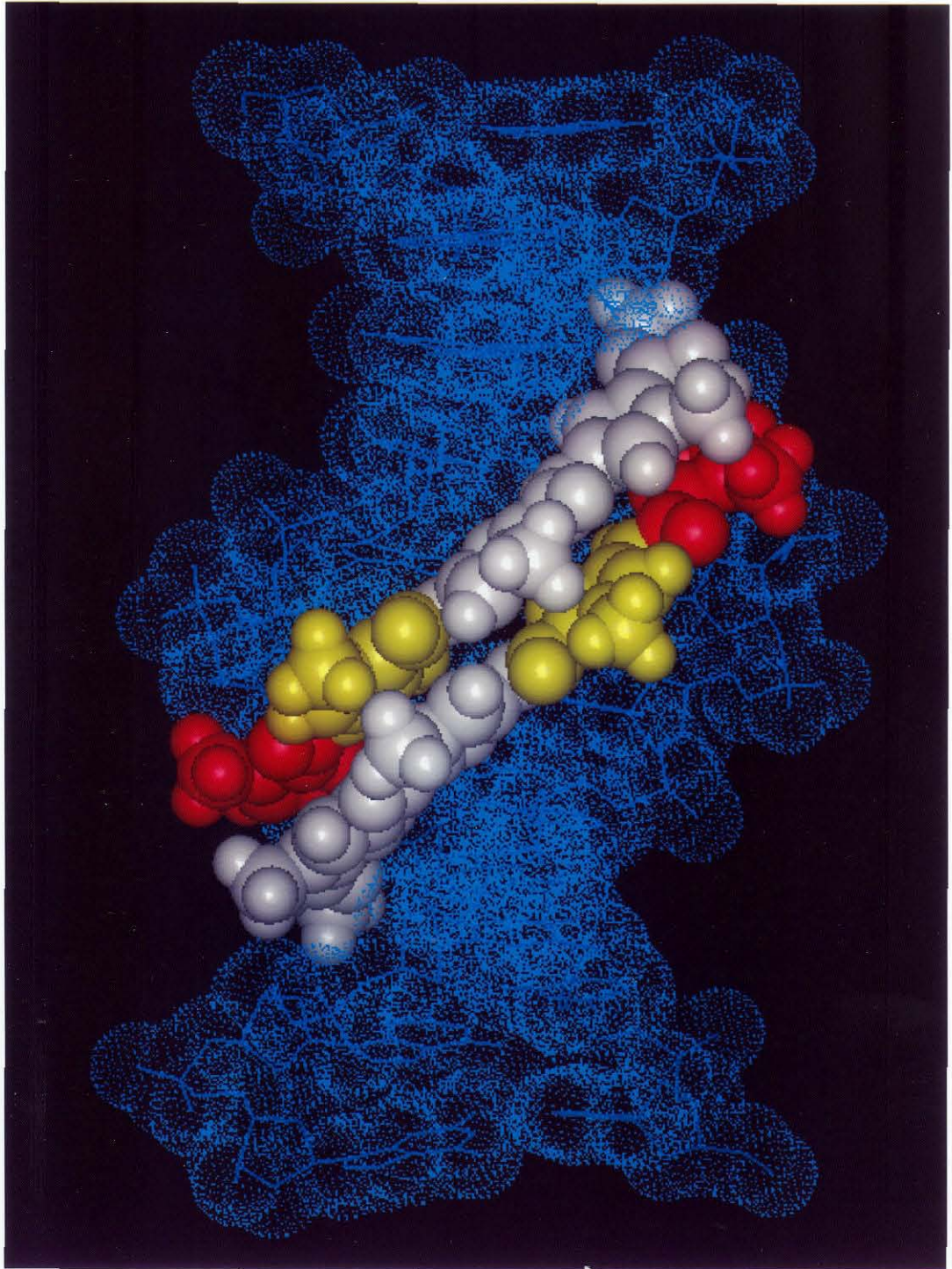
Two prior approaches for the development of synthetic transcriptional antagonists have been reported. Oligodeoxynucleotides which recognize the major groove of double helical DNA via triple helix formation bind a broad sequence repertoire with high affinity and specificity.<sup>22</sup> Although oligonucleotides and their analogs have been shown to interfere with gene expression,<sup>23</sup> the triple helix approach is limited to purine tracks and suffers from poor cellular uptake. There are a few examples of cell-permeable carbohydrate based ligands that interfere with transcription factor function.<sup>24</sup> However, oligosaccharides are not yet amenable to recognition of a broad range of predetermined DNA sequences.

Because of the small size and hydrophobic nature of polyamides (MW  $\approx$  1200) and because the parent ligand Distamycin is itself cell-permeable, these ligands have the potential to underpin a new field of small molecule regulation of gene expression. It remained to be determined if low molecular weight pyrrole-imidazole polyamides could be constructed which recognize predetermined DNA sites at subnanomolar concentrations without compromising sequence-selectivity.

Methodology is described here for the machine-assisted solid phase synthesis of the DNA-binding polyamides developed by Peter Dervan. The new synthetic procedures have enabled biological and biophysical studies of the polyamides by reducing the time required for preparation of polyamides from months to days. Using the solid phase methodology, the Dervan group has recently completed the DNA recognition code (Figure 1.9), preparing polyamides that achieve affinities and specificities comparable to natural DNA-binding proteins. Studies performed in collaboration with the Gottesfeld and Mosier groups at the Scripps Research Institute have demonstrated that these synthetic ligands inhibit HIV-1 transcription in cell-free assays and virus replication in isolated human peripheral blood lymphocytes.



**Figure 1.9.** (Above) Small molecule polyamide, reads each of the four base-pairs in the DNA minor groove. (Next page) Space filling representation of the polyamide dimer ImHpPyPy- $\beta$ -Dp bound in the minor groove of DNA. The figure was prepared using InsightII software and is derived from a high-resolution x-ray co-crystal structure of a polyamide dimer bound to DNA which was obtained in collaboration with the Rees group at the California Institute of Technology.



## References

1. Watson, J.D. *Gene* **1993**, 135, 309.
2. Roeder, R.G. *TIBS* **1996**, 9, 327.
3. Tjian, R. *Sci. Am.* **1995**, 2, 54.
4. Gottesfeld, J.M.; Neely, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1997**, 387, 202.
5. Watson, J.D.; Crick, F.H.C. *Nature* **1953**, 171, 737.
6. Dickerson, R.E.; Drew, H.R.; Conner, B.N.; Wing, M.; Fratini, A.V.; Koopka, M.L. *Science* **1982**, 216, 475.
7. Principles of Nucliec Acid Stucture Saenger, W.; Springer-Verlag, New York, 1984.
8. Steitz, T.A. *Quart. Rev. Biophys.* **23**, 205.
9. (a) Gao, X.L.; Mirau, P.; Patel, D.J. *J. Mol. Biol.* **1992**, 223, 259; (b) Kamitori, S.; Takusagawa, F. *J. Mol. Biol.* **1992**, 225, 445; (c) Paloma, L.G.; Smith, J.A.; Chazin, W. J.; Nicolaou, K.C. *J. Am. Chem. Soc.* **1994**, 116, 3697.
10. Coll, M.; Frederick, C.A.; Wang, A.H.J.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, 84, 8385.
11. (a) Zimmer, C. *Prog. Nucleic Acid Res.* **1975**, 15, 285; (b) Baguley, B.C. *Molecular and Cellular Biochem.* **1982**, 43, 167; (c) Zimmer, C.; Wahnert, U, *Prog. Biophy. Mol. Biol.* **1986**, 47, 31.
12. Dervan, P.B. *Science* **1986**, 232, 464.
13. (a) Pelton, J.G.; Wemmer, D.E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, 86, 5723; (b) Pelton, J. G.; Wemmer, D.E. *J. Am. Chem. Soc.* **1990**, 112, 1393; (c) Keilkopf, C.L.; Baird, E.E.; Dervan, P.B.; Rees, D.C. *Nature Struct. Biol.* **1998**, 5, 104.
14. (a) Wade, W.S.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1992**, 114, 8783; (b) Mrksich, M., Wade, W.S., Dwyer, T.J., Geirstanger, B.H., Wemmer, D.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 7586; (c) Wade, W.S., Mrksich, M.; Dervan, P.B. *Biochemistry* **1993**, 32, 11385.
15. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, 115, 2572; (b) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, 266, 646; (c) Mrksich, M.; Dervan, P.B.; *J. Am. Chem. Soc.* **1995**, 117, 3325.
16. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, 115, 9892; (b) Dwyer, T.J.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, 115, 9900. (c) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, 116, 3663; (d) Chen,

- 
- Y.H.; Lown, J.W. *J. Am. Chem. Soc.* **1994**, *116*, 6995; (e) Alsaid, N.H.; Lown, J.W. *Tett. Lett.* **1994**, *35*, 7577; (f) Alsaid, N.H.; Lown, J.W. *Synth. Comm.* **1995**, *25*, 1059; (g) Chen, Y.H.; Lown, J.W. *Biophys. J.* **1995**, *68*, 2041; (h) Chen, Y.H.; Liu, J.X.; Lown, J.W. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 2223; (i) Chen, Y.H.; Yang, Y.W.; Lown, J.W. *J. Biomol. Struct. Dyn.* **1996**, *14*, 341; (j) Singh, M.P.; Wylie, W.A.; Lown, J.W. *Mag. Res. Chem.* **1996**, *34*, S55.
17. Cho, J.; Parks, M.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.*, **1995**, *92*, 10389.
  18. (a) Mrksich, M.; Parks, M.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 7983; (b) Parks, M.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6147; (c) Parks, M.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6153; (d) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Chem. & Biol.* **1996**, *3*, 369; (e) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 8198; (f) Pilch, D.S.; Poklar, N.A.; Gelfand, S.A.; Law, S.M.; Breslauer, K.J.; Baird, E.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306; (h) White, S.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, *119*, 8756. (i) White, S.; Baird, E.E.; Dervan, P.B. *Chem & Biol.* **1997**, *4*, 569.
  19. de Clairec, R.P.L.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1997**, *119*, 7909.
  20. Clemens, K.R.; Zhang, P.H.; Liao, X.B.; McBryant, S.J.; Wright, P.E.; Gottesfeld, J.M. *J. Mol. Biol.* **1994**, *244*, 23.
  21. (a) Jamieson, A.C.; Kim, S.H.; Wells, J.A. *Biochemistry* **1994**, *33*, 5689; (b) Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11168; (c) Greisman, H.A.; Pabo, C.O. *Science* **1997**, *275*, 657.
  22. (a) Moser, H.E.; Dervan, P.B. *Science* **1987**, *238*, 645; (b) Thuong, N.T.; Helene, C. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 666.
  23. (a) Maher, J.L.; Dervan, P.B.; Wold, B. *Biochemistry* **1992**, *31*, 70; (b) Duvalvalentin, G.; Thuong, N.T.; Helene, C. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 504.
  24. (a) Ho, S.N.; Boyer, S. H.; Schreiber, S.L.; Danishefsky, S. J.; Crabtree, G. R. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9203; (b) Liu, C. *et al. Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 940.



## CHAPTER 2

# Solid Phase Synthesis of DNA-Binding Polyamides Containing Aromatic Amino Acids

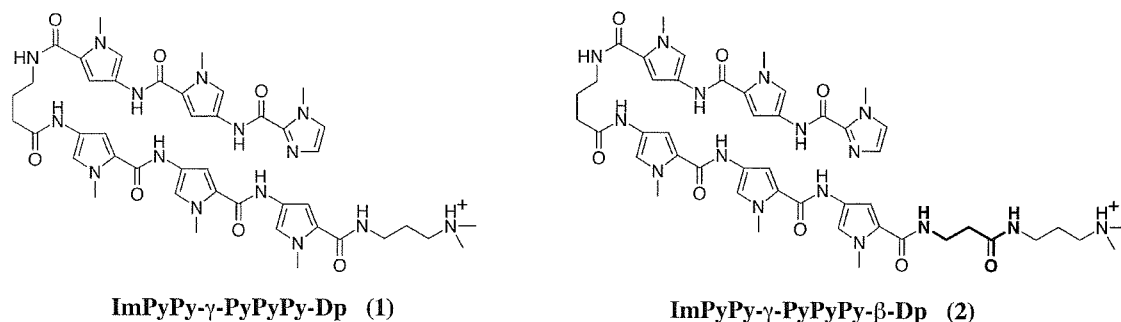
**Abstract:** *The solid phase synthesis of sequence specific DNA binding polyamides containing N-methylimidazole (Im) and N-methylpyrrole (Py) amino acids is described. Two monomer building blocks, Boc-Py-OBt ester and Boc-Im acid, are prepared on a 50 g scale without column chromatography. Using commercially available Boc-β-alanine-Pam resin, cycling protocols were optimized to afford high stepwise coupling yields (> 99%). Deprotection by aminolysis affords up to 100 milligram quantities of polyamide. Solid phase methodology increases both the number and complexity of minor groove binding polyamides which can be synthesized and analyzed with regards to DNA binding affinity and sequence specificity. The solid phase synthesis of a representative eight-residue polyamide is reported.*

**Publication:** Baird & Dervan *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.

**Introduction.** Recent efforts to discover a universal set of simple chemical rules for the digital readout of double helical DNA by artificial molecules have met with encouraging success.<sup>1,2</sup> The natural products netropsin and distamycin A are crescent shaped polyamides that bind in the minor groove of DNA at sites of at least four successive A•T base pairs.<sup>3</sup> Efforts to design distamycin analogs *specific* for G,C containing sequences were completely unsuccessful<sup>4</sup> until the recent discovery that polyamides containing *N*-methylimidazole (Im) and *N*-methylpyrrole (Py) amino acids can be combined in antiparallel side-by-side dimeric complexes with the minor groove of DNA.<sup>2,5,6</sup> The DNA-binding sequence-specificity of these small molecules can be controlled by the linear sequence of Py and Im amino acids. An Im ring on one ligand complemented by a Py ring on the second ligand recognizes a G•C base pair, while a Py/Im combination targets a C•G base pair.<sup>2,6</sup> A Py/Py pair is degenerate for A•T/T•A base pairs.<sup>2,5,6</sup> The utility of the 2:1 model is demonstrated by the four-ring polyamide ImPyImPy-Dp (Dp = dimethylaminopropylamine) which binds the four base pair core sequence 5'-GCGC-3', a complete reversal of the natural specificity of netropsin and distamycin.<sup>7</sup>

Covalently linking polyamide heterodimers and homodimers within the 2:1 motif has led to designed ligands with both increased affinity and specificity.<sup>8,9</sup> A polyamide "hairpin" motif with  $\gamma$ -aminobutyric acid ( $\gamma$ ) serving as a "turn monomer" provides a synthetically accessible method of linking polyamide units within the 2:1 motif.<sup>9</sup> The polyamide ImPyPy- $\gamma$ -PyPyPy-Dp (**1**) was found to bind the designated target site 5'-TGTTA-3' with a 300-fold binding enhancement over the individual unlinked polyamide pair; ImPyPy and PyPyPy.<sup>9</sup>

While exploring the limits of the 2:1 model for the design of polyamides for the recognition of any sequence of any site size, the synthetic effort emerged as a limiting step. The process of expanding the 2:1 motif to include longer sequences recognized by increasingly complex polyamides is demanding. For example, using previously described multi-step solution phase chemistry, the total synthesis of hairpin polyamides such as ImPyPy- $\gamma$ -PyPyPy-Dp **1** and ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **2** ( $\beta$  =  $\beta$  - alanine) would require more than a month's effort for each polyamide (Figure 2.1). We report here general protocols for manual and machine-assisted Boc-

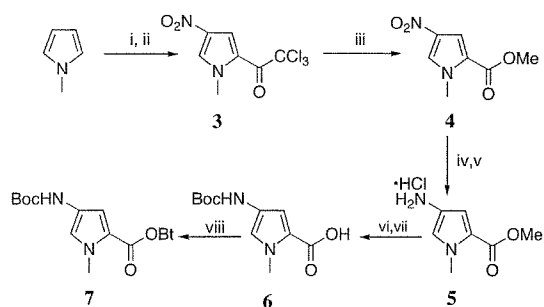


**Figure 2.1.** (Left) ImPyPy-γ-PyPyPy-Dp prepared by multi-step solution phase synthesis and the solid phase analog (Right) ImPyPy-γ-PyPyPy-β-Dp containing a C-terminal β-alanine residue to facilitate synthesis.

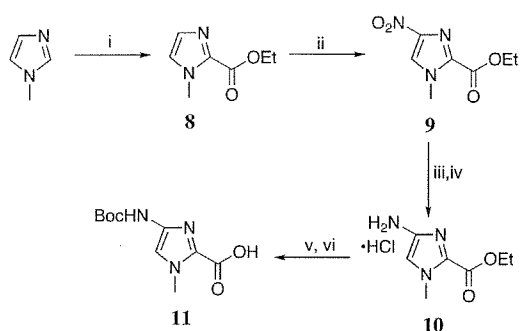
chemistry solid phase synthesis of the pyrrole-imidazole polyamides that reduce this synthetic investment from months to days.

**Synthesis of Pyrrole- Imidazole Polyamides.** Distamycin and its analogs have previously been considered targets of traditional multi-step synthetic organic chemistry,<sup>10</sup> not solid phase synthesis. This is because the repeating amide of distamycin is formed from an aromatic carboxylic acid and an aromatic amine, both of which have proven problematic for solution phase coupling reactions. The aromatic acid is often unstable to decarboxylation and the aromatic amines have been found to be air and light sensitive.<sup>11</sup> It seemed likely that the variable coupling yields, long reaction times (often > 24 h), numerous side products, and reactive intermediates (acid chlorides and trichloroketones) characteristic of the traditional solution phase coupling reactions would make the solid phase synthesis of the aromatic carboxamides problematic.<sup>12</sup> Therefore, it might not be possible to achieve the high yields required to prevent failure sequences from accumulating, precluding either the synthesis or purification of the desired polyamides.

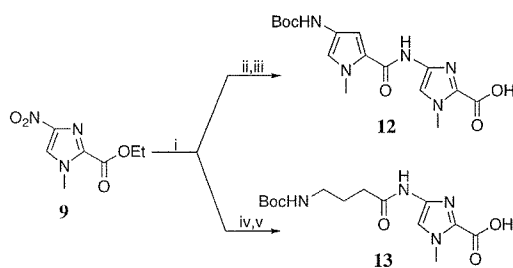
**Solid Phase Synthesis.** The solid phase approach has been successfully developed for the synthesis of a variety of proteins,<sup>13</sup> oligonucleotides,<sup>14</sup> peptoids,<sup>15</sup> oligosaccharides,<sup>16</sup> and small non-polymeric molecules.<sup>17</sup> In order to implement an efficient solid phase methodology for the synthesis of the pyrrole-imidazole polyamides the following components were developed: (1) a synthesis which provides large quantities of appropriately protected monomer or dimer building



**Figure 2.2.** (a) (i) trichloroacetyl chloride, ethyl ether; (ii) 90% nitric acid,  $\text{Ac}_2\text{O}$ ; (iii)  $\text{NaOMe}/\text{MeOH}$ ; (iv) 1 atm of  $\text{H}_2$ , 10%  $\text{Pd}/\text{C}$ ,  $\text{EtOAc}$ ; (v)  $\text{HCl}$ , ethyl ether; (vi) 10%  $\text{K}_2\text{CO}_3$  (aq), Boc-anhydride; (vii)  $\text{NaOH}$ ,  $\text{MeOH}$ , water,  $60^\circ\text{C}$ ; (viii)  $\text{DCC}$ ,  $\text{HOBt}$ ,  $\text{DMF}$ .



**Figure 2.3.** (i) ethyl chloroformate, TEA, CH<sub>3</sub>CN, -20°C; (ii) H<sub>2</sub>SO<sub>4</sub> (conc.), 90% nitric acid; (iii) 1 atm of H<sub>2</sub>, 10% Pd/C, EtOAc/EtOH; (iv) HCl, ethyl ether; (v) Boc-anhydride, DIEA, DMF, 40°C; (vi) 1M NaOH (aq.).

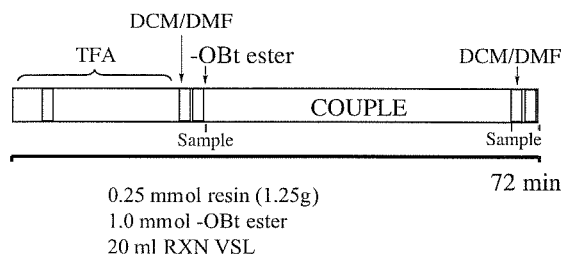


**Figure 2.4.** (i) 500 psi H<sub>2</sub>, 10% Pd/C, DMF; (ii) Boc-Pyrrole-acid (activated *in situ* with DCC/HOBt), DIEA, DMF, 40°C; (iii) NaOH, MeOH, water, 60°C; (vi) Boc-γ-aminobutyric acid (activated *in situ* with DCC/HOBt), DIEA, DMF, 40°C; (vii) NaOH, MeOH, water, 60°C.

blocks in high purity, (2) optimized protocols for forming an amide in high yield from a support bound aromatic amine and an aromatic carboxylic acid, (3) methods for monitoring reactions on the solid support, (4) a stable resin linkage agent that can be cleaved in high yield upon completion of the synthesis.

**Monomer Syntheses.** The synthesis of Boc-Py-OBt ester **7**<sup>18</sup> and Boc-Im acid **11**<sup>19</sup> has been previously described. Available procedures<sup>18-23</sup> provide only milligram to gram quantities of monomer while requiring difficult column chromatography and the use of toxic chlorofluorophosgene for introduction of the Boc group. An optimized synthesis, using inexpensive starting materials, has been developed allowing Boc-Py-OBt ester **7** and Boc-Im acid **11** monomers to be prepared on 50 g scale without the use of column chromatography (Figure 2.2 and 2.3). Two dimeric building blocks have also been prepared, Boc-Py-Im acid **12** and Boc- $\gamma$ -Im acid **13** (Figure 2.4).

**Resin Linkages.** For solid phase synthesis, the polyamide is attached to an insoluble matrix by a linkage that is cleaved by a single



**Figure 2.5.** Protocol for the solid phase synthesis of a pyrrole-imidazole polyamide.

step process that introduces a positive charge into the polyamide.

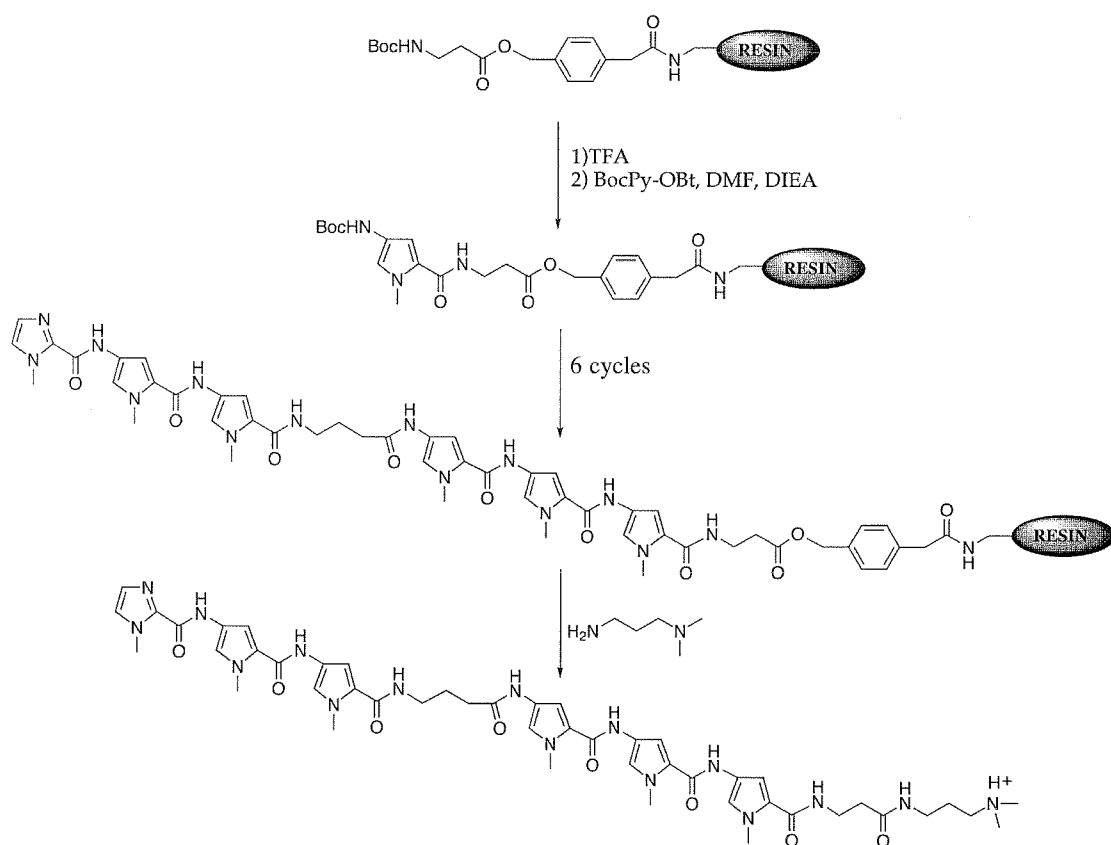
The addition of an aliphatic amino acid at the C-terminus of the pyrrole-imidazole polyamides allows the use of Boc- $\beta$ -alanine-Pam-Resin resin which is

commercially available (Available in bulk from: Peptides International, Louisville, Kentucky) in appropriate substitution levels (0.2 mmol/gram) (Figure 2.1).<sup>24</sup> Aminolysis of the resin ester linkage provides a simple and efficient method for cleaving the polyamide from the support. Attempts to isolate polyamides by direct aminolysis of Pyrrole-Pam resin were unsuccessful. The DNA binding affinity and sequence-specificity of polyamides containing a C-terminal  $\beta$ -alanine spacer such as ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **2** are not greatly altered from solution phase polyamides such as ImPyPy- $\gamma$ -PyPyPy-Dp **1**. Detailed thermodynamic characterization is described elsewhere (see Chapter 2).<sup>25</sup>

**Table 1.1.** Standard Protocol for Manual Solid Phase Synthesis of Pyrrole- Imidazole Polyamides.

	Synthesis Cycle	Time/mode
1) Deprotect	80% TFA/DCM, 0.5M PhSH	1 min shake
		30 s flow
		20 min shake
2) Wash	DCM	1 min flow
	DMF	30 s flow
	(take resin sample for monitoring)	
3) Couple	-OBt ester, DIEA	45 min shake
	(take resin sample for monitoring)	
4) Wash	DMF	2 x 30 s flow
	DCM	30 s flow

**Solid Phase Polyamide Synthesis Protocols.** Solid phase polyamide synthesis protocols were modified from the *in situ* neutralization Boc-chemistry protocols recently reported by Kent and coworkers.<sup>26</sup> Coupling cycles are rapid, 72 min. per residue for manual synthesis or 180 min. per residue for machine-assisted synthesis, and require no special precautions beyond those used for ordinary solid phase peptide synthesis. Manual solid phase synthesis of a pyrrole-imidazole polyamide consists of a dichloromethane (DCM) wash, removal of the Boc group with trifluoroacetic acid (TFA)/DCM/thiophenol (PhSH), a DCM wash, a DMF wash, taking a resin sample for analysis, addition of activated monomer, addition of DIEA if necessary, coupling for



**Figure 2.6.** Solid phase synthetic scheme for ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp starting from commercially available Boc- $\beta$ -Pam-resin: (i) 80% TFA/DCM, 0.4M PhSH; (ii) BocPy-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4M PhSH; (iv) BocPy-OBt, DIEA, DMF; (v) 80% TFA/DCM, 0.4M PhSH; (vi) BocPy-OBt, DIEA, DMF; (vii) 80% TFA/DCM, 0.4M PhSH; (viii) Boc- $\gamma$ -aminobutyric acid (HBTU, DIEA); (ix) 80% TFA/DCM, 0.4M PhSH; (x) BocPy-OBt, DIEA, DMF; (xi) 80% TFA/DCM, 0.4M PhSH; (xii) BocPy-OBt, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4M PhSH; (xiv) Imidazole-2-carboxylic acid (HBTU/DIEA); (xv) *N,N*-dimethylaminopropyl-amine, 55°C.

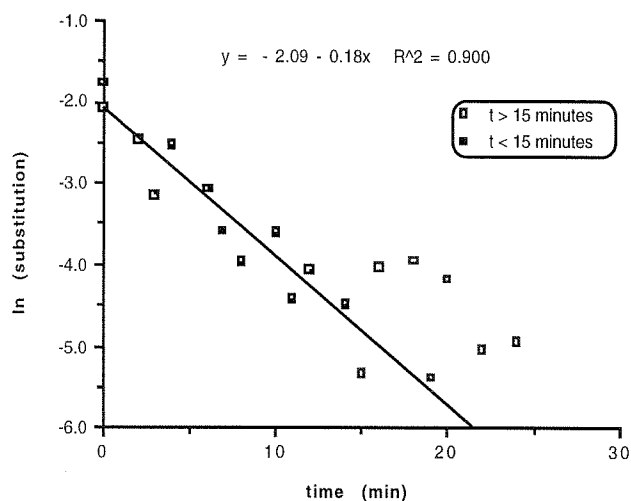
45 min, taking a resin sample for analysis, and a final DMF wash (Figure 2.5, Table I). In addition, the manual solid phase protocol for synthesis of pyrrole-imidazole polyamides has been adapted for use on a ABI 430A peptide synthesizer.

**Monitoring the synthesis.** The aromatic amine of the pyrrole and imidazole do not react in the quantitative ninhydrin test.<sup>27</sup> Stepwise cleavage of a sample of resin and analysis by HPLC indicates that high stepwise yields (> 99%) are routinely achieved. We note that acylation of imidazole amine with Boc-Py-OBt ester was not satisfactory. However, acylation with Boc-Py symmetrical anhydride/DMAP ester (DCC, DMAP, DCM) proceeds to completion within 3 h. Alternatively, the preparation of a Boc-PyIm-OBt dimer unit avoids the difficult coupling of pyrrole to imidazole.

**Synthesis of Eight Residue Polyamide 2.** ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp was prepared in 14 steps using the protocols described in the Experimental Section (Figure 2.6). The yield of each individual coupling step was established as >98% by HPLC analysis. The resin was cleaved in high yield (> 90%) by aminolysis with *N,N*-dimethylaminopropylamine. A single HPLC separation of the eight residue polyamide was sufficient to obtain a final purity greater than 98% as determined by a combination of analytical HPLC, <sup>1</sup>H NMR and mass spectroscopy.

**Rates and Efficiency of Coupling Reactions.** Under the standard coupling conditions the efficiency of coupling reactions appears to be as follows (activated ester/free amine). Py/Gly = Im/Gly > Gly/Py > Im/Py > Py/Py > Gly/Im > Im/Im > Py/Im. All couplings except for Im/Im and Py/Im are > 99.8% complete in the recommended 42 min. coupling times. The faster couplings are more than 99.8% complete within 5 min. For the Im/Im and Py/Im couplings, extended reaction times are recommended in order to assure complete reaction. Coupling rates are estimated based on picric acid titration data, and ninhydrin tests when possible.

No correction is made for the change in substitution of the resin resulting from the addition of a monomer, because the effect is very small for the low substitution resins used for synthesis here. For the Py/Py, Im/Py, Py/Gly, and Gly/Py couplings it was attempted to measure rates using the picric acid test at 1 min. time intervals. The Im/Py, Py/Gly, and Gly/Py couplings

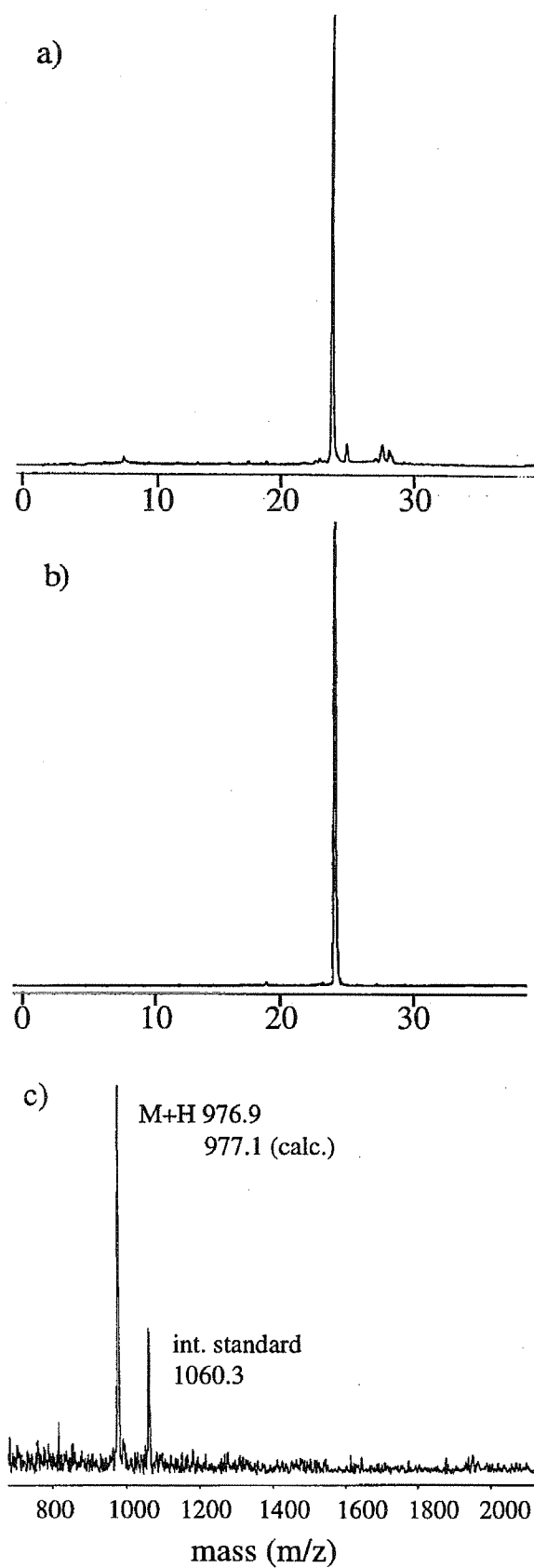


**Figure 2.7.** Monitoring the time course of coupling Boc-Py-OBt to  $\text{PyNH}_2$  by Picric acid titration. Linear curve fit through data obtained during the first 15 min. of the coupling reaction.

all reached completion too rapidly to measure an accurate rate. For the Py/Py coupling, reasonably accurate data was obtained for monitoring the disappearance of amine. From the slope of a plot of  $\ln(\text{meq. amine})$  v time, it is possible to estimate a rate of reaction of  $0.18\text{h}^{-1}$  which corresponds to a 3.9 min. half life, and indicates that 25.6 min. are required for 99% reaction, and 38.4 min. for 99.9% reaction. The 45 min. coupling time recommended here is chosen to be conservative. The short length of the peptides synthesized here, combined with an inaccurate method of monitoring coupling yields beyond 90% (as can clearly be seen from the scatter in the picric acid data after 15 min. of reaction), both support the use of long reaction times.

**Conclusion.** Pyrrole-imidazole polyamide-DNA complexes provide a potentially general model for the design of non-natural ligands for the sequence-specific recognition of the minor groove of DNA. The large number of polyamides made available by solid phase synthetic methodology should accelerate the elucidation of the scope and limitations of this approach.





**Figure 2.8.** Characterization of the synthesis of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp. (a) Analytical HPLC analysis of crude synthesis products from the machine-assisted solid phase synthesis, (b) Analytical HPLC analysis of purified polyamide, C18 chromatography, 0.1% (wt/v) TFA, gradient elution 1.25% CH<sub>3</sub>CN/min monitored at 254 nm. (c) MALDI-TOF mass spectral analysis of purified polyamide.

## Experimental

**Materials.** Boc- $\beta$ -alanine-(4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc- $\beta$ -Pam-Resin), dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBt), 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU), Boc-glycine, and Boc- $\beta$ -alanine were purchased from Peptides International. *N,N*-diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF), *N*-methylpyrrolidone (NMP), and DMSO/NMP were purchased from Applied Biosystems. Boc- $\gamma$ -aminobutyric acid was from NOVA Biochem, dichloromethane (DCM) and triethylamine (TEA) was reagent grade from EM, thiophenol (PhSH), dimethylaminopropylamine, trichloroacetyl chloride, *N*-methylpyrrole, and *N*-methylimidazole from Aldrich, and trifluoroacetic acid (TFA) from Halocarbon. All reagents were used without further purification.

$^1\text{H}$  NMR were recorded on a GE 300 instrument operating at 300 MHz. Chemical shifts are reported in ppm relative to the solvent residual signal. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. IR spectra were recorded on a Perkin-Elmer FTIR spectrometer. High-resolution FAB mass spectra were recorded at the Mass Spectroscopy Laboratory at the University of California, Riverside. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed either on a HP 1090M analytical HPLC or a Beckman Gold system using a RAINEN  $\text{C}_{18}$ , Microsorb MV, 5 $\mu\text{m}$ , 300 x 4.6 mm reversed phase column in 0.1% (wt/v) TFA with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% acetonitrile/min. Preparatory HPLC was carried out on a Beckman HPLC using a Waters DeltaPak 25 x 100 mm, 100 $\mu\text{m}$   $\text{C}_{18}$  column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. 18M $\Omega$  water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 $\mu\text{m}$  filtered. Thin-layer chromatography (TLC) was performed on silica gel 60 F<sub>254</sub> precoated plates. Reagent-grade chemicals were used unless otherwise stated.

## Monomer Syntheses

**4-Nitro-2-trichloroacetyl-1-methylpyrrole (3).** To a well stirred solution of trichloroacetyl chloride (1 kg, 5.5 mole) in 1.5 liter ethyl ether in a 12 liter flask was added dropwise over a period of 3 h a solution of *N*-methylpyrrole (0.45 kg, 5.5 mole) in 1.5 liter anhydrous ethyl ether. The reaction was stirred for an additional 3 hours and quenched by the dropwise addition of a solution of 400 g potassium carbonate in 1.5 liters water. The layers were separated and the ether layer concentrated *in vacuo* to provide 2-(trichloroacetyl)pyrrole (1.2 kg, 5.1 mol) as a yellow crystalline solid sufficiently pure to be used without further purification. To a cooled (-40°C) solution of 2-(trichloroacetyl) pyrrole (1.2 kg, 5.1 mol) in acetic anhydride (6 L) in a 12 L flask equipped with a mechanical stirrer was added 440 mL fuming nitric acid over a period of 1 hour while maintaining a temperature of (-40°C). The reaction was carefully allowed to warm to room temperature and stir an additional 4 h. The mixture was cooled to -30°C, and isopropyl alcohol (6 L) added. The solution was stirred at -20°C for 30 min during which time a white precipitate forms. The solution was allowed to stand for 15 min and the resulting precipitate collected by vacuum filtration to provide **3** (0.8 kg, 54% yield) TLC (7:2 benzene/ethyl acetate) R<sub>f</sub> 0.7; <sup>1</sup>H NMR (DMSO-d<sub>6</sub>) δ 8.55 (d, 1 H, *J* = 1.7 Hz), 7.77 (d, 1 H, *J* = 1.7 Hz), 3.98 (s, 3 H); <sup>13</sup>C NMR (DMSO-d<sub>6</sub>) δ 173.3, 134.7, 133.2, 121.1, 116.9, 95.0, 51.5; IR (KBr) 1694, 1516, 1423, 1314, 1183, 1113, 998, 750. FABMS *m/e* 269.936 (M + H 269.937 calc. for C<sub>7</sub>H<sub>5</sub>N<sub>2</sub>O<sub>3</sub>Cl<sub>3</sub>).

**Methyl 4-nitropyrrole-2-carboxylate (4).** To a solution of **3** (800 g, 2.9 mol) in 2.5 L methanol in a 4 L Erlenmeyer flask equipped with a mechanical stirrer was added dropwise a solution of NaH (60% dispersion in oil) (10g, 0.25 mol) in 500 mL methanol. The reaction was stirred 2 h. at room temperature, and quenched by the addition of conc. sulfuric acid (25 mL). The reaction was then heated to reflux, allowed to slowly cool to room temperature as **4** crystallizes as white needles which were collected by vacuum filtration and dried *in vacuo*. (450 g, 47% yield). TLC (ethyl acetate) R<sub>f</sub> 0.8; <sup>1</sup>H NMR (DMSO-d<sub>6</sub>) δ 8.22 (d, 1 H, *J* = 1.7 Hz), 7.22 (d, 1 H, *J* = 1.6 Hz), 3.88 (s, 3 H), 3.75 (s, 3 H); <sup>13</sup>C NMR (DMSO-d<sub>6</sub>) δ 37.8, 52.2, 112.0, 123.0, 129.9, 134.6, 160.3;

IR(KBr) 3148, 1718, 1541, 1425, 1317, 1226, 1195, 1116, 753. FABMS  $m/e$  184.048 (M + H 184.048 calc. for  $C_7H_8N_2O_4$ ).

**Methyl 4-amino-1-methyl-pyrrole-2-carboxylate hydrochloride (5).** Methyl-4-nitropyrrole-2-carboxylate **4** (450g, 2.8 mol) was dissolved in ethyl acetate (8 L). A slurry of 40 g of 10% Pd/C in 800 mL ethyl acetate was then added and the mixture stirred under a slight positive pressure of hydrogen (c.a. 1.1 atm) for 48 h. Pd/C was removed by filtration through Celite, washed 1 x 50 mL ethyl acetate, and the volume of the mixture reduced to c.a. 500 mL. 7 L of cold ethyl ether was added and HCl gas gently bubbled through the mixture. The precipitated amine hydrochloride was then collected by vacuum filtration to yield (380 g, 81.6%) of **5** as a white powder. TLC (ethyl acetate) Rf(amine) 0.6, Rf salt (0.0),  $^1H$  NMR (DMSO- $d_6$ )  $\delta$  10.23 (br s, 3H), 7.24 (d, 1H  $J$  = 1.9), 6.79 (d, 1H,  $J$  = 2.0), 3.83 (s, 3H), 3.72 (s, 3H)  $^{13}C$  NMR (DMSO- $d_6$ )  $\delta$  160.8, 124.3, 121.2, 113.4, 112.0, 51.8, 37.1; IR (KBr) 3095, 2693, 1709, 1548, 1448, 1266, 1102, 802, 751. FABMS  $m/e$  154.075 (154.074 calc. for  $C_7H_{10}N_2O_2$ ).

**4-[(*tert*-Butoxycarbonyl)amino]-1-methylpyrrole-2-carboxylic acid (6).** The hydrochloride salt of the pyrrole amine **5** (340 g, 1.8 mol) was dissolved in 1 L of 10% aqueous sodium carbonate in a 3 L flask equipped with a mechanical stirrer, di-*t*-butyldicarbonate (400 g, 2.0 mmol) slurried in 500 mL of dioxane was added over a period of thirty min maintaining a temperature of 20°C. The reaction was allowed to proceed for three h and was determined complete by TLC, cooled to 5°C for 2 h and the resulting white precipitate collected by vacuum filtration. The Boc-pyrrole ester contaminated with Boc-anhydride was dissolved in 700 mL MeOH, 700 mL of 2M NaOH was added and the solution heated at 60°C for 6 h. The reaction was cooled to room temperature, washed with ethyl ether (4 x 1000 mL), the pH of the aqueous layer reduced to c.a. 3 with 10% (v/v)  $H_2SO_4$ , and extracted with ethyl acetate (4 x 2000 mL). The combined ethyl acetate extracts were dried (sodium sulfate) and concentrated *in vacuo* to provide a tan foam. The foam was dissolved in 500 mL of DCM and 2 L petroleum ether added, the resulting slurry was concentrated *in vacuo*. The reaction was redissolved and concentrated three additional times to provide (320 g, 78% yield) of **6** as a fine white powder. TLC (7:2 benzene/ethyl acetate v/v) Rf

(ester) 0.8, Rf (acid) 0.1. (ethyl acetate), Rf (acid) 0.6,  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  12.10 (s, 1H), 9.05 (s, 1H), 7.02 (s, 1H), 6.55 (s, 1H), 3.75 (s, 3H), 1.41 (s, 9H)  $^{13}\text{C}$  NMR (DMSO- $d_6$ )  $\delta$  162.4, 153.2, 123.3, 120.1, 119.2, 107.9, 78.9, 36.6, 28.7.; IR(KBr) 3350, 2978, 1700, 1670, 1586, 1458, 1368, 1247, 1112, 887, 779. FABMS  $m/e$  241.119 (M + H 241.119 calc. for  $\text{C}_{11}\text{H}_{17}\text{N}_2\text{O}_4$ ).

**1,2,3-Benzotriazol-1-yl 4-[(*tert*-butoxycarbonyl)-amino]-1-methylpyrrole-2-carboxylate (7).**

Boc-Py-acid **6** (31 g, 129 mmol) was dissolved in 500 mL DMF, HOBt (17.4 g, 129 mmol) was added followed by DCC (34 g, 129 mmol). The reaction was stirred for 24 h and then filtered dropwise into a well stirred solution of 5 L of ice water. The precipitate was allowed to sit for 15 min at 0°C and then collected by filtration. The wet cake was dissolved in 500 mL DCM, and the organic layer added slowly to a stirred solution of cold petroleum ether (4°C). The mixture was allowed to stand at -20°C for 4 h and then collected by vacuum filtration and dried *in vacuo* to provide (39 g, 85% yield) of **7** as a finely divided white powder. TLC (7:2 benzene/ ethyl acetate v/v) Rf 0.6,  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  9.43 (s, 1H), 8.12 (d, 1H,  $J$  = 8.4 Hz), 7.80 (d, 1H,  $J$  = 8.2Hz), 7.64 (t, 1H,  $J$  = 7.0 Hz), 7.51 (m, 2H), 7.18 (s, 1H), 3.83 (s, 3H), 1.45 (s, 9H),  $^{13}\text{C}$  NMR (DMSO- $d_6$ )  $\delta$  156.5, 153.3, 143.2, 129.6, 129.2, 125.7, 125.2, 124.6, 120.3, 112.8, 110.3, 109.8, 79.5, 36.8, 28.6.; IR (KBr) 3246, 3095, 2979, 1764, 1711, 1588, 1389, 1365, 1274, 1227, 1160, 1101, 999, 824, 748.; FABMS  $m/e$  358.152 (M + H 358.151 calc. for  $\text{C}_{17}\text{H}_{20}\text{N}_5\text{O}_4$ ).

**Ethyl 1-methylimidazole-2-carboxylate (8).** *N*-methylimidazole (320 g, 3.9 mol) was combined with 2 L acetonitrile and 1 L triethylamine in a 12 L flask equipped with a mechanical stirrer and the solution cooled to -20°C. Ethyl chloroformate (1000 g, 9.2 mol) was added with stirring, keeping the temperature between -20°C and -25°C. The reaction was allowed to *slowly* warm to room temperature and stir for 36 h. Precipitated triethylamine hydrochloride was removed by filtration and the solution concentrated *in vacuo* at 65°C. The resulting oil was purified by distillation under reduced pressure (2 torr, 102°C) to provide **8** as a white solid (360 g, 82% yield). TLC (7:2 benzene/ ethyl acetate v/v) Rf 0.2;  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  7.44 (d, 1 H,  $J$  = 2.8 Hz), 7.04 (d, 1 H,  $J$  = 2.8 Hz), 4.26 (q, 2 H,  $J$  = 3.5 Hz), 3.91 (s, 3 H), 1.26 (t, 3 H,  $J$  = 3.5 Hz);  $^{13}\text{C}$  NMR (DMSO- $d_6$ )  $\delta$  159.3, 129.1, 127.7, 61.0, 36.0, 14.5; IR(KBr) 3403, 3111, 2983, 1713,

1480, 1422, 1262, 1134, 1052, 922, 782, 666; FABMS  $m/e$  155.083 ( $M + H$  155.083 calc. for  $C_7H_{11}N_2O_2$ ).

**Ethyl 1-methyl-4-nitroimidazole-2-carboxylate (9).** Compound **8** was carefully dissolved in 1000 mL of concentrated sulfuric acid cooled to 0°C. 90% nitric acid (1 L) was slowly added maintaining a temperature of 0°C. The reaction was then refluxed with an efficient condenser (-20 °C) in a well ventilated hood for 50 min. The reaction was cooled with an ice bath, and quenched by pouring onto 10 L ice. The resulting blue solution was then extracted with 20 L DCM, the combined extracts dried (sodium sulfate) and concentrated *in vacuo* to yield a tan solid which was recrystallized from 22 L of 21:1 carbon tetrachloride/ethanol. The resulting white crystals are collected by vacuum filtration to provide pure **9** (103 g, 22% yield). TLC (7:2 benzene/ ethyl acetate v/v)  $R_f$  0.5,  $^1H$  NMR (DMSO- $d_6$ )  $\delta$  8.61 (s, 1 H), 4.33 (q, 2 H,  $J = 6.4$  Hz), 3.97 (s, 3 H), 1.29 (t, 3 H,  $J = 6.0$  Hz),  $^{13}C$  NMR (DMSO- $d_6$ )  $\delta$  158.2, 145.4, 135.3, 127.4, 62.2, 37.3, 14.5; IR(KBr) 3139, 1719, 1541, 1498, 1381, 1310, 1260, 1122, 995, 860, 656.; FABMS  $m/e$  200.066 ( $M + H$  200.067 calc. for  $C_7H_{10}N_3O_4$ ).

**Ethyl 4-amino-1-methylimidazole-2-carboxylate hydrochloride (10).** The nitro imidazole ethyl ester **9** (103g, 520 mmol) was dissolved in 5 L of 1:1 ethanol/ethyl acetate, 20g 10% Pd/C slurried in 500 mL ethyl acetate was added and the mixture stirred under a slight positive pressure of hydrogen (c.a. 1.1 atm) for 48 h. The reaction mixture was filtered, concentrated *in vacuo* to a volume of 500 mL and 5 L of cold anhydrous ethyl ether added. Addition of HCl gas provided a white precipitate. The solution was cooled at -20°C for 4 h and the precipitate collected by vacuum filtration and dried *in vacuo* to provide (75 g, 78% yield) of **10** as a fine white powder. TLC (7:2 benzene: ethyl acetate)  $R_f$  (amine) 0.3,  $R_f$  (salt) 0.0.  $^1H$  NMR (DMSO- $d_6$ )  $\delta$  10.11 (br s, 3H), 7.43 (s, 1H), 4.28 (q, 2H,  $J = 7.1$  Hz), 3.92 (s, 1H), 1.28 (t, 3H,  $J = 7.1$  Hz)  $^{13}C$  NMR (DMSO- $d_6$ )  $\delta$  157.6, 132.6, 117.4, 117.3, 61.8, 36.6, 14.5; IR(KBr) 3138, 2883, 1707, 1655, 1492, 1420, 1314, 1255, 1152, 1057, 837, 776.; FABMS  $m/e$  169.085 (169.084 calc. for  $C_7H_{11}N_3O_2$ ).

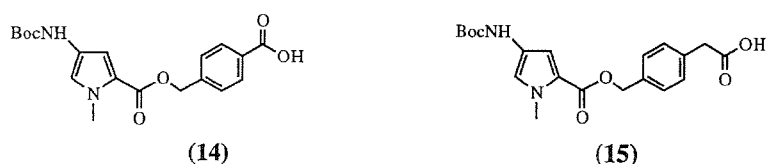
**4-[(*tert*-butoxycarbonyl)amino]-1-methylimidazole-2-carboxylic acid (11).** The imidazole amine **10** (75 g, 395 mmol) was dissolved in 200 mL DMF. DIEA (45 mL, 491 mmol) was added followed by di-*t*-butyldicarbonate (99 g, 491 mmol). The mixture was shaken at 60°C for 18 h, allowed to assume room temperature, and partitioned between 500 mL brine, 500 mL ethyl ether. The ether layer was extracted (2 x 200 mL each) 10% citric acid, brine, satd. sodium bicarbonate, brine, dried over sodium sulfate and concentrated *in vacuo* to yield the Boc-ester contaminated with 20% Boc-anhydride as indicated by <sup>1</sup>H NMR. The Boc-ester, used without further purification, was dissolved in 200 mL 1M NaOH. The reaction mixture was allowed to stand for 3 h at 60°C with occasional agitation. The reaction mixture was cooled to 0°C, and carefully neutralized with 1 M HCl to pH 2, at which time a white gel forms. The gel was collected by vacuum filtration, frozen before drying, and remaining water lyophilized to yield **10** as a white powder (51 g, 54% yield). <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>) δ 9.47 (s, 1H), 7.13 (s, 1H), 3.85 (s, 3H), 1.41 (s, 9H). <sup>13</sup>C NMR (DMSO-*d*<sub>6</sub>) δ 160.9, 152.9, 137.5, 134.5, 112.4, 79.5, 35.7, 28.6; IR(KBr) 3448, 2982, 1734, 1654, 1638, 1578, 1357, 1321, 1249, 1163, 799.; FABMS *m/e* 241.105 (241.106 calc. for C<sub>10</sub>H<sub>15</sub>N<sub>3</sub>O<sub>4</sub>).

**4-[(*tert*-butoxycarbonyl)amino]-1-methylpyrrole-2-(4-carboxamido-1-methyl imidazole)-2-carboxylic acid. 12** was prepared as described below for **13** substituting Boc-Pyrrole acid for Boc-γ-aminobutyric acid (4.1 g, 91% yield). <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>) δ 10.58 (s, 1 H), 9.08 (s, 1 H), 7.57 (s, 1 H), 6.97 (s, 1 H), 6.89 (s, 1 H), 3.89 (s, 3 H), 3.75 (s, 3 H), 1.35 (s, 9 H); <sup>13</sup>C NMR (DMSO-*d*<sub>6</sub>) δ 160.36, 159.1, 153.4, 137.9, 132.3, 122.8, 122.3, 118.5, 115.5, 105.5, 105.4, 78.8, 28.7, 24.9; IR 3346, 2929, 1685, 1618, 1529, 1342, 1274, 1179, 997, 761. FABMS *m/e* 364.161 (364.162 calc. for C<sub>16</sub>H<sub>22</sub>N<sub>5</sub>O<sub>5</sub>).

**γ-[(*tert*-butoxycarbonyl)amino]-butyric acid -(4-carboxamido-1-methyl-imidazole)-2-carboxylic acid 13.** To a solution of Boc-γ-aminobutyric acid (10 g, 49 mmol) in 40 mL DMF was added 1.2 eq HOBt (7.9 g, 59 mmol) followed by 1.2 eq DCC (11.9 g, 59 mmol). The solution was stirred for 24 h, and the DCU removed by filtration. Separately, to a solution of ethyl 4-nitro-1-methylimidazole-2-carboxylate (9.8 g, 49 mmol) in 20 mL DMF was added Pd/C

catalyst (10%, 1 g), and the mixture was hydrogenated in a Parr bomb apparatus (500 psi H<sub>2</sub>) for 2 h. The catalyst was removed by filtration through celite and filtrate immediately added to the OBt ester solution. An excess of DIEA (15 mL) was then added and the reaction stirred at 37°C for 48 h. The reaction mixture was then added dropwise to a stirred solution of ice water and the resulting precipitate collected by vacuum filtration to provide crude ethyl  $\gamma$ -[[tert-butoxy)carbonyl]amino]-butyric acid -(4-carboxamido-1-methyl-pyrrole)-2-carboxylate (5 g, 14.1 mmol). To the crude ester dissolved in 50 mL methanol was added 50 mL 1M KOH and the resulting mixture allowed to stir for 6 h at 37°C. Excess methanol was removed *in vacuo* and the resulting solution acidified by the addition of 1 M HCl. The resulting precipitate was collected by vacuum filtration and dried *in vacuo* to yield a brown powder. (4.4g, 89% yield). <sup>1</sup>H NMR (DMSO-d<sub>6</sub>)  $\delta$  10.50 (s, 1 H), 7.45 (s, 1 H), 6.82 (t, 1 H, *J* = 3.6 Hz), 3.86 (s, 3 H), 2.86 (q, 2 H, *J* = 4.6 Hz), 2.22 (t, 2 H, *J* = 7.4 Hz), 1.57 (quintet, 2 H, *J* = 5.9 Hz), 1.29 (s, 9 H); IR 3416, 2950, 2841, 1650, 1538 1449, 1392, 1250, 1165, 1108; FABMS *m/e* 326.160 (326.159 calc. for C<sub>14</sub>H<sub>22</sub>N<sub>4</sub>O<sub>5</sub>).

## Resin Synthesis



**Figure 2.9.** Linkers prepared for the Synthesis of Pyrrole-PAM resin.

**4-(Bromomethyl)benzoic Acid Phenacyl Ester.** Triethylamine (16 ml, 115 mmol) and bromoacetophenone (22.9 g, 115 mmol) were dissolved in 450 ml ethyl acetate. 4-(Bromomethyl)benzoic Acid (17.5 g, 155 mmol) was added in seven equal portions over a three hour period to the stirred solution at 50°C. Stirring was continued for an additional 8 hours at the same temperature. Precipitated triethylaminehydrobromide was removed by filtration, and the



ethyl acetate solution washed (3 x 150 ml each) 10% citric acid, brine, satd. sodium bicarbonate, brine. The organic phase was dried, sodium sulfate, and concentrated *in vacuo*. The residue was recrystallized from dichloromethane-petroleum ether to give fine white crystals (10.2 g, 27% yield).  $^1\text{H}$  NMR  $\delta$  7.99 (m, 4H), 7.69-7.54 (m, 5H), 5.74 (s, 2H), 4.77 (s, 2H)  $^{13}\text{C}$  NMR (DMSO- $d_6$ )  $\delta$  193.7, 165.9, 144.6, 134.9, 131.2, 131.0, 130.7, 130.6, 130.2, 129.9, 128.8, 128.6, 68.2, 34.0.

**Boc-Pyrrolyl-4-(oxymethyl)benzoic Acid Phenacyl Ester.** A solution of Boc-Pyrrole-OH **8** (2.9 g 12 mmol), 4-(bromomethyl)phenylacetic Acid Phenacyl Ester (4 g, 12 mmol), and diisopropylethylamine (3.0 ml, 16.8 mmol) in 60 ml DMF were stirred at 50°C for 6 hours. The solution was cooled and partitioned between 400 ml water and 400 ml ethyl ether. The ether layer was washed (2 x 200 ml each) 10% citric acid, brine, satd. sodium bicarbonate, brine. The organic phase was dried, sodium sulfate, and concentrated *in vacuo* to yield a light white foam which was used without further purification. (5.4g 97% yield) TLC (2:3 hexane/ethyl acetate v/v) Rf 0.6.  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  9.14 (s, 1H), 8.03 (m, 4H), 7.67 (m, 1H), 7.55 (m, 4H), 7.13 (s, 1H), 6.72 (d, 1H,  $J$  = 1.5 Hz), 5.74 (s, 1H), 5.32 (s, 1H), 3.79 (s, 3H), 1.42 (s, 9H)  $^{13}\text{C}$  NMR (DMSO- $d_6$ )  $\delta$  193.2, 165.5, 160.4, 153.2, 143.1, 134.5, 130.1, 129.5, 128.3, 128.2, 123.8, 120.3, 118.8, 108.2, 79.1, 67.7, 64.6, 36.7, 28.6.

**Boc-Pyrrolyl-4-(oxymethyl)phenylacetic Acid Phenacyl Ester.** Prepared as described for Boc-Pyrrolyl-4-(oxymethyl)benzoic Acid Phenacyl Ester. Product purified by crystallization hexane - ethyl acetate (3:1, v/v) as long needles (6.1 g, 44.5%) TLC (3:1 hexane/ethyl acetate) Rf 0.2  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  9.11, (s, 1H), 7.93 (d, 2H,  $J$ = 8.2), 7.67 (t, 1H  $J$ = 7.0), 7.52 (t, 2H,  $J$ = 7.9), 7.35 (m, 4H), 7.10 (s, 1H), 6.67 (s, 1H), 5.50 (s, 2H), 5.22 (s, 2H), 5.19 (s, 2H), 3.83 (s, 3H), 1.42 (s, 9H)  $^{13}\text{C}$  NMR (DMSO- $d_6$ )  $\delta$  193.1, 171.2, 160.6, 153.2, 135.7, 134.4, 130.1, 129.4, 128.5, 128.3, 123.7, 120.0, 119.0, 108.0, 79.0, 67.4, 65.1, 36.7, 28.6.

**Boc-Pyrrolyl-4-(oxymethyl)benzoic Acid (**14**)** Zinc dust was activated as previously described. Boc-Pyrrolyl-4-(oxymethyl)benzoic Acid Phenacyl Ester (3 g, 5.9 mmol) was dissolved in 90 ml Acetic acid/water (80:20, v/v). Zinc dust (9.6g 147 mmol) was added and the reaction stirred for

18 hours at room temperature, the zinc removed by filtration, and the reaction mixture partitioned between 200 ml ethyl ether and 200 ml water, the layers separated, and the aqueous layer extracted with 200 ml ethyl ether, the combined ether layers are washed (5 x 100 ml) with water. The combined organics were dried, sodium sulfate, concentrated *in vacuo*, and azeotroped (6 x 100 ml benzene). The product, contaminated with acetophenone, was purified by flash chromatography, with a gradient of 2:1 hexane ethyl acetate to ethyl acetate to give a yellow oil. (1.9 g, 54%) TLC ethyl acetate Rf 0.7

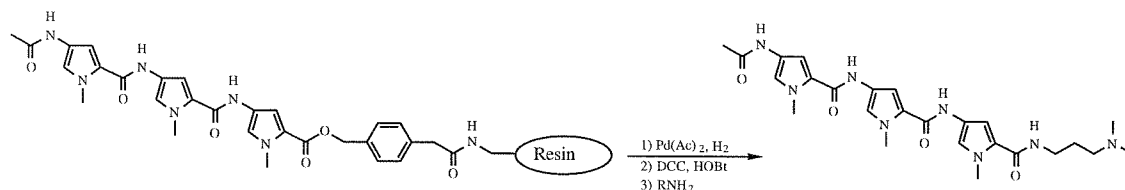
**Boc-Pyrrolyl-4-(oxymethyl)phenylacetic Acid** was prepared in a manner analogous to (14) as a yellow oil in 78% yield.

**Boc-aminoacyl-Pyrrolyl-4-(oxymethyl)-BAM-resin.** Bam linker acid (1 g, 2.6 mmol) was dissolved in 6.5 ml DMF. HOBT (382 mg, 2.8 mmol) followed by DCC (735 mg, 2.8 mmol) was added and the reaction mixture shaken for 4 hours at room temperature. The precipitated DCU was filtered and the reaction mixture added to 3 grams aminomethyl-polystyrene-resin (0.7 mmol/gram substitution) previously swollen for 30 min. in DMF. Diisopropylethylamine (913  $\mu$ l, 5.3 mmol) was added and the reaction shaken for 12 hours. The resin was determined by ninhydrin test to be approximately 0.3 mmol/gram substituted at this point, the resin was washed with DMF and the remaining amine groups capped by acetylation (2x) with excess acetic anhydride capping solution. The resin was washed with DMF, DCM, and MeOH and dried *in vacuo*.

**Boc-aminoacyl-Pyrrolyl-4-(oxymethyl)-Pam-resin.** Pyrrole Pam resin was prepared in 0.3 mmol/g substitution as described for Bam resin.

**Cleavage from the resin.** Acetylated tripyrrole was chosen as the target compound for experiments involving resin cleavage.<sup>8</sup> Successful cleavage was achieved from PAM and BAM pyrrole resin with Pd(Ac)<sub>2</sub> in DMF under a pressurized atmosphere of hydrogen (100psi, 8 hours) (Figure 3.10). The palladium black is then filtered, and the pyrrole acid activated with DCC/HOBT and reacted with a large excess of dimethylaminopropyl amine to give the HPLC

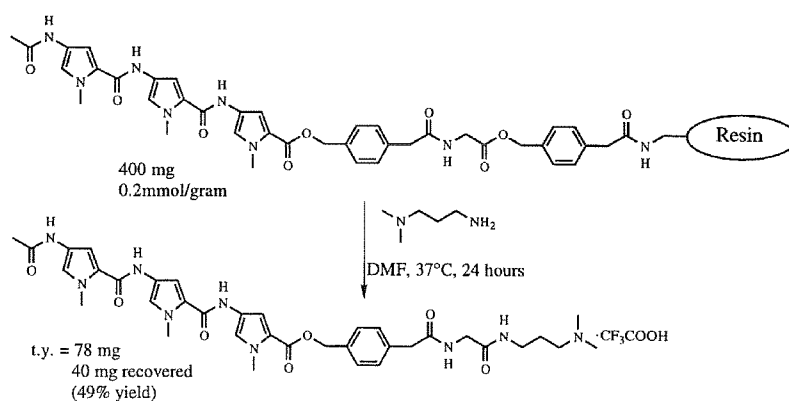
purified acetylated tripyrrole in 5% overall yield, Figure 3.10. This approach may be limited by the potential insolubility of longer peptide-acids.



**Figure 3.10.** Synthesis of AcP<sub>3</sub> by a 2 step cleavage/activation procedure, hydrogenolysis to liberate AcP<sub>3</sub>COOH, followed by DCC/HOBt coupling to the amine side chain.

Cleavage from PAM and BAM resins by aminolysis was unsuccessful at 37°C and 100°C in 1:1 amine:DMF or neat amine for 24 hours. Treatment of BAM resin with neat amine at 100°C for 3 days resulted in a 12% recovery of acetylated tripyrrole. The presence of a large number of decomposition products indicates that cleavage at 100°C for 3 days is too harsh to be generally useful.

As a control the Boc-pyrrole-Pam linker was coupled to Boc-Gly-Pam-Resin in a manner analogous to the synthesis of Boc-Py-Pam-Resin, Figure 8. Two additional pyrroles were then added to the resin with the standard protocols to give **16**. The 'safety catch' resin was subjected to treatment with a 1:1 mixture of dimethylaminopropyl amine: DMF at 37°C for 12 hours. Two



**Figure 3.11.** Selective cleavage of a glycine benzyl ester in the presence of a pyrrole benzyl ester. 37°C, 1:1 RNH<sub>2</sub>:DMF.

products AcPyPyPy-PAM-Gly-DMAPA **17** (98% of products) and the failure sequence AcPyPy-PAM-Gly-DMAPA (2%) were identified by  $^1\text{H}$  NMR. Recovery of the product was very high, almost 50% of the theoretical yield, indicating that the glycine-PAM, aliphatic benzyl ester is cleaved with much higher recovery than the pyrrole-PAM aromatic benzyl ester.

**AcPy<sub>3</sub>-PAM-Gly-DMAPA. 16** 180 mg (29  $\mu\text{mole}$ ) of AcPy<sub>3</sub>-PAM-Gly-PAM-Resin was treated with 1.5 ml DMF followed by 1.5 ml dimethylaminopropylamine and the reaction mixture shaken for 12 hours, and purified by prep HPLC to give AcPyPyPy-PAM-Gly-DMAPA in 49% yield.  $^1\text{H}$  NMR ( $\text{DMSO-d}_6$ )  $\delta$  9.90 (m, 2H), 9.83 (s, 1H), 9.3 (br s, 1H), 8.37 (t, 1H,  $J = 5.7$  Hz), 8.05 (t, 1H,  $J = 5.8$  Hz), 7.44 (d, 1H,  $J = 1.7$  Hz), 7.32 (q, 4H,  $J = 8.2$  Hz), 7.20 (d, 1H,  $J = 1.7$  Hz), 7.13 (d, 1H,  $J = 1.7$  Hz), 7.04 (d, 1H,  $J = 1.7$  Hz), 6.95 (d, 1H,  $J = 1.9$  Hz), 6.83 (d, 1H,  $J = 1.8$  Hz), 5.19 (s, 2H), 3.82 (s, 3H), 3.82 (s, 3H), 3.81 (s, 3H), 3.63 (d, 2H,  $J = 6.1$  Hz), 3.48 (s, 2H), 3.11 (q, 2H,  $J = 6.1$  Hz), 2.96 (m, 2H), 2.67 (d, 6H,  $J = 4.8$  Hz), 1.95 (s, 3H), 1.71 (quintet, 2H,  $J = 7.4$  Hz) A failure sequence was also isolated from the reaction mixture.  $^1\text{H}$  NMR ( $\text{DMSO-d}_6$ )  $\delta$  9.91 (m, 2H), 9.80 (s, 1H), 9.3 (br s, 1H), 8.40 (t, 1H,  $J = 5.7$  Hz), 8.08 (t, 1H,  $J = 5.8$  Hz), 7.44 (d, 1H,  $J = 1.7$  Hz), 7.38 (q, 4H,  $J = 8.4$  Hz), 7.15 (d, 1H,  $J = 1.7$  Hz), 6.96 (d, 1H,  $J = 1.8$  Hz), 6.85 (d, 1H,  $J = 1.7$  Hz), 5.15 (s, 2H), 3.84 (s, 3H), 3.82 (s, 3H), 3.69 (d, 2H,  $J = 5.4$  Hz), 3.51 (s, 2H), 3.19 (m, 2H), 3.04 (m, 2H), 2.74 (d, 6H,  $J = 4.2$  Hz), 1.97 (s, 3H), 1.77 (m, 2H).

## Solid Phase Syntheses

**Activation of Imidazole-2-carboxylic acid,  $\gamma$ -aminobutyric acid, Boc-glycine, and Boc- $\beta$ -alanine.** The appropriate amino acid or acid (2 mmol) was dissolved in 2 mL DMF. HBTU (720 mg, 1.9 mmol) was added followed by DIEA (1 mL) and the solution lightly shaken for at least 5 min.

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

**Activation of Boc- $\gamma$ -Imidazole acid and Boc-Pyrrole-Imidazole acid.** The appropriate dimer (1 mmol) and HBTU (378 mg, 1 mmol) are combined in 2 mL DMF. DIEA (1 mL) is then added and the reaction mixture allowed to stand for 5 min.

**Activation of Boc-Pyrrole acid. (for coupling to Imidazole amine)** Boc-Pyrrole acid (514 mg, 2 mmol) was dissolved in 2 mL dichloromethane, DCC (420 mg, 2 mmol) added, and the solution allowed to stand for 10 min, DMAP (101 mg, 1 mmol) was added and the solution allowed to stand for 1 min.

**Acetylation Mix.** 2 mL DMF, DIEA (710  $\mu$ L, 4.0 mmol), and acetic anhydride (380  $\mu$ L, 4.0 mmol) were combined immediately before use.

**Manual Synthesis Protocol.** Boc- $\beta$ -alanine-Pam-Resin (1.25 g, 0.25 mmol) is placed in a 20 mL glass reaction vessel, shaken in DMF for 5 min and the reaction vessel drained. The resin was washed with DCM (2 x 30 s.) and the Boc group removed with 80% TFA/DCM/0.5M PhSH, 1 x 30 s., 1 x 20 min. The resin was washed with DCM (2 x 30 s.) followed by DMF (1 x 30 s.). A resin sample (5 - 10 mg) was taken for analysis. The vessel was drained completely and activated monomer added, followed by DIEA if necessary. The reaction vessel was shaken vigorously to make a slurry. The coupling was allowed to proceed for 45 min., and a resin sample taken. The reaction vessel was then washed with DCM, followed by DMF.

**Machine-Assisted Protocols.** Machine-assisted synthesis was performed on a ABI 430A synthesizer on a 0.18 mmol scale (900 mg resin; 0.2 mmol/gram). Each cycle of amino acid addition involved: deprotection with approximately 80% TFA/DCM/0.4M PhSH for 3 min., draining the reaction vessel, and then deprotection for 17 min.; 2 dichloromethane flow washes; an NMP flow wash; draining the reaction vessel; coupling for 1 hour with *in situ* neutralization, addition of dimethyl sulfoxide (DMSO)/NMP, coupling for 30 min., addition of DIEA, coupling for 30 min.; draining the reaction vessel; washing with DCM, taking a resin sample for evaluation of the progress of the synthesis by HPLC analysis; capping with acetic anhydride/DIEA in DCM for 6 min.; and washing with DCM. A double couple cycle is employed when coupling aliphatic amino acids to imidazole, all other couplings are performed with single couple cycles.

The ABI 430A synthesizer was left in the standard hardware configuration for NMP-HOBt protocols. Reagent positions 1 and 7 were DIEA, reagent position 2 was TFA/0.5M thiophenol, reagent position 3 was 70% ethanolamine/methanol, reagent position 4 was acetic anhydride, reagent position 5 was DMSO/NMP, reagent position 6 was methanol, and reagent position 8 was DMF. New activator functions were written, one for direct transfer of the cartridge contents to the concentrator (switch list 21, 25, 26, 35, 37, 44), and a second for transfer of reagent position 8 directly to the cartridge (switch list 37, 39, 45, 46).

Boc-Py-OBt ester (357 mg, 1 mmol) was dissolved in 2 ml DMF and filtered into a synthesis cartridge. Boc-Im acid monomer was activated (DCC/HOBt), filtered, and placed in a synthesis cartridge. Imidazole-2-carboxylic acid was added manually. At the initiation of the coupling cycle the synthesis was interrupted, the reaction vessel vented and the activated monomer added directly to the reaction vessel through the resin sampling loop via syringe. When manual addition was necessary an empty synthesis cartridge was used. Aliphatic amino acids (2 mmol) and HBTU (1.9 mmol) were placed in a synthesis cartridge. 3 ml of DMF was added using a calibrated delivery loop from reagent bottle 8, followed by calibrated delivery of 1 ml DIEA from reagent bottle 7, and a 3 minute mixing of the cartridge.

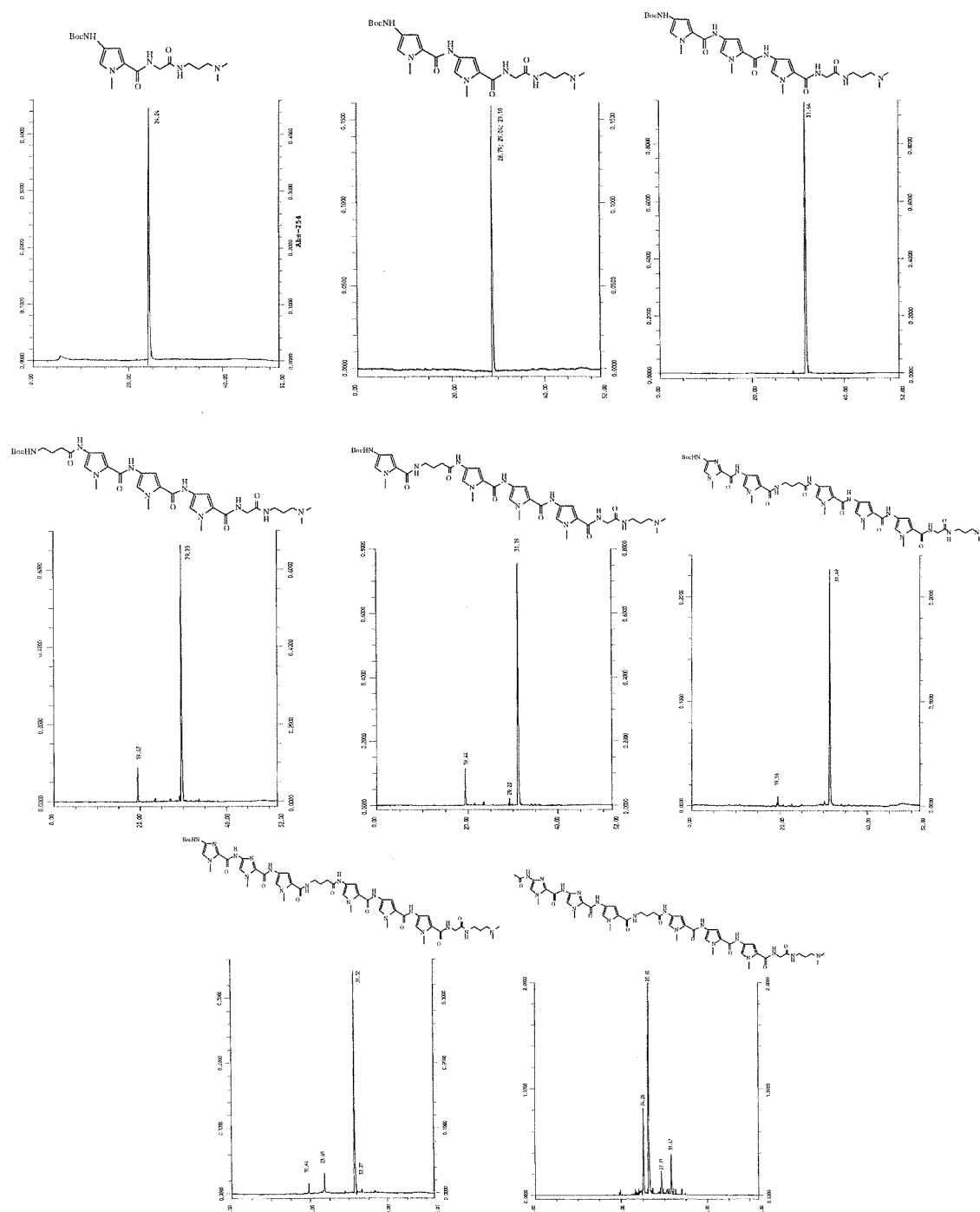
The activator cycle was written to transfer activated monomer directly from the cartridge to the concentrator vessel, bypassing the activator vessel. After transfer, 1 ml of DIEA was measured into the cartridge using a calibrated delivery loop, and the DIEA solution combined with the activated monomer solution in the concentrator vessel. The activated ester in 2:1 DMF/DIEA was then transferred to the reaction vessel. All lines were emptied with argon before and after solution transfers.

**ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp (2).** ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Pam-Resin was prepared by machine-assisted synthesis protocols. A sample of resin (1 g, 0.17 mmol[Resin substitution is calculated as  $L_{\text{new}}(\text{mmol/g}) = L_{\text{old}} / (1 + L_{\text{old}}(W_{\text{new}} - W_{\text{old}}) \times 10^{-3})$ ; L is the loading, and W is the molecular weight of the polyamide attached to the resin]<sup>28</sup>) was placed in a 20 mL glass scintillation vial, 4 mL of dimethylaminopropylamine added, and the solution heated at 55°C for 18 h. Resin is removed by

filtration through a disposable propylene filter and 16 mL of water added. The polyamide/amine mixture was purified directly by preparatory HPLC and the appropriate fractions lyophilized to yield a white powder. (103 mg, 61% recovery) HPLC r.t. 24.1, UV  $\lambda_{\text{max}}(\text{H}_2\text{O})$  ( $\epsilon$ ), 234 nm (39,300), 304 nm (52,000);  $^1\text{H}$  NMR ( $\text{DMSO-d}_6$ ); 10.47 (s, 1 H), 9.91 (s, 1 H), 9.89 (s, 1 H), 9.87 (s, 1 H), 9.84 (s, 1 H), 9.2 (br s, 1 H), 8.08 (m, 3 H), 7.38 (s, 1 H), 7.26 (d, 1 H,  $J = 1.0$  Hz), 7.20 (d, 1 H,  $J = 1.0$  Hz), 7.14 (m, 4 H), 7.04 (d, 1 H,  $J = 1.1$  Hz), 7.02 (d, 1 H,  $J = 1.1$  Hz), 6.89 (d, 1 H,  $J = 1.0$  Hz), 6.85 (m, 2 H), 3.97 (s, 3 H), 3.82 (m, 6 H), 3.81 (s, 3 H), 3.77 (m, 6 H), 3.34 (m, 2 H,  $J = 3.9$  Hz), 3.18 (m, 2 H,  $J = 5.5$  Hz), 3.06 (m, 2 H,  $J = 5.7$  Hz), 2.95 (m, 2 H,  $J = 4.9$  Hz), 2.71 (d, 6 H,  $J = 4.6$  Hz), 2.30 (m, 6 H), 1.75 (m, 4 H); MALDI-TOF MS 978.0 (978.1 calc. for  $\text{M}+\text{H}$ ).

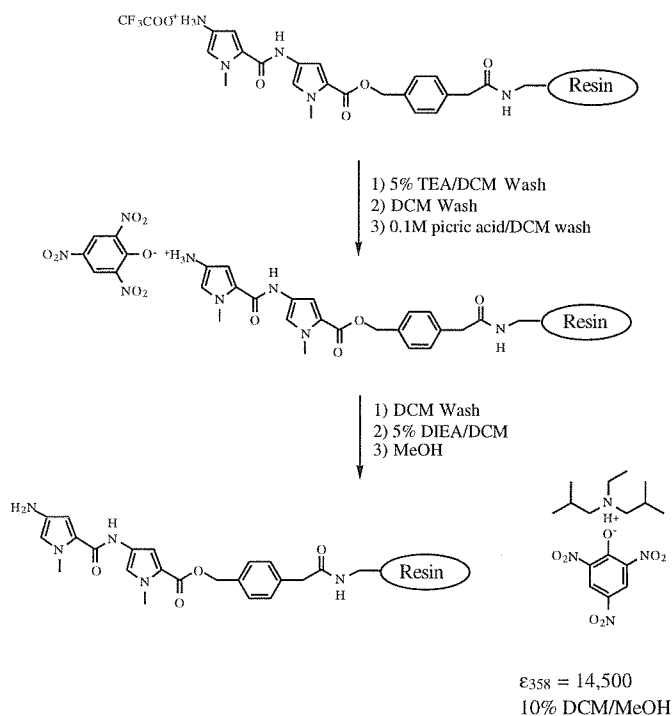
**Stepwise HPLC analysis.** A resin sample (c.a. 4 mg) was placed in a 4 mL glass test tube, 200  $\mu\text{L}$  of *N,N*-dimethylaminopropylamine was added and the mixture heated at 100°C for 5 min. The cleavage mixture was filtered and a 25  $\mu\text{L}$  sample analyzed by analytical HPLC at 254 nm.

This method suffers from the possibility of impurities with unusually large extinction coefficients providing misleading results. On more than one occasion a major side product, as indicated by analytical HPLC, has been found when isolated to constitute less than 1% of the sample by weight. Stepwise HPLC analysis has been applied to a large number of the syntheses reported here, and would indicate that very high yields (> 99%) are routinely achieved in most coupling reactions. HPLC does indicate that failures and side reactions up to 5% (beyond the detection limit of the picric acid titration) still occasionally occur. The use of stepwise HPLC analysis is an effective way to obtain detailed information on the progress of a synthesis, allowing the exact step that results in a side reaction or deletion product to be readily identified (Figure 3.12).



**Figure 3.12.** Monitoring the synthesis of AcImImPy- $\gamma$ -PyPyPy-G-Dp by analytical HPLC. Reverse Phase  $C_{18}$  chromatography, 0.1% TFA (wt/v), 1.5%/min.  $CH_3CN$ .





**Figure 3.13.** Picric acid titration for determination of free amine groups.

a reaction is >90% complete. The picric acid test is useful for immediate monitoring of coupling reactions, but not applicable for accurately documenting reaction yields.

**Picric Acid Titration.** For coupling to aromatic amines, the Picric acid titration has proven useful for estimating coupling yields. The picric acid titration involves washing a resin sample (c.a. 5 mg) successively with 5% TEA/DCM, DCM, 0.1M picric acid/DCM, followed by DCM. The picrate salt thus formed is eluted with 5% DIEA/DCM, and quantitated from the reported extinction coefficient, Figure 3.13. The picric acid test is inaccurate for low concentrations of amine, thus it is only possible to tell if

## References

1. (a) Moser, H.E.; Dervan, P.B. *Science* **1987**, *238*, 645. (b) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhouh, N.; Decout, J.L.; Thoung, N.T.; Lhomme, J.; Helene, C. *Nucleic Acids Res.* **1987**, *15*, 7749. (c) For a review, see: Thoung, N.T.; Helene, C. *Angew. Chemie Int. Ed.* **1993**, *32*, 666.
2. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D.E.; Dervan, P. B. *Proc. Natl. Acad. Sci., U.S.A.* **1992**, *89*, 7586. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385.
3. (a) Van Dyke, D.M.; Hertzberg, R.P.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5470. (b) Taylor, J.S.; Schultz, P.G.; Dervan, P.B. *Tetrahedron* **1984**, *40*, 457. (c) Schultz, P.G.; Dervan, P.B. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1134. (d) Klevit, R.E.; Wemmer, D.E.; Reid, B.R. *Biochemistry* **1986**, *25*, 3296.
4. (a) Lown, J.W.; Krowicki, K.; Bhat, U.G.; Ward, B.; Dabrowiak, J.C. *Biochemistry* **1986**, *25*, 7408. (b) Kissinger, K.; Krowicki, K.; Dabrowiak, J.C.; Lown, J.W. *Biochemistry* **1987**, *26*, 5590. (c) Lee, M.; Chang, D.K.; Hartley, J.A.; Pon, R.T.; Krowicki, K.; Lown, J.W. *Biochemistry* **1987**, *27*, 445. (d) Lee, M.; Rhodes, A.L.; Wyatt, M.D.; Forrow, S.; Hartley, J.A. *Biochemistry* **1993**, *32*, 4237.
5. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S.T.; Sundaralingham, M. *Struct. Biol. Nature* **1994**, *1*, 169.
6. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (b) Geierstanger, B.H.; Jacobsen, J.P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E.; *Biochemistry* **1994**, *33*, 3055. (c) Geierstanger, B.H.; Dwyer, T.J.; Bathini, Y.; Lown, J.W.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.
7. (a) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646. (b) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 3325.
8. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 9892. (b) Dwyer, T.J.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 9900. (c) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 3663. (d) Chen, Y.H.; Lown, J.W. *J. Am. Chem. Soc.* **1994**, *116*, 6995.
9. Mrksich, M.; Parks, M.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 7983.
10. (a) Weiss, M.J.; Webb, J.S.; Smith, J.M. *J. Am. Chem. Soc.* **1957**, *79*, 1266. (b) Arcamone, F.; Orezzi, P.G.; Barbieri, W.; Nicoletta, V.; Penco, S. *Gazz. Chim. Ital.* **1967**, *97*, 1097. (c) Penco, S.; Redaelli, S.; Arcamone, F. *Gazz. Chim. Ital.* **1967**, *97*, 1110. (d) Bailer, M.; Yagen, B.; Machoulam, P. *Tetrahedron* **1978**, *34*, 2389
11. Lown, J.W.; Krowicki, K. *J. Org. Chem.* **1985**, *50*, 3774.
12. (a) Church, K.M.; Wurdeman, R.L.; Zhang, Y.; Chen, F.; Gold, B. *Biochemistry* **1990**, *29*, 6827. (b) He, G.; Browne, K.A.; Gropee, J.C.; Blasko, A.; Mei, H.; Bruice, T.C. *J. Am. Chem. Soc.* **1993**, *115*, 7061.

13. (a) Merrifield, R.B. *J. Am. Chem. Soc.* **1963**, 85, 2149 (b) Merrifield, B. *Science*, **1986**, 232, 341. (c) Kent, S.B.H. *Ann. Rev. Biochem.* **1988**, 57, 957.
14. (a) Caruthers, M.H.; Barone, A.D.; Beaucage, S.L.; Dodds, D.R.; Fisher, E.F. *Methods. Enzymol.* **1987**, 154, 287 (b) Caruthers, M.H. *Acc. Chem. Res.* **1991**, 24, 278.
15. (a) Simon, R.J.; Kania, R.S.; Zuckermann, R.N.; Hueber, V.D.; Jewell, D.A.; Banville, S.; Ng, S.; Wang, L.; Rosenberg, S.; Marlowe, C.K.; Spellmeyer, D.C.; Frankel, A.D.; Santi, D.V.; Cohen, F.E.; Bartlett, P.A.; *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 9367. (b) Zuckermann, R.N.; Kerr, J.M.; Kent, S.B.H.; Moos, W.H.; *J. Am. Chem. Soc.* **1992**, 114, 10646.
16. (a) Danishefsky, S.J.; McClure, K.F.; Randolph, J.T.; Ruggeri, R.B. *Science*, **1993**, 260, 1307. (b) Roberge, J.Y.; Beebe, X.; Danishefsky, S.J. *Science*, **1995**, 269, 202.
17. Bunin, B.A.; Ellman, J.A. *J. Amer. Chem. Soc.* **1992**, 113, 10997.
18. Grehn, L.; Ragnarsson, U. *J. Org. Chem.* **1981**, 46, 3492.
19. Grehn, L.; Ding, L.; Ragnarsson U. *Acta. Chim. Scand.* **1990**, 44, 67.
20. Bailey, D.M.; Johnson, R.E.; Albertson, N.F. *Org. Synth.* **1971**, 51, 101.
21. Nishiwaki, E.; Tanaka, S.; Lee, H.; Shibuya, M. *Heterocycles* **1988**, 27, 1945.
22. Bailey, C.; Pommery, N.; Houssin, R.; Henichart, J. *J. Pharm. Sci.* **1989**, 78, 910.
23. Krowicki, K.; Lown, W.J. *J. Org. Chem.* **1987**, 52, 3493.
24. Mitchell, A.R.; Kent, S.B.H.; Engelhard, M.; Merrifield, R.B. *J. Org. Chem.* **1978**, 43, 2845.
25. Parks, M.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, 118, 6147.
26. (a) Schnolzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S.B.H. *Int. J. Peptide. Protein. Res.* **1992**, 40, 180. (b) Milton, R.C. deL.; Milton, S.C.F.; Kent, S.B.H. *Science* **1992**, 256, 1445.
27. Sarin, V.K.; Kent, S.B.H.; Tam, J.P.; Merrifield, R.B. *Anal. Biochem.* **1981**, 117, 147.
28. Barlos, K.; Chatzi, O.; Gatos, D.; Stravropoulos, G. *Int. J. Peptide Protein Res.* **1991**, 37, 513.

## CHAPTER 3

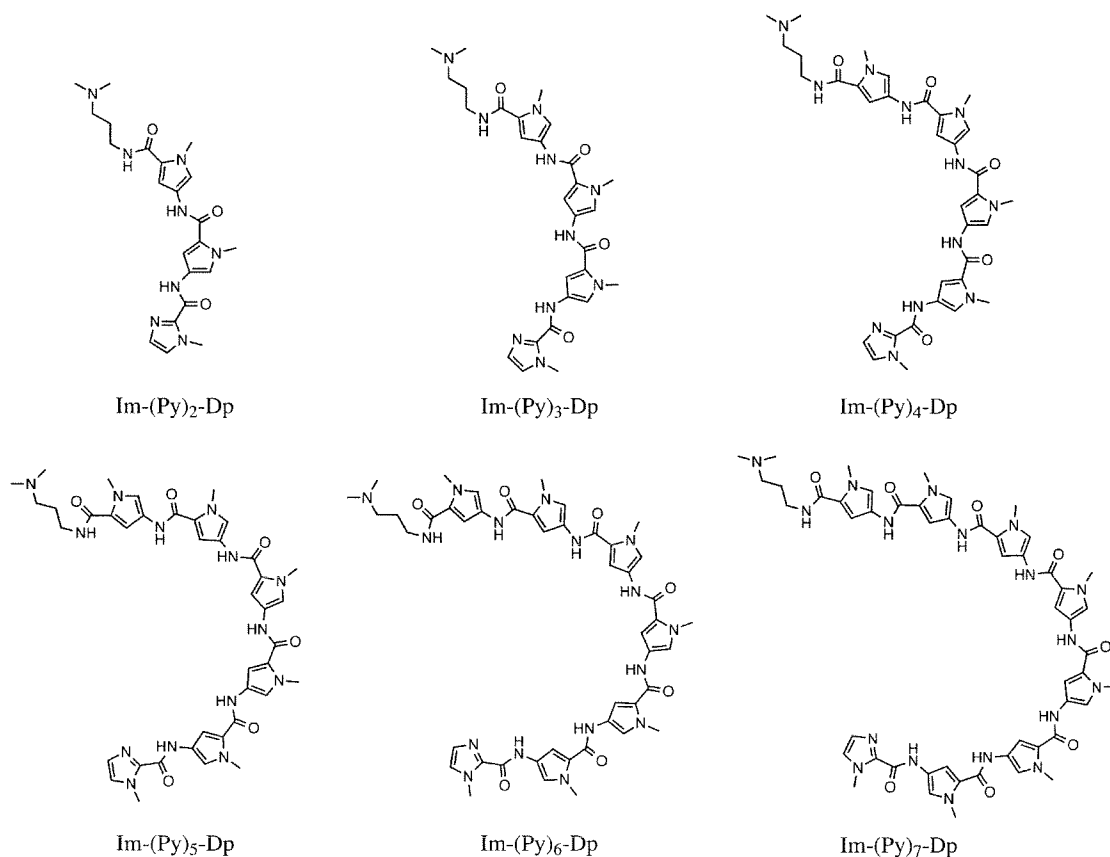
# Extending the Binding Site Size Limit of the 2:1 Pyrrole-Imidazole Polyamide-DNA Motif

**Abstract:** Polyamides containing Imidazole (Im) and Pyrrole (Py) amino acids can be combined in antiparallel side-by-side dimeric complexes for sequence-specific recognition in the minor groove of DNA. For polyamides containing three to eight rings, affinity maximizes and is similar at ring sizes of five, six, and seven. Sequence specificity decreases as the length of the polyamides and the corresponding DNA-binding sites increases. A  $\beta$ -alanine residue is found to join polyamide subunits in an extended conformation, providing a structural motif for the design of new polyamides targeted to sequences longer than 7 base pairs. It is found that there exists at least a 20-fold preference for placement of the  $\beta/\beta$  pair opposite A•T/T•A relative to G•C/C•G. These results implicate the  $\beta/\beta$  combination as both a flexible spacer unit and a sequence-specific DNA binding element. A high-resolution x-ray crystal structure of a four-ring Py-Im polyamide specifically bound as a dimer to a six-base DNA site reveals that the polyamide rise-per-aromatic amino acid residue matches the pitch of the B-DNA helix. However, the curvature of the polyamide is over-wound with respect to the DNA consistent with unfavorable interaction for complexes comprised of more than five consecutive rings. The results described here expand the binding site size targetable with Py-Im polyamides and provide structural elements that will facilitate the design of new polyamides targeted to a variety of DNA sequence.

**Publications:** Kelly, Baird, & Dervan *Proc. Natl. Acad. Sci. U.S.A.* **1996**, 93, 6981.  
Trauger, Baird, Mrksich & Dervan *J. Am. Chem. Soc.* **1996**, 118, 6160.  
Swalley, Baird & Dervan *Chem. Eur. J.* **1997**, 3, 1600.  
Trauger, Baird & Dervan *J. Am. Chem. Soc.* **1998**, 120, 3534.  
Keilkopf, Baird, Rees & Dervan *Nat. Struct. Biol* **1998**, 5, 104.

**Binding Site Size Limit of the 2:1 Pyrrole- Imidazole Polyamide-DNA Motif.** Recent examples of 2:1 polyamide-DNA complexes have created new models for the design of nonnatural ligands for recognition of a broad sequence repertoire in the minor groove of DNA.<sup>1</sup> The side-by-side combination of one Im ring on one ligand and a Py on the second ligand is specific for G•C, while a Py/Im pair targets a C•G base pair.<sup>1,2</sup> A Py/Py pair is partially degenerate and binds A•T or T•A base pairs.<sup>1,2</sup> The DNA sequence specificities of these small molecules can be controlled by the linear sequences of Py and Im amino acids. The three-ring polyamide Im-(Py)<sub>2</sub>-Dp was shown to specifically bind 5 base pair 5'-(A,T)G(A,T)C(A,T)-3' sequences while the four-ring polyamide ImPyImPy-Dp was shown to bind 6 base pair 5'-(A,T)GCGC(A,T)-3' sites as side-by-side antiparallel dimers in the minor groove.<sup>1,3</sup>

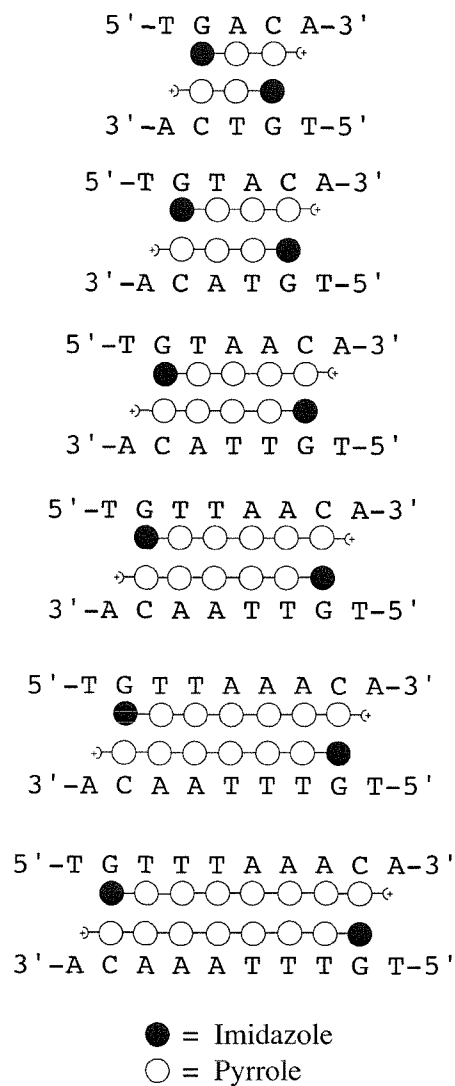
A major goal of our efforts in evaluating the scope and limitations of the 2:1 polyamide-DNA motif was to extend specific recognition to larger binding sites. In order to determine the effect of polyamide length on binding site size, binding affinity, and sequence specificity within the motif, a series of six Im-Py polyamides containing three to eight rings were synthesized. The polyamide series is based on Im-(Py)<sub>2</sub>-Dp (**1**) with pyrrolicarboxamide moieties added sequentially to the C-termini to afford Im-(Py)<sub>3</sub>-Dp (**2**), Im-(Py)<sub>4</sub>-Dp (**3**), Im-(Py)<sub>5</sub>-Dp (**4**), Im-(Py)<sub>6</sub>-Dp (**5**) and Im-(Py)<sub>7</sub>-Dp (**6**) which are designed to bind five to ten base pair sites, respectively, as side-by-side antiparallel dimers (Figures 3.1, 3.2 and 3.3). The DNA binding sites are based on a 5'-TGACA-3' core sequence and contain sequential A,T base pair inserts in the center of the binding sites that will be recognized by the additional pyrrolicarboxamides. This combination of polyamides and DNA binding site sequences was chosen to satisfy several criteria. The presence of G•C and C•G base pairs in the binding sites are expected to lock the polyamides in the designated binding sites by preventing them from binding in undesired slipped conformations on the DNA.<sup>4,5</sup> The alternating A,T tract in the center of the binding sites is designed to favor 2:1 binding in contrast to pure A-tracts which generally favor 1:1 polyamide-DNA complex formation.<sup>6,7</sup> The binding site size was determined by MPE•Fe(II) footprinting<sup>8</sup> and the apparent first order binding affinity and sequence specificity of each polyamide was



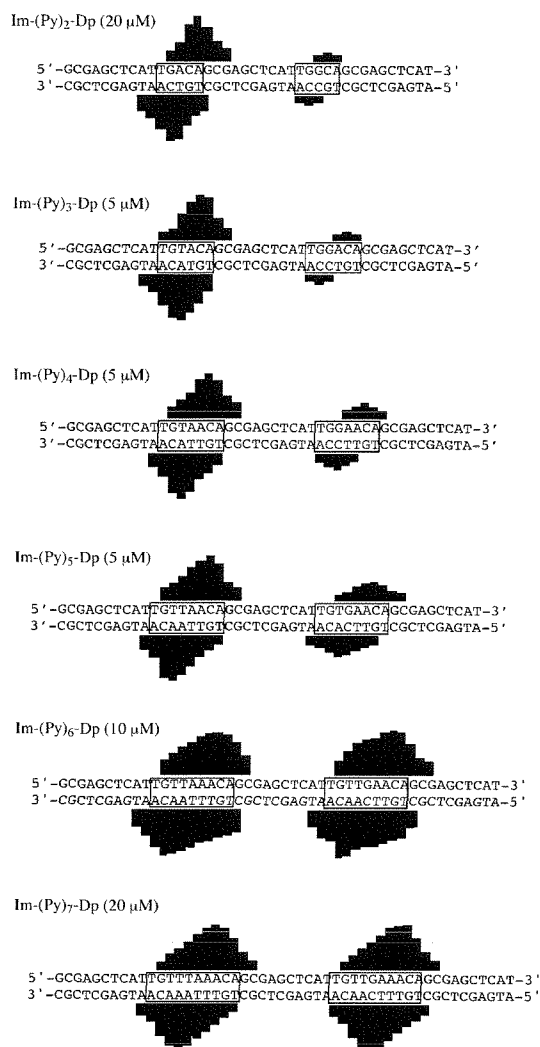
**Figure 3.1.** Structures of the imidazole-pyrrole containing polyamides Im-(Py)<sub>2</sub>-Dp, Im-(Py)<sub>3</sub>-Dp, Im-(Py)<sub>4</sub>-Dp, Im-(Py)<sub>5</sub>-Dp, Im-(Py)<sub>6</sub>-Dp, and Im-(Py)<sub>7</sub>-Dp.

determined by quantitative DNase I footprint titration experiments<sup>9</sup> on a series of restriction fragments containing a match site and a single base pair mismatch site for each polyamide.

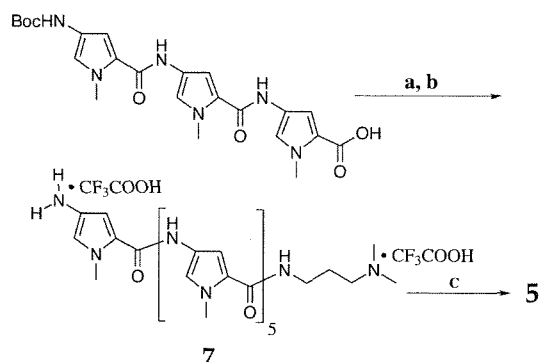
Polyamides **3-6** were prepared in two steps from previously described intermediates exemplified for Im-(Py)<sub>6</sub>-Dp **5** (Figure 3.4). Activation of *N*-(*tert*-butoxycarbonyl)-tris(*N*-methylpyrrole-carboxamide)<sup>10</sup> with HBTU followed by coupling to aminotris(*N*-methylpyrrole-carboxamide)<sup>11</sup> provided (*tert*-butoxycarbonyl)-hexa(*N*-methylpyrrolecarboxamide). The amine was deprotected by the addition of neat TFA to the crude reaction mixture, and the ditrifluoroacetate salt of aminohexa(*N*-methylpyrrolecarbox-amide) isolated by preparatory HPLC. Capping with 1-methylimidazole-2-carboxylic acid (**1**) (DCC/HOBt) provided Im-(Py)<sub>6</sub>-Dp (**5**) after HPLC purification.



**Figure 3.2.** Proposed 2:1 binding models of Im-(Py)<sub>2</sub>-Dp, Im-(Py)<sub>3</sub>-Dp, Im-(Py)<sub>4</sub>-Dp, Im-(Py)<sub>5</sub>-Dp, Im-(Py)<sub>6</sub>-Dp and Im-(Py)<sub>7</sub>-Dp binding to five, six, seven, eight, nine and ten base pair match sites, respectively. The imidazole and pyrrole rings are represented as shaded and unshaded spheres, respectively.



**Figure 3.3.** Histograms of cleavage protection from MPE•Fe(II) for Im-(Py)<sub>2</sub>-Dp, Im-(Py)<sub>3</sub>-Dp, Im-(Py)<sub>4</sub>-Dp, Im-(Py)<sub>5</sub>-Dp, Im-(Py)<sub>6</sub>-Dp and Im-(Py)<sub>7</sub>-Dp binding to the *Eco* RI/*Pvu* II restriction fragments from plasmids pJK5, pJK6, pJK7, pJK8, pJK9 and pJK10, respectively. The individual bar heights are proportional to the protection from MPE•Fe(II) cleavage at each nucleotide. The match and mismatch sites indicated by boxes.



**Figure 3.4.** Synthesis of ImPyPyPyPyPyPy-Dp: (a) (i) HBTU, DIEA; (ii) aminotris(1-methylpyrrole-2-carboxamide), DIEA, (b) TFA, (c) 1-methylimidazole-2-carboxylic acid, HOBT, DCC, DIEA.

pyrrolocarboxamide moieties to the C-termini of the polyamides increases the preferred DNA binding site size by one base pair (Figure 3.3). For all six polyamides the observed MPE•Fe(II) protection patterns are 3'-shifted consistent with 2:1 polyamide-DNA complex formation in the minor groove.

The apparent first-order binding affinities<sup>12</sup> for the match and single base pair mismatch sites for the six polyamides was determined by quantitative DNase I footprint titration experiments (10 mM Tris•HCl pH 7.0, 10 mM KCl, 5 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>, 22°C) in the absence of unlabeled carrier DNA. The  $\theta_{app}$  points for each polyamide were fit by a cooperative binding isotherm consistent with 2:1 polyamide-DNA complex formation.<sup>3</sup> Analysis of the match site data reveals that Im-(Py)<sub>2</sub>-Dp binds the five base pair 5'-TGACA-3' site with an affinity of  $1.3 \times 10^5 \text{ M}^{-1}$  while Im-(Py)<sub>3</sub>-Dp binds the six base pair 5'-TGTACA-3' site with an affinity of  $8.6 \times 10^6 \text{ M}^{-1}$ , corresponding to a 66-fold enhancement in affinity (Table I). The five-ring polyamide Im-(Py)<sub>4</sub>-Dp binds the seven base pair site 5'-TGTAACA-3' with an affinity of  $4.5 \times 10^7 \text{ M}^{-1}$ , a five-fold gain in affinity. Binding affinity levels off with Im-(Py)<sub>5</sub>-Dp binding the eight base pair 5'-TGTTAACA-3' site ( $K_a = 5.3 \times 10^7 \text{ M}^{-1}$ ) and Im-(Py)<sub>6</sub>-Dp binding the nine base pair 5'-TGTTAACA-3' site ( $K_a = 4.7 \times 10^7 \text{ M}^{-1}$ ) with the same affinity as Im-(Py)<sub>4</sub>-Dp for a seven base

Binding site sizes were determined by MPE•Fe(II) footprinting of **1-6** on six restriction fragments from plasmids pJK5-pJK10, respectively (Figure 2.3). Each restriction fragment contains a match and a single base pair mismatch site separated by 10 base pairs. The single base pair mismatch is generated by a G•C base pair replacing an A•T or T•A base pair in the center of the binding site. Analysis of the MPE•Fe(II) protection patterns reveals that the sequential addition of



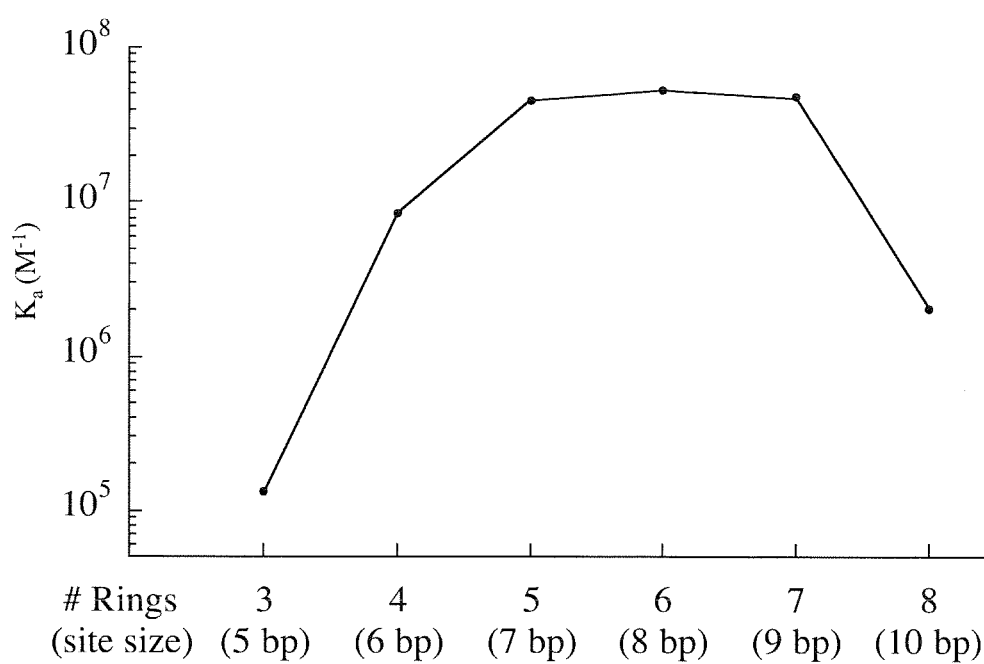
**Table 3.1.** Apparent First Order Binding Affinities ( $M^{-1}$ )<sup>a,b</sup>.

Polyamide	Rings	Binding Site Size (bp)	Match	Mismatch	Specificity <sup>c</sup>
Im-(Py) <sub>2</sub> -Dp	3	5	$1.3 \times 10^5$ (0.3)	$< 2 \times 10^4$ <sup>d</sup>	$> 6.5$
Im-(Py) <sub>3</sub> -Dp	4	6	$8.5 \times 10^6$ (1.3)	$1.6 \times 10^6$ (0.2)	5.3
Im-(Py) <sub>4</sub> -Dp	5	7	$4.5 \times 10^7$ (1.1)	$7.9 \times 10^6$ (1.8)	5.7
Im-(Py) <sub>5</sub> -Dp	6	8	$5.3 \times 10^7$ (0.5)	$< 2 \times 10^7$ <sup>d</sup>	$> 2.7$
Im-(Py) <sub>6</sub> -Dp	7	9	$4.7 \times 10^7$ (0.4)	$1.7 \times 10^7$ (0.7)	2.8
Im-(Py) <sub>7</sub> -Dp	8	10	$< 2 \times 10^6$	$< 2 \times 10^6$	1

<sup>a</sup> Values reported are the mean values from at least three footprint titration experiments. Numbers in parentheses indicate the standard deviation for each data set. <sup>b</sup> The assays were performed at 22°C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. <sup>c</sup> Defined as the ratio of the match site binding affinity to the binding affinity of the single base pair mismatch site.

pair site. Binding affinity drops dramatically for Im-(Py)<sub>7</sub>-Dp binding to a ten base pair 5'-TGTTTAAACA-3' site ( $K_a < 2 \times 10^6 M^{-1}$ ).

The 66-fold enhancement in apparent first-order binding affinity for Im-(Py)<sub>3</sub>-Dp binding to a six base pair 5'-TGTACA-3' site compared to Im-(Py)<sub>2</sub>-Dp binding to a five base pair 5'-TGACA-3' site corresponds to an increase in the free energy of binding ( $\Delta\Delta G$ ) of 2.4 kcal/mol (22°C). This increase in the free energy of binding is likely a result of the additional Py/Py pair in the Im-(Py)<sub>3</sub>-Dp dimer recognizing the T•A base pair in the context of the 5'-TGTACA-3' sequence. A significant fraction of the increase in binding affinity for Im-(Py)<sub>3</sub>-Dp presumably results from the hydrogen bonds and van der Waals contacts that likely form when the additional pyrrolicarboxamides in the ligand bind the additional T•A base pair in the core of the binding site. The five-fold enhancement in affinity of Im-(Py)<sub>4</sub>-Dp for the seven base pair 5'-TGTAACA-3' site over Im-(Py)<sub>3</sub>-Dp binding to 5'-TGTACA-3' corresponds to an increase in the free energy of binding of 1.0 kcal/mol. This relatively modest change in binding free energy suggests that the addition of a pyrrolicarboxamide moiety to the four ring Im-(Py)<sub>3</sub>-Dp is less favorable. The observation that the affinities of Im-(Py)<sub>5</sub>-Dp and Im-(Py)<sub>6</sub>-Dp for the 8 and 9 base pair 5'-TGTTAACA-3' and 5'-TGTTAAACA-3' sites, respectively, are the same as Im-(Py)<sub>4</sub>-Dp binding



**Figure 3.5.** Effects of polyamide length on DNA-binding affinity. Data taken from table I. Polyamides are of the form Im(Py)<sub>x</sub>-Dp, binding sites are of the form 5'-WG(W)<sub>x-1</sub>CW-3'.

5'-TGTAACA-3' indicates that further lengthening of the polyamides from five to seven rings and the DNA binding site from seven to nine base pairs has little energetic benefit. The dramatic decrease in binding affinity of Im-(Py)<sub>7</sub>-Dp for the ten base pair 5'-TGTTTAAACA-3' site ( $K_a < 2 \times 10^6 M^{-1}$ ) suggests the presence of an additional pyrrolicarboxamide in the ligand creates an unfavorable contribution to the binding energy of Im-(Py)<sub>7</sub>-Dp for the 5'-TGTTTAAACA-3' site.

Comparison of the apparent first-order binding affinities for the match and the single base pair mismatch site for each polyamide demonstrates that the specificity of 2:1 polyamide-DNA complex formation, defined as the ratio of match site binding affinity to the affinity for the single base pair mismatch site, generally decreases with increasing polyamide length (Table I). Im-(Py)<sub>2</sub>-Dp prefers binding the 5'-TGACA-3' site over the 5'-TGGCA-3' site by a factor of at least 6.5. Im-(Py)<sub>3</sub>-Dp and Im-(Py)<sub>4</sub>-Dp bind the mismatch sites with five- and six-fold lower affinity, respectively. Im-(Py)<sub>5</sub>-Dp displays at least a three-fold preference for the match site while Im-(Py)<sub>6</sub>-Dp binds the mismatch site with 2.7-fold lower affinity. At concentrations below 1  $\mu M$  Im-(Py)<sub>7</sub>-Dp shows no binding to DNA. At concentrations 1  $\mu M$  and above equal protection from

cleavage by DNase I of both the match and mismatch sites are observed indicating that the specificity as defined here is approximately one.

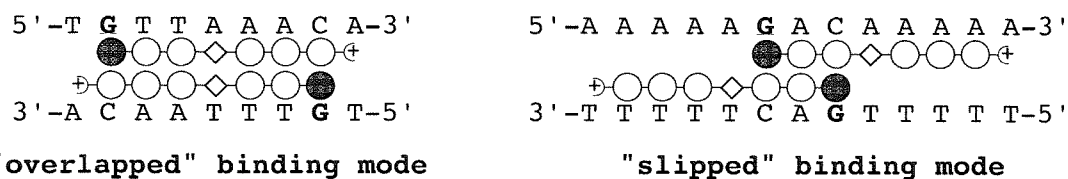
One explanation for these observations is that three and four ring polyamides Im-(Py)<sub>2</sub>-Dp and Im-(Py)<sub>3</sub>-Dp, respectively, are in phase with their DNA binding sites. For the longer polyamides beginning with Im-(Py)<sub>4</sub>-Dp where binding affinities level off at Im-(Py)<sub>5</sub>-Dp and ultimately decrease at Im-(Py)<sub>7</sub>-Dp the ligands may be falling out of register with the DNA binding sites. In these 2:1 complexes the curvature of the polyamides does not match the curvature of the minor groove surfaces of the longer binding sites.<sup>13</sup> One likely consequence would be that the energetic benefit of the hydrogen bonds and van der Waals interactions that stabilize the 2:1 complexes would decrease as the lack of correspondence between the binding surfaces of the polyamides and the DNA sites became more pronounced. In the mismatch sites a G•C base pair replaces an A•T or T•A base in the core of the binding site. The 2-amino group of guanine protrudes from the floor of the minor groove and therefore is expected to introduce a bump on the surface of the binding sites that may interfere with the hydrogen bonding of the Py NH of the polyamides to N3 of the guanine base. If the longer polyamides are out of register with the DNA, then the effect of this unfavorable interaction on the free energy of binding at the mismatch site may become less significant. This would lower the overall free energy difference between binding the match and mismatch sites and thereby reduce specificity.

We note that sequence-dependent DNA structural features such as intrinsic minor groove width, minor groove flexibility, and inherent curvature may differ between each of the binding sites and could therefore contribute to the measured difference in the binding affinities of the six polyamides.<sup>14</sup> One possibility is that as the length of the A,T-tract in the binding sites increases, the minor groove width and/or flexibility may decrease which would likely impose an energetic penalty for 2:1 polyamide-DNA complex formation and lower the apparent binding affinities for these sites. These effects may also contribute to the lower sequence specificity observed for the longer polyamides in the series. Insertion of G•C base pairs into an A•T tract may increase the minor groove width and/or flexibility which may reduce the difference in binding free energy

between the match and mismatch sites thereby lowering specificity. These considerations of sequence-dependent DNA structure suggest that the sequence composition of the DNA target site may be an important factor in determining polyamide affinity and specificity in the 2:1 motif.

The results of this study demonstrate that DNA sequences up to nine base pairs in length can be specifically recognized by Py-Im containing polyamides containing three to seven rings by 2:1 polyamide-DNA complex formation in the minor groove. Recognition of a nine base pair site defines the new lower limit of the binding site size that can be recognized by polyamides containing exclusively Im and Py amino acids. The binding affinity reaches a maximum value for the five-ring polyamide Im-(Py)<sub>4</sub>-Dp and addition of up to two additional pyrrolicarboxamides has no effect on the observed binding affinity. These results and the failure of an eight-ring polyamide to specifically recognize a ten base pair site suggests that a new class of polyamides is needed for extension of the 2:1 polyamide-DNA motif to sequences longer than nine base pairs. Replacement of a central Py or Im amino acid with a more flexible spacer amino acid subunit should allow the antiparallel dimer to reset the register for continued gain in affinity and specificity.

**Extension of Sequence-Specific Recognition in the Minor Groove of DNA by Pyrrole-Imidazole Polyamides to 9-13 Base-Pairs.** The binding affinity and sequence-specificity of a non-covalent antiparallel homodimeric or heterodimeric polyamide-DNA complex can be increased by covalently linking the two polyamides.<sup>11,15,16</sup> The DNA-binding properties of the polyamides ImPyPy-G-PyPyPy-Dp, ImPyPy- $\beta$ -PyPyPy-Dp, and ImPyPy- $\gamma$ -PyPyPy-Dp, in which the terminal carboxyl group of ImPyPy and the terminal amine of PyPyPy-Dp are connected with glycine (G),  $\beta$ -alanine ( $\beta$ ), and  $\gamma$ -aminobutyric acid ( $\gamma$ ), respectively, were reported.<sup>11</sup> The  $\gamma$ -aminobutyric acid-linked polyamide bound the designated target site 5'-TGTTA-3' with high

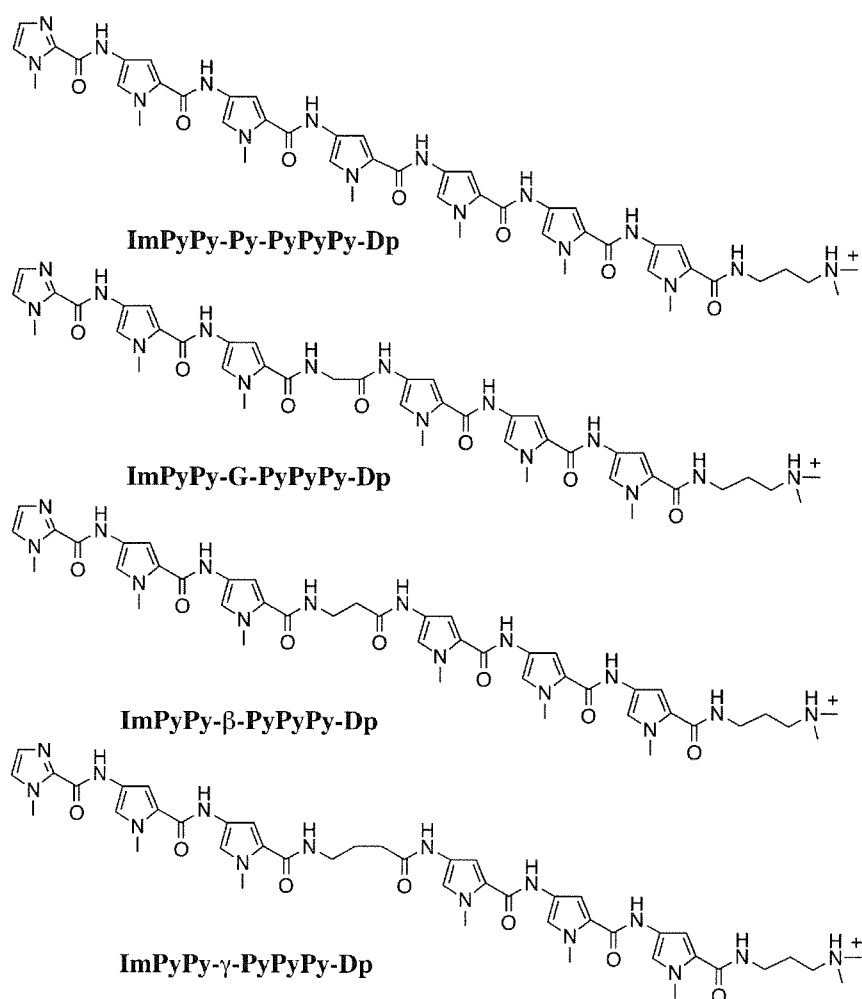


**Figure 3.6.** Expected complexes of ImPyPy-X-PyPyPy-Dp, where X = Py, G, and  $\beta$ , with 5'-TGTTAAACA-3' and 5'-AAAAAGACAAAA-3'. The shaded and light circles represent Im and Py rings, respectively, and the diamond represents the internal amino acid X. The specifically targeted guanines are highlighted.

affinity and sequence-specificity, and exhibited a Langmuir binding isotherm in quantitative footprinting experiments consistent with formation of an intramolecular “hairpin” complex in which the polyamide folds back on itself.<sup>11</sup> Modeling suggested that the glycine- and  $\beta$ -alanine-linked polyamides could not favorably bind as “hairpins” in the minor groove of DNA. Moreover, these polyamides exhibited cooperative binding isotherms in quantitative footprinting experiments, consistent with two polyamides binding in extended conformations as *intermolecular dimers*.<sup>3,4</sup> It appears that the glycine- and  $\beta$ -alanine-linked polyamides disfavor binding in the hairpin conformation and are restricted to binding in an extended conformation. In a formal sense, there are multiple extended binding motifs (and hence multiple binding site sequences) for polyamides of sequence composition Im(Py)<sub>x</sub>-Dp, as discussed below.

We report here the DNA-binding affinities of four polyamides having the general sequence ImPyPy-X-PyPyPy-Dp, where X = Py, G,  $\beta$ , or  $\gamma$ , to the 9 bp site 5'-TGTTAAACA-3' and to the 13 bp sites 5'-AAAAAGACAAAA-3' and 5'-ATATAGACATATA-3'. The polyamides having internal Py, glycine and  $\beta$ -alanine residues were anticipated to bind the 9 and 13 bp sites in an extended conformation. It was not clear at the outset if the  $\gamma$ -aminobutyric acid-linked polyamide ImPyPy- $\gamma$ -PyPyPy-Dp would bind in an extended or “hairpin” conformation to the targeted sites.

For ImPyPy-X-PyPyPy-Dp polyamides binding in an extended conformation, the polyamide-DNA complexes expected to form at the 9 bp and 13 bp target sites represent two distinct binding modes, which we refer to as “overlapped” and “slipped,” respectively. In the



**Figure 3.7.** Structures of the four polyamides ImPyPy-X-PyPyPy-Dp, where X = *N*-methylpyrrole (Py), glycine (G),  $\beta$ -alanine ( $\beta$ ) and  $\gamma$ -aminobutyric acid ( $\gamma$ ).

“overlapped” (9 bp) binding mode, two ImPyPy-X-PyPyPy-Dp polyamides bind directly opposite one another (Figures 3.6 and 3.7).

The “slipped” (13 bp) binding mode integrates the 2:1 and 1:1 polyamide-DNA binding motifs at a single site. In this binding mode, the ImPyPy moieties of two ImPyPy-X-PyPyPy-Dp polyamides bind the central 5'-AGACA-3' sequence in a 2:1 manner as in the ImPyPy homodimer,<sup>1</sup> and the PyPyPy moieties of the polyamides bind the all-A,T flanking sequences as in the 1:1 complexes of distamycin (Figures 3.7).<sup>17</sup> The structure of the complex formed by the polyamide ImPyPy-G-PyPyPy-Dp with a 13 bp target site has been characterized by 2D NMR.<sup>4</sup>



**Table 5.2.** Apparent First Order Association Constants ( $M^{-1}$ ) for Polyamides ImPyPy-X-PyPyPy-Dp, Where X = *N*-methylpyrrole (Py), Glycine (G),  $\beta$ -alanine ( $\beta$ ), and  $\gamma$ -aminobutyric acid ( $\gamma$ ).<sup>a,b</sup>

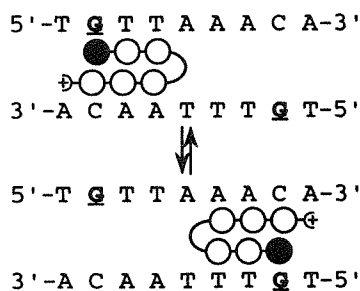
Binding Site	Polyamide			
	X = Py	X = G	X = $\beta$	X = $\gamma$
5'-TGTTAAACA-3'	$9.7 (\pm 2.3) \times 10^7$	$1.4 (\pm 0.1) \times 10^8$	<b><math>7.8 (\pm 0.6) \times 10^8</math></b>	$1.4 (\pm 0.3) \times 10^8$
5'-AAAAAGACAAAAA-3'	$5.4 (\pm 1.5) \times 10^7$	$1.1 (\pm 0.1) \times 10^8$	<b><math>&gt; 4.7 (\pm 0.7) \times 10^9</math></b>	$6.4 (\pm 0.6) \times 10^6$
5'-ATATAGACATATA-3'	$3.6 (\pm 0.5) \times 10^7$	$6.6 (\pm 0.4) \times 10^6$	<b><math>1.0 (\pm 0.1) \times 10^9</math></b>	$4.6 (\pm 0.5) \times 10^6$

<sup>a</sup>The reported association constants are the mean values obtained from three DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses. <sup>b</sup>The assays were carried out at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

distamycin analog Ac-PyPyPy-Dp (Ac = acetyl) was shown to bind the sites 5'-AATAA-3' and 5'-TTAAT-3' with 2-fold and 14-fold lower affinity, respectively, relative to the site 5'-AAAAA-3'.<sup>1c</sup> On the basis of this result, we anticipated that the 13 bp site 5'-AAAAAGACAAAAA-3' may be bound with higher affinity than the 13 bp site 5'-ATATAGACATATA-3'.

The binding affinities of the four polyamides (Figure 3.8) for the three targeted sites 5'-TGTTAAACA-3', 5'-AAAAAGACAAAAA-3', and 5'-ATATAGACATATA-3' were determined by quantitative DNase I footprint titration experiments.<sup>9</sup> Quantitative DNase I footprinting on the 3'-<sup>32</sup>P-labelled 281-bp pJT3 *Afl* II/*Fsp* I restriction fragment (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>, pH 7.0, 22°C) reveals that of the four polyamides ImPyPy-X-PyPyPy-Dp, three (X = Py, G,  $\beta$ ) bind to both the 9 bp “overlapped” site 5'-TGTTAAACA-3' and the 13 bp “slipped” site 5'-AAAAAGACAAAAA-3' with affinities greater than  $5 \times 10^7 M^{-1}$  (Table 2),<sup>5</sup> and display cooperative binding isotherms at these sites *consistent with binding as intermolecular dimers*. The fact that the polyamides ImPyPy-G-PyPyPy-Dp and ImPyPy- $\beta$ -PyPyPy-Dp bind in the 9 bp “overlapped” binding mode *indicates that the internal glycine and  $\beta$ -alanine amino acids are accommodated opposite a second ligand in a 2:1 polyamide-DNA complex*.





**Figure 3.9.** Model for the complex formed by ImPyPy- $\gamma$ -PyPyPy-Dp with 5'-TGTAAACA-3' (5 bp, "hairpin").

The polyamide ImPyPy- $\gamma$ -PyPyPy-Dp binds the site 5'-TGTAAACA-3' with an affinity of  $1 \times 10^8 \text{ M}^{-1}$ , and also binds the site 5'-AAAAAGACAAAAA-3' with lower affinity ( $6 \times 10^6 \text{ M}^{-1}$ ). This compound displays Langmuir binding isotherms at these sites, *consistent with binding as an intramolecular hairpin* to the 5-bp "matched" sites 5'-TGTTA-3' and 5'-AAACA-3' (Figure 3.9) and to the 5-bp "single base-

pair mismatch" site 5'-AGACA-3'. Significantly, it appears from these results that ImPyPy- $\gamma$ -PyPyPy-Dp does not effectively link polyamide subunits in an extended conformation.

Comparison of the binding affinities of the four polyamides ImPyPy-X-PyPyPy-Dp, where X = Py, G,  $\beta$ , and  $\gamma$ , reveals that the internal amino acid X has a dramatic effect on complex stability (Table 5.2). The formally Py-linked polyamide ImPyPy-Py-PyPyPy-Dp binds 5'-TGTAAACA-3' and 5'-AAAAAGACAAAAA-3' with affinities of  $1 \times 10^8 \text{ M}^{-1}$  and  $5 \times 10^7 \text{ M}^{-1}$ , respectively. The glycine-linked polyamide ImPyPy-G-PyPyPy-Dp binds both 5'-TGTAAACA-3' and 5'-AAAAAGACAAAAA-3' with affinities similar (equal and 2-fold higher, respectively) to ImPyPy-Py-PyPyPy-Dp. In contrast, the  $\beta$ -alanine linked polyamide ImPyPy- $\beta$ -PyPyPy-Dp binds 5'-TGTAAACA-3' and 5'-AAAAAGACAAAAA-3' with affinities higher than ImPyPy-Py-PyPyPy-Dp by factors of approximately 8 ( $K_a = 8 \times 10^8 \text{ M}^{-1}$ ) and 85 ( $K_a = 5 \times 10^9 \text{ M}^{-1}$ ), respectively. Relative to ImPyPy-Py-PyPyPy-Dp, the hairpin-forming polyamide ImPyPy- $\gamma$ -PyPyPy-Dp binds 5'-TGTAAACA-3' (a matched hairpin binding site) and 5'-AAAAAGACAAAAA-3' (a mismatched hairpin binding site) with equal and 8-fold lower affinities, respectively.

Comparison of the binding affinities of the four polyamides at the 13 bp sites 5'-AAAAAGACAAAAA-3' and 5'-ATATAGACATATA-3' indicates that the specificity between

the sites depends on the internal amino acid (Table 2). ImPyPy-G-PyPyPy-Dp and ImPyPy- $\beta$ -PyPyPy-Dp are approximately 20-fold and 5-fold specific, respectively, for 5'-AAAAAGACAAAAA-3' versus 5'-ATATAGACATATA-3'. The polyamides ImPyPy-PyPyPy-Dp and ImPyPy- $\gamma$ -PyPyPy-Dp bind 5'-AAAAAGACAAAAA-3' and 5'-ATATAGACATATA-3' with similar affinities.

These results show that  *$\beta$ -alanine is an optimal linker for joining two three-ring subunits in an extended conformation, providing a useful structural motif for the design of new polyamides targeted to sites longer than 7 bp*. Recently, it has been shown that as the length of a polyamide having the general sequence Im(Py)<sub>x</sub>-Dp increases beyond 5 rings (corresponding to a 7 bp binding site), the binding affinity ceases to increase with increasing polyamide length, indicating that the Im and Py residues fail to maintain an appropriate base-pair register across the entire length of the polyamide-DNA complex.<sup>18</sup> The higher binding affinities observed here for ImPyPy- $\beta$ -PyPyPy-Dp relative to ImPyPy-Py-PyPyPy-Dp indicate that the flexible  $\beta$ -alanine linker relieves the register mismatch, allowing both three-ring subunits to bind optimally. Notably, higher binding affinities are observed for ImPyPy- $\beta$ -PyPyPy-Dp versus ImPyPy-Py-PyPyPy-Dp despite the higher conformational entropy and lower aromatic surface area of the  $\beta$ -alanine-linked polyamide. The observation here that  $\beta$ -alanine effectively links polyamide subunits within 2:1 polyamide-DNA complexes is consistent with the previously reported finding that  $\beta$ -alanine effectively links polyamide subunits within 1:1 polyamide-DNA complexes.<sup>19</sup>

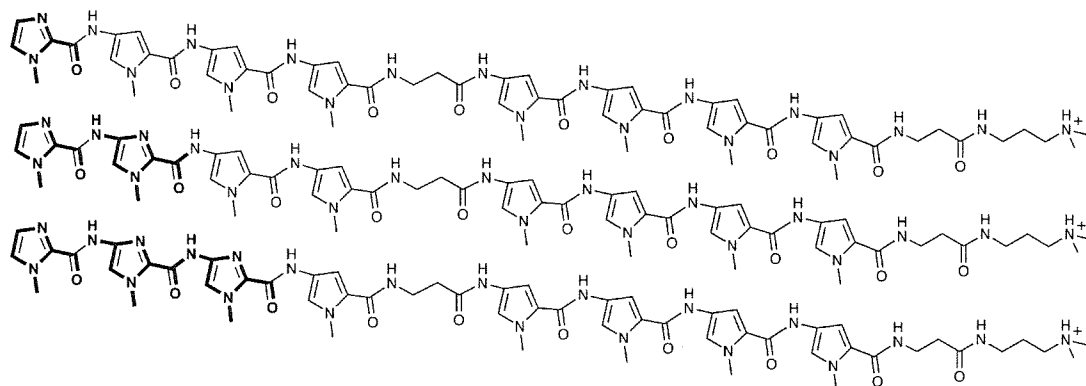
From the standpoint of binding specificity, the observation here that a single compound can bind in multiple binding modes is problematic. The next generation of Py-Im polyamides targeted to binding sites greater than 8 bp in length should incorporate constraints that specify a single binding mode.

The previously described  $\gamma$ -aminobutyric acid-based “hairpin” motif<sup>11</sup> complements the  $\beta$ -alanine-based “extended” motif described here. For ImPyPy- $\gamma$ -PyPyPy-Dp, the binding

isotherms and affinities observed here are consistent with the previous report that  $\gamma$ -aminobutyric acid effectively links polyamide subunits in a “hairpin” conformation<sup>11</sup> and indicate that  $\gamma$ -aminobutyric acid does not effectively link polyamide subunits in an extended conformation. Significantly, these observations indicate that (1) extended binding modes will not compromise the sequence-specificity of hairpin-forming,  $\gamma$ -aminobutyric acid-linked polyamides, and (2)  $\beta$ -alanine and  $\gamma$ -aminobutyric acid linkers could be used within a single polyamide with predictable results.

**A Pyrrole-Imidazole Polyamide Motif for Recognition of 11 Base Pair Sequences in the Minor Groove of DNA.** Fully overlapped 2:1 polyamide dimers containing five contiguous rings and solely Py and Im amino acids optimally recognize a maximum binding site size of 7 base pairs.<sup>18</sup> Insertion of a flexible  $\beta$ -alanine ( $\beta$ ) linker allows the recognition of longer sites.<sup>5</sup> Six-ring polyamides based on three-ring subunits linked C-N by a central  $\beta$ -alanine bind to 9 base pair sequences as “overlapped”  $(3\text{-}\beta\text{-}3)_2\cdot\text{DNA}$  complexes.<sup>5</sup> The observation that four-ring polyamides bind as 2:1 complexes with 70-fold higher affinity relative to three-ring polyamides suggests an eight-ring  $(4\text{-}\beta\text{-}4)_2\cdot\text{DNA}$  motif for recognition of 11 base pair sequences, defining a new upper limit for minor groove binding by fully-overlapped polyamide dimers.

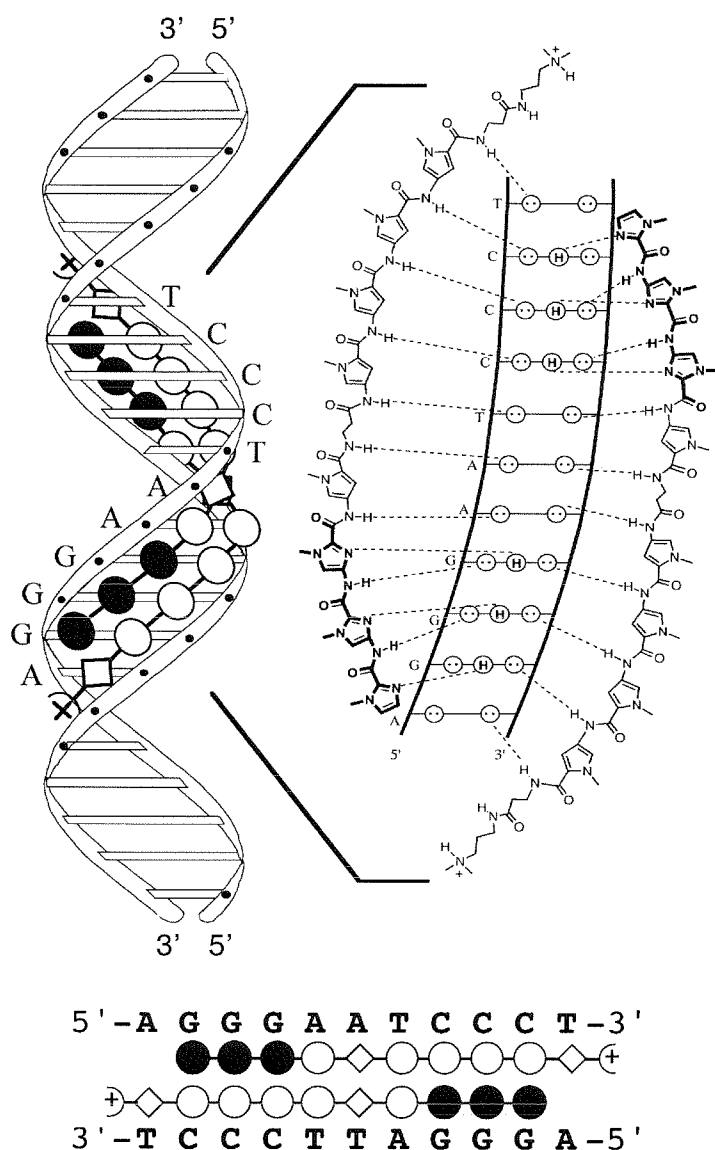
Three 4- $\beta$ -4 polyamides, ImPyPyPy- $\beta$ -PyPyPyPy- $\beta$ -Dp, ImImPyPy- $\beta$ -PyPyPyPy- $\beta$ -Dp and ImImImPy- $\beta$ -PyPyPyPy- $\beta$ -Dp were synthesized by solid phase methods (Figure 3.10 and 3.11).<sup>20</sup> We report here the affinities, relative sequence specificity, and binding site size of the three 4- $\beta$ -4 polyamides as determined by MPE•Fe(II)<sup>8</sup> and DNase I<sup>9</sup> footprinting. Binding site



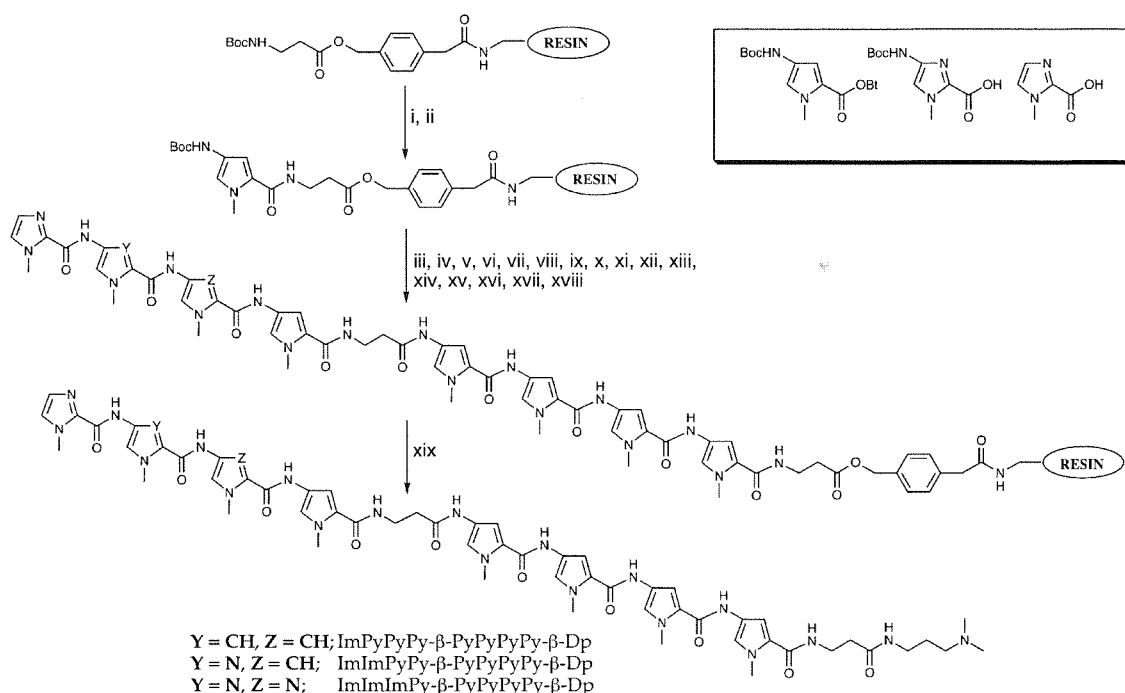
**Figure. 3.10.** Structures of ImPyPyPy-β-PyPyPyPy-β-Dp (Top), ImImPyPy-β-PyPyPyPy-β-Dp (Middle), and ImImImPy-β-PyPyPyPy-β-Dp (Bottom).

size is more accurately determined by MPE•Fe(II) footprinting, while quantitative DNase I footprint titration is more suitable for measurement of apparent equilibrium association constants ( $K_a$ ) for the polyamide binding to match and mismatch sequences. Finally, the sequence preference of the central β-β pairing was examined by testing the polyamide ImImImPy-β-PyPyPyPy-β-Dp against two potential 11-bp match sites, 5'-AGGGAATCCCT-3' and 5'-AGGGAGTCCCT-3' which differ by a single central base pair.

The polyamides ImPyPyPy-β-PyPyPyPy-β-Dp, ImImPyPy-β-PyPyPyPy-β-Dp, and ImImImPy-β-PyPyPyPy-β-Dp were synthesized in 18 steps each from Boc-β-alanine-Pam resin (1 g resin/0.2 mmol/g substitution) using previously described solid phase methods (Fig. 3).<sup>20</sup> A sample of polyamide-resin (240 mg) was cleaved by a single step aminolysis reaction with ((dimethylamino)propyl)amine (55°C, 18 h) and subsequently purified by reversed phase HPLC chromatography. The trifluoroacetate salts of eight-ring 4-β-4 polyamides are stable at room temperature, and soluble in aqueous solution at concentrations < 1 mM.

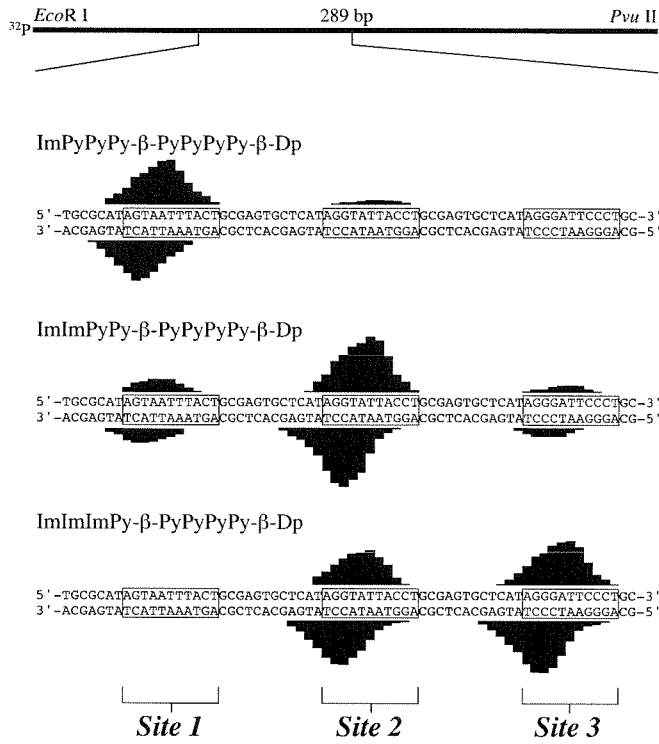


**Figure 3.11.** Models of the 2:1 polyamide:DNA complex formed with the eight-ring polyamide ImImImPy-β-PyPyPyPy-β-Dp and the minor groove of double stranded B-form DNA. (left) Ribbon model for the 2:1 complex formed between the match sequence 5'-AGGGAATCCCT-3' and ImImImPy-β-PyPyPyPy-β-Dp. *N*-methylimidazole carboxamides are represented by unfilled circles, *N*-methylpyrrole carboxamides are represented by empty circles, and β-alanine amino acids are represented by unfilled triangles. (right) Hydrogen-bonding model for the 2:1 complex formed between the match sequence 5'-AGGGAATCCCT-3' and ImImImPy-β-PyPyPyPy-β-Dp. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. (bottom) Ball and stick representation of the dimerization of ImImImPy-β-PyPyPyPy-β-Dp with a 5'-AGGGAATCCCT-3' site. The amino acids are represented as in the ribbon model.



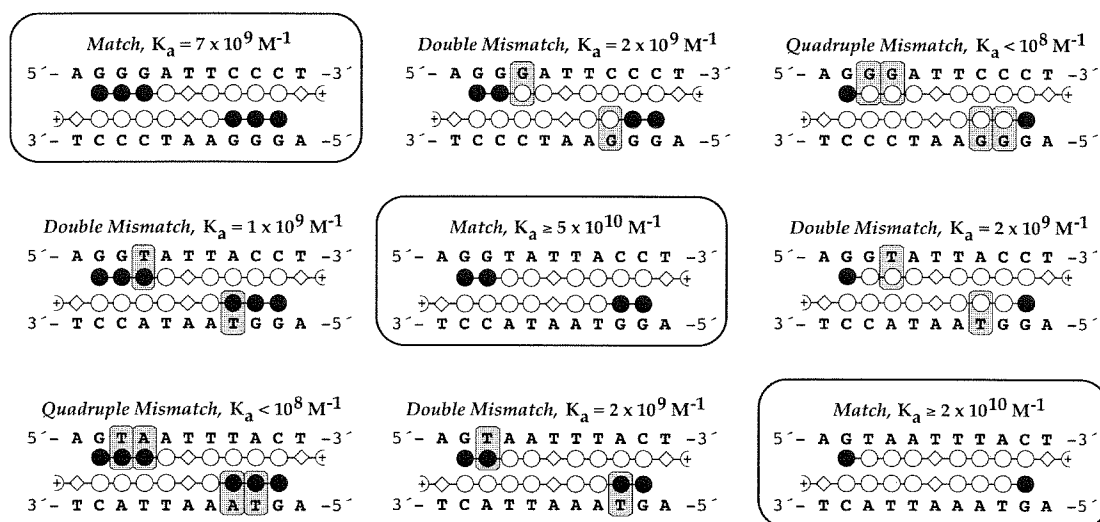
**Figure. 3.12.** (Box) Pyrrole and imidazole monomers for polyamide synthesis; Boc-Pyrrole-OBt ester, Boc-Imidazole-acid, and imidazole-2-carboxylic acid.<sup>1a</sup> Solid phase synthetic scheme for ImPyPyPy-β-PyPyPyPy-β-Dp, ImImPyPy-β-PyPyPyPy-β-Dp, and ImImImPy-β-PyPyPyPy-β-Dp prepared from commercially available Boc-β-alanine-Pam-resin (0.2 mmol/gram): (i) 80% TFA/DCM, 0.4M PhSH; (ii) BocPy-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4M PhSH; (iv) BocPy-OBt, DIEA, DMF; (v) 80% TFA/DCM, 0.4M PhSH; (vi) BocPy-OBt, DIEA, DMF; (vii) 80% TFA/DCM, 0.4M PhSH; (viii) BocPy-OBt, DIEA, DMF; (ix) 80% TFA/DCM, 0.4M PhSH; (x) Boc-β-alanine (HBTU, DIEA), DMF; (xi) 80% TFA/DCM, 0.4M PhSH;3233 (xii) BocPy-OBt, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4M PhSH; (xiv) BocPy-OBt, DIEA, DMF; (for **1** and **2**); BocIm-OBt (DCC/HOBt), DIEA, DMF (for **3**) (xv) 80% TFA/DCM, 0.4M PhSH; (xvi) BocPy-OBt, DIEA, DMF (for **1**); BocIm-OBt (DCC/HOBt), DIEA, DMF (for **2** and **3**) (xvii) 80% TFA/DCM, 0.4M PhSH; (xviii) imidazole-2-carboxylic acid (HBTU/DIEA); (xix) *N,N*-((dimethylamino)propyl)amine, 55°C.

A plasmid was constructed, pSES123, which contains match sites according to the pairing rules for each of the three polyamides: 5'-atAGTAATTTACTgc-3' (site 1), 5'-atAGGTATTACCTgc-3' (site 2), and 5'-atAGGGATTCCCTgc-3' (site 3). MPE•Fe(II) footprinting on the 3'- and 5'-<sup>32</sup>P end-labeled 289 base pair restriction fragment from this plasmid (25 mM Tris-acetate, 10 mM NaCl, 100 μM calf thymus DNA, pH 7.0 and 22°C) reveals



**Figure 3.13.** (Top) Illustration of the 289 bp restriction fragment with the position of the sequence indicated. Bar heights are proportional to the relative protection from cleavage at each band. Boxes represent equilibrium binding sites determined by the published model, and only sites that were quantitated by DNase I footprint titrations are boxed. Shown with brackets are Site 1, 5'-AGTAATTTACT-3'; Site 2, 5'-AGGTATTACCT-3'; and Site 3, 5'-AGGGATTCCCT-3'.

that all three polyamides bind to their designated match sites (Figure 3.13). The polyamide ImPyPyPy-β-PyPyPyPy-β-Dp which contains a single Im amino acid residue binds to its match sequence 5'-AGTAATTTACT-3' (site 1) and weakly to the double mismatch sequence 5'-AGGTATTACCT-3' (site 2) but not to the quadruple mismatch sequence, 5'-AGGGATTCCCT-3' (site 3). Similarly, polyamide ImImPyPy-β-PyPyPyPy-β-Dp reveals a strong footprint to its match sequence 5'-AGGTATTACCT-3', (site 2) and only weak footprints on the double mismatch sequences, sites 1 and 3. Lastly, the polyamide with three contiguous Im residues ImImImPy-β-PyPyPyPy-β-Dp recognizes its match sequence, 5'-AGGGATTCCCT-3' (site 3) and its double mismatch sequence, site 2, but not site 1. The size of the footprint cleavage protection for the polyamides are consistent with binding sites 11 bp in size.



**Figure 3.14.** Ball and stick models of ImPyPyPy-β-PyPyPyPy-β-Dp (left), ImImPyPy-β-PyPyPyPy-β-Dp (middle), and ImImImPy-β-PyPyPyPy-β-Dp (right) for each binding site, with the corresponding equilibrium association constants shown above each individual model. The binding sites shown are 5'-AGTAATTTACT-3' (top), 5'-AGGTATTACCT-3' (middle), and 5'-AGGGATTCCCT-3' (bottom). Shaded and nonshaded circles denote Im and Py carboxamides, respectively. Shaded diamonds represent the β-alanine residue. Formally mismatched base pairs are boxed. Values reported are the mean values measured from four DNase I footprinting titration experiments, with the standard deviation for each data set indicated in parentheses. The assays were performed at 22°C at pH 7.0 in the presence of 10 mM tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

Quantitative DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub> and 5 mM CaCl<sub>2</sub>, pH 7.0 and 22°C)<sup>9</sup> were performed in order to determine the apparent equilibrium association constants ( $K_a$ ) of the polyamides for the three designed sites. ImPyPyPy-β-PyPyPyPy-β-Dp recognizes its match site, 5'-atAGTAATTTACTgc-3', with an apparent equilibrium association constant of  $K_a = 2.4 \times 10^{10} M^{-1}$ . The double mismatch 5'-atAGGTATTACCTgc-3' is bound with 10-fold lower affinity. ImImPyPy-β-PyPyPyPy-β-Dp, binds its match site 5'-atAGGTAATACCTgc-3' with the highest affinity of any of the three polyamides studied,  $K_a = 5.4 \times 10^{10} M^{-1}$ . The double mismatches 5'-atAGGGATTCCCTgc-3' and 5'-atAGTAATTTACTgc-3' are bound with at least 27-fold and 30-fold lower affinity, respectively. ImImImPy-β-PyPyPyPy-β-Dp binds its match site 5'-atAGGGATTCCCTgc-3'



with  $K_a = 7.4 \times 10^9 \text{ M}^{-1}$ . The double base pair mismatch, 5'-atAGGTAATACCTgc-3', is bound with 7-fold lower affinity.

A plasmid containing two 11-bp sites which differ by a single base pair in the sixth position, 5'-aaAGGGAATCCCTta-3' and 5'-aaAGGGAGTTCCCTta-3', was constructed in order to explore the sequence-specificity of a  $\beta/\beta$  pairing within the polyamide ImImImPy- $\beta$ -PyPyPyPy- $\beta$ -Dp. The relative binding affinity of ImImImPy- $\beta$ -PyPyPyPy- $\beta$ -Dp for the sites was determined on a 3'-<sup>32</sup>P end-labeled 294 base pair restriction fragment derived from this plasmid. Quantitative DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub> and 5 mM CaCl<sub>2</sub>, pH 7.0 and 22°C) reveal the apparent equilibrium association constants ( $K_a$ ) of the polyamide for the two sites. ImImImPy- $\beta$ -PyPyPyPy- $\beta$ -Dp binds the site with the A•T base pair in the sixth position with an equilibrium association constant of  $K_a = 1.4 \times 10^{10} \text{ M}^{-1}$ . The sequence 5'-AGGGAGTTCCCT-3', which has a G•C base pair in the center position opposite the  $\beta$ - $\beta$  pair, is bound with 20-fold lower affinity ( $K_a = 6.9 \times 10^8 \text{ M}^{-1}$ ) (Figure 3.15). Binding of polyamide **3** to an additional double mismatch site, 5'-AGGTATTACCT-3', also occurs with reduced affinity ( $K_a = 3.1 \times 10^9 \text{ M}^{-1}$ ).

MPE•Fe(II) footprinting reveals that the three 4- $\beta$ -4 polyamides bind with highest affinity to only binding sites expected from the pairing rules (Figure 3.13). The asymmetric 3'-shifts in the footprint protection patterns are consistent with minor groove binding.<sup>8</sup> DNase I footprint titrations of the polyamides containing 1, 2 or 3 Im residues demonstrate the versatility of the dimeric 4- $\beta$ -4 motif. Each polyamide binds its respective target site with apparent equilibrium association constants ranging from  $K_a = 7 \times 10^9 \text{ M}^{-1}$  to  $K_a = 5 \times 10^{10} \text{ M}^{-1}$ , and with DNA mismatch specificities ranging from 7-fold to > 30-fold. Polyamides ImPyPyPy- $\beta$ -PyPyPyPy- $\beta$ -Dp and ImImPyPy- $\beta$ -PyPyPyPy- $\beta$ -Dp each recognize their respective target sites containing 4 or 6 G,C base pairs with  $K_a = 2 \times 10^{10} \text{ M}^{-1}$ . ImImImPy- $\beta$ -PyPyPyPy- $\beta$ -Dp, which contains an ImImImPy sequence component, binds with 3-fold lower affinity compared to

polyamide ImPyPyPy- $\beta$ -PyPyPyPy- $\beta$ -Dp and at least 7-fold lower than ImImPyPy- $\beta$ -PyPyPyPy- $\beta$ -Dp. Most likely, the Im ring located in the third position from the polyamides amino-terminus does not maintain precise ring-base pair register with the DNA and hence is not positioned optimally for hydrogen bond formation.<sup>13</sup>

ImImImPy- $\beta$ -PyPyPyPy- $\beta$ -Dp binds to 5'-AGGGATCCCT-3' with an apparent equilibrium association constant of  $K_a = 1 \times 10^{10} \text{ M}^{-1}$ . For recognition of core 5'-GGG(A,T)-3' sequences, four-ring ImImImPy subunits appear optimal. C-N covalent head-to-tail linkage of polyamide subunits by a  $\gamma$ -aminobutyric acid ( $\gamma$ ) linker to form "hairpin" polyamides increases affinity by 100-fold relative to the unlinked subunits.<sup>16</sup> The eight-ring hairpin polyamide ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp binds to a 5'-AGGGAA-3' target site with an equilibrium association constant of  $K_a = 4 \times 10^8 \text{ M}^{-1}$ .<sup>21</sup> The eight-ring 4- $\beta$ -4 polyamide ImImImPy- $\beta$ -PyPyPyPy- $\beta$ -Dp binds to a 5'-AGGGAATCCCT-3' target site with 25-fold enhanced affinity. As a control, the affinity of the eight-ring hairpin polyamide ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp was measured on the 5'-AGGGAATCCCT-3' site and found to be  $K_a \sim 3 \times 10^8 \text{ M}^{-1}$ . The 11 base pair site contains two overlapping 6 base pair 5'-AGGGA(A,T)-3' sites. The affinity of the 4- $\gamma$ -4 hairpin was measured for over the entire 11 base pairs, since the two individual 6 base pair sites cannot be considered separately. However, the determined value of  $K_a \sim 3 \times 10^8 \text{ M}^{-1}$  is consistent with the previously published finding of  $K_a = 4 \times 10^8 \text{ M}^{-1}$  for the single site 5'-AGGGAA-3'.<sup>21</sup> Internal  $\beta$  and  $\gamma$  amino acids are specific "guide residues" for binding of polyamides in extended and turn conformations, respectively.<sup>4,22</sup> Supporting this, the deletion of a single methylene unit (MW = 14) from the linker region of ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp to ImImImPy- $\beta$ -PyPyPyPy- $\beta$ -Dp changes the binding motif from 1:1 (hairpin-DNA) to 2:1 (dimer-DNA) and thereby enlarges the targeted binding site size from 6 to 11 base pairs.

The polyamide ImImImPy- $\beta$ -PyPyPyPy- $\beta$ -Dp binds the sequence 5'-AGGGAGTCCCT-3' with 20-fold lower affinity than it does 5'-AGGGAATCCCT-3', demonstrating that a  $\beta/\beta$

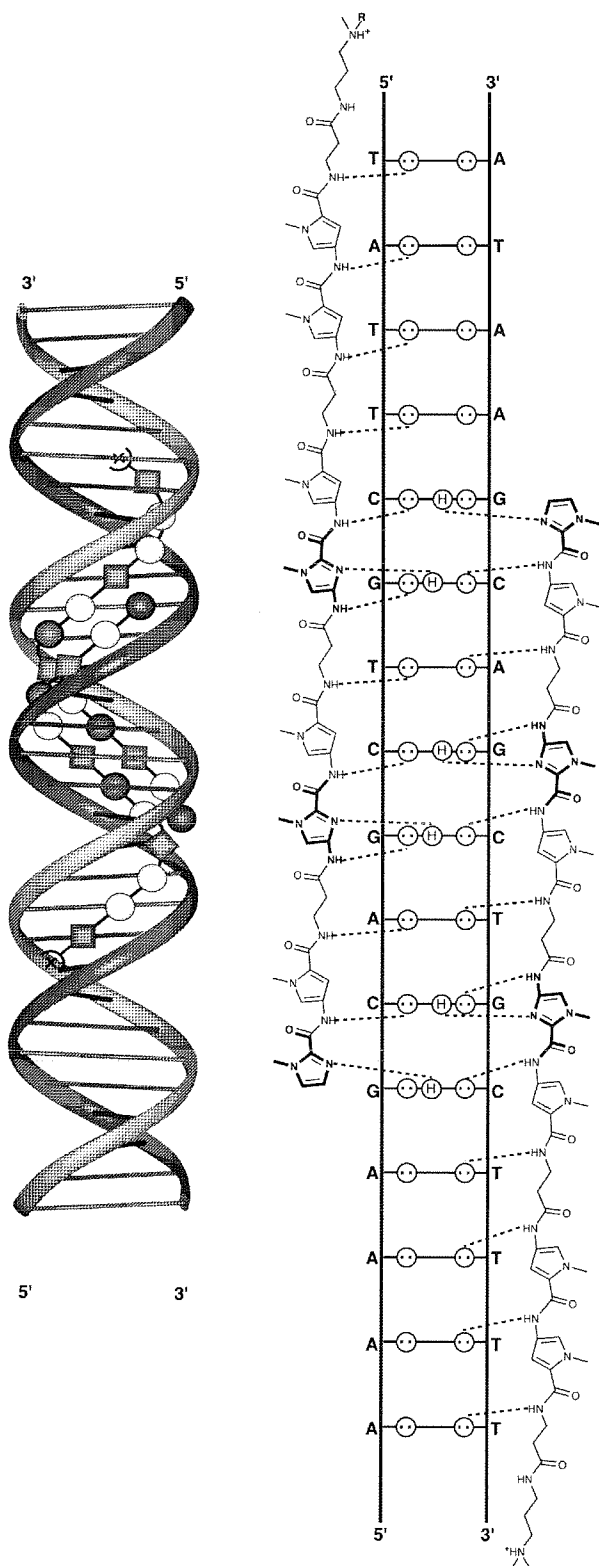
pairing has a preference for binding opposite an A•T or T•A base pair relative to a G•C or C•G base pair (Figure 3.15). The exocyclic amino group of guanine presents a steric bulk in the minor groove of B-form DNA. The specificity of a Py-Py pair for A,T base pairs relative to G,C base pairs likely arises from a steric clash between a Py C3-H and the guanine exocyclic 2-amino group. An antiparallel-oriented  $\beta/\beta$  pair may present an even more sterically demanding surface by placing 4 hydrogens toward the floor of the groove, as compared to the 2 hydrogens presented by a Py-Py pair. Although generality has yet to be examined, these results suggest  *$\beta$ -alanine as a sequence-specific DNA binding element in the overlapped 2:1 motif, rather than a sequence neutral spacer residue.*

Py-Im polyamides bind the minor groove of DNA as dimers with high affinity and specificity. Specific base pair recognition is governed by a simple set of ring-ring pairing rules. Here we expand these pairing rules by demonstrating that side-by-side  $\beta$ -alanine residues are a sequence-specific DNA recognition element with preference for A•T/T•A relative to a G•C/C•G base pair. Fully overlapped 4- $\beta$ -4 dimers bind 11 base pair sequences at subnanomolar concentrations, expanding the DNA-binding site size limit for synthetic polyamides. Unlike the structural complexity of protein-DNA complexes,<sup>14</sup> simple aromatic and aliphatic amino acids (Im, Py, and  $\beta$ ), rationally placed within a (4- $\beta$ -4)<sub>2</sub>•DNA motif, recognize predetermined DNA sequences. The specific recognition of 11 base pairs by synthetic ligands is another pivotal step in the development of synthetic ligands for regulation of gene-expression in living cells.

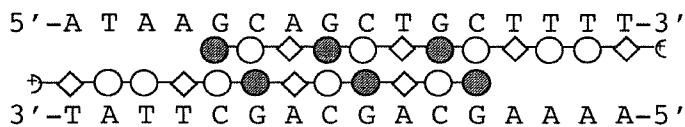
**Recognition of 16 Base Pairs in the Minor Groove of DNA by a Py-Im Polyamide Dimer.**

Recently, an eight-ring hairpin Py-Im polyamide that binds 6 base pairs of DNA was shown to inhibit transcription of a specific gene in cell culture.<sup>23</sup> Polyamides recognizing longer DNA sequences should provide more specific biological activity. To specify a single site within the 3 billion base pair human genome, ligands which specifically recognize 15–16 base pairs are necessary.<sup>24</sup> For this reason, recognition of 16 base pairs represents a milestone in the development of chemical approaches to DNA recognition.<sup>24,25</sup> We examine here the affinity and specificity of a Py-Im polyamide dimer which targets 16 contiguous base pairs in the minor groove of DNA.<sup>26</sup>

As the length of a polyamide dimer having the general sequence ImPy<sub>2-6</sub> increases beyond five rings (corresponding to a 7 base pair binding site), the DNA-binding affinity ceases to increase with polyamide length.<sup>18</sup> A structural basis for this observation is provided by the recently determined x-ray crystal structure of a four-ring homodimer in complex with DNA, which reveals a perfect match of polyamide rise-per-residue with the pitch of the DNA duplex, but over-wound ligand curvature.<sup>13</sup> The curvature mismatch explains the observation that  $\beta$ -alanine residues reset an optimum fit of polyamide dimers with the DNA helix at long binding sites.<sup>5</sup>



**Figure 3.16.** Model of the complex of ImPy- $\beta$ -ImPy- $\beta$ -ImPy- $\beta$ -PyPy- $\beta$ -Dp (Im = *N*-methylimidazole, Py = *N*-methylpyrrole,  $\beta$  =  $\beta$ -alanine, Dp = dimethylaminopropylamide) with 5'-ATAAGCAGCT-GCTTTT-3'. The shaded and open circles represent Im and Py rings, respectively, and the diamonds represents  $\beta$ -alanine. Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines, and circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dashed lines.



**Figure 3.16.** The sequence of the restriction fragment in the region of the 16 base pair target site and a model of the polyamide•DNA complex are shown. The shaded and open circles represent Im and Py rings, respectively, and the diamonds represents  $\beta$ -alanine.

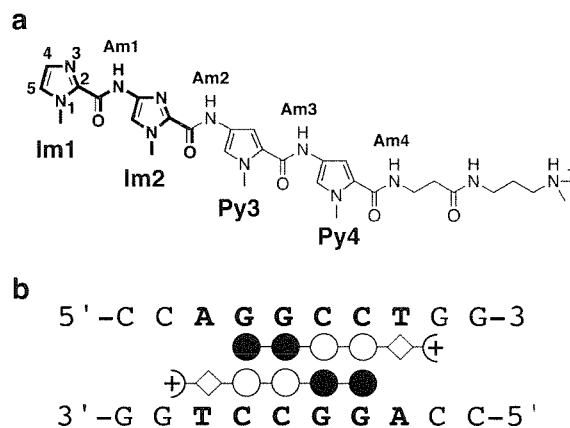
We chose as our binding site the 16 base pair sequence 5'-ATAAGCAGCTGCTTTT-3' present in the regulatory region of the HIV-1 genome.<sup>27</sup> Consideration of the polyamide ring pairing rules,<sup>1</sup> the A,T specificity of  $\beta/\beta$  pairs, and the “slipped” dimer motif<sup>5</sup> suggested that the eight-ring polyamide ImPy- $\beta$ -ImPy- $\beta$ -ImPy- $\beta$ -PyPy- $\beta$ -Dp would specifically bind the target sequence as a cooperative antiparallel dimer (Figure 3.16).

The polyamide was synthesized using purity confirmed by <sup>1</sup>H NMR, analytical HPLC and MALDI-TOF MS. A quantitative DNase I footprinting experiment carried out on a 245 base pair 3'-<sup>32</sup>P-end-labeled restriction fragment revealed that the polyamide specifically binds its target site at subnanomolar concentrations (apparent monomeric association constant,  $K_a = 3.5 \times 10^{10} \text{ M}^{-1}$ ). The binding data were well-fit by a cooperative binding isotherm, consistent with formation of a cooperative 2:1 complex. To provide further evidence that **1** binds as an extended dimer, an affinity cleavage experiment<sup>2a</sup> was carried out with the polyamide-EDTA•Fe(II) conjugate **1-E** (Figure 2b). Cleavage was observed at each end of the match sequence, consistent with a dimeric, antiparallel binding mode. With regard to sequence specificity, there is a proximal two base pair mismatch site, 5'-cAGATGCTGCATATa-3', to the 5' side of the <sup>32</sup>P-labeled strand which is bound with at least 35-fold lower affinity than the match site. However, other mismatch sites on the restriction fragment are bound with 10–20-fold lower affinity, revealing limitations of this first effort at 16 base pair recognition. Undoubtedly there is ample room for further optimization of sequence specificity.

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

### **Structural Basis for G•C Recognition in the DNA Minor Groove.**

The antiparallel dimer in the DNA minor groove formed by a polyamide containing both Im and Py was confirmed by NMR structure studies.<sup>1</sup> Subsequent high resolution x-ray studies of distamycin, (PyPyPy)<sub>2</sub>• DNA, have provided the structure of the Py/Py pair bound opposite A•T/T•A.<sup>2c,2e</sup> The Im/Py and Py/Im pairings, which allowed the design of small molecules which recognize any predetermined DNA sequence at subnanomolar concentration, have not yet been characterized. Because these ligands have been shown to be cell permeable and to inhibit the transcription of specific genes in cell culture, it is important to understand the molecular basis for this non-biological recognition.<sup>23</sup>



**Figure 3.17.** Diagram of polyamide ImImPyPy-β-Dp (Im = *N*-methylimidazole carboxamide, Py = *N*-methylpyrrole carboxamide, β = β-alanine, Dp = dimethylaminopropylamide, Am = amide). Rings are numbered starting from the N-terminal imidazole (Im1, Am1) to the C-terminal pyrrole (Py4, Am4). **b**, Schematic DNA-binding model, with the target site indicated in bold within the sequence of the

We report here the atomic resolution structure of a polyamide of sequence composition ImImPyPy bound to an oligonucleotide duplex containing its cognate 5'-WGGCCW-3' site as predicted by the pairing rules. In addition to providing a molecular basis for predictable read-out of the DNA minor groove, the structure addresses key issues for the design of enhanced second generation polyamides: the register of the polyamide dimer with respect to the DNA double helix critical for expanding binding site size, the position of the paired polyamide rings and the base pairs for the design of novel pairings to break the A•T/T•A degeneracy,<sup>2</sup> and the determinants of high affinity binding in order to increase affinity for sequences which have been difficult to target.<sup>30</sup>

The Py-Im polyamide ImImPyPy-β-Dp was co-crystallized with the oligonucleotide 5'-CCAGGCCTGG-3' (Figure 1). The polyamide selectively binds the six base pair target sequence, 5'-WGGCCW-3' (where W represents either A or T), with a equilibrium dissociation constant,  $K_d = 42$  nM.<sup>30</sup> The structure was solved by MIR with multiple-crystal form and noncrystallographic symmetry-averaging to improve the initial phase information (Table 1). One



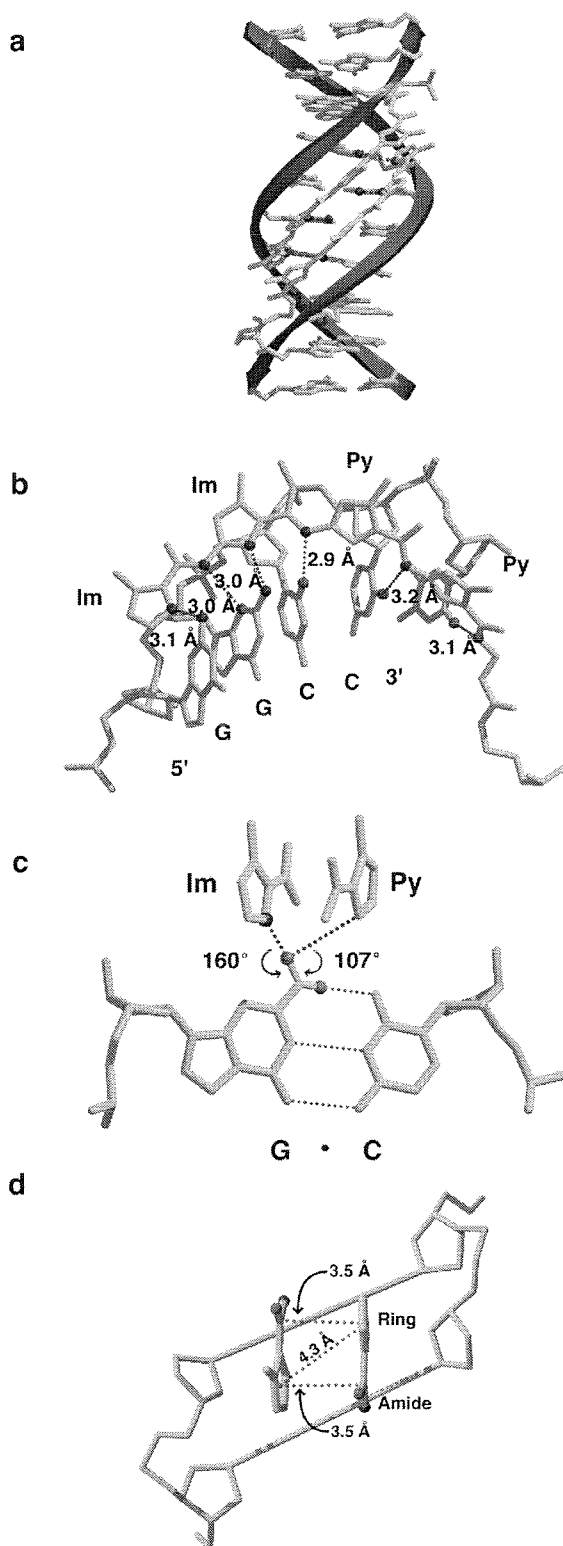
of the brominated derivatives diffracts to 2.0 Å resolution with synchrotron radiation, and was used for the refinement of the structure. There are two independent polyamide molecules and one duplex oligonucleotide present in the asymmetric unit, although the conformations of the polyamides are very similar. Unlike the crystal structure of the distamycin dimer,<sup>2c</sup> the crystal contacts are not mediated by the polyamide molecules.

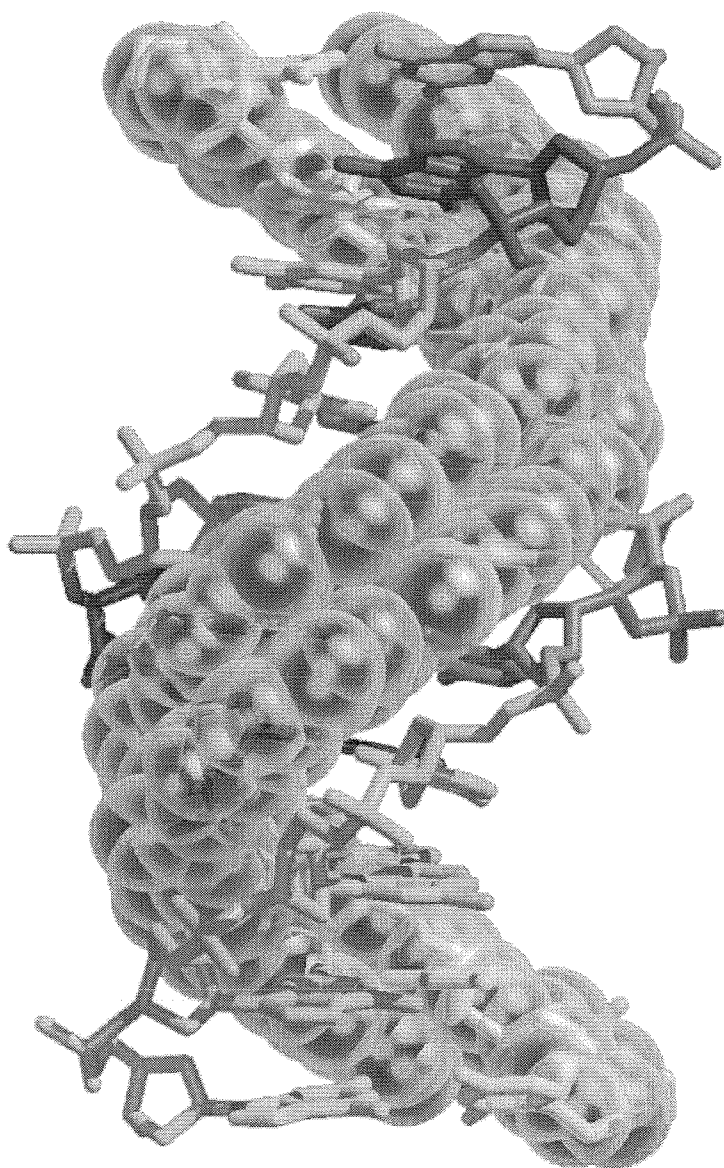
Consistent with NMR studies of other Py-Im polyamides, the polyamide ImImPyPy-β-Dp is bound as an antiparallel head-to-tail dimer, burying 39% of the total available surface area deep in the minor groove of the B-form oligonucleotide (Figure 3a). The N- to C-terminal orientation of each polyamide recognizes the adjacent 5' to 3' nucleic acid strand, a distinct orientation preference that is crucial for a given Py-Im polyamide to predictably target a single site.<sup>32</sup> Minor groove recognition is accomplished by two aromatic residues, each from independent molecules in the dimer, that complement each target base pair of the duplex DNA. The four Im/Py, Im/Py, Py/Im, and Py/Im pairs center over and read the core of the oligonucleotide, G•C, G•C, C•G, C•G, with a series of hydrogen bonds and van der Waals interactions. Flanking the central GGCC sequence, the smooth minor grooves of the A•T/T•A base pairs accommodate the hydrophobic β-alanines.

The ImImPyPy-β-Dp molecules are fully overlapped: the last residue of one polyamide is paired with the first residue of the other in the plane of the bases. The binding mode with the C-terminal pyrroles unpaired, referred to as slipped, is prevented because the resulting Im/Im pair is disfavored for DNA binding.<sup>31</sup> The fully-overlapped conformation of the designed polyamide contrasts with the staggered register of distamycin and ImIm homodimers observed in crystallographic models,<sup>33</sup> although it is consistent with the NMR characterization of the designed polyamide ImPyImPy-β-Dp.<sup>3</sup> In practice, slipped binding modes are disfavored by linking antiparallel molecules with a γ-aminobutyric acid hairpin turn.<sup>11</sup>

**Figure 3.18.** Structure of the (ImImPyPy- $\beta$ -Dp)<sub>2</sub> / DNA complex.

(a), Overall view of the complex. The polyamide dimer completely fills the minor groove. (b), Isolated view of one of the ImImPyPy- $\beta$ -Dp molecules and hydrogen bonds with the adjacent DNA strand. The hydrogen bond lengths given are the average of the ncs-related strands. (c) Geometry of the Im/Py polyamide pair interacting with the G5•C16 base pair. An Im on the cytosine side of the base pair would be unable to form an optimum hydrogen bond with the  $sp^2$ -hybridized hydrogen on the guanine exocyclic amine. This allows the Im/Py pair to discriminate the G•C base pair from C•G. (d), The aromatic ring of one molecule is closest to the carbonyl carbon of the other, although the rings of the polyamide are paired in the planes of the bases.





**Figure 3.19.** An extended polyamide helix. Least squares minimized, mean coordinates for a polyamide monomer were translated and rotated by the parameters of the polyamide helix. After a period of approximately seven residues, the  $14^\circ$  greater twist of the polyamide with respect to B-form DNA manifests as clash with the walls of the minor groove. The number of contacts closer than  $3.4 \text{ \AA}$  between the DNA nucleotides and the polyamides is indicated by a shading gradient. Unfavorable contact would occur towards the end of the helix with the polyamide completely passing through the DNA helix.

The Imidazoles and the amide groups of the polyamides form a series of hydrogen bonds with the edges of the bases on the adjacent DNA strand (Figure 3b). Contributing to the affinity, if not specificity, of the polyamide for DNA, the amides linking the aromatic rings contribute hydrogen bonds to the purine N3 or pyrimidine O2. As opposed to the three-center hydrogen bonds donated by the amides of the netropsin monomers to successive A•T pairs of the binding site,<sup>34</sup> each amide is within hydrogen bond distance of a single base ( $3.0 \pm 0.2$  Å), and beyond contact with the preceding or following N3 or O2 ( $5.1 \pm 0.2$  Å). Although the  $\beta$ -alanines of one molecule in the dimer and the dimethylaminopropylamide tails are not well-ordered, the  $\beta$ -alanine of the other polyamide molecule crosses over to contribute a hydrogen bond from the C-terminal amide to the adenine N3 of the opposite DNA strand.

The Im/Py pairing targets the G•C base pair by packing the Im in the minor groove with favorable geometry to form a hydrogen bond with the free hydrogen of the guanine exocyclic amine (N2). Although the guanine N2 is approximately in the center of the minor groove<sup>35</sup> ( $0.3 \pm 0.2$  Å closer to the Gua C1' atom, on average, when measured in the central 5'-GGCC-3' sequence of the liganded and free<sup>36</sup> structures of the oligonucleotide duplex), an Im/Py pair has a 100 fold greater affinity for a G•C base pair than does a Py/Im pair. Individual hydrogen atoms cannot be discriminated at 2.0 Å resolution; however, hydrogens with geometry consistent with small molecule structures can be assigned to the molecules using the program X-PLOR.<sup>37</sup> The Gua N2-H – Im N3 angle ( $\sim 160^\circ$ , Figure 3c) falls within most probable range of hydrogen bond angles, whereas the potential hydrogen bond angle to an Im over the cytosine of the base pair, extrapolated from the angle to the opposite Py ( $\sim 107^\circ$ ), is less than the most acute angle ( $132^\circ$ ) observed in a survey of two-center hydrogen bonds.<sup>38</sup> Furthermore, with an Im, instead of the Py, on the cytosine side of the G•C base pair, the ring would be expected to move deeper into the minor groove to form a hydrogen bond, further compressing the angle. The potential hydrogen

bond angle ( $\sim 93^\circ$ ) to the other hydrogen of the guanine amine is even less favorable, and in addition the hydrogen is already engaged in the Watson-Crick base pair. The energetic preference for a linear hydrogen bond, coupled with the unfavorable angle to an Im over the cytosine side of the base pair, provides a structural basis for the ability of an Im/Py pair to specifically discriminate G•C from C•G.

Although the rings of the polyamide dimer are adjacent in the planes of the base pairs, when viewed perpendicular to the long axis of the polyamide, the residues are offset such that the rings of one molecule stack over the carboxamides of the other. Because the molecules track the minor groove, the normals of the Im and Py planes are inclined  $50^\circ$  with respect to the base pairs. By simple geometric constraints, the centers of the Im or Py of one molecule are expected to be closest to the bond between the carbonyl carbon and the aromatic ring of the other. Within error, this prediction is observed, although the carbonyl carbons are actually closer than the bonds to the centers of aromatic rings. In the NMR and crystallographic models of polyamide dimers, the aromatic rings of one molecule also stack over the carbonyl carbons of the other, and despite the differences between the overall overlap of the molecules, the C-terminal three residues of the ImImPyPy- $\beta$ -Dp dimer superimpose with 0.3 Å difference on those of the distamycin dimer. The stacking interaction between Py-Im polyamides is a consequence of, and may contribute to, side-by-side pairing and cooperative binding to B-form DNA.

To evaluate the structural changes associated with polyamide binding, the DNA conformation can be compared with the structure of the same sequence in the absence of ligand,<sup>36</sup> and to an oligonucleotide with a central 5'-GGCC-3' sequence and an isomorphous crystal lattice.<sup>39</sup> The twist, pitch, bend and sugar pucker of the DNA helix is B-form and equivalent between the structures. By contrast, the central two G•C base pairs are pried apart by  $\sim 20^\circ$  in the minor groove by the ImImPyPy- $\beta$ -Dp dimer. The Gua-N2H to Cyt-O2 Watson-Crick hydrogen bonds are elongated by  $0.3 \pm 0.1$  Å on average by the reorientation of these base pairs to

optimize the intermolecular hydrogen bonds with the imidazoles as well as to the opposing nucleotide strand. Perhaps due to the slightly different conformation of the leading Im ring, described below, the G•C base pairs at the borders of the central GGCC binding site do not open, and the hydrogen bonds with the first Im ring are slightly longer than those to the second.

The minor groove of G•C rich sequences ( $\sim 6.4$  Å) is inherently wider than that of ideal fiber B-form DNA ( $\sim 5.4$  Å).<sup>36,39</sup> With respect to the unliganded oligonucleotides, the ImImPyPy- $\beta$ -Dp dimer expands the minor groove by only  $\sim 1$  Å. Instead, the polyamides are enclosed tightly within the nucleic acid backbone, engaged by interactions between the deoxyribose O4's and the first Im or the third Py residues. The electron density between the sugar oxygens and these aromatic rings is continuous, and the distances between the nucleic acid oxygens and the rings is in some cases less than the sum of the atomic van der Waals radii. The distance from the sugar O4' is shortest to the center rather than the atoms of the leading Im, although the methylated nitrogen and adjacent carbon of one of the polyamides are also close. The sugar O4' is closest to the third Py C3 atom than to any of the other atoms or the center of the ring. However, the predicted C-H to O4' angle is too acute ( $70.8^\circ$ ) to be a C-H hydrogen bond,<sup>38</sup> and is probably an interaction similar to that between the sugar O4' and the leading Im. Close interactions between the DNA O4' and the drug aromatic rings were also observed in the structures of the distamycin dimer and ImIm bound to DNA.<sup>33</sup> Intimate deoxyribose O4' contacts with aromatic systems are well-documented and have been suggested to play important roles in the stability of many diverse nucleic acid structures.<sup>40</sup>

Calculations based on PyPyPy•DNA crystal structures led to the notion that the rise-per-Py was greater than the pitch of the DNA helix.<sup>41</sup> The four consecutive Py-Im polyamide residues allow calculation of the best-fit helical parameters of the polyamide dimer for comparison with the DNA helix. Strikingly, the rise-per-aromatic amino acid residue is very similar to the rise-per-base pair. However, the twist of the polyamide helix is  $14^\circ$  overwound with respect to the

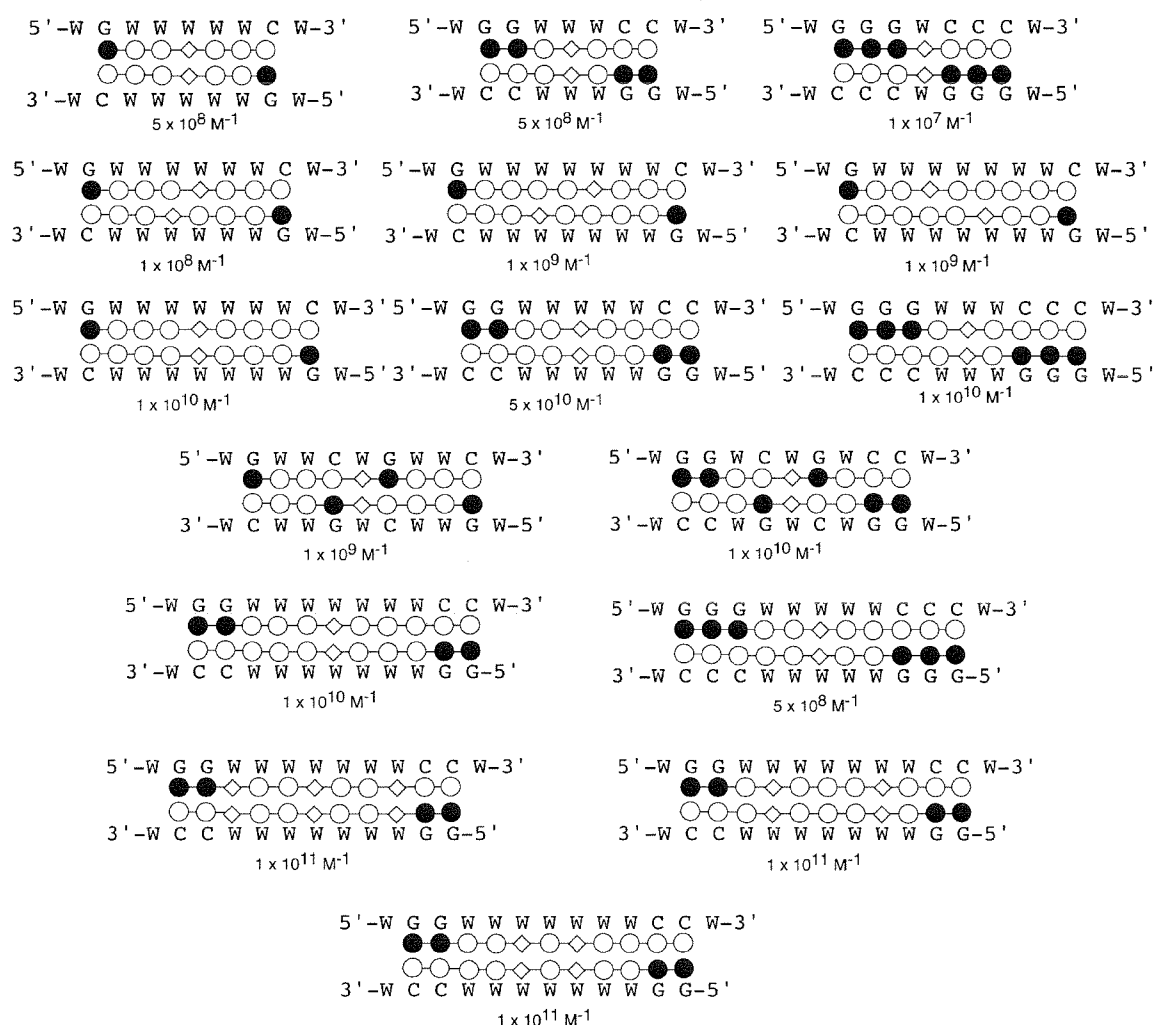
twist of the binding site. When the polyamide is extended by application of the helical transformations to the best fit coordinates of the monomer, a steric clash with the B-form oligonucleotide results. The fifth residue is apparently acceptable, but the sixth has bad contacts, and by the seventh residue, the polyamide is penetrating the base pairs. Quantitative DNase I footprint titrations demonstrate that Py-Im polyamide binding affinity increases with length up to five residues, plateaus between five and seven residues, then plummets as the size is increased further.<sup>18</sup> The insertion of a  $\beta$ -alanine linker within a series of Im or Py-carboxamide subunits restores the increase in binding affinity with length.<sup>5</sup> The helical parameters of the polyamide provides a structural basis for these observations: rather than correcting a translational discrepancy, the flexible linker resets the rotation of the relatively rigid Py or Im carboxamides to fit the curvature of B-DNA.

When the rotation of the residues of ImImPyPy- $\beta$ -Dp are compared with the corresponding monomers of the idealized polyamide helix, the obvious outliers are the leading Im's. The first Im residue is rotated by 20° with respect to the ideal coordinates in one polyamide, and 14° in the other. The singularity of the N-terminal Im's can be explained in terms of their unique interactions with the nucleic acid backbone. These include the sugar O4'-aromatic contact discussed above, and in addition, an interaction between the edge of the first aromatic ring and the preceding deoxyribose O4'. The latter interaction could be described as an interaction with the quadrupole moment of the aromatic ring,<sup>42</sup> or as a C-H hydrogen bond.<sup>38</sup> The C-H to O4' angle is nearly linear, and the carbon to oxygen distance is less than the sum of the atomic van der Waals radii. For the idealized monomer, the distances between the centers and the edges of the leading Imidazoles to the sugar O4's are much longer. The Im with the largest deviation from the idealized coordinates is also the one with the closest interactions. Since the leading Im twists in order to achieve close contacts with the DNA backbone, it is probable that these contacts are energetically favorable.

The hydrogen bond array necessary for specific recognition of the DNA helix by Py-Im polyamides has been predicted through a combination of thermodynamic and NMR studies.<sup>1-5,16,21,22</sup> The crystal structure reported here provides the first direct confirmation that all potential hydrogen bonds are formed between the Im N3 and the exocyclic amine of guanine as well as hydrogen bonds between the Im- and Pycarboxamide NH and the respective N3 of Gua and the O2 of Cyt. The perfect match of the rise-per-aromatic amino acid residue with the pitch of the DNA duplex puts to rest a long standing question regarding curvature and register of the ligand•DNA complex.<sup>41</sup> The overwound curvature of more than four consecutive rings explains the observation that flexible  $\beta$ -alanine residues reset an optimum fit of the polyamide dimer with the DNA helix at long binding sites.<sup>5</sup> Additionally, these results suggest that selectively placed 'β-spring' amino acids may allow for improved positioning of Im residues in order to recognize certain problematic G,C-rich sequences.<sup>30</sup> A wealth of biochemical and structural information has identified general principles of protein-DNA recognition, with a versatile recognition code emerging from studies of certain zinc finger-DNA complexes.<sup>43</sup> A similar database of high resolution small molecule•DNA structures is notably absent. The knowledge of the position of each atom within a DNA-bound synthetic G•C-specific ligand provides a new baseline model of molecular design for minor groove recognition.



**The  $\beta$ -Alanine Extended Polyamide Motif Recognizes a Broad Sequence Repertoire.** The extended, fully overlapped polyamide-DNA motif, provides a versatile template for recognition of symmetric sequences containing from 9 to 16 base pairs in the minor groove (Figure 3.20). Equilibrium association constants for cooperative complex formation range from  $K_a = 1 \times 10^7 \text{ M}^{-1}$  to  $K_a > 1 \times 10^{11} \text{ M}^{-1}$ . Specificities are found to range from 2-fold to >20-fold for discrimination of single base pair mismatch sites.



**Figure 3.20.** A schematic representation of the  $\beta$ -linked fully overlapped polyamide complexes synthesized and studied to date.

## Methods

All crystal growing, structure determination, and data refinement were performed by C. Kielkopf as described elsewhere.<sup>13</sup> All gel electrophoresis and footprinting assays were performed by John Trauger,<sup>5,44</sup> Sue Swalley,<sup>28</sup> and Jim Kelly<sup>18</sup> as described elsewhere.

**Resin Substitutions.** Resin substitution was corrected according to:  $L_{\text{new}}(\text{mmol/g}) = L_{\text{old}} / (1 + L_{\text{old}}(W_{\text{new}} - W_{\text{old}}) \times 10^{-3})$ , where  $L$  is the loading (mmol of amine per gram of resin), and  $W$  is the weight ( $\text{gmol}^{-1}$ ) of the polyamide.<sup>45</sup>

**Aminohexa-(*N*-methylpyrrolicarboxamide) ditrifluoroacetate (7).** To a solution of *N*-(*tert*-butoxycarbonyl)-tris(*N*-methylpyrrolicarboxamide) (20 mg, 41  $\mu\text{mol}$ ) in DMF (100  $\mu\text{l}$ ) was added HBTU (26 mg, 69  $\mu\text{mol}$ ) followed by DIEA (50  $\mu\text{l}$ , 288  $\mu\text{mol}$ ). The reaction was allowed to stand for 5 minutes, agitated, and allowed to stand for an additional 5 minutes. Aminotris-(*N*-methylpyrrolicarboxamide) (24 mg, 41  $\mu\text{mol}$ ) was then added followed by DIEA (50  $\mu\text{l}$ , 288  $\mu\text{mol}$ ) and the reaction agitated for 2 hours. The reaction mixture was concentrated *in vacuo* and TFA (10 ml) added. After 2 minutes the TFA was removed *in vacuo*. Purification of the resulting brown oil by reversed phase HPLC afforded the diamine **7** as a white powder. Yield: 26 mg (58%);  $^1\text{H}$  NMR ( $\text{DMSO-}d_6$ )  $\delta$  10.06 (s, 1 H), 9.95 (m, 2 H), 9.91 (s, 1 H), 9.84 (s, 1 H), 9.44 (br s, 1 H), 8.16 (t, 1 H,  $J = 4.0$  Hz), 7.22 (m, 4 H), 7.16 (d, 1 H,  $J = 1.7$  Hz), 7.10 (s, 1 H,  $J = 1.7$  Hz), 7.07 (m, 3 H), 6.98 (s, 1 H,  $J = 1.7$  Hz), 6.93 (s, 1 H,  $J = 1.8$  Hz), 3.88 (m, 6 H), 3.84 (m, 12 H), 3.79 (m, 6 H), 3.21 (m, 2 H), 3.04 (m, 2 H), 2.77 (d, 6 H,  $J = 4.8$  Hz), 1.80 (m, 2 H); FABMS  $m/e$  835.412 ( $M + H$ , 835.416 calc. for  $\text{C}_{41}\text{H}_{51}\text{N}_{14}\text{O}_6$ ).

**ImPyPyPyPyPyPy-Dp 5.** *N*-methyl-Imidazole-2-carboxylic acid (100 mg, 741  $\mu\text{mol}$ ) and HOBt (72 mg, 500  $\mu\text{mol}$ ) were suspended in 500  $\mu\text{l}$  DMF. Upon addition of DCC (100 mg, 500  $\mu\text{mol}$ ) the reaction mixture became a homogeneous solution. The activation was allowed to stand for 12 hours, precipitated dicyclohexylurea removed by filtration and **7** (10 mg, 9.4  $\mu\text{mol}$ ) added

followed by DIEA (100  $\mu$ l, 576  $\mu$ mol), and the reaction allowed to stand for 2 hours. Reversed phase HPLC purification of the reaction mixture afforded **5** as a white powder. Yield: 6.3 mg (62%); HPLC, r.t. 27.4 min; UV  $\lambda_{\text{max}}$  ( $\epsilon$ ), 246 (34,100), 304 (56,600) nm;  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.46 (s, 1 H), 9.55 (s, 1 H), 9.94 (m, 3 H), 9.90 (s, 1 H), 9.20 (br s, 1 H), 8.14 (t, 1 H,  $J = 7.2$  Hz), 7.38 (s, 1 H), 7.28 (d, 1 H,  $J = 1.4$  Hz), 7.26 (d, 1 H,  $J = 1.4$  Hz), 7.23 (m, 4 H), 7.08 (m, 5 H), 7.04 (s, 1 H,  $J = 1.2$  Hz), 6.93 (d, 1 H,  $J = 1.6$  Hz), 3.98 (s, 3 H), 3.84 (m, 15 H), 3.83 (s, 1 H), 3.30 (q, 2 H,  $J = 7.4$  Hz), 3.21 (t, 2 H,  $J = 7.1$  Hz), 2.77 (d, 6H,  $J = 4.1$  Hz), 1.74 (m, 2 H); MALDI-TOF MS 944.21 (M+H 944.04 calc.); FABMS  $m/e$  965.430 (M + Na, 965.426 calc. for  $\text{C}_{46}\text{H}_{54}\text{N}_{16}\text{O}_7\text{Na}$ ).

**ImPyPy-G-PyPyPy-G-Dp.** Polyamide was prepared by manual solid phase methods and obtained as a white powder upon cleavage of 240 mg resin. (initial substitution of 0.2 mmol Boc-Glycine/gram) with dimethylaminopropylamine (11.9 mg, 29% recovery). HPLC, r.t. 26.9 min.; UV  $\lambda_{\text{max}}$  ( $\epsilon$ ), 246 (41,100), 312 (48,400) nm;  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.49 (s, 1 H), 9.98 (s, 1 H), 9.95 (s, 1 H), 9.92 (s, 1 H), 9.89 (s, 1 H), 9.2 (br s, 1 H), 8.30 (m, 2 H), 8.06 (t, 1 H,  $J = 5.8$  Hz), 7.40 (s, 1 H), 7.24 (d, 1 H,  $J = 1.7$  Hz), 7.23 (m, 3 H), 7.17 (m, 2 H), 7.06 (m, 2 H), 6.94 (m, 3 H), 3.99 (s, 3 H), 3.89 (d, 2 H,  $J = 5.8$  Hz), 3.84 (s, 3 H), 3.84 (s, 3 H), 3.83 (s, 3 H), 3.81 (s, 3 H), 3.80 (s, 3 H), 3.72 (d, 2 H,  $J = 4.3$  Hz), 3.13 (q, 2 H,  $J = 5.7$  Hz), 3.01 (q, 2 H,  $J = 5.2$  Hz), 2.76 (d, 6 H,  $J = 4.3$  Hz), 1.77 (quintet, 2 H,  $J = 7.4$  Hz); MALDI-TOF MS 935.7 (M + H 936.0 calc for  $\text{C}_{44}\text{H}_{55}\text{N}_{16}\text{O}_8$ ); FABMS  $m/e$  935.433 (M + H 935.439 calcd. for  $\text{C}_{44}\text{H}_{55}\text{N}_{16}\text{O}_8$ ).

**ImPyPy-G-PyPyPy- $\beta$ -Dp:** Polyamide was prepared by manual solid phase methods as a white powder upon cleavage of 180 mg resin (initial substitution of 0.2 mmol Boc- $\beta$ -alanine/gram) with dimethylaminopropylamine (2 ml, 55°C). (12.3 mg, 38% recovery after HPLC purification). HPLC, r.t. 25.5, UV  $\lambda_{\text{max}}$  ( $\epsilon$ ), 246 (39,500), 312 (52,000) nm;  $^1\text{H}$  NMR (DMSO- $d_6$ ); 10.46 (s, 1 H), 9.96 (s, 1 H), 9.90 (s, 1 H), 9.88 (m, 2 H), 9.21 (br s, 1 H), 8.27 (t, 1 H,  $J = 4.2$  Hz), 8.06 (m, 2 H), 7.39 (s, 1 H), 7.28 (d, 1 H,  $J = 1.6$  Hz), 7.23 (d, 1 H,  $J = 1.7$  Hz), 7.20 (d, 1 H,  $J = 1.5$  Hz),

7.15 (m, 3 H), 7.04 (m, 2 H), 7.03 (d, 1 H,  $J = 1.6$  Hz), 6.94 (d, 1 H,  $J = 1.7$  Hz), 6.92 (d, 1 H,  $J = 1.4$  Hz), 3.98 (s, 3 H), 3.88 (d, 2 H,  $J = 5.6$  Hz), 3.83 (s, 3 H), 3.82 (m, 6 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.36 (q, 2 H,  $J = 5.3$  Hz), 3.09 (q, 2 H,  $J = 6.0$  Hz), 2.75 (q, 2 H,  $J = 5.2$  Hz), 2.72 (d, 6 H,  $J = 4.8$  Hz), 2.30 (t, 2 H,  $J = 6.3$  Hz), 1.72 (quintet, 2 H,  $J = 5.7$  Hz) MALDI-TOF MS 950.1 (950.0 calc for M + H); FABMS  $m/e$  949.462 (M + H 949.455 calc. for  $C_{45}H_{57}N_{16}O_8$ ).

**ImPyPy- $\beta$ -PyPyPy-G-Dp:** Polyamide was prepared by automated solid phase methods as a white powder upon cleavage of 180 mg resin (initial substitution of 0.2 mmol Boc-Glycine/gram) with dimethylaminopropylamine (2 ml, 55°C) (17.2 mg, 57% recovery after HPLC purification). HPLC, r.t. 26.5; UV  $\lambda_{max}$  ( $\epsilon$ ), 246 (46,500), 312 (54,800) nm;  $^1H$  NMR (DMSO- $d_6$ )  $\delta$  10.54 (s, 1 H), 9.92 (s, 1 H), 9.90 (m, 3 H), 9.23 (br s, 1 H), 8.27 (t, 1 H,  $J = 5.5$  Hz), 8.06 (t, 1 H,  $J = 6.3$  Hz), 8.03 (t, 1 H,  $J = 6.2$  Hz), 7.39 (s, 1 H), 7.26 (d, 1 H,  $J = 1.7$  Hz), 7.20 (m, 2 H), 7.17 (m, 3 H), 7.13 (m, 2 H), 7.04 (d, 1 H,  $J = 1.5$  Hz), 6.87 (d, 1 H,  $J = 1.8$  Hz), 6.83 (d, 1 H,  $J = 1.8$  Hz), 3.97 (s, 3 H), 3.82 (m, 15 H), 3.78 (d, 2 H,  $J = 3.4$  Hz), 3.27 (m, 4 H), 3.15 (m, 2 H), 3.79 (m, 2 H), 2.76 (d, 6 H,  $J = 4.9$  Hz), 1.78 (quintet, 2 H,  $J = 6.6$  Hz) MALDI-TOF MS 950.2 (950.0 calc. for M + H); FABMS  $m/e$  949.458 (M + H 949.455 calc. for  $C_{45}H_{57}N_{16}O_8$ ).

**ImPyPy- $\beta$ -PyPyPy- $\beta$ -Dp:** Polyamide was prepared by automated solid phase methods as a white powder upon cleavage of 240 mg resin (initial substitution of 0.2 mmol Boc- $\beta$ -alanine/gram) with dimethylaminopropylamine (2 ml, 55°C). (19.0 mg, 43% recovery after HPLC purification). HPLC, r.t. 26.8; UV  $\lambda_{max}$  ( $\epsilon$ ), 246 (42,100), 312 (53,900) nm;  $^1H$  NMR (DMSO- $d_6$ )  $\delta$  10.56 (s, 1 H), 9.90 (s, 1 H), 9.89 (m, 2 H), 9.87 (s, 1 H), 9.21 (br s, 1 H), 8.24 (t, 1 H,  $J = 5.2$  Hz), 8.04 (t, 1 H,  $J = 6.1$  Hz), 8.01 (t, 1 H,  $J = 6.0$  Hz), 7.35 (s, 1 H), 7.26 (d, 1 H,  $J = 1.6$  Hz), 7.23 (m, 3 H), 7.16 (m, 3 H), 7.12 (m, 1 H), 7.02 (d, 1 H,  $J = 1.5$  Hz), 6.85 (d, 1 H,  $J = 1.9$  Hz), 6.80 (d, 1 H,  $J = 1.8$  Hz), 3.96 (s, 3 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.36 (q, 2 H,  $J = 5.3$  Hz), 3.09 (q, 2 H,  $J = 6.0$  Hz), 2.75 (q, 2 H,  $J = 5.0$  Hz), 2.72 (d, 6 H,  $J = 4.7$  Hz), 2.30 (t, 2 H,  $J = 6.1$  Hz), 1.72 (quintet, 2 H,  $J = 5.5$  Hz); MALDI-TOF MS 964.2 (964.1 calc. for M + H).

**ImPyPy-Py-PyPyPy-G-Dp:** Polyamide was prepared by manual solid phase methods. Recovery is based on cleavage of 180 mg resin (initial substitution of 0.2 mmol Boc-Glycine/gram) with dimethylaminopropylamine (2 ml, 55°C). (8 mg, 24% recovery after HPLC purification). A small quantity of the failure heptamide AcPyPyPyPyPyPy-Dp was found in the initial preparation and was removed by a second preparatory HPLC purification to afford pure ImPyPy-Py-PyPyPy-G-Dp as a white powder (1.2 mg). HPLC, r.t. 28.5, UV  $\lambda_{\text{max}}$  ( $\epsilon$ ), 246 (34,600), 312 (55,300);  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.55 (s, 1 H), 10.02(s, 1 H), 10.00 (m, 4 H), 9.3 (br s, 1 H), 8.32 (t, 1 H,  $J$  = 6.2 Hz), 8.06 (t, 1 H,  $J$  = 5.9 Hz), 7.44 (s, 1 H), 7.31 (d, 1 H,  $J$  = 1.7 Hz), 7.26 (m, 5H), 7.19 (d, 1 H,  $J$  = 1.8 Hz), 7.10 (m, 5 H), 6.97 (d, 1 H,  $J$  = 1.7 Hz), 4.01 (s, 3 H), 3.87 (m, 15 H), 3.82 (s, 3 H), 3.73 (d, 2 H,  $J$  = 5.5 Hz), 3.16 (q, 2 H,  $J$  = 6.2 Hz), 3.03 (q, 2 H,  $J$  = 5.2 Hz), 2.74 (d, 6 H,  $J$  = 4.9 Hz), 1.77 (quintet, 2 H,  $J$  = 6.7 Hz); MALDI-TOF MS 1000.5; FABMS  $m/e$  1001.471 ( $M + H$  1001.473 calcd. for  $\text{C}_{48}\text{H}_{59}\text{N}_{17}\text{O}_8$ ).

**ImPyPy-Py-PyPyPy- $\beta$ -Dp:** Polyamide was prepared by machine assisted solid phase synthesis to afford a white powder upon cleavage of 800 mg resin (initial substitution of 0.2 mmol Boc- $\beta$ -alanine/gram) with dimethylaminopropylamine (2 ml, 55°C). (56 mg, 36% recovery after HPLC purification) ( $\epsilon$ ) 246 (34,800), 308 (57,000); HPLC r.t. 27.9 min.;  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.47 (s, 1 H), 9.95 (m, 4 H), 9.89 (s, 1 H), 9.2 (br s, 1 H), 8.03 (m, 2 H), 7.39 (s, 1 H), 7.27 (d, 1 H,  $J$  = 1.3 Hz), 7.22 (m, 4 H), 7.15 (m, 2 H), 7.07 (m, 4 H), 7.03 (d, 1 H,  $J$  = 1.4 Hz), 6.86 (d, 1 H,  $J$  = 1.0 Hz), 3.97 (s, 3 H), 3.84 (m, 12 H), 3.82 (s, 3 H), 3.77 (s, 3 H), ( $\beta$ -ala quartet covered by water.), 3.11 (q, 2 H,  $J$  = 5.1 Hz), 3.08 (q, 2 H,  $J$  = 6.0 Hz), 2.72 (d, 6 H,  $J$  = 4.8 Hz), 2.34 (t, 2 H,  $J$  = 4.4 Hz), 1.7 (m, 2 H); MALDI-TOF-MS, 1014.7 (1015.1 calc for  $M+H$ ).

**ImPyPy-G-PyPyPy-G-Dp:** Polyamide was prepared by manual solid phase methods as a white powder upon cleavage of 240 mg resin with *N*-methyl-bis(aminopropyl)amine (2 ml, 55°C) (19.0 mg, 44% recovery after HPLC purification).  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.49 (s, 1 H), 9.97 (s, 1 H), 9.93 (s, 1 H), 9.91 (s, 1 H), 9.89 (s, 1 H), 9.7 (br s, 1 H), 8.27 (m, 2 H), 8.04 (t, 1 H,  $J$  = 5.1 Hz),

7.88 (br s, 3 H), 7.39 (s, 1 H), 7.27 (d, 1 H,  $J = 1.6$  Hz), 7.21 (m, 3 H), 7.15 (m, 2 H), 7.05 (m, 2 H), 6.93 (m, 3 H), 3.97 (s, 3 H), 3.96 (m, 6 H), 3.92 (m, 9 H), 3.72 (m, 4 H), 3.14 (m, 6 H), 3.05 (q, 2 H,  $J = 5.4$  Hz), 2.73 (d, 3 H,  $J = 3.3$  Hz), 1.88 (quintet, 2 H,  $J = 4.6$  Hz), 1.75 (quintet, 2 H,  $J = 6.3$  Hz). MALDI-TOF-MS, 979.0 (979.1 calc for M+H).

**ImPyPy-G-PyPyPy- $\beta$ -Dp:** Polyamide was prepared by manual solid phase methods as a white powder upon cleavage of 240 mg resin with *N*-methyl-bis(aminopropyl)amine (2 ml, 55°C) (25 mg, 55% recovery). HPLC, r.t. 22.0;  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.53 (s, 1 H), 10.00 (s, 1 H), 9.98 (s, 1 H), 9.93 (s, 1 H), 9.92 (s, 1 H), 9.7 (br s, 1 H), 8.31 (t, 1 H,  $J = 5.7$  Hz), 8.12 (t, 1 H,  $J = 5.5$  Hz), 8.04 (t, 1 H,  $J = 5.6$  Hz), 7.9 (br s, 3 H), 7.41 (s, 1 H), 7.29 (d, 1 H,  $J = 1.7$  Hz), 7.23 (d, 1 H,  $J = 1.5$  Hz), 7.22 (d, 1 H,  $J = 1.4$  Hz), 7.16 (m, 3 H), 7.07 (d, 1 H,  $J = 1.2$  Hz), 7.03 (d, 1 H,  $J = 1.3$  Hz), 6.94 (d, 1 H,  $J = 1.6$  Hz), 6.93 (d, 1 H,  $J = 1.5$  Hz), 6.86 (d, 1 H,  $J = 1.4$  Hz), 3.98 (s, 3 H), 3.88 (d, 2 H,  $J = 5.6$  Hz), 3.83 (s, 3 H), 3.82 (m, 6 H), 3.80 (s, 3 H), 3.78 (s, 3 H), 3.37 (q, 2 H,  $J = 6.4$  Hz), 3.11 (m, 6 H), 2.86 (q, 2 H,  $J = 6.1$  Hz), 2.70 (d, 3 H,  $J = 4.6$  Hz), 2.32 (t, 2 H,  $J = 7.2$  Hz), 1.87 (quintet, 2 H,  $J = 7.4$  Hz), 1.75 (quintet, 2 H,  $J = 6.0$  Hz), MALDI-TOF-MS, 993.3 (993.1 calc for M+H).

**ImPyPy- $\beta$ -PyPyPy-G-Dp:** Polyamide was prepared by automated solid phase methods as a white powder upon cleavage of 240 mg resin with *N*-methyl-bis(aminopropyl)amine (2 ml, 55°C) (23.0 mg, 53% recovery). HPLC, r.t. 20.6;  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.45 (s, 1 H), 9.95 (s, 1 H), 9.92 (m, 3 H), 9.6 (br s, 1 H), 8.27 (t, 1 H,  $J = 4.7$  Hz), 8.11 (m, 2 H), 7.9 (s, 3 H), 7.38 (s, 1 H), 7.26 (d, 1 H,  $J = 1.7$  Hz), 7.21 (m, 2 H), 7.17 (m, 2 H), 7.13 (d, 1 H,  $J = 1.8$  Hz), 7.05 (m, 2 H), 6.93 (d, 1 H,  $J = 1.6$  Hz), 6.88 (d, 1 H,  $J = 1.6$  Hz), 6.83 (d, 1 H,  $J = 1.7$  Hz), 3.97 (s, 3 H), 3.82 (s, 9 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.73 (m, 2 H), 3.44 (q, 2 H,  $J = 5.5$  Hz), 3.2 (m, 6 H), 2.85 (q, 2 H,  $J = 5.8$  Hz), 2.73 (d, 3 H,  $J = 4.5$  Hz), 1.89 (quintet, 2 H,  $J = 6.4$  Hz), 1.77 (quintet, 2 H,  $J = 6.9$  Hz) MALDI-TOF-MS, 992.9 (993.1 calc for M+H).

**ImPyPy-G-PyPyPy-G-Dp-EDTA:** EDTA-dianhydride (50 mg) was dissolved in 1 mL DMSO/NMP solution and 1 mL DIEA by heating at 55 °C for 5 min. The dianhydride solution was added to ImPyPy-G-PyPyPy-G-Bp (12.0 mg, 11  $\mu$ mol) dissolved in 750  $\mu$ L DMSO. The mixture was heated at 55°C for 25 min., and treated with 3 mL 0.1M NaOH, and heated at 55 °C for 10 min. 0.1% TFA was added to adjust the total volume to 8 mL and the solution purified directly by preparatory HPLC chromatography to provide ImPyPy-G-PyPyPy-G-Bp-EDTA as a white powder. (4.7 mg, 31% recovery after HPLC purification); HPLC, r.t. 28.8;  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.49 (s, 1 H), 9.97 (s, 1 H), 9.91 (s, 1 H), 9.89 (m, 2 H), 9.4 (br s, 1 H), 8.42 (t, 1 H,  $J$  = 5.0 Hz), 8.31 (t, 1 H,  $J$  = 5.5 Hz), 8.00 (m, 2 H), 7.38 (s, 1 H), 7.26 (d, 1 H,  $J$  = 1.5 Hz), 7.22 (d, 1 H,  $J$  = 1.4 Hz), 7.20 (d, 1 H,  $J$  = 1.4 Hz), 7.14 (m, 3 H), 7.03 (m, 2 H), 6.92 (d, 1 H,  $J$  = 1.5 Hz), 3.95 (s, 3 H), 3.85 (m, 4 H), 3.84 (s, 3 H), 3.80 (m, 6 H), 3.78 (s, 3 H), 3.76 (s, 3 H), 3.69 (m, 6 H), 3.55 (q, 2 H,  $J$  = 5.7 Hz), 3.3-3.0 (m, 12 H), 2.69 (d, 3 H,  $J$  = 3.9 Hz), 2.31 (t, 2 H,  $J$  = 6.8 Hz), 1.73 (m, 4 H); MALDI-TOF-MS, 1254.8 (1254.3 calc for M+H).

**ImPyPy-G-PyPyPy- $\beta$ -Bp-EDTA:** Polyamide was prepared from ImPyPy-G-PyPyPy- $\beta$ -Bp (20 mg) as described for ImPyPy-G-PyPyPy-G-Bp-EDTA. (13.0 mg, 55% recovery after HPLC purification). HPLC, r.t. 27.3;  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.49 (s, 1 H), 9.97 (s, 1 H), 9.91 (s, 1 H), 9.89 (m, 2 H), 9.4 (br s, 1 H), 8.42 (t, 1 H,  $J$  = 5.0 Hz), 8.31 (t, 1 H,  $J$  = 5.5 Hz), 8.00 (m, 2 H), 7.38 (s, 1 H), 7.26 (d, 1 H,  $J$  = 1.5 Hz), 7.22 (d, 1 H,  $J$  = 1.4 Hz), 7.20 (d, 1 H,  $J$  = 1.4 Hz), 7.14 (m, 3 H), 7.03 (m, 2 H), 6.92 (d, 1 H,  $J$  = 1.5 Hz), 3.95 (s, 3 H), 3.85 (m, 4 H), 3.84 (s, 3 H), 3.80 (m, 6 H), 3.78 (s, 3 H), 3.76 (s, 3 H), 3.69 (m, 6 H), 3.55 (q, 2 H,  $J$  = 5.7 Hz), 3.3-3.0 (m, 12 H), 2.69 (d, 3 H,  $J$  = 3.9 Hz), 2.31 (t, 2 H,  $J$  = 6.8 Hz), 1.73 (m, 4 H); MALDI-TOF-MS, 1268.5 (1268.3 calc for M+H).

**ImPyPy- $\beta$ -PyPyPy-G-Bp-EDTA:** Polyamide was prepared from ImPyPy- $\beta$ -PyPyPy-G-Bp (12 mg) as described for ImPyPy-G-PyPyPy-G-Bp-EDTA. (6 mg, 42% recovery after HPLC purification). HPLC, r.t. 28.0;  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.46 (s, 1 H), 9.95 (s, 1 H), 9.93 (m, 3 H),

9.9 (br s, 1 H), 8.43 (t, 1 H,  $J = 5.1$  Hz), 8.28 (t, 1 H,  $J = 5.3$  Hz), 8.03 (m, 2 H), 7.38 (s, 1 H), 7.26 (m, 2 H), 7.21 (d, 1 H,  $J = 1.6$  Hz), 7.17 (d, 1 H,  $J = 1.8$  Hz), 7.12 (d, 1 H,  $J = 1.8$  Hz), 7.10 (s, 1 H), 7.04 (d, 1 H,  $J = 1.6$  Hz), 6.93 (m, 2 H), 6.88 (d, 1 H,  $J = 1.6$  Hz), 6.84 (d, 1 H,  $J = 1.4$  Hz), 3.97 (s, 3 H), 3.87 (m, 4 H), 3.82 (m, 9 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.68 (m, 6 H), 3.3-2.9 (m, 16 H), 2.71 (d, 3 H,  $J = 4.1$  Hz), 1.78 (m, 4 H); MALDI-TOF-MS, 1268.9 (1268.3 calc for M+H).

**ImPyPyPy- $\beta$ -PyPyPyPy- $\beta$ -Dp:** A sample of ImPyPyPy- $\beta$ -PyPyPyPy- $\beta$ -resin prepared by machine-assisted solid phase synthesis (240 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial, and treated with ((dimethylamino)propyl)amine (2 mL) at 55°C for 18 hours. Resin was removed by filtration, and the filtrate diluted to a total volume of 8 mL with 0.1 % (wt/v) aqueous TFA. The resulting crude polyamide/amine solution was purified directly by reversed phase HPLC to provide the trifluoroacetate salt of ImPyPyPy- $\beta$ -PyPyPyPy- $\beta$ -Dp (31 mg, 40% recovery) as a white powder.  $^1\text{H}$  NMR (300 MHz,  $[\text{D}_6]\text{DMSO}$ , 20°C):  $\delta = 10.49$  (s, 1H; NH), 9.97 (s, 1H; NH), 9.95 (s, 1H; NH), 9.94 (s, 1H; NH), 9.93 (m, 2H; NH), 9.91 (s, 1H; NH), 9.4 (br s, 1H;  $\text{CF}_3\text{COOH}$ ), 8.10 (m, 3H; NH), 7.38 (s, 1H; CH), 7.28 (d,  $^2J(\text{H,H}) = 1.6$  Hz, 1H; CH), 7.22 (m, 3H; CH), 7.19 (m, 2H; CH), 7.16 (m, 2H; CH), 7.09 (m, 2H; CH), 7.04 (m, 2H; CH), 6.87 (d,  $^2J(\text{H,H}) = 1.6$  Hz, 1H; CH), 6.86 (d,  $^2J(\text{H,H}) = 1.6$  Hz, 1H; CH), 6.84 (d,  $^2J(\text{H,H}) = 1.5$  Hz, 1H; CH), 3.97 (s, 3H;  $\text{NCH}_3$ ), 3.82 (m, 15H;  $\text{NCH}_3$ ), 3.80 (s, 3H;  $\text{NCH}_3$ ), 3.78 (s, 3H;  $\text{NCH}_3$ ), 3.4 (m, 6H;  $\text{CH}_2$ ), 3.10 (q,  $^4J(\text{H,H}) = 5.4$  Hz, 2H;  $\text{CH}_2$ ), 2.98 (q,  $^4J(\text{H,H}) = 5.3$  Hz, 2H;  $\text{CH}_2$ ), 2.72 (d,  $^2J(\text{H,H}) = 4.7$  Hz, 6H;  $\text{N}(\text{CH}_3)_2$ ), 2.33 (t,  $^3J(\text{H,H}) = 7.0$  Hz, 2H;  $\text{CH}_2$ ), 1.71 (q,  $^5J(\text{H,H}) = 6.4$  Hz, 2H;  $\text{CH}_2$ ); UV/VIS ( $\text{H}_2\text{O}$ )  $\lambda_{\text{max}}$  ( $\epsilon$ ) = 312 (66,600, calculated from  $\epsilon = 8,333/\text{ring}$ ), 244 nm; MALDI-TOF-MS  $[\text{M}^+ - \text{H}]$  1208.2: calc. 1208.3.

**ImImPyPy- $\beta$ -PyPyPyPy- $\beta$ -Dp:** A sample of ImImPyPy- $\beta$ -PyPyPyPy- $\beta$ -resin prepared by machine-assisted solid phase synthesis (240 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial, and treated with ((dimethylamino)propyl)amine (2 mL) at 55°C for 18



hours. Resin was removed by filtration, and the filtrate diluted to a total volume of 8 mL with 0.1 % (wt/v) aqueous TFA. The resulting crude polyamide/amine solution was purified directly by reversed phase HPLC to provide the trifluoroacetate salt of ImImPyPy- $\beta$ -PyPyPyPy- $\beta$ -Dp (31 mg, 40% recovery) as a white powder.  $^1\text{H}$  NMR (300 MHz,  $[\text{D}_6]\text{DMSO}$ , 20 °C):  $\delta$  = 10.38 (s, 1H; NH), 9.95 (s, 1H; NH), 9.93 (s, 1H; NH), 9.91 (s, 1H; NH), 9.90 (m, 2H; NH), 9.76 (s, 1H; NH), 9.4 (br s, 1H;  $\text{CF}_3\text{COOH}$ ), 8.09 (m, 3H; NH), 7.56 (s, 1H; CH), 7.46 (s, 1H; CH), 7.27 (d,  $^2J(\text{H,H})$  = 1.8 Hz, 1H; CH), 7.21 (d,  $^2J(\text{H,H})$  = 1.7 Hz, 1H; CH), 7.20 (d,  $^2J(\text{H,H})$  = 1.9 Hz, 1H; CH), 7.19 (d,  $^2J(\text{H,H})$  = 1.9 Hz, 1H; CH), 7.16 (d,  $^2J(\text{H,H})$  = 1.9 Hz, 1H; CH), 7.15 (d,  $^2J(\text{H,H})$  = 1.6 Hz, 1H; CH), 7.14 (d,  $^2J(\text{H,H})$  = 1.9 Hz, 1H; CH), 7.12 (d,  $^2J(\text{H,H})$  = 1.6 Hz, 1H; CH), 7.07 (s, 1H; CH), 7.05 (d,  $^2J(\text{H,H})$  = 1.5 Hz, 1H; CH), 6.87 (d,  $^2J(\text{H,H})$  = 1.9 Hz, 1H; CH), 6.86 (d,  $^2J(\text{H,H})$  = 1.6 Hz, 1H; CH), 6.84 (d,  $^2J(\text{H,H})$  = 1.6 Hz, 1H; CH), 3.99 (m, 6H;  $\text{NCH}_3$ ), 3.82 (m, 12H;  $\text{NCH}_3$ ), 3.80 (s, 3H;  $\text{NCH}_3$ ), 3.78 (s, 3H;  $\text{NCH}_3$ ), 3.4 (m, 6H;  $\text{CH}_2$ ), 3.09 (q,  $^4J(\text{H,H})$  = 5.6 Hz, 2H;  $\text{CH}_2$ ), 2.97 (q,  $^4J(\text{H,H})$  = 5.2 Hz, 2H;  $\text{CH}_2$ ), 2.71 (d,  $^2J(\text{H,H})$  = 4.2 Hz, 6H;  $\text{N}(\text{CH}_3)_2$ ), 2.32 (t,  $^3J(\text{H,H})$  = 5.1 Hz, 2H;  $\text{CH}_2$ ), 1.71 (q,  $^5J(\text{H,H})$  = 7.4 Hz, 2H;  $\text{CH}_2$ ); UV/VIS ( $\text{H}_2\text{O}$ )  $\lambda_{\text{max}}$  ( $\epsilon$ ) = 306 (66,600, calculated from  $\epsilon$  = 8,333/ring), 243 nm; MALDI-TOF-MS  $[\text{M-H}]$  1209.1: calc. 1209.3.

**ImImImPy- $\beta$ -PyPyPyPy- $\beta$ -Dp:** A sample of ImImImPy- $\beta$ -PyPyPyPy- $\beta$ -resin prepared by machine-assisted solid phase synthesis (240 mg, 0.16 mmol/gram<sup>[17]</sup>) was placed in a 20 mL glass scintillation vial, and treated with ((dimethylamino)propyl)amine (2 mL) at 55°C for 18 hours. Resin was removed by filtration, and the filtrate diluted to a total volume of 8 mL with 0.1 % (wt/v) aqueous TFA. The resulting crude polyamide/amine solution was purified directly by reversed phase HPLC to provide the trifluoroacetate salt of ImImImPy- $\beta$ -PyPyPyPy- $\beta$ -Dp (31 mg, 40% recovery) as a white powder.  $^1\text{H}$  NMR (300 MHz,  $[\text{D}_6]\text{DMSO}$ , 20°C):  $\delta$  = 10.37 (s, 1H; NH), 10.12 (s, 1H; NH), 9.95 (s, 1H; NH), 9.94 (s, 1H; NH), 9.93 (s, 1H; NH), 9.92 (s, 1H; NH), 9.59 (s, 1H; NH), 9.4 (br s, 1H;  $\text{CF}_3\text{COOH}$ ), 8.09 (m, 3H; NH), 7.65 (s, 1H; CH), 7.56 (s, 1H;

CH), 7.45 (s, 1H; CH), 7.27 (d,  $^2J(\text{H,H}) = 1.3$  Hz, 1H; CH), 7.22 (m, 2H; CH), 7.18 (d,  $^2J(\text{H,H}) = 1.2$  Hz, 1H; CH), 7.16 (d,  $^2J(\text{H,H}) = 1.0$  Hz, 1H; CH), 7.07 (m, 2H; CH), 6.95 (d,  $^2J(\text{H,H}) = 1.1$  Hz, 1H; CH), 6.88 (d,  $^2J(\text{H,H}) = 1.4$  Hz, 1H; CH), 6.86 (d,  $^2J(\text{H,H}) = 1.3$  Hz, 1H; CH), 4.01 (s, 3H; NCH<sub>3</sub>), 3.98 (m, 2H; NCH<sub>3</sub>), 3.83 (s, 3H; NCH<sub>3</sub>), 3.82 (m, 6H; NCH<sub>3</sub>), 3.80 (s, 3H; NCH<sub>3</sub>), 3.78 (s, 3H; NCH<sub>3</sub>), 3.4 (m, 6H; CH<sub>2</sub>), 3.11 (q,  $^4J(\text{H,H}) = 5.2$  Hz, 2H; CH<sub>2</sub>), 2.94 (q,  $^4J(\text{H,H}) = 5.3$  Hz, 2H; CH<sub>2</sub>), 2.69 (d,  $^2J(\text{H,H}) = 4.4$  Hz, 6H; N(CH<sub>3</sub>)<sub>2</sub>), 2.33 (t,  $^3J(\text{H,H}) = 5.4$  Hz, 2H; CH<sub>2</sub>), 1.75 (q,  $^5J(\text{H,H}) = 7.1$  Hz, 2H; CH<sub>2</sub>); UV/VIS (H<sub>2</sub>O)  $\lambda_{\text{max}}$  ( $\epsilon$ ) = 304 (66,600, calculated from  $\epsilon = 8,333/\text{ring}$ ), 241 nm; MALDI-TOF-MS [M<sup>+</sup>–H] 1210.4: calc. 1210.3.

## REFERENCES

1. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783; (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci., U.S.A.* **1992**, *89*, 7586; (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385.
2. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci., U.S.A.* **1989**, *86*, 5723; (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393; (c) Chen, X.; Ramakrishnan, Rao, B. S. T.; Sundaralingham, M. *Nature Struct. Biol.* **1994**, *1*, 169; (d) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532. (e) Chen, X.; Ramakrishnan, B. & Sundaralingum, M. *J. Mol. Biol.* **267**, 1157 (1997).
3. (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science*, **1994**, *266*, 646. (b) Mrksich, M.; Dervan, P. B.; *J. Am. Chem. Soc.* **1995**, *117*, 3325.
4. Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Nature Struct. Biol.* **1996**, *3*, 321.
5. Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160.
6. Schultz, P. G.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6834.
7. Youngquist, R. S.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2565.
8. (a) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373; (b) Van Dyke, M. W., Dervan, P. B. *Nucleic Acids Res.* **1983**, *11*, 5555.
9. (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132; (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci., U.S.A.* **1986**, *83*, 8462; (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.
10. Baily, C.; Pommery, N.; Houssin, R.; and Henichart, J.P. *J. Pharm. Sci.* **1989**, *78*, 910.
11. Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983
12. Cantor, C.R.; Schimmel, P. R. *Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules*. W.H. Freeman and Company, New York, NY, **1980**. pp. 863.
13. Keilkopf, C.L.; Baird, E.E.; Dervan, P.B.; Rees, D.C. *Nature Struct. Biol.* **1998**, *5*, 104.
14. Steitz, T.A. *Quart. Rev. Biophys* **1990**, *23*, 205.
15. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 9892. (b) Dwyer, T.J.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 9900, (c) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 3663-3664. (d) Singh, M.P.; Plouvier, B.; Hill, G. C.; Gueck, J.; Pon, R.T.; Lown, J.W. *J. Am. Chem. Soc.* **1994**, *116*, 2006.

16. (a) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147; (b) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6153; (c) Pilch, D. S.; Poklar, N. A.; Gelfaned, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.*, **1996**, *93*, 8306.
17. (a) Breslauer, K.J.; Remeta, D.P.; Chou, W.-Y.; Ferrante, R.; Curry, J.; Zaunczkowski, D.; Snyder, J.G.; Marky, L.A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8922, (b) Marky, L.A.; Breslauer, K.J.; *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4359. (c) Marky, L.A.; Kupke, K.J. *Biochemistry* **1989**, *28*, 9982.
18. Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981
19. Youngquist, R.S.; Dervan, P.B. *J. Am. Chem. Soc.* **1987**, *109*, 7564.
20. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.
21. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198.
22. Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Chem. & Biol.* **1996**, *3*, 369.
23. Gottesfeld, J.M.; Neely, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1997**, 387, 202.
24. (a) Dervan, P.B. *Science* **1986**, *232*, 464. (b) Dervan, P.B. In *The Robert A. Welch Foundation Conference on Chemical Research XXXI. Design of Enzymes and Enzyme Models*; Houston, Texas, November 2–4, 1987; pp. 93. (c) Dervan, P.B. In *Nucleic Acids and Molecular Biology*, Vol. 2; Springer-Verlag: Heidelberg, 1988; pp. 49.
25. (a) Moser, H.E.; Dervan, P.B. *Science* **1987**, *238*, 645. (b) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhouh, N.; Decout, J.L.; Thoung, N.T.; Lhomme, J.; Helene, C. *Nucleic Acids Res.* **1987**, *15*, 7749. (c) Strobel, S.A.; Doucetestamm, L.A.; Riba, L.; Housman, D.E.; Dervan, P.B. *Science* **1991**, *254*, 1639. (d) Thuong, N.T.; Helene, C. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 666.
26. (a) Youngquist, R.S.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2565. (b) Youngquist, R.S.; Dervan, P.B. *J. Am. Chem. Soc.* **1987**, *109*, 7564.
27. (a) Jones, K.A.; Peterlin, B.M. *Ann. Rev. Biochem.* **1994**, *63*, 717. (b) Frech, K.; Brack-Werner, R; Werner, T. *Virology* **1996**, *224*, 256.
28. Swalley, S.E.; Baird, E.E.; Dervan, P.B. *Chem. Eur. J.* **1997**, *3*, 1600.
29. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559.
30. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953.
31. White, S.; Baird, E.E.; Dervan, P.B. *Chem. & Biol.* **1997**, *4*, 569.
32. White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756.

33. Kopka, M. L.; Goodsell, D. S.; Han, G. W.; Chiu, T. K.; Lown, J. W.; Dickerson, R. E. *Structure* **1998**, 5, 1033.
34. Kopka, M. L.; Yoon, C.; Goodsell, D.; Psura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1995**, 82, 1376.
35. Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. USA*, **1976**, 73, 804.
36. Heinemann, U. *J. Mol. Biol.* **1989**, 223, 369.
37. Brunger, A. T. *X-PLOR Version 3.1. A system for x-ray Crystallography and NMR* (University Press, New Haven, 1992).
38. Jeffrey, G. A. *An Introduction to Hydrogen Bonding* (Oxford University Press, New York, 1997).
39. Goodsell, D. S.; Kopka, M. L.; Cascio, D.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1993**, 90, 2930.
40. Egli, M.; Gessner, R. V. *Proc. Natl. Acad. Sci. USA* **1995**, 92, 180.
41. Goodsell, D.; Dickerson, R. E. *J. Med. Chem.* **1986**, 29, 727.
42. Mecozzi, S.; West, A. P.; Dougherty, D. *Proc. Natl. Acad. Sci. USA* **1996**, 93, 10566.
43. Choo, Y. Klug, A. *Curr. Opin. Struct. Biol.* **1997**, 7, 117.
44. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, 120, 3534.
45. Barlos, K.; Chatzi, O.; Gatos, D.; Stravropoulos, G. *Int. J. Peptide Protein Res.* **1991**, 37, 513.

## CHAPTER 4

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

**Abstract:** *In order to optimize the hairpin design for ligands which bind the minor groove of DNA, a series of four Py-Im polyamides substituted at the C-terminus with aliphatic amino acids were prepared using solid phase synthetic methodology. Addition of a C-terminal  $\beta$ -alanine residue was found to enhance both the DNA binding affinity and sequence specificity, while addition of a C-terminal glycine residue was found to reduce DNA binding affinity and sequence specificity. These effects were modulated by the addition of an N-terminal acetyl group. Insertion of a C-terminal aliphatic amino acid residue makes the hairpin-polyamide motif compatible with solid phase synthetic methods, allowing the rapid design of new polyamides for high affinity specific recognition of a broad sequence repertoire in the minor groove of DNA.*

*A series of hairpin polyamides were designed for recognition of 5'-(A,T)GG(A,T)<sub>2</sub>-3' and 5'-(A,T)GGG(A,T)-3' sequences in the minor groove of DNA. Quantitative footprint titration experiments reveal that the optimal polyamide ImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp binds a designated 5'-TGGTT-3' match site with an association constant of  $K_a = 1.0 \times 10^8 \text{ M}^{-1}$  and the single base pair mismatch sites, 5'-TGTTA-3' and 5'-GGGTA-3', with 50-fold and 100-fold lower affinity, respectively. Polyamides of sequence composition AcImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp and AcPyPyPy- $\gamma$ -ImImPy- $\beta$ -Dp, which differ only by the position of the  $\gamma$ -linker, bind with reduced affinity and sequence-specificity. Recognition of sequences containing contiguous G•C base pairs expands the sequence repertoire available for targeting DNA with Py-Im polyamides.*

*We set out to determine the relative energetics of pairings of Im/Py, Py/Im, Im/Im, and Py/Py for targeting G•C and A•T base pairs. A key specificity issue which has not been previously*

addressed is whether an Im/Im pair is energetically equivalent to an Im/Py pair for targeting G•C base pairs. Equilibrium association constants were determined at two five base pair sites for a series of four six-ring hairpin polyamides in order to test the relative energetics of the four aromatic amino acid pairings opposite G•C and A•T base pairs in the central position. We observed that a G•C base pair is effectively targeted with Im/Py but not Py/Im, Py/Py, or Im/Im. The A•T base pair is targeted with Py/Py but not Im/Py, Py/Im, or Im/Im. An Im/Im pairing is energetically disfavored for recognition of both A•T and G•C. This specificity will create limitations on slipped motifs available for unlinked dimers in the minor groove resulting in enhanced predictability of the current pairing rules for specific molecular recognition of double helical DNA.

In order to determine the sequence-specificity of the  $\gamma$ -turn and C-terminal  $\beta$ -tail amino acids, the DNA-binding properties of the hairpin polyamide ImImPy- $\gamma$ -ImPyPy- $\beta$ -Dp were analyzed on a DNA-restriction fragments containing the eight possible 5'-ATGGCNA-3' and 5'-ANGGCTA-3' sites (N = A, T, G or C; 5-bp hairpin site is underlined). Quantitative footprint titrations demonstrate that both the turn and tail amino acids have a > 200-400 fold preference for A•T/T•A relative to G•C base pairs at these primary turn and tail positions. These results indicate that the turn and tail amino acids do not simply act as neutral linker residues, but in fact are sequence-specific recognition elements with predictable DNA-binding specificity.

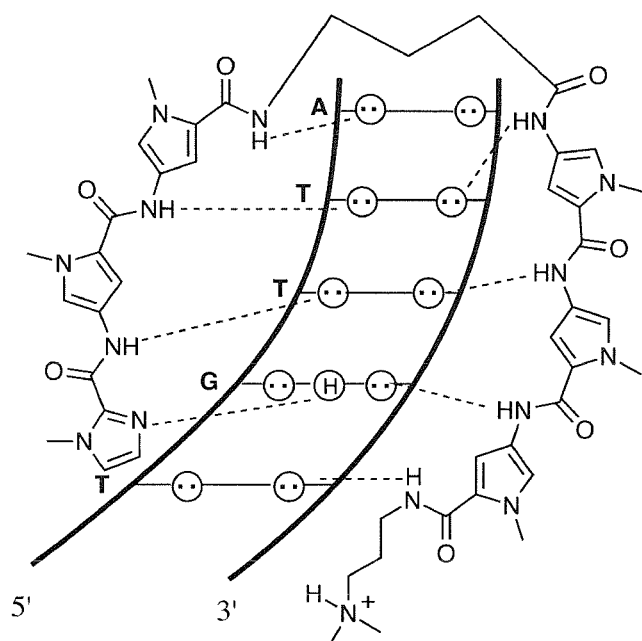
**Publication:** Parks, Baird & Dervan *J. Am. Chem. Soc.* **1996**, *118*, 6147.

Parks, Baird & Dervan *J. Am. Chem. Soc.* **1996**, *118*, 6153.

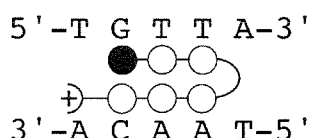
Swalley, Baird & Dervan *J. Am. Chem. Soc.* **1996**, *118*, 8198.

Pilch, Poklar, Gelfand, Law, Breslauer, Baird & Dervan  
*Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306.

White, Baird & Dervan *Chem & Biol* **1997**, *4*, 569.



ImPyPy- $\gamma$ -PyPyPy-Dp • TGTTA

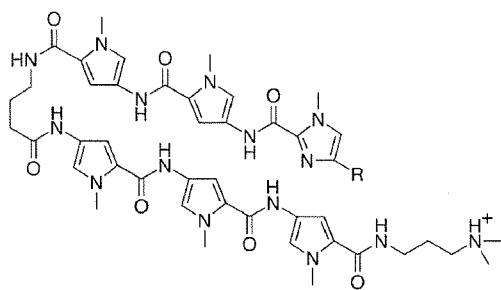


**Figure 4.1.** Hairpin-polyamide binding model for the complex formed between ImPyPy- $\gamma$ -PyPyPy-Dp and a 5'-TGTTA-3' sequence (Top). Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. (Bottom) Schematic binding model, the imidazole and pyrrole rings are represented as shaded and unshaded circles respectively.

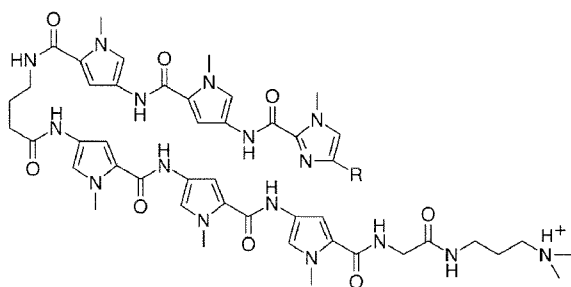
Recently described 2:1 Py-Im polyamide-DNA complexes combined with a convenient solid phase synthesis provide a powerful new paradigm for the design of artificial molecules for the sequence specific recognition of DNA.<sup>1,2</sup> Polyamides containing *N*-methylimidazole and *N*-methylpyrrole amino acids can be combined in antiparallel side-by-side dimeric complexes with the minor groove of DNA.<sup>1,3-5</sup> The DNA sequence specificity of these small molecules can be controlled by the linear sequence of pyrrole and imidazole amino acids. An Im ring on one ligand complemented by a Py ring on the second ligand recognizes a G•C base pair, while a Py/Im combination targets a C•G base pair.<sup>1,4</sup> A Py/Py pair is degenerate for A•T or T•A base pairs.<sup>1,3-4</sup>

A polyamide hairpin motif with  $\gamma$ -aminobutyric acid ( $\gamma$ ) serving as a turn monomer provides a synthetically accessible method of covalently linking polyamide units within the 2:1 motif (Figure 4.1).<sup>6</sup> Covalently linked polyamide heterodimers and homodimers have both increased affinities and sequence specificity.<sup>6,7</sup> The polyamide ImPyPy- $\gamma$ -PyPyPy-Dp **1** was

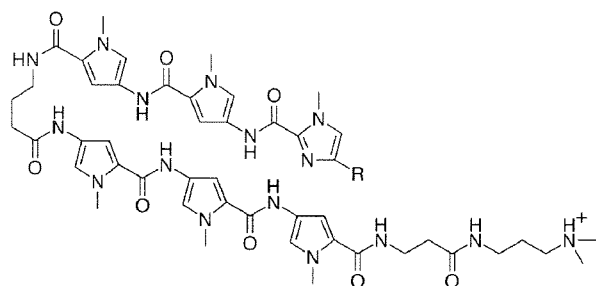




**ImPyPy- $\gamma$ -PyPyPy-Dp; R = H**  
**Ac-ImPyPy- $\gamma$ -PyPyPy-Dp; R = NHCOCH<sub>3</sub>**



**ImPyPy- $\gamma$ -PyPyPy-G-Dp; R = H**  
**Ac-ImPyPy- $\gamma$ -PyPyPy-G-Dp; R = NHCOCH<sub>3</sub>**



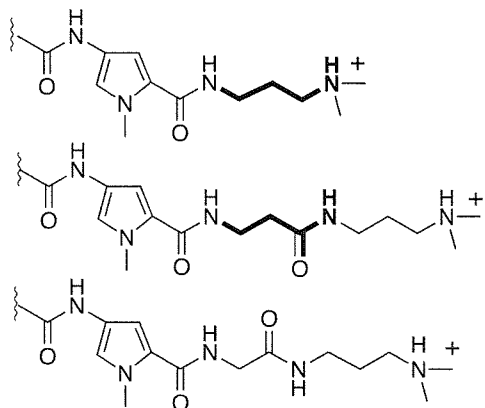
**ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp; R = H**  
**Ac-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp; R = NHCOCH<sub>3</sub>**

**Figure 4.2.** Hairpin pyrrole-imidazole polyamides, with C-terminal -Dp **1** and **2**, G-Dp **3** and **4**, and  $\beta$ -Dp, **5** and **6** end groups.

found to bind a designated 5'-TGTTA-3' target site with high specificity and 300-fold binding enhancement over the individual unlinked polyamide pair; ImPyPy and PyPyPy.<sup>6</sup>

During the process of developing solid phase methodology for the synthesis of the Py-Im polyamides, it was necessary to identify an appropriate resin linkage. Because the reactivity of an imidazole aromatic ester, a pyrrole aromatic ester, and the ester of an aliphatic amino acid are different, separate novel synthetic linkage agents would have to be developed to prepare polyamides with C-terminal imidazole and C-terminal pyrrole residues. The addition of an

aliphatic amino acid at the C-terminus of the Py-Im polyamides provides a convenient alternative to developing new synthetic chemistry for linking aromatic esters to a solid support. *Tert*-butyloxycarbonylaminoacyl-4-(oxymethyl)phenylacetamidomethyl-resin (PAM resin) has been reported to be cleaved in high yield by aminolysis.<sup>8</sup> Preloaded Boc- $\beta$ -alanine-Pam-Resin and Boc-glycine-Pam-Resin are commercially available in a variety of substitution levels.



**Figure 4.3.** Modification of the C-terminus of pyrrole-imidazole polyamides: (top) -Dp, (middle) G-Dp, (bottom)  $\beta$ -Dp.

A series of four Py-Im polyamides containing either a C-terminal glycine or  $\beta$ -alanine residue were prepared by solid phase methodology (Figure 4.2). We find that *N*-terminal acetylation can alter the relative affinities of single base pair mismatches. We report here that the polyamide synthesized from Boc- $\beta$ -alanine-Pam-Resin, ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **5** binds with both enhanced affinity and specificity relative to the parent compound,

ImPyPy- $\gamma$ -PyPyPy-Dp **1**,<sup>6</sup> which lacks the C-terminal  $\beta$ -alanine residue.

MPE•Fe (II) footprinting<sup>9</sup> on the 3' and 5'-<sup>32</sup>P-end-labeled 135 base pair *EcoRI/BsrBI* restriction fragments (25 mM Tris-acetate, 10 mM NaCl, 100  $\mu$ M/ base pair calf thymus DNA at pH 7.0 and 22°C) reveal that the polyamides all bind to the 5'-TGTTA-3' match site. Single base pair mismatch site preferences varied for the individual polyamides.

Quantitative DNase I footprinting<sup>10</sup> on the 3'-<sup>32</sup>P-end-labeled 135 base pair *EcoRI/BsrBI* restriction fragment (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub> at pH 7.0 and 22°C) reveals the following first order association constants for the polyamides studied: ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **5** ( $2.9 \times 10^8 \text{ M}^{-1}$ ) > ImPyPy- $\gamma$ -PyPyPy-Dp **1** ( $7.6 \times 10^7 \text{ M}^{-1}$ ) > AcImPyPy- $\gamma$ -PyPyPy-Dp **2** ( $6.4 \times 10^7 \text{ M}^{-1}$ ) > AcImPyPy- $\gamma$ -PyPyPy-G-Dp **4** ( $4.0 \times 10^7 \text{ M}^{-1}$ ) > AcImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **6** ( $2.9 \times 10^7 \text{ M}^{-1}$ ) > ImPyPy- $\gamma$ -PyPyPy-G-Dp **3** ( $3.3 \times 10^6 \text{ M}^{-1}$ ) (Table 4.1). First order association constants for the single base pair mismatch site 5'-TGACA-3' for ImPyPy- $\gamma$ -PyPyPy-Dp **1** and ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **5** are  $3.2 \times 10^6 \text{ M}^{-1}$  and  $4.8 \times 10^6 \text{ M}^{-1}$ , respectively. AcImPyPy- $\gamma$ -PyPyPy-Dp **2**, ImPyPy- $\gamma$ -PyPyPy-G-Dp **3**, AcImPyPy- $\gamma$ -PyPyPy-G-Dp **4**, and AcImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **6** bind to 5'-ATTCG-3' with affinities of  $2.8 \times 10^7 \text{ M}^{-1}$ ,  $1.0 \times 10^6 \text{ M}^{-1}$ ,  $1.2 \times 10^7 \text{ M}^{-1}$ , and  $1.8 \times 10^7 \text{ M}^{-1}$ , respectively.

Among the six polyamides, ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **5** binds the targeted 5'-TGTTA-3'

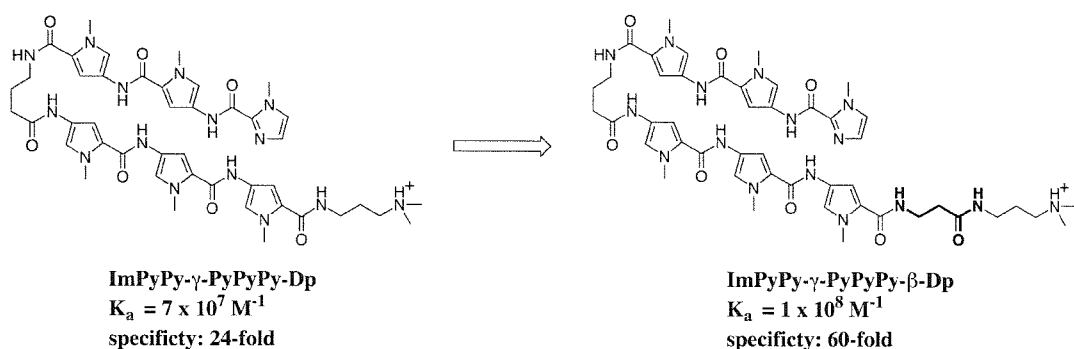
**Table 4.1** Binding Constants ( $M^{-1}$ )<sup>a,b</sup>

Polyamide	5'-TGTTA-3'	mismatch <sup>c</sup>
PyPyPy- $\gamma$ -ImPyPy-G-Dp	$5.1 \times 10^6$ (1.2)	$\leq 1 \times 10^5$
PyPyPy- $\gamma$ -ImPyPy- $\beta$ -Dp	$2.0 \times 10^7$ (0.5)	$1.2 \times 10^7$ (0.4)
AcPyPyPy- $\gamma$ -ImPyPy-G-Dp	$5.1 \times 10^7$ (1.8)	$1.6 \times 10^7$ (0.1)
AcPyPyPy- $\gamma$ -ImPyPy- $\beta$ -Dp	$6.8 \times 10^7$ (2.1)	$8.7 \times 10^6$ (2.9)

<sup>a</sup>Values are the mean values measured from three footprint titration experiments, with the standard deviation for each data set indicated in parentheses. <sup>b</sup>The assays were performed at 22°C at pH 7.0 in the presence of 10 mM tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. <sup>c</sup> PyPyPy- $\gamma$ -ImPyPy-G-Dp and PyPyPy- $\gamma$ -ImPyPy- $\beta$ -Dp bind a 5'-TGACA-3' site; AcPyPyPy- $\gamma$ -ImPyPy-G-Dp and AcPyPyPy- $\gamma$ -ImPyPy- $\beta$ -Dp bind to a 5'-ACATT-3' site.

site with the highest affinity. This suggests that addition of a C-terminal linker  $\beta$ -alanine residue to facilitate solid phase polyamide synthesis is not merely an acceptable strategy, but serendipitously, *designates an optimized hairpin polyamide*. ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **5** binds with a first order association constant,  $K_a = 3 \times 10^8 M^{-1}$ , a factor of four greater than the parent polyamide, ImPyPy- $\gamma$ -PyPyPy-Dp **1**,  $K_a = 8 \times 10^7 M^{-1}$ . A C-terminal glycine residue *reduces* binding affinity at the 5'-TGTTA-3' match site by a factor of 88 for ImPyPy- $\gamma$ -PyPyPy-G-Dp **3** binding relative to ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **5**. The glycine residue may create a steric clash placing the glycine carbonyl in the floor of the minor groove.

A high resolution NMR study of the 2:1 polyamide:DNA complex for ImPyPy-G-PyPyPy-Dp indicates considerable distortion of the polyamides to avoid placing the glycine carbonyl in the floor of the minor groove.<sup>11</sup> In contrast to glycine, a C-terminal  $\beta$ -alanine residue presents a steric surface which resembles that of the C-terminus of the original hairpin polyamide (Figure 4.3). The modest increased binding affinity of the C-terminal  $\beta$ -alanine polyamide may result from an additional hydrogen bond between the  $\beta$ -alanine carboxamide and a 'sixth' base pair of the binding site.



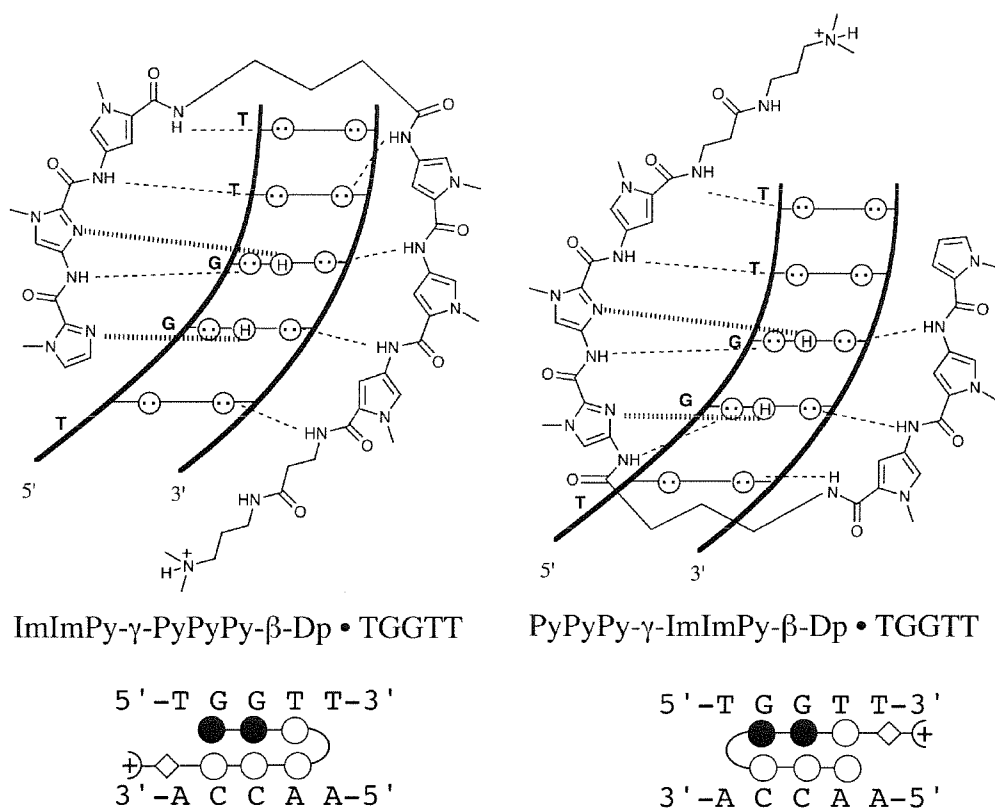
**Figure 4.4.** The chemical structures of the polyamide of the prior art ImPyPy- $\gamma$ -PyPyPy-Dp, and the optimized hairpin polyamide ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp.

All six polyamides of common sequence component, ImPyPy- $\gamma$ -PyPyPy, but different substitutions at the N and C -termini were found to be highly specific for the designated 5'-TGTTA-3' target site, indicating the pyrrole-imidazole polyamides can tolerate a variety of substitutions. The optimal was ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **5** which binds the target 5'-TGTTA-3' match site with 60-fold specificity relative to a single base pair 5'-TGACA-3' mismatch site. This can be compared with the parent polyamide ImPyPy- $\gamma$ -PyPyPy-Dp **1** which has a 24-fold specific binding relative to the same sites.<sup>6</sup>

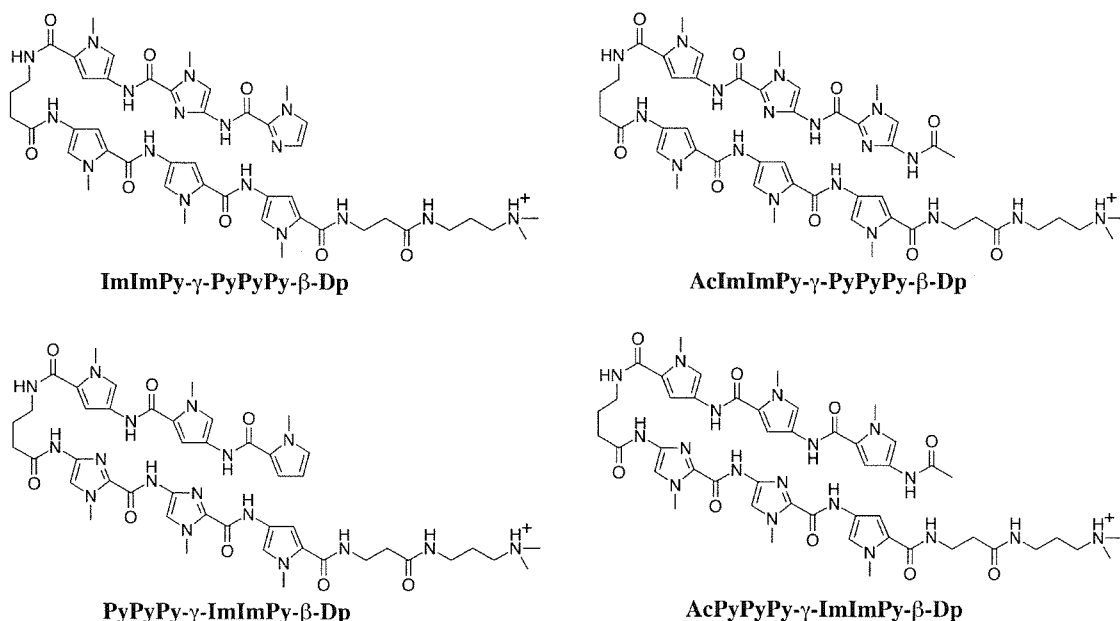
An N-terminal acetyl group or a C-terminal glycine group reduces the observed relative sequence specificity to between 1 and 3 fold. The steric bulk of the acetyl group or glycine residue may simply be preventing the polyamide from sitting deeply in the minor groove, reducing the affinity for the match, while increasing the tolerance at the mismatch. Structural studies will be necessary to describe this interaction and are in progress in collaboration with the Wemmer group at UC Berkeley.

**Optimization of the placement of the  $\gamma$ -turn amino acid for recognition of the minor groove by hairpin polyamides.** Formation of a hairpin-polyamide by covalently linking a polyamide heterodimer with a  $\gamma$ -aminobutyric acid ( $\gamma$ ) residue provides a 300-fold enhancement in affinity over the unlinked polyamides.<sup>6</sup> Moreover, the specificity of the hairpin is greatly improved. The initial placement of the  $\gamma$ -amino acid turn was chosen for synthetic ease and was not varied. With the development of solid phase methodology for polyamide synthesis, we can now rapidly assess the effect of varying the position of the  $\gamma$ -turn monomer.

In order to explore the recognition of 5'-(A,T)GG(A,T)-3' sequences, a series of four head-to-tail linked hairpin polyamides containing *neighboring* imidazole rings, ImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp **7**, PyPyPy- $\gamma$ -ImImPy- $\beta$ -Dp **8**, AcImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp **9**, and AcPyPyPy- $\gamma$ -



**Figure 4.5.** Binding model for the complexes formed between polyamides ImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp **7** (left) and PyPyPy- $\gamma$ -ImImPy- $\beta$ -Dp **8** (right) and a 5'-TGGTT-3' sequence. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. Shaded and nonshaded circles denote Im and Py carboxamides.

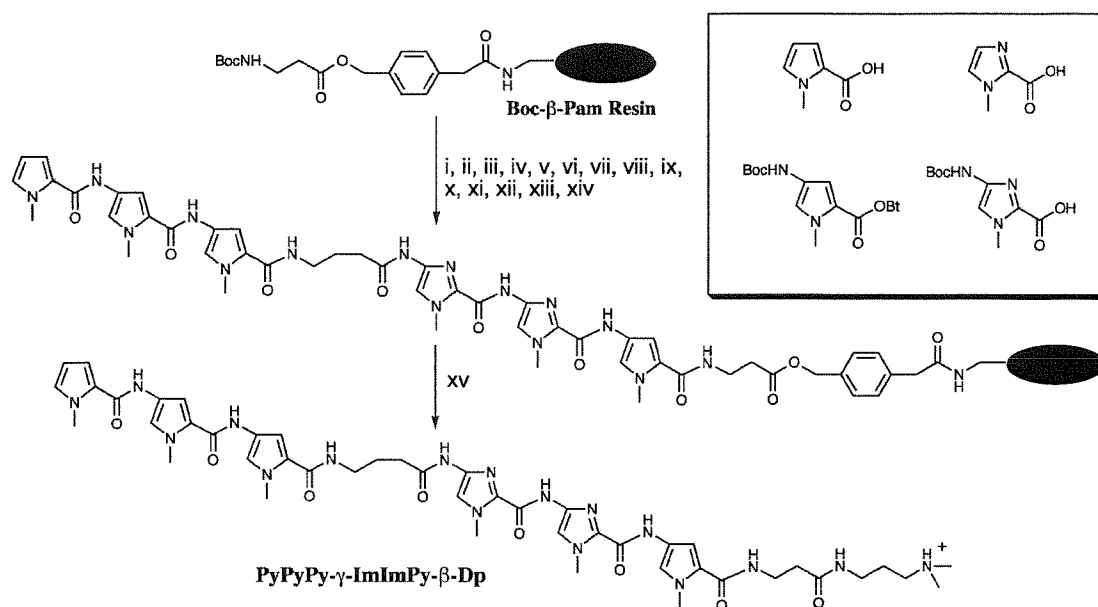


**Figure 4.6.** Series of polyamides synthesized using solid phase methodology.

ImImPy- $\beta$ -Dp **10**, were prepared using solid phase methods (Figures 4.5 and 4.6).<sup>7</sup> The polyamides are all synthesized with Boc- $\beta$ -alanine-Pam resin, previously shown as optimal for polyamides.<sup>12</sup> Each imidazole is expected to form a specific hydrogen bond with a guanine amino group allowing the recognition of contiguous G•C base pairs (Figure 4.5). In addition, the linker turn position is varied within the nonacetylated and acetylated pairs of polyamides to determine the effect on the sequence specificity and binding affinity.

We report here the binding specificity and affinity of the polyamides as determined by the complementary techniques, MPE•Fe(II) footprinting<sup>9</sup> and quantitative DNase I footprinting.<sup>10</sup> MPE•Fe(II) footprinting verifies that sequence specific recognition of the expected 5'-TGGTT-3' target site has been achieved. In addition, DNase I quantitative footprint titration experiments reveal that the position of the  $\gamma$ -linker does not dramatically affect either affinity or specificity of polyamides, especially the pair containing acetylated N-termini.

The polyamides ImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp **7**, PyPyPy- $\gamma$ -ImImPy- $\beta$ -Dp **8**, AcImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp **9**, and AcPyPyPy- $\gamma$ -ImImPy- $\beta$ -Dp **10** were prepared by solid phase methodology (Figure 4.7). Four unique pyrrole and imidazole building blocks were combined in a stepwise



**Figure 4.7.** (Box) Pyrrole and imidazole monomers for synthesis of all compounds described here; pyrrole-2-carboxylic acid, imidazole-2-carboxylic acid, Boc-Pyrrole-OBt ester, and Boc-Imidazole-acid. Solid phase synthetic scheme for PyPyPy-γ-ImImPy-β-Dp starting from commercially available Boc-β-Pam-resin: (i) 80% TFA/DCM, 0.4M PhSH; (ii) BocPy-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4M PhSH; (iv) BocIm-OBt (DCC/HOBt), DIEA, DMF; (v) 80% TFA/DCM, 0.4M PhSH; (vi) BocIm-OBt (DCC/HOBt), DIEA, DMF; (vii) 80% TFA/DCM, 0.4M PhSH; (viii) Boc-γ-aminobutyric acid (HBTU, DIEA); (ix) 80% TFA/DCM, 0.4M PhSH; (x) BocPy-OBt, DIEA, DMF; (xi) 80% TFA/DCM, 0.4M PhSH; (xii) BocPy-OBt, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4M PhSH; (xiv); pyrrole-2-carboxylic acid (HBTU/DIEA); (xv) *N,N*-dimethylaminopropylamine, 55°C.

manner on a solid support using Boc- chemistry protocols. For example, polyamide **8**, PyPyPy-γ-ImImPy-γ-Dp, was prepared in 16 steps on the resin, and then cleaved with a single step aminolysis reaction (Figure 4.7). All polyamides were found to be soluble up to at least 1 mM concentration in aqueous solution.

MPE•Fe(II) footprinting on a 3'- or 5'-<sup>32</sup>P end-labeled 266 base pair *EcoRI/PvuII* restriction fragment from plasmid pMEPGG (25 mM tris-acetate, 100 μM bp calf thymus DNA, 10 mM NaCl) reveals that the synthetic polyamides **7-10**, at 10 μM concentration, bind the designated target site 5'-TGGTT-3'. In addition, several single base pair mismatch sites are bound with lower affinity.

Quantitative DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl,

**Table 4.2.** First Order Binding Constants (M<sup>-1</sup>)<sup>a,b</sup>.

Polyamide	Match Site	Single Mismatch Sites	
	5'-aTGGTTt -3'	5'-tTGTTAt-3'	5'-tGGGTAg-3'
ImImPy-γ-PyPyPy-β-Dp	1.0 x 10 <sup>8</sup> (0.1)	1.7 x 10 <sup>6</sup> (0.6)	≤ 1 x 10 <sup>6</sup>
PyPyPy-γ-ImImPy-β-Dp	1.6 x 10 <sup>7</sup> (0.2)	< 1 x 10 <sup>5</sup>	< 1 x 10 <sup>5</sup>
AcImImPy-γ-PyPyPy-β-Dp	1.3 x 10 <sup>7</sup> (0.7)	1.6 x 10 <sup>6</sup> (1.1)	1.3 x 10 <sup>6</sup> (0.8)
AcPyPyPy-γ-ImImPy-β-Dp	2.0 x 10 <sup>7</sup> (0.4)	1.3 x 10 <sup>6</sup> (0.5)	≤ 1 x 10 <sup>6</sup>

<sup>a</sup>Values are the mean values measured from at least three footprint titration experiments, with the standard deviation for each data set indicated in parentheses. <sup>b</sup>Assays were performed at 22 °C at pH 7.0 in the presence of 10 mM tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

10mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>, pH 7.0 and 22°C) were performed to determine the first order binding affinities of the four polyamides **7-10** for a designated match site, 5'-TGGTT-3', as well as for two single base pair mismatch sites, 5'-TGTTA-3' and 5'-GGGTA-3' (Table 4.2). The polyamide ImImPy-γ-PyPyPy-β-Dp binds the target site 5'-TGGTT-3' with the highest affinity (first order association constant  $K_a = 1.0 \times 10^8 \text{ M}^{-1}$ ). The remaining polyamides have lower but approximately equal first-order association constants of  $K_a = \sim 2 \times 10^7 \text{ M}^{-1}$  for the target site. The non-acetylated polyamides in this series are > 50-fold specific for the 5'-TGGTT-3' match site over either of the single base pair mismatch sites analyzed. The acetylated pair of polyamides exhibit lower sequence specificity for the analyzed sites.

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

Quantitative DNase I footprint titration experiments reveal that ImImPy-γ-PyPyPy-β-Dp

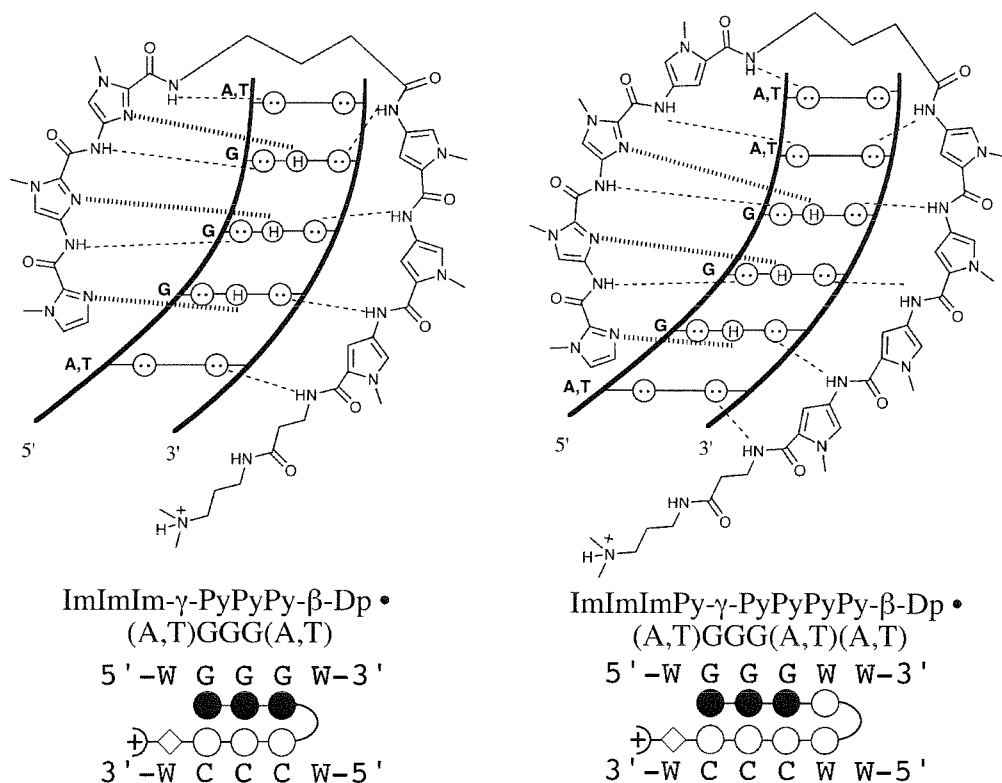


**7** is optimal within this series of four polyamides. This hairpin binds a 5'-TGGTT-3' match site with an apparent first-order association constant of  $K_a = 1 \times 10^8 \text{ M}^{-1}$ , while the corresponding hairpin PyPyPy- $\gamma$ -ImImPy- $\beta$ -Dp **8**, which differs only in the position of the  $\gamma$ -turn, shows lower affinity ( $K_a = \sim 2 \times 10^7 \text{ M}^{-1}$ ) for the 5'-TGGTT-3' site. Both unacetylated polyamides demonstrate good specificity ( $> 10$ -fold) for the target match site over the single base pair mismatch sites. The acetylated polyamides are similar in affinity to PyPyPy- $\gamma$ -ImImPy- $\beta$ -Dp **8**, but exhibit lower specificity. AcImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp **9** and AcPyPyPy- $\gamma$ -ImImPy- $\beta$ -Dp **10** are virtually indistinguishable from each other based on affinity and specificity for the analyzed target sequences, indicating little preference for turn position.

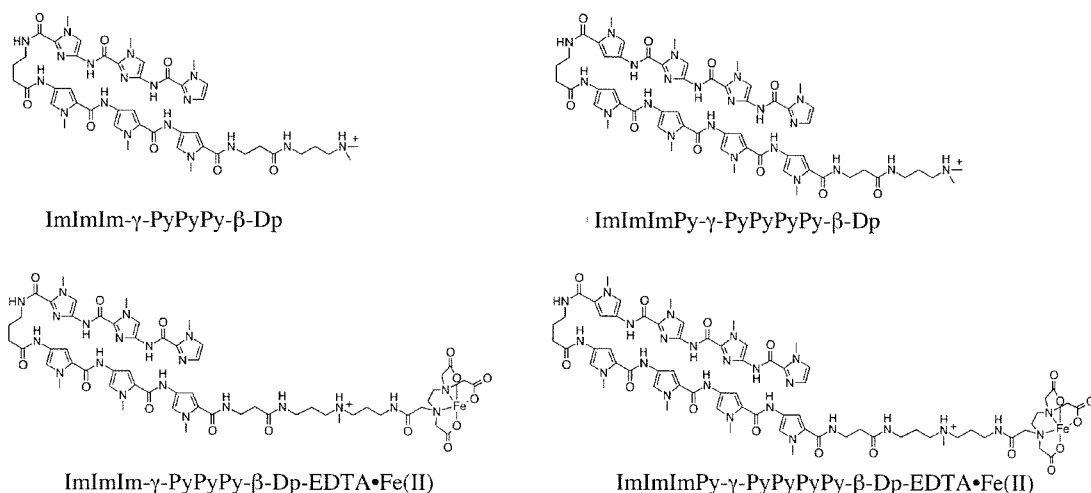
This series of contiguous imidazole-containing polyamides is remarkably similar in affinity and specificity to the single imidazole-containing hairpin polyamide, ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp, indicating little or no energetic penalty in this system for *adjacent* imidazoles.<sup>12</sup> Importantly, the position of the hairpin turn does not significantly affect the recognition of the target 5'-TGGTT-3' match site, although single base pair mismatch relative affinities are altered. The 2:1 motif has previously been used to specifically target several sequences: 5'-TGTC A-3',<sup>1</sup> 5'-TGTT A-3',<sup>6</sup> 5'-AAGTT-3',<sup>4</sup> and 5'-TGCGCA-3'.<sup>5</sup> The results reported herein add sequences containing two contiguous G•C base pairs to the list, expanding the sequence repertoire for DNA recognition by polyamides. Furthermore, turn position showed minimal effects on the specificity and affinity of the polyamides, indicating a new degree of flexibility within the 2:1 motif.

**Recognition of a 5'-(A,T)GGG(A,T)<sub>2</sub>-3' Sequence in the Minor Groove of DNA by an Eight-Ring Hairpin Polyamide.** To further expand the sequence repertoire available with the hairpin motif, two polyamides containing three contiguous Im rings, ImImIm- $\gamma$ -PyPyPy- $\beta$ -Dp (**11**) and ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp (**12**), and the corresponding affinity cleaving analogs, ImImIm- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA (**11-E**) and ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp-EDTA (**12-E**), were synthesized using solid phase synthetic protocols (Figures 4.8 and 4.9).

Specific hydrogen bonds are expected to form between each Im N3 and one of the three individual guanine 2-amino groups on the floor of the minor groove (Figure 4.8).



**Figure 4.8.** Binding model for the complexes between DNA and the six-ring hairpin ImImIm- $\gamma$ -PyPyPy- $\beta$ -Dp (left) or the eight-ring hairpin ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp (right). Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. Ball and stick models are also shown. Shaded and nonshaded circles denote Im and Py, respectively. Diamonds represent the  $\beta$ -alanine residue. W represents either an A or T base.



**Figure 4.9.** Structures of the polyamides and their EDTA derivatives synthesized using solid phase methodology.

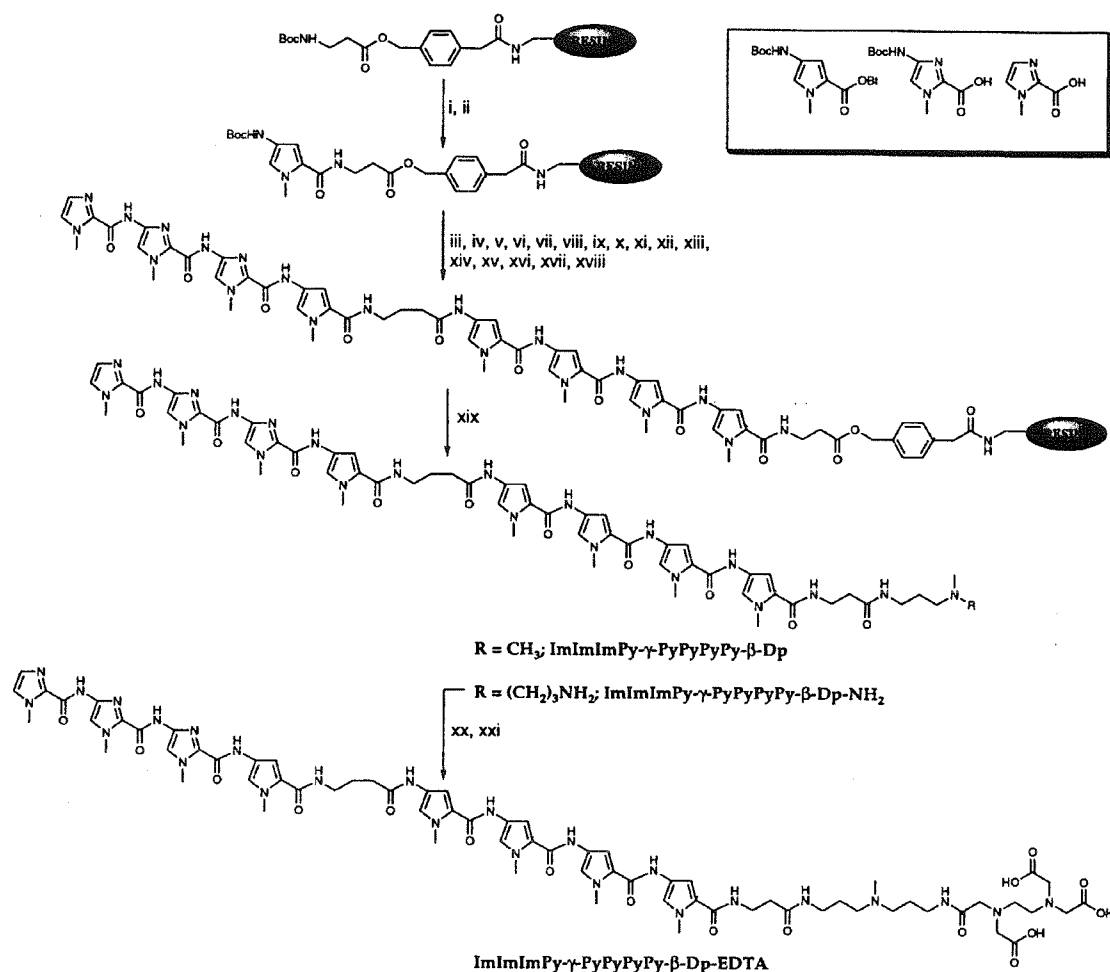
We report here the affinities and relative specificities of these polyamides as determined by three separate techniques: MPE•Fe(II) footprinting, DNase I footprinting, and affinity cleaving. Information about binding site size is gained from MPE•Fe(II) footprinting, while quantitative DNase I footprint titrations allow the determination of equilibrium association constants ( $K_a$ ) for the polyamides to a variety of match and single base pair mismatch sequences. Affinity cleavage studies confirm the binding orientation and stoichiometry of the 1:1 hairpin:DNA complex.

The polyamides ImImIm- $\gamma$ -PyPyPy- $\beta$ -Dp (**11**) and ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp (**12**) were synthesized in a stepwise manner from Boc- $\beta$ -alanine-Pam resin using recently described Boc-chemistry protocols in 14 and 18 steps, respectively. The polyamides were then cleaved by a single step aminolysis reaction with dimethylaminopropylamine and subsequently purified by HPLC chromatography. The syntheses of the polyamides, ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp (**12**), ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (**12-NH<sub>2</sub>**), and ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp-EDTA (**12-E**), are outlined (Figure 4.10). For the synthesis of the EDTA analog, a sample of resin is treated with 3,3'-diamino-*N*-methyldipropylamine (55°C) and then purified by preparatory HPLC to provide **12-NH<sub>2</sub>**. The polyamide ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (**12-NH<sub>2</sub>**) provides a free aliphatic primary amine group suitable for modification. This polyamide-amine is then treated

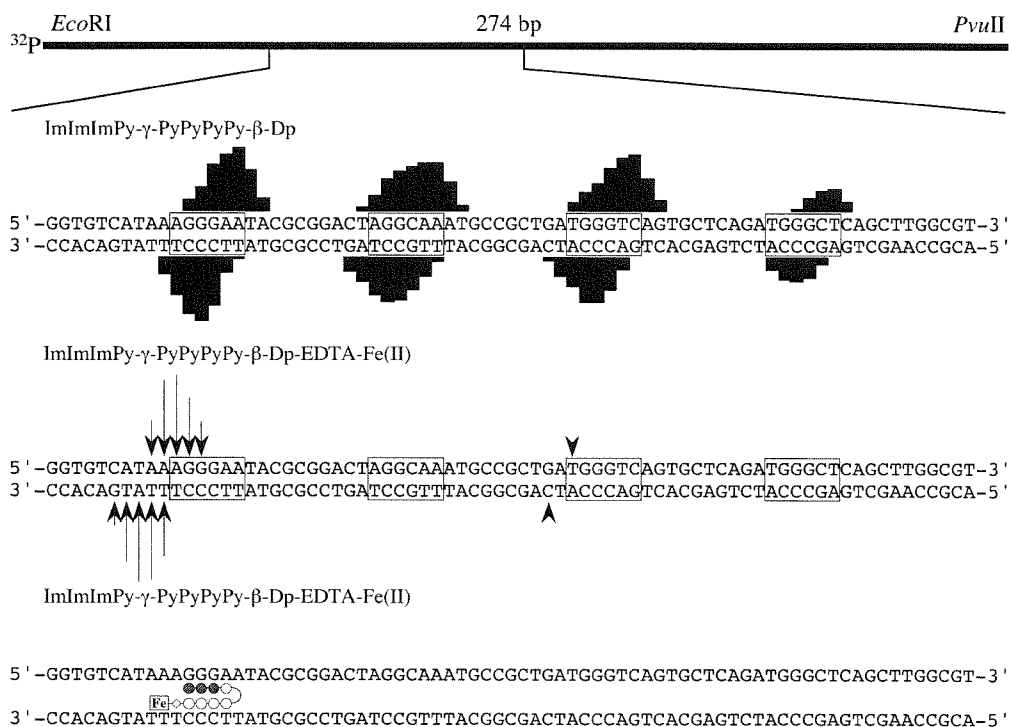
with an excess of the dianhydride of EDTA (DMSO/NMP, DIEA, 55°C) and the remaining anhydride hydrolyzed (0.1M NaOH, 55°C). The EDTA modified polyamide is then isolated by preparatory HPLC.

MPE•Fe(II) footprinting on the 3'- and 5'-<sup>32</sup>P end-labeled 274 base pair *EcoRI/PvuII* restriction fragment from plasmid pSES1 (25 mM Tris-acetate, 10 mM NaCl, 100 μM calf thymus DNA, pH 7.0 and 22°C) reveals that the polyamides, each at 10 μM concentration, are binding to the four designed sites, 5'-AGGGA(A)-3', 5'-AGGCA(A)-3', 5'-TGGGT(C)-3' and 5'-TGGGC(T)-3' (Figure 4.11). No footprinting is seen with either compound at the single base pair mismatch site, 5'-AGGAA(A)-3'. The footprinting patterns for the six-ring polyamide are consistent with five base pair binding sites, while the patterns for the eight-ring polyamide are indicative of six base pair binding sites.

Affinity cleavage experiments on the same restriction by the six-ring and eight-ring EDTA•Fe(II) analogs reveal cleavage patterns that are 3'-shifted and appear on only one side of each binding site. ImImIm-γ-PyPyPy-β-Dp-EDTA•Fe(II), at 1 μM, and ImImImPy-γ-PyPyPyPy-β-Dp-EDTA•Fe(II), at 100 nM, show cleavage patterns that demonstrate recognition of the same binding sites identified by MPE•Fe(II) footprinting. No carrier DNA was used in these experiments, and thus the relative cleavage intensities indicate that the six-ring polyamide binds most strongly to the two match sites 5'-AGGGA-3' and 5'-TGGGT-3', followed by the end mismatch 5'-TGGGC-3'. The core mismatch 5'-AGGCA-3', with little appreciable cleavage at 1 μM concentration, is bound more weakly. Similarly, the eight-ring polyamide binds most strongly at 100 nM to 5'-AGGGAA-3', much less strongly to 5'-TGGGTC-3', and not significantly to 5'-TGGGCT-3' and 5'-AGGCAA-3'.



**Figure 4.10.** (Box) Boc-Py-OBt ester, Boc-Im-acid, and imidazole-2-carboxylic acid. Solid phase synthetic scheme for ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp, ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp- $\text{NH}_2$  and ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp-EDTA prepared from commercially available Boc- $\beta$ -alanine-Pam-resin (0.2 mmol/gram): (i) 80% TFA/DCM, 0.4M PhSH; (ii) BocPy-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4M PhSH; (iv) BocPy-OBt, DIEA, DMF; (v) 80% TFA/DCM, 0.4M PhSH; (vi) BocPy-OBt, DIEA, DMF; (vii) 80% TFA/DCM, 0.4M PhSH; (viii) BocPy-OBt, DIEA, DMF; (ix) 80% TFA/DCM, 0.4M PhSH; (x) Boc- $\gamma$ -aminobutyric acid (HBTU, DIEA), DMF; (xi) 80% TFA/DCM, 0.4M PhSH; (xii) BocPy-OBt, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4M PhSH; (xiv) BocIm-OBt (DCC/HOBt), DIEA, DMF; (xv) 80% TFA/DCM, 0.4M PhSH; (xvi) BocIm-OBt (DCC/HOBt), DIEA, DMF; (xvii) 80% TFA/DCM, 0.4M PhSH; (xviii) imidazole-2-carboxylic acid (HBTU/DIEA); (xix) *N,N*-dimethylaminopropylamine or 3,3'-diamino-*N*-methyldipropylamine, 55°C; (xx) EDTA-dianhydride, DMSO/NMP, DIEA, 55°C. (xxi) 0.1M NaOH.



**Figure 4.11.** Results from MPE•Fe(II) footprinting of ImImImPy-γ-PyPyPyPy-β-Dp and affinity cleavage of ImImImPy-γ-PyPyPyPy-β-Dp-EDTA•Fe(II). (Top) Illustration of the 274 bp restriction fragment with the position of the sequence indicated. (second) MPE•Fe(II) protection patterns for polyamides at 10 μM concentration. Bar heights are proportional to the relative protection from cleavage at each band. (third) Affinity cleavage pattern of ImImImPy-γ-PyPyPyPy-β-Dp-EDTA•Fe(II) at 100 nM, arrow heights are proportional to the relative cleavage intensities at each base pair. (bottom) Ball and stick binding models for binding by the eight-ring EDTA•Fe(II) analog. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent the β-alanine residue. The boxed Fe denotes the EDTA•Fe(II) cleavage moiety.

Quantitative DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub> and 5 mM CaCl<sub>2</sub>, pH 7.0 and 22°C) were performed to determine the equilibrium association constants of the polyamides for the four bound sites (Table 4.3). ImImIm-γ-PyPyPy-β-Dp binds the two match sites 5'-AGGGA-3' and 5'-TGGGT-3' with equilibrium association constants of  $K_a = 4.6 \times 10^6 \text{ M}^{-1}$  and  $K_a = 7.6 \times 10^6 \text{ M}^{-1}$ , respectively. The sequence 5'-TGGGC-

**Table 4.3.** Equilibrium Association Constants ( $M^{-1}$ )<sup>a, b, c</sup>.

Match Sites	End Mismatch	Core Mismatch		
Polyamide	5'-AGGGA-3'	5'-TGGGT-3'	5'-TGGGC-3'	5'-AGGCA-3'
ImImIm- $\gamma$ -	$4.6 \times 10^6$	$7.6 \times 10^6$	$1.3 \times 10^6$	$8.6 \times 10^5$
PyPyPy- $\beta$ -Dp				

Match Site	End Mismatch	Core Mismatches		
Polyamide	5'-AGGGAA-3'	5'-TGGGTC-3'	5'-TGGGCT-3'	5'-AGGCAA-3'
ImImImPy- $\gamma$ -	$3.7 \times 10^8$	$1.4 \times 10^7$	$1.7 \times 10^6$	$2.9 \times 10^6$
PyPyPyPy- $\beta$ -Dp				

<sup>a</sup>Values reported are the mean values from a minimum of three DNase I footprint titration experiment. <sup>b</sup>The assays were performed at 22°C at pH 7.0 in the presence of 10 mM tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. <sup>c</sup>Base pairs that are in bold represent formal mismatches.

3', which has a mismatch in the final position is bound with 6-fold lower affinity than the best match, while the mismatch sequence 5'-AGG**C**A-3' is bound with 9-fold lower affinity. ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp binds the match site 5'-AGGGAA-3' with an equilibrium association constant of  $K_a = 3.7 \times 10^8 M^{-1}$ . The mismatch 5'-TGGG**T**C-3' is bound with 26-fold lower affinity, and the two mismatches, 5'-AGG**C**AA-3' and 5'-TGGG**C**T-3', are bound with 130-fold and 220-fold lower affinity, respectively. Neither polyamide shows any appreciable binding to the mismatch 5'-AGG**A**(A)-3'.

The two hairpin polyamides recognize the targeted 5'-AGGGA(A)-3' sequence, as determined by MPE•Fe(II) footprinting and affinity cleaving, demonstrating specific recognition by polyamides of sequences containing three contiguous G•C base pairs. Affinity cleavage data indicate that each polyamide is binding in the minor groove in a single orientation, consistent with the hairpin binding model. The relative intensities of the cleavage patterns are consistent with quantitative DNase I footprint titration results.

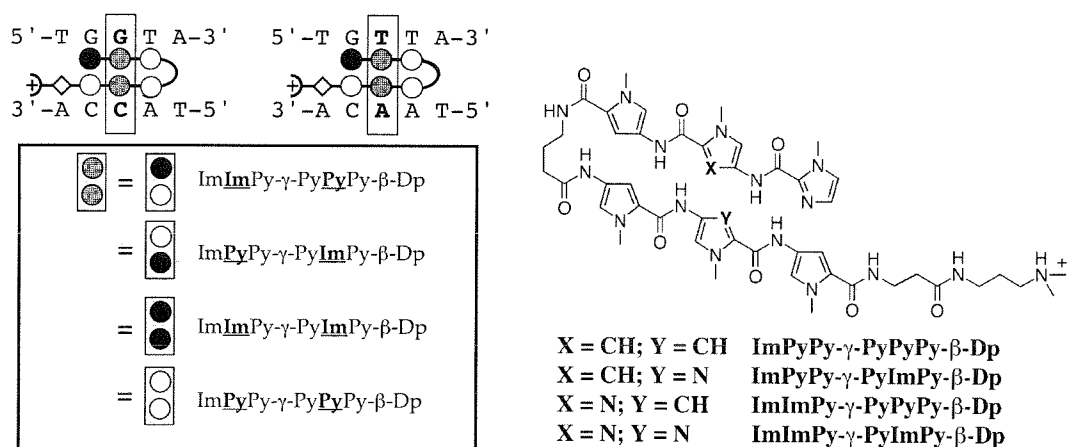
Quantitative DNase I footprint titrations reveal that ImImIm- $\gamma$ -PyPyPy- $\beta$ -Dp binds the match sites 5'-AGGGA-3' and 5'-TGGGT-3' with equilibrium association constants of  $K_a = 5 \times 10^6 \text{ M}^{-1}$  and  $K_a = 8 \times 10^6 \text{ M}^{-1}$ , respectively. For comparison, the analogous six-ring hairpins containing only one and two contiguous imidazoles, ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp and ImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp, bind their respective match sites with affinities of  $K_a = 10^8 \text{ M}^{-1}$ . Examination of the eight ring hairpin, ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp, reveals increased affinity, the match site 5'-AGGGAA-3' being bound with an equilibrium association constant of  $K_a = 4 \times 10^8 \text{ M}^{-1}$ . The 80-fold increase in affinity in expanding from a three-ring to four-ring hairpin polyamide mirrors the 66-fold enhancement of ImPyPyPy-Dp over ImPyPy-Dp that was observed in the 2:1 homodimeric polyamide:DNA motif.

Py-Im polyamides have been used to recognize a variety of target sequences containing A•T and G•C base pairs. By recognizing sequences containing *three* contiguous G•C base pairs, 5'-(A,T)GGG(A,T)-3' and 5'-(A,T)GGG(A,T)<sub>2</sub>-3', this work expands the sequence-composition repertoire targetable by the hairpin polyamide motif. Both affinity and specificity for a G,C rich sequence are increased by the use of an eight-ring hairpin polyamide. This ability to enlarge the sequence repertoire, combined with rapid solid phase synthesis, brings us one step closer to a universal approach for the recognition of any desired DNA sequence by strictly chemical methods.

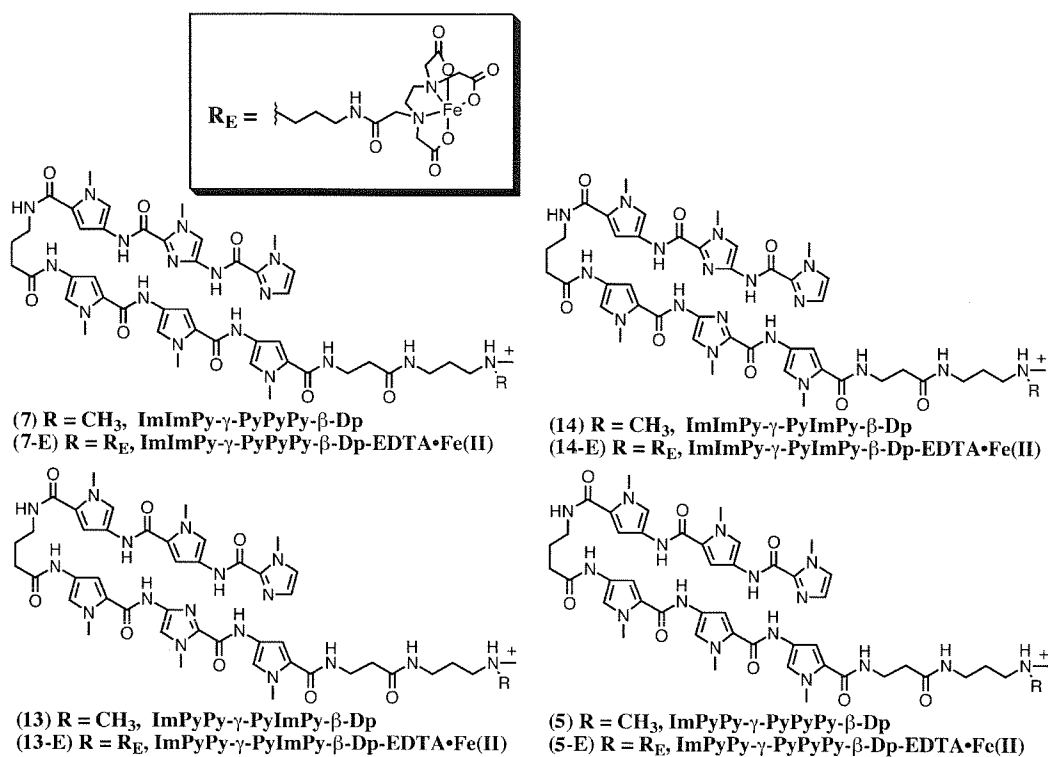


**An Im/Im Pairing is Disfavored.** Recognition of G•C by the Im/Py pairing requires precise positioning for the key hydrogen bond between the Im N3 and the exocyclic amine of guanine.<sup>13</sup> Given the central location of the guanine exocyclic amine group in the DNA minor groove,<sup>14-16</sup> the question arises whether an Im/Im pairing might also be expected to target G•C.<sup>17</sup> Remarkably, even in the first report on the binding specificity of the three ring polyamide homodimer (ImPyPy-Dp)<sub>2</sub>, there was qualitative data to suggest that there was indeed a binding preference for placement of the Im/Py pair opposite G•C.<sup>1</sup>

It would be useful to determine the generality of the aromatic amino acid pairing preferences and to compare the relative energetics of the four possible pairings of Im and Py for recognition of G•C and T•A base pairs. Therefore, we describe here an experimental design utilizing a single oriented six-ring hairpin polyamide which allows the relative energetic preferences of four different binary combinations, Im/Py, Py/Im, Im/Im, and Py/Py, to be tested opposite a G•C or an A•T base pair (Figure 4.12).



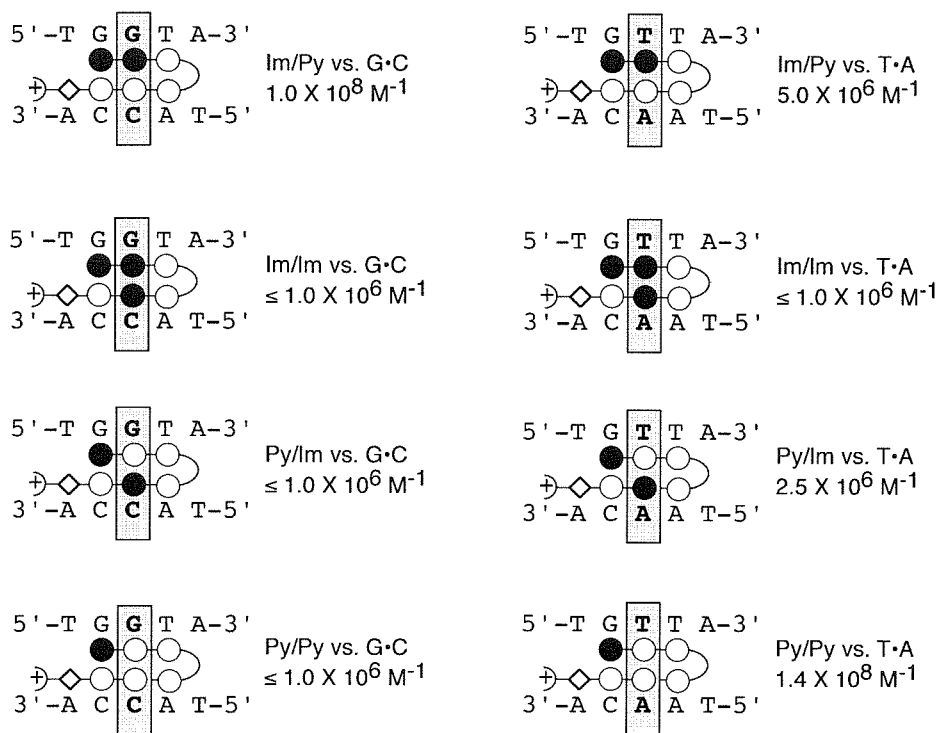
**Figure. 4.12** (left) Models of the expected “hairpin” complex of ImImPy-γ-PyPyPy-β-Dp **7**, ImPyPy-γ-PyImPy-β-Dp **13**, ImImPy-γ-PyImPy-β-Dp **14**, and ImPyPy-γ-PyPyPy-β-Dp **5** in complex with 5'-TGGTA-3' and 5'-TGTTA-3'. Unfilled white circles may be either Im or Py residues which are represented by filled and unfilled circles respectively. Diamonds represent β-alanine, and γ-aminobutyric acid is represented as a curved line. The central pairing of Im/Py, Py/Im, Im/Im and Py/Py with G•C and T•A is highlighted with a gray box. (right) Structures of the corresponding hairpin polyamides.



**Figure. 4.13.** Structures of polyamides ImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp **7**, ImPyPy- $\gamma$ -PyImPy- $\beta$ -Dp **13**, ImImPy- $\gamma$ -PyImPy- $\beta$ -Dp **14**, and ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **5**, ImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA•Fe(II) **7-E•Fe(II)**, ImPyPy- $\gamma$ -PyImPy- $\beta$ -Dp-EDTA•Fe(II) **13-E•Fe(II)**, ImImPy- $\gamma$ -PyImPy- $\beta$ -Dp-EDTA•Fe(II) **14-E•Fe(II)**, and ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA•Fe(II) **5-E•Fe(II)**.

Three-ring polyamide subunits covalently coupled C-N by a central  $\gamma$ -aminobutyric acid linker form hairpin structures at 5-bp target sequences.<sup>6</sup> For example, according to the pairing rules, polyamides of sequence composition ImImPy- $\gamma$ -PyPyPy- $\beta$  and ImPyPy- $\gamma$ -PyPyPy- $\beta$  (central amino acid pairing underlined) would be expected to bind to 5'-TGGTA-3' and 5'-TGTTA-3' sequences respectively. Selective substitution of the central amino acid of each three ring polyamide subunit allows for four-ring pairings at a unique central location within the hairpin structure placed opposite a G•C or T•A base pair (Figure 4.12).

Four six-ring polyamides, ImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp **7**, ImPyPy- $\gamma$ -PyImPy- $\beta$ -Dp **13**, ImImPy- $\gamma$ -PyImPy- $\beta$ -Dp **14**, and ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **5**, containing central amino pairings of Im/Py, Py/Im, Im/Im, and Py/Py in a hairpin structure were synthesized by solid phase methods



**Figure 4.14.** Models of the expected "hairpin" complex of ImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp **7**, ImPyPy- $\gamma$ -PyImPy- $\beta$ -Dp **13**, ImImPy- $\gamma$ -PyImPy- $\beta$ -Dp **14**, and ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **5** in complex with 5'-TGGTA-3' and 5'-TGTTA-3'. Filled and unfilled circles represent Im and Py rings, respectively. Diamonds represent  $\beta$ -alanine, and  $\gamma$ -aminobutyric acid is represented as a curved line. The central pairing of Im/Py, Py/Im, Im/Im and Py/Py with G•C and T•A is highlighted with a gray box. The binding affinity for each complex as determined from quantitative DNase I footprinting is shown to the right of each complex.

(Figure 4.13). The corresponding EDTA analogs ImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA **7-E**, ImPyPy- $\gamma$ -PyImPy- $\beta$ -Dp-EDTA **13-E**, ImImPy- $\gamma$ -PyImPy- $\beta$ -Dp **14-E**, and ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **5-E** were also constructed in order to confirm the single binding orientation of each hairpin:DNA complex.

We report here the DNA-binding affinities, orientations, and sequence-selectivity of the four polyamides for two five base pair binding sites, 5'-TGTTA-3' and 5'-TGGTA-3', which vary at one unique third position. Three separate techniques are used to characterize the DNA-binding properties of the polyamides: affinity cleaving, MPE•Fe(II) and DNase I footprinting.

Affinity cleavage studies determine the specific binding orientation and stoichiometry of each hairpin:DNA complex. Binding location and site size is accurately determined by MPE•Fe(II) footprinting, while quantitative DNase I footprint titration is more suitable for measurement of equilibrium association constants ( $K_a$ ) for the polyamide binding to designated sequences.

MPE•Fe(II) footprinting on 3' and 5'-<sup>32</sup>P end labeled 302 base pair restriction fragments (25 mM tris-acetate, 10 mM NaCl, 100  $\mu$ M bp calf thymus DNA, 5 mM DTT, pH 7.0 and 22°C) reveals that the polyamides bind and discriminate the two five base pair sites, 5'-TGTTA-3' and 5'-TGGTA-3'. Polyamides **7** and **13** at  $\mu$ M concentrations bind in decreasing affinity to 5'-TGGTA-3' > 5'-TGTTA-3'. Polyamides **5** and **14** at  $\mu$ M concentrations bind in reverse order with decreasing affinity 5'-TGTTA-3' > 5'-TGGTA-3'.

	G•C	T•A
<b>Im/Py</b>	<b>+</b>	<b>-</b>
<b>Py/Im</b>	<b>-</b>	<b>-</b>
<b>Im/Im</b>	<b>-</b>	<b>-</b>
<b>Py/Py</b>	<b>-</b>	<b>+</b>

**Figure 4.15.** Pairing rules for Im and Py aromatic amino acids.

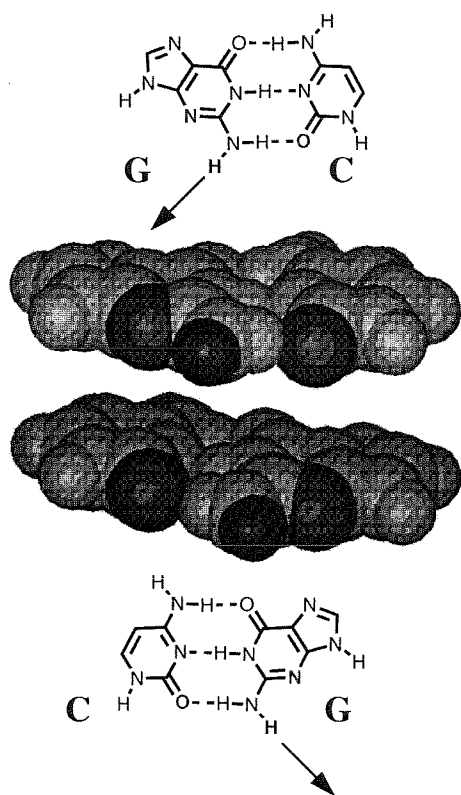
Affinity cleaving experiments using hairpin polyamides modified with EDTA•Fe(II) at the C-terminus were used to determine polyamide binding orientation and stoichiometry. Experiments were performed on the same 3' and 5'-<sup>32</sup>P end labeled 302 base restriction fragments (25 mM tris-acetate, 20 mM NaCl, 100  $\mu$ M bp calf thymus DNA, pH 7, 22°C, 10 mM DTT, 10  $\mu$ M Fe(II)). The observed cleavage patterns are in all cases 3'-shifted, consistent with minor groove occupancy. A single cleavage locus proximal to the 5'-side of both the 5'-TGTTA-3' and 5'-TGGTA-3' binding sites confirm that the four polyamides bind each discrete site with a single orientation. The observation of a single cleavage locus is consistent only with an oriented 1:1

complex in the minor groove of DNA and rules out dimeric overlapped or extended binding motifs. A 1:1 oriented but extended motif would require at least an eight base pair binding site, which is inconsistent with the high-resolution MPE footprinting data on both target sites. The hairpin complex ImPyPy- $\gamma$ -PyPyPy-Dp•5'-TGTTA-3' has recently been characterized by direct NMR methods.<sup>18</sup>

MPE•Fe(II) footprinting combined with affinity cleaving experiments indicate that each polyamide binds the designated five base pair target site as a 1:1 hairpin complex in the minor groove. This single and consistent mode of binding allows valid thermodynamic comparison for the central ring amino acid pairing of each polyamide recognizing the central base pair of each designated target site. Quantitative DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5mM CaCl<sub>2</sub>, pH 7.0, 22°C) were performed to determine the equilibrium association constants  $K_a$  for recognition of the bound sites. The 5'-TGGTA-3' site is bound by the polyamides with decreasing affinity: ImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp (**7**) >> ImPyPy- $\gamma$ -PyImPy- $\beta$ -Dp (**13**)  $\approx$  ImImPy- $\gamma$ -PyImPy- $\beta$ -Dp (**14**)  $\approx$  ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp (**5**). The 5'-TGTTA-3' site is bound with decreasing affinity ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp (**5**) > ImPyPy- $\gamma$ -PyImPy- $\beta$ -Dp (**13**)  $\approx$  ImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp (**7**) > ImImPy- $\gamma$ -PyImPy- $\beta$ -Dp (**14**). Remarkably, the association constant for recognition of each site varies 100-fold between the four polyamides, indicating a remarkable sensitivity to a single atomic substitution within the central ring amino acids (Figure 4.14).

Among the four ligands, ImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp **7** (central Im/Py pairing) binds to the 5'-TGGTA-3' site which contains a central G•C base with the highest affinity ( $K_a = 9.0 \times 10^7 \text{ M}^{-1}$ ). This selectivity indicates that Im/Py is the optimal ring pairing for recognition of G•C. The sequence-specificity of the Im/Py pairing is underscored by the 50-fold reduced affinity ( $K_a = 1.7 \times 10^6 \text{ M}^{-1}$ ) for placement of the Im/Py pair opposite a T•A base pair at 5'-TGTTA-3'.

The polyamide ImPyPy- $\gamma$ -PyImPy- $\beta$ -Dp **13** (central Py/Im pairing) binds to the 5'-TGGTA-3' site (central G•C base pair) with 100-fold reduced affinity relative to polyamide **7**



**Figure 4.16.** Space-filling model of the G•C and C•G base pairs as viewed from the minor groove of DNA, generated using B-from DNA coordinates provided in InsightII. The central location of the 2-amino nitrogen is highlighted in dark gray. The asymmetrically directed amino proton is highlighted as well.

(central Im/Py pairing). Remarkably, given the central location of the exocyclic 2-amino group of guanine in the minor groove, the Py/Im pairing is disfavored relative to the Im/Py pair for recognition of G•C. Although the nitrogen of the 2-amino group is displayed in a similar location for both G•C and C•G, the proton available for hydrogen bond recognition in the minor groove has a *strand specific directionality* (Figure 4.16). The directionality requirements for effective hydrogen bond formation allow discrimination of G•C by the Im/Py and Py/Im pairs. This interpretation has been validated by high resolution x-ray structure studies.<sup>13</sup> Placement of the Py/Im pair of **13** opposite T•A at the 5'-TGTTA-3' target site results in similar affinity ( $K_a = 1.7 \times 10^6 \text{ M}^{-1}$ ) to placement of the Im/Py pair **5** opposite T•A. It should be noted that the "Py/Im" hairpin polyamide **13** recognizes a 5'-AGCTT-3' match site present on the restriction fragment (central C•G pair) with high affinity.

The polyamide ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **5** (central Py/Py pairing) binds to the 5'-TGTTA-3' site (central T•A base pair) with 100-fold enhanced affinity relative to the 5'-

TGGTA-3' site (central G•C base pair). A Py/Py pairing discriminates A•T/T•A base pairs from G•C/C•G, likely due to the exocyclic amine groups of guanine which present a steric hindrance to deep polyamide binding in the minor groove. However, Im rich hairpin polyamides recognize G,C sequences with affinities and specificities similar to Py rich polyamides that recognize A•T rich sequences indicating, as will become evident below, that additional energetic parameters may be important for high affinity recognition.<sup>19</sup>

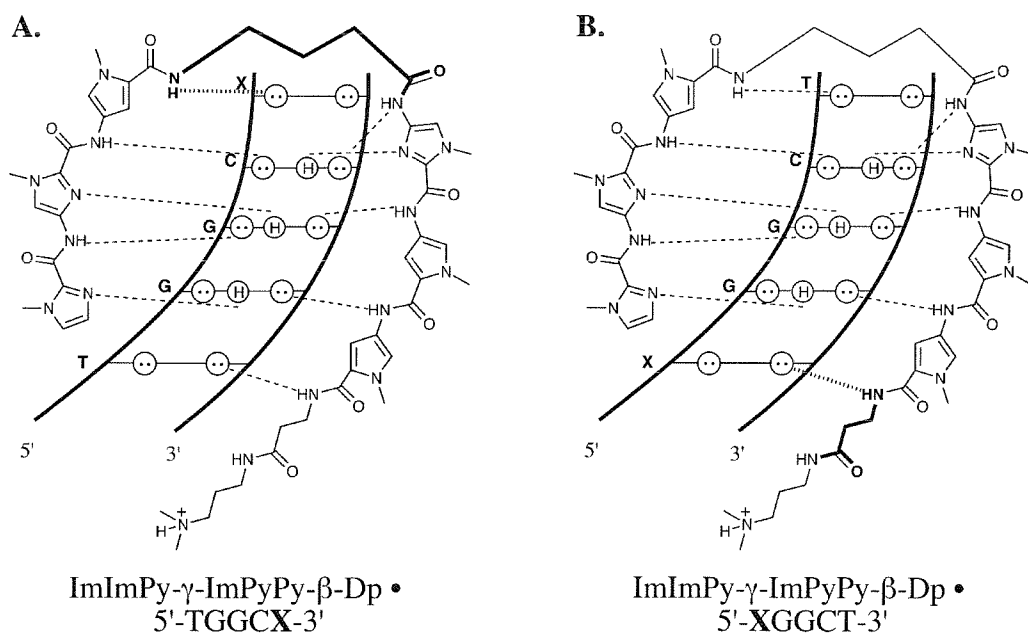
The hairpin ImImPy- $\gamma$ -PyImPy- $\beta$ -Dp (central Im/Im pair) binds both designated sites, 5'-TGTTA-3' and 5'-TGGTA-3', with > 100-fold reduced affinity relative to polyamides **7** and **5** (central Py/Py and Im/Py pair, respectively). The reduced binding energetics of the Im/Im pair may result from desolvation of the N-3 of Im upon binding to DNA.<sup>20</sup> Such unfavorable desolvation could be compensated when a hydrogen bond is formed between Im and the exocyclic 2-amino group of G, but not compensated upon placement of Im opposite A, T, or C.

Alternatively, specificity may result from electrostatic interactions between the aromatic ring carboxamides and the floor of the minor groove.<sup>21</sup> The DNA minor groove displays a negative electrostatic potential at A•T and T•A base pairs.<sup>22</sup> The only positive potential is at the 2-amino group of G.<sup>22</sup> The Py ring displays a positive potential (calculated with Spartan<sup>23</sup>) across the interface with the floor of the minor groove, providing complementary electrostatic interactions at A, T, and C bases. The Im ring displays a negative potential which can interact favorably with the 2-amino group of G, but may interact unfavorably with the other three bases.<sup>24</sup>

The results described here demonstrate the sequence-specificity of the four individual ring amino acid pairings: (1) The Im/Py pairing recognizes G•C, but is disfavored when placed opposite T•A. (2) The Py/Im pairing targets C•G, but is disfavored for placement opposite G•C and T•A base pairs. (3) The Py/Py pairing recognizes T•A/A•T base pairs but is disfavored opposite G•C/C•G. (4) An Im/Im pairing is disfavored at all four base pairs breaking a potential degeneracy for recognition by preventing unlinked polyamide dimers from binding in certain slipped motifs. These results create baseline energetic parameters which will guide further second-generation polyamide design for DNA recognition.

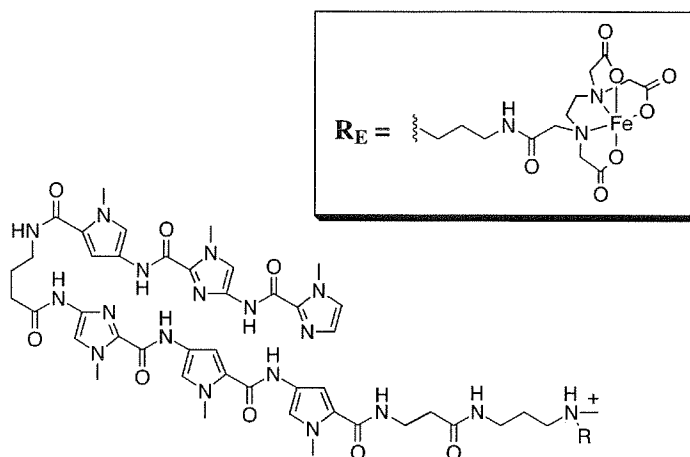
**Effects of  $\gamma$ -Turn and  $\beta$ -Tail Amino Acids on the Sequence-Specific Recognition of the DNA Minor Groove by Hairpin Polyamides.** For a polyamide containing a number of consecutive ring pairings,  $n$ , the binding site size will be  $n + 2$  base pairs.<sup>1</sup> Each side-by-side ring pairing recognizes a central base pair, and the outer base pairs are formally recognized by the terminal amides.<sup>1</sup> For example, according to the “pairing rules” the six-ring hairpin-polyamide ImImPy- $\gamma$ -ImPyPy- $\beta$ -Dp (Dp = dimethylaminopropylamine) will bind to a five base pair 5′-(A,T)GGC(A,T)-3′ sequence, where the GGC “core” is recognized by the Im/Py and Py/Im ring pairings (Figure 4.17).<sup>2</sup> The sequence-specificity of the  $\gamma$ -turn and  $\beta$ -tail amino acids for the (A,T) and (T,A) base pairs, which are not placed directly opposite aromatic ring pairings, has not been examined in detail.

Eight-ring hairpin polyamides have recently been shown to be cell permeable and to inhibit the transcription of specific genes.<sup>25</sup> Understanding the effects of the  $\gamma$ -turn and  $\beta$ -tail amino acids on the sequence-specificity of hairpin polyamides for the base pairs surrounding the “core” sequence could potentially enhance the predictability of sequence-targeting with hairpin polyamides.<sup>26</sup> In order to investigate the sequence-specific effects of the  $\gamma$ -turn and  $\beta$ -tail amino



**Figure 4.17.** Binding models for the complex formed between ImImPy- $\gamma$ -ImPyPy- $\beta$ -Dp (**15**) and either (a) 5′-TGGCX-3′, with the tail position varied or (b) 5′-XGGCT-3′, with the turn position varied. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.





(15)  $R = \text{CH}_3$ , ImImPy- $\gamma$ -ImPyPy- $\beta$ -Dp

(15-E•Fe(II))  $R = R_E$ , ImImPy- $\gamma$ -ImPyPy- $\beta$ -Dp-EDTA•Fe(II)

**Figure 4.18.** Structures of the six-ring hairpin polyamides **15** and the corresponding Fe(II)•EDTA affinity cleaving derivative.

acids, a heterodimeric hairpin polyamide, ImImPy- $\gamma$ -ImPyPy- $\beta$ -Dp (**15**), and the corresponding affinity cleaving analog (**15-E**) were prepared (Figures 4.18). The single binding orientation predicted for this hairpin-polyamide<sup>27</sup> allows the effects of the  $\gamma$  and  $\beta$  amino acids to be determined at the eight primary 3' and 5' base pairs, 5'-ATGGCNA-3' and 5'-ANGGCTA-3'; 5-bp hairpin recognition site is underlined. It was unknown whether adjacent base pairs would also be discriminated by the linker amino acids. Thus, the binding of polyamide **1** was also examined on DNA-restriction fragments containing the eight possible 5'-ATGGCTN-3' and 5'-NTGGCTA-3' sequences. A hairpin polyamide ImImPy-( $R$ )<sup>H<sub>2</sub>N</sup> $\gamma$ -ImPyPy-PrOH (**17**) (where ( $R$ )<sup>H<sub>2</sub>N</sup> $\gamma$  = ( $R$ )-2,4-diaminobutyric acid and PrOH = propanolamide),<sup>28</sup> was analyzed as a "tail-less" control to exclude the possibility of sequence dependent DNA microstructure effects on the tail specificity and to determine the effect of the C-terminal amide bond.

We report here the affinities, binding locations, and relative selectivities of these polyamides as determined by three separate techniques: MPE•Fe(II) footprinting, affinity cleaving, and DNase I footprinting. MPE•Fe(II) footprinting (25 mM Tris-acetate, 10 mM NaCl, 100  $\mu$ M/base pair calf thymus DNA, pH 7.0 and 22°C) was performed on the 3'- and 5'-<sup>32</sup>P end-labeled 286 base pair restriction fragments from the cloned plasmids pSES-TN1 and pSES-TL1. pSES-TN1 contains the four possible 5'-ATGGCNA-3' (turn specificity) sites while pSES-TL1 contains the four possible 5'-ANGGCTA-3' (tail-specificity) sites, with identical sequences

between sites. MPE•Fe(II) footprinting reveals that polyamide **15**, at 10  $\mu$ M concentration, binds to three of the four binding sites: 5'-ATGGCTA-3', 5'-ATGGCAA-3' and 5'-ATGGCCA-3' (A•T, T•A, and C•G) base pairs at the turn position but not 5'-ATGGCGA-3', which has a G•C base pair at the turn position. For the tail position, MPE•Fe(II) footprinting reveals that polyamide **15**, at 10  $\mu$ M concentration, binds to only two of the four binding sites on the restriction fragment, 5'-ATGGCTA-3' and 5'-AAGGCTA-3' (A•T and T•A at the tail position), but not 5'-AGGGCTA-3' and 5'-ACGGCTA-3' (G•C and C•G at the turn position). The size and 3'-shift of the protection patterns is consistent with the polyamide binding as a hairpin in the DNA minor groove.

Affinity cleavage experiments were performed (25 mM Tris-acetate, 200 mM NaCl, 50  $\mu$ g/mL glycogen, pH 7.0 and 22°C)<sup>15</sup> on the “turn” restriction fragment pSES-TN1 which contains the four possible 5'-ATGGCNA-3' sites. The affinity cleavage assays reveal cleavage patterns that are 3'-shifted and appear on only the 5' side of the 5'-ATGGCTA-3' and 5'-ATGGCAA-3' sites, as is expected for hairpin formation. Strong affinity cleavage is seen at only the 5'-side of the two recognized sites, consistent with formation of an oriented hairpin polyamide complex in the minor groove. Equal amounts of cleavage is seen on both sides of the 5'-ATGGCCA-3' site, which can be explained by the fact the site contains the palindromic sequence 5'-TGGCCA-3', where two identical sites are present, 5'-ATGGCCA-3' and 5'-ATGGCCA-3'.

Quantitative DNase I footprint titrations (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub> and 5 mM CaCl<sub>2</sub>, pH 7.0 and 22°C) were performed to determine the equilibrium association constants ( $K_a$ ) of polyamide **15** for the eight sites on the two restriction fragments (Table 4.4). For the turn position, the 5'-ATGGCAA-3' and 5'-ATGGCTA-3' sites are bound with equivalent equilibrium association constants of  $K_a = 2.0 \times 10^7 \text{ M}^{-1}$  and  $2.1 \times 10^7 \text{ M}^{-1}$ , respectively. The 5'-ATGGCCA-3' site is bound with 28-fold lower affinity ( $K_a = 7.4 \times 10^5 \text{ M}^{-1}$ ), while no binding is observable at the 5'-ATGGCGA-3' site ( $K_a < 5 \times 10^4 \text{ M}^{-1}$ ). The  $\beta$ -tail residue has comparable sequence preference with the 5'-AAGGCTA-3' and 5'-ATGGCTA-3' sites being bound with similar affinities,  $K_a = 3.1 \times 10^7 \text{ M}^{-1}$  and  $K_a = 2.1 \times 10^7 \text{ M}^{-1}$ , respectively. No binding is seen at either the 5'-ACGGCTA-3' and 5'-AGGGCTA-3' sites ( $K_a < 10^5 \text{ M}^{-1}$ ), such that turn-sites containing A•T and T•A base pairs are bound with > 200-fold higher affinity than those with G•C and C•G base pairs.

**Table 4.4.** Equilibrium Association Constants ( $M^{-1}$ )<sup>a,b</sup>.

Binding Site	ImImPy- $\gamma$ -ImPyPy- $\beta$ -Dp	Specificity <sup>c</sup>
5'- <u>ATGGCAA</u> -3'	$2.0 \times 10^7$ (0.3)	> 400
5'- <u>ATGGCTA</u> -3'	$2.1 \times 10^7$ (0.1)	> 420
5'- <u>ATGGCCA</u> -3'	$7.4 \times 10^5$ (1.4) <sup>d</sup>	> 15
5'- <u>ATGGCGA</u> -3'	$< 5 \times 10^4$ <sup>d</sup>	1
5'- <u>AAGGCTA</u> -3'	$3.1 \times 10^7$ (1.1)	> 310
5'- <u>ATGGCTA</u> -3'	$2.1 \times 10^7$ (0.8)	> 210
5'- <u>ACGGCTA</u> -3'	$< 10^5$ <sup>d</sup>	1
5'- <u>AGGGCTA</u> -3'	$< 10^5$ <sup>d</sup>	1

<sup>a</sup>Values reported are the mean values measured from at least three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. <sup>b</sup>The assays were performed at 22°C at pH 7.0 in the presence of 10 mM tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. <sup>c</sup>Specificity is calculated as  $K_a(5'-\text{ATGGC}\underline{\text{N}}\text{A}-3')$  /  $K_a(5'-\text{ATGGC}\underline{\text{G}}\text{A}-3')$  for the data from pSES-TN1 and as  $K_a(5'-\text{A}\underline{\text{N}}\text{GGCTA}-3')$  /  $K_a(5'-\text{A}\underline{\text{N}}\text{GGCTA}-3')$  for the data from pSES-TL1. <sup>d</sup>These equilibrium association constants were measured in a 40  $\mu\text{L}$  reaction volume instead of the standard 400  $\mu\text{L}$  (see Experimental).

Footprinting experiments were also performed on the 286 base pair restriction fragments (pSES-TN2 and pSES-TL2) which contain the 8 possible 5'-NTGGCTA-3' and 5'-ATGGCTN-3' sites (Table 4.5). ImImPy- $\gamma$ -ImPyPy- $\beta$ -Dp (**15**) again shows preference for A•T and T•A at the secondary turn position. However, the magnitude of the discrimination at this position is greatly reduced with equilibrium association constants at the four sites of: 5'-ATGGCTG-3' ( $K_a = 1.2 \times 10^7 M^{-1}$ ) < 5'-ATGGCTC-3' ( $K_a = 2.6 \times 10^7 M^{-1}$ ) < 5'-ATGGCTT-3' ( $K_a = 3.6 \times 10^7 M^{-1}$ ) < 5'-ATGGCTA-3' ( $K_a = 4.1 \times 10^7 M^{-1}$ ). Similar effects are seen for polyamide **15** binding at the secondary tail position with a preference for A•T and T•A over G•C and C•G base pairs. The 5'-TTGGCTA-3' ( $K_a = 2.8 \times 10^7 M^{-1}$ ) and 5'-ATGGCTA-3' ( $K_a = 6.2 \times 10^7 M^{-1}$ ) sites are bound with the highest affinity. The 5'-CTGGCTA-3' ( $K_a = 2.0 \times 10^6 M^{-1}$ ) and 5'-GTGGCTA-3' ( $K_a =$

**Table 4.5.** Equilibrium Association Constants ( $M^{-1}$ )<sup>a,b</sup>.

Binding Site	ImImPy- $\gamma$ -ImPyPy- $\beta$ -Dp	Specificity <sup>c</sup>
5'-ATGGCTA-3'	$4.1 \times 10^7$ (0.9)	3
5'-ATGGCTT-3'	$3.6 \times 10^7$ (1.3)	3
5'-ATGGCTC-3'	$2.6 \times 10^7$ (0.6)	2
5'-ATGGCTG-3'	$1.2 \times 10^7$ (0.3)	1
5'-ATGGCTA-3'	$6.2 \times 10^7$ (0.2)	44
5'-TTGGCTA-3'	$2.8 \times 10^7$ (0.4)	20
5'-CTGGCTA-3'	$2.0 \times 10^6$ (0.8)	1
5'-GTGGCTA-3'	$1.4 \times 10^6$ (1.4)	1

<sup>a</sup>Mean values measured from at least three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. <sup>b</sup>The assays were performed at 22°C at pH 7.0 in the presence of 10 mM tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. <sup>c</sup>Specificity is calculated as  $K_a(5'-\text{ATGGCTN-3}') / K_a(5'-\text{ATGGCTG-3}')$  for data from pSES-TN2 and as  $K_a(5'-\text{NTGGCTA-3}') / K_a(5'-\text{GTGGCTA-3}')$  pSES-TL2.

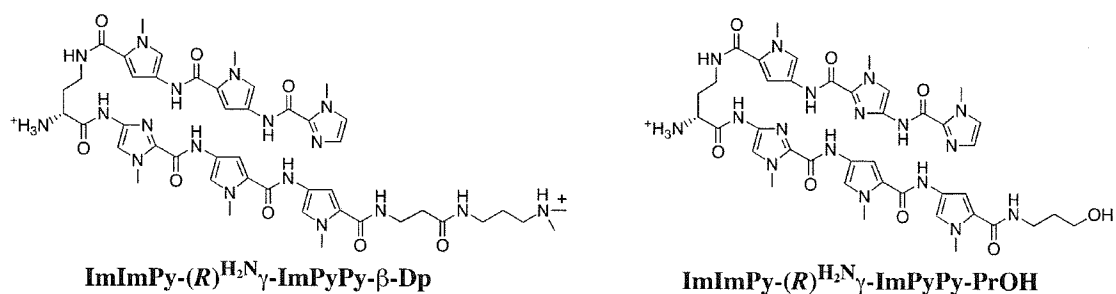
$1.4 \times 10^6 M^{-1}$ ) sites are bound with 20-fold and 44-fold reduced affinity, respectively (relative to the 5'-GTGGCTA-3' site).

ImImPy- $\gamma$ -ImPyPy- $\beta$ -Dp (**15**) was examined by quantitative DNase I footprinting against the eight possible 5'-ATGGCNA-3' (primary) and 5'-ATGGCTN-3' (secondary) sequences. At the primary turn position (5'-ATGGCNA-3'), the hairpin polyamide demonstrates a large preference for A,T base pairs, with both 5'-ATGGCTA-3' and 5'-ATGGCAA-3' bound with an equilibrium association constant of  $K_a = 2 \times 10^7 M^{-1}$ . A C•G base pair (5'-ATGGCCA-3') sequence is recognized with 28-fold reduced affinity. A sequence with a G•C base pair in this position, 5'-ATGGCGA-3', is bound with > 400-fold reduced affinity. The preference of the  $\gamma$ -turn for A•T and T•A base pairs is not unexpected, due to the steric factors, and has been noted

before. Presumably, the amide at the C-terminal end of the first subunit makes a hydrogen bond to the N3 of A or O2 of T at the turn position as observed for C-terminal aromatic amide in the crystal structure of the dimeric complex of ImImPyPy.<sup>13</sup> The preference for C•G over G•C base pairs has not been observed before. This effect is consistent with the preference of paired Py amide residues for placement opposite T, A, or C relative to G. Placement of the turn-amide opposite G could result in unfavorable steric interaction or an unfavorable electrostatic effect. Alternatively, the DNA-binding preference could result from sequence-dependant microstructure effects.

At the secondary turn position (5'-ATGGCTN-3'), the hairpin polyamide demonstrates a much smaller preference for A,T base pairs than at the primary turn position (5'-ATGGCNA-3'). Both 5'-ATGGCTA-3' and 5'-ATGGCTT-3' are bound with 2-fold higher affinity than 5'-ATGGCTC-3' and 3-fold higher affinity than 5'-ATGGCTG-3'. It is interesting to note that the relative preference ( $A \approx T > C > G$ ) is observed for polyamide **15** for binding to both 5'-ATGGCTN-3' and 5'-ATGGCNA-3' sequences. The reduced magnitude of the discrimination is consistent with the compact  $\gamma$ -turn observed in an NMR structure, indicating that the  $\gamma$ -turn should make significantly greater contact with the primary relative to the secondary base pair.

ImImPy- $\gamma$ -ImPyPy- $\beta$ -Dp (**15**) was also examined against the eight possible 5'-ANGGCTA-3' (primary) and 5'-NTGGCTA-3' (secondary) tail sequences. As seen for the  $\gamma$ -turn, the  $\beta$ -tail amino acid is specific for A•T and T•A base pairs at the primary position. The 5'-AAGGCTA-3' and 5'-ATGGCTA-3' sequences are bound with equilibrium association constants of  $3 \times 10^7 \text{ M}^{-1}$  and  $2 \times 10^7 \text{ M}^{-1}$ , respectively, while the 5'-AGGGCTA-3' and 5'-ACGGCTA-3' sites are not bound appreciably (>300-fold reduced affinity). The preference C over G observed for the turn is not observed for the tail. This most likely is due to the increased steric bulk of the extended  $\beta$ -alanine residue which can potentially adopt many conformations relative to the compact folded  $\gamma$ -turn residue.



**Figure 4.19.** Structures of ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPy-β-Dp (**16**) and ImImPy-(R)<sup>H<sub>2</sub>N</sup>γ-ImPyPy-PrOH (**17**) prepared by solid phase synthesis.

**Table 4.6.** Equilibrium Association Constants (M<sup>-1</sup>)<sup>a,b</sup>.

Binding Site	ImImPy-(R) <sup>H<sub>2</sub>N</sup> γ-ImPyPy-β-Dp	Specificity <sup>c</sup>
5'- <u>ATGGCTA</u> -3'	2.3 x 10 <sup>9</sup> (0.8)	23
5'- <u>TTGGCTA</u> -3'	1.1 x 10 <sup>9</sup> (0.4)	11
5'- <u>CTGGCTA</u> -3'	1.0 x 10 <sup>8</sup> (0.8)	1
5'- <u>GTGGCTA</u> -3'	1.0 x 10 <sup>8</sup> (0.3)	1
Binding Site	ImImPy-(R) <sup>H<sub>2</sub>N</sup> γ-ImPyPy-PrOH	Specificity <sup>c</sup>
5'- <u>ATGGCTA</u> -3'	1.1 x 10 <sup>9</sup> (0.3)	2
5'- <u>TTGGCTA</u> -3'	3.4 x 10 <sup>8</sup> (0.8)	0.5
5'- <u>CTGGCTA</u> -3'	2.7 x 10 <sup>8</sup> (0.4)	0.4
5'- <u>GTGGCTA</u> -3'	7.0 x 10 <sup>8</sup> (2.1)	1

<sup>a</sup>Values reported are from at least three DNase I footprint experiments, the standard deviation for each data set indicated in parentheses. <sup>b</sup>The assays were performed at 22°C at pH 7.0 in the presence of 10 mM tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. <sup>c</sup>Specificity is calculated as K<sub>a</sub>(5'-NTGGCTA-3') / K<sub>a</sub>(5'-GTGGCTA-3') for both compounds.

Specificity is also observed at the secondary tail position 5'-NTGGCTA-3'. Sequences with A•T and T•A base pairs at this position, 5'-ATGGCTA-3' and 5'-TTGGCTA-3', are bound with the highest affinity. The sequences containing G•C and C•G base pairs 5'-GTGGCTA-3' and 5'-CTGGCTA-3' are bound with 20-fold lower affinity compared to 5'-TTGGCTA-3'. In the first study of a hairpin polyamide with a C-terminal  $\beta$ -Dp group, it was observed that the hairpin polyamide ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp bound a 5'-TTGTTAG-3' sequence with 3-fold higher affinity than the corresponding hairpin ImPyPy- $\gamma$ -PyPyPy-Dp which has a C-terminal -Dp group. It was proposed based on this observed binding enhancement that the C-terminal tail could make a favorable interaction with the secondary “sixth” base pair of the binding site. Subsequently the hairpin ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp was found to bind a 5'-GTGTTAC-3' sequence with 10-fold reduced affinity relative to binding at 5'-TTGTTAG-3'. Crystal structure analysis of the homodimer of ImImPyPy- $\beta$ -Dp bound to a 5'-GTGGCCAC-3' sequence revealed that the tails which are placed opposite G•C were disordered and did not sit deeply within the minor groove.<sup>13</sup> No corresponding structure of a polyamide with a secondary A•T or T•A base pair is available; however, these previous observations combined with the study reported here indicate that there are favorable interactions of the polyamide tail with A•T rich flanking sequences.

In order to confirm that favorable interactions exist between the  $\beta$ -Dp tail and flanking A•T or T•A base pairs, the polyamide ImImPyPy-(*R*)<sup>H2N</sup> $\gamma$ -ImPyPy-PrOH (**17**) was prepared such that the  $\beta$ -dimethylaminopropylamide tail is replaced by a smaller propanolamide moiety. In order to maintain a positive charge on the resulting polyamide, (*R*)-2,4-diaminobutyric acid was substituted for the  $\gamma$ -turn.<sup>13</sup> A control polyamide which contains only the turn substitution, ImImPyPy-(*R*)<sup>H2N</sup> $\gamma$ -ImPyPy- $\beta$ -Dp (**16**), shows similar tail specificity to polyamide **1** with a 20-fold preference for A•T and T•A base pairs observed at the secondary tail position. The binding affinity of polyamide **16** is increased by approximately 50-fold at each of the 5'-NTGGCTA-3' sites consistent with the substitution of (*R*)<sup>H2N</sup> $\gamma$  for  $\gamma$ .<sup>28</sup> Interestingly, the presence of the chiral  $\alpha$ -amino group on the  $\gamma$ -turn was found to have little effect on the relative turn specificity.

In contrast to polyamides **15** and **16**, the polyamide, ImImPyPy-(*R*)<sup>H2N</sup> $\gamma$ -ImPyPy-PrOH (**17**), tolerates all four base pairs in the secondary tail position (5'-NTGGCTA-3') with no sequence specificity (Table 4.5). Furthermore, elimination of the Dp moiety reduces the binding affinity of

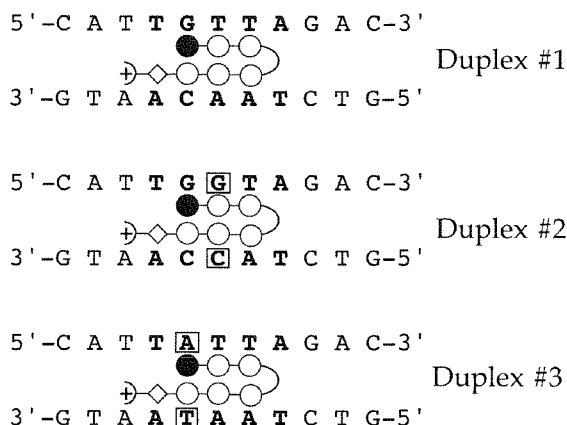
polyamide **17** by 3-fold relative to polyamide **16** for binding at the 5'-ATGGCTA-3' and 5'-TTGGCTA-3' sites. This effect is similar in magnitude to the previously observed difference for the polyamides ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp and ImPyPy- $\gamma$ -PyPyPy-Dp binding to a 5'-TTGTTAG-3' sequence,<sup>8</sup> and is consistent with the  $\beta$ -Dp tail of polyamides **15** and **16** making favorable contact with A•T rich flanking sequences.

Py-Im polyamides provide a versatile chemical method for targeting predetermined DNA sequences. The recent demonstration that hairpin polyamides are cell permeable and inhibit the transcription of specific genes increases the importance of understanding the sequence-specificity of the hairpin structure in detail. Both the  $\gamma$ -turn and  $\beta$ -tail amino acids show strong preferences for A•T and T•A base pairs at the turn and tail positions, respectively. Additionally, the  $\beta$ -Dp tail results in additional specificity at the secondary position. This binding effect can be regulated by elimination of the C-terminal amide bond. The recent demonstration that certain ring pairings of 3-hydroxypyrrole with pyrrole can discriminate A•T from T•A,<sup>26</sup> combined with the observation that polyamide turns and tails, like the Py/Py pair, are degenerate for A•T, T•A base pairs, provides an interesting challenge to engineer additional A•T specificity into these linker amino acids. The polyamide end-effects observed here explain the previously reported 10-fold difference in affinity for the hairpin polyamide ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp binding to 5'-TTGTTAG-3' and 5'-GTGTTAC-3' sequences providing a cautionary note that both sequence dependant DNA microstructure effects as well as polyamide end-effects will have to be considered for successful use of the polyamide pairing rules.



**Binding of a Hairpin Polyamide in the Minor Groove of DNA: Sequence-Specific Enthalpic Discrimination.** We have used a combination of calorimetric and spectroscopic techniques to characterize the binding of the hairpin polyamide, ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp, to three 11mer DNA duplexes, whose base sequences are presented in Figure 4.20. The central five base pair sequence of one of the three duplexes (duplex #1) is 5'-TGTTA-3', the designated match site as defined by the pairing rules and footprinting studies.<sup>1,29</sup> The other two duplexes contain single base pair changes to produce 5'-TGGTA-3' (duplex #2) and 5'-TATTA-3' (duplex #3), which we designate as mismatch sites. Our studies at 20°C reveal that ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp exhibits a 1-2 kcal/mol greater affinity for the 5'-TGTTA-3' target site than for either the 5'-TATTA-3' or the 5'-TGGTA-3' single base pair mismatch site, with this enhanced affinity being entirely enthalpic in origin.

The hairpin polyamide binds to and enhances the thermal stabilities of each DNA duplex in a manner that is sensitive to single base pair changes in the target sequence. UV melting experiments were conducted in the absence and presence of ligand to assess the impact, if any, of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp on the thermal stabilities of the three 11mer DNA duplexes studied here. As the [total ligand] to [duplex] ratio ( $r_{\text{Dup}}$ ) increases from 0 to 1.0, the thermal stabilities of all three host duplexes increase concomitantly. Higher polyamide to duplex ratios do not result in further increases in the  $T_m$  of either duplex #1 or duplex #2, while inducing only marginal increases in the  $T_m$  of duplex #3. This observation is suggestive of secondary binding to duplex #3 at high polyamide concentrations. In this work, we will focus exclusively on the one-to-one complex, which is the one observed in footprinting and NMR studies.<sup>18,27</sup> The polyamide-induced changes in duplex thermal stability noted above are consistent with ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp binding to each duplex, with a preference for the duplex versus single-stranded state.



**Figure 4.17.** Base sequences for the three 11mer DNA duplexes used in this study (denoted #1, #2, and #3). Schematic binding model of putative complexes between the hairpin polyamide and each duplex. The Im and Py rings are represented as shaded and unshaded spheres, respectively, while the  $\beta$ -alanine residue is represented as an unshaded diamond. The boxes indicate the base pairs in duplexes #2 and #3 that have been changed relative to the match duplex #1. The central five base pair binding sites of each duplex are presented in bold face.

ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-induced enhancement in duplex thermal stability follows the hierarchy:

$$\text{Duplex \#1} > \text{Duplex \#3} > \text{Duplex \#2}$$

At a  $r_{\text{Dup}}$  ratio of 1.0, hairpin polyamide binding increases the thermal stabilities of duplexes #1, #3, and #2 by approximately 11, 6, and 2°C, respectively. Thus, as measured by differences in  $\Delta T_m$ , the hairpin polyamide is able to distinguish between duplex targets that differ by only a single base pair.

In addition to the UV thermal denaturation studies described above, CD spectropolarimetry provides a second means for detecting and characterizing the DNA binding of the hairpin polyamide. Neither free ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp nor any of the ligand-free duplexes exhibit CD signals between 300 and 380 nm. However, substantial CD signals arise in this wavelength range upon addition of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp to a solution of any one of the

**Table 4.7.** Calorimetrically-derived binding enthalpies ( $\Delta H_b$ ) for the interactions of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp with the three 11mer DNA duplexes at 20°C\*

Duplex	$^\dagger \Delta H_b$ (kcal/mol)
#1 (5'-TGTTA-3')	$-6.7 \pm 0.6$
#2 (5'-TGGTA-3')	$-4.6 \pm 0.8$
#3 (5'-TATTA-3')	$-4.4 \pm 0.6$

\*Solution conditions are 10 mM sodium cacodylate (pH 6.9), 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

$^\dagger \Delta H_b$  values were determined at [total ligand] to [duplex] ratios ( $r_{Dup}$ ) of 1.0, with the indicated uncertainties corresponding to the sum of the standard deviations from three separate mixing experiments (DNA-ligand, ligand-buffer, and buffer-buffer) of at least 18 independent injections each.

three duplexes. These induced CD signals are indicative of interactions between ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp and each of the host DNA duplexes and can be used to detect and to monitor *CD-active* DNA binding mode(s).

The magnitudes of the induced CD signals differ in a manner that depends on the host duplex. These differences in CD signal suggest that ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp adopts different structural/electronic properties when bound to each duplex, a reasonable expectation given the differences in the binding sites. The magnitude of the induced CD signal follows a similar hierarchy to that defined above based on the  $\Delta T_m$  data; namely, duplex #1 > duplex #3 > duplex #2. This correlation also extends to the hierarchy of the ligand binding constants at 20°C ( $K_{20}$ ), which are presented in a later section and suggests that the duplex binding strength of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp is correlated with its induced chirality and its ability to thermally stabilize the host duplex. Isothermal, stopped-flow mixing calorimetry was used to measure the binding enthalpies ( $\Delta H_b$ ) for ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp complexation with the three 11mer DNA duplexes studied here. The resulting  $\Delta H_b$  values are listed in Table 4.7. Inspection of these data reveals the enthalpy for hairpin polyamide binding to duplex #1, which contains the primary 5'-TGTTA-3'

match site as defined by footprinting is 7 kcal/mol. By contrast, the enthalpies for hairpin polyamide binding to the 5'-TGGTA-3' and 5'-TATTA-3' mismatch sites are only -4.6 kcal/mol and -4.4 kcal/mol, respectively, values which are essentially indistinguishable. Thus, the enthalpy data are consistent with the observed hairpin polyamide binding preference for the 5'-TGTTA-3' site.

The hairpin polyamide binds to the 5'-TGTTA-3' match site with a greater affinity than it binds to the 5'-TGGTA-3' and 5'-TATTA-3' single mismatch sites. We used the  $\Delta T_m$  approach described below to assess, by a single method, the relative strength of polyamide binding to all three duplexes, since the magnitude of the binding to duplex #1 precluded a Scatchard analysis of the optical data. Significantly, both the  $\Delta T_m$  and Scatchard methods yield similar binding constants for duplex #2, thereby validating our use of the  $\Delta T_m$  method for the systems studied here. This validation is consistent with previous reports in which the  $\Delta T_m$  method was successfully used to determine ligand binding affinities for both oligomeric and polymeric host duplexes.<sup>30-33</sup>

Measured ligand-induced changes in the thermal stabilities of the three 11mer duplexes were used in conjunction with a binding site size,  $n_{app}$ , defined by both footprinting studies on a virtually identical ligand and NMR studies on a similar polyamide without the hairpin link to estimate apparent ligand-duplex association constants at  $T_m$  ( $K_{T_m}$ ) from the equation

$$\frac{1}{T_m^\circ} - \frac{1}{T_m} = \frac{R}{n_{app}(\Delta H_{W-C})} \ln[1 + (K_{T_m})af] \quad (1)$$

where  $T_m^\circ$  and  $T_m$  are the melting temperatures of the ligand-free and ligand-saturated duplexes, respectively;  $\Delta H_{W-C}$  is the enthalpy change for the melting of a Watson-Crick (W-C) base pair in the absence of bound ligand (values we determined independently for each of the three target duplexes using DSC); and  $af$  is the free ligand activity for a W-C transition, which we estimated as one-half the total ligand concentration. We then used the binding enthalpies ( $\Delta H_b$ ) listed in Table 4.7 to extrapolate these calculated binding constants at  $T_m$  to a common reference temperature of 20°C.

**Table 4.8.**  $\Delta T_m$ -derived Binding Affinities of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp for the Three 11mer DNA Duplexes at 20°C\*.

Duplex	$^{\dagger}T_m^{\circ}$ (°C)	$^{\dagger}T_m$ (°C)	$^{\ddagger}K_{20}$ (M <sup>-1</sup> )
#1 (5'-TGTTA-3')	42.7 $\pm$ 0.3	53.7 $\pm$ 0.7	7.3 $\pm$ 1.3 $\times 10^6$
#2 (5'-TGGTA-3')	46.1 $\pm$ 0.3	48.2 $\pm$ 0.5	1.6 $\pm$ 0.7 $\times 10^5$
#3 (5'-TATTA-3')	38.1 $\pm$ 0.3	43.8 $\pm$ 0.5	8.6 $\pm$ 0.9 $\times 10^5$

\*Solution conditions are as described in the footnote to Table 4.7.

$^{\dagger}T_m$  values were derived from UV melting profiles at 5  $\mu$ M duplex (D) in the absence ( $T_m^{\circ}$ ) and presence of ligand (L) at 1L:1D stoichiometric ratios. Each  $T_m$  value is an average derived from two independent experiments, with the indicated errors corresponding to the average deviation from the mean.

$^{\ddagger}$ Binding constants at 20°C ( $K_{20}$ ) were determined using equation (1), a footprinting-derived apparent binding site size ( $n_{app}$ ) of 5 bp/ligand, the appropriate values of  $\Delta H_b$  listed in Table 4.7, and the following calorimetrically-determined duplex-to-single strand transition enthalpies ( $\Delta H_{W-C}$ ) for the three host duplexes: 77.1 kcal/mol for duplex #1, 73.5 kcal/mol for duplex #2, and 60.1 kcal/mol for duplex #3. The uncertainties reflect the maximum errors in  $K_{20}$  that result from the corresponding uncertainties noted above in  $T_m$  and  $T_m^{\circ}$ , as propagated through equation (1).

The resulting  $K_{20}$  values are listed in Table 4.8. Inspection of these data reveals that the apparent binding affinities of the hairpin polyamide follows the hierarchy: Duplex #1 (5'-TGTTA-3') > Duplex #3 (5'-TATTA-3') > Duplex #2 (5'-TGGTA-3'). Note the agreement between this hierarchy and that noted above for binding-induced enhancement in duplex thermal stability. Thus, given the binding enthalpies listed in Table 4.7, the relative extent to which ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp thermally stabilizes the target duplex is correlated with its relative binding affinity.

The approximately 9- to 47-fold relative higher affinity of the hairpin polyamide for the 5'-TGTTA-3' site observed here is in agreement with the footprinting results of Mrksich et al.<sup>6</sup> While the absolute binding affinity is roughly an order of magnitude lower than that determined by Mrksich et al. is not surprising given the short length (11 bp) of our DNA target relative to the

**Table 4.9.** Thermodynamic parameters for the binding of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp to the three 11mer DNA duplexes\*.

Duplex	$^{\dagger}\Delta H_b$ (kcal/mol)	$^{\S}T\Delta S_b$ (kcal/mol)	$^{\ddagger}\Delta G_{b-20}$ (kcal/mol)	$^{\dagger}K_{20}$ (M $^{-1}$ )
#1 (5'-TGTTA-3')	$-6.7 \pm 0.6$	$+2.5 \pm 0.4$	$-9.2 \pm 0.1$	$7.3 \pm 1.3 \times 10^6$
#2 (5'-TGGTA-3')	$-4.6 \pm 0.8$	$+2.4 \pm 0.4$	$-7.0 \pm 0.2$	$1.6 \pm 0.7 \times 10^5$
#3 (5'-TATTA-3')	$-4.4 \pm 0.6$	$+3.6 \pm 0.5$	$-8.0 \pm 0.1$	$8.6 \pm 0.9 \times 10^5$

\*Solution conditions are as described in the footnote to Table 4.7.

$^{\dagger}$ The indicated errors in  $\Delta H_b$  and  $K_{20}$  are as described in the footnotes to Tables 4.7 and 4.8, respectively.

$^{\ddagger}\Delta G_{b-20}$  is the binding free energy at 20°C, as determined using equation (2) in the text and the corresponding value of  $K_{20}$ . The indicated uncertainties reflect the errors in  $\Delta G_{b-20}$  that result from the corresponding uncertainties noted above in  $K_{20}$ , as propagated through equation (2).

$^{\S}\Delta S_b$  is the binding entropy, as determined using equation (3) in the text and the corresponding values of  $\Delta H_b$  and  $\Delta G_{b-20}$ . The uncertainties reflect the maximum possible errors in  $\Delta S_b$  that result from the corresponding uncertainties noted above in  $\Delta H_b$  and  $\Delta G_{b-20}$ , as propagated through equation (3).

135 bp DNA fragment used in the footprinting studies and the large differences (1,000- to 10,000-fold) in the DNA concentrations used in footprinting relative to our optical/calorimetric studies. The significant feature is that both the biophysical and the footprinting studies independently reveal that the hairpin polyamide binds preferentially to the 5'-TGTTA-3' site.

The Preferential Binding of the Hairpin Polyamide to the 5'-TGTTA-3' Match Site Is Enthalpic in Origin. Armed with the binding constants listed in Table 4.8, we calculated the corresponding binding free energies ( $\Delta G_b$ ) using the standard relationship

$$\Delta G_b = -RT \ln K \quad (2)$$

These binding free energies, coupled with our calorimetrically determined binding enthalpies, also allowed us to calculate the corresponding binding entropies ( $\Delta S_b$ ) using

$$\Delta S_b = \frac{\Delta H_b - \Delta G_b}{T} \quad (3)$$

These calculations enabled us to generate complete thermodynamic profiles for the binding of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp to each of the three 11mer duplexes studied here. These profiles are summarized in Table 4.9. Inspection of these data reveals that, at 20°C, the preferential binding

of the hairpin polyamide to duplex #1 (5'-TGTTA-3') is primarily ( $\approx 73\%$ ) enthalpically driven, while the reduced binding to either duplex #2 (5'-TGGTA-3') or duplex #3 (5'-TATTA-3') is due to less favorable binding enthalpies. In fact, relative to duplex #1, the reduced binding to duplex #3 occurs despite a favorable entropic contribution to binding [ $\Delta(T\Delta S)$ ], which is overcompensated by the enthalpy loss. This favorable entropic contribution may reflect binding-induced desolvation of the all-AT minor groove that is present only in duplex #3.<sup>34-37</sup> Thus, the preferential binding to the 5'-TGTTA-3' match site is *enthalpic in origin*.

Single base pair changes in the high-affinity 5'-TGTTA-3' site reduce the hairpin polyamide binding affinity by  $\approx 1$ -2 kcal/mol, with this reduction being entirely enthalpic in origin. The data listed in Table 4.9 allow us to evaluate the thermodynamic consequences on hairpin polyamide binding of single base pair changes in the high-affinity 5'-TGTTA-3' site (Table 4.10). Inspection of these data reveals that the single base pair changes which produce duplex #2 (T•A to G•C at position 3) and duplex #3 (G•C to A•T at position 2) result in losses of 2.1 and 2.3 kcal/mol of binding enthalpy, respectively, while resulting in entropy changes that depend on the nature of the alteration.

These losses in binding enthalpy, coupled with the corresponding entropy changes, translate into losses in binding free energy of 2.2 and 1.2 kcal/mol, respectively, which reflect an approximately 9- to 47-fold binding preference for duplex #1 relative to duplexes #2 and #3. Thus, the  $\approx 1$ -2 kcal/mol enhanced affinity exhibited by the hairpin polyamide for the 5'-TGTTA-3' site relative to two sites with single base pair changes is entirely enthalpic in origin.

The structural picture which emerges from the NMR studies is shown in Figure 4.1. Note that two classes of ligand-base hydrogen bonds are proposed that may be critical to the sequence specificity exhibited by the dimeric ligand complex. One class of hydrogen bonds involves an imidazole nitrogen and the 2-amino hydrogen of guanine, while a second class involves an amide hydrogen and either the N3 nitrogen of adenine, the O2 oxygen of thymine, or the O2 oxygen of cytosine. The differences we observe in the thermodynamics of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp binding to the three duplex targets studied here may be related to the ability of the ligand to form either or

**Table 4.10.** Thermodynamic Consequences of Single Base Pair Changes on the Binding of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp to the 5'-TGTTA-3' Match Site.

Duplex	* $\Delta\Delta H_b$ (kcal/mol)	* $\Delta(T\Delta S_b)$ (kcal/mol)	* $\Delta\Delta G_{b-20}$ (kcal/mol)
#1 (5'-TGTTA-3')	--	--	--
#2 (5'-TGGTA-3')	+2.1	-0.1	+2.2
#3 (5'-TATTA-3')	+2.3	+1.1	+1.2

\* $\Delta\Delta H_b$ ,  $\Delta(T\Delta S_b)$ , and  $\Delta\Delta G_{b-20}$  were determined by subtracting values of  $\Delta H_b$ ,  $T\Delta S_b$ , and  $\Delta G_{b-20}$  for duplex #1 from the corresponding  $\Delta H_b$ ,  $T\Delta S_b$ , and  $\Delta G_{b-20}$  values for either duplex #2 or duplex #3.

both of these types of hydrogen bonds with its duplex target, although potential differences in van der Waals contacts and solvation also contribute. According to the NMR-derived structure model (Figure 4.1), when the hairpin is complexed with the 5'-TGTTA-3' site of duplex #1 (Figure 4.17), it should be able to form one Im(N)-(H2)guanine hydrogen bond, two amide-(O2)thymine hydrogen bonds, one amide-(O2)cytosine hydrogen bond, and four amide-(N3)adenine hydrogen bonds. By contrast, when complexed with the 5'-TATTA-3' site of duplex #3, the polyamide would be unable to form the Im(N)-(H2)guanine hydrogen bond (see Figure 4.1), while when complexed with the 5'-TGGTA-3' site of duplex #2, the polyamide would be unable to form one of the amide-(O2)thymine hydrogen bonds noted above. The polyamide may not be able to compensate this latter loss by forming a hydrogen bond between the unbonded amide hydrogen and the N3 atom of guanine due to steric interference from the neighboring 2-amino group.

The inability of the hairpin polyamide to form the hydrogen bonds noted above when complexed with the 5'-TATTA-3' and 5'-TGGTA-3' sites of duplexes #3 and #2, respectively, may give rise to its reduced binding free energy relative to that which it exhibits when complexed with the 5'-TGTTA-3' site of duplex #1 (Table 4.10). The enthalpic origin we observe for this reduction in binding free energy may reflect the enthalpic cost of failing to form these hydrogen bonds. We recognize that other factors, such as hydrophobic interactions and differential hydration of the polyamide and the DNA duplexes in their free and complexed states, also may contribute to the observed thermodynamic differences.



## Experimental Section.

All calorimetry, and CD spectrometry were performed by Dani Pilch as described elsewhere.<sup>38</sup> All gel electrophoresis and footprinting assays were performed by Michelle Parks,<sup>12,29</sup> Sue Swalley,<sup>39</sup> and Sarah White<sup>24</sup> as described elsewhere.

**Materials.** 0.2 mmol/gram Boc-Glycine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc-G-Pam-Resin), 0.2 mmol/gram Boc- $\beta$ -alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc- $\beta$ -Pam-Resin), Dicyclohexylcarbodiimide (DCC), Hydroxybenzo-triazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU), Boc-Glycine and Boc- $\beta$ -alanine were purchased from Peptides International. *N,N*-diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF), *N*-methylpyrrolidone (NMP), DMSO/NMP, Acetic anhydride ( $\text{Ac}_2\text{O}$ ), and 0.0002M potassium cyanide/pyridine were purchased from Applied Biosystems. Boc- $\gamma$ -aminobutyric acid was from NOVA Biochem, dichloromethane (DCM) and triethylamine (TEA) was reagent grade from EM, thiophenol (PhSH), dimethylaminopropylamine from Aldrich, trifluoroacetic acid (TFA) from Halocarbon, phenol from Fisher, and ninhydrin from Pierce. All reagents were used without further purification.

Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. and are used for filtration of DCU. Disposable polypropylene filters are also used for washing resin for ninhydrin and picric acid tests, and for filtering pre-dissolved amino acids into synthesis cartridges. A shaker for manual solid phase synthesis was obtained from Milligen. A rotary evaporator was also modified for use as a shaker. Screw- cap glass peptide synthesis reaction vessels (5 ml and 20 ml) with a #2 sintered glass frit were made at the Caltech glass shop as described by Kent.<sup>40</sup>  $^1\text{H}$  NMR were recorded in  $\text{d}_6$ -DMSO on a GE 300 instrument operating at 300 MHz. Chemical shifts are reported in ppm relative to the solvent residual signal. UV spectra

were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed either on a HP 1090M analytical HPLC or a Beckman Gold system using a RAINEN C<sub>18</sub>, Microsorb MV, 5µm, 300 x 4.6 mm reversed phase column in 0.1% (wt/v) TFA with acetonitrile as eluent and a flow rate of 1.0 ml/min., gradient elution 1.25% acetonitrile/min. Preparatory HPLC was carried out on a Beckman HPLC using a Waters DeltaPak 25 x 100 mm, 100µm C<sub>18</sub> column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. 18MΩ water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2µm filtered. Reagent-grade chemicals were used unless otherwise stated.

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

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

**Calculation of Resin Substitution.** Resin substitution is calculated as  $L_{\text{new}}(\text{mmol/g}) = L_{\text{old}} / (1 + L_{\text{old}}(W_{\text{new}} - W_{\text{old}}) \times 10^{-3})$ ; L is the loading, and W is the molecular weight of the polyamide attached to the resin.<sup>41</sup>

**Typical Manual Synthesis Protocol: AcPyPyPy-γ-ImImPy-β-Dp.** Boc-β-Pam-resin (1.25 g, 0.25 mmol amine) was shaken in DMF for 30 min. and drained. The N-Boc group removed by washing with DCM for 2 x 30 s, followed by a 1 min shake in 80% TFA/DCM/0.5M PhSH, draining the reaction vessel and a brief 80% TFA/DCM/ 0.5 M PhSH wash, and 20 min. shaking in 80% TFA/DCM/0.5M PhSH solution. The resin was washed 1 min. with DCM and 30 s with DMF. A resin sample (8-10 mg) was taken for analysis. The resin was drained completely and Boc-pyrrole-OBt monomer (357 mg, 1 mmol) dissolved in 2 ml DMF added followed by DIEA (1 ml) and the resin shaken vigorously to make a slurry. The coupling was allowed to proceed for

45 min. A resin sample (8-10 mg) was taken after 40 min. to check reaction progress. The reaction vessel was washed with DMF for 30 s and dichloromethane for 1 min. to complete a single reaction cycle. Six additional cycles were performed adding, BocIm-OH (DCC/HOBt), BocIm-OH (DCC/HOBt), Boc- $\gamma$ -aminobutyric acid (HBTU/DIEA) and allowed to couple for 2 hours, BocPy-OBt, BocPy-OBt, and pyrrole-2-carboxylic acid (HBTU/DIEA). The resin was washed with DMF, DCM, MeOH, and ethyl ether and then dried *in vacuo*. PyPyPy- $\gamma$ -ImImPy- $\beta$ -Pam-Resin (180 mg, 29  $\mu$ mol) was weighed into a glass scintillation vial, 1.5 ml of *N,N*-dimethylaminopropylamine added, and the mixture heated at 55°C for 18 hours. The resin was removed by filtration through a disposable polypropylene filter and washed with 5 ml of water, the amine solution and the water washes combined, and the solution loaded on a C<sub>18</sub> preparatory HPLC column, the column allowed to wash for 4 min. in 0.1% TFA at 8 ml/min., the polyamide was then eluted in 100 min. as a well defined peak with a gradient of 0.25% acetonitrile per min. The polyamide was collected in four separate 8 ml fractions, the purity of the individual fractions verified by HPLC and <sup>1</sup>H NMR, to provide purified PyPyPy- $\gamma$ -ImImPy- $\beta$ -Dp **8**. (11.2 mg, 39% recovery), UV  $\lambda_{\text{max}}$ , 246 (31,100), 312 (51,200) HPLC, r.t. 23.6, <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.30 (s, 1 H), 10.26 (s, 1 H), 9.88 (s, 1 H), 9.80 (s, 1 H), 9.30 (s, 1 H), 9.2 (br s, 1 H), 8.01 (m, 3 H), 7.82 (br s 1 H), 7.54 (s, 1 H), 7.52 (s, 1 H), 7.20 (d, 1 H, *J* = 1.3 Hz), 7.18 (d, 1 H, *J* = 1.2 Hz), 7.15 (d, 1 H, *J* = 1.3 Hz), 7.01 (d, 1 H, *J* = 1.4 Hz), 6.96 (d, 1 H, *J* = 1.4 Hz), 6.92 (d, 1 H, *J* = 1.8 Hz), 6.89 (m, 2 H), 6.03 (t, 1 H, *J* = 2.4 Hz), 3.97 (s, 3 H), 3.96 (s, 3 H), 3.85 (s, 3 H), 3.82 (s, 3 H), 3.78 (m, 6 H), 3.37 (m, 2 H), 3.20 (q, 2 H, *J* = 5.7 Hz), 3.08 (q, 2 H *J* = 6.6 Hz), 2.94 (q, 2 H *J* = 5.3 Hz), 2.71 (d, 6 H *J* = 5.8 Hz), 2.32 (m, 4 H), 1.83 (m, 4 H); MALDI-TOF-MS, 978.7 (979.1 calc. for M+H).

**ImPyPy- $\gamma$ -PyPyPy-G-Dp (3).** (12 mg, 40% recovery). HPLC, r.t. 26.9, UV $\lambda_{\text{max}}$  (H<sub>2</sub>O), 246 (41,100), 306 (51,300) <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.50 (s, 1 H), 9.95 (s, 1 H), 9.93 (s, 1 H), 9.92 (s, 1 H), 9.86 (s, 1 H), 9.2 (br s, 1H), 8.29 (t, 1 H, *J* = 4.4 Hz), 8.07 (t, 1 H, *J* = 5.2 Hz), 8.03 (t, 1 H, *J* = 5.4 Hz), 7.39 (s, 1 H), 7.27, (d, 1 H, *J* = 1.6 Hz), 7.22 (m, 2 H), 7.16 (m, 2 H), 7.04 (m, 2 H), 6.92 (d, 1 H, *J* = 1.6 Hz), 6.89 (d, 1 H, *J* = 1.7 Hz), 6.86 (d, 1 H, *J* = 1.6 Hz), 3.97 (s, 3 H), 3.82

(m, 6 H), 3.81 (s, 3 H), 3.78 (m, 6 H), 3.70 (d, 2 H,  $J = 5.7$  Hz), 3.20 (q, 2 H,  $J = 5.7$ ), 3.11 (q, 2 H,  $J = 4.2$  Hz), 3.00 (q, 2 H,  $J = 4.4$  Hz), 2.76 (d, 6 H,  $J = 4.7$  Hz), 2.24 (t, 2 H,  $J = 4.8$  Hz), 1.77 (m, 4 H); MALDI-TOF-MS, 964.3 (964.1 calc. for M+H).

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

**ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp (5).** ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Pam-Resin was prepared by machine-assisted synthesis protocols. A sample of resin (1 g, 0.17 mmol) was placed in a 20 mL glass scintillation vial, 4 mL of dimethylaminopropylamine added, and the solution heated at 55°C for 18 h. Resin is removed by filtration through a disposable propylene filter and 16 mL of water added. The polyamide/amine mixture was purified directly by preparatory HPLC and the appropriate fractions lyophilized to yield a white powder. (103 mg, 61% recovery) HPLC r.t. 24.1, UV  $\lambda_{\max}$ (H<sub>2</sub>O) ( $\epsilon$ ), 234 nm (39,300), 304 nm (52,000); <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>); 10.47 (s, 1 H), 9.91 (s, 1 H), 9.89 (s, 1 H), 9.87 (s, 1 H), 9.84 (s, 1 H), 9.2 (br s, 1 H), 8.08 (m, 3 H), 7.38 (s, 1 H), 7.26 (d, 1 H,  $J = 1.0$  Hz), 7.20 (d, 1 H,  $J = 1.0$  Hz), 7.14 (m, 4 H), 7.04 (d, 1 H,  $J = 1.1$  Hz), 7.02 (d, 1 H,  $J = 1.1$  Hz), 6.89 (d, 1 H,  $J = 1.0$  Hz), 6.85 (m, 2 H), 3.97 (s, 3 H), 3.82 (m, 6 H), 3.81 (s, 3 H), 3.77 (m, 6 H), 3.34 (m, 2 H,  $J = 3.9$  Hz), 3.18 (m, 2 H,  $J = 5.5$  Hz), 3.06 (m, 2 H,  $J = 5.7$  Hz), 2.95 (m, 2 H,  $J = 4.9$  Hz), 2.71 (d, 6 H,  $J = 4.6$  Hz), 2.30 (m, 6 H), 1.75 (m, 4 H); MALDI-TOF MS 978.0 (978.1 calc. for M+H).

**AcImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp (6).** (9.2 mg 31% yield), UV  $\lambda_{\max}$  (H<sub>2</sub>O), 246 (42,800), 312 (50,400) HPLC, r.t. 24.9, <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.25 (s, 1 H), 10.01 (s, 1 H), 9.92 (m, 3 H), 9.86 (s, 1 H), 9.3 (br s, 1 H), 8.10 (m, 3 H), 7.42 (s, 1 H), 7.25 (d, 1 H,  $J = 1.5$  Hz), 7.20 (d, 1 H,  $J = 1.6$

Hz), 7.16 (m, 3 H), 7.12 (d, 1 H,  $J = 1.4$  Hz), 7.03 (d, 1 H  $J = 1.7$ ), 6.89 (d, 1 H,  $J = 1.6$  Hz), 6.86 (m, 2 H), 3.92 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.80 (s, 6H), 3.78 (s, 3 H), 3.35 (q, 2 H,  $J = 5.5$  Hz), 3.20 (q, 2 H,  $J = 3.8$  Hz), 3.08 (q, 2 H,  $J = 3.3$  Hz), 2.97 (q, 2 H,  $J = 3.8$  Hz), 2.75 (d, 6 H  $J = 4.8$  Hz), 2.34 (t, 2 H,  $J = 5.0$  Hz), 2.24 (t, 2 H,  $J = 4.4$  Hz), 2.00 (s, 3 H), 1.71 (m, 4 H); MALDI-TOF-MS, 1035.4 (1035.1 calc. for M+H).

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

**AcImImPy- $\gamma$ -PyPyPy- $\beta$ -Dp (9).** Polyamide was prepared by manual solid phase protocols and isolated as a white powder. (8 mg, 28% recovery), UV  $\lambda_{\max}$ , 246 (43,400), 312 (50,200) HPLC, r.t. 24.8,  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.35 (s, 1 H), 10.30 (s, 1 H), 9.97 (s, 1 H), 9.90 (s, 1 H), 9.82 (s, 1 H), 9.30 (s, 1 H), 9.2 (br s, 1H), 8.02 (m, 3 H), 7.52 (s, 1 H). 7.48 (s, 1 H), 7.21 (m, 2H), 7.16 (d, 1 H,  $J = 1.1$  Hz), 7.11 (d, 1 H,  $J = 1.2$  Hz), 7.04 (d, 1 H,  $J = 1.1$  Hz), 6.97 (d, 1 H,  $J = 1.3$  Hz), 6.92 (d, 1 H,  $J = 1.4$  Hz), 6.87 (d, 1 H,  $J = 1.2$  Hz), 3.99 (s, 3 H), 3.97 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.80 (s, 3 H), 3.79 (s, 3 H), 3.47 (q, 2 H,  $J = 4.7$  Hz), 3.30 (q, 2 H,  $J = 4.6$  Hz), 3.20 (q, 2 H,  $J = 5.0$  Hz), 3.05 (q, 2 H,  $J = 5.1$  Hz), 2.75 (d, 6 H,  $J = 4.1$  Hz), 2.27 (m, 4 H), 2.03 (s, 3 H), 1.74 (m, 4 H) MALDI-TOF-MS, 1036.4 (1036.1 calc. for M+H).

**AcPyPyPy- $\gamma$ -ImImPy- $\beta$ -Dp (10).** Polyamide was prepared by machine assisted solid phase methods protocols as a white powder. (14 mg, 48% recovery), UV  $\lambda_{\max}$ , 246 (44,400), 312 (52,300) HPLC, r.t. 23.8,  $^1\text{H}$  NMR (DMSO- $d_6$ ) 10.32 (s, 1 H), 10.28 (s, 1 H), 9.89 (m, 2 H), 9.82 (s, 1 H), 9.18 (s, 1 H), 9.10 (br s, 1 H), 8.03 (m, 3 H), 7.55 (s, 1 H), 7.52 (s, 1 H), 7.21 (d, 1 H,  $J =$

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

**ImImIm- $\gamma$ -PyPyPy- $\beta$ -Dp (11).** The product was synthesized by manual solid phase protocols and recovered as a white powder (2.4 mg, 4% recovery). UV  $\lambda_{\max}$  312 (48,500);  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.09 (s, 1 H), 9.89 (s, 1 H), 9.88 (s, 1 H), 9.83 (s, 1 H), 9.57 (s, 1 H), 9.19 (br s, 1 H), 8.36 (t, 1 H,  $J = 5.6$  Hz), 8.03 (m, 2 H), 7.64 (s, 1 H), 7.51 (s, 1 H), 7.45 (s, 1 H), 7.20 (d, 1 H,  $J = 1.0$  Hz), 7.15 (d, 1 H,  $J = 2.0$  Hz), 7.14 (s, 1 H), 7.08 (s, 1 H), 7.04 (s, 1 H), 6.87 (d, 2 H,  $J = 2.2$  Hz), 4.01 (s, 3 H), 3.99 (s, 3 H), 3.95 (s, 3 H), 3.82 (s, 3 H) 3.82 (s, 3 H), 3.79 (s, 3 H), 3.37 (q, 2 H,  $J = 5.8$  Hz), 3.26 (q, 2 H,  $J = 6.1$  Hz), 3.10 (q, 2 H,  $J = 6.1$  Hz), 2.99 (m, 2 H), 2.73 (d, 6 H,  $J = 4.8$  Hz), 2.34 (t, 2 H,  $J = 7.2$  Hz), 2.27 (t, 2 H,  $J = 7.3$  Hz), 1.79 (m, 4 H); MALDI-TOF-MS, 980.1 (980.1 calc. for M+H).

**ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp (12).** The product was synthesized by manual solid phase protocols and recovered as a white powder (7.6 mg, 11% recovery). UV  $\lambda_{\max}$ , 248 (42,000), 312 (48,500);  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.32 (s, 1 H), 10.13 (s, 1 H) 9.93 (s, 1 H), 9.90 (s, 1 H), 9.89 (s, 1 H), 9.84 (s, 1 H), 9.59 (s, 1 H), 9.23 (br s, 1 H), 8.09 (t, 1 H,  $J = 5.3$  Hz), 8.04 (m, 2 H), 7.65 (s, 1 H), 7.57 (s, 1 H), 7.46 (d, 1 H,  $J = 0.6$  Hz) 7.22 (m, 3 H), 7.16 (s, 2 H), 7.09 (d, 1 H,  $J = 0.8$  Hz), 7.06 (d, 2 H,  $J = 1.1$  Hz), 7.00 (d, 1 H,  $J = 1.7$ ), 6.88 (d, 1 H,  $J = 1.8$ ), 6.87 (d, 1 H,  $J = 1.8$  Hz), 4.02 (s, 3 H), 4.00 (s, 3 H), 3.99 (s, 3 H), 3.84 (s, 3 H), 3.83 (s, 3 H), 3.83 (s, 3 H), 3.80 (s, 3 H), 3.79 (s, 3 H), 3.37 (q, 2 H,  $J = 6.2$  Hz), 3.21 (q, 2 H,  $J = 6.4$  Hz), 3.10 (q, 2 H,  $J = 6.2$  Hz), 3.00 (m, 2 H), 2.73 (d, 6 H,  $J = 4.9$  Hz), 2.34 (t, 2 H,  $J = 7.2$  Hz), 2.28 (t, 2 H,  $J = 7.0$  Hz), 1.76 (m, 4 H); MALDI-TOF-MS, 1225.9 (1224.3 calc. for M+H).

**ImImIm- $\gamma$ -PyPyPy- $\beta$ -Dp-NH<sub>2</sub> (11-NH<sub>2</sub>).** A sample of machine-synthesized resin (350 mg, 0.17 mmol/gram) was placed in a 20 mL glass scintillation vial, and treated with 2 mL 3,3'-diamino-*N*-methyldipropylamine at 55°C for 18 hours. The resin was removed by filtration

through a disposable propylene filter, and the resulting solution dissolved with water to a total volume of 8 mL, and purified directly by preparatory reversed phase HPLC to provide ImImIm- $\gamma$ -PyPyPy- $\beta$ -Dp-NH<sub>2</sub> (28 mg, 41% recovery) as a white powder. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.14 (s, 1 H), 9.89 (s, 1 H), 9.88 (s, 1 H), 9.83 (s, 1 H), 9.6 (br s, 1 H), 9.59 (s, 1 H), 8.36 (t, 1 H, *J* = 5.5 Hz), 8.09 (t, 1 H, *J* = 5.0 Hz), 8.03 (t, 1 H, *J* = 5.0 Hz), 7.9 (br s, 3 H), 7.63 (s, 1 H), 7.50 (s, 1 H), 7.44 (s, 1 H), 7.19 (d, 1 H, *J* = 1.2 Hz), 7.13 (m, 2 H), 7.08 (d, 1 H, *J* = 1.3 Hz), 7.02 (d, 1 H, *J* = 1.2 Hz), 6.85 (m, 2 H), 4.01 (s, 3 H), 3.99 (s, 3 H), 3.97 (m, 6 H), 3.80 (s, 3 H), 3.77 (s, 3 H), 3.34 (q, 2 H, *J* = 5.3 Hz), 3.23 (q, 2 H, *J* = 6.0 Hz), 3.05 (m, 6 H), 2.83 (q, 2 H, *J* = 5.0 Hz), 2.70 (d, 3 H, *J* = 4.0 Hz), 2.32 (t, 2 H, *J* = 6.9 Hz), 2.25 (t, 2 H, *J* = 6.9 Hz), 1.90 (m, 2 H), 1.77 (m, 4 H). MALDI-TOF-MS, 1022.8 (1023.1 calc. for M+H).

**ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (12-NH<sub>2</sub>).** A sample of machine-synthesized resin (350 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial, and treated with 2 mL 3,3'-diamino-*N*-methyldipropylamine at 55°C for 18 hours. The resin was removed by filtration through a disposable propylene filter, and the resulting solution dissolved with water to a total volume of 8 mL, and purified directly by preparatory reversed phase HPLC to provide ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (31 mg, 40% recovery) as a white powder. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.37 (s, 1 H), 10.16 (s, 1 H), 9.95 (s, 1 H), 9.93 (s, 1 H), 9.91 (s, 1 H), 9.86 (s, 1 H), 9.49 (br s, 1 H), 9.47 (s, 1 H), 8.12 (m, 3 H), 8.0 (br s, 3 H), 7.65 (s, 1 H), 7.57 (s, 1 H), 7.46 (s, 1 H), 7.20 (m, 3 H), 7.16 (m, 2 H), 7.09 (d, 1 H, *J* = 1.5 Hz), 7.05 (m, 2 H), 7.00 (d, 1 H, *J* = 1.6 Hz), 6.88 (m, 2 H), 4.01 (s, 3 H), 3.99 (s, 3 H), 3.98 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.36 (q, 2 H, *J* = 5.3 Hz), 3.21- 3.05 (m, 8 H), 2.85 (q, 2 H, *J* = 4.9 Hz), 2.71 (d, 3 H, *J* = 4.4 Hz), 2.34 (t, 2 H, *J* = 5.9 Hz), 2.26 (t, 2 H, *J* = 5.9 Hz), 1.85 (quintet, *J* = 5.7 Hz), 1.72 (m, 4 H). MALDI-TOF-MS, 1267.1 (1267.4 calc. for M+H).

**ImImIm- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA (11-E).** EDTA-dianhydride (50 mg) was dissolved in 1 mL DMSO/NMP solution and 1 mL DIEA by heating at 55°C for 5 min. The dianhydride solution was added to ImImIm- $\gamma$ -PyPyPy- $\beta$ -Dp-NH<sub>2</sub> (11-NH<sub>2</sub>) (8.0 mg, 7  $\mu$ mol) dissolved in 750  $\mu$ L DMSO. The mixture was heated at 55 °C for 25 min., and treated with 3 mL 0.1M NaOH, and

heated at 55°C for 10 min. 0.1% TFA was added to adjust the total volume to 8 mL and the solution purified directly by preparatory HPLC chromatography to provide **9-E** as a white powder. (3.3 mg, 30% recovery)  $^1\text{H}$  NMR ( $\text{DMSO-}d_6$ )  $\delta$  10.14 (s, 1 H), 9.90 (s, 1 H), 9.89 (s, 1 H), 9.85 (s, 1 H), 9.58 (s, 1 H), 9.3 (br s, 1 H), 8.40 (m, 2 H), 8.02 (m, 2 H), 7.65 (s, 1 H), 7.51 (s, 1 H), 7.45 (s, 1 H), 7.20 (d, 1 H,  $J = 1.5$  Hz), 7.15 (m, 2 H), 7.08 (d, 1 H,  $J = 1.1$  Hz), 7.04 (d, 1 H,  $J = 1.5$  Hz), 6.86 (m, 2 H), 4.00 (s, 3 H), 3.98 (s, 3 H), 3.94 (s, 3 H), 3.87 (m, 4 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.78 (s, 3 H), 3.72 (m, 4 H), 3.4-3.0 (m, 16 H), 2.71 (d, 3 H,  $J = 4.2$  Hz), 2.33 (t, 2 H,  $J = 5.1$  Hz), 2.25 (t, 2 H,  $J = 5.9$  Hz), 1.75 (m, 6 H). MALDI-TOF-MS, 1298.4 (1298.3 calc. for  $\text{M}+\text{H}$ ).

**ImImImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp-EDTA (12-E).** Compound **12-E** was prepared as described for compound **11-E**. (3.8 mg, 40%).  $^1\text{H}$  NMR ( $\text{DMSO-}d_6$ )  $\delta$  10.34 (s, 1 H), 10.11 (s, 1 H), 9.92 (s, 1 H), 9.90 (s, 1 H), 9.89 (s, 1 H), 9.84 (s, 1 H), 9.57 (s, 1 H), 8.42 (m, 1 H), 8.03 (m, 3 H), 7.64 (s, 1 H), 7.56 (s, 1 H), 7.44 (s, 1 H), 7.20 (m, 3 H), 7.15 (m, 2 H), 7.07 (d, 1 H,  $J = 1.6$  Hz), 7.05 (m, 2 H), 6.99 (d, 1 H,  $J = 1.6$  Hz), 6.87 (m, 2 H), 4.00 (s, 3 H), 3.98 (s, 3 H), 3.97 (s, 3 H), 3.83 (m, 4 H), 3.82 (s, 6 H), 3.79 (s, 3 H), 3.78 (s, 6 H), 3.67 (m, 4 H), 3.4- 3.0 (m, 16 H), 2.71 (d, 3 H,  $J = 4.2$  Hz), 2.34 (t, 2 H,  $J = 5.4$  Hz), 2.25 (t, 2 H,  $J = 5.9$  Hz), 1.72 (m, 6 H). MALDI-TOF-MS, 1542.2 (1542.6 calc. for  $\text{M}+\text{H}$ ).



## References

1. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D.E.; Dervan, P. B. *Proc. Natl. Acad. Sci., U.S.A.* **1992**, *89*, 7586. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385.
2. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.
3. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S.T.; Sundaralingham, M. *Struct. Biol. Nature* **1994**, *1*, 169.
4. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (b) Geierstanger, B.H.; Jacobsen, J.P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E.; *Biochemistry* **1994**, *33*, 3055. (c) Geierstanger, B.H.; Dwyer, T.J.; Bathini, Y.; Lown, J.W.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.
5. (a) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646. (b) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 3325.
6. Mrksich, M.; Parks, M.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 7983.
7. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 9892. (b) Dwyer, T.J.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 9900. (c) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 3663. (c) Chen, Y.H.; Lown, J.W. *J. Am. Chem. Soc.* **1994**, *116*, 6995.
8. Mitchell, A.R.; Kent, S.B.H.; Engelhard, M.; Merrifield, R.B. *J. Org. Chem.* **1978**, *43*, 2845.
9. (a) Van Dyke, M.W.; Hertzberg, R.P.; and Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5470. (b) Van Dyke, M.W.; Dervan, P.B. *Science* **1984**, *225*, 1122.
10. (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8462. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.
11. Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Nature Struct.Biol.* **1996**, *3*, 321.
12. Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6153.
13. Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nature Struct. Biol.* **1998**, *5*, 104.
14. Wing, R.; Drew, T.; Takano, C.; Broka, S.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Nature* **1980**, *287*, 755.
15. Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804.
16. Steitz, T. A. *Quart. Rev. Biophys.* **1990**, *23*, 205.

17. Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 5911.
18. de Clairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909.
19. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953-6961.
20. Singh, S. B.; Wemmer, D. E.; Kollman, P.A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7673.
21. Zimmer, C.; Wahnert, U. *Prog. Biophys. molec. Biol.* **1986**, *47*, 31-112.
22. Pullman, B. *Advances in Drug Research* **1990**, *18*, 1.
23. Mecozzi, S.; West, A. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10566.
24. White, S.; Baird, E. E.; Dervan, P. B. *Chem & Biol* **1997**, *4*, 569.
25. Gottesfield, J. M.; Nealy, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1997**, *387*, 202.
26. White, S. E.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468.
27. White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756.
28. Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382.
29. Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147.
30. Snyder, J. G.; Hartman, N. G.; D'Estantoit, B. L.; Kennard, O.; Remeta, D. P.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 3968.
31. Marky, L. A.; Curry, J.; Breslauer, K. J. in *Molecular Basis of Cancer, Part B: Macromolecular Recognition, Chemotherapy, and Immunology*, **1989** ed. Rein, R. (Alan R. Liss, Inc.; New York), pp. 155-173.
32. Chou, W. Y.; Marky, L. A.; Zaunczkowski, D.; Breslauer, K. J. *J. Biomol. Struct. Dyn.* **1987**, *5*, 345.
33. Breslauer, K. J.; Freire, E; Straume, M. *Meth. Enzymol.* **1992**, *211*, 533.
34. Drew, H. R.; Dickerson, R. E. *J. Mol. Biol.* **1981**, *151*, 535.
35. Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376.
36. Marky, L. A.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4359.
37. Chalikian, T. V.; Plum, G. E.; Sarvazyan, A. P.; Breslauer, K. J. *Biochemistry* **1984**, *33*, 8629.

38. Pilch, D. S.; Poklar, N. A.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, 93, 8306.
39. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, 118, 8198.
40. Kent, S.B.H. *Ann. Rev. Biochem.* **1988**, 57, 957.
41. Barlos, K.; Chatzi, O.; Gatos, D.; Stravropoulos, G. *Int. J. Peptide Protein Res.* **1991**, 37, 513.

## CHAPTER 5

### Stereochemical Control of the DNA-Binding Affinity, Sequence-Specificity, and Orientation Preference of Hairpin Polyamides in the Minor Groove

**Abstract:** Three-ring Py-Im polyamides covalently coupled by  $\gamma$ -aminobutyric acid ( $\gamma$ ) form six-ring hairpins that recognize five base pair sequences in the minor groove of DNA. In order to determine whether there is an orientation preference of Py-Im polyamide dimers with respect to the 5'-3' direction of the DNA helix backbone, association constants ( $K_a$ ) were determined for a series of six-ring hairpin polyamides which differ with respect to substitution at the N and C termini. Affinity cleaving experiments using the hairpin polyamides ImPyPy- $\gamma$ -PyPyPy with an EDTA•Fe(II) moiety at the C-terminus reveal a single binding orientation at each formal match site, 5'-(A,T)G(A,T)<sub>3</sub>-3' and 5'-(A,T)C(A,T)<sub>3</sub>-3'. A positive charge at the C-terminus and no substitution at the N-terminus imidazole affords the maximum binding orientation preference, with the N-terminal end of each three-ring subunit located toward the 5' side of the target DNA strand. Removal of the positive charge, rearrangement of the positive charge to the N-terminus or substitution at the N-terminal imidazole decreases the orientation preference. These results suggest that second generation design principles superimposed on the simple pairing rules can further optimize the sequence-specificity of Py-Im polyamides for double helical DNA.

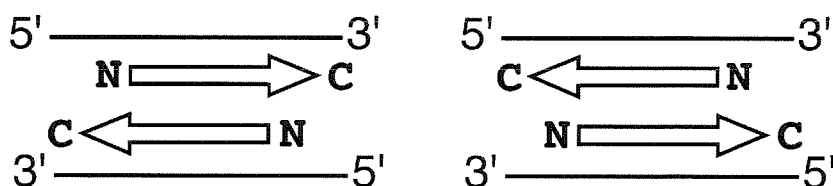
Selective chiral substitution of the " $\gamma$ -turn" is found to enhance the properties of polyamide-hairpins with regard to DNA affinity and sequence-specificity. Polyamides of core sequence composition ImPyPy- $\gamma$ -PyPyPy which differ by selective stereochemical substitution of the prochiral  $\alpha$ -position in the  $\gamma$ -turn were prepared. The DNA binding properties of two chiral polyamides were analyzed by footprinting and affinity cleavage on a DNA fragment

containing 5'-TGTTA-3', 5'-ACATT-3', and 5'-TGTC A-3' sites. Quantitative footprint titrations demonstrate that replacement of  $\gamma$ -aminobutyric acid by (R)-2,4-diaminobutyric acid enhances DNA binding affinity for the 5'-TGTTA-3' match site 13-fold. The enhanced affinity is achieved without a compromise in sequence selectivity, which in fact increases and is found to be 100-fold higher relative to binding at a single base pair mismatch sequence, 5'-TGTC A-3'. An (S)-2,4-diaminobutyric acid linked hairpin binds with 170-fold reduced affinity relative to the (R)-2,4-diaminobutyric acid linked hairpin, and only 5-fold sequence specificity versus a 5'-ACATT-3' reversed orientation site. These effects are modulated by acetylation of the chiral amine substituents. This study identifies structural elements which should facilitate the design of new hairpin-polyamides with improved DNA binding affinity, sequence-specificity, and orientational selectivity.

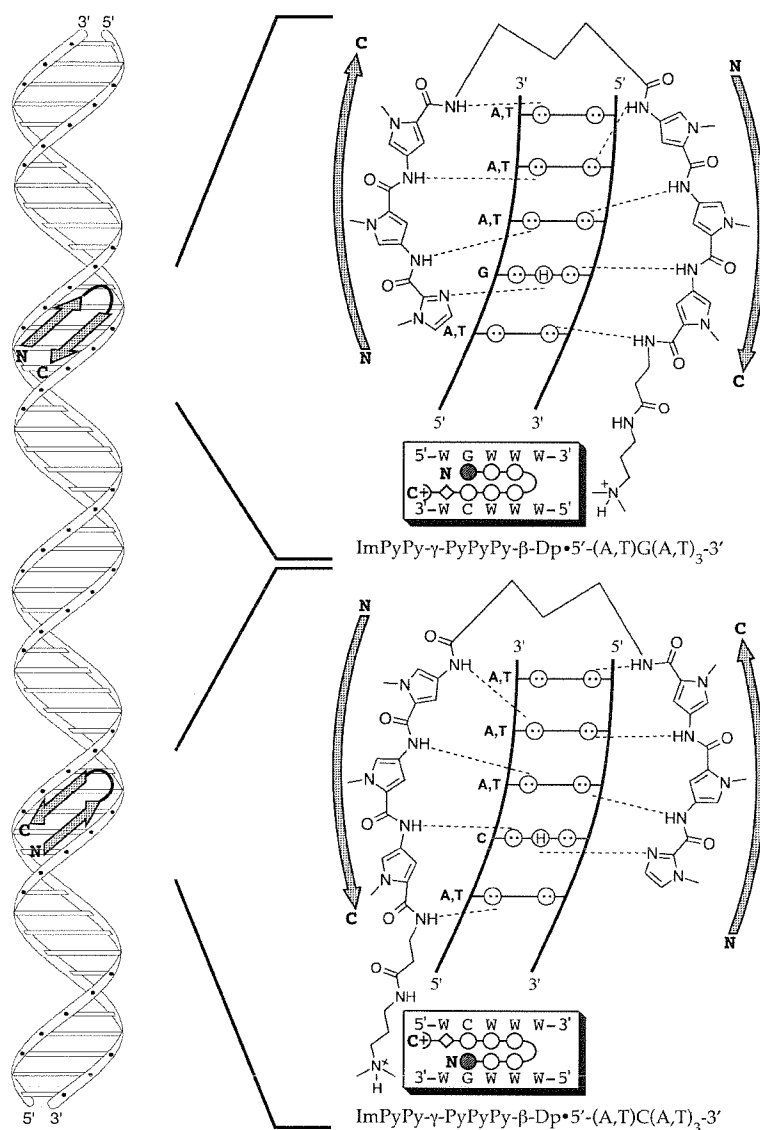
**Publications:** White, Baird & Dervan *J. Am. Chem. Soc.* **1997**, 119, 8756.  
Herman, Baird & Dervan *J. Am. Chem. Soc.* **1998**, 120, 1382.

**Orientation Preferences of Pyrrole-Imidazole Polyamides in the Minor Groove of DNA.** Polyamides containing pyrrole (Py) and imidazole (Im) amino acids bind cooperatively as antiparallel dimers in the minor groove of the DNA helix.<sup>1,2</sup> Sequence-specificity depends on the side-by-side pairings of *N*-methylpyrrole and *N*-methylimidazole amino acids.<sup>1</sup> A pairing of Im opposite Py targets a G•C base-pair, while Py opposite Im targets a C•G base-pair<sup>1</sup>. A Py/Py combination is degenerate and targets both T•A and A•T base-pairs.<sup>2</sup> Py-Im polyamides have been shown to be cell permeable and to inhibit the transcription of genes in cell culture.<sup>3</sup> This provides impetus to develop second generation polyamide design rules that provide for enhanced sequence-specificity and perhaps optimal biological regulation.

Although the polyamides bind DNA antiparallel to each other, the “pairing rules” do not distinguish whether there should be any energetic preference for alignment of each polyamide (N-C) with respect to the backbone (5'-3') of the DNA double helix (Figure 5.1). In a formal sense the homodimer (ImPyPy)<sub>2</sub> could bind 5'-WGWCW-3' or 5'-WCWGW-3' and still not violate the binary code. Remarkably, even in the first report on the binding specificity of the three-ring polyamide ImPyPy-Dp, there were qualitative data to suggest that there was indeed a binding preference 5'-WGWCW-3' > 5'-WCWGW-3'.<sup>1a</sup> This suggested that pyrrole-imidazole polyamide dimers align N-C with the 5'-3' direction of the DNA strand.



**Figure 5.1.** Antiparallel polyamide subunits are depicted as arrows. Arrowheads correspond to C-terminal end of the polyamide. (left) Polyamide binding with N-terminal end located towards 5'-side of the targeted DNA strand. (right) Binding with the C-terminal end of the polyamide located towards 5'-side of binding site.



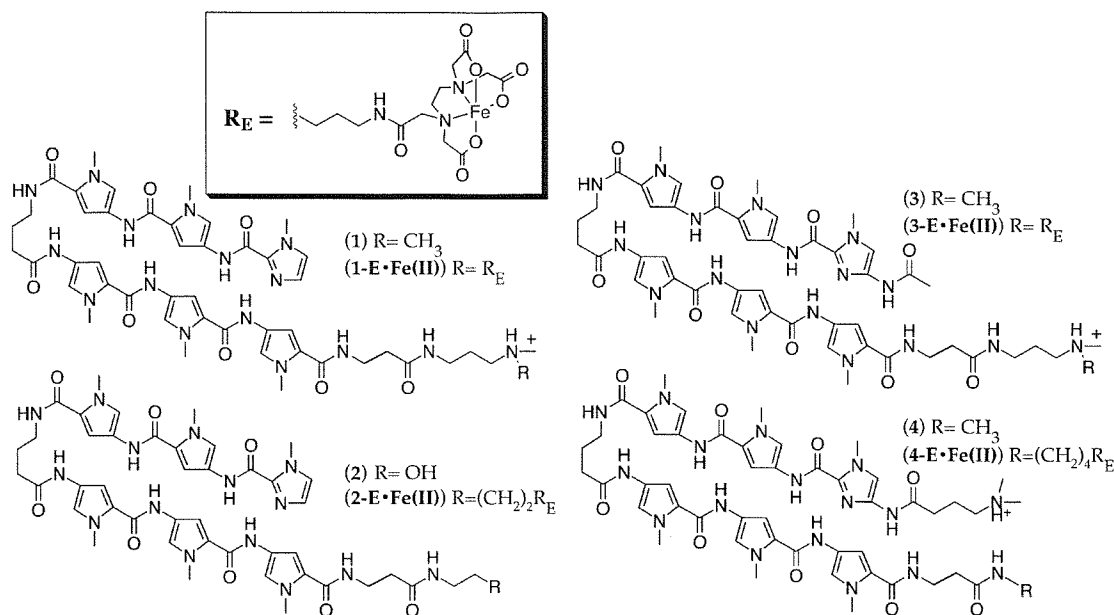
**Figure 5.2.** Binding models for ImPyPy-γ-PyPyPy-β-Dp in complex with 5'-WGWW-3' (Top) and 5'-WCWW-3' (Bottom) (W = A or T). Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. For schematic binding models, the imidazole and pyrrole rings are represented as shaded and unshaded spheres respectively, the curved line represents γ-aminobutyric acid, and the β-alanine residue is represented as an unshaded diamond. Shaded arrows represent the orientation of individual polyamide subunits. Arrowheads represent the polyamide C-terminus. (Left) A ribbon model depicting the hairpin structure bound in the minor groove of the DNA helix with either the N or C-terminus located at the 5' side of the binding site.

This orientation preference superimposed on the pairing rules confers added specificity by breaking a potential degeneracy for recognition. It would be useful to find out whether this preference is general and which aspects of the ligand design control the energetics of orientation preference. Therefore, we describe here a study to address the influence on orientation of (1) positive charge or lack of, (2) position of the positive charge at the N or C-terminus, and (3) substitution of the terminal imidazole.

Three-ring polyamide subunits covalently coupled by a  $\gamma$ -aminobutyric acid linker form 6-ring hairpin structures that bind to 5-bp target sequences with enhanced affinity and specificity relative to the unlinked polyamide pair.<sup>4,5</sup> In principle, a hairpin polyamide:DNA complex can form at two different DNA sequences depending on the N-C alignment of the polyamide with the walls of the minor groove of DNA (5'-3'). A six-ring hairpin polyamide of core sequence composition ImPyPy- $\gamma$ -PyPyPy which places the N-terminus of each three-ring polyamide subunit at the 5'-side of each recognized DNA strand would bind 5'-TGTTA-3'. Placement of the polyamide N-terminus at the 3' side of each recognized strand would result in targeting of a 5'-TCTTA-3' sequence (Figure 5.2).

Four six-ring hairpin polyamides, ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **1**, ImPyPy- $\gamma$ -PyPyPy- $\beta$ -EtOH **2**, Ac-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp **3**, and Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Me **4** were synthesized by solid phase methods (Figure 5.3).<sup>6</sup> The corresponding EDTA analogs ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA **1-E**, ImPyPy- $\gamma$ -PyPyPy- $\beta$ -C7-EDTA **2-E**, Ac-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA **3-E**, and Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -C7-E **4-E** were also constructed in order to confirm a single orientation of each hairpin:DNA complex. We report here the DNA-binding affinity, orientation, and sequence-selectivity of the four polyamides for the two match five base pair binding sites, 5'-TGTTA-3' and 5'-TCTTA-3'. Three separate techniques are used to characterize the DNA-binding properties of the polyamides: affinity cleaving<sup>7</sup> reveals the binding orientation and stoichiometry of each hairpin:DNA complex. Binding site size is accurately determined by MPE•Fe(II) footprinting,<sup>8</sup> while quantitative



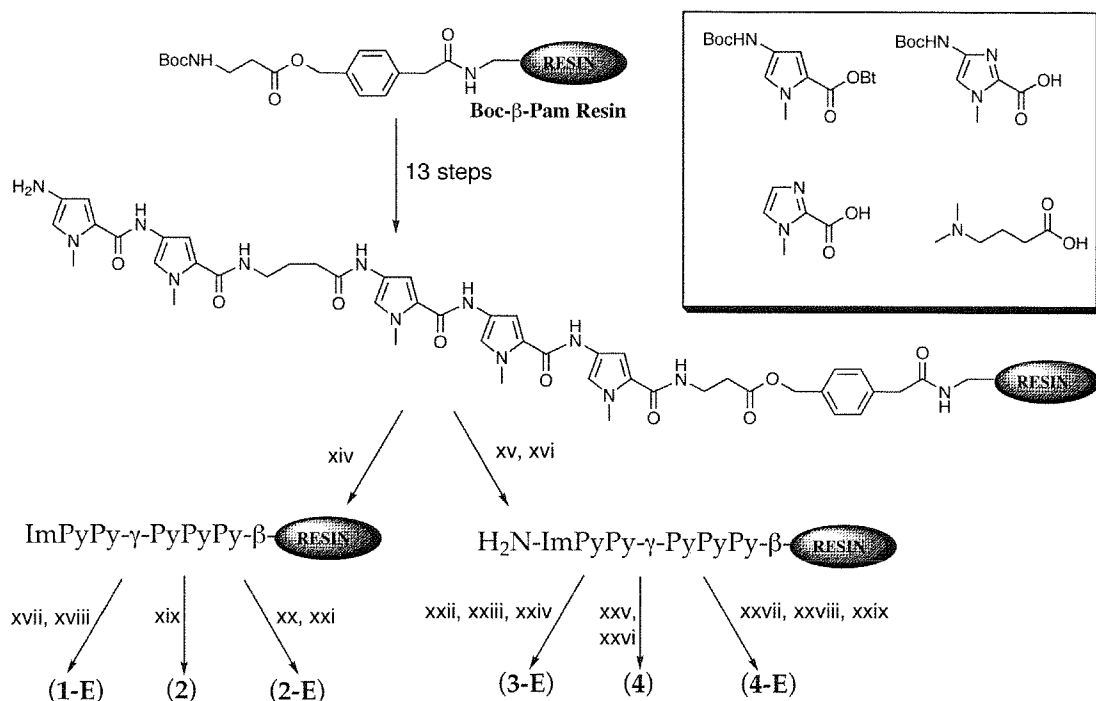


**Figure 5.3.** Structure of the hairpin polyamides **1-4** and the corresponding EDTA-modified derivatives; ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp (**1**), ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA•Fe(II) (**1-E**), ImPyPy- $\gamma$ -PyPyPy- $\beta$ -EtOH (**2**), ImPyPy- $\gamma$ -PyPyPy- $\beta$ -C7-EDTA (**3-E**), AcImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp (**3**), AcImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA•Fe(II) (**3-E**), Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Me (**4**), Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -C7-EDTA (**4-E**).

DNase I footprint titration<sup>9</sup> is more suitable for measurement of equilibrium association constants ( $K_a$ ) for the polyamide binding to designated sequences.

Polyamides were synthesized from Boc- $\beta$ -alanine-Pam-Resin (0.2 mmol/gram substitution) using stepwise solid phase methods (Figure 5.4).<sup>6</sup> A C-terminal  $\beta$ -alanine residue facilitates solid phase synthesis and increases hairpin-polyamide DNA-binding affinity and sequence specificity.<sup>5a</sup> The synthesis of polyamides **1** and **3** was described in Chapter 4.

ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Pam-Resin was cleaved by a single-step aminolysis reaction with neat ethanolamine (55°C, 16 h) to provide polyamide **2** after HPLC purification. Cleavage of Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Resin with a satd. solution of methylamine in DMF (55°C, 80 psi, 48 h) provided polyamide **4** after HPLC purification.

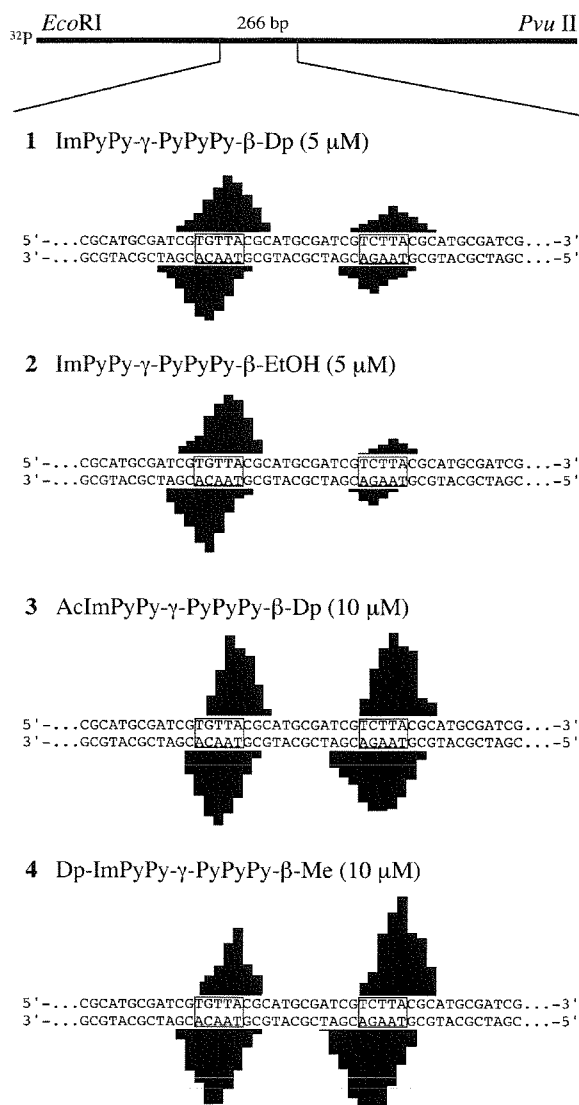


**Figure 5.4.** (Box) Pyrrole and imidazole monomers used for synthesis of the polyamides described here; Boc-Pyrrole-OBt ester, Boc-Imidazole-acid, and imidazole-2-carboxylic acid.<sup>1a</sup> Solid phase synthetic scheme for ImPyPy-γ-PyPyPy-β-Dp-EDTA (**1-E**), ImPyPy-γ-PyPyPy-β-EtOH (**2**), ImPyPy-γ-PyPyPy-C7-EDTA (**2-E**), AcImPyPy-γ-PyPyPy-β-Dp-EDTA (**3-E**), Dp-ImPyPy-γ-PyPyPy-β-Me (**4**), and Dp-ImPyPy-γ-PyPyPy-β-C7-EDTA (**4-E**). Synthesis is initiated from Boc-β-alanine-Pam-resin: (i) 80% TFA/DCM, 0.4M PhSH; (ii) BocPy-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4M PhSH; (iv) BocPy-OBt, DIEA, DMF; (v) 80% TFA/DCM, 0.4M PhSH; (vi) BocPy-OBt, DIEA, DMF; (vii) 80% TFA/DCM, 0.4M PhSH; (viii) Boc-γ-aminobutyric acid (HBTU, DIEA), DMF; (ix) 80% TFA/DCM, 0.4M PhSH; (x) BocPy-OBt, DIEA, DMF; (xi) 80% TFA/DCM, 0.4M PhSH; (xii) BocPy-OBt, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4M PhSH; (xiv) imidazole-2-carboxylic acid (HBTU/DIEA); (xv) BocIm-OBt (DCC/HOBt), DIEA, DMF; (xvi) 80% TFA/DCM, 0.4M PhSH; (xvii) 3,3'-diamino-*N*-methyldipropylamine, 55°C; (xviii) EDTA-dianhydride, DMSO/NMP, DIEA, 55°C; 0.1M NaOH; (xix) dimethylaminopropylamine, 55°C; (xx) 1,7-diaminoheptane, 55°C; (xxi) EDTA-dianhydride, DMSO/NMP, DIEA, 55°C; 0.1M NaOH (xxii) Acetic Anhydride, DMF, DIEA; (xxiii) 3,3'-diamino-*N*-methyldipropylamine, 55°C; (xxiv) EDTA-dianhydride, DMSO/NMP, DIEA, 55°C; 0.1M NaOH; (xxv) dimethylamino-γ, (HBTU/DIEA); (xxvi) MeNH<sub>2</sub> (satd.), DMF, 80 psi, 55°C (xxvii) dimethylamino-γ, (HBTU/DIEA); (xxviii) 1,7-diaminoheptane, 55°C; (xxix) EDTA-dianhydride, DMSO/NMP, DIEA, 55°C; 0.1M NaOH.

For the synthesis of analogs modified with EDTA, a sample of resin was cleaved with 3,3'-diamino-*N*-methyldipropylamine or 1,7-diaminoheptane (55°C, 18h) for polyamides **1-E** and **3-E** and polyamides **2-E** and **4-E** respectively. The amine modified polyamides were purified by reverse phase HPLC and then treated with an excess of the dianhydride of EDTA (DMSO/NMP, DIEA, 55°C, 30 min) and the remaining anhydride hydrolyzed (0.1 M NaOH, 55°C, 10 min). The EDTA modified polyamides **1-E**, **2-E**, **3-E** and **4-E** were then isolated by reverse phase HPLC. Polyamides were characterized by a combination of analytical HPLC, <sup>1</sup>H NMR spectroscopy, and MALDI-TOF mass spectroscopy.

Quantitative DNase I footprint titration experiments<sup>9</sup> (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5mM CaCl<sub>2</sub>, pH 7.0, 22°C) were performed to determine the equilibrium association constants  $K_a$  for recognition of the bound sites. The 5'-TGTTA-3' site is bound by the polyamides with decreasing affinity: ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp (**1**) > ImPyPy- $\gamma$ -PyPyPy- $\beta$ -EtOH (**2**) > AcImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp (**3**) > Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Me (**4**). The 5'-TCTTA-3' site is bound with decreasing affinity: Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Me (**4**) > AcImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp (**3**) > ImPyPy- $\gamma$ -PyPyPy- $\beta$ -EtOH (**2**) > ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp (**1**). Remarkably, the ratio of association constants for each site varies from 16 to 1-fold between the four polyamides indicates a sensitivity to substitution at the N and C terminus (Table 5.1).

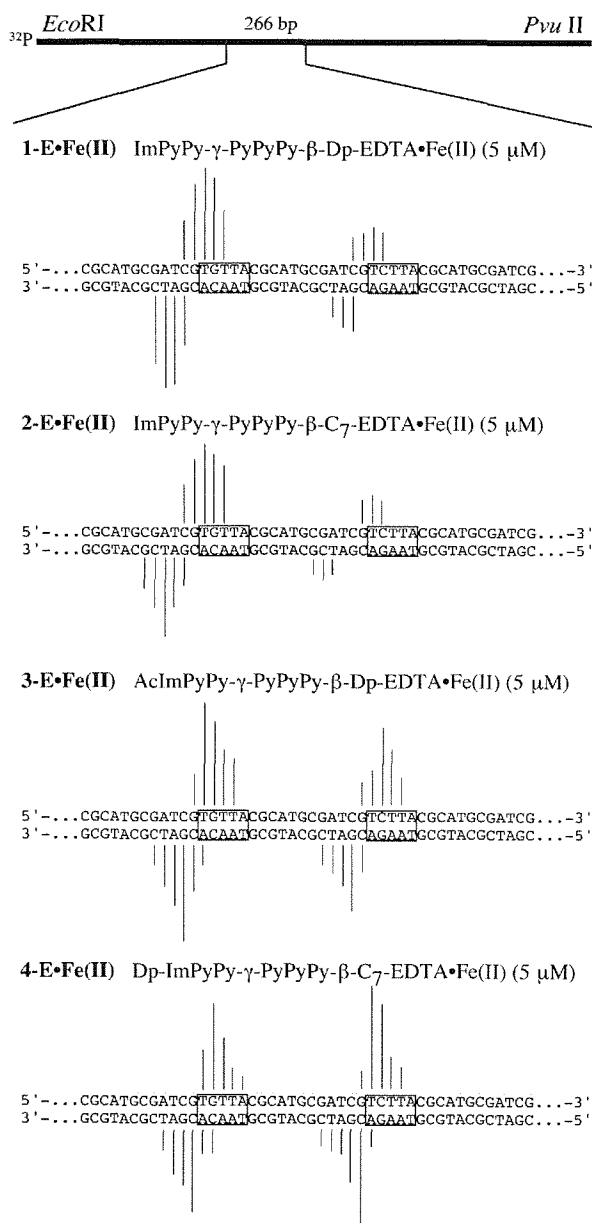
MPE•Fe(II) footprinting reveals that the four polyamides of core sequence composition ImPyPy- $\gamma$ -PyPyPy bind with high affinity to both match sites 5'-TGTTA-3' and 5'-TCTTA-3' (Figure 5.5). Affinity cleavage experiments using polyamides with Fe(II)•EDTA at the carboxy terminus confirm that the four polyamides bind each discrete site with a single orientation (Figure 5.6 and 5.7). Asymmetric 3'-shifted cleavage patterns are consistent with the location of the 1:1 polyamide:DNA complex in the minor groove. The observation of a single cleavage locus is consistent only with an oriented 1:1 complex and rules out any 2:1 overlapped or extended binding motifs.<sup>10</sup> A 1:1 oriented but extended motif would require at least an eight base pair binding site, which is inconsistent with MPE



**Figure 5.5.** (Top) Illustration of the 266 bp restriction fragment with the position of the sequence indicated. (Bottom) MPE•Fe(II) protection patterns of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp (1) or ImPyPy- $\gamma$ -PyPyPy- $\beta$ -EtOH (2) for 5  $\mu$ M concentration and AcImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp (3), or Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Me (4) for 10  $\mu$ M concentration. Bar heights are proportional to the relative protection from cleavage at each band.

footprinting data on both target sites. The hairpin structure is supported by direct NMR structure studies on a six-ring hairpin polyamide of sequence composition ImPyPy- $\gamma$ -PyPyPy binding a core five base pair 5'-TGTTA-3' site.<sup>5f</sup>

A single cleavage locus is observed proximal to the 5'-side of both the 5'-TGTTA-3' and 5'-TCTTA-3' binding sites, indicating that the carboxy terminus of the polyamide is located at the 5'-side of each site. Furthermore, the relative location of the observed cleavage maxima is unchanged for recognition of 5'-TGTTA-3' and 5'-TCTTA-3', indicating similar placement of the polyamide C-termini at both sites.



**Figure 5.6.** Results from affinity cleavage with ImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II) (**1-E**), ImPyPy-γ-PyPyPy-β-C7-EDTA (**2-E**), AcImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II) (**3-E**), and Dp-ImPyPy-γ-PyPyPy-β-C7-EDTA (**4-E**) at 5 μM concentration. (Top) Illustration of the 266 bp restriction fragment with the position of the sequence indicated. Only the two designated target sites are boxed. Arrow heights are proportional to the relative cleavage intensities at each base pair.



**Figure 5.7.** Affinity cleavage patterns and ball and stick models of the six-ring EDTA•Fe(II) analog **1-E•Fe(II)** bound to the match sites, 5'-TGTTA-3' and 5'-TCTTA-3'. Bar heights are proportional to the relative cleavage intensities at each base pair. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent the β-alanine residue. The boxed **Fe** denotes the EDTA•Fe(II) cleavage moiety. Cleavage patterns show a single binding orientation at each match site.

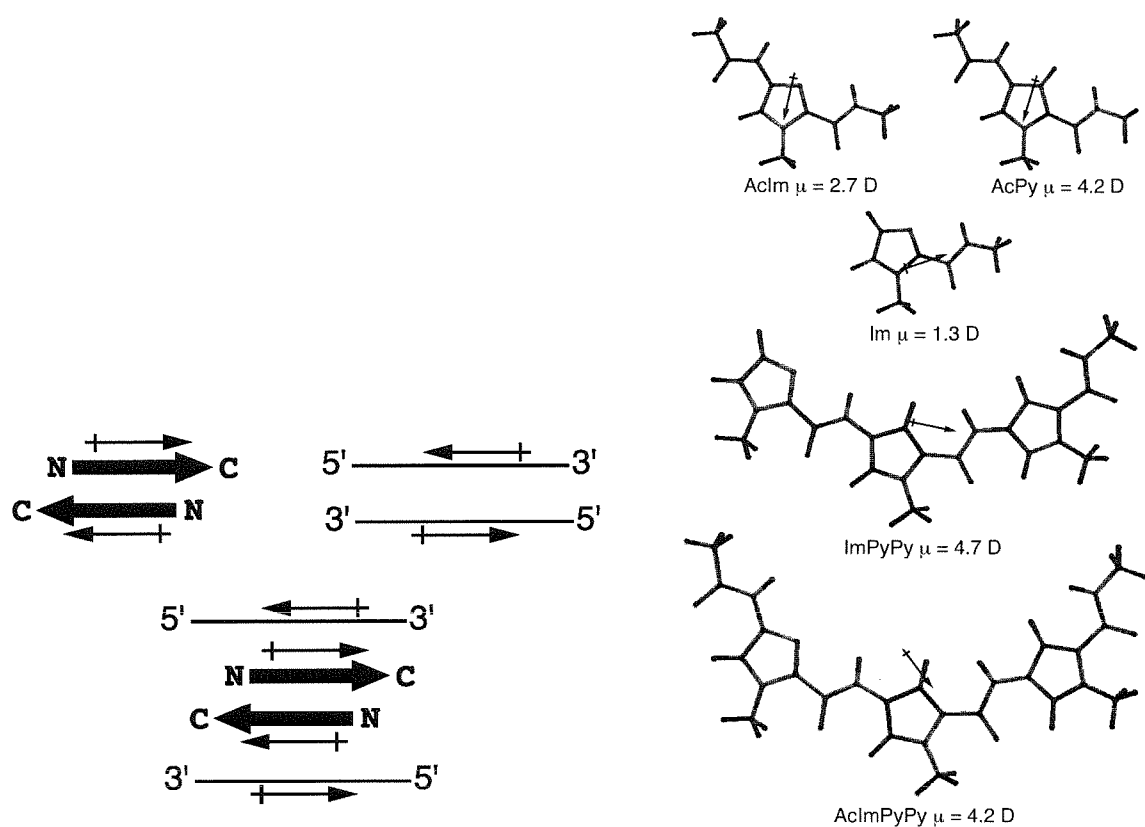
*These results indicate that all four polyamides may adapt two unique binding orientations.* Each binding orientation may represent a unique and distinguishable hairpin fold (Figure 5.2). Mismatched pairing of Im/Py opposite C•G cannot be ruled out for recognition of 5'-TCTTA-3' sites; however, mismatches of this class are energetically unfavorable and have been found to reduce DNA binding by at least 3 kcal/mol.

Four polyamides of core sequence composition ImPyPy-γ-PyPyPy-β, but varying at the N- and C- terminus bind both match sites 5'-TGTTA-3' and 5'-TCTTA-3' as a 1:1 complex consistent with the hairpin motif. However, the relative discrimination between the two sites varies by 16-fold. Among the four ligands, polyamide (1) binds the 5'-TGTTA-3' site with the highest affinity ( $K_a = 1.4 \times 10^7 \text{ M}^{-1}$ ) and orientation specificity (16-fold). Replacement of the charged dimethylaminopropylamide tail group with an uncharged ethoxyamide group as in polyamide (2) results in a negligible decrease in affinity ( $K_a = 1.1 \times 10^7 \text{ M}^{-1}$ ) but reduced specificity (9-fold). The decrease in orientational specificity indicates that the cationic tail group is necessary but not sufficient for optimal oriented polyamide binding.

**Table 5.1.** Equilibrium Association Constants ( $\text{M}^{-1}$ )<sup>a-c</sup>.

Polyamide (N→C)		5'-TGTTA -3'	5'-TCTTA-3'	Orientation <sup>c</sup>
ImPyPy-γ-PyPyPy-β-Dp	1	$1.4 \times 10^7$ (0.4)	$8.8 \times 10^5$ (1.0)	16
ImPyPy-γ-PyPyPy-β-EtOH	2	$1.1 \times 10^7$ (0.2)	$1.2 \times 10^6$ (1.0)	9
AcImPyPy-γ-PyPyPy-β-Dp	3	$7.6 \times 10^6$ (1.1)	$1.9 \times 10^6$ (1.2)	4
Dp-ImPyPy-γ-PyPyPy-β-Me	4	$7.2 \times 10^6$ (1.4)	$9.1 \times 10^6$ (0.1)	0.8

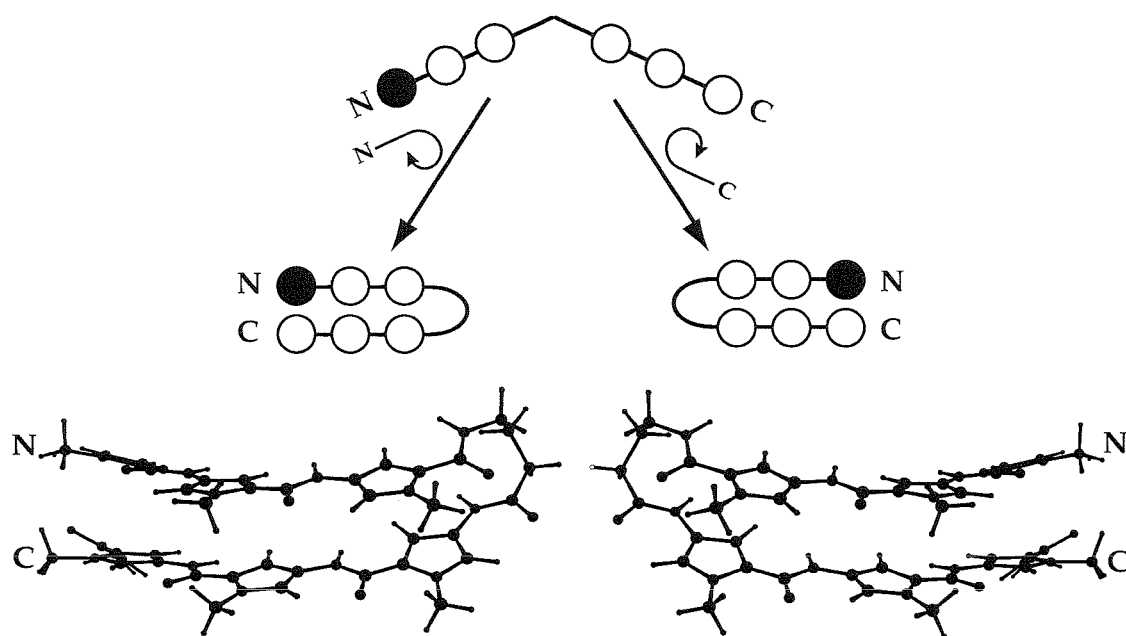
<sup>a</sup>Values reported are the mean values measured from at least three footprint titration experiments, with the standard deviation for each data set indicated in parentheses. <sup>b</sup>The assays were performed at 22°C in the presence of 10 mM tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. <sup>c</sup>Orientation preference calculated from  $K_a(5'-TGTTA-3')/K_a(5'-TCTTA-3')$ .



**Figure 5.8.** (left) Schematic model for how a polyamide dipole (N-C) and a helix dipole (3'-5') could interact to stabilize one particular polyamide-DNA-binding orientation. (right) Dipoles calculated for individual polyamide monomers and subunits. All C-termini are the methylamide derivatives, all N-termini are either unmodified, Im and ImPyPy or acetamide derivatives, AcIm, AcPy, and AcImPyPy. Arrows depict direction of dipole moments.

We have observed that N-terminal acetylation reduces both the binding affinity and the sequence preference for target sites.<sup>5</sup> The acetylated polyamide (**3**) binds the 5'-TGTTA-3' site with reduced affinity and orientational specificity compared to (**1**). Replacement of the N-terminal acetyl group with a charged dimethylaminopropyl group results in a 4-fold increase in affinity at the 5'-TCTTA-3' site but a complete loss of orientational specificity for polyamide (**4**). These results indicate that the N or C terminus position of charged substituents plays an important role in orientation preference.

Dipole-dipole interactions could potentially account in part for the observed DNA-binding orientation preference (Figure 5.8). Dipoles were calculated using MacSpartan v 1.0.<sup>11</sup> Calculated dipoles for the individual monomer units AcIm, AcPy, and Im, as well as the



**Figure 5.9.** Model for the two DNA-binding hairpin folds that are related by mirror symmetry.

three-ring subunits ImPyPy and AcImPyPy reveal that only monomers and subunits containing an unsubstituted Im residue could have a dipole antiparallel to the walls of the helix. Subunit dipoles could stabilize side-by-side antiparallel placement of polyamide subunits in the minor groove. In addition, dipole interactions between the DNA backbone and the polyamide could lead to 5' to 3' aligned with N to C oriented binding *if there exists a net dipole on each antiparallel strand of the DNA double helix*. (B. Norden, personal communication)

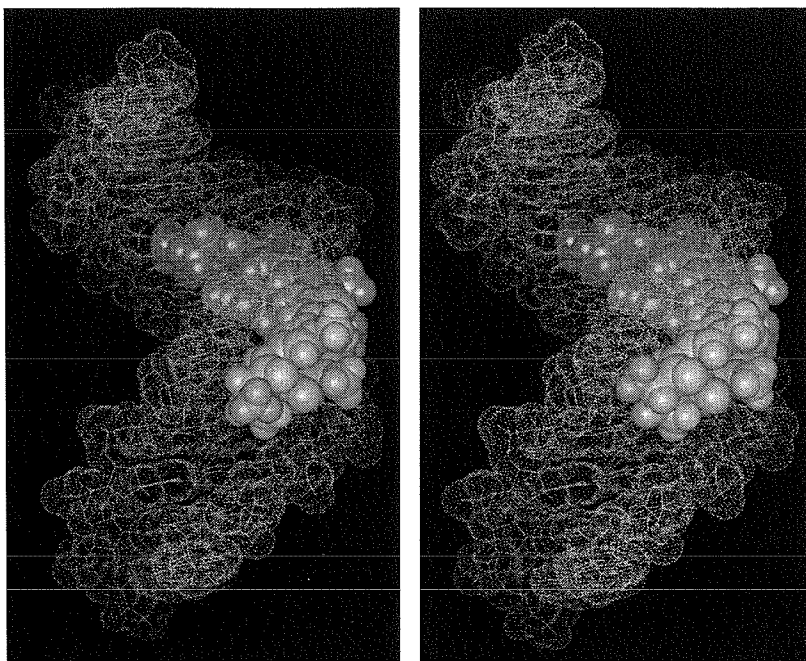
Py-Im polyamides are achiral molecules that form chiral complexes with DNA. In principle there exists two non-superimposable hairpin folds which are related by mirror plane symmetry (Figure 5.9). Fold 1 is responsible for the preferred 5' to 3' N to C orientation for Py-Im polyamide DNA recognition. Fold 2 corresponds to the 3' to 5' N to C recognition observed here. In the absence of DNA each fold should be energetically equivalent. Upon binding to DNA, adjacent Py and Im monomers twist to accommodate the right handed B-form DNA helix. Twisting of polyamide monomers results in an induced asymmetry upon



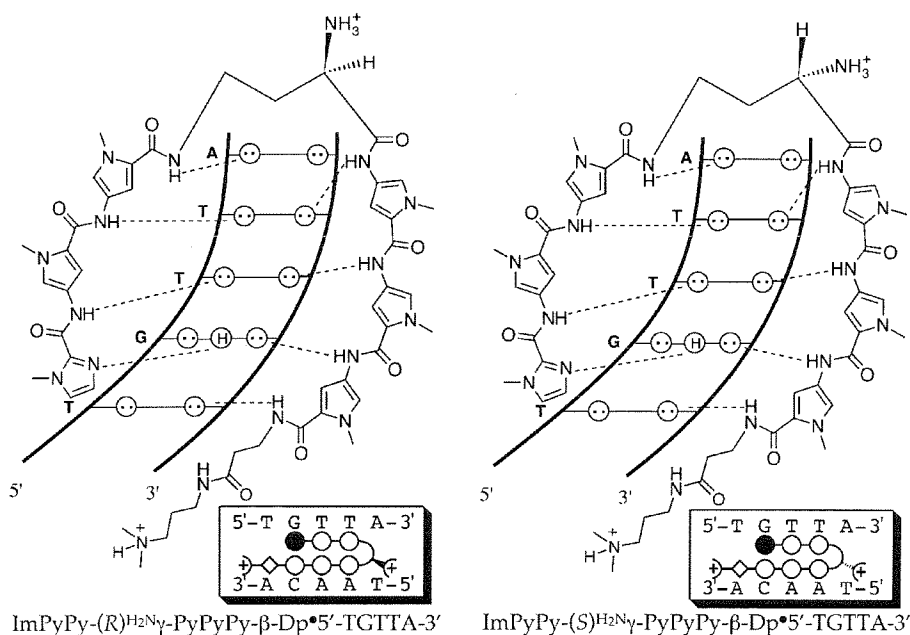
binding. A hairpin polyamide-DNA complex composed of an asymmetrically folded polyamide may be expected to display differential energetics for oriented binding.

The results reported here establish that a hairpin polyamide binds with at least a 16-fold orientation preference with the N-termini of the polyamide subunits located at the 5'-end of the targeted DNA strand. Although the cationic dimethylaminopropyl end group is not required for oriented hairpin polyamide recognition in the minor groove of DNA, the N or C terminus position of the charge appears important. Py-Im polyamide DNA-binding orientation preference defines a new design feature which must be considered for application of the pairing rules for DNA targeting. For optimal orientation and, hence sequence-specificity, a positive charge at the C-terminus and no substitution of an N-terminal imidazole is preferred.

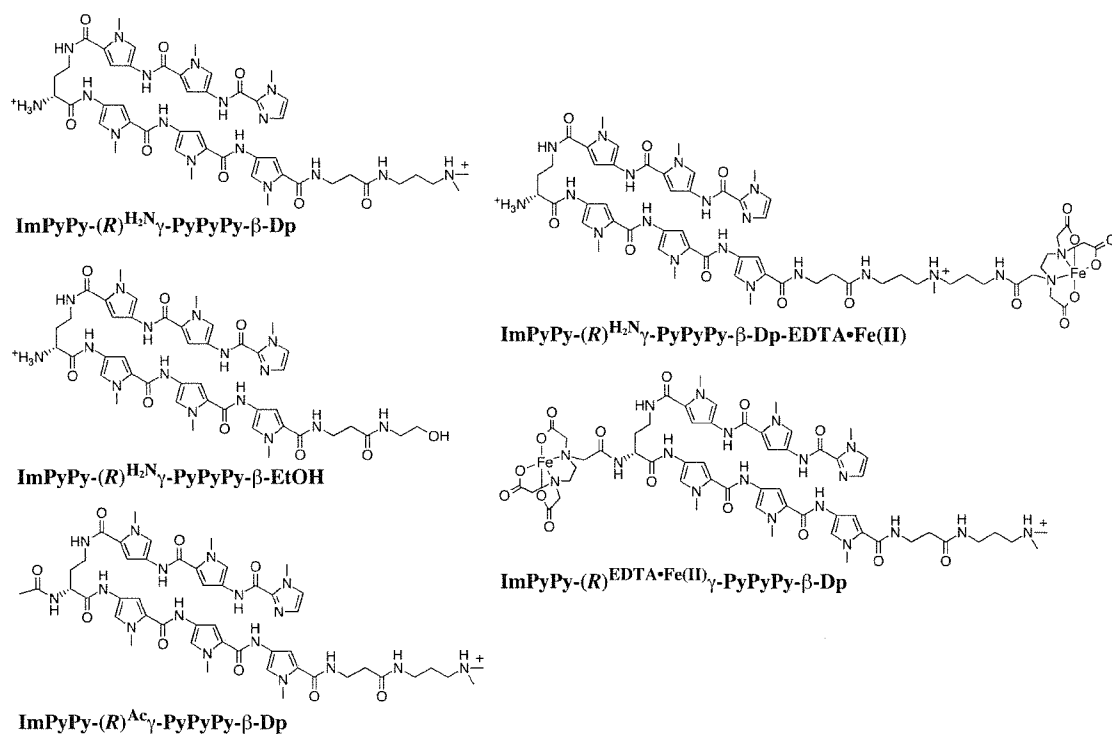
**Chiral Hairpin Polyamides.** The relationship between biological regulation by hairpin-polyamides, and the placement, frequency, and nature of charged moieties within the hairpin structure has yet to be determined. This provided impetus to elucidate the effects on hairpin polyamide DNA-binding by selectively placed substituents within the  $\gamma$ -turn. Analysis of the NMR structure of a hairpin polyamide of sequence composition ImPyPy- $\gamma$ -PyPyPy complexed with a 5'-TGTTA-3' target site, suggested that substitutions at the  $\alpha$ -position of the  $\gamma$ -aminobutyric acid residue could be accommodated within the hairpin-DNA complex.<sup>5f</sup> Modeling indicated that replacing the  $\alpha$ -H of  $\gamma$  with an amino group that confers an *R*-configuration at the  $\alpha$ -carbon could be accommodated within the floor and walls of the minor groove (Figure 5.10 and 5.11). Formally this is accomplished by substitution of the  $\gamma$ -residue with (*R*)-2,4,-diaminobutyric acid ( $((R)^{H_2N})\gamma$ ). (*S*)-2,4,-diaminobutyric acid ( $((S)^{H_2N})\gamma$ ) is predicted to clash with the walls of the minor groove of the DNA helix (Figure 5.10 and 5.11). We describe here the synthesis of a new class of chiral hairpin polyamides and their characterization with regards to DNA binding affinity and sequence specificity.



**Figure 5.10.** Computer generated models of (left) ImPyPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-PyPyPy-β-Dp (gray) and (right) ImPyPy-(*S*)<sup>H<sub>2</sub>N</sup>γ-PyPyPy-β-Dp (gray) bound in the minor groove of double stranded DNA. Models are derived from the NMR structure coordinates of ImPyPy-γ-PyPyPy-Dp•5'-TGTTA-3'<sup>91</sup>.



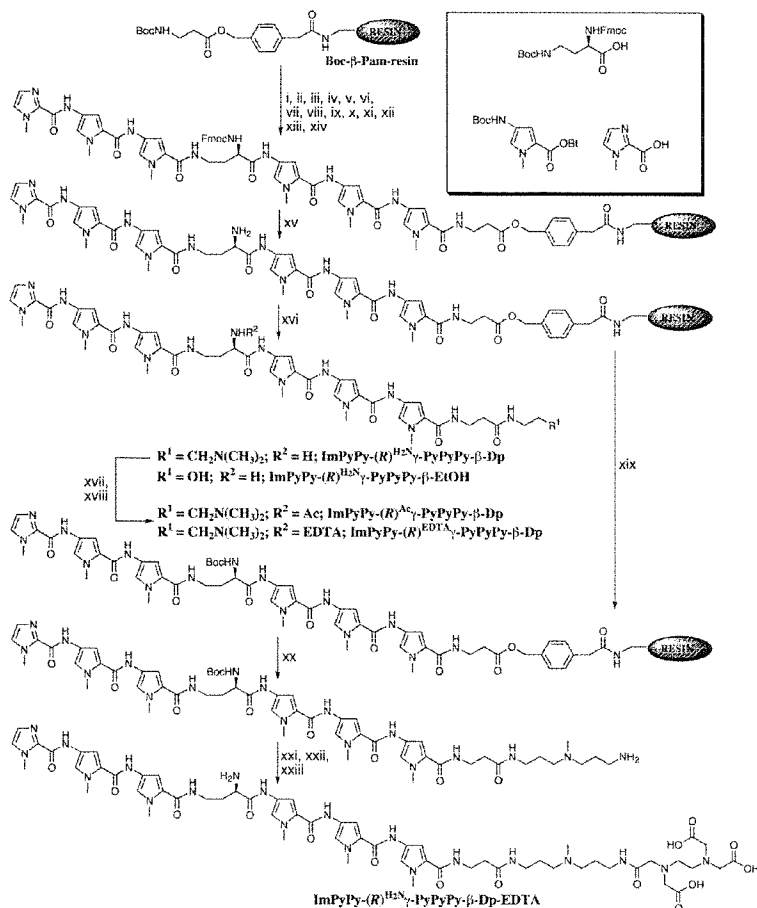
**Figure 5.11.** (Top) Hydrogen bonding model of the 1:1 polyamide:DNA complex formed between the hairpin polyamide (left) ImPyPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-PyPyPy-β-Dp (**5-R**) and (right) ImPyPy-(*S*)<sup>H<sub>2</sub>N</sup>γ-PyPyPy-β-Dp (**5-S**) with a 5'-TGTTA-3' site.



**Figure 5.12.** Structures of the six-ring hairpin polyamides ImPyPy-(R)<sup>H2N</sup>γ-PyPyPy-β-Dp **5-R**, ImPyPy-(R)<sup>H2N</sup>γ-PyPyPy-β-EtOH **6-R**, ImPyPy-(R)<sup>Ac</sup>γ-PyPyPy-β-Dp **7-R**, ImPyPy-(R)<sup>H2N</sup>γ-PyPyPy-β-Dp-EDTA•Fe(II) **8-R•Fe(II)**, and ImPyPy-(R)<sup>EDTA•Fe(II)</sup>γ-PyPyPy-β-Dp **9-R•Fe(II)**. Structures of the corresponding (S)<sup>H2N</sup>γ-linked hairpins are not shown.

Substitution of the prochiral γ-turn with either enantiomer of 2,4-diaminobutyric acid provides the dicationic six-ring polyamides ImPyPy-(R)<sup>H2N</sup>γ-PyPyPy-β-Dp (**5-R**) and ImPyPy-(S)<sup>H2N</sup>γ-PyPyPy-β-Dp (**5-S**) which were synthesized by solid phase methods.

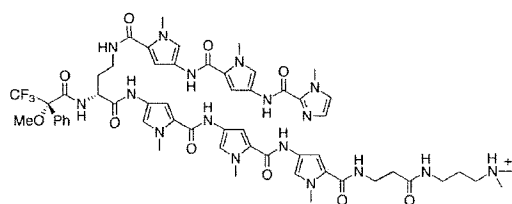
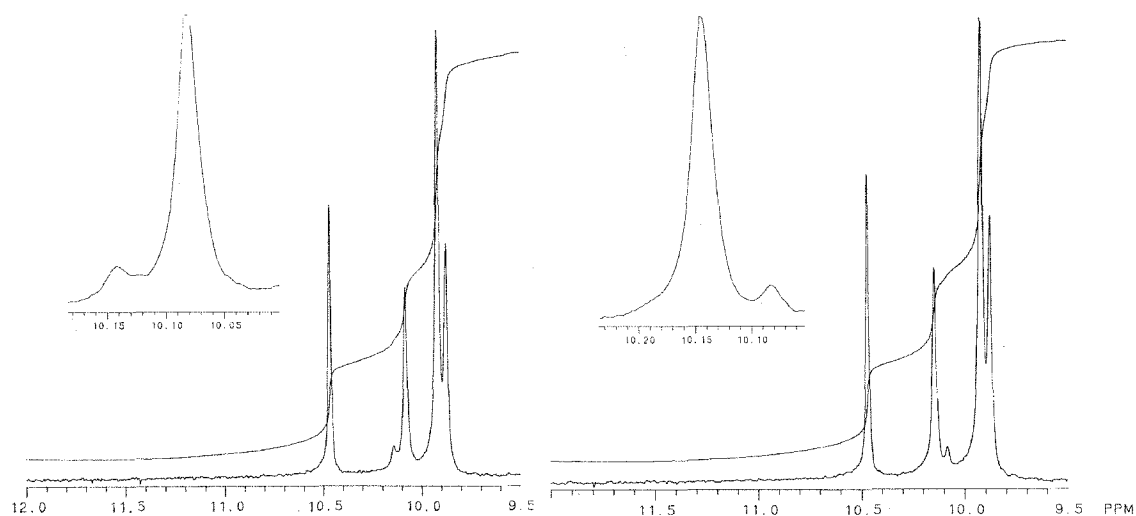
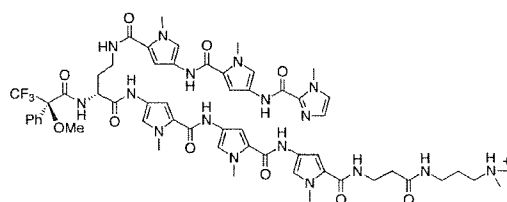
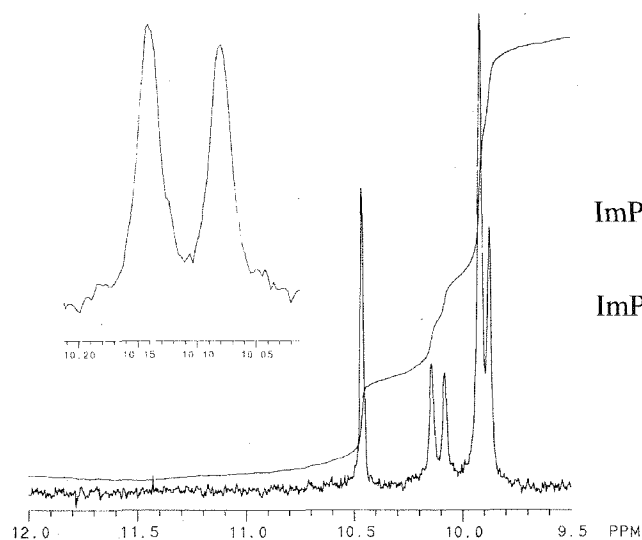
As a control, the monocationic polyamide ImPyPy-(R)<sup>H2N</sup>γ-PyPyPy-β-EtOH (**6-R**) which lacks a charge at the C-terminus was prepared. In order to further probe steric effects, the α-acetamido polyamides ImPyPy-(R)<sup>Ac</sup>γ-PyPyPy-β-Dp (**7-R**) and ImPyPy-(S)<sup>Ac</sup>γ-PyPyPy-β-Dp (**7-S**) were also studied (Figure 5.12). The EDTA analogs ImPyPy-(R)<sup>H2N</sup>γ-PyPyPy-β-Dp-EDTA•Fe(II) (**8-R•Fe(II)**), ImPyPy-(S)<sup>H2N</sup>γ-PyPyPy-β-Dp-EDTA•Fe(II) (**8-S•Fe(II)**), ImPyPy-(R)<sup>EDTA•Fe(II)</sup>γ-PyPyPy-β-Dp (**9-R•Fe(II)**), and ImPyPy-(S)<sup>EDTA•Fe(II)</sup>γ-PyPyPy-β-Dp (**9-S•Fe(II)**) were constructed to confirm the binding orientation of the modified hairpins (Figure 5.12).



**Figure 5.13.** Solid phase synthetic scheme exemplified for ImPyPy-(R)<sup>H2N</sup>-γ-PyPyPy-β-Dp **5-R**, ImPyPy-(R)<sup>H2N</sup>-γ-PyPyPy-β-EtOH **6-R**, ImPyPy-(R)<sup>Ac</sup>-γ-PyPyPy-β-Dp **7-R**, ImPyPy-(R)<sup>H2N</sup>-γ-PyPyPy-β-Dp-EDTA **8-R**, and ImPyPy-(R)<sup>EDTA</sup>-γ-PyPyPy-β-Dp **9-R**: (i) 80% TFA/DCM, 0.4 M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4 M PhSH; (iv) Boc-Py-OBt, DIEA, DMF; (v) 80% TFA/DCM, 0.4 M PhSH; (vi) Boc-Py-OBt, DIEA, DMF; (vii) 80% TFA/DCM, 0.4 M PhSH; (viii) Fmoc-α-Boc-γ-diaminobutyric acid (HBTU, DIEA); (ix) 80% TFA/DCM, 0.4 M PhSH; (x) Boc-Py-OBt, DIEA, DMF; (xi) 80% TFA/DCM, 0.4 M PhSH; (xii) Boc-Py-OBt, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4 M PhSH; (xiv) imidazole-2-carboxylic acid (HBTU/DIEA); (xv) 80% Piperidine:DMF (25 °C, 30 min) (xvi) *N,N*-dimethylaminopropylamine (55°C, 18 h) for **5-R**; ethanolamine (55°C, 18 h) for **6-R**; (xvii) Ac<sub>2</sub>O (for **7-R**), EDTA-dianhydride (for **9-R**), DIEA, DMF (55°C, 30 min.); (xviii) 0.1 NaOH (55°C, 10 min.); (xix) Boc<sub>2</sub>O, DIEA, DMF; (xx) 3,3'-diamino-*N*-methyldipropylamine (55°C, 18 h); (xxi) EDTA-dianhydride, DMSO, NMP, DIEA (55°C, 30 min.); (xxii) 0.1M NaOH, (55°C, 10 min.) (xxiii) TFA; (Inset) Pyrrole, Imidazole, and diaminobutyric acid monomers for solid phase synthesis: (R)-Fmoc-α-Boc-γ-diaminobutyric acid, Boc-Pyrrole-OBt ester (Boc-Py-OBt), and Imidazole-2-Carboxylic acid<sup>2a</sup>.

We report here the DNA-binding affinity, orientation, and sequence selectivity of five six-ring hairpin polyamides **5-R**, **5-S**, **6-R**, **7-R**, and **7-S** for 5-bp 5'-TGTTA-3', 5'-ACATT-3' and 5'-TGTC A-3' sequences. Three separate techniques are used to characterize the DNA-binding properties of the designed polyamides: MPE•Fe(II) footprinting, affinity cleaving, and DNase I footprinting.

Two polyamide-resins, ImPyPy-(*R*)<sup>Fmoc</sup>γ-PyPyPy-β-Pam-resin and ImPyPy-(*S*)<sup>Fmoc</sup>γ-PyPyPy-Q-Pam-resin, were synthesized in 14 steps from Boc-β-alanine-Pam-resin (1 g resin, 0.2 mmol/g substitution) using previously described Boc-chemistry machine-assisted protocols (Figure 5.4).<sup>6</sup> (*R*)- and (*S*)-2,4-diaminobutyric acid residues were introduced as orthogonally protected *N*-α-Fmoc-*N*-γ-Boc derivatives (HBTU, DIEA). Fmoc protected polyamide resins, ImPyPy-(*R*)<sup>Fmoc</sup>γ-PyPyPy-β-Pam-resin and ImPyPy-(*S*)<sup>Fmoc</sup>γ-PyPyPy-β-Pam-resin, were treated with 1:4 DMF:Piperidine (22°C, 30 min.) to provide ImPyPy-(*R*)<sup>H2N</sup>γ-PyPyPy-β-Pam-resin and ImPyPy-(*S*)<sup>H2N</sup>γ-PyPyPy-β-Pam-resin, respectively. A single-step aminolysis of the resin ester linkage was used to cleave the polyamide from the solid support. A sample of resin (240 mg) was treated with either dimethylaminopropylamine (55°C, 18 h) to provide **5-R**, **5-S**, **7-R**, and **7-S** or ethanolamine (55°C, 18 h) to provide **6-R**. Resin cleavage products were purified by reverse phase HPLC to provide ImPyPy-(*R*)<sup>H2N</sup>γ-PyPyPy-β-Dp (**5-R**), ImPyPy-(*S*)<sup>H2N</sup>γ-PyPyPy-β-Dp (**5-S**), and ImPyPy-(*R*)<sup>H2N</sup>γ-PyPyPy-β-EtOH (**6-R**). The stereochemical purity of **5-R** was determined to be > 98% by Mosher amide analysis.<sup>12</sup> **5-R,R** and **5-R,S** Mosher amides were prepared by reaction of **5-R** with HOBt activated esters generated *in situ* from (*R*)-α-methoxy-α-(trifluoromethyl)phenylacetic acid and (*S*)-α-methoxy-α-(trifluoromethyl)phenylacetic acid (Figure 5.14). For synthesis of analogs modified with EDTA at the carboxy-terminus, the amine-resin was treated with Boc-anhydride (DMF, DIEA, 55°C, 30 min) to provide ImPyPy-(*R*)<sup>Boc</sup>γ-PyPyPy-β-Pam-resin and ImPyPy-(*S*)<sup>Boc</sup>γ-PyPyPy-β-Pam-resin (Figure 5.13).

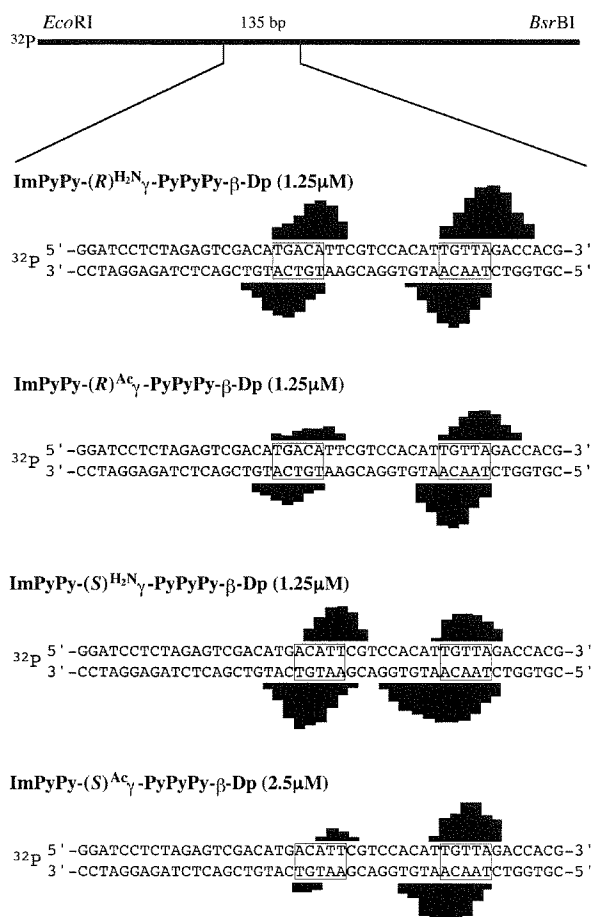
ImPyPy-(R)(S)MTPA- $\gamma$ -PyPyPy- $\beta$ -Dp (**5-R,S**)ImPyPy-(R)(R)MTPA- $\gamma$ -PyPyPy- $\beta$ -Dp (**5-R,R**)ImPyPy-(R)(R)MTPA- $\gamma$ -PyPyPy- $\beta$ -Dp (**5-R,R**)

+

ImPyPy-(R)(S)MTPA- $\gamma$ -PyPyPy- $\beta$ -Dp (**5-R,S**)**Figure 5.14.** Mosher amide analysis of chiral hairpin polyamide **5-R**.

A sample of Boc-resin was then cleaved with 3,3'-diamino-*N*-methyldipropylamine (55°C, 18 h) and purified by reversed phase HPLC to provide either ImPyPy-(*R*)<sup>Boc</sup>γ-PyPyPy-β-Dp-NH<sub>2</sub> (**5-*R*-Boc-NH<sub>2</sub>**) or ImPyPy-(*S*)<sup>Boc</sup>γ-PyPyPy-β-Dp-NH<sub>2</sub> (**5-*S*-Boc-NH<sub>2</sub>**) which afford free primary amine groups at the C-terminus suitable for post-synthetic modification. The polyamide-amines **5-*R*-Boc-NH<sub>2</sub>** and **5-*S*-Boc-NH<sub>2</sub>** were treated with an excess of EDTA-dianhydride (DMSO/NMP, DIEA, 55°C, 15 min) and the remaining anhydride hydrolyzed (0.1 M NaOH, 55°C, 10 min). The Boc protected EDTA modified polyamides ImPyPy-(*R*)<sup>Boc</sup>γ-PyPyPy-β-Dp-EDTA (**8-*R*-Boc**) and ImPyPy-(*S*)<sup>Boc</sup>γ-PyPyPy-β-Dp-EDTA (**8-*S*-Boc**) were isolated by HPLC. Individual Boc-EDTA-polyamides were deprotected with neat TFA (22°C, 1 h) to provide the respective C-terminal EDTA derivatives, ImPyPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-PyPyPy-β-Dp-EDTA (**8-*R***) and ImPyPy-(*S*)<sup>H<sub>2</sub>N</sup>γ-PyPyPy-β-Dp-EDTA (**8-*S***). For the synthesis of acetamide-turn or EDTA-turn derivatives, a sample of the α-amino polyamide ImPyPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-PyPyPy-β-Dp (**5-*R***) or ImPyPy-(*S*)<sup>H<sub>2</sub>N</sup>γ-PyPyPy-β-Dp (**5-*S***) was treated with an excess of either acetic anhydride or EDTA-dianhydride (DMSO/NMP, DIEA 55°C, 30 min) and the remaining anhydride hydrolyzed (0.1 M NaOH, 55°C, 10 min). The polyamides ImPyPy-(*R*)<sup>Ac</sup>γ-PyPyPy-β-Dp (**7-*R***), ImPyPy-(*S*)<sup>Ac</sup>γ-PyPyPy-β-Dp (**7-*S***), ImPyPy-(*R*)<sup>EDTA</sup>γ-PyPyPy-β-Dp (**8-*R***) and ImPyPy-(*S*)<sup>EDTA</sup>γ-PyPyPy-β-Dp (**8-*S***) were then isolated by reverse phase HPLC.

MPE•Fe(II) footprinting reveals that the polyamides bind with highest affinity to the 5'-TGTTA-3' match site, the 5'-TGACA-3' single base pair mismatch site for polyamides **5-*R*** and **7-*R***; and the 5'-ACATT-3' reverse orientation match site for polyamides **5-*S*** and **7-*S*** (Figure 5.6). Each polyamide at 1 μM concentration bind to the 5'-TGTTA-3' match site (25 mM Tris-acetate, 10 mM NaCl, 100 μM/base pair calf thymus DNA, pH 7.0 and 22°C). Compounds **5-*R*** and **7-*R***, each at 1.25 μM, protect both the cognate 5'-TGTA-3' site and the single base pair mismatch sequence 5'-TGTC-3'. Remarkably, binding sequence preferences vary for the

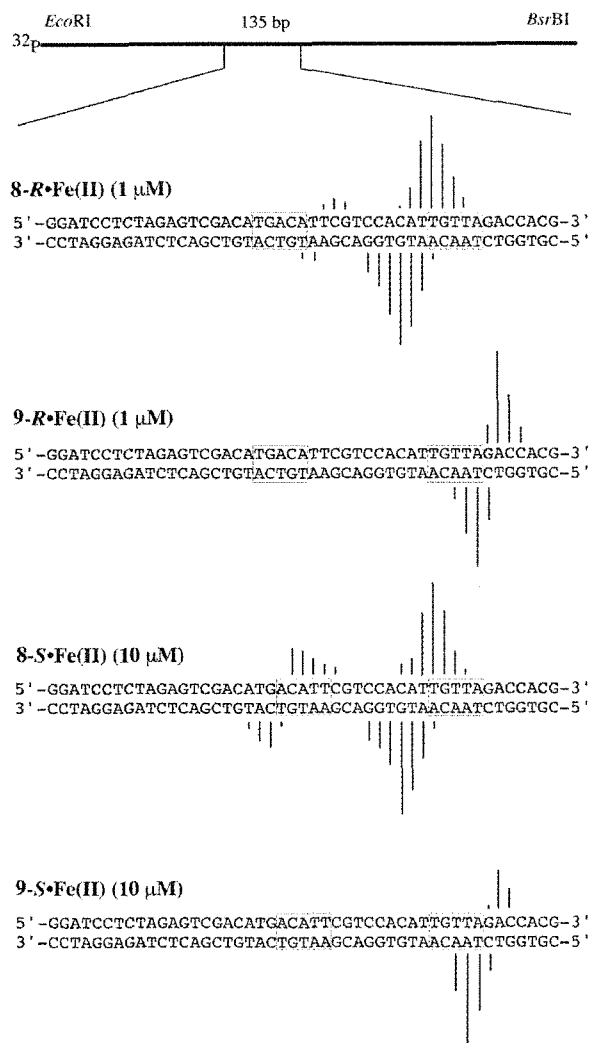


**Figure 5.15.** (Top) Illustration of the 135 bp restriction fragment with the position of the sequence indicated. Bar heights are proportional to the relative protection from cleavage at each band. Binding sites are boxed. (Bottom) MPE•Fe(II) protection patterns of 1.25 μM ImPyPy-(R)<sup>H<sub>2</sub>N</sup>γ-PyPyPy-β-Dp **5-R**, 1.25 μM ImPyPy-(R)<sup>Ac</sup>γ-PyPyPy-β-Dp **7-R**, 1.25 μM ImPyPy-(S)<sup>H<sub>2</sub>N</sup>γ-PyPyPy-β-Dp **5-S**, and 2.5 μM ImPyPy-(S)<sup>Ac</sup>γ-PyPyPy-β-Dp **7-S**.

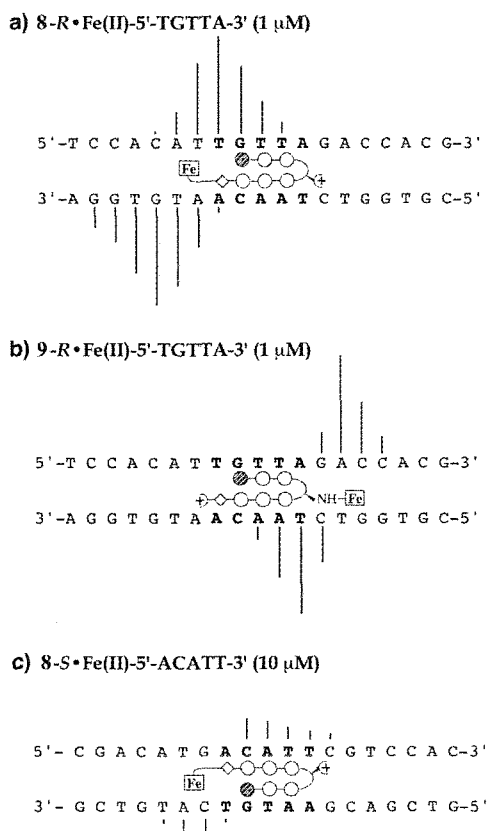
polyamides depending on the stereochemistry of the amine substituent. At 1.25 μM and 2.5 μM concentration respectively, polyamides **5-S** and **7-S** bind a 5'-ACATT-3' reverse orientation match site in addition to the target match site 5'-TGTTA-3'. The sizes of the asymmetrically 3'-shifted footprint cleavage protection patterns for the polyamides are consistent with 5 base pair binding sites.

Affinity cleavage experiments using hairpin polyamides modified with EDTA•Fe(II), at either the C-terminus or on the γ-turn, were used to determine polyamide binding orientation and stoichiometry. Affinity cleavage experiments were performed on the same 3'- and 5'-<sup>32</sup>P end-labeled 135 base pair restriction fragment (25 mM Tris-acetate, 10 mM NaCl, 100 μM/base pair calf thymus DNA, pH 7.0 and 22°C).





**Figure 5.16.** (Top) Illustration of the 135 bp restriction fragment with the position of the sequence indicated. Line heights are proportional to the relative cleavage at each band. Binding sites determined by MPE•Fe(II) footprinting and quantitated by DNase I footprint titrations are boxed. (Bottom) Affinity cleavage patterns for ImPyPy-(R)<sup>H2N</sup>γ-PyPyPy-β-Dp-EDTA•Fe(II) (**8-R•Fe(II)**) and ImPyPy-(R)<sup>EDTA•Fe(II)</sup>γ-PyPyPy-β-Dp (**9-R•Fe(II)**) at 1  $\mu$ M concentration; ImPyPy-(S)<sup>H2N</sup>γ-PyPyPy-β-Dp-EDTA•Fe(II) (**8-S•Fe(II)**) and ImPyPy-(S)<sup>EDTA•Fe(II)</sup>γ-PyPyPy-β-Dp (**9-S•Fe(II)**) at 10  $\mu$ M concentration.



**Figure 5.17.** Affinity cleavage patterns and ball and stick models of the six-ring EDTA•Fe(II) analogs. Bar heights are proportional to the relative cleavage intensities at each base pair. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent β-alanine. The boxed **Fe** denotes the EDTA•Fe(II) cleavage moiety. (a) ImPyPy-(R)<sup>H2N</sup>γ-PyPyPy-β-Dp-EDTA•Fe(II) **8-R•Fe(II)**•5'-TGTTA-3' at 1  $\mu$ M concentration; (b) 1  $\mu$ M ImPyPy-(R)<sup>EDTA•Fe(II)</sup>γ-PyPyPy-β-Dp **9-R•Fe(II)**•5'-TGTTA-3'; and (c) 10  $\mu$ M ImPyPy-(S)<sup>H2N</sup>γ-PyPyPy-β-Dp-EDTA•Fe(II) **8-S•Fe(II)**•5'-ACATT-3'.

**Table 5.2.** Equilibrium Association Constants ( $M^{-1}$ )<sup>a,b</sup>.

	match site 5'-TGTTA-3'	reverse site 5'-ACATT-3'	mismatch site 5'-TGACA-3'	specificity <sup>c</sup>
ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp	$2.9 \times 10^8$		$4.8 \times 10^6$	60
ImPyPy-( <i>R</i> ) <sup>H<sub>2</sub>N</sup> $\gamma$ -PyPyPy- $\beta$ -Dp	$3.8 \times 10^9$	n/d	$3.5 \times 10^7$	100
ImPyPy-( <i>S</i> ) <sup>H<sub>2</sub>N</sup> $\gamma$ -PyPyPy- $\beta$ -Dp	$2.2 \times 10^7$	$4.6 \times 10^6$	n/d	5
ImPyPy-( <i>R</i> ) <sup>H<sub>2</sub>N</sup> $\gamma$ -PyPyPy- $\beta$ -EtOH	$3.3 \times 10^9$	n/d	$3.1 \times 10^7$	100
ImPyPy-( <i>R</i> ) <sup>Ac</sup> $\gamma$ -PyPyPy- $\beta$ -Dp	$3.0 \times 10^8$	n/d	$< 5.0 \times 10^6$	$> 60$
ImPyPy-( <i>S</i> ) <sup>Ac</sup> $\gamma$ -PyPyPy- $\beta$ -Dp	$< 5.0 \times 10^6$	$< 5.0 \times 10^6$	n/d	n/d

<sup>a</sup>The reported association constants are the average values obtained from three DNase I footprint titration experiments. The standard deviation for each data set is indicated in parentheses. The assays were carried out at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. <sup>b</sup>The five base-pair binding sites are in capital letters. <sup>c</sup>Specificity is calculated by  $K_a(5'-TGTTA-3')/K_a(5'-ACATT-3' \text{ or } 5'-TGACA-3')$ .

symmetric single base pair mismatch sequence 5'-TGACA-3' in two distinct orientations. Polyamides linked with the *S*-enantiomer of 2,4-diaminobutyric acid bind to a 5'-ACATT-3' reverse orientation match sequence as revealed by a unique cleavage loci at the 5' side of the site.

In the presence of 3.3  $\mu$ M of **8-*R*•Fe(II)** and 10  $\mu$ M **8-*S*•Fe(II)** which have an EDTA•Fe(II) moiety at the C-terminus, a single cleavage locus proximal to the 5' side of the 5'-TGTTA-3' match sequence is revealed. In the presence of 3.3  $\mu$ M **9-*R*•Fe(II)** and 10  $\mu$ M **5-*S*•Fe(II)** which have an EDTA•Fe(II) moiety appended to the  $\gamma$ -turn, a single cleavage locus is revealed proximal to the 3' side of the 5'-TGTTA-3' match sequence. Cleavage loci are more concise for the  $\gamma$ -turn EDTA•Fe(II) placement relative to carboxy terminal placement, consistent with the shorter tether. Cleavage loci are observed at both the 5' and 3' side of the 5'-TGTTA-3' single base pair mismatch site in the presence of 10  $\mu$ M of **8-*R*•Fe(II)**. The cleavage patterns observed at the 3' side of the site is approximately 3-fold more intense than cleavage at the 5' side. For polyamide **8-*S*•Fe(II)** at 10  $\mu$ M concentration, a single cleavage locus is revealed

proximal to the 5' side of the 5'-ACATT-3' reverse orientation match site.

Quantitative DNase I footprint titrations (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub> and 5 mM CaCl<sub>2</sub>, pH 7.0 and 22°C) were performed to determine the equilibrium association constant ( $K_a$ ) of each six-ring hairpin polyamide for the three resolved sites. The 5'-TGTTA-3' site is bound by polyamides in the order: ImPyPy-(*R*)<sup>H2N</sup>γ-PyPyPy-β-Dp (**5-R**) ( $K_a = 3.8 \times 10^9 \text{ M}^{-1}$ ) = ImPyPy-(*R*)<sup>H2N</sup>γ-PyPyPy-β-EtOH (**6-R**) ( $K_a = 3.3 \times 10^9 \text{ M}^{-1}$ ) > ImPyPy-(*R*)<sup>Ac</sup>γ-PyPyPy-β-Dp (**7-R**) ( $K_a = 3.0 \times 10^8 \text{ M}^{-1}$ ) = ImPyPy-γ-PyPyPy-β-Dp ( $K_a = 2.9 \times 10^8 \text{ M}^{-1}$ ) > ImPyPy-(*S*)<sup>H2N</sup>γ-PyPyPy-β-Dp (**5-S**) ( $K_a = 2.2 \times 10^7 \text{ M}^{-1}$ ) > ImPyPy-(*S*)<sup>Ac</sup>γ-PyPyPy-β-Dp (**7-S**) ( $K_a < 5.0 \times 10^6 \text{ M}^{-1}$ ). Equilibrium association constants for recognition of the 5'-TGACT-3' single base pair mismatch site are: ImPyPy-(*R*)<sup>H2N</sup>γ-PyPyPy-β-Dp (**5-R**) ( $K_a = 3.5 \times 10^7 \text{ M}^{-1}$ ) = ImPyPy-(*R*)<sup>H2N</sup>γ-PyPyPy-β-EtOH (**6-R**) ( $K_a = 3.1 \times 10^7 \text{ M}^{-1}$ ) > ImPyPy-(*R*)<sup>Ac</sup>γ-PyPyPy-β-Dp (**7-R**) ( $K_a < 5 \times 10^6 \text{ M}^{-1}$ ) = ImPyPy-γ-PyPyPy-β-Dp ( $K_a = 4.8 \times 10^6 \text{ M}^{-1}$ ). The polyamides ImPyPy-(*S*)<sup>H2N</sup>γ-PyPyPy-β-Dp (**5-S**) and ImPyPy-(*S*)<sup>Ac</sup>γ-PyPyPy-β-Dp (**7-S**) recognize the 5'-ACATT-3' reverse orientation sequence with  $K_a = 4.6 \times 10^6 \text{ M}^{-1}$  and  $K_a < 5 \times 10^6 \text{ M}^{-1}$  respectively. It should be noted that a detailed comparison of the relative mismatch binding energetics cannot be made since the 5'-TGACA-3' and 5'-ACATT-3' binding sites overlap. The relative affinity of 5'-TGTTA-3' match site binding varies from 100-fold to 5-fold depending on the stereochemistry of the γ-turn substitutions (Table 5.2).

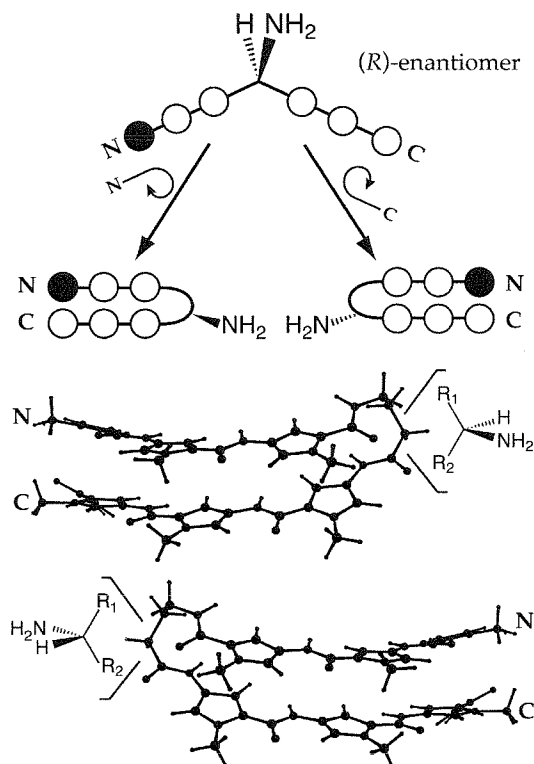
All six polyamides bind to the 5'-TGTTA-3' target site with binding isotherms consistent with binding as an intramolecular hairpin. However, the relative match site binding affinity varies by nearly 1000-fold depending on the stereochemistry and nature of chiral substituents. Among the six polyamides, ImPyPy-(*R*)<sup>H2N</sup>γ-PyPyPy-β-Dp **5-R** binds to the targeted 5'-TGTTA-3' site with the highest affinity. ImPyPy-(*R*)<sup>H2N</sup>γ-PyPyPy-β-Dp binds with an association constant,  $K_a = 3 \times 10^9 \text{ M}^{-1}$ , a factor of 10 greater than that of the parent polyamide, ImPyPy-γ-PyPyPy-β-Dp ( $K_a = 3 \times 10^8 \text{ M}^{-1}$ ). Replacement of the C-terminal dimethylaminopropylamine tail

group of **5-R**, with an ethoxyamide group as in ImPyPy- $\gamma$ -PyPyPy- $\beta$ -EtOH (**6-R**), results in no decrease in binding affinity ( $K_a = 3 \times 10^9 \text{ M}^{-1}$ ). Acetylation of the  $\gamma$ -turn amino group, as in ImPyPy-(*R*)<sup>Ac</sup> $\gamma$ -PyPyPy- $\beta$ -Dp (**7-R**), reduces binding affinity 10-fold ( $K_a = 3 \times 10^8 \text{ M}^{-1}$ ) relative to **5-R**.

The observation that polyamides which differ only by replacement of the dimethylaminopropylamide group **5-R** with an ethoxyamide group **6-R** bind with similar affinity indicates that interactions between the cationic dimethylaminopropyl tail group with anionic phosphate residues or the negative electrostatic potential in the floor of the minor groove<sup>13</sup> may not contribute substantially to the energetics of hairpin-DNA binding. Furthermore, these results indicate that the observed binding enhancement is not simply the difference between a monocationic and dicationic ligand binding to the polycationic DNA helix.<sup>13</sup> The modest increased binding affinity of polyamide **5-R** may result from electrostatic interactions between the precisely placed amine group and the floor of the minor groove. Alternately the increased affinity could indicate a reduction in the degrees of freedom accessible to the free hairpin in solution resulting from a steric effect, or an electrostatic interaction between the positively charged amine group and the negative potential of the  $\alpha$ -carbonyl group.

Polyamides linked with the *S*-enantiomer of 2,4-diaminobutyric acid, ImPyPy-(*S*)<sup>H<sub>2</sub>N</sup> $\gamma$ -PyPyPy- $\beta$ -Dp (**5-S**) and ImPyPy-(*S*)<sup>Ac</sup> $\gamma$ -PyPyPy- $\beta$ -Dp (**7-S**), bind to the 5'-TGTTA-3' match site with 100-fold ( $K_a = 2 \times 10^7 \text{ M}^{-1}$ ) and 1000-fold ( $K_a < 5 \times 10^6 \text{ M}^{-1}$ ) reduced affinity relative to the *R*-enantiomer linked polyamide **5-R**. These results demonstrate that the DNA-binding affinity of chiral hairpin polyamides can be predictably regulated as a function of the stereochemistry of the turn residue.

Polyamides with a variety of chiral substitutions at the  $\gamma$ -turn bind preferentially to the 5'-TGTTA-3' match site, while overall specificity versus binding at reverse orientation and mismatch sites is modified. Replacing the  $\alpha$ -proton in the  $\gamma$ -aminobutyric acid residue of



**Figure 5.9.** Model for chiral hairpin folding: filled and unfilled circles represent Py and Im residues respectively,  $\alpha$ -amino, and  $\alpha$ -H are highlighted and shown in the  $(R)^{H^{2N}}\gamma$ . Folding pathways leading to hairpin structures suitable for (left) polyamide (N-C) recognition of DNA ‘forward orientation’ (5’-3’) and (right) polyamide (C-N) recognition of DNA ‘reverse orientation’ (5’-3’). The corresponding stereochemistry of the  $\alpha$ -position of the  $\gamma$ -turn is highlighted for each fold. ‘Forward’ hairpin structure model for ImPyPy- $\gamma$ -PyPyPy was generated from NMR structure coordinates using Chem3D software. ‘Reverse’ hairpin structure (bottom) was generated by performing a mirror transformation of the ‘forward’ hairpin.

ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp, with an amino group that confers a chiral  $\alpha$ -hydrogen ( $R$ ) conformation, provides the most specific polyamide ImPyPy- $(R)^{H^{2N}}\gamma$ -PyPyPy- $\beta$ -Dp (**5-R**). The ImPyPy- $(R)^{H^{2N}}\gamma$ -PyPyPy- $\beta$ -Dp•5’-TGTTA-3’ complex forms with 100-fold preference relative to the ImPyPy- $(R)^{H^{2N}}\gamma$ -PyPyPy- $\beta$ -Dp•5’-TGTC A-3’ mismatch complex. Substitution of the charged dimethylaminopropyl tail group with an ethoxyamide group as in (**6-R**) does not alter binding specificity. The modest increase in specificity against single base mismatch sequences for polyamides **5-R** and **6-R** (100-fold) relative to the parent unsubstituted hairpin polyamide (60-fold) implicates chiral hairpin polyamides as an optimized class of small molecules for

recognition of the DNA minor groove.

In principle, a polyamide:DNA complex can form at two different DNA sequences depending on the alignment of the polyamide (N-C) with the walls of the minor groove (5'-3'). A six-ring hairpin polyamide of core sequence composition ImPyPy- $\gamma$ -PyPyPy which places the N-terminus of each three-ring polyamide subunit at the 5'-side of individual recognized DNA strands would bind to 'forward match' 5'-WGWW-3' sequences (W = A or T). Placement of the polyamide N-terminus at the 3'-side of each recognized strand would result in targeting 'reverse match' 5'-WCWW-3' sequences. For hairpin polyamides there is an energetic preference for 'forward' alignment of each polyamide subunit (N-C) with respect to the backbone (5'-3') of the DNA double helix.<sup>2f</sup>

In addition to decreasing the affinity for the 5'-TGTTA-3' match site, replacing the  $\alpha$ -proton of  $\gamma$ -aminobutyric acid in ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp with an amino group that confers a chiral  $\alpha$ -carbon (*S*) conformation changes the mismatch sequence preference from 5'-TGTC A-3' to binding at a reverse match site 5'-ACATT-3' as compared to the single base mismatch 5'-TGACA-3' site bound by the (*R*)-linked polyamides. Binding to the reverse 5'-ACATT-3' site may result from the presence of the steric bulk of the amino or acetamido groups in the floor of the minor groove preventing the deep polyamide binding required for specific DNA recognition. However, an analysis of hairpin folding requirements for 'forward' and 'reverse' binding reveals an additional model.

In principle, there exists two non-superimposable hairpin folds which are related by mirror plane symmetry (Figure 5.9). One hairpin fold is responsible for the preferred 5' to 3' N to C orientation, while the other fold corresponds to the 3' to 5' N to C reverse orientation binding. In the absence of DNA, each non-superimposable fold of an achiral hairpin polyamide should be energetically equivalent. However, an asymmetrically folded hairpin polyamide with a chiral substituent may display differential energetics for oriented binding. In the forward folded hairpin

the (*R*) enantiomer of 2,4-diaminobutyric acid directs the amine functionality away from the floor of the DNA helix, while for the (*S*) enantiomer, the amine functionality is predicted to direct the amine into the floor of the DNA helix. For the ‘reverse’ fold hairpin, the (*S*) enantiomer of 2,4-diaminobutyric acid directs the amine functionality away from the floor of the DNA helix, while the (*R*) enantiomer is predicted to clash with the floor of the helix. The modest enhanced specificity of chiral polyamides **5-*R*** and **6-*R*** relative to the unsubstituted achiral parent hairpin may result from stabilization of the forward binding mode and/or destabilization of the reverse binding hairpin fold.

The results presented here reveal properties of chiral structure elements that will guide future polyamide design: (i) Amine substituents conferring (*R*) chirality on the  $\alpha$ -carbon of the  $\gamma$ -turn amino acid enhance DNA-binding affinity and specificity relative to the unsubstituted parent hairpin, providing for an optimized class of hairpin polyamides. (ii) (*R*)- $\alpha$ -acetamido substituents at the  $\gamma$ -turn do not compromise affinity or specificity relative to the parent achiral hairpin, providing a convenient synthetic attachment point at the ‘capped’ end of the molecule. (iii)  $\alpha$ -amino substituents with an (*S*) configuration at the  $\gamma$ -turn bind with enhanced affinity to reverse orientation sights relative to the parent hairpin, indicating chiral hairpin substituents as a possible general approach for regulation of hairpin polyamide binding orientation. These results set the stage for preparation of a variety of new chiral hairpin polyamide structures for specific recognition in the DNA minor groove.

## Experimental Section

All gel electrophoresis and footprinting assays were for increasing for increasing performed by Sarah White<sup>16</sup> and 'Dread Boy' Herman<sup>17</sup> as described elsewhere.

Dicyclohexylcarbodiimide (DCC), Hydroxybenzotriazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU) and 0.2 mmol/gram Boc- $\beta$ -alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc- $\beta$ -Pam-Resin) and Boc- $\gamma$ -aminobutyric acid were purchased from Peptides International. *N,N*-diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF), *N*-methylpyrrolidone (NMP), DMSO/NMP, Acetic anhydride (Ac<sub>2</sub>O), and 0.0002 M potassium cyanide/pyridine were purchased from Applied Biosystems. Dichloromethane (DCM) and triethylamine (TEA) was reagent grade from EM, thiophenol (PhSH) and dimethylaminopropylamine from Aldrich, (*R*)-2-Fmoc-4-Boc-diaminobutyric acid, (*S*)-2-Fmoc-4-Boc-diaminobutyric acid, and (*R*)-2-amino-4-Boc-diaminobutyric acid were from Bachem, Mosher acids: (*R*)- $\alpha$ -methoxy- $\alpha$ -(trifluoromethyl)phenylacetic acid and (*S*)- $\alpha$ -methoxy- $\alpha$ -(trifluoromethyl)phenylacetic acid were from Aldrich, trifluoroacetic acid (TFA) from Halocarbon, phenol from Fisher, and ninhydrin from Pierce. All reagents were used without further purification.

Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. and were used for filtration of DCU. Disposable polypropylene filters were also used for washing resin for ninhydrin and picric acid tests, and for filtering pre-dissolved amino acids into reaction vessels. A shaker for manual solid phase synthesis was obtained from St. John Associates, Inc. Screw-cap glass peptide synthesis reaction vessels (5 mL and 20 mL) with a #2 sintered glass frit were made as described by Kent.<sup>14</sup> <sup>1</sup>H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz in DMSO-*d*<sub>6</sub>, with chemical shifts reported in parts per million relative



to residual solvent. UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on either a HP 1090M analytical HPLC or a Beckman Gold system using a RAINEN C<sub>18</sub>, Microsorb MV, 5 $\mu$ m, 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 reverse phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100  $\mu$ m C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. 18M $\Omega$  water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2  $\mu$ m filtered.

**Resin Substitution:** Resin substitution can be calculated as  $L_{\text{new}}(\text{mmol/g}) = L_{\text{old}} / (1 + L_{\text{old}}(W_{\text{new}} - W_{\text{old}}) \times 10^{-3})$ , where L is the loading (mmol of amine per gram of resin), and W is the weight (gmol<sup>-1</sup>) of the growing polyamide attached to the resin.<sup>15</sup>

**ImPyPy- $\gamma$ -PyPyPy- $\beta$ -EtOH (2).** ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Pam-Resin was synthesized by machine-assisted solid phase methods. A sample of resin (240 mg, 0.18 mmol/gram) was placed in a glass scintillation vial and treated with neat ethanolamine (2 mL). The reaction mixture was placed in an oven and periodically agitated (55°C, 24 h). Upon completion of polyamide cleavage, the reaction mixture was filtered to remove resin, 0.1% (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC chromatography. ImPyPy- $\gamma$ -PyPyPy- $\beta$ -EtOH is recovered upon lyophilization as a white powder (6.8 mg, 16% recovery). UV $\lambda_{\text{max}}$  246, 306 (50,000); <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.47 (s, 1 H), 9.90 (s, 1 H), 9.88 (s, 1 H), 9.87 (s, 1 H), 9.82 (s, 1 H), 8.03 (t, 1 H, *J* = 5.2 Hz), 7.95 (t, 1 H, *J* = 6.0 Hz), 7.85 (t, 1 H, *J* = 5.2 Hz), 7.38 (s, 1 H), 7.25 (d, 1 H, *J* = 1.4 Hz), 7.20 (d, 1 H, *J* = 1.3 Hz), 7.16 (d, 1 H, *J* = 1.4 Hz), 7.14 (m, 2 H), 7.11 (d, 1 H, *J* = 1.5 Hz), 7.05 (s, 1 H), 7.00 (d, 1 H, *J* = 1.3 Hz), 6.87 (d, 1 H, *J* = 1.4 Hz), 6.84 (d, 1 H, *J* = 1.4 Hz), 6.78 (d, 1 H, *J* = 1.4 Hz), 3.96 (s, 3 H), 3.81 (s, 3 H), 3.80 (s, 3 H), 3.79 (s, 3 H)

3.77 (s, 3 H), 3.76 (s, 3 H), 3.34 (m, 4 H), 3.17 (q, 2 H,  $J = 5.7$  Hz), 3.05 (q, 2 H,  $J = 5.9$  Hz), 2.27 (m, 4 H), 1.72 (quintet, 2 H,  $J = 6.9$  Hz); MALDI-TOF-MS, 936.3 (937.0 calc. for M+H).

**Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Me (4).** Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Pam-Resin was synthesized by machine-assisted solid phase methods. A sample of resin (240 mg, 0.18 mmol/gram) was placed in a glass scintillation vial and treated with a saturated solution of methylamine in DMF (20 ml). The reaction mixture was placed in sealed Parr apparatus and periodically agitated (55°C, 24 h, 80 psi). Upon completion of cleavage, the reaction mixture was cooled to room temperature and filtered to remove resin. Excess DMF was removed *in vacuo*, 0.1% (wt/v) TFA added (6 mL), and the resulting solution purified by reversed phase HPLC chromatography. Upon lyophyllization Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Me is recovered as a white powder (20.6 mg, 47% recovery). UV  $\lambda_{\text{max}}$ , 248, 312 (50,000);  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.34 (s, 1 H), 9.53 (s, 1 H) 9.90 (s, 3 H), 9.85 (s, 1 H), 9.5 (br s, 1 H), 8.03 (m, 2 H), 7.81 (q, 1 H,  $J = 4.2$  Hz), 7.42 (s, 1 H), 7.24 (d, 1 H,  $J = 1.6$  Hz), 7.22 (d, 1 H,  $J = 1.5$  Hz), 7.17 (d, 1 H,  $J = 1.5$  Hz), 7.17 (m, 2 H), 7.12 (d, 1 H,  $J = 1.5$  Hz), 7.00 (d, 1 H,  $J = 1.6$  Hz), 6.88 (d, 1 H,  $J = 1.6$  Hz), 6.85 (d, 1 H,  $J = 1.6$  Hz), 6.80 (d, 1 H,  $J = 1.5$  Hz), 3.93 (s, 3 H), 3.83 (s, 3 H), 3.81 (m, 6 H), 3.77 (m, 6 H), 3.34 (q, 2 H,  $J = 6.1$  Hz), 3.18 (q, 2 H,  $J = 5.4$  Hz), 3.03 (q, 2 H,  $J = 5.2$  Hz), 2.76 (d, 6 H,  $J = 4.5$  Hz), 2.54 (d, 3 H,  $J = 4.3$  Hz), 2.37 (t, 2 H,  $J = 6.3$  Hz), 2.26 (m, 4 H), 1.87 (quintet, 2 H,  $J = 6.6$  Hz), 1.76 (quintet, 2 H,  $J = 6.8$  Hz), MALDI-TOF-MS, 1034.4 (1035.2 calc. for M+H).

**ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-NH<sub>2</sub> (1-NH<sub>2</sub>).** A sample of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Pam-resin (350 mg, 0.18 mmol/gram) was placed in a glass scintillation vial and treated with neat 3,3'-diamino-*N*-methyldipropylamine (2 ml). The reaction mixture was placed in an oven and periodically agitated (55°C, 24 h). Resin was removed by filtration through a disposable propylene filter, and the resulting solution diluted with 0.1% (wt/v) TFA to a total volume of 8 mL, and purified directly by reversed phase HPLC to provide ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-NH<sub>2</sub> (28 mg, 40% recovery) as a white powder  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.47 (s, 1 H), 9.91 (s, 1 H), 9.90 (s, 1 H),

9.88 (s, 1 H), 9.84 (s, 1 H), 9.2 (br s, 1 H), 8.0 (m, 3 H), 7.8 (br s, 3 H), 7.33 (s, 1 H), 7.26 (d, 1 H,  $J = 1.2$  Hz), 7.20 (d, 1 H,  $J = 1.4$  Hz), 7.14 (m, 4 H), 7.03 (m, 2 H), 6.88 (d, 1 H,  $J = 1.4$  Hz), 6.85 (m, 2 H), 3.97 (s, 3 H), 3.82 (m, 9 H), 3.78 (m, 6 H), 3.33 (q, 2 H,  $J = 5.7$  Hz), 3.2- 3.0 (m, 8 H), 2.81 (q, 2 H,  $J = 5.8$  Hz), 2.71 (d, 3 H,  $J = 4.4$  Hz), 2.39 (t, 2 H,  $J = 5.8$  Hz), 2.22 (t, 2 H,  $J = 6.1$  Hz), 1.85 (quintet, 2 H,  $J = 6.2$  Hz), 1.78 (m, 4 H). MALDI-TOF-MS, 1022.1 (1021.2 calc. for M+H).

**ImPyPy- $\gamma$ -PyPyPy- $\beta$ -C7-NH<sub>2</sub> (2-NH<sub>2</sub>).** A sample of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Pam-resin (350 mg, 0.18 mmol/gram) was placed in a glass scintillation vial and treated with neat 1,7-diaminoheptane (2 ml). The reaction mixture was placed in an oven and periodically agitated (55°C, 24 h). Resin was removed by filtration through a disposable propylene filter, and the resulting solution dissolved with 0.1% (wt/v) TFA to a total volume of 8 mL, and purified directly by preparatory reversed phase HPLC to provide ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-NH<sub>2</sub> (28 mg, 39% recovery) as a white powder. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.62 (s, 1 H), 9.95 (s, 1 H), 9.92 (s, 1 H), 9.91 (s, 1 H), 9.87 (s, 1 H), 8.09 (t, 1 H,  $J = 5.3$  Hz), 8.05 (t, 1 H,  $J = 5.6$  Hz), 7.90 (t, 1 H,  $J = 5.7$  Hz), 7.6 (br s, 3 H), 7.46 (s, 1 H), 7.29 (d, 1 H,  $J = 1.7$  Hz), 7.22 (d, 1 H,  $J = 1.5$  Hz), 7.16 (m, 3 H), 7.13 (d, 1 H,  $J = 1.3$  Hz), 7.02 (d, 1 H,  $J = 1.6$  Hz), 6.88 (d, 1 H,  $J = 1.6$  Hz), 6.86 (d, 1 H,  $J = 1.6$  Hz), 6.82 (d, 1 H,  $J = 1.6$  Hz), 3.98 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.34 (q, 2 H,  $J = 5.9$  Hz), 3.18 (q, 2 H,  $J = 6.2$  Hz), 3.03 (q, 2 H,  $J = 5.9$  Hz), 2.73 (m, 2 H), 2.27 (m, 4 H), 1.77 (t, 2 H,  $J = 5.8$  Hz), 1.46 (m, 2 H), 1.35 (m, 2 H), 1.23 (m, 6 H); MALDI-TOF-MS, 1006.6 (1006.2 calc. for M+H).

**AcImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-NH<sub>2</sub> (3-NH<sub>2</sub>).** AcImPyPy- $\gamma$ -PyPyPy- $\beta$ -resin was synthesized by machine assisted methods. A sample of resin (350 mg, 0.18 mmol/gram) was placed in a 20 mL glass scintillation vial, and treated with 2 mL 3,3'-diamino-*N*-methyldipropylamine (55°C, 18 h). Resin was removed by filtration through a propylene filter, and the resulting solution diluted with 0.1% (wt/v) TFA to a total volume of 8 mL, and purified directly by reversed phase HPLC to

provide AcImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-NH<sub>2</sub> (29 mg, 43% recovery) as a white powder. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.26 (s, 1 H), 10.17 (s, 1 H), 9.92 (m, 2 H), 9.90 (s, 1 H), 9.87 (s, 1 H), 9.5 (br s, 1 H), 8.12 (t, 2 H, *J* = 5.6 Hz), 8.09 (m, 2 H), 7.9 (br s, 3 H), 7.41 (s, 1 H), 7.25 (d, 1 H, *J* = 1.5 Hz), 7.21 (d, 1 H, *J* = 1.3 Hz), 7.16 (m, 3 H), 7.11 (d, 1 H, *J* = 1.6 Hz), 7.03 (d, 1 H, *J* = 1.3 Hz), 6.89 (d, 1 H, *J* = 1.6 Hz), 6.85 (m, 2 H), 3.93 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.37 (q, 2 H, *J* = 5.9 Hz), 3.2- 3.0 (m, 8 H), 2.84 (q, 2 H, *J* = 5.7 Hz), 2.71 (d, 3 H, *J* = 4.3 Hz), 2.33 (t, 2 H, *J* = 6.6 Hz), 2.26 (t, 2 H, *J* = 6.8 Hz), 2.00 (s, 3 H), 1.88 (quintet, 2 H, *J* = 6.8 Hz), 1.77 (m, 4 H). MALDI-TOF-MS, 1078.0 (1078.2 calc. for M+H).

**Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -C7-NH<sub>2</sub> (4-NH<sub>2</sub>).** A sample of Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -resin (350 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial and treated with 2 mL 1,7-diaminoheptane (55°C, 18 h). Resin was removed by filtration through a disposable propylene filter, and the resulting solution diluted with 0.1% (wt/v) TFA to a total volume of 8 mL, and purified directly by reversed phase HPLC to provide Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -C7-NH<sub>2</sub> (37 mg, 55% recovery) as a white powder. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.40 (s, 1 H), 9.98 (s, 1 H), 9.92 (d, 2 H), 9.91 (s, 1 H), 9.87 (s, 1 H), 9.6 (br s, 1 H), 8.09 (m, 1 H), 8.05 (t, 1 H, *J* = 5.1 Hz), 7.88 (t, 1 H, *J* = 5.4 Hz), 7.6 (br s, 3 H), 7.43 (s, 1 H), 7.25 (d, 1 H, *J* = 1.6 Hz), 7.21 (d, 1 H, *J* = 1.5 Hz), 7.17 (m, 3 H), 7.13 (d, 1 H, *J* = 1.6 Hz), 7.02 (d, 1 H, *J* = 1.6 Hz), 6.89 (d, 1 H, *J* = 1.7 Hz), 6.86 (d, 1 H, *J* = 1.6 Hz), 6.83 (d, 1 H, *J* = 1.6 Hz), 3.93 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.80 (s, 3 H), 3.78 (s, 3 H), 3.32 (q, 2 H, *J* = 5.7 Hz), 3.21 (q, 2 H, *J* = 6.2 Hz), 3.05 (m, 4 H), 2.76 (d, 6 H, *J* = 3.9 Hz), 2.72 (m, 2 H), 2.39 (t, 2 H, *J* = 6.1 Hz), 2.28 (m, 4 H), 1.91 (quintet, 2 H, *J* = 6.6 Hz), 1.77 (quintet, 2 H, *J* = 6.4 Hz), 1.47 (m, 2 H), 1.36 (m, 2 H), 1.24 (m, 6 H). MALDI-TOF-MS, 1133.8 (1134.3 calc. for M+H).

**ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA (1-E).** EDTA-dianhydride (50 mg) was dissolved in 1 mL DMSO/NMP solution and 1 mL DIEA by heating at 55°C for 5 min. The dianhydride solution was added to ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-NH<sub>2</sub> (**1-NH<sub>2</sub>**) (8.0 mg, 7  $\mu$ mol) dissolved in 750  $\mu$ L

DMSO. The mixture was heated at 55°C for 25 minutes, and treated with 3 mL 0.1M NaOH, and heated at 55°C for 10 minutes. 0.1% aqueous TFA was added to adjust the total volume to 8 mL and the solution purified directly by reversed phase HPLC chromatography to provide **1-E** as a white powder. (4 mg, 40% recovery)  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.45 (s, 1 H), 9.89 (s, 1 H), 9.88 (s, 1 H), 9.86 (s, 1 H), 9.82 (s, 1 H), 9.2 (br s, 1 H), 8.39 (t, 2 H,  $J = 6.0$  Hz), 8.06 (m, 3 H), 7.37 (s, 1 H), 7.25 (d, 1 H,  $J = 1.5$  Hz), 7.17 (d, 1 H,  $J = 1.3$  Hz), 7.13 (m, 4 H), 7.02 (m, 2 H), 6.87 (d, 1 H,  $J = 1.2$  Hz), 6.85 (m, 2 H), 3.96 (s, 3 H), 3.81 (s, 3 H), 3.80 (s, 3 H), 3.79 (s, 3H), 3.77 (s, 3 H), 3.76 (s, 3 H), 3.68 (m, 4 H), 3.33 (q, 2 H,  $J = 5.3$  Hz), 3.3-3.0 (m, 16 H), 2.85 (m, 2 H), 2.69 (d, 3 H,  $J = 4.3$  Hz), 2.33 (t, 2 H,  $J = 5.5$  Hz), 2.24 (t, 2 H,  $J = 6.6$  Hz), 1.73 (m, 6 H). MALDI-TOF-MS, 1295.4 (1295.3 calc. for M+H).

**ImPyPy- $\gamma$ -PyPyPy- $\beta$ -C7-EDTA (2-E).** Compound **2-E** was prepared from compound **2-NH<sub>2</sub>** (8.0 mg, 7  $\mu\text{mol}$ ) as described for **1-E**. (2 mg, 20% recovery)  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.51 (s, 1 H), 9.94 (s, 1 H), 9.92 (s, 1 H), 9.91 (s, 1 H), 9.86 (s, 1 H), 8.33 (t, 1 H,  $J = 5.5$  Hz), 8.09 (t, 1 H,  $J = 5.8$  Hz), 8.04 (t, 1 H,  $J = 5.0$  Hz), 7.87 (t, 1 H,  $J = 5.2$  Hz), 7.40 (s, 1 H), 7.27 (d, 1 H,  $J = 1.6$  Hz), 7.21 (d, 1 H,  $J = 1.5$  Hz), 7.17 (m, 3 H), 7.14 (d, 1 H,  $J = 1.7$  Hz), 7.05 (s, 1 H), 7.02 (d, 1 H,  $J = 1.5$  Hz), 6.89 (d, 1 H,  $J = 1.6$  Hz), 6.86 (d, 1 H,  $J = 1.5$  Hz), 6.82 (d, 1 H,  $J = 1.5$  Hz), 3.97 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.67 (m, 4 H), 3.33 (q, 2 H,  $J = 5.6$  Hz), 3.22 (m, 4 H), 3.10 (m, 4 H), 3.04 (m, 4 H), 2.92 (m, 4 H), 1.77 (t, 2 H,  $J = 6.4$  Hz), 1.35 (m, 4 H), 1.22 (m, 6 H). MALDI-TOF-MS, 1280.4 (1280.3 calc. for M+H).

**AcImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA (3-E).** Compound **3-E** was prepared from compound **3-NH<sub>2</sub>** (8.0 mg, 7  $\mu\text{mol}$ ) as described for **1-E**. (3 mg, 30%).  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.25 (s, 1 H), 10.00 (s, 1 H), 9.91 (m, 2 H), 9.90 (s, 1 H), 9.85 (s, 1 H), 8.42 (t, 1 H,  $J = 5.6$  Hz), 8.08 (m, 3 H), 7.41 (s, 1 H), 7.26 (d, 1 H,  $J = 1.5$  Hz), 7.20 (d, 1 H,  $J = 1.5$  Hz), 7.16 (m, 3 H), 7.14 (d, 1 H,  $J = 1.5$  Hz), 7.04 (d, 1 H,  $J = 1.2$  Hz), 6.88 (d, 1 H,  $J = 1.6$  Hz), 6.86 (m, 4 H), 3.93 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.71 (m, 4 H), 3.58 (q, 2 H,  $J = 5.2$

Hz), 3.4- 3.0 (m, 18 H), 2.71 (d, 3 H,  $J = 4.3$  Hz), 2.33 (t, 2 H,  $J = 5.4$  Hz), 2.26 (t, 2 H,  $J = 5.7$  Hz), 2.00 (s, 3 H), 1.74 (m, 6 H). MALDI-TOF-MS, 1352.8 (1352.4 calc. for M+H).

**Dp-ImPyPy- $\gamma$ -PyPyPy- $\beta$ -C7-EDTA (4-E).** Compound 4-E was prepared from compound 4-NH<sub>2</sub> (8.0 mg, 7  $\mu$ mol) as described for 1-E.. (4 mg, 40% recovery) <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.36 (s, 1 H), 9.92 (s, 1 H), 9.88 (m, 2 H), 9.87 (s, 1 H), 9.82 (s, 1 H), 9.2 (br s, 1 H), 8.03 (m, 3 H), 7.98 (m, 1 H), 7.41 (s, 1 H), 7.22 (d, 1 H,  $J = 1.5$  Hz), 7.20 (d, 1 H,  $J = 1.6$  Hz), 7.14 (m, 4 H), 7.10 (d, 1 H,  $J = 1.6$  Hz), 7.00 (d, 1 H,  $J = 1.5$  Hz), 6.87 (d, 1 H,  $J = 1.3$  Hz), 6.83 (d, 1 H,  $J = 1.4$  Hz), 6.78 (d, 1 H,  $J = 1.6$  Hz), 3.92 (s, 3 H), 3.82 (s, 3 H), 3.80 (s, 3 H), 3.79 (s, 3 H), 3.77 (s, 3 H), 3.76 (s, 3 H), 3.41 (m, 4 H), 3.3-3.0 (m, 16 H), 2.74 (d, 6 H,  $J = 4.3$  Hz), 2.71 (m, 2 H), 2.31 (t, 2 H,  $J = 5.4$  Hz), 2.22 (t, 2 H,  $J = 5.6$  Hz), 1.77 (m, 2 H), 1.39 (m, 2 H), 1.28 (m, 4 H), 1.20 (m, 6 H). MALDI-TOF-MS, 1407.7 (1408.6 calc. for M+H).

**ImPyPy-(R)<sup>H2N</sup> $\gamma$ -PyPyPy- $\beta$ -Dp (5-R).** ImPyPy-(R)<sup>Fmoc</sup> $\gamma$ -PyPyPy- $\beta$ -Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods. (R)-2-Fmoc-4-Boc-diaminobutyric acid (0.7 mmol) was incorporated as previously described for Boc- $\gamma$ -aminobutyric acid. ImPyPy-(R)<sup>Fmoc</sup> $\gamma$ -PyPyPy- $\beta$ -Pam-Resin was placed in a glass 20 mL peptide synthesis vessel and treated with DMF (2 mL), followed by piperidine (8 mL) and agitated (22°C, 30 min.). ImPyPy-(R)<sup>H2N</sup> $\gamma$ -PyPyPy- $\beta$ -Pam-resin was isolated by filtration, and washed sequentially with an excess of DMF, DCM, MeOH, and ethyl ether and the amine-resin dried in vacuo. A sample of ImPyPy-(R)<sup>H2N</sup> $\gamma$ -PyPyPy- $\beta$ -Pam-resin (240 mg, 0.18 mmol/gram) was placed in a glass 20 mL scintillation vial and treated with neat dimethylaminopropylamine (2 mL). The reaction mixture was placed in an oven and periodically agitated (55°C, 16 h). Upon completion of polyamide resin cleavage, the reaction mixture was filtered to remove resin, 0.1% (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC chromatography. ImPyPy-(R)<sup>H2N</sup> $\gamma$ -PyPyPy- $\beta$ -Dp is recovered upon lyophilization as a white powder (32 mg, 66% recovery).  $[\alpha]_D^{20} +14.6$  (c 0.05, H<sub>2</sub>O); UV  $\lambda_{max}$  246, 310 (50,000); <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.56 (s, 1 H), 10.47 (s,

1 H), 9.97 (s, 1 H), 9.94 (s, 1 H), 9.88 (s, 1 H), 9.4 (br s, 1 H), 8.28 (s, 3 H), 8.22 (m, 1 H), 8.03 (m, 2 H), 7.38 (s, 1 H), 7.25 (d, 1 H,  $J = 1.6$  Hz), 7.22 (d, 1 H,  $J = 1.5$  Hz), 7.19 (d, 1 H,  $J = 1.5$  Hz), 7.16 (d, 1 H,  $J = 1.6$  Hz), 7.14 (d, 1 H,  $J = 1.8$  Hz), 7.12 (d, 1 H,  $J = 1.7$  Hz), 7.03 (m, 2 H), 6.95 (d, 1 H,  $J = 1.6$  Hz), 6.91 (d, 1 H,  $J = 1.6$  Hz), 6.85 (d, 1 H,  $J = 1.6$  Hz), 3.96 (s, 3 H), 3.83 (s, 3 H), 3.81 (m, 6 H), 3.79 (s, 3 H), 3.76 (s, 3 H), 3.33 (q, 2 H,  $J = 6.3$  Hz), 3.25 (q, 2 H,  $J = 5.7$  Hz), 3.05 (q, 2 H,  $J = 5.9$  Hz), 2.96 (q, 2 H,  $J = 5.3$  Hz), 2.71 (d, 6 H,  $J = 4.9$  Hz), 2.32 (t, 2 H,  $J = 7.1$  Hz), 1.95 (q, 2 H,  $J = 5.9$  Hz), 1.70 (quintet, 2 H,  $J = 7.3$  Hz); MALDI-TOF-MS (monoisotopic), 992.5 (992.5 calc. for  $C_{47}H_{62}N_{17}O_8$ ). Mosher acid (117 mg, 0.5 mmol) and HOBT (70 mg, 0.5 mmol) were dissolved in DMF (1 mL), DCC (100 mg, 0.5 mmol) added and the solution agitated for 30 min at 22°C. A sample of activated ester solution (100  $\mu$ L, 0.05 mmol) was added to ImPyPy-(*R*)<sup>H<sub>2</sub>N</sup> $\gamma$ -PyPyPy- $\beta$ -Dp **5-R** (10 mg, 0.01 mmol), DIEA (50  $\mu$ L) added, and the solution agitated for 3h at 22°C. DMF (1 mL) followed by 0.1% (wt/v) TFA (6 mL) was added to the reaction mixture and the resulting solution purified by reversed phase HPLC chromatography (1% acetonitrile/min.) under conditions which do not separate the diastereomers. This is supported by the equal **5-S,R:5-R,R** and **5-S,S:5-R,S** ratios observed for the two preparations as well as analytical HPLC analysis of crude and purified reaction mixtures. ImPyPy-(*R*)<sup>(R)MTPA</sup> $\gamma$ -PyPyPy- $\beta$ -Dp **5-R,R** (6 mg, 53% recovery). <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.50 (s, 1 H), 10.14 (s, 1 H), 9.92 (s, 2 H), 9.88 (s, 1 H), 9.2 (br s, 1 H), 8.43 (d, 1 H,  $J = 7.0$  Hz), 8.02 (m, 3 H), 7.92 (m, 1 H), 7.47 (m, 2 H), 7.41 (m, 2 H), 7.36 (s, 1 H), 7.24 (m, 1 H), 7.19 (m, 1 H), 7.15 (m, 1 H), 7.12 (m, 3 H), 7.01 (m, 2 H), 6.90 (m, 3 H), 6.83 (m, 1 H), 4.46 (q, 1 H,  $J = 5.5$  Hz), 3.94 (s, 3 H), 3.79 (m, 9 H), 3.75 (m, 6 H), 3.32 (m, 4 H), 3.05 (m, 2 H), 2.94 (m, 2 H), 2.68 (d, 6H,  $J = 4.0$  Hz), 2.28 (t, 2 H,  $J = 6.3$  Hz), 1.93 (q, 2 H,  $J = 6.1$  Hz), 1.66 (quintet, 2 H,  $J = 6.0$  Hz), 1.18 (s, 3 H); MALDI-TOF-MS (monoisotopic), 1208.5 (1208.5 calc. for  $C_{57}H_{68}F_3N_{17}O_{10}$ ). ImPyPy-(*R*)<sup>(S)MTPA</sup> $\gamma$ -PyPyPy- $\beta$ -Dp **5-R,S** (5 mg, 45% recovery). <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.47 (s, 1 H), 10.08 (s, 1 H), 9.92 (s, 2 H), 9.88 (s, 1 H), 9.2 (br s, 1 H), 8.43 (d, 1 H,  $J = 6.9$  Hz), 8.02 (m, 3 H), 7.46 (m, 2 H), 7.40 (m, 2 H), 7.36 (s, 1 H), 7.23 (m, 1 H), 7.19 (m, 1 H), 7.14 (m, 1 H),

7.12 (m, 3 H), 7.01 (m, 2 H), 6.87 (m, 3 H), 6.83 (m, 1 H), 4.44 (q, 1 H,  $J = 6.5$  Hz), 3.94 (s, 3 H), 3.79 (m, 9 H), 3.75 (m, 6 H), 3.28 (m, 4 H), 3.06 (m, 4 H), 2.94 (m, 2 H), 2.69 (d, 6H,  $J = 4.5$  Hz), 2.28 (t, 2 H,  $J = 6.5$  Hz), 1.93 (q, 2 H,  $J = 6.1$  Hz), 1.66 (quintet, 2 H,  $J = 6.0$  Hz), 1.18 (s, 3 H); MALDI-TOF-MS (monoisotopic), 1209.0 (1208.5 calc. for  $C_{57}H_{68}F_3N_{17}O_{10}$ ).

**ImPyPy-(S)<sup>H2N</sup>γ-PyPyPy-β-Dp (5-S).** ImPyPy-(S)<sup>H2N</sup>γ-PyPyPy-β-Dp was prepared as described for **5-R**. (23 mg, 49% recovery).  $[\alpha]_D^{20}$  -14.2 ( $c$  0.04,  $H_2O$ ); MALDI-TOF-MS (monoisotopic), 992.5 (992.5 calc. for  $C_{47}H_{62}N_{17}O_8$ ).

**ImPyPy-(R)<sup>H2N</sup>γ-PyPyPy-β-EtOH (6-R).** A sample of ImPyPy-(R)<sup>H2N</sup>γ-PyPyPy-β-Pam-resin (240 mg, 0.18 mmol/gram) was placed in a glass 20 mL scintillation vial and treated with neat ethanolamine (2 mL). The reaction mixture was placed in an oven and periodically agitated (55°C, 16 h). Upon completion of polyamide resin cleavage, the reaction mixture was filtered to remove resin, 0.1% (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC chromatography. ImPyPy-(R)<sup>H2N</sup>γ-PyPyPy-β-Dp is recovered upon lyophilization as a white powder (21 mg, 46% recovery).  $[\alpha]_D^{20}$  +18.6 ( $c$  0.04,  $H_2O$ ); UV  $\lambda_{max}$  246, 310 (50,000); <sup>1</sup>H NMR (DMSO- $d_6$ )  $\delta$  10.55 (s, 1 H), 10.48 (s, 1 H), 9.97 (s, 1 H), 9.94 (s, 1 H), 9.89 (s, 1 H), 8.24 (m, 4 H), 8.00 (t, 1 H,  $J = 4.1$  Hz), 7.89 (t, 1 H,  $J = 5.8$  Hz), 7.38 (s, 1 H), 7.25 (d, 1 H,  $J = 1.6$  Hz), 7.22 (d, 1 H,  $J = 1.6$  Hz), 7.21 (d, 1 H,  $J = 1.5$  Hz), 7.16 (m, 2 H), 7.14 (d, 1 H,  $J = 1.6$  Hz), 7.03 (d, 1 H,  $J = 1.7$  Hz), 6.99 (d, 1 H,  $J = 1.4$  Hz), 6.95 (d, 1 H,  $J = 1.6$  Hz), 6.91 (d, 1 H,  $J = 1.5$  Hz), 6.78 (d, 1 H,  $J = 1.5$  Hz), 5.33 (m, 1 H), 3.95 (s, 3 H), 3.83 (s, 3 H), 3.81 (m, 6 H), 3.79 (s, 3 H), 3.76 (s, 3 H), 3.37 (q, 2 H,  $J = 6.2$  Hz), 3.07 (q, 2 H,  $J = 5.9$  Hz), 2.29 (t, 2 H,  $J = 7.1$  Hz), 1.93 (q, 2 H,  $J = 5.8$  Hz), 1.20 (m, 4 H); MALDI-TOF-MS (monoisotopic), 951.4 (951.4 calc. for  $C_{44}H_{55}N_{16}O_9$ ).

**ImPyPy-(R)<sup>Ac</sup>γ-PyPyPy-β-Dp (7-R).** A sample of ImPyPy-(R)<sup>H2N</sup>γ-PyPyPy-β-Dp (4 mg) in DMSO (1 mL) was placed in a 20 mL glass scintillation vial and treated with DMF (1 mL), Acetic Anhydride (1 mL), and DIEA (1 mL) and placed in oven (55°C, 30 min.) Remaining



anhydride was hydrolyzed (0.1 M NaOH, 1 mL, 55°C, 10 min.), 0.1 % (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC chromatography. ImPyPy-(R)<sup>H<sup>2</sup>N</sup>γ-PyPyPy-β-Dp is recovered upon lyophilization as a white powder (2 mg, 50% recovery).  $[\alpha]_D^{20} +33.1$  (*c* 0.04, H<sub>2</sub>O); <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>) δ 10.49 (s, 1 H), 10.06 (s, 1 H), 9.94 (m, 2 H), 9.00 (s, 1 H), 9.4 (br s, 1 H), 8.21 (d, 1 H, *J* = 7.8 Hz), 8.06 (m, 2 H), 8.00 (t, 1 H, *J* = 6.2 Hz), 7.39 (s, 1 H), 7.27 (d, 1 H, *J* = 1.7 Hz), 7.21 (d, 1 H, *J* = 1.6 Hz), 7.18 (m, 2 H), 7.14 (m, 2 H), 7.03 (m, 2 H), 6.90 (d, 1 H, *J* = 1.6 Hz), 6.86 (m, 2 H), 4.43 (q, 1 H, *J* = 7.5 Hz), 3.96 (s, 3 H), 3.82 (m, 9 H), 3.73 (m, 6 H), 3.37 (q, 2 H, *J* = 5.8 Hz), 3.11 (q, 2 H, *J* = 6.9 Hz), 2.98 (q, 2 H, *J* = 5.4 Hz), 2.79 (q, 2 H, *J* = 5.3 Hz), 2.71 (d, 6 H, *J* = 4.7 Hz), 2.33 (t, 2 H, *J* = 6.2 Hz), 1.97 (s, 3 H), 1.70 (quintet, 2 H, *J* = 6.0 Hz) MALDI-TOF-MS (average), 1035.1 (1035.2 calc. for M+H).

**ImPyPy-(S)<sup>Ac</sup>γ-PyPyPy-β-Dp (7-S).** ImPyPy-(S)<sup>Ac</sup>γ-PyPyPy-β-Dp was prepared as described for **7-R**. (2 mg, 50% recovery).  $[\alpha]_D^{20} -30$  (*c* 0.05, H<sub>2</sub>O); MALDI-TOF-MS (monoisotopic), 1034.6 (1034.5 calc. for C<sub>49</sub>H<sub>64</sub>N<sub>17</sub>O<sub>9</sub>).

**ImPyPy-(R)<sup>Boc</sup>γ-PyPyPy-β-Dp-NH<sub>2</sub> (8-R-Boc-NH<sub>2</sub>).** A sample of ImPyPy-(R)<sup>H<sup>2</sup>N</sup>γ-PyPyPy-β-Pam-resin (300 mg, 0.18 mmol/gram) was placed in a 20 mL glass scintillation vial and treated with DMF (4 mL), Boc Anhydride (500 mg), and DIEA (1 mL) and placed in oven (55°C, 30 min.). ImPyPy-(R)<sup>Boc</sup>γ-PyPyPy-β-Pam-resin was isolated by filtration, and washed sequentially with an excess of DMF, DCM, MeOH, and ethyl ether and the Boc-resin dried *in vacuo*. A sample of ImPyPy-(R)<sup>Boc</sup>γ-PyPyPy-β-Pam-resin (240 mg, 0.18 mmol/gram) was placed in a glass 20 mL scintillation vial and treated with neat 3,3'-diamino-*N*-methyldipropylamine (2 mL). The reaction mixture was placed in an oven and periodically agitated (55°C, 16 h). Upon completion of polyamide resin cleavage, the reaction mixture was filtered to remove resin, 0.1% (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC chromatography. ImPyPy-(R)<sup>Boc</sup>γ-PyPyPy-β-Dp-NH<sub>2</sub> is recovered upon lyophilization as a white powder. (18 mg, 36 recovery); <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>) δ 10.59 (s, 1 H), 10.16 (s, 1 H), 10.04 (m, 2 H), 10.00 (s, 1

H), 9.4 (br s, 1 H), 8.31 (d, 1 H,  $J = 7.8$  Hz), 8.16 (m, 2 H), 8.10 (t, 1 H,  $J = 6.2$  Hz), 7.89 (t, 1 H,  $J = 5.8$  Hz), 7.49 (s, 1 H), 7.37 (d, 1 H,  $J = 1.7$  Hz), 7.22 (d, 1 H,  $J = 1.6$  Hz), 7.21 (d, 1 H,  $J = 1.5$  Hz), 7.16 (m, 2 H), 7.14 (d, 1 H,  $J = 1.6$  Hz), 7.03 (d, 1 H,  $J = 1.7$  Hz), 6.99 (d, 1 H,  $J = 1.4$  Hz), 6.95 (d, 1 H,  $J = 1.6$  Hz), 6.91 (d, 1 H,  $J = 1.5$  Hz), 6.78 (d, 1 H,  $J = 1.5$  Hz), 5.33 (m, 1 H), 3.95 (s, 3 H), 3.83 (s, 3 H), 3.81 (m, 6 H), 3.79 (s, 3 H), 3.76 (s, 3 H), 3.37 (q, 2 H,  $J = 6.2$  Hz), 3.07 (q, 2 H,  $J = 5.9$  Hz), 2.29 (t, 2 H,  $J = 7.1$  Hz), 1.93 (q, 2 H,  $J = 5.8$  Hz), 1.20 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1135.3 (1135.6 calc. for  $C_{54}H_{75}N_{18}O_{10}$ ).

**ImPyPy-(S)<sup>Boc</sup>γ-PyPyPy-β-Dp-NH<sub>2</sub> (8-S-Boc-NH<sub>2</sub>).** ImPyPy-(S)<sup>Boc</sup>γ-PyPyPy-β-Dp was prepared as described for **8-R**. (16 mg, 32% recovery). MALDI-TOF-MS (monoisotopic), 1135.4 (1135.6 calc. for  $C_{54}H_{75}N_{18}O_{10}$ ).

**ImPyPy-(R)<sup>Boc</sup>γ-PyPyPy-β-Dp-EDTA (8-R-Boc).** EDTA-dianhydride (50 mg) was dissolved in DMSO/NMP (1 mL) and DIEA (1 mL) by heating at 55°C for 5 min. The dianhydride solution was added to ImPyPy-(R)<sup>Boc</sup>γ-PyPyPy-β-Dp-NH<sub>2</sub> (10.4 mg, 10 μmol) dissolved in DMSO (750 μL). The mixture was heated (55°C, 25 min.) and the remaining anhydride hydrolyzed (0.1M NaOH, 3 mL, 55°C, 10 min.). 0.1% aqueous TFA was added to adjust the total volume to 8 mL and the solution purified directly by reversed phase HPLC chromatography to provide ImPyPy-(R)<sup>Boc</sup>γ-PyPyPy-β-Dp-EDTA **8-R-Boc** as a white powder (4 mg, 40% recovery). MALDI-TOF-MS (monoisotopic), 1409.6 (1409.7 calc. for  $C_{64}H_{89}N_{20}O_{17}$ ).

**ImPyPy-(S)<sup>Boc</sup>γ-PyPyPy-β-Dp-EDTA (8-S-Boc).** ImPyPy-(S)<sup>Boc</sup>γ-PyPyPy-β-Dp-NH<sub>2</sub> (12.0 mg, 12 μmol) was converted in to **8-S-Boc** as described (4 mg, 33% recovery). MALDI-TOF-MS (monoisotopic), 1409.7 (1409.7 calc. for  $C_{64}H_{89}N_{20}O_{17}$ ).

**ImPyPy-(R)<sup>H<sub>2</sub>N</sup>γ-PyPyPy-β-Dp-EDTA (8-R).** A sample of ImPyPy-(R)<sup>Boc</sup>γ-PyPyPy-β-Dp-EDTA (2.1 mg) in DMSO (750 μL) was placed in a 50 mL flask and treated with TFA (15 mL, 22°C, 2 h). Excess TFA was removed *in vacuo*, water added (6 mL) and the resulting solution purified by reversed phase HPLC chromatography. ImPyPy-(R)<sup>H<sub>2</sub>N</sup>γ-PyPyPy-β-Dp-EDTA is

recovered upon lyophilization as a white powder (1.3 mg, 50% recovery). MALDI-TOF-MS (monoisotopic), 1309.5 (1309.6 calc. for  $C_{59}H_{81}N_{20}O_{15}$ ).

**ImPyPy-(S)<sup>H<sub>2</sub>N</sup>- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA (8-S).** ImPyPy-(S)<sup>Boc</sup>- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA (3.0 mg) was converted in to **8-S** as described (1 mg, 33% recovery). MALDI-TOF-MS (monoisotopic), 1309.5 (1309.6 calc. for  $C_{59}H_{81}N_{20}O_{15}$ ).

**ImPyPy-(R)<sup>EDTA</sup>- $\gamma$ -PyPyPy- $\beta$ -Dp (9-R).** EDTA-dianhydride (50 mg) was dissolved in DMSO/NMP (1 mL) and DIEA (1 mL) by heating at 55°C for 5 min. The dianhydride solution was added to ImPyPy-<sup>H<sub>2</sub>N</sup>- $\gamma$ -PyPyPy- $\beta$ -Dp (1.0 mg, 1  $\mu$ mol) dissolved in DMSO (750  $\mu$ L). The mixture was heated (55°C, 25 min.) and the remaining anhydride hydrolyzed (0.1M NaOH, 3mL, 55°C, 10 min). 0.1% aqueous TFA was added to adjust the total volume to 8 mL and the solution purified directly by reversed phase HPLC chromatography to provide **9-R** as a white powder (0.6 mg, 60% recovery). MALDI-TOF-MS (monoisotopic), 1266.4 (1266.6 calc. for  $C_{57}H_{76}N_{19}O_{15}$ ).

**ImPyPy-(S)<sup>EDTA</sup>- $\gamma$ -PyPyPy- $\beta$ -Dp (9-S).** ImPyPy-(S)<sup>EDTA</sup>- $\gamma$ -PyPyPy- $\beta$ -Dp was prepared as described for **9-R** (6.8 mg, 16% recovery). MALDI-TOF-MS (monoisotopic), 1266.5 (1266.6 calc. for  $C_{57}H_{76}N_{19}O_{15}$ ).

## References

1. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385. (d) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572 (e) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559.
2. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingham, M. *Nature Struct. Biol.* **1994**, *1*, 169. (d) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532.
3. Gottesfield, J. M.; Nealy, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202.
4. Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983.
5. (a) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147. (b) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6153. (c) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Chem. & Biol.* **1996**, *3*, 369. (d) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198. (e) Pilch, D. S.; Pokar, N. A.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306. (f) de Claire, R. P. L.; Geierstanger B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909.
6. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.
7. (a) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457. (b) Dervan, P. B. *Science* **1986**, *232*, 464.
8. (a) Van Dyke, M. W.; Dervan, P. B. *Nucl. Acids Res.* **1983**, *11*, 5555. (b) Van Dyke, M. W.; Dervan, P. B. *Science* **1984**, *225*, 1122.
9. (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8462. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.
10. (a) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160. (b) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Nature Struct. Biol.* **1996**, *3*, 321.
11. Mecozzi, S.; West, A. P.; Dougherty, D. A. *J. Am. Chem. Soc.* **1996**, *93*, 10566-10571.
12. (a) Dale, J. A.; Mosher, H. S. *J. Am. Chem. Soc.* **1973**, *95*, 512. (b) Yamaguchi, S. *Asymmetric Synthesis (Vol. 1), Analytical Methods* p.125-152, J. D. Morrison (ed.) Academic Press (1983).

13. (a) Zimmer, C.; Wahnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31. (b) Pullman, B. *Adv. Drug. Res.* **1990**, *18*, 1. (c) Breslauer, K. J.; Ferrante, R.; Marky, L. A.; Dervan, P. B. ; Youngquist, R. S. *Structure and Expression (Vol. 2), DNA and Its Drug Complexes* p. 273-289, R. H. Sarma and M. H. Sarma (eds.) Academic Press (1988).
14. Kent, S.B.H. *Annu. Rev. Biochem.* **1988**, *57*, 957.
15. Barlos, K.; Chatzi, O.; Gatos, D.; Stravropoulos, G. *Int. J. Peptide Protein Res.* **1991**, *37*, 513.
16. White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756.
17. Herman, D.M.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1998**, *120*, 1382.

## CHAPTER 6

# Recognition of DNA by Designed Ligands at Subnanomolar Concentrations

**Abstract** Small molecules that specifically bind with high affinity to any predetermined DNA sequence in the human genome would be useful tools in molecular biology and potentially in human medicine. Simple rules have been developed to rationally control the sequence-specificity of minor groove binding polyamides containing *N*-methylimidazole (Im) and *N*-methylpyrrole (Py) amino acids. Two eight-ring Py-Im polyamides differing in sequence by a single amino acid bind specifically to respective 6-bp target sites which differ in sequence by a single base pair. Binding is observed at subnanomolar concentrations of ligand. The replacement of a single nitrogen atom with a C-H regulates affinity and specificity by two orders of magnitude.

Eight-ring hairpin polyamides which differ only by the linear arrangement of Py and Im amino acids were designed for recognition of six base pair DNA sequences containing four contiguous G,C base pairs. The respective DNA binding properties of three polyamides, ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp, ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp, and ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp, were analyzed by footprinting and affinity cleavage on a DNA fragment containing the respective match sites 5'-TGGCCA-3', 5'-TGCGCA-3', and 5'-TGGGGA-3'. Quantitative footprint titrations demonstrate that ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp binds the match site 5'-TGGCCA-3' with  $K_a = 1 \times 10^{10} M^{-1}$  and >250-fold specificity versus the mismatch sequences, 5'-TGCGCA-3' and 5'-TGGGGA-3'. The polyamides ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp and ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp recognize their respective 5'-TGCGCA-3' and 5'-TGGGGA-3' match sites with reduced affinity relative to ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp, but again with high specificity with regard to mismatch sites. These results expand the DNA sequence repertoire targeted by Py-Im polyamides

and identify sequence composition effects which will guide further second-generation polyamide design for DNA recognition.

Ten-ring hairpin polyamides were designed for recognition of 7-bp sequences in the minor groove of DNA. The DNA binding properties of two polyamides, ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp and ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp, were analyzed by footprinting and affinity cleavage on a DNA fragment containing the respective match sites 5'-TGTAACA-3' and 5'-TGGAACA-3'. Quantitative footprint titrations demonstrate that ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp binds the 7-bp match sequence 5'-TGTAACA-3' with  $K_a = 1.2 \times 10^{10} M^{-1}$  and 18-fold specificity versus the single base pair mismatch sequence 5'-TGGAACA-3'. ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp differs from ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp by a single amino acid substitution and binds its match 5'-TGGAACA-3' site with  $K_a = 3.6 \times 10^9 M^{-1}$  and 300-fold specificity versus its corresponding single base pair mismatch sequence 5'-TGTAACA-3'. Ten-ring hairpin polyamides have binding affinities similar to those of eight-ring hairpin polyamides. These results indicate that the affinity of hairpin binding ceases to increase as the length of the polyamide subunits increases beyond four rings, analogous to the behavior of unlinked subunits. Therefore, recognition of 7 base pairs by a ten-ring hairpin polyamide most likely represents an upper limit to the effective targetable site size of the hairpin polyamide-DNA motif.

Selective placement of an aliphatic  $\beta$ -alanine ( $\beta$ ) residue paired side-by-side with either a Py or Im aromatic amino acid is found to compensate for sequence composition effects for recognition of the minor groove of DNA by hairpin Py-Im polyamides. A series of polyamides were prepared which contain Py and Im aromatic amino acids, as well as  $\gamma$ -aminobutyric acid ( $\gamma$ ) "turn" and  $\beta$ -alanine "spring" aliphatic amino acid residues. The binding affinities and specificities of these polyamides are regulated by the placement of paired  $\beta/\beta$ , Py/ $\beta$  and Im/ $\beta$  residues. Quantitative footprint titrations demonstrate that replacing two Py/Py pairings in a 12-ring hairpin (6- $\gamma$ -6) with two Py/ $\beta$  pairings affords 10-fold enhanced affinity and similar

sequence specificity for an 8-bp target sequence. The 6- $\gamma$ -6 hairpin ImPyImPyPyPy- $\gamma$ -ImPyPyPyPyPy- $\beta$ -Dp, which contains six consecutive amino acid pairings, is unable to discriminate a single-base-pair mismatch site 5'-TGTTAACA-3' from a 5'-TGTGAACA-3' match site. The hairpin polyamide Im- $\beta$ -ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Py- $\beta$ -Dp binds to the 8-bp match sequence 5'-TGTGAACA-3' with an equilibrium association constant of  $K_a = 2.4 \times 10^{10} M^{-1}$  and  $> 48$ -fold specificity versus the 5'-TGTTAACA-3' single-base-pair mismatch site. Modeling indicates that the  $\beta$ -alanine residue relaxes ligand curvature, providing for optimal hydrogen bond formation between the floor of the minor groove and both Im-residues within the Im- $\beta$ -Im polyamide subunit. This observation provided the basis for design of a hairpin polyamide, Im- $\beta$ -ImPy- $\gamma$ -Im- $\beta$ -ImPy- $\beta$ -Dp, which incorporates Im/ $\beta$  pairings to recognize a "problematic" 5'-GCGC-3' sequence at subnanomolar concentrations. These results identify Im/ $\beta$  and  $\beta$ /Im pairings that respectively discriminate G•C and C•G from A•T/T•A as well as Py/ $\beta$  and  $\beta$ /Py pairings that discriminate A•T/T•A from G•C/C•G. These aliphatic/aromatic amino acid pairings will facilitate the design of hairpin polyamides which recognize both a larger binding site size as well as a more diverse sequence repertoire.

We report here a cooperative six-ring extended hairpin polyamide which dimerizes to specifically bind a predetermined ten base pair sequence. For example, we found that the eight-ring hairpin polyamide ImPy- $\beta$ -ImPyPy-(R)<sup>H<sub>2</sub>N</sup>- $\gamma$ -PyImPy-C<sub>2</sub>-OH binds the twelve base pair match site 5'-AAGCAGCTGCTT-3' with subnanomolar affinity, and 100-fold specificity for this site versus the double-base pair mismatch site 5'-CAGATGCTGCAT-3'. The DNA-binding affinity and specificity of the cooperative eight-ring extended-hairpin polyamide for its ten base pair binding site are typical of fully overlapped eight-ring hairpins which recognize six base pairs. Thus, use of the cooperative hairpin dimer motif doubles the binding site size relative to the standard hairpin motif without sacrificing affinity or specificity, and without increasing the molecular weight of the ligand. The results reported here show that by using a novel cooperative



*hairpin dimer motif, relatively low molecular weight pyrrole-imidazole polyamides (MW 950-1,200) can specifically recognize 10-12 base pairs of DNA. Such polyamides will be useful in determining the optimal ligand*

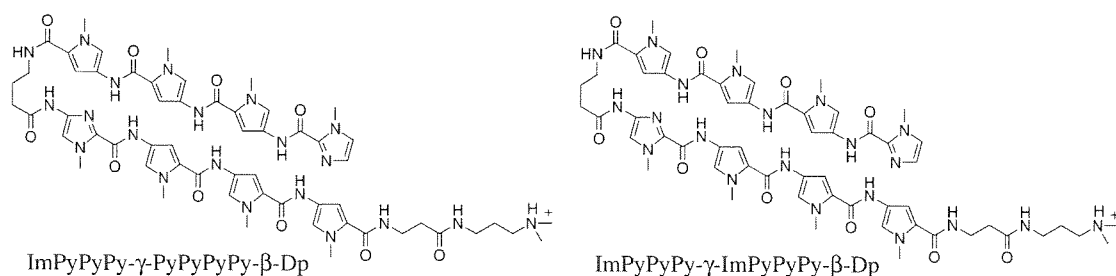
*size and optimal binding site size required for specific biological activity.*

*The broad range of sequences that can be specifically targeted with Py-Im polyamides, coupled with an efficient solid-phase synthesis methodology, identify a powerful class of small molecules for sequence-specific recognition of double-helical DNA.*

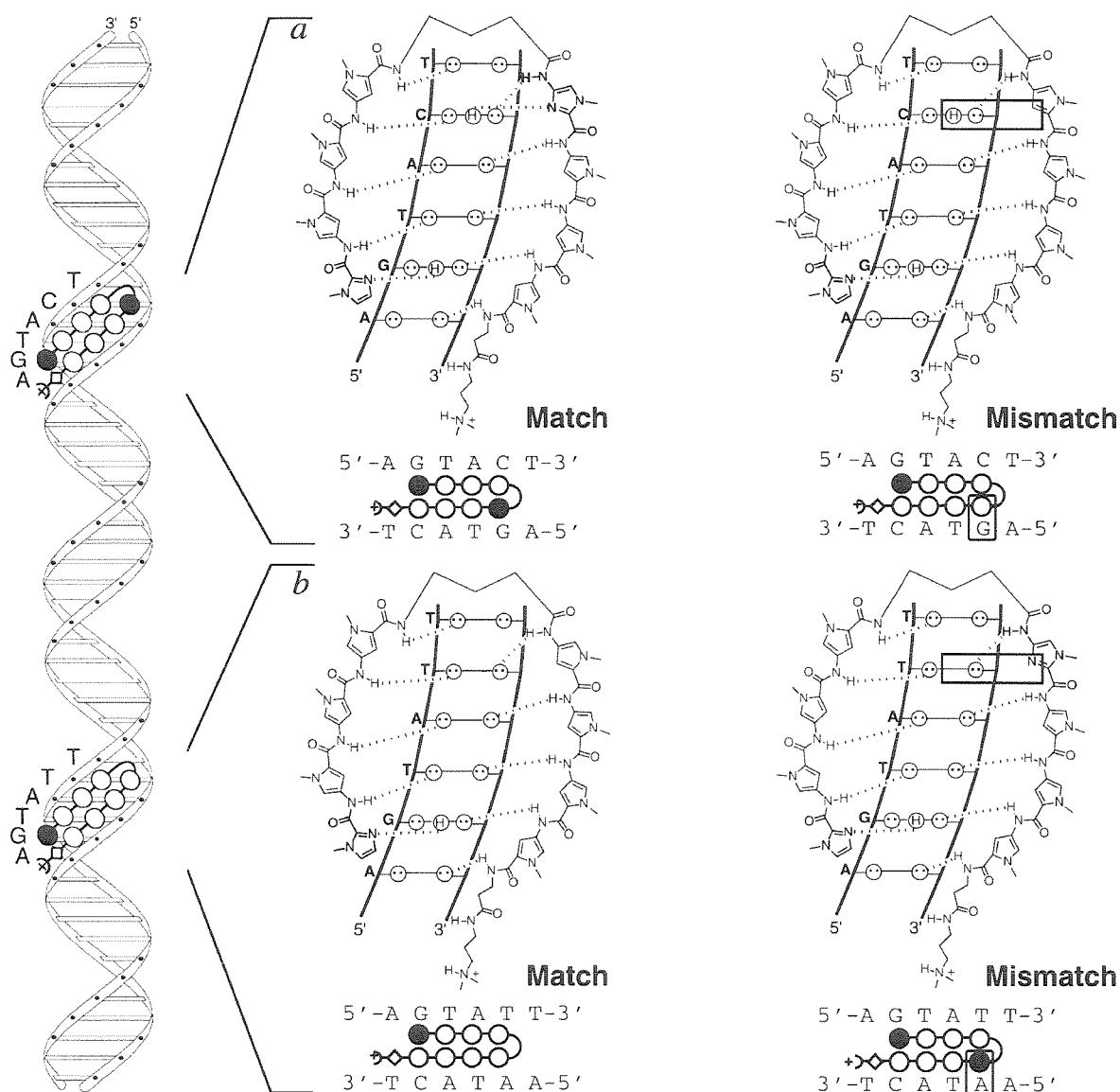
- Publications:** Trauger, Baird & Dervan *Nature* **1996**, 382, 559.  
Trauger, Baird & Dervan *Chem. & Biol.* **1996**, 3, 369.  
Swalley, Baird & Dervan *J. Am. Chem. Soc.* **1997**, 119, 6953.  
Turner, Baird & Dervan *J. Am. Chem. Soc.* **1997**, 119, 7636.  
Turner, Swalley, Baird & Dervan *J. Am. Chem. Soc.* **1998**, 120, 6219.  
Trauger, Baird & Dervan *Angew. Chemie. Int. Ed.* **1998**, 37, 1421.

For side-by-side complexes of Py-Im polyamides in the minor groove of DNA, the DNA-binding sequence-specificity depends on the sequence of amino acid pairings. A pairing of Im opposite Py (Im/Py pair) targets a G•C base-pair, while Py/Im targets a C•G base-pair.<sup>1</sup> A Py/Py combination is degenerate and targets both T•A and A•T base-pairs.<sup>1,2</sup> Specificity for G,C base-pairs results from the formation of a hydrogen bond between the imidazole N3 and the exocyclic amino group of guanine.<sup>3,4</sup> The generality of these pairing rules has been demonstrated by targeting a wide variety of sequences and is supported directly by several NMR and x-ray structure studies.<sup>1,4</sup>

In parallel with the elucidation of the scope and limitations of the pairing rules described above have been efforts to increase the DNA-binding affinity and specificity of pyrrole-imidazole polyamides by covalently linking polyamide subunits.<sup>5,6</sup> The polyamide ImPyPy- $\gamma$ -PyPyPy-Dp containing a “turn” amino acid  $\gamma$ -aminobutyric acid ( $\gamma$ ) was shown to specifically bind the designated target site 5'-TGTTA-3' in a “hairpin” conformation with an equilibrium association constant,  $K_a = 8 \times 10^7 \text{ M}^{-1}$ , an increase of 300-fold relative to unlinked three-ring polyamide dimers.<sup>6b</sup> A key issue was to determine if low molecular weight (M.W. = 1200) Py-Im polyamides could be constructed which would bind DNA at subnanomolar concentration without compromising sequence selectivity.



**Figure 6.1.** Structures of polyamides ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp (1) and ImPyPyPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp (2). The identity and purity of the polyamides was verified by <sup>1</sup>H NMR, MALDI-TOF MS, and analytical HPLC. MALDI-TOF MS: **1**, 1223.4 (1223.3 calculated for M+H); **2**, 1222.3 (1222.3 calculated for M+H).



**Figure 6.2.** Binding models for (a) 5'-AGTACT-3' in complex with polyamides **1** (match) and **2** (mismatch), and (b) 5'-AGTATT-3' in complex with polyamides **2** (match) and **1** (mismatch). Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines, and circles containing an H represent the N2 hydrogen of G. Putative hydrogen bonds are illustrated by dashed lines. Filled and unfilled circles represent Im and Py rings, respectively, the curved line represents  $\gamma$ , and the diamond represents  $\beta$ . Mismatches are highlighted.

<b>Table 6.1.</b> Equilibrium Association Constants ( $M^{-1}$ ).		
Binding Site	<b>1</b>	<b>2</b>
5'-ttAGTACTtg-3'	$3.7 \times 10^{10}$ (0.8)	$5.0 \times 10^8$ (0.5)
5'-ttAGTATTtg-3'	$4.1 \times 10^8$ (0.5)	$3.5 \times 10^9$ (0.8)

Association constants are average values obtained from three DNase I footprint titration experiments. The standard deviation for each data set is indicated in parentheses. Assays were carried out in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub> at pH 7.0 and 22°C. The six base-pair binding sites are in capital letters, with flanking sequences in lower-case letters.

We report the DNA-binding affinities of two eight-ring hairpin polyamides, ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp (**1**) and ImPyPyPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp (**2**), which differ by a single amino acid, for two 6 base pair (bp) target sites, 5'-AGTACT-3' and 5'-AGTATT-3', which differ by a single base pair. Based on the pairing rules for polyamide-DNA complexes, the sites 5'-AGTACA-3' and 5'-AGTATT-3' are for polyamide **1** “match” and “single base pair mismatch” sites, respectively, and for polyamide **2** “single base pair mismatch” and “match” sites, respectively (Figures 6.1 and 6.2).

Polyamides **1** and **2** were synthesized by solid phase methods and purified by reversed phase HPLC.<sup>7</sup> Equilibrium association constants for complexes of **1** and **2** with match and mismatch 6 base pair binding sites on a 3'-<sup>32</sup>P-labeled 229 bp restriction fragment were determined by quantitative DNase I footprint titration experiments (Table 6.1).<sup>8</sup> Polyamide **1** binds its match site 5'-AGTACT-3' at 0.03 nM concentration and its single base pair mismatch site 5'-AGTATT-3' with nearly 100-fold lower affinity. Polyamide **2** binds its designated match site 5'-AGTATT-3' at 0.3 nM concentration and its single base pair mismatch site 5'-AGTACT-3' with nearly 10-fold lower affinity. The specificity of **1** and **2** for their respective match sites results from very small structural changes (Figure 6.1). Replacing a single nitrogen atom in **1** with C-H (as in **2**) reduces the affinity of the polyamide•5'-AGTACT-3' complex by ~75-fold

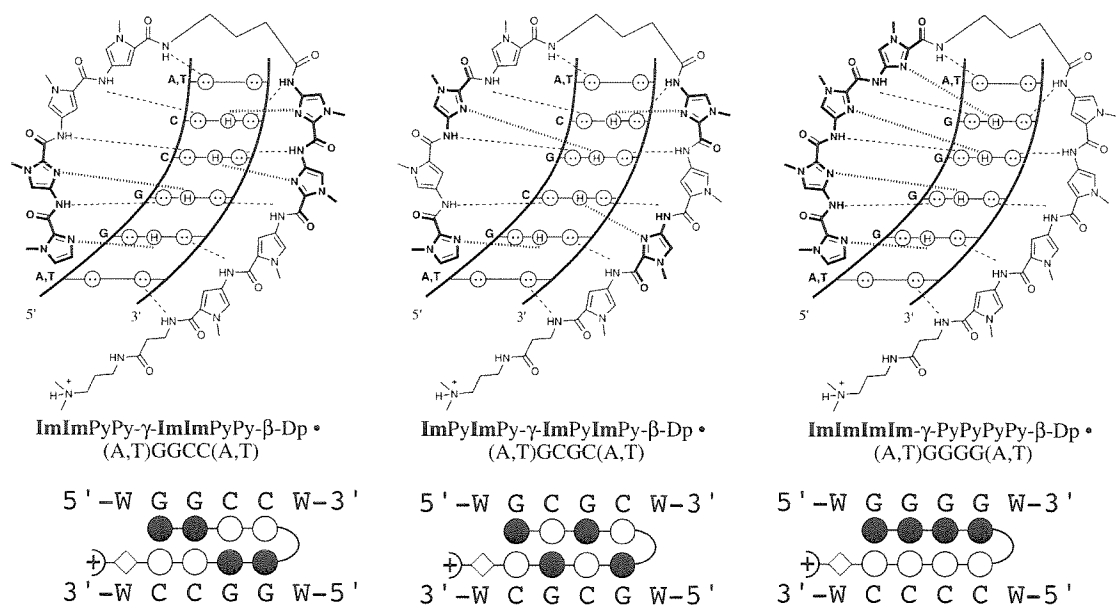
representing a free energy difference of  $\sim 2.5$  kcal/mole. Similarly, replacing a C-H in **2** with N (as in **1**) reduces the affinity of the polyamide•5'-AGTATT-3' complex  $\sim 10$ -fold, a loss in binding energy of  $\sim 1.3$  kcal/mol.

Crystal structures of protein-DNA complexes reveal that nature chose a combinatorial approach for specific DNA recognition. Although there are several highly conserved structural modules that bind DNA, no single motif exists which generates an amino acid-base pair code for all DNA sequences.<sup>9</sup> One encouraging success is the development of zinc finger libraries with different DNA binding specificities based on a one finger-three nucleotide code.<sup>10-15</sup> The three zinc finger DNA binding domain of the transcription factor Zif268 has a dissociation constant  $K_d = 0.5 - 3.0$  nM for its 9 bp binding site and is 2-12 fold specific over single base pair mismatches.<sup>11-15</sup> Similarly, the three zinc finger transcription factor Sp1 has a  $K_d = 0.5$  nM for its 9 bp recognition sequence and is 3-30 fold specific over single base pair mismatches.<sup>16</sup> Using a simple molecular shape and a two letter aromatic amino acid code, pyrrole-imidazole polyamides achieve affinities and specificities comparable to these DNA-binding proteins and, in addition, have the potential to be general for any desired DNA sequence. This nonbiological approach to DNA recognition could provide an underpinning for the design of cell-permeable molecules for the control of gene-specific regulation *in vivo*.

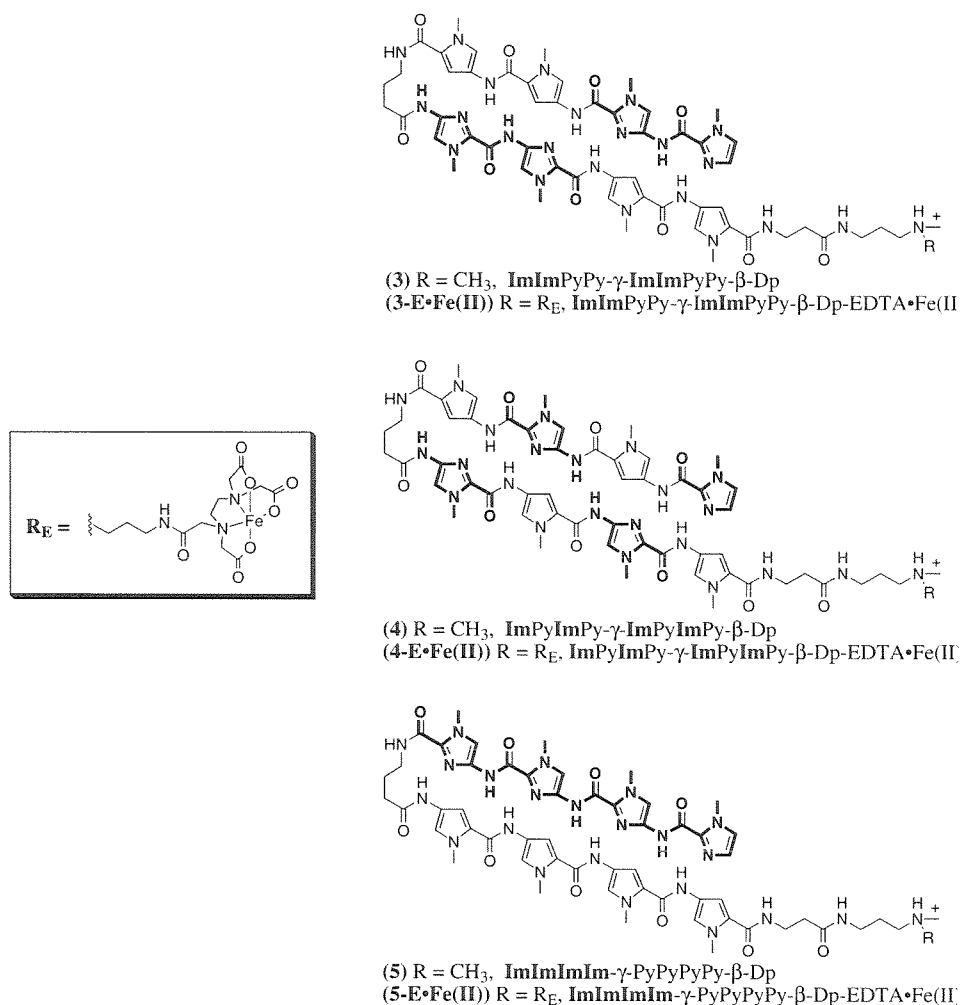
**Discrimination of 5'-GGGG-3', 5'-GCGC-3', and 5'-GGCC-3' Sequences in the Minor Groove of DNA by Eight-Ring Hairpin Polyamides.** Discrimination of G•C from C•G by the Im/Py pair requires precise positioning for the key hydrogen bond between the imidazole N3 and the exocyclic amine of guanine.<sup>4</sup> Given the structural requirement for binding the exocyclic amine groups of contiguous G,C base pairs, it remained to be determined if polyamides could be designed to recognize pure G,C sequences with very high affinity. Two four-ring polyamides, ImPyImPy-Dp and ImPyPyPy-Dp, which differ by a single amino acid substitution, bind as 2:1

homodimers to the respective match sites 5'-WGCGCW-3' ( $K_a \sim 1 \times 10^5 \text{ M}^{-1}$ ) and 5'-WGWWCW-3' ( $K_a = 9 \times 10^6 \text{ M}^{-1}$ ) with equilibrium association constants  $K_a$  that differ by a factor of  $\sim 90$  ( $W = A \cdot T$  or  $T \cdot A$ ).<sup>17,18</sup> One possible explanation for this difference is that the presence of four exocyclic amino groups in the minor groove presents a steric hindrance to deep polyamide binding, such that any polyamide designed for sequences containing four or more G•C base pairs will bind with low affinity.

As a minimum first step to high affinity binding of G,C sequences, C-N covalently coupled polyamide subunits, i.e., “hairpin polyamides” may be used. To examine whether a core sequence of purely G,C base pairs could be recognized with high affinity and specificity, three eight-ring hairpin polyamides containing the maximum four imidazole amino acids, ImImPyPy- $\gamma$ -



**Figure 6.3.** Binding models for the complexes formed between the DNA and the three isomeric eight-ring hairpin polyamides, with the imidazole amino acids in bold: (left) **ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp**, (middle) **ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp**, (right) **ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp**. (Top) Hydrogen bonding models of 1:1 polyamide:DNA recognition. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. (Bottom) Ball and stick models of hairpin-DNA interaction. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent the  $\beta$ -alanine residue. W represents either an A or T base.



**Figure 6.4.** Structures of the eight-ring hairpin polyamides **3-5** and the corresponding Fe(II)•EDTA affinity cleaving derivatives. The imidazole amino acids are in bold print.

ImImPyPy- $\beta$ -Dp (**3**), ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp (**4**), and ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp (**5**), were synthesized by solid phase methods (Figures 6.3 and 6.4).<sup>11</sup> The corresponding EDTA analogs, ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp-EDTA (**3-E**), ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp-EDTA (**4-E**), and ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp-EDTA (**5-E**), were also constructed to confirm the orientation of the hairpin motif.

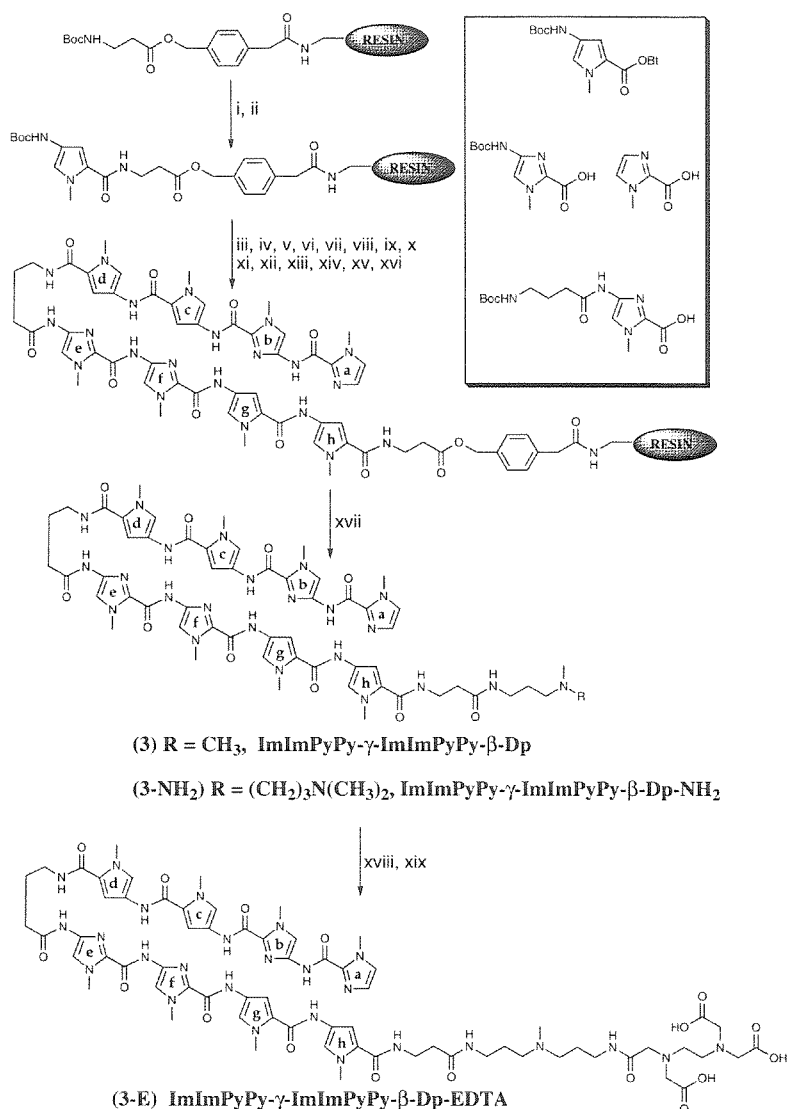
Specific hydrogen bonds are expected to form between each imidazole N3 and one of the four individual guanine 2-amino groups on the floor of the minor groove (Figure 6.3).<sup>3,4</sup> Covalent head-to-tail linkage of polyamide subunits by a  $\gamma$ -aminobutyric acid ( $\gamma$ ) linker to form "hairpin"

polyamides has been shown to increase affinity by >100-fold relative to the unlinked subunits.<sup>6</sup> The C-terminal  $\beta$ -alanine ( $\beta$ ) residue increases both the affinity and specificity of the hairpin motif and facilitates solid phase synthesis.<sup>6b</sup> We report here the affinities, binding orientations, and relative selectivities of the three polyamides as determined by three separate techniques: MPE•Fe(II) footprinting,<sup>19</sup> affinity cleaving,<sup>20</sup> and DNase I footprinting.<sup>8</sup> Information about binding site size and location is provided by MPE•Fe(II) footprinting, while binding orientations are demonstrated by affinity cleaving. Quantitative DNase I footprint titrations allow the determination of equilibrium association constants ( $K_a$ ) of the polyamides for their respective match sites.

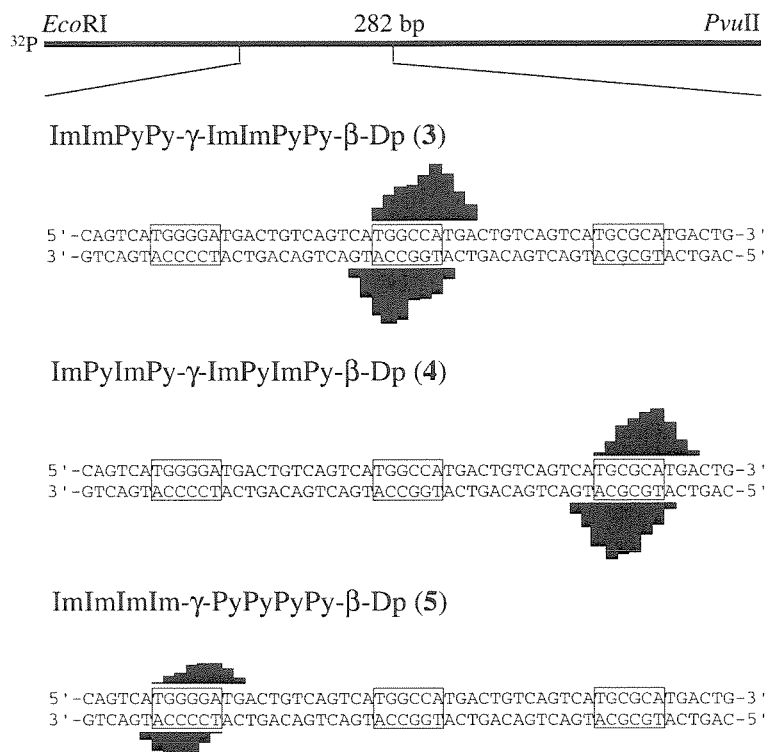
The polyamides ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp (**3**), ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp (**4**), and ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp (**5**) were synthesized in a stepwise manner from Boc- $\beta$ -alanine-Pam resin (1 g resin/0.2 mmol/g substitution) using Boc-chemistry machine-assisted protocols in 16, 12, and 18 steps, respectively (Figure 6.5).<sup>7</sup> The polyamides ImImPyPy- $\beta$ -Dp (**6**) and ImPyImPy- $\beta$ -Dp (**7**) were synthesized by manual solid phase methods in 8 and 6 steps respectively. The  $\gamma$ -Im and PyIm subunits were introduced to polyamides **1**, **2** and **5** as dimer-blocks in order to avoid the slow coupling of  $\gamma$  or Py to Im. A sample of resin (240 mg) was then cleaved by a single-step aminolysis reaction with ((dimethylamino)propyl)amine (55°C, 18 h) and subsequently purified by reverse phase HPLC chromatography to provide ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp (**3**) (26 mg), ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp (**4**) (19 mg), and ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp (12 mg) (**5**), ImImPyPy- $\beta$ -Dp (**6**) (14 mg), and ImPyImPy- $\beta$ -Dp (**7**) (5 mg).

Reaction of a sample of resin (240 mg) with the symmetric triamine 3,3'-diamino-*N*-methyldipropylamine (55°C, 18 h) and subsequent purification by reverse phase HPLC chromatography provides polyamides modified with the trifluoroacetate salt of a free aliphatic primary amine group suitable for post-synthetic modification.





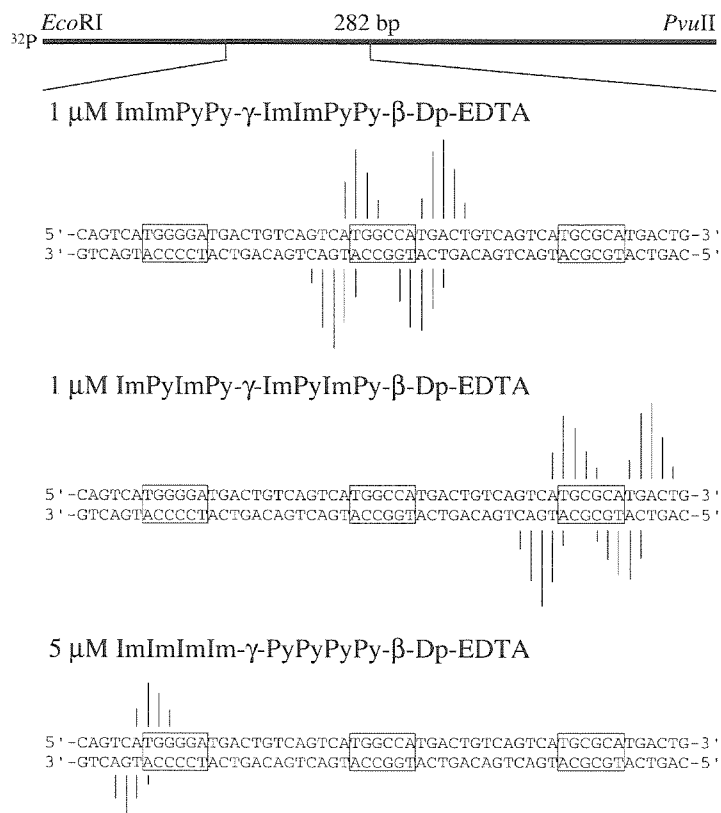
**Figure 6.5.** Solid phase synthetic scheme exemplified for ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp, ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp-NH<sub>2</sub>, and ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp-EDTA starting from commercially available Boc- $\beta$ -Pam-resin: (i) 80% TFA/DCM, 0.4 M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4 M PhSH; (iv) Boc-Py-OBt, DIEA, DMF; (v) 80% TFA/DCM, 0.4 M PhSH; (vi) Boc-Im-OBt (DCC, HOBt), DIEA, DMF; (vii) 80% TFA/DCM, 0.4 M PhSH; (viii) Boc- $\gamma$ -Im-OBt (HBTU, DIEA); (ix) 80% TFA/DCM, 0.4 M PhSH; (x) Boc-Py-OBt, DIEA, DMF; (xi) 80% TFA/DCM, 0.4 M PhSH; (xii) Boc-Py-OBt, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4 M PhSH; (xiv) Boc-Im-OBt (DCC, HOBt), DIEA, DMF; (xv) 80% TFA/DCM, 0.4 M PhSH; (xvi) Im-OBt (HBTU, DIEA), DMF; (xvii) *N,N*-dimethylaminopropylamine (55°C, 18 h) for **3**; 3,3'-diamino-*N*-methyldipropylamine (55°C, 18h) for **3-NH<sub>2</sub>**; (xviii) EDTA-dianhydride, DMSO, NMP, DIEA (55°C, 1 h); (xix) 0.1M NaOH, (55°C, 30 min). (Inset) Pyrrole and Imidazole monomers for solid phase synthesis: Boc-Pyrrole-OBt ester (Boc-Py-OBt), Boc-Imidazole-acid (Boc-Im-OH), Imidazole-2-Carboxylic acid (Im-OH),<sup>1a</sup> Boc- $\gamma$ -Imidazole acid (Boc- $\gamma$ -Im-OH).



**Figure 6.6.** MPE•Fe(II) protection patterns for ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp, ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp and ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp at 10  $\mu$ M concentration. (Top) Illustration of the 282 bp restriction fragment with the position of the sequence indicated. Bar heights are proportional to the relative protection from cleavage at each band.

The polyamide amines ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp-NH<sub>2</sub> (**3-NH<sub>2</sub>**), ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp-NH<sub>2</sub> (**4-NH<sub>2</sub>**), and ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (**5-NH<sub>2</sub>**) were then treated with an excess of the dianhydride of EDTA (DMSO/NMP, DIEA, 55°C) and the remaining anhydride hydrolyzed (0.1 M NaOH, 55°C). The respective EDTA modified polyamides, ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp-EDTA (**3-E**), ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp-EDTA (**4-E**), and ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp-EDTA (**5-E**), were then isolated by reversed phase HPLC. Eight-ring hairpin polyamides are stable for storage at room temperature as the trifluoroacetate salt and can be synthesized in sufficient quantity (20-200 mg) and purity for biological applications.<sup>21</sup>

MPE•Fe(II) footprinting (25 mM Tris-acetate, 10 mM NaCl, 100  $\mu$ M/base pair calf thymus DNA, pH 7.0 and 22°C)<sup>19</sup> on the 3'- and 5'-<sup>32</sup>P end-labeled 282 base pair *Eco*RI/*Pvu*II



**Figure 6.7.** Affinity cleavage patterns for ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp-EDTA and ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp-EDTA at 1  $\mu$ M concentration, and ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp-EDTA at 5  $\mu$ M concentration. (Top) Illustration of the 282 bp restriction fragment with the position of the sequence indicated. Bar heights are proportional to the relative cleavage at each band.

restriction fragment from the cloned plasmid pSES9hp (Figure 6.6) reveals that polyamides **3-5** are protecting their targeted match sequences, 5'-TGGCCA-3', 5'-TGCGCA-3', and 5'-TGGGGA-3', respectively. Polyamide **3** also recognizes two singly-mismatched sequences, 5'-TGGCGT-3' and 5'-TGGTCA-3'. The footprinting patterns for the eight-ring hairpin polyamides are consistent with binding sites that are six base pairs in length.

Affinity cleavage assays (25 mM Tris-acetate, 10 mM NaCl, 100  $\mu$ M/base pair calf thymus DNA, pH 7.0 and 22°C)<sup>20</sup> were performed in order to identify the binding orientations of the EDTA analogues of the three hairpin polyamides: ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp-EDTA (**3-E**), ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp-EDTA (**4-E**), and ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp-EDTA (**5-E**)

**Table 6.2** Equilibrium Association Constants ( $M^{-1}$ ) <sup>a,b,c</sup>

Polyamide	5'-TGGCCA-3'	5'-TGCGCA-3'	5'-TGGGGA-3'
ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp (3)	<b><math>9.7 \times 10^9</math> (1.3)</b>	$< 2 \times 10^7$	$< 2 \times 10^7$
ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp (4)	$< 10^7$	<b><math>3.7 \times 10^7</math> (0.9)</b>	$< 10^7$
ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp (5)	$< 5 \times 10^6$	$< 5 \times 10^6$	<b><math>2.8 \times 10^7</math> (0.2)</b>
ImImPyPy- $\beta$ -Dp (6)	<b><math>2.4 \times 10^7</math> (0.6)</b>	$< 10^5$	$< 10^5$
ImPyImPy- $\beta$ -Dp (7)	$< 2 \times 10^5$	<b><math>&lt; 2 \times 10^5</math></b>	$< 2 \times 10^5$

<sup>a</sup>Values reported are the mean values measured from at least three DNase I footprint titration experiments, with the standard deviation for each data set indicated in parentheses. <sup>b</sup>The assays were performed at 22 °C at pH 7.0 in the presence of 10 mM tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. <sup>c</sup>The data in bold type represents the formal match site for each compound, as described by the model.

(Figure 6.7). Polyamides **3-E** and **4-E** recognize their respective palindromic match sequences, 5'-TGGCCA-3' and 5'-TGCGCA-3', in two equivalent orientations, consistent with hairpin formation. In contrast, polyamide **5-E** recognizes a non-palindromic sequence, 5'-TGGGGA-3', in a single orientation with cleavage visible only on the 5'-side of the site, as predicted by the hairpin model.

Quantitative DNase I footprint titrations (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub> and 5 mM CaCl<sub>2</sub>, pH 7.0 and 22°C) were performed to determine the equilibrium association constant ( $K_a$ ) of each polyamide to its respective match sequence (Table 6.2). Also, equilibrium association constants were determined for ImImPyPy- $\beta$ -Dp (6), the unlinked homodimer of polyamide **3**, as well as for ImPyImPy- $\beta$ -Dp (7), the unlinked homodimer of polyamide **4**. This allows a direct comparison between the linked and unlinked polyamides for identical DNA sequences. ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp recognizes its match site 5'-caTGGCCAtg-3' with an equilibrium association constant of  $K_a = 9.7 \times 10^9 M^{-1}$ , an increase of 400-fold over the unlinked

ImImPyPy- $\beta$ -Dp ( $K_a = 2.4 \times 10^7 \text{ M}^{-1}$ ). The single base pair mismatch sites identified by MPE•Fe(II) footprinting, 5'-TGGCGT-3' and 5'-TGGTCA-3', are positioned too close together to allow accurate quantitation of each individual site. Neither site is bound by ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp with an equilibrium association constant greater than  $\sim 4 \times 10^8 \text{ M}^{-1}$ , indicating at least 25-fold specificity towards these two single base pair mismatch sequences.

In contrast to polyamide **3**, both ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp and ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp bind their respective match sites with much lower affinity: the former recognizes the site 5'-caTGCGCAtg-3' with a  $K_a = 3.7 \times 10^7 \text{ M}^{-1}$  (an increase of 190-fold over the unlinked ImPyImPy- $\beta$ -Dp), while the latter binds the site 5'-caTGGGGATg-3' with a  $K_a = 2.8 \times 10^7 \text{ M}^{-1}$ . The hairpin polyamides **3** and **4** are bound with approximately two orders of magnitude higher affinity relative to the unlinked homodimer polyamides **6** and **7**, consistent with results from earlier systems.<sup>17,18</sup> Each of the three hairpin polyamides is specific for its individual target sequence. Polyamide **3** recognizes the two designed double mismatch sequences 5'-caTGCGCAtg-3' and 5'-caTGGGGATg-3' with 250-fold and 350-fold lower affinity, respectively, compared to its recognition of the match site 5'-caTGGCCATg-3'. Similarly, polyamide **4** is at least 4-fold specific for its match site compared to the double base pair mismatch sites, and polyamide **5** is at least 6-fold specific.

It was thought that pyrrole-imidazole polyamides, which sit deeply in the minor groove of DNA, would recognize G,C rich sequences with low binding affinity due to steric hinderance with the exocyclic amines of the guanine bases. It has also been noted that the lower negative electrostatic potential of a G,C rich minor groove relative to an A,T rich minor groove might prohibit high affinity binding.<sup>22</sup> Quantitative DNase I footprint titrations reveal that the eight-ring hairpin polyamide ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp binds with subnanomolar affinity ( $K_a = 1 \times 10^{10} \text{ M}^{-1}$ ) to a 5'-TGGCCA-3' sequence, an affinity that equals those seen for eight-ring hairpin polyamides that recognize six base pair sequences containing only one or two G,C base pairs.<sup>23</sup>

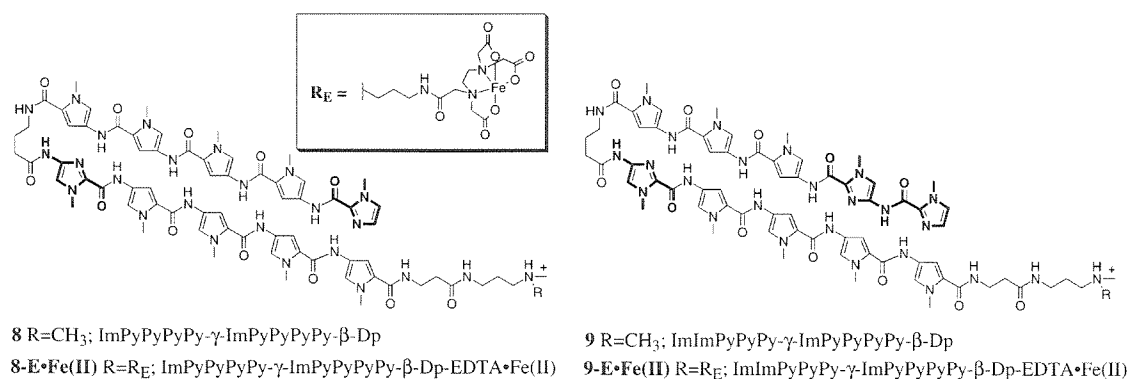
Therefore, it seems likely that neither the electrostatic properties nor the steric bulk of the four exocyclic guanine amino groups interfere with polyamide binding at G,C rich sites.

Two isomeric polyamides, differing only in the ordering of the Im residues, ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp and ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp, bind their match sequences with far lower affinity than does ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp. Sequence-dependent DNA structural features, such as intrinsic minor groove width, minor groove flexibility, and inherent curvature, may differ between the three binding sites and could contribute to the range of binding affinities.<sup>24,25</sup> However, the observed binding affinities do *not* correlate with the number of purine-pyrimidine or purine-purine steps, suggesting that the *positions* of the Im amino acids are critical to high affinity DNA recognition. One possible explanation is that the Im residues located at the C-terminal end of each four-ring polyamide subunit are out of register with the DNA helix.

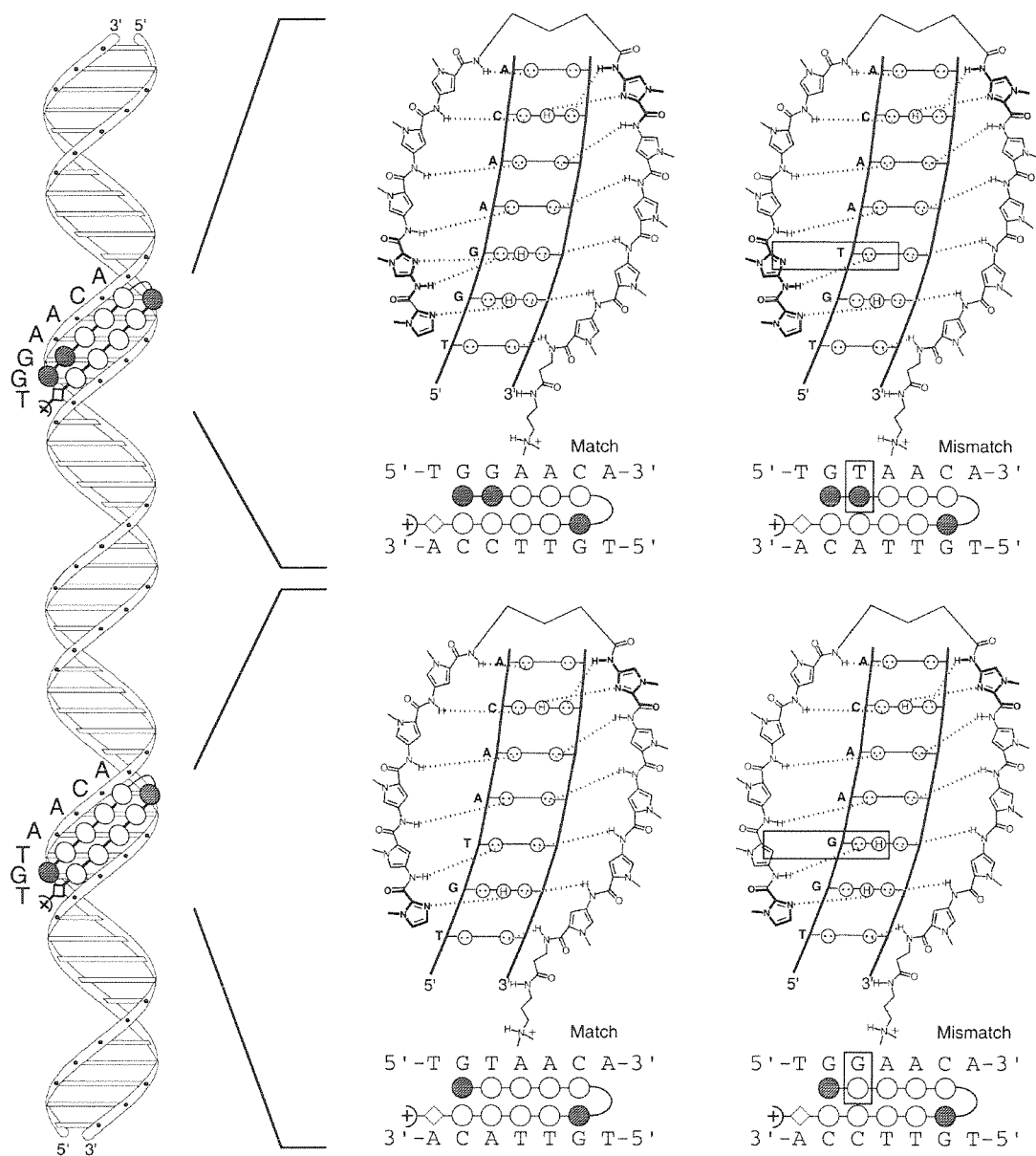
Pyrrole-imidazole polyamides provide a versatile chemical method for the targeting of any predetermined DNA sequence. However, prior to this study, subnanomolar recognition of sequences containing *only* G,C base pairs in the core sequence had not been demonstrated. Here, an eight-ring hairpin polyamide is used to recognize a 5'-TGGCCA-3' sequence with subnanomolar affinity. This work represents an important step in the understanding of polyamide-DNA recognition, since it demonstrates that the positioning of the Im amino acids has a profound effect on the binding affinities of pyrrole-imidazole polyamides. If the register of the ligand is responsible for these position-dependent effects, then perhaps binding affinity can be restored by the design of hairpin polyamides where specific ring residues have been substituted by more flexible spacer amino acids.

**Recognition of 7 Base Pair Sequences in the Minor Groove of DNA by Ten-Ring Pyrrole-Imidazole Polyamide Hairpins.** Four-ring polyamides covalently coupled to form eight-ring hairpin structures bind specifically to 6-bp target sequences and, importantly, have been shown to be cell permeable and to inhibit transcription of specific genes in cell culture.<sup>21</sup> The optimal binding site size for effective biological regulation has yet to be determined. This provides impetus to explore the binding site size limitations of the hairpin motif for recognition in the minor groove of DNA. Given the failure of more than five contiguous rings to maintain precise register with the DNA helix<sup>18</sup> combined with the constraint of covalent coupling, it remained to be determined if ten-ring hairpin polyamides could be successfully designed to recognize 7-bp sequences without compromising DNA-binding affinity or sequence specificity.

To expand the targetable binding site size and sequence repertoire of the hairpin polyamide motif, two polyamides containing either two or three Im amino acid residues, ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (**8**) and ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (**9**), were synthesized by solid phase methods (Figure 6.8).<sup>9</sup> The corresponding EDTA analogs ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-EDTA (**8-E**) and ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-EDTA (**9-E**) were also constructed to confirm the orientation of the hairpin motif at each binding site.



**Figure 6.8.** Structures of the ten-ring polyamides and their EDTA derivatives synthesized by solid phase methods.

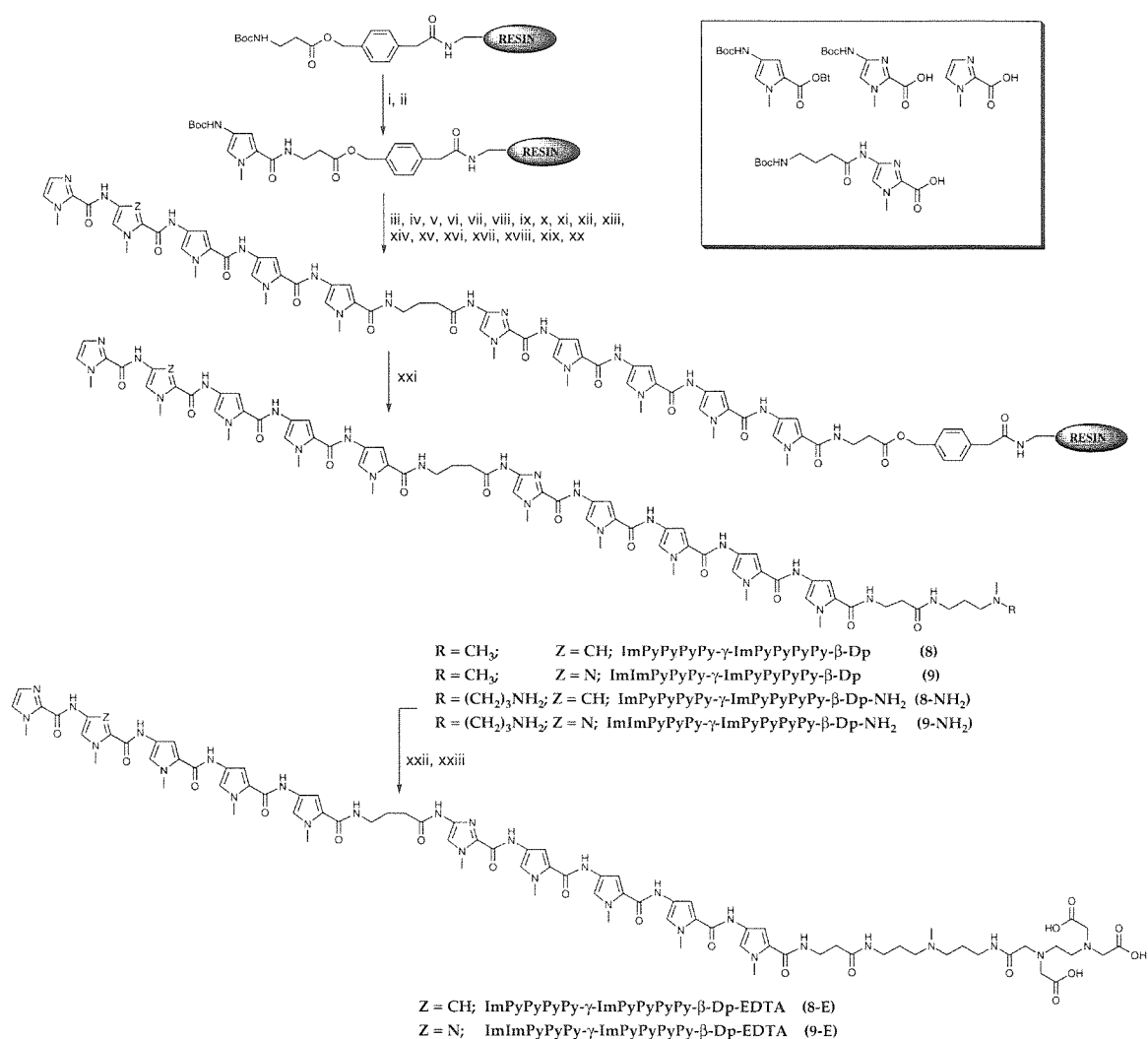


**Figure 6.9.** Binding model for the complexes formed between the DNA and either ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (top) or ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (bottom). Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. Ball and stick models are also shown. Shaded and nonshaded circles denote Im and Py carboxamides, respectively. Nonshaded diamonds represent the  $\beta$ -alanine residue.

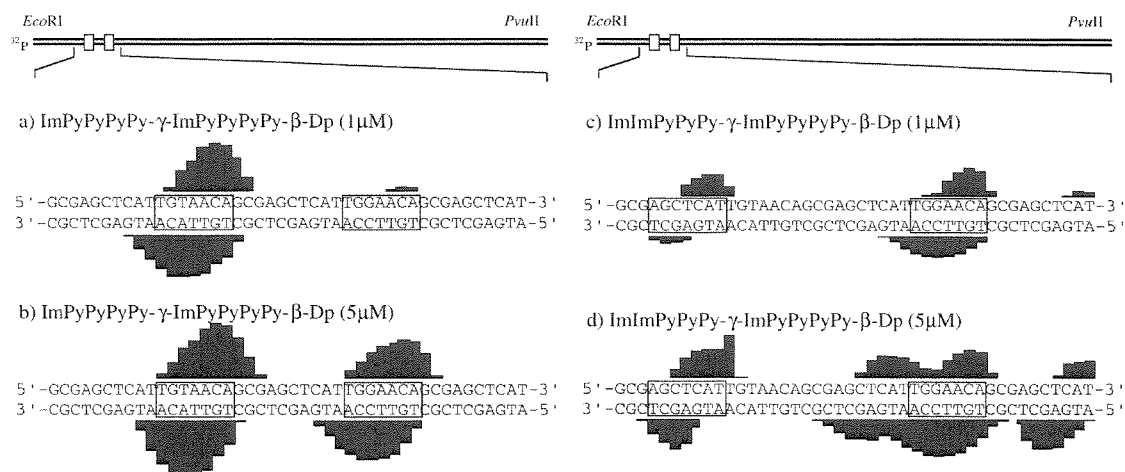


We report here the DNA-binding affinity, orientation, and sequence selectivity of two ten-ring hairpin polyamides, ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (**8**) and ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (**9**), for two respective 7-bp target sequences, 5'-TGTAACA-3' and 5'-TGGAACA-3' (Figure 6.9). Three separate techniques are used to characterize the DNA-binding properties of the designed polyamides: Methidiumpropyl-EDTA•Fe(II) (MPE•Fe(II)) footprinting, affinity cleaving, and DNase I footprinting.

The polyamides ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (**8**) and ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (**9**) were synthesized in a stepwise manner from Boc- $\beta$ -alanine-Pam resin (1 g resin, 0.2 mmol/g substitution) using Boc-chemistry machine-assisted protocols in 21 steps (Figure 6.10). The  $\gamma$ -Im subunit was introduced to both polyamides as a dimer-block in order to avoid the slow coupling of  $\gamma$  to Im. A sample of resin (240 mg) was then cleaved by a single-step aminolysis reaction with ((dimethylamino)propyl)amine (55°C, 18 h) and subsequently purified by reversed phase HPLC to provide ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (**8**, 13 mg) and ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (**9**, 11 mg). For the synthesis of analogs modified with EDTA, a sample of resin was cleaved with 3,3'-diamino-*N*-methyldipropylamine (55°C) and purified by reversed phase HPLC to provide either ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (**8**-NH<sub>2</sub>, 31 mg) or ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (**9**-NH<sub>2</sub>, 29 mg). **8**-NH<sub>2</sub> and **9**-NH<sub>2</sub> afford a primary amine group suitable for post-synthetic modification. The polyamide amine was treated with an excess of EDTA-dianhydride (DMSO/NMP, DIEA, 55°C, 15 min) and the remaining anhydride hydrolyzed (0.1 M NaOH, 55°C, 10 min). The EDTA modified polyamides ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-EDTA (**8**-E, 5 mg) and ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-EDTA (**9**-E, 6 mg) were then isolated by HPLC chromatography. The synthesis of **8**, **8**-NH<sub>2</sub>, **8**-E, **9**, **9**-NH<sub>2</sub>, and **9**-E is outlined in Figure 6.10.



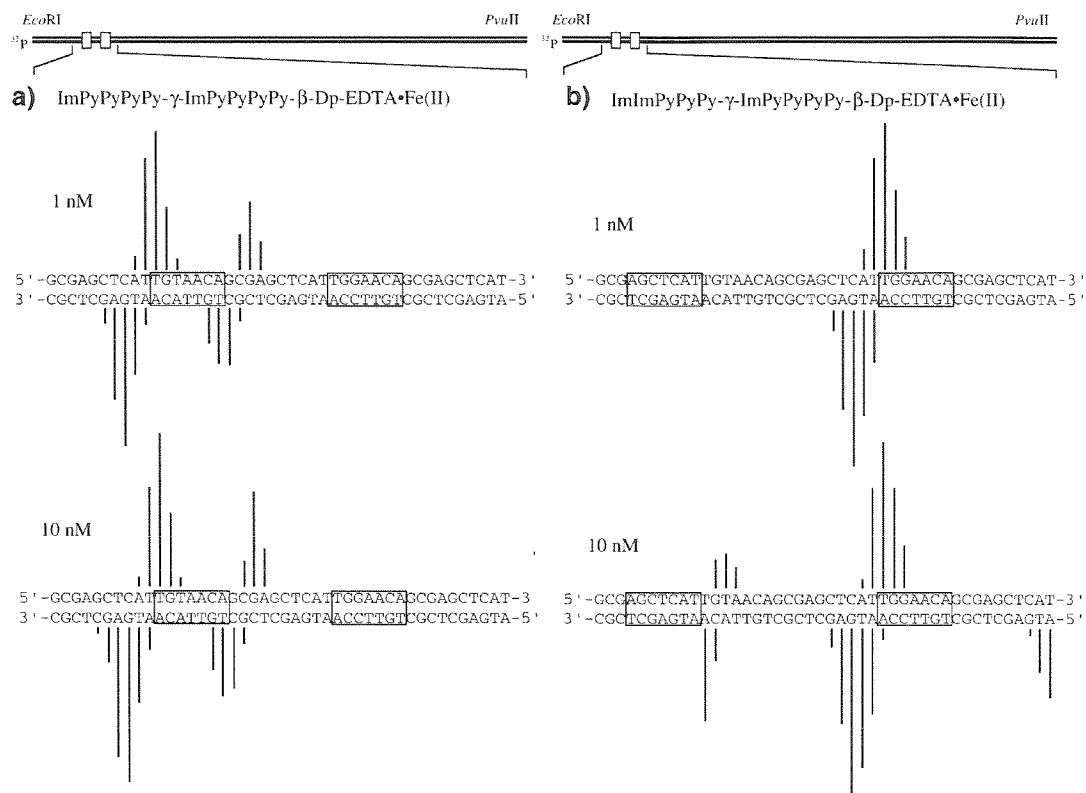
**Figure 6.10.** (Box) Py and Im monomers for synthesis of all compounds described here; Boc-Pyrrole-OBt ester, Boc-Imidazole-acid, imidazole-2-carboxylic acid,<sup>1a</sup> and Boc- $\gamma$ -Imidazole-acid. Solid phase synthetic scheme for ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp, ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp, ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-NH<sub>2</sub>, ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-NH<sub>2</sub>, ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-EDTA and ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-EDTA prepared from commercially available Boc- $\beta$ -alanine-Pam-resin (0.2 mmol/gram): (i) 80% TFA/DCM, 0.4M PhSH; (ii) BocPy-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4M PhSH; (iv) BocPy-OBt, DIEA, DMF; (v) 80% TFA/DCM, 0.4M PhSH; (vi) BocPy-OBt, DIEA, DMF; (vii) 80% TFA/DCM, 0.4M PhSH; (viii) BocPy-OBt, DIEA, DMF; (ix) 80% TFA/DCM, 0.4M PhSH; (x) Boc- $\gamma$ -Im-OOH, (HBTU, DIEA), DMF; (xi) 80% TFA/DCM, 0.4M PhSH; (xii) BocPy-OBt, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4M PhSH; (xiv) BocPy-OBt, DIEA, DMF; (xv) 80% TFA/DCM, 0.4M PhSH; (xvi) BocPy-OBt, DIEA, DMF; (xvii) 80% TFA/DCM, 0.4M PhSH; (xviii) BocPy-OBt, DIEA, DMF for **1**, Boc-Im-COOH (DCC, HOBt) for **2**; (xix) 80% TFA/DCM, 0.4M PhSH; (xx) imidazole-2-carboxylic acid (HBTU/DIEA); (xxi) *N,N*-dimethylaminopropylamine for **8** or **9**, or 3,3'-diamino-*N*-methyldipropylamine for **8-NH<sub>2</sub>** or **9-NH<sub>2</sub>**, 55°C; (xxii) EDTA-dianhydride, DMSO/NMP, DIEA, 55°C; (xxiii) 0.1M NaOH.



**Figure 6.11.** Results from MPE•Fe(II) footprinting of ImPyPyPyPy-γ-ImPyPyPyPy-β-Dp and ImImPyPyPy-γ-ImPyPyPyPy-β-Dp. (Top) Illustration of the 252 bp restriction fragment with the position of the sequence indicated. Boxes represent equilibrium binding sites determined by the published model. Only sites that were quantitated by DNase I footprint titrations are boxed. (a) and (c): MPE•Fe(II) protection patterns for polyamides at 1 μM concentration. (b) and (d): MPE•Fe(II) protection patterns for polyamides at 5 μM concentration. Bar heights are proportional to the relative protection from cleavage at each band.

MPE•Fe(II) footprinting on the 3'- and 5'-<sup>32</sup>P end-labeled 252 base pair *Eco*RI/*Pvu*II restriction fragment from the plasmid pJK7<sup>18</sup> (25 mM Tris-acetate, 10 mM NaCl, 100 μM calf thymus DNA, pH 7.0, 22°C) reveals that the polyamides, each at 1 μM concentration, are binding to their designated match sites (Figures 6.11). Mismatch sites are observed at higher polyamide concentration.

Footprinting patterns reveal asymmetrically 3'-shifted protection of 7-bp sites (Figure 6.11). MPE•Fe(II) protection patterns are consistent with formation of a 1:1 hairpin polyamide-DNA complex in the minor groove. Polyamide 8 at 5 μM concentration protects both the cognate 5'-TGTAACA-3' site and the single-base-pair mismatch 5'-TGGAACA-3' site. At 5 μM concentration, polyamide 9 protects its cognate 5'-TGGAACA-3' site; however, no protection is observed at the 5'-TGTAACA-3' single base pair mismatch site. Instead polyamide 9 is found to

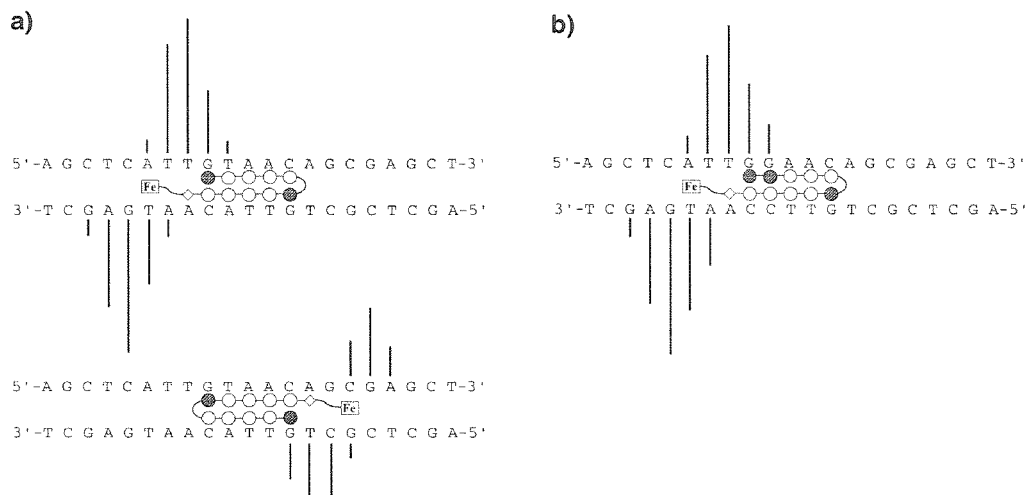


**Figure 6.12.** Affinity cleavage patterns of: (a) ImPyPyPyPy-γ-ImPyPyPyPy-β-Dp-EDTA•Fe(II), and (b) ImImPyPyPy-γ-ImPyPyPyPy-β-Dp-EDTA•Fe(II) at 1 nM and at 10 nM. Bar heights are proportional to the relative cleavage intensities at each base pair.

preferentially protect a 5'-AGCTCAT-3' double base pair mismatch site present in the repeated 10-bp sequence which flanks each of the designated binding sites.

Affinity cleavage experiments were performed on a 3'- and 5'-<sup>32</sup>P end-labeled EcoRI/PvuII restriction fragment from the plasmid pJK7 (20 mM HEPES, 200 mM NaCl, 50 μg/mL glycogen, pH 7.0, 22°C) (Figures 6.12). Affinity cleavage experiments were performed with analogs of polyamides **8** and **9** modified with an EDTA•Fe(II) moiety at the carboxy terminus. The observed cleavage patterns are in all cases 3' shifted, consistent with location of the polyamide in the minor groove.<sup>11</sup>

Affinity cleavage patterns reveal that at 1 nM and 10 nM concentrations, the homodimeric polyamide **8-E** binds its symmetrical target site 5'-TGTAACA-3' in two distinct orientations, as expected (Figure 6.12). Despite the symmetry of the 7-bp target site around the



**Figure 6.13.** Affinity cleavage patterns and ball and stick models of the ten-ring EDTA•Fe(II) analogs **8-E** (a) and **9-E** (b) bound to their formal match sites. Bar heights are proportional to the relative cleavage intensities at each base pair. Shaded and nonshaded circles denote imidazole and pyrrole carboxamides, respectively. Nonshaded diamonds represent the  $\beta$ -alanine residue. The boxed **Fe** denotes the EDTA•Fe(II) cleavage moiety. (a): symmetric binding to a formal match sequence by ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-EDTA•Fe(II) at 1 nM. (b): Single orientation binding to a formal match sequence by ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-EDTA•Fe(II) at 1 nM.

central A•T base pair, an apparent 2-fold orientational preference is observed, indicating that the sequences flanking the 7-bp site affect complex stability.

The orientation preference is likely due to sequence-dependent interactions of the polyamide tail which extends beyond the 7-bp binding site. Binding of **8-E** with the tail at the 5'-side of the 5'-catTGTAACAgcg-3' binding site places the C-terminus approximately opposite a T•A base pair, whereas binding with the tail of the polyamide opposite the 3'-side of the binding site places the C-terminal  $\beta$ -alanine residue opposite a G•C base pair. The observed orientation preference may result from a steric interaction between the C-terminal  $\beta$ -alanine residue and the exocyclic 2-amino group of guanine. It is unlikely that the observed orientation preference results from a specific interaction with the central A•T base pair of the binding site.<sup>3d</sup>

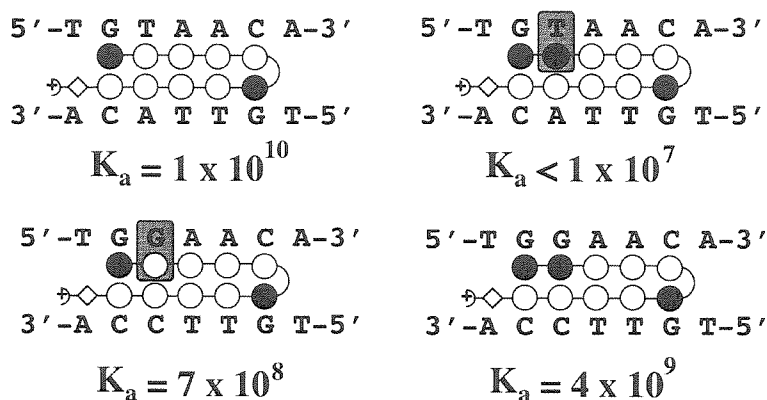
In the presence of 1 nM **9-E** a single cleavage locus is observed proximal to the 5'-side of the binding site for the heterodimeric polyamide **9-E** binding its designated target site 5'-

TGGAACA-3, indicating that the terminus of the polyamide is located at the 5'-side of the binding site. The observation of a single cleavage locus is consistent only with an oriented 1:1 complex and rules out any 2:1 overlapped or extended binding motifs.<sup>14</sup> In the presence of 10 nM polyamide **2-E** additional cleavage loci are revealed which result from polyamide binding in the 10-bp intervening sequence between binding sites. The observed cleavage patterns combined with the MPE•Fe(II) footprinting results described above allow putative assignment of the mismatch polyamide binding site as the double base pair mismatch site 5'-AGCTCAT-3'.

Quantitative DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub> and 5 mM CaCl<sub>2</sub>, pH 7.0, 22°C) were performed to determine the equilibrium association constants of the polyamides for the two bound sites (Table 6.3). ImPyPyPyPy-γ-ImPyPyPyPy-β-Dp binds its match site 5'-TGTAACA-3' with  $K_a = 1.2 \times 10^{10} \text{ M}^{-1}$ . The sequence 5'-TGGAACA-3' is bound with 18-fold lower affinity ( $K_a = 6.8 \times 10^8 \text{ M}^{-1}$ ) (Table 6.3). ImImPyPyPy-γ-ImPyPyPyPy-β-Dp binds its 5'-TGGAACA-3' match site with an equilibrium association constant of  $K_a = 3.6 \times 10^9 \text{ M}^{-1}$ . The internal mismatch 5'-TGTAACA-3' site is bound with at least 300-fold lower affinity ( $K_a < 1 \times 10^7 \text{ M}^{-1}$ ). In addition to the designated match and mismatch sites, polyamide **2** is observed to bind the unexpected 5'-AGCTCAT-3' mismatch site identified by affinity cleavage and MPE•Fe(II) footprinting experiments described above with  $K_a = 7.5 \times 10^8 \text{ M}^{-1}$ . The reduced overall specificity and binding affinity of polyamide **9** relative to **8** may result from the presence of a 5'-GA-3' step in the designated target site. A 5'-GA-3' step has recently been observed to reduce polyamide binding affinities by approximately 10-fold.<sup>2d</sup>

Based on the pairing rules for polyamide-DNA complexes, the sites 5'-TGTAACA-3' and 5'-TGGAACA-3' are for polyamide **8** 'match' and 'single-base-pair-mismatch sites' respectively and for polyamide **9** 'single-base-pair-mismatch' and 'match sites' respectively. The specificity of polyamides **8** and **9** for their respective match sites results from very small structural changes (Figure 6.9). Replacing a single C-H in **8** with a nitrogen atom as in **9** reduces the affinity of the ImImPyPyPy-γ-ImPyPyPyPy-β-Dp•5'-TGTAACA-3' complex relative to the

**Table 6.3.** Equilibrium Association Constants for Polyamide **8** and **9** Bound to Their Respective Match and Mismatch Sites.

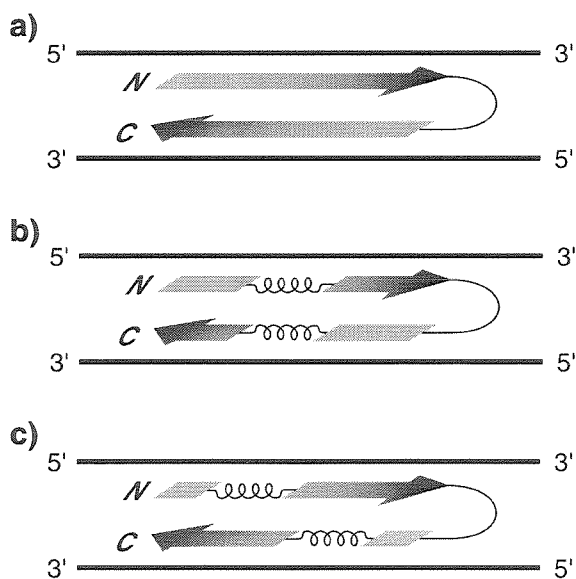


ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp•5'-TGTAACA-3' complex by > 1200-fold, a free energy difference of at least 4 kcal/mol. Similarly, replacing a N in **9** with a C-H as in **8** reduces the affinity of the ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp•5'-TGGAACA-3' complex relative to the ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp•5'-TGGAACA-3' complex by a factor of 5-fold, a loss in binding energy of ~1 kcal/mol.

The results presented here reveal that ten-ring polyamides based on  $\gamma$ -linked five-ring subunits expand the binding site size of the hairpin motif, providing a model for the recognition of 7-bp sequences. Recognition of 7 base pairs by a ten-ring hairpin represents an upper limit of contiguous rings which will match the curvature of the DNA helix without severe energetic penalty. In support of this, we find that a twelve-ring hairpin polyamide motif decreases with regard to both affinity and specificity compared with eight and ten-ring hairpin polyamides. This observation is consistent with failure of six-ring subunits containing solely Py and Im amino acids to maintain register as a dimer across the entire length of the minor groove of the helix.<sup>4,18</sup> Modified hairpin motifs which incorporate flexible linkers to reset the polyamide register with the DNA helix will be reported in due course.

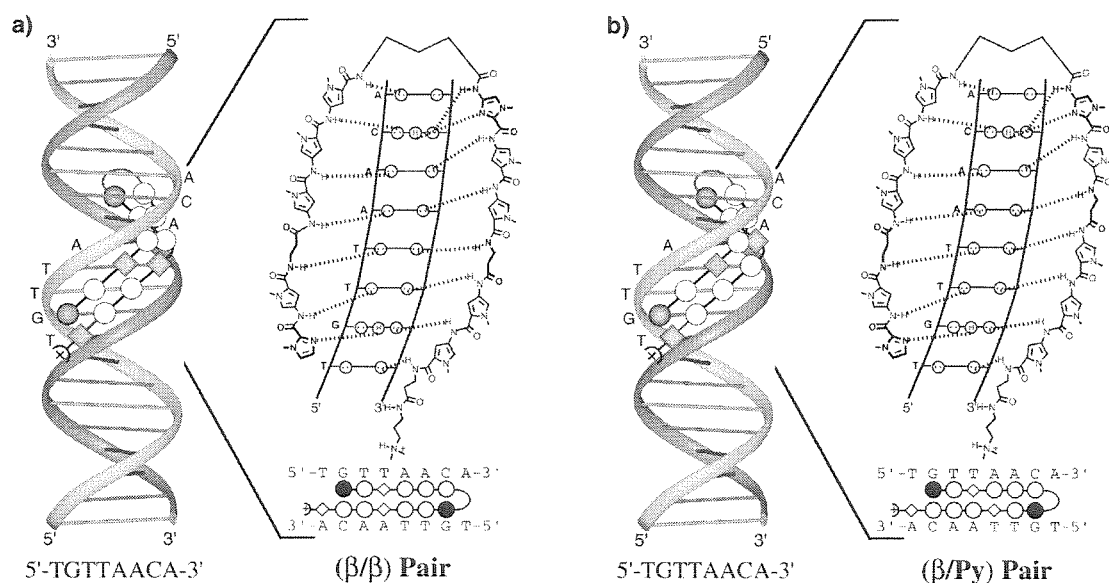
**Aliphatic/Aromatic Amino Acid Pairings for Polyamide Recognition in the Minor Groove of DNA.** Given the sequence dependent microstructure of the DNA helix,<sup>24</sup> it is surprising that a simple recognition code can be developed at all.<sup>1</sup> In both published and unpublished work, over a hundred Py-Im polyamides have been synthesized which recognize predetermined sequences. However, within this group certain difficult Py-Im/DNA base-pair sequences have emerged. Sequence dependent DNA structure features such as intrinsic minor groove width, minor groove flexibility, and inherent DNA curvature may reduce polyamide binding at certain sites.<sup>24</sup> However, it may also be possible to identify polyamide structural elements which will restore affinity at difficult sequences by providing an optimal fit between the hydrogen bond donors and acceptors displayed on the edges of both the Watson-Crick base pairs and the crescent shaped polyamide dimers.

In contrast to unlinked dimers which can adopt a variety of “slipped” binding modes, *the hairpin structure locks the relative positions of the individual subunits* and allows greater control of amino acid ring pairings. Three, four, and five-ring polyamides covalently coupled to form six,



**Figure 6.14.** Ribbon model for aliphatic amino acids in hairpin Py-Im polyamides. Arrows represent aromatic Im and Py amino acids, while springs represent  $\beta$ -alanine residues. (a): Six consecutive aromatic amino acid pairings; (b): One  $\beta$ / $\beta$  pairing; (c): Two aromatic/ $\beta$  pairings.

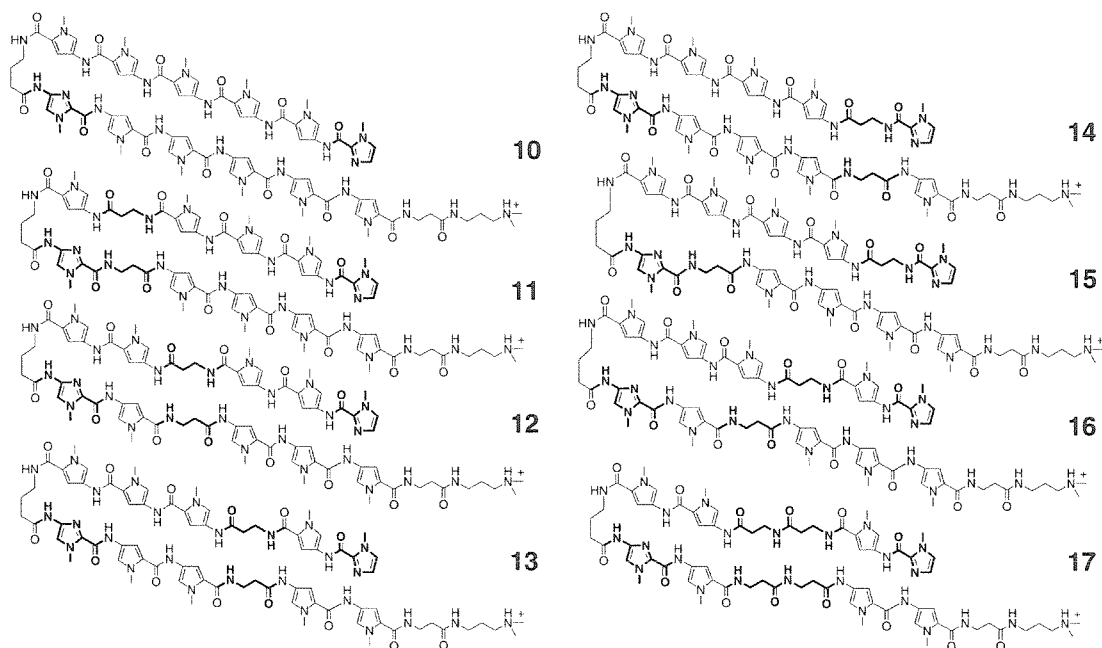




**Figure 6.15.** Binding model for the complexes formed between the DNA and (a): ImPy-β-PyPyPy-γ-ImPyPy-β-PyPy-β-Dp; (b): ImPy-β-PyPyPy-γ-ImPy-β-PyPyPy-β-Dp. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. Ball and stick models are also shown. Shaded and nonshaded circles denote Im and Py carboxamides, respectively. Nonshaded diamonds represent β-alanine residues.

eight, and ten-ring hairpin structures bind specifically to 5, 6, and 7-bp target sequences, respectively.<sup>6,23,26</sup> Recognition of seven base pairs by a ten-ring hairpin represents an upper limit of five contiguous ring pairings which will match the curvature of the DNA helix without energetic penalty.<sup>26</sup>

For complexes of fully overlapped 2:1 polyamide dimers in the minor groove, a flexible β-alanine (β) spring was found to form A,T specific β/β pairings which were necessary for recognition of longer binding sites.<sup>27</sup> It remained to be determined: (i) if paired β-alanine residues could be accommodated in the hairpin structure to restore register of the hairpin with the DNA helix for recognition of larger binding sites, (ii) if the antiparallel polyamide dimer could accommodate aromatic amino acids paired side-by-side with aliphatic β-amino acids to give β/Py and β/Im ring pairings (Figures 6.14 and 6.15).



**Figure 6.16.** Structures of the hairpin polyamides **10-17** synthesized by solid phase methods.

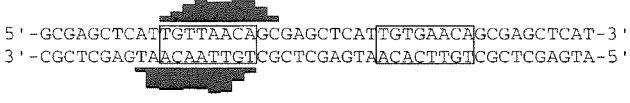
Eight polyamides were synthesized containing either solely ring amino acids, a side-by-side pairing of two  $\beta$ -alanine residues ( $\beta/\beta$ ), or a side-by-side pairing of a pyrrole and a  $\beta$ -alanine residue (Py/ $\beta$ ).

We report here the DNA-binding affinity and sequence selectivity of the eight hairpin polyamides, ImPyPyPyPyPy- $\gamma$ -ImPyPyPyPyPy- $\beta$ -Dp (**10**), ImPyPyPy- $\beta$ -Py- $\gamma$ -Im- $\beta$ -PyPyPyPy- $\beta$ -Dp (**11**), ImPyPy- $\beta$ -PyPy- $\gamma$ -ImPy- $\beta$ -PyPyPy- $\beta$ -Dp (**12**), ImPy- $\beta$ -PyPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPy- $\beta$ -Dp (**13**), Im- $\beta$ -PyPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Py- $\beta$ -Dp (**14**), Im- $\beta$ -PyPyPyPy- $\gamma$ -Im- $\beta$ -PyPyPyPy- $\beta$ -Dp (**15**), ImPy- $\beta$ -PyPyPy- $\gamma$ -ImPy- $\beta$ -PyPyPy- $\beta$ -Dp (**16**), and ImPy- $\beta$ - $\beta$ -PyPy- $\gamma$ -ImPy- $\beta$ - $\beta$ -PyPy- $\beta$ -Dp (**17**) for an 8-bp 5'-TGTTAACA-3' target sequence and a 5'-TGTGAACA-3' single-base-pair mismatch sequence. Polyamides **11-14** vary the position of a single ( $\beta/\beta$ ) pairing, while polyamides **15** and **16** have two (Py/ $\beta$ ) pairings. As a control, polyamide **17** was synthesized with two consecutive ( $\beta/\beta$ ) pairings (Figure 6.16). Additional polyamides ImPyImPyPyPy- $\gamma$ -ImPyPyPyPyPy- $\beta$ -Dp (**18**) and Im- $\beta$ -ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Py- $\beta$ -Dp (**19**) were designed to target the 8-bp sequence 5'-TGTGAACA-3'. A polyamide, Im- $\beta$ -ImPy- $\gamma$ -Im- $\beta$ -ImPy- $\beta$ -Dp (**20**),

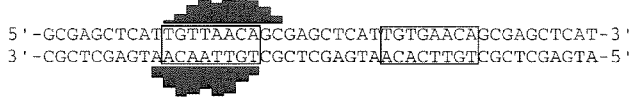
10 ImPyPyPyPyPy-γ-ImPyPyPyPyPy-β-Dp



13 ImPy-β-PyPyPy-γ-ImPyPy-β-PyPy-β-Dp



16 ImPy-β-PyPyPy-γ-ImPy-β-PyPyPy-β-Dp

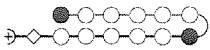
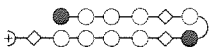
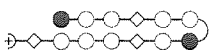
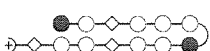
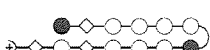
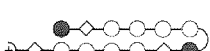
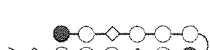
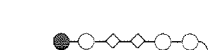


**Figure 6.17.** Results from MPE•Fe(II) footprinting of ImPyPyPyPyPy-γ-ImPyPyPyPyPy-β-Dp, ImPy-β-PyPyPyPy-γ-ImPyPy-β-PyPy-β-Dp, and ImPy-β-PyPyPyPy-γ-ImPy-β-PyPyPy-β-Dp. Boxes represent equilibrium binding sites determined by the published model. Only sites that were quantitated by DNase I footprint titrations are boxed. Bar heights are proportional to the relative protection from cleavage at each band.

which incorporates Im/β and β/Im pairings, was also synthesized and its DNA binding affinity and specificity determined for a 5'-TGCGCA-3' site.

MPE•Fe(II) footprinting on a <sup>32</sup>P end-labeled 254 base pair DNA restriction fragment (25 mM Tris-acetate, 10 mM NaCl, 100 mM calf thymus DNA, pH 7.0, 22°C) reveals that each polyamide is binding to the 5'-TGTTAACA-3' match site (Figure 6.17). Footprinting patterns reveal asymmetrically 3'-shifted protection of the 8-bp sites, consistent with formation of a 1:1 hairpin polyamide-DNA complex in the minor groove. Polyamide **10** at 10 μM concentration protects both the 5'-TGTTAACA-3' match site and the single-base-pair mismatch 5'-TGTGAACA-3' site. Polyamides **13** and **16** each at 10 μM concentration protect their cognate 5'-TGTTAACA-3' match site; however, no protection is observed at the 5'-TGTGAACA-3' single-base-pair mismatch site. Quantitative DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub> and 5 mM CaCl<sub>2</sub>, pH 7.0, 22°C) were performed to determine the equilibrium association constants of the polyamides for the two bound sites (Table 6.4).

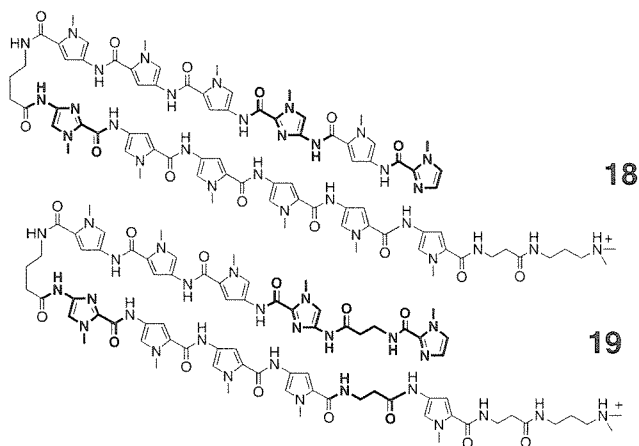
**Table 6.4.** Equilibrium association constants ( $M^{-1}$ ) for polyamides.<sup>a-c</sup>

Polyamide	5'-TGTTAACA-3'	5'-TGTGAACA-3'	Specificity <sup>d</sup>
	$2.5 \times 10^9$	$3.9 \times 10^8$	6
	$1.3 \times 10^9$	$2.0 \times 10^8$	7
	<b><math>1.7 \times 10^{10}</math></b>	$2.7 \times 10^9$	6
	<b><math>1.2 \times 10^{11}</math></b>	$2.2 \times 10^9$	<b>55</b>
	<b><math>6.6 \times 10^9</math></b>	$2.5 \times 10^8$	<b>26</b>
	<b><math>4.5 \times 10^{10}</math></b>	$7.7 \times 10^9$	6
	<b><math>2.7 \times 10^{10}</math></b>	$5.7 \times 10^9$	5
	$\leq 1 \times 10^8$	$\leq 1 \times 10^8$	1

<sup>a</sup>Values reported are the mean values obtained from three DNase I footprint titration experiments. <sup>b</sup>The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. <sup>c</sup>Match site association constants and specificities higher than the parent hairpin are shown in boldtype. <sup>d</sup>Specificity is calculated as  $K_a(\text{match}) / K_a(\text{mismatch})$ .

The 5'-TGTTAACA-3' match site was bound by the polyamides with decreasing affinity:

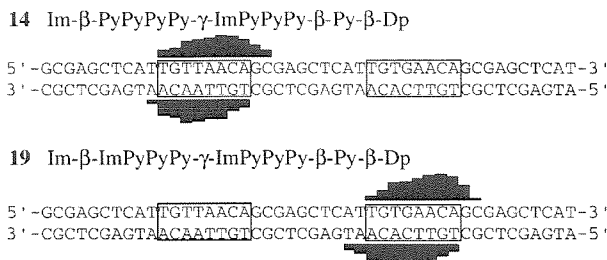
ImPy-β-PyPyPy-γ-ImPyPy-β-PyPy-β-Dp (**13**,  $K_a = 1.2 (\pm 0.1) \times 10^{11} M^{-1}$ ) > Im-β-PyPyPyPy-γ-Im-β-PyPyPyPy-β-Dp (**15**,  $K_a = 4.5 (\pm 2.7) \times 10^{10} M^{-1}$ ) > ImPy-β-PyPyPy-γ-ImPy-β-PyPyPy-β-Dp (**16**,  $K_a = 2.7 (\pm 1.5) \times 10^{10} M^{-1}$ ) > ImPyPy-β-PyPy-γ-ImPy-β-PyPyPy-β-Dp (**12**,  $K_a = 1.7 (\pm 0.4) \times 10^{10} M^{-1}$ ) > Im-β-PyPyPyPy-γ-ImPyPyPy-β-Py-β-Dp (**14**,  $K_a = 6.6 (\pm 1.9) \times 10^9 M^{-1}$ ) > ImPyPyPyPyPy-γ-ImPyPyPyPyPy-β-Dp (**10**,  $K_a = 2.5 (\pm 0.6) \times 10^9 M^{-1}$ ) > ImPyPyPy-β-Py-γ-Im-β-PyPyPyPy-β-Dp (**11**,  $K_a = 1.3 (\pm 0.4) \times 10^9 M^{-1}$ ) > ImPy-β-β-PyPy-γ-ImPy-β-β-PyPy-β-Dp (**17**,  $K_a = 1 \times 10^8 M^{-1}$ ). The 5'-TGTGAACA-3' site was bound with decreasing affinity: Im-β-PyPyPyPy-γ-Im-β-PyPyPyPy-β-Dp (**15**,  $K_a = 7.7 (\pm 4.5) \times 10^9 M^{-1}$ ) > ImPy-β-PyPyPy-γ-ImPy-β-



**Figure 6.18.** Structures of the hairpin polyamides ImPyImPyPyPy- $\gamma$ -ImPyPyPyPyPy- $\beta$ -Dp (**18**) and Im- $\beta$ -ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Py- $\beta$ -Dp (**19**) synthesized by solid phase methods. Polyamides were synthesized using Boc-chemistry machine-assisted solid phase protocols.

PyPyPy- $\beta$ -Dp (**16**,  $K_a = 5.7 (\pm 4.1) \times 10^9 \text{ M}^{-1}$ ) > ImPyPy- $\beta$ -PyPy- $\gamma$ -ImPy- $\beta$ -PyPyPy- $\beta$ -Dp (**12**,  $K_a = 2.7 (\pm 1.5) \times 10^9 \text{ M}^{-1}$ )  $\cong$  ImPy- $\beta$ -PyPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPy- $\beta$ -Dp (**13**,  $K_a = 2.2 (\pm 1.9) \times 10^9 \text{ M}^{-1}$ ) > ImPyPyPyPyPy- $\gamma$ -ImPyPyPyPyPy- $\beta$ -Dp (**10**,  $K_a = 3.9 (\pm 1.7) \times 10^8 \text{ M}^{-1}$ ) > Im- $\beta$ -PyPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Py- $\beta$ -Dp (**14**,  $K_a = 2.5 (\pm 0.4) \times 10^8 \text{ M}^{-1}$ )  $\cong$  ImPyPyPy- $\beta$ -Py- $\gamma$ -Im- $\beta$ -PyPyPyPy- $\beta$ -Dp (**11**,  $K_a = 2.0 (\pm 0.1) \times 10^8 \text{ M}^{-1}$ ) > ImPy- $\beta$ - $\beta$ -PyPy- $\gamma$ -ImPy- $\beta$ - $\beta$ -PyPy- $\beta$ -Dp (**17**,  $K_a = 1 \times 10^8 \text{ M}^{-1}$ ). Remarkably, the equilibrium association constant for the 5'-TGTTAACA-3' match site varied 100-fold between polyamides **10-16**, indicating a sensitivity to the position relative to the  $\gamma$ -turn for placement of ( $\beta/\beta$ ) or (Py/ $\beta$ ) pairings. A structural basis for the  $\beta$ -alanine mediated affinity enhancement awaits high resolution x-ray and NMR studies which are in progress.

Polyamide **10** places a Py/Py pair opposite T•A in the fourth position and A•T in the fifth position in the sequence 5'-TGTTAACA-3'. The mismatch sequence 5'-TGTTGAACA-3' (Py/Py opposite G•C in the fourth position and A•T in the fifth position) is discriminated by 6-fold. For polyamide **7**, each of these Py/Py pairs is replaced sequentially with a  $\beta$ /Py or Py/ $\beta$  pair. Similar sequence specificity is observed for polyamides **10** and **16**, indicating that the single Py/ $\beta$  and  $\beta$ /Py pairings discriminate A•T, T•A from G•C, C•G at least as effectively as the Py/Py pair.

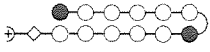
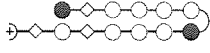
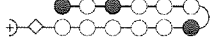



**Figure 6.19.** Results from MPE•Fe(II) footprinting of Im-β-PyPyPyPy-γ-ImPyPyPy-β-Py-β-Dp and Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-β-Dp. Boxes represent equilibrium binding sites determined by the published model. Only sites that were quantitated by DNase I footprint titrations are boxed. Bar heights are proportional to the relative protection from cleavage at each band.

Based on the pairing rules for polyamide-DNA complexes, the sites 5'-TGTTAACA-3' and 5'-TGTGAACA-3' are for polyamides **10-17** “match” and “single-base-pair mismatch” sites, respectively. The full-ring hairpin polyamide ImPyImPyPyPy-γ-ImPyPyPyPyPy-β-Dp (**18**) (Figure 6.18) recognizes the same 5'-TGTGAACA-3' match site with reduced affinity ( $K_a = 5 \times 10^9 \text{ M}^{-1}$ ) and no specificity versus the 5'-TGTTAACA-3' single-base-pair mismatch site. This may be due to excessive curvature of the ligand preventing both imidazoles of the ImPyIm subunit from making favorable contacts with the DNA.<sup>4</sup> The substitution of Im for Py in Im-β-PyPyPyPy-γ-ImPyPyPy-β-Py-β-Dp to give Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-β-Dp (**19**), formally the single atomic substitution of N for C-H, regulates affinity and specificity by more than an order of magnitude.

MPE•Fe(II) footprinting on a <sup>32</sup>P end-labeled 254 base pair DNA restriction fragment (Figure 6.19) reveals asymmetrically 3'-shifted protection of 8-bp sites, consistent with formation of a 1:1 hairpin polyamide-DNA complex in the minor groove. Polyamide **14** protects its cognate 5'-TGTTAACA-3' match site at 1 μM concentration, while no protection is observed at the cognate 5'-TGTGAACA-3' mismatch site. Polyamide **19** protects its cognate 5'-TGTGAACA-3' match site at 1 μM concentration; however, no protection is observed at the 5'-TGTTAACA-3' single-base-pair mismatch site. Quantitative DNaseI footprint titrations (Table 6.5) were

**Table 6.5.** Equilibrium association constants ( $M^{-1}$ ) for polyamides.<sup>a-c</sup>

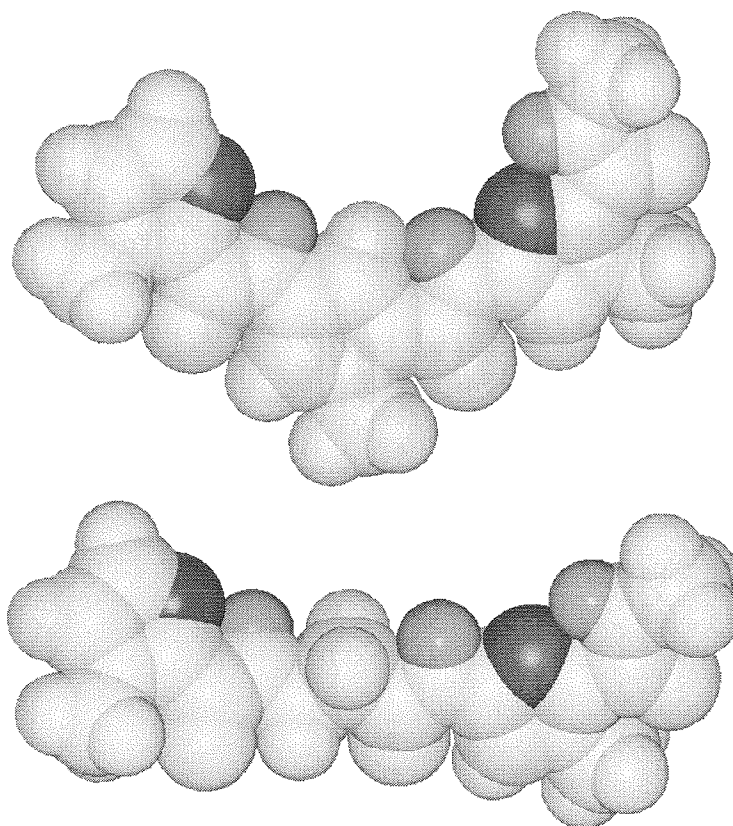
Polyamide	5'-TGTTAACA-3'	5'-TGTGAACA-3'	Specificity <sup>d</sup>
	$2.5 \times 10^9$	$3.9 \times 10^8$	6
	<b><math>6.6 \times 10^9</math></b>	$2.5 \times 10^8$	<b>26</b>
	$5 \times 10^9$	$5 \times 10^9$	1
	$\leq 5 \times 10^8$	<b><math>2.4 \times 10^{10}</math></b>	<b><math>\geq 48</math></b>

<sup>a</sup>Values reported for **10**, **14**, and **19** are the mean values obtained from three DNase I footprint titration experiments. <sup>b</sup>The assays were carried out at 22 °C at pH 7.0 in the presence of 10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. <sup>c</sup>Match site association constants and specificities higher than parent hairpins are shown in boldtype. <sup>d</sup>Specificity is calculated as  $K_a(\text{match}) / K_a(\text{mismatch})$ .

performed in order to determine the equilibrium association constants for polyamide **19** binding its 5'-TGTGAACA-3' match site and 5'-TGTTAACA-3' mismatch site. Im-β-ImPyPyPy-γ-ImPyPyPy-β-Py-β-Dp (**19**) recognizes its 5'-TGTGAACA-3' match site with  $K_a = 2.4 (\pm 0.2) \times 10^{10} M^{-1}$ . The 5'-TGTTAACA-3' mismatch sequence is bound with 48-fold ( $K_a = 5 \times 10^8 M^{-1}$ ) reduced affinity, respectively (Table 6.5).

The specificity and affinity of **19** suggests that all three imidazole residues make favorable DNA contacts. Previous studies have demonstrated that internal amino acids located in the third or fourth position relative to the subunit N-terminus do not form optimal hydrogen bonds.<sup>26a</sup> Modeling suggests that the increased affinity and specificity of hairpin **19** is due to a partial straightening of the ligand (Figure 6.20), allowing both imidazoles to adopt a curvature which better matches that of the DNA helix. These results identify Im-β-Im subunits as a general structure motif for the design of ligands that target certain “problematic” sequences which require an imidazole at the third position.

In an effort to target “core” G•C rich sequences, we had reported previously that the eight-ring hairpin polyamide ImPyImPy-γ-ImPyImPy-β-Dp (**20**) (Figure 6.21) binds to its 5'-

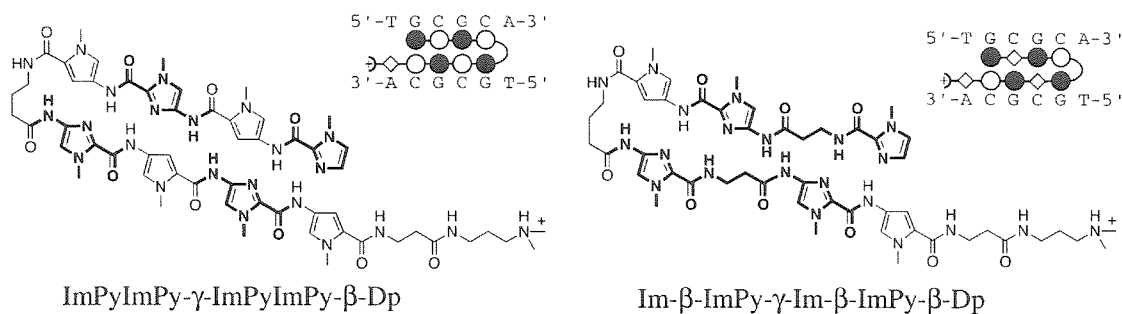


**Figure 6.20.** Model of the curvature adopted by (a): ImPyIm; and (b): Im-β-Im subunits. The N3 of imidazole is shown in dark gray, and amide NH groups are shown light gray. For simplicity, the C-terminus of both subunits is modeled as the methyl amide. Structures were minimized using MM2 calculations in Chem3D Pro v. 3.5 running on a Power Computing 180e computer. Minimized coordinates were then used to generate space-filling structures using Insight II v. 2.2 software on a Silicon Graphics Workstation.

TGCGCA-3' match sequence with good specificity but with only a modest affinity ( $K_a = 3.7 (\pm 0.9) \times 10^7 \text{ M}^{-1}$ ) compared to other eight-ring hairpin polyamides which bind their target sites with subnanomolar affinity.<sup>23,26</sup> Because paired β-residues reset the imidazole residues of polyamide **19** for recognition of 5'-GWG-3' sequences, it seemed plausible that Im/β and β/Im pairings could be used to optimize the positioning of the imidazole amino acids of **20** in a similar manner (Figure 6.21).

In order to determine whether Im/β pairs would be G•C specific recognition elements, the polyamide Im-β-ImPy-γ-Im-β-ImPy-β-Dp (**21**) was synthesized and its DNA binding properties





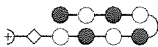
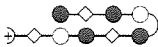
**Figure 6.21.** Structures of the hairpin polyamides ImPyImPy-γ-ImPyImPy-β-Dp (**20**) and Im-β-ImPy-γ-Im-β-ImPy-β-Dp (**21**) synthesized by solid phase methods. Polyamides were synthesized using Boc-chemistry machine-assisted solid phase protocols. Ball and stick models are also shown. Shaded and nonshaded circles denote Im and Py carboxamides, respectively. Nonshaded diamonds represent β-alanine residues.

compared with ImPyImPy-γ-ImPyImPy-β-Dp (**20**). MPE•Fe(II) footprinting (25 mM Tris-acetate, 10 mM NaCl, 100 μM/base pair calf thymus DNA, pH 7.0 and 22°C) was performed on the 3'- and 5'-<sup>32</sup>P end-labeled 263 base pair restriction fragments from the plasmid pSES11. The assays reveal identical footprinting protection at the designed 5'-TGCGCA-3' match site for polyamides **20** and **21** and are consistent with 6 base pair recognition.

Quantitative DNaseI footprint titrations were performed in order to determine the equilibrium association constants for polyamide **21** binding to three G,C sequences: 5'-TGCGCA-3', 5'-TGGCCA-3', and 5'-TGGGGA-3'. Im-β-ImPy-γ-Im-β-ImPy-β-Dp (**21**) recognizes its 5'-TGCGCA-3' match site with  $K_a = 3.7 (\pm 1.5) \times 10^9 \text{ M}^{-1}$ , a 100-fold increase in affinity over eight-ring polyamide **20**. (Table 6.6) Im/β pairings are found to be sequence specific for G•C base pairs relative to C•G, A•T, and T•A. For example, the mismatch sequences 5'-TGGCCA-3' and 5'-TGGGGA-3' are bound with 26-fold ( $K_a = 1.4 (\pm 1.0) \times 10^8 \text{ M}^{-1}$ ) and 34-fold ( $K_a = 1.1 (\pm 0.6) \times 10^8 \text{ M}^{-1}$ ) reduced affinity, respectively.

Substitution of β residues for ring residues does not always result in a gain in binding energy. For example, in unpublished results it was determined that Im-β-PyPy-γ-Im-β-PyPy-β-Dp binds to a 5'-TGTACA-3' match site with lower affinity than ImPyPyPy-γ-ImPyPyPy-β-Dp.

**Table 6.6.** Equilibrium association constants ( $M^{-1}$ ) for polyamides.<sup>a</sup>

Polyamide	5'-TGCGCA-3'	5'-TGGCCA-3'	5'-TGGGGA-3'
	$3.7 \times 10^7$	$< 10^7$	$< 10^7$
	$3.7 \times 10^9$	$1.4 \times 10^8$	$1.1 \times 10^8$

<sup>a</sup> Values reported are the mean values obtained from a minimum of three DNase I footprint titration experiments.

Therefore, for 5'-TGTACA-3' recognition, four-ring subunits containing exclusively ring amino acids are optimal, while for 5'-TGCGCA-3' recognition,  $\beta$ -alanine “spring” amino acids are necessary in order to reset the register of the internal Im amino acids. These results indicate that the preferred polyamide template for sequence recognition is not determined merely by site size, but also by sequence composition.

The results presented here reveal that hairpin polyamides based on optimally spaced subunits provide a useful design for recognition of 8-bp binding sites in the hairpin motif. The high binding affinity of the paired-( $\beta/\beta$ ) hairpin polyamides described here indicates that the Py and Im amino acid residues make energetically favorable contact *across all eight base pairs* of the DNA helix. The specificity of polyamides which differ by a single atomic substitution, Im- $\beta$ -ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Py- $\beta$ -Dp and Im- $\beta$ -PyPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Py- $\beta$ -Dp, for discrimination of sites which differ by a single base-pair, identifies the Im- $\beta$ -Im subunit as a structure motif for the design of ligands that target certain “problematic” sequences which require an imidazole at the third position. The use of the Im- $\beta$ -Im subunit in both the 6-bp (Im- $\beta$ -ImPy- $\gamma$ -Im- $\beta$ -ImPy- $\beta$ -Dp) and 8-bp (Im- $\beta$ -ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Py- $\beta$ -Dp) hairpin templates indicates that  $\beta$ -alanine “springs” introduced as either aromatic/aliphatic or aliphatic/aliphatic pairs may offer a general approach to broaden the accessible sequence repertoire for hairpin polyamides. The reduced binding affinity of polyamide **17** suggests that consecutive side-by-side pairing of  $\beta$ -alanine creates unfavorable polyamide:DNA interactions. These results are

**Table 6.7.** Aliphatic/Aromatic pairing code for minor groove recognition\*

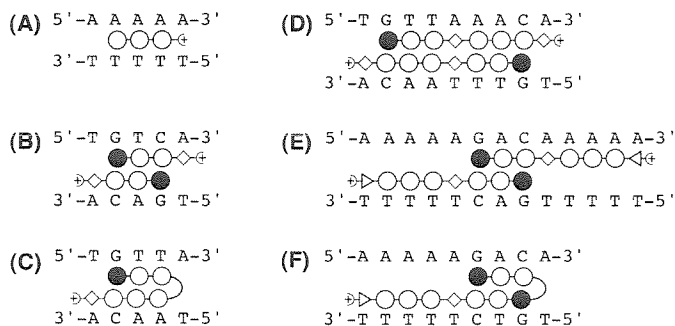
Pair	G•C	C•G	T•A	A•T
β/β	-	-	+	+
β/Py	-	-	+	+
Py/β	-	-	+	+
Im/β	+	-	-	-
β/Im	-	+	-	-

\* favored (+), disfavored (-)

consistent with the observation that motifs for recognition of longer sequences can use alternating sets of aromatic ring pairings separated by a single β/β pair.<sup>27c</sup> The sequence-dependent combination of rings and β-alanine required for optimal polyamide binding to longer sequences awaits further studies.

Polyamides which incorporate Py/β, β/Py, Im/β, and β/Im pairings provide the first example of aromatic/aliphatic amino acid pairings for minor groove recognition. Py/β and β/Py pairings are found to recognize A•T/T•A relative to G•C/C•G. The Im/β pairing is found to target G•C relative to C•G, A•T, and T•A, while the β/Im pair targets C•G relative to G•C, A•T, and T•A (Table 6.7). The results described here provide guidelines for placement of Im/β, β/Im, Py/β, β/Py and β/β pairings within the hairpin template, setting the stage for investigation of the effects of polyamide molecular weight, affinity, and binding site size on targeted gene regulation in living cells.

**Extended Hairpin Polyamide Motif.** The polyamide ImPyPy- $\gamma$ -PyPyPy-Dp, containing a “ $\gamma$ -turn” amino acid specifically binds the sequence 5'-TGTTA-3' in a “hairpin” conformation with a  $K_a$  of  $8 \times 10^7 \text{ M}^{-1}$ .<sup>6</sup> The six-ring polyamide ImPyPy- $\beta$ -PyPyPy-Dp containing an internal  $\beta$ -residue specifically binds as a dimer in an extended conformation to two designated target sites, 5'-TGTTAAACA-3' (9 bp) and 5'-AAAAAGACAAAAA-3' (13 bp), with  $K_a = 8 \times 10^8 \text{ M}^{-1}$  and  $K_a = 5 \times 10^9 \text{ M}^{-1}$ , respectively.<sup>27</sup> This represents increases of 10-fold for the 9 bp sequence and 100-fold for the 13 bp sequence over the formally Py-linked polyamide ImPyPy-Py-PyPyPy-Dp. Addition of a C-terminal amino acid regulates the specificity between the 13 bp and 9 bp binding modes. Two ImPyPy- $\beta$ -PyPyPy- $\beta$ -Dp polyamides, each containing a C-terminal  $\beta$ -residue, can bind directly opposite one another, whereas two ImPyPy- $\beta$ -PyPyPy-G-Dp polyamides, each containing a C-terminal glycine residue strongly favor binding in the 13 bp binding mode. The 13 bp binding mode integrates the 2:1 and 1:1 polyamide-DNA binding motifs at a single site.<sup>27</sup> The ImPyPy moieties of two ImPyPy- $\beta$ -PyPyPy-G-Dp polyamides bind the central 5'-AGACA-3' sequence in a 2:1 manner as in the ImPyPy homodimer, and the PyPyPy



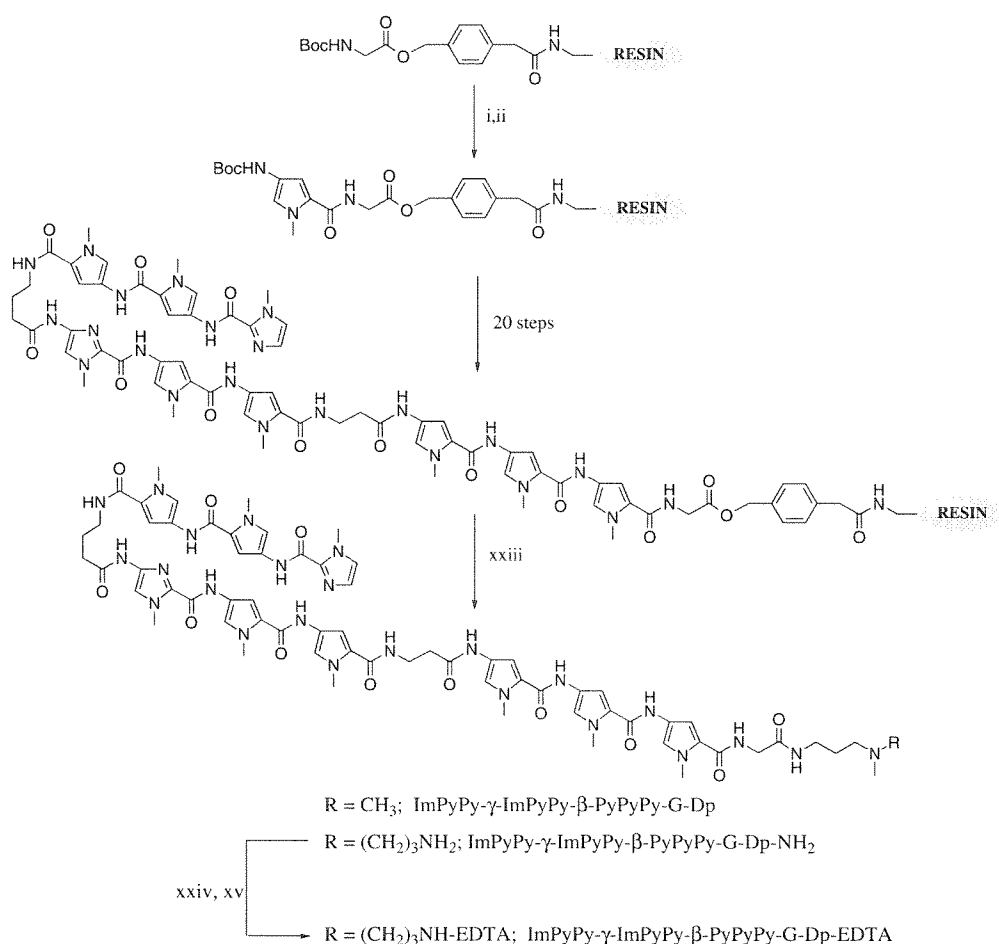
**Figure 6.22.** Schematic models of polyamide-DNA complexes: (A) 1:1 complex of distamycin (PyPyPy), (B) 2:1 complex of ImPyPy-Dp, (C) “hairpin” complex of ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp, (D) 9 bp and (E) 13 bp “extended” complexes formed by ImPyPy- $\beta$ -PyPyPy- $\beta$ -Dp and ImPyPy- $\beta$ -PyPyPy-G-Dp, respectively, and (F) 9bp “extended hairpin” complex of ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp. Filled and unfilled circles represent Im and Py rings, triangles and diamonds represent glycine and  $\beta$ , respectively, and curved lines represent  $\gamma$ .

moieties of the polyamides bind the A,T flanking sequences as in the 1:1 complexes of distamycin (Figure 6.22).

It has been demonstrated that  $\gamma$ -aminobutyric does not optimally link polyamide subunits in extended conformations, and  $\beta$ -alanine does not optimally link polyamide subunits in hairpin conformations.<sup>27</sup> The polyamide ImPyPy- $\beta$ -PyPyPy-Dp binds the hairpin site 5'-TGTTAgacc-3' with an association constant of  $K_a \leq 2 \times 10^6 \text{ M}^{-1}$ , a decrease of  $\geq 40$ -fold relative to ImPyPy- $\gamma$ -PyPyPy-Dp, and displays a cooperative binding isotherm in quantitative footprinting experiments at this site, consistent with mismatched binding as an intermolecular dimer. The polyamide ImPyPy- $\gamma$ -PyPyPy-Dp binds the 13 bp site 5'-AAAAAGACAAAA-3' with an association constant of  $K_a = 6 \times 10^6 \text{ M}^{-1}$ , a decrease of 700-fold relative to ImPyPy- $\beta$ -PyPyPy-Dp, and displays a Langmuir binding isotherm at this site consistent with mismatched binding as a hairpin. The site 5'-TGTTAAACA-3' is a match site for both hairpin (5'-TGTTA-3' and 5'-AAACA-3') and extended binding by ImPyPy-X-PyPyPy-Dp polyamides. ImPyPy- $\gamma$ -PyPyPy-Dp and ImPyPy- $\beta$ -PyPyPy-Dp both bind this sequence with high affinity ( $K_a = 1 \times 10^8 \text{ M}^{-1}$  and  $K_a = 8 \times 10^8 \text{ M}^{-1}$ , respectively), but display, respectively, a Langmuir isotherm consistent with hairpin binding and a cooperative isotherm consistent with extended binding. Remarkably, given their simplicity and similarity,  $\gamma$ -aminobutyric acid and  $\beta$ -alanine selectively link polyamide subunits in hairpin and extended conformations, respectively.

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

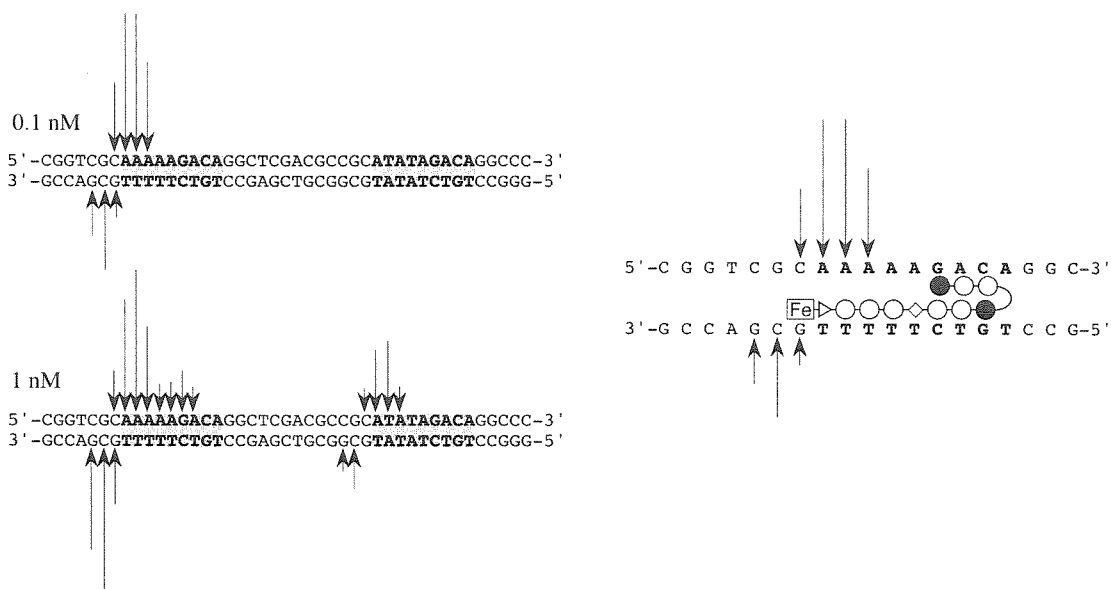




**Figure 6.24.** Synthetic scheme for polyamides **23**, **23-NH<sub>2</sub>** and **23-E**. Cycling protocols consist of TFA deprotection, followed by coupling with -Obt activated ester, steps i-xxii. The resin is then cleaved by treatment with a primary amine, step xxiii, and purified by reversed phase HPLC. Polyamide **23-NH<sub>2</sub>** is subsequently post-synthetically modified with EDTA dianhydride, xxiv, xxv, to provide the affinity cleaving analog, **23-E**.

experiments. We also carried out affinity cleavage experiments using the EDTA•Fe(II)-polyamide ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp-EDTA•Fe(II) (**23-E**) to determine the DNA-binding orientation and qualitative DNA-binding affinity and specificity of this polyamide.

All polyamides were prepared in high purity using a solid phase synthetic methodology. Polyamides **22** and **23** were assembled in a stepwise manner on Boc- $\beta$ -alanine-Pam resin and Boc-glycine-Pam-resin, respectively. Polyamides **22**, **23** and **23-NH<sub>2</sub>** were cleaved from the



**Figure 6.25.** Compound **2-E** cleaves the 247 bp pJT4 *Afl* II/*Fsp* I restriction fragment at specific sites. (right) Schematic representation of affinity cleaving experiment. Arrow lengths are proportional to the amount of cleavage at the indicated base. (left) Observed cleavage intensities at the designated match sites at 0.1 nM and 1 nM concentrations.

support with an appropriate primary amine and purified by reversed-phase HPLC to provide 10–30 mg of polyamide. Polyamide **23-NH<sub>2</sub>** was treated with an excess of the dianhydride of EDTA, unreacted anhydride was hydrolyzed, and polyamide **23-E** isolated by reversed-phase HPLC (Figure 6.24).

We carried out affinity cleavage experiments with ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp-EDTA•Fe(II) (**23-E**) on the 5'- or 3'-<sup>32</sup>P end-labeled 247 bp pJT4 *Afl* II/*Fsp* I restriction fragment. This polyamide selectively binds to the 5'-AAAAAGACA-3' and 5'-ATATAGACATATA-3' target sequences at subnanomolar concentrations. A single 3'-shifted cleavage pattern is observed at each 9 bp binding site, indicating that the polyamide is bound in one orientation with the C-terminus at the 5' end of the 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' sequences (Figure 6.25).

Quantitative DNase I footprint titration experiments on the 3'-<sup>32</sup>P-labeled 247 bp restriction fragment (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>, pH 7.0, 22°C)



**Table 6.8.** Equilibrium association constants ( $M^{-1}$ )<sup>a,b,c,d</sup>.

Binding Site	Polyamide	
	ImPyPy- $\gamma$ -ImPyPy- $\beta$ -Dp	ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp
Match sites:		
5'- <u>AAAAA</u> <b>GACA</b> -3'	$5.2 \times 10^7$ (0.5)	$2.0 \times 10^{10}$ (0.4)
5'- <u>ATATA</u> <b>GACA</b> -3'	$9.1 \times 10^7$ (1.1)	$8.1 \times 10^9$ (0.1)
Mismatch sites:		
5'- <u>GAATT</u> <b>CACT</b> -3'	$\leq 8 \times 10^6$	$4.5 \times 10^9$ (1.0)
5'- <u>CGTTT</u> <b>TACA</b> -3'	$9.2 \times 10^6$ (1.6)	$1.6 \times 10^9$ (0.3)
5'- <u>GTTTT</u> <b>CCCA</b> -3'	$< 10^6$	$2.5 \times 10^9$ (0.7)
5'-GGCG <u>ATTA</u> <b>AGTTG</b> -3'	$< 10^6$	$8.9 \times 10^8$ (0.6)
5'-TCGCT <u>ATTAC</u> <b>GCCA</b> -3'	$< 10^6$	$1.5 \times 10^9$ (0.2)

<sup>a</sup>Values reported are the mean values obtained from four DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses. <sup>b</sup>The assays were carried out at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>. <sup>c</sup>The 5 base pair ImPyPy- $\gamma$ -ImPyPy- $\beta$ -Dp binding sites are in bold type. <sup>d</sup>The portion of each site occupied by the C-terminal PyPyPy subunit of ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp, as evidenced by affinity cleavage experiments with ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp-EDTA•Fe(II), is underlined.

revealed that ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp specifically binds the 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' target sequences with equilibrium association constants of  $K_a \sim 2 \times 10^{10} M^{-1}$  and  $K_a = 8 \times 10^9 M^{-1}$ , respectively (Table 6.8). The polyamide also binds to additional sites on the restriction fragment with lower affinity. The six-ring hairpin polyamide ImPyPy- $\gamma$ -ImPyPy- $\beta$ -Dp binds 5'-aaaaAGACA-3' and 5'-atatAGACA-3' with association constants of  $K_a = 5 \times 10^7 M^{-1}$  and  $K_a = 9 \times 10^7 M^{-1}$ , respectively.

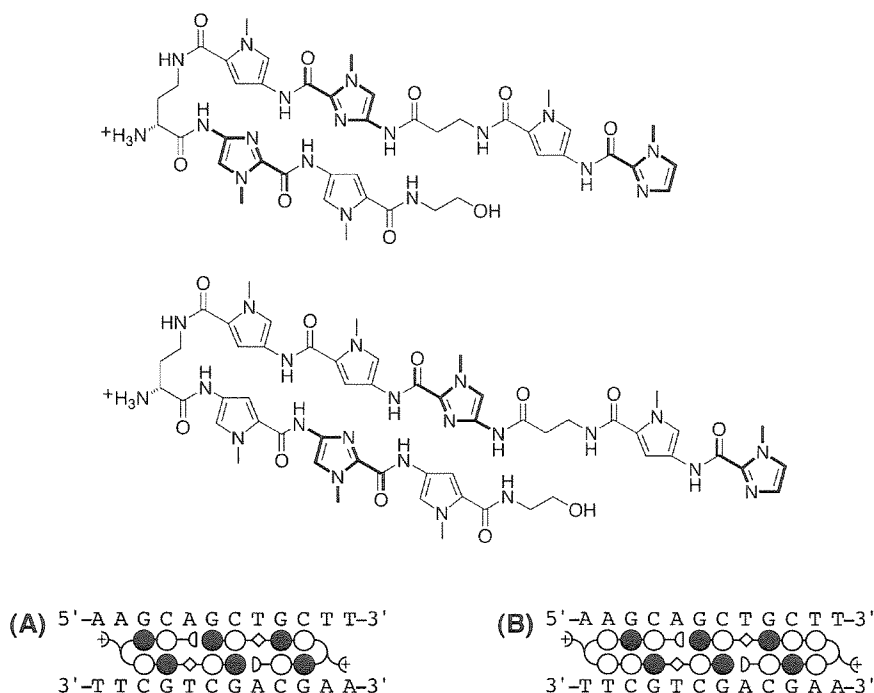
Relative to the six-ring polyamide ImPyPy- $\gamma$ -ImPyPy- $\beta$ -Dp, the nine-ring polyamide ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp binds 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' with ~400-fold and ~100-fold higher affinity, respectively. Addition of a C-terminal PyPyPy subunit using a  $\beta$ -alanine linker is thus an effective strategy for increasing the DNA-binding affinity of

hairpin polyamides that bind adjacent to an (A,T)<sub>4</sub> sequence.

Polyamide ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp binds to several mismatch sites present on the 247 bp restriction fragment with high affinity (Table 6.8). The two highest affinity mismatch sites, 5'-GAATTCACT-3' ( $4.5 \times 10^9 \text{ M}^{-1}$ ) and 5'-GTTTTCCCA-3' ( $2.5 \times 10^9 \text{ M}^{-1}$ ), are bound with at least 5-fold reduced affinity relative to the optimal match site 5'-AAAAAGACA-3' (formally mismatched base-pairs are highlighted). This value may, however, be a lower limit due to the uncertainty in the very high association constant for the optimal match site. In contrast, the six-ring polyamide ImPyPy- $\gamma$ -ImPyPy- $\beta$ -Dp binds more strongly to the match site 5'-AGACA-3' than the single base pair mismatch sites 5'-ATTCA-3' and 5'-TTTACA-3' by a factor of 10. The encouraging sequence-specificity of the extended hairpin motif suggests utility of extended hairpin polyamides to develop a cooperative hairpin motif for recognition of predetermined DNA-sequences.

**Cooperative Hairpin Dimers.** Small molecules that permeate cells and bind predetermined DNA sequences have the potential to control the expression of specific genes. Recently, an eight-ring polyamide that binds to a 6 base pair target site was shown to inhibit gene transcription in cell culture.<sup>21</sup> Polyamides recognizing longer DNA sequences should provide more specific biological activity,<sup>31</sup> which could be achieved by synthesizing larger hairpins.<sup>26,30</sup> However, the upper limit of polyamide size with regard to efficient cell permeation is not known.

Alternatively, a more biomimetic approach is to bind larger DNA sequences while maintaining the size of the polyamide. Nature's transcription factors often bind large DNA sequences by formation of cooperative protein dimers at adjacent half-sites.<sup>32</sup> For cooperatively binding extended Py-Im polyamide dimers, the two ligands can slip sideways with respect to one another, allowing recognition of other sequences.<sup>27</sup> Hairpin polyamides utilizing the turn-specific  $\gamma$ -aminobutyric acid linker<sup>6</sup> are constrained to be fully overlapped and preclude the "slipped



**Figure 6.26.** (top) Structures of polyamides **24**, ImPy- $\beta$ -ImPy-( $R$ )<sup>H<sub>2</sub>N</sup> $\gamma$ -ImPy-C<sub>2</sub>-OH and **25**, ImPy- $\beta$ -ImPyPy-( $R$ )<sup>H<sub>2</sub>N</sup> $\gamma$ -PyImPy-C<sub>2</sub>-OH. (bottom) Models of (A) the 10 base pair (**24**)<sub>2</sub>•5'-AGCAGCTGCT-3' complex and (B) the 12 base pair (**25**)<sub>2</sub>•5'-AAGCAGCTGCTT-3' complex.

motif" option. Here we report a cooperative six-ring extended hairpin polyamide which dimerizes to specifically bind a predetermined 10 base pair sequence.

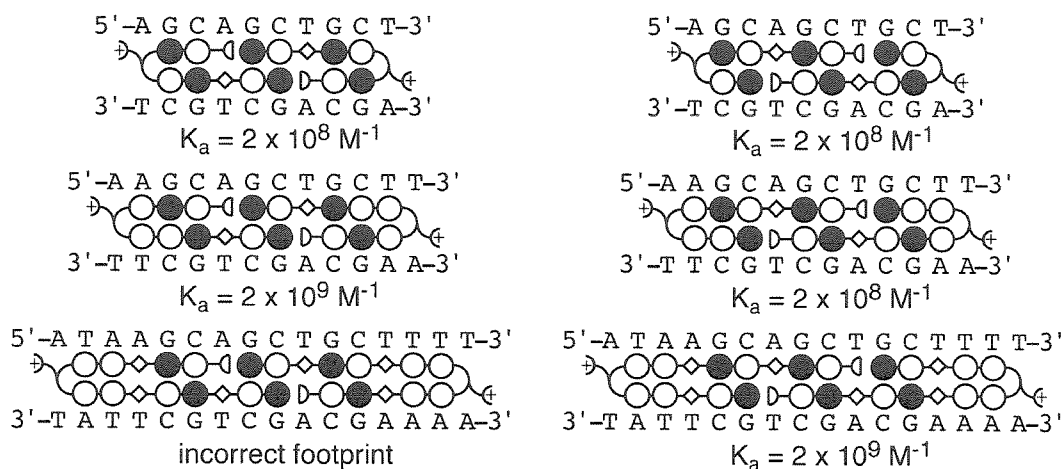
For our target site, we chose a sequence contained in the regulatory region of the HIV-1 genome.<sup>33</sup> To design the ligand we considered the polyamide ring pairing rules, the need for  $\beta$ -alanine to relax ligand curvature,<sup>27</sup> and the preference of  $\gamma$ -aminobutyric acid for a "hairpin turn" conformation within polyamide-DNA complexes. This analysis suggested that the six-ring polyamide having the core sequence ImPy- $\beta$ -ImPy- $\gamma$ -ImPy might bind the target sequence 5'-AGCAGCTGCT-3' through formation of a cooperative hairpin dimer. To avoid a collision between the N-terminal end of one ligand and the C-terminal end of the second within the complex, the positively-charged  $\beta$ -alanine-dimethylaminopropylamide C-terminus used in

standard polyamides has been replaced with the shorter, uncharged (CH<sub>2</sub>)<sub>2</sub>OH group (C<sub>2</sub>-OH). The cationic “turn” residue (*R*)-2,4-diaminobutyric acid ((*R*)<sup>H<sub>2</sub>N</sup>γ)<sup>34</sup> maintains the overall +1 charge needed for optimal solubility in water.

Polyamide **24**, ImPy-β-ImPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-ImPy-C<sub>2</sub>-OH and **25**, ImPy-β-ImPyPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-PyImPy-C<sub>2</sub>-OH were synthesized using solid-phase methods on glycine-PAM resin, reductively cleaved from the solid support using LiBH<sub>4</sub>,<sup>35</sup> and purified by HPLC (reverse-phase). The identity and purity of polyamides **24** and **25** was confirmed by <sup>1</sup>H NMR, analytical HPLC, and MALDI-TOF MS. MALDI-TOF MS (monoisotopic) (M+H): **24**, obsd 953.3, calcd (C<sub>42</sub>H<sub>53</sub>N<sub>18</sub>O<sub>9</sub>) 953.4; **25**, obsd 1197.5, calcd (C<sub>54</sub>H<sub>65</sub>N<sub>22</sub>O<sub>11</sub>) 1197.5.

Quantitative DNase I footprinting on a 245 base pair 3'-<sup>32</sup>P-end-labeled restriction fragment showed that **24** binds its match site 5'-AGCAGCTGCT-3' at nanomolar concentrations (apparent monomeric association constant, K<sub>a</sub> = 1.9 (±0.3) × 10<sup>8</sup> M<sup>-1</sup>), and also binds a single-base pair mismatch site 5'-AGATGCTGCA-3' with 9-fold lower affinity, K<sub>a</sub> = 2.2 (±0.5) × 10<sup>7</sup> M<sup>-1</sup>. The binding data for match and single-base pair mismatch sites were well-fit by cooperative binding isotherms, consistent with formation of cooperative 2:1 polyamide-DNA complexes. A double base pair mismatch site, 5'-AGCTGCATCC-3', is also bound with 65-fold lower affinity. The fact that this mismatch site, which contains the “half-site” 5'-AGCTGCA-3', is not effectively bound indicates that recognition of the match site occurs through cooperative dimerization, and not due to formation of 1:1 hairpin complexes.

Further study of the generality and sequence specificity of this motif is in progress and will be reported in due course. For example, we found that the eight-ring ImPy-β-ImPyPy-(*R*)<sup>H<sub>2</sub>N</sup>γ-PyImPy-C<sub>2</sub>-OH (**25**) binds the 12 base pair match site 5'-AAGCAGCTGCTT-3' with 10-fold higher affinity than **24**, and is ~100-fold specific for this site versus the double base pair mismatch site 5'-CAGATGCTGCAT-3'.



**Figure 6.27.** Models of complexes of cooperative hairpin polyamides with the site 5'-ATAAGCAGCTGCTTTT-3'. (left) Cooperative hairpins with N-terminal dimerization domain. (right) C-terminal dimerization domain.

It is interesting to note that the DNA-binding affinity and specificity of the six-ring polyamide **24** for its ten base pair binding site are typical of standard six-ring hairpins that recognize five base pairs.<sup>6</sup> Thus, use of a the cooperative hairpin dimer motif doubles the binding site size relative to the standard hairpin motif without sacrificing affinity or specificity, and without increasing the molecular weight of the ligand. The results reported here show that by using a novel cooperative hairpin dimer motif, relatively low molecular weight pyrrole-imidazole polyamides (MW ~ 950-1,200) can specifically recognize 10-12 base pairs of DNA. Such polyamides will be useful in determining the optimal ligand size and optimal binding site size required for specific biological activity.

Four additional cooperative hairpin polyamides were synthesized and their equilibrium association constants for the site 5'-ATAAGCAGCTGCTTTT-3' determined by quantitative DNase I footprinting experiments (Figure 6.27). Out of the series of six polyamides, three are at least 5-fold specific for the match site versus all other sites on the restriction fragment used.

## Experimental Section

All gel electrophoresis and footprinting assays were performed by John Trauger,<sup>6d,23,37</sup> Sue Swalley,<sup>26a,30</sup> and Jim Turner<sup>26b,30</sup> as described elsewhere.

Dicyclohexylcarbodiimide (DCC), Hydroxybenzo-triazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU) and 0.2 mmol/gram Boc- $\beta$ -alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc- $\beta$ -Pam-Resin) were purchased from Peptides International. *N,N*-diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF), *N*-methylpyrrolidone (NMP), DMSO/NMP, Acetic anhydride ( $\text{Ac}_2\text{O}$ ), and 0.0002 M potassium cyanide/pyridine were purchased from Applied Biosystems. Boc- $\gamma$ -aminobutyric acid was from NOVA Biochem, dichloromethane (DCM) and triethylamine (TEA) was reagent grade from EM, thiophenol (PhSH), dimethylaminopropylamine from Aldrich, trifluoroacetic acid (TFA) from Halocarbon, phenol from Fisher, and ninhydrin from Pierce. All reagents were used without further purification.

Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. and were used for filtration of DCU. A shaker for manual solid phase synthesis was obtained from St. John Associates, Inc. Screw-cap glass peptide synthesis reaction vessels (5 mL and 20 mL) with a #2 sintered glass frit were made as described by Kent.<sup>28</sup> <sup>1</sup>H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz in  $\text{DMSO-}d_6$ , with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on either a HP 1090M analytical HPLC or a Beckman Gold system using a RAINEN  $\text{C}_{18}$ , Microsorb MV, 5 $\mu\text{m}$ , 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 reverse phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100  $\mu$ m C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. 18M $\Omega$  water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2  $\mu$ m filtered.

**Resin Substitution.** Resin substitution can be calculated as  $L_{\text{new}}(\text{mmol/g}) = L_{\text{old}} / (1 + L_{\text{old}}(W_{\text{new}} - W_{\text{old}}) \times 10^{-3})$ , where L is the loading (mmol of amine per gram of resin), and W is the weight (gmol<sup>-1</sup>) of the growing polyamide attached to the resin.<sup>29</sup>

**ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp (1).** Polyamide ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp was prepared by machine-assisted solid phase methods as a white powder (17 mg, 56% recovery). HPLC, r.t.: 26.1 min; UV,  $\lambda_{\text{max}}$  ( $\epsilon$ ): 234 nm (39,300), 312 nm (53,200); <sup>1</sup>H NMR (DMSO-d<sub>6</sub>):  $\delta$  10.53 (s, 1 H), 10.27 (s, 1 H), 10.04 (s, 1 H), 9.96 (s, 1 H), 9.94 (s, 1 H), 9.2 (br s, 1 H), 8.08 (m, 3 H), 7.49 (s, 2 H), 7.44 (s, 1 H), 7.31 (d, 1 H,  $J$  = 1.0 Hz), 7.23 (d, 1 H,  $J$  = 1.1 Hz), 7.19 (m, 3 H), 7.10 (s, 1 H), 6.92 (d, 1 H,  $J$  = 1.1 Hz), 6.90 (d, 1 H,  $J$  = 1.1 Hz), 4.01 (s, 3 H), 3.97 (s, 3 H), 3.86 (m, 6 H), 3.82 (m, 6 H), 3.41 (q, 2 H,  $J$  = 6.0 Hz), 3.22 (q, 2 H,  $J$  = 5.9 Hz), 3.13 (q, 2 H,  $J$  = 5.9 Hz), 3.0 (q, 2 H,  $J$  = 5.6 Hz), 2.76 (d, 6 H,  $J$  = 4.8 Hz), 2.37 (m, 4 H), 1.78 (m, 4 H); MALDI-TOF MS: 1223.4 (1223.3 calc. for M+H).

**ImPyPyPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp (2).** The polyamide ImPyPyPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp was prepared by machine-assisted solid phase methods as a white powder (12 mg, 19% recovery). HPLC, r.t.: 29.5 min; UV,  $\lambda_{\text{max}}$  ( $\epsilon$ ): 238 nm (53,900), 312 nm (71,100); <sup>1</sup>H NMR (DMSO-d<sub>6</sub>):  $\delta$  10.46 (s, 1 H), 10.24 (s, 1 H), 9.96 (s, 1 H), 9.90 (m, 5 H), 9.2 (br s, 1 H), 8.25 (m, 1 H), 8.00 (m, 3 H), 7.44 (s, 1 H), 7.39 (s, 1 H), 7.26 (d, 1 H,  $J$  = 1.3 Hz), 7.24 (d, 1 H,  $J$  = 1.5 Hz), 7.20 (m, 2 H), 7.16 (m, 2 H), 7.13 (m, 2 H), 7.11 (d, 1 H,  $J$  = 1.4 Hz), 7.05 (d, 1 H,  $J$  = 1.4 Hz), 7.03 (d, 1 H,  $J$  = 1.5 Hz), 6.93 (d, 1 H,  $J$  = 1.3 Hz), 6.87 (m, 2 H), 6.84 (d, 1 H,  $J$  = 1.5 Hz), 3.97 (s, 3 H), 3.92 (s, 3 H), 3.82 (m, 9 H), 3.79 (m, 6 H), 3.76 (m, 6 H), 3.73 (m, 2 H), 3.44 (q, 2 H,  $J$  = 5.0 Hz), 3.17

(m, 4 H), 3.03 (m, 2 H), 2.74 (d, 6 H,  $J = 4.8$  Hz), 2.50 (m, 2 H) 2.33 (t, 2 H,  $J = 6.7$  Hz), 1.77 (m, 4 H); MALDI-TOF MS: 1222.3 (1222.3 calc for M+H).

**ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp (3).** ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -PAM-Resin was prepared from 0.2 mmol/gram Boc- $\beta$ -PAM-resin by machine assisted synthesis. The  $\gamma$ -Im step was introduced using Boc- $\gamma$ -Im acid (HBTU, DIEA), all other residues were added as appropriate activated Boc protected monomer units. A sample of resin (250 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial, 2 mL dimethylaminopropylamine added and the mixture allowed to stand at 55°C for 18 hours. Resin was removed by filtration through a disposable propylene filter, and the resulting solution diluted with water to a total volume of 8 mL, and purified directly by reversed phase HPLC to provide ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp (26 mg, 45% recovery) as a white powder. UV $\lambda_{\text{max}}$ (H<sub>2</sub>O) 248, 312 (66,000); <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.34 (m, 2 H); 10.32 (m, 2 H); 9.73 (m, 2 H); 9.5 (br s, 1 H), 9.32 (s, 1 H); 8.10 (m, 3 H); 7.55 (m, 2 H); 7.52 (s, 1 H); 7.44 (s, 1 H); 7.23 (m, 2 H), 7.14 (m, 4 H); 7.06 (d, 1 H,  $J = 1.4$  Hz); 6.86 (m, 2 H); 3.98 (m, 9 H); 3.95 (s, 3 H); 3.81 (m, 6 H); 3.77 (m, 6 H); 3.31 (m, 2 H); 3.17 (t, 2 H,  $J = 5.5$  Hz) 3.06 (m, 2 H,  $J = 5.7$  Hz); 2.93 (m, 2 H,  $J = 4.7$  Hz); 2.74 (d, 6 H,  $J = 4.4$  Hz); 2.30 (m, 4 H); 1.74 (m, 4 H); MALDI-TOF-MS, 1224.9 (1225.3 calc. for M+H).

**ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp (4).** ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -PAM-Resin was prepared from 0.2 mmol/gram Boc- $\beta$ -PAM-resin by manual polyamide synthesis. The Py-Im and  $\gamma$ -Im steps were introduced using Boc- $\gamma$ -Im acid and Boc-Py-Im acid (HBTU, DIEA), all other residues were added as appropriate activated Boc protected monomer units. A sample of resin (250 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial, 2 mL dimethylaminopropylamine added and the mixture allowed to stand at 55°C for 18 hours. Resin was removed by filtration through a disposable propylene filter, and the resulting solution diluted with water to a total volume of 8 mL, and purified directly by reversed phase HPLC to provide ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp (19 mg, 32% recovery) as a white powder. UV $\lambda_{\text{max}}$ (H<sub>2</sub>O) 246, 312 (66,000);



$^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  10.33 (m, 2 H); 10.25 (m, 2 H); 10.04 (m, 2 H); 9.95 (s, 1 H); 9.5 (br s, 1 H); 8.10 (m, 3 H); 7.57 (m, 2 H); 7.48 (s, 1 H); 7.42 (s, 1 H); 7.40 (s, 1 H); 7.23 (m, 2 H); 7.17 (d, 1 H;  $J = 1.5$  Hz); 7.03 (d, 1 H,  $J = 1.5$  Hz); 6.98 (m, 3 H); 4.02 (s, 3 H); 3.99 (m, 6 H); 3.97 (s, 3 H); 3.88 (m, 6 H); 3.83 (m, 6 H); 3.42 (m, 2 H); 3.18 (t, 2 H,  $J = 5.2$  Hz) 3.06 (m, 2 H,  $J = 5.5$  Hz); 2.80 (m, 2 H,  $J = 4.7$  Hz); 2.76 (d, 6 H,  $J = 4.4$  Hz); 2.38 (m, 4 H); 1.93 (m, 4 H); MALDI-TOF-MS, 1225.2. (1225.3 calc. for M+H).

**ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp (5).** ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -PAM-Resin was prepared from 0.2 mmol/gram Boc- $\beta$ -PAM-resin by manual polyamide synthesis. All residues were added as appropriate activated Boc protected monomer units. A sample of resin (250 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial, 2 mL dimethylaminopropylamine added and the mixture allowed to stand at 55°C for 18 hours. Resin was removed by filtration through a disposable propylene filter, and the resulting solution diluted with water to a total volume of 8 mL, and purified directly by reversed phase HPLC to provide ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp (12 mg, 21% recovery) as a white powder. UV  $\lambda_{\text{max}}$ (H<sub>2</sub>O) 246, 314 (66,000);  $^1\text{H}$  NMR (DMSO- $d_6$ )  $\delta$  9.91 (m, 2 H); 9.89 (m, 4 H); 9.83 (s, 1 H); 9.60 (s, 1 H); 9.5 (br s, 1 H); 8.34 (m, 1 H); 8.10 (m, 2 H); 7.63 (m, 2 H); 7.50 (s, 1 H); 7.42 (s, 1 H); 7.19 (m, 2 H); 7.13 (m, 2 H); 7.04 (m, 2 H); 6.86 (m, 2 H); 3.98 (m, 6 H); 3.96 (s, 3 H); 3.93 (s, 3 H); 3.81 (m, 6 H); 3.77 (s, 3 H); 3.73 (s, 3 H); 3.30 (m, 2 H); 3.10 (t, 2 H,  $J = 5.3$  Hz) 3.09 (m, 2 H,  $J = 5.5$  Hz); 2.91 (m, 2 H,  $J = 4.6$  Hz); 2.71 (d, 6 H,  $J = 4.2$  Hz); 2.32 (m, 4 H); 1.70 (m, 4 H); MALDI-TOF-MS, 1225.6 (1225.3 calc. for M+H).

**ImImPyPy- $\beta$ -Dp (6).** ImImPyPy- $\beta$ -PAM-Resin was prepared from 0.2 mmol/gram Boc- $\beta$ -PAM-resin by machine assisted synthesis. Residues were added as appropriate activated Boc protected monomer units. A sample of resin (250 mg, 0.18 mmol/gram) was placed in a 20 mL glass scintillation vial, 2 mL dimethylaminopropylamine added and the mixture allowed to stand at 55°C for 18 hours. Resin was removed by filtration through a disposable propylene filter, and the

resulting solution diluted with water to a total volume of 8 mL and purified directly by reversed phase HPLC to provide ImImPyPy- $\beta$ -Dp (14 mg, 54% recovery) as a white powder. UV  $\lambda_{\text{max}}(\text{H}_2\text{O})$  248, 308 (33,000);  $^1\text{H}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  10.35 (s, 1 H); 9.91 (s, 1 H); 9.72 (s, 1 H); 9.3 (br s, 1 H), 9.32 (s, 1 H); 8.03 (m, 2 H); 7.55 (s, 1 H); 7.44 (s, 1 H); 7.24 (d, 1 H,  $J = 1.6$  Hz); 7.14 (d, 1 H,  $J = 1.6$  Hz), 7.11 (d, 1 H,  $J = 1.7$  Hz); 7.05 (d, 1 H,  $J = 1.6$  Hz); 6.84 (d, 1 H,  $J = 1.6$  Hz); 3.97 (s, 3 H); 3.81 (s, 3 H); 3.77 (s, 3 H); 3.45 (q, 2 H,  $J = 5.6$  Hz); 3.36 (q, 2 H,  $J = 6.1$  Hz); 3.07 (q, 2 H,  $J = 6.2$  Hz); 2.70 (d, 6 H,  $J = 4.9$  Hz); 2.31 (t, 2 H,  $J = 6.9$  Hz); 1.70 (quintet, 2 H,  $J = 7.3$  Hz); MALDI-TOF-MS, 649.1 (649.6 calc. for M+H).

**ImPyImPy- $\beta$ -Dp (7).** ImPyImPy- $\beta$ -PAM-Resin was prepared from 0.2 mmol/gram Boc- $\beta$ -PAM-resin by manual polyamide synthesis. The Py-Im step was introduced using Boc-Py-Im acid (HBTU, DIEA), and all other residues were added as appropriate activated Boc protected monomer units. A sample of resin (250 mg, 0.18 mmol/gram) was placed in a 20 mL glass scintillation vial, 2 mL dimethylaminopropylamine added and the mixture allowed to stand at 55°C for 18 hours. Resin was removed by filtration through a disposable propylene filter, and the resulting solution diluted with water to a total volume of 8 mL, and purified directly by reversed phase HPLC to provide ImPyImPy- $\beta$ -Dp (5 mg, 19% recovery) as a white powder. UV  $\lambda_{\text{max}}(\text{H}_2\text{O})$  252, 312 (33,000);  $^1\text{H}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  10.35 (s, 1 H); 10.25 (s, 1 H); 10.05 (s, 1 H); 9.3 (br s, 1 H), 8.10 (m, 2 H); 7.58 (s, 1 H); 7.42 (s, 1 H); 7.40 (d, 1 H,  $J = 1.5$  Hz); 7.24 (d, 1 H,  $J = 1.5$  Hz), 7.17 (d, 1 H;  $J = 1.5$  Hz); 7.07 (d, 1 H;  $J = 1.5$  Hz); 6.93 (d, 1 H,  $J = 1.5$  Hz); 4.01 (s, 3 H); 3.99 (s, 3 H); 3.88 (s, 3 H); 3.82 (s, 3 H); 3.41 (q, 2 H,  $J = 5.8$  Hz); 3.31 (m, 2 H); 3.03 (m, 2 H); 2.73 (d, 6 H,  $J = 4.4$  Hz); 2.32 (t, 2 H,  $J = 6.1$  Hz); 1.74 (m, 2 H); MALDI-TOF-MS, 650.3 (649.6 calc. for M+H).

**ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp-NH<sub>2</sub> (3-NH<sub>2</sub>).** A sample of ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -PAM-Resin (250 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial, 2 mL 3,3'-diamino-*N*-methyldipropylamine was added and the mixture allowed to stand at 55°C for 18

hours. Resin was removed by filtration through a disposable propylene filter, and the resulting solution diluted with water to a total volume of 8 mL, and purified directly by reversed phase HPLC to provide ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp-NH<sub>2</sub> (30 mg, 54% recovery) as a white powder.

<sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.35 (m, 2 H); 10.32 (m, 2 H); 9.71 (m, 2 H); 9.6 (br s, 1 H), 9.30 (s, 1 H); 8.0 (m, 3 H); 7.8 (br s, 3 H); 7.55 (m, 2 H); 7.48 (s, 1 H); 7.41 (s, 1 H); 7.22 (d, 2 H, *J* = 1.2 Hz), 7.14 (m, 4 H); 7.04 (d, 1 H, *J* = 1.2 Hz); 6.86 (d, 2 H, *J* = 1.3 Hz); 4.00 (m, 9 H); 3.94 (s, 3 H); 3.80 (m, 12 H); 3.57 (q, 2 H, *J* = 4.5 Hz), 3.2- 3.0 (m, 8 H), 2.83 (q, 2 H, *J* = 4.5 Hz), 2.71 (d, 3 H, *J* = 4.8 Hz), 2.46 (t, 2 H, *J* = 6.0 Hz), 2.32 (t, 2 H, *J* = 5.9 Hz), 1.87 (quintet, 2 H, *J* = 5.7 Hz), 1.76 (m, 4 H). MALDI-TOF-MS, 1267.9 (1268.4 calc. for M+H).

**ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp-NH<sub>2</sub> (4-NH<sub>2</sub>).** A sample of ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -PAM-Resin (250 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial, 2 mL 3,3'-diamino-*N*-methyldipropylamine added and the mixture allowed to stand at 55°C for 18 hours. Resin was removed by filtration through a disposable propylene filter, and the resulting solution diluted with water to a total volume of 8 mL, and purified directly by reversed phase HPLC to provide ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp-NH<sub>2</sub> (24 mg, 43% recovery) as a white powder. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.33 (m, 2 H); 10.22 (m, 2 H); 10.01 (m, 2 H); 9.96 (s, 1 H); 9.8 (br s, 1 H), 8.0 (m, 3 H); 7.8 (br s, 3H); 7.55 (m, 2 H); 7.48 (s, 1 H); 7.41 (s, 1 H); 7.23 (d, 2 H; *J* = 1.6 Hz), 7.17 (d, 1 H; *J* = 1.4 Hz); 7.02 (d, 1 H, *J* = 1.6 Hz); 6.96 (m, 3 H); 4.03 (s, 3 H); 3.99 (m, 6 H); 3.96 (s, 3 H); 3.90 (m, 6 H); 3.81 (m, 6 H); 3.51 (q, 2 H, *J* = 5.1 Hz); 3.2-3.0 (m, 8 H); 2.80 (q, 2 H, *J* = 5.1 Hz), 2.71 (d, 3 H, *J* = 4.6 Hz), 2.41 (t, 2 H, *J* = 5.9 Hz), 2.28 (t, 2 H, *J* = 5.9 Hz), 1.88 (quintet, 2 H, *J* = 6.2 Hz), 1.73 (m, 4 H). MALDI-TOF-MS, 1268.0 (1268.4 calc. for M+H).

**ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (5-NH<sub>2</sub>).** A sample of ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -PAM-Resin (250 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial, 2 mL 3,3'-diamino-*N*-methyldipropylamine added and the mixture allowed to stand at 55°C for 18 hours. Resin was removed by filtration through a disposable propylene filter, and the resulting solution

diluted with water to a total volume of 8 mL and purified directly by preparatory reversed phase HPLC to provide ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp (18 mg, 17% recovery) as a white powder. <sup>1</sup>H NMR (DMSO-d<sub>6</sub>)  $\delta$  9.90 (m, 4 H); 9.81 (s, 1 H); 9.60 (s, 1 H); 9.5 (br s, 1 H); 8.30 (m, 1 H); 8.0 (m, 2 H); 7.8 (br s, 3H) 7.65 (m, 2 H); 7.55 (s, 1 H); 7.40 (s, 1 H); 7.17 (m, 2 H), 7.10 (m, 2 H); 7.03 (m, 2 H); 6.80 (m, 2 H); 4.01 (m, 6 H); 3.95 (s, 3 H); 3.90 (s, 3 H); 3.80 (m, 6 H); 3.77 (s, 3 H); 3.73 (s, 3 H); 3.55 (q, 2 H, J = 5.4 Hz); 3.2-3.0 (m, 8 H); 2.84 (q, 2 H, J = 5.4 Hz), 2.71 (d, 3 H, J = 4.8 Hz), 2.44 (t, 2 H, J = 5.8 Hz), 2.20 (t, 2 H, J = 6.0 Hz), 1.86 (quintet, 2 H J = 6.1 Hz), 1.76 (m, 4 H). MALDI-TOF-MS, 1267.4 (1268.4 calc. for M+H).

**ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp-EDTA (3-E).** EDTA-dianhydride (50 mg) was dissolved in 1 mL DMSO/NMP solution and 1 mL DIEA by heating at 55 °C for 5 min. The dianhydride solution was added to ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -Dp-NH<sub>2</sub> (**3-NH<sub>2</sub>**) (10 mg, 7  $\mu$ mol) dissolved in 750  $\mu$ L DMSO. The mixture was heated at 55 °C for 25 min., and treated with 3 mL 0.1M NaOH, and heated at 55 °C for 10 min. 0.1% TFA was added to adjust the total volume to 8 mL and the solution purified directly by preparatory HPLC chromatography to provide **3-E** as a white powder. (3 mg, 26% recovery) MALDI-TOF-MS, 1543.0 (1543.6 calc. for M+H).

**ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp-EDTA (4-E).** Polyamide **4-E** was prepared from ImPyImPy- $\gamma$ -ImPyImPy- $\beta$ -Dp-NH<sub>2</sub> (**4-NH<sub>2</sub>**) (10 mg, 7  $\mu$ mol) as described for **3-E**. (1 mg, 9% recovery) MALDI-TOF-MS, 1543.2 (1543.6 calc. for M+H).

**ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp-EDTA (5-E).** Polyamide **5-E** was prepared from ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (**5-NH<sub>2</sub>**) (10 mg, 7  $\mu$ mol) as described for **3-E**. (2.5 mg, 22% recovery) MALDI-TOF-MS, 1542.6 (1543.6 calc. for M+H).

**ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (8).** ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -resin was prepared by machine-assisted solid phase synthesis on a 430A Applied Biosystems peptide synthesizer. A sample of resin (240 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial, and treated with dimethylaminopropylamine (2 mL) at 55°C for 18 hours. Resin was removed by

filtration, and the filtrate diluted to a total volume of 8 mL with 0.1% (wt/v) aqueous TFA. The resulting crude polyamide/amine solution was purified directly by reversed phase HPLC to provide the trifluoroacetate salt of ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (13 mg, 22% recovery) as a white powder. UV (H<sub>2</sub>O)  $\lambda_{\text{max}}$  250, 316 ( $\epsilon$ ) 83300 (calculated based on  $\epsilon = 8,333/\text{ring}^{8f}$ ); <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.52 (s, 1 H), 10.29 (s, 1 H), 10.04 (s, 1 H), 10.00 (s, 1 H), 9.97 (m, 3 H), 9.92 (m, 2 H), 9.2 (br s, 1 H), 8.06 (m, 3 H), 7.46 (s, 1 H), 7.41 (s, 1 H), 7.29 (d, 1 H, *J* = 1.5 Hz), 7.27 (d, 1 H), 7.23 (m, 4 H), 7.17 (m, 4 H), 7.07 (m, 5 H), 6.90 (d, 1 H), 6.88 (d, 1 H), 3.99 (s, 3 H), 3.94 (s, 3 H), 3.85 (m, 18 H), 3.79 (s, 6 H), 3.38 (q, 2 H, *J* = 6.0 Hz), 3.20 (q, 2 H, *J* = 5.1 Hz), 3.11 (q, 2 H, *J* = 6.3 Hz), 3.00 (q, 2 H, *J* = 4.7 Hz), 2.72 (d, 6 H, *J* = 4.8 Hz), 2.35 (m, 4 H), 1.75 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1466.0 (1466.7 calc. for M+H).

**ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (9).** ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Pam-Resin was prepared as described for **8**. A sample of resin (240 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial and treated with dimethylaminopropylamine (2 mL) at 55°C for 18 hours. Resin was removed by filtration, and the filtrate diluted to a total volume of 8 mL with 0.1% (wt/v) aqueous TFA. The resulting crude polyamide/amine solution was purified directly by reversed phase HPLC to provide the trifluoroacetate salt of ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp as a white powder (11 mg, 18% recovery). UV (H<sub>2</sub>O)  $\lambda_{\text{max}}$  246, 318 ( $\epsilon$ ) 83300 (calculated based on  $\epsilon = 8,333/\text{ring}^{8f}$ ); <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.38 (s, 1 H), 10.28 (s, 1 H), 10.02 (s, 1 H), 9.99 (s, 1 H), 9.96 (s, 1 H), 9.95 (s, 1 H), 9.92 (s, 1 H), 9.91 (s, 2 H), 9.76 (s, 1 H), 9.2 (br s, 1 H), 8.05 (m, 3 H), 7.57 (s, 1 H), 7.46 (m, 2 H), 7.28 (d, 1 H), 7.26 (d, 1 H), 7.16 (m, 4 H), 7.08 (m, 5 H), 6.89 (d, 2 H), 6.88 (d, 2 H), 4.00 (s, 6 H), 3.94 (s, 3 H), 3.85 (m, 15 H), 3.79 (s, 6 H), 3.54 (m, 2 H), 3.19 (m, 2 H), 3.12 (q, 2 H, *J* = 5.7 Hz), 2.99 (q, 2 H, *J* = 5.1 Hz), 2.73 (d, 6 H, *J* = 4.8 Hz), 2.34 (m, 4 H), 1.75 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1466.9 (1466.6 calc. for M+H).

**ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (8-NH<sub>2</sub>).** A sample of ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -resin resin (240 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial, and treated

with 3,3-diamino-*N*-methyldipropylamine (2 mL) at 55 °C for 18 hours. Resin was removed by filtration, and the filtrate diluted to a total volume of 8 mL with 0.1 % (wt/v) aqueous TFA. The resulting crude polyamide/amine solution was purified directly by reversed phase HPLC to provide the trifluoroacetate salt of ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (31 mg, 50% recovery) as a white powder. UV  $\lambda_{\text{max}}$  241, 316; <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.53 (s, 1 H), 10.28 (s, 1 H), 10.03 (s, 1 H), 9.99 (s, 1 H), 9.96 (m, 3 H), 9.92 (m, 2 H), 9.6 (br s, 1 H), 8.09 (t, 2 H, *J* = 6.1 Hz), 8.07 (m, 2 H), 7.9 (br s, 3 H), 7.45 (s, 1 H), 7.41 (s, 1 H), 7.29 (d, 1 H), 7.26 (d, 1 H), 7.23 (m, 4 H), 7.16 (m, 5 H), 7.08 (m, 2 H), 7.06 (m, 2 H), 6.89 (d, 1 H), 6.87 (d, 1 H), 3.98, (s, 3 H), 3.94 (s, 3 H), 3.84, (m, 18 H), 3.79 (s, 6 H), 3.35 (q, 2 H, *J* = 6.3 Hz), 3.2-3.0 (m, 8 H), 2.85 (q, 2 H, *J* = 5.6 Hz), 2.72 (d, 3 H, *J* = 4.2 Hz), 2.34 (m, 4 H), 1.91 (quintet, 2 H, *J* = 7.3 Hz), 1.78 (m, 4 H). MALDI-TOF MS (monoisotopic), 1509.8 (1509.7 calc. for M+H).

**ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (9-NH<sub>2</sub>).** A sample of ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Pam resin resin (240 mg, 0.16 mmol/gram) was placed in a 20 mL glass scintillation vial, and treated with 3,3-diamino-*N*-methyldipropylamine (2 mL) at 55°C for 18 hours. Resin was removed by filtration, and the filtrate diluted to a total volume of 8 mL with 0.1% (wt/v) aqueous TFA. The resulting crude polyamide/amine solution was purified directly by reversed phase HPLC to provide the trifluoroacetate salt of ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (29 mg, 47% recovery). <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.39 (s, 1 H), 10.28 (s, 1 H), 10.03 (s, 1 H), 10.01 (s, 1 H), 10.00 (s, 1 H), 9.97 (s, 1 H), 9.96 (s, 1 H), 9.92 (m, 1 H), 9.82 (s, 1 H), 9.7 (br s, 1 H), 8.11 (t, 1 H, *J* = 5.5 Hz), 8.07 (m, 2 H), 7.9 (br s, 3 H), 7.57 (s, 1 H), 7.46 (s, 1 H), 7.45 (s, 1 H), 7.28 (d, 1 H, *J* = 1.6 Hz), 7.26 (d, 1 H, *J* = 1.1 Hz), 7.23 (m, 2 H), 7.22 (d, 1 H, *J* = 1.3 Hz), 7.16 (m, 4 H), 7.09 (m, 2 H), 7.07 (d, 1 H, *J* = 1.5 Hz), 7.06 (d, 1 H, *J* = 1.6 Hz), 6.89 (d, 1 H, *J* = 1.4 Hz), 6.87 (d, 1 H), 4.00 (s, 3 H), 3.99 (s, 3 H), 3.94 (s, 3 H), 3.84 (m, 12 H), 3.83 (s, 3 H), 3.79 (m, 6 H), 3.35 (q, 2 H, *J* = 6.1 Hz), 3.2 - 3.0 (m, 8 H), 2.86 (q, 2 H, *J* = 6.1 Hz), 2.72 (d, 2 H, *J* = 4.4 Hz),

2.34 (m, 4 H), 1.90 (quintet, 2 H,  $J = 7.0$  Hz), 1.78 (m, 4 H). MALDI-TOF-MS (monoisotopic), 1510.9 (1510.7 calc. for M+H).

**ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-EDTA (8-E).** EDTA-dianhydride (50 mg) was dissolved by heating at 55°C for 5 min. in a solution of DMSO/NMP (1 ml) and DIEA (1 mL). The dianhydride solution was added to ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (**8-NH<sub>2</sub>**) (10 mg, 6  $\mu$ mol) dissolved in DMSO (750  $\mu$ L). The mixture was heated at 55°C for 25 minutes, and treated with 0.1M NaOH (3 mL), and heated at 55°C for 10 minutes. Aqueous 0.1% (wt/v) TFA was added to adjust the total volume to 8 mL and the solution purified directly by preparatory HPLC chromatography to provide **8-E** as a white powder. (2.8 mg, 25% recovery) MALDI-TOF-MS (monoisotopic), 1784.7 (1783.8 calc. for M+H).

**ImImPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp-EDTA (9-E).** Compound **9-E** was prepared as a white powder as described for compound **8-E**. (6 mg, 42% recovery). MALDI-TOF-MS (monoisotopic), 1785.2 (1784.8 calc. for M+H).

**ImPyPy- $\gamma$ -ImPyPy- $\beta$ -Dp (22).** Polyamide was prepared by machine-assisted solid phase methods as a white powder. (17 mg, 56% recovery). HPLC r.t. 26.1, UV  $\lambda_{\text{max}}$  ( $\epsilon$ ), 234 (39,300), 312 (53,200) nm; <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>);  $\delta$  10.53 (s, 1 H), 10.27 (s, 1 H), 10.04 (s, 1 H), 9.96 (s, 1 H), 9.94 (s, 1 H), 9.2 (br s, 1 H), 8.08 (m, 3 H), 7.49 (s, 2 H), 7.44 (s, 1 H), 7.31 (d, 1 H,  $J = 1.0$  Hz), 7.23 (d, 1 H,  $J = 1.1$  Hz), 7.19 (m, 3 H), 7.10 (s, 1 H), 6.92 (d, 1 H,  $J = 1.1$  Hz), 6.90 (d, 1 H,  $J = 1.1$  Hz), 4.01 (s, 3 H), 3.97 (s, 3 H), 3.86 (m, 6 H), 3.82 (m, 6 H), 3.41 (q, 2 H,  $J = 6.0$  Hz), 3.22 (q, 2 H,  $J = 5.9$  Hz), 3.13 (q, 2 H,  $J = 5.9$  Hz), 3.0 (q, 2 H,  $J = 5.6$  Hz), 2.76 (d, 6 H,  $J = 4.8$  Hz), 2.37 (m, 4 H), 1.78 (m, 4 H); MALDI-TOF MS 979.3 (979.1 calc. for M+H).

**ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp (23)** Polyamide was prepared by machine-assisted solid phase methods as a white powder. (12 mg 19% recovery). HPLC r.t. 29.5, UV  $\lambda_{\text{max}}$  ( $\epsilon$ ), 238 (53,900), 312 (71,100) nm; <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>);  $\delta$  10.46 (s, 1 H), 10.24 (s, 1 H), 9.96 (s, 1 H), 9.90 (m, 5 H), 9.2 (br s, 1 H), 8.25 (m, 1 H), 8.00 (m, 3 H), 7.44 (s, 1 H), 7.39 (s, 1 H), 7.26 (d, 1

H,  $J = 1.3$  Hz), 7.24 (d, 1 H,  $J = 1.5$  Hz), 7.20 (m, 2 H), 7.16 (m, 2 H), 7.13 (m, 2 H), 7.11 (d, 1 H,  $J = 1.4$  Hz), 7.05 (d, 1 H,  $J = 1.4$  Hz), 7.03 (d, 1 H,  $J = 1.5$  Hz), 6.93 (d, 1 H,  $J = 1.3$  Hz), 6.87 (m, 2 H), 6.84 (d, 1 H,  $J = 1.5$  Hz), 3.97 (s, 3 H), 3.92 (s, 3 H), 3.82 (m, 9 H), 3.79 (m, 6 H), 3.76 (m, 6 H), 3.73 (m, 2 H), 3.44 (q, 2 H,  $J = 5.0$  Hz), 3.17 (m, 4 H), 3.03 (m, 2 H), 2.74 (d, 6 H,  $J = 4.8$  Hz), 2.50 (m, 2 H), 2.33 (t, 2 H,  $J = 6.7$  Hz), 1.77 (m, 4 H). MALDI-TOF MS 1402.2 (1402.5 calc for M+H).

**ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp-NH<sub>2</sub> (23-NH<sub>2</sub>).** Polyamide was prepared by machine-assisted solid phase methods as a white powder. (29 mg 59% recovery). HPLC r.t. 21.5, <sup>1</sup>H NMR (DMSO-d<sub>6</sub>);  $\delta$  10.50 (s, 1 H), 10.27 (s, 1 H), 9.96 (s, 1 H), 9.93 (m, 5 H), 9.2 (br s, 1 H), 8.27 (t, 1 H,  $J = 5.1$  Hz), 8.03 (m, 3 H), 7.90 (s, 3 H), 7.45 (s, 1 H), 7.40 (s, 1 H), 7.27 (d, 1 H,  $J = 1.3$  Hz), 7.25 (d, 1 H,  $J = 1.4$  Hz), 7.22 (m, 2 H), 7.18 (m, 2 H), 7.17 (d, 1 H,  $J = 1.4$  Hz), 7.14 (d, 1 H,  $J = 1.3$  Hz), 7.11 (m, 2 H), 7.06 (d, 1 H,  $J = 1.5$  Hz), 6.94 (d, 1 H,  $J = 1.3$  Hz), 6.88 (m, 2 H), 6.84 (d, 1 H,  $J = 1.4$  Hz), 3.97 (s, 3 H), 3.93 (s, 3 H), 3.83 (m, 9 H), 3.80 (m, 6 H), 3.76 (m, 6 H), 3.72 (d, 2 H,  $J = 5.2$  Hz), 3.43 (q, 2 H,  $J = 5.0$  Hz), 3.17 (m, 6 H), 3.11 (q, 2 H,  $J = 5.3$  Hz), 2.85 (q, 2 H,  $J = 5.2$  Hz), 2.73 (d, 3 H,  $J = 3.9$  Hz), 2.51 (t, 2 H,  $J = 6.5$  Hz), 2.35 (t, 2 H,  $J = 6.7$  Hz), 1.92 (quintet, 2 H,  $J = 6.8$  Hz), 1.78 (m, 4 H). MALDI-TOF MS 1445.6 (1445.6 calc for M+H).

**ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp-EDTA (23-E).** EDTA-dianhydride (50 mg) was dissolved in 1 mL DMSO/NMP solution and 1 mL DIEA by heating at 55°C for 5 min. The dianhydride solution was added to ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp-NH<sub>2</sub> (9.0 mg, 5  $\mu$ mol) dissolved in 750  $\mu$ L DMSO. The mixture was heated at 55°C for 25 min., and treated with 3 mL 0.1M NaOH, and heated at 55°C for 10 min. 0.1% TFA was added to adjust the total volume to 8 mL and the solution purified directly by reversed-phase HPLC to provide ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy-G-Dp-EDTA as a white powder. (3 mg, 30% recovery after HPLC purification); MALDI-TOF MS 1720.1 (1719.8 calc for M+H).



## References

1. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385. (d) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (e) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.
2. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingham, M. *Nature Struct. Biol.* **1994**, *1*, 169. (d) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532.
3. Pilch, D. S.; Poklar, N. A.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306.
4. Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nature Struct. Biol.* **1988**, *5*, 104.
5. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663. (b) Chen, Y. H.; Lown, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 6995.
6. (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983. (b) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147. (c) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6153. (d) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Chem. & Biol.* **1996**, *3*, 369. (e) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198. (f) de Claire, R. P. L.; Geierstanger B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909. (g) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756. (h) White, S.; Baird, E. E.; Dervan, P. B. *Chem & Biol.* **1997**, *4*, 569.
7. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.
8. (a) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods Enzymol.* **1986**, *130*, 132. (b) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8462. (c) Senear, D.F.; Brenowitz, M.; Shea, M.A.; Ackers, G.K. *Biochemistry* **1986**, *25*, 7344.
9. Steitz, T.A. *Quart. Rev. Biophys.* **1990**, *23*, 205.
10. Desjarlais, J.R.; Berg, J.M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7345.
11. Desjarlais, J.R.; Berg, J.M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2256.
12. Jamieson, A.C.; Kim, S.-H.; Wells, J.A. *Biochemistry* **1994**, *33*, 5689.
13. Rebar, R.J.; Pabo, C.O. *Science* **1994**, *263*, 671.
14. Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11163.
15. Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11168

16. Letovsky, J.; Dynan, W.S. *Nucleic Acids Res.* **1989**, *17*, 2639.
17. (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science*, **1994**, *266*, 646. (b) Mrksich, M.; Dervan, P. B.; *J. Am. Chem. Soc.* **1995**, *117*, 3325.
18. Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981.
19. (a) Van Dyke, M.W.; Dervan, P.B. *Biochemistry* **1983**, *22*, 2373. (b) Van Dyke, M.W.; Dervan, P.B. *Nucl. Acids Res.* **1983**, *11*, 5555.
20. (a) Schultz, P.G.; Taylor, J.S.; Dervan, P.B. *J. Am. Chem. Soc.* **1982**, *104*, 6861. (b) Schultz, P.G.; Dervan, P.B. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1133. (c) Taylor, J.S.; Schultz, P.B.; Dervan, P.B. *Tetrahedron* **1984**, *40*, 457.
21. Gottesfield, J. M.; Nealy, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1997**, *387*, 202.
22. Pullman, B. *Advances Drug Res.* **1989**, *18*, 1.
23. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559.
24. (a) Wu, H.; Crothers, D. M.; *Nature*, **1984**, *308*, 509.; (b) Steitz, T. A. *Ann. Rev. Biophys.* **1990**, *23*, 205. (c) Goodsell, D. S.; Kopka, M. L.; Cascio, D.; Dickerson, R. E. *Proc. Natl. Sci. U.S.A.* **1993**, *90*, 2930. (d) Paoletta, D. N.; Palmer, R.; Schepartz, A. *Science* **1994**, *264*, 1130. (e) Kahn, J. D.; Yun, E.; Croothers, D. M. *Nature* **1994**, *368* 163. (f) Geierstanger, B. H.; Wemmer, D. E. *Ann. Rev. Biochem.* **1995**, *24*, 463. (g) Hansen, M. R.; Hurley, L. H. *Acc. Chem. Res.* **1996**, *29*, 249.
25. (a) Koopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553. (b) Coll, M.; Fredrick, C. A.; Wang, A. H.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8385.
26. (a) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953. (b) Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 7636.
27. (a) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160. (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1997**, *3*, 1600. (c) Trauger, J.W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534.
28. Kent, S.B.H. *Annu. Rev. Biochem.* **1988**, *57*, 957.
29. Barlos, K.; Chatzi, O.; Gatos, D.; Stravropoulos, G. *Int. J. Peptide Protein Res.* **1991**, *37*, 513.
30. Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219.
31. Dervan, P.B. *Science* **1986**, *232*, 464.

32. (a) Ptashne, M. *A Genetic Switch*, Blackwell Scientific Publications and Cell Press: Palo Alto, CA, **1986**; (b) Pabo, C.O.; Sauer, R.T. *Ann. Rev. Biochem.* **1992**, *61*, 1053; (c) Marmorstein, R.; Carey, M.; Ptashne, M.; Harrison, S.C. *Nature* **1992**, *356*, 408; (d) Klemm, J.D.; Rould, M.A.; Aurora, R.; Herr, W.; Pabo, C.O. *Cell* **1994**, *77*, 21; (e) Bellon, S.F.; Rodgers, K.K.; Schatz, D.G.; Coleman, J.E.; Steitz, T.A. *Nature Struct. Biol.* **1997**, *4*, 586.
33. (a) Jones, K.A.; Peterlin, B.M. *Ann. Rev. Biochem.* **1994**, *63*, 717–743; (b) Frech, K.; Brack-Werner, R.; Werner, T. *Virology* **1996**, *224*, 256–267.
34. Herman, D.M.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1998**, *120*, 1382.
35. (a) Mitchell, A.R.; Kent, S.B.; Engelhard, M.; Merrifield, R.B.J. *J. Org. Chem.* **1978**, *43*, 2845; (b) Stewart, J.M.; Young, J.D. *Solid Phase Peptide Synthesis*, Pierce Chemical Company, Rockford, IL, **1984**.
36. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Angew. Chemie. Int. Ed. Eng.* **1998** *37*, 1421.

## CHAPTER 7

### Regulation of Gene Expression by Cell-Permeable DNA-Binding Small Molecules

**Abstract** *Small molecules that target specific DNA sequences offer a potentially general approach for the control of gene-expression. Ligands designed for therapeutic applications must bind any predetermined DNA sequence with high affinity and permeate living cells. Synthetic polyamides containing Im and Py aromatic amino acids have affinities and specificities comparable to DNA-binding proteins. We report here that an eight-ring hairpin polyamide targeted to a specific region of the transcription factor TFIIIA binding site interferes with 5S RNA gene expression in Xenopus kidney cells. The results suggest that Py-Im polyamides are cell-permeable and can inhibit the transcription of specific genes.*

*Multiple cellular DNA-binding transcription factors are required by HIV-1 for RNA synthesis. Two Py-Im polyamides were designed to bind DNA sequences immediately adjacent to binding sites for the transcription factors Ets-1, LEF-1, and TBP. These synthetic ligands specifically inhibit DNA-binding of each transcription factor and HIV-1 transcription in cell-free assays. When used in combination, the polyamides inhibit virus replication by greater than 99% in isolated human peripheral blood lymphocytes, with no detectable cell toxicity. The ability of small molecules to target predetermined DNA sequences located within RNA polymerase II promoters suggests a general approach for regulation of gene expression, as well as a mechanism for the inhibition of viral replication.*

**Publications:** Gottesfield, Nealy, Trauger, Baird & Dervan *Nature* **1997**, 387, 202.

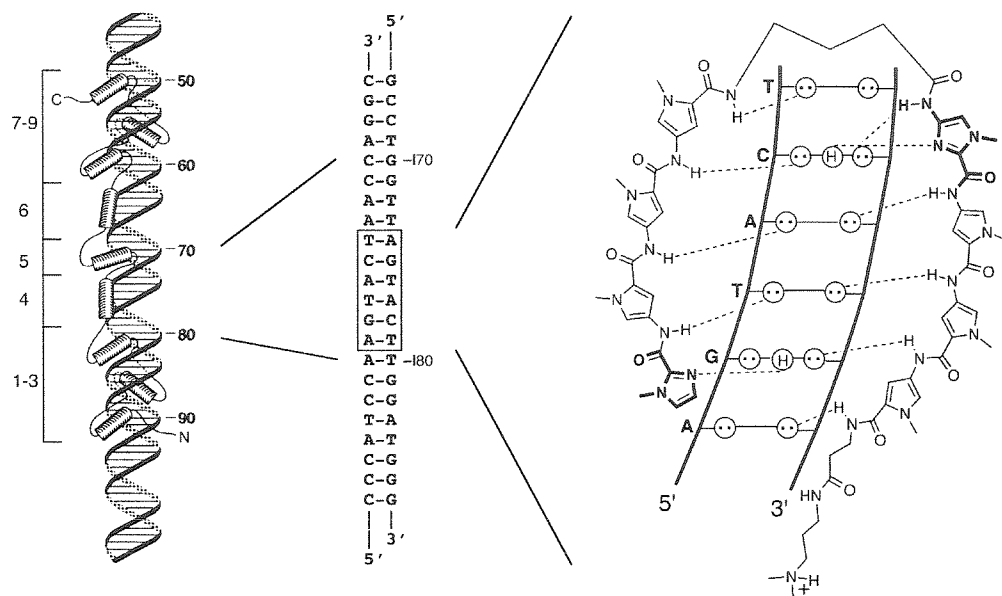
Nealy, Trauger, Baird, Dervan & Gottesfeld *J. Mol. Biol.* **1997**, 274, 439.

Dickinson, Gulizia, Trauger, Baird, Mosier, Gottesfeld & Dervan *Proc. Natl. Acad. Sci. U.S.A.* *in press*.

Two approaches for the development of synthetic transcriptional antagonists have been reported.<sup>1-3</sup> Oligodeoxynucleotides which recognize the major groove of double helical DNA via triple helix formation bind a broad sequence repertoire with high affinity and specificity.<sup>1</sup> Although oligonucleotides and their analogs have been shown to interfere with gene expression,<sup>2</sup> the triple helix approach is limited to purine tracks and suffers from poor cellular uptake. There are a few examples of cell-permeable carbohydrate-based ligands that interfere with transcription factor function.<sup>3</sup> However, oligosaccharides are not yet amenable to recognition of a broad range of DNA sequences.

Py-Im polyamides represent the only class of small molecules to date that can bind any predetermined DNA sequence. DNA recognition depends on side-by-side amino acid pairings in the minor groove.<sup>4</sup> A pairing of imidazole (Im) opposite pyrrole (Py) targets a G•C base pair, while Py/Im targets a C•G base pair.<sup>5</sup> A Py/Py combination is degenerate and targets both T•A and A•T base pairs.<sup>4,5</sup> The generality of these pairing rules has been demonstrated by targeting a variety of sequences 5-13 base pairs in size and is supported directly by NMR and x-ray structural studies.<sup>4-9</sup>

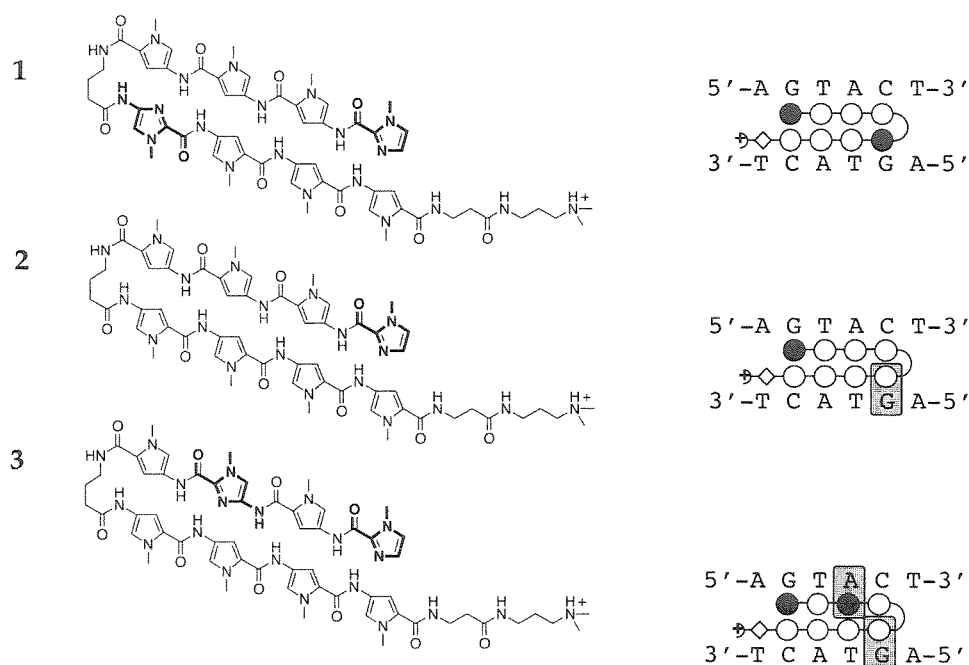
The relatively small number of genes transcribed by RNA polymerase III provides a simple system to explore the selectivity and efficiency of polyamides as regulators of gene expression. Well established methods exist for assessing in living cells the status of RNA polymerase III transcription complexes on the genes encoding the small 5S ribosomal RNA.<sup>10</sup> The role of the nine zinc finger protein TFIIIA in transcriptional regulation of the 5S RNA gene by RNA polymerase III has been extensively characterized.<sup>11</sup> Zinc fingers 1-3, 5, and 7-9 bind the internal control region (ICR) of the gene through base-specific interactions in the major groove.<sup>12</sup> Fingers 4 and 6 are essential for high affinity DNA binding and bind in or across the minor groove<sup>12</sup> (Figure 7.1). The six base pair DNA sequence 5'-AGTACT-3' is within the binding site for finger 4. According to the pairing rules, this sequence would be bound in a hairpin motif by the eight-ring polyamide **1** of sequence composition ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp.



**Figure 7.1.** (a) (left) Model of the nine zinc finger protein TFIIIA with the 5S RNA gene internal control region (ICR). (middle) Sequence of the ICR recognized by zinc finger 4 in the minor groove. (right) Complex of hairpin polyamide **1** with its target site, 5'-AGTACT-3'. Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine.

Quantitative footprint titration experiments reveal that polyamide **1** selectively binds the six base pair target sequence with a dissociation constant,  $K_d = 0.03 \text{ nM}^9$ , a higher affinity than TFIIIA for its 50 base pair site ( $K_d \sim 1 \text{ nM}$ )<sup>12</sup>. Mismatch eight-ring polyamides **2** and **3** have 100-fold ( $K_d = 2.0 \text{ nM}$ ) and 1000-fold ( $K_d = 33 \text{ nM}$ ) lower affinities, respectively, for the 5'-AGTACT-3' site (Figure 1).

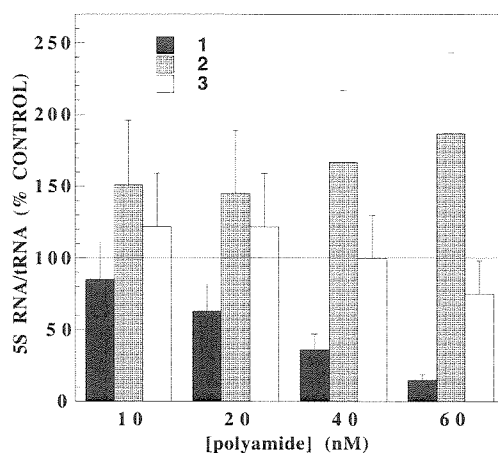
The effect of polyamide **1** on TFIIIA binding to a restriction fragment isolated from a 5S RNA gene-containing plasmid was examined. Zf1-3, a recombinant TFIIIA analog missing fingers 4-9, binds in the major groove of the C-block promoter element<sup>12</sup> (see Figure 7.1). DNase I footprinting demonstrates that zf1-3 and polyamide **1** can co-occupy the same DNA molecule. When 5 nM polyamide **1** was pre-incubated with the same DNA target, the binding of nine finger TFIIIA was inhibited by >90%. The differential inhibition of zf1-3 and full-length TFIIIA



**Figure 7.2.** (left) Structures of polyamides ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp (1), ImPyPyPy-PyPyPyPy- $\beta$ -Dp (2), and ImPyImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp (3). (Dp = dimethylaminopropylamide). (c) Binding models. Red and blue circles represent imidazole and pyrrole rings, respectively, the curved line represents  $\gamma$ -aminobutyric acid, and the diamond represents  $\beta$ -alanine. Hydrogen bond mismatches are highlighted.

provides evidence that finger 4 interacts with or is placed in the minor groove. Polyamide 1 does not inhibit TFIID binding to 5S RNA.

Transcription of the 5S RNA gene in an *in vitro* system was monitored in the presence of increasing concentrations (10–60 nM) of polyamide 1. In these experiments, polyamide 1 was added to a 5S RNA gene-containing plasmid prior to the addition of exogenous TFIID (12 nM) and a crude extract derived from unfertilized *Xenopus* eggs. As a control, a tyrosine tRNA gene was included on a separate plasmid in these reactions. The tRNA gene has an upstream binding site for 1, but lacks a predicted protein-polyamide interaction.<sup>13</sup> Both genes are actively transcribed in this system, either individually or in mixed template reactions. Addition of 60 nM polyamide 1 inhibits 5S gene transcription by >80%. Only a small degree of non-specific

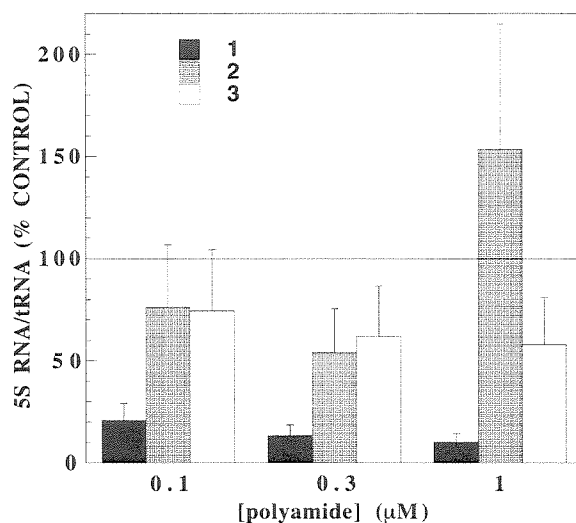


**Figure 7.3.** Inhibition of 5S RNA gene transcription *in vitro*. DNA templates were incubated with polyamide for 30 min. prior to the addition of TFI<sub>IIA</sub> and a cytoplasmic extract prepared from unfertilized *Xenopus* eggs. Reactions contained either the somatic-type 5S RNA gene, a tRNA<sup>tyrD</sup> gene, or a mixture of both genes. Transcription reactions were for 1 h at ambient temperature and RNA products were analyzed on a denaturing polyacrylamide gel. Graphic representation of inhibition results are expressed as the ratio of 5S RNA to tRNA transcription relative to the ratio obtained in the absence of polyamide. Error bars represent estimated standard deviation.

inhibition of tRNA transcription is observed at the concentrations of polyamide **1** required for efficient 5S RNA inhibition (Figure 7.3). The targeted 5S RNA gene is inhibited approximately 10-fold more effectively than the control tRNA gene. Mismatch polyamides **2** and **3** do not inhibit 5S RNA transcription at concentrations up to 60 nM. If the TFI<sub>IIA</sub>-DNA complex is first allowed to form, 30 nM polyamide **1** added, and the mixture incubated for 90 min. prior to adding egg extract, efficient inhibition (80%) of 5S RNA transcription is also observed. Shorter incubation times result in less inhibition. The required incubation time of 90 min. is similar to the measured half-life of the TFI<sub>IIA</sub>-DNA complex<sup>12</sup> and supports that **1** forms a more stable complex with DNA than does TFI<sub>IIA</sub>.

The effect of the polyamides on 5S gene transcription *in vivo* was monitored. *Xenopus* kidney-derived fibroblasts were grown in the presence of increasing concentrations of polyamide





**Figure 7.4.** Inhibition of 5S RNA gene transcription complex formation *in vivo*. Nuclei were prepared from *Xenopus* kidney-derived fibroblasts grown in culture in the presence or absence of polyamides. Polyamides were included in culture medium (in 2.5 mL of media per 25 cm<sup>2</sup> flask) for 24 hours prior to harvesting cells and isolation of nuclei. Equal amounts of nuclei (containing 5 μg of DNA) were incubated for 2 hours in 20 μL reactions containing *Xenopus* RNA polymerase III and labeled and unlabeled nucleoside triphosphates. RNA was isolated from these reactions and analyzed on a denaturing polyacrylamide gel. Graphic representation of the effects of polyamides 1, 2 and 3 on 5S RNA and tRNA transcription expressed as the ratio of 5S RNA to tRNA gene transcription relative to the ratio obtained in the absence of polyamide. Error bars represent estimated standard deviation.

1 in the culture medium for various times. We found that concentrations of polyamide up to 1 μM were not toxic, as measured by cell density, if growth was limited to less than 72 hours.

Nuclei were prepared from cells by hypotonic lysis and equivalent amounts of the isolated nuclei from control, and treated cells were used as templates for transcription with exogenous RNA polymerase III and labeled and unlabeled nucleoside triphosphates. This experiment monitors the occupancy of class III genes with active transcription complexes. 5S RNA transcription can easily be assessed since the repetitive 5S genes give rise to a prominent band on a denaturing polyacrylamide gel. Figure 7.4 shows a graphical representation of the effect of the polyamides on 5S gene transcription relative to transcription of the tRNA genes.

Concentrations of polyamide **1** as low as 100 nM have a pronounced and selective effect on 5S transcription. At higher polyamide concentration, a general decrease in the transcriptional activity of the nuclei is observed; however, at each concentration tested, the effects of the polyamide are much greater on 5S RNA transcription than on tRNA transcription. Having established that nearly maximal inhibition of 5S transcription is achieved with 1  $\mu$ M polyamide **1**, we monitored nuclear transcription after various times of cell growth in the presence of the polyamide. No inhibition is observed for zero time incubation with polyamide **1** at 1  $\mu$ M concentration, indicating that disruption of transcription complexes does not occur during or after the isolation or work-up of cell nuclei. Statistically equivalent levels of 5S transcription inhibition were observed when the cells were exposed to polyamide **1** for 24, 48, or 72 hours.

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

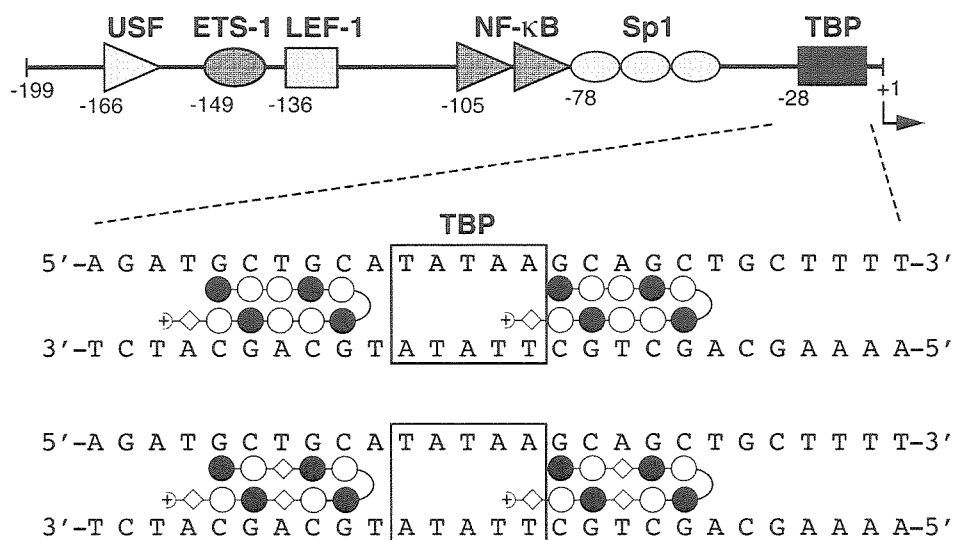
Selective inhibition of RNA polymerase III transcription of a 5S RNA gene relative to a tRNA gene is a first step towards exploring the potential of polyamides as artificial regulators of eukaryotic gene expression. It remains to be determined if a broad panel of genes can be selectively targeted within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

**Targeting the HIV-1 TATA box.** Most protein-encoding genes in mammalian cells are transcribed by RNA Polymerase II.<sup>14</sup> Furthermore, the promoters of most tissue-specific genes and viral protein-encoding genes contain binding sites (TATA boxes) for the TATA-binding protein (TBP). Binding of the TBP subunit of TFIID to the TATA box is known to nucleate assembly of the RNA Polymerase II transcription machinery in TATA box-containing promoters.<sup>14</sup> The x-ray crystal structure of TBP bound to DNA shows that TBP binds in the minor groove, inducing a severe bend and unwinding of the DNA.<sup>15</sup> This information indicates that minor-groove-binding polyamides which block TBP/TFIID-DNA interactions in a gene-specific manner could provide a general approach to regulation of gene expression by small molecules.

To test whether polyamides could specifically regulate RNA Polymerase II transcription by inhibiting with TBP/TFIID-DNA interactions, we set out to design a polyamide to bind with subnanomolar affinity to a target sequence immediately adjacent to the HIV-1 TATA box. Inspection of the sequence of a typical HIV-1 promoter revealed 7 base pair 5'-(A,T)GC(A,T)GC(A,T)-3' sequences immediately flanking the TATA box on both sides (Figure 7.5).<sup>16</sup> We were thus challenged to design a polyamide which binds 7 base pair 5'-(A,T)GC(A,T)GC(A,T)-3' sequences with subnanomolar affinity.

The polyamide pairing rules function within the context of a binding motif, a “molecular template” which correctly positions imidazole and pyrrole rings allowing specific polyamide-DNA contacts. The first binding motif for DNA recognition by polyamides consisted of three-ring polyamides which bind 5 bp sites as antiparallel dimers.<sup>4</sup> Subsequent efforts led to the development of motifs which have increased the binding affinity and specificity, binding site size, and sequence repertoire of polyamides.<sup>8,9</sup>

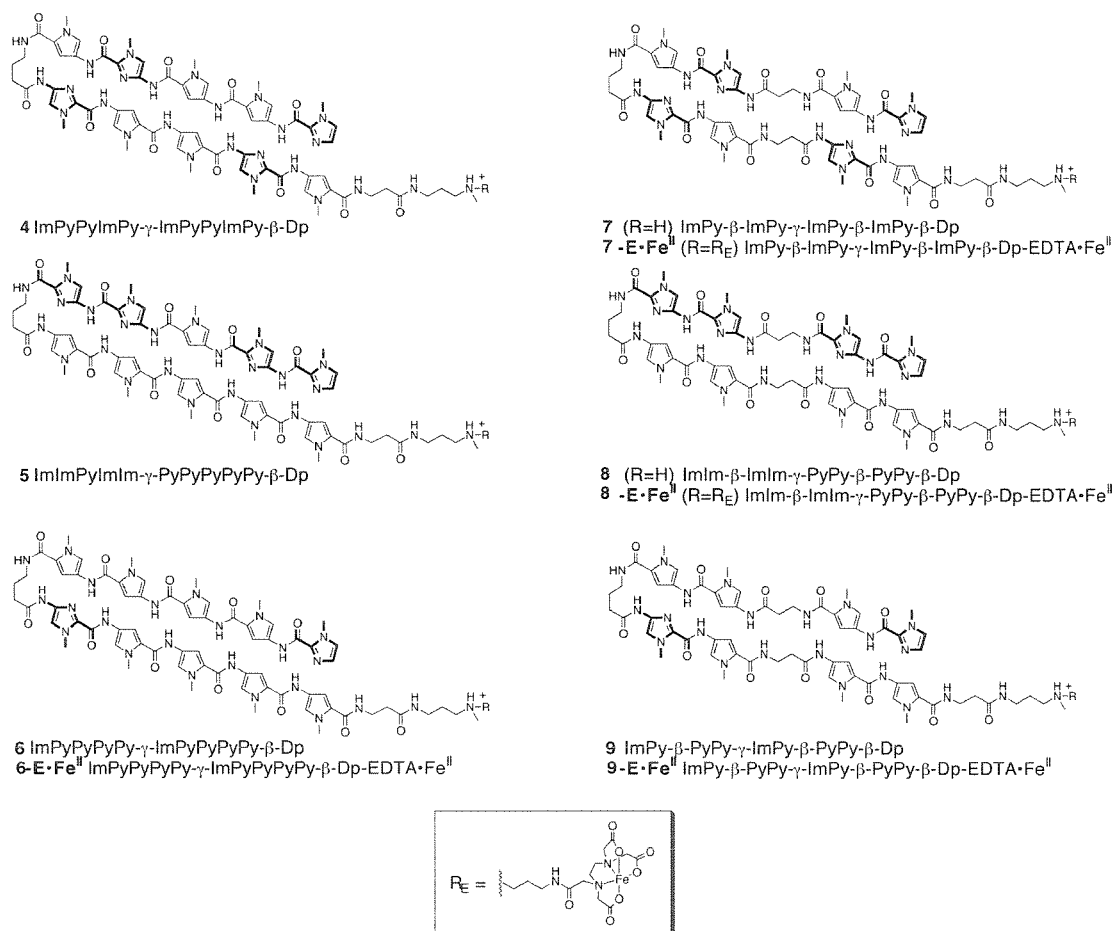
Recently, it was shown that the ten-ring, “5- $\gamma$ -5 motif” polyamide ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp (**3**) specifically binds a 7 base pair 5'-TGTTACA-3' target sequence with subnanomolar affinity in a “hairpin” conformation (Figure 2).<sup>7</sup> Applying the polyamide pairing



**Figure 7.5.** (Top) Schematic representation of the HIV-1 promoter/enhancer. (Bottom) The nucleotide sequence of the TATA box region from a typical HIV-1 promoter is shown, along with binding models for polyamides.

rules to the 5-γ-5 molecular template suggested that the ten-ring hairpin polyamide ImPyPyImPy-γ-ImPyPyImPy-β-Dp would bind the HIV-1 target sequences 5'-(A,T)GC(A,T)GC(A,T)-3' (Figure 7.5). However, as reported below, this all ring polyamide specifically binds the target site 5'-TGCTGCA-3', but with low affinity, necessitating development of a second-generation polyamide.

The amino acid β-alanine has been employed effectively as a single base pair-spanning linker in several different polyamide motifs.<sup>17</sup> In many cases in which a polyamide binds with relatively low affinity due to an apparent register mismatch between polyamide residues and their target DNA bases, substitution of one or more pyrrole residues with β-alanine results in marked increases in binding affinity. Furthermore, it was demonstrated that paired β-residues could be incorporated in a hairpin polyamide.<sup>18</sup> These results suggested that the polyamide ImPy-β-ImPy-γ-ImPy-β-ImPy-β-Dp would specifically bind the 5'-(A,T)GC(A,T)GC(A,T)-3' HIV-1 target sequences with high affinity using a 2-β-2-γ-2-β-2 hairpin motif.



**Figure 7.6.** Structures of polyamides 4-9.

We report here equilibrium association constants determined by quantitative DNase I footprint titration experiments<sup>19</sup> of the formally matched (according to the polyamide pairing rules) polyamides **4** and **7** and the formally mismatched polyamides for the HIV-1 target sequence 5'-TGCTGCA-3'.

In addition, to further explore the generality of the 2- $\beta$ -2- $\gamma$ -2- $\beta$ -2 motif, we investigate binding of each polyamide to their respective formal match sequences 5'-TGGTGGA-3' and 5'-TGTTACA-3'. We also report the results of MPE- $\text{Fe}^{\text{II}}$  footprinting<sup>20</sup> and affinity cleavage studies for a subset of these compounds. Finally, we compare the equilibrium association constants of polyamide **7** for the HIV-1 target sequence 5'-TGCTGCA-3' determined at either

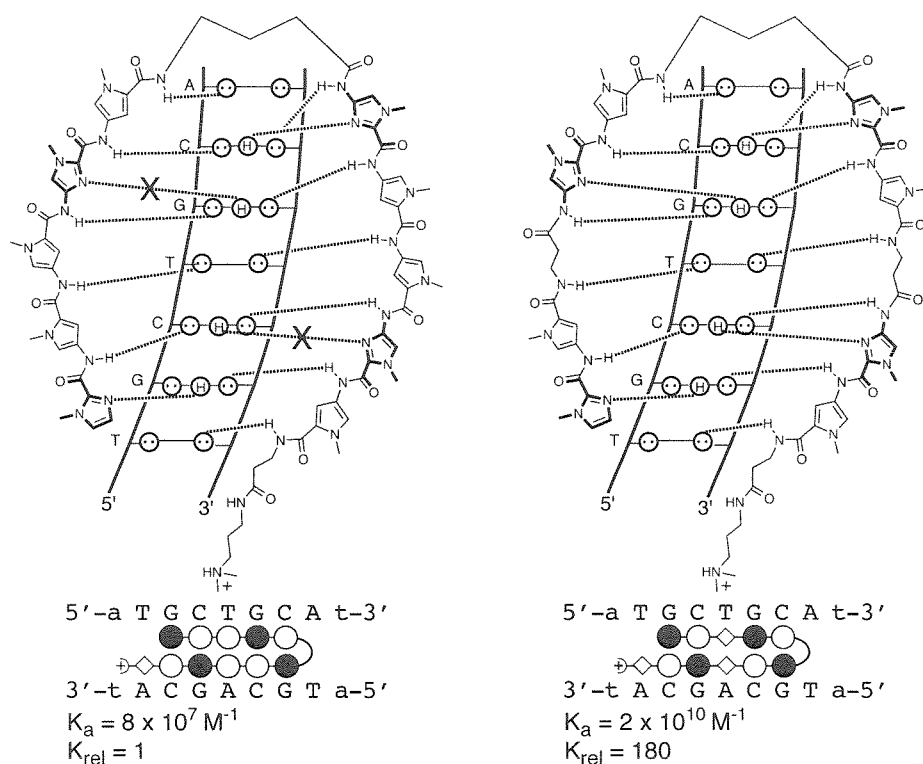
**Table 7.1.** Equilibrium association constants (M<sup>-1</sup>)<sup>a</sup>.

Polyamide	5'-aTGCTGCA <sub>T</sub> -3'	5'-aTGGTGGA <sub>T</sub> -3'	5'-aTGTTACA <sub>T</sub> -3
<i>5-γ-5 motif:</i>			
<b>4</b> ImPyPyImPy-γ-ImPyPyImPy-β-Dp	<b>8.3 x 10<sup>7</sup> (1.5)</b>	< 1 x 10 <sup>7</sup>	< 1 x 10 <sup>7</sup>
<b>5</b> ImImPyImIm-γ-PyPyPyPyPy-β-Dp	< 5 x 10 <sup>7</sup>	<b>&lt; 5 x 10<sup>7</sup></b>	< 5 x 10 <sup>7</sup>
<b>6</b> ImPyPyPyPy-γ-ImPyPyPyPy-β-Dp	1.2 x 10 <sup>9</sup> (0.4)	< 5 x 10 <sup>8</sup>	<b>5.1 x 10<sup>10</sup> (1.1)</b>
<i>2-β-2-γ-2-β-2 motif:</i>			
<b>7</b> ImPy-β-ImPy-γ-ImPy-β-ImPy-β-Dp <sup>b</sup>	<b>1.5 x 10<sup>10</sup> (0.3)</b>	< 5 x 10 <sup>8</sup>	< 5 x 10 <sup>8</sup>
<b>8</b> ImIm-β-ImIm-γ-PyPy-β-PyPy-β-Dp	1.7 x 10 <sup>8</sup> (0.3)	<b>7.6 x 10<sup>9</sup> (1.0)</b>	< 1 x 10 <sup>8</sup>
<b>9</b> ImPy-β-PyPy-γ-ImPy-β-PyPy-β-Dp	< 1 x 10 <sup>8</sup>	< 1 x 10 <sup>8</sup>	<b>2.6 x 10<sup>9</sup> (1.0)</b>
<i>Unlinked dimer motif:</i>			
ImPyPyPyPy-β-Dp	1.1 x 10 <sup>7</sup> (0.2)	< 1 x 10 <sup>7</sup>	<b>3.9 x 10<sup>8</sup> (0.6)</b>
ImPy-β-ImPy-β-Dp	<b>3.0 x 10<sup>6</sup> (0.2)</b>	< 2 x 10 <sup>5</sup>	1.5 x 10 <sup>6</sup> (0.3)
ImPy-β-PyPy-β-Dp	< 5 x 10 <sup>5</sup>	< 5 x 10 <sup>5</sup>	<b>2.8 x 10<sup>6</sup> (0.1)</b>

<sup>a</sup>Values reported are the mean values from at least three DNase I footprint titration experiments. Nucleotides flanking polyamide binding sites are in lowercase type. Association constants corresponding to formally matched complexes are in bold type. <sup>b</sup>This polyamide binds the sites 5'-TCGTCGA-3' and 5'-tAGCTGTTt-3' with equilibrium association constants of 2.1 x 10<sup>9</sup> M<sup>-1</sup> (0.1) and 1.7 x 10<sup>9</sup> M<sup>-1</sup> (0.1), respectively.

24°C or 37°C, and using either standard polyamide assay solution conditions or model intracellular solution conditions.

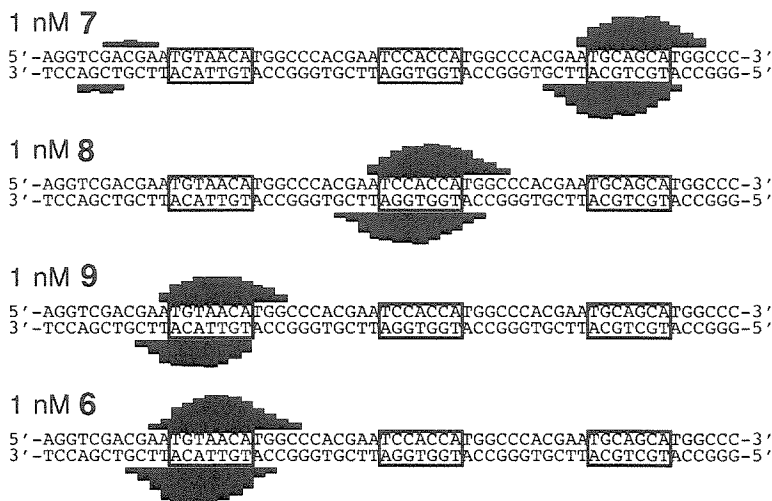
Quantitative DNase I footprint titration experiments (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>, pH 7.0, 24°C) were carried out on the 282 bp, 3'-<sup>32</sup>P end-labeled pJT2B2 *Eco*RI/*Pvu*II restriction fragment which contains the three target sites 5'-TGCTGCA-3' (the HIV-1 target sequence), 5'-TGGTGGA-3', and 5'-TGTTACA-3'. These experiments reveal that the 5-γ-5 motif polyamide **4** binds the HIV target sequence 5'-TGCTGCA-3' with a relatively low equilibrium association constant, K<sub>a</sub> = 8 x 10<sup>7</sup> M<sup>-1</sup>, while no binding of the 5-γ-5 polyamide **5** to its target site 5'-TGGTGGA-3' is observed (K<sub>a</sub> < 5 x 10<sup>7</sup> M<sup>-1</sup>). In contrast, 5-γ-5 polyamide **6** binds its target site 5'-TGTTACA-3' with very high affinity (K<sub>a</sub> = 5 x 10<sup>10</sup> M<sup>-1</sup>) (Table 7.1).



**Figure 7.7.** Models and experimentally determined equilibrium association constants for the complexes of polyamides **1** (left) and **2** (right) with the sequence 5'-TGCTGCA-3'.

The three 2- $\beta$ -2- $\gamma$ -2- $\beta$ -2 motif polyamides **7**, **8** and **9** selectively target their respective match sequences 5'-TGCTGCA-3' (the HIV-1 target sequence), 5'-TGGTGGA-3', and 5'-TGTTACA-3' with equilibrium association constants greater than  $10^9 \text{ M}^{-1}$ . These results demonstrate that the 2- $\beta$ -2- $\gamma$ -2- $\beta$ -2 motif allows specific targeting of a range of 7 base pair sequences with subnanomolar affinity, and is potentially general for all 5'-(A,T)NN(A,T)NN(A,T)-3' sequences. Polyamide **7** binds the HIV-1 target sequence 5'-TGCTGCA-3' with subnanomolar affinity ( $K_a = 2 \times 10^{10} \text{ M}^{-1}$ ) and is 10-fold specific for this site relative to the single base pair mismatch site 5'-AGCTGTT-3'. Thus, **7** represents a solution to the polyamide design problem we set out to solve here.

The failure of polyamide **1** to bind the HIV target sequence with high affinity is consistent with previous results. For example, the 4- $\gamma$ -4 motif polyamides ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp,

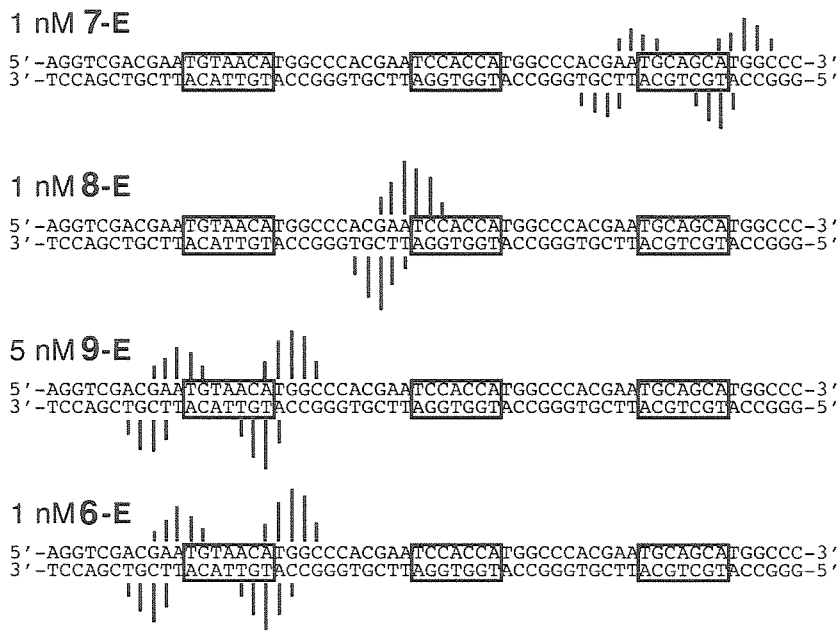


**Figure 7.8.** (a) Results of MPE•Fe<sup>II</sup> footprinting experiments: binding of respective 7 bp target sequences by polyamides 7, 8, 9, and 6. Bar heights are proportional to the extent of cleavage protection at the indicated base.

ImPyPyPy-γ-PyPyPyPy-β-Dp and ImImPyPy-γ-ImImPyPy-β-Dp bind respective 6 base pair target sequences with subnanomolar affinity ( $K_a \geq 10^{10} \text{ M}^{-1}$ ), while ImPyImPy-γ-ImPyImPy-β-Dp and ImImImIm-γ-PyPyPyPy-β-Dp bind respective target sequences with substantially lower affinity ( $K_a < 10^8 \text{ M}^{-1}$ ).<sup>21</sup> Based on these and additional results, the common feature of hairpin polyamides that bind with high affinity appears to be the placement of imidazole residues only at the first and second positions of polyamide subunits, counting from the N-terminal end. Apparently imidazoles placed beyond the second position of a polyamide subunit are not positioned optimally for specific ligand-DNA contacts.

Polyamides 4 and 5, which have imidazole residues beyond the second position of a polyamide subunit, have equilibrium association constants for their match sites more than 100-fold lower than their β-alanine-substituted analogs 7 and 8, respectively (Table 7.1). Thus, *the 2-β-2-γ-2-β-2 motif is essential for high-affinity recognition in these cases.* In contrast, polyamide 6 having no imidazoles beyond the second position of a polyamide subunit binds its match site with an equilibrium association constant ~20-fold higher than its β-alanine-substituted analog 9.

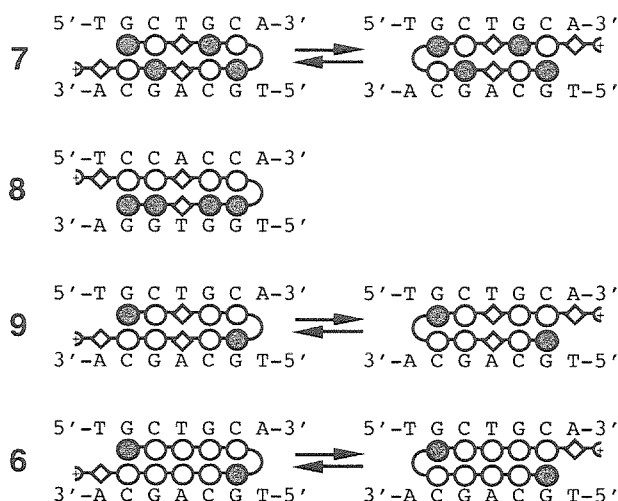




**Figure 7.9.** Results of affinity cleavage experiments. Line heights are proportional to the extent of cleavage at the indicated base.

The following model is consistent with our data: 1) In 2- $\beta$ -2- $\gamma$ -2- $\beta$ -2 polyamides **7** and **8** the flexible  $\beta$ -alanine linkers correctly position the imidazole residues following them for specific hydrogen bond contacts with their target guanine bases, while in 5- $\gamma$ -5 polyamides **4** and **5** the analogous hydrogen bonds cannot form. 2) In contrast, for both 2- $\beta$ -2- $\gamma$ -2- $\beta$ -2 polyamide **9** and 5- $\gamma$ -5 polyamide **6**, the pyrroles beyond the second positions within polyamide subunits are correctly positioned, and **6** binds with lower affinity than **9** due to its greater conformational entropy.

These results indicate that the optimal binding positions, or “binding register,” is different for imidazole and pyrrole residues, consistent with previous results which suggest that, for subunits composed entirely of rings (i.e., without  $\beta$ -alanine linkers), imidazole residues go out of register after 2-3 residues with substantial drops in affinity,<sup>21</sup> while pyrrole residues go out of register after 5 rings with a gradual leveling of affinity before a substantial drop at 8 rings.<sup>22</sup> Finally, these results indicate that the 2- $\beta$ -2- $\gamma$ -2- $\beta$ -2 and 5- $\gamma$ -5 motifs are complementary for recognition

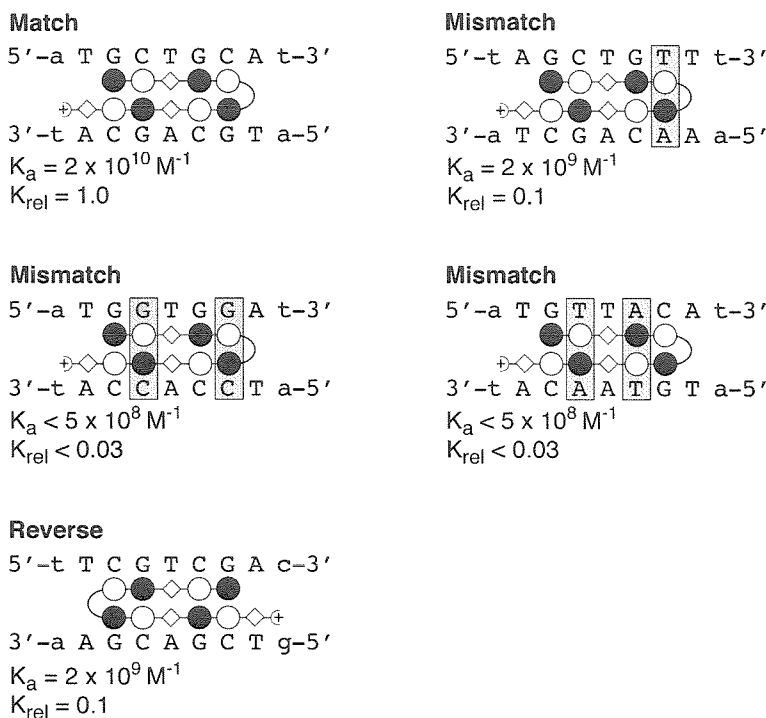


**Figure 7.10.** Polyamide binding models.

of 7 base pair sequences, with the optimal motif dependent on the sequence composition of the desired target site.

Equilibrium association constants for the “unlinked dimer motif” polyamides were measured to allow comparison with their corresponding  $\gamma$ -aminobutyric acid-linked polyamides. The hairpin polyamides, which are composed of 2- $\beta$ -2 subunits, bind with  $\sim 6000$ -fold and  $\sim 1000$ -fold higher affinity, respectively, relative to their respective unlinked analogs. Polyamide 6, which is composed of five-ring subunits, binds with  $\sim 100$ -fold higher affinity than its unlinked analog.

MPE•Fe<sup>II</sup> footprinting experiments, carried out with the polyamides which have subnanomolar binding affinities, confirm that these compounds specifically bind their respective 7 base pair match sites (Figure 7.8). Affinity cleavage experiments were performed with the EDTA-polyamides 7-E, 8-E, 9-E, and 6-E to identify the location of the C-termini of these polyamides when bound to their target sites (Figure 7.9). The symmetric polyamides 7-E, 9-E, and 6-E produce cleavage patterns at both sides of their binding sites as expected, consistent with two distinct binding orientations. It appears that, in general, side-by-side polyamide-DNA complexes preferentially bind in the 5' to 3' (N- to C-terminus) orientation. Recent studies with hairpin polyamides are consistent with formation of 3' to 5'-oriented polyamide-DNA complexes,



**Figure 7.11.** Association constants and binding models for complexes of polyamide 7.

and suggest that such “reversed orientation” complexes typically have affinities ~10-fold lower than analogous 5’ to 3’-oriented complexes.<sup>8g</sup> Consistent with this trend, we find here that polyamide 7 binds the “reversed orientation” match site 5’-TCGTCGA-3’ with an affinity ~10-fold lower than it binds the site 5’-TGCTGCA-3’. However, the asymmetric polyamide 8-E produces a cleavage pattern consistent with the polyamide binding preferentially in the reversed orientation (3’ to 5’, N- to C-terminus) as illustrated (for the parent compound 8) in Figure 7.10.

Quantitative DNase I footprinting experiments indicate that increasing the equilibration temperature from 24°C to 37°C, and changing the solution conditions from standard polyamide assay conditions (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>, pH 7.0 at 24°C) to conditions modeling those encountered within a typical mammalian cell<sup>23</sup> (140 mM KCl, 10 mM NaCl, 1 mM MgCl<sub>2</sub>, 1 mM spermine, pH 7.2) has very little effect on polyamide binding affinity.

**Table 7.2.** Equilibrium association constants ( $M^{-1}$ ) of polyamide 7 for its match site 5'-TGCTGCA-3'<sup>a</sup>.

Buffer <sup>b</sup>	Temp.	5'-aTGCTGCA-3'
A	24 °C	$1.5 \times 10^{10}$ (0.3)
	37 °C	$8.4 \times 10^9$ (2.3)
B	24 °C	$1.9 \times 10^{10}$ (0.6)
	37 °C	$1.1 \times 10^{10}$ (0.2)

<sup>a</sup>Values reported are the mean values from at least three DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses.

<sup>b</sup>Buffer A: 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>, pH 7.0 at 24°C; Buffer B: 10 mM HEPES•HCl, 140 mM KCl, 10 mM NaCl, 1 mM MgCl<sub>2</sub>, 1 mM spermine, pH 7.2.

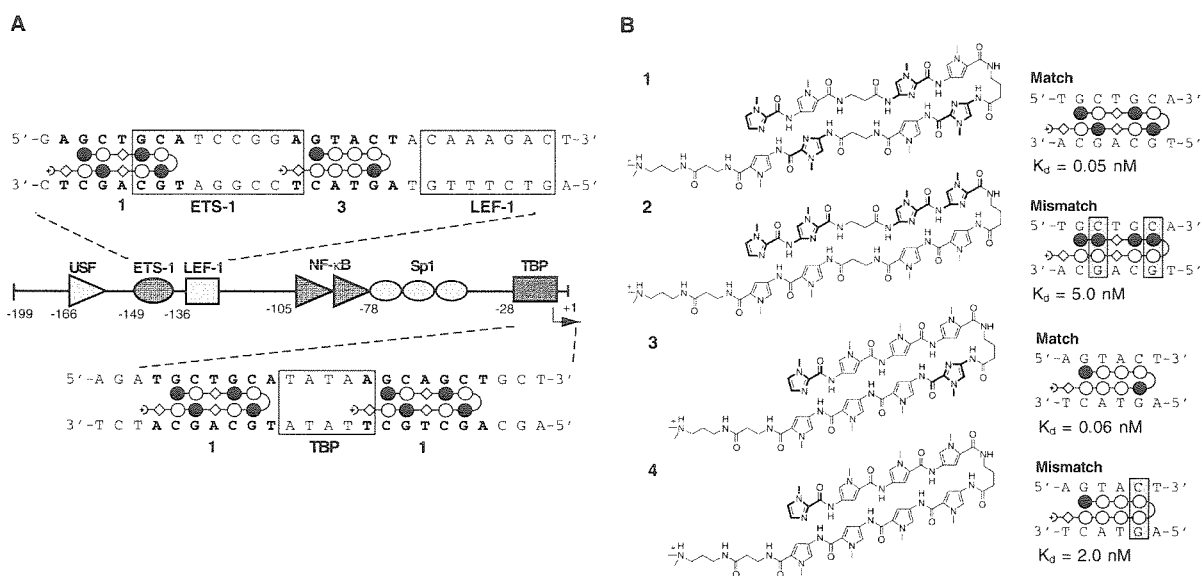
Cell-permeable, sequence-specific DNA-binding ligands have the potential to regulate gene expression *in vivo*. For this approach to be generally useful, a class of ligands is needed which can target *predetermined* DNA sequences with high affinity and specificity. We report here the development of a small molecule which binds with subnanomolar affinity to a conserved 7 base pair sequence within the HIV-1 promoter using a novel polyamide motif, and demonstrate the unique ability of this motif to target certain (G,C)-rich sequences including the HIV-1 target sequence. These results demonstrate the power of state-of-the-art design rules to guide the development of polyamides which bind predetermined sequences with high affinity and specificity. Finally, these results provide a basis for investigation of the ability of polyamides targeted adjacent to a TATA box to inhibit RNA Polymerase II-mediated transcription, and to block viral replication.

**Inhibition of RNA Polymerase II Transcription and HIV-1 Viral Replication in Human Cells by Synthetic DNA-Binding Ligands.** Py-Im polyamides represent the only class of synthetic small molecules that can bind predetermined DNA sequences with affinities and specificities comparable to DNA-binding proteins.<sup>9</sup> The DNA-binding activity of the 5S RNA gene-specific transcription factor TFIID was inhibited by an eight-ring hairpin polyamide that bound within the recognition site of zinc finger four in the DNA minor groove. As a result, transcription of 5S RNA genes by RNA polymerase III was suppressed *in vitro* and in cultured *Xenopus* cells.<sup>24</sup> The question arises whether polyamides can permeate human cells and specifically regulate genes transcribed by pol II. As a first case study, we examined the ability of polyamides to inhibit HIV-1 transcription in cell-free assays and viral replication in human lymphocytes.

The HIV-1 enhancer/ promoter element contains binding sites for the cell-encoded proteins USF, Ets-1, LEF-1, the nuclear factors NF- $\kappa$ B, Sp1 and TBP.<sup>16</sup> In order to shut-down the promoter, polyamides were designed to target the transcription factors TBP, LEF-1 and Ets-1 simultaneously. TBP is indispensable for initiation of HIV-1 transcription, and LEF-1, considered to be an architectural protein, plays a central role in coordinating activities of multiple transcription factors. Both TBP and LEF-1 bind the minor groove of DNA and are likely to be inhibited by the minor groove binding polyamides. Ets-1 predominantly contacts the major groove in the center of its binding site, with additional flanking contacts that are possibly in the minor groove. The DNA-recognition sites of these transcription factors are not optimal polyamide target sequences because they are found in the promoters of many cellular protein-coding genes. However, the sequences immediately flanking these transcription factor binding sites are often conserved for a particular gene, providing an address for gene-specific targeting. This study demonstrates how gene-specific polyamide inhibition can be achieved by targeting sequences that are located adjacent to, but do not coincide with, binding sites for transcriptional regulators, and

how strategic targeting of multiple transcription factors can result in effective inhibition of a pol II promoter and viral replication in human cells.

The sequence 5'-(A,T)GC(A,T)GC(A,T)-3' is present on both sides of the TATA-box and immediately upstream of the Ets-1 binding site. The sequence 5'-AGTACT-3' is found between the recognition sites for Ets-1 and LEF-1. These sequences are conserved for most reported strains of HIV-1.<sup>25</sup> Although the propensity for mutation at these sites is unknown, allowed sequence changes in the promoter could be targeted with new polyamides designed by the pairing rules.



**Figure 7.12.** Polyamide and transcription factor binding sites. (A) Schematic of the HIV-1 promoter (nucleotide positions -199 to +1) showing binding sites for polyamides **1** and **3** and the transcription factors USF, Ets-1, LEF-1, NF-κB, Sp1 and TFIID (TBP). For polyamide binding models, shaded and unshaded circles represent Im and Py rings, respectively. (B) Structures of polyamides ImPy-β-ImPy-γ-ImPy-β-ImPy-β-Dp (**HIV-1**), ImIm-β-ImIm-γ-PyPy-β-PyPy-β-Dp (**HIV-2**), ImPyPyPy-γ-ImPyPyPy-β-Dp (**HIV-3**), and ImPyPyPy-γ-PyPyPyPy-β-Dp (**HIV-4**). Binding models and measured dissociation constants are shown. Mismatches are highlighted.

According to these rules, 5'-(A,T)GC(A,T)GC(A,T)-3' sequences may be targeted by hairpin polyamide **1** having sequence composition ImPy- $\beta$ -ImPy- $\gamma$ -ImPy- $\beta$ -ImPy- $\beta$ -Dp. Since the  $\beta/\beta$  pairing recognizes both A•T and T•A base pairs,<sup>26</sup> polyamide **HIV-1** is expected to bind to all three sites. Quantitative footprint titration experiments revealed that polyamide **HIV-1** binds to each of these sites with a  $K_d$  of 0.05 nM. A mismatch control polyamide ImIm- $\beta$ -ImIm- $\gamma$ -PyPy- $\beta$ -PyPy- $\beta$ -Dp (**HIV-2**), which differs only in the placement of the Im and Py amino acids, binds the 5'-(A,T)GC(A,T)GC(A,T)-3' sites with 100-fold reduced affinity relative to polyamide **HIV-1**. According to the pairing rules, the sequence 5'-AGTACT-3', between the LEF-1 and Ets-1 binding sites, will be bound by polyamide **HIV-3** of sequence composition ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp. Quantitative footprint titration experiments reveal that polyamide **HIV-3** binds this site with  $K_d = 0.06$  nM. Mismatch polyamide **HIV-4** binds this sequence with >100-fold reduced affinity.

Binding of the TBP subunit of TFIID in the minor groove nucleates assembly of the pol II transcription machinery for TATA-containing genes.<sup>27</sup> TBP binds the HIV-1 TATA element with a  $K_d$  of ~1-3 nM. A gel mobility shift assay revealed that polyamide **HIV-1** inhibits TBP binding to a double stranded oligonucleotide corresponding to the HIV-1 TATA box region, while no inhibition is observed for control polyamide **HIV-2**. Additionally, polyamide **HIV-1** does not inhibit TBP binding to the TATA box region of the adenovirus major late (AdML) promoter (5'-GGGGGCTATAAAAGGGGGT-3'), which contains mismatch flanking sequences. The half-life of the polyamide **HIV-1**-DNA complex was determined by competition experiments to be in excess of 2.5 hours.

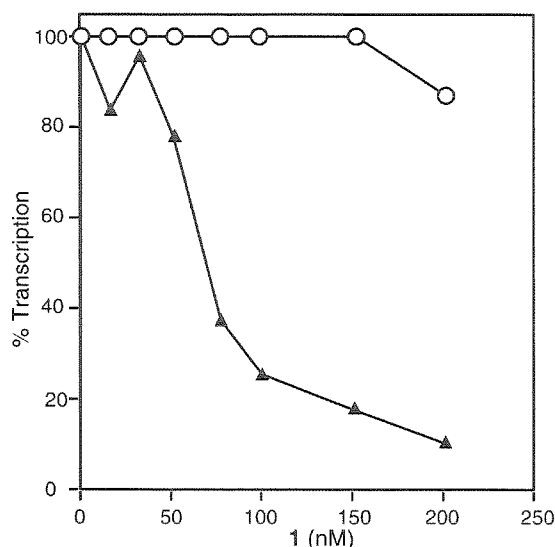
The Ets-1 recognition site in the HIV-1 enhancer is flanked by binding sites for polyamides **HIV-1** and **HIV-3**. The isolated Ets-1 DNA-binding domain,  $\Delta$ N331 (amino acids 331 to 416), bound to the HIV-1 enhancer with a  $K_d$  of approximately 0.5 nM.<sup>28</sup> When polyamides were preincubated with the labeled, double-stranded HIV-1 oligonucleotide before

adding  $\Delta N331$ , polyamide **HIV-3** had no effect on Ets-1 DNA-binding. Polyamide **HIV-1**, however, prevented the Ets-1/DNA complex formation. The two mismatch polyamides **HIV-2** and **HIV-4** did not prevent complex formation. DNase I footprinting revealed that polyamide **HIV-1** prevents Ets-1 DNA-binding while polyamide **HIV-3** co-occupies the DNA with Ets-1. Polyamide **HIV-1** likely inhibits Ets-1 binding by steric interaction with protein contacts in the minor groove, while polyamide **HIV-3**, which does not inhibit Ets-1 binding, must occupy the minor groove beyond the Ets-1 contacts.

LEF-1 is a member of the HMG family of minor-groove binding proteins<sup>29</sup> and has been shown to be essential for HIV-1 transcription and replication in lymphoid cells. Three sites on the HIV-1 promoter/enhancer restriction fragment are bound by the 86-amino acid LEF-1 DNA binding domain, sites L1, L2, and L3 with  $K_d = 1.4$ , 5.8, and 4.9 nM, respectively. The LEF-1 footprint at site L1, characterized by a marked DNase I hypersensitive site (HSS), changes to the polyamide footprint in the presence of match polyamide **HIV-3**. LEF-1 binding is inhibited 50% at a polyamide **HIV-3** concentration of approximately 60 pM. Thus, polyamide **HIV-3**, located in the minor groove immediately adjacent to site L1, inhibits LEF-1 binding to this site. Polyamide **HIV-3** inhibits LEF-1 binding to the sites L2 and L3 at markedly higher polyamide concentrations, since there are no high affinity polyamide recognition sites adjacent to sites L2 and L3. Inhibition was observed either by adding the polyamide to the DNA before LEF-1 or after preincubation of the DNA with LEF-1, consistent with the relative dissociation constants for the two binding reactions. Mismatch polyamide **HIV-4** failed to inhibit LEF-1 binding.

The effects of polyamides **HIV-1** and **HIV-2** on HIV-1 transcription were tested in an in vitro transcription assay with a HeLa cell nuclear extract. Polyamide **HIV-1** inhibited basal transcription from the HIV-1 promoter but not from the CMV major intermediate early promoter (MIEP), which contains a mismatched TATA-flanking sequence (5'-GAGGTCTATATAAGCAGA-3'). The mismatch polyamide **HIV-2** did not inhibit transcription from either promoter. We performed titrations of polyamide **HIV-1** over a wide range of

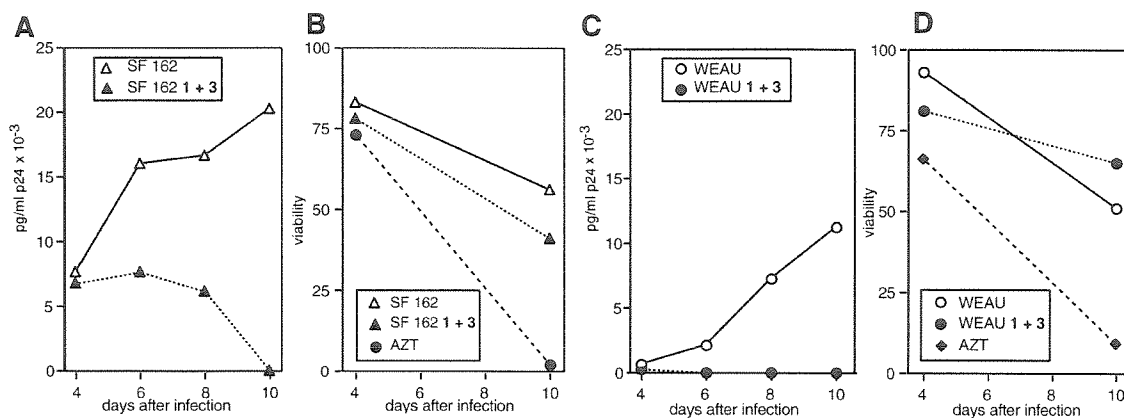




**Figure 4.13.** Polyamide inhibition of HIV-1 transcription in vitro. Relative levels of HIV-1 transcription (closed triangles) compared to CMV control transcription (open circles) are plotted as a function of polyamide **HIV-1** concentration, from mixed template reactions with HeLa nuclear extract.

concentrations (1 to 200 nM) with both the HIV-1 and CMV templates in the same reaction. Under these conditions, we observed 50% inhibition of HIV-1 transcription at 60 nM polyamide **HIV-1**, which corresponds to a 6- to 10-fold excess of polyamide over binding sites. No significant inhibition of CMV transcription was observed.

In order to test the effects of polyamides **HIV-3** and **HIV-4** on LEF-1 activated HIV-1 transcription, we supplemented the HeLa nuclear extract with a human lymphoid H9 cell extract, which contains high levels of LEF-1 protein (as determined by western blots). The mixture of H9 and HeLa cell extracts stimulated HIV-1 transcription 2.5-3-fold over the level of transcription observed with the HeLa extract alone. Immunodepletion of LEF-1 protein from the H9 extract abolished this activated transcription. Polyamide **HIV-3** inhibited HIV-1 transcription in this system with a 50% reduction of transcription observed at 10-30 nM polyamide. Polyamide **HIV-3** failed to inhibit HIV-1 transcription in the LEF-1-depleted extract. We also monitored the activity of the CMV MIEP in both the mock-depleted and LEF-1-depleted H9 cell extract, with

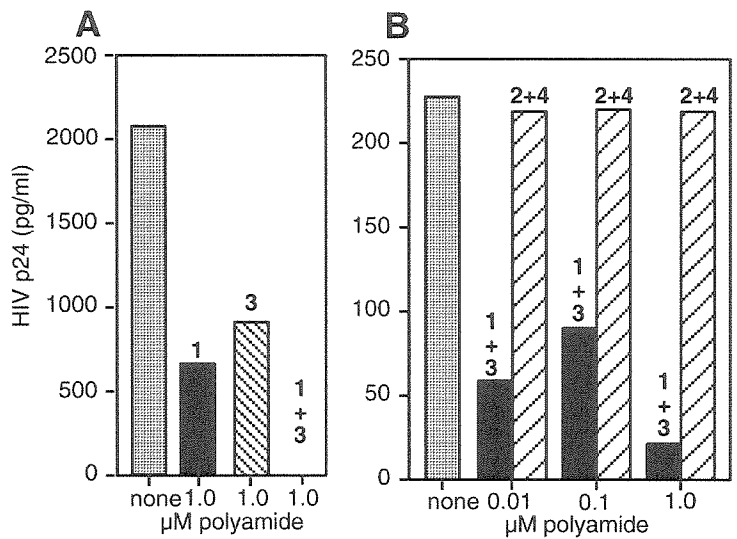


**Figure 4.14.** Kinetics of polyamide inhibition of HIV-1 replication and effects on cell viability. PBMC were infected with the macrophage-tropic HIV-1 isolate SF162 (panels A and B) or the T-cell tropic isolate WEAU 1.6 (panels C and D). Polyamides **1** and **3** were added in combination at 1  $\mu$ M each. Virus replication was monitored by p24 ELISA (panels A and C). Cell viability was determined by trypan blue exclusion in the presence of polyamides or AZT (panels B and D).

the result that LEF-1 depletion had no effect on CMV transcription. As additional controls, we tested the effect of the mismatch polyamide **HIV-4** on HIV-1 transcription and polyamides **HIV-3** and **HIV-4** on CMV transcription. No potential binding sites for either polyamide **HIV-3** or **HIV-4** are present in the CMV MIEP sequence. As expected, polyamide **3** fails to inhibit CMV transcription and polyamide **HIV-4** fails to inhibit either HIV-1 or CMV transcription.

Notably, polyamide **HIV-3** does not inhibit basal transcription although binding sites for this polyamide are present at the start-site for transcription and within the HIV-1 RNA coding sequence. These observations suggest that pol II can transcribe DNA with a polyamide bound in the minor groove and that polyamides are only inhibitory to transcription when these compounds interfere with the DNA binding activity of a required transcription factor.

Since there are multiple spliced and unspliced species of HIV-1 RNA with different turnover kinetics, we examined HIV-1 transcription indirectly by measuring HIV-1 replication in human PBMC in culture. PBMC were infected with the T cell-tropic HIV-1 strain WEAU1.6,<sup>30,31</sup> or the macrophage-tropic strain SF162.<sup>32,33</sup> Polyamides were added to culture medium and levels



**Figure 4.15.** Polyamide inhibition of HIV-1 replication. (A) and (B) depict independent experiments in which polyamides alone or in combination were added to cultures of PBMCs, obtained from two separate donors, infected with the primary HIV-1 isolate WEAU 1.6 (kindly provided by G. Shaw). Replication was measured 6 or 8 days after infection by release of p24 capsid antigen into the medium. Assays were performed in duplicate and showed less than 5% variation from the mean.

of HIV-1 p24 viral capsid protein (primarily as virions) were determined on subsequent days post infection using a standard ELISA assay. In control PBMC cultures with no added polyamide, viral replication of both strains resulted in increasing p24 levels between 4 and 10 days of culture. Polyamide **HIV-1** at 1  $\mu$ M concentration caused an 80% reduction in virus, while polyamide **HIV-3** at 1  $\mu$ M concentration caused a 60% reduction after 6 to 8 days.

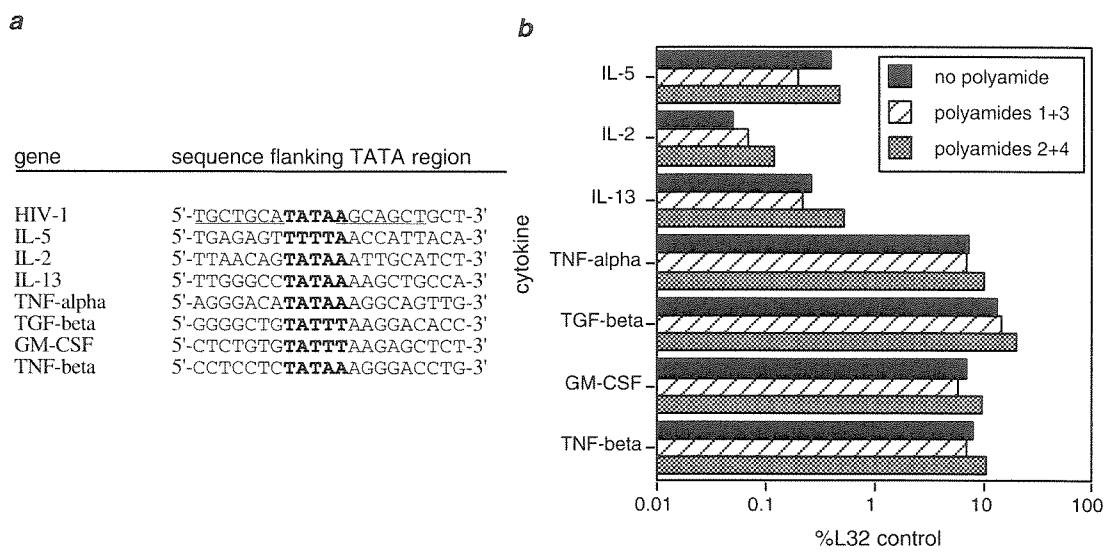
For virus replication experiments, human peripheral blood mononuclear cells (PBMC) were separated from whole blood collected from normal adult volunteers by density gradient centrifugation as described.<sup>34,35</sup> Donors were provided by the General Clinical Research Center of The Scripps Research Institute, which is supported by NIH grant MO1 RR00833. PBMC were activated with 2  $\mu$ g/ml phytohemagglutinin and 20 units/ml of interleukin-2 for 2-3 days prior to HIV-1 infection. Each culture of  $5 \times 10^5$  PBMC was infected with  $10^3$  tissue culture infectious

doses of HIV-1 for 24 hours; free virus was removed by washing the cells in medium, and polyamides added to the culture. Virus replication in culture was measured by HIV-1 p24 viral capsid antigen ELISA (Dupont Medical Products, Boston, MA). Assays of HIV-1 replication were performed five times with five human PBMC donors.

The combination of polyamides **HIV-1** + **HIV-3** inhibited HIV-1 replication at 10 nM to 1  $\mu$ M concentration, but the closely related polyamides **HIV-2** + **HIV-4** did not. The combination of polyamides **HIV-1** and **HIV-3** at 1  $\mu$ M each acted in synergy to reduce viral p24 levels to below the threshold of detection after 6-8 days for WEAU (<10 pg/ml; greater than 99.9% inhibition of viral replication), and were as effective as 1  $\mu$ M azidothymidine (AZT) in blocking HIV-1 replication. The macrophage-tropic SF162 isolate, which replicates in both macrophages and CD4<sup>+</sup> T-lymphocytes, was not efficiently inhibited by either polyamide alone, but the combination of 1  $\mu$ M each polyamide **HIV-1** and **HIV-3** reduced and eventually blocked its replication after 10 days. Addition of mismatch polyamides **HIV-2** and **HIV-4**, which differ by a single atomic substitution (**HIV-4**), or a rearrangement of the Im and Py amino acids (**HIV-2**), from the match polyamides, had no effect on the level of virus in the medium, either alone, or in combination.

The inhibitory effects of polyamides was not due to obvious toxicity. No significant decrease in cell viability was apparent in PBMC cultures treated with polyamides **HIV-1** and **HIV-3** for 10 days, in contrast to 90% mortality observed for PBMC cells treated with 1  $\mu$ M AZT for the same. Cell viability was slightly higher in WEAU-infected cultures that were treated with polyamides **HIV-1** and **HIV-3** than in untreated cultures (panel D, filled circles versus open circles), probably because the cytopathic effect of HIV-1 infection was completely reversed. Cell recovery was not impacted by polyamide treatment, but was reduced by AZT treatment.

The observed polyamide inhibition of virus replication is likely due to interference with the DNA-binding activities of TBP and Ets-1 by polyamide **HIV-1** and the binding activity of



**Figure 4.15.** Control gene expression. (A) The TATA-box (bold) and flanking sequences (from GenBank listings) of each of the cytokine/growth factor genes examined are shown. The polyamide **1** binding site is underlined. (B) Ribonuclease protection assays for the indicated mRNAs. PBMCs were cultured and either left untreated, or treated with 10  $\mu$ M of polyamides **HIV-1 + HIV-3** or **HIV-2 + HIV-4** for six days. Data are expressed as the intensity of each RNA relative to the intensity of the ribosomal L32 RNA band to standardize for RNA loading.

LEF-1 by polyamide **HIV-3**, but it is possible that inhibition of cellular genes involved in T-cell activation could have an indirect effect on HIV-1 replication.

To assess this possibility, we performed an RNAase protection assay for transcripts of a number of cytokine and growth factor genes, including IL-2, IL-5, and IL-13 which differ in the target sequences flanking the TATA box. Four other cytokine genes that lack binding sites for either polyamide **HIV-1** or **HIV-3** in their promoters were also examined. The results show that exposure of activated human PBMC to a combination of either polyamides **HIV-1 + HIV-3** or **HIV-2 + HIV-4** (1  $\mu$ M each) for 6 days failed to inhibit RNA expression of all the genes examined. There was no difference in the intensity of CD4 and CD8 RNA bands (not shown), indicating equivalent recovery of CD4 and CD8 T cells in treated and untreated cells. This lack of inhibition of cytokine gene transcription suggests that the polyamides reduce virus replication in cells by a direct effect on HIV-1 RNA transcription.

Our present studies show that polyamides designed to target DNA sequences 6-7 bp in length are effective inhibitors of gene transcription in cell-free systems and viral replication in human cells. Because sequences of these lengths would be highly redundant in the human genome, it had seemed likely that these ligands would have deleterious effects on cell metabolism due to interference with the activity of cellular genes. However, the results described here indicate that a set of two polyamides which recognize 6-7 bp sequences can be sufficient for gene-specific regulation *in vivo*. It is interesting to compare these small molecule transcription repressors to eukaryotic transcriptional regulatory proteins that also recognize multiple short sequences in order to increase functional specificity. The observations that polyamides do not interfere with pol II elongation, and that polyamides can bind simultaneously with certain major groove proteins, should further enhance gene-specificity.<sup>36</sup> In addition, polyamides are not limited to 6-7 base-pair recognition. For example, polyamides of similar size to those described here have been shown to bind as cooperative dimers to sites 10-16 base pairs in length.<sup>37,38</sup> The polyamide binding site size required to elicit optimal biological function will be reported in due course.

The specific inhibition of genes transcribed by pol II represents an important first step toward asking whether cell-permeable small molecule transcription antagonists might regulate gene expression in complex organisms. We have chosen TBP and two additional key regulators of HIV-1 transcription, Ets-1 and LEF-1, as targets for inhibition of pol II-driven transcription. Since most tissue-specific cellular genes and viral genes contain TATA elements as well as enhancer factor binding sites, this approach may be generally applicable for the inhibition of most target genes.

All gel electrophoresis and footprinting assays were performed by J. Trauger, L. Neely, and L. Dickenson, all *in vitro* transcription experiments were performed by L. Neely, L. Dickenson, and J. Gottesfeld, *in vivo* TFIIIA inhibition was performed by J. Gottesfeld, and HIV-1 virus replication assays were performed by P. Guzilia.

## References

1. (a) Moser, H.E.; Dervan, P.B. *Science* **1987**, 238, 645–650. (b) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.L.; Thoung, N.T.; Lhomme, J.; Helene, C. *Nucleic Acids Res.* **1987**, 15, 7749. (c) Strobel, S.A.; Doucetestamm, L.A.; Riba, L.; Housman, D.E.; Dervan, P.B. *Science* **1991**, 254, 1639–1642. (d) Thuong, N.T.; Helene, C. *Angew. Chem. Int. Ed. Engl.* **1993**, 32, 666–690.
2. (a) Maher, J. L.; Dervan, P. B.; Wold, B. *Biochemistry* **1992**, 31, 71. (b) Duvalentin, G.; Thuong, N.T.; Helene, C.; *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 504.
3. (a) Ho, S. N.; Boyer, S. H.; Schreiber, S. L.; Danishefsky, S. J.; Crabtree, G.R. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, 91, 9203. (b) Liu, C.; Smith, B. M.; Ajito, K.; Komatsu, H.; Gomezpaloma, L.; Li, T. H.; Theodorakis, E. A.; Nicolaou, K. C.; Vogt, P. K. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, 93 940.
4. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, 114, 8783. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 7586. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, 32, 11385. (d) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, 115, 2572. (e) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, 115, 4474.
5. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, 86, 5723. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, 112, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingham, M. *Nature Struct. Biol.* **1994**, 1, 169. (d) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, 35, 12532.
6. Pilch, D. S.; Poklar, N. A.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, 93, 8306.
7. Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nature Struct. Biol.* **1988**, 5, 104.
8. (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, 116, 7983. (b) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, 118, 6147. (c) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, 118, 6153. (d) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Chem. & Biol.* **1996**, 3, 369. (e) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, 118, 8198. (f) de Claire, R. P. L.; Geierstanger B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, 119, 7909. (g) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, 119, 8756. (h) White, S.; Baird, E. E.; Dervan, P. B. *Chem & Biol.* **1997**, 4, 569.
9. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, 382, 559.
10. Schlissel, M.S.; Brown, D.D. *Cell* **1984**, 37, 903–913.
11. (a) Engelke, D.R.; Ng, S.-Y.; Shastry, B.S.; Roeder, R.G. *Cell* **1980**, 19, 717. (b) Wolffe, A.P.; Brown, D. D. *Science* **1988**, 241, 1626–1632.
12. (a) Fairall, L.; Rhodes, D.; Klug, A. *J. Mol. Biol.* **1986**, 192, 577. (b) Sakonju, S.; Brown, D. D. *Cell* **1982**, 31, 395. (c) Clemens, K.R.; Liao, X.B.; Wolf, V.; Wright, P.E.;



- Gottesfeld, J.M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 10822–10826. (d) Hayes, J.J.; Tullius, T.D. *J. Mol. Biol.* **1992**, 227, 407–417. (e) Clemens, K.R. *et al. J. Mol. Biol.* **1994**, 244, 23–35.
13. Stutz, F.; Gouilloud, E.; Clarkson, S.G. *Genes Dev.* **1989**, 3, 1190.
14. Goodrich, L.A.; Tjian, R. *Curr. Op. Cell Biol.* **1994**, 6, 403–409.
15. Kim, J.L.; Nikolev, D.B.; Burley, S.K. *Nature* **1993**, 365, 520–527.
16. Jones, K.A.; Peterlin, B.M. *Ann. Rev. Biochem.* **1994**, 63, 717–743.
17. (a) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, 118, 6160. (b) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Chem. & Biol.* **1996**, 3, 369. (c) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer D. E.; *Nature Struct. Biol.* **1996**, 3, 321.
18. Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, 120, 6219 .
19. Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods Enzymol.* **1986**, 130, 132. (b) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, 83, 8462. (c) Senear, D.F.; Brenowitz, M.; Shea, M.A.; Ackers, G.K. *Biochemistry* **1986**, 25, 7344.
20. (a) Van Dyke, M. W.; Dervan, P. B. *Nucl. Acids Res.* **1983**, 11, 5555. (b) Van Dyke, M. W.; Dervan, P. B. *Science* **1984**, 225, 1122.
21. Swalley, S.E.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, 119, 6953.
22. Kelly, J.J.; Baird, E.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, 91, 9203.
23. Jones, R.J.; Lin, K.Y.; Milligan, J.F.; Wadwani, S.; Matteucci, M.D. *J. Org. Chem.* **1993**, 58, 2983.
24. Gottesfeld, J.M.; Neely, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1997**, 387, 202.
25. Frech, K.; Brack-Werner, R.; Werner, T. *Virology* **1996**, 224, 256.
26. Swalley, S.E.; Baird, E.E.; Dervan, P.B. *Chem. Eur. J.* **1997**, 3, 1600.
27. Burley, S. K.; Roeder, R. G. *Ann. Rev. Biochem.* **1996**, 65, 769.
28. Kim, J.; Gonzales-Scarano, F.; Zeichner, S.; Alwine, J. J. *Virol* **1993**, 67, 1658.
29. Klotman, M. E.; Kim, S.; Buchbinder, A.; DeRossi, A.; Baltimore, D.; Wong-Staal, F. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, 88, 5011.
30. Clark, S. J.; Saag, M. S.; Decker, W. D.; Campbell-Hill, S.; Roberson, J. L.; Veldkamp, P. J.; Kappes, J. C.; Hahn, B. H.; Shaw, G. M. *N. Engl. J. Med.* **1991**, 324, 954.

31. Borrow, P.; Lewicki, H.; Wei, X.; Horwitz, M. S.; Pfeffer, N.; Meyers, H.; Nelson, J. A.; Gairin, J. E.; Hahn, B. H.; Oldstone, M. B.; Shaw, G. M. *Nature Med.* **1997**, *3*, 205.
32. Cheng-Mayer, C.; Seto, D.; Tateno, M.; Levy, J. A. *Science* **1988**, *240*, 80.
33. Cheng-Mayer, C.; Weiss, C.; Seto, D.; Levy, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8575.
34. Hobbs, M. V.; Weigle, W. O.; Noonan, D. J.; Torbett, B. E.; McEvilly, R. J.; Koch, R. J.; Cardenas, G. J.; Ernst, D. N. *J. Immunol.* **1993**, *150*, 3602.
35. Rochford, R.; Miller, C. L.; Cannon, M. J.; Izumi, K. M.; Kieff, E.; Longnecker, R. *Arch. Virol.* **1997**, *142*, 707.
36. Oakley, M. G.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10969.
37. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534.
38. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Angew. Chemie. Int. Ed.* **1998**, *37*, 1421.

## CHAPTER 8

### Recognition of the Four Watson-Crick Base Pairs in the DNA Minor Groove by Synthetic-Ligands

**Abstract** *The design of synthetic ligands that read the information stored in the DNA double helix has been a long standing goal at the interface of chemistry and biology. Cell-permeable small molecules which target predetermined DNA sequences offer a potential approach for the regulation of gene-expression. Oligodeoxynucleotides that recognize the major groove of double-helical DNA via triple-helix formation bind to a broad range of sequences with high affinity and specificity. Although oligonucleotides and their analogs have been shown to interfere with gene expression, the triple helix approach is limited to purine tracks and suffers from poor cellular uptake. The subsequent development of pairing rules for minor groove binding polyamides containing pyrrole (Py) and imidazole (Im) amino acids offers a second code to control sequence specificity. An Im/Py pair distinguishes G•C from C•G and both of these from A•T/T•A base pairs. A Py/Py pair specifies A,T from G,C but appears not to distinguish A•T from T•A. To test the extent of this degeneracy, the affinity and binding orientation of the hairpin polyamide ImPyPy-γ-PyPyPy-β-Dp was measured for eight possible five base pair 5'-TG(A,T)<sub>3</sub>-3' match sites. Affinity cleavage experiments using a polyamide with an EDTA•Fe(II) moiety at the carboxy terminus, ImPyPy-γ-PyPyPy-β-Dp-EDTA•Fe(II), are consistent with formation of an oriented 1:1 hairpin polyamide complex at all eight 5'-TG(A,T)<sub>3</sub>-3' binding sites. Quantitative DNase I footprint titration experiments reveal that ImPyPy-γ-PyPyPy-β-Dp binds all eight 5'-TG(A,T)<sub>3</sub>-3' target sites with only a 12-fold difference in the equilibrium association constants between the strongest site, 5'-TGTTT-3' ( $K_a = 2.1 \times 10^8 \text{ M}^{-1}$ ) and the weakest site, 5'-TGAAT-3' ( $K_a = 1.8 \times 10^7 \text{ M}^{-1}$ ). This relatively small range indicates that the Py/Py pair is approximately degenerate*

for recognition of A,T base pairs, affording generality with regard to targeting sequences of mixed A•T/T•A composition.

*In order to break this degeneracy, a new aromatic amino acid, 3-hydroxypyrrole (Hp), has been added to the repertoire to test for pairings which discriminate A•T from T•A. We find that replacement of a single hydrogen atom with a hydroxy group in a Hp/Py pairing regulates affinity and specificity by an order of magnitude. By incorporation of a third amino acid, hydroxypyrrole-imidazole-pyrrole polyamides form four ring-pairings (Im/Py, Py/Im, Hp/Py, and Py/Hp) which distinguish all four Watson-Crick base pairs in the minor groove of DNA.*

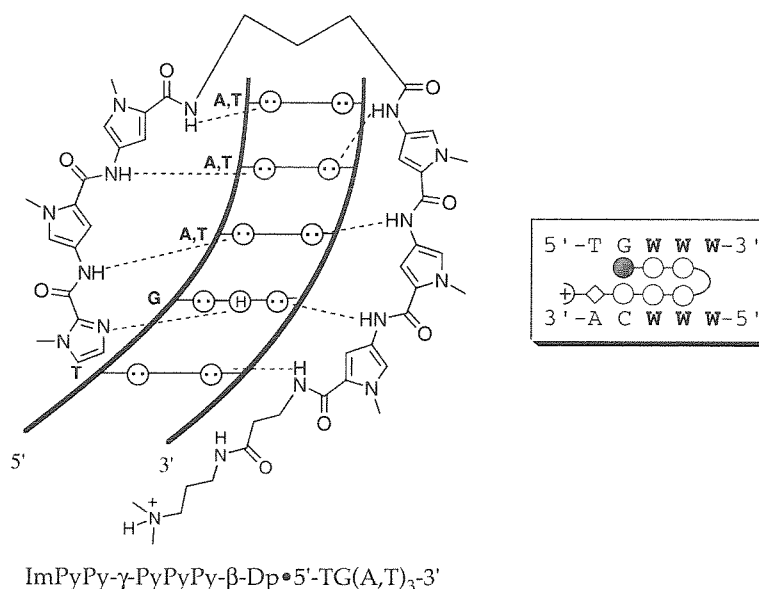
**Publications:** White, Baird, & Dervan *Biochemistry* **1996**, 35, 12532-12537.

White, Szewczyk, Turner, Baird, & Dervan *Nature* **1998**, 391, 468.

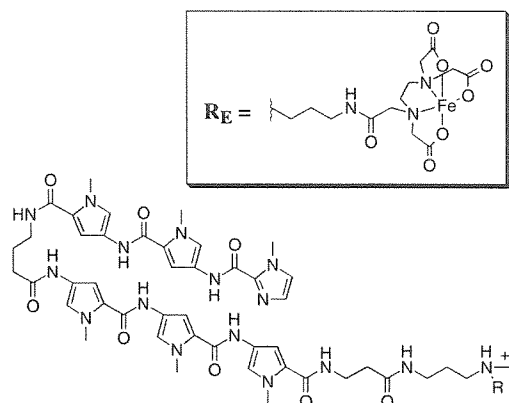
Kielkopf, White, Szewczyk, Turner, Baird, Dervan, & Rees  
*Science* **1998**, 282, 111.

**Effects of A•T/T•A Degeneracy for Recognition of the Minor Groove by Py-Im Polyamides.** Py-Im polyamide-DNA complexes provide a paradigm for the design of artificial molecules for recognition of double helical DNA.<sup>1-7</sup> Polyamides containing Im and Py amino acids can be combined in antiparallel side-by-side dimeric complexes with the minor groove of DNA. The DNA-binding sequence specificity of these small molecules depends on the sequence of side-by-side amino acid pairings.<sup>1</sup> A pairing of Im opposite Py recognizes a G•C base pair, while a Py/Im combination targets a C•G base pair.<sup>1</sup> A Py/Py pair has apparent degeneracy for A•T/ T•A base pairs.<sup>1,3</sup>

The discrimination of G•C from C•G base pairs and both of these from A•T/ T•A base pairs by pyrrole-imidazole polyamides has been demonstrated.<sup>1</sup> A key issue is to determine if a hairpin polyamide of core sequence composition ImPyPy- $\gamma$ -PyPyPy would bind all possible 5'-(A,T)G(A,T)<sub>3</sub>-3' target sequences. Alternatively, given the sequence dependent variation in DNA groove width,<sup>5</sup> perhaps only a smaller subset of 5'-(A,T)G(A,T)<sub>3</sub>-3' sequences would



**Figure 8.1.** (Top) Model for the complex formed between the hairpin polyamide ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp with a 5'-TG(A,T)<sub>3</sub>-3' site. Circles with dots represent lone pairs of N3 purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. W is either A•T or T•A.



$R = \text{CH}_3$  ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp

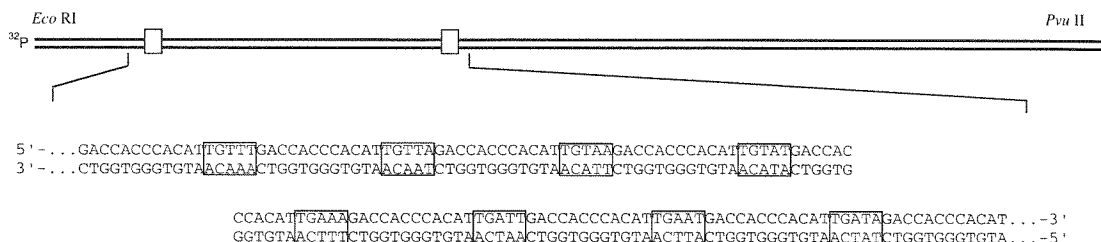
$R = R_E$  ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA•Fe(II)

**Figure 8.2.** Structure of the hairpin polyamides ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp and ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA•Fe(II).

be structurally compatible with polyamide-DNA complex formation.

To address this question, a plasmid was designed containing eight binding sites of the form 5'-TG(A,T)<sub>3</sub>-3'; 5'-TGTTT-3', 5'-TGTTA-3', 5'-TGTA-3', 5'-TGTAT-3', 5'-TGAAA-3', 5'-TGATT-3', 5'-TGAAT-3', and 5'-TGATA-3', with each site flanked by the same 12 base pair sequence (Figure 8.3). Quantitative DNase I footprint titration experiments afford a comparison of the equilibrium association constant for binding of each site by the polyamide ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp, previously optimized for a hairpin motif.<sup>4</sup> For controls, affinity cleavage experiments with ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA•Fe(II) confirm that the polyamide binds each five base pair site in a single orientation, supporting the hairpin model. We report here that ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp binds *all sites* of the form 5'-TG(A,T)<sub>3</sub>-3' with a 12-fold range between the highest and lowest observed affinities.

Affinity Cleavage experiments<sup>8</sup> using a polyamide with Fe(II)•EDTA at the carboxy terminus confirm that ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp binds each discrete site with a *single orientation*. Cleavage experiments with ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp-EDTA•Fe(II) were performed on the 370 base pair restriction fragment radiolabeled at either the 5' or 3' end (20 mM HEPES, 200 mM NaCl, 50 mg/ml glycogen, pH 7.0, 22°C, 5 mM DTT, 1 mM Fe(II)). A single cleavage locus is observed proximal to the 5'-side of each of the eight 5'-



**Figure 8.3.** Partial sequence of the 370 base pair *Eco* RI/ *Pvu* II restriction fragment. Eight 5 base pair binding sites having the sequence 5'-TG(A,T)<sub>3</sub>-3' proximal to the <sup>32</sup>P label at the *Eco* RI site were analyzed by quantitative footprint titration analysis.

TG(A,T)<sub>3</sub>-3' binding sites indicating that the carboxy terminus of the polyamide is located at the 5'-side of each binding site. A 3'-shifted asymmetric cleavage patterns is consistent with location of the 1:1 polyamide complex in the minor groove. The observation of a single cleavage locus is consistent only with an oriented 1:1 complex and rules out any 2:1 overlapped or extended binding motifs.<sup>6</sup> A 1:1 oriented but extended motif would require at least an 8 base pair binding site which is inconsistent with high resolution MPE footprinting.<sup>4b</sup> The hairpin structure is supported by direct NMR structure studies.

DNase I footprinting<sup>9</sup> on the 3'-<sup>32</sup>P end-labeled 370 base pair *Eco*RI/ *Pvu* II restriction fragment from the plasmid pDEH1 (10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>, pH 7.0, 22°C) reveals the equilibrium association constants for ImPyPy-γ-PyPyPy-β-Dp binding each of the 5 base pair sites, 5'-TG(A,T)<sub>3</sub>-3', ranges from  $K_a = 2.1 \times 10^8 \text{ M}^{-1}$  to  $K_a = 1.8 \times 10^7 \text{ M}^{-1}$  in decreasing order: 5'-TGTTT-3' > 5'-TGTTA-3' > 5'-TGTA-3' > 5'-TGTAT-3' > 5'-TGATT-3' > 5'-TGATA-3' > 5'-TGAAA-3' > 5'-TGAAT-3' (Table 8.1). The polyamide displays a binding isotherm consistent with binding as an intramolecular hairpin at all sites.

The affinities of ImPyPy-γ-PyPyPy-β-Dp for binding sites of the type 5'-TG(A,T)<sub>3</sub>-3' may be grouped into two sets according to sequence composition: 5'-TGT(A,T)<sub>2</sub>-3' and 5'-TGA(A,T)<sub>2</sub>-3'. ImPyPy-γ-PyPyPy-β-Dp binds 5'-TGT(A,T)<sub>2</sub>-3' sites with between 2-fold and 12-fold higher affinity than 5'-TGA(A,T)<sub>2</sub>-3' sites. Overall, sequence composition

**Table 8.1.** Equilibrium Association Constants for ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp<sup>a,b</sup>.

Binding Site	$K_a$ (M <sup>-1</sup> )	$K_a$ (rel) <sup>c</sup>	$\Delta G$ (kcal/mol)
5'-TGTTT-3'	$2.1 (\pm 0.7) \times 10^8$	12	-11.2
5'-TGTTA-3'	$1.5 (\pm 0.4) \times 10^8$	8.4	-11.0
5'-TGTA-3'	$7.3 (\pm 1.0) \times 10^7$	4.1	-10.6
5'-TGTAT-3'	$4.7 (\pm 0.8) \times 10^7$	2.6	-10.4
5'-TGATT-3'	$3.9 (\pm 1.0) \times 10^7$	2.2	-10.3
5'-TGATA-3'	$2.5 (\pm 0.9) \times 10^7$	1.4	-10.0
5'-TGAAA-3'	$2.2 (\pm 0.9) \times 10^7$	1.2	-9.9
5'-TGAAT-3'	$1.8 (\pm 0.8) \times 10^7$	1	-9.8

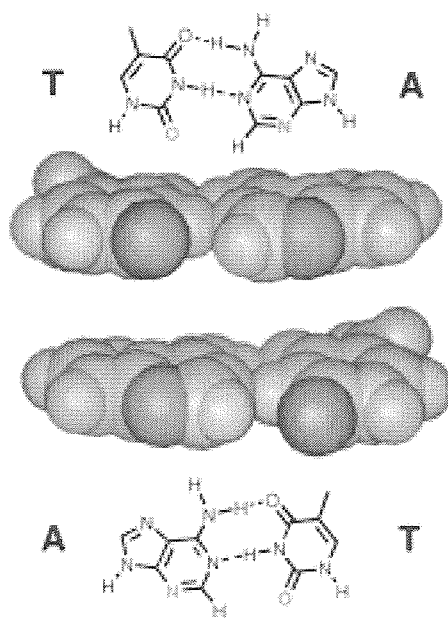
<sup>a</sup>Mean values measured from four footprint titration experiments, with the standard deviation for each data set indicated in parentheses. <sup>b</sup>The assays were performed at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM MgCl<sub>2</sub>, 10 mM KCl, and 5 mM CaCl<sub>2</sub>. <sup>c</sup> $K_a$  (rel) =  $K_a(5'-TGWW-3')/K_a(5'-TGAAT-3')$ , W is either A or T.

results in a difference of up to 1.5 kcal/mole for recognition of 5'-TG(A,T)<sub>3</sub>-3' sites. X-ray diffraction data suggest that a G-A step acts to narrow the minor groove of B-form DNA.<sup>10</sup> A decrease in minor groove width and flexibility would act to disfavor binding of a hairpin polyamide, which prefers a wide, flexible minor groove for favorable binding.<sup>7</sup>

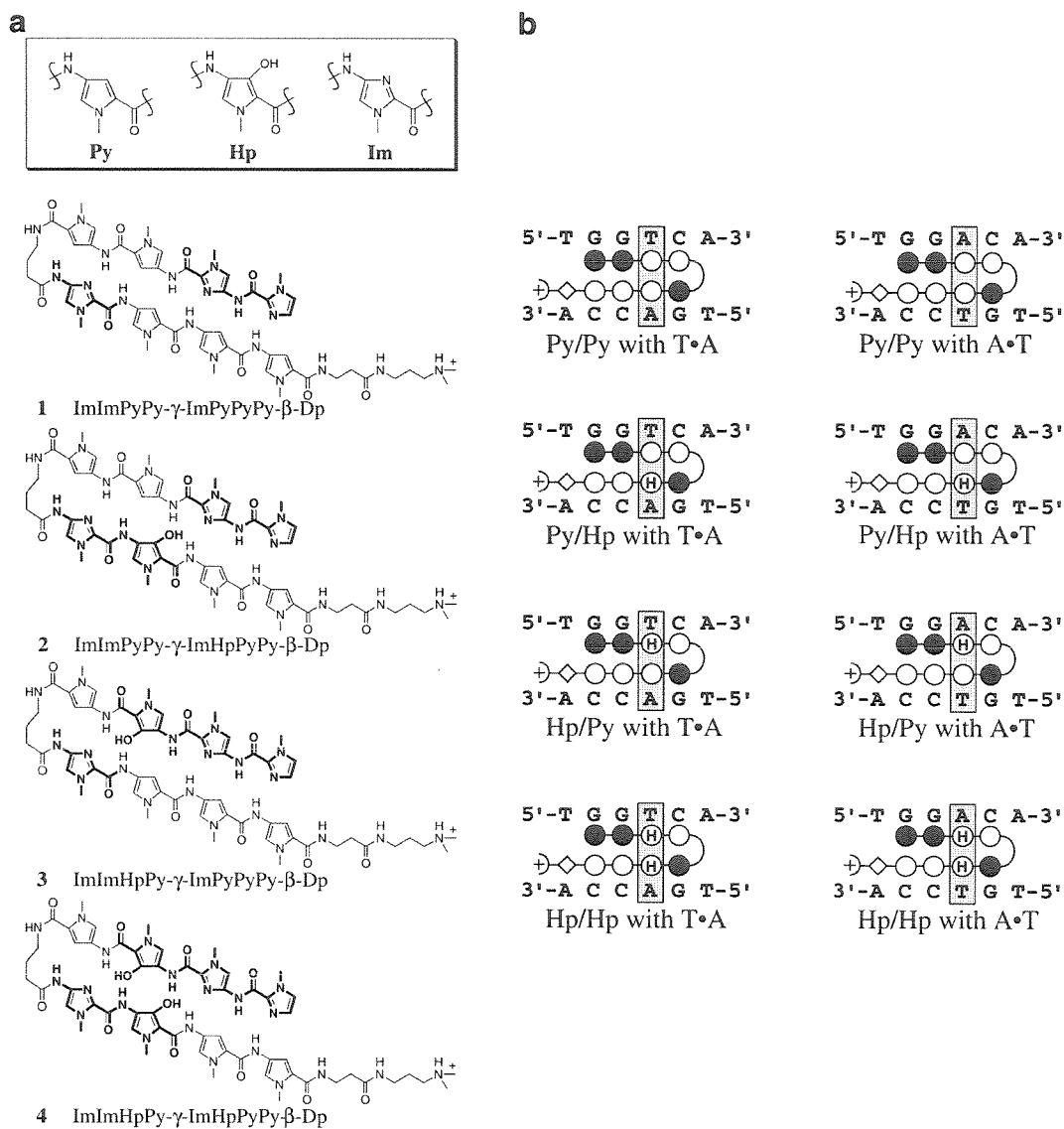
The results reported here indicate that A•T and T•A base pairs are degenerate in the hairpin polyamide-DNA motif with a 12-fold difference in binding affinities. These results indicate that at least a 10-fold *range* of binding affinities and sequence specificities will be observed for a polyamide binding to a designated set of match sites containing A•T base pairs. The similarity of the polyamide binding affinities for the eight 5'-TG(A,T)<sub>3</sub>-3' match sites reflects a limit to the specificity of the hairpin polyamide binding motif. Because G•C is distinct from C•G, the most specific recognition will be G,C rich sequences. In order to increase the specificity of pyrrole-imidazole polyamides for sequences of rich A,T composition, methods of recognition to discriminate A•T and T•A base pairs are needed.



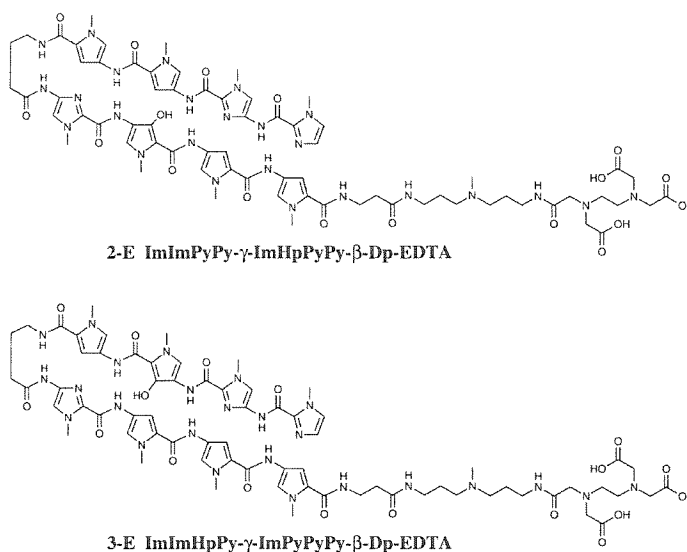
**Recognition of the Four Watson-Crick Base Pairs in the DNA Minor Groove by Synthetic-Ligands.** Due to degeneracy of the hydrogen bond donors and acceptors displayed on the edges of the base pairs, the minor groove was thought to lack sufficient information for a complete recognition code.<sup>1</sup> However, despite the central placement of the guanine exocyclic N2 amine group in the G,C minor groove,<sup>11-12</sup> Py/Im and Im/Py pairings distinguish energetically G•C and C•G.<sup>1-7</sup> The neighboring Py packs an Im to one side of the minor groove resulting in a precisely placed hydrogen bond between Im N3 and guanine N2 for specific recognition.<sup>13-14</sup> This remarkable sensitivity to single atomic replacement indicates that substitution at the 3 position of one Py within a Py/Py pair can compliment small structural differences at the edges of the base pairs in the center of the minor groove.



**Figure 8.4.** Chemical structures and space-filling models of the T•A and A•T base pairs as viewed from the minor groove of DNA. Models generated using B-form DNA coordinates provided in InsightII. The hydrogen-bond acceptors (N3 of adenine and O2 of thymine) are in dark gray.



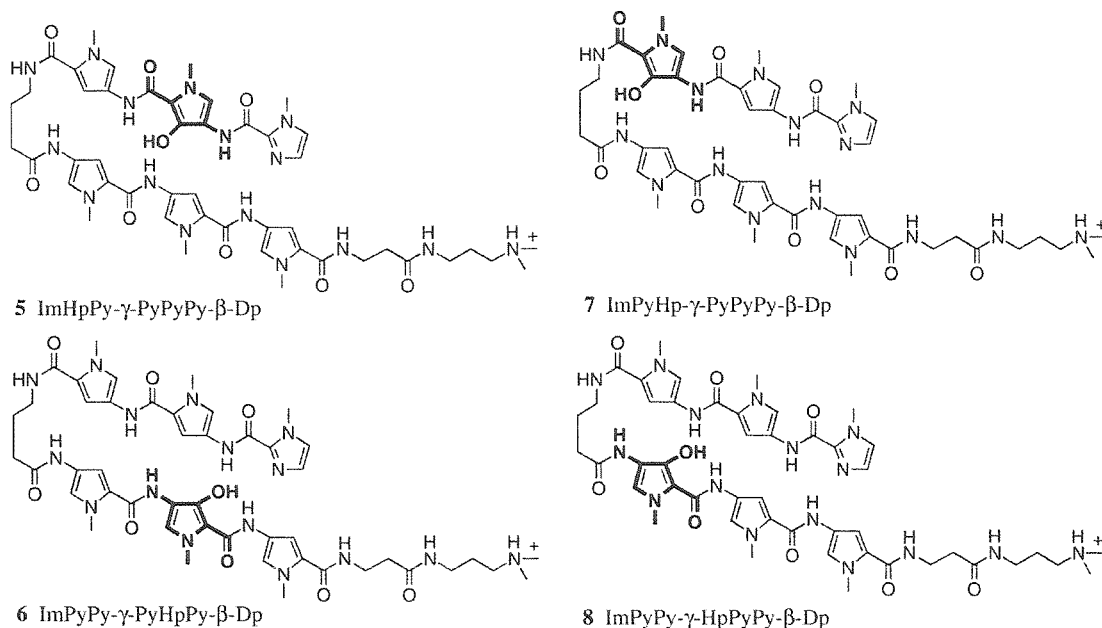
**Figure 8.5.** Eight-ring hairpin polyamides containing three aromatic amino acids (Py, Hp, Im). **a**, Structures of polyamides ImImPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp (1), ImImPyPy- $\gamma$ -ImHpPyPy- $\beta$ -Dp (2), ImImHpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp (3), and ImImHpPy- $\gamma$ -ImHpPyPy- $\beta$ -Dp (4). (Hp = 3-hydroxypyrrole, Im = imidazole, Py = pyrrole,  $\beta$  =  $\beta$ -alanine,  $\gamma$  =  $\gamma$ -aminobutyric acid, Dp = dimethylaminopropylamide). **b**, Binding models for polyamides 1-4 in complex with 5'-TGGTCA-3' and 5'-TGGACA-3' (A•T and T•A in the fourth position highlighted). Filled and unfilled circles represent imidazole and pyrrole rings respectively; circles containing an H represent 3-hydroxypyrrole, the curved line connecting the polyamide subunits represents  $\gamma$ -aminobutyric acid, the diamond represents  $\beta$ -alanine, and the + represents the positively charged dimethylaminopropylamide tail group.



**Figure 8.6.** Chemical structures of affinity cleaving polyamides ImImPyPy-γ-ImHpPyPy-β-Dp-EDTA **2-E** and ImImHpPy-γ-ImPyPyPy-β-Dp-EDTA **3-E**.

For A,T base pairs, the hydrogen bond acceptors at N3 of adenine and O2 of thymine are almost identically placed in the minor groove, making hydrogen bond discrimination a challenge (Figure 8.4).<sup>11</sup> The existence of an asymmetrically placed cleft on the minor groove surface between the thymine O2 and the adenine 2H suggests a possible shape-selective mechanism for A•T recognition.<sup>15</sup> We reasoned that substitution of C3-H by C3-OH within a Py/Py pair would create 3-hydroxypyrrole (Hp)/Py pairings to discriminate T•A from A•T (Figure 8.5). Selectivity could potentially arise from steric destabilization of polyamide binding via placement of Hp opposite A or stabilization by a specific hydrogen bond between Hp and T.

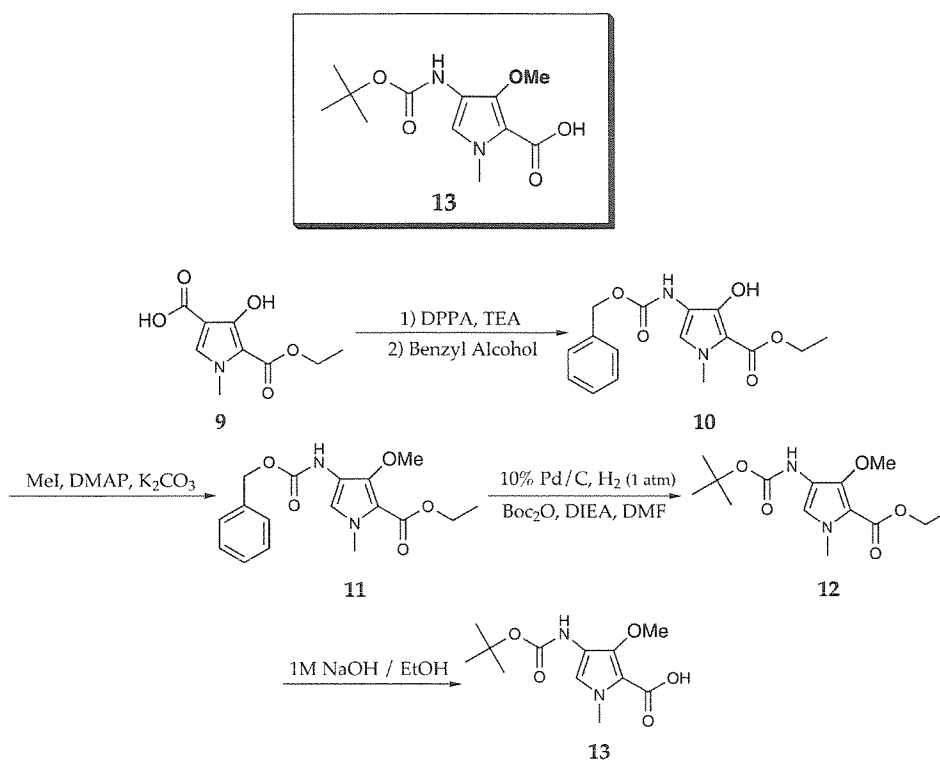
Four-ring polyamide subunits, covalently coupled to form eight-ring hairpin structures, bind specifically to 6-bp target sequences at subnanomolar concentrations.<sup>5</sup> We report here the DNA-binding affinities of three eight-ring hairpin polyamides containing pairings of Im/Py, Py/Im opposite G•C, C•G and either Py/Py, Hp/Py or Py/Hp at a common single point opposite T•A and A•T (Figure 8.5). Equilibrium association constants ( $K_a$ ) for ImImPyPy-γ-ImPyPyPy-β-Dp **1**, ImImPyPy-γ-ImHpPyPy-β-Dp **2**, and



**Figure 8.7.** Chemical structures of 6-ring hairpin hydroxypyrrole-imidazole-pyrrole polyamides. ImPyPy- $\gamma$ -PyPyPy- $\beta$ -Dp has been previously reported.<sup>13</sup>

ImImHpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp **3** were determined by quantitative DNase I footprint titration experiments<sup>9</sup> on a 3' <sup>32</sup>P-labeled 250-bp DNA fragment containing the target sites, 5'-TGGACA-3' and 5'-TGGTCA-3', which differ by a single A,T base pair in the fourth position (Figure 8.9).

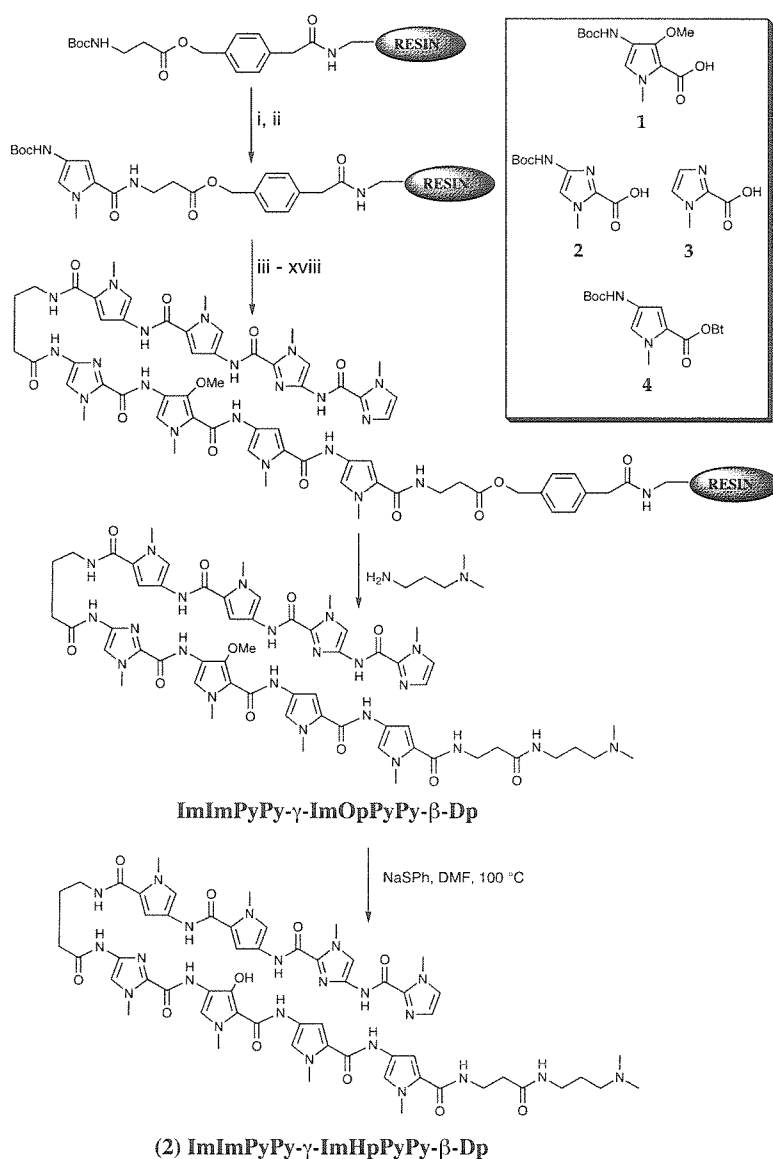
The binding orientations and stoichiometries of the corresponding EDTA analogs ImImPyPy- $\gamma$ -ImHpPyPy- $\beta$ -Dp-EDTA•Fe(II) 2-E•Fe(II) and ImImHpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp-EDTA•Fe(II) 3-E•Fe(II) were determined by affinity cleaving footprint titration experiments (Figures 5.3, 5.8).<sup>8</sup> To test whether a Hp/Hp pair is degenerate for recognition of A•T and T•A base pairs, the polyamide ImImHpPy- $\gamma$ -ImHpPyPy- $\beta$ -Dp **4** (Figure 8.5) was studied for binding versus the sites 5'-TGGACA-3' and 5'-TGGTCA-3' (Table 8.2). In addition, a series of six-ring hairpins containing a single Hp/Py ring pairing was studied versus T•A and A•T base pairs in a variety of sequence contexts to test the generality of the 3 amino acid pairing rules. Equilibrium association constants were determined for ImHpPy- $\gamma$ -PyPyPy- $\beta$ -Dp **5**, ImPyPy- $\gamma$ -PyHpPy- $\beta$ -Dp **6**, ImPyHp- $\gamma$ -PyPyPy- $\beta$ -Dp **7**, and ImPyPy- $\gamma$ -



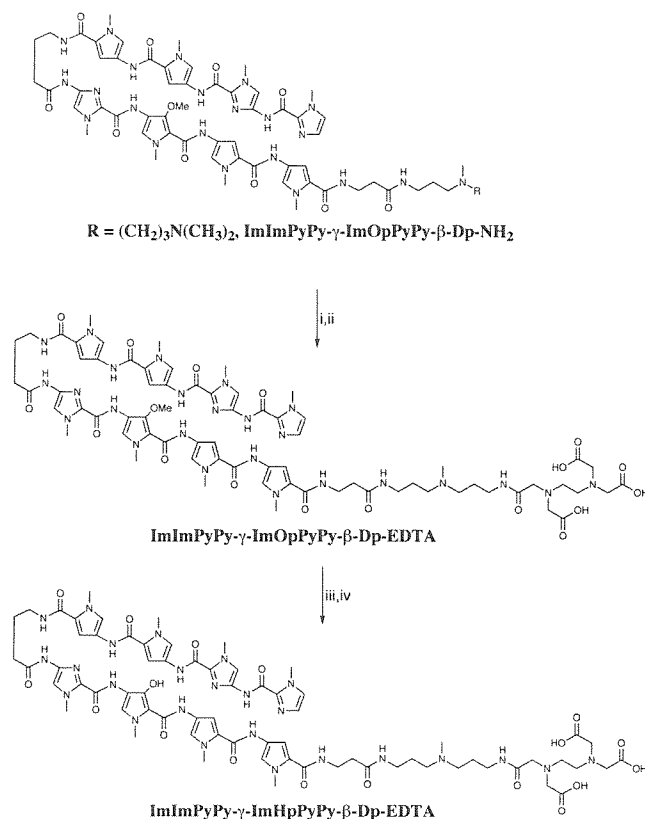
**Figure 8.8** The synthetic scheme for 3-O-methyl-N-Boc protected pyrrole-2-carboxylate **13**. The hydroxypyrrole monoester **9** can be prepared in 0.5 kg quantity using published procedures on enlarged scale.<sup>24</sup>

HpPyPy-β-Dp **8** on a 3'-<sup>32</sup>P end-labeled DNA fragment containing eight possible five base pair 5'-TG(A,T)<sub>3</sub>-3' match sites (Figure 8.3).<sup>16</sup>

Synthesis of 3-hydroxypyrrole amino acid was as follows. The 3-O-methyl-N-Boc protected pyrrole-2-carboxylate was synthesized on multigram scale starting from the previously described hydroxypyrrole monoester **9** (Figure 8.8).<sup>17</sup> First, the 4-carboxy-3-hydroxypyrrole monoester was converted to 4-benzyl carbamate 3-hydroxypyrrole **10** using DPPA (TEA, CH<sub>3</sub>CN, reflux, 5 h), followed by the addition of benzyl alcohol (reflux, 12 h). Then, the 3-hydroxy group of the pyrrole was protected as the methyl ether, giving **11** via alkylation of **10** using methyl iodide (anhydrous K<sub>2</sub>CO<sub>3</sub>, DMAP, 22°C, 12 h).



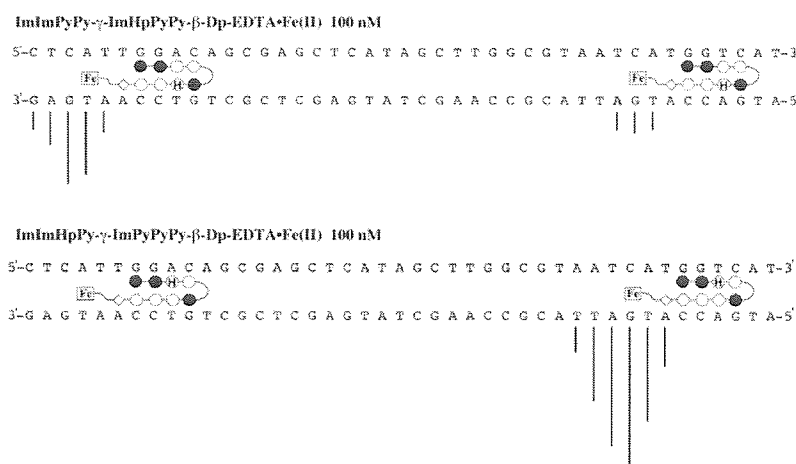
**Figure 8.9** Solid phase synthetic scheme for ImImPyPy- $\gamma$ -ImHpPyPy- $\beta$ -Dp starting from Boc- $\beta$ -Pam-Resin: (i) 80% TFA/DCM, 0.4 M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4 M PhSH; (iv) Boc-Py-OBt, DIEA, DMF; (v) 80% TFA/DCM, 0.4 M PhSH; (vi) Boc-3-OMe-Py-OH, HBTU, DMF, DIEA; (vii) 80% TFA/DCM, 0.4 M PhSH; (viii) Boc-Im-OH, DCC, HOBT; (ix) 80% TFA/DCM, 0.4 M PhSH; (x) Boc- $\gamma$ -aminobutyric acid, DIEA, DMF; (xi) 80% TFA/DCM, 0.4 M PhSH; (xii) Boc-Py-OBt, DIEA, DMF; (xiii) 80% TFA/DCM, 0.4 M PhSH; (xiv) Boc-Py-OBt, DIEA, DMF; (xv) 80% TFA/DCM, 0.4 M PhSH; (xvi) Boc-Im-OH, DCC, HOBT; (xvii) 80% TFA/DCM, 0.4 M PhSH; (xviii) imidazole-2-carboxylic acid, HBTU, DIEA; (xix) dimethylaminopropylamine, 55°C, 18 h. Purification by reversed phase HPLC provides ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Dp. (Op = 3-methoxypyrrole). Treatment of the 3-methoxypyrrole polyamide with thiophenol, NaH, DMF at 100°C for 120 min provides polyamide **2** after purification.



**Figure 8.10.** Synthesis of bifunctional Hp-Py-Im polyamide. Treatment of ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Pam-Resin with 3,3'-diamino-*N*-methyldipropylamine, 55°C, 18 h followed by HPLC purification provides the Op polyamide. Treatment with (i) EDTA-dianhydride, DMSO/NMP, DIEA, 55°C; (ii) 0.1 M NaOH, followed by HPLC purification provides the Op-polyamide-EDTA conjugate. Treatment with thiophenol, NaH, DMF, at 100°C for 2 h provides polyamide **2-E** after reversed phase HPLC purification.

The 4-benzyl carbamate 3-methoxypyrrole **11** was converted to the 4-Boc protected 3-methoxypyrrole **12** by removal of the benzyl group via reduction, then *in situ* reaction of the 4-amine group with Boc anhydride ( $H_2$ , 1 atm, Pd/C, DIEA, 2.1 h). Finally, the 2-ethyl ester of Boc-protected 3-methoxypyrrole **12** was hydrolysed (1M NaOH, 22°C, 4 days) to give the 3-O-methyl-N-Boc protected pyrrole 2-carboxylate **13** for use in solid phase protocols.

Hydroxypyrrole-imidazole-pyrrole polyamides were then synthesized using previously described solid phase protocols starting from the commercially available Boc- $\beta$ -Pam-Resin (Figure 5.6).<sup>18</sup> In machine synthesis protocols, 3-hydroxypyrrole-Boc-amino acid was incorporated by placing the amino acid and an equivalent of HBTU in a machine



**Figure 8.11.** Affinity cleaving with hairpin polyamides ImImPyPy-γ-ImHpPyPy-β-Dp-EDTA•Fe(II) **2-E•Fe(II)** and ImImHpPy-γ-ImPyPyPy-β-Dp-EDTA•Fe(II) **3-E•Fe(II)**. Affinity cleavage patterns of **2-E•Fe(II)** and **3-E•Fe(II)** at 100 nM bound to 5'-TGGACA-3' and 5'-TGGTCA-3'. Bar heights are proportional to the relative cleavage intensities at each base pair.

synthesis cartridge. Upon automated delivery of DMF (2 mL) and DIEA (1 mL) activation occurs. The methyl-protected polyamides **1-8** were synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc-β-Pam-Resin, cleaved from the resin with dimethylaminopropylamine, and purified by reversed phase HPLC (Figure 5.6).

The 3-methoxypyrrole amino acids were deprotected by treating the polyamide with sodium thiophenoxide in DMF at 100°C for 2 h. For affinity cleaving analogs of polyamides **2** and **3** (Figure 5.7), however, the polyamides were cleaved from the resin using neat 3,3'-diamino-*N*-methyldipropylamine (55°C, 16 h). The methyl-protected amine modified polyamides **2-NH<sub>2</sub>** and **3-NH<sub>2</sub>**, after purification by reversed phase HPLC, were reacted with excess EDTA-dianhydride (DMSO/NMP, DIEA, 55°C, 25 min.) and the remaining anhydride was hydrolyzed (0.1 M NaOH, 55°C, 10 min.). After HPLC purification, the EDTA modified polyamides were deprotected with sodium thiophenoxide to give the affinity cleaving analogs **2-E** and **3-E**. Polyamide identity and purity were confirmed by <sup>1</sup>H NMR, analytical HPLC, and MALDI-TOF mass spectrometry.



**Table 8.2.** Equilibrium Association Constants ( $M^{-1}$ )<sup>a</sup>.

Polyamide <sup>b</sup>		$K_a$		$K_{rel}$ <sup>c</sup>
		5'-TGG <u>T</u> CA-3'	5'-TGG <u>A</u> CA-3'	
1	Py/Py	$1.3 (\pm 0.9) \times 10^{10}$	$6.8 (\pm 0.4) \times 10^9$	2
2	Py/Hp	$6.8 (\pm 1.0) \times 10^7$	$1.2 (\pm 0.9) \times 10^9$	0.06
3	Hp/Py	$2.1 (\pm 0.7) \times 10^9$	$2.7 (\pm 0.9) \times 10^7$	77
4	Hp/Hp	$\leq 1 \times 10^7$	$\leq 1 \times 10^7$	--

<sup>a</sup>The reported association constants are the average values obtained from three DNase I footprint titration experiments. The standard deviation for each data set is less than 15% of the reported number. Assays were carried out in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub> at pH 7.0 and 22°C. <sup>b</sup>Ring pairing opposite T•A and A•T in the fourth position of six base pair binding site. <sup>c</sup> $K_{rel}$  is calculated as  $K_a$  (TGGACA)/  $K_a$  (TGGTCA).

Affinity cleaving experiments using hairpin polyamides modified with EDTA•Fe(II) at the C-terminus were used to determine polyamide binding orientation and stoichiometry (Figure 5.8). Affinity cleaving experiments were performed on a 3'-<sup>32</sup>P end-labeled 250-bp pJK6 *Eco*RI/ *Pvu*II restriction fragment (25 mM Tris-acetate, 20 mM NaCl, 100 μM/bp calf thymus DNA, pH 7.0). The observed cleavage patterns are in both cases 3'-shifted, indicating minor groove occupancy. A single cleavage locus proximal to the 5'-side of the binding site is consistent with a 1:1 hairpin:DNA complex in a single orientation at each site.

DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>, pH 7.0, 22°C) were performed to determine the equilibrium association constants  $K_a$  for recognition of bound sites (Figure 5.9, Table 5.1). Based on the pairing rules for polyamide-DNA complexes, the sequences 5'-TGGACA-3' and 5'-TGGTCA-3' are match sites for control polyamide 1 which places a Py/Py pairing opposite A•T and T•A at both sites. We find that polyamide 1 (Py/Py) binds to 5'-TGGTCA-3' and 5'-TGGACA-3' within a factor of 2 ( $K_a = 1.3 (\pm 0.9) \times 10^{10} M^{-1}$  or  $6.8 (\pm 0.4) \times 10^9 M^{-1}$ , respectively). In contrast, polyamide 2 (Py/Hp) binds to 5'-TGGTCA-3' and 5'-TGGACA-3'

with dissociation constants which differ by a factor of 18 ( $K_a = 6.8 (\pm 1.0) \times 10^7 \text{ M}^{-1}$  and  $1.2 (\pm 0.9) \times 10^9 \text{ M}^{-1}$ , respectively). By reversing the pairing in polyamide **3** (Hp/Py) the association constants differ again in the opposite direction by a factor of 77 ( $K_a = 2.1 (\pm 0.7) \times 10^9 \text{ M}^{-1}$  and  $2.7 (\pm 0.9) \times 10^7 \text{ M}^{-1}$  respectively). Control experiments performed on separate DNA fragments reveal that neither a 5'-TGGGCA-3' or a 5'-TGGCCA-3' site is bound by polyamide **2** or **3** with a  $K_a \leq 1.0 \times 10^7 \text{ M}^{-1}$ , indicating that the Hp/Py and Py/Hp ring pairings do not bind opposite G•C or C•G. Additional experiments with polyamide ImImHpPy- $\gamma$ -ImHpPyPy- $\beta$ -Dp **4**, which contains a central Hp/Hp pair, show that the polyamide binds with a  $K_a \leq 1.0 \times 10^7 \text{ M}^{-1}$  to either 5'-TGGTCA-3' or 5'-TGGACA-3', indicating that a Hp/Hp pair is not degenerate for A•T/T•A base pairs and is in fact disfavored (Table 8.2).

The discrimination between A•T vs. T•A. is achieved when the two neighboring base pairs are G•C and C•G (GTC vs. GAC). A general rule would require that the same discrimination be observed when adjacent A•T/T•A base pairs are present in the target sequence. Further sequence composition studies of six-ring hydroxypyrrole-imidazole-pyrrole hairpin polyamides confirm A•T/T•A discrimination at sequence contexts which include TTT, TAA, TAT, TTA, ATT, GTT, GAT, and GTA (Tables 8.3-8.10).

The specificity of **2** and **3** for sites which differ by a single A•T/T•A base pair results from small chemical changes. Replacing the Py/Py pair in **1** with a Py/Hp pairing as in **2**, a single substitution of C3-OH for C3-H, destabilizes interaction with 5'-TGGTCA-3' by 191-fold, a free energy difference of  $3.1 \text{ kcal mol}^{-1}$  ( $13.0 \text{ kJ mol}^{-1}$ ). Interaction of **2** with 5'-TGGACA-3' is destabilized only 6-fold relative to **1**, a free energy difference of  $1.1 \text{ kcal mol}^{-1}$  ( $4.6 \text{ kJ mol}^{-1}$ ). Similarly, replacing the Py/Py pair in **1** with Hp/Py as in **3** destabilizes interaction with 5'-TGGACA-3' by 252-fold, a free energy difference of  $3.2 \text{ kcal mol}^{-1}$ . Interaction of **3** with 5'-TGGTCA-3' is destabilized only 6-fold relative to **1**, a free energy difference of  $1.0 \text{ kcal mol}^{-1}$ .

**Table 8.3.** Discrimination of 5'-TGATT-3' and 5'-TGTTT-3'\*.

Pair <sup>†</sup>	5'-TGATT-3'	5'-TGTTT-3'	$K_{rel}^{\ddagger}$
Py/Py	5'-T G <b>A</b> T T-3'  3'-A C <b>T</b> A A-5' $K_a = 3.8 \times 10^7 \text{ M}^{-1}$	5'-T G <b>T</b> T T-3'  3'-A C <b>A</b> A A-5' $K_a = 2.0 \times 10^8 \text{ M}^{-1}$	5.2
Py/Hp	5'-T G <b>A</b> T T-3'  3'-A C <b>T</b> A A-5' $K_a = 1.5 \times 10^6 \text{ M}^{-1}$	5'-T G <b>T</b> T T-3'  3'-A C <b>A</b> A A-5' $K_a = 8.8 \times 10^5 \text{ M}^{-1}$	0.58
Hp/Py	5'-T G <b>A</b> T T-3'  3'-A C <b>T</b> A A-5' $K_a = 1.3 \times 10^6 \text{ M}^{-1}$	5'-T G <b>T</b> T T-3'  3'-A C <b>A</b> A A-5' $K_a = 1.8 \times 10^8 \text{ M}^{-1}$	133

\*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3' <sup>32</sup>P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment.<sup>13</sup> The assays were carried out at 22°C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

<sup>†</sup>Ring pairing opposite T•A and A•T in the second position

<sup>‡</sup>Calculated as  $K_a(5'\text{-TGTTT-3}')/K_a(5'\text{-TGATT-3}')$ .

**Table 8.4.** Discrimination of 5'-TGATA-3' and 5'-TGTTA-3'\*.

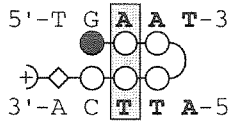
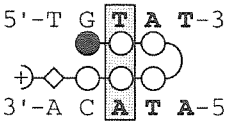
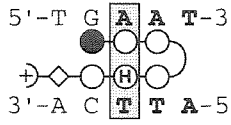
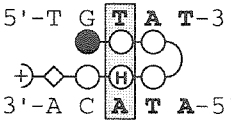
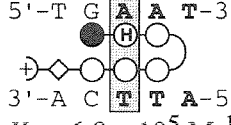
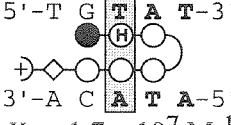
Pair <sup>†</sup>	5'-TGATA-3'	5'-TGTTA-3'	$K_{rel}^{\ddagger}$
Py/Py	5'-T G <b>A</b> T A-3'  3'-A C <b>T</b> A T-5' $K_a = 2.5 \times 10^7 \text{ M}^{-1}$	5'-T G <b>T</b> T A-3'  3'-A C <b>A</b> A T-5' $K_a = 1.4 \times 10^8 \text{ M}^{-1}$	5.7
Py/Hp	5'-T G <b>A</b> T A-3'  3'-A C <b>T</b> A T-5' $K_a = 5.6 \times 10^5 \text{ M}^{-1}$	5'-T G <b>T</b> T A-3'  3'-A C <b>A</b> A T-5' $K_a = 5.6 \times 10^5 \text{ M}^{-1}$	1.02
Hp/Py	5'-T G <b>A</b> T A-3'  3'-A C <b>T</b> A T-5' $K_a = 7.0 \times 10^5 \text{ M}^{-1}$	5'-T G <b>T</b> T A-3'  3'-A C <b>A</b> A T-5' $K_a = 5.9 \times 10^7 \text{ M}^{-1}$	84

\*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3' <sup>32</sup>P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment.<sup>13</sup> The assays were carried out at 22°C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

<sup>†</sup>Ring pairing opposite T•A and A•T in the second position.

<sup>‡</sup>Calculated as  $K_a(5'\text{-TGTTA-3}')/K_a(5'\text{-TGATA-3}')$ .

**Table 8.5.** Discrimination of 5'-TGAAT-3' and 5'-TGTAT-3'\*

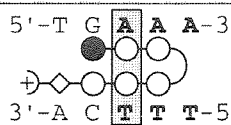
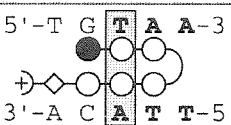
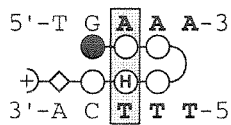
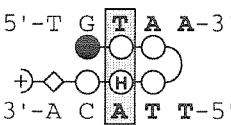
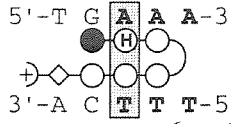
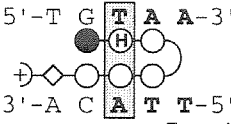
Pair <sup>†</sup>	5'-TGAAT-3'	5'-TGTAT-3'	$K_{rel}^{\ddagger}$
Py/Py	5'-T G <b>A</b> A T-3'  3'-A C <b>T</b> T A-5' $K_a = 1.8 \times 10^7 \text{ M}^{-1}$	5'-T G <b>T</b> A T-3'  3'-A C <b>A</b> T A-5' $K_a = 4.8 \times 10^7 \text{ M}^{-1}$	2.7
Py/Hp	5'-T G <b>A</b> A T-3'  3'-A C <b>T</b> T A-5' $K_a = 3.8 \times 10^5 \text{ M}^{-1}$	5'-T G <b>T</b> A T-3'  3'-A C <b>A</b> T A-5' $K_a = 2.8 \times 10^5 \text{ M}^{-1}$	0.73
Hp/Py	5'-T G <b>A</b> A T-3'  3'-A C <b>T</b> T A-5' $K_a = 6.9 \times 10^5 \text{ M}^{-1}$	5'-T G <b>T</b> A T-3'  3'-A C <b>A</b> T A-5' $K_a = 1.7 \times 10^7 \text{ M}^{-1}$	25

\*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3' <sup>32</sup>P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment<sup>13</sup>. The assays were carried out at 22°C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

<sup>†</sup>Ring pairing opposite T•A and A•T in the second position.

<sup>‡</sup>Calculated as  $K_a(5'\text{-TGTAT-3}')/K_a(5'\text{-TGAAT-3}')$ .

**Table 8.6.** Discrimination of 5'-TGAAA-3' and 5'-TGTA-3'\*

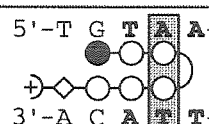
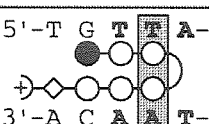
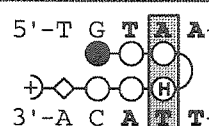
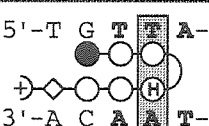
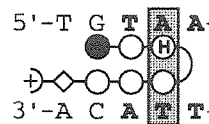
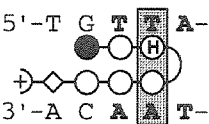
Pair <sup>†</sup>	5'-TGAAA-3'	5'-TGTA-3'	$K_{rel}^{\ddagger}$
Py/Py	5'-T G <b>A</b> A A-3'  3'-A C <b>T</b> T T-5' $K_a = 2.2 \times 10^7 \text{ M}^{-1}$	5'-T G <b>T</b> A A-3'  3'-A C <b>A</b> T T-5' $K_a = 7.1 \times 10^7 \text{ M}^{-1}$	3.3
Py/Hp	5'-T G <b>A</b> A A-3'  3'-A C <b>T</b> T T-5' $K_a = 6.5 \times 10^5 \text{ M}^{-1}$	5'-T G <b>T</b> A A-3'  3'-A C <b>A</b> T T-5' $K_a = 3.7 \times 10^5 \text{ M}^{-1}$	0.57
Hp/Py	5'-T G <b>A</b> A A-3'  3'-A C <b>T</b> T T-5' $K_a = 1.0 \times 10^6 \text{ M}^{-1}$	5'-T G <b>T</b> A A-3'  3'-A C <b>A</b> T T-5' $K_a = 1.9 \times 10^7 \text{ M}^{-1}$	18

\*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3' <sup>32</sup>P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment<sup>13</sup>. The assays were carried out at 22°C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

<sup>†</sup>Ring pairing opposite T•A and A•T in the second position.

<sup>‡</sup>Calculated as  $K_a(5'\text{-TGTA-3}')/K_a(5'\text{-TGAAA-3}')$ .

**Table 8.7. Discrimination of 5'-TGTA-3' and 5'-TGTTA-3'.**

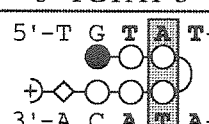
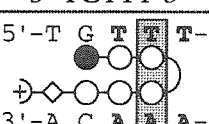
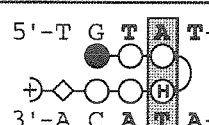
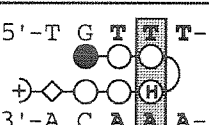
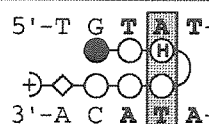
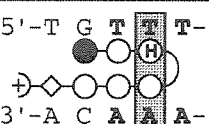
Pair <sup>†</sup>	5'-TGTA-3'	5'-TGTTA-3'	$K_{rel}^{\ddagger}$
Py/Py	5'-T G T A A-3'  3'-A C A T T-5' $K_a = 7.1 \times 10^7 \text{ M}^{-1}$	5'-T G T T A-3'  3'-A C A A T-5' $K_a = 1.4 \times 10^8 \text{ M}^{-1}$	2.0
Py/Hp	5'-T G T A A-3'  3'-A C A T T-5' $K_a = 5.0 \times 10^6 \text{ M}^{-1}$	5'-T G T T A-3'  3'-A C A A T-5' $K_a = 1.8 \times 10^6 \text{ M}^{-1}$	0.36
Hp/Py	5'-T G T A A-3'  3'-A C A T T-5' $K_a = 2.5 \times 10^5 \text{ M}^{-1}$	5'-T G T T A-3'  3'-A C A A T-5' $K_a = 3.6 \times 10^6 \text{ M}^{-1}$	14

\*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3' <sup>32</sup>P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment.<sup>13</sup> The assays were carried out at 22°C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

<sup>†</sup>Ring pairing opposite T•A and A•T in the third position.

<sup>‡</sup>Calculated as  $K_a(5'-TGTTA-3')/K_a(5'-TGTA-3')$ .

**Table 8.8. Discrimination of 5'-TGTAT-3' and 5'-TGTTT-3'.**

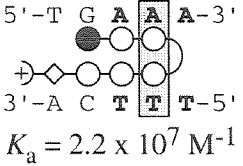
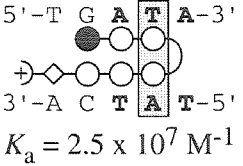
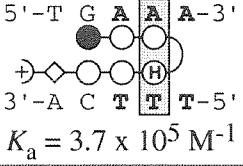
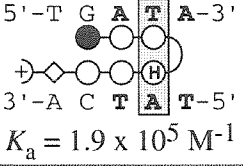
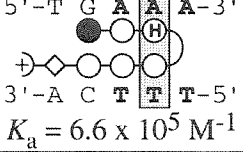
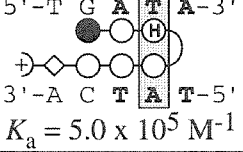
Pair <sup>†</sup>	5'-TGTAT-3'	5'-TGTTT-3'	$K_{rel}^{\ddagger}$
Py/Py	5'-T G T A T-3'  3'-A C A T A-5' $K_a = 4.8 \times 10^7 \text{ M}^{-1}$	5'-T G T T T-3'  3'-A C A A A-5' $K_a = 2.0 \times 10^8 \text{ M}^{-1}$	4.2
Py/Hp	5'-T G T A T-3'  3'-A C A T A-5' $K_a = 1.2 \times 10^6 \text{ M}^{-1}$	5'-T G T T T-3'  3'-A C A A A-5' $K_a = 2.4 \times 10^6 \text{ M}^{-1}$	2.0
Hp/Py	5'-T G T A T-3'  3'-A C A T A-5' $K_a = 3.6 \times 10^5 \text{ M}^{-1}$	5'-T G T T T-3'  3'-A C A A A-5' $K_a = 7.7 \times 10^6 \text{ M}^{-1}$	21

\*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3' <sup>32</sup>P labeled 370-bp pDEH1 *Eco*RI/*Pvu*II DNA restriction fragment.<sup>13</sup> The assays were carried out at 22°C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

<sup>†</sup>Ring pairing opposite T•A and A•T in the third position.

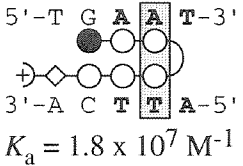
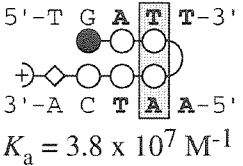
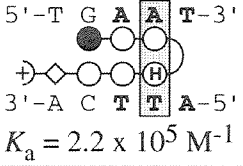
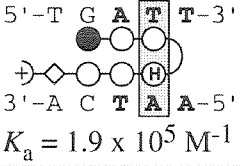
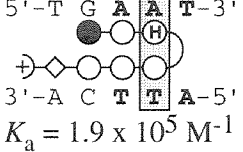
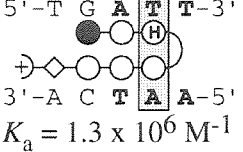
<sup>‡</sup>Calculated as  $K_a(5'-TGTTT-3')/K_a(5'-TGTAT-3')$ .

**Table 8.9.** Discrimination of 5'-TGAAA-3' and 5'-TGATA-3'\*.

Pair†	5'-TGAAA-3'	5'-TGATA-3'	$K_{rel}^{\ddagger}$
Py/Py	 $K_a = 2.2 \times 10^7 \text{ M}^{-1}$	 $K_a = 2.5 \times 10^7 \text{ M}^{-1}$	1.1
Py/Hp	 $K_a = 3.7 \times 10^5 \text{ M}^{-1}$	 $K_a = 1.9 \times 10^5 \text{ M}^{-1}$	0.53
Hp/Py	 $K_a = 6.6 \times 10^5 \text{ M}^{-1}$	 $K_a = 5.0 \times 10^5 \text{ M}^{-1}$	0.76

\*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3' <sup>32</sup>P labeled 370-bp pDEH1 *Eco*RI/ *Pvu*II DNA restriction fragment.<sup>13</sup> The assays were carried out at 22 °C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.  
†Ring pairing opposite T•A and A•T in the third position.  
‡Calculated as  $K_a(5'-TGATA-3')/K_a(5'-TGAAA-3')$ .

**Table 8.10.** Discrimination of 5'-TGAAT-3' and 5'-TGATT-3'\*.

Pair†	5'-TGAAT-3'	5'-TGATT-3'	$K_{rel}^{\ddagger}$
Py/Py	 $K_a = 1.8 \times 10^7 \text{ M}^{-1}$	 $K_a = 3.8 \times 10^7 \text{ M}^{-1}$	2.2
Py/Hp	 $K_a = 2.2 \times 10^5 \text{ M}^{-1}$	 $K_a = 1.9 \times 10^5 \text{ M}^{-1}$	0.88
Hp/Py	 $K_a = 1.9 \times 10^5 \text{ M}^{-1}$	 $K_a = 1.3 \times 10^6 \text{ M}^{-1}$	6.7

\*The reported equilibrium dissociation constants are the mean values obtained from three DNase I footprint titration experiments on a 3' <sup>32</sup>P labeled 370-bp pDEH1 *Eco*RI/ *Pvu*II DNA restriction fragment.<sup>13</sup> The assays were carried out at 22 °C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.  
†Ring pairing opposite T•A and A•T in the third position.  
‡Calculated as  $K_a(5'-TGATT-3')/K_a(5'-TGAAT-3')$ .

**Table 8.11.** Pairing Code for Minor Groove Recognition.

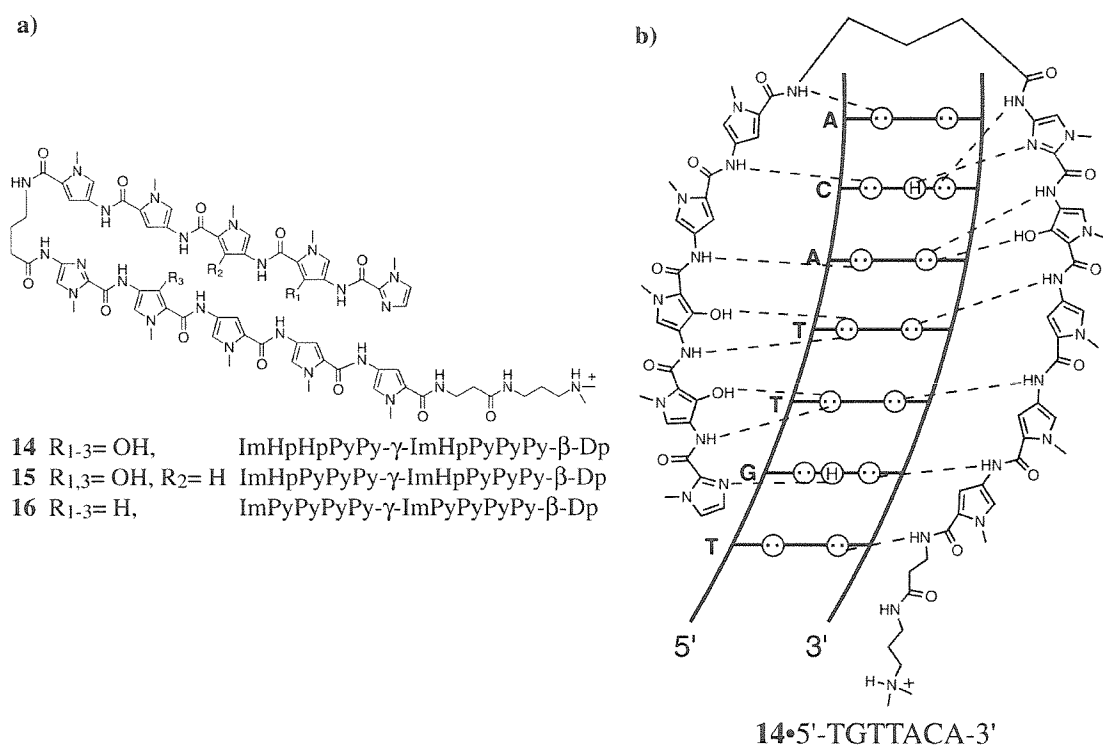
Pair	G•C	C•G	T•A	A•T
Im/Py	+	--	--	--
Py/Im	--	+	--	--
Hp/Py	--	--	+	--
Py/Hp	--	--	--	+

Favored (+), disfavored (--).

Molecular recognition in nature occurs via an ensemble of stabilizing and destabilizing forces.<sup>12</sup> For example the DNA-binding transcription factor, TBP, recognizes 5'-TATA-3' sequences in the DNA minor groove via specific-contacts and a large DNA-bend.<sup>19</sup> For DNA-bending proteins such as TBP, practical sequence-specificity can be enhanced via the sequence-dependent energetics of DNA distortion.<sup>20</sup> Homeodomain proteins recognize target sequences via a combination of specific interactions with both the major and minor grooves.<sup>21</sup> An N-terminal arm recognizes 5'-TAAT-3' sequences in the minor groove such that a single substitution of T•A for A•T reduces binding at 5'-TTAT-3' by a factor of 7.<sup>22</sup> However, no single protein structure motif has been identified which provides a general amino acid-base pair code for the minor groove, although there has been remarkable progress with zinc fingers in the major groove.<sup>23</sup>

Polyamides use a single molecular shape to provide for coded targeting of predetermined DNA sequences with affinity and specificity comparable to sequence-specific DNA binding proteins.<sup>5</sup> Hydroxypyrrole-imidazole-pyrrole polyamides complete the minor groove recognition code using three aromatic amino acids, which combine to form four ring pairings (Im/Py, Py/Im, Hp/Py, and Py/Hp) that complement the four Watson-Crick base pairs (Table 8.11).

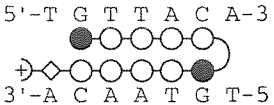
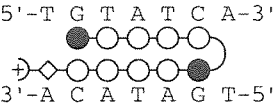
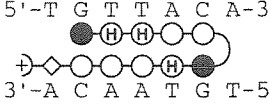
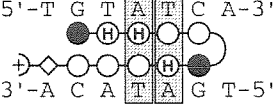
**Hydoxypyrrole-Imidazole-Pyrrole Hairpin Polyamides Coded for Recognition of a 5'-TGTTACA-3' Sequence in the Minor Groove of DNA.** The T•A selectivity of the Hp/Py pair likely arises from a combination of differential destabilization of polyamide binding via placement of Hp/Py opposite A•T or T•A, and specific hydrogen bonds between the 3-hydroxy and 4-carboxamido groups of Hp with the O2 of thymine (Figure 5.11).<sup>24</sup> A general pairing rule would require the same discrimination to be observed for the recognition of *multiple* T•A base pairs within other sequence contexts, including A•T rich sequences. It remains to be determined whether consecutive Hp/Py pairings could target binding sites varying in their A•T base pair sequence composition without compromising polyamide affinity and sequence specificity. We report here that two and three Hp/Py pairs can be



**Figure 8.12.** a) Structures of the ten-ring hairpin polyamides used for this study b) Proposed binding model for the complex formed between the DNA and ImHpHpPyPy-γ-ImHpPyPyPy-β-Dp **14**. Circles with dots represent the lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.



**Table 8.12.** Equilibrium Association Constants ( $K_a$ ).

5'-TTA-3' <sup>d</sup>	5'-TAT-3'	Specificity <sup>e</sup>
		4-fold
$1.0 (\pm 0.4) \times 10^{10} \text{ M}^{-1}$	$2.5 (\pm 0.9) \times 10^9 \text{ M}^{-1}$	
		≥ 94-fold
$9.4 (\pm 1.0) \times 10^8 \text{ M}^{-1}$	$\leq 1 \times 10^7 \text{ M}^{-1}$	

<sup>d</sup>Central 3-base pairs of the 7-base pair binding sites.  
<sup>e</sup>Specificity is calculated as  $K_a$  (5'-TGTTACA-3') /  $K_a$  (5'-TGTATCA-3').

combined within a hairpin template to distinguish core sequence 5'-TTA-3' from 5'-TAT-3' in the DNA minor groove.

Polyamides ImHpHpPyPy-γ-ImHpPyPyPy-β-Dp (**14**) and ImHpPyPyPy-γ-ImHpPyPyPy-β-Dp (**15**) were synthesized by solid-phase methods using Boc-protected 3-methoxypyrrole, imidazole, and pyrrole aromatic amino acids. Polyamide identity and purity were verified by <sup>1</sup>H NMR, analytical HPLC, and matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS-monoisotopic): **14** 1514.7 (1514.7 calculated), **15** 1498.0 (1498.7 calculated).

Application of the pairing rules predicts that the ten-ring hairpin polyamides ImHpHpPyPy-γ-ImHpPyPyPy-β-Dp (**14**) and ImHpPyPyPy-γ-ImHpPyPyPy-β-Dp (**15**) will bind to the 7-base pair site 5'-TGTTACA-3' but will be mismatched with the site 5'-TGTATCA-3' (Figure 5.11). Quantitative DNase I footprint titration experiments performed on a 361-base pair DNA fragment revealed that polyamide **14** prefers the match site 5'-TGTTACA-3' ( $K_a = 9.4 (\pm 1.0) \times 10^8 \text{ M}^{-1}$ ) over 5'-TGTATCA-3' ( $K_a \leq 1 \times 10^7 \text{ M}^{-1}$ ) by a factor of at least 94-fold (Table 8.12).

Table 8.13. Equilibrium Association Constants<sup>a,b</sup>.

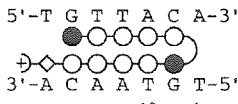
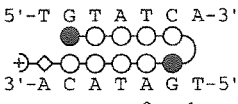
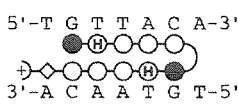
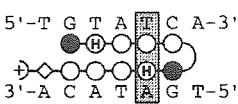
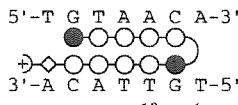
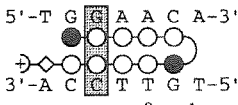
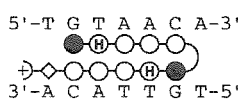
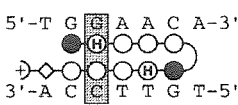
5'-TTA-3' <sup>c</sup>	5'-TAT-3'	Specificity <sup>d</sup>
 $1.2 \times 10^{10} \text{ M}^{-1}$	 $3.0 \times 10^9 \text{ M}^{-1}$	4-fold
 $4.8 \times 10^9 \text{ M}^{-1}$	 $1.5 \times 10^8 \text{ M}^{-1}$	32-fold

Table 8.14. Equilibrium Association Constants<sup>a,b</sup>.

5'-TAA-3' <sup>c</sup>	5'-GAA-3'	Specificity <sup>d</sup>
 $1.2 \times 10^{10} \text{ M}^{-1}$	 $6.8 \times 10^8 \text{ M}^{-1}$	18-fold
 $2.9 \times 10^9 \text{ M}^{-1}$	 $\leq 1 \times 10^6 \text{ M}^{-1}$	$\geq 2900\text{-fold}$

<sup>a</sup>The assays were carried out at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.  
<sup>b</sup>Binding models for polyamides 1 and 2. <sup>c</sup>Central 3-base pairs of the 7-base pair binding sites. <sup>d</sup>Specificity is calculated as  $K_a(5'\text{-TGTTACA-3'}) / K_a(5'\text{-TGTATCA-3'})$ .

Control experiments on a separate fragment reveal that polyamide 14 binds to a single base mismatch site 5'-TGTTCCA-3' with a  $K_a \leq 1 \times 10^7 \text{ M}^{-1}$ , indicating that the Hp/Py pairing is disfavored for placement opposite a G•C base pair.

In comparison, polyamide 15 prefers the match site 5'-TGTTACA-3' ( $K_a = 4.8 \times 10^9 \text{ M}^{-1}$ ) over single base pair mismatch site 5'-TGTATCA-3' ( $K_a = 1.5 \times 10^8 \text{ M}^{-1}$ ) by a factor of 32-fold (Table 8.13). Remarkably, polyamide 15 exhibits much increased specificity for the match site 5'-TGTAACA-3' versus a single G•C base pair mismatch site 5'-TGGAACA-3' ( $K_a \leq 1 \times 10^6 \text{ M}^{-1}$ ), a preference of *over 2900-fold* (Table 8.14).

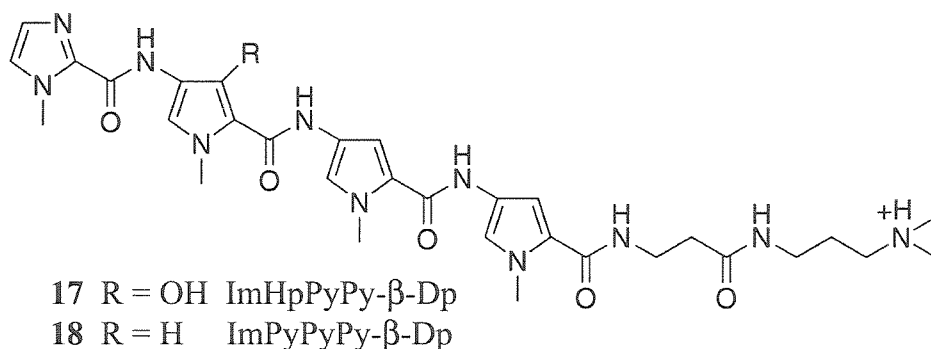
A symmetric pairing of Py/Py is degenerate for placement opposite A•T and T•A base pairs in the minor groove of DNA.<sup>16</sup> We find that control polyamide ImPyPyPyPy- $\gamma$ -ImPyPyPyPy- $\beta$ -Dp **16**, which contains Py/Py pairs rather than Hp/Py pairings, binds within a factor of four to the sites 5'-TGTTACA-3' ( $K_a = 1.0 (\pm 0.4) \times 10^{10} \text{ M}^{-1}$ ) and 5'-TGTATCA-3' ( $K_a = 2.5 (\pm 0.9) \times 10^9 \text{ M}^{-1}$ ), consistent with the assignment of Py/Py as degenerate for A•T and T•A (Table 5.11). In addition, only 18-fold specificity is observed when polyamide **16** is placed opposite the site 5'-TGGAACA-3' versus 5'-TGTAACA-3' (Table 8.12).<sup>25</sup>

In the original report, we observed that a single Hp/Py pair replacing a Py/Py pair destabilizes an eight-ring hairpin polyamide by 5-fold for an identical match site.<sup>24</sup> High resolution x-ray structure analysis reveals that Hp/Py polyamides bind undistorted B-form DNA; however, a localized 0.5 Å melting of the T•A Watson-Crick base pair is observed that is potentially responsible for the energetic destabilization of Hp/Py relative to the Py/Py pair.<sup>26</sup> Remarkably, three Hp/Py pairs destabilize polyamide **14** for 5'-TGTTACA-3' by only 11-fold relative to polyamide **16**, while two Hp/Py pairs destabilize polyamide **15** for 5'-TGTTACA-3' by 3-fold. Therefore, multiple Hp/Py pair substitutions within this ten-ring polyamide do not appear to have an additive effect on binding affinity.

Hydroxypyrrole-imidazole-pyrrole polyamides complete the minor groove recognition code using the four ring pairings (Im/Py, Py/Im, Hp/Py, and Py/Hp) to complement the four Watson-Crick base pairs. The observation of substantially increased A•T/T•A specificity without severe energetic penalty by incorporation of multiple Hp/Py pairings is a minimal first step toward full integration of multiple Hp/Py ring pairings for Py/Py pairs. Because the sequence dependent microstructure of DNA is still not well understood, the scope and limitations of the 3-ring pairing rules with regards to energetics and specificity remains to be elucidated and will be reported in due course.

**X-ray Structure Analysis Reveals Two Hp H-Bonds to T.** Before the first structure of a molecule bound to DNA had been determined, specific recognition of double helical B-form DNA was predicted to occur primarily in the major, rather than the minor groove.<sup>11</sup> This proposal was based upon the observation that for A,T base pairs, the hydrogen bond acceptors at N3 of adenine and O2 of thymine are almost symmetrically placed in the minor groove (Figure 8.4).<sup>11</sup> Subsequent structures of DNA-binding domains co-crystallized with DNA supported this idea, since most of the specific contacts were made with the major groove.<sup>12</sup> The principle that ‘the major groove is a better candidate for sequence-specific recognition than the minor groove’<sup>27</sup> continues to provide the basis for strategies to decipher rules for protein-DNA recognition. Although there has been remarkable progress in the design of zinc fingers to recognize the major groove,<sup>23</sup> no protein structure motif has been identified which provides an  $\alpha$ -amino acid-base pair code for the minor groove.

The discovery of pairing rules for dimers of minor groove binding polyamides containing three aromatic amino acids, pyrrole (Py), imidazole (Im), and 3-hydroxypyrrole (Hp), affords a recognition code for the discrimination of all four Watson-Crick base pairs in the minor groove of DNA.<sup>24</sup> The side-by-side pairing of the aromatic residues in the polyamide dimer determines the DNA sequence recognized. An Im/Py pair distinguishes G•C



**Figure 8.13.** Chemical structures of the polyamides ImHpPyPy-β-Dp 17 and ImPyPyPy-β-Dp 18.

**Table 8.15. Equilibrium Association Constants ( $K_a$ )<sup>a,b</sup>.**

Match	Mismatch	Specificity <sup>c</sup>
5'-A G T A C T-3'  3'-T C A T G A-5' $2.1 (\pm 0.1) \times 10^7 \text{ M}^{-1}$	5'-A G T A T T-3'  3'-T C A T A A-5' $1.4 (\pm 0.3) \times 10^6 \text{ M}^{-1}$	15-fold
5'-A G T A C T-3'  3'-T C A T G A-5' $2.9 (\pm 0.7) \times 10^6 \text{ M}^{-1}$	5'-A G T A T T-3'  3'-T C A T A A-5' $7.6 (\pm 0.6) \times 10^5 \text{ M}^{-1}$	4-fold

<sup>a</sup>Values reported are the mean values obtained from three DNase I footprint titration experiments. The standard deviation for each value is indicated in parentheses.

<sup>b</sup>The assays were carried out at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

<sup>c</sup>Specificity is calculated as  $K_a(5'\text{-AGTACT-3}') / K_a(5'\text{-AGTATT-3}')$ .

from C•G and both of these from A•T/T•A,<sup>1</sup> and the structural basis of this discrimination has been determined through x-ray structure studies.<sup>13</sup> However, a structural understanding of how an Hp/Py pair distinguishes T•A from A•T, and both from G•C/C•G, has yet to be established. To address this question, the co-crystal structure of a polyamide of sequence ImHpPyPy-β-Dp **17** (Figure 8.13) bound as a dimer to a self-complementary ten base pair oligonucleotide containing all four Watson-Crick base pairs, 5'-CCAGTACTGG-3' (binding site in bold) has been determined.

The structure of the polyamide ImPyPyPy-β-Dp **18**, containing Py/Py pairs which do not distinguish T•A from A•T,<sup>16</sup> bound to the same duplex was solved for comparison. In addition, the equilibrium association constants ( $K_a$ ) for polyamides ImHpPyPy-β-Dp **17** and ImPyPyPy-β-Dp **18** (Figure 8.13) were determined using DNase I quantitative footprint titration experiments in order to correlate energetic analysis of minor groove recognition with three-dimensional structural information (Tables 8.15, 8.16).

**Table 8.16.** Equilibrium Association Constants ( $K_a$ )<sup>a,b</sup>.

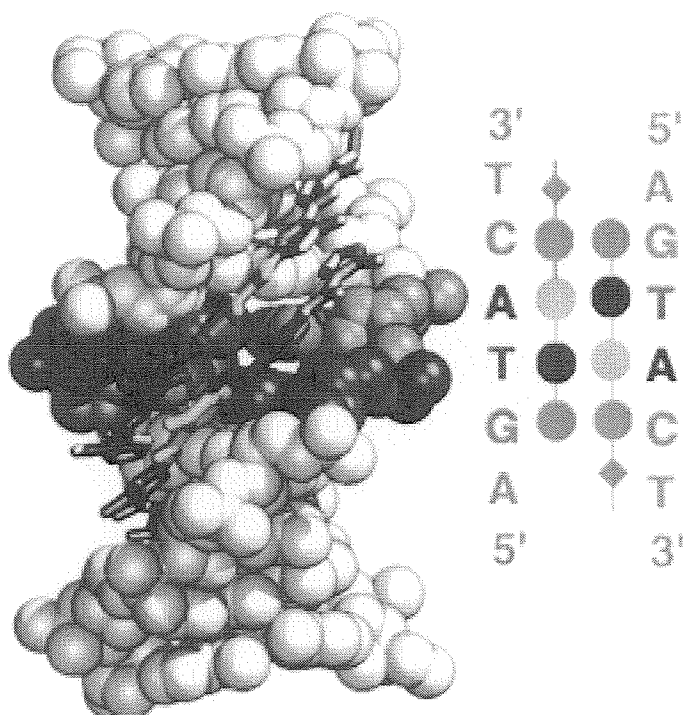
GTAC	GAAC	GATC	Specificity
 $2.5 \times 10^7 \text{ M}^{-1}$	 $4.0 \times 10^6 \text{ M}^{-1}$	 $4.0 \times 10^6 \text{ M}^{-1}$	6-fold
 $4.0 \times 10^6 \text{ M}^{-1}$	 $3.0 \times 10^5 \text{ M}^{-1}$	 $1.0 \times 10^5 \text{ M}^{-1}$	13-40-fold

<sup>a</sup>The assays were carried out at 22°C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>.

<sup>b</sup>Specificity is calculated as  $K_a$  (5'-AGTACT-3') /  $K_a$  (5'-AGAACT-3') and  $K_a$  (5'-AGTACT-3') /  $K_a$  (5'-AGATCT-3').

The polyamides ImHpPyPy-β-Dp **17** and ImPyPyPy-β-Dp **18** were synthesized by solid-phase methods using Boc-protected 3-methoxypyrrole, imidazole, and pyrrole amino acids. The identity and purity of the polyamides were confirmed by <sup>1</sup>H NMR and matrix-assisted laser-desorption ionization time-of-flight mass spectrometry. Synthetic deoxyoligonucleotides were synthesized with the 5'-trityl on, and purified with two rounds of reverse phase on a C8 column.

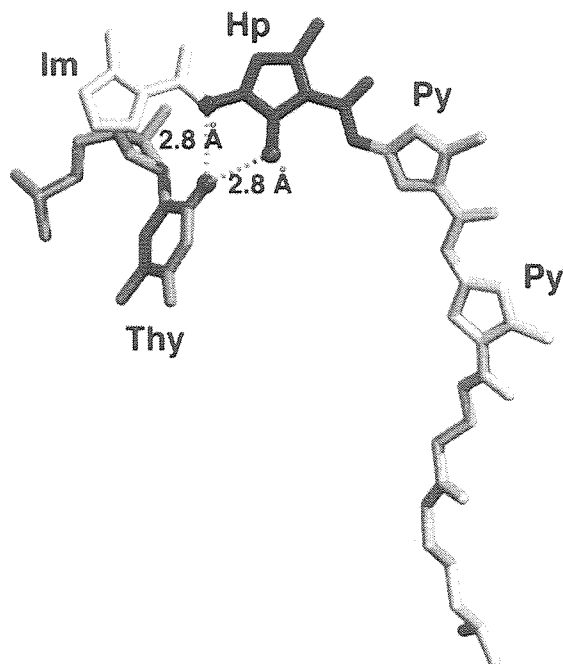
DNase I footprint titration experiments (10 mM Tris-HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub>, pH 7.0, 22°C) were performed to determine the equilibrium association constants  $K_a$  for recognition of bound sites. The hydroxypyrrole-containing polyamide, ImHpPyPy-β-Dp **17**, prefers the match site 5'-AGTACT-3' ( $K_a = 2.9 (\pm 0.7) \times 10^6 \text{ M}^{-1}$ ) over the single base pair mismatch site 5'-AGTATT-3' ( $K_a = 7.6 (\pm 0.6) \times 10^5 \text{ M}^{-1}$ ) by a factor of 4 (Table 8.15). Polyamide **17** has increased A•T/T•A specificity, preferring the match site 5'-AGTACT-3' ( $K_a = 2.9 (\pm 0.7) \times 10^6 \text{ M}^{-1}$ ) over the mismatch sites 5'-AGAACT-3' ( $K_a = 3.0 \times 10^5 \text{ M}^{-1}$ ) and 5'-AGATCT-3' ( $K_a = 1.0 \times 10^5 \text{ M}^{-1}$ ) by a factor of 10- and 29-fold respectively (Table 8.16). The control polyamide ImPyPyPy-β-Dp **18**, containing only Py/Py pairs which are approximately degenerate for A•T and T•A base pairs, binds to all three sites 5'-AGTACT-3', 5'-AGAACT-3', and 5'-AGATCT-3' within a factor



**Figure 8.14.** Structure of  $(\text{ImHpPyPy-}\beta\text{-Dp})_2 \cdot 5'\text{-CCAGTACTGG-3'}$ . Adenosine is very dark gray and thymidine is medium gray. A schematic DNA-binding model is shown to the side, with the residues of the polyamide indicated by filled circles. Im/Py pairs = medium gray, Hp = black, Py = light gray. The overall structure of  $(\text{ImPyPyPy-}\beta\text{-Dp})_2 \cdot 5'\text{-CCAGTACTGG-3'}$  is similar.

of six (Table 8.16). In addition,  $\text{ImPyPyPy-}\beta\text{-Dp}$  **18** prefers the site  $5'\text{-AGTACT-3'}$  ( $K_a = 2.1 (\pm 0.1) \times 10^7 \text{ M}^{-1}$ ) over the site  $5'\text{-AGTATT-3'}$  ( $K_a = 1.4 (\pm 0.3) \times 10^6 \text{ M}^{-1}$ ) by at least 15-fold (Table 8.15). Within these polyamides, the replacement of a Py/Py pair by a Hp/Py pair decreases the G•C specificity of an adjacent Im/Py pair. Also, there is a loss in binding affinity of about 7-fold with the substitution of two Py/Py pairs in polyamide **18** by two Hp/Py pairs to give polyamide **17**. The structural studies described herein provide insight into how hydroxypyrrole-imidazole-pyrrole polyamides discriminate T•A from A•T base pairs via hydrogen bond formation and differential destabilization.

In both the  $\text{ImHpPyPy}$  and  $\text{ImPyPyPy}$  structures, the polyamides bind as antiparallel dimers centered over the target AGTACT sequence in the minor groove of a B-form DNA duplex (Figure 8.14). The N to C-terminal orientation of each fully overlapped polyamide is



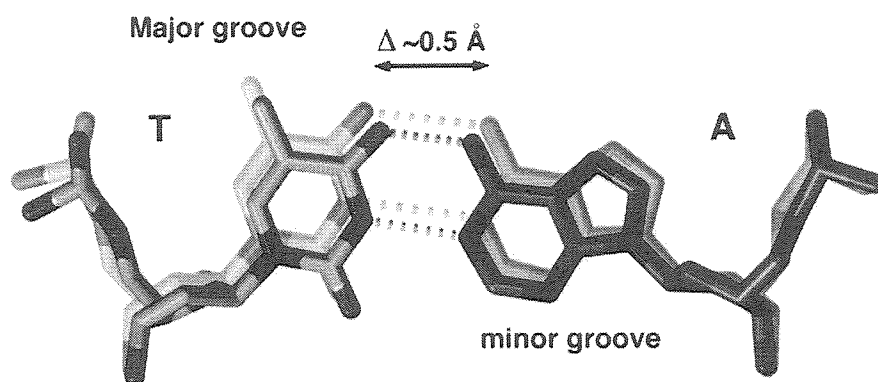
**Figure 8.15.** Diagram of the hydrogen bonds between the hydroxyl and the preceding amide and the two lone pairs of thymine-O2.

parallel to the adjacent 5' to 3' strand of DNA, consistent with previous chemical<sup>1</sup> and structural studies of polyamide dimers.<sup>13</sup>

Although the functional groups of adenine and thymine are very similar in the minor groove, the number of lone pairs on the hydrogen bond acceptors is different: a thymine-O2 has two free lone pairs, whereas an adenine-N3 has only one (Figure 8.4). The amide nitrogens of the ligand form hydrogen bonds with the purine-N3 (A or G) or pyrimidine-O2 (T or C). Therefore, the hydrogen bond potential of adenine-N3 is filled when an imidazole-pyrrole polyamide is bound, but the thymine-O2 has the capacity to accept an additional hydrogen bond. We find that both the hydroxyl group of the Hp, and the amide-NH of the preceding residue, form hydrogen bonds with the target thymine-O2 of the adjacent DNA strand (Figure 8.15). A similar interaction between the Hp and the adenine-N3 would be impossible without the loss of the hydrogen bond from the preceding amide-NH.

In addition to the difference in number of lone pairs of the adenine-N3 versus thymine-O2, adenine is also distinguished from thymine by a bulkier aromatic ring. Although the

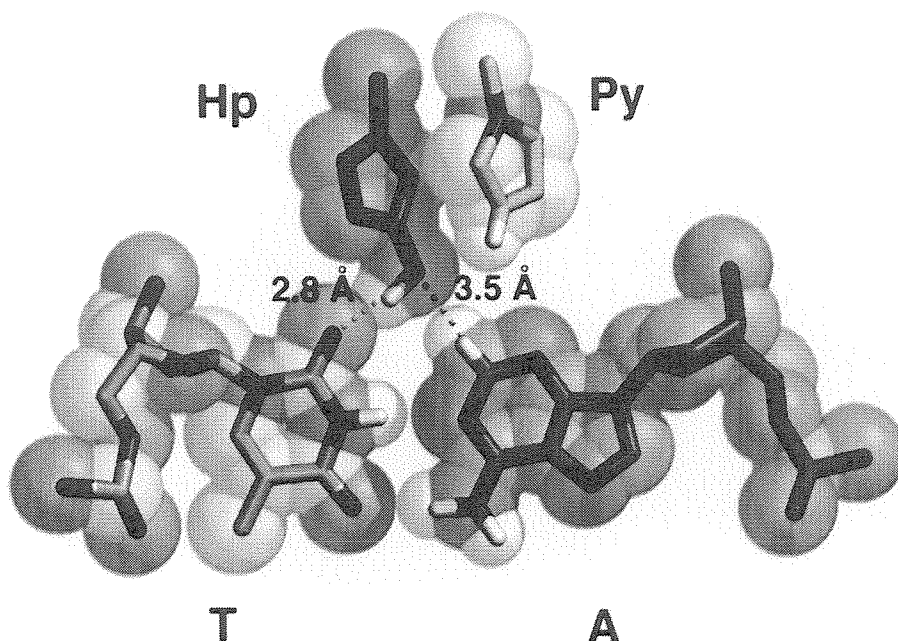




**Figure 8.16.** T•A base pair of the ImPyPyPy structure superimposed on the corresponding base pair from the ImHpPyPy structure, showing the slight elongation of the hydrogen bonds of the target T•A base pair.

adenine-C2-H does not protrude into the minor groove like the guanine exocyclic amine, the additional carbon results in an asymmetric cleft in the minor groove of a T•A base pair (Figure 8.4).<sup>15</sup> The adenine-C2 of the ImHpPyPy structure contacts the Hp hydroxyl. Modeling the target thymine as an adenine reveals that the C2 carbon of a mismatch ‘adenine’ opposite an Hp residue would sterically overlap the hydroxyl by 1-2 Å (depending on the hydrogen positions). Shape selective recognition of the asymmetric cleft is the second feature that allows the Hp/Py pair to discriminate T•A from A•T.

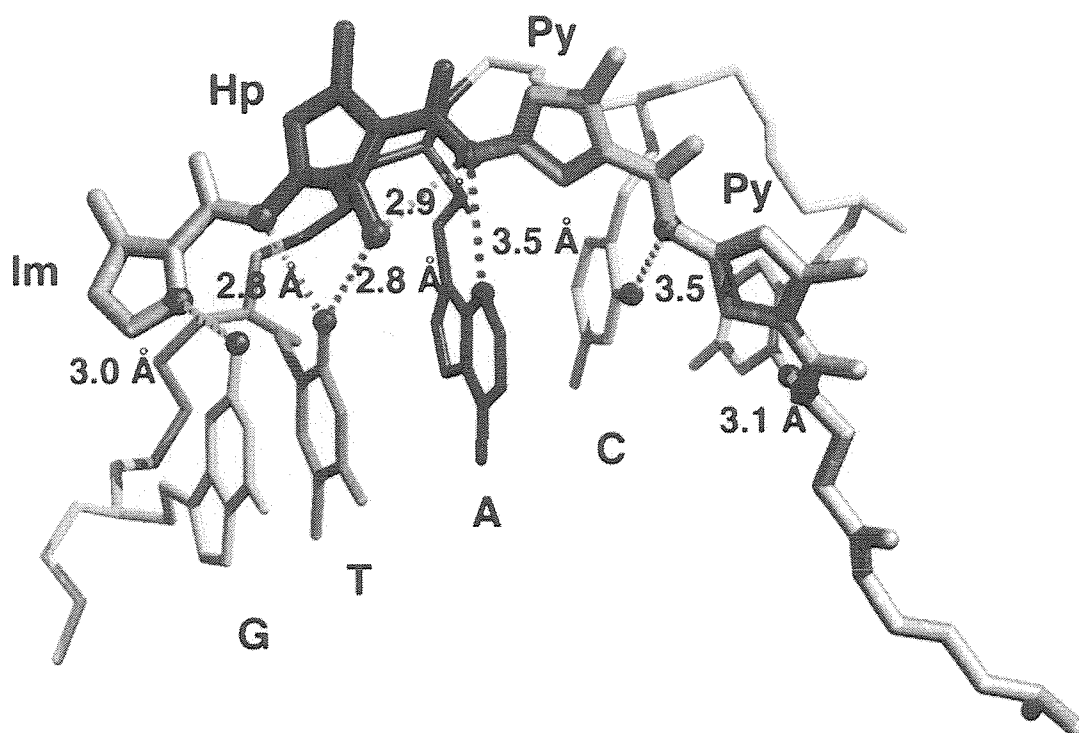
In both structures, the oligonucleotides have standard B-DNA features, 35° twist, 3.4 Å rise per residue, and C2'-endo sugar pucker. However, the oligonucleotide structure deviates from ideal B-form by having a strong propeller twist and opening of the target T•A base pairs. The Hp/Py pairs induce a change in the T•A base pairs from no shear (displacement between the bases in the base pair, perpendicular to the helix axis) to a large positive shear. The movement of the bases past one another may result from the Hp-O6 contact with the adenine-C2 pressing the adenine of the target base pair back into the major groove. The increased displacement between the bases stretches the Watson-Crick hydrogen bonds between them by 0.5 Å, on average (Figure 8.16). Although the specificity of hydroxypyrrole-containing polyamides is greatly increased for T•A over A•T, the affinities



**Figure 8.17.** Interaction of Hp/Py pair with T•A base pair.

are slightly reduced relative to the pyrrole counterparts (Table 8.15). For example, ImHpPyPy- $\beta$ -Dp and ImPyPyPy- $\beta$ -Dp bind a 5'-AGTACT-3' site with equilibrium association constants of  $2.9 (\pm 0.7) \times 10^6 \text{ M}^{-1}$  and  $2.1 (\pm 0.1) \times 10^7 \text{ M}^{-1}$  respectively, a 7-fold difference. The energetic penalty due to the partial 'melting' of the target T•A base pairs could account for the 1.2 kcal/mol reduction in binding affinity.<sup>28</sup>

The hydrogen bonds between the amides of each ImPyPyPy polyamide and the purine-N3 or pyrimidine-O2 of the adjacent DNA strand are maintained for the ImHpPyPy polyamide. However, the hydrogen bonds between the DNA and the ImHpPyPy amides are longer for the residues that follow the Hp than those observed for the ImPyPyPy complex. The hydroxyl forms an intramolecular hydrogen bond with the following amide, causing the hydrogen bond of that amide with the adenine-N3 to become bifurcated and therefore weaker. This may be an additional source of the slightly decreased affinity of the Hp-containing polyamides relative to the pyrrole counterparts.



**Figure 8.18.** Crystal structure reveals that all potential aromatic amino acid – base pair H-bonds are formed.

These structural studies have established how a designed ligand can predictably discriminate A•T from T•A in the minor groove, utilizing the double hydrogen bond acceptor potential of thymine-O2 and the asymmetry of the adenine-C2 cleft. The structure eliminates the possibilities that a bulky substitution at the pyrrole 3-position might 1) sterically clash with the thymine-O2,<sup>29</sup> or 2) cause a gross distortion of the DNA duplex.<sup>30</sup>

## Experimental Section

All gel electrophoresis, footprinting assays, and polyamide deprotection reactions were performed by Sarah White.<sup>16,24,26</sup> Monomer synthesis was by Jason Szewczyk and Jim Turner.<sup>24,26</sup> Polyamide **14** was prepared by Jim Turner. All crystal growing and structure analysis was performed by Clara Kielkopf.<sup>26</sup>

Dicyclohexylcarbodiimide (DCC), Hydroxybenzo-triazole (HOBt), 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate (HBTU) and 0.2 mmol/gram Boc- $\beta$ -alanine-(-4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc- $\beta$ -Pam-Resin) were purchased from Peptides International. *N,N*-diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF), *N*-methylpyrrolidone (NMP), DMSO/NMP, Acetic anhydride ( $\text{Ac}_2\text{O}$ ), and 0.0002 M potassium cyanide/pyridine were purchased from Applied Biosystems. Boc- $\gamma$ -aminobutyric acid was from NOVA Biochem, dichloromethane (DCM) and triethylamine (TEA) was reagent grade from EM, thiophenol (PhSH), dimethylaminopropylamine from Aldrich, trifluoroacetic acid (TFA) from Halocarbon, phenol from Fisher, and ninhydrin from Pierce. All reagents were used without further purification.

Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. and were used for filtration of DCU. A shaker for manual solid phase synthesis was obtained from St. John Associates, Inc. Screw-cap glass peptide synthesis reaction vessels (5 mL and 20 mL) with a #2 sintered glass frit were made as described by Kent.<sup>31</sup>  $^1\text{H}$  NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz in  $\text{DMSO-}d_6$ , with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology.

HPLC analysis was performed on either a HP 1090M analytical HPLC or a Beckman Gold system using a RAINEN C<sub>18</sub>, Microsorb MV, 5µm, 300 x 4.6 mm reversed phase column in 0.1% (wt/v) TFA with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% acetonitrile/min. Preparatory reverse 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. 18MΩ water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 µm filtered.

**Resin Substitution.** Resin substitution can be calculated as  $L_{\text{new}}(\text{mmol/g}) = L_{\text{old}} / (1 + L_{\text{old}}(W_{\text{new}} - W_{\text{old}}) \times 10^{-3})$ , where L is the loading (mmol of amine per gram of resin), and W is the weight (gmol<sup>-1</sup>) of the growing polyamide attached to the resin.<sup>32</sup>

**Ethyl 4-[benzyloxycarbonyl]amino]-3-hydroxy-1-methylpyrrole-2-carboxylate 10** Ethyl-4-carboxy-3-hydroxy-1-methylpyrrole-2-carboxylate **9** (60 g, 281.7 mmol) was dissolved in 282 mL acetonitrile. TEA (28.53 g, 282 mmol) was added, followed by diphenylphosphorylazide (77.61 g, 282 mmol). The mixture was refluxed for 5 hours, followed by addition of benzyl alcohol (270 mL) and reflux continued overnight. The solution was cooled and volatiles removed *in vacuo*. The residue was absorbed onto silica and chromatographed, 4:1 hexanes:ethyl acetate, to give a white solid (21.58 g, 24%). <sup>1</sup>H NMR (DMSO-d<sub>6</sub>) δ 8.73 (s, 1H), 8.31 (s, 1H), 7.31 (m, 5H), 6.96 (s, 1H), 5.08 (s, 2H), 4.21 (q, 2H, J = 7.1 Hz), 3.66 (s, 3H), 1.25 (t, 3H, J = 7.1 Hz); MS *m/e* 319.163 (M+H 319.122 calcd. for C<sub>16</sub>H<sub>18</sub>N<sub>2</sub>O<sub>5</sub>).

**Ethyl 4-[(*tert*-butoxycarbonyl)amino]-3-methoxy-1-methylpyrrole-2-carboxylate 12** Ethyl 4-[(benzyloxycarbonyl)amino]-3-hydroxy-1-methylpyrrole-2-carboxylate **10** (13.4 g, 42.3 mmol) was dissolved in 110 mL acetone. Anhydrous K<sub>2</sub>CO<sub>3</sub> (11.67 g, 84.5 mmol) was added, followed by methyl iodide (5.96 g, 42.3 mmol) and dimethylaminopyridine (0.5 g, 4.23 mmol) and the mixture stirred overnight. The solid K<sub>2</sub>CO<sub>3</sub> was removed by filtration and 200 mL water added. Volatiles were removed *in vacuo* and the solution made acidic with addition of 1N H<sub>2</sub>SO<sub>4</sub>. The aqueous layer was extracted

with diethyl ether. Organic layers were combined, washed with 10%  $\text{H}_2\text{SO}_4$ , dried over  $\text{MgSO}_4$ , and dried to give a white solid **11**. The solid was used without further purification and dissolved in 38 mL DMF. DIEA (11 mL), Boc anhydride (9.23 g, 42.3 mmol), and 10% Pd/C (500 mg) were added and the solution stirred under hydrogen (1 atm) for 2.1 h. The slurry was filtered through celite which was washed with methanol. Water (250 mL) was added and volatiles removed *in vacuo*. The aqueous layer was extracted with ether. Organic layers were combined, washed with water and brine, and dried over  $\text{MgSO}_4$ . Solvent was removed *in vacuo* to give a white solid **12** (8.94 g, 71%).  $^1\text{H}$  NMR ( $\text{DMSO-d}_6$ )  $\delta$  8.43 (s, 1H), 7.03 (s, 1H), 4.19 (q, 2H,  $J = 7.1$  Hz), 3.70 (s, 3H), 3.67 (s, 3H), 1.42 (s, 9H), 1.26 (t, 3H,  $J = 7.1$  Hz); MS  $m/e$  299.161 (M+H 299.153 calcd. for  $\text{C}_{14}\text{H}_{22}\text{N}_2\text{O}_5$ ).

**Ethyl 4-[(*tert*-butoxycarbonyl)amino]-3-hydroxy-1-methylpyrrole-2-carboxylate **13**** Ethyl 4-[(*tert*-butoxycarbonyl)amino]-3-methoxy-1-methylpyrrole-2-carboxylate **12** (9.0 g, 30.2 mmol) was dissolved in 30 mL ethanol. NaOH (30 mL, 1 M, aq.) was added and the solution stirred for 4 days. Water (200 mL) was added and ethanol removed *in vacuo*. The solution was extracted with diethyl ether, aqueous layer acidified to pH = 2-3, and extracted again with diethyl ether. Organic layers were dried over  $\text{MgSO}_4$ , and solvent removed *in vacuo* to give a white solid **13** (6.0 g, 20.5 mmol, 87% based on recovered SM).  $^1\text{H}$  NMR ( $\text{DMSO-d}_6$ )  $\delta$  12.14 (s, 1H), 8.37 (s, 1H), 6.98 (s, 1H), 3.69 (s, 3H), 3.66 (s, 3H), 1.42 (s, 9H); MS  $m/e$  293.112 (M+H 293.104 calcd. for  $\text{C}_{12}\text{H}_{18}\text{N}_2\text{O}_5$ ).

**ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp** ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc- $\beta$ -Pam-Resin (0.66 mmol/g).<sup>25</sup> 3-hydroxypyrrole-Boc amino acid was incorporated by placing the amino acid (0.5 mmol) and HBTU (0.5 mmol) in a machine synthesis cartridge. Upon automated delivery of DMF (2 mL) and DIEA (1 mL), activation occurs. A sample of ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Pam-Resin (400 mg, 0.40 mmol/gram) was placed in a glass 20 mL peptide synthesis vessel and treated with neat dimethylaminopropylamine (2 mL) and heated (55°C) with periodic agitation for 16 h. The reaction mixture was then filtered to

remove resin, 0.1% (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC. ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp is recovered upon lyophilization of the appropriate fractions as a white powder (97 mg, 49% recovery). UV (H<sub>2</sub>O)  $\lambda_{\text{max}}$  246, 316 (66,000); <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.24 (s, 1 H), 10.14 (s, 1 H), 9.99 (s, 1 H), 9.94 (s, 1 H), 9.88 (s, 1 H), 9.4 (br s, 1 H), 9.25 (s, 1 H), 9.11 (s, 1 H), 8.05 (m, 3 H), 7.60 (s, 1 H), 7.46 (s, 1 H), 7.41 (s, 1 H), 7.23 (d, 1 H), 7.21 (d, 1 H), 7.19 (d, 1 H), 7.13 (m, 2 H), 7.11 (m, 2 H), 7.02 (d, 1 H), 6.83 (m, 2 H), 3.96 (s, 6 H), 3.90 (s, 3 H), 3.81 (m, 6 H), 3.79 (s, 3 H), 3.75 (d, 9 H), 3.33 (q, 2 H, *J* = 5.4 Hz), 3.15 (q, 2 H, *J* = 5.5 Hz), 3.08 (q, 2 H, *J* = 6.0 Hz), 2.96 (quintet, 2 H, *J* = 5.6 Hz), 2.70 (d, 6 H, *J* = 4.5 Hz), 2.32 (m, 4 H), 1.71 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1253.5 (1253.6 calcd. for C<sub>58</sub>H<sub>72</sub>N<sub>22</sub>O<sub>11</sub>).

**ImImHpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp 3** In order to remove the methoxy protecting group, a sample of ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp (5 mg, 3.9  $\mu$ mol) was treated with sodium thiophenoxide at 100°C for 2 h. DMF (1000  $\mu$ L) and thiophenol (500  $\mu$ L) were placed in a (13 x 100 mm) disposable Pyrex screw cap culture tube. A 60% dispersion of sodium hydride in mineral oil (100 mg) was slowly added. Upon completion of the addition of the sodium hydride, ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp (5 mg) dissolved in DMF (500 mL) was added. The solution was agitated, and placed in a 100°C heat block, and deprotected for 2 h. Upon completion of the reaction, the culture tube was cooled to 0°C, and 7 mL of a 20% (wt/v) solution of trifluoroacetic acid added. The aqueous layer is separated from the resulting biphasic solution and purified by reversed phase HPLC. ImImHpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp is recovered as a white powder upon lyophilization of the appropriate fractions (3.8 mg, 77% recovery). UV (H<sub>2</sub>O)  $\lambda_{\text{max}}$  246, 312 (66,000); <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>)  $\delta$  10.34 (s, 1 H), 10.24 (s, 1 H), 10.00 (s, 2 H), 9.93 (s, 1 H), 9.87 (s, 1 H), 9.83 (s, 1 H), 9.4 (br s, 1 H), 9.04 (s, 1 H), 8.03 (m, 3 H), 7.58 (s, 1 H), 7.44 (s, 1 H), 7.42 (s, 1 H), 7.23 (s, 1 H), 7.20 (m, 3 H), 7.12 (m, 2 H), 7.05 (d, 1 H), 7.02 (d, 1 H), 6.83 (s, 1 H), 6.79 (s, 1 H), 3.96 (s, 6 H), 3.90 (s, 3 H), 3.81 (s, 6 H), 3.79 (s, 3 H), 3.75 (d, 6 H), 3.33 (q, 2 H, *J* = 5.4 Hz), 3.14 (q, 2 H, *J* = 5.4 Hz), 3.08 (q, 2 H, *J* = 6.1 Hz), 2.99 (quintet, 2 H, *J* = 5.4 Hz), 2.69 (d, 6 H, *J* = 4.2 Hz),

2.31 (m, 4 H), 1.72 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1239.6 (1239.6 calcd. for  $C_{57}H_{71}N_{22}O_{11}$ ).

**ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Dp** ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc- $\beta$ -Pam-Resin (0.66 mmol/g) as described for ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp.<sup>25</sup> A sample of ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Pam-Resin (400 mg, 0.40 mmol/gram) was placed in a glass 20 mL peptide synthesis vessel and treated with neat dimethylaminopropylamine (2 mL) and heated (55°C) with periodic agitation for 16 h. The reaction mixture was then filtered to remove resin, 0.1% (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC. ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Dp is recovered upon lyophilization of the appropriate fractions as a white powder (101 mg, 50% recovery).

**ImImPyPy- $\gamma$ -ImHpPyPy- $\beta$ -Dp 2** A sample of ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Dp (5 mg, 3.9  $\mu$ mol) was treated with sodium thiophenoxide and purified by reversed phase HPLC as described for ImImHpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp. ImImPyPy- $\beta$ -ImHpPyPy- $\beta$ -Dp is recovered upon lyophilization of the appropriate fractions as a white powder (3.2 mg, 66% recovery). UV ( $H_2O$ )  $\lambda_{max}$  246, 312 (66,000); MALDI-TOF-MS (monoisotopic), 1239.6 (1239.6 calcd. for  $C_{57}H_{71}N_{22}O_{11}$ ).

**ImPyPy- $\gamma$ -OpPyPy- $\beta$ -Dp** ImPyPy- $\gamma$ -OpPyPy- $\beta$ -Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc- $\beta$ -Pam-Resin (0.66 mmol/g).<sup>25</sup> 3-hydroxypyrrole-Boc amino acid was incorporated by placing the amino acid (0.5 mmol) and HBTU (0.5 mmol) in a machine synthesis cartridge. Upon automated delivery of DMF (2 mL) and DIEA (1 mL), activation occurs. A sample of ImPyPy- $\gamma$ -OpPyPy- $\beta$ -Pam-Resin (400 mg, 0.45 mmol/gram) was placed in a glass 20 mL peptide synthesis vessel and treated with neat dimethylaminopropylamine (2 mL) and heated (55°C) with periodic agitation for 16 h. The reaction mixture was then filtered to remove resin, 0.1% (wt/v) TFA added (6 mL) and the resulting solution purified by reversed phase HPLC. ImPyPy- $\gamma$ -OpPyPy- $\beta$ -Dp is



recovered upon lyophilization of the appropriate fractions as a white powder (45 mg, 25% recovery). UV (H<sub>2</sub>O)  $\lambda_{\text{max}}$  246, 310 (50,000); <sup>1</sup>H NMR (DMSO-d<sub>6</sub>)  $\delta$  10.45 (s, 1 H), 9.90 (s, 1 H), 9.82 (s, 1 H), 9.5 (br s, 1 H), 9.38 (s, 1 H), 9.04 (s, 1 H), 8.02 (m, 3 H), 7.37 (s, 1 H), 7.25 (m, 2 H), 7.15 (d, 1 H, *J* = 1.6 Hz), 7.11 (m, 2 H), 7.09 (d, 1 H), 7.03 (d, 1 H), 6.99 (d, 1 H), 6.87 (d, 1 H), 6.84 (d, 1 H), 3.96 (s, 3 H), 3.81 (s, 6 H), 3.77 (s, 6 H), 3.76 (s, 3 H), 3.74 (s, 1 H), 3.34 (q, 2 H, *J* = 5.6 Hz), 3.20 (q, 2 H, *J* = 5.8 Hz), 3.09 (q, 2 H, *J* = 6.1 Hz), 2.97 (quintet, 2 H, *J* = 5.3 Hz), 2.70 (d, 6 H, *J* = 3.9 Hz), 2.34 (m, 4 H), 1.73 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1007.6 (1007.5 calcd. for C<sub>48</sub>H<sub>63</sub>N<sub>16</sub>O<sub>9</sub>).

**ImPyPy- $\gamma$ -HpPyPy- $\beta$ -Dp 8** In order to remove the methoxy protecting group, a sample of ImPyPy- $\gamma$ -OpPyPy- $\beta$ -Dp (5 mg, 4.8 mmol) was treated with sodium thiophenoxide 100°C for 2 h and reverse phase HPLC purified using the same protocol as for ImImHpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp. ImPyPy- $\gamma$ -HpPyPy- $\beta$ -Dp is recovered upon lyophilization of the appropriate fractions as a white powder (2.5 mg, 52% recovery). UV (H<sub>2</sub>O)  $\lambda_{\text{max}}$  246, 310 (50,000); <sup>1</sup>H NMR (DMSO-d<sub>6</sub>)  $\delta$  10.44 (s, 1 H), 10.16 (s, 1 H), 9.90 (s, 1 H), 9.77 (s, 1 H), 9.5 (br s, 1 H), 9.00 (s, 1 H), 8.03 (m, 3 H), 7.37 (s, 1 H), 7.26 (m, 2 H), 7.14 (d, 1 H, *J* = 1.7 Hz), 7.12 (m, 2 H), 7.02 (d, 1 H), 6.93 (d, 1 H), 6.88 (d, 1 H), 6.82 (d, 1 H), 6.72 (d, 1 H), 3.96 (s, 3 H), 3.81 (s, 6 H), 3.77 (s, 3 H), 3.76 (s, 3 H), 3.74 (s, 1 H), 3.36 (q, 2 H, *J* = 5.6 Hz), 3.22 (q, 2 H, *J* = 5.9 Hz), 3.09 (q, 2 H, *J* = 5.5 Hz), 2.98 (quintet, 2 H, *J* = 5.3 Hz), 2.70 (d, 6 H, *J* = 4.3 Hz), 2.34 (m, 4 H), 1.78 (m, 4 H); MALDI-TOF-MS (monoisotopic), 994.2 (993.5 calcd. for C<sub>47</sub>H<sub>61</sub>N<sub>16</sub>O<sub>9</sub>).

**ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp-NH<sub>2</sub>** ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc- $\beta$ -Pam-Resin (0.66 mmol/g) as described previously for hydroxypyrrole-imidazole-pyrrole polyamides.<sup>25</sup> A sample of ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Pam-Resin (400 mg, 0.40 mmol/g) was placed in a glass 20 mL peptide synthesis vessel and treated with neat 3,3'-diamino-*N*-methyldipropylamine (2 mL) and heated (55°C) with periodic agitation for 16 h. The reaction mixture was then filtered to remove resin, 0.1% (wt/v) TFA added (6 mL) and the

resulting solution purified by reversed phase HPLC. ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp-NH<sub>2</sub> is recovered upon lyophilization of the appropriate fractions as a white powder (93 mg, 46% recovery). UV (H<sub>2</sub>O)  $\lambda_{\text{max}}$  246, 316 (66,000); <sup>1</sup>H NMR (DMSO-d<sub>6</sub>)  $\delta$  10.34 (s, 1 H), 10.30 (br s, 1 H), 10.25 (s, 1 H), 9.96 (s, 1 H), 9.95 (s, 1 H), 9.89 (s, 1 H), 9.24 (s, 1 H), 9.11 (s, 1 H), 8.08 (t, 1 H, *J* = 5.6 Hz), 8.0 (m, 5 H), 7.62 (s, 1 H), 7.53 (s, 1 H), 7.42 (s, 1 H), 7.23 (d, 1 H, *J* = 1.2 Hz), 7.21 (m, 2 H), 7.15 (m, 2 H), 7.13 (d, 1 H), 7.11 (m, 2 H), 7.04 (d, 1 H), 6.84 (m, 3 H), 3.98 (s, 3 H), 3.97 (s, 3 H), 3.92 (s, 3 H), 3.82 (m, 6 H), 3.80 (s, 3 H), 3.77 (d, 6 H), 3.35 (q, 2 H, *J* = 5.8 Hz), 3.0-3.3 (m, 8 H), 2.86 (q, 2 H, *J* = 5.4 Hz), 2.66 (d, 3 H, *J* = 4.5 Hz), 2.31 (m, 4 H), 1.94 (quintet, 2 H, *J* = 6.2 Hz), 1.74 (m, 4 H); MALDI-TOF-MS (monoisotopic), 1296.0 (1296.6 calcd. for C<sub>60</sub>H<sub>78</sub>N<sub>23</sub>O<sub>11</sub>).

**ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp-EDTA** Excess EDTA-dianhydride (50 mg) was dissolved in DMSO/NMP (1 mL) and DIEA (1 mL) by heating at 55°C for 5 min. The dianhydride solution was added to ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp-NH<sub>2</sub> (13 mg, 10  $\mu$ mol) dissolved in DMSO (750  $\mu$ L). The mixture was heated (55°C, 25 min.) and the remaining EDTA-dianhydride hydrolyzed (0.1 M NaOH, 3 mL, 55°C, 10 min.). Aqueous TFA (0.1%, wt/v) was added to adjust the total volume to 8 mL and the solution purified directly by reversed phase HPLC to provide ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp-EDTA as a white powder upon lyophilization of the appropriate fractions (5.5 mg, 40% recovery). MALDI-TOF-MS (monoisotopic), 1570.9 (1570.7 calcd. for C<sub>70</sub>H<sub>92</sub>N<sub>25</sub>O<sub>18</sub>).

**ImImHpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp-EDTA 3-E** In order to remove the methoxy protecting group, a sample of ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp-EDTA (5 mg, 3.1  $\mu$ mol) was treated with sodium thiophenoxide at 100°C for 2 h and purified as before to give as a white powder ImImHpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp-EDTA (3.2 mg, 72% recovery). UV (H<sub>2</sub>O)  $\lambda_{\text{max}}$  246, 312 (66,000); MALDI-TOF-MS (monoisotopic), 1556.6 (1556.7 calcd. for C<sub>69</sub>H<sub>90</sub>N<sub>25</sub>O<sub>18</sub>).

**ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Dp-NH<sub>2</sub>** ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Pam-Resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc- $\beta$ -Pam-Resin (0.66 mmol/g),<sup>25</sup> and the polyamide was cleaved from the resin using neat 3,3'-

diamino-*N*-methyldipropylamine (2 mL) as before. After purification by reversed phase HPLC, ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Dp-NH<sub>2</sub> is recovered upon lyophilization of the appropriate fractions as a white powder (104 mg, 54% recovery). UV (H<sub>2</sub>O)  $\lambda_{\text{max}}$  246, 316 (66,000); MALDI-TOF-MS (monoisotopic), 1296.6 (1296.6 calcd. for C<sub>60</sub>H<sub>78</sub>N<sub>23</sub>O<sub>11</sub>).

**ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Dp-EDTA** Excess EDTA-dianhydride (50 mg) was dissolved in DMSO/NMP (1 mL) and DIEA (1 mL) by heating at 55°C for 5 min. The dianhydride solution was added to ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Dp-NH<sub>2</sub> (13 mg, 10  $\mu$ mol) dissolved in DMSO (750  $\mu$ L). The mixture was heated (55°C, 25 min.) and the remaining EDTA-dianhydride hydrolyzed (0.1 M NaOH, 3 mL, 55°C, 10 min.). Aqueous TFA (0.1%, wt/v) was added to adjust the total volume to 8 mL and the solution purified directly by reversed phase HPLC to provide ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Dp-EDTA as a white powder upon lyophilization of the appropriate fractions (5.9 mg, 42% recovery). MALDI-TOF-MS (monoisotopic), 1570.8 (1570.7 calcd. for C<sub>70</sub>H<sub>92</sub>N<sub>25</sub>O<sub>18</sub>).

**ImImPyPy- $\gamma$ -ImHpPyPy- $\beta$ -Dp-EDTA 2-E** In order to remove the methoxy protecting group, a sample of ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Dp-EDTA (5 mg, 3.1  $\mu$ mol) was treated with sodium thiophenoxide at 100°C for 2 h and purified as before to give as a white powder ImImHpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp-EDTA (3.2 mg, 72% recovery). UV (H<sub>2</sub>O)  $\lambda_{\text{max}}$  246, 312 (66,000); MALDI-TOF-MS (monoisotopic), 1555.9 (1556.7 calcd. for C<sub>69</sub>H<sub>90</sub>N<sub>25</sub>O<sub>18</sub>).

**Table 8.17.** Mass spectral characterization of Op and Hp polyamides, synthesized and purified as described for ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp and ImImHpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp.

POLYAMIDE	FORMULA	(M+H)CALCD	FOUND
ImOpPy- $\gamma$ -PyPyPy- $\beta$ -Dp	C <sub>48</sub> H <sub>63</sub> N <sub>16</sub> O <sub>9</sub>	1007.5	1007.5
ImHpPy- $\gamma$ -PyPyPy- $\beta$ -Dp	C <sub>47</sub> H <sub>61</sub> N <sub>16</sub> O <sub>9</sub>	993.5	993.2
ImPyOp- $\gamma$ -PyPyPy- $\beta$ -Dp	C <sub>48</sub> H <sub>63</sub> N <sub>16</sub> O <sub>9</sub>	1007.5	1007.5
ImPyHp- $\gamma$ -PyPyPy- $\beta$ -Dp	C <sub>47</sub> H <sub>61</sub> N <sub>16</sub> O <sub>9</sub>	993.5	993.4
ImPyPy- $\gamma$ -OpPyPy- $\beta$ -Dp	C <sub>48</sub> H <sub>63</sub> N <sub>16</sub> O <sub>9</sub>	1007.5	1007.6
ImPyPy- $\gamma$ -HpPyPy- $\beta$ -Dp	C <sub>47</sub> H <sub>61</sub> N <sub>16</sub> O <sub>9</sub>	993.5	993.2
ImPyPy- $\gamma$ -PyOpPy- $\beta$ -Dp	C <sub>48</sub> H <sub>63</sub> N <sub>16</sub> O <sub>9</sub>	1007.5	1007.5
ImPyPy- $\gamma$ -PyHpPy- $\beta$ -Dp	C <sub>47</sub> H <sub>61</sub> N <sub>16</sub> O <sub>9</sub>	993.5	993.4
ImImOpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp	C <sub>58</sub> H <sub>72</sub> N <sub>22</sub> O <sub>11</sub>	1253.6	1253.5
ImImHpPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp	C <sub>57</sub> H <sub>71</sub> N <sub>22</sub> O <sub>11</sub>	1239.6	1239.6
ImImPyPy- $\gamma$ -ImOpPyPy- $\beta$ -Dp	C <sub>58</sub> H <sub>72</sub> N <sub>22</sub> O <sub>11</sub>	1253.6	1253.6
ImImPyPy- $\gamma$ -ImHpPyPy- $\beta$ -Dp	C <sub>57</sub> H <sub>71</sub> N <sub>22</sub> O <sub>11</sub>	1239.6	1239.6
ImImOpPy- $\gamma$ -ImOpPyPy- $\beta$ -Dp	C <sub>59</sub> H <sub>75</sub> N <sub>22</sub> O <sub>12</sub>	1283.6	1283.6
ImImHpPy- $\gamma$ -ImHpPyPy- $\beta$ -Dp	C <sub>57</sub> H <sub>71</sub> N <sub>22</sub> O <sub>12</sub>	1255.6	1255.5
ImOpPyPyPy- $\gamma$ -ImOpPyPyPy- $\beta$ -Dp	C <sub>72</sub> H <sub>88</sub> N <sub>25</sub> O <sub>14</sub>	1526.7	1526.6
ImHpPyPyPy- $\gamma$ -ImHpPyPyPy- $\beta$ -Dp	C <sub>70</sub> H <sub>84</sub> N <sub>25</sub> O <sub>14</sub>	1498.7	1498.0
ImHpHpPyPy- $\gamma$ -ImHpPyPyPy- $\beta$ -Dp	C <sub>71</sub> H <sub>87</sub> N <sub>26</sub> O <sub>14</sub>	1514.7	1514.7

## References

1. (a) Wade, W.S.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (b) Wade, W.S.; Mrksich, M.; Dervan, P.B. *Biochemistry* **1993**, *32*, 11385. (c) Mrksich, M.; Wade, W.S.; Dwyer, T.J.; Geierstanger, B.H.; Wemmer, D.H.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586.
2. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (b) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 3325. (c) Geierstanger, B.H.; Dwyer, T.J.; Bathini, Y.; Lown, J.W.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 4474. (d) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646. (e) Geierstanger, B.H.; Jacobsen, J.P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Biochemistry* **1994**, *33*, 3055.
3. (a) Pelton, J.G.; Wemmer, D.E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723. (b) Pelton, J.G.; Wemmer, D.E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S.T.; Sundaralingham, M. *Nature Struct. Biol.* **1994**, *1*, 169.
4. (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983. (b) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147. (c) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6153. (d) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Chem. & Biol.* **1996**, *3*, 369. (e) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198. (f) de Claire, R. P. L.; Geierstanger B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909. (g) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756. (h) White, S.; Baird, E. E.; Dervan, P. B. *Chem & Biol.* **1997**, *4*, 569.
5. (a) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559. (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953. (c) Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 7636. (d) Turner, J. M. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219.
6. (a) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160. (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1997**, *3*, 1600. (c) Trauger, J.W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534.
7. De Clairac, R.P.L.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1997**, *119*, 7906.
8. (a) Schultz, P.G.; Taylor, J.S.; Dervan, P.B. *J. Am. Chem. Soc.* **1982**, *104*, 6861. (b) Schultz, P.G.; Dervan, P.B. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1133. (c) Taylor, J.S.; Schultz, P.B.; Dervan, P.B. *Tetrahedron* **1984**, *40*, 457.
9. (a) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods Enzymol.* **1986**, *130*, 132. (b) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8462. (c) Senear, D.F.; Brenowitz, M.; Shea, M.A.; Ackers, G.K. *Biochemistry* **1986**, *25*, 7344.
10. (a) Larsen, T.A.; Kopka, M.L.; Dickerson, R.E. *Biochemistry* **1991**, *30*, 4443. (b) Narayana, N.; Ginell, S.L.; Russu, I.M.; Berman, H.M. *Biochemistry* **1991**, *30*, 4449. (c) Yanagi, K.; Prive, G.G.; Dickerson, R.E. *J. Mol. Biol.* **1991**, *217*, 201.
11. Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804-808.
12. Steitz, T. A. *Quart. Rev. Biophys.* **1990**, *23*, 203-280.

13. Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nature Struct. Biol.* **1998**, *5*, 104.
14. Pilch, D. S. *et al.* *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306.
15. Wong, J. M.; Bateman, E. *Nucl. Acids. Res.* **1994**, *22*, 1890.
16. White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532.
17. Momose, T.; Tamaka, T.; Yokota, T.; Nagamoto, N.; Yamada, K. *Chem. Pharm. Bull.* **1978**, *26*, 2224.
18. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.
19. Kim, Y.; Geiger, J. H.; Hahn, S.; Sigler, P. B. *Nature* **1993**, *365*, 512.
20. Gartenberg, M. R.; Crothers, D. M. *Nature* **1988**, *333*, 824.
21. Sluka, J. P.; Horvath, S. J.; Glasgow, A. C.; Simon, M. I.; Dervan, P. B. *Biochemistry* **1990**, *29*, 6551.
22. Ades, S. E.; Sauer, R. T. *Biochemistry* **1995**, *34*, 14601-14608.
23. Choo, Y.; Klug, A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 117-125.
24. White, S.; Szewczyk, J.W.; Turner, J.M.; Baird, E.E.; Dervan, P.B. *Nature* **1998**, *391*, 468.
25. Turner, J.M.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1997**, *119*, 7636.
26. Kielkopf, C.L.; White, S.; Szewczyk, J.W.; Turner, J.M.; Baird, E.E.; Dervan, P.B.; Rees, D.C. **1998**, *282*, 111.
27. Branden, C.; Tooze, J. *Introduction to Protein Structure*; Garland: New York, 1991.
28. Brameld, K.; Dasgupta, S.; Goddard, W.A. *J. Amer. Phys. Chem. B.* **1997**, *101*, 4851.
29. Kopka, M.L.; Goodsell, D.S.; Han, G.W.; Chiu, T.K.; Lown, J.W.; Dickerson, R. *Structure* **1997**, *5*, 1033.
30. Kim, J.L.; Nikolov, D.B.; Burley, S.K. *Nature* **1993**, *365*, 520.
31. Kent, S.B.H. *Ann. Rev. Biochem.* **1988**, *57*, 957.
32. Barlos, K.; Chatzi, O.; Gatos, D.; Stravropoulos, G. *Int. J. Peptide Protein Res.* **1991**, *37*, 513.

## CHAPTER 9

### Sequence Specific Recognition of Double Helical DNA by Major-Minor Groove Binding Ligands

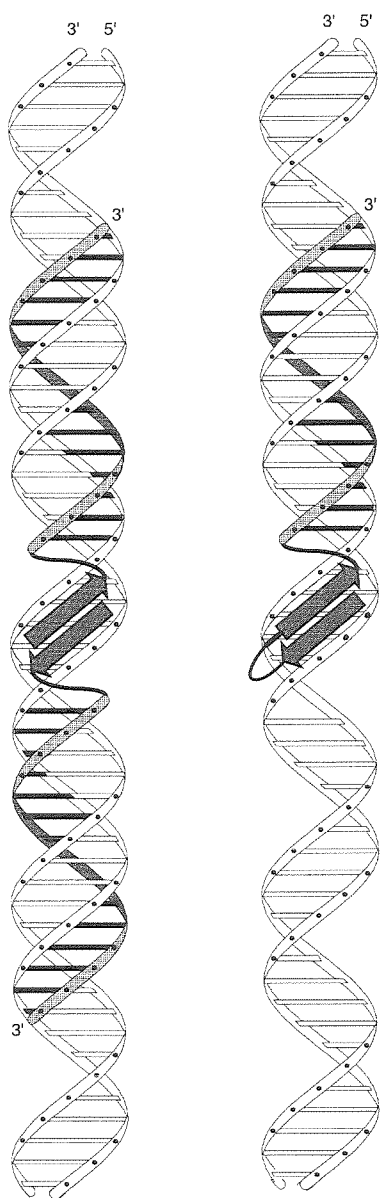
**Abstract:** *The two most powerful chemical approaches to date for the sequence specific recognition of double helical DNA are the recently described 2:1 polyamide- DNA complexes which recognize the minor groove and oligonucleotide directed triple helix formation in the major groove. We report here a pyrrole-imidazole polyamide oligonucleotide which specifically and simultaneously recognizes both the major and minor groove at subnanomolar concentrations. Binding affinity of the polyamide- oligonucleotide is enhanced by at least a factor of 250 over either of the individual oligonucleotide or the polyamide alone. Protocols for the solid phase synthesis of Py-Im polyamides are expanded to include polyamide-oligonucleotide conjugates.*

**Publications:** Szewczyk, Baird & Dervan *Angew. Chemie.* **1996**, 35, 1487-1489.  
Szewczyk, Baird & Dervan *J. Am. Chem. Soc.* **1996**, 118, 6778-6779.

Recently described 2:1 polyamide-DNA complexes<sup>1-7</sup> and oligonucleotide directed triple helix formation<sup>8,9</sup> provide powerful *chemical* approaches for the sequence specific recognition of double helical DNA. In one triple helical model, a pyrimidine oligonucleotide binds specifically to a purine tract in the major groove of double helical DNA parallel to the purine Watson-Crick (W-C) strand.<sup>8</sup> Specificity is derived from thymine (T) recognition of adenine-thymine base pairs (T•AT base triplets) and protonated cytosine (C<sup>+</sup>) recognition of guanine-cytosine base pairs (C<sup>+</sup>•GC base triplets).<sup>8</sup> In the 2:1 polyamide- DNA model, polyamides containing *N*-methylimidazole (Im) and *N*-methylpyrrole (Py) amino acids are combined in antiparallel side-by-side dimeric complexes with the minor groove of DNA.<sup>1</sup> The DNA sequence specificity of these small molecules can be controlled by the linear sequence of pyrrole and imidazole amino acids. A pairing of imidazole (Im) opposite pyrrole (Py) targets a G•C base-pair, while pyrrole opposite imidazole targets a C•G base-pair.<sup>1</sup> A pyrrole/pyrrole combination is partially degenerate and targets both T•A and A•T base-pairs.<sup>1-3</sup> Efficient solid phase synthetic methodology makes pyrrole- imidazole polyamides routinely available.<sup>10</sup> A key issue is to determine if a simple linker domain could combine both models within a single molecule without compromising the sequence specificity of either recognition element.

In one design, two polyamide-oligonucleotide conjugates bind antiparallel to a sequence of duplex DNA, with binding mediated by the dimerization of the individual polyamide moieties in the minor groove of DNA (Figure 9.1). In a second design, a single hairpin-polyamide-oligonucleotide conjugate binds a designated sequence (Figure 9.1). In both designs specificity will be derived from specific contacts in the major groove from the pyrimidine motif triple helix and in the minor groove from the 2:1 polyamide:DNA complex. We report here the synthesis and DNA-binding properties of both classes of polyamide-oligonucleotide conjugates.

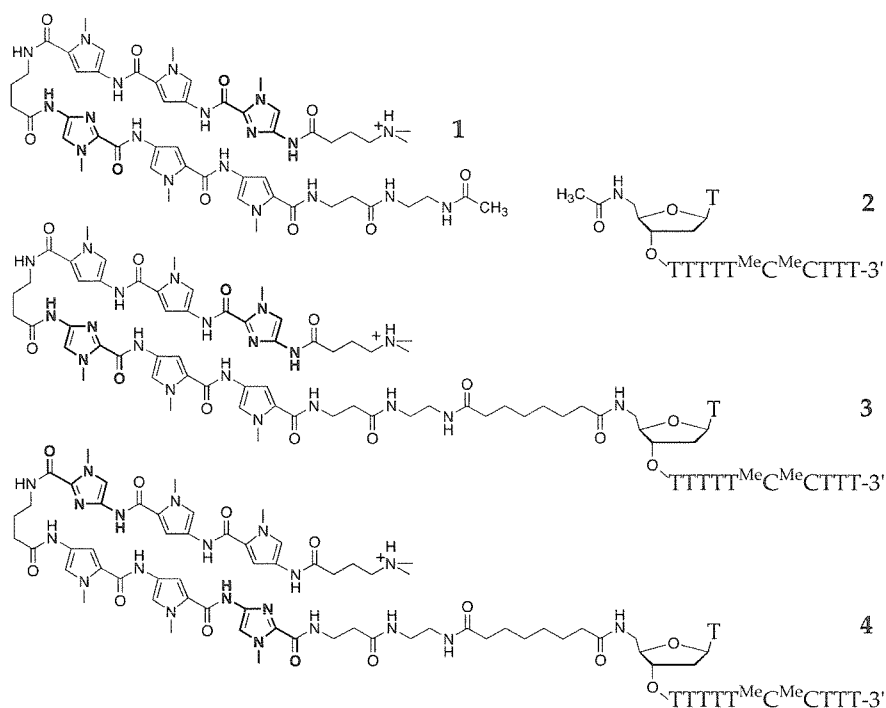




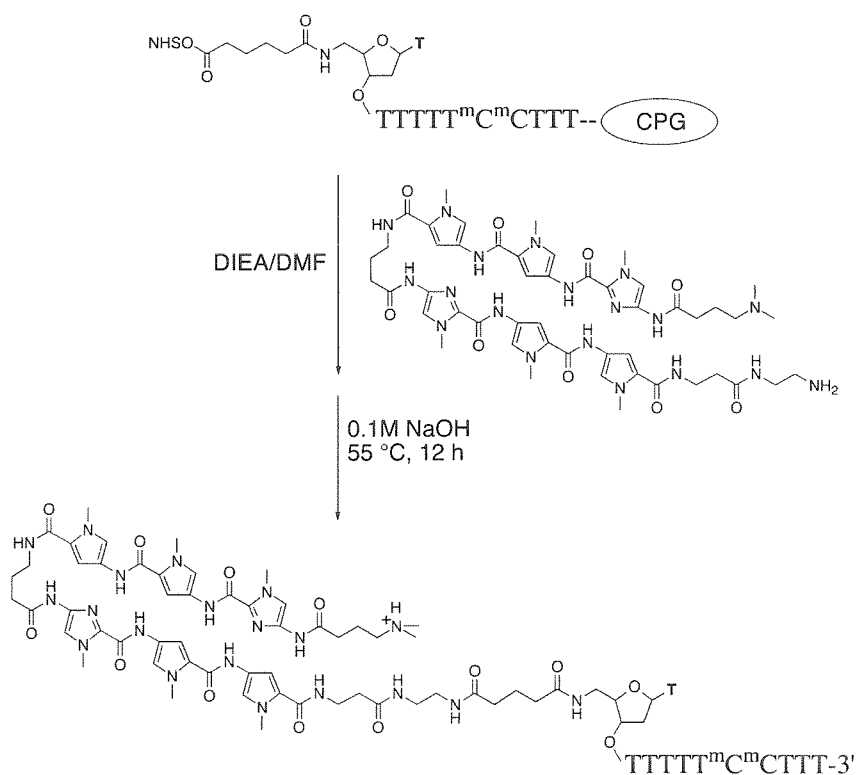
**Figure 9.1.** The recognition of double helical DNA by an oligonucleotide-polyamide conjugate (left) oligonucleotide directed triple helix formation in the major groove mediated by cooperative polyamide-dimerization in the minor groove. (right) Directed binding of a head-to-tail polyamide dimer in the minor groove mediated by oligonucleotide directed triple helix formation in the major groove.

**Solid Phase Methods.** A number of methods exist for the synthesis of oligonucleotide-peptide conjugates, based on post-synthetic modification,<sup>11</sup> assembly of a peptide on controlled pore glass followed by oligonucleotide synthesis,<sup>12</sup> and synthesis of an amine-modified oligonucleotide followed by solid phase synthesis of a peptide. Both post-synthetic modification as well as synthesis of an oligonucleotide, followed by solid phase assembly of the polyamide, seemed the most convenient for the purposes desired here.

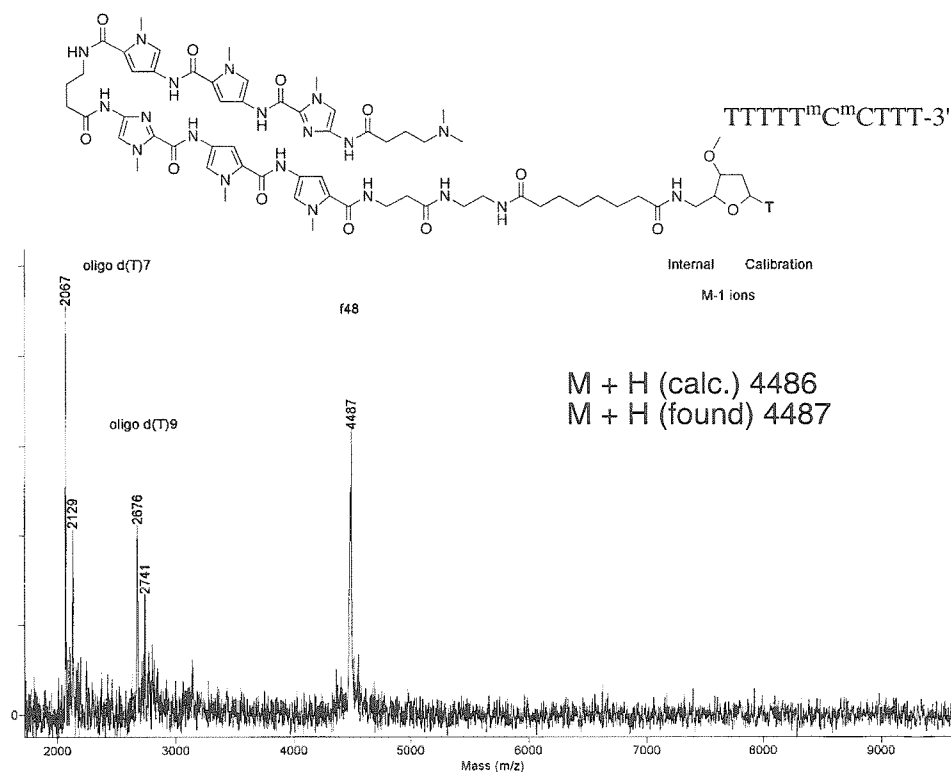
**Synthesis of Hairpin-Polyamide-Oligonucleotide.** The polyamide ImPyPy- $\gamma$ -ImPyPy- $\beta$ , modified with a free primary amine group at the carboxy terminus, was prepared by solid phase synthetic methodology and allowed to react with either acetic anhydride to provide hairpin **1** or a support bound *N*-hydroxysuccinimide ester of a suitably 5' modified 11 mer oligonucleotide 5'-TTTTTT<sup>Me</sup>C<sup>Me</sup>CTTT-3' to provide the conjugate, ImPyPy- $\gamma$ -ImPyPy- $\beta$ -(linker)-TTTTTT<sup>Me</sup>C<sup>Me</sup>CTTT-3' **3** upon deprotection (Figure 9.2 and 9.3). Similarly, a control conjugate **4** was synthesized with a reverse



**Figure 9.2.** Structures of hairpin polyamide **1**, oligonucleotide **2**, match polyamide-oligonucleotide **3**, and control polyamide-oligonucleotide **4**.



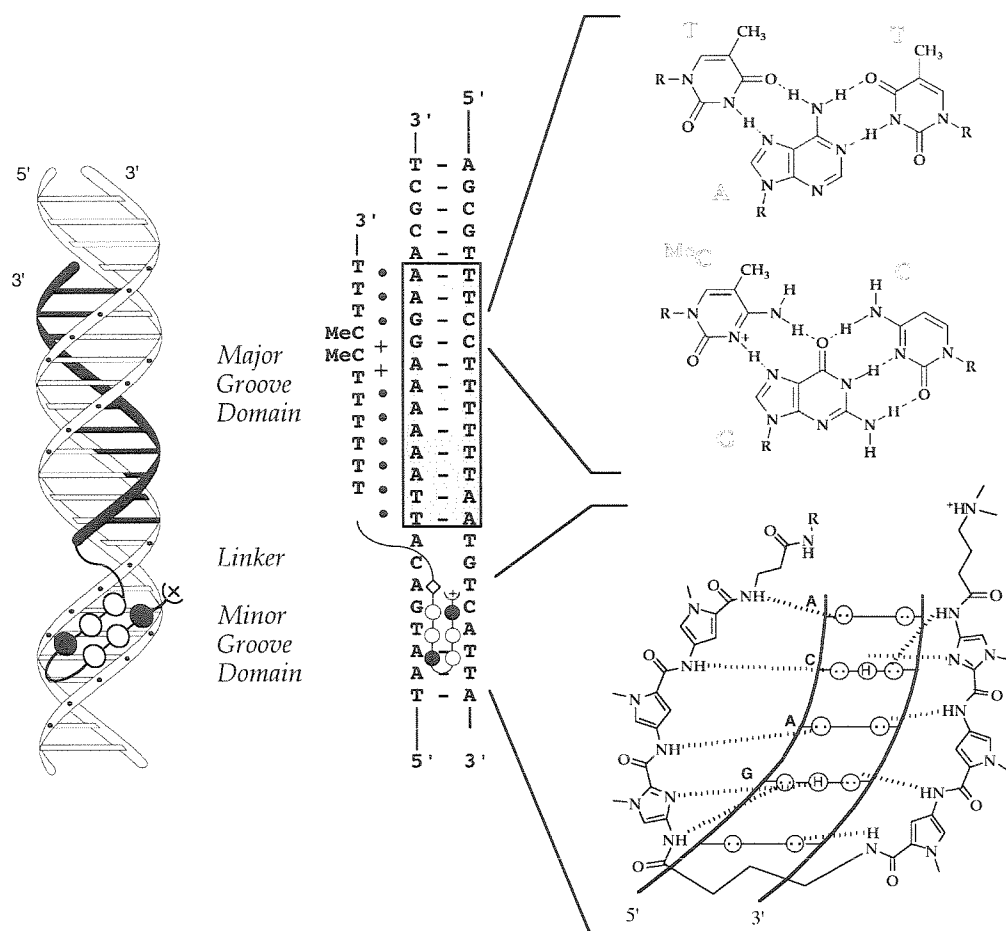
**Figure 9.3.** Synthetic Scheme for match polyamide-oligonucleotide **3**. Purified hairpin polyamide is coupled to protected oligonucleotide on a solid support, the resulting conjugate is simultaneously cleaved and deprotected and then isolated by HPLC.



**Figure 9.4.** MALDI-TOF mass-spectral analysis of purified polyamide-oligonucleotide **3** and Oligo d(T) molecular weight standards.

orientation polyamide component PyPyIm- $\gamma$ -PyPyIm- $\beta$ . Polyamide-oligonucleotides **3** and **4** were purified by HPLC and the molecular composition verified by MALDI-TOF MS (Figure 9.4), U/V, and HPLC analysis of enzymatic phosphodiester hydrolysis products.

We report that a hairpin polyamide linked to an 11 mer oligonucleotide specifically and simultaneously binds *both the major and minor grooves* of DNA at *subnanomolar* concentration. A pyrimidine oligonucleotide, 5'-TTTTTT<sup>Me</sup>C<sup>Me</sup>CTTT-3', covalently tethered by a twelve atom linker to the hairpin polyamide of sequence composition ImPyPy- $\gamma$ -ImPyPy- $\beta$  ( $\beta$  =  $\beta$ -alanine) binds to 16 base pairs of double helical DNA by formation of a specific pyrrole-imidazole polyamide-DNA complex in the minor groove and a specific triple helix in the major groove (Figure 9.2 and 9.3). In a formal sense the hybrid ligand-DNA complex affords 10 discrete hydrogen bonds for recognition of 5 base pairs in



**Figure 9.5.** (left) Ribbon model depicting a hairpin Py-Im polyamide-oligonucleotide simultaneously binding the major and minor grooves of double helical DNA. Filled and unfilled circles represent Im and Py rings, and the diamond represents  $\beta$ . (right) 2-D models depicting the molecular interactions responsible for sequence specific recognition. (top) The T•AT and C+GC base triplets formed by Hoogsteen hydrogen bonding of T to a Watson-Crick AT base pair and protonated  $^{Me}C$  to a Watson-Crick GC base pair. (bottom) Expected complex of ImPyPy- $\gamma$ -ImPyPy- $\beta$  with the site 5'-TGACA-3'.

the minor groove (5'-TGACA-3') and 22 discrete hydrogen bonds for recognition of 11-base pairs in the major groove (5'-AAAAAACC AAA-3') (Figure 9.5).

A DNA fragment was constructed which contains the 18-base pair site 5'-TGACATTAAAAAAGGAAA-3, composed of a 5-base pair polyamide binding site separated by 2-base pairs from an 11-base pair triple helix site. For controls, isolated 5-base pair polyamide and 11-base pair triplex binding sites were also present on the fragment. Quantitative DNase I footprint titration experiments<sup>13</sup> on a 302-base pair restriction

**Table 9.1.** Equilibrium Association Constants.

Ligand	$K_a$ ( $M^{-1}$ )
polyamide <b>1</b>	$1.5 (\pm 0.2) \times 10^8$
oligonucleotide <b>2</b>	$3.0 (\pm 1.0) \times 10^5$
polyamide-oligonucleotide <b>3</b>	$> 4.0 \times 10^{10}$
(mismatch) polyamide-oligonucleotide <b>4</b>	$4.0 (\pm 0.4) \times 10^8$

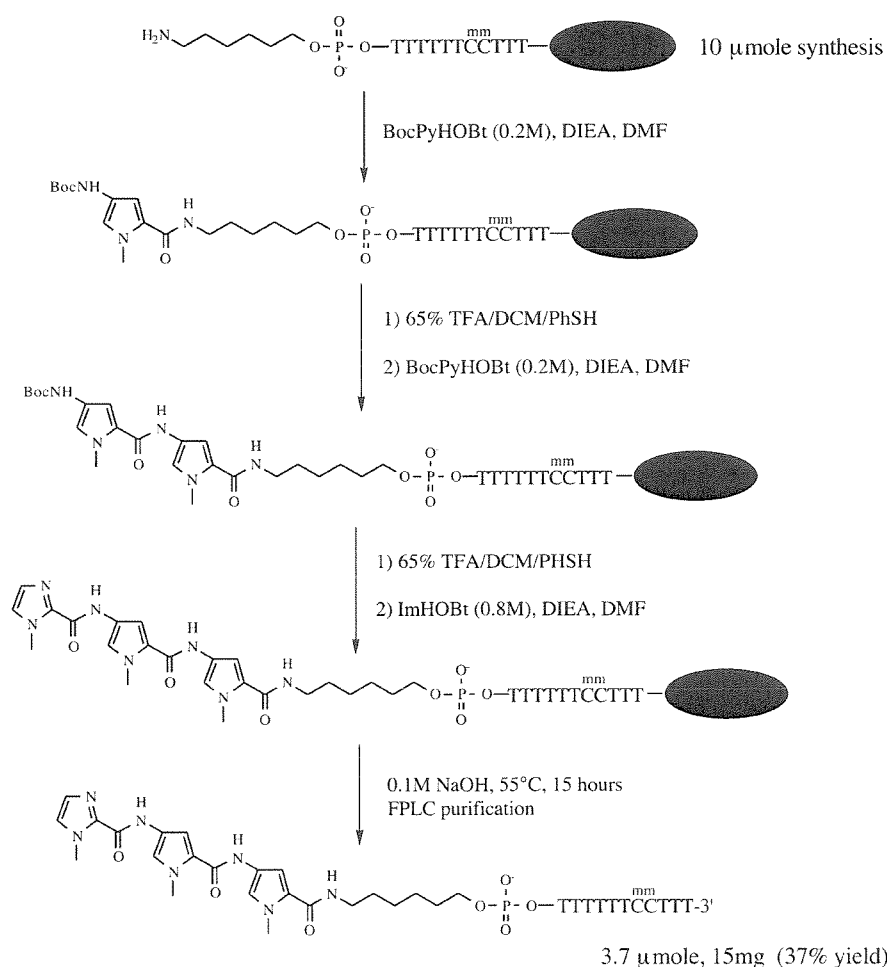
fragment afford the binding affinity of the designed ligands (Table 9.1). Polyamide **1** binds to 5 base pairs of the target site, 5'-TGACATTAAAAAAGGAAA-3, with an apparent first order association constant,  $K_a = 1.5 \times 10^8 M^{-1}$ . Unmodified oligonucleotide **2** binds 11 base pairs, 5'-TGACATTAAAAAAGGAAA-3, with a first order association constant,  $K_a = 3.0 \times 10^5 M^{-1}$  (Table 1). The hybrid polyamide-oligonucleotide **3** binds the designated target site 5'-TGACATTAAAAAAGGAAA-3 with an apparent first order equilibrium association constant,  $K_a > 4.0 \times 10^{10} M^{-1}$ ,<sup>[1]</sup> representing at least a 250-fold increase in binding affinity relative to the unlinked subunits. In controls, the binding affinity of **3** at either the isolated polyamide, 5'-TGACATTAGCTCGTAATG-3', ( $K_a < 1.0 \times 10^5 M^{-1}$ ) or oligonucleotide, 5'-ACCGGTTAAAAAACCAAA-3', ( $K_a = 5.5 \times 10^8 M^{-1}$ ) binding sites is reduced by approximately 100-fold relative to binding at the designated match site. Mismatch polyamide-oligonucleotide **4** binds the 18 base pair target, 5'-TGACATTAAAAAAGGAAA-3, ( $K_a = 4 \times 10^8 M^{-1}$ ) with 100-fold reduced affinity relative to match conjugate **3**.

Polyamide- oligonucleotide **3** binds with subnanomolar affinity *only* at the designated target site, indicating that the 3.3 kcal/mol binding enhancement results from simultaneous sequence specific recognition of *both* the major and the minor grooves. By mimicking the function of certain sequence specific DNA binding proteins,<sup>14</sup> hairpin polyamide-oligonucleotides expand the sequence composition repertoire targeted by artificial methods and create a new motif for the design of molecules for the sequence-specific recognition of both grooves of the DNA double helix.

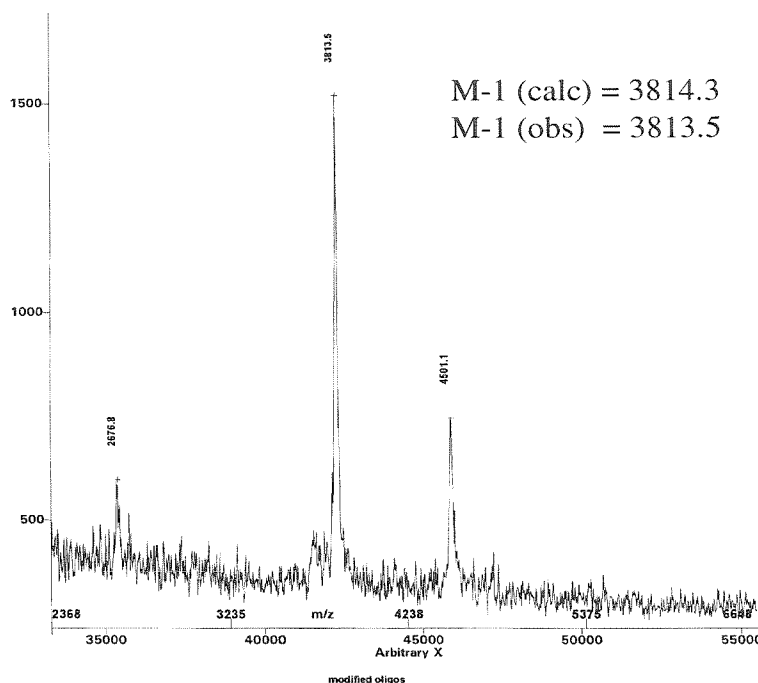
### Cooperative Sequence Specific Major - Minor Groove DNA Binding Ligands.

Cooperative interactions between DNA binding ligands are critical to their specificity, affinity, and biological activity.<sup>14</sup> We report here that a Py-Im polyamide-oligonucleotide can cooperatively and simultaneously bind both the major and minor groove of a double stranded DNA template.

Polyamide-oligonucleotides were prepared by automated synthesis of the amino-modified oligonucleotide, followed by manual stepwise Boc-chemistry solid phase synthesis of



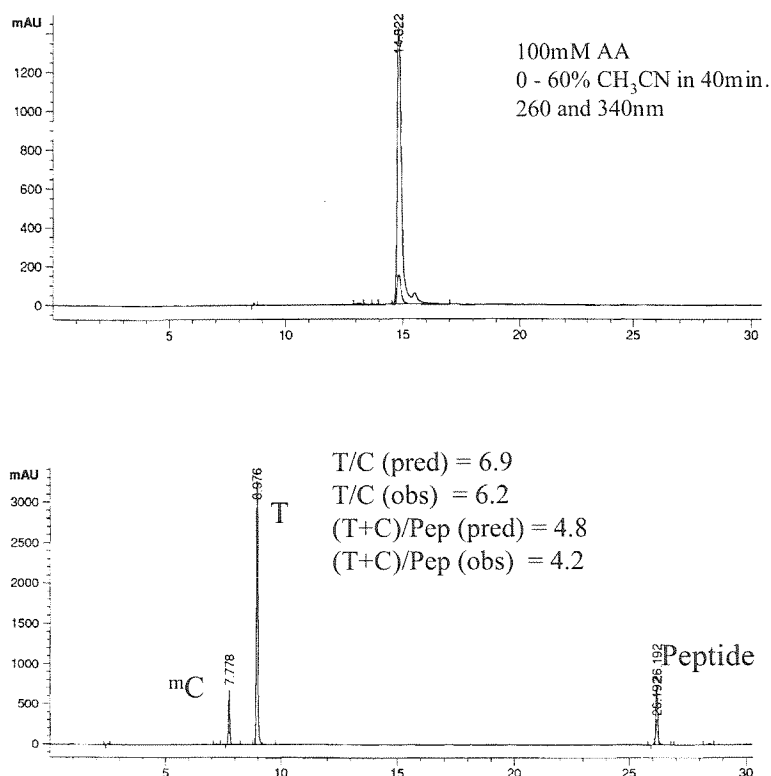
**Figure 9.6.** The Solid Phase Synthesis of polyamide oligonucleotide conjugate ImPyPy-C<sub>6</sub>-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3'. The oligonucleotide portion of the molecule is synthesized by standard machine assisted oligonucleotide chemistry and the polyamide added with manual solid phase Boc-chemistry.



**Figure 8.7.** MALDI-TOF mass spectrum of a minor groove polyamide-oligonucleotide conjugate. Ions at 4501.1 and 2676.8 correspond to internal standards, ion at 3813.5 is consistent with full length product.

the polyamide (Figure 9.6 and 9.7). An oligonucleotide was assembled at 10  $\mu$ mol scale on an automated DNA synthesizer using standard DNA cycles, a 5'-MMT-C6-amino modifier was attached using an extended 10 min. synthesis cycle, and the synthesis removed from the synthesizer.

The MMT group was removed from the modified oligonucleotide by manual treatment with 3% TCA in dichloromethane until no yellow color was observed in the wash. The controlled pore glass support was then removed from the synthesis cartridge and transferred to a standard peptide synthesis reaction vessel. The oligonucleotide was reacted with a 0.2M solution of Boc-Py-OBt in DMF/0.4M DIEA for 45 min. The reaction was determined to be complete by the quantitative ninhydrin test which showed a distinct blue color for the oligo amine, consistent with a 0.05mmol/gram loading, and a lack of a blue color after a 1 hour reaction time. The Boc group was removed with 65% TFA/DCM/0.5M PhSH for 20 min., and a second pyrrole coupled to from the aromatic carboxamide. The Boc group is again removed with TFA, and the peptide capped with N-methylimidazole-2-carboxylic acid. The

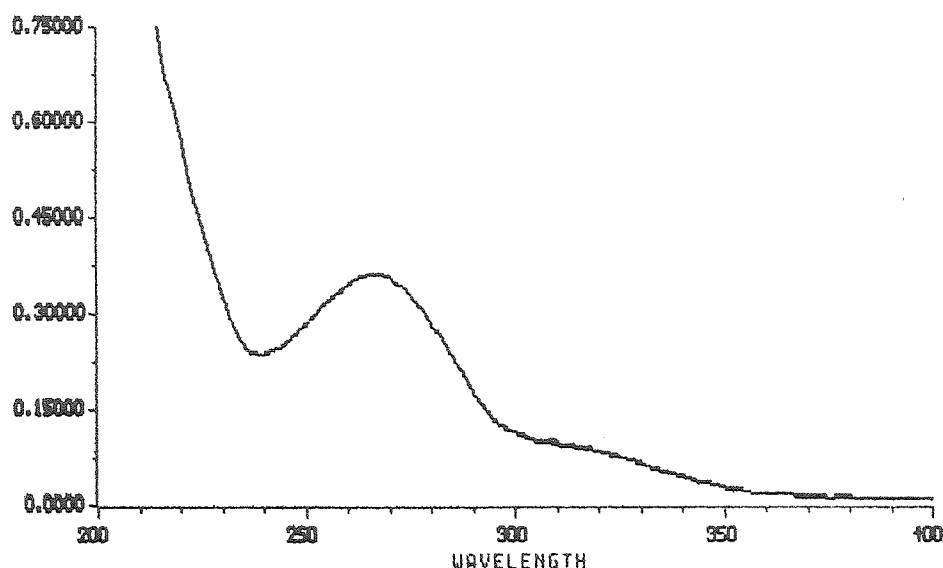


**Figure 9.8.** HPLC analysis of purified peptide- oligonucleotide conjugate. a) intact conjugate, major absorbance measured at 260nm, minor absorbance at 340nm. b) Enzymatic digest of oligonucleotide conjugate, observed ratios of digestion products are consistent with predicted ratios.

oligonucleotide- peptide conjugate is then simultaneously cleaved from the resin and deprotected by treatment with 0.1M NaOH at 55°C for 15 hours and purified by FPLC chromatography.

The molecular composition of purified polyamide-oligonucleotides was verified by MALDI-TOF mass spectroscopy, and HPLC analysis of enzymatic phosphodiester hydrolysis products. A single reversed phase purification yields up to 15 mg of peptide conjugate, ImPyPy-C6-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3', of reasonably high purity. The product obtained is characterized by a number of techniques. MALDI- time of flight mass spectroscopy shows a single peak corresponding to a molecular mass of 3813.5, in excellent agreement with the predicted mass of 3814.3, indicating that full length product has indeed been isolated. Reverse phase HPLC analysis of 10 nmoles of the conjugate indicates one major product, absorbing at



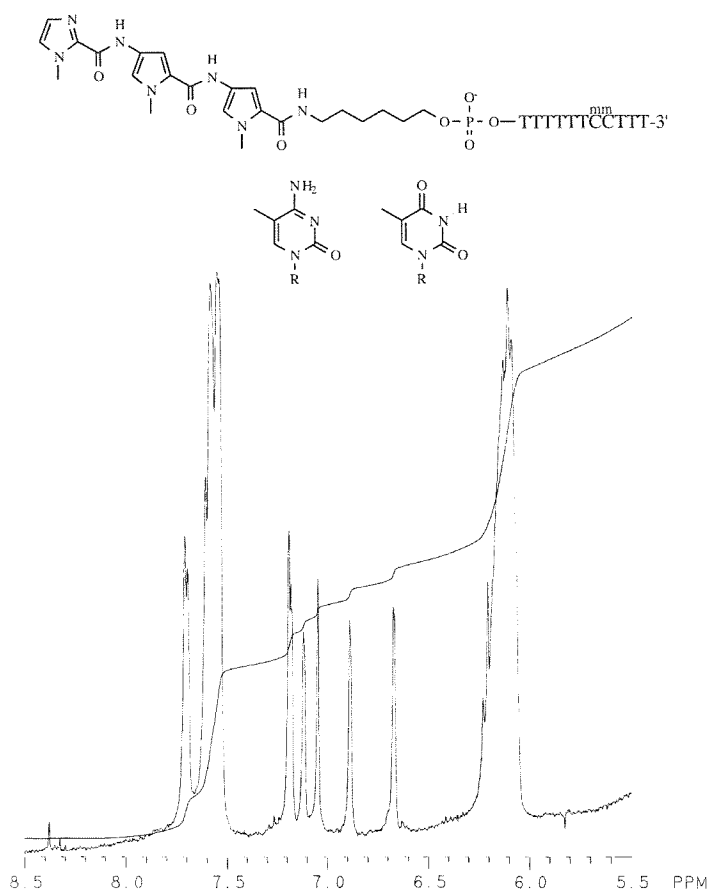


**Figure 9.9.** U/V Visible spectrum of the peptide- oligonucleotide conjugate. Spectra is consistent with the addition of the U/V spectra of the individual peptide and oligonucleotide components.

both oligonucleotide wavelength (260) and peptide wavelength (340). Enzymatic digestion and subsequent HPLC analysis of a 10 nmole sample of conjugate is consistent with the proposed composition of the oligonucleotide.

U/V spectroscopy, Figure 9.9, indicates an additive spectra as might be expected for a conjugate of 2-Imidazole Netropsin and an 11-mer thymidine rich oligonucleotide. From the extinction coefficient of the bases, 8,800 for T at 260 nm and 5,700 for MeC at 260 nm and the reported extinction coefficient of 2-ImN at 255 nm, 19,000 and 302 nm, 26,000 it is possible to predict the ratio of the extinction coefficients at 260 nm.<sup>15</sup> Assuming a contribution from the oligo of 90,600 and 19,000 from the polyamide, a ratio of 4.2 is expected, and a ratio of 3.7 is observed.

A 14 mg sample of the polyamide conjugate was dissolved in 700 $\mu$ l of deuterium oxide and analyzed by  $^1\text{H}$  NMR spectroscopy at 300 MHz. Most of the spectrum is complex; however, the aromatic region of the spectrum was readily interpreted. The protons expected in the aromatic region correspond to the polyamide ring protons, and the  $\text{C}_6$  ring protons of thymidine and 5-methylcytidine. The observed spectra is consistent with the predicted sequence ImPyPy-TTTTTT<sup>m</sup>C<sup>m</sup>CTT-3', with 2 protons observed at 7.7 most likely

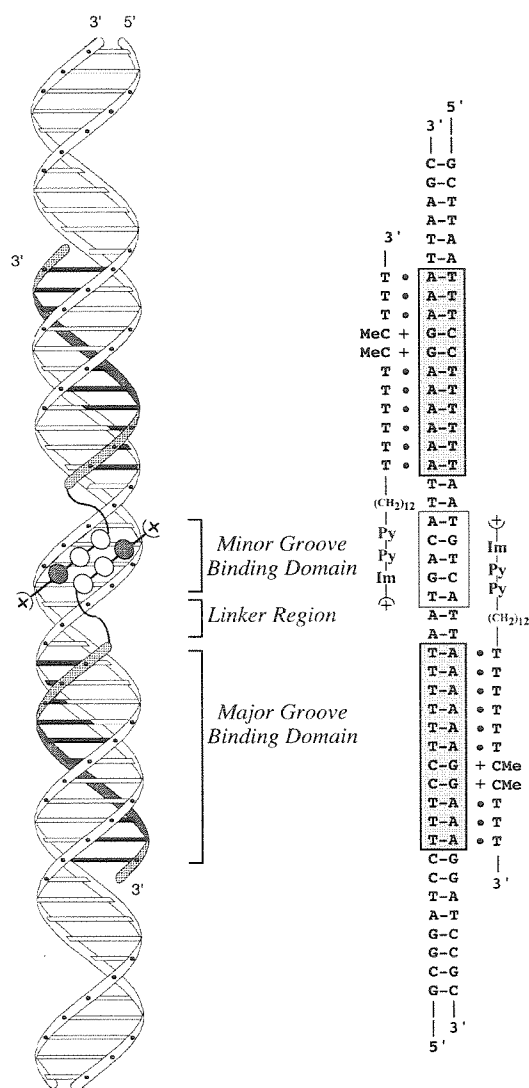


**Figure 9.10.** Aromatic region of the  $^1\text{H}$  NMR spectra of the minor-groove peptide oligonucleotide conjugate  $\text{ImPyPy-C}_6\text{-TTTTTTCCTTT-3'}$ . The sequence content of the product is consistent with the spectrum, 2 C, 9 T, 1 Im, 2 Py (4H), and 11 anomeric protons.

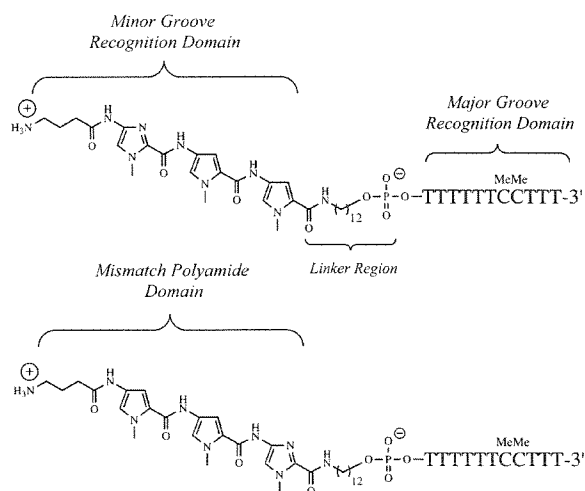
corresponding to the cytidine, 9 protons at 7.6 most likely corresponding to the thymidine contribution, 2 protons at 7.3 corresponding to the imidazole ring, four pyrrole doublets at 7.2, 7.1, 6.9, and 6.7 corresponding to four protons, and 11 anomeric protons at 6.2, Figure 8.10. The purity of the sample, as determined by NMR, is  $> 98\%$ . The ability to rapidly obtain NMR data (30 minutes of acquisition) on a molecule of this size (3 kD) warrants a synthesis scale such as the one chosen here.

A DNA fragment was constructed containing the symmetrical 31 base pair target site  $5'\text{-T}_3\text{C}_2\text{T}_6\text{A}_2\text{TGTCAT}_2\text{A}_6\text{G}_2\text{A}_3\text{-3'}$ , composed of two 11 base pair triple helix sites separated by a 9 base pair region containing the 5 base pair polyamide binding site. Quantitative DNase

I footprint titration experiments on the 268 base pair restriction fragment afford the binding affinities of the designed polyamide-oligonucleotides **1** and **2**.



**Figure 9.11.** Ribbon model depicting two Py-Im polyamide-oligonucleotides simultaneously binding as a homodimer in the major and minor grooves of double helical DNA.



**Figure 9.12.** The cognate Py-Im polyamide-oligonucleotide **1** and Py-Im polyamide-oligonucleotide **2** containing a mismatch polyamide moiety.

We find that polyamide-oligonucleotide **1** binds the designated 31 base pair target with an apparent first order equilibrium association constant,  $K_a = 1.7 \times 10^8 \text{ M}^{-1}$ , representing a 100-fold increase in apparent affinity over both the mismatch polyamide-oligonucleotide **2** and unmodified oligonucleotide, 5'-d(TTTTTT<sup>Me</sup>C<sup>Me</sup>CTTT)-3' (Table 9.2).

The untethered polyamide ImPyPy was found to bind its 5'-TGTC A-3' target site with an apparent association constant,  $K_a = 1.5 (\pm 0.3) \times 10^5 \text{ M}^{-1}$ . Polyamide-oligonucleotide **2** contains a mismatch in the polyamide moiety and is found to bind without positive cooperativity, demonstrating that the observed cooperative interaction is mediated by sequence specific recognition in the minor groove.

**Table 9.2.** Equilibrium Association Constants ( $M^{-1}$ )<sup>a,b</sup>.

Hybrid	K <sub>a</sub> ( $M^{-1}$ )
<b>1</b>	$1.7 (\pm 0.9) \times 10^8$
<b>2</b>	$2.5 (\pm 0.3) \times 10^6$
5'-d(TTTTTT <sup>Me</sup> C <sup>Me</sup> CTTT)-3'	$1.7 (\pm 0.1) \times 10^6$

<sup>a</sup>Experiments performed at 22°C in the presence of 10 mM NaCl, 10 mM Bis Tris•HCl (pH 7.0), and 250  $\mu$ M spermine. <sup>b</sup> Values reported are the mean values measured from three or more footprint titration experiments.

The 2.7 kcal/mol binding enhancement reported here is achieved with an unconstrained 12 carbon linker domain for crossing from the major to the minor groove. The enhanced affinity of the designed homodimeric complex is only consistent with simultaneous recognition of *both* the major and minor grooves. It is interesting to compare this artificial major-minor groove ligand-DNA complex with certain sequence specific DNA binding proteins which also make specific contacts with both grooves of the DNA double helix.<sup>16</sup> By mimicking such DNA binding proteins, cooperative polyamide- oligonucleotide complexes expand the sequence composition repertoire of oligonucleotide directed triple helix formation. Pyrrole-imidazole polyamide- oligonucleotides provide a powerful new paradigm for the design of artificial molecules for the sequence-specific recognition of both grooves of duplex DNA.

## Experimental Section.

All gel electrophoresis and synthesis of novel linkers was performed by Jason Szewczyk as described.<sup>17,18</sup>

**Materials.** 6-(4-Monomethoxytritylamino)propyl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite, 5'-Amino-Modifier C6, 12-(4-Monomethoxytritylamino)propyl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite, 5'-Amino-Modifier C12, dT CE (2-cyanoethyl) phosphoramidite, 0.45M sublimed tetrazole in acetonitrile, THF/Lutidine/Ac<sub>2</sub>O (8:1:1), 10% MeIm in THF, 0.1M I<sub>2</sub> in THF/Pyridine/H<sub>2</sub>O, 3% TCA/DCM, 5-Methylcytidine CE (2-cyanoethyl) phosphoramidite, and bulk 500Å dT-Icaa-CPG were purchased from Glen Research. All 10 µmole synthesis columns were packed manually from bulk support. 1,2,3- Benzotriazol-1-yl-4[[tert- Butyloxy)carbonyl]-amino]-1-methylpyrrole-2-carboxylate and 1,2,3- Benzotriazol-1-yl-4[[tert- Butyloxy)carbonyl]-amino]-1-methylimidazole-2-carboxylate were prepared as previously described.<sup>13,17</sup> Trifluoroacetic acid (TFA) was purchased from Halocarbon. *N,N*-diisopropylethylamine (DIEA), *N,N*-dimethylformamide (DMF), 0.0002M potassium cyanide/pyridine, and Acetic Anhydride (Ac<sub>2</sub>O) were purchased from Applied Biosystems, and Dicyclohexylcarbodiimide (DCC) and Hydroxybenzotriazole (HOBt) were from Peptides International. Boc-GABA was from NOVABiochem, dichloromethane (DCM), reagent grade from EM. Thiophenol was purchased from Aldrich. A shaker for manual solid phase synthesis was obtained from Milligen. Quik-Sep polypropylene disposable filters were purchased from Isolab Inc. and are used for filtration of CPG. Screw-cap glass peptide synthesis reaction vessels (5 ml) with a #2 sintered glass frit were made at the Caltech glass shop as described by Kent.<sup>15</sup> NMR were recorded on a GE 300 instrument operating at 300MHz (<sup>1</sup>H). Chemical shifts are reported in

ppm relative to the solvent residual signal; UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed either on a HP 1090 M analytical HPLC or a Beckman Gold system using a RAINEN C18, Microsorb MV, 5 $\mu$ m, 300 x 4.6 mm reversed phase column in 100 mM Ammonium Acetate, pH 4.9 with acetonitrile as eluent and a flow rate of 1.0 ml/min, gradient elution 1.0% acetonitrile/min. Oligonucleotide conjugates were purified by FPLC (Pharmacia) on a ProRPC HR 10/10 reversed phase column using a linear gradient from 0 to 40% acetonitrile in 55 min., 100 mM triethylammonium acetate, pH 7.0.

**ImPyPy-C<sub>6</sub>-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3'** The oligonucleotide DMT-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3' was prepared on an Applied Biosystems Model 394B DNA synthesizer using a manually prepared 10  $\mu$ mole synthesis column and a standard 10  $\mu$ mole synthesis cycle. C<sub>6</sub>-AminoModifier-MMT (100  $\mu$ mole) was dissolved in 1,100 $\mu$ l of anhydrous acetonitrile, vortexed vigorously, and placed on the synthesizer. The amino modifier was added by machine synthesis using a modified 10  $\mu$ mole synthesis cycle with an extended 10 min. coupling time and the MMT group left on. The column is manually washed with 50 ml of 3% TCA/dichloromethane for a period of 12 min., until yellow color is no longer observed in the wash. The column is washed with 15 ml dichloromethane and dried *in vacuo*. The CPG is transferred to a 5 ml glass manual peptide reaction vessel and washed with DMF (30 sec.). A sample is taken for ninhydrin test and an absorbance consistent with 50  $\mu$ mole/ gram substitution found. Boc-Pyrrole-OBt ester (70 mg, 200  $\mu$ mole) is dissolved in DMF (600 $\mu$ l) and DIEA (68 $\mu$ l) added. The coupling mixture is added to the reaction vessel and allowed to shake for 60 min. A sample is taken for ninhydrin test and no blue color observed, indicating at least 90% reaction. The resin is washed with DMF (30 sec.) followed by dichloromethane (30 sec.) and 65% TFA/DCM/0.5M PhSH (30 sec.). The resin is shaken in 65% TFA/DCM/0.5M PhSH for 20 min., drained, washed with DCM (30 sec.) followed by DMF 30sec. A second equivalent of

Boc-Pyrrole-OBt is added under identical conditions to the first, and the reaction shaken for 1 hour, washed DMF (30 sec.), DCM (30 sec.) and treated with 65% TFA/DCM/0.5M PhSH as described for the first deprotection. N-methylimidazole-2-carboxylic acid (133 mg) is activated in 1 ml of DMF with HOBt/DCC as previously described,<sup>13</sup> added to the reaction vessel with DIEA (200μl) and the reaction allowed to shake for 2 hours. The CPG is washed with DMF (30 sec.), DCM (30 sec.), and dried *in vacuo*. The entire CPG (c.a. 210 mg) is placed in 2 ml of 0.1M NaOH and heated at 55°C for 15 hours. The CPG is removed by filtration through a polypropylene filter, and 1 ml of 1M Triethylammonium acetate, pH 7, added and the reaction diluted to 10 ml total volume with water. The mixture is purified in two separate portions by FPLC. In each run the conjugate is triple injected in 1.5 ml volume portions. Collected fractions are lyophilized, and re-lyophilized from water to yield the desired conjugate ImPyPy-C<sub>6</sub>-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3' (15 mg, 37% yield). UV 260 (118,000), 304 (33,000). Analytic HPLC (10 nmole), r.t., 14.8 min., 18,800u, Enzymatic digestion (10 nmole), 7.8 min. (mC, 2441u), 9.0 min. (T, 15,235u) 26.2 min. (polyamide, 4257u (260), 1843u (340)) <sup>1</sup>H NMR (D<sub>2</sub>O) δ 7.71 (s, 1H), 7.69 (s, 1H), 7.50 (m, 9H), 7.19 (s, 1H), 7.18 (s, 1H), 7.12 (d, 1H, *J* = 1.2 Hz), 7.05 (d, 1H, *J* = 1.3Hz), 6.89 (d, 1H, *J* = 1.3Hz), 6.67 (d, 1H, *J* = 1.3Hz), 6.11 (m, 11H), 4.89, 4.76, 4.70, 4.22, 4.01, 3.91, 3.88, 3.75, 3.42, 3.11, 2.40, 2.25, 2.20, 1.80, 1.75, 1.41, 1.23 MALDI-TOF MS, calc. M-1 3814.3, found 3813.5.

**AcImPyPy-C<sub>12</sub>-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3'** BocPyPy-C<sub>12</sub>-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT--CPG is assembled as described for the C<sub>6</sub> compound. The N-boc group removed with 65% TFA/DCM/0.5M PhSH, and the resin treated with a solution of Boc-Im-OBt (70 mg) DIEA (68μl) and DMF (600μl), and shaken for 1 hour. The CPG is washed, DMF (30 sec.), DCM (30 sec), and dried *in vacuo*. One third of the CPG, 70 mg, is removed from the synthesis and placed in a 10 μmole DNA synthesis column. Two syringes are used simultaneously to manipulate reagents into and out of the column. The column is washed with DCM and *carefully* treated with 65% TFA/DCM/0.5M PhSH to remove the N-Boc group. The column is carefully washed with DCM (20 ml) and DMF (20 ml) and then treated with acetylation mixture for 1 hour, washed

with DMF (20 ml) and DCM (20 ml) and the cartridge dried *in vacuo*. The resin is removed from the column and placed in 1 ml 0.1M NaOH at 55°C for 15 hours. The CPG is removed by filtration, 1 ml of 1M pH 7 triethylammonium acetate added, and the mixture purified by FPLC. The appropriate fractions are collected and concentrated *in vacuo* to give AcImPyPy-C<sub>12</sub>-TTTTTTmCmCTTT-3' (336 nmole, 10% yield). UV<sub>260</sub>(110,000), 304 (36,000). Analytic HPLC (10 nmole), r.t., 19.7 min., 13,000u, Enzymatic digestion (10 nmole), 7.7 min. (mC, 1551u), 8.9 min. (T, 11374u) 33.2 min. (polyamide, 1646 u (260), 1281u (340)) MALDI-TOF MS, calc. M-1 3953.5, found 3952.9.

**Boc-GABA-ImPyPy-C<sub>12</sub>-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3'** A sample of BocImPyPy-C<sub>12</sub>-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT--CPG, 140 mg, is placed in a 10 μmole DNA synthesis column. Two syringes are used simultaneously to manipulate reagents into and out of the column. The column is washed with DCM and *carefully* treated with 65% TFA/DCM/0.5M PhSH to remove the N-Boc group. The column is carefully washed with DCM (20 ml) and DMF (20 ml) and then treated with the HOBt ester of Boc-GABA prepared *in situ* (1 mmol, 200μl DIEA, 1ml DMF) and allowed to react for 1 hour.<sup>13</sup> The CPG is washed with DMF (20 ml), DCM (20 ml), and dried *in vacuo*. The resin is removed from the column and placed in 1 ml 0.1M NaOH at 55°C for 15 hours. The CPG is removed by filtration, 1 ml of 1M pH 7 triethylammonium acetate added and the mixture purified by FPLC. The appropriate fractions are collected and concentrated *in vacuo* to give Boc-GABA-ImPyPy-C<sub>12</sub>-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3' (343 nmole, 5% yield). UV<sub>260</sub>(120,000), 304 (34,000). Analytic HPLC (5 nmole), r.t., 23.0 min., 6,000u, Enzymatic digestion (5 nmole), 7.7 min. (mC, 779u), 8.9 min. (T, 5873 u) 42.1 min. (polyamide, 767u (260), 597u (340)) MALDI-TOF MS, calc. M-1 4094.7, found 4096.8.

**H<sub>2</sub>N-GABA-ImPyPy-C<sub>12</sub>-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3'**

**Boc-GABA-ImPyPy-C<sub>12</sub>-**

TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3' (330 nmole) is treated with 400μl of 65% TFA/DCM/0.5M PhSH for 30 min. 2 ml of 1M pH 7 triethylammonium acetate is added and 5 ml of water. The reaction is vortexed, frozen, and lyophilized. The reaction is dissolved in 5 ml of 100 mM



triethylammonium acetate and purified by FPLC, appropriate fractions are collected and concentrated *in vacuo* to give H<sub>2</sub>N-GABA-ImPyPy-C<sub>12</sub>-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3' (240 nmole, 70% yield). MALDI-TOF MS, calc. M-1 3995.5, found 3999.7.

**EDTA-GABA-ImPyPy-C<sub>12</sub>-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3'**

H<sub>2</sub>N-GABA-ImPyPy-C<sub>12</sub>-

TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3' is dissolved in 500μl 500 mM Carbonate buffer, pH 9.5. 10 mg of the monoanhydride of EDTA is added and the reaction allowed to proceed for 15 minutes. 1 ml of triethylammonium acetate, pH 7.0 is added with 4 ml of water and the reaction immediately purified by FPLC. The appropriate fractions are collected and concentrated *in vacuo* to give EDTA-GABA-ImPyPy-C<sub>12</sub>-TTTTTT<sup>m</sup>C<sup>m</sup>CTTT-3' (70 nmole, 41% yield).

## References

1. (a) Wade, W.S.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (b) Wade, W.S.; Mrksich, M.; Dervan, P.B. *Biochemistry* **1993**, *32*, 11385. (c) Mrksich, M.; Wade, W.S.; Dwyer, T.J.; Geierstanger, B.H.; Wemmer, D.H.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586.
2. (a) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (b) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 3325. (c) Geierstanger, B.H.; Dwyer, T.J.; Bathini, Y.; Lown, J.W.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 4474. (d) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646. (e) Geierstanger, B.H.; Jacobsen, J.P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Biochemistry* **1994**, *33*, 3055.
3. (a) Pelton, J.G.; Wemmer, D.E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723. (b) Pelton, J.G.; Wemmer, D.E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S.T.; Sundaralingham, M. *Nature Struct. Biol.* **1994**, *1*, 169.
4. (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983. (b) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147. (c) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6153. (d) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Chem. & Biol.* **1996**, *3*, 369. (e) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198. (f) de Claire, R. P. L.; Geierstanger B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909. (g) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756. (h) White, S.; Baird, E. E.; Dervan, P. B. *Chem & Biol.* **1997**, *4*, 569.
5. (a) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559. (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953. (c) Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 7636. (d) Turner, J. M. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219.
6. (a) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160. (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1997**, *3*, 1600. (c) Trauger, J.W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534.
7. De Clairac, R.P.L.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1997**, *119*, 7906.
8. (a) Moser, H.E.; Dervan, P.B. *Science* **1987**, *238*, 645–650. (b) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhou, N.; Decout, J.L.; Thoung, N.T.; Lhomme, J.; Helene, C. *Nucleic Acids Res.* **1987**, *15*, 7749. (c) Strobel, S.A.; Doucettstamm, L.A.; Riba, L.; Housman, D.E.; Dervan, P.B. *Science* **1991**, *254*, 1639–1642. (d) Thuong, N.T.; Helene, C. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 666–690.
9. (a) Maher, J. L.; Dervan, P. B.; Wold, B. *Biochemistry* **1992**, *31*, 71. (b) Duvalentin, G.; Thuong, N.T.; Helene, C.; *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 504.
10. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.

11. (a) Ede, N.; Tregear, G.W.; and Haralambidis, J.; *Bioconj. Chem.* **1994**, *5*, 373-378, (b) Haralambidis, J.; Lagniton, L.; and Treger, G.W. *Bioorg. and Med. Chem. Let.* **1993**, *4*, 1005-1010.
12. (a) Haralambidis, J.; Duncan, L.; Angus, K.; and Tregear, G.W. *Nuc. Acid. Res.* **1990**, *18*, 493-499, (b) Haralambidis, J. Duncan, L.; and Tregear, G.W. *Tet. Lett.* **1987**, *28*, 5199-5202, (c) Tong, G.; Lawlor, J.M.; Tregear, G.W.; Haralambidis, J. *J. Org. Chem.* **1993**, *58*, 2223-2231 (d) Tung, C.; Rudolph, J.; and Stein, S. *Bioconj. Chem.* **1991**, *2*, 464-465 (e) Bongratz, J., Aubertin, A.; Milhaud, P.G.; and Lebleu, B. *Nuc. Acid. Res.* **1994**, *22*, 4681-4688 (f) Zhu, T.; Stein, S. *Bioconj. Chem.* **1994**, *5*, 312-315.
13. (a) Breslauer, K.J.; Remeta, D.P.; Chou, W.-Y.; Ferrante, R.; Curry, J.; Zaunczkowski, D.; Snyder, J.G.; Marky, L.A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8922, (b) Marky, L.A.; Breslauer, K.J.; *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4359. (c) Marky, L.A.; Kupke, K.J. *Biochemistry* **1989**, *28*, 9982.
14. (a) Ptashne, M. *A Genetic Switch*, Blackwell Scientific Publications and Cell Press: Palo Alto, CA, **1986**; (b) Pabo, C.O.; Sauer, R.T. *Ann. Rev. Biochem.* **1992**, *61*, 1053; (c) Marmorstein, R.; Carey, M.; Ptashne, M.; Harrison, S.C. *Nature* **1992**, *356*, 408; (d) Klemm, J.D.; Rould, M.A.; Aurora, R.; Herr, W.; Pabo, C.O. *Cell* **1994**, *77*, 21; (e) Bellon, S.F.; Rodgers, K.K.; Schatz, D.G.; Coleman, J.E.; Steitz, T.A. *Nature Struct. Biol.* **1997**, *4*, 586.
15. Colocci, N.; Distefano, M.D.; and Dervan, P.B. *J. Am Chem. Soc.* **1993**, *115*, 4468-4473.
16. (a) Sluka, J.P.; Horvath, S.J.; Bruist, M.F.; Simon, M.I.; Dervan, P.B. *Science* **1987**, *238*, 1129. (b) Sluka, J.P.; Horvath, S.J.; Glasgow, A.C.; Simon, M.I.; Dervan, P.B. *Biochemistry* **1990** *29*, 6551. (c) Feng, J.; Johnson, R.C.; Dickerson, R.E. *Science* **1994**, *263*, 348.
17. Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *Angew. Chemie.* **1996**, *35*, 1487-1489.
18. Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6778-6779.